Imprecise Statistical Methods for Accelerated Life Testing

ABDULLAH ALI H AHMADINI

Imprecise Statistical Methods for Accelerated Life Testing

Abdullah Ali H. Ahmadini

A Thesis presented for the degree of
Doctor of Philosophy

Statistics and Probability
Department of Mathematical Sciences
University of Durham
England

April 2019

Dedicated to

To Allah my God.

To my father and mother.

To my son Mohammed.

Imprecise statistical methods for accelerated life testing

Abdullah Ali H. Ahmadini

Submitted for the degree of Doctor of Philosophy
April 2019

Abstract

Accelerated Life Testing (ALT) is frequently used to obtain information on the lifespan of devices. Testing items under normal conditions can require a great deal of time and expense. To determine the reliability of devices in a shorter period of time, and with lower costs, ALT can often be used. In ALT, a unit is tested under levels of physical stress (e.g. temperature, voltage, or pressure) greater than the unit will experience under normal operating conditions. Using this method, units tend to fail more quickly, requiring statistical inference about the lifetime of the units under normal conditions via extrapolation based on an ALT model.

This thesis presents a novel method for statistical inference based on ALT data. The method quantifies uncertainty using imprecise probabilities, in particular it uses Nonparametric Predictive Inference (NPI) at the normal stress level, combining data from tests at that level with data from higher stress levels which have been transformed to the normal stress level. This has been achieved by assuming an ALT model, with the relation between different stress levels modelled by a simple parametric link function. We derive an interval for the parameter of this link function, based on the application of classical hypothesis tests and the idea that, if data from a higher stress level are transformed to the normal stress level, then these transformed data and the original data from the normal stress level should not be distinguishable. In this thesis we consider two scenarios of the methods. First, we present this approach with the assumption of Weibull failure time distributions at each stress level using the likelihood ratio test to obtain the interval for the pa-

rameter of the link function. Secondly, we present this method without an assumed parametric distribution at each stress level, and using a nonparametric hypothesis test to obtain the interval.

To illustrate the possible use of our new statistical method for ALT data, we present an application to support decisions on warranties. A warranty is a contractual commitment between consumer and producer, in which the latter provides post-sale services in case of product failure. We will consider pricing basic warranty contracts based on the information from ALT data and the use of our novel imprecise probabilistic statistical method.

Declaration

The work in this thesis is based on research carried out in the Department of Mathematical Sciences at Durham University. No part of this thesis has been submitted elsewhere for any degree or qualification.

Acknowledgements

Alhamdullilah, first and foremost. I am truly grateful to Allah my God for the countless blessings He has bestowed on me, both generally and in accomplishing this thesis especially. I would like to express my warmest gratitude and deepest appreciation to my academic supervisor, Professor Frank Coolen, for his untiring support, patience, kindness, enthusiasm, calm advice, suggestions with immense knowledge, and guidance throughout my work and during the writing of this thesis. He is a very supportive, nice person, who has very much helped and supported me during my stay in Durham. Whilst Professor Frank has been like a second father to me, he has also always listened to me and respected my situation when my wife and my son were far away. I will not forget his help and encouragement which will have a significant impact on my future. Also, I would like to thank Dr. Tahani Coolen-Maturi for her ideas and support with my computational and research questions.

I am truly grateful to my father and my mother who have been a great source of motivation, constant support, love, and prayers, and have always stood by me and been proud of me. I would like to express my sincere gratitude to my brothers and sisters for their confidence, encouragement, and helping my wife and my son for all their needs while I have been abroad.

Special thanks and warm appreciation also go to my beloved wife Reem Ahmadini for her patience, love, prayers, standing by me, her kindness, and taking care of our son Mohammed while I was abroad. Thanks to my son Mohammed.

I gratefully acknowledge the financial support from Jazan University in Saudi Arabia and the Saudi Arabian Cultural Bureau (SACB) in London to allow me to pursue my PhD studies at Durham University.

Many thanks also to Durham University for providing me with an enjoyable

academic atmosphere, for offering me great opportunities to participate and present my results at international conferences, and for the facilities that have enabled me to study comfortably.

Finally, many thanks go to my colleagues, friends, and to everyone who has assisted me, stood by me or contributed to my educational progress in any way.

Contents

Chapter 1

Introduction

The aim of accelerated life testing (ALT) is to allow researchers to estimate the useful lifespans of products or components and to inform advances in product reliability [50, 60]. The unique characteristic of ALT is to determine the reliability of products in a shorter period of time. The process of testing components' expected lifespans under normal operational conditions may be time-consuming and expensive. ALT is frequently used to determine the reliability level of devices in a shorter period of time. The basic premise of ALT is to test a unit under greater-stress conditions than the stress levels encountered under normal usage. ALT represents an important technique for gathering product failure information under a variety of conditions within a reasonable time and budget [50, 60].

In this thesis, we present a new imprecise statistical inference method for ALT data, where nonparametric predictive inferences (NPI) at normal stress levels are integrated with a parametric link function, combining data from tests at that level with data from higher stress levels which have been transformed to the normal stress level. The method includes imprecision which provides robustness with regard to the model assumptions. The imprecision leads to observations at increased stress levels being transformed into interval-valued observations at the normal stress level, where the width of an interval is larger for observations from higher stress levels.

Generally, uncertainty, within the field of statistics and probability, tends to be measured by classical probability following Kolmogorov's axioms [12]. Generalisation of Kolmogorov's axioms can bring forth possible solutions when information

and knowledge are limited or incomplete, where classical probability is considered too restrictive [12]. Such generalisation includes the use of imprecise probabilities, which is mainly characterized by using lower and upper bounds for probabilities instead of the standard theory of 'precise' or 'single valued' probability [8, 12, 70]. Further, the domain of imprecise probability, which has seen a wealth of research over the past two decades, has served as a motivation to researchers in various areas of statistics and engineering. This has resulted in a project website (The Society for Imprecise Probability: Theories and Applications - www.sipta.org) and biennial conferences [8, 48].

Over recent years, a number of methods of quantifying uncertainty and assessing reliability have appeared in the literature, which provide many benefits compared to classical probability. Thus, these methods and their respective practical applications represent a key area of cutting-edge research in this domain. For instance, interval probability [74, 75] and the theory of imprecise probabilities [72] provide techniques for reliability analysis. Regarding imprecise probabilties, Coolen [20] has investigated a range of problems related to imprecise reliability by assessing various tools designed to apply imprecise reliability to many practical applications.

The interesting work on imprecise probability has paved the road to the development of new methods of statistical inference such as nonparametric predictive inference (NPI) [9]. Many researchers have presented applications of NPI related to new approaches on different types of data.

The rest of this chapter is organized as follows. Section 1.1 provides the motivation for the work in this thesis. In Section 1.2, we provide the outline of this thesis.

1.1 Motivation

Because of the complex nature of ALT scenarios, there can be quite complicated modelling for statistical inference, which provides many challenges. Thus, the main object of this thesis is to provide a straightforward, robust model for quantifying imprecision in which can be widely used in practical applications. This model will

generate interval-based probabilities rather than exact probabilities. However, if these interval-based data fail to provide clear insight into ways to overcome practical application issues, we can modify our model's assumptions, collect more data or include specialist opinion to refine our approach. The starting point of this work is that the paper of Yin et al. [79], where it was first suggested.

Yin et al. [79] introduced an imprecise statistical method for ALT data using the power-Weibull model. In their paper, they developed the imprecision in the power-law link function by considering an interval around the parameter estimate, leading observations at stress levels other than the normal level to be transformed into intervals at the normal level [79]. Yin et al. [79] did not give an argument, other than simulation studies, for the amount of imprecision in the parameter. Building on the work by Yin et al. [79], we introduce the use of classical statistical tests between pairwise stress levels to obtain the interval for the parameter of the link function. This use of frequentist statistical tests to determine the level of imprecision is the main contribution presented in this thesis.

We obtain an interval for the parameter of the link function which is assumed at each stress level by (i) applying classical hypothesis testing between the pairwise stress levels to determine the level of imprecision, and (ii) assuming that if data from a higher stress level are then transformed to a normal stress level, then the transformed data and the original data (i.e. from the normal stress level) should, in theory, be indistinguishable. Note that each observation at the higher stress level is transformed to an interval at the normal stress level, where the interval tends to be larger if a data point was originally derived from a higher stress level.

We present this method using the assumption of Weibull failure time distributions at each stress level using the likelihood ratio test to obtain the interval; as well as without such an assumption, but using a nonparametric hypothesis test to obtain such interval for the parameter of the link function. We explore imprecision in the link function, which will allow observations of increased stress levels to be transformed to interval-valued observations at normal stress levels. We then present simulation studies to investigate the performance of the proposed method to establish appropriate links between stress test levels.

1.2 Outline of the thesis

In this thesis, we develop novel important extensions for the use of the imprecise statistical predictive inference method for accelerated life testing data. Our new approach provides robustness for the predictive inference by using statistical tests between the pairwise stress levels, to obtain the intervals for the values of the parameter of the link function. Thus, this thesis provides an innovative approach by using frequentist statistical tests to determine the level of imprecision.

This thesis is structured as follows. Chapter 2 introduces and summarizes key background concepts from the literature, applicable to the topics investigated in this thesis. It provides a brief overview of the ALT concepts and some of the life distribution functions which are applicable in reliability applications. We briefly discuss stress loading methods which can be applied in accelerated testing, including constant, step, and progressive stress loading. Chapter 2 also briefly review the statistical tests which are used in the thesis. Then, we address general notions of imprecise probability and the main idea of nonparametric predictive inference (NPI).

In Chapter 3, we present a new predictive inference method based on ALT data and the likelihood ratio test. It assumes a failure time distribution with a parametric link function at each stress level. We apply the pairwise likelihood ratio test to create an interval for the parameter of the link function. We assume a distribution model for all levels and derive an interval for the parameter of the link function using pairwise likelihood ratio tests. We use this interval to transform observations of increased stress levels to interval-valued observations at the normal stress level.

Also, we use NPI at the normal stress level to achieve predictive inference on the failure time of a particular unit operating under normal stress levels using the original data at the normal stress level and interval-valued data transformed from higher stress levels. We investigate the performance of our method via simulations. A paper presenting the results in this chapter has been submitted for publication [6]. Also, the results of this chapter have been presented at several seminars and conferences, and we have also published short papers in related conference proceedings. Parts of Chapter 3 have been presented at the 2017 Research Students' Conference in Probability and Statistics at the University of Durham, and at the 10th

IMA International Conference on Modelling in Industrial Maintenance and Reliability (Manchester, 2018), and a related short paper was published in the conference proceedings [3]. The results of Chapter 3 have also been presented at the 3rd International Conference on System Reliability and Safety (Barcelona, 2018), and a related short paper was published in the conference proceedings [5].

Chapter 4, presents a similar method to in Chapter 3 but without the assumption of a parametric distribution for the failure times at each stress level. The method is largely nonparametric, with a basic parametric function to link different stress levels. We use the log-rank test to provide adequate imprecision for the link function parameter, and conduct simulation studies to investigate the performance of our proposed approach. A paper based on this chapter has been submitted for publication [26]. This was presented at the Soft Methods in Probability and Statistics (SMPS) conference (Compigne, 2018), and a short paper was published in the conference proceedings [4].

In Chapter 5 we explore the use of our new methods, as presented in Chapters 3 and 4, for decision making with regard to warranties. This work will be presented at the ESREL conference (Hannover, 2019), a short paper has been submited for the conference proceedings [7].

Chapter 2

Background

This chapter provides an overview of concepts from the literature relevant to the topics investigated in this thesis. Section 2.1 present an overview of reliability testing. Section 2.2 provides a brief introduction to the main concepts of accelerated life testing (ALT). We present a brief overview of ALT and discuss some of the commonly used failure distributions and ALT link functions between stress levels. In Section 2.3, we briefly review basic statistical tests used in this thesis, namely the likelihood ratio test and the log-rank test. The likelihood ratio test and log-rank test will be employed in Chapters 3 and 4, respectively. Finally, we give an overview of the main aspects of NPI in Section 2.4.

2.1 Reliability testing

Reliability analysis techniques play an important role in the engineering field and the manufacturing industry in terms of product design and development processes. However, product reliability analysis is renowned for being a time-consuming process. Reliability analysis is used to model a product's time to failure in most applications, and falls into two categories: complete (where all failure data are made available) and censored (where some data are omitted) [41, 60].

In terms of complete data sets, exact test-unit failure times are used, as these data are both measured and known [41, 60]. However, not all (or indeed any) units may fail during a single testing cycle; such data are called censored data [41, 60].

Thus, because of these two different conditions under which censoring may occur, such data can be further divided into time-censored data (type I-censored) and failure-censored data (type II-censored). Type I-censored data is typically obtained when the censoring time is pre-set, and the number of failures is expressed as a random variable. In terms of type II-censored data, testing is completed following a pre-specified number of failures [41, 60]. In this case, the time period in which a specified number of failures will occur is expressed as a random variable [60].

Reliability analysis for any new product at time $t = 0$ and consequent failure at time T employs two terms: the reliability function $R(t)$ and the hazard function $h(t)$, where T represents the reliability period for the product in question [41]. The reliability function $R(t)$ represents the probability that the product will survive until time t. It is defined as

$$R(t) = P(T > t) = 1 - F(t). \tag{2.1.1}$$

Note that the reliability function $R(t)$ is a monotonically decreasing function:

$$R(t_1) \geq R(t_2) \; \forall \; t_1 < t_2. \tag{2.1.2}$$

where $R(0) = P(T > 0) = 1$.

A concise definition of the hazard rate is the immediate potential per unit time for an event to occur based on the assumption that an individual unit has withstood testing up until time t [41]. This results from the determination of the survival function, because when t increases, the survival rate will never increase [41]. The hazard function $h(t)$ is defined as

$$h(t) = \lim_{\triangle t \to 0} \frac{P(t \leq T < t + \triangle t | T \geq t)}{\triangle t}. \tag{2.1.3}$$

Maximum likelihood estimation (MLE) is a well-known and widely used statistical method used to estimate the parameters of probability distributions. Likelihood functions are formed from observed data and the chosen distribution for these data, which are assumed to be independent and identically distributed [1, 60, 62]. The likelihood function has been discussed a great deal in the literature, and applied to both failure data and censored data [60].

The method of MLE can be used for estimating the unknown parameters of lifetime distributions. The likelihood function for a chosen failure time distribution with parameter θ is defined as

$$L(\theta; t) = \prod_{i=1}^{n} f(t_i; \theta) \prod_{j=1}^{u} S(c_j; \theta),$$

where $f(.)$ is the probability density function, t_i is the observed failure times, and $i = 1, ..., n$. $S(.)$ is the survival function, c_j is the censored times, and $j = 1, ..., u$.

When assessing the reliability of a new product, the foremost challenge is to determine reliability in a relatively short period of time. Typically, we use lifespan to assess a product, system or component. In traditional testing methods, researchers perform these life tests only under normal operating conditions, to directly determine a product's failure time distribution and the parameters associated with failure. However, this method of obtaining life data at least in its current form is often not viable in present-day industrial testing because current products are typically more reliable than in the past. Further, there are also greater time pressures nowadays where new products are required to be launch-ready very soon after they have been designed. This issue has provided the motivation to create a new lifespan testing method, namely accelerated life testing (ALT), which can provide meaningful product failure data in a short space of time. ALT will be introduced in the next section.

2.2 Accelerated life testing

ALT is frequently used to gather information on the expected lifespans of a range of devices; it represents an efficient way of testing, if testing under normal conditions requires a great deal of time and expense. Thus, ALT enables testing the reliability of devices in shorter time periods and at lower cost. ALT involves testing a particular unit under varying degrees of physical stress (e.g. temperature, voltage, or pressure), greater than normal operating conditions. This approach causes devices to tend to fail more quickly, enabling testers to estimate devices' expected lifetimes under normal operating conditions via extrapolation using an ALT model.

In recent decades, various methods for analyzing ALT data and a variety of methods for assessing the reliability with different ALT scenarios have been introduced. An excellent introduction to ALT was given by Nelson [60]. This was followed by a large literature on ALT. In what follows, we we briefly introduce some designs of ALT tests, life time distribution, and acceleration models [60].

A wide range of test designs can be applied in accelerated testing, the main ones are: constant-, step- and progressive-stress testing [60]. The basis for classification of the stress loading is the dependency of the stress with respect to time t.

Constant stress loading (which is time-independent) is the most common form of stress loading, where products operate at constant stresses. In constant-stress testing, units are tested at a specific stress level $K_0, K_1, ..., K_m$ until they either fail at that particular stress level or testing is ended for another reason [60].

Constant stress loading has several advantages [60]. When the experiment is designed, extrapolation at the constant use level condition is more accurate than time-varying loading. This is because all units under constant stress are maintained at the same voltage, temperature, or pressure for a given stress level, which makes it easier and simpler to run each test. However, the failure of units under constant stress usually takes a long time, especially at normal levels of stress, which makes it more time-consuming than testing under time-dependent stress levels [60]. We only consider the constant stress loading in this thesis, because it is the simplest stress loading for which to develop new methods. Note that we do not consider design of ALT tests in this thesis, we just assume the data are given.

When developing an acceleration model, to determine the life-stress relationship in ALT, two important components need to be considered. First, we need to describe the relationship between failure time and stress level. In particular, they should reflect the way that various levels of stress affect how quickly the failure mechanisms occur, and how stress affects the overall lifetime of a given device. In real life if one increases the stress, the failure occurs faster, and it may lead to different modes of failure that would need to be modelled. However, in this thesis, we assume that the modes of failure and their relative frequency are unchanged by stress and we only study the statistical development of the methods. The model should also include a

lifetime distribution for each fixed stress level (e.g. Weibull or log-normal) [50, 60].

Variations among products, in terms of performance, lifetime and quality, can be modelled by a statistical distribution. Life time distribution functions describe distributions for products which are applicable to failure data analysis and reliability estimations [60]. Life time distributions can be used to model increasing or decreasing failure rate, and analyse the differences between the products, and to model reliability [50, 60]. One of the most commonly used and widely applicable distributions used for life data analysis is the Weibull distribution, which we use in this thesis.

The Weibull distribution [60] is a popular way to model reliability, and is thus used for examining product lifetime. It is also used to describe the failure properties of electronic components and the breaking strengths of the material(s) they are composed of [60]. Breaking strengths (electronical or mechanical) in accelerated test form one example [60]. The Weibull distribution is commonly utilised in the accelerated testing of roller bearings [50, 60]. The Weibull distribution [60], is given by

$$f(t) = \frac{\beta}{\alpha}(\frac{t}{\alpha})^{\beta-1} \exp\Big[-(\frac{t}{\alpha})^{\beta}\Big],$$

where $t > 0$. The unknown parameters of the Weibull distribution are the shape parameter $\beta > 0$, and the scale parameter $\alpha > 0$, and its survival function is

$$S(t) = \exp\Big[-\Big(\frac{t}{\alpha}\Big)^{\beta}\Big].$$

The hazard rate of the Weibull distribution is

$$h(t) = f(t)/S(t) = \Big(\frac{\beta}{\alpha}\Big)\Big(\frac{t}{\alpha}\Big)^{\beta-1},$$

where the $h(t)$ is the hazard rate at age t. In short, the hazard rate is used to discribe whether the failure rate of a product increases or decreases with product age [60]. For $\beta > 1$, the hazard rate is a strictly increasing function with respect to time and when $\beta < 1$, the Weibull distribution shows a decreasing hazard rate.

For $\beta = 1$, the Weibull distribution has a constant hazard rate and is equal to the Exponential distribution [50, 60]

Some other distributions used for life data analysis are the Exponential, Log-normal, and Gamma distributions, but we do not consider these in this thesis. Note that they can be used instead of the Weibull distribution in Chapter 3, as long as these would be combined with a link function such that it enables transformation as has been done in Chapter 3.

Determining the life-stress relationship in ALT usually involves using either a physical or an empirical acceleration model [68]. Physical or chemical theory based models are attractive for some failure mechanisms [68]. These types of models describe the process that causes device failure over a range of stress levels, in order to facilitate extrapolation to the normal stress level [68]. Failure mechanisms and accelerating variables usually have a complicated relationship [68]. In most situations, a simple model will not be sufficient to describe the failure process and its causes [50, 60, 68].

In this thesis, we do not aim at complex models but we are exploring simple models combined with imprecision to get robustness around the simple models. We have not found such an ALT method combined with imprecise probability methods in the literature. Statistical inference for ALT data tends to focus on parameter estimates. We comment on this in Chapter 6.

ALT is used for extrapolating information to the normal stress level about failure time distributions, and the reliability of a given product [60, 68]. Acceleration factors are used to determine the failure time at a particular stress level, which can subsequently be used to predict the failure time at different levels of operating stress [60,68]. This is known as an acceleration model. While statistical distributions identify the lifetimes of a particular type of unit at each stress level and distribution for items, the acceleration model derives the scale parameter, or the shape parameter of a life distribution, as a function of the applied stress [50, 60, 68]. Some of the most important and commonly used acceleration models are the Arrhenius, Eyring,

and power-law models, which we use in this thesis [17, 34, 50, 60, 71].

The Arrhenius life-stress relationship model is one of the most widely used for testing when the stimulus or accelerated variable is thermal stress [60, 68]. It is often an appropriate choice if a unit's failure mechanism is driven by temperature [60, 68]. The physics-based Arrhenius law for chemical reaction rates explains that as temperature increases, this induces increased levels of atomic movement, and so the processes that cause failure speed up [60, 68]. Therefore, the Arrhenius model explains the reaction rate α of a unit as a function of applied temperature [17, 34, 50, 60, 71]. This model can be expressed as

$$\alpha = A \exp\left(\frac{-E_A}{k_B.K}\right), \tag{2.2.1}$$

where α is the reaction rate, A is a constant characteristic of the united failure mechanism and test condition, E_A is the activation energy in electron-volts, k_B is the Boltzmann's constant (8.6171×10^{-5} electron-volts per $^\circ C$), and K is the absolute temperature (Kelvin) for the Arrhenius relationship (accelerated stress).

By taking the natural logarithm of (2.2.1), it yields the linear relationship

$$\ln(\alpha) = \gamma_0 + \frac{\gamma}{K}, \tag{2.2.2}$$

where $\gamma = \frac{E_A}{k_B}$, $\gamma_0 = \ln(A)$.

The natural logarithm of the scale parameter at the normal stress level is

$$\ln(\alpha_0) = \gamma_0 + \frac{\gamma}{K_0},$$

and the natural logarithm of the scale parameter at the higher stress level is

$$\ln(\alpha_i) = \gamma_0 + \frac{\gamma}{K_i}.$$

The Weibull distributions for different stress levels are assumed to have different scale parameters $\alpha_i > 0$ for level i. The Arrhenius link function for scale parameters is

$$\alpha_i = \alpha_0 \exp\Big(\frac{\gamma}{K_i} - \frac{\gamma}{K_0}\Big). \tag{2.2.3}$$

Where using this model with temperature as stress levels, K_0 is the normal temperature (Kelvin) at stress level 0, K_i is the higher temperature (Kelvin) at stress level i, and $\gamma > 0$ is the parameter of the Arrhenius link function model.

In this thesis, we use the Arrhenius model under the Weibull distribution to link the scale parameter at different failure times. The Arrhenius model, however, cannot be applied to all temperature issues [60, 68]. In some cases, it is sufficient over only a limited temperature range [60, 68]. According to Nelson [60], in particular real-world application (e.g. motor insulation), the Arrhenius model may not fit the data well [60].

The Eyring model provides an alternative to the Arrhenius model. It also uses temperature as the accelerating variable, and is based on quantum mechanics [50,60, 71]. In this model, the relationship between mean time to failure α and temperature K is defined by

$$\alpha = \frac{A}{K} \exp(\frac{\lambda}{K}),$$

where $\lambda = \frac{E_A}{k_B}$. A and $\lambda > 0$ are constant characteristic of the united failure mechanism and test condition, k_B is the Boltzmann's constant, (8.6171×10^{-5} electron-volts per $^\circ C$), and K is the absolute temperature (Kelvin).

The scale parameter of a unit at the normal stress level is

$$\alpha_0 = \frac{A}{K_0} \exp(\frac{\lambda}{K_0}),$$

and the scale parameter at the higher stress level is

$$\alpha_i = \frac{A}{K_i} \exp(\frac{\lambda}{K_i}). \tag{2.2.4}$$

The Eyring link function for the Weibull scale parameters is

$$\alpha_i = \alpha_0 \times (K_0/K_i) \times \exp\Big[(\lambda/K_i - \lambda/K_0)\Big]$$

Where using this model with temperatures as stress levels, K_0 is the normal temperature (Kelvin) at stress level 0, K_i is the higher temperature (Kelvin) at stress level i, and $\lambda > 0$ is the parameter of the Eyring link function model.

The Eyring model can be applied to testing capacitors, electro-migration failure, and solid rupturing [60]. It should be noted that both the Eyring and Arrhenius models derive similar results in many applications, and both fit failure times related to temperature [60].

The power-law model is commonly used for testing when the stimulus or accelerated variable is voltage stress. It is often an appropriate choice if a unit's failure mechanism is driven by voltage to analyse lifetime data as a function of the ALT model [60]. In the power-law model, the relationship between scale parameter α and voltage K is

$$\alpha = \frac{1}{(C.K)^\gamma},$$

where C and γ are parameters representing characteristics of the product and test method, and K represents the stress level in terms of voltage. The scale parameter at normal stress level is

$$\alpha_0 = \frac{1}{(C.K_0)^\gamma},$$

and the scale parameter at the higher stress level is

$$\alpha_i = \frac{1}{(C.K_i)^\gamma}.$$

The power-law link function of scale parameter α_i should be used for establishing a connection between different stress levels i, and is assumed to satisfy the function

$$\alpha_i = \alpha_0 \Big(\frac{K_0}{K_i}\Big)^\gamma.$$

Where using this model with different voltages as the stress levels, K_0 is the normal voltage at stress level zero, K_i is the higher voltage at stress levels, and γ is the parameter of the power-law link function model.

Some common applications of the power-law model are for testing electrical insulation, dielectrics in voltage endurance evaluations and materials, bearings, and electronic devices, to determine their useful lifespans and reliability [60]. For more details, see [50, 60].

As examples of applications, Fan et al. [30] presented the maximum likelihood estimation (MLE) and Bayesian inference on all parameters of ALT models under an exponential distribution, with a linear link function between the failure rate and the stress variables under the Box-Cox transformation. Fard and Li [31] presented an optimal step stress to obtain the optimal hold time at which the stress level is changed for step stress ALT design for reliability prediction. They assumed a Weibull distribution for the failure time at any constant stress level, and the scale parameter of the Weibull distribution was assumed to be a log-linear function of the stress level. Elsayed and Zhang [29] proposed an optimal multiple-stress-type ALT plan using a proportional hazards model to obtain failure time data rapidly in a short period of time. Sha and Pan [67] introduced step-stress ALT with Bayesian analysis for the Weibull proportional hazard model. Nasir and Pan [59] assumed a Bayesian optimal design criterion and presented acceleration model selection ALT studies, while Han [36] conducted research into temporally and financially constrained constant-stress and step-stress ALT.

Whilst ALT models typically consider failure times as the events of interest, there have also been important contributions with more detailed modelling, in particular exploiting methods to mathematically model degradation processes. For example, Liao and Tseng [44] proposed an optimal design for step-stress accelerated degradation tests with the degradation process modelled as a stochastic diffusion process. Pan et al. [64] presented a bivariate constant stress accelerated degradation model and related inference. They assumed a device which has two performance characteristics which are modelled by a Wiener process, to determine the relibility of high

quality devices with a time scale transformation, and the Frank copula is used to model dependence of the two performance characteristics. Duan and Wang [28] proposed a bivariate constant stress accelerated degradation model with inference based on the inverse Gaussian process. It is important to note here that the modelling of degradation processes does require much information about the engineering process and physical properties of the equipment, which may come from detailed measurements of the process or expert judgements. While this is an important development for real world accelerated life testing, we do not address such approaches further in this thesis and only assume information about the failure times to be present.

2.3 Basic Statistical Methods

Comparing the survival function or the probability distribution of two independent groups, possibly including right-censored observations often requires classical statistical tests. In this section we present a brief overview of the statistical tests used in Chapters 3 and 4 of this thesis. Subsections 2.3.1 and 2.3.2, introduce the likelihood ratio test and the log-rank test, respectively. In Chapters 3 and 4 of this thesis, we will use these test statistics for the pairwise stress levels to find the interval of values of the parameter of the link function for which we do not reject the null hypothesis of two or more groups of failure data, possibly including right censored data, coming from the same underlying distribution. Indeed, for our methods in Chapters 3 and 4, other statistical tests could be used, and of course if we use different tests these may lead to slightly different results.

2.3.1 Likelihood ratio test

Hypothesis testing is one of the main methods in statistical inference and its applications. Testing equality of the probability distribution of two or more independent groups often involves parametric statistical tests. A general classical hypothesis test that can be used to test equality of the survival distributions, is the likelihood ratio test [42, 63]. The likelihood ratio test is used to compare two independent failure data groups, possibly including right-censored observations (e.g. resulting from two

ALT test), in assumed parametric models.

The probability density function of a statistical distribution is assumed to describe the failure time at a fixed stress level, and its parameters are therefore maximized in the likelihood ratio test based on the idea of hypothesis testing for which we do not reject the null hypothesis of two groups of data, coming from the same underlying distribution. To obtain the likelihood ratio test, we need to compute the difference between the log likelihood of the alternative hypothesis $\hat{L}_1$ and the likelihood of the null hypothesis $\hat{L}_0$ [42].

Suppose that L_1 and L_0 be the maximized log-likelihood under the alternative and the null hypothesis respectively. Then the likelihood ratio (LR) test statistic is $LR = 2(\hat{L}_1 - \hat{L}_0)$, which under the null hypothesis, follows a χ^2 distribution with degree of freedom which is equal to the difference in the number of parameters of each of the two models. The likelihood ratio test is discussed in more detail in all good introductory statistics books (see e.g. [42]).

2.3.2 Log-rank test

The general advantege of nonparametric tests is that they are "distribution-free". The use of nonparametric statistical tests often involve comparison of two groups of data, e.g. data resulting from an experiment with two different groups in which units are studied in accelerated test (e.g. temperature, voltage, humidity, pressure). Testing equality of the survival distribution of two independent groups often involves nonparametric statistical tests. There are several nonparametric test procedures that can be used to test equality of the survival distributions. One popular nonparametric test for equality of the survival distributions of two groups is the log-rank test [46, 65]. The log-rank test is a nonparametric test and one of the most widely used for comparing the survival distribution of $m \geq 2$ groups [65]. Sometimes, the log-rank test is called the Mantel-Cox test [33, 47, 66]. It compares the observed numbers of failures for two groups with the expected numbers of failures [66].

Conditionally on the number at risk in the groups [41], the log-rank test statistics can be determined by using observed and expected values for each group, and comparing the hazard rates between the two groups throughout [65, 66]. The log-rank

test obtains the expected number of failure for expected number of failure for group $i \in \{1, 2\}$ as:

$$e_{ij} = \left(\frac{n_{i,j}}{n_{1,j} + n_{2,j}}\right) \times (m_{1,j} + m_{1,j}). \tag{2.3.1}$$

Suppose that there are two groups consisting of n_i $(i = 1, 2, ..., m)$ individuals. Individuals in each group have either a failure or right-censoring time. Let $0 < t_{(1)} < t_{(2)} < ... < t_{(k)} < \infty$ denoted the failure time, for ease of notation let $t_0 = 0$ and $t_{k+1} = \infty$. For ease of presentation, we assume that no ties occur among the observed value in the combined group. Suppose that $m_{1,j}$ denoted the number of failures occur at t_j in group one, $m_{2,j}$ denoted the number of failures occur at t_j in group two. Let $n_{i,j}$ be the number of units at risk just prior to time $t_{(j)}$ $(j = 1, 2, ..., k)$.

The log-rank test statistics is obtained by

$$\textit{log-rank} = \frac{(O_2 - E_2)^2}{\textit{Var}(O_2 - E_2)}, \tag{2.3.2}$$

where $O_2 - E_2 = \sum_{j=1}^{k}(m_{2,j} - e_{2,j})$.

Therefore, the variance of $\textit{Var}(O_i - E_i)$ is defined by

$$\textit{Var}(O_i - E_i) = \sum_{j=1}^{k} \frac{n_{1,j}n_{2,j}(m_{1,j} + m_{2,j})(n_{1,j} + n_{2,j} - m_{1,j} - m_{2,j})}{(n_{1,j} + n_{2,j})^2(n_{1,j} + n_{2,j} - 1)}. \tag{2.3.3}$$

The test statistics approximately follows a χ^2 distribution and a degree of freedom which is equal to the difference in the number of groups [42]. Alternative tests statistics can be used instead of the log-rank test statistics to test equality of survival functions [33].

2.4 Nonparametric predictive inference

In terms of imprecise probability, classical probability is generalised as uncertainties relating to events are measured not with single numbers but intervals [20]. For example, in classical probability, event A would be ascribed a single number $P(A) \in$

$[0,1]$, where P is a probability measured by classical probability which operates within the bounds of Kolmogorovs axioms. In terms of quantifying uncertainty, imprecise probability as a statistical concept was first proposed by Boole [15] in 1854 and has had a long history since Hampel [35]. Over the past few years, a number of alternative approaches to quantifying uncertainty have been proposed including interval probability theory, introduced by Walley [72] and Weichselberger [74] which proposes that probabilities have lower and upper range of probabilities i.e. $[\underline{P}(A), \overline{P}(A)]$ respectively, with $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$ while classical probability theory includes a complete lack of data about the event in question i.e. $\underline{P}(A) = 0$ and $\overline{P}(A) = 1$. Here, $\underline{P}(A)$ represents the lower probability of event A, and $\overline{P}(A)$ represents the upper probability for event A. The imprecision for event A is $\triangle(A) = \overline{P}(A) - \underline{P}(A)$. For the lifetime failure observation, Coolen [18] presented the lower and upper predictive probabilities. These probabilities form part of a wider statistical methodology called Nonparametric Predictive Inference (NPI), which will be addressed briefly in this section.

We review Nonparametric Predictive Inference (NPI), closely following [8,48,58]. NPI is a statistical method which provides lower and upper survival functions for a future observation based on past data using imprecise probability [10, 24]. Hill [37] proposed an assumption which gives direct conditional probabilities for a future random quantity which depend on the values of related random qualities [9, 23, 24]. It proposes that the rank of a future observation among the values already observed will be equally likely to have each possible value $1, ..., n+1$ [48,58]. Suppose that $X_1, X_2, ..., X_n, X_{n+1}$ represent exchangeable and continuous real-valued possible random quanities, then the ranked observed values of $X_1, X_2, ..., X_n$ can be denoted by $x_{(1)} < x_{(2)} < ... < x_{(n)}$. Let $x_{(0)} = 0$ and $x_{(n+1)} = \infty$. The assumption $A_{(n)}$ is

$$P(X_{n+1} \in (x_{(j-1)}, x_{(j)})) = 1/(n+1)$$

for all $j = 1, 2, ..., n+1$. Here, no tied observations are included for convenience, any tied values can be dealt with by assuming that tied observations differ by a small amount which tends to zero [38, 58].

Inferences which are based on $A_{(n)}$ are nonparametric and predictive [58]. They

can be considered suitable if there is hardly any knowledge about the random quantity of interest, except for the n observations, or if one does not want to use any such further information [58]. The $A_{(n)}$ assumption is not sufficient to derive precise probabilities for many events of interest [58]. However, this approach does yield optimal bounds for probabilities through the 'fundamental theorem of probability' [27], which are lower and upper probabilities in the imprecise probability theory [9, 10].

The lower and upper probabilities for event A are denoted by $\underline{P}(A)$ and $\overline{P}(A)$, respectivily. These are open to interpretation in various ways [10]. For instance, $\underline{P}(A)$ can be assumed to be the supremum buying price for a gamble on event A, such that if A occurs then 1 is paid, if not then 0 is paid. This can also simply be interpreted as the maximum lower bound for the probability of A, which derives from the assumptions made. Similarly, $\overline{P}(A)$ can be interpreted as the minimum selling price for the gamble on A, or the minimum upper bound based on the assumptions made. We have $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$, and the conjugacy property $\underline{P}(A) = 1 - \overline{P}(A^c)$ where, A^c is the complimentary event of A [9, 10].

The NPI lower and upper survival functions for a future observation X_{n+1} are

$$\underline{S}_{X_{n+1}}(t) = \frac{n-j}{n+1}, \text{ for } t \in (x_j, x_{j+1}), \ j = 0, ..., n. \tag{2.4.1}$$

$$\overline{S}_{X_{n+1}}(t) = \frac{n+1-j}{n+1}, \text{ for } t \in (x_j, x_{j+1}), \ j = 0, ..., n. \tag{2.4.2}$$

Events of interest in reliability and survival analysis are usually failure [19,25,79]. However, such data are often right-censoring, which means for some units, they are only known that the events have not yet failed at specific time of observations [79]. The $A_{(n)}$ assumption cannot handle right-censored observation, and demands fully observed data [79]. Coolen and Yan [21] presented a generalization of $A_{(n)}$, called rc-$A_{(n)}$, which is suitable for right-censored data with nonparametric predictive inference [79]. Moreover, rc-$A_{(n)}$ uses the additional assumption that, at the moment of censoring, the residual lifetime to failure of a right-censored unit is exchangeable with the residual lifetime to failure of all other unites that have not yet failed or been censored [48, 79].

Suppose that n units consisting of u units failed during the experiment at different times $x_{(1)} < x_{(2)} < ... < x_{(u)}$. Also, $n-u$ right-censored units $c_{(1)} < c_{(2)} < ... < c_{(n-u)}$ and we set $x_{(0)} = -\infty$ and $x_{(u+1)} = \infty$. Suppose further that there are s_i right-censored observations in the interval (x_i, x_{i+1}), denoted by $c_1^i < c_2^i < ... < c_{s_i}^i$, so $\sum_{n=0}^{u} s_i = n - u$. Let d_j^i be the number of event at the failure or censoring time, with $d_0^i = x_i$ and $d_j^i = c_j^i$ for $i = 1, 2, ..., u$ and $j = 1, 2, ..., s_i$, and set $\tilde{n}_{c_u}$ and $\tilde{n}_{d_j^i}$ number of subjects in the risk set just before to time c_u and d_j^i, respectively, corresponding to the definition $\tilde{n}_0 = n + 1$ [48, 79].

Coolen and Yan [21] presented the lower and upper survival functions of the NPI for the lifetime failure observation, so the NPI lower survival function $\underline{S}_{X_{n+1}}(t)$ and the corresponding NPI upper survival function $\overline{S}_{X_{n+1}}(t)$, in case of right-censored data, respectively [48, 49]. Also, according to the previous notation, let $d_{s_i+1}^i = d_0^{i+1} = x_{i+1}$ for $i = 1, 2, ..., u-1$. Therefore, for $t \in [d_j^i, d_{j+1}^i)$ with $i = 1, 2, ..., u$ and $j = 1, 2, ..., s_i$, and for $t \in [x_i, x_{i+1})$ with $i = 1, 2, ..., u$. The lower and upper survival functions can expressed as [48, 49]

$$\underline{S}_{X_{n+1}}(t) = \frac{1}{n+1}\tilde{n}_{d_j^i} \prod_{r:c_r \leq d_j^i} \left(\frac{\tilde{n}_{c_r} + 1}{\tilde{n}_{c_r}} \right) \tag{2.4.3}$$

$$\overline{S}_{X_{n+1}}(t) = \frac{1}{n+1}\tilde{n}_{x_i} \prod_{r:c_r \leq x_i} \left(\frac{\tilde{n}_{c_r} + 1}{\tilde{n}_{c_r}} \right). \tag{2.4.4}$$

Equations (2.4.1) - (2.4.4) play an important role in imprecise probability theory [27]. The imprecision, the difference between the upper and lower survival functions, reflects the amount of information in the data. This imprecision is non-zero because of the limited inferential assumptions made, and reflects the amount of information in the data as explained previously. Note that, where we introduce NPI lower and upper survival functions above in the case of right-censored data, the lower survival function decreases at each observation, and the upper survival function only at the observed failure times. This beautifully illustrates an attractive informal interpretation of lower and upper probabilities: the lower probability for event A reflects the information in support of event A, the upper probability (actually, $1 - \overline{P}(A)$) reflects the information against event A - hence also in support

of the complementary event, which agrees with the conjugacy property. A failure observation is clearly information against survival, therefore further reducing information supporting survival, so both lower and upper survival functions decrease. A right-censored observation reduces information in support of survival past that point (smaller number of items known to survive), but it does not provide more support against survival as the item did not fail.

Chapter 3

Statistical inference based on the likelihood ratio test

3.1 Introduction

The development of ALT statistical modelling is typically complex and brings challenges for modelling and statistical inference. The literature tends towards creating ever more complex models [28, 29, 32, 45]. While these may be of theoretical interest, we believe that practical application of ALT can be best served by widely applicable, easy-to-use statistical methods which feature in-built robustness. In this chapter, we develop a new likelihood ratio test based method for analysing ALT data with imprecise probabilities, where nonparametric predictive inference (NPI) at the normal stress level is integrated with a parametric Arrhenius-Weibull model. This use of a frequentist statistical test to determine the level of imprecision is the main novelty in this chapter. This new method consists of two steps. First, we assume the Arrhenius link function for all levels and we derive an interval for the parameter of the Arrhenius link function by pairwise likelihood ratio tests. This interval of parameter values enables observations of increased stress levels to be transformed to interval-valued observations at the normal stress level, where we assume that we have data at the normal stress level. Secondly, we use NPI at the normal stress level for predictive inference on the failure time of a future unit operating under normal stress, using the original data at the normal stress level and interval-valued data

transformed from higher stress levels.

This chapter is organized as follows. Section 3.2, we outline the model description. In Section 3.3, our novel method of imprecise statistical inference is introduced. In Section 3.4 we illustrate our proposed method in three examples. Section 3.5 presents results of simulation studies that investigate the performance of the proposed method using the Arrhenius link function. Section 3.6 illustrates our method using the power-law link function and illustrate the method using the power-law link function in two examples. Section 3.7 presents results of simulation studies that investigate the performance of the proposed method, where the data are simulated using the assumed power-Weibull model with different shape parameters β_i at different stress levels. Section 3.8 presents some concluding remarks.

3.2 The model

In this chapter, we consider the Arrhenius model and a Weibull lifetime distribution for a constant-stress ALT. The Arrhenius model is based on physical or chemical theory, and is often an appropriate model to use when the failure mechanism is driven by temperature [60]. The Weibull distribution is often suitable for examining component, system or product life. All these models have been introduced briefly in Section 2.2. The Arrhenius-Weibull model is adopted to the current research on imprecise statistical approaches to establish the use of imprecision in modelling the nature of the relationship between stress levels and unit failure rates. The main issue here is how to extrapolate the failure data from units tested at higher-than-normal stress levels to units operating at the normal stress level [79]. Note that we also consider the power-law and the Weibull lifetime distribution for a constant-stress ALT in Section 3.6. In that section, we apply our method with the different scale parameters α_i and different shape parameters β_i for the Weibull distributions for the different stress levels i. We comments on this assumption in Section 3.6.

The model at each stress level (the two-parameter Weibull distribution) is

$$f(t) = \frac{\beta}{\alpha_i}(\frac{t}{\alpha_i})^{\beta-1} \exp\Big[- (\frac{t}{\alpha_i})^{\beta}\Big]. \tag{3.2.1}$$

The unknown parameters of the Weibull distribution are the shape parameter β, and the scale parameters α_i at stress level i, where $\alpha_i > 0$ and $\beta > 0$, and its survival function is

$$P(T > t) = \exp\Big[- \Big(\frac{t}{\alpha_i}\Big)^{\beta}\Big].$$

The Arrhenius-Weibull model is specified as follows. K_0 represents the stress at the normal level. There are $m \geq 1$ increased stress levels, with stress K_i at level $i \in \{1, ..., m\}$, we assume that K_i increases as a function of i. In this chapter, the Weibull distributions for different stress levels are assumed to have different scale parameters $\alpha_i > 0$ for level i, but the same shape parameter β. In Section 3.6 we also consider the generalization with different shape parameters β_i for level i for these Weibull distributions combined with power-law link function. The Arrhenius link function for the scale parameters is

$$\alpha_i = \alpha_0 \exp\Big(\frac{\gamma}{K_i} - \frac{\gamma}{K_0}\Big). \tag{3.2.2}$$

Using this model with varying temperatures (in Kelvin) as stress levels, K_0 is the normal temperature at stress level 0, K_i is the higher temperature at stress level i, and $\gamma > 0$ is the parameter of the Arrhenius link function model.

Using this link function model, an observation t^i at the stress level i, subject to stress K_i, can be transformed to stress level 0. For fixed γ the transformed observation denoted by $t^{i\to 0}(\gamma)$ from level i to level 0 is given by the equation

$$t^{i\to 0}(\gamma) = t^i \exp\left(\frac{\gamma}{K_0} - \frac{\gamma}{K_i}\right). \tag{3.2.3}$$

Now, we define the model through the probability density function as:

$$f(t^i; \alpha, \beta, \gamma, \underline{K}) = \frac{\beta}{\alpha_i}\left(\frac{t^{i\to 0}(\gamma)}{\alpha_i}\right)^{\beta-1} \exp\left(-\left(\frac{t^{i\to 0}(\gamma)}{\alpha_i}\right)^{\beta}\right), \tag{3.2.4}$$

where the Arrhenius link function for scale parameters α_i should be identified to establish a connection between the different stress levels i.

Consider $\underset{\sim}{t} = \{t_1^0, ...t_{n_0}^0, t_1^1, ...t_{n_1}^1, ..., t_1^m, ...t_{n_m}^m\}$ and $\underline{K} = \{K_0, ..., K_m\}$, where $\underset{\sim}{t}$ is the data and it should be used for the whole data if denoted including the transformations, and $\underline{K}$ denotes the stress applied to each level.

The likelihood function is defined as

$$L(\underset{\sim}{t}; \alpha, \beta, \gamma, \underline{K}) = \prod_{j=0}^{m} \prod_{i=1}^{n_j} f\left(t_i^j; \alpha, \beta, \gamma, \underline{K}\right).$$

So there are in total three parameters that need to be estimated to fit the complete model, α_0, β, and γ. In this chapter, we will apply the pairwise likelihood ratio test to create an interval for the parameter γ of the link function, as presented in Section 3.3. We assume the same β for all stress levels, so differences between stress levels are only modelled through the α_i, and hence in the model through the α_0 and γ parameters of the link function between different stress levels. For ease of presentation, we assume that there are no right-censored observations and that there are failure observations at the normal stress level. We briefly comment on these assumption in Section 3.8.

3.3 ALT inference using likelihood ratio tests

To investigate equality of two independent failure data groups, possibly including right-censored observations, the likelihood ratio test can be used [63]. This is a popular statistical test that can be applied to investigate equality of the probability distribution of two independent groups, which has been briefly introduced in Section 2.3.1 [2, 60].

In this section we present new predictive inference based on ALT data and the likelihood ratio test. We use NPI at the normal stress level, with the fully parametric model used in our new statistical method analysing data from ALT. The use of NPI here provides lower and upper survival functions for a future observation at the normal stress level, based on all failure data.

This new statistical method for data in ALT divides into two steps. First, the basic Arrhenius-Weibull model is adopted [60], and the pairwise likelihood ratio test is used between the stress levels K_i and stress level K_0, to obtain the intervals for the

parameter γ for which we do not reject the null hypothesis that the data transformed from stress level i to normal stress level 0, and the original data obtained at the normal stress level 0, come from the same Weibull distribution. The hypothesis test we use in this chapter is

$$\begin{cases} H_0 : \gamma = \gamma\prime \\ H_1 : \gamma \neq \gamma\prime, \end{cases}$$

where $\gamma \in \mathbb{R}$. The test statistics can be defined as

$$LR = \frac{L(\underset{\sim}{t}; \tilde{\alpha}, \tilde{\beta}, \gamma\prime, \underline{K})}{L(\underset{\sim}{t}; \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \underline{K})}$$

where $\tilde{\alpha}, \tilde{\beta}$ are such that $\sup_{(\alpha,\beta)\in\mathbb{R}^+\times\mathbb{R}^+} L(\underset{\sim}{t}; \alpha, \beta, \gamma\prime, \underline{K}) = L(\underset{\sim}{t}; \tilde{\alpha}, \tilde{\beta}, \gamma\prime, \underline{K})$ and $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$ are such that $\sup_{(\alpha,\beta,\gamma)\in\mathbb{R}^+\times\mathbb{R}^+\times\mathbb{R}} L(\underset{\sim}{t}; \alpha, \beta, \gamma, \underline{K}) = L(\underset{\sim}{t}; \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \underline{K})$.

The probability density function of the Arrhenius-Weibull model is assumed in this chapter to describe the failure time at a fixed stress level, and its parameters are therefore maximized in the likelihood ratio test. There are in total three parameters that need to be estimated under the alternative hypothesis; α_0, β, and γ. But, we only estimate two parameters which are α_0 and β, and fix the parameter γ under the null hypothesis. Under the null hypothesis, the LR follows a ${\chi_1}^2$ distribution. To get $[\ \underline{\gamma}_i, \overline{\gamma}_i]$, for each value of γ we would have different $\hat{\alpha}_0$ and $\hat{\beta}$ but we do not use these any further in our method.

For each $i = 1, ..., m$, we find $\underline{\gamma}_i$, the smallest value for $\gamma\prime$ for which we do not reject the null hypothesis, and $\overline{\gamma}_i$, the largest value for $\gamma\prime$ for which we do not reject the null hypothesis. Then we define $\underline{\gamma} = \max\{\min \underline{\gamma}_i, 0\}$ and $\overline{\gamma} = \max \overline{\gamma}_i$. Note that, because of the physical interpretation of generally faster failures with increased stress levels, we exclude negative values which leads to some $\underline{\gamma}$ values being set at 0. In this thesis, we will always restrict $\underline{\gamma}$ to non-negative values. We find the $\underline{\gamma}_i$ and $\overline{\gamma}_i$ numerically using the statistical software R.

Note that, we do not make a confidence statement for the final NPI lower and upper survival functions, so we do not explicitly quantify the prediction accuracy. If indeed the main assumption is valid, that increased stress tends to decreased failure

times, then negative $\underline{\gamma}_i$ are typically resulting from statistical variation and would disappear for larger samples.

One may wish to allow $\underline{\gamma}$ to be negative, which may e.g. be reasonable if it turned out that higher stress level could possibly improve a unit's failure time. However, in normal ALT applications there tends to be sufficient knowledge about the effect of the stress on the failure times that such cases would be rare, hence we do not consider negative values for $\underline{\gamma}$ values in this thesis. It should be emphasized though that our inferential method could still be used if negative values for lower $\underline{\gamma}$ were allowed.

In the second step, we apply the data transformation using $\underline{\gamma}$ for all levels $i = 1, ..., m$ to get transformed data at level 0, which are then used together with the original data at the normal stress level, to derive the NPI lower survival function $\underline{S}$. Similarly, we apply the data transformation using $\overline{\gamma}$ for all levels $i = 1, ..., m$ to get transformed data at level 0, which are then used together with the original data at the normal stress level, to derive the NPI upper survival function $\overline{S}$. Note that each observation at an increased stress level transforms into an interval-valued observation at the normal stress level 0, where the width of an interval is larger for an observation from a higher stress level.

Note that, if the model fits really well, we expect most $\underline{\gamma}_i$ values to be quite similar, as well as most $\overline{\gamma}_i$ values. If the model fits poorly, $\underline{\gamma}_i$ are most probably very different, or $\overline{\gamma}_i$ are very different, or both. Hence, in case of poor model fit, the resulting interval $[\underline{\gamma}, \overline{\gamma}]$ tends to be wider than in the case of good model fit. If the model assumed is not too far from reality, we would expect the widest interval for the parameter γ to come from the likelihood ratio test applied to levels 1 and 0.

If the model assumed is not too far from reality, we would expect the widest interval for the parameter γ to come from the likelihood ratio test applied to levels 1 and 0. If the model fits well, a level 1 observation is transferred to a smaller interval on level 0 than a level 2 interval, if the transferred intervals are close, in particular if they are overlapping. In the overlapping case, because the level 2 interval is wider, the left and right end points of these intervals from level 2 are further apart, which implies that the γ in the null hypothesis will be rejected in more cases. Hence, the

interval $[\underline{\gamma}, \overline{\gamma}]$ form level 1 will be wider than the interval $[\underline{\gamma}, \overline{\gamma}]$ form level 2 (in most cases, due to variability in the samples not in all cases). If the model is worse than we would expect more often that $\underline{\gamma} = \underline{\gamma}_i$ for an $i \neq 1$, or $\overline{\gamma} = \overline{\gamma}_i$ for an $i \neq 1$. If the model assumptions are not fully correct, for example, using some misspecification cases or there is a lot of overlap between the data, then latter can happen.

Although our method involved multiple pairwise tests, we do not aim to combine these into a single test of a hypothesis involving all groups simultaneously. A multiple testing (comparison) procedure arises in many scenarios, where two groups of data are compared over time with each other [16, 61]. A well known scenario that one may be interested in testing is the comparison between two groups based on confidence level [61]. While we use pairwise tests in our approach, we do not combine these into an overall confidence level statement for the resulting inference. Instead we use NPI to derive the lower and upper predictive survival functions and we investigate the performance of our predictive method separately via simulations.

If the assumed model is fully correct then the lower and upper γ will form an interval with at least $1 - \alpha$ confidence level, with α the significance level for each pairwise test. However, we explicitly develop our method for robust inference as the basic assumed model will in practice not be 'correct', and acknowledging this makes confidence statements hard to justify. This is an interesting topic for future research. The method proposed in this section is illustrated by three examples using simulated data and real data in Section 3.4, and studied in more detailed by simulation in Section 3.5.

3.4 Examples

In this section we present three examples to illustrate the method presented in Section 3.3. In Example 3.4.1 we simulate $n = 10$ observations at all levels using the same model for the simulation as assumed for the analysis with our proposed method in Section 3.3. In Example 3.4.2, we simulate $n = 20$ observations at all levels to show the effect of the larger sample that correspond to the model for the link function we assume for the analysis in the Example 3.4.1, and applied our proposed

methods in Section 3.3. In Example 3.4.3 we use a data set from the literature. In this thesis, our method does not require the same number of observations at every level. However, in many situations we show examples where n_i are all the same, so in this case we call it n.

Example 3.4.1. This example consists of two cases. In Case 1 we simulate data at all levels using a Weibull distribution at each stress level and we assume the Arrhenius link function we assume for the analysis. In Case 2 we change these data such that the assumed link function will not provide a good fit anymore and we investigate the effect on the interval $[\underline{\gamma}, \overline{\gamma}]$ and on the corresponding lower and upper predictive survival functions for a future observation at the normal stress level.

We assume the normal temperature level to be $K_0 = 283$, and the increased temperature levels to be $K_1 = 313$ and $K_2 = 353$ Kelvin. We generate ten observations from a Weibull distribution at each stress level linked by the Arrhenius link function. The Weibull distribution at level K_0 has shape parameter $\beta = 3$ and scale parameter $\alpha_0 = 7000$, and the Arrhenius parameter's set at $\gamma = 5200$. We keep the same shape parameter at each temperature, and the scale parameters are linked by the Arrhenius relation, which leads to $\alpha_1 = 1202.942$ at level K_1 and $\alpha_2 = 183.091$ at level K_2. Ten failure times were simulated at each temperature, so data for a total of 30 units is used in the study. The failure times are given in Table 3.1.

To illustrate the method discussed in Section 3.3 using these data, we assume the Weibull distribution at each stress level and the Arrhenius link function for the data. To obtain the intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ of the values for γ for which we do not reject the null hypothesis, we used the pairwise likelihood ratio test between K_i for $i = 1, 2$ and K_0. The resulting intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ for three significance levels are given in Table 3.2. Note that we transformed the data using the overall values $[\underline{\gamma}, \overline{\gamma}]$, derived as the minimum and maximum corresponding values for the pairwise tests, respectively.

In the second step of our method, we transformed the data using the $[\underline{\gamma}, \overline{\gamma}]$ values. All observations at the increased stress levels were transformed to the normal stress level. Therefore, the observations at the increased stress levels K_1 and K_2 are transformed to interval-valued observations at the normal stress level K_0. We

Case	$K_0 = 283$	$K_1 = 313$	$K_2 = 353$	$K_1 = 313$ ($\times 1.4$)	$K_2 = 353$ ($\times 0.4$)
1	2692.596	241.853	74.557	338.595	29.823
2	3208.336	759.562	94.983	1063.387	37.993
3	3324.788	769.321	138.003	1077.050	55.201
4	5218.419	832.807	180.090	1165.930	72.036
5	5417.057	867.770	180.670	1214.878	72.279
6	5759.910	1066.956	187.721	1493.739	75.088
7	6973.130	1185.382	200.828	1659.535	80.331
8	7690.554	1189.763	211.913	1665.668	84.765
9	8189.063	1401.084	233.529	1961.517	93.412
10	9847.477	1445.231	298.036	2023.323	119.214

Table 3.1: Failure times at three temperature levels (first three columns) and changed failure times (last two columns), Example 3.4.1.

Significance level	0.01		0.05		0.10	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
Case 1: K_1 K_0	4060.018	6605.752	4424.881	6261.168	4593.700	6100.653
K_2 K_0	4377.043	5602.321	4550.205	5434.908	4630.511	5357.037
Case 2: $K_1 \times (1.4)$, K_0	3066.539	5612.273	3431.402	5267.689	3600.221	5107.174
$K_2 \times (0.4)$, K_0	5684.708	6909.985	5857.870	6742.573	5938.175	6664.701

Table 3.2: $[\underline{\gamma}_i, \overline{\gamma}_i]$ for Example 3.4.1.

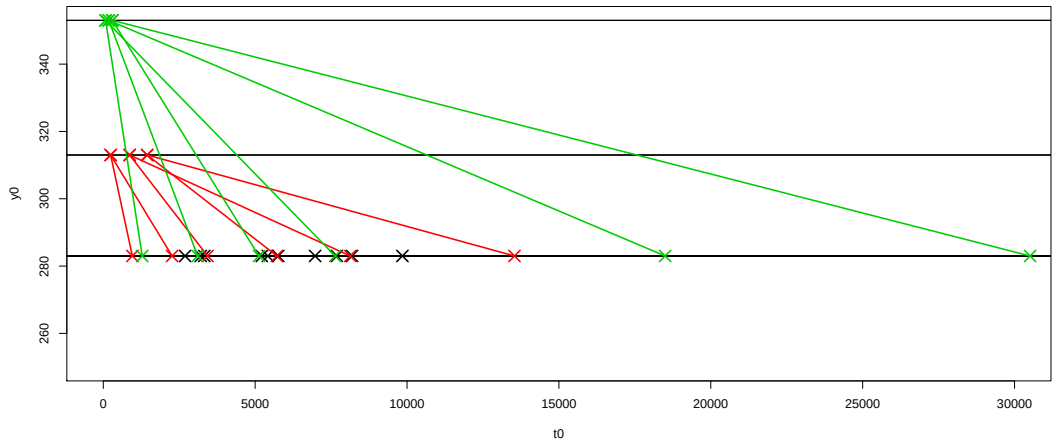


Figure 3.1: Some transformed data using $[4060.018, 6605.752]$, Example 3.4.1.

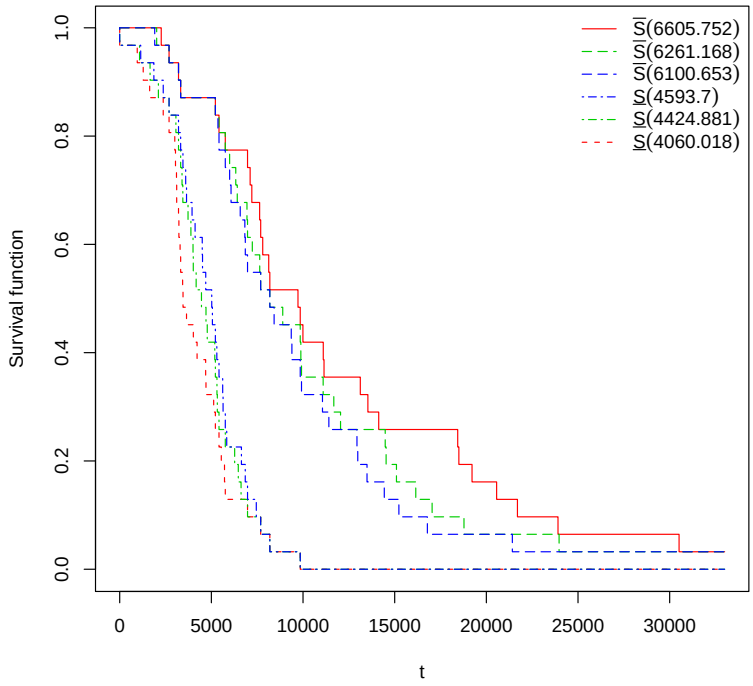
briefly illustrate this in Figure 3.1, using only three points of the data at each level, therefore, we have six lines going down from each level. We transformed the data from the higher stress levels K_1 and K_2 using $[\underline{\gamma}, \overline{\gamma}] = [4060.018, 6605.752]$ with 0.01 level of significance, mixed with the original data at the normal stress level K_0. Note that, in Figure 3.1 the two largest transformed data points are the $\underline{\gamma}$ and $\overline{\gamma}$ transformations of the largest observation from level K_2. So, this illustrates a key property of our method, that data transformed from higher levels tend to be wider intervals at the normal level.

The NPI lower survival function is based on the original data at level 0 together with the transformed data from the stress levels K_1 to K_0 and K_2 to K_0 using $\underline{\gamma}$. Similarly, the NPI upper survival function is based on the original data at level 0 together with the transformed data from the stress levels K_1 to K_0 and K_2 to K_0 using $\overline{\gamma}$. The $\underline{\gamma}$ transformed the points to the smallest values and therefore is the most pessimistic case, which leads to the lower survival function $\underline{S}$. The $\overline{\gamma}$ transformed the points to the largest values and therefore is the most optimistic case, which leads to the upper survival function $\overline{S}$. In Case 1, we have $\underline{\gamma} = 4060.018$ and $\overline{\gamma} = 6605.752$, $\underline{\gamma} = 4424.881$ and $\overline{\gamma} = 6261.168$, and $\underline{\gamma} = 4593.700$ and $\overline{\gamma} = 6100.653$, which they are all equal to the values $[\underline{\gamma}, \overline{\gamma}]$ that results from the pairwise test K_1, K_0 with significance levels 0.01, 0.05 and 0.10, respectively. We used all the above $\underline{\gamma}$ and $\overline{\gamma}$ values to transform the data to the normal stress level 0, see Figure 3.2(a). In

this figure, the lower survival function $\underline{S}$ is labeled as $\underline{S}$ $(\underline{\gamma}_i)$ and the upper survival function $\overline{S}$ is labeled as $\overline{S}$ $(\overline{\gamma}_i)$. This figure shows that lower significance level leads to more imprecision and nesting for the NPI lower and upper survival functions.

In Case 2, we illustrate our method in case of misspecification. The goal of our method is to study whether or not a simple model will give robustness in case of misspecification. We multiply the data at level K_1 by 1.4 and in addition we multiply the data at level K_2 by 0.4. The corresponding data values are given in the last two columns in Table 3.1. In this case, we have $\underline{\gamma} = 3066.539$ and $\overline{\gamma} = 6909.985$, $\underline{\gamma} = 3431.402$ and $\overline{\gamma} = 6742.573$, and $\underline{\gamma} = 3600.221$ and $\overline{\gamma} = 6664.701$ for significance levels 0.01, 0.05 and 0.10, respectively. We used all the above $\underline{\gamma}$ and $\overline{\gamma}$ values to transform the data to the normal stress level 0, see Figure 3.2(b). Figure 3.2(b) also shows that lower significance level results in more imprecision for the NPI lower and upper survival functions. In Case 2, we can see that the $[\underline{\gamma_i}, \overline{\gamma_i}]$ intervals for the two pairwise comparisons are fully disjoint unlike in Case 1. Note that in Case 2, the observations at level K_1 have increased, leading to smaller $\underline{\gamma}_1$ and $\overline{\gamma}_1$ values, which, in turn, leads to the lower and upper survival functions to decrease in comparison to Case 1. Also, the observations at stress level K_2 in Case 2 have decreased, resulting in larger values for $\underline{\gamma}_2$ and $\overline{\gamma}_2$, and this leads to the lower and upper survival functions to increase in comparison to Case 1. Therefore, it is obvious that using the $\underline{\gamma}$ and $\overline{\gamma}$ values gives substantially more imprecision in our NPI method in Case 2 than in Case 1. This illustrates that, in case of poor model fit, the NPI lower and upper survival functions in Figure 3.2(b), using our method as discussed in Section 3.3, have more imprecision and more nesting than if the model fits well.

(a) Case 1



(b) Case 2

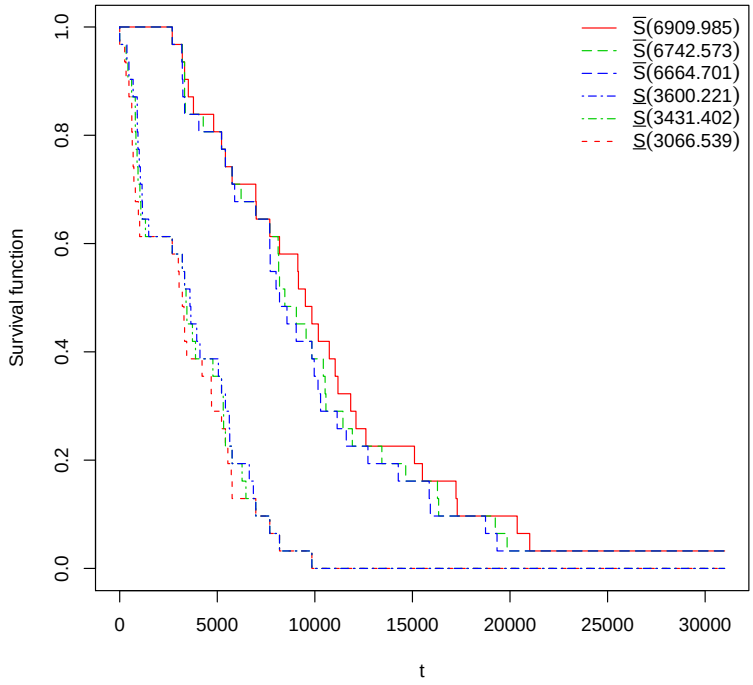


Figure 3.2: The NPI lower and upper survival functions, Example 3.4.1.

Example 3.4.2. The previous example illustrated our methods using a data set with $n = 10$ observations at each stress level. In this example, we use a larger data set to explain the effect of the amount of the data on our inference. We simulated twenty failure times at each of three stress levels, so a total of 60 failure times is used in the study. Note that we use the same temperature stress levels as in Example 3.4.1. The Arrhenius-Weibull model also keeps the same (α, β, and γ) parameters in the simulation as in Example 3.4.1. The failure times are given in Table 3.3.

The resulting intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ for three significance levels using our method are given in Table 3.4. Comparing the resulting intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ in Tables 3.2 and 3.4, shows that based on more data at each stress level, there is less imprecision for the larger value of n, which is due to the likelihood ratio test, so fewer values of γ in the null hypothesis are not rejected. In this example, we have $\underline{\gamma} = 4117.376$ and $\overline{\gamma} = 5712.309$, $\underline{\gamma} = 4324.934$ and $\overline{\gamma} = 5506.789$, and $\underline{\gamma} = 4425.681$ and $\overline{\gamma} = 5406.786$, which they are all equal to the values $[\underline{\gamma}, \overline{\gamma}]$ that results from the pairwise test K_1 to K_0 with significance levels 0.01, 0.05 and 0.10, respectively. Note further that the NPI part also has less imprecision for the larger of n. Figure 3.3 shows the NPI lower and upper survival functions which are based on the original data at level 0 together with the transformed data from the stress levels K_1 to K_0 and K_2 to K_0 using $\underline{\gamma}$ and $\overline{\gamma}$, respectively, which also shows less imprecision and nesting compared to Figure 3.2. This example shows that based on more data the imprecision in the resulting intervals $[\underline{\gamma}, \overline{\gamma}]$ decreases.

Example 3.4.3. The method proposed in Section 3.3 is now applied to a data set from the literature [71], resulting from a temperature-accelerated life test. The author did not provide a detailed description of the set-up of the experiment. The time-to-failure data were collected at three temperatures (in Kelvin): $K_0 = 393, K_1 = 408$, and $K_2 = 423$, with 393 the normal temperature for the process of interest. Ten units were tested at each temperature, so a total of 30 units where used in the study. All units failed during the experiment. The failure times, in hours, are given in Table 3.5.

For the data in Table 3.5, we have assumed the same model as discussed in Section 3.1, so with a Weibull failure time distributions at each stress level and the

Case	$K_0 = 283$	$K_1 = 313$	$K_2 = 353$
1	1407.360	489.851	97.156
2	2692.596	624.058	106.588
3	3208.336	690.951	113.232
4	3324.788	737.428	114.521
5	4419.943	825.673	124.121
6	4476.732	888.376	131.099
7	4846.159	906.701	136.292
8	5049.613	962.071	138.762
9	5218.419	1071.215	153.764
10	5417.057	1082.967	157.405
11	5759.910	1156.457	159.132
12	6208.689	1183.222	165.528
13	6897.815	1187.228	165.589
14	6923.310	1233.356	171.741
15	6973.130	1319.475	176.731
16	7690.554	1392.307	191.887
17	8152.997	1430.277	205.179
18	8189.063	1534.328	241.959
19	8409.894	1570.710	253.354
20	9847.477	1958.147	284.655

Table 3.3: Failure times at three temperature levels, Example 3.4.2.

Significance level	0.01		0.05		0.10	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
K_1 K_0	4117.376	5712.309	4324.934	5506.789	4425.681	5406.786
K_2 K_0	4731.700	5492.383	4830.244	5392.011	4878.140	5346.543

Table 3.4: $[\underline{\gamma}_i, \overline{\gamma}_i]$ for Example 3.4.2.

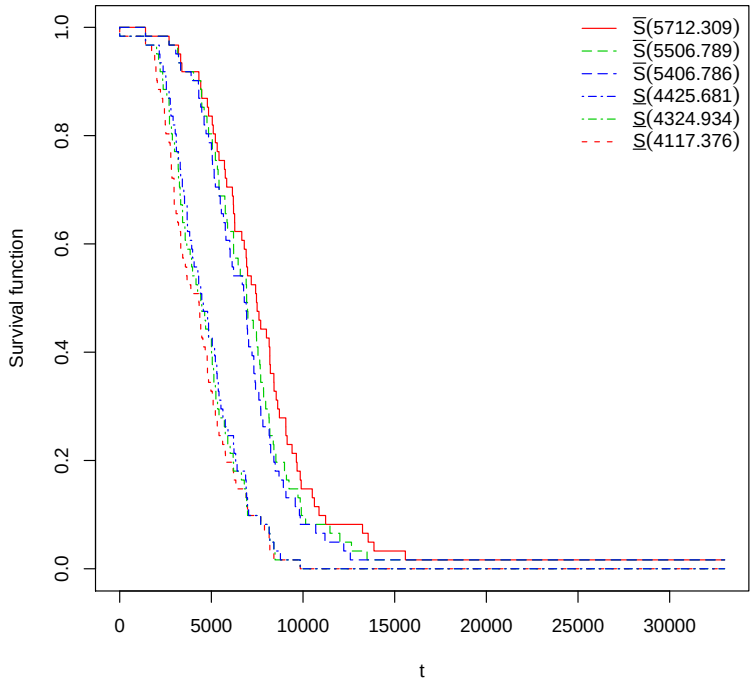


Figure 3.3: NPI lower and upper survival functions, Example 3.4.2.

Arrhenius link function between different stress levels. Then the pairwise likelihood ratio test is used separately between level K_i and K_0, for $i = 1, 2$, to derive the intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ for the value of the parameter γ of the Arrhenius link function, such that the null hypothesis that two groups of failure data (the transformed data from level i and the real data from level 0) come from the same underlying distribution, is not rejected for values of γ in this interval. The resulting intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ are given in Table 3.6, for three test significance levels.

Note that, because the data corresponding to the different stress levels already have quite some overlap, the likelihood ratio tests even did not rule out negative values for γ. However, because of the physical interpretation of generally faster failures with increased temperature, we exclude negative values for $\underline{\gamma}$, so we define $\underline{\gamma_i} = \max\{\min \underline{\gamma}_i, 0\}$, as discussed in Section 3.3. Following the method presented in this chapter, we transformed the data using the overall values $[\underline{\gamma}, \overline{\gamma}]$, derived as the smallest and largest corresponding values for the pairwise tests, respectively. For the significance levels $0.01, 0.05$, and 0.1, the values of $\underline{\gamma}$ are always 0 in this

Case	$K_0 = 393$	$K_1 = 408$	$K_2 = 423$
1	3850	3300	2750
2	4340	3720	3100
3	4760	4080	3400
4	5320	4560	3800
5	5740	4920	4100
6	6160	5280	4400
7	6580	5640	4700
8	7140	6120	5100
9	7980	6840	5700
10	8960	7680	6400

Table 3.5: Failure times at three temperature levels, Example 3.4.3.

Significance level	0.01		0.05		0.10	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
K_1 K_0	-1585.607	4881.225	-692.940	3988.558	-276.575	3572.193
K_2 K_0	188.348	3540.639	651.091	3077.896	866.927	2862.060

Table 3.6: $[\underline{\gamma}_i, \overline{\gamma}_i]$ for, Example 3.4.3.

case, and for $\overline{\gamma}$ we have, 4881.225, 3988.558 and 3572.193, respectively. Based on the original data at level 0 together with the data transformed from the stress levels K_1 to K_0 and K_2 to K_0 using $\underline{\gamma}$, the NPI approach provides the NPI lower survival function. Similarly, but using $\overline{\gamma}$ for the transformation, we derived the NPI upper survival function.

The resulting lower and upper survival functions are presented in Figure 3.4, where of course the three different significant levels lead to the same lower survival function due to the restriction for the γ values to be non-negative. In this figure, the lower survival function $\underline{S}$, denoted by $\underline{S}$ (0), is the same for all these significance levels. The upper survival function $\overline{S}$ is labeled as $\overline{S}$ ($\overline{\gamma}_i$). This figure shows that a smaller significance level leads to more imprecision for the NPI lower and upper

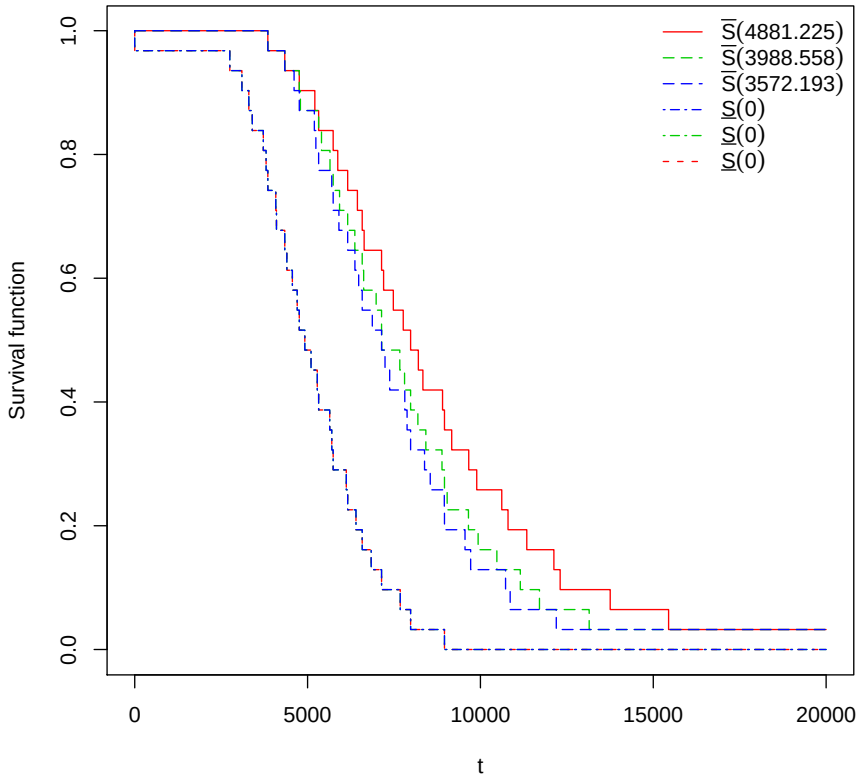


Figure 3.4: The NPI lower and upper survival functions, Example 3.4.3.

survival functions, which is directly resulting from the fact that the intervals of not rejected values of γ_i in the pairwise tests will be nested, becoming larger if the significance level is decreased. These lower and upper survival functions can also be used to deduce corresponding lower and upper values for percentiles, which are found in the usual way by inverting the respective functions. These functions can also be used as inputs into decision processes, for example with regard to setting warranty policies, this is the topic of Chapter 5.

3.5 Simulation studies

In this section, we investigate the performance of the imprecise predictive statistical method for ALT data in this chapter. In this section, we first study the performance of our proposed method by simulations when considering the assumed Arrhenius-

Weibull model as the true underlying model. Thereafter, we consider cases of misspecification.

To investigate the predictive inference performance of our new method for ALT data, we have conducted a simulation study. We assumed the temperature stress levels to be $K_0 = 283$, $K_1 = 313$, and $K_2 = 353$. In this analysis, we ran the simulation 10,000 times with the data simulated from the model given in Section 3.1, with the Weibull distribution with $\beta = 3$, $\alpha = 7000$, and the Arrhenius link function parameter $\gamma = 2000$. We have simulated $n = 10, 50, 100$ observations at each stress level. We applied the method described in Section 3.3, with levels of significance 0.01, 0.05, and 0.10. To evaluate the performance of the method, we check the results by simulating a future observation at the normal stress level K_0, and we consider if it mixes well among the data at level K_0, where we use both the actual data at level K_0 and the data transformed to the normal stress level K_0. We examined the performance by considering the quartiles of NPI lower and upper survival functions for $q = 0.25, 0.50, 0.75$, and whether or not the future observation at the normal stress level has exceeded these quartiles in the right proportions. One can similarly use different quantiles, but the quartiles provide a good indicator of the overall performance of our method.

In Case A, we conducted the simulation with the assumed Arrhenius-Weibull model, hence with data generated from the same model as used for our method, hence we expect a good performance, which will require that the future observation for each run at the normal stress level has exceeded the first, second, and third quartiles of the NPI lower survival functions just over proportions $0.75, 0.50$ and 0.25 of all runs, respectively, and for the NPI upper survival functions just under proportions $0.75, 0.50$ and 0.25. Table 3.7 and Figures 3.5-3.7 present the results of the performance of our method for this simulation. All cases in Table 3.7 and Figures 3.5-3.7 show an overall good performance of the proposed method, with levels of significance 0.01, 0.05, and 0.10, and with sample sizes $n = 10, 50, 100$. Note that the first, second, and third quartiles in these figures are denoted $qL0.25$ and $qU0.25$, $qL0.50$ and $qU0.50$, and $qL0.75$ and $qU0.75$ corresponding to the NPI lower and upper survival functions, respectively. We note that for corresponding

proportions with larger values of n, the differences between lower and upper survival functions tend to decrease, as shown in Figure 3.7. That means that when have more data, the imprecision in the NPI lower and upper survival functions decreases. As mentioned in Section 3.3, if the model fits very well or perfectly, then in most cases $\underline{\gamma} = \underline{\gamma}_1$, and $\overline{\gamma} = \overline{\gamma}_1$. Table 3.8 shows that when we simulate from the assumed Arrhenius-Weibull model out of 10,000 runs with 0.05 level of significance, hence with data generated from the same model, that in most cases we have indeed that $\underline{\gamma} = \underline{\gamma}_1$, and $\overline{\gamma} = \overline{\gamma}_1$. From these simulations, we conclude that using our approach with the Arrhenius-Weibull model, which illustrated our method achieves suitable predictive inference if the model assumptions are fully correct.

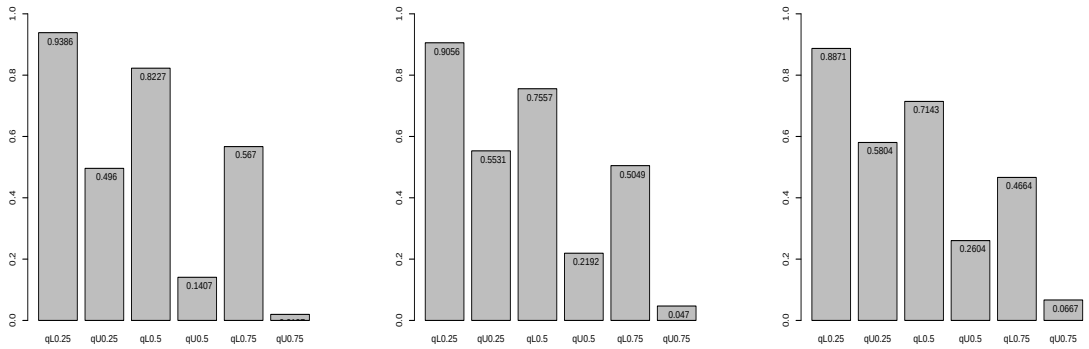
K_1K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9386	0.4960	0.8565	0.6287	0.8293	0.6717
	0.50	0.8227	0.1407	0.6726	0.3208	0.6349	0.3684
	0.75	0.5670	0.0197	0.4314	0.0900	0.3818	0.1245
0.05	0.25	0.9058	0.5531	0.8326	0.6585	0.8131	0.6938
	0.50	0.7557	0.2192	0.6322	0.3660	0.6028	0.4000
	0.75	0.5049	0.0470	0.3966	0.1238	0.3511	0.1531
0.1	0.25	0.8871	0.5804	0.8214	0.6730	0.8048	0.7030
	0.50	0.7143	0.2604	0.6137	0.3866	0.5880	0.4155
	0.75	0.4664	0.0667	0.3755	0.1415	0.3351	0.1681
K_2K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.8795	0.5054	0.8088	0.6923	0.7974	0.7132
	0.50	0.6974	0.2869	0.5919	0.4072	0.5712	0.4361
	0.75	0.4588	0.0649	0.3525	0.1599	0.3162	0.1838
0.05	0.25	0.8538	0.6418	0.7945	0.7073	0.7881	0.7220
	0.50	0.6524	0.3392	0.5699	0.4287	0.5552	0.4518
	0.75	0.4073	0.1035	0.3287	0.1792	0.2998	0.2005
0.1	0.25	0.8398	0.6589	0.7881	0.7146	0.7835	0.7260
	0.50	0.6285	0.3644	0.5594	0.4411	0.5470	0.4605
	0.75	0.3818	0.1239	0.3175	0.1900	0.2917	0.2079
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9427	0.4925	0.8577	0.6273	0.8299	0.6708
	0.50	0.8277	0.1352	0.6749	0.3189	0.6363	0.3672
	0.75	0.5732	0.0149	0.4333	0.0886	0.3834	0.1235
0.05	0.25	0.9122	0.5459	0.8347	0.6562	0.8144	0.6920
	0.50	0.7664	0.2087	0.6362	0.3625	0.6058	0.3982
	0.75	0.5150	0.0376	0.4003	0.1203	0.3538	0.1500
0.1	0.25	0.8957	0.5714	0.8238	0.6700	0.8074	0.7002
	0.50	0.7299	0.2485	0.6202	0.3818	0.5920	0.4122
	0.75	0.4792	0.0546	0.3812	0.1366	0.3384	0.1642

Table 3.7: Proportion of runs with future observation greater than the quartiles. Case *A*.

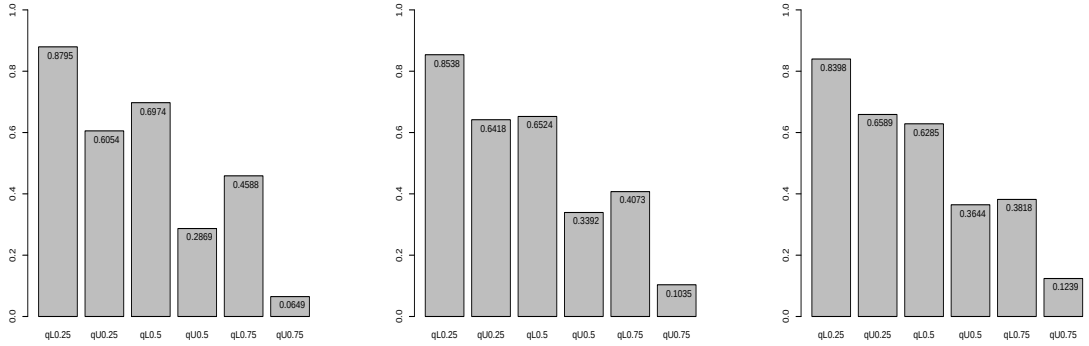
min $\underline{\gamma}$ and the max $\overline{\gamma}$	$n = 10$	$n = 50$	$n = 100$
$\underline{\gamma} = \underline{\gamma}_1$	8699	8963	9022
$\overline{\gamma} = \overline{\gamma}_1$	8675	9003	9087
$\underline{\gamma} = \underline{\gamma}_1$ and $\overline{\gamma} = \overline{\gamma}_1$	7374	7966	8109

Table 3.8: Number of the simulation runs with $\underline{\gamma} = \underline{\gamma_1}$ or $\overline{\gamma} = \overline{\gamma_1}$ or both out of 10,000 simulation runs with 0.05 level of significance. Case A.

(a) K_1 and K_0 (0.01) (b) K_1 and K_0 (0.05) (c) K_1 and K_0 (0.1)



(d) K_2 and K_0 (0.01) (e) K_2 and K_0 (0.05) (f) K_2 and K_0 (0.1)



(g) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.01) (h) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.05) (i) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.1)

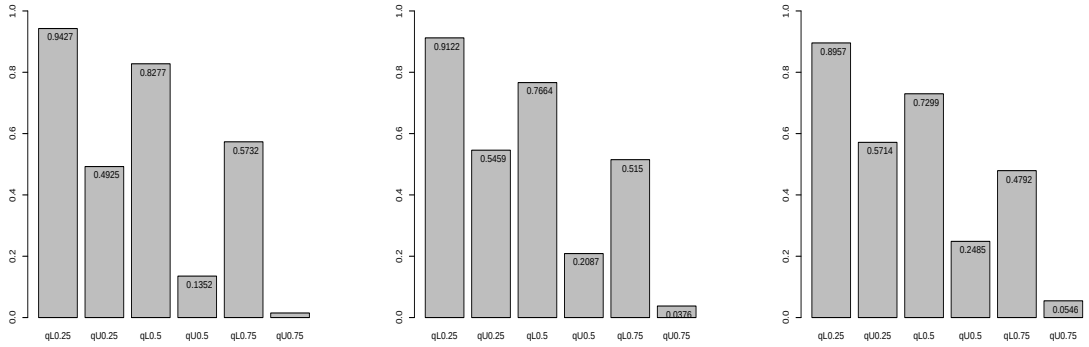
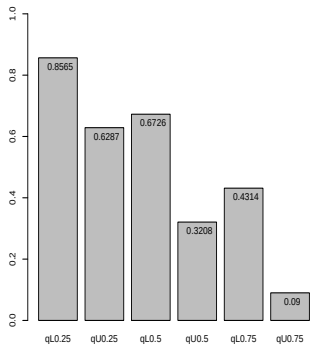
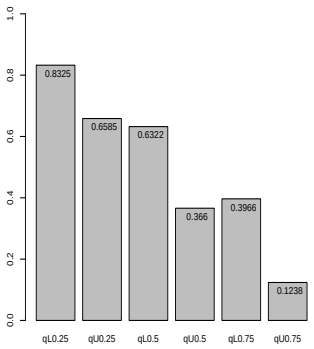


Figure 3.5: Proportion of runs with future observation greater than the quartiles, $n = 10$. Case A.

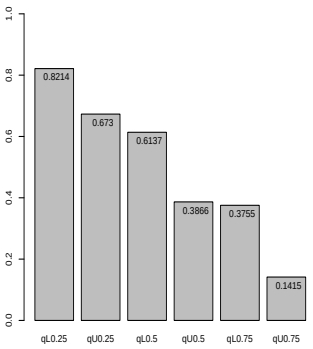
(a) K_1 and K_0 (0.01)



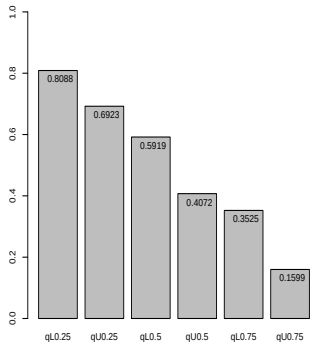
(b) K_1 and K_0 (0.05)



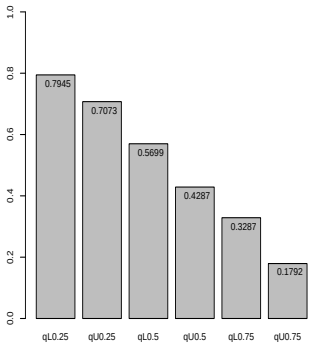
(c) K_1 and K_0 (0.1)



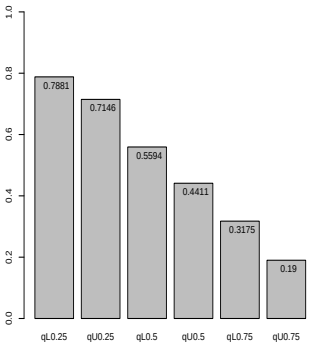
(d) K_2 and K_0 (0.01)



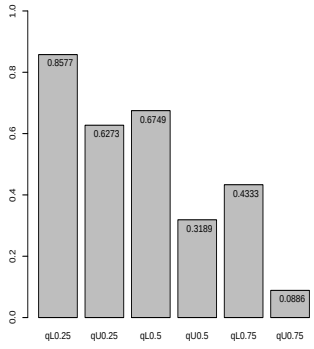
(e) K_2 and K_0 (0.05)



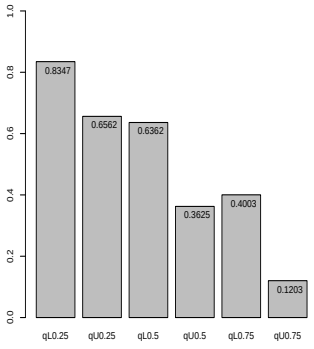
(f) K_2 and K_0 (0.1)



(g) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.01)



(h) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.05)



(i) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.1)

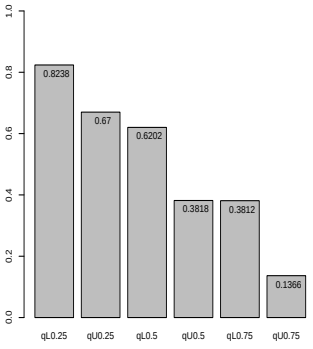
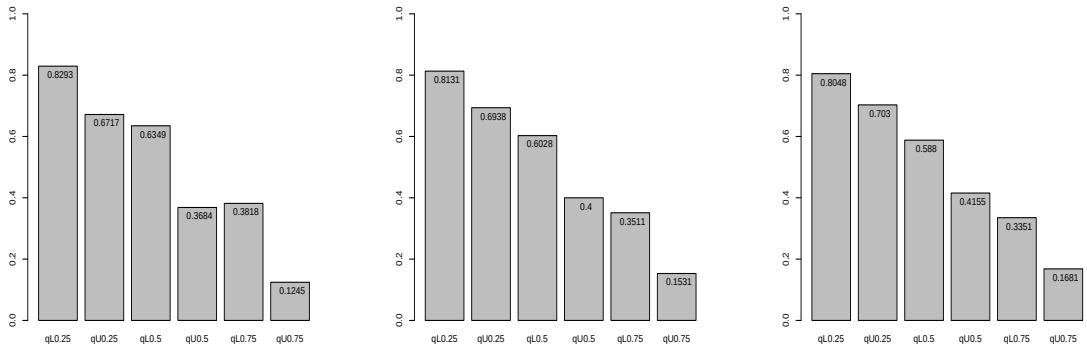
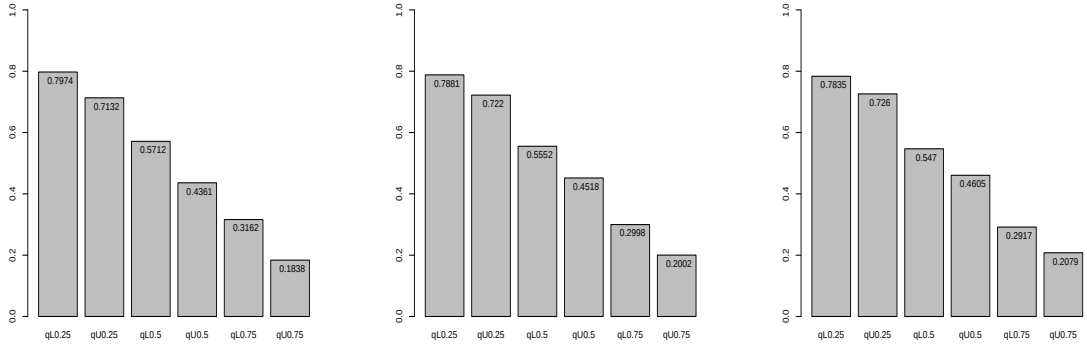


Figure 3.6: Proportion of runs with future observation greater than the quartiles, $n = 50$. Case A.

(a) K_1 and K_0 (0.01) (b) K_1 and K_0 (0.05) (c) K_1 and K_0 (0.1)



(d) K_2 and K_0 (0.01) (e) K_2 and K_0 (0.05) (f) K_2 and K_0 (0.1)



(g) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.01) (h) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.05) (i) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.1)

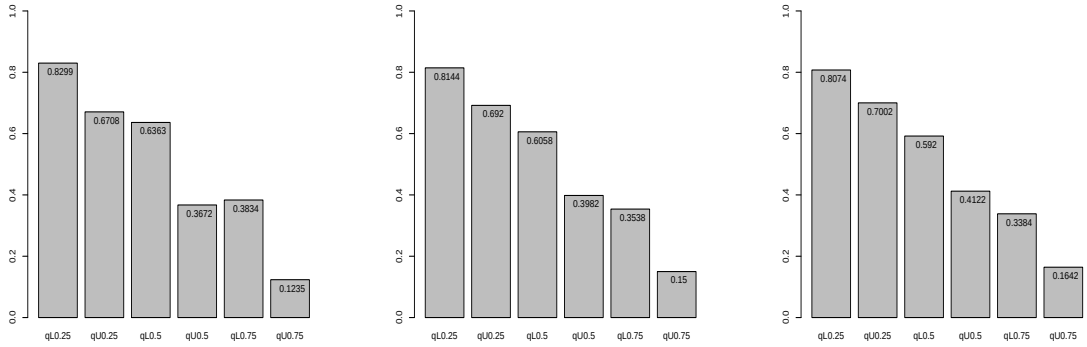


Figure 3.7: Proportion of runs with future observation greater than the quartiles, $n = 100$. Case A.

As mentioned before, the goal in this chapter is to develop a quite straightforward method of predictive inference based on few assumptions, where the imprecision in the link function between different stress levels provides robustness against the necessary assumptions. Next, we investigate robustness in case of model misspecification in the next three simulation cases, and study whether or not a simple model will give robustness in case of misspecification. In Case 1, we used the same scenario as in the Case A simulation, but with the shape parameter $\beta = 2$ for each level in the analysis, while in Case 2 we changed to the simulated data in Case A such that the assumed link function may not provide a good fit. In Case 3, we look a situation that is more likely in practice, namely where the Arrhenius model is assumed for the analysis but the data is generated from the Eyring-Weibull model.

In Case 1, we are dealing with a particular misspecification of the model for statistical inference for ALT data [50, 52]. Table 3.9 presents the results of the predictive performance of our method. We used the same scenario as in the Case A simulation so we simulate from a Weibull distribution with $\beta = 3$ but we assumed wrongly that $\beta = 2$ for each level in the analysis. Now only the scale parameter α and the parameter γ of the link function are explicitly maximized in the likelihood ratio. In comparison to the Case A simulation results where the shape parameter β was estimated, there is a slight more imprecision on the quartiles in Table 3.9. From Table 3.7 and Table 3.9, the effect of fixing the shape parameter $\beta = 2$ in the analysis can also be seen on the quartiles of the NPI lower and upper survival functions for $q = 0.25, 0.50, 0.75$, which is reflected by more imprecision for this case than in Case A, using the resulting interval $[\underline{\gamma}, \overline{\gamma}]$ in our method, as presented in Section 3.3. Also, we checked whether or not the simulated future observation at the normal stress level has exceeded these quartiles in the right proportions. The imprecision in the NPI lower and upper survival functions with larger values of n in Table 3.9 for this case is wider, in comparison to the imprecision in the NPI lower and upper survival functions with larger values of n in Table 3.7 for Case A. Throughout this simulation, when we fixed the shape parameter $\beta = 2$ in the fitted model, our method provided levels of robustness against the misspecification.

In Case 2, we investigate the predictive performance of our method for ALT data,

with a change to the data such that the assumed link function may not provide a good fit. We show the change to the resulting quartiles of NPI lower and upper survival functions for $q = 0.25, 0.50, 0.75$, and we check if the simulated future observation at the normal stress level K_0 has exceeded these quartiles in the right proportions. To illustrate this, we generated data sets as before with $n = 10, 50, 100$ observations at each stress level using the Arrhenius-Weibull model. But now all the data at stress level K_1 are multiplied by 1.2, and all the data at level K_2 are multiplied by 0.8. This is similar to what was done in Example 3.4.1 to illustrate our method in case of misspecification. We ran the simulation 10,000 times. Using these generated data, we again applied our method as described in Section 3.3, with levels of significance 0.01, 0.05, and 0.10. We performed the likelihood ratio test within the statistical software R to the simulated data set. All the results in Case 2 provide an insight into whether or not the presented method shows some robustness against the misspecification case considered. Table 3.10 presents the results of these simulations with $n = 10, 50, 100$.

As explained previously, we required that the future observation for each run at the normal stress level has exceeded the first, second, and third quartiles of the NPI lower survival functions just over proportions $0.75, 0.50$ and 0.25 of all runs, respectively, and for the NPI upper survival functions just under proportions $0.75, 0.50$ and 0.25. However, for $n = 50, 100$ in this simulation, there are a few cases for which the future observation for each run at the normal stress level has exceeded the first, second, and third quartile ($qU0.25$, $qU0.50$, $qU0.75$). These correspond to the NPI upper survival functions just over $0.75, 0.50, 0.25$ of the pairwise level $K_1 \times (1.2)$ to K_0, respectively, see Table 3.10, and they are highlighted by bold font in Table 3.10, whereas they are only just over these proportions ($qU0.25$, $qU0.50$, $qU0.75$). Exceeding the first, second, and third quartile ($qU0.25$, $qU0.50$, $qU0.75$) of the NPI upper survival function just over $0.75, 0.50, 0.25$ is due to the use of this misspecification case where the data at stress level K_1 (K_0) are multiplied by 1.2 (0.8). There are slight increases for $qU0.25$, $qU0.50$, $qU0.75$, which means that we have too many observations passing these quartiles from the upper $0.75, 0.50, 0.25$. It seems that the points $0.75, 0.50, 0.25$ occurred a bit earlier, which is related to

the effect of multiplying the data by 1.2 for the stress levels $K_1 \times (1.2)$ to K_0. This is in line with our expectation, which is mainly due to the misspecification case we assumed. Note that, the smaller significance level leads to more imprecision for the NPI lower and upper survival functions. Further note that in this simulation, the observations at the stress level K_1 have increased compared to the earlier simulations, resulting in smaller $\underline{\gamma_1}$ and $\overline{\gamma_1}$, and hence possibly smaller $\underline{\gamma}$ and $\overline{\gamma}$, compared to the simulation results in Table 3.7 for Case A where the assumed model assumptions for the analysis were fully correct. Also, the data at level $K_2 \times (0.8)$ have decreased, resulting in larger $\underline{\gamma_2}$ and $\overline{\gamma_2}$ and hence possibly larger $\underline{\gamma}$ and $\overline{\gamma}$, in comparison to the simulation results in Table 3.7.

Table 3.11 shows the numbers of the simulation runs with $\underline{\gamma} = \underline{\gamma_1}$ or $\overline{\gamma} = \overline{\gamma_1}$ (or both) for this case considered, out of 10,000 simulation runs with 0.05 level of significance. It shows that, when we use the simulated data for this case, most values come from level K_1 but only few of the $\overline{\gamma}$ come from level K_1 in comparison to the simulation in Table 3.8 which Table 3.11 is related to. Note that, the resulting intervals at level $K_2 \times (0.8)$ become larger, which explains why there are more cases where the $\overline{\gamma}$ comes from $\overline{\gamma}_2$ in comparison to Case A simulation results in Table 3.8. Therefore, in case of worse model fit, the NPI lower and upper survival functions for $q = 0.25, 0.50, 0.75$, using our proposed method discussed in Section 3.3, have more imprecision.

K_1K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9524	0.4640	0.8727	0.6028	0.8599	0.6357
	0.50	0.8432	0.0915	0.7047	0.2859	0.6913	0.3120
	0.75	0.5807	0.0028	0.4629	0.0795	0.4400	0.0831
0.05	0.25	0.9169	0.5211	0.8396	0.6336	0.8339	0.6669
	0.50	0.7686	0.1542	0.6447	0.3336	0.6450	0.3625
	0.75	0.5056	0.0112	0.4019	0.0995	0.3910	0.1193
0.1	0.25	0.8941	0.5413	0.8193	0.6518	0.8218	0.6945
	0.50	0.7230	0.1878	0.6076	0.3521	0.6217	0.4096
	0.75	0.4600	0.0221	0.3752	0.1151	0.3663	0.1618
K_2K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.8828	0.5851	0.8196	0.6791	0.8118	0.6982
	0.50	0.7026	0.2563	0.6096	0.3971	0.6012	0.4088
	0.75	0.4514	0.0447	0.3700	0.1518	0.3472	0.1565
0.05	0.25	0.8524	0.6135	0.7998	0.6960	0.7976	0.7135
	0.50	0.6491	0.3076	0.5780	0.4161	0.5740	0.4368
	0.75	0.3939	0.0862	0.3338	0.1683	0.3186	0.1834
0.1	0.25	0.8356	0.6029	0.7869	0.7038	0.7910	0.7240
	0.50	0.6169	0.3132	0.5587	0.4258	0.5622	0.4584
	0.75	0.3601	0.1266	0.3165	0.1781	0.3068	0.2042
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9539	0.4623	0.8748	0.5958	0.8602	0.6356
	0.50	0.8468	0.0889	0.7081	0.2757	0.6915	0.3118
	0.75	0.5844	0.0027	0.4667	0.0656	0.4402	0.0826
0.05	0.25	0.9216	0.5089	0.8450	0.6303	0.8357	0.6663
	0.50	0.7779	0.1484	0.6522	0.3267	0.6481	0.3614
	0.75	0.5184	0.0222	0.4085	0.0948	0.3938	0.1177
0.1	0.25	0.9017	0.5064	0.8258	0.6476	0.8243	0.6896
	0.50	0.7369	0.1727	0.6182	0.3474	0.6258	0.4008
	0.75	0.4782	0.0580	0.3824	0.1108	0.3714	0.1519

Table 3.9: Proportion of runs with future observation greater than the quartiles, $\beta = 2$ in the fitted model. Case 1.

$K_1 \times (1.2), K_0$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9684	0.5460	0.9179	0.7178	0.8995	**0.7575**
	0.50	0.8544	0.2169	0.7662	0.4495	0.7398	**0.5044**
	0.75	0.5674	0.0501	0.4701	0.1909	0.4467	0.2421
0.05	0.25	0.9487	0.6200	0.9013	0.7468	0.8857	**0.7770**
	0.50	0.8145	0.3172	0.7370	0.4931	0.7183	**0.5373**
	0.75	0.5236	0.1020	0.4466	0.2316	0.4271	**0.2689**
0.1	0.25	0.9353	0.6551	0.8933	**0.7608**	0.8785	**0.7883**
	0.50	0.7872	0.3607	0.7208	**0.5141**	0.7072	**0.5530**
	0.75	0.4972	0.1333	0.4350	0.2499	0.4169	**0.2815**
$K_2 \times (0.8), K_0$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.8679	0.5767	0.7896	0.6642	0.7778	0.6918
	0.50	0.6690	0.2470	0.5596	0.3710	0.5368	0.3960
	0.75	0.4135	0.0481	0.3106	0.1303	0.2788	0.1479
0.05	0.25	0.8379	0.6135	0.7756	0.6793	0.7648	0.7011
	0.50	0.6209	0.3024	0.5377	0.3928	0.5204	0.4171
	0.75	0.3675	0.0793	0.2864	0.1470	0.2637	0.1621
0.1	0.25	0.8232	0.6307	0.7690	0.6869	0.7594	0.7055
	0.50	0.5976	0.3282	0.5260	0.4041	0.5130	0.4233
	0.75	0.3465	0.0971	0.2747	0.1563	0.2557	0.1701
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9807	0.5149	0.9483	0.6253	0.9368	0.6541
	0.50	0.8619	0.1722	0.7977	0.3107	0.7756	0.3328
	0.75	0.5681	0.0182	0.4721	0.0780	0.4498	0.0952
0.05	0.25	0.9675	0.5637	0.9358	0.6430	0.9261	0.6666
	0.50	0.8309	0.2346	0.7764	0.3369	0.7619	0.3503
	0.75	0.5263	0.0393	0.4506	0.0973	0.4332	0.1063
0.1	0.25	0.9585	0.5868	0.9283	0.6505	0.9198	0.6730
	0.50	0.8125	0.2648	0.7635	0.3478	0.7547	0.3592
	0.75	0.5020	0.0536	0.4398	0.1042	0.4239	0.1131

Table 3.10: Proportion of runs with future observation greater than the quartiles. $K_1 \times (1.2)$ and $K_2 \times (0.8)$. Case 2.

min $\underline{\gamma}$ and the max $\overline{\gamma}$	$n = 10$	$n = 50$	$n = 100$
$\underline{\gamma} = \underline{\gamma}_1$	9993	10000	10000
$\overline{\gamma} = \overline{\gamma}_1$	1821	12	5
$\underline{\gamma} = \underline{\gamma}_1$ and $\overline{\gamma} = \overline{\gamma}_1$	1814	12	5

Table 3.11: Number of the simulation runs with $\underline{\gamma} = \underline{\gamma_1}$ or $\overline{\gamma} = \overline{\gamma_1}$ or both out of 10,000 simulation runs with 0.05 level of significance.$K_1 \times (1.2)$ and $K_2 \times (0.8)$. Case 2.

As introduced in Section 2.2, the Eyring model provides an alternative to the Arrhenius model [60]. It also uses temperature as the accelerating variable [50, 60, 71]. Where using this model with temperatures as stress levels, K_0 is the normal temperature (Kelvin) at stress level 0, K_i is the higher temperature (Kelvin) at stress level i, and $\lambda > 0$ is the parameter of the Eyring link function model. The Weibull distributions for different stress levels are assumed to have different scale parameters $\alpha_i > 0$ for level i, but the same shape parameter β. The Eyring link function for the Weibull scale parameters is

$$\alpha_i = \alpha_0 \times (K_0/K_i) \times \exp\left[(\lambda/K_i - \lambda/K_0)\right] \tag{3.5.1}$$

In Case 3, we investigate the robustness and the performance of our predictive inference against the necessary assumptions, where the imprecision in the Arrhenius link function between different stress level provide robustness against the model assumptions. To conduct this, we simulated the data from the Eyring-Weibull model [60] with the parameters $\alpha_0 = 7000, \beta = 3$ and $\lambda = 2000$. This case performs a more likely practical cases, namely where the Arrhenius-Weibull is assumed for the inference but this is actually not fully in line with the data generating mechanism [58].

We conducted the simulation with the assumed Arrhenius-Weibull model for the analysis. In this simulation, the data was generated from the Eyring-Weibull model, hence we hope that our method will provide a good performance. We applied the method described in Section 3.3, with 10, 000 simulation runs. To illustrate this,

we generated data sets as before with $n = 10, 50, 100$ observations at each stress level using the Eyring-Weibull model. We used the similar scenario as in the Case A simulations where the model assumptions were fully correct but with data generated from the Eyring-Weibull model. Using these generated data, we again applied our method as described in Section 3.3, with levels of significance 0.01, 0.05, and 0.10 and we checked whether or not the future observation at the normal stress level K_0 has exceeded these quartiles in the right proportion.

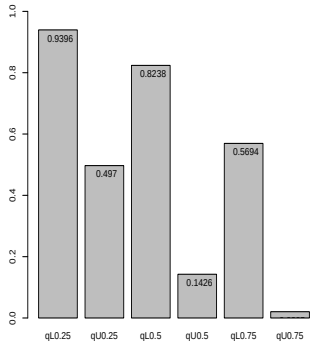
In comparison with the simulation where the model assumptions are fully correct, the results in Table 3.10 are similar to those in Table 3.7. Also, we can see the similarity of these results from Figures 3.5-3.7 and Figures 3.8-3.10 of the quartiles of NPI lower and upper survival functions for $q = 0.25, 0.50, 0.75$. These results show that the proposed approach provides robustness in predictive inference against the model assumptions in case of this specific model misspecification.

The main findings drawn from the above simulations are: the future observation at the normal stress level K_0 has exceeded the quartiles that we considered in the right proportions. Using our approach with both the Arrhenius-Weibull model and the power-Weibull model, achieves suitable predictive inference if the model assumptions are fully correct and the end resulting intervals $[\underline{\gamma}, \overline{\gamma}]$ have reasonable imprecision. However, in the case of model misspecification, the end resulting intervals $[\underline{\gamma}, \overline{\gamma}]$ have wider imprecision compared to if the model assumptions are correct. One can similarly investigate other cases of model misspecification. Of course, in the case of huge misspecification, no method would give meaningful inferences; in our model it would most likely lead to large imprecision, which would reflect that there is a problem of model fit. We have seen that when the number of observations at each stress level is $n = 100$, the imprecision between the NPI lower and upper survival functions tends to decrease compared to when the number of observations at each stress level $n = 10$.

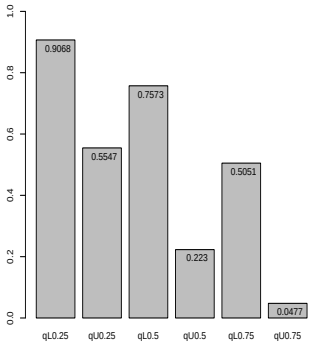
K_1K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9396	0.4970	0.8583	0.6306	0.8332	0.6718
	0.50	0.8238	0.1426	0.6762	0.3254	0.6426	0.3706
	0.75	0.5694	0.0205	0.4338	0.0938	0.3889	0.1251
0.05	0.25	0.9068	0.5547	0.8355	0.6612	0.8162	0.6942
	0.50	0.7573	0.2230	0.6350	0.3700	0.6089	0.4007
	0.75	0.5051	0.0477	0.3988	0.1264	0.3570	0.1544
0.1	0.25	0.8887	0.5822	0.8244	0.6763	0.8084	0.7035
	0.50	0.7184	0.2630	0.6176	0.3908	0.5944	0.4165
	0.75	0.4692	0.0694	0.3785	0.1457	0.3418	0.1686
K_2K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.8782	0.6038	0.8072	0.6911	0.7959	0.7116
	0.50	0.6965	0.2848	0.5905	0.4053	0.5704	0.4336
	0.75	0.4571	0.0648	0.3486	0.1576	0.3140	0.1820
0.05	0.25	0.8529	0.6407	0.7943	0.7054	0.7875	0.7204
	0.50	0.6497	0.3381	0.5673	0.4271	0.5540	0.4501
	0.75	0.4047	0.1020	0.3265	0.1775	0.2970	0.1971
0.1	0.25	0.8379	0.6558	0.7871	0.7131	0.7822	0.7243
	0.50	0.6261	0.3618	0.5571	0.4393	0.5455	0.4578
	0.75	0.3795	0.1222	0.3153	0.1883	0.2902	0.2056
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9433	0.4934	0.8583	0.6286	0.8338	0.6705
	0.50	0.8281	0.1366	0.6777	0.3229	0.6435	0.3690
	0.75	0.5749	0.0153	0.4349	0.0916	0.3902	0.1238
0.05	0.25	0.9133	0.5472	0.8376	0.6583	0.8176	0.6915
	0.50	0.7671	0.2116	0.6384	0.3650	0.6117	0.3975
	0.75	0.5142	0.0373	0.4016	0.1222	0.3592	0.1503
0.1	0.25	0.8966	0.5723	0.8261	0.6727	0.8103	0.6999
	0.50	0.7334	0.2495	0.6229	0.3848	0.5980	0.4116
	0.75	0.4800	0.0566	0.3831	0.1386	0.3445	0.1634

Table 3.12: Proportion of runs with future observation greater than the quartiles, simulation from Eyring model. Case 3.

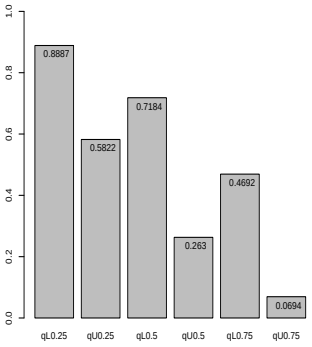
(a) K_1 and K_0 (0.01)



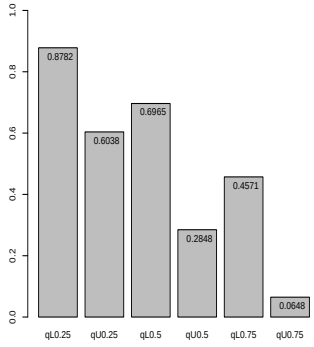
(b) K_1 and K_0 (0.05)



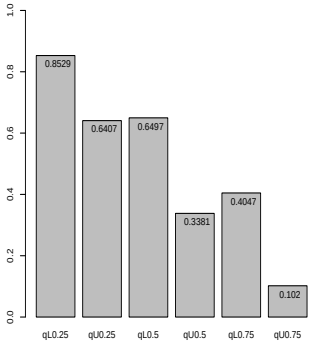
(c) K_1 and K_0 (0.1)



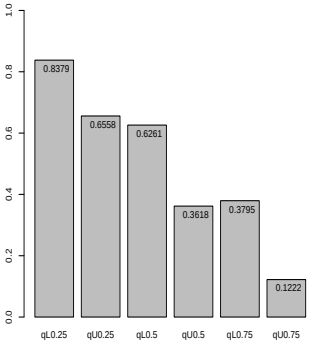
(d) K_2 and K_0 (0.01)



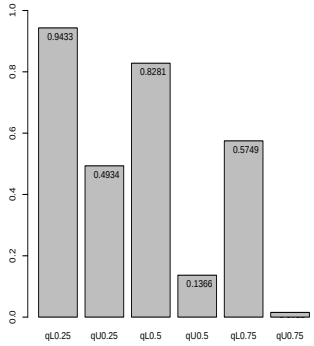
(e) K_2 and K_0 (0.05)



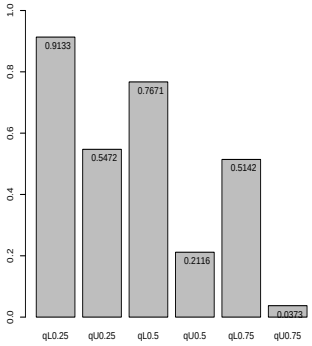
(f) K_2 and K_0 (0.1)



(g) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.01)



(h) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.05)



(i) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.1)

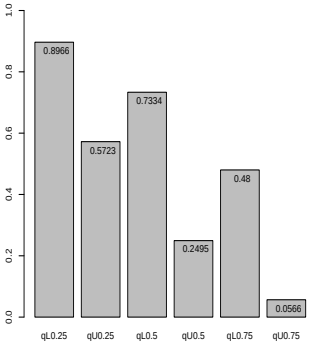
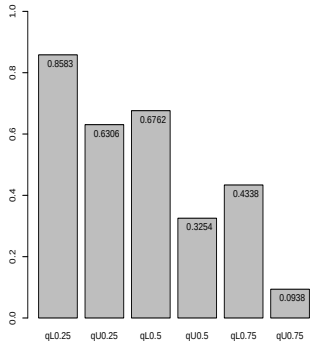
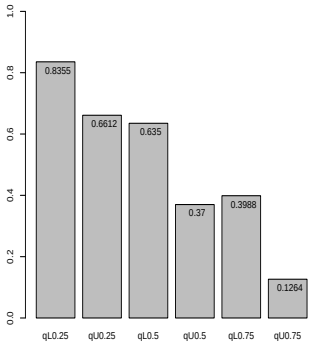


Figure 3.8: Proportion of runs with future observation greater than the quartiles, $n = 10$. Case 3

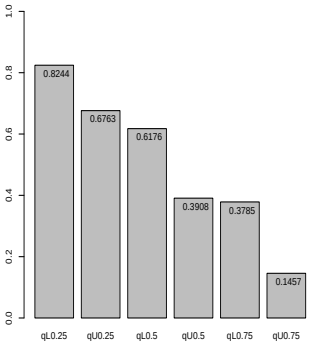
(a) K_1 and K_0 (0.01)



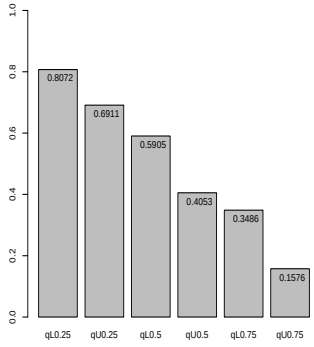
(b) K_1 and K_0 (0.05)



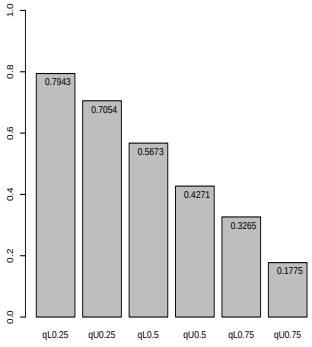
(c) K_1 and K_0 (0.1)



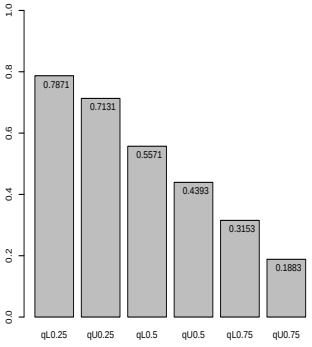
(d) K_2 and K_0 (0.01)



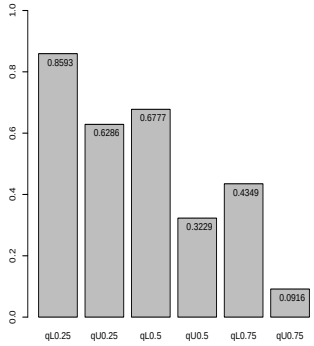
(e) K_2 and K_0 (0.05)



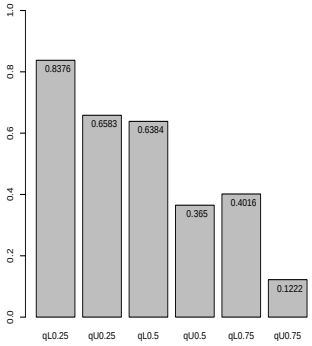
(f) K_2 and K_0 (0.1)



(g) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.01)



(h) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.05)



(i) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.1)

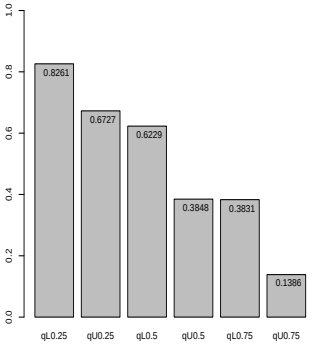
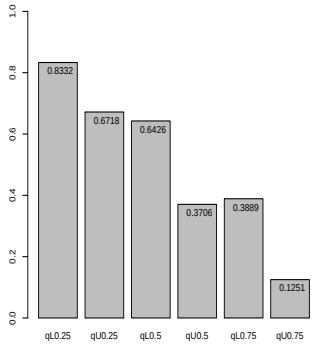
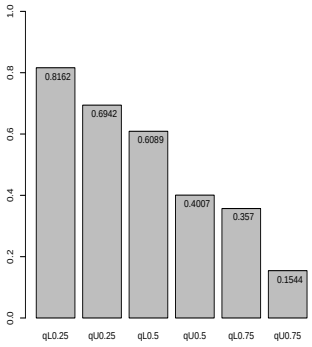


Figure 3.9: Proportion of runs with future observation greater than the quartiles, $n = 50$. Case 3

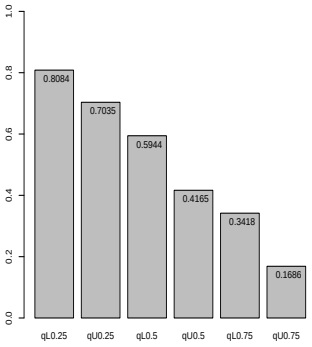
(a) K_1 and K_0 (0.01)



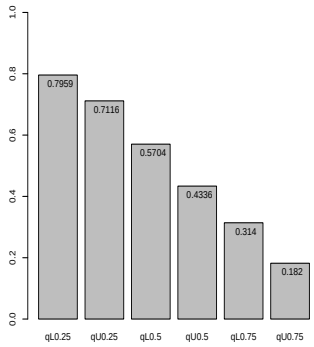
(b) K_1 and K_0 (0.05)



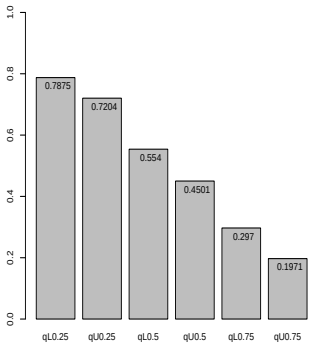
(c) K_1 and K_0 (0.1)



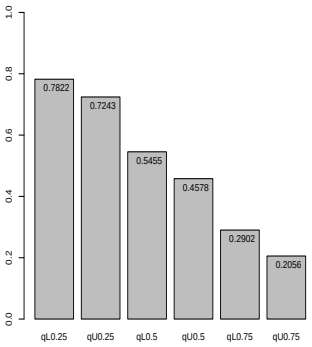
(d) K_2 and K_0 (0.01)



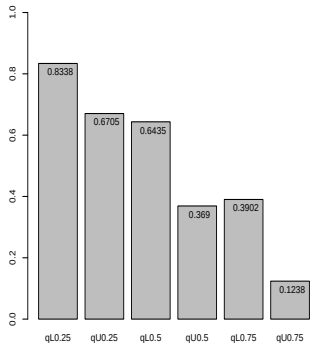
(e) K_2 and K_0 (0.05)



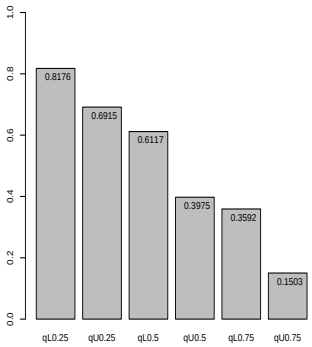
(f) K_2 and K_0 (0.1)



(g) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.01)



(h) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.05)



(i) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.1)

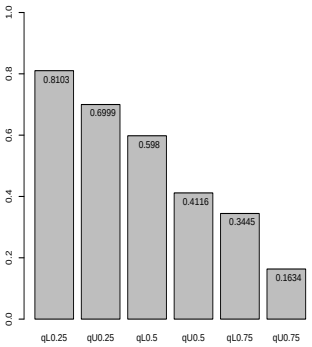


Figure 3.10: Proportion of runs with future observation greater than the quartiles, $n = 100$. Case 3

3.6 Inference using the power-law link function

Yin et al. [79], introduced the statistical approach on which our work is based, as discussed in Section 3.3. They used the power-law link function. In addition to our use of the likelihood ratio test to determine the imprecision, we also use this model to illustrate the use of our method without the assumption of equal shape parameter for the Weibull distribution at all stress levels. In this section we apply our method with the different scale parameters α_i and different shape parameters β_i for the Weibull distributions for the different stress levels i. The power-law link function for scale parameters α_i should be identified to establish a connection between the different stress levels i. Regarding to this link function model, an observation t^i at the stress level i, subject to stress K_i, can be transformed to stress level 0, with the transformed to an observation from level i to level 0 represented by the equation

$$t^{i\to 0} = t^i\Big(\frac{K_i}{K_0}\Big)^{\gamma}, \tag{3.6.1}$$

as used by Yin et al. [79] also is introduced briefly in Section 2.2.

Moreover, a transformation link function, generalizing Equation (3.6.1), can also be derived if we allow different shape parameters β_i for each level i, so our method can be generalized in this way as well. Then we get

$$t^{i\to 0} = \alpha_0\Big[\Big(\frac{t^i}{\alpha_0\Big(\frac{K_0}{K_i}\Big)^{\gamma}}\Big)^{\beta_i}\Big]^{\frac{1}{\beta_0}}. \tag{3.6.2}$$

The proposed method in Section 3.3 is illustrated by two examples using simulated data, but we use the power-law link function.

Example 3.6.1. In this example we present two cases, similar to Example 3.4. In Case 1 we simulate data at all levels using a Weibull distribution at each stress level and the power-law link function we assume for the analysis. In Case 2 we change these data such that the assumed link function will not provide a good fit anymore and we investigate the effect on the interval $[\underline{\gamma}, \overline{\gamma}]$ and on the corresponding lower and upper predictive survival functions for a future observation at the normal stress level.

Case	$K_0 = 50$	$K_1 = 53$	$K_2 = 59$	$K_1 = 53$ ($\times 1.2$)	$K_2 = 59$ ($\times 0.5$)
1	384.657	112.266	77.804	134.719	38.902
2	458.334	352.582	99.120	423.098	49.559
3	474.970	357.112	144.012	428.534	72.006
4	745.488	386.581	187.932	463.898	93.966
5	773.865	402.811	188.568	483.373	94.284
6	822.844	495.271	195.895	594.326	97.948
7	996.161	550.243	209.573	660.292	104.787
8	1098.651	552.277	221.141	662.733	110.571
9	1169.866	650.370	243.699	780.444	121.849
10	1406.782	670.863	311.014	805.035	155.507

Table 3.13: Failure times at three voltage levels (first three columns) and changed failure times (last two columns), Example 3.6.1.

In this example, we assume the voltage levels to be $K_0 = 50$, $K_1 = 53$ and $K_2 = 59$ kilovolts. We generate ten observations from a Weibull distribution at each stress level, linked with the power-law link function. The Weibull distribution at K_0 has $\beta = 3$ and $\alpha_0 = 1000$, and the power-law parameter is set at $\gamma = 10$. In this example we assume $\beta_i = \beta$ for all levels i, for which we used Equation 3.6.1, so this model keeps the same shape parameter at each voltage, and the scale parameters are linked by the power-law relation, which leads to $\alpha_1 = 558.3948$ at K_1 and $\alpha_2 = 191.0645$ at K_2. There are in total three parameters that need to be estimated, α_0, β, and γ. Ten units were tested at each voltage level, so a total of 30 failure times are simulated for this example. The failure times are given in Table 3.13.

We use the pairwise likelihood ratio test between the stress levels K_i and K_0 to obtain the intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ of the values γ for which we do not reject the null hypothesis. The intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ are given in Table 3.14.

The NPI lower and upper survival functions are based on the original data at level 0 together with the transformed data using $\underline{\gamma}$ and $\overline{\gamma}$, respectively. In Case 1, we have $\underline{\gamma} = 3.374$ and the $\overline{\gamma} = 18.179$, $\underline{\gamma} = 5.495$ and $\overline{\gamma} = 16.167$, and $\underline{\gamma} = 6.476$ and $\overline{\gamma} = 15.234$ for significance levels 0.01, 0.05, and 0.1, respectively. These values

Significance level	0.01		0.05		0.1	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
Case 1: K_1 K_0	3.374	18.179	5.495	16.167	6.476	15.234
K_2 K_0	6.516	11.703	7.250	10.994	7.590	10.664
Case 2: $K_1 \times (1.2)$, K_0	0.246	15.050	2.366	13.038	3.347	12.105
$K_2 \times (0.5)$, K_0	10.704	15.891	11.437	15.182	11.777	14.852

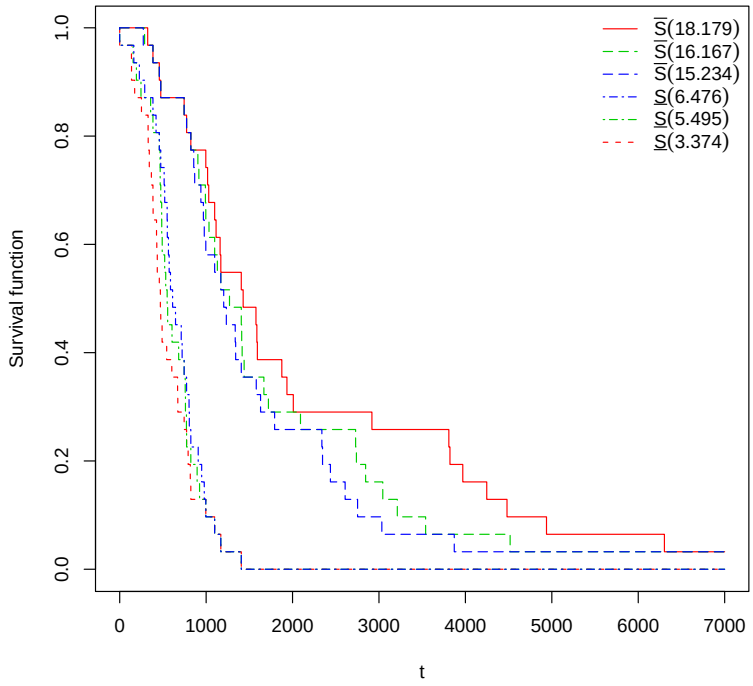
Table 3.14: $[\underline{\gamma}_i, \overline{\gamma}_i]$ for Example 3.6.1.

are all equal to the corresponding $\underline{\gamma}_1$ and $\overline{\gamma}_1$ of the pairwise comparison of K_1 and K_0. We used all the above $\underline{\gamma}$ and $\overline{\gamma}$ values to transform the data to the normal stress level 0, see Figure 3.11(a). As mentioned in Section 3.3, if the model fits perfectly, in most cases we will have $\underline{\gamma} = \underline{\gamma}_1$ and $\overline{\gamma} = \overline{\gamma}_1$, which is therefore also illustrated here. If the model does not fit well, then it is most likely that $\underline{\gamma}$ (or $\overline{\gamma}$) is equal to $\underline{\gamma}_i$ (or $\overline{\gamma}_i$) from another stress level.

In Case 2, we illustrate our method in case of misspecification, to study whether or not a simple model will give robustness in case of misspecification. We multiply the data at level K_1 by 1.2 and we multiply the data at level K_2 by 0.5. The simulated data values are given in the last two columns in Table 3.13. We have $\underline{\gamma} = 0.246$ and $\overline{\gamma} = 15.891$, $\underline{\gamma} = 2.366$ and $\overline{\gamma} = 15.182$, and $\underline{\gamma} = 3.347$ and $\overline{\gamma} = 14.852$ at significance levels 0.01, 0.05, and 0.1, respectively. Now, the $\underline{\gamma}$ values are all equal to the corresponding $\underline{\gamma}_1$, but the $\overline{\gamma}$ values are all equal to the corresponding $\overline{\gamma}_2$. We used all the above $\underline{\gamma}$ and $\overline{\gamma}$ values to transform the data to the normal stress level 0, the corresponding lower and upper survival functions are presented in Figure 3.11(b). In these figures, the lower survival function $\underline{S}$ is labeled as $\underline{S}$ $(\underline{\gamma}_i)$ and the upper survival function $\overline{S}$ is labeled as $\overline{S}$ $(\overline{\gamma}_i)$. These figures show that smaller significance level leads to more imprecision for the NPI lower and upper survival functions. Note that in Case 2, the observations at level K_1 have increased, leading to smaller $\underline{\gamma}_1$ and $\overline{\gamma}_1$ values, and this leads to the lower and upper survival functions decreasing in comparison to Case 1. Also, the observations at K_2 stress level in Case 2, have decreased, resulting in larger values for $\underline{\gamma}_2$ and $\overline{\gamma}_2$, and this leads to the lower and upper survival functions increasing in comparison to Case 1.

As mentioned in Section 3.3, in case of poor model fit, the NPI lower and upper survival functions in Figures 3.11(b), have more imprecision than in Case 1 where the model fit was perfectly.

(a) Case 1



(b) Case 2

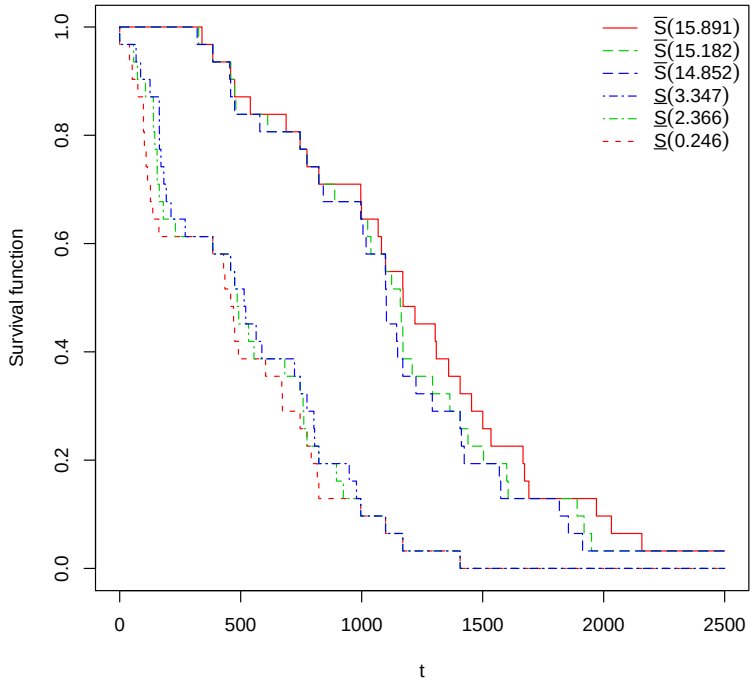


Figure 3.11: The NPI lower and upper survival functions, Example 3.6.1.

Case	$K_0 = 50$	$K_1 = 53$	$K_2 = 59$
1	384.657	81.454	59.190
2	458.334	321.606	81.174
3	474.970	326.571	132.139
4	745.488	359.170	186.989
5	773.865	377.340	187.815
6	822.844	483.530	197.389
7	996.161	548.627	215.555
8	1098.651	551.062	231.203
9	1169.866	670.509	262.431
10	1406.782	695.941	360.729

Table 3.15: Failure times at three voltage levels, Example 3.6.2.

Example 3.6.2. Example 3.6.1 illustrated our methods with the assumption of equal shape parameter for the Weibull distributions at all stress levels using the power-law link function. In this example, we also use this model to illustrate the use of our method but without the assumption of equal shape parameter for the Weibull distribution at all stress levels using Equation 3.6.2. Note that we use the same voltage stress levels as in Example 3.6.1. We simulated ten failure times at each of three stress levels from the Weibull distribution and using different shape parameters β_0, β_1, and β_2 for stress levels K_0, K_1 and K_2, respectively.

The Weibull distribution at K_0 has $\beta_0 = 3$, at K_1 has $\beta_1 = 2.5$, and at K_2 has $\beta_2 = 2.3$ and $\alpha_0 = 1000$, and the power-law parameter sets at $\gamma = 10$. Ten failure times are simulated at each voltage level. The failure times are given in Table 3.15.

Using our method, the intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ are given in Table 3.16, using Equation 3.6.2. This analysis led to parameters estimates of $\hat{\beta}_0 = 2.884081$, $\hat{\beta}_1 = 2.714822$, $\hat{\alpha_0} = 937.854795$, and $\hat{\gamma} = 11.006434$, resulting from the pairwise levels K_1 and K_0. Similarly, this analysis led to parameters estimates of $\hat{\beta}_0 = 2.884084$, $\hat{\beta}_2 = 2.479445$, $\hat{\alpha_0} = 937.854418$, and $\hat{\gamma} = 8.872149$, resulting from the pairwise levels K_2 and K_0.

For the significance levels $0.01, 0.05, 0.1$, the values of $\underline{\gamma}$ are 2.765 corresponding

Significance level	0.01		0.05		0.1	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
K_1 K_0	2.765	19.245	5.010	16.997	6.064	15.944
K_2 K_0	5.845	11.941	6.664	11.104	7.049	11.941

Table 3.16: $[\underline{\gamma_i}, \overline{\gamma_i}]$ for Example 3.6.2.

to parameter estimates $\hat{\beta}_0 = 2.182308$ and $\hat{\beta}_1 = 2.410073$, 5.010 corresponding to $\hat{\beta}_0 = 2.361527$ and $\hat{\beta}_1 = 2.620797$, and 6.064 corresponding to $\hat{\beta}_0 = 2.461963$ and $\hat{\beta}_1 = 2.689885$ in this case, and for $\overline{\gamma}$ we have, 19.245 corresponding to parameter estimates $\hat{\beta}_0 = 2.802078$ and $\beta_1 = 1.870777$, 16.997 corresponding to $\hat{\beta}_0 = 2.891816$ and $\hat{\beta}_1 = 2.141687$, and 15.944 corresponding to $\hat{\beta}_0 = 2.928766$ and $\hat{\beta}_1 = 2.265410$, respectively. Therefore, based on the original data at level 0 together with the data transformed from the stress levels K_1 to K_0 and K_2 to K_0 using $\underline{\gamma}$ and the corresponding to parameter estimates $\hat{\beta}_i$, the NPI approach provides the NPI lower survival function. Similarly, but using $\overline{\gamma}$ for the transformation, we derived the NPI upper survival function, see Figure 3.12. These cases are all equal to the $\underline{\gamma}_1$ and $\overline{\gamma}_1$ of the pairwise comparison of K_1 and K_0. Comparing this example with Example 3.6.1, we notice that there is slightly more imprecision in the resulting intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ in this example when we assumed to have different ($\beta_i = 3$, 2.5, 2.3) for stress levels K_0, K_1 and K_2, respectively, than in Example 3.6.1, where we assumed $\beta_i = \beta$ for all levels i, which was indeed used for the simulation. Note that, in Example 3.6.1 the data was generated using same shape parameter $\beta = 3$, but in this example we used different shape parameters $\beta_i = 3, 2.5, 2.3$.

3.7 Simulation studies

To investigate the predictive performance of our new inference method for ALT data, we conducted simulation studies. We assumed the voltage stress levels to be $K_0 = 40$, $K_1 = 50$, and $K_2 = 55$ kilovolts. In this analysis, we ran the simulation 10,000 times with the data simulated from the Weibull distribution and using the scale parameter $\alpha_0 = 7000$, and the link function parameter $\gamma = 10$, us-

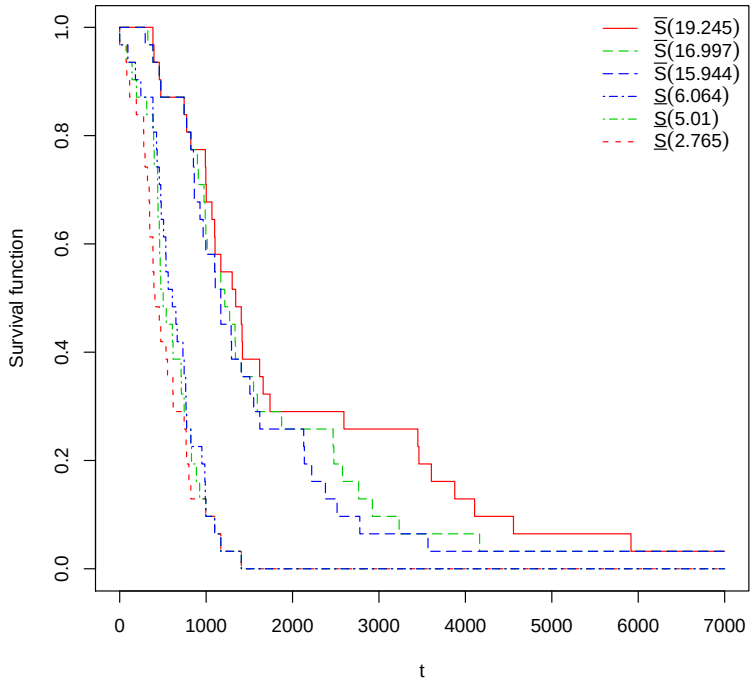


Figure 3.12: The NPI lower and upper survival functions, Example 3.6.2.

ing the assumed power-law link function model between different stress levels, with $n = 10, 30, 50$ data at each stress level. We applied similar method described in Section 3.5, with levels of significance $0.01, 0.05, 0.1$, but now using the power-law link function Equation 3.6.2 between different stress level. In addition, we also simulated one more observation at the normal stress level 0, which is used to as a future observation to investigate the predictive performance. For good performance of our method, we require that the future observation for each run at the normal stress level exceeds the first, second, and third quartiles of the NPI lower survival functions just over proportions $0.75, 0.50$ and 0.25 of all runs, respectively, and that the future observation exceeds the first, second, and third quartiles for the NPI upper survival functions just under proportions $0.75, 0.50$ and 0.25.

In Case A, we investigated the predictive performance of our new method for ALT data where we assume different shape parameters β_i for different levels. In this analysis, we ran the simulation 10,000 times with the data simulated from the Weibull distribution while using different shape parameters $\beta_i = (\beta_0, \beta_1, \beta_2)$ for stress

levels K_0, K_1 and K_2, respectively.

To illustrate this, we generated new data set as before with $n = 10, 30, 50$ observations at each stress levels using different shape parameters $\beta_i = 3$, 2.5, 2.3, for stress levels K_0, K_1 and K_2, respectively, and assumed the power-Weibull model for the analysis, where the parameters α_0, β_0, β_1, β_2, and the γ parameter of the link function were estimated. This is similar as done in Example 3.6.2 to illustrate our method in case of perfect model fit.

We applied our inferential method according to the null hypothesis we used. Then we used these $\underline{\gamma}$ and $\overline{\gamma}$ values to transform the data to the normal stress level K_0 using Equation 3.6.2. Note that, the shape parameters β_i were estimated in the analysis. There are estimating values for the shape parameters $\beta_i = (\beta_0, \beta_1, \beta_2)$ with each corresponding $\underline{\gamma}_i$ and $\overline{\gamma}_i$, respectively, out of 10.000 simulation runs. Table 3.17 presents the results of these simulations with $n = 10, 30, 50$, with attention on prediction of the simulated future observation at level K_0. Table 3.17 shows an overall good performance of the proposed method. From these proportions, which are indeed achieved, we conclude that the proposed approach using the power-Weibull model provides sufficient predictive inference if the model assumptions are fully correct.

Now, we investigate robustness in case of model misspecification in the next three simulation cases. In Case 1, we investigate the predictive performance of our method for ALT data, where we assumed different shape parameters for the generated data. Table 3.18 presents the results for this simulation when we generated the data using different shape parameters $\beta_0 = 3$, $\beta_1 = 2.5$, and $\beta_2 = 2.3$ for stress levels K_0, K_1 and K_2, respectively. However, in the fitted model using the power-Weibull model we only assumed a single estimated shape parameter β for all levels, hence only the parameters α_0, β, and the γ parameter of the link function were estimated by the MLE method. There is more imprecision on the quartiles of NPI lower and upper survival functions for $q = 0.25, 0.50, 0.75$ in comparison to the Case A where the shape parameters β_0, β_1, and β_2 corresponding to the stress levels K_0, K_1 and K_2, respectively, were estimated in the fitted model for the analysis.

All the results in Case 1 provide an insight into whether or not the presented

method shows a level of robustness against the misspecification case considered. Note that, in Table 3.18 for $n = 50$ with 0.1 level of significance, there is one case for which the future observation for each run at the normal stress level has exceeded the first quartile ($qU0.25$) corresponding to the NPI upper survival functions just over 0.75 resulting from the pairwise levels K_2 to K_0. It is highlighted by use of the bold font in Table 3.18. In this case for the pairwise levels K_2 to K_0 we only transformed the observations at the stress level K_2 to the normal stress level K_0 while the observations at the normal stress level K_0 were not transformed with $\beta_0 = 3$. In most cases in this simulation, the estimate of this single β in the fitted model was less than 2.5, and because the observations at level K_0, with $\beta_0 = 3$, were not transformed, therefore the β_0 for the normal stress level influenced the estimation which caused the future observation at the normal stress to slightly exceed the first quartile ($qU0.25$) of the NPI upper survival functions just over the 0.75. There is a slight increase in $qU0.25$, which means that we have too many observations passing this quartile from the upper 0.75. It seems that the point 0.75 occurred a bit earlier and it should be related to the effect of estimating only a single β for the stress levels K_2 to K_0 in the fitted model while the data were generated using different β_i. As mentioned, this is in line with our expectation, which is mainly due to the misspecification case we assumed. Note that, the smaller significance level leads to more imprecision for the NPI lower and upper survival functions.

In Case 2, we investigated the predictive performance of our method for ALT data where we assumed different shape parameters β_i for the generated data. But, in the fitted model using the power-Weibull model we only assumed a single fixed shape parameter $\beta = 2$ for all levels, hence only the parameters α_0 and the γ parameter of the link function were estimated. Table 3.19 shows the results with same scenario as in the previous simulation study. In comparison to the previous case considered where we assumed the shape parameter as a single estimated β in the fitted model for the analysis, there are more imprecision on the quartiles of NPI lower and upper survival functions for $q = 0.25, 0.50, 0.75$. In Case 2 most of $[\underline{\gamma}, \overline{\gamma}]$ have more imprecision than in Case 1, so we see again the method works well in the sense that it leads to more imprecision if it is more misspecified.

In Case 3, we also look at what happens if we set the shape parameter $\beta = 3$ for all levels in Case 2 instead of the shape parameter $\beta = 2$. Then there are few cases for which the future observation for each run at the normal stress level has exceeded the first quartile ($qU0.25$) corresponding to the NPI upper survival functions just over 0.75 of the pairwise level K_2 and K_0 compared to Case 1, see Table 3.20. These simulations show that the proposed approach provides robustness in the predictive inference against the mistake in the model assumptions in case of these specific model misspecifications.

In summary, all cases in the above simulations show an overall good performance of the proposed method, with levels of significance 0.01, 0.05 and 0.1, and with sample sizes $n = 10, 30, 50$. Note that, in all cases the data was generated using different shape parameters $\beta_i = 3$, 2.5, 2.3 at stress levels K_0, K_1, and K_2, respectively, but in the analysis we used different scenarios as presented. In Cases 1, 2, and 3, from Tables 3.18-3.20, the imprecision in the resulting intervals $[\underline{\gamma}, \overline{\gamma}]$ becomes wider than in Case A. This happened when we made the mistake of only assuming a single estimated shape parameter β for all levels in Case 1, a single fixed shape parameter $\beta = 2$ for all levels in Case 2, and a single fixed shape parameter $\beta = 3$ for all levels in Case 3. In Case A, however, we assumed different shape parameters β_i for each stress level.

K_1K_0		$n=10$		$n=30$		$n=50$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9077	0.5332	0.8525	0.6296	0.8326	0.6532
	0.50	0.7655	0.1940	0.6726	0.3154	0.6339	0.3618
	0.75	0.5407	0.0311	0.4299	0.0917	0.4005	0.1201
0.05	0.25	0.8790	0.5892	0.8333	0.6600	0.8148	0.6784
	0.50	0.7040	0.2664	0.6339	0.3571	0.6050	0.3931
	0.75	0.4751	0.0655	0.3881	0.1249	0.3703	0.1469
0.1	0.25	0.8620	0.6152	0.8231	0.6748	0.8049	0.6894
	0.50	0.6733	0.3002	0.6139	0.3789	0.5880	0.4102
	0.75	0.4396	0.0870	0.3650	0.1421	0.3495	0.1620
K_2K_0		$n=10$		$n=30$		$n=50$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.8842	0.5826	0.8351	0.6572	0.8174	0.6787
	0.50	0.7142	0.2562	0.6333	0.3573	0.6049	0.3905
	0.75	0.4876	0.0529	0.3902	0.1182	0.3687	0.1464
0.05	0.25	0.8588	0.6246	0.8182	0.6792	0.8022	0.6970
	0.50	0.6648	0.3165	0.6055	0.3907	0.5822	0.4154
	0.75	0.4274	0.0878	0.3567	0.1468	0.3422	0.1679
0.1	0.25	0.8452	0.6426	0.8079	0.6925	0.7933	0.7056
	0.50	0.6412	0.3471	0.5886	0.4083	0.5698	0.4281
	0.75	0.3993	0.1096	0.3387	0.1636	0.3295	0.1787
$\underline{\gamma}$ and $\overline{\gamma}$		$n=10$		$n=30$		$n=50$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9158	0.5171	0.8573	0.6207	0.8365	0.6477
	0.50	0.7785	0.1688	0.6803	0.3036	0.6411	0.3519
	0.75	0.5570	0.0182	0.4406	0.0801	0.4079	0.1115
0.05	0.25	0.8900	0.5704	0.8398	0.6499	0.8205	0.6717
	0.50	0.7227	0.2384	0.6464	0.3418	0.6147	0.3824
	0.75	0.4970	0.0435	0.4013	0.1116	0.3789	0.1377
0.1	0.25	0.8757	0.5934	0.8306	0.6653	0.8106	0.6823
	0.50	0.6960	0.2748	0.6282	0.3619	0.5982	0.3977
	0.75	0.4636	0.0619	0.3808	0.1290	0.3617	0.1500

Table 3.17: Proportion of runs with future observation greater than the quartiles. Correct model, Case A.

K_1K_0		$n=10$		$n=30$		$n=50$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9413	0.5657	0.8949	0.6756	0.8792	0.7025
	0.50	0.8107	0.1990	0.7213	0.3333	0.6814	0.3844
	0.75	0.5547	0.0234	0.4469	0.0795	0.4155	0.1101
0.05	0.25	0.9164	0.6259	0.8759	0.7077	0.8643	0.7248
	0.50	0.7474	0.2790	0.6812	0.3817	0.6492	0.4209
	0.75	0.4868	0.0560	0.4011	0.1143	0.3819	0.1404
0.1	0.25	0.9017	0.6546	0.8664	0.7244	0.8563	0.7364
	0.50	0.7168	0.3187	0.6599	0.4054	0.6324	0.4395
	0.75	0.4513	0.0780	0.3786	0.1341	0.3641	0.1570
K_2K_0		$n=10$		$n=30$		$n=50$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9215	0.6203	0.8814	0.7070	0.8663	0.7293
	0.50	0.7585	0.2681	0.6806	0.3805	0.6535	0.4179
	0.75	0.5014	0.0415	0.4057	0.1099	0.3842	0.1389
0.05	0.25	0.8998	0.6649	0.8662	0.7285	0.8551	0.7451
	0.50	0.7077	0.3358	0.6489	0.4200	0.6270	0.4468
	0.75	0.4426	0.0786	0.3691	0.1413	0.3569	0.1631
0.1	0.25	0.8892	0.6885	0.8588	0.7411	0.8474	**0.7528**
	0.50	0.6848	0.3691	0.6343	0.4395	0.6152	0.4627
	0.75	0.4123	0.1007	0.3505	0.1571	0.3419	0.1752
$\underline{\gamma}$ and $\overline{\gamma}$		$n=10$		$n=30$		$n=50$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9489	0.5519	0.8998	0.6638	0.8837	0.6979
	0.50	0.8268	0.1767	0.7264	0.3240	0.6904	0.3761
	0.75	0.5732	0.0130	0.4592	0.0720	0.4249	0.1031
0.05	0.25	0.9280	0.6070	0.8832	0.6920	0.8705	0.7190
	0.50	0.7712	0.2553	0.6878	0.3652	0.6608	0.4101
	0.75	0.5144	0.0366	0.4196	0.1058	0.3924	0.1304
0.1	0.25	0.9156	0.6343	0.8758	0.7096	0.8627	0.7295
	0.50	0.7454	0.2936	0.6682	0.3890	0.6465	0.4273
	0.75	0.4808	0.0544	0.3962	0.1246	0.3755	0.1443

Table 3.18: Proportion of runs with future observation greater than the quartiles. Case 1.

K_1K_0		$n=10$		$n=30$		$n=50$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9377	0.4939	0.8776	0.6039	0.8580	0.6357
	0.50	0.8309	0.1399	0.7212	0.2784	0.6770	0.3343
	0.75	0.6038	0.0105	0.4869	0.0648	0.4419	0.0976
0.05	0.25	0.9103	0.5489	0.8551	0.6403	0.8361	0.6638
	0.50	0.7670	0.2205	0.6774	0.3306	0.6390	0.3769
	0.75	0.5352	0.0346	0.4335	0.1029	0.4041	0.1324
0.1	0.25	0.8938	0.5796	0.8426	0.6601	0.8253	0.6801
	0.50	0.7296	0.2591	0.6529	0.3587	0.6199	0.3967
	0.75	0.4992	0.0563	0.4061	0.1224	0.3863	0.1511
K_2K_0		$n=10$		$n=30$		$n=50$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9077	0.5602	0.8549	0.6492	0.8351	0.6709
	0.50	0.7611	0.2270	0.6689	0.3384	0.6364	0.3810
	0.75	0.5291	0.0348	0.4288	0.1079	0.4006	0.1385
0.05	0.25	0.8835	0.6046	0.8361	0.6734	0.8195	0.6938
	0.50	0.7068	0.2896	0.6350	0.3791	0.6094	0.4111
	0.75	0.4746	0.0680	0.3893	0.1396	0.3714	0.1639
0.1	0.25	0.8707	0.6264	0.8267	0.6872	0.8108	0.7041
	0.50	0.6825	0.3216	0.6174	0.4004	0.5944	0.4263
	0.75	0.4443	0.0888	0.3679	0.1576	0.3568	0.1753
$\underline{\gamma}$ and $\overline{\gamma}$		$n=10$		$n=30$		$n=50$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9405	0.4888	0.8802	0.6008	0.8597	0.6328
	0.50	0.8365	0.1321	0.7253	0.2740	0.6811	0.3303
	0.75	0.6110	0.0079	0.4921	0.0614	0.4451	0.0945
0.05	0.25	0.9161	0.5408	0.8596	0.6353	0.8398	0.6608
	0.50	0.7775	0.2066	0.6843	0.3231	0.6462	0.3705
	0.75	0.5501	0.0259	0.4433	0.0960	0.4101	0.1275
0.1	0.25	0.9025	0.5690	0.8489	0.6546	0.8296	0.6758
	0.50	0.7453	0.2449	0.6619	0.3487	0.6286	0.3892
	0.75	0.5164	0.0430	0.4173	0.1156	0.3935	0.1434

Table 3.19: Proportion of runs with future observation greater than the quartiles. Case 2.

K_1K_0		$n=10$		$n=30$		$n=50$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9369	0.5912	0.8920	0.6894	0.8757	0.7138
	0.50	0.7938	0.2395	0.7123	0.3569	0.6747	0.4022
	0.75	0.5341	0.0321	0.4337	0.0953	0.4070	0.1237
0.05	0.25	0.9158	0.6383	0.8735	0.7176	0.8624	0.7332
	0.50	0.7410	0.3011	0.6747	0.3991	0.6449	0.4334
	0.75	0.4760	0.0646	0.3942	0.1264	0.3767	0.1513
0.1	0.25	0.9013	0.6627	0.8647	0.7305	0.8549	0.7425
	0.50	0.7120	0.3350	0.6564	0.4175	0.6305	0.4487
	0.75	0.4407	0.0864	0.3733	0.1448	0.3587	0.1657
K_2K_0		$n=10$		$n=30$		$n=50$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9157	0.6485	0.8775	0.7222	0.8643	0.7403
	0.50	0.7412	0.3110	0.6708	0.4059	0.6456	0.4391
	0.75	0.4742	0.0632	0.3932	0.1309	0.3756	0.1543
0.05	0.25	0.8969	0.6839	0.8635	0.7409	0.8537	**0.7537**
	0.50	0.6988	0.3623	0.6440	0.4378	0.6252	0.4616
	0.75	0.4281	0.0966	0.3624	0.1568	0.3533	0.1766
0.1	0.25	0.8883	0.7020	0.8581	**0.7507**	0.8460	**0.7606**
	0.50	0.6781	0.3860	0.6306	0.4547	0.6143	0.4739
	0.75	0.4024	0.1174	0.3461	0.1698	0.3391	0.1881
$\underline{\gamma}$ and $\overline{\gamma}$		$n=10$		$n=30$		$n=50$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9433	0.5802	0.8972	0.6831	0.8811	0.7092
	0.50	0.8102	0.2239	0.7213	0.3460	0.6831	0.3943
	0.75	0.5524	0.0220	0.4465	0.0865	0.4150	0.1172
0.05	0.25	0.9254	0.6227	0.8819	0.7083	0.8682	0.7280
	0.50	0.7637	0.2811	0.6864	0.3852	0.6569	0.4222
	0.75	0.5010	0.0473	0.4095	0.1151	0.3876	0.1412
0.1	0.25	0.9143	0.6463	0.8745	0.7209	0.8613	0.7366
	0.50	0.7379	0.3110	0.6718	0.4035	0.6449	0.4361
	0.75	0.4689	0.0652	0.3900	0.1316	0.3717	0.1541

Table 3.20: Proportion of runs with future observation greater than the quartiles. Case 3.

3.8 Concluding remarks

This chapter has presented statistical methods for ALT using the Arrhenius-Weibull model under constant stress testing, using theory of imprecise probability [9, 10], where the imprecision results from the use of the likelihood ratio test [63]. The proposed method applies a classical test for comparison of the survival distributions at different stress levels. The observations at the increased stress levels were transformed to interval-valued observations at the normal stress level, by developing imprecision in the link function of the Arrhenius model via the likelihood ratio test between the pairwise stress levels. Using the Arrhenius model, we linked the data at different stress levels to the normal stress level, after which NPI can be used at normal stress level. We found that using an interval of values for the parameter in the link function between different stress levels enabled us to achieve a greater level of robustness than if we were to use a single point for the parameter. Using the proposed approach, the intervals $[\underline{\gamma}, \overline{\gamma}]$ for the parameter γ for the link function have adequate imprecision if the model fits well. However, the intervals $[\underline{\gamma}, \overline{\gamma}]$ for the parameter γ for the link function get wider if the model fits poorly. The latter can happen if the model assumptions are not fully correct, for example, using some misspecification cases. However, if we have huge imprecision, the remaining inferences are probably of no use at all. Therefore, it will be a strong recommendation to do more detailed modelling or sample more data. We also comment on this in Chapter 6. Regarding the choice of the values of the factors for assessing robustness of the methods, we only show that any suggested form of misspecification can be included in simulations to then study the level of robustness.

The application of our novel method in this chapter assumed that we have failure times observed at the normal stress level K_0. The assumption of having failure data at the normal stress level K_0 may not be realistic in real world applications. In this case, we can apply our method to a higher stress level than the normal stress level K_0. The combined data at that level can then be transformed all together to the normal stress level K_0. Investigating this is a topic for future research.

Usually, events of interest in reliability and survival analysis are failure times [19, 25], such data often includes right-censored observations. The $A_{(n)}$ assumption

cannot handle right-censored observations, and demands fully observed data. Coolen and Yan [21] presented a generalization of $A_{(n)}$, called rc-$A_{(n)}$, which is suitable for NPI with right-censored data, as discussed in Section 2.4. This method can be used at the second step in our approach if there are right-censored data, and the likelihood ratio test can just be applied in the first step. So generalizing this method to data including right-censored observations is straightforward, which will be illustrated in the next chapter.

We have considered the power-Weibull model to illustrate the use of our method without the assumption of equal shape parameter for the Weibull distribution at all stress levels too, and that other lifetime distributions and link functions can be applied, as long as the transformation of the failure times from higher stress levels to the normal stress level is monotonously increasing function. The approach presented in this chapter with the assumption of the Weibull failure time distributions at each stress level can be deleted and using nonparametric hypothesis tests instead, which will be illustrated in Chapter 4. In that chapter, we will also explain why we use pairwise tests instead of one test on all stress levels simultaneously.

Chapter 4

Imprecise inference based on the log-rank test

4.1 Introduction

In this chapter, we develop a new imprecise statistical method for ALT data related to the method introduced in Chapter 3. There, we assumed an explicit model at each stress level, namely the Weibull failure time distribution. In this chapter, we do not assume a failure time distribution at each stress level, only a specific parametric link function between the levels.

In this chapter, we assume the Arrhenius model for the analysis of ALT with failure data under a constant level of stress also used in Section 3.3. If the Arrhenius model provides a realistic link between the different stress levels, then the observations transformed from the increased stress levels to the normal stress level should be indistinguishable. For clarity, according to this model, an observation t^i at stress level i, subject to stress K_i, can be transformed to an observation at the normal stress level K_0, by the equation

$$t^{i\to 0} = t^i \exp\Big(\frac{\gamma}{K_0} - \frac{\gamma}{K_i}\Big), \tag{4.1.1}$$

where K_i is the accelerated temperature at level i (Kelvin), K_0 is the normal temperature at level 0 (Kelvin), and γ is the parameter of the Arrhenius model.

Note that in this chaper there is no distribution assumed at each stress level, hence we replace the likelihood ratio test used in Chapter 3, by the log-rank test. In this chapter, we propose a new log-rank test based method for predictive inference on a future unit functioning at the normal stress level. Testing the equality of the survival distribution of two or more independent groups, as we do in this chapter, is possible by using a nonparametric statistical test. There are several nonparametric test procedures that can be used to test the equality of the survival distributions; a popular one is the log-rank test [46, 65]. We use the log-rank test to find the intervals of values of the parameter γ of the Arrhenius link function for which we do not reject the null hypothesis that two groups of failure data, possibly including right-censored observations are, derived from the same underlying distribution. This can be interpreted such that, for such values of γ, the combined data at stress level K_0 are well mixed. This interval of values of the parameter γ of the Arrhenius link function is used to transform the data from the increased stress levels to the normal stress level. Then, the ultimate aim is inference at the normal stress level. We consider nonparametric predictive inference at the normal stress level combined with the Arrhenius model linking observations at different stress levels. Note we also assume that we have failure data at the normal stress level, as discussed in Chapter 3. Note also that this method follows the same procedure as in Chapter 3, except that we use a different classical hypothesis test because we do not assume the Weibull failure time distribution at each stress level.

This chapter is organized as follows. Section 4.2 introduces the main idea of imprecise predictive inference based on ALT and the log-rank test. The main novelty of the approach in this chapter is that by using a classical nonparametric test, we do not need to assume a parametric failure time distribution at each stress level. This should make the method more widely applicable than the method presented in Chapter 3. In Section 4.3 we explain why we do not use a single log-rank test on all stress levels. Section 4.4 illustrates our method in seven examples using simulated data and data from the literature. Section 4.5 presents results of simulation studies that investigate the performance of the proposed method using the Arrhenius link function. Section 4.6 presents results of simulation studies of robustness

that investigate the performance of the proposed method when we consider cases of misspecification. Section 4.7 presents some concluding remarks.

4.2 Predictive Inference Based on Log-Rank Test

In this section we present a new semi-parametric statistical method based on ALT data. The idea of the method is similar as that of the method introduced in Section 3.3, but we do not assume a parametric failure distribution at each stress level and we define the interval of values for the parameter γ of the Arrhenius link function based on the log-rank test, as reviewed in Section 3.3. The proposed new semi-parametric method consists of two steps. First, the pairwise log-rank test is used between stress levels K_i and K_0, to calculate the intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ of values γ of the Arrhenius link function for which we do not reject the null hypothesis that the data transformed from level i to level 0, and the original data from level 0, are derived from the same underlying distribution, where $i = 1, ..., m$. With these m pairs $(\underline{\gamma}_i, \overline{\gamma}_i)$, we define $\underline{\gamma}$ = max $\{\min\ \underline{\gamma}_i, 0\}$ and $\overline{\gamma} = \max\ \overline{\gamma}_i$.

As the second step of this method, each observation at an increased stress level is transformed to an interval at level 0. If the model fits relatively well, we expect most $\underline{\gamma}_i$ to be quite similar, and also most $\overline{\gamma}_i$ to be quite similar. The NPI lower survival function $\underline{S}$ is attained when all data observations at increased stress levels are transformed to the normal stress level using $\underline{\gamma}$, and the NPI upper survival function $\overline{S}$ results from the use of $\overline{\gamma}$. However, if the model fits poorly, the $\underline{\gamma}_i$ or the $\overline{\gamma}_i$, or both, are likely to differ considerably. Hence, in case of poor model fit, the resulting interval $[\underline{\gamma}, \overline{\gamma}]$ tends to be wider than in case of good model fit. The main novelty of the method in this chapter is that there is no parametric failure time distribution at each stress level and we also expect more imprecision in the resulting intervals $[\underline{\gamma}, \overline{\gamma}]$ of the parameter γ of the Arrhenius link function than in Chapter 3, if the method in Chapter 3 was applied with correct model.

4.3 Imprecision based on the pairwise test

Generally, testing the equality of a survival distribution of two or more independent groups often requires a nonparametric hypothesis statistical test. In our novel method discussed in Section 4.2, we use pairwise log-rank tests between stress level K_i and K_0. An alternative would be to use one log-rank test for the data at all stress levels combined. We now explain why this would not lead to a sensible method of imprecise statistical inference. If we have combined groups derived from the same underlying distribution, it is easy to reject the null hypothesis of equality. For example, if we assume we have three groups of data: if we do the pairwise comparison the confidence interval for any pairwise comparison for the data are higher and we may not be able to reject the null hypothesis. However, if we have a combined comparison for all the groups, it is unlikely that the test will show you that the probability distributions are the same, which makes it easier to reject the null hypothesis.

Let us briefly explain why we use the pairwise tests for each increased stress level and the normal stress level, instead of a single test for all stress levels combined. Suppose we wish to test the null hypothesis that data from all stress levels, all transformed to level 0 using parameter value γ_a, originate from the same underlying distribution. Let $[\underline{\gamma}_a, \overline{\gamma}_a]$ be the interval of the values γ_a for which this hypothesis is not rejected. If the model fits well, we would expect $\underline{\gamma}_a$ to be close to the $\underline{\gamma}$ from Section 4.2 and also $\overline{\gamma}_a$ to be close to $\overline{\gamma}$. If however, the model fits poorly, the interval $[\underline{\gamma}_a, \overline{\gamma}_a]$ may be very small or even empty. Therefore, this leads to less imprecision if the model fits poorly, which is opposite to the effect we wish to achieve by including imprecision. Thus, by performing pairwise testing on levels K_i and K_0, and taking the minimum and the maximum of the $\underline{\gamma}_i$ and $\overline{\gamma}_i$, respectively, we achieve more imprecision in the case of worse model fit, as discussed in Section 4.2. It would be of interest to study any relations between the lower and upper survival functions corresponding to the $[\underline{\gamma_i}, \overline{\gamma_i}]$ and $[\underline{\gamma_a}, \overline{\gamma_a}]$, this is left as a topic for future research.

For sensible imprecision, we explain why we only use the pairwise tests including the normal level data K_0 and not tests on two higher stress levels. We are obviously putting trust in the data if we have them available for the real level K_0. We choose

only to perform the pairwise test with level K_i and K_0. The level K_0 is always present because we assume that the data are available in the normal stress level K_0 and there are other data transformed to it from other levels K_i, so we wish to use the normal stress level K_0 as the true basis for the comparison.

The Examples 4.4.1-4.4.7 in Section 4.4 illustrate the method proposed by Section 4.2 as well as the problem if we wish to use the combined approach for all levels in Examples 4.4.1-4.4.3, as explained in Section 4.3.

4.4 Examples

In this section, we present seven examples in order to investigate the performance of our method and its robustness. In Example 4.4.1 we simulate data at all levels that correspond to the model for the link function which we assume for the analysis. In Example 4.4.2 we change these data such that the assumed link function no longer provides a good fit. In Example 4.4.3 we illustrate our new predictive inference method as presented in Section 4.2, using a data set from the literature. Examples 4.4.1, 4.4.2, and 4.4.3 also illustrate the problem that would occur if we would use the log-rank test on all stress levels combined, as discussed in Section 4.3. Example 4.4.4 briefly illustrates our proposed method in Chapters 3 and our method in this chapter. In Example 4.4.5 we illustrate our method for a data set including some right-censored data. Examples 4.4.6 and 4.4.7 apply our method to larger data sets, to illustrate the effect of the number of available data.

Example 4.4.1. The method proposed in Section 4.2 is illustrated data simulated at three temperatures. The normal temperature condition is $K_0 = 283$ and the increased temperatures in the stress levels are $K_1 = 313$ and $K_2 = 353$ Kelvin. Ten observations were simulated from a fully specified model using the Arrhenius link function in combination with a Weibull distribution at each temperature. This is the same model presented in Section 4.1 and used to simulate data in Examples 4.4.6 and 4.4.7. The Arrhenius parameter was set at $\gamma = 5200$, and the Weibull distribution at K_0 had shape parameter $\beta = 3$ and scale parameter $\alpha_0 = 7000$. This model has the same shape parameter at each temperature but the scale parameters are linked

Case	K_0= 283	K_1=313	K_2= 353	K_1=313 ($\times$1.4)	K_2= 353 ($\times$0.8)
1	2692.596	241.853	74.557	338.595	59.645
2	3208.336	759.562	94.983	1063.387	75.987
3	3324.788	769.321	138.003	1077.050	110.402
4	5218.419	832.807	180.090	1165.930	144.072
5	5417.057	867.770	180.670	1214.878	144.560
6	5759.910	1066.956	187.721	1493.739	150.176
7	6973.130	1185.382	200.828	1659.535	160.662
8	7690.554	1189.763	211.913	1665.668	169.531
9	8189.063	1401.084	233.529	1961.517	186.823
10	9847.477	1445.231	298.036	2023.323	238.429

Table 4.1: Failure times at three temperature levels (columns 1-3 and) corresponding failure times with misspecification (columns 4-5), Example 4.4.1.

by the Arrhenius relation, which led to scale parameter 1202.942 at K_1 and 183.091 at K_2. Failure times are simulated for ten units tested at each temperature (so for a total of 30 units in the study). The failure times are given in Table 4.1. In this example, to illustrate our method we assume that there are no right-censored data.

To illustrate our method using these data, we assume the Arrhenius link function for the data. Note that our method does not assume a parametric distribution at each stress level. The pairwise log-rank test is used between levels K_1 and K_0 and between levels K_2 and K_0 to derive the intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ of values γ for which we do not reject the null hypothesis. The resulting intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ are given in the first two rows of Table 4.2, for three test significance levels. Of course, for lower significance level the $[\underline{\gamma}_i, \overline{\gamma}_i]$ intervals become wider.

We obtain the NPI lower and upper survival functions by defining $\underline{\gamma} = \min \underline{\gamma}_i =$ 3901.267 and $\overline{\gamma} = \max \overline{\gamma}_i =$ 6563.545 at significance level 0.01, $\underline{\gamma} = \min \underline{\gamma}_i =$ 4254.053 and $\overline{\gamma} = \max \overline{\gamma}_i =$ 6251.168 at significance level 0.05, and $\underline{\gamma} = \min \underline{\gamma}_i =$ 4486.491 and the $\overline{\gamma} = \max \overline{\gamma}_i =$ 6017.435 at significance level 0.10. Note that the NPI lower survival function is based on the original data at level 0 together with

Significance level	0.01		0.05		0.10	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
K_1 K_0	3901.267	6563.545	4254.053	6251.168	4486.491	6017.435
K_2 K_0	4161.086	5555.130	4499.174	5442.667	4638.931	5353.034
Stress level	$\underline{\gamma}_a$	$\overline{\gamma}_a$	$\underline{\gamma}_a$	$\overline{\gamma}_a$	$\underline{\gamma}_a$	$\overline{\gamma}_a$
K_2 K_1 K_0	4156.263	5652.662	4464.828	5478.451	4499.174	5419.662

Table 4.2: $[\underline{\gamma}_i, \overline{\gamma}_i]$ for Example 4.4.1.

the transformed data from the stress levels K_1 and K_2 to K_0 using $\underline{\gamma}$. Similarly, the NPI upper survival function is based on the original data at level 0 together with the transformed data from the stress levels K_1 and K_2 to K_0 using $\overline{\gamma}$. In Figure 4.1, the lower survival function $\underline{S}$ is labeled as $\underline{S}$ $(\underline{\gamma}_i)$ and the upper survival function $\overline{S}$ is labeled as $\overline{S}$ $(\overline{\gamma}_i)$. This figure shows that smaller significance levels lead to more imprecision for the NPI lower and upper survival functions, which directly results from the fact that the intervals $[\underline{\gamma}, \overline{\gamma}]$ are nested, becoming larger if the significance level is decreased. The NPI lower and upper survival functions can be derived using the these $[\underline{\gamma}, \overline{\gamma}]$ intervals, as showed in Figure 4.1(a).

As discussed in Section 4.3, the use of a single log-rank test involving the data from all stress levels simultaneously will not lead to the desired effect that a worse model fit should lead to more imprecision. This is illustrated by the values in the final row in Table 4.2, which are derived by such a simultaneous log-rank test for this example. From this interval we can again obtain the lower and upper survival functions using NPI; these are presented in Figure 4.1(b). In this example, the data were simulated precisely with the link function as assumed in our method, so there is not much difference between the lower and upper survival functions for the corresponding significance levels in Figures 4.1(a) and 4.1(b). Example 4.4.2 illustrates what happens if the model does not fit well.

Example 4.4.2. To illustrate our method in a case where the assumed Arrhenius link function does not fit the data well, and also to show what would have happened if we had used the joint log-rank test in our method instead of the pairwise tests, we use the same simulated data as in Example 4.4.1, but we change the data from level

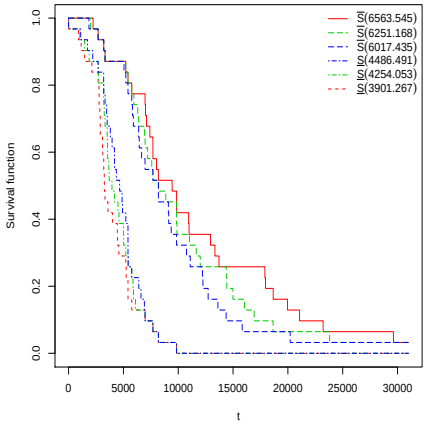
Significance level	0.01		0.05		0.10	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
$K_1 \times 1.4$, K_0	2907.787	5570.065	3260.574	5257.689	3493.011	5023.956
K_2 K_0	4161.086	5597.978	4499.174	5442.667	4638.931	5353.034
Stress level	$\underline{\gamma}_a$	$\overline{\gamma}_a$	$\underline{\gamma}_a$	$\overline{\gamma}_a$	$\underline{\gamma}_a$	$\overline{\gamma}_a$
K_2 K_1 K_0	4455.573	5568.468	4638.930	5368.780	4742.958	5257.689

Table 4.3: $[\underline{\gamma}_i, \overline{\gamma}_i]$ for Example 4.4.2, Scenario 1.

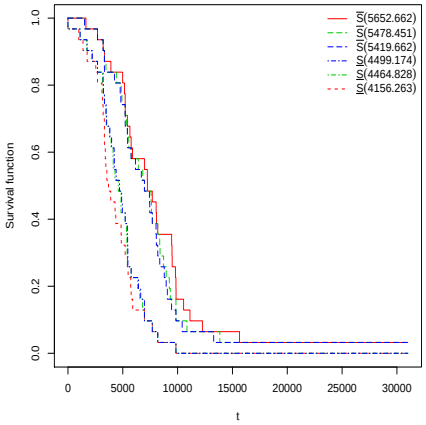
K_1 in Scenario 1 and from levels K_1 and K_2 in Scenario 2. In Scenario 1 (indicated as Ex 4.4.2-1 in Figure 4.1), we multiply the data at level K_1 by 1.4. In Scenario 2 (Ex 4.4.2-2 in Figure 4.1), we do the same while we also multiply the data at level K_2 by 0.8. The resulting data values are given in the last two columns in Table 4.1. This is similar to our Example 3.4.1 in Chapter 3 but in this example, we define the interval of values of the parameter γ of the Arrhenius link function based on the nonparametric log-rank test.

For these two scenarios, we have repeated the analysis as described in Example 4.4.1. The resulting intervals of $[\underline{\gamma}_i, \overline{\gamma}_i]$ are given in the first two rows of Tables 4.3 and 4.4. The NPI lower and upper survival functions in Figures 4.1(c) and 4.1(e), using our method as discussed in Section 4.2, have more imprecision than the NPI lower and upper survival functions in Figures 4.1(a), as shown in Example 4.4.1 of this chapter. Note that the lower survival function is identical in both scenarios as the same $\underline{\gamma}$ is used; this is because the increased values at K_1 have resulted in smaller values for $\underline{\gamma}_1$ and $\overline{\gamma}_1$ and the $\underline{\gamma}$ in our method is equal to the $\underline{\gamma}_1$ in these cases. In Scenario 2, the observations at level K_2 have decreased, leading to larger $\underline{\gamma}_2$ and $\overline{\gamma}_2$ values, and this leads to the upper survival functions increasing in comparison to Scenario 1.

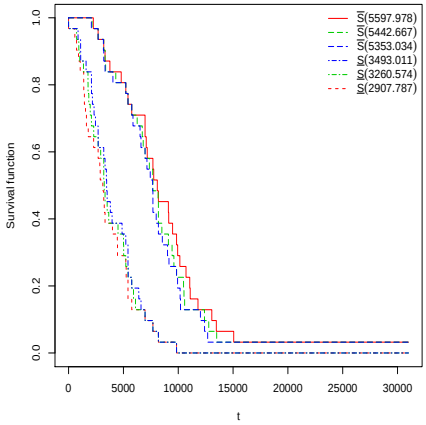
(a) Using $\underline{\gamma}$, $\overline{\gamma}$ (Ex 4.4.1)



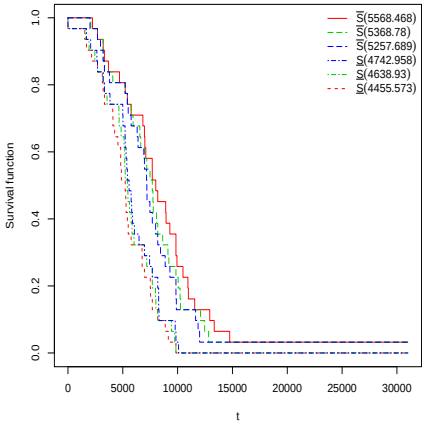
(b) Using $\underline{\gamma}_a$, $\overline{\gamma}_a$ (Ex 4.4.1)



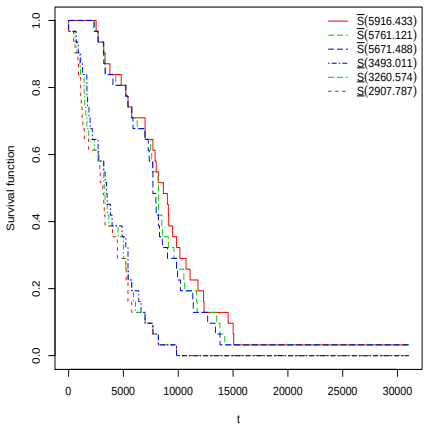
(c) Using $\underline{\gamma}$, $\overline{\gamma}$ (Ex 4.4.2-1)



(d) Using $\underline{\gamma}_a$, $\overline{\gamma}_a$ (Ex 4.4.2-1)



(e) Using $\underline{\gamma}$, $\overline{\gamma}$ (Ex 4.4.2-2)



(f) Using $\underline{\gamma}_a$, $\overline{\gamma}_a$ (Ex 4.4.2-2)

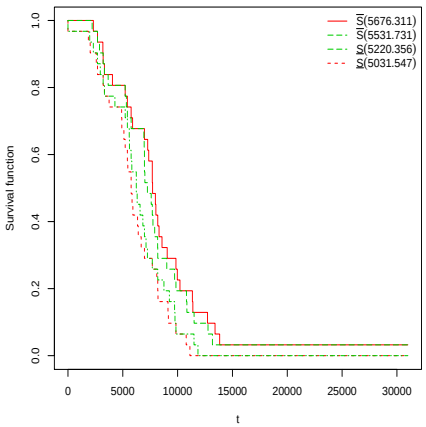


Figure 4.1: The NPI lower and upper survival functions. Examples 4.4.1 and 4.4.2.

Significance level	0.01		0.05		0.10	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
$K_1 \times 1.4$, K_0	2907.787	5570.065	3260.574	5257.689	3493.011	5023.956
$K_2 \times 0.8$, K_0	4479.541	5916.433	4817.629	5761.121	4957.386	5671.488
Stress level	$\underline{\gamma}_a$	$\overline{\gamma}_a$	$\underline{\gamma}_a$	$\overline{\gamma}_a$	$\underline{\gamma}_a$	$\overline{\gamma}_a$
K_2 K_1 K_0	5031.547	5676.311	5220.356	5531.731	$\emptyset$	

Table 4.4: $[\underline{\gamma}_i, \overline{\gamma}_i]$ for Example 4.4.2, Scenario 2.

If we had used the joint long-rank test instead of the pairwise tests, as discussed in Section 4.3, then imprecision would have decreased in these two scenarios, as seen from Figures 4.1(d) and 4.1(f). Note that in Figure 4.1(f) there are no lower and upper survival functions corresponding to the use of the joint log-rank test for significance level 0.10, as this leads to an empty interval of γ values, so the interval from $\underline{\gamma}_a$ to the $\overline{\gamma}_a$ is an empty set, so clearly our method would not work if we had used this joint test instead of the pairwise tests. As mentioned in Section 4.3, if the model does not fit well, then we may reject the null hypothesis for the joint test of the three levels together, see Tables 4.3 and 4.4. Therefore, we have a smaller range of values for which we do not reject the null hypothesis. However, if the model fits poorly, as shown in Figures 4.1(d) and 4.1(f), we actually want a larger range of values of γ, and therefore increased imprecision in the lower and upper survival functions. Therefore, it is obvious that this is achieved by taking the minimum of the $\underline{\gamma}_i$ and the maximum of the $\overline{\gamma}_i$ of the pairwise tests, which will have increased imprecision, hence this is our proposed method as discussed in Section 4.2. This is illustrated by Figures 4.1(a), 4.1(c) and 4.1(e). Examples 4.4.3 and 4.4.4 will illustrate our predictive inference method using data sets from the literature.

Example 4.4.3. The methods proposed in Sections 4.2 and 4.3 are illustrated using the same data set as Example 3.4.3 in Chapter 3, resulting from a temperature-accelerated life test. The time-to-failure data were collected at the normal temperature $K_0 = 393$ Kelvin and at two increased temperature stress levels, $K_1 = 408$ and $K_2 = 423$. Ten units were tested at each temperature, so a total of 30 units were used in the study. All of the units failed during the experiment. The failure times,

Significance level	0.01		0.05		0.1	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
K_1 K_0	-1874.191	5169.809	-1108.280	4403.899	-624.387	3920.005
K_2 K_0	38.751	3690.236	435.786	3293.202	686.627	3042.360
Stress level	$\underline{\gamma}_a$	$\overline{\gamma}_a$	$\underline{\gamma}_a$	$\overline{\gamma}_a$	$\underline{\gamma}_a$	$\overline{\gamma}_a$
K_2 K_1 K_0	-384.624	3980.852	153.698	3575.289	430.064	3298.923

Table 4.5: $[\underline{\gamma}_i, \overline{\gamma}_i]$ for Example 4.4.3.

in hours, are given in Table 3.5.

To illustrate our method using these data, we assume the Arrhenius link function for the data. Then we use the pairwise log-rank test separately between K_i and K_0 to find the intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ of values of γ for which we do not reject the null hypothesis with regard to the well-mixed data transformation. The resulting intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ are given in the first two rows of Table 4.5, for three test significance levels.

According to the $[\underline{\gamma}_i, \overline{\gamma}_i]$ for log-rank test intervals in Table 4.5, we can obtain the NPI lower and upper survival functions as described in Section 2.4. The lower and upper survival functions for levels K_1 and K_0 at level of significance 0.01, 0.05 and 0.10 values are shown in Figure 4.2(a). The latter figure shows that the NPI lower survival functions have not transformed because of the $\underline{\gamma} = \max\{\min \underline{\gamma}_i, 0\} = 0$ at significance levels $0.01, 0.05, 0.10$ are always 0 in this case, and for $\overline{\gamma_1}$ we have, 5169.809, 44.3.899 and 3920.005, respectively, for the NPI upper survival functions. The resulting intervals $[\underline{\gamma}_2, \overline{\gamma}_2]$ of values γ_2 are given in the second row in Table 4.5. From these intervals, the NPI lower and upper survival functions for all levels K_2 and K_0 with levels of significance 0.01, 0.05 and 0.10 values are shown in Figure 4.2(b).

Also, we illustrate the effect of using a single test of the null hypothesis that all the transformed data from all stress levels come from the same underlying distribution as the observations at the normal stress level. The resulting intervals $[\underline{\gamma}_a, \overline{\gamma}_a]$ of values γ_a are given in the third row in Table 4.5. From these intervals, the NPI lower and upper survival functions for all levels K_2, K_1 and K_0 with levels of significance 0.01, 0.05 and 0.10 values are shown in Figure 4.2(c).

$\overline{\gamma}$	Transformed data				
$\overline{\gamma} = 5169.809$	3850	4340	4760	5320	5352.439
	5740	6033.658	6160	6580	6617.560
$\alpha = 0.01$	6990.559	7140	7396.097	7880.266	7979.999
	7980	8563.902	8642.873	8960	9147.804
	9659.681	9926.341	10422.288	11094.145	11184.894
	11947.501	12456.584	12964.309	14489.522	16268.937
$\overline{\gamma} = 4403.899$	3850	4340	4760	4982.353	5320
	5616.471	5740	6088.136	6160	6160.000
$\alpha = 0.05$	6580	6862.990	6884.706	7140	7428.235
	7527.151	7971.765	7980	8412.698	8515.294
	8960	9076.858	9240.000	9741.018	10327.059
	10405.179	11290.726	11595.294	12619.047	14168.754
$\overline{\gamma} = 3920.005$	3850	4340	4760	4761.842	5320
	5367.894	5579.042	5740	5887.368	6160
$\alpha = 0.1$	6289.102	6579.999	6580	6897.724	7099.473
	7140	7618.947	7709.221	7980	8138.420
	8317.844	8831.052	8926.467	8960	9535.090
	9869.999	10346.587	11082.104	11563.832	12983.952

Table 4.6: Transformed data for the upper survival functions.

The resulting intervals $[\underline{\gamma}, \overline{\gamma}]$ of values γ are given in the first row in Table 4.5. Figure 4.2(d) shows only our new predictive inference method, which only taking the minimum of the $\underline{\gamma}_i$ and the maximum of the $\overline{\gamma}_i$ of the pairwise tests, to have more imprecision. In this figure, the lower survival function $\underline{S}$ is labeled as $\underline{S}$ $(\underline{\gamma}_i)$ and the upper survival function $\overline{S}$ is labeled as $\overline{S}$ $(\overline{\gamma}_i)$. This figure shows that lower significance levels leads to more imprecision and nesting for the NPI lower and upper survival functions.

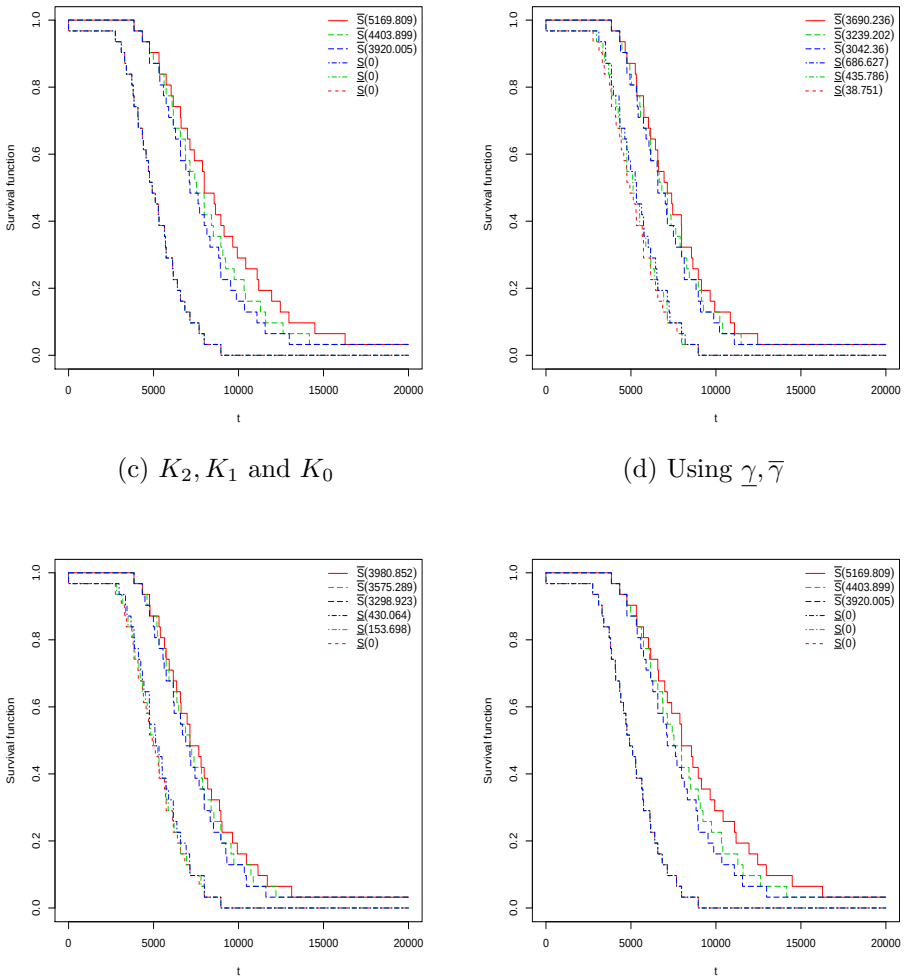
Table 4.6 shows the comparison for all transformed observations with levels K_1 and K_0 for the transformed data of the upper survival function. We know that the transformed data of the lower survival function end at 8960 at significance levels $0.01, 0.05, 0.10$ because of the $\underline{\gamma} = 0$, therefore, we only have the data as given in Table 3.5 in Example 3.4.3. From Table 4.6, the observations transformed with the 0.01 significance level are wider than the transformed observations with the 0.05 and

0.10 significance levels. Note that, the intervals $[\underline{\gamma_2}, \overline{\gamma_2}]$ of the pairwise stress levels K_2 and K_0 are within the intervals $[\underline{\gamma_1}, \overline{\gamma_1}]$ of the pairwise stress levels K_1 and K_0.

In this method the intervals occur when two values are identical (up to numerical rounding), these values are underlined in Table 4.6. For example, the point 6160 appears consecutively in Table 4.6 with significance levels 0.05. When this occurs or when the difference between the two points is extremely small, we find the interval $[\underline{\gamma}_i, \overline{\gamma}_i]$ of values γ parameter of the link function. One of these points is an observation from the real level K_0 and the other data point is transformed to the level K_0 from the level K_1. Consequently, when we have a γ which is at the point where the p-value changes, you must have one data point from the higher level K_i which is about equal to one of the data points from the normal level K_0.

As mentioned we used the same data set as Example 3.4.3 in Chapter 3, where we use the assumption of the Weibull distribution at each stress level, so the resulting intervals $[\underline{\gamma}, \overline{\gamma}]$ in Table 3.6 shows less imprecision compared to the resulting intervals $[\underline{\gamma}, \overline{\gamma}]$ in this example in Table 4.5 where the log-rank test is used.

(a) K_1 and K_0 (b) K_2 and K_0



(c) K_2, K_1 and K_0 (d) Using $\underline{\gamma}, \overline{\gamma}$

Figure 4.2: The NPI lower and upper survival functions. Example 4.4.3.

Example 4.4.4. In this example, we compare our method presented in this chapter with the method presented in Chapter 3 when they assumed to have a failure time distribution at each stress level and use the likelihood ratio test to derive the interval for the parameter of the link function. Suppose the data correspond to the ALT under the constant stress level published in [53, 77], which presents the voltage-accelerated lifespan test. Three voltage test levels were established at $K_0 = 80, K_1 = 100$, and $K_2 = 120$. The normal voltage is $K_0 = 80$ and the increased voltage stress levels are $K_1 = 100$ and $K_2 = 120$ Voltages. At each voltage level, $n = 8$ failure times were under analysis, for a total of 24 failure times used in the study. All of the observations failed during the experiment. The failure times, in hours, are given

Case	$K_0 = 80$	$K_1 = 100$	$K_2 = 120$
1	1770	1090	630
2	2448	1907	848
3	3230	2147	1121
4	3445	2645	1307
5	3538	2903	1321
6	5809	3357	1357
7	6590	4135	1984
8	6744	4381	2331

Table 4.7: Failure time of surface-mounted electrolytic capacitor. Example 4.4.4.

Significance level	0.01		0.05		0.1	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
K_1 K_0	-0.567	4.329	0.153	3.642	0.480	3.330
K_2 K_0	1.438	4.192	1.836	3.803	2.018	3.625

Table 4.8: $[\underline{\gamma_i}, \overline{\gamma_i}]$ for the likelihood ratio test. Example 4.4.4

in Table 4.7.

Applying our new method as discussed in Chapter 3 when we assumed a failure time distribution at each stress level using the data in Table 4.7, we have assumed the Weibull failure time distributions at each stress level, with the power-law link function between different stress levels. The power-Weibull model for different stress levels is assumed to have different scale parameters, α_i and same shape parameters $\beta_i = \beta$, for level $i = 0, 1, ..., m$.

Then, we derive the intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ of the parameter γ values of the power law link function based on the pairwise likelihood ratio test for which the null hypothesis that the two groups of failure data come from the same underlying distribution, is not rejected [63]. The resulting intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ are given in Table 4.8, for three test significance levels.

Furthermore, applying our new method as discussed in Section 4.2 using these data, first, we assume the pairwise log-rank test used between the stress level K_i and the normal stress level K_0 to obtain the intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ of the values γ of the

Significance level	0.01		0.05		0.1	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
K_1 K_0	-1.077	4.460	-0.699	3.777	0	3.126
K_2 K_0	1.030	4.057	1.427	3.990	2.139	3.954

Table 4.9: $[\underline{\gamma_i}, \overline{\gamma_i}]$ for log-rank test. Example 4.4.4

power-law link function. The resulting $[\underline{\gamma}_i, \overline{\gamma}_i]$ intervals values of the parameter γ of the link function we assumed are given in Table 4.9. We have found that the resulting intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ in Table 4.9 have more imprecision than in the resulting intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ in Table 4.8 where we assumed the Weibull assumption at each stress level.

In the second step of our method presented in Chapters 3 and 4, we transform the data using the $[\underline{\gamma}, \overline{\gamma}]$ values, respectively. The NPI lower and upper survival functions can be obtained according to the $[\underline{\gamma_i}, \overline{\gamma_i}]$ intervals, as explained in Chapters 3 and 4, see Figures 4.3 and 4.4, respectively.

Note that there is a lot of overlap between the data in Table 4.7 and as mentioned in Chapters 3 and in this chapter as well, if the model assumed is not too far from reality, we would expect the widest interval for the parameter γ to come from the likelihood ratio test and the log-rank test, respectively, applied to levels 1 and 0. However, if the model is worse then we would expect more often that $\underline{\gamma} = \underline{\gamma}_i$ for an $i \neq 1$, or $\overline{\gamma} = \overline{\gamma}_i$ for an $i \neq 1$.

Example 4.4.5. In the previous examples, we studied the effect of observed failure times on our statistical inference method for a future observation. However, there may be right-censored observations among the data. In this example, we use the same data set as Example 3.4.3 in Chapter 3 again, as done in Example 4.4.3, but we assume a few cases of right censoring and study the effect of the censoring in the original data in Table 3.5 by showing the NPI lower and upper survival functions for the failure time of the future unit based on the combined information. Then we compare the original NPI lower and upper survival functions in the Example 4.4.3 with the NPI lower and upper survival functions in this example based on where the right-censored observations are and the effect of these right-censored observations

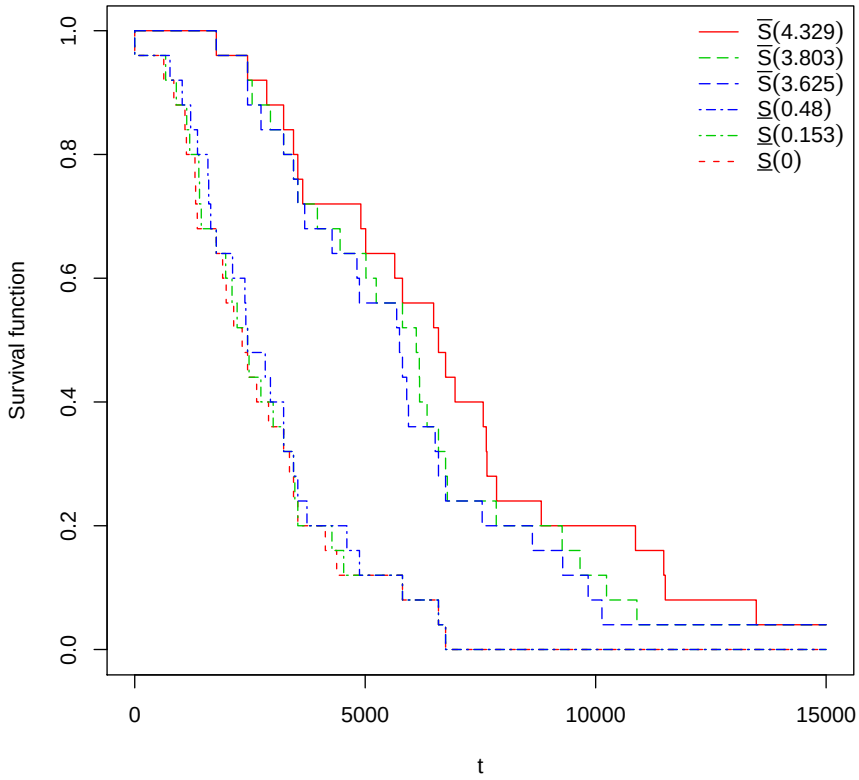


Figure 4.3: NPI lower and upper survival functions. Likelihood ratio test, Example 4.4.4

on the log-rank test. We consider three cases for the right-censoring observations which are given in Table 4.10.

In Case 1, we assume that the 4 observations with + sign in Table 4.10 are instead right-censored observations at 4000. In Case 2, we assume that the 5 observations with $*$ sign in Table 4.10 are instead right-censored observations at 6600. In Case 3, all the right-censored observations of Cases 1 and 2 occur.

Again, we obtain the lower and upper interval of the γ values according to the null hypotheses with levels of significance $0.01, 0.05, 0.1$ using the log-rank test for all the three cases. All the resulting intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ of values γ of Cases 1, 2, and 3 are given in Table 4.11. The NPI lower and upper survival functions using these intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ for the γ parameter of the Arrhenius link function, as presented Figure 4.5. The intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ for the first, second, and third cases in this figure are

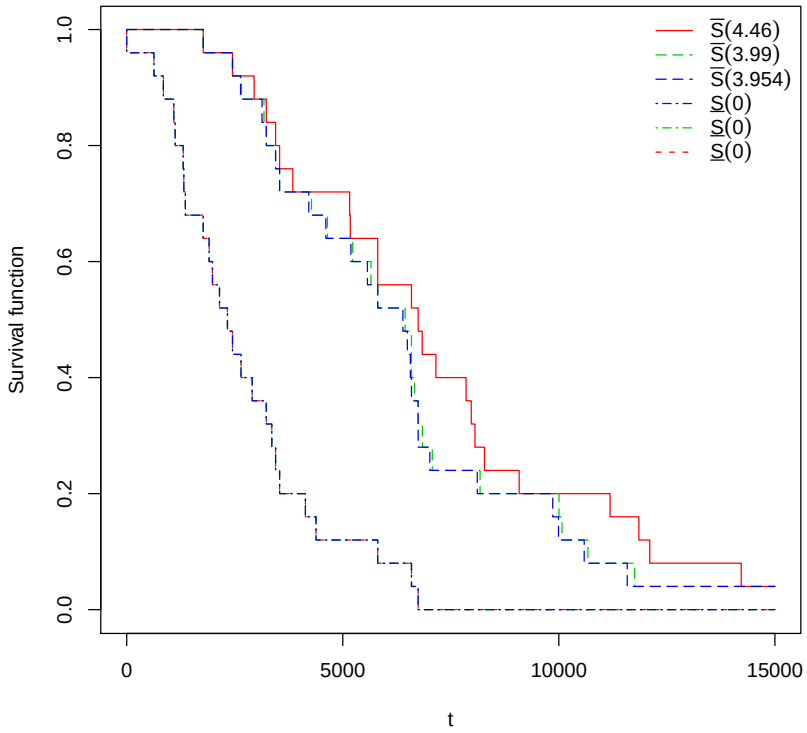


Figure 4.4: NPI lower and upper survival functions. Log-rank test, Example 4.4.4

$[\underline{\gamma}, \overline{\gamma}] = [0, 5982.068]$, $[\underline{\gamma}, \overline{\gamma}] = [0, 8258.461]$, and $[\underline{\gamma}, \overline{\gamma}] = [0, 10677.282]$ at significance level 0.01 corresponding to the NPI lower and upper survival functions, respectively. Note that the NPI lower and upper survival functions, in Figure 4.2, in Example 4.4.3, decrease at observed failure times, because all the units transformed to the normal level as well as the observations at level 0, were failure times. However, in Example 4.4.5, Cases 1, 2 and 3, the NPI lower survival functions in Figure 4.5 decrease at observed failure times and also decrease by small amounts at right-censored observations, while the NPI upper survival functions in Figure 4.5 only decrease at observed failure times, as introduced in Section 2.4. In these Cases 1, 2 and 3, at level K_0, we have right-censored observations and these observations never get transformed which means that the NPI lower survival functions decrease slightly due to the censoring, whilst the NPI upper survival functions decrease only at the observed failure times. Note that the $[\underline{\gamma}_i, \overline{\gamma}_i]$ intervals become wider when we have more censored observations, see Table 4.11 and Figure 4.5.

Case	$K_0 = 393$	$K_1 = 408$	$K_2 = 423$
1	3850	3300	2750
2	4340+	3720	3100
3	4760+	4080+	3400
4	5320+	4560	3800
5	5740	4920	4100
6	6160	5280	4400
7	6580	5640	4700
8	7140*	6120	5100
9	7980*	6840*	5700
10	8960*	7680*	6400

Table 4.10: Failure times at three temperature levels, Example 4.4.5.

From Figure 4.5, we can find out that the upper survival function for Case 2 moved over the upper survival functions for Case 1 under certain circumstance. This can happen if the order of an observed event and a right-censored observation is different under two different transformations. It is important if a right-censored observation is before or after a fully observed event time, because the probability mass that is divided at a right-censoring time among the intervals to the right of it of course depends on the number of observations to the right of the right-censoring time. If there are fewer observations to the right of the right-censoring time, the intervals between them all get a bit more probability mass. So, that means that in the right tail there will be some probability mass if earlier a right-censoring had occurred to the right of an observation. More examples are presented in [**?**] which apply our method to larger data sets, to illustrate the effect of the number of available data. Increasing the values of n at each stress level, lead to less imprecision in the resulting intervals $[\underline{\gamma}, \overline{\gamma}]$, so fewer values of γ in the null hypothesis are not rejected, compare to, for example, when we have $n = 10$ observations at each stress level. Note further that the NPI part also has less imprecision for the larger values of n.

Significance level	0.01		0.05		0.1	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
Case 1: K_1 K_0	−1119.318	5982.068	−353.408	4948.068	409.614	4636.459
K_2 K_0	606.301	4332.095	1200.640	3805.070	1222.635	3575.298
Case 2: K_1 K_0	−3673.884	8185.461	−2357.513	6408.005	−1652.449	5653.132
K_2 K_0	38.751	5239.500	435.786	4332.095	795.557	3940.782
Case 3: K_1 K_0	−2357.513	10677.282	−1119.319	7377.025	−414.253	7220.264
K_2 K_0	686.627	6545.213	1352.626	5287.021	1864.493	4834.417

Table 4.11: $[\underline{\gamma}_i, \overline{\gamma}_i]$ for Example 4.4.5.

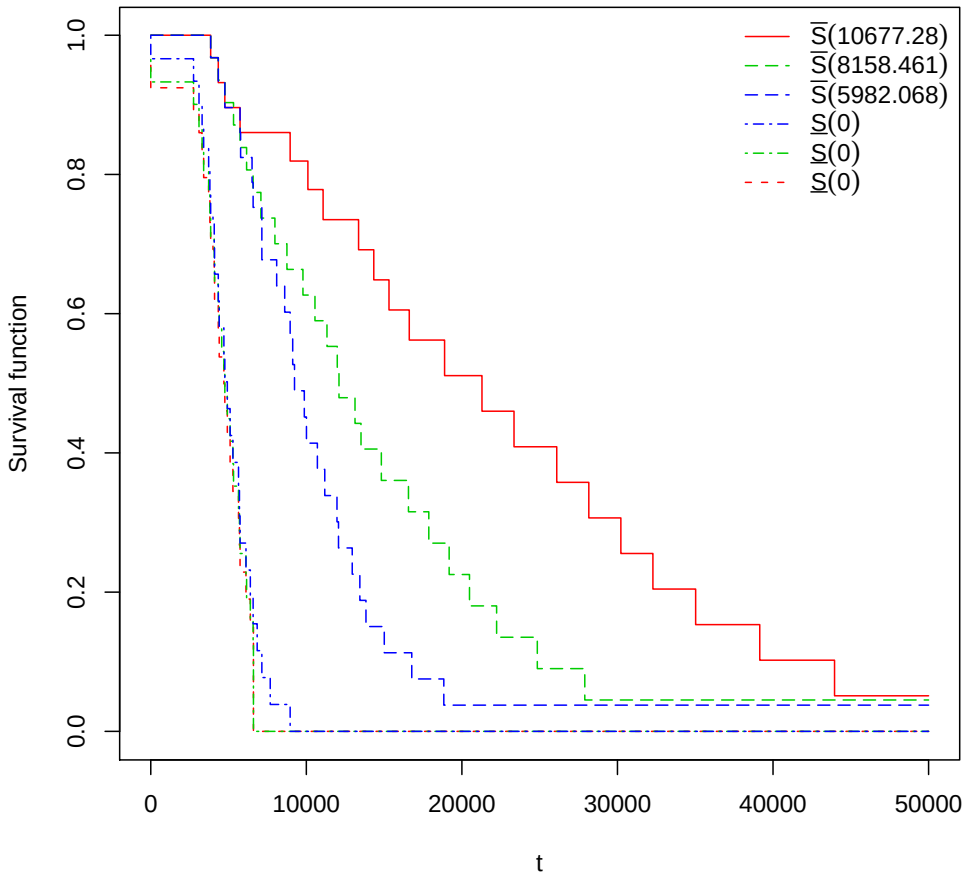


Figure 4.5: The NPI lower and upper survival functions, Case 1 (blue line), Case 2 (green line), and Case 3 (red line). Example 4.4.5.

Significance level	0.01		0.05		0.10	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
K_1 K_0	3901.115	6023.391	4174.068	5712.041	4224.630	5695.763
K_2 K_0	4763.596	5620.123	4870.732	5542.659	4946.343	5479.204

Table 4.12: $[\underline{\gamma_i}, \overline{\gamma_i}]$ for Example 4.4.6.

Example 4.4.6. To apply our method in Section 4.2 and show the effect of a larger data set, the failure data are again simulated at three temperatures as used in Example 4.4.1, but we simulate $n_0 = 10$, $n_1 = 20$, and $n_2 = 50$ observations from a Weibull distribution at levels K_0, K_1, and K_2, respectively, so a total of 80 observations. For the simulation we used again the Weibull distribution at level K_0 with shape parameter $\beta = 3$ and scale parameter $\alpha_0 = 7000$, and the Arrhenius parameter $\gamma = 5200$. These values lead to $\alpha_1 = 1202.942$ at level K_1 and $\alpha_2 = 183.0914$ at level K_2. To have some idea about the range of the data for each level for this example note that the smallest observation at stress level K_0 is 2692.596 and the largest observation is 9847.477, the smallest observation at stress level K_1 is 241.853 and the largest observation is 1958.147, and the smallest observation at stress level K_2 is 62.709 and the largest observation is 284.655.

The resulting intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ for three significance levels using the pairwise log-rank tests are given in Table 4.12. In this example, we have $\underline{\gamma} = 3901.115$ and $\overline{\gamma} = 6023.391$, $\underline{\gamma} = 4174.068$ and $\overline{\gamma} = 5712.041$, and $\underline{\gamma} = 4224.630$ and $\overline{\gamma} = 5695.763$ at significance levels 0.01, 0.05 and 0.10, respectively. We used the above $\underline{\gamma}$ and $\overline{\gamma}$ values to transform the data to the normal stress level 0, the resulting NPI lower and upper survival functions are presented in Figure 4.6. This figure shows that smaller significance level leads to more imprecision for the NPI lower and upper survival functions. In addition, increasing the values of n_0, n_1, and n_2, lead to less imprecision in Figure 4.6 compare to Figure 4.1(a), when we have $n = 10$ observations at each stress level.

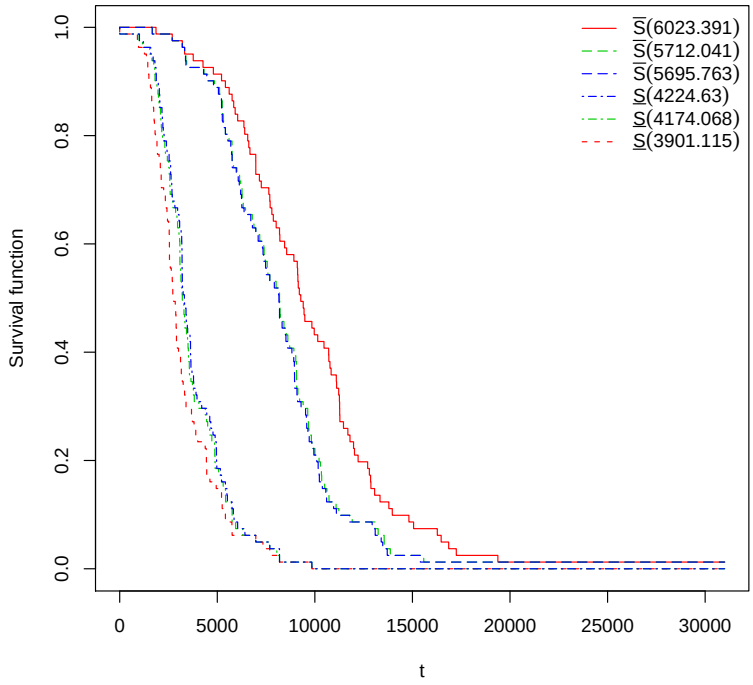


Figure 4.6: NPI lower and upper survival functions. Example 4.4.6.

Example 4.4.7. In the previous example, we studied the effect of $n_0 = 10$, $n_1 = 20$, and $n_2 = 50$ observations at levels K_0, K_1, and K_2, respectively. In this example we simulate $n = 50$ observations from a Weibull distribution at all levels K_0, K_1, and K_2, so total of 150 failure times are simulated. The failure data are again simulated at three temperatures as used in Example 4.4.1, with the same parameters of the Arrhenius-Weibull used. The range of the data for each level for this example note that the smallest observation at stress level K_0 is 1407.360 and the the largest observation is 11394.589, the smallest observation at stress level K_1 is 412.0113 and the the largest observation is 1701.799, and the smallest observation at stress level K_2 is 35.587 and the the largest observation is 298.579.

The resulting intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ for three significance levels using the pairwise log-rank tests are given in Table 4.13. Based on the original data at level 0 together with the data transformed from the stress levels K_1 to K_0 and K_2 to K_0 using $\underline{\gamma}$, the NPI approach provides the NPI lower survival functions. Similarly, but using $\overline{\gamma}$ for the transformation, we derived the NPI upper survival functions. For the significance

Significance level	0.01		0.05		0.10	
Stress level	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$	$\underline{\gamma}_i$	$\overline{\gamma}_i$
K_1 K_0	4691.036	5679.580	4808.502	5584.119	4852.349	5543.948
K_2 K_0	4893.003	5404.910	4971.533	5332.027	4994.124	5310.658

Table 4.13: $[\underline{\gamma_i}, \overline{\gamma_i}]$ for Example 4.4.7.

levels $0.01, 0.05, 0.1$, the values of $\underline{\gamma}$ are 4691.036, 4808.502 and 4852.349 in this case, and for $\overline{\gamma}$ we have, 5679.580, 5584.119 and 5543.948, respectively. The corresponding NPI lower and upper survival functions are presented in Figure 4.7. Comparing the imprecision in the survival functions in this example as shown in Figure 4.7, with the imprecision in the survival functions in the Example 4.4.6, it shows that when we generated $n = 50$ observation at each stress level, the imprecision is smaller than with fewer observations in Example 4.4.6.

4.5 Simulation studies

In this section, we present the results of a simulation study to investigate the performance of the imprecise predictive inference method for ALT data proposed in this chapter. For the simulation of data, we applied the same method as described in Section 3.5. However, for the analysis we do not assume a failure time distribution at each stress level, only a specific parametric link function between the levels, and we derive the interval of values of the parameter of this link function using log-rank tests. In Case A, we assumed the temperature stress levels to be $K_0 = 283, K_1 = 313$, and $K_2 = 353$ Kelvin. We then generated random samples from the Arrhenius-Weibull model with the scale parameter $\alpha = 7000$, shape parameter $\beta = 3$, and the Arrhenius link function parameter $\gamma = 5200$. Then we used the pairwise log-rank test with the simulated data sets separately between K_1 and K_0 and between K_2 and K_0 to derive the intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ of γ values, as described in Section 4.2, where we have chosen different values of $\alpha = 0.01, 0.05, 0.1$ for the level of significance.

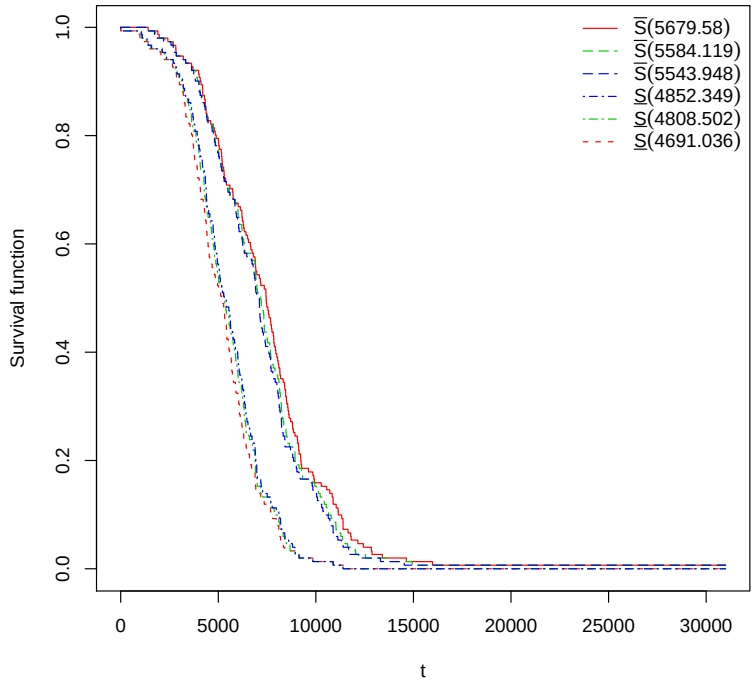


Figure 4.7: NPI lower and upper survival functions. Example 4.4.7.

As was done in Section 3.5, a future observation was simulated at stress level K_0 and this was used to investigate the predictive performance of our method. For each simulation, we ran the simulation 10,000 times with $n = 10, 50, 100$ observations at each stress level. We examined the performance at the quartiles of NPI lower and upper survival functions, checking whether or not the future observation at the normal stress level has exceeded these quartiles in the right proportion of all simulation runs. For our method to perform well, the future observation for each run at the normal stress level should exceed the first, second, and third quartiles of the NPI lower survival functions in a fraction just over $0.75, 0.50$ and 0.25 of the runs and in a fraction just under these values for the NPI upper survival functions.

Tables 4.14 and Figures 4.8-4.10 present the results of the performance of our method when the Arrhenius model is the assumed model for the generated data and for the analysis, so the model assumed for the analysis is the same as used for the data simulation. All of these figures show that the proposed method performs well overall. Note that the first, second, and third quartiles in these figures are denoted by $qL0.25$

and $qU0.25$, $qL0.50$ and $qU0.50$, and $qL0.75$ and $qU0.75$ corresponding to the NPI lower and upper survival functions, respectively. We note that for corresponding proportions with larger values of n, the differences between the lower and upper survival functions tend to decrease. This means that when based on more data, the NPI lower and upper survival functions allow less imprecise inference. However, when $n = 100$ there remains quite a bit of imprecision at 0.1 significance level. This is due to the effect caused by performing the nonparametric test statistics, and therefore, we still see quite a bit of imprecision here even with a larger data set. The intervals $[\underline{\gamma_2}, \overline{\gamma_2}]$ of the pairwise stress levels K_2 to K_0 always seems to be within the the intervals $[\underline{\gamma_1}, \overline{\gamma_1}]$ of the pairwise stress levels K_1 to K_0 once. So the K_1 to K_0 simply has more imprecision, as described in Example 4.4.3. Taking the $[\underline{\gamma}, \overline{\gamma}]$ of the pairwise stress levels K_1 to K_0 or K_2 to K_0 with $0.01, 0.05, 0.10$ significance level will have more imprecision, see Figures 4.8-4.10 (g, h, i).

In Case B, we used a similar simulation scenario, with $n = 10, 50, 100$ observations at each stress level, however, we generate the data from the power-Weibull model and for the analysis, as introduced in Section 3.6 using Equation 3.6.1, so the model assumed for the analysis is the same as used for the data simulation with $\alpha_0 = 1500$, shape parameter $\beta = 3$, and the power-law link function parameter $\gamma = 10$. We assume three different stress levels $K_0 = 50, K_1 = 80$, and $K_2 = 120$ kilovolts. The power-Weibull model for different stress levels are assumed to have different scale parameters, α_i and same shape parameters $\beta_i = \beta$, for level $i = 0, 1, ..., m$.

For this simulation, we have repeated the same analysis as just described in Case A simulation. The results are presented in Table 4.15. Again, the results for this simulation study support the same conclusion as those just described, with attention on the prediction of one future observation at the normal stress level K_0 and how well it mixes with actual data at the normal stress level.

The use of pairwise tests is discussed in Section 3.3, where we assumed a Weibull distribution at each stress level and use the parametric likelihood ratio test instead of the log-rank test. The argument for the use of the pairwise test is the same: if the model fits poorly, a single test on all stress levels would result in less imprecision while our proposed method, combining pairwise tests tends to result in more

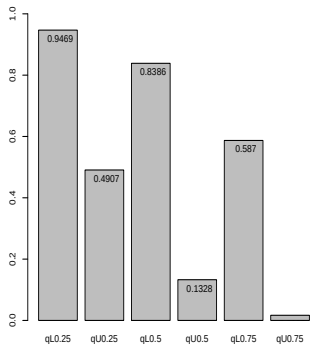
imprecision. For comparison, using the same simulated data set from the Arrhenius-Weibull model in Case A, we now assume the Weibull distribution at each stress level and the likelihood ratio test is used between the stress levels K_i and K_0 in the analysis of Case C, to get the intervals $[\underline{\gamma}_i, \overline{\gamma}_i]$ of values γ for which we do not reject the null hypothesis that the data transformed from level i to level 0, and the original data from level 0, is derived from the same underlying distribution, where $i = 1, 2$. The results for this simulation are presented in Table 4.16. We can see from these results, in Table 4.14 where we did not assume a Wibull distribution at each stress level and use the log-rank test, there appears be more imprecision where we use a small sample size $n = 10$ than in Table 4.16 where we did assume the Weibull distribution at each stress level and use the likelihood ratio test. However, with $n = 50$ and $n = 100$ the imprecision is quite similar in both simulations.

Throughout these simulations in this section, the proposed predictive inference method provides insight into whether or not the presented method shows predictive inference if the model assumptions are fully correct. All of these simulations show that the proposed method performs well. Using imprecision around the link function provides more robustness against the model assumptions. We have found that using $[\underline{\gamma}, \overline{\gamma}]$ from the pairwise stress levels K_i and K_0 provides adequate imprecision for the link function parameter if the model assumptions are correct. In the next section, we will apply the proposed method to investigate robustness in the case of model misspecification.

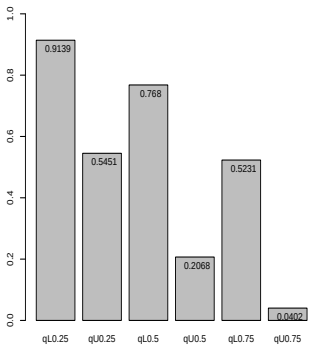
K_1K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9469	0.4907	0.8553	0.6264	0.8300	0.6708
	0.50	0.8386	0.1328	0.6736	0.3218	0.6346	0.3687
	0.75	0.5870	0.0168	0.4326	0.0888	0.3822	0.1253
0.05	0.25	0.9139	0.5451	0.8343	0.6567	0.8144	0.6938
	0.50	0.7680	0.2068	0.6331	0.3635	0.6029	0.4014
	0.75	0.5231	0.0402	0.3975	0.1222	0.3497	0.1545
0.1	0.25	0.8957	0.5765	0.8214	0.6722	0.8049	0.7029
	0.50	0.7303	0.2488	0.6147	0.3858	0.5879	0.4163
	0.75	0.4845	0.0619	0.3763	0.1415	0.3352	0.1677
K_2K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.8956	0.5960	0.8049	0.6917	0.7965	0.7140
	0.50	0.7122	0.2740	0.5932	0.4073	0.5703	0.4375
	0.75	0.4782	0.0580	0.3521	0.1599	0.3159	0.1856
0.05	0.25	0.8593	0.6336	0.7949	0.7076	0.7869	0.7224
	0.50	0.6609	0.3294	0.5701	0.4280	0.5556	0.4522
	0.75	0.4220	0.0960	0.3299	0.1774	0.3004	0.2005
0.1	0.25	0.8450	0.6517	0.7880	0.7149	0.7830	0.7273
	0.50	0.6357	0.3545	0.5588	0.4409	0.5469	0.4616
	0.75	0.3943	0.1165	0.3197	0.1879	0.2922	0.2090
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9498	0.4857	0.8565	0.6251	0.8302	0.6696
	0.50	0.8415	0.1237	0.6756	0.3192	0.6358	0.3673
	0.75	0.5917	0.0113	0.4344	0.0869	0.3840	0.1244
0.05	0.25	0.9197	0.5360	0.8363	0.6539	0.8156	0.6919
	0.50	0.7777	0.1939	0.6374	0.3596	0.6058	0.3988
	0.75	0.5323	0.0306	0.4006	0.1189	0.3528	0.1511
0.1	0.25	0.9036	0.5650	0.8239	0.6692	0.8075	0.7007
	0.50	0.7437	0.2327	0.6205	0.3811	0.5916	0.4130
	0.75	0.4960	0.0483	0.3815	0.1356	0.3387	0.1639

Table 4.14: Proportion of runs with future observation greater than the quartiles, Case *A*.

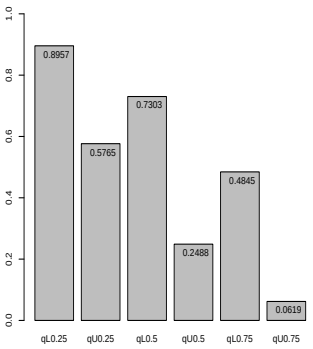
(a) K_1 and K_0 (0.01)



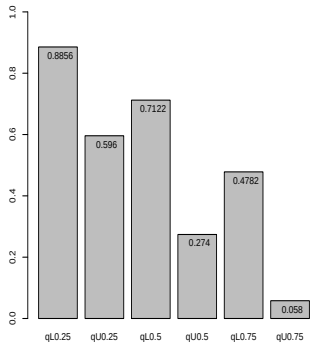
(b) K_1 and K_0 (0.05)



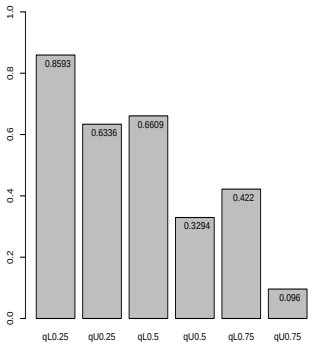
(c) K_1 and K_0 (0.1)



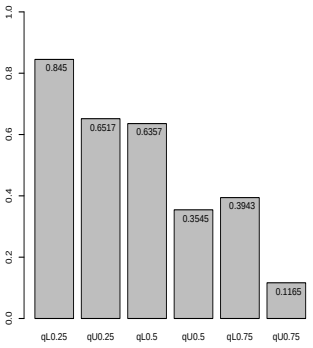
(d) K_2 and K_0 (0.01)



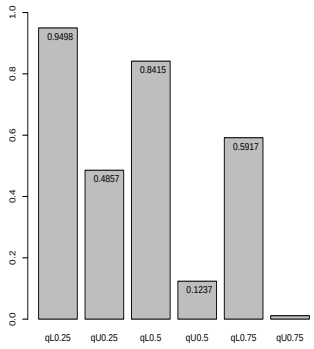
(e) K_2 and K_0 (0.05)



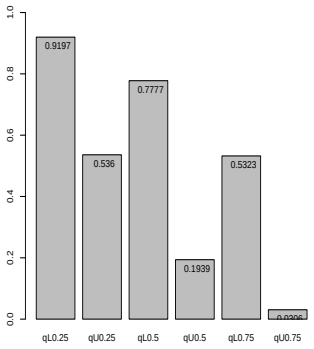
(f) K_2 and K_0 (0.1)



(g) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.01)



(h) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.05)



(i) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.1)

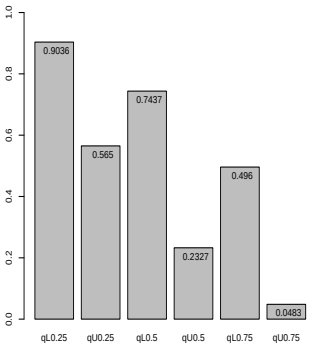
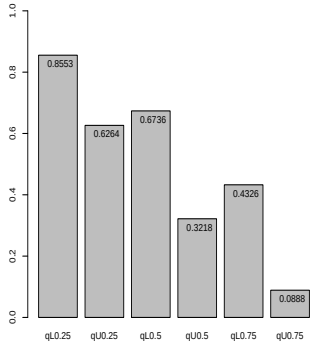
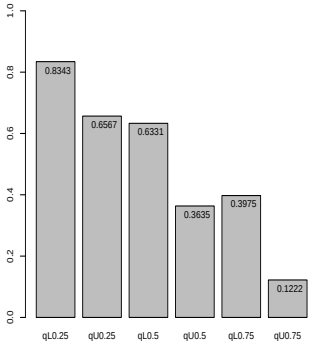


Figure 4.8: Proportion of runs with future observation greater than the quartiles, $n = 10$. Case A.

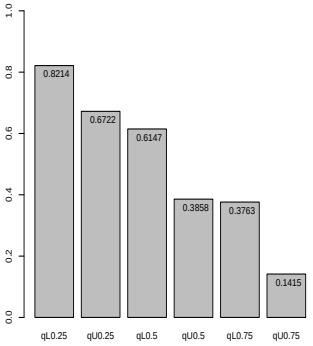
(a) K_1 and K_0 (0.01)



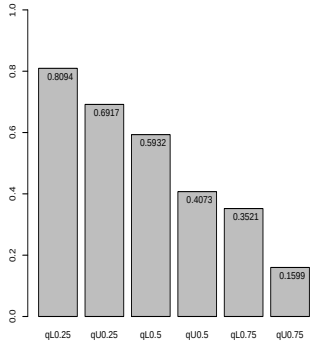
(b) K_1 and K_0 (0.05)



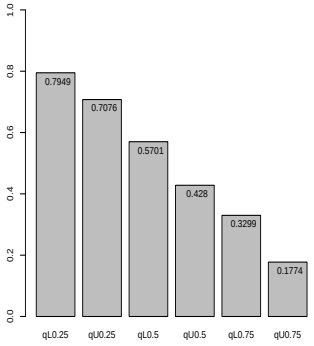
(c) K_1 and K_0 (0.1)



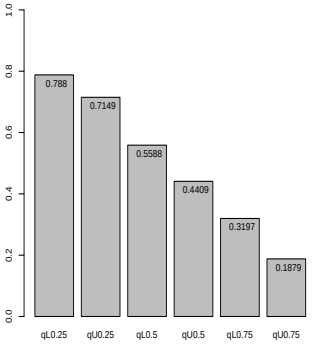
(d) K_2 and K_0 (0.01)



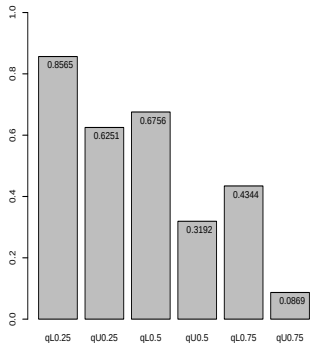
(e) K_2 and K_0 (0.05)



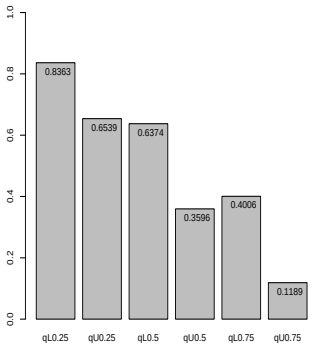
(f) K_2 and K_0 (0.1)



(g) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.01)



(h) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.05)



(i) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.1)

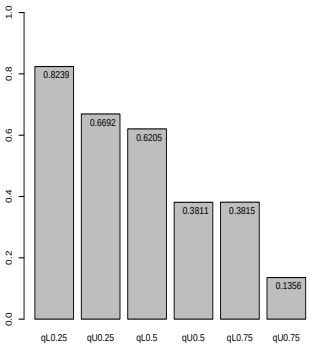
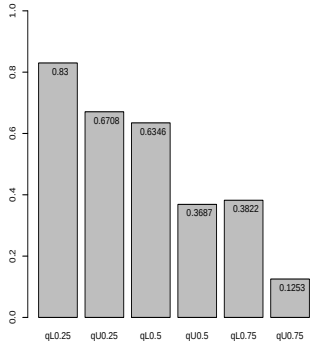
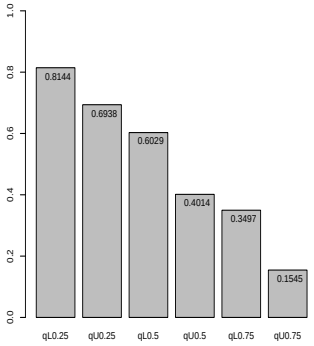


Figure 4.9: Proportion of runs with future observation greater than the quartiles, $n = 50$. Case A.

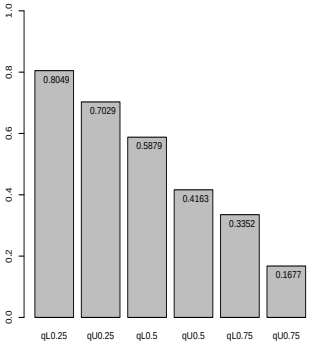
(a) K_1 and K_0 (0.01)



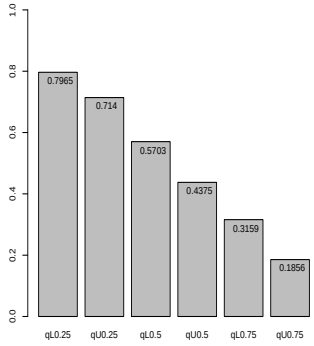
(b) K_1 and K_0 (0.05)



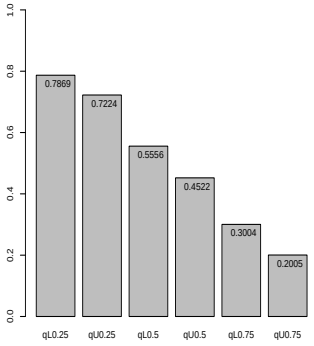
(c) K_1 and K_0 (0.1)



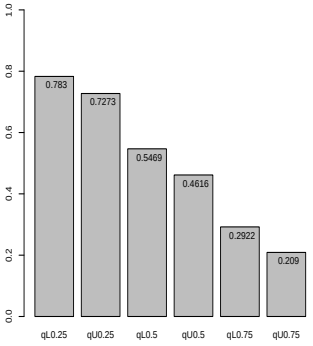
(d) K_2 and K_0 (0.01)



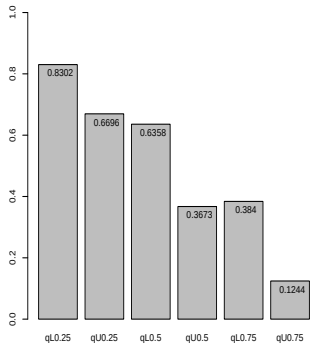
(e) K_2 and K_0 (0.05)



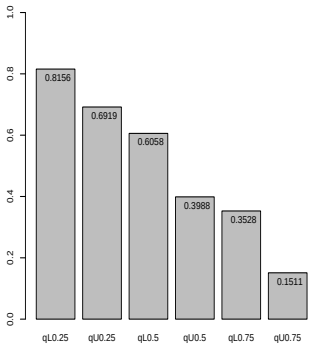
(f) K_2 and K_0 (0.1)



(g) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.01)



(h) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.05)



(i) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.1)

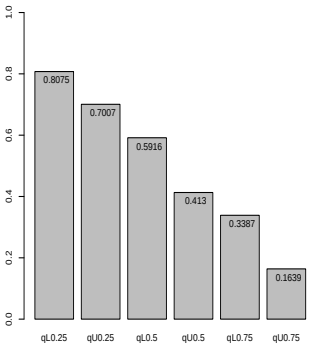


Figure 4.10: Proportion of runs with future observation greater than the quartiles, $n = 100$. Case A.

K_1K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9399	0.4807	0.8575	0.6223	0.8364	0.6636
	0.50	0.8314	0.1286	0.6783	0.3145	0.6454	0.3548
	0.75	0.5872	0.0158	0.4389	0.0861	0.3946	0.1147
0.05	0.25	0.9092	0.5418	0.8384	0.6519	0.8206	0.6841
	0.50	0.7617	0.2051	0.6389	0.3576	0.6152	0.3892
	0.75	0.5206	0.0416	0.4030	0.1197	0.3620	0.1428
0.1	0.25	0.8926	0.5749	0.8259	0.6673	0.8112	0.6957
	0.50	0.7216	0.2508	0.6190	0.3795	0.5984	0.4063
	0.75	0.4818	0.0616	0.3815	0.1368	0.3459	0.1582
K_2K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.8913	0.5770	0.8258	0.6727	0.8121	0.6982
	0.50	0.7272	0.2493	0.6173	0.3809	0.5993	0.4054
	0.75	0.4947	0.0473	0.3774	0.1386	0.3458	0.1566
0.05	0.25	0.8657	0.6194	0.8089	0.6904	0.8027	0.7065
	0.50	0.6767	0.3122	0.5934	0.4060	0.5802	0.4226
	0.75	0.4386	0.0820	0.3536	0.1594	0.3260	0.1719
0.1	0.25	0.8527	0.6400	0.8013	0.6972	0.7954	0.6443
	0.50	0.6495	0.3396	0.5805	0.4173	0.5666	0.3895
	0.75	0.4097	0.1062	0.3406	0.1711	0.3119	0.2324
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9434	0.4745	0.8595	0.6203	0.8375	0.6622
	0.50	0.8360	0.1160	0.6818	0.3110	0.6483	0.3526
	0.75	0.5930	0.0102	0.4409	0.0835	0.3971	0.1126
0.05	0.25	0.9158	0.5323	0.8412	0.6486	0.8229	0.6815
	0.50	0.7744	0.1900	0.6450	0.3518	0.6195	0.3844
	0.75	0.5330	0.0298	0.4080	0.1142	0.3670	0.1382
0.1	0.25	0.9016	0.5634	0.8304	0.6612	0.8149	0.6281
	0.50	0.7387	0.2322	0.6268	0.3713	0.6041	0.3621
	0.75	0.4976	0.0468	0.3895	0.1304	0.3516	0.2056

Table 4.15: Proportion of runs with future observation greater than the quartiles, Case B.

K_1K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9387	0.4952	0.8563	0.6270	0.8307	0.6693
	0.50	0.8225	0.1406	0.6741	0.3185	0.6379	0.3654
	0.75	0.5697	0.0185	0.4331	0.0880	0.3847	0.1228
0.05	0.25	0.9058	0.5542	0.8347	0.6570	0.8158	0.6921
	0.50	0.7550	0.2188	0.6340	0.3628	0.6062	0.3967
	0.75	0.5046	0.0464	0.3964	0.1236	0.3529	0.1518
0.1	0.25	0.8878	0.5804	0.8217	0.6729	0.8057	0.7013
	0.50	0.7157	0.2609	0.6149	0.3832	0.5898	0.4124
	0.75	0.4669	0.0666	0.3762	0.1393	0.3378	0.1656
K_2K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.8808	0.6051	0.8112	0.6901	0.7992	0.7110
	0.50	0.6983	0.2842	0.5960	0.4030	0.5760	0.4324
	0.75	0.4606	0.0643	0.3540	0.1555	0.3229	0.1794
0.05	0.25	0.8552	0.6405	0.7964	0.7061	0.7897	0.7185
	0.50	0.6518	0.3386	0.5736	0.4250	0.5590	0.4471
	0.75	0.4082	0.1035	0.3334	0.1758	0.3043	0.1956
0.1	0.25	0.8400	0.6566	0.7900	0.7133	0.7850	0.7231
	0.50	0.6286	0.3621	0.5623	0.4374	0.5519	0.4555
	0.75	0.3831	0.1220	0.3198	0.1862	0.2957	0.2044
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9424	0.4917	0.8576	0.6258	0.8315	0.6681
	0.50	0.8276	0.1349	0.6763	0.3163	0.6395	0.3645
	0.75	0.5762	0.0142	0.4351	0.0866	0.3863	0.1212
0.05	0.25	0.9130	0.5470	0.8366	0.6553	0.8169	0.6894
	0.50	0.7653	0.2076	0.6390	0.3594	0.6099	0.3941
	0.75	0.5150	0.0374	0.4011	0.1199	0.3566	0.1484
0.1	0.25	0.8960	0.5714	0.8247	0.6696	0.8087	0.6979
	0.50	0.7313	0.2483	0.6212	0.3791	0.5944	0.4081
	0.75	0.4797	0.0543	0.3818	0.1343	0.3421	0.1618

Table 4.16: Proportion of runs with future observation greater than the quartiles, Arrhenius-Weibull model, Case C.

4.6 Simulation study of robustness

In the previous section simulation showed that when we generated the data from the Arrhenius model and power-law, hence with the analysis from the same models, our method performs well. In this section we report on a simulation study to investigate the robustness of the method. As mentioned before, our aim of our new method is to develop a quite straightforward method of predictive inference based on few assumptions, where imprecision in the link function between different stress levels provides robustness against the model assumptions. To illustrate this we have three cases. In Case 1, we simulated new data sets as before from the Arrhenius-Weibull model with scale parameter $\alpha = 7000$, shape parameter $\beta = 3$, and Arrhenius parameter $\gamma = 5200$, with the three temperature levels set at $K_0 = 283, K_1 = 313$, and $K_2 = 353$ Kelvin corresponding to the normal stress level K_0, and the increased stress levels K_1 and K_2, respectively. We ran the simulation 10,000 times, with different sample sizes $n = 10, 50, 100$ observations at each stress level. However, in this simulation study, all the samples at the stress level K_1 are multiplied by 1.2. Using these generated data, so with data multiplied by 1.2 at stress level K_1, we again applied our method as before.

Table 4.17 of Case 1 presents the results of these simulations with $n = 10, 50, 100$, hence all the samples at stress level K_1 are multiplied by 1.2, with attention to the prediction of the simulated future observation at level K_0. This table shows some robustness for the imprecision in our method when this misspecification case is considered. For $n = 50, 100$ there are a few cases for which the simulated future observation for each run at the normal stress level has exceeded the first, second, and third quartiles of the NPI upper survival functions just over $0.75, 0.50$ and 0.25, respectively, see Table 4.17 of the pairwise level $K_1 \times 1.2$ to K_0. Also, the simulated future observation for each run at the normal stress level has exceeded the NPI lower survival functions in just under $0.75, 0.50$ and 0.25 of the first, second, and third quartiles, respectively, see Table 4.17 of the pairwise level K_2 to K_0. Note that in this simulation, because of the effect of multiplying the data at level K_1 by 1.2 which makes the data larger, the lower and upper $[\underline{\gamma_1}, \overline{\gamma_1}]$ become smaller compared to the earlier simulations in Section 4.5 (Case A). Note that the transformation

in Table 4.17 for the pairwise level $K_1 \times 1.2$ to K_0 with $0.01, 0.05, 0.1$ significance levels are based on the $[\underline{\gamma}, \overline{\gamma}] = [\underline{\gamma_1}, \overline{\gamma_1}]$ intervals and the transformation in Table 4.17 with $0.01, 0.05, 0.1$ significance levels are based on the $[\underline{\gamma}, \overline{\gamma}] = [\underline{\gamma_2}, \overline{\gamma_2}]$ intervals for pairwise stress levels K_2 and K_0. However, using our proposed approach in Section 4.2, we take the minimum $\underline{\gamma_i}$ and the maximum $\overline{\gamma_i}$ of the pairwise levels $K_1 \times 1.2$ and K_0 or K_2 and K_0 with $0.01, 0.05, 0.1$ significance levels and go in the widest case to achieve more imprecision, where $i = 1, 2$. As mentioned in Section 4.2, in case of poor model fit, the resulting interval $[\underline{\gamma}, \overline{\gamma}]$ in this case tends to be wider than in Case A in Section 4.5 of good model fit.

Table 4.18 of Case 2 shows the results of a similar simulation as before, however, all the samples at stress level K_1 are multiplied by 0.8. Note that in this simulation, because of the effect of multiplying the data at level K_1 by 0.8 which makes the data smaller, the lower and upper $[\underline{\gamma_1}, \overline{\gamma_1}]$ become larger compared to the earlier simulations in Section 4.5 (Case A). Again, the results for this simulation study support the same conclusion as those just described in Case 1 in this section, with attention on the prediction of one future observation at the normal stress level K_0 and how well it mixes with actual data at the normal stress level. From these simulations, we show that our new proposed method provides some robustness in predictive inference against the model assumptions in the case of model misspecifcation.

$K_1 \times (1.2), K_0$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9742	0.5342	0.9197	0.7153	0.9000	**0.7566**
	0.50	0.8663	0.1962	0.7669	0.4465	0.7404	**0.5047**
	0.75	0.5844	0.0409	0.4711	0.1872	0.4473	0.2408
0.05	0.25	0.9562	0.6102	0.9018	0.7452	0.8864	0.7779
	0.50	0.8245	0.3008	0.7373	0.4905	0.7189	0.5389
	0.75	0.5378	0.0926	0.4474	0.2287	0.4275	0.2708
0.1	0.25	0.9441	0.6460	0.8938	**0.7594**	0.8800	0.7895
	0.50	0.7992	0.3529	0.7220	**0.5122**	0.7067	0.5536
	0.75	0.5121	0.1241	0.4367	0.2491	0.4176	0.2818
K_2K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.8682	0.5452	0.7789	0.6458	0.7638	0.6754
	0.50	0.6658	0.2125	0.5376	0.3453	0.5146	0.3690
	0.75	0.3973	0.0310	0.2791	0.1079	0.2513	0.1273
0.05	0.25	0.8361	0.5895	0.7641	0.6619	0.7533	0.6871
	0.50	0.6119	0.2644	0.5145	0.3691	**0.4988**	0.3871
	0.75	0.3544	0.0596	0.2555	0.1271	**0.2371**	0.1399
0.1	0.25	0.8187	0.6073	0.7572	0.6696	**0.7472**	0.6928
	0.50	0.5884	0.2922	0.5029	0.3805	**0.4911**	0.3965
	0.75	0.3305	0.0773	**0.2446**	0.1346	**0.2294**	0.1460
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9742	0.4957	0.9197	0.6443	0.9000	0.6754
	0.50	0.8663	0.1446	0.7669	0.3438	0.7404	0.3690
	0.75	0.5844	0.0145	0.4711	0.1058	0.4473	0.1273
0.05	0.25	0.9563	0.5554	0.9018	0.6612	0.8864	0.6871
	0.50	0.8248	0.2219	0.7373	0.3687	0.7189	0.3871
	0.75	0.5390	0.0389	0.4474	0.1267	0.4275	0.1399
0.1	0.25	0.9446	0.5802	0.8938	0.6691	0.8800	0.6928
	0.50	0.7999	0.2599	0.7220	0.3802	0.7067	0.3965
	0.75	0.5129	0.0574	0.4367	0.1344	0.4176	0.1460

Table 4.17: Proportion of runs with future observation greater than the quartiles, Case 1.

$K_1 \times (0.8), K_0$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9000	0.4928	0.7805	0.5695	0.7502	0.6024
	0.50	0.7544	0.1179	0.5346	0.2154	**0.4880**	0.2560
	0.75	0.5321	0.0070	0.2793	0.0151	**0.2186**	0.0382
0.05	0.25	0.8571	0.5168	0.7550	0.5919	**0.7345**	0.6224
	0.50	0.6594	0.1560	**0.4965**	0.2522	**0.4585**	0.2682
	0.75	0.4324	0.0148	**0.2330**	0.0268	**0.1867**	0.0401
0.1	0.25	0.8316	0.5367	**0.7430**	0.6050	**0.7263**	0.6322
	0.50	0.6165	0.1833	**0.4755**	0.2704	**0.4429**	0.2904
	0.75	0.3782	0.0211	**0.2085**	0.0364	**0.1717**	0.0475
K_2K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9087	0.6670	0.8504	**0.7504**	0.8343	**0.7699**
	0.50	0.7657	0.3566	0.6603	0.4882	0.6484	**0.5201**
	0.75	0.5486	0.0965	0.4224	0.2207	0.3953	**0.2509**
0.05	0.25	0.8871	0.7015	0.8401	**0.7619**	0.8283	**0.7782**
	0.50	0.7217	0.4049	0.6421	**0.5095**	0.6332	**0.5349**
	0.75	0.4971	0.1432	0.4007	0.2441	0.3789	**0.2662**
0.1	0.25	0.8769	0.7177	0.8334	**0.7691**	0.8252	**0.7818**
	0.50	0.6991	0.4344	0.6316	**0.5187**	0.6253	**0.5431**
	0.75	0.4676	0.1693	0.3908	**0.2565**	0.3679	**0.2745**
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9250	0.4927	0.8505	0.5695	0.8343	0.6024
	0.50	0.8026	0.1179	0.6605	0.2154	0.6484	0.2560
	0.75	0.5832	0.0068	0.4227	0.0151	0.3953	0.0382
0.05	0.25	0.8978	0.5168	0.8401	0.5919	0.8283	0.6224
	0.50	0.7419	0.1553	0.6422	0.2522	0.6332	0.2782
	0.75	0.5210	0.0145	0.4007	0.0268	0.3789	0.0401
0.1	0.25	0.8850	0.5365	0.8334	0.6050	0.8252	0.6322
	0.50	0.7152	0.1826	0.6317	0.2704	0.6253	0.2904
	0.75	0.4847	0.0207	0.3908	0.0364	0.3679	0.0475

Table 4.18: Proportion of runs with future observation greater than the quartiles, Case 2.

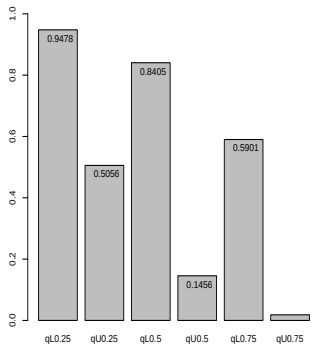
In Case 3, using our new method in Section 4.2, we investigate the robustness and the performance of our predictive inference against the necessary assumptions, where the imprecision in the Arrhenius link function between different stress levels provides robustness against the model assumptions. To perform this, we simulated the data from the Eyring-Weibull model [60] with the parameters $\alpha_0 = 7000, \beta = 3$ and $\lambda = 5200$ using the Eyring link function for the Weibull scale parameters Equation 3.5.1 introduced in section 3.5.

In this simulation, we used the assumed Arrhenius link function model for the analysis. We applied the method described in Section 4.2, with levels of significance 0.01, 0.05, and 0.10, with $10,000$ simulation runs. The results presented in Table 4.19 and Figures 4.11-4.13. These reveal that the proposed method performs well overall, which allows us to conclude that our method shows robustness in predictive inferences. In comparison with the simulation where the model assumptions are fully correct in Case A in Section 4.5, the results in Table 4.19 are very similar to those in Table 4.14, which means that from the preceding investigation, the Eyring model and the Arrhenius model lead to similar conclusions.

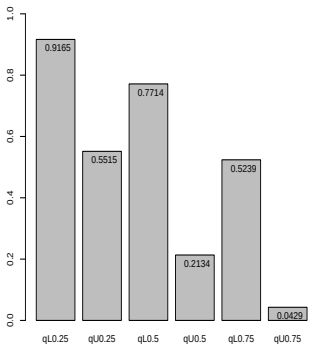
K_1K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9478	0.5056	0.8577	0.6282	0.8320	0.6740
	0.50	0.8405	0.1456	0.6772	0.3242	0.6394	0.3733
	0.75	0.5901	0.0183	0.4342	0.0919	0.3859	0.1291
0.05	0.25	0.9156	0.5515	0.8366	0.6593	0.8168	0.6960
	0.50	0.7714	0.2134	0.6376	0.3663	0.6065	0.4059
	0.75	0.5239	0.0429	0.4013	0.1268	0.3539	0.1577
0.1	0.25	0.8971	0.5818	0.8246	0.6744	0.8069	0.7052
	0.50	0.7340	0.2532	0.6189	0.3892	0.5930	0.4209
	0.75	0.4867	0.0649	0.3805	0.1444	0.3382	0.1724
K_2K_0		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.8840	0.5947	0.8079	0.6900	0.7953	0.7135
	0.50	0.7115	0.2715	0.5909	0.4053	0.5677	0.4348
	0.75	0.4759	0.0575	0.3492	0.1583	0.3136	0.1828
0.05	0.25	0.8585	0.6323	0.7937	0.7054	0.7862	0.7215
	0.50	0.6610	0.3246	0.5684	0.4258	0.5527	0.4506
	0.75	0.4200	0.0941	0.3272	0.1761	0.2983	0.1986
0.1	0.25	0.8444	0.6498	0.7874	0.7127	0.7816	0.7256
	0.50	0.6337	0.3509	0.5569	0.4382	0.5445	0.4597
	0.75	0.3915	0.1150	0.3171	0.1858	0.2907	0.2071
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	q	qL	qU	qL	qU	qL	qU
0.01	0.25	0.9504	0.5003	0.8588	0.6265	0.8322	0.6725
	0.50	0.8429	0.1355	0.6788	0.3213	0.6402	0.3715
	0.75	0.5941	0.0125	0.4357	0.0896	0.3868	0.1277
0.05	0.25	0.9214	0.5416	0.8384	0.6559	0.8179	0.6938
	0.50	0.7804	0.1993	0.6409	0.3622	0.6089	0.4017
	0.75	0.5325	0.0316	0.4035	0.1219	0.3566	0.1542
0.1	0.25	0.9047	0.5692	0.8266	0.6704	0.8088	0.7023
	0.50	0.7466	0.2362	0.6234	0.3833	0.5958	0.4163
	0.75	0.4968	0.0495	0.3846	0.1370	0.3412	0.1671

Table 4.19: Proportion of runs with future observation greater than the quartiles, Case 3.

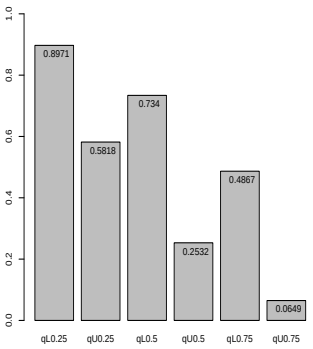
(a) K_1 and K_0 (0.01)



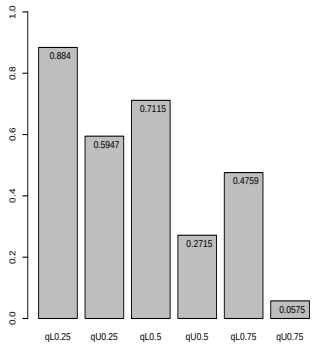
(b) K_1 and K_0 (0.05)



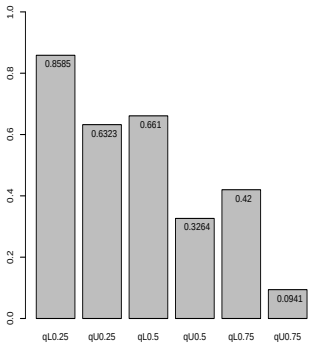
(c) K_1 and K_0 (0.1)



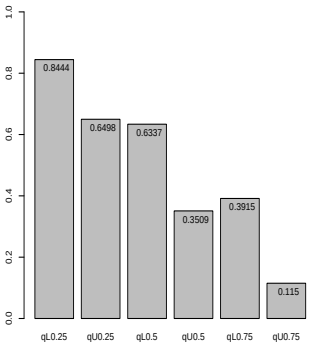
(d) K_2 and K_0 (0.01)



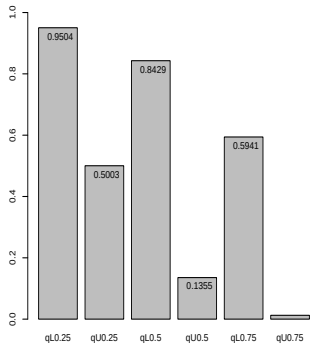
(e) K_2 and K_0 (0.05)



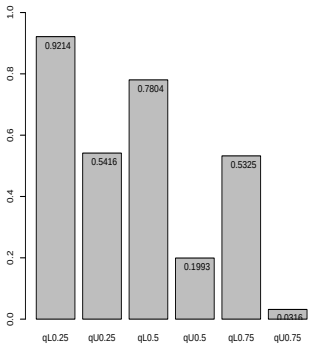
(f) K_2 and K_0 (0.1)



(g) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.01)



(h) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.05)



(i) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.1)

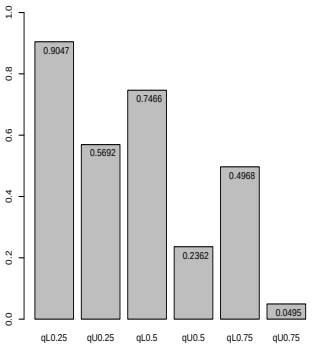
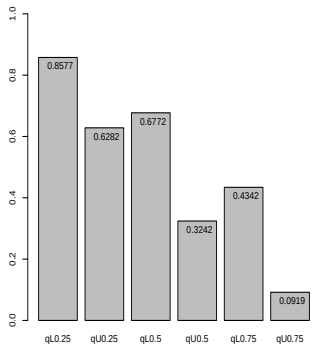
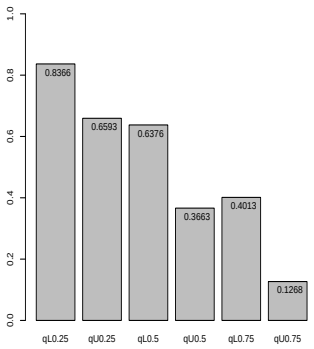


Figure 4.11: Proportion of runs with future observation greater than the quartiles, $n = 10$. Case 3.

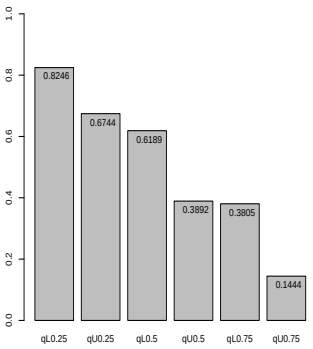
(a) K_1 and K_0 (0.01)



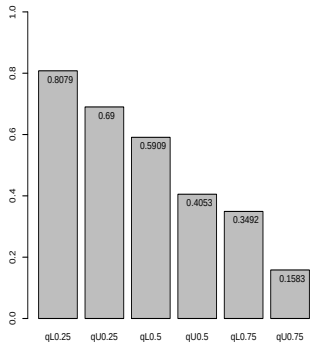
(b) K_1 and K_0 (0.05)



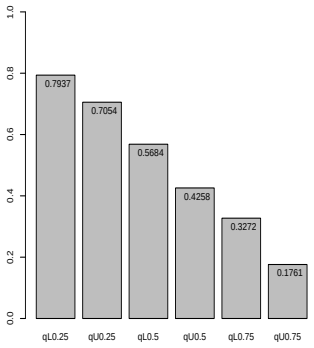
(c) K_1 and K_0 (0.1)



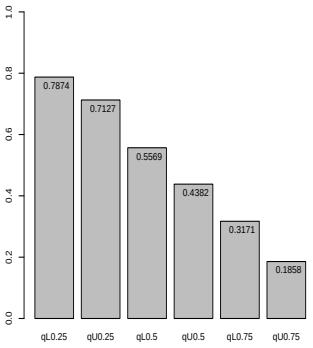
(d) K_2 and K_0 (0.01)



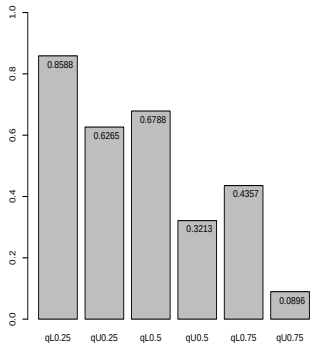
(e) K_2 and K_0 (0.05)



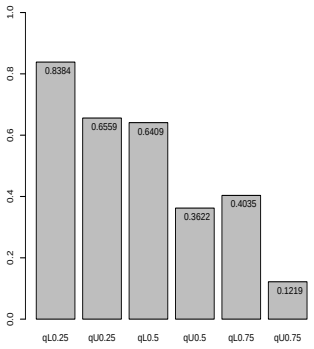
(f) K_2 and K_0 (0.1)



(g) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.01)



(h) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.05)



(i) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.1)

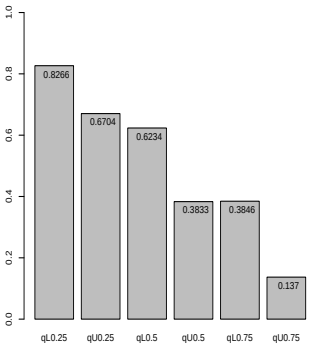
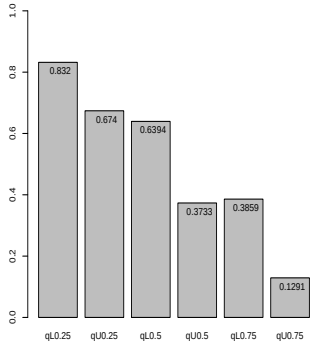
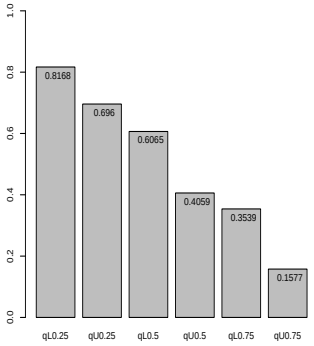


Figure 4.12: Proportion of runs with future observation greater than the quartiles, $n = 50$. Case 3.

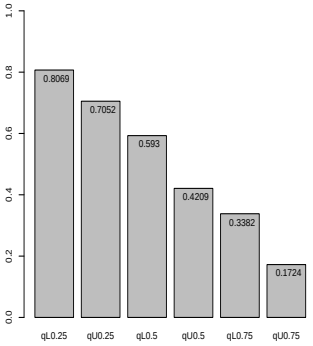
(a) K_1 and K_0 (0.01)



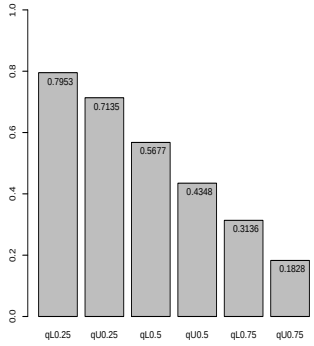
(b) K_1 and K_0 (0.05)



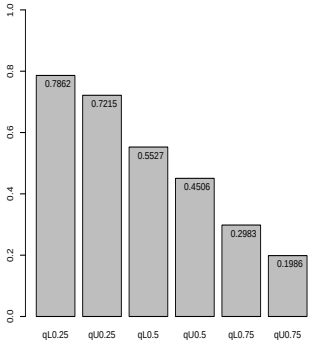
(c) K_1 and K_0 (0.1)



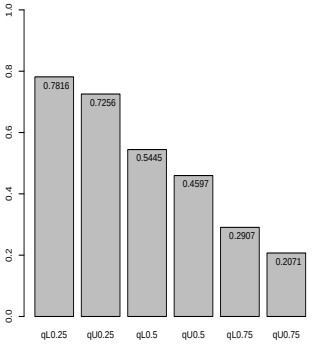
(d) K_2 and K_0 (0.01)



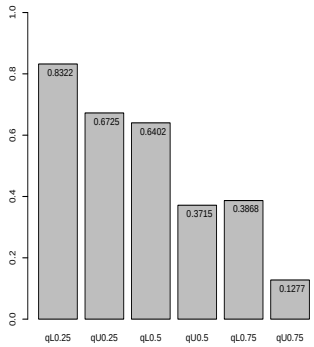
(e) K_2 and K_0 (0.05)



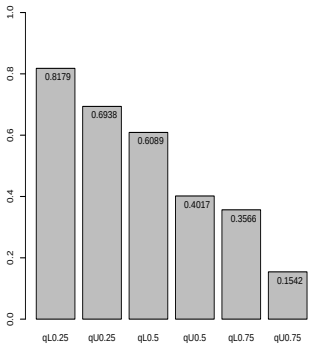
(f) K_2 and K_0 (0.1)



(g) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.01)



(h) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.05)



(i) Using $\underline{\gamma}$, $\overline{\gamma}$, (0.1)

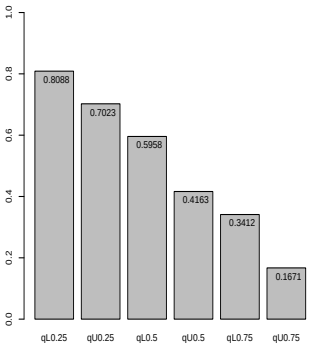


Figure 4.13: Proportion of runs with future observation greater than the quartiles, $n = 100$. Case 3.

4.7 Conclusions

In this chapter we presented a novel statistical method of imprecise semi-parametric inference for ALT data. In this chapter, we do not assume a failure time distribution at each stress level. The proposed method applies the use of the log-rank test to compare the survival distribution of pairwise stress levels, in combination with the Arrhenius model to find the interval of γ values. We developed imprecision through the use of nonparametric tests for the parameter of the link function between different stress levels, which enabled us to transform the observations at increased stress levels to interval-valued observations at the normal stress level and achieve robustness. The main findings drawn from this chapter are: we obtain an interval for the parameter of the link function, which is assumed at each stress level, by applying classical hypothesis testing between the pairwise stress levels to determine the level of imprecision. We showed why, in our method, we use the imprecision from combined pairwise log-rank tests, and not from a single log-rank test on all stress levels together. The latter would lead to less imprecision if the model fits poorly, while our proposed method leads to more imprecision. We have found that the end resulting $[\underline{\gamma}_i, \overline{\gamma}_i]$ intervals get wider when we have more censored observations.

Throughout this research, we have presented two main contributions. First, Chapter 3 presented a new imprecise statistical method for ALT data with imprecision based on the likelihood ratio test to define the interval of values of the parameter γ of the Arrhenius link function. Secondly, Chapter 4, presented a similar method, but we defined the interval of values of the parameter γ of the Arrhenius link function based on the log-rank test. Comparing these two scenarios, the results in the examples and the simulations show that we have more imprecision when we apply the nonparametric test than when we apply the likelihood ratio test with the assumption of a Weibull distribution at each stress level.

As with any novel statistical method developed for real-world applications, the real value of our method should be shown in practical applications. To implement the methods, no more is needed on the modelling side than for the classic inference methods with the same model assumptions, rather the main question is how one can use the resulting lower and upper survival functions to support real-world decisions.

Further, we investigate this important aspect in the context of warranty contracts, which will be discussed in the next chapter.

Chapter 5

Study of warranties with ALT data

5.1 Introduction

In this chapter, we will illustrate a possible application of our new method using both approaches presented in Chapters 3 and 4. In particular, we focus on warranties and illustrate how our predictive inference can be used for inference on expected costs of warranty contracts. This section briefly introduces basic warranties considered in this chapter.

Products which include a warranty incur added costs to the manufacturer (or the consumer on occasion) for honouring the terms of the warranty: the warranty cost. This cost is related to a number of factors; the reliability of the product being the key factor. Products which fail within the warranty period entail the manufacturer taking responsibility for honouring the warranty, usually either by refunding or replacing faulty goods [57]. Generally, a warranty guarantees that a given product will provide reliable service for a defined period of time [69]. A warranty represents a contractual relationship between the manufacturer and the consumer that a specific product will provide reliable service and is absent of material or manufacturing defects, and, that if such defects cause the product to fail, it will be refunded, repaired or replaced at the manufacturer's expense [13, 14]. However, a warranty is non-binding if the product has been used outside of certain specified conditions and manufacturers have no obligation to service the product in this case. Further, a warranty also outlines the limits of the manufacturers' liability when a product is

not used as intended.

Warranties are of equal importance to both consumers and manufacturers [54]. For example, consumers want to be confident the product they have purchased will function well. Warranties reassure consumers a product is of suitable quality and unlikely to develop a fault due to standardization issues, design faults or workmanship. On the other hand, manufacturers or distributors use warranties to safeguard their reputation and increase sales [54]. For example, providing a warranty lowers customers' sense of risk in buying a product and encourages trust in the manufacturer's products. Warranties can also increase sales by offering guaranteed reliability [54]. Offering a replacement or a refund of the customer's original purchase price is an effective way of promoting a brand and increasing consumer demand [54].

Warranties also provide manufacturers with a level of protection against unfair demands for a refund or replacement by stating their responsibilities [54]. For example, while the manufacturer guarantees the consumer will receive a particular standard of performance from a product, this reduces unreasonable consumer demands that cause a financial loss. Finally, warranties also help manufacturers to gather consumer information for use in marketing and identify potential quality or workmanship issues [54].

Accelerated life testing (ALT) plays a key role in the manufacturing industry in terms of product design and development processes [45]. Indeed, the growth in competition within design innovation and the drive to slash product development timescales also underline how important ALT-based approaches are in product design and development [45]. At present, products are checked under hard conditions to cause the types of failures that occur in real-life applications [45]. This produces an amount of data including failure mechanisms, causes, and aspects of probability distributions of failure times which indicate a product's reliability in the field under normal use. These data can also be useful for highlighting further design modifications to enhance reliability [45]. However, determining a product's reliability under normal conditions from the ALT data requires extrapolation in the form of a life-stress relationship [45, 60, 76], as described in the introduction of Chapter 1 and Section 2.2.

ALT is widely used for reliability testing, predicting warranty cost, and assessment, and the comparison of a range of different approaches to solving product design issues [45]. Here, a well-rounded knowledge of statistical data analysis and validation techniques are key; indeed, the complexity of statistical models has led to the recruitment of researchers from a wide range of related fields, and this has become multidisciplinary with computational mathematics, statistics, and engineering all taking part [45].

A literature on warranties is available with focus on different perspectives. For example, Blischke and Murthy [14] provide an overview of warranty cost analysis. Various methods for analyzing and pricing warranty contracts can be found in review articles [39, 55, 56, 69]. Recently, researchers have been considering pricing warranty contracts based on ALT data. Yang [78] presented a design for accelerated life testing plans to predict warranty costs, assuming that the manufacturer offers a free replacement warranty policy [78]. He developed a test plan to minimize the analysis of the asymptotic variance of the maximum likelihood estimate of the warranty cost [78]. Meeker et al. [51] propose a simple use-rate model to predict the failure time distribution for a future component using accelerated life tests results. Zhao and Xie [82] use ALT data to predict warranty cost and risk warranty under imperfect repair. Their goal is to predict the expected warranty cost and provide confidence intervals for it [82].

Generally, in imprecise probability theory [10], lower and upper expectations of a real-valued random quantity X, denoted by $\underline{E}X$ and $\overline{E}X$ respectively, can be interpreted in terms of prices as follows. The lower expectation can be regarded as the maximum buying price for X, meaning that one would be willing to pay any amount up to $\underline{E}X$ in order to receive the random amount X. The upper expectation can be regarded as the minimum selling price for X, meaning that one would be willing to sell the random amount X for any price greater than $\overline{E}X$. Whilst these interpretations may sometimes be somewhat difficult to link to reality, in our setting of warranties, one can use them and consider them, for example, as insurance prices. If a producer takes on a warranty with a random cost X, then they would prefer to pay a fixed cost up to the lower expectation of X instead of having to pay the

random X. Similarly, they would certainly prefer to pay out the random amount X instead of any fixed amount greater than the upper expectation.

Reliability and liability are the most important factors for products. To sell new products to the consumer, producers must predict the expected warranty cost to indicate future warranty claims [13, 82]. In this chapter, we consider pricing basic warranty contracts based on information from ALT data and the use of our novel imprecise probabilistic statistical method, are described in Chapters 3 and 4. The new statistical methods we introduced in this thesis include imprecision based on the likelihood ratio test and log-rank test which provide robustness with regard to the model assumptions. In this chapter, we derive bounds for expected costs of warranties based on ALT data and using the NPI lower and upper survival functions resulting from our new statistical methods.

This chapter is organized as follows. In Section 5.2, we consider a warranty policy with a fixed penalty cost based on the ALT data and our predictive method, as an example of how our new methods can be applied to support decisions on warranties, followed by three examples which illustrate the proposed method. In Section 5.3, we consider a warranty policy with the penalty cost per unit of time, followed by two examples which illustrates the proposed method. In Section 5.4, we present some concluding remarks.

5.2 Policy A: fixed penalty cost

We will explore the use of our new methods presented above for decision making with regard to warranties. In this section, we consider a simple warranty contract and consider the expected warranty cost based on information from ALT data and the use of our novel imprecise predictive probabilistic statistical methods. Our goal is to predict the lower expected warranty cost $\underline{E}C$ and the upper expected warranty cost $\overline{E}C$.

Suppose that T_w is the warranty period. Once the period of the warranty T_w is complete, the warranty on the product expires. The basic contract states that manufacturer agrees to pay the penalty cost W if the product fails by the warranty

period T_w. Under this basic policy, the manufacturer agrees to refund a fixed penalty cost independent of the failure time, if a product fails by time T_w.

The expected value of the cost of this contract, in a single application, is

$$EC(T_w, W) = W \times (1 - S(T_w)).$$

where $S(.)$ is the item's survival function. We do not know $S(.)$ but we use our methods and assumed ALT data to have bounds.

Obviously, EC is monotonously decreasing as a function of $S(T_w)$. Therefore, the lower expected cost $\underline{E}C$ is related to the NPI upper survival function $\overline{S}$, and the upper expected cost $\overline{E}C$ is related to the NPI lower survival function $\underline{S}$. So, the lower expected cost of the payment under the warranty contract is

$$\underline{E}C(T_w, W) = W \times (1 - \overline{S}_{X_{n+1}}(T_w)). \tag{5.2.1}$$

and the upper expected cost of the payment under the warranty contract is

$$\overline{E}C(T_w, W) = W \times (1 - \underline{S}_{X_{n+1}}(T_w)). \tag{5.2.2}$$

Example 5.2.1. Consider the Example 4.4.4 in Section 4.4 and the warranty policy A mentioned in this section 5.2: incorporating a fixed penalty failure cost. In this example we consider two cases. Note that the period of the warranty T_w for all policies in Case 1 and Case 2 is the same. In Case 1, we use the data from Table 4.7, where $n = 10$ observations at each stress level, and our lower and upper survival functions in Figures 4.3 in Example 4.4.4 using the interval $[\underline{\gamma}, \overline{\gamma}] = [0.153, 3.642]$ at significance level 0.05. We wish to predict the lower expected cost $\underline{E}C(T_w, W)$ and upper expected cost $\overline{E}C(T_w, W)$ to the producer of the warranty if applied to one future product at normal stress level using Equations 5.2.1 and 5.2.2, respectively. To do so, in Table 5.1, we give the lower and upper expected costs $\underline{E}C$ and $\overline{E}C$ for a range of different scenarios.

In Case 2, we use the data from Table 4.7 and our lower and upper survival functions in Figures 4.4 in Example 4.4.4 using the interval $[\underline{\gamma}, \overline{\gamma}] = [0, 3.990]$ at

Case	T_w	W	$\overline{S}(T_w)$	$\underline{S}(T_w)$	$\underline{E}C(T_w, W)$	$\overline{E}C(T_w, W)$
1	5000	500	0.5594954	0.1203636	220.2523	439.8182
2	5000	1000	0.5594954	0.1203636	440.5046	879.6364
3	5000	2500	0.5594954	0.1203636	1101.2615	2199.091

Table 5.1: $\underline{EC}(T_w, W)$ and $\overline{EC}(T_w, W)$. Example 5.2.1, Case 1.

significance level 0.05. Again, we derive the lower expected cost $\underline{E}C(T_w, W)$ and upper expected cost $\overline{E}C(T_w, W)$ to the producer of the warranty if applied to one future product at normal stress level using Equations 5.2.1 and 5.2.2, respectively, which are given in Tables 5.2.

In Tables 5.1 and 5.2, we present the results of these studies. As mentioned in Section 5.2, that the lower expected cost $\underline{E}C(T_w, W)$ is achieved from multiplying W which is the price of the product by $1-\overline{S}(T_w)$. Similarly, the upper expected cost $\overline{E}C(T_w, W)$ is achieved from multiplying the cost of the product W by $1 - \underline{S}(T_w)$.

Following the general imprecise probability theory [10] discussion in Section 5.1, we can give the following interpretation of the lower expected costs $\underline{E}C(T_w, W)$ and upper expected costs $\overline{E}C(T_w, W)$ to the producer of a warranty. On the basis of the data observations and our new statistical method, we have lower and upper expectations for the expected costs of the policies. So, in terms of the warranty contract in Table 5.1, for the second policy of Case 1, if a producer takes on a warranty with random cost $C_2(T_w, W)$, the producer would prefer to pay a fixed cost up to the lower expectation of 440.5046 instead of having to pay the random $C_2(T_w, W)$. Similarly, they would certainly prefer to pay out the random amount $C_2(5000, 1000)$ instead of any fixed amount greater than the upper expectation of 879.6364, and similar for policies 1 and 3 in Table 5.1. The results for the warranty contract (Case 2) in Table 5.2 support the same conclusion as those just described in Case 1.

If the upper expected cost for policy 1 $\overline{E}C_1(T_w, W)$ less than the lower expected cost for policy 2 $\underline{E}C_2(T_w, W)$ then we strongly prefer policy 1 over policy 2. If the lower expected cost for policy 1 $\underline{E}C_1(T_w, W)$ less than the lower expected cost for policy 2 $\underline{E}C_2(T_w, W)$ and the upper expected cost for policy 1 $\overline{E}C_1(T_w, W)$ less

Policy	T_w	W	$\overline{S}(T_w)$	$\underline{S}(T_w)$	$\underline{E}C(T_w, W)$	$\overline{E}C(T_w, W)$
1	5000	500	0.6403675	0.1201611	179.81625	439.91945
2	5000	1000	0.6403675	0.1201611	359.6325	879.8389
3	5000	2500	0.6403675	0.1201611	899.08125	2199.59725

Table 5.2: $\underline{EC}(T_w, W)$ and $\overline{EC}(T_w, W)$. Example 5.2.1, Case 2.

Policy	T_w	W	$\overline{S}(T_w)$	$\underline{S}(T_w)$	$\underline{E}C(T_w, W)$	$\overline{E}C(T_w, W)$
1	3000	1000	0.9352566	0.8397405	64.7434	160.2595
2	5000	1500	0.8704304	0.5151495	194.344	727.27575
3	5000	2500	0.8704304	0.5151495	323.924	1212.12625
4	7000	2500	0.5471789	0.1594751	1132.05275	2101.31225

Table 5.3: $\underline{E}C(T_w, W)$ and $\overline{E}C(T_w, W)$. Example 5.2.2.

than the upper expected cost for policy 2 $\overline{E}C_2(T_w, W)$ then it make sense to also prefer (perhaps 'weakly') policy 1 over policy 2, although the unknown $C_1(5000, 500)$ could have an expected value greater than $C_2(5000, 1000)$. Of course we got a lot of imprecision between the lower expected costs $\underline{E}C(T_w, W)$ and upper expected costs $\overline{E}C(T_w, W)$ because that the number of observations in this example is not that large. Note that our methods do not give guidance on whether you should select an amount within the interval $[\underline{E}C(T_w, W), \overline{E}C(T_w, W)]$ compared to the fix costs.

Example 5.2.2. Consider the Example 3.4.1 in Section 3.4 and the warranty policy A mentioned in this section 5.2 (fixed penalty failure cost). We use the data from Table 3.1 and our lower and upper survival functions in Figure 3.2(a) in Example 3.4.1 using the interval $[\underline{\gamma}, \overline{\gamma}] = [4593.700, 6100.653]$ at significance level 0.1. We derive the lower expected cost $\underline{E}C(T_w, W)$ and upper expected cost $\overline{E}C(T_w, W)$ to the producer of the warranty if applied to one future product at normal stress level using Equations 5.2.1 and 5.2.2, respectively. To do so, in Table 5.3, we give the lower and upper expected costs $\underline{E}C$ and $\overline{E}C$ for a range of different scenarios.

Using the same interpretation of the lower expected costs $\underline{E}C(T_w, W)$ and upper expected costs $\overline{E}C(T_w, W)$ to producers of the warranty in Example 5.2.1. There-

Case	T_w	W	$\overline{S}(T_w)$	$\underline{S}(T_w)$	$\underline{E}C(T_w, W)$	$\overline{E}C(T_w, W)$
1	3000	1000	0.966794	0.8191574	33.206	180.8426
2	5000	1500	0.7709172	0.4262666	343.6242	860.6001
3	5000	2500	0.7709172	0.4262666	572.707	1434.3335
4	7000	2500	0.4188237	0.1236196	1452.94075	2190.951

Table 5.4: $\underline{E}C(T_w, W)$ and $\overline{E}C(T_w, W)$. Example 5.2.3.

fore, in terms of the warranty contract in Table 5.3, for the first policy, if a producer takes on a warranty with a random cost $C_1(T_w, W)$, then they would prefer to pay a fixed cost up to the lower expectation of 64.7434 instead of having to pay the random $C_1(T_w, W)$. Similarly, they would certainly prefer to pay out the random amount $C_1(T_w, W)$ instead of any fixed amount greater than the upper expectation of 160.2595, and similar for policies $2, 3$, and 4 in Table 5.3.

Example 5.2.3. Consider the Example 3.4.2 in Section 3.4 and the warranty policy A mentioned in this section 5.2 (fixed penalty failure cost). We use the data from Table 3.3, where $n = 20$ observations at each stress level, and our lower and upper survival functions in Figure 3.3 in Example 3.4.2 using the interval $[\underline{\gamma}, \overline{\gamma}]$ $= [4425.681, 5406.786]$ at significance level 0.1. We derive the lower expected cost $\underline{E}C(T_w, W)$ and upper expected cost $\overline{E}C(T_w, W)$ to the producer of the warranty if applied to one future product at normal stress level using Equations 5.2.1 and 5.2.2, respectively. To do so, in Table 5.4, we give the lower and upper expected costs $\underline{E}C$ and $\overline{E}C$ for a range of different scenarios.

Using the same interpretation of the lower expected costs $\underline{E}C(T_w, W)$ and upper expected costs $\overline{E}C(T_w, W)$ to producers of the warranty in the Example 5.2.1. Thus, in terms of the warranty contract in Table 5.4, for the first policy, if a producer takes on a warranty with a random cost $C_1(T_w, W)$, then they would prefer to pay a fixed cost up to the lower expectation of 33.206 instead of having to pay the random $C_1(T_w, W)$. Similarly, they would certainly prefer to pay out the random amount $C_1(T_w, W)$ instead of any fixed amount greater than the upper expectation of 180.8426, and similar for policies $2, 3, 4$, and 5 in Table 5.4. Comparing the impre-

cision in the lower expected costs $\underline{E}C(T_w, W)$ and upper expected costs $\overline{E}C(T_w, W)$ to producers of the warranty in this example as shown in Table 5.4, with the imprecision in the lower expected costs $\underline{E}C(T_w, W)$ and upper expected costs $\overline{E}C(T_w, W)$ to producers of the warranty in the Example 5.2.1 as shown in Table 5.1, it shows that when we generated $n = 20$ observations at each stress level, the imprecision is smaller than with fewer observations in Example 5.2.1.

5.3 Policy B: time dependent penalty cost

In this section, we consider another warranty contract and again consider the expected warranty cost based on information from ALT data and the use of our novel imprecise predictive probabilistic statistical methods. Our goal is to predict the lower expected warranty cost $\underline{E}C$ and the upper expected warranty cost $\overline{E}C$. In policy B, we are looking if a product fails at a certain time then the producer will need to pay a penalty amount (but the penalty is per time unit) until the end of the policy EC, which is the warranty period T_w.

Generally, where the product fails at time t, and if there would be a distribution, we would have to take an integral, given by the following equation

$$EC(T_w, w) = \int_0^{T_w} (T_w - t) w f(t) dt,$$

where $f(t)$ is the PFD of the product failure times. However, for our case, we know that the lower and upper survival functions $\underline{S}$ and $\overline{S}$, respectively, which are related to the upper and lower costs $\overline{E}C$ and $\underline{E}C$, are discrete, which simplifies the computation as $\underline{S}(.)$ and $\overline{S}(.)$ are step functions.

Suppose that T_w is the warranty period. The penalty cost per unit of time that the product is not working is denoted by w. Therefore, if the penalty failure cost that needs to be paid if the product fails before a fixed time T_w, then the penalty cost is equal to $w(T_w - t)$, where t is random. Once the period of the warranty T_w is complete, the warranty on the product expires. This contract states that manufacturer agrees to pay the penalty cost $w(T_w - t)$ if the product fails by

Policy	T_w	w	$\underline{E}C(T_w, w)$	$\overline{E}C(T_w, w)$
1	3000	5	168.260	1463.384
2	5000	2	549.663	2751.085
3	5000	5	1374.160	6878.436
4	7000	2	1365.657	6061.928
5	7000	5	3414.145	15154.822

Table 5.5: $\underline{E}C(T_w, w)$ and $\overline{E}C(T_w, w)$. Example 5.3.1.

the warranty period T_w. The lower expected cost $\underline{E}C$ is related to the NPI upper survival function $\overline{S}$, which is a simple discrete distribution given by

$$\sum_{t=t_1}^{t_{T_w}} P_{\overline{S}_{X_{n+1}}}(t) \times (T_w - t) \times w, \tag{5.3.1}$$

and the upper expected cost $\overline{E}C$ is related to the NPI lower survival function $\underline{S}$, which is a simple discrete distribution given by

$$\sum_{t=t_1}^{t_{T_w}} P_{\underline{S}_{X_{n+1}}}(t) \times (T_w - t) \times w. \tag{5.3.2}$$

Example 5.3.1. Consider the Example 3.4.1 in Section 3.4 and the warranty policy B (time-dependent penalty failure cost). In this example we use the data from Table 3.1 and our lower and upper survival functions in Figures 3.1 in Example 3.4.1 using the interval $[\underline{\gamma}, \overline{\gamma}] = [4060.018, 6605.752]$ at significance level 0.01, as used for Example 5.2.1. We derive the lower expected cost $\underline{E}C(T_w, w)$ and upper expected cost $\overline{E}C(T_w, w)$ to the producer of the warranty if applied to one future product at normal stress level using Equations 5.3.1 and 5.3.2, respectively. To do so, in Table 5.5, we give the lower and upper expected costs $\underline{E}C(T_w, w)$ and $\overline{E}C(T_w, w)$ for a range of different scenarios. For the period of time $[0, 3000]$, we assume that the penalty failure cost per unit of time is $w = 5$, and for the period of time $[0, 5000]$ and $[0, 7000]$, we assume that the penalty failure costs per unit of time are $w = 2$ and 5 for both, respectively.

Again, following the general imprecise probability theory [10] discussion in Section 5.1, we can give the following interpretation of the lower expected costs $\underline{E}C(T_w, w)$ and upper expected costs $\overline{E}C(T_w, w)$ to the producers of warranties. On the basis of the data observations, we have lower and upper expectations for the expected values. So, in terms of the warranty contract in Table 5.5, for the first policy, if a producer takes on a warranty with a random cost $C_1(T_w, w)$, then they would prefer to pay a fixed cost up to the lower expectation of 168.260 instead of having to pay the random $C_1(T_w, w)$. Similarly, they would certainly prefer to pay out the random amount $C_1(T_w, w)$ instead of any fixed amount greater than the upper expectation of 1463.384, and similar for policies $2, 3, 4$ and 5 in Table 5.5.

Example 5.3.2. Consider the Example 4.4.7 in Section 4.4 and the warranty policy B (time-dependent penalty failure cost). In this example we use the same data as in Example 4.4.7 and our lower and upper survival functions in Figures 4.7 using the interval $[\underline{\gamma}, \overline{\gamma}] = [4808.502, 5584.119]$ at significance level 0.05. We derive the lower expected cost $\underline{E}C(T_w, w)$ and upper expected cost $\overline{E}C(T_w, w)$ to the producer of the warranty if applied to one future product at normal stress level using Equations 5.3.1 and 5.3.2, respectively. To do so, in Table 5.6, we give the lower and upper expected costs $\underline{E}C(T_w, w)$ and $\overline{E}C(T_w, w)$ for a range of different scenarios. For the periods of time $[0, 3000]$, $[0, 4000]$, $[0, 5000]$ and $[0, 7000]$, we assume that the penalty failure costs per unit of time are $w = 2$ and 5 for all, respectively.

On the basis of the data observations, we have lower and upper expectations for the expected values. So, in terms of the warranty contract in Table 5.6, for the first policy, if a producer takes on a warranty with a random cost $C_1(T_w, w)$, then they would prefer to pay a fixed cost up to the lower expectation of 76.6025 instead of having to pay the random $C_1(T_w, w)$. Similarly, they would certainly prefer to pay out the random amount $C_1(T_w, w)$ instead of any fixed amount greater than the upper expectation of 158.2098, and similar for policies $2, 3, 4, 5, 6, 7$ and 8 in Table 5.6. Comparing the imprecision in the lower expected costs $\underline{E}C(T_w, W)$ and upper expected costs $\overline{E}C(T_w, W)$ to producers of the warranty in this example as shown in Table 5.6, with the imprecision in the lower expected costs $\underline{E}C(T_w, W)$ and upper expected costs $\overline{E}C(T_w, W)$ to producers of the warranty in the Example 5.3.1 as

Policy	T_w	w	$\underline{E}C(T_w, w)$	$\overline{E}C(T_w, w)$
1	3000	2	76.6025	197.9448
2	3000	5	191.5063	494.8620
3	4000	2	217.0351	500.1913
4	4000	5	542.5878	1250.4784
5	5000	2	550.5390	1191.9109
6	5000	5	1376.3475	2979.7774
7	7000	2	1950.4091	3739.4323
8	7000	5	4876.0228	9348.5809

Table 5.6: $\underline{E}C(T_w, w)$ and $\overline{E}C(T_w, w)$. Example 5.3.2.

shown in Table 5.5, we see that when we generated $n = 50$ observations at each stress level, the imprecision is smaller than with $n = 10$ observations at each stress level.

5.4 Concluding remarks

This chapter has shown a basic application of our new methods to pricing of warranties defined by imprecise probabilities based on ALT data. More specifically, it uses the NPI lower and upper survival functions at the normal stress level to support decisions on warranties. This development has been achieved by considering two steps based on ALT data, namely (*i*) an ALT model with the link between different stress levels modelled by a simple parametric link function, e.g. the power law or the Arrhenius relation, with the application of classical hypothesis tests, e.g. the likelihood ratio test and the log-rank test to obtain such intervals for the link function, and (*ii*) it uses Nonparametric Predictive Inference (NPI) at the normal stress level, combining data from tests at that level with data from higher stress levels which have been transformed to the normal stress level.

Based on the results from the ALT data and our novel imprecise predictive probabilistic statistical inference methods, both with and without assuming a Weibull failure time distribution at each stress level, we derive bounds for expected costs

of warranties. Other ALT link functions, lifetime distributions, and statistical tests with different types of data can be used straightforwardly with the presented approach, while it is also interesting to consider the use of our method for other decision-support scenarios. For example, different cost structures, modeling failure and cost for different types of business in warranty contract, safeguarding maintenance on the long term warranty costs.

Chapter 6

Concluding Remarks

The present chapter introduces a brief overview of our key results and outlines a number of tasks for future research. The main novelty of this thesis is that the imprecision results are derived based on the classical tests and the idea that, if data from a higher stress level are transformed to the normal stress level, then these transformed data and the original data from the normal stress level should not be distinguishable. This thesis presents a novel method for statistical inference based on ALT data and imprecise probability. First, we proposed a development for ALT using the Arrhenius-Weibull model under constant stress testing, using the theory of imprecise probability, where the imprecision results are derived from a likelihood ratio test. Secondly, we proposed the development of the use of a novel statistical method providing imprecise semi-parametric inference for ALT data, where the imprecision is related to the log-rank test. Both developments apply the use of the classical tests to compare the survival distribution of pairwise stress levels. In this thesis, we have considered the use of the imprecise probability, and, in particular, we have considered the Nonparametric Predictive Inference (NPI) at the normal stress level combined with the link function between the different stress levels modelled by a simple parametric link function, e.g. the Arrhenius relation, where we used the lower and upper intervals for the parameter of the link function. We further present an initial study of the use of the methods mentioned above to support manufacturers' decisions on warranties.

Chapter 3 presents a new imprecise statistical inference method for ALT data, where NPI at normal stress level is integrated with a parametric Arrhenius-Weibull model. The method includes imprecision based on the likelihood ratio test which provides robustness with regard to the model assumptions. We applied the use of the likelihood ratio test to obtain an interval for the parameter of the Arrhenius link function providing imprecision into the method. The imprecision leads to observations at increased stress levels being transformed into interval-valued observations at the normal stress level, where the width of an interval is larger for observations from higher stress levels. Note that in Chapter 3, at each stress level, we have assumed the Weibull distribution with the Arrhenius link function between different stress levels which transform the data from the increased stress levels to the normal stress level. Examples have been presented in this chapter, namely, a simulated data set and real data set from the literature. Moreover, the transformation link function has been derived if we allow different shape parameters β_i for each level i. An investigation of the performance of the proposed method has been illustrated using both the correct model assumption and the model of misspecification via simulations. In terms of the latter, our proposed methods achieved a suitable level of robustness with regard to the model assumptions.

In Chapter 4, the assumption of the Weibull distribution at each stress level we assumed in Chapter 3, is deleted. In this chapter, we consider an imprecise predictive inference method for ALT. The method is largely nonparametric, with a basic parametric function to link different stress levels. Based on the log-rank test, we provide adequate imprecision for the parameter of the assumed link function. According to the null hypothesis, we applied the use of the log-rank test to compare the survival distribution of pairwise stress levels. Therefore, using the assumed link function between different stress levels with the use of the log-rank test, we drive the interval of the parameter of the link function. The observations from the higher stress levels are then transformed into interval-valued observations at the normal stress level using this interval to achieve further robustness. We have also shown why the imprecision from a single log-rank test should not be used on all stress levels simultaneously. Therefore, the imprecision from the combined pairwise stress levels

using the minimum of lower value and the maximum of the upper value of the parameter of the link function provide substantially more imprecision in our proposed method. Generalizing this method to data including right-censored observations has been presented. Examples and simulation studies have been illustrated in this chapter using both the correct model assumption and the model of misspecification. Based on the results of the latter, our methods provide an overall good performance.

In both the approaches in Chapters 3 and 4, the argument for use of the pairwise test is the same: if the model fits poorly, a single test on all stress levels would result in less imprecision while our proposed method, combining pairwise tests tends to result in more imprecision. The most pessimistic case, which leads to the lower survival function $\underline{S}$, uses $\underline{\gamma}$ to transform the data points to the smallest values at the normal stress level. Unlike the most optimistic case, which uses $\overline{\gamma}$ to transform the data points to the largest values at the normal stress level.

Building imprecision through the use of parametric and nonparametric tests for the parameter of the link function between different stress levels, which enabled us to transform the observations at increased stress levels to interval-valued observations at the normal stress level, achieved further robustness against the necessary assumptions. Clearly the method used in Chapter 3 is preferable if one has a good knowledge about the failure times distribution per stress level and Chapter 4 if not. Chapter 4 typically leads to wider intervals $[\ \underline{\gamma}_i, \overline{\gamma}_i]$.

By comparing the use of the pairwise with the assumption of the Weibull distribution at each stress level and using the nonparametric log-rank test in Chapter 4, we confirm that using the log-rank test proposed in Chapter 4 results in more imprecision where we have a small sample size than using the assumption of the Weibull distribution at each stress level and using the parametric likelihood ratio test instead of the log-rank test. However, with large sample size, e.g. $n = 100$ the imprecision is quite similar in both approaches.

Since we began this research, we have not found ALT methods with imprecision in the literature. The classical methods as presented in the literature seem effectively to stop at parameter estimations, so no predictions and certainly not predictions explicitly at the normal stress level are considered while there are more research

projects of imprecise statistical methods for different reliability issues [19, 25, 70]. The main approach presented in this thesis, that is taking a simple model and adding imprecision to a parameter, then using corresponding transformed data for imprecision at the level of interest, has not been presented before for any inference problem.

In Chapter 5, we consider pricing basic warranty contracts based on information from ALT data and the use of our novel imprecise probabilistic statistical method, are described in Chapters 3 and 4. The new statistical methods we introduced in this thesis include imprecision based on the likelihood ratio test and log-rank test which provide robustness with regard to the model assumptions. In Chapter 5, we derive bounds for expected costs of warranties based on ALT data and using the NPI lower and upper survival functions resulting from our new statistical methods.

Chapter 5 has illustrated the first exploration of our new methods to warranties. For our novel statistical methods, one can point out how this can be implemented in real-world scenarios. For instance, we can derive bounds for expected costs of warranties based on ALT data and using the NPI lower and upper survival functions resulting from our new statistical methods in Chapters 3 and 4. We have illustrated some examples involved in inference on warranty and explained how we can calculate the expected warranty cost for a product.

An interesting future research challenges, including more investigation of further simulation studies will be of interest. For example, we will consider other distributions instead of the Weibull model as the lifetime distribution, using the theory of imprecise probability for ALT to extend and develop the use of different ways of transforming the data from different stress levels. It would be particularly interesting to work on imprecise methods for analysing ALT data into different ALT scenarios, where we aim to develop a similar approach, e.g. for the case of step-stress testing. Two such alternative approaches that may be of interest in Chapters 3 and 4 are that the likelihood ratio test and log-rank test in these approaches could be replaced by other classical tests, where classical statistical tests could be replaced by imprecise statistical methods to infer whether or not the transformed data and actual data at the normal stress level are well mixed or even the use of tests based

on imprecise probability theory could be explored. It would be of interest to consider different statistical approaches and methods to transformed the failure times from higher stress levels to the normal stress level, as we did, and also, of course, different cases of misspecification. One can consider to use similar methodology in a fully parametric ALT (with possible imprecision) as an interesting topic for future research. It will be interesting to explore if the approach presented in this thesis can also be applied with degradation models. The modelling of degradation processes does require much information about the engineering process and physical properties of the equipment, which may come from detailed measurements of the process or expert judgements.

The final NPI based inference does not have sufficient assumptions to guide choice of design. In addition, the assumptions for the model derived $[\underline{\gamma}, \overline{\gamma}]$ intervals are not strong enough for design. We do not think people will have studied design issues if an assumed model turns out to fit the real world data poorly, so design with the explicit aim of meaningful application of our robust method is a challenge. In this thesis, to illustrate the main idea of our novel method, we assumed that the failure data are available at all stress levels including the normal stress level. This may not be realistic. If there are no failure data at the normal stress level, or only right-censored observations, then we can apply our method using a higher stress level as the basis for the combinations, so transform data to that stress level. Then the combined data at that level could be transformed all together to the normal stress level. This is a topic for future research.

The aim for this research project as introduced in Chapters 1 and 2, is to use simple models with imprecision. If the end results are sufficient to answer the practical questions, then there is no need for more detailed modelling or more data. If not, then one needs to model in more detail or gather more data (or both). Note that data are often problematic in real-word contexts. For example, if it involves testing on prototypes the number of items available is likely to be limited.

Bibliography

[1] Abdel Ghaly, A.A., Aly, H.M., Salah, R.N. (2015). Different estimation methods for constant stress accelerated life test under the family of the exponentiated distributions. *Quality and Reliability Engineering International*, **32**, 1095–1108.

[2] Alam, K., Sudhir, P. (2015). Testing equality of scale parameters of two Weibull distributions in the presence of unequal shape parameters. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, **19**, 11–26.

[3] Ahmadini, A.A.H., Coolen, F.P.A. (2018). Imprecise statistical inference for accelerated life testing data: imprecision related to the likelihood ratio test. *Proceedings of the 10th IMA International Conference on Modelling in Industrial Maintenance and Reliability (MIMAR)*, Manchester, P. Scarf., S. Wu., P. Do (Eds), pp. 1–6.

[4] Ahmadini, A.A.H., Coolen, F.P.A. (2018). Imprecise statistical inference for accelerated life testing data: imprecision related to log-rank test. In: *International Conference Series on Soft Methods in Probability and Statistics*, Advances in Intelligent Systems and Computing (AISC). **823**, pp. 1–8. Springer.

[5] Ahmadini, A.A.H., Coolen, F.P.A. (2019). On the use of an imprecise statistical method for accelerated life testing data using the power-law link function. In: *Proceedings of the 2018 3rd International Conference on System Reliability and Safety (ICSRS)*, Barcelona, pp. 244-249. IEEE.

[6] Ahmadini, A.A.H., Coolen, F.P.A. Statistical inference for the Arrhenius-Weibull accelerated life testing model with imprecision based on the likelihood ratio test. In submission.

[7] Ahmadini, A.A.H., Coolen, F.P.A. Nonparametric predictive inference for warranties based on accelerated life testing data. In submission.

[8] Aboalkhair, A.M. (2012). *Nonparametric Predictive Inference for System Reliability*. PhD Thesis, Durham University, available from www.npi-statistics.com.

[9] Augustin, T., Coolen, F.P.A. (2004). Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, **124**, 251–272.

[10] Augustin, T., Coolen, F.P.A., de Cooman, G., Troffaes, M.C.M. (2014). *Introduction to Imprecise Probabilities*. Wiley, Chichester.

[11] Baker, R.M., Coolen, F.P.A. (2010). Nonparametric predictive category selection for multinomial data. *Journal of Statistical Theory and Practice*, **4**, 509–526.

[12] Baker, R.M. (2010). *Multinomial Nonparametric Predictive Inference: Selection, Classification and Subcategory Data*. PhD Thesis, Durham University, available from www.npi-statistics.com.

[13] Blischke, W.R., Murthy, D.N.P. (1992). Product warranty managementI: A taxonomy for warranty policies. *European Journal of Operational Research*, **62**, 127–148.

[14] Bischke, W.R., Murthy, D.N.P. (1994). *Warrant Cost Analysis*.CRC Press: New York.

[15] Boole, G. (1854). *An Investigation of the Laws of Thought on which are founded the Mathematical Theories of Logic and Probabilities*. Walton and Maberley, London.

[16] Bretz, F., Genz, A., A. Hothorn, L. (2001). On the numerical availability of multiple comparison procedures. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, **43**, 645-656.

[17] Caruso, H., Dasgupta, A. (1998). A fundamental overview of accelerated testing analytical models. *Journal of the IEST*, **41**, 16–20.

[18] Coolen, F.P.A. (1998). Low structure imprecise predictive inference for bayes' problem. *Statistics and Probability Letters*, **36**, 349–357.

[19] Coolen, F.P.A., Coolen-Schrijner, P., Yan, K.J. (2002). Nonparametric predictive inference in reliability. *Reliability Engineering and System Safety*, **78**, 185–193.

[20] Coolen, F.P.A. (2004). On the use of imprecise probabilities in reliability. *Quality and Reliability Engineering International*, **20**, 193–202.

[21] Coolen, F.P.A., Yan, K.J. (2004). Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference*, **126**, 25–54.

[22] Coolen, F.P.A., Augustin, T. (2005). Learning from multinomial data: a nonparametric predictive alternative to the imprecise dirichlet model. In *ISIPTA*, 125–134.

[23] Coolen, F.P.A. (2006). On nonparametric predictive inference and objective Bayesianism. *Journal of Logic, Language and Information*, **15**, 21–47.

[24] Coolen, F.P.A. (2011). Nonparametric predictive inference. *International Encyclopedia of Statistical Science*, pp. 968–970. Springer, Berlin.

[25] Coolen, F.P.A., Utkin, L.V. (2011). *Imprecise reliability.* In: Lovric, M. (ed.), *International Encyclopedia of Statistical Science*, pp. 649–650. Springer, Berlin.

[26] Coolen, F.P.A., Ahmadini, A.A.H., Maturi, T.A. Imprecise semi-parametric inference for accelerated life testing. In submission.

[27] De Finette, B. (1974). *Theory of Probability.* Wiley, Chichester.

[28] Duan, F., Wang, G. (2018). Bivariate constant-stress accelerated degradation model and inference based on the inverse Gaussian process. *Journal of Shanghai Jiaotong University (Science)*, **23**, 784–790.

[29] Elsayed, E.A., Zhang, H. (2007). Design of PH-based accelerated life testing plans under multiple-stress-type. *Reliability Engineering and System Safety*, **92**, 286–292.

[30] Fan, T., Wang, W., Balakrishnan, N. (2008). Exponential progressive step-stress life-testing with link function based on Box–Cox transformation. *Journal of Statistical Planning and Inference*, **138**, 2340–2354.

[31] Fard, N., Li, C. (2009). Optimal simple step stress accelerated life test design for reliability prediction. *Journal of Statistical Planning and Inference*, **139**(5), 1799–1802.

[32] Fan, T.H., Yu, C.H. (2013). Statistical inference on constant stress accelerated life tests under generalized gamma lifetime distributions. *Quality and Reliability Engineering International*, **29**, 631–638.

[33] Gehan, E. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, **52**, 203–224.

[34] Groebel, D.J., Mettas, A., Sun, F.B. (2001). Determination and interpretation of activation energy using accelerated-test data. *In Reliability and Maintainability Symposium. Proceedings. Annual*, 58–63. IEEE.

[35] Hampel, F. (2009). Nonadditive probabilities in statistics. *Journal of Statistical Theory and Practice*, **3**, 11–23.

[36] Han, D. (2015). Time and cost constrained optimal designs of constant-stress and step-stress accelerated life tests. *Reliability Engineering and System Safety*, **140**, 1-14.

[37] Hill, B.M. (1968). Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, **63**, 677–691.

[38] Hill, B,M. (1988). De Finetti's theorem, induction, and Bayesian nonparametric predictive inference (with discussion). *Bayesian Statistics 3*, (Editors) J.M. Bernando, M.H. DeGroot, D.V. Lindley and A. Smith, pp. 211–241. Oxford University Press.

[39] Karim M.R., Suzuki K. (2005). Analysis of warranty claim data: A literature review. *International Journal of Quality and Reliability Management*. **22**, 667–686

[40] Kim, C.M., Bai, D.S. (2002). Analyses of accelerated life test data under two failure modes. *International Journal of Reliability, Quality and Safety Engineering*, **9**, 111–125.

[41] Klein, J.P., Moeschberger, M.L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science and Business Media.

[42] Kleinbaum, D.G., Klein, M. (2005). *Survival Analysis. Statistics for Biology and Health*. Spinger-Verlag. New York

[43] Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. Wiley, New York.

[44] Liao, C.M., Tseng, S.T. (2006). Optimal design for step-stress accelerated degradation tests. *IEEE Transactions on Reliability*, **55**, 59–66.

[45] Limon, S., Yadav, O.P., Liao, H. (2017). A literature review on planning and analysis of accelerated testing for reliability assessment. *Quality and Reliability Engineering International* **33**, 2361–2383.

[46] Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports.*, **50**, 163–170.

[47] Mantel, N. (1967). Ranking procedures for arbitrarily restricted observation.*Biometrics*, **8**, 65-78.

[48] Maturi, T.A. (2010). *Nonparametric Predictive Inference for Multiple Comparisons*. PhD Thesis, Durham University, available from www.npi-statistics.com.

[49] Maturi, T.A., Coolen-Schrijner, P., Coolen, F.P.A. (2010). Nonparametric predictive inference for competing risks. *Journal of Risk and Reliability*, **224**, 11–26.

[50] Meeker, W.Q., Escobar L.A. (1998). *Statistical Methods for Reliability Data*. Wiley, New York.

[51] Meeker, W.Q., Escobar, L.A., Hong, Y. (2009). Using accelerated life tests results to predict product field reliability. *Technometrics*. **51**, 146–161.

[52] Meeker, W.Q., Sarakakis, G., Gerokostopoulos, A. (2013). More pitfalls in conducting and interpreting the results of accelerated tests. *Journal of Quality Technology*, **45**, 213–222.

[53] Méndez-González, L.C., Rodríguez-Picón, L.A., Valles-Rosales, D.J., Romero-López, R., Quezada-Carreón, A.E. (2017). Reliability analysis for electronic devices using betaWeibull distribution. *Quality and Reliability Engineering International*. **33**, 2521–2530.

[54] Mitra, A., Patankar, J.G. (1997). Market share and warranty costs for renewable warranty programs. *International Journal of Production Economics*. **50**, 155–168.

[55] Murthy, D.N.P., Djamaludin, I. (2002). New product warranty: A literature review. *International Journal of Production Economics*. **79**, 231–260.

[56] Murthy, D.N.P. (2006). Product warranty and reliability. *Annals of Operations Research*. **143**, 133–146.

[57] Murthy, D.P., Blischke, W.R. (2006). *Warranty management and product manufacture*. Springer Science and Business Media.

[58] Muhammad, N. (2016). *Predictive Inference with Copulas for Bivariate Data*. PhD Thesis, Durham University, available from www.npi-statistics.com.

[59] Nasira, E.A., Pan, R. (2015). Simulation-based Bayesian optimal ALT designs for model discrimination. *Reliability Engineering and System Safety*, **134**, 1–9.

[60] Nelson, W.B. (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*. Wiley, Hoboken, New Jersey.

[61] O'Brien, P. C., Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, 549-556.

[62] Olteanu, D., Freeman, L (2010). The evaluation of median-rank regression and maximum likelihood estimation techniques for a two-parameter Weibull distribution. *Quality Engineering*, **22**, 256–272.

[63] Pawitan, Y. (2001). *In all Likelihood: Statistical Modelling and Inference using Likelihood*. Oxford University Press.

[64] Pan, Z., Balakrishnan, N., Sun, Q. (2011). Bivariate constant-stress accelerated degradation model and inference. *Communications in Statistics - Simulation and Computation*, **40**, 247–257.

[65] Peto, R., Peto, J (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society, Series A*, **135**, 185–207.

[66] Shanahan, J. (2013). *A New Method For the Comparison Of Survival Distributions*. Thesis, University of South Carolina.

[67] Sha, N., Pan, R. (2014). Bayesian analysis for step-stress accelerated life testing using Weibull proportional hazard model. *Statistical Papers*, **55**, 715–726.

[68] Thiraviam, A.R. (2010). *Accelerated Life Testing Of Subsea Equipment Under Hydrostatic Pressure*. PhD Thesis, University of Central Florida.

[69] Thomas M.U., Rao, S.S. (1999). Warranty economic decision models: A summary and some suggested directions for future research. *Operations Research*, **47**, 807–820.

[70] Utkin, L.V., Coolen, F.P.A. (2007). Imprecise reliability: an introductory overview. In: G. Levitin (Ed.), *Computational Intelligence in Reliability Engineering, Volume 2: New Metaheuristics, Neural and Fuzzy Techniques in Reliability*, 261–306. New York: Springer.

[71] Vassilious, A., Mettas, A. (2001). Understanding accelerated life-testing analysis. In *Annual Reliability and Maintainability Symposium, Tutorial Notes*, 1–21.

[72] Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

[73] Walley, P. (1996). Measures of uncertainty in expert systems. *Artificial Intelligence*, **38**, 1–58.

[74] Weichselberger, K. (2000). The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, **24**, 149–170.

[75] Weichselberger, K. (2001). *Elementare Grundbegriffe einer Allgemeineren Wahrscheinlichkeitsrechnung: I. Intervallwahrscheinlichkeit als Umfassendes Konzept*. Physika, Heidelberg.

[76] Yang, G., Zaghati, Z. (2006). Accelerated life tests at higher usage rates: a case study. In *Reliability and Maintainability Symposium, RAMS'06. Annual*, 313–317. IEEE.

[77] Yang G. (2007). *Life Cycle Reliability Engineering*. Wiley, Hoboken, New Jersey.

[78] Yang, G. (2010). Accelerated life test plans for predicting warranty cost. *IEEE Transactions on Reliability*. **59**, 628–634.

[79] Yin, Y., Coolen, F.P.A., Coolen-Maturi, T.A. (2017). An imprecise statistical method for accelerated life testing using the power-Weibull model. *Reliability Engineering and System Safety*, **167**, 158-167.

[80] Zhang, J.P., Wang, R.T. (2009). Reliability life prediction of VFD by constant temperature stress accelerated life tests and maximum likelihood estimation. *Journal of Testing and Evaluation*, **37**, 1–5.

[81] Zhu, Y., Elsayed, E.A. (2011). Design of equivalent accelerated life testing plans under different stress applications. *Quality Technology and Quantitative Management*, **8**, 463–478.

[82] Zhao, X., Xie, M. (2017). Using accelerated life tests data to predict warranty cost under imperfect repair. *Computers and Industrial Engineering*. **107**, 223–234.