Robust Statistical Methods for Step-Stress Accelerated Life Test Data

Sultan Eid S. Albalwy

A Thesis presented for the degree of
Doctor of Philosophy

Department of Mathematical Sciences
Durham University
England

August 2025

Dedicated to

My beloved parents and children

Robust Statistical Methods for Step-Stress Accelerated Life Test Data

Sultan Eid S. Albalwy

Submitted for the degree of Doctor of Philosophy
August 2025

Abstract

Accelerated Life Testing (ALT) is commonly implemented to derive insights into experimental items' reliability. Conducting tests under typical usage conditions can be both time-consuming and expensive. In ALT, an experimental item is examined under levels of physical stress, such as temperature, voltage, or pressure, higher than the experimental item will experience under normal operation levels. Step-stress accelerated life testing (SSALT) is a special type of accelerated life test designed to gradually increase stress levels, thereby accelerating the failure process and allowing for data collection in a shorter period.

The main contribution of this thesis lies in the development of predictive and robust methods based on imprecise probability theories. These methods provide interval-based results that reflect uncertainty in both the data and the model, unlike traditional approaches that rely on exact values and strong assumptions. By allowing for imprecision, the proposed methods offer more realistic and cautious predictions, which are necessary when normal stress data is limited or the model is uncertain.

This thesis presents three robust statistical methods for the analysis of SSALT data based on theories of imprecise probability. In the first method, imprecision is incorporated based on the likelihood ratio test within the cumulative exposure model. This method consists of three steps. First, failure times occurring under different strategies at higher stress levels are transformed to the normal stress level. Second, imprecision is introduced based on the likelihood ratio test applied to the accelerating parameter under the null

hypothesis that all failure times originate from the same distribution. This imprecision allows failure times to be transformed into interval values at the normal stress level, where the transformed failure times are assumed to be indistinguishable from those observed under normal conditions. Third, Nonparametric Predictive Inference (NPI) is applied to the transformed data to provide robust predictive inference.

In the second method, imprecision is introduced based on the log-rank test applied to a parametric link function. This imprecision facilitates the transformation of data from higher stress levels into interval-valued observations at the normal stress level. The transformation is performed using the parametric link function framework, after which the transformed data are combined with Nonparametric Predictive Inference (NPI) at the normal stress level to construct lower and upper survival functions. These methods incorporate imprecision to enhance robustness with regard to model assumptions. The results demonstrate that imprecision increases for observations derived from higher stress levels, leading to more imprecise data at the normal stress level.

In the third method, a robust Bayesian framework is developed to analyze SSALT data while incorporating imprecision in prior knowledge. This method models uncertainty by considering a class of prior distributions, where the extreme bounds of this class reflect minimal prior information about the model parameters. By using these extreme bounds, lower and upper posterior predictive distributions are derived separately, enabling the prediction of future failure times at the normal stress level. This approach also facilitates the construction of lower and upper predictive survival functions, ensuring robustness in predictive inference under model uncertainty.

The performance of the proposed methods is evaluated through simulation studies. The findings indicate that imprecision increases across all methods when the assumed likelihood function or link function is misspecified. Among the three methods, the robust Bayesian approach exhibits relatively more imprecision under model misspecification, whereas the first two methods primarily show increased imprecision due to data from higher stress strategies. Additionally, the results demonstrate that as the number of observations increases, the imprecision decreases, highlighting the impact of sample size on predictive performance.

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, and it is all my own work unless referenced to the contrary in the text.

Acknowledgements

All thanks and praise is due to Allah, the Most Merciful. His guidance, mercy, and countless blessings have been my source of strength throughout this journey. Without His will, nothing would have been possible. I ask Him to accept this thesis and make it beneficial for knowledge and humanity.

I would like to express my heartfelt gratitude to my academic supervisors, Professor Frank Coolen and Dr. Jonathan Cumming, for their unwavering support, invaluable guidance, and continuous encouragement throughout my PhD journey. Their patience, calm advice, and insightful discussions have been instrumental in shaping my research. I deeply appreciate their willingness to always offer constructive feedback and their encouragement to present my work at conferences, which has significantly contributed to my academic growth and research experience. Their expertise and kindness have made this journey both enriching and inspiring, and for that, I am truly grateful.

I am deeply grateful to my beloved parents for their endless prayers, love, and guidance, which have paved the way for my success. I also extend my heartfelt appreciation to my dear brothers and sisters for their unwavering encouragement and support. I sincerely thank my wife, Hajar, for her sacrifices, and patience throughout this journey. A heartfelt appreciation goes to my precious children, Bashayer and Abdulaziz, for their patience and cheerful presence throughout this journey.

I gratefully acknowledge the financial support from Tabuk University, Saudi Arabia, and the Saudi Arabian Cultural Bureau in London, enabling me to pursue my PhD studies at Durham University. Many thanks also to Durham University for providing an exceptional research environment and academic resources that have greatly contributed to my work.

I am also grateful to my colleagues, friends, and everyone who has supported me throughout my academic journey.

Contents

Chapter 1

Introduction

Accelerated life testing (ALT) is designed to support statistical analysts to determine the expected durability of products or components and to contribute to improvements in product reliability [48, 52]. The defining feature of ALT lies in its ability to assess product reliability more rapidly than traditional testing methods. Traditional lifespan testing under standard usage conditions can be protracted and costly. ALT is commonly employed to expedite this process, providing a quicker assessment of a product's reliability. By exposing a product to stress levels that exceed typical usage conditions, ALT facilitates the collection of failure data across different conditions efficiently, both in terms of time and cost [48, 52]. Step stress accelerated life test (SSALT) is a special type of accelerated life test, which aims to increase the levels of stress gradually to accelerate or hasten an object's life in elevated manner [52].

In this thesis, we present three robust statistical approaches for the analysis of SSALT data based on the theory of imprecise probability. In the first approach, we incorporate imprecision based on the likelihood ratio test in a parametric link function. This imprecision allows to transform data from higher stress levels into interval-valued observations at the normal stress level. This transformation is achieved within the cumulative exposure model. In the second approach, we incorporate imprecision based on the log-rank test in a parametric link function only. This imprecision similarly facilitates the transformation of data from elevated stress levels into interval-valued observations at the normal condition. This transformation is achieved within a parametric link function only. The transformed data are integrated with the nonparametric predictive inferences (NPI) at normal stress levels to construct lower and upper survival functions. These approaches

incorporate imprecision to create robustness against model assumptions. Such imprecision transforms observations from elevated higher stress levels into interval-valued data at the normal stress level, with intervals widening correspondingly for observations derived from higher stress levels. In the third approach, we create imprecision based on robust Bayesian analysis. This imprecision is obtained based on class of priors identified by its extreme prior distribution. A class of priors refers to a set of prior distributions defined over a range of plausible values, reflecting uncertainty or imprecision in prior knowledge rather than specifying a single fixed prior. Based on these extreme bounds, we obtain the lower and upper posterior predictive distributions separately, which allow us to predict future failure times at the normal stress level. This approach enables the construction of lower and upper predictive survival functions at the normal stress level.

In the field of statistics and probability, uncertainty is traditionally quantified using classical probability based on Kolmogorov's axioms [5]. Extensions of these axioms can provide viable solutions in situations where information is scarce or incomplete, and where traditional probability measures may be overly constricting [5]. Such extensions include the adoption of imprecise probabilities, notably distinguished by the assignment of lower and upper probability bounds rather than precise, singular values [1, 5, 63]. The field of imprecise probability, which has flourished with research activity over the past two decades, has inspired scholars across various statistical and engineering disciplines, leading to the establishment of a dedicated society and website (The Society for Imprecise Probability: Theories and Applications) [60] as well as biennial conferences [1, 46].

In recent times, numerous methods for quantifying uncertainty and evaluating reliability have been proposed, offering several advantages over classical probability. These methods, along with their practical implementations, are at the forefront of current research within this field. Techniques such as interval probability [66, 67] and the broader theory of imprecise probabilities [65] have been instrumental in the analysis of reliability. Within this context, Coolen [14] has explored various challenges in imprecise reliability, examining an array of tools for implementing imprecise reliability in numerous practical settings.

The advancing research in imprecise probability has set the stage for the creation of novel statistical inference methods, with Nonparametric Predictive Inference (NPI) [3] being a notable approach. Various researchers have brought forward applications of NPI,

exploring new techniques applicable to diverse datasets.

The subsequent sections of this chapter are structured as follows. Section 1.1 provides the motivation behind the research conducted for this thesis. Section 1.2 outlines the organization of the thesis.

1.1 Motivation

The complexity inherent in accelerated life testing (ALT) scenarios often necessitates sophisticated statistical modelling for inference, presenting numerous challenges. The primary goal of this thesis is to propose a model that is both simple and robust, capable of quantifying imprecision in a way that is broadly applicable in practical situations. This model will produce interval probabilities rather than precise ones. In case these interval-based probabilities prove insufficient in addressing practical application challenges, then adjustments to the model, additional data collection, or expert input may be employed to refine the methodology. This research is initiated by building on the work of Yin et al. [69], who proposed an imprecise statistical method for ALT data.

Yin et al. [69] utilized the power-Weibull model in their methodology and introduced imprecision through intervals around parameter estimates, thereby transforming observations from various higher stress levels into interval data at the normal stress level [69]. However, they did not prove the extent of imprecision in the parameter beyond simulations. Expanding upon Yin et al. [69]'s work, Ahamadini [2] integrated classical statistical tests across different stress levels to derive intervals for the link function's parameters.

Ahamadini [2] determined an interval for the parameter inherent to the link function, assumed for each level of stress, by applying established hypothesis testing methods to pairs of stress levels to measure imprecision, and hypothesising that when data from a higher stress level is transformed to a normal stress level, the resulting data should theoretically be indistinguishable from the original data collected at the normal level. It should be noted that the intervals of the transformed data tend to be greater when it is transformed from higher stress levels.

In this thesis, we extended these robust approaches with a new transformation method for the analysis of step-stress accelerated life testing data where the cumulative exposure model is implemented. This transformation method is more appropriate to be imple-

mented for SSALT data as this method will mix the transformed data with the original data at the normal stress to achieve the assumption that the transformed times of failures should not be distinguishable from failure times occurring at the same level.

Moreover, we develop a new robust Bayesian approach for the analysis of SSALT data considered by imprecise probabilities. This approach incorporates imprecision to model imprecise prior knowledge on the parameters involved in the assumed model where it allows to obtain lower and upper bounds for the posterior distribution. This modelling of imprecise prior knowledge is defined based on a class of prior distributions, while this class is identified by its extremes under complete lack of knowledge about the parameters. Based on these extreme bounds, we obtain the lower and upper posterior predictive distribution in a separate manner, which allow us to predict future failure times at the normal stress level. This approach enables us to construct lower and upper predictive survival functions at the normal stress level.

The performance of the proposed robust approaches is evaluated by extensive simulation studies. The analysis examines how a future observation aligns with and is comparable to the existing observations at the normal stress level. These simulations consider both cases where model assumptions hold true and instances of model misspecification. Data were generated to assess the robustness and predictive performance of the proposed methods. The simulations were also used to compare the three approaches across various scenarios, illustrating their ability to perform effectively under a range of conditions.

The predictive and robust methods developed in this thesis offer practical value for real-world engineering reliability analysis. In many engineering systems, data from normal stress levels may be limited, incomplete, or unavailable. Traditional methods often rely on strong model assumptions and precise priors, which may not hold in practice. By incorporating imprecise probability theory, the proposed methods provide interval-based predictions that reflect both model and data uncertainty. This robustness enhances the reliability of predictions under complex and uncertain testing conditions, offering engineers more cautious and realistic assessments for decision-making and system design.

1.2 Outline of the thesis

This thesis introduces significant advancements in applying imprecise statistical predictive inference to step-stress accelerated life testing data. Our innovative methodology enhances the robustness of predictive inferences by employing statistical tests across pairs of stress levels, which facilitates the derivation of intervals for the parameter values of the link function. Furthermore, a robust Bayesian approach is developed considered by imprecise probabilities to obtain lower and upper bounds for the posterior distribution. Consequently, this thesis supplements advanced approaches by implementing frequentist statistical tests and robust Bayesian analysis to determine the amount of imprecision.

This thesis is organised as follows. Chapter 2 presents foundational concepts from the literature that informs the approaches developed in this thesis. It provides an overview of reliability testing, accelerated life testing (ALT) and step-stress accelerated life testing (SSALT). It discusses the Cumulative exposure model (CEM) and acceleration models related to ALT. It reviews basic statistical methods implemented in this thesis, namely the maximum likelihood estimation, Bayesian estimation, Markov Chain Monte Carlo, and the Metropolis-Hastings algorithm. Also, it reviews basic statistical tests, specifically the likelihood ratio test and the log-rank test. Lastly, it presents an overview of nonparametric predictive inference (NPI).

Chapter 3 presents a new robust statistical approach based on the likelihood ratio test for SSALT data. This chapter was presented at the 12^{th} workshop on principle and methods for statistical inference with interval probability, (WPMSIIP, 9-12 September 2109, Durham University). Also, this chapter was presented at the 11^{th} IMA International Conference on Modelling in Industrial Maintenance and Reliability (MIMAR), in June 2021.

Chapter 4 introduces a new robust statistical approach based on the log-rank test for SSALT data. This chapter was presented in the Reliability Meeting at the Department of Mathematical Sciences (9 June 2022, Durham University). It was also presented at the Second International Workshop on Reliability Engineering and Computational Intelligence (14-15 November 2022, Delft, the Netherlands) and at the 15th International Conference of the ERCIM WG on Computational and Methodological Statistics and 16th International Conference on Computational and Financial Econometrics, (17-19 Decem-

ber 2022, King's College London, UK).

Chapters 3 and 4 were presented at the Institute for Statistics and Mathematics, RWTH Aachen University, Germany, on 29 March 2023.

Chapter 5 presents a new robust Bayesian statistical approach based on a class of priors for SSALT data. This chapter was presented at the 12th IMA International Conference on Modelling in Industrial Maintenance and Reliability (MIMAR), Nottingham, UK, from 4-6 July 2023. Additionally, it was presented at the 3rd International Workshop on Reliability Engineering and Computational Intelligence (RECI), held on 6-8 November 2024 in Žilina, Slovakia.

Chapter 2

Preliminaries

In this chapter, we present foundational concepts from the literature that inform the approaches developed in this thesis. Section 2.1 provides an overview of reliability testing. Section 2.2 introduces the main concepts of accelerated life testing (ALT). Section 2.3 introduces step-stress accelerated life testing (SSALT). Section 2.4 discusses the Cumulative exposure model (CEM). Section 2.5 discusses acceleration models related to ALT. In Section 2.6, we review basic statistical methods implemented in this thesis, namely the maximum likelihood estimation, Bayesian estimation, Markov Chain Monte Carlo, and the Metropolis-Hastings algorithm. Also, we review basic statistical tests, specifically the likelihood ratio test and the log-rank test. Lastly, Section 2.7 presents an overview of nonparametric predictive inference (NPI).

2.1 Reliability testing

Reliability testing provides comprehensive applications to engineers in terms of obtaining perfect and optimal designs, systems, services, or products. These applications process data in order to gain useful information to interpret different situations. It measures the success or failure of engineering outcomes, and consistency or life span of products under specific conditions. It also assesses the life cycle of products as it examines each factor affecting the life span. As a result, such calculations can detect reasons of failure, enhance designs, or implement innovative modifications. This leads to improved quality and affordability of products. Testing life time data is essential in order to make major decisions about the reliability of products, and statistical methodologies have developed

a variety of engineering approaches to deal with these decisions.

The reliability of an object is its ability to perform necessary and required processes in a successful manner. It is defined as the probability that the object is functioning successfully for a specified period of time [47, 52, 61]. Suppose T represents a positive random variable, which describes the lifetime of the object by a life time distribution. The cumulative distribution function (CDF), which is expressed by $F(t)$, is the probability that the object fails before or at time t. $F(t)$ is given by

$$F(t) = P[T \leq t], \tag{2.1}$$

where $F(0) = 0$. Therefore, the reliability of the object is the probability that it will survive beyond time t, which is given by the survival function

$$S(t) = P[T > t] = 1 - F(t). \tag{2.2}$$

In addition, suppose $f(t)$ represents the probability density function (PDF) of $F(t)$. Suppose $\lambda(t)$ represents the hazard rate function, which measures the rate at which failing items in the complete population occur at time t [61]. The hazard function $\lambda(t)$, also known as the conditional failure rate, is given by

$$\lambda(t) = \frac{f(t)}{S(t)}. \tag{2.3}$$

Moreover, there are many statistical distributions, which express the variation in the lifetime of objects. The Weibull distribution is implemented to demonstrate any approaches and methodologies will be presented in this thesis as discussed in later sections. The Weibull distribution is one of the most widely used models in reliability engineering due to its flexibility in characterizing various types of failure behavior. Its shape parameter, often denoted by β, allows the distribution to model decreasing, constant, or increasing failure rates, which correspond to early-life failures, random failures, and wear-out failures, respectively. This adaptability makes the Weibull model suitable for diverse components and systems across industries **??**.

In traditional reliability analysis, it is required to collect sufficient data about the lifetime of experimental units to estimate the lifetime parameters under normal conditions. Failure times of new highly reliable products will occur after a long period of time which requires time and cost to collect such data [61]. These difficulties have led to an engineering approach, which is called accelerated life testing (ALT).

2.2 Accelerated life testing

Accelerated life testing (ALT) is defined as the process of estimating failure times of products exposed to abnormal conditions. It provides testing methods to obtain sufficient information of a product's life span, but in a much shorter time period [52]. The majority of highly reliable products will survive for a long period of time under normal conditions. This long period of time poses significant difficulties in quantifying and characterising the life time of products. The purpose of ALT is to accelerate failure occurrences to shorten the life time of products or to hasten and degrade product performance, in order to produce the desired information quickly. This acceleration applies stress to various factors influencing the product's reliability, allowing failure times to be observed under different stress levels within a shorter time frame. As a result, ALT generates lifetime data in a compressed time period, which can then be generalized to estimate real-world performance or assess product reliability under normal operating conditions.

Accelerated life testing (ALT) is applied in testing various materials, products, and degradation mechanisms [52]. It is widely used across multiple industries to assess the reliability and longevity of materials such as metals, polymers, ceramics, adhesives, and coatings, as well as complex products ranging from simple components to advanced electronic systems [52]. Additionally, ALT is implemented to evaluate degradation mechanisms and assess product performance under different stress factors. Overall, the extensive application of ALT highlights its crucial role in ensuring reliability and durability in engineering and manufacturing.

Accelerated life tests and performance degradation tests are conducted for several reasons and purposes. Firstly, a design of a product can be improved when ALT data assess the design failure modes and suggest appropriate modes to be taken into consideration. These data also provide comparisons and assessments between different designs, components, suppliers and operating conditions. It investigates any variables which may affect the reliability of a product. It can therefore adopt suggestions to optimize product reliability and limitations. Moreover, it measures reliability of a product in order to permit to release a design to a manufacturer or a product to a customer. These processes, which are called "Burn-In", aim to eliminate early failures in shorter time [52]. Furthermore, it can establish quality control of a product to assure its reliability and take corrective

action when it is needed. It forms service policy to decide when to inspect the product for replacement or maintenance. Finally, these measurements of reliability estimate warranty and service expenses as a result of estimating failure rates, mean time to failure and degradation rates [52].

Accelerated test data can be classified into two types: Measures of Performance Data and Life Data which are the interesting characteristics of products to be observed [52].

ALT aims to examine how products' performance degrades with age. Such testing involves an experiment to measure the relationship between performance, age, and stress. These experiments include higher stresses to observe products degradation of performance over time. Such observations provide measures of performance data to fit a model to estimate the relationship between performance, age, and stress levels under normal use [52].

Life data consist of observations which measure the exact life (failure age) of each sample unit of a product. Life data is classified into two types, complete and incomplete [52]. Most life data are incomplete in that the exact failure times of some units are unknown which therefore provides incomplete information of the failure times.

There are many examples of incomplete life data. These include:

- **Censored data**, which means that some units of a product, which are expected to fail, are still running while the data is analysed. Such censored data arise when:
  - Some units are removed from the test or service before failure,
  - Some units are still running at the time of the data analysis,
  - Some units are removed from test or service because failure has occurred from an extraneous cause. Precisely, these data are recognised as being **censored on the right**.
- Similarly, **censored data on the left** consider failure times of some units known to be before a certain time.
- In case all survival units have a specific running time and all failure times are before data analysis, such data are called **singly censored on the right**. Singly censored data arise when units are started together at a test condition and the data are analysed before all units fail. Such data are singly time censored when the

censoring time is fixed, thus such period of fixed time reflect the number of failures to be random.

- Another example of incomplete life data is **multiply censored data**, which is censored data on the right and have different running times intermixed with the failure times. Multiply censored data may arise when units are being tested at different times. Thus, such units have different running times when the data are recorded [52].

Competing failure modes is another type of data which occurs when sample units fail from different causes. Such causes can be categorised into failure modes which are described as a mix of competing failure modes. These data on a specific failure mode record the failure times of units which are failed by that mode. Such data will be dealt with as multiply censored. Additionally, quantal-response data investigate whether the failure time of a unit is before or after a specific time. Each observation is either censored on the right or on the left. Such life data arise when each unit is inspected once to test whether it has failed or not. There is also interval data which aims to inspect failure time of a unit within an interval, which means each unit is inspected periodically when a unit failed in an interval between inspections. It divides the time as intervals or groups. Such data consist of right and left censored observations. Finally, there is mixture data which contains a mixture of all of the above types of data. To conclude, these types of data have similar purposes whether it is complete or incomplete data as it derives similar analyses. These purposes will be generally estimating model parameters, fitting a distribution of the product life, or predicting future observations [52].

There are several types of acceleration of tests, which are high usage rate, censoring, degradation, specimen design and stress loading. The selection of the type of test depends on the purpose of the test. For example, a high usage rate is a simple way to accelerate a product's life, when it is run intensively at a higher usage rate. There are two different ways to run a product to compress the time of tests. These are either to run a product faster or to reduce the off time. Censoring tests are used to test a product's life before all sample units run to failure in order to shorten test time or to obtain prior data of failure time. Another type of acceleration reset is a degradation test [52]. This type measures a product's performance before failure times occur, which aims to observe performance degradation over time rather than the product's life. A specimen test looks

at characteristics of the whole specimen design that can effect and accelerate failure of a product. Such characteristics are size and shape. Finally, there are stress loading tests. They are methods of running a product at higher levels than normal. These stresses, which aim to shorten product life or to degrade product performance faster, are temperature, voltage, mechanical load, thermal cycling, humidity, and vibration, etc. There are various types of stress loading tests including constant, cyclic, step, progressive, and random stress loading. These tests can be conducted by a single stress variable or multiple stress variables [52].

In this thesis, we focus solely on step-stress loading for the development of our novel robust statistical approaches. It's important to note that the design of Accelerated Life Testing (ALT) tests is not within the scope of this thesis; rather, it is assumed that the data are provided for the analysis.

2.3 Step stress accelerated life testing (SSALT)

Step stress accelerated life test (SSALT) is a special type of accelerated life tests, which aims to increase the levels of stress gradually to accelerate or hasten an object's life in elevated manner [52]. This manner of acceleration ensures more failures of experimental items in a much shorter period of time when it is compared to constant stress accelerated life tests (CSALT). Figure 2.1 illustrates the concept of SSALT.

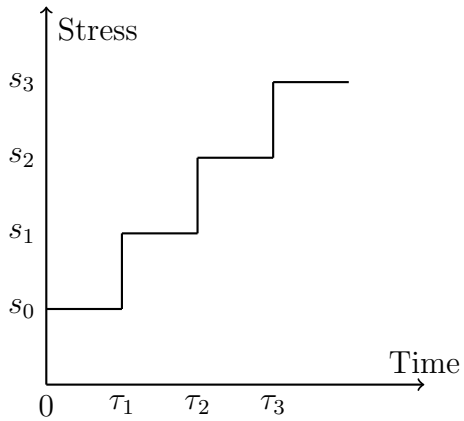


Figure 2.1: Step stress accelerated life testing

Suppose that SSALT experiment is conducted with k number of stress levels $s_0, s_1, s_2, ..., s_k$, which are predetermined to test n experimental items. These k stress levels are

increased in elevated manner at prespecified points of time $\tau_1 < \tau_2 < ... < \tau_k$. In a basic practice of SSALT, n experimental items are placed on the experiment, when these items are initially exposed to the stress level s_0 (normal use or initial stress level) for a period of time $t \in (0, \tau_1)$ then the stress level is increased from s_0 to s_1 at the point of time τ_1 for a period of time $t \in (\tau_1, \tau_2)$. Likewise, the level of stress is increased from s_2 to s_3 at the point of time τ_2 for the period of time $t \in (\tau_2, \tau_3)$. In general, the level of stress is increased from s_{k-1} to s_k at the point of time τ_{k-1} for the period of time $t \in (\tau_{k-1}, \tau_k)$.

This basic SSALT experiment starts at time $\tau_0 = 0$ and ends at time $\tau_k = \infty$ or when all the experimental items placed in the experiment fail or are censored. Also, this experiment considers that the increasing times of the levels of stress are fixed. The times of item failures are recorded sequentially in an ordered manner. Suppose that the number of failing items n_i, which occur before the time point τ_i for $i = 1, 2, \ldots, k$, results in a complete data set as

$$t_{1:n} < \cdots < t_{n_1:n} < \tau_1 < t_{n_1+1:n} < \cdots < t_{n_2:n} < \tau_2 < \cdots < \tau_{k-1} < t_{n_k-1+1:n} < \cdots < t_{n:n} < \tau_k.$$

Here, t represents the failure time. In this notation, $t_{i:n}$ denotes the i-th order statistic from a sample of size n, that is, the i-th smallest observed failure time. Specifically, $t_{n_1:n}$ represents the failure time of the n_1-th failed item, occurring before the stress level changes at time τ_1.

To model and analyse data of step stress experiments, it is required to relate the cumulative distribution functions of the lifetime of the experimental items at each stress level to the normal stress level under normal operating conditions. There are several models, which have been developed in the literature to express this concept. The most commonly implemented and proposed models are cumulative exposure models, tampered random variable models, tampered failure rate models, and cumulative risk models [41].

The cumulative exposure model, which was introduced by Sediakin [57] and then extended by Nelson [51, 52]. This model assumes that the residual life of the experimental items depends only on the cumulative exposure the items have been exposed to, regardless how these exposures were accumulated through stresses.

The tampered random variable model, which was introduced by Goel [28, 29] and then extended by DeGroot and Goel [21], assumes that the consequence of increasing the level of stress to a higher level at a specific point of time is equivalent to multiply the residual life of the experimental items by an unknown positive constant.

The tampered failure rate model, which was proposed by Bhattacharyya and Soejoeti [10], adopts the concept of Cox's proportional hazards model [19]. It assumes that the consequence of increasing the stress level is a multiplicative effect on the failure rate function at the level of stress s_0.

The cumulative risk model, which was introduced by Van Dorp, J.R. et al. [64] and then extended by Kannan et al. [39], assumes that the previous step stress models have a lag or latency period of time to count the effect of increasing the level of stress. It is also assumed that the related hazard functions of these models are not continuous as the increase of the level of stress is instantaneous. This discontinuity is considered unrealistic for many applications, however the cumulative risk model solves this problem.

In this thesis, the cumulative exposure model is implemented to develop our novel robust statistical approaches for step stress ALT data. The cumulative exposure model (CEM) is preferred in this work due to its key assumption that residual life depends only on total accumulated exposure, regardless of the path of stress application. This makes it suitable for step-stress experiments where stress changes occur in stages.

2.4 Cumulative exposure model (CEM)

The cumulative exposure model (CEM) is considered as the most commonly used model to analyse the step stress experiments. This model assumes that the residual life of the experimental items depends only on the cumulative exposure the items have been exposed to, regardless how these exposures were accumulated through stresses. Furthermore, the step stress model contains a piecewise continuous function of constant stress levels at different consecutive periods of time, and there is a life distribution for each of these stress levels. The failing of items occurs according to the cumulative distribution at the current level of stress, however it is accumulated to the previous fraction failed. The failure of items occurs according to the cumulative distribution at the current level of stress, while accounting for the fraction of items that have already failed at the previous stress levels. Specifically, the remaining survival fraction at each step is adjusted by the cumulative proportion of failures observed in the earlier stress periods. This way of accumulation contains a shift parameter h between the overall CDF of the step stress and the CDF's of constant stress levels the experimental items are being tested at. This

shift parameter h is shown horizontally from the right to the left, where it is illustrated in Figure 2.2.

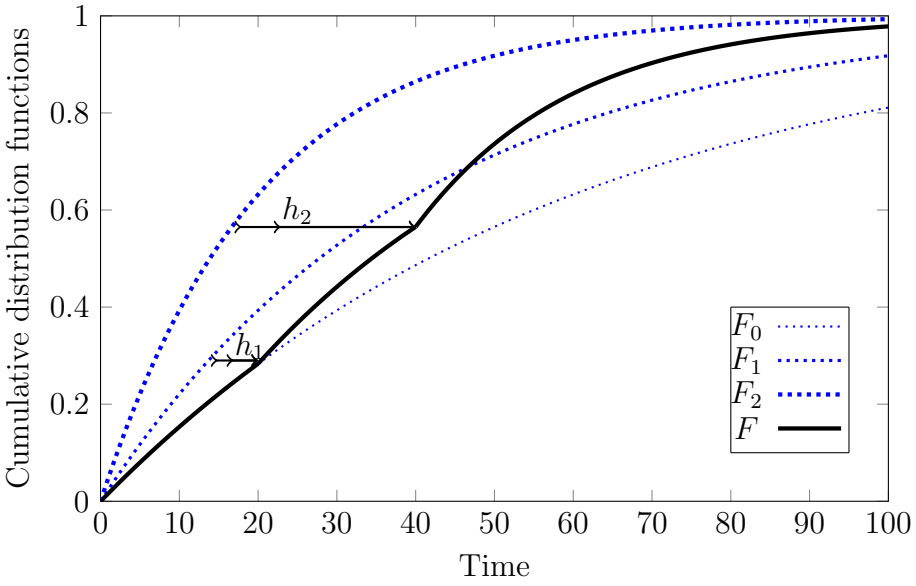


Figure 2.2: Cumulative exposure model (CEM) in a step-stress accelerated life test with three stress levels. The blue dotted curves show the CDFs under constant stresses s_0, s_1, and s_2. The solid black curve represents the overall CDF, adjusted at each stress level by shift parameters h_1 and h_2 to ensure continuity.

To clarify, suppose a step stress experiment is conducted with three stress levels s_0, s_1, s_2 for periods of time τ_1, τ_2, τ_3, which have lifetime distributions F_0, F_1, F_2 respectively. These life distributions should belong to a specific distribution family. These life distributions should belong to a specific distribution family, which will be specified in the subsequent sections. The experiment begins with n identical items, and all items are exposed to an initial stress level s_0 with lifetimes expressed by the CDF of $F_0(t)$. Failures, which occur at stress s_0 will be collected and the remaining experimental units will be stressed under stress level s_1 with lifetimes expressed by the CDF of $F_1(t)$. Failures, which occur at stress s_1, will be accumulated to the previous fraction failed. Similarly, when the stress is increased from s_1 to s_2 with lifetimes expressed by the CDF of $F_2(t)$.

Let $F(t)$ represent the overall CDF of the lifetime of experimental units in SSALT experiment under the cumulative exposure model. In the first step, experimental units

fail according the CDF of the initial stress level s_0, which is

$$F(t) = F_0(t) \qquad \text{for} \qquad \tau_0 \leq t \leq \tau_1. \tag{2.4}$$

In the second step, the consequence of increasing the level of stress from s_0 to s_1 at the point of time τ_1 is equivalent to changing the CDF from $F_1(t)$ to $F_1(t - h_1)$ at the stress level s_1. Note that $F_1(t)$ represents the lifetime of experimental units when subjected to a constant stress level. The cumulative exposure model contains a function of constant stress and the shift parameter h_1 maintains the assumption that the failures at stress level s_1 are accumulated to the previous fraction failed at the stress level s_0. Therefore, experimental units fail according the CDF of the stress level s_1, which is

$$F(t) = F_1(t - h_1) \qquad \text{for} \qquad \tau_1 \leq t \leq \tau_2, \tag{2.5}$$

where h_1 is a shift parameter for the stress level s_1, and is obtained by solving the following equation

$$F_1(\tau_1 - h_1) = F_0(\tau_1). \tag{2.6}$$

In the third step, the consequence of increasing the level of stress from s_1 to s_2 at the point of time τ_2 is equivalent to change the CDF from $F_2(t)$ to $F_2(t - h_2)$ at the stress level s_2. Therefore, experimental units fail according the CDF of the stress level s_2, which is

$$F(t) = F_2(t - h_2) \qquad \text{for} \qquad \tau_2 \leq t \leq \tau_3, \tag{2.7}$$

where h_2 is a shift parameter for the stress level s_2, which is obtained by solving the following equation

$$F_2(\tau_2 - h_2) = F_1(\tau_2 - h_1). \tag{2.8}$$

Since $F(t)$ represents a cumulative distribution function (CDF), it is assumed to be continuous at τ_1. This means there is no sudden jump or gap in the function when transitioning from one stress level to another. The continuity ensures that the failures observed at the new stress level are accumulated smoothly without disrupting the overall failure process. Because this holds at each transition point, the structure of the cumulative exposure model remains consistent from the initial time until the end of the testing period. Therefore, the general expression of $F(t)$ follows this pattern across all stress levels.

As $F(t)$ is a continuous function at τ_1, these processes continue in this manner from $\tau_0 = 0$ to $\tau_k = \infty$. Therefore, the general expression of $F(t)$ is given as:

$$F(t) = F_i(t - h_i), \quad \text{for} \quad \tau_i \leq t \leq \tau_{i+1}, \quad i = 0, 1, ..., k-1, \tag{2.9}$$

where $h_0 = 0$ and h_i for $i = 0, 1, ..., k-1$, is obtained by solving the following equation

$$F_i(\tau_i - h_i) = F_{i-1}(\tau_i - h_{i-1}). \tag{2.10}$$

The corresponding probability density function (PDF) is given by

$$f(t) = f_i(t - h_i), \quad \text{for} \quad \tau_i \leq t \leq \tau_{i+1}, \quad i = 0, 1, ..., k-1. \tag{2.11}$$

Furthermore, the shifting parameter $h_i = (\tau_i - \tau_i^*)$, where it starts to shift at the point of time τ_i^* to τ_i. Alternatively, $F(t)$ can be written as

$$F(t) = F_i(t - \tau_i + \tau_i^*), \quad \text{for} \quad \tau_i \leq t \leq \tau_{i+1}, \quad i = 0, 1, ..., k-1. \tag{2.12}$$

Also, the corresponding probability density function (PDF) can be written as

$$f(t) = f_i(t - \tau_i + \tau_i^*), \quad \text{for} \quad \tau_i \leq t \leq \tau_{i+1}, \quad i = 0, 1, ..., k-1. \tag{2.13}$$

In this thesis, the cumulative exposure model is implemented using the Weibull distribution, which is widely used for reliability analysis [52]. The Weibull distribution is extensively utilized to quantify reliability and is thus integral in evaluating the lifespan of experimental units. It describes the failure characteristics of electronic components and the structural integrity of their constituent materials. Notably, the Weibull distribution is commonly employed in accelerated life testing scenarios.

The Weibull distribution includes a scale parameter $\theta > 0$ and a shape parameter $\beta > 0$. The probability density function (PDF) is given by

$$f_i(t) = \left(\frac{\beta}{\theta_i^\beta}\right) t^{\beta-1} \exp\left[-\left(\frac{t}{\theta_i}\right)^\beta\right], \quad \text{for} \quad t \geq 0, \tag{2.14}$$

and the cumulative distribution function (CDF)is given by

$$F_i(t) = 1 - \exp\left[-\left(\frac{t}{\theta_i}\right)^\beta\right], \quad \text{for} \quad t \geq 0, \tag{2.15}$$

where θ_i is the scale parameter at the stress level s_i for $i = 1, 2, .., k-1$, and β is the shape parameter which is assumed to be constant at each stress level. Consider $k-1$ steps of

SSALT experiment based of the assumption of the cumulative exposure model while the lifetime of experimental units follow Weibull distribution. In the first step, experimental units fail according the CDF at the initial stress level s_0, which is $F_0(t)$ for $t \in (\tau_0, \tau_1)$. In the second step, the CDF at s_1 is $F_1(t-h_1)$ for $t \in (\tau_1, \tau_2)$,where h_1 is the solution of $F_1(\tau_1 - h_1) = F_0(\tau_1)$. Consequently, h_1 is obtained by solving the following equation

$$1 - \exp\left[-\left(\frac{\tau_1 - h_1}{\theta_1}\right)^{\beta}\right] = 1 - \exp\left[-\left(\frac{\tau_1}{\theta_0}\right)^{\beta}\right]. \tag{2.16}$$

Therefore, $h_1 = (1 - \theta_1/\theta_0)\tau_1$, and the CDF at s_1 becomes

$$F_1(t) = F_1(t-h_1) = 1 - \exp\left[-\left(\frac{t-\tau_1}{\theta_1} + \frac{\tau_1}{\theta_0}\right)^{\beta}\right], \quad \text{for} \quad \tau_1 \le t \le \tau_2 \tag{2.17}$$

Similarly, in the third step, the CDF at s_2 is $F_2(t-h_2)$ for $t \in (\tau_2, \tau_3)$,where h_2 is the solution of $F_2(\tau_2 - h_2) = F_1(\tau_2 - h_1)$. Consequently, h_2 is obtained by solving the following equation

$$1 - \exp\left[-\left(\frac{\tau_2 - h_2}{\theta_2}\right)^{\beta}\right] = 1 - \exp\left[-\left(\frac{\tau_2 - \tau_1}{\theta_1} + \frac{\tau_1}{\theta_0}\right)^{\beta}\right]. \tag{2.18}$$

Therefore, $h_2 = (1 - \theta_2/\theta_2)\tau_2 + (1/\theta_1 - 1/\theta_0)\theta_2\tau_1$, and the CDF of s_2 becomes

$$F_2(t) = F_2(t-h_2) = 1 - \exp\left[-\left(\frac{t-\tau_2}{\theta_2} + \frac{\tau_2-\tau_1}{\theta_1} + \frac{\tau_1}{\theta_0}\right)^{\beta}\right], \text{ for } \tau_2 \le t \le \tau_3. \tag{2.19}$$

In general, the shifting parameter h_i is given by $h_i = (\tau_i - \tau_i^*)$, where τ_i^* follows a recursive structure:

$$\tau_1^* = \left[\frac{\theta_1}{\theta_0}\tau_1\right], \tag{2.20}$$

$$\tau_2^* = \left[\frac{\theta_2}{\theta_1}\left(\tau_2 - \tau_1 + \frac{\theta_1}{\theta_0}\tau_1\right)\right], \tag{2.21}$$

$$\tau_3^* = \left[\frac{\theta_3}{\theta_2}\left(\tau_3 - \tau_2 + \frac{\theta_2}{\theta_1}\left(\tau_2 - \tau_1 + \frac{\theta_1}{\theta_0}\tau_1\right)\right)\right]. \tag{2.22}$$

Thus, τ_i^* is given recursively as:

$$\tau_i^* = \left[\frac{\theta_i}{\theta_{i-1}}\left(\tau_i - \tau_{i-1} + \tau_{i-1}^*\right)\right], \quad \text{for} \quad \tau_i \le t \le \tau_{i+1}, \tag{2.23}$$

with the base case:

$$\tau_0^* = 0. \tag{2.24}$$

In addition, the cumulative distribution function (CDF) of an SSALT experiment with $k-1$ stress transitions, while the lifetime of experimental units follows a Weibull distribution, is given by:

$$F(t) = \begin{cases} 1-\exp[-(\frac{t}{\theta_0})^{\beta}] & \text{if } \tau_0 \le t \le \tau_1 \\ 1-\exp[-(\frac{t-\tau_1}{\theta_1}+\frac{\tau_1}{\theta_0})^{\beta}] & \text{if } \tau_1 \le t \le \tau_2 \\ 1-\exp[-(\frac{t-\tau_2}{\theta_2}+\frac{\tau_2-\tau_1}{\theta_1}+\frac{\tau_1}{\theta_0})^{\beta}] & \text{if } \tau_2 \le t \le \tau_3 \\ \quad . & \quad . \quad . \\ \quad . & \quad . \quad . \\ 1-\exp[-(\frac{t-\tau_{k-1}}{\theta_{k-1}}+\frac{\tau_{k-1}-\tau_{k-2}}{\theta_{k-2}}+.....+\frac{\tau_1}{\theta_0})^{\beta}] & \text{if } \tau_{k-1} \le t \le \tau_k \end{cases} \tag{2.25}$$

Alternatively, the CDF at $k-1$ SSALT experiment of Weibull distribution, can be written as:

$$F_i(t) = 1-\exp\left[-\left(\frac{t-\tau_i}{\theta_i}+\sum_{j=1}^{i}\frac{\tau_j-\tau_{j-1}}{\theta_{j-1}}\right)^{\beta}\right], \quad \text{for } \tau_i \le t \le \tau_{i+1}. \tag{2.26}$$

For $i=0$, the summation term is not defined. Thus, the base case is given as:

$$F_0(t) = 1-\exp\left[-\left(\frac{t}{\theta_0}\right)^{\beta}\right], \quad \text{for } \tau_0 \le t \le \tau_1. \tag{2.27}$$

The probability density function (PDF) is given by

$$f_i(t) = \left(\frac{\beta}{\theta_i^{\beta}}\right)(t-(\tau_i-\tau_i^*))^{\beta-1}\exp\left[-\left(\frac{t-\tau_i}{\theta_i}+\sum_{j=1}^{i}\frac{\tau_j-\tau_{j-1}}{\theta_{j-1}}\right)^{\beta}\right], \text{ for } \tau_i \le t \le \tau_{i+1}. \tag{2.28}$$

For $i=0$, the summation term is not defined. Thus, the base case for the probability density function (PDF) is given as:

$$f_0(t) = \frac{\beta}{\theta_0^{\beta}}t^{\beta-1}\exp\left[-\left(\frac{t}{\theta_0}\right)^{\beta}\right], \quad \text{for } \tau_0 \le t \le \tau_1. \tag{2.29}$$

Besides these statistical models, there are acceleration models that aim to model the failure mechanism in terms of physical and chemical factors to extrapolate failure times at the normal stress level.

2.5 Acceleration models

There are specific characteristics of the lifetime data when stresses are applied. Life time tends to be longer at low stress than at high stress, and the variation in life time is greater at low stress than at high stress levels. At each stress level, there is a statistical distribution of life that requires it to be linked to other stress levels in order to mimic the normal life time of experimental items at the normal stress level [52]. Establishing the life-stress relationship in accelerated life testing (ALT) typically requires the utilization of either a physical or empirical acceleration model [62].

Physical acceleration models, which are based on physical or chemical theory, explain the causes of the failure process over a range of stress levels and enable extrapolation to failure conditions at normal use. The relationship between the accelerating variable and the failure mechanism tends to be highly complex, where basic models, in many scenarios, are not sufficient to define the failure causes and process. The Eyring model [52], which is one illustration of physical acceleration models, was built based on the theory of quantum mechanics [62].

Empirical acceleration models are derived based on a mathematical framework to fit the observed data. These models aim to describe the failure mechanism, when it is difficult to develop a physical acceleration model. It aims to link the scale and shape parameters at higher stress levels in terms of acceleration variables to predict failure times at different stress levels. Examples of empirical acceleration models are the linear and log-linear functions and the inverse power link function [42, 52].

In this thesis, we aim to develop robust statistical approaches to analyse SSALT data based on the theory of imprecise probability. These robust statistical approaches aim to deal with this complicity of the modelling of SSALT data and provide easy-to-use methods based on a few modelling assumptions. We will discuss this further in the later chapters. We will implement a combination of the Weibull cumulative exposure model and the Arrhenius and Eyring acceleration models to illustrate our robust statistical approaches to analyse SSALT data.

The Arrhenius life-temperature relationship is usually implemented to model product life as a function of temperature [52]. Applications include electrical insulations and dielectrics, solid state and semiconductor devices battery cells, lubricants and greases, plas-

tics, and incandescent lamp filaments. Based on the Arrhenius Law for simple chemical-reaction rates, the relationship is implemented to express many products that fail as a result of degradation because of chemical reactions or metal diffusion [12, 30, 52]. The relationship is adequate over some range of temperature. The Arrhenius rate law equation is given by

$$\mathrm{R} = A \cdot \exp\left(-\frac{E}{K_B \cdot K}\right), \tag{2.30}$$

where R represents the reaction rate that explains the speed of chemical reaction and the metal diffusion, A is a constant that describes the product failure mechanism and test conditions, E is the activation energy of the reaction, K_B is Boltzmann's constant as 8.617×10^{-5} electron volts per $^{\circ}C$, and K is the absolute Kelvin temperature that equals to Centigrade temperature plus 273 degrees [52].

The relationship explains a simple view of failure in a product as a result of chemical reaction or diffusion. The product is expected to fail at a critical amount of chemical reaction or diffusion. This critical amount is given by (Critical amount) = (rate) $\times$ (time to failure). This leads to time to failure θ being considered as inversely proportional to the rate [23, 52]. Therefore, the Arrhenius life-stress relationship is given by

$$\theta = A \cdot \exp\left(\frac{E}{K_B \cdot K}\right). \tag{2.31}$$

The log-linear function of the Arrhenius relationship for the normal stress level is given by

$$\ln(\theta_0) = \gamma_0 + \frac{\gamma}{K_0}, \tag{2.32}$$

where θ_0 is the Weibull scale parameter at the normal stress level, $\gamma_0 = \ln(A)$, $\gamma = \frac{E}{K_B}$ and $\gamma > 0$ [52]. Also, the log-linear function of the Arrhenius relationship for the higher stress level is given by

$$\ln(\theta_i) = \gamma_0 + \frac{\gamma}{K_i}. \tag{2.33}$$

Thus, the Arrhenius link function between the higher stress level k_i and the normal stress level k_0 of temperature (Kelvin) is given by [52]

$$\theta_i = \theta_0 \exp\left(\frac{\gamma}{k_i} - \frac{\gamma}{k_0}\right). \tag{2.34}$$

The Eyring model is considered an alternative approach to the Arrhenius model as it also models the accelerating variable in terms of temperature [23, 52]. This model establishes the correlation between the mean time to failure θ and the acceleration variable temperature K. It is important to note that the Weibull scale parameter θ does not generally equal the mean time to failure (MTTF), except when the shape parameter $\beta = 1$, corresponding to the exponential distribution. For $\beta \neq 1$, the MTTF is given by

$$\text{MTTF} = \theta\,\Gamma\left(1 + \frac{1}{\beta}\right),$$

where $\Gamma(\cdot)$ is the gamma function [52].

The Eyring model describes the correlation between the mean time to failure and the accelerating variable temperature K as given by

$$\theta = \frac{A}{K}\exp\left(\frac{\lambda}{K}\right), \tag{2.35}$$

where $\theta = \frac{E_A}{k_B}$. A and $\theta > 0$ are constants that describe the characteristics of the combined failure mechanisms and experimental conditions, k_B represents the Boltzmann's constant, (8.6171×10^{-5} electron-volts per $^{\circ}C$), and K represents the absolute temperature (Kelvin). The scale parameter of experimental items at the normal stress level is

$$\theta = \frac{A}{K_0}\exp\left(\frac{\lambda}{K_0}\right), \tag{2.36}$$

and the scale parameters at the higher stress levels are

$$\theta_i = \frac{A}{K_i}\exp\left(\frac{\lambda}{K_i}\right). \tag{2.37}$$

The Eyring link function for the higher stress level k_i and the normal stress level k_0 of temperature (Kelvin) is given by

$$\alpha_i = \alpha_0 \times (K_0/K_i) \times \exp\left[(\lambda/K_i - \lambda/K_0)\right], \tag{2.38}$$

where $\theta > 0$ is the acceleration parameter of the Eyring link function model. Applications include testing capacitors, electro-migration failure, and solid rupturing [52].

It should be emphasised that the Arrhenius and Eyring link functions predict equal extrapolations in numerous applications, however, the Arrhenius link function is not appropriate and can be misleading for some applications. The Arrhenius model may not be suitable in certain scenarios where multiple competing chemical reactions with

different activation energies are involved. In such cases, the Arrhenius relationship may not accurately describe the overall reaction rate [23].

In such cases, the Eyring model is often preferred because it accounts for additional stress factors beyond temperature, such as humidity, voltage, or mechanical load. This makes the Eyring relationship more suitable for reliability studies where temperature is not the primary cause of stress failure. An example of this is breakdown in semiconductor devices, where factors such as electric fields, voltage stress, and mechanical stress, in addition to temperature, contribute to the activation energy leading to device failure [23, 52].

2.6 Basic statistical methods

In this section, basic statistical methods will be discussed as these methods will be implemented to construct our novel robust approaches to analyse SSALT data. In general, the basic statistical methods to analyse survival data are parameters estimation and hypothesis testing.

2.6.1 Parameter estimation

In classical statistics, maximum likelihood estimation (MLE) is a method used to estimate the parameters of a model. It seeks the parameter values that maximize the likelihood function, which is the product of the probability densities at each data point [31]. Let $f(t_i;\theta)$ denote the probability density function for the observed failure times t_i with $i = 1,\ldots,n$, and $S(c_j;\theta)$ represent the survival function for censored times c_j, with $j = 1,\ldots,u$. The likelihood function, given a specific failure time distribution with parameter $\underline{\theta}$, is expressed as:

$$L(\underline{\theta};\mathbf{t}) = \prod_{i=1}^{n} f(t_i;\underline{\theta}) \prod_{j=1}^{u} S(c_j;\underline{\theta}), \tag{2.39}$$

In Chapter 5, the focus is primarily on Bayesian methods for (SSALT) data. Bayesian inference is based on the principle that prior knowledge or subjective beliefs about parameters can be updated with new data to form posterior beliefs. This Bayesian approach begins with establishing prior beliefs about an event, which are then systematically revised into posterior beliefs as additional data is collected. Unlike frequentist inference,

where model parameters are considered fixed entities, Bayesian inference assigns probabilities to these parameters, allowing for a dynamic updating of beliefs based on new evidence [31, 59]. As detailed by Hamada, M.S. et al. [31], Bayesian inference melds prior knowledge about parameters - typically represented through a probability density function - with experimental data to refine these beliefs. This combinations is expressed through Bayes' theorem, which calculates the posterior distribution, $\pi(\underline{\theta} \mid \underline{x})$, as follows:

$$\pi(\underline{\theta} \mid \underline{x}) = \frac{\pi(\underline{\theta})L(\underline{x} \mid \underline{\theta})}{\int \pi(\underline{\theta})L(\underline{x} \mid \underline{\theta})d\underline{\theta}},$$

where $\underline{\theta}$ symbolizes the vector of parameters, $\underline{x}$ represents the observed data, $\pi(\underline{\theta})$ is the prior distribution,$L(\underline{x} \mid \underline{\theta})$ denotes the likelihood function of a model $f(\underline{x} \mid \underline{\theta})$, and the denominator serves as a normalizing constant [31, 59].

The unnormalized posterior distribution is expressed as follows:

$$\pi(\underline{\theta} \mid \underline{x}) \propto \pi(\underline{\theta})L(\underline{x} \mid \underline{\theta}),$$

$$\text{posterior} \propto \text{prior} \times \text{likelihood}.$$

The choice of prior distributions for model parameters stands as a widely discussed issue within Bayesian statistics. Subjective priors, informed by either expert judgment or historical data, allow for the incorporation of specific insights into the statistical model. For instance, engineers with expertise in the lifespan of products, drawing on either the physical basis of failure or accumulated failure data, can integrate this knowledge into the analysis through the selection of tailored subjective priors [48]. Conversely, the use of objective priors, often described in literature as non-informative, flat, or vague, presents an alternative approach. Such priors, including the uniform, reference, Jeffreys, and maximal data information priors, are explored in detail by Berger, J.O. et al. [7], who provide insights and discussions on the use of objective priors. Additionally, Yang and Berger [68] offer a comprehensive list of non-informative priors suitable for a range of distributions and models, with a preference for reference and Jeffreys priors. It is critical to note the importance of verifying the appropriateness of the posterior distribution when utilizing objective priors, as many of them are known to be improper [31, 59].

The posterior mean of a parameter, given observed data, can be formally expressed as an equation derived from the posterior distribution. This equation is given by:

$$E[\underline{\theta}|\underline{x}] = \int \underline{\theta}\, \pi(\underline{\theta}|\underline{x})\, d\underline{\theta}, \tag{2.40}$$

where $E[\underline{\theta}|\underline{x}]$ denotes the expected value (mean) of the parameter vector $\underline{\theta}$ given the data $\underline{x}$, and $\pi(\underline{\theta}|\underline{x})$ represents the posterior distribution of $\underline{\theta}$ conditioned on $\underline{x}$. The integral is taken over all possible values of $\underline{\theta}$, effectively weighting each possible parameter value by its probability under the posterior distribution. This equation shows the Bayesian parameter estimation by quantifying the expected value of a parameter based on the posterior distribution, which incorporates both prior knowledge and the likelihood of the observed data [31, 59].

The posterior predictive distribution is a fundamental concept in Bayesian statistics that allows for predictions of future observations based on the data already observed. It is denoted by $p(x^*|\underline{x})$, where x^* represents a new, future data point, and $\underline{x}$ denotes the observed data set. The posterior predictive distribution is given by the integral:

$$p(x^*|\underline{x}) = \int p(x^*|\underline{\theta})\pi(\underline{\theta}|\underline{x})d\underline{\theta}. \tag{2.41}$$

This equation relies on the assumption that the future data point x^* is conditionally independent of the observed data $\underline{x}$, given the model parameters $\underline{\theta}$. In other words, x^* and $\underline{x}$ are independent when conditioned on $\underline{\theta}$, which can be expressed as $x^* \perp \underline{x}|\underline{\theta}$.

Under this assumption, $p(x^*|\underline{\theta})$ represents the likelihood of observing the new data point x^* given the parameters $\underline{\theta}$, while $\pi(\underline{\theta}|\underline{x})$ is the posterior distribution of the parameters given the observed data $\underline{x}$. The integral aggregates over all possible parameter values $\underline{\theta}$, weighing the likelihood of the new observation by the updated beliefs about the parameters after observing $\underline{x}$ [31].

In many advanced Bayesian models, encountering an intractable posterior distribution is a common challenge, where simplifying the distribution to a closed form is not feasible. This situation complicates the computation of marginal posterior distributions and the identification of the normalizing constant, thus posing significant obstacles to Bayesian inference. To address these issues, Markov Chain Monte Carlo (MCMC) methods have been introduced. These computational strategies enable sampling from complex posterior distributions that are difficult to manage directly [31, 59].

A Markov chain is characterized as a sequence of random variables $X^{(1)}, X^{(2)}, \ldots$, with the notable property that the state at time $t+1$, $X^{(t+1)}$, is influenced solely by the state at time t, $X^{(t)}$, and is independent of any preceding states. This relationship is

formalized as:

$$P\left(X^{(t+1)} = x \mid X^{(1)} = x^{(1)}, X^{(2)} = x^{(2)}, \ldots, X^{(t)} = x^{(t)}\right) = P\left(X^{(t+1)} = x \mid X^{(t)} = x^{(t)}\right). \tag{2.42}$$

Starting from an arbitrary initial point $X^{(0)}$, it is possible to construct a Markov chain $\{X^{(t)}\}$ that, by following a transition distribution possessing a stationary distribution f, ensures the chain's alignment with f [56]. In simple terms, a transition distribution defines how the chain moves from one state to another, while the stationary distribution f represents the target distribution we want to simulate from. An ergodic Markov chain is one that, after many steps, forgets its starting point and explores the entire space such that the samples eventually represent draws from f. This approach facilitates the computation of the expected value of a function $g(x)$, expressed as:

$$\mathrm{E}[g(x)] = \int g(x) f(x) dx, \tag{2.43}$$

which can be estimated through the Monte Carlo average obtained from MCMC samples:

$$\mathrm{E}[g(x)] \approx \frac{1}{M} \sum_{m=1}^{M} g\left(X^{(m)}\right), \tag{2.44}$$

given a sufficiently large number M [31, 59]. In this thesis, the Metropolis-Hastings sampler is implemented as an MCMC method in the development of our robust approach in Chapter 5.

The Metropolis-Hastings algorithm stands as a widely utilized MCMC sampler, respected for its straightforward approach and versatility. It finds its roots in the foundational works of Metropolis et al. [49] and Hastings [33], which introduced the concept of initiating parameter value sampling from approximate distributions and then refining these samples to more closely represent the desired posterior distribution. Over time, these approximate distributions are adjusted to align with the posterior distribution, and through a process of sequential sampling - where each new sample depends only on the preceding one - a Markov chain is formed [31, 59].

The core principle of MCMC algorithms is the simulation of parameter values directly from the posterior distribution for inferential purposes, using these simulated values as the basis for analysis. Suppose $\pi(\underline{\theta} \mid \underline{x})$ is the posterior distribution, also known as the target density, encompassing a parameter vector θ with dimensions $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$.

We denote the t^{th} sample in the sequence as $\underline{\theta}^{(t)}$. As t increases, the samples' distribution approaches the target density [31, 59].

The steps for executing the Metropolis-Hastings algorithm are as follows:

1. Begin with an initial value, $\underline{\theta}^{(0)}$, selected such that $\pi\left(\underline{\theta}^{(0)} \mid \underline{x}\right) > 0$. This initial value may be chosen based on prior knowledge or a preliminary estimate [27]. In this thesis, the maximum likelihood estimation is implemented as the preliminary estimate for each parameter.

2. A proposed point $\underline{\theta}^*$ is generated from the current state $\underline{\theta}^{(t-1)}$ using a proposal distribution, $q\left(\underline{\theta}^* \mid \underline{\theta}^{(t-1)}\right)$. This distribution is a conditional density at time step t and is responsible for suggesting the next potential move in the parameter space. The probability of proposing a move from $\underline{\theta}^*$ back to $\underline{\theta}^{(t-1)}$ is denoted as $q\left(\underline{\theta}^{(t-1)} \mid \underline{\theta}^*\right)$. The chosen proposal distribution must be one from which sampling is straightforward, and it should fulfil specific requirements to ensure that the resulting Markov chain is both irreducible and aperiodic, thus facilitating convergence to the target distribution.

3. Calculate the acceptance probability, denoted by ρ, for the candidate to become the new sample:

$$\rho = \min\left(1, \frac{\pi\left(\underline{\theta}^* \mid \underline{x}\right) q\left(\underline{\theta}^{(t-1)} \mid \underline{\theta}^*\right)}{\pi\left(\underline{\theta}^{(t-1)} \mid \underline{x}\right) q\left(\underline{\theta}^* \mid \underline{\theta}^{(t-1)}\right)}\right).$$

   The ratio $\frac{\pi(\underline{\theta}^*|\underline{x})}{\pi\left(\underline{\theta}^{(t-1)}|\underline{x}\right)}$ encourages transitions to parameter values with higher posterior probabilities, and the ratio $\frac{q\left(\underline{\theta}^{(t-1)}|\underline{\theta}^*\right)}{q\left(\underline{\theta}^*|\underline{\theta}^{(t-1)}\right)}$ indicates the proposal distribution's preference for certain parameter values.

4. Draw a value u from a uniform distribution $U(0,1)$. Accept the candidate θ^* as the new sample if $u \leq \rho$; otherwise, retain the previous sample $\theta^{(t-1)}$. The transition is thus defined by:

$$\underline{\theta}^{(t)} = \begin{cases} \underline{\theta}^* & \text{with probability } \rho \\ \underline{\theta}^{(t-1)} & \text{with probability } 1-\rho. \end{cases}$$

5. Proceed to the next step in the sequence by updating t such that t becomes $t+1$, and then revisit Step 2. Continue this process for a predefined large number

of iterations specified by N_{MH}, which will construct the Markov chain sequence $\underline{\theta}^{(1)}, \underline{\theta}^{(2)}, \ldots, \underline{\theta}^{(N_{MH})}$ [31, 59].

Furthermore, to numerically obtain the posterior predictive distribution and generate future data points, the following steps are performed:

1. Posterior samples are obtained in a manner that the Metropolis-Hastings algorithm is utilized to draw a large number of samples from the posterior distribution of the parameters, $\underline{\theta}$, given the observed data, $\underline{x}$. These samples are drawn according to the distribution $\pi(\underline{\theta} \mid \underline{x})$.

2. Future data are simulated in a manner that for each sample $\underline{\theta}^{(i)}$ drawn from the posterior, simulate a data point x^* from the likelihood $p(x^*|\underline{\theta}^{(i)})$. This step is repeated for each posterior sample, thus generating a set of future observations x^* based on the parameter values specified by $\underline{\theta}^{(i)}$.

3. Posterior predictive distribution is constructed by aggregating the simulated data points x^* to form the posterior predictive distribution $p(x^* \mid \underline{x})$. This distribution is a composite of the likelihoods of all possible future observations, weighted by the posterior probabilities of the parameters that generated them.

4. Finally, the simulated future observations are compiled into a predictive dataset. This dataset is a numerical representation of the posterior predictive distribution and reflects the uncertainty inherent in the predictions based on the observed data $\underline{x}$ and the prior information summarised in $\pi(\underline{\theta})$ [31].

2.6.2 Hypothesis testing

Statistical hypothesis tests are commonly implemented to compare survival functions or probability distributions between two or more independent groups of survival data, including those with right-censored data. This section provides an introduction of the likelihood ratio test and the log-rank test, which are implemented in Chapter 3 and 4, respectively. This implementation aims to create an interval of the acceleration parameter in the acceleration link function. The statistical tests are executed in a pairwise manner that two or more data sets of different SSALT settings are coming from the same underlying distribution under the null hypothesis, where the interval represents the range of

parameter values for which the null hypothesis cannot be rejected. In fact, this concept can be applied with other statistical tests and certainly these tests may show largely similar results. Hypothesis testing plays a critical role in statistical analysis, especially when comparing whether two or more independent groups have the same probability distribution. Such comparisons often require the use of parametric tests. One widely recognized test for assessing the equality of survival distributions between groups of data is the likelihood ratio test. This test is particularly practical for analysing two sets of failure data that may have right-censored data, such as from Step-stress accelerated life testing (SSALT), within predefined parametric models.

The probability density function of a particular assumed statistical model defines failure times occurred at a strategy of predefined and fixed experimental settings of SSALT data where the test aims therefore to maximise the parameters of the model based on the idea of the hypothesis testing. The null hypothesis (H_0) assumes that the two data sets are originated from the same underlying statistical model. To perform the likelihood ratio test, the difference between the log likelihood under the null and alternative hypothesis are required to be computed. This comparison is based on the ratio of the likelihood functions values to reject the null hypothesis (H_0) when the ratio is sufficiently small [53].

Suppose that l_0 represents the maximized log likelihood function on the parameter space under the null hypothesis (H_0) and l_1 represents the maximized log likelihood function over the entire parameter space under the alternative hypothesis (H_1). The likelihood ratio test statistic is given by $LR = 2(l_1 - l_0)$. This statistic follows a χ^2 distribution where the degree of freedom is the difference in the number of the parameters between the two models [53]. Detailed discussions on the likelihood ratio test can be found in numerous introductory statistics books (see e.g. [40, 55]).

The Log-rank test is a nonparametric statistical test used to compare survival distributions between two or more groups $m \geq 2$ [44, 54]. This test is also called the Mantel-Cox test [26, 45, 58]. It is widely used in survival analysis to test the hypothesis that there is no difference in survival data between the groups being compared.

The log-rank test evaluates whether there is a significant difference between the survival distributions of two groups by comparing the observed numbers of failures that occurred against the expected numbers of failures [58]. Based on the number of individ-

uals at risk within each group, the log-rank test calculates its statistics by utilizing the observed and expected event counts for each group and assessing the difference in hazard rates across the two groups over the entire study period [37].

Suppose that $m_{1,j}$ represent the number of failures occur at t_j in group one, $m_{2,j}$ represent the number of failures occur at t_j in group two. Let $n_{i,j}$ represent the count of units at risk immediately before the time point $t_{(j)}$ $(j = 1, 2, \ldots, k)$. The log-rank test calculates the expected number of failures for a given group $i \in \{1, 2\}$ as

$$e_{ij} = \left(\frac{n_{i,j}}{n_{1,j} + n_{2,j}} \right) \times (m_{1,j} + m_{1,j}) \,. \tag{2.45}$$

Let's consider two groups, each containing $n_i (i = 1, 2, \ldots, m)$ individuals. Each individual in the groups can be either a failure event or is subject to right-censoring at a certain time point. Let $0 < t_{(1)} < t_{(2)} < \ldots < t_{(k)} < \infty$ represent the distinct times at which failures occur, where $t_0 = 0$ and $t_{k+1} = \infty$. For simplicity in this explanation, it is assumed that there are no tied observations within the combined dataset of both groups.

The log-rank test statistic ζ is calculated as follows:

$$\zeta = \frac{(O_2 - E_2)^2}{\mathrm{Var}\,(O_2 - E_2)}, \tag{2.46}$$

where $O_2 - E_2 = \sum_{j=1}^{k} (m_{2,j} - e_{2,j})$. Hence, the variance of $\mathrm{Var}\,(O_i - E_i)$ is determined as follows:

$$\mathrm{Var}\,(O_i - E_i) = \sum_{j=1}^{k} \frac{n_{1,j} n_{2,j} \,(m_{1,j} + m_{2,j})\,(n_{1,j} + n_{2,j} - m_{1,j} - m_{2,j})}{(n_{1,j} + n_{2,j})^2 \,(n_{1,j} + n_{2,j} - 1)}. \tag{2.47}$$

The test statistic ζ approximately follows a chi-squared distribution with degrees of freedom corresponding to the number of groups minus one [40]. This approximation becomes accurate in large samples due to the asymptotic properties of the log-rank test. However, for small sample sizes or when assumptions are violated, the distribution of ζ may deviate from the chi-squared distribution, which can affect the validity of p-values and subsequent hypothesis testing. In such cases, alternative test statistics or exact methods may be considered [26]. In such cases, alternative test statistics or exact methods may be considered, such as Gehan's generalized Wilcoxon test, which applies different weightings to account for censored data and may be appropriate [26].

2.7 Nonparametric predictive inference (NPI)

Within the framework of imprecise probability, classical probability is extended so that uncertainties associated with events are quantified not by singular values but rather by ranges of numbers [14]. For instance, under classical probability theory, a distinct probability measure $P(A) \in [0, 1]$ is attributed to an event A, with P being a probability measure defined in accordance with Kolmogorov's foundational axioms. The idea of imprecise probability as a means to quantify uncertainty was originally introduced by Boole [11] in 1854, and it has been developed extensively since in the field of statistics following Hampel's contributions [32].

In the last few years, diverse methodologies for assessing uncertainty have been developed, including the concept of interval probability theory by Walley and Weichselberger [65, 66], which suggests that probabilities should be expressed as a range with a lower and upper bound, denoted as $[\underline{P}(A), \overline{P}(A)]$ respectively, where $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$. This contrasts with classical probability theory, which in the case of complete absence of data regarding an event, assigns it the full range from $\underline{P}(A) = 0$ to $\overline{P}(A) = 1$. In this context, $\underline{P}(A)$ signifies the lower probability of event A, and $\overline{P}(A)$ signifies the upper probability for event A. The imprecision associated with event A is measured by $\triangle(A) = \overline{P}(A) - \underline{P}(A)$. Coolen [13] introduced the concept of lower and upper predictive probabilities for analysing survival data. This approach is a component of the broader statistical framework known as Nonparametric Predictive Inference (NPI), which is briefly discussed in this section.

Nonparametric Predictive Inference (NPI) is reviewed in this section narrowly guided by [1, 46, 50]. NPI is a statistical methodology which construct lower and upper predictive survival functions for a future event within the framework of imprecise probability [4, 16]. This methodology is based on Hill's assumption [35], which provides direct conditional probabilities for a future random quantity which rely on the values of associated random quantities [3, 15, 16].

The concept suggests that the rank of a future observation relative to values already observed is equally likely to be of each possible value $1, \ldots, n+1$ [46, 50]. Assume that $X_1, X_2, \ldots, X_n, X_{n+1}$ symbolise exchangeable and continuous real-valued possible random quanities, subsequently the ranked observed values of $X_1, X_2, \ldots, X_n$ can be symbolised

by $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$. Let $x_{(0)} = 0$ and $x_{(n+1)} = \infty$. The assumption $A_{(n)}$ is

$$P\left(X_{n+1} \in \left(x_{(j-1)}, x_{(j)}\right)\right) = \frac{1}{n+1}, \tag{2.48}$$

for all $j = 1, 2, \ldots, n+1$. For simplicity, this explanation assumes there are no tied observations. However, if ties are present, this scenario can be managed by assuming the tied values are separated by a small difference that converges towards zero [34, 50]. Inferences derived from $A_{(n)}$ are characterised as nonparametric and predictive [50]. This approach is appropriate when minimal knowledge is known about the random quantity under consideration beyond the n observed values, or when there is a deliberate choice to not utilize any additional information [50]. While the $A_{(n)}$ assumption does not lead to precise probabilities for many numerous events of interest [50], it does establish the best possible bounds for probabilities, as per the 'fundamental theorem of probability' [20]. These bounds are expressed as lower and upper probabilities within the framework of imprecise probability theory [3, 4].

These probabilities are suggesting a scope of interpretations [4]. As an illustration, $\underline{P}(A)$ may be conceptualized to the maximum selling price for a bet concerning event A, it would be such that a payment of 1 is made if event A takes place, otherwise, a payment of 0 is made if A does not occur. This concept extends basically to be interpreted as the highest lower bound for the probability of A based on the given assumptions. In a similar manner, $\overline{P}(A)$ can be interpreted as the lowest selling price for the a bet concerning event A, or the lowest upper bound based on the given assumptions. It holds that $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$, with the property of conjugacy $\underline{P}(A) = 1 - \overline{P}\left(A^c\right)$ where, A^c is the complimentary event of A [3, 4].

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

$$\underline{S}_{X_{n+1}}(t) = \frac{n-j}{n+1}, \text{ for } t \in \left(x_j, x_{j+1}\right), j = 0, \ldots, n. \tag{2.49}$$

$$\overline{S}_{X_{n+1}}(t) = \frac{n+1-j}{n+1}, \text{ for } t \in \left(x_j, x_{j+1}\right), j = 0, \ldots, n. \tag{2.50}$$

In the fields of reliability and survival analysis, failure events are often the primary focus [17, 69]. Nonetheless, the data in such studies frequently involve right-censoring, indicating that for some observations, there is only evidence that the failure events have

not occurred by a certain time [69]. The $A_{(n)}$ assumption is not applicable for such right-censored data, it requires data where all the failure events are fully observed [69]. To extend its applicability, Coolen and Yan [18] introduced an extension of $A_{(n)}$, defined as rc- $A_{(n)}$, proposed to adapt right-censored data within the framework of nonparametric predictive inference [69]. Further, rc- $A_{(n)}$ includes the supplementary assumption that, at the time of censoring, the remaining time until a failure occurs of a right-censored unit is exchangeable among the remaining time to failure of any other unit that is yet to fail or be censored [46, 69].

Consider that n units under observation, out of which u units fail at distinct times $x_{(1)} < x_{(2)} < \ldots < x_{(u)}$. There are also $n-u$ that are right-censored units at times $c_{(1)} < c_{(2)} < \ldots < c_{(n-u)}$ and we assign $x_{(0)} = -\infty$ and $x_{(u+1)} = \infty$. Furthermore, consider that there are s_i right-censored units within each interval (x_i, x_{i+1}), signified by $c_1^i < c_2^i < \ldots < c_{s_i}^i$, consequently $\sum_{n=0}^{u} s_i = n-u$. Let d_j^i be the number of event at the failure or censoring time, where $d_0^i = x_i$ and $d_j^i = c_j^i$ for $i = 1, 2, \ldots, u$ and $j = 1, 2, \ldots, s_i$. Lastly, denote $\tilde{n}_{c_u}$ and $\tilde{n}_{d_j^i}$ as the number of units at risk immediately before the censoring time c_u and d_j^i, respectively, in line with the definition that $\tilde{n}_0 = n+1$ [46, 69].

Coolen and Yan [18] introduced the lower and upper survival functions within the Nonparametric Predictive Inference (NPI) framework for lifetime failure data. In the presence of right-censored data, the NPI lower survival function is denoted as $\underline{S}_{X_{n+1}}(t)$ and and the upper survival function is denoted as $\overline{S}_{X_{n+1}}(t)$ [46]. Correspondingly, consistent with the previously established notation, let $d_{s_i+1}^i = d_0^{i+1} = x_{i+1}$ for $i = 1, 2, \ldots, u-1$. Thus, for $t \in \left[d_j^i, d_{j+1}^i\right)$ where $i = 1, 2, \ldots, u$ and $j = 1, 2, \ldots, s_i$, and for $t \in [x_i, x_{i+1})$ where $i = 1, 2, \ldots, u$. The NPI lower and upper survival functions are given as

$$\underline{S}_{X_{n+1}}(t) = \frac{1}{n+1}\tilde{n}_{d_j^i} \prod_{r:c_r \leq d_j^i} \left(\frac{\tilde{n}_{c_r}+1}{\tilde{n}_{c_r}}\right). \tag{2.51}$$

$$\overline{S}_{X_{n+1}}(t) = \frac{1}{n+1}\tilde{n}_{x_i} \prod_{r:c_r \leq x_i} \left(\frac{\tilde{n}_{c_r}+1}{\tilde{n}_{c_r}}\right). \tag{2.52}$$

Equations (2.49) - (2.52) are crucial in the domain of imprecise probability theory because both lower and upper bounds for the survival probability are provided, allowing for uncertainty in survival data to be accounted for. This approach enables more cautious inferences when the available data is incomplete or uncertain, making the analysis more robust and less reliant on precise assumptions [20]. The difference between the upper and

lower survival functions, termed imprecision, mirrors the level of information contained within the data. This imprecision exists and is non-zero due to the restrained inferential assumptions applied, and is indicative of the amount of information in the data as previously discussed. It is noteworthy that in the context of right-censored data, where we introduced NPI's lower and upper survival functions, the lower survival function decreases at each recorded observation, whereas the upper survival function decreases solely at the times of actual failures. This elegantly demonstrates a distinct informal interpretation of lower and upper probabilities: the lower probability for event A represents the information in support of event A, the upper probability (actually, $1 - \overline{P}(A)$) represents the information against event A - hereafter also in support of the complementary event, which consistent with the principle of conjugacy property. The occurrence of a failure naturally counts as evidence against the prospect of survival, thereby decreasing the support for survival in both the lower and upper survival functions. Meanwhile, a right-censored event reduces the assurance of survival beyond that point (since fewer units are verified to survive), yet it does not enhance the evidence against survival, given that no failure has occurred.

Chapter 3

Robust statistical inference using the likelihood ratio test

3.1 Introduction

The construction of statistical models for step-stress accelerated life testing (SSALT) often involves complex processes and introduces challenges in modelling and deriving statistical inferences. This complexity arises primarily from the assumptions made in three main areas. Firstly, the statistical models must account for the effects of changing stress levels over time. Secondly, physical and chemical models, such as the Eyring and Arrhenius models, add complexity by describing how stress influences failure times. Thirdly, the assumptions about the lifetime distribution, such as using Weibull or exponential distributions, must accurately represent the variation in failure times. The combination of these assumptions makes SSALT models challenging to implement and validate.

Existing research frequently focuses on developing more complex models [22, 24, 42]. Despite their theoretical attraction, we argue that for practical implementation, SSALT is best analysed by statistical methods that are readily applicable, straightforward to implement, and equipped with essential robustness.

The aim of this chapter is to develop an easy-to-use method that incorporates essential robustness while relying on fewer model assumptions. We propose a predictive approach that allows for the extrapolation of failure times at the normal stress level. This approach enables making inferences at the normal stress level, providing practical applicability and reliable results without relying on overly complex models.

In this chapter, we introduce a novel approach that integrates a likelihood ratio test-based method for the analysis of SSALT data within the framework of imprecise probabilities. This approach combines nonparametric predictive inference (NPI) at the normal stress level with a parametric Arrhenius cumulative exposure model for higher stress levels. The primary innovation of this method lies in the application of the likelihood ratio test to measure the level of imprecision between two sets of SSALT data. Additionally, we introduce a new data transformation technique that allows failure times from higher stress levels to be extrapolated to the normal stress level. Together, these elements contribute to a novel and comprehensive framework for SSALT analysis.

This chapter is organised as follows. Section 3.2 briefly presents the cumulative exposure model and its characteristic for the analysis in the proposed method. Section 3.3 presents the proposed method which is based on the likelihood ratio test. Section 3.4 presents illustrative examples of the proposed method. Section 3.5 presents simulations studies to evaluate the performance of the proposed method. Section 3.6 briefly presents the cumulative exposure model with different shape parameters at each strategy of experimental settings. Section 3.7 presents an illustrative example of the proposed method with different shape parameters. Section 3.8 presents simulation studies to evaluate the performance of the proposed method with different shape parameters. Section 3.9 presents concluding remarks on this chapter.

3.2 The model

In this section, we recall the Weibull cumulative exposure model, which was explained in Section 2.4. This model is implemented to analyse step stress ALT data, while it is assumed that failure times follow the Weibull distribution and the Weibull scale parameters at different stress levels are linked by the Arrhenius function. First, the probability distribution function of the model under $s-1$ steps of SSALT strategies is given by:

$$f_i(t) = \left(\frac{\beta}{\theta_i^{\beta}}\right)\left(t-(\tau_i-\tau_i^*)\right)^{\beta-1}\exp\left[-\left(\frac{t-\tau_i}{\theta_i}+\sum_{j=1}^{i}\frac{\tau_j-\tau_{j-1}}{\theta_{j-1}}\right)^{\beta}\right], \tag{3.1}$$

for $\tau_i \le t \le \tau_{i+1}$, where τ_i^* is given by

$$\tau_i^* = \left[\frac{\theta_i}{\theta_{i-1}}\left(\tau_i-\tau_{i-1}+\tau_{i-1}^*\right)\right] \text{ for } \tau_i \le t \le \tau_{i+1}, \tag{3.2}$$

where $\tau_0 = 0$ and $\tau_s = \infty$, for $i = 0, 1, .., s-1$, i represents the stress level, with $i = 0$ denoting the normal stress level, the shape parameter $\beta > 0$ is assumed to be constant and the same at each stress level, and the scale parameters θ_i, for stress levels i, are linked by the Arrhenius function. This function expresses the relationship between the lifetime and the applied stresses in terms of temperature to link the Weibull scale parameters at different stress levels, which is given by

$$\theta_i = \theta_0 \exp\left(\frac{\gamma}{k_i} - \frac{\gamma}{k_0}\right), \tag{3.3}$$

where γ is the accelerating parameter between the normal stress level k_0 and the higher stress level k_i in terms of temperature (Kelvin), which links the scale parameters θ_i at different stress levels.

This model can be generalized to represent various strategies, where there are a number of different experiments conducted under different settings.

Suppose that m represents the number of strategies conducted, and the probability density function is generally expressed for stress levels $i = 0, 1, ..., s-1$ as

$$f_i^m(t) = \left(\frac{\beta}{(\theta_i^m)^\beta}\right)\left(t - (\tau_i^m - \tau_i^{m*})\right)^{\beta-1} \exp\left[-\left(\frac{t-\tau_i^m}{\theta_i^m} + \sum_{j=1}^{i}\frac{\tau_j^m - \tau_{j-1}^m}{\theta_{j-1}^m}\right)^\beta\right], \tag{3.4}$$

where τ_i^{m*} is given by:

$$\tau_i^{m*} = \left[\frac{\theta_i^m}{\theta_{i-1}^m}\left(\tau_i^m - \tau_{i-1}^m + \tau_{i-1}^{m*}\right)\right], \tag{3.5}$$

for $\tau_i \leq t \leq \tau_{i+1}$, where t_i^m represents failure times which occurred at strategy m and stress level s_i, where $m = 0, 1, 2, ..., z$ and $i = 0, 1, ..., s-1$ for $t_i^m \in [\tau_i^m, \tau_{i+1}^m]$. Note that t_0^0 represents failures times occurring at the normal stress level without acceleration which follows the Weibull model, and it is assumed in this study that there are failure data at the normal stress level. The Arrhenius link function is given by

$$\theta_i^m = \theta_0^0 \exp\left(\frac{\gamma}{k_i^m} - \frac{\gamma}{k_0^{m_0}}\right), \tag{3.6}$$

where θ_0^1, γ and β are the three parameters need to be estimated.

Suppose that t represents a data set of failure times which occurred at different strategies, where $t = \{t_{0:1}^0, ..., t_{0:n_0^0}^0, t_{0:1}^1, ..., t_{0:n_0^1}^1, ..., t_{0:1}^2, ..., t_{0:n_0^2}^2, ..., t_{i:1}^2, ..., t_{i:n_0^2}^2, ..., t_{i:1}^m, ..., t_{i:n_i^m}^m\}$. where:

- $t^m_{i:j}$ denotes the j-th observed failure time at stress level i under strategy m,
- n^m_i is the number of failure times at stress level i within strategy m,
- n is the total number of failure times across all strategies and stress levels, i.e., $n = \sum_{m=0}^{z} \sum_{i=0}^{s-1} n^m_i$.

Thus, the likelihood function is given as

$$L(t; \theta^m_i, \beta, \gamma, k^m_i) = \prod_{m=0}^{z} \prod_{i=0}^{s-1} \prod_{n_i=1}^{n^m_i} f_i(t^m_{i:n^m_i}; \theta^m_i, \beta, \gamma, k^m_i). \tag{3.7}$$

In this case of the model, it is assumed that the data set is complete, where it can also involve right censored data.

Different strategies in the analysis refer to the various experimental designs and conditions under which failure times are observed. These strategies involve altering the experimental settings, such as applying different higher stress levels, to evaluate how failure times vary under each condition. The goal of incorporating multiple strategies is to measure the uncertainty and imprecision introduced by the different experimental setups, providing a more comprehensive understanding of how these variables affect failure behaviour.

3.3 The method

The method proceeds in three key steps. First, imprecision is introduced by applying the likelihood ratio test to the accelerating parameter, under the null hypothesis that all failure times originate from the same distribution. This provides a measure of uncertainty between different SSALT data sets. Second, failure times from different experimental strategies at higher stress levels are transformed to the normal stress level. Through this transformation, failure times are transformed as interval values at the normal stress level, with the assumption that these transformed values are indistinguishable from actual failure times at the normal stress level. Third, nonparametric predictive inference (NPI) is applied to the transformed data to provide robust predictive inference. These steps are explained as follows.

3.3.1 Imprecision using the likelihood ratio test

First, the likelihood ratio test is used to determine the amount of imprecision in the accelerating parameter γ in the Arrhenius link function. This implementation of the likelihood ratio test is typically to examine the equality of two independent failure times data sets, possibly involving right-censored data. This is a common statistical test, that can be applied to examine the equality of the probability distribution of two independent failure times data sets, which has been briefly explained in Section 2.6.2.

The likelihood ratio test is applied in a pairwise manner between strategy m_i and strategy m_0 to provide imprecision for the parameter γ. This imprecision for γ leads to lower and upper values for which the null hypothesis is not rejected under the assumption that all failure times come from the same underlying distribution. The null hypothesis, denoted as $H_0 : \gamma = \gamma^*$, assumes that the parameter γ is equal across strategies, while the alternative hypothesis $H_1 : \gamma \neq \gamma^*$ assumes that there is a difference in γ. Here, γ represents the accelerating parameter within the Arrhenius model and $\gamma \in \mathbb{R}^+$. The test statistic is defined as

$$\Lambda(T) = \frac{L(t; \tilde{\theta_0^0}, \beta, \gamma^*, k_i^m)}{L(t; \hat{\theta_0^0}, \beta, \hat{\gamma}, k_i^m)} \tag{3.8}$$

where $\tilde{\theta}_0^0 = \sup_{\theta_0^0 \in \mathbb{R}^+} L(T; \theta_0^0, \beta, \gamma^*, k_i^m)$ and $(\hat{\theta}_0^0, \hat{\gamma}) = \sup_{(\theta_0^0, \gamma) \in \mathbb{R}^+ \times \mathbb{R}^+} L(T; \theta_0^0, \beta, \gamma, k_i^m)$. Under H_0, the likelihood function is maximized on a restricted parameter space, where the parameter θ_0^0 is estimated, and γ^* is held fixed. The value γ^* is then compared to the likelihood maximized over the entire parameter space, yielding a range $[\underline{\gamma}, \overline{\gamma}]$ where the null hypothesis is not rejected. In this context, l_0 represents the log-likelihood maximized over the restricted space under H_0, and l_1 represents the log-likelihood maximized over the full parameter space. The likelihood ratio test statistic can alternatively be written as

$$LR = 2(l_1 - l_0) \tag{3.9}$$

This statistic follows a χ^2 distribution where the degree of freedom is the difference in the number of the parameters between the two models.

This concept can be extended to obtain different estimates of the lower and upper bounds for γ through pairwise comparisons between strategies. Specifically, these pairwise comparisons are performed between strategy $m = 0$ and each other strategy m, where $m = 1, 2, \ldots, z$. Each comparison results in lower and upper bounds, denoted as $\underline{\gamma}_{0,m}$ and

$\overline{\gamma}_{0,m}$, respectively. The overall lower bound for γ is determined by taking the minimum of these pairwise lower bounds, i.e., $\underline{\gamma} = \min(\underline{\gamma}_{0,m})$, and the overall upper bound is determined by taking the maximum of the upper bounds, i.e., $\overline{\gamma} = \max(\overline{\gamma}_{0,m})$. This approach captures the greater imprecision introduced through the pairwise comparisons between the strategies for the parameter γ, as it determines the overall lower and upper bounds for γ based on all the comparisons.

For each interval $\left[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}\right]$ derived from the pairwise tests, corresponding lower and upper bounds for the scale parameter $\hat{\theta}_0$ and the shape parameter $\hat{\beta}$ are obtained. However, these interval-based bounds for $\hat{\theta}_0$ and $\hat{\beta}$ are not used in the subsequent steps of the method. Instead, the method proceeds with the precise estimates of these parameters, $\hat{\theta}_0$ and $\hat{\beta}$, which are considered fixed and not affected by the interval estimations of γ. The scale parameter at the normal stress level is fixed to allow the imprecision in the model to directly reflect the uncertainty in the acceleration relationship across different stress levels.

3.3.2 Transformation of SSALT data

In the second step, a new method is introduced for the analysis of SSALT data and the corresponding statistical inference, enabling the transformation of failure times from higher stress levels to the normal stress level. First, this method assumes that the cumulative distribution functions (CDF's) of failure times at higher stress levels can be equated to those at the normal stress level. By using the inverse function of the CDF, failure times at higher stress levels are transformed to the normal stress level. Secondly, the transformed failure times are assumed to be indistinguishable from those occurring at the normal stress level. Thirdly, when the model fits perfectly, the transformed failure times should yield equivalent inferences to those from failure times observed at the normal stress level.

To clarify, let t_i denote failures times occurred at stress level s_i where $i = 0, 1, ..., s-1$ and $t_i \in [\tau_i, \tau_{i+1}]$. Also, let $t_i^{\rightarrow j}$ denote failures times which occurred at stress level i and are transformed to lower stress level s_j where $j = 0, 1, ..., i$ and $t_i^{\rightarrow j} \in [\tau_j, \infty)$. These lower levels include the normal stress level s_0 (normal use), which is the primary stress level to which failure times are typically transformed, as it represents the conditions under normal operation. Therefore, $t_i^{\rightarrow 0}$ are failures times which occurred at higher stress level

s_i and are transformed to the normal stress level s_0, where $t_i^{\to 0} \in [0, \infty)$.

The first assumption is that the probability that an experimental unit fails at time t_i at stress level s_i is equated to the probability that the experimental unit fails at time t_j at stress level s_j. This allows us to transform the failure times from higher stress levels s_i to lower stress levels s_j in a manner that $F_j(t_i^{\to j}) = F_i(t_i)$, where $t_i^{\to j}$ is therefore given by $t_i^{\to j} = F_j^{-1}(F_i(t_i))$. In particular, $t_i^{\to 0}$ is given by $t_i^{\to 0} = F_0^{-1}(F_i(t_i))$, when failures times are transformed from higher stress level s_i to stress level s_0.

The second assumption is that the transformed times of failures from higher stress levels s_i to lower stress levels s_j should not be distinguishable from original failures times at the stress level s_j. To verify this assumption, the Wilcoxon rank sum test is suggested as a nonparametric hypothesis test to determine whether the transformed failure times and the actual failure time samples at the normal stress level come from the same distribution. To verify this assumption, the Wilcoxon rank sum test is suggested as a nonparametric hypothesis test to determine whether the transformed failure times and the actual failure time samples at the normal stress level come from the same distribution. The null hypothesis states that the two data sets — transformed and actual failure times — are drawn from the same underlying distribution. If the assumption holds, the null hypothesis should not be rejected, confirming that the transformation preserves the same statistical properties and allow for the same statistical inferences as the real data distribution.

The third assumption is that once the model fits correctly, the transformed failure times will preserve the same statistical properties and allow for the same statistical inferences as the real data distribution. This implies that if the transformed data is combined with the actual data at the normal stress level, the overall data set will maintain the same statistical properties and allow for the same statistical inferences as the real data distribution, ensuring consistency in the model's predictions at the normal stress levels.

Furthermore, this concept of transformation can be generalized to transform data from higher strategies of SSALT experiments to lower strategies, including the normal stress level s_0. Each strategy is designed to independently examine a sample of experimental units, with these units subjected to stress levels s_i, for $i = 0, 1, 2, \ldots, s-1$.

Let t_i^m denote failure times that occurred under strategy m and stress level s_i, where $m = 0, 1, 2, \ldots, z$ and $i = 0, 1 \ldots, s-1$, for $t_i^m \in [\tau_i^m, \tau_{i+1}^m]$. Also, let $t_i^{(m)\to(m^*,i^*)}$ denote

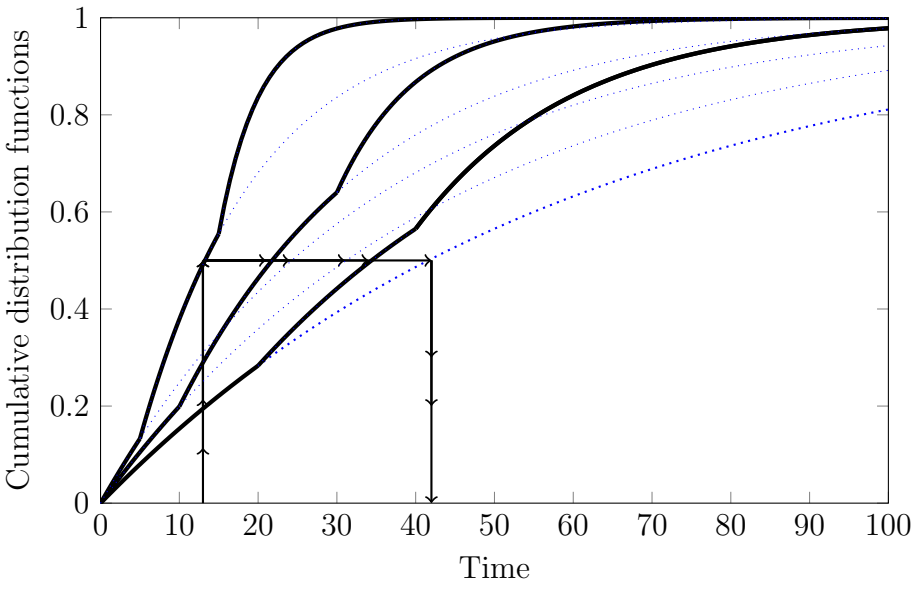


Figure 3.1: Illustration of the transformation method.

failure times that occurred under strategy m and stress level s_i, transformed to strategy m^* and stress level i^*, where $m^* = 0, 1, 2, \ldots, z$ and $i^* = 0, 1, \ldots, s-1$, for $t_i^{(m)\to(m^*,i^*)} \in [\tau_{i^*}^{m^*}, \infty)$. This transformation is achieved such that $F_{(i^*,m^*)}(t_i^{(m)\to(m^*,i^*)}) = F_{(i,m)}(t_i^m)$, and hence $t_i^{(m)\to(m^*,i^*)}$ is given by $t_i^{(m)\to(m^*,i^*)} = F^{-1}_{(i^*,m^*)}(F_{(i,m)}(t_i^m))$. In particular, $t_i^{(m)\to(0^*,0^*)}$ is given by $t_i^{(m)\to(0^*,0^*)} = F^{-1}_{(0^*,0^*)}(F_{(i,m)}(t_i^m))$, when failure times are transformed from a higher stress level s_i under strategy m to the normal stress level s_0 and the strategy $m = 0$ (normal use).

Figure 3.1 illustrates the transformation process where a unit failed under strategy 3 and stress level 2 at time $t_2^3 = 13$, corresponding to a cumulative probability of $F(t_2^3) = 0.50$. This failure time is transformed to several other times under different stress levels and strategies, maintaining the same cumulative probability of failure. For instance, $t_2^3 = 13$ is transformed to $t_2^{3\to(3,0)} = 21$, $t_2^{3\to(2,1)} = 24$, $t_2^{3\to(2,0)} = 31$, $t_2^{3\to(1,1)} = 34$, and $t_2^{3\to(0,0)} = 42$. These transformed times reflect the equivalent failure probabilities at different stress levels and strategies. The concept of equated probabilities refers to the process by which the cumulative distribution function (CDF) values at different stress levels are equated, ensuring that failure times transformed from one stress level to another maintain the same probability of failure.

In this section, the method of data transformation is applied when failure times follow a Weibull distribution. The cumulative distribution function (CDF) of the Weibull cumu-

lative exposure model (CEM), where failure times occur at stress level i under strategy m, is given by

$$F_{(i,m)}(t_i^m) = 1 - \exp\left[-\left(\frac{t_i^m - \tau_i^m}{\theta_i^m} + \sum_{j=1}^{i} \frac{\tau_j^m - \tau_{j-1}^m}{\theta_{j-1}^m}\right)^{\beta}\right], \quad \text{for} \quad \tau_i^m \leq t \leq \tau_{i+1}^m \quad (3.10)$$

where t_i^m are failures times occurred under strategy m and stress level s_i, where $m = 0, 1, 2, \ldots, z$ and $i = 0, 1, \ldots, s-1$ for $t_i^m \in [\tau_i^m, \tau_{i+1}^m]$. Also, $t_i^{(m)\to(m^*,i^*)}$ refers to failure times, which occurred at strategy m and stress level s_i, are transformed to strategy m^* and stress level i^* where $m^* = 0, 1, 2, \ldots, z$ and $i^* = 0, 1, \ldots, s-1$ for $t_i^{(m)\to(m^*,i^*)} \in [\tau_{i^*}^{m^*}, \infty)$. This transformation is achieved in a manner that $F_{(i^*,m^*)}(t_i^{(m)\to(m^*,i^*)}) = F_{(i,m)}(t_i^m)$, where $t_i^{(m)\to(m^*,i^*)} = F_{(i^*,m^*)}^{-1}(F_{(i,m)}(t_i^m))$. Therefore, $t_i^{(m)\to(m^*,i^*)}$ is given by

$$t_i^{(m)\to(m^*,i^*)} = \left[\left(\frac{t_i^m - \tau_i^m}{\theta_i^m} + \sum_{j=1}^{i} \frac{\tau_j^m - \tau_{j-1}^m}{\theta_{j-1}^m} - \sum_{j^*=1}^{i^*} \frac{\tau_{j^*}^{m^*} - \tau_{j^*-1}^{m^*}}{\theta_{j^*-1}^{m^*}}\right)\right] \theta_{i^*}^{m^*} + \tau_{i^*}^{m^*}, \quad \text{for} \quad \tau_{i^*}^{m^*} \leq t_{i^*}^{m^*} \leq \infty \quad (3.11)$$

In particular, $t_i^{(m)\to(0^*,0^*)} = F_{(0^*,0^*)}^{-1}(F_{(i,m)}(t_i^m))$ are failure times transformed from stress level s_i under strategy m to the stress level s_0 under $m = 0$. Therefore, $t_i^{(m)\to(0^*,0^*)}$ is given by

$$t_i^{(m)\to(0^*,0^*)} = \left[\left(\frac{t_i^m - \tau_i^m}{\theta_i^m} + \sum_{j=1}^{i} \frac{\tau_j^m - \tau_{j-1}^m}{\theta_{j-1}^m}\right)\right] \theta_{0^*}^{0^*}, \qquad \text{for} \qquad 0 \leq t_i^{(m)\to(0^*,0^*)} \leq \infty \quad (3.12)$$

Moreover, when failure times are transformed from higher stress levels to the stress level s_i under a given strategy, let t_i denote the failure times occurring at stress level s_i, where $i = 0, 1, \ldots, s-1$ and $t_i \in [\tau_i, \tau_{i+1}]$. The transformed failure times from stress level i to a lower stress level s_j, denoted as $t_i^{\to j}$, occur at $t_i^{\to j} \in [\tau_j, \infty)$, where $j = 0, 1, \ldots, i$. Specifically, when failure times are transformed from a higher stress level s_i to the normal stress level s_0, the transformed failure times, $t_i^{\to 0} \in [0, \infty)$. The expression for $t_i^{\to 0}$ is given by

$$t_i^{\to 0} = \left[\left(\frac{t_i - \tau_i}{\theta_i} + \sum_{j=1}^{i} \frac{\tau_j - \tau_{j-1}}{\theta_{j-1}}\right)\right] \theta_0, \quad \text{for} \quad 0 \leq t_i^{\to 0} \leq \infty. \quad (3.13)$$

This method transforms all failure times occurring at higher stress strategies to the normal stress level using the interval estimates $[\underline{\gamma}, \overline{\gamma}]$ obtained from the likelihood ratio test. Each failure time from a higher strategy is transformed to an interval-valued observation at the normal stress level s_0, with the width of the interval reflecting the imprecision introduced by the comparison between strategies. The interval is wider for

failure times originating from higher stress levels, capturing the greater imprecision in these transformations at the normal stress level.

3.3.3 Implementation of NPI

In the third step, nonparametric predictive inference (NPI) is implemented, as explained in Section 2.7, to provide robust predictive inference for future failure times, accounting for the imprecision introduced in the parameter γ through the likelihood ratio test. NPI is particularly suitable for this method because it does not require strong assumptions about the underlying distribution of failure times, offering a more adaptable approach to handling the transformed data.

NPI uses the observed data and transformed failure times to calculate lower and upper survival functions, denoted as $\underline{S}(t)$ and $\overline{S}(t)$, respectively. These functions provide lower and upper bounds for the survival probability of a future unit at a given time t, reflecting the imprecision in the data due to uncertainties in the parameter γ.

The final NPI lower and upper survival functions are constructed based on the minimum of the lower bounds and the maximum of the upper bounds for the transformed failure times and the original failure times at the normal stress level. The transformed data are obtained by applying the overall lower bound $\underline{\gamma}$ and upper bound $\overline{\gamma}$ to the transformation process, ensuring that the imprecision introduced by the pairwise likelihood ratio tests is reflected in the final survival functions.

This approach accounts for the imprecision across all pairwise likelihood ratio tests used to estimate γ. As explained earlier, pairwise comparisons are performed between strategy m_0 (the normal stress strategy) and each other strategy m, where $m = 1, 2, \ldots, z$. Each comparison yields lower and upper bounds for γ, denoted as $\underline{\gamma}_{0,m}$ and $\overline{\gamma}_{0,m}$, respectively. The overall lower bound for γ is determined by taking the minimum of the pairwise lower bounds, i.e., $\underline{\gamma} = \min(\underline{\gamma}_{0,m})$, and the overall upper bound by taking the maximum of the pairwise upper bounds, i.e., $\overline{\gamma} = \max(\overline{\gamma}_{0,m})$. This approach captures the greater imprecision introduced by the comparisons, determining the overall bounds for γ based on all the pairwise comparisons.

Therefore, the NPI lower survival function, $\underline{S}(t)$, is constructed using the transformed data from all higher strategies, based on the overall lower bound $\underline{\gamma}$, combined with the observed failure times at the normal stress level. Similarly, the NPI upper survival func-

tion, $\overline{S}(t)$, is constructed using the transformed data from all higher strategies, based on the overall upper bound $\overline{\gamma}$, alongside the observed failure times at the normal stress level.

In this method, pairwise tests are used for each strategy compared to the normal stress level, rather than performing a single test across all strategies. The rationale for using pairwise comparisons lies in the goal of preserving greater imprecision in the interval of γ. When conducting a single test to compare all strategies simultaneously, the result is often a more decisive rejection of the null hypothesis, which could lead to a smaller interval for γ, reducing the measure of uncertainty. This is because the combined data from all higher stress levels tends to make it easier to reject the assumption that failure times across different levels come from the same distribution.

By contrast, performing separate pairwise tests between each higher strategy and the normal stress level ensures that the null hypothesis is harder to reject in each individual test. This approach allows for the measurement of imprecision between the two strategies, resulting in a larger interval for γ. The larger interval captures the inherent variability and uncertainty between different strategies. Therefore, the pairwise method provides a more flexible and broader way to account for imprecision, which is essential when the goal is to reflect the differences between strategies with the larger interval for γ.

Each observation from a higher stress strategy is transformed into an interval-valued observation at the normal stress level, with the width of the interval increasing for observations from higher stress strategies. If the model is reasonably accurate and correct, the widest interval for the parameter γ is expected to come from the likelihood ratio test applied between strategy m_1 and the normal stress strategy m_0. When the model fits well, a transformed observation from strategy m_1 will typically result in a narrower interval at the normal stress level compared to an observation from strategy m_2, particularly when the intervals are close or overlapping. In overlapping cases, since the interval from strategy m_2 is generally wider, the endpoints of this interval will be further apart, increasing the probability of rejecting the null hypothesis for γ. Such a scenario is likely to occur in cases of model misspecification or significant overlap between the data across different strategies.

Although we employ pairwise tests in our approach, we do not aggregate them into an overall confidence level statement for the final inference. Instead, we use NPI to derive the lower and upper predictive survival functions and examine the performance of our

predictive method independently through simulations. If the model is perfectly accurate and correct, the lower and upper bounds for γ will form an interval with a confidence level of at least $1-\alpha$, where α represents the significance level for each pairwise test. However, we develop our method with the understanding that the assumed model may not be fully accurate and correct in practice, and this acknowledgment makes traditional confidence statements less appropriate to apply.

The proposed method is explained by illustrative examples in Section 3.4 and the performance of this method is investigated via simulation studies presented and discussed in Section 3.5.

3.4 Illustrative examples

In this section, three examples are presented to demonstrate the proposed method outlined in Section 3.3. Example 3.4.1 involves three generated data sets, each with a sample size of $n = 10$ across all strategies. These strategies include two accelerating strategies, along with data generated at the normal stress level. In Example 3.4.2, the sample sizes are increased to $n = 100$ to evaluate the impact of larger sample sizes on the results. It should be noted that while the method does not require equal sample sizes at each strategy, equal sample sizes are used in all the examples presented in this chapter for simplicity and consistency. If the sample sizes are unequal when comparing two datasets using the likelihood ratio test, the method remains valid; however, the level of imprecision may increase, potentially resulting in wider intervals for the parameter estimates.

Example 3.4.1 This example consists of four cases. In Case 1, the shape parameter is assumed to be constant and known for each strategy. The Arrhenius link function is implemented to connect the scale parameters at all stress levels within the strategies, both for the simulated data and for the analysis. The Arrhenius link function is implemented to connect the scale parameters at all stress levels within the strategies, both for the simulated data and for the analysis. In Case 2, the shape parameter is considered unknown and is estimated during the analysis. In Case 3, the Arrhenius link function is replaced by the Eyring link function. This replacement is done to evaluate the model's performance when the assumed link function does not provide a good fit. The focus of Case 3 is to examine the interval $[\underline{\gamma}, \overline{\gamma}]$ and its effect on the corresponding NPI lower

strategies	Stress level	Data sets	Failures times
m_0	s_0	t_0	1755.83, 4149.09, 4799.74, 5095.24, 7454.25, 7563.30, 8245.62, 10385.32, 11166.58, 12411.59
m_1	s_0^1	t_0^1	185.63, 222.74
m_1	s_1^1	t_1^1	300.66, 320.37, 325.00, 341.08, 342.94
m_1	s_2^1	t_2^1	359.43, 360.06, 371.28
m_2	s_0^3	t_0^2	14.28, 42.54
m_2	s_1^2	t_1^2	101.73, 105.56, 108.97, 120.80, 123.50
m_2	s_2^2	t_2^2	138.93, 140.52, 146.53

Table 3.1: A simulated data of Example 3.4.1.

and upper survival functions at the normal stress level. Finally, in Case 4, the shape parameter is set to 1, assuming to an exponential distribution, to test the method under an additional case of model misspecification using the Eyring link function. This case evaluates the method's robustness and the implications of the exponential assumption on the model's performance.

Three data sets are generated, each corresponding to a different strategy. In the first strategy, the normal temperature level is assumed to be $k_0^0 = 300$ Kelvin, with the scale parameter set to $\theta_0^0 = 7000$. This strategy represents the normal use stress level. Ten observations are generated from the Weibull distribution for this strategy. The Arrhenius link function is used to link the scale parameters across all strategies, with the accelerating parameter set to $\gamma = 5000$.

In the second strategy, the temperature levels are assumed to be $k_0^1 = 350$, $k_1^1 = 400$, and $k_2^1 = 450$ Kelvin for stress levels s_0^1, s_1^1, and s_2^1, respectively. The stress level increases from s_0^1 to s_1^1 at $\tau_1^1 = 300$, and from s_1^1 to s_2^1 at $\tau_2^1 = 350$. The corresponding scale parameters are $\theta_0^1 = 647.23$, $\theta_1^1 = 108.52$, and $\theta_2^1 = 8.72$ for s_0^1, s_1^1, and s_2^1, respectively.

In the third strategy, the temperature levels are assumed to be $k_0^2 = 380$, $k_1^2 = 420$, and $k_2^2 = 460$ Kelvin for stress levels s_0^2, s_1^2, and s_2^2, respectively. The stress level increases from s_0^2 to s_1^2 at $\tau_1^2 = 100$, and from s_1^2 to s_2^2 at $\tau_2^2 = 130$. The corresponding scale parameters are $\theta_0^2 = 209.53$, $\theta_1^2 = 59.84$, and $\theta_2^2 = 21.25$ for s_0^2, s_1^2, and s_2^2, respectively. The generated failure times for each strategy are shown in Table 3.1.

Cases	Significance Level	0.01		0.05		0.10	
	strategies	$\underline{\gamma}$	$\overline{\gamma}$	$\underline{\gamma}$	$\overline{\gamma}$	$\underline{\gamma}$	$\overline{\gamma}$
Case 1	m_1, m_0	4434.32	5674.48	4627.98	5577.03	4723.69	5524.58
	m_2, m_0	4483.03	5538.61	4645.16	5454.20	4725.78	5408.93
Case 2	m_1, m_0	4446.87	5987.43	4641.56	5747.77	4733.31	5639.55
	m_2, m_0	4419.83	6021.34	4616.62	5648.72	4709.40	5648.72
Case 3	m_1, m_0	4094.42	5639.92	4289.67	5399.47	4381.69	5290.88
	m_2, m_0	4062.62	5669.93	4260.06	5411.82	4353.14	5295.93
Case 4	m_1, m_0	3868.23	6352.40	4234.90	6166.11	4417.23	6064.27
	m_2, m_0	3818.89	6074.06	4137.48	5901.83	4296.78	5806.65

Table 3.2: $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ for Example 3.4.1.

To analyse the data sets given in Table 3.1, the Weibull cumulative exposure model is assumed for the accelerating strategies m_1 and m_2, and the Weibull model for the normal stress strategy m_0. First, the pairwise likelihood ratio test is applied between m_0 and each of the other strategies m (where $m = 1, 2$), to determine the intervals $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ for which the null hypothesis is not rejected. The resulting intervals $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ for three significance levels are shown in Table 3.2. It should be noted that the transformed data are based on the overall values $[\underline{\gamma}, \overline{\gamma}]$, obtained as the minimum and maximum values from the pairwise tests. All failure times at the higher stress strategies are transformed to the normal stress level. Consequently, the failure times at the higher stress strategies m_1 and m_2 are transformed into interval-valued data at the normal stress level s_0 in strategy m_0.

Figure 3.2 illustrates some of the data at each strategy being transformed into intervals at the normal stress level. It shows that the largest interval of the transformed data is derived from the highest stress level s_2^2. This highlights an essential property of the proposed method: transformed data from higher stress levels tend to produce larger intervals at the normal stress level. This is crucial because larger intervals reflect greater uncertainty at higher stress levels, capturing the increased variability and imprecision in the data as the stress level rises.

The analysis was performed for four different cases, and the results are summarized as follows. In Case 1, it is assumed that the shape parameter is constant and known. The analysis shows that $\hat{\theta}_0^0 = 7843.90$ and $\hat{\gamma} = 5196.17$ for the first and second strategies, while

for the first and third strategies, $\hat{\theta}_0^0 = 7865.73$ and $\hat{\gamma} = 5129.62$. The scale parameter from the initial analysis is used to transform all data instead of using the scale parameters of the corresponding estimation in the test. The estimation of the scale parameter at the normal stress level is based on the data at the normal stress level and the transformed data, derived from the pairwise comparison of the first and second strategies, yielding $\hat{\theta} = 7843.90$. This confirms an essential result in the proposed method: the transformed data, when combined with the original stress level data, will result in an equivalent estimation of the scale parameter if fitted to the Weibull model or Weibull cumulative exposure model. The estimation of the scale parameter at the normal stress level based on the transformed data and the original data from the pairwise comparison of the first and third strategies yields $\hat{\theta} = 7865.73$. Table 3.3 provides the transformed data for Case 1 in Example 3.4.1, showing the failure times from different strategies transformed using the lower bound $\underline{\gamma}$, the point estimate $\hat{\gamma}$, and the upper bound $\overline{\gamma}$ across three significance levels: 0.01, 0.05, and 0.10. The results, as shown in Table 3.3 and Figure 3.2, clearly illustrate that the transformed data from higher stress levels (s_1^1 and s_2^1 for m_1 and s_1^2 and s_2^2 for m_2) produce larger intervals at the normal stress level (m_0). This aligns with the method discussed earlier, where higher stress levels tend to yield wider intervals due to the increased variability and imprecision in the data.

Case 2 assumes that the constant shape parameter is unknown. The analysis reveals that $\hat{\theta}_0^0 = 7892.93$, $\hat{\gamma} = 5168.44$, and $\hat{\beta} = 2.17$ for the first and second strategies, and $\hat{\theta}_0^0 = 7805.66$, $\hat{\gamma} = 5154.11$, and $\hat{\beta} = 1.82$ for the first and third strategies.

Case 3 also assumes the shape parameter is unknown but replaces the Arrhenius link function with the Eyring link function to assess the method's performance under misspecification. The analysis indicates that $\hat{\theta}_0^0 = 7875.32$, $\hat{\gamma} = 4817.98$, and $\hat{\beta} = 2.16$ for the first and second strategies, while for the first and third strategies, $\hat{\theta}_0^0 = 7780.42$, $\hat{\gamma} = 4797.22$, and $\hat{\beta} = 1.81$.

In Case 4, the shape parameter is set to 1, assuming an exponential model to test the method under another misspecification. The results show that $\hat{\theta}_0^0 = 8442.35$, $\hat{\gamma} = 5326.69$ for the first and second strategies, and $\hat{\theta}_0^0 = 7982.48$, $\hat{\gamma} = 5102.45$ for the first and third strategies.

After performing the steps of the analysis, it is important to recall how the NPI survival functions are derived. The NPI lower survival function $\underline{S}$ is obtained by combining

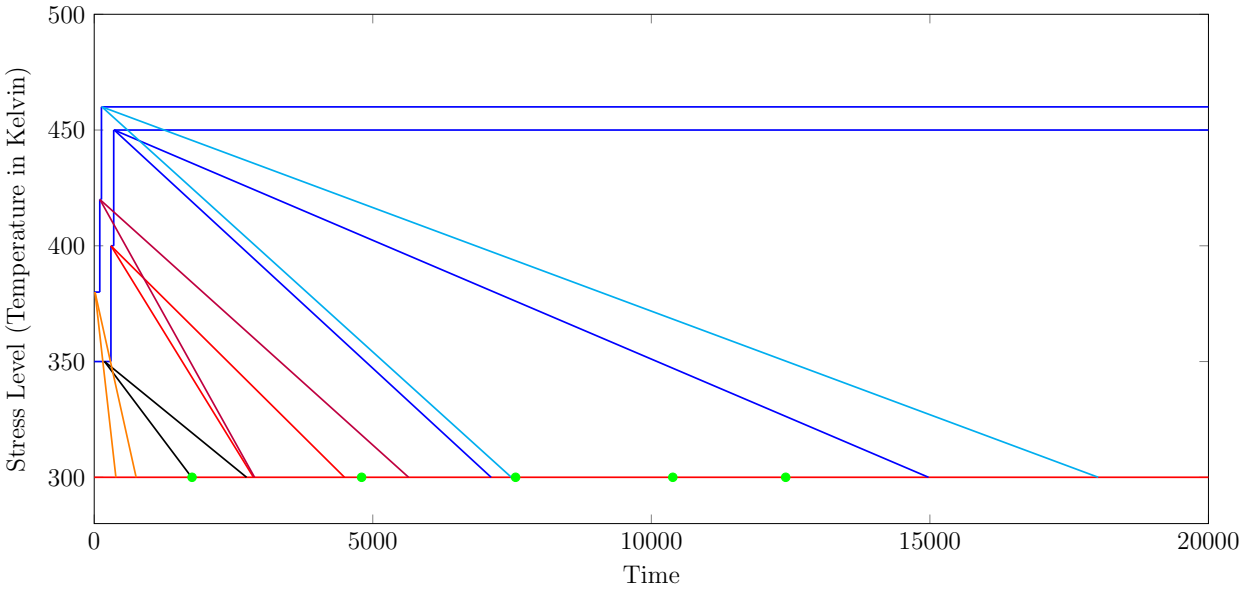


Figure 3.2: This figure illustrates how some failure times from higher stress levels are transformed into intervals at the normal stress level. The green dots represent the original failure times observed directly at the normal stress level. In Case 1 of Example 3.4.1.

the data at the normal stress level with the transformed data from the higher stress strategies, specifically from m_1 to m_0 and from m_2 to m_0, using the overall lower bound $\underline{\gamma}$. The lower points of the transformed data represent the pessimistic case, resulting in the NPI lower survival function $\underline{S}$. Similarly, the NPI upper survival function $\overline{S}$ is derived by combining the data at the normal stress level with the transformed data from the higher stress strategies, again from m_1 to m_0 and from m_2 to m_0, but using the overall upper bound $\overline{\gamma}$. The upper points of the transformed data represent the optimistic case, which provides the NPI upper survival function $\overline{S}$. The difference between the lower and upper NPI survival functions reflects the amount of imprecision in the data, capturing the uncertainty associated with the model.

The Table 3.3 also highlights the effect of different significance levels on the transformation of failure times. For instance, the transformation based on the 0.01 significance level results in larger intervals for both $\underline{\gamma}$ and $\overline{\gamma}$ compared to those based on the 0.10 significance

Data set	Failures times	$\underline{\gamma}_{0.01}$	$\underline{\gamma}_{0.05}$	$\underline{\gamma}_{0.10}$	$\hat{\gamma}$	$\overline{\gamma}_{0.10}$	$\overline{\gamma}_{0.05}$	$\overline{\gamma}_{0.01}$
t_0^1	185.64	1523.07	1672.70	1748.26	2175.40	2734.40	2888.94	3265.30
t_0^1	222.74	1827.47	2007.01	2097.66	2610.17	3280.90	3466.33	3917.90
t_1^1	300.66	2487.77	2734.29	2858.88	3564.83	4492.48	4749.65	5377.21
t_1^1	320.38	3271.79	3658.04	3856.88	5027.90	6675.62	7153.27	8355.29
t_1^1	325.01	3456.01	3875.08	4091.38	5371.67	7188.57	7718.03	9055.03
t_1^1	341.08	4095.25	4628.24	4905.07	6564.54	8968.53	9677.74	11483.13
t_1^1	342.94	4169.40	4715.60	4999.46	6702.91	9175.01	9905.08	11764.79
t_2^1	359.44	5731.23	6640.59	7124.14	10170.08	14975.47	16471.41	20424.00
t_2^1	360.07	5816.78	6747.05	7242.15	10366.62	15310.58	16852.40	20930.99
t_2^1	371.28	7339.20	8641.57	9342.38	13864.29	21274.56	23632.67	29953.82
t_0^2	14.29	317.66	364.70	389.23	531.80	752.46	815.96	977.34
t_0^2	42.54	945.86	1085.94	1158.99	1583.49	2240.55	2429.63	2910.16
t_1^2	101.73	2340.18	2693.52	2878.25	3957.23	5643.13	6131.34	7377.48
t_1^2	105.56	2597.86	3004.32	3217.76	4475.79	6473.70	7058.44	8561.87
t_1^2	108.97	2827.39	3281.16	3520.18	4937.69	7213.51	7884.24	9616.84
t_1^2	120.81	3623.92	4241.89	4569.65	6540.62	9780.88	10750.01	13277.91
t_1^2	123.51	3805.87	4461.35	4809.38	6906.77	10367.32	11404.62	14114.19
t_2^2	138.94	5745.40	6876.06	7487.14	11306.32	18021.94	20117.57	25744.04
t_2^2	140.52	6011.22	7210.01	7859.02	11929.09	19126.98	21380.89	27446.21
t_2^2	146.53	7021.43	8479.14	9272.27	14295.81	23326.44	26181.84	33914.93

Table 3.3: Failure times represent the original data, while $\underline{\gamma}_{0.01}$, $\underline{\gamma}_{0.05}$, $\underline{\gamma}_{0.10}$, $\hat{\gamma}$, $\overline{\gamma}_{0.10}$, $\overline{\gamma}_{0.05}$, and $\overline{\gamma}_{0.01}$ are the transformed data based on the estimated value $\hat{\gamma}$ and its lower and upper bounds at significance levels 0.01, 0.05, and 0.10. Example 3.4.1.

level. This observation holds consistently across all the failure times and strategies, confirming the expected outcome: smaller significance levels yield wider intervals, reflecting greater uncertainty in the pairwise comparison of strategies. As a result, the transformed failure times are more spread out, especially at the 0.01 level, compared to the narrower intervals observed at the 0.10 level. For example, for the failure time $t_1^1 = 341.08$, the transformed values range from 4095.25 to 11483.13 at the 0.01 significance level, which produces the widest interval. In contrast, the 0.10 significance level results in a narrower range (4905.07 to 8968.53).

This pattern is consistent across all failure times and strategies, reinforcing the con-

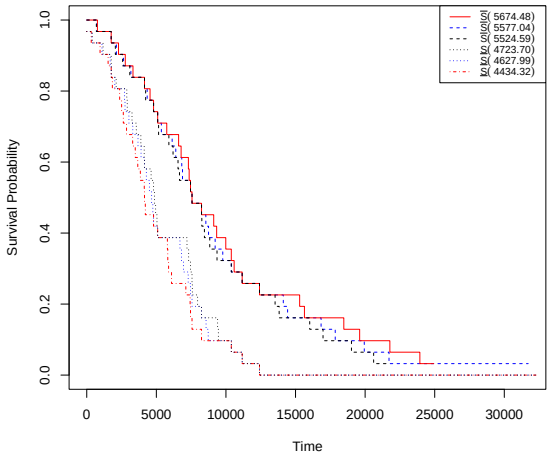
clusion that lower significance levels preserve greater imprecision in the intervals, which is critical for reflecting the inherent variability and uncertainty in the transformed data. Additionally, the results in Table 3.3 support the robustness of the proposed method in capturing the effect of higher stress levels on the transformed failure times, particularly when preserving imprecision in the estimation of γ.

The analysis across the four cases reveals important insights regarding the performance of the proposed method, as shown in the results from Figure 3.3 and Table 3.2. In Case 1, where the shape parameter is constant and known, the transformed data produces relatively narrower intervals for γ compared to the other cases. The survival functions for Case 1, as illustrated in Figure 3.3a, show less imprecision between the lower and upper bounds, suggesting that the model fits well, and there is less uncertainty in the transformed data. This result aligns with the expectation that a known shape parameter provides more precision in estimating the scale parameter at the normal stress level.

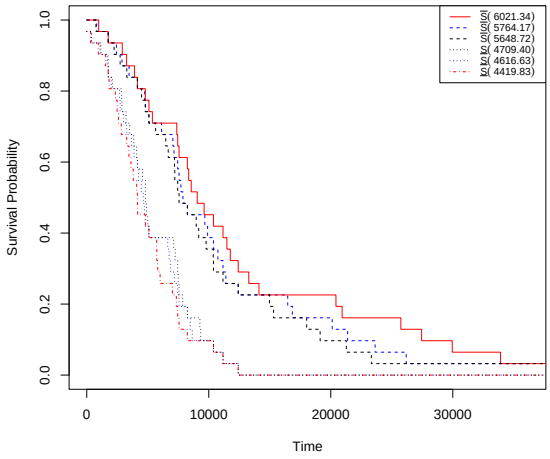
In Case 2, where the shape parameter is estimated rather than known, the intervals for γ widen slightly, reflecting the added uncertainty from estimating both the shape and scale parameters. As seen in Figure 3.3b, the wider intervals between the NPI lower and upper survival functions indicate greater imprecision in the transformed data. This additional imprecision is expected, given the increased complexity of the estimation process when both parameters are unknown. The pairwise comparison of strategies ensures that this uncertainty is captured effectively.

Case 3, which introduces the Eyring link function to demonstrate model misspecification, results in even wider intervals for γ. The survival functions in Figure 3.3c show substantial divergence between the lower and upper NPI survival functions, highlighting the impact of misspecification. The method captures this misspecification through the increased imprecision in the intervals, demonstrating its robustness in scenarios where the assumed link function does not fit well.

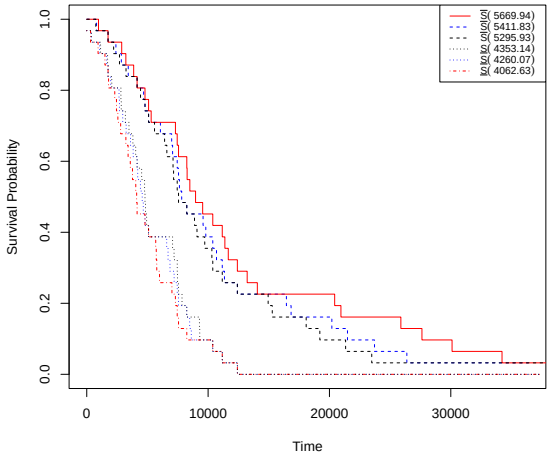
Case 4, in which an exponential Eyring cumulative exposure model is used to further examine model misspecification, shows the wider intervals between the lower and upper NPI survival functions, as illustrated in Figure 3.3d. The high imprecision observed in this case underscores the method's sensitivity to misspecification. The difference between the NPI lower and upper survival functions suggests that the exponential assumption poorly fits the data, emphasizing the need for further model evaluation. In



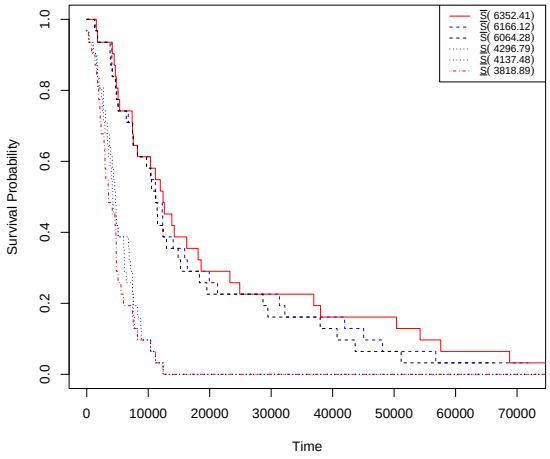
(a) Case 1



(b) Case 2



(c) Case 3



(d) Case 4

Figure 3.3: NPI lower and upper survival functions for all four cases in Example 3.4.1, with red, blue, and black lines representing significance levels 0.01, 0.05, and 0.10, respectively.

cases where such high imprecision is observed, it is advisable to either investigate the model assumptions in greater detail or collect more data to improve estimation accuracy.

This analysis demonstrates that while the proposed method, particularly through the use of NPI, effectively captures imprecision via pairwise comparisons, the model fit and how well the assumptions align with reality significantly influence the width of the intervals. Therefore, when the model shows considerable imprecision, particularly under

misspecification, it is essential to re-evaluate the model or gather additional data to ensure robust and reliable predictions.

Example 3.4.2 The previous example applied the methods to a data set with $n = 10$ observations per strategy. In this case, a larger data set is used to assess the impact of the data quantity on the resulting inference. A total of 100 failure times for each of the three strategies were simulated, yielding 300 failure times in the study. The experimental conditions remain consistent with those described in Example 3.4.1. The resulting intervals $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ for three significance levels are presented in Table 3.4.

The analysis of the larger data set, with $n = 100$ at each strategy, reveals several key differences compared to the results from the smaller data set ($n = 10$). In particular, the larger sample size provides more precise estimates of the parameters and narrows the intervals for γ, which directly impacts the inference drawn from the model.

In Case 1, with $n = 100$, the shape parameter is still assumed to be constant and known. The results show that $\hat{\theta}_0^0 = 6830.57$ and $\hat{\gamma} = 4901.79$ for the first and second strategies, and $\hat{\theta}_0^0 = 6741.04$ and $\hat{\gamma} = 5033.41$ for the first and third strategies. Compared to the smaller data set, the larger sample size leads to narrower intervals for γ, which demonstrates greater precision in estimating the scale parameter at the normal stress level. This result is consistent with the expectation that increased data quantity reduces uncertainty. The NPI lower and upper survival functions, as shown in Figure 3.4a, are closer together in this case, indicating reduced imprecision in the predictions compared to the smaller sample.

In Case 2, where the shape parameter is unknown and estimated, the larger data set leads to improved parameter estimation. The analysis results in $\hat{\theta}_0^0 = 6847.35$, $\hat{\gamma} = 4895.15$, and $\hat{\beta} = 2.04$ for the first and second strategies, and $\hat{\theta}_0^0 = 6694.25$, $\hat{\gamma} = 5044.70$, and $\hat{\beta} = 1.90$ for the first and third strategies. Compared to the smaller data set, the intervals for γ are narrower, indicating that the increased data has enhanced the precision of the estimates. Figure 3.4b shows that the NPI lower and upper survival functions are much closer than in the previous example, reflecting the greater precision provided by the larger sample.

In Case 3, where the Eyring link function is used to introduce model misspecification, the larger data set again results in narrower intervals for γ. The results show $\hat{\theta}_0^0 = 6829.40$, $\hat{\gamma} = 4543.31$, and $\hat{\beta} = 2.03$ for the first and second strategies, and $\hat{\theta}_0^0 = 6679.91$, $\hat{\gamma} =$

Cases	Significance Level	0.01		0.05		0.10	
	strategies	$\underline{\gamma}$	$\overline{\gamma}$	$\underline{\gamma}$	$\overline{\gamma}$	$\underline{\gamma}$	$\overline{\gamma}$
Case 1	m_1, m_0	4673.92	5124.74	4728.96	5071.64	4756.96	5044.43
	m_2, m_0	4831.47	5233.29	4880.12	5185.53	4904.91	5161.11
Case 2	m_1, m_0	4672.39	5124.61	4725.86	5067.90	4753.06	5039.41
	m_2, m_0	4831.57	5265.34	4882.74	5210.88	4908.79	5183.51
Case 3	m_1, m_0	4320.87	4773.86	4374.42	4717.03	4401.66	4688.49
	m_2, m_0	4477.47	4912.81	4528.80	4858.12	4554.93	4830.65
Case 4	m_1, m_0	4612.38	5478.31	4717.50	5375.81	4771.05	5323.33
	m_2, m_0	4579.79	5399.41	4678.21	5301.26	4728.46	5251.13

Table 3.4: $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ for Example 3.4.2.

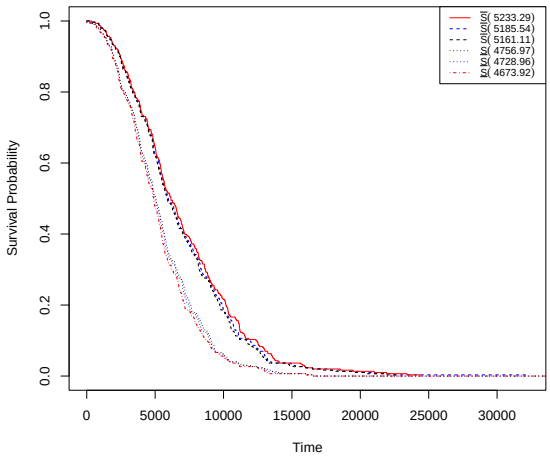
4691.24, and $\hat{\beta} = 1.89$ for the first and third strategies. While the model misspecification still leads to increased imprecision compared to Case 1 and 2, the larger data set overcomes the extent of this imprecision. The NPI lower and upper survival functions, as shown in Figure 3.4c, demonstrate less imprecision compared to the smaller data set, confirming that the increased sample size reduces the impact of misspecification.

In Case 4, where the shape parameter is set to 1 (assuming an exponential Eyring cumulative exposure model), the results are $\hat{\theta}_0^0 = 7426.50$ and $\hat{\gamma} = 5048.43$ for the first and second strategies, and $\hat{\theta}_0^{02} = 6770.28$ and $\hat{\gamma} = 4990.53$ for the first and third strategies. The exponential model, which was tested under another misspecification, produces a high level of imprecision, but this imprecision is reduced compared to the smaller data set. The NPI lower and upper survival functions, as shown in Figure 3.4d, still show significant imprecision, though they are narrower compared to the results from the previous example.

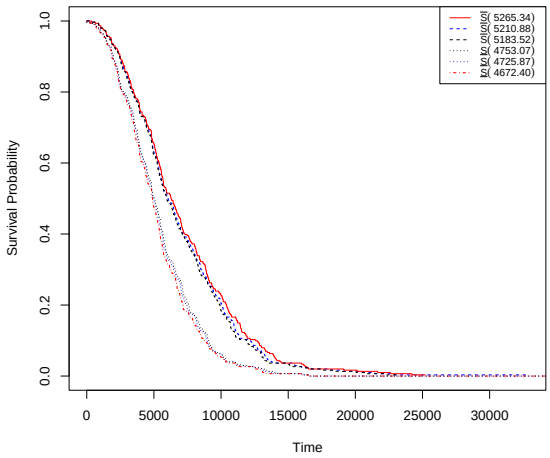
The larger sample size significantly reduces the imprecision in the estimates across all cases. The narrower intervals for γ across all significance levels, as shown in Table 3.4, illustrate the improved precision in parameter estimation with increased data. The NPI survival functions across all cases demonstrate less imprecision between the lower and upper bounds, which reflects the reduced uncertainty in the predictions. This result aligns with the expectation that larger data sets yield more reliable and precise inferences, especially in scenarios with model misspecification, as seen in Case 3 and 4.

In conclusion, the comparison of the two examples demonstrates that increasing the

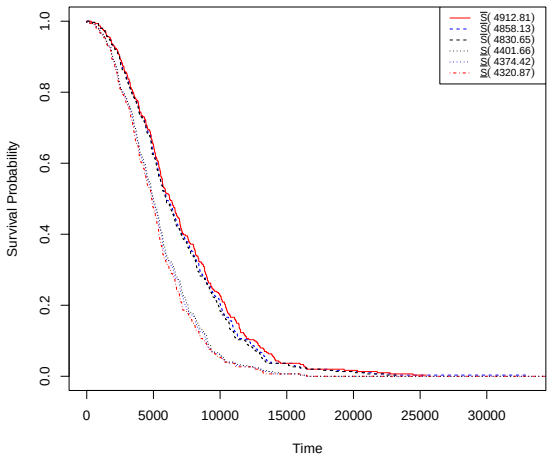
sample size has a substantial effect on reducing imprecision and improving the accuracy



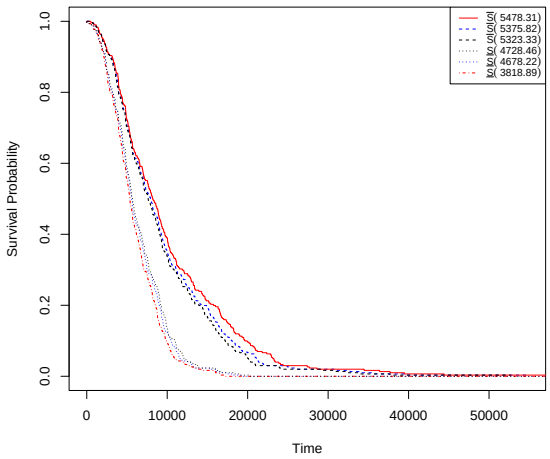
(a) Case 1



(b) Case 2



(c) Case 3



(d) Case 4

Figure 3.4: NPI lower and upper survival functions for all four cases in Example 3.4.2, with red, blue, and black lines representing significance levels 0.01, 0.05, and 0.10, respectively.

of parameter estimates, as well as narrowing the NPI survival function intervals. This further emphasizes the importance of sufficient data quantity when applying the proposed method to achieve robust and reliable predictions.

3.5 Simulation studies

In this section, simulation studies are performed to assess the performance of the proposed method outlined in Section 3.3. The simulation is carried out using three strategies of SSALT data, following the experimental setup described in Example 3.4.1. In the first strategy, the normal stress level is defined by assuming $\theta_0^0 = 7000$ and a normal temperature of $k_0^0 = 300$. In the second strategy, three temperature stress levels are introduced: $k_0^1 = 350$, $k_1^1 = 400$, and $k_2^1 = 450$, with the stress levels increasing at $\tau_1^1 = 300$ and $\tau_2^1 = 350$. The third strategy also considers three temperature stress levels: $k_0^2 = 380$, $k_1^2 = 420$, and $k_2^2 = 460$, with the stress levels increasing at $\tau_1^2 = 100$ and $\tau_2^2 = 130$. Additionally, the shape parameter is set to $\beta = 2$, and the Arrhenius link function parameter γ between all stress levels under higher strategies is assumed to be $\gamma = 5000$.

This simulation was repeated 10,000 times using data generated from the proposed model, with varying sample sizes of $n = 20$, $n = 50$, and $n = 100$ for each strategy, and significance levels set at 0.01, 0.05, and 0.10. The method's performance is assessed by simulating a future observation at the normal stress level m_0 and examining how well this future observation mixes with the transformed data from the higher stress strategies, as well as with the data at the normal stress level. The assessment focuses on whether the future observation falls within the quartile range of the NPI lower and upper survival functions. The performance is evaluated by determining whether the future observation surpasses the quartiles of the NPI lower and upper survival functions. The quartiles provide a useful measure for evaluating the overall performance, though other quantiles could also be used in a similar manner for further analysis.

First, the proportions are calculated to determine whether the future observation exceeds the quartiles of the NPI lower and upper survival functions in the right proportions. For a good performance, the first, second, and third quartiles of the lower NPI survival functions are expected to be greater than 0.75, 0.50, and 0.25, respectively. Similarly, the first, second, and third quartiles of the upper NPI survival functions should be less than 0.75, 0.50, and 0.25, respectively.

The simulation studies were conducted across three distinct cases to evaluate the performance of the proposed method under various conditions. In Case 1, we consider

the scenario previously discussed in Example 3.4.1, where the shape parameter is constant and known. Table 3.5 and Figures 3.5–3.7 illustrate the outcomes of this simulation for different significance levels ($\alpha = 0.01, 0.05, 0.10$) and sample sizes ($n = 10, 50, 100$). In these figures, the first, second, and third quartiles are denoted as $qL0.25$, $qU0.25$, $qL0.50$, $qU0.50$, $qL0.75$, and $qU0.75$, representing the NPI lower and upper survival functions, respectively. These quartiles provide insight into how well the future observations align with the predicted intervals, revealing several key outcomes from the simulation.

The overall results demonstrate satisfactory performance across all settings, with the proportions of future observations consistently meeting the expected thresholds of 0.75, 0.50, and 0.25 for the lower NPI survival function quartiles, and remaining below these thresholds for the upper NPI survival function quartiles. The differences in proportions between the lower and upper survival function quartiles serve as indicators of the level of imprecision in predictive inferences, with larger differences reflecting greater imprecision in the predictions.

A larger sample size generally reduces imprecision, leading to more consistent proportions between the quartiles. Specifically, for a larger sample size ($n = 100$), the quartile proportions are closer to their expected values compared to smaller sample sizes ($n = 20$), indicating a less imprecision in the predictive inferences. This reduction in imprecision for larger samples is evident in Figure 3.7, where the quartiles become narrower and more closely aligned with the theoretical expectations.

Data from higher stress level strategies, such as m_2, exhibit higher imprecision compared to lower stress level strategies like $m_1 m_0$. Here, m_0 and m_1 refer to two distinct data sets obtained from specific experimental strategies, each corresponding to unique combinations of stress levels and step durations designed to simulate accelerated life conditions. The larger differences in proportions for higher stress levels indicate that predictive inferences are less precise in these scenarios, likely due to greater variability inherent at higher stress conditions. Additionally, the overall results for the lower and upper γ parameters show a higher degree of imprecision than those derived from individual strategies $m_1 m_0$ and $m_2 m_0$. This trend underscores the impact of stress levels on the predictive performance of the method, with higher stress conditions generally leading to increased imprecision in the predictive inferences.

The influence of significance level (α) on predictive precision is less pronounced than

that of sample size or stress level, though it remains notable. Higher significance levels ($\alpha = 0.10$) result in slightly narrower intervals of γ, reflecting a marginally narrower level of imprecision. Conversely, lower significance levels ($\alpha = 0.01$) produce larger intervals of γ with a slightly higher level of imprecision in the NPI lower and upper survival functions. However, the effect of significance level is mitigated with larger sample sizes, where the impact on predictive imprecision is comparatively smaller. This finding suggests that while the choice of α can influence interval width of γ, sufficient sample size plays a more substantial role in enhancing imprecision.

The results of the simulation studies illustrate that the proposed method provides robust predictive inference when model assumptions are correct, especially with larger sample sizes. The figures and tables consistently show that increasing the sample size reduces the imprecision between the NPI lower and upper survival functions and brings the quartile proportions closer to their theoretical values. This reduction in imprecision is particularly clear in the largest sample sizes, where the quartiles show the narrowest intervals and the most consistent alignment between the lower and upper survival functions.

This chapter presents a predictive inference method designed to be robust against model misspecification by incorporating imprecision in the link function and distributional assumptions across stress levels. The approach is based on minimal assumptions, aiming to improve inference under varying experimental conditions.

In Case 2, robustness is examined under conditions of model misspecification. Specifically, this scenario investigates how a simple model might retain robustness when the shape parameter differs between the sampling and analysis models. In this case, the shape parameter β is set to 3 in the sampling model but is assumed to be 2 in the analysis. Table 3.6 and Figures A.1–A.3 present the predictive performance outcomes of the proposed method under these conditions. The simulation follows the same setup as in Case 1, with data generated according to the model described in Section 3.2, where $\beta = 3$, while the analysis incorrectly assumes $\beta = 2$ for each strategy.

Both Case 1 and Case 2 assume a known and constant shape parameter; however, Case 2 introduces additional imprecision in the proportions of future observations exceeding the specified quartiles, as depicted in Figures A.1–A.3. This imprecision is particularly noticeable in the NPI lower and upper survival functions for Case 2, where the difference in

proportions remains relatively large, even with larger sample sizes (n), as shown in Table 3.6. Typically, increased sample sizes reduce imprecision, but in Case 2, the imprecision remains relatively high compared to Case 1, suggesting that model misspecification affects the size of imprecision in predictive inferences.

These findings indicate that, while the proposed method demonstrates robustness, Case 2 shows a slight increase in imprecision due to the misspecification of the shape parameter. Despite this added imprecision, the method maintains reasonable robustness, even under the conditions of model misspecification presented in Case 2.

In Case 3, robustness is again evaluated under model misspecification, specifically by examining the effect of using an incorrect link function. In this case, the Eyring link function is assumed in the sampling model, while the Arrhenius link function is used in the analysis. This setup aims to explore the robustness of the method when the link function does not provide a good fit, a situation commonly encountered in practice.

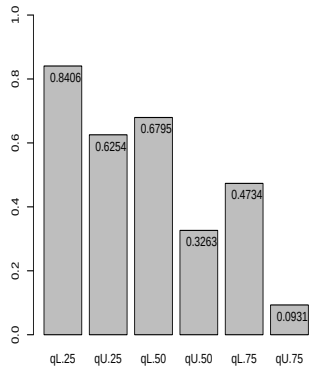
As discussed in Section 2.5, the Eyring link function serves as an alternative to the Arrhenius link function for modeling the accelerating parameter concerning temperature. In this simulation, we used a scenario similar to Case 1, where the model assumptions were entirely correct, except for replacing the link function with the Eyring link function in the sampling model. Table 3.7 and Figures A.3-A.5 present the outcomes of this simulation. When compared with cases where the model assumptions are fully correct (as shown in Table 3.5 and Table 3.6), the results reveal only a slight increase in imprecision due to the misspecification of the link function.

The similarity of these results is also apparent when comparing Figures 3.5-3.7, Figures A.1-A.3, and Figures A.4-A.6 for the quartiles $qL0.25$, $qU0.25$, $qL0.50$, $qU0.50$, $qL0.75$, and $qU0.75$ of the NPI lower and upper survival functions for $1-q = 0.75, 0.50, 0.25$. These results confirm that the proposed approach retains robustness in predictive inference, even when subjected to model misspecification in the link function.

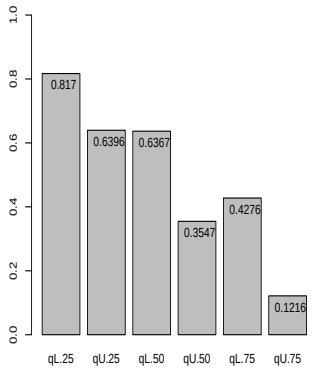
The main findings from the simulation studies demonstrate that the generated future observations at the normal stress level successfully exceeded the quartiles in the expected proportions, confirming the robustness of the proposed predictive inference method. This approach offers good performance with reasonable imprecision when model assumptions hold, particularly within intervals defined by $\underline{\gamma}$ and $\overline{\gamma}$.

m_1m_0		$n=10$		$n=50$		$n=100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.6254	0.8406	0.7081	0.7906	0.7268	0.7824
	0.50	0.3263	0.6795	0.4301	0.5761	0.4533	0.5598
	0.25	0.0931	0.4734	0.1692	0.3488	0.1919	0.3182
0.05	0.75	0.6396	0.8170	0.7163	0.7809	0.7327	0.7774
	0.50	0.3547	0.6367	0.4440	0.5588	0.465	0.5453
	0.25	0.1216	0.4276	0.1861	0.3254	0.2038	0.3028
0.10	0.75	0.6474	0.8062	0.7213	0.7756	0.7362	0.7737
	0.50	0.3706	0.6145	0.4520	0.5495	0.4706	0.5377
	0.25	0.1364	0.4013	0.1958	0.3128	0.2119	0.2931
m_2m_0		$n=10$		$n=50$		$n=100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.6064	0.8530	0.6989	0.7965	0.7218	0.7884
	0.50	0.3121	0.6929	0.4213	0.5817	0.4444	0.5641
	0.25	0.0835	0.4736	0.1681	0.3525	0.1917	0.3193
0.05	0.75	0.6266	0.8321	0.7082	0.7857	0.7294	0.7816
	0.50	0.3424	0.6483	0.4375	0.5619	0.4599	0.5515
	0.25	0.1151	0.4315	0.1854	0.3307	0.2048	0.3050
0.10	0.75	0.6370	0.8192	0.7146	0.7801	0.7319	0.7774
	0.50	0.3616	0.6264	0.4479	0.5513	0.4660	0.5434
	0.25	0.1327	0.4102	0.1953	0.3194	0.2114	0.2962
$\underline{\gamma}$ and $\overline{\gamma}$		$n=10$		$n=50$		$n=100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.5853	0.8862	0.6823	0.8148	0.7100	0.7973
	0.50	0.2319	0.7460	0.3822	0.6156	0.4164	0.5897
	0.25	0.0289	0.5478	0.1252	0.3952	0.1594	0.3512
0.05	0.75	0.6115	0.8625	0.6957	0.8020	0.7199	0.7901
	0.50	0.2747	0.6954	0.4059	0.5909	0.4346	0.5717
	0.25	0.0516	0.4851	0.1467	0.3647	0.1776	0.3317
0.10	0.75	0.6251	0.8508	0.7031	0.7954	0.7240	0.7867
	0.50	0.2971	0.6686	0.4175	0.5804	0.4430	0.5615
	0.25	0.0675	0.4571	0.1582	0.3508	0.187 0	0.3207

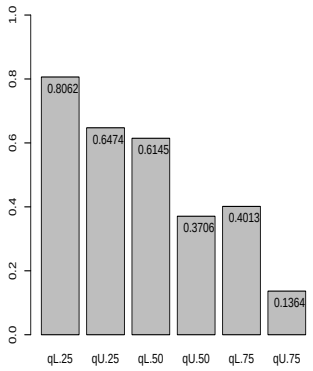
Table 3.5: Proportion of runs with future observation greater than the quartiles, Case 1.



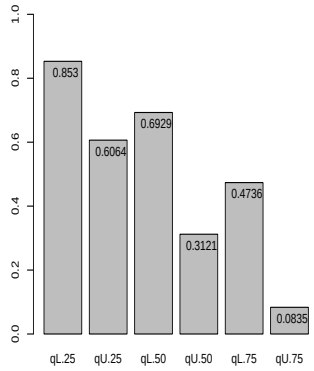
(a) m_1 to $m_0(0.01)$



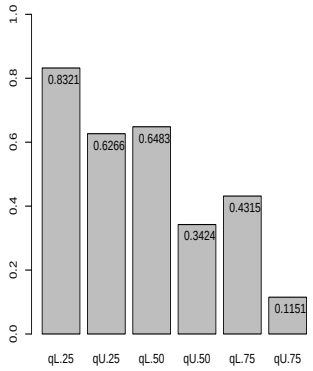
(b) m_1 to $m_0(0.05)$



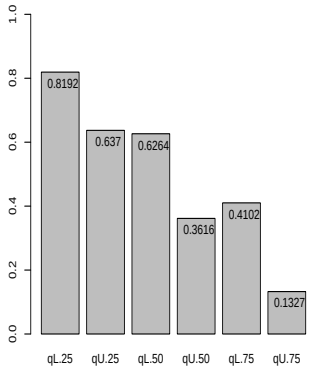
(c) m_1 to $m_0(0.10)$



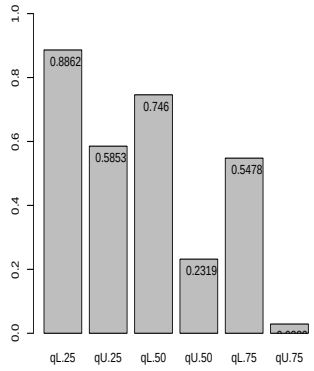
(d) m_2 to m_0 (0.01)



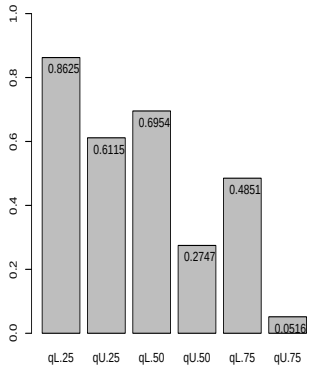
(e) m_2 to m_0 (0.05)



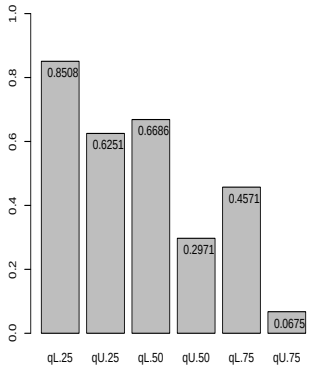
(f) m_2 to m_0 (0.10)



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

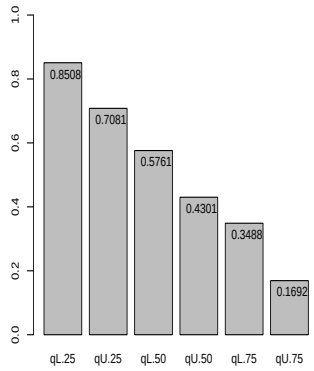


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

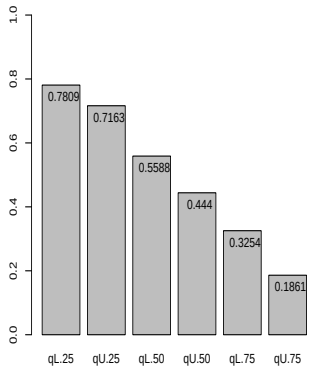


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

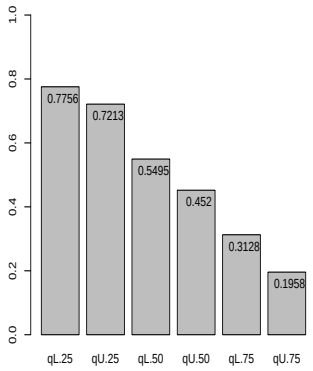
Figure 3.5: Proportion of runs with future observation greater than the quartiles, Case 1, $n = 10$.



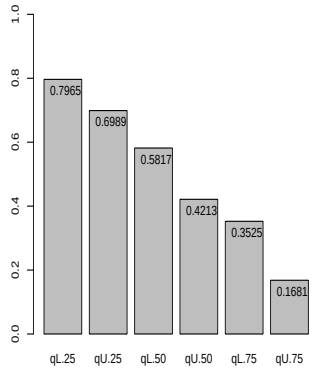
(a) m_1 to m_0 (0.01)



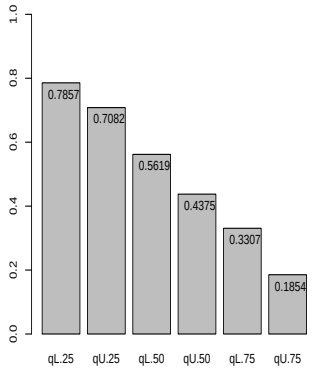
(b) m_1 to m_0 (0.05)



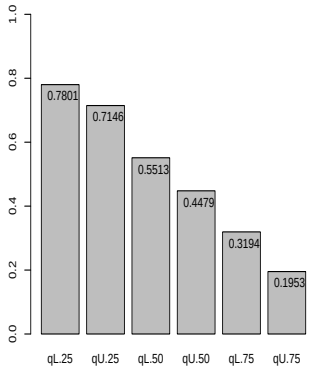
(c) m_1 to m_0 (0.10)



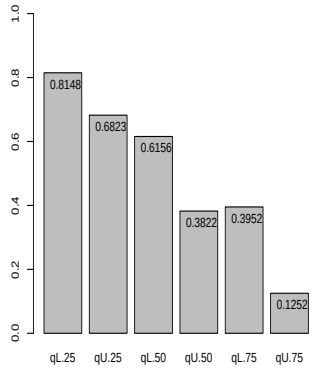
(d) m_2 to m_0(0.01)



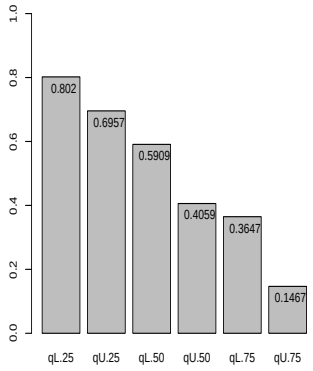
(e) m_2 to m_0(0.05)



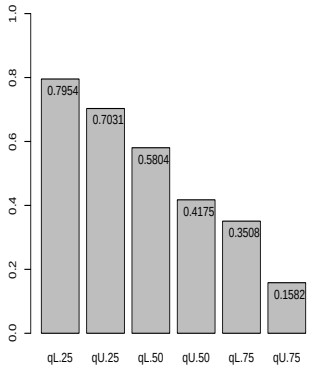
(f) m_2 to m_0(0.10)



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

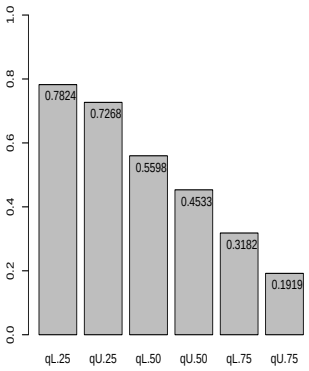


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

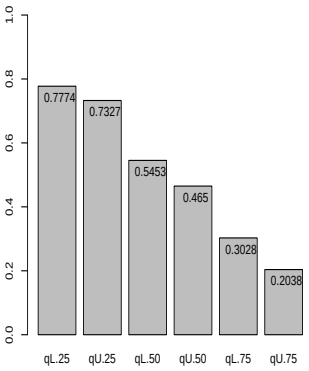


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

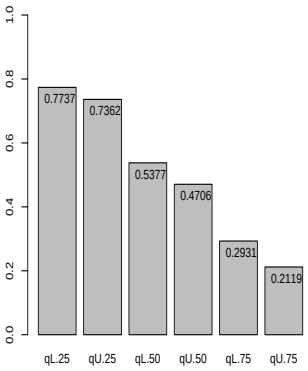
Figure 3.6: Proportion of runs with future observation greater than the quartiles, Case 1, $n = 50$.



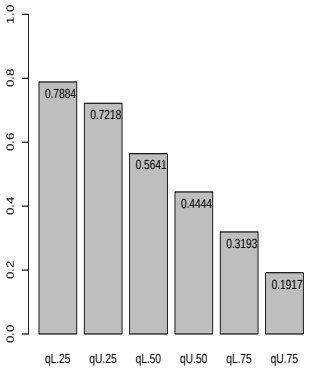
(a) m_1 to m_0 (0.01)



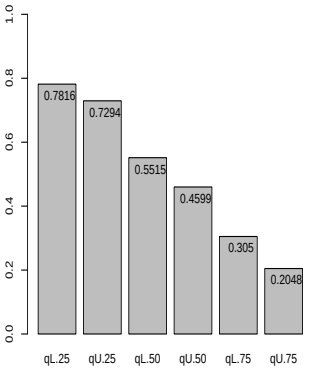
(b) m_1 to m_0 (0.05)



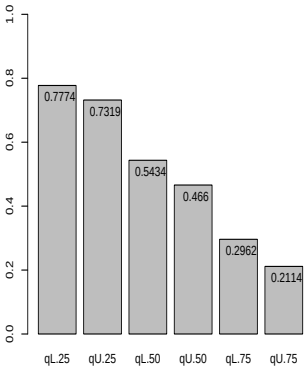
(c) m_1 to m_0 (0.10)



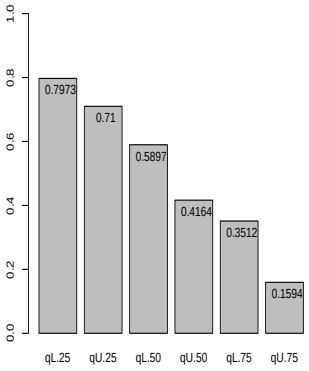
(d) m_2 to m_0 (0.01)



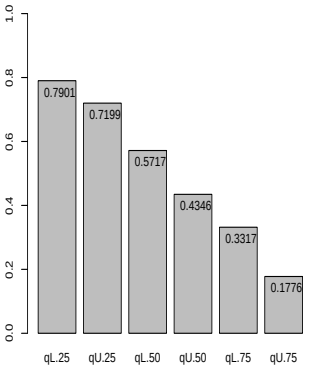
(e) m_2 to m_0 (0.05)



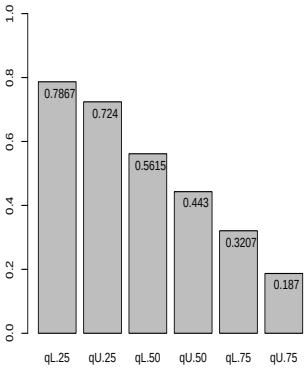
(f) m_2 to m_0 (0.10)



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



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



(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

Figure 3.7: Proportion of runs with future observation greater than the quartiles, Case 1, $n = 100$.

m_1m_0		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.6212	0.8387	0.7881	0.7051	0.7244	0.7805
	0.50	0.3237	0.6767	0.5746	0.4264	0.4492	0.5585
	0.25	0.0909	0.4745	0.1664	0.3466	0.1899	0.3148
0.05	0.75	0.6352	0.8159	0.7142	0.7776	0.7303	0.7750
	0.50	0.3505	0.6348	0.4396	0.5574	0.4604	0.5430
	0.25	0.1184	0.4245	0.1835	0.322	0.2030	0.2990
0.10	0.75	0.6442	0.805	0.7176	0.7728	0.7328	0.7715
	0.50	0.3670	0.6106	0.4488	0.5482	0.4678	0.5353
	0.25	0.1344	0.4001	0.1944	0.3106	0.2098	0.2905
m_2m_0		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.5999	0.8537	0.6936	0.7954	0.7178	0.7872
	0.50	0.3081	0.6914	0.4191	0.5810	0.4407	0.5641
	0.50	0.0814	0.4743	0.1657	0.351	0.1890	0.3174
0.05	0.75	0.6216	0.8314	0.7043	0.7843	0.7263	0.7800
	0.50	0.3396	0.6478	0.4346	0.5612	0.4560	0.5503
	0.25	0.1136	0.4309	0.1835	0.3280	0.2037	0.3032
0.10	0.75	0.6327	0.8182	0.7098	0.7776	0.7302	0.7752
	0.50	0.3584	0.625	0.4438	0.5506	0.4640	0.5420
	0.25	0.1312	0.4082	0.1939	0.3177	0.2105	0.2940
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.5755	0.8860	0.6763	0.8124	0.7053	0.7956
	0.50	0.2258	0.7465	0.3763	0.6136	0.4109	0.5887
	0.25	0.0278	0.546	0.1250	0.3908	0.1581	0.3470
0.05	0.75	0.6024	0.8608	0.6911	0.7991	0.7150	0.7882
	0.50	0.2683	0.6949	0.4005	0.5900	0.4294	0.5705
	0.25	0.0504	0.484	0.1454	0.3616	0.1756	0.3270
0.10	0.75	0.6171	0.8492	0.6980	0.7928	0.7204	0.7847
	0.50	0.2911	0.6679	0.4125	0.5775	0.4390	0.5603
	0.25	0.0658	0.4503	0.1567	0.3459	0.1850	0.3167

Table 3.6: Proportion of runs with future observation greater than the quartiles, Case 2, $\beta = 3$.

m_1m_0		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.5643	0.8721	0.6584	0.7932	0.6886	0.7790
	0.50	0.2252	0.7285	0.3534	0.5784	0.3838	0.5465
	0.25	0.0228	0.5163	0.0974	0.3476	0.1228	0.3019
0.05	0.75	0.5836	0.8386	0.6742	0.7778	0.6976	0.7690
	0.50	0.2618	0.6688	0.3782	0.549	0.3987	0.5271
	0.25	0.0479	0.4567	0.1195	0.3130	0.1406	0.2772
0.10	0.75	0.5956	0.82	0.6823	0.7687	0.7020	0.7641
	0.50	0.2876	0.6355	0.3924	0.5349	0.4090	0.5184
	0.25	0.0647	0.4225	0.1303	0.2970	0.1501	0.2678
m_2m_0		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.5494	0.8937	0.6570	0.8100	0.6893	0.7929
	0.50	0.2152	0.7557	0.3554	0.5987	0.3866	0.5697
	0.25	0.0197	0.5246	0.1061	0.3702	0.1349	0.3233
0.05	0.75	0.5736	0.8571	0.6727	0.7911	0.6999	0.7808
	0.50	0.2610	0.6938	0.3831	0.5692	0.4054	0.5495
	0.25	0.0468	0.4749	0.1315	0.3377	0.1549	0.3011
0.10	0.75	0.5863	0.8401	0.6809	0.7826	0.705	0.7741
	0.50	0.2862	0.6619	0.3970	0.5544	0.4160	0.5368
	0.25	0.0673	0.4434	0.1435	0.3211	0.1637	0.2874
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.4814	0.9180	0.6065	0.8241	0.6473	0.7988
	0.50	0.0954	0.8057	0.2644	0.6249	0.3176	0.5856
	0.25	0.0013	0.6138	0.0462	0.4039	0.0779	0.3440
0.05	0.75	0.5183	0.8874	0.6304	0.8049	0.6646	0.7865
	0.50	0.1466	0.7398	0.3050	0.5887	0.3457	0.5601
	0.25	0.0074	0.5336	0.0705	0.3613	0.0982	0.3144
0.10	0.75	0.5432	0.8704	0.6405	0.7942	0.6721	0.7791
	0.50	0.1803	0.7032	0.3258	0.5715	0.3596	0.5452
	0.25	0.0156	0.4884	0.0848	0.3383	0.1104	0.2986

Table 3.7: Proportion of runs with future observation greater than the quartiles, Case 3.

However, under model misspecification, particularly in cases of significant imprecision from the true model, the method exhibits increased imprecision, indicating either a potential model fit issue or the need for additional data to enhance predictive accuracy. As sample size increases across different strategies, the imprecision level generally becomes smaller, emphasizing the role of larger sample sizes in achieving more precise predictions. Higher significance levels in the likelihood test also contribute to narrower intervals for $\underline{\gamma}$ and $\overline{\gamma}$, further reducing imprecision. These results suggest that while the proposed method retains robustness under minor misspecifications, substantial imprecision necessitate for model reassessment or supplementary data to maintain less imprecise predictive inferences.

3.6 The model with different shape parameters

In this section, the model described in Section 2.4 is extended to allow different shape parameters at each stress level. Here, β_i represents the shape parameter at stress level s_i for $i = 0, 1, 2, \ldots, s-1$. The probability density function (PDF) of the failure time at stress level s_i is given by

$$f_i(t) = \left(\frac{\beta_i}{\theta_i^{\beta_i}}\right) t^{\beta_i - 1} \exp\left[-\left(\frac{t}{\theta_i}\right)^{\beta_i}\right], \quad \text{for } t \geq 0,$$

and the cumulative distribution function (CDF) is

$$F_i(t) = 1 - \exp\left[-\left(\frac{t}{\theta_i}\right)^{\beta_i}\right], \quad \text{for } t \geq 0.$$

Consider an $s-1$ step SSALT experiment under the assumption of the cumulative exposure model, where the lifetime of experimental units follows a Weibull distribution.

At the initial stress level s_0, the cumulative distribution function (CDF) of the lifetime of experimental units is given by $F_0(t)$ for $t \in (\tau_0, \tau_1)$. Failures occur according to this distribution during this time interval. When the stress level increases to stress level s_1, the cumulative distribution function (CDF) for the lifetime of experimental units is given by $F_1(t - h_1)$ for $t \in (\tau_1, \tau_2)$, where h_1 is the shift parameter. This shift parameter h_1 is determined by solving the equation $F_1(\tau_1 - h_1) = F_0(\tau_1)$, ensuring continuity between the stress levels. Consequently, h_1 is obtained by solving the equation

$$1-\exp\left[-\left(\frac{\tau_1-h_1}{\theta_1}\right)^{\beta_1}\right]=1-\exp\left[-\left(\frac{\tau_1}{\theta_0}\right)^{\beta_0}\right]$$

leading to $h_1=-((\frac{\tau_1}{\theta_0})^{\beta_0})^{\frac{1}{\beta_1}}\theta_1+\tau_1$. Then, the CDF at stress level s_1 becomes:

$$F_1(t)=F_1(t-h_1)=1-\exp\left[-\left(\frac{t-\tau_1}{\theta_1}+\left(\frac{\tau_1}{\theta_0}\right)^{\frac{\beta_0}{\beta_1}}\right)^{\beta_1}\right],\quad \text{for } \tau_1\le t\le\tau_2.$$

Similarly, at the stress level s_2, the CDF is $F_2(t-h_2)$ for $t\in(\tau_2,\tau_3)$, where h_2 is the solution of $F_2(\tau_2-h_2)=F_1(\tau_2-h_1)$. Consequently, h_2 is obtained by solving the equation

$$1-\exp\left[-\left(\frac{\tau_2-h_2}{\theta_2}\right)^{\beta_2}\right]=1-\exp\left[-\left(\frac{t-\tau_1}{\theta_1}+\left(\frac{\tau_1}{\theta_0}\right)^{\frac{\beta_0}{\beta_1}}\right)^{\beta_1}\right].$$

Therefore, $h_2=-((\frac{\tau_2-\tau_1}{\theta_1}+(\frac{\tau_1}{\theta_0})^{\frac{\beta_0}{\beta_1}})^{\frac{\beta_1}{\beta_2}})\theta_2+\tau_2$. Then, the CDF of s_2 becomes

$$F_2(t)=F_2(t-h_2)=1-\exp\left[-\left(\tfrac{t-\tau_2}{\theta_2}+\left(\left(\tfrac{\tau_2-\tau_1}{\theta_1}+\left(\tfrac{\tau_1}{\theta_0}\right)^{\frac{\beta_0}{\beta_1}}\right)^{\frac{\beta_1}{\beta_2}}\right)\right)^{\beta_2}\right],\quad \text{for}\quad \tau_2\le t\le\tau_3$$

The shifting parameter is $h_i=(\tau_i-\tau_i^*)$, where τ_i^* has the following general expression:

$$\begin{aligned}
\tau_1^*&=\left(\frac{\tau_1}{\theta_0}\right)^{\frac{\beta_0}{\beta_1}}\theta_1,\\
\tau_2^*&=\left(\frac{\tau_2-\tau_1}{\theta_1}+\left(\frac{\tau_1}{\theta_0}\right)^{\frac{\beta_0}{\beta_1}}\right)^{\frac{\beta_1}{\beta_2}}\theta_2,\\
\tau_3^*&=\left(\frac{\tau_3-\tau_2}{\theta_2}+\left(\frac{\tau_2-\tau_1}{\theta_1}+\left(\frac{\tau_1}{\theta_0}\right)^{\frac{\beta_0}{\beta_1}}\right)^{\frac{\beta_1}{\beta_2}}\right)^{\frac{\beta_2}{\beta_3}}\theta_3,
\end{aligned}$$

where $\tau_0^*=0$ and $\tau_i^*$is given by the following expression:

$$\tau_i^*=\theta_i\left(\frac{\tau_i-\tau_{i-1}+\tau_{i-1}^*}{\theta_{i-1}}\right)^{\frac{\beta_{i-1}}{\beta_i}}\quad \text{for}\quad \tau_i\le t\le\tau_{i+1}\tag{3.14}$$

In addition, the CDF of the failure time for an $s-1$ SSALT experiment with a Weibull

distribution is given by

$$F(t) = \begin{cases} 1 - \exp\left[-\left(\frac{t}{\theta_0}\right)^{\beta_0}\right] & \text{if} \quad \tau_0 \le t \le \tau_1 \\ 1 - \exp\left[-\left(\frac{t-\tau_1}{\theta_1} + \left(\frac{\tau_1}{\theta_0}\right)^{\frac{\beta_0}{\beta_1}}\right)^{\beta_1}\right] & \text{if} \quad \tau_1 \le t \le \tau_2 \\ 1 - \exp\left[-\left(\frac{t-\tau_2}{\theta_2} + \left(\frac{\tau_2-\tau_1}{\theta_1} + \left(\frac{\tau_1}{\theta_0}\right)^{\frac{\beta_0}{\beta_1}}\right)^{\frac{\beta_1}{\beta_2}}\right)^{\beta_2}\right] & \text{if} \quad \tau_2 \le t \le \tau_3 \\ 1 - \exp\left[-\left(\frac{t-\tau_3}{\theta_3} + \left(\frac{\tau_3-\tau_2}{\theta_2} + \left(\frac{\tau_2-\tau_1}{\theta_1} + \left(\frac{\tau_1}{\theta_0}\right)^{\frac{\beta_0}{\beta_1}}\right)^{\frac{\beta_1}{\beta_2}}\right)^{\frac{\beta_2}{\beta_3}}\right)^{\beta_3}\right] & \text{if} \quad \tau_3 \le t \le \infty \end{cases} \tag{3.15}$$

Alternatively, the cumulative distribution function for the failure time in an $(s-1)$ step SSALT experiment, assuming a Weibull distribution, can be expressed as:

$$F(t) = 1 - \exp\left[-\left(\frac{t - \tau_i + \tau_i^*}{\theta_i}\right)^{\beta_i}\right], \quad \text{for} \quad \tau_i \le t \le \tau_{i+1}. \tag{3.16}$$

The corresponding probability density function is given by:

$$f(t) = \left(\frac{\beta_i}{\theta_i^{\beta_i}}\right)(t - (\tau_i - \tau_i^*))^{\beta_i - 1} \exp\left[-\left(\frac{t - \tau_i + \tau_i^*}{\theta_i}\right)^{\beta_i}\right], \quad \text{for} \quad \tau_i \le t \le \tau_{i+1}. \tag{3.17}$$

3.6.1 Data transformation with different shape parameters

In this section, the transformation method is applied to the model with varying shape parameters, as described in Section 3.3.2. Specifically, failure times, denoted by $t_i^{(m)\to(0,0)}$, are transformed from a higher stress level s_i under strategy m to the normal stress level s_0 under the lowest strategy $m = 0$. This transformation to the normal stress level is typically essential for analyzing failure times under standard conditions. The transformed time $t_i^{(m)\to(0,0)}$ is therefore given by

$$t_i^{(m)\to(1^*,0^*)} = \left[\left(\frac{t_i^m - \tau_i^m + \tau_i^{*m}}{\theta_i^m}\right)^{\frac{\beta_i^m}{\beta_0^1}}\right]\theta_0^0, \quad \text{for} \quad 0 \le t_i^{(m)\to(0,0)} \le \infty, \tag{3.18}$$

where τ_i^{m*} is given by the following expression:

$$\tau_i^{m*} = \left[\theta_i\left(\frac{\tau_i^{(m)} - \tau_{i-1}^{(m)} + \tau_{i-1}^{m*}}{\theta_{i-1}^{m*}}\right)^{\frac{\beta_{i-1}^{m*}}{\beta_i^{m*}}}\right], \quad \text{for} \quad \tau_i^m \le t \le \tau_{i+1}^m. \tag{3.19}$$

When failure times are transformed from a higher stress level to stress level s_i, the transformed time $t_i^{\to 0}$ is given by

$$t_i^{\to 0} = \left[\left(\frac{t_i - \tau_i + \tau_i^*}{\theta_i} \right)^{\frac{\beta_i}{\beta_0}} \right] \theta_0, \quad \text{for} \quad 0 \leq t_i^{\to 0} \leq \infty. \tag{3.20}$$

When transforming failure times from a higher stress level, $t_0^{(1)\to(0,0)}$ is given by

$$t_0^{(1)\to(0,0)} = \left(\frac{t_0^1}{\theta_0^1} \right)^{\frac{\beta_0^1}{\beta_0^0}} \theta_0^0, \quad \text{for} \quad 0 \leq t_0^{(1)\to(0,0)} \leq \infty. \tag{3.21}$$

The failure times $t_1^{(1)\to(0,0)}$ are transformed as follows:

$$t_1^{1\to(0,0)} = \left(\frac{t_1^1 - \tau_1^1}{\theta_1^1} + \left(\frac{\tau_1^1}{\theta_0^1} \right)^{\frac{\beta_0^1}{\beta_1^1}} \right)^{\frac{\beta_1^1}{\beta_0^0}} \theta_0^0, \quad \text{for} \quad 0 \leq t_1^{1\to(0,0)} \leq \infty. \tag{3.22}$$

Similarly, the failure times $t_2^{1\to(0,0)}$ are transformed as:

$$t_2^{1\to(0,0)} = \left(\frac{t_2^1 - \tau_2^1}{\theta_2^1} + \left(\frac{\tau_2^1 - \tau_1^1}{\theta_1^1} + \left(\frac{\tau_1^1}{\theta_0^1} \right)^{\frac{\beta_0^1}{\beta_1^1}} \right)^{\frac{\beta_1^1}{\beta_2^1}} \right)^{\frac{\beta_2^1}{\beta_0^1}} \theta_0^0, \quad \text{for} \quad 0 \leq t_2^{1\to(0,0)} \leq \infty. \tag{3.23}$$

3.7 Illustrative examples with different shape parameters

This section presents an example to illustrate the proposed method outlined in Section 3.3, demonstrating the application of varying shape parameters across different stress levels.

Example 3.7.1 This example consists of four cases to demonstrate the impact of different assumptions on shape parameters and link functions. In Case 1, shape parameters are assumed to be constant and known for each strategy, with the Arrhenius link function applied to connect scale parameters across all stress levels in both the simulated data and the analysis. Case 2 considers constant but unknown shape parameters, which are estimated during the analysis. Case 3, the Arrhenius link function is replaced by the Eyring link function to examine model performance when the assumed link function does not fit well. This change allows investigation of the resulting interval $[\underline{\gamma}, \overline{\gamma}]$ and the

corresponding lower and upper NPI survival functions at the normal stress level. In Case 4, all shape parameters are set to 1, assuming an exponential model, with the Eyring link function employed to further explore outcomes under misspecification.

Three data sets are simulated under three different experimental strategies. In the first strategy, the experiment is set with $k_0^0 = 300$, $\theta_0 = 7000$, $\gamma = 5000$, and $\beta_0^0 = 2$. This strategy represents the normal use stress level, and twenty observations are generated from the Weibull distribution. The Arrhenius link function is used to connect scale parameters across all strategies, with the accelerating parameter set to $\gamma = 5000$.

Strategies	Stress level	Data sets	Failures times
m_0	s_0^0	t_0^0	1997.16, 2726.58, 2752.79, 4841.75, 5629.64 5799.14, 6302.59, 7921.43, 7951.54, 9040.98
m_2	s_0^1	t_0^1	231.63
m_2	s_1^1	t_1^1	304.48, 307.45, 317.95, 318.85, 348.18
m_2	s_2^1	t_2^1	360.79, 367.25, 372.93, 381.58
m_3	s_0^2	t_0^2	26.64, 90.29
m_3	s_1^2	t_1^2	100.18, 108.72, 116.36
m_3	s_2^2	t_2^2	130.035, 132.35, 132.82, 135.75, 145.07

Table 3.8: Simulated data of Example 3.7.1

In the second strategy, the accelerated experiment is set with increased temperature levels $k_0^1 = 350$, $k_1^1 = 400$, and $k_2^1 = 450$ Kelvin for stress levels s_0^1, s_1^1, and s_2^1, respectively. The stress level increases from s_0^1 to s_1^1 at $\tau_1^1 = 300$, and from s_1^1 to s_2^1 at $\tau_2^1 = 350$. The shape parameters for this strategy are set as $\beta_0^1 = 1.9$, $\beta_1^1 = 1.7$, and $\beta_2^1 = 1.5$.

In the third strategy, the accelerated experiment uses temperature levels $k_0^2 = 380$, $k_1^2 = 420$, and $k_2^2 = 460$ Kelvin for stress levels s_0^2, s_1^2, and s_2^2, respectively. The stress level increases from s_0^2 to s_1^2 at $\tau_1^2 = 100$ and from s_1^2 to s_2^2 at $\tau_2^2 = 130$. The shape parameters are set to $\beta_0^2 = 1.95$, $\beta_1^2 = 1.60$, and $\beta_2^2 = 1.40$.

The generated failure times for each strategy are shown in Table 3.8. For this example (Example 3.7.1), three data sets are generated, each with a sample size of $n = 10$ across all strategies, as in Example 3.4.1. These strategies include two accelerating strategies

Cases	Significance Level	0.01		0.05		0.10	
	strategies	$\underline{\gamma}$	$\overline{\gamma}$	$\underline{\gamma}$	$\overline{\gamma}$	$\underline{\gamma}$	$\overline{\gamma}$
Case 1	m_1, m_0	3941.86	5541.91	4155.23	5360.95	4260.66	5268.53
	m_2, m_0	3965.32	5298.97	4146.49	5151.56	4235.71	5075.87
Case 2	m_1, m_0	2370.54	6502.28	2372.55	6226.90	3064.92	6051.10
	m_2, m_0	1410.29	5935.39	1653.10	5834.07	1855.89	5756.98
Case 3	m_1, m_0	2013.82	6164.18	2034.14	5882.83	2758.95	5693.86
	m_2, m_0	1011.98	5614.96	1249.00	5532.66	1385.33	5452.11
Case 4	m_1, m_0	3766.95	6330.52	4134.82	6136.62	4317.68	6028.17
	m_2, m_0	3780.71	6068.67	4100.31	5894.94	4260.47	5798.39

Table 3.9: $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ for Example 3.7.1.

alongside data collected at the normal stress level.

To analyze the data sets in Table 3.8, the Weibull cumulative exposure model is assumed for the accelerating strategies m_1 and m_2, while the Weibull model is used for the normal stress strategy m_0. A pairwise likelihood ratio test is applied between m_0 and each of m_i (where $i = 1, 2$), identifying intervals $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ within which the null hypothesis is not rejected. The intervals $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ for three significance levels are summarized in Table 3.9.

The transformed data points rely on overall values $[\underline{\gamma}, \overline{\gamma}]$, determined as the minimum and maximum bounds from the pairwise tests. All failure times at the elevated stress strategies are transformed to the normal stress level, resulting in interval-valued failure times at m_0 for both m_1 and m_2.

The NPI lower survival function is derived by the data at the normal stress level combined with the transformed data from the higher strategies m_1 to m_0 and m_2 to m_0 using $\underline{\gamma}$.The lower points of the transformed data represent the pessimistic case which leads to the lower survival function $\underline{S}$. Similarly, The NPI upper survival function is derived by the data at the normal stress level combined with the transformed data from the higher strategies m_1 to m_0 and m_2 to m_0 using $\overline{\gamma}$. The upper points of the transformed data represent the optimistic case which leads to the upper survival function $\overline{S}$.

In Case 1, the shape parameters are assumed to be constant and known. The analysis results indicate that $\hat{\theta}_0^1 = 6056.00$ and $\hat{\gamma}_1 = 4780.22$ for the comparison between the first

Data set	Failures times	$\underline{\gamma}_{0.01}$	$\underline{\gamma}_{0.05}$	$\underline{\gamma}_{0.10}$	$\hat{\gamma}$	$\overline{\gamma}_{0.10}$	$\overline{\gamma}_{0.05}$	$\overline{\gamma}_{0.01}$
t_0^1	231.63	1622.22	1779.56	1852.86	243.35	2956.34	3082.57	3345.54
t_1^1	304.49	2196.48	2418.01	2521.63	530.92	4111.63	4296.87	4684.88
t_1^1	307.46	2276.83	2511.68	2621.77	667.39	4328.64	4529.41	4951.12
t_1^1	317.96	2557.22	2838.31	2970.88	1117.91	5081.80	5336.12	5873.91
t_1^1	318.86	2580.99	2865.98	3000.44	1154.84	5145.38	5404.20	5951.74
t_1^1	348.18	3336.58	3744.75	3938.98	2276.84	7152.05	7551.71	8404.10
t_2^1	360.80	4140.47	4711.34	4986.36	3333.19	9793.54	10420.07	11774.04
t_2^1	367.25	4571.73	5229.41	5547.48	3879.16	11198.43	11944.38	13561.94
t_2^1	372.94	4940.61	5671.91	6026.44	4339.52	12388.65	13234.92	15073.68
t_2^1	381.59	5485.06	6324.12	6731.99	5010.61	14130.50	15122.50	17282.39
t_0^2	26.65	452.04	519.98	552.71	744.76	1120.46	1193.60	1350.93
t_0^2	90.29	1485.67	1708.95	1816.53	2447.71	3682.49	3922.87	4439.94
t_0^2	100.18	1649.66	1897.78	2017.33	3094.10	4091.82	4359.16	4934.25
t_1^2	108.72	2036.27	2351.18	2503.36	16805.24	5178.02	5526.38	6278.06
t_1^2	116.37	2367.13	2738.93	2918.87	26738.08	6103.24	6520.28	7421.55
t_2^2	130.04	2933.80	3402.71	3630.05	42574.10	7682.96	8216.91	9372.71
t_2^2	132.35	3138.49	3646.15	3892.59	46575.89	8312.24	8897.36	10165.68
t_2^2	132.83	3179.73	3695.17	3945.46	47377.77	8438.68	9034.05	10324.92
t_2^2	135.76	3429.41	3991.86	4265.32	52205.16	9201.94	9859.03	11285.55
t_2^2	145.07	4177.01	4879.21	5221.48	66448.72	11469.96	12309.11	14135.41

Table 3.10: Tranformation of Case 1, Example 3.7.1.

and second strategies, and $\hat{\theta}_0^2 = 5671.22$ and $\hat{\gamma}_2 = 4671.58$ for the comparison between the first and third strategies. The scale parameters from the initial analysis are applied to transform all data, rather than using the scale parameters estimated during the test. Table 3.10 displays the transformed data using $\hat{\gamma}$, $\underline{\gamma}$, and $\overline{\gamma}$ across different significance levels. As noted, the minimum and maximum values of $\underline{\gamma}$ and $\overline{\gamma}$ across all strategies are used to transform the failure times at higher stress levels for Case 1.The NPI lower and upper survival function is presented in Figure 3.8a.

In Case 2, the shape parameters are assumed to be constant but unknown. The analysis results indicate that $\hat{\theta}_0^1 = 6265.02$, $\hat{\gamma}_1 = 5133.40$, with shape parameters $\hat{\beta}_0^0 = 2.66$, $\hat{\beta}_0^1 = 3.88$, $\hat{\beta}_1^1 = 0.70$, and $\hat{\beta}_2^1 = 1.20$ for the first and second strategies. For the first and third strategies, the estimates are $\hat{\theta}_0^1 = 6999.99$, $\hat{\gamma}_2 = 5000.00$, with shape

parameters $\hat{\beta}_0^0 = 2.68$, $\hat{\beta}_0^2 = 1.74$, $\hat{\beta}_1^2 = 0.22$, and $\hat{\beta}_2^2 = 2.70$. The scale parameter from the initial analysis is used to transform all data, rather than relying on the scale parameters estimated during the pairwise tests. The NPI lower and upper survival functions for Case 2 are shown in Figure 3.8b.

In Case 3, the shape parameters are assumed to be constant but unknown, and the Arrhenius link function is replaced by the Eyring link function to introduce model misspecification. This setup is used to examine the performance of the proposed method under conditions where the assumed link function does not fit well.

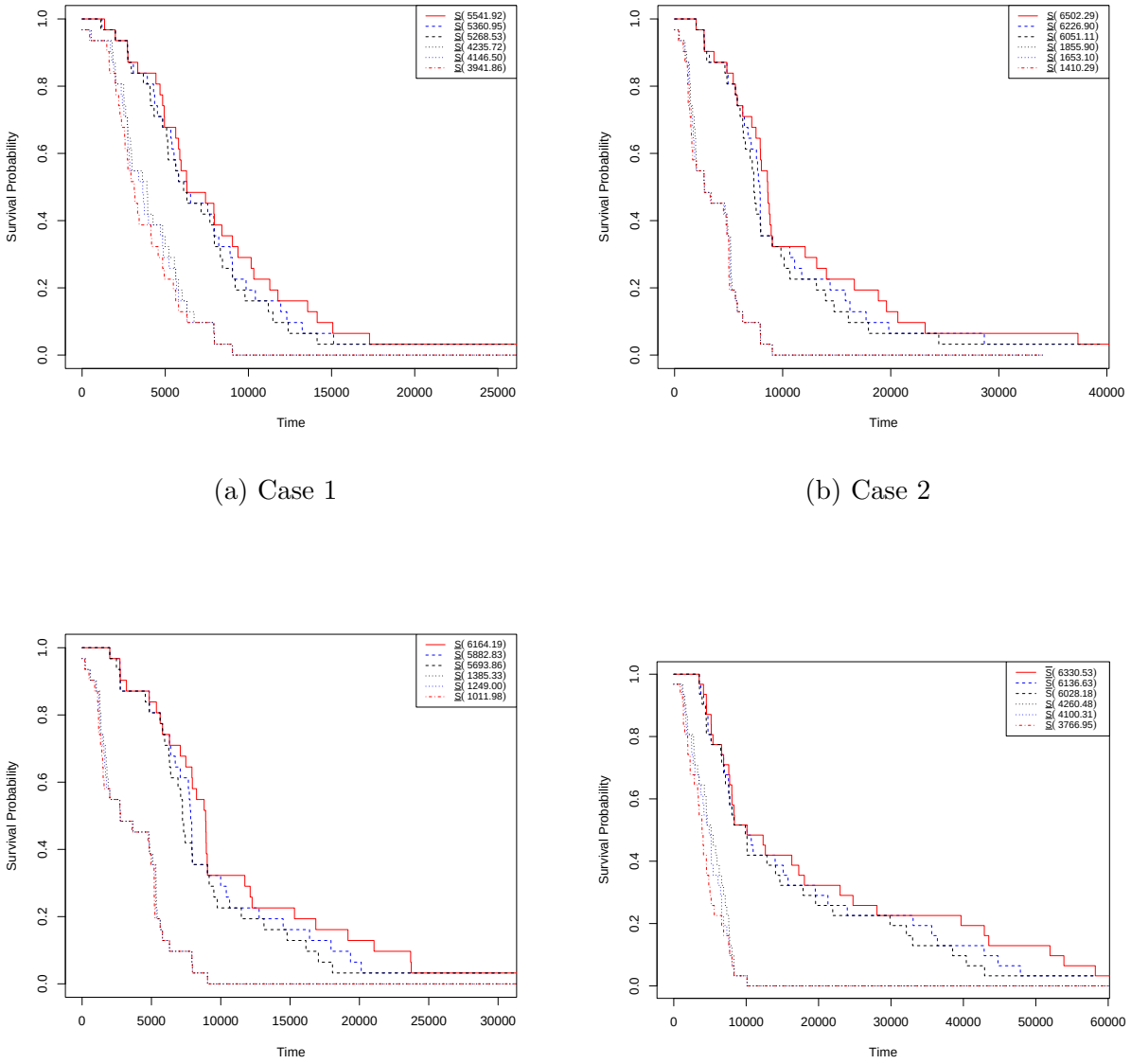
The analysis yields the following results: for the comparison between the first and second strategies, $\hat{\theta}_0^1 = 6129.22$, $\hat{\gamma}_1 = 4613.37$, with shape parameters $\hat{\beta}_0^0 = 2.62$, $\hat{\beta}_0^1 = 3.55$, $\hat{\beta}_1^1 = 0.66$, and $\hat{\beta}_2^1 = 1.37$. For the first and third strategies, the estimates are $\hat{\theta}_0^2 = 6999.99$, $\hat{\gamma}_2 = 4999.99$, with shape parameters $\hat{\beta}_0^0 = 2.68$, $\hat{\beta}_0^2 = 2.08$, $\hat{\beta}_1^2 = 0.19$, and $\hat{\beta}_2^2 = 1.90$. The NPI lower and upper survival functions for Case 3 are shown in Figure 3.8c.

In Case 4, an additional level of model misspecification is introduced by setting all shape parameters to 1, effectively assuming an exponential model, while using the Eyring link function. This case is intended to assess the performance of the proposed method under more pronounced misspecification conditions, specifically with a cumulative exposure model that differs significantly from the true underlying distribution.

The results of the analysis for the first and second strategies are $\hat{\theta}_0 = 7814.97$ and $\hat{\gamma}_1 = 5228.79$. For the first and third strategies, the estimates are $\hat{\theta}_0 = 7790.73$ and $\hat{\gamma}_2 = 5074.71$.

The outcomes of Case 4 highlight increased imprecision due to the exponential assumption, especially in the transformed data at the normal stress level, where intervals become wider. The NPI lower and upper survival functions for this case, illustrated in Figure 3.8d, reflect the added variability from this model misspecification.

To compare the four cases in Example 3.7.1, consider the resulting intervals for γ and the corresponding NPI lower and upper survival functions. In Case 1, where the shape parameters are constant and known, the intervals for γ are relatively narrower, as shown in Table 3.9. This results in less imprecision between the NPI lower and upper survival functions, reflecting a good model fit. In Case 2, where the shape parameters



(a) Case 1

(b) Case 2

(c) Case 3

(d) Case 4

Figure 3.8: NPI lower and upper survival functions for all four cases in Example 3.7.1, with red, blue, and black lines representing significance levels 0.01, 0.05, and 0.10, respectively.

are unknown and estimated, the intervals widen, increasing the imprecision in the NPI survival functions due to the additional uncertainty in parameter estimation. Case 3, which introduces the Eyring link function to demonstrate model misspecification, results in even wider intervals for γ and increased imprecision in the survival functions, indicating the effect of an incorrect model assumption. Finally, Case 4 assumes an exponential model, also a form of misspecification. This case yields the widest intervals and the

most substantial imprecision in the NPI survival functions, highlighting the sensitivity of the method to misspecification. Overall, the width of the intervals and the degree of imprecision increase from Case 1 through Case 4, illustrating the method's robustness in capturing uncertainty under different model assumptions and parameter settings.

3.8 Simulation studies with different shape parameters

In this section, simulation studies are performed to evaluate the performance of the proposed method, which incorporates distinct shape parameters at each stress level. The simulations utilize three SSALT (Step-Stress Accelerated Life Testing) data strategies, following the experimental settings outlined in Example 3.7.1. Three separate data sets are generated to represent each strategy.

In the first strategy, the experiment is set at a normal-use stress level with $k_0^0 = 300$ Kelvin and a scale parameter $\theta_0^0 = 7000$. Ten observations are generated from the Weibull distribution under these conditions. The Arrhenius link function is used to link the scale parameters in all strategies, with an acceleration parameter $\gamma = 5000$.

In the second strategy, an accelerated experiment is conducted with increased temperature levels: $k_0^1 = 350$, $k_1^1 = 400$, and $k_2^1 = 450$ Kelvin for stress levels s_0^1, s_1^1, and s_2^1, respectively. The stress level transitions from s_0^1 to s_1^1 at time $\tau_1^1 = 300$, and from s_1^1 to s_2^1 at time $\tau_2^1 = 350$. The corresponding scale parameters for this strategy are $\theta_0^1 = 647.23$, $\theta_1^1 = 108.52$, and $\theta_2^1 = 8.72$.

In the third strategy, the experiment further accelerates with temperature levels $k_0^2 = 380$, $k_1^2 = 420$, and $k_2^2 = 460$ Kelvin for stress levels s_0^2, s_1^2, and s_2^2, respectively. The stress level increases from s_0^2 to s_1^2 at time $\tau_1^2 = 100$, and from s_1^2 to s_2^2 at time $\tau_2^2 = 130$. The corresponding scale parameters assumed in this strategy are $\theta_0^2 = 209.53$, $\theta_1^2 = 59.84$, and $\theta_2^2 = 21.25$.

This simulation was repeated 10,000 times, generating data from the proposed model with varying sample sizes at each strategy, specifically $n = 20$, 50, and 100, and using significance levels of 0.01, 0.05, and 0.10. The performance of the simulation is assessed by simulating a future observation at the baseline stress level, m_0. This assessment evaluates whether the future observation at the normal stress level is indistinguishable from the

transformed data at higher stress levels and from the data at the normal stress level. Performance is determined by examining if the future observation surpasses the quartiles of NPI lower and upper survival functions for $1 - q = 0.75$, 0.50, and 0.25.

To quantify performance, the proportions of instances where the future observation exceeded the quartiles of NPI lower and upper survival functions are calculated. Ideal performance is indicated when the first, second, and third quartiles of the lower NPI survival functions exceed 0.75, 0.50, and 0.25, respectively. Similarly, the upper NPI survival function quartiles are expected to remain below 0.75, 0.50, and 0.25.

In Case 1, the scenario explained in Example 3.7.1 is considered, where the shape parameters are assumed to be constant and known. Table 3.11 and Figures A.7–A.9 present the simulation outcomes for varying significance levels ($\alpha = 0.01, 0.05, 0.10$) and sample sizes ($n = 20, 50, 100$). In these figures, the first, second, and third quartiles are denoted as $qL0.25$, $qU0.25$, $qL0.50$, $qU0.50$, $qL0.75$, and $qU0.75$, representing the NPI lower and upper survival functions, respectively. These quartiles offer insight into the alignment of future observations with the predicted intervals, providing key performance indicators. The results demonstrate that the imprecision between the lower and upper survival functions reduces, as shown in Figure A.9. Overall, these findings indicate that the proposed method achieves suitable predictive inference when the model assumptions hold true.

The results show satisfactory performance across all settings, as the proportions of future observations consistently reach the expected thresholds of 0.75, 0.50, and 0.25 for the lower NPI survival function quartiles, while remaining below these levels for the upper quartiles. Variations in proportions between the lower and upper survival function quartiles indicate levels of imprecision in the predictive inferences, with larger differences pointing to greater imprecision in predictions.

In general, a larger sample size reduces imprecision, resulting in more stable quartile proportions. For instance, with a larger sample size ($n = 100$), quartile proportions more closely match the expected values than those with smaller samples ($n = 20$), indicating reduced imprecision in predictive inferences. This improvement is illustrated in Figure A.9, where the quartiles narrow and align more closely with theoretical expectations.

Data from strategies at higher stress levels, such as m_2, show greater imprecision compared to those at lower stress levels like m_1 and m_0. The larger differences in quartile

proportions observed at higher stress levels suggest that predictive inferences are less accurate in these cases, likely due to the increased variability inherent at elevated stress levels. Furthermore, the overall results for the lower and upper γ parameters reveal greater imprecision than those obtained from individual strategies m_1 and m_2. This pattern highlights the influence of stress levels on predictive accuracy, with higher stress levels generally contributing to increased imprecision in predictive inferences.

The effect of the significance level (α) on predictive precision is less substantial than that of sample size or stress level, though it remains noticeable. Higher significance levels ($\alpha = 0.10$) lead to slightly narrower intervals for γ, indicating a marginal reduction in imprecision. In contrast, lower significance levels ($\alpha = 0.01$) yield larger intervals for γ, corresponding with a slight increase in imprecision for the NPI lower and upper survival functions. Notably, the influence of α is minimized when sample sizes increase, as larger samples tend to reduce the impact of α on predictive imprecision. This suggests that while α selection affects the interval width for γ, a sufficient sample size is more crucial in reducing imprecision.

The simulation study results demonstrate that the proposed method achieves robust predictive inference under correct model assumptions, particularly with larger sample sizes. The figures and tables consistently reveal that increasing sample size reduces imprecision between the NPI lower and upper survival functions, bringing quartile proportions closer to their theoretical expectations. This reduction in imprecision is especially apparent with the largest sample sizes, where the quartile intervals narrow and the alignment between the lower and upper survival functions is most consistent.

This chapter aims to establish a straightforward predictive inference approach, founded on minimal assumptions, where imprecision in the link function and statistical model across different stress levels enhances robustness to key model assumptions and potential misspecification.

In Case 2, robustness is evaluated under conditions of model misspecification. Specifically, this case examines how a simplified model maintains robustness when the shape parameter differs between the sampling and analysis models. Here, the shape parameter β is set to 3 in the sampling model but is assumed to be 2 in the analysis. Table 3.12 and Figures A.10–A.12 display the predictive performance outcomes of the proposed method in this scenario.

Following the setup in Case 1, data are generated according to the model outlined in Section 3.6, where the shape parameters are specified as $\beta_0^1 = 2.5$, $\beta_1^1 = 2.3$, $\beta_2^1 = 2.1$, $\beta_0^2 = 2.4$, $\beta_1^2 = 2.2$, $\beta_2^2 = 2.0$, and $\beta_0^0 = 3$. In the analysis, however, these values are incorrectly assumed as $\beta_0^1 = 2.0$, $\beta_1^1 = 1.8$, $\beta_2^1 = 1.6$, $\beta_0^2 = 1.9$, $\beta_1^2 = 1.7$, $\beta_2^2 = 1.5$, and $\beta_0^0 = 2.5$. In both Case 1 and Case 2, the shape parameter is considered constant and known, though the introduced misspecification in Case 2 results in a small additional imprecision in the quartiles, as seen in Figures A.10–A.12. This imprecision becomes more pronounced in the NPI lower and upper survival functions as sample sizes increase, as reflected in Table 3.12 when compared to Table 3.6. The simulation results confirm that the proposed method retains a level of robustness even under these misspecified conditions. These findings suggest that, although the proposed method exhibits robustness, Case 2 introduces a modest increase in imprecision resulting from the shape parameter's misspecification. Nevertheless, the method upholds a satisfactory level of robustness, even with the model misspecifications introduced in Case 2.

In Case 3, robustness is assessed once more under model misspecification, this time by examining the impact of using an incorrect link function. Specifically, the Eyring link function is applied in the sampling model, while the Arrhenius link function is assumed in the analysis. This configuration tests the robustness of the proposed method in scenarios where the link function may not provide a perfect fit—a situation frequently encountered in practice.

As noted in Section 2.5, the Eyring link function offers an alternative to the Arrhenius function for modeling the acceleration parameter with respect to temperature. Here, a similar setup to Case 1 is applied, where model assumptions are fully appropriate, except for replacing the sampling model's link function with the Eyring function. Table 3.13 presents the simulation results, which, when compared to the fully aligned model assumptions shown in Table 3.11, reveal a similar pattern with slight increases in imprecision, as observed in Tables 3.12 and 3.13. Figures A.7-A.9, A.10-A.12, and A.13-A.15 further illustrate this consistency, showing the NPI lower and upper survival function quartiles for $1-q$ values of 0.75, 0.50, and 0.25. These findings confirm that the proposed method maintains robustness in predictive inference, even when model assumptions around the link function are misspecified.

$m_1 m_0$		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.6052	0.8593	0.7223	0.8093	0.7183	0.7923
	0.50	0.3071	0.6993	0.4392	0.6002	0.4622	0.5572
	0.25	0.0840	0.4812	0.1801	0.3531	0.2211	0.3151
0.05	0.75	0.6212	0.8313	0.7303	0.7973	0.7243	0.7803
	0.50	0.3441	0.6593	0.4592	0.5822	0.4752	0.5462
	0.25	0.1120	0.4352	0.1991	0.3281	0.2281	0.2951
0.10	0.75	0.6323	0.8223	0.7353	0.7913	0.7333	0.7763
	0.50	0.3591	0.6273	0.4672	0.5712	0.4792	0.5402
	0.25	0.1271	0.4022	0.2091	0.3121	0.2371	0.2941
$m_2 m_0$		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.5822	0.8643	0.6973	0.8033	0.6843	0.7813
	0.50	0.2731	0.7083	0.4092	0.5832	0.4062	0.5392
	0.25	0.0780	0.4692	0.1651	0.3201	0.1891	0.2811
0.05	0.75	0.6102	0.8373	0.7053	0.7923	0.6963	0.7703
	0.50	0.3141	0.6593	0.4292	0.5662	0.4222	0.5262
	0.25	0.1090	0.4212	0.1781	0.2991	0.2001	0.2731
0.10	0.75	0.6283	0.8233	0.7123	0.7893	0.7003	0.7653
	0.50	0.3441	0.6343	0.4382	0.5572	0.4322	0.5192
	0.25	0.1261	0.3982	0.1831	0.2861	0.2081	0.2701
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.5062	0.9034	0.6733	0.8493	0.6733	0.8133
	0.50	0.1551	0.7723	0.3391	0.6693	0.3872	0.6102
	0.25	0.0180	0.5582	0.1160	0.4242	0.1591	0.3701
0.05	0.75	0.5422	0.8784	0.6833	0.8353	0.6813	0.8043
	0.50	0.2171	0.7173	0.3741	0.6493	0.3962	0.5912
	0.25	0.0340	0.5032	0.1281	0.3952	0.1731	0.3501
0.10	0.75	0.5592	0.8713	0.6973	0.8253	0.6843	0.8043
	0.50	0.2461	0.6913	0.3912	0.6333	0.4062	0.5842
	0.25	0.0520	0.4672	0.1421	0.3822	0.1801	0.3411

Table 3.11: Proportion of runs with future observation greater than the quartiles. Case 1.

m_1m_0		$n=10$		$n=50$		$n=100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.6413	0.8633	0.7203	0.8323	0.7393	0.8093
	0.50	0.3521	0.7153	0.4672	0.6413	0.4862	0.5892
	0.25	0.1170	0.5062	0.2011	0.4042	0.2341	0.3611
0.05	0.75	0.6623	0.8483	0.7273	0.8183	0.7423	0.8003
	0.50	0.3862	0.6803	0.4752	0.6142	0.4882	0.5762
	0.25	0.1521	0.4732	0.2271	0.3782	0.2421	0.3461
0.10	0.75	0.6723	0.8333	0.7333	0.8143	0.7473	0.7963
	0.50	0.4032	0.6533	0.4822	0.6042	0.4932	0.5712
	0.25	0.1621	0.4352	0.2551	0.3651	0.2471	0.3371
m_2m_0		$n=10$		$n=50$		$n=100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.6032	0.8643	0.7093	0.8043	0.6963	0.7813
	0.50	0.2961	0.7033	0.4262	0.5892	0.4252	0.5482
	0.25	0.0960	0.4822	0.1711	0.4262	0.2061	0.3061
0.05	0.75	0.6273	0.8343	0.7193	0.7943	0.7053	0.7723
	0.50	0.3381	0.6633	0.4452	0.5702	0.4382	0.5312
	0.25	0.1190	0.4382	0.1941	0.3141	0.2191	0.2871
0.10	0.75	0.6423	0.8253	0.7253	0.7933	0.7083	0.7683
	0.50	0.3581	0.6433	0.4502	0.5592	0.4462	0.5232
	0.25	0.1411	0.4122	0.2001	0.3041	0.2211	0.2831
$\underline{\gamma}$ and $\overline{\gamma}$		$n=10$		$n=50$		$n=100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.5362	0.9164	0.7083	0.8683	0.6973	0.8373
	0.50	0.1911	0.7963	0.4022	0.7103	0.4192	0.6593
	0.25	0.0390	0.5952	0.1511	0.4802	0.1891	0.4282
0.05	0.75	0.5702	0.8954	0.7173	0.8553	0.7053	0.8283
	0.50	0.2401	0.7493	0.4292	0.6833	0.4382	0.6413
	0.25	0.0570	0.5472	0.1731	0.4532	0.2031	0.4122
0.10	0.75	0.5872	0.8824	0.7263	0.8483	0.7083	0.8243
	0.50	0.2691	0.7273	0.4432	0.6743	0.4462	0.6313
	0.25	0.0700	0.5132	0.1811	0.4412	0.2111	0.4002

Table 3.12: Proportion of runs with future observation greater than the quartiles. Case 2.

m_1m_0		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.5692	0.8433	0.6694	0.7715	0.6672	0.7615
	0.50	0.2871	0.6923	0.4122	0.5763	0.4141	0.5272
	0.25	0.0780	0.4962	0.1861	0.3682	0.2171	0.3231
0.05	0.75	0.5892	0.8093	0.6764	0.7615	0.6712	0.7515
	0.50	0.3151	0.6453	0.4303	0.5553	0.4291	0.5182
	0.25	0.1220	0.4512	0.2051	0.3342	0.2291	0.3031
0.10	0.75	0.5972	0.7893	0.6824	0.7585	0.6742	0.7506
	0.50	0.3391	0.6172	0.4383	0.5493	0.4341	0.5132
	0.25	0.1461	0.4232	0.2131	0.3182	0.2331	0.2931
m_2m_0		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.5432	0.8453	0.6534	0.7735	0.6422	0.7713
	0.50	0.2581	0.6993	0.3702	0.5653	0.3801	0.5182
	0.25	0.0760	0.4842	0.1731	0.3362	0.1931	0.2991
0.05	0.75	0.5682	0.8113	0.6664	0.7615	0.6552	0.7596
	0.50	0.2891	0.6483	0.4022	0.5413	0.3911	0.5052
	0.25	0.1130	0.4442	0.1871	0.3162	0.2091	0.2881
0.10	0.75	0.5832	0.7963	0.6714	0.7575	0.6622	0.7521
	0.50	0.3171	0.6313	0.4152	0.5323	0.4021	0.5004
	0.25	0.1371	0.4222	0.1921	0.3082	0.2151	0.2801
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.4722	0.8904	0.6154	0.8025	0.6212	0.7923
	0.50	0.1511	0.7623	0.3182	0.6394	0.3601	0.5692
	0.25	0.0160	0.5752	0.1191	0.4263	0.1650	0.3671
0.05	0.75	0.5112	0.8583	0.6374	0.7905	0.6342	0.7804
	0.50	0.2091	0.7123	0.3502	0.6044	0.3741	0.5482
	0.25	0.0410	0.5202	0.1391	0.3962	0.1791	0.3441
0.10	0.75	0.5292	0.8413	0.6434	0.7825	0.6392	0.7804
	0.50	0.2371	0.6833	0.3642	0.5894	0.3801	0.5422
	0.25	0.0560	0.4872	0.1551	0.3842	0.1861	0.3301

Table 3.13: Proportion of runs with future observation greater than the quartiles. case 3.

The primary results from the simulation studies indicate that future observations at the normal stress level consistently exceeded the quartile thresholds in expected proportions, underscoring the robustness of the proposed predictive inference method. This approach demonstrates strong performance with reasonable levels of imprecision when model assumptions hold, especially within the intervals bounded by $\underline{\gamma}$ and $\overline{\gamma}$. However, under conditions of model misspecification—particularly with substantial deviations from the true model—imprecision increases, indicating either a need for model reassessment or the collection of additional data to improve predictive accuracy. Consistent with the findings in Section 3.5, the level of imprecision generally diminishes with larger sample sizes across various strategies, affirming the importance of ample sample sizes for achieving more precise predictions. Additionally, higher significance levels in the likelihood test lead to narrower intervals for $\underline{\gamma}$ and $\overline{\gamma}$, further reducing imprecision. These results imply that the proposed method maintains robustness under minor misspecifications, while significant imprecision may necessitate model refinement or supplemental data to ensure accuracy in predictive inferences.

3.9 Concluding remarks

In this chapter, a new robust statistical method for step-stress accelerated life testing is presented, utilizing the Arrhenius-Weibull cumulative exposure model. This method was developed based on the likelihood ratio test, applied here to compare failure times across various stress level strategies. Imprecision in the acceleration parameter within the Arrhenius link function is introduced through pairwise likelihood ratio tests between different stress level strategies. This imprecision facilitates the transformation of failure times into interval values at the normal stress level, under the assumption that these transformed failure times are indistinguishable from those occurring at the normal stress level. The transformation is achieved through an inversion between the cumulative distribution functions of two strategies, enhancing robustness by providing an interval estimate for the acceleration parameter instead of a single value. Specifically, the pairwise likelihood ratio tests are used to compare two datasets from different stress levels, obtaining an interval for the acceleration parameter within which the null hypothesis—stating that all data originate from the same underlying distribution—cannot be rejected. This ap-

proach allows for a more robust inference by acknowledging and quantifying potential imprecision in the acceleration parameter.

Nonparametric predictive inference (NPI) is subsequently applied to the transformed data, enabling robust predictive inference. Increased imprecision may arise when data from higher stress levels are used or when the assumed model does not adequately fit the data, especially under model misspecification. An investigation into scenarios with differing shape parameters at each strategy also revealed a slight increase in imprecision for the acceleration parameter.

This method assumes the availability of data at the normal stress level; however, in practical applications, such data may not always be accessible. In such cases, performing tests between data from higher stress levels is recommended to estimate an interval for the acceleration parameter. Additionally, while complete datasets are assumed here for simplicity, these methods are readily extendable to right-censored data. Tests can incorporate right-censored data, which are then transformed to the normal stress level; in the final stage, the NPI method for right-censored data, as outlined in Section 2.7, can be applied. As a result, the overall methodology remains unchanged.

Chapter 4

Robust statistical inference using the log-rank test

4.1 Introduction

In Chapter 4, a novel nonparametric approach to step-stress accelerated life testing (SSALT) data analysis is introduced, expanding upon the robust framework established in Chapter 3 with a central enhancement: the elimination of specific assumptions regarding the failure time distribution. While Chapter 3's method relied on the Arrhenius-Weibull cumulative exposure model, the following approach provides a fully nonparametric methodology, removing the need for predefined failure time distributions. Instead, this approach only utilizes a parametric link function to relate stress levels, omitting any additional assumptions about the distribution of failure times. This adaptability makes the proposed method particularly beneficial in scenarios where the underlying failure time distribution is unknown, highly variable, or complex to define accurately.

The motivation for this method arises from practical limitations observed in traditional SSALT modeling. Often, complete information about failure mechanisms and the influence of stress is unavailable, making parametric assumptions difficult to justify. In response, Chapter 4 seeks to create a framework that enhances predictive robustness by overcoming these limitations. The introduction of imprecision via the log-rank test, rather than relying on a predefined statistical model, allows for an interval-based prediction that accounts for the variability and uncertainty inherent in SSALT data across different stress levels. This shift toward a nonparametric predictive inference framework

offers flexibility, making it applicable to a wider range of SSALT applications where the robustness of inference is essential, particularly when exact model assumptions cannot be met.

This new methodology thus extends the utility of SSALT data analysis by accommodating real-world complexity and variability, enhancing its robustness under model misspecification and practical constraints. By enabling data transformation and analysis without a predetermined lifetime distribution model, Chapter 4 supports more adaptive inference with fewer assumptions, enhancing the applicability of SSALT in contexts where traditional models may not fully capture the underlying process.

The key development in this method is the introduction of imprecision through a pairwise log-rank test on the accelerating parameter under the null hypothesis that all failure times originate from the same distribution. This imprecision supports transforming failure times from higher stress levels into interval values at the normal stress level, assuming these transformed values to be indistinguishable from failure times at this baseline level. To achieve this, the method uses the inversion of the Arrhenius link function, enabling data transformation from higher stress levels to the normal stress level. An interval estimate of the accelerating parameter is then obtained, identifying the range where the transformed data and baseline data distributions are statistically similar (i.e., the null hypothesis is not rejected). In the final stage, nonparametric predictive inference (NPI) is applied to the interval-transformed data, providing robust predictive inferences while minimizing reliance on strict model assumptions.

This chapter is organized as follows. Section 4.2 introduces the proposed method, developed based on the log-rank test for analyzing SSALT data. Section 4.3 provides an illustrative example to demonstrate the application of this method. Section 4.4 presents simulation studies conducted to assess the method's performance under various conditions. Finally, Section 4.5 summarizes the key findings of this chapter.

4.2 Imprecise statistical inference based on the log-rank test

In this section, a new imprecise nonparametric statistical method is introduced for step-stress accelerated life testing (SSALT) data. Unlike the previous method, which relies

on explicit models for each strategy, this approach does not assume a specific statistical distribution for the failure times at different stress levels. Previous approaches often use models such as the Weibull cumulative exposure model with the Arrhenius link function for accelerated failure times, and a Weibull distribution at the normal stress level. In contrast, this method employs only a parametric link function, specifically the Arrhenius function, to connect the effects of different stress levels. The Arrhenius link function characterizes the relationship between failure times and applied stress strategies, represented through temperature, by linking the scale parameters across the varying stress conditions in the experiment.

This method consists of three steps. First, it transforms failure times arising from different experimental strategies at higher stress levels to the normal stress level. Second, it introduces imprecision based on the log-rank test applied to the accelerating parameter, under the null hypothesis that all failure times are drawn from the same distribution. This imprecision enables the transformation of failure times into interval values at the normal stress level, where these transformed failure times are assumed to be indistinguishable from those occurring at the normal stress level. Third, nonparametric predictive inference is applied to the transformed data to provide robust predictive inference. This method results in greater imprecision when data from higher stress levels are used or in cases of model misspecification.

While the log-rank test in the second step is used to assess whether the failure times across different strategies plausibly arise from the same underlying distribution, it does not assume or enforce any specific parametric form for the data. Instead, it is applied to identify and quantify potential imprecision in the transformation process. This use of the log-rank test is consistent with the overall nonparametric nature of the method. The imprecision resulting from this step reflects the uncertainty in assuming distributional similarity across stress levels, and it is used in the transformation of failure times to intervals at the normal stress level. The subsequent inference is based on nonparametric predictive inference (NPI), which does not rely on any parametric assumptions about the form of the lifetime distribution.

In this method, the Arrhenius link function is used to analyze step-stress accelerated life testing data, as outlined in Section 3.3. The first step of the method involves transforming failure times generated from various experimental strategies at higher stress

levels to those at the normal stress level. This transformation is carried out through an inversion based on the Arrhenius link function, applied between two temperature stress levels. Specifically, if the Arrhenius link function is used in the analysis, then an observation t_i^m at stress level i and strategy m, subject to stress K_i^m, is transformed into an observation at the normal stress level K_0^0 according to the equation:

$$t_i^{(m)\to(0,0)} = t_i^m \exp\left(\frac{\gamma}{K_0^0} - \frac{\gamma}{K_i^m}\right) \quad \text{for,} \quad t_i^{(m)\to(0,0)} \in (0,\infty),$$

where K_i^m represents the accelerated temperature at stress level i in strategy m, K_0^0 is the normal temperature at stress level 0, and γ is the Arrhenius model parameter.

In the second step, imprecision is introduced based on the log-rank test applied to the accelerating parameter, where the null hypothesis, that all failure times originate from the same distribution, is not rejected. This imprecision allows to transform failure times into interval values at the normal stress level where it is assumed that these transformed failure times are not distinguishable from failure times occurring at the normal stress level. The log-rank test, which was discussed in Section 2.6.2, is generally used to examine the equality of the survival distributions of two or more independent groups. The test is used to compare equality of the distribution between the transformed failure times and the failure times occurring at the normal stress. The interval $[\underline{\gamma}, \overline{\gamma}]$ for the acceleration parameter in the Arrhenius link function is obtained under the condition that the null hypothesis is not rejected. This hypothesis states that the failure times at the normal stress level and the transformed failure times, including any right-censored observations, originate from the same underlying distribution. This is interpreted in a manner that, for the obtained interval of γ, both the transformed data and the actual data at stress level K_0 are well mixed. This imprecision allows us to transform failure times into interval values at the normal stress level where it is assumed that these transformed failure times are not distinguishable from failure times occurring at the normal stress level.

In this method, we use the pairwise log-rank test between strategies m_i and m_0 instead of combing all data in a single test. This single test may not lead to a sensible method of imprecise statistical inference as it shows that the null hypothesis of equality is easy to be rejected. For illustration, if it is assumed there are three data sets: if the pairwise test was performed, the confidence interval will be wider, and it is it is less likely to reject the null hypothesis. If the test is performed and all data sets are compared in a single

test then it is more likely that the test will show you that the probability distributions are different, where the null hypothesis becomes more easy to be rejected.

In this method, the pairwise log-rank test is employed to compare each strategy m_i individually with the strategy m_0, rather than combining all data into a single test. Conducting a single test may lead to less robust imprecise statistical inference, as it can make it easier to reject the null hypothesis of equality due to cumulative differences across multiple strategies. By testing each strategy separately against m_0, we maintain a higher level of imprecision, making it less likely for the null hypothesis to be rejected.

For instance, with three distinct data sets, performing individual pairwise tests between each data set and the baseline strategy results in broader confidence intervals, providing a stronger framework for capturing imprecision. Conversely, if all data sets were analyzed within a single test, the probability distributions would more likely appear different, thus increasing the likelihood of rejecting the null hypothesis. This pairwise approach better accommodates the complexity of data across multiple stress levels and enhances the robustness of the inference.

In this method, pairwise comparisons are conducted between the baseline strategy $m = 0$ and each other strategy m, where $m = 1, 2, \ldots, z$. Each comparison yields lower and upper bounds for γ, denoted by $\underline{\gamma}_{0,m}$ and $\overline{\gamma}_{0,m}$, respectively. The overall lower bound for γ is obtained by taking the minimum of these pairwise lower bounds, $\underline{\gamma} = \min(\underline{\gamma}_{0,m})$, while the overall upper bound is determined by taking the maximum of the upper bounds, $\overline{\gamma} = \max(\overline{\gamma}_{0,m})$. This approach, implemented in Chapters 3 and 4, effectively captures the imprecision introduced through the pairwise comparisons across strategies. By consolidating the minimum and maximum bounds from all pairwise comparisons, a comprehensive interval for γ is established, reflecting imprecision across the various experimental strategies.

It is noteworthy that if the model is a good fit, most values of the lower bound $\underline{\gamma}_{0,m}$ and all values of the upper bound $\overline{\gamma}_{0,m}$ are expected to be approximately similar. In contrast, if the model fit is poor, the values of $\underline{\gamma}_{0,m}$ and $\overline{\gamma}_{0,m}$ may vary substantially from one another. Consequently, in cases of inadequate model fit, the resulting interval $[\underline{\gamma}, \overline{\gamma}]$ is anticipated to be wider than it would be under a perfect model fit, reflecting greater imprecision. This expanded interval for $[\underline{\gamma}, \overline{\gamma}]$ subsequently produces wider intervals for the transformed failure times, indicating increased uncertainty. If the chosen model

represents the data well, the widest interval for the accelerating parameter $[\underline{\gamma}, \overline{\gamma}]$ is likely to be derived from the pairwise log-rank test conducted between strategies m_2 and m_1.

In the third step, nonparametric predictive inference (NPI) is implemented, as detailed in Section 2.7, to construct the NPI lower and upper survival functions based on all available data, which includes both the transformed and actual data at the normal stress level. The NPI lower survival function, denoted as $\underline{S}(t)$, is derived using the transformed failure times at the normal stress level with the overall lower bound $\underline{\gamma}$, along with the actual data at this stress level. Similarly, the NPI upper survival function, $\overline{S}(t)$, is constructed using the transformed failure times based on the upper bound $\overline{\gamma}$, ensuring that the imprecision due to uncertainties in the parameter γ is fully incorporated. This approach provides robust predictive intervals that capture the imprecision introduced by the transformation process.

In this method, the objective is not to combine all multiple tests across all strategies into a single test. Instead, pairwise tests are performed separately between two strategies at a time. Suppose that a single test were conducted under the null hypothesis, transforming data from all strategies to the normal stress level K_0 based on a specific parameter value γ_a derived from the equivalent original distribution. Let $[\underline{\gamma_a}, \overline{\gamma_a}]$ represent the interval of γ_a values for which the null hypothesis is not rejected in this single-test scenario. Under a good model fit, $[\underline{\gamma_a}, \overline{\gamma_a}]$ would be close to $[\underline{\gamma}, \overline{\gamma}]$. Conversely, with a poor model fit, the interval $[\underline{\gamma_a}, \overline{\gamma_a}]$ could be extremely narrow or even null, contradicting our goal of capturing imprecision. Therefore, rather than a single test, this method applies pairwise comparisons between each m_i and m_0, adopting the minimum and maximum bounds across all intervals $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$. This approach enhances imprecision when the model fit is inadequate, ensuring robustness through the interval $[\underline{\gamma}, \overline{\gamma}]$.

The subsequent inference is based on NPI, yielding both lower and upper predictive survival functions that incorporate the introduced imprecision. The robustness and practical applicability of this predictive method are evaluated through comprehensive simulation studies, demonstrating its suitability for real-world SSALT data analysis under varying model assumptions.

4.3 Illustrative example

This section presents an example to illustrate the proposed method. Similar experimental settings are adopted from Example 3.4.1 to facilitate a comparison between the two methods introduced in this chapter. Three data sets are generated, each with a sample size of $n = 10$ across all strategies. These strategies include two accelerated testing strategies, alongside a data set collected at the normal stress level.

Example 4.3.1 This example is structured into three cases. In Case 1, the shape parameter is assumed to be constant and known across each strategy, and the Arrhenius link function is applied to connect the scale parameters at all stress levels within each strategy for the simulated data. In Case 2, the test is performed in a unified manner by transforming all data sets to be compared directly with the data at the normal stress level. In Case 3, the Arrhenius link function is replaced with the Eyring link function, representing a scenario where the chosen link function does not fit the data well. This substitution allows for an examination of the resulting interval $[\underline{\gamma}, \overline{\gamma}]$ and its impact on the corresponding lower and upper NPI survival functions at the normal stress level.

Three data sets were generated, each corresponding to a different experimental strategy. In the first strategy, the experiment was set to a normal temperature level of $k_0^0 = 300$ Kelvin, with a scale parameter of $\theta_0^0 = 7000$, representing the normal use stress level. Ten observations were generated from a Weibull distribution for this strategy. The Arrhenius link function was applied to connect the scale parameters across all strategies, with an acceleration parameter of $\gamma = 5000$.

In the second strategy, the accelerated experiment was configured with temperature levels of $k_0^1 = 350$, $k_1^1 = 400$, and $k_2^1 = 450$ Kelvin for stress levels s_0^1, s_1^1, and s_2^1, respectively. The stress level increased from s_0^1 to s_1^1 at $\tau_1^1 = 300$, and from s_1^1 to s_2^1 at $\tau_2^1 = 350$. The corresponding scale parameters were $\theta_0^1 = 647.23$, $\theta_1^1 = 108.52$, and $\theta_2^1 = 8.72$ for s_0^1, s_1^1, and s_2^1.

In the third strategy, the accelerated experiment involved temperature levels of $k_0^2 = 380$, $k_1^2 = 420$, and $k_2^2 = 460$ Kelvin for stress levels s_0^2, s_1^2, and s_2^2, respectively. The stress level increased from s_0^2 to s_1^2 at $\tau_1^2 = 100$ and from s_1^2 to s_2^2 at $\tau_2^2 = 130$. The corresponding scale parameters were $\theta_0^2 = 209.53$, $\theta_1^2 = 59.84$, and $\theta_2^2 = 21.25$ for s_0^2, s_1^2, and s_2^2. The generated failure times for each strategy are presented in Table 4.1.

Strategies	Stress level	Data sets	Failures times
m_0	s_0	t_0	1755.83, 4149.09, 4799.74, 5095.24, 7454.25, 7563.30, 8245.62, 10385.32, 11166.58, 12411.59
m_1	s_0^1	t_0^1	185.63, 222.7402
m_1	s_1^1	t_1^1	300.66, 320.37, 325.00, 341.08, 342.94
m_1	s_2^1	t_2^1	359.43, 360.06, 371.28
m_2	s_0^3	t_0^2	14.28, 42.54
m_2	s_1^2	t_1^2	101.73, 105.56, 108.97, 120.80, 123.50
m_2	s_2^2	t_2^2	138.93, 140.52, 146.53

Table 4.1: A simulated data of Example 4.3.1.

The NPI lower survival function, $\underline{S}$, is constructed by combining the data at the normal stress level with the transformed data from the higher stress strategies, specifically transforming data from m_1 to m_0 and m_2 to m_0 using the lower bound $\underline{\gamma}$. This transformation represents a pessimistic scenario, as it uses the lower points of the interval, thus defining the lower survival bound.

Similarly, the NPI upper survival function, $\overline{S}$, is obtained by combining the normal stress level data with the transformed data from the higher stress strategies using the upper bound $\overline{\gamma}$. This setup reflects the optimistic scenario, where the upper points of the transformed data define the upper survival bound.

To analyze the data sets presented in Table 4.1, we assume the Arrhenius link function for the accelerated strategies m_1 and m_2 and link them to the data at the normal stress level, m_0. First, the pairwise log-rank test is applied between m_0 and each of the other strategies m (where $m = 1, 2$), to determine the intervals $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ for which the null hypothesis is not rejected. The resulting intervals $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ at three significance levels are shown in Table 4.2.

It is important to note that the transformed data are based on the overall bounds $[\underline{\gamma}, \overline{\gamma}]$, derived as the minimum and maximum values across the pairwise tests. Consequently, all failure times at the higher stress strategies are transformed to the normal stress level, resulting in interval-valued data at the normal stress level, m_0, for the strategies m_1 and m_2.

Cases	Significance Level	0.01		0.05		0.10	
	Stratgies	$\underline{\gamma}$	$\overline{\gamma}$	$\underline{\gamma}$	$\overline{\gamma}$	$\underline{\gamma}$	$\overline{\gamma}$
Case 1	m_1, m_0	2712.69	4313.12	2818.03	4186.27	2998.07	4157.158
	m_2, m_0	3401.54	4894.44	3512.16	4818.28	3675.03	4729.59
Case 2	m_2, m_1, m_0	2818.03	4752.82	3025.67	4451.96	3060.86	4371.04
Case 3	m_1, m_0	2347.77	3967.90	2453.11	13841.05	2633.15	3811.93
	m_2, m_0	3032.87	4541.15	3143.49	4464.99	3306.36	4376.29

Table 4.2: $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ for Example 4.3.1.

In Case 1, the two accelerated strategies m_1 and m_2 are compared with the data at the normal stress level, m_0. The results indicate that $\hat{\gamma}_{0,1} = 3182.17$ for the comparison between the first and second strategies, and $\hat{\gamma}_{0,2} = 3845.95$ for the comparison between the first and third strategies. In both tests, the p-value is equal to 1, suggesting no significant difference between the distributions at the chosen stress levels. Table 4.3 shows the transformed data values, derived using γ, $\underline{\gamma}$, and $\overline{\gamma}$ across different significance levels. As mentioned previously, the minimum and maximum values of $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ among all strategies are applied to transform all failure times at higher stress levels in this case. The NPI lower and upper survival functions for this case are shown in Figure 4.1.

In Case 2, the two accelerated strategies m_1 and m_2 are combined in a single test against the data at the normal stress strategy m_0, as discussed earlier. The analysis yields an estimated value of $\hat{\gamma} = 3725.91$ for the combined strategies, with a p-value equal to 1, indicating no significant difference in failure time distributions across these stress levels. Table 4.4 presents the transformed data values obtained using $\hat{\gamma}$, $\underline{\gamma_a}$, and $\overline{\gamma_a}$ across different significance levels. As mentioned previously, the minimum and maximum values of $\underline{\gamma_a}$ and $\overline{\gamma_a}$ from all strategies are used to transform the failure times at higher stress levels in this case. The NPI lower and upper survival functions for this case are shown in Figure 4.2.

In Case 3, the Arrhenius link function is replaced with the Eyring link function to intentionally examine the effects of a model that does not provide a good fit to the data. This allows for an assessment of imprecision when the assumed link function differs from the actual stress-response relationship. The two accelerated strategies m_1 and m_2 are individually compared to the normal stress strategy m_0, yielding estimated

Data set	Failures times	$\underline{\gamma}_{0.01}$	$\underline{\gamma}_{0.05}$	$\underline{\gamma}_{0.10}$	$\hat{\gamma}$	$\overline{\gamma}_{0.10}$	$\overline{\gamma}_{0.05}$	$\overline{\gamma}_{0.01}$
t_0^1	185.64	675.57	710.32	773.91	844.82	1765.14	1841.30	1909.30
t_0^1	222.74	810.59	852.29	928.58	1013.66	2117.93	2209.30	2290.89
t_1^1	300.66	2882.96	3147.47	3656.96	4263.32	15480.24	16667.81	17759.92
t_1^1	320.38	3071.97	3353.83	3896.72	4542.83	16495.17	17760.60	18924.32
t_1^1	325.01	3116.39	3402.31	3953.05	4608.51	16733.64	18017.37	19197.91
t_1^1	341.08	3270.49	3570.56	4148.54	4836.40	17561.14	18908.35	20147.27
t_1^1	342.94	3288.37	3590.08	4171.21	4862.84	17657.13	19011.70	20257.40
t_2^1	359.44	7322.03	8231.19	10054.04	12336.08	68846.84	75977.55	82686.66
t_2^1	360.07	7334.87	8245.62	10071.66	12357.71	68967.53	76110.74	82831.61
t_2^1	371.28	7563.30	8502.42	10385.33	12742.57	71115.41	78481.08	85411.26
t_0^2	14.29	95.87	103.22	117.13	212.35	394.79	420.14	443.21
t_0^2	42.54	285.46	307.36	348.76	632.31	1175.53	1251.02	1319.70
t_1^2	101.73	1347.33	1489.51	1768.12	3964.66	9197.96	10008.72	10761.64
t_1^2	105.56	1398.03	1545.56	1834.65	4113.84	9544.07	10385.33	11166.58
t_1^2	108.97	1443.19	1595.49	1893.91	4246.73	9852.35	10720.79	11527.28
t_1^2	120.81	1599.90	1768.74	2099.57	4707.87	10922.20	11884.94	12779.01
t_1^2	123.51	1635.70	1808.31	2146.55	4813.21	11166.58	12150.86	13064.93
t_2^2	138.94	3226.62	3645.77	4492.05	12005.40	33443.80	37066.18	40487.95
t_2^2	140.52	3263.35	3687.27	4543.18	12142.05	33824.49	37488.10	40948.82
t_2^2	146.53	3402.93	3844.98	4737.50	12661.39	35271.21	39091.52	42700.26

Table 4.3: Transformation of Case 1, Example 4.3.1.

values $\hat{\gamma}_{1,0} = 2821.171$ for the comparison between the first and second strategies, and $\hat{\gamma}_{2,0} = 3492.332$ for the first and third strategies. In both comparisons, the p-value is 1, indicating no significant difference in failure distributions across these stress levels. Table 4.3 presents the transformed data values obtained using using γ, $\underline{\gamma}$, and $\overline{\gamma}$ across different significance levels. As outlined previously, the minimum and maximum values from the pairwise comparisons—namely, $\underline{\gamma}_{0,m}$ and $\overline{\gamma}_{0,m}$ where $m = 1, 2$—are applied to transform all failure times at the higher stress levels. The NPI lower and upper survival functions for this case are shown in Figure 4.3, capturing the imprecision introduced by the misspecified link function due to the Eyring model.

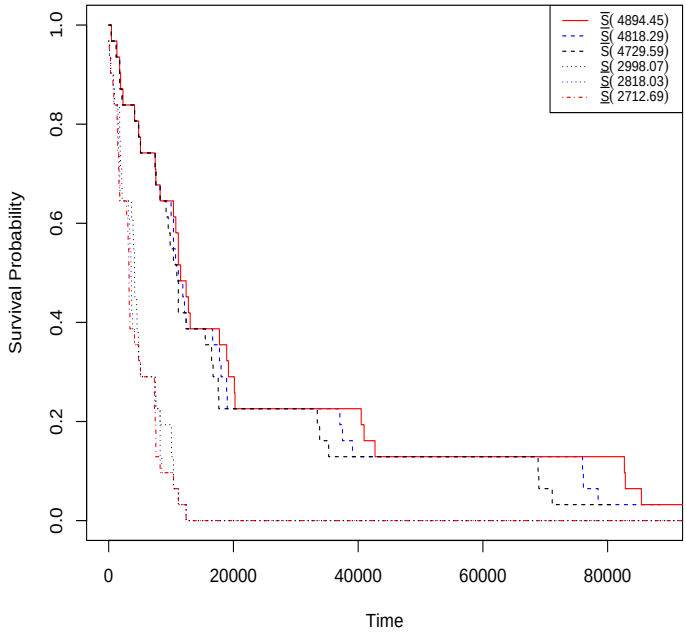


Figure 4.1: NPI lower and upper survival functions for Case 1 in Example 4.3.1, with red, blue, and black lines representing significance levels 0.01, 0.05, and 0.10, respectively.

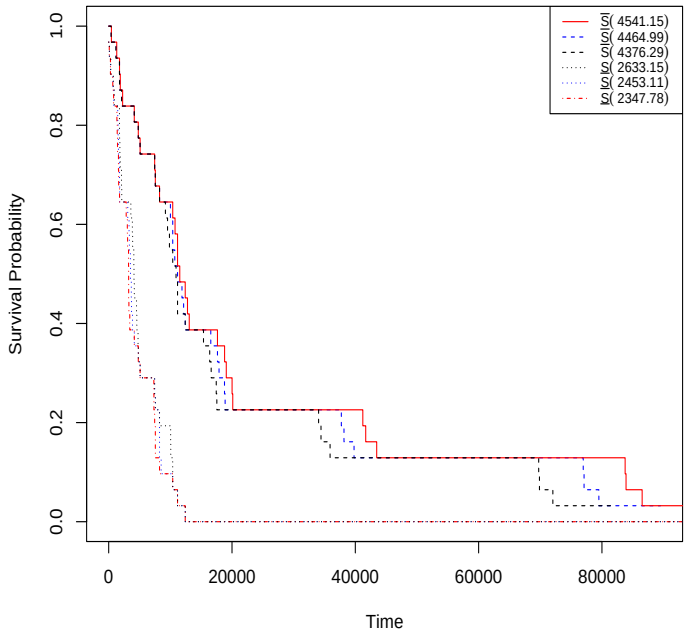


Figure 4.3: NPI lower and upper survival functions for Case 3 in Example 4.3.1, with red, blue, and black lines representing significance levels 0.01, 0.05, and 0.10, respectively.

Data set	Failures times	$\underline{\gamma}_{0.01}$	$\underline{\gamma}_{0.05}$	$\underline{\gamma}_{0.10}$	$\hat{\gamma}$	$\overline{\gamma}_{0.10}$	$\overline{\gamma}_{0.05}$	$\overline{\gamma}_{0.01}$
t_0^1	185.64	710.32	784.15	797.40	1094.49	1488.09	1546.56	1784.78
t_0^1	222.74	852.29	940.87	956.77	1313.24	1785.50	1855.65	2141.49
t_1^1	300.66	3147.47	3742.05	3853.40	6707.08	11481.93	12282.95	15782.88
t_1^1	320.38	3353.83	3987.39	4106.04	7146.82	12234.72	13088.26	16817.66
t_1^1	325.01	3402.31	4045.03	4165.40	7250.14	12411.60	13277.48	17060.79
t_1^1	341.08	3570.56	4245.06	4371.39	7608.66	13025.36	13934.06	17904.47
t_1^1	342.94	3590.08	4268.27	4395.28	7650.25	13096.56	14010.23	18002.34
t_2^1	359.44	8231.19	10367.15	10780.52	22571.40	46223.94	50572.85	70647.33
t_2^1	360.07	8245.62	10385.33	10799.42	22610.97	46304.98	50661.51	70771.18
t_2^1	371.28	8502.42	10708.76	11135.75	23315.15	47747.07	52239.28	72975.23
t_0^2	14.29	103.22	119.42	122.40	195.20	306.97	324.90	401.28
t_0^2	42.54	307.36	355.58	364.47	581.22	914.03	967.44	1194.85
t_1^2	101.73	1489.51	1815.21	1877.08	3536.35	6537.19	7060.95	9403.76
t_1^2	105.56	1545.56	1883.52	1947.71	3669.42	6783.18	7326.64	9757.61
t_1^2	108.97	1595.49	1944.36	2010.62	3787.95	7002.28	7563.30	10072.80
t_1^2	120.81	1768.74	2155.49	2228.95	4199.27	7762.65	8384.59	11166.58
t_1^2	123.51	1808.31	2203.72	2278.82	4293.23	7936.34	8572.19	11416.43
t_2^2	138.94	3645.77	4638.12	4831.26	10445.58	22068.67	24239.55	34356.97
t_2^2	140.52	3687.27	4690.92	4886.26	10564.48	22319.87	24515.46	34748.05
t_2^2	146.53	3844.98	4891.56	5095.25	11016.34	23274.53	25564.02	36234.27

Table 4.4: Tranformation of Case 2, Example 4.3.1.

Based on Table 4.2, a comparison of the resulting intervals $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ reveals that Case 1 exhibits wider intervals than Case 2, indicating greater imprecision in the parameter estimates. In Case 3, where the Arrhenius link function is replaced by the Eyring link function, the imprecision further increases, as the method does not align well with the data distribution under the Eyring model. This added imprecision suggests a higher degree of uncertainty due to the link function's misspecification, making Case 3's intervals the widest across significance levels. Therefore, the choice of link function clearly impacts the interval bounds and highlights the robustness of the method in capturing model fit variations.

In terms of visual comparison, Figures 4.1, 4.2, and 4.3 illustrate the NPI lower and upper survival functions across Cases 1, 2, and 3, respectively. In Case 1 (Figure 4.1),

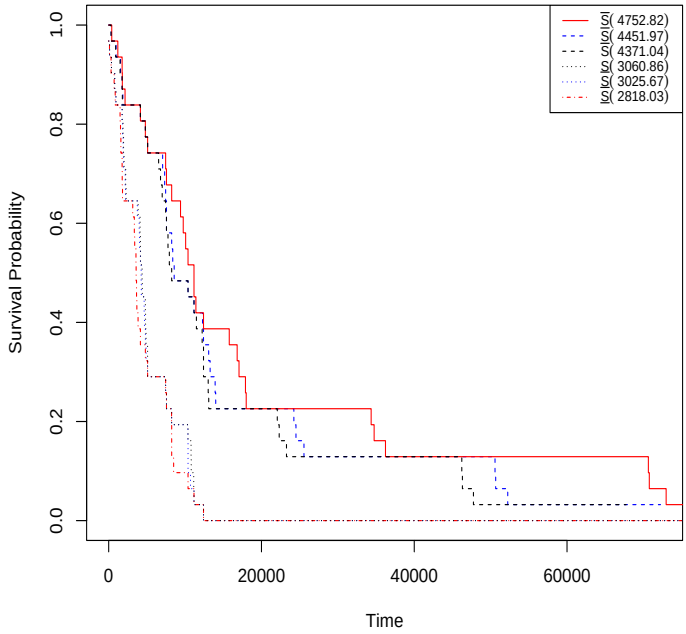


Figure 4.2: NPI lower and upper survival functions for Case 2 in Example 4.3.1, with red, blue, and black lines representing significance levels 0.01, 0.05, and 0.10, respectively.

the survival functions demonstrate moderate imprecision, with distinct upper and lower bounds. Figure 4.2 for Case 2, where tests were combined in a single transformation step, shows a reduced gap between the bounds, reflecting less imprecision. Lastly, in Figure 4.3 for Case 3, the survival function exhibits the widest range, highlighting the increased imprecision from the Eyring link function's incompatibility. These figures effectively depict how model fit and test configuration impact survival function robustness in SSALT data analysis.

In addition, we compare the results of the interval $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ from this example with those in Example 3.4.1, which was analyzed using the likelihood ratio test in Chapter 3. The results indicate that the interval $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ generated by the log-rank test in Chapter 4 demonstrates slightly more imprecision, as reflected in the resulting lower and upper NPI survival functions, compared to the intervals obtained using the likelihood ratio test, particularly for cases with higher stress levels.

In Chapter 3, two cases were examined: one where the shape parameter was held constant and another where it varied across strategies. In the constant shape parameter case, the interval $[\underline{\gamma}_{0,m}, \overline{\gamma}_{0,m}]$ showed more precision due to the consistency in parameter

Data set	Failures times	$\underline{\gamma}_{0.01}$	$\underline{\gamma}_{0.05}$	$\underline{\gamma}_{0.10}$	$\hat{\gamma}$	$\overline{\gamma}_{0.10}$	$\overline{\gamma}_{0.05}$	$\overline{\gamma}_{0.01}$
t_0^1	185.64	662.45	696.52	758.87	829.95	1740.46	1815.54	1882.59
t_0^1	222.74	794.84	835.73	910.54	995.83	2088.30	2178.40	2258.85
t_1^1	300.66	2836.02	3096.22	3597.41	4207.62	15376.39	16555.99	17640.78
t_1^1	320.38	3021.95	3299.22	3833.27	4483.48	16384.51	17641.46	18797.37
t_1^1	325.01	3065.64	3346.92	3888.69	4548.30	16621.38	17896.50	19069.12
t_1^1	341.08	3217.24	3512.42	4080.99	4773.22	17443.33	18781.50	20012.11
t_1^1	342.94	3234.83	3531.62	4103.30	4799.31	17538.68	18884.16	20121.50
t_2^1	359.44	7322.03	8231.19	10054.04	12389.92	69741.71	76965.10	83761.41
t_2^1	360.07	7334.87	8245.62	10071.66	12411.64	69863.97	77100.02	83908.25
t_2^1	371.28	7563.30	8502.42	10385.33	12798.18	72039.77	79501.18	86521.43
t_0^2	14.29	94.00	101.21	114.84	209.87	390.26	415.32	438.12
t_0^2	42.54	279.89	301.37	341.95	624.91	1162.04	1236.67	1304.56
t_1^2	101.73	1332.50	1473.12	1748.65	3963.44	9197.96	10008.71	10761.64
t_1^2	105.56	1382.64	1528.55	1814.45	4112.58	9544.07	10385.33	11166.58
t_1^2	108.97	1427.30	1577.92	1873.06	4245.42	9852.35	10720.79	11527.28
t_1^2	120.81	1582.29	1749.27	2076.46	4706.43	10922.20	11884.94	12779.01
t_1^2	123.51	1617.70	1788.41	2122.92	4811.73	11166.58	12150.86	13064.93
t_2^2	138.94	3240.69	3661.66	4511.63	12216.75	34045.31	37732.83	41216.15
t_2^2	140.52	3277.58	3703.34	4562.98	12355.82	34432.84	38162.34	41685.31
t_2^2	146.53	3417.76	3861.74	4758.15	12884.29	35905.58	39794.60	43468.25

Table 4.5: Tranformation of Case 3, Example 4.3.1.

assumptions, making it more suitable for cumulative exposure models with fixed shape parameters. However, in the varying shape parameter case, a larger degree of imprecision was noted, reflecting the added uncertainty in estimating γ under differing shape parameter conditions.

By comparison, the log-rank test in Chapter 4 yields slightly larger intervals for the overall lower and upper γ, as evidenced by the wider lower and upper NPI survival functions, even with misspecified link functions such as the Eyring model. This increased interval width in the NPI survival functions suggests that the log-rank test captures a broader range for γ, introducing slightly more variability compared to the likelihood ratio test. Thus, while the log-rank approach in Chapter 4 is advantageous because it does not require additional assumptions, the increased imprecision arises due to the inherent

nature of nonparametric tests.

4.4 Simulation studies

In this section, simulation studies are conducted to evaluate the performance of the proposed method, as outlined in Section 4.2. These simulations are designed around three strategies for SSALT data, using experimental settings consistent with Example 4.3.1.

In the first strategy, representing the normal stress level, the scale parameter θ_0^0 is set to $\theta_0^0 = 7000$, with the temperature $k_0^0 = 300$. In the second strategy, three increasing stress levels are assumed, with temperatures $k_0^1 = 350$, $k_1^1 = 400$, and $k_2^1 = 450$, transitioning at times $\tau_1^1 = 300$ and $\tau_2^1 = 350$. Similarly, the third strategy includes three stress levels: $k_0^2 = 380$, $k_1^2 = 420$, and $k_2^2 = 460$ Kelvin, with transitions at $\tau_1^2 = 100$ and $\tau_2^2 = 130$. For all strategies, the shape parameter is set to $\beta = 2$, and the Arrhenius link function parameter, γ, is assumed to be $\gamma = 5000$. Note that while the cumulative exposure model and Weibull distribution are used for data generation, only the link functions are applied in the analysis. Although the data are generated using the cumulative exposure model (CEM) under a Weibull distribution, the CEM is not employed during the analysis; instead, each simulation iteration relies solely on the link function and the log-rank test.

The simulations follow procedures similar to those in Section 3.5, with the addition of the imprecise statistical inference approach based on the log-rank test, as detailed in Section 4.2. This approach enables quantifying imprecision in the parameter γ through interval bounds derived from pairwise comparisons, reflecting how the log-rank test manages model assumptions, stress level effects, and imprecision in transformed data across varied experimental settings.

Each simulation was repeated 10,000 times, with data generated from the proposed model and varying sample sizes for each strategy ($n = 20$, 50, and 100) and significance levels of 0.01, 0.05, and 0.10. The performance of the method is assessed by simulating a future observation at the normal strategy m_0, testing whether it mixes indistinguishably with transformed data from higher stress strategies as well as with data at the normal stress level. To evaluate robustness, the future observation is analyzed against the quartiles of the NPI lower and upper survival functions, with significance levels set

at $1 - q = 0.75$, 0.50, and 0.25.

In assessing the method's robustness, the proportion of cases is computed where the future observation exceeds the quartiles of the NPI lower and upper survival functions. For ideal performance, the first, second, and third quartiles of the lower NPI survival functions are expected to exceed 0.75, 0.50, and 0.25, respectively, while those of the upper survival functions should remain below these values.

The simulation studies consider two primary cases. In Case 1, the conditions specified in Example 4.3.1 are applied, assuming a constant, known shape parameter. Table 4.6 and Figures A.16–A.18 present the performance outcomes across significance levels of 0.01, 0.05, and 0.10, and sample sizes of $n = 10, 50$, and 100. The figures denote the first, second, and third quartiles of the NPI lower and upper survival functions as qL0.25, qU0.25, qL0.50, qU0.50, qL0.75, and qU0.75, respectively. The results indicate a general trend toward reduced imprecision between lower and upper survival functions, as seen in Figure A.17, suggesting that the proposed method achieves robust predictive inference when model assumptions are accurately met.

In Case 2, robustness is evaluated under model misspecification, examining whether the method maintains robustness when assumptions differ from the underlying model. Here, data are generated using the Eyring link function while the analysis assumes the Arrhenius link function, testing performance when the assumed link function may not accurately represent the true relationship. As outlined in Section 2.5, the Eyring link function serves as an alternative to the Arrhenius link function for modeling the temperature-dependent accelerating parameter. Table 4.7 and Figures A.19–A.21 display the outcomes. Comparatively, these results align closely with those in Tables 4.6 and 4.7 (where model assumptions are entirely accurate). Similarly, quartiles of the NPI lower and upper survival functions in Figures A.16–A.18, and A.19–A.21 exhibit high consistency across significance levels of $1 - q = 0.75$, 0.50, and 0.25. These findings confirm the proposed method's robustness, effectively preserving predictive inference even when model assumptions are not met.

Tables 4.6 and 4.7 display the proportion of runs where a future observation exceeds specified quartiles of the NPI lower and upper survival functions for Cases 1 and 2, respectively, with Figures A.16–A.18 illustrating results for Case 1 and Figures A.19–A.21 for Case 2.

The tables and figures reveal how differences in significance levels (α = 0.01, 0.05, and 0.10) and sample sizes (n = 20, 50, and 100) affect prediction performance. In both cases, larger sample sizes (e.g., n = 100) generally yield more consistent quartile coverage, enhancing robustness as sample size increases. Conversely, smaller significance levels (α = 0.01) widen the intervals, resulting in a more evenly distributed spread of proportions, highlighting that lower α values yield broader confidence intervals and thus increase imprecision in predictions.

In Case 1, where model assumptions are fully accurate, observed proportions closely align with expected values, reflecting a high level of consistency. This suggests the proposed method achieves reasonable predictive performance when assumptions are fully met. Case 2 examines robustness under model misspecification, where the Eyring link function is used in data generation, and the Arrhenius link function is applied in analysis. Despite the variation, proportions remain within acceptable ranges, with slightly increased imprecision in quartile coverage. This outcome illustrates the method's robustness, offering reasonably accurate prediction intervals even when model assumptions are not entirely accurate.

Data from higher stress strategies, such as m_2, exhibits greater imprecision, as reflected in the differences between the observed proportions. This trend indicates that higher stress conditions introduce additional variability, leading to less precise predictive intervals. In particular, the overall γ parameter analysis displays the highest level of imprecision, underscoring the increased uncertainty at elevated stress levels. In comparison, the $m_1 m_0$ test provides a more controlled setting, resulting in lower imprecision in the predictive intervals. These findings demonstrate the method's capacity to capture the variability inherent in different stress strategies, with greater imprecision arising as stress levels increase.

The results from the tables and figures collectively demonstrate that the proposed method maintains robust predictive inference, with robustness improving with larger sample sizes and only marginally impacted by model misspecification, as observed in Case 2. These findings confirm the method's adaptability and accuracy for practical applications, even in scenarios with varying model assumptions.

A comparative analysis between Chapter 3 and Chapter 4 for Case 1 demonstrates the method's robustness under similar conditions, focusing on consistency in predictive

m_1m_0		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.6092	0.8868	0.7031	0.8121	0.7122	0.7962
	0.50	0.2548	0.6720	0.4268	0.5853	0.4719	0.5835
	0.25	0.0445	0.5267	0.0950	0.2904	0.0887	0.2902
0.05	0.75	0.6473	0.8577	0.7126	0.7991	0.7197	0.7847
	0.50	0.3266	0.6256	0.4506	0.5837	0.4795	0.5832
	0.25	0.0803	0.4466	0.1015	0.2490	0.0941	0.2663
0.10	0.75	0.6615	0.8415	0.7178	0.7912	0.7238	0.7791
	0.50	0.3613	0.6064	0.4573	0.5825	0.4826	0.5831
	0.25	0.0998	0.4010	0.1055	0.2265	0.0980	0.2500
m_2m_0		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.6119	0.8962	0.7098	0.8243	0.7230	0.8065
	0.50	0.2586	0.6855	0.4117	0.6221	0.4692	0.6249
	0.25	0.0466	0.5120	0.1246	0.2825	0.1183	0.2906
0.05	0.75	0.6476	0.8672	0.7194	0.8110	0.7303	0.7962
	0.50	0.3246	0.6466	0.4439	0.6184	0.4877	0.6229
	0.25	0.0881	0.4347	0.1280	0.2736	0.1216	0.2660
0.10	0.75	0.6623	0.8510	0.7238	0.8033	0.7351	0.7925
	0.50	0.3595	0.6296	0.4567	0.6167	0.4945	0.6215
	0.25	0.1062	0.3869	0.1314	0.2553	0.1232	0.2542
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.5649	0.9670	0.6824	0.9395	0.6935	0.9367
	0.50	0.0950	0.8747	0.3124	0.7841	0.3287	.7620
	0.25	0.0022	0.5859	0.0307	0.3630	0.0417	0.3361
0.05	0.75	0.6307	0.9542	0.6972	0.9331	0.7049	0.9317
	0.50	0.1793	0.8322	0.3310	0.7628	0.3369	0.7440
	0.25	0.0114	0.4944	0.0469	0.3482	0.0526	0.3334
0.10	0.75	0.6491	0.9472	0.7052	0.9306	0.7112	0.9296
	0.50	0.2204	0.8085	0.3396	0.7506	0.3412	0.7353
	0.25	0.0200	0.4582	0.0553	0.3421	0.0596	0.3332

Table 4.6: Proportion of runs with future observation greater than the quartiles, Case 1.

m_1m_0		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.6545	0.8403	0.7044	0.8212	0.7108	0.7944
	0.50	0.3447	0.5731	0.4200	0.4569	0.4685	0.4749
	0.25	0.0919	0.4266	0.0775	0.3484	0.0622	0.2327
0.05	0.75	0.6459	0.8390	0.7180	0.8091	0.7195	0.7857
	0.50	0.3119	0.5866	0.4210	0.4464	0.4685	0.4749
	0.25	0.0679	0.4790	0.0955	0.3073	0.0821	0.2056
0.10	0.75	0.6601	0.8265	0.7243	0.8029	0.7259	0.7809
	0.50	0.3357	0.5566	0.4212	0.4428	0.4685	0.4749
	0.25	0.0829	0.4373	0.1081	0.2866	0.0908	0.1984
m_2m_0		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.5977	0.8705	0.7084	0.8299	0.7211	0.8048
	0.50	0.2616	0.6648	0.4404	0.4809	0.4263	0.4382
	0.25	0.0426	0.5307	0.0855	0.3429	0.0908	0.2518
0.05	0.75	0.5977	0.8705	0.7230	0.8170	0.7307	0.7936
	0.50	0.2616	0.6648	0.4418	0.4714	0.4263	0.4367
	0.25	0.0426	0.5307	0.1068	0.3039	0.1116	0.2351
0.10	0.75	0.6354	0.8490	0.7293	0.8112	0.7378	0.7857
	0.50	0.3164	0.5988	0.4423	0.4688	0.4263	0.4367
	0.25	0.0752	0.4668	0.1186	0.2833	0.1203	0.2223
$\underline{\gamma}$ and $\overline{\gamma}$		$n = 10$		$n = 50$		$n = 100$	
α	$1-q$	qU	qL	qU	qL	qU	qL
0.01	0.75	0.5401	0.9508	0.6771	0.8866	0.7028	0.8837
	0.50	0.0973	0.8484	0.2645	0.7982	0.2892	0.7578
	0.25	0.0060	0.6119	0.0495	0.3781	0.0637	0.2645
0.05	0.75	0.5977	0.9350	0.6937	0.8785	0.7124	0.8813
	0.50	0.1710	0.8238	0.2838	0.7771	0.3044	0.7426
	0.25	0.0221	0.5131	0.0639	0.3355	0.0733	0.2446
0.10	0.75	0.6202	0.9249	0.7019	0.8755	0.7187	0.8805
	0.50	0.2061	0.8058	0.2926	0.7654	0.3116	0.7355
	0.25	0.0321	0.4673	0.0715	0.3157	0.0805	0.2382

Table 4.7: Proportion of runs with future observation greater than the quartiles, Case 2.

intervals. Results from this chapter exhibit slightly more imprecision compared to those in Chapter 3, with observed proportions aligning more closely with expected quartile values, especially for larger sample sizes. This increase in imprecision reflects the broader intervals derived from the overall lower and upper γ values, indicating the method's sensitivity in Chapter 4 to variability under model assumptions.

The comparison between Case 3 in Chapter 3 (using the Eyring link function for data generation but analyzing with the Arrhenius link function) and Case 2 in Chapter 4 (implementing the pairwise log-rank test) reveals that Chapter 4 results (Table 4.7) show slightly greater imprecision than those in Chapter 3 (Table 3.7). This increased imprecision is particularly noticeable for smaller sample sizes ($n = 10$ and $n = 50$), where wider intervals based on the overall lower and upper γ values introduce added imprecision, especially under model misspecification. In contrast, Chapter 3 results display comparatively narrower bounds between qU and qL, though both chapters benefit from larger sample sizes ($n = 100$) in reducing imprecision.

While larger sample sizes ($n = 100$) reduce imprecision in both chapters, Chapter 4 consistently shows slightly more imprecision due to its derivation from the overall lower and upper γ values. This broader interval range reflects Chapter 4's robustness in capturing imprecision introduced by misspecification, supporting the method's capability to account for model uncertainty effectively across all cases.

4.5 Concluding remarks

This chapter introduces a new approach to nonparametric inference with imprecise probability for analyzing SSALT data. Unlike traditional methods, this approach does not depend on predefined failure time distributions for each stress level. Instead, the log-rank test is employed to assess pairwise stress level survival distributions, while the Arrhenius model is used to determine the range of γ values. By applying nonparametric tests to the Arrhenius model's link function parameter across different stress levels, this technique introduces imprecision, allowing transformation of data from higher stress levels into interval-valued data at the normal stress level, thus enhancing robustness through the use of the lower and upper NPI survival functions.

Key insights from this chapter include establishing an interval for the link function's

parameter at each stress level. This is achieved by conducting nonparametric hypothesis tests between pairs of stress levels to quantify imprecision. Preference is given to imprecision derived from combined pairwise log-rank tests rather than a single test applied across all stress levels, as the pairwise approach increases imprecision when the model fit is inadequate.

In comparing methods, it was concluded that the imprecision resulting from the log-rank test is slightly greater than that from the likelihood ratio test under the statistical model. This conclusion is supported by examples and simulation results. Performance evaluations through simulation studies indicate a high level of robustness in cases of model misspecification. However, the increased imprecision from the log-rank test suggests that further model investigation or additional data collection may be advisable in cases with substantial imprecision.

These methods assume the availability of data at the normal stress level, which may not always be practical. In such cases, it is recommended to perform tests between two data sets from higher stress strategies to establish an interval for the acceleration parameter. While complete data sets were assumed for simplicity, implementing these methods for right-censored data is straightforward. Right-censored data can be incorporated in the tests, then transformed to the normal stress level, allowing for the application of NPI techniques for right-censored data, as discussed in Section 2.7.

Chapter 5

Robust Bayesian inference using a class of priors

5.1 Introduction

In practical data analysis, particularly in reliability and step-stress accelerated life testing (SSALT), precise prior knowledge about model parameters is often unavailable or difficult to define. Traditional Bayesian methods require specific assumptions about these priors, which may be difficult to justify in cases where data is limited or when knowledge of the parameters is based on diverse sources. This reliance on exact priors can lead to conclusions that are highly sensitive to slight changes in assumptions, particularly when the likelihood or model specification is uncertain [8, 36]. To address these challenges, robust Bayesian analysis introduces imprecision into model assumptions, providing a framework to quantify the effects of varying assumptions on the analysis. Instead of relying on a single prior, this approach defines a range of plausible priors, leading to the derivation of lower and upper bounds for posterior and posterior predictive distributions. These bounds capture the uncertainty and imprecision inherent in the assumptions, allowing the analysis to reflect the imprecision introduced by different plausible scenarios. The primary objective of robust Bayesian analysis is to measure this imprecision explicitly, ensuring that the conclusions account for uncertainties in priors, likelihoods, and model assumptions, thereby leading to robust inferences.

This chapter introduces a robust Bayesian method that leverages a class of priors, particularly Beta distributions, to account for imprecise prior knowledge of the scale and

acceleration parameters in SSALT models. The Beta distribution is chosen due to its flexibility in modeling uncertainty over bounded parameters, which is appropriate for the scale and acceleration parameters in SSALT models that are constrained to lie within a known interval. Its shape can easily be adjusted to reflect varying degrees of belief or imprecision, making it a natural choice for robust Bayesian inference under prior uncertainty. Using a class of priors, rather than relying on a single, precise prior, allows the method to represent a broader spectrum of possible beliefs about these parameters, capturing diverse information sources and permitting a full range of values in cases where prior knowledge is absent [25]. By accommodating this flexibility, the analysis is able to address potential prior misspecification while supporting a more comprehensive framework that can integrate various perspectives and imprecise prior information.

The primary aim of this chapter is to develop a robust Bayesian method for analyzing step-stress accelerated life testing (SSALT) data by addressing sensitivity to prior misspecification and accounting for uncertainty in prior knowledge. This method utilizes imprecise probabilities to model vague prior knowledge about the model parameters, allowing for the derivation of both lower and upper bounds for the posterior distribution. By representing this imprecise knowledge with a class of prior distributions, where the extremes reflect a complete lack of knowledge about the parameters, the method provides a framework to quantify the effects of varying assumptions on the analysis. Consequently, it measures the imprecision arising from the prior, likelihood, and link function, offering a comprehensive evaluation of how these components influence the results. These bounds enable the derivation of lower and upper posterior predictive distributions, which facilitate predictions of future failure times at the normal stress level.

A significant objective of this approach is the construction of lower and upper predictive survival functions at normal stress conditions, providing robust predictive intervals that capture both data-driven variability and the imprecision arising from prior distributions, likelihoods, and link functions. This is particularly relevant in SSALT applications, where assumptions about these components may only approximate the real-world complexity of failure mechanisms [38]. By explicitly incorporating imprecision, the proposed method enhances the robustness of predictions by providing interval probabilities for future observations under uncertain conditions [6]. These interval probabilities are derived from the lower and upper posterior predictive distributions, offering a range of plausible

outcomes that reflect both data-driven variability and the imprecision introduced by all model components. The performance of the method is further evaluated through simulation studies, demonstrating its applicability in addressing variability and uncertainty in reliability and SSALT data analysis.

To implement this approach, Markov Chain Monte Carlo (MCMC) methods are employed to estimate the posterior distributions across the class of priors. This empirical Bayesian analysis quantifies the imprecision introduced by model assumptions and data variability, offering a robust framework for predictions at normal stress levels. An empirical Bayesian analysis is a statistical approach in which prior distributions are partially specified and then completed using information derived from the observed data. This allows the method to incorporate both prior structure and data-driven estimation in the inference process. In this work, MCMC methods are used because the analytical form of the posterior distribution is not tractable due to the complexity introduced by the class of priors and the SSALT model structure. Therefore, MCMC provides a practical numerical solution for approximating the posterior distributions under each prior within the class. The resulting methodology not only accounts for prior uncertainty but also allows practitioners to incorporate interval-based survival estimates in reliability applications, providing robust inference that accommodates uncertainties in complex modeling scenarios and offers a practical tool for robust Bayesian analysis [9, 43].

This chapter is organised as follows. Section 5.2 introduces the cumulative exposure model used for step-stress accelerated life testing (SSALT) data, describing its structure and the Bayesian framework applied to it. Section 5.3 details the Bayesian inference process, utilizing the Metropolis-Hastings algorithm to estimate model parameters and posterior distributions. Section 5.4 presents the proposed robust Bayesian method. Section 5.6 presents illustrative examples of the proposed method. Section 5.7 presents simulations studies to evaluate the performance of the proposed method. Section 5.8 presents concluding remarks.

5.2 The Bayesian model for SSALT Data

This section recalls the Weibull cumulative exposure model, as described in Section 2.4. This model is implemented to analyze step-stress accelerated life testing (SSALT) data,

where failure times are assumed to follow a Weibull distribution. The Weibull scale parameters at different stress levels are linked through the Arrhenius function, which relates lifetime to stress levels, typically represented as temperature.

The number of experimental strategies conducted under different stress settings is denoted by m. The probability density function (PDF) for failure times at stress levels $i = 0, 1, \ldots, s-1$ is expressed as:

$$f_i(t) = \left(\frac{\beta}{(\theta_i^m)^\beta}\right)\left(t - (\tau_i^m - \tau_{i^m}^*)\right)^{\beta-1} \exp\left[-\left(\frac{t-\tau_i^m}{\theta_i^m} + \sum_{j=1}^{i}\frac{\tau_j^m - \tau_{j-1}^m}{\theta_{j-1}^m}\right)^\beta\right], \tag{5.1}$$

for $\tau_i^m \leq t \leq \tau_{i+1}^m$, where t represents the failure time, and τ_i^m is the time at which the stress level changes during the m-th strategy. The term τ_i^{m*}, used to calculate the shift parameter, adjusts for the cumulative effects of previous stress levels in the Weibull cumulative exposure model, as explained in Section 2.4. This term is given by:

$$\tau_i^{m*} = \left[\frac{\theta_i^m}{\theta_{i-1}^m}\left(\tau_i^m - \tau_{i-1}^m + \tau_{i-1}^{m*}\right)\right] \tag{5.2}$$

for $\tau_i \leq t \leq \tau_{i+1}$. The shape parameter $\beta > 0$ is assumed to be constant across all stress levels. The scale parameter at each stress level θ_i^m is linked to the normal stress level scale parameter θ_0^0 through the Arrhenius function, expressed as:

$$\theta_i^m = \theta_0^0 \exp\left(\frac{\gamma}{k_i^m} - \frac{\gamma}{k_0^0}\right), \tag{5.3}$$

where k_i^m represents the temperature in Kelvin applied at the i-th stress level during strategy m, k_0^m is the temperature at the normal stress level, and γ is the acceleration parameter that quantifies the relationship between stress and the scale parameter.

The dataset of observed failure times is denoted by t, where

$$t = \{t_{0:1}^0, ..., t_{0:n_0^0}^0, t_{0:1}^1, ..., t_{0:n_0^1}^1, ..., t_{0:1}^2, ..., t_{0:n_0^2}^2, ..., t_{i:1}^2, ..., t_{i:n_0^2}^2, ..., t_{i:1}^m, ..., t_{i:n_i^m}^m\}, \tag{5.4}$$

with $t_{i:n_i^m}^m$ representing failure times observed at stress level i during strategy m. Each failure time $t_{i:n_i^m}^m$ is observed in the interval $[\tau_i^m, \tau_{i+1}^m]$, where τ_i^m denotes the time of the i-th stress level change, and τ_{i+1}^m denotes the start of the next stress level or the end of the test. Failure times at the normal stress level (t_0^0) follow the Weibull distribution without acceleration effects.

The likelihood function for the dataset t, given the parameters θ_i^m, β, γ, and k_i^m, is expressed as:

$$L(t;\theta_i^m,\beta,\gamma,k_i^m)=\prod_{m=0}^{z}\prod_{i=0}^{s-1}\prod_{n_i=1}^{n_i^m} f_i(t_{i:n_i^m}^m;\theta_i^m,\beta,\gamma,k_i^m), \tag{5.5}$$

where n_i^m is the number of failure times observed at stress level i during strategy m, and n represents the total number of failure times across all strategies.

The parameters θ_0^0, γ, and β are estimated to characterize the Weibull model under SSALT conditions. The time points τ_i^m, representing the predetermined changes in stress levels, and the predetermined stress levels k_i^m, applied during the experiments, are essential components defined by the experimental design.

The Bayesian approach provides a structured and principled framework for incorporating prior knowledge about model parameters and updating this knowledge in light of observed data. This process results in a posterior distribution, which represents a refined understanding that combines prior beliefs with evidence derived from the data. In this context, the analysis focuses on estimating three primary parameters: the scale parameter at normal stress (θ_0), the acceleration parameter (γ), and the shape parameter (β), which collectively characterize the failure time distribution under step-stress accelerated life testing (SSALT) conditions.

To model prior knowledge about the parameters θ_0, γ, and β, Beta distributions are assumed as priors. The Beta distribution is particularly well-suited for this purpose due to its flexibility in representing parameters constrained to a bounded interval [8]. Its ability to accommodate a wide range of shapes makes it an ideal choice for capturing diverse prior beliefs. For instance, a uniform prior can be specified by selecting equal shape parameters, while skewed or concentrated distributions can be represented by adjusting the shape parameters accordingly. This adaptability enables the modeling of prior uncertainty in a manner that aligns with theoretical considerations or expert judgment.

By adopting Beta priors, it becomes possible to construct a class of priors that encapsulate varying degrees of imprecision about the parameters θ_0, γ, and β. This approach not only accounts for potential uncertainty in the prior specification but also supports the exploration of sensitivity to different prior assumptions. Further details on the development and implementation of this class of priors are presented in Section 5.4, where the methodology for addressing prior imprecision is discussed in depth. This provides a ro-

bust foundation for Bayesian inference, enhancing the robustness of parameter estimation and predictive analysis in the presence of uncertain prior knowledge.

The scale parameter at normal stress, θ_0, is modeled using a Beta prior distribution. To ensure the parameter lies within the interval $[0, 1]$, a linear transformation is applied:

$$\theta_0^* = \frac{\theta_0 - a_{\theta_0}}{b_{\theta_0} - a_{\theta_0}}, \quad \theta_0 \in [a_{\theta_0}, b_{\theta_0}], \tag{5.6}$$

where a_{θ_0} and b_{θ_0} are the lower and upper bounds of the parameter range for θ_0. The transformed parameter θ_0^* follows a Beta prior:

$$\pi(\theta_0^*) = \frac{\theta_0^{*p_{\theta_0}-1}(1-\theta_0^*)^{q_{\theta_0}-1}}{B(p_{\theta_0}, q_{\theta_0})}, \quad \text{for } \theta_0^* \in [0, 1], \tag{5.7}$$

where $p_{\theta_0} > 0$ and $q_{\theta_0} > 0$ are the shape parameters of the Beta distribution, and $B(p_{\theta_0}, q_{\theta_0})$ is the Beta function, expressed as:

$$B(p_{\theta_0}, q_{\theta_0}) = \frac{\Gamma(p_{\theta_0})\Gamma(q_{\theta_0})}{\Gamma(p_{\theta_0} + q_{\theta_0})}. \tag{5.8}$$

The Gamma function, $\Gamma(p_{\theta_0})$, used in the Beta function, is given by:

$$\Gamma(p_{\theta_0}) = \int_0^\infty t^{p_{\theta_0}-1} e^{-t}\, dt, \quad \text{for } p_{\theta_0} > 0. \tag{5.9}$$

The acceleration parameter γ, quantifying the relationship between stress levels and failure times, is similarly modeled. The transformation to the interval $[0, 1]$ is given by:

$$\gamma^* = \frac{\gamma - a_\gamma}{b_\gamma - a_\gamma}, \quad \gamma \in [a_\gamma, b_\gamma], \tag{5.10}$$

with a Beta prior:

$$\pi(\gamma^*) = \frac{\gamma^{*p_\gamma-1}(1-\gamma^*)^{q_\gamma-1}}{B(p_\gamma, q_\gamma)}, \quad \text{for } \gamma^* \in [0, 1], \tag{5.11}$$

where $p_\gamma > 0$ and $q_\gamma > 0$ are the shape parameters of the Beta distribution.

The shape parameter β, which determines the form of the failure time distribution (e.g., exponential-like or heavy-tailed behavior), is assigned a uniform prior. Since the main objective of the analysis is to assess imprecision in the scale and acceleration parameters, assuming a non-informative prior for β ensures that the inference focuses on the parameters of interest without introducing additional assumptions or constraints through the prior on β. The transformation for β is expressed as:

$$\beta^* = \frac{\beta - a_\beta}{b_\beta - a_\beta}, \quad \beta \in [a_\beta, b_\beta], \tag{5.12}$$

with the uniform prior given by:

$$\pi(\beta^*) = \frac{\beta^{*1-1}(1-\beta^*)^{1-1}}{B(1,1)} = 1, \quad \text{for } \beta^* \in [0,1], \tag{5.13}$$

where the shape parameters of the Beta distribution are $p_\beta = 1$ and $q_\beta = 1$, corresponding to a uniform distribution over the interval $[0,1]$. This assumption reflects minimal prior information about β, as the focus is on incorporating imprecision in the scale (θ_0) and acceleration (γ) parameters, which are modeled using flexible Beta distributions. For β, a uniform prior is used across its plausible range, maintaining consistency in the analysis while enabling the measurement of imprecision in the other parameters, as will be discussed in Section 5.4.

Assuming independence among θ_0, γ, and β, the joint prior distribution for the transformed parameters is expressed as:

$$\pi(\theta_0^*, \gamma^*, \beta^*) = \pi(\theta_0^*) \cdot \pi(\gamma^*) \cdot \pi(\beta^*), \tag{5.14}$$

where:

$$\pi(\theta_0^*) = \frac{\theta_0^{*p_{\theta_0}-1}(1-\theta_0^*)^{q_{\theta_0}-1}}{B(p_{\theta_0}, q_{\theta_0})}, \quad \pi(\gamma^*) = \frac{\gamma^{*p_\gamma-1}(1-\gamma^*)^{q_\gamma-1}}{B(p_\gamma, q_\gamma)}, \quad \pi(\beta^*) = 1. \tag{5.15}$$

The bounds $a_{\theta_0}, b_{\theta_0}$, a_γ, b_γ, and a_β, b_β are chosen based on theoretical considerations, prior empirical evidence, and expert knowledge. These bounds ensure that the priors adequately represent the plausible range for each parameter while accommodating uncertainty.

This formulation establishes Beta priors for θ_0 and γ, allowing for the exploration of varying assumptions, while the uniform prior for β reflects the focus on constructing a class of priors primarily for the scale and acceleration parameters. Further details on sensitivity to prior imprecision are provided in Section 5.4.

The posterior distribution combines the likelihood and prior distributions, normalized by the marginal likelihood (denoted as W). The posterior distribution is expressed as:

$$\pi(\theta_0, \gamma, \beta \mid t) = \frac{L(t \mid \theta_0, \gamma, \beta) \cdot \pi\left(\frac{\theta_0 - a_{\theta_0}}{b_{\theta_0} - a_{\theta_0}}\right) \cdot \pi\left(\frac{\gamma - a_\gamma}{b_\gamma - a_\gamma}\right) \cdot \pi\left(\frac{\beta - a_\beta}{b_\beta - a_\beta}\right)}{W}, \tag{5.16}$$

where $L(t \mid \theta_0, \gamma, \beta)$ represents the likelihood function for the data t given the parameters θ_0, γ, and β. The terms $\pi\left(\frac{\theta_0 - a_{\theta_0}}{b_{\theta_0} - a_{\theta_0}}\right)$, $\pi\left(\frac{\gamma - a_\gamma}{b_\gamma - a_\gamma}\right)$, and $\pi\left(\frac{\beta - a_\beta}{b_\beta - a_\beta}\right)$ correspond to the priors for θ_0, γ, and β, respectively, modeled using Beta and uniform distributions through

linear transformations. The marginal likelihood W, also referred to as the normalizing constant, ensures the posterior distribution integrates to one and is defined as:

$$W = \int_{a_\beta}^{b_\beta} \int_{a_\gamma}^{b_\gamma} \int_{a_{\theta_0}}^{b_{\theta_0}} L(t \mid \theta_0, \gamma, \beta) \cdot \pi\left(\frac{\theta_0 - a_{\theta_0}}{b_{\theta_0} - a_{\theta_0}}\right) \cdot \pi\left(\frac{\gamma - a_\gamma}{b_\gamma - a_\gamma}\right) \cdot \pi\left(\frac{\beta - a_\beta}{b_\beta - a_\beta}\right) d\theta_0 \, d\gamma \, d\beta. \tag{5.17}$$

This formulation integrates the likelihood and prior distributions, enabling the posterior to capture both prior information and evidence from the observed data.

The posterior distribution can be expressed as:

$$\pi(\theta_0, \gamma, \beta \mid t) \propto L(t \mid \theta_0, \gamma, \beta) \cdot \pi\left(\frac{\theta_0 - a_{\theta_0}}{b_{\theta_0} - a_{\theta_0}}\right) \cdot \pi\left(\frac{\gamma - a_\gamma}{b_\gamma - a_\gamma}\right) \cdot \pi\left(\frac{\beta - a_\beta}{b_\beta - a_\beta}\right). \tag{5.18}$$

To focus on individual parameters, the marginal posterior distributions are obtained by integrating out the other two parameters from the joint posterior distribution, as shown in the following equations:

$$\pi(\theta_0 \mid t) = \int_{a_\beta}^{b_\beta} \int_{a_\gamma}^{b_\gamma} \pi(\theta_0, \gamma, \beta \mid t) \, d\gamma \, d\beta, \tag{5.19}$$

$$\pi(\gamma \mid t) = \int_{a_\beta}^{b_\beta} \int_{a_{\theta_0}}^{b_{\theta_0}} \pi(\theta_0, \gamma, \beta \mid t) \, d\theta_0 \, d\beta, \tag{5.20}$$

$$\pi(\beta \mid t) = \int_{a_\gamma}^{b_\gamma} \int_{a_{\theta_0}}^{b_{\theta_0}} \pi(\theta_0, \gamma, \beta \mid t) \, d\theta_0 \, d\gamma. \tag{5.21}$$

The Bayesian estimation of the parameters is derived from the posterior distributions. The posterior means for θ_0, γ, and β are given by:

$$E[\theta_0 \mid t] = \int_{a_{\theta_0}}^{b_{\theta_0}} \theta_0 \cdot \pi(\theta_0 \mid t) \, d\theta_0, \tag{5.22}$$

$$E[\gamma \mid t] = \int_{a_\gamma}^{b_\gamma} \gamma \cdot \pi(\gamma \mid t) \, d\gamma., \tag{5.23}$$

$$E[\beta \mid t] = \int_{a_\beta}^{b_\beta} \beta \cdot \pi(\beta \mid t) \, d\beta. \tag{5.24}$$

Since the posterior distribution does not have a closed-form solution, numerical methods are essential to estimate the parameters effectively. Bayesian estimation typically relies on minimizing the squared error loss function, which corresponds to calculating the expected value of the posterior distribution. When analytical posterior distributions are

unavailable, numerical methods become essential. In such cases, sampling-based techniques like the Metropolis-Hastings algorithm, a Markov Chain Monte Carlo (MCMC) method, are employed to approximate the posterior distributions. This algorithm is used to sample from the posterior distribution and estimate parameter values. The parameter vectors $(\theta_0, \gamma, \beta)$ are estimated using this approach, as described in Section 5.3.

The Bayesian framework also facilitates the prediction of new data points based on the observed data. The posterior predictive distribution is derived by integrating the likelihood of new data over the posterior distribution of the parameters. Assuming that t^* is conditionally independent of the observed data t given $(\theta_0, \gamma, \beta)$. For a new data point t^*, the posterior predictive distribution is given by:

$$p(t^* \mid t) = \int_{a_\beta}^{b_\beta} \int_{a_\gamma}^{b_\gamma} \int_{a_{\theta_0}}^{b_{\theta_0}} p(t^* \mid \theta_0, \gamma, \beta) \cdot \pi(\theta_0, \gamma, \beta \mid t)\, d\theta_0\, d\gamma\, d\beta. \tag{5.25}$$

The computation of the posterior predictive distribution requires numerical methods, as an analytical solution is generally not available. Samples from the posterior distribution are generated using the Metropolis-Hastings algorithm, a Markov Chain Monte Carlo (MCMC) method, to approximate the posterior predictive distribution. This approach enables the prediction of future failure times at the normal stress level. The implementation of this method is discussed in detail in Section 5.3.

The bounds $a_{\theta_0}, b_{\theta_0}$, a_γ, b_γ, and a_β, b_β are critical components in defining the prior distributions and are selected based on theoretical considerations, prior empirical evidence, and expert knowledge [31]. These bounds ensure that the priors adequately capture the plausible range of each parameter while addressing inherent uncertainty in their values. While the theoretical range of these parameters may extend from 0 to infinity, such unbounded intervals are not practical for the modeling process. Assigning an infinite range to the parameters would lead to overly vague priors that fail to effectively incorporate any prior knowledge.

Instead, practical bounds are chosen to reflect the realistic and plausible range of the parameter values based on domain-specific insights. For instance, the lower bounds are selected to be values close to zero, while the upper bounds are set to large numbers, such as one million $(1,000,000)$, to approximate a state of minimal prior knowledge about the parameters. With the chosen bounds $a_{\theta_0}, b_{\theta_0}$, a_γ, b_γ, and a_β, b_β, the initial hyperparameters of the Beta distributions are uniformly set to 1 (i.e., $p_{\theta_0} = q_{\theta_0} = 1$,

$p_\gamma = q_\gamma = 1$, $p_\beta = q_\beta = 1$). This assumption reflects a uniform prior distribution over the transformed parameter space $[0, 1]$, ensuring that the priors exert minimal influence on the posterior distribution. The use of uniform priors allows the observed data to play a dominant role in precise parameter estimation and in identifying the plausible range for the parameters.

Moreover, the hyperparameters of the Beta distributions are used to establish a class of priors that encapsulates varying levels of belief about the parameters within these bounds. This methodology forms an essential aspect of the Bayesian approach presented in this work, as it enables the modeling of uncertainty and imprecision in the priors. The selection of these bounds is particularly important for constructing robust and flexible prior distributions that can be used to evaluate the sensitivity of the posterior to different prior assumptions.

To assess the plausibility of these bounds and the corresponding class of priors, posterior predictive checks are conducted using the Metropolis-Hastings algorithm within the Markov Chain Monte Carlo (MCMC) framework. These checks evaluate how well the selected priors align with the observed data and help refine the bounds to measure the imprecision resulting from assumptions in the posterior distribution, such as the likelihood function and the link function. Through simulations, the full range of plausible parameter values is explored, allowing for a comprehensive assessment of their impact on the posterior distribution. The implementation and analysis of these checks are discussed in detail in Section 5.3. This methodological approach underscores the significance of carefully chosen parameter bounds in addressing uncertainty and imprecision within the Bayesian framework presented in this study.

5.3 Bayesian parameter estimation using MCMC

The posterior distributions derived in Bayesian inference often lack closed-form solutions, necessitating numerical methods for estimation. One such method is Markov Chain Monte Carlo (MCMC), a powerful class of algorithms used to generate samples from complex posterior distributions. Among these, the Metropolis-Hastings algorithm is widely recognized for its flexibility and effectiveness in Bayesian analysis. This algorithm enables the approximation of posterior distributions by constructing a Markov chain whose

stationary distribution matches the target posterior distribution.

This section focuses on implementing the Metropolis-Hastings algorithm to approximate posterior distributions of the parameters θ_0, γ, and β, which characterize the Weibull cumulative exposure model for step-stress accelerated life testing (SSALT). Additionally, the algorithm is employed to compute the posterior predictive distribution and facilitate posterior predictive checks, providing insights into the plausibility of model assumptions.

5.3.1 Utilizing the Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm is a widely used Markov Chain Monte Carlo (MCMC) method for sampling from posterior distributions when direct analytical solutions are intractable. In this study, it is applied to generate samples from the posterior distribution $\pi(\theta_0, \gamma, \beta \mid t)$, where the parameters θ_0, γ, and β characterize the Weibull cumulative exposure model. The algorithm iteratively proposes new parameter values and accepts or rejects them based on their likelihood under the target posterior distribution. This ensures the Markov chain converges to the true posterior distribution after sufficient iterations.

For this implementation, a Normal distribution is chosen as the proposal distribution for each parameter. A Normal distribution is selected as the proposal distribution due to its symmetric shape and ease of implementation. It provides efficient random walk proposals for continuous parameters like θ_0, γ, and β. Additionally, the variance of the Normal proposal can be tuned to control the step size, which directly affects the mixing and convergence behavior of the chain. At each iteration i, the proposed parameter values θ_0', γ', and β' are drawn as:

$$\theta_0' \sim \mathcal{N}(\theta_0^{(i-1)}, \sigma_{\theta_0}^2), \quad \gamma' \sim \mathcal{N}(\gamma^{(i-1)}, \sigma_\gamma^2), \quad \beta' \sim \mathcal{N}(\beta^{(i-1)}, \sigma_\beta^2),$$

where $\sigma_{\theta_0}^2$, σ_γ^2, and σ_β^2 are the variances of the proposal distributions for θ_0, γ, and β, respectively. These variances are critical tuning parameters that influence the efficiency of the algorithm.

The choice of variances is crucial. Large variances may result in frequent rejection of proposed values, leading to poor mixing, while small variances may slow the exploration of the posterior distribution. To ensure a balance between efficient exploration of the parameter space and an acceptance rate within the optimal range of 20% to 40%, the variances

$\sigma^2_{\theta_0}$, σ^2_γ, and σ^2_β are adjusted dynamically rather than being fixed at a specific value [27]. This adaptive approach allows the algorithm to maintain an appropriate acceptance rate throughout the sampling process. Initial values for the parameters, $(\theta_0^{(0)}, \gamma^{(0)}, \beta^{(0)})$, are set to their maximum likelihood estimates (MLEs), ensuring a reasonable starting point for the Markov chain. The Metropolis-Hastings algorithm, as detailed in Algorithm 1, proceeds as follows:

Algorithm 1 Metropolis-Hastings Algorithm for Posterior Approximation

Require: Observed data t, initial parameter values $(\theta_0^{(0)}, \gamma^{(0)}, \beta^{(0)})$, proposal distributions $q(\cdot \mid \cdot)$, number of iterations N_{MH}, burn-in period N_{burn}.

Ensure: Posterior samples $(\theta_0^{(i)}, \gamma^{(i)}, \beta^{(i)})$ for $i = 1, \ldots, N_{MH}$.

1: Initialize the chain with starting values $(\theta_0^{(0)}, \gamma^{(0)}, \beta^{(0)})$.

2: **for** iteration $i = 1$ to N_{MH} **do**

3: Propose new parameter values $(\theta_0', \gamma', \beta')$ from the normal proposal distributions:

$$\theta_0' \sim \mathcal{N}(\theta_0^{(i-1)}, \sigma^2_{\theta_0}), \quad \gamma' \sim \mathcal{N}(\gamma^{(i-1)}, \sigma^2_\gamma), \quad \beta' \sim \mathcal{N}(\beta^{(i-1)}, \sigma^2_\beta).$$

4: Calculate the acceptance ratio:

$$\rho = \frac{\pi(\theta_0', \gamma', \beta' \mid t) \cdot q(\theta_0^{(i-1)}, \gamma^{(i-1)}, \beta^{(i-1)} \mid \theta_0', \gamma', \beta')}{\pi(\theta_0^{(i-1)}, \gamma^{(i-1)}, \beta^{(i-1)} \mid t) \cdot q(\theta_0', \gamma', \beta' \mid \theta_0^{(i-1)}, \gamma^{(i-1)}, \beta^{(i-1)})}. \tag{5.26}$$

5: Generate a random value $u \sim \text{Uniform}(0, 1)$.

6: **if** $u \leq \min(1, \rho)$ **then**

7: Accept the proposal: $(\theta_0^{(i)}, \gamma^{(i)}, \beta^{(i)}) = (\theta_0', \gamma', \beta')$.

8: **else**

9: Retain the current values: $(\theta_0^{(i)}, \gamma^{(i)}, \beta^{(i)}) = (\theta_0^{(i-1)}, \gamma^{(i-1)}, \beta^{(i-1)})$.

10: **end if**

11: **end for**

12: **return** Posterior samples $(\theta_0^{(i)}, \gamma^{(i)}, \beta^{(i)})$ for $i = 1, \ldots, N_{MH}$.

Once the Markov chain has been generated, the initial portion of the chain, known as the burn-in period, is discarded to eliminate the influence of the starting values. For this analysis, a burn-in period of $N_{burn} = 1000$ iterations is applied. The remaining

$N_{MH} - N_{burn}$ samples are used for posterior analysis, with $N_{MH} = 10,000$ ensuring sufficient mixing and accurate estimation. To support the claim of sufficient convergence and mixing for the MCMC procedure used in Example 5.6.1 (Case 1, Chapter 5), trace plots for the parameters θ_0 and γ are provided in Figures B.1 and B.2 in the Appendix. These plots visually confirm good mixing behavior and stability of the posterior chains after burn-in.

The posterior means for the parameters are calculated as the average of the retained samples:

$$\hat{\theta}_0 = \frac{1}{N_{MH} - N_{burn}} \sum_{i=N_{burn}+1}^{N_{MH}} \theta_0^{(i)}, \tag{5.27}$$

$$\hat{\gamma} = \frac{1}{N_{MH} - N_{burn}} \sum_{i=N_{burn}+1}^{N_{MH}} \gamma^{(i)}, \tag{5.28}$$

$$\hat{\beta} = \frac{1}{N_{MH} - N_{burn}} \sum_{i=N_{burn}+1}^{N_{MH}} \beta^{(i)}. \tag{5.29}$$

These posterior means provide point estimates for the parameters, enabling precise inference based on the observed data and the chosen prior distributions.

5.3.2 Generating future data from the posterior predictive distribution

The generation of future datasets from the posterior predictive distribution involves sampling new data points t^* by integrating over the posterior samples of the parameters. The steps for numerically generating future datasets are outlined below.

This algorithm uses the posterior samples obtained via the Metropolis-Hastings algorithm to generate future datasets. Each new data point t_j^* is drawn from the likelihood function conditional on a randomly selected posterior sample of the parameters. The process is repeated N_{future} times to generate the required number of future data points.

The predictive distribution of the future data points reflects both the uncertainty in the parameter estimates and the variability in the data generation process. This ensures that the generated datasets are consistent with the observed data and the Bayesian model assumptions.

Algorithm 2 Generating Future Data via Posterior Predictive Distribution

Require: Posterior samples of the parameters $(\theta_0^{(i)}, \gamma^{(i)}, \beta^{(i)})$ for $i = 1, \ldots, N_{MH}$, likelihood function $p(t^* \mid \theta_0, \gamma, \beta)$, number of future data points N_{future}.

Ensure: Future data samples $t^* = \{t_1^*, t_2^*, \ldots, t_{N_{future}}^*\}$.

1: Initialize an empty set for future data samples t^*.
2: **for** $j = 1$ to N_{future} **do**
3: Randomly select a posterior sample $(\theta_0^{(i)}, \gamma^{(i)}, \beta^{(i)})$, where $i \sim \text{Uniform}(1, N_{MH})$.
4: Sample a new data point t_j^* from the likelihood function $p(t^* \mid \theta_0^{(i)}, \gamma^{(i)}, \beta^{(i)})$.
5: Append t_j^* to the set t^*.
6: **end for**
7: **return** Future data samples $t^* = \{t_1^*, t_2^*, \ldots, t_{N_{future}}^*\}$.

5.3.3 Posterior predictive checks

Posterior predictive checks are used to assess the plausibility of parameter bounds and refine prior distributions. This process involves generating replicated datasets from the posterior predictive distribution and analyzing their properties. Specifically, the approach focuses on extreme datasets—those with the earliest failure times (minimum datasets) and delayed failure times (maximum datasets)—to refine the ranges for the scale parameter (θ_0), acceleration parameter (γ), and shape parameter (β). Posterior predictive checks are an essential aspect of Bayesian data analysis, allowing for the evaluation of model adequacy and sensitivity to assumptions. As discussed by Gelman et al. [27], these checks involve comparing replicated datasets generated from the posterior predictive distribution to the observed data, helping assess the plausibility of the model while refining prior assumptions and parameter bounds.

In this analysis, both the scale parameter and the acceleration parameter are assigned uniform prior distributions to reflect minimal prior knowledge $\theta_0 \sim \text{Unif}(1, 1{,}000{,}000)$ and $\gamma \sim \text{Unif}(1, 1{,}000{,}000)$. For the shape parameter (β), a uniform distribution is assumed over the range $[0.5, 5]$. These prior distributions aim to represent plausible parameter values based on empirical evidence and are validated through posterior predictive checks. This methodology provides a flexible framework to explore imprecision in prior assumptions and improve parameter estimates through an iterative refinement process.

The process, as outlined in Algorithm 3, involves generating replicated datasets from

the posterior predictive distribution. The following algorithm outlines the procedure:

1. Generate replicated datasets from the posterior predictive distribution.
2. Sort the datasets to identify the maximum and minimum datasets.
3. Fit these extreme datasets to the posterior distribution to obtain refined parameter ranges.
4. Update the parameter bounds based on these refined ranges.

Replicated datasets are generated to explore the plausible ranges for the parameters. Sorting these datasets into extremes provides insights into the behavior of the model under early and delayed failure scenarios. The minimum dataset ($t_{\min}$) represents scenarios where failures occur early, leading to smaller acceleration parameter values (γ) and scale parameter values (θ_0), indicating rapid failure rates. Meanwhile, the shape parameter (β) is larger due to the compressed failure times. In contrast, the maximum dataset ($t_{\max}$), which is characterized by delayed failures, exhibits larger acceleration (γ) and scale (θ_0) parameters, reflecting slower failure rates, while the shape parameter (β) is smaller, indicating extended failure times.

This iterative process generates extreme datasets, which are then fitted back into the model using Algorithm 1 to refine the parameter bounds by assessing imprecision introduced by vague priors and the likelihood function. By replicating data from the posterior predictive distribution, the method explores the variability of unobserved datasets, capturing plausible ranges for the parameters. This approach not only quantifies the impact of prior assumptions but also provides a means to validate the adequacy of the model.

This methodology aligns with the Bayesian framework and posterior predictive checks outlined by Gelman et al. [27]. Through careful replication and evaluation, it ensures that the priors reflect realistic ranges for the parameters while maintaining sufficient flexibility for robust posterior inference.

Algorithm 3 Refining Parameter Bounds Using Posterior Predictive Checks

Require: Posterior samples $(\theta_0^{(i)}, \gamma^{(i)}, \beta^{(i)})$, number of replicated datasets $R = 10,000$.

Ensure: Refined parameter bounds for θ_0, γ, and β.

1: Generate R replicated datasets $t^{*\text{rep},r}$ from the posterior predictive distribution:

$$t^{*\text{rep},r} \sim p(t^{*\text{rep}} \mid \theta_0^{(i)}, \gamma^{(i)}, \beta^{(i)}), \quad r = 1, \ldots, R.$$

2: Identify the maximum dataset ($t_{\max}$) and the minimum dataset ($t_{\min}$) by sorting $t^{*\text{rep},r}$.

3: Fit $t_{\max}$ and $t_{\min}$ to the posterior distribution to obtain:

- **Minimum dataset ($t_{\min}$):** Yields:
  - Extreme lower bounds for γ, and θ_0.
  - Extreme upper bounds for β.
- **Maximum dataset ($t_{\max}$):** Yields:
  - Extreme upper bounds for γ, and θ_0.
  - Extreme lower bounds for β.

4: Refine the prior bounds $a_{\theta_0}, b_{\theta_0}, a_\gamma, b_\gamma, a_\beta, b_\beta$ based on the estimation of the parameters from the extreme data sets using Algorithm 1.

5: **return** Updated parameter bounds.

5.4 Imprecision based on a class of prior distributions

The empirical Bayesian approach employed in this study begins with the absence of prior knowledge about the parameters of interest: the scale parameter (θ_0), acceleration parameter (γ), and shape parameter (β). In such cases, vague priors are used to define a broad parameter space. Beta priors with hyperparameters $p = q = 1$ (representing a

uniform distribution and equal weight across all values within the bounds) are chosen to initiate posterior predictive checks. These checks allow for the exploration of plausible parameter ranges by replicating data from the posterior predictive distribution, as described in Section 5.3.3.

This section introduces a robust Bayesian framework designed to address imprecision in parameter estimation by leveraging a class of prior distributions. The methodology combines the adaptability of empirical Bayesian analysis with the robustness of imprecise probabilities, facilitating the derivation of lower and upper posterior distributions. A key innovation of this approach lies in employing extreme priors to define a class of Beta distributions for each parameter, thereby capturing the uncertainty and imprecision associated with Bayesian inference for SSALT data. These lower and upper posterior distributions form the foundation for defining bounds on posterior predictive distributions, which are instrumental in predicting future datasets at the normal stress level.

After determining the plausible lower and upper bounds for the parameters θ_0 (scale), γ (acceleration), and β (shape) through posterior predictive checks, a class of Beta priors, denoted as $\mathcal{B}$, is introduced. The use of Beta distributions provides the flexibility to model prior knowledge in terms of imprecise probabilities. Beta distributions can represent a range of plausible beliefs about the parameters. The shape parameters p and q control the degree of skewness and concentration. For vague priors, setting $p = q = 1$ results in a uniform distribution over the defined range. Alternatively, when expert-informed priors are available, p and q can be adjusted to reflect the expert's beliefs, skewing the distribution toward specific parameter values. This approach enables the incorporation of expert knowledge when available, while still allowing for the exploration of imprecise priors for robust Bayesian analysis.

The Beta distribution is used to represent prior beliefs on bounded parameters, with the total fixed evidence defined as the sum of shape parameters $v = p + q$. This total reflects the strength of prior information while allowing flexibility in expressing different beliefs. For example, fixing $v = 10$, the priors Beta$(5, 5)$, Beta$(8, 2)$, and Beta$(2, 8)$ reflect the same level of confidence but express beliefs centered at 0.5, 0.8, and 0.2 respectively, illustrating how prior opinions can shift toward extreme scenarios while maintaining the same weight of evidence.

This class is formally defined as:

$$\mathcal{B} = \{\text{Beta}(p, q) \mid p + q = v,\ p > 0,\ q > 0\},$$

where p and q are the shape parameters of the Beta distribution, and $v = p+q$ represents the total fixed evidence or information encapsulated by the prior [4]. This total reflects the strength of prior information while allowing flexibility in expressing different beliefs. For example, fixing $v = 10$, the priors $\text{Beta}(5, 5)$, $\text{Beta}(8, 2)$, and $\text{Beta}(2, 8)$ reflect the same level of confidence but express beliefs centered at 0.5, 0.8, and 0.2 respectively, illustrating how prior opinions can shift toward extreme scenarios while maintaining the same weight of evidence. The scale parameter, transformed to the interval $[0, 1]$, follows a Beta distribution such that $(\theta_0 - a_{\theta_0})/(b_{\theta_0} - a_{\theta_0}) \sim \text{Beta}(p_{\theta_0}, q_{\theta_0})$, where the lower and upper bounds for the scale parameter, denoted as a_{θ_0} and b_{θ_0}, respectively, define the parameter space. The class of priors for the scale parameter is defined as:

$$\mathcal{B}_{\theta_0} = \{\text{Beta}(p_{\theta_0}, q_{\theta_0}) \mid p_{\theta_0} + q_{\theta_0} = v_{\theta_0},\ p_{\theta_0} > 0,\ q_{\theta_0} > 0\},$$

where p_{θ_0} and q_{θ_0} are the shape parameters of the Beta distribution, and $v_{\theta_0} = p_{\theta_0} + q_{\theta_0}$ represents the total fixed evidence or information encapsulated by the prior for the scale parameter. This flexible formulation allows the Beta distribution to vary between extreme cases, capturing a range of beliefs about θ_0.

This class of priors contributes to the posterior and posterior predictive distributions by capturing the impact of extreme scenarios. When the shape parameter of the Beta distribution satisfies $q_{\theta_0} \gg p_{\theta_0}$, the prior is heavily skewed toward the lower bound a_{θ_0}, reflecting smaller scale values. This configuration leads to posterior and posterior predictive distributions that favor early failure times, which are consistent with compressed failure mechanisms. Conversely, when $p_{\theta_0} \gg q_{\theta_0}$, the prior is heavily skewed toward the upper bound b_{θ_0}, reflecting larger scale values. In this case, the posterior and posterior predictive distributions favor delayed failure times, corresponding to longer lifetimes.

By systematically varying p_{θ_0} and q_{θ_0} within the class of Beta priors $\mathcal{B}_{\theta_0}$, this approach quantifies the imprecision in the posterior distribution caused by uncertainty in the prior assumptions. Specifically, a lower posterior distribution based on priors skewed toward the lower bound predicts smaller scale values, leading to shorter expected lifetimes. On the other hand, an upper posterior distribution based on priors skewed toward the upper bound predicts larger scale values, resulting in longer expected lifetimes.

The class of priors for the acceleration parameter γ is developed to capture the imprecision associated with estimating this parameter, which represents the activation energy affecting failure mechanisms. The acceleration parameter is transformed to the interval $[0, 1]$ and modeled using a Beta distribution, where $(\gamma - a_\gamma)/(b_\gamma - a_\gamma)$ follows a Beta distribution with shape parameters p_γ and q_γ. Here, a_γ and b_γ represent the lower and upper bounds of the acceleration parameter, respectively, defining its range. The class of Beta priors for γ is expressed as:

$$\mathcal{B}_\gamma = \{\text{Beta}(p_\gamma, q_\gamma) \mid p_\gamma + q_\gamma = v_\gamma,\ p_\gamma > 0,\ q_\gamma > 0\},$$

where p_γ and q_γ are the shape parameters of the Beta distribution, and $v_\gamma = p_\gamma + q_\gamma$ represents the total fixed evidence encapsulated by the prior. This structure allows flexibility to represent a range of beliefs about γ, from vague to concentrated priors.

This class of priors influences the posterior and posterior predictive distributions by capturing the effects of extreme values of γ. When $p_\gamma \gg q_\gamma$, the prior is heavily skewed toward the upper bound b_γ, which corresponds to scenarios with lower acceleration values. As a result, failure times are delayed, leading to longer lifetimes. Consequently, the posterior and posterior predictive distributions favor outcomes where lifetimes are extended due to slower failure mechanisms.

In contrast, when $q_\gamma \gg p_\gamma$, the prior is heavily skewed toward the lower bound a_γ, reflecting scenarios with higher acceleration values. This configuration results in increased failure rates and earlier failure times, as the posterior and posterior predictive distributions align with more rapid failure mechanisms.

By varying p_γ and q_γ within the defined class of priors $\mathcal{B}_\gamma$, the approach measures the imprecision introduced by uncertain prior assumptions. Specifically, a lower posterior distribution, derived from priors skewed toward the lower bound, predicts lower values of γ and shorter expected lifetimes. Conversely, an upper posterior distribution, derived from priors skewed toward the upper bound, predicts higher values of γ and longer expected lifetimes.

The initial estimation of the shape parameter β is performed under a uniform prior within the range $[0.5, 5]$. Once estimated, the shape parameter is fixed to its posterior mean value, as the aim is to measure the imprecision related to the failure mechanisms resulting from the scale (θ_0) and acceleration (γ) parameters, in line with the developed approaches in Chapter 3 and Chapter 4. By fixing β, the focus shifts to modeling the

uncertainty in the scale and acceleration parameters, which are allowed to vary within the defined class of Beta priors. This analysis reflects the methodological emphasis on exploring the impact of imprecise priors for parameters that are most sensitive in the likelihood function and the link function.

The posterior distributions for the scale (θ_0) and acceleration (γ) parameters allow for the derivation of Bayesian estimates under varying prior assumptions. These posterior distributions are obtained separately, with the lower posterior distribution derived using priors heavily skewed toward the lower bounds for the scale parameter and the lower bounds for the acceleration parameter. The resulting estimates, denoted as $E_{\text{lower}}[\theta_0, \gamma \mid t]$, correspond to early failure mechanisms, as smaller scale values and acceleration values lead to compressed failure times under accelerated failure conditions.

Conversely, the upper posterior distribution is derived using priors heavily skewed toward the upper bounds for the scale parameter and the upper bounds for the acceleration parameter. This configuration produces estimates $E_{\text{upper}}[\theta_0, \gamma \mid t]$, reflecting delayed failure mechanisms. Here, larger scale values and larger acceleration values correspond to extended lifetimes and reduced failure rates, indicative of slower failure mechanisms.

The posterior predictive distributions, derived from these separately obtained posterior distributions, encapsulate the range of possible outcomes. The lower posterior predictive distribution, $p_{\text{lower}}(t^* \mid t)$, predicts failure times aligned with accelerated and early failure scenarios. In contrast, the upper posterior predictive distribution, $p_{\text{upper}}(t^* \mid t)$, predicts failure times associated with delayed failure scenarios, driven by slower failure mechanisms and longer lifetimes.

This methodology provides a robust framework for analyzing uncertainty in step-stress accelerated life testing (SSALT) data. By systematically deriving the lower and upper posterior distributions separately, this approach captures the full range of potential failure scenarios. These predictions reflect both accelerated and delayed failure conditions, enabling comprehensive analysis of the imprecision associated in SSALT data.

5.5 Lower and upper posterior predictive empirical survival functions

The construction of empirical survival functions based on the posterior predictive distributions is a critical step in quantifying uncertainty in survival probabilities at the normal stress level. These predictive empirical survival functions enable the prediction of future observations under normal stress conditions, providing robust estimates of survival probabilities while explicitly accounting for the imprecision associated with the data. This approach explicitly models the uncertainty resulting from failure mechanisms in step-stress accelerated life testing (SSALT) data.

The lower posterior predictive empirical survival function, $\underline{S}(t)$, is derived from the lower posterior predictive distribution. The process begins by generating multiple replicated datasets, t^*_{lower}, at the normal stress level. Specifically, n_d replicated datasets are generated to capture the uncertainty in the lower posterior predictive distribution. Each replicated dataset consists of n_o predicted failure times, representing observations at the normal stress level. For a single dataset, the survival probability at time t is computed as the proportion of the n_o predicted failure times that exceed t. To incorporate the variability across all n_d datasets, the average survival probability at time t is calculated. This results in the construction of the lower posterior predictive empirical survival function as:

$$\underline{S}(t) = \frac{1}{n_d} \sum_{i=1}^{n_d} P[T^*_{\text{lower},i,j} \geq t], \tag{5.30}$$

where $T^*_{\text{lower},i,j}$ denotes the j-th predicted failure time from the i-th replicated dataset, t is the time of interest, n_d is the total number of replicated datasets, and n_o is the number of observations in each dataset. This survival function captures early failure scenarios, as it is based on smaller scale and smaller acceleration parameter values, which compress the failure times.

Similarly, the upper posterior predictive empirical survival function, $\overline{S}(t)$, is constructed using the upper posterior predictive distribution. The process begins by generating multiple replicated datasets, t^*_{upper}, at the normal stress level. Specifically, n_d replicated datasets are generated to capture the uncertainty in the upper posterior predictive distribution. For each dataset, survival probabilities are computed for n_o observations,

and the average survival probability at time t across all n_d datasets is calculated. This results in the upper posterior predictive empirical survival function, defined as:

$$\overline{S}(t) = \frac{1}{n_d} \sum_{i=1}^{n_d} P[T^*_{\text{upper},i,j} \geq t], \tag{5.31}$$

where $T^*_{\text{upper},i,j}$ denotes the j-th predicted failure time from the i-th replicated dataset, t is the time of interest, n_d is the total number of replicated datasets, and n_o is the number of observations in each dataset. This survival function captures delayed failure scenarios, as it is based on larger scale and larger acceleration parameter values, which extend the failure times. Together, these survival functions encapsulate the range of uncertainty in survival probabilities at the normal stress level.

The process for constructing the lower and upper posterior predictive empirical survival functions is outlined in the following algorithm:

Algorithm 4 Construction of Empirical Survival Functions

Require: Posterior predictive distribution (lower or upper), number of replicated datasets n_d, number of observations per dataset n_o, time points t

Ensure: Empirical survival function $S(t)$

1: Initialize an empty list to store survival probabilities across datasets.
2: **for** $i = 1$ to n_d **do**
3: Generate a new dataset t_0^{*0} consisting of n_o failure times from the specified posterior predictive distribution.
4: Compute the survival probability $P[T_{i,j} \geq t]$ for each observation in the dataset.
5: Compute the survival probability for the dataset as the proportion of observations exceeding t.
6: Append the dataset survival probability to the list.
7: **end for**
8: Compute the average survival probability across all datasets:

$$S(t) = \frac{1}{n_d} \sum_{i=1}^{n_d} P[T_{i,j} \geq t].$$

9: **return** Empirical survival function $S(t)$

The lower empirical survival function highlights early failure scenarios caused by smaller scale and acceleration parameters, while the upper empirical survival function

captures delayed failure scenarios associated with larger scale and acceleration parameters. The lower survival function represents a pessimistic scenario indicating early failures, while the upper survival function reflects an optimistic scenario, indicating extended lifetimes and delayed failures. The difference between these two survival functions quantifies the imprecision arising from the uncertainty in the data, the assumed likelihood, and the link function. If the assumed likelihood and link function are correct, the imprecision will be minimal, suggesting consistency between the model and the observed data. Conversely, significant differences between the lower and upper survival functions may indicate potential misspecification in the likelihood or the link function.

It should be emphasized that the resulting inferences of this method indicate that the lower posterior distribution is always less than the upper posterior distribution due to the effect of the extreme priors. This relationship propagates to the posterior predictive distributions, where the lower posterior predictive distribution is less than the upper posterior predictive distribution. Consequently, the lower empirical predictive survival function is always less than the upper empirical predictive survival function, which can be formally expressed as:

$$\underline{S}(t) < \overline{S}(t), \quad \forall t.$$

This inequality reflects the imprecision introduced by the extreme priors and will be illustrated through several examples in Section 5.6. The derivation of this result requires a mathematical proof, which we propose as a topic for future research.

Furthermore, the performance of this approach will be thoroughly evaluated via simulation studies in Section 5.7. These studies will explore the behavior of the predictive empirical survival functions under various assumptions about the likelihood, link function, and parameter ranges, providing deeper insights into the robustness of the proposed method.

5.6 Illustrative examples

This section presents two examples to illustrate the proposed method outlined in Section 5.4. These examples align with the experimental settings and data used in Chapters 3 and 4 to facilitate a comprehensive comparison of the results obtained using the three methods.

In Example 5.6.1, three datasets are generated with $n = 10$ observations per stress strategy. These strategies consist of two accelerated stress levels, along with a dataset collected at the normal stress level. In Example 5.6.2, the sample size is increased to $n = 100$ to evaluate the impact of larger sample sizes on the proposed method.

In previous chapters, comparisons between datasets relied on pairwise assumptions, as seen in the likelihood ratio test and log-rank test methods presented in Chapters 3 and 4. These methods required distinct datasets across stress levels to establish comparisons, making the inclusion of specific datasets, such as those at the normal stress level, essential for analysis.

In contrast, the method introduced in this chapter does not require pairwise comparisons or the inclusion of specific datasets, such as those at the normal stress level. Instead, this approach allows for greater flexibility in dataset selection, enabling comparisons that are not restricted to predefined datasets. While incorporating data at the normal stress level can enhance the analysis and provide additional insights, it is not a requirement for the proposed method, highlighting its adaptability and broader applicability. This flexibility makes the approach particularly advantageous when dealing with limited data or even a single dataset. However, in this chapter, pairwise datasets are included, as in the previous chapters, purely for comparative purposes. It is important to emphasize that such pairwise datasets are not necessary for the implementation of the proposed method, demonstrating its adaptability to diverse data scenarios.

Furthermore, while the method does not require equal sample sizes at each stress strategy, for consistency and simplicity in these examples, equivalent sample sizes n are assumed across all stress levels. This setup ensures that the comparisons are straightforward and the results are directly comparable across the different methods. By maintaining consistent datasets across Chapters 3, 4, and this chapter, the examples provide a clear basis for comparing the performance and conclusions of the three methods under identical experimental settings.

Example 5.6.1 This example consists of four cases. In Case 1, the shape parameter is assumed to be constant and known for each strategy. The Arrhenius link function is implemented to connect the scale parameters at all stress levels within the strategies, both for the simulated data and for the analysis. In Case 2, the shape parameter is considered unknown and is estimated during the analysis. In Case 3, the Arrhenius link function is

replaced by the Eyring link function. This replacement is done to evaluate the model's performance when the assumed link function does not provide a good fit. The focus of Case 3 is to examine the interval $[\underline{\gamma}_{m,0}, \overline{\gamma}_{m,0}]$ and its impact on the related lower and upper empirical predictive survival functions at the normal stress level. In Case 4, the shape parameter is set to 1, assuming an exponential model to test the method under another misspecification with the Eyring link function. This case tests the method's performance and robustness under a misspecified likelihood and link function.

Three data sets were generated corresponding to three different strategies. In the first strategy, the experiment was set such that the normal temperature level was $k_0^0 = 300$ in Kelvin, with the scale parameter $\theta_0^0 = 7000$. This strategy represents the normal use stress level. Ten observations were generated from the Weibull distribution. The Arrhenius link function was assumed to relate the scale parameters across all strategies, with the acceleration parameter set to $\gamma = 5000$.

In the second strategy, the accelerated experiment was set with increased temperature levels of $k_0^1 = 350$, $k_1^1 = 400$, and $k_2^1 = 450$ in Kelvin for s_0^1, s_1^1, and s_2^1, respectively. The stress level increased from s_0^1 to s_1^1 at $\tau_1^1 = 300$, and from s_1^1 to s_2^1 at $\tau_2^1 = 350$. The corresponding scale parameters were $\theta_0^1 = 647.23$, $\theta_1^1 = 108.52$, and $\theta_2^1 = 8.72$ for s_0^1, s_1^1, and s_2^1, respectively.

In the third strategy, the accelerated experiment was set with increased temperature levels of $k_0^2 = 380$, $k_1^2 = 420$, and $k_2^2 = 460$ in Kelvin for s_0^2, s_1^2, and s_2^2, respectively. The stress level increased from s_0^2 to s_1^2 at $\tau_1^2 = 100$, and from s_1^2 to s_2^2 at $\tau_2^2 = 130$. The corresponding scale parameters were $\theta_0^2 = 209.53$, $\theta_1^2 = 59.84$, and $\theta_2^2 = 21.25$ for s_0^2, s_1^2, and s_2^2, respectively. The generated failure times are presented in Table 5.1.

To analyze the data sets provided in Table 5.1, the Weibull cumulative exposure model is applied to the accelerated stress strategies m_1 and m_2, while the Weibull model is used for the normal stress strategy m_0. The analysis is conducted in a pairwise manner between m_i $(i = 1, 2)$ and m_0 to estimate the parameter intervals $[\underline{\gamma}_{m,0}, \overline{\gamma}_{m,0}]$ and $[\underline{\theta}_{m,0}, \overline{\theta}_{m,0}]$. The following procedure outlines the analysis specifically for m_1 and m_0.

In the first step, uniform prior distributions are assigned to the scale parameter (θ_0) and the acceleration parameter (γ) over the range $[1, 100000]$. The investigation of plausible parameter values is carried out through posterior predictive checks, which serve to assess and refine the initial parameter ranges. This iterative process ensures that the

Strategies	Stress level	Data sets	Failures times
m_0	s_0	t_0	1755.83, 4149.09, 4799.74, 5095.24, 7454.25, 7563.30, 8245.62, 10385.32, 11166.58, 12411.59
m_1	s_0^1	t_0^1	185.63, 222.74
m_1	s_1^1	t_1^1	300.66, 320.37, 325.00, 341.08, 342.94
m_1	s_2^1	t_2^1	359.43, 360.06, 371.28
m_2	s_0^3	t_0^2	14.28, 42.54
m_2	s_1^2	t_1^2	101.73, 105.56, 108.97, 120.80, 123.50
m_2	s_2^2	t_2^2	138.93, 140.52, 146.53

Table 5.1: A simulated data of Example 5.6.1

priors capture prior knowledge while adequately representing plausible parameter values derived from the data. The posterior predictive checks involve several key steps to refine the parameter bounds iteratively. First, replicated datasets are generated from the posterior predictive distribution to capture the variability and uncertainty in the predictions. These datasets are then sorted to identify the maximum and minimum datasets, representing optimistic and pessimistic failure scenarios, respectively. Next, the extreme datasets are fitted to the posterior distribution to obtain refined parameter ranges.

The analysis of extreme datasets provides valuable insights into the range of uncertainty in parameter estimates. The minimum dataset represents early failure scenarios and yields pessimistic parameter estimates, while the maximum dataset reflects delayed failure scenarios, resulting in optimistic parameter estimates. For the minimum dataset, the failure times for the normal stress strategy m_0 were $t_{0,\min} = \{12.72, 246.45, 643.95, 1052.23, 1761.57, 2061.41, 3179.23, 4075.14, 4701.15, 5781.78\}$, and for the accelerated stress strategy m_1, the failure times were $t_{1,\min} = \{0.50, 19.88, 59.15, 79.89, 151.36, 208.63, 272.01, 304.45, 315.65, 334.40\}$. After fitting the minimum dataset to the posterior distribution, the lower bounds for the parameters were obtained as $a_{\theta_0} = 2800.22$ and $a_\gamma = 3753.12$.

Conversely, for the maximum dataset, the failure times for the normal stress strategy m_0 were $t_{0,\max} = \{8124.85, 9435.63, 10754.89, 10794.74, 12531.77, 14480.72, 15341.75, 16211.28, 20707.48, 26759.62\}$, and for the accelerated stress strategy m_1, the failure times were $t_{1,\max} = \{353.97, 357.18, 357.88, 361.99, 369.39, 373.78, 383.47, 388.92, 397.74, 453.91\}$.

After fitting the maximum dataset to the posterior distribution, the upper bounds for the parameters were determined as $b_{\theta_0} = 15979.51$ and $b_\gamma = 4965.96$.

This analysis highlights the range of uncertainty in the scale parameter θ_0 and the acceleration parameter γ. The lower bounds a_{θ_0} and a_γ correspond to pessimistic scenarios associated with early failures, while the upper bounds b_{θ_0} and b_γ capture optimistic scenarios reflecting delayed failures.

These results are integral to the robust Bayesian framework developed in this chapter, as they quantify the imprecision inherent in the failure mechanisms under normal and accelerated stress conditions. After determining the plausible ranges of the parameters, imprecision is incorporated by introducing a class of priors for the parameters. This approach enables the derivation of separate lower and upper posterior distributions, which include the observed data and consider extreme failure scenarios for robust analysis.

The choice of the Beta distribution with shape parameters $p = 1$ and $q = 10$ is intended to introduce heavy skewness in the prior, which reflects a form of strong prior belief concentrated near one end of the parameter range. This skewed distribution is used to determine the bounds of the class of priors, capturing a wide range of potential values to obtain extreme prior scenarios and quantify the impact of prior uncertainty.

The estimation process begins with the identification of the lower and upper bounds of the scale and acceleration parameters, which are based on these extreme classes of priors. The lower posterior distribution is derived by setting the scale parameter such that its transformation, $(\theta_0 - a_{\theta_0})/(b_{\theta_0} - a_{\theta_0})$, follows a Beta distribution with shape parameters $p = 1$ and $q = 10$. The choice of the Beta distribution with shape parameters $p = 1$ and $q = 10$ is intended to introduce heavy skewness in the prior, which reflects a form of strong prior belief concentrated near one end of the parameter range. This skewed distribution is used to determine the bounds of the class of priors, capturing a wide range of potential values to obtain extreme prior scenarios and quantify the impact of prior uncertainty. Similarly, the acceleration parameter is modeled using the transformation $(\gamma - a_\gamma)/(b_\gamma - a_\gamma)$, with the same Beta distribution parameters. This configuration captures early failure scenarios, emphasizing smaller values of the scale and acceleration parameters.

After deriving the lower posterior distribution, parameter values for the lower scale and acceleration parameters are estimated. These estimates are then used to construct

the posterior predictive distribution, which informs the construction of the lower empirical predictive survival function. This survival function reflects the pessimistic scenario associated with early failures, incorporating imprecision arising from variability in the likelihood and link functions.

For the upper posterior distribution, the scale parameter is modeled such that its transformation, $(\theta_0 - a_{\theta_0})/(b_{\theta_0} - a_{\theta_0})$, follows a Beta distribution with shape parameters $p = 10$ and $q = 1$. The acceleration parameter is similarly modeled with the transformation $(\gamma - a_\gamma)/(b_\gamma - a_\gamma)$, using the same Beta distribution parameters. This configuration highlights delayed failure scenarios, characterized by larger values of the scale and acceleration parameters

Following the estimation of the upper posterior distribution, parameter values for the upper scale and acceleration parameters are obtained. These values are then used to construct the upper posterior predictive distribution, leading to the construction of the upper empirical predictive survival function. This survival function reflects the optimistic scenario of delayed failures, accounting for imprecision in the failure mechanisms.

Additionally, a precise predictive survival function is constructed to represent a uniform belief across the parameter space. For this purpose, the scale parameter is modeled such that its transformation, $(\theta_0 - a_{\theta_0})/(b_{\theta_0} - a_{\theta_0})$, follows a Beta distribution with shape parameters $p = 1$ and $q = 1$. The acceleration parameter is similarly modeled using the transformation $(\gamma - a_\gamma)/(b_\gamma - a_\gamma)$, with identical Beta distribution parameters.

The iterative process ensures the derivation of robust estimates for the lower, upper, and precise survival functions, capturing the full range of imprecision introduced by variability in the likelihood, link functions, and parameter uncertainties. This comprehensive framework provides a nuanced evaluation of survival probabilities under extreme scenarios.

It should be noted that, in the previous chapters, the transformed data are based on the overall values $[\underline{\gamma}, \overline{\gamma}]$, obtained as the minimum and maximum matching values for the pairwise tests, correspondingly. All failure times at the higher strategies are transformed to the normal stress level based on the overall values $[\underline{\gamma}, \overline{\gamma}]$. Consequently, the failure times at the higher strategies m_1 and m_2 are transformed into interval-valued data at the normal stress level s_0 in strategy m_0. In this chapter, we developed a similar technique to examine whether there is a difference in the analysis obtained from data sets coming from

Cases	Strategies	a_γ	$\underline{\gamma}$	γ	$\overline{\gamma}$	b_γ	a_{θ_0}	$\underline{\theta_0}$	θ_0	$\overline{\theta_0}$	b_{θ_0}	a_β	β	b_β
Case 1	m_1m_0	3753.12	3985.73	4772.22	4905.07	4965.95	2800.22	4619.18	7449.79	10400.70	15979.51	2	2	2
	m_2m_0	3861.584	4085.23	4934.18	4940.36	4947.77	2550.402	4451.85	6809.65	10397.66	16103.02	2	2	2
Case 2	m_1m_0	3981.75	4498.09	5217.13	6128.03	6694.78	4733.33	6179.69	8455.22	13174.73	14272.1	1.10	2.14	3.24
	m_2m_0	3867.85	3938.93	5200.14	6368.05	6401.81	4239.51	4351.42	8380.49	18308.52	18421.65	0.74	1.77	3.01
Case 3	m_1m_0	3547.35	3890.77	4850.17	5625.487	6014.52	4480.60	5634.27	8331.25	12532.23	13856.21	1.014	2.16	3.78
	m_2m_0	3708.34	4235.13	4813.74	5605.31	6141.73	4668.55	5819.96	8092.90	12090.35	12888.59	0.92	1.75	3.34
Case 4	m_1m_0	3178.74	4249.40	5506.00	6762.83	7284.22	3485.91	5960.11	11179.69	24782.76	27705.37	1	1	1
	m_2m_0	3160.57	4093.27	5103.00	6332.44	6776.257	3276.34	5282.97	8882.02	21691.27	24096.78	1	1	1

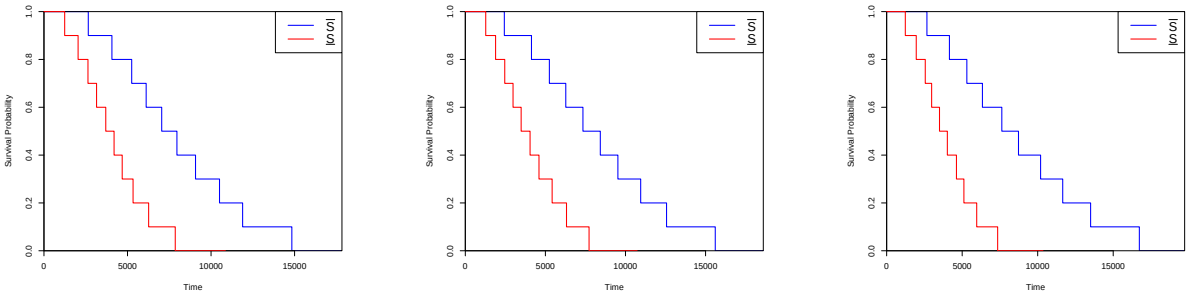
Table 5.2: $[\underline{\theta}_{m,0}, \overline{\theta}_{m,0}]$ and $[\underline{\gamma}_{m,0}, \overline{\gamma}_{m,0}]$ for Example 5.6.1.

different experimental settings or not. This technique aims to obtain the lower posterior predictive in a manner that we obtain the minimum values of the parameters of the two lower posteriors distributions in the posterior predictive distribution. Similarly, the upper posterior predictive is obtained in a manner that we obtain the maximum values of the parameters of the two upper posterior distributions in the posterior predictive distribution. We define these posterior predictive distributions as the minimum and maximum posterior predictive distributions.

In Case 1, where the shape parameter is assumed to be constant and known, the results demonstrate relatively narrow intervals for both the scale parameter (θ_0) and the acceleration parameter (γ) compared to the other cases, as presented in Table 5.2. This indicates low imprecision in the parameter estimates across all cases.

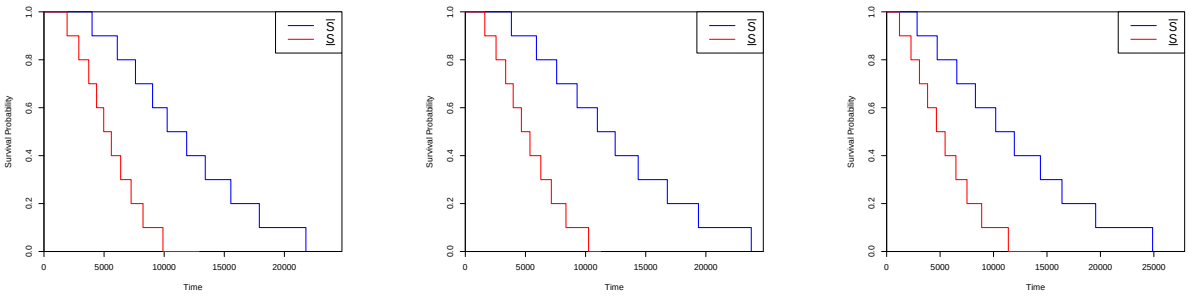
Figure 5.1 illustrates the lower and upper empirical predictive survival functions derived for the data coming from m_0 and m_1, as well as m_0 and m_2. The survival functions exhibit similar patterns, reflecting consistency in the estimated failure times across different stress strategies. Additionally, the survival functions constructed using the minimum and maximum posterior predictive distributions align closely with those obtained for m_1 and m_2, further supporting the observation of minimal imprecision.

When comparing these results with those obtained in Chapter 3 (Example 3.4.1), it is observed that the imprecision in the lower and upper empirical predictive survival functions in this chapter is relatively large. A similar observation can be made by comparing the results with Chapter 4 (Example 4.3.1). The imprecision illustrated in this case is relatively greater than that observed in Chapter 4. The comparisons highlight that the method developed here captures a wider range of imprecision and uncertainty in survival probabilities.



(a) m_0 and m_1. (b) m_0 and m_2 (c) Using min and max

Figure 5.1: Lower and upper empirical predictive survival functions for Case 1, Example 5.6.1.



(a) m_0 and m_1 (b) m_0 and m_2 (c) Using min and max

Figure 5.2: Lower and upper empirical predictive survival functions for Case 2, Example 5.6.1.

In Case 2, where the shape parameter is treated as unknown and estimated, slightly wider intervals for the scale and acceleration parameters are observed compared to Case 1. Despite this increase in imprecision, the results remain consistent across stress strategies m_1 and m_2, as shown in Table 5.2. This indicates that the inclusion of additional variability in the parameter estimates does not significantly alter the predictive performance of the survival functions.

The survival functions in Figure 5.2 for m_0 and m_1, as well as m_0 and m_2, reveal minor differences compared to Case 1, reflecting the slight increase in imprecision. The minimum and maximum posterior predictive distributions exhibit a marginally less imprecision, but the overall behavior of the survival functions remains consistent with those observed in Case 1, suggesting robustness of the method under this scenario.

In Case 3, the Arrhenius link function is replaced with the Eyring link function to as-

sess the method's performance under a misspecified link function. The results, presented in Table 5.2, reveal moderate intervals for the scale and acceleration parameters, with a noticeable increase in imprecision compared to Cases 1 and 2. The introduction of the Eyring link function contributes to additional imprecision, particularly in the survival functions derived from the minimum and maximum posterior predictive distributions.

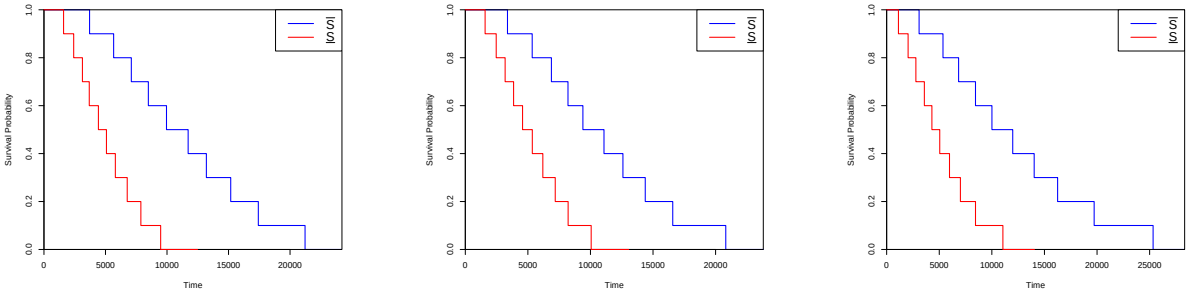
From Figure 5.3, the survival functions for the data coming from m_0 and m_1, as well as m_0 and m_2, remain largely consistent with those observed in the previous cases. However, the survival functions constructed from the minimum and maximum posterior predictive distributions exhibit minimal imprecision, suggesting that the proposed method maintains robustness despite the link function misspecification.

Notably, in contrast to the results obtained using the methods presented in Chapter 3 (Example 3.4.1) and Chapter 4 (Example 4.3.1), which revealed significant differences due to the increased imprecision resulting from link function misspecification, the NPI approach developed in these chapters effectively captures and quantifies the imprecision and uncertainties associated with such misspecifications. However, the method developed in this chapter demonstrates greater consistency in incorporating the inherent imprecision in the comparisons across different cases.

Although the cases appear similar in the current analysis, the comparisons reveal that the imprecision captured by the lower and upper NPI survival functions in Chapters 3 and 4 is more pronounced. This results in a wider gap between the lower and upper bounds in those chapters, whereas in the current approach, the imprecision is captured more comprehensively and consistently. As a consequence, the developed approach in this chapter leads to more balanced and stable estimates, ensuring that the observed differences between the lower and upper empirical predictive survival functions remain nearly consistent across various experimental settings and different cases.

In Case 4, where the shape parameter is fixed to 1 (exponential model) and the Eyring link function is applied, the results reveal considerably wider intervals for the scale and acceleration parameters, as presented in Table 5.2. The combination of the fixed shape parameter and the misspecified link function introduces substantial imprecision in the predictive survival functions across all scenarios.

Figure 5.4 shows that the survival functions corresponding to the data from m_0 and m_1, as well as m_0 and m_2, exhibit a high degree of imprecision, similar to those de-

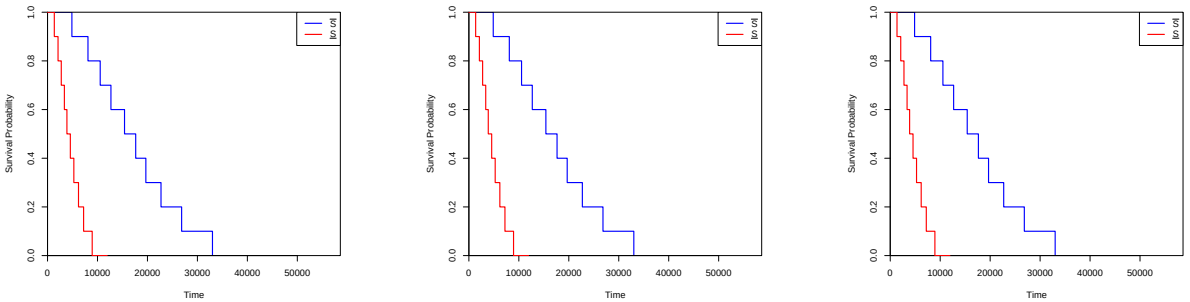


(a) m_0 and m_1 (b) m_0 and m_2 (c) Using min and max

Figure 5.3: Lower and upper empirical predictive survival functions for Case 3, Example 5.6.1.

rived from the minimum and maximum posterior predictive distributions. Despite the overall high imprecision observed across all scenarios, the survival functions remain relatively consistent without significant differences between the strategies. This suggests that the combined effect of the fixed shape parameter and the misspecified link function introduces substantial uncertainty uniformly across all settings, reinforcing the need for careful consideration of model assumptions in the analysis.

A comparison with the corresponding analysis in Chapter 3 (Example 3.4.1) suggests that the imprecision observed in the predictive survival functions across all scenarios in this chapter is greater than that observed in Chapter 3. This indicates that the combined effects of the misspecified likelihood and link function introduce a higher level of uncertainty in the current analysis compared to the previous approach in Chapter 3.



(a) m_0 and m_1 (b) m_0 and m_2 (c) Using min and max

Figure 5.4: Lower and upper empirical predictive survival functions for Case 4, Example 5.6.1.

Cases	Strategies	a_γ	$\underline{\gamma}$	γ	$\overline{\gamma}$	b_γ	a_{θ_0}	$\underline{\theta_0}$	θ_0	$\overline{\theta_0}$	b_{θ_0}	a_β	β	b_β
Case 1	m_1m_0	4764.51	4809.84	4951.63	5083.55	5160.00	6453.74	6533.5	7014.83	7260.73	7319.21	2	2	2
	m_2m_0	4167.64	4655.66	4862.01	5044.09	5305.97	5203.43	5343.02	6171.00	7113.75	7173.04	2	2	2
Case 2	m_1m_0	3038.84	3028.65	4841.44	5092.85	5166.72	6021.37	6029.34	6744.45	7561.74	7605.30	2.00	2.14	2.66
	m_2m_0	4827.37	4862.64	4966.77	5062.13	5085.58	6371.30	6596.19	6817.75	7060.09	7467.20	2.00	2.16	2.56
Case 3	m_1m_0	4235.65	4249.52	4551.45	4917.91	4956.58	6141.92	6181.21	6940.48	7732.75	7763.74	2	2.14	2.72
	m_2m_0	4339.11	4422.30	4603.60	4756.77	4860.89	6122.61	6324.10	6903.83	7740.49	7871.46	2.00	2.16	2.78
Case 4	m_1m_0	4511.74	4766.26	5129.55	5541.21	5774.66	5157.36	6008.17	7412.14	9495.93	10182.38	1	1	1
	m_2m_0	4230.48	4456.34	4820.57	5152.42	5280.67	5119.85	5768.00	7284.37	9355.90	10127.56	1	1	1

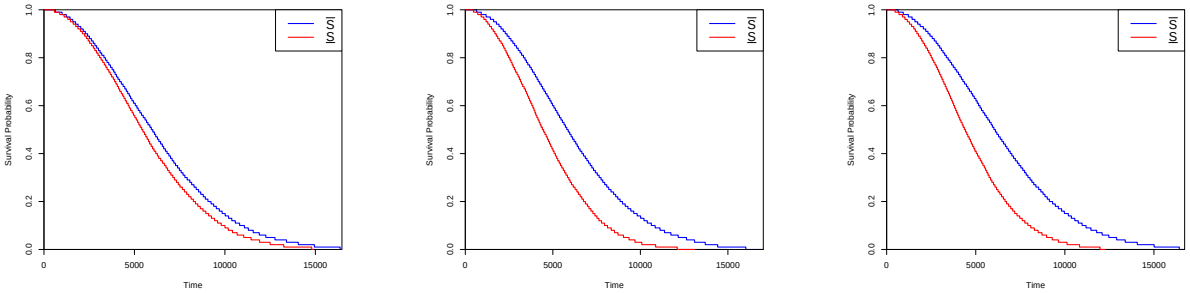
Table 5.3: $[\underline{\theta}_{m,0}, \overline{\theta}_{m,0}]$ and $[\underline{\gamma}_{m,0}, \overline{\gamma}_{m,0}]$ for Example 5.6.2.

Across the four cases, the comparison of survival functions demonstrates the increasing impact of parameter imprecision and model misspecification on the predictive performance. Cases 1 through 3 show relatively consistent results, with no significant differences in the survival functions across strategies, reflecting the robustness of the method under these scenarios. However, in Case 4, the imprecision is substantially higher, leading to greater divergence in the survival functions across all scenarios.

Example 5.6.2 This example extends the previous analysis by generating a larger data set with $n = 100$ observations per strategy for all cases. The objective is to evaluate the impact of increased data quantity on the resulting inference. The experimental conditions remain the same as described in Example 5.6.1, and the resulting parameter intervals for all cases are presented in Table 5.3.

The analysis of the larger data set demonstrates several key differences compared to the smaller data set ($n = 10$). Notably, the increased sample size results in narrower intervals for the acceleration parameter γ, leading to improved precision in parameter estimation. This effect is evident across all cases and suggests that the increased data quantity enhances the precision of the results.

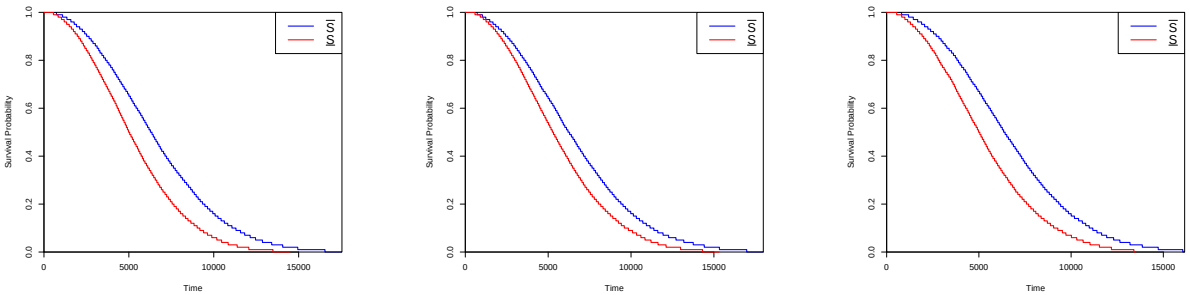
In Case 1, where the shape parameter is assumed constant and known, the larger data set results in narrower parameter intervals, reducing imprecision in the empirical predictive survival functions. Figure 5.5 shows that the lower and upper empirical predictive survival functions exhibit less divergence, indicating more precise failure time estimates.



(a) m_0 and m_1 (b) m_0 and m_2 (c) Using min and max

Figure 5.5: Lower and upper empirical predictive survival functions for Case 1 with $n = 100$, Example 5.6.2.

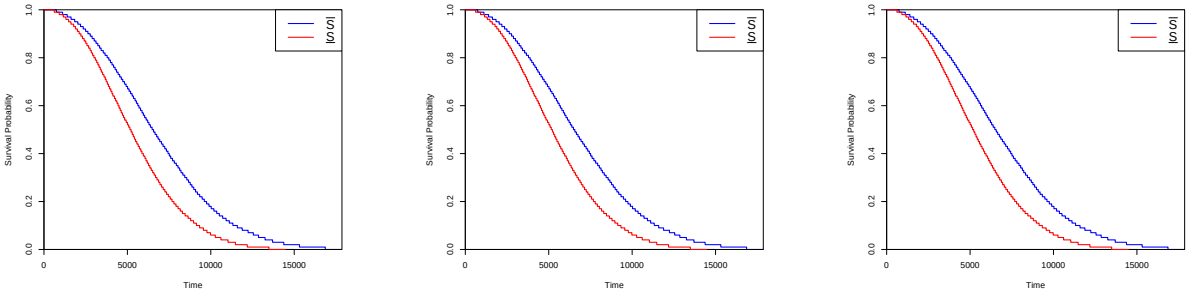
In Case 2, where the shape parameter is estimated, the increased sample size similarly results in narrower intervals, as shown in Figure 5.6, demonstrating improved precision compared to the smaller sample.



(a) m_0 and m_1 (b) m_0 and m_2 (c) Using min and max

Figure 5.6: Lower and upper empirical predictive survival functions for Case 2 with $n = 100$, Example 5.6.2.

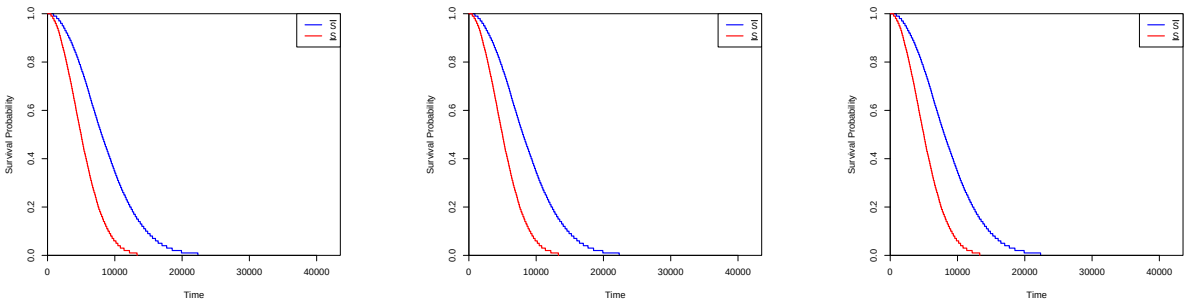
In Case 3, where the Eyring link function introduces model misspecification, the increased data set mitigates the impact of misspecification, resulting in reduced imprecision in the survival function estimates. Figure 5.7 confirms that the predictive survival functions align more closely compared to the smaller sample scenario.



(a) m_0 and m_1 (b) m_0 and m_2 (c) Using min and max

Figure 5.7: Lower and upper empirical predictive survival functions for Case 3 with $n = 100$, Example 5.6.2.

In Case 4, where the shape parameter is fixed to 1, assuming an exponential model with the Eyring link function, the larger data set reduces but does not eliminate the substantial imprecision introduced by the misspecified likelihood and link function. Figure 5.8 illustrates a noticeable improvement in the precision of the survival functions.



(a) m_0 and m_1 (b) m_0 and m_2 (c) Using min and max

Figure 5.8: Lower and upper empirical predictive survival functions for Case 4 with $n = 100$, Example 5.6.2.

This example demonstrates that increasing the number of observations reduces the imprecision in the lower and upper empirical predictive survival functions, leading to narrower parameter intervals and more precise estimates.

5.7 Simulation studies

In this section, simulation studies are performed to assess the performance of the proposed method outlined in this chapter. The simulation follows the same procedure as in the previous chapters, utilizing three SSALT data strategies under the experimental settings described in Example 5.6.1.

In the first strategy, the normal stress level is defined by assuming $\theta_0^0 = 7000$ with the normal temperature level set at $k_0^0 = 300$. The second strategy incorporates three stress levels: $k_0^1 = 350$, $k_1^1 = 400$, and $k_2^1 = 450$, with stress transitions occurring at $\tau_1^1 = 300$ and $\tau_2^1 = 350$. The third strategy also includes three stress levels: $k_0^2 = 380$, $k_1^2 = 420$, and $k_2^2 = 460$, with stress level changes at $\tau_1^2 = 100$ and $\tau_2^2 = 130$. Additionally, the shape parameter is set to $\beta = 2$, and the Arrhenius link function parameter γ between all stress levels in the accelerated strategies is assumed to be $\gamma = 5000$.

This simulation was run 10,000 times with the data simulated from the proposed model with different numbers of observations at each strategy, which are $n = 10$, 50, and 100. The performance of this simulation is investigated by the result of simulating a future observation at the normal strategy m_0. This is to examine whether the future observation of the normal stress level mixes without distinction with the predicted data at the normal strategy. The performance is inspected by considering if the future observation has surpassed the quartiles of the lower and upper survival functions for $1 - q$= 0.75, 0.50, and 0.25. In this method, the quartiles have been used for the overall performance as providing a suitable indicator as well as other different quantiles similarly can be used.

First, the proportions are computed to determine whether the future observation exceeds the quartiles of the lower and upper survival functions in the expected proportions. For a well-performing method, the first, second, and third quartiles of the lower survival functions should be greater than 0.75, 0.50, and 0.25, respectively. Similarly, the first, second, and third quartiles of the upper survival functions should be less than 0.75, 0.50, and 0.25, respectively.

The simulation studies are conducted under three different cases. In Case 1, the scenario described in Example 5.6.1 is considered, where the shape parameter is assumed to be constant and known. Table 5.4 and Figures 5.9–5.11 present the performance of the proposed method for this setting across sample sizes of $n = 10, 50, 100$. These figures

m_1m_0	$n = 10$		$n = 50$		$n = 100$	
$1-q$	qU	qL	qU	qL	qU	qL
0.75	0.4387	0.9323	0.5948	0.8665	0.6483	0.8425
0.50	0.1898	0.7605	0.3265	0.6245	0.4058	0.5991
0.25	0.0671	0.5803	0.1374	0.3732	0.1788	0.3636
m_2m_0	$n = 10$		$n = 50$		$n = 100$	
$1-q$	qU	qL	qU	qL	qU	qL
0.75	0.4345	0.9355	0.5942	0.8615	0.6391	0.8367
0.50	0.1963	0.7539	0.3280	0.6220	0.3582	0.5966
0.25	0.0652	0.5785	0.1345	0.3703	0.1725	0.3582
$\underline{\gamma}$ and $\overline{\gamma}$	$n = 10$		$n = 50$		$n = 100$	
$1-q$	qU	qL	qU	qL	qU	qL
0.75	0.4445	0.9394	0.5951	0.8627	0.6472	0.8408
0.50	0.1916	0.7572	0.3303	0.6207	0.4076	0.5941
0.25	0.0687	0.5815	0.1386	0.3665	0.1725	0.3641

Table 5.4: Proportion of runs with future observation greater than the quartiles, Case 1

denote the first, second, and third quartiles as $qL0.25$, $qU0.25$, $qL0.50$, $qU0.50$, $qL0.75$, and $qU0.75$, corresponding to the lower and upper survival functions, respectively.

The results indicate that the imprecision between the lower and upper survival functions tends to decrease, as shown in Figure 5.11. This demonstrates that the proposed method achieves robust predictive inference when the model assumptions are fully satisfied.

The results indicate that the level of imprecision remains consistent across different scenarios, whether the data are obtained from m_0m_1, m_0m_2, or through the minimum lower and maximum upper predictive survival functions. Additionally, the results demonstrate that as the number of observations increases, the imprecision decreases.

A comparison with the corresponding case in Chapter 3 reveals that the imprecision observed in Chapter 4 and 5 is slightly larger when the sample size is small ($n = 10$). However, when the number of observations increases to $n = 100$, the results in Chapter 5 exhibit reduced imprecision, which is relatively similar to the findings in Chapter 3 and 4.

The objective of this chapter is to develop a straightforward predictive inference method based on minimal assumptions. The imprecision in the accelerating and link

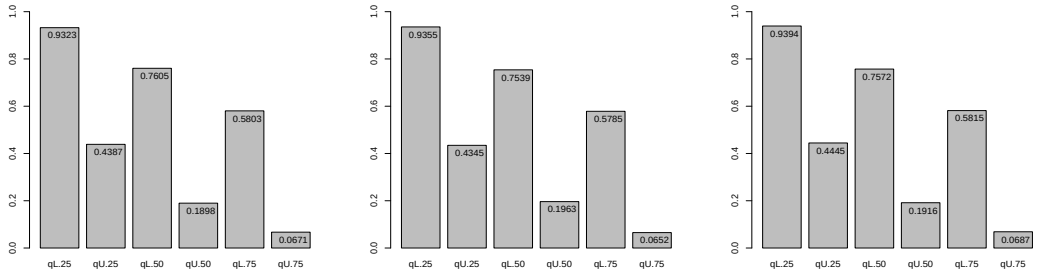
parameters across different strategies serves as an indicator of robustness against deviations from the assumed model assumptions.

In Case 2, robustness is evaluated under model misspecification to assess whether a simplified model can retain robustness despite incorrect parameter assumptions. In this scenario, the shape parameter in the data-generating process is set to $\beta = 3$, whereas the analysis mistakenly assumes $\beta = 2$. Table 5.5 and Figures A.22–A.24 present the predictive performance results of the proposed method under this misspecification. The experimental setup remains identical to that of Case 1, with data generated from the model described in Section 3.2 using $\beta = 3$, while the analysis incorrectly assumes $\beta = 2$ for each strategy.

Both Case 1 and Case 2 consider a known and constant shape parameter. However, the results in Figures A.22–A.24 indicate that Case 2 exhibits slightly greater imprecision in the predictive survival functions due to the misspecified shape parameter. Additionally, Table 5.5 shows that imprecision in the lower and upper survival functions remains more pronounced in Case 2 across larger sample sizes (n) compared to Case 1. Despite this increase in imprecision, the proposed method demonstrates reasonable robustness against model misspecification.

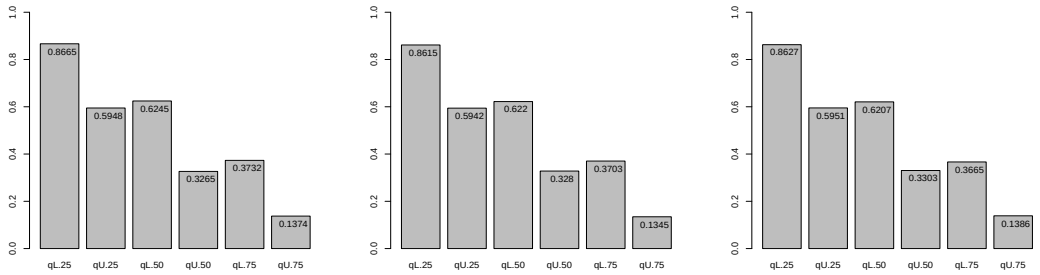
A comparison with the corresponding results in Chapter 3 reveals that the imprecision observed in this case is larger than in Chapter 3, particularly when the sample size is small ($n = 10$). However, as the sample size increases to $n = 100$, the results in this chapter show a reduction in imprecision, aligning more closely with the findings in Chapter 5. These findings indicate that while misspecification introduces additional uncertainty, the proposed method retains reasonable robustness, particularly when sufficient data are available.

In Case 3, robustness is again examined under model misspecification, specifically by assessing the effect of using an incorrect link function. This case investigates whether a simple model can still maintain robustness when the assumed link function does not align with the true underlying failure mechanism. In this scenario, the data are generated using the Eyring link function, while the analysis assumes the Arrhenius link function. This setting is particularly relevant in practice, where the correct link function is often uncertain.



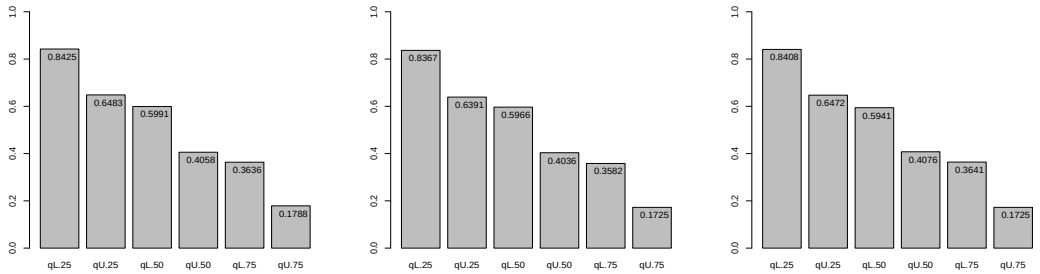
(a) m_1 to m_0 (b) m_2 to m_0 (c) Using min and max

Figure 5.9: Proportion of runs with future observation greater than the quartiles, Case 1, $n = 10$.



(a) m_1 to m_0 (b) m_2 to m_0 (c) Using min and max

Figure 5.10: Proportion of runs with future observation greater than the quartiles, Case 1, $n = 50$.



(a) m_1 to m_0 (b) m_2 to m_0 (c) Using min and max

Figure 5.11: Proportion of runs with future observation greater than the quartiles, Case 1, $n = 100$.

m_1m_0	$n = 10$		$n = 50$		$n = 100$	
$1-q$	qU	qL	qU	qL	qU	qL
0.75	0.4506	0.9829	0.7155	0.9441	0.7641	0.9237
0.50	0.1063	0.8518	0.3343	0.7426	0.4057	0.7183
0.25	0.0121	0.6476	0.0671	0.4155	0.0987	0.3391
m_2m_0	$n = 10$		$n = 50$		$n = 100$	
$1-q$	qU	qL	qU	qL	qU	qL
0.75	0.4518	0.9815	0.7175	0.9435	0.7658	0.9263
0.50	0.1022	0.8535	0.3304	0.7473	0.4041	0.7102
0.25	0.0129	0.6474	0.0631	0.4156	0.0958	0.3400
$\underline{\gamma}$ and $\overline{\gamma}$	$n = 10$		$n = 50$		$n = 100$	
$1-q$	qU	qL	qU	qL	qU	qL
0.75	0.4512	0.9830	0.7143	0.9435	0.7638	0.9266
0.50	0.1043	0.8511	0.3347	0.7415	0.4083	0.7116
0.25	0.0115	0.6492	0.0705	0.4108	0.0958	0.3325

Table 5.5: Proportion of runs with future observation greater than the quartiles, Case 2.

As discussed in Section 2.5, the Eyring link function serves as an alternative to the Arrhenius link function for modeling the acceleration parameter as a function of temperature. The simulation setup follows the same conditions as in Case 1, where the model assumptions were entirely correct, except that the link function in the sampling model was replaced with the Eyring link function. Table 5.6 and Figures A.25-A.27 present the results of the simulation.

A comparison with the fully specified model in Case 1, as shown in Table 5.4 and Table 5.5, indicates that the impact of link function misspecification introduces large additional imprecision in the predictive survival functions. This similarity is also evident when comparing Figures 5.9–5.11, Figures A.22–A.24, and Figures A.25–A.27 for the quartiles of the lower and upper predictive survival functions at $1 - q = 0.75, 0.50, 0.25$.

However, when comparing the results of this chapter with those in Chapters 3 and 4, it is evident that the imprecision observed here is larger than in both previous chapters. The increase in imprecision is particularly noticeable when the sample size is small ($n = 10$), though it decreases as the sample size grows. This suggests that while the proposed method retains robustness against link function misspecification, the effects of model misspecification accumulate across different assumptions, leading to greater un-

$m_1 m_0$	$n = 10$		$n = 50$		$n = 100$	
$1 - q$	qU	qL	qU	qL	qU	qL
0.75	0.2488	0.9527	0.4648	0.9034	0.5148	0.8523
0.50	0.0582	0.8049	0.1772	0.7203	0.2272	0.6692
0.25	0.0113	0.6515	0.0461	0.5135	0.0982	0.4624
$m_2 m_0$	$n = 10$		$n = 50$		$n = 100$	
$1 - q$	qU	qL	qU	qL	qU	qL
0.75	0.2459	0.9508	0.4627	0.8974	0.5127	0.8463
0.50	0.0576	0.8023	0.1778	0.7214	0.2278	0.6703
0.25	0.0118	0.6516	0.0467	0.5103	0.0989	0.4592
$\underline{\gamma}$ and $\overline{\gamma}$	$n = 10$		$n = 50$		$n = 100$	
$1 - q$	qU	qL	qU	qL	qU	qL
0.75	0.2455	0.9512	0.4629	0.8941	0.5157	0.8429
0.50	0.0559	0.8019	0.1735	0.7214	0.2257	0.6703
0.25	0.0111	0.6549	0.0455	0.5126	0.0977	0.4615

Table 5.6: Proportion of runs with future observation greater than the quartiles, Case 3.

certainty in the predictive inferences. Despite this, the overall predictive performance remains relatively stable, demonstrating the method's adaptability under various modeling conditions.

The main results of the previous simulations conclude that the generated future observation at the normal stress strategy has surpassed the quartiles which were measured in the right proportions. This approach shows an overall good performance as an appropriate predictive inference with reasonable imprecision, if the model assumptions are completely true. Conversely, it shows relatively larger imprecision in the instances of model misspecification, and obviously, in cases of massive misspecification, meaningful inferences are hard to be driven by any method. In this case, our method will show larger imprecision which will indicate that there is a problem with the model fit or it is required to collect more data. Moreover, it shows, when the number of observations tends to be larger in the generated data at different strategies, the imprecision decreases. To compare the results of this chapter with the results obtained in Chapter 3, we conclude that the results are equivalent to the results based on the likelihood ratio test with significant level equals 0.10 where the minimum and maximum values of γ are used to transform the data.

5.8 Concluding remarks

This chapter introduced a robust Bayesian approach for analyzing SSALT data within the framework of imprecise probabilities. The proposed method incorporates imprecision to account for uncertainty in prior knowledge about the model parameters, allowing for the derivation of lower and upper bounds for the posterior distribution. Imprecise prior knowledge is modeled using a class of prior distributions, where this class is defined by its extreme cases in the absence of complete knowledge about the parameters. Based on these extreme bounds, the lower and upper posterior predictive distributions are obtained separately, enabling the prediction of future failure times at the normal stress level. This approach further facilitates the construction of lower and upper predictive survival functions at the normal stress level. The effectiveness of the method was assessed through comprehensive simulation studies.

The results demonstrate that the proposed approach achieves considerable robustness, particularly when data from higher stress levels are incorporated. Furthermore, the method remains effective under instances of model misspecification, even when the assumed model does not perfectly fit the observed data.

A key distinction of this approach from the methods employed in previous chapters is that it does not rely on pairwise comparisons of datasets from different strategies and experimental settings. In contrast, the earlier chapters applied a pairwise analysis framework for statistical testing. When comparing the results obtained in this chapter with those in Chapter 3, we find that the outcomes are consistent with those derived from the likelihood ratio test at a significance level of $\alpha = 0.10$, where the minimum and maximum values of γ were used for data transformation.

Moreover, the findings indicate that the imprecision observed in Chapter 5 is relatively larger when the link function is misspecified compared to the results in Chapters 3 and 4. This suggests that while the Bayesian approach remains robust under model misspecification, incorrect assumptions about the link function contribute to increased imprecision in predictive survival functions. The comparison across chapters further reinforces the importance of selecting an appropriate link function to mitigate the impact of imprecision in accelerated life testing analysis.

Chapter 6

Conclusions

In this chapter, we summarise and compare the primary findings of the thesis in Section 6.1, highlight practical implementation and offer recommendations in Section 6.2, and outline potential areas for future research in Section 6.3, including suggestions for extending our robust statistical methods for step-stress accelerated life testing data.

6.1 Summary and comparisons

The main novelties of this thesis are divided in two parts. Frist, we have incorporated imprecision in the assumed models based on likelihood ratio test, log-rank test and robust Bayesian analysis in three different approaches. Second, we have developed new transformation and extrapolation methods to predict future failures times at the normal stress level. In the first two approaches, we have introduced imprecision based on classical tests in a manner that the transformed data should not be distinguishable from failure times at the normal stress level. In the third approach, we have introduced imprecision based on a class of priors identified by its extreme prior distribution in a robust Bayesian analysis, where the data are extrapolated at the normal stress level via the posterior predictive distribution.

First, we have adopted the Arrhenius Weibull-cumulative exposure model for step-stress accelerated life testing data, where the imprecision is obtained based on a likelihood ratio test. Second, we have adopted the Arrhenius link function only for step-stress accelerated life testing data, where the imprecision is obtained based on the log-rank test. This method provides an imprecise non-parametric inference for SSALT data. Both ap-

proaches employ classical tests to examine whether the survival distributions of two data sets coming from different experimental settings are from the same underlying distribution. Based on this comparison, we obtained an interval of the acceleration parameter where the null hypothesis is not rejected. This interval for the parameters within the link function allows the transformation of data into interval-valued observations at the normal stress level. In the first approach, we transform the data by an inversion between two cumulative distribution functions and in the second approach, we transform the data by an inversion between two stress levels within the link function such as the Arrhenius link function. In terms of imprecise probability, both approaches rely on the Nonparametric Predictive Inference (NPI) for the transformed data, based on the lower and upper values of the acceleration parameter, to construct the lower and upper NPI survival functions. Third, we have adopted the Arrhenius Weibull-cumulative exposure model for step-stress accelerated life testing data, where the imprecision is obtained based on a class of priors. This class is identified by its extreme prior distribution in a robust Bayesian analysis, where the data are extrapolated at normal stress level via the posterior predictive distribution.

In Chapter 3, we have presented a new robust statistical method for step-stress accelerated life testing, where the Arrhenius-Weibull cumulative exposure model is implemented. This method was developed based on the likelihood ratio test where the test is applied for comparison of the failure times at different strategies of stress levels. This imprecision is developed for the acceleration parameter in the link function of the Arrhenius model, by the pairwise likelihood ratio test between the pairwise strategies of stress levels. This imprecision allows the transformation of failure times into interval values at the normal stress level where it is assumed that these transformed failure times are not distinguishable from failure times occurring at the normal stress level. This transformation is performed by an inversion between the cumulative distribution functions of two strategies. This achieves a high level of robustness rather than a single value of the acceleration parameter. The nonparametric predictive inference is applied to the transformed data to provide robust predictive inference. This method leads to more imprecision if data are used from higher stress levels and if the assumed model does not fit the data well for instances of model misspecification. We also investigated our method in the case of different shape parameters at each strategy and we found there is slightly more

imprecision in the acceleration parameter.

In Chapter 4, we have introduced a new approach to semi-parametric inference with imprecise probability for analysing SSALT data. Unlike traditional methods, our approach does not rely on predefined failure time distributions for each stress level. Instead, we employ the log-rank test to assess pairwise stress level survival distributions and use the Arrhenius model to determine the range of γ values. This novel technique introduces imprecision by applying nonparametric tests to the Arrhenius model's link function parameter across different stress levels. This allows us to transform data from higher stress levels into interval-valued data at the normal stress level, enhancing the method's robustness. The key insights from this chapter include establishing an interval for the link function's parameter at each stress level. This is accomplished by conducting nonparametric hypothesis tests between pairs of stress levels to quantify the imprecision. We explain our preference for imprecision derived from combined pairwise log-rank tests rather than a singular test applied across all stress levels, as the former approach increases imprecision when the model is not appropriate. To compare these two methods, we conclude that the imprecision resulting from the log-rank test is less than the imprecision resulting from the likelihood ratio test when we implement the statistical model. This was illustrated by the examples and the results of the simulations. The performance of these methods was evaluated by simulation studies. These methods show a high level of robustness in cases of model misspecification when it was investigated via simulations.

In these methods, we assumed that we have data at the normal stress level, but these data may not be available in practice. Therefore, we recommend performing the tests between two data sets coming from higher strategies of stress levels to determine the interval of the acceleration parameter. Furthermore, in this chapter, we assumed complete data sets for ease of presentation, while the implementation of these methods for right-censored data is straightforward. It is possible to perform the tests by including the right-censored data and then transform these data to the normal stress level where in the final stage the NPI with right-censored can be implemented. The procedure for the implementation is as follows. First, the pairwise log-rank test is performed, including the right-censored data, to determine the interval for the acceleration parameter. Second, both the observed and right-censored failure times are transformed to the normal stress level using the obtained parameter interval. Third, nonparametric predictive inference

(NPI) for right-censored data is applied to construct the lower and upper predictive survival functions based on the transformed data.

In Chapter 5, we have presented a new robust Bayesian approach for the analysis of SSALT data considered by imprecise probabilities. This approach incorporates imprecision to model imprecise prior knowledge on the parameters involved in the assumed model where it allows to obtain lower and upper bounds for the posterior distribution. This modelling of imprecise prior knowledge is defined based on a class of prior distributions, while this class is identified by its extremes under complete lack of knowledge about the parameters. Based on these extreme bounds, we obtain the lower and upper posterior predictive distribution in a separate manner, which permit to predict future failure times at the normal stress level. This approach enables to construct lower and upper predictive survival functions at the normal stress level. Finally, we evaluated the performance of this method by simulation studies.

This approach achieves sufficient robustness if data are used from higher stress levels and if the assumed model does not fit the data well for instances of model misspecification. We also investigated our method in the case of different shape parameters at each strategy and we found there is slightly more imprecision in the lower and upper predictive survival functions at the normal stress level. In the previous chapters, we assumed that we perform the statistical tests in a pairwise manner where we compare two data sets coming from different strategies of different experimental settings, while in this approach, the analysis is not required to be performed in a pairwise manner. To compare the results of this chapter with the results obtained in Chapter 3, we conclude that the results are equivalent to the results based on the likelihood ratio test with significant level equals 0.10 where the minimum and maximum values of the lower and upper values of γ are used to transform the data.

The findings of the simulation studies show that imprecision increases in all three methods when the assumed likelihood or link function is misspecified. The robust Bayesian method tends to produce relatively more imprecision under such model misspecification. On the other hand, the first two methods show higher imprecision mainly when the data are taken from higher stress levels. The results also indicate that imprecision becomes smaller as the number of observations increases, which highlights the positive effect of larger sample sizes on predictive performance.

The choice of method depends on the available information and the level of certainty about the statistical model. The first method, based on the likelihood ratio test, is most suitable when there is confidence in the assumed statistical model, such as the Arrhenius-Weibull cumulative exposure model. In contrast, the second method, which uses the log-rank test, is recommended when there is uncertainty about the model structure, as it offers a nonparametric approach that does not rely on specific distributional assumptions. Finally, the robust Bayesian method is particularly useful when only a single dataset is available or when there is imprecise prior knowledge. This method allows flexible modelling through a class of prior distributions and supports predictive inference even under limited or uncertain information.

6.2 Applications and recommendations

The methods proposed in this thesis contribute to the advancement of reliability assessment for engineering systems by explicitly accounting for uncertainty and imprecision. Their ability to generate robust predictions without depending on exact model specifications or precise prior knowledge makes them suitable for applications where experimental data are limited or partially observed. In practice, this allows reliability engineers to better assess system behavior under stress, improve planning for product lifetimes, and reduce risk in decision-making. Overall, these developments represent a step toward more flexible and applicable reliability methodologies in complex real-world environments.

This section expands the discussion to emphasize the applicability of the proposed methods within the context of engineering system reliability assessment. Although the primary focus of Chapters 3, 4, and 5 was on methodological development and simulation-based evaluation, the practical relevance of the methods is considered here. The methods introduced in this thesis are particularly suited for systems where time-to-failure data under normal conditions are limited or costly to obtain. Following the principles of accelerated life testing, such as those described in Nelson [52], the use of robust and predictive methods that incorporate imprecision is especially advantageous in systems exposed to varying levels of physical stress, such as temperature, voltage, or load. These include electronic devices, mechanical components, and structural materials. In such cases, accelerated testing combined with cautious, interval-based predictions can offer

timely reliability insights while explicitly accounting for model and data uncertainty, thus supporting more informed decision-making in engineering practice.

Despite the strengths and practical advantages of the proposed methods, there are several limitations that should be acknowledged. The methods proposed in this thesis currently focus on binary failure outcomes, where systems are considered either operational or failed. While this assumption simplifies modelling and inference, it may not fully reflect the range of degradation states observed in practical engineering systems, where partial or progressive failures are common. Additionally, the first two approaches developed in this thesis, based on the likelihood ratio and log-rank tests, require pairwise comparisons between two data sets from different stress levels. This requirement can be limiting in situations where only one experimental group is available or when stress levels are not clearly separated. Another limitation is that these methods assume the presence of only one unknown parameter within the link function, such as the acceleration parameter γ. If the link function involves multiple unknown parameters, the implementation of these classical test-based methods becomes less straightforward. In contrast, the robust Bayesian method introduced in Chapter 5 overcomes these limitations. It allows for analysis using a single dataset and can handle more complex models involving multiple parameters through its flexible prior structure and posterior predictive inference. This highlights the broader applicability and greater modelling power of the Bayesian approach in complex or data-limited reliability settings.

For the practical application of our robust statistical approaches, warranty analysis is an important example. Ahmadini [2] presented a method to estimate the lower and upper costs of warranty policies using accelerated life testing (ALT) data. By applying nonparametric predictive inference (NPI), lower and upper survival functions are calculated, which then help estimate the time a product is likely to survive under normal conditions. These survival bounds are used to estimate how many items may fail within the warranty period. From this, lower and upper estimates of total warranty expenses can be calculated, giving companies a more cautious and informed way to plan for future costs when product lifetimes are uncertain. This approach can be directly applied using the robust predictive methods proposed in this thesis.

6.3 Future research work

There are several future research tasks based on the developments presented in this thesis. First, further simulation studies could be conducted to explore the performance of the proposed robust approaches using different lifetime models beyond the Weibull distribution. Additionally, the cumulative exposure model adopted here can be replaced by alternative step-stress models, such as the tampered failure rate model, to examine the robustness of the proposed methods under different stress-acceleration assumptions. Second, it would be interesting to extend the proposed methodologies to more complex accelerated life testing (ALT) scenarios, including progressive, cycling, and random stress designs. These experimental setups are more reflective of real-world testing environments and would allow for a broader assessment of the methods' applicability.

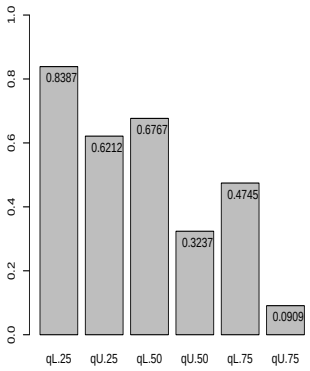
Moreover, it would be valuable to consider degradation models and to examine the methods in contexts where no failures have yet occurred. Constructing models for degradation processes requires detailed information about the engineering system and material properties, often obtained from physical measurements or expert judgment. Such models could provide early predictions of failure and further enhance the practicality of the proposed approaches. In systems that experience partial or progressive failures, the proposed framework could be extended by categorizing failure states into multiple levels and treating them as ordinal outcomes. Predictive and robust methods based on imprecise probability could be adapted to account for varying degrees of failure severity. In such cases, the transformation and inference procedures would need to be generalized to accommodate multilevel or continuous failure indicators. This represents a promising direction for future work.

As outlined in Chapters 1 and 2, the objective of this research is to employ straightforward models that incorporate imprecision. If the derived inferences adequately address and solve the real-world issues, there is no necessity for additional modelling investigations or data collection. Otherwise, a more modelling investigations or data collection, or perhaps a combination of both, would be essentially required. It should be noted that data collection often poses challenges in practical scenarios. For instance, when conducting tests on prototypes, the quantity of samples available for analysis is typically constrained [2].

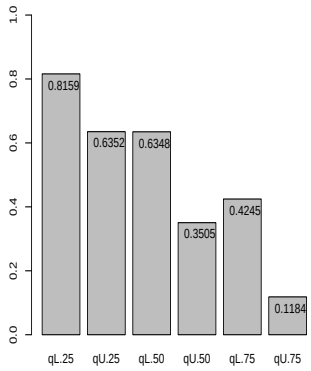
Appendix A

Simulation Figures

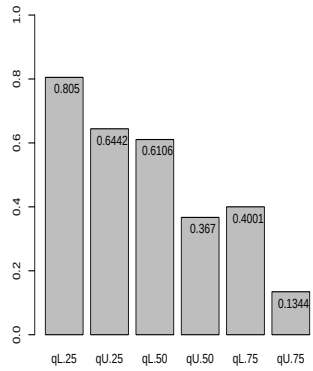
This section provides the plots of the simulation studies discussed in Chapters 3, 4, and 5, corresponding to Sections 3.5, 3.8, 4.4, and 5.7.



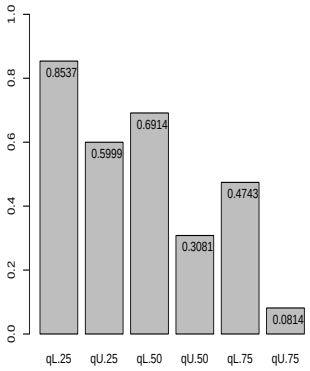
(a) m_1 to m_0 (0.01)



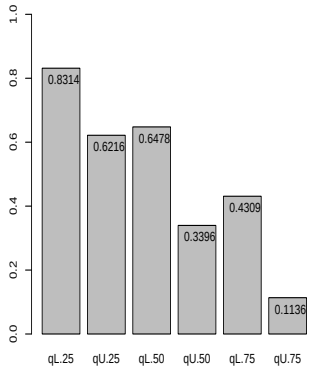
(b) m_1 to m_0 (0.05)



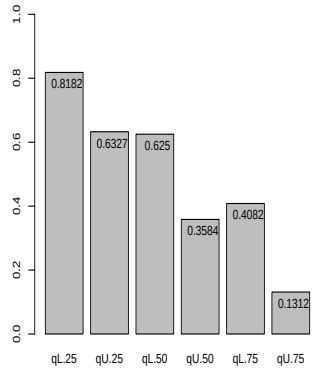
(c) m_1 to m_0 (0.10)



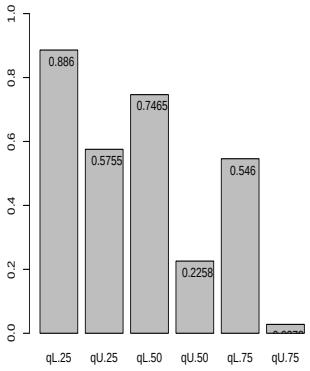
(d) m_2 to m_0 (0.01)



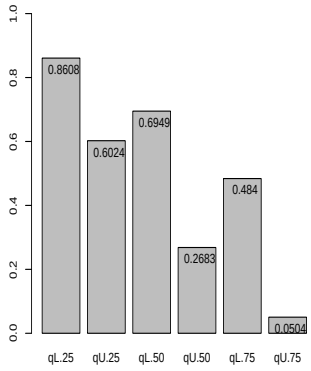
(e) m_2 to m_0 (0.05)



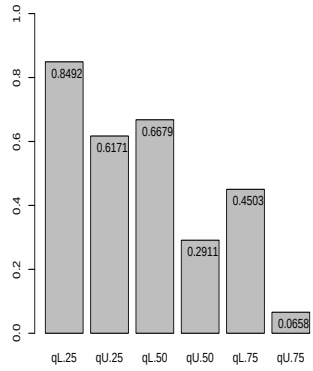
(f) m_2 to m_0 (0.10)



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

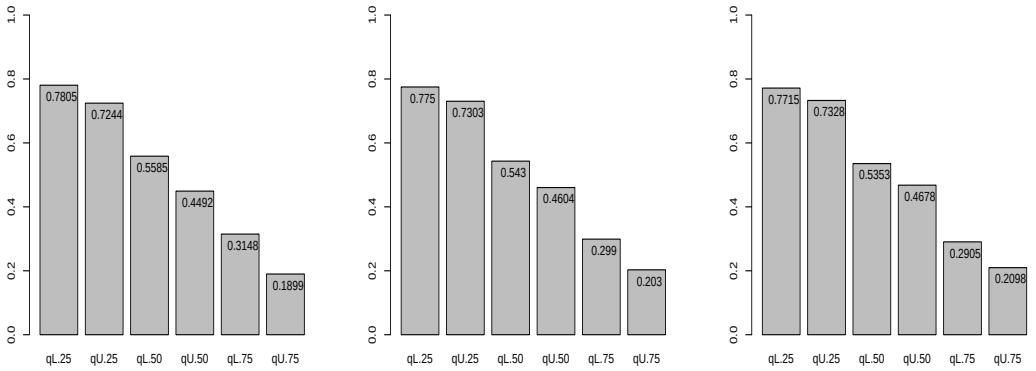


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

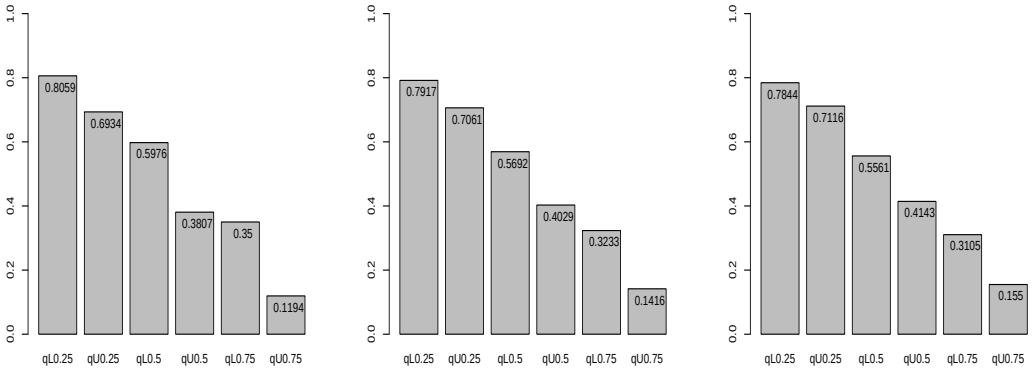


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

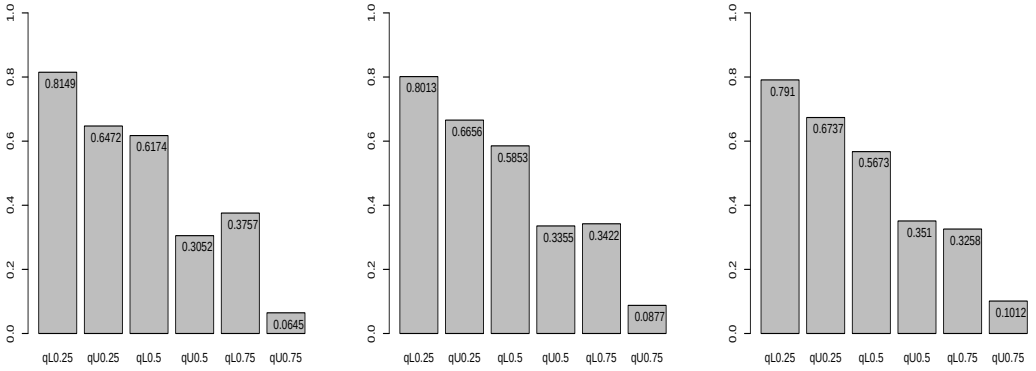
Figure A.1: Proportion of runs with future observation greater than the quartiles, Case 2, $\beta = 3$, $n = 20$. Section 3.5.



(a) m_1 to m_0 (0.01) (b) m_1 to m_0 (0.05) (c) m_1 to m_0 (0.10)

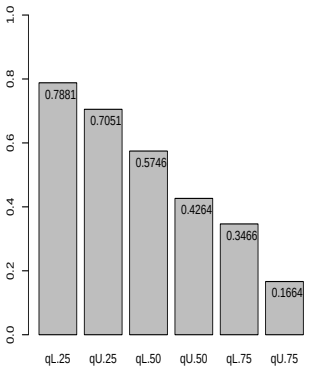


(d) m_2 to m_0 (0.01) (e) m_2 to m_0 (0.05) (f) m_2 to m_0 (0.10)

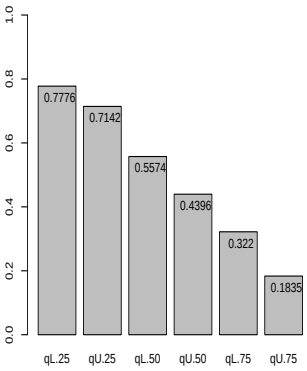


(g) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.01) (h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05) (i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

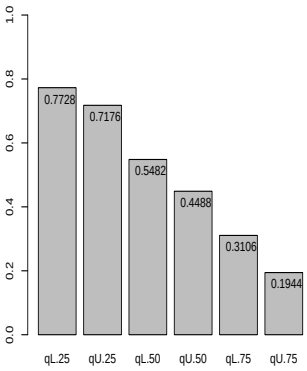
Figure A.2: Proportion of runs with future observation greater than the quartiles, Case 2, $\beta = 3$, $n = 50$. Section 3.5.



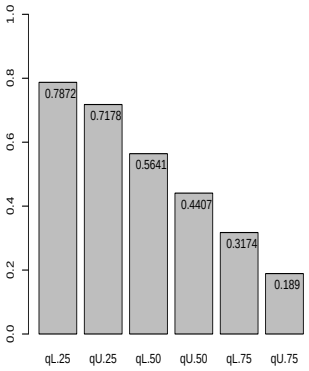
(a) m_1 to m_0 (0.01)



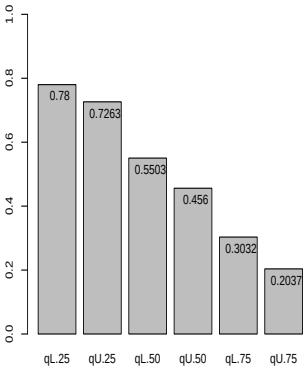
(b) m_1 to m_0 (0.05)



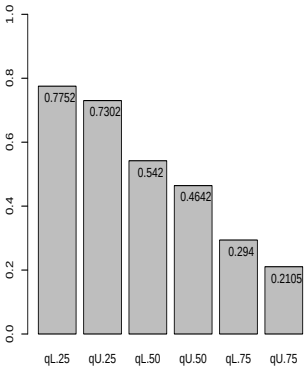
(c) m_1 to m_0 (0.10)



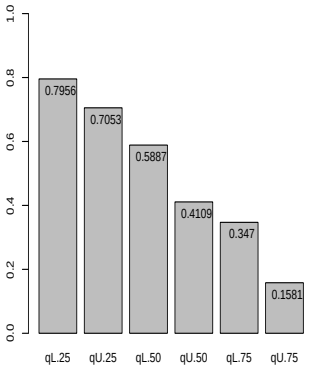
(d) m_2 to m_0 (0.01)



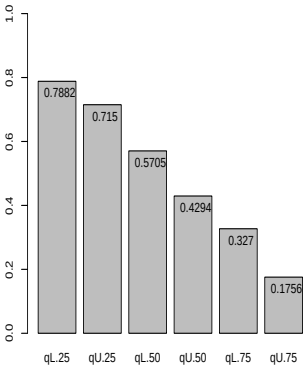
(e) m_2 to m_0 (0.05)



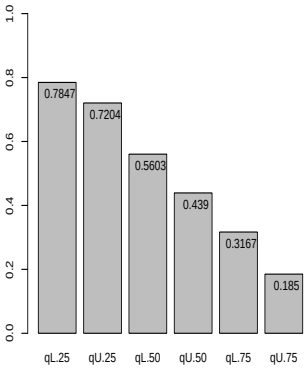
(f) m_2 to m_0 (0.10)



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

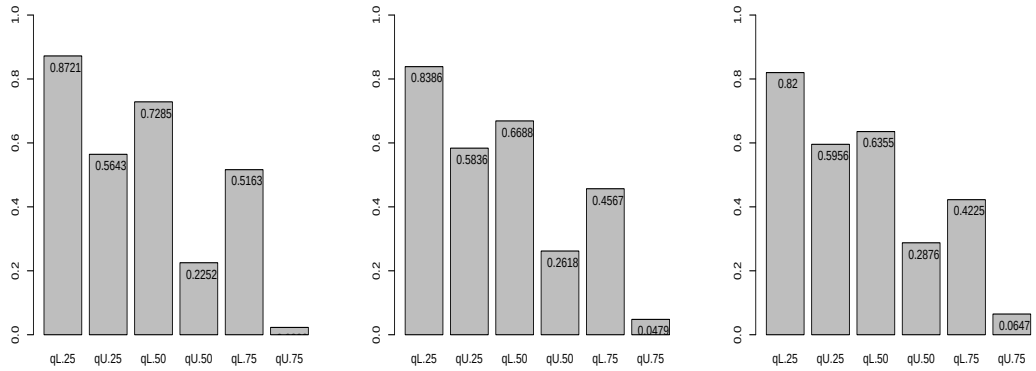


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

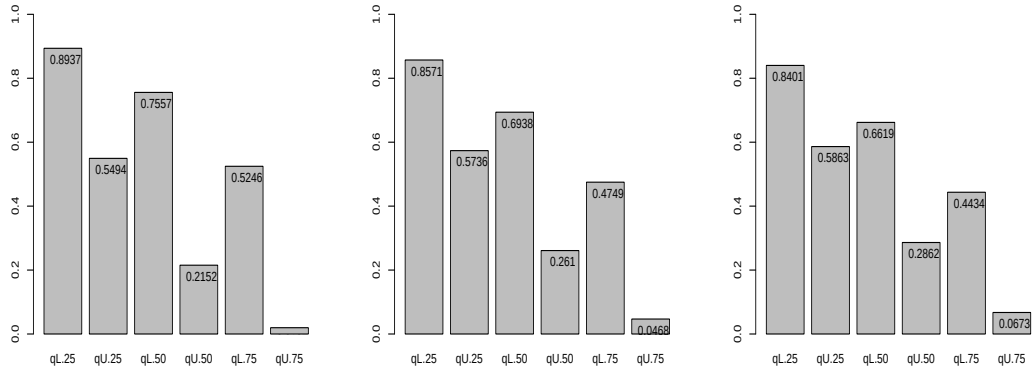


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

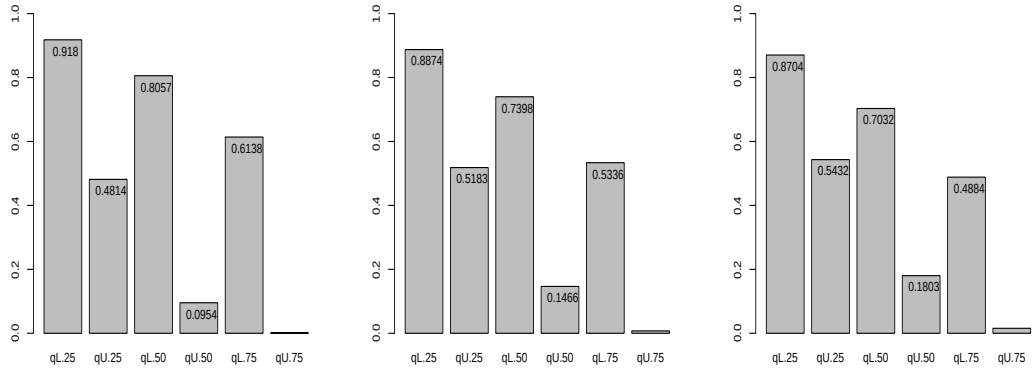
Figure A.3: Proportion of runs with future observation greater than the quartiles, Case 2, $\beta = 3$, $n = 100$. Section 3.5.



(a) m_1 to m_0(0.01) (b) m_1 to m_0 (0.05) (c) m_1 to m_0 (0.10)

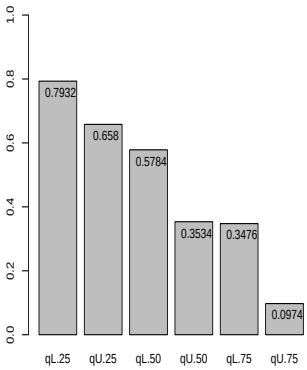


(d) m_2 to m_0 (0.01) (e) m_2 to m_0 (0.05) (f) m_2 to m_0 (0.10)

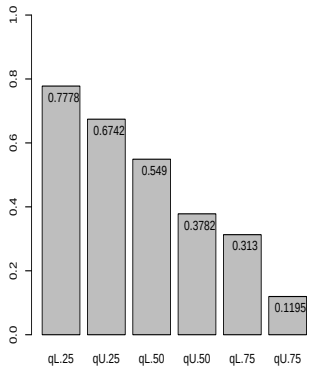


(g) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.01) (h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05) (i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

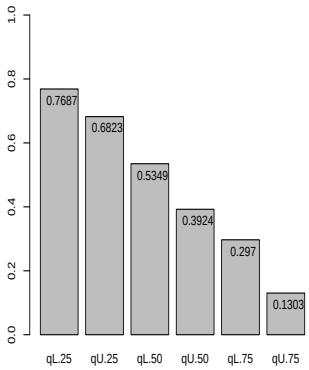
Figure A.4: Proportion of runs with future observation greater than the quartiles, Case 3, $n = 10$. Section 3.5.



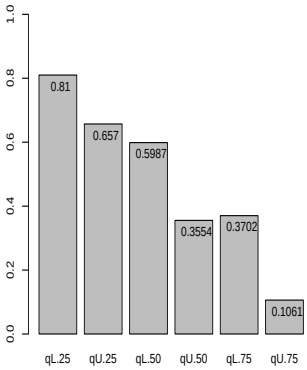
(a) m_1 to m_0 (0.01)



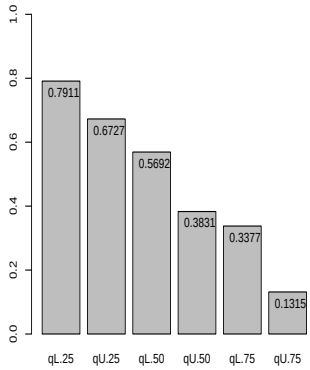
(b) m_1 to m_0 (0.05)



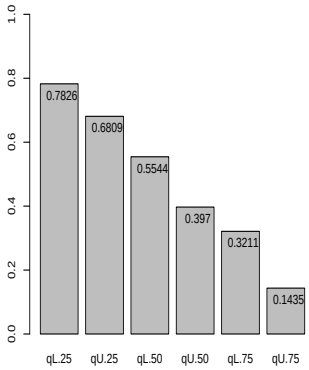
(c) m_1 to m_0 (0.10)



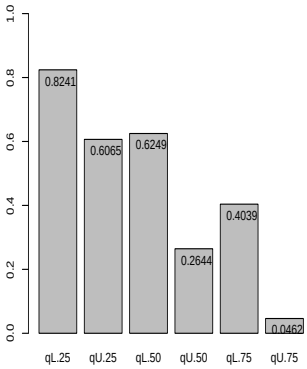
(d) m_2 to m_0 (0.01)



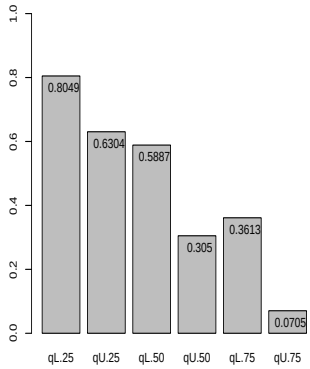
(e) m_2 to m_0 (0.05)



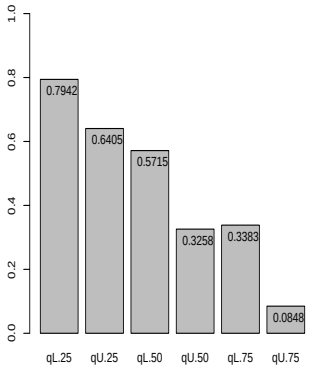
(f) m_2 to m_0 (0.10)



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

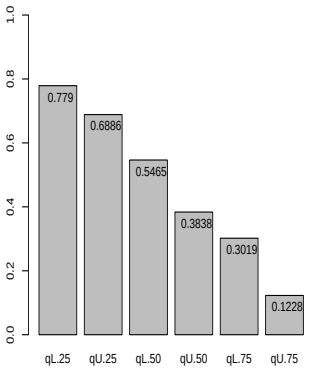


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

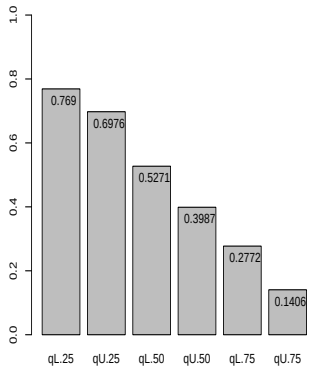


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

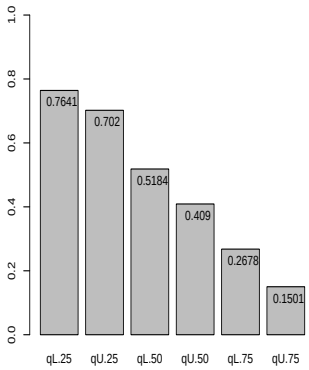
Figure A.5: Proportion of runs with future observation greater than the quartiles, Case 3, $n = 50$. Section 3.5.



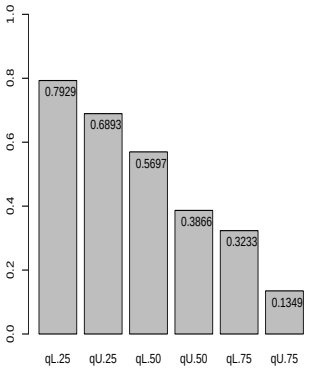
(a) m_1 to m_0 (0.01)



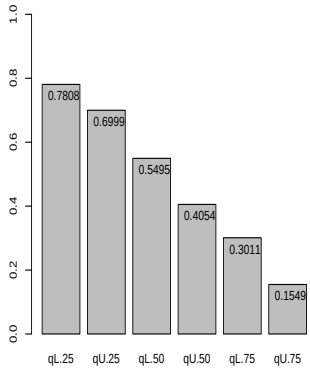
(b) m_1 to m_0 (0.05)



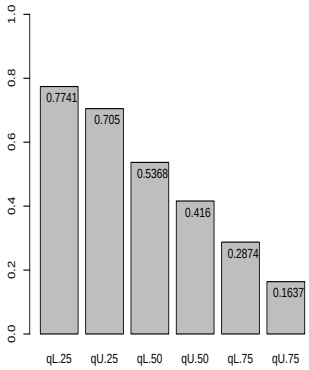
(c) m_1 to m_0 (0.10)



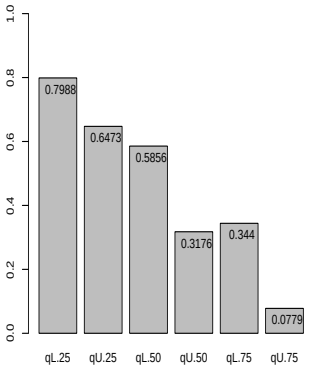
(d) m_2 to m_0(0.01)



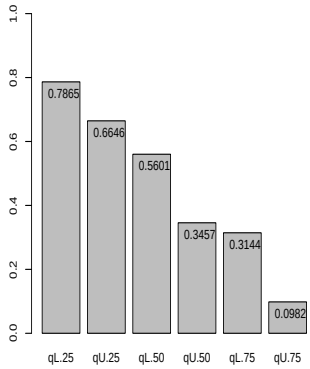
(e) m_2 to m_0 (0.05)



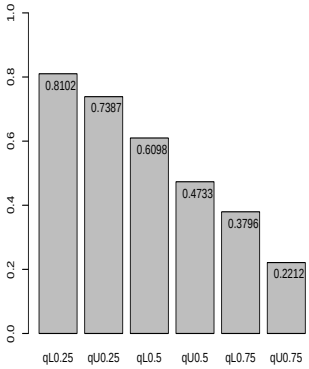
(f) m_2 to m_0 (0.10)



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

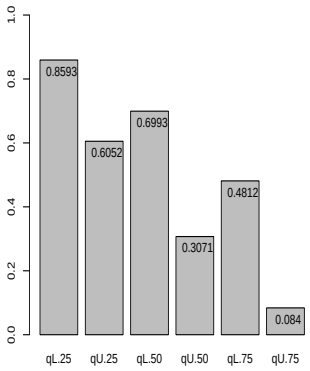


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

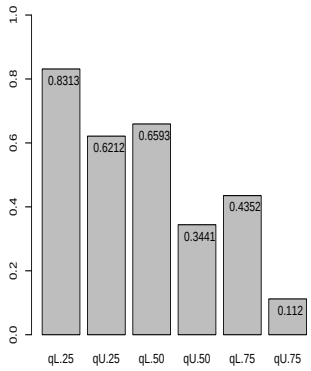


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

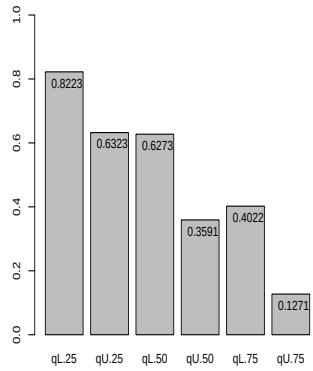
Figure A.6: Proportion of runs with future observation greater than the quartiles, Case 3, $n = 100$. Section 3.5.



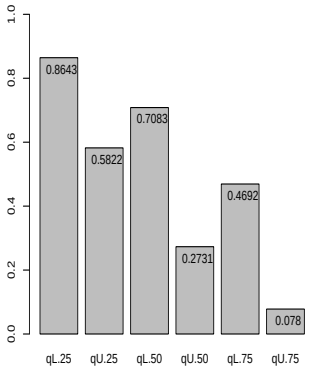
(a) m_1 to m_0 (0.01)



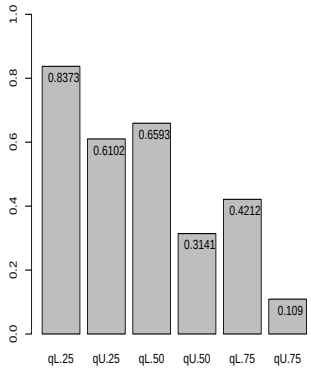
(b) m_1 to m_0 (0.05)



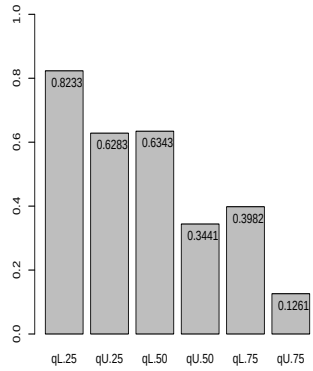
(c) m_1 to m_0 (0.10)



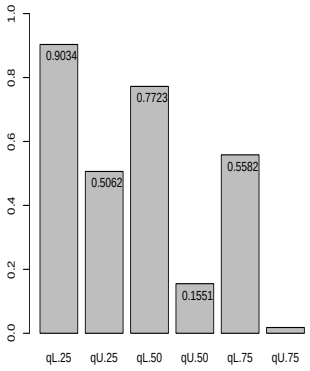
(d) m_2 to m_0 (0.01)



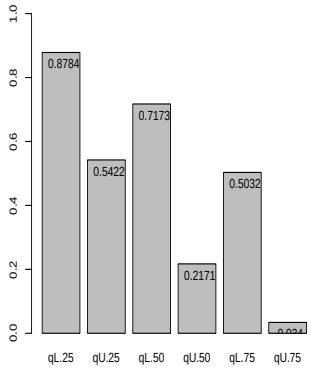
(e) m_2 to m_0 (0.05)



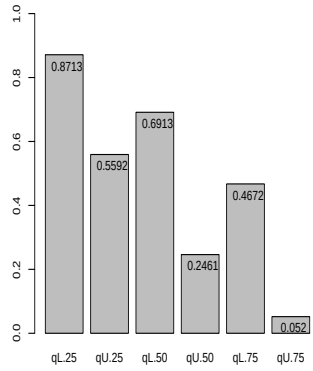
(f) m_2 to m_0 (0.10)



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

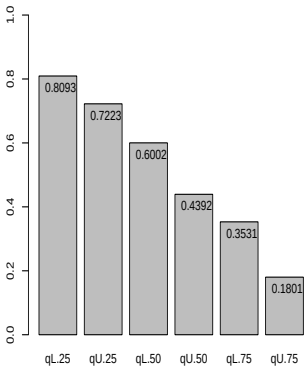


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

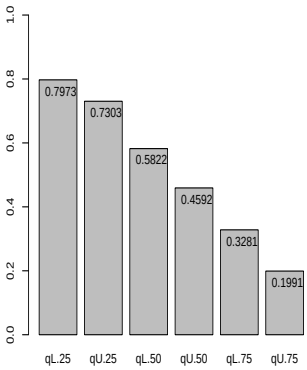


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

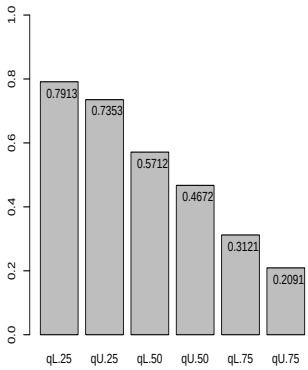
Figure A.7: Proportion of runs with future observation greater than the quartiles, Case 1, $n = 10$. Section 3.8.



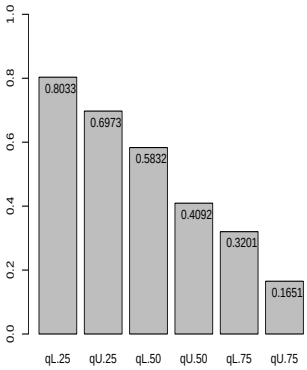
(a) m_1 to m_0 (0.01)



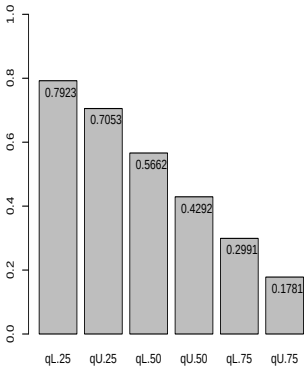
(b) m_1 to m_0 (0.05)



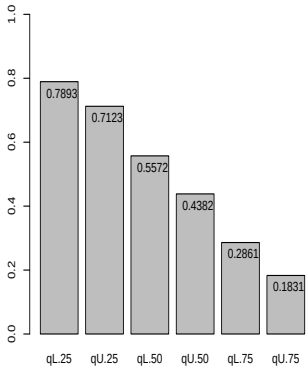
(c) m_1 to m_0 (0.10)



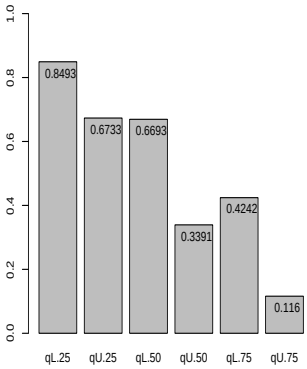
(d) m_2 to m_0 (0.01)



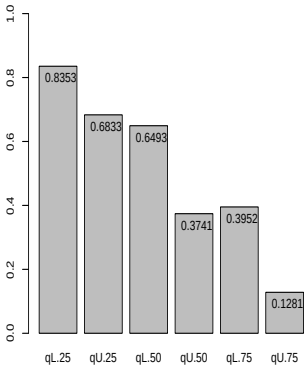
(e) m_2 to m_0 (0.05)



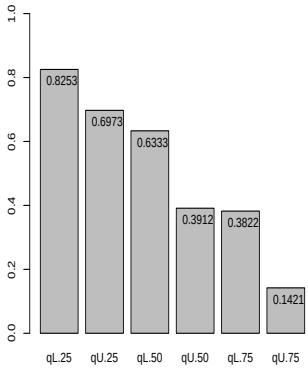
(f) m_2 to m_0 (0.10)



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

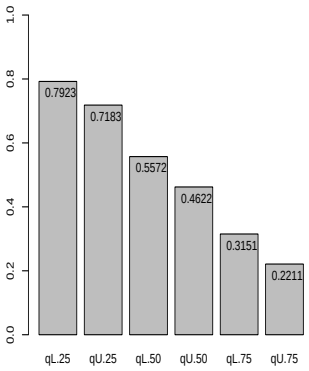


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

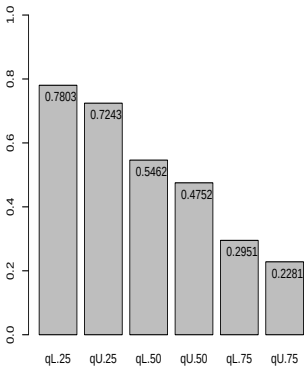


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

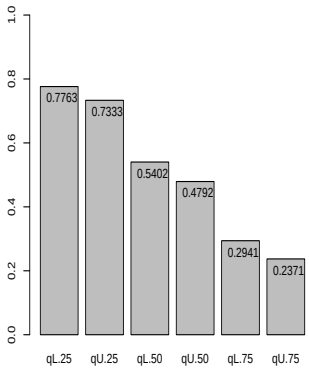
Figure A.8: Proportion of runs with future observation greater than the quartiles, Case 1, $n = 50$. Section 3.8.



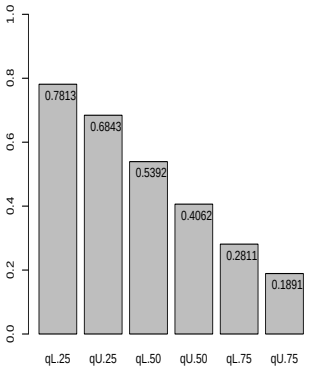
(a) m_1 to m_0 (0.01)



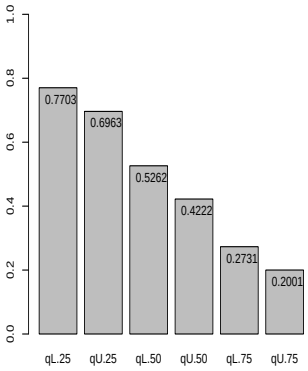
(b) m_1 to m_0 (0.05)



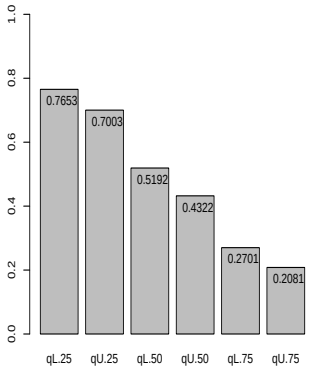
(c) m_1 to m_0 (0.10)



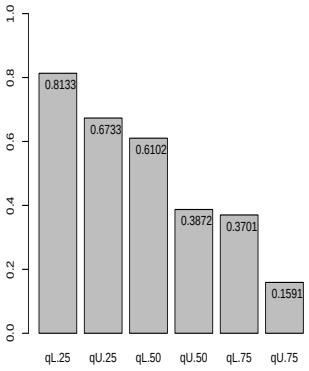
(d) m_2 to m_0 (0.01)



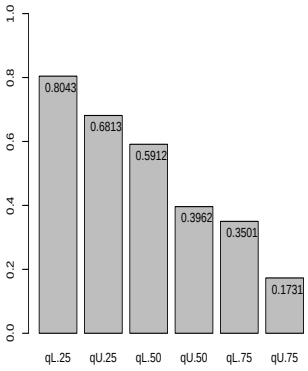
(e) m_2 to m_0 (0.05)



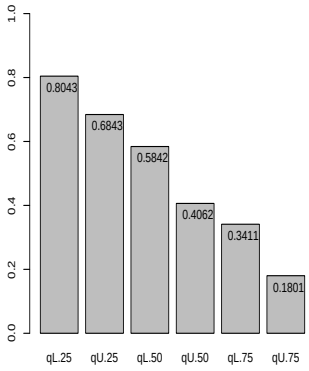
(f) m_2 to m_0 (0.10)



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

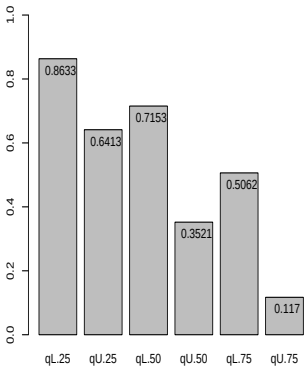


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

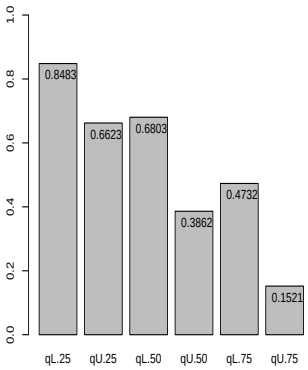


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

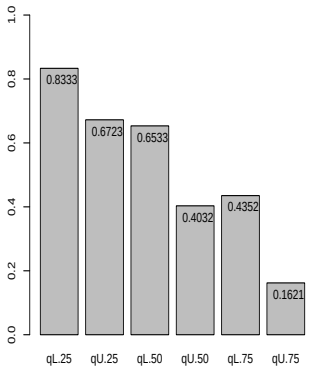
Figure A.9: Proportion of runs with future observation greater than the quartiles, Case 1, $n = 100$. Section 3.8.



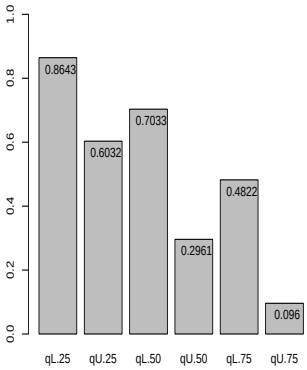
(a) m_1 to m_0 (0.01)



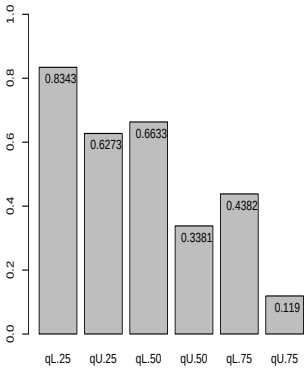
(b) m_1 to m_0 (0.05)



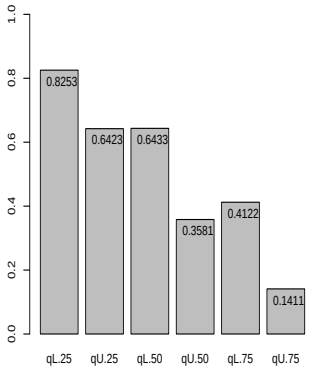
(c) m_1 to m_0 (0.10)



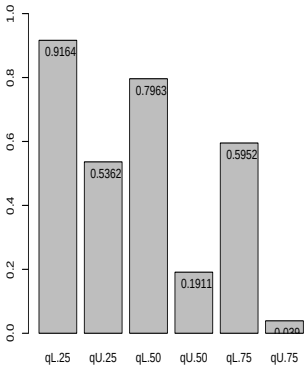
(d) m_2 to m_0 (0.01)



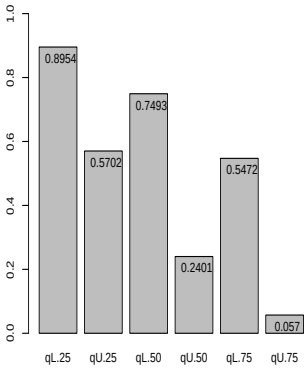
(e) m_2 to m_0 (0.05)



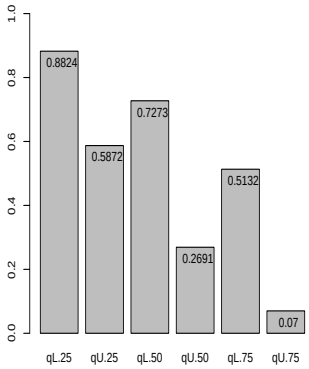
(f) m_2 to m_0(0.10)



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

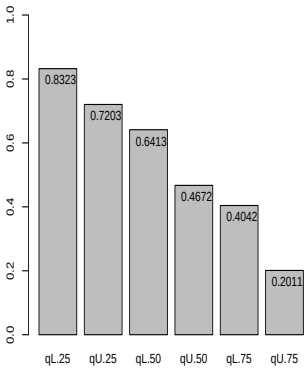


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

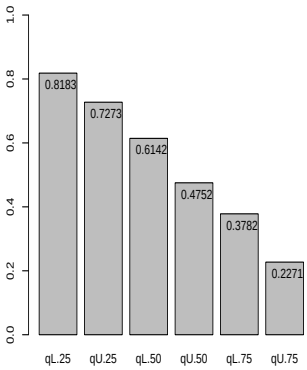


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

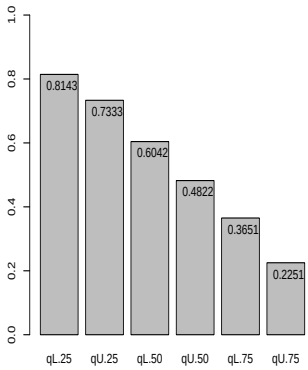
Figure A.10: Proportion of runs with future observation greater than the quartiles, Case 2, $n = 10$. Section 3.8.



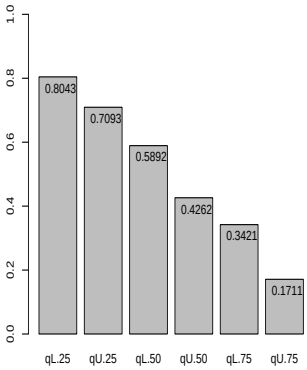
(a) m_1 to m_0 (0.01)



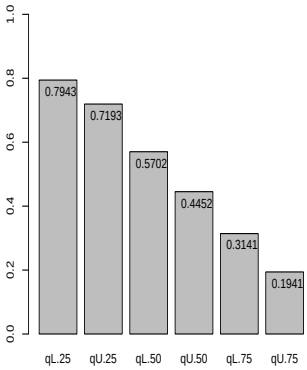
(b) m_1 to m_0 (0.05)



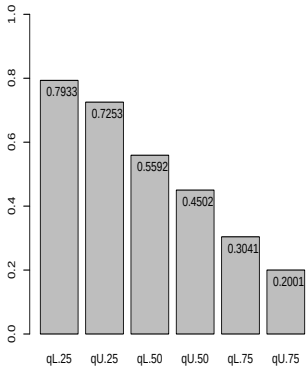
(c) m_1 to m_0 (0.10)



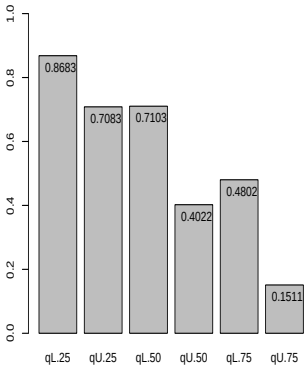
(d) m_2 to m_0 (0.01)



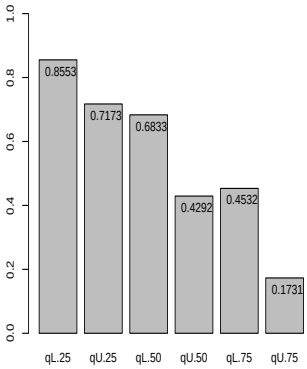
(e) m_2 to m_0 (0.05)



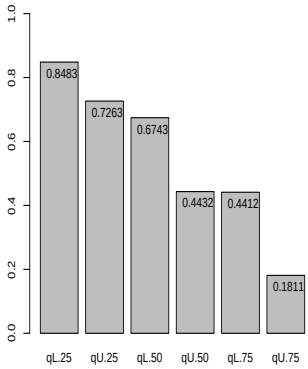
(f) m_2 to m_0 (0.10)



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

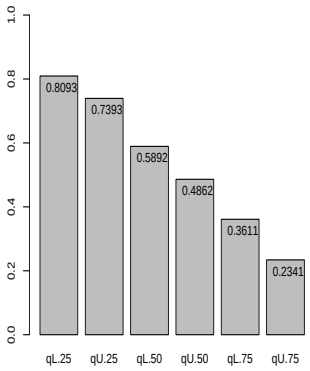


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

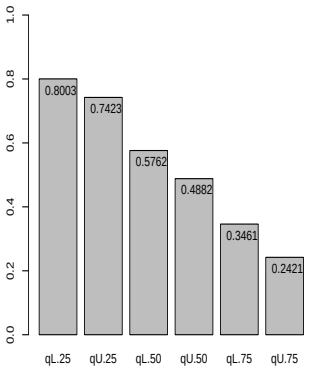


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

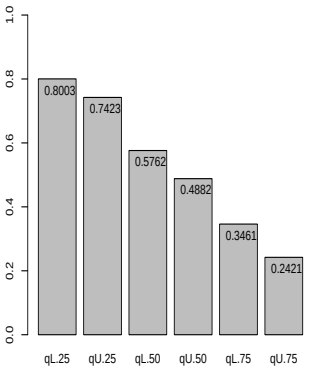
Figure A.11: Proportion of runs with future observation greater than the quartiles, Case 2, $n = 50$. Section 3.8.



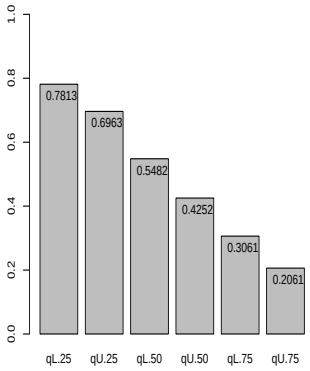
(a) m_1 to m_0 (0.01)



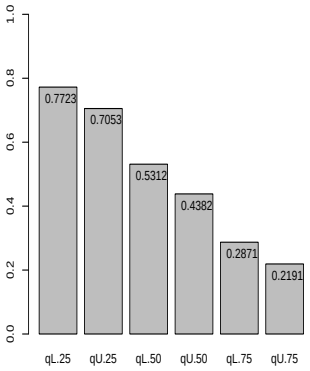
(b) m_1 to m_0 (0.05)



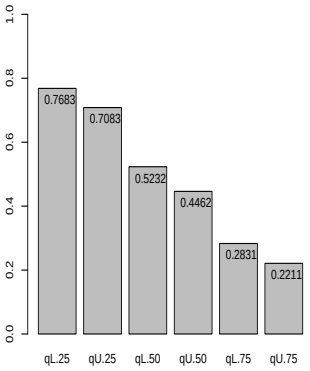
(c) m_1 to m_0 (0.10)



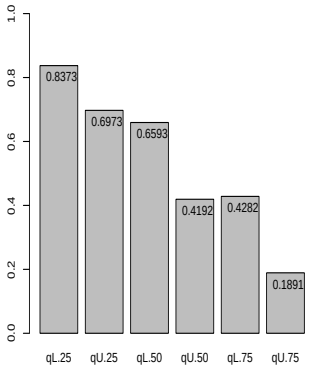
(d) m_2 to m_0 (0.01)



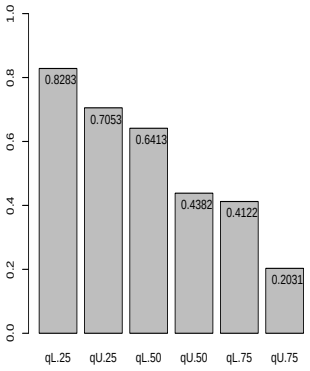
(e) m_2 to m_0 (0.05)



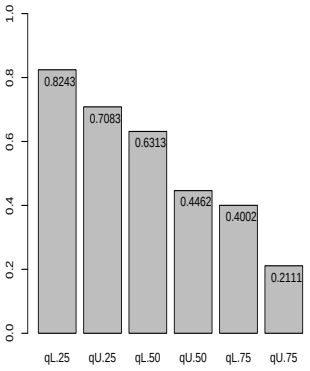
(f) m_2 to m_0 (0.10)



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

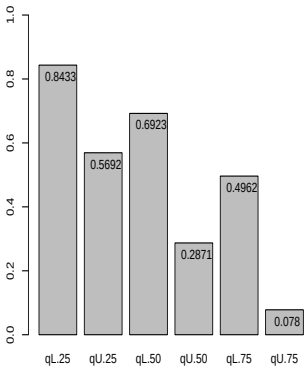


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

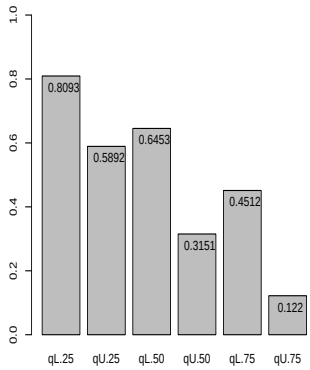


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

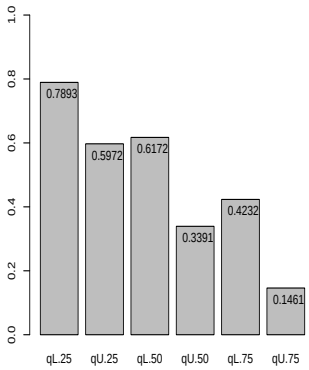
Figure A.12: Proportion of runs with future observation greater than the quartiles, Case 2, $n = 100$. Section 3.8.



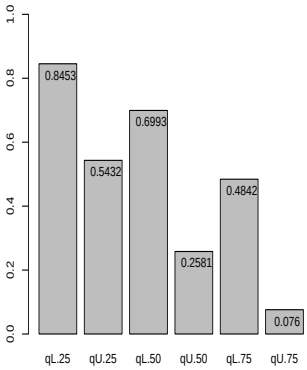
(a) m_1 to m_0 (0.01)



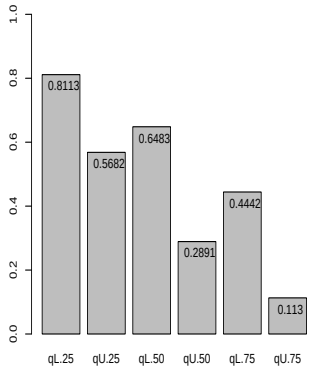
(b) m_1 to m_0 (0.05)



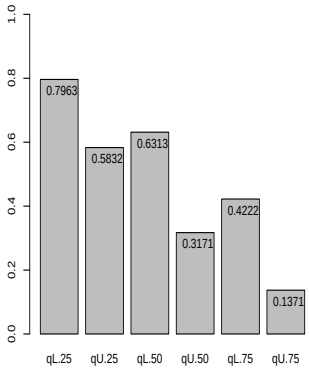
(c) m_1 to m_0 (0.10)



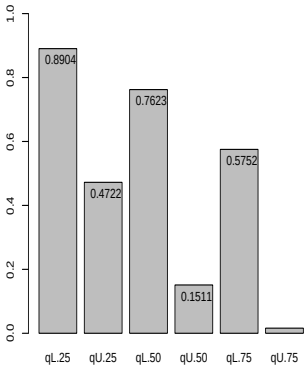
(d) m_2 to m_0 (0.01)



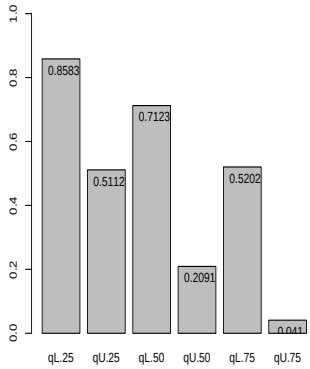
(e) m_2 to m_0 (0.05)



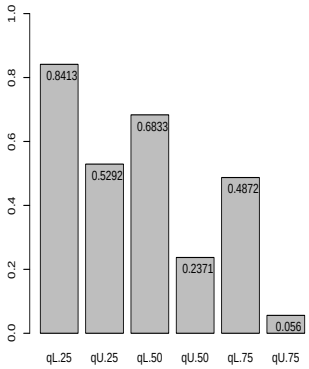
(f) m_2 to m_0 (0.10)



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

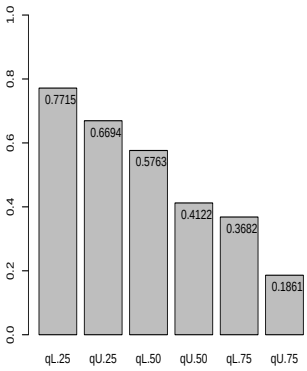


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

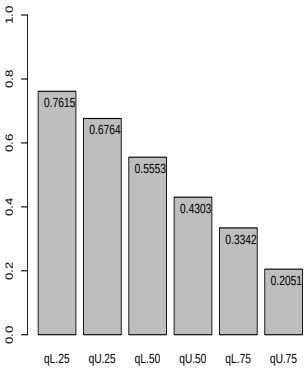


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

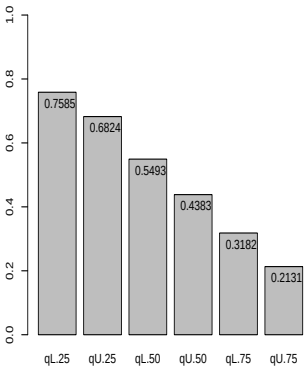
Figure A.13: Proportion of runs with future observation greater than the quartiles, Case 3, $n = 10$. Section 3.8.



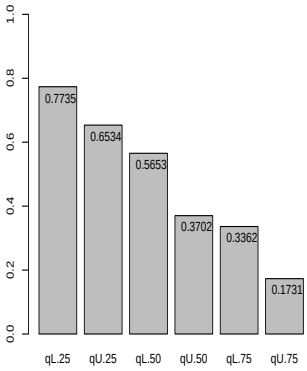
(a) m_1 to m_0 (0.01)



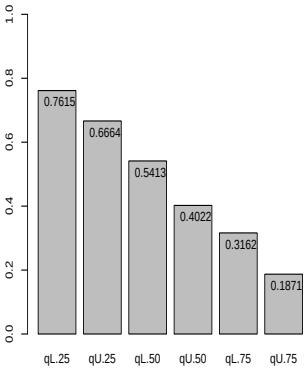
(b) m_1 to m_0 (0.05)



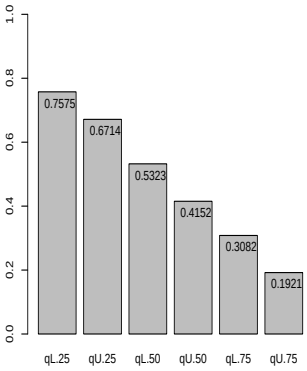
(c) m_1 to m_0 (0.10)



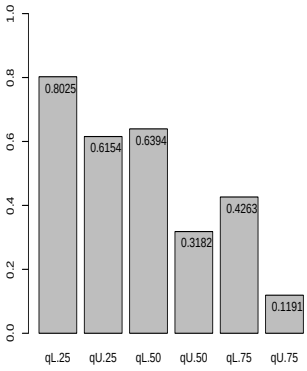
(d) m_2 to m_0 (0.01)



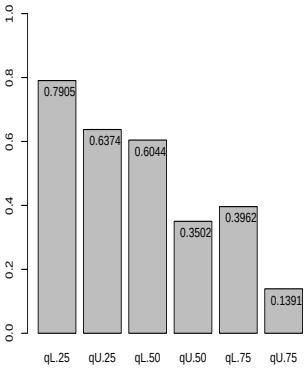
(e) m_2 to m_0 (0.05)



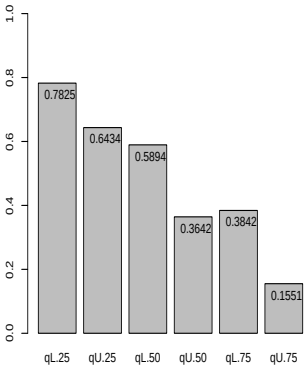
(f) m_2 to m_0 (0.10)



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

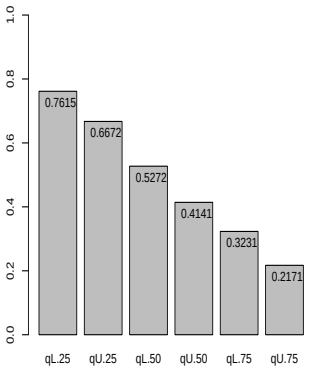


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

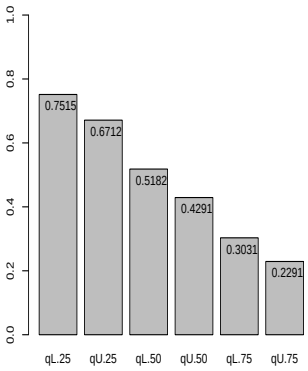


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

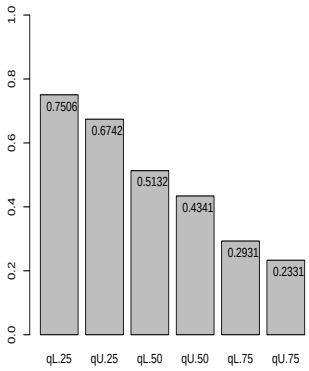
Figure A.14: Proportion of runs with future observation greater than the quartiles, Case 3, $n = 50$. Section 3.8.



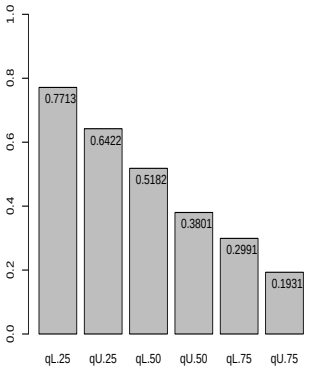
(a) m_1 to m_0 (0.01)



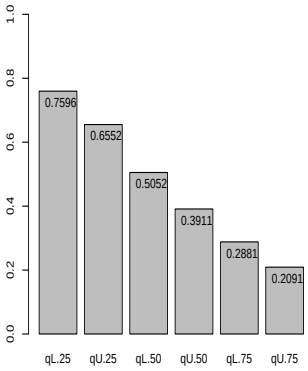
(b) m_1 to m_0 (0.05)



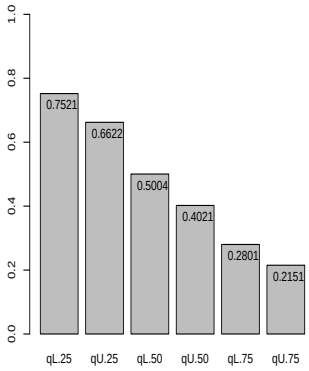
(c) m_1 to m_0 (0.10)



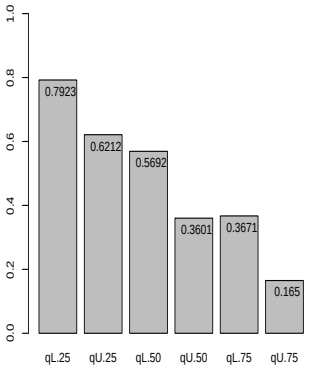
(d) m_2 to m_0 (0.01)



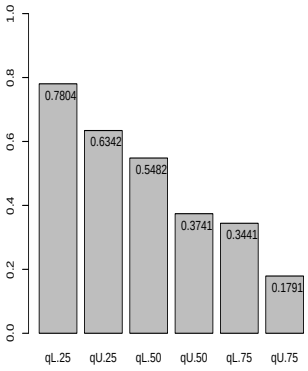
(e) m_2 to m_0 (0.05)



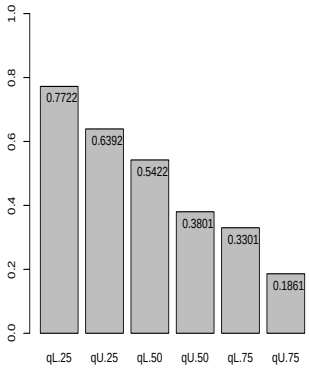
(f) m_2 to m_0 (0.10)



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

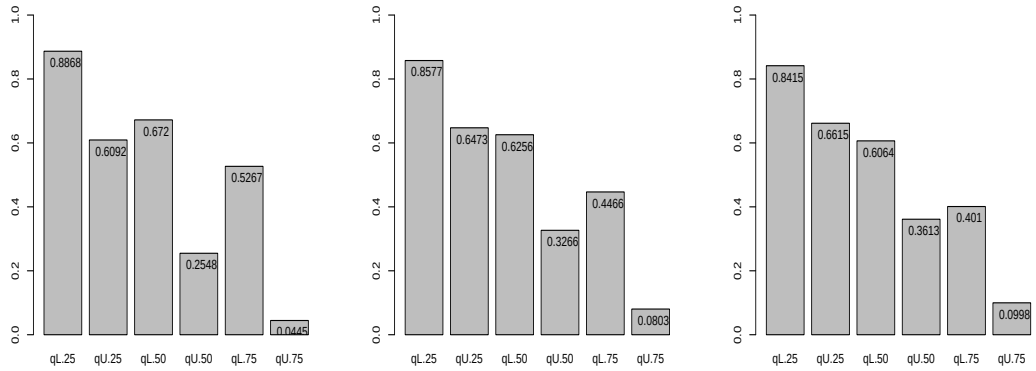


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

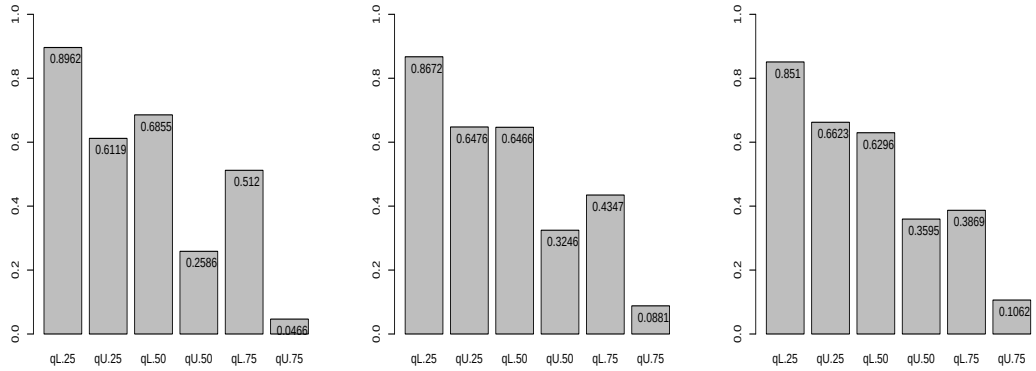


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

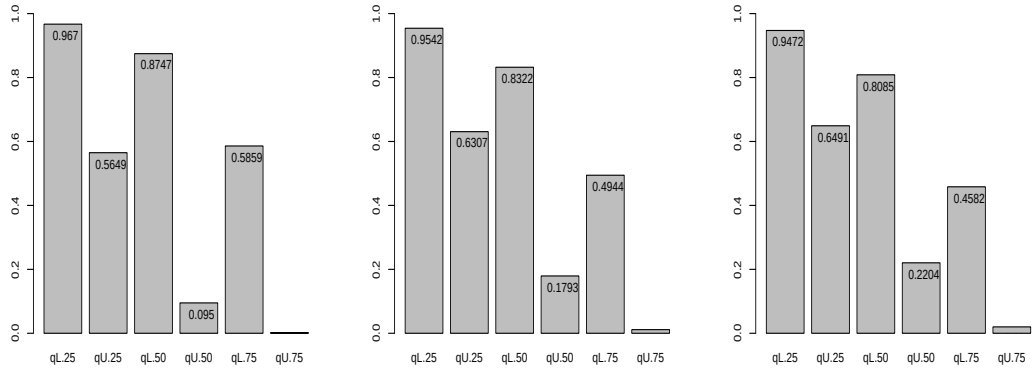
Figure A.15: Proportion of runs with future observation greater than the quartiles, Case 3, $n = 100$. Section 3.8.



(a) m_1 to m_0 (0.01) (b) m_1 to m_0 (0.05) (c) m_1 to m_0 (0.10)

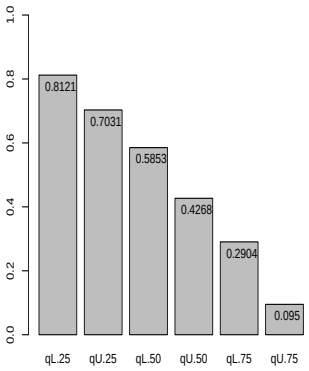


(d) m_2 to m_0 (0.01) (e) m_2 to m_0 (0.05) (f) m_2 to m_0 (0.10)

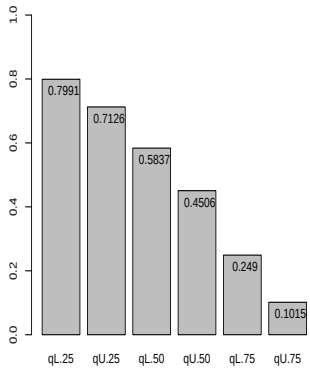


(g) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.01) (h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05) (i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

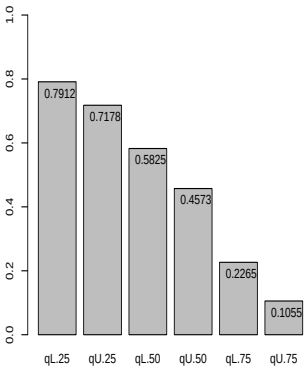
Figure A.16: Proportion of runs with future observation greater than the quartiles, Case 1, $n = 10$. Section 4.4.



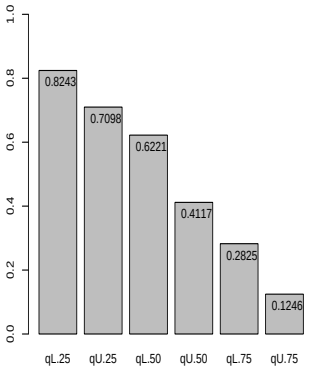
(a) m_1 to m_0 (0.01)



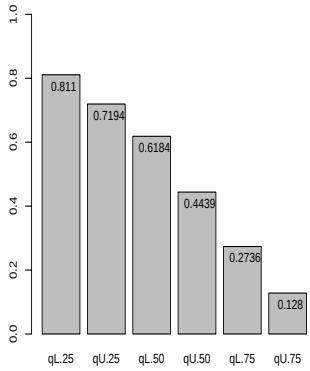
(b) m_1 to m_0 (0.05)



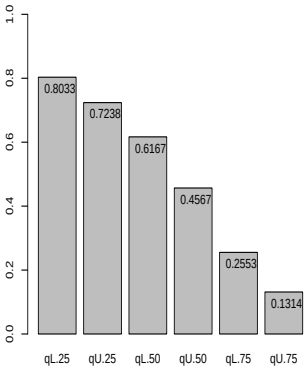
(c) m_1 to m_0 (0.10)



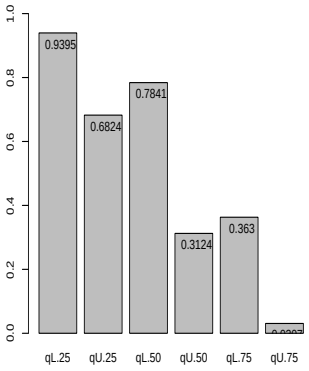
(d) m_2 to m_0 (0.01)



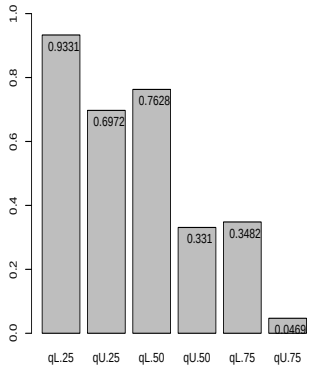
(e) m_2 to m_0 (0.05)



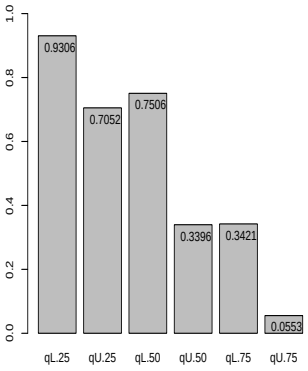
(f) m_2 to m_0 (0.10)



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

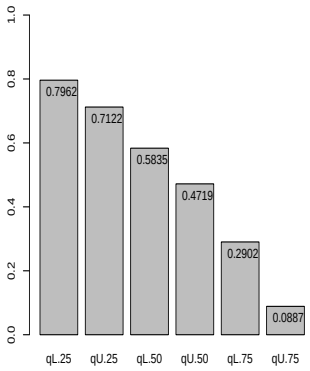


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

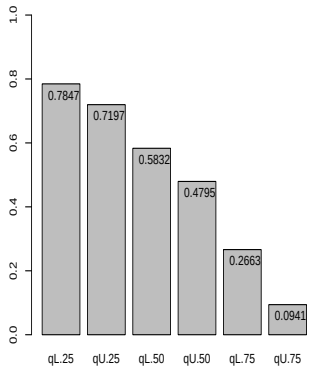


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

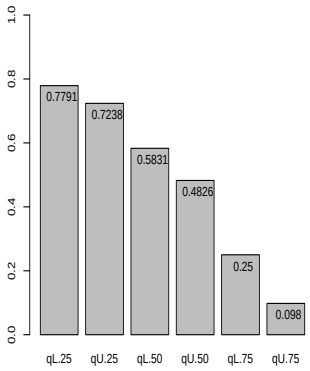
Figure A.17: Proportion of runs with future observation greater than the quartiles, Case 1, $n = 50$. Section 4.4.



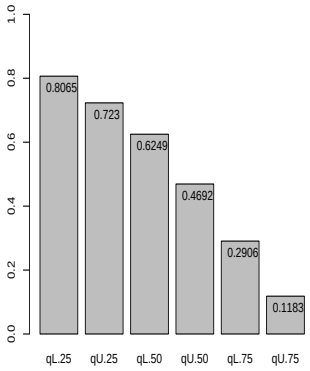
(a) m_1 to m_0 (0.01)



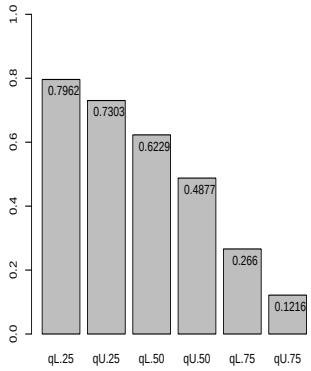
(b) m_1 to m_0 (0.05)



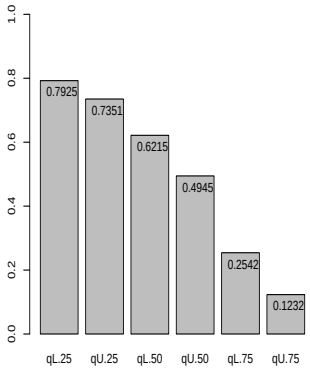
(c) m_1 to m_0 (0.10)



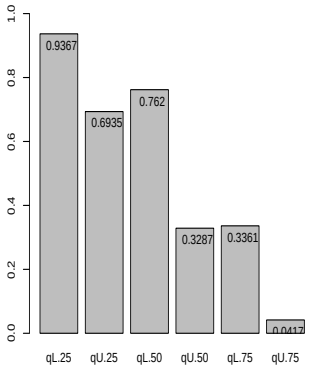
(d) m_2 to m_0 (0.01)



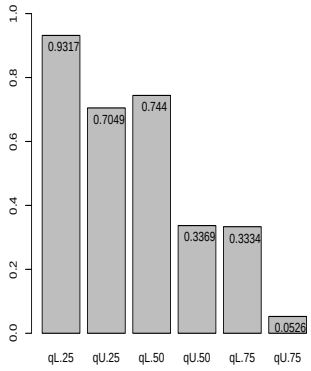
(e) m_2 to m_0 (0.05)



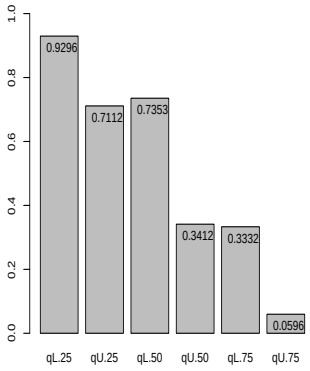
(f) m_2 to m_0 (0.10)



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

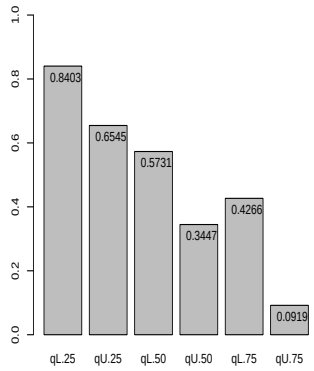


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

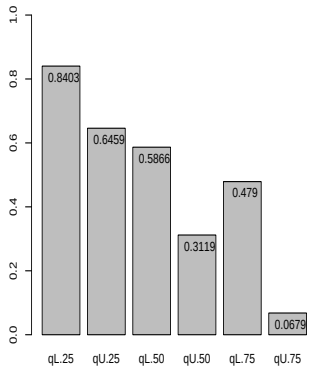


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

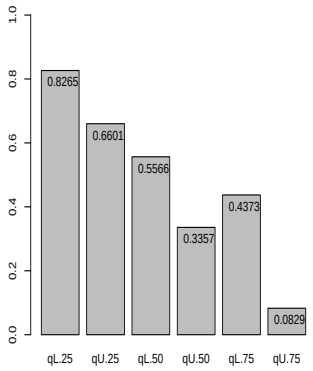
Figure A.18: Proportion of runs with future observation greater than the quartiles, Case 1 ,$n = 100$. Section 4.4.



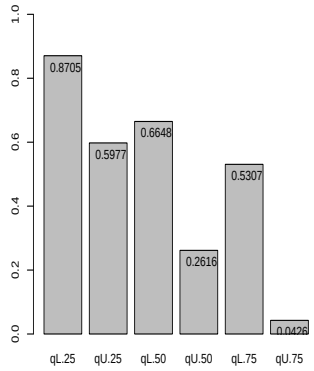
(a) m_1 to m_0 (0.01)



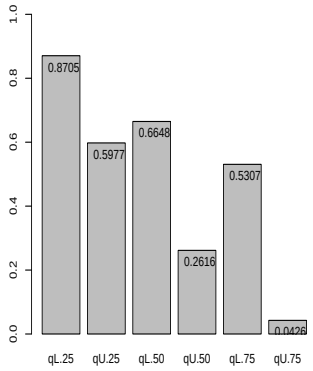
(b) m_1 to m_0 (0.05)



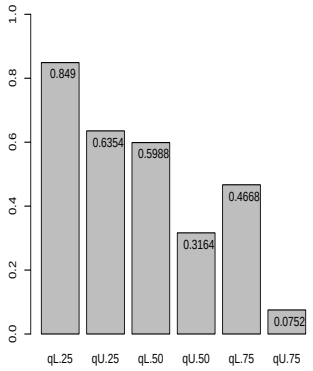
(c) m_1 to m_0 (0.10)



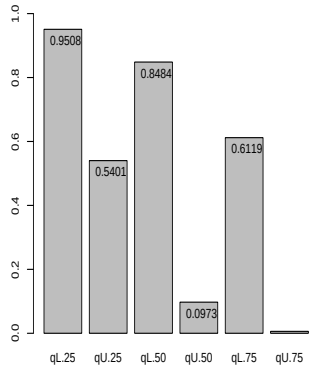
(d) m_2 to m_0 (0.01)



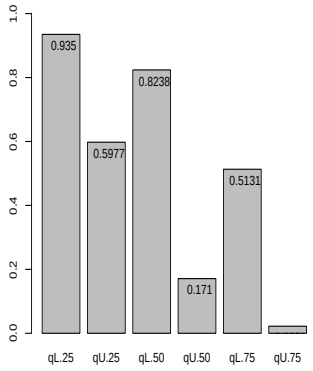
(e) m_2 to m_0 (0.05)



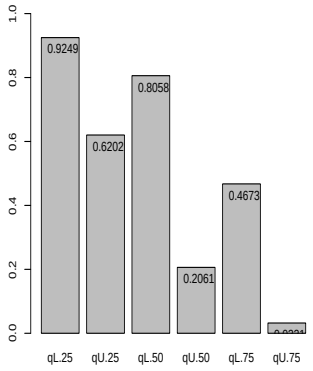
(f) m_2 to m_0 (0.10)



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

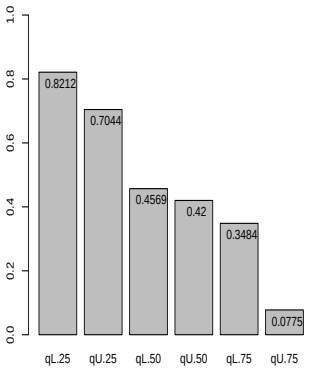


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)

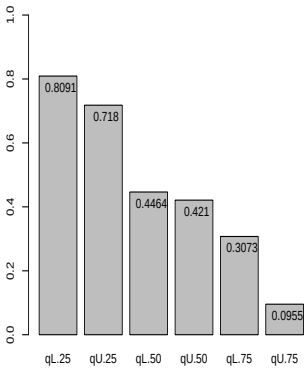


(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

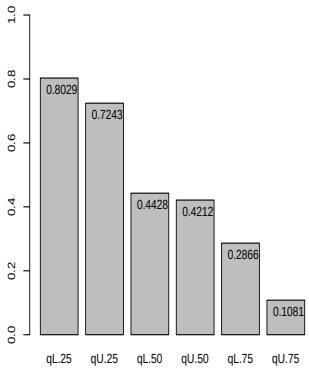
Figure A.19: Proportion of runs with future observation greater than the quartiles, Case 2, $n = 10$. Section 4.4.



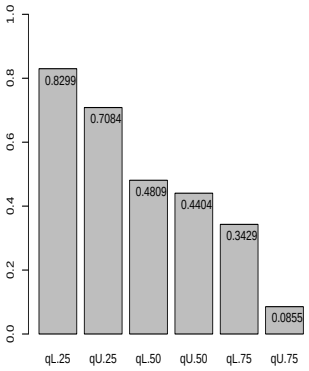
(a) m_1 to m_0 (0.01)



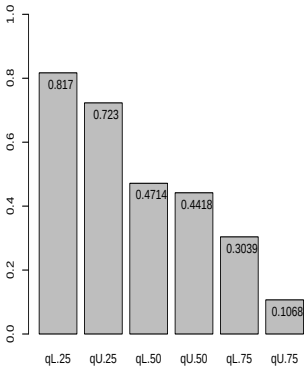
(b) m_1 to m_0 (0.05)



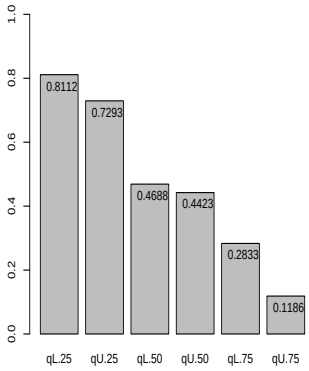
(c) m_1 to m_0(0.10)



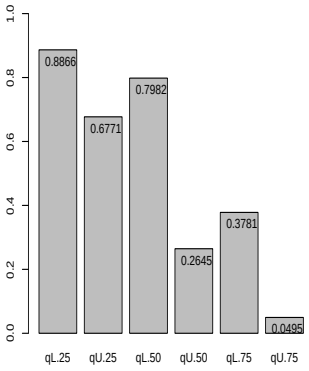
(d) m_2 to m_0(0.01)



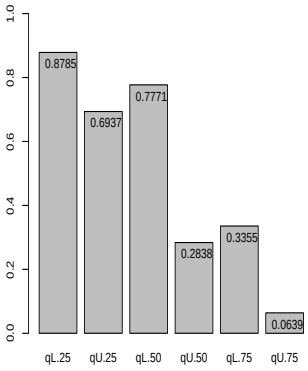
(e) m_2 to m_0 (0.05)



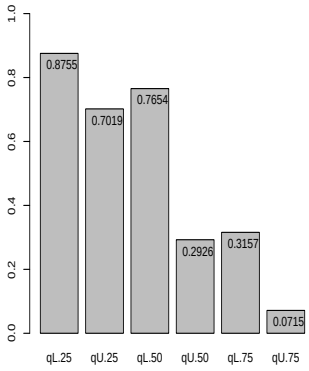
(f) m_2 to m_0 (0.10)



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

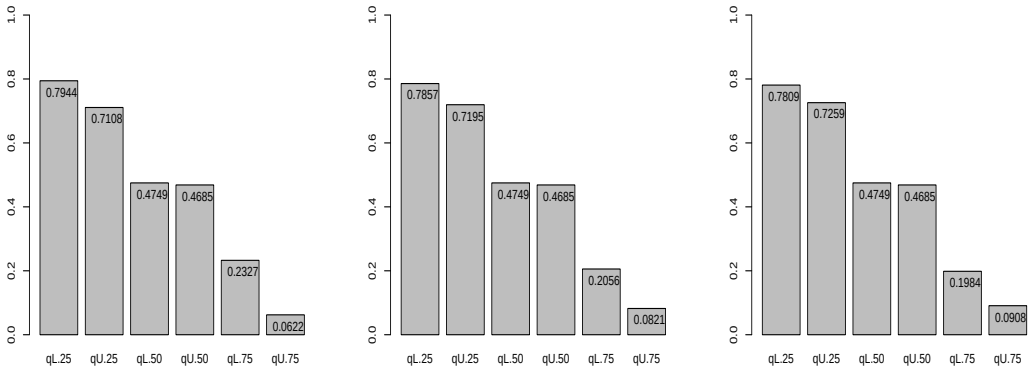


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)



(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

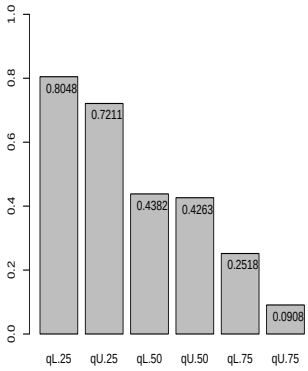
Figure A.20: Proportion of runs with future observation greater than the quartiles, Case 2, $n = 50$. Section 4.4.



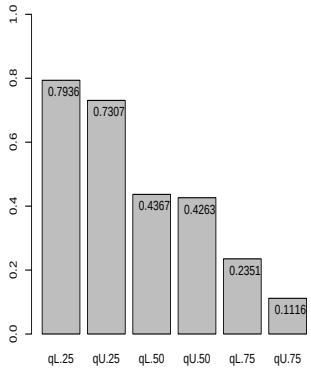
(a) m_1 to m_0 (0.01)

(b) m_1 to m_0 (0.05)

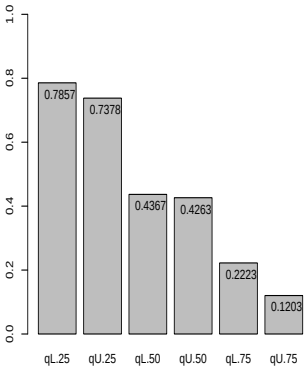
(c) m_1 to m_0 (0.10)



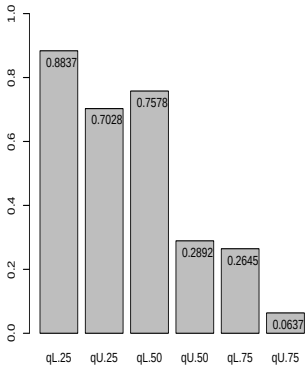
(d) m_2 to m_0 (0.01)



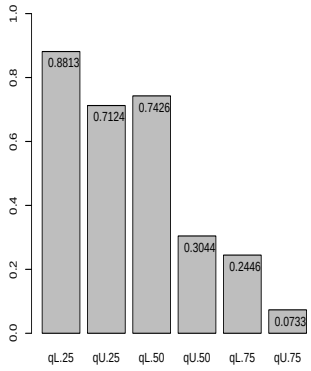
(e) m_2 to m_0 (0.05)



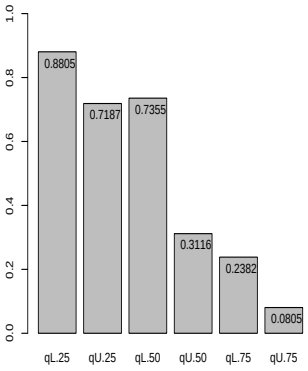
(f) m_2 to m_0 (0.10)



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

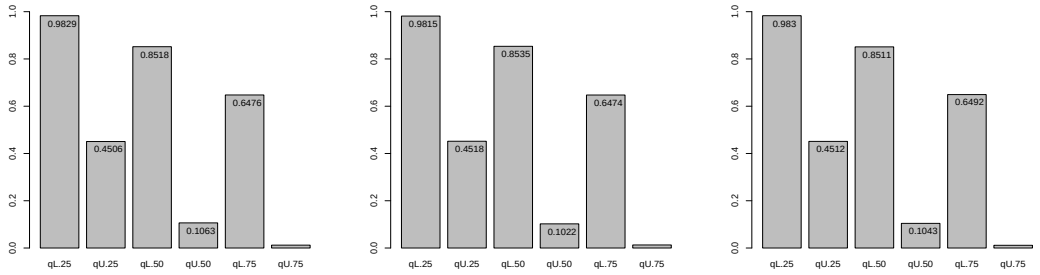


(h) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.05)



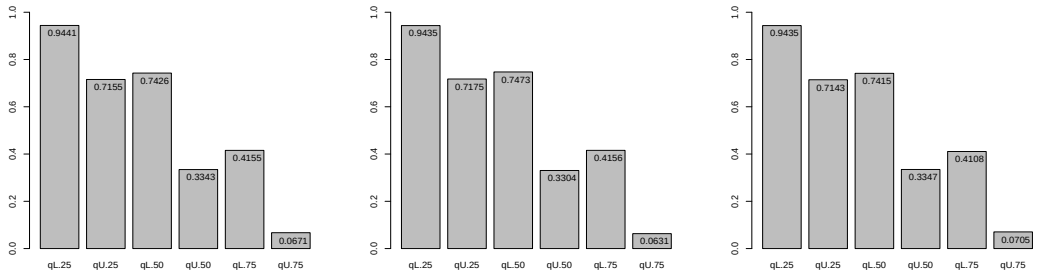
(i) Using $\underline{\gamma}$ and $\overline{\gamma}$ (0.10)

Figure A.21: Proportion of runs with future observation greater than the quartiles, Case 2, $n = 100$. Section 4.4.



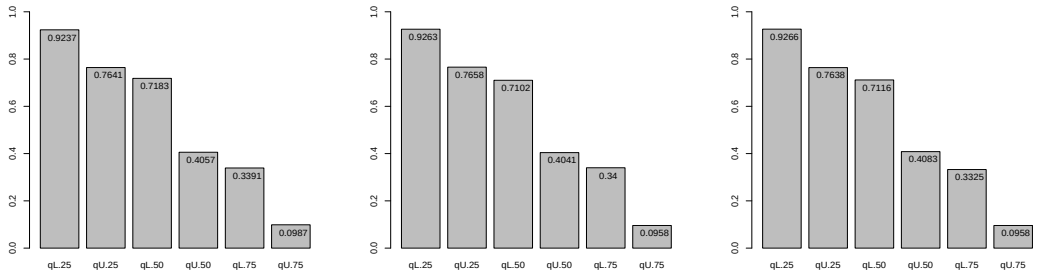
(a) m_1 to m_0 (b) m_2 to m_0 (c) Using min and max

Figure A.22: Proportion of runs with future observation greater than the quartiles, Case 2, $n = 10$. Section 5.7.



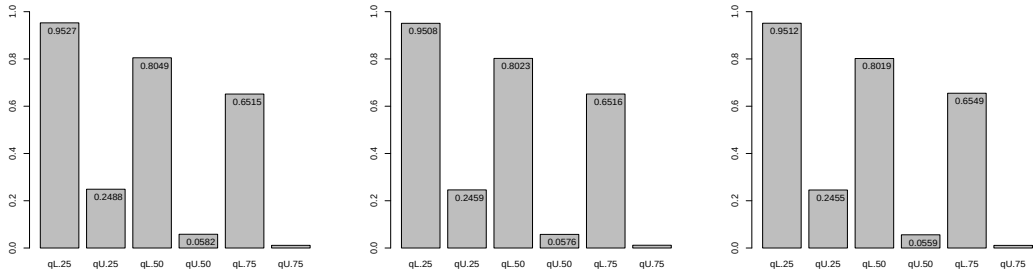
(a) m_1 to m_0 (b) m_2 to m_0 (c) Using min and max

Figure A.23: Proportion of runs with future observation greater than the quartiles, Case 2, $n = 50$. Section 5.7.



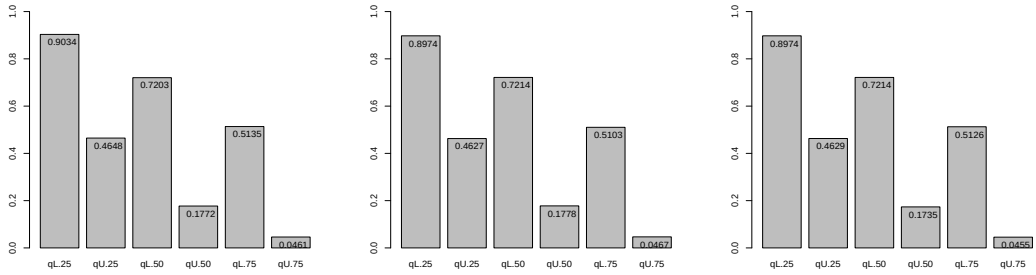
(a) m_1 to m_0 (b) m_2 to m_0 (c) Using min and max

Figure A.24: Proportion of runs with future observation greater than the quartiles, Case 2, $n = 100$. Section 5.7.



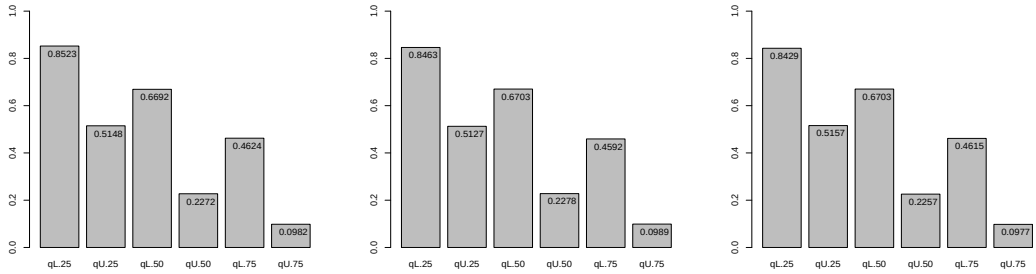
(a) m_1 to m_0 (b) m_2 to m_0 (c) Using min and max

Figure A.25: Proportion of runs with future observation greater than the quartiles, Case 3, $n = 10$. Section 5.7.



(a) m_1 to m_0 (b) m_2 to m_0 (c) Using min and max

Figure A.26: Proportion of runs with future observation greater than the quartiles, Case 3, $n = 50$. Section 5.7.



(a) m_1 to m_0 (b) m_2 to m_0 (c) Using min and max

Figure A.27: Proportion of runs with future observation greater than the quartiles, Case 3, $n = 100$. Section 5.7.

Appendix B

MCMC Trace Plots

This section presents the trace plots of the MCMC samples obtained in Chapter 5 for the Bayesian robust method. These plots serve to evaluate convergence and mixing of the posterior samples for each parameter (e.g., θ_0, and γ). Good mixing and lack of apparent trends suggest that the MCMC sampler has adequately explored the posterior distribution.

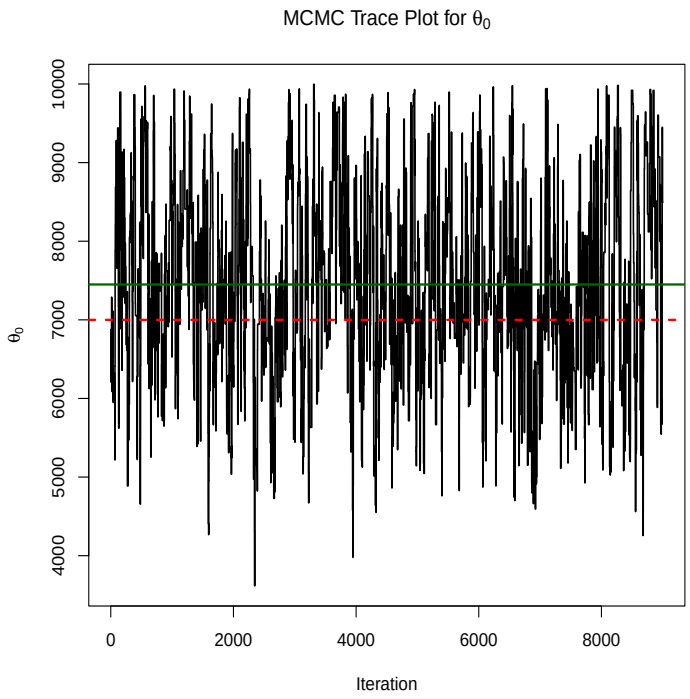


Figure B.1: MCMC trace plot for the parameter θ_0. The x-axis represents the MCMC iterations, and the black line shows the sampled values of θ_0 over time. The green solid line indicates the estimated value of the parameter θ_0 obtained from the posterior mean, while the red dashed line represents the true value used in the simulation. This result corresponds to Example 5.6.1 in Chapter 5, Case 1.

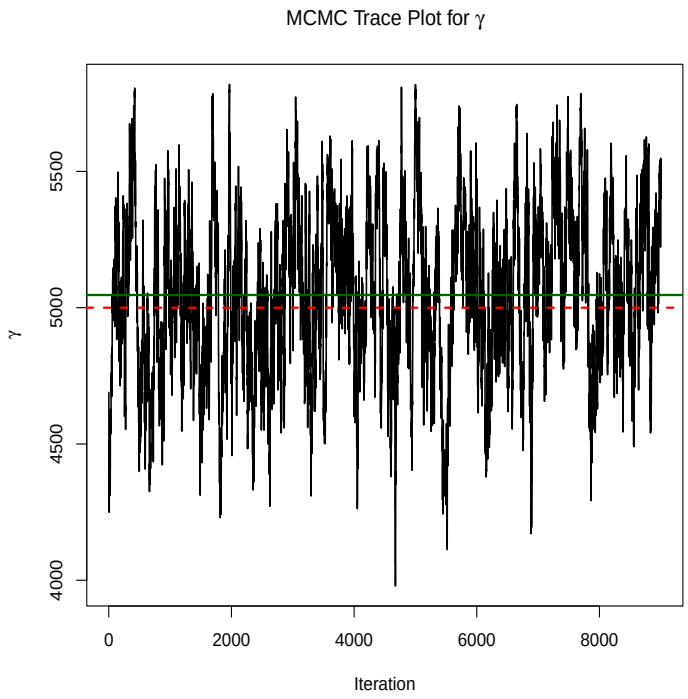


Figure B.2: MCMC trace plot for the parameter γ. The x-axis represents the MCMC iterations, and the black line shows the sampled values of γ over time. The green solid line indicates the estimated value of the parameter γ obtained from the posterior mean, while the red dashed line represents the true value used in the simulation. This result corresponds to Example 5.6.1 in Chapter 5, Case 1.

Bibliography

[1] Aboalkhair, A. (2012). *Nonparametric Predictive Inference for System Reliability*. PhD thesis, Durham University. Available from www.npi-statistics.com.

[2] Ahmadini, A. (2019). *Imprecise Statistical Methods for Accelerated Life Testing*. PhD thesis, Durham University. Available from www.npi-statistics.com.

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

[4] Augustin, T., Coolen, F.P.A., De Cooman, G. and Troffaes, M.C.M. (2014). *Introduction to imprecise probabilities*. Wiley.

[5] Baker, R. (2010). *Multinomial Nonparametric Predictive Inference: selection, classification and subcategory data*. PhD thesis, Durham University. Available from www.npi-statistics.com.

[6] Basu, S. (2000). Bayesian robustness and Bayesian nonparametrics. In *Robust Bayesian Analysis*. Springer-Verlag, New York.

[7] Berger, J.O., Bernardo, J.M. and Sun, D. (2015). Overall objective priors. *Bayesian Analysis*, 10:189–221.

[8] Berger, J.O., Moreno, E., Pericchi, L.R., Bayarri, M.J., Bernardo, J.M., Cano, J.A., De la Horra, J., Martín, J., Ríos-Insúa, D. and Betrò, B. (1994). An overview of robust bayesian analysis. *Test*, 3:5–124.

[9] Betrò, B., Ruggeri, F. and Meczarski, M. (1994). Robust bayesian analysis under generalized moments conditions. *Journal of Statistical Planning and Inference*, 41:257–266.

[10] Bhattacharyya, G.K. and Soejoeti, Z. (1989). A tampered failure rate model for step-stress accelerated life test. *Communications in Statistics Theory and Methods*, 18:1627–1643.

[11] Boole, G. (1854). *An Investigation of the Laws of Thought: on which are Founded the Mathematical Theories of Logic and Probabilities*. Walton and Maberly, London.

[12] Caruso, H. and Dasgupta, A. (1998). A fundamental overview of accelerated testing analytical models. *Journal of the Institute of Environmental Sciences and Technology*, 41:16–20.

[13] Coolen, F.P.A. (1998). Low structure imprecise predictive inference for Bayes' problem. *Statistics & Probability Letters*, 36:349–357.

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

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

[16] Coolen, F.P.A (2011). Nonparametric predictive inference. Technical report, Durham University. Available from www.npi-statistics.com.

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

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

[19] Cox, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34:187–202.

[20] De Finetti, B. (1974). *Theory of Probability*. Wiley, Chichester.

[21] DeGroot, M.H. and Goel, P.K. (1979). Bayesian estimation and optimal designs in partially accelerated life testing. *Naval Research Logistics Quarterly*, 26:223–235.

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

[23] Escobar, L.A. and Meeker, W.Q. (2006). A review of accelerated test models. *Statistical Science*, 21:552–577.

[24] Fan, T.H. and 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.

[25] Fortini, S. and Ruggeri, F. (2000). On the use of the concentration function in bayesian robustness. In Insua, D.R. and Ruggeri, F., editors, *Robust Bayesian Analysis*, pages 109–126. Springer, New York.

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

[27] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2003). *Bayesian Data Analysis*. Chapman and Hall, New York, 2nd edition.

[28] Goel, P.K. (1971). Some estimation problems in the study of tampered random variables. Technical Report Technical Report No. 50, Department of Statistics, Carnegie-Mellon University, Pittsburgh, Pennsylvania.

[29] Goel, P.K. (1975). Consistency and asymptotic normality of maximum likelihood estimators. *Scandinavian Actuarial Journal*, 1975:109–118.

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

[31] Hamada, M.S., Martz, H.F., Reese, C.S. and Wilson, A.G. (2008). *Bayesian Reliability*. Springer, New York.

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

[33] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109.

[34] Hill, B. M. (1988). De Finetti's theorem, induction, and Bayesian nonparametric predictive inference (with discussion). In Bernardo, J. M., DeGroot, M.H., Lindley

D.V. and Smith, A., editors, *Bayesian Statistics 3*, pages 211–241. Oxford University Press.

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

[36] Insua, D.R. and Criado, R. (2000). *Topics on the foundations of robust Bayesian analysis*. Springer, New York.

[37] John, P.K. and Melvin, L.M. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. Statistics for Biology and Health. Springer, New York, 2 edition.

[38] Kadane, J., Salinetti, G. and Srinivasan, C. (2000). *Stability of Bayes decisions and applications*, pages 187–196. Springer, New York.

[39] Kannan, N., Kundu, D. and Balakrishnan, N. (2010). *Survival models for step-stress experiments with lagged effects*, pages 355–369. Birkhäuser, Boston, MA.

[40] Kleinbaum, D.G. and Klein, M. (2005). *Survival analysis*. Statistics for Biology and Health. Springer-Verlag, New York.

[41] Kundu, D. and Ganguly, A. (2017). *Analysis of Step-stress Models: Existing Results and Some Recent Developments*. Academic Press, Cambridge, MA.

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

[43] MacEachern, S. and Müller, P. (2000). *Efficient "MCMC" schemes for robust model extensions using encompassing Dirichlet process mixture models*, pages 295–315. Springer, New York.

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

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

[46] Maturi, T. (2010). *Nonparametric Predictive inference for Multiple comparisons*. PhD thesis, Durham University. Available from www.npi-statistics.com.

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

[48] Meeker, W.Q., Escobar, L.A. and Pascual, F.G. (2022). *Statistical methods for reliability data*. Wiley, New York.

[49] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N, Teller, A.H and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087–1092.

[50] Muhammad, N. (2016). *Predictive Inference with copulas for Bivariate Data*. PhD thesis, Durham University. Available from www.npi-statistics.com.

[51] Nelson, W.B. (1980). Accelerated life testing step-stress models and data analyses. *IEEE Transactions on Reliability*, 29:103–108.

[52] Nelson, W.B. (2009). *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*, volume 344. Wiley, New Jersey.

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

[54] Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society: Series A*, 135:185–198.

[55] Rice, J.A. (2007). *Mathematical statistics and data analysis*. Thomson/Brooks/Cole Belmont, CA.

[56] Robert, C.P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer, New York.

[57] Sedyakin, N. (1966). On one physical principle in reliability theory. *Technical Cybernetics*, 3:80–87.

[58] Shanahan, J. (2013). A new method for the comparison of survival distributions. Master's thesis, University of South Carolina. Available from https://scholarcommons.sc.edu/etd/555.

[59] Smit, N. (2021). *Accelerated life testing using the Eyring model for the Weibull and Birnbaum-Saunders distributions*. PhD thesis, North-West University, South Africa. Available from https://orcid.org/0000-0002-4570-033X.

[60] Society for Imprecise Probability: Theories and Applications (2025). SIPTA - The Society for Imprecise Probability: Theories and Applications. Available at http://www.sipta.org.

[61] Srivastava, P.W. (2016). *Optimum Accelerated Life Testing Models with Time-Varying Stresses*. World Scientific Publishing Co. Pte. Ltd., Singapore.

[62] Thiraviam, A. and Malone, L. (2011). Accelerated life testing of subsea equipment under hydrostatic pressure. *Marine Technology Society Journal*, 45:42–54.

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

[64] Van Dorp, J.R., Mazzuchi, T.A., Fornell, G.E. and Pollock, L.R. (1996). A Bayes approach to step-stress accelerated life testing. *IEEE Transactions on Reliability*, 45:491–498.

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

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

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

[68] Yang, R. and Berger, J.O. (1996). A catalog of noninformative priors. Technical report, Institute of Statistics and Decision Sciences, Duke University, Durham, NC, USA.

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