Bayesian accelerated life tests for the Weibull distribution under non-informative priors

A project submitted in fulfilment of the requirements for the degree of

Masters of Science

in

Mathematical Statistics

Department of Statistics

Rhodes University

by

Philip Mostert

November 2020

Supervisor: Mr S. Izally

Declaration

I, the undersigned, declare that the work contained in this thesis is my own work, except for references specifically indicated in the text, and that I have not previously submitted it elsewhere for degree purposes.

_____
Signature

_____
Date

Abstract

In a competitive world where products are designed to last for long periods of time, obtaining time-to-failure data is both difficult and costly. Hence for products with high reliability, accelerated life testing is required to obtain relevant life-data quickly. This is done by placing the products under higher-than-use stress levels, thereby causing the products to fail prematurely. Part of the analysis of accelerated life-data requires a life distribution that describes the lifetime of a product at a given stress level and a life-stress relationship – which is some function that describes the way in which the life distribution changes across different stress levels. In this thesis it is assumed that the underlying life distribution is the well-known Weibull distribution, with shape parameter constant over all stress levels and scale parameter as a log-linear function of stress. The primary objective of this thesis is to obtain estimates from Bayesian analysis, and this thesis considers five types of non-informative prior distributions: Jeffreys' prior, reference priors, maximal data information prior, uniform prior and probability matching priors. Since the associated posterior distribution under all the derived non-informative priors are of an unknown form, the propriety of the posterior distributions is assessed to ensure admissible results. For comparison purposes, estimates obtained via the method of maximum likelihood are also considered. Finding these estimates requires solving non-linear equations, hence the Newton-Raphson algorithm is used to obtain estimates. A simulation study based on the time-to-failure of accelerated data is conducted to compare results between maximum likelihood and Bayesian estimates. As a result of the Bayesian posterior distributions being analytically intractable, two methods to obtain Bayesian estimates are considered: Markov chain Monte Carlo methods and Lindley's approximation technique. In the simulation study the posterior means and the root mean squared error values of the estimates under the symmetric squared error loss function and the two asymmetric loss functions: the LINEX loss function and general entropy loss function, are considered. Furthermore the coverage rates for the Bayesian Markov chain Monte Carlo and maximum likelihood estimates are found, and are compared by their average interval lengths. A case study using a dataset based on accelerated time-to-failure of an insulating fluid is considered. The fit of these data for the Weibull distribution is studied and is compared to that of other popular life distributions. A full simulation study is conducted to illustrate convergence of the proper posterior distributions. Both maximum likelihood and Bayesian estimates are found for these data. The deviance information criterion is used to compare Bayesian estimates between the prior distributions. The case study is concluded by

finding reliability estimates of the data at use-stress levels.

Contents

Appendix C: Additional results for Chapter 5   **173**

List of Figures

List of Tables

List of Abbreviations

AF Acceleration factor

AIC Akaike information criterion

ALT Accelerated life testing

ARS Adaptive rejection sampling

BGR Brooks-Gelman-Ruben

BIC Bayesian information criterion

BUGS Bayesian inference using Gibbs sampling

CDF Cumulative density function

DIC Deviance information criterion

GELF General entropy loss function

HALT Highly accelerated life testing

KS Kolmogorov-Smirnov

kV Kilovolts

LINEX Linear exponential

MCMC Markov chain Monte Carlo

MCSE Monte Carlo standard error

MDI Maximal data information

MH Metropolis-Hastings

MLE Maximum likelihood estimate

NR Newton-Raphson

PALT Progressive-stress accelerated tests

PDF Probability density function

PMP Probability matching prior

PSRF Potential scale reduction factor

QQ Quantile-Quantile

QualAt Qualitative accelerated life tests

QuanAT Quantitative accelerated life tests

RMSE Root mean squared error

SSALT Step-stress accelerated tests

STRIFE Stress life

WPP Weibull probability plot

List of Notations

T Random variable associated with time

$\boldsymbol{t}' = (t_1, t_2, ..., t_n)'$ Vector of n observed time variables

$f(\cdot)$ Probability density function

$F(\cdot)$ Cumulative density function

$R(\cdot)$ Reliability function

$\lambda(\cdot)$ Hazard function

$\Lambda(\cdot)$ Cumulative hazard function

$E(\cdot)$ Expected value

$Var(\cdot)$ Variance function

$\boldsymbol{s}' = (S_1, S_2, ..., S_p)'$ Vector of p Stressors

S_0 Use stress level

$\nu(\cdot)$ Life characteristic of a given life distribution

$\mu(S_i)$ Stress function for stress level i

$\boldsymbol{\theta}' = (\theta_1, \theta_2, ..., \theta_p)'$ Vector of p parameters

$\boldsymbol{\psi}' = (\psi_1, \psi_2, ..., \psi_p)'$ Vector of p parameters subject to transformation

$L(\boldsymbol{\theta} \mid \boldsymbol{t})$ Likelihood function for $\boldsymbol{\theta}'$

$\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{t})$ Log likelihood function for $\boldsymbol{\theta}'$

$\hat{\boldsymbol{\theta}}' = \left(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_p\right)'$ Vector of p estimators for $\boldsymbol{\theta}'$

$\nabla'(\cdot)$ Gradient vector

$H(\cdot)$ Hessian matrix

$\Sigma(\cdot)$ Variance-covariance matrix

$I\left(\cdot\right)$ Fisher information matrix

$\pi\left(\cdot\right)$ Prior distribution

$\pi\left(\boldsymbol{\theta}\mid\boldsymbol{t}\right)$ Posterior distribution for $\boldsymbol{\theta}'$

$K\left(\cdot\right)$ Kullback-Leibler distance

$\mathcal{H}\left(\cdot\right)$ Negative Shannon entropy

X_t Markov chain at position t

U Uniform random variable on interval $[0,1]$

$\boldsymbol{\varsigma}' = \left(\varsigma_1, \varsigma_2, ..., \varsigma_m\right)'$ Vector of m Markov chains

$L\left(\theta, \hat{\theta}\right)$ Loss function for $\hat{\theta}$ estimating θ

$\Gamma\left(\cdot\right)$ Gamma function

γ Euler's constant

$\mathcal{O}\left(\cdot\right)$ Big O function

$P\left(\theta_{\left(\frac{\alpha}{2}\right)} < \theta < \theta_{\left(1-\frac{\alpha}{2}\right)} \mid \boldsymbol{t}\right)$ Bayesian credibility interval

$P\left(\theta_{\left(\frac{\alpha}{2}\right)} < \theta < \theta_{\left(1-\frac{\alpha}{2}\right)} \mid \psi\right)$ Frequentist confidence interval

ℓ Average interval length

$\Psi\left(a, b\right)$ Indicator function on interval $\left(a, b\right)$

Acknowledgments

Chapter 1

Introduction

1.1 Overview

Reliability analysis is a field that lies on the interface of mathematics and engineering, and is used to understand the performance of a product under given conditions. However with higher consumer expectations and demands, products have been developed to have very high reliability under standard-use conditions, resultingly making the process of obtaining relevant time-to-failure data both costly and difficult.

A common way to deal with this problem is to expose the products to a range of higher-than-use stress levels, causing the product to fail faster, and thus obtaining time-to-failure data in a reasonable time frame – a process deemed accelerated life testing (ALT). One then uses the accelerated data to try and make inference on the products under use-stress levels.

When designing an accelerated life test, there are many factors which need to be considered, such as: the type of dataset used, the type and method of applying stress to the products and the life distribution used to model the products' failures.

In this thesis it is assumed that the time-to-failure of a product follows a Weibull distribution, where the scale parameter of the distribution is a log-linear function of stress, and the shape parameter is constant regardless of the level of stress applied. Furthermore it is assumed that the stress applied to the products is constant at every stress level, and that all failures are observed before the test is completed, that is, considering a complete dataset.

Statistical analysis requires finding estimates via some procedure, and in this thesis, the Bayesian approach is of primary consideration.

The Bayesian approach combines prior information on a test with data obtained from a test to form a posterior distribution – the distribution used to make inference. Often prior information of an experiment is not readily available; in these cases, non-informative prior distributions are used. This thesis considers five types of non-informative priors: Jeffreys' prior, reference priors, maximal data information (MDI) prior, uniform prior and probability matching priors (PMPs). When using non-informative priors, the form of the resultant

posterior is often not that of a known statistical distribution, hence the propriety of the posterior distribution needs to be considered before any inference is made.

A further difficulty with Bayesian estimates is that they are often analytically difficult to solve. Hence both Markov chain Monte Carlo (MCMC) techniques and a method proposed by Lindley (1980) will be used to find Bayesian estimates. The MCMC techniques used in this thesis are: slice sampling and the adaptive rejection sampling (ARS) technique. The simulation of the models was conducted on the Bayesian statistical software, $WinBUGS$ (Lunn et al., 2000).

In Bayesian literature it is common to contrast the results obtained with the results obtained via an alternative estimation technique. Therefore estimates obtained via the method of maximum likelihood are also considered in this thesis. The method of maximum likelihood often requires solving difficult, non-linear equations, thus the iterative Newton-Raphson (NR) algorithm is considered to approximate these equations and obtain estimates.

1.2 Objectives

The objectives of this thesis are given as:

- Form the likelihood function of the Weibull distribution, assuming its shape parameter is constant over all stress levels and its scale parameter is a log-linear time transformation function.

- Derive the equations to find the maximum likelihood estimates (MLEs) for the Weibull distribution under the log-linear time transformation function.

- Derive Jeffreys' prior for the Weibull distribution under the log-linear time transformation function.

- Derive the reference priors for the Weibull distribution under the log-linear time transformation function.

- Derive the MDI prior for the Weibull distribution under the log-linear time transformation function.

- Derive the uniform prior for the Weibull distribution under the log-linear time transformation function.

- Form the posterior distributions for the Weibull distribution under the log-linear time transformation function, using the above-mentioned prior distributions.

- Derive the second-order PMPs for the Weibull distribution under the log-linear time transformation function for all parameters of interest.

- Show which of the above-mentioned prior distributions are the second-order probability matching priors, regardless of the parameter of interest.

- Show properness of the posterior distribution for the Weibull distribution under the log-linear time transformation function using Jeffreys' prior and the reference priors.

- Show improperness of the posterior distribution for the Weibull distribution under the log-linear time transformation function using the MDI prior and uniform prior.

- Complete a convergence study for the MCMC Bayesian estimates for the Weibull distribution under the log-linear time transformation function.

- Complete a case study for the Weibull distribution under the log-linear time transformation function for a complete dataset, assuming its scale parameter is an inverse power law function of stress, using maximum likelihood and MCMC Bayesian estimates.

- Find the form of the Bayesian estimates for the Weibull distribution under the log-linear time transformation function with the approximation technique suggested by Lindley (1980).

- Complete a simulation study for the Weibull distribution under the log-linear time transformation function for maximum likelihood, MCMC and Lindley's estimates subject to the squared error loss function, linear exponential (LINEX) loss function and general entropy loss function (GELF).

- Find the coverage rates and average interval lengths for the Weibull distribution under the log-linear time transformation function for the maximum likelihood and MCMC estimates.

1.3 Contributions

The contributions from this thesis are as follows:

- Derive the posterior distribution for the Weibull distribution under the log-linear time transformation function using the MDI prior, and show that it is an improper distribution.

- Derive the posterior distribution for the Weibull distribution under the log-linear time transformation function using the uniform prior, and show that it is an improper distribution.

- Use the approximation technique from Lindley (1980) to find estimates from the Weibull distribution under the log-linear time transformation function.

- Complete a simulation study with the Weibull distribution under the log-linear time transformation function with both symmetric loss functions and asymmetric loss functions.

- Complete a full simulation study showing that the MCMC models for the posterior distribution of the Weibull under the log-linear time transformation function converge.

1.4   Thesis outline

The structure of this thesis is as follows:

Chapter 2 provides the necessary definitions and explanations required for this thesis. The main topics discussed in this chapter are: an introduction to reliability analysis and one of its branches, ALT, and an introduction to statistical estimation techniques, notably the Bayesian estimation technique and the many methods to derive estimates using Bayes' theorem.

In **Chapter 3** the Weibull distribution subject to a k-level constant stress accelerated life test is considered. We furthermore assume that the scale parameter of the Weibull is a log-linear time transformation function, and its shape parameter is constant for all levels of stress. Since the form of the likelihood distribution is difficult to work with, a transformation proposed by Xu et al. (2015) is considered. Following this transformation, maximum likelihood and non-informative Bayesian estimates are considered. Five non-informative Bayesian priors are formed, namely: Jeffreys' prior, reference priors, the MDI prior, the uniform prior and PMPs. The posterior distributions under each of these priors is formed and the properness of the resulting posteriors is considered.

Chapter 4 considers a simulation study for the Weibull distribution. Firstly, the form of Bayesian estimates with the approximation technique suggested by Lindley (1980) are found. The simulation study assumes a three-level constant-stress accelerated life test for the Weibull distribution, assuming its scale parameter is dependent on stress related to temperature, that is, the Arrhenius model is used. The simulation study finds both the estimated values and the root mean squared error (RMSE) values for the MLEs, and for non-informative Bayesian estimates using two approximation techniques: those found via an MCMC simulation test, and those found using the approximation technique suggested by Lindley (1980). The estimates are found using three loss functions: the squared error loss function, LINEX loss function, and GELF. The second part of the simulation study considers finding the coverage rates for both the maximum likelihood and MCMC Bayesian estimates.

In **Chapter 5** a case study for the Weibull distribution is considered using a complete dataset from Nelson (1990), by finding maximum likelihood and MCMC Bayesian estimates. Here it is assumed that the scale parameter is an inverse power law function of stress. A full convergence study is completed to ensure admissible results from the Bayesian MCMC estimates. The second part of the case study considers finding reliability estimates for the dataset under use-stress levels.

Chapter 6 provides concluding remarks of the results and relevant findings of this thesis. It also provides possibilities for future research of this thesis.

Appendix A provides additional results related to **Chapter 3** of this thesis. Notably it provides extensive derivations to some of the proofs provided by Xu et al. (2015), as well as preliminaries given by Abramowitz and Stegun (1964) and Ramos et al. (2020) required for the proofs of properness of the posterior distributions.

Appendix B provides additional results related to **Chapter 4** of this thesis. It provides derivations showing the log-concavity of one of the conditional posteriors under the general reference prior, additional tables related to posterior means, root mean squared errors and coverage rates as well as the $R^®$ (R Core Team, 2013) and $WinBUGS$ (Lunn et al., 2000) code used in both **Chapter 4** and **Chapter 5**.

Appendix C provides additional results related to **Chapter 5** of this thesis. It contains additional plots used in the convergence study as well as additional estimation results.

Chapter 2

Literature review

2.1 Introduction to reliability analysis

Rapid increases in technology, global competition and consumer expectations have put great pressure on manufacturers to produce high-quality and reliable products. It is expected, with high probability, that these products should operate under standard operating conditions without flaws for a sustained period of time.

Reliability analysis is a field that lies on the interface of mathematics and engineering, and uses these tools to understand the performance of a product. The topic of reliability (and other similar topics, such as: risk and survival analysis) has appeared in literature as early as the 17th century in the form of morality tables, however in the 20th century the field experienced explosive growth, being represented in a wide variety of academic journals – predominantly in statistics (Singpurwalla, 2006).

According to Soyer et al. (2012), the field of reliability is composed of three basic components, namely a:

1. structural,

2. stochastic and

3. statistical component.

The structural component is concerned with how the product is put together, and is often concerned with how well the system is designed in comparison to another. The stochastic component is concerned with modeling the uncertainties inherent in the product. And the statistical component is concerned with methods for learning about the performance of a product, given observed data.

The variable of interest in reliability analysis is clearly time, which is non-negative and can be either continuous or discrete. Our aim is to determine the performance of a product based on the frequency of failures over some period of time (Zacks, 2012). Cox and Oakes (1984) provide three requirements required before the frequency of failure is obtained, namely:

1. a well defined time origin for each product,

2. a scale to measure the passage of time and

3. the root-cause of failure.

The rest of this section is devoted to discussing in detail:

- the relevant functions used in reliability analysis and

- the different types of datasets used in a reliability analysis.

2.1.1 Relevant functions in reliability analysis

Let T, some positive continuous random variable, denote the interval of time from a well-defined specific starting point to the occurrence of some event. If this event is time-to-failure of some item, then T is termed the reliability time.

The most important function in any statistical analysis is the probability density function (PDF), which denotes the probability that a failure occurs with in an interval $\triangle t$ (Rausand and Høyland, 2003). Mathematically it can be expressed as:

$$f\left(t\right) = \frac{P\left(t \leq T \leq t + \triangle t\right)}{\triangle t},$$

(2.1)

where:

- $f\left(t\right) \geq 0 \underset{t}{\forall}$ and

- $\int\limits_{0}^{\infty} f\left(t\right) dt = 1.$

Then define the cumulative density function (CDF) of T – denoted by $F\left(t\right)$ – as the probability that the lifetime of some item does not exceed some time t (Rausand and Høyland, 2003). Notationally this is given by:

$$F\left(t\right) = P\left(T \leq t\right) = \int\limits_{0}^{t} f\left(x\right) dx,$$

(2.2)

where $F\left(t\right)$ is a monotonically increasing function with $\underset{t \to 0}{lim} F\left(t\right) = 0$ and $\underset{t \to \infty}{lim} F\left(t\right) = 1$. Furthermore equation 2.2 implies that:

$$f\left(t\right) = \frac{d}{dt} F\left(t\right).$$

(2.3)

The primary interest of reliability analysis is to find the lifetime of a given item. Hence define the reliability function of T – denoted by $R\left(t\right)$ – which is the probability that the lifetime of

some item exceeds some time t (Rausand and Høyland, 2003). Notationally this is given by:

$$R(t) = P(T \geq t) = \int\limits_t^\infty f(x) \, dx. \tag{2.4}$$

From equation 2.4 above, it is obvious that:

$$R(t) = 1 - F(t), \tag{2.5}$$

and from equation 2.3 that:

$$f(t) = -\frac{d}{dt} R(t). \tag{2.6}$$

The reliability function is a monotonically decreasing function with $\lim\limits_{t \to 0} R(t) = 1$ and $\lim\limits_{t \to \infty} R(t) = 0$.

The hazard function – denoted by $\lambda(t)$ – is perhaps the most valuable function in reliability analysis, and is defined as the frequency with which an item fails (Rausand and Høyland, 2003). Mathematically it is given by:

$$\lambda(t) = \lim\limits_{t \to \infty} \frac{P(t \leq T \leq t + \triangle t \mid T > t)}{\triangle t} = \frac{f(t)}{R(t)}, \tag{2.7}$$

with the following properties:

- $\lambda(t) \geq 0 \underset{t}{\forall}$ and

- $\int\limits_0^\infty \lambda(t) \, dt = \infty.$

Therefore the hazard function is defined as the probability that an item will fail in the interval $(t, t + \triangle t]$ given that the item has survived at time t.

Using equation 2.3, it can be shown that:

$$\lambda(t) = -\frac{R'(t)}{R(t)} = -\frac{d}{dt} log(R(t)) \tag{2.8}$$

(Rausand and Høyland, 2003). Since $R(0) = 1$ it implies that:

$$\int\limits_0^t \lambda(x) \, dx = -log(R(t)). \tag{2.9}$$

Therefore:

$$R(t) = exp\left\{-\int\limits_0^t \lambda(x) \, dx\right\}. \tag{2.10}$$

Furthermore, a related quantity to the hazard function is the cumulative hazard function,

given by:

$$\Lambda(t) = \int_0^t \lambda(x)\,dx \tag{2.11}$$

(Klein and Moeschberger, 2006).

The average time until failure (or mean time to failure) is given by:

$$E(t) = \int_0^\infty t f(t)\,dt, \tag{2.12}$$

provided that this integral is finite, and the variance is given by:

$$Var(t) = \int_0^\infty (t - E(t))^2 f(t)\,dt \tag{2.13}$$

(Rausand and Høyland, 2003).

2.2 Types of data

The inferential process of a reliability analysis is dependent on how the data has been collected. In this section, two types of datasets considered in reliability analysis are discussed:

- complete datasets and

- censored datasets.

It is assumed that the dataset takes the form: $T_1, T_2, ..., T_n$ where T_i is given as the time-to-failure of item i, and have observed values given by: $t_1, t_2, ..., t_n$. It is also assumed that these items are censored independently, that is censoring occurs independent of any information gained from previously failed items in the same test.

2.2.1 Complete dataset

A dataset is defined to be complete if we are able to observe the failures of all n items being tested. We often order the dataset in an increasing sequence:

$$T_{(1)} \leq T_{(2)} \leq ... \leq T_{(n)}, \tag{2.14}$$

where $T_{(i)}$ is called the ith order statistic (Klein and Moeschberger, 2006).
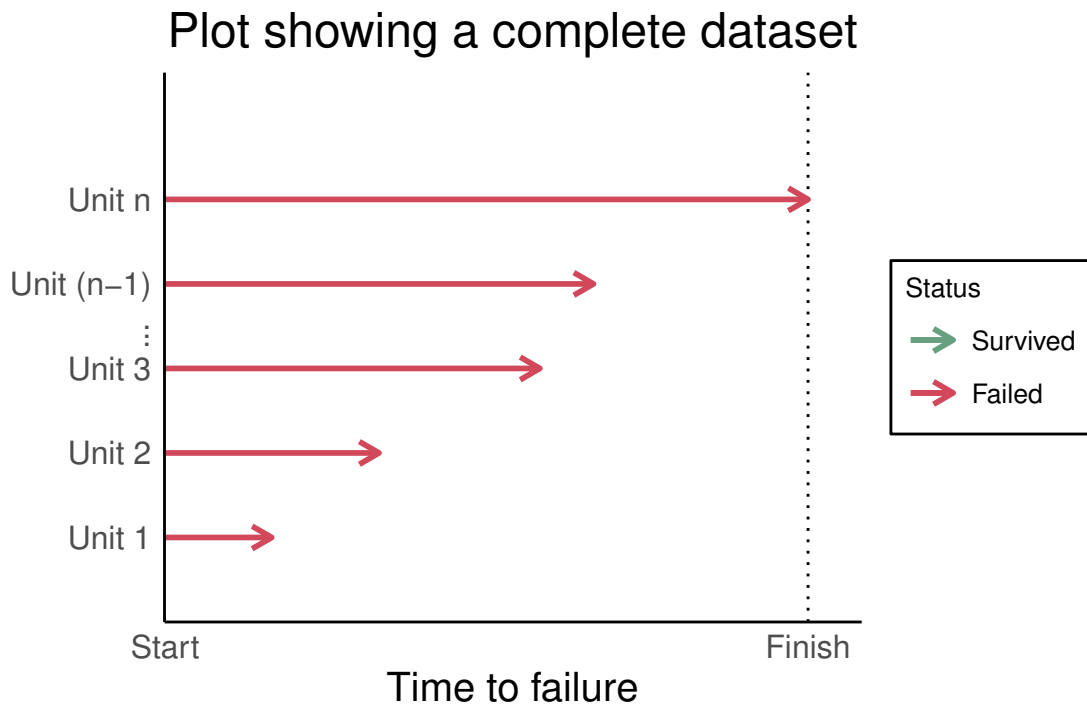
Figure 2.1: Plot illustrating a complete dataset.

2.2.2 Censored data set

Often due to time and cost considerations a test is terminated prematurely, resulting in some items not observing failure. These items are thus considered censored data. Klein and Moeschberger (2006) provided various categories of censoring, such as:

- Left-censored data,

- Right-censored data,

- Interval-censored data and

- Truncation.

There are two main types of censoring, type I and type II censoring, which will also be discussed below.

2.2.2.1 Left-censored data

A dataset is said to be left-censored when the failure time of an item is only known to be before a certain time. Notationally, the exact lifetime of a unit, T_i is considered left-censored if it is less than or equal to some value C_l – denoting the censoring time (Klein and Moeschberger, 2006).

In the case that $T_i \geq C_l$, the exact lifetime of the item is known. By introducing a pair of random variables (X, ϵ) – where ϵ indicates whether the lifetime of an item X is observed

($\epsilon = 1$) or was censored ($\epsilon = 0$) – and where T_i is equal to X if the lifetime is observed or C_l if the item was censored. That is, the time-to-failure of item i is given as $T_i = max\,(X, C_l)$. Meeker and Escobar (2014) provide a method for constructing a likelihood function under left-censored data:

$$L_{Left}\,(\boldsymbol{\theta} \mid \boldsymbol{t}) = \prod_{i=1}^{n} f\,(t_i)^{\epsilon_i}\,F\,(t_i)^{1-\epsilon_i}\,. \tag{2.15}$$
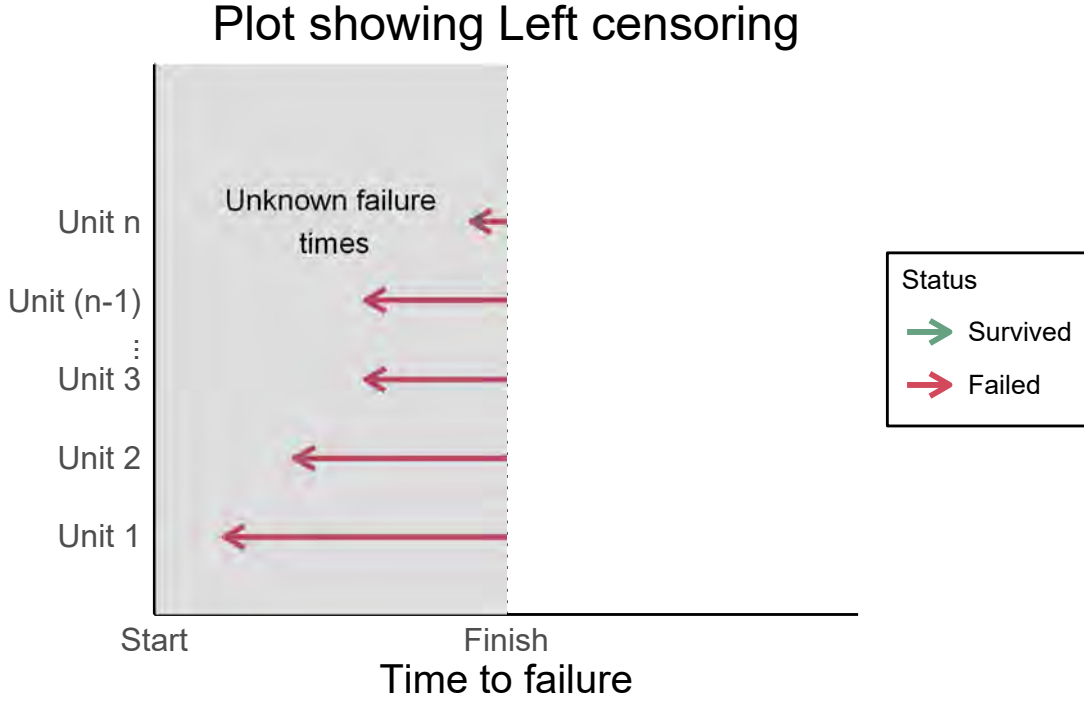
Figure 2.2: Plot illustrating left-censored data.

2.2.2.2   Right-censored data

A dataset is said to be right-censored when some items have survived the experiment and their failure times are known only to be beyond their present running times. Notationally, the exact lifetime of an item, T_i will only be known if it is less than or equal to some value C_r – denoting the censoring time (Klein and Moeschberger, 2006).

In the event that $T_i \geq C_r$ the item is said to have survived the test, and its test time is censored at C_r. By introducing a pair of random variables (X, δ) – where δ indicates whether unit X failed during the test ($\delta = 1$) or was censored ($\delta = 0$) – and T_i is equal to X if the lifetime is observed or C_r if the unit was censored. That is, the time-to-failure of item i is given as $T_i = min\,(X, C_r)$.

Meeker and Escobar (2014) provide a method for constructing a likelihood function under right-censored data:

$$L_{Right}\,(\boldsymbol{\theta} \mid \boldsymbol{t}) = \prod_{i=1}^{n} f\,(t_i)^{\delta_i}\,[1 - F\,(t_i)]^{1-\delta_i}\,. \tag{2.16}$$
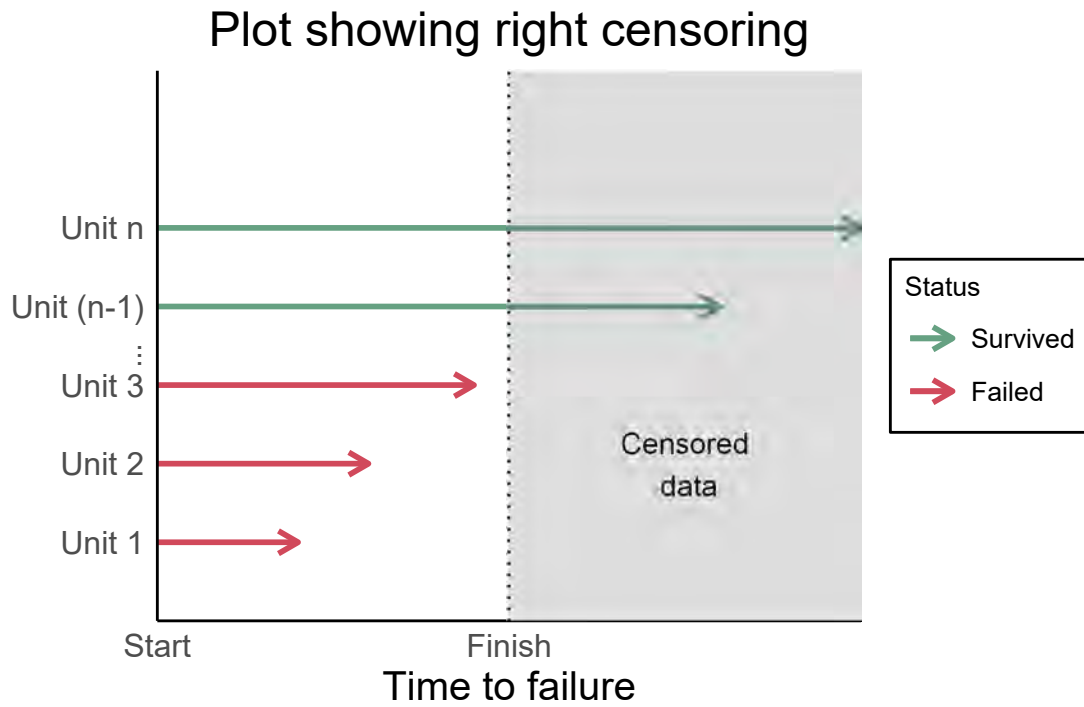
Figure 2.3: Plot illustrating right-censored data.

2.2.2.3 Interval-censored data

Interval censoring is a more general type of censoring when the failure time of an item is observed only within a certain interval, denoted by $(L_i, R_i]$, otherwise the item's failure time is censored (Klein and Moeschberger, 2006).

Meeker and Escobar (2014) provide a method for constructing a likelihood function under interval-censored data:

$$L_{Int}\left(\boldsymbol{\theta} \mid \boldsymbol{t}\right) = \prod_{i=1}^{n} f\left(t_i\right)^{\gamma_i} \left[F\left(r_i\right) - F\left(l_i\right)\right]^{1-\gamma_i}, \tag{2.17}$$

where $(\gamma_i = 1)$ if the failure time was observed within the interval $(L_i, R_i]$, and $(\gamma = 0)$ if the failure time was censored.

Plot showing Interval censoring
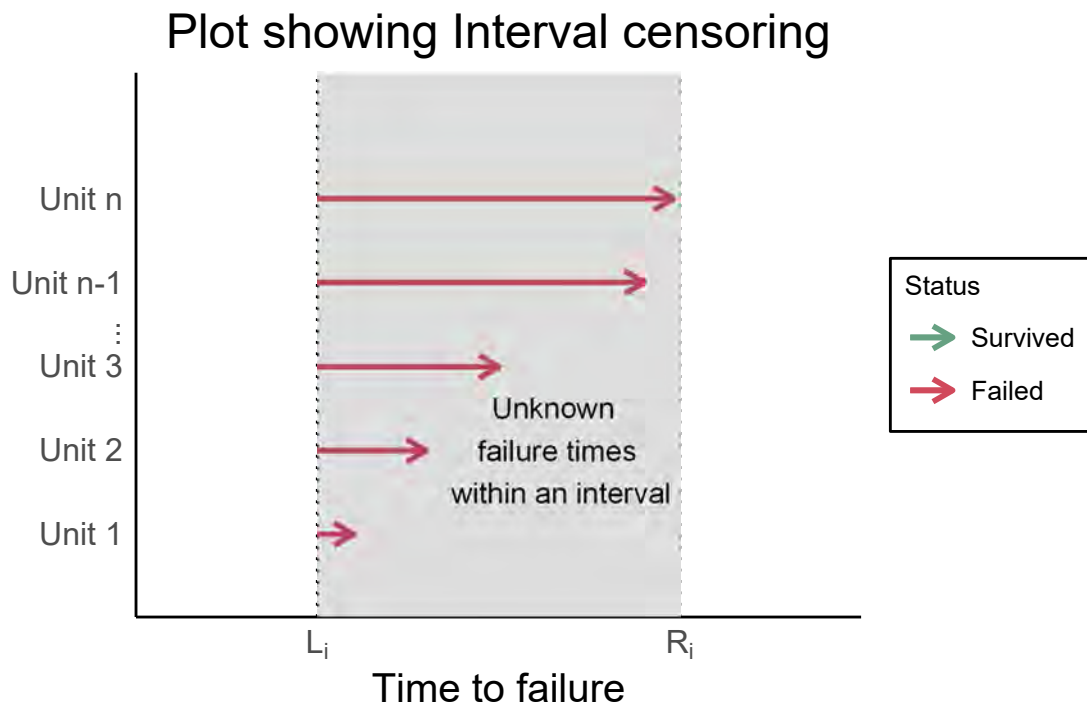
Figure 2.4: Plot illustrating interval-censored data.

2.2.2.4 Truncation

Another concept similar to censoring is interval truncation (or just truncation), which occurs only when an item's failure lies within a certain observed interval, denoted by (Y_L, Y_R) (Klein and Moeschberger, 2006). If an item's failure is not observed within this interval, no information on the item is available.

Plot showing truncation
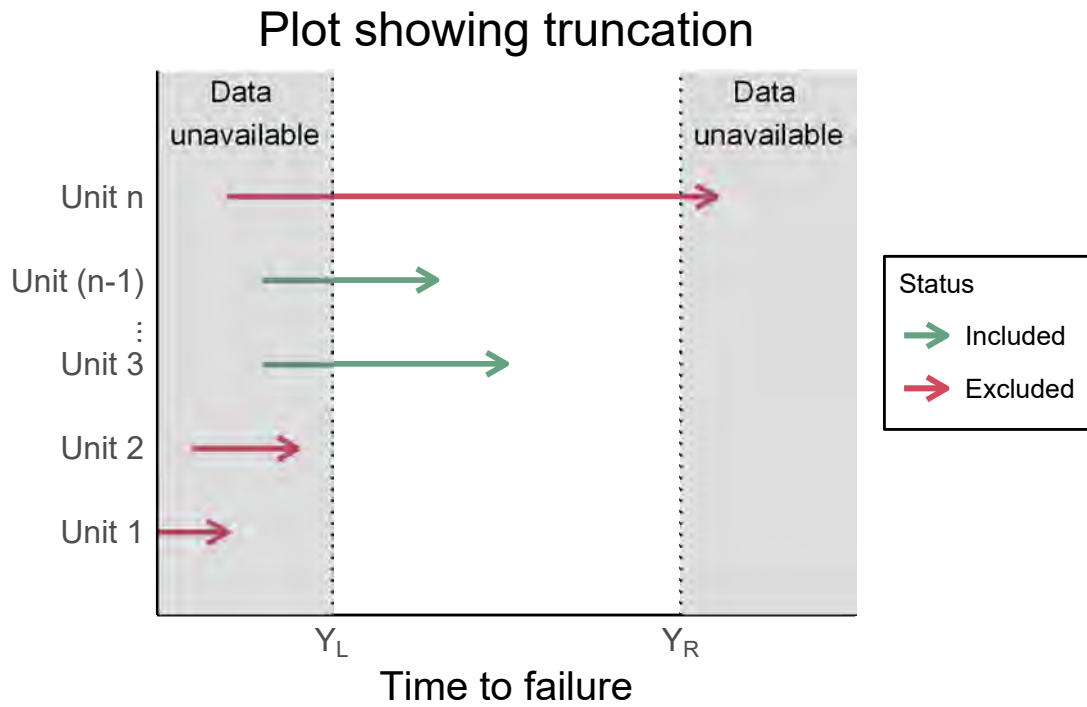
Figure 2.5: Plot illustrating truncation.

Truncation differs from censoring since in the latter partial information is available on each item. Left truncation occurs when an item enters a study at a particular time and is tested until the item has failed or has been censored, that is Y_R in the interval is infinite and we gain information on a unit when the time-to-failure is larger than Y_L.

Plot showing Left truncation
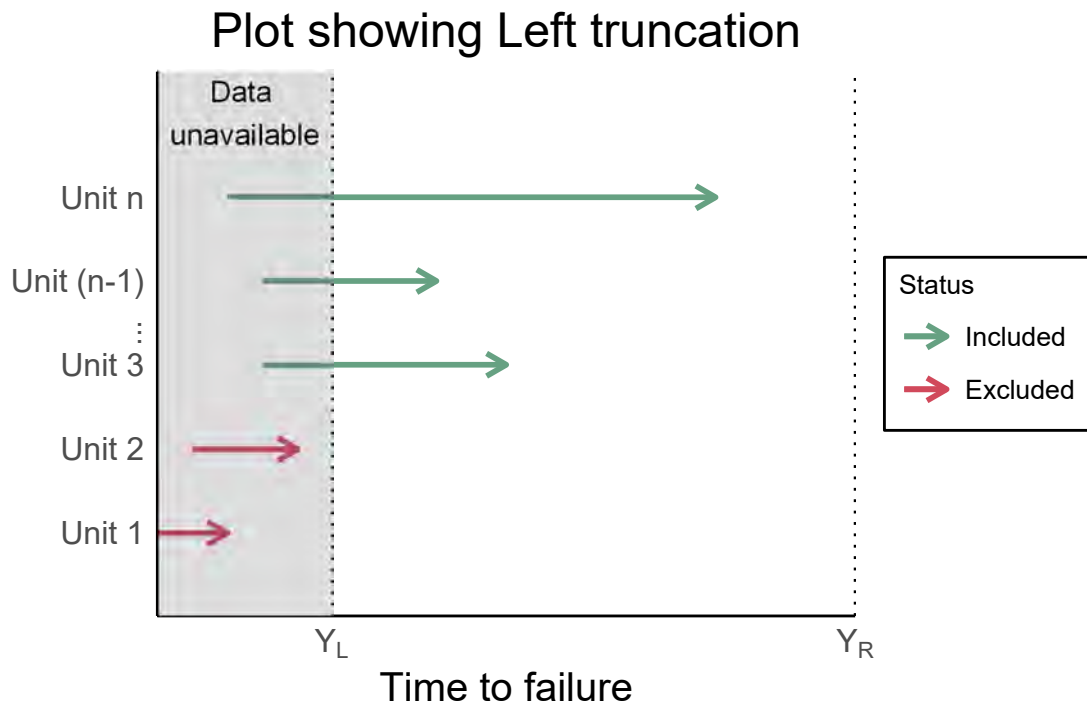
Figure 2.6: Plot illustrating left truncation.

In right truncation Y_L in the interval is 0 and we gain information on an item when the time-to-failure is less than Y_R.
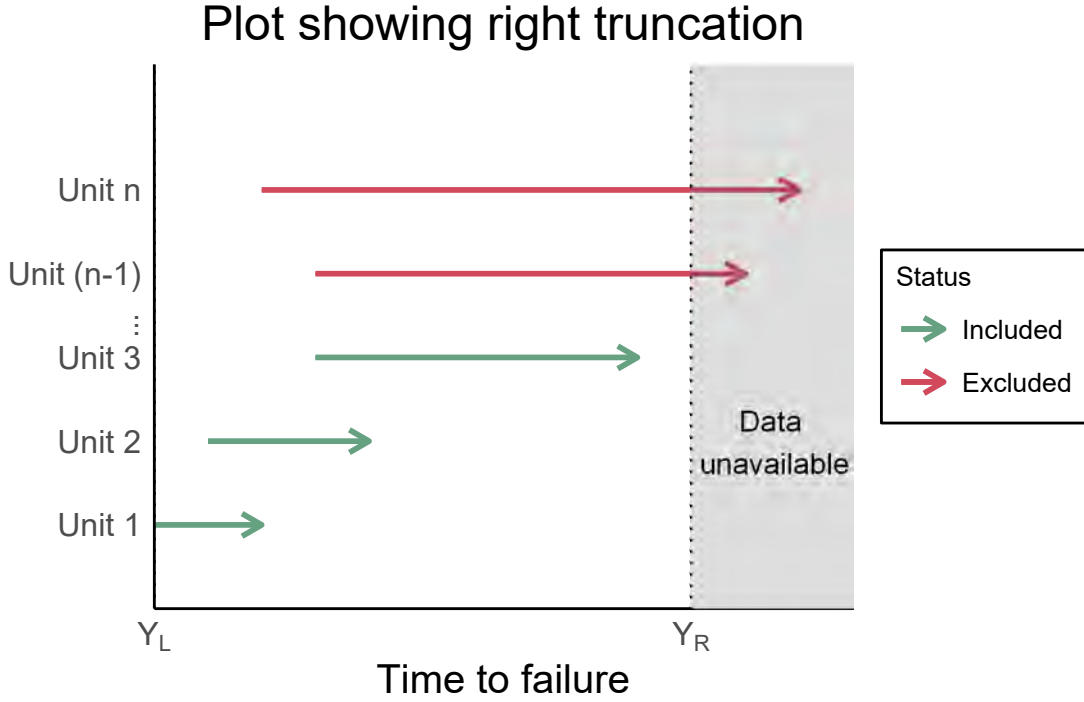
Figure 2.7: Plot illustrating right truncation.

2.2.2.5   Type I censoring

Type I censoring is a process in which all items began a test at time $t = t_0$ and are run until all items have failed or until a prespecified time – denoted by $t = t_f$ – when the test has been terminated. That is, after the test the lifetime of items failed before time t_f are known exactly. The information in this type of dataset is given by:

$$T_{(1)} \leq T_{(2)} \leq ... \leq T_{(s)}, \tag{2.18}$$

where s is the number of items that failed the test before time t_f, with $s \leq n$. Therefore it is known that $(n-s)$ items survived the test.

Rausand and Høyland (2003) provide rebuttal against type I censoring: they claim that since s is stochastic there is a chance none or few items will fail before time t_f, therefore reducing the information available on time-to-failure.

Klein and Moeschberger (2006) provide a method to constructing a likelihood under type I censoring:

$$L_I(\boldsymbol{\theta} \mid \boldsymbol{t}) = c \prod_{i=1}^{n} f(t_i)^{\delta_i} [1 - F(t_i)]^{1-\delta_i}, \tag{2.19}$$

where:

- c is an ordering constant which does not depend on the parameters $\boldsymbol{\theta}$,

- $(\delta_i = 1)$ if the lifetime of the item was observed and

- $(\delta_i = 0)$ if the lifetime of the item was not observed.
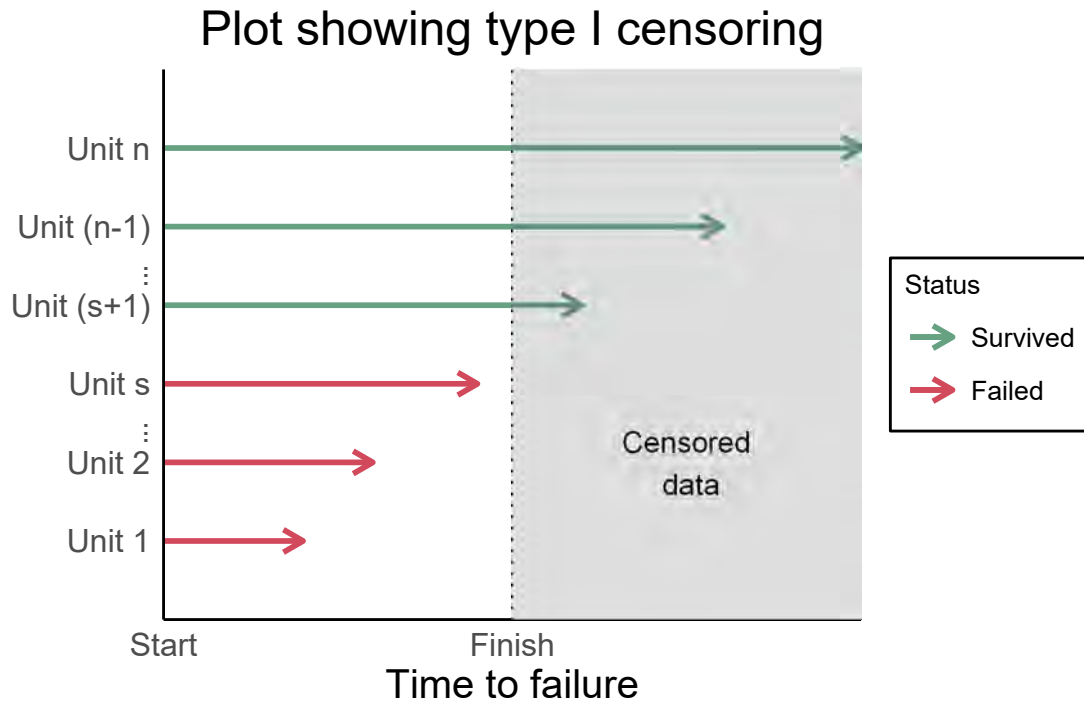
Figure 2.8: Plot illustrating type I data.

2.2.2.6   Type II censoring

In type II censoring the test is run until a prespecified number of failures are observed – denoted by r, for $0 < r < n$. As with Type I censoring, the test begins at time $t = t_0$. The information in this type of dataset is given by:

$$T_{(1)} \leq T_{(2)} \leq ... \leq T_{(r)}. \tag{2.20}$$

Therefore $(n - r)$ items are known to have survived the test.

Rausand and Høyland (2003) again provide criticism of such a test: since the time at which the rth failure occurs is stochastic, the time the test takes to finish cannot be determined.

Klein and Moeschberger (2006) provide a method to constructing a likelihood under type II censoring:

$$L_{II}\left(\boldsymbol{\theta} \mid \boldsymbol{t}\right) = \frac{n!}{(n - r)!} \prod_{i=1}^{r} \left[f\left(t_i\right)\right] \left[1 - F\left(t_i\right)\right]^{n-r}, \tag{2.21}$$

where: $\frac{n!}{(n-r)!}$ is the ordering constant.

Plot showing type II censoring
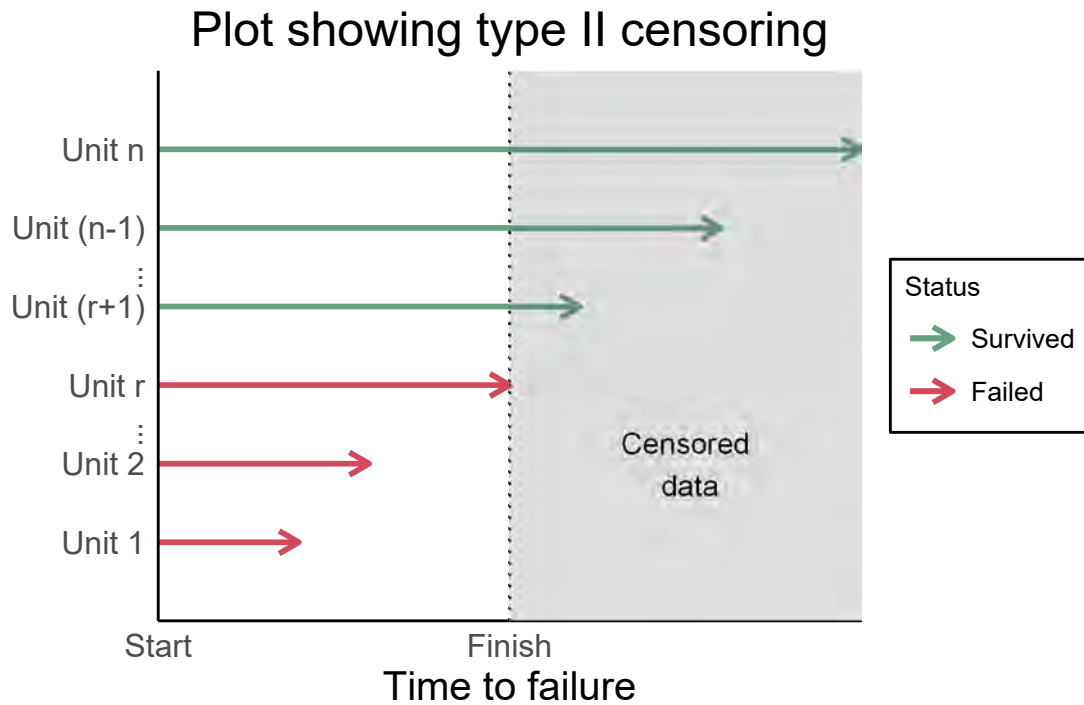
Figure 2.9: Plot illustrating type *II* data.

2.2.2.7 Other types of censoring

Rausand and Høyland (2003) introduce two other forms of censoring which haven't been discussed much in literature, namely: type *III* censoring and type *IV* censoring.

In Type *III* censoring is a combination of type *I* and type *II* censoring. The test is terminated either at time $t = t_f$ or after r failures have been observed, whichever occurs first. Therefore both t_f and r need to be specified prior to the start of the test.

In type *IV* censoring, n units begin testing at different prespecified points at time. If the time for censoring of unit i is stochastic, then the censoring is said to be of type *IV*.
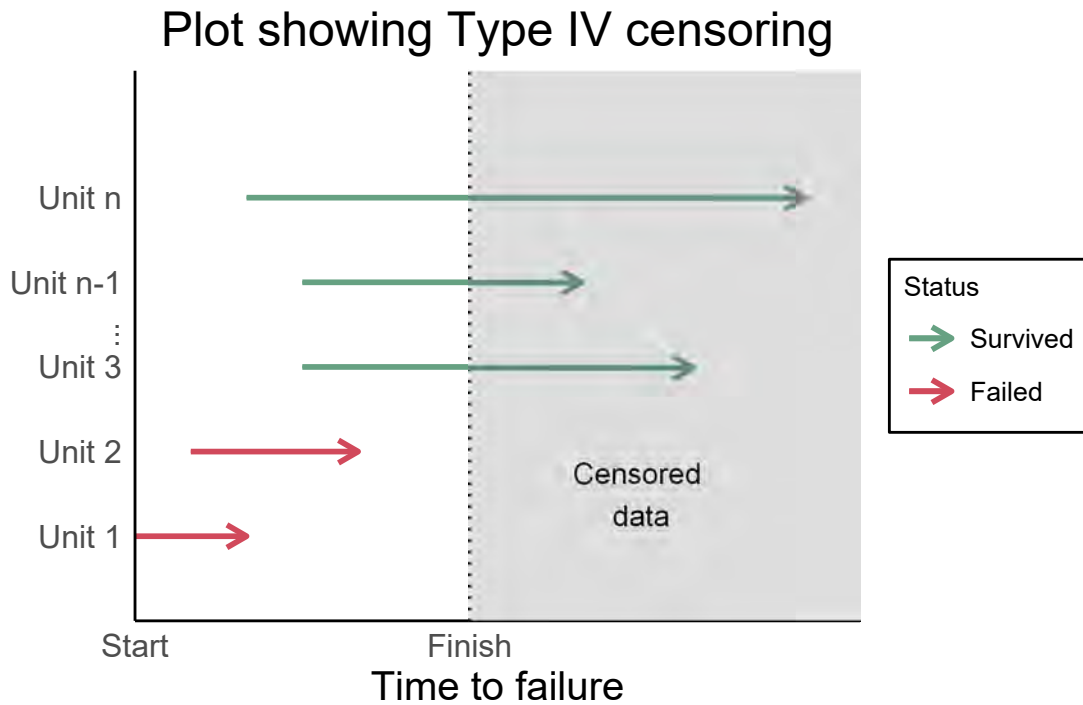
Figure 2.10: Plot illustrating type IV data.

2.3 Accelerated life testing

Items built today often have very high reliability under their normal use conditions, resulting in lengthy times-to-failure. Therefore finding reliability data on these items is impractical, because once their reliability data has been collected, they are already out of date and replaced by new-and-improved items (Rausand and Høyland, 2003). Sometimes long testing procedures are costly and hence to economize, it is desirable to speed up the testing process.

ALT is a process of exposing an item to various degrees of stress higher than under its use (nominal) stress levels in order to obtain reliability data faster. This data is then used to make inference on the item subject to the use stress level. Depending on the type of item, there are various techniques used to speed up the failure process, such as: increasing the temperature, voltage, pressure, vibration and so on. These variables are called the stressors, and are often represented by the vector, $\boldsymbol{s}' = (S_1, S_2, ..., S_p)'$ such that $S_0 < S_1 < ... < S_p$, where S_0 represents the level of stress on an item at use levels, and S_p denotes the most severe stress applied to an item (Rausand and Høyland, 2003).

This section aims at discussing the two different type of ALT: qualitative and quantitative, and the different engineering considerations required for running an ALT.

2.3.1 Qualitative versus Quantitative accelerated life testing

Qualitative accelerated life tests (QualAt) are tests used to find the underlying reasons for an item's failure under higher-than-usual stress levels (Escobar and Meeker, 2006). The design

of such a test is simple: run the item at a higher-than-use stress level of a given variable, and if the item survives, it passes the test (Nelson, 1990). Otherwise if the item fails, a probable failure mode is revealed and appropriate action to redesign the item is required. These tests are referred to by many names, including:

- HALT (Highly accelerated life testing),

- elephant tests,

- STRIFE (Stress life),

- environmental stress testing and

- shake and bake tests.

On the other hand, quantitative accelerated life tests (QuanAT), unlike qualitative life tests, are designed to obtain quantitative information on the failure rate of the data through a life distribution (usually: exponential, Weibull or lognormal) at the use level of a unit and thereby provide useful reliability data (Escobar and Meeker, 2006). In these types of tests probable failure modes, some knowledge that describes the relationship between failure mechanisms and the accelerating variables (*Vida infra*) are known *a priori.*

2.3.2   Time transformation functions

A time transformation function is one which describes the relationship between the lifetime of an item and the stress level considered, and is expressed as a distribution's life characteristic. Depending on the life distribution assumed, different life characteristics are considered, for example: for the Weibull – the life characteristic is the scale parameter, for the exponential – the life characteristic is the mean life and for the lognormal distribution the life characteristic is the median.

Assume that the life characteristic of a given distribution – denoted by ν – is related to a vector of specified stress-functions given as $\boldsymbol{\mu} = (\mu_0(S_i), \mu_1(S_i), ..., \mu_{p-1}(S_i))'$ where the initial stress function $\mu_0 \equiv 1$. Then the general form of the time transformation function is given by:

$$\nu(S_i) = exp\{\boldsymbol{\theta}'\boldsymbol{\mu}\}, \tag{2.22}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)'$ is a vector of unknown parameters which need to be estimated.

The choice of $\boldsymbol{\mu}$ is often determined by the underlying physics of failure, and an ill-chosen time transformation can often lead to inadmissible results (Singpurwalla, 2006).

A more popular way of expressing the life-time stress function is by a log-linear relationship, given by:

$$log(\nu(S_i)) = \boldsymbol{\theta}'\boldsymbol{\mu}. \tag{2.23}$$

In this literature review, three of the most popular models will be discussed, namely:

- the Arrhenius,

- the Eyring and

- the inverse power law model.

2.3.2.1 Acceleration factor

The acceleration factor (AF) is the ratio of nominal life between the item's use stress level, ν_{use} and a higher stress level, ν_a, notationally given by:

$$AF = \frac{\nu_{use}}{\nu_a}, \tag{2.24}$$

(Nelson, 1990). This is analogous to saying that the rate of failure at the higher stress level is AF times faster than the rate of failure at the use stress level.

2.3.2.2 Linearity assumption

Nelson (1990) claims that testing for the linearity of the life-stress relationship is vital in performing an ALT. Non-linearity of the relationship may be caused by:

- the life-test not being run properly (possibly malfunctioning test equipment),

- several failure methods acting on the item at once or

- by choosing a life-stress relationship that is inherently nonlinear.

A subjective test for linearity is plotting the level of stress against the observed values on log-log paper, and then drawing a line through these points (Nelson, 1990). If the line is straight then the assumption of linearity holds. Otherwise if data from a given stress level are out of line from other stress levels, it might imply that the data in that stress level are in error. The reason for error should be examined to determine whether those data are valid and should thus be included in the study.

2.3.2.3 Arrhenius model

The Arrhenius model is based on the law derived by Nobel laureate Arrhenius (1889), and is used to describe the failure of an item due to chemical reactions caused by changes in temperature.

Nelson (1990) provides a form of the Arrhenius life-stress relationship, and it is given by:

$$\nu\left(S_i\right) = A \times exp\left\{-\frac{E_{act}}{kS_i}\right\}, \tag{2.25}$$

where:

- A is a constant that is characteristic of the product failure mechanism and test conditions,

- S_i is the stress applied given in absolute Kelvin temperature (K),

- k is Boltzmann's constant $= 8.6171{\times}10^{-5}$ electron-volts per Kelvin and

- E_{act} is the activation energy of the reaction measured in electron-volts.

To counter the naive assumption that time-to-failure is inversely proportional to the failure rate, Nelson (1990) claims that the following model is sometimes used:

$$\nu\left(S_i\right) = A \times exp\left\{\frac{E_{act}}{kS_i}\right\}. \tag{2.26}$$

Taking the natural logarithm of equation 2.26 yields:

$$log\left(\nu\left(S_i\right)\right) = \theta_1 + \theta_2 \times \mu\left(S_i\right), \tag{2.27}$$

where:

- $\theta_1 = log\left(A\right)$,

- $\theta_2 = \frac{E_{act}}{k}$ and

- $\mu\left(S_i\right) = \frac{1}{S_i}$.

The AF of the model presented in equation 2.26 is thus given by:

$$AF_A = exp\left\{\frac{E_{act}}{k}\left(\frac{1}{S_{use}} - \frac{1}{S_a}\right)\right\}, \tag{2.28}$$

where S_{use} is the temperature applied at the use stress level, and S_a is the temperature applied at the higher stress level.

2.3.2.4   Eyring model

Similar to the Arrhenius model, the Eyring model developed by Glasstone et al. (1941) is used to describe the failure of an item due to chemical reactions caused by not only changes in temperature, but by other factors such as humidity.

Nelson (1990) provides a form of the Eyring life-stress relationship, and it is given by:

$$\nu\left(S_i\right) = \frac{A}{S_i} \times exp\left\{\frac{B}{kS_i}\right\}, \tag{2.29}$$

where:

- A and B are constants that are characteristic of the product failure mechanism and test conditions,

- S_i is the stress applied given in absolute Kelvin temperature (K) and

- k is Boltzmann's constant $= 8.6171 \times 10^{-5}$ electron-volts per Kelvin.

Nelson (1990) claims that for small ranges of absolute temperature, the ratio $\frac{\theta_1}{S_i}$ is essentially constant and equation 2.29 is approximately the Arrhenius model.

The AF is thus given by:

$$AF_E = exp\left\{\frac{B}{k}\left(\frac{1}{S_{use}} - \frac{1}{S_a}\right)\right\}, \tag{2.30}$$

where S_{use} is the temperature applied at the use stress level, and S_a is the temperature applied at the higher stress level.

2.3.2.5 Inverse power law model

The inverse power law model, or simply the power law, is used to describe failure of an item as a result of degradation due to changes in non-thermal stress, for example voltage.

Nelson (1990) provides a form of the inverse power law relationship, and it is given by:

$$\nu\left(S_i\right) = \frac{A}{S_i^B}, \tag{2.31}$$

where:

- A and B are constants that are characteristic of the product failure mechanism and test conditions and

- S_i is the ith stress level.

Taking the natural logarithm of equation 2.31 yields:

$$log\left(\nu\left(S_i\right)\right) = \theta_1 + \theta_2 \times \mu\left(S_i\right), \tag{2.32}$$

where:

- $\theta_1 = log\left(A\right)$,

- $\theta_2 = B$ and

- $\mu\left(S_i\right) = -log\left(S_i\right)$.

The AF of the inverse power law given in equation 2.31 is given by:

$$AF_I = \left(\frac{S_a}{S_u}\right)^B, \tag{2.33}$$

where S_u is the use stress level and S_a is the higher stress level.

2.3.3   Accelerated life testing designs

Before an ALT is run, the method of loading stress on to the item needs to be considered. This section aims at discussing the various ways to design stress loading for an ALT, assuming that only one stressor is applied to the tested item.

2.3.3.1   Constant stress

In this method an item is run at a constant stress level throughout the test, and is perhaps the most common way of loading stress. Nelson (1990) claims that the advantages of using such a model is for its simplicity and because it mimics use stress levels accurately.

Rausand and Høyland (2003) provide a method to running a constant stress test subject to no censoring:

1. Choose a random stress level denoted by S_i for $i = 1, 2, ..., p$ and choose a random sample of test items of the same type among all available test items.

2. Run the random sample of test items at stress level S_i until all the items have failed.

3. Choose another stress level S_j for $j = 1, 2, ..., p$ and repeat steps 1. and 2. until all p stress levels have been considered.
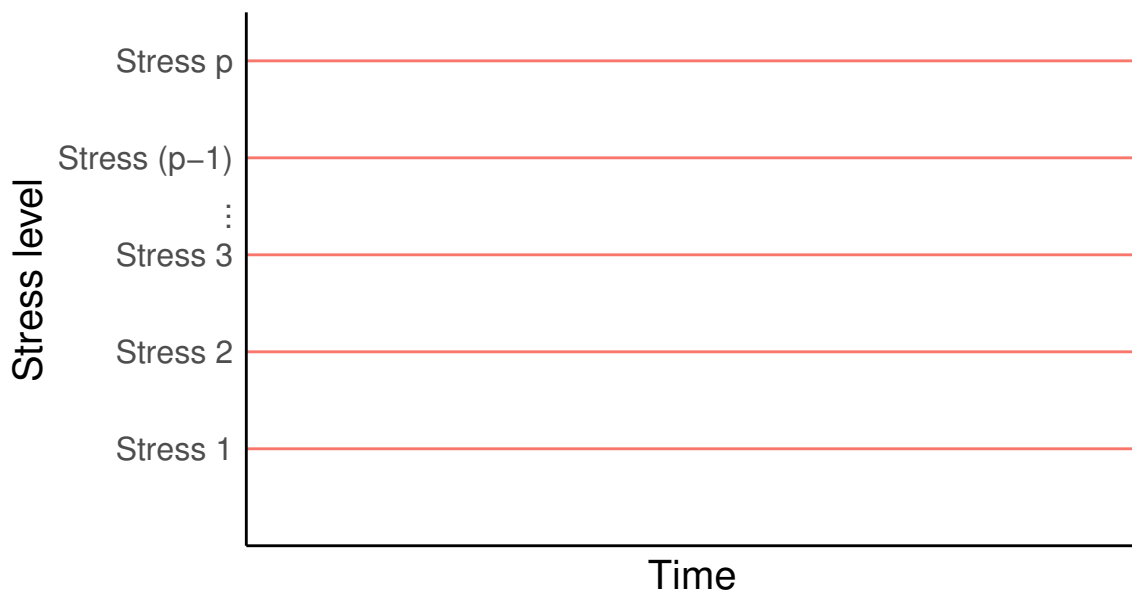
Figure 2.11: Plot illustrating items subject to constant stress.

2.3.3.2   Step-Stress

Step-stress accelerated tests (SSALT) is a process in which an item is exposed to successively higher levels of stress. This method is used because it yields failure rates faster than under

the constant stress method (Nelson, 1990). However in practical use most items are exposed to constant stress, and the SSALT models are more complicated in obtaining reliability data. Rausand and Høyland (2003) provide a method to running a SSALT subject to no censoring:

1. Fix p points in time such that $0 < t_1 < t_2 < ... < t_p < t$ where t is the length of the test and randomly choose n items of the same type to be tested starting at time $t = 0$.

2. In the interval $(0, t_1]$ run the items on the constant stress level S_1, and at the time t_1 remove all items which have failed.

3. In the interval $(t_1, t_2]$ run the items that have not failed on the constant stress level S_2, again removing all the failed items from the test.

4. Repeat the test on the interval $(t_{k-1}, t_k]$ at constant stress level S_k for $k = 2, 3, ..., p$.

5. In the time interval $(t_p, \infty]$, the items which have not yet failed are run at constant stress S_{k+1} until they eventually fail.
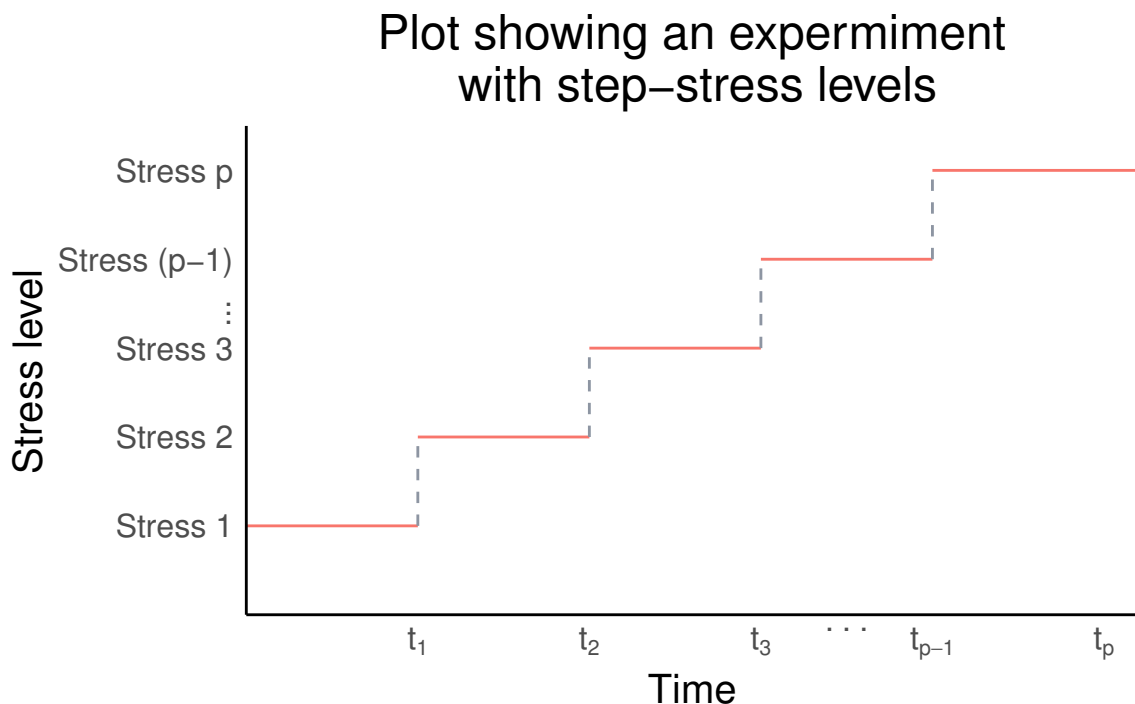
Figure 2.12: Plot illustrating items subject to step-stress.

2.3.3.3   Varying-stress levels

There are three methods which apply varying stress levels on items: progressive-stress accelerated tests (PALT), cyclic stress and random stress testing. PALT is a process in which the items are exposed to stress level $S(t)$ – some known increasing function of time, until the test is terminated. In cyclic stress testing the stress level changes cyclically, for example following a sinusoidal pattern, and in random stress testing the items are exposed to a stress

level changing randomly with time. However Nelson (1990) claims that it is difficult to control the varying stress accurately enough to obtain adequate reliability data, and hence these methods are not often used.
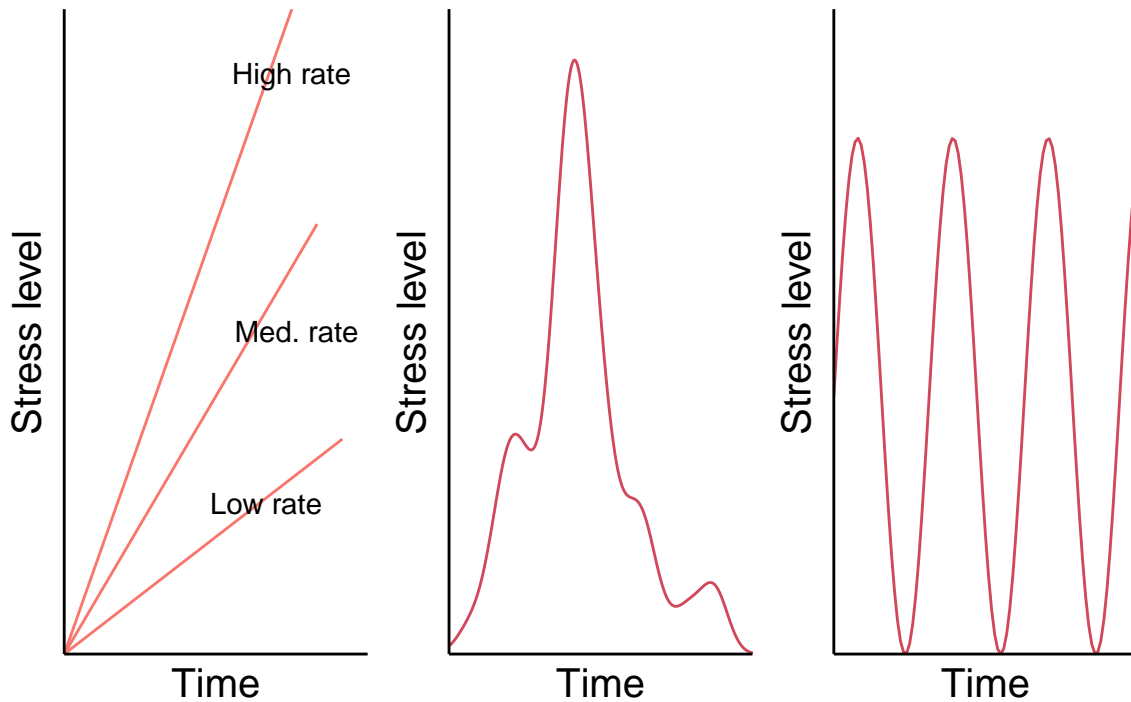
Figure 2.13: (From left to right) Plot illustrating: PALT, random and cyclic stress.

2.4 Parameter estimation

Parameter estimation is perhaps the most important part of any statistical analysis, and involves finding numeric estimates via some procedure given sample data. In this thesis two parameter estimation methods will be discussed, namely:

- the method of maximum likelihood and

- Bayesian methods.

2.4.1 Method of maximum likelihood

Fisher (1922) was the first to formally introduce the method of maximum likelihood as a means of parameter estimation, and has since been one of the most significant statistical developments. Consider the joint density function of n independent and identically distributed observations $\boldsymbol{t} = (t_1, t_2, ..., t_n)'$ with parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)'$, denoted by $f(\boldsymbol{t} \mid \boldsymbol{\theta})$. Then the likelihood function is given by:

$$L(\boldsymbol{\theta} \mid \boldsymbol{t}) = \prod_{i=1}^{n} f(t_i \mid \boldsymbol{\theta}). \tag{2.34}$$

Often it is preferable to work with the log of the likelihood function, denoted by:

$$log\left(L\left(\boldsymbol{\theta}\mid\boldsymbol{t}\right)\right) = \mathcal{L}\left(\boldsymbol{\theta}\mid\boldsymbol{t}\right). \tag{2.35}$$

With respect to its name, the MLEs are the parameter values which maximize equation 2.35, mathematically:

$$\hat{\boldsymbol{\theta}} = \underset{\theta}{argmax}\left(\mathcal{L}\left(\boldsymbol{\theta}\mid\boldsymbol{t}\right)\right).$$

Often finding the values of $\hat{\boldsymbol{\theta}}$ is difficult since it requires solving a system of non-linear equations. The most common method to solve these equations is via the iterative-gradient method, often referred to by its synecdoche, the NR algorithm. The steps used to compile the NR algorithm are given by Klein and Moeschberger (2006):

1. Choose initial values for $\hat{\boldsymbol{\theta}}$ given by $\hat{\boldsymbol{\theta}}_0$.

2. Find the gradient vector of the log-likelihood, given by:

$$\nabla'\left(\boldsymbol{\theta}\right) = \left.\frac{\partial\mathcal{L}\left(\boldsymbol{\theta}\mid\boldsymbol{t}\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{2.36}$$

   and the Hessian matrix of the log-likelihood:

$$H\left(\boldsymbol{\theta}\right) = \left.\frac{\partial^2\mathcal{L}\left(\boldsymbol{\theta}\mid\boldsymbol{t}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{2.37}$$

3. Compute:
$$\hat{\boldsymbol{\theta}}_{p+1} = \hat{\boldsymbol{\theta}}_p - s_p \times H^{-1}\left(\boldsymbol{\theta}\right)\nabla\left(\boldsymbol{\theta}\right),\, p = 0, 1, 2, \dots \tag{2.38}$$

   until $\left|\hat{\boldsymbol{\theta}}_{p+1} - \hat{\boldsymbol{\theta}}_p\right| < \epsilon$, where:

   - s_p is an arbitrary constant, usually unity $\underset{p}{\forall}$ and

   - ϵ is some small, prespecified constant.

Gelman et al. (2013) claim that the choice of starting values $\hat{\boldsymbol{\theta}}_0$ is crucial since the algorithm is not guaranteed to converge for all starting values. He suggests that potential starting values could be: crude parameter estimates or values generated from conditional maximization.

Other gradient methods described by Rinne (2008) include: the method of steepest ascent where the Hessian matrix described in equation 2.37 is replaced with the identity matrix, or the method of scoring where the Hessian matrix is replaced with the observed Fisher information matrix (*Vide infra*) and $s_p = -1 \underset{p}{\forall}$. Despite the fact that the method of steepest ascent is by far the easiest to compute, it often cannot be recommended since it may converge slowly if the maximum is defined on a long and narrow ridge.

2.4.1.1   Properties of maximum likelihood estimators

Rinne (2008) provides some important asymptotic properties of MLEs, namely they are:

- Consistent, that is as $n \to \infty$ the estimator converges to its true value with probability unity.

- Have normality, that is as $n \to \infty$ the estimator approaches a normal distribution with properties:
$$\hat{\boldsymbol{\theta}} \sim Normal_p\left(\boldsymbol{\theta}, \Sigma\right), \tag{2.39}$$
  where $\Sigma = I^{-1}\left(\boldsymbol{\theta}\right)$, which, provided that it exists, is the inverse of the Fisher information matrix, given as:
$$I\left(\boldsymbol{\theta}\right) = -E_t\left(\frac{\partial^2 log\left(L\left(\boldsymbol{t} \mid \boldsymbol{\theta}\right)\right)}{\partial \theta_i \partial \theta_j}\right). \tag{2.40}$$

- Efficient, that is it achieves equality in the Cramér-Rao lower bound, which for an unbiased estimator is:
$$Var\left(\hat{\boldsymbol{\theta}}\right) \geq \frac{1}{nI\left(\boldsymbol{\theta}\right)}. \tag{2.41}$$

- Functionally invariant to transformations.

2.4.2   Bayesian Statistics

The Bayesian paradigm was established in the posthumous work of Bayes (1763), who provided a theory based on personalistic beliefs in the context of uncertainty. This belief comes in the form of prior knowledge of the unknown parameters postulated before data from an experiment is available. When data is available, it is combined with the prior knowledge resulting in updated beliefs on the unknown parameters. This rationalist theory has since resulted in the nomenclature of *prior* and *posterior distributions* commonly used by Bayesians today.

The essential element of the Bayesian approach is *Bayes' theorem*, sometimes referred to in literature as the *principle of inverse probability*. Let $f\left(\boldsymbol{t} \mid \boldsymbol{\theta}\right)$ denote the joint PDF for a continuous random vector of n observations $\boldsymbol{t}$ and p unknown parameters $\boldsymbol{\theta}$. The likelihood function of this PDF is then given by $L\left(\boldsymbol{\theta} \mid \boldsymbol{t}\right)$, which represents information carried by the observations. Let $\pi\left(\boldsymbol{\theta}\right)$ be the prior distribution describing the uncertainties of the unknown parameters. Then Bayes' theorem states:

$$\pi\left(\boldsymbol{\theta} \mid \boldsymbol{t}\right) = \frac{L\left(\boldsymbol{\theta} \mid \boldsymbol{t}\right) \times \pi\left(\boldsymbol{\theta}\right)}{\int L\left(\boldsymbol{\theta} \mid \boldsymbol{t}\right) \times \pi\left(\boldsymbol{\theta}\right) d\boldsymbol{\theta}}, \tag{2.42}$$

where $\pi\left(\boldsymbol{\theta} \mid \boldsymbol{t}\right)$ is called the posterior distribution. Since the denominator of equation 2.42 does not depend on $\boldsymbol{\theta}$, the integral may treated as a constant. Therefore Bayes' theorem may

be reduced to the following relation:

$$\pi\left(\boldsymbol{\theta}\mid\boldsymbol{t}\right)\propto L\left(\boldsymbol{\theta}\mid\boldsymbol{t}\right)\times\pi\left(\boldsymbol{\theta}\right). \tag{2.43}$$

That is, Bayes' theorem states that the probability distribution of the unknown parameters $\boldsymbol{\theta}$ conditioned on data $\boldsymbol{t}$ is merely proportional to the product of the prior distribution of $\boldsymbol{\theta}$ and the likelihood of $\boldsymbol{\theta}$ conditioned on $\boldsymbol{t}$, simply put:

$$\pi\left(\boldsymbol{\theta}\mid\boldsymbol{t}\right)\propto Likelihood\,function\,\times\,prior\,density. \tag{2.44}$$

Defining the prior distribution is perhaps the most important part of any Bayesian analysis. Press (2009) describes three types of prior distributions, namely:

- informative,

- conjugate and

- non-informative priors.

2.4.3 Informative priors

An informative prior (or subjective prior) is one which is based on either sample evidence from previous experiments, or from mere expert reasoning (Jeffreys and Zellner, 1989). These types of priors are necessary as they force experimenters to explicitly state their beliefs and prejudices, encouraging others to form their own opinion on the data (Kadane, 1995).

There is however a fair amount of criticism against these priors. Garthwaite et al. (2005) argue that even expert-judgments are subject to severe error and systematic bias, which will resultingly impact the outcome of an experiment. Press (2009) adds that it is sometimes difficult to express an informative prior as a mathematical expression. Furthermore subjective views may differ greatly between experimenters, making it difficult to compare results on the same experiment.

2.4.4 Conjugate priors

A popular informative prior is the conjugate prior introduced by Schlaifer and Raiffa (1961). By definition, a family of distributions is said to be a conjugate family, if for a given likelihood and prior distribution in the family, the posterior distribution also belongs to the family. That is, a conjugate prior is one which produces a posterior distribution of a known form and hence are chosen due to their mathematical convenience.

Press (2009) provides a method to form a conjugate prior:

1. Form the likelihood function for the underlying statistical distribution.

2. Change the roles of the random variables and parameters in the likelihood function.

3. "Enrich" the parameters of the resulting density kernel of the distribution such that the parameters do not depend on the sample data.

4. Identify the distribution corresponding to the resulting kernel, and adjoin the appropriate normalizing constant to ensure the density integrates to unity.

2.4.5   Non-informative priors

Sometimes before an experiment is run no information is available on the unknown parameters, and hence an informative prior cannot be used. In these cases non-informative priors (often known by other names such as: objective, default or reference) can be used, which are defined as priors which provides little or no information about any of the unknown parameters (Meeker and Escobar, 2014). These priors have been used throughout the history of the Bayesian theory: Bayes (1763) himself used a uniform prior for the Bernoulli distribution, and Laplace (1820) used a uniform prior for the mean of the Gaussian distribution.

The primary purpose of non-informative priors is that their PDFs are constant (or approximately constant) over the range of the model parameters such that the data dominates the effect of the prior, and they are chosen by convention or through structural rules, such that they can be viewed as a standard of reference (Singpurwalla, 2006).

The difficulty with using non-informative priors is that they often produce improper PDFs, that is, they do not integrate to a finite quantity and result in improper probability theory (Meeker and Escobar, 2014). However using improper prior PDFs suffices provided that the corresponding posterior PDF is proper. Press (2009) continues with the rebuttal claiming that non-informative priors follow the naive assumption that parameters are independent, which is rare in a real-life scenario.

Five classes of non-informative priors will be discussed in this thesis, namely:

- Jeffreys' prior,

- reference priors,

- maximal data information prior,

- uniform prior and

- probability matching priors.

2.4.5.1   Jeffreys' prior

Jeffreys (1961) provided a rule to form a prior distribution, which is perhaps the most widely

used non-informative prior by Bayesian statisticians. The prior is denoted by:

$$\pi_J\left(\boldsymbol{\theta}\right) \propto \left[det\left(I\left(\boldsymbol{\theta}\right)\right)\right]^{\frac{1}{2}}, \tag{2.45}$$

in the case where the prior is developed for a density function with p unknown variables $\boldsymbol{\theta}$ and $I\left(\boldsymbol{\theta}\right)$ is the Fisher information matrix associated with the likelihood function for the data, as described in equation 2.40.

The prior is well used due to its invariance property under monotonic transformations, and when there is little information known about the unknown parameters – and is thus considered a weak prior.

Although Jeffreys' prior is successful in one-parameter models, it often experiences difficulties in the presence of nuisance parameters (Berger and Bernardo, 1992).

2.4.5.2   Reference prior

The reference prior, initially proposed by Bernardo (1979) and later revised by Berger and Bernardo (1992), was developed to derive a non-informative prior in the presence of nuisance parameters. A reference prior, denoted by π_R, is a prior which maximizes the expected posterior about the parameters provided by independent replications of an experiment, given by:

$$\Phi\left(\boldsymbol{\theta}\right) = E_t\left(K\left(\pi\left(\boldsymbol{\theta}\mid\boldsymbol{t}\right), \pi\left(\boldsymbol{\theta}\right)\right)\right), \tag{2.46}$$

where

$$K\left(\pi\left(\boldsymbol{\theta}\mid\boldsymbol{t}\right), \pi\left(\boldsymbol{\theta}\right)\right) = \int \pi\left(\boldsymbol{\theta}\mid\boldsymbol{t}\right) log\left(\frac{\pi\left(\boldsymbol{\theta}\mid\boldsymbol{t}\right)}{\pi\left(\boldsymbol{\theta}\right)}\right) d\boldsymbol{\theta} \tag{2.47}$$

is the Kullback-Leibler distance, which measures how much information is lost when a theoretical distribution is used to approximate its true distribution.

The posterior distribution under the reference prior has some interesting properties described by Bernardo and Smith (1994), notably:

- invariance under one-to-one transformations,

- consistent sampling properties and

- consistent marginalization.

When one parameter is considered and the posterior distribution under the reference prior is asymptotically normal, the reference prior is identical to Jeffreys' prior (Bernardo, 1979). However in multivariate problems the reference prior can be different to Jeffreys' prior since it depends on the ordering and grouping of the parameters according to their inferential interest.

By dividing the vector of parameters $\boldsymbol{\theta}$ into parameters of interest and nuisance parameters, the reference prior amends the deficiencies of Jeffreys' prior by removing any ad hoc modifications needed in the multiparameter case. Jeffreys' prior is *de facto* the reference prior

that ensues when all the parameters are grouped together, but should only be considered if all the parameters are of equal interest.

Berger and Bernardo (1992) provide a general algorithm for developing a reference prior in the multiparameter case, and is given as follows:

1. For a vector of p unknown parameters $\boldsymbol{\theta}$, derive the Fisher information matrix for the underlying probability model, as defined in equation 2.40, and, provided that it exists, find its inverse, denoted by $I^{-1}(\boldsymbol{\theta})$.

2. Define: I_j^{-1} as the upper left $(N_j \times N_j)$ corner of $I^{-1}(\boldsymbol{\theta})$ and $H_j \equiv I_j$. Then the matrices h_j are the lower right $(n_j \times n_j)$ of $H_j, j = 1, 2, ..., p$.

3. Separate the parameters in $\boldsymbol{\theta}$ into m groups of size n_i, denoted by $\boldsymbol{\theta}_{(m)}$ for $i = 1, 2, ..., m$, and define:
$$\boldsymbol{\theta}_{[j]} = \left(\boldsymbol{\theta}_{(1)}, \boldsymbol{\theta}_{(2)}, ..., \boldsymbol{\theta}_{(j)} \right)$$
$$\boldsymbol{\theta}_{[\sim j]} = \left(\boldsymbol{\theta}_{(j+1)}, \boldsymbol{\theta}_{(j+2)}, ..., \boldsymbol{\theta}_{(p)} \right).$$

4. Choose a nested sequence $\left\{ \Omega^l \right\}$ of compact sets of Ω, the parameter space of $\boldsymbol{\theta}$, such that $\bigcup\limits_{l=1}^{\infty} \Omega^l = \Omega$. If $\Omega^l \subset \Omega$, define:
$$\Omega^l \left(\theta_{[j]} \right) = \left\{ \boldsymbol{\theta}_{[j+1]} : \left(\boldsymbol{\theta}_{[j]}, \boldsymbol{\theta}_{[j+1]}, \boldsymbol{\theta}_{[\sim j+1]} \right) \epsilon \Omega^\star \exists \boldsymbol{\theta}_{[\sim j+1]} \right\}.$$

5. Define:
$$\pi_p^l \left(\boldsymbol{\theta}_{[\sim(m-1)]} \mid \boldsymbol{\theta}_{[m-1]} \right) = \pi_m^l \left(\boldsymbol{\theta}_{[m]} \mid \boldsymbol{\theta}_{[m-1]} \right)$$
$$= \frac{det \left(h_m(\boldsymbol{\theta}) \right)^{\frac{1}{2}} I_{\Omega'\left(\boldsymbol{\theta}_{[m-1]} \right)} \left(\boldsymbol{\theta}_{(m)} \right)}{\int\limits_{\Omega'\left(\boldsymbol{\theta}_{[m-1]} \right)} det \left(h_m(\boldsymbol{\theta}) \right)^{\frac{1}{2}} d\boldsymbol{\theta}_{[m]}}.$$

6. For $j = m - 1, m - 2, ..., 1$ define:
$$\pi_j^l \left(\boldsymbol{\theta}_{[\sim(j-1)]} \mid \boldsymbol{\theta}_{[j-1]} \right) = \frac{\pi_{j+1}^l \left(\boldsymbol{\theta}_{[\sim(j)]} \mid \boldsymbol{\theta}_{[j]} \right) exp \left\{ \frac{1}{2} E_j^l \left[\left(log \left(h_j(\boldsymbol{\theta}) \right) \mid \boldsymbol{\theta}_{[j]} \right) \right] \right\} I_{\Omega'\left(\boldsymbol{\theta}_{[j-1]} \right)} \left(\boldsymbol{\theta}_{(j)} \right)}{\int\limits_{\Omega'\left(\boldsymbol{\theta}_{[j-1]} \right)} exp \left\{ \frac{1}{2} E_j^l \left[\left(log \left(h_j(\boldsymbol{\theta}) \right) \mid \boldsymbol{\theta}_{[j]} \right) \right] \right\} d\boldsymbol{\theta}_{[j]}},$$

where:
$$E_j^l \left[\left(log \left(h_j(\boldsymbol{\theta}) \right) \mid \boldsymbol{\theta}_{[j]} \right) \right] = \int\limits_{\left\{ \boldsymbol{\theta}_{[\sim j]} : \left(\boldsymbol{\theta}_{[j]}, \boldsymbol{\theta}_{[\sim j]} \epsilon \Omega^l \right) \right\}} log \left(h_j(\boldsymbol{\theta}) \right) \pi_{j+1}^l \left(\boldsymbol{\theta}_{[\sim j]} \mid \boldsymbol{\theta}_{[j]} \right) d\boldsymbol{\theta}_{[\sim j]}.$$

7. Thus, assuming the limit exists, the reference prior is defined as:
$$\pi_R(\boldsymbol{\theta}) = \lim_{l \to \infty} \frac{\pi_1^l(\boldsymbol{\theta})}{\pi_1^l(\boldsymbol{\theta}^\star)}, \tag{2.48}$$

where $\boldsymbol{\theta}^{\star}$ is any fixed point in Ω with positive density for all π_1^l.

2.4.5.3 Maximal data information prior

Zellner (1977) introduced the MDI prior as one which emphasizes the information of the unknown parameters from the data in comparison to the prior density. The MDI prior does so by maximizing the average information from the likelihood function of the data density in relation to information known *a priori* about the unknown parameters. Let:

$$\mathcal{H}\left(\boldsymbol{\theta}\right) = \int f\left(\boldsymbol{t} \mid \boldsymbol{\theta}\right) log\left(f\left(\boldsymbol{t} \mid \boldsymbol{\theta}\right)\right) d\boldsymbol{t} \tag{2.49}$$

be the negative Shannon entropy of the distribution function $f\left(\boldsymbol{t} \mid \boldsymbol{\theta}\right)$, which is some measure of information. The functional criterion employed in the MDI approach is given by:

$$G\left(\pi\left(\boldsymbol{\theta}\right)\right) = \int_a^b \mathcal{H}\left(\boldsymbol{\theta}\right) \pi\left(\boldsymbol{\theta}\right) d\boldsymbol{\theta} - \int_a^b \pi\left(\boldsymbol{\theta}\right) log\left(\pi\left(\boldsymbol{\theta}\right)\right) d\boldsymbol{\theta}, \tag{2.50}$$

which is the prior average information in the data density less the information from the prior density. $G\left(\pi\left(\boldsymbol{\theta}\right)\right)$ is maximized by selecting $\pi\left(\boldsymbol{\theta}\right)$ subject to $\int_a^b \pi\left(\boldsymbol{\theta}\right) d\boldsymbol{\theta} = 1$, and is given by:

$$\pi_{MDI}\left(\boldsymbol{\theta}\right) = \frac{1}{c} \times exp\left\{\mathcal{H}\left(\boldsymbol{\theta}\right)\right\} \quad, \tag{2.51}$$

where:

- $a \leq \boldsymbol{\theta} \leq b$ and

- c^{-1} is is the normalizing constant.

The MDI prior however has some invariant limitations, as it is only invariant to linear transformations of either $\boldsymbol{t}$ or $\boldsymbol{\theta}$.

2.4.5.4 Uniform prior

The uniform prior is the simplest non-informative prior considered. Under the uniform prior, the unknown parameters are assigned a prior distribution on the interval (a, b) using the uniform distribution. Notationally it is given by:

$$\pi_U\left(\boldsymbol{\theta}\right) \propto c, \tag{2.52}$$

for some constant c. Since the uniform prior is a constant function, it implies that all possible values are equally likely *a priori*.

2.4.5.5 Probability matching prior

The PMP is a prior distribution which aims to connect frequentist methods to the Bayesian approach by ensuring that non-informative priors have desirable frequentist properties.

Peers (1965) provided a definition of a PMP for a parameter of interest: if $T_1, T_2, ..., T_n \mid \boldsymbol{\theta}$ are independent and identically distributed continuous random variables with density function $f(\boldsymbol{t} \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_i, \boldsymbol{\theta}')$ is a vector of parameters with θ_i, a scalar component of interest. Letting $\theta_i^{(1-\alpha)}(\pi, \boldsymbol{T})$ be the $(1-\alpha)^{th}$ posterior quantile of θ_i under a prior $\pi(\theta_i)$, then $\pi(\theta_i)$ is considered the rth-order PMP if it satisfies:

$$P_{\theta_i}\left(\theta_i \leq \theta_i^{(1-\alpha)}(\pi, \boldsymbol{T})\right) = 1 - \alpha + \mathcal{O}\left(n^{-\frac{r}{2}}\right), \tag{2.53}$$

where:

- P_{θ_i} is the frequentist confidence interval,

- n the sample size,

- $r > 0$ and

- for some $0 < \alpha < 1$.

That is, a prior is deemed the rth-order PMP for a parameter of interest if it provides a coverage probability of a Bayesian credible interval that is asymptotically equivalent to the coverage probability of the frequentist confidence interval up to a remainder term that converges to zero faster than $n^{-\frac{r}{2}}$. Hence larger values of r are more desirable since the actual coverage converges faster to the nominal level, and thus provide confidence intervals with more accurate coverage.

Rationale for such a prior is given by Datta and Sweeting (2005): they are attractive to Bayesians as they produce a suitable candidate for an objective prior, and to frequentists as a means of getting approximate confidence intervals with a Bayesian interpretation.

Datta and Ghosh (1995a) provide a method for developing second-order probability matching priors for a one-sided credibility interval:

1. For a vector of p unknown parameters $\boldsymbol{\theta}$, derive the Fisher information matrix as defined in equation 2.40, and, provided that it exists find its inverse, denoted by $I^{-1}(\boldsymbol{\theta})$.

2. Let $t(\boldsymbol{\theta})$ be a vector of real-valued twice continuously differentiable parametric functions of interest, and let the gradient of $t(\boldsymbol{\theta})$ be defined as: $\nabla'_t(\boldsymbol{\theta}) = \left(\frac{\partial t(\boldsymbol{\theta})}{\partial \theta_1}, \frac{\partial t(\boldsymbol{\theta})}{\partial \theta_2}, ..., \frac{\partial t(\boldsymbol{\theta})}{\partial \theta_p}\right)'$.

3. Define: $\eta(\boldsymbol{\theta}) = \frac{\nabla'_t(\boldsymbol{\theta})I^{-1}(\boldsymbol{\theta})}{\sqrt{\nabla'_t(\boldsymbol{\theta})I^{-1}(\boldsymbol{\theta})\nabla_t(\boldsymbol{\theta})}}$.

4. The prior $\pi(\boldsymbol{\theta})$ is considered a probability matching prior if and only if it is a solution to the following differential equation:

$$\sum_{i=1}^{p} \frac{\partial}{\partial \theta_i} \{\eta_i(\boldsymbol{\theta}) \pi(\boldsymbol{\theta})\} = 0. \tag{2.54}$$

The merit of using this method to solve probability matching priors is that it is invariant under monotonic transformations (Datta and Ghosh, 1995b).

2.4.6 Bayesian inference

Bayesian inference requires finding expectations related to the parameters from the posterior distribution. The posterior expectation of some function of $\boldsymbol{\theta}$, given here by $w(\boldsymbol{\theta})$, may be written in the form:

$$E_{post}(w(\boldsymbol{\theta}) \mid \boldsymbol{t}) = \frac{\int w(\boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid \boldsymbol{t}) d\boldsymbol{\theta}}{\int \pi(\boldsymbol{\theta} \mid \boldsymbol{t}) d\boldsymbol{\theta}}, \tag{2.55}$$

for some posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{t})$. Solving equation 2.55 is often analytically impossible, hence some approximation technique needs to be used. This section is devoted to introducing two approximation techniques, namely:

- Lindley's and

- Tierney-Kadane approximation.

Monte-Carlo procedures are another popular method used to find approximations of equation 2.55, but are discussed in detail in section 2.5.

2.4.6.1 Lindley's approximation

Lindley (1980) provided a method to approximate the ratio of two integrals of the form given in equation 2.55. Consider the posterior expectation of some function $w(\boldsymbol{\theta})$, which may be written as:

$$E_{post}(w(\boldsymbol{\theta}) \mid \boldsymbol{t}) = \frac{\int w(\boldsymbol{\theta}) \times exp\{\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{t}) + \rho(\boldsymbol{\theta})\} d\boldsymbol{\theta}}{\int exp\{\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{t}) + \rho(\boldsymbol{\theta})\} d\boldsymbol{\theta}}, \tag{2.56}$$

where $\rho(\boldsymbol{\theta}) = log(\pi(\boldsymbol{\theta}))$ for some prior distribution $\pi(\boldsymbol{\theta})$. Expanding $\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{t})$ and $\rho(\boldsymbol{\theta})$ into a Taylor Series expansion around the MLEs of $\boldsymbol{\theta}$, Lindley (1980) found an approximation for the posterior expectation, given by:

$$\begin{aligned} E_{post}(w(\boldsymbol{\theta}) \mid \boldsymbol{t}) \approx &w\left(\hat{\boldsymbol{\theta}}\right) + \frac{1}{2} \sum_{i=1}^{p}\sum_{j=1}^{p} (w_{ij} + 2w_i\rho_j)\sigma_{ij} + \\ &+ \frac{1}{2}\left(\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} \mathcal{L}_{ijk}w_l\sigma_{ij}\sigma_{kl}\right) + \mathcal{O}\left(n^{-1}\right), \end{aligned} \tag{2.57}$$

where:

- $\hat{\boldsymbol{\theta}}=$ are the MLEs of $\boldsymbol{\theta}$,

- $\mathcal{L}_{ijk}=\left.\frac{\partial^3 \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{t})}{\partial \theta^i \partial \theta^j \partial \theta^k}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, where $i,j,k=0,1,2,3$ and $i+j+k=3$,

- $\sigma_{ij}\left(\hat{\boldsymbol{\theta}}\right)$ is the $(i,j)^{th}$ element of the minus inverse Hessian evaluated at $\hat{\boldsymbol{\theta}}$,

- $w_i\left(\hat{\boldsymbol{\theta}}\right)=\left.\frac{\partial w(\boldsymbol{\theta})}{\partial \theta_i}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, where θ_i is the ith element in $\boldsymbol{\theta}$ for $i=1,2,...,p$,

- $w_{ij}\left(\hat{\boldsymbol{\theta}}\right)=\left.\frac{\partial^2 w(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ and

- $\rho_i\left(\hat{\boldsymbol{\theta}}\right)=\left.\frac{\partial \rho(\boldsymbol{\theta})}{\partial \theta_i}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$.

Since third-order partial derivatives need to be calculated, this method can be difficult for models with many variables. Press (2009) suggests that this method should only be used when the number of parameters ≤ 5.

2.4.6.2   Tierney-Kadane approximation

Tierney and Kadane (1986), like Lindley (1980), proposed a method to approximate the ratio of two integrals. Their method is based on the Laplace method, which is used to approximate an integral of the form:

$$I=\int_a^b b\left(\theta\right)\times exp\left\{-nh\left(\theta\right)\right\}d\theta, \tag{2.58}$$

where:

- n is some large, positive number,

- $h\left(\theta\right)$ is smooth and has its minimum value on the interval $a\leq\theta\leq b$ at $\theta=\hat{\theta}$ and

- $b\left(\theta\right)$ is smooth.

Following Tierney et al. (1989), the Laplace expansion of I is:

$$\begin{aligned} I\approx &\sqrt{2\pi\sigma^2}\times exp\left\{-n\hat{h}\right\}\times \\ &\times\left[\hat{b}+\frac{1}{2n}\left(\sigma^2\hat{b}''-\sigma^4\hat{b}'\hat{h}'''+\frac{5}{12}\hat{b}\left(\hat{h}'''\right)^2\sigma^6-\frac{1}{4}\hat{b}\hat{h}^{(4)}\sigma^4\right)\right]+\mathcal{O}\left(n^{-2}\right) \end{aligned} \tag{2.59}$$

where:

- $\hat{b}$ and $\hat{h}$ denote the values of $b\left(\theta\right)$ and $h\left(\theta\right)$ and their derivatives at $\theta=\hat{\theta}$ respectively and

- $\sigma^2=\left[h''\left(\tilde{\theta}\right)\right]^{-1}$.

Thus using the Laplace method, Tierney and Kadane (1986) claimed that the posterior expectation of some function $w(\theta)$ may be written as:

$$E_{post}(w(\theta) \mid t) = \frac{\int exp\{ng^\star(\theta)\} \, d\theta}{\int exp\{ng(\theta)\} \, d\theta}, \qquad (2.60)$$

where:

- $g(\theta) = \frac{1}{n}(log(\pi(\theta)) + \mathcal{L}(\theta \mid t))$,

- $g^\star(\theta) = g(\theta) + \frac{1}{n}log(w(\theta))$ and

- n is the sample size.

Following the expansion in equation 2.59, the posterior expectation reduces to:

$$E_{post}(w(\theta) \mid x) \approx \frac{\sigma^\star}{\sigma} exp\left\{n\left(g^\star\left(\hat{\theta}^\star\right) - g\left(\hat{\theta}\right)\right)\right\}, \qquad (2.61)$$

where:

- $g\left(\hat{\theta}\right)$ and $g^\star\left(\hat{\theta}^\star\right)$ have their maximum at $\hat{\theta}$ and $\hat{\theta}^\star$ respectively,

- $\sigma = \left(g''\left(\hat{\theta}\right)\right)^{-\frac{1}{2}}$ and

- $\sigma^\star = \left(g''^\star\left(\hat{\theta}^\star\right)\right)^{-\frac{1}{2}}$.

In the multiparameter case equation 2.61 becomes:

$$E_{post}(w(\boldsymbol{\theta}) \mid t) \approx \left(\frac{det\left(H^{-1}\left(\hat{\boldsymbol{\theta}}^\star\right)\right)}{det\left(H^{-1}\left(\hat{\boldsymbol{\theta}}\right)\right)}\right)^{\frac{1}{2}} \times exp\left\{n\left(g^\star\left(\hat{\boldsymbol{\theta}}^\star\right) - g\left(\hat{\boldsymbol{\theta}}\right)\right)\right\}, \qquad (2.62)$$

where:

- $g\left(\hat{\boldsymbol{\theta}}\right)$ and $g^\star\left(\hat{\boldsymbol{\theta}}^\star\right)$ have their maximum at $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^\star$ respectively and

- $H^{-1}(\boldsymbol{\theta})$ and $H^{-1}(\boldsymbol{\theta}^\star)$ are minus the inverse Hessians of $g\left(\hat{\boldsymbol{\theta}}\right)$ and $g^\star\left(\hat{\boldsymbol{\theta}}^\star\right)$ at $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^\star$ respectively.

Robert (2007) believes that this method is better than that of Lindley (1980) since it only requires solving the second-order partial derivatives of $w(\boldsymbol{\theta})$ and, despite only being justified asymptotically, appears to perform well in most cases. However he also notes that the feasibility of using this method is dependent on the difficulty of solving for the maximum values of both g and $g^\star$.

2.5   Monte Carlo procedures

Bayesians often need to integrate over complex and high-dimensional posterior distributions in order to make inference on estimates. The approximation techniques described in section 2.4.6 are known to work well, however in high-parameter models they can be incredibly tedious to compute. MCMC, a form of Monte Carlo integration using Markov chains, provides a reliable method to both finding the form of these complex posterior distributions and providing a method to solve equation 2.4.6.

Algorithms used to compile MCMC have been around since the 1950s, but only came into mainstream statistical practices 40 years later due to the rise in computing power (Gilks et al., 1995). Since then, MCMC methods have revolutionized Bayesian research, and has extended the breadth of its practical applications.

This section is devoted to discussing the idea behind Markov chains, the different algorithms used in MCMC and to describing convergence of the chains.

2.5.1   Markov chains

Markov chains were introduced by Russian mathematician Markov (1906) who extended the theory of probability in a new way – to sequences of linked events.

Consider a discrete time stochastic process $\{X_0, X_1, X_2, ...\}$ such that the next value in the process X_{t+1}, $t \geq 0$ is dependent only on the current state of the chain and not on the history of the chain $\{X_0, X_1, ..., X_{t-1}\}$, then this process is called a Markov chain. MCMC is continuous-valued generalization of a Markov chain, and mathematically it can be written as:

$$\pi\left(X_{t+1} \in A \mid X_0, X_1, ..., X_t\right) = \pi\left(X_{t+1} \mid X_t\right), \tag{2.63}$$

for any set A and where π denotes the transition kernel. A Markov chain is said to have reached its stationary distribution, the MCMCs target distribution, if the conditional distribution of X_{t+1} given X_t does not depend on t, and the distribution of X_t converges to a stationary distribution if it satisfies three properties, namely the chain is:

1. Irreducible, that is, it can reach any other state with some non-zero probability after a certain number of iterations.

2. Aperiodic, that is, the chain does not move between states in periodic intervals.

3. Positive recurrent, that is, the expected amount of time to return to a state is finite.

Finally after discarding m iterations of the chain as a burn-in period (*Vida infra*), estimates from the MCMC can be summarized by ergodic averages in the form:

$$E_{post}\left(w\left(\boldsymbol{\theta}\right) \mid \boldsymbol{t}\right) \approx \frac{1}{n-m} \sum_{t=m+1}^{n} w\left(X_t \mid \boldsymbol{t}\right) \tag{2.64}$$

where w is some real valued function of the parameters $\boldsymbol{\theta}$ that need to be estimated. Provided that the samples are independent and n is sufficiently large, the law of large numbers ensures that this approximation is accurate (Gilks et al., 1995). Furthermore, the Monte Carlo standard error (MCSE) of the chain is given by:

$$SE_{post}\left(w\left(\boldsymbol{\theta}\right)\mid\boldsymbol{t}\right)=\sqrt{\frac{\sum\limits_{t=m+1}^{n}\left(w\left(X_t\mid\boldsymbol{t}\right)-E_{post}\left(w\left(\boldsymbol{\theta}\right)\mid\boldsymbol{t}\right)\right)^2}{\left(n-m\right)\left(n-m-1\right)}}. \tag{2.65}$$

2.5.2 Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithm was first introduced by Metropolis et al. (1953) who used it in a mechanical physics context. The algorithm was later generalized by Hastings (1970) for statistical applications, and is a commonly used algorithm in MCMC.

The idea behind the MH algorithm is that at each time t, the next state in the chain X_{t+1} is chosen by first sampling a candidate point Y_t from some proposal density given by $q\left(\cdot\mid X_t\right)$. This proposal density can have a variety of forms, but needs to be aperiodic, irreducible and a good approximation of the target distribution (Gilks et al., 1995). The candidate Y_t generated from the proposal density gets accepted as X_{t+1}, the next state of the chain with acceptance probability:

$$P=min\left\{1,\frac{\pi\left(Y_t\right)q\left(X_t\mid Y_t\right)}{\pi\left(X_t\right)q\left(Y_t\mid X_t\right)}\right\}, \tag{2.66}$$

otherwise if the candidate is rejected the chain does not move and thus $X_{t+1}=X_t$. Note that when q is symmetric, that is $q\left(Y_t\mid X_t\right)=q\left(X_t\mid Y\right)$, the acceptance probability reduces to:

$$P=min\left\{1,\frac{\pi\left(Y_t\right)}{\pi\left(X_t\right)}\right\}. \tag{2.67}$$

This was an assumption in the initial Metropolis algorithm proposed by Metropolis et al. (1953), however the updated MH algorithm allows q to be an asymmetric function.

Gilks et al. (1995) provide caution when choosing a proposal density. If the scale of q is too small, there will be a high acceptance rate and the chain move slowly. If the scale of q is too large, there will be a low acceptance rate which may cause the chain to be stuck. Therefore q needs to be scaled appropriately to avoid both of these issues.

The general MH algorithm is given by Brooks et al. (2011) as:

1. Initialize the iteration counter to $t=0$ and choose an arbitrary initial value given by X_0.

2. Generate a value Y_t using the proposal density given by $q\left(\cdot\mid X_t\right)$.

3. Generate a uniform random variable U on the interval $(0,1)$.

4. If $U\leq P=min\left\{1,\frac{\pi(Y_t)q(X_t\mid Y_t)}{\pi(X_t)q(Y_t\mid X_t)}\right\}$, then set $X_{t+1}=Y_t$, otherwise set $X_{t+1}=X_t$.

5. Set $t = t + 1$ and repeat steps $2 - 5$ many times.

Robert and Casella (1999) claim that the algorithm depends on only two ratios, $\frac{\pi(Y_t)}{\pi(X_t)}$ and $\frac{q(X_t|Y_t)}{q(Y_t|X_t)}$, therefore being independent of any normalizing constant, assuming that q is independent of Y_t. However Lawson (2009) claims that a difficulty with the algorithm is that it does not guarantee the acceptance of a new value, resulting in convergence difficulties.

2.5.3 Gibbs sampling

The Gibbs sampler was first introduced by Geman and Geman (1984), who used it to sample from high-dimensional complex distributions used in image restoration. The sampler was later introduced to statistics by Gelfand and Smith (1990), and since then has been a prominent technique used by Bayesians to sample from the posterior.

To define the Gibbs sampler let the random variables of the target distribution be $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)'$, where the θ_i's are either uni- or multidimensional. The fundamental idea behind Gibbs sampling is to solve a large problem by breaking it down into smaller pieces and solving each piece individually.

To do this we generate samples from the target distribution by repeatedly taking samples from the conditional distribution of its random variables conditioned on all the other random variables, thus forming a Markov chain. The Gibbs sampler is actually a special case of the MH algorithm, where the proposal density for updating the ith component of $\boldsymbol{\theta}$ is given as the full conditional distribution $\pi(\theta_i \mid \theta_{-i})$, where θ_{-i} represents all the components of $\boldsymbol{\theta}$ except for θ_i.

The general Gibbs sampler algorithm is given by Brooks et al. (2011) as:

1. Choose arbitrary initial values, $\theta_1^{(0)}, \theta_2^{(0)}, ..., \theta_p^{(0)}$.

2. Generate a sample from:

   2.1 $\qquad \theta_1^{(i)}$ from $\pi\left(\theta_1 \mid \theta_2^{(i-1)}, \theta_3^{(i-1)}, ..., \theta_p^{(i-1)}\right)$,

   2.2 $\qquad \theta_2^{(i)}$ from $\pi\left(\theta_2 \mid \theta_1^{(i)}, \theta_3^{(i-1)}, ..., \theta_p^{(i-1)}\right)$,

   ... ...

   2.p $\qquad \theta_p^{(i)}$ from $\pi\left(\theta_p \mid \theta_1^{(i)}, \theta_2^{(i)}, ..., \theta_{p-1}^{(i)}\right)$.

3. Increment i by 1 and repeat step 2 many times.

A benefit of using the Gibbs sampler is found by substituting the proposal density into equation 2.66, which always results in an acceptance probability of unity. Therefore the Gibbs sampler guarantees the acceptance of a new value for each iteration, amending the convergence issue with the MH algorithm. However despite the Gibbs sampler converging faster than the MH algorithm, it requires the full conditional distribution of θ_i which is often difficult to obtain or computationally expensive to run (Lawson, 2009). In those cases the MH algorithm is preferred.

2.5.4   Slice sampling

The Gibbs sampling technique requires that the conditional posterior distributions of a distribution are readily available, which often is not the case.

Slice sampling (or sometimes, the slice Gibbs sampler) is a sampling technique based on the Gibbs sampler, but used when the conditional distributions are not of a convenient form (Ntzoufras, 2011). This method works by augmenting the parameter space of a distribution by adding a set of auxiliary random variables that convert the conditional distributions to a convenient form whilst keeping the marginal posterior distribution the same.

Consider some distribution $\pi(\theta)$ – either uni- or multiparameter – which may be written by the following product:

$$\pi(\theta) = \prod_{i=1}^{p} f_i(\theta), \tag{2.68}$$

where: $f_i(\theta) > 0 \; \underset{i}{\forall}$.

Then $\pi(\theta)$ may be written as:

$$\pi(\theta) = \int \prod_{i=1}^{p} \Psi\left(0 < \omega_i < f_i(\theta)\right) df_1 df_2 ... df_p, \tag{2.69}$$

where:

- the w_i's are a set of auxiliary variables chosen based on convenience and

- $\Psi(\cdot)$ is an indicator function taking the value unity if $0 < \omega_i < f_i(\theta)$ or 0 elsewhere.

Robert and Casella (1999) give the slice sampler algorithm as:

1. Choose an arbitrary initial value, given by: $\theta^{(0)}$.

2. At iteration i, simulate:

   2.1 $\omega_1^{(i)}$ from a uniform random variable U on the interval $(0, f_1(\theta))$,

   2.2 $\omega_2^{(i)}$ from a uniform random variable U on the interval $(0, f_2(\theta))$,

   ...   ...

   2.p $\omega_p^{(i)}$ from a uniform random variable U on the interval $(0, f_p(\theta))$.

   2.p+1   Update $\theta^{(j)}$ from a uniform random variable U on the interval $A^{(i)}$ with $A^{(i)} = \left\{\xi : f(\xi) \geq \omega_j^{(i)} \; j = 1, 2, ..., p\right\}$.

3. Increment i by 1 and repeat step 2 many times.

Ntzoufras (2011) claims that the benefits of using the slice sampler is that it requires neither the full conditional distribution of a Gibbs sampler nor the specification of a proposal density as used in the MH algorithm. However the major disadvantage for using such a sampling

scheme is that it often results in highly correlated chains as a result of the augmented param-
eter space. Furthermore Robert and Casella (1999) claim that as p increases finding the set
$A^{(i)}$ becomes more challenging. To amend this issue, Neal (2003) proposed an adaptive slice
sampling algorithm, which is used in the popular Bayesian statistical software, $WinBUGS$
($vide\ infra$).

2.5.5   Rejection sampling

2.5.5.1   Non-adaptive rejection sampling

Non-adaptive rejection sampling, or alternatively the accept-reject method, is a method used
to sample from a complex target distribution. Suppose that the target density has the form
$\pi(x)$ and we can find a density $g(x)$, deemed the envelope density, and some positive constant
c, deemed the envelope constant, such that the following inequality holds:

$$c \geq \frac{\pi(x)}{g(x)}. \tag{2.70}$$

Although the target density may be computationally expensive to evaluate, it only needs to
be known up to a multiplicative constant (Robert and Casella, 1999).

Furthermore the envelope density must be picked such that it is both close to the target
density and computationally easy to sample from. The value of c is merely chosen to ensure
that the envelope density is above the target density, and its choice is determined by how
close $g(x)$ is to $\pi(x)$ (Gamerman and Lopes, 2006). When $g(x)$ is close to $\pi(x)$, the constant
c needs only be slightly bigger than unity, however when $g(x)$ is substantially different from
$\pi(x)$, c needs to be far larger than unity.

To compile the method, a random value y from $g(x)$ is generated and is accepted with
probability:

$$P = \frac{\pi(x)}{cg(x)}. \tag{2.71}$$

If the random value y is accepted set $x = y$, otherwise if y is not accepted it is discarded
and a further value is generated from $g(x)$. This process continues until at least one y is
accepted.

The general algorithm for non-adaptive rejection sampling is given by Gilks et al. (1995) as:

1. Generate a value y from the density $g(x)$.

2. Generate a uniform random variable U on the interval $(0, 1)$.

3. If $U \leq \pi(x)/cg(x)$, set $x = y$, otherwise return to step 1 and repeat many times.

The issue with non-adaptive rejection sampling is that its effectiveness relies on the choice of
c and $g(x)$. Finding a suitable $g(x)$ is challenging since it requires that the constraint given

in equation 2.70 is satisfied, which involves tiresome maximization techniques (Gilks et al., 1995).

2.5.5.2 Adaptive rejection sampling

ARS is a sampling method proposed by Gilks and Wild (1992) and is based off the non-adaptive sampling algorithm with an added assumption that $h(x) = log(\pi(x))$ is concave. This assumption is useful since if the target distribution is log-concave, it allows the maximization techniques needed in the non-adaptive rejection sampling method to be avoided.

There are two types of ARS, namely: the tangent method and the derivative-free method. The former will be discussed in this literature review. The fundamental idea of the tangent method is to partition the domain of $\pi(x)$, given by D, into separate intervals and construct an envelope function on each of these intervals. That is suppose $h(x)$ and $h'(x)$ have been evaluated at a set of k abscissae $S = (x_1, x_2, ..., x_k) \subset D$ sorted in ascending order. Gilks and Wild (1992) showed that by letting $T_k = \{x_i : i = 1, 2, ..., k\}$, a rejection envelope on T_k can be defined as $exp\{u_k(x)\}$ where $u_k(x)$ is a piecewise linear upper hull formed from the tangents to $h(x)$ at the abscissae in T_k, given as:

$$u_k(x) = h(x_j) - (x + x_j) h'(x_j), \ x \in [z_{j-1}, z_j], \tag{2.72}$$

where for z_j represents the intersection point of the tangents at x_j and x_{j+1} for $j = 1, 2, ..., k-1$. A squeezing function on T_k can be defined as $exp\{l_k(x)\}$, where $l_k(x)$ is a linear lower hull formed from the chords between adjacent abscissae in T_k, given as:

$$l_k(x) = \frac{(x_{j+1} - x) h(x_j) + (x - x_j) h(x_{j+1})}{x_{j+1} - x_j}, \ x \in [x_j, x_{j+1}]. \tag{2.73}$$

And since $h(x)$ is concave, the following inequality holds:

$$exp\{l_k(x)\} \leq \pi(x) \leq exp(u_k(x)), \ \forall x \in D. \tag{2.74}$$

A general algorithm for ARS is given by Gilks et al. (1995) as:

1. Initialize S

2. Generate a random value y from the density $exp(u_k(x))$.

3. Generate a uniform random variable U on the interval $(0, 1)$.

4. Compute the following:

    (a) If $U \leq exp\{l_k(x)\}/exp(u_k(x))$, set $x = y$, include y in S and return to step 1, otherwise;

    (b) If $U \leq \pi(x)/exp(u_k(x))$, set $x = y$, include y in S and return to step 1, otherwise;

(c) Return to step 1.

The merit of using ARS is that it can be used adaptively. That is the set S is updated every time the random value Y is accepted, resulting in an increase in accuracy for the two envelope functions, and thus progressively reducing both the number of evaluations required for $\pi(x)$ and the risk of further rejections (Robert and Casella, 1999).

2.5.6   Software and packages

The BUGS (Bayesian inference using Gibbs sampling) project is courtesy of Lunn et al. (2000), and is one of the most versatile options available for performing Bayesian analysis via MCMC methods. *WinBUGS* is a statistical-software based on the BUGS project that is accessible to all due to its user-friendly, 'point-and-click' environment. Bayesian analysis in this paper will be performed on $R^{\circledR}$ (R Core Team, 2013) via the *R2WinBUGS* (Sturtz et al., 2005) package. $R^{\circledR}$ has become the *lingua franca* of contemporary applied statistics, and using it in Bayesian analysis is often more convenient and practical than on *WinBUGS* itself. Furthermore, Bayesian graphics were created on the $R^{\circledR}$ packages: *coda* (Plummer et al., 2006) and *ggmcmc* (Fernández-i Marín, 2016).

2.6   Convergence of a model

This section is devoted to talking about the many ways to test for convergence of a Markov chain, as well as methods used to improve convergence. Assessing the convergence of a model is incredibly important, as without it the results obtained will be unacceptable.

Although it is important to check for converge, it is also possible that one could fall into the trap of pseudo-convergence – when the chain appears to have converged when it in fact has not (Brooks et al., 2011). This phenomenon occurs when parts of the state space are poorly connected – that is, it takes too many iterations to get from one point to another. As a result the Markov chain converges to its equilibrium distribution conditioned on the part in which the chain was started, but not to its true equilibrium distribution.

2.6.1   Graphical methods

Graphical methods are an informal method to assessing the convergence of a model. Despite not being backed by any valid statistical reasoning, these tests are necessary for a quick and easy approach in showing whether the model stabilizes. Three graphical methods will be discussed:

- autocorrelation,

- trace and

  - running mean plots.

2.6.1.1   Autocorrelation plots

An autocorrelation plot is a histogram used to show the correlation between a sample value and its successive value, and hence shows dependence between samples in a Markov chain. The plot shows the relationship between the lag of the sampled value and the magnitude of the autocorrelation, which is some value between ± 1, where the first bar representing lag zero has autocorrelation unity. Autocorrelation plots are incredibly important as they show the mixing rate of the chain (Lesaffre and Lawson, 2012).

When autocorrelation decreases slowly, the mixing rate is slow and hence so is the rate of convergence. Autocorrelation plots are also useful because they can indicate the minimum number of iterations required before the chain reaches its stationary distribution. Despite being a significant tool in assessing convergence autocorrelation plots should not be used as a convergence diagnostic, since even if autocorrelation is high in a model it does not imply an absence of convergence, but rather a low mixing rate.

2.6.1.2   Trace plots

A trace plot for a parameter under consideration plots the number of iterations of the model against the generated value of a parameter, and is used to show convergence of a model (Lesaffre and Lawson, 2012). A chain has reached convergence when the plot looks like a random scatter around a stationary mean, or informally, like a "hairy caterpillar" (Lunn et al., 2012).

An obvious issue which may occur is that the chain appears to have reached convergence, but rather got trapped in some region rather than exploring the entire posterior. Another method to show convergence with trace plots is to simulate two or more chains with significantly different initial values, and claim convergence when the chains come together and behave in a similar manner.

2.6.1.3   Running mean plot

A running mean plot shows how well the chain has mixed by plotting the iteration against the mean of all draws up to and including that iteration (Lesaffre and Lawson, 2012). Initially the variability of the running mean plot should be high, but as the number of iterations increase the plot should stabilize to its mean value as the stationary distribution is reached.

2.6.2   Diagnostic tests

Diagnostic tests are used as a formal method throughout MCMC to ensure model accuracy, and as Cowles and Carlin (1996) state: "a weak diagnostic test is better than no diagnostic

at all." This literature review discusses two of the most popular diagnostic tests used in MCMC:

- Brooks-Gelman-Ruben (BGR) diagnostic and

- Geweke diagnostic.

2.6.2.1   Brooks-Gelman-Ruben diagnostic

The BGR diagnostic was initially proposed by Gelman and Rubin (1992) and later modified by Brooks and Gelman (1998), and has since become one of the most popular convergence tests for a multi-modal posterior distribution. The test runs multiple chains and evaluates convergence by comparing the estimated between-chain and within-chain variances for the model parameters. If there is a small difference between these variances, convergence has occurred.

To compute a BGR, consider running m chains, $\boldsymbol{\varsigma}' = (\varsigma_1, \varsigma_2, ..., \varsigma_m)'$, starting from significantly different starting points, and run them for a post-burn-in period of n iterations. The mean of an individual chain is $\bar{\varsigma}_i = \frac{1}{n} \sum_{i=1}^{n} \varsigma_i$ and the overall mean of all the chains is $\bar{\varsigma} = \frac{1}{m} \sum_{i=1}^{m} \bar{\varsigma}_i$. The between-chains variance can be computed as:

$$B = \frac{n}{m-1} \sum_{i=1}^{m} (\bar{\varsigma}_i - \bar{\varsigma})^2 , \tag{2.75}$$

and the within-chain variance is given by:

$$W = \frac{1}{m} \sum_{i=1}^{m} s_i^2, \tag{2.76}$$

where: $s_i^2 = \frac{1}{n} \sum_{k=1}^{n} \left(\varsigma_i^k - \bar{\varsigma}_i \right)$. Under the stationary distribution, the unbiased pooled posterior variance is given by:

$$\hat{V} = \frac{n-1}{n} W + \frac{1}{n} B. \tag{2.77}$$

Finally the potential scale reduction factor (PSRF) can be computed as:

$$\hat{R} = \frac{d+3}{d+1} \times \frac{\hat{V}}{W}, \tag{2.78}$$

where $d \approx 2\hat{V}/v\hat{a}r\left(\hat{V}\right)$ is used to account for sampling variability. Brooks and Gelman (1998) suggests that if $\hat{R} < 1.2$, the model has reached convergence for all model parameters, otherwise the chain needs to run longer.

A graphical version of the BGR was also proposed by Brooks and Gelman (1998) and provides a quick method to evaluating convergence. The BGR graph plots the relationship between $\hat{R}$ and the number of iterations, and convergence occurs when $\hat{R}$ approaches unity.

2.6.2.2   Geweke diagnostic

Geweke (1992) provided a diagnostic test to check for convergence. The test uses a frequentist significance test for the equality of means of an early part of the Markov chain, usually the first 10% of values, and a late part of the Markov chain, usually the last 50% of values. It is assumed that the late part of the chain has converged, and the early part has not. Therefore if the two mean values are close to each other we assume that the two different parts of the chain are in a similar location in the state space and thus come from a similar distribution.

Let n values of one chain ς_i be split into two separate parts, ς_A representing the n_A elements of the first part of the chain, and ς_B, representing the n_B elements of the late part of the chain. Let their respective posterior means be represented by $\bar{\varsigma}_A$ and $\bar{\varsigma}_B$, which can be compared with the unpaired Z-test given by:

$$Z = \frac{\bar{\varsigma}_A - \bar{\varsigma}_B}{\sqrt{\frac{s_A^2}{n_A} - \frac{s_B^2}{n_B}}}, \tag{2.79}$$

where:

- s_A^2 and s_B^2 are the classical estimates of the respective variances for ς_A and ς_B, and

- Z approximately follows a standard normal distribution as $n_a, n_b \to \infty$.

For large values of Z it is concluded that the two parts come from different distributions and hence a longer burn-in period is required. A plot of the diagnostic described by Lesaffre and Lawson (2012) is also possible, which plots the Z statistic for successively smaller segments of the last part of the chain.

2.6.3   Improving convergence

The rate of convergence of a model is heavily dependent on the choice of model, and making simple changes within a model can help with the convergence process. Whilst none of these techniques considered are mandatory and some may not work in the context of all problems, they might aid in obtaining accurate estimates and reduce required processing power.

2.6.3.1   Burn-in period

An adequate burn-in period is required to insure accurate and independent samples from a chain (Lesaffre and Lawson, 2012). A burn-in period is the initial part of the chain that is discarded before the chain reaches stochastic stability and any inference on estimates is made. Determining the length of the burn-in period is difficult since rates of convergence of different models may vary considerably, more complex models require a longer burn-in period than their less complex counterparts.

Brooks (1998) provides a method estimating the length of the burn-in period. Given geometric ergodicity, the t-step transition distribution denoted by $\pi'(\boldsymbol{t},\cdot)$ is such that:

$$|\pi'(\boldsymbol{t},\cdot) - \pi(\cdot)| \leq M(\boldsymbol{t})\,p', \tag{2.80}$$

for some $M, p \in \mathbb{R}$. The chain is stopped when $|\pi'(\boldsymbol{t},\cdot) - \pi(\cdot)| \leq \epsilon$, for some $\epsilon > 0$ in which case the length of the burn-in period is given by:

$$t^\star = \frac{log\,(\epsilon/M\,(\boldsymbol{t}))}{log\,(p)}. \tag{2.81}$$

However this method is impractical since it is difficult to prove the existence of a geometric rate of convergence and since many commonly used algorithms fail to converge geometrically quickly.

Other informal methods used to determine convergence have been discussed, such as the graphical *thick pen technique* from Gelfand et al. (1990), which claims convergence occurs when the difference in density estimates obtained via univariate density plots is less than the width of a felt-tip pen.

2.6.3.2   Stopping time

Deciding the length of time a chain must run for is also an important matter. Despite being computationally expensive to run a chain for a large number of iterations, running a chain for too few iterations results in imprecise estimates. A proposal to determining the length of the chain is made by Gilks et al. (1995): run several chains of length n in parallel, and compare the values of the estimates. If the values are significantly different, increase the value of n.

2.6.3.3   Initial values

Gelman and Rubin (1992) claim that well selected starting points with respect to the target distribution need to be selected to insure convergence of the model. A poor choice of initial values results in the chain becoming stuck in a certain area, and this issue can be observed on a trace plot (Lesaffre and Lawson, 2012).

The trace plot of a chain with poorly chosen initial values will have an increasing or decreasing trend resulting in the posterior probability being low in that area. An obvious remedy is to rechoose new values until the convergence improves.

Brooks (1998) suggests some *ad hoc* methods to selecting initial values, namely:

- setting hyperparameters to fixed values,

- simplifying the model,

- choosing maximum likelihood estimates as the initial values or by

- ignoring missing values.

More elaborate techniques to choosing initial values are presented by Gelman and Rubin (1992), who proposed a mode-finding algorithm to find areas of high density, and then sample from a t-distribution to find initial values in those areas, and by Brooks and Morgan (1994) who used a simulated annealing algorithm to generate initial values.

2.6.3.4 Number of chains

One of the most contentious topics in implementing MCMC techniques is in choosing whether to run one long chain or several short chains in parallel. Proponents of one long chain claim that the chain will be closer to the target posterior distribution than any short chains, and that short chains are wasteful after the initial burn-in period is considered Brooks (1998).

Rebuters argue that multiple chains help guard against a single chain leaving a significant proportion of a sample space unexplored. Also using multiple chains helps to determine how well the chains have mixed, that is how indistinguishable results from different chains are, thus protecting the results of any bias.

An alternative method to running multiple chains is proposed by Brooks (1998). He proposes a regenerative method which involves running one chain which is restarted at appropriate regeneration times. This essentially splits the one long chain into multiple replications that are closer to the stationary distribution in comparison to multiple independent chains since observations are taken from the end of a long chain. However Brooks (1998) finds fault with the method; it can be both more difficult and computationally expensive than simply running multiple chains.

2.6.3.5 Thinning

Thinning is a process in which only every m^{th} value ($m > 1$) in the chain is retained (Lesaffre and Lawson, 2012). The method of thinning is given notationally by Gilks et al. (1995): let X_t denote the value of the state at time t, then form the binary process:

$$
Z_t = \begin{cases} 1 & X_t \leq u \\ 0 & X_t > u, \end{cases}
$$

where u is some function of the Markov chain. Here Z_t is derived from a Markov chain, but is not one itself.

Finally we can form the new process $Z_t^m = Z_{1+(t-1)m}$, which is a Markov chain consisting of every m^{th} value of the original chain, provided that m is sufficiently large. The value of m is determined by comparing likelihood ratio test statistics and Bayesian information criterion (BIC) (*Vide infra*) values between chains with different values of m.

The main reason for thinning is to reduce high levels of autocorrelation in a model (Lesaffre and Lawson, 2012). This method may even minimize the autocorrelation in chains with lags greater than one to an autocorrelation of zero. An additional bonus to thinning a Markov chain is that it helps save computer memory and reduces processing times. However thinning is not without its critics. Link and Eaton (2012) claim that thinning is inefficient and results in less precise estimates and Lesaffre and Lawson (2012) state that thinning a chain results in a higher MCSE value in comparison to a chain that has not been thinned.

2.6.3.6   Transformations

Transforming helps improve convergence when there are high levels of multicollinearity or great differences in magnitude between the independent variables of a model (Lesaffre and Lawson, 2012). Solutions to reduce these problems is to render the independent variables unit-free by dividing by standardizing them or by utilizing the Gram-Schmidt orthogonalization technique to reduce high correlations. However Brooks (1998) remarks that in higher dimensions orthogonalization is unfeasible.

2.7   Model comparison

It is natural for us to want to compare different statistical models in order to choose the one which models the data "best". This thesis introduces two methods for comparing different models, namely:

- information criteria and

- loss functions.

2.7.1   Information criteria

2.7.1.1   Akaike information criterion

The Akaike information criterion (AIC) was suggested by Akaike (1973) to predict the accuracy of point estimates, typically the MLEs of a model, by approximating the expected Kullback-Leibler distance (*Vide ante*) between the estimated model and its true model. The AIC thus measures the amount of information lost by a model due to estimation, and hence, given a set of models, the model with the lowest AIC values is deemed best.

Let M_i be the ith model with estimates $\hat{\boldsymbol{\theta}}_i$ then the AIC for M_i is defined as:

$$AIC\left(i\right) = 2k_i - 2\mathcal{L}\left(\hat{\boldsymbol{\theta}}_i \mid \boldsymbol{t}\right), \tag{2.82}$$

where k_i represents the number of parameters estimated in M_i and $\mathcal{L}(\cdot)$ represents the maximum log-likelihood value for $\boldsymbol{\theta}_i$. Here k acts as a penalty, since increasing the number of parameters in a model naturally increases the goodness of fit of a model.

Burnham and Anderson (2004) claim that individual AIC values are arbitrary and often uninterpretable, and hence defined:

$$\triangle_i = AIC_i - AIC_{min}, \tag{2.83}$$

as a method to rescale AIC values, where AIC_{min} is the smallest value among a set of AIC values, and $\triangle_i$ represents the information lost from using model M_i as opposed to the best model. Therefore large $\triangle_i$ values imply a less plausible fit in comparison to the best model. Burnham and Anderson (2004) continue with a rule-of-thumb that if $\triangle_i \leq 2$ model M_i has substantial supporting evidence against the remaining models, if $4 \leq \triangle_i \leq 7$ model M_i has considerably less evidence and if $\triangle_i > 10$ model M_i has no support and should not be considered.

2.7.1.2   Bayesian information criterion

Schwarz (1978) proposed the BIC, closely related to the AIC, as a method to compare models. Let M_i be the ith model with estimates $\hat{\boldsymbol{\theta}}_i$ then the BIC for M_i is defined as:

$$BIC(i) = 2k_i log(n) - 2\mathcal{L}\left(\hat{\boldsymbol{\theta}}_i \mid \boldsymbol{t}\right), \tag{2.84}$$

where n represents the number of data points in $\boldsymbol{t}$, and hence for larger datasets, gives a larger penalty per parameter in comparison to the AIC. Therefore the BIC favors more parsimonious models, and similar to the AIC, the best model has the lowest BIC value.

2.7.1.3   Deviance information criterion

Spiegelhalter et al. (2002) provided the deviance information criterion (DIC), a Bayesian generalization of the AIC. It does so by replacing the MLE $\hat{\boldsymbol{\theta}}$ with the posterior mean $\hat{\boldsymbol{\theta}}_{Bayes} = E(\boldsymbol{\theta} \mid \boldsymbol{t})$, and by replacing the constant k with a data-based bias correction. To define the DIC first let:

$$D(\boldsymbol{\theta}) = -2\mathcal{L}\left(\hat{\boldsymbol{\theta}} \mid \boldsymbol{t}\right) + 2log(f(\boldsymbol{t})) \tag{2.85}$$

be the deviance evaluated at the posterior mean, where $f(\boldsymbol{t})$ is some fully specified standardizing term dependent on the data alone. Then define p_d, the effective number of parameters as:

$$p_d = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}), \tag{2.86}$$

where $\overline{D(\boldsymbol{\theta})}$ denotes the posterior mean of $D(\bar{\boldsymbol{\theta}})$. Then the DIC for a model is given by:

$$DIC = p_d + \overline{D(\boldsymbol{\theta})}, \tag{2.87}$$

or equivalently as:

$$DIC = D\left(\bar{\boldsymbol{\theta}}\right) + 2p_d, \tag{2.88}$$

such that it closely resembles the AIC. Like the AIC and the BIC, the best model is the one with the lowest DIC value. A benefit of using this method in a Bayesian context is that both p_d and DIC can be calculated from an MCMC simulation (Spiegelhalter et al., 2002). By letting $\theta_1, \theta_2, ..., \theta_m$ represent m chains post burn-in, $\overline{D\left(\boldsymbol{\theta}\right)}$ can be estimated by $\frac{1}{m}\sum\limits_{i=1}^{m} D\left(\theta_i\right)$ and $D\left(\bar{\boldsymbol{\theta}}\right)$ as $D\left(\frac{1}{m}\sum\limits_{i=1}^{m} D\left(\theta_i\right)\right).$

Despite being the most popular information criterion in Bayesian analysis, Spiegelhalter et al. (2014) provides some criticism of the DIC, mainly:

- p_d is not invariant to reparameterization,

- it lacks consistency,

- unlike the AIC, the DIC is not based on a proper predictive criterion and

- it has weak theoretical justification.

2.7.2   Loss functions

Loss functions are objective, non-negative functions used in analysis to measure how well a model does in predicting the expected outcome (Press, 2009). A loss function, denoted by: $L\left(\theta, \hat{\theta}\right)$ thus measures how much loss is incurred by estimating the expected parameter value θ of the model using $\hat{\theta}$. Furthermore if:

- $\theta = \hat{\theta}$ then we say that no loss has occurred,

- $\theta < \hat{\theta}$ we claim that $\hat{\theta}$ has overestimated θ and

- $\theta > \hat{\theta}$ we claim that $\hat{\theta}$ underestimated θ.

The objective here is to choose a $\hat{\theta}$ which minimizes the loss function.

Three loss functions will be discussed in this thesis:

- squared error,

- LINEX and

- the GELF.

2.7.2.1   Squared error loss function

The most well used loss function in Bayesian analysis is the squared error (or otherwise, quadratic) loss function, given by:

$$L_S\left(\theta, \hat{\theta}\right) = c\left(\theta - \hat{\theta}\right)^2, \tag{2.89}$$

where: c is some positive constant, usually chosen as unity (Press, 2009). This loss function is symmetric – that is, underestimates are as consequential as overestimates, which is often not a realistic assumption. Depending on the test, overestimation may be more severe than underestimation, or vice versa.

The Bayes estimator under the squared error loss function is given by:

$$\hat{\theta}_S = E_\theta\left(\theta \mid \boldsymbol{t}\right), \tag{2.90}$$

that is – provided the expectation exists and is finite, the Bayes estimator of the squared error loss function is identical to the mean of the posterior distribution.
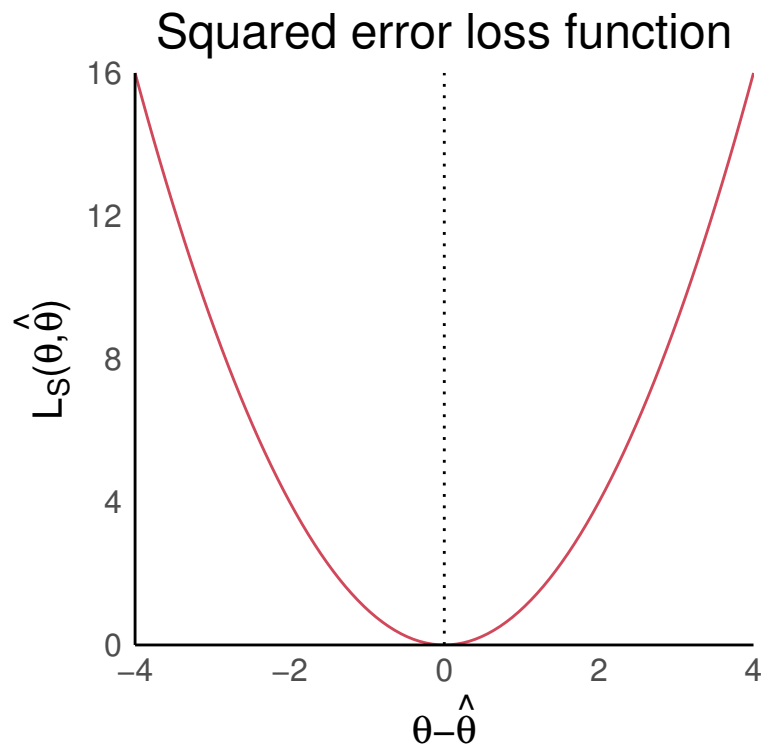
Figure 2.14: Figure illustrating the squared error loss function.

2.7.2.2   Linear exponential loss function

The LINEX loss function is an asymmetric loss function due to Jeffreys and Zellner (1989) and is given by:

$$L_L\left(\theta,\hat{\theta}\right) = b\left(exp\left\{a\left(\hat{\theta}-\theta\right)\right\} - a\left(\hat{\theta}-\theta\right) - 1\right),\tag{2.91}$$

where:

- $a \neq 0$ and

- $b > 0$.

Furthermore when:

- $a > 0$ overestimation is deemed more serious than underestimation,

- $a < 0$ underestimation is deemed more serious than overestimation and when

- $a \approx 0$ the LINEX loss function is approximately the quadratic loss function.

Therefore depending on the value of a, the LINEX loss function rises approximately linearly on one side of zero, and approximately exponentially on the other side of zero.

The posterior expectation of the LINEX loss function according to Zellner (1996) is:

$$E_\theta\left(L_L\left(\theta,\hat{\theta}\right)\right) \propto exp\left\{a\hat{\theta}\right\}E_\theta\left(exp\left\{-a\theta\right\}\right) - a\left(\hat{\theta} - E_\theta\left(\theta\right)\right) - 1.\tag{2.92}$$

The Bayes estimator under the LINEX loss function is the value of $\hat{\theta}$ which maximizes equation 2.92, and is given by:

$$\hat{\theta}_L = -a^{-1}log\left(E_\theta\left(exp\left\{-a\theta\right\}\right)\right),\tag{2.93}$$

provided that the expectation exists and is finite.
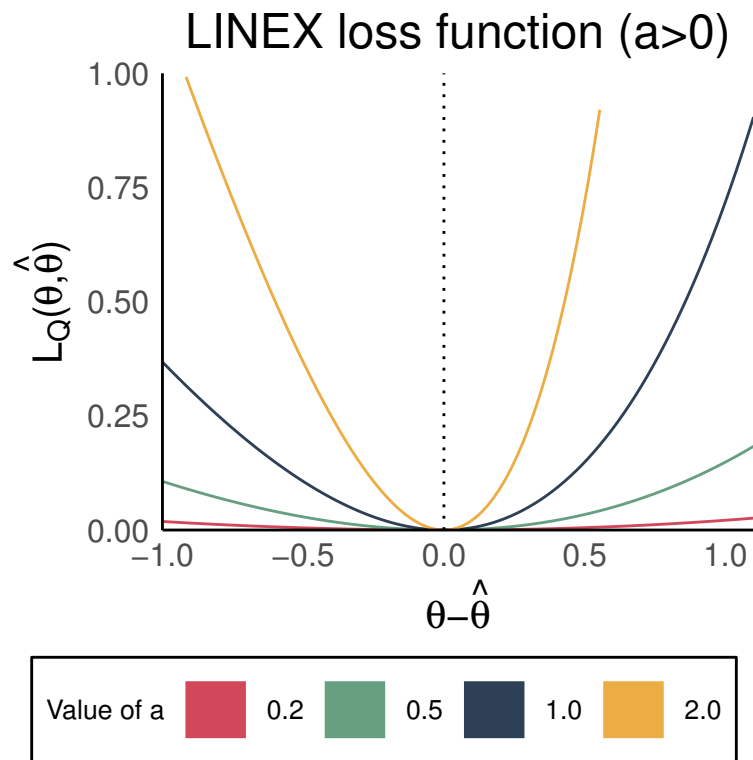
Figure 2.15: Figure illustrating the LINEX loss function $(a > 0)$.
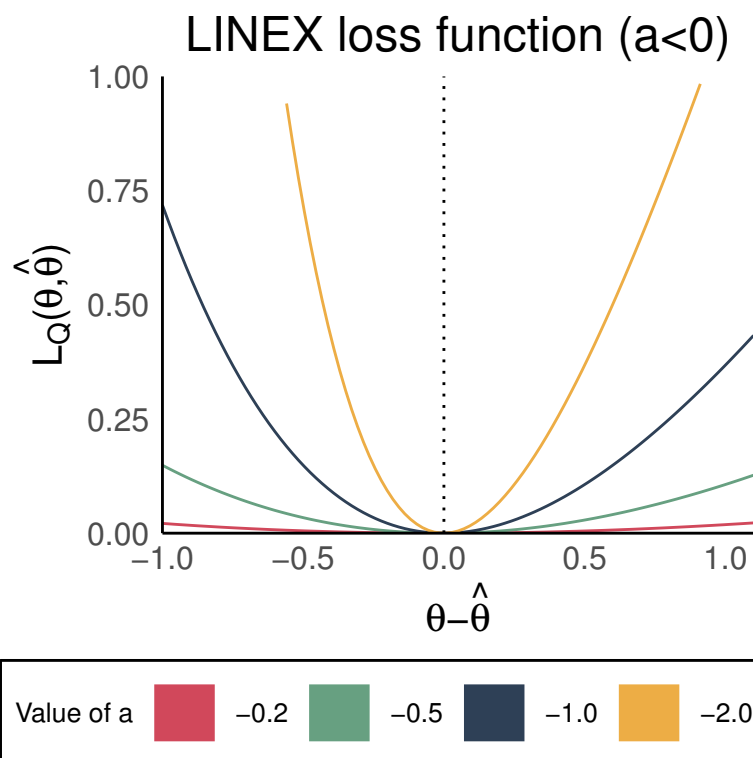
Figure 2.16: Figure illustrating the LINEX loss function $(a < 0)$.

2.7.2.3 General entropy loss function

Calabria and Pulcini (1994) introduced the asymmetric GELF as a loss function expressed

in terms of the ratio of θ and $\hat{\theta}$ given by:

$$L_G\left(\theta,\hat{\theta}\right) = \left(\frac{\hat{\theta}}{\theta}\right)^k - k\log\left(\frac{\hat{\theta}}{\theta}\right) - 1, \tag{2.94}$$

where: $k \neq 0$.

Like the LINEX loss function, when:

- $k > 0$, overestimation is deemed more serious than underestimation,

- $k < 0$, underestimation is deemed more serious than overestimation and when

- $k \approx 0$ the loss function is approximately symmetric.

The Bayes estimator under the GELF is given by:

$$\hat{\theta}_G = \left(E_\theta\left(\theta^{-k}\right)\right)^{-\frac{1}{k}}, \tag{2.95}$$

provided that the expectation exists and is finite.
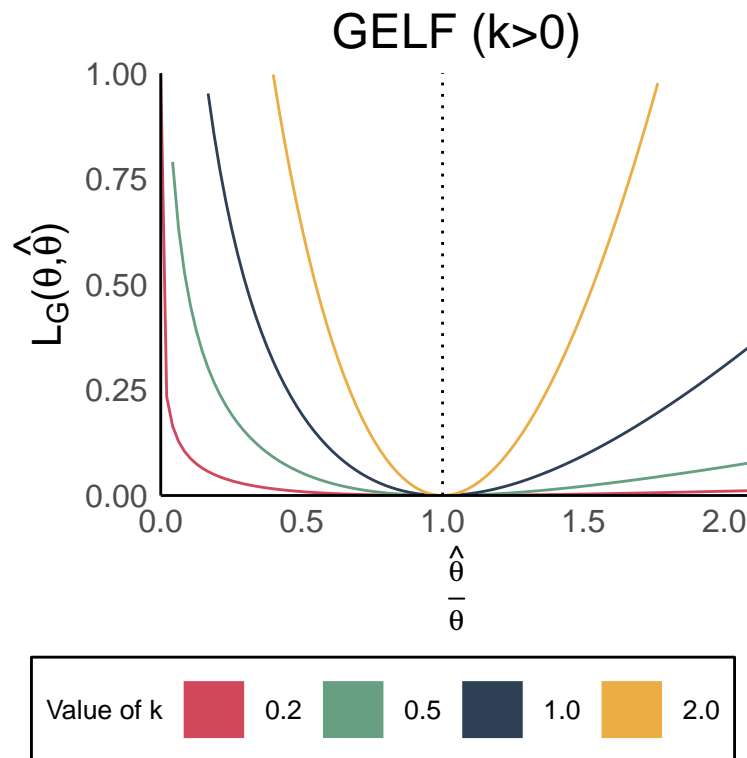
Figure 2.17: Figure illustrating the GELF $(k > 0)$.
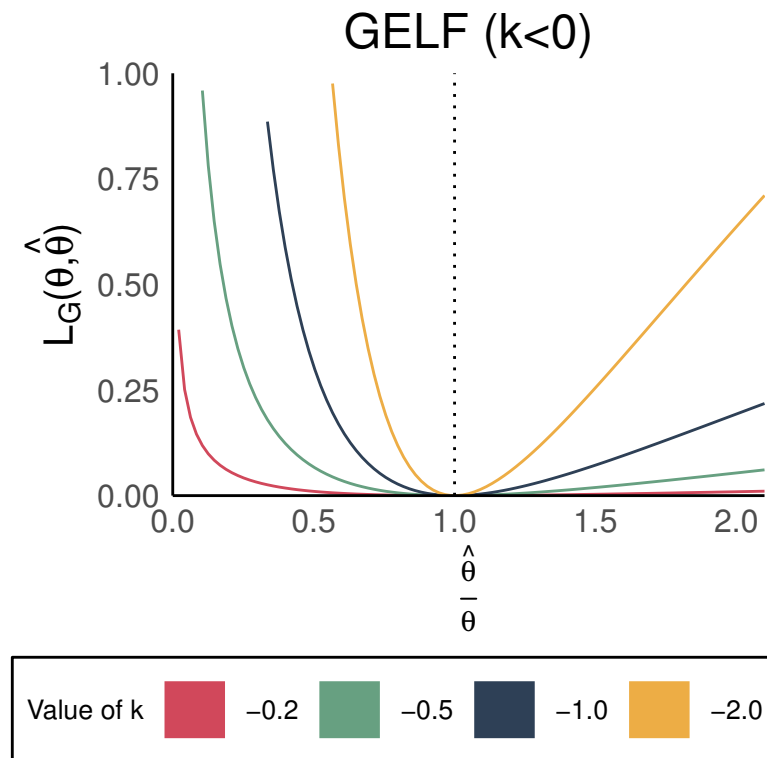
Figure 2.18: Figure illustrating the GELF $(k < 0)$.

2.8 Life distributions

This section introduces two probability models that are helpful in modeling reliability data – although naturally any distribution of non-negative random variables can be used to describe time. The two life distributions mentioned are the:

- exponential and

- Weibull distribution.

2.8.1 Exponential distribution

The exponential distribution is a well used life distribution due to its simplicity. The time-to-failure T is said to have an exponential distribution with rate parameter $\theta > 0$ if the PDF is given by:

$$f\left(t \mid \theta\right) = \begin{cases} \frac{1}{\theta} exp\left\{-\frac{t}{\theta}\right\} & t \geq 0 \\ 0 & t < 0. \end{cases} \tag{2.96}$$

The model above will be denoted in the rest of this thesis by: $T_i \mid \theta \sim Exponential\left(\theta\right)$.

The corresponding cumulative distribution function is:

$$F(t \mid \theta) = \begin{cases} 1 - exp\left\{-\frac{t}{\theta}\right\} & t \geq 0 \\ 0 & t < 0, \end{cases} \qquad (2.97)$$

the reliability function is given by:

$$R(t \mid \theta) = \begin{cases} exp\left\{-\frac{t}{\theta}\right\} & t \geq 0 \\ 0 & t < 0, \end{cases} \qquad (2.98)$$

and the hazard function is given by:

$$\lambda(t \mid \theta) = \begin{cases} \frac{1}{\theta} & t \geq 0 \\ 0 & t < 0. \end{cases} \qquad (2.99)$$

Exponential plots

PDF plot

CDF plot

Reliability plot

Hazard plot

Rate parameter: θ=0.5 θ=1.0 θ=1.5 θ=2.0
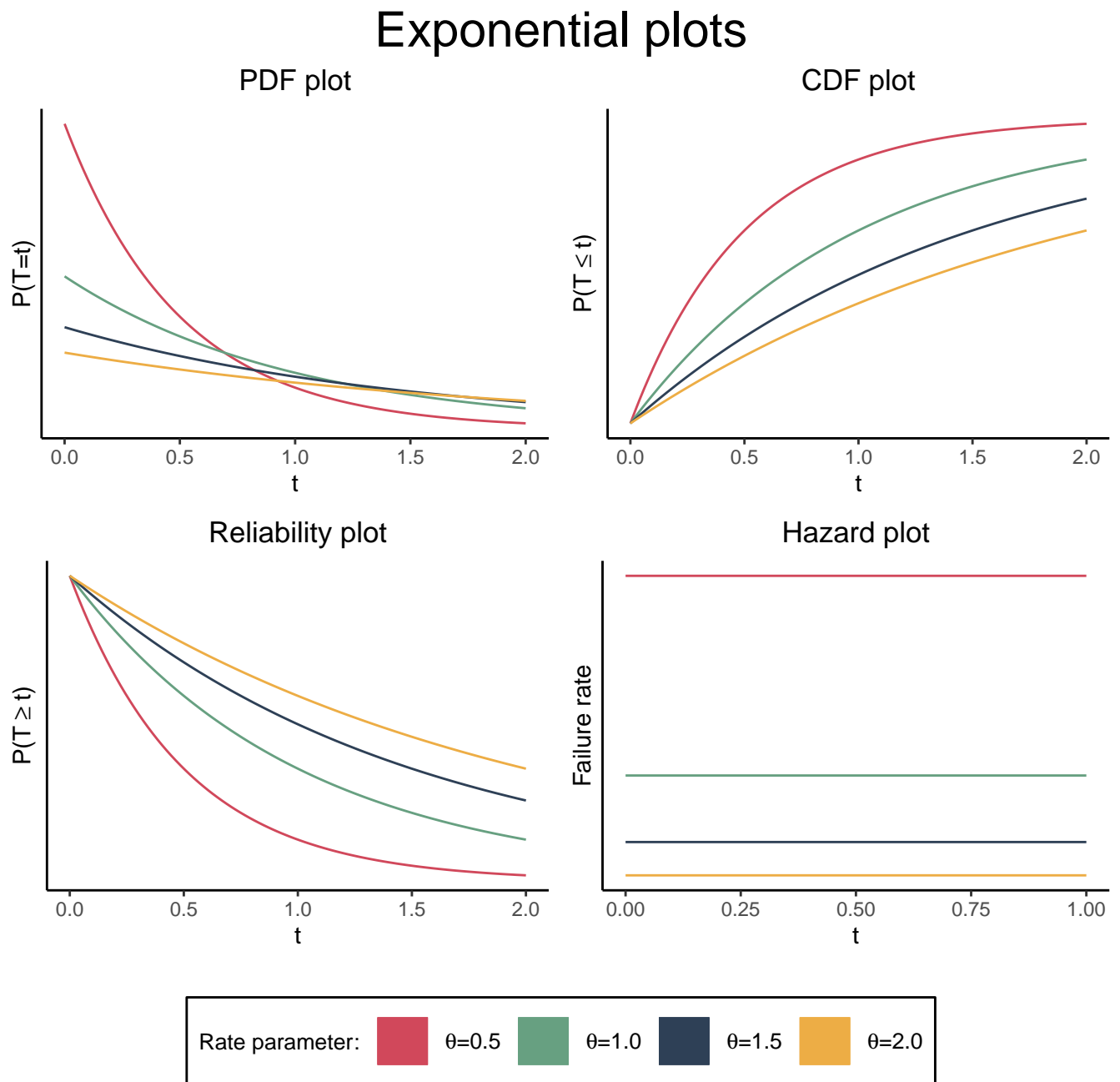
Figure 2.19: Plot showing the various functions of the exponential distribution under different rate parameter values.

The hazard function is constant regardless of the rate parameter, which highlights potential issues when using the distribution.

The expected value of the exponential is given by:

$$E\left(t\right) = \theta, \tag{2.100}$$

and the variance is given by:

$$Var\left(t\right) = \theta^2 \tag{2.101}$$

(Nelson, 1990). Other parameterizations of the exponential distribution are possible. A common parameterization of the exponential distribution is formed by setting the rate parameter $\theta = \frac{1}{\lambda}$, which is used predominantly used for items with high reliability (Nelson, 1990). This parameterization of the exponential distribution results in a CDF of:

$$F\left(t \mid \theta\right) = \begin{cases} 1 - exp\left\{-\lambda t\right\} & t \geq 0 \\ 0 & t < 0. \end{cases} \tag{2.102}$$

The two exponential models presented are equivalent, however depending on the context a particular parameterization might be more appropriate than the other.

2.8.2 Weibull distribution

The Weibull distribution was eponymously named after the Swedish physicist Weibull (1939), who used it extensively in modeling the strength of materials. Since then, the Weibull has become one of the most cited and important distributions in statistical literature. A compendium of uses and applications of the Weibull distribution is presented by Rinne (2008). The time-to-failure T is said to have a Weibull distribution with parameters θ and β, if the PDF is given by:

$$f\left(t \mid \theta, \beta\right) = \begin{cases} \frac{\beta t^{\beta-1}}{\theta^{\beta}} exp\left\{-\left(\frac{t}{\theta}\right)^{\beta}\right\} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{2.103}$$

where $\theta > 0$ is the scale parameter and $\beta > 0$ is the shape parameter. The model above will be denoted in the rest of this thesis by: $T_i \mid \theta, \beta \sim Weibull\left(\theta, \beta\right).$ When $\beta = 1$ the Weibull distribution is identical the exponential distribution with rate parameter θ.

The corresponding cumulative distribution function is:

$$F\left(t \mid \theta, \beta\right) = \begin{cases} 1 - exp\left\{-\left(\frac{t}{\theta}\right)^{\beta}\right\} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{2.104}$$

the reliability function is given by:

$$R\left(t \mid \theta, \beta\right) = \begin{cases} exp\left\{-\left(\frac{t}{\theta}\right)^{\beta}\right\} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{2.105}$$

and the hazard function is given by:

$$\lambda\left(t \mid \theta, \beta\right) = \begin{cases} \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{2.106}$$

The Weibull distribution is used significantly in reliability analysis due to its ability to reflect

different failure rates. When $\beta > 1$ it implies that the failure rate is a increasing function, when $\beta < 1$ it implies that the failure rate is a decreasing function and when $\beta = 1$ it implies that the failure rate is a constant function.
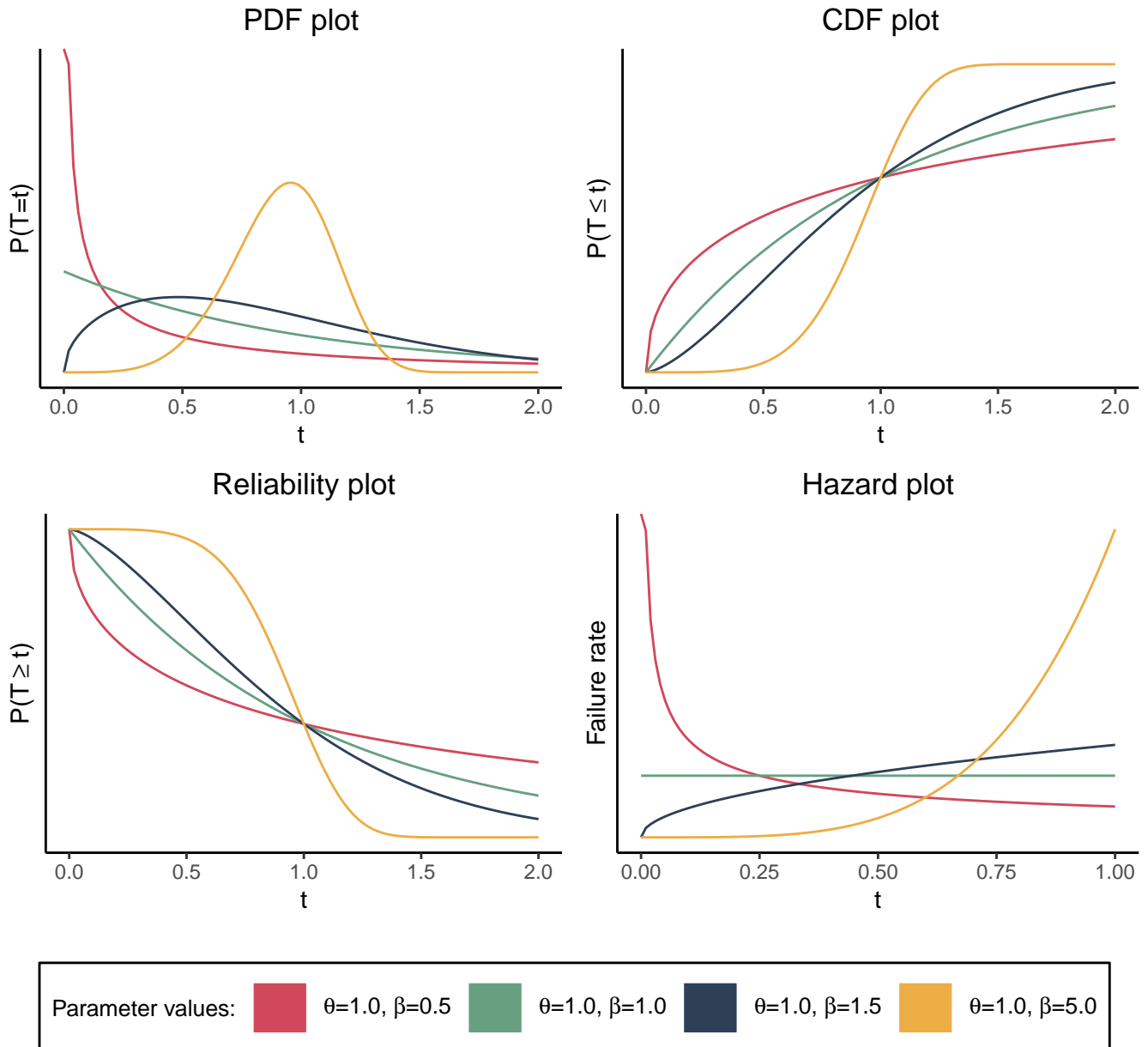
Figure 2.20: Plot showing the various functions of the Weibull distribution under different parameter values.

The expected value of the Weibull distribution is given by:

$$E\left(t\right) = \theta \Gamma \left(1 + \frac{1}{\beta}\right), \tag{2.107}$$

where $\Gamma(x)$ is the gamma function defined by:

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du = (x-1)!, \tag{2.108}$$

and the variance is given by:

$$Var(t) = \theta^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \left(\Gamma\left(1 + \frac{1}{\beta}\right)\right)^2 \right] \tag{2.109}$$

(Nelson, 1990). Other parameterizations of the Weibull distribution are possible. A common parameterization of the Weibull distribution is formed by setting the scale parameter $\theta = \frac{1}{\lambda}$, analogous in a reliability context as "failures per unit of time." This parameterization of the Weibull results in a CDF of:

$$F(t \mid \lambda, \beta) = \begin{cases} 1 - exp\left\{ -(\lambda t)^\beta \right\} & t \geq 0 \\ 0 & t < 0. \end{cases} \tag{2.110}$$

The two Weibull models presented are equivalent, however depending on the context a particular parameterization might be more appropriate than the other (Murthy et al., 2004).

2.8.2.1 Testing the Weibull assumption

Before doing any statistical analysis with the Weibull distribution, tests should be conducted to determine if the underlying data is indeed Weibull. A graphical method to assess the assumption is by plotting a Weibull probability plot (WPP).

A WPP is a plot of the empirical CDF, $\hat{F}(t)$ of the Weibull distribution on special scales such that if the underlying data is indeed Weibull the points will be linear (Rinne, 2008). The values for $\hat{F}(t)$ are estimated by either the mean plotting position, given by:

$$\hat{F}(t_i) \approx \frac{i}{n},$$

where i is the rank of the dataset and n is the sample size, or by the median plotting position given by:

$$\hat{F}(t_i) \approx \frac{i - 0.3}{n + 0.5}.$$

According to Rinne (2008) a WPP's advantages are:

- it is fast and simple to use with sufficient accuracy,

- it presents data in a easily understandable form, helping the researcher to make conclusions and

- helps to spot unusual data and outliers.

Rinne (2008) presents a method to construct a WPP for complete datasets:

1. Arrange the data in ascending order: $t_1 \leq t_2 \leq ... \leq t_n$.

2. Convert the data to logarithms: $x_i = log\left(t_i\right), 1 \leq i \leq n$.

3. Compute: $y_i = log\left(-log\left(1 - \hat{F}\left(t_i\right)\right)\right).$

4. Plot y_i on the ordinate against x_i on the abscissa.

5. Determine the best straight line by either a regression or least-squares method.

6. Claim the dataset comes from a Weibull distribution if the points scatter around the straight line.

However this method is subjective, hence for more formal results Murthy et al. (2004) suggests performing:

- Chi-Square tests,

- Kolmogorov-Smirnov (KS) tests or

- Cramèr-von Mises tests.

Chapter 3

Derivations for Weibull distribution

3.1   Introduction

This chapter introduces an ALT Weibull distribution, where the scale parameter is a log-linear function of stress and the shape parameter is constant at each level of stress. Two forms of estimates are derived for this model: maximum likelihood and non-informative Bayesian estimates. Moreover, five types of non-informative Bayesian priors are considered, namely: Jeffreys' prior, reference priors, the MDI prior, the uniform prior and PMPs, and thus the relevant posterior distributions are constructed. Since the posterior distributions under all the non-informative priors have an unknown form, the properness of each posterior is considered before any inference is made later in this thesis.

3.2   Log-linear ALT Weibull model

Assume that the life of a unit T_i follows a Weibull distribution with scale parameter $\nu\left(S_i\right) > 0$ dependent on some function of stress and shape parameter $\beta > 0$ under a k-level constant-stress ALT. At each stress level S_i, n_i units are tested until failure. That is, a complete dataset is considered for this model.

Let: $n = n_1 + n_2 + .. + n_k$ be the total number of units under consideration, and let S_0 denote the use stress level of the units. Furthermore denote the vector of observations under stress level of i as: $\boldsymbol{t_i} = \left(t_{i1}, t_{i2}, ..., t_{in_i}\right)'$, and the vector of all observations as: $\boldsymbol{t} = \left(\boldsymbol{t_1}, \boldsymbol{t_2}, ..., \boldsymbol{t_k}\right)'$.

Therefore under S_0 and accelerated stress S_i, $i = 1, 2, ..., k$, the PDF of the Weibull distribution under consideration is given by:

$$f_i\left(t_i \mid \beta_i, \nu\left(S_i\right)\right) = \begin{cases} \frac{\beta_i t_i^{\beta_i - 1}}{\nu(S_i)^{\beta_i}} exp\left\{-\left(\frac{t_i}{\nu(S_i)}\right)^{\beta_i}\right\} & t_i \geq 0 \\ 0 & t_i < 0. \end{cases} \tag{3.1}$$

Assume that the scale parameter is the log-linear time transformation function, and is given

63

by:

$$log\left(\nu\left(S_i\right)\right) = \theta_1 + \theta_2 \times \mu\left(S_i\right), \tag{3.2}$$

where: $\theta_1 > 0$ and $\theta_2 > 0$ are unknown parameters which need to be estimated and $\mu\left(S_i\right)$ is some decreasing-function of stress at stress level S_i which needs to be chosen appropriately depending on the underlying physics of failure.

The failure mechanism is assumed constant at each stress level, that is $\beta_0 = \beta_1 = ... = \beta_k = \beta$. Assuming conditional independence of the failure times t_{ij} given stress level S_i and the parameters β and $\nu\left(S_i\right)$, the likelihood function is given by:

$$L\left(\beta, \nu\left(S_i\right) \mid \boldsymbol{t}\right) = \prod_{j=1}^{n_i} \frac{\beta t_{ij}^{\beta-1}}{exp\left\{\beta\left(\theta_1 + \theta_2\mu\left(S_i\right)\right)\right\}} exp\left\{-\frac{t_{ij}^{\beta}}{exp\left\{\beta\left(\theta_1 + \theta_2\mu\left(S_i\right)\right)\right\}}\right\}. \tag{3.3}$$

The likelihood function under k stress levels is thus given by:

$$L\left(\beta, \nu\left(S_i\right) \mid \boldsymbol{t}\right) = \prod_{i=1}^{k} \frac{\beta^{n_i}}{exp\left\{\beta n_i\left(\theta_1 + \theta_2\mu\left(S_i\right)\right)\right\}} \prod_{j=1}^{n_i} t_{ij}^{\beta-1} exp\left\{-\frac{\sum\limits_{j=1}^{n_i} t_{ij}^{\beta}}{exp\left\{\beta\left(\theta_1 + \theta_2\mu\left(S_i\right)\right)\right\}}\right\}. \tag{3.4}$$

Following the method of Xu et al. (2015), define a transformation of the scale parameter given by:

$$\begin{aligned} \lambda_i &= \frac{1}{\nu\left(S_i\right)}, \ i = 1, 2, ..., k \\ &= exp\left\{-\left[\theta_1 + \theta_2\mu\left(S_i\right)\right]\right\} \\ &= \lambda_0 \times exp\left\{\theta_2\left[\mu\left(S_0\right) - \mu\left(S_i\right)\right]\right\} \\ &= \lambda_0 \times \eta^{\delta_i}, \end{aligned} \tag{3.5}$$

where:

- λ_0 denotes the use stress level,

- $\eta = exp\left\{\theta_2\left(\mu\left(S_0\right) - \mu\left(S_1\right)\right)\right\} = \frac{\lambda_1}{\lambda_0}$ denotes the acceleration factor from S_0 to S_1 and

- $\delta_i = \frac{\mu(S_0)-\mu(S_i)}{\mu(S_0)-\mu(S_1)}, \ i = 1, 2, ..., k.$

Since $\mu\left(S_i\right)$ is a decreasing function, it implies that $\delta_k > \delta_{k-1} > ... > \delta_1 = 1$. The new parameters chosen for the model are: λ_0, η and β where the inverse transformations are given as follows:

- $\lambda_0 = exp\left\{-\left[\theta_1 + \theta_2\mu\left(S_0\right)\right]\right\} \iff \theta_1 = -log\left(\lambda_0\right) - \frac{log(\eta)\mu(S_0)}{\mu(S_0)-\mu(S_1)}$

- $\eta = exp\left\{\theta_2\left[\mu\left(S_0\right) - \mu\left(S_1\right)\right]\right\} \iff \theta_2 = \frac{log(\eta)}{\mu(S_0)-\mu(S_1)}$

- $\beta = \beta \iff \beta = \beta.$

And hence the Jacobian matrix denoted by J is given by:

$$
J = \begin{bmatrix} \frac{-1}{\lambda_0} & \frac{\mu(S_0)}{\eta[\mu(S_0)-\mu(S_1)]} & 0 \\ 0 & \frac{1}{\eta[\mu(S_0)-\mu(S_1)]} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.6}
$$

where the determinant of this matrix is: $det(J) = \frac{-1}{\lambda_0\eta[\mu(S_0)-\mu(S_1)]}$. Therefore this transformation is one-to-one. By defining $\boldsymbol{\psi}' = (\lambda_0, \eta, \beta)'$, Xu et al. (2015) claim that likelihood function given in equation 3.4 may be rewritten as:

$$
L(\boldsymbol{\psi} \mid \boldsymbol{t}) \propto \beta^n \lambda_0^{n\beta} \eta^{\beta\bar{\delta}} \prod_{i=1}^{k}\prod_{j=1}^{n_i} t_{ij}^{\beta-1} exp\left\{-\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta}\eta^{\beta\delta_i}\right\}, \tag{3.7}
$$

where: $\bar{\delta} = \sum_{i=1}^{k} n_i\delta_i$.

Note that under this transformation $T_{ij} \mid \nu(S_i)^{\star}, \beta^{\star} \sim Weib\left(\nu(S_i)^{\star} = \lambda_0\eta^{\delta_i}, \beta^{\star} = \beta\right)$, which is the alternative parameterization of the Weibull distribution as defined by equation 2.110. Let $\mathcal{L} = log\left(L(\boldsymbol{\psi} \mid \boldsymbol{t})\right)$ be the log-likelihood of equation 3.7, which is given by:

$$
\mathcal{L} = nlog(\beta) + (n\beta)log(\lambda_0) + \left(\bar{\delta}\beta\right)log(\eta) +
$$
$$
+ (\beta-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i} log(t_{ij}) - \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta}\eta^{\beta\delta_i}. \tag{3.8}
$$

3.3 Maximum likelihood estimation

The method of maximum likelihood is perhaps the most popular method for deriving estimates. To find these estimates, consider the first-order partial derivatives of the log-likelihood function given in equation 3.8:

$$
\frac{\partial\mathcal{L}}{\partial\lambda_0} = \frac{n\beta}{\lambda_0} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\beta\lambda_0^{\beta-1}\eta^{\beta\delta_i}, \tag{3.9}
$$

$$
\frac{\partial\mathcal{L}}{\partial\eta} = \frac{\bar{\delta}\beta}{\eta} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta}\beta\delta_i\eta^{\beta\delta_i-1}, \tag{3.10}
$$

$$
\frac{\partial\mathcal{L}}{\partial\beta} = \frac{n}{\beta} + nlog(\lambda_0) + \bar{\delta}log(\eta) + \sum_{i=1}^{k}\sum_{j=1}^{n_i} log(t_{ij}) -
$$
$$
- \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta}\eta^{\beta\delta_i}log\left(t_{ij}\lambda_0\eta^{\delta_i}\right). \tag{3.11}
$$

The MLEs of the Weibull distribution are found by setting these partial derivatives equal to zero and solving for the parameter. However given that these equations are non-linear, some

iterative procedure will be required to approximate the estimates.

For simplicity, note that from equation 3.9, the estimate of λ_0 can be found by:

$$\hat{\lambda}_0 = \left(\frac{n}{\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} t_{ij}^{\hat{\beta}} \hat{\eta}^{\hat{\beta}\delta_i}} \right)^{\frac{1}{\beta}}. \tag{3.12}$$

Therefore the log-likelihood of $\boldsymbol{\psi}^\star = \lambda_0\left(\beta, \eta\right)$ given in equation 3.8 may be written as:

$$\mathcal{L}^\star = nlog\left(\beta\right) + nlog\left(n\right) - nlog\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta} \eta^{\beta\delta_i} \right) + \\ + \left(\bar{\delta}\beta\right) log\left(\eta\right) + \left(\beta - 1\right) \sum_{i=1}^{k}\sum_{j=1}^{n_i} log\left(t_{ij}\right) - n. \tag{3.13}$$

Finding the first- and second-order partial derivatives of equation 3.13 yield:

$$\frac{\partial \mathcal{L}^\star}{\partial \beta} = \frac{n}{\beta} + \bar{\delta}log\left(\eta\right) + \sum_{i=1}^{k}\sum_{j=1}^{n_i} log\left(t_{ij}\right) - \frac{n \sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} t_{ij}^{\beta} \eta^{\beta\delta_i} log\left(t_{ij}\eta^{\delta_i}\right)}{\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} t_{ij}^{\beta}\eta^{\beta\delta_i}}, \tag{3.14}$$

$$\frac{\partial \mathcal{L}^\star}{\partial \eta} = \frac{\bar{\delta}\beta}{\eta} - \frac{n \sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} \beta\delta_i t_{ij}^{\beta}\eta^{\beta\delta_i-1}}{\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} t_{ij}^{\beta}\eta^{\beta\delta_i}}, \tag{3.15}$$

$$\frac{\partial^2 \mathcal{L}^\star}{\partial \beta^2} = -\frac{n}{\beta^2} - \frac{n}{\left(\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} t_{ij}^{\beta}\eta^{\beta\delta_i}\right)^2} \left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\eta^{\beta\delta_i} \left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\eta^{\beta\delta_i} log\left(t_{ij}\eta^{\delta_i}\right) \left(log\left(t_{ij}\eta^{\delta_i}\right)\right) \right] - \ldots \right. \\ \left. - \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\eta^{\beta\delta_i} log\left(t_{ij}\eta^{\delta_i}\right) \left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\eta^{\beta\delta_i} \left(log\left(t_{ij}\eta^{\delta_i}\right)\right) \right] \right), \tag{3.16}$$

$$\frac{\partial^2 \mathcal{L}^\star}{\partial \eta^2} = -\frac{\bar{\delta}\beta}{\eta} - \frac{n}{\left(\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} t_{ij}^{\beta}\eta^{\beta\delta_i}\right)^2} \left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\eta^{\beta\delta_i} \left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} \beta\delta_i\left(\beta\delta_i - 1\right) t_{ij}^{\beta}\eta^{\beta\delta_i-2} \right] - \ldots \right. \\ \left. - \left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} \beta\delta_i t_{ij}^{\beta}\eta^{\beta\delta_i-1} \right)^2 \right), \tag{3.17}$$

$$\frac{\partial^2 \mathcal{L}^\star}{\partial \beta \partial \eta} = \frac{\bar{\delta}}{\eta} + \frac{n}{\left(\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i}\right)^2} \left(\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i} \left[\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i - 1} \delta_i \left(\beta log \left(t_{ij}\eta^{\delta_i} \right) + 1 \right) \right] - ... \right.$$

$$\left. - \sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i} log \left(t_{ij}\eta^{\delta_i} \right) \left[\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} \beta\delta_i t_{ij}^\beta \eta^{\beta\delta_i - 1} \right] \right).$$

(3.18)

Hence the gradient vector of equation 3.13 is given by:

$$\nabla' \left(\boldsymbol{\psi}^\star \right) = \left[\frac{\partial \mathcal{L}^\star}{\partial \beta}, \frac{\partial \mathcal{L}^\star}{\partial \eta} \right]'$$

(3.19)

and the associated Hessian matrix is given by:

$$H \left(\boldsymbol{\psi}^\star \right) = \begin{bmatrix} \frac{\partial^2 \mathcal{L}^\star}{\partial \beta^2} & \frac{\partial^2 \mathcal{L}^\star}{\partial \beta \partial \eta} \\ \frac{\partial^2 \mathcal{L}^\star}{\partial \beta \partial \eta} & \frac{\partial^2 \mathcal{L}^\star}{\partial \eta^2} \end{bmatrix}.$$

(3.20)

Therefore the NR algorithm as described in section 2.4.1 may be used to approximate the MLEs.

To find the variance-covariance matrix of the MLEs, the Fisher information matrix needs to be derived.

Theorem 3.1. *The Fisher information matrix for the Weibull distribution under the log-linear time transformation function is given by:*

$$I \left(\boldsymbol{\psi} \right) = \begin{bmatrix} \frac{n\beta^2}{\lambda_0^2} & \frac{\bar{\delta}\beta^2}{\lambda_0 \eta} & \frac{c_1}{\lambda_0} \\ \frac{\bar{\delta}\beta^2}{\lambda_0 \eta} & \frac{\beta^2 c_4}{\eta^2} & \frac{c_3}{\eta} \\ \frac{c_1}{\lambda_0} & \frac{c_3}{\eta} & \frac{n+c_2}{\beta^2} \end{bmatrix},$$

(3.21)

where:

- $c_1 = \sum\limits_{i=1}^{k} c_{1i} = n \left(-\gamma + 1 \right),$

- $c_2 = \sum\limits_{i=1}^{k} c_{2i} = n \left(2\gamma + \gamma^2 + \frac{\pi^2}{6} \right),$

- $c_3 = \sum\limits_{i=1}^{k} c_{1i}\delta_i = \bar{\delta} \left(-\gamma + 1 \right),$

- $c_4 = \sum\limits_{i=1}^{k} n_i \delta_i^2$ *and*

$\gamma \approx 0.5772$ *is Euler's constant (Abramowitz and Stegun, 1964).*

Proof. The proof is provided in Xu et al. (2015). For working out, see appendix A.1. $\quad\square$

And therefore the inverse of the Fisher information matrix is given by:

$$I^{-1}\left(\boldsymbol{\psi}\right) = \frac{1}{M} \begin{bmatrix} \frac{\left(nc_4 + c_2c_4 - c_3^2\right)\lambda_0^2}{\beta^2} & \frac{\left(c_1c_3 - n\bar{\delta} - c_2\bar{\delta}\right)\lambda_0\eta}{\beta^2} & \left(c_3\bar{\delta} - c_1c_4\right)\lambda_0 \\ \frac{\left(c_1c_3 - n\bar{\delta} - c_2\bar{\delta}\right)\lambda_0\eta}{\beta^2} & \frac{\left(n^2 + nc_2 - c_1^2\right)\eta^2}{\beta^2} & \left(c_1\bar{\delta} - nc_3\right)\eta \\ \left(c_3\bar{\delta} - c_1c_4\right)\lambda_0 & \left(c_1\bar{\delta} - nc_3\right)\eta & \left(\bar{\delta}^2 - nc_4\right)\beta^2 \end{bmatrix},$$ (3.22)

where: $M = c_4n^2 + n\left(c_2c_4 - c_3^2 - \bar{\delta}^2\right) + 2\bar{\delta}c_1c_3 - \bar{\delta}^2 - c_1^2c_4$.

For simplicity, the inverse of the Fisher information matrix given in equation 3.22 will be written as:

$$I^{-1}\left(\boldsymbol{\psi}\right) = \frac{1}{M} \begin{bmatrix} \frac{a_{11}\lambda_0^2}{\beta^2} & \frac{a_{12}\lambda_0\eta}{\beta^2} & a_{13}\lambda_0 \\ \frac{a_{21}\lambda_0\eta}{\beta^2} & \frac{a_{22}\eta^2}{\beta^2} & a_{23}\eta \\ a_{31}\lambda_0 & a_{32}\eta & a_{33}\beta^2 \end{bmatrix},$$ (3.23)

where $a_{ij}\, i, j = 1, 2, 3$ are the corresponding multipliers of the parameters left in the matrix given in equation 3.22.

Hence, via the properties described in section 2.4.1.1 the MLEs have a variance-covariance matrix given by:

$$\Sigma = I^{-1}\left(\boldsymbol{\psi}\right) = \begin{bmatrix} var\left(\hat{\lambda}_0\right) & cov\left(\hat{\lambda}_0, \hat{\eta}\right) & cov\left(\hat{\lambda}_0, \hat{\beta}\right) \\ cov\left(\hat{\lambda}_0, \hat{\eta}\right) & var\left(\hat{\eta}\right) & cov\left(\hat{\eta}, \hat{\beta}\right) \\ cov\left(\hat{\lambda}_0, \hat{\beta}\right) & cov\left(\hat{\eta}, \hat{\beta}\right) & var\left(\hat{\beta}\right) \end{bmatrix}.$$ (3.24)

3.4 Non-informative Bayesian derivations

3.4.1 Jeffreys' prior

Jeffreys' prior is the most discussed prior in non-informative Bayesian analysis and is described in section 2.4.5.1. The prior is given by:

$$\pi_J\left(\boldsymbol{\psi}\right) \propto \left[det\left(I\left(\boldsymbol{\psi}\right)\right)\right]^{\frac{1}{2}},$$ (3.25)

where the Fisher information matrix is given in equation 3.21.

Theorem 3.2. *Jeffreys' prior for the Weibull distribution under the log-linear time transformation function is given by:*

$$\pi_J\left(\boldsymbol{\psi}\right) \propto \frac{\beta}{\lambda_0\eta}.$$ (3.26)

Proof. Finding the determinant of the Fisher information matrix defined in equation A.25

yields:

$$
\begin{aligned}
det\left(I\left(\boldsymbol{\psi}\right)\right) &= \frac{n\beta^2}{\lambda_0^2} \times \begin{vmatrix} \frac{\beta^2 c_4}{\eta^2} & \frac{c_3}{\eta} \\ \frac{c_3}{\eta} & \frac{n+c_2}{\beta^2} \end{vmatrix} - \frac{\bar{\delta}\beta^2}{\lambda_0\eta} \times \begin{vmatrix} \frac{\bar{\delta}\beta^2}{\lambda_0\eta} & \frac{c_3}{\eta} \\ \frac{c_1}{\lambda_0} & \frac{n+c_2}{\beta^2} \end{vmatrix} + \frac{c_1}{\lambda_0} \times \begin{vmatrix} \frac{\bar{\delta}\beta^2}{\lambda_0\eta} & \frac{\beta^2 c_4}{\eta^2} \\ \frac{c_1}{\lambda_0} & \frac{c_3}{\eta} \end{vmatrix} \\
&= \frac{n\beta^2}{\lambda_0^2} \times \left[\frac{c_4\beta^2\left(n+c_2\right)}{\eta^2\beta^2} - \frac{c_3^2}{\eta^2}\right] - \frac{\bar{\delta}\beta^2}{\lambda_0\eta} \times \left[\frac{\bar{\delta}\beta^2\left(n+c_2\right)}{\lambda_0\eta\beta^2} - \frac{c_3 c_1}{\eta\lambda_0}\right] + ... \\
&\quad + \frac{c_1}{\lambda_0} \times \left[\frac{\bar{\delta}\beta^2 c_3}{\lambda_0\eta^2} - \frac{c_1 c_4\beta^2}{\lambda_0\eta^2}\right] \\
&= \frac{\beta^2}{\lambda_0^2\eta^2} \times \left[c_4 n^2 + n c_4 c_2 - \bar{\delta}^2 n - \bar{\delta}^2 c_2 + \bar{\delta} c_3 c_1 + \bar{\delta} c_1 c_2 - c_1^2 c_4\right] \\
&= \frac{\beta^2 M}{\lambda_0^2\eta^2} \\
&\propto \frac{\beta^2}{\lambda_0^2\eta^2}.
\end{aligned}
\tag{3.27}
$$

Hence Jeffreys' prior is given by:

$$
\begin{aligned}
\pi_J\left(\boldsymbol{\psi}\right) &\propto \left[det\left(I\left(\boldsymbol{\psi}\right)\right)\right]^{\frac{1}{2}} \\
&= \frac{\beta}{\lambda_0\eta}.
\end{aligned}
\tag{3.28}
$$

$\square$

Therefore the posterior distribution of $\boldsymbol{\psi}$ under Jeffreys' prior is given by:

$$
\begin{aligned}
\pi_J\left(\boldsymbol{\psi} \mid \boldsymbol{t}\right) &= c_J^{-1} \times \pi_J\left(\boldsymbol{\psi}\right) \times L\left(\boldsymbol{\psi} \mid \boldsymbol{t}\right) \\
&\propto \pi_J\left(\boldsymbol{\psi}\right) \times L\left(\boldsymbol{\psi} \mid \boldsymbol{t}\right) \\
&= \frac{\beta}{\lambda_0\eta} \times \beta^n \lambda_0^{n\beta} \eta^{\beta\bar{\delta}} \prod_{i=1}^{k}\prod_{j=1}^{n_i} t_{ij}^{\beta-1} exp\left\{-\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta}\eta^{\beta\delta_i}\right\} \\
&= \beta^{n+1} \lambda_0^{n\beta-1} \eta^{\beta\bar{\delta}-1} \prod_{i=1}^{k}\prod_{j=1}^{n_i} t_{ij}^{\beta-1} exp\left\{-\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta}\eta^{\beta\delta_i}\right\},
\end{aligned}
\tag{3.29}
$$

where c_J is the normalizing constant, and is given by:

$$
c_J = \int\limits_{0}^{\infty}\int\limits_{0}^{\infty}\int\limits_{1}^{\infty} \beta^{n+1} \lambda_0^{n\beta-1} \eta^{\beta\bar{\delta}-1} \prod_{i=1}^{k}\prod_{j=1}^{n_i} t_{ij}^{\beta-1} exp\left\{-\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta}\eta^{\beta\delta_i}\right\} d\lambda_0 d\beta d\eta.
\tag{3.30}
$$

Since the posterior distribution under Jeffreys' prior is of an unknown form, the properness needs to be considered to ensure admissible results.

3.4.2   Reference prior

The reference prior approach is discussed in section 2.4.5.2, and is an alternative to Jeffreys' prior for multiparameter models. To compile a reference prior, the algorithm described in

section 2.4.5.2 needs to be followed.

Theorem 3.3. *The reference prior for the grouping order $\{\lambda_0, \eta, \beta\}$ is given by:*

$$\pi_{R1}(\boldsymbol{\psi}) \propto \frac{1}{\lambda_0 \eta \beta}. \tag{3.31}$$

Proof. The proof is provided in Xu et al. (2015). For working out see appendix A.2.1 $\qquad \square$

Theorem 3.4. *The reference prior for the grouping $\{\lambda_0, (\eta, \beta)\}$ is given by:*

$$\pi_{R2}(\boldsymbol{\psi}) \propto \frac{1}{\lambda_0 \eta}. \tag{3.32}$$

Proof. The proof is provided in Xu et al. (2015). For working out see appendix A.2.2 $\qquad \square$

Theorem 3.5. *The reference prior for the grouping order $\{\eta, (\lambda_0, \beta)\}$ is given by:*

$$\pi_{R2}(\boldsymbol{\psi}) \propto \frac{1}{\lambda_0 \eta}. \tag{3.33}$$

Proof. The proof can be found in appendix A.2.3. $\qquad \square$

Theorem 3.6. *The reference prior for the grouping $\{\beta, (\lambda_0, \eta)\}$ is given by:*

$$\pi_{R1}(\boldsymbol{\psi}) \propto \frac{1}{\lambda_0 \eta \beta}. \tag{3.34}$$

Proof. The proof can be found in appendix A.2.4. $\qquad \square$

Table 3.1 provides a summary of the possible reference priors derived above. Since Jeffreys' prior is *de facto* the reference prior when all parameters are of equal importance, it is also included in table 3.1.

Table 3.1: Possible reference priors dependent on the grouping of the parameters in $\boldsymbol{\psi}$.

Grouping order	Reference prior for $\boldsymbol{\psi}$
$\{(\lambda_0, \eta, \beta)\}$	$\pi_J(\boldsymbol{\psi}) = \frac{\beta}{\lambda_0 \eta}$
$\{\lambda_0, \eta, \beta\}$, $\{\beta, (\lambda_0, \eta)\}$, $\{(\lambda_0, \eta), \beta\}$	$\pi_{R1}(\boldsymbol{\psi}) = \frac{1}{\lambda_0 \eta \beta}$
$\{\lambda_0, (\eta, \beta)\}$, $\{(\eta, \beta), \lambda_0\}$, $\{\eta, (\lambda_0, \beta)\}$, $\{(\lambda_0, \beta), \eta\}$	$\pi_{R2}(\boldsymbol{\psi}) = \frac{1}{\lambda_0 \eta}$

For convenience, the general reference prior can be defined as:

$$\pi_R(\boldsymbol{\psi}) = \frac{1}{\lambda_0 \eta \beta^m}, \, m = -1, 0, 1. \tag{3.35}$$

Therefore the posterior distribution of $\boldsymbol{\psi}$ under the prior $\pi_R(\boldsymbol{\psi})$ is given by:

$$
\begin{aligned}
\pi_R(\boldsymbol{\psi} \mid \boldsymbol{t}) &= c_R^{-1} \times \pi_R(\boldsymbol{\psi}) \times L(\boldsymbol{\psi} \mid \boldsymbol{t}) \\
&\propto \pi_R(\boldsymbol{\psi}) \times L(\boldsymbol{\psi} \mid \boldsymbol{t}) \\
&= \frac{1}{\lambda_0 \eta \beta^m} \times \beta^n \lambda_0^{n\beta} \eta^{\beta\bar{\delta}} \prod_{i=1}^{k} \prod_{j=1}^{n_i} t_{ij}^{\beta-1} exp\left\{ -\sum_{i=1}^{k} \sum_{j=1}^{n_i} t_{ij}^{\beta} \lambda_0^{\beta} \eta^{\beta\delta_i} \right\} \\
&= \beta^{n-m-1} \lambda_0^{n\beta-1} \eta^{\beta\bar{\delta}-1} \prod_{i=1}^{k} \prod_{j=1}^{n_i} t_{ij}^{\beta-1} exp\left\{ -\sum_{i=1}^{k} \sum_{j=1}^{n_i} t_{ij}^{\beta} \lambda_0^{\beta} \eta^{\beta\delta_i} \right\},
\end{aligned}
\tag{3.36}
$$

where $m = -1, 0, 1$ and c_R is the normalizing constant, and is given by:

$$
c_R = \int_0^{\infty} \int_0^{\infty} \int_1^{\infty} \beta^{n-m-1} \lambda_0^{n\beta-1} \eta^{\beta\bar{\delta}-1} \prod_{i=1}^{k} \prod_{j=1}^{n_i} t_{ij}^{\beta-1} exp\left\{ -\sum_{i=1}^{k} \sum_{j=1}^{n_i} t_{ij}^{\beta} \lambda_0^{\beta} \eta^{\beta\delta_i} \right\} d\lambda_0 d\beta d\eta.
\tag{3.37}
$$

The posterior distribution under the general reference prior has an unknown form, therefore the properness needs to be considered to ensure admissible results.

3.4.3 Maximal data information prior

The MDI prior is described in section 2.4.5.3, and is given by:

$$
\pi_{MDI}(\boldsymbol{\psi}) \propto exp\left\{ \mathcal{H}(\boldsymbol{\psi}) \right\},
\tag{3.38}
$$

where: $\mathcal{H}(\boldsymbol{\psi}) = \int_0^{\infty} f(\boldsymbol{t} \mid \boldsymbol{\psi}) log(f(\boldsymbol{t} \mid \boldsymbol{\psi})) d\boldsymbol{t}$ is the negative Shannon entropy.

Theorem 3.7. *The MDI prior for the Weibull distribution under the log-linear time transformation function is given as:*

$$
\pi_{MDI}(\boldsymbol{\psi}) \propto \beta\left(\lambda_0 \eta^{\delta_i}\right) exp\left\{ \frac{\gamma}{\beta} \right\},
\tag{3.39}
$$

where: $\gamma \approx 0.5772$ is Euler's constant (Abramowitz and Stegun, 1964).

Proof. The PDF of the Weibull distribution under the log-linear time transformation function for any t_{ij} may be written as:

$$
f(t_{ij} \mid \boldsymbol{\psi}) = \beta\left(\lambda_0 \eta^{\delta_i}\right)^{\beta} t_{ij}^{\beta-1} exp\left\{ -\left(t_{ij} \lambda_0 \eta^{\delta_i}\right)^{\beta} \right\} \quad t_{ij} \geq 0.
\tag{3.40}
$$

Consider the negative Shannon entropy of the Weibull distribution under the log-linear time

transformation function, denoted by $\mathcal{H}(\boldsymbol{\psi})$:

$$
\begin{aligned}
\mathcal{H}(\boldsymbol{\psi}) &= \int_0^\infty f(t_{ij} \mid \boldsymbol{\psi}) \log(f(t_{ij} \mid \boldsymbol{\psi})) \, dt_{ij} \\
&= \int_0^\infty \beta \left(\lambda_0 \eta^{\delta_i}\right)^\beta t_{ij}^{\beta-1} exp\left\{-\left(\lambda_0 \eta^{\delta_i}\right)^\beta t_{ij}^\beta\right\} \left[\log(\beta) + \beta \log(\lambda_0) + \delta_i \beta \log(\eta) + \ldots\right. \\
&\qquad \left. + (\beta-1)\log(t_{ij}) - \left(t_{ij}\lambda_0 \eta^{\delta_i}\right)^\beta\right] dt_{ij} \\
&= \beta \left(\lambda_0 \eta^{\delta_i}\right)^\beta \log(\beta) \int_0^\infty t_{ij}^{\beta-1} exp\left\{-\left(\lambda_0 \eta^{\delta_i}\right)^\beta t_{ij}^\beta\right\} dt_{ij} + \ldots \\
&\quad + \beta \left(\lambda_0 \eta^{\delta_i}\right)^\beta \beta \log(\lambda_0) \int_0^\infty t_{ij}^{\beta-1} exp\left\{-\left(\lambda_0 \eta^{\delta_i}\right)^\beta t_{ij}^\beta\right\} dt_{ij} + \ldots \\
&\quad + \beta \left(\lambda_0 \eta^{\delta_i}\right)^\beta \delta_i \beta \log(\eta) \int_0^\infty t_{ij}^{\beta-1} exp\left\{-\left(\lambda_0 \eta^{\delta_i}\right)^\beta t_{ij}^\beta\right\} dt_{ij} + \ldots \\
&\quad + \beta \left(\lambda_0 \eta^{\delta_i}\right)^\beta (\beta-1) \int_0^\infty \log(t_{ij}) t_{ij}^{\beta-1} exp\left\{-\left(\lambda_0 \eta^{\delta_i}\right)^\beta t_{ij}^\beta\right\} dt_{ij} - \ldots \\
&\quad - \int_0^\infty \left(t_{ij}\lambda_0 \eta^{\delta_i}\right)^\beta \beta \left(\lambda_0 \eta^{\delta_i}\right)^\beta t_{ij}^{\beta-1} exp\left\{-\left(\lambda_0 \eta^{\delta_i}\right)^\beta t_{ij}^\beta\right\} dt_{ij} \\
&= \log(\beta) + \beta \log(\lambda_0) + \delta_i \beta \log(\eta) + A - B,
\end{aligned}
\tag{3.41}
$$

where:

$$
\begin{aligned}
A &= \beta \left(\lambda_0 \eta^{\delta_i}\right)^\beta (\beta-1) \int_0^\infty \log(t_{ij}) t_{ij}^{\beta-1} exp\left\{-\left(\lambda_0 \eta^{\delta_i}\right)^\beta t_{ij}^\beta\right\} dt_{ij}, \\
B &= \int_0^\infty \left(t_{ij}\lambda_0 \eta^{\delta_i}\right)^\beta \beta \left(\lambda_0 \eta^{\delta_i}\right)^\beta t_{ij}^{\beta-1} exp\left\{-\left(\lambda_0 \eta^{\delta_i}\right)^\beta t_{ij}^\beta\right\} dt_{ij}.
\end{aligned}
\tag{3.42}
$$

Consider A, making the substitution:

- $u_{ij} = \left(\lambda_0 \eta^{\delta_i}\right)^\beta t_{ij}^\beta \Longrightarrow du_{ij} = \beta \left(\lambda_0 \eta^{\delta_i}\right)^\beta t_{ij}^{\beta-1} dt_{ij},$

which gives:

$$
\begin{aligned}
A &= \frac{(\beta-1)}{\beta} \int_0^\infty \left[\log(u_{ij}) - \beta \log\left(\lambda_0 \eta^{\delta_i}\right)\right] exp\left\{-u_{ij}\right\} du_{ij} \\
&= \frac{(\beta-1)}{\beta} \left[-\gamma - \beta \log\left(\lambda_0 \eta^{\delta_i}\right)\right],
\end{aligned}
\tag{3.43}
$$

where:

$$-\gamma = \int\limits_0^\infty log\left(u_{ij}\right) exp\left\{-u_{ij}\right\} du_{ij}. \tag{3.44}$$

Consider B, making the transformation:

- $y_{ij} = t_{ij}^\beta,$

then $y_{ij} \sim Exp\left(\theta^\star = \left(\lambda_0\eta^{\delta_i}\right)^\beta\right).$
Thus:

$$\begin{aligned} B &= \left(\lambda_0\eta^{\delta_i}\right)^\beta \int\limits_0^\infty E\left[\sim Exp\left(\theta^\star = \left(\lambda_0\eta^{\delta_i}\right)^\beta\right)\right] dy_{ij} \\ &= \left(\lambda_0\eta^{\delta_i}\right)^\beta \left(\lambda_0\eta^{\delta_i}\right)^{-\beta} \\ &= 1. \end{aligned} \tag{3.45}$$

Therefore the MDI prior can be formed as:

$$\begin{aligned} \pi_{MDI}\left(\boldsymbol{\psi}\right) &= exp\left\{\mathcal{H}\left(\boldsymbol{\psi}\right)\right\} \\ &= exp\left\{log\left(\beta\right) + \beta log\left(\lambda_0\right) + \delta_i\beta log\left(\eta\right) + \frac{\left(\beta-1\right)}{\beta}\left[-\gamma - \beta log\left(\lambda_0\eta^{\delta_i}\right)\right] - 1\right\} \\ &= \beta\left(\lambda_0\eta^{\delta_i}\right)^\beta\left(\lambda_0\eta^{\delta_i}\right)^{-(\beta-1)} exp\left\{-\frac{\left(\beta-1\right)}{\beta}\gamma - 1\right\} \\ &\propto \beta\left(\lambda_0\eta^{\delta_i}\right) exp\left\{\frac{\gamma}{\beta}\right\}. \end{aligned} \tag{3.46}$$

$\square$

Therefore the posterior distribution of $\boldsymbol{\psi}$ under the MDI prior for k stress levels is given by:

$$\begin{aligned} \pi_{MDI}\left(\boldsymbol{\psi}\mid\boldsymbol{t}\right) &= c_{MDI}^{-1}\times\pi_{MDI}\left(\boldsymbol{\psi}\right)\times L\left(\boldsymbol{\psi}\mid\boldsymbol{t}\right) \\ &\propto \pi_{MDI}\left(\boldsymbol{\psi}\right)\times L\left(\boldsymbol{\psi}\mid\boldsymbol{t}\right) \\ &= \beta\lambda_0\eta^{\delta_i}exp\left\{\frac{\gamma}{\beta}\right\}\times\beta^n\lambda_0^{n\beta}\eta^{\beta\bar{\delta}}\prod_{i=1}^k\prod_{j=1}^{n_i}t_{ij}^{\beta-1}exp\left\{-\sum_{i=1}^k\sum_{j=1}^{n_i}t_{ij}^\beta\lambda_0^\beta\eta^{\beta\delta_i}\right\} \\ &= exp\left\{\frac{\gamma}{\beta}\right\}\beta^{n+1}\lambda_0^{n\beta+1}\eta^{\bar{\delta}(\beta+1)}\prod_{i=1}^k\prod_{j=1}^{n_i}t_{ij}^{\beta-1}exp\left\{-\sum_{i=1}^k\sum_{j=1}^{n_i}t_{ij}^\beta\lambda_0^\beta\eta^{\beta\delta_i}\right\}. \end{aligned} \tag{3.47}$$

where c_{MDI} is the normalizing constant, and is given by:

$$c_{MDI} = \int\limits_0^\infty\int\limits_0^\infty\int\limits_1^\infty exp\left\{\frac{\gamma}{\beta}\right\}\beta^{n+1}\lambda_0^{n\beta+1}\eta^{\bar{\delta}(\beta+1)}\prod_{i=1}^k\prod_{j=1}^{n_i}t_{ij}^{\beta-1}exp\left\{-\sum_{i=1}^k\sum_{j=1}^{n_i}t_{ij}^\beta\lambda_0^\beta\eta^{\beta\delta_i}\right\}d\lambda_0 d\beta d\eta. \tag{3.48}$$

The posterior distribution under the MDI prior is of an unknown form, therefore the properness of the posterior needs to be considered to ensure admissible results.

3.4.4 Uniform prior

The uniform prior is defined in section 2.4.5.4, and is given by:

$$\pi_U(\boldsymbol{\psi}) \propto c, \tag{3.49}$$

where c is an arbitrary constant. Therefore the posterior distribution of $\boldsymbol{\psi}$ under the uniform prior is merely proportional to likelihood function, and is given by:

$$
\begin{aligned}
\pi_U(\boldsymbol{\psi} \mid \boldsymbol{t}) &= c_U^{-1} \times \pi_U(\boldsymbol{\psi}) \times L(\boldsymbol{\psi} \mid \boldsymbol{t}) \\
&\propto \pi_U(\boldsymbol{\psi}) \times L(\boldsymbol{\psi} \mid \boldsymbol{t}) \\
&= \beta^n \lambda_0^{n\beta} \eta^{\beta\bar{\delta}} \prod_{i=1}^{k}\prod_{j=1}^{n_i} t_{ij}^{\beta-1} exp\left\{ -\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta}\eta^{\beta\delta_i} \right\}.
\end{aligned}
\tag{3.50}
$$

where c_U is the normalizing constant, and is given by:

$$
c_U = \int_0^{\infty}\int_0^{\infty}\int_1^{\infty} \beta^n \lambda_0^{n\beta} \eta^{\beta\bar{\delta}} \prod_{i=1}^{k}\prod_{j=1}^{n_i} t_{ij}^{\beta-1} exp\left\{ -\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta}\eta^{\beta\delta_i} \right\} d\lambda_0 d\beta d\eta.
\tag{3.51}
$$

The posterior distribution under the uniform prior is of an unknown form, therefore the properness needs to be considered to ensure admissible results.

3.4.5 Probability matching prior

The PMP is described in 2.4.5.5 and is used to connect the Bayesian and frequentist methods. The algorithm to construct a second-order PMP for a parameter of interest is given in section 2.4.5.5.

The PMPs were found in Xu et al. (2015), although no working out was shown. Hence this section is devoted to showing full derivations of the probability matching priors for the Weibull distribution under the log-linear time transformation function.

Theorem 3.8. *The probability matching prior when λ_0 is the parameter of interest is given by:*

$$\pi_{M1}(\boldsymbol{\psi}) = g_1\left(\eta\lambda_0^{-\frac{a_{12}}{a_{11}}}, \frac{a_{11} + a_{13}\beta log(\lambda_0)}{a_{13}\beta} \right) \lambda_0^{-1-\frac{a_{12}}{a_{11}}} \beta^{-1}, \tag{3.52}$$

where g_1 is an arbitrary positive continuous function.

Proof. Consider λ_0 as the parameter of interest. Following the algorithm in section 2.4.5.5 choose $t'(\boldsymbol{\psi}) = (\lambda_0, 0, 0)'$, resulting in the gradient vector $\nabla'_t(\boldsymbol{\psi}) = (1, 0, 0)'$. Therefore:

$$\eta_i(\boldsymbol{\psi}) = \frac{\nabla'_t(\boldsymbol{\psi}) I^{-1}(\boldsymbol{\psi})}{\sqrt{\nabla'_t(\boldsymbol{\psi}) I^{-1}(\boldsymbol{\psi}) \nabla_t(\boldsymbol{\psi})}}, \tag{3.53}$$

which further reduces to:

$$\eta_i\left(\boldsymbol{\psi}\right) = \frac{I_{1j}^{-1}\left(\boldsymbol{\psi}\right)}{\sqrt{I_{11}^{-1}\left(\boldsymbol{\psi}\right)}}, \tag{3.54}$$

where: $I_{1j}^{-1}\left(\boldsymbol{\psi}\right)$ is the first row of the inverse Fisher information matrix.

Therefore the probability matching prior $\pi_{M1}\left(\boldsymbol{\psi}\right)$ should satisfy the following differential equation:

$$\sum_{i=1}^{3} \frac{\partial}{\partial\psi_i}\left\{\eta_i\left(\boldsymbol{\psi}\right)\pi_{M1}\left(\boldsymbol{\psi}\right)\right\} = 0, \tag{3.55}$$

Where: ψ_i is the ith element in $\boldsymbol{\psi}$.

Equation 3.55 becomes:

$$\frac{\partial}{\partial\lambda_0}\left[\sqrt{\frac{a_{11}}{M}}\frac{\lambda_0}{\beta}\pi_{M1}\left(\boldsymbol{\psi}\right)\right] + \frac{\partial}{\partial\eta}\left[\frac{a_{12}}{\sqrt{Ma_{11}}}\frac{\eta}{\beta}\pi_{M1}\left(\boldsymbol{\psi}\right)\right] + \frac{\partial}{\partial\beta}\left[\frac{a_{13}}{\sqrt{Ma_{11}}}\beta\pi_{M1}\left(\boldsymbol{\psi}\right)\right] = 0, \tag{3.56}$$

where: a_{ij} are the values found in the inverse Fisher information matrix, given in equation 3.23.

Equation 3.56 simplifies to:

$$\frac{\partial}{\partial\lambda_0}\left[a_{11}\frac{\lambda_0}{\beta}\pi_{M1}\left(\boldsymbol{\psi}\right)\right] + \frac{\partial}{\partial\eta}\left[a_{12}\frac{\eta}{\beta}\pi_{M1}\left(\boldsymbol{\psi}\right)\right] + \frac{\partial}{\partial\beta}\left[a_{13}\beta\pi_{M1}\left(\boldsymbol{\psi}\right)\right] = 0. \tag{3.57}$$

Using the method of characteristics, the differential characteristic equations are as follows:

$$\frac{\beta\partial\lambda_0}{a_{11}\lambda_0} = \frac{\beta\partial\eta}{a_{12}\eta} = \frac{\partial\beta}{a_{13}\beta} = \frac{\partial_{\pi_{M1}}}{-\pi_{M1}\left(\frac{a_{11}}{\beta} + \frac{a_{12}}{\beta} + a_{13}\right)}. \tag{3.58}$$

A first family of characteristics comes from:

$$\frac{\beta\partial\lambda_0}{a_{11}\lambda_0} = \frac{\beta\partial\eta}{a_{12}\eta} \implies c_1 = \eta\lambda_0^{-\frac{a_{12}}{a_{11}}}. \tag{3.59}$$

A second family of characteristics comes from:

$$\frac{\beta\partial\lambda_0}{a_{11}\lambda_0} = \frac{\partial\beta}{a_{13}\beta} \implies c_2 = \frac{a_{11} + a_{13}\beta log\left(\lambda_0\right)}{a_{13}\beta}. \tag{3.60}$$

A third family of characteristics comes from:

$$\frac{\beta\partial\lambda_0}{a_{11}\lambda_0} = \frac{\partial_{\pi_{M1}}}{-\pi_{M1}\left(\frac{a_{11}}{\beta} + \frac{a_{12}}{\beta} + a_{13}\right)}, \tag{3.61}$$

and since $c_2 = \frac{a_{11}+a_{13}\beta log(\lambda_0)}{a_{13}\beta}$, this can be simplified to:

$$\frac{\left(a_{11} + a_{12} + \frac{a_{11}}{(c_2-log(\lambda_0))}\right)\partial\lambda_0}{\lambda_0} = \frac{a_{11}\partial_{\pi_{M1}}}{-\pi_{M1}}. \tag{3.62}$$

By setting $u = c_2 - log(\lambda_0) \implies du = -\lambda_0^{-1}$ and noting that $c_2 - log(\lambda_0) = \frac{a_{11}}{a_{13}\beta}$ the above equation becomes:

$$(a_{11} + a_{12}) log(\lambda_0) - a_{11} log\left(\frac{a_{11}}{a_{13}\beta}\right) + c_3 = -a_{11} log(\pi_{M1}). \tag{3.63}$$

Solving in terms of π_{M1} yields:

$$\pi_{M1} = \lambda_0^{-1-\frac{a_{12}}{a_{11}}} \beta^{-1} c_3^\star, \tag{3.64}$$

where $c_3^\star = \left(\frac{a_{11}}{a_{13}}\right) exp\left\{-\frac{c_3}{a_{11}}\right\}$. Therefore

$$\begin{aligned}
\pi_{M1}(\boldsymbol{\psi}) &= g_1(c_1, c_2) \lambda_0^{-1-\frac{a_{12}}{a_{11}}} \beta^{-1} \\
&= g_1\left(\eta \lambda_0^{-\frac{a_{12}}{a_{11}}}, \frac{a_{11} + a_{13}\beta log(\lambda_0)}{a_{13}\beta}\right) \lambda_0^{-1-\frac{a_{12}}{a_{11}}} \beta^{-1},
\end{aligned} \tag{3.65}$$

where g_1 is is any arbitrary positive function differentiable on the two variables. $\qquad\square$

Theorem 3.9. *The probability matching prior when η is the parameter of interest is given by:*

$$\pi_{M2}(\boldsymbol{\psi}) = g_2\left(\eta \lambda_0^{-\frac{a_{22}}{a_{21}}}, \frac{a_{21} + a_{23}\beta log(\lambda_0)}{a_{23}\beta}\right) \lambda_0^{-1-\frac{a_{22}}{a_{21}}} \beta^{-1}, \tag{3.66}$$

where g_2 is an arbitrary positive continuous function.

Proof. Consider η as the parameter of interest. Following the algorithm in section 2.4.5.5 choose $t'(\boldsymbol{\psi}) = (0, \eta, 0)'$, resulting in the gradient vector $\nabla_t'(\boldsymbol{\psi}) = (0, 1, 0)'$. Therefore:

$$\eta_i(\boldsymbol{\psi}) = \frac{I_{2j}^{-1}(\boldsymbol{\psi})}{\sqrt{I_{22}^{-1}(\boldsymbol{\psi})}}, \tag{3.67}$$

where: $I_{2j}^{-1}(\boldsymbol{\psi})$ is the second row of the inverse Fisher information matrix.

Therefore the probability matching prior $\pi_{M2}(\boldsymbol{\psi})$ should satisfy the following differential equation:

$$\frac{\partial}{\partial \lambda_0}\left[\sqrt{\frac{a_{21}}{M}}\frac{\lambda_0}{\beta}\pi_{M2}(\boldsymbol{\psi})\right] + \frac{\partial}{\partial \eta}\left[\frac{a_{22}}{\sqrt{Ma_{22}}}\frac{\eta}{\beta}\pi_{M2}(\boldsymbol{\psi})\right] + \frac{\partial}{\partial \beta}\left[\frac{a_{23}}{\sqrt{Ma_{22}}}\beta\pi_{M2}(\boldsymbol{\psi})\right] = 0, \tag{3.68}$$

which is simplified to:

$$\frac{\partial}{\partial \lambda_0}\left[a_{21}\frac{\lambda_0}{\beta}\pi_{M2}(\boldsymbol{\psi})\right] + \frac{\partial}{\partial \eta}\left[a_{22}\frac{\eta}{\beta}\pi_{M2}(\boldsymbol{\psi})\right] + \frac{\partial}{\partial \beta}[a_{23}\beta\pi_{M2}(\boldsymbol{\psi})] = 0. \tag{3.69}$$

Using the method of characteristics, the differential characteristic equations are as follows:

$$\frac{\beta\partial\lambda_0}{a_{21}\lambda_0} = \frac{\beta\partial\eta}{a_{22}\eta} = \frac{\partial\beta}{a_{23}\beta} = \frac{\partial_{\pi_{M2}}}{-\pi_{M2}\left(\frac{a_{21}}{\beta} + \frac{a_{22}}{\beta} + a_{23}\right)}. \tag{3.70}$$

Following an approach similar to when λ_0 was the parameter of interest, the probability matching prior $\pi_{M2}(\psi)$ is given by:

$$
\begin{aligned}
\pi_{M2}(\psi) &= g_2(c_1, c_2) \lambda_0^{-1-\frac{a_{22}}{a_{21}}} \beta^{-1} \\
&= g_2\left(\eta\lambda_0^{-\frac{a_{22}}{a_{21}}}, \frac{a_{21} + a_{23}\beta log(\lambda_0)}{a_{23}\beta}\right) \lambda_0^{-1-\frac{a_{22}}{a_{21}}} \beta^{-1},
\end{aligned}
\tag{3.71}
$$

where g_2 is any arbitrary positive function differentiable on the two variables. $\qquad\square$

Theorem 3.10. *The probability matching prior when β is the parameter of interest is given by:*

$$
\pi_{M3}(\psi) = g_3\left(\eta\lambda_0^{-\frac{a_{32}}{a_{31}}}, \frac{a_{31} + a_{33}\beta log(\lambda_0)}{a_{33}\beta}\right) \lambda_0^{-1-\frac{a_{32}}{a_{31}}} \beta^{-1},
\tag{3.72}
$$

where g_3 is an arbitrary positive continuous function.

Proof. Consider β as the parameter of interest. Following the algorithm in section 2.4.5.5 choose $t'(\psi) = (0, 0, \beta)'$, resulting in the gradient vector $\nabla'_t(\psi) = (0, 0, 1)'$. Therefore:

$$
\eta_i(\psi) = \frac{I_{3j}^{-1}(\psi)}{\sqrt{I_{33}^{-1}(\psi)}},
\tag{3.73}
$$

where: $I_{3j}^{-1}(\psi)$ is the third row of the inverse Fisher information matrix.

Therefore the probability matching prior $\pi_{M3}(\psi)$ should satisfy the following differential equation:

$$
\frac{\partial}{\partial\lambda_0}\left[\sqrt{\frac{a_{31}}{M}}\frac{\lambda_0}{\beta}\pi_{M3}(\psi)\right] + \frac{\partial}{\partial\eta}\left[\frac{a_{32}}{\sqrt{M a_{33}}}\frac{\eta}{\beta}\pi_{M3}(\psi)\right] + \frac{\partial}{\partial\beta}\left[\frac{a_{33}}{\sqrt{M a_{33}}}\beta\pi_{M3}(\psi)\right] = 0,
\tag{3.74}
$$

which is simplified to:

$$
\frac{\partial}{\partial\lambda_0}\left[a_{31}\frac{\lambda_0}{\beta}\pi_{M3}(\psi)\right] + \frac{\partial}{\partial\eta}\left[a_{32}\frac{\eta}{\beta}\pi_{M3}(\psi)\right] + \frac{\partial}{\partial\beta}\left[a_{33}\beta\pi_{M3}(\psi)\right] = 0.
\tag{3.75}
$$

Using the method of characteristics, the differential characteristic equations are as follows:

$$
\frac{\beta\partial\lambda_0}{a_{31}\lambda_0} = \frac{\beta\partial\eta}{a_{32}\eta} = \frac{\partial\beta}{a_{33}\beta} = \frac{\partial_{\pi_{M3}}}{-\pi_{M3}\left(\frac{a_{31}}{\beta} + \frac{a_{32}}{\beta} + a_{33}\right)}.
\tag{3.76}
$$

Following an approach similar to when λ_0 was the parameter of interest, the probability matching prior $\pi_{M2}(\psi)$ is given by:

$$
\begin{aligned}
\pi_{M3}(\psi) &= g_3(c_1, c_2) \lambda_0^{-1-\frac{a_{32}}{a_{31}}} \beta^{-1} \\
&= g_3\left(\eta\lambda_0^{-\frac{a_{32}}{a_{31}}}, \frac{a_{31} + a_{33}\beta log(\lambda_0)}{a_{33}\beta}\right) \lambda_0^{-1-\frac{a_{32}}{a_{31}}} \beta^{-1},
\end{aligned}
\tag{3.77}
$$

where g_3 is any arbitrary positive function differentiable on the two variables. $\qquad\square$

Finally, it will be shown that the reference priors: $\pi_J(\boldsymbol{\psi})$, $\pi_{R1}(\boldsymbol{\psi})$ and $\pi_{R2}(\boldsymbol{\psi})$ and the priors: $\pi_{MDI}(\boldsymbol{\psi})$ and $\pi_U(\boldsymbol{\psi})$ are (or are not) the second-order probability matching priors.

Theorem 3.11. *The reference prior $\pi_J(\boldsymbol{\psi})$ is not a second-order probability matching prior regardless of which parameter is of interest.*

Proof. Substituting $\pi_J(\boldsymbol{\psi})$ into equation 3.56 yields:

$$\frac{\partial}{\partial \lambda_0}\left[\sqrt{\frac{a_{11}}{M}}\frac{1}{\eta}\right] + \frac{\partial}{\partial \eta}\left[\frac{a_{12}}{\sqrt{Ma_{11}}}\frac{1}{\lambda_0}\right] + \frac{\partial}{\partial \beta}\left[\frac{a_{13}}{\sqrt{Ma_{11}}}\frac{\beta^2}{\lambda_0\eta}\right] = \frac{2a_{13}\beta}{\sqrt{Ma_{11}}\lambda_0\eta} \neq 0. \qquad (3.78)$$

Similar answers are given if $\pi_J(\boldsymbol{\psi})$ is substituted into equations 3.68 and 3.74. Therefore the reference prior $\pi_J(\boldsymbol{\psi})$ is not the probability matching prior regardless of which parameter is of interest. $\qquad\square$

Theorem 3.12. *The reference prior $\pi_{R1}(\boldsymbol{\psi})$ is not a second-order probability matching regardless of which parameter is of interest.*

Proof. Substituting π_{R1} into equation 3.56 yields:

$$\frac{\partial}{\partial \lambda_0}\left[\sqrt{\frac{a_{11}}{M}}\frac{1}{\beta\eta}\right] + \frac{\partial}{\partial \eta}\left[\frac{a_{12}}{\sqrt{Ma_{11}}}\frac{1}{\lambda_0\beta}\right] + \frac{\partial}{\partial \beta}\left[\frac{a_{13}}{\sqrt{Ma_{11}}}\frac{\beta}{\lambda_0\eta}\right] = \frac{a_{13}}{\sqrt{Ma_{11}}\lambda_0\eta} \neq 0. \qquad (3.79)$$

Similar answers are given if $\pi_{R1}(\boldsymbol{\psi})$ is substituted into equations 3.68 and 3.74. Therefore the reference prior $\pi_{R1}(\boldsymbol{\psi})$ is not the probability matching prior regardless of which parameter is of interest. $\qquad\square$

Theorem 3.13. *The reference prior $\pi_{R2}(\boldsymbol{\psi})$ is a second-order probability matching prior regardless of which parameter is of interest.*

Proof. Substituting $\pi_{R2}(\boldsymbol{\psi})$ into equation 3.56 yields:

$$\frac{\partial}{\partial \lambda_0}\left[\sqrt{\frac{a_{11}}{M}}\frac{1}{\eta\beta^2}\right] + \frac{\partial}{\partial \eta}\left[\frac{a_{12}}{\sqrt{Ma_{11}}}\frac{1}{\lambda_0\beta^2}\right] + \frac{\partial}{\partial \beta}\left[\frac{a_{13}}{\sqrt{Ma_{11}}}\frac{1}{\lambda_0\eta}\right] = 0. \qquad (3.80)$$

Similar answers are given if $\pi_{R2}(\boldsymbol{\psi})$ is substituted into equations 3.68 and 3.74. Therefore the reference prior $\pi_{R2}(\boldsymbol{\psi})$ is the probability matching prior regardless of which parameter is of interest. $\qquad\square$

Theorem 3.14. *The prior $\pi_{MDI}(\boldsymbol{\psi})$ is not a second-order probability matching prior regardless of which parameter is of interest.*

Proof. Substituting $\pi_{MDI}(\boldsymbol{\psi})$ into equation 3.56 yields:

$$\frac{\partial}{\partial \lambda_0}\left[\sqrt{\frac{a_{11}}{M}}\lambda_0^2\eta^{\delta_i}exp\left\{\frac{\gamma}{\beta}\right\}\right] + \frac{\partial}{\partial \eta}\left[\frac{a_{12}\lambda_0\eta^{\delta_i+1}}{\sqrt{Ma_{11}}}exp\left\{\frac{\gamma}{\beta}\right\}\right] + \frac{\partial}{\partial \beta}\left[\frac{a_{13}\beta\lambda_0\eta^{\delta_i}}{\sqrt{Ma_{11}}}exp\left\{\frac{\gamma}{\beta}\right\}\right] = k \neq 0,$$
$$\qquad (3.81)$$

where: $k = 2\sqrt{\frac{a_{11}}{M}}\lambda_0\eta^{\delta_i}exp\left\{\frac{\gamma}{\beta}\right\} + (\delta_i + 1)\frac{a_{12}}{\sqrt{Ma_{11}}}\lambda_0\eta^{\delta_i}exp\left\{\frac{\gamma}{\beta}\right\} + \frac{a_{13}}{\sqrt{Ma_{11}}}\lambda_0\eta^{\delta_i}\left(\frac{\beta-\gamma}{\beta}\right)exp\left\{\frac{\gamma}{\beta}\right\}$.
Similar answers are given if $\pi_{MDI}(\boldsymbol{\psi})$ is substituted into equations 3.68 and 3.74. Therefore the prior $\pi_{MDI}(\boldsymbol{\psi})$ is not the probability matching prior regardless of which parameter is of interest. $\qquad\square$

Theorem 3.15. *The prior $\pi_U(\boldsymbol{\psi})$ is not a second-order probability matching prior regardless of which parameter is of interest.*

Proof. Substituting $\pi_U(\boldsymbol{\psi})$ into equation 3.56 yields:

$$\frac{\partial}{\partial\lambda_0}\left[\sqrt{\frac{a_{11}}{M}}\frac{\lambda_0}{\beta}\right] + \frac{\partial}{\partial\eta}\left[\frac{a_{12}}{\sqrt{Ma_{11}}}\frac{\eta}{\beta}\right] + \frac{\partial}{\partial\beta}\left[\frac{a_{13}}{\sqrt{Ma_{11}}}\beta\right] = \sqrt{\frac{a_{11}}{M\beta^2}} + \frac{a_{12}}{\sqrt{Ma_{11}\beta^2}} + \frac{a_{13}}{\sqrt{Ma_{11}}} \neq 0. \tag{3.82}$$

Similar answers are given if $\pi_U(\boldsymbol{\psi})$ is substituted into equations 3.68 and 3.74. Therefore the prior $\pi_U(\boldsymbol{\psi})$ is not the probability matching prior regardless of which parameter is of interest. $\qquad\square$

3.4.6 Properness of posterior distributions

For Bayesian analysis to be meaningful the posterior distributions need to be proper, that is, they integrate to a finite constant. The posterior distribution under each prior derived above are all of an unknown form, and hence their properness needs to be considered. This section is devoted proving proneness (or improperness) of the posterior distributions under the non-informative priors derived above. The necessary lemmas and propositions required can be found in appendix A.3.

This section includes a proof for the properness of the posterior distributions under: Jeffreys' prior, the reference priors, the MDI prior and the uniform prior. An alternative proof for the properness of the posterior distributions for the three reference priors can be found in Xu et al. (2015).

Theorem 3.16. *The posterior distribution for the log-linear Weibull distribution under the reference priors is proper provided that $\bar{\delta} < n\,\delta_{max}$ and $n > m + 1$.*

Proof. For the posterior distribution to be proper the following must hold:

$$c_R \int\limits_0^\infty \int\limits_0^\infty \int\limits_1^\infty \pi_R(\boldsymbol{\psi} \mid \boldsymbol{t})\, d\beta d\lambda_0 d\eta = 1, \tag{3.83}$$

for some normalizing constant c_R. For this to be true it suffices to show:

$$\int\limits_0^\infty \int\limits_0^\infty \int\limits_1^\infty \pi_R(\boldsymbol{\psi} \mid \boldsymbol{t})\, d\beta d\lambda_0 d\eta < \infty. \tag{3.84}$$

Consider the above integral:

$$\propto \int\limits_0^\infty \int\limits_0^\infty \int\limits_1^\infty \beta^{n-m} \lambda_0^{n\beta-1} \eta^{\beta\bar{\delta}-1} \prod_{i=1}^k \prod_{j=1}^{n_i} t_{ij}^\beta exp\left\{ -\sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \lambda_0^\beta \eta^{\beta\delta_i} \right\} d\beta d\lambda_0 d\eta. \tag{3.85}$$

Since $\lambda_0^\beta \sim Gamma\left(\alpha^\star = n, \beta^\star = \sum\limits_{i=1}^k \sum\limits_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i}\right)$ the integral in equation 3.85 may be written as:

$$\int\limits_0^\infty \int\limits_1^\infty \Gamma\left(n\right) \beta^{n-m-1} \eta^{\beta\bar{\delta}-1} \prod_{i=1}^k \prod_{j=1}^{n_i} t_{ij}^\beta \left(\sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i}\right)^{-n} d\beta d\eta$$

$$\propto \int\limits_0^\infty \int\limits_1^\infty \beta^{n-m-1} \eta^{\beta\bar{\delta}-1} \prod_{i=1}^k \prod_{j=1}^{n_i} t_{ij}^\beta \left(\sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i}\right)^{-n} d\beta d\eta, \tag{3.86}$$

where $\Gamma\left(\cdot\right)$ is the gamma function defined in equation 2.108.

Following Ramos et al. (2020), let t_{max} and δ_{max} denote the largest t_{ij} and δ_i values respectively. Then for all $t_{ij} < t_{max}$ and $\delta_i < \delta_{max}$:

$$\lim_{\eta,\beta\to\infty} \frac{t_{ij}^\beta \eta^{\beta\delta_i}}{t_{max}^\beta \eta^{\beta\delta_{max}}} = 0. \tag{3.87}$$

Therefore it follows that,

$$\lim_{\eta,\beta\to\infty} \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i}}{t_{max}^\beta \eta^{\beta\delta_{max}}} = 1 \Rightarrow \sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i} \underset{\eta,\beta\to\infty}{\propto} t_{max}^\beta \eta^{\beta\delta_{max}}. \tag{3.88}$$

Furthermore:

$$\lim_{\eta\to 1} \lim_{\beta\to 0} \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i}}{t_{max}^\beta \eta^{\beta\delta_{max}}} = n \Rightarrow \sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i} \underset{\eta\to 1, \beta\to 0}{\propto} t_{max}^\beta \eta^{\beta\delta_{max}}. \tag{3.89}$$

Therefore from proposition A.3, it follows that:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i} \propto t_{max}^\beta \eta^{\beta\delta_{max}}, \tag{3.90}$$

in the interval $\beta \in [0, \infty)$ and $\eta \in [1, \infty)$. Therefore the integral in equation 3.86 is:

$$
\begin{aligned}
&\propto \int_0^\infty \int_1^\infty \beta^{n-m-1} \eta^{\beta\bar{\delta}-1} \prod_{i=1}^k \prod_{j=1}^{n_i} t_{ij}^\beta \left(t_{max}^\beta \eta^{\beta\delta_{max}}\right)^{-n} d\beta d\eta \\
&\propto \int_0^\infty \int_1^\infty \beta^{n-m-1} exp\{-c\beta\} \eta^{\beta\bar{\delta}-n\beta\delta_{max}-1} d\beta d\eta,
\end{aligned}
\tag{3.91}
$$

where $c = log\left(\dfrac{t_{max}^n}{\prod\limits_{i=1}^k \prod\limits_{j=1}^{n_i} t_{ij}}\right) > 0$ provided that $t_{max} \neq t_{ij} \underset{i,j}{\forall}$. Now consider three separate cases:

1. Consider the case where $\beta\bar{\delta} - n\beta\delta_{max} - 1 \geq 0$, then equation 3.91 becomes:

$$
\int_0^\infty \beta^{n-m-1} exp\{-c\beta\} \int_1^\infty \eta^{\beta\bar{\delta}-n\beta\delta_{max}-1} d\eta d\beta = \infty.
\tag{3.92}
$$

2. Consider the case where $-1 \leq \beta\bar{\delta} - n\beta\delta_{max} - 1 < 0$. By letting $p^\star = \beta\bar{\delta} - n\beta\delta_{max} - 1$, then equation 3.91 becomes:

$$
\int_0^\infty \beta^{n-m-1} exp\{-c\beta\} \int_1^\infty \eta^{p^\star} d\eta d\beta = \infty.
\tag{3.93}
$$

This is because the p-series integral diverges for $-1 \leq p^\star < 0$.

3. Consider the case where $\beta\bar{\delta} - n\beta\delta_{max} - 1 < -1$. By letting $p^\star = \beta\bar{\delta} - n\beta\delta_{max} - 1$, then equation 3.91 becomes:

$$
\begin{aligned}
&\int_0^\infty \beta^{n-m-1} exp\{-c\beta\} \int_1^\infty \eta^{p^\star} d\eta d\beta \\
&= \int_0^\infty \frac{\beta^{n-m-1} exp\{-c\beta\}}{\beta\left(\bar{\delta} - n\delta_{max}\right)} d\beta \\
&\propto \int_0^\infty \beta^{n-m-2} exp\{-c\beta\} d\beta,
\end{aligned}
\tag{3.94}
$$

since the p-series integral converges when $p^\star < -1$. The integral in equation 3.94 converges for $n > m + 1$ since $\beta \sim Gamma\left(\alpha^\star = n - m - 1, \beta^\star = c\right)$, which is a distribution of a known-form and therefore has a finite-normalizing constant.

Therefore the posterior distributions under the reference priors are proper provided that $\bar{\delta} < n\,\delta_{max}$ and $n > m + 1$, and hence further analysis may be conducted with them.   □

Theorem 3.17. *The posterior distribution for the log-linear Weibull distribution under the uniform prior is improper for* $\underset{n}{\forall}$.

Proof. For the posterior distribution to be proper the following must hold:

$$c_U \int\limits_0^\infty \int\limits_0^\infty \int\limits_1^\infty \pi_U \left(\boldsymbol{\psi} \mid \boldsymbol{t} \right) d\beta d\lambda_0 d\eta = 1, \tag{3.95}$$

for some normalizing constant c_U. For this to be true it suffices to show:

$$\int\limits_0^\infty \int\limits_0^\infty \int\limits_1^\infty \pi_U \left(\boldsymbol{\psi} \mid \boldsymbol{t} \right) d\beta d\lambda_0 d\eta < \infty. \tag{3.96}$$

Consider the above integral:

$$\propto \int\limits_0^\infty \int\limits_0^\infty \int\limits_1^\infty \beta^n \lambda_0^{n\beta} \eta^{\beta\bar{\delta}} \prod_{i=1}^k \prod_{j=1}^{n_i} t_{ij}^\beta exp \left\{ - \sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \lambda_0^\beta \eta^{\beta\delta_i} \right\} d\beta d\lambda_0 d\eta. \tag{3.97}$$

Since $\lambda_0^\beta \sim Gamma \left(\alpha^\star = \left(n + \frac{1}{\beta} \right), \beta^\star = \sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i} \right)$, the integral in equation 3.97 may be written as:

$$\int\limits_0^\infty \int\limits_1^\infty \beta^{n-1} \Gamma \left(n + \frac{1}{\beta} \right) \prod_{i=1}^k \prod_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i} \left(\sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i} \right)^{-\left(n + \frac{1}{\beta} \right)} d\beta d\eta, \tag{3.98}$$

where $\Gamma \left(\cdot \right)$ is the gamma function defined in equation 2.108.

Following Ramos et al. (2020), let t_{max} and δ_{max} denote the largest t_{ij} and δ_i values respectively. Then for all $t_{ij} < t_{max}$ and $\delta_i < \delta_{max}$:

$$\lim_{\eta, \beta \to \infty} \frac{t_{ij}^\beta \eta^{\beta\delta_i}}{t_{max}^\beta \eta^{\beta\delta_{max}}} = 0. \tag{3.99}$$

Therefore it follows that,

$$\lim_{\eta, \beta \to \infty} \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i}}{t_{max}^\beta \eta^{\beta\delta_{max}}} = 1 \Rightarrow \sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i} \underset{\eta, \beta \to \infty}{\propto} t_{max}^\beta \eta^{\beta\delta_{max}}. \tag{3.100}$$

Furthermore:

$$\lim_{\eta \to 1} \lim_{\beta \to 0} \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i}}{t_{max}^\beta \eta^{\beta\delta_{max}}} = n \Rightarrow \sum_{i=1}^k \sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i} \underset{\eta \to 1, \beta \to 0}{\propto} t_{max}^\beta \eta^{\beta\delta_{max}}. \tag{3.101}$$

Therefore from proposition A.3, it follows that:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\eta^{\beta\delta_i} \propto t_{max}^{\beta}\eta^{\beta\delta_{max}}, \tag{3.102}$$

in the interval $\beta \in [0,\infty)$ and $\eta \in [1,\infty)$. Therefore the integral in equation 3.98 is:

$$\begin{aligned}
&\propto \int_0^{\infty}\int_1^{\infty} \beta^{n-1}\Gamma\left(n+\frac{1}{\beta}\right) \prod_{i=1}^{k}\prod_{j=1}^{n_i} t_{ij}^{\beta}\eta^{\beta\delta_i} \left(t_{max}^{\beta}\eta^{\beta\delta_{max}}\right)^{-\left(n+\frac{1}{\beta}\right)} d\beta d\eta \\
&\propto \int_0^{\infty}\int_1^{\infty} \beta^{n-1}\Gamma\left(n+\frac{1}{\beta}\right) exp\left\{-c\beta\right\} \eta^{\beta\bar{\delta}-n\beta\delta_{max}-\delta_{max}} d\beta d\eta,
\end{aligned} \tag{3.103}$$

where $c = log\left(\frac{t_{max}^n}{\prod\limits_{i=1}^{k}\prod\limits_{j=1}^{n_i} t_{ij}}\right) > 0$ provided that $t_{max} \neq t_{ij} \underset{i,j}{\forall}$. Now consider three separate cases:

1. Consider the case where $\beta\bar{\delta} - n\beta\delta_{max} - \delta_{max} \geq 0$, then equation 3.103 becomes:

$$\int_0^{\infty} \beta^{n-1}\Gamma\left(n+\frac{1}{\beta}\right) exp\left\{-c\beta\right\} \int_1^{\infty} \eta^{\beta\bar{\delta}-n\beta\delta_{max}-\delta_{max}} d\eta d\beta = \infty. \tag{3.104}$$

2. Consider the case where $-1 \leq \beta\bar{\delta} - n\beta\delta_{max} - \delta_{max} < 0$. By letting $p^{\star} = \beta\bar{\delta} - n\beta\delta_{max} - \delta_{max}$, then equation 3.103 becomes:

$$\int_0^{\infty} \beta^{n-1}\Gamma\left(n+\frac{1}{\beta}\right) exp\left\{-c\beta\right\} \int_1^{\infty} \eta^{p^{\star}} d\eta d\beta = \infty. \tag{3.105}$$

   This is because the p-series integral diverges for $-1 \leq p^{\star} < 0$.

3. Consider the case where $\beta\bar{\delta} - n\beta\delta_{max} - \delta_{max} < -1$. By letting $p^{\star} = \beta\bar{\delta} - n\beta\delta_{max} - \delta_{max}$, then equation 3.103 becomes:

$$\begin{aligned}
&\int_0^{\infty} \beta^{n-1}\Gamma\left(n+\frac{1}{\beta}\right) exp\left\{-c\beta\right\} \int_1^{\infty} \eta^{p^{\star}} d\eta d\beta \\
&\propto \int_0^{\infty} \frac{\beta^{n-1}\Gamma\left(n+\frac{1}{\beta}\right) exp\left\{-c\beta\right\}}{\beta\left(n\delta_{max}-\bar{\delta}\right) + \delta_{max} - 1} d\beta,
\end{aligned} \tag{3.106}$$

   since the p-series integral converges when $p^{\star} < -1$. Since $\delta_{max} - 1 \geq 0$, it implies that

the integral in equation 3.106 is larger than or equal to:

$$\int_0^\infty \frac{\beta^{n-1}\Gamma\left(n+\frac{1}{\beta}\right)exp\left\{-c\beta\right\}}{\beta\left(n\delta_{max}-\bar{\delta}\right)}d\beta$$

$$\propto \int_0^\infty \beta^{n-2}\Gamma\left(n+\frac{1}{\beta}\right)exp\left\{-c\beta\right\}d\beta. \tag{3.107}$$

Consider the substitution $\omega = n + \frac{1}{\beta} \Rightarrow d\omega = -\frac{1}{\beta^2}d\beta$. Then equation 3.107 becomes:

$$\propto \int_n^\infty (\omega-n)^{-n}\,\Gamma(\omega)\,exp\left\{-c(w-n)^{-1}\right\}d\omega. \tag{3.108}$$

Following Stirling's approximation as given in corollary A.1:

$$\Gamma(\omega) \sim \sqrt{2\pi}\omega^{\omega+\frac{1}{2}}exp\left\{-\omega\right\}. \tag{3.109}$$

Therefore:

$$(\omega-n)^{-n}\,\Gamma(\omega)\,exp\left\{-c(w-n)^{-1}\right\} \sim (\omega-n)^{-n}\,\sqrt{2\pi}\omega^{\omega+\frac{1}{2}}exp\left\{-c(w-n)^{-1}-\omega\right\}. \tag{3.110}$$

Therefore as $\omega \to \infty$ equation 3.107 becomes:

$$\mathcal{O}\left(\omega^\omega\right)\;\omega > 0. \tag{3.111}$$

That is equation 3.107 is not finite as $\omega \to \infty$.

Therefore the posterior distribution under the uniform prior is an improper distribution. $\qquad\square$

Theorem 3.18. *The posterior distribution for the log-linear Weibull distribution under the MDI prior is improper for $\underset{n}{\forall}$.*

Proof. For the posterior distribution to be proper the following must hold:

$$c_{MDI}\int_0^\infty\int_0^\infty\int_1^\infty \pi_{MDI}\left(\boldsymbol{\psi}\mid\boldsymbol{t}\right)d\beta d\lambda_0 d\eta = 1, \tag{3.112}$$

for some normalizing constant c_{MDI}. For this to be true it suffices to show:

$$\int_0^\infty\int_0^\infty\int_1^\infty \pi_{MDI}\left(\boldsymbol{\psi}\mid\boldsymbol{t}\right)d\beta d\lambda_0 d\eta < \infty. \tag{3.113}$$

Consider the above integral:

$$\propto \int\limits_0^\infty \int\limits_0^\infty \int\limits_1^\infty exp\left\{\frac{\gamma}{\beta}\right\} \beta^{n+1}\lambda_0^{n\beta+1}\eta^{\bar{\delta}(\beta+1)} \prod_{i=1}^k\prod_{j=1}^{n_i} t_{ij}^{\beta-1}exp\left\{-\sum_{i=1}^k\sum_{j=1}^{n_i} t_{ij}^\beta\lambda_0^\beta\eta^{\beta\delta_i}\right\} d\lambda_0 d\beta d\eta. \tag{3.114}$$

Since $\lambda_0^\beta \sim Gamma\left(\alpha^\star = \left(n+\frac{2}{\beta}\right), \beta^\star = \sum_{i=1}^k\sum_{j=1}^{n_i} t_{ij}^\beta\eta^{\beta\delta_i}\right)$, the integral in equation 3.114 may be written as:

$$\int\limits_1^\infty \int\limits_0^\infty exp\left\{\frac{\gamma}{\beta}\right\} \beta^n\Gamma\left(n+\frac{2}{\beta}\right) \prod_{i=1}^k\prod_{j=1}^{n_i} t_{ij}^\beta\eta^{\beta\delta_i+\delta_i} \left(\sum_{i=1}^k\sum_{j=1}^{n_i} t_{ij}^\beta\eta^{\beta\delta_i}\right)^{-\left(n+\frac{2}{\beta}\right)} d\beta d\eta, \tag{3.115}$$

where $\Gamma(\cdot)$ is the gamma function defined in equation 2.108.

Note that $exp\left\{\frac{\gamma}{\beta}\right\} > 1 \,\forall\, \beta > 0$. Therefore the integral in equation 3.115 is greater than:

$$\int\limits_1^\infty \int\limits_0^\infty \beta^n\Gamma\left(n+\frac{2}{\beta}\right) \eta^{\bar{\delta}} \prod_{i=1}^k\prod_{j=1}^{n_i} t_{ij}^\beta\eta^{\beta\delta_i} \left(\sum_{i=1}^k\sum_{j=1}^{n_i} t_{ij}^\beta\eta^{\beta\delta_i}\right)^{-\left(n+\frac{2}{\beta}\right)} d\beta d\eta. \tag{3.116}$$

Following Ramos et al. (2020), let t_{max} and δ_{max} denote the largest t_{ij} and δ_i values respectively. Then for all $t_{ij} < t_{max}$ and $\delta_i < \delta_{max}$:

$$\lim_{\eta,\beta\to\infty} \frac{t_{ij}^\beta\eta^{\beta\delta_i}}{t_{max}^\beta\eta^{\beta\delta_{max}}} \quad = 0. \tag{3.117}$$

Therefore it follows that,

$$\lim_{\eta,\beta\to\infty} \frac{\sum_{i=1}^k\sum_{j=1}^{n_i} t_{ij}^\beta\eta^{\beta\delta_i}}{t_{max}^\beta\eta^{\beta\delta_{max}}} = 1 \Rightarrow \sum_{i=1}^k\sum_{j=1}^{n_i} t_{ij}^\beta\eta^{\beta\delta_i} \underset{\eta,\beta\to\infty}{\propto} t_{max}^\beta\eta^{\beta\delta_{max}}. \tag{3.118}$$

Furthermore:

$$\lim_{\eta\to1}\lim_{\beta\to0} \frac{\sum_{i=1}^k\sum_{j=1}^{n_i} t_{ij}^\beta\eta^{\beta\delta_i}}{t_{max}^\beta\eta^{\beta\delta_{max}}} = n \Rightarrow \sum_{i=1}^k\sum_{j=1}^{n_i} t_{ij}^\beta\eta^{\beta\delta_i} \underset{\eta\to1,\beta\to0}{\propto} t_{max}^\beta\eta^{\beta\delta_{max}}. \tag{3.119}$$

Therefore from proposition A.3, it follows that:

$$\sum_{i=1}^k\sum_{j=1}^{n_i} t_{ij}^\beta\eta^{\beta\delta_i} \propto t_{max}^\beta\eta^{\beta\delta_{max}}, \tag{3.120}$$

in the interval $\beta \in [0,\infty)$ and $\eta \in [1,\infty)$. Therefore the integral in equation 3.116 may be

written as:

$$\propto \int_0^\infty \int_1^\infty \beta^n \Gamma\left(n + \frac{2}{\beta}\right) \eta^{\bar{\delta}(\beta+1)} \prod_{i=1}^k \prod_{j=1}^{n_i} t_{ij}^\beta \left(t_{max}^\beta \eta^{\beta\delta_{max}}\right)^{-\left(n+\frac{2}{\beta}\right)} d\beta dd\eta$$

$$\propto \int_0^\infty \int_1^\infty \beta^n \Gamma\left(n + \frac{2}{\beta}\right) exp\left\{-c\beta\right\} \eta^{\bar{\delta}(\beta+1)-n\beta\delta_{max}-\delta_{max}} d\beta d\eta,$$

(3.121)

where $c = log\left(\frac{t_{max}^n}{\prod_{i=1}^k \prod_{j=1}^{n_i} t_{ij}}\right) > 0$ provided that $t_{max} \neq t_{ij} \; \underset{i,j}{\forall}$. Now consider three separate cases:

1. Consider the case where $\bar{\delta}(\beta+1) - n\beta\delta_{max} - \delta_{max} \geq 0$, then equation 3.103 becomes:

$$\int_0^\infty \beta^n \Gamma\left(n + \frac{2}{\beta}\right) exp\left\{-c\beta\right\} \int_1^\infty \eta^{\bar{\delta}(\beta+1)-n\beta\delta_{max}-\delta_{max}} d\eta d\beta = \infty.$$

(3.122)

2. Consider the case where $-1 \leq \bar{\delta}(\beta+1) - n\beta\delta_{max} - \delta_{max} < 0$. By letting $p^\star = \bar{\delta}(\beta+1) - n\beta\delta_{max} - \delta_{max}$, then equation 3.103 becomes:

$$\int_0^\infty \beta^n \Gamma\left(n + \frac{2}{\beta}\right) exp\left\{-c\beta\right\} \int_1^\infty \eta^{p^\star} d\eta d\beta = \infty.$$

(3.123)

This is because the p-series integral diverges for $-1 \leq p^\star < 0$.

3. Consider the case where $\bar{\delta}(\beta+1) - n\beta\delta_{max} - \delta_{max} < -1$. By letting $p^\star = \bar{\delta}(\beta+1) - n\beta\delta_{max} - \delta_{max}$, then equation 3.103 becomes:

$$\int_0^\infty \beta^n \Gamma\left(n + \frac{2}{\beta}\right) exp\left\{-c\beta\right\} \int_1^\infty \eta^{p^\star} d\eta d\beta$$

$$\propto \int_0^\infty \frac{\beta^n \Gamma\left(n + \frac{2}{\beta}\right) exp\left\{-c\beta\right\}}{\beta\left(n\delta_{max} - \bar{\delta}\right) - \bar{\delta} + \delta_{max} - 1} d\beta,$$

(3.124)

since the p-series integral converges when $p^\star < -1$. Since $\delta_{max} - 1 \geq 0$, it implies that the integral in equation 3.124 is larger than or equal to:

$$\int_0^\infty \frac{\beta^n \Gamma\left(n + \frac{2}{\beta}\right) exp\left\{-c\beta\right\}}{\beta\left(n\delta_{max} - \bar{\delta}\right) - \bar{\delta}} d\beta.$$

(3.125)

Consider the substitution $\omega = n + \frac{2}{\beta} \Rightarrow d\omega = -\frac{1}{\beta^2}d\beta$. Then equation 3.125 becomes:

$$\propto \int_n^\infty \frac{\left(\frac{\omega-n}{2}\right)^{-(n+2)} \Gamma\left(\omega\right) exp\left\{-c\left(\frac{w-n}{2}\right)^{-1}\right\}}{\left(\frac{\omega-n}{2}\right)^{-1}\left(n\delta_{max} - \bar\delta\right) - \bar\delta} d\omega. \tag{3.126}$$

Following Stirling's approximation as given in corollary A.1:

$$\Gamma\left(\omega\right) \sim \sqrt{2\pi}\omega^{\omega+\frac{1}{2}}exp\left\{-\omega\right\}. \tag{3.127}$$

Therefore:

$$\left(\frac{\omega-n}{2}\right)^{-(n+2)} \Gamma\left(\omega\right) exp\left\{-c\left(\frac{w-n}{2}\right)^{-1}\right\} y \sim \left(\frac{\omega-n}{2}\right)^{-(-n+2)} \sqrt{2\pi}\omega^{\omega+\frac{1}{2}} \times$$
$$\times exp\left\{-c\left(\frac{w-n}{2}\right)^{-1} - \omega\right\} y, \tag{3.128}$$

where: $y = \left(\frac{\omega-n}{2}\right)^{-1}\left(n\delta_{max} - \bar\delta\right) - \bar\delta$.

Therefore as $\omega \to \infty$ equation 3.125 becomes:

$$\mathcal{O}\left(\omega^\omega\right) \omega > 0. \tag{3.129}$$

That is equation 3.125 is not finite as $\omega \to \infty$.

Therefore the posterior distribution under the MDI prior is an improper distribution.   □

As a result the posterior distributions under the reference prior will be considered for the remainder of this thesis, whereas the posterior distributions under the MDI and uniform prior will not.

Chapter 4

Simulation study for Weibull distribution

4.1 Introduction

This chapter is devoted to running a simulation study using the estimates derived in chapter 3. Along with the maximum likelihood and MCMC Bayesian estimates, estimates obtained using Lindley's approximation technique are also found. The simulation study considers a three-level constant-stress ALT, where the stress is related to temperature. That is, the Arrhenius model is required. The first part of the simulation study compares the RMSE values of the estimates under three different loss functions: the squared error, LINEX and GELF. The second part of the simulation study finds the coverage rates for the MLEs and the MCMC Bayesian estimates, as well as their average interval lengths.

4.2 Conditional posterior distributions

Consider the posterior distribution under the prior distribution $\pi_R(\boldsymbol{\psi})$ as given in equation 3.36. The conditional posterior of λ_0 given β and η is given by:

$$\pi_R(\lambda_0 \mid \beta, \eta, \boldsymbol{t}) \propto \lambda_0^{n\beta-1} exp\left\{-\lambda_0^\beta \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i}\right\}, \tag{4.1}$$

therefore: $\lambda_0^\beta \sim Gamma\left(\alpha^\star = n, \beta^\star = \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^\beta \eta^{\beta\delta_i}\right).$

Sampling for the models will be completed using $WinBUGS$ (Spiegelhalter et al., 2002), which is unable to generate samples from λ_0^β. Hence the slice sampler will be required to obtain samples from λ_0.

Furthermore consider the conditional posterior of η given λ_0 and β:

$$\pi_R(\eta \mid \lambda_0, \beta, \boldsymbol{t}) \propto \eta^{\beta\bar{\delta}-1} exp\left\{-\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^\beta \lambda_0^\beta \eta^{\beta\delta_i}\right\} \Psi(1, \infty), \tag{4.2}$$

where: $\Psi(1, \infty)$ is an indicator function and the conditional posterior is of an unknown form. Since the conditional posterior distribution is of an unknown form, the slice sampler will be used to simulate observations of η.

Finally consider the conditional posterior of β given λ_0 and η:

$$\pi_R\left(\beta \mid \lambda_0, \eta, \boldsymbol{t}\right) \propto \beta^{n-m-1} \lambda_0^{n\beta} \eta^{\beta\bar{\delta}} \prod_{i=1}^{k}\prod_{j=1}^{n_i} t_{ij}^{\beta} exp\left\{ - \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta} \lambda_0^{\beta} \eta^{\beta\delta_i} \right\}, \qquad (4.3)$$

which is of an unknown form.

Since the conditional posterior of β given λ_0 and η provides a posterior distribution of an unknown form, The ARS sampling method will be used. The ARS sampling method requires that the conditional posterior of $\beta \mid \lambda_0, \eta$ derived above is log-concave, which will be shown in appendix B.4.

4.2.1   Specifying prior distributions in $WinBUGS$

MCMC Bayesian estimates will be found using the statistical software, $WinBUGS$ (Spiegelhalter et al., 2002), and the code used for sampling is provided in appendix: B.8, B.9 and B.10.

When sampling we define independent priors for the unknown parameters, and therefore the posterior distribution under the prior $\pi_R(\boldsymbol{\psi})$ may be written as:

$$\pi_R\left(\boldsymbol{\psi} \mid \boldsymbol{t}\right) \propto \pi_R\left(\lambda_0\right) \times \pi_R\left(\eta\right) \times \pi_R\left(\beta\right) \times L\left(\boldsymbol{\psi} \mid \boldsymbol{t}\right), \qquad (4.4)$$

where:

- $L\left(\boldsymbol{\psi} \mid \boldsymbol{t}\right)$ is the likelihood function for the Weibull distribution under the log-linear transformation function,

- $\pi_R\left(\lambda_0\right) = \frac{1}{\lambda_0}$ is the prior distribution for λ_0,

- $\pi_R\left(\eta\right) = \frac{1}{\eta}$ is the prior distribution for η and

- $\pi_R\left(\beta\right) = \frac{1}{\beta^m}$, $m = -1, 0, 1$ is the prior distribution for β.

When sampling with $WinBUGS$, the distributions of the priors need to be fully specified. Following Lunn et al. (2012), when sampling from: $\pi_R\left(\lambda_0\right)$, $\pi_R\left(\eta\right)$ and $\pi_R\left(\beta\right)$ when $m = 1$, we approximate the prior distributions with the a gamma distribution with small parameters to ensure that the distribution is reasonably flat. The prior distributions in this case will be approximated by: $\psi_i \sim Gamma\left(0.001, 0.001\right)$, where: ψ_i is the ith parameter in $\boldsymbol{\psi}$.

When sampling from $\pi_R\left(\beta\right)$ when $m = 0$, the prior may be approximated with a uniform prior with a suitably wide range between parameters, say: $\beta \sim Uniform\left(0, 100\right)$ (Lunn et al., 2012).

Finally, when sampling from $\pi_R(\beta)$ when $m = -1$, there is no statistical distribution found in $WinBUGS$ which may approximate the prior. Therefore the "zeros trick" will be implemented to specify a new distribution to approximate this prior (Lunn et al., 2012).

To use this "trick", first invent an arbitrary observation $z = 0$, assumed to come from a $\sim Poisson(\phi = -log(\pi_R(\beta)))$. The "trick" is based on the observation that a single Poisson observation equal to zero with mean: $\phi = -log(\pi_R(\beta))$ contributes a term: $\pi_R(\phi) = exp\{-\phi\}$ to the likelihood. Therefore, when ϕ is replaced with a flat prior for β, the correct prior distribution results.

Lunn et al. (2012) states that the issues with using this method is that it may lead to:

- slow computations,

- poor convergence,

- high autocorrelation and

- high MC error values.

4.3 Sampling steps for simulations

The following steps will be used to sample from the posterior under the prior $\pi_R(\boldsymbol{\psi})$ under stress level S_i:

1. Generate n_i values from a $Weibull\left(\nu(S_i)^\star = \lambda_0\eta^{\delta_i}, \beta^\star = \beta\right)$.

2. Choose initial values given by $\boldsymbol{\psi}^{(0)} = \left(\lambda_0^{(0)}, \eta^{(0)}, \beta^{(0)}\right)$.

3. Sample from $\pi_R\left(\lambda_0^{(i)} \mid \beta^{(i-1)}, \eta^{(i-1)}, \boldsymbol{t}\right)$ via the slice sampler.

4. Sample from $\pi_R\left(\eta^{(i)} \mid \lambda_0^{(i-1)}, \beta^{(i-1)}, \boldsymbol{t}\right)$ via the slice sampler.

5. Sample from $\pi_R\left(\beta^{(i)} \mid \lambda_0^{(i-1)}, \eta^{(i-1)}, \boldsymbol{t}\right)$ via the ARS algorithm.

6. Repeat steps 3, 4 and 5 for every value of i.

4.4 Lindley's approximation

Lindley's approximation is a method to finding Bayesian estimates, and is described in detail in section 2.4.6.1. According to Jung and Chung (2018), equation 2.57 reduces to the following

when a model with three parameters is considered:

$$
E\left(w\left(\boldsymbol{\psi}\right)\mid\boldsymbol{t}\right)\approx w\left(\hat{\boldsymbol{\psi}}\right)+U\left(\hat{\boldsymbol{\psi}}\right)+\rho_1\left(\hat{\boldsymbol{\psi}}\right)A_{123}+\rho_2\left(\hat{\boldsymbol{\psi}}\right)A_{213}+\rho_3\left(\hat{\boldsymbol{\psi}}\right)A_{321}+\ldots
$$
$$
+\frac{1}{2}\left[\mathcal{L}_{300}B_{123}+\mathcal{L}_{030}B_{213}+\mathcal{L}_{003}B_{321}+2\mathcal{L}_{111}\left(C_{123}+C_{213}+C_{312}\right)+\ldots\right.
$$
$$
\left.+\ \mathcal{L}_{210}D_{123}+\mathcal{L}_{201}D_{132}+\mathcal{L}_{120}D_{213}+\mathcal{L}_{102}D_{312}+\mathcal{L}_{021}D_{231}+\mathcal{L}_{012}D_{321}\right],
$$

$$(4.5)$$

where:

- $\hat{\boldsymbol{\psi}}=\left(\hat{\lambda}_0,\hat{\eta},\hat{\beta}\right)$ are the MLEs of $\boldsymbol{\psi}$,

- $\mathcal{L}_{ijk}=\left.\frac{\partial^3\mathcal{L}}{\partial\lambda_0^i\partial\eta^j\partial\beta^k}\right|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}}$, where $i,j,k=0,1,2,3$ and $i+j+k=3$,

- $\sigma_{ij}\left(\hat{\boldsymbol{\psi}}\right)$ is the $(i,j)^{th}$ element of the minus inverse Hessian evaluated at $\hat{\boldsymbol{\psi}}$,

- $U\left(\hat{\boldsymbol{\psi}}\right)=\frac{1}{2}\sum\limits_{i=1}^{3}\sum\limits_{j=1}^{3}w_{ij}\left(\hat{\boldsymbol{\psi}}\right)\sigma_{ij}\left(\hat{\boldsymbol{\psi}}\right)$,

- For $i,j,k=1,2,3$:

$$
A_{ijk}=w_i\sigma_{ii}\left(\hat{\boldsymbol{\psi}}\right)+w_j\sigma_{ji}\left(\hat{\boldsymbol{\psi}}\right)+w_k\sigma_{ki}\left(\hat{\boldsymbol{\psi}}\right)
$$
$$
B_{ijk}=\sigma_{ii}\left(\hat{\boldsymbol{\psi}}\right)\left[w_i\sigma_{ii}\left(\hat{\boldsymbol{\psi}}\right)+w_j\sigma_{ij}\left(\hat{\boldsymbol{\psi}}\right)+w_k\sigma_{ik}\left(\hat{\boldsymbol{\psi}}\right)\right]
$$
$$
C_{ijk}=w_i\left[\sigma_{ii}\left(\hat{\boldsymbol{\psi}}\right)\sigma_{jk}\left(\hat{\boldsymbol{\psi}}\right)+2\sigma_{ij}\left(\hat{\boldsymbol{\psi}}\right)\sigma_{ik}\left(\hat{\boldsymbol{\psi}}\right)\right]
$$
$$
D_{ijk}=3w_i\sigma_{ii}\left(\hat{\boldsymbol{\psi}}\right)\sigma_{ij}\left(\hat{\boldsymbol{\psi}}\right)+w_j\left(\sigma_{ii}\left(\hat{\boldsymbol{\psi}}\right)\sigma_{ij}\left(\hat{\boldsymbol{\psi}}\right)+2\sigma_{ij}^2\left(\hat{\boldsymbol{\psi}}\right)\right)+\ldots
$$
$$
+\ w_k\left(\sigma_{ii}\left(\hat{\boldsymbol{\psi}}\right)\sigma_{jk}\left(\hat{\boldsymbol{\psi}}\right)+2\sigma_{ij}\left(\hat{\boldsymbol{\psi}}\right)\sigma_{ik}\left(\hat{\boldsymbol{\psi}}\right)\right),
$$

- $w_i\left(\hat{\boldsymbol{\psi}}\right)=\left.\frac{\partial w(\boldsymbol{\psi})}{\partial\psi_i}\right|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}}$, where ψ_i is the ith element in $\boldsymbol{\psi}$ for $i=1,2,3$,

- $w_{ij}\left(\hat{\boldsymbol{\psi}}\right)=\left.\frac{\partial^2 w(\boldsymbol{\psi})}{\partial\psi_i\partial\psi_j}\right|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}}$,

- $\rho_i\left(\hat{\boldsymbol{\psi}}\right)=\left.\frac{\partial\rho(\boldsymbol{\psi})}{\partial\psi_i}\right|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}}$ and

- $\rho=log\left(\pi\left(\boldsymbol{\psi}\right)\right)$, for some prior distribution $\pi\left(\boldsymbol{\psi}\right)$.

Consider the third-order partial derivatives of the log-likelihood described in equation 3.8:

$$
\frac{\partial^3\mathcal{L}}{\partial\lambda_0^3}=\frac{2n\beta}{\lambda_0^3}-\sum_{i=1}^{k}\sum_{j=1}^{n_i}t_{ij}^{\beta}\beta\left(\beta-1\right)\left(\beta-2\right)\lambda_0^{\beta-3}\eta^{\beta\delta_i},
$$

$$(4.6)$$

$$
\frac{\partial^3\mathcal{L}}{\partial\eta^3}=\frac{2\bar{\delta}\beta}{\eta^3}-\sum_{i=1}^{k}\sum_{j=1}^{n_i}t_{ij}^{\beta}\lambda_0^{\beta}\beta\delta_i\left(\beta\delta_i-1\right)\left(\beta\delta_i-2\right)\eta^{\beta\delta_i-3},
$$

$$(4.7)$$

$$\frac{\partial^3 \mathcal{L}}{\partial \beta^3} = \frac{2n}{\beta^3} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(t_{ij}\lambda_0\eta^{\delta_i}\right)^{\beta} log^3\left(t_{ij}\lambda_0\eta^{\delta_i}\right) \quad, \tag{4.8}$$

$$\frac{\partial^3 \mathcal{L}}{\partial \lambda_0^2 \partial \eta} = -\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\beta^2\delta_i\left(\beta-1\right)\lambda_0^{\beta-2}\eta^{\beta\delta_i-1}, \tag{4.9}$$

$$\frac{\partial^3 \mathcal{L}}{\partial \lambda_0^2 \partial \beta} = -\frac{n}{\lambda_0^2} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta-2}\eta^{\beta\delta_i}\left(\left(\left(log\left(t_{ij}\lambda_0\eta^{\delta_i}\right)\right)\left(\beta-1\right)+2\right)\beta-1\right), \tag{4.10}$$

$$\frac{\partial^3 \mathcal{L}}{\partial \eta^2 \partial \lambda_0} = -\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta-1}\beta^2\delta_i\left(\beta\delta_i-1\right)\eta^{\beta\delta_i-2}, \tag{4.11}$$

$$\frac{\partial^3 \mathcal{L}}{\partial \eta^2 \partial \beta} = -\frac{\bar{\delta}}{\eta^2} - \frac{1}{\eta^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \delta_i t_{ij}^{\beta}\lambda_0^{\beta}\eta^{\beta\delta_i}\left[\beta\left(\delta_i\left(log\left(\eta\right)\left(\delta_i\beta-1\right)+2\right)+\left(\delta_i\beta-1\right)log\left(t_{ij}\lambda_0\right)\right)-1\right], \tag{4.12}$$

$$\frac{\partial^3 \mathcal{L}}{\partial \beta^2 \partial \lambda_0} = -\frac{1}{\lambda_0}\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta}\eta^{\beta\delta_i} log\left(t_{ij}\lambda_0\eta^{\delta_i}\right)\left(\beta log\left(t_{ij}\lambda_0\eta^{\delta_i}\right)+2\right), \tag{4.13}$$

$$\frac{\partial^3 \mathcal{L}}{\partial \beta^2 \partial \eta} = -\frac{1}{\eta}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \delta_i t_{ij}^{\beta}\lambda_0^{\beta}\eta^{\beta\delta_i} log\left(t_{ij}\lambda_0\eta^{\delta_i}\right)\left(\beta log\left(t_{ij}\lambda_0\eta^{\delta_i}\right)+2\right), \tag{4.14}$$

$$\frac{\partial^3 \mathcal{L}}{\partial \lambda_0 \partial \eta \partial \beta} = -\frac{1}{\lambda_0\eta}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \beta\delta_i t_{ij}^{\beta}\lambda_0^{\beta}\eta^{\beta\delta_i}\left(\beta log\left(t_{ij}\lambda_0\eta^{\delta_i}\right)+2\right). \tag{4.15}$$

Therefore, under any prior, posterior estimates for a parameter can be estimated by choosing $w\left(\boldsymbol{\psi}\right)$ appropriately.

When estimates for the squared error loss function are obtained, $w\left(\hat{\psi}_i\right) = \hat{\psi}_i$ for any parameter $\hat{\psi}_i$ in $\hat{\boldsymbol{\psi}}$. Then $w_i\left(\hat{\psi}_i\right)$ and $w_{ii}\left(\hat{\psi}_i\right)$ become:

- $\frac{\partial w(\psi_i)}{\partial \psi_i} = 1$ and

- $\frac{\partial^2 w(\psi_i)}{\partial \psi_{ii}} = 0$,

When any parameter not $\hat{\psi}_i$ in $\hat{\boldsymbol{\psi}}$, denoted by $\hat{\psi}_{-i}$, $w_{-i}\left(\hat{\psi}_{-i}\right)$ and $w_{ij}\left(\hat{\psi}_{-i}\right)$ become:

- $\frac{\partial w(\psi_{-i})}{\partial \psi_{-i}} = 0$ and

- $\frac{\partial w(\psi_{-i})}{\partial \psi_{ij}} = 0$.

When estimates for the LINEX loss function are obtained, $w\left(\hat{\psi}_i\right) = exp\left\{-a\hat{\psi}_i\right\}$ for any parameter $\hat{\psi}_i$ in $\hat{\boldsymbol{\psi}}$. Then $w_i\left(\hat{\psi}_i\right)$ and $w_{ii}\left(\hat{\psi}_i\right)$ become:

- $\frac{\partial w(\psi_i)}{\partial \psi_i} = -a \times exp\left\{-a\psi_i\right\}$ and

- $\frac{\partial^2 w(\psi_i)}{\partial \psi_{ii}} = a^2 \times exp\left\{-a\psi_i\right\}$,

for any $a \neq 0$. When any parameter not $\hat{\psi}_i$ in $\hat{\boldsymbol{\psi}}$, denoted by $\hat{\psi}_{-i}$, $w_{-i}\left(\hat{\psi}_{-i}\right)$ and $w_{ij}\left(\hat{\psi}_{-i}\right)$ become:

- $\frac{\partial w(\psi_{-i})}{\partial \psi_{-i}} = 0$ and

- $\frac{\partial w(\psi_{-i})}{\partial \psi_{ij}} = 0$.

When estimates for the GELF are obtained, $w\left(\hat{\psi}_i\right) = \hat{\psi}_i^{-k}$ for any parameter $\hat{\psi}_i$ in $\hat{\boldsymbol{\psi}}$. Then $w_i\left(\hat{\psi}_i\right)$ and $w_{ii}\left(\hat{\psi}_i\right)$ become:

- $\frac{\partial w(\psi_i)}{\partial \psi_i} = -k \times \psi_i^{-k-1}$ and

- $\frac{\partial^2 w(\psi_i)}{\partial \psi_{ii}} = -k \times (-k-1) \times \psi_i^{-k-2}$,

for any $k \neq 0$. When any parameter not $\hat{\psi}_i$ in $\hat{\boldsymbol{\psi}}$, denoted by $\hat{\psi}_{-i}$, $w_{-i}\left(\hat{\psi}_{-i}\right)$ and $w_{ij}\left(\hat{\psi}_{-i}\right)$ become:

- $\frac{\partial w(\psi_{-i})}{\partial \psi_{-i}} = 0$ and

- $\frac{\partial w(\psi_{-i})}{\partial \psi_{ij}} = 0$.

4.5   Simulation study

Consider an example with a three-level, constant stress ALT using a complete dataset. Assume that the data comes from a Weibull distribution with constant shape parameter β and scale parameter $\nu(S_i)$ dependent on the stress level S_i.

The three levels under consideration are:

$$\boldsymbol{s}' = (S_1, S_2, S_3)' = (200, 250, 300)', \tag{4.16}$$

where the stress applied is temperature measured in Kelvin (K), and it is assumed that the use stress level is $S_0 = 150$. Since the data is related to temperature, the Arrhenius model is required. Therefore the shape parameter is given by:

$$log\left(\nu\left(S_i\right)\right) = \theta_1 + \frac{\theta_2}{S_i}, \tag{4.17}$$

where it is assumed that $\theta_1 = 3$ and $\theta_2 = 1000$.

Therefore $\nu(S_i)$ values are given by: $\nu(S_1) = 2980.9580$, $\nu(S_2) = 1096.6330$ and $\nu(S_3) = 563.0302$. Moreover it is also assumed that the shape parameter $\beta = 1$, suggesting a constant failure rate.

Following the transformations used in chapter 3, $\lambda_0 = 6.3361 \times 10^{-5}$, $\eta = 5.2945$ and $\beta = 1$. Under this transformation the data follows a $Weibull\left(\nu\left(S_i\right)^\star = \lambda_0 \eta^{\delta_i}, \beta^\star = \beta\right)$, where:

$$\delta_i = \frac{\frac{1}{150} - \frac{1}{S_i}}{\frac{1}{150} - \frac{1}{200}}, \quad i = 1, 2, 3. \tag{4.18}$$

For this simulation study, the sizes of n_i are equal for $i = 1, 2, 3$ and the values considered are: $n_i = (10, 20, 30, 40, 50)$, that is $n = (30, 60, 90, 120, 150)$.

The aim of this section is to compare the estimates of the data obtained from the posteriors under the priors: $\pi_J\left(\psi\right)$, $\pi_{R1}\left(\psi\right)$ and $\pi_{R2}\left(\psi\right)$ (from here on using the short-hand notation: π_J, π_{R1} and π_{R2}), using three loss functions: squared error, LINEX and the GELF. The comparison among the estimates were made using their RMSE values, defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{N}\left(\hat{\psi}_i - \psi\right)^2}, \tag{4.19}$$

where $\hat{\psi}_i$ is the ith estimator of ψ and n is the total number of estimates obtained.

Two classes of estimates will be considered: Bayesian estimates and MLEs. Bayesian estimates will be obtained via both an MCMC procedure and by Lindley's approximation method. MLEs will be found using the R package $maxLik$ (Henningsen and Toomet, 2011), via the Newton-Raphson method. The MCMC Bayes' estimates were found using the $R^{®}$ package $R2WinBUGS$ (Sturtz et al., 2005), a package that allows one to operate $WinBUGS$ within $R^{®}$. Here the MLEs will be denoted by ψ_{MLE}, Bayesian estimates obtained via the MCMC procedure as ψ_{MC} and Bayesian estimates found via Lindley's approximation technique as ψ_{LIN}.

The estimates under the squared error loss function will be denoted by $\psi_{(S)}$, the estimates under the LINEX loss function will be denoted by $\psi_{(L)}$ and the estimates under the GELF will be denoted by $\psi_{(G)}$. The values of a for the LINEX loss parameter considered are: $a = \pm 0.5$ and $a = \pm 1.5$, and the values of k for the general entropy loss parameter are: $k = \pm 0.5$ and $k = \pm 1.5$.

Furthermore the estimated values highlighted in blue represent the smallest RMSE values under a given sample size, and the estimated values highlighted in red represent the largest RMSE values.

A full test for convergence for the Markov Bayes' estimates is provided in chapter 5. Each test scheme was repeated a total of 2000 times before results were obtained.

This section shows the results for the posterior under the prior π_J; the results for the posteriors under prior π_{R1} and π_{R2} can be found in the appendix.

4.5.1   Estimated values

Tables: 4.1, 4.2 and 4.3 present the posterior means and the expected values for the three parameters subject to different parameter estimation methods for the posterior distributions under π_J. As n gets large all three parameters tend towards their true values.

For all three parameters the estimates found using Lindley's method are closest to the true value, followed by the MLE values and then by the MCMC estimates.

Tables: B.1, B.2 and B.3 provide the posterior means and expected values for π_{R1} and the tables B.9, B.10 and B.11 provide the posterior means and expected values for π_{R2}. The results for these two priors are the same as the results under prior π_J (Lindley's methods provides the estimates closest to the true value; MCMC estimates are the furthest away).

Table 4.1: Mean values for λ_0 $\left(\times 10^{-5}\right)$ under prior π_J.

n	MCMC	Lindley	MLE
30	10.3067	8.3629	8.6676
60	8.1239	7.2846	7.4481
90	7.3056	6.8966	7.0031
120	7.1508	6.8673	6.9508
150	7.0207	6.6950	6.7550

Table 4.2: Mean values for η under prior π_J.

n	MCMC	Lindley	MLE
30	6.0776	5.6198	5.7617
60	5.8411	5.4145	5.5028
90	5.7122	5.3925	5.4583
120	5.5547	5.2947	5.3439
150	5.4958	5.3440	5.3755

Table 4.3: Mean values for β under prior π_J.

n	MCMC	Lindley	MLE
30	1.0728	1.0499	1.0646
60	1.0304	1.0233	1.0318
90	1.0192	1.0024	1.0201
120	1.0172	1.0105	1.0150
150	1.0120	1.0083	1.0124

4.5.2 Root mean squared error values under different loss functions

This section shows the RMSE values for the estimates under the three above-mentioned loss functions. Again as expected, as n becomes larger so the RMSE values decrease.

The RMSE values for λ_0 as illustrated in tables: 4.4, B.4 and B.12 show the lowest RMSE values for all priors tends to be provided by the MCMC estimates with the GELF when the loss parameter $k > 0$. Since the value of the loss parameter is positive it implies that the GELF overestimated the parameter. Furthermore the largest values in all cases are provided by the MCMC estimates GELF when the loss parameter $k < 0$. The range between largest RMSE value and lowest RMSE value is very large for the small dataset $(n = 30)$, however as n becomes larger so that range decreases as well. In all cases the lowest RMSE was provided by the posterior under prior π_{R2}, whereas the highest value was generally provided by π_{R1}.

Table 4.4: RMSE for λ_0 $\left(\times 10^{-5}\right)$ under prior π_J.

n		30	60	90	120	150
$\hat{\lambda}_{0MLE(S)}$		8.5076	4.3328	3.2901	2.6357	2.2828
$\hat{\lambda}_{0MC(S)}$		8.4508	4.9190	3.3356	2.7575	2.3733
$\hat{\lambda}_{0LIN(S)}$		8.2171	4.2107	3.2247	2.5892	2.2541
$\hat{\lambda}_{0MC(L)}$	$a, k = 0.5$	9.1128	4.9185	3.3355	2.7575	2.3733
$\hat{\lambda}_{0MC(G)}$		5.8583	3.7925	2.7969	2.3925	2.1081
$\hat{\lambda}_{0LIN(L)}$		8.0293	4.1314	3.1244	2.5645	2.2261
$\hat{\lambda}_{0LIN(G)}$		8.3094	4.1011	3.0293	2.6432	2.3237
$\hat{\lambda}_{0MC(L)}$	$a, k = 1.5$	8.9180	4.9183	3.3354	2.7574	2.3733
$\hat{\lambda}_{0MC(G)}$		4.5186	3.3305	2.5956	2.3472	2.0104
$\hat{\lambda}_{0LIN(L)}$		7.4882	3.8414	2.9226	2.6232	2.2322
$\hat{\lambda}_{0LIN(G)}$		9.1369	4.3342	3.0996	2.5480	2.2185
$\hat{\lambda}_{0MC(L)}$	$a, k = -0.5$	6.4800	4.9191	3.3356	2.7576	2.3733
$\hat{\lambda}_{0MC(G)}$		7.7308	4.4871	3.1254	2.6170	2.2717
$\hat{\lambda}_{0LIN(L)}$		7.6218	4.3472	3.1756	2.6166	2.2280
$\hat{\lambda}_{0LIN(G)}$		8.0586	4.2457	3.1373	2.6999	2.1050
$\hat{\lambda}_{0MC(L)}$	$a, k = -1.5$	7.8730	4.9192	3.3357	2.7576	2.3733
$\hat{\lambda}_{0MC(G)}$		10.0719	5.4059	3.3575	2.9157	2.4873
$\hat{\lambda}_{0LIN(L)}$		7.4817	4.1113	3.0465	2.6533	2.2004
$\hat{\lambda}_{0LIN(G)}$		8.0111	4.2837	3.1185	2.6551	2.1923

The lowest RMSE values for η (as provided by tables: 4.5, B.5 and B.13) are all provided by

the MCMC estimates with the LINEX loss function when the loss parameter $a > 0$. This, like the estimates for λ_0, implies overestimation of the parameter values. The largest RMSE values in all cases are provided by the MCMC estimates with the LINEX loss function when the loss parameter $a < 0$. The range between largest and lowest RMSE values is very large and is even greater in percent terms in comparison to the range for λ_0. The RMSE values were usually the lowest for the posterior under prior π_J, and were always the highest when the prior π_{R2} was used.

Table 4.5: RMSE for η under prior π_J.

n		30	60	90	120	150
$\hat{\eta}_{MLE(S)}$		2.7736	1.8674	1.4545	1.2004	1.0985
$\hat{\eta}_{MC(S)}$		2.4974	1.8997	1.4932	1.2745	1.1029
$\hat{\eta}_{LIN(S)}$		2.7034	1.8358	1.4344	1.1893	1.0915
$\hat{\eta}_{MC(L)}$	$a, k = 0.5$	1.7760	1.4868	1.2147	1.0996	0.9803
$\hat{\eta}_{MC(G)}$		2.1933	1.7231	1.3750	1.2036	1.0572
$\hat{\eta}_{LIN(L)}$		2.8638	1.7533	1.4198	1.2462	1.0822
$\hat{\eta}_{LIN(G)}$		2.6669	1.7732	1.4588	1.1982	1.0598
$\hat{\eta}_{MC(L)}$	$a, k = 1.5$	1.7980	1.4798	1.2039	1.0965	0.9083
$\hat{\eta}_{MC(G)}$		2.0895	1.6542	1.3265	1.1758	1.0338
$\hat{\eta}_{LIN(L)}$		2.8833	1.8029	1.4556	1.1794	1.1045
$\hat{\eta}_{LIN(G)}$		2.7296	1.7647	1.4724	1.1800	1.0344
$\hat{\eta}_{MC(L)}$	$a, k = -0.5$	4.8633	3.0725	2.1871	1.6896	1.3808
$\hat{\eta}_{MC(G)}$		2.3779	1.8314	1.4479	1.2470	1.0833
$\hat{\eta}_{LIN(L)}$		2.8932	1.8380	1.4132	1.2019	1.1112
$\hat{\eta}_{LIN(G)}$		2.8222	1.7609	1.4071	1.2042	1.0843
$\hat{\eta}_{MC(L)}$	$a, k = -1.5$	9.1119	6.1807	4.4204	3.3172	2.5684
$\hat{\eta}_{MC(G)}$		2.6327	1.9769	1.5441	1.3058	1.1252
$\hat{\eta}_{LIN(L)}$		2.9172	1.7189	1.3877	1.2053	1.1013
$\hat{\eta}_{LIN(G)}$		2.8938	1.8080	1.4174	1.2099	1.0783

Tables: 4.6, B.6 and B.14 show the RMSE values for β. The lowest RMSE value is generally given by the MCMC estimates with the GELF when the loss parameter $k > 0$. Again suggesting overestimation of the parameter values. The highest RMSE values are provided by the Lindley estimates when using both the GELF and LINEX loss function with loss parameters $a, k > 0$. The major difference between the RMSE values for β in comparison to λ_0 and η is that all the RMSE are similar – possibly suggesting symmetry in the distribution. The prior π_{R2} generally provided the lowest RMSE values, whereas the highest RMSE values

were generally given by π_J.

Table 4.6: RMSE for β under prior π_J.

n		30	60	90	120	150
$\hat{\beta}_{MLE(S)}$		0.1735	0.1110	0.0870	0.0742	0.0657
$\hat{\beta}_{MC(S)}$		0.1749	0.1127	0.0865	0.0745	0.0667
$\hat{\beta}_{LIN(S)}$		0.1664	0.1089	0.0850	0.0729	0.0648
$\hat{\beta}_{MC(L)}$	$a, k = 0.5$	0.1707	0.1141	0.0858	0.0740	0.0664
$\hat{\beta}_{MC(G)}$		0.1661	0.1099	0.0850	0.0734	0.0660
$\hat{\beta}_{LIN(L)}$		0.1769	0.1091	0.0860	0.0745	0.0648
$\hat{\beta}_{LIN(G)}$		0.1748	0.1091	0.0903	0.0751	0.0661
$\hat{\beta}_{MC(L)}$	$a, k = 1.5$	0.1629	0.1090	0.0846	0.0731	0.0658
$\hat{\beta}_{MC(G)}$		0.1609	0.1084	0.0843	0.0728	0.6570
$\hat{\beta}_{LIN(L)}$		0.1613	0.1088	0.0845	0.0761	0.0639
$\hat{\beta}_{LIN(G)}$		0.1715	0.1085	0.0869	0.0741	0.0651
$\hat{\beta}_{MC(L)}$	$a, k = -0.5$	0.1792	0.1141	0.0872	0.0750	0.0671
$\hat{\beta}_{MC(G)}$		0.1718	0.1117	0.0860	0.0741	0.0665
$\hat{\beta}_{LIN(L)}$		0.1679	0.1099	0.0890	0.0766	0.0667
$\hat{\beta}_{LIN(G)}$		0.1848	0.1086	0.0849	0.0764	0.0655
$\hat{\beta}_{MC(L)}$	$a, k = -1.5$	0.1886	0.1170	0.0887	0.0769	0.0678
$\hat{\beta}_{MC(G)}$		0.1781	0.1137	0.0870	0.0749	0.0670
$\hat{\beta}_{LIN(L)}$		0.1786	0.1147	0.0911	0.0771	0.0658
$\hat{\beta}_{LIN(G)}$		0.1810	0.1123	0.0859	0.0777	0.0667

4.6   Coverage rate

It is desirable to identify regions of the parameter space that are likely to contain the true parameter value. To do this, after observing the data we can construct an in interval $\left[\psi_{\left(\frac{\alpha}{2}\right)}, \psi_{\left(1-\frac{\alpha}{2}\right)}\right]$, such that the probability that: $\psi_{\left(\frac{\alpha}{2}\right)} < \psi < \psi_{\left(1-\frac{\alpha}{2}\right)}$ is large.
Hoff (2009) provides a definition of Bayesian coverage:

Definition 4.1. An interval $\left[\psi_{\left(\frac{\alpha}{2}\right)}, \psi_{\left(1-\frac{\alpha}{2}\right)}\right]$ based on the observed data has $(1-\alpha)\,\%$ Bayesian coverage for ψ if:

$$P\left(\psi_{\left(\frac{\alpha}{2}\right)} < \psi < \psi_{\left(1-\frac{\alpha}{2}\right)} \mid \boldsymbol{t}\right) = 1 - \alpha. \tag{4.20}$$

This differs from the definition of frequentist coverage, which is given by:

Definition 4.2. A random interval $\left[\psi_{\left(\frac{\alpha}{2}\right)}, \psi_{\left(1-\frac{\alpha}{2}\right)}\right]$ has $(1-\alpha)\,\%$ frequentist coverage for ψ, if before data is collected:

$$P\left(\psi_{\left(\frac{\alpha}{2}\right)} < \psi < \psi_{\left(1-\frac{\alpha}{2}\right)} \mid \psi\right) = 1 - \alpha. \tag{4.21}$$

The coverage rate is defined as the proportion of times that the Bayesian credibility or frequentist confidence interval contains the true parameter value. Therefore by defining:

$$P\left(\psi_{\left(\frac{\alpha}{2}\right)} < \psi < \psi_{\left(1-\frac{\alpha}{2}\right)} \mid \psi\right) = \begin{cases} 1 & \psi \in \left[\psi_{\left(\frac{\alpha}{2}\right)}, \psi_{\left(1-\frac{\alpha}{2}\right)}\right] \\ 0 & \psi \notin \left[\psi_{\left(\frac{\alpha}{2}\right)}, \psi_{\left(1-\frac{\alpha}{2}\right)}\right], \end{cases} \tag{4.22}$$

the coverage for ψ is the found by finding the proportion of times that ψ is in the interval: $\left[\psi_{\left(\frac{\alpha}{2}\right)}, \psi_{\left(1-\frac{\alpha}{2}\right)}\right]$.

The coverage rate should be approximately equal to the nominal coverage rate (Newcombe, 1998). In situations where over-coverage occurs the results are too conservative and more simulations are required. On the other hand if under-coverage occurs there is over-confidence in the parameter estimates, and again suggests more simulations are required.

The average lengths of the interval will also be considered in this study, which for parameter ψ is defined as:

$$\ell = \frac{1}{n} \sum_{i=1}^{n} \left(\psi_{i\left(1-\frac{\alpha}{2}\right)} - \psi_{i\left(\frac{\alpha}{2}\right)}\right), \tag{4.23}$$

where n is the number of intervals constructed. Due to the computational complexity of the simulations, the value of n chosen for this simulation study was $n = 2000$.

The preferred credibility and confidence interval is the one with the shortest length and with the a coverage rate closest to its nominal value.

4.6.1   Coverage rate results

This section is devoted to showing the coverage rates for the Bayesian MCMC estimates and MLEs when $\alpha = 0.05$ and $\alpha = 0.1$. That is, obtaining both coverage rates and average interval lengths from 95% and 90% credibility and confidence intervals.

Tables: 4.7, B.7, B.15 and B.17 provide results for the posterior distribution under the priors: π_J, π_{R1} and π_{R2} and results for the MLEs respectively when $\alpha = 0.05$. We expect the coverage rate to be close to its nominal value, that is we expect the coverage rate to be around 0.95. For all the estimates, the results are close to this value.

Table 4.7: 95% Coverage rate for π_J.

n	30	60	90	120	150
λ_0	0.9510	0.9450	0.9500	0.9535	0.9445
$\ell\ (\times 10^{-5})$	(27.8113)	(16.3479)	(11.9246)	(10.0267)	(8.7919)
η	0.9450	0.9420	0.9485	0.9480	0.9485
ℓ	(9.5392)	(7.1794)	(5.8251)	(4.8879)	(4.3290)
β	0.9410	0.9435	0.9475	0.9485	0.9445
ℓ	(0.6033)	(0.4078)	(0.3277)	(0.2835)	(0.2519)

Figures: 4.1, 4.2 and 4.3 illustrate the average length for the three parameters under the different estimation techniques. The average lengths are similar for all the estimates, however when n is small the MLE lengths tend to be larger than their Bayesian counterparts. As n becomes larger the interval lengths become smaller, and the difference in interval length between the different estimators becomes negligible. However the average interval lengths for λ_0 are generally shorter when estimated with the posterior distribution under prior π_J, for η under prior π_J and for β under prior π_{R2}. The percentage change in average length between successive sample sizes also decreases as n becomes larger.
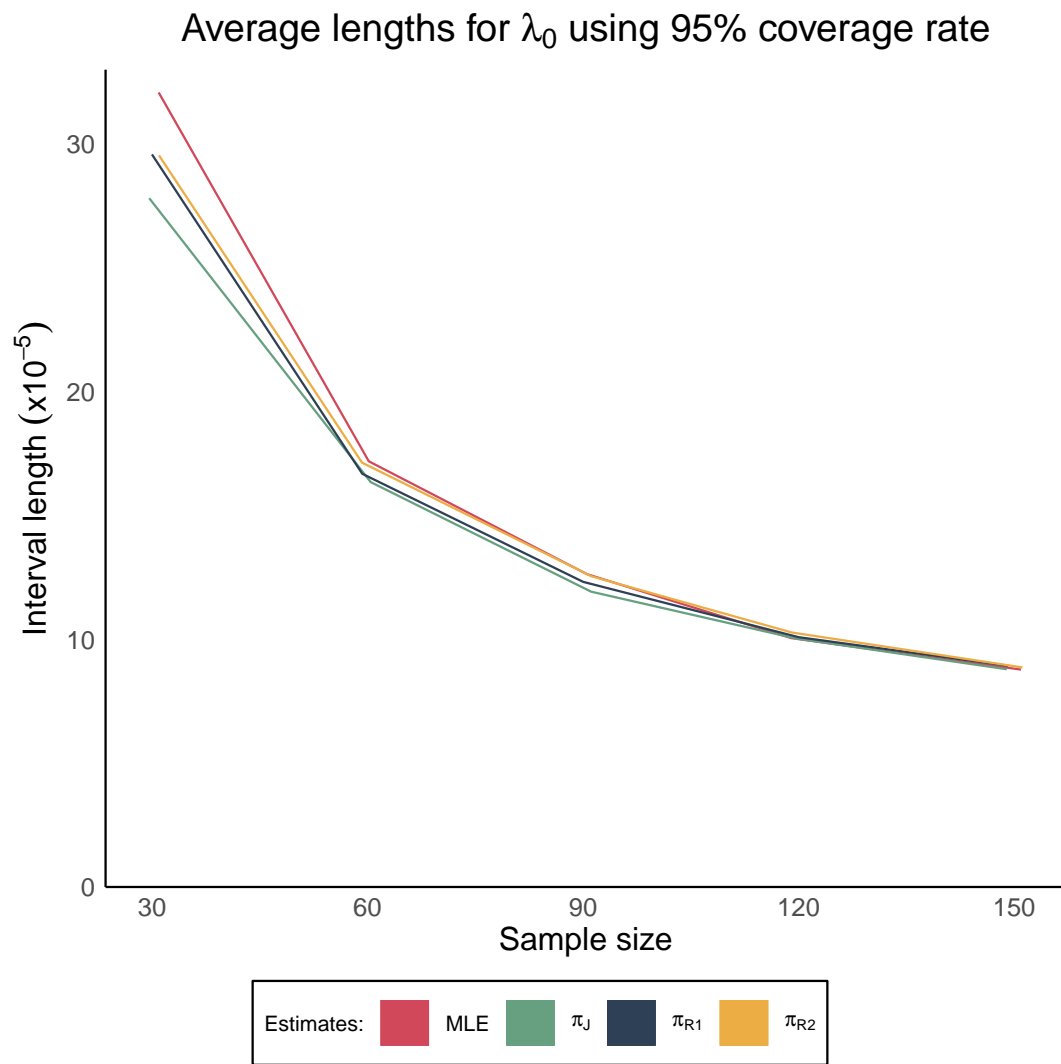
Figure 4.1: Average length for λ_0 for the 95% coverage rate.
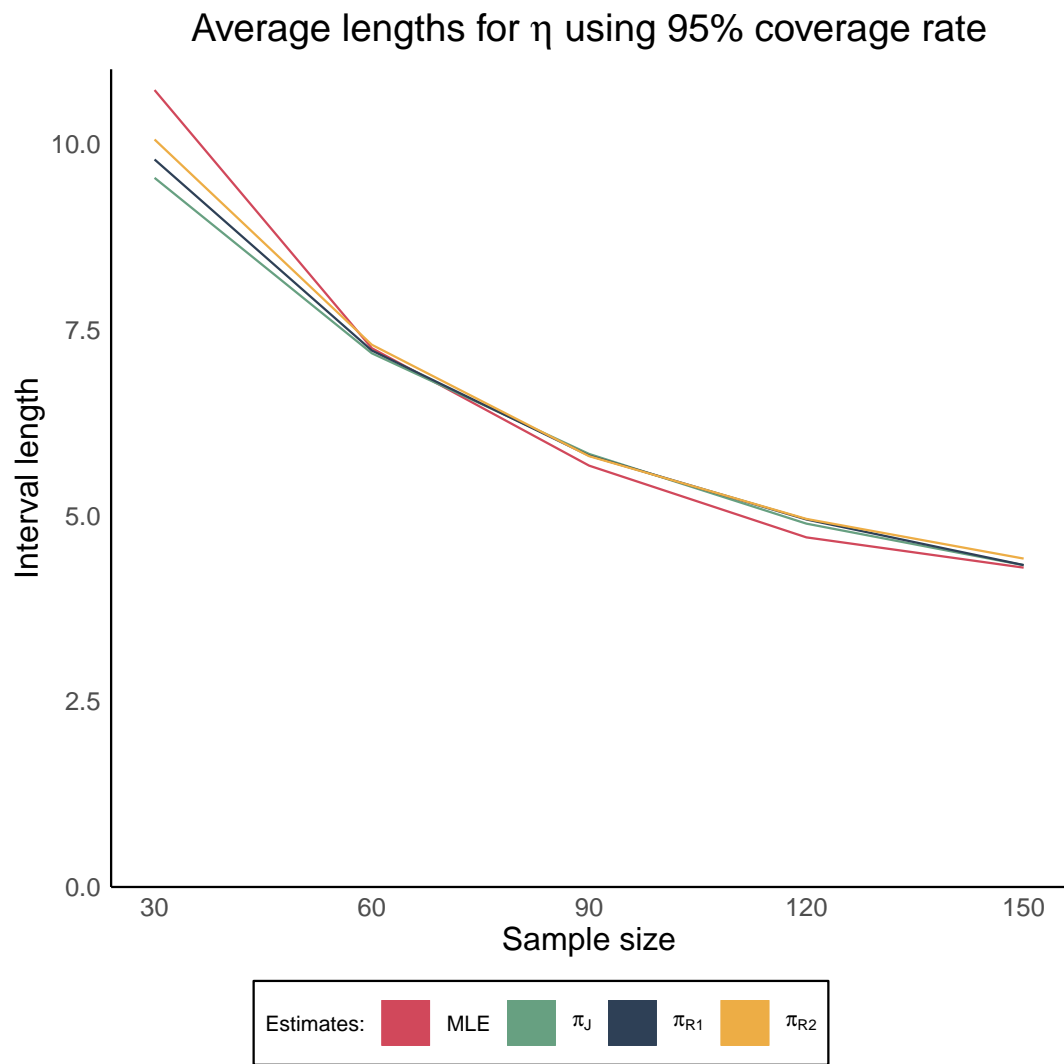
Figure 4.2: Average length for η for the 95% coverage rate.
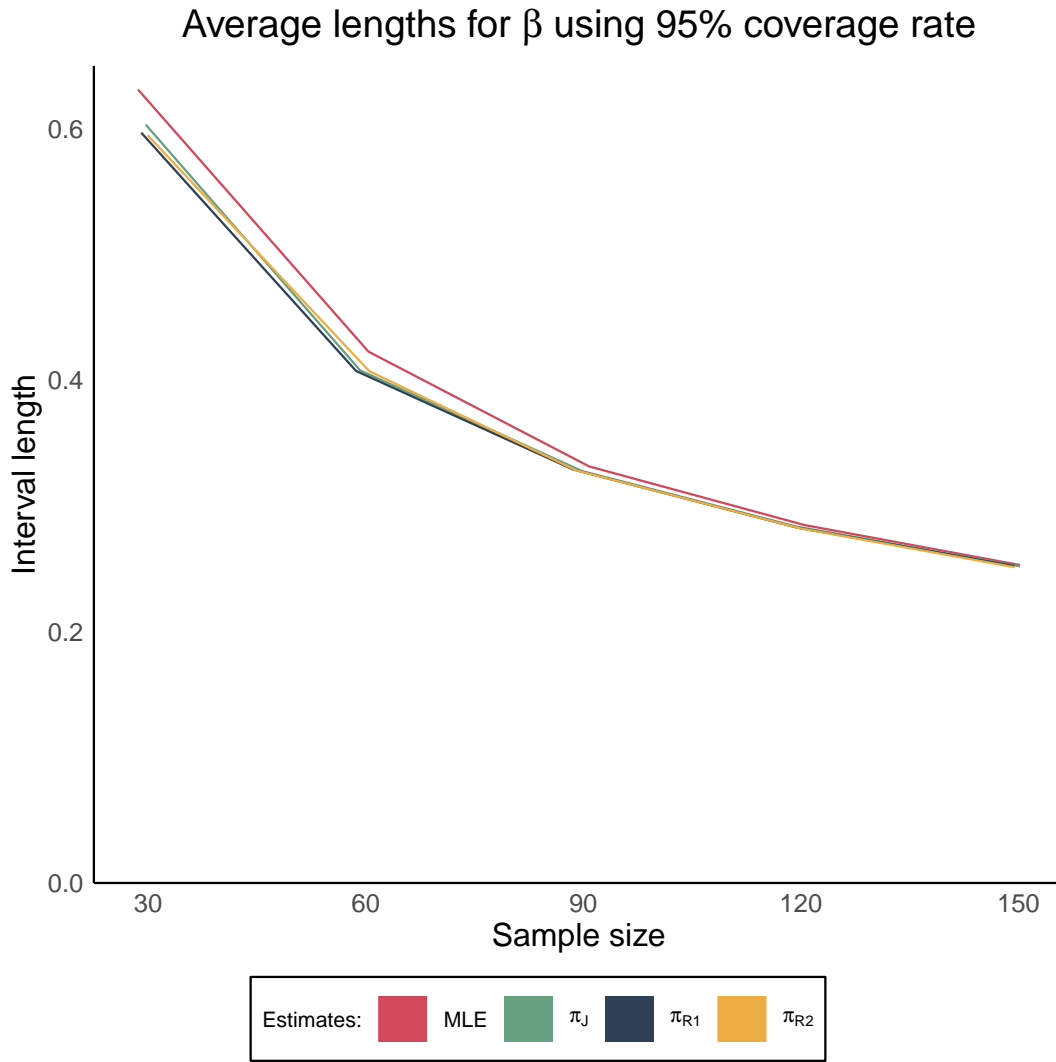
Figure 4.3: Average length for β for the 95% coverage rate.

90% coverage rates and the associated average lengths are presented in tables: 4.8, B.8, B.16 and B.18. The results are again good for all estimators, as the coverage rate is around 0.9.

Table 4.8: 90% Coverage rate for π_J.

n	30	60	90	120	150
λ_0	0.8895	0.8895	0.9015	0.8950	0.9075
ℓ $(\times 10^{-5})$	(23.8845)	(13.1178)	(9.7068)	(8.2354)	(7.2580)
η	0.8940	0.8915	0.9005	0.8940	0.9040
ℓ	(7.9284)	(5.9321)	(4.8183)	(4.0532)	(3.5964)
β	0.8930	0.8935	0.8975	0.9020	0.8910
ℓ	(0.5075)	(0.3431)	(0.2746)	(0.2385)	(0.2119)

Average length plots for the 90% coverage rates are found in figures: 4.4, 4.5 and 4.6. Again, initially the MLE average interval lengths are longer than the Bayesian average interval lengths, but as n becomes large so the difference becomes negligible. However the average

interval lengths for λ_0 are generally shorter when estimated with the posterior distribution under prior π_J, for η under prior π_J and for β under prior π_{R2}.
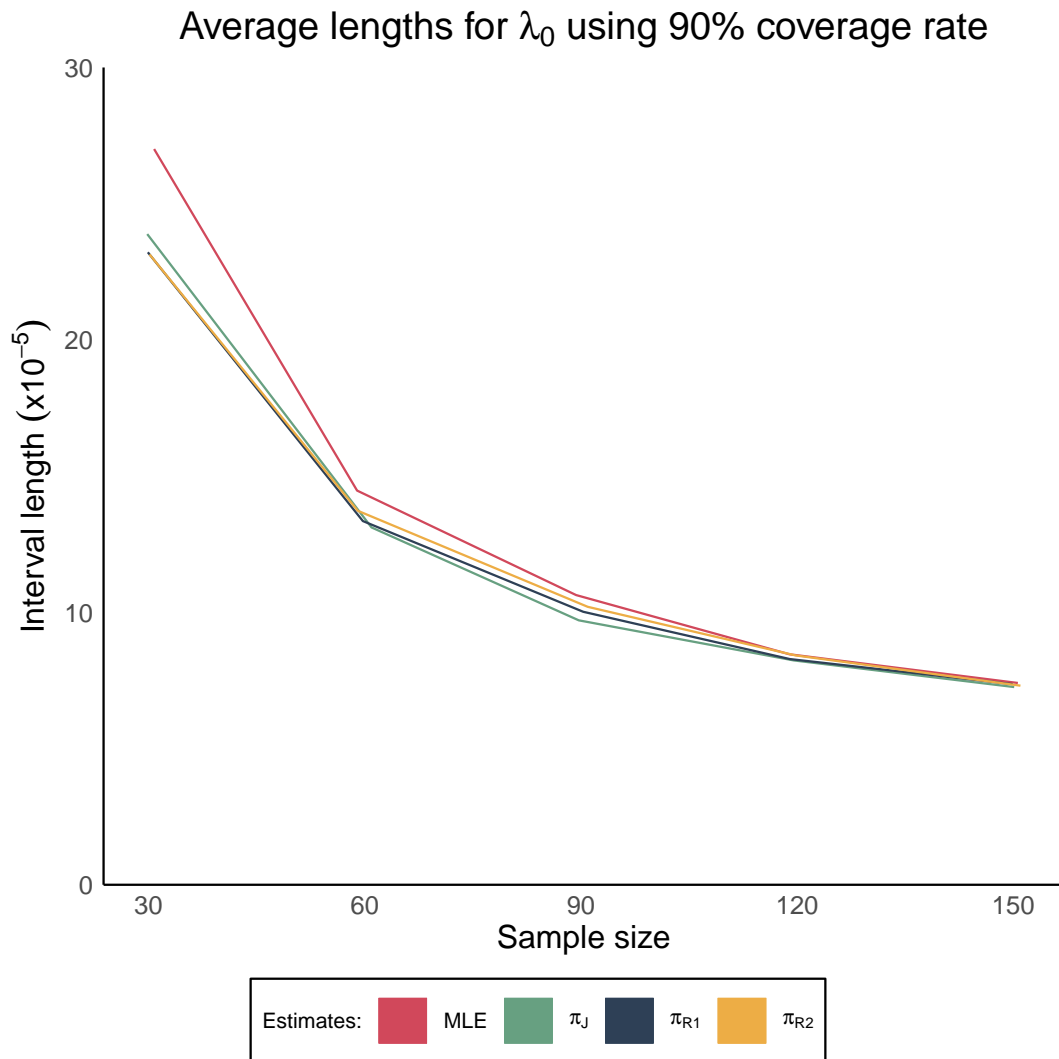
Figure 4.4: Average length for λ_0 for the 90% coverage rate.

Average lengths for η using 90% coverage rate
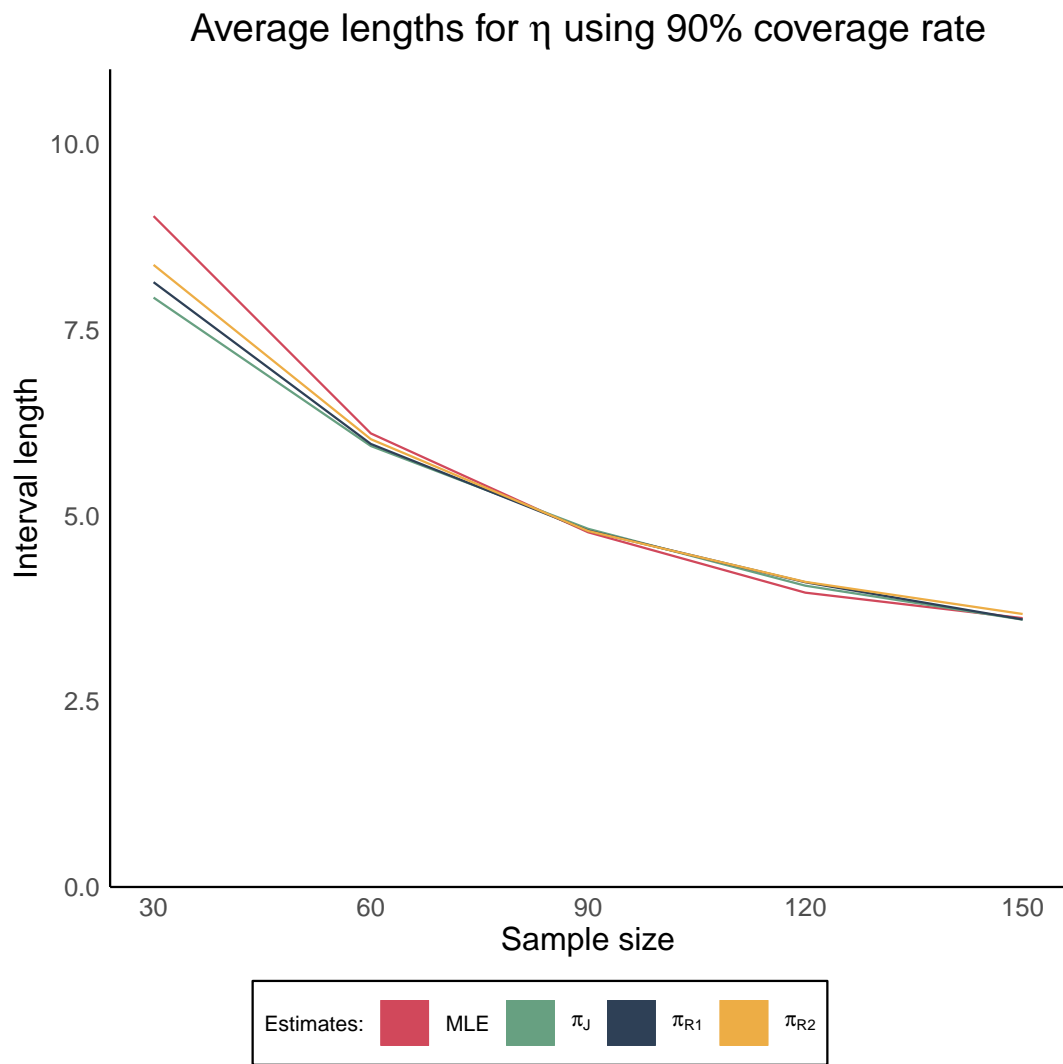
Figure 4.5: Average length for η for the 90% coverage rate.
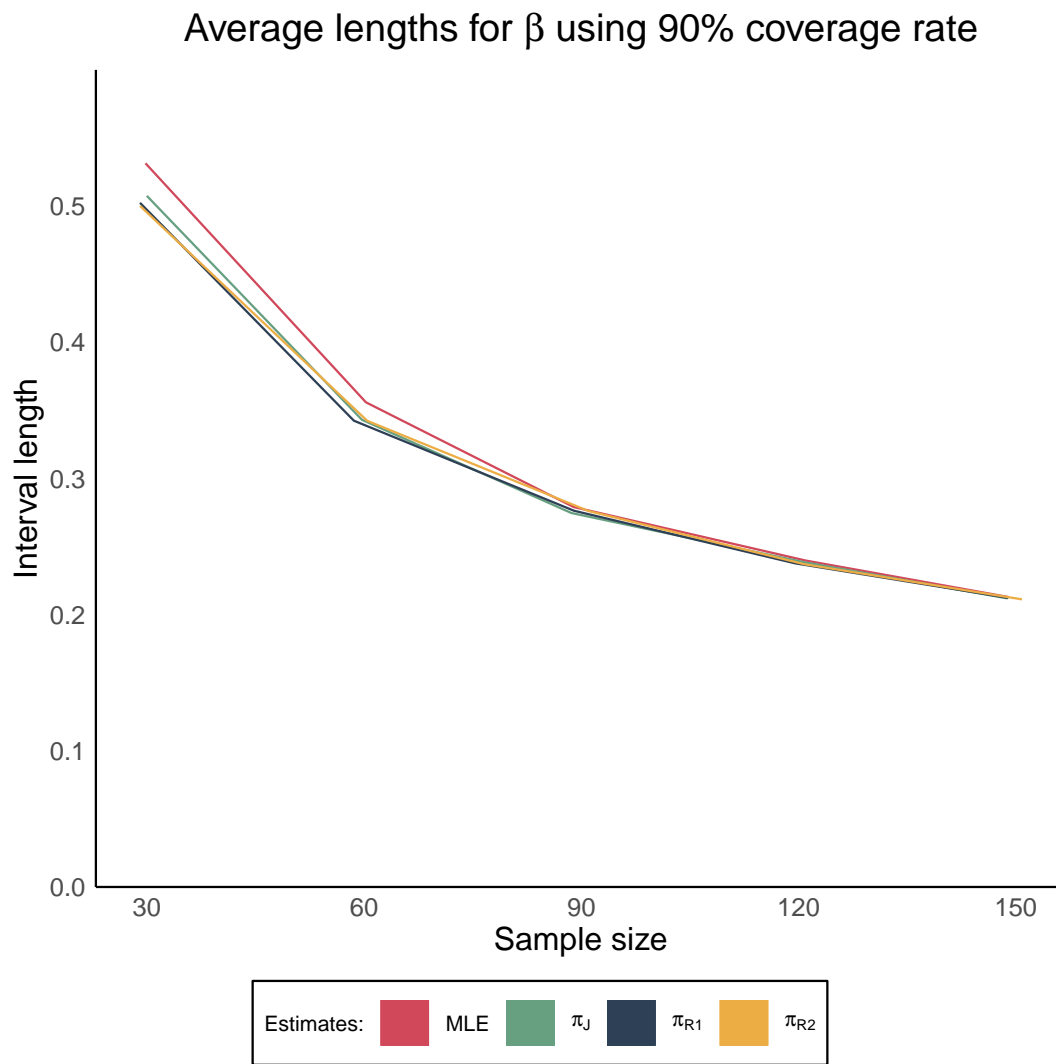
Figure 4.6: Average length for β for the 90% coverage rate.

Chapter 5

Case Study for Weibull distribution

5.1   Introduction

This section is devoted to applying the Weibull distribution to a dataset initially found in Nelson (1972). The data – as given in table 5.1 – consists of the time-to-failure of an insulating fluid subject to a constant elevated test voltage, and at each elevated level of voltage a number of times-to-failure were observed.

This experiment was run long enough such that all failures were observed, and hence this is a complete dataset. Nelson (1972) used seven elevated stress levels in his experiment, represented by the stress vector:

$$\boldsymbol{s}' = (S_1, S_2, S_3, S_4, S_5, S_{6,}, S_7)' = (26, 28, 30, 32, 34, 36, 38)', \tag{5.1}$$

where the stress applied is voltage measured in kilovolts (kV). The aim of the experiment is to estimate both the relationship between the distribution of time-to-failure and stress, and the model at the use stress level, denoted by S_0 at 20kV.

Nelson (1972) provided some assumptions of the model, namely:

1. For any constant, positive stress the life distribution is Weibull,

2. the shape parameter β of the distribution is constant at each level of stress – that is, independent of stress – and

3. the scale parameter $\nu(S_i)$ is an inverse power law function at stress level S_i, that is:

$$log(\nu(S_i)) = \theta_1 - \theta_2 \times log(S_i), \tag{5.2}$$

where: θ_1 and θ_2 are unknown parameters that need to be estimated.

Therefore the model may be written as $T_{ij} \mid \nu(S_i), \beta \sim Weibull(\nu(S_i), \beta)$ with observed values denoted by t_{ij}. Estimates for this model will be presented by the parameters $\boldsymbol{\psi}' =$

$(\lambda_0, \eta, \beta)'$, as used in chapter 3. Under the power law, δ_i is given as:

$$\delta_i = \frac{\log(20) - \log(S_i)}{\log(20) - \log(26)}, \quad i = 1, 2, ..., 7. \tag{5.3}$$

In this case study estimates from the model with be derived from two methods of parameter estimation: maximum likelihood and MCMC Bayesian methods.

Before any estimates are found, the assumptions provided above need to be tested to ensure that the results obtained are valid.

Table 5.1: Failure data of an insulating fluid under various stress levels.

	Time to failure (minutes)						
Voltage (kV)	26	28	30	32	34	36	38
	5.79	68.85	7.74	0.27	0.19	0.35	0.09
	1579.52	108.29	17.05	0.40	0.78	0.59	0.39
	2323.70	110.29	20.46	0.69	0.96	0.96	0.47
		426.07	21.02	0.79	1.31	0.99	0.73
		1067.60	22.66	2.75	2.78	1.69	0.74
			43.40	3.91	3.16	1.97	1.13
			47.30	9.88	4.15	2.07	1.40
			139.07	13.95	4.67	2.58	2.38
			144.12	15.93	4.85	2.71	
			175.88	27.80	6.50	2.90	
			194.90	53.24	7.35	3.67	
				82.85	8.01	3.99	
				89.29	8.27	5.35	
				100.58	12.06	13.77	
				215.10	31.75	25.50	
					32.52		
					33.91		
					36.71		
					72.89		
n_i	3	5	11	15	19	15	8

5.1.1 Comparing different distributions

Firstly, the assumption that the dataset comes from a Weibull distribution needs to be considered. Hence, the dataset will need to be compared to other popular life distributions, namely: the exponential, gamma and log-normal distribution, to ensure that the fit is the best comparatively. Here two methods of comparing distributions will be considered: quantile-quantile (QQ) plots and AIC values.

Figure 5.1 considers QQ plots for the dataset under each stress level. A QQ-plot is a graphical tool which helps one assess if the data comes from a given distribution (Rinne, 2008). A QQ-plot plots theoretical quantiles on the abscissa and sample quantiles on the ordinate, and

the sample is deemed to follow the given distribution if the points roughly follow a straight line. Theoretical quantiles for the dataset under each stress level were computed using the R package $fitdistrplus$ (Delignette-Muller et al., 2015).

The QQ-plots shows that the fit is good for the four life distributions, especially under the stress levels with larger n_i (here: S_4, S_5 and S_6). The fit under stress levels with smaller n_i (S_1, S_2 and S_3) appear slightly suspect, however the erratic behavior of these curves is most likely due to the small sample size. The fit under stress level S_7 appears reasonable despite having a comparatively small sample size ($n_7 = 8$). The fit between the Weibull, gamma and exponential distribution appear to be very similar under the stress levels, whereas the fit of the data with the log- normal model appears slightly different – with the exception of the fit under the stress level S_7.

QQ–plots
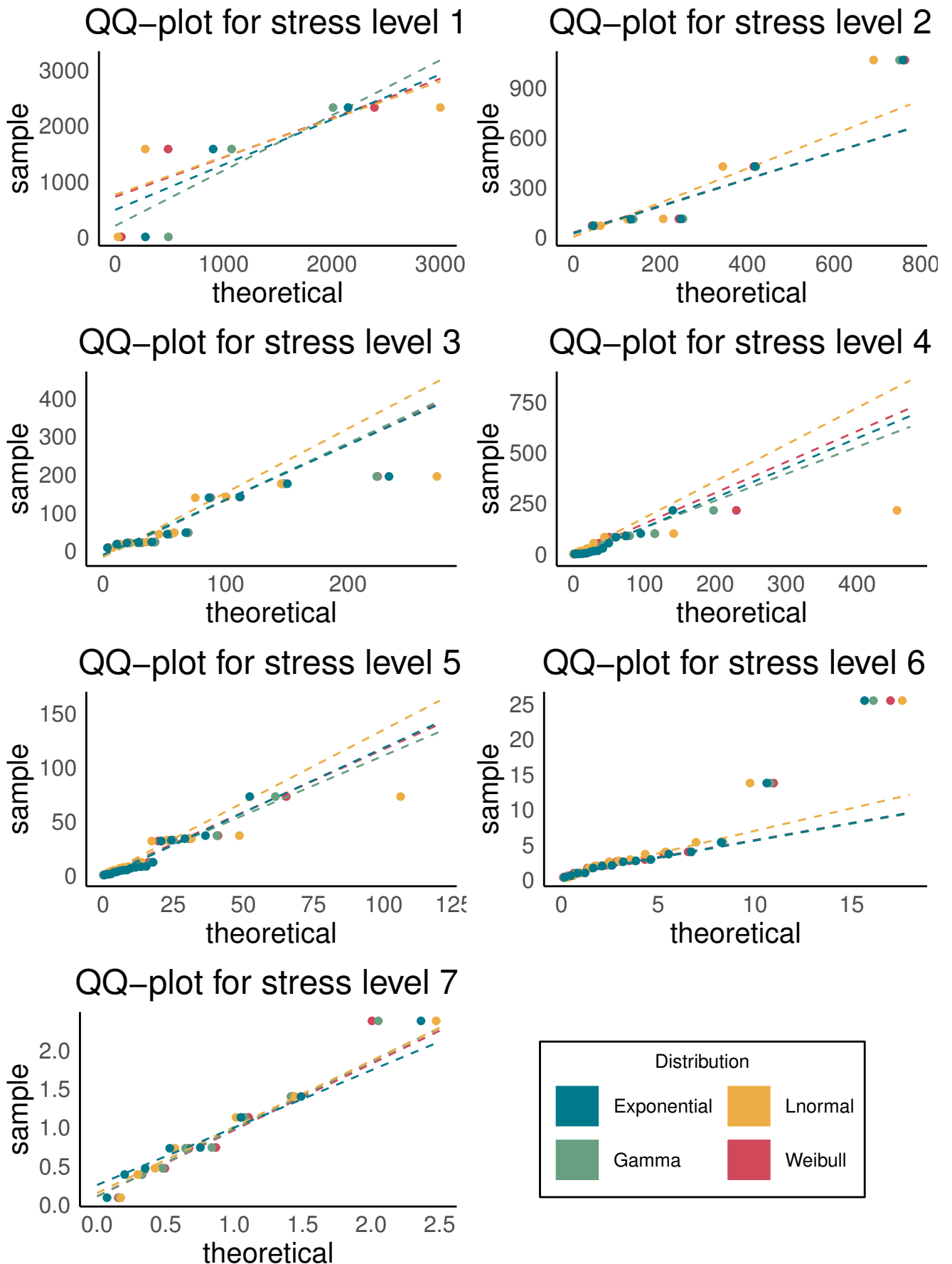
Figure 5.1: QQ-plots for the dataset under different distributions.

Table 5.2 shows the AIC values for the dataset under different distributions. The lowest AIC value is for the log-normal distribution, implying that it provides the best fit of the data. However Shaked and Singpurwalla (1982) (who also assessed the fit for this dataset) claim that from a practical standpoint – despite the log-normal distribution fitting better for this dataset – the Weibull distribution is preferred due to its monotonic-failure rate. This is in comparison the the log-normal's failure rate which first increases and then decreases, which is an unrealistic description of the time-to-failure. Hence the Weibull distribution will be considered from here on.

Table 5.2: AIC values of the dataset under different distributions.

Distribution	AIC Value	$\triangle_i$ AIC
Exponential	620.1166	4.3326
Gamma	626.5542	10.7702
Log-normal	615.7840	0
Weibull	618.4192	2.6352

5.1.2   Testing Weibull assumption

To test for the Weibull distribution, consider figure 5.2 which shows a WPP of the data. If the data points are close to the trend line then we claim that the dataset comes from a Weibull distribution. From the plot the data appears to come from a Weibull distribution, with the exception of the data from stress level S_1. Again the issue may be due to the small sample size.
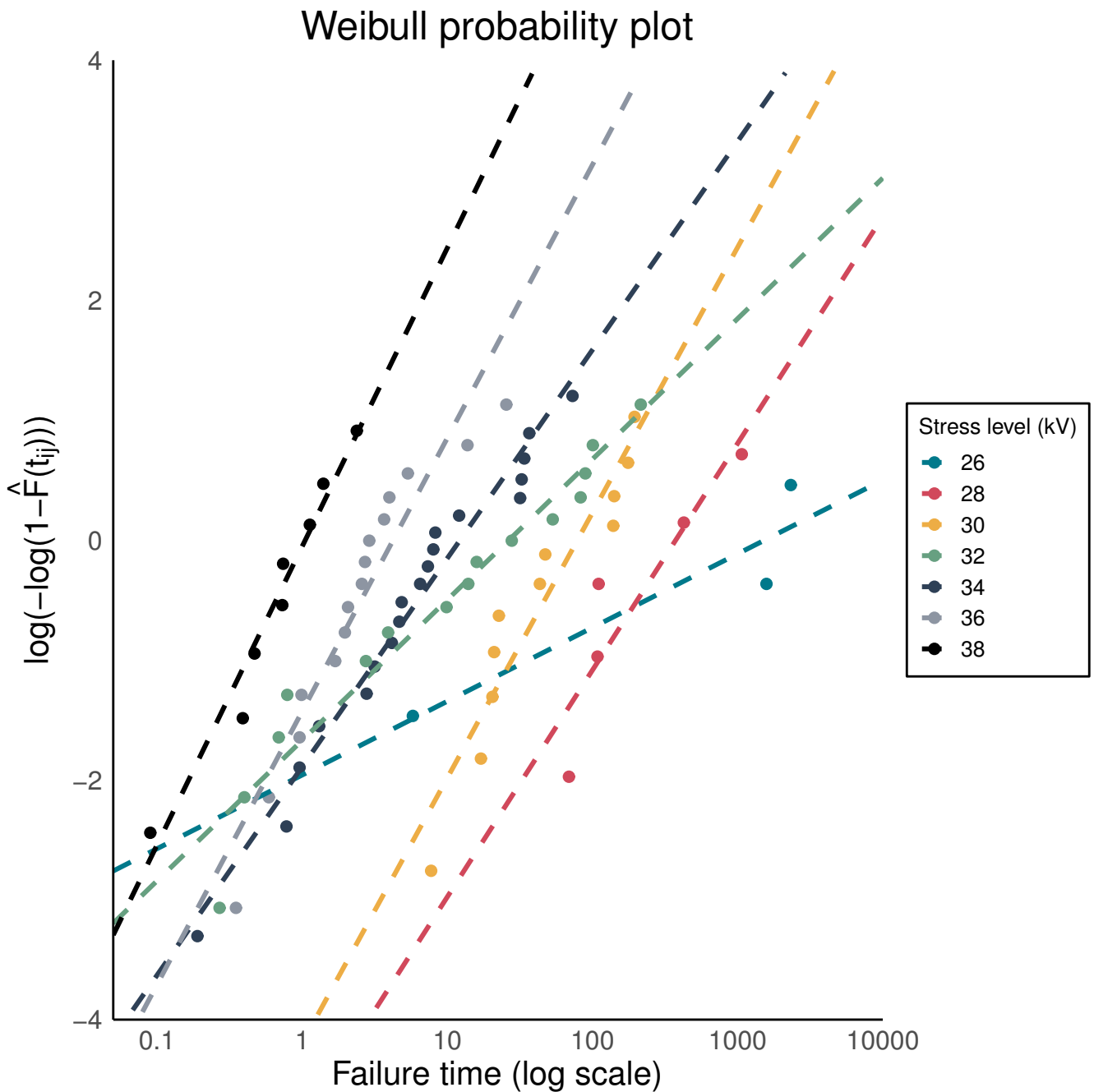
Figure 5.2: Weibull plot for the dataset under each stress level.

Nelson (1990) claims that the slope of the WPP is identical to the Weibull's shape parameter β. Since the slopes for each stress level are not the same, it implies that the shape parameter is not constant. However, to meet the naive assumption that the shape parameter is constant, a second Weibull-plot is given in figure 5.3 such that the slopes of all the curves are constant. Graphically, the assumption of Weibull distribution appears to be valid, although less so in comparison to figure 5.2, suggesting that the shape parameter may not actually be constant as was initially assumed.
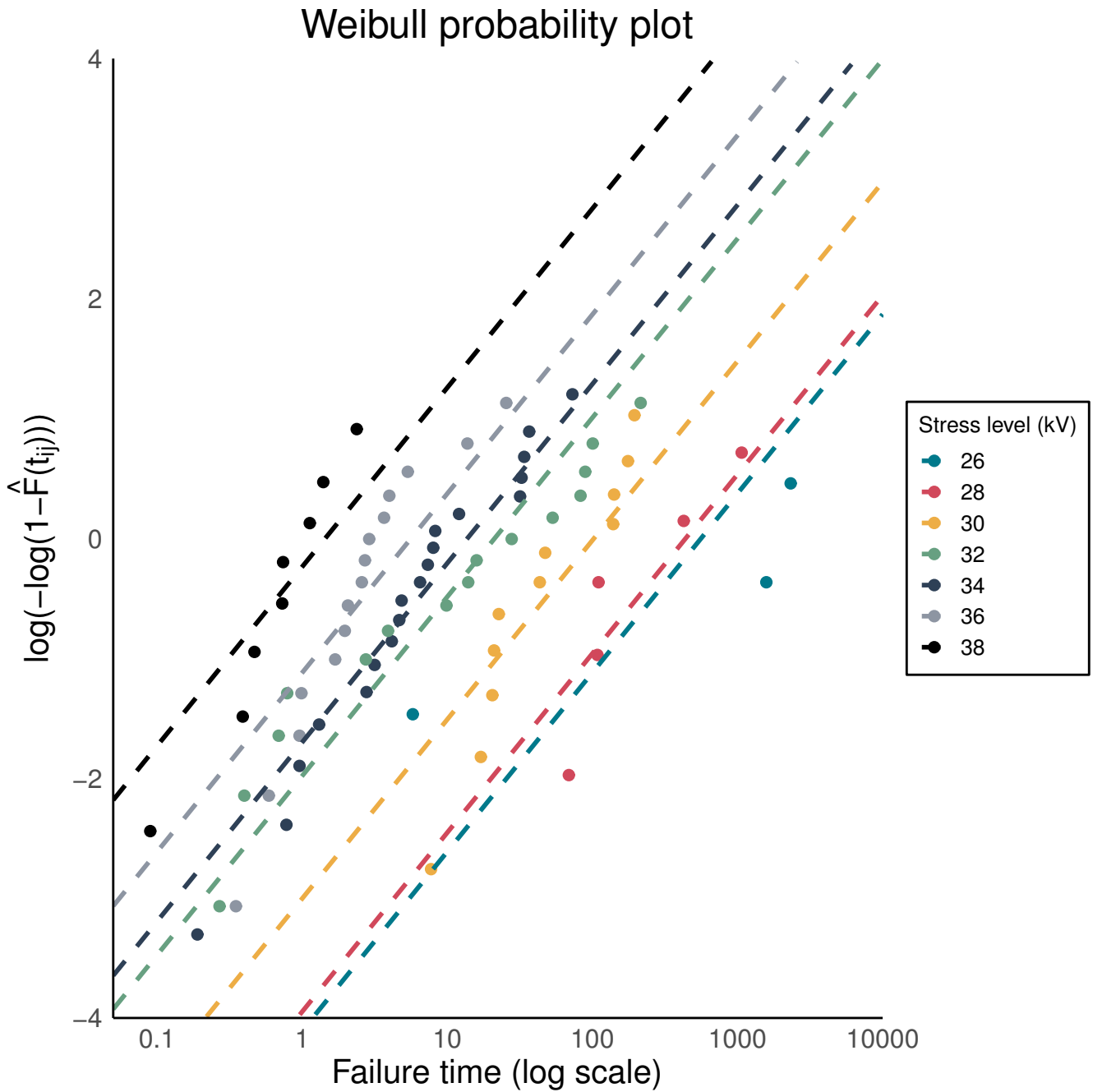
Figure 5.3: Weibull plot for the dataset under each stress level (assuming shape parameter constant).

5.1.3   Testing linearity assumption

This model assumes that the transformed life-stress relationship is linear. A graphical assessment between characteristic life and stress is provided in figure 5.4. It appears that the life-stress relationship is reasonably linear, with the exception of a kink at stress level S_4. This appears to be caused by the faster than expected failure times at this stress level, suggesting that some of the failure times at this stress level may be outliers.
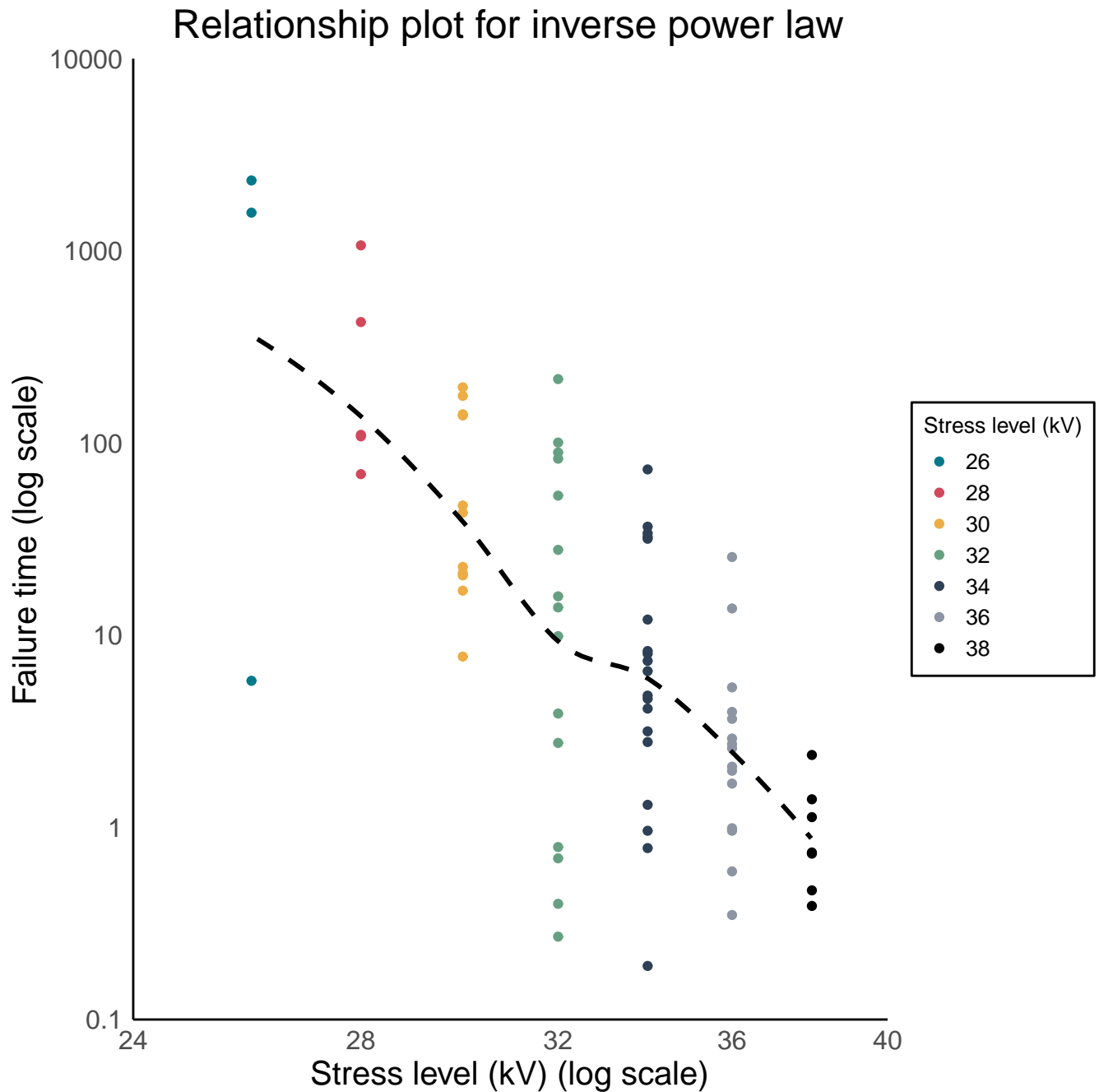
Figure 5.4: Plot illustrating linearity of the life-stress relationship.

5.2   Maximum likelihood estimates

Maximum likelihood estimates for this dataset were found, and their summaries are described in table 5.3. The value of the shape parameter $\beta < 1$ suggesting that the failure rate is decreasing.

The asymptotic properties as described in section 2.4.1.1 were used in construction the 95% confidence intervals, which are given by:

- $\hat{\lambda}_0 \pm z_{\frac{\alpha}{2}} \times \sqrt{var\left(\hat{\lambda}_0\right)},$

- $\hat{\eta} \pm z_{\frac{\alpha}{2}} \times \sqrt{var\left(\hat{\eta}\right)}$ and

- $\hat{\beta} \pm z_{\frac{\alpha}{2}} \times \sqrt{var\left(\hat{\beta}\right)}$,

where: $z_{\frac{\alpha}{2}}$ is a standard normal variate.

Table 5.3: MLE values for the parameters and their respective standard errors.

Parameter	Estimate	Se	95% confidence interval
$\lambda_0(\times 10^{-6})$	8.0357	1.1870	$(5.7093, 10.3621)$
η	104.6360	7.9737	$(88.3716, 119.6284)$
β	0.7766	0.06474	$(0.6507, 0.9045)$

5.3   Bayesian analysis

This section is devoted to finding Bayesian estimates for the dataset in table 5.1 using the non-informative priors derived in chapter 3. Before estimates are found a simulation study needs to be conducted to ensure convergence of the MCMC algorithm.

Three separate chains, denoted here by $\boldsymbol{\varsigma}' = (\varsigma_1, \varsigma_2, \varsigma_3)'$, with well-dispersed starting points given by:

- $\varsigma_1^{(0)} = \left(\lambda_0^{(0)} = 1 \times 10^{-6}, \eta^{(0)} = 100, \beta^{(0)} = 0.1\right)$,

- $\varsigma_2^{(0)} = \left(\lambda_0^{(0)} = 3 \times 10^{-6}, \eta^{(0)} = 200, \beta^{(0)} = 0.3\right)$ and

- $\varsigma_3^{(0)} = \left(\lambda_0^{(0)} = 5 \times 10^{-6}, \eta^{(0)} = 500, \beta^{(0)} = 0.5\right)$

were run for 250000 iterations. A burn-in period of 150000 iterations was considered before estimates were found.

In this study, five graphical methods are used to show convergence, namely:

- trace plots,

- running mean plots,

- autocorrelation plots,

- BGR plots and

- Geweke plots.

The plots for the posterior under the prior π_J are shown in this section, and the plots for the posteriors under the priors π_{R1} and π_{R2} are provided in the appendix.

5.3.1   Trace plots

Consider the trace plots presented in figures: 5.5, C.1 and C.2. It can be seen that the three chains for β have mixed well as the plots look like white noise and seem to be centered around a stationary mean. For the other two parameters, (λ_0, η), the three chains do not appear as stable as the chains for β. However the values appear to fluctuate heavily around their mean values.
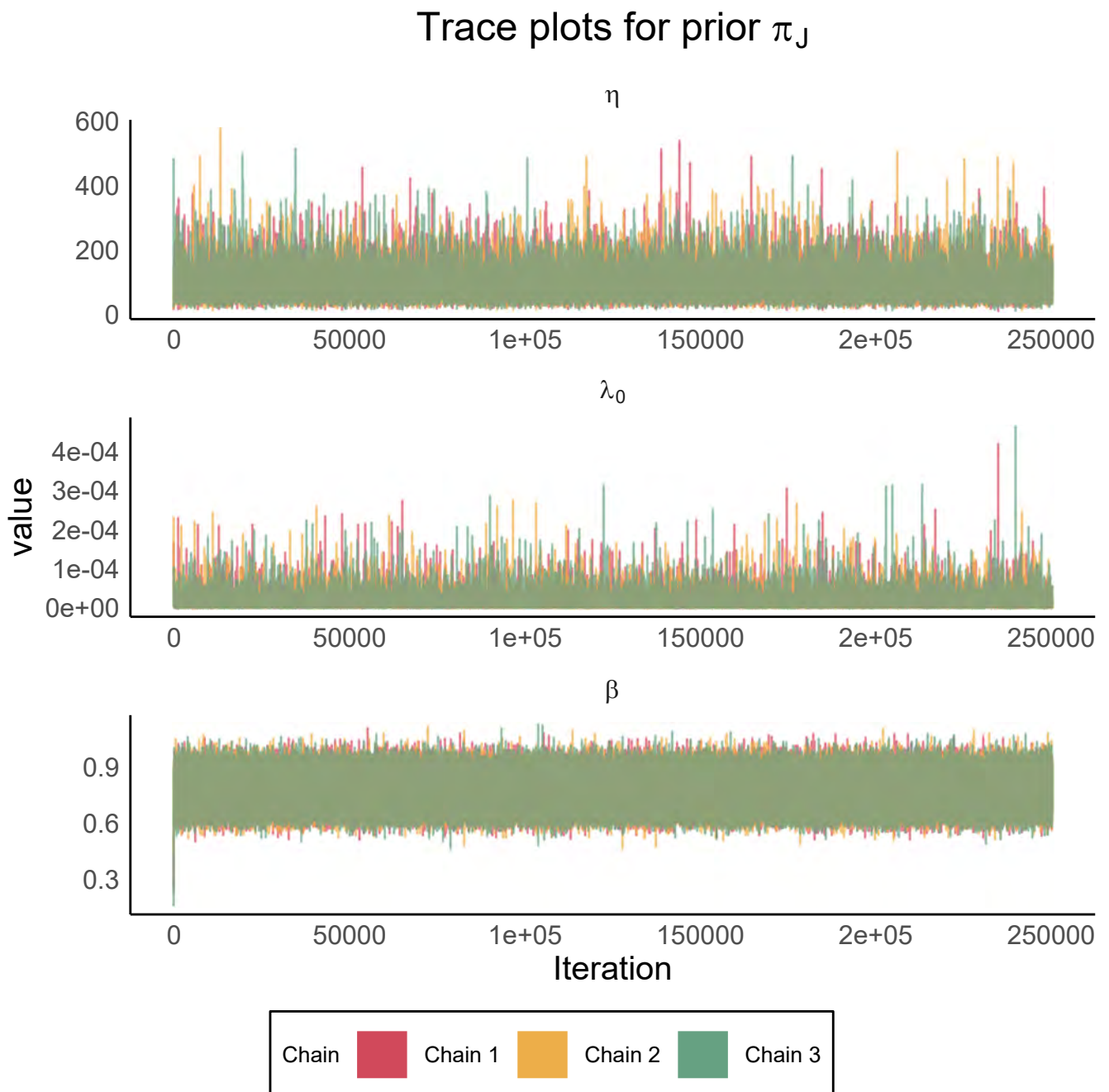
Figure 5.5: Trace plots for the chains under prior π_J.

5.3.2   Running mean plots

Running mean plots are given in figures: 5.6, C.3 and C.4. The horizontal lines on the plot represent the mean values for the parameters.

It appears that β approaches a stationary mean almost immediately, whereas λ_0 and η approach a stationary mean far slower. Although after around 150000 iterations, the lines for all parameter stabilizes suggesting that the parameters have reached their true mean value.
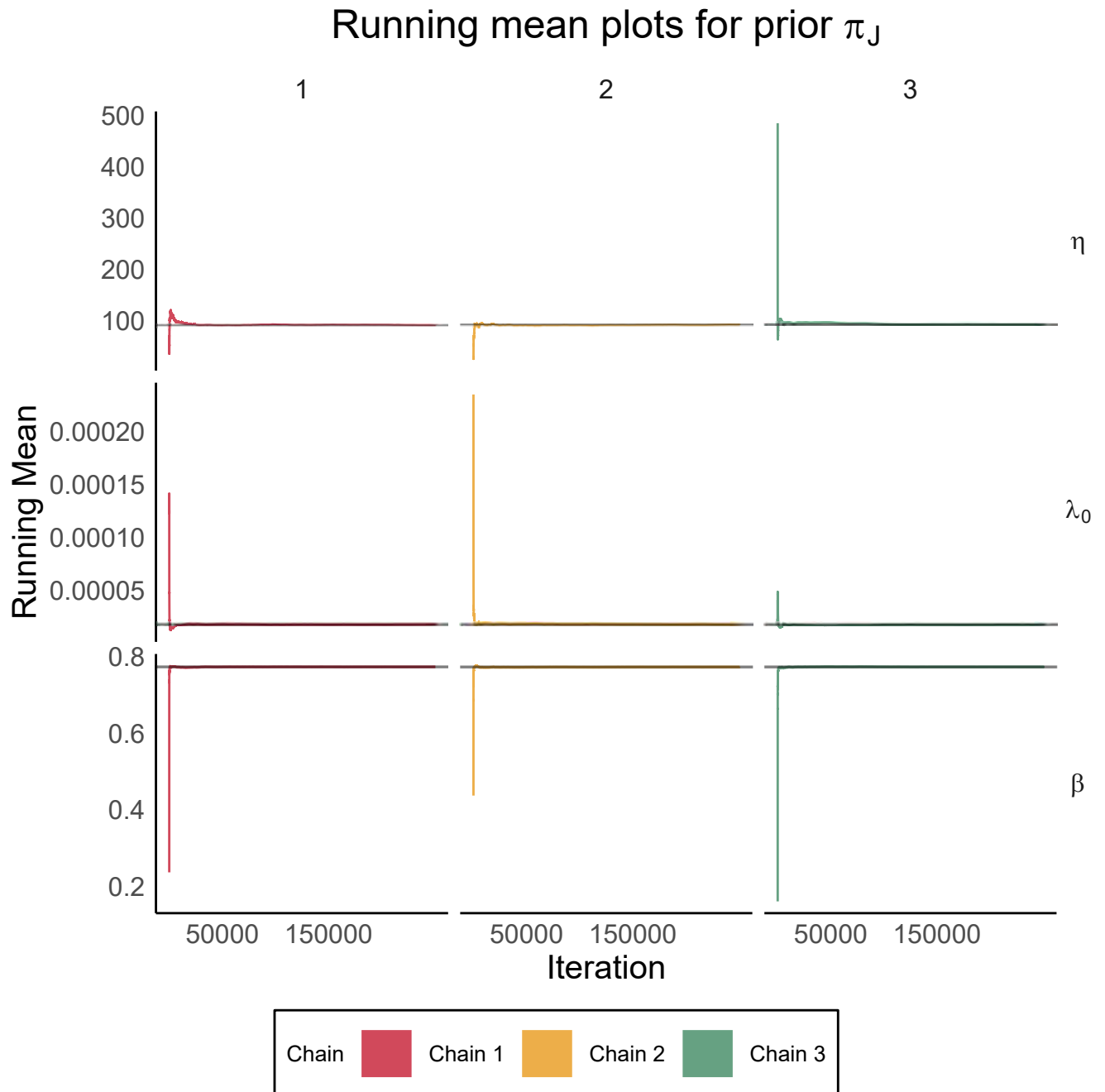
Figure 5.6: Running mean plots for the parameters under prior π_J.

5.3.3   Autocorrelation plots

Autocorrelation plots can be found in figures: 5.7, C.5 and C.6. For all plots the mixing rate of β is incredibly quick, suggesting that it finds its stationary distribution almost immediately. The mixing rate for λ_0 and η is far slower, although they eventually approach lag 0.
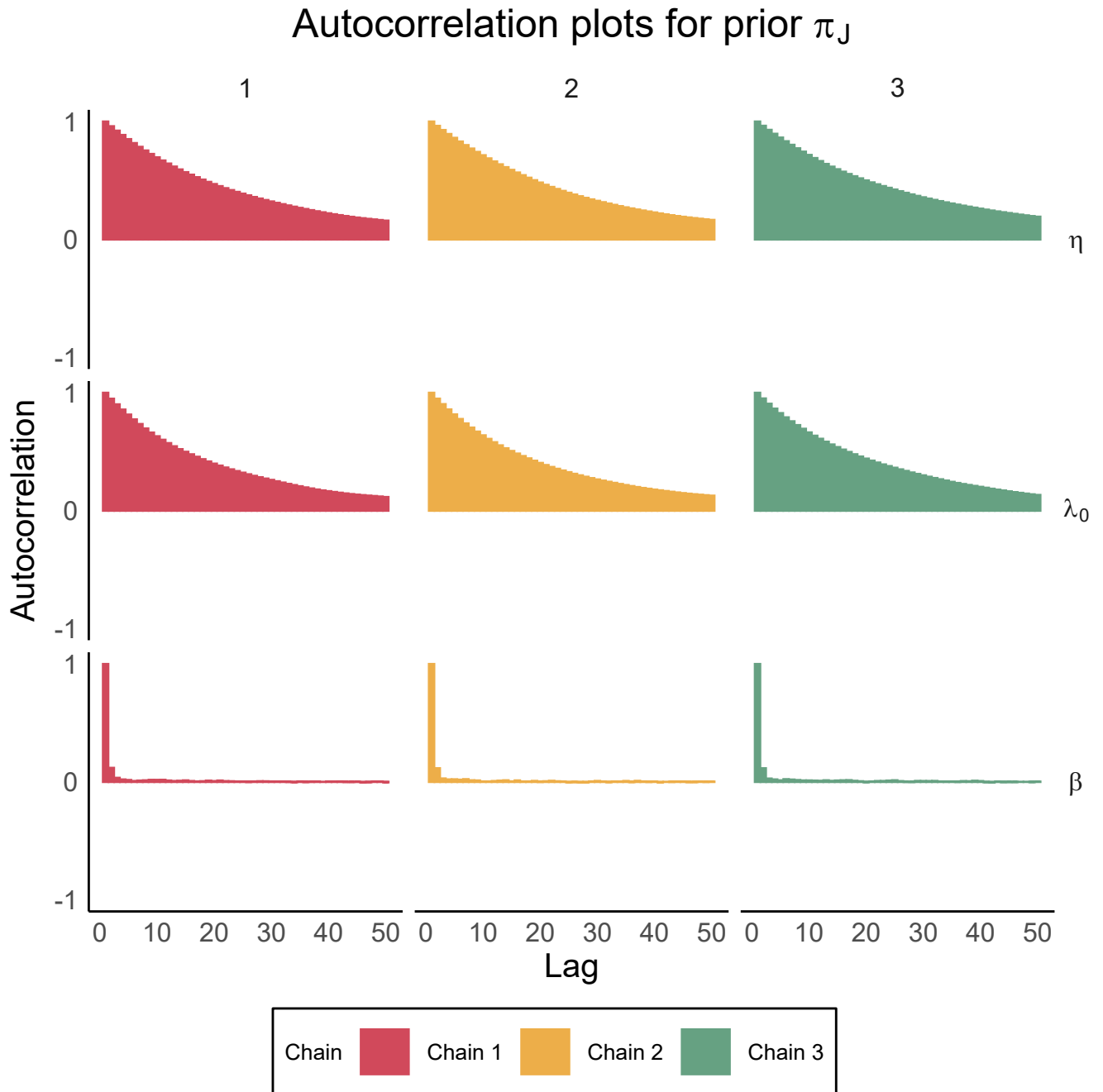
Figure 5.7: Autocorrelation plots for the parameters under prior π_J.

5.3.4   BGR plots

The BGR plots in figures: 5.8, C.7 and C.8 plot the $\hat{R}$ line as described in section 2.6.2.1. If $\hat{R} < 1.2$ we claim that the chains have converged to a stationary mean.

For all plots at around the 50000th iteration $\hat{R} < 1.2$, suggesting convergence.
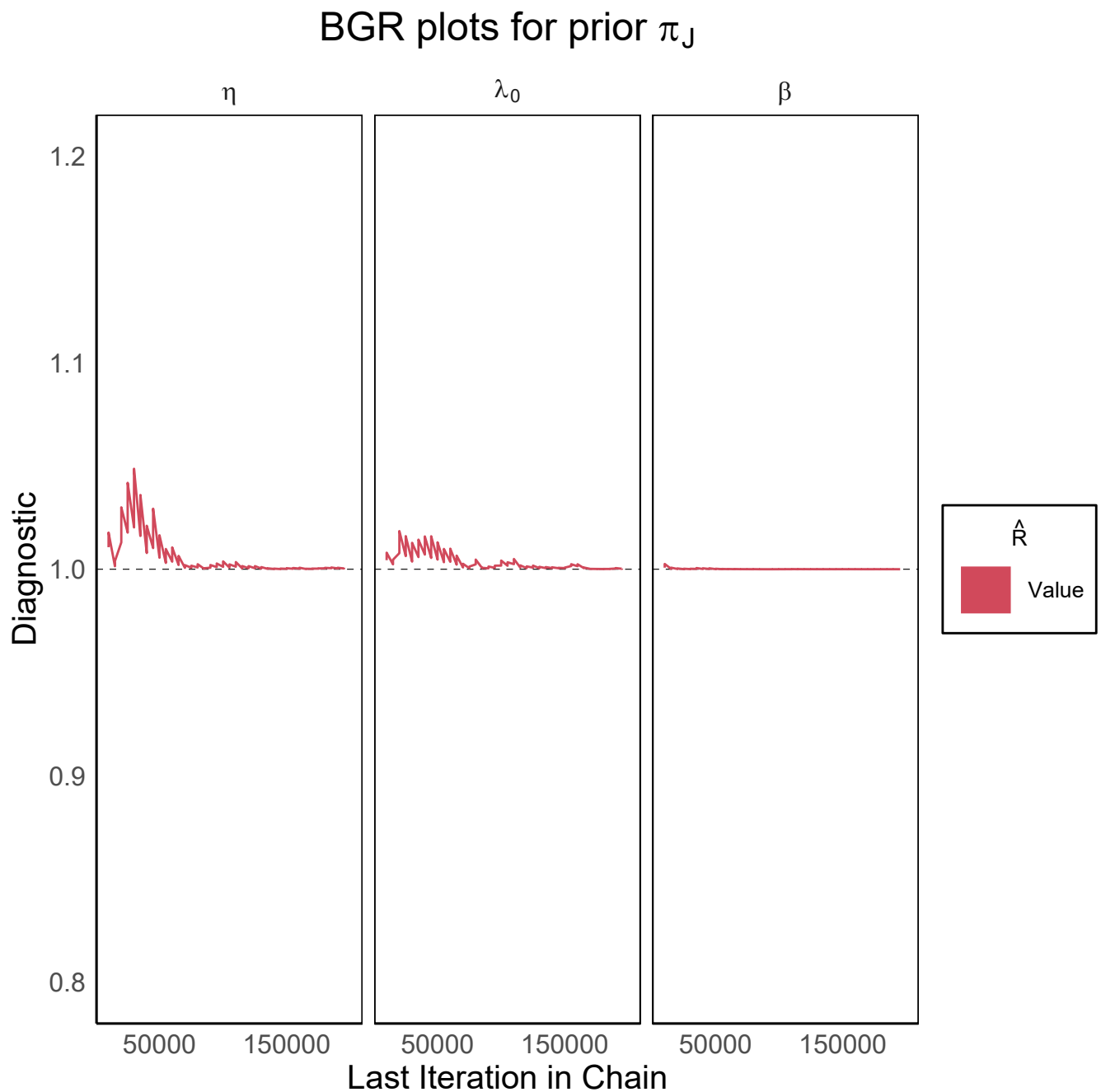
Figure 5.8: BGR plots for the parameters under prior π_J.

5.3.5   Geweke plots

Geweke plots presented in figures: 5.9, C.9 and C.10 illustrate the hypothesis: H_0 : the mean values of the first and last part of the chain are the same against H_a : the mean values of the first and last part of the chain are different, for successively smaller segments of the last part of the chain. We reject the null hypothesis at a 95% level of significance if the dots on the plot fall out of the ± 1.96 area.

For most segments of the last part of the chain, we fail to reject null hypothesis and claim

that the first part of the chain and the last part of the chain are significantly similar – which suggests that the chains have converged.
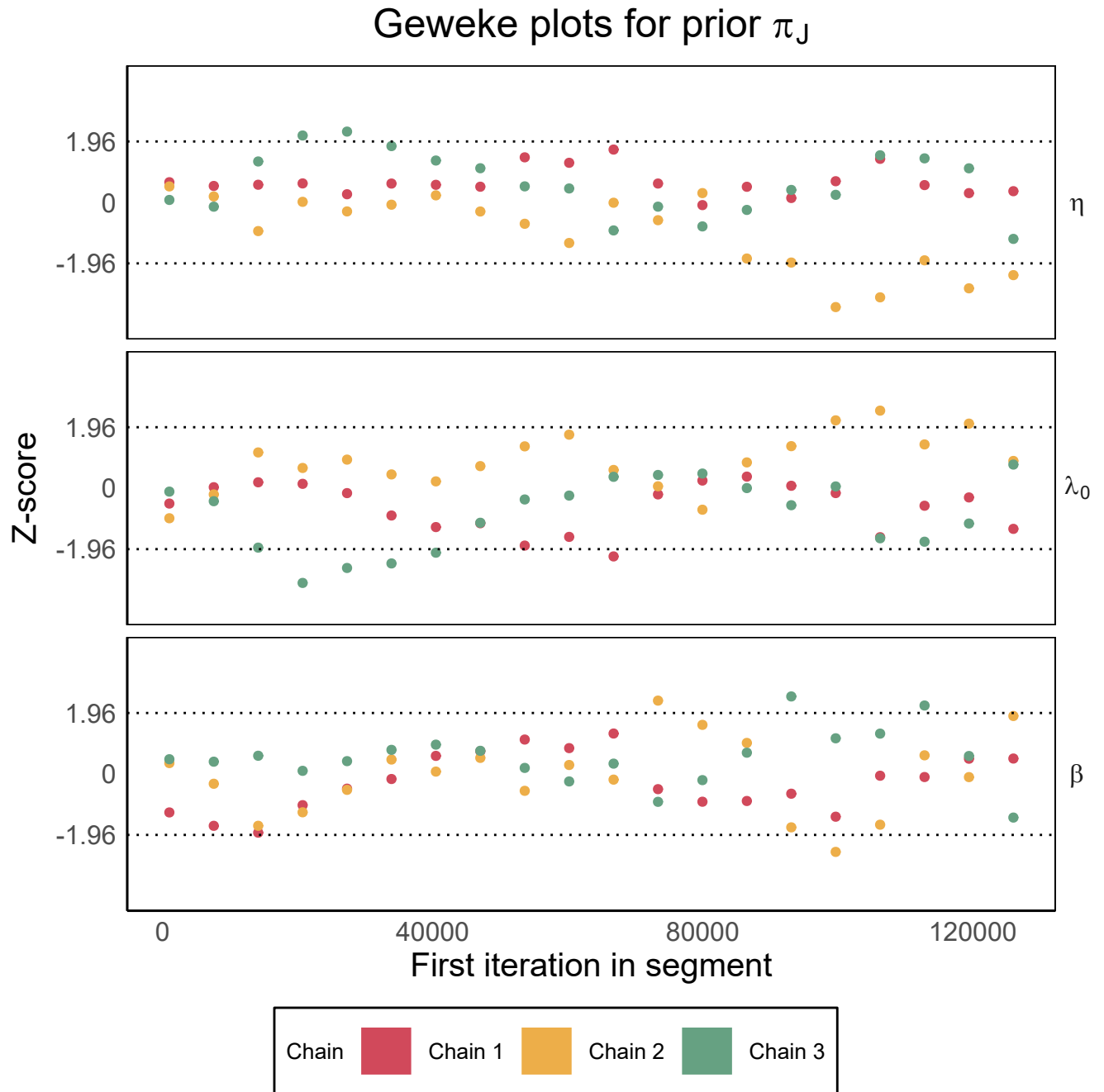
Figure 5.9: Geweke plots for the parameters under prior π_J.

5.3.6 Density plots

Figures: 5.10, C.11 and C.12 give density plots for the three parameters in $\boldsymbol{\psi}$ for each chain. For all parameters, the density plot for each chain appears to be similar suggesting that the estimates under each chain are similar.

The density plots for β tend to be symmetric whereas the density plots for λ_0 and η both tend to be skewed to the right. Therefore the estimates for these two parameters are dif-

ferent to those obtained via the method of maximum likelihood which assumes a symmetric distribution under the normal asymptotic assumptions. The skewness is most likely due to the small sample size under certain stress levels, for example, the sample size when stress level S_1 is used is only three. Therefore the estimates obtained via the method of maximum likelihood may be misleading in comparison to the Bayesian estimates.
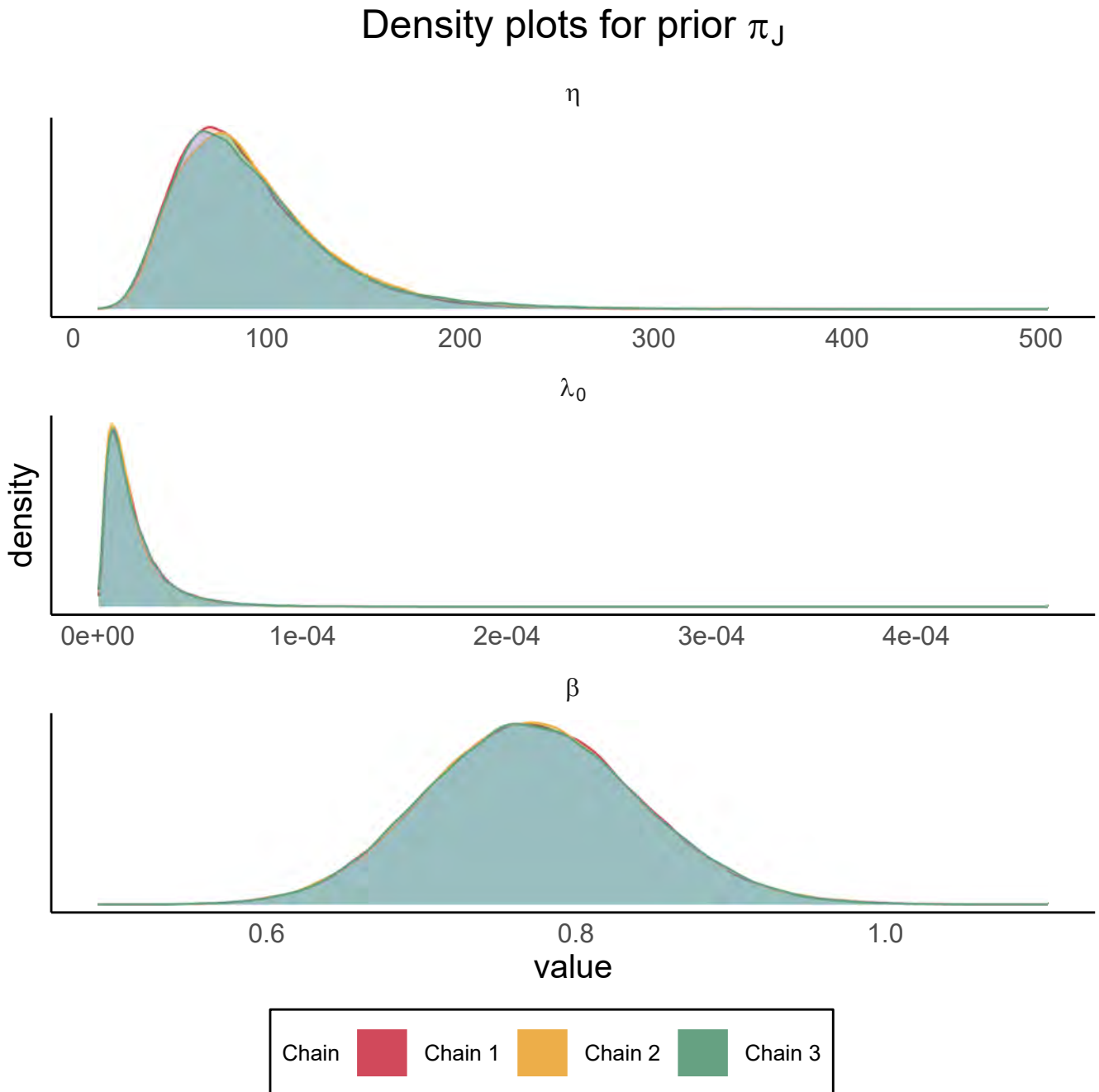
Figure 5.10: Density plots for the parameters under prior π_J.

5.3.7 Bayesian estimates

Bayesian estimates are given in tables: 5.4, C.19 and C.20. The estimates under the three priors are all very similar. As expected, the estimates for λ_0 and η are slightly different to

the MLE values as a result of the MLE values assuming normal asymptotic assumptions.

Table 5.4: Estimates under prior π_J.

Parameter	Mean	Sd	MCMC error	95% credible interval	Median
$\lambda_0(\times 10^{-6})$	11.0300	10.1500	0.1250	$(1.5910, 38.0600)$	8.0700
η	113.6000	50.1800	0.6628	$(46.4400, 237.4000)$	103.7000
β	0.7756	0.0684	0.0002	$(0.6455, 0.9337)$	0.7742

DIC values are given in table 5.5. Again the values for the three priors are similar with π_{R1} having the lowest value – that is, implying the best fit. However after employing the $\triangle_i$ rule as mentioned in section 2.7.1.1, the difference between fit in the three models is negligible.

Table 5.5: DIC values for the Bayesian priors.

Prior	$\overline{D(\psi)}$	$D(\bar{\psi})$	p_d	DIC	$\triangle_i$ DIC
π_J	605.0650	614.027	-8.9620	596.1020	1.2780
π_{R1}	604.5910	614.3580	-9.7670	594.8240	0
π_{R2}	604.5950	614.0850	-9.4900	595.1050	0.2810

One usually expects a positive p_d value, however Lesaffre and Lawson (2012) claim that p_d may be negative when the likelihood function is non-log-concave or when the posterior mean is not a good summary measure. The latter may occur when the posterior distribution for a parameter is either extremely asymmetric, or symmetric but bimodal.

Lindley's approximations of the estimates are found in table 5.6. The estimates were found using the derivations from section 4.4. The values obtained are more closely related to the MLE values in comparison to their Bayesian MCMC counterparts.

Table 5.6: Lindley's approximation of the parameters.

Parameter	π_J	π_{R1}	π_{R2}
$\lambda_0(\times 10^{-6})$	8.0345	8.0351	8.0358
η	104.6316	104.6398	104.6481
β	0.7766	0.7767	0.7768

5.4 Estimates under use-stress level

This section is devoted to obtaining reliability plots and relevant percentiles for the data under the use stress level $(20kV)$. The reliability function for the data at any stress level i is

given as:

$$R\left(t \mid \boldsymbol{\psi}\right) = \begin{cases} exp\left\{-\left(\lambda_0 \eta^{\delta_i} t\right)^\beta\right\} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{5.4}$$

therefore reliability under the MLE values can be obtained by merely replacing the parameters in 5.4 with the estimated values obtained in table 5.3.

Reliability for the Bayesian estimates can be found by the equation:

$$R\left(\boldsymbol{\psi} \mid t\right) = \int\limits_0^\infty \int\limits_1^\infty \int\limits_0^\infty R\left(t \mid \boldsymbol{\psi}\right) \pi_R\left(\boldsymbol{\psi} \mid t\right) d\lambda_0 d\eta d\beta, \tag{5.5}$$

where:

- $R\left(t \mid \boldsymbol{\psi}\right)$ is given in equation 5.4 and

- $\pi_R\left(\boldsymbol{\psi} \mid t\right)$ is the posterior distribution under the general reference prior as given in equation 3.36.

Solving equation 5.5 analytically is difficult, however according to Soyer et al. (2014) reliability estimates for the Bayesian estimates can be solved via ergodic averages with the equation:

$$R\left(\boldsymbol{\psi} \mid t\right) \approx \frac{1}{n} \sum_{i=1}^n R\left(t \mid \lambda_0^{(i)}, \eta^{(i)}, \beta^{(i)}\right), \tag{5.6}$$

where:

- n is the number of iterations of the Markov chain, post burn-in and

- $\lambda_0^{(i)}, \eta^{(i)}$ and $\beta^{(i)}$ are the Bayesian estimates of $\lambda_0 \, \eta$ and β at iteration i.

Figure 5.11 illustrates the reliability curves of the data under the estimates found above. The four reliability curves all appear to be very similar. The MLEs initially provide estimates with the longest reliability, however as time becomes larger the MLEs become smaller in comparison to the Bayesian estimates. Among the Bayesian estimates, the reliability curve under the priors π_{R1} and π_{R2} are practically identical (the estimates under π_{R1} are marginally larger) and appear to give the largest time estimates in comparison to the reliability curve under the prior π_J.
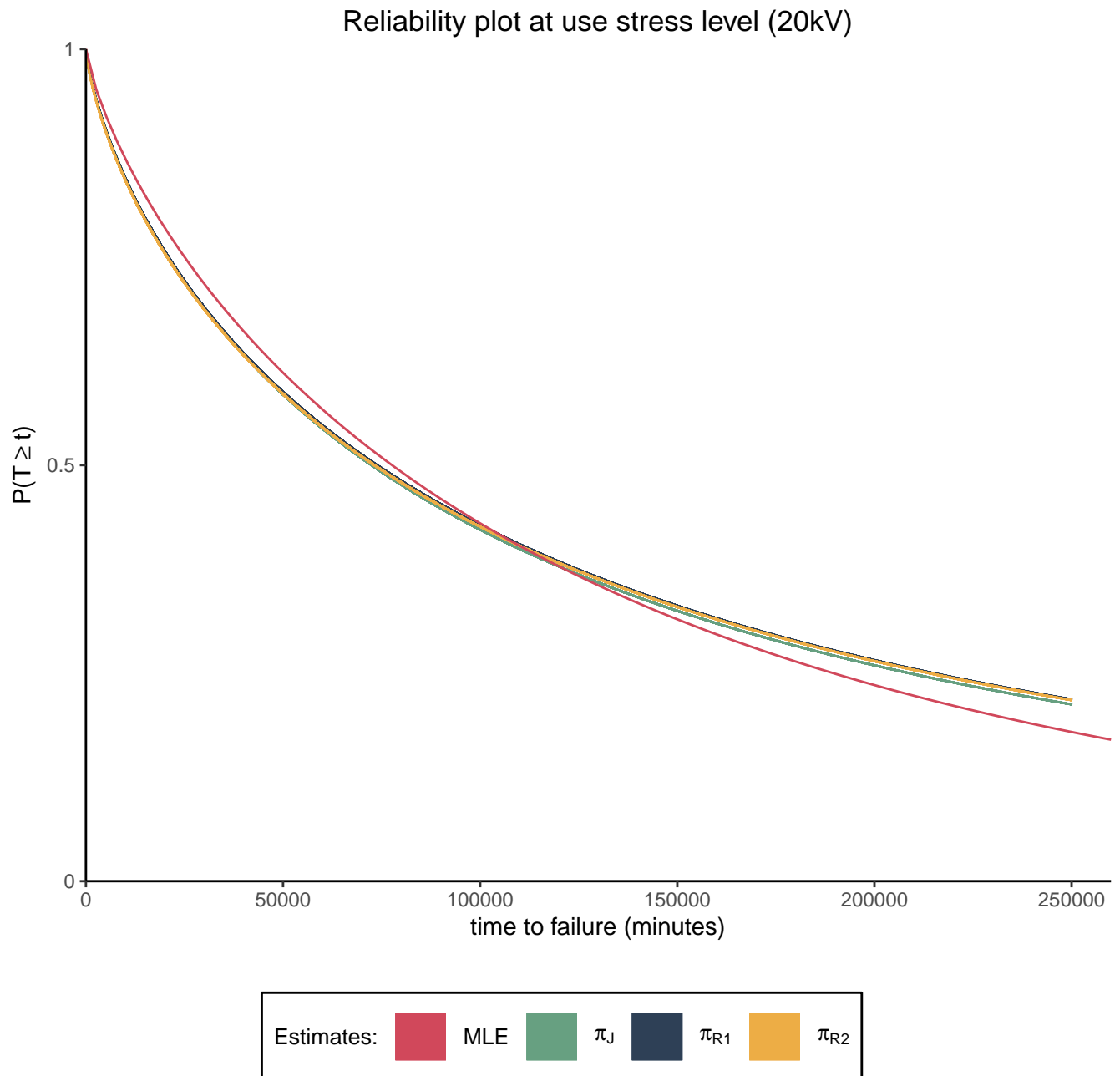
Figure 5.11: Reliability plot for the estimates at use stress level (20kV).

Furthermore table 5.7 provides quantiles for the reliability estimates under each estimate. These percentiles again confirm that as time becomes large the Bayesian estimates provide the largest reliability at use stress levels in comparison to the MLEs. The nth percentile can be interpreted as: "the time it took for $n\%$ of the insulating fluids to break down." The 63.2^{th} percentile of the Weibull distribution is significant as it represents the characteristic life of the distribution, and is identical to the scale parameter of the model, denoted by: $\nu\left(S_i\right)$ (Nelson, 1990).

Table 5.7: Table of percentiles for the reliability curves under the different estimates (minutes).

	MLE	π_J	π_{R1}	π_{R2}
10th	6907.2130	5381.6710	5419.4700	5388.9040
63.2^{th}	124359.2000	124308.7000	126690.7000	126460.4000
90th	364884.1000	492174.9000	508531.2000	499657.0000

Chapter 6

Conclusion

6.1 Concluding remarks

6.1.1 Conclusions for Chapter 3

The premier focus of this thesis was deriving non-informative Bayesian estimates for the Weibull distribution subject to constant-stress ALT, and assuming that a complete dataset is used in analysis. The Weibull distribution was assumed to have a constant shape parameter, regardless of the stress level, and a scale parameter that was a log-linear function of stress. The likelihood of the described distribution was constructed, and since finding estimates from it would be difficult, a transformation proposed by Xu et al. (2015) was considered.

The five non-informative Bayesian priors derived for this thesis were: Jeffreys' prior, reference priors, the MDI prior, the uniform prior and PMPs. These priors are considered when little information prior on an experiment prior to it being run is available.

The properness of the posterior distributions under these non-informative priors were also considered. The posterior under Jeffreys' prior, and the two reference priors were found to be proper distributions, and hence were used for analysis in both a simulation – and case study. The posterior distributions under the MDI prior and uniform prior were found to be improper distributions, and hence were not considered further in this thesis.

To contrast the non-informative Bayesian estimates, maximum likelihood estimates were also derived. The MLEs required solving difficult, non-linear equations and hence an iterative procedure was required to approximate them. First, these equations were simplified by forming the MLE for λ_0 as a function of the MLEs of η and β, then re-forming the log-likelihood with λ_0 as a function of η and β, and finally using the NR algorithm to approximate the estimates of η and β from this new log-likelihood.

6.1.2 Conclusions for Chapter 4

Chapter 4 considered a simulation study using the Weibull distribution formed in chapter 3 of this thesis. It contained simulating data from hypothetical, complete datasets under three elevated stress levels associated with temperature. Therefore the Arrhenius model was used.

The main focus was finding parameter estimates and root mean squared error values for the parameters subject to different estimation techniques. The estimation techniques considered were maximum likelihood estimation, and two types of Bayesian estimates: those found via an MCMC simulation, and those found via the technique suggested by Lindley (1980). The RMSE values were found under three loss functions: the symmetric squared error loss function and the asymmetric LINEX loss function and the GELF.

In terms of parameter estimates the Lindley estimates provided results closest to the true parameter values, followed by the MLEs and then by the MCMC estimates.

The RMSE values for λ_0 were smallest for the MCMC Bayesian estimates when the GELF was used and the loss parameter k was positive. Thus suggesting that the general entropy loss function overestimated the parameter values. The lowest RMSE values tended to be provided by the posterior distribution under prior π_{R2}.

For η the lowest RMSE values were provided by the MCMC Bayesian estimates under the LINEX loss function with the loss parameter a being positive. The lowest RMSE values tended to be provided by the posterior distribution under the prior π_J.

For β the lowest RMSE values were provided by the MCMC Bayesian estimates under the GELF with loss parameter k being positive. The posterior distribution under the prior π_{R2} provided the estimates with the lowest RMSE values.

The second part of the simulation study dealt with finding the 95% and 90% coverage rates for the maximum likelihood and MCMC Bayesian estimates. Both types of estimates did well since their coverage rates were approximately their nominal values, although generally the MCMC Bayesian estimates performed better than the MLEs since their average interval lengths were generally shorter.

6.1.3 Conclusions for Chapter 5

In chapter 5 a case study for the Weibull distribution was considered by using a dataset from Nelson (1990), containing the time-to-failure of an insulating fluid subject to various higher levels of stress. The stress for these data was related to voltage and hence the inverse power law was used. A full test to show that the assumptions of the model were met and that the data does indeed follow a Weibull distribution was considered. Moreover, a comparison of the fit of the data between other common life distributions was also completed.

A full convergence test was considered for the Bayesian MCMC estimates to ensure that the models reached convergence, and that their results were admissible. The fit of the data between the non-informative Bayesian estimates were compared with DIC values. The fit was

best for the posterior distribution under the reference prior π_{R1}, although the difference between DIC values for the three posterior distributions considered was negligible. Finally, estimates under use-stress level were considered using maximum likelihood and MCMC Bayesian estimates; and using the resultant estimates a reliability plot was formed.

6.2   Future research

In this paper two life-stress relationships were considered: the Arrhenius relationship used in the simulation study in chapter 4 and the inverse power law relationship used in chapter 5. There are however other life-stress relationships that are not often discussed in literature, for example the Eyring relationship, which could be considered in future research.

Figure 5.3 found in chapter 5 of this thesis suggested that the shape parameter of the Weibull distribution from the Nelson (1972) dataset was not constant at each elevated stress level. Hence future research could consider a Weibull distribution with both scale and shape parameter dependent on stress.

The sole distribution used in this paper was the Weibull distribution – perhaps the most commonly used life distribution in literature. Other life distributions could be considered to compare against the Weibull distribution – for example, the fit of the data used in chapter 5 was found to be best under the log-normal distribution, and hence should be considered further. It would also be interesting to derive the estimates for these alternative distributions using the transformation from Xu et al. (2015).

This paper considered estimates found from complete datasets, however in most practical situations complete datasets are not used. Hence future research should apply this log-linear Weibull distribution to the many types of censoring available (see section 2.2) – Xu et al. (2015) applied this ALT Weibull model to type II censoring.

Shafiq et al. (2018) claims that life-data obtained via accelerated life testing is inevitably imprecise and hence recommends using fuzzy data – the theory that continuous data is never precise and should thus be modeled with stochastic properties. Using fuzzy data in analysis could reduce information lost in data and hence could make results more meaningful (Viertl, 2011). Therefore using fuzzy data should be considered by any future research related to this topic.

Estimates in this paper were found via three methods: the method of maximum likelihood, MCMC Bayesian estimates and Bayesian estimates found using the approximation technique from Lindley (1980). Future research can consider other methods of parameter estimation, such as the least squares approach, method of moments or even graphical estimation methods such as those used by Nelson (1990).

Tierney and Kadane (1986) proposed a method to obtain Bayesian estimates, which should also be considered to compare against the approximations of Lindley (1980) and MCMC estimates. *A priori* one would expect the estimates of Tierney and Kadane (1986) to perform

better than those of Lindley (1980) since the former are of order $\mathcal{O}\left(n^{-2}\right)$ in comparison to the latter's $\mathcal{O}\left(n^{-1}\right)$.

Finally a comparison between non-informative Bayesian estimates and subjective Bayesian estimates should also be considered – for example, Soyer et al. (2014) found estimates from the Nelson (1972) dataset using subjective Bayesian priors.

Bibliography

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, volume 55. US Government Printing Office, 1st edition.

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In *2nd International Symposium on Information Theory, 1973*, pages 267–281. Akademiai Kaido.

Arrhenius, S. (1889). Influence of Temperature on the Rate of Inversion of Sucrose. *Z. Physik. Chem*, 4(1):226–227.

Bayes, T. (1763). Lii. An Essay Towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, FRS Communicated by Mr. Price, in a Letter to John Canton, AMFR S. *Philosophical Transactions of the Royal Society of London*, 1(53):370–418.

Berger, J. O. and Bernardo, J. M. (1992). On the Development of the Reference Prior Method. *Bayesian Statistics*, 4(4):35–60.

Bernardo, J. and Smith, A. (1994). *Bayesian Theory*. Wiley, New York, 1st edition.

Bernardo, J. M. (1979). Reference Posterior Distributions for Bayesian Inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):113–128.

Brooks, S. (1998). Markov Chain Monte Carlo Method and its Application. *Journal of the Royal Statistical Society: Series D (the Statistician)*, 47(1):69–100.

Brooks, S., Gelman, A., Jones, G., and Meng, X. (2011). *Handbook of Markov Chain Monte Carlo*. CRC Press, 1st edition.

Brooks, S. P. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.

Brooks, S. P. and Morgan, B. J. (1994). Automatic Starting Point Selection for Function Optimization. *Statistics and Computing*, 4(3):173–177.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods and Research*, 33(2):261–304.

Calabria, R. and Pulcini, G. (1994). An Engineering Approach to Bayes Estimation for the Weibull Distribution. *Microelectronics Reliability*, 34(5):789–802.

Cowles, M. K. and Carlin, B. P. (1996). Markov Chain Monte Carlo Convergence Diagnostics: a Comparative Review. *Journal of the American Statistical Association*, 91(434):883–904.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. CRC Press, 1st edition.

Datta, G. S. and Ghosh, J. K. (1995a). On Priors Providing Frequentist Validity for Bayesian Inference. *Biometrika*, 82(1):37–45.

Datta, G. S. and Ghosh, M. (1995b). Some Remarks on Noninformative Priors. *Journal of the American Statistical Association*, 90(432):1357–1363.

Datta, G. S. and Sweeting, T. J. (2005). Probability Matching Priors. *Handbook of Statistics*, 25(1):91–114.

Delignette-Muller, M. L., Dutang, C., et al. (2015). Fitdistrplus: an R Package for Fitting Distributions. *Journal of Statistical Software*, 64(4):1–34.

Escobar, L. A. and Meeker, W. Q. (2006). A Review of Accelerated Test Models. *Statistical Science*, 21(4):552–577.

Fernández-i Marín, X. (2016). ggmcmc: Analysis of MCMC Samples and Bayesian Inference. *Journal of Statistical Software*, 70(9):1–20.

Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Sochastic Simulation for Bayesian Inference*. CRC Press, 1st edition.

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005). Statistical Methods for Eliciting Probability Distributions. *Journal of the American Statistical Association*, 100(470):680–701.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. (1990). Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *Journal of the American Statistical Association*, 85(412):972–985.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85(410):398–409.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. CRC Press, 3rd edition.

Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(6):721–741.

Geweke, J. (1992). Evaluating the Accurating of Sampling-Based Approaches to the Calculation of Posterior Moments. *Bayesian Statistics*, 4(144):169–193.

Gilks, W., Richardson, S., and Spiegelhalter, D. (1995). *Markov Chain Monte Carlo in Practice*. CRC Press, 1st edition.

Gilks, W. R. and Wild, P. (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41(2):337–348.

Glasstone, S., Laidler, K. J., and Eyring, H. (1941). *The Theory of Rate Processes*. McGraw-Hill, 1st edition.

Hastings, W. K. (1970). Monte carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97–109.

Henningsen, A. and Toomet, O. (2011). maxlik: a Package for Maximum Likelihood Estimation in R. *Computational Statistics*, 26(3):443–458.

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer, 1st edition.

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, 3rd edition.

Jeffreys, H. and Zellner, A. (1989). *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*. R.E. Krieger Publishing Company, 2nd edition.

Jung, M. and Chung, Y. (2018). Bayesian Inference of Three-Parameter Bathtub-Shaped Lifetime Distribution. *Communications in Statistics-Theory and Methods*, 47(17):4229–4241.

Kadane, J. B. (1995). Prime Time for Bayes. *Controlled Clinical Trials*, 16(5):313–318.

Klein, J. P. and Moeschberger, M. L. (2006). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science and Business Media, 1st edition.

Laplace, P. S. (1820). *Théorie Analytique des probabilités*. Courcier, 1st edition.

Lawson, A. (2009). *Bayesian Disease Mapping*. CRC Press, 1st edition.

Lesaffre, E. and Lawson, A. B. (2012). *Bayesian Biostatistics*. John Wiley and Sons, 1st edition.

Lindley, D. V. (1980). Approximate Bayesian Methods. *Trabajos de Estadística y de Investigación Operativa*, 31(1):223–245.

Link, W. A. and Eaton, M. J. (2012). On Thinning of Chains in MCMC. *Methods in Ecology and Evolution*, 3(1):112–115.

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2012). *The BUGS Book: a Practical Introduction to Bayesian Analysis*. CRC Press, 1st edition.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs - a Bayesian Modelling Framework: Concepts, Structure, and Extensibility. *Statistics and Computing*, 10(4):325–337.

Markov, A. A. (1906). Extension of the Law of Large Numbers to Dependent Quantities. *Izv. Fiz.-Matem. Obsch. Kazan Univ.(2nd Ser)*, 15(1):135–156.

Meeker, W. Q. and Escobar, L. A. (2014). *Statistical Methods for Reliability Data*. John Wiley and Sons, 3rd edition.

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(6):1087–1092.

Murthy, D. P., Xie, M., and Jiang, R. (2004). *Weibull Models*. John Wiley and Sons, 1st edition.

Neal, R. M. (2003). Slice Sampling. *Annals of Statistics*, 31(3):705–741.

Nelson, W. (1972). Graphical Analysis of Accelerated Life Test Data with the Inverse Power Law Model. *IEEE Transactions on Reliability*, 21(1):2–11.

Nelson, W. B. (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*. John Wiley and Sons, 1st edition.

Newcombe, R. G. (1998). Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods. *Statistics in Medicine*, 17(8):857–872.

Ntzoufras, I. (2011). *Bayesian Modeling Using WinBUGS*. John Wiley and Sons, 1st edition.

Peers, H. (1965). On Confidence Points and Bayesian Probability Points in the Case of Several Parameters. *Journal of the Royal Statistical Society: Series B (Methodological)*, 27(1):9–16.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11.

Press, S. J. (2009). *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications.* John Wiley and Sons, 2nd edition.

R Core Team (2013). R: a Language and Environment for Statistical Computing.

Ramos, E., Ramos, P. L., and Louzada, F. (2020). Posterior Properties of the Weibull Distribution for Censored Data. *Statistics and Probability Letters*, 166(1):1–8.

Rausand, M. and Høyland, A. (2003). *System Reliability Theory: Models, Statistical Methods, and Applications.* John Wiley and Sons, 1st edition.

Rinne, H. (2008). *The Weibull Distribution: a Handbook.* CRC Press, 1st edition.

Robert, C. (2007). *The Bayesian Choice: from Decision-Theoretic Foundations to Computational Implementation.* Springer Science and Business Media, 1st edition.

Robert, C. and Casella, G. (1999). *Monte Carlo Statistical Methods.* Springer Science and Business Media, 1st edition.

Schlaifer, R. and Raiffa, H. (1961). *Applied Statistical Decision Theory.* Division of Research, Graduate School of Business Administration, Harvard University, 1st edition.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Shafiq, M., Atif, M., and Viertl, R. (2018). Beyond Precision: Accelerated Life Testing for Fuzzy Life Time Data. *Soft Computing*, 22(22):7355–7365.

Shaked, M. and Singpurwalla, N. D. (1982). Nonparametric Estimation and Goodness-of-fit Testing of Hypotheses for Distributions in Accelerated Life Testing. *IEEE Transactions on Reliability*, 31(1):69–74.

Singpurwalla, N. D. (2006). *Reliability and Risk: a Bayesian Perspective.* John Wiley and Sons, 1st edition.

Soyer, R., Erkanli, A., and Merrick, J. R. (2014). Accelerated Life Tests: Bayesian Models. *Encyclopedia of Statistics in Quality and Reliability*, 1(1):20–30.

Soyer, R., Mazzuchi, T. A., and Singpurwalla, N. D. (2012). *Mathematical reliability: an Expository Perspective.* Springer Science and Business Media, 1st edition.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van der Linde, A. (2014). The Deviance Information Criterion: 12 Years On. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 76(3):485–493.

Sturtz, S., Ligges, U., and Gelman, A. (2005). R2winbugs: a Package for Running WinBUGS from R. *Journal of Statistical Software*, 12(3):1–16.

Sun, D. and Berger, J. O. (1998). Reference Priors with Partial Information. *Biometrika*, 85(1):55–71.

Sun, D., Ghosh, M., and Basu, A. P. (1998). Bayesian Analysis for a Stress-Strength System under Noninformative Priors. *Canadian Journal of Statistics*, 26(2):323–332.

Tierney, L. and Kadane, J. B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association*, 81(393):82–86.

Tierney, L., Kass, R. E., and Kadane, J. B. (1989). Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions. *Journal of the American Statistical Association*, 84(407):710–716.

Viertl, R. (2011). *Statistical Methods for Fuzzy Data*. John Wiley and Sons, 1st edition.

Weibull, W. (1939). A Statistical Theory of Strength of Materials. *IVB-Handl.*, 18(3):293–297.

Xu, A., Fu, J., Tang, Y., and Guan, Q. (2015). Bayesian Analysis of Constant-Stress Accelerated Life Test for the Weibull Distribution Using Noninformative Priors. *Applied Mathematical Modelling*, 39(20):6183–6195.

Zacks, S. (2012). *Introduction to Reliability Analysis: Probability Models and Statistical Methods*. Springer Science and Business Media, 1st edition.

Zellner, A. (1977). Maximal Data Information Prior Distributions. *New Developments in the Applications of Bayesian Methods*, 1(1):211–232.

Zellner, A. (1996). Models, Prior Information, and Bayesian Analysis. *Journal of Econometrics*, 75(1):51–68.

Appendix A: Additional results for Chapter 3

A.1 Fisher information matrix for the log-linear Weibull distribution

Proof. Consider the second-order partial derivatives of the log-likelihood defined in equation 3.8:

$$\frac{\partial^2 \mathcal{L}}{\partial \lambda_0^2} = -\frac{n\beta}{\lambda_0^2} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\beta\left(\beta - 1\right)\lambda_0^{\beta-2}\eta^{\beta\delta_i}, \tag{A.1}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \eta^2} = -\frac{\bar{\delta}\beta}{\eta^2} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta}\beta\delta_i\left(\beta\delta_i - 1\right)\eta^{\beta\delta_i-2}, \tag{A.2}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta^2} = -\frac{n}{\beta^2} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}log^2\left(t_{ij}\right)\left(\lambda_0\eta^{\delta_i}\right)^{\beta} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\left(\lambda_0\eta^{\delta_i}\right)^{\beta}log^2\left(\lambda_0\eta^{\delta_i}\right) -$$
$$-2\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}log\left(t_{ij}\right)\left(\lambda_0\eta^{\delta_i}\right)^{\beta}log\left(\lambda_0\eta^{\delta_i}\right), \tag{A.3}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \lambda_0 \partial \beta} = \frac{n}{\lambda_0} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\left[\lambda_0^{\beta-1}\eta^{\beta\delta_i} + \frac{\beta}{\lambda_0}\left(\lambda_0\eta^{\delta_i}\right)^{\beta}log\left(\lambda_0\eta^{\delta_i}\right)\right] -$$
$$-\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}log\left(t_{ij}\right)\left(\beta\lambda_0^{\beta-1}\eta^{\beta\delta_i}\right), \tag{A.4}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \eta \partial \beta} = \frac{\bar{\delta}}{\eta} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\left[\delta_i\lambda_0\eta^{\beta\delta_i-1} + \beta\delta_i\lambda_0^{\beta}\eta^{\beta\delta_i-1}log\left(\lambda_0\eta^{\delta_i}\right)\right] -$$
$$-\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}log\left(t_{ij}\right)\left[\beta\delta_i\lambda_0\eta^{\beta\delta_i-1}\right], \tag{A.5}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \lambda_0 \partial \eta} = -\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\beta^2\delta_i\lambda_0^{\beta-1}\eta^{\beta\delta_i-1}. \tag{A.6}$$

Define the transformation:

$$Y_i = \left(\frac{X_i}{\theta_i}\right)^{\beta} = \left(X_i \lambda_0 \eta^{\delta_i}\right)^{\beta}, \tag{A.7}$$

then Y_i is an exponential random variable with rate parameter given by $\theta^{\star} = 1$ and hence an expected value of unity.

Define:

$$c_{1i} = E\left[\sum_{j=1}^{n_i} Y_{ij} log\left(Y_{ij}\right)\right] \tag{A.8}$$

and

$$c_{2i} = E\left[\sum_{j=1}^{n_i} Y_{ij} log^2\left(Y_{ij}\right)\right]. \tag{A.9}$$

Furthermore, note that:

$$Y_{ij} log\left(Y_{ij}\right) = \left(X_i \lambda_0 \eta^{\delta_i}\right)^{\beta} \beta \left[log\left(X_i\right) + log\left(\lambda_0 \eta^{\delta_i}\right)\right], \tag{A.10}$$

and that

$$Y_{ij} log^2\left(Y_{ij}\right) = \left(X_i \lambda_0 \eta^{\delta_i}\right)^{\beta} \beta^2 \left[log^2\left(X_i\right) + 2log\left(X_i\right) log\left(\lambda_0 \eta^{\delta_i}\right) + log^2\left(\lambda_0 \eta^{\delta_i}\right)\right]. \tag{A.11}$$

Consider the expectation of equation A.10:

$$\begin{aligned}
E\left[Y_{ij} log\left(Y_{ij}\right)\right] &= \int_0^{\infty} y_{ij} log\left(y_{ij}\right) \times \sim Exp\left(\theta^{\star} = 1\right) dy_{ij} \\
&= \int_0^{\infty} y_{ij} log\left(y_{ij}\right) exp\left(-y_{ij}\right) dy_{ij}.
\end{aligned} \tag{A.12}$$

This can be solved by parts by letting:

- $v_{ij} = y_{ij} log\left(y_{ij}\right) \implies dv_{ij} = \left(log\left(y_{ij}\right) + 1\right) dy_{ij}$ and

- $du_{ij} = exp\left\{-y_{ij}\right\} \implies u_{ij} = -exp\left\{-y_{ij}\right\}$.

Hence:

$$\begin{aligned}
E\left[Y_{ij} log\left(Y_{ij}\right)\right] &= \int_0^{\infty} \left(y_{ij} log\left(y_{ij}\right) + 1\right) exp\left\{-y_{ij}\right\} dy_{ij} \\
&= \int_0^{\infty} y_{ij} log\left(y_{ij}\right) exp\left\{-y_{ij}\right\} dy_{ij} + \int_0^{\infty} exp\left\{-y_{ij}\right\} dy_{ij} \\
&= -\gamma + 1,
\end{aligned} \tag{A.13}$$

where:

$$\psi\left(1\right) = -\gamma = \int_0^{\infty} log\left(y_{ij}\right) exp\left\{-y_{ij}\right\} dy_{ij} \approx -0.5772, \tag{A.14}$$

is Euler's constant (Abramowitz and Stegun, 1964).

Therefore:

$$c_{1i} = E\left[\sum_{j=1}^{n_i} Y_{ij}log\left(Y_{ij}\right)\right] = n_i\left(-\gamma + 1\right). \tag{A.15}$$

Now consider the expectation of equation A.11:

$$
\begin{aligned}
E\left[Y_{ij}log^2\left(Y_{ij}\right)\right] &= \int_0^\infty y_{ij}log^2\left(y_{ij}\right) \cdot \sim Exp\left(\theta^\star = 1\right)dy_{ij} \\
&= \int_0^\infty y_{ij}log^2\left(y_{ij}\right)exp\left(-y_{ij}\right)dy_{ij}.
\end{aligned}
\tag{A.16}
$$

Solving by parts, let:

- $v_{ij} = y_{ij}log^2\left(y_{ij}\right) \implies dv_{ij} = \left(log^2\left(y_{ij}\right) + 2log\left(y_{ij}\right)\right)dy_{ij}$ and

- $du_{ij} = exp\left\{-y_{ij}\right\} \implies u_{ij} = -exp\left\{-y_{ij}\right\}$.

Hence:

$$
\begin{aligned}
E\left[Y_{ij}log^2\left(Y_{ij}\right)\right] &= \int_0^\infty \left[2log\left(Y_{ij}\right) + log^2\left(y_{ij}\right)\right]exp\left\{-y_{ij}\right\}dy_{ij} \\
&= 2\int_0^\infty log\left(y_{ij}\right)exp\left\{-y_{ij}\right\}dy_{ij} + \int_0^\infty log^2\left(y_{ij}\right)exp\left\{-y_{ij}\right\}dy_{ij} \\
&= 2\gamma + \gamma^2 + \frac{\pi^2}{6},
\end{aligned}
\tag{A.17}
$$

where:

$$\int_0^\infty log^2\left(y_{ij}\right)exp\left\{-y_{ij}\right\}dy_{ij} = \gamma^2 + \frac{\pi^2}{6} \tag{A.18}$$

(Abramowitz and Stegun, 1964).

Therefore:

$$c_{2i} = E\left[\sum_{j=1}^{n_i} Y_{ij}log^2\left(Y_{ij}\right)\right] = n_i\left(2\gamma + \gamma^2 + \frac{\pi^2}{6}\right).$$

Finally define the following products:

- $c_1 = \sum_{i=1}^k c_{1i} = n\left(-\gamma + 1\right),$

- $c_2 = \sum_{i=1}^k c_{2i} = n\left(2\gamma + \gamma^2 + \frac{\pi^2}{6}\right),$

- $c_3 = \sum_{i=1}^k c_{1i}\delta_i = \bar{\delta}\left(-\gamma + 1\right),$

- $c_4 = \sum\limits_{i=1}^{k} n_i \delta_i^2$.

Thus, consider the following expected values of the second-order partial derivatives:

$$
\begin{aligned}
E_{t_{ij}}\left[\frac{\partial^2 \mathcal{L}}{\partial \lambda_0 \partial \beta}\right] =& \frac{n}{\lambda_0} - E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\left[\lambda_0^{\beta-1}\eta^{\beta\delta_i} + \frac{\beta}{\lambda_0}\left(\lambda_0\eta^{\delta_i}\right)^{\beta} log\left(\lambda_0\eta^{\delta_i}\right)\right]\right] - \\
& - E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta} log\left(t_{ij}\right)\beta\lambda_0^{\beta-1}\eta^{\beta\delta_i}\right] \\
=& \frac{n}{\lambda_0} - \frac{1}{\lambda_0}E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(t_{ij}\lambda_0\eta^{\delta_i}\right)^{\beta}\beta\left[log\left(t_{ij}\right) + log\left(\lambda_0\eta^{\delta_i}\right)\right]\right] - \\
& - \frac{1}{\lambda_0}E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(t_{ij}\lambda_0\eta^{\delta_i}\right)^{\beta}\right] \\
=& \frac{n}{\lambda_0} - \frac{c_1}{\lambda_0} - \frac{\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} E\left[\sim Exp\left(\theta^{\star}=1\right)\right]}{\lambda_0} \\
=& \frac{n}{\lambda_0} - \frac{c_1}{\lambda_0} - \frac{n}{\lambda_0} \\
=& -\frac{c_1}{\lambda_0}.
\end{aligned}
\tag{A.19}
$$

$$
\begin{aligned}
E_{t_{ij}}\left[\frac{\partial^2 \mathcal{L}}{\partial \eta \partial \beta}\right] =& \frac{\bar{\delta}}{\eta} - E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\left[\delta_i\lambda_0\eta^{\beta\delta_i-1} + \beta\delta_i\lambda_0^{\beta}\eta^{\beta\delta_i-1}log\left(\lambda_0\eta^{\delta_i}\right)\right]\right] - \\
& - E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta} log\left(t_{ij}\right)\left[\beta\delta_i\lambda_0\eta^{\beta\delta_i-1}\right]\right] \\
=& \frac{\bar{\delta}}{\eta} - \frac{1}{\eta}E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(t_{ij}\lambda_0\eta^{\delta_i}\right)^{\beta}\beta\delta_i\left[log\left(t_{ij}\right) + log\left(\lambda_0\eta^{\delta_i}\right)\right]\right] - \\
& - \frac{1}{\eta}E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(t_{ij}\lambda_0\eta^{\delta_i}\right)^{\beta}\delta_i\right] \\
=& \frac{\bar{\delta}}{\eta} - \frac{c_3}{\eta} - \frac{\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i}\delta_i E\left[\sim Exp\left(\theta^{\star}=1\right)\right]}{\eta} \\
=& \frac{\bar{\delta}}{\eta} - \frac{c_3}{\eta} - \frac{\bar{\delta}}{\eta} \\
=& -\frac{c_3}{\eta}.
\end{aligned}
\tag{A.20}
$$

$$
\begin{aligned}
E_{t_{ij}}\left[\frac{\partial^2 \mathcal{L}}{\partial \beta^2}\right] =& -\frac{n}{\beta^2} - E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta} log^2\left(t_{ij}\right)\right]\left(\lambda_0 \eta^{\delta_i}\right)^{\beta} - E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\left(\lambda_0 \eta^{\delta_i}\right)^{\beta} log^2\left(\lambda_0 \eta^{\delta_i}\right)\right] - \\
& -2E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta} log\left(t_{ij}\right)\right]\left(\lambda_0 \eta^{\delta_i}\right)^{\beta} log\left(\lambda_0 \eta^{\delta_i}\right) \\
=& -\frac{n}{\beta^2} - E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(t_{ij}\lambda_0\eta^{\delta_i}\right)^{\beta}\beta^2\left[log^2\left(t_{ij}\right) + 2log\left(t_{ij}\right)log\left(\lambda_0\eta^{\delta_i}\right) + log^2\left(\lambda_0\eta^{\delta_i}\right)\right]\right] \\
=& -\frac{n}{\beta^2} - \frac{c_2}{\beta^2} \\
=& -\frac{n+c_2}{\beta^2}.
\end{aligned}
$$

$$\text{(A.21)}$$

It is known that $\dfrac{2\sum_{j=1}^{n_i} t_{ij}^{\beta}}{\theta_i^{\beta}}$ follows a χ^2 with $2n_i$ degrees of freedom, and therefore

$E_{t_{ij}}\left[\sum_{j=1}^{n_i} t_{ij}^{\beta}\right] = n_i\theta_i^{\beta} = \dfrac{n_i}{\left(\lambda_0\eta^{\delta_i}\right)^{\beta}}$. Hence:

$$
\begin{aligned}
E_{t_{ij}}\left[\frac{\partial^2 \mathcal{L}}{\partial \lambda_0^2}\right] =& -\frac{n\beta}{\lambda_0^2} - E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\beta\left(\beta-1\right)\lambda_0^{\beta-2}\eta^{\beta\delta_i}\right] \\
=& -\frac{n\beta}{\lambda_0^2} - \frac{n\beta\left(\beta-1\right)}{\lambda_0^2} \\
=& -\frac{n\beta^2}{\lambda_0^2}.
\end{aligned}
$$

$$\text{(A.22)}$$

$$
\begin{aligned}
E_{t_{ij}}\left[\frac{\partial^2 \mathcal{L}}{\partial \eta^2}\right] =& -\frac{\bar{\delta}\beta}{\eta^2} - E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\lambda_0^{\beta}\beta\delta_i\left(\beta\delta_i-1\right)\eta^{\beta\delta_i-2}\right] \\
=& -\frac{\bar{\delta}\beta}{\eta^2} - \frac{\beta^2 c_4 - \bar{\delta}\beta}{\eta^2} \\
=& -\frac{\beta^2 c_4}{\eta^2}.
\end{aligned}
$$

$$\text{(A.23)}$$

$$
\begin{aligned}
E_{t_{ij}}\left[\frac{\partial^2 \mathcal{L}}{\partial \lambda_0 \partial \eta}\right] =& -E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta}\beta^2\delta_i\lambda_0^{\beta-1}\eta^{\beta\delta_i-1}\right] \\
=& -\frac{\bar{\delta}\beta^2}{\lambda_0\eta}.
\end{aligned}
$$

$$\text{(A.24)}$$

Hence, the Fisher information matrix of $\boldsymbol{\psi}$ is given as:

$$
I\left(\boldsymbol{\psi}\right) = \begin{bmatrix}
\frac{n\beta^2}{\lambda_0^2} & \frac{\bar{\delta}\beta^2}{\lambda_0\eta} & \frac{c_1}{\lambda_0} \\
\frac{\bar{\delta}\beta^2}{\lambda_0\eta} & \frac{\beta^2 c_4}{\eta^2} & \frac{c_3}{\eta} \\
\frac{c_1}{\lambda_0} & \frac{c_3}{\eta} & \frac{n+c_2}{\beta^2}
\end{bmatrix}.
$$

$$\text{(A.25)}$$

$\square$

A.2   Derivations for reference priors

A.2.1   Reference prior for the grouping $\{\lambda_0, \eta, \beta\}$

Proof. Let the reference prior for the grouping $\{\lambda_0, \eta, \beta\}$ be $\pi_{R1}(\boldsymbol{\psi})$. This grouping order implies that λ_0 is regarded as the most important parameter, η is regarded as second in importance and β is of least importance.

The inverse of the fishers information matrix is given by:

$$
I^{-1}(\boldsymbol{\psi}) = \frac{1}{M}
\begin{bmatrix}
\frac{\left(nc_4 + c_2 c_4 - c_3^2\right)\lambda_0^2}{\beta^2} & \frac{\left(c_1 c_3 - n\bar{\delta} - c_2\bar{\delta}\right)\lambda_0\eta}{\beta^2} & \left(c_3\bar{\delta} - c_1 c_4\right)\lambda_0 \\
\frac{\left(c_1 c_3 - n\bar{\delta} - c_2\bar{\delta}\right)\lambda_0\eta}{\beta^2} & \frac{\left(n^2 + nc_2 - c_1^2\right)\eta^2}{\beta^2} & \left(c_1\bar{\delta} - nc_3\right)\eta \\
\left(c_3\bar{\delta} - c_1 c_4\right)\lambda_0 & \left(c_1\bar{\delta} - nc_3\right)\eta & \left(\bar{\delta}^2 - nc_4\right)\beta^2
\end{bmatrix}
= \frac{1}{M}
\begin{bmatrix}
\frac{a_{11}\lambda_0^2}{\beta^2} & \frac{a_{12}\lambda_0\eta}{\beta^2} & a_{13}\lambda_0 \\
\frac{a_{21}\lambda_0\eta}{\beta^2} & \frac{a_{22}\eta^2}{\beta^2} & a_{23}\eta \\
a_{31}\lambda_0 & a_{32}\eta & a_{33}\beta^2
\end{bmatrix},
$$

(A.26)

where a_{ij}, $i,j = 1,2,3$ are the corresponding multipliers of the parameters left in the larger matrix.

Following the notation and procedure of Sun et al. (1998), write the inverse of the Fisher information matrix as:

$$
I^{-1}(\boldsymbol{\psi}) =
\begin{bmatrix}
j_{11} & j_{12} & j_{13} \\
j_{21} & j_{22} & j_{23} \\
j_{31} & j_{32} & j_{33}
\end{bmatrix}^{-1}
=
\begin{bmatrix}
j^{11} & j^{12} & j^{13} \\
j^{21} & j^{22} & j^{23} \\
j^{31} & j^{32} & j^{33}
\end{bmatrix},
$$

(A.27)

that is, the elements j_{ik}, $i,k = 1,2,3$ represent the $(i,k)^{th}$ element of the Fisher information matrix, and the elements j^{ik}, $i,k = 1,2,3$ represent the $(i,k)^{th}$ element of the inverse of the Fisher information matrix.

To find h_1, form a 1×1 matrix of the top-left hand corner of A.26, and invert that resulting matrix. Therefore the value h_1 is given by the inverse of the first element of the inverse Fisher information matrix:

$$
h_1 = \frac{1}{j^{11}} = \frac{M\beta^2}{a_{11}\lambda_0^2}.
$$

(A.28)

To find h_2, form a 2×2 matrix of the top-left hand corner of A.26, and invert that resulting matrix. That 2×2 matrix is given by:

$$
I_2^{-1}(\boldsymbol{\psi}) =
\begin{bmatrix}
j^{11} & j^{12} \\
j^{21} & j^{22}
\end{bmatrix}
= \frac{1}{M}
\begin{bmatrix}
\frac{a_{11}\lambda_0^2}{\beta^2} & \frac{a_{12}\lambda_0\eta}{\beta^2} \\
\frac{a_{21}\lambda_0\eta}{\beta^2} & \frac{a_{22}\eta^2}{\beta^2}
\end{bmatrix}.
$$

(A.29)

And hence the inverse of the matrix given in equation A.29 is:

$$
I_2(\boldsymbol{\psi}) = \frac{M\beta^4}{k\lambda_0^2\eta^2}
\begin{bmatrix}
\frac{a_{22}\eta^2}{\beta^2} & -\frac{a_{21}\lambda_0\eta}{\beta^2} \\
-\frac{a_{12}\lambda_0\eta}{\beta^2} & \frac{a_{11}\lambda_0^2}{\beta^2}
\end{bmatrix},
$$

(A.30)

where $k = a_{11}a_{22} - a_{12}a_{21}$.

Therefore h_2 is given as the bottom right-hand element of the matrix given in equation A.30:

$$h_2 = \frac{a_{11} M \beta^2}{k \eta^2}, \tag{A.31}$$

To find h_3, form a 3×3 matrix of the top-left hand corner of A.26, and invert that resulting matrix. The 3×3 matrix of the top-left hand corner of A.26 is *de facto* the entire matrix of the inverse of the Fisher information matrix, which inverted is the Fisher information matrix itself. Hence the value of h_3 is given by:

$$h_3 = j_{33} = \frac{n + c_2}{\beta^2}. \tag{A.32}$$

Choose $\Omega_s = [a_{1s}, b_{1s}] \times [a_{2s}, b_{2s}] \times [a_{3s}, b_{3s}]$ as a collection of compact sets in $(0, \infty) \times (1, \infty) \times (0, \infty)$, such that: $a_{1s}, a_{3s} \to 0$, $a_{2s} \to 1$, $b_{1s}, b_{2s}, b_{3s} \to \infty$.

The conditional prior of β given λ_0 and η is given by:

$$
\begin{aligned}
\pi_3^s (\beta \mid \lambda_0, \eta) &= \frac{\sqrt{h_3} I_{[a_{3s}, b_{3s}]} (\beta)}{\int\limits_{a_{3s}}^{b_{3s}} \sqrt{h_3} d\beta} \\
&= \frac{\sqrt{\frac{n + c_2}{\beta^2}} I_{[a_{3s}, b_{3s}]} (\beta)}{\sqrt{n + c_2} \int\limits_{a_{3s}}^{b_{3s}} \frac{1}{\beta} d\beta} \\
&= \frac{I_{[a_{3s}, b_{3s}]} (\beta)}{\beta log \left(\frac{b_{3s}}{a_{3s}} \right)}.
\end{aligned}
\tag{A.33}
$$

Given λ_0, the conditional prior of (β, η) is given by:

$$
\begin{aligned}
\pi_2^s (\beta, \eta \mid \lambda_0) &= \frac{\pi_3^s (\beta \mid \lambda_0, \eta) \, exp \left\{ \frac{1}{2} E \left[log (h_2) \mid \lambda_0, \eta \right] \right\} I_{[a_{2s}, b_{2s}]} (\eta)}{\int\limits_{a_{2s}}^{b_{2s}} exp \left\{ \frac{1}{2} E \left[log (h_2) \mid \lambda_0, \eta \right] \right\} d\eta} \\
&= \frac{\pi_3^s (\beta \mid \lambda_0, \eta) \, exp \left\{ \frac{1}{2} log \left(\frac{a_{11} M \beta^2}{k \eta^2} \right) \right\} I_{[a_{2s}, b_{2s}]} (\eta)}{\int\limits_{a_{2s}}^{b_{2s}} exp \left\{ \frac{1}{2} log \left(\frac{a_{11} M \beta^2}{k \eta^2}, \right) \right\} d\eta} \\
&\propto \frac{I_{[a_{2s}, b_{2s}] \times [a_{3s}, b_{3s}]} (\eta, \beta)}{\eta \beta log \left(\frac{b_{3s}}{a_{3s}} \right) log \left(\frac{b_{2s}}{a_{2s}} \right)}.
\end{aligned}
\tag{A.34}
$$

The joint prior of (λ_0, η, β) on Ω_s is given by:

$$
\begin{aligned}
\pi_1^s(\lambda_0, \beta, \eta) &= \frac{\pi_2^s(\beta, \eta \mid \lambda_0) \, exp\left\{\frac{1}{2}E\left[log(h_1) \mid \lambda_0\right]\right\} I_{[a_{1s}, b_{1s}]}(\lambda_0)}{\displaystyle\int_{a_{1s}}^{b_{1s}} exp\left\{\frac{1}{2}E\left[log(h_1) \mid \lambda_0\right]\right\} d\lambda_0} \\[2em]
&= \frac{\pi_2^s(\beta, \eta \mid \lambda_0) \, exp\left\{\frac{1}{2}log\left(\frac{M\beta^2}{a_{11}\lambda_0^2}\cdot\right)\right\} I_{[a_{1s}, b_{1s}]}(\lambda_0)}{\displaystyle\int_{a_{1s}}^{b_{1s}} exp\left\{\frac{1}{2}log\left(\frac{M\beta^2}{a_{11}\lambda_0^2}\cdot\right)\right\} d\lambda_0} \\[2em]
&\propto \frac{I_{[a_{1s}, b_{1s}]\times[a_{2s}, b_{2s}]\times[a_{3s}, b_{3s}]}(\lambda_0, \eta, \beta)}{\lambda_0 \eta \beta log\left(\frac{b_{3s}}{a_{3s}}\right) log\left(\frac{b_{2s}}{a_{2s}}\right) log\left(\frac{b_{1s}}{a_{1s}}\right)}.
\end{aligned}
\tag{A.35}
$$

Hence, the reference prior is given as:

$$
\pi_{R1}(\boldsymbol{\psi}) = \lim_{s \to \infty} \frac{\pi_1^s(\lambda_0, \beta, \eta)}{\pi_1^s(1, 2, 1)} = \frac{1}{\lambda_0 \eta \beta},
\tag{A.36}
$$

assuming that $(1, 2, 1)$ is a point on $[a_{1s}, b_{1s}] \times [a_{2s}, b_{2s}] \times [a_{3s}, b_{3s}]$. $\square$

A.2.2   Reference prior for the grouping $\{\lambda_0, (\eta, \beta)\}$

Proof. Let the reference prior for the grouping $\{\lambda_0, (\eta, \beta)\}$ be $\pi_{R2}(\boldsymbol{\psi})$. This grouping implies that λ_0 is regarded as the most important parameter, whereas η and β are assumed to be nuisance parameters.

Following the method found in Sun and Berger (1998), first define a subset of the Fisher information matrix for η and β as:

$$
I_2(\boldsymbol{\psi}) = \begin{bmatrix} j_{22} & j_{23} \\ j_{32} & j_{33} \end{bmatrix} = \begin{bmatrix} \frac{\beta^2 c_4}{\eta^2} & \frac{c_3}{\eta} \\ \frac{c_3}{\eta} & \frac{n+c_2}{\beta^2} \end{bmatrix}.
\tag{A.37}
$$

Then the value of h_1 is given by:

$$
h_1 = \frac{det(I(\boldsymbol{\psi}))}{det(I_2(\boldsymbol{\psi}))} = \frac{M\beta^2}{\left(c_4(n+c_2) - c_3^2\right)\lambda_0^2},
\tag{A.38}
$$

and also:

$$
h_2 = det(I_2(\boldsymbol{\psi})) = \frac{c_4(n+c_2) - c_3^2}{\eta^2}.
\tag{A.39}
$$

Thus the conditional prior of (η, β) given λ_0 is given by:

$$
\begin{aligned}
\pi_2^k (\eta, \beta \mid \lambda_0) &= \frac{\sqrt{h_2} I_{[a_{2s}, b_{2s}] \times [a_{3s}, b_{3s}]} (\eta, \beta)}{\int\limits_{a_{3s}}^{b_{3s}} \int\limits_{a_{2s}}^{b_{2s}} \sqrt{h_2} d\beta d\eta} \\
&= \frac{\sqrt{\frac{c_4 (n + c_2) - c_3^2}{\eta^2}} I_{[a_{2s}, b_{2s}] \times [a_{3s}, b_{3s}]} (\eta, \beta)}{\sqrt{c_4 (n + c_2) - c_3^2} \int\limits_{a_{3s}}^{b_{3s}} \int\limits_{a_{2s}}^{b_{2s}} \frac{1}{\eta} d\beta d\eta} \\
&= \frac{I_{[a_{2s}, b_{2s}] \times [a_{3s}, b_{3s}]} (\eta, \beta)}{\eta \, log \left(\frac{b_{2s}}{a_{2s}} \right) (b_{3s} - a_{3s})}.
\end{aligned}
\tag{A.40}
$$

And the marginal prior of λ_0 is given by:

$$
\begin{aligned}
\pi_1^k (\lambda_0) &\propto exp \left\{ \frac{1}{2} E \left[log (h_1) \mid \lambda_0 \right] \right\} I_{[a_{1s}, b_{1s}]} (\lambda_0) \\
&= exp \left\{ \frac{1}{2} \int\limits_{a_{3s}}^{b_{3s}} \int\limits_{a_{2s}}^{b_{2s}} \pi_2^k (\eta, \beta \mid \lambda_0) \, log (h_1) \, d\eta d\beta \right\} I_{[a_{1s}, b_{1s}]} (\lambda_0) \\
&\propto \frac{I_{[a_{1s}, b_{1s}]} (\lambda_0)}{\lambda_0}.
\end{aligned}
\tag{A.41}
$$

Thus the reference prior for $\{\lambda_0, (\eta, \beta)\}$ is given by:

$$
\pi_{R2} (\boldsymbol{\psi}) = \lim_{s \to \infty} \frac{\pi_2^k (\eta, \beta \mid \lambda_0) \pi_1^k (\lambda_0)}{\pi_2^k (2, 1 \mid 1) \pi_1^k (1)} = \frac{1}{\lambda_0 \eta}.
\tag{A.42}
$$

$\square$

A.2.3   Reference prior for the grouping $\{\eta (\lambda_0, \beta)\}$

Proof. Let the reference prior for the grouping $\{\eta, (\lambda_0, \beta)\}$ be $\pi_{R2} (\boldsymbol{\psi})$. This implies that η is regarded as the most important parameter, whereas λ_0 and β are assumed to be nuisance parameters.

First define a subset of the Fisher information matrix for λ_0 and β as:

$$
I_2 (\boldsymbol{\psi}) = \begin{bmatrix} j_{11} & j_{13} \\ j_{31} & j_{33} \end{bmatrix} = \begin{bmatrix} \frac{n\beta^2}{\lambda_0^2} & \frac{c_1}{\lambda_0} \\ \frac{c_1}{\lambda_0} & \frac{n + c_2}{\beta^2} \end{bmatrix}.
\tag{A.43}
$$

Then:

$$
h_1 = \frac{det (I (\boldsymbol{\psi}))}{det (I_2 (\boldsymbol{\psi}))} = \frac{M\beta^2}{(n (n + c_2) - c_1) \eta^2},
\tag{A.44}
$$

and

$$
h_2 = det (I_2 (\boldsymbol{\psi})) = \frac{n (n + c_2) - c_1}{\lambda_0^2}.
\tag{A.45}
$$

Thus the conditional prior of (λ_0, β) given η is given by:

$$
\begin{aligned}
\pi_2^k (\lambda_0, \beta \mid \eta) &= \frac{\sqrt{h_2} I_{[a_{1s}, b_{1s}] \times [a_{3s}, b_{3s}]} (\lambda_0, \beta)}{\int\limits_{a_{3s}}^{b_{3s}} \int\limits_{a_{1s}}^{b_{1s}} \sqrt{h_2} \, d\beta d\lambda_0} \\
&= \frac{\sqrt{\frac{n(n+c_2)-c_1}{\lambda_0^2}} I_{[a_{1s}, b_{1s}] \times [a_{3s}, b_{3s}]} (\lambda_0, \beta)}{\sqrt{n(n+c_2)-c_1} \int\limits_{a_{3s}}^{b_{3s}} \int\limits_{a_{1s}}^{b_{1s}} \frac{1}{\lambda_0} d\beta d\lambda_0} \\
&= \frac{I_{[a_{1s}, b_{1s}] \times [a_{3s}, b_{3s}]} (\lambda_0, \beta)}{\lambda_0 log \left(\frac{b_{1s}}{a_{1s}} \right) (b_{3s} - a_{3s})}.
\end{aligned}
\tag{A.46}
$$

And the marginal prior of η is given by:

$$
\begin{aligned}
\pi_1^k (\eta) &\propto exp \left\{ \frac{1}{2} E \left[log (h_1) \mid \eta \right] \right\} I_{[a_{2s}, b_{2s}]} (\eta) \\
&= exp \left\{ \frac{1}{2} \int\limits_{a_{3s}}^{b_{3s}} \int\limits_{a_{1s}}^{b_{1s}} \pi_2^k (\lambda_0, \beta \mid \eta) \, log (h_1) \, d\lambda_0 d\beta \right\} I_{[a_{2s}, b_{2s}]} (\eta) \\
&\propto \frac{I_{[a_{2s}, b_{2s}]} (\eta)}{\eta}.
\end{aligned}
\tag{A.47}
$$

Thus the reference prior for $\{\eta, (\lambda_0, \beta)\}$ is given by:

$$
\pi_{R2} (\boldsymbol{\psi}) = \lim_{s \to \infty} \frac{\pi_2^k (\lambda_0, \beta \mid \eta) \pi_1^k (\eta)}{\pi_2^k (1, 1 \mid 2) \pi_1^k (1)} = \frac{1}{\lambda_0 \eta}.
\tag{A.48}
$$

$\square$

A.2.4 Reference prior for the grouping $\{\beta, (\lambda_0, \eta)\}$

Proof. Let the reference prior for the grouping $\{\beta, (\lambda_0, \eta)\}$ be $\pi_{R1} (\boldsymbol{\psi})$. Then β is regarded as the most important parameter, whereas λ_0 and η are assumed to be nuisance parameters.

Furthermore define a subset of the Fisher information matrix for λ_0 and η as:

$$
I_2 (\boldsymbol{\psi}) = \begin{bmatrix} j_{11} & j_{12} \\ j_{21} & j_{22} \end{bmatrix} = \begin{bmatrix} \frac{n\beta^2}{\lambda_0^2} & \frac{\bar{\delta}\beta^2}{\lambda_0 \eta} \\ \frac{\bar{\delta}\beta^2}{\lambda_0 \eta} & \frac{\beta^2 c_4}{\eta^2} \end{bmatrix}.
\tag{A.49}
$$

Then:

$$
h_1 = \frac{det (I (\boldsymbol{\psi}))}{det (I_2 (\boldsymbol{\psi}))} = \frac{M}{\beta^2 (nc_4 - \bar{\delta})},
\tag{A.50}
$$

and

$$
h_2 = det (I_2 (\boldsymbol{\psi})) = \frac{\beta^4 (nc_4 - \bar{\delta})}{\lambda_0^2 \eta^2}.
\tag{A.51}
$$

Thus the conditional prior of (λ_0, η) given β is given by:

$$
\pi_2^k (\lambda_0, \eta \mid \beta) = \frac{\sqrt{h_2} I_{[a_{1s}, b_{1s}] \times [a_{2s}, b_{2s}]} (\lambda_0, \eta)}{\int\limits_{a_{2s}}^{b_{2s}} \int\limits_{a_{1s}}^{b_{1s}} \sqrt{h_2} d\eta d\lambda_0}
$$

$$
= \frac{\sqrt{\frac{\beta^4 \left(nc_4 - \bar{\delta}\right)}{\lambda_0^2 \eta^2}} I_{[a_{1s}, b_{1s}] \times [a_{2s}, b_{2s}]} (\lambda_0, \eta)}{\sqrt{\beta^4 \left(nc_4 - \bar{\delta}\right)} \int\limits_{a_{2s}}^{b_{2s}} \int\limits_{a_{1s}}^{b_{1s}} \frac{1}{\lambda_0 \eta} d\eta d\lambda_0} \tag{A.52}
$$

$$
= \frac{I_{[a_{1s}, b_{1s}] \times [a_{3s}, b_{3s}]} (\lambda_0, \beta)}{\lambda_0 \eta \log\left(\frac{b_{1s}}{a_{1s}}\right) \left(\frac{b_{2s}}{a_{2s}}\right)}.
$$

And the marginal prior of β is given by:

$$
\pi_1^k (\beta) \propto exp \left\{ \frac{1}{2} E \left[\log (h_1) \mid \beta \right] \right\} I_{[a_{3s}, b_{3s}]} (\beta)
$$

$$
= exp \left\{ \frac{1}{2} \int\limits_{a_{2s}}^{b_{2s}} \int\limits_{a_{1s}}^{b_{1s}} \pi_2^k (\lambda_0, \eta \mid \beta) \log (h_1) d\eta d\lambda_0 \right\} I_{[a_{3s}, b_{3s}]} (\beta) \tag{A.53}
$$

$$
\propto \frac{I_{[a_{3s}, b_{3s}]} (\beta)}{\beta}.
$$

Thus the reference prior for $\{\beta, (\lambda_0, \eta)\}$ is given by:

$$
\pi_{R1} (\boldsymbol{\psi}) = \lim_{s \to \infty} \frac{\pi_2^k (\lambda_0, \eta \mid \beta) \pi_1^k (\beta)}{\pi_2^k (1, 2 \mid 1) \pi_1^k (1)} = \frac{1}{\lambda_0 \eta \beta}. \tag{A.54}
$$

$\square$

A.3 Preliminaries for properness of priors

Corollary A.1. *Let $\Gamma (z)$ be a gamma function as described in equation 2.108. Then, as $z \to \infty$, Stirling's approximation is:*

$$
\Gamma (z) \sim exp \left\{ -z \right\} z^{z - \frac{1}{2}} (2\pi)^{\frac{1}{2}} \left[1 + \frac{1}{12z} + \frac{1}{288z^2} - \frac{139}{51840z^3} - \frac{571}{2488320z^4} + ... \right],
$$

where the sign $\sim$ indicates that the two quantities are asymptotic (Abramowitz and Stegun, 1964).

Definition A.2. *Let $\bar{\mathbb{R}}$ denote the extended real number line $R \cup \{-\infty, \infty\}$ and let the subscript $\star$ in $\mathbb{R}$ and $\bar{\mathbb{R}}$ denote the exclusion of 0 in these sets. Let $g : U \to \bar{\mathbb{R}}_\star^+$ and $h : U \to \bar{\mathbb{R}}_\star^+$ where $U \subset \mathbb{R}$. We say that $g (x) \propto h (x)$ if there exists $c_0 \in \mathbb{R}_\star^+$ and $c_1 \in \mathbb{R}_\star^+$ such that $c_0 \leq g (x) \leq c_1 h (x)$ for all $x \in U$ (Ramos et al., 2020).*

Let $a \in \bar{R}$, $g : U \to \mathbb{R}^+$ and $h : U \to \mathbb{R}^+$, where $U \subset \mathbb{R}$. We say that $g(x) \underset{x \to a}{\propto} h(x)$ if:

$$\lim_{x \to a} \inf \left(\frac{g(x)}{h(x)} \right) > 0,$$

where $\inf(\cdot)$ is the infimum of the subset and:

$$\lim_{x \to a} \sup \left(\frac{g(x)}{h(x)} \right) < \infty,$$

where $\sup(\cdot)$ is the supremum of the subset (Ramos et al., 2020).

Proposition A.3. *Let $g : (a, b) \to \mathbb{R}^+$ and $h : (a, b) \to \mathbb{R}^+$ be continuous functions on $(a, b) \subset \mathbb{R}$, where $a, b \in \bar{\mathbb{R}}$. Then $g(x) \propto h(x)$ if and only if $g(x) \underset{x \to a}{\propto} h(x)$ and $g(x) \underset{x \to b}{\propto} h(x)$ (Ramos et al., 2020).*

Proposition A.4. *Let $g : (a, b) \to \mathbb{R}^+$ and $h : (a, b) \to \mathbb{R}^+$ be continuous functions on $(a, b) \subset \mathbb{R}$, where $a, b \in \bar{\mathbb{R}}$, and let $c \in (a, b)$. Then if either $g(x) \underset{x \to a}{\propto} h(x)$ or $g(x) \underset{x \to b}{\propto} h(x)$, it follows that:*

$$\int_a^c g(x)\,dx \propto \int_a^c h(x)\,dx,$$

or that:

$$\int_c^b g(x)\,dx \propto \int_c^b h(x)\,dx$$

(Ramos et al., 2020).

Appendix B: Additional results for Chapter 4

B.4 Log-concavity of the conditional posterior under the general reference prior

The conditional posterior of β given λ_0 and η is log-concave if and only if the second-order partial derivative of the log conditional distribution with respect to β is negative. Consider the conditional posterior of $\beta \mid \lambda_0, \eta$:

$$\pi\left(\beta \mid \lambda_0, \eta\right) \propto \beta^{n-m-1} \lambda_0^{n\beta} \eta^{\beta\bar{\delta}} \prod_{i=1}^{k}\prod_{j=1}^{n_i} t_{ij}^{\beta} exp\left\{-\sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta} \lambda_0^{\beta} \eta^{\beta\delta_i}\right\}.$$

The log of the conditional posterior is given by:

$$log\left(\pi\left(\beta \mid \lambda_0, \eta\right)\right) \propto (n-m-1)log\left(\beta\right) + n\beta log\left(\lambda_0\right) + \beta\bar{\delta} log\left(\eta\right) + \beta \sum_{i=1}^{k}\sum_{j=1}^{n_i} log\left(t_{ij}\right) - \sum_{i=1}^{k}\sum_{j=1}^{n_i} t_{ij}^{\beta} \lambda_0^{\beta} \eta^{\beta\delta_i}.$$

Differentiating with respect to β gives:

$$\frac{\partial log\left(\pi\left(\beta \mid \lambda_0, \eta\right)\right)}{\partial\beta} \propto \frac{(n-m-1)}{\beta} + nlog\left(\lambda_0\right) + \bar{\delta} log\left(\eta\right) + \sum_{i=1}^{k}\sum_{j=1}^{n_i} log\left(t_{ij}\right) - \sum_{i=1}^{k}\sum_{j=1}^{n_i} log\left(t_{ij}\lambda_0\eta^{\delta_i}\right) t_{ij}^{\beta} \lambda_0^{\beta} \eta^{\beta\delta_i}.$$

Taking the second derivative with respect to β gives:

$$\frac{\partial^2 log\left(\pi\left(\beta \mid \lambda_0, \eta\right)\right)}{\partial\beta^2} \propto -\frac{(n-m-1)}{\beta^2} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(log\left(t_{ij}\lambda_0\eta^{\delta_i}\right)\right)^2 t_{ij}^{\beta} \lambda_0^{\beta} \eta^{\beta\delta_i},$$

which is negative since: β, η, $\lambda_0 > 0$, $n - m - 1 > 0$ and $t_{ij} > 0 \; \underset{ij}{\forall}$. Therefore the conditional posterior $\pi\left(\beta \mid \lambda_0, \eta\right)$ is log-concave.

B.5 Additional tables for prior π_{R1}

B.5.1 Posterior mean tables

Table B.1: Posterior means for λ_0 $\left(\times 10^{-5}\right)$ under prior π_{R1}.

n	MCMC	Lindley	MLE
30	10.6756	8.3723	8.6676
60	8.1918	7.2867	7.4481
90	7.5083	6.8958	7.0031
120	7.1446	6.8677	6.9508
150	7.0839	6.6953	6.7550

Table B.2: Posterior means for η under prior π_{R1}.

n	MCMC	Lindley	MLE
30	6.1370	5.6241	5.7617
60	5.8053	5.4157	5.5028
90	5.6659	5.3921	5.4583
120	5.5863	5.2949	5.3439
150	5.4762	5.3442	5.3755

Table B.3: Posterior means for β under prior π_{R1}.

n	MCMC	Lindley	MLE
30	1.0454	1.0341	1.0646
60	1.0196	1.0145	1.0318
90	1.0135	1.0090	1.0201
120	1.0098	1.0063	1.0150
150	1.0088	1.0054	1.0124

B.5.2   Root mean squared error tables

Table B.4: RMSE for λ_0 $\left(\times 10^{-5}\right)$ under prior π_{R1}.

n		30	60	90	120	150
$\hat{\lambda}_{0MLE(S)}$		8.5076	4.3328	3.2901	2.6357	2.2828
$\hat{\lambda}_{0MC(S)}$		9.3036	4.8864	3.4549	2.7796	2.4190
$\hat{\lambda}_{0LIN(S)}$		8.2135	4.2123	3.2241	2.5894	2.2542
$\hat{\lambda}_{0MC(L)}$	*a, k = 0.5*	9.5327	4.8863	3.4549	2.7796	2.4190
$\hat{\lambda}_{0MC(G)}$		5.9675	3.7121	2.8433	2.4074	2.1349
$\hat{\lambda}_{0LIN(L)}$		7.3231	4.1075	3.1091	2.5029	2.3260
$\hat{\lambda}_{0LIN(G)}$		8.1081	4.1864	2.8800	2.5173	2.2170
$\hat{\lambda}_{0MC(L)}$	*a, k = 1.5*	9.3741	4.8861	3.4548	2.7795	2.4190
$\hat{\lambda}_{0MC(G)}$		4.4839	3.2786	2.5970	2.6018	2.0157
$\hat{\lambda}_{0LIN(L)}$		7.9294	4.4246	3.2412	2.6048	2.2238
$\hat{\lambda}_{0LIN(G)}$		8.2816	4.0344	3.0684	2.5381	2.2651
$\hat{\lambda}_{0MC(L)}$	*a, k = −0.5*	6.2442	4.8865	3.4549	2.7796	2.4091
$\hat{\lambda}_{0MC(G)}$		8.0577	4.3552	3.2202	2.6362	2.3109
$\hat{\lambda}_{0LIN(L)}$		8.0147	3.8610	3.1882	2.5218	2.2770
$\hat{\lambda}_{0LIN(G)}$		7.5453	4.3588	3.0493	2.6676	2.2044
$\hat{\lambda}_{0MC(L)}$	*a, k = −1.5*	8.0397	4.8867	3.4500	2.7797	2.4091
$\hat{\lambda}_{0MC(G)}$		10.6726	5.3691	3.7190	2.9412	2.5396
$\hat{\lambda}_{0LIN(L)}$		8.3092	4.4564	3.0511	2.6058	2.1922
$\hat{\lambda}_{0LIN(G)}$		7.8492	4.0490	3.2057	2.6373	2.1798

Table B.5: RMSE for η under prior π_{R1}.

n		30	60	90	120	150
$\hat{\eta}_{MLE(S)}$		2.7736	1.8674	1.4545	1.2004	1.0985
$\hat{\eta}_{MC(S)}$		2.5537	1.8279	1.4981	1.2784	1.1082
$\hat{\eta}_{LIN(S)}$		2.7060	1.8361	1.4343	1.1894	1.0916
$\hat{\eta}_{MC(L)}$		1.7952	1.4124	1.2339	1.0961	0.9984
$\hat{\eta}_{MC(G)}$	$a, k = 0.5$	2.2230	1.6544	1.3883	1.2039	1.0608
$\hat{\eta}_{LIN(L)}$		2.7837	1.7633	1.4026	1.2264	1.0507
$\hat{\eta}_{LIN(G)}$		2.6871	1.7176	1.4065	1.2355	1.0997
$\hat{\eta}_{MC(L)}$		2.9901	1.4659	1.2385	1.0871	0.9961
$\hat{\eta}_{MC(G)}$	$a, k = 1.5$	2.1163	1.5918	1.3455	1.1741	1.0434
$\hat{\eta}_{LIN(L)}$		2.8825	1.7888	1.4064	1.1745	1.0994
$\hat{\eta}_{LIN(G)}$		2.5618	1.7116	1.3883	1.2123	1.0746
$\hat{\eta}_{MC(L)}$		5.0748	3.0273	2.1827	1.7039	1.3811
$\hat{\eta}_{MC(G)}$	$a, k = -0.5$	2.4271	1.7599	1.4555	1.2496	1.0896
$\hat{\eta}_{LIN(L)}$		2.8227	1.7394	1.3763	1.1877	1.1054
$\hat{\eta}_{LIN(G)}$		2.8467	1.7606	1.4197	1.2002	1.0567
$\hat{\eta}_{MC(L)}$		9.4864	6.2390	4.4830	3.3692	2.5131
$\hat{\eta}_{MC(G)}$	$a, k = -1.5$	2.6963	1.9055	1.5465	1.3110	1.1297
$\hat{\eta}_{LIN(L)}$		2.6488	1.7587	1.4492	1.2231	1.0604
$\hat{\eta}_{LIN(G)}$		2.7543	1.7911	1.3887	1.2377	1.0534

Table B.6: RMSE for β under priorπ_{R1}.

n		30	60	90	120	150
$\hat{\beta}_{MLE(S)}$		0.1735	0.1110	0.0870	0.0742	0.0657
$\hat{\beta}_{MC(S)}$		0.1645	0.1083	0.0858	0.0714	0.0665
$\hat{\beta}_{LIN(S)}$		0.1627	0.1060	0.0844	0.0723	0.0648
$\hat{\beta}_{MC(L)}$	$a, k = 0.5$	0.1551	0.1072	0.0852	0.0710	0.0662
$\hat{\beta}_{MC(G)}$		0.1581	0.1063	0.0847	0.0704	0.0660
$\hat{\beta}_{LIN(L)}$		0.1607	0.1104	0.0842	0.0737	0.0655
$\hat{\beta}_{LIN(G)}$		0.1728	0.1077	0.0878	0.0729	0.0653
$\hat{\beta}_{MC(L)}$	$a, k = 1.5$	0.1681	0.1054	0.0842	0.0704	0.0657
$\hat{\beta}_{MC(G)}$		0.1547	0.1052	0.0841	0.0703	0.0657
$\hat{\beta}_{LIN(L)}$		0.1586	0.1087	0.0863	0.0757	0.0650
$\hat{\beta}_{LIN(G)}$		0.1752	0.1070	0.0872	0.0734	0.0646
$\hat{\beta}_{MC(L)}$	$a, k = -0.5$	0.1681	0.1094	0.0864	0.0719	0.0668
$\hat{\beta}_{MC(G)}$		0.1622	0.1075	0.0854	0.0711	0.0663
$\hat{\beta}_{LIN(L)}$		0.1739	0.1107	0.0878	0.0725	0.0662
$\hat{\beta}_{LIN(G)}$		0.1777	0.1104	0.0852	0.0758	0.0647
$\hat{\beta}_{MC(L)}$	$a, k = -1.5$	0.1760	0.1118	0.0877	0.0726	0.0667
$\hat{\beta}_{MC(G)}$		0.1670	0.1091	0.0862	0.0716	0.0667
$\hat{\beta}_{LIN(L)}$		0.1769	0.1128	0.0888	0.0760	0.0662
$\hat{\beta}_{LIN(G)}$		0.1738	0.1103	0.0893	0.0746	0.0668

B.5.3 Coverage rate tables

Table B.7: 95% Coverage rate for π_{R1}.

n	30	60	90	120	150
λ_0	0.9490	0.9550	0.9515	0.9495	0.9525
ℓ $(\times 10^{-5})$	(29.5801)	(16.6838)	(12.3201)	(10.0862)	(8.9308)
η	0.9505	0.9510	0.9490	0.9520	0.9485
ℓ	(9.7866)	(7.2169)	(5.8028)	(4.9439)	(4.3306)
β	0.9485	0.9460	0.9510	0.9515	0.9465
ℓ	(0.5969)	(0.4073)	(0.3290)	(0.2826)	(0.2522)

Table B.8: 90% Coverage rate for π_{R1}.

n	30	60	90	120	150
λ_0	0.8960	0.9020	0.9030	0.9020	0.9035
ℓ $(\times 10^{-5})$	(23.2169)	(13.3509)	(10.0207)	(8.2854)	(7.3640)
η	0.90250	0.9015	0.8970	0.9065	0.9050
ℓ	(8.1356)	(5.9596)	(4.7938)	(4.0992)	(3.5988)
β	0.8960	0.8990	0.9125	0.9065	0.8960
ℓ	(0.5024)	(0.3424)	(0.2763)	(0.2377)	(0.2123)

B.6   Additional tables for prior π_{R2}

B.6.1   Posterior mean tables

Table B.9: Posterior means for λ_0 $\left(\times 10^{-5}\right)$ under prior π_{R2}.

n	MCMC	Lindley	MLE
30	10.5412	8.3831	8.6676
60	8.3295	7.2888	7.4481
90	7.5861	6.8967	7.0031
120	7.2413	6.8682	6.9508
150	6.9776	6.6956	6.7550

Table B.10: Posterior means for η under prior π_{R2}.

n	MCMC	Lindley	MLE
30	6.2025	5.6285	5.7617
60	5.8252	5.4167	5.5028
90	5.6240	5.3923	5.4583
120	5.5638	5.2951	5.3439
150	5.5452	5.3443	5.3755

Table B.11: Posterior means for β under prior π_{R2}.

n	MCMC	Lindley	MLE
30	1.0230	1.0183	1.0646
60	1.0116	1.0057	1.0318
90	1.0085	1.0024	1.0201
120	1.0043	1.0020	1.0150
150	1.0028	1.0024	1.0124

B.6.2 Root mean squared error tables

Table B.12: RMSE for λ_0 $\left(\times 10^{-5}\right)$ under prior π_{R2}.

n		30	60	90	120	150
$\hat{\lambda}_{0MLE(S)}$		8.5076	4.3328	3.2901	2.6357	2.2828
$\hat{\lambda}_{0MC(S)}$		8.6230	5.0849	3.4657	2.9105	2.3819
$\hat{\lambda}_{0LIN(S)}$		8.2346	4.2140	3.2247	2.5898	2.2543
$\hat{\lambda}_{0MC(L)}$	$a, k = 0.5$	8.6226	5.0847	3.4656	2.9104	2.2960
$\hat{\lambda}_{0MC(G)}$		5.3747	3.8072	2.8305	2.5115	2.1150
$\hat{\lambda}_{0LIN(L)}$		8.6606	4.3911	3.2997	2.5458	2.2601
$\hat{\lambda}_{0LIN(G)}$		6.7768	4.1456	2.9887	2.6748	2.2155
$\hat{\lambda}_{0MC(L)}$	$a, k = 1.5$	8.6219	5.0846	3.4656	2.9104	2.2961
$\hat{\lambda}_{0MC(G)}$		4.0603	3.2770	2.5729	2.3430	2.0103
$\hat{\lambda}_{0LIN(L)}$		7.9261	3.8581	3.2622	2.5958	2.1608
$\hat{\lambda}_{0LIN(G)}$		8.3617	4.2752	3.0440	2.6019	2.2335
$\hat{\lambda}_{0MC(L)}$	$a, k = -0.5$	8.6234	5.0850	3.4567	2.9105	2.2965
$\hat{\lambda}_{0MC(G)}$		7.3879	4.5978	3.2232	2.7589	2.2789
$\hat{\lambda}_{0LIN(L)}$		8.5946	3.9175	3.1306	2.5899	2.2655
$\hat{\lambda}_{0LIN(G)}$		7.3903	3.9842	3.2742	2.5171	2.2279
$\hat{\lambda}_{0MC(L)}$	$a, k = -1.5$	8.6241	5.0851	3.4658	2.9106	2.2966
$\hat{\lambda}_{0MC(G)}$		9.9968	5.6302	3.5391	3.0797	2.4983
$\hat{\lambda}_{0LIN(L)}$		7.2914	4.5670	3.1459	2.5136	2.1443
$\hat{\lambda}_{0LIN(G)}$		7.8841	4.0976	3.0197	2.5594	2.2457

Table B.13: RMSE for η under prior π_{R2}.

n		30	60	90	120	150
$\hat{\eta}_{MLE(S)}$		2.7736	1.8674	1.4545	1.2004	1.0985
$\hat{\eta}_{MC(S)}$		2.5800	1.9122	1.4722	1.2870	1.1313
$\hat{\eta}_{LIN(S)}$		2.7088	1.8366	1.4344	1.1895	1.0916
$\hat{\eta}_{MC(L)}$	$a, k = 0.5$	1.7860	1.4844	1.2330	1.1055	0.9920
$\hat{\eta}_{MC(G)}$		2.2398	1.7360	1.3708	1.2126	1.0733
$\hat{\eta}_{LIN(L)}$		2.6483	1.7864	1.4336	1.2051	1.0730
$\hat{\eta}_{LIN(G)}$		2.6555	1.7899	1.4215	1.2443	1.0535
$\hat{\eta}_{MC(L)}$	$a, k = 1.5$	1.9384	1.5161	1.2504	1.0249	0.9785
$\hat{\eta}_{MC(G)}$		2.1249	1.6713	1.3345	1.1833	1.0492
$\hat{\eta}_{LIN(L)}$		2.7636	1.8371	1.4366	1.2294	1.0960
$\hat{\eta}_{LIN(G)}$		2.6378	1.8110	1.4172	1.2644	1.0643
$\hat{\eta}_{MC(L)}$	$a, k = -0.5$	5.2631	3.1508	2.1424	1.7243	1.4326
$\hat{\eta}_{MC(G)}$		2.4456	1.8433	1.4323	1.2582	1.1091
$\hat{\eta}_{LIN(L)}$		2.5479	1.7613	1.3804	1.1969	1.1023
$\hat{\eta}_{LIN(G)}$		2.7818	1.7562	1.4429	1.2104	1.0692
$\hat{\eta}_{MC(L)}$	$a, k = -1.5$	9.7880	6.3622	4.4885	3.3887	2.5691
$\hat{\eta}_{MC(G)}$		2.5796	1.9907	1.5182	1.3198	1.1565
$\hat{\eta}_{LIN(L)}$		2.5406	1.7263	1.4358	1.2094	1.1025
$\hat{\eta}_{LIN(G)}$		2.7615	1.7562	1.4392	1.2681	1.0327

Table B.14: RMSE for β under prior π_{R2}.

n		30	60	90	120	150
$\hat{\beta}_{MLE(S)}$		0.1735	0.1110	0.0870	0.0742	0.0656
$\hat{\beta}_{MC(S)}$		0.1570	0.1082	0.0837	0.0723	0.0665
$\hat{\beta}_{LIN(S)}$		0.1603	0.1050	0.0835	0.0721	0.0643
$\hat{\beta}_{MC(L)}$	$a, k = 0.5$	0.1542	0.1074	0.0832	0.0721	0.0663
$\hat{\beta}_{MC(G)}$		0.1525	0.1068	0.0829	0.0719	0.0663
$\hat{\beta}_{LIN(L)}$		0.1612	0.1043	0.0826	0.0755	0.0638
$\hat{\beta}_{LIN(G)}$		0.1641	0.1058	0.0862	0.0727	0.0662
$\hat{\beta}_{MC(L)}$	$a, k = 1.5$	0.1506	0.1059	0.0825	0.0717	0.0661
$\hat{\beta}_{MC(G)}$		0.1497	0.1062	0.0824	0.0716	0.0663
$\hat{\beta}_{LIN(L)}$		0.1726	0.1069	0.0848	0.0740	0.0652
$\hat{\beta}_{LIN(G)}$		0.1647	0.1071	0.0873	0.0728	0.0662
$\hat{\beta}_{MC(L)}$	$a, k = -0.5$	0.1599	0.1091	0.0842	0.0726	0.0666
$\hat{\beta}_{MC(G)}$		0.1553	0.1077	0.0834	0.0722	0.0664
$\hat{\beta}_{LIN(L)}$		0.1766	0.1089	0.0866	0.0747	0.0655
$\hat{\beta}_{LIN(G)}$		0.1703	0.1085	0.0874	0.0737	0.0671
$\hat{\beta}_{MC(L)}$	$a, k = -1.5$	0.1664	0.1112	0.0853	0.0732	0.0670
$\hat{\beta}_{MC(G)}$		0.1588	0.1085	0.0840	0.0725	0.0666
$\hat{\beta}_{LIN(L)}$		0.1773	0.1153	0.0861	0.0759	0.0660
$\hat{\beta}_{LIN(G)}$		0.1722	0.1090	0.0869	0.7340	0.0653

B.6.3 Coverage rate tables

Table B.15: 95% Coverage rate for π_{R2}.

n	30	60	90	120	150
λ_0 ℓ ($\times 10^{-5}$)	0.9520 (29.5406)	0.9410 (17.1340)	0.9580 (12.5445)	0.9465 (10.2768)	0.9480 (8.8650)
η ℓ	0.9455 (10.0555)	0.9415 (7.2939)	0.9495 (5.7954)	0.9470 (4.9515)	0.9555 (4.4177)
β ℓ	0.9505 (0.5949)	0.9460 (0.4072)	0.9540 (0.3284)	0.9455 (0.2821)	0.9425 (0.2512)

Table B.16: 90% Coverage rate for π_{R2}.

n	30	60	90	120	150
λ_0	0.9020	0.8935	0.9000	0.8890	0.9060
ℓ ($\times 10^{-5}$)	(23.1381)	(13.7289)	(10.2026)	(8.4199)	(7.3083)
η	0.8995	0.8895	0.9100	0.8975	0.8970
ℓ	(8.3685)	(6.0240)	(4.7884)	(4.1041)	(3.6725)
β	0.9065	0.8950	0.9075	0.8940	0.8870
ℓ	(0.5001)	(0.3423)	(0.2761)	(0.2370)	(0.2112)

B.7 Additional tables for MLEs

B.7.1 Coverage rate tables

Table B.17: 95% Coverage rate for MLEs.

n	30	60	90	120	150
λ_0	0.9490	0.9445	0.9405	0.9385	0.9450
ℓ ($\times 10^{-5}$)	(32.0809)	(17.1885)	(12.6323)	(10.0496)	(8.7791)
η	0.9435	0.9505	0.9515	0.9570	0.9485
ℓ	(10.7205)	(7.2478)	(5.6681)	(4.7031)	(4.2959)
β	0.9350	0.9370	0.9460	0.9500	0.9410
ℓ	(0.6312)	(0.4227)	(0.3314)	(0.2848)	(0.2529)

Table B.18: 90% Coverage rate for MLEs.

n	30	60	90	120	150
λ_0	0.9350	0.9265	0.9195	0.9105	0.9095
ℓ ($\times 10^{-5}$)	(27.0069)	(14.4699)	(10.6344)	(8.4601)	(7.4074)
η	0.9170	0.9195	0.9130	0.9105	0.9125
ℓ	(9.0249)	(6.10153)	(4.7716)	(3.9592)	(3.6164)
β	0.8920	0.8915	0.8895	0.8985	0.9000
ℓ	(0.5314)	(0.3559)	(0.2790)	(0.2397)	(0.2129)

B.8 BUGS script for prior π_J

model{

for(i in 1:N){

#Define likelihood

```
#Complete data
X[i] ~ dweib(z, l[i])
#Define scale parameter
#For Arrhenius law
l[i] <- pow(lam, z) * pow(eta, z *((1/l0 - 1/S[i])/ (1/l0 - 1/l1)))
#For Power law
l[i] <- pow(lam, z) * pow(eta, z *(log(lO)/20 - log(S[i])/20)/ (log(lO)/20 - log(l1)/20))
}
#Define priors
zero <- 0
z ~ dunif(0,100)
phi <- -log(z)
zero ~ dpois(phi)
lam ~dgamma(0.001,0.001)
eta ~ dgamma(0.001,0.001)I(1,)
}
#INITS list()
#DATA list()
```

B.9   BUGS script for prior π_{R1}

```
model{
for(i in 1:N){
#Define likelihood
#Complete data
X[i] ~ dweib(z, l[i])
#Define scale parameter
#For Arrhenius law
l[i] <- pow(lam, z) * pow(eta, z *((1/l0 - 1/S[i])/ (1/l0 - 1/l1)))
#For Power law
l[i] <- pow(lam, z) * pow(eta, z *(log(lO)/20 - log(S[i])/20)/ (log(lO)/20 - log(l1)/20))
}
#Define priors
z ~ dunif(0,100)
lam ~dgamma(0.001,0.001)
```

eta ~ dgamma(0.001,0.001)I(1,)

}

#INITS list()

#DATA list()

B.10   BUGS script for prior π_{R2}

model{

for(i in 1:N){

#Define likelihood

#Complete data

X[i] ~ dweib(z, l[i])

#Define scale parameter

#For Arrhenius law

l[i] <- pow(lam, z) * pow(eta, z *((1/l0 - 1/S[i])/ (1/l0 - 1/l1)))

#For Power law

l[i] <- pow(lam, z) * pow(eta, z *(log(lO)/20 - log(S[i])/20)/ (log(lO)/20 - log(l1)/20))

}

#Define priors

z ~ dgamma(0.001,0.001)

lam ~dgamma(0.001,0.001)

eta ~ dgamma(0.001,0.001)I(1,)

}

#INITS list()

#DATA list()

B.11   R script for finding MLE values

```
#Load packages
library(maxLik)
#Define functions
f1 = function(beta,eta){

    A = sum(X1^beta * eta^(beta * Stan_d[1]))
    B = sum(X2^beta * eta^(beta * Stan_d[2]))
    C = sum(X3^beta * eta^(beta * Stan_d[3]))
    A + B + C
```

```r
    }

f2 = function(beta,eta){

    A = sum(X1^beta * eta^(beta * Stan_d[1]) * (Stan_d[1]*log(eta
        ) + log(X1)))
    B = sum(X2^beta * eta^(beta * Stan_d[2]) * (Stan_d[2]*log(eta
        ) + log(X2)))
    C = sum(X3^beta * eta^(beta * Stan_d[3]) * (Stan_d[3]*log(eta
        ) + log(X3)))
    A + B + C

    }

f3 = function(beta,eta){

    A = sum(X1^beta * eta^(beta * Stan_d[1] -1) * beta*Stan_d[1])
    B = sum(X2^beta * eta^(beta * Stan_d[2] -1) * beta*Stan_d[2])
    C = sum(X3^beta * eta^(beta * Stan_d[3] -1) * beta*Stan_d[3])
    A + B + C

    }

#Define likelihood
ll = function(theta){
    beta = theta[1]
    eta = theta[2]

    ll = n*log(beta) + n*log(n/f1(beta,eta)) + D_bar*beta*log(eta
        ) + (beta-1)*sum(log(X)) - n
    return(ll)

    }

#Define gradient
grad = function(theta){
    beta = theta[1]
    eta = theta[2]

    grad = c(n/beta + D_bar*log(eta) + sum(log(X)) - n *(f2(beta,
        eta)/f1(beta,eta)),
                D_bar*beta/eta - n*(f3(beta,eta)/f1(beta,eta))
    )

    return(grad)
    }
```

```r
#Empty vectors
l = c()
e = c()
b = c()


#Loop 2000 times
for(i in 1:2000){


#Sample size
n = ##


#Stress vector
S = c(rep(300,n/3), rep(250,n/3), rep(200,n/3))


#Data
X = c(rweibull(n/3,shape = 1, scale = (6.3361e-5*5.2945^2)^-1),
        rweibull(n/3,shape = 1, scale = (6.3361e-5*5.2945^1.6)
            ^-1),
        rweibull(n/3,shape = 1, scale = (6.3361e-5*5.2945)^-1))



#Subset data
X1 = X[#1st item stress level 3:#last item stress level 3]
X2 = X[#1st item stress level 2:#last item stress level 2]
X3 = X[#1st item stress level 1:#last item stress level 1]


US = unique(S)


#Find delta[i]
d = function(q){
    ((1/150)-(1/q))/((1/150)-(1/200))
  }


Stan_d = sapply(US,d)


Lus = c(n/3,n/3,n/3)


#Find delta bar
D_bar = Lus * Stan_d
D_bar = sum(D_bar)


#Obtain MLE for beta and eta
mle = maxLik(ll,grad, start = c(1,5.2945), method = "NR")
beta = mle$estimate[1]
eta = mle$estimate[2]


#Obtain MLE for lam
lam = (n/f1(mle$estimate[1],mle$estimate[2]))^(1/mle$estimate[1])
```

```
#Store results in vector
l[i] = lam
b[i] = beta
e[i] = eta


}
```

Listing B.1: MLE values

B.12 R script for finding coverage rates

```
#Load packages

library(R2WinBUGS)

#Sample size
n = ##


# Define stress vector

S = c(rep(300,n/3), rep(250,n/3), rep(200,n/3))


# Define initial values

inits = function(){

list(z = 1, eta = 1, lam = 0.00001)

}


#Empty vectors

z = c()

lam = c()

eta = c()


#For loop

for(i in 1:2000){
```

```r
#Gen random values

X = c(rweibull(n/3,shape = 1, scale = (6.3361e-5*5.2945^2)^-1),

rweibull(n/3,shape = 1, scale = (6.3361e-5*5.2945^1.6)^-1),

rweibull(n/3,shape = 1, scale = (6.3361e-5*5.2945)^-1))


#Store in data frame

data = data.frame(X,S)


#Call WinBUGS

sim = bugs(data,inits = inits, model.file = ,

parameters.to.save = c('z','lam','eta'), n.chains = 1,

n.iter = 250000, n.burnin = 150000, n.thin = 2,

bugs.directory = )


#Store values

z[i] = sim$summary[1,1]

z2.5[i] = as.numeric(quantile(sort(sim$sims.list$z),0.025))

z97.5[i] = as.numeric(quantile(sort(sim$sims.list$z),0.975))

z5[i] = as.numeric(quantile(sort(sim$sims.list$z),0.05))

z95[i] = as.numeric(quantile(sort(sim$sims.list$z),0.95))


lam[i] = sim$summary[2,1]

lam2.5[i] = as.numeric(quantile(sort(sim$sims.list$lam),0.025))

lam97.5[i] = as.numeric(quantile(sort(sim$sims.list$lam),0.975))

lam5[i] = as.numeric(quantile(sort(sim$sims.list$lam),0.05))

lam95[i] = as.numeric(quantile(sort(sim$sims.list$lam),0.95))
```

```r
eta[i] = sim$summary[3,1]

eta2.5[i] = as.numeric(quantile(sort(sim$sims.list$eta),0.025))

eta97.5[i] = as.numeric(quantile(sort(sim$sims.list$eta),0.975))

eta5[i] = as.numeric(quantile(sort(sim$sims.list$eta),0.05))

eta95[i] = as.numeric(quantile(sort(sim$sims.list$eta),0.95))

}

#Calculate relevant values

z_true = 1
lam_true = 6.3361e-5
eta_true = 5.2954

#Beta
sum(z2.5 < z_true & z97.5 > z_true)/2000
mean(z97.5 - z2.5)
sum(z5 < z_true & z95 > z_true)/2000
mean(z95 - z5)

#Lam
sum(lam2.5 < lam_true & lam97.5 > lam_true)/2000
mean(lam97.5 - lam2.5)
sum(lam5 < lam_true & lam95 > lam_true)/2000
mean(lam95 - lam5)

#Eta
sum(eta2.5 < eta_true & eta97.5 > eta_true)/2000
mean(eta97.5 - eta2.5)
sum(eta5 < eta_true & eta95 > eta_true)/2000
mean(eta95 - eta5)
```

Listing B.2: Coverage rate

B.13   R script for Lindley's approximation

```r
#Define answer vector
fff = c()

#Loop
```

```r
  for(i in 1:2000){

##Load MLE values############
beta = b[i]
lam = l[i]
eta = e[i]

    ##Define Functions####
    p = # log prior
    p1 = #derivative of p wrt lam
    p2 = #derivative of p wrt eta
    p3 = #derivative of p wrt beta
    w =   #Function of interest
    w1 =  #derivative of w wrt lam
    w2 =  #derivative of w wrt eta
    w3 =#derivative of w wrt beta

    w12 = #derivative of lam and eta
    w13 = #derivative of lam and beta
    w23 =  #derivative of eta and beta
    w11 =  #derivative of lam and lam
    w22 = #derivative of eta and eta
    w33 = #derivative of beta and beta

X = c(rweibull(n/3,shape = 1, scale = (6.3361e-5*5.2945^2)^-1),
      rweibull(n/3,shape = 1, scale = (6.3361e-5*5.2945^1.6)^-1),
      rweibull(n/3,shape = 1, scale = (6.3361e-5*5.2945)^-1))

n = length(S)

X1 = X[#1st item stress level 3:#last item stress level 3]
X2 = X[#1st item stress level 2:#last item stress level 2]
X3 = X[#1st item stress level 1:#last item stress level 1]


US = unique(S)

d = function(q){
    ((1/150)-(1/q))/((1/150)-(1/200))
  }

Stan_d = sapply(US,d)

Lus = c(n/3,n/3,n/3)

D_bar = Lus * Stan_d
D_bar = sum(D_bar)
```

```r
    ##Functions for Hessian####
f1 = function(beta,eta,lam){
    A = sum(beta*((beta-1)*eta^(beta*Stan_d[1])*X1^beta*lam^
        beta+n))
    B = sum(beta*((beta-1)*eta^(beta*Stan_d[2])*X2^beta*lam^
        beta+n))
    C = sum(beta*((beta-1)*eta^(beta*Stan_d[3])*X3^beta*lam^
        beta+n))
    A + B + C
  }


f2 = function(beta,eta,lam){
    A = sum(beta^2*Stan_d[1]*lam^(beta-1)*X1^beta*eta^(beta*
        Stan_d[1]-1))
    B = sum(beta^2*Stan_d[2]*lam^(beta-1)*X2^beta*eta^(beta*
        Stan_d[2]-1))
    C = sum(beta^2*Stan_d[3]*lam^(beta-1)*X3^beta*eta^(beta*
        Stan_d[3]-1))
    A + B + C
  }


f3 = function(beta,eta,lam){
    A = sum(eta^(beta*Stan_d[1])*lam^beta*X1^beta*((log(X1*lam*
        eta^Stan_d[1]))*beta+1)-n)
    B = sum(eta^(beta*Stan_d[2])*lam^beta*X2^beta*((log(X2*lam*
        eta^Stan_d[2]))*beta+1)-n)
    C = sum(eta^(beta*Stan_d[3])*lam^beta*X3^beta*((log(X3*lam*
        eta^Stan_d[3]))*beta+1)-n)
    A + B + C
  }


f4 = function(beta,eta,lam){
    A = sum(beta*((beta*Stan_d[1]^2-Stan_d[1])*lam^beta*X1^beta
        *eta^(beta*Stan_d[1])+D_bar))
    B = sum(beta*((beta*Stan_d[2]^2-Stan_d[2])*lam^beta*X2^beta
        *eta^(beta*Stan_d[2])+D_bar))
    C = sum(beta*((beta*Stan_d[3]^2-Stan_d[3])*lam^beta*X3^beta
        *eta^(beta*Stan_d[3])+D_bar))
    A + B + C
  }


f5 = function(beta,eta,lam){
    A = sum(Stan_d[1]*eta^(Stan_d[1]*beta)*lam^beta*X1^beta*((
        log(X1*lam*eta^Stan_d[1]))*beta+1)-D_bar)
```

```r
        B = sum(Stan_d[2]*eta^(Stan_d[2]*beta)*lam^beta*X2^beta*((
            log(X2*lam*eta^Stan_d[2]))*beta+1)-D_bar)
        C = sum(Stan_d[3]*eta^(Stan_d[3]*beta)*lam^beta*X3^beta*((
            log(X3*lam*eta^Stan_d[3]))*beta+1)-D_bar)
        A + B + C
    }

f6 = function(beta,eta,lam){
        A = sum(eta^(beta*Stan_d[1])*lam^beta*X1^beta*(log(X1*lam*
            eta^Stan_d[1])^2))
        B = sum(eta^(beta*Stan_d[2])*lam^beta*X2^beta*(log(X2*lam*
            eta^Stan_d[2])^2))
        C = sum(eta^(beta*Stan_d[3])*lam^beta*X3^beta*(log(X3*lam*
            eta^Stan_d[3])^2))
        A + B + C

    }




h11 = (-1/lam^2*f1(beta,eta,lam))

h12 = (-f2(beta,eta,lam))

h13 = (-1/lam * f3(beta,eta,lam))

h21 = h12

h22 = (-1/eta^2 *f4(beta,eta,lam))

h23 = (-1/eta * f5(beta,eta,lam))

h31 = h13

h32 = h23

h33 = (-n/beta^2 -f6(beta,eta,lam))

m = matrix(c(h11,h12,h13,h21,h22,h23,h31,h32,h33), nrow = 3)
m = solve(m)

h11 = -m[1,1]

h12 = -m[1,2]

h13 = -m[1,3]
```

```r
h21 = -m[2,1]

h22 = -m[2,2]

h23 = -m[2,3]

h31 = -m[3,1]

h32 = -m[3,2]

h33 = -m[3,3]




g1 = function(beta,eta,lam){
    A = sum(X1^beta * beta * (beta-1) * (beta - 2) * lam^(beta
        -3) * eta^(beta*Stan_d[1]))
    B = sum(X2^beta * beta * (beta-1) * (beta - 2) * lam^(beta
        -3) * eta^(beta*Stan_d[2]))
    C = sum(X3^beta * beta * (beta-1) * (beta - 2) * lam^(beta
        -3) * eta^(beta*Stan_d[3]))
    A + B + C
}

g2 = function(beta,eta,lam){
    A = sum(X1^beta * lam^beta * beta * Stan_d[1] * (beta*Stan_
        d[1]-1) * (beta*Stan_d[1]-2) * eta^(beta*Stan_d[1]-3))
    B = sum(X2^beta * lam^beta * beta * Stan_d[2] * (beta*Stan_
        d[2]-1) * (beta*Stan_d[2]-2) * eta^(beta*Stan_d[2]-3))
    C = sum(X3^beta * lam^beta * beta * Stan_d[3] * (beta*Stan_
        d[3]-1) * (beta*Stan_d[3]-2) * eta^(beta*Stan_d[3]-3))
    A + B + C
}

g3 = function(beta,eta,lam){
    A = sum(X1^beta * lam^beta * eta^(beta*Stan_d[1]) * (log(X1
        *lam*eta^Stan_d[1]))^3)
    B = sum(X2^beta * lam^beta * eta^(beta*Stan_d[2]) * (log(X2
        *lam*eta^Stan_d[2]))^3)
    C = sum(X3^beta * lam^beta * eta^(beta*Stan_d[3]) * (log(X3
        *lam*eta^Stan_d[3]))^3)
    A + B + C
}

g4 = function(beta,eta,lam){
    A = sum(X1 * beta^2 * Stan_d[1] * (beta - 1) * lam^(beta-2)
        * eta^(beta*Stan_d[1]-1))
```

```
    B = sum(X2 * beta^2 * Stan_d[2] * (beta - 1) * lam^(beta-2)
        * eta^(beta*Stan_d[2]-1))
    C = sum(X3 * beta^2 * Stan_d[3] * (beta - 1) * lam^(beta-2)
        * eta^(beta*Stan_d[3]-1))
    A + B + C
}


g5 = function(beta,eta,lam){
    A = sum(X1^beta * lam^(beta - 2) * eta^(beta*Stan_d[1]) *
        (((log(X1*lam*eta^Stan_d[1]))*(beta-1)+2)*beta-1))
    B = sum(X2^beta * lam^(beta - 2) * eta^(beta*Stan_d[2]) *
        (((log(X2*lam*eta^Stan_d[2]))*(beta-1)+2)*beta-1))
    C = sum(X3^beta * lam^(beta - 2) * eta^(beta*Stan_d[3]) *
        (((log(X3*lam*eta^Stan_d[3]))*(beta-1)+2)*beta-1))
    A + B + C
}


g6 = function(beta,eta,lam){
    A = sum(X1^beta * lam^(beta-1)* beta^2 * Stan_d[1] * (beta*
        Stan_d[1]-1) * eta^(beta*Stan_d[1]-2))
    B = sum(X2^beta * lam^(beta-1)* beta^2 * Stan_d[2] * (beta*
        Stan_d[2]-1) * eta^(beta*Stan_d[2]-2))
    C = sum(X3^beta * lam^(beta-1)* beta^2 * Stan_d[3] * (beta*
        Stan_d[3]-1) * eta^(beta*Stan_d[3]-2))
    A + B + C
}


g7 = function(beta,eta,lam){

    A = sum(X1^beta * lam^beta * Stan_d[1] * eta^(beta*Stan_d
        [1])* (beta*(Stan_d[1]*(log(eta)*(Stan_d[1]*beta-1)+2)+
        log(X1)*(Stan_d[1]*beta-1)+log(lam)*(Stan_d[1]*beta-1))
        -1))
    B = sum(X2^beta * lam^beta * Stan_d[2] * eta^(beta*Stan_d
        [2])* (beta*(Stan_d[2]*(log(eta)*(Stan_d[2]*beta-1)+2)+
        log(X2)*(Stan_d[2]*beta-1)+log(lam)*(Stan_d[2]*beta-1))
        -1))
    C = sum(X3^beta * lam^beta * Stan_d[3] * eta^(beta*Stan_d
        [3])* (beta*(Stan_d[3]*(log(eta)*(Stan_d[3]*beta-1)+2)+
        log(X3)*(Stan_d[3]*beta-1)+log(lam)*(Stan_d[3]*beta-1))
        -1))
    A + B + C



}


g8 = function(beta,eta,lam){
```

```r
    A = sum(X1^beta * lam ^ beta * eta^(beta*Stan_d[1]) * log(
        X1*lam*eta^Stan_d[1])* (beta*log(X1*lam*eta^Stan_d[1]))
        +2)
    B = sum(X2^beta * lam ^ beta * eta^(beta*Stan_d[2]) * log(
        X2*lam*eta^Stan_d[2])* (beta*log(X2*lam*eta^Stan_d[2]))
        +2)
    C = sum(X3^beta * lam ^ beta * eta^(beta*Stan_d[3]) * log(
        X3*lam*eta^Stan_d[3])* (beta*log(X3*lam*eta^Stan_d[3]))
        +2)
    A + B + C
  }

g9 = function(beta,eta,lam){
    A = sum(X1^beta * Stan_d[1] * lam^beta * eta^(beta*Stan_d
        [1]) * log(X1*lam*eta^Stan_d[1])*(beta*log(X1*lam*eta^
        Stan_d[1])+2))
    B = sum(X2^beta * Stan_d[2] * lam^beta * eta^(beta*Stan_d
        [2]) * log(X2*lam*eta^Stan_d[2])*(beta*log(X2*lam*eta^
        Stan_d[2])+2))
    C = sum(X3^beta * Stan_d[3] * lam^beta * eta^(beta*Stan_d
        [3]) * log(X3*lam*eta^Stan_d[3])*(beta*log(X3*lam*eta^
        Stan_d[3])+2))
    A + B + C
  }

g10 = function(beta,eta,lam){
    A = sum(X1^beta * beta * Stan_d[1] * lam^beta * eta^(beta*
        Stan_d[1]) * (beta*log(X1*lam*eta^Stan_d[1])+2))
    B = sum(X2^beta * beta * Stan_d[2] * lam^beta * eta^(beta*
        Stan_d[2]) * (beta*log(X2*lam*eta^Stan_d[2])+2))
    C = sum(X3^beta * beta * Stan_d[3] * lam^beta * eta^(beta*
        Stan_d[3]) * (beta*log(X3*lam*eta^Stan_d[3])+2))
    A + B + C
  }




    ##Third-order derivatives#####
l300 = 2*n*beta/lam^3 - g1(beta,eta,lam)

l030 = 2*D_bar*beta/eta^3 - g2(beta,eta,lam)

l003 = 2*n/beta^3 - g3(beta,eta,lam)
```

```
l210 =  - g4(beta,eta,lam)

l201 = -n/lam^2 - g5(beta,eta,lam)

l120 = - g6(beta,eta,lam)

l021 = -D_bar/eta^2 -1/eta^2 * g7(beta,eta,lam)

l102 = -1/lam * g8(beta,eta,lam)

l012 = -1/eta * g9(beta,eta,lam)

l111 = - (lam*eta)^(-1) * g10(beta,eta,lam)



    ##Define U(psi)##

##Define Aijk####
A123 = w1*h11 + w2*h21 + w3*h31

A213 = w2*h22 + w1*h12 + w3*h32

A321 = w3*h33 + w2*h23 + w1*h13

##Define Bijk####
B123 = h11 *(w1*h11 + w2*h12 + w3*h13)

B213 = h22 *(w2*h22 + w1*h21 + w3*h23)

B321 = h33 *(w3*h33 + w2*h32 + w1*h31)

##Define Cijk####
C123 = w1 * (h11*h23 + 2*h12*h13)

C213 = w2 * (h22*h13 + 2*h21*h23)

C312 = w3 * (h33*h12 + 2*h31*h32)

##Define Dijk####
D123 = 3*w1*h11*h12 + w2*(h11*h22 + 2*h12^2) + w3*(h11*h23 + 2*
    h12*h13)

D132 = 3*w1*h11*h13 + w3*(h11*h33 + 2*h13^2) + w2*(h11*h32 + 2*
    h13*h12)

D213 = 3*w2*h22*h21 + w1*(h22*h11 + 2*h21^2) + w3*(h22*h13 + 2*
    h21*h23)
```

```
D312 = 3*w3*h33*h31 + w1*(h33*h11 + 2*h31^2) + w2*(h33*h12 + 2*
    h31*h32)


D231 = 3*w2*h22*h23 + w3*(h22*h33 + 2*h23^2) + w1*(h22*h31 + 2*
    h23*h21)


D321 = 3*w3*h33*h32 + w2*(h33*h22 + 2*h32^2) + w1*(h33*h21 + 2*
    h32*h31)


##Def final sum####
u_theta = w12*h12 + w13*h13 + w23*h23 + 0.5*(w11*h11 + w22*h22 +
    w33*h33)


fff[i] = w + u_theta + p1*A123 + p2*A213 + p3*A321 + 0.5*(l300*
    B123 + l030*B213 + l003*B321 + 2*l111*(C123 + C213 + C312) +
    l210*D123 + l201*D132 + l120*D213 + l102*D312 + l021*D231 +
    l012*D321)


}
```

Listing B.3: Lindley's approximation

Appendix C: Additional results for Chapter 5

C.14   Additional trace plots
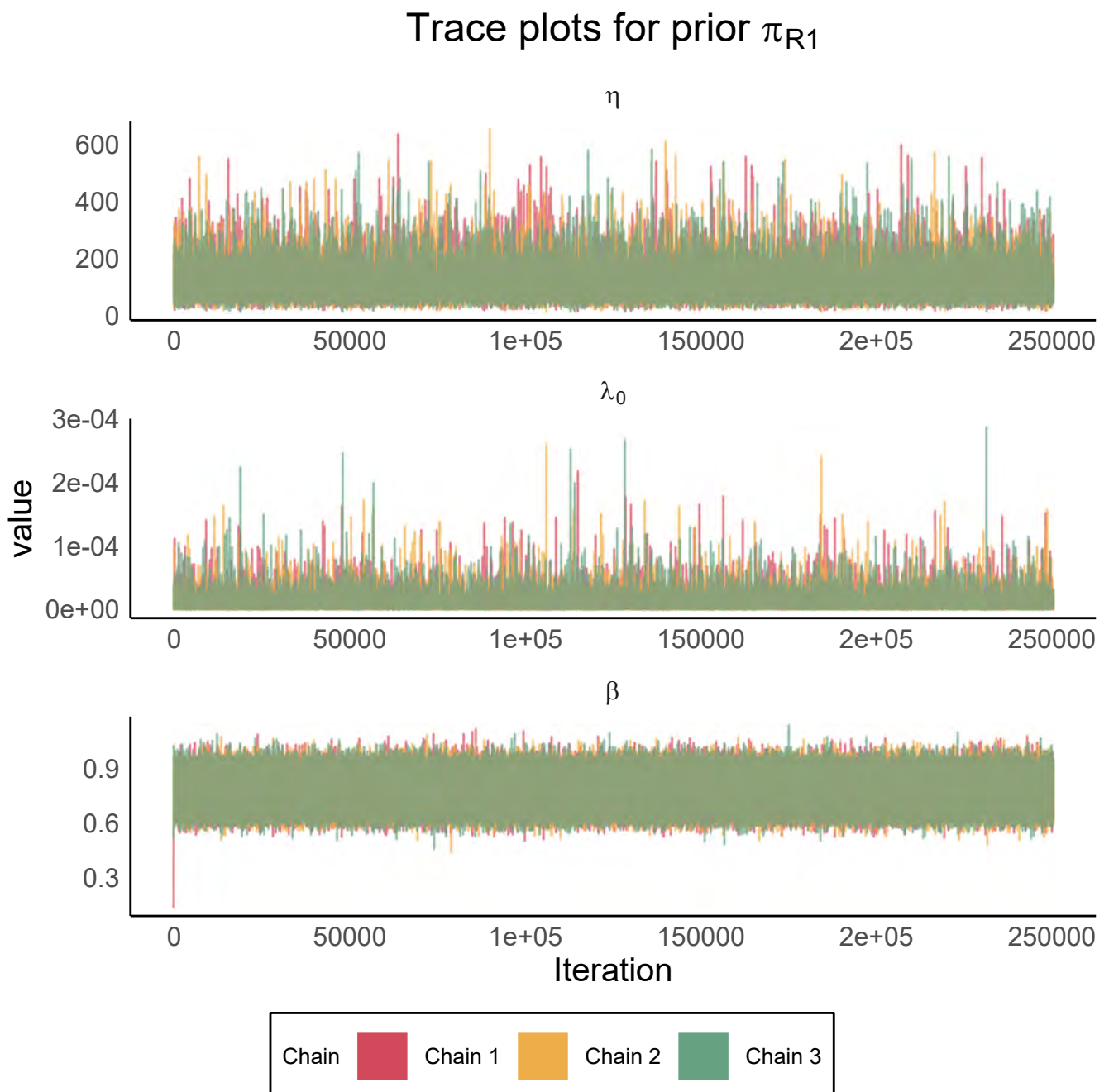
Figure C.1: Trace plots for the chains under prior π_{R1}.

Trace plots for prior π_{R2}
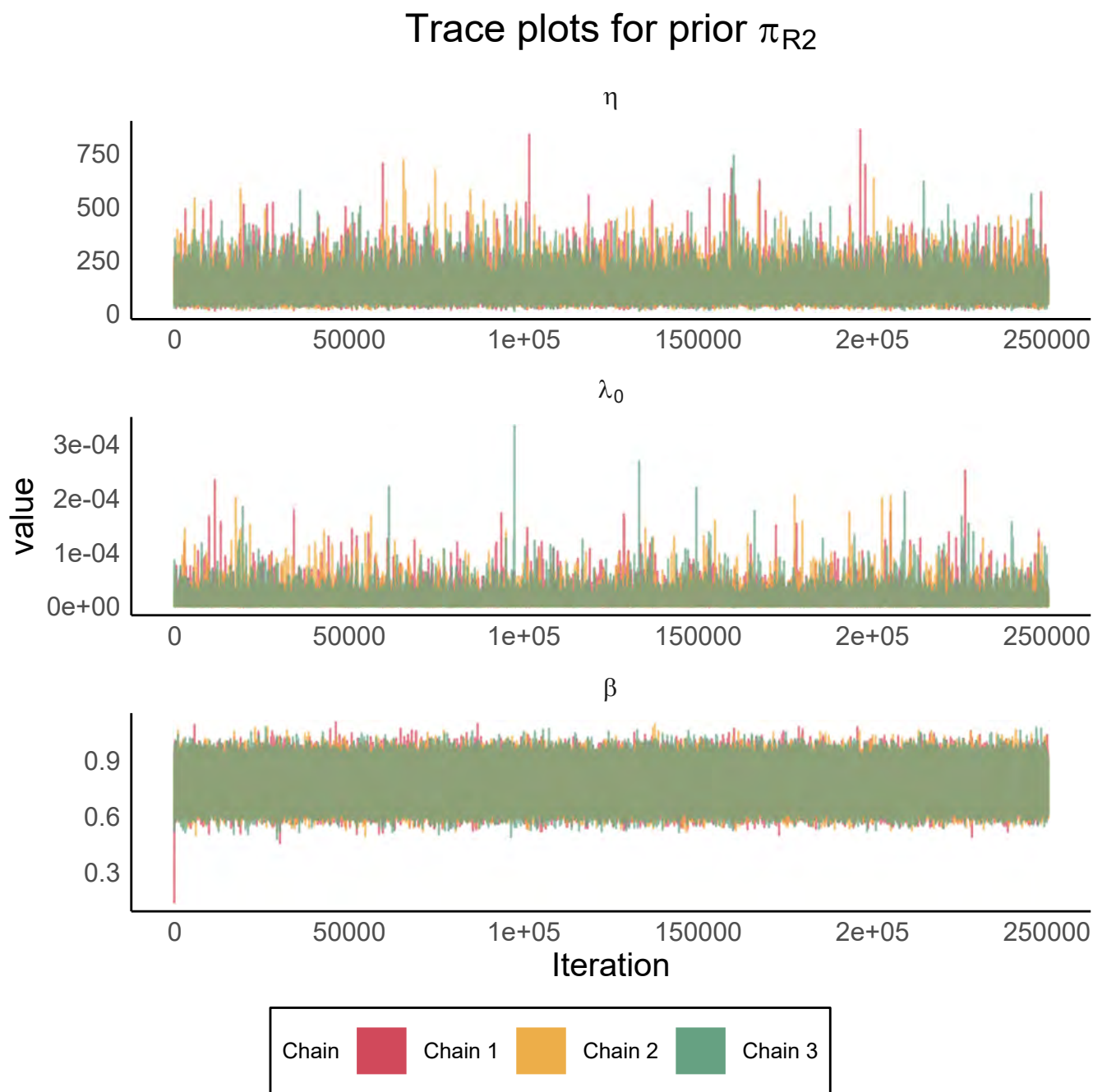
Figure C.2: Trace plots for the chains under prior π_{R2}.

C.15 Additional running mean plots
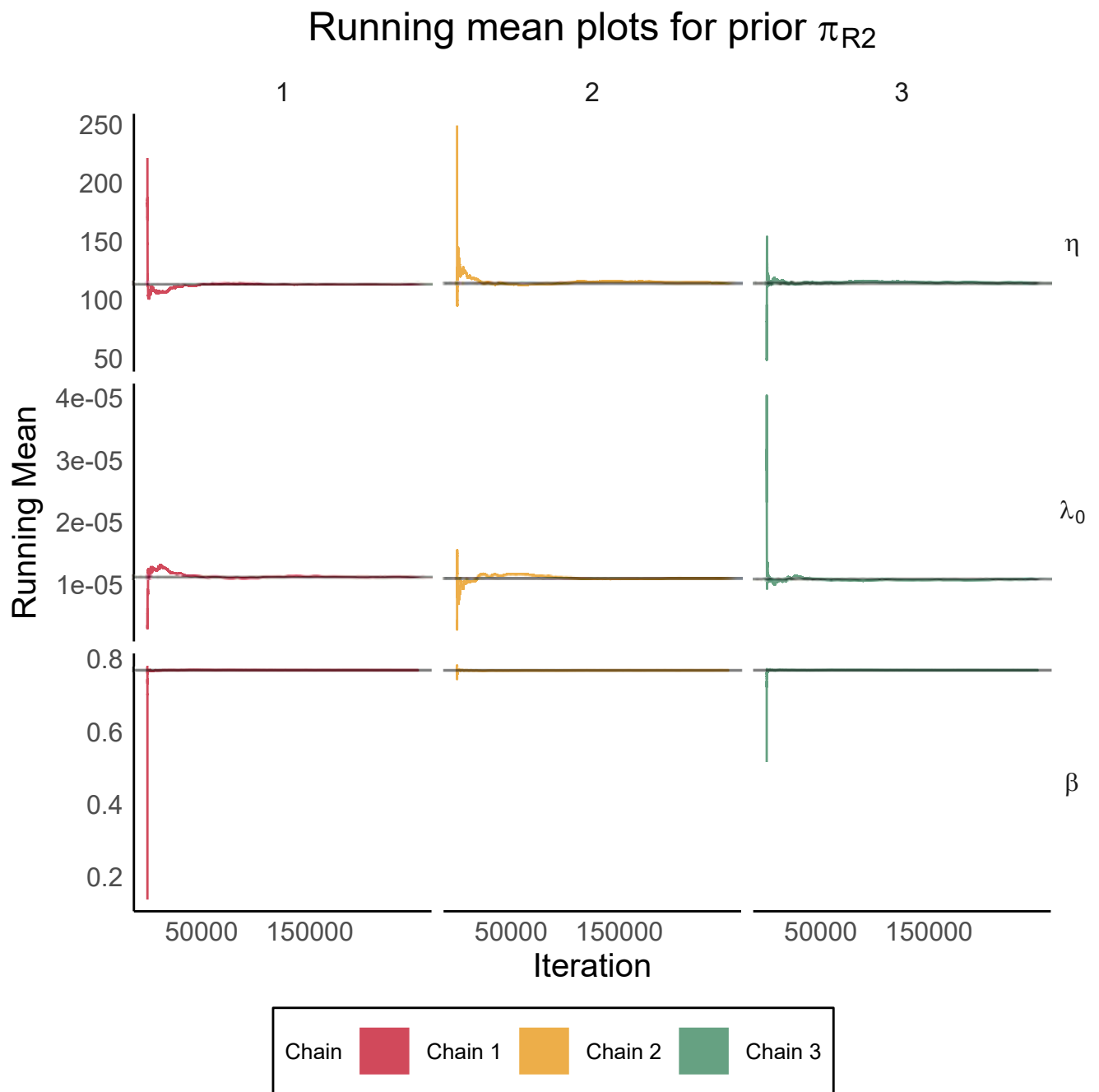
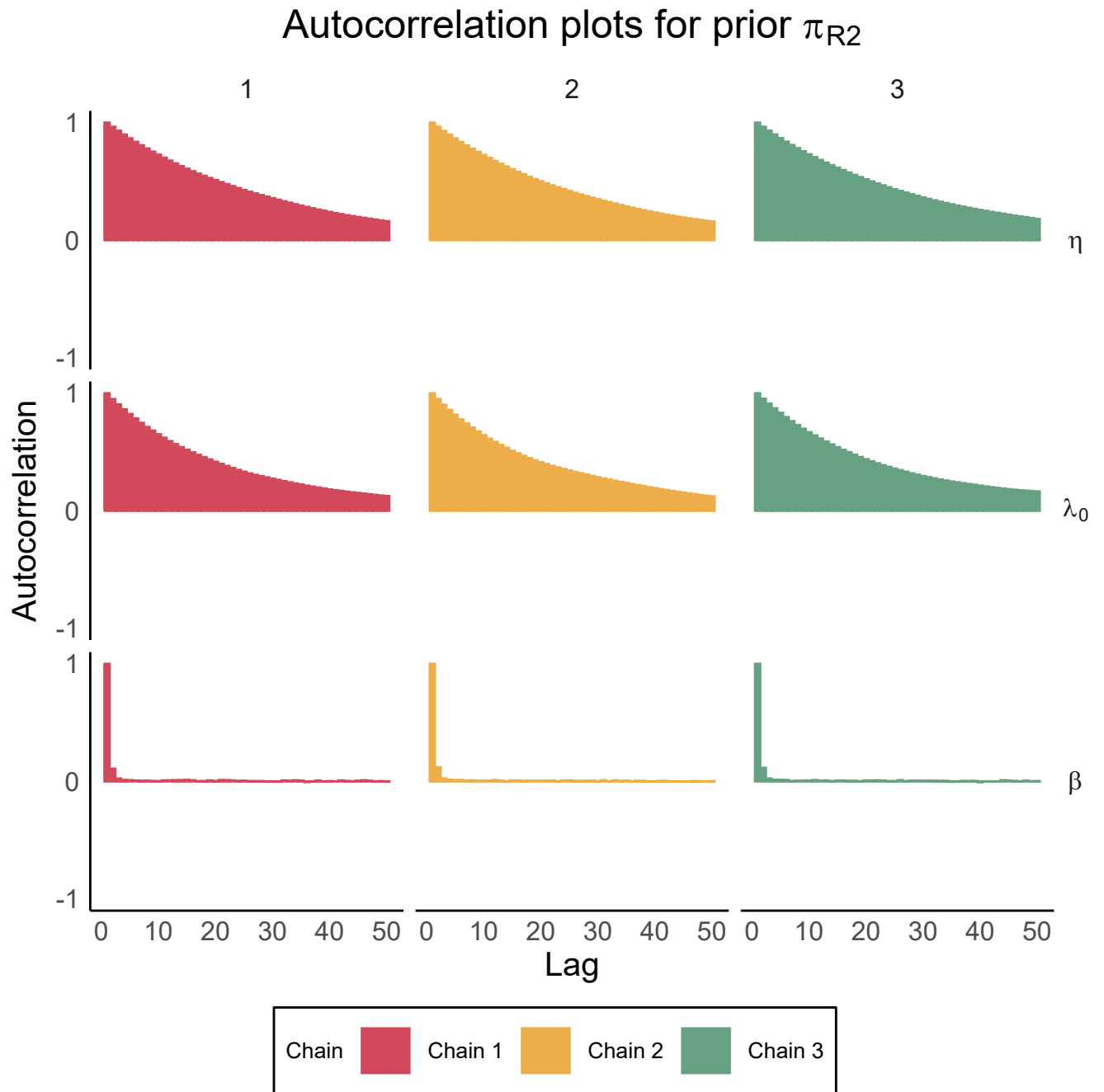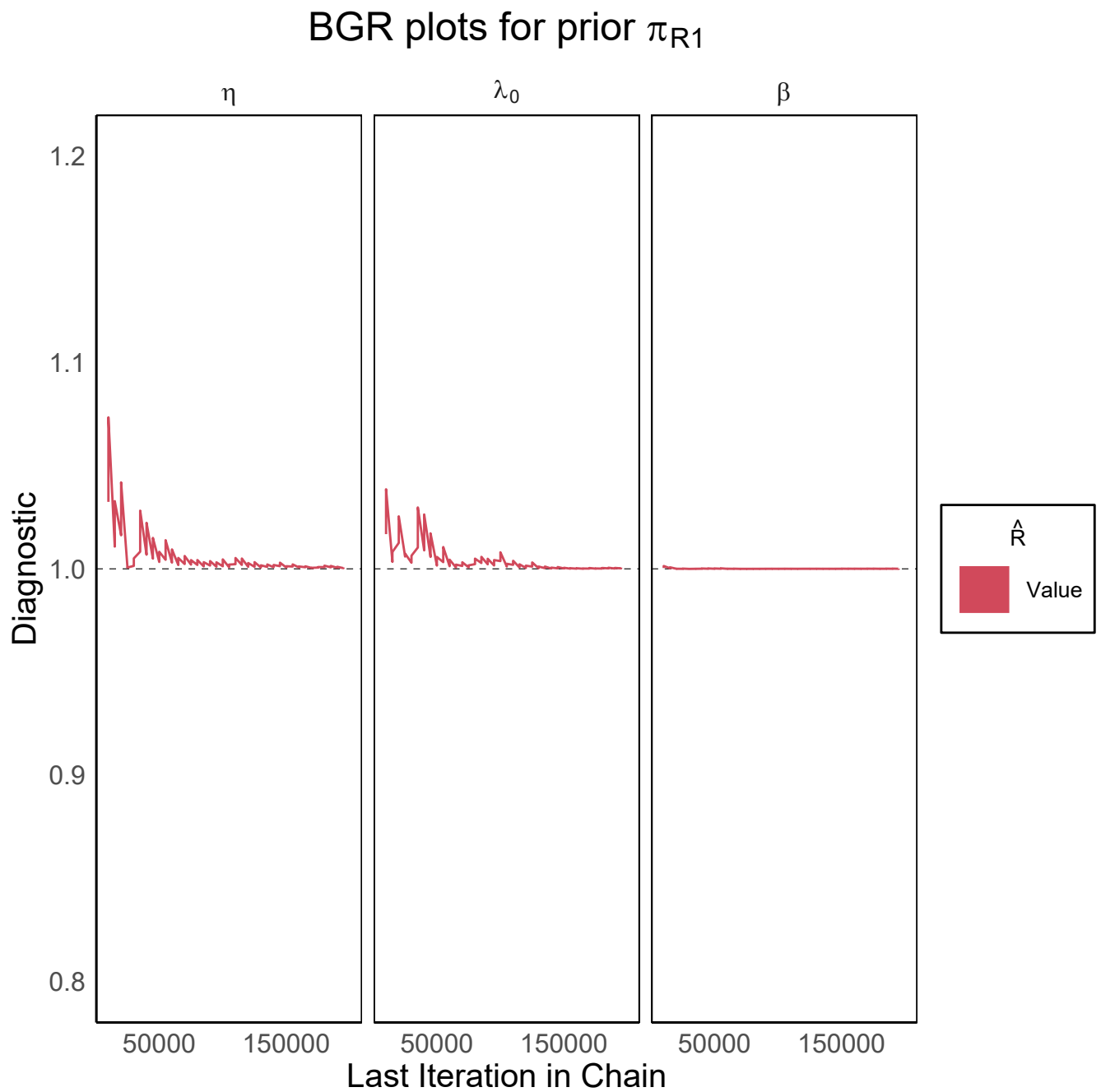**Figure C.3:** Running mean plots for the parameters under prior π_{R1}.
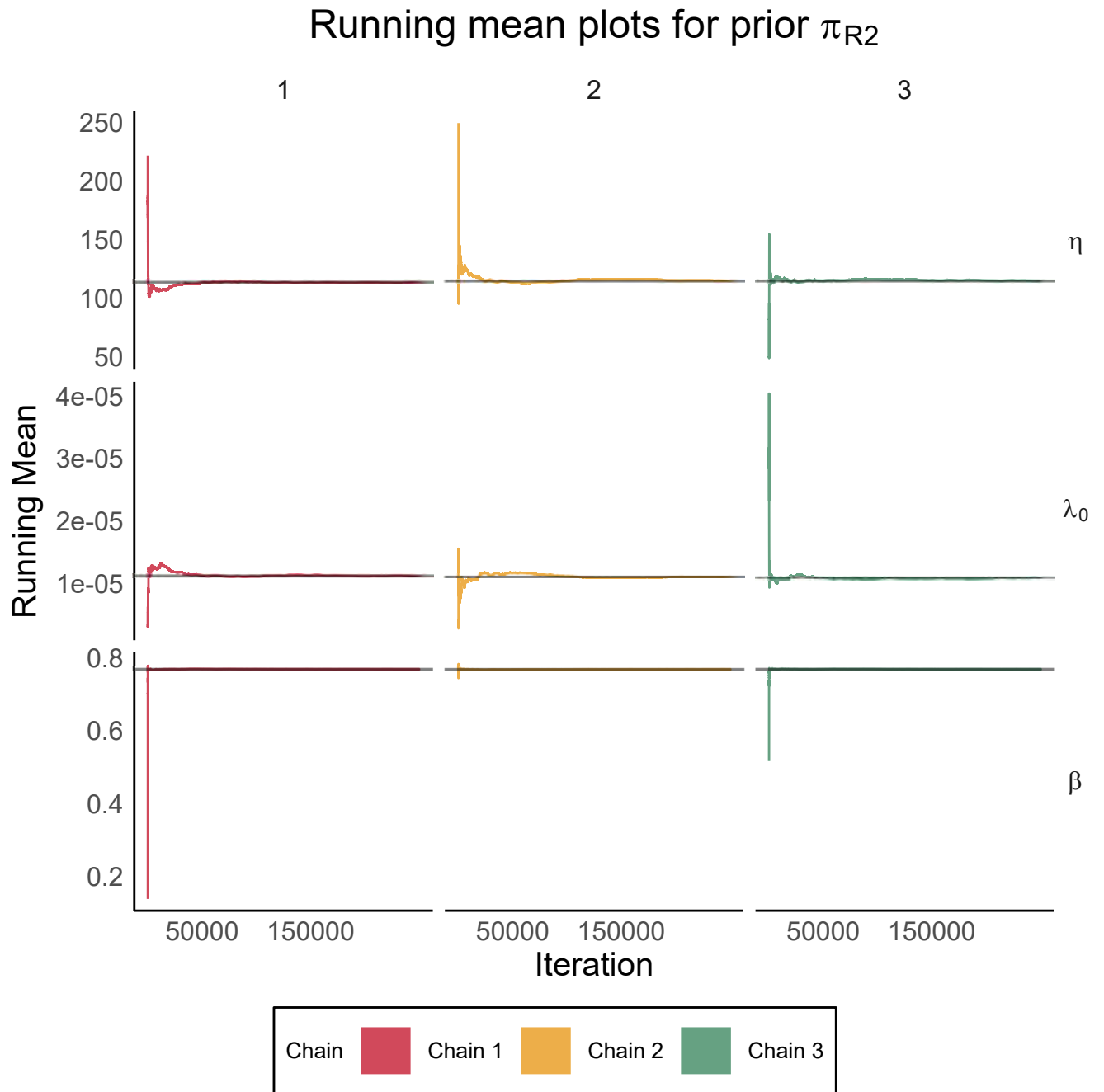
Figure C.4: Running mean plots for the parameters under prior π_{R2}.

C.16   Additional autocorrelation plots
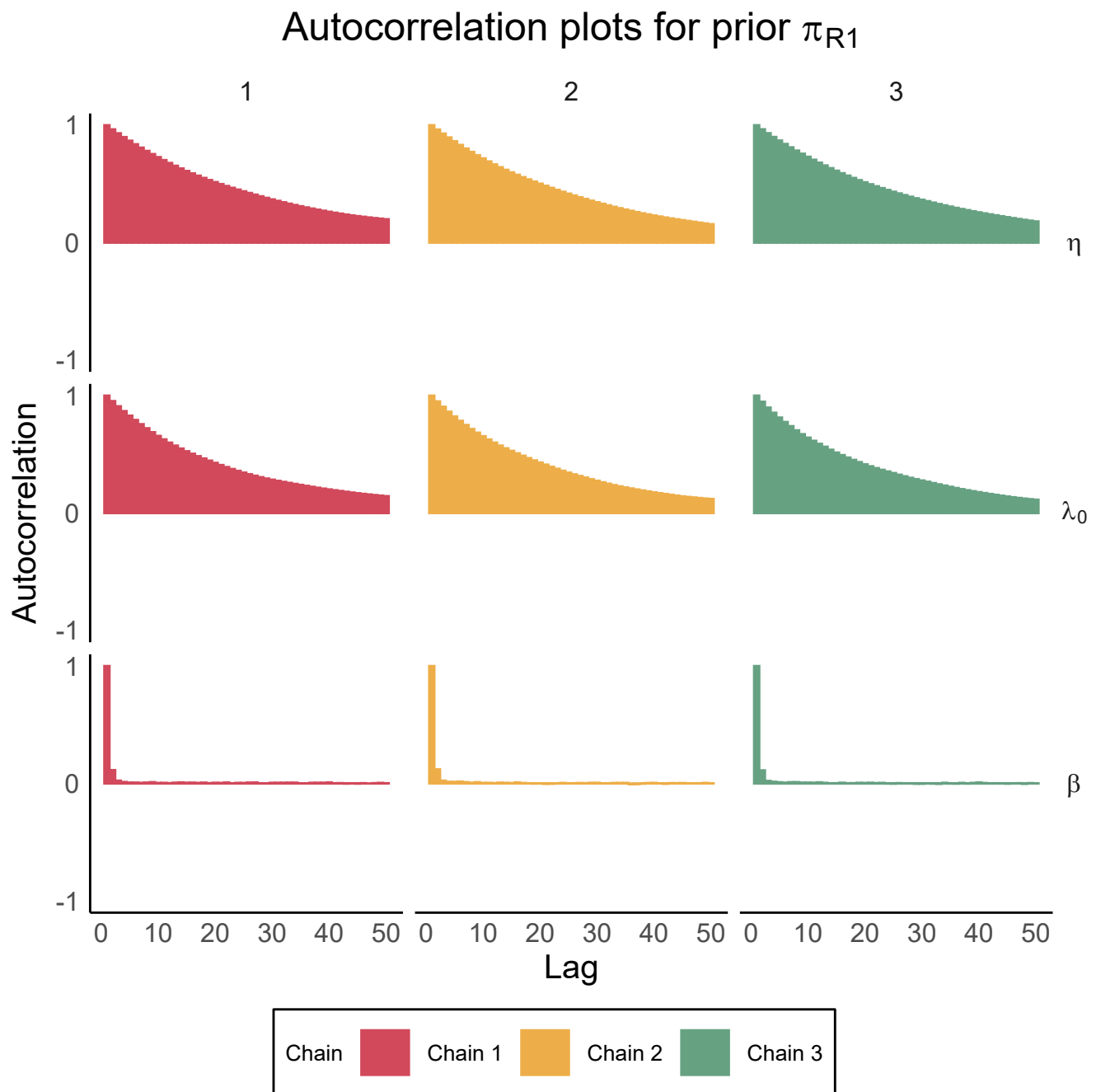
Figure C.5: Autocorrelation plots for the parameters under prior π_{R1}.

Figure C.6: Autocorrelation plots for the parameters under prior π_{R2}.

C.17 Additional BGR plots

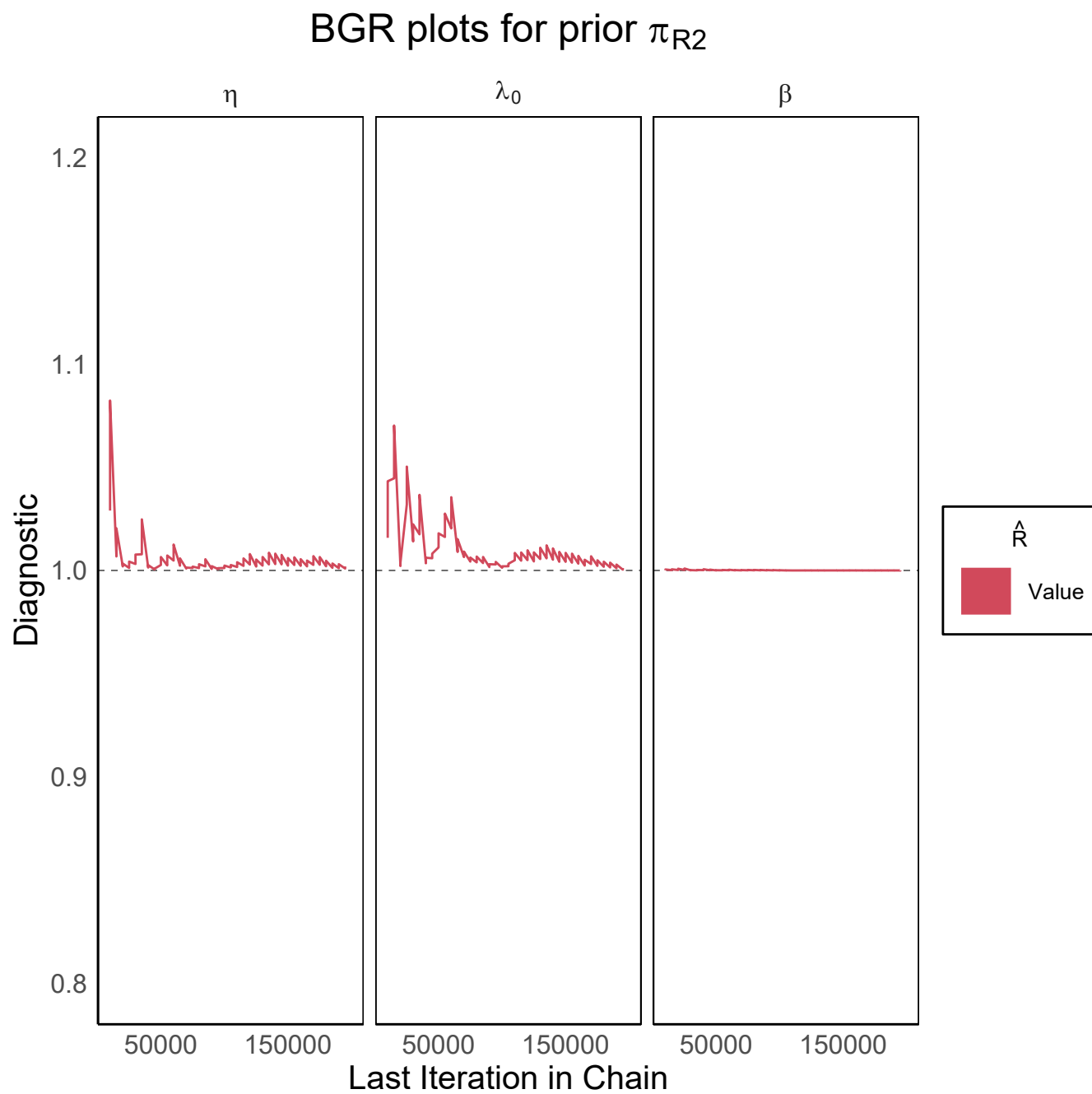**Figure C.7:** BGR plots for the parameters under prior π_{R1}.

Figure C.8: BGR plots for the parameters under prior π_{R2}.
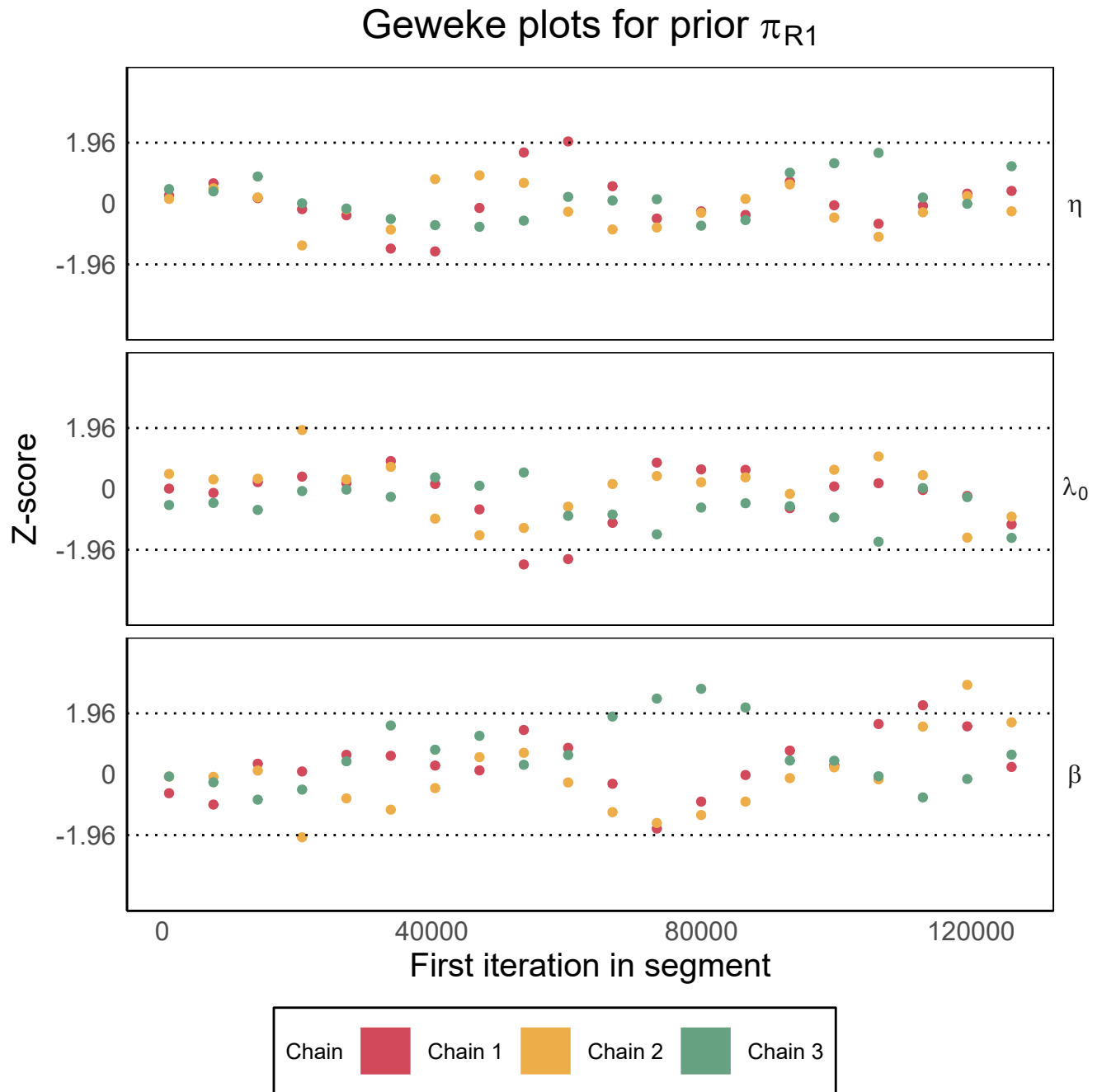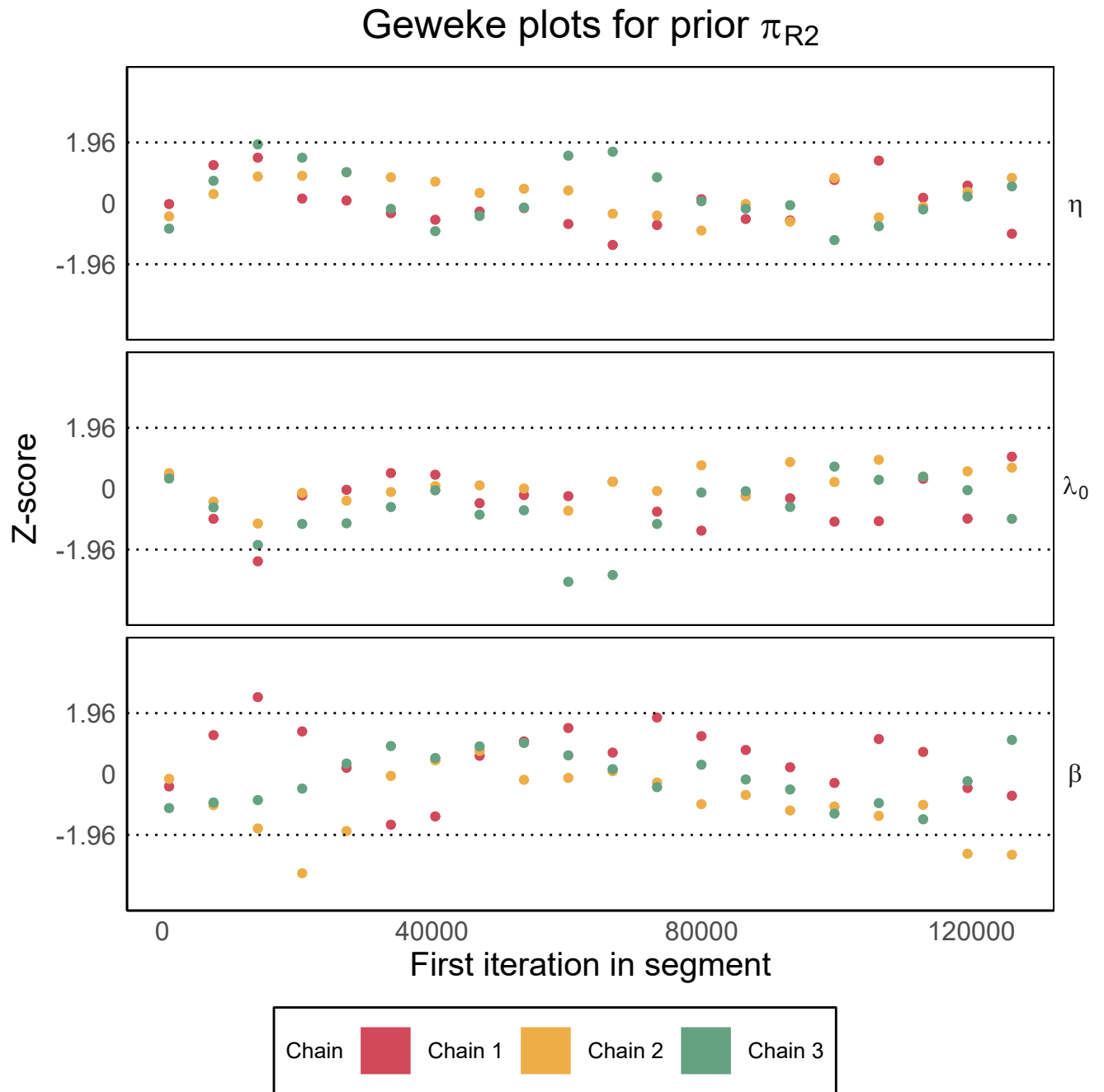
C.18   Additional Geweke plots
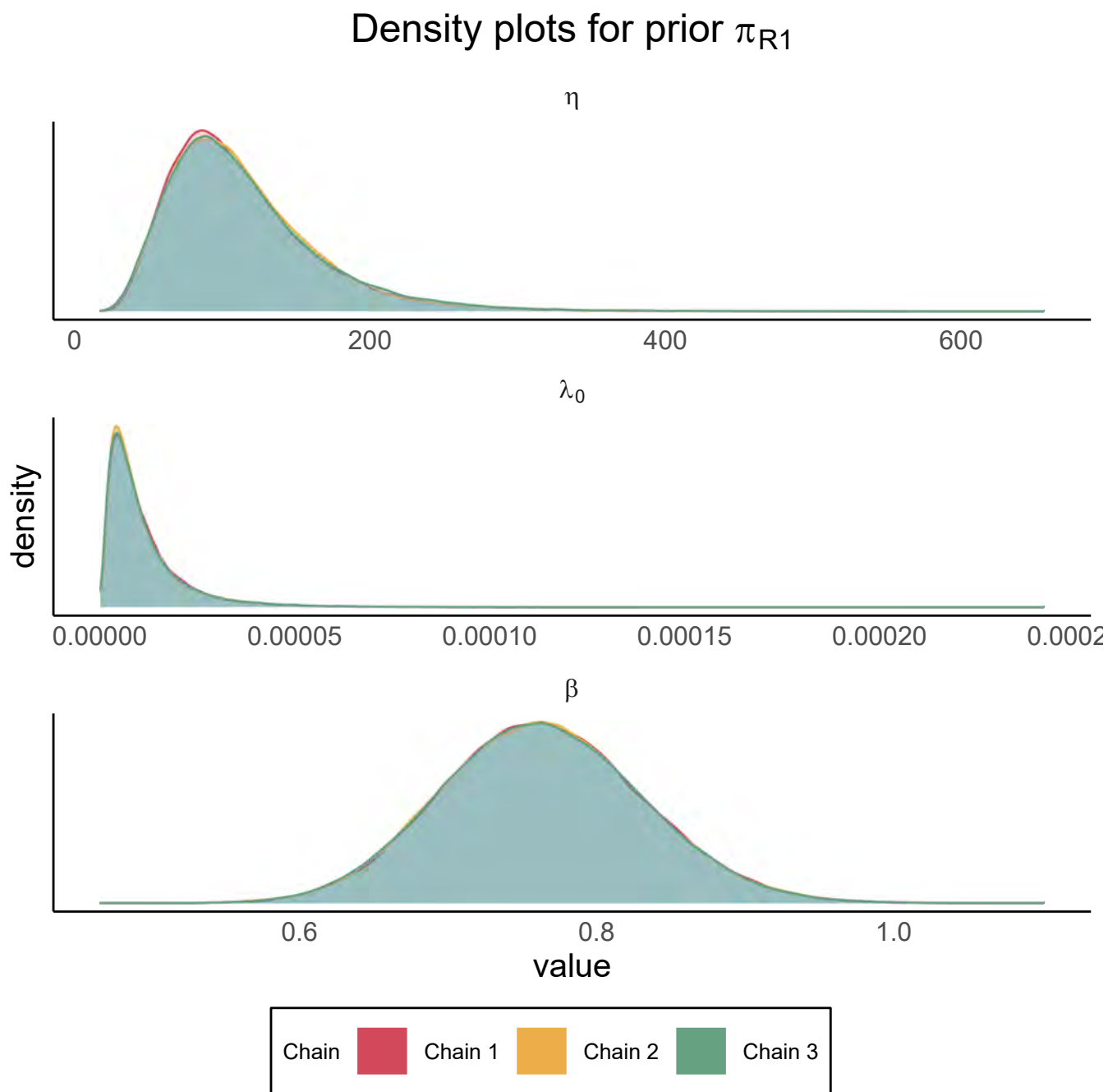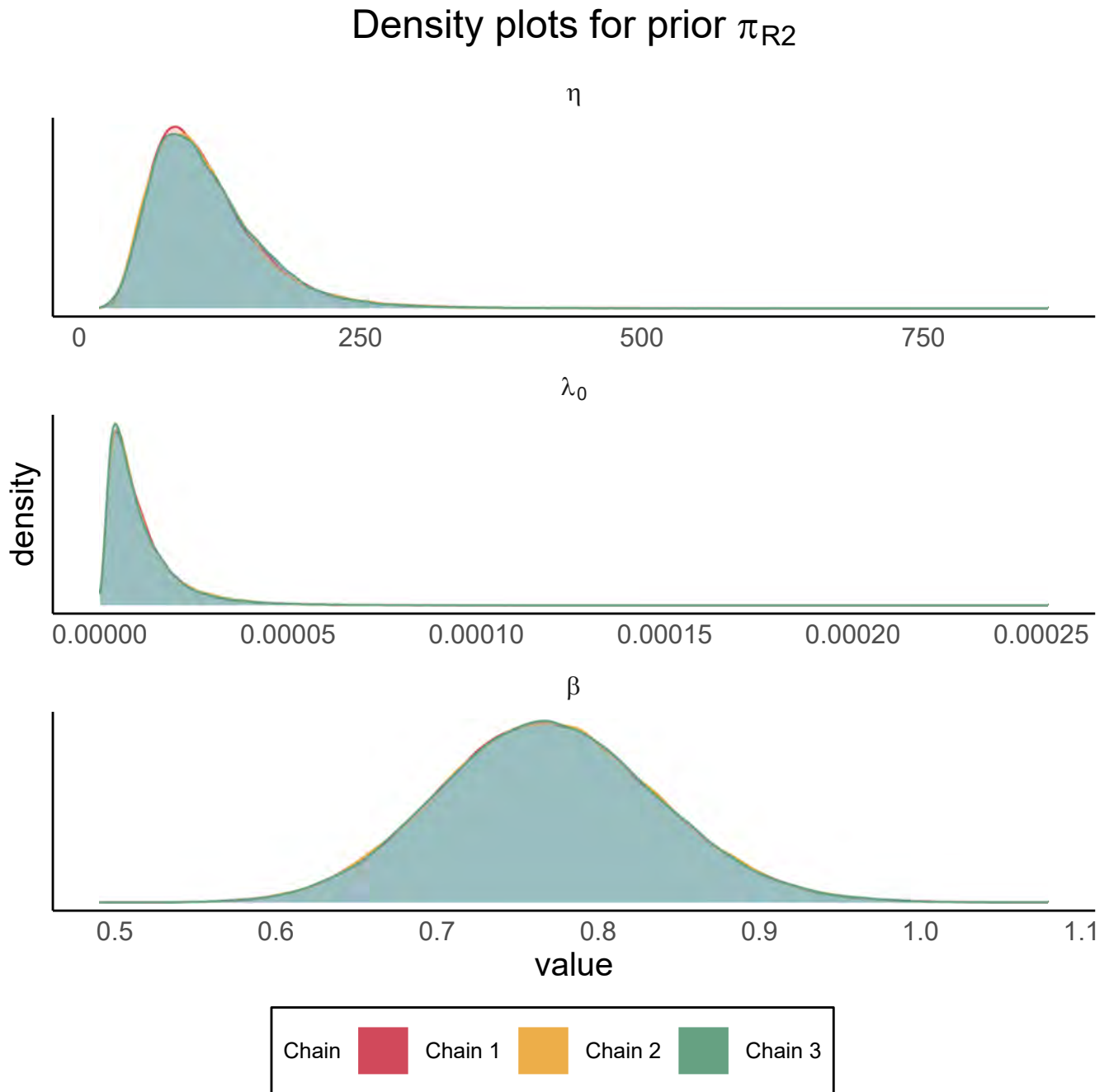

Figure C.9: Geweke plots for the parameters under prior π_{R1}.

Geweke plots for prior π_{R2}

Figure C.10: Geweke plots for the parameters under prior π_{R2}.

C.19   Additional density plots

Figure C.11: Density plots for the parameters under prior π_{R1}.

Density plots for prior π_{R2}

η

λ_0

β

| Chain | | Chain 1 | | Chain 2 | | Chain 3 |

Figure C.12: Density plots for the parameters under prior π_{R2}.

C.20 Additional Bayesian estimate tables

Table C.19: Estimates under prior π_{R1}.

Parameter	Mean	Sd	MCMC error	95% credible interval	Median
$\lambda_0(\times 10^{-6})$	11.2400	11.3800	0.1360	$(1.5200, 39.6600)$	8.0890
η	114.2000	51.6600	0.6799	$(45.4300, 243.4000)$	103.9000
β	0.7696	0.0682	0.0002	$(0.6396, 0.9073)$	0.7682

Table C.20: Estimates under prior π_{R2}.

Parameter	Mean	Sd	MCMC error	95% credible interval	Median
$\lambda_0(\times 10^{-6})$	11.4100	11.3900	0.1387	$(1.4740, 41.2300)$	8.1080
η	114.8000	53.2800	0.7309	$(44.6300, 248.3000)$	104.0000
β	0.7634	0.0682	0.0002	$(0.6336, 0.9008)$	0.7617