Accelerated life testing and the Burr XII distribution.

Johnson, Richard

Accelerated Life Testing and the Burr XII Distribution

by

Richard Johnson, BSc (Hons) (University of Wales, Swansea)

Thesis

submitted to the University of Wales
in candidature for the degree of

PHILOSOPHIÆ DOCTOR

European Business Management School
University of Wales Swansea
Swansea SA2 8PP
United Kingdom

February 2003

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

RICHARD JOHNSON
1 February 2003

Statement

This thesis is the result of my own investigation, except where acknowledgement of other sources is given.

RICHARD JOHNSON
1 February 2003

Statement

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan (subject to the law of copyright), and for the title and summary to be made available to outside organisations.

RICHARD JOHNSON
1 February 2003

Acknowledgements

First and foremost, I would like to express my considerable gratitude to my supervisor Dr. Alan Watkins, without whose unending advice, guidance (and patience!), this thesis would never have been possible.

I would also like to thank my sponsor, the EPSRC, for their continued financial support. Also, I would like to express my thanks to my colleague and good friend Andrea John, who has acted as a superb sounding board over the last couple of years.

In a similar vein, the warmth and sincerity offered by my friends at the European Business Management School, has provided an environment that has been a pleasure to work in; thank you each and every one.

Finally, I would like to voice my heart-felt appreciation to my family for their continued support throughout my life.

RICHARD JOHNSON

University of Wales
February 2003

Summary

This thesis looks at extending previous work in the field of accelerated life testing experiments. Hitherto, much investigation in this field has centred on a few standard statistical lifetime distributions, the Weibull being particularly popular. We consider a more flexible distribution, the Burr XII, and compare theoretical and simulated results; we also examine a number of examples. This comparison is interesting particularly since there is a limiting relationship between the two distributions, a property that is exploited in this thesis.

Having laid down the necessary groundwork, we then proceed to fit the Weibull and Burr XII models to completely failed, published data sets and compare results. In order to assess our ability to make small sample theoretical inspections, we then establish the expected Fisher information matrices for the accelerated and non-accelerated Burr XII models, validating our results through simulations.

The limiting link between the two distributions is then investigated, where we see that we can determine whether the Burr XII distribution will provide a better fit to a given data set than the Weibull, by fitting the Weibull distribution and then calculating a simple discriminating function.

Type I censoring for the accelerated models is then considered. As for complete data, we formulate the expected Fisher information matrices. We then examine theoretical agreement between these results and those obtained for completely failed data. We also examine the agreement between Type I simulated and theoretical results.

Finally, we investigate the practical applications of our work, and consider in particular extrapolations to lower operating stresses and the expected lifetimes of items tested at those stresses. Our investigations, although based on limited parameter values, illustrate useful conclusions on the conduct of such experiments and, consequently, are of potential value to a practitioner who, prior to carrying out an experiment, would like to know what combination of stresses and sample sizes would return the most information about the running time of items at the normal operating stress.

After summarising our results and conclusions, some ideas for future research are detailed.

To my parents.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Role Of Lifetime Modelling

Things fail! It's a fact of life. Look around you and every day you see components expiring, light bulbs exploding, the washing machine breaking down and so on. Even at a more personal level, we see animals dying and our loved ones passing on due to some biological malfunction. Quite simply, as soon as something is created it is already on the slow journey towards death. How soon it reaches its destination will depend on a vast array of factors. For example, the rapidity of death for an organism will depend on its climate, diet, threat from predators and genetic makeup inherited from its parents to name but a few factors; a light bulb might fail because of a poorly made tungsten filament, a surge of electricity or overuse. The complexity and interaction of components in a washing machine makes one wonder how such a mechanism gets through a single day without something failing. But the fact of the matter is, it does! Indeed, in a large majority of cases such appliances can last five, ten or even twenty years without failing.

The reason for this is quite simple: all the components that make up the appliance have been tested to breaking point. Having determined what caused the failure, improvements are then made in the manufacture of the component with the aim of increasing its longevity. However, in many cases the experimenter cannot easily simulate the normal operating conditions of the component. For example, a light bulb has an estimated lifetime of some 10000 hours. When conducting the experiment, it is implausible simply to switch on a bulb and wait 10000 hours, or so, for it to fail. Besides, during the course of a bulb's life it will be switched on and off and will remain in each of these states for varying lengths of time. Somehow, the normal ageing process of the bulb needs to be mimicked during testing, but at a faster rate.

This raises the following question: if we increase the rate at which the bulb is used, (switch it on and off more frequently, pass a greater current through it, et cetera), with the aim of inducing a quicker failure, how do we then ascertain the lifetime of the bulb under normal operating conditions?

1.1.1 Accelerated Life Testing

Accelerated Life Testing, (hereafter abbreviated to ALT), is the name given to the process by which we exert a product - such as a light bulb - to one or more external stresses - such as temperature, current or humidity - above and beyond those found in normal operating conditions in an attempt to induce early failure. Although in reality a combination of factors tend to interact to have an overall effect on the product, ALT in practice usually examines the effect of external factors individually and independently of other factors. For example, the Coffin-Manson relationship, which concerns fatigue failure of metals, gives the number of cycles to failure as a function of the temperature range, Palmgren's equation specifies life of roller ball bearings - in millions of revolutions - as a function of load, and Taylor's model for lifetime of cutting tools is a function of the cutting velocity.

These, and other, examples are discussed by Nelson (1990,pp86-87), who links a number of life distributions with various life-stress relationships; see below. Nelson (1990) summarises work to that date, and may be regarded as a convenient starting point for our discussion; however, his approach is essentially a practical one - for instance, complete data is examined using least squares analysis - although the analysis of censored data (involving items that have not failed at the time of analysis) involves maximum likelihood methods. Even with this more theoretical approach, Nelson says "*for most models, the formulas for these variances and covariances* [of model parameters] *are too complex to express analytically.*"(p369). Nelson also considers experimental design and, briefly, the subject of step stress testing; see also the review by Meeker and Escobar (1993). We shall look to adopt this approach for other models, with the primary aim of developing the asymptotic expressions necessary for a detailed examination of maximum likelihood theory. We concern ourselves solely with constant stresses; the field of step-stress testing requiring a different discussion altogether.

Thus, we view a typical ALT experiment as a process in which we select a range of values for a single stress factor, x, and at each of $k\,(\geq 2)$ specified stress levels, $(x_1, x_2, \cdots, x_k)$, we test batches of identical components $(n_1, n_2, \cdots, n_k)$. For example, we might expect, under normal conditions, for a ball bearing to operate in oil at a temperature of $50°C$; temperature is then the stress x and $50°C$ is the *design stress,* denoted by x_d. In setting up the ALT experiment, we might decide to immerse ten bearings each in three vats of oil at temperatures of $x_1 = 100°C, x_2 = 150°C$ and $x_3 = 200°C$. In order to describe the lifetime - or, more usually, the $100p^{th}$ percentile of life, typically denoted by B_{100p} (see, for example, Nelson, 1990,pp86-87) - under normal operating conditions, we need to be able to model the resultant lifetime data, using an appropriate statistical distribution and life-stress relationship. In the above scenario, these two stages will allow us to analyse the data obtained at temperatures of $100°C, 150°C$ and $200°C$, and then to extrapolate results back to estimate performance at a temperature of $50°C$.

Hitherto, work in this field has centred around the negative exponential and, more

commonly, Weibull distributions; the latter is particularly popular in the field of actuarial studies since it can model both 'wear-out' and 'infant mortality' patterns of failure, albeit not simultaneously. More recently, investigations have been conducted using the Burr XII distribution; this has more parameters - and hence greater flexibility - than either the negative exponential or Weibull. We introduce these distributions in more detail in the next section.

1.2 Basic Reliability Distributions

There are many references which provide an introduction to the statistical analysis of quantitative reliability data, and which present the fundamental results for basic reliability distributions. For instance, Patel, Kapadia and Owen (1976) present a concise overview of properties, such as moment generating functions and order statistics, of many widely used statistical distributions, including Normal, Cauchy, Logistic and Weibull. The Burr XII distribution is mentioned briefly, but there is little discussion about its properties. Johnson and Kotz (1970) provide more details on key reliability distributions, including the Weibull, but only a brief outline of the Burr XII distribution, together with Burr's other eleven distributions.

1.2.1 Weibull and Burr XII Distributions

The cumulative distribution function (hereafter abbreviated to CDF) for a random variable D following the two-parameter Weibull distribution is

$$G(d; B, \phi) = 1 - \exp\left\{-\left(\frac{d}{\phi}\right)^B\right\}, \tag{1.1}$$

for $d \geq 0$, with associated probability density function (hereafter abbreviated to PDF)

$$g(d; B, \phi) = \frac{Bd^{B-1}}{\phi^B} \exp\left\{-\left(\frac{d}{\phi}\right)^B\right\}, \tag{1.2}$$

in which the positive quantities B, ϕ are shape and scale parameters respectively. Since

$$Z = \left(\frac{D}{\phi}\right)^B$$

has a standard negative exponential distribution, we can use

$$E[Z^r] = \int_0^\infty z^r \exp(-z)\, dz = \Gamma(1+r)$$

to write

$$E[D^s] = \phi^s \Gamma\left(1 + B^{-1}s\right),$$

where $\Gamma(\cdot)$ denotes the usual gamma function. This defines all moments of D in terms of the gamma function; we summarise useful properties of this function below.

The basic two-parameter Burr XII distribution, introduced by Burr (1942) has parameters a and τ with CDF

$$1-(1+d^{\tau})^{-a}, \tag{1.3}$$

for $d \geq 0$, with associated PDF

$$a\tau d^{\tau-1}(1+d^{\tau})^{-(a+1)} \tag{1.4}$$

for $d \geq 0$, where both a and τ are positive shape parameters. We can introduce a third - scaling - parameter into (1.3) in several ways and Tadikamalla (1980) discusses various options. The most natural seems to be to define

$$Y = \theta D$$

where $\theta > 0$ is a natural scale parameter, and D now follows (1.3). Then Y has CDF

$$F(y; a, \tau, \theta) = 1-\left\{1+\left(\frac{y}{\theta}\right)^{\tau}\right\}^{-a} \tag{1.5}$$

for $y \geq 0$. The moments of Y are

$$E[Y^s] = \frac{\theta^s \Gamma\left(1+\frac{s}{\tau}\right)\Gamma\left(a-\frac{s}{\tau}\right)}{\Gamma(a)}; \tag{1.6}$$

see, for example, Watkins (1997).

1.2.2 Link between Weibull and Burr XII Random Variables

Adopting a parameterisation used by Hogg and Klugman (1984), we now write $\theta^{\tau} = \lambda$, so that the CDF (1.5) becomes

$$1-\left(1+\frac{y^{\tau}}{\lambda}\right)^{-a}; \tag{1.7}$$

note that both (1.5) and (1.7) reduce to (1.3) when $\theta = \lambda = 1$. We can now write (1.7) as

$$1-\left(1+\frac{y^{\tau}}{\lambda}\right)^{-\left(\frac{a}{\lambda}\right)\lambda} = 1-\left\{\left(1+\frac{y^{\tau}}{\lambda}\right)^{\lambda}\right\}^{-\left(\frac{a}{\lambda}\right)},$$

so that on letting $\lambda \to \infty$ with $\frac{a}{\lambda}$ remaining finite, this CDF approaches

$$1-\{\exp(y^{\tau})\}^{-\left(\frac{a}{\lambda}\right)} = 1-\exp\left\{-\frac{ay^{\tau}}{\lambda}\right\},$$

on recognising the limiting form of $\exp(y^\tau)$. With $\lambda = \theta^\tau$ this can be written as

$$1 - \exp\left\{-a\left(\frac{y}{\theta}\right)^\tau\right\} = 1 - \exp\left\{-\left(\frac{y}{a^{-1/\tau}\theta}\right)^\tau\right\} = G\left(y; \tau, \frac{\theta}{a^{1/\tau}}\right).$$

Hence, the shape parameter of this limiting Weibull distribution is τ, while the scale parameter is

$$\frac{\theta}{a^{1/\tau}} = \left(\frac{\lambda}{a}\right)^{1/\tau}. \tag{1.8}$$

We shall return to this very important fact later. Here, we see the Burr XII shape parameter τ corresponds to the Weibull shape parameter B, while a is effectively an extra shape parameter.

1.2.3 Alternative Reliability Distributions

In addition to the Weibull and Burr XII models, there are many other lifetime distributions; see Richards and McDonald (1987) for a discussion on relationships between them. We briefly list some of these models, which are discussed in greater depth by Nelson (1990, Chapter 2);

Extreme Value Distribution

With CDF

$$F(t) = 1 - \exp\{-\exp[(t-\varepsilon)/\delta]\}$$

for $-\infty < t < \infty$, with location parameter ε $(-\infty < \varepsilon < \infty)$ and scale parameter δ $(\delta > 0)$ the extreme value distribution is useful in situations where a system consists of several components and fails when the weakest component fails.

Birnbaum-Saunders Distribution

Eponymously proposed (see Birnbaum and Saunders, 1969) to describe metal fatigue, and derived 'from a model for crack propagation', this distribution has CDF

$$F(t) = \Phi\left\{\left[\sqrt{t/\beta} - \sqrt{\beta/t}\right]/\alpha\right\}$$

for $t > 0$. Here, β $(\beta > 0)$ is the median of the distribution, and so is a location parameter, while α $(\alpha > 0)$ is a shape parameter; $\Phi(\cdot)$ is the standard Normal CDF.

Lognormal Distribution

In this distribution, the probability of failing by time $t > 0$ is given by the CDF

$$F(t) = \Phi\{[\ln(t) - \mu]/\sigma\}$$

where $\mu\,(-\infty < \mu < \infty)$ is the mean of the log of life, $\sigma\,(\sigma > 0)$ is the standard deviation of the log of life and - as throughout our work - ln is the natural logarithm. Alternative formulations use logarithms to base ten. This distribution is commonly used in solid state experiments (semiconductors, diodes, et cetera) and metal fatigue.

Nonparametric models

Although these are frequently used in the biosciences, they are less popular in engineering circles. One problem is the relative inaccuracy of the nonparametric model compared to a correctly specified one; we cannot easily extrapolate to a lower design stress with a nonparametric model. The most commonly used nonparametric model is the proportional hazards model (also known as the Cox model) where, despite the use of a parametric equation linking life and stress, no underlying lifetime distribution is assumed, see, for example, Kalbfleisch and Prentice (1980).

1.3 Data Available and Censoring Regimes

Usually, the data available for analysis consists of the failures measured in units of time. Equally, however, failures can be satisfactorily measured in terms of some other aspect of life or usability, such as pressure or cycles per second, for example, as we shall see later. Similarly, the stress factor, which is often measured in units of temperature, can also be measured using other factors that affect lifetime or usability, such as humidity or length, for example. Now, frequently when carrying out an experiment to collect reliability data, time constraints may dictate that we cannot wait until all items have failed; that is to say, wait until we have a *complete sample.* Under these conditions we can adopt and adapt one of several well known approaches to obtaining observed life time data; particularly common is Type I censoring. We define the overall sample size to be N with individual failure times in a complete sample then given by $D_1, D_2, \cdots, D_N$.

1.3.1 Type I Censoring

Under Type I censoring, the observation period ends at the pre-specified time c, and all those items which have not failed yield the censored time c; thus, the number of observed failures is a random variable, denoted by n, and, without loss of generality, we may denote the data available for analysis by the n times to failure $D_1, \ldots, D_n \leq c$, and m censored values $D_{n+1} = \cdots = D_{n+m} = c$. Naturally then, $N = n + m$. Clearly, $c \to \infty$ gives a complete sample, with $n \to N$.

Type II censoring is also used in the analysis of failure data. Here, we wait until a certain number of items have failed; the stopping time is then the random variable. However, we can imagine that in reality, an experiment is more likely to be time-constrained, with results typically being required by the end of the day, or the end of the week, for example.

Consequently, we shall focus our attention on Type I censoring, although we discuss some possible approaches to Type II censoring, when we come to look at Type I in more detail later.

1.4 Useful Functions

We now summarise some results used to develop the theory associated with accelerated lifetime experiments. Many are neatly presented by Abramowitz and Stegun (1972), another key reference in this discussion.

1.4.1 The Gamma Function

We have already introduced

$$\Gamma(u) = \int_0^\infty t^{u-1} \exp(-t)\, dt,$$

for $u > 0$. Some useful properties of the Gamma function include $\Gamma(u+1) = u!$ for integer values of u, while, more generally, we have the recurrence relation

$$\Gamma(u+1) = u\Gamma(u);$$

differentiating this with respect to u, we obtain

$$\Gamma'(u+1) = u\Gamma'(u) + \Gamma(u). \tag{1.9}$$

Since

$$\left.\frac{d\Gamma(u)}{du}\right|_{u=1} = \Gamma'(1) = -\gamma$$

where $\gamma = 0.5772156\cdots$ is Euler's constant, using (1.9) with $u = 1, 2$, we also obtain

$$\Gamma'(2) = \Gamma'(1) + \Gamma(1) = 1 - \gamma$$

and

$$\Gamma'(3) = 2\Gamma'(2) + \Gamma(2) = 3 - 2\gamma.$$

Extending this to second derivatives, we have

$$\Gamma''(u+1) = u\Gamma''(u) + 2\Gamma'(u), \tag{1.10}$$

and

$$\Gamma''(1) = \frac{\pi^2}{6} + \gamma^2,$$

so that, taking $u = 1, 2$ in (1.10), we now have.

$$\Gamma''(2) = \Gamma''(1) + 2\Gamma'(1) = \frac{\pi^2}{6} + \gamma^2 - 2\gamma$$

and

$$\Gamma''(3) = 2\Gamma''(2) + 2\Gamma'(2) = \frac{\pi^2}{3} + 2\left(\gamma^2 - 3\gamma + 1\right).$$

1.4.2 Psi or Digamma Function

The results on derivatives of the Gamma function can also be expressed in terms of the Psi or Digamma Function, defined as

$$\psi(u) = \frac{d\left[\ln \Gamma(u)\right]}{du} = \frac{\Gamma'(u)}{\Gamma(u)}$$

1.4.3 Incomplete Gamma Function

This is defined as

$$\gamma(u, x) = \int_0^x t^{u-1} \exp(-t)\, dt$$

for $u, x > 0$; clearly, as $x \to \infty$, the incomplete Gamma function tends to the Gamma function.

1.4.4 Beta Function

This is defined as

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1}\, dt = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)},$$

writing this in terms of the Gamma function.

1.4.5 Incomplete Beta Function

The generalisation of the Beta function to an arbitrary upper limit gives us the incomplete Beta function, defined as.

$$B_z(a, b) = \int_0^z t^{a-1} (1-t)^{b-1}\, dt$$

1.4.6 Generalized Hypergeometric Function

This is defined as

$$F_{p,q}(a_1, \ldots, a_p; b_1, \ldots, b_q; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k \ldots (a_p)_k}{(b_1)_k \ldots (b_q)_k} \frac{z^k}{k!}$$

where $(a)_k$ is Pochhammer's Symbol, defined by

$$(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)}.$$

A specific case, frequently used, has $p = 2, q = 1$, which then gives us

$$F_{2,1}(a_1, a_2; b_1; z) = \frac{\Gamma(b_1)}{\Gamma(a_1)\Gamma(a_2)} \sum_{k=0}^{\infty} \frac{\Gamma(a_1+k)\Gamma(a_2+k)}{\Gamma(b_1+k)} \frac{z^k}{k!},$$

where we now write terms in the summation explicitly in terms of the Gamma function.

1.5 Kaplan-Meier Plots

The purpose of a Kaplan-Meier plot is to display estimates of the survival function for survival data. It presents the Empirical Cumulative Distribution Function (hereafter abbreviated to ECDF) for a set of sample data, as illustrated by Newton (1991). The ECDF at time t_i is defined as

$$M(t_i) = 1 - \prod_{j=1}^{h_i} \left(\frac{s_j - v_j}{s_j} \right) \tag{1.11}$$

where s_i denotes the number of survivors immediately before time t_i, v_i is the number of failures at t_i, and h_i the number of distinct failure times before t_i. We shall illustrate the use of this function later when we look at specific examples.

1.6 Links Between Stress And Scale

In order to use results from an accelerated lifetime experiment, we need to link the failures observed to an external stress level. Typically, this is done by defining the scale parameter, in our case ϕ or θ, to be a function of the stress variable. Ideally, we look for a function that is flexible enough both to deal with many different scenarios yet is also sufficiently tractable to allow for the development of supporting theory. Many such functions exist - for example, the Arrhenius, Eyring, Power Law and Exponential - and Nelson (1990, Chapter 2) presents a detailed discussion on these. In summary, the Arrhenius model is very good at modelling the effect of temperature on product life, the Eyring model is similar to the Arrhenius and was developed for use in quantum mechanics, while the widely-used Power Law model presents the log of the scale of lifetimes as a function of the log of the accelerating stress. Finally, the Exponential relationship (or log-linear, as we will term it) has been used to model the lifetimes of dielectrics and other such electronic components. The adaptability of this final model lends itself well to the current field of study, and we base our discussion on the function

$$\text{scale parameter of lifetimes at } x_i = \exp(\alpha + \beta x_i) \tag{1.12}$$

where α and β are two new parameters.

To illustrate the effect of (1.12), we now recall the example above in which ball bearings were immersed in oil, and we wished to use three stress levels of $100°C, 150°C$ and $200°C$. Fitting a Weibull ALT model to the observed data now requires that we first estimate three parameters, B, α_w, β_w, from which we would then calculate three separate scale parameters:

$$\phi_1 = \exp(\alpha_w + \beta_w x_1), \phi_2 = \exp(\alpha_w + \beta_w x_2) \text{ and } \phi_3 = \exp(\alpha_w + \beta_w x_3),$$

and retain the estimate of the single shape parameter B. Similarly, fitting a Burr XII ALT model to the data we would now estimate four parameters, $a, \tau, \alpha_b, \beta_b$, which, in turn, would give us the three scale parameters

$$\theta_1 = \exp(\alpha_b + \beta_b x_1), \theta_2 = \exp(\alpha_b + \beta_b x_2) \text{ and } \theta_3 = \exp(\alpha_b + \beta_b x_3),$$

while retaining the two shape parameters a, τ. From this, it is clear that, if we are considering only $k = 2$ stress levels, then the replacement of ϕ_1, ϕ_2 or θ_1, θ_2 with α_w, β_w or α_b, β_b, in the Weibull or Burr XII models, respectively is simply a reparameterisation of the model. For $k \geq 3$ stress levels, there is a reduction of the number of model parameters.

As an alternative to these strictly parametric stress-scale relationships, Qiu and Tsokos (2000) presents a more generalised model. Effectively, it is a trade-off between the more rigorous parametric models and non-parametric ones, with the Eyring, Arrhenius and Power-Law being special cases. They propose relaxing the need to specify completely the function of the stress variable and this, in turn, provides a better interpolation to lower stress levels. However, since we have a strictly linear relationship between log of life and function of stress - which in (1.12) is simply the stress factor x_i and is consequently completely specified - we shall adopt (1.12) as the basis of our work.

1.7 Computational Issues

Algorithms for fitting models, which are outlined in later chapters and given explicitly in the appendix, were written using the statistical software package SAS, (and in particular the module IML, see, for example, SAS Inc., 1990) These algorithms usually involve nested procedures, requiring iteration within iteration. Consequently, to obtain 10000 replications to assess agreement with asymptotic results, each replication - using a single set of parameter values - took up to eight hours to run.

1.8 Outline of Future Chapters

In this chapter we illustrated the need for ALT modelling and the role it plays in our daily lives. We then reviewed the background theory necessary for analysing data arising from an accelerated lifetime experiment. Having determined the limiting relationship between

the Burr XII and Weibull models, we made the distinction between Type I and Type II censoring. To develop further the theory in later chapters, we then defined all necessary functions and concluded by highlighting the link between the non-accelerated and the accelerated frameworks. In the next chapter we outline previous work and consider using the Burr XII distribution in an accelerated framework for complete data sets only. Chapter three establishes the expected Fisher information matrices for the models in Chapter two. Chapter four considers an accelerated extension to a method for determining which of the Weibull or Burr XII models provides the better fit to a data set. All appropriate extensions to Type I censored data are the focus of Chapter five. Chapter six is then concerned with the design of experiments for both complete data and Type I censored data. Chapter seven presents summaries and conclusions.

Chapter 2

Fitting Reliability Distributions To Complete Data

Throughout, we are only concerned with estimating parameters by maximum likelihood methods; see Cox and Hinkley (1974) for a detailed discussion of this approach and of the main theoretical properties of these estimators (hereafter abbreviated to MLE.)

We begin this chapter by reviewing various aspects of fitting Weibull non-ALT and Weibull ALT models to complete data sets; thus, here we have $n = N$, the number of items available for inspection. We then present details for fitting the three parameter Burr XII distribution, as in (1.5), to data, see Watkins (1998a). Finally, we present a further extension to this work, and fit an accelerated Burr XII model to data using an algorithm detailed by Watkins and Johnson (1999). In addition to theoretical results, we outline the associated computational algorithms and illustrate various stages with reference to published data sets. We conclude by comparing the fitted Weibull and Burr XII models, and consider the difference in the maximised log-likelihoods for given data sets.

2.1 Fitting Weibull Models To Data

We first outline the procedures for fitting models based on the Weibull distribution.

2.1.1 Weibull Non-ALT

From (1.2), the likelihood of observing n independent failures $d_1, \ldots, d_n$ is

$$L_w(B, \phi) = \prod_{i=1}^{n} g(d_i; B, \phi) = \prod_{i=1}^{n} \frac{B d_i^{B-1}}{\phi^B} \exp\left\{ -\left(\frac{d_i}{\phi}\right)^B \right\},$$

from which the log-likelihood is

$$l_w(B, \phi) = n \ln B + (B-1) \sum_{i=1}^{n} \ln d_i - Bn \ln \phi - \phi^{-B} \sum_{i=1}^{n} d_i^B. \tag{2.1}$$

It is straightforward to show that the first order partial derivatives of l_w are

$$\begin{aligned}\frac{\partial l_w}{\partial B} &= nB^{-1} - n\ln\phi + \sum_{i=1}^{n}\ln d_i - \phi^{-B}\sum_{i=1}^{n} d_i^B \ln d_i + \phi^{-B}\ln\phi\sum_{i=1}^{n} d_i^B \\ \frac{\partial l_w}{\partial \phi} &= -nB\phi^{-1} + B\phi^{-B-1}\sum_{i=1}^{n} d_i^B,\end{aligned}$$

and it is well known that the likelihood equations based on these derivatives have no analytical solution. However, we note that we can solve $\frac{\partial l_w}{\partial \phi} = 0$ for ϕ in terms of B. This gives

$$\left(\frac{1}{n}\sum_{i=1}^{n} d_i^B\right)^{\frac{1}{B}} \tag{2.2}$$

as the estimator of ϕ for known B; we can now substitute this in l_w to obtain the *profile log-likelihood*

$$l_w^*(B) = n\ln B - n\ln\left(\sum_{i=1}^{n} d_i^B\right) + (B-1)\sum_{i=1}^{n}\ln d_i.$$

We still need to find the estimate $\widehat{B}$ of B numerically, and much evidence points to the Newton-Raphson iterative scheme as an effective method for convergence from a suitable starting value. Farnum and Booth (1997) propose a technique for finding suitable starting values for this iterative scheme. We can then find $\widehat{\phi}$ by setting B equal to $\widehat{B}$ in (2.2).

2.1.1.1 Example: Carbon Fibre Rod Data

We consider the complete data set presented in Table 2.1, consisting of failure stresses (in GPa) of single carbon fibre rods at four different lengths; see Crowder, Kimber, Smith and Sweeting (1991). So, for this example, our failures are measured in units of pressure (GPa), while the stress factor is measured in units of length (mm). Using the method of maximum likelihood to fit a Weibull distribution to each subset of data gives the results in Table 2.2. Having established the separate scale estimates we may now ask ourselves whether the log-linear link is justifiable. It appears from Figure 2.1 that it may be difficult to pass a straight line through all points; note that inference may be problematical when based on only four data points. Figures 2.2 to 2.5 indicate the quality of fit of the fitted model for each subset of data. In each figure, the continuous line is based on the fitted Weibull CDF, given by (1.1), while that based on the ECDF, indicated by $\times$, is obtained using (1.11), using the failure times given in Table 2.1. The steps required in using the Kaplan-Meier function are presented in Table 2.3 for early failures of the 1mm rod data. We plot $\ln[-\ln\{1-G(t)\}]$ against $\ln(t)$ since, the form of (1.1) implies that the plot should result in a straight line with slope B and intercept $-B\ln(\phi)$ on the $\ln(t)$ axis. The figures show that there is reasonable agreement between the observed lifetimes and the fitted distributions. For all but the lowest stress level, there is a tendency for observations to drift away from

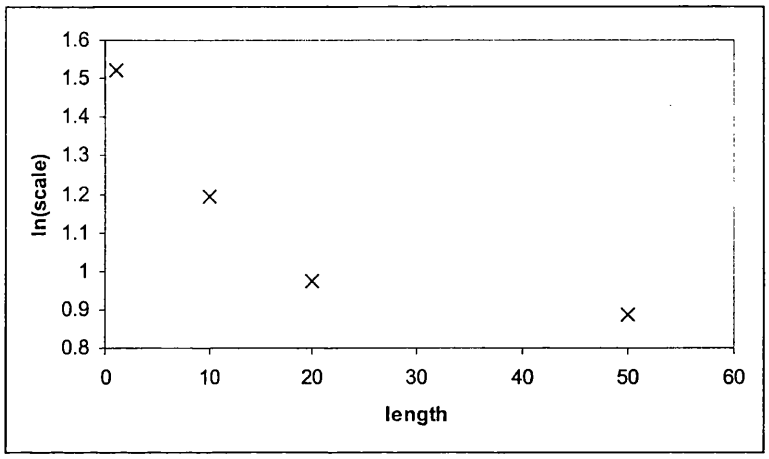

Figure 2.1: Log-linear plot illustrating relationship between log of scale estimates and stress for carbon fibre rod data.

their expected counterparts at earlier failure times. Taken together, in Figure 2.6, these plots illustrate a relatively constant shape parameter across fibre length, highlighted by reasonably consistent gradients. The results in Table 2.2 also indicate a decreasing scale parameter with increasing fibre length. This is perhaps not too surprising since a longer carbon fibre rod is likely to snap at a lower pressure.

2.1.1.2 Example: Aluminium Coupon Data

This example is taken from a data set presented in Owen and Padgett (2000) and represents the fatigue life (in cycles) to failure of coupons (rectangular strips) cut from aluminium sheeting. The data is presented in a scaled form and we based our analyses on this scaled version of the data set. The data set is presented in Table 2.4. Fitting a Weibull distribution to each subset in turn gives us the results in Table 2.5. Using the Kaplan-Meier function (1.11) and fitted CDF from (1.1), shown in Figures 2.7 to 2.10, we observe rather good agreement between observed and theoretical results for each stress level; the only exceptions being the extreme end points in each case.

2.1.1.3 Example: Electrode Data

This example is based on data from Nelson (1990,p232), given in Table 2.6, and concerns disc electrodes immersed in an insulating oil. A voltage was passed across the electrodes at three different but constant rates, and the voltage at oil breakdown was recorded; it is these values, measured in volts, that are then presented in Table 2.6. Fitting a Weibull distribution to each subset of data for the 9 square inch electrodes, we obtain the results given in Table 2.7. Using the Kaplan-Meier function (1.11) and fitted CDF (1.1), shown

1mm									
2.248	2.640	2.842	2.908	3.099	3.126	3.245	3.328	3.355	3.383
3.572	3.581	3.681	3.726	3.727	3.728	3.783	3.785	3.786	3.896
3.912	3.964	4.050	4.063	4.082	4.111	4.118	4.141	4.216	4.251
4.262	4.326	4.402	4.457	4.466	4.519	4.542	4.555	4.614	4.632
4.634	4.636	4.678	4.698	4.738	4.835	4.924	5.043	5.099	5.134
5.359	5.473	5.571	5.684	5.721	5.998	6.060			
10mm									
1.901	2.132	2.203	2.228	2.257	2.35	2.361	2.396	2.397	2.445
2.454	2.454	2.474	2.581	2.522	2.525	2.532	2.575	2.614	2.616
2.618	2.624	2.659	2.675	2.738	2.74	2.856	2.917	2.928	2.937
2.937	2.977	2.996	3.030	3.125	3.139	3.145	3.22	3.223	3.235
3.243	3.264	3.272	3.294	3.332	3.346	3.377	3.408	3.435	3.493
3.501	3.537	3.554	3.562	3.628	3.852	3.871	3.886	3.971	4.014
4.027	4.225	4.395	5.020						
20mm									
1.312	1.314	1.479	1.552	1.700	1.803	1.861	1.865	1.944	1.958
1.966	1.997	2.006	2.021	2.027	2.055	2.063	2.098	2.140	2.179
2.224	2.240	2.253	2.270	2.272	2.274	2.301	2.339	2.359	2.382
2.382	2.426	2.434	2.435	2.478	2.490	2.511	2.514	2.535	2.554
2.566	2.570	2.586	2.629	2.633	2.642	2.648	2.684	2.697	2.726
2.770	2.773	2.800	2.809	2.818	2.821	2.848	2.880	2.954	3.012
3.067	3.084	3.090	3.096	3.128	3.233	3.433	3.585	3.585	
50mm									
1.339	1.434	1.549	1.574	1.589	1.613	1.746	1.753	1.764	1.807
1.812	1.840	1.852	1.852	1.862	1.864	1.931	1.952	1.974	2.019
2.051	2.055	2.058	2.088	2.125	2.162	2.171	2.172	2.180	2.194
2.211	2.270	2.272	2.280	2.299	2.308	2.335	2.349	2.356	2.386
2.390	2.410	2.430	2.431	2.458	2.471	2.497	2.514	2.558	2.577
2.593	2.601	2.604	2.620	2.633	2.670	2.682	2.699	2.705	2.735
2.785	2.785	3.020	3.042	3.116	3.174				

Table 2.1: Failure stresses (in GPa) of single carbon fibre rods of varying lengths, taken from Crowder, Kimber, Smith and Sweeting (1991)

Length/mm	**1**	**10**	**20**	**50**
n	57	64	70	66
$\widehat{B}$	5.59297	5.02663	5.52444	6.03828
$\widehat{\phi}$	4.57525	3.30519	2.64819	2.42443
$l_w\left(\widehat{B},\widehat{\phi}\right)$	-71.02396	-62.9666	-49.92875	-36.16505

Table 2.2: Summaries of the MLEs obtained for the Weibull distribution for the four carbon fibre rod data subsets in Table 2.1

i	t_i	s_i	v_i	$\left(\frac{s_i-v_i}{s_i}\right)$	$M(t_i) = 1 - \prod_{j=1}^{h_i}\left(\frac{s_j-v_j}{s_j}\right)$
1	2.248	57	1	$\frac{57-1}{57} = \frac{56}{57}$	$1 - \frac{56}{57} = \frac{1}{57} = 0.0175\cdots$
2	2.640	56	1	$\frac{56-1}{56} = \frac{55}{56}$	$1 - \frac{55}{56} \times \frac{56}{57} = 1 - \frac{55}{57} = \frac{2}{57} = 0.0351\cdots$
3	2.842	55	1	$\frac{55-1}{55} = \frac{54}{55}$	$1 - \frac{54}{55} \times \cdots \times \frac{56}{57} = 1 - \frac{54}{57} = \frac{3}{57} = 0.0526$
4	2.908	54	1	$\frac{54-1}{54} = \frac{53}{54}$	$1 - \frac{53}{54} \times \cdots \times \frac{56}{57} = 1 - \frac{53}{57} = \frac{4}{57} = 0.0702$
5	3.099	53	1	$\frac{53-1}{53} = \frac{52}{53}$	$1 - \frac{52}{53} \times \cdots \times \frac{56}{57} = 1 - \frac{52}{57} = \frac{5}{57} = 0.0877$
6	3.126	52	1	$\frac{52-1}{52} = \frac{51}{52}$	$1 - \frac{51}{52} \times \cdots \times \frac{56}{57} = 1 - \frac{51}{57} = \frac{6}{57} = 0.1055$
7	3.245	51	1	$\frac{51-1}{51} = \frac{50}{51}$	$1 - \frac{50}{51} \times \cdots \times \frac{56}{57} = 1 - \frac{50}{57} = \frac{7}{57} = 0.1228$
8	3.328	50	1	$\frac{50-1}{50} = \frac{49}{50}$	$1 - \frac{49}{50} \times \cdots \times \frac{56}{57} = 1 - \frac{49}{57} = \frac{8}{57} = 0.1404$
...	...	...	...	...	...

Table 2.3: An example of calculations in the Kaplan-Meier function (1.11) for early failures in the 1mm carbon fibre rod data subset

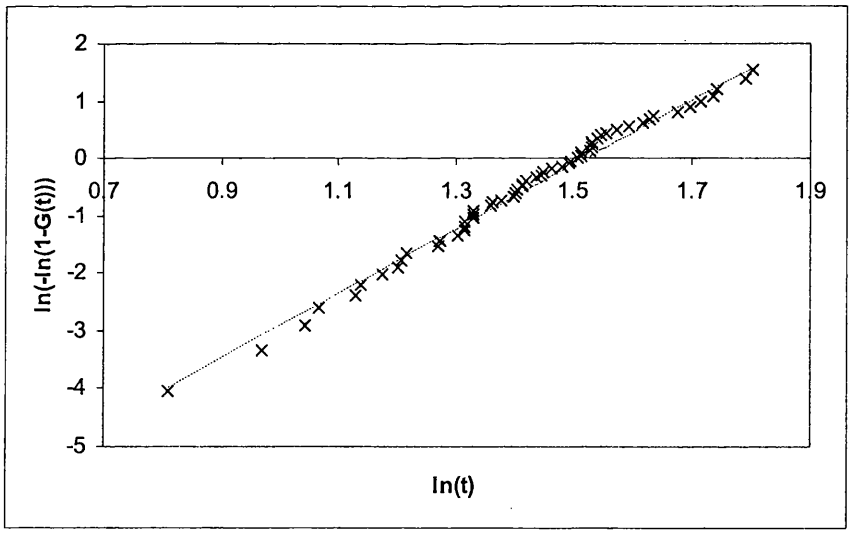

Figure 2.2: Weibull plot for 1mm carbon fibre rods. The points marked × are based on the ECDF while the continuous line is based on the theoretical Weibull CDF.

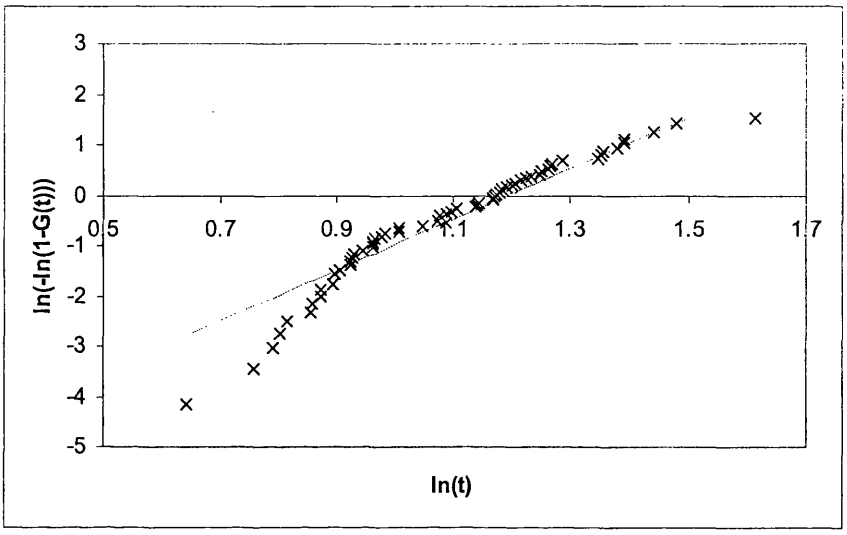

Figure 2.3: Weibull plot for 10mm carbon fibre rods.

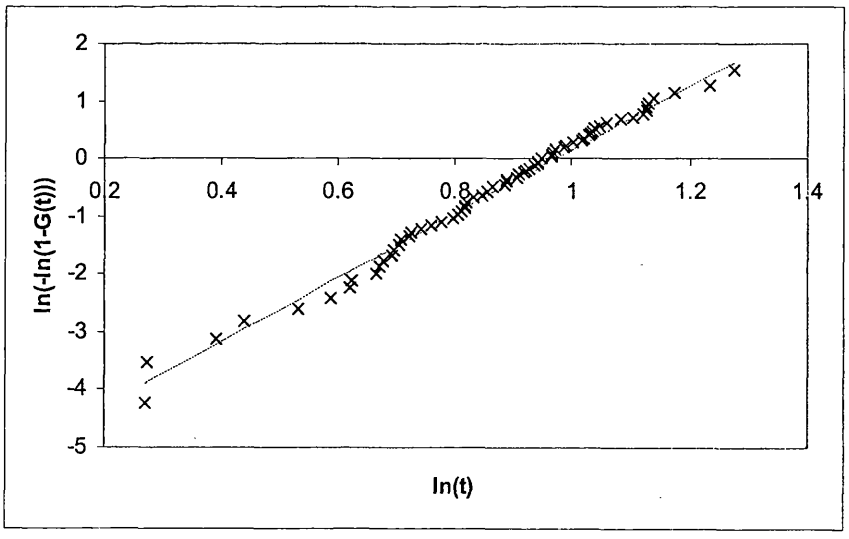

Figure 2.4: Weibull plot for 20mm carbon fibre rods.

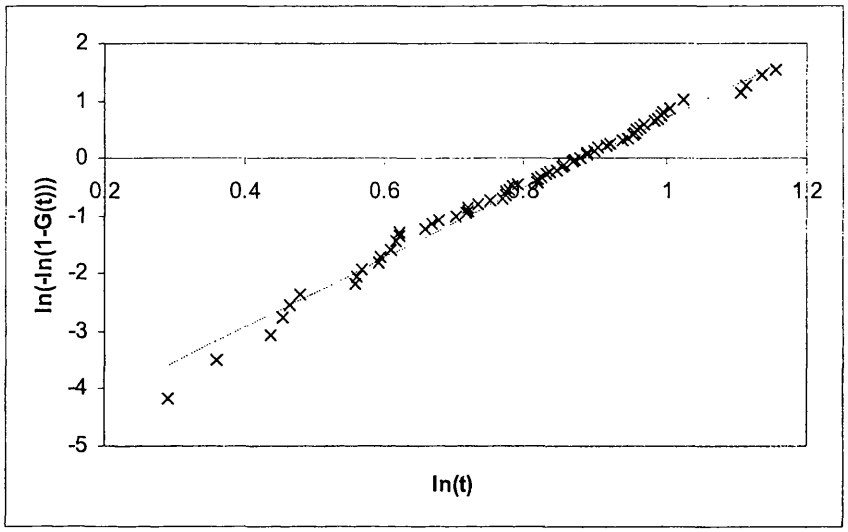

Figure 2.5: Weibull plot for 50mm carbon fibre rods.

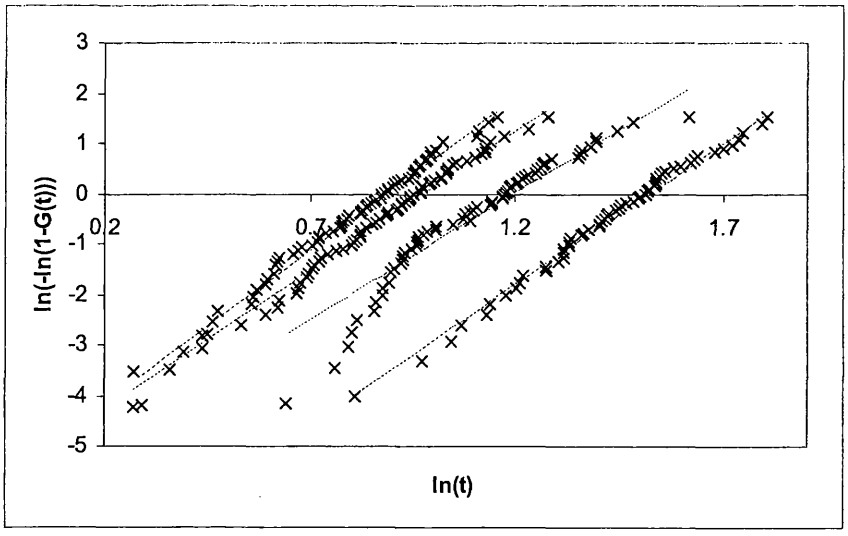

Figure 2.6: Combined Weibull plot for the four subsets of the carbon fibre rod data.

2.1 psi/cycle								
3.70	7.06	7.06	7.46	7.85	7.97	8.44	8.55	8.58
8.86	8.86	9.30	9.60	9.88	9.90	10.00	10.10	10.16
10.18	10.20	10.55	10.85	11.02	11.02	11.08	11.15	11.20
11.34	11.40	11.99	12.00	12.00	12.03	12.22	12.35	12.38
12.52	12.58	12.62	12.69	12.70	12.90	12.93	13.00	13.10
13.13	13.15	13.30	13.55	13.90	14.16	14.19	14.20	14.20
14.50	14.52	14.75	14.78	14.81	14.85	15.02	15.05	14.13
15.22	15.22	15.30	15.40	15.60	15.67	15.78	15.94	16.02
16.04	16.08	16.30	16.42	16.74	17.30	17.50	17.50	17.63
17.68	17.81	17.82	17.92	18.20	18.68	18.81	18.90	18.93
18.95	19.10	19.23	19.40	19.45	20.23	21.00	21.30	22.15
22.68	24.40							
2.6 psi/cycle								
2.33	2.58	2.68	2.76	2.90	3.10	3.12	3.15	3.18
3.21	3.21	3.29	3.35	3.36	3.38	3.38	3.42	3.42
3.42	3.44	3.49	3.50	3.50	3.51	3.51	3.52	3.52
3.56	3.58	3.58	3.60	3.62	3.63	3.66	3.67	3.70
3.70	3.72	3.72	3.74	3.75	3.76	3.79	3.79	3.80
3.82	3.89	3.89	3.95	3.96	4.00	4.00	4.00	4.03
4.04	4.06	4.08	4.08	4.10	4.12	4.14	4.16	4.16
4.16	4.20	4.22	4.23	4.26	4.28	4.32	4.32	4.33
4.33	4.37	4.38	4.39	4.39	4.43	4.45	4.45	4.52
4.56	4.56	4.60	4.64	4.66	4.68	4.70	4.70	4.73
4.74	4.76	4.76	4.86	4.88	4.89	4.901	4.91	5.03
5.17	5.40	5.60						
3.1 psi/cycle								
0.70	0.90	0.96	0.97	0.99	1.00	1.03	1.04	1.04
1.05	1.07	1.08	1.08	1.08	1.09	1.09	1.12	1.12
1.13	1.14	1.14	1.14	1.16	1.19	1.20	1.20	1.20
1.21	1.21	1.23	1.24	1.24	1.24	1.24	1.24	1.28
1.28	1.29	1.29	1.30	1.30	1.30	1.31	1.31	1.31
1.31	1.31	1.32	1.32	1.32	1.33	1.34	1.34	1.34
1.34	1.34	1.36	1.36	1.37	1.38	1.38	1.38	1.39
1.39	1.41	1.41	1.42	1.42	1.42	1.42	1.42	1.42
1.44	1.44	1.45	1.46	1.48	1.48	1.49	1.51	1.51
1.52	1.55	1.56	1.57	1.57	1.57	1.57	1.58	1.59
1.62	1.63	1.63	1.64	1.66	1.66	1.68	1.70	1.74
1.96	2.12							

Table 2.4: Fatigue life of coupons cut from aluminium sheeting, from Owen and Padgett (2000).

psi/cycle	2.1	2.6	3.1
n	101	102	101
$\widehat{B}$	4.0298862	7.0075353	6.0734031
$\widehat{\phi}$	15.426431	4.2437821	1.4316699
$l_w\left(\widehat{B},\widehat{\phi}\right)$	-279.4588	-98.07688	2.807636

Table 2.5: Summaries of the MLEs obtained for the Weibull distribution for the three aluminium coupon data subsets in Table 2.4

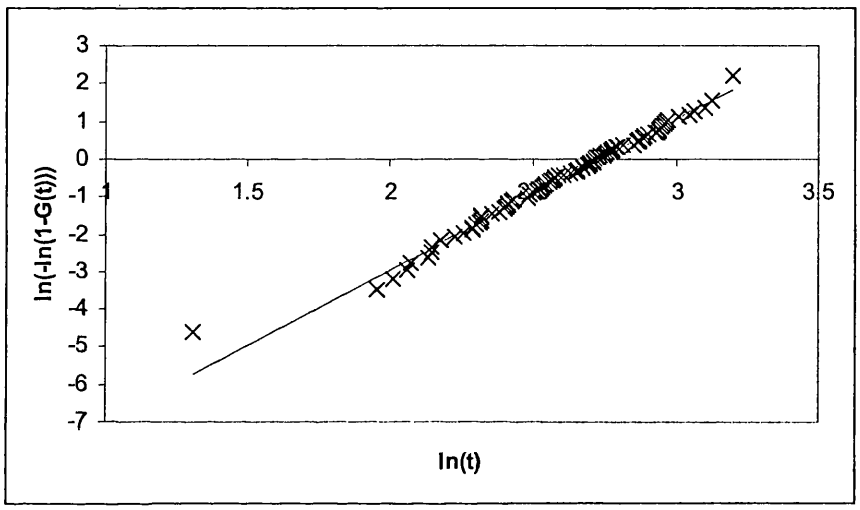

Figure 2.7: Weibull plot for 2.1 psi/cycle subset from aluminium coupon data. The points marked $\times$ are based on the ECDF while the continuous line is based on the theoretical Weibull CDF

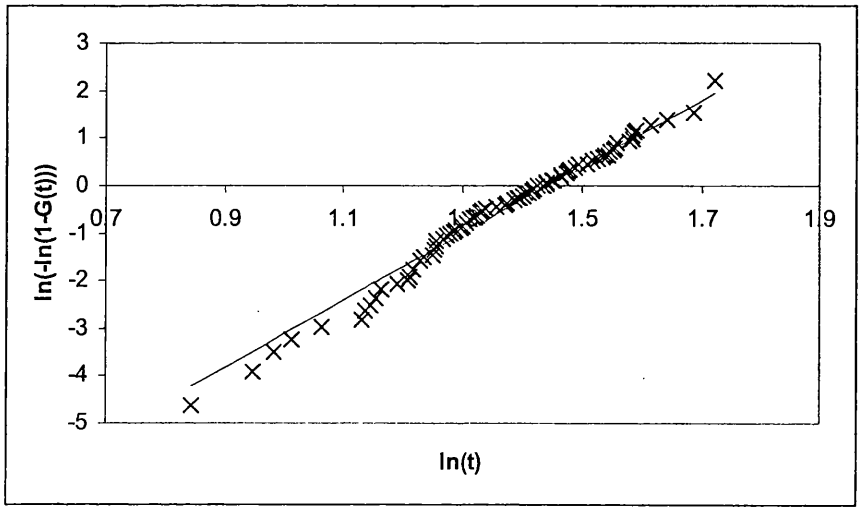

Figure 2.8: Weibull plot for 2.6 psi/cycle from aluminium coupon data.

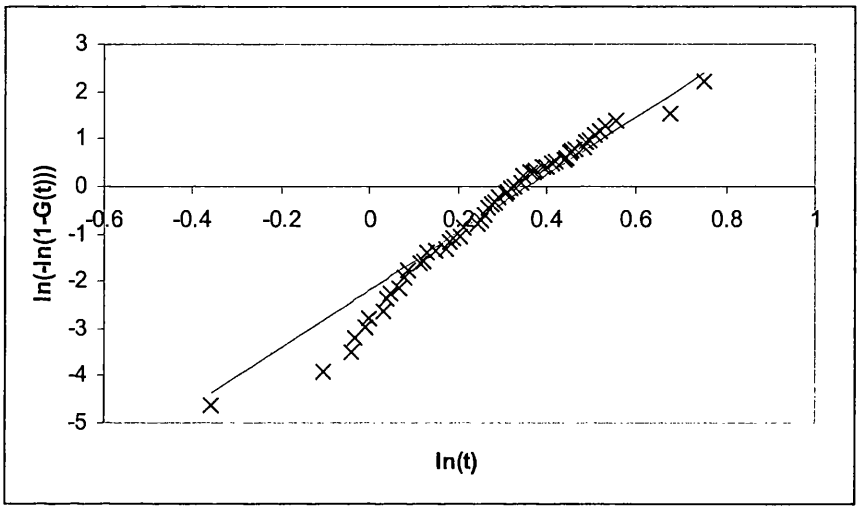

Figure 2.9: Weibull plot for 3.1 psi/cycle from aluminium coupon data.

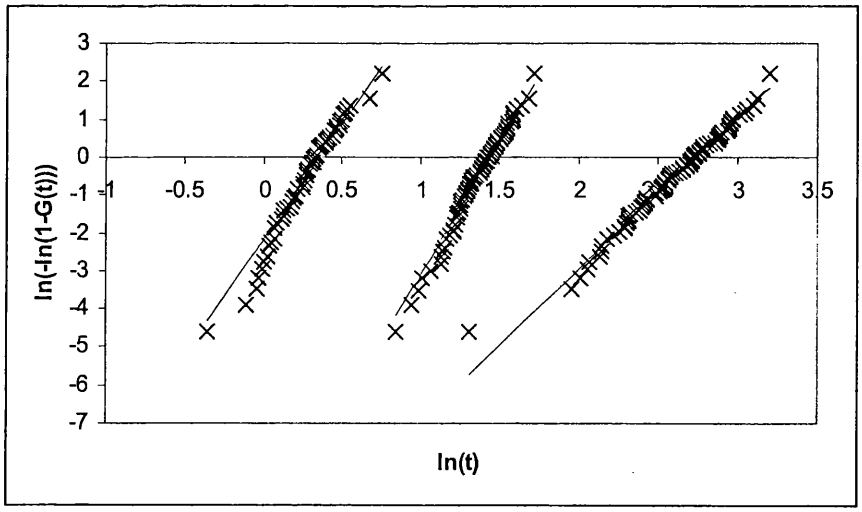

Figure 2.10: Combined Weibull plot for aluminium coupon data

10 volts/second									
33	37	38	38	38	37	27	42	39	38
42	32	42	40	32	38	36	42	20	37
43	40	38	43	39	41	35	41	40	32
38	40	37	29	31	41	38	36	35	40
37	41	36	39	43	42	43	43	41	44
37	43	38	40	40	38	33	40	35	41
100 volts/second									
43	42	45	48	38	44	37	44	43	42
43	49	44	45	50	44	44	45	41	48
45	48	43	49	50	45	45	46	47	42
47	48	47	48	39	49	44	47	34	41
45	48	44	47	45	50	40	47	47	43
49	45	45	45	47	39	44	37	47	48
1000 volts/second									
50	53	50	49	53	51	47	44	53	42
49	46	50	38	48	43	52	53	52	48
45	53	52	50	55	50	43	52	50	54
51	40	52	53	47	45	53	47	54	50
32	48	53	52	45	48	48	51	53	48
54	51	50	54	35	56	51	48	48	46

Table 2.6: Voltage (in volts) at oil breakdown data for disc electrodes immersed in oil, from Nelson (1990).

volts/second	**10**	**100**	**1000**
n	60	60	60
$\widehat{B}$	12.220089	16.450886	14.681733
$\widehat{\phi}$	39.695289	46.247051	50.860083
$l_w\left(\widehat{B},\widehat{\phi}\right)$	-165.5837	-155.4434	-169.9556

Table 2.7: Summaries of the MLEs obtained for the Weibull distribution for the three electrode data subsets in Table 2.6

in Figures 2.11 to 2.14, we again observe a fairly good agreement for middle to late failure times at each of the three stress levels. However, in all cases there is a tendency for the theoretical Weibull model to underestimate the observed times to failure in the tails. This may indicate the need to include a non-zero location parameter in the model; this example is therefore used for illustrative purposes only.

These examples use the SAS IML program *weibull*, given in Appendix A. To indicate the rate of convergence to the maximum likelihood estimates (MLEs) of the parameters, we give details in Table 2.8 of intermediate values of $\widehat{B}$ for the electrode data example. The value of $\widehat{\phi}$ was subsequently calculated using (2.2).

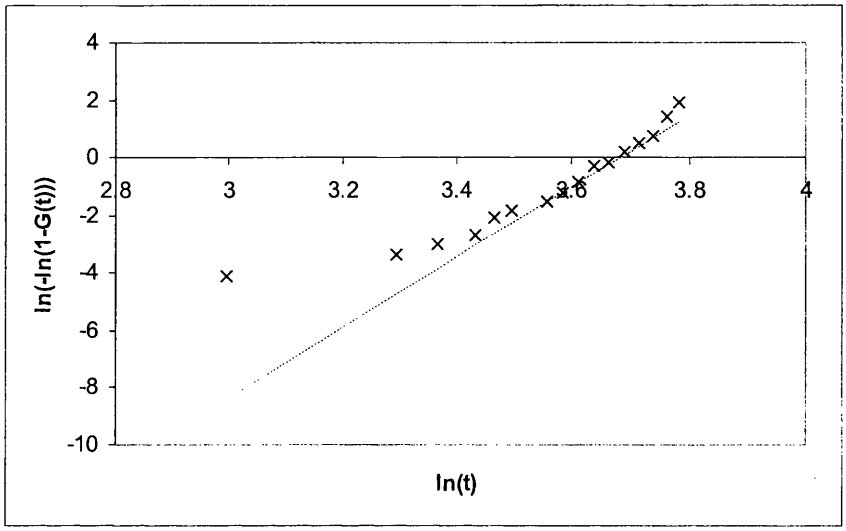

Figure 2.11: Weibull plot for 10 volts/second subset from voltage data. The points marked $\times$ are based on the ECDF while the continuous line is based on the theoretical Weibull CDF.

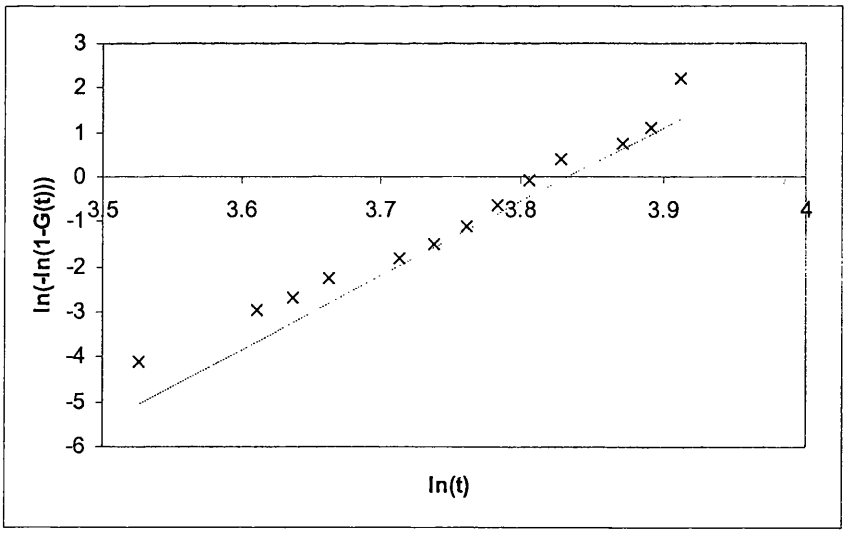

Figure 2.12: Weibull plot for 100 volts/second subset from electrode data.

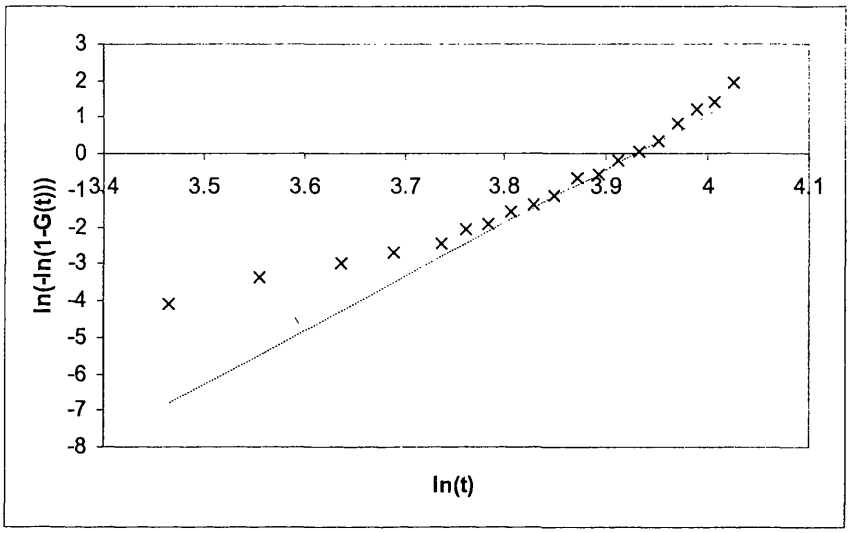

Figure 2.13: Weibull plot for 1000 volts/second subset from electrode data.

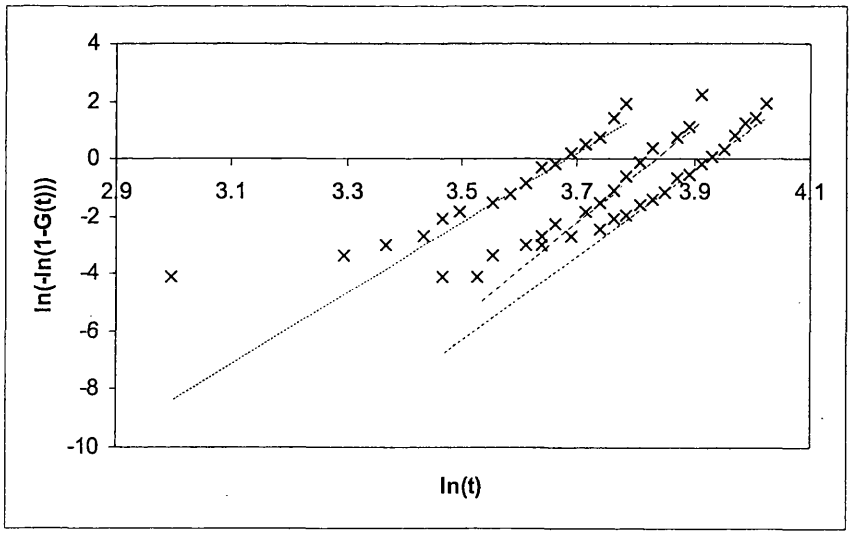

Figure 2.14: Combined Weibull plot for electrode data.

10 volts/second			
Iteration	**B**	$\frac{\partial l_w^*}{\partial B}$	$\frac{\partial^2 l_w^*}{\partial B^2}$
1	1	59.111533	−60.77302
2	1.9726608	28.85765	−16.03335
3	3.7725121	13.411621	−4.658881
4	6.6512339	5.4689624	−1.665878
5	9.9341654	1.6129986	−0.838521
6	11.85779	0.2211976	−0.625788
7	12.21126	0.0052628	−0.596401
8	12.220084	0.0000031	−0.595697
9	12.220089	1.052×10^{-12}	−0.595696
10	12.220089	0	−0.595696
100 volts/second			
Iteration	**B**	$\frac{\partial l_w^*}{\partial B}$	$\frac{\partial^2 l_w^*}{\partial B^2}$
1	1	59.627328	−60.35685
2	1.9879132	29.471656	−15.51103
3	3.8879588	14.144786	−4.24992
4	7.2162066	6.2024477	−1.371043
5	11.740096	2.1422993	−0.59927
6	15.314944	0.4187068	−0.390203
7	16.387993	0.0219737	−0.350441
8	16.450696	0.0000663	−0.348328
9	16.450886	6.079×10^{-10}	−0.348322
10	16.450886	-2.84×10^{-14}	−0.348322
1000 volts/second			
Iteration	**B**	$\frac{\partial l_w^*}{\partial B}$	$\frac{\partial^2 l_w^*}{\partial B^2}$
1	1	59.395644	−60.55007
2	1.9809344	29.189233	−15.75314
3	3.8338494	13.808445	−4.429014
4	6.9515742	5.8980792	−1.479318
5	10.9386	1.9574428	−0.670272
6	13.858972	0.3544481	−0.451654
7	14.643751	0.0156449	−0.412773
8	14.681653	0.0000329	−0.41104
9	14.681733	1.458×10^{-10}	−0.411037
10	14.681733	0	−0.400137

Table 2.8: Convergence to $\widehat{B}$ for subsets of electrode data in example 2.1.1.3. The derivatives of the profile log-likelihood for each subset show that $\widehat{B}$ is a local maximum.

2.1.2 Weibull ALT

Watkins (1994) summarises the techniques involved in fitting a Weibull ALT model to given data sets. There, the theory and examples accommodate censored items and the computational algorithm was implemented in FORTRAN. We shall summarise the salient issues below, noting that a subsequent implementation of the algorithm in SAS IML is included in Appendix A as *weibull_alt*. Using (1.12) to introduce the accelerating parameters α_w and β_w and adapting (2.1), we can write the log-likelihood for complete data obtained at the i^{th} stress level as

$$\begin{aligned} & n_i \ln B + (B-1) \sum_{j=1}^{n_i} \ln d_{ij} - B n_i \ln \phi_i - \phi_i^{-B} \sum_{j=1}^{n_i} d_{ij}^B \\ = \; & n_i \ln B + (B-1) \sum_{j=1}^{n_i} \ln d_{ij} - B n_i \ln \{\exp(\alpha_w + \beta_w x_i)\} - \exp(\alpha_w + \beta_w x_i)^{-B} \sum_{j=1}^{n_i} d_{ij}^B, \end{aligned}$$

where we now use the double subscript d_{ij} to indicate the observed time to failure of the j^{th} item at the i^{th} stress level, for $i = 1, \cdots, k$. Then, the log-likelihood for the complete data set is

$$l_W(B, \alpha_w, \beta_w) = S_n \ln B + (B-1) S_e - (\alpha_w S_n + \beta_w S_x) B - \exp(-\alpha_w B) S_{0,0} \qquad (2.3)$$

where we find it convenient to write

$$\begin{aligned} S_n &= \sum_{i=1}^{k} n_i \\ S_x &= \sum_{i=1}^{k} n_i x_i \\ S_e &= \sum_{i=1}^{k} \sum_{j=1}^{n_i} \ln d_{ij} \end{aligned}$$

and

$$S_{g,h}(B, \beta_w) = \sum_{i=1}^{k} \left\{ x_i^g \exp(-\beta_w B x_i) \left[\sum_{j=1}^{n_i} d_{ij}^B (\ln d_{ij})^h \right] \right\}.$$

Now, for fixed B and β_w, the maximising value of α_w in (2.3) is given by

$$B^{-1} \ln \left(\frac{S_{0,0}}{S_n} \right), \qquad (2.4)$$

and substituting this into (2.3) and ignoring terms independent of model parameters, gives the profile log-likelihood

$$l_W^*(B, \beta_w) = S_n \ln B + (B-1) S_e - S_n \ln S_{0,0}(B, \beta_w) - \beta_w B S_x.$$

Watkins (1994) establishes two useful recurrence relations regarding derivatives of $S_{g,h}$, leading to first and second order partial derivatives of l_W^*. The relations are

$$\begin{aligned}\frac{\partial S_{g,h}}{\partial B} &= S_{g,h+1} - \beta_w S_{g+1,h},\\ \frac{\partial S_{g,h}}{\partial \beta_w} &= -B S_{g+1,h},\end{aligned}$$

and hence we can obtain

$$\begin{aligned}\frac{\partial l_W^*}{\partial B} &= \frac{S_n}{B} + S_e - \frac{S_n}{S_{0,0}}\left(S_{0,1} - \beta_w S_{1,0}\right) - \beta_w S_x,\\ \frac{\partial l_W^*}{\partial \beta_w} &= S_n B \frac{S_{1,0}}{S_{0,0}} - B S_x\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2 l_W^*}{\partial B^2} &= -\frac{S_n}{B^2} - \frac{S_n}{S_{0,0}}\left(S_{0,2} - 2\beta_w S_{1,1} + \beta_w^2 S_{2,0}\right) + \frac{S_n}{S_{0,0}^2}\left(S_{0,1} - \beta_w S_{1,0}\right)^2,\\ \frac{\partial^2 l_W^*}{\partial B \partial \beta_w} &= \frac{S_n}{S_{0,0}}\left(B S_{1,1} + S_{1,0} - \beta_w B S_{2,0}\right) - \frac{S_n}{S_{0,0}^2} B\left(S_{0,1} - \beta_w S_{1,0}\right) S_{1,0} - S_x,\\ \frac{\partial^2 l_W^*}{\partial \beta_w^2} &= S_n B^2 \left(\frac{S_{1,0}^2}{S_{0,0}^2} - \frac{S_{2,0}}{S_{0,0}}\right).\end{aligned}$$

Example 2.1.1.1 [Carbon Fibre Rod Data] revisited

We now summarise the fitting of this Weibull ALT model to the carbon fibre rod failure data in Table 2.1; convergence to $\widehat{B}$ and $\widehat{\beta}_w$ is given in Table 2.9. The initial starting values for $\widehat{B}$ and $\widehat{\beta}_w$ were 5.0 and -0.01, respectively. The starting value for $\widehat{B}$ is based on the separate estimates of $\widehat{B}$ in Table 2.2; using the estimates of the separate $\widehat{\phi}$ in Table 2.2, the starting value for $\widehat{\beta}_w$ is established through (1.12) as

$$\frac{\widehat{\phi}_4}{\widehat{\phi}_1} = \frac{\exp\left\{\widehat{\alpha}_w + \widehat{\beta}_w x_4\right\}}{\exp\left\{\widehat{\alpha}_w + \widehat{\beta}_w x_1\right\}} = \exp\left\{\widehat{\beta}_w\left(x_4 - x_1\right)\right\} = \exp\left(\widehat{\beta}_w\left(50 - 1\right)\right),$$

and then solving for $\widehat{\beta}_w$ giving approximately -0.01. The final maximum likelihood estimates are then $\widehat{B} = 4.5646885, \widehat{\alpha}_w = 1.3932556$, and $\widehat{\beta}_w = -0.011704$, and hence $\widehat{\phi}_1 = 3.9810743, \widehat{\phi}_2 = 3.5830576, \widehat{\phi}_3 = 3.1873105$ and $\widehat{\phi}_4 = 2.2435517$, with associated log-likelihood, $l_W = -268.6926$. For reassurance, we check the second derivatives. For the final iteration, they were observed to be $\frac{\partial^2 l_W^*}{\partial B^2} = -22.19443$, $\frac{\partial^2 l_W^*}{\partial B \partial \beta_w} = -389.9967$ and $\frac{\partial^2 l_W^*}{\partial \beta_w^2} = -2453561$, and for the parameter estimates to be maximums, we also require the following conditions to be satisfied:

$$\frac{\partial^2 l_W^*}{\partial B^2} < 0, \tag{2.5}$$

Iteration	$\widehat{B}$	$\widehat{\beta}_w$	$\frac{\partial l_W^*}{\partial B}$	$\frac{\partial l_W^*}{\partial \beta_w}$
1	5.0	−0.01	−11.80104	−5024.646
2	4.6188314	−0.011528	−1.282296	−462.0631
3	4.5651463	−0.011700	−0.011645	−9.470263
4	4.5646888	−0.011704	−0.000008	−0.00121
5	4.5646885	−0.011704	2.913×10^{-13}	-2.0×10^{-10}
6	4.5646885	−0.011704	1.6×10^{-14}	-2.1×10^{-11}

Table 2.9: Convergence of $\widehat{B}$ and $\widehat{\beta}_w$ when fitting a Weibull ALT model to Carbon Fibre Rod data.

and

$$\frac{\partial^2 l_W^*}{\partial B^2} \times \frac{\partial^2 l_W^*}{\partial \beta_w^2} - \left(\frac{\partial^2 l_W^*}{\partial B \partial \beta_w}\right)^2 > 0; \tag{2.6}$$

see, for example, Weber (1982). Clearly, (2.5) is satisfied, while substituting results into (2.6), gives the positive value 5.4303×10^7, therefore confirming that the likelihood estimates correspond to a maximum of l_W^*. Comparison between these MLEs and those in Table 2.2 show similar values for the single $\widehat{B}$ here and the separate counterparts there. We see that the more parsimonious ALT model also provides good approximations to the separate scale parameters in Table 2.2.

Example 2.1.1.2 [Aluminium Coupon Data] revisited

Fitting a Weibull ALT model to the aluminium coupon data, in Table 2.4, using the starting values for $\widehat{B}$ and $\widehat{\beta}_w$ of 5.0 and −2.37495, we obtain the parameter estimates $\widehat{B} = 4.91691$, $\widehat{\alpha}_w = 7.795572$ and $\widehat{\beta}_w = -2.41544$; and hence $\widehat{\phi}_1 = 15.227945, \widehat{\phi}_2 = 4.5512789$ and $\widehat{\phi}_3 = 1.3602715$, with associated log-likelihood based on (2.3), $l_W = -415.5829$. As in the previous example, we make the necessary checks on second derivatives. For the final iteration, they were observed to be $\frac{\partial^2 l_W^*}{\partial B^2} = -23.02807$, $\frac{\partial^2 l_W^*}{\partial B \partial \beta_w} = -15.83327$ and $\frac{\partial^2 l_W^*}{\partial \beta_w^2} = -1454.296$, and so, with (2.5) and (2.6) satisfied, we conclude that these values maximise l_W^*. Convergence to $\widehat{B}$ and $\widehat{\beta}_w$ is shown in Table 2.10. Comparison between these results and those seen for individual subsets, in Table 2.5, shows an ALT value of $\widehat{B}$ of the same order of magnitude as the separate estimates of B. We also note that the ALT estimates of the $\widehat{\phi}_i$ agree with the separate estimates of the $\widehat{\phi}_i$.

Example 2.1.1.3 [Electrode Data] revisited

In fitting a Weibull ALT model to the 9 inch square electrode data, in Table 2.6, we obtain the parameter estimates $\widehat{B} = 11.280264$, $\widehat{\alpha}_w = 3.7627961$ and $\widehat{\beta}_w = 0.0001656$, and hence $\widehat{\phi}_1 = 43.140079, \widehat{\phi}_2 = 43.787981$ and $\widehat{\phi}_3 = 50.827037$, with associated log-likelihood based on (2.3), $l_W = -530.2297$. For the final iteration, the second derivatives were seen to be $\frac{\partial^2 l_W^*}{\partial B^2} = -2.241696$, $\frac{\partial^2 l_W^*}{\partial B \partial \beta_w} = -8380.925$ and $\frac{\partial^2 l_W^*}{\partial \beta_w^2} = -4.155 \times 10^9$; as in the previous two examples (2.5) is clearly satisfied, while substituting results into (2.6) returns the positive

Iteration	$\widehat{B}$	$\widehat{\beta}_w$	$\frac{\partial l_W^*}{\partial B}$	$\frac{\partial l_W^*}{\partial \beta_w}$
1	5.0	−2.37495	−3.104938	−62.28286
2	4.93986	−2.41481	−0.537156	−1.293024
3	4.91685	−2.41543	0.0011966	−0.011898
4	4.91691	−2.41544	-1.051×10^{-8}	3.1251×10^{-7}
5	4.91691	−2.41544	2.274×10^{-13}	1.364×10^{-12}
6	4.91691	−2.41544	-2.27×10^{-13}	-9.09×10^{-13}

Table 2.10: Convergence of $\widehat{B}$ and $\widehat{\beta}_w$.when fitting a Weibull ALT model to aluminium coupon data.

Iteration	$\widehat{B}$	$\widehat{\beta}_w$	$\frac{\partial l_W^*}{\partial B}$	$\frac{\partial l_W^*}{\partial \beta_w}$
1	11.28	0.0001	−2.587184	330221.72
2	12.064757	0.0001614	−1.633037	12699.373
3	11.239251	0.0001653	0.0946997	1607.3278
4	11.280169	0.0001656	0.0001917	−9.931507
5	11.280264	0.0001656	-2.69×10^{-10}	0.0002082
6	11.280264	0.0001656	2.487×10^{-13}	-3.49×10^{-10}
7	11.280264	0.0001656	-1.3×10^{-15}	-4.1×10^{-11}

Table 2.11: Convergence of $\widehat{B}$ and $\widehat{\beta}_w$ when fitting a Weibull ALT model to electrode data.

value 9.2440×10^9, and as before, we conclude that these values maximise l_W^*. The rate of convergence of $\widehat{B}$ and $\widehat{\beta}_w$ is shown in Table 2.11. Comparison with results in Table 2.7 shows that the single estimate of B here is of the same order of magnitude as the three separate estimates there. We also observe the ALT $\widehat{\phi}_i$ here to be similar to the individual estimates in Table 2.7, although the degree of acceleration observed in this example is rather limited.

2.2 Three Parameter Burr XII Model

In this section we consider the extension to the basic two parameter Burr XII distribution by way of introducing a scaling parameter θ. From (1.5), the likelihood for this model for complete data is given by

$$\prod_{i=1}^{n} \frac{a\tau\theta^{-\tau} d_i^{\tau-1}}{\left\{1+\left(\frac{d_i}{\theta}\right)^{\tau}\right\}^{a+1}}$$

for which the log-likelihood is

$$l_b(a,\tau,\theta) = n\ln(a\tau) - n\tau\ln\theta + (\tau-1)\sum_{i=1}^{n}\ln d_i - (a+1)\sum_{i=1}^{n}\ln\left\{1+\left(\frac{d_i}{\theta}\right)^{\tau}\right\}. \quad (2.7)$$

Watkins (1999) presents the associated first and second order partial derivatives as

$$\begin{aligned}
\frac{\partial l_b}{\partial a} &= na^{-1} - \sum_{i=1}^{n} \ln\left\{1 + \left(\frac{d_i}{\theta}\right)^{\tau}\right\} \\
\frac{\partial l_b}{\partial \tau} &= n\tau^{-1} - n\ln\theta + \sum_{i=1}^{n} \ln d_i - (a+1)\sum_{i=1}^{n} \frac{\left(\frac{d_i}{\theta}\right)^{\tau} \ln\left(\frac{d_i}{\theta}\right)}{1 + \left(\frac{d_i}{\theta}\right)^{\tau}} \\
\frac{\partial l_b}{\partial \theta} &= -n\tau\theta^{-1} + \tau(a+1)\theta^{-1} \sum_{i=1}^{n} \frac{\left(\frac{d_i}{\theta}\right)^{\tau}}{1 + \left(\frac{d_i}{\theta}\right)^{\tau}}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 l_b}{\partial a^2} &= -na^{-2} \\
\frac{\partial^2 l_b}{\partial \tau^2} &= -n\tau^{-2} - (a+1)\sum_{i=1}^{n} \frac{\left(\frac{d_i}{\theta}\right)^{\tau} \left\{\ln\left(\frac{d_i}{\theta}\right)\right\}^2}{\left\{1 + \left(\frac{d_i}{\theta}\right)^{\tau}\right\}^2} \\
\frac{\partial^2 l_b}{\partial \theta^2} &= n\tau\theta^{-2} - \tau(a+1)\theta^{-2} \sum_{i=1}^{n} \frac{\left(\frac{d_i}{\theta}\right)^{\tau}}{1 + \left(\frac{d_i}{\theta}\right)^{\tau}} - \tau^2(a+1)\theta^{-2} \sum_{i=1}^{n} \frac{\left(\frac{d_i}{\theta}\right)^{\tau}}{\left\{1 + \left(\frac{d_i}{\theta}\right)^{\tau}\right\}^2} \\
\frac{\partial^2 l_b}{\partial a \partial \tau} &= -\sum_{i=1}^{n} \frac{\left(\frac{d_i}{\theta}\right)^{\tau} \ln\left(\frac{d_i}{\theta}\right)}{1 + \left(\frac{d_i}{\theta}\right)^{\tau}} \\
\frac{\partial^2 l_b}{\partial a \partial \theta} &= \tau\theta^{-1} \sum_{i=1}^{n} \frac{\left(\frac{d_i}{\theta}\right)^{\tau}}{1 + \left(\frac{d_i}{\theta}\right)^{\tau}} \\
\frac{\partial^2 l_b}{\partial \tau \partial \theta} &= -n\theta^{-1} + (a+1)\theta^{-1} \sum_{i=1}^{n} \frac{\left(\frac{d_i}{\theta}\right)^{\tau}}{1 + \left(\frac{d_i}{\theta}\right)^{\tau}} + \tau(a+1)\theta^{-1} \sum_{i=1}^{n} \frac{\left(\frac{d_i}{\theta}\right)^{\tau} \log\left(\frac{d_i}{\theta}\right)}{\left\{1 + \left(\frac{d_i}{\theta}\right)^{\tau}\right\}^2}
\end{aligned}$$

We note that we can solve $\frac{\partial l_b}{\partial a} = 0$ for a in terms of θ and τ; this gives

$$\frac{n}{\sum_{i=1}^{n} \ln\left\{1 + \left(\frac{d_i}{\theta}\right)^{\tau}\right\}}. \tag{2.8}$$

From this, we then obtain the profile log likelihood as

$$\begin{aligned}
l_b^*(\tau, \theta) &= n\ln\tau + n\ln\left\{\frac{n}{\sum_{i=1}^{n} \ln\left\{1 + \left(\frac{d_i}{\theta}\right)^{\tau}\right\}}\right\} - n\tau\ln\theta + (\tau - 1)\sum_{i=1}^{n} \ln d_i \\
&\quad - \left(1 + \frac{n}{\sum_{i=1}^{n} \ln\left\{1 + \left(\frac{d_i}{\theta}\right)^{\tau}\right\}}\right) \sum_{i=1}^{n} \ln\left\{1 + \left(\frac{d_i}{\theta}\right)^{\tau}\right\}.
\end{aligned}$$

Since the likelihood equations above have no analytical solution, Watkins (1998a) provides a numerical algorithm; the SAS IML code is given in Appendix A as *burr3*. The approach, which utilises the Newton-Raphson algorithm, converges on estimates of a and τ for each iterated value of θ. We now outline the key steps in the algorithm, illustrating them on the 50mm carbon fibre rod length data subset above (see Table 2.1.) As programmed, the algorithm terminates when the absolute values of all first derivatives ($\frac{\partial l_b}{\partial a}$, $\frac{\partial l_b}{\partial \tau}$ and $\frac{\partial l_b}{\partial \theta}$) are less than 10^{-9}. This tolerance level is deemed sufficiently small to allow us to regard the current estimates as MLEs, and is used throughout. We note that alternative stopping criteria are available; for instance, we could use

$$\left| \begin{pmatrix} \frac{\partial l_b}{\partial a} \\ \frac{\partial l_b}{\partial \tau} \\ \frac{\partial l_b}{\partial \theta} \end{pmatrix}' \begin{pmatrix} \frac{\partial^2 l_b}{\partial a^2} & \frac{\partial^2 l_b}{\partial a \partial \tau} & \frac{\partial^2 l_b}{\partial a \partial \theta} \\ \frac{\partial^2 l_b}{\partial a \partial \tau} & \frac{\partial^2 l_b}{\partial \tau^2} & \frac{\partial^2 l_b}{\partial \tau \partial \theta} \\ \frac{\partial^2 l_b}{\partial a \partial \theta} & \frac{\partial^2 l_b}{\partial \tau \partial \theta} & \frac{\partial^2 l_b}{\partial \theta^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial l_b}{\partial a} \\ \frac{\partial l_b}{\partial \tau} \\ \frac{\partial l_b}{\partial \theta} \end{pmatrix} \right| < 10^{-9}.$$

Step 1 Fit the two parameter Weibull model (1.1) to the data set and obtain parameter estimates $\widehat{B}, \widehat{\phi}$. Using the method outlined above, we see that $\widehat{B}$ is found iteratively, while $\widehat{\phi}$ can be found using $\widehat{B}$ in (2.2).

Step 2 Scale the original data set by dividing throughout by $\widehat{\phi}$; thus re-fitting the Weibull model would then give the same value for $\widehat{B}$ and $\widehat{\phi} = 1$. More generally, the rescaled data values are now approximately equal to one.

Step 3 Set Burr XII scaling parameter to $\theta_1 = 1$ and obtain initial estimates of shape parameters a_1, τ_1 for the rescaled data. Here, τ_1 is found iteratively from a starting value of $\widehat{B}$, and we obtain a_1 by using θ_1 and τ_1 in (2.8).

Step 4 Evaluate first and second derivatives, together with the log-likelihood, given by (2.7), at a_1, τ_1 and θ_1. This is effectively the first stage in the iterative procedure. Using a_1 and τ_1, fit a three parameter Burr XII model to the rescaled data, and obtain a new estimate of θ, θ_2, using the Newton-Raphson iterative scheme and the full matrix of second derivatives, as described in Watkins (1999). Scale the rescaled data set by θ_2 and obtain new estimates of a and τ, denoted as a_2, τ_2. The starting value of τ here is τ_1, the final value of τ in the previous iteration.

Step 5 Repeat step 4 until convergence to $\widehat{\theta}$, $\widehat{a}$ and $\widehat{\tau}$. By convergence we mean until the first derivatives - $\frac{\partial l_b}{\partial a}$, $\frac{\partial l_b}{\partial \tau}$ and $\frac{\partial l_b}{\partial \theta}$ - were zero to at least eight decimal places and the principle minors of the matrix of second derivatives alternate in sign, starting with a negative; see, for example, Weber (1982).

Step 6 Undo the effect of the scaling at step 2, by multiplying $\widehat{\theta}$ by $\widehat{\phi}$ to obtain $\widehat{\theta}$ for the original unscaled data.

We see here the double nested nature of the algorithm, as mentioned in Chapter 1. For a given value of θ we iterate on τ, then obtain a new estimate of θ, iterate on τ again and so on.

Example 2.1.1.1 revisited (50mm carbon fibre rod subset)

We illustrate the above algorithm using this data set; see Table 2.1.

Step 1 From Table 2.2, we have $\widehat{B} = 6.03828$, with $\widehat{\phi} = 2.42443$.

Step 2 Scaling the original data set by $\widehat{\phi} = 2.42443$ gives the following re-scaled data (to three decimal places)

0.552	0.591	0.638	0.649	0.655	0.665	0.720	0.723	0.727	0.745
0.747	0.758	0.763	0.763	0.768	0.768	0.796	0.805	0.814	0.832
0.845	0.847	0.848	0.861	0.876	0.891	0.895	0.895	0.899	0.904
0.911	0.936	0.937	0.940	0.948	0.951	0.963	0.968	0.971	0.984
0.985	0.994	1.002	1.002	1.013	1.019	1.029	1.036	1.055	1.062
1.069	1.072	1.074	1.080	1.086	1.101	1.106	1.113	1.115	1.128
1.148	1.148	1.245	1.254	1.285	1.309				

The rescaled data lies in the range $[0.552, 1.309]$, with mean 0.928 ($\simeq 1$) and standard deviation 0.1723.

Step 3 Initially - with $\theta_1 = 1$ - the intermediate values of τ_1 (starting at $\widehat{B}$) and the corresponding first derivative were observed to be

τ_1	7.59147	8.03676	8.06345	8.06353	8.06353	8.06353
$\frac{\partial l_b^*}{\partial \tau}$	4.11433	0.77411	0.04153	0.00013	1.443×10^{-9}	-1.11×10^{-15}

Thus, these iterations converge to $\tau_1 = 8.0635$, at which $\frac{\partial l_b^*}{\partial \tau} = -1.11 \times 10^{-15}$, (and so is sufficiently close to zero); the final value of $\frac{\partial^2 l_b^*}{\partial \tau^2}$ is -1.546109. Using $\theta_1 = 1, \tau_1 = 8.0635$ in (2.8) gives $a_1 = 1.6449003$.

Step 4 (1^{st}run) The relevant first and second derivatives of l_b, evaluated at a_1, τ_1 and θ_1, are $\frac{\partial l_b}{\partial a} = 0$, $\frac{\partial^2 l_b}{\partial a^2} = -24.3929$, $\frac{\partial l_b}{\partial \theta} = 9.6653$ and $\frac{\partial^2 l_b}{\partial \theta^2} = -1941.563$. The new value of θ is then found via the full matrix of second derivatives. So we evaluate

$$\begin{pmatrix} a_2 \\ \tau_2 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ \tau_1 \\ \theta_1 \end{pmatrix} - \begin{pmatrix} \frac{\partial^2 l_b}{\partial a^2} & & \\ \frac{\partial^2 l_b}{\partial a \partial \tau} & \frac{\partial^2 l_b}{\partial \tau^2} & \\ \frac{\partial^2 l_b}{\partial a \partial \theta} & \frac{\partial^2 l_b}{\partial \tau \partial \theta} & \frac{\partial^2 l_b}{\partial \theta^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial l_b}{\partial a} \\ \frac{\partial l_b}{\partial \tau} \\ \frac{\partial l_b}{\partial \theta} \end{pmatrix}$$

$$= \begin{pmatrix} 1.6449 \\ 8.0635 \\ 1 \end{pmatrix} - \begin{pmatrix} -24.392 & & \\ -0.813 & -1.573 & \\ 204.869 & -6.674 & -1941.56 \end{pmatrix}^{-1} \times \begin{pmatrix} 0 \\ -1.33 \times 10^{-15} \\ 9.665 \end{pmatrix}$$

and extract the third element, giving us θ_2 as the new estimate of θ. This turns out to be

$$\theta_2 = 1.0 - (-0.0939025) = 1.0939025$$

Now, divide the 'Weibull-scaled' data set by θ_2 and determine a_2, τ_2. The intermediate values of τ_2 (starting at $\tau_1 = 8.0635$) and associated derivatives were observed to be

τ_2	7.3428	7.3945	7.3949	7.3949	7.3949
$\frac{\partial l_b^*}{\partial \tau}$	−1.3190	0.11123	0.00066	2.37×10^{-8}	-8.88×10^{-16}

So, these iterations converge to $\tau_2 = 7.3949$, at which $\frac{\partial l_b^*}{\partial \tau} = -8.88 \times 10^{-16}$ (and is sufficiently close to zero); the final value of $\frac{\partial^2 l_b^*}{\partial \tau^2}$ is −2.125441. Using $\theta_2 = 1.0939025, \tau_2 = 7.3949$ in (2.8) gives $a_2 = 2.6318855$. The relevant first and second derivatives of l_b, evaluated at a_2, τ_2 and θ_2 are $\frac{\partial l_b}{\partial a} = 0, \frac{\partial^2 l_b}{\partial a^2} = -9.52818, \frac{\partial l_b}{\partial \theta} = 3.17956$ and $\frac{\partial^2 l_b}{\partial \theta^2} = -1718.541$. This process is then repeated, with the first ten iterations presented below.

Iteration	a_i	τ_i	θ_i
1	1.6449	8.0635	1
2	2.6318	7.3949	1.0939
3	3.6975	7.0331	1.1683
4	4.5416	6.8569	1.2159
5	4.8978	6.7994	1.2338
6	4.9428	6.7927	1.2360
7	4.9434	6.7926	1.2360
8	4.9434	6.7926	1.2360
9	4.9434	6.7926	1.2360
10	4.9434	6.7926	1.2360

Step 5 For the 10^{th} iteration of step 4 - where θ_{11} was found to be 1.2360724 - the intermediate values of τ_{11} and the associated derivatives were observed to be

τ_{11}	6.7926	6.79262	6.79262	6.79262	6.79262
$\frac{\partial l_b^*}{\partial \tau}$	5.329×10^{-15}	1.776×10^{-15}	-3.55×10^{-15}	1.776×10^{-15}	-3.55×10^{-15}

So, these iterations quickly converge to $\tau_{11} = 6.79262$, at which $\frac{\partial l_b^*}{\partial \tau} = -3.55 \times 10^{-15}$ (and so is sufficiently close to zero).

Using $\theta_{11} = 1.2360724, a_{11} = 4.9434832$ in (2.8) gives $a_{12} = 4.9434832$, $\frac{\partial l_b}{\partial a} = -1.78 \times 10^{-15}$ and $\frac{\partial l_b}{\partial \theta} = 0$, and, since all first derivatives are practically zero, the new value of θ (θ_{12}) will also be unchanged from the previous iteration. We therefore regard the iterative process as complete and take $\widehat{a} = a_{11} = 4.9434832$, $\widehat{\tau} = \tau_{11} = 6.79262$ and $\widehat{\theta} = \theta_{11} = 1.2360724$ to be the assumed MLEs of the 'Weibull-scaled' data. To ensure that these estimates are true maximums, we calculate the principal minors of the matrix of second derivatives

$$\begin{pmatrix} \frac{\partial^2 l_b}{\partial a^2} & & \\ \frac{\partial^2 l_b}{\partial a \partial \tau} & \frac{\partial^2 l_b}{\partial \tau^2} & \\ \frac{\partial^2 l_b}{\partial a \partial \theta} & \frac{\partial^2 l_b}{\partial \tau \partial \theta} & \frac{\partial^2 l_b}{\partial \theta^2} \end{pmatrix} = \begin{pmatrix} -2.70078 & & \\ 1.73364 & -3.90630 & \\ 61.02344 & -47.69785 & -1414.859 \end{pmatrix}.$$

The first principal minor is given by $\frac{\partial^2 l_b}{\partial a^2}$, which is indeed negative. The second is given by the determinant,

$$\begin{vmatrix} \frac{\partial^2 l_b}{\partial a^2} & \\ \frac{\partial^2 l_b}{\partial a \partial \tau} & \frac{\partial^2 l_b}{\partial \tau^2} \end{vmatrix} = (-2.70078) \times (-3.90630) - (1.73364)^2$$
$$= 7.54427,$$

and is positive, while the third principal minor is given by the determinant of the full 3×3 matrix, and is found to be -69.3623, and therefore negative. Since we have alternating signs for the principle minors (starting with a negative) we conclude that the assumed MLEs above, are indeed maximums. This method is, of course, the natural extension of the two-parameter case which was used above when fitting the Weibull ALT model.

Step 6 Finally, we calculate $\widehat{\theta}$ for the original unscaled data; this is given by the product of $\widehat{\phi}$ ($= 2.42443$) and θ_{11} in previous step ($= 1.2360724$). Therefore, we have

$$\begin{aligned} \widehat{\theta} &= 2.996771 \\ \widehat{a} &= 4.943483 \\ \widehat{\tau} &= 6.792626 \end{aligned}$$

as the MLEs for the original data. These results are included in Table 2.12, which also contains the value of the maximised log-likelihood for this subset.

Example 2.1.1.1 [Carbon Fibre Rod Data] revisited

Fitting the Burr XII three parameter model to the individual subsets of the carbon fibre rod data in Table 2.1 produces the results given in Table 2.12, where we have also included the maximised Weibull log-likelihood for comparison. We see from this table that, on the basis of these maximised likelihoods, the Burr XII model is a better fit to each data subset

Length/mm	1	10	20	50
n	57	64	70	66
$\widehat{a}$	2.588745	0.9826893	3.3372807	4.943483
$\widehat{\tau}$	6.963570	8.6832448	6.6199838	6.792626
$\widehat{\theta}$	5.014342	2.972029	3.064437	2.996771
l_b	−70.30808	−58.61848	−49.13677	−35.83991
l_w	−71.02396	−62.9666	−49.92875	−36.16505

Table 2.12: Burr XII fit to carbon fibre rod data subsets

psi/cycle	2.1	2.6	3.1
n	101	102	101
$\widehat{a}$	19.235884	2.831573	1.5708883
$\widehat{\tau}$	4.1505101	8.615969	9.3872102
$\widehat{\theta}$	31.15297	4.632881	1.4174932
l_b	−279.4295	−96.59839	10.014003
l_w	−279.4588	−98.07688	2.807636

Table 2.13: Burr XII fit to aluminium coupon data subsets

than the Weibull model. We also note the relatively consistent value of $\widehat{\tau}$ across subsets and the expected decrease in $\widehat{\theta}$ as fibre length increases. We consider these results in greater detail later in this chapter.

Example 2.1.1.2 [Aluminium Coupon Data] revisited

Fitting the Burr XII three parameter model to the individual subsets of the aluminium coupon data (Table 2.4) gives the parameter estimates shown in Table 2.13. The associated first derivatives - $\frac{\partial l_b}{\partial a}$, $\frac{\partial l_b}{\partial \tau}$ and $\frac{\partial l_b}{\partial \theta}$ - were all below the tolerance level of -1.0×10^{-9} (stated earlier) for the final iteration at each stress level. As in the previous example, we include the maximised Weibull log-likelihood for comparison, and observe that the Burr XII is preferred to the Weibull distribution at all stress levels. We shall return to these results in greater detail later.

Example 2.1.1.3 [Electrode Data] revisited

Attempting to fit the Burr XII three parameter model to the electrode data (Table 2.6) leads to the parameter values shown in Table 2.14. In this case, the method fails to converge; estimates of a for each subset steadily rose. This was also the case with θ; the values presented are the last values observed when the algorithm completed 500 iterations. Again, we include the maximised Weibull log-likelihood, and using this as a basis for comparison, conclude that the Weibull model is a better fit to each data subset than the Burr XII model.

We shall make further comments on this phenomenon and comparisons with corresponding Weibull results in section 2.4 below.

volts/second	10	100	1000
n	60	60	60
a	253895932	1.31×10^{10}	3.48×10^{9}
τ	12.220089	16.450873	14.681725
θ	193.47247	190.70582	227.82527
l_b	−165.584	−155.4438	−170.0088
l_w	−165.5837	−155.4434	−169.9556

Table 2.14: Burr XII fit to electrode data subsets

2.3 Burr XII ALT Model

A popular choice of model in the field of accelerated testing has been the Weibull model, featuring a single shape parameter - B - and two accelerating parameters - α_w, β_w. We now consider the logical extension of fitting a Burr XII model to an accelerated data set, as outlined by Watkins and Johnson (1999). This will require specifying scale parameters θ_i in terms of the accelerating parameters α_b and β_b using (1.12). We begin by considering the log-likelihood and its first and second order partial derivatives, and then present an algorithm to determine the maximum likelihood estimates of the parameters. Finally, we illustrate the approach through examples.

2.3.1 The Model

Following our approach with the Weibull distribution, we now adopt the log-linear model (1.12), together with (1.5) to give us the Burr XII ALT CDF as,

$$F(d; a, \tau, \alpha_b, \beta_b) = 1 - \left\{1 + \left(\frac{d}{\exp(\alpha_b + \beta_b x_i)}\right)^{\tau}\right\}^{-a}, \tag{2.9}$$

from which the likelihood of observing a complete data set at stress x_i is then

$$\prod_{j=1}^{n_i} \frac{a\tau \left\{\exp(\alpha_b + \beta_b x_i)\right\}^{-\tau} d_{ij}^{\tau-1}}{\left[1 + \left\{\frac{d_{ij}}{\exp(\alpha_b+\beta_b x_i)}\right\}^{\tau}\right]^{a+1}}$$

for which the log-likelihood is

$$n_i \ln(a\tau) + (\tau - 1)\sum_{j=1}^{n_i} \ln d_{ij} - \tau n_i (\alpha_b + \beta_b x_i) - (a+1)\sum_{j=1}^{n_i} \ln\left\{1 + \left(\frac{d_{ij}}{\exp(\alpha_b + \beta_b x_i)}\right)^{\tau}\right\}.$$

It then follows that the complete log-likelihood, based on all k subsets of data, is

$$l_B(a, \tau, \alpha_b, \beta_b) = S_n \ln(a\tau) + (\tau - 1) S_e - \tau(\alpha_b S_n + \beta_b S_x) - (a+1) S_d^* \tag{2.10}$$

where S_n, S_x and S_e are as before, and

$$S_d^* = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \ln\left[1+\left\{\frac{d_{ij}}{\exp\left(\alpha_b+\beta_b x_i\right)}\right\}^{\tau}\right]. \tag{2.11}$$

In considering the derivatives of l_B, we will also find it convenient to write

$$S_{xx} = \sum_{i=1}^{k} n_i x_i^2,$$

and

$$S_{fghl} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_i^f \frac{\left\{\frac{d_{ij}}{\exp(\alpha_b+\beta_b x_i)}\right\}^{g\tau}\left[\ln\left\{\frac{d_{ij}}{\exp(\alpha_b+\beta_b x_i)}\right\}\right]^h}{\left(1+\left\{\frac{d_{ij}}{\exp(\alpha_b+\beta_b x_i)}\right\}^{\tau}\right)^l}. \tag{2.12}$$

At this stage we can give some useful definitions and recurrence relations. We have

$$\begin{aligned}
\frac{\partial S_d^*}{\partial \tau} &= S_{0111} \\
\frac{\partial S_d^*}{\partial \alpha_b} &= -\tau S_{0101} \\
\frac{\partial S_d^*}{\partial \beta_b} &= -\tau S_{1101},
\end{aligned}$$

while, more generally, we have

$$\begin{aligned}
\frac{\partial S_{fghl}}{\partial \tau} &= g \times S_{fg(h+1)l} - l \times S_{f(g+1)(h+1)(l+1)} \\
\frac{\partial S_{fghl}}{\partial \alpha_b} &= -h \times S_{fg(h-1)l} - g\tau \times S_{fghl} + l\tau S_{f(g+1)h(l+1)} \\
\frac{\partial S_{fghl}}{\partial \beta_b} &= -h \times S_{(f+1)g(h-1)l} - g\tau \times S_{(f+1)ghl} + l\tau \times S_{(f+1)(g+1)h(l+1)}
\end{aligned}$$

2.3.1.1 First and Second Order Partial Derivatives of l_B

These follow from (2.10) and the definitions above. We have

$$\frac{\partial l_B}{\partial a} = S_n a^{-1} - S_d^*, \tag{2.13}$$

$$\frac{\partial l_B}{\partial \tau} = S_n \tau^{-1} + S_e - \alpha_b S_n - \beta_b S_x + S_e - (a+1)\, S_{0111}, \tag{2.14}$$

$$\frac{\partial l_B}{\partial \alpha_b} = -S_n \tau + (a+1)\, \tau S_{0101}, \tag{2.15}$$

$$\frac{\partial l_B}{\partial \beta_b} = -S_x \tau + (a+1)\, \tau S_{1101}, \tag{2.16}$$

while the second derivatives are defined by

$$\frac{\partial^2 l_B}{\partial a^2} = -S_n a^{-2} \tag{2.17}$$

$$\frac{\partial^2 l_B}{\partial a \partial \tau} = -S_{0111} \tag{2.18}$$

$$\frac{\partial^2 l_B}{\partial a \partial \alpha_b} = \tau \times S_{0101} \tag{2.19}$$

$$\frac{\partial^2 l_B}{\partial a \partial \beta_b} = \tau \times S_{1101} \tag{2.20}$$

and

$$\begin{aligned}
\frac{\partial^2 l_B}{\partial \tau^2} &= -S_n \tau^{-2} - (a+1)(S_{0121} - S_{0222}) \\
\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b} &= -S_n + (a+1)(S_{0101} + \tau \times S_{0111} - \tau \times S_{0212}) \\
\frac{\partial^2 l_B}{\partial \tau \partial \beta_b} &= -S_x + (a+1)(S_{1101} + \tau \times S_{1111} - \tau \times S_{1212})
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 l_B}{\partial \alpha_b^2} &= (a+1)\tau^2 (S_{0202} - S_{0101}) \\
\frac{\partial^2 l_B}{\partial \alpha_b \partial \beta_b} &= (a+1)\tau^2 (S_{1202} - S_{1101})
\end{aligned}$$

and, finally,

$$\frac{\partial^2 l_B}{\partial \beta_b^2} = (a+1)\tau^2 (S_{2202} - S_{2101}).$$

2.3.2 Algorithm for fitting a Burr XII ALT model.

Watkins and Johnson (1999) extend previous work on the Burr XII non-ALT model, discussed above, to cover the Burr XII ALT model. The algorithm proceeds in a similar vein to that for the non-accelerated Burr XII model, in that we first divide the data subsets by the corresponding Weibull ALT scale estimates, ϕ_i. The advantage of this step is that we can largely eliminate the effect of the stress factor, thereby introducing a degree of numerical stability into the procedure. Having converged on the maximum likelihood estimates of the model parameters for rescaled data, we can then reverse the effect of the scaling to obtain parameter estimates for the unscaled data set. We now give a more detailed explanation of the algorithm.

Step 1 Fit a Weibull ALT model to the data set to obtain parameter estimates, $\widehat{B}, \widehat{\alpha}_w, \widehat{\beta}_w$.

Step 2 Use $\widehat{\alpha}_w, \widehat{\beta}_w$ to obtain separate estimates $\widehat{\phi}_i$ of the scale for each data subset, and then re-scale the full data set by dividing the i^{th} data subset by $\widehat{\phi}_i$.

Step 3 Fit a three parameter Burr XII distribution, (as detailed above), to the rescaled data. This is effectively fitting the full ALT model under the constraint $\beta_b = 0$, and leads to parameter estimates $\widehat{a}, \widehat{\tau}, \widehat{\theta}$.

Step 4 Fit the accelerated model using (1.12) and Burr XII distribution to re-scaled data. Here, we re-admit the information on the stress factor, x_i, but expect this covariate to have only a minor role due to the re-scaling. The Newton-Raphson technique is thus found to be fairly robust and to provide quick convergence on maximum likelihood estimates. Suitable starting values are $\widehat{a}$ and $\widehat{\tau}$ from step 3, together with $\widehat{\alpha}_b = \log \widehat{\theta}$ and $\widehat{\beta}_b = 0$.

Step 5 Finally, reverse the effect of the scaling at step 2; the estimates of the scale parameters in (1.12) are found by adding $\widehat{\alpha}_w$, $\widehat{\beta}_w$ to $\widehat{\alpha}_b, \widehat{\beta}_b$ respectively. The estimates of the two shape parameters, $\widehat{a}, \widehat{\tau}$, remain unchanged from step 4.

Example 2.1.1.1 [Carbon Fibre Rods] revisited

Step 1 Recall from Table 2.9 that the final parameter estimates were

$$\widehat{B} = 4.564688,\ \widehat{\alpha}_w = 1.39326,\ \widehat{\beta}_w = -0.01170.$$

Step 2 Using $\widehat{\alpha}_w$ and $\widehat{\beta}_w$ together with the stress levels - $1, 10, 20, 50$ - gives separate Weibull scale estimates as

$$\widehat{\phi}_1 = 3.9811, \widehat{\phi}_2 = 3.5832, \widehat{\phi}_3 = 3.1875, \widehat{\phi}_4 = 2.2440.$$

The data set resulting from dividing each original data subset by $\widehat{\phi}_i$ is given below; entries are given to three decimal places. We also present some summary statistics to highlight the precise effect of the scaling on each data subset.

	1mm	**10mm**	**20mm**	**50mm**
Mean	1.063	0.851	0.769	1.003
Standard Deviation	0.209	0.172	0.155	0.186
Minimum	0.564	0.530	0.411	0.596
Maximum	1.522	1.400	1.124	1.414

1mm									
0.564	0.663	0.713	0.730	0.778	0.785	0.815	0.835	0.842	0.849
0.897	0.899	0.924	0.935	0.936	0.936	0.950	0.950	0.950	0.978
0.982	0.995	1.017	1.020	1.025	1.032	1.034	1.040	1.059	1.067
1.070	1.086	1.105	1.119	1.121	1.135	1.140	1.144	1.158	1.163
1.163	1.164	1.175	1.180	1.190	1.214	1.236	1.266	1.280	1.289
1.346	1.374	1.399	1.427	1.437	1.506	1.522			
10mm									
0.530	0.594	0.614	0.621	0.629	0.655	0.658	0.668	0.668	0.682
0.684	0.684	0.690	0.720	0.703	0.704	0.706	0.718	0.729	0.730
0.730	0.732	0.742	0.746	0.764	0.764	0.797	0.814	0.817	0.819
0.819	0.830	0.836	0.845	0.872	0.876	0.877	0.898	0.899	0.902
0.905	0.910	0.913	0.919	0.929	0.933	0.942	0.951	0.958	0.974
0.977	0.987	0.991	0.994	1.012	1.070	1.080	1.084	1.093	1.120
1.123	1.176	1.226	1.400						

20mm									
0.411	0.412	0.464	0.486	0.533	0.565	0.583	0.585	0.609	0.614
0.616	0.626	0.629	0.634	0.635	0.644	0.647	0.658	0.671	0.683
0.697	0.702	0.706	0.712	0.712	0.713	0.721	0.733	0.740	0.747
0.747	0.761	0.763	0.763	0.777	0.781	0.787	0.788	0.795	0.801
0.805	0.806	0.811	0.824	0.826	0.828	0.830	0.842	0.846	0.855
0.962	0.967	0.969	0.971	0.981	1.014	1.077	1.124	1.124	
50mm									
0.596	0.639	0.690	0.701	0.708	0.718	0.778	0.780	0.786	0.805
0.807	0.819	0.825	0.825	0.829	0.830	0.860	0.869	0.879	0.899
0.913	0.915	0.917	0.930	0.946	0.963	0.967	0.967	0.971	0.977
0.985	1.011	1.012	1.016	1.024	1.028	1.040	1.046	1.049	1.063
1.065	1.073	1.082	1.083	1.095	1.101	1.112	1.120	1.139	1.148
1.241	1.241	1.345	1.355	1.388	1.414				

Step 3 Recall that this step involves fitting a three parameter Burr XII distribution and hence follows the procedure outlined above. With $\theta_1 = 1$, the intermediate values of τ_1 (starting at $\widehat{B}$) and the corresponding first derivatives were observed to be

τ_1	5.87366	6.28297	6.31166	6.31179	6.31179	6.31179
$\frac{\partial l_b^*}{\partial \tau}$	23.75528	4.73885	0.29229	0.00128	2.497×10^{-8}	-4.44×10^{-15}

Thus, $\tau_1 = 6.31179$, $\frac{\partial l_b^*}{\partial \tau} = -4.44 \times 10^{-15}$, while we also have $\frac{\partial^2 l_b^*}{\partial \tau^2} = -10.09775$. Hence, $a_1 = 1.65616$, $\frac{\partial l_b}{\partial a} = 0$, $\frac{\partial^2 l_b}{\partial a^2} = -93.6968$, $\frac{\partial l_b}{\partial \theta} = 6.5773$ and $\frac{\partial^2 l_b}{\partial \theta^2} = -4571.532$. We then use these to find $\theta_2 = 1.04153$. After 10 iterations, θ_{11} was found to be 1.0534. The intermediate

values of τ_{11} and the corresponding first derivatives were observed to be

τ_{11}	6.011827	6.011827	6.011827
$\frac{\partial l_b^*}{\partial \tau}$	-1.97×10^{-13}	-4.5×10^{-15}	-4.5×10^{-15}

Thus $\tau_{11} = 6.011827$, $\frac{\partial l_b^*}{\partial \tau} = -4.5 \times 10^{-15}$, while $a_{11} = 2.04742$, $\frac{\partial l_b}{\partial a} = 0$ and $\frac{\partial l_b}{\partial \theta} = -3.24 \times 10^{-13}$. We can then find $\theta_{12} = 1.0534$. To ensure that these parameter estimates are maximums, we determine the principal minors of the full 3×3 matrix of second derivatives, as for the three-parameter Burr XII case above. The matrix is given by

$$\begin{pmatrix} \frac{\partial^2 l_b}{\partial a^2} & & \\ \frac{\partial^2 l_b}{\partial a \partial \tau} & \frac{\partial^2 l_b}{\partial \tau^2} & \\ \frac{\partial^2 l_b}{\partial a \partial \theta} & \frac{\partial^2 l_b}{\partial \theta \partial \tau} & \frac{\partial^2 l_b}{\partial \theta^2} \end{pmatrix} = \begin{pmatrix} -61.3081 & & \\ 0.21596 & -12.0506 & \\ 481.26298 & -61.7825 & -4170.317 \end{pmatrix}.$$

Clearly, the first principal minor is negative (-61.3081), while the second is found to be positive (738.753) and the third principal minor is found to be negative (-68569.6), confirming the parameter estimates to be maximums. Therefore, the Burr XII parameter estimates were found to be $\widehat{\tau} = 6.0118$, $\widehat{a} = 2.04742$ and $\widehat{\theta} = 1.0534$

Step 4 We now set $\alpha_b = \ln\left(\widehat{\theta}\right) = \ln(1.0534) = 0.05202$ and $\beta_b = 0$. Using the final estimates of τ and a from the previous step as a starting point, we then converge on the final parameter estimates for the rescaled data. The table below shows the values obtained at the end of each iteration, and their corresponding first derivatives.

Iteration	**1**	**2**	**3**	**4**	**5**	**6**
a	1.9221	1.8990	1.8993	1.8993	1.8993	1.8993
τ	6.1130	6.1235	6.1234	6.1234	6.1234	6.1234
α_b	0.0211	0.0177	0.0177	0.0177	0.0177	0.0177
β_b	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
$\frac{\partial l_B}{\partial a}$	-1.6×10^{-12}	0.0592	0.0096	2.9×10^{-6}	-1.3×10^{-11}	0
$\frac{\partial l_B}{\partial \tau}$	2.01×10^{-13}	-0.0775	0.0003	2.7×10^{-7}	1.6×10^{-12}	-8.4×10^{-15}
$\frac{\partial l_B}{\partial \alpha_b}$	1.4×10^{-11}	-1.2147	-0.0697	-9.1×10^{-6}	-8.1×10^{-11}	0
$\frac{\partial l_B}{\partial \beta_b}$	1268.1273	-11.782	-1.3843	-0.0001	-1.6×10^{-9}	7.2×10^{-12}

For the final iteration, the full 4×4 matrix of second derivatives is given by

$$\begin{pmatrix} \frac{\partial^2 l_B}{\partial a^2} & & & \\ \frac{\partial^2 l_B}{\partial a \partial \tau} & \frac{\partial^2 l_B}{\partial \tau^2} & & \\ \frac{\partial^2 l_B}{\partial a \partial \alpha_b} & \frac{\partial^2 l_B}{\partial \tau \partial \alpha_b} & \frac{\partial^2 l_B}{\partial \alpha_b^2} & \\ \frac{\partial^2 l_B}{\partial a \partial \beta_b} & \frac{\partial^2 l_B}{\partial \tau \partial \beta_b} & \frac{\partial^2 l_B}{\partial \alpha_b \partial \beta_b} & \frac{\partial^2 l_B}{\partial \beta_b^2} \end{pmatrix}$$

$$= \begin{pmatrix} -71.23 & & & \\ -1.17 & -11.33 & & \\ 542.77 & -55.57 & -4623.08 & \\ 11398.37 & -1192.47 & -100654.1 & -3965323 \end{pmatrix}.$$

Extending previous results for 3×3 matrices, we now determine the four principal minors of this matrix. Clearly, the first is negative (-71.23), while the second is found to be positive (805.667), the third negative (-96314.8) and the fourth positive $\left(1.616 \times 10^{11}\right)$. Hence, for the re-scaled data, the parameter estimates are maximums and are $\widehat{a} = 1.8993$, $\widehat{\tau} = 6.1234$, $\widehat{\alpha}_b = 0.0177$ and $\widehat{\beta}_b = 0.0007$.

We can see that the estimates of the shape parameters a and τ are similar to those in step 3, while the estimates of the two scale parameters α_b and β_b are close to their starting values of $\ln\left(\widehat{\theta}\right)$ and zero respectively.

Step 5 The shape parameters remain at $\widehat{\tau} = 6.1234$ and $\widehat{a} = 1.8993$. Upon adding $\widehat{\alpha}_w$ and $\widehat{\beta}_w$ to $\widehat{\alpha}_b$ and $\widehat{\beta}_b$ respectively we then obtain parameter estimates for the unscaled data of $\widehat{\alpha}_b = 1.41096$ and $\widehat{\beta}_b = -0.011$.

The SAS IML program that executes this algorithm can be found in Appendix A as *burr_alt.*

2.3.3 Examples

Example 2.1.1.1 [Carbon Fibre Rod Data] revisited

Summarising the results above, the maximum likelihood estimates of the parameters from the Burr XII ALT model are $\widehat{a} = 1.8993$, $\widehat{\tau} = 6.1234$, $\widehat{\alpha}_b = 1.41096$ and $\widehat{\beta}_b = -0.011$. Hence, $\widehat{\theta}_1 = 4.05503, \widehat{\theta}_2 = 3.67282, \widehat{\theta}_3 = 3.29023$ and $\widehat{\theta}_4 = 2.36543$, with associated log-likelihood $l_B = -262.7479$.

Example 2.1.1.2 [Aluminium Coupon Data] revisited

Fitting a Burr XII ALT model to the aluminium coupon data we obtain parameter estimates of $\widehat{a} = 1.8200143$, $\widehat{\tau} = 7.0197622$, $\widehat{\alpha}_b = 7.6785289$ and $\widehat{\beta}_b = -2.361454$. Hence, $\widehat{\theta}_1 = 15.1724$, $\widehat{\theta}_2 = 4.65876$ and $\widehat{\theta}_3 = 1.4305$, with associated log-likelihood $l_B = -393.5695$.

Example 2.1.1.3 [Electrode Data] revisited

In attempting to establish the maximum likelihood estimates of the parameters in the above way, we observed similar behaviour to that found in fitting the three parameter Burr XII model to the entire data set; see section 2.2 above. Here, the iterative values for a did not converge on a finite maximum likelihood estimate for this parameter; the values of a, and also l_B, steadily rose. Again, we discuss this phenomenon in greater detail in the next section. For completeness, we record the final values as $a = 1000000$, $\tau = 11.28027$,

$\alpha_b = 4.987465$ and $\beta_b = 0.0001656$, and using these values we then find approximations to the scales θ_1, θ_2 and θ_3 as $146.8073, 149.0117$ and 172.9608 respectively, with associated log-likelihood $l_B = -530.2298$.

2.4 Contrast between Weibull and Burr XII

In this final section, we examine the maximum likelihood estimates of the parameters in each of the models outlined above. In addition, we shall compare the associated maximised log-likelihoods and present some graphs to illustrate our work.

Example 2.1.1.1 [Carbon Fibre Rod Data] revisited

In Table 2.15 we present an overview of the parameter estimates and maximised log-likelihoods from fitting the distributions to each subset of data. It is worth noting that $\widehat{\tau}$ is not too far removed from $\widehat{B}$ in each case, while $\widehat{\theta}$ is also similar to $\widehat{\phi}$ for each subset. In addition, $\widehat{\phi}$ gradually decreases as the length of the fibre increases. We see that, across all subsets of data, the Burr XII model outperforms the Weibull model in terms of maximised log-likelihood; albeit occasionally only marginally.

Moving on to consider results from an accelerated framework, we obtained the maximum likelihood estimates detailed in Table 2.16. We see that, in both models, the scale estimates $\widehat{\phi}, \widehat{\theta}$ have a tendency to decrease with increasing fibre length, as expected, while the four values of $\widehat{\phi}$ are numerically comparable to their $\widehat{\theta}$ counterparts. We also note, that $\widehat{\tau}$ is similar in value to $\widehat{B}$, a facet that was noticed for the non-accelerated models in Table 2.15. On the same theme, we see that scale estimates obtained from fitting the accelerated models (Weibull or Burr XII) are in-line with their non-accelerated counterparts. As in the non-accelerated case, the maximised log-likelihood for the Burr XII model is greater than that for the Weibull model. However, in both instances the maximised log-likelihoods are less than the sum of those for the separate analyses under the non-ALT models; this is to be expected, since we are essentially using two parameters in the log-linear link (1.12) to determine four scale values, while in the non-ALT case we are estimating four individual scale values.

Example 2.1.1.2 [Aluminium Coupon Data] revisited

Table 2.17 presents the parameter estimates and maximised log-likelihoods when fitting the two distributions to each subset of the aluminium coupon data. As with the previous example, we note that $\widehat{\tau}$ is similar in magnitude to the $\widehat{B}$ for all stress levels. At the lowest stress level ($x_1 = 2.1$) we see the value of $\widehat{a}$ cited is actually quite large, perhaps larger than would be deemed normal in practice. Nevertheless, the value was converged upon by the algorithm and is a true maximum likelihood estimate. Comparing $\widehat{\theta}$ and $\widehat{\phi}$, we see that there is rather good agreement at the two highest stress levels, but there is some disparity at the

	Length/mm	1	10	20	50
	n	57	64	70	66
Weibull	$\widehat{B}$	5.59297	5.02663	5.52444	6.03828
Non-ALT	$\widehat{\phi}$	4.57525	3.30519	2.64819	2.42443
Burr XII	$\widehat{a}$	2.588745	0.9826893	3.3372807	4.943483
Non-ALT	$\widehat{\tau}$	6.963570	8.6832448	6.6199838	6.792626
	$\widehat{\theta}$	5.014342	2.972029	3.064437	2.996771
Likelihoods	l_w	−71.02396	−62.9666	−49.92875	−36.16505
	l_b	−70.30808	−58.61848	−49.13677	−35.83991

Table 2.15: Comparison of Weibull and Burr XII three-parameter models for carbon fibre rod data

	Parameter Estimates		Scale Estimates	
Weibull	$\widehat{B}$	4.5646	$\widehat{\phi}_1 = 3.98107$	$\widehat{\phi}_2 = 3.58305$
ALT	$\widehat{\alpha}_w$	1.3932	$\widehat{\phi}_3 = 3.18731$	$\widehat{\phi}_4 = 2.24355$
	$\widehat{\beta}_w$	−0.011		
Burr XII	$\widehat{a}$	1.8993	$\widehat{\theta}_1 = 4.05503$	$\widehat{\theta}_2 = 3.67282$
ALT	$\widehat{\tau}$	6.1234	$\widehat{\theta}_3 = 3.29023$	$\widehat{\theta}_4 = 2.36543$
	$\widehat{\alpha}_b$	1.4109		
	$\widehat{\beta}_b$	−0.011		
Likelihoods	l_W	−268.6926		
	l_B	−262.7480		

Table 2.16: Summaries of Weibull and Burr XII ALT models for carbon fibre rod data.

	psi/cycle n	2.1 101	2.6 102	3.1 101
Weibull	$\widehat{B}$	4.0298862	7.0075353	6.0734031
Non-ALT	$\widehat{\phi}$	15.426431	4.2437821	1.4316699
Burr XII	$\widehat{a}$	19.235884	2.831573	1.5708883
Non-ALT	$\widehat{\tau}$	4.1505101	8.615969	9.3872102
	$\widehat{\theta}$	31.15297	4.632881	1.4174932
Likelihoods	l_w	−279.4588	−98.07688	2.807636
	l_b	−279.4295	−96.59839	10.014003

Table 2.17: Comparison of Weibull and Burr XII three-parameter models for aluminium coupon data

	Parameter Estimates		Scale Estimates
Weibull	$\widehat{B}$	4.91691	$\widehat{\phi}_1 = 15.227945$
ALT	$\widehat{\alpha}_w$	7.795572	$\widehat{\phi}_2 = 4.5512789$
	$\widehat{\beta}_w$	−2.41544	$\widehat{\phi}_3 = 1.3602715$
Burr XII	$\widehat{a}$	1.8200143	$\widehat{\theta}_1 = 15.1724$
ALT	$\widehat{\tau}$	7.0197622	$\widehat{\theta}_2 = 4.65876$
	$\widehat{\alpha}_b$	7.6785289	$\widehat{\theta}_3 = 1.4305$
	$\widehat{\beta}_b$	−2.361454	
Likelihoods	l_W	−415.5829	
	l_B	−393.5695	

Table 2.18: Summaries of Weibull and Burr XII ALT models for aluminium coupon data.

lowest level; the value of $\widehat{\theta}$ for $x_1 = 2.1$ is clearly influenced by the value of $\widehat{a}$. Nevertheless, we observe both $\widehat{\theta}$ and $\widehat{\phi}$ to decrease as the psi/cycle increases. In terms of log-likelihood, the Burr XII model is seen to outperform the Weibull model at all three stress levels.

Results for the corresponding accelerated models are given in Table 2.18. Here, we clearly see the Burr XII ALT model providing a superior fit to the data set than its Weibull counterpart. As for the non-accelerated case, we see $\widehat{\tau}$ is similar to $\widehat{B}$, while the Weibull scale estimates are close to their Burr XII counterparts. These scale estimates are also seen to be very similar in value to the separate scale estimates obtained for the non-ALT Weibull model.

Example 2.1.1.3 [Electrode Data] revisited

Table 2.19 summarises our results for the electrode data, (Table 2.6); the carets over the parameters are omitted here since, in this case, we do not strictly speaking have maximum likelihood estimates. As with the previous example, we see τ taking approximately the same value as $\widehat{B}$ for each data subset. However, as noticed previously, the value of a does not correspond to convergence; rather, the value for this parameter continued to rise steadily until the algorithm was stopped after 500 iterations, and it is this final value that is presented. With regards to maximised log-likelihoods, the Weibull model provides a slightly

	volts/second n	10 60	100 60	1000 60
Weibull	$\widehat{B}$	12.220089	16.450886	14.681733
Non-ALT	$\widehat{\phi}$	39.695289	46.247051	50.860083
Burr XII	a	253895932	1.31×10^{10}	3.48×10^{9}
Non-ALT	τ	12.220089	16.450873	14.681725
	θ	193.47247	190.70582	227.82527
Likelihoods	l_w	−165.5837	−155.4434	−169.9556
	l_b	−165.584	−155.4438	−170.0088

Table 2.19: Comparison of Weibull and Burr XII three-parameter models for electrode data,

better fit than the Burr XII three-parameter model in this case.

The results from the accelerated framework are shown in Table 2.20. Again, the log-likelihood for the Weibull model is marginally higher than that for the Burr XII model, with τ assuming a value very close to $\widehat{B}$. The value of a was pre-determined as a stopping point for the numerical procedure: with values for a failing to converge on a finite value, but simply getting gradually larger, it was decided to force the algorithm to take a at progressively higher values and then to use the Newton-Raphson procedure to converge on the remaining three parameters. This is in contrast to the non-accelerated case, where we simply let the algorithm continue to try to home in on the MLEs. There (Table 2.19) we saw a and θ get progressively larger, while τ tended to $\widehat{B}$. We also saw a gradually increasing log-likelihood that approached, but never attained, the value of its counterpart for the Weibull model. For the ALT case, we observe a similar behaviour for progressively larger, fixed values of a. In particular, we see τ is very close to $\widehat{B}$. We also see β_b having a similar value to $\widehat{\beta}_w$ and the log-likelihoods being very similar. The asymptotic properties of these parameters is illustrated in Figures 2.15 to 2.18, where a ranges from 5 to 1000000 and, for the sake of clarity, we present the natural logarithm of a on the horizontal axis. We see that l_B seems to attain its maximum value when $\ln(a) \to \infty$, while, in the same limit, $\alpha_b \to \infty$, $\tau \to \widehat{B}$ and $\beta_b \to \widehat{\beta}_w$. We shall return to these figures in Chapter four.

Consideration of this asymptotic behaviour has been undertaken previously by Crowder and Kimber (1997), and Watkins (2001) for a non-accelerated framework, and has links to the limiting behaviour outlined in (1.8). From there, we recall that

$$\frac{\theta}{a^{1/\tau}} \to \phi$$

when θ - and consequently a - tend to infinity. As a result, a and θ could be unbounded as l_b approaches its maximised value. However, Watkins' approach to problems where MLEs were difficult to obtain (such as the electrode example above) differed from ours. While we chose to let the algorithm attempt to determine MLEs for all three model parameters, Watkins chose to fix the scale parameter at progressively larger values and then determine estimates of the two shape parameters for each value of the scale. We employed a similar approach

	Shape Estimates		Scale Estimates
Weibull	$\widehat{B}$	11.280264	$\widehat{\phi}_1 = 43.140079$
ALT	$\widehat{\alpha}_w$	3.7627961	$\widehat{\phi}_2 = 43.787981$
	$\widehat{\beta}_w$	0.0001656	$\widehat{\phi}_3 = 50.827037$
Burr XII	a	1000000	$\widehat{\theta}_1 = 146.8073$
ALT	τ	11.28027	$\widehat{\theta}_2 = 149.0117$
	α_b	4.987465	$\widehat{\theta}_3 = 172.9608$
	β_b	0.0001656	
Likelihoods	l_W	−530.2297	
	l_B	−530.2298	

Table 2.20: Summaries of Weibull and Burr XII ALT models for electrode data.

when making the extension to acceleration. However, here we fixed a at progressively larger values and let the algorithm determine parameter estimates for τ, α_b and β_b. Nevertheless, the asymptotic results we observe in the non-ALT case, as a and θ are seen to increase, are consistent with those observed by Watkins for increasing, user-defined, values of the scale parameter, and are analogous when we extend our work to the accelerated framework.

2.5 Discussion

In this chapter we have examined three examples. Firstly, we had the situation whereby the Burr XII three parameter and Burr XII ALT models provided a better fit to the data set than the corresponding Weibull models. There, all parameter estimates were finite and easily converged upon by the algorithm. The second example exhibited the same properties, with the Burr XII models providing a better fit than their Weibull counterparts. Finally, we saw an example where the Weibull was a better fit than the Burr XII for each data subset and, similarly, the Weibull ALT was preferred to the Burr XII ALT model for the data set as a whole. In this case, there were difficulties in converging on a value for $\widehat{a}$, and the log-likelihood for the Burr XII models was found to rise steadily with increasing values of a, eventually approaching, but never quite attaining, its Weibull counterpart. We shall consider this phenomenon in more detail in Chapter four, while in the next chapter we investigate the properties of the parameter estimates, and in particular, we are interested in determining the variance of the MLEs. This we do by formulating the expected Fisher information matrices for the array of models considered in this chapter.

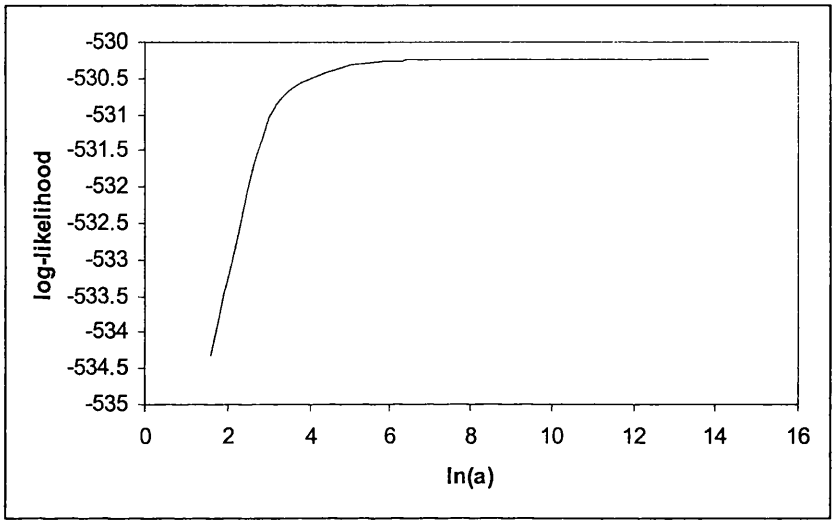

Figure 2.15: Burr XII ALT log-likelihood versus $\ln(a)$ for electrode data.

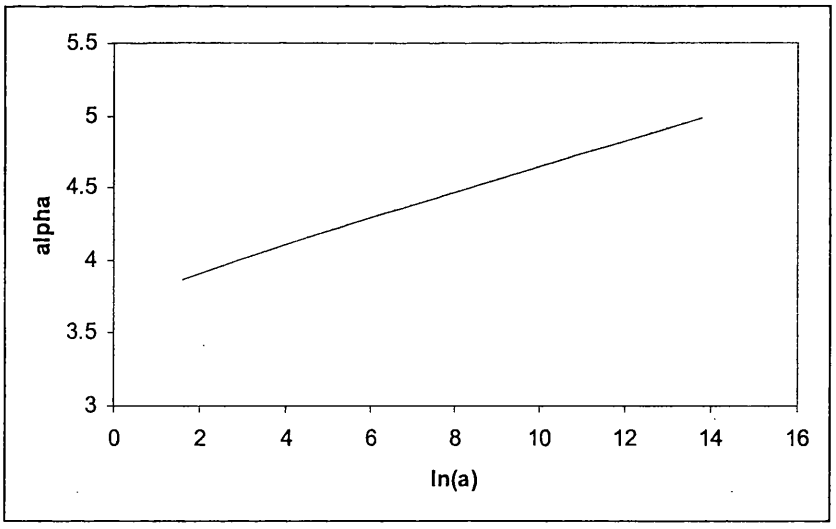

Figure 2.16: α_b versus $\ln(a)$ for electrode data.

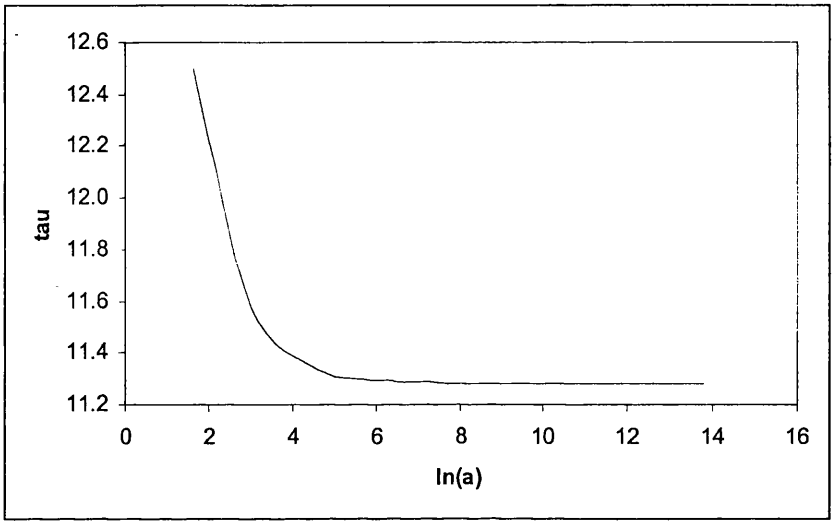

Figure 2.17: τ versus $\ln(a)$ for electrode data.

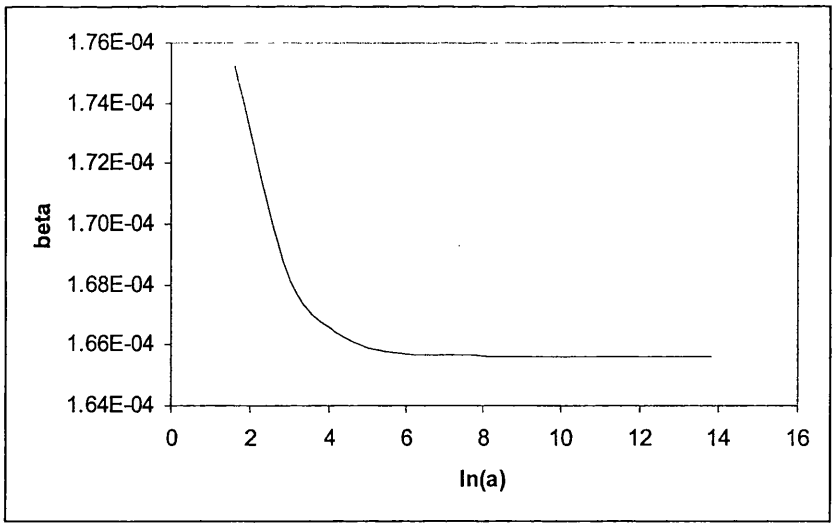

Figure 2.18: β_b versus $\ln(a)$ for electrode data.

Chapter 3

The Expected Fisher Information Matrices for Burr XII Models

In the previous chapter, we considered the maximum likelihood estimates of the parameters in the various Weibull and Burr XII models. We now consider the expected Fisher information (hereafter abbreviated to EFI) matrices needed to make large sample approximations to the variances of the MLEs of the parameters. In this chapter, we again only consider complete data sets. Cox and Hinkley (1974) discuss the theoretical properties of MLEs, with particular attention on asymptotic results. In general, the EFI matrix of second order partial derivatives is symmetric about the main diagonal; see, for example, Nelson (1990). We exploit this symmetry by presenting only the lower triangle of such matrices.

Watkins (1998b) presents expectations of second derivatives for the two parameter Weibull model, while Watkins (1991) considers the extension to the accelerated Weibull framework; this paper also considers the asymptotic validity of formulae derived from the expressions for simulated data sets with varying sample sizes.

We begin by outlining previous work for the Burr XII two and three parameter models and then extend this to accomodate the accelerated Burr XII model. Finally we shall provide evidence of the agreement with theoretical results, using simulations.

3.1 EFI Matrix for Various Burr XII Distributions

3.1.1 The Basic Two Parameter Distribution

In a further paper, Watkins (1997) presents a series of compact expressions for expectations of second derivatives of l_b; these results will form a basis for the accelerated model later. The EFI matrix is given as:

$$\begin{bmatrix} na^{-2} & \\ \frac{n\{1-\gamma-\psi(a)\}}{\tau(a+1)} & n\tau^{-2}\left[1+\frac{a}{a+2}\left\{\begin{array}{c}\frac{\pi^2}{6}+\gamma^2-2\gamma+2(\gamma-1)\psi(a+1)\\ +[\psi(a+1)]^2+\psi'(a+1)\end{array}\right\}\right] \end{bmatrix}$$

3.1.2 Non-ALT Three Parameter Distribution

The introduction of a scaling parameter into the Burr XII model is also accommodated by Watkins (1997). The resultant EFI matrix for the parameterisation (1.5) is

$$\begin{bmatrix} na^{-2} & & \\ \frac{n\{1-\gamma-\psi(a)\}}{\tau(a+1)} & n\tau^{-2}\left[1+\frac{a}{a+2}\left\{\begin{array}{c}\frac{\pi^2}{6}+\gamma^2-2\gamma \\ +2(\gamma-1)\psi(a+1) \\ +[\psi(a+1)]^2+\psi'(a+1)\end{array}\right\}\right] & \\ -\frac{n\tau}{\theta(a+1)} & -\frac{na\{1-\gamma-\psi(a+1)\}}{\theta(a+2)} & \frac{n\tau^2 a}{\theta^2(a+2)} \end{bmatrix}, \tag{3.1}$$

where the expectations of $\frac{\partial^2 l_b}{\partial a^2}, \frac{\partial^2 l_b}{\partial a \partial \tau}$ and $\frac{\partial^2 l_b}{\partial \tau^2}$ are as for the two parameter model above.

3.1.3 Burr XII ALT Distribution

In order to consider the accelerated model for the Burr XII distribution in a more tractable form, we first note that the transformation

$$Y_{ij} = \frac{D_{ij}}{\theta_i} = \frac{D_{ij}}{\exp(\alpha_b + \beta_b x_i)}$$

produces variables which follow the basic two parameter Burr XII distribution, given by (1.3). This transformation will enable us to write down simplified expressions for expectations using results from Watkins (1997) For instance, if Y follows (1.3), then $1+Y^\tau$ follows a Pareto distribution and consequently, $\log(1+Y^\tau)$ follows a negative exponential distribution. Therefore, we immediately have

$$E[\ln(1+Y^\tau)] = a^{-1}, \tag{3.2}$$

while we observed at (1.6) that

$$E[Y^s] = \frac{\Gamma\left(1+\frac{s}{\tau}\right)\Gamma\left(a-\frac{s}{\tau}\right)}{\Gamma(a)}, \tag{3.3}$$

defined for $s < a\tau$. The study of the first and second derivatives of (2.10), in Chapter 2 above, shows that we only need to find the expectations of a few fundamental functions in order to write down the expectations of the more complicated expressions required. These fundamental functions are $Y^{i\tau}(\ln Y)^j$ for $i, j \leq 2$. From (3.3), we have

$$E[Y^\tau] = \frac{\Gamma(2)\Gamma(a-1)}{\Gamma(a)} = \frac{1}{a-1} \tag{3.4}$$

and

$$E[Y^{2\tau}] = \frac{\Gamma(3)\Gamma(a-2)}{\Gamma(a)} = \frac{2}{(a-1)(a-2)}, \tag{3.5}$$

while differentiating (3.3) with respect to s gives,

$$E\left[Y^s \ln Y\right] = \frac{\Gamma'\left(1+\frac{s}{\tau}\right)\Gamma\left(a-\frac{s}{\tau}\right) - \Gamma\left(1+\frac{s}{\tau}\right)\Gamma'\left(a-\frac{s}{\tau}\right)}{\tau\Gamma(a)}$$

which, for $s = 0$, gives

$$E\left[\ln Y\right] = \frac{-\{\gamma + \psi(a)\}}{\tau} \tag{3.6}$$

while, for $s = \tau$, we have

$$E\left[Y^\tau \ln Y\right] = \frac{\Gamma'(2)\Gamma(a-1) - \Gamma(2)\Gamma'(a-1)}{\tau\Gamma(a)} = \frac{1-\gamma-\psi(a-1)}{(a-1)\tau}. \tag{3.7}$$

Differentiating (3.3) a second time with respect to s gives,

$$E\left[Y^s (\ln Y)^2\right] = \frac{\Gamma''\left(1+\frac{s}{\tau}\right)\Gamma\left(a-\frac{s}{\tau}\right) - 2\Gamma'\left(1+\frac{s}{\tau}\right)\Gamma'\left(a-\frac{s}{\tau}\right) + \Gamma\left(1+\frac{s}{\tau}\right)\Gamma''\left(a-\frac{s}{\tau}\right)}{\tau^2\Gamma(a)}.$$

Setting $s = 0$, we obtain

$$E\left[(\ln Y)^2\right] = \frac{\frac{\pi^2}{6} + \gamma^2 + 2\gamma\psi(a) + \{\psi(a)\}^2 + \psi'(a)}{\tau^2},$$

while setting $s = \tau$, we get

$$E\left[Y^\tau (\ln Y)^2\right] = \frac{1}{\tau^2(a-1)}\left[\begin{array}{c}\left\{\frac{\pi^2}{6} + \gamma^2 - 2\gamma\right\} + 2(\gamma-1)\psi(a-1) \\ +\{\psi(a-1)\}^2 + \psi'(a-1)\end{array}\right]. \tag{3.8}$$

With $s = 2\tau$ we have

$$E\left[Y^{2\tau} (\ln Y)^2\right] = \frac{1}{\tau^2(a-1)(a-2)}\left[\begin{array}{c}\frac{\pi^2}{3} + 2\left(\gamma^2 - 3\gamma + 1\right) \\ -2(3-2\gamma)\psi(a-2) \\ +2\left[\{\psi(a-2)\}^2 + \psi'(a-2)\right]\end{array}\right] \tag{3.9}$$

upon simplification. It is now straightforward to show that

$$E\left[S_d^*\right] = E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} \ln\left(1 + Y_{ij}^\tau\right)\right] = S_n E\left[\ln\left(1+Y^\tau\right)\right] = \frac{S_n}{a}, \tag{3.10}$$

using (3.2), and

$$E\left[S_e\right] = E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} \ln d_{ij}\right] = \sum_{i=1}^{k} n_i\left\{\ln\theta_i + E\left[\ln Y\right]\right\} = S_n\alpha_b + S_x\beta_b - \frac{S_n\{\gamma+\psi(a)\}}{\tau}, \tag{3.11}$$

using (3.6) and (1.12).

3.1.3.1 Expectations of Derived Expressions

In the following, we present expectations with the subscript a to emphasise the role of this shape parameter in the expectation. The following theorem covers many of the terms we consider below.

Theorem 3.1 *Let Y follow a two parameter Burr XII distribution with CDF given by (1.3). Then, for $m \geq 0$,*

$$E_a\left[\frac{Y^{i\tau}(\ln Y)^j}{(1+Y^\tau)^m}\right] = \frac{a}{a+m}E_{a+m}\left[Y^{i\tau}(\ln Y)^j\right].$$

Proof

By definition, we have, from (1.4),

$$E_a\left[\frac{Y^{i\tau}(\ln Y)^j}{(1+Y^\tau)^m}\right] = \int_0^\infty \frac{a\tau y^{\tau-1}}{(1+y^\tau)^{a+1}} \times \frac{y^{i\tau}(\ln y)^j}{(1+y^\tau)^m} dy = \int_0^\infty \frac{a\tau y^{(i+1)\tau-1}(\ln y)^j}{(1+y^\tau)^{a+1+m}} dy$$

while

$$\frac{a}{a+m}E_{a+m}\left[Y^{i\tau}(\ln Y)^j\right] = \frac{a}{a+m}\int_0^\infty \frac{(a+m)\tau y^{\tau-1}}{(1+y^\tau)^{a+1+m}} \times y^{i\tau}(\ln y)^j\, dy$$

which is just

$$\int_0^\infty \frac{a\tau y^{(i+1)\tau-1}(\ln y)^j}{(1+y^\tau)^{a+1+m}} dy,$$

as above. This completes the proof.

3.1.3.1.1 Terms Not Involving Logarithms Now, by applying Theorem 3.1 with $i = 1, j = 0, m = 1$, we have,

$$E_a\left[\frac{Y^\tau}{1+Y^\tau}\right] = \frac{a}{a+1}E_{a+1}\left[Y^\tau\right] = \frac{1}{a+1}, \tag{3.12}$$

using (3.4); similarly, we have, with $i = 2, j = 0, m = 2$,

$$E_a\left[\frac{Y^{2\tau}}{(1+Y^\tau)^2}\right] = \frac{a}{a+2}E_{a+2}\left[Y^{2\tau}\right] = \frac{2}{(a+1)(a+2)}, \tag{3.13}$$

using (3.5).

3.1.3.1.2 Terms Involving Logarithms The same manipulations can be applied to the following two expressions in this category, which are

$$E_a\left[\frac{Y^\tau \ln Y}{1+Y^\tau}\right] = \frac{a}{a+1}E_{a+1}\left[Y^\tau \ln Y\right] = \frac{1-\gamma-\psi(a)}{(a+1)\tau} \tag{3.14}$$

and

$$E_a\left[\frac{Y^{2\tau}\ln Y}{(1+Y^\tau)^2}\right]=\frac{a}{a+2}E_{a+2}\left[Y^{2\tau}\ln Y\right]=\frac{3-2\gamma-2\psi(a)}{\tau(a+1)(a+2)}, \tag{3.15}$$

which, we note, also make use of Theorem 3.1.

3.1.3.1.3 Terms Involving Squares Of Logarithms We will need only one term involving a square of a logarithm, which is

$$E_a\left[\frac{Y^{2\tau}(\ln Y)^2}{(1+Y^\tau)^2}\right]=\frac{a}{a+2}E_{a+2}\left[Y^{2\tau}(\ln Y)^2\right];$$

using Theorem 3.1 and, replacing a with $a+2$ in (3.9), this is

$$\frac{1}{\tau^2(a+1)(a+2)}\left[\left\{\frac{\pi^2}{3}+2\gamma^2-6\gamma+2\right\}-2(3-2\gamma)\psi(a)+2\left\{\psi^2(a)+\psi'(a)\right\}\right]. \tag{3.16}$$

Now, to be able to write down the expectations of all first and second derivatives of l_B, as given in section 2.3.1, we need to establish expectations of the summations given therein. However, it is quite straightforward to link the expectations of the derived expressions above to the required summations. For instance, we see that

$$\begin{aligned}E[S_{0101}] &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{Y_{ij}^\tau}{1+Y_{ij}^\tau}\right]=S_nE_a\left[\frac{Y^\tau}{1+Y^\tau}\right]=\frac{aS_n}{a+1}E_{a+1}[Y^\tau]\\ &= \frac{S_n}{a+1},\end{aligned}$$

from (3.12). The introduction of the stress factor is also straightforward to accommodate; for example, we have

$$E[S_{1202}]=E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{x_iY_{ij}^{2\tau}}{\left(1+Y_{ij}^\tau\right)^2}\right]=\frac{2S_x}{(a+1)(a+2)},$$

using (3.13). Other expectations can be found through recurrence relations: for instance, we have, after some routine algebra,

$$S_{0111}-S_{0212}=S_{0112}$$

and

$$S_{0121}-S_{0222}=S_{0122}$$

where

$$E[S_{0112}] = S_nE_a\left[\frac{Y^\tau\ln Y}{(1+Y^\tau)^2}\right]=S_n\frac{a}{a+2}E_{a+2}[Y^\tau\ln Y]$$

$$= \frac{S_n a \{1-\gamma-\psi(a+1)\}}{(a+1)(a+2)\tau} \tag{3.17}$$

and

$$E[S_{0122}] = S_n a \left[\frac{\frac{\pi^2}{6}+\gamma^2-2\gamma-2(1-\gamma)\psi(a+1)+\{\psi(a+1)\}^2+\psi'(a+1)}{\tau^2(a+1)(a+2)} \right]. \tag{3.18}$$

These examples illustrate a general procedure for writing down expectations for all required summations.

3.1.3.2 Expectation Of The Score

Before proceeding to determining expectations of second derivatives, we first demonstrate that the expectations of score functions are zero, as we require. From (2.13), we have

$$E\left[\frac{\partial l_B}{\partial a}\right] = S_n a^{-1} - E[S_d^*] = S_n a^{-1} - S_n a^{-1} = 0,$$

while, from (2.14),

$$\begin{aligned} E\left[\frac{\partial l_B}{\partial \tau}\right] &= S_n \tau^{-1} - \alpha_b S_n - \beta_b S_x + E[S_e] - (a+1)E[S_{0111}] \\ &= S_n \tau^{-1} - \alpha_b S_n - \beta_b S_x + \alpha_b S_n + \beta_b S_x - \frac{S_n\{\gamma+\psi(a)\}}{\tau} - \frac{S_n\{1-\gamma-\psi(a)\}}{\tau} \\ &= 0, \end{aligned}$$

and from (2.15)

$$\begin{aligned} E\left[\frac{\partial l_B}{\partial \alpha_b}\right] &= -S_n\tau + (a+1)\tau E[S_{0101}] \\ &= -S_n\tau + (a+1)\tau\frac{S_n}{a+1} \\ &= 0 \end{aligned}$$

and, finally, from (2.16)

$$\begin{aligned} E\left[\frac{\partial l_B}{\partial \beta_b}\right] &= -S_x\tau + (a+1)\tau E[S_{1101}] \\ &= -S_x\tau + (a+1)\tau\frac{S_x}{a+1} \\ &= 0 \end{aligned}$$

Thus, the expectations of these first derivatives are zero for all values of n.

3.1.3.3 Expectations Of Second Derivatives of l_B

We now consider the expectations of second derivatives; note that these expressions then appear in formulae based on asymptotic theory, so that subsequent calculations derived from these expectations hold only for large sample sizes. Since $\frac{\partial^2 l_B}{\partial a^2}$ at (2.17) is deterministic, we now recall (2.18) to (2.20), and write

$$E\left[\frac{\partial^2 l_B}{\partial a^2}\right] = -\frac{S_n}{a^2}, \tag{3.19}$$

$$E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right] = -E\left[S_{0111}\right] = -\frac{S_n\left\{1-\gamma-\psi\left(a\right)\right\}}{\left(a+1\right)\tau}, \tag{3.20}$$

$$E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right] = \tau E\left[S_{0101}\right] = \frac{\tau S_n}{a+1},$$

and

$$E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right] = \tau E\left[S_{1101}\right] = \frac{\tau S_x}{a+1}.$$

Next, employing (3.18), we also have

$$\begin{aligned}
E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right] &= -S_n\tau^{-2} - \left(a+1\right)E\left[S_{0121} - S_{0222}\right] \\
&= -S_n\tau^{-2} - \left(a+1\right)E\left[S_{0122}\right] \\
&= -S_n\tau^{-2} - \frac{S_n a}{\tau^2\left(a+2\right)}\left[\begin{array}{c}\frac{\pi^2}{6}+\gamma^2-2\gamma-2\left(1-\gamma\right)\psi\left(a+1\right) \\ +\left\{\psi\left(a+1\right)\right\}^2+\psi'\left(a+1\right)\end{array}\right] \\
&= -\frac{S_n}{\tau^2\left(a+2\right)}\left\{2+a\left[\begin{array}{c}\frac{\pi^2}{6}+\gamma^2-2\gamma+1-2\left(1-\gamma\right)\psi\left(a+1\right) \\ +\left\{\psi\left(a+1\right)\right\}^2+\psi'\left(a+1\right)\end{array}\right]\right\} \\
&= -\frac{S_n}{\tau^2\left(a+2\right)}\left\{2+a\left[\begin{array}{c}\frac{\pi^2}{6}+\left(1-\gamma\right)^2-2\left(1-\gamma\right)\psi\left(a+1\right) \\ +\left\{\psi\left(a+1\right)\right\}^2+\psi'\left(a+1\right)\end{array}\right]\right\},
\end{aligned}$$

and, through the use of (3.17), we have

$$\begin{aligned}
E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right] &= -S_n + \left(a+1\right)\left(E\left[S_{0101}\right] + \tau E\left[S_{0111} - S_{0212}\right]\right) \\
&= -S_n + \left(a+1\right)\left(\frac{S_n}{a+1} + \tau E\left[S_{0112}\right]\right) \\
&= \left(a+1\right)\tau E\left[S_{0112}\right] \\
&= \frac{S_n a\left\{1-\gamma-\psi\left(a+1\right)\right\}}{\left(a+2\right)}
\end{aligned}$$

and

$$\begin{aligned}
E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right] &= -S_x + \left(a+1\right)\left(E\left[S_{1101}\right] + \tau E\left[S_{1111} - S_{1212}\right]\right) \\
&= -S_x + \left(a+1\right)\left(\frac{S_x}{a+1} + \tau E\left[S_{1112}\right]\right)
\end{aligned}$$

$$
\begin{aligned}
&= (a+1)\tau E[S_{1112}] \\
&= \frac{S_x a\{1-\gamma-\psi(a+1)\}}{(a+2)}.
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right] &= (a+1)\tau^2 E[S_{0202}-S_{0101}] \\
&= (a+1)\tau^2\left\{\frac{2S_n}{(a+1)(a+2)}-\frac{S_n}{(a+1)}\right\} \\
&= \tau^2 S_n\left\{\frac{2-(a+2)}{(a+2)}\right\} \\
&= -\frac{\tau^2 a S_n}{(a+2)};
\end{aligned}
$$

similarly,

$$
\begin{aligned}
E\left[\frac{\partial^2 l_B}{\partial \alpha_b \partial \beta_b}\right] &= (a+1)\tau^2 E[S_{1202}-S_{1101}] \\
&= -\frac{\tau^2 a S_x}{a+2},
\end{aligned}
$$

and, finally,

$$
\begin{aligned}
E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right] &= (a+1)\tau^2 E[S_{2202}-S_{2101}] \\
&= -\frac{\tau^2 a S_{xx}}{a+2}.
\end{aligned}
$$

We now have all the terms we need to form the EFI matrix for a Burr XII ALT model.

3.2 Simulation 1: Three Stress Levels

3.2.1 Background

It is now appropriate to reinforce the theoretical results above with some simulations. Initially, we take $k = 3$, stress levels $(x_1, x_2, x_3) = (150, 170, 190)$, with equal sample sizes at those stress levels. The choice of parameter values is based on failure times for Class-B insulation data, given in Nelson (1990,p158). Although the real life experiment featured censored items, we start with the assumption that the experiment is complete. We shall consider six variations on this experiment, each time making modifications to the values of the two shape parameters a and τ, keeping all other parameter values constant. For each value of $a\,(=2$ or $5)$ we choose a value for $\tau\,(=3, 5$ or $7)$, with the intention of covering a suitable portion of the parameter space. We also take $\alpha_b = 17.60139$ and $\beta_b = -0.056282$, determined from (1.12) in which we take $\theta_1 = 9500$ and $\theta_3 = 1000$. We shall perform 10000

replications for $n_1 = 30, 100, 1000$ and 3000. Now, with the EFI matrix being defined as

$$-\begin{bmatrix} E\left[\frac{\partial^2 l_B}{\partial a^2}\right] & & & \\ E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right] & E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right] & & \\ E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right] & E\left[\frac{\partial^2 l_B}{\partial \alpha_b \partial \tau}\right] & E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right] & \\ E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right] & E\left[\frac{\partial^2 l_B}{\partial \beta_b \partial \tau}\right] & E\left[\frac{\partial^2 l_B}{\partial \alpha_b \partial \beta_b}\right] & E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right] \end{bmatrix}, \tag{3.21}$$

we can write the expectations of second derivatives for simulations and theory and then compare results. Some simplification is now possible, partly because we have constant subsample sizes, and partly because the parameter values are now specified. For instance, we can now write

$$\begin{aligned} S_n &= k \times n_1 = 3n_1 \\ S_x &= (150 + 170 + 190)\, n_1 = 510 n_1 \\ S_{xx} &= \left(150^2 + 170^2 + 190^2\right) n_1 = 87500 n_1, \end{aligned}$$

and then substituting these expressions into the expectations of second derivatives, given in section 3.2.4.3 above, we see (3.21) reduces to

$$\begin{bmatrix} \frac{3n_1}{a^2} & & & \\ \frac{3n_1\{1-\gamma-\psi(a)\}}{(a+1)\tau} & \frac{3n_1}{\tau^2}\left\{\begin{array}{c} 2 + a\left[(1-\gamma)^2 + \psi'(a+1)\right. \\ -2(1-\gamma)\psi(a+1) \\ \left. + [\psi(a+1)]^2 + \frac{\pi^2}{6}\right] \end{array}\right\} & & \\ -\frac{3n_1\tau}{a+1} & -\frac{3an_1(1-\gamma-\psi(a+1))}{a+2} & \frac{3an_1\tau^2}{a+2} & \frac{510an_1\tau^2}{a+2} \\ -\frac{510n_1\tau}{a+1} & -\frac{510an_1\{1-\gamma-\psi(a+1)\}}{a+2} & \frac{510an_1\tau^2}{a+2} & \frac{87500an_1\tau^2}{a+2} \end{bmatrix}$$

$$= n_1 \begin{bmatrix} \frac{3}{a^2} & & & \\ \frac{3\{1-\gamma-\psi(a)\}}{(a+1)\tau} & \frac{3}{\tau^2}\left\{\begin{array}{c} 2 + a\left[(1-\gamma)^2 + \psi'(a+1)\right. \\ -2(1-\gamma)\psi(a+1) \\ \left. + [\psi(a+1)]^2 + \frac{\pi^2}{6}\right] \end{array}\right\} & & \\ -\frac{3\tau}{a+1} & -\frac{3a(1-\gamma-\psi(a+1))}{a+2} & \frac{3a\tau^2}{a+2} & \frac{510a\tau^2}{a+2} \\ -\frac{510\tau}{a+1} & -\frac{510a\{1-\gamma-\psi(a+1)\}}{a+2} & \frac{510a\tau^2}{a+2} & \frac{87500a\tau^2}{a+2} \end{bmatrix}.$$

Substituting $a = 2$ and $\tau = 3$ into this expression, we then have the EFI matrix, in our first example as

$$n_1 \begin{bmatrix} \frac{3}{4} & & & \\ 0 & \frac{1}{6}\left\{\begin{array}{c} 1 - \frac{5}{4} + (1-\gamma)^2 - 2(1-\gamma)\left(\frac{3}{2} - \gamma\right) \\ \left. + \left(\frac{3}{2} - \gamma\right)^2 + \frac{\pi^2}{3}\right. \end{array}\right\} & & \\ -3 & \frac{3}{4} & \frac{27}{2} & \\ -510 & \frac{255}{2} & 2295 & 393750 \end{bmatrix}$$

$$= n_1 \begin{bmatrix} 0.75 & & & \\ 0 & 0.548311\cdots & & \\ -3 & 0.75 & 13.5 & \\ -510 & 127.5 & 2295 & 393750 \end{bmatrix} \tag{3.22}$$

where the second diagonal element reduces to $\frac{\pi^2}{18}$, and is presented to six decimal places and all others are exact. Taking the inverse of (3.22) and simplifying gives the approximate variance-covariance matrix as

$$\begin{aligned} &\frac{1}{n_1} \begin{bmatrix} \frac{36-48\pi^2}{27-4\pi^2} & & & \\ \frac{144}{27-4\pi^2} & \frac{72}{4\pi^2-27} & & \\ \frac{32\pi^2}{12\pi^2-81} & \frac{36}{27-4\pi^2} & \frac{7803-1252\pi^2}{972-144\pi^2} & \\ 0 & 0 & -\frac{17}{360} & \frac{1}{3600} \end{bmatrix} \\ &= \frac{1}{n_1} \begin{bmatrix} 35.0798\cdots & & & \\ -11.5399\cdots & 5.76996\cdots & & \\ 8.43663\cdots & -2.88498\cdots & 10.1369\cdots & \\ 0 & 0 & -0.0472222\cdots & 0.000277778\cdots \end{bmatrix}. \end{aligned} \tag{3.23}$$

With the diagonal elements representing approximations to the variances of the maximum likelihood estimates, we can see that increasing the sample size n_1 reduces the variances, as expected. Naturally, taking the square roots of elements on the main diagonal of (3.23) gives the approximate standard deviations, and we can write

$$\begin{aligned} \sqrt{n_1} \times \text{standard deviation}\,(\widehat{a}) &= \sqrt{35.0798} = 5.92282\cdots, \\ \sqrt{n_1} \times \text{standard deviation}\,(\widehat{\tau}) &= \sqrt{5.76996} = 2.40207\cdots, \\ \sqrt{n_1} \times \text{standard deviation}\,(\widehat{\alpha}_b) &= \sqrt{10.1369} = 3.18385\cdots, \end{aligned}$$

and

$$\sqrt{n_1} \times \text{standard deviation}\left(\widehat{\beta}_b\right) = \sqrt{0.000277778} = 0.016666\cdots.$$

It is also straightforward to find the approximate correlation matrix from (3.23) as

$$\begin{bmatrix} 1 & & & \\ -0.811125\cdots & 1 & & \\ 0.447391\cdots & -0.377227\cdots & 1 & \\ 0 & 0 & -0.889906\cdots & 1 \end{bmatrix},$$

which is clearly independent of n_1.

3.2.2 Results

In Table 3.1, we present elements from the theoretical EFI matrix (3.22) divided by n_1, together with simulated counterparts for our four values of n_1. Table 3.2 gives the corresponding correlations, approximations to the means and standard deviations of parameter estimates and the skewness of the parameter estimates. Now, in some cases, particularly for smaller sample sizes, the maximum likelihood estimates were not obtained; in these instances a and α_b were usually seen to increase without limit. Consequently, these values are not included in the sample, and we indicate the percentage of cases (out of 10000 replications) that were included in the sample.

For the majority of cases, the terms in the EFI matrix get closer to their theoretical counterparts with increasing sample size. Repeated experimental runs suggested that, for those terms whose simulated values do not get progressively closer to their theoretical values, the cause was simulation variation: different runs with the same parameter values gave slightly different results each time. Since correlations and standard deviations in Table 3.2 are based on the inverse of the EFI matrix, such variations will naturally propagate and consequently we occasionally observe fluctuations in simulated values as we increase the sample size.

It is worth noting the similarity between correlations observed here for the Burr XII ALT model and those illustrated by Watkins (1991) for the Weibull ALT model. There, Watkins noted that the theoretical correlation between maximum likelihood estimates of B and β_w was zero, while there was a considerable negative correlation between the MLEs of α_w and β_w (-0.8957). This is consistent with the Burr XII ALT model, where the asymptotic theoretical correlations between MLEs of shape parameters and β_b are zero, while between the MLEs of α_b and β_b, the correlations are negative. Considering the zero correlations further, we now show that this is always true. Firstly, since the correlations are zero then the covariances are also zero. Now, with the EFI matrix given as (3.21), it is sufficient to show that the two minors are zero. The first is given by the determinant of

$$-\begin{bmatrix} E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right] & E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right] & E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right] \\ E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right] & E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right] & E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right] \\ E\left[\frac{\partial^2 l_B}{\partial \alpha_b \partial \tau}\right] & E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right] & E\left[\frac{\partial^2 l_B}{\partial \alpha_b \partial \beta_b}\right] \end{bmatrix},$$

which is,

$$\begin{aligned} &-E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right] \times \left\{E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right] E\left[\frac{\partial^2 l_B}{\partial \alpha_b \partial \beta_b}\right] - E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right] E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right]\right\} \\ &+E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right] \times \left\{E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right] E\left[\frac{\partial^2 l_B}{\partial \alpha_b \partial \beta_b}\right] - E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right] E\left[\frac{\partial^2 l_B}{\partial \alpha_b \partial \tau}\right]\right\} \\ &-E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right] \times \left\{E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right] E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right] - E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right] E\left[\frac{\partial^2 l_B}{\partial \alpha_b \partial \tau}\right]\right\}, \end{aligned}$$

and, in full, is,

$$\begin{aligned}
&\frac{S_n\left(1-\gamma-\psi\left(a+1\right)\right)}{\tau\left(a+1\right)}\left\{\frac{S_n a\left(1-\gamma-\psi\left(a+1\right)\right)}{a+2}\times\left(-\frac{S_x\tau^2 a}{a+2}\right)\right\}\\
&-\frac{S_n\left(1-\gamma-\psi\left(a+1\right)\right)}{\tau\left(a+1\right)}\left\{\frac{S_x a\left(1-\gamma-\psi\left(a+1\right)\right)}{a+2}\times\left(-\frac{S_n\tau^2 a}{a+2}\right)\right\}\\
&+\frac{S_n\tau}{a+1}\left\{\frac{S_n}{\tau^2\left(a+2\right)}\left[2+a\left\{\begin{array}{c}\frac{\pi^2}{6}+\left(1-\gamma\right)^2-2\left(1-\gamma\right)\psi\left(a+1\right)\\+\left[\psi\left(a+1\right)\right]^2+\psi'\left(a+1\right)\end{array}\right\}\right]\times\left(-\frac{S_x\tau^2 a}{a+2}\right)\right\}\\
&-\frac{S_n\tau}{a+1}\left\{\frac{S_x a\left(1-\gamma-\psi\left(a+1\right)\right)}{a+2}\times\frac{S_n a\left(1-\gamma-\psi\left(a+1\right)\right)}{a+2}\right\}\\
&-\frac{S_x\tau}{a+1}\left\{\frac{S_n}{\tau^2\left(a+2\right)}\left[2+a\left\{\begin{array}{c}\frac{\pi^2}{6}+\left(1-\gamma\right)^2-2\left(1-\gamma\right)\psi\left(a+1\right)\\+\left[\psi\left(a+1\right)\right]^2+\psi'\left(a+1\right)\end{array}\right\}\right]\times\left(-\frac{S_n\tau^2 a}{a+2}\right)\right\}\\
&+\frac{S_x\tau}{a+1}\left\{\frac{S_n a\left(1-\gamma-\psi\left(a+1\right)\right)}{a+2}\times\frac{S_n a\left(1-\gamma-\psi\left(a+1\right)\right)}{a+2}\right\}.
\end{aligned}$$

Through inspection, we see that the first term cancels with the final term, the second term cancels with the fourth and the third term cancels with the fifth, giving the final result of zero, as required. The second minor is given by the determinant of

$$-\begin{bmatrix}
E\left[\frac{\partial^2 l_B}{\partial a^2}\right] & E\left[\frac{\partial^2 l_B}{\partial a\partial\alpha_b}\right] & E\left[\frac{\partial^2 l_B}{\partial a\partial\beta_b}\right]\\
E\left[\frac{\partial^2 l_B}{\partial a\partial\tau}\right] & E\left[\frac{\partial^2 l_B}{\partial\tau\partial\alpha_b}\right] & E\left[\frac{\partial^2 l_B}{\partial\tau\partial\beta_b}\right]\\
E\left[\frac{\partial^2 l_B}{\partial a\partial\alpha_b}\right] & E\left[\frac{\partial^2 l_B}{\partial\alpha_b^2}\right] & E\left[\frac{\partial^2 l_B}{\partial\alpha_b\partial\beta_b}\right]
\end{bmatrix},$$

and is written as,

$$\begin{aligned}
&-E\left[\frac{\partial^2 l_B}{\partial a^2}\right]\times\left\{E\left[\frac{\partial^2 l_B}{\partial\tau\partial\alpha_b}\right]E\left[\frac{\partial^2 l_B}{\partial\alpha_b\partial\beta_b}\right]-E\left[\frac{\partial^2 l_B}{\partial\tau\partial\beta_b}\right]E\left[\frac{\partial^2 l_B}{\partial\alpha_b^2}\right]\right\}\\
&+E\left[\frac{\partial^2 l_B}{\partial a\partial\alpha_b}\right]\times\left\{E\left[\frac{\partial^2 l_B}{\partial a\partial\tau}\right]E\left[\frac{\partial^2 l_B}{\partial\alpha_b\partial\beta_b}\right]-E\left[\frac{\partial^2 l_B}{\partial\tau\partial\beta_b}\right]E\left[\frac{\partial^2 l_B}{\partial a\partial\alpha_b}\right]\right\}\\
&-E\left[\frac{\partial^2 l_B}{\partial a\partial\beta_b}\right]\times\left\{E\left[\frac{\partial^2 l_B}{\partial a\partial\tau}\right]E\left[\frac{\partial^2 l_B}{\partial\alpha_b^2}\right]-E\left[\frac{\partial^2 l_B}{\partial\tau\partial\alpha_b}\right]E\left[\frac{\partial^2 l_B}{\partial a\partial\alpha_b}\right]\right\},
\end{aligned}$$

and, in full, is,

$$\begin{aligned}
&\frac{S_n}{a^2}\frac{S_n a\left(1-\gamma-\psi\left(a+1\right)\right)}{a+2}\left(-\frac{S_x\tau^2 a}{a+2}\right)-\frac{S_n}{a^2}\frac{S_x a\left(1-\gamma-\psi\left(a+1\right)\right)}{a+2}\left(-\frac{S_n\tau^2 a}{a+2}\right)\\
&+\frac{S_n\tau}{a+1}\frac{S_n\left(1-\gamma-\psi\left(a\right)\right)}{\tau\left(a+1\right)}\left(-\frac{S_x\tau^2 a}{a+2}\right)-\frac{S_n\tau}{a+1}\frac{S_x a\left(1-\gamma-\psi\left(a+1\right)\right)}{a+2}\left(-\frac{S_n\tau}{a+1}\right)\\
&-\frac{S_x\tau}{a+1}\frac{S_n\left(1-\gamma-\psi\left(a\right)\right)}{\tau\left(a+1\right)}\left(-\frac{S_n\tau^2 a}{a+2}\right)+\frac{S_x\tau}{a+1}\frac{S_n a\left(1-\gamma-\psi\left(a+1\right)\right)}{a+2}\left(-\frac{S_n\tau}{a+1}\right).
\end{aligned}$$

Again, by inspection we can see that the first term cancels the second, the third cancels the fifth and the fourth cancels the sixth, giving zero as the required final result. Therefore,

	$n_1^{-1}\times$ Average Simulated EFI Element				Theoretical
	$n_1 = 30$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
$-E\left[\frac{\partial^2 l_B}{\partial a^2}\right]$	0.75	0.75	0.75	0.75	0.75
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right]$	0.000115035	−0.000067	0.000072	0.000024	0
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right]$	−3.00108	−3.00092	−3.00001	−3.00028	−3
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right]$	−510.193	−510.173	−510	−510.04	−510
$-E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right]$	0.548153	0.548259	0.548309	0.548287	0.548311
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right]$	0.747823	0.749375	0.749934	0.74957	0.75
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right]$	127.129	127.378	127.486	127.436	127.5
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right]$	13.502	13.5056	13.4993	13.5008	13.5
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b \beta_b}\right]$	2295.38	2296.06	2294.88	2295.13	2295
$-E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right]$	393823	393950	393732	393772	393750

Table 3.1: Simulated and theoretical elements from the Burr XII ALT EFI matrix, with shape parameters $a = 2$ and $\tau = 3$.

since the determinants of these two 3×3 matrices are zero, the inverse of the full 4×4 matrix will give zeros at positions $(4, 1)$ and $(4, 2)$.

It is perhaps intuitive that the correlations between a and β_b and between τ and β_b will be zero, since we do not naturally expect the coefficient of the stress factor - effectively a scale parameter - to have any bearing on the shape of the distribution. Further discussion on interpretation of the correlations is omitted, since this investigation will require a more detailed examination of the theoretical matrices than is appropriate here.

3.2.3 Other Results For $a = 2$

We now perform a similar analysis using the shape parameters $a = 2$ and $\tau = 5$. Results are presented in Tables 3.3 and 3.4. As with the previous example, the extent to which the simulated values strictly approach their theoretical counterparts with increasing sample size, varies. Nevertheless, in all cases the simulated values are close to the theoretical values. Our next example considers the same experiment but with shape parameters $a = 2$ and $\tau = 7$. The results are presented in Tables 3.5 and 3.6. Both tables show rather good agreement between simulated and theoretical values, and provide further evidence supporting the expressions for the theoretical expectations established in section 3.2.4.3.

3.2.4 Results For $a = 5$

Our next example uses the same parameter values for k, n_i, x_i and for α_b and β_b as above but with $a = 5$ and $\tau = 3$. The results are shown in Tables 3.7 and 3.8. Again, simulated and theoretical values match up rather well in all cases. Performing the same analysis

	Simulated Value				Theoretical
	$n_1 = 30$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
% valid MLEs	92.71	99.97	100.00	100.00	100.00
mean$(\widehat{a})$	3.93102	2.31662	2.01941	2.00779	2.0
mean$(\widehat{\tau})$	3.04897	3.01106	3.00183	3.00013	3.0
mean$(\widehat{\alpha}_b)$	17.70504	17.63806	17.60246	17.60169	17.60139
mean$\left(\widehat{\beta}_b\right)$	−0.05625	−0.056288	−0.056272	−0.056276	−0.056282
$\sqrt{n_1}\times$s.d.$(\widehat{a})$	52.47341	12.13306	6.044758	6.033097	5.92282
$\sqrt{n_1}\times$s.d.$(\widehat{\tau})$	2.29877	2.47633	2.39734	2.39736	2.40207
$\sqrt{n_1}\times$s.d.$(\widehat{\alpha}_b)$	3.40227	3.33985	3.14820	3.15134	3.18385
$\sqrt{n_1}\times$s.d.$\left(\widehat{\beta}_b\right)$	0.016958	0.016827	0.016503	0.016549	0.0166667
skewness$(\widehat{a})$	13.798	9.999	0.461	0.281	0
skewness$(\widehat{\tau})$	0.523	0.379	0.111	0.099	0
skewness$(\widehat{\alpha}_b)$	0.266	0.162	0.019	0.005	0
skewness$\left(\widehat{\beta}_b\right)$	−0.039	0.018	0.001	−0.018	0
corr$(\widehat{a}, \widehat{\tau})$	−0.766	−0.806	−0.799	−0.803	−0.811125
corr$(\widehat{a}, \widehat{\alpha}_b)$	0.446	0.461	0.430	0.426	0.447391
corr$\left(\widehat{a}, \widehat{\beta}_b\right)$	0.020	−0.002	0.003	0.016	0
corr$(\widehat{\tau}, \widehat{\alpha}_b)$	−0.366	−0.386	−0.362	−0.358	−0.377227
corr$\left(\widehat{\tau}, \widehat{\beta}_b\right)$	−0.012	−0.003	−0.005	−0.016	0
corr$\left(\widehat{\alpha}_b, \widehat{\beta}_b\right)$	−0.846	−0.858	−0.879	−0.875	−0.889906

Table 3.2: Simulated and theoretical standard deviations and correlations of the MLEs from the Burr XII ALT model, with shape parameters $a = 2$ and $\tau = 3$.

	$n_1^{-1}\times$ Average Simulated EFI Element				Theoretical
	$n_1 = 30$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
$-E\left[\frac{\partial^2 l_B}{\partial a^2}\right]$	0.75	0.75	0.75	0.75	0.75
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right]$	0.0006301	0.0000519	0.0000519	0.000031	0
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right]$	−5.00847	−5.00053	−5.00089	−5.00027	−5
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right]$	−851.413	−850.114	−850.153	−850.06	−850
$-E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right]$	0.197346	0.197347	0.197381	0.197385	0.1973921
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right]$	0.741897	0.748954	0.749423	0.749703	0.75
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right]$	126.145	127.31	127.395	127.44	127.5
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right]$	37.514	37.5006	37.5057	37.5	37.5
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b \beta_b}\right]$	6377.1	6375.21	6375.94	6375.1	6375
$-E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right]$	1094060	1093800	1093910	1093780	1093750

Table 3.3: Simulated and theoretical elements from the Burr XII ALT EFI matrix, with shape parameters $a = 2$ and $\tau = 5$.

	Simulated Value				Theoretical
	$n_1 = 30$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
% valid MLEs	94.20	99.91	100.00	100.00	100.00
mean$(\widehat{a})$	4.28442	2.29158	2.021652	2.006107	2
mean$(\widehat{\tau})$	5.09002	5.02995	5.003214	5.001002	5
mean$(\widehat{\alpha}_b)$	17.66571	17.61951	17.602796	17.60098	17.60139
mean$\left(\widehat{\beta}_b\right)$	−0.056262	−0.056277	−0.056279	−0.056276	−0.056282
$\sqrt{n_1}\times$s.d.$(\widehat{a})$	76.96717	11.33064	6.099483	5.83556	5.92282
$\sqrt{n_1}\times$s.d.$(\widehat{\tau})$	4.00383	4.11344	4.016026	3.90727	4.00346
$\sqrt{n_1}\times$s.d.$(\widehat{\alpha}_b)$	2.09438	1.98445	1.908771	1.89548	1.91031
$\sqrt{n_1}\times$s.d.$\left(\widehat{\beta}_b\right)$	0.010210	0.010013	0.009993	0.010011	0.01
skewness$(\widehat{a})$	19.581	6.176	0.442	0.246	0
skewness$(\widehat{\tau})$	0.535	0.292	0.088	0.084	0
skewness$(\widehat{\alpha}_b)$	0.310	0.158	0.059	0.024	0
skewness$\left(\widehat{\beta}_b\right)$	−0.010	−0.052	−0.022	−0.041	0
corr$(\widehat{a},\widehat{\tau})$	−0.785	−0.800	−0.799	−0.792	−0.811125
corr$(\widehat{a},\widehat{\alpha}_b)$	0.474	0.458	0.429	0.407	0.447391
corr$\left(\widehat{a},\widehat{\beta}_b\right)$	0.006	0.006	0.004	0.016	0
corr$(\widehat{\tau},\widehat{\alpha}_b)$	−0.394	−0.398	−0.368	−0.338	−0.377227
corr$\left(\widehat{\tau},\widehat{\beta}_b\right)$	−0.004	0.007	0.004	−0.018	0
corr$\left(\widehat{\alpha}_b,\widehat{\beta}_b\right)$	−0.837	−0.854	−0.879	−0.885	−0.889906

Table 3.4: Simulated and theoretical standard deviations and correlations of the MLEs from the Burr XII ALT model, with shape parameters $a = 2$ and $\tau = 5$.

	$n_1^{-1}\times$ Average Simulated EFI Element				Theoretical
	$n_1 = 30$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
$-E\left[\frac{\partial^2 l_B}{\partial a^2}\right]$	0.75	0.75	0.75	0.75	0.75
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right]$	0.00005	−0.00006	0.00001	0.00002	0
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right]$	−7.00697	−6.99839	−7.00052	−7.004	−7
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right]$	−1191.33	−1189.76	−1190.09	−1190.08	−1190
$-E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right]$	0.1007	0.10072	0.100714	0.10071	0.10071
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right]$	0.74706	0.751014	0.749844	0.749697	0.75
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right]$	126.922	127.644	127.47	127.441	127.5
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right]$	73.5637	73.4946	73.5025	73.499	73.5
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b \beta_b}\right]$	12506.7	12493.8	12495.5	12494.8	12495
$-E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right]$	2145920	2143490	2143840	2143730	2143750

Table 3.5: Simulated and theoretical elements from the Burr XII ALT EFI matrix, with shape parameters $a = 2$ and $\tau = 7$.

	Simulated Value				Theoretical
	$n_1 = 30$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
% valid MLEs	93.98	99.96	100.00	100.00	100.00
mean$(\widehat{a})$	4.25586	2.31116	2.02143	2.00725	2
mean$(\widehat{\tau})$	7.14507	7.03608	7.00207	7.00091	7
mean$(\widehat{\alpha}_b)$	17.64848	17.61519	17.60268	17.60209	17.60139
mean$\left(\widehat{\beta}_b\right)$	−0.056274	−0.056278	−0.056282	−0.056283	−0.056282
$\sqrt{n_1}\times$s.d.$(\widehat{a})$	71.85744	13.66082	6.12376	6.05583	5.92282
$\sqrt{n_1}\times$s.d.$(\widehat{\tau})$	12.15888	5.822027	5.56507	5.66461	5.60484
$\sqrt{n_1}\times$s.d.$(\widehat{\alpha}_b)$	1.48763	1.437580	1.36734	1.35997	1.36451
$\sqrt{n_1}\times$s.d.$\left(\widehat{\beta}_b\right)$	0.0071848	0.0072137	0.0070804	0.0071132	0.00714286
skewness$(\widehat{a})$	18.190	13.537	0.535	0.341	0
skewness$(\widehat{\tau})$	67.527	0.370	0.116	0.038	0
skewness$(\widehat{\alpha}_b)$	0.318	0.165	0.045	0.013	0
skewness$\left(\widehat{\beta}_b\right)$	0.015	0.012	−0.014	0.000	0
corr$(\widehat{a}, \widehat{\tau})$	−0.783	−0.803	−0.793	−0.800	−0.811125
corr$(\widehat{a}, \widehat{\alpha}_b)$	0.484	0.474	0.445	0.430	0.447391
corr$\left(\widehat{a}, \widehat{\beta}_b\right)$	−0.010	−0.012	−0.011	0.007	0
corr$(\widehat{\tau}, \widehat{\alpha}_b)$	−0.404	−0.394	−0.366	−0.358	−0.377227
corr$\left(\widehat{\tau}, \widehat{\beta}_b\right)$	0.017	0.001	0.002	−0.012	0
corr$\left(\widehat{\alpha}_b, \widehat{\beta}_b\right)$	−0.841	−0.856	−0.878	−0.877	−0.889906

Table 3.6: Simulated and theoretical standard deviations and correlations of the MLEs from the Burr XII ALT model, with shape parameters $a = 2$ and $\tau = 7$.

	$n_1^{-1}\times$ Average Simulated EFI Element				Theoretical
	$n_1 = 30$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
$-E\left[\frac{\partial^2 l_B}{\partial a^2}\right]$	0.12	0.12	0.12	0.12	0.12
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right]$	−0.18057	−0.180595	−0.180585	−0.180568	−0.180556
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right]$	−1.49814	−1.49976	−1.50018	−1.49986	−1.5
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right]$	−254.71	−254.937	−255.029	−254.978	−255
$-E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right]$	0.922473	0.922547	0.922165	0.922213	0.922191
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right]$	2.75391	2.75158	2.74993	2.7504	2.75
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right]$	468.08	467.799	467.49	467.56	467.5
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right]$	19.2683	19.2841	19.2886	19.2848	19.2857
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b \beta_b}\right]$	3275.77	3278.04	3279.05	3278.43	3278.57
$-E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right]$	562050	562366	562580	562479	562500

Table 3.7: Simulated and theoretical elements from the Burr XII ALT EFI matrix, with shape parameters $a = 5$ and $\tau = 3$.

for $a = 5$ and $\tau = 5$, we obtain the results shown in Tables 3.9 and 3.10. As with all previous examples, the agreement between simulated and theoretical results is good and seems to improve as we increase sample size. Our final adjustment for this example has shape parameters of $a = 5$ and $\tau = 7$. The results are presented in Tables 3.11 and 3.12. Yet again, we observe good agreement between simulated and theoretical results, which seems to improve with increasing sample size.

3.3 Simulation 2: Four Stress Levels

We now consider a second experiment, with a different number of stress levels and suitably large range of parameter values. The choice of the shape parameters is now $a = 2$ or 5 and $\tau = 2, 4$ or 6; with $\beta_b = -0.01$ and $\alpha_b = 5.5$. This time, we have $k = 4$ stress levels, $(x_1, x_2, x_3, x_4) = (10, 80, 150, 240)$, with equal sub-sample sizes, and n_1 taken to be 25, 100, 1000 and 3000. The approximate scale parameters are then $(\theta_1, \theta_2, \theta_3, \theta_4) = (221, 109, 54, 22)$. As with the previous simulation, we present the EFI matrix - analogous to (3.22) - and its inverse - analogous to (3.23) - for $a = 2$ and $\tau = 2$ as

$$n_1 \begin{bmatrix} & 1 & & & \\ & 0 & 1.64493\cdots & & \\ -2.66667\cdots & & 1 & 8 & \\ -320 & & 120 & 960 & 173200 \end{bmatrix}$$

	Simulated Value				Theoretical
	$n_1 = 30$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
% valid MLEs	74.88	94.70	100.00	100.00	100.00
mean$(\widehat{a})$	10.04393	9.35930	5.21678	5.06651	5
mean$(\widehat{\tau})$	3.18404	3.02549	3.00003	3.000164	3
mean$(\widehat{\alpha}_b)$	17.59748	17.67008	17.61153	17.60368	17.60139
mean$\left(\widehat{\beta}_b\right)$	−0.056255	−0.056250	−0.056284	−0.056277	−0.056282
$\sqrt{n_1}\times$s.d.$(\widehat{a})$	135.7306	186.6651	33.62628	29.96307	28.4309
$\sqrt{n_1}\times$s.d.$(\widehat{\tau})$	2.2849	2.21533	2.26944	2.27526	2.25459
$\sqrt{n_1}\times$s.d.$(\widehat{\alpha}_b)$	3.2309	3.89157	3.52540	3.46710	3.43394
$\sqrt{n_1}\times$s.d.$\left(\widehat{\beta}_b\right)$	0.014193	0.014020	0.0138704	0.013937	0.0139443
skewness$(\widehat{a})$	14.656	8.859	1.402	0.598	0
skewness$(\widehat{\tau})$	0.865	0.445	0.079	0.055	0
skewness$(\widehat{\alpha}_b)$	0.287	0.715	0.270	0.052	0
skewness$\left(\widehat{\beta}_b\right)$	0.003	0.049	0.012	0.037	0
corr$(\widehat{a}, \widehat{\tau})$	−0.712	−0.762	−0.790	−0.784	−0.793006
corr$(\widehat{a}, \widehat{\alpha}_b)$	0.607	0.715	0.710	0.708	0.7818193
corr$\left(\widehat{a}, \widehat{\beta}_b\right)$	0.001	0.008	0.001	0.001	0
corr$(\widehat{\tau}, \widehat{\alpha}_b)$	−0.459	−0.568	−0.603	−0.594	−0.604278
corr$\left(\widehat{\tau}, \widehat{\beta}_b\right)$	−0.005	−0.026	0.006	0.000	0
corr$\left(\widehat{\alpha}_b, \widehat{\beta}_b\right)$	−0.747	−0.637	−0.661	−0.665	−0.690325

Table 3.8: Simulated and theoretical standard deviations and correlations of the MLEs from the Burr XII ALT model, with shape parameters $a = 5$ and $\tau = 3$.

	$n_1^{-1}\times$ Average Simulated EFI Element				Theoretical
	$n_1 = 30$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
$-E\left[\frac{\partial^2 l_B}{\partial a^2}\right]$	0.12	0.12	0.12	0.12	0.12
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right]$	−0.108347	−0.108277	−0.108338	−0.108325	−0.108333
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right]$	−2.5014	−2.49945	−2.50018	−2.49998	−2.5
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right]$	−425.277	−424.908	−425.031	−424.994	−425
$-E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right]$	0.33183	0.331993	0.331955	0.331998	0.331989
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right]$	2.74753	2.74978	2.74969	2.74994	2.75
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right]$	466.993	467.452	467.444	467.491	467.5
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right]$	53.5977	53.5528	53.5758	53.5693	53.5714
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b \beta_b}\right]$	9111.8	9103.89	9107.86	9106.75	9107.14
$-E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right]$	1563340	1561930	1562620	1562430	1562500

Table 3.9: Simulated and theoretical elements from the Burr XII ALT EFI matrix, with shape parameters $a = 5$ and $\tau = 5$.

	Simulated Value				Theoretical
	$n_1 = 30$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
% valid MLEs	75.34	95.18	100.00	100.00	100.00
mean$(\widehat{a})$	10.96528	9.99898	5.21488	5.05974	5
mean$(\widehat{\tau})$	5.28461	5.04795	5.00167	5.00040	5
mean$(\widehat{\alpha}_b)$	17.61227	17.64996	17.60698	17.603012	17.60139
mean$\left(\widehat{\beta}_b\right)$	−0.056330	−0.056301	−0.056281	−0.056282	−0.056282
$\sqrt{n_1}\times$s.d.$(\widehat{a})$	170.32033	252.8482	33.5082	29.57480	28.4309
$\sqrt{n_1}\times$s.d.$(\widehat{\tau})$	3.97966	3.73813	3.81067	3.75346	3.75765
$\sqrt{n_1}\times$s.d.$(\widehat{\alpha}_b)$	1.98491	2.39457	2.11219	2.07430	2.06037
$\sqrt{n_1}\times$s.d.$\left(\widehat{\beta}_b\right)$	0.008584	0.0085505	0.0084108	0.008355	0.0083666
skewness$(\widehat{a})$	12.302	12.120	1.636	0.640	0
skewness$(\widehat{\tau})$	0.594	0.439	0.119	0.065	0
skewness$(\widehat{\alpha}_b)$	0.350	0.799	0.236	0.099	0
skewness$\left(\widehat{\beta}_b\right)$	0.031	0.019	0.034	0.024	0
corr$(\widehat{a}, \widehat{\tau})$	−0.713	−0.762	−0.784	−0.779	−0.793006
corr$(\widehat{a}, \widehat{\alpha}_b)$	0.607	0.710	0.703	0.699	0.718193
corr$\left(\widehat{a}, \widehat{\beta}_b\right)$	−0.003	0.002	0.006	−0.004	0
corr$(\widehat{\tau}, \widehat{\alpha}_b)$	−0.466	−0.582	−0.588	−0.589	−0.604278
corr$\left(\widehat{\tau}, \widehat{\beta}_b\right)$	0.010	0.007	−0.011	0.014	0
corr$\left(\widehat{\alpha}_b, \widehat{\beta}_b\right)$	−0.749	−0.647	−0.666	−0.678	−0.690325

Table 3.10: Simulated and theoretical standard deviations and correlations of the MLEs from the Burr XII ALT model, with shape parameters $a = 5$ and $\tau = 5$.

	$n_1^{-1}\times$Average Simulated EFI Element				Theoretical
	$n_1 = 30$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
$-E\left[\frac{\partial^2 l_B}{\partial a^2}\right]$	0.12	0.12	0.12	0.12	0.12
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right]$	−0.0774859	−0.0773846	−0.0773706	−0.0773881	−0.077381
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right]$	−3.49935	−3.50237	−3.49992	−3.50012	−3.5
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right]$	−594.907	−595.39	−594.992	−595.029	−595
$-E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right]$	0.169362	0.169363	0.169363	0.169379	0.169382
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right]$	2.75252	2.74806	2.74989	2.75005	2.75
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right]$	467.943	467.185	467.473	467.5	467.5
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right]$	105.04	105.054	104.993	105.006	105
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b \beta_b}\right]$	17857.7	17858.9	17848.9	17851.3	17850
$-E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right]$	3063960	3063980	3063980	3062760	3062500

Table 3.11: Simulated and theoretical elements from the Burr XII ALT EFI matrix, with shape parameters $a = 5$ and $\tau = 7$.

	Simulated Value				Theoretical
	$n_1 = 30$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
% valid MLEs	75.43	95.16	100.00	100.00	100.00
mean$(\widehat{a})$	10.06161	9.73120	5.21719	5.06650	5
mean$(\widehat{\tau})$	7.43269	7.06477	6.99958	7.00037	7
mean$(\widehat{\alpha}_b)$	17.600753	17.63377	17.60603	17.60279	17.60139
mean$\left(\widehat{\beta}_b\right)$	−0.056287	−0.056283	−0.056284	−0.056282	−0.056282
$\sqrt{n_1}\times$s.d.$(\widehat{a})$	145.95794	232.7660	33.55176	29.78438	28.4309
$\sqrt{n_1}\times$s.d.$(\widehat{\tau})$	5.445185	5.25221	5.31631	5.23886	5.2607
$\sqrt{n_1}\times$s.d.$(\widehat{\alpha}_b)$	1.381558	1.70159	1.50634	1.47534	1.47169
$\sqrt{n_1}\times$s.d.$\left(\widehat{\beta}_b\right)$	0.006070	0.0060918	0.005953	0.0059689	0.00597614
skewness$(\widehat{a})$	10.831	12.993	1.663	0.683	0
skewness$(\widehat{\tau})$	0.894	0.439	0.092	0.040	0
skewness$(\widehat{\alpha}_b)$	0.286	0.693	0.216	0.121	0
skewness$\left(\widehat{\beta}_b\right)$	−0.024	0.031	0.028	0.027	0
corr$(\widehat{a},\widehat{\tau})$	−0.716	−0.767	−0.783	−0.772	−0.793006
corr$(\widehat{a},\widehat{\alpha}_b)$	0.612	0.722	0.708	0.703	0.718193
corr$\left(\widehat{a},\widehat{\beta}_b\right)$	−0.006	−0.011	−0.009	0.003	0
corr$(\widehat{\tau},\widehat{\alpha}_b)$	−0.461	−0.596	−0.589	−0.584	−0.604278
corr$\left(\widehat{\tau},\widehat{\beta}_b\right)$	−0.010	0.016	0.000	−0.002	0
corr$\left(\widehat{\alpha}_b,\widehat{\beta}_b\right)$	−0.749	−0.644	−0.671	−0.667	−0.690325

Table 3.12: Simulated and theoretical standard deviations and correlations of the MLEs from the Burr XII ALT model, with shape parameters $a = 5$ and $\tau = 7$, for three stress levels.

	$n_1^{-1}\times$ Average Simulated EFI Element				Theoretical
	$n_1 = 25$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
$-E\left[\frac{\partial^2 l_B}{\partial a^2}\right]$	1	1	1	1	1
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right]$	−0.0005759	0.0000372	0.0001039	0.0000029	0
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right]$	−2.66629	−2.66667	−2.66668	−2.66655	−2.66667
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right]$	−319.927	−319.959	−319.995	−319.976	−320
$-E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right]$	1.64468	1.64488	1.64496	1.64499	1.64493
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right]$	1.00037	0.999544	0.999924	1.00023	1
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right]$	120.095	120.047	119.998	120.065	120
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right]$	8.0012	7.99962	7.99964	7.99963	8
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b \beta_b}\right]$	960.192	959.997	959.92	959.98	960
$-E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right]$	173253	173190	173182	173197	173200

Table 3.13: Simulated and theoretical elements from the Burr XII ALT EFI matrix, with shape parameters $a = 2$ and $\tau = 2$, for four stress levels.

and

$$\frac{1}{n_1}\begin{bmatrix} 26.3099\cdots & & & \\ -5.76996\cdots & 1.92332\cdots & & \\ 9.49121\cdots & -2.16374\cdots & 3.80748\cdots & \\ 0 & 0 & -0.00206\cdots & 0.0000172\cdots \end{bmatrix},$$

respectively. With regard to simulations, we perform 10000 replications and the results are presented in Tables 3.13 and 3.14. We can see that, with the exception of the standard deviation of $\widehat{a}$, there is rather good agreement between simulated and theoretical values, even for smaller sample sizes.

3.3.1 Other Results For $a = 2$

We now consider the same set of parameter values but with the minor alteration, $\tau = 4$. The results are given in Tables 3.15 and 3.16, and, as before, we observe good agreement between simulated and theoretical results. We then consider the same set of parameter values but now with the minor change, $\tau = 6$. The results are given in Tables 3.17 and 3.18 and the good agreement between simulated and theoretical results further increases our confidence in the theoretical results established in section 3.2.4.3.

3.3.2 Results For $a = 5$.

Using the values $a = 5$ and $\tau = 2$, keeping all other parameter values the same, we perform the same analysis; the results are presented in Tables 3.19 and 3.20. Again, we observe good agreement between simulated and theoretical results throughout both tables. The

	Simulated Value				Theoretical
	$n_1 = 25$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
% valid MLEs	94.63	100.00	100.00	100.00	
mean$(\widehat{a})$	3.66738	2.19165	2.01658	2.005133	2
mean$(\widehat{\tau})$	2.01770	2.00973	2.00088	2.000259	2
mean$(\widehat{\alpha}_b)$	5.67372	5.42892	5.503561	5.501142	5.5
mean$\left(\widehat{\beta}_b\right)$	−0.01001	−0.009997	−0.0100001	−0.0100001	−0.01
$\sqrt{n_1}\times$s.d.$(\widehat{a})$	36.34330	7.79892	5.279216	5.167248	5.12932
$\sqrt{n_1}\times$s.d.$(\widehat{\tau})$	1.32552	1.42297	1.37503	1.390346	1.38684
$\sqrt{n_1}\times$s.d.$(\widehat{\alpha}_b)$	2.45769	2.17297	1.96850	1.953105	1.95128
$\sqrt{n_1}\times$s.d.$\left(\widehat{\beta}_b\right)$	0.004227	0.004184	0.0040947	0.0041266	0.00415227
skewness$(\widehat{a})$	13.790	3.844	0.432	0.203	0
skewness$(\widehat{\tau})$	0.458	0.293	0.081	0.079	0
skewness$(\widehat{\alpha}_b)$	1.499	0.844	0.177	0.054	0
skewness$\left(\widehat{\beta}_b\right)$	−0.012	−0.019	0.015	−0.020	0
corr$(\widehat{a}, \widehat{\tau})$	−0.772	−0.807	−0.795	−0.801	−0.811125
corr$(\widehat{a}, \widehat{\alpha}_b)$	0.938	0.945	0.947	0.946	0.948295
corr$\left(\widehat{a}, \widehat{\beta}_b\right)$	0.013	0.008	−0.007	−0.001	0
corr$(\widehat{\tau}, \widehat{\alpha}_b)$	−0.767	−0.796	−0.785	−0.789	−0.799576
corr$\left(\widehat{\tau}, \widehat{\beta}_b\right)$	0.006	−0.002	0.009	−0.002	0
corr$\left(\widehat{\alpha}_b, \widehat{\beta}_b\right)$	−0.234	−0.230	−0.243	−0.239	−0.25537

Table 3.14: Simulated and theoretical standard deviations and correlations of the MLEs from the Burr XII ALT model, with shape parameters $a = 2$ and $\tau = 2$.

	$n_1^{-1}\times$ Average Simulated EFI Element				Theoretical
	$n_1 = 25$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
$-E\left[\frac{\partial^2 l_B}{\partial a^2}\right]$	1	1	1	1	1
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right]$	0.0007184	0.0000805	0.0000201	−0.0000237	0
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right]$	−5.3344	−5.33411	−5.33349	−5.3333	−5.33333
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right]$	−640.028	−639.784	−640.045	−639.95	−640
$-E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right]$	0.411312	0.411174	0.411235	0.411237	0.411234
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right]$	0.999164	0.998617	0.999972	1.0001	1
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right]$	119.954	120.167	119.989	120.053	120
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right]$	31.9866	32.0012	32.0006	32.0005	32
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b \beta_b}\right]$	3839.24	3838.63	3840.31	3839.97	3840
$-E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right]$	692752	692438	692874	692797	692800

Table 3.15: Simulated and theoretical elements from the Burr XII ALT EFI matrix, with shape parameters $a = 2$ and $\tau = 4$.

	Simulated Value				Theoretical Value
	$n_1 = 25$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	
% valid MLEs	95.22	100.00	100.00	100.00	
mean$(\widehat{a})$	3.98581	2.18991	2.016092	2.004496	2
mean$(\widehat{\tau})$	4.05985	4.02181	4.000866	4.001116	4
mean$(\widehat{\alpha}_b)$	5.58090	5.41198	5.501714	5.500339	5.5
mean$\left(\widehat{\beta}_b\right)$	−0.010006	−0.0099985	−0.0099997	−0.009999	−0.01
$\sqrt{n_1}\times$s.d.$(\widehat{a})$	55.06266	7.61164	5.311191	5.192484	5.12932
$\sqrt{n_1}\times$s.d.$(\widehat{\tau})$	2.76497	2.87793	2.776967	2.783386	2.77368
$\sqrt{n_1}\times$s.d.$(\widehat{\alpha}_b)$	1.30002	1.09047	0.984610	0.980858	0.975638
$\sqrt{n_1}\times$s.d.$\left(\widehat{\beta}_b\right)$	0.002131	0.0020762	0.0020804	0.0020791	0.00207614
skewness$(\widehat{a})$	17.119	2.839	0.489	0.204	0
skewness$(\widehat{\tau})$	0.505	0.322	0.113	0.099	0
skewness$(\widehat{\alpha}_b)$	1.606	0.709	0.222	0.064	0
skewness$\left(\widehat{\beta}_b\right)$	−0.004	−0.029	−0.011	−0.005	0
corr$(\widehat{a}, \widehat{\tau})$	−0.789	−0.808	−0.796	−0.796	−0.811125
corr$(\widehat{a}, \widehat{\alpha}_b)$	0.943	0.948	0.944	0.946	0.948295
corr$\left(\widehat{a}, \widehat{\beta}_b\right)$	−0.010	−0.001	0.015	0.004	0
corr$(\widehat{\tau}, \widehat{\alpha}_b)$	−0.783	−0.798	−0.782	−0.785	−0.799576
corr$\left(\widehat{\tau}, \widehat{\beta}_b\right)$	0.016	0.006	−0.023	−0.006	0
corr$\left(\widehat{\alpha}_b, \widehat{\beta}_b\right)$	−0.245	−0.223	−0.229	−0.236	−0.255357

Table 3.16: Simulated and theoretical standard deviations and correlations of the MLEs from the Burr XII ALT model, with shape parameters $a = 2$ and $\tau = 4$.

	$n_1^{-1}\times$ Average Simulated EFI Element				Theoretical Value
	$n_1 = 25$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	
$-E\left[\frac{\partial^2 l_B}{\partial a^2}\right]$	1	1	1	1	1
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right]$	0.0002117	0.0000625	−0.0000158	0.0000424	0
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right]$	−8.00052	−8.00207	−8.00049	−8.00035	−8
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right]$	−960.012	−960.103	−960.124	−960.058	−960
$-E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right]$	0.182783	0.182753	0.182759	0.182774	0.18277
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right]$	0.999404	0.998339	0.99973	0.999611	1
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right]$	119.971	119.925	119.912	119.961	120
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right]$	71.9844	72.0099	72.007	71.9973	72
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b \beta_b}\right]$	8637.08	8639.82	8641.23	8639.8	8640
$-E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right]$	1558380	1558500	1559000	1558780	1558800

Table 3.17: Simulated and theoretical elements from the Burr XII ALT EFI matrix, with shape parameters $a = 2$ and $\tau = 6$.

	Simulated Value				Theoretical
	$n_1 = 25$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
% valid MLEs	95.21	99.98	100.00	100.00	
mean$(\widehat{a})$	4.05465	2.18172	2.014147	2.004573	2
mean$(\widehat{\tau})$	6.10230	6.02721	6.002621	6.000535	6
mean$(\widehat{\alpha}_b)$	5.55021	5.51121	5.500894	5.500382	5.5
mean$\left(\widehat{\beta}_b\right)$	−0.009988	−0.010003	−0.0100001	−0.0100001	−0.01
$\sqrt{n_1}\times$s.d.$(\widehat{a})$	74.29613	7.711701	5.253496	5.1538510	5.12932
$\sqrt{n_1}\times$s.d.$(\widehat{\tau})$	4.17629	4.290739	4.143423	4.136447	4.16052
$\sqrt{n_1}\times$s.d.$(\widehat{\alpha}_b)$	0.871887	0.713529	0.656403	0.653009	0.650425
$\sqrt{n_1}\times$s.d.$\left(\widehat{\beta}_b\right)$	0.0014102	0.001365	0.0014022	0.0014021	0.00138409
skewness$(\widehat{a})$	29.021	6.905	0.429	0.233	0
skewness$(\widehat{\tau})$	0.466	0.296	0.084	0.046	0
skewness$(\widehat{\alpha}_b)$	1.652	0.752	0.175	0.108	0
skewness$\left(\widehat{\beta}_b\right)$	0.008	−0.037	−0.035	−0.003	0
corr$(\widehat{a}, \widehat{\tau})$	−0.790	−0.804	−0.800	−0.798	−0.811125
corr$(\widehat{a}, \widehat{\alpha}_b)$	0.944	0.947	0.943	0.944	0.948295
corr$\left(\widehat{a}, \widehat{\beta}_b\right)$	−0.016	0.007	−0.009	−0.021	0
corr$(\widehat{\tau}, \widehat{\alpha}_b)$	−0.775	−0.797	−0.787	−0.784	−0.799576
corr$\left(\widehat{\tau}, \widehat{\beta}_b\right)$	0.008	0.005	0.013	0.008	0
corr$\left(\widehat{\alpha}_b, \widehat{\beta}_b\right)$	−0.250	−0.226	−0.253	−0.264	−0.255357

Table 3.18: Simulated and theoretical standard deviations and correlations of the MLEs from the Burr XII ALT model, with shape parameters $a = 2$ and $\tau = 6$.

	$n_1^{-1}\times$ Average Simulated EFI Element				Theoretical
	$n_1 = 25$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
$-E\left[\frac{\partial^2 l_B}{\partial a^2}\right]$	0.16	0.16	0.16	0.16	0.16
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right]$	−0.361204	−0.361146	−0.361076	−0.361127	−0.361111
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right]$	−1.33286	−1.33333	−1.33347	−1.33337	−1.33333
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right]$	−160.061	−160.02	−160.052	−160.026	−160
$-E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right]$	2.76709	2.76684	2.76654	2.76654	2.76657
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right]$	3.669	3.6672	3.66605	3.66664	3.66667
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right]$	440.088	440.043	439.796	439.936	440
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right]$	11.4266	11.4289	11.429	11.429	11.4286
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b \beta_b}\right]$	1372.17	1371.73	1371.68	1371.61	1371.43
$-E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right]$	247659	247520	247477	247464	247429

Table 3.19: Simulated and theoretical elements from the Burr XII ALT EFI matrix, with shape parameters $a = 5$ and $\tau = 5$.

next example takes the same parameter values as in the previous example, but now $a = 5$ and $\tau = 4$. The results are presented in Tables 3.21 and 3.22. As before, we observe good conformity between simulated and theoretical results, thereby increasing our confidence in the theoretical results established above. Our final example takes the same parameter values as in the previous example, but with $a = 5$ and $\tau = 6$. The results are presented in Tables 3.23 and 3.24. In this final example, as with all previous examples, we observe good agreement between simulated and theoretical results, which improves as the sample size increases.

3.4 Discussion on Simulations

For both experiments above, and in every combination of shape parameter values considered, we observed larger means and standard deviations of $\widehat{a}$ for smaller sample sizes. These became more consistent with their theoretical counterparts as the sample size increased. However, for the other three parameter estimates, $\widehat{\tau}, \widehat{\alpha}_b$ and $\widehat{\beta}_b$, the means and standard deviations were quite close to their theoretical counterparts, even for the smallest sample size considered. It is also worth noting that as the sample size decreased, not only did the mean of $\widehat{a}$ rise, but so did the mean of $\widehat{\alpha}_b$. We have observed something similar to this before. It is now appropriate to recall the electrode data (example 2.1.1.3) from Chapter two, where the algorithm failed to converge on the maximum likelihood estimates of a and α_b, rather the values rose without limit. This asymptotic behaviour of $\widehat{a}$ and $\widehat{\alpha}_b$ shall be examined in more detail in the next chapter. Also worthy of note at this stage, is the

	Simulated Value				Theoretical
	$n_1 = 25$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
% valid MLEs	80.93	97.82	100.00	100.00	
mean$(\widehat{a})$	8.89231	7.99056	5.153282	5.048465	5
mean$(\widehat{\tau})$	2.10739	2.01124	2.000153	2.000187	2
mean$(\widehat{\alpha}_b)$	5.50616	5.59573	5.510123	5.503365	5.5
mean$\left(\widehat{\beta}_b\right)$	−0.010009	−0.010001	−0.0100005	−0.010001	−0.01
$\sqrt{n_1}\times$s.d.$(\widehat{a})$	78.91692	118.4836	28.22751	25.945914	24.6219
$\sqrt{n_1}\times$s.d.$(\widehat{\tau})$	1.29510	1.28485	1.308019	1.318133	1.30169
$\sqrt{n_1}\times$s.d.$(\widehat{\alpha}_b)$	2.79073	4.04626	3.412059	3.345328	3.25422
$\sqrt{n_1}\times$s.d.$\left(\widehat{\beta}_b\right)$	0.003532	0.0035073	0.0034713	0.0034957	0.0034740
skewness$(\widehat{a})$	6.320	8.056	1.203	0.567	0
skewness$(\widehat{\tau})$	0.798	0.388	0.108	0.082	0
skewness$(\widehat{\alpha}_b)$	0.941	1.357	0.453	0.238	0
skewness$\left(\widehat{\beta}_b\right)$	−0.032	−0.057	0.012	−0.018	0
corr$(\widehat{a}, \widehat{\tau})$	−0.711	−0.774	−0.782	−0.780	−0.793006
corr$(\widehat{a}, \widehat{\alpha}_b)$	0.971	0.983	0.983	0.983	0.984487
corr$\left(\widehat{a}, \widehat{\beta}_b\right)$	0.013	0.010	0.002	−0.010	0
corr$(\widehat{\tau}, \widehat{\alpha}_b)$	−0.748	−0.809	−0.820	−0.817	−0.828333
corr$\left(\widehat{\tau}, \widehat{\beta}_b\right)$	−0.021	−0.005	0.002	0.011	0
corr$\left(\widehat{\alpha}_b, \widehat{\beta}_b\right)$	−0.147	−0.110	−0.117	−0.132	−0.128106

Table 3.20: Simulated and theoretical standard deviations and correlations of the MLEs from the Burr XII ALT model, with shape parameters $a = 5$ and $\tau = 2$.

	$n_1^{-1}\times$ Average Simulated EFI Element				Theoretical
	$n_1 = 25$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
$-E\left[\frac{\partial^2 l_B}{\partial a^2}\right]$	0.16	0.16	0.16	0.16	0.16
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right]$	−0.180554	−0.180583	−0.180545	−0.180555	−0.180556
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right]$	−2.66797	−2.66812	−2.6662	−2.66644	−2.66667
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right]$	−320.01	−320.275	−320.009	−319.958	−320
$-E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right]$	0.691542	0.691481	0.691642	0.691667	0.691643
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right]$	3.66478	3.66463	3.66713	3.66706	3.66667
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right]$	439.831	439.515	439.94	440.054	440
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right]$	45.7316	45.7364	45.7076	45.7113	45.7143
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b \beta_b}\right]$	5485	5489.24	5485.66	5485.09	5485.71
$-E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right]$	989383	990455	989748	989587	989714

Table 3.21: Simulated and theoretical elements from the Burr XII ALT EFI matrix, with shape parameters $a = 5$ and $\tau = 4$.

	Simulated Value				Theoretical Value
	$n_1 = 25$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	
% valid MLEs	77.79	97.08	100.00	100.00	
mean$(\widehat{a})$	10.73362	8.47798	5.158571	5.051951	5
mean$(\widehat{\tau})$	4.22528	4.01973	4.000266	4.000248	4
mean$(\widehat{\alpha}_b)$	5.50909	5.55052	5.505426	5.501820	5.5
mean$\left(\widehat{\beta}_b\right)$	−0.010004	−0.009996	−0.010000	−0.009999	−0.01
$\sqrt{n_1}\times$s.d.$(\widehat{a})$	138.10812	168.1969	27.55780	25.63805	24.6219
$\sqrt{n_1}\times$s.d.$(\widehat{\tau})$	2.64845	2.60758	2.581774	2.612391	2.60337
$\sqrt{n_1}\times$s.d.$(\widehat{\alpha}_b)$	1.49861	2.09453	1.673018	1.661179	1.62711
$\sqrt{n_1}\times$s.d.$\left(\widehat{\beta}_b\right)$	0.001785	0.0017520	0.0017287	0.001735	0.0017370
skewness$(\widehat{a})$	9.208	13.316	1.091	0.481	0
skewness$(\widehat{\tau})$	0.794	0.357	0.093	0.018	0
skewness$(\widehat{\alpha}_b)$	1.097	1.490	0.460	0.189	0
skewness$\left(\widehat{\beta}_b\right)$	0.020	0.039	0.014	0.011	0
corr$(\widehat{a}, \widehat{\tau})$	−0.726	−0.772	−0.773	−0.783	−0.793006
corr$(\widehat{a}, \widehat{\alpha}_b)$	0.972	0.983	0.983	0.984	0.984487
corr$\left(\widehat{a}, \widehat{\beta}_b\right)$	−0.009	−0.001	−0.002	−0.024	0
corr$(\widehat{\tau}, \widehat{\alpha}_b)$	−0.761	−0.807	−0.809	−0.819	−0.828333
corr$\left(\widehat{\tau}, \widehat{\beta}_b\right)$	0.017	0.002	−0.003	0.019	0
corr$\left(\widehat{\alpha}_b, \widehat{\beta}_b\right)$	−0.167	−0.118	−0.123	−0.143	−0.128106

Table 3.22: Simulated and theoretical standard deviations and correlations of the MLEs from the Burr XII ALT model, with shape parameters $a = 5$ and $\tau = 4$.

	$n_1^{-1}\times$ Simulated EFI Element				Theoretical Value
	$n_1 = 25$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	
$-E\left[\frac{\partial^2 l_B}{\partial a^2}\right]$	0.16	0.16	0.16	0.16	0.16
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right]$	−0.120358	−0.120333	−0.12036	−0.120367	−0.12037
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right]$	−4.00196	−4.00113	−4.00049	−3.99992	−4
$-E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right]$	−479.714	−480.004	−480.072	−479.984	−480
$-E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right]$	0.307369	0.307342	0.307386	0.307401	0.307397
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right]$	3.66542	3.66437	3.66587	3.66667	3.66667
$-E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right]$	440.549	439.929	439.855	440.009	440
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right]$	102.895	102.868	102.863	102.855	102.857
$-E\left[\frac{\partial^2 l_B}{\partial \alpha_b \beta_b}\right]$	12339.4	12342.2	12343.6	12342.5	12342.9
$-E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right]$	2225420	2226530	2226870	2226760	2226860

Table 3.23: Simulated and theoretical elements from the Burr XII ALT EFI matrix, with shape parameters $a = 5$ and $\tau = 6$.

	Simulated Value				Theoretical
	$n_1 = 25$	$n_1 = 100$	$n_1 = 1000$	$n_1 = 3000$	Value
% valid MLEs	77.09	97.00	100.00	100.00	
mean$(\widehat{a})$	11.25251	9.13891	5.15754	5.041633	5
mean$(\widehat{\tau})$	6.32542	6.03081	6.000474	6.000920	6
mean$(\widehat{\alpha}_b)$	5.50751	5.53496	5.503609	5.500804	5.5
mean$\left(\widehat{\beta}_b\right)$	−0.010004	−0.009999	−0.0099998	−0.0099998	−0.01
$\sqrt{n_1}\times$s.d.$(\widehat{a})$	162.40586	234.3730	27.49136	25.68123	24.6219
$\sqrt{n_1}\times$s.d.$(\widehat{\tau})$	3.94142	3.92945	3.92094	3.907596	3.90506
$\sqrt{n_1}\times$s.d.$(\widehat{\alpha}_b)$	1.00712	1.43324	1.12062	1.103259	1.80474
$\sqrt{n_1}\times$s.d.$\left(\widehat{\beta}_b\right)$	0.001176	0.001168	0.0011632	0.0011558	0.0011580
skewness$(\widehat{a})$	11.210	13.811	0.990	0.587	0
skewness$(\widehat{\tau})$	0.761	0.450	0.121	0.073	0
skewness$(\widehat{\alpha}_b)$	1.160	1.627	0.373	0.261	0
skewness$\left(\widehat{\beta}_b\right)$	−0.013	0.007	−0.013	0.037	0
corr$(\widehat{a}, \widehat{\tau})$	−0.725	−0.773	−0.785	−0.783	−0.793006
corr$(\widehat{a}, \widehat{\alpha}_b)$	0.973	0.984	0.984	0.983	0.984487
corr$\left(\widehat{a}, \widehat{\beta}_b\right)$	−0.005	0.014	0.007	−0.006	0
corr$(\widehat{\tau}, \widehat{\alpha}_b)$	−0.762	−0.810	−0.820	−0.820	−0.828333
corr$\left(\widehat{\tau}, \widehat{\beta}_b\right)$	0.008	−0.003	−0.011	0.008	0
corr$\left(\widehat{\alpha}_b, \widehat{\beta}_b\right)$	−0.160	−0.102	−0.110	−0.126	−0.128106

Table 3.24: Simulated and theoretical standard deviations and correlations of the MLEs from the Burr XII ALT model, with shape parameters $a = 5$ and $\tau = 6$.

relationship between the sample size and the percentage of valid MLEs. We see that as one decreases, so does the other. Furthermore, when the algorithm failed to converge, the estimates for a and α_b rose without limit; in terms of maximised log-likelihood, the Weibull distribution is preferred to the Burr XII. We shall consider these results in greater detail in the next chapter. However, we may note this result is perhaps to be expected: when the sample size is small, it seems sensible to expect the Weibull distribution (with only one shape parameter) sometimes to be a better fit than the Burr XII distribution, with two shape parameters. Further, given the limiting relationship between the Weibull and Burr XII in mind, for some sets of data we can also expect values of $\widehat{a}$ to be rather larger in smaller samples than with larger sample sizes; these large values of $\hat{a}$ help to explain the systematic bias seen in our simulations, particularly with small sample sizes.

With all simulations above, we observed a heavily right-skewed distribution for $\widehat{a}$, for small sample sizes; the distribution becoming more Normal with increasing sample size. The skewness in the distribution of the other parameter estimates was less marked, even for small sample sizes.

Similar theoretical correlations were observed for both simulations; for example, the two shape parameters had strong inverse correlation, as did the two scale parameters; while $\widehat{\alpha}_b$ was positively correlated with $\widehat{a}$ and negatively correlated with $\widehat{\tau}$.

Clearly, the two experiments considered here - which in total constitute twelve separate examples - are just a small fraction of the possible combinations of stresses and parameter values that could be studied. For example, we could go further and examine the effect of five or more stress levels, unequal sample sizes across stress levels and a whole host of different values for the shape parameters. However, we regard the array of results presented here as sufficient to endorse the theoretical results obtained above.

In this chapter we have written down the EFI matrix for the Burr XII non-ALT model. We then extended the Burr XII model into an accelerated environment and presented theoretical results. To validate these results we then presented a series of simulations for varying parameter values and sample sizes. Due to the improved agreement between simulated and theoretical results as the sample size increased, we concluded that the theory was sound.

In the next chapter, we investigate further the asymptotic behaviour occasionally exhibited by $\widehat{a}$ and $\widehat{\alpha}_b$. We shall examine, in more detail, the link between this behaviour and our preference for the Weibull distribution over the Burr XII. Consequently, we derive a result that allows us to determine if the Burr XII will provide a superior fit than the Weibull, to a given data set, simply through fitting the Weibull distribution.

Chapter 4

Weibull or Burr XII?

In Chapter two we presented three examples: two where the Burr XII models (either ALT or non-ALT) provided a better fit, in terms of maximized log-likelihood, than their Weibull counterparts, (example 2.1.1.1: Carbon Fibre Rod Data and example 2.1.1.2: Aluminium Coupon Data), and one where the Weibull models produced a better fit than the analogous Burr XII models, (example 2.1.1.3: Electrodes in Oil Data). In the final case, we saw that when attempting to fit a Burr XII model in the non-ALT (ALT) frameworks, the approximations to $\widehat{a}$ and $\widehat{\theta}$ ($\widehat{a}$ and $\widehat{\alpha}_b$) gradually rose without limit, while the log-likelihood approached but never quite attained the Weibull equivalent.

In this chapter, we consider the circumstances under which a given ALT data set will exhibit the behaviour observed in example 2.1.1.3., as opposed to that observed in examples 2.1.1.1. and 2.1.1.2. We shall show that we can determine which model will provide a better fit to a given data set through one simple expression, acting as a discriminating factor. This extends the work of Watkins (2001), which considers the non-ALT model.

We first derive the discriminating factor for complete data from first principles and then apply it to the examples we studied in Chapter two. Our attention then turns to the theoretical properties of the discriminating factor; in particular we are interested in the probability of specifying the correct model, when the data comes from a Weibull distribution or alternatively from a Burr XII distribution. We are also interested in the probability of misspecifying the model; that is to say the probability of preferring to fit a Weibull distribution when the data is from a Burr XII distribution and, conversely, preferring to fit a Burr XII distribution when the data is from a Weibull distribution. This investigation is further motivated by observations made regarding simulations at the end of the last chapter. These suggest that there are instances, seemingly more common for smaller sample sizes, when we prefer to fit the Weibull distribution, even though the data has been generated from a Burr XII distribution.

4.1 Derivation Of The Discriminating Factor: Non-ALT Case

In this section, we review the steps presented in Watkins (2001) to determine the discriminating factor for non-ALT models. The results come from firstly recalling, in (1.8), that as $\lambda\ (=\theta^{\tau}) \to \infty$, the shape parameter of the limiting Weibull model was τ with scale parameter,

$$\frac{\theta}{a^{1/\tau}}$$

which tended to ϕ in the same limit. Consequently, we also have $a \to \infty$ in this limit; thus, two of the three parameters in the Burr XII CDF may be unbounded as l_b approaches its maximum. In addition to this, the limiting behaviour of l_b - as illustrated in Watkins (2001) - means that all derivatives will tend to zero in this limit, making it difficult to distinguish between the two types of behaviour on the basis of these derivatives. To overcome this, Watkins considered a reparameterisation of the Burr XII distribution, using

$$\theta = -\ln\omega$$

and

$$\psi = \frac{\theta}{a^{1/\tau}},$$

so that

$$a = \left(\frac{\theta}{\psi}\right)^{\tau} = \left(\frac{-\ln\omega}{\psi}\right)^{\tau}.$$

Consequently, as $a, \theta \to \infty$, we have $\omega \to 0$ and $\psi \to \phi$, and interest then centres on the behaviour of the associated score function at $\tau = \widehat{B}, \psi = \widehat{\phi}$ and as $\omega \to 0$. For these first derivatives of the reparameterised likelihood, now a function of the parameters τ, ω and ψ, it was shown that

$$\lim_{\omega\to 0} \left.\frac{\partial l_b}{\partial \psi}\right|_{\tau=\widehat{B},\psi=\widehat{\phi}} = 0$$

and

$$\lim_{\omega\to 0} \left.\frac{\partial l_b}{\partial \tau}\right|_{\tau=\widehat{B},\psi=\widehat{\phi}} = 0.$$

To simplify the consideration of $\frac{\partial l_b}{\partial \omega}$, it is convenient to write $\lambda = (-\ln\omega)^{\tau}$; the behaviour of $\frac{\partial l_b}{\partial \omega}$ is then determined by

$$\psi^{-\tau}\lambda \sum_{i=1}^{n} \ln\left\{1 + \left(\frac{d_i}{\lambda^{1/\tau}}\right)^{\tau}\right\} - \left(\psi^{-\tau}\lambda + 1\right) \sum_{i=1}^{n} \frac{\left(\frac{d_i}{\lambda^{1/\tau}}\right)^{\tau}}{1 + \left(\frac{d_i}{\lambda^{1/\tau}}\right)^{\tau}}.$$

Applying a Taylor series expansion to these summations, together with the result

$$\phi^{B}\big|_{B=\widehat{B},\phi=\widehat{\phi}} = \frac{1}{n}\sum_{i=1}^{n} d_i^{\widehat{B}},$$

Length/mm	1	10	20	50
n	57	64	70	66
Discriminating Factor	4.2429678	17.06521	6.1955223	3.57413

Table 4.1: Discriminating factor for subsets of carbon fibre rod data

at (2.2), it can then be shown that the sign of

$$\lim_{\omega \to 0} \frac{\partial l_b}{\partial \omega}\bigg|_{\tau=\widehat{B},\psi=\widehat{\phi}}$$

is determined by

$$\frac{S_0\left(2\widehat{B}\right)}{2} - \frac{\left\{S_0\left(\widehat{B}\right)\right\}^2}{n} \tag{4.1}$$

where

$$S_j(t) = \sum_{i=1}^{n} d_i^t (\ln d_i)^j$$

and, as before, $\widehat{B}$ is the maximum likelihood estimate of the shape parameter in the Weibull distribution. The expression (4.1) thus defines the discriminating factor, based on the criterion of greater maximised likelihood; a value greater than zero indicates a preference for fitting the Burr XII model over the Weibull, while a value less than zero indicates the Weibull model is preferred to the Burr XII. Note that any rescaling of the data set leaves the sign of (4.1) unchanged, but may introduce some numerical stability into the calculation of $S_0\left(2\widehat{B}\right)$ and $S_0\left(\widehat{B}\right)$. Typically, one rescales the data set using $d_i \to \frac{d_i}{\widehat{\phi}}$.

4.1.1 Examples Revisited

Example 2.1.1.1 [Carbon Fibre Rod Data] revisited

We return to the complete data set consisting of failure stresses of single carbon fibre rods at four different lengths; see Table 2.1. We can now determine the value of (4.1), suitably rescaled, for each subset of data, and these are presented in Table 4.1. It appears that we favour the three parameter Burr XII model over the Weibull model for each of the four data subsets. This is consistent with results seen in Table 2.15 where we observed the maximised log-likelihoods for the Burr XII model exceeded those for the Weibull model.

Example 2.1.1.2 [Aluminium Coupon Data] revisited

Recall, this data set consisted of the fatigue life to failure of coupons cut from aluminium sheeting, see Table 2.4. The values of (4.1) for each subset of data are presented in Table 4.2. Again, this is consistent with results seen in Table 2.17, where we saw the Burr XII distribution providing a superior fit to the Weibull at all stress levels.

psi/cycle	2.1	2.6	3.1
n	101	102	101
Discriminating Factor	1.159937	11.16959	40.01791

Table 4.2: Discriminating factor for subsets of aluminium coupon data

volts/second	10	100	1000
n	60	60	60
Discriminating Factor	−7.804207	−3.739976	−7.660377

Table 4.3: Discriminating factor for subsets of electrode data for 9 sq in electrodes.

Example 2.1.1.3 [Electrode Data] revisited

We recall that this data set concerns disc electrodes immersed in an insulating oil; see Table 2.6. The observed values of (4.1) are presented in Table 4.3. It seems the Weibull model is favoured over the Burr XII model. Again, this is consistent with results in Table 2.19 where the maximised log-likelihoods were greater for the Weibull model.

4.2 Extension to ALT Models

As in the non-ALT case discussed by Watkins (2001), we shall now consider a reparameterisation of the Burr XII ALT model Our approach will be analogous to Watkins' paper based on the behaviour seen in Chapter two, where, as a and α_b tended to infinity, τ and β_b remained bounded - in particular, τ appeared to tend to the Weibull ALT shape parameter, with β_b close to zero for the rescaled data. We begin by examining the behaviour of the log-likelihood, key partial derivatives and estimates of model parameters when we steadily increase the value of a. This we do for the carbon fibre rod data and the electrode data; results for the aluminium coupon data are omitted for the sake of brevity. The graphs then assist us in determining a suitable reparameterisation for the Burr XII ALT model, which we then pursue through to the ultimate goal of determining a discriminating factor for an ALT data set.

4.2.1 Examples Revisited

Example 2.1.1.1 [Carbon Fibre Rod Data] revisited

In Figures 4.1 to 4.6, we choose values of a from 0.9 to 5000000, at gradually increasing increments. For instance, for a in the range 0.9 to 50, we take increments of 1, rising to 50 until a reaches 1000, while for values of a over 100000, the incrementation is in steps of 100000. For any particular value of a, we determine the maximum likelihood estimates of the remaining three parameters in the model - τ, α_b and β_b - together with the maximised log-likelihood. We then plot each MLE and maximised log-likelihood, independently, against the corresponding value of a. However, for clarity, we use the natural logarithm of a on the

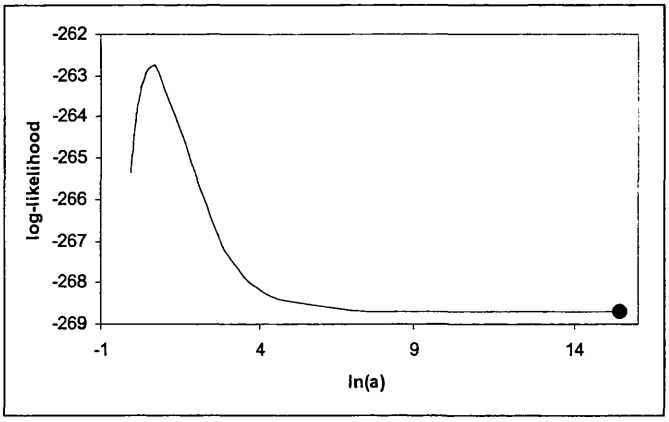

Figure 4.1: Maximised log-likelihood for Burr XII ALT (continuous line) and Weibull ALT (single point) for the carbon fibre rod data.

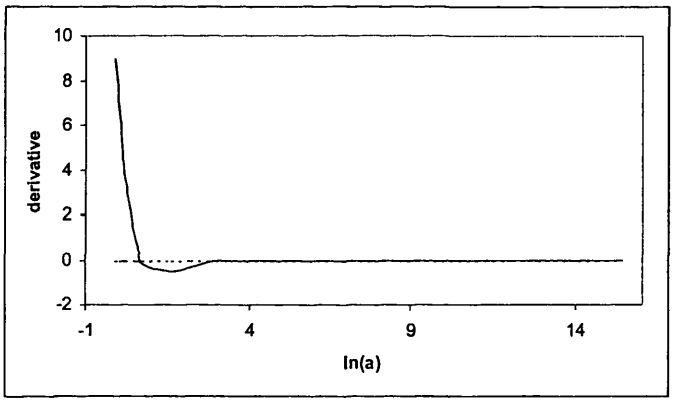

Figure 4.2: The derivative $\frac{\partial l_B}{\partial a}$ of the log-likelihood l_B maximised with respect to τ, α_b and β_b for a given a, against a, for the carbon fibre rod data.

horizontal axis. Figures 4.1 and 4.2 show that the Burr XII log-likelihood is maximised at finite a. In addition, as $a \to \infty$ we see $\widehat{\tau} \to \widehat{B}$ (in Figure 4.3), $\widehat{\beta}_b \to \widehat{\beta}_w$ (in Figure 4.4) and $\widehat{\alpha}_b \to \infty$ (in Figure 4.5). We also observe in Figure 4.6, $\frac{\exp(\widehat{\alpha}_b)}{a^{1/\widehat{\tau}}} \to \exp(\widehat{\alpha}_w)$, which is analogous to the limiting behaviour in the non-ALT case, since $\exp(\widehat{\alpha}_b)$ is the estimated base scale ignoring the effect of stress.

Example 2.1.1.3 [Electrode Data] revisited

In Figures 4.7 to 4.12, a was taken between 15 and 5000000. A similar incrementation was used here as in the previous example. For instance, for values of a up to 1000 we used incrementations of 20, while for values of a greater than 100000, we used steps of 100000. The figures are as before: we choose a value for a and then determine the maximum likelihood estimates of the other three model parameters, together with the maximised log-likelihood.

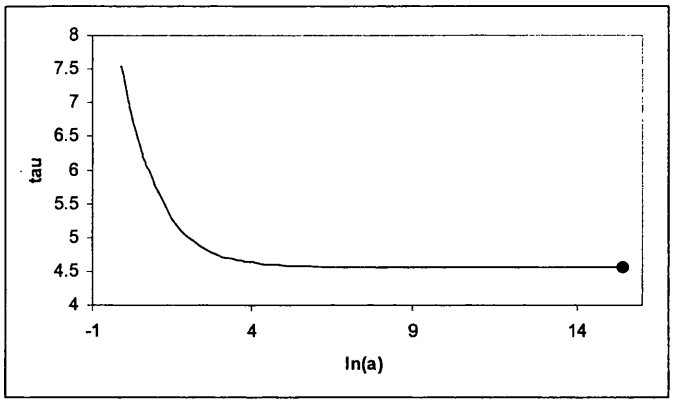

Figure 4.3: $\widehat{\tau}$ (continuous line) against a, and $\widehat{B}$ (single point), for the carbon fibre rod data.

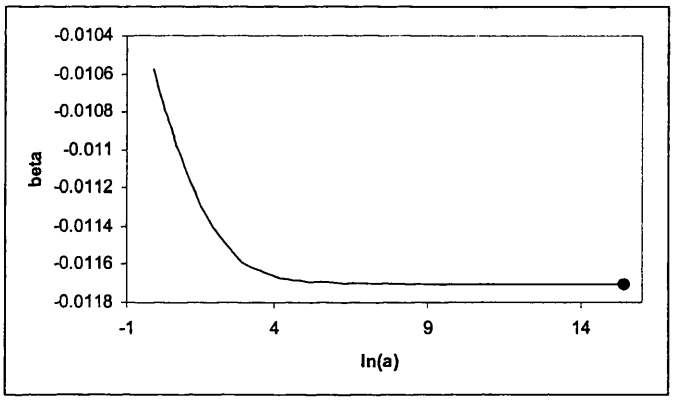

Figure 4.4: $\widehat{\beta}_b$ (continuous line) against a, and $\widehat{\beta}_w$ (single point), for the carbon fibre rod data.

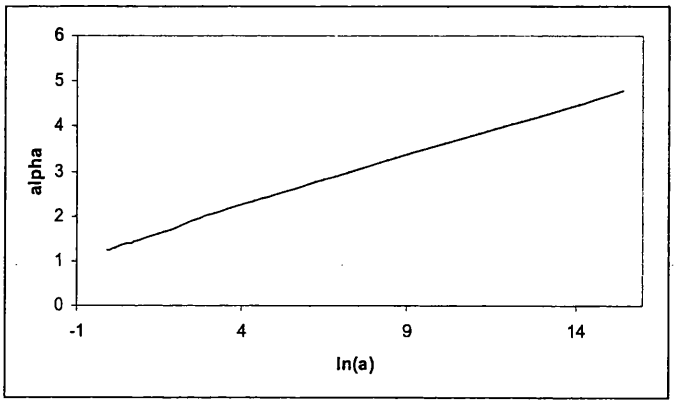

Figure 4.5: $\widehat{\alpha}_b$ against a for the carbon fibre rod data.

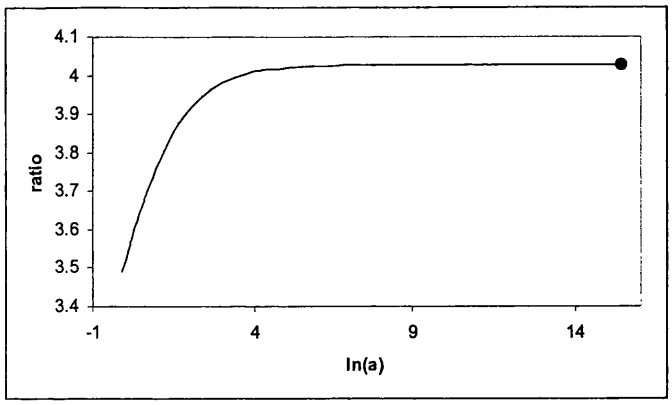

Figure 4.6: $\frac{\exp(\widehat{\alpha}_b)}{a^{1/\widehat{\tau}}}$ (continuous line) against a, and $\exp(\widehat{\alpha}_w)$ (single point), for the carbon fibre rod data.

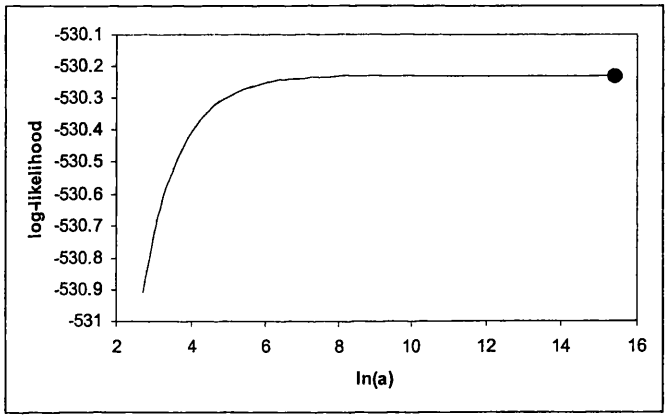

Figure 4.7: Maximised log-likelihood for Burr XII ALT (continuous line) and Weibull ALT (single point) for the electrode data.

Again, a natural logarithm scale was chosen for clarity. Figures 4.7 and 4.8 clearly show that the Burr XII log-likelihood is maximised at infinite a, and indeed tends to the Weibull log-likelihood with increasing values of a. As with example 2.1.1.1, when $a \to \infty$ we see $\widehat{\tau} \to \widehat{B}$, $\widehat{\beta}_b \to \widehat{\beta}_w$ and $\widehat{\alpha}_b \to \infty$; we also observe $\frac{\exp(\widehat{\alpha}_b)}{a^{1/\widehat{\tau}}} \to \exp(\widehat{\alpha}_w)$; see Figures 4.9 to 4.12.

4.2.2 Reparameterisations of the Burr XII ALT Distribution

For the non-accelerated case, Watkins (2001) observed two of the three model parameters tending to infinity - namely ϕ and a - while τ remained finite. He initially chose to reparameterise the scale parameter ϕ and, from this, to derive a suitable reparameterisation for a. An equally viable approach would be to begin by reparameterising the shape parameter a and from this, establish a suitable reparameterisation for the scale ϕ.

However, this choice should have no bearing upon the final result. Similarly, the manner

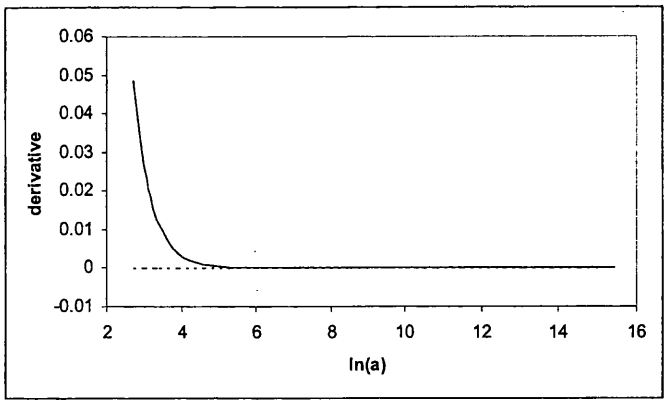

Figure 4.8: The derivative $\frac{\partial l_B}{\partial a}$ of the log-likelihood l_B maximised with respect to τ, α_b and β_b for a given a, against a, for the electrode data.

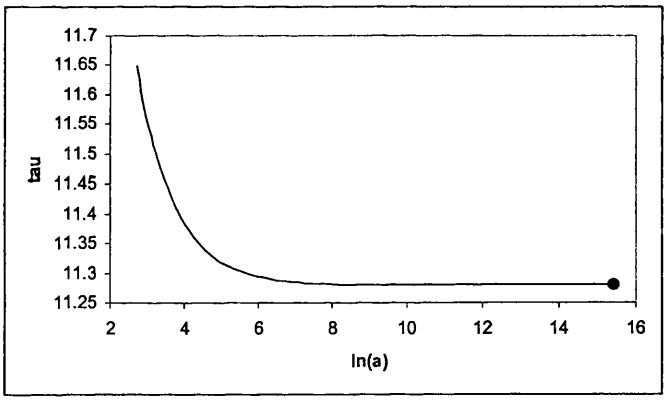

Figure 4.9: $\widehat{\tau}$ (continuous line) against a, and $\widehat{B}$ (single point), for the electrode data.

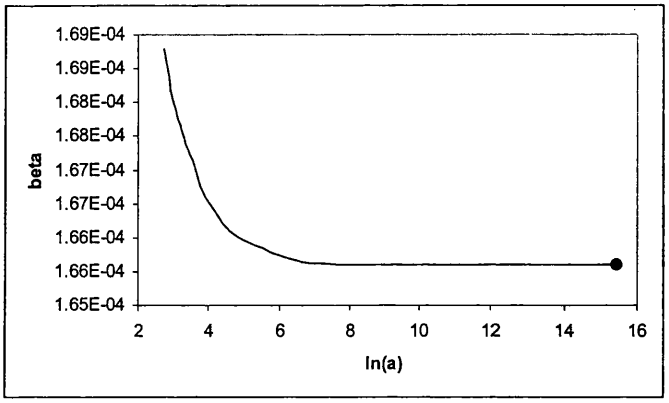

Figure 4.10: $\widehat{\beta}_b$ (continuous line) against a, and $\widehat{\beta}_w$ (single point), for the electrode data.

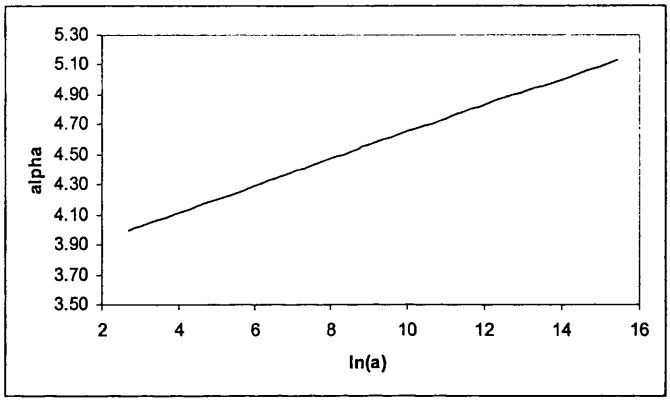

Figure 4.11: $\widehat{\alpha}_b$ against a for the electrode data.

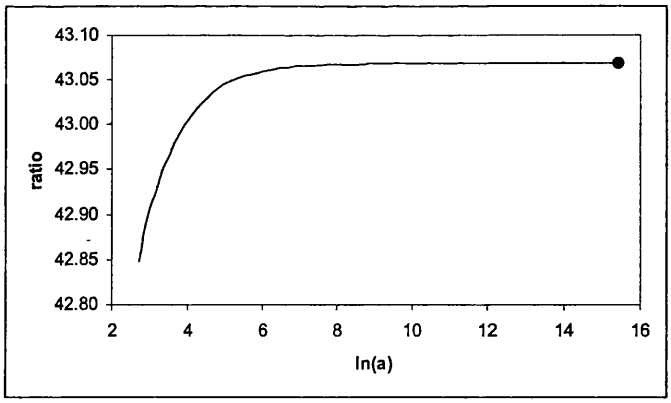

Figure 4.12: $\frac{\exp(\widehat{\alpha}_b)}{a^{1/\widehat{\tau}}}$ (continuous line) against a, and $\exp(\widehat{\alpha}_w)$ (single point), for the electrode data.

in which we reparameterise in the ALT case should have no impact upon the outcome. To emphasise this point, we shall work through two reparameterisations: one which considers an initial reparamerisation of the shape parameter a and another, analogous to Watkins' approach, which considers a reparameterisation of α_b.

Reparameterising the Shape Parameter a

If we define

$$a = \ln(-\ln\omega), \tag{4.2}$$

then, as $a \to \infty$, $\omega \to 0$. Now, we observed that $\frac{\exp(\alpha_b)}{a^{1/\tau}}$ tended to a finite limit - seemingly $\exp(\widehat{\alpha}_w)$ - so we shall write

$$\frac{\exp(\alpha_b)}{a^{1/\tau}} = \psi. \tag{4.3}$$

We use a double logarithmic expression in (4.2) to compensate for the inclusion of the exponential function in (4.3). Substituting (4.2) into (4.3) and rewriting gives us

$$\alpha_b = \ln(\psi) + \tau^{-1}\ln[\ln\{-\ln(\omega)\}].$$

Under this reparameterisation, l_B at (2.10) now becomes,

$$\begin{aligned} l_B(\omega,\psi,\tau,\beta_b) &= S_n\ln\tau + S_n\ln(\ln(-\ln\omega)) + (\tau-1)S_e \\ &\quad -\tau S_n\left[\ln\psi + \tau^{-1}\ln(\ln(-\ln\omega))\right] - \tau\beta_b S_x - (1+\ln(-\ln\omega))\,t_*^d(\psi,\omega), \end{aligned} \tag{4.4}$$

where we now define

$$t_*^d(\psi,\omega) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln\left\{1+\left(\frac{\frac{d_{ij}}{\exp(\beta_b x_i)}}{\psi[\ln(-\ln\omega)]^{1/\tau}}\right)^{\tau}\right\}. \tag{4.5}$$

We are interested in the behaviour of the score function for (4.4) at $\tau = \widehat{B}, \psi = \exp(\widehat{\alpha}_w), \beta_b = \widehat{\beta}_w$ and as $\omega \to 0$. We make the following definitions, analogous to those in Watkins (2001),

$$s_{v,w}^d(t,\beta_b) = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_i^w\left(\frac{d_{ij}}{\exp(\beta_b x_i)}\right)^t\left(\ln\left\{\frac{d_{ij}}{\exp(\beta_b x_i)}\right\}\right)^v, \tag{4.6}$$

and

$$t_{v,w}^d(\psi,\omega) = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_i^w\left[\frac{\left(\frac{\frac{d_{ij}}{\exp(\beta_b x_i)}}{\psi[\ln(-\ln\omega)]^{1/\tau}}\right)^{\tau}\left(\ln\left\{\frac{\frac{d_{ij}}{\exp(\beta_b x_i)}}{\psi[\ln(-\ln\omega)]^{1/\tau}}\right\}\right)^v}{1+\left(\frac{\frac{d_{ij}}{\exp(\beta_b x_i)}}{\psi[\ln(-\ln\omega)]^{1/\tau}}\right)^{\tau}}\right]. \tag{4.7}$$

Before proceeding to examine first derivatives of l_B it is convenient to consider the Weibull ALT log-likelihood in terms of this new notation. Firstly, we note that we can rewrite (2.4)

as

$$\{\exp(\widehat{\alpha}_w)\}^{-\widehat{B}} S_{0,0} = S_n, \tag{4.8}$$

while the log-likelihood (2.3) becomes

$$l_W = S_n \ln B + (B-1)\, s^d_{1,0}(0,0) - \alpha_w B S_n - \beta_w B S_x - \exp(-\alpha_w B)\, s^d_{0,0}(B, \beta_w),$$

which gives

$$l^*_W = S_n \ln B + (B-1)\, s^d_{1,0}(0,0) - S_n \ln\left\{\frac{s^d_{0,0}(B,\beta_w)}{S_n}\right\} - \beta_w B S_x - S_n$$

as the profile log-likelihood. Differentiating l^*_W with respect to B and equating to zero yields

$$\frac{\partial l^*_W}{\partial B} = S_n \widehat{B}^{-1} + s^d_{1,0}(0,0) - \widehat{\beta}_w S_x - S_n \left(\frac{s^d_{1,0}\left(\widehat{B},\widehat{\beta}_w\right)}{s^d_{0,0}\left(\widehat{B},\widehat{\beta}_w\right)}\right) = 0,$$

from which

$$S_n \left(\frac{s^d_{1,0}\left(\widehat{B},\widehat{\beta}_w\right)}{s^d_{0,0}\left(\widehat{B},\widehat{\beta}_w\right)}\right) = S_n \widehat{B}^{-1} + s^d_{1,0}(0,0) - \widehat{\beta}_w S_x. \tag{4.9}$$

We shall see the importance of (4.8) and (4.9) below, where we examine the first derivatives of l_B.

First Derivatives Taking first derivatives of (4.4) with respect to all parameters gives

$$\begin{aligned}
\frac{\partial l_B}{\partial \psi} &= -\tau S_n \psi^{-1} + (1 + \ln(-\ln\omega))\, \tau \psi^{-1} t^d_{0,0}(\psi,\omega), \\
\frac{\partial l_B}{\partial \tau} &= S_n \tau^{-1} + S_e - S_n \ln\psi - \beta_b S_x - (1+\ln(-\ln\omega)) \left\{ \begin{array}{c} \tau^{-1} \ln(\ln(-\ln\omega)) \times \\ t^d_{0,0}(\psi,\omega) + t^d_{1,0}(\psi,\omega) \end{array} \right\}, \\
\frac{\partial l_B}{\partial \beta_b} &= -\tau S_x + (1+\ln(-\ln\omega))\, \tau t^d_{0,1}(\psi,\omega)
\end{aligned}$$

and

$$\frac{\partial l_B}{\partial \omega} = -\frac{1}{\omega \ln\omega} t^d_*(\psi,\omega) + \frac{\{1+\ln(-\ln\omega)\}\, t^d_{0,0}(\psi,\omega)}{\omega \ln\omega \ln(-\ln\omega)}.$$

We now make the substitution

$$\lambda = \psi^\tau [\ln(-\ln\omega)],$$

so that, with ψ and τ finite, $\omega \to 0$ implies $\lambda \to \infty$. The first derivatives then become

$$\begin{aligned}
\frac{\partial l_B}{\partial \psi} &= -\tau S_n \psi^{-1} + \left(1 + \lambda \psi^{-\tau}\right) \tau \psi^{-1} t^d_{0,0}(\lambda), \\
\frac{\partial l_B}{\partial \beta_b} &= -\tau S_x + \left(1 + \lambda \psi^{-\tau}\right) \tau t^d_{0,1}(\lambda),
\end{aligned}$$

$$\frac{\partial l_B}{\partial \tau} = S_n \tau^{-1} + S_e - S_n \ln \psi - \beta_b S_x - \left(1 + \lambda \psi^{-\tau}\right) \left\{ \begin{array}{c} \tau^{-1} \ln\left(\lambda \psi^{-\tau}\right) t^d_{0,0}(\lambda) \\ + t^d_{1,0}(\lambda) \end{array} \right\}$$

and

$$\frac{\partial l_B}{\partial \omega} = \frac{1}{\omega\left(-\log \omega\right)} \left\{ t^d_*(\lambda) - \frac{1 + \lambda \psi^{-\tau}}{\lambda \psi^{-\tau}} t^d_{0,0}(\lambda) \right\}; \tag{4.10}$$

where we now write the $t^d_{v,w}$ and t^d_* with a λ notation, to indicate the role of the parameter in the summations. We now examine the behaviour of the first derivatives. Since our summations are in essence the same as those in Watkins (2001), we use some of the lemmas there, which we now present without proof.

Lemma 1. As $\lambda \to \infty$, we have

$$\lambda t^d_*(\lambda) = s^d_{0,0}(\tau, \beta_b) - \frac{1}{2\lambda} s^d_{0,0}(2\tau, \beta_b) + O\left(\lambda^{-2}\right).$$

Remark 1. From Lemma 1, we have, as $\lambda \to \infty$

$$\lambda t^d_*(\lambda) = s^d_{0,0}(\tau, \beta_b) + O\left(\lambda^{-1}\right).$$

Lemma 2. As $\lambda \to \infty$, we have

$$\lambda t^d_{1,0}(\lambda) = s^d_{1,0}(\tau, \beta_b) - \tau^{-1} (\ln \lambda) s^d_{0,0}(\tau, \beta_b) + O\left(\lambda^{-1}\right).$$

Remark 2. Adapting the argument in Lemma 2, we have, as $\lambda \to \infty$,

$$t^d_{1,0}(\lambda) = O\left(\lambda^{-1}\right).$$

Lemma 3. We have, as $\lambda \to \infty$,

$$\lambda t^d_{0,0}(\lambda) = s^d_{0,0}(\tau, \beta_b) - \lambda^{-1} s^d_{0,0}(2\tau, \beta_b) + O\left(\lambda^{-2}\right).$$

Behaviour of $\frac{\partial l_B}{\partial \psi}$ We write

$$\frac{\partial l_B}{\partial \psi} = \tau \psi^{-1} \left[\lambda \psi^{-\tau} t^d_{0,0}(\lambda) + t^d_{0,0}(\lambda) - S_n \right];$$

adapting Lemma 3 gives

$$\lambda t^d_{0,0}(\lambda) = s^d_{0,0}(\tau, \beta_b) + O\left(\lambda^{-1}\right), \tag{4.11}$$

while

$$t^d_{0,0}(\lambda) = \lambda^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{\left(\frac{d_{ij}}{\exp(\beta_b x_i)}\right)^{\tau}}{1 + \frac{\left(\frac{d_{ij}}{\exp(\beta_b x_i)}\right)^{\tau}}{\lambda}} = O\left(\lambda^{-1}\right).$$

Now, employing (4.8), we have

$$\frac{\partial l_B}{\partial \psi} = \tau \psi^{-1} \left[\psi^{-\tau} s^d_{0,0}(\tau, \beta_b) - S_n + O\left(\lambda^{-1}\right) \right].$$

which gives

$$\psi^{-\tau} s^d_{0,0}(\tau, \beta_b)\Big|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = \left\{\exp(\widehat{\alpha}_w)\right\}^{-\widehat{B}} S_{0,0} = S_n \tag{4.12}$$

and hence we have

$$\lim_{\lambda\to\infty} \frac{\partial l_B}{\partial \psi}\Big|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = \lim_{\omega\to 0} \frac{\partial l_B}{\partial \psi}\Big|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = 0.$$

Behaviour of $\frac{\partial l_B}{\partial \beta_b}$ A similar argument applies to $\frac{\partial l_B}{\partial \beta_b}$, since we can write this derivative as

$$\frac{\partial l_B}{\partial \beta_b} = \tau \left[\lambda \psi^{-\tau} t^d_{0,1}(\lambda) + t^d_{0,1}(\lambda) - S_x \right].$$

Here,

$$t^d_{0,1}(\lambda) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_i \left[\frac{\frac{\left(\frac{d_{ij}}{\exp(\beta_b x_i)}\right)^{\tau}}{\lambda}}{1 + \frac{\left(\frac{d_{ij}}{\exp(\beta_b x_i)}\right)^{\tau}}{\lambda}} \right] = O\left(\lambda^{-1}\right)$$

and

$$\lambda t^d_{0,1}(\lambda) = s^d_{0,1}(\tau, \beta_b) + O\left(\lambda^{-1}\right),$$

by analogy with (4.11), while, adapting (4.12), and writing in full, we have

$$\psi^{-\tau} s^d_{0,1}(\tau, \beta_b) = \psi^{-\tau} \left[\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\frac{d_{ij}}{\exp(\beta_b x_i)} \right)^{\tau} x_i \right] = \sum_{i=1}^{k} n_i x_i = S_x.$$

Thus, we also have

$$\lim_{\lambda\to\infty} \frac{\partial l_B}{\partial \beta_b}\Big|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = \lim_{\omega\to 0} \frac{\partial l_B}{\partial \beta_b}\Big|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = 0.$$

Behaviour of $\frac{\partial l_B}{\partial \tau}$ For this derivative we have

$$\begin{aligned}
\frac{\partial l_B}{\partial \tau} &= S_n \tau^{-1} - S_n \ln \psi + s^d_{1,0}(0,0) - \beta_b S_x - \left(1 + \lambda \psi^{-\tau}\right) \left(\tau^{-1} \ln \lambda - \ln \psi\right) \left\{ t^d_{0,0}(\lambda) \right\} \\
&\quad - \left(1 + \lambda \psi^{-\tau}\right) t^d_{1,0}(\lambda) \\
&= S_n \tau^{-1} - S_n \ln \psi + s^d_{1,0}(0,0) - \beta_b S_x - \psi^{-\tau} \left(\tau^{-1} \ln \lambda - \ln \psi\right) \left\{ s^d_{0,0}(\tau, \beta_b) \right. \\
&\quad \left. + O\left(\lambda^{-1}\right) \right\} - \psi^{-\tau} \left\{ s^d_{1,0}(\tau, \beta_b) - \tau^{-1} \ln \lambda s^d_{0,0}(\tau, \beta_b) + O\left(\lambda^{-1}\right) \right\} \\
&\quad - \tau \ln\left(\lambda \psi^{-\tau}\right) O\left(\lambda^{-1}\right) + O\left(\lambda^{-1}\right),
\end{aligned}$$

using Lemma2, Lemma 3 and Remark2. Simplification yields

$$\begin{aligned}\frac{\partial l_B}{\partial \tau} &= S_n\tau^{-1} + s^d_{1,0}(0,0) - \beta_b S_x - \psi^{-\tau} s^d_{1,0}(\tau,\beta_b) + \ln\psi\left\{\psi^{-\tau} s^d_{0,0}(\tau,\beta_b) - S_n\right\} \\ &\quad + O\left(\lambda^{-1}\right) \\ &= S_n\tau^{-1} + s^d_{1,0}(0,0) - \beta_b S_x - \psi^{-\tau} s^d_{1,0}(\tau,\beta_b) + O\left(\lambda^{-1}\right) \qquad (4.13)\end{aligned}$$

using (4.12) for the term in brackets. In order to show that this expression is zero in the limit, we use the result at (4.9), together with the fact that we can write (4.12) as

$$\psi^{-\tau}\big|_{\psi=\exp(\widehat{\alpha}_w),\tau=\widehat{B}} = \frac{S_n}{s^d_{0,0}\left(\widehat{B},\widehat{\beta}_w\right)}.$$

Substituting into (4.13) gives

$$\lim_{\lambda\to\infty}\left.\frac{\partial l_B}{\partial\tau}\right|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = \lim_{\omega\to 0}\left.\frac{\partial l_B}{\partial\tau}\right|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = 0.$$

Behaviour of $\frac{\partial l_B}{\partial\omega}$ Finally, we turn our attention to $\frac{\partial l_b}{\partial\omega}$. From (4.10), we see the behaviour of this derivative is determined by the expression

$$\begin{aligned}& t^d_*(\lambda) - \frac{1+\lambda\psi^{-\tau}}{\lambda\psi^{-\tau}} t^d_{0,0}(\lambda) = \frac{\lambda\psi^{-\tau}\left\{t^d_*(\lambda)\right\} - \left(1+\lambda\psi^{-\tau}\right) t^d_{0,0}(\lambda)}{\lambda\psi^{-\tau}} \\ =\ & \lambda^{-1}\left[\lambda\left\{t^d_*(\lambda)\right\} - \lambda\left\{t^d_{0,0}(\lambda)\right\} - \psi^{\tau} t^d_{0,0}(\lambda)\right]\end{aligned}$$

as $\lambda\to\infty$. Using Lemma 1 to deal with $t^d_*(\lambda)$ and Lemma 3 for the other terms, this expression becomes

$$\begin{aligned}& \lambda^{-1}\left[\begin{array}{c} s^d_{0,0}(\tau,\beta_b) - \frac{1}{2\lambda} s^d_{0,0}(2\tau,\beta_b) + O\left(\lambda^{-2}\right) - \left\{s^d_{0,0}(\tau,\beta_b) - \frac{1}{\lambda} s^d_{0,0}(2\tau,\beta_b)\right. \\ \left. + O\left(\lambda^{-2}\right)\right\} - \frac{\psi^{\tau}}{\lambda}\left\{s^d_{0,0}(\tau,\beta_b) + O\left(\lambda^{-1}\right)\right\}\end{array}\right] \\ =\ & \lambda^{-2}\left[\frac{1}{2} s^d_{0,0}(2\tau,\beta_b) - \psi^{\tau} s^d_{0,0}(\tau,\beta_b)\right] + O\left(\lambda^{-3}\right).\end{aligned}$$

So, as $\lambda\to\infty$, the sign of

$$\left.\frac{\partial l_B}{\partial\omega}\right|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w}$$

is determined by

$$\Delta_d = \frac{s^d_{0,0}\left(2\widehat{B},\widehat{\beta}_w\right)}{2} - \frac{\left[s^d_{0,0}\left(\widehat{B},\widehat{\beta}_w\right)\right]^2}{S_n} \qquad (4.14)$$

where, we recall, we have

$$s^d_{0,0}\left(r,\widehat{\beta}_w\right) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{d_{ij}}{\exp\left(\widehat{\beta}_w x_i\right)}\right)^r.$$

The function Δ_d now acts as a test statistic for determining which of the Burr XII and Weibull ALT models provides the better fit - in terms of greater maximised log-likelihood - to a given data set; it is the natural extension to the discriminating factor derived by Watkins (2001). A positive Δ_d implies l_B will increase with ω, and the maximum value of l_B will be found at positive ω, corresponding to finite a and the Burr XII ALT model. A negative Δ_d implies l_B will decrease with increasing ω, and the maximum value of l_B will be found at $\omega = 0$, corresponding to infinite a and the Weibull ALT model. For convenience, we can rescale the value of Δ_d by dividing each d_{ij} by $\exp(\widehat{\alpha}_w)$, giving

$$\begin{aligned}\Delta_s\left(\widehat{B},\widehat{\alpha}_w,\widehat{\beta}_w\right) &= \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{d_{ij}}{\exp\left(\widehat{\alpha}_w+\widehat{\beta}_w x_i\right)}\right)^{2\widehat{B}} \\ &\quad -\frac{1}{S_n}\left\{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{d_{ij}}{\exp\left(\widehat{\alpha}_w+\widehat{\beta}_w\right)}\right)^{\widehat{B}}\right\}^2 .\end{aligned} \tag{4.15}$$

Reparameterising the scale parameter α_b

This reparameterisation is based on that in Watkins (2001); we begin by letting

$$\exp(\alpha_b) = -\ln\omega$$

so that

so that

$$\alpha_b = \ln(-\ln\omega).$$

Th

Thus, as $\alpha_b \to \infty$, $\omega \to 0$. Rewriting (4.3) gives us

$$a = \left(\frac{\exp(\alpha_b)}{\psi}\right)^{\tau} = \left(\frac{-\ln\omega}{\psi}\right)^{\tau}$$

as the reparameterised expression for a. Under this reparameterisation, l_B at (2.10) becomes

$$\begin{aligned} l_B(\omega,\psi,\tau,\beta_b) &= S_n\ln\tau+\tau S_n\ln(-\ln\omega)-\tau S_n\ln\psi-\tau S_n\ln(-\ln\omega)+(\tau-1)S_e \\ &\quad -\tau\beta_b S_x-\left\{1+\left(\frac{-\ln\omega}{\psi}\right)^{\tau}\right\}S_d^*(\omega) \\ &= S_n\ln\tau-\tau S_n\ln\psi+(\tau-1)S_e-\tau\beta_b S_x \\ &\quad -S_d^*(\omega)\left\{1+\left(\frac{-\ln\omega}{\psi}\right)^{\tau}\right\} \end{aligned} \tag{4.16}$$

where we use $S_d^*(\omega)$ to indicate the role of ω in the expression (2.11), which in full is

$$S_d^*(\omega) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln\left[1+\left\{\frac{\frac{d_{ij}}{\exp(\beta_b x_i)}}{(-\ln\omega)}\right\}^{\tau}\right].$$

We also define

$$t^d_{v,w}(\omega) = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_i^w \left[\frac{\left(\frac{d_{ij}}{\exp(\beta_b x_i)}\right)^\tau \left\{ \ln\left(\frac{d_{ij}}{\exp(\beta_b x_i)}\right)\right\}^v}{1+\left(\frac{d_{ij}}{\exp(\beta_b x_i)}\right)^\tau} \right],$$

which will prove convenient when we examine the score of l_B. Note the similarity between this expression and that at (4.7), defined for the previous reparameterisation. The $\psi\left[\ln\left(-\ln\omega\right)\right]^{1/\tau}$ there is now replaced by $(-\ln\omega)$.

First Derivatives The first derivatives of (4.16) are

$$\begin{aligned}
\frac{\partial l_B}{\partial \psi} &= \tau\psi^{-1}\left[\left(\frac{-\ln\omega}{\psi}\right)^\tau S^*_d(\omega) - S_n\right], \\
\frac{\partial l_B}{\partial \tau} &= S_n\tau^{-1} - S_n\ln\psi + S_e - \beta_b S_x - S^*_d(\omega)\left(\frac{-\ln\omega}{\omega}\right)^\tau \ln\left(\frac{-\ln\omega}{\psi}\right) \\
&\quad - \left\{1+\left(\frac{-\ln\omega}{\psi}\right)^\tau\right\} t^d_{1,0}(\omega), \\
\frac{\partial l_B}{\partial \omega} &= \frac{\tau}{\omega(-\ln\omega)}\left[\left(\frac{-\ln\omega}{\psi}\right)^\tau S^*_d(\omega) - \left\{1+\left(\frac{-\ln\omega}{\psi}\right)^\tau\right\} t^d_{0,0}(\omega)\right]
\end{aligned}$$

and

$$\frac{\partial l_B}{\partial \beta_b} = -\tau S_x + \tau t^d_{0,1}(\omega)\left\{1+\left(\frac{-\ln\omega}{\psi}\right)^\tau\right\}.$$

If we now make the substitution

$$\lambda = (-\ln\omega)^\tau \tag{4.17}$$

then we have

$$\ln\left(-\ln\omega\right) = \tau^{-1}\ln\lambda$$

and all Lemmas and Remarks above remain valid here.

Behaviour of $\frac{\partial l_B}{\partial \psi}$ Using (4.17) we can write $\frac{\partial l_B}{\partial \psi}$ as

$$\frac{\partial l_B}{\partial \psi} = \tau\psi^{-1}\left[\lambda\psi^{-\tau}S^*_d(\omega) - S_n\right].$$

Using Remark 1, with S^*_d in place of t^d_*, gives us

$$\frac{\partial l_B}{\partial \psi} = \tau\psi^{-1}\left[\psi^{-\tau}S^d_{0,0}(\tau,\beta_b) - S_n + O\left(\lambda^{-1}\right)\right],$$

while substituting in the result (4.12) gives

$$\lim_{\lambda\to\infty}\left.\frac{\partial l_B}{\partial\psi}\right|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w}=0$$

as the required limiting result.

Behaviour of $\frac{\partial l_B}{\partial\beta_b}$ Using (4.17), we can write this derivative as

$$\begin{aligned}\frac{\partial l_B}{\partial\beta_b} &= -\tau S_x+\tau\left\{1+\lambda\psi^{-\tau}\right\}t^d_{0,1}(\omega)\\ &= \tau\left[\lambda\psi^{-\tau}t^d_{0,1}(\lambda)+t^d_{0,1}(\lambda)-S_x\right],\end{aligned}$$

where we now write $t^d_{0,1}$with the λ notation to indicate the role of this parameter in the expression. We recognise this derivative as having the same form as in the previous reparameterisation and consequently the same arguments can be used to arrive at the same result; namely

$$\lim_{\lambda\to\infty}\left.\frac{\partial l_B}{\partial\beta_b}\right|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w}=0.$$

Behaviour of $\frac{\partial l_B}{\partial\tau}$ Using (4.17), we can write this derivative as

$$\begin{aligned}\frac{\partial l_B}{\partial\tau} &= S_n\tau^{-1}-S_n\ln\psi+S_e-\beta_bS_x-S^*_d(\lambda)\lambda\psi^{-\tau}\left\{\tau^{-1}\ln\lambda-\ln\psi\right\}\\ &\quad -t^d_{1,0}(\lambda)\left\{1+\lambda\psi^{-\tau}\right\}.\end{aligned}$$

As with the previous reparameterisation, we use Remark 1, Remark 2 and Lemma 2 to write

$$\begin{aligned}\frac{\partial l_B}{\partial\tau} &= S_n\tau^{-1}-S_n\ln\psi+S_e-\beta_bS_x-\psi^{-\tau}\left\{\tau^{-1}\ln\lambda-\ln\psi\right\}\left[S^d_{0,0}(\tau,\beta_b)+O\left(\lambda^{-1}\right)\right]\\ &\quad -\psi^{-\tau}\left[S^d_{1,0}(\tau,\beta_b)-\tau^{-1}\ln\lambda S^d_{0,0}(\tau,\beta_b)+O\left(\lambda^{-1}\right)\right]+O\left(\lambda^{-1}\right)\\ &= S_n\tau^{-1}+S_e-\beta_bS_x-\psi^{-\tau}S^d_{1,0}(\tau,\beta_b)+\ln\psi\left\{\psi^{-\tau}S^d_{0,0}(\tau,\beta_b)-S_n\right\}+O\left(\lambda^{-1}\right),\end{aligned}$$

upon simplification. Using (4.12) for $\psi^{-\tau}S^d_{0,0}(\tau,\beta_b)-S_n$, together with (4.9) - upon recognising that $s^d_{1,0}(0,0)$ is simply S_e here - we have

$$\lim_{\lambda\to\infty}\left.\frac{\partial l_B}{\partial\tau}\right|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w}=0$$

as the required limiting result.

Behaviour of $\frac{\partial l_B}{\partial\omega}$ We can see that this derivative is effectively the same as for that derived by Watkins (2001) for the non-accelerated case. Examination of the terms involved shows that, in making the extension to the ALT framework, we need only replace the single

sample size m by the sum of the sub-samples S_n and the generic data value d by $\frac{d}{\exp(\beta_b x)}$. Since we can employ the same Remarks and Lemmas, as used by Watkins, we will inevitably end up with same result as for the non-ALT case, now adapted to accomodate acceleration. As such, the behaviour of $\frac{\partial l_B}{\partial \omega}$ is governed by

$$\Delta_d = \frac{s^d_{0,0}\left(2\widehat{B}, \widehat{\beta}_w\right)}{2} - \frac{\left[s^d_{0,0}\left(\widehat{B}, \widehat{\beta}_w\right)\right]^2}{S_n}$$

for the raw data values. This is the same result as that obtained at (4.14).

4.2.3 Examples

Example 2.1.1.1 [Carbon Fibre Rod Data] revisited

Through the SAS IML code *burr_alt*, (see Appendix A), we found the value of Δ_s to be

$$\Delta_s = 30.283411,$$

while the corresponding result for the raw data was

$$\Delta_d = 10121331.$$

This suggests that for the carbon fibre rod data the Burr XII ALT model is a superior fit than the Weibull ALT model. This is consistent with the respective likelihoods in Table 2.16.

Example 2.1.1.2 [Aluminium Coupon Data] revisited

The value of Δ_s for the aluminium coupon data was

$$\Delta_s = 55.769813,$$

while the corresponding result for the raw data was

$$\Delta_d = 8.3439 \times 10^{35},$$

suggesting the Burr XII ALT model is preferred to the Weibull ALT model; a result that is consistent with the log-likelihoods observed in Table 2.18.

Example 2.1.1.3 [Electrode Data] revisited

For the electrode data, the value of Δ_s was

$$\Delta_s = -9.412877,$$

while the corresponding result for the raw data was

$$\Delta_d = -6.939 \times 10^{37},$$

suggesting the Weibull ALT model provides a better fit to the data set than the Burr XII ALT model. Again, this is consistent with the likelihoods in Table 2.20.

These examples are encouraging but only illustrate results which are expected. We observed for the carbon fibre rod data subsets that the discriminating factor was positive for all data subsets, see Table 4.1, while, in Table 4.3, the discriminating factor was negative for all electrode data subsets. As in the case of the carbon fibre rod data, for the aluminium coupon data, the discriminating factor was seen to be positive for all data subsets, see Table 4.2. It should therefore come as no surprise to observe that for the corresponding ALT models, Δ was positive for the carbon fibre rod data and the aluminium coupon data, while Δ was negative for the electrode data. Furthermore, it is possible to have a data set in which (4.1) is negative for each subset, while Δ is positive for the ALT model. We consider this scenario next, before moving on to look at an example whereby some data subsets exhibit a positive value for the discriminating factor and others a negative value.

Example 4.2.1.1 [Nelson (1990,p232)]

As with Example 2.1.1.3, the data here concerns disc electrodes immersed in an insulating oil, at three different temperatures. The experiment is conducted in exactly the same way, except now the electrodes are 1 square inch in area. The maximum likelihood estimates of the Weibull ALT parameters are $\widehat{B} = 11.345806$, $\widehat{\alpha}_w = 3.8844387$ and $\widehat{\beta}_w = 0.000228$. Consequently, the estimates of the scales are $\left(\widehat{\phi}_1, \widehat{\phi}_2, \widehat{\phi}_3\right) = (48.750648, 49.761233, 61.093996)$, with log-likelihood, based on (2.3), $l_W = -550.5821$. Meanwhile, the maximum likelihood estimates of the Burr XII ALT parameters are $\widehat{a} = 34.690467$, $\widehat{\tau} = 11.533921$, $\widehat{\alpha}_b = 4.1895109$ and $\widehat{\beta}_b = 0.0002293$, with scale estimates $\left(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3\right) = (66.14199, 67.52115, 82.99753)$. The log-likelihood was $l_B = -550.5636$. We see that the log-likelihood for the Burr XII model is slightly greater than that for the Weibull model, suggesting a preference for the former model. However, it should be noted that the Weibull shape parameter is quite high, suggesting that this data set may not follow a two parameter Weibull distribution, with the data suggesting a non-zero location parameter. Again, we include this example purely for illustrative purposes.

The values of the discriminating factor observed for the individual scaled data subsets are shown in Table 4.4 and are seen to be negative in each case, suggesting the Weibull distribution is preferred to the Burr XII distribution for each data subset. The corresponding result for the ALT model is

$$\Delta_s = 1.30661;$$

volts/second	10	100	1000
n	60	60	60
Discriminating Factor	−2.46619	−4.76249	−0.36956

Table 4.4: Discriminating factor for subsets of voltage data, for 1 sq in electrodes.

for the raw data we have

$$\Delta_d = 2.4928 \times 10^{38}.$$

The positive values of Δ_s and Δ_d, on the other hand, indicate a preference for the Burr XII ALT model over the Weibull ALT model.

Example 4.2.1.2 [Simulated Data]

This example considers a data set, shown in Table 4.5, generated from a Weibull ALT distribution, with arbitrarily chosen parameter values, $B = 3$ and β_w, α_w chosen such that ϕ_1, ϕ_2 and ϕ_3 are approximately 100, 60 and 30 respectively, (we use $\beta_w = -0.012206$ and $\alpha_w = 4.621869$.) Stress levels $(x_1, x_2, x_3) = (10, 50, 100)$ and equal subset sample size $n_1 = 30$ were also chosen arbitrarily. The resultant Weibull and Burr XII ALT parameter estimates, together with maximised log-likelihoods and values of the discriminating factor for the scaled data subsets are given in Table 4.6. The maximum likelihood estimates for the Weibull ALT parameters were $\widehat{B} = 3.0701136$, $\widehat{\alpha}_w = 4.6119034$ and $\widehat{\beta}_w = -0.011191$, with maximised log-likelihood $l_W = -384.2829$. The estimates of the scales were then $\left(\widehat{\phi}_1, \widehat{\phi}_2, \widehat{\phi}_3\right) = (90.016778, 57.533703, 32.879174)$. The maximum likelihood estimates for the Burr XII ALT parameters were $\widehat{a} = 5.844251$, $\widehat{\tau} = 3.42111$, $\widehat{\alpha}_b = 5.09533$ and $\widehat{\beta}_b = -0.011302$, with maximised log-likelihood $l_B = -383.2033$. The estimates of the scales were then $\left(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3\right) = (145.8108, 92.7798, 52.7271)$. The corresponding ALT value of Δ_s was then found to be

$$\Delta_s = 6.20051$$

while for the raw data set, we have

$$\Delta_d = 1.2326 \times 10^{13}.$$

We can see that even though two of the data subsets returned a negative value for the discriminating factor, the ALT values of Δ_s and Δ_d are positive. The extent to which the magnitude and sign of the discriminating factor for data subsets influence the ALT value of Δ, (and our ability to predict this value), is beyond the scope of the current discussion. We include this example simply to illustrate the possible outcomes of an experiment.

x_1=10									
21.16	88.95	84.47	90.18	76.79	43.38	86.25	71.30	59.73	79.14
52.18	89.79	89.06	82.73	89.48	49.58	131.66	141.01	89.51	95.89
33.93	108.76	72.80	109.91	93.71	29.00	114.71	39.30	136.08	98.94
x_2=50									
25.86	58.21	38.69	40.12	57.64	40.81	51.69	65.87	30.50	81.91
56.72	44.76	56.94	72.75	48.60	29.72	29.26	95.99	66.91	37.95
17.99	37.53	31.78	61.25	52.07	71.87	50.71	25.44	42.34	55.05
x_3=100									
18.38	43.61	41.45	23.52	37.35	27.22	60.28	34.70	18.41	28.14
32.70	20.77	29.38	27.76	26.28	31.25	21.87	13.51	25.55	18.99
16.78	40.95	32.11	24.47	37.77	26.84	37.63	31.42	36.70	40.32

Table 4.5: Data set simulated from Weibull ALT distribution. Parameter values, stress levels and subset sample size chosen arbitrarily. β_w and α_w chosen to tie in with stress levels.

Stress	**10**	**50**	**100**
n	30	30	30
$\widehat{B}$	3.0101439	2.974225	4.1709285
$\widehat{\phi}$	91.441267	55.210235	32.211097
l_w	-144.5186	-128.4725	-104.5113
$\widehat{a}$	16.77744	3.307549	1.884797
$\widehat{\tau}$	3.119576	3.545053	4.55167
$\widehat{\theta}$	223.2216	72.28226	35.07435
l_b	-144.6594	-128.1714	-109.5815
Discriminating Factor	-2.235591	2.436463	-2.858899

Table 4.6: Weibull and Burr XII parameter estimates, together with the discriminating factor for data subsets when data is simulated from a Weibull ALT distribution

4.3 Behaviour of Δ

In most of the previous examples, the data were observed in real life experiments, and thus may not follow any underlying distribution. All we can do is fit models to the data and observe, through maximised log-likelihoods and the concept of model misspecification, which provides the best fit from a potentially infinite set of models. Moreover, in our study, we are only concerned with which of the Weibull and Burr XII models provides the better fit, and we have seen that we can make this determination using a single algebraic expression, labelled Δ.

In this section, we are interested in the general behaviour of Δ, in particular its distribution, expectation, standard deviation and skewness. Here, since our concern lies with the Burr XII and Weibull distributions, we shall focus on the properties of Δ when the data comes from each of these distributions in turn. Now, since we can write Δ as a summation, we would expect, by the Central Limit Theorem, that with increasing sample sizes, the distribution of Δ will be approximately Normal. As such, determining the expectation and variance of Δ would tell us everything about its distribution, at least for large sample sizes; although it is also possible that small sample approximations will be accurate enough to satisfy our needs.

To investigate this possibility further we consider two scenarios: one where the data is generated from a Weibull distribution and the other where the data is generated from a Burr XII distribution. For both cases, we calculate the value of Δ_s for an array of possible parameter values and sample sizes and report the proportion of times Δ_s returned a positive value. We perform 10000 replications and assess Normality through Q-Q plots and the Kolmogorov-Smirnov (K-S) Test Statistic, see for example Lawless (1982), which also summarises some critical values of the test. To assist in the interpretation of the tables below, we state here the $90^{th}, 95^{th}$ and 99^{th} quantiles for the K-S test as $1.224, 1.358$ and 1.628 respectively. With Δ_s offering more numerical stability than Δ_d, we shall concentrate on this value throughout the investigations that follow. However, to maintain some consistency across each of the 10000 separate simulations, we use the known value of the parameter α_w, rather than its MLE - $\widehat{\alpha}_w$ - in the scaled version of Δ_s, when the data come from a Weibull distribution, and the known value α_b in place of $\widehat{\alpha}_w$ when the data comes from a Burr XII distribution.

4.3.1 When Data is From a Weibull ALT Distribution

For our first example the parameter values were chosen to be $k = 3$, $B = 3, \alpha_w = 17.60139$ and $\beta_w = -0.056282$. The stress levels were chosen to be $(x_1, x_2, x_3) = (150, 180, 190)$ and, consequently, the scales are $(\theta_1, \theta_2, \theta_3) = (9500, 1755, 1000)$. Six equally weighted sample sizes are considered, with values $n_1 = 30, 50, 100, 500, 1000$ and 5000. The results are presented in Table 4.7.

Figures 4.13 to 4.18, together with summaries in Table 4.7, seem to suggest that as the

	Summary Statistics for Δ_s				
n_1	Mean	S.D.	Skewness	K-S Test	Percent > 0
30	−6.6766	206.8953	−7.107	37.363	38.25
50	−4.0722	110.1310	−10.699	33.758	40.75
100	−2.5717	38.1255	−3.390	21.872	42.23
500	−1.2499	31.3618	0.210	6.681	46.55
1000	−0.2394	39.3449	0.171	4.374	49.31
5000	−1.9916	79.2920	0.091	1.361	48.05

Table 4.7: Example 1: Summary statistics for properties of Δ_s for varying sample sizes when data is from a Weibull ALT distribution.

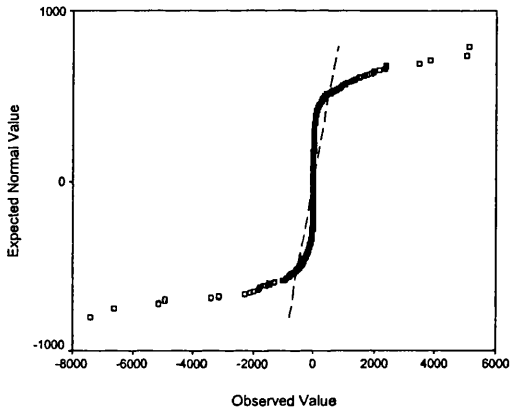

Figure 4.13: Distribution of Δ_s for example one: $n_1 = 30$. Here, as throughout, the theoretical Normal quantiles are indicated by the broken line.

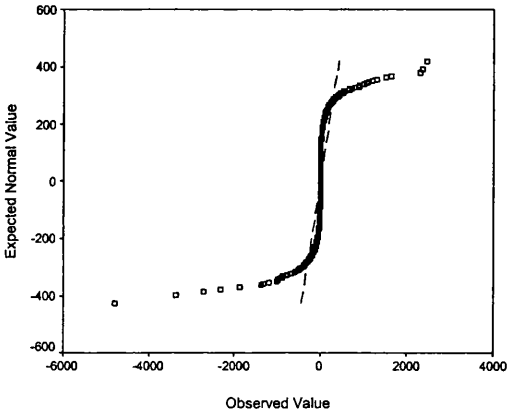

Figure 4.14: Distribution of Δ_s for example one: $n_1 = 50$

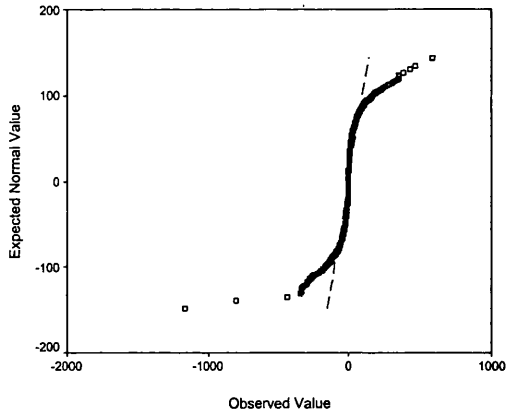

Figure 4.15: Distribution of Δ_s for example one: $n_1 = 100$

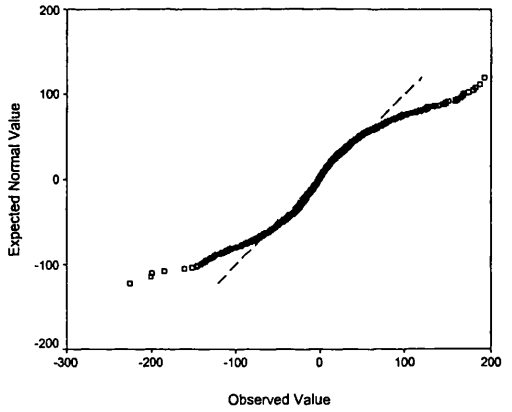

Figure 4.16: Distribution of Δ_s for example one: $n_1 = 500$

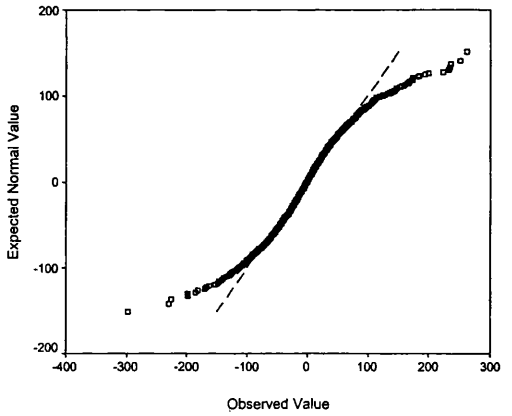

Figure 4.17: Distribution of Δ_s for example one: $n_1 = 1000$

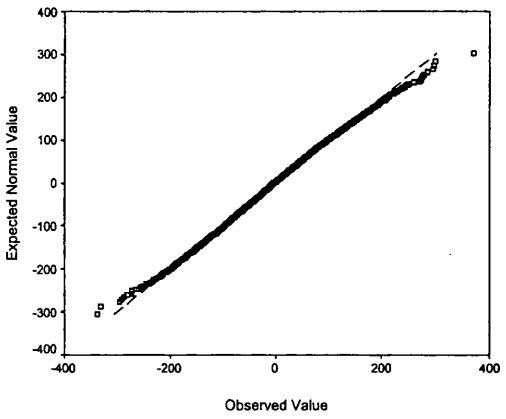

Figure 4.18: Distribution of Δ_s for example one: $n_1 = 5000$

sample size increases, the distribution of Δ_s becomes more and more Normal - as indicated by the decreasing K-S statistic - and more centred around zero, as shown by the final column in the table, where we see the percentage of values greater than zero approach 50% as the sample size increases. However, for no sample size considered is the Normal distribution regarded as a suitable model for the distribution of Δ_s, as indicated by the K-S test statistics. As such, it seems the best we can say is that half of the time we would prefer to fit the Weibull distribution and half the time we would prefer to fit the Burr XII distribution. We now consider some alternative experimental designs to determine the extent to which this is true in general.

The next example is based on a data set from Nelson (1990, p158), with parameter values $B = 2.5554$, $\alpha_w = 21.9087$ and $\beta_w = -0.0563$; these being the maximum likelihood estimates derived by, for example, Watkins (1994). The associated stress levels were $(x_1, x_2, x_3, x_4, x_5) = (200, 215, 230, 245, 260)$. To ensure the same total sample size as in the previous example, we define $n_1 = 18, 30, 60, 300, 600$ and 3000 and tabulate the same collection of summary statistics as before, see Table 4.8. From Figures 4.19 to 4.24, we tend to observe the same pattern as in the previous example, with disagreement between theoretical and observed results in the tails becoming less pronounced as the sample size increases; consequently, the distribution of Δ_s becomes more Normal in this limit. This is supported by the K-S test statistics; although the Normal distribution is not regarded as satisfactory for all bar the largest sample size considered. Also, the percentage of values that return a positive Δ_s steadily rises towards 50% as the sample size increases. As before, we conclude that we prefer to fit the Weibull distribution half the time and the Burr XII distribution half the time.

In our final example we consider parameter values based on the electrode data considered in example 2.1.1.3. For the purposes of illustration, we shall use the maximum likelihood estimates of the parameters, given in section 2.1, as the true parameter values. Consequently,

	Summary Statistics for Δ_s				
n_1	**Mean**	**S.D.**	**Skewness**	**K-S Test**	**Percent > 0**
18	−27.2598	1153.0783	−15.245	43.119	39.15
30	−5.9759	157.0178	−3.882	35.387	40.40
60	−2.3821	46.3722	0.080	23.591	43.39
300	−0.6752	33.5111	0.641	8.526	47.48
600	−1.6654	40.4739	0.178	4.741	47.78
3000	−2.0531	79.0591	0.052	1.339	48.56

Table 4.8: Example 2: Summary statistics for properties of Δ_s for varying sample sizes when data is from a Weibull ALT distribution.

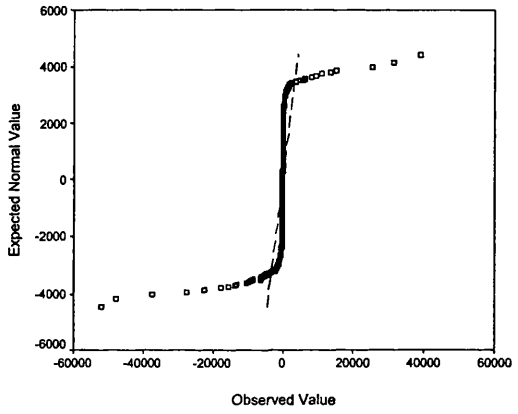

Figure 4.19: Distribution of Δ_s for example two: $n_1 = 18$

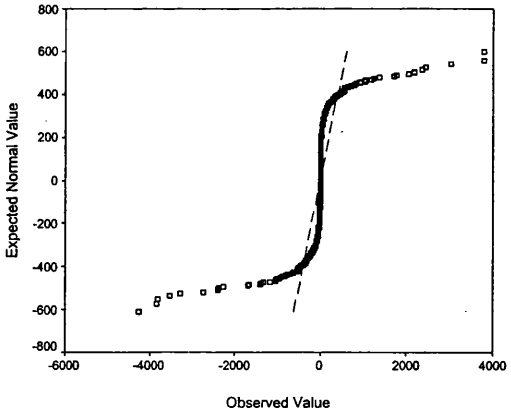

Figure 4.20: Distribution of Δ_s for example two: $n_1 = 30$

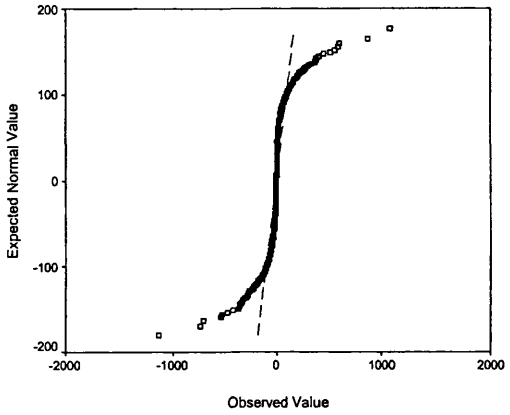

Figure 4.21: Distribution of Δ_s for example two: $n_1 = 60$

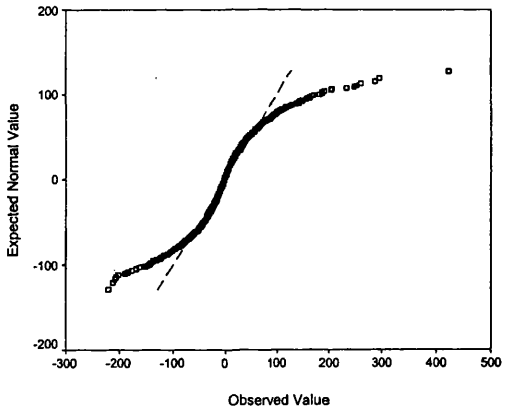

Figure 4.22: Distribution of Δ_s for example two: $n_1 = 300$

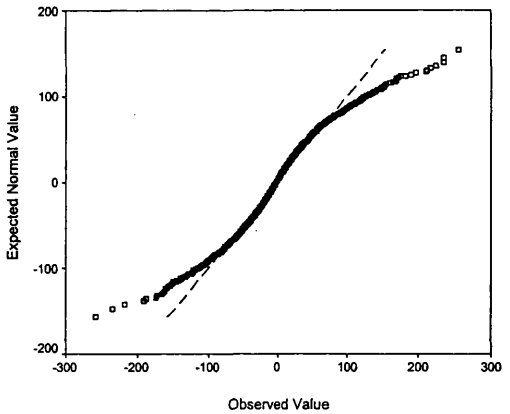

Figure 4.23: Distribution of Δ_s for example two: $n_1 = 600$

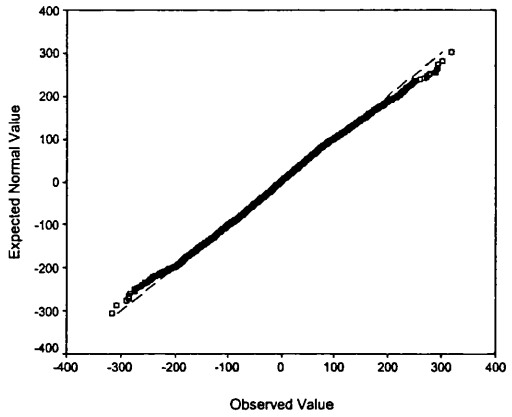

Figure 4.24: Distribution of Δ_s for example two: $n_1 = 3000$

$B = 11.280264$, $\alpha_w = 3.7627961$ and $\beta_w = 0.0001656$, with $(x_1, x_2, x_3) = (10, 100, 1000)$. As in the first example above, we take $n_1 = 30, 50, 100, 500, 1000$ and 5000. The relevant summary statistics are presented in Table 4.9. Clearly, we can only draw the same conclusions as in the previous two examples, with the distribution of Δ_s becoming more Normal as the sample size increases. This is apparent from the Q-Q plots, (Figures 4.25 to 4.30), which are then re-enforced by the K-S statistics. We see the mean for each sample size at around -1 with associated standard deviation ranging from 5.6 to 76.9. As before, the percentage of values of Δ_s observed to be greater than zero steadily rises to around 50%, with increasing sample size. Naturally we are drawn to the same conclusion as before: in order for the distribution of Δ_s to be Normal, we require a very large sample size; but for such a sample size, the chances of preferring to fit a Weibull distribution as opposed to a Burr XII are at best 50:50.

Now, our initial hope was that the distribution of Δ_s would be approximately Normal, even for relatively small sample sizes, thereby allowing us to draw conclusions about the probability of obtaining a positive or negative value of Δ_s, using only the mean and standard deviation. However, these three examples show that our ability to accurately predict the sign of Δ_s for a given set of parameter values is very limited. As such, there is limited benefit in finding the theoretical mean and variance of Δ_s, at least when the data comes from a Weibull ALT distribution. Granted, for any individual data set we can say whether the Burr XII ALT distribution will provide a superior fit, simply by fitting a Weibull ALT distribution; but are in no position to say which distribution will, in probability, provide the better fit for a specific set of parameter values.

4.3.2 When Data is From a Burr XII ALT Distribution

We now perform a similar series of investigations, this time with data generated from a Burr XII ALT distribution. Our first example is similar to the first considered above,

	Summary Statistics for Δ_s				
n_1	**Mean**	**S.D.**	**Skewness**	**K-S Test**	**Percent > 0**
30	−0.9781	5.6356	0.823	5.221	38.72
50	−0.9333	7.4024	0.583	4.059	41.34
100	−1.0416	10.6799	0.507	3.638	42.60
500	−0.9689	24.6386	0.283	2.375	46.20
1000	−0.7566	34.1049	0.193	1.437	47.99
5000	0.2013	76.9361	0.111	1.370	48.99

Table 4.9: Example 3 : Summary statistics for properties of Δ_s for varying sample sizes when data is from a Weibull ALT distribution.

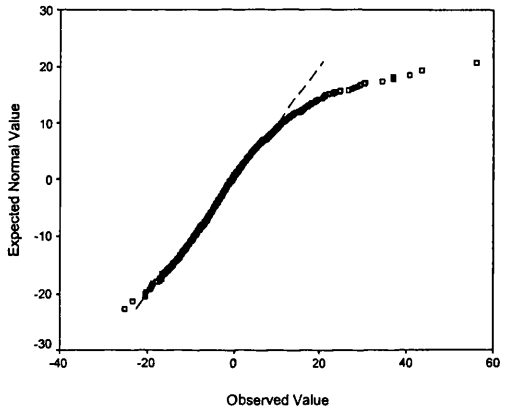

Figure 4.25: Distribution of Δ_s for example three: $n_1 = 30$

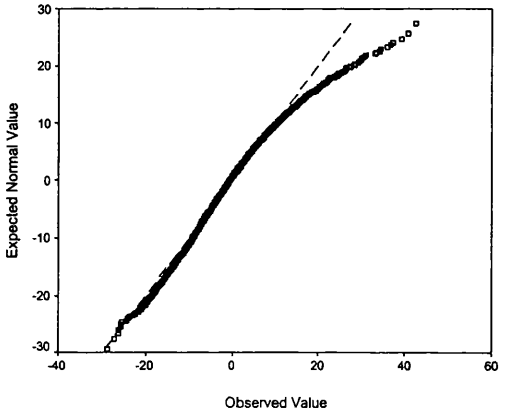

Figure 4.26: Distribution of Δ_s for example three: $n_1 = 50$

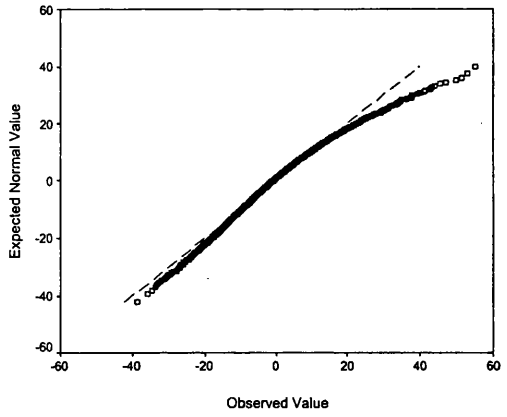

Figure 4.27: Distribution of Δ_s for example three: $n_1 = 100$

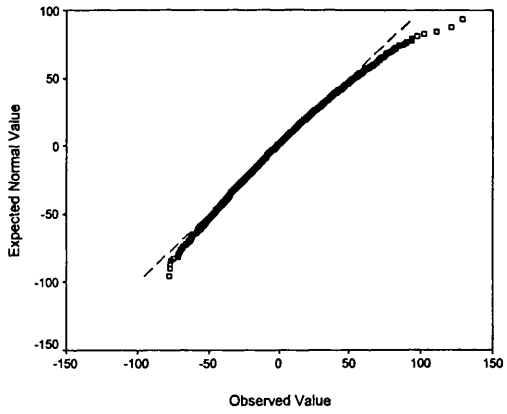

Figure 4.28: Distribution of Δ_s for example three: $n_1 = 500$

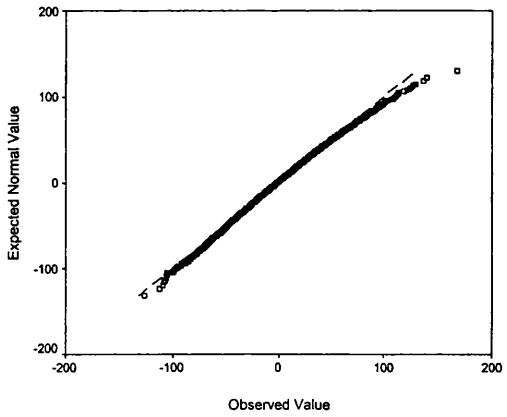

Figure 4.29: Distribution of Δ_s for example three: $n_1 = 1000$

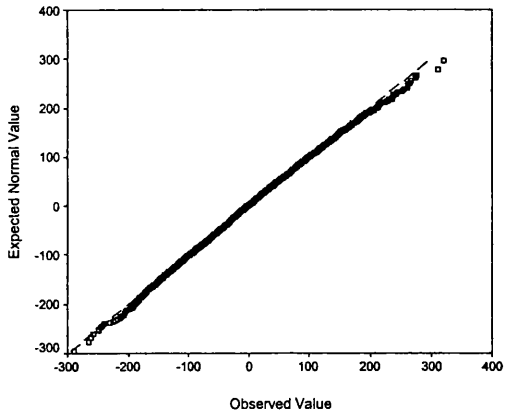

Figure 4.30: Distribution of Δ_s for example three: $n_1 = 5000$

	Summary Statistics for Δ_s				
n_1	**Mean**	**S.D.**	**Skewness**	**K-S Test**	**Percent > 0**
30	313.5905	3045.6818	24.067	44.942	95.09
50	178.6159	1012.1509	19.818	42.572	99.34
100	172.7275	570.7698	11.456	38.192	100.00
500	488.0724	691.1044	7.576	25.731	100.00
1000	889.3060	884.6590	8.057	20.605	100.00
5000	4299.1489	2126.3733	4.847	15.414	100.00

Table 4.10: Example 4: Summary of statistics for properties of Δ_s for varying sample sizes when data is from a Burr XII ALT distribution.

where the data came from a Weibull distribution, and takes the parameter values $a = 2$, $\tau = 5$, $\alpha_b = 17.60139$ and $\beta_b = -0.056282$; with $(x_1, x_2, x_3) = (150, 180, 190)$. We take $n_1 = 30, 50, 100, 500, 1000$ and 5000. The summary statistics of the results are presented in Table 4.10. We see from the K-S test statistics, that for any given sample size, the distribution is clearly non-Normal; a result which is then endorsed by the associated Q-Q plots, see Figures 4.31 to 4.36.

From this initial example, it appears that as the sample size increases we become more confident that we shall obtain a positive value of Δ_s, with the percentage of positive values observed rapidly reaching 100%, while, in the same limit, the distribution of Δ_s tends to become more Normal (on the basis of the K-S statistic) albeit never significantly for the sample sizes considered.

We now consider some alternative experimental designs to assess the extent to which these conclusions can be considered typical. Consider the randomly chosen shape parameters $a = 3$ and $\tau = 3.5$, with $k = 4$ and stress levels $(x_1, x_2, x_3, x_4) = (100, 150, 200, 250)$. In order to observe a suitable degree of acceleration, we determined the scale parameters to be $\alpha_b = 8.44285$ and $\beta_b = -0.015351$, giving the scales $(\theta_1, \theta_2, \theta_3, \theta_4) = (1000, 464.14, 215.43, 100)$, to two decimal places. To ensure total sample sizes as close to those used in the previous

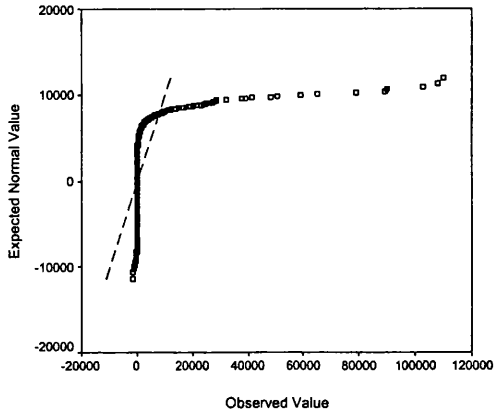

Figure 4.31: Distribution of Δ_s for example four: $n_1 = 30$

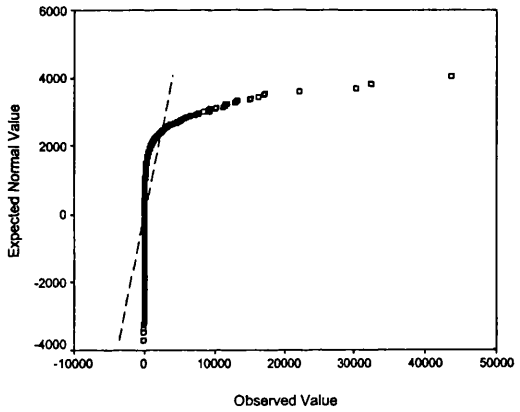

Figure 4.32: Distribution of Δ_s for example four: $n_1 = 50$

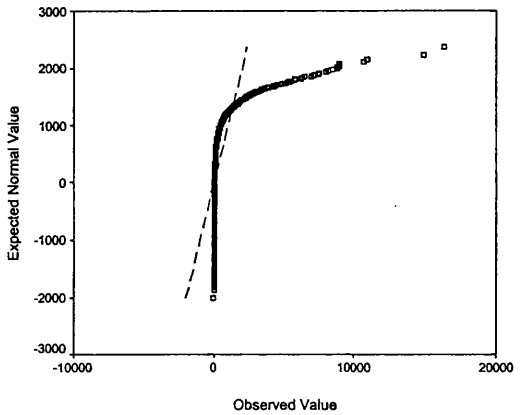

Figure 4.33: Distribution of Δ_s for example four: $n_1 = 100$

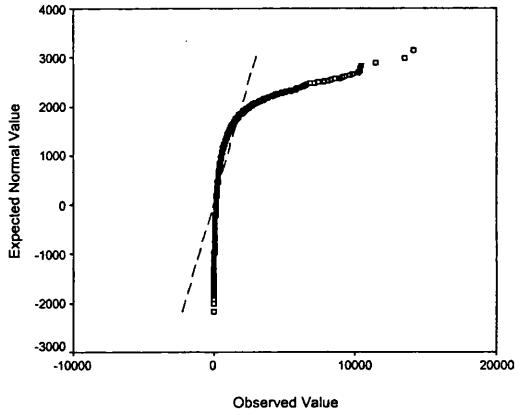

Figure 4.34: Distribution of Δ_s for example four: $n_1 = 500$

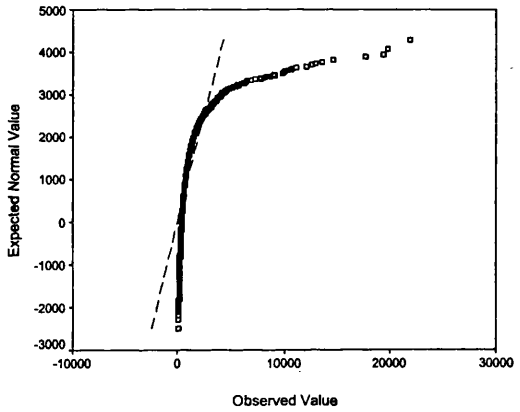

Figure 4.35: Distribution of Δ_s for example four: $n_1 = 1000$

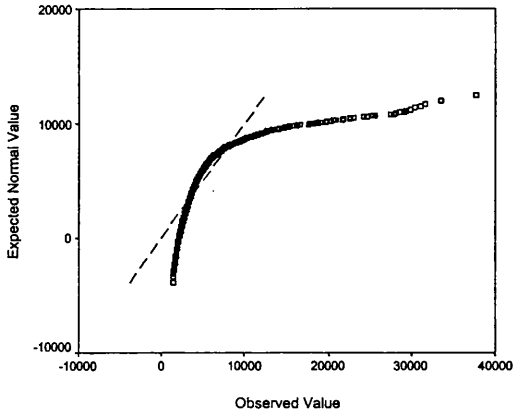

Figure 4.36: Distribution of Δ_s for example four: $n_1 = 5000$

	Summary Statistics for Δ_s				
n_1	Mean	S.D.	Skewness	K-S Test	Percent > 0
22	3.6085	6.5427	4.924	21.301	86.63
37	6.6389	10.0280	6.553	21.116	95.35
74	14.1011	15.9297	6.333	18.759	99.40
370	77.1711	42.9796	3.599	13.407	100.00
740	157.6681	66.4984	2.567	12.942	100.00
3700	815.1305	204.8269	3.393	12.145	100.00

Table 4.11: Example 5: Summary statistics for properties of Δ_s for varying sample sizes when data is from a Burr XII ALT distribution.

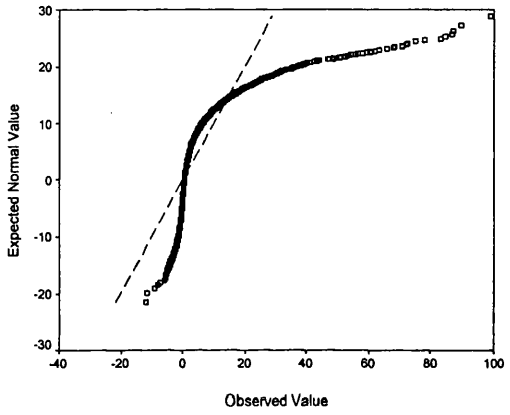

Figure 4.37: Distribution of Δ_s for example five: $n_1 = 22$

example, we set $n_1 = 22, 37, 74, 370, 740$ and 3700. The resultant summary statistics are presented in Table 4.11, while the Q-Q plots are presented in Figures 4.37 to 4.42. As in the previous example, we see the distribution of Δ_s becomes more Normal (but never significantly so) as the sample size increases. At the same time, we see the percentage of positive values of Δ_s rapidly increase to 100%. With both the mean and the standard deviation of Δ_s rising with increasing sample size - the standard deviation at a slower rate than the mean - we see that we are, at least, more likely to observe a positive value of Δ_s in this limit. However, as in the previous case, the consistent lack of Normality makes further inferential progress troublesome.

We now consider one final example. Somewhat arbitrarily, we define $a = 5.4$, $\tau = 2.1$ and $(x_1, x_2, x_3) = (10, 30, 50)$. We define the scales θ_i to be $(1000, 500, 100)$, thereby giving us an suitable acceleration factor. Using (1.12), we determine $\beta_b = -0.05756$ and $\alpha_b = 7.9415$. Substituting the values of α_b and β_b in (1.12) then gives us the scales explicitly as (1581,500,158). We use the equally weighted sample sizes $n_1 = 30, 50, 100, 500, 1000$ and 5000. The summary statistics are presented in Table 4.12 with the Q-Q plots being presented in Figures 4.43 to 4.48. Yet again, we fail to observe Normality in the distribution of Δ_s for any sample size considered. However, the distributions become more Normal as

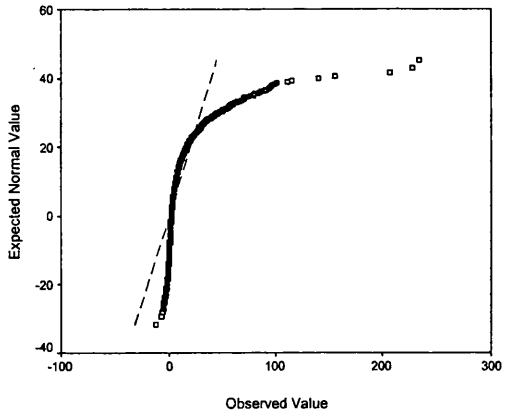

Figure 4.38: Distribution of Δ_s for example five: $n_1 = 37$

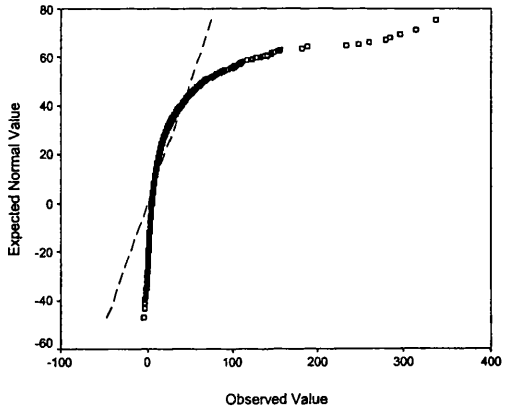

Figure 4.39: Distribution of Δ_s for example five: $n_1 = 74$

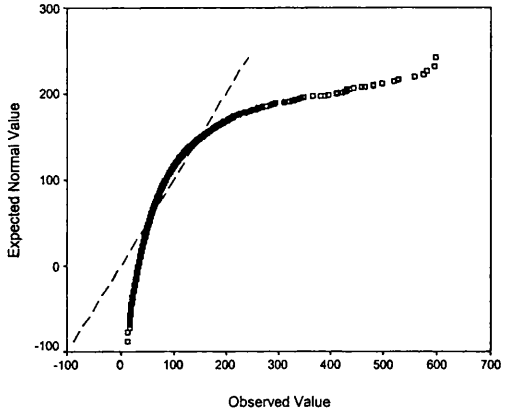

Figure 4.40: Distribution of Δ_s for example five: $n_1 = 370$

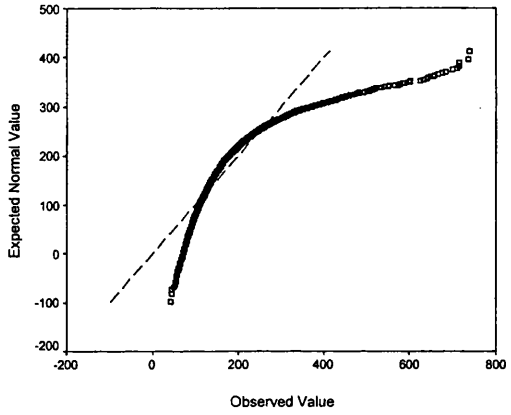

Figure 4.41: Distribution of Δ_s for example five: $n_1 = 740$

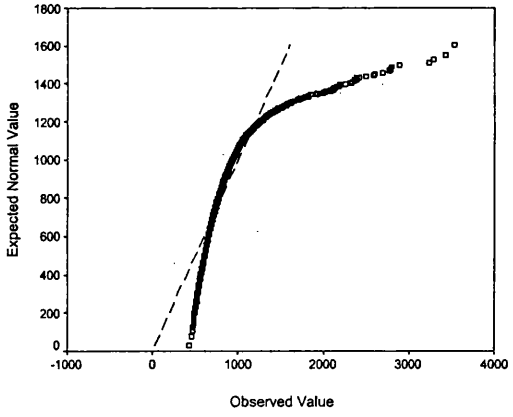

Figure 4.42: Distribution of Δ_s for example five: $n_1 = 3700$

	Summary Statistics for Δ_s				
n_1	**Mean**	**S.D.**	**Skewness**	**K-S Test**	**Percent > 0**
30	0.3692	0.7620	4.021	16.545	71.11
50	0.6935	1.0304	3.687	14.040	81.30
100	1.5656	1.6254	3.065	12.434	93.27
500	8.7408	4.3681	2.180	9.277	99.97
1000	17.5020	6.1779	1.782	8.192	100.00
5000	88.7497	14.4137	1.077	5.244	100.00

Table 4.12: Example 6 : Summary statistics for properties of Δ_s for varying sample sizes when data is from a Burr XII ALT distribution.

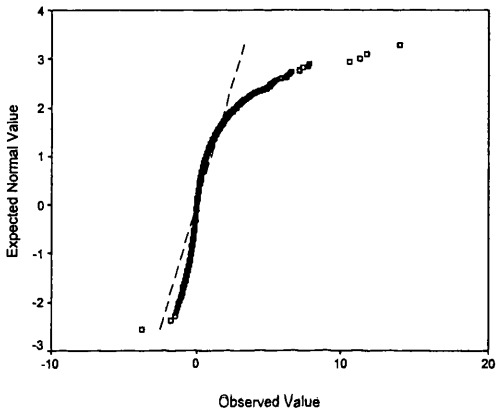

Figure 4.43: Distribution of Δ_s for example six: $n_1 = 30$

the sample size increases, while, in the same limit, the mean and standard deviation of Δ_s increases, along with the percentage of positive values of Δ_s observed. So, as the sample size increases, we are clearly more confident that the sign of Δ_s will be positive, but are unable to make any further inferential progress due to the consistent lack of Normality in the distribution of Δ_s.

4.3.3 Discussion on Results

In summary, our hope, when considering the behaviour of Δ_s, was that with increasing sample size, its distribution would be more Normal. As such, we would only need to find its expectation and variance in order to make inferences about the probability of Δ_s being positive or negative. However, we saw that, for practically every case considered, we failed to obtain Normality in the distribution of Δ_s, thereby inhibiting further inferences.

When the data was generated from a Weibull ALT distribution, we observed an approximately 50:50 split in the percentage of times we preferred to fit the Weibull distribution to the Burr XII, for larger sample sizes. It was also apparent that the standard deviation of Δ_s was considerably larger than its mean, for the whole range of sample sizes considered. Now, as the sample size increased, the approximate mean of Δ_s was -1. Naturally, as the

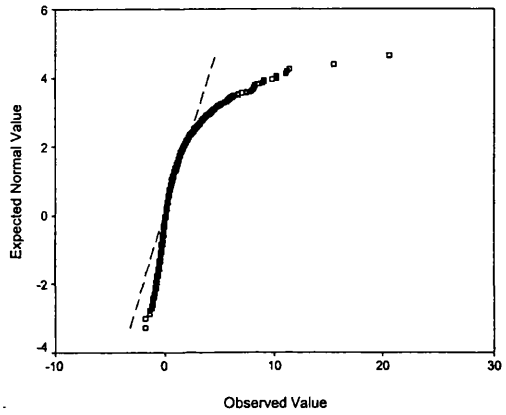

Figure 4.44: Distribution of Δ_s for example six: $n_1 = 50$

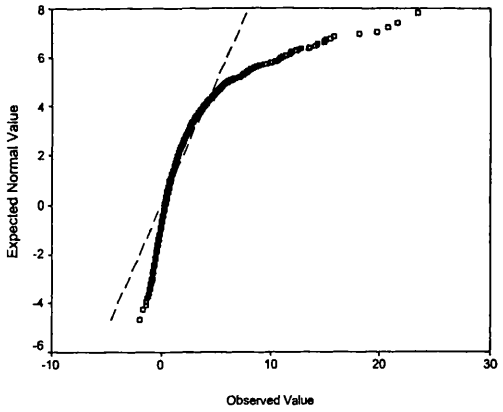

Figure 4.45: Distribution of Δ_s for example six: $n_1 = 100$

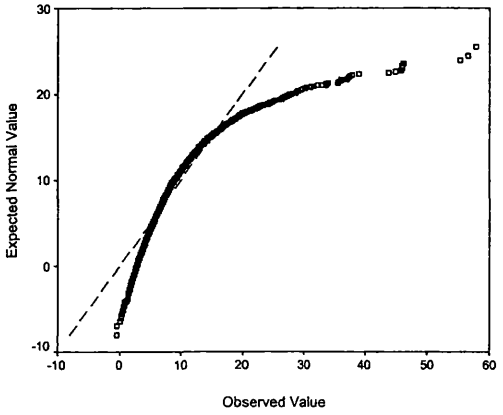

Figure 4.46: Distribution of Δ_s for example six: $n_1 = 500$

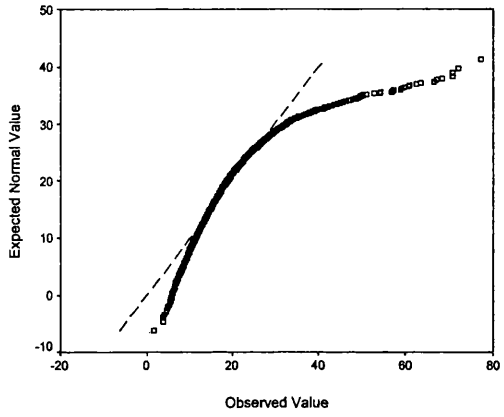

Figure 4.47: Distribution of Δ_s for example six: $n_1 = 1000$

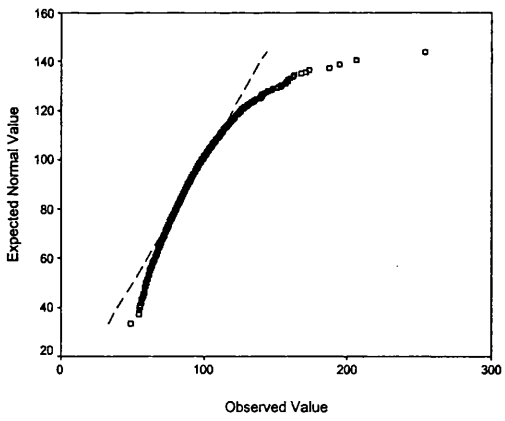

Figure 4.48: Distribution of Δ_s for example six: $n_1 = 5000$

sample size increases further we would expect to eventually obtain Normality in the distribution of Δ_s, with mean of around -1 and large standard deviation. Consequently, we would expect around half the observations to be greater than -1 and half to be less than -1, thereby re-enforcing the 50:50 split observed in the percentage of positive and negative values of Δ_s.

To aid in the interpretation of the results when data was generated from a Burr XII distribution, we use the *coefficient of variation*, given by

$$C_v = \frac{\text{standard deviation}}{\text{mean}} \times 100;$$

see, for example, Anderson, Sweeney and Williams (1996). Clearly, this indicates the spread in the data relative to the mean, and, when the data is generated from a Burr XII distribution, we can easily show that, for our examples, C_v decreases as the sample size increases, which, together with the fact that the mean of Δ_s increases in the same limit, suggests that we are more confident of obtaining a positive value of Δ_s for larger sample sizes. This is endorsed by the final column in the tables, where we saw the percentage of cases observed to be positive, increase with sample size.

The observations made above, with regard to the percentage of positive and negative values of Δ_s, are exemplified by taking 95% confidence intervals around the mean. We use *Chebyshev's Inequality* for the mean of Δ_s, defined by

$$\Pr\left(\left|\overline{\Delta}_s - \mu_{\Delta_s}\right| \leq k\frac{\sigma_{\Delta_s}}{\sqrt{n}}\right) \geq 1 - \frac{1}{k^2},$$

for $k > 1$; it states that at least $\left(1 - 1/k^2\right) \times 100\%$ of the values of $\overline{\Delta}_s$ lie within k standard deviations of the mean μ_{Δ_s}; see, for example, Anderson, Sweeney and Williams (1996). From this we can formulate the 95% confidence interval as

$$\overline{\Delta}_s \pm 4.47\frac{\sigma_{\Delta_s}}{\sqrt{n}}.$$

We now consider the three examples where data originated from a Weibull distribution, and take 95% confidence intervals around the mean of Δ_s for the largest sample size considered. They are $(-4.8855, 0.9023)$, $(-4.9385, 0.8323)$ and $(-2.6066, 3.0092)$ respectively. For the three examples where the data was generated from a Burr XII distribution, again for the largest sample size considered, the confidence intervals are $(4221.5418, 4376.7559)$, $(807.6045, 822.6564)$ and $(88.2236, 89.2757)$ respectively. Clearly, when the data comes from a Weibull distribution, we cannot rule out the possibility of obtaining a positive or negative value for the mean of Δ_s, while for data generated from a Burr XII distribution, the mean of Δ_s has a positive upper and lower confidence limit, in keeping with the observations we made above.

It seems, then, that there is limited practicality in finding the theoretical expectation

and variance of Δ_s, whether the data was generated from a Weibull or Burr XII distribution. If the data came from a Weibull distribution then, for a sufficiently large sample size and a given set of parameter values, we have a 50:50 chance of preferring to fit the Weibull to the Burr XII. Conversely if the data came from a Burr XII then, for a sufficiently large sample size and set of parameter values, we preferred to fit the Burr XII to the Weibull 100% of the time. That is not to say that Δ_s has limited practical use; as we remarked earlier, for a specific data set we can say with certainty whether the Burr XII will provide a superior fit to the Weibull, simply by fitting the more tractable Weibull distribution and then calculating Δ_s. The problem lies with trying to draw general conclusions about our preference of models, based simply on a particular set of parameter values. In light of this result, this is an appropriate point to end our discussion on the behaviour of Δ_s; further theoretical investigations being considered beyond the scope of the current discussion.

4.4 Summary

In this chapter we began by considering an extension to the non-accelerated discriminating factor, by deriving Δ_s for an accelerated framework. In order to obtain this result, we looked at two reparameterisations of the Burr XII ALT distribution: the first reparameterised the shape parameter a, while the second considered a reparameterisation of α_b. We showed that both reparameterisations led to the same result. Some real life examples were then investigated. Finally, we sought theoretical results on the behaviour of Δ_s, hoping to decide which of the two models would provide a better fit for a given set of parameter values. Through simulations, we concluded that the lack of Normality in the distribution of Δ_s, failed to facilitate further theoretical advances; certainly to an extent deemed appropriate in the current context. Of course, it may be possible to find some alternative model for the distribution of Δ , although the extended nature of the investigation seemingly required - particularly in light of that needed to assess whether Δ follows a Normal distribution - renders this inquiry beyond the scope of the current discussion.

Hitherto, we have examined completely failed data sets only. In Chapter two we presented the algorithm for fitting a Burr XII ALT model to complete data and proceeded to consider three published data sets, for which we presented summaries of parameter values and log-likelihoods. In Chapter three, we established the expected Fisher information matrices for the Burr XII ALT model and referenced the corresponding matrices for other models under consideration. This would allow us to make large sample approximations to the standard deviations of parameter estimates in the models. Finally, in this chapter, we determined an expression for accelerated data, that would enable us to determine which of the Burr XII or Weibull models would provide the superior fit to given data sets, without having to trouble ourselves with actually fitting the Burr XII model. In the next chapter, we make the necessary extensions to Type I censored data: we seek to develop the expected Fisher information matrix for the Burr XII ALT model, extend Δ for Type I censoring, and

give illustrative examples throughout. We also show that theoretical results agree with the complete results of Chapter three when the stopping time tends to infinity.

Chapter 5

Censoring

This chapter is concerned with the extension of previous results for complete data to the censored case. Our discussion shall focus on Type I censoring, although we briefly discuss the possible practical approaches to Type II censoring. We begin by outlining the principles involved in Type I censoring and proceed to extend our notation to accomodate accelerated lifetime models. A published data set shall be considered, which shall be examined later in this chapter. We develop theory - from the construction of the model and EFI matrix - for the Burr XII model, beginning with the two and three parameter, non-ALT models, before making the natural extension to acceleration. During the course of deriving the terms that form the EFI matrices, we assess theoretical agreement with simulated data. For completeness and comparison of results we then present the theory for the Weibull ALT model. Our analysis then moves on to consider some practical applications of the theory; we look at the effect upon the standard deviations of parameter values for changing values of c Finally, we present a detailed investigation into the asymptotic agreement between Type I censored results and complete results, as $c \to \infty$.

Chapter one briefly outlines the principles of Type I censoring in a non-accelerated environment. There, the failure times were the n items $D_1, \cdots, D_n \leq c$ and the $m = N - n$ censored values $D_{n+1} = \cdots = D_N = c$. Now, the extension to accomodate acceleration will require us to use a double subscripted notation, D_{ij}, to indicate the j^{th} failure at stress level x_i $(1 \leq i \leq k)$; while we now introduce the notation C_{ij} to represent the j^{th} censored item at the same stress level. Consequently, we now have n_i times to failure and $m_i = N_i - n_i$ censored items at stress level x_i.

5.1 Type I Censoring in Practice and Theory

To illustrate the data we are now dealing with, we consider the data set in Table 5.1, taken from Nelson (1990, p158). The data comprises the times to failure of electric motors with a certain class of insulation. Ten motors were tested at each of four temperatures: 150,170,190 and 220 degrees Celsius. This example clearly shows the need for both censoring

x_i	150	170	190	220
N_i	10	10	10	10
n_i	0	8	9	10
d_{ij}		1764,2772, 3444,3542, 3780,4860, 5196,6206	408,408,1344, 1344,1440,1920, 2256,2352,2596	408,408,504,504, 504,600,600,648, 648,696
m_i	10	2	1	0
c_{ij}	9429,9429,9429, 9429,9429,9429, 9429,9429,9429, 9429	6792,6792	3120	

Table 5.1: Time to failure of electric motors with Class-B insulation; from Nelson (1990)

and acceleration: if we were to choose one stress level of 150 degrees Celsius and wait for all items to fail, we would be waiting in excess of 9000 hours. The additional stress levels have presented us with some earlier failure times and more information for our analysis, whereas censoring has ensured that we do not have to wait until all items have failed at lower stress levels before we can begin the analysis. Typically, an ALT experiment with Type I censoring will require us to choose the stress levels and the allocation of sample items to those stress levels, although the manner in which we stop the experiment is optional. We could choose to set some items to run at the lowest stress, and then at predetermined times in the future set a number of items to run at increased stress levels. We would then stop the experiment at some future point in time; this is what appears to have taken place in this example; see Table 5.1. Alternatively, we could set all items at all stress levels to run at the same time and then treat each stress level separately, giving us k stopping times. Another option would be to set all items, at all stress levels, to run at the same time, wait a predetermined time c and then stop the experiment and collect our results. So, we already have three ways of carrying out Type I censoring in practice. In terms of modelling the data, we can treat each resulting data set in the same way and proceed to run the algorithm and fit the distribution of interest. However, from a theoretical angle, we must be particular about which approach we are adopting, since each method carries its own assumptions with regard to starting and stopping times, which, in turn, have an important bearing on how we determine expectations of results. The final approach to Type I censoring discussed, is perhaps both practically and algebraically the simplest of the options mentioned, and shall be the manner in which we consider a Type I censoring regime.

5.2 Type II Censoring in Practice

It is clear that for any non-accelerated distribution, the practical application of Type II censoring is straightforward. Out of our total sample size N, we terminate the experiment

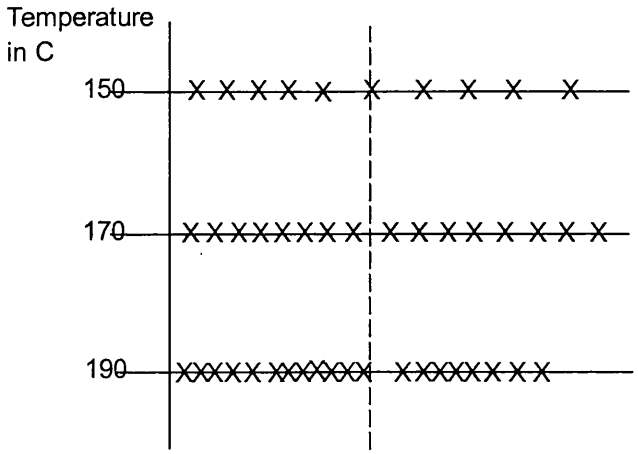

Figure 5.1: Visualisation of an accelerated life test, with failures represented by $\times$. The experiment is stopped once the 6^{th} item fails at the lowest stress level.

once the n^{th} item has failed; the stopping time is then the random variable. Thus, we have a known number of failures n and a known number of censored items $N - n$. However, the extension to an accelerated framework is not immediately apparent. For instance, consider an experiment with, say, $k = 3$ stress levels. We could stop the experiment once a certain number of items have failed at the lowest stress level, that is, until n_1 have failed out of N_1 - this is illustrated in Figure 5.1 - or wait until a certain number of items have failed at the highest stress level, that is, until n_3 have failed out of N_3 - as illustrated in Figure 5.2. Furthermore, we could even choose to stop the experiment once a certain number of items have failed out of $N_1 + N_2 + N_3$, irrespective of which stress level the failures came from; see Figure 5.3. For each approach outlined, we would need to determine the expected number of failures at every stress level and the expectations of various related summations; for the third approach, this is rather complicated: we would need to determine the expected value of the n^{th} order statistic when, in turn, each of the previous $n - 1$ failures could have occurred at any of the three stress levels. However, the other two approaches are perhaps, more likely to be adopted in practice. We see from this brief discussion that consideration of Type II censoring would require us to be very specific about the way in which we set up the experiment, and this, in turn, affects the way in which we consider theoretical expectations. Clearly, this requires a very detailed investigation, to be considered elsewhere.

5.3 Burr XII Two-Parameter Model

Since many of the expectations required for the EFI matrix for the Burr XII ALT model, with Type I censoring, are based on expectations for the non-ALT three-parameter model, which is a subsequent extension of the two-parameter model, we begin by considering theory for the Burr XII two-parameter model. So, using (1.3) and (1.4), the likelihood for this

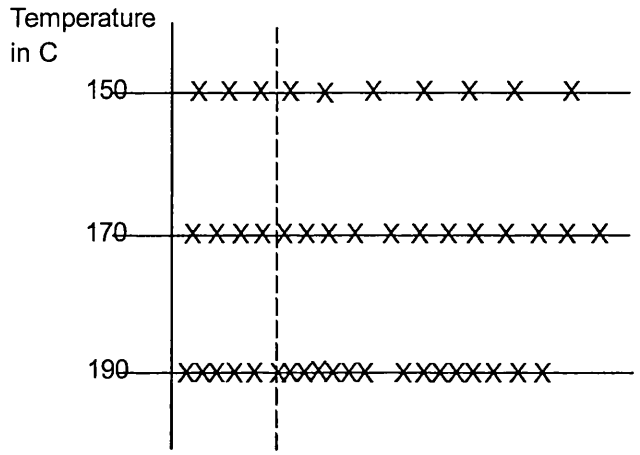

Figure 5.2: The same experiment, now stopped once the 6^{th} item fails at the highest stress level.

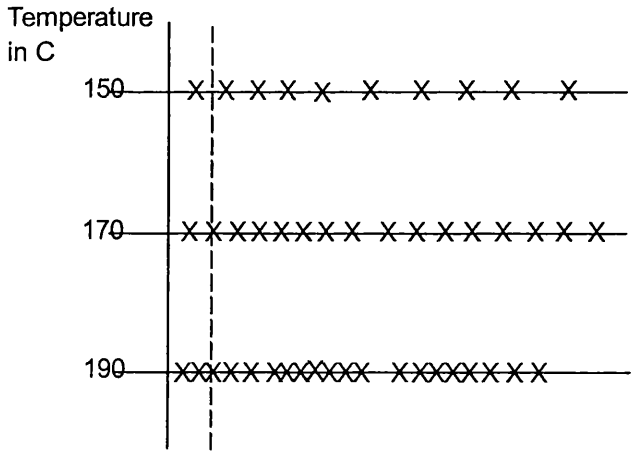

Figure 5.3: The same experiment, now stopped once the 6^{th} item out of all items on test fails.

particular model, for failed data $d_1, \cdots d_n$ and censored data $c_1, \cdots c_m$, is given by

$$\left[\prod_{i=1}^{n} \frac{a\tau d_i^{\tau-1}}{[1+d_i^{\tau}]^{a+1}}\right]\left[\prod_{i=1}^{m}(1+c_i^{\tau})^{-a}\right] = \left[\frac{(a\tau)^n \prod_{i=1}^{n} d_i^{\tau-1}}{\prod_{i=1}^{n}[1+d_i^{\tau}]^{a+1}}\right](1+c^{\tau})^{-(N-n)a},$$

so that the log-likelihood is

$$l(a,\tau) = n\ln(a\tau) + (\tau-1)\, s_e - (a+1)\, t_*^d - a t_*^c \tag{5.1}$$

where we define

$$\begin{aligned} s_e &= \sum_{i=1}^{n} \ln d_i, \\ t_*^d &= \sum_{i=1}^{n} \ln(1+d_i^{\tau}), \end{aligned}$$

and

$$t_*^c = \sum_{i=n+1}^{N} \ln(1+d_i^{\tau}) = (N-n)\ln(1+c^{\tau}),$$

under Type I censoring. We also identify the probability an item fails in $(0,c)$; from (1.3), this is

$$1 - \frac{1}{(1+c^{\tau})^a} = q_{c,a},$$

say. Thus, the two first derivatives of (5.1) are

$$\frac{\partial l}{\partial a} = na^{-1} - \left(t_*^d + t_*^c\right) \tag{5.2}$$

and

$$\frac{\partial l}{\partial \tau} = n\tau^{-1} + s_e - (a+1)\, t_{111}^d - a t_{111}^c \tag{5.3}$$

where

$$\begin{aligned} t_{111}^d &= \sum_{i=1}^{n} \frac{d_i^{\tau} \ln d_i}{1+d_i^{\tau}}, \\ t_{111}^c &= \sum_{i=n+1}^{N} \frac{d_i^{\tau} \ln d_i}{1+d_i^{\tau}} = \frac{(N-n)\, c^{\tau} \ln c}{1+c^{\tau}}; \end{aligned}$$

while the second derivatives are

$$\frac{\partial^2 l}{\partial a^2} = -na^{-2}, \tag{5.4}$$

$$\frac{\partial^2 l}{\partial \tau^2} = -n\tau^{-2} - (a+1)\, t_{122}^d - a t_{122}^c \tag{5.5}$$

and

$$\frac{\partial^2 l}{\partial a \partial \tau} = -\left(t^d_{111} + t^c_{111}\right), \tag{5.6}$$

where we now define

$$\begin{aligned} t^d_{122} &= \sum_{i=1}^{n} \frac{d_i^\tau (\ln d_i)^2}{\{1+d_i^\tau\}^2} \\ t^c_{122} &= \sum_{i=n+1}^{N} \frac{d_i^\tau (\ln d_i)^2}{\{1+d_i^\tau\}^2} = \frac{(N-n)\, c^\tau (\ln c)^2}{(1+c^\tau)^2}. \end{aligned}$$

5.3.1 Expectations In Derivatives

Firstly, since the number of failures n is a Binomial random variable, with parameters $N, q_{c,a}$, we have

$$\begin{aligned} E[n] &= N q_{c,a} \\ E[N-n] &= N(1-q_{c,a}). \end{aligned}$$

To consider the derivatives above, we will need the expectations of terms of the form $\ln(1+Y^\tau), \ln Y$,

$$\frac{Y^\tau \ln Y}{1+Y^\tau}$$

and

$$\frac{Y^\tau (\ln Y)^2}{(1+Y^\tau)^2},$$

for Y following the truncated Burr XII distribution

$$\frac{a\tau y^{\tau-1}(1+y^\tau)^{-(a+1)}}{q_{c,a}},$$

for $0 \le c \le y$. Thus, the expectation of $\ln(1+Y^\tau)$ is

$$E[\ln(1+Y^\tau)] = \int_0^c \ln(1+y^\tau) \times \frac{a\tau y^{\tau-1}}{q_{c,a}(1+y^\tau)^{a+1}} dy,$$

where we now let $u = a\ln(1+y^\tau)$, so that $0 \le y \le c \Leftrightarrow 0 \le u \le a\ln(1+c^\tau)$, and

$$du = \frac{a\tau y^{\tau-1}}{1+y^\tau} dy,$$

from which

$$E[\ln(1+Y^\tau)] = \frac{1}{a q_{c,a}} \int_0^{a\ln(1+c^\tau)} u \exp(-u)\, du = \frac{\gamma\{2, a\ln(1+c^\tau)\}}{a q_{c,a}},$$

Various equivalent forms of this result exist; for instance, we have

$$\begin{aligned} E\left[\ln\left(1+Y^{\tau}\right)\right] &= \frac{\left[1-\{1+a\ln(1+c^{\tau})\}(1+c^{\tau})^{-a}\right]}{aq_{c,a}} \\ &= \frac{1}{a}-\frac{\ln(1+c^{\tau})}{q_{c,a}(1+c^{\tau})^{a}} \\ &= \frac{1}{a}-\frac{\ln(1+c^{\tau})}{(1+c^{\tau})^{a}-1}, \end{aligned} \tag{5.7}$$

We now consider

$$E\left[Y^{r}\right]=\frac{a}{q_{c,a}}\int_{0}^{c}\frac{\tau y^{r+\tau-1}}{\left(1+y^{\tau}\right)^{a+1}}dy,$$

and define

$$v=\frac{y^{\tau}}{1+y^{\tau}}$$

so that $0\leq y\leq c\Leftrightarrow 0\leq v\leq \frac{c^{\tau}}{1+c^{\tau}}=c'$, say, and

$$dv=\frac{\tau y^{\tau-1}}{\left(1+y^{\tau}\right)^{2}}dy$$

and hence obtain

$$E\left[Y^{r}\right]=\frac{a}{q_{c,a}}\int_{0}^{c'}v^{\frac{r}{\tau}}\left(1-v\right)^{a-\frac{r}{\tau}-1}dv=\frac{aB_{c'}\left(r_{1},a-r_{0}\right)}{q_{c,a}}, \tag{5.8}$$

defining

$$r_{i}=\frac{r}{\tau}+i.$$

Two of the required expectations may now be found by differentiating (5.8) with respect to r, and then taking $r=0,\tau$. Again, various equivalent forms of (5.8) may be found, chiefly through linking the incomplete Beta function to hypergeometric functions; for instance, from paragraphs [6.6.8] and [15.3.4] in Abramowitz and Stegun (1972), we have

$$B_{c'}\left(r_{1},a-r_{0}\right)=\frac{\left(c'\right)^{r_{1}}}{r_{1}}F_{2,1}\left(r_{1},r_{1}-a;r_{2};c'\right)=\frac{c^{r+\tau}}{r_{1}}F_{2,1}\left(r_{1},a+1;r_{2};-\omega\right), \tag{5.9}$$

where $\omega=c^{\tau}$ and the negative argument arises from an integral representation of the hypergeometric function on the complex plane cut along the real axis from 1 to ∞. The final form is useful here, since substituting (5.9) in (5.8) allows us to write

$$E\left[Y^{r}\right]=\frac{a\omega}{q_{c,a}}\times\frac{c^{r}}{r_{1}}F_{2,1}\left(r_{1},a+1;r_{2};-c^{\tau}\right)=\frac{a\omega}{q_{c,a}}\times\lambda_{0,1}f_{1}\left(r_{1},a+1,-\omega\right), \tag{5.10}$$

where

$$\lambda_{0,1}=\frac{c^{r}}{r_{1}}$$

and we introduce the notation

$$f_{q}\left(a,b,z\right)=F_{q+1,q}\left(\{a,...,a,b\};\{a+1,...,a+1\};z\right)$$

to represent the generalised hypergeometric function with repeated and related arguments. Then, differentiating (5.10) with respect to r yields

$$E\left[Y^{\tau}\ln Y\right]=\frac{a\omega}{\tau q_{c,a}}\times\left[\lambda_{1,1}f_1\left(r_1,a+1,-\omega\right)+\lambda_{1,2}f_2\left(r_2,a+2,-\omega\right)\right], \tag{5.11}$$

where

$$\lambda_{1,1}=\frac{c^r\left(r_1\tau\ln c-\tau^{-1}\right)}{r_1^2}$$

and

$$\lambda_{1,2}=\frac{-\left(a+1\right)c^{r+\tau}}{r_1r_2^2};$$

on differentiating (5.10) a second time with respect to r we also have

$$E\left[Y^{\tau}\left(\ln Y\right)^2\right]=\frac{ac^{\tau}}{\tau^2q_{c,a}}\times\left[\begin{array}{c}\lambda_{2,1}f_1\left(r_1,a+1,-\omega\right)\\+\lambda_{2,2}f_2\left(r_2,a+2,-\omega\right)\\+\lambda_{2,3}f_3\left(r_3,a+3,-\omega\right)\end{array}\right] \tag{5.12}$$

where

$$\lambda_{2,1}=\frac{c^r}{r_1^3}\left\{1+\left(\tau r_1\ln c-1\right)^2\right\},$$

$$\lambda_{2,2}=\frac{-2\left(a+1\right)c^{r+\tau}\left\{r_2\left(\tau r_1\ln c-1\right)-r_1\right\}}{r_1^2r_2^3}$$

and

$$\lambda_{2,3}=\frac{2\left(a+1\right)\left(a+2\right)c^{r+2\tau}}{r_1r_2r_3^3}.$$

From (5.11), we have (on letting $r\rightarrow 0$, so $r_i\rightarrow i$),

$$E\left[\ln Y\right]=\tfrac{a\omega}{\tau q_{c,a}}\times\left[\left(\ln\omega-1\right)f_1\left(1,a+1,-\omega\right)-\tfrac{(a+1)\omega f_2(2,a+2,-\omega)}{4}\right] \tag{5.13}$$

while setting $r=\tau$ (so $r_i=i+1$), we obtain

$$E\left[Y^{\tau}\ln Y\right]=\frac{a\omega^2}{2\tau q_{c,a}}\times\left[\frac{\left(2\ln\omega-1\right)f_1\left(2,a+1,-\omega\right)}{2}-\frac{\left(a+1\right)\omega f_2\left(3,a+2,-\omega\right)}{9}\right]; \tag{5.14}$$

from (5.12), we have (again letting $r\rightarrow 0$)

$$E\left[\left(\ln Y\right)^2\right]=\frac{a\omega}{\tau^2q_{c,a}}\times\left[\begin{array}{c}\left\{1+\left(\ln\omega-1\right)^2\right\}f_1\left(1,a+1,-\omega\right)\\-\frac{1}{4}\left(a+1\right)\omega\left(2\ln\omega-3\right)f_2\left(2,a+2,-\omega\right)\\+\frac{1}{27}\left(a+1\right)\left(a+2\right)\omega^2f_3\left(3,a+3,-\omega\right)\end{array}\right],$$

while setting $r = \tau$ (so $r_i = i + 1$), we obtain

$$E\left[Y^{\tau}(\ln Y)^2\right] = \frac{a\omega^2}{\tau^2 q_{c,a}} \times \begin{bmatrix} \frac{1}{8}\left\{1 + (2\ln\omega - 1)^2\right\} f_1(2, a+1, -\omega) \\ -\frac{1}{54}(a+1)\,\omega\,(6\ln\omega - 5)\, f_2(3, a+2, -\omega) \\ +\frac{1}{192}(a+1)(a+2)\,\omega^2 f_3(4, a+3, -\omega) \end{bmatrix}. \tag{5.15}$$

5.3.2 Derived Expectations

We can use the above expectations to derive further expectations. In particular, we now consider

$$E\left[\frac{Y^{\tau}\ln Y}{1+Y^{\tau}}\right] \equiv E_a\left[\frac{Y^{\tau}\ln Y}{1+Y^{\tau}}\right], \tag{5.16}$$

To ensure that censoring is fully taken into account here, we write this expectation as an integral involving $q_{c,a}$. As such,

$$E_a\left[\frac{Y^{\tau}\ln Y}{1+Y^{\tau}}\right] = \frac{a q_{c,a+1}}{(a+1)\,q_{c,a}} E_{a+1}\left[Y^{\tau}\ln Y\right]; \tag{5.17}$$

or, equivalently,

$$E_a\left[\frac{Y^{\tau}\ln Y}{1+Y^{\tau}}\right] = E_a\left[\ln Y\right] - \frac{a q_{c,a+1}}{(a+1)\,q_{c,a}} E_{a+1}\left[\ln Y\right]. \tag{5.18}$$

Using (5.14) in (5.17), we see that the expectation (5.16) is

$$E\left[\frac{Y^{\tau}\ln Y}{1+Y^{\tau}}\right] = \frac{a\omega^2}{2\tau q_{c,a}}\left[\frac{(2\ln\omega - 1)}{2} f_1(2, a+2, -\omega) - \frac{(a+2)\omega}{9} f_2(3, a+3, -\omega)\right]; \tag{5.19}$$

the second recurrence relation (5.18) produces

$$E\left[\frac{Y^{\tau}\ln Y}{1+Y^{\tau}}\right] = \frac{a\omega}{\tau q_{c,a}} \begin{bmatrix} (\ln\omega - 1)\left\{f_1(1, a+1, -\omega) - f_1(1, a+2, -\omega)\right\} \\ -\frac{\omega}{4}\left\{(a+1) f_2(2, a+2, -\omega) - (a+2) f_2(2, a+3, -\omega)\right\} \end{bmatrix}.$$

Similarly, we have

$$E\left[\frac{Y^{\tau}(\ln Y)^2}{(1+Y^{\tau})^2}\right] \equiv E_a\left[\frac{Y^{\tau}(\ln Y)^2}{(1+Y^{\tau})^2}\right] = \frac{a q_{c,a+2}}{(a+2)\,q_{c,a}} E_{a+2}\left[Y^{\tau}(\ln Y)^2\right];$$

this expectation can also be expressed as

$$\frac{a q_{c,a+1}}{(a+1)\,q_{c,a}} E_{a+1}\left[(\ln Y)^2\right] - \frac{a q_{c,a+2}}{(a+2)\,q_{c,a}} E_{a+2}\left[(\ln Y)^2\right].$$

Using (5.15), we have

$$E\left[\frac{Y^{\tau}(\ln Y)^2}{(1+Y^{\tau})^2}\right] = \frac{a\omega^2}{\tau^2 q_{c,a}} \times \begin{bmatrix} \frac{1}{8}\left\{1+(2\ln\omega-1)^2\right\} f_1(2,a+3,-\omega) \\ -\frac{1}{54}(a+3)\,\omega\,(6\ln\omega-5)\, f_2(3,a+4,-\omega) \\ +\frac{1}{192}(a+3)(a+4)\,\omega^2 f_3(4,a+5,-\omega) \end{bmatrix} \tag{5.20}$$

So we now have all the expectations we require to enable us to establish expectations in first and second derivatives of the log-likelihood for a Burr XII 2-parameter model.

Simplifications

With the few expectations already considered, we have seen that we require an array of generalised hypergeometric functions, each taking a variety of arguments. As an aid to later work, where we consider limiting properties of hypergeometric functions with c tending to infinity, we now present a series of simplifications that will allow us to write down all expectations using just a few hypergeometric functions. We start with two useful recurrence relationships between 'neighbouring' hypergeometric functions; see Watkins and Johnson (2002) for proofs, which are also detailed in Appendix B.

$$f_q(a,b,z) - f_{q-1}(a,b,z) = -\frac{za^{q-1}b}{(a+1)^q} f_q(a+1,b+1,z) \tag{5.21}$$

$$f_q(a,b+1,z) - f_q(a,b,z) = \frac{za^q}{(a+1)^q} f_q(a+1,b+1,z) \tag{5.22}$$

Using (5.21) and (5.22) we are able to express our expectations in terms of

$$\begin{aligned} \kappa_0 &= (1+\omega)^{-a-1} = 1 - q_{c,a+1} = p_{c,a+1}, \\ \kappa_1 &= f_1(1,a+1,-\omega) = \frac{q_{c,a}}{a\omega}, \end{aligned} \tag{5.23}$$

and

$$\kappa_2 = f_2(1,a+1,-\omega). \tag{5.24}$$

We shall also find it useful to write

$$\kappa_3 = f_3(1,a+1,-\omega). \tag{5.25}$$

We note that

$$f_1(2,a+2,-\omega) = \frac{2}{(a+1)\,\omega}\left[\kappa_1 - \kappa_0\right], \tag{5.26}$$

while

$$f_2(2,a+2,-\omega) = \frac{4(\kappa_2-\kappa_1)}{(a+1)\,\omega}, \tag{5.27}$$

and

$$f_2(3,a+3,-\omega) = \frac{9\left[2\kappa_2 - 3\kappa_1 + \kappa_0\right]}{(a+1)(a+2)\,\omega^2} \tag{5.28}$$

These simplifications and recurrence relations then yield

$$E[\ln Y] = \frac{a\omega}{\tau q_{c,a}}\left[(\ln\omega - 1)\frac{q_{c,a}}{a\omega} - [\kappa_2 - \kappa_1]\right] = \frac{1}{\tau}\left[\ln\omega - \frac{\kappa_2}{\kappa_1}\right]. \tag{5.29}$$

We also have

$$f_1(1, a+2, -\omega) - f_1(1, a+1, -\omega) = -\frac{\omega}{2} f_1(2, a+2, -\omega) = -\frac{(\kappa_1 - \kappa_0)}{(a+1)}$$

from (5.22), so that

$$f_1(1, a+2, -\omega) = \kappa_1 - \frac{1}{(a+1)}[\kappa_1 - \kappa_0] = \frac{a\kappa_1 + \kappa_0}{a+1}$$

and

$$\kappa_1 - f_1(1, a+2, -\omega) = \frac{(\kappa_1 - \kappa_0)}{(a+1)}.$$

Furthermore,

$$\begin{aligned} f_2(2, a+3, -\omega) - f_2(2, a+2, -\omega) &= -\frac{4\omega}{9} f_2(3, a+3, -\omega) \\ &= -\frac{4\omega}{9} \times \frac{9[2\kappa_2 - 3\kappa_1 + \kappa_0]}{(a+1)(a+2)\omega^2} \\ &= -\frac{4[2\kappa_2 - 3\kappa_1 + \kappa_0]}{(a+1)(a+2)\omega} \end{aligned}$$

so that

$$f_2(2, a+3, -\omega) = \frac{4[a(\kappa_2 - \kappa_1) + (\kappa_1 - \kappa_0)]}{(a+1)(a+2)\omega}.$$

We now consider the two expressions for (5.16). The first becomes

$$\frac{a\omega}{\tau(a+1)q_{c,a}}[(\kappa_1 - \kappa_0)(\ln\omega) - (\kappa_2 - \kappa_1)],$$

while the second, on simplification, also reduces to

$$\frac{a\omega}{\tau(a+1)q_{c,a}}[(\kappa_1 - \kappa_0)(\ln\omega) - (\kappa_2 - \kappa_1)].$$

5.3.3 Expectations of the Score

We are now in a position to check that the expectations of the score functions are in fact zero. For (5.2), we have

$$\begin{aligned} E\left[t_*^d\right] &= E[n] \times E[\ln(1 + D^\tau)] = Nq_{c,a} \times \left[\frac{1}{a} - \frac{\ln(1+\omega)}{q_{c,a}(1+\omega)^a}\right] \\ &= \frac{Nq_{c,a}}{a} - \frac{N\ln(1+\omega)}{(1+\omega)^a} \end{aligned}$$

while

$$E\left[t_*^c\right] = E\left[N-n\right] \times \ln\left(1+\omega\right) = \frac{N\ln\left(1+\omega\right)}{\left(1+\omega\right)^a}$$

so that

$$E\left[t_*^d\right] + E\left[t_*^c\right] = \frac{Nq_{c,a}}{a}.$$

Thus, taking expectations in (5.2), we have

$$E\left[\frac{\partial l}{\partial a}\right] = a^{-1}E\left[n\right] - E\left[t_*^d + t_*^c\right] = \frac{Nq_{c,a}}{a} - \frac{Nq_{c,a}}{a} = 0,$$

as required. For (5.3), we have, from (5.13)

$$E\left[s_e\right] = E\left[n\right] \times E\left[\ln D\right] = \frac{Na\omega}{\tau}\left[\left(\ln\omega - 1\right) f_1\left(1, a+1, -\omega\right) - \frac{\left(a+1\right)\omega f_2\left(2, a+2, -\omega\right)}{4}\right],$$

so that

$$\tau^{-1}E\left[n\right] + E\left[s_e\right] = \frac{Nac^\tau}{\tau}\left[\left(\ln\omega\right) f_1\left(1, a+1, -\omega\right) - \frac{\left(a+1\right)\omega f_2\left(2, a+2, -\omega\right)}{4}\right];$$

we also need

$$E\left[t_{111}^d\right] = E\left[n\right] \times E\left[\frac{D^\tau \ln D}{1+D^\tau}\right] = \frac{Na\omega^2}{2\tau}\left[\begin{array}{c} \frac{\left(2\ln\omega - 1\right)}{2} f_1\left(2, a+2, -\omega\right) \\ -\frac{\left(a+2\right)\omega}{9} f_2\left(3, a+3, -\omega\right) \end{array}\right]$$

and

$$E\left[t_{111}^c\right] = E\left[N-n\right] \times \frac{\omega\ln c}{1+\omega} = \frac{N\omega\ln c}{\left(1+\omega\right)^{a+1}}.$$

Thus, from (5.3), we have

$$\begin{aligned} E\left[\frac{\partial l}{\partial \tau}\right] &= \frac{Na\omega}{\tau}\left[\left(\ln\omega\right) f_1\left(1, a+1, -\omega\right) - \frac{\left(a+1\right)\omega f_2\left(2, a+2, -\omega\right)}{4}\right] \\ &\quad - \frac{Na\left(a+1\right)\omega^2}{\tau}\left[\begin{array}{c} \frac{\left(2\ln\omega - 1\right)}{4} f_1\left(2, a+2, -\omega\right) \\ -\frac{\left(a+2\right)\omega}{18} f_2\left(3, a+3, -\omega\right) \end{array}\right] - \frac{Na\omega\ln c}{\left(1+\omega\right)^{a+1}} \end{aligned}$$

which becomes

$$\begin{aligned} E\left[\frac{\partial l}{\partial \tau}\right] &= \frac{Na\left(a+1\right)c^{2\tau}}{4\tau}\left[\begin{array}{c} f_1\left(2, a+2, -\omega\right) - f_2\left(2, a+2, -\omega\right) \\ +\frac{2\left(a+2\right)c^\tau}{9} f_2\left(3, a+3, -\omega\right) \end{array}\right] \\ &\quad + Nac^\tau\left(\ln c\right)\left[\begin{array}{c} f_1\left(1, a+1, -\omega\right) \\ -\frac{\left(a+1\right)c^\tau}{2} f_1\left(2, a+2, -\omega\right) - \frac{1}{\left(1+c^\tau\right)^{a+1}} \end{array}\right] \end{aligned} \tag{5.30}$$

We now show that both terms in (5.30) reduce to zero. For the first bracket, we use (5.27) and (5.28) to write

$$\frac{2(a+2)\omega}{9} f_2(3, a+3, -\omega) - f_2(2, a+2, -\omega) = \frac{2\left[a\omega(1+\omega)^{-a-1} - q_{c,a}\right]}{a(a+1)\omega^2};$$

given the form of $f_1(2, a+2, -\omega)$ at (5.26), we see that the first bracket reduces to zero, as we require. For the second bracket, we now use (5.23) and (5.26) to see that

$$f_1(1, a+1, -\omega) - \frac{(a+1)\omega}{2} f_1(2, a+2, -\omega) = \frac{1}{(1+\omega)^{a+1}},$$

so that the second bracket reduces to zero, as we require; this completes our consideration of $E\left[\frac{\partial l}{\partial \tau}\right]$.

5.3.4 Expectations of Second Derivatives

We can now write the expectations of the second derivatives, (5.4), (5.5) and (5.6) - using the results (5.19) and (5.20) - as

$$\begin{aligned}
E\left[\frac{\partial^2 l}{\partial a^2}\right] &= -N q_{c,a} a^{-2}, \\
E\left[\frac{\partial^2 l}{\partial \tau^2}\right] &= -N q_{c,a} \tau^{-2} - \frac{aN(1-q_{c,a})\, c^\tau (\ln c)^2}{(1+c^\tau)^2}, \\
&\quad -\frac{a(a+1)\omega^2}{\tau^2 q_{c,a}} \times \begin{bmatrix} \frac{1}{8}\left\{1 + (2\ln\omega - 1)^2\right\} f_1(2, a+3, -\omega) \\ -\frac{1}{54}(a+3)\omega(6\ln\omega - 5) f_2(3, a+4, -\omega) \\ +\frac{1}{192}(a+3)(a+4)\omega^2 f_3(4, a+5, -\omega) \end{bmatrix},
\end{aligned}$$

and

$$\begin{aligned}
E\left[\frac{\partial^2 l}{\partial a \partial \tau}\right] &= -\frac{a\omega^2}{2\tau q_{c,a}}\left[\frac{(2\ln\omega - 1)}{2} f_1(2, a+2, -\omega) - \frac{(a+2)\omega}{9} f_2(3, a+3, -\omega)\right] \\
&\quad -\frac{N(1-q_{c,a})\, c^\tau \ln c}{1+c^\tau}.
\end{aligned}$$

5.4 Burr XII Three-Parameter Model

We considered the introduction of the scaling parameter in Chapter one, with the CDF at (1.5). If we let $Y = \theta D$ and denote c/θ by C, then we can write down the modified log-likelihood and first and second derivatives as

$$\begin{aligned}
l_b(a, \tau, \theta) &= n\ln(a\tau) - n\tau\ln\theta + (\tau - 1)\sum_{i=1}^{n} \ln y_i - (a+1)\sum_{i=1}^{n} \ln\left\{1 + \left(\frac{y_i}{\theta}\right)^\tau\right\} \\
&\quad -a(N-n)\ln(1+C^\tau)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial l_b}{\partial a} &= na^{-1} - \sum_{i=1}^{n} \ln\{1+d_i^\tau\} - (N-n)\ln(1+C^\tau) \\
\frac{\partial l_b}{\partial \tau} &= n\tau^{-1} + \sum_{i=1}^{n} \ln d_i - (a+1)\sum_{i=1}^{n} \frac{d_i^\tau \ln d_i}{1+d_i^\tau} - a(N-n)\frac{C^\tau \ln C}{1+C^\tau} \\
\frac{\partial l_b}{\partial \theta} &= -n\tau\theta^{-1} + \tau(a+1)\theta^{-1}\sum_{i=1}^{n} \frac{d_i^\tau}{1+d_i^\tau} + a(N-n)\tau\theta^{-1}\frac{C^\tau}{1+C^\tau}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 l_b}{\partial a^2} &= -na^{-2} \\
\frac{\partial^2 l_b}{\partial \tau^2} &= -n\tau^{-2} - (a+1)\sum_{i=1}^{n} \frac{d_i^\tau \{\ln d_i\}^2}{\{1+d_i^\tau\}^2} - a(N-n)\frac{C^\tau \{\ln C\}^2}{\{1+C^\tau\}^2} \\
\frac{\partial^2 l_b}{\partial a \partial \tau} &= -\sum_{i=1}^{n} \frac{d_i^\tau \ln d_i}{1+d_i} - (N-n)\frac{C^\tau \ln C}{1+C^\tau} \\
\frac{\partial^2 l_b}{\partial a \partial \theta} &= \tau\theta^{-1}\sum_{i=1}^{n} \frac{d_i^\tau}{1+d_i^\tau} + (N-n)\tau\theta^{-1}\frac{C^\tau}{1+C^\tau} \qquad (5.31) \\
\frac{\partial^2 l_b}{\partial \tau \partial \theta} &= -n\theta^{-1} + (a+1)\theta^{-1}\sum_{i=1}^{n} \frac{d_i^\tau}{1+d_i^\tau} + \tau(a+1)\theta^{-1}\sum_{i=1}^{n} \frac{d_i^\tau \ln d_i}{[1+d_i^\tau]^2} \\
&\quad + a(N-n)\theta^{-1}\frac{C^\tau}{1+C^\tau} + a(N-n)\theta^{-1}\tau\frac{C^\tau \ln C}{[1+C^\tau]^2} \qquad (5.32) \\
\frac{\partial^2 l_b}{\partial \theta^2} &= n\tau\theta^{-2} - \tau(a+1)\theta^{-2}\sum_{i=1}^{n} \frac{d_i^\tau}{1+d_i^\tau} - \tau^2(a+1)\theta^{-2}\sum_{i=1}^{n} \frac{d_i^\tau}{[1+d_i^\tau]^2} \\
&\quad - \tau(N-n)a\theta^{-2}\frac{C^\tau}{1+C^\tau} - \tau^2 a(N-n)\theta^{-2}\frac{C^\tau}{[1+C^\tau]^2} \qquad (5.33)
\end{aligned}$$

We see that derivatives with respect to the shape parameters alone are essentially the same as for the two-parameter model previously considered; the only difference being the inclusion of the scale parameter θ. Consequently, when looking at expectations of the score functions below, we can concentrate on the remaining first derivative, $\frac{\partial l_b}{\partial \theta}$.

5.4.1 Expectations in Derivatives

We note that we now need to find expectations of the following functions to be able to write down expectations of all derivatives.

$$\frac{D^\tau}{1+D^\tau}, \quad \frac{D^\tau}{(1+D^\tau)^2} \text{ and } \frac{D^\tau \ln D}{(1+D^\tau)^2}.$$

Making the necessary adjustment to Theorem 3.2 to accomodate censoring, we have

$$E_a\left[\frac{D^{i\tau}\left(\ln D\right)^j}{\left(1+D^\tau\right)^k}\right] = \frac{aq_{c,a+k}}{(a+k)\,q_{c,a}} E_{a+k}\left[D^{i\tau}\left(\ln D\right)^j\right]. \tag{5.34}$$

Therefore, we have,

$$\begin{aligned}
E\left[\frac{D^\tau}{1+D^\tau}\right] &= E_a\left[\frac{D^\tau}{1+D^\tau}\right] = \frac{aq_{c,a+1}}{(a+1)\,q_{c,a}} E_{a+1}\left[D^\tau\right] \\
E\left[\frac{D^\tau}{(1+D^\tau)^2}\right] &= E_a\left[\frac{D^\tau}{(1+D^\tau)^2}\right] = \frac{aq_{c,a+2}}{(a+2)\,q_{c,a}} E_{a+2}\left[D^\tau\right] \\
E\left[\frac{D^\tau \ln D}{(1+D^\tau)^2}\right] &= E_a\left[\frac{D^\tau \ln D}{(1+D^\tau)^2}\right] = \frac{aq_{c,a+2}}{(a+2)\,q_{c,a}} E_{a+2}\left[D^\tau \ln D\right]
\end{aligned}$$

So,

$$\begin{aligned}
E\left[\frac{D^\tau}{1+D^\tau}\right] &= \frac{aq_{c,a+1}}{(a+1)\,q_{c,a}} E_{a+1}\left[D^\tau\right] \\
&= \frac{aq_{c,a+1}}{(a+1)\,q_{c,a}} \frac{(a+1)\,\omega}{q_{c,a+1}} \frac{c^\tau}{2} f_1\left(2, a+2, -\omega\right)
\end{aligned}$$

using (5.10);

$$= \frac{a\omega^2}{2q_{c,a}}\left[\frac{2}{(a+1)\,\omega}\left\{\frac{q_{c,a}}{\omega a} - \frac{1}{(1+\omega)^{a+1}}\right\}\right] \tag{5.35}$$

using (5.26). Secondly,

$$\begin{aligned}
E\left[\frac{D^\tau}{(1+D^\tau)^2}\right] &= \frac{aq_{c,a+2}}{(a+2)\,q_{c,a}} E_{a+2}\left[D^\tau\right] \\
&= \frac{aq_{c,a+2}}{(a+2)\,q_{c,a}} \frac{(a+2)\,\omega}{q_{c,a+2}} \frac{c^\tau}{2} f_1\left(2, a+3, -\omega\right)
\end{aligned}$$

using (5.10);

$$= \frac{a\omega^2}{2q_{c,a}}\left[\frac{2}{(a+2)\,\omega}\left\{\frac{q_{c,a+1}}{\omega\,(a+1)} - \frac{1}{(1+\omega)^{a+2}}\right\}\right] \tag{5.36}$$

using results from section 5.3.2; and finally

$$\begin{aligned}
E\left[\frac{D^\tau \ln D}{(1+D^\tau)^2}\right] &= \frac{aq_{c,a+2}}{(a+2)\,q_{c,a}} E_{a+2}\left[D^\tau \ln D\right] \\
&= \frac{a}{q_{c,a}}\left\{E_{a+2}\left[D^\tau \ln D\right] \frac{q_{c,a+2}}{(a+2)}\right\} \\
&= \frac{a\omega^2}{2\tau q_{c,a}}\left\{\left[\begin{array}{c} \frac{(2\log\omega - 1) f_1(2, a+3, -\omega)}{2} \\ -\frac{(a+3)\omega f_2(3, a+4, -\omega)}{9} \end{array}\right]\right\}
\end{aligned} \tag{5.37}$$

from (5.14). It now remains to show that the expectation of $\frac{\partial l_b}{\partial \theta}$ is zero. With $C' = \frac{C^\tau}{1+C^\tau}$,

$$\begin{aligned}
E\left[\frac{\partial l_b}{\partial \theta}\right] &= -E[n]\tau\theta^{-1} + \tau(a+1)\theta^{-1}E[n]\,E\left[\frac{D^\tau}{1+D^\tau}\right] + aE[N-n]\tau\theta^{-1}\frac{C^\tau}{1+C^\tau} \\
&= -\tau\theta^{-1}\left(N - \frac{N}{\{1+C^\tau\}^a}\right) + a\left(\frac{N}{\{1+C^\tau\}^a}\right)\tau\theta^{-1}\frac{C^\tau}{1+C^\tau} \\
&\quad +\tau(a+1)\theta^{-1}\left(N - \frac{N}{\{1+C^\tau\}^a}\right)\frac{a}{1-(1+C^\tau)^{-a}}B_{C'}(2,a) \\
&= -\tau\theta^{-1}\left(N - \frac{N}{\{1+C^\tau\}^a}\right) + \tau(a+1)\theta^{-1}\left(N - \frac{N}{\{1+C^\tau\}^a}\right)\frac{a}{1-(1+C^\tau)^{-a}} \\
&\quad \times\frac{1}{2}(C')^2 F(2,1-a,3,C') + a\left(\frac{N}{\{1+C^\tau\}^a}\right)\tau\theta^{-1}\frac{C^\tau}{1+C^\tau}, \\
&= \tau\theta^{-1}\left(N - \frac{N}{\{1+C^\tau\}^a}\right)\left[-1 + \frac{-2\left(-1+(1-C')^a + a(1-C')^a C'\right)}{2\left(1-(1+C^\tau)^{-a}\right)}\right] \\
&\quad +a\left(\frac{N}{\{1+C^\tau\}^a}\right)\tau\theta^{-1}\frac{C^\tau}{1+C^\tau}
\end{aligned}$$

It can be shown that this then simplifies to

$$-\frac{(1+C^\tau)^{-1-a}\{1+(1+a)C^\tau\}\{-1+(1+C^\tau)^{-a}(1+C^\tau)^a\}N\tau}{\theta}$$

where we note that $(1+C^\tau)^{-a}(1+C^\tau)^a$ is 1, giving zero in the third bracket in the numerator and so the expectation is zero. Hence,

$$E\left[\frac{\partial l}{\partial \theta}\right] = 0.$$

5.4.2 Expectations of second derivatives

The remaining second derivatives, those with respect to the scale parameter θ, namely (5.31), (5.32) and (5.33), are considered here. Expectations of second derivatives with respect to shape parameters alone, were given for the two-parameter model

$$\begin{aligned}
E\left[\frac{\partial^2 l_b}{\partial a \partial \theta}\right] &= \tau\theta^{-1}E\left[t^d_{101}\right] + \tau\theta^{-1}E[N-n]\frac{C^\tau}{1+C^\tau} \\
E\left[\frac{\partial^2 l_b}{\partial \tau \partial \theta}\right] &= -n\theta^{-1} + (a+1)\theta^{-1}E\left[t^d_{101}\right] + \tau(a+1)\theta^{-1}E\left[t^d_{112}\right] \\
&\quad +a\theta^{-1}E[N-n]\frac{C^\tau}{1+C^\tau} + a\theta^{-1}\tau E[N-n]\frac{C^\tau \ln C}{(1+C^\tau)^2} \\
E\left[\frac{\partial^2 l_b}{\partial \theta^2}\right] &= n\tau\theta^{-2} - \tau(a+1)\theta^{-2}E\left[t^d_{101}\right] - \tau^2(a+1)\theta^{-2}E\left[t^d_{102}\right] \\
&\quad -\tau a\theta^{-2}E[N-n]\frac{C^\tau}{1+C^\tau} - \tau^2 a\theta^{-2}E[N-n]\frac{C^\tau}{(1+C^\tau)^2}
\end{aligned}$$

It is now just a straightforward matter of substituting in the appropriate expectations - (5.35), (5.36) and (5.37). Inevitably there will be several ways of expressing the expectations; some forms simpler than others. However, since we are interested in these expressions only in terms of the inverse of the EFI matrix, we do not write down the expectations explicitly.

5.5 Burr XII ALT Model

Here, the d_{ij} are observations from a truncated Burr XII distribution with PDF

$$\frac{a\tau\left\{\exp\left(\alpha_b+\beta_b x_i\right)\right\}^{-\tau} d^{\tau-1}\left(1+\left(\frac{d}{\exp(\alpha_b+\beta_b x_i)}\right)^{\tau}\right)^{-(a+1)}}{q_{c,i,a}},$$

where

$$q_{c,i,a}=1-\left(1+\left(\frac{c}{\exp\left(\alpha_b+\beta_b x_i\right)}\right)^{\tau}\right)^{-a}.$$

The likelihood for data obtained at the i^{th} stress level is then given by

$$\begin{aligned}&\left[\prod_{i=1}^{n_i}\frac{a\tau\theta_i^{-\tau}d_{ij}^{\tau-1}}{\left\{1+\left(\frac{d_{ij}}{\theta_i}\right)^{\tau}\right\}^{a+1}}\right]\left[\prod_{i=n_i+1}^{N_i}\left\{1+\left(\frac{d_{ij}}{\theta_i}\right)^{\tau}\right\}^{-a}\right]\\ =&\left[\frac{(a\tau)^{n_i}\theta_i^{-n_i\tau}\prod_{i=1}^{n_i}d_{ij}^{\tau-1}}{\prod_{i=1}^{n_i}\left\{1+\left(\frac{d_{ij}}{\theta_i}\right)^{\tau}\right\}^{a+1}}\right]\left[\prod_{i=1}^{m_i}\left\{1+\left(\frac{c_{ij}}{\theta_i}\right)^{\tau}\right\}^{-a}\right]\end{aligned}$$

so that the log-likelihood for the entire sample is

$$l_B\left(a,\tau,\alpha_b,\beta_b\right)=S_n\ln\left(a\tau\right)+\left(\tau-1\right)S_e-\tau\alpha_b S_n-\tau\beta_b S_x-\left(a+1\right)S_*^d-aS_*^c \qquad (5.38)$$

where

$$S_*^c=\sum_{i=1}^{k}\sum_{j=1}^{m_i}\ln\left\{1+\left(\frac{c_{ij}}{\exp\left(\alpha_b+\beta_b x_i\right)}\right)^{\tau}\right\}$$

and we extend the notation in (2.12) to include

$$S_{klmn}^p=\sum_{i=1}^{k}\sum_{j}\left(\frac{x_i^k\left\{\frac{p_{ij}}{\exp(\alpha_b+\beta_b x_i)}\right\}^{l\tau}\left[\ln\left\{\frac{p_{ij}}{\exp(\alpha_b+\beta_b x_i)}\right\}\right]^m}{\left(1+\left\{\frac{p_{ij}}{\exp(\alpha_b+\beta_b x_i)}\right\}^{\tau}\right)^n}\right)$$

where the upper limit of the summations of j is n_i for $p=d$ and m_i for $p=c$. All other functions are as defined in sections 2.1 and 2.3. The first derivatives of (5.38) are

$$\frac{\partial l_B}{\partial\tau}=S_n\tau^{-1}+S_e-\alpha_b S_n-\beta_b S_x-\left(a+1\right)S_{0111}^d-aS_{0111}^c, \qquad (5.39)$$

$$\frac{\partial l_B}{\partial a} = S_n a^{-1} - S^d_* - S^c_*, \tag{5.40}$$

$$\frac{\partial l_B}{\partial \alpha_b} = -\tau S_n + \tau\,(a+1)\,S^d_{0101} + \tau a S^c_{0101}, \tag{5.41}$$

$$\frac{\partial l_B}{\partial \beta_b} = -S_x\tau + (a+1)\,\tau S^d_{1101} + a\tau S^c_{1101}; \tag{5.42}$$

while second derivatives of (5.38) are given by,

$$\frac{\partial^2 l_B}{\partial a^2} = -S_n a^{-2}, \tag{5.43}$$

$$\frac{\partial^2 l_B}{\partial \tau^2} = -S_n\tau^{-2} - (a+1)\left[S^d_{0121} - S^d_{0222}\right] - a\left[S^c_{0121} - S^c_{0222}\right], \tag{5.44}$$

$$\frac{\partial^2 l_B}{\partial \alpha_b^2} = (a+1)\,\tau^2\left[S^d_{0202} - S^d_{0101}\right] + a\tau^2\left[S^c_{0202} - S^c_{0101}\right], \tag{5.45}$$

$$\frac{\partial^2 l_B}{\partial \beta_b^2} = (a+1)\,\tau^2\left[S^d_{2202} - S^d_{2101}\right] + a\tau^2\left[S^c_{2202} - S^c_{2101}\right], \tag{5.46}$$

and

$$\frac{\partial^2 l_B}{\partial a \partial \tau} = -S^d_{0111} - S^c_{0111}, \tag{5.47}$$

$$\frac{\partial^2 l_B}{\partial a \partial \alpha_b} = \tau S^d_{0101} + \tau S^c_{0101}, \tag{5.48}$$

$$\frac{\partial^2 l_B}{\partial a \partial \beta_b} = \tau S^d_{1101} + \tau S^c_{1101}, \tag{5.49}$$

$$\begin{aligned}\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b} &= -S_n + (a+1)\left[S^d_{0101} + \tau S^d_{0111} - \tau S^d_{0212}\right] \\ &\quad + a\left[S^c_{0101} + \tau S^c_{0111} - \tau S^c_{0212}\right],\end{aligned} \tag{5.50}$$

$$\begin{aligned}\frac{\partial^2 l_B}{\partial \tau \partial \beta_b} &= -S_x + (a+1)\left[S^d_{1101} + \tau S^d_{1111} - \tau S^d_{1212}\right] \\ &\quad + a\left[S^c_{1101} + \tau S^c_{1111} - \tau S^c_{1212}\right],\end{aligned} \tag{5.51}$$

$$\frac{\partial^2 l_B}{\partial \alpha_b \partial \beta_b} = (a+1)\,\tau^2\left[S^d_{1202} - S^d_{1101}\right] + a\tau^2\left[S^c_{1202} - S^c_{1101}\right]. \tag{5.52}$$

To establish the maximum likelihood estimates of the parameters in our model we use the same basic algorithm as outlined for complete data in Chapter two. The SAS program *burrcens* - given in Appendix A - accounts for censoring.

Expectations in Derivatives

Now, for the complete case, expectations for accelerated elements were based on expectations for non accelerated data. Similarly, we can define expectations for ALT Type I censored items in terms of the non ALT Type I terms.

So, we require expectations of the following terms

$$S_e, S_*^p, S_{0111}^p, S_{0101}^p, S_{1101}^p, S_{0121}^p, S_{0222}^p, S_{0212}^p, \quad (5.53)$$
$$S_{1111}^p, S_{1212}^p, S_{0202}^p, S_{1202}^p, S_{2202}^p, S_{2101}^p$$

where $p = d$ or c. Now, recall for the non-accelerated case, the expected number of failed items was the total number of items in the sample multiplied by $q_{c,a}$, the probability an items fails in (0,c), (see section 5.3.1.) Similarly, here,

$$\begin{aligned} E[n_i] &= N_i \times q_{c,i,a} \\ E[S_n] &= \sum_{i=1}^{k} N_i q_{c,i,a} \\ E[m_i] &= E[N_i - n_i] = N_i(1 - q_{c,i,a}) \end{aligned}$$

On letting $Y_{ij} = \frac{d_{ij}}{\exp(\alpha_b+\beta_b x_i)}$ and $\omega_i = \left(\frac{c}{\exp(\alpha_b+\beta_b x_i)}\right)^\tau$ and adopting the techniques used in Abramowitz and Stegun (1972), together with terms derived in Watkins (1997), we can determine these expectations as detailed below. Figures 5.4 to 5.17 illustrate the agreement between simulated and theoretical results, giving the average of 10000 replications. We also remind the reader of the analogous complete results and the resulting asymptotic agreement. For illustration, we use the parameter values $a = 2$, $\tau = 5$, $\alpha_b = 17.60139$, $\beta_b = -0.056282$, with the equally weighted sample size $N_1 = 1000$ and stress levels $(x_1, x_2, x_3) = (150, 170, 190)$. This gives scale values $(\theta_1, \theta_2, \theta_3) \approx (9500, 3082, 1000)$. We also present a sample set of values for $q_{c,i,a}$ for each stress level and various values of c, to illustrate the probability of an item failing by time c at each of the stress levels. They are,

	Stress		
c	150	170	190
4000	0.025945	0.954349	0.999999
8000	0.506425	0.999929	1
12000	0.943716	0.999999	1
16000	0.995275	1	1
20000	0.999442	1	1

from which we can see that we would expect practically all items to have failed at every stress level setting by time $c = 16000$.

$\mathbf{E}[S_*^d]$ The complete case is given at (3.10), where we have

$$E\left[S_*^d\right] = \frac{S_n}{a},$$

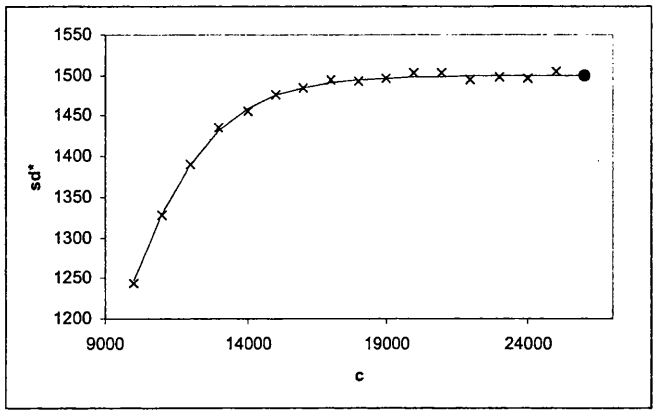

Figure 5.4: Theoretical and simulated values of $E\left[S_*^d\right]$ versus c.

while for Type I,

$$\begin{aligned} E\left[S_*^d\right] &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln\left(1+Y_{ij}^{\tau}\right)\right] \\ &= \sum_{i=1}^{k} q_{c,i,a}N_i E\left[\ln\left(1+Y^{\tau}\right)\right] \\ &= \sum_{i=1}^{k} q_{c,i,a}N_i\left[\frac{1}{a}-\frac{\ln\left(1+\omega_i\right)}{\left(1+\omega_i\right)^a-1}\right], \end{aligned}$$

using (5.7); Figure 5.4 presents a comparison of the Type I simulated expectation ($\times$), Type I theoretical expected result, (continuous line), and complete result, ($\cdot$) as $c \to \infty$, for the chosen parameters.

E$[S_e]$ From (3.11), the complete result is

$$E\left[S_e\right] = S_n\alpha_b + S_x\beta_b - \frac{S_n\left\{\gamma+\psi\left(a\right)\right\}}{\tau},$$

while the Type I censored result is

$$\begin{aligned} E\left[S_e\right] &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln d_{ij}\right] \\ &= \sum_{i=1}^{k} E\left[n_i\right]\left\{\ln\theta_i + E\left[\ln Y\right]\right\} \\ &= \sum_{i=1}^{k}\left\{\alpha_b q_{c,i,a}N_i + \beta_b q_{c,i,a}N_i x_i + q_{c,i,a}N_i\left[\frac{1}{\tau}\left(\ln\omega_i - \frac{f_2\left(1,a+1,-\omega_i\right)}{f_1\left(1,a+1,-\omega_i\right)}\right)\right]\right\}, \end{aligned}$$

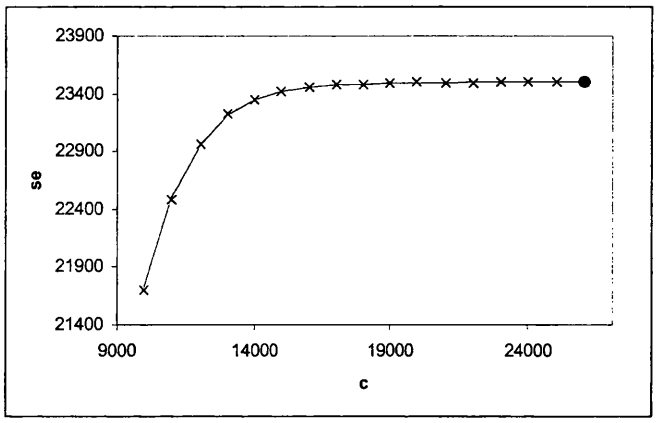

Figure 5.5: Theoretical and simulated values of $E[S_e]$ versus c.

using the full form of (5.29); the asymptotic behaviour of this function is presented in Figure 5.5.

$\mathbf{E}\left[S_{0111}^{d}\right]$ For the complete result, we have

$$E\left[S_{0111}^{d}\right] = \frac{S_n\left\{1-\gamma-\psi\left(a\right)\right\}}{\left(a+1\right)\tau},$$

from (3.14). The Type I censored result is given by

$$\begin{aligned} E\left[S_{0111}^{d}\right] &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{Y_{ij}^{\tau}\ln Y_{ij}}{1+Y_{ij}^{\tau}}\right] \\ &= \sum_{i=1}^{k}\left\{q_{c,i,a}N_i\left\{\frac{a\omega_i^2}{2\tau q_{c,i,a}}\left[\begin{array}{c}\frac{\left(2\ln\omega_i-1\right)}{2}f_1\left(2,a+2,-\omega_i\right)\\ -\frac{(a+2)\omega_i}{9}f_2\left(3,a+3,-\omega_i\right)\end{array}\right]\right\}\right\}, \end{aligned}$$

using (5.19); the asymptotic behaviour of this function is presented in Figure 5.6.

$\mathbf{E}\left[S_{0101}^{d}\right]$ From (3.12) we have the result for complete data as

$$E\left[S_{0101}^{d}\right] = \frac{S_n}{a+1},$$

while the result for Type I censoring is

$$E\left[S_{0101}^{d}\right] = E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{Y_{ij}^{\tau}}{1+Y_{ij}^{\tau}}\right]$$

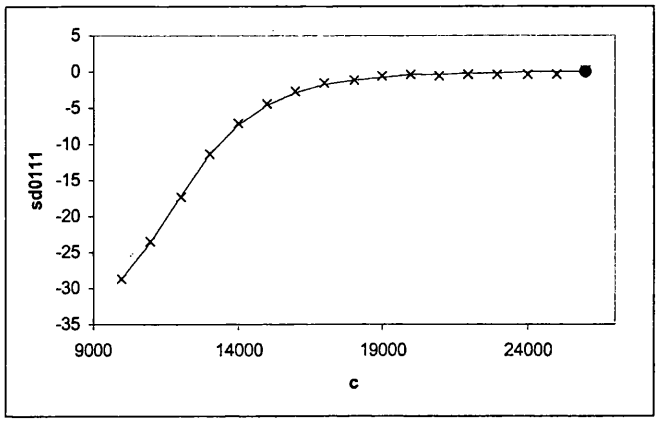

Figure 5.6: Theoretical and simulated values of $E\left[S^d_{0111}\right]$ versus c.

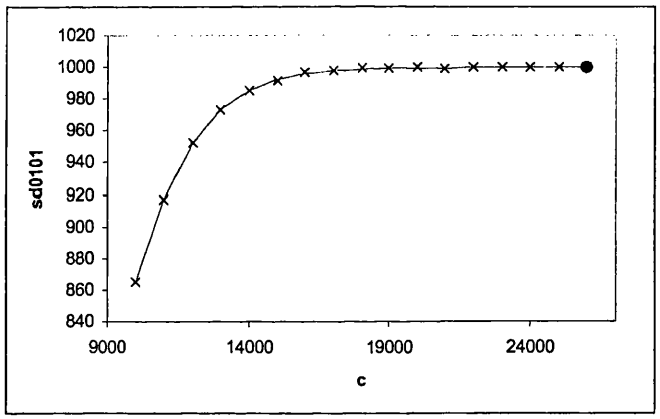

Figure 5.7: Theoretical and simulated values of $E\left[S^d_{0101}\right]$ versus c.

$$= \sum_{i=1}^{k}\left\{q_{c,i,a}N_i\left[\frac{a\omega_i^2}{2q_{c,i,a}}\left[\frac{2}{(a+1)\,\omega_i}\left\{\frac{q_{c,i,a}}{\omega_i a}-\frac{1}{(1+\omega_i)^{a+1}}\right\}\right]\right]\right\},$$

from (5.35); the asymptotic behaviour of this function is presented in Figure 5.7.

$\mathbf{E}\left[S^d_{0121}\right]$ Here, the complete result is

$$E\left[S^d_{0121}\right]=\frac{S_n}{\tau^2\,(a+1)}\left[\begin{array}{c}\left\{\frac{\pi^2}{6}+\gamma^2-2\gamma\right\}+2\,(\gamma-1)\,\psi\,(a)\\ +\left\{\psi\,(a)\right\}^2+\psi'\,(a)\end{array}\right],$$

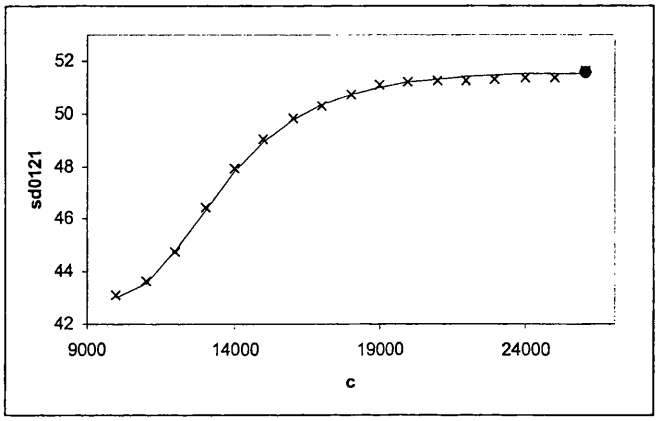

Figure 5.8: Theoretical and simulated values of $E\left[S^d_{0121}\right]$ versus c.

using (3.8), while, by (5.15), the Type I result is

$$\begin{aligned} E\left[S^d_{0121}\right] &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{Y_{ij}^{\tau}\left(\ln Y_{ij}\right)^2}{1+Y_{ij}^{\tau}}\right] \\ &= \frac{a}{a+1}\sum_{i=1}^{k}\frac{q_{c,i,a+1}}{q_{c,i,a}}E\left[n_i\right]E_{a+1}\left[Y_{ij}^{\tau}\left(\ln Y_{ij}\right)^2\right] \\ &= \sum_{i=1}^{k}q_{c,i,a}N_i\left[\frac{aq_{c,i,a+1}}{(a+1)\,q_{c,i,a}}E_{a+1}\left[Y^{\tau}\left(\ln Y\right)^2\right]\right] \\ &= \sum_{i=1}^{k}N_i\frac{a\omega_i^2}{\tau^2}\times\begin{bmatrix}\frac{1}{8}\left\{1+\left(2\ln\omega_i-1\right)^2\right\}f_1\left(2,a+2,-\omega_i\right)\\ -\frac{1}{54}\left(a+2\right)\omega_i\left(6\ln\omega_i-5\right)f_2\left(3,a+3,-\omega_i\right)\\ +\frac{1}{192}\left(a+2\right)\left(a+3\right)\omega_i^2 f_3\left(4,a+4,-\omega_i\right)\end{bmatrix}; \end{aligned}$$

the asymptotic behaviour of this function is presented in Figure 5.8.

$\mathbf{E}\left[S^d_{0222}\right]$ From (3.16), we have the complete result as

$$E\left[S^d_{0222}\right]=\frac{S_n}{\tau^2\left(a+1\right)\left(a+2\right)}\begin{bmatrix}\left\{\frac{\pi^2}{3}+2\gamma^2-6\gamma+2\right\}-2\left(3-2\gamma\right)\psi\left(a\right)\\ +2\left\{\psi^2\left(a\right)+\psi'\left(a\right)\right\}\end{bmatrix}.$$

Using the result at (5.12) and simplifying, the Type I result is

$$\begin{aligned} E\left[S^d_{0222}\right] &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{Y_{ij}^{2\tau}\left(\ln Y_{ij}\right)^2}{\left[1+Y_{ij}^{\tau}\right]^2}\right] \\ &= \sum_{i=1}^{k}q_{c,i,a}N_i\left[\frac{a}{a+2}\frac{q_{c,i,a+2}}{q_{c,i,a}}E_{a+2}\left[Y^{2\tau}\left(\ln Y\right)^2\right]\right] \end{aligned}$$

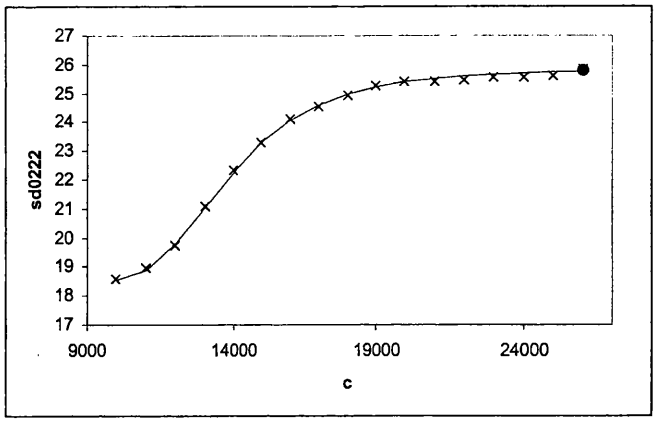

Figure 5.9: Theoretical and simulated values of $E\left[S^d_{0222}\right]$ versus c.

$$= \sum_{i=1}^{k} \frac{a\omega_i N_i}{\tau^2} \left\{ \begin{array}{c} \frac{(a+3)(a+4)\omega_i^4}{750} f_3\left(5, a+5, -\omega_i\right) \\ -\frac{(a+3)\omega_i^3}{288}\left(-3+4\left(-1+3\ln\omega_i\right)\right) f_2\left(4, a+4, -\omega_i\right) \\ +\frac{1}{27}\omega_i^2\left(1+\left(-1+3\ln\omega_i\right)^2\right) f_1\left(3, a+3, -\omega_i\right) \end{array} \right\};$$

the asymptotic behaviour of this function is presented in Figure 5.9.

$\mathbf{E}\left[S^d_{0212}\right]$ The complete result is

$$E\left[S^d_{0212}\right] = \frac{S_n\left\{3-2\gamma-2\psi\left(a\right)\right\}}{\tau\left(a+1\right)\left(a+2\right)}.$$

from (3.15), while the Type I result is given by

$$\begin{aligned} E\left[S^d_{0212}\right] &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} \frac{Y_{ij}^{2\tau}\ln Y_{ij}}{\left[1+Y_{ij}^{\tau}\right]^2}\right] \\ &= \sum_{i=1}^{k} q_{c,i,a} N_i \left[\frac{a}{a+2}\frac{q_{c,i,a+2}}{q_{c,i,a}} E_{a+2}\left[Y^{2\tau}\ln Y\right]\right] \\ &= \sum_{i=1}^{k} \frac{\omega_i a N_i}{\tau}\left\{\frac{\omega_i^2\left(3\ln\omega_i-1\right)}{9} f_1\left(3, a+3, -\omega_i\right) - \frac{\left(a+3\right)\omega_i^3}{48} f_2\left(4, a+4, -\omega_i\right)\right\}, \end{aligned}$$

using (5.11) and simplifying; the asymptotic behaviour of this function is presented in Figure 5.10.

$\mathbf{E}\left[S^d_{0202}\right]$ The result at (3.13) gives us the expression for complete data as

$$E\left[S^d_{0202}\right] = \frac{2S_n}{\left(a+1\right)\left(a+2\right)};$$

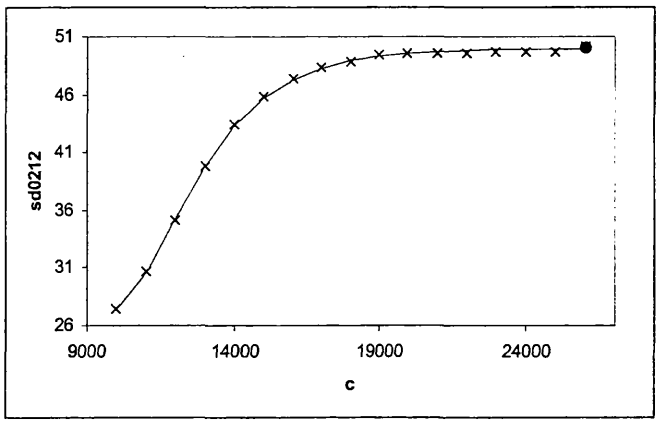

Figure 5.10: Theoretical and simulated values of $E\left[S^d_{0212}\right]$ versus c.

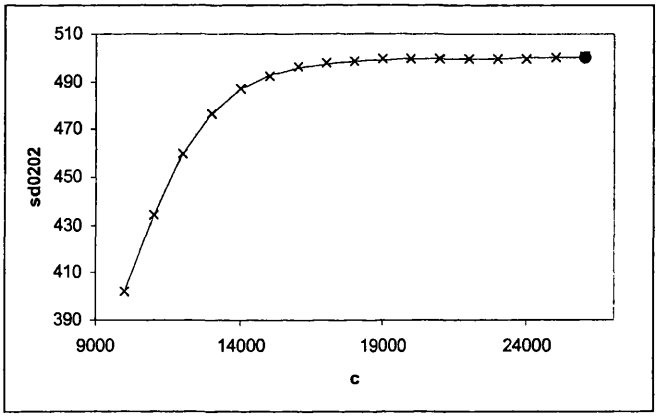

Figure 5.11: Theoretical and simulated values of $E\left[S^d_{0202}\right]$ versus c.

while the analogous result for Type I censored data is

$$\begin{aligned} E\left[S^d_{0202}\right] &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{Y_{ij}^{2\tau}}{\left[1+Y_{ij}^{\tau}\right]^2}\right] \\ &= \sum_{i=1}^{k} q_{c,i,a}N_i\left[\frac{a}{a+2}\frac{q_{c,i,a+2}}{q_{c,i,a}}E_{a+2}\left[Y^{2\tau}\right]\right] \\ &= \sum_{i=1}^{k}\frac{a\omega_i^3 N_i}{3}f_1\left(3,a+3,-\omega_i\right), \end{aligned}$$

using (5.10); the asymptotic behaviour of this function is presented in Figure 5.11.

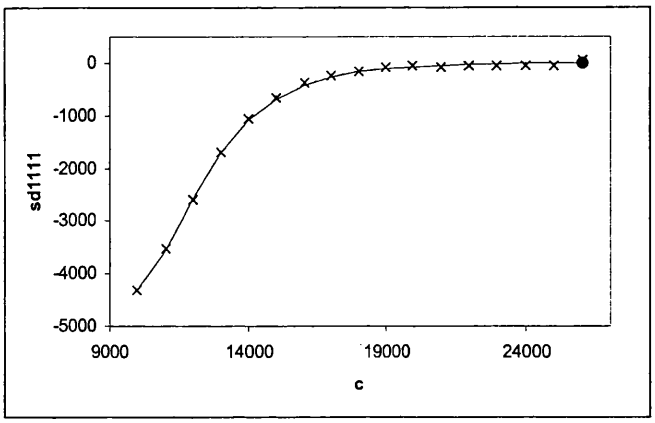

Figure 5.12: Theoretical and simulated values of $E\left[S^d_{1111}\right]$ versus c.

$\mathbf{E}\left[S^d_{1111}\right]$ We have the complete result as

$$E\left[S_{1111}\right] = \frac{S_x\left\{1-\gamma-\psi\left(a\right)\right\}}{\tau\left(a+1\right)},$$

using the result at (3.14). Using the modified form of S^d_{0111} above, we have the Type I censored result as

$$\begin{aligned} E\left[S^d_{1111}\right] &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{x_i Y_{ij}^{\tau}\ln Y_{ij}}{1+Y_{ij}^{\tau}}\right] \\ &= \sum_{i=1}^{k}\left\{N_i x_i\left\{\frac{a\omega_i^2}{2\tau}\left[\begin{array}{c}\frac{(2\ln\omega_i-1)}{2}f_1\left(2,a+2,-\omega_i\right)\\ -\frac{(a+2)\omega_i}{9}f_2\left(3,a+3,-\omega_i\right)\end{array}\right]\right\}\right\}; \end{aligned}$$

the asymptotic behaviour of this function is presented in Figure 5.12.

$\mathbf{E}\left[S^d_{1212}\right]$ Adapting the result for S^d_{0212} above, we have the complete result,

$$E\left[S_{1212}\right] = \frac{S_x\left\{3-2\gamma-2\psi\left(a\right)\right\}}{\tau\left(a+1\right)\left(a+2\right)}$$

while, using the result S^d_{0212} above, we have the Type I censored result.

$$\begin{aligned} E\left[S^d_{1212}\right] &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{x_i Y_{ij}^{2\tau}\ln Y_{ij}}{\left[1+Y_{ij}^{\tau}\right]^2}\right] \\ &= \sum_{i=1}^{k}\frac{\omega_i a N_i x_i}{\tau}\left\{\begin{array}{c}\frac{\omega_i^2(3\ln\omega_i-1)}{9}f_1\left(3,a+3,-\omega_i\right)\\ -\frac{(a+3)\omega_i^3}{48}f_2\left(4,a+4,-\omega_i\right)\end{array}\right\}; \end{aligned}$$

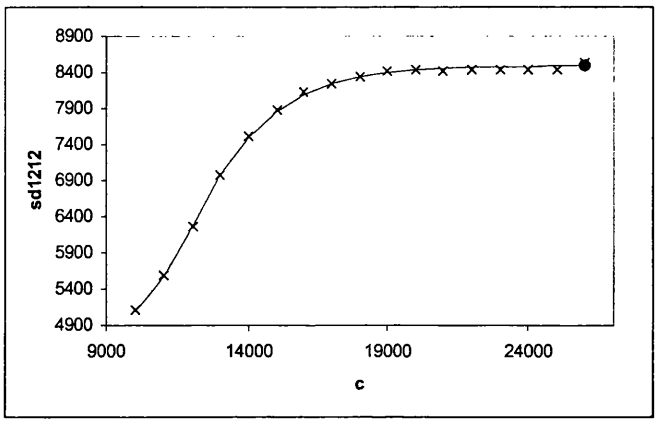

Figure 5.13: Theoretical and simulated values of $E\left[S^d_{1212}\right]$ versus c.

the asymptotic behaviour of this function is presented in Figure 5.13.

$\mathbf{E}\left[S^d_{1202}\right]$ Adapting S^d_{0202} above, we have the complete result.

$$E\left[S_{1202}\right] = \frac{2S_x}{(a+1)(a+2)}.$$

The Type I censored result is given by

$$\begin{aligned} E\left[S^d_{1202}\right] &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{x_i Y_{ij}^{2\tau}}{\left[1+Y_{ij}^{\tau}\right]^2}\right] \\ &= \sum_{i=1}^{k}\frac{a\omega_i^3 N_i x_i}{3} f_1\left(3, a+3, -\omega_i\right), \end{aligned}$$

adapting the form of S^d_{0202} above; the asymptotic behaviour of this function is presented in Figure 5.14.

$\mathbf{E}\left[S^d_{2202}\right]$ Adapting S^d_{0202} above, we have the complete result

$$E\left[S_{2202}\right] = \frac{2S_{xx}}{(a+1)(a+2)},$$

while the Type I result is

$$E\left[S^d_{2202}\right] = E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{x_i^2 Y_{ij}^{2\tau}}{\left[1+Y_{ij}^{\tau}\right]^2}\right]$$

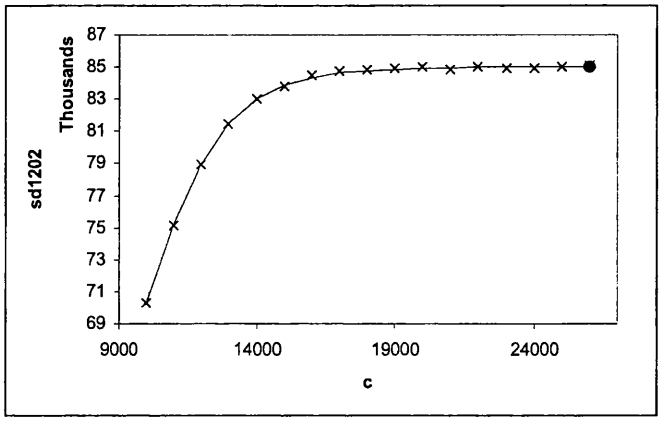

Figure 5.14: Theoretical and simulated values of $E\left[S^d_{1202}\right]$ versus c.

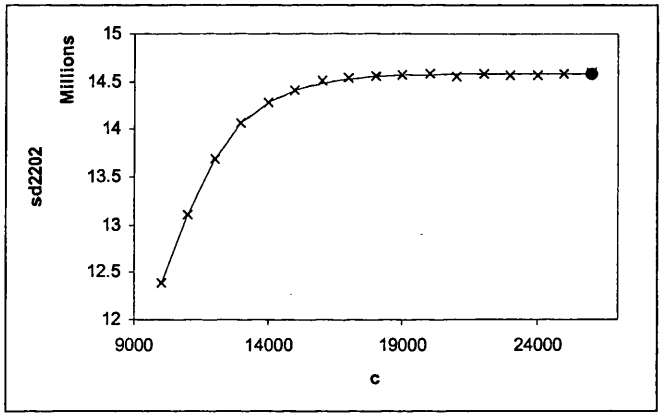

Figure 5.15: Theoretical and simulated values of $E\left[S^d_{2202}\right]$ versus c.

$$= \sum_{i=1}^{k} \frac{a\omega_i^3 N_i x_i^2}{3} f_1\left(3, a+3, -\omega_i\right),$$

making the minor modification to S^d_{0202} above; the asymptotic behaviour of this function is presented in Figure 5.15.

$\mathbf{E}[S^d_{2101}]$ Adapting S^d_{0101} we have the complete result

$$E\left[S_{2101}\right] = \frac{S_{xx}}{a+1}.$$

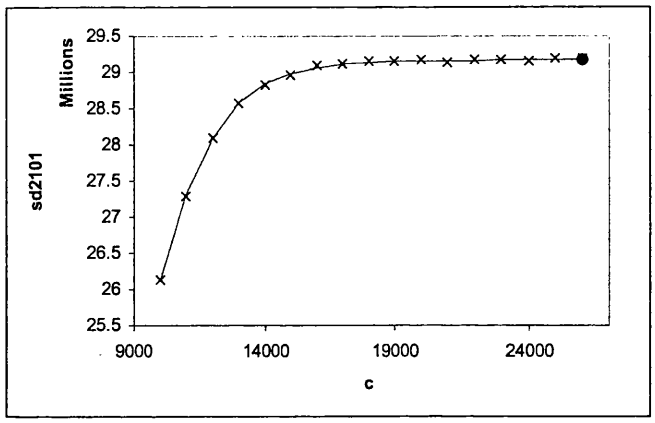

Figure 5.16: Theoretical and simulated values of $E\left[S^d_{2101}\right]$ versus c.

Adapting the form of S^d_{0101} above, the Type I result is given by

$$\begin{aligned} E\left[S^d_{2101}\right] &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{x_i^2 Y_{ij}^{\tau}}{1+Y_{ij}^{\tau}}\right] \\ &= \sum_{i=1}^{k}\left\{q_{c,i,a}N_i x_i^2\left[\frac{a\omega_i^2}{2q_{c,i,a}}\left[\frac{2}{(a+1)\,\omega_i}\left\{\frac{q_{c,i,a}}{\omega_i a}-\frac{1}{(1+\omega_i)^{a+1}}\right\}\right]\right]\right\}; \end{aligned}$$

the asymptotic behaviour of this function is presented in Figure 5.16.

$\mathbf{E}[S^d_{1101}]$ Adapting S^d_{0101} we have the complete result

$$E\left[S_{1101}\right] = \frac{S_x}{a+1},$$

while the Type I censoring result is

$$\begin{aligned} E\left[S^d_{1101}\right] &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{x_i Y_{ij}^{\tau}}{1+Y_{ij}^{\tau}}\right] \\ &= \sum_{i=1}^{k}\left\{q_{c,i,a}N_i x_i\left[\frac{a\omega_i^2}{2q_{c,i,a}}\left[\frac{2}{(a+1)\,\omega_i}\left\{\frac{q_{c,i,a}}{\omega_i a}-\frac{1}{(1+\omega_i)^{a+1}}\right\}\right]\right]\right\}, \end{aligned}$$

adapting the result for S^d_{0101} above; the asymptotic behaviour of this function is presented in Figure 5.17.For the censored data (superscript c) we would have, for example, in (5.53),

$$E\left[S^c_{2101}\right] = E\left[\sum_{i=1}^{k} m_i \frac{x_i^2\left(\frac{c}{\exp(\alpha_b+\beta_b x_i)}\right)^{\tau}}{1+\left(\frac{c}{\exp(\alpha_b+\beta_b x_i)}\right)^{\tau}}\right]$$

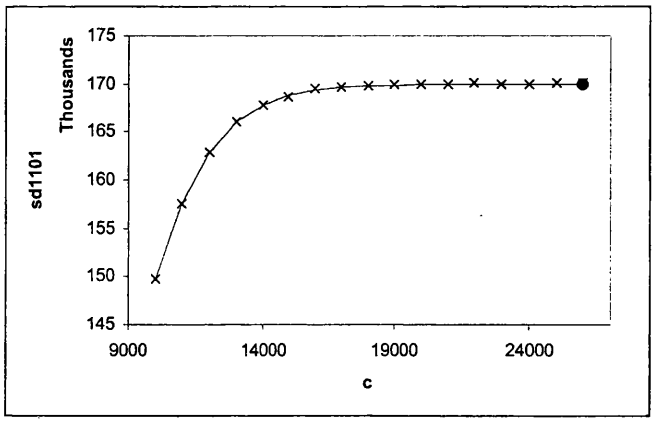

Figure 5.17: Theoretical and simulated values of $E\left[S^d_{1101}\right]$ versus c.

$$= \sum_{i=1}^{k} N_i \left(1 - q_{c,i,a}\right) \frac{x_i^2 \omega_i}{1+\omega_i},$$

where we recall $\omega_i = \left(\frac{c}{\exp(\alpha_b + \beta_b x_i)}\right)^{\tau}$; all other expectations are similarly dealt with.

Summary We can see that in each case the simulated data, on the whole, agrees with expected results. The figures show that the Type I expectation tends to the complete expectation when $c \to \infty$.

5.5.1 Expectations of the Score

Let us first consider $E\left[\frac{\partial l_B}{\partial a}\right]$. We note:

$$E\left[S^c_*\right] = \sum_{i=1}^{k} N_i \left(1 - q_{c,i,a}\right) \ln\left(1+\omega_i\right)$$

Hence, from (5.39), (5.40), (5.41) and (5.42):

$$\begin{aligned}
E\left[\frac{\partial l_B}{\partial a}\right] &= a^{-1} E\left[S_n\right] - E\left[S^d_*\right] - E\left[S^c_*\right] \\
&= a^{-1} \sum_{i=1}^{k} N_i q_{c,i,a} - \sum_{i=1}^{k} q_{c,i,a} N_i \left[\frac{1}{a} - \frac{\ln\left(1+\omega_i\right)}{\left(1+\omega_i\right)^a - 1}\right] \\
&\quad - \sum_{i=1}^{k} N_i \left(1 - q_{c,i,a}\right) \ln\left(1+\omega_i\right) \\
&= a^{-1} \sum_{i=1}^{k} N_i q_{c,i,a} - \sum_{i=1}^{k} N_i \left[\frac{1}{a} - \frac{\ln\left(1+\omega_i\right)}{\left(1+\omega_i\right)^a - 1} + \left(1 - q_{c,i,a}\right) \ln\left(1+\omega_i\right)\right] \\
&= 0,
\end{aligned}$$

$$
\begin{aligned}
E\left[\frac{\partial l_B}{\partial \alpha_b}\right] &= -\tau E\left[S_n\right] + \tau\left(a+1\right) E\left[S^d_{0101}\right] + \tau a E\left[S^c_{0101}\right] \\
&= -\tau \sum_{i=1}^{k} N_i q_{c,i,a} + \tau a \sum_{i=1}^{k} N_i \left(1 - q_{c,i,a}\right) \frac{\omega_i}{\left(1+\omega_i\right)} \\
&\quad + \tau\left(a+1\right) \sum_{i=1}^{k} \left\{ q_{c,i,a} N_i \left[\frac{a\omega_i^2}{2q_{c,i,a}} \left[\frac{2}{\left(a+1\right)\omega_i} \left\{ \frac{q_{c,i,a}}{\omega_i a} - \frac{1}{\left(1+\omega_i\right)^{a+1}} \right\} \right] \right] \right\} \\
&= \sum_{i=1}^{k} \left[a\tau N_i \omega_i \left\{ \frac{q_{c,i,a}}{a\omega_i} - \frac{1}{\left(1+\omega_i\right)^{a+1}} \right\} - \tau N_i q_{c,i,a} + a\tau N_i \left(1 - q_{c,i,a}\right) \frac{\omega_i}{1+\omega_i} \right] \\
&= -a\tau \sum_{i=1}^{k} N_i \frac{\omega_i}{\left(1+\omega_i\right)^{a+1}} \left\{ 1 - \left(1+\omega_i\right)^a + q_{c,i,a} \left(1+\omega_i\right)^a \right\} \\
&= -a\tau \sum_{i=1}^{k} N_i \frac{\omega_i}{\left(1+\omega_i\right)^{a+1}} \left\{ 1 - \left(1+\omega_i\right)^a + \left[1 - \left(1+\omega_i\right)^{-a} \right] \left(1+\omega_i\right)^a \right\} \\
&= 0;
\end{aligned}
$$

similarly

$$
\begin{aligned}
E\left[\frac{\partial l_B}{\partial \beta_b}\right] &= -\tau E\left[S_x\right] + \left(a+1\right)\tau E\left[S^d_{1101}\right] + a\tau E\left[S^c_{1101}\right] \\
&= -\tau \sum_{i=1}^{k} N_i x_i q_{c,i,a} + a\tau \sum_{i=1}^{k} N_i \left(1 - q_{c,i,a}\right) \frac{x_i \omega_i}{\left(1+\omega_i\right)} \\
&\quad + \left(a+1\right)\tau \sum_{i=1}^{k} \left\{ q_{c,i,a} N_i x_i \left[\frac{a\omega_i^2}{2q_{c,i,a}} \left[\frac{2}{\left(a+1\right)\omega_i} \left\{ \frac{q_{c,i,a}}{\omega_i a} - \frac{1}{\left(1+\omega_i\right)^{a+1}} \right\} \right] \right] \right\} \\
&= 0,
\end{aligned}
$$

by analogy with $E\left[\frac{\partial l_B}{\partial \alpha_b}\right]$. Finally, we have

$$
\begin{aligned}
E\left[\frac{\partial l_B}{\partial \tau}\right] &= \tau^{-1} E\left[S_n\right] + E\left[S_e\right] - \alpha_b E\left[S_n\right] - \beta_b E\left[S_x\right] - \left(a+1\right) E\left[S^d_{0111}\right] - a E\left[S^c_{0111}\right] \\
&= \tau^{-1} \sum_{i=1}^{k} N_i q_{c,i,a} + \sum_{i=1}^{k} \left\{ \alpha_b q_{c,i,a} N_i + \beta_b q_{c,i,a} N_i x_i + q_{c,i,a} N_i \left[\frac{a\omega_i}{\tau q_{c,i,a}} \times \right. \right. \\
&\quad \left. \left. \left(\left(\ln \omega_i - 1\right) f_1\left(1, a+1, -\omega_i\right) - \frac{\omega_i \left(a+1\right) f_2\left(2, a+2, -\omega_i\right)}{4} \right) \right] \right\} \\
&\quad - \alpha_b \sum_{i=1}^{k} N_i q_{c,i,a} - \beta_b \sum_{i=1}^{k} N_i x_i q_{c,i,a} - a \sum_{i=1}^{k} N_i \left(1 - q_{c,i,a}\right) \frac{\omega_i \ln \omega_i^{1/\tau}}{\left(1+\omega_i\right)} \\
&\quad - \left(a+1\right) \sum_{i=1}^{k} \left\{ q_{c,i,a} N_i \left\{ \frac{a\omega_i^2}{2\tau q_{c,i,a}} \left[\begin{array}{c} \frac{\left(2\ln\omega_i - 1\right)}{2} f_1\left(2, a+2, -\omega_i\right) \\ -\frac{\left(a+2\right)\omega_i}{9} f_2\left(3, a+3, -\omega_i\right) \end{array} \right] \right\} \right\}
\end{aligned}
$$

Now, using (5.21) and (5.23) we have

$$f_1(2,a+2,-\omega) - f_2(2,a+2,-\omega) = \frac{-2\omega(a+2)}{9} f_2(3,a+3,-\omega) \quad (5.54)$$

$$f_1(1,a+1,-\omega) - \frac{(a+1)\omega}{2} f_1(2,a+2,-\omega) = \frac{1}{(1+\omega)^{a+1}} \quad (5.55)$$

$$a\omega f_1(1,a+1,-\omega) = q_{c,i,a}, \quad (5.56)$$

and, through cancelling and grouping terms, we obtain

$$\begin{aligned} E\left[\frac{\partial l_B}{\partial \tau}\right] &= \sum_{i=1}^{k} \frac{a\omega_i N_i}{\tau}\left\{-\frac{1}{4}(a+1)\omega_i f_2(2,a+2,-\omega_i) + f_1(1,a+1,-\omega_i)(-1+\ln\omega_i)\right\} \\ &\quad -\sum_{i=1}^{k} \frac{a(a+1)\omega_i^2 N_i}{2\tau}\left\{\begin{array}{c} -\frac{1}{9}(a+2)\omega_i f_2(3,a+3,-\omega_i) \\ +\frac{1}{2} f_1(2,a+2,-\omega_i)(-1+2\ln\omega_i) \end{array}\right\} \\ &\quad -\sum_{i=1}^{k} \frac{a\omega_i N_i \ln\omega_i (1-q_{c,i,a})}{\tau(1+\omega_i)} + \sum_{i=1}^{k} \frac{N_i q_{c,i,a}}{\tau} \end{aligned} \quad (5.57)$$

Using (5.54) we see that we can write:

$$\begin{aligned} & -\frac{1}{9}(a+2)\omega f_2(3,a+3,-\omega) + \frac{1}{2} f_1(2,a+2,-\omega)(-1+2\ln\omega) \\ = \; & -\frac{1}{2} f_2(2,a+2,-\omega) + f_1(2,a+2,-\omega)\ln\omega \end{aligned}$$

Substituting into (5.57) and collecting like terms we have:

$$\begin{aligned} E\left[\frac{\partial l_B}{\partial \tau}\right] &= \sum_{i=1}^{k}\left\{ f_2(2,a+2,-\omega_i)\left[-\frac{a\omega_i^2 N_i(a+1)}{4\tau} - \frac{a(a+1)\omega_i^2 N_i}{2\tau}\left(-\frac{1}{2}\right)\right]\right. \\ &\quad + f_1(1,a+1,-\omega_i)(-1+\ln\omega_i)\frac{a\omega_i N_i}{\tau} - \frac{a(a+1)\omega_i^2 N_i}{2\tau} f_1(2,a+2,-\omega_i)\ln\omega_i \\ &\quad \left. -\frac{a\omega_i N \ln\omega_i (1-q_{c,i,a})}{\tau(1+\omega_i)} + \frac{N_i q_{c,i,a}}{\tau}\right\} \\ &= \sum_{i=1}^{k}\left\{\frac{a\omega_i N_i}{\tau}\ln\omega_i\left[f_1(1,a+1,-\omega_i) - \frac{(a+1)\omega_i}{2} f_1(2,a+2,-\omega_i)\right]\right. \\ &\quad \left. -a\omega_i f_1(1,a+1,-\omega_i)\frac{N_i}{\tau} - \frac{a\omega_i N_i \ln\omega_i}{\tau(1+\omega_i)}(1-q_{c,i,a}) + \frac{N_i q_{c,i,a}}{\tau}\right\} \\ &= \sum_{i=1}^{k}\left\{\frac{a\omega_i N_i \ln\omega_i}{\tau(1+\omega_i)^{a+1}} - \frac{N_i q_{c,i,a}}{\tau} - \frac{a\omega_i N_i \ln\omega_i (1-q_{c,i,a})}{\tau(1+\omega_i)} + \frac{N_i q_{c,i,a}}{\tau}\right\} \\ &= 0 \end{aligned}$$

using (5.55) and (5.56). Having established that the expectations of the score functions are zero, we now move on to consider the second derivatives.

5.5.2 Expectations of Second Derivatives

The expectations of individual functions detailed above use several hypergeometric functions. It would be helpful to be able to express them using as few hypergeometric functions as possible. Examination of these expectations in conjunction with (5.21) and (5.22) shows that we can write down all expectations using only marginally adapted forms of (5.23), (5.24) and (5.25). So, we now write

$$\begin{aligned}\kappa_{0,i} &= (1+\omega_i)^{-a-1} = p_{c,i,a+1},\\ \kappa_{1,i} &= f_1(1,a+1,-\omega_i) = \frac{q_{c,i,a}}{a\omega_i},\\ \kappa_{2,i} &= f_2(1,a+1,-\omega_i),\end{aligned}$$

and

$$\kappa_{3,i} = f_3(1,a+1,-\omega_i).$$

The expectations are clearly still cumbersome; to outline the logical steps required to arrive at these final forms would inevitably require many pages of rather long-winded algebra. Since, in essence, we are primarily interested in the inverse of the EFI matrix for a given set of parameter values, we include these theoretical results merely as information for the reader. As with the individual functions above, we also provide evidence of the agreement between these theoretical results and simulated values, using the same parameter values as before: $a = 2$, $\tau = 5$, $\alpha_b = 17.60139$, $\beta_b = -0.056282$, with the equally weighted sample size $N_1 = 1000$ and stress levels $(x_1, x_2, x_3) = (150, 170, 190)$. We can clearly see from the results below, that in each case there is good agreement between simulated values and their theoretical counterparts. The limiting agreement with the corresponding complete result is also evident. An analogous consideration for elements of the inverse of the EFI matrix shall be considered later, where we make further enquiries into the practical applications of the theory with various sample sizes. So, taking the expectation of (5.43) we have,

$$E\left[\frac{\partial^2 l_B}{\partial a^2}\right] = -a^{-2}\sum_{i=1}^{k} N_i q_{c,i,a};$$

the limiting behaviour of this function is presented in Figure 5.18. The expectation of (5.44) is,

$$\begin{aligned}E\left[\frac{\partial^2 l_B}{\partial \tau^2}\right] &= \frac{1}{2(a+1)(a+2)\tau^2}\sum_{i=1}^{k}(1+\omega_i)^{-2-a}\left\{-a(a+1)N_i\omega_i^2(21+5a\right.\\ &+2(a+1)[\ln\omega_i]^2\Big) - N_i(1+\omega_i)^a(-2a(a+1)(a+2)\times\\ &(-1+q_{c,i,a})\tau^2\omega_i\left[\ln\left(\omega_i^{1/\tau}\right)\right]^2 + (1+\omega_i)^2[-2a(1+a)\omega_i\ln\omega_i\times\\ &(2(-\kappa_{0,i}+\kappa_{1,i}-a\kappa_{1,i}+a\kappa_{2,i}) + a(\kappa_{0,i}-\kappa_{1,i})\ln\omega_i)\end{aligned}$$

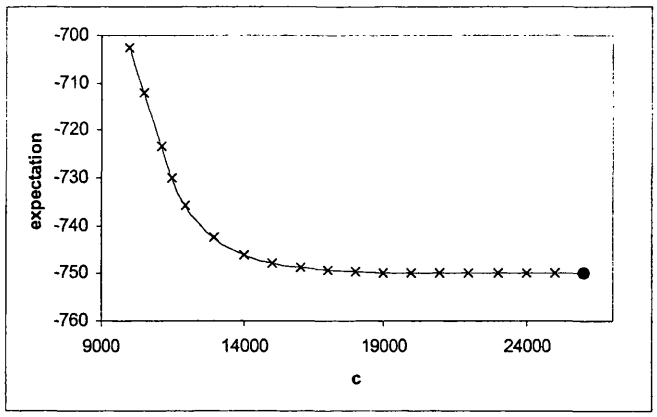

Figure 5.18: Theoretical and simulated values of $E\left[\frac{\partial l_B^2}{\partial a^2}\right]$ versus c..

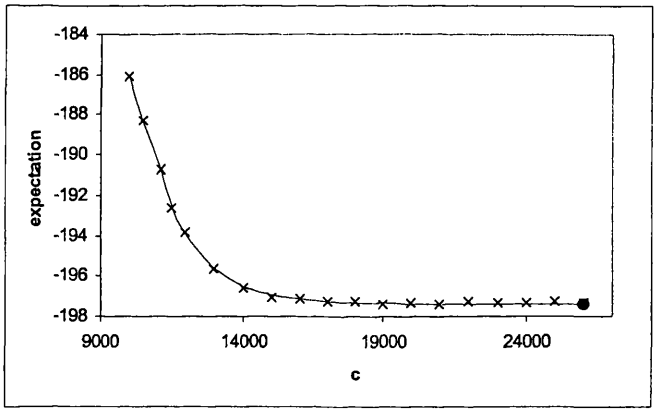

Figure 5.19: Theoretical and simulated values of $E\left[\frac{\partial l_B^2}{\partial \tau^2}\right]$ versus c

$$+2(a+1)(a+2)q_{c,i,a}+a(9(21+5a)\kappa_{0,i}-739\kappa_{1,i}+1054\kappa_{2,i}$$
$$-504\kappa_{3,i}+a(-179\kappa_{1,i}+(250-4a)\kappa_{2,i}+4(-29+a)\kappa_{3,i})\omega_i)])\},$$

upon simplification; the limiting behaviour of this function is presented in Figure 5.19. The expectation of (5.45) is given by,

$$E\left[\frac{\partial^2 l_B}{\partial \alpha_b^2}\right] = (a+2)^{-1}\sum_{i=1}^{k} N_i\tau^{-2}(1+\omega_i)^{-a-2}\{a\omega_i(2+a+\omega_i)$$
$$-(1+\omega_i)^a\begin{pmatrix} q_{c,i,a}\left(2+a-\left(-4+a^2\right)\omega_i+(2+a)\omega_i^2\right) \\ +a\omega_i\left(a+2\left(1+\kappa_{0,i}(1+\omega_i)^2-\kappa_{1,i}(1+\omega_i)^2\right)\right)\end{pmatrix}\Bigg\};$$

the limiting behaviour of this function is presented in Figure 5.20. The expectation of

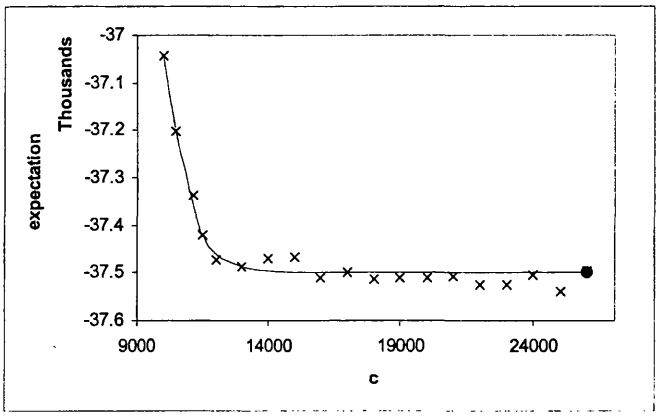

Figure 5.20: Theoretical and simulated values of $E\left[\frac{\partial l_B^2}{\partial \alpha_b^2}\right]$ versus c

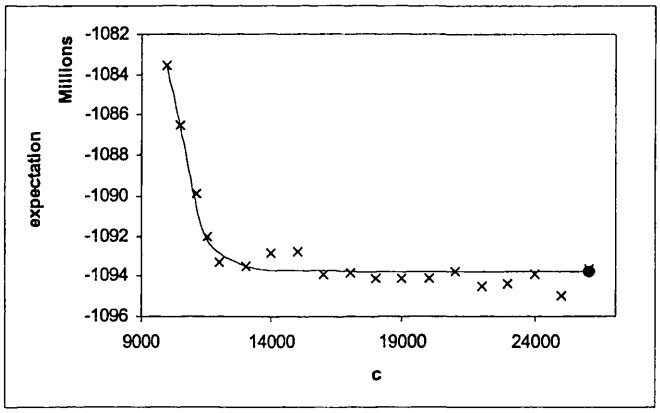

Figure 5.21: Theoretical and simulated values of $E\left[\frac{\partial l_B^2}{\partial \beta_b^2}\right]$ versus c

(5.46) is similarly defined as,

$$E\left[\frac{\partial^2 l_B}{\partial \beta_b^2}\right] = (a+2)^{-1}\sum_{i=1}^{k} x_i^2 N_i \tau^{-2}(1+\omega_i)^{-a-2}\left\{a\omega_i(2+a+\omega_i)\right.$$
$$\left. -(1+\omega_i)^a\begin{pmatrix} q_{c,i,a}\left(2+a-\left(-4+a^2\right)\omega_i+(2+a)\omega_i^2\right) \\ +a\omega_i\left(a+2\left(1+\kappa_{0,i}(1+\omega_i)^2-\kappa_{1,i}(1+\omega_i)^2\right)\right)\end{pmatrix}\right\};$$

the limiting behaviour of this function is presented in Figure 5.21. We also have

$$E\left[\frac{\partial^2 l_B}{\partial a \partial \tau}\right] = \sum_{i=1}^{k} N_i\omega_i\left(\frac{(q_{c,i,a}-1)\ln\left(\omega_i^{1/\tau}\right)}{\omega_i+1} - \frac{a\left(\kappa_{1,i}-\kappa_{2,i}+(\kappa_{1,i}-\kappa_{0,i})\ln\omega_i\right)}{\tau(a+1)}\right),$$

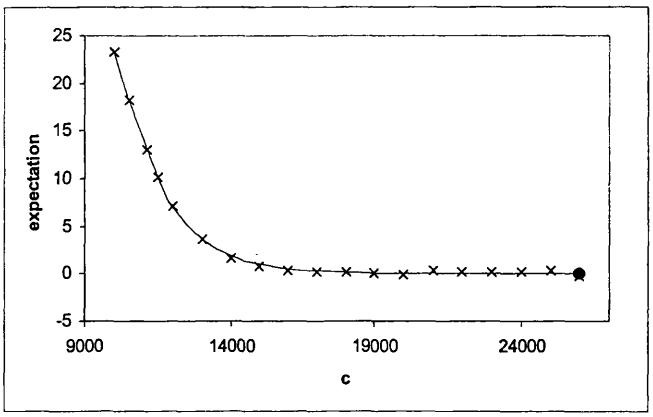

Figure 5.22: Theoretical and simulated values of $E\left[\frac{\partial l_B^2}{\partial a \partial \tau}\right]$ versus c

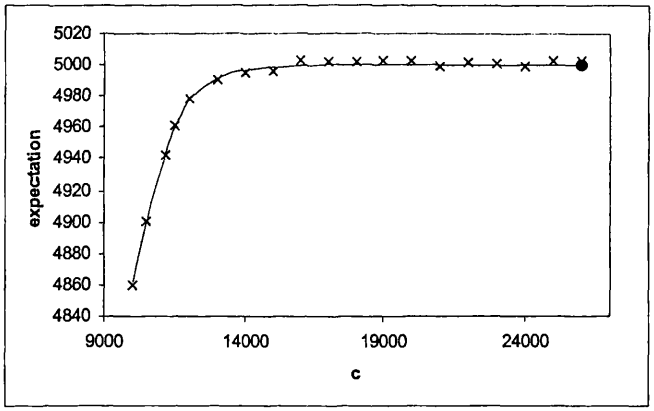

Figure 5.23: Theoretical and simulated values of $E\left[\frac{\partial l_B^2}{\partial a \partial \alpha_b}\right]$ versus c

upon taking the expectation of (5.47); the limiting behaviour of this function is presented in Figure 5.22. The expectation of (5.48) is

$$E\left[\frac{\partial^2 l_B}{\partial a \partial \alpha_b}\right] = (a+1)^{-1} \sum_{i=1}^{k} N_i \tau \left(1+\omega_i\right)^{-1-a} \left(-a\omega_i + (1+\omega_i)^a \left(\begin{array}{c} q_{c,i,a} + \omega_i \\ +a\omega_i - a q_{c,i,a} \omega_i \end{array} \right)\right);$$

the limiting behaviour of this function is presented in Figure 5.23. The expectation of (5.49) is

$$E\left[\frac{\partial^2 l_B}{\partial a \partial \beta_b}\right] = (a+1)^{-1} \sum_{i=1}^{k} x_i N_i \tau \left(1+\omega_i\right)^{-1-a} \left(-a\omega_i + (1+\omega_i)^a \left(\begin{array}{c} q_{c,i,a} + \omega_i \\ +a\omega_i - a q_{c,i,a} \omega_i \end{array} \right)\right);$$

the limiting behaviour of this function is presented in Figure 5.24. Also,

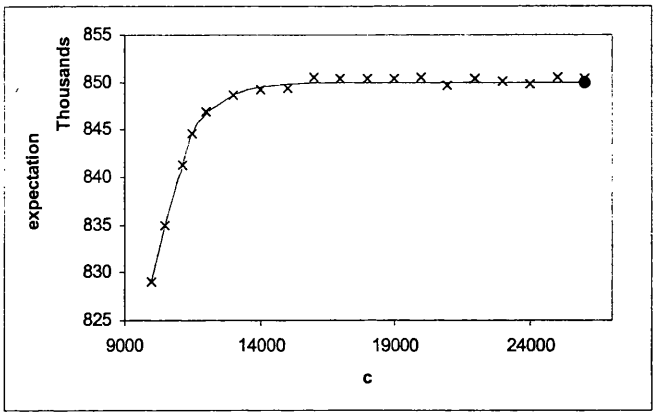

Figure 5.24: Theoretical and simulated values of $E\left[\frac{\partial l_B^2}{\partial a \partial \beta_b}\right]$ versus c

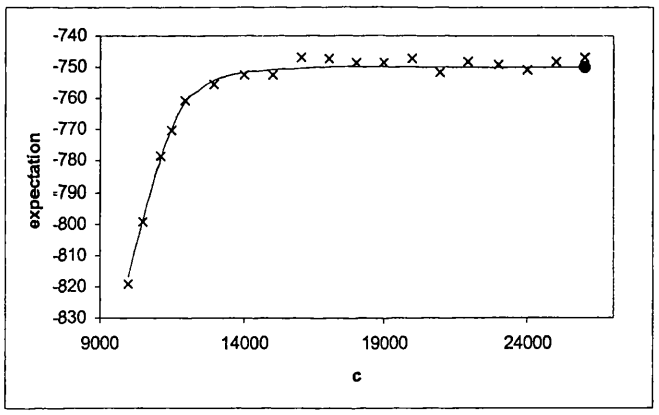

Figure 5.25: Theoretical and simulated values of $E\left[\frac{\partial l_B^2}{\partial \tau \partial \alpha_b}\right]$ versus c

$$E\left[\frac{\partial^2 l_B}{\partial \tau \partial \alpha_b}\right] = (a+2)^{-1}\sum_{i=1}^{k} aN_i\omega_i\left(1+\omega_i\right)^{-2-a}\left\{-2-a-2\omega_i-a\omega_i+a\omega_i+\omega_i\ln\omega_i\ln\omega_i\right.$$
$$\left.-\left(1+\omega_i\right)^{a+1}\left(\begin{array}{c}\left(\begin{array}{c}-2+\kappa_{1,i}+2q_{c,i,a}+\kappa_{1,i}\omega_i-\kappa_{0,i}\left(1+\omega_i\right)\\ +a\left(\begin{array}{c}-1+\kappa_{2,i}+q_{c,i,a}\\ +\kappa_{2,i}\omega_i-\kappa_{1,i}\left(1+\omega_i\right)\end{array}\right)\\ +a\left(\kappa_{0,i}-\kappa_{1,i}\right)\left(1+\omega_i\right)\ln\omega_i\end{array}\right)\\ +\left(1+\omega_i\right)^{-1}\left[\left(a+2\right)\left(-1+q_{c,i,a}\right)\tau\ln\left(\omega_i^{1/\tau}\right)\right]\end{array}\right)\right\},$$

with its limiting behaviour being presented in Figure 5.25, and

$$E\left[\frac{\partial^2 l_B}{\partial \tau \partial \beta_b}\right] = (a+2)^{-1}\sum_{i=1}^{k} ax_iN_i\omega_i\left(1+\omega_i\right)^{-2-a}\left\{-2-a-2\omega_i+\omega_i\ln\omega_i+a\omega_i\ln\omega_i\right.$$

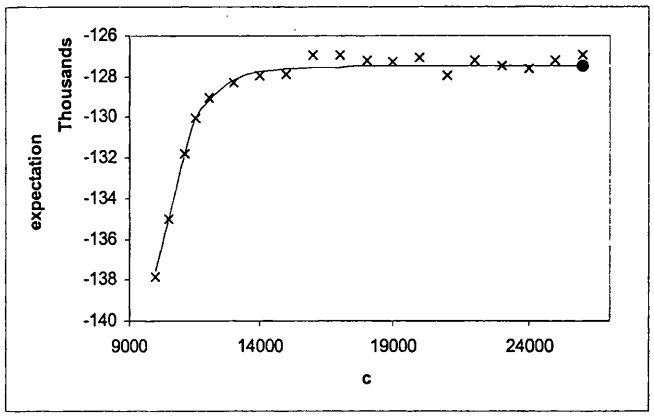

Figure 5.26: Theoretical and simulated values of $E\left[\frac{\partial l_B^2}{\partial \tau \partial \beta_b}\right]$ versus c

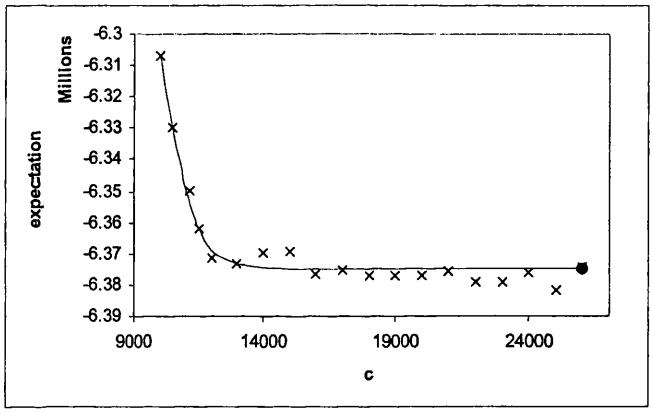

Figure 5.27: Theoretical and simulated values of $E\left[\frac{\partial l_B^2}{\partial \alpha_b \partial \beta_b}\right]$ versus c

$$-a\omega_i-(1+\omega_i)^{a+1}\left\{\left(\begin{array}{c}-2+\kappa_{1,i}+2q_{c,i,a}+\kappa_{1,i}\omega_i-\kappa_{0,i}(1+\omega_i)\\ +a\left(\begin{array}{c}-1+\kappa_{2,i}+q_{c,i,a}\\ +\kappa_{2,i}\omega_i-\kappa_{1,i}(1+\omega_i)\end{array}\right)\\ +a(\kappa_{0,i}-\kappa_{1,i})(1+\omega_i)\ln\omega_i\end{array}\right)+(1+\omega_i)^{-1}\left[(a+2)(-1+q_{c,i,a})\tau\ln\left(\omega_i^{1/\tau}\right)\right]\right\},$$

taking expectations of (5.50) and (5.51) respectively; the limiting behaviour of this function is presented in Figure 5.26. Finally, taking the expectation of (5.52) gives,

$$E\left[\frac{\partial^2 l_B}{\partial\alpha_b\partial\beta_b}\right]=a\tau\sum_{i=1}^{k}x_iN_i\omega_i\left(\frac{q_{c,i,a}-1}{(1+\omega_i)^2}+(a+2)^{-1}\tau\left(\begin{array}{c}-2\kappa_{0,i}+2\kappa_{1,i}-\frac{q_{c,i,a}(a+2)}{a\omega_i}\\ +(1+\omega_i)^{-2-a}(2+a+\omega_i)\end{array}\right)\right),$$

and the limiting behaviour of this final function is presented in Figure 5.27.

5.6 Weibull ALT Model

In section 2.1.2, we discussed an accelerated Weibull model for complete data sets. Watkins (1994) accounts for censored items in the model; although we ignored Type I censoring at the time. In this section we briefly highlight the necessary changes to the model that allow for the inclusion of censoring. The first point to note is

$$\begin{aligned}\Pr\left\{\text{item at } i^{th} \text{ stress level fails in } (0,c)\right\} &= G(c;B,\phi_i) = 1-\exp\left\{-\left(\frac{c}{\phi_i}\right)^B\right\} \\ &= 1-\exp(-C_i)\end{aligned}$$

where $C_i = (c/\phi_i)^B$; so that n_i - the number of failures at the i^{th} stress level - follows a Binomial distribution with parameters N_i - the total sample size at the i^{th} stress level - and $q_{c,i} = 1-\exp(-C_i)$. Then, the observed times to failure d_{ij} for the i^{th} stress level are from a truncated Weibull distribution with PDF,

$$\frac{B\left(\frac{d}{\phi_i}\right)^B d^{-1}\exp\left\{-\left(\frac{d}{\phi_i}\right)^B\right\}}{1-\exp(-C_i)}.$$

Hence, $Z_{ij} = \left(\frac{d_{ij}}{\phi_i}\right)^B$ are observations from a truncated negative exponential distribution with PDF

$$\frac{\exp(-Z)}{1-\exp(-C_i)}$$

for $0 \le Z \le C_i$. The likelihood for data obtained at the i^{th} stress level is then given by

$$\begin{aligned}&\left[\prod_{j=1}^{n_i} g(d_{ij};B,\phi_i)\right]\cdot\left[\prod_{j=1}^{m_i}\{1-G(c_{ij};B,\phi_i)\}\right] \\ =\ &\left[\prod_{j=1}^{n_i} B\left(\frac{d_{ij}}{\phi_i}\right)^B d_{ij}^{-1}\exp\left\{-\left(\frac{d_{ij}}{\phi_i}\right)^B\right\}\right]\left[\prod_{j=1}^{m_i}\exp\left\{-\left(\frac{c_{ij}}{\phi_i}\right)^B\right\}\right]\end{aligned}$$

so that, on recalling that $\phi_i = \exp(\alpha_w + \beta_w x_i)$, the log-likelihood for the entire data set is

$$\begin{aligned}l_W(B,\alpha_w,\beta_w) &= -\sum_{i=1}^{k}\sum_{j=1}^{n_i}\{\alpha_w+\beta_w x_i + B^{-1}\ln Z_{ij}\} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} Z_{ij} \\ &\quad + S_n \ln B + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln Z_{ij} - \sum_{i=1}^{k}\sum_{j=1}^{m_i} C_i\end{aligned}$$

It is now straightforward to obtain the first and second derivatives as

$$\begin{aligned}
\frac{\partial l_W}{\partial B} &= S_n B^{-1} + B^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln Z_{ij} - B^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Z_{ij}\ln Z_{ij} \\
&\quad -B^{-1}\sum_{i=1}^{k}\sum_{j=1}^{m_i} C_i \ln C_i \\
\frac{\partial l_W}{\partial \alpha_w} &= -S_n B + B\sum_{i=1}^{k}\sum_{j=1}^{n_i} Z_{ij} + B\sum_{i=1}^{k}\sum_{j=1}^{m_i} C_i \\
\frac{\partial l_W}{\partial \beta_w} &= -S_x B + B\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_i Z_{ij} + B\sum_{i=1}^{k}\sum_{j=1}^{m_i} x_i C_i
\end{aligned}$$

and,

$$\begin{aligned}
\frac{\partial^2 l_W}{\partial B^2} &= -S_n B^{-2} - B^{-2}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Z_{ij}(\ln Z_{ij})^2 - B^{-2}\sum_{i=1}^{k}\sum_{j=1}^{m_i} C_i(\ln C_i)^2 \\
\frac{\partial^2 l_W}{\partial \alpha_w^2} &= -B^2\sum_{i=1}^{k}\sum_{j=1}^{n_i} Z_{ij} - B^2\sum_{i=1}^{k}\sum_{j=1}^{m_i} C_i \\
\frac{\partial^2 l_W}{\partial \beta_w^2} &= -B^2\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_i^2 Z_{ij} - B^2\sum_{i=1}^{k}\sum_{j=1}^{m_i} x_i^2 C_i \\
\frac{\partial^2 l_W}{\partial B \partial \alpha_w} &= -S_n + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\{Z_{ij} + Z_{ij}\ln Z_{ij}\} + \sum_{i=1}^{k}\sum_{j=1}^{m_i}\{C_i + C_i \ln C_i\} \\
\frac{\partial^2 l_W}{\partial B \partial \beta_w} &= -S_x + \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_i\{Z_{ij} + Z_{ij}\ln Z_{ij}\} + \sum_{i=1}^{k}\sum_{j=1}^{m_i} x_i\{C_i + C_i \ln C_i\} \\
\frac{\partial^2 l_W}{\partial \alpha_w \partial \beta_w} &= -B^2\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_i Z_{ij} - B^2\sum_{i=1}^{k}\sum_{j=1}^{m_i} x_i C_i.
\end{aligned}$$

The algorithm used to converge on the maximum likelihood estimates of the parameters is given in the SAS program *weibull_alt*, and is presented in Appendix A.

Example 5.1.1.1 [Electric Motor Data]

We first fit a two-parameter Weibull model to data subsets in Table 5.1, making the minor modification to both algebra and algorithm, outlined in section 2.1, to accomodate censoring. We also present maximum likelihood estimates of parameters coming from a three-parameter Burr XII model, for data subsets. The algorithm employed is the same as that discussed in section 2.2, but of course now adapted to accomodate censoring. The SAS IML code is given in Appendix A as *weibull_ALT*. The maximum likelihood estimates for both models, together with the maximised log-likelihoods are given in Table 5.2. Naturally, we must omit

	Temperature/$^\circ C$		
	170	**190**	**220**
$\widehat{B}$	2.45663	1.85823	6.76171
$\widehat{\phi}$	5389.8982	2004.8873	592.29927
$\widehat{a}$	0.60712	11.43485	16.25742
$\widehat{\tau}$	4.07215	1.92794	6.96966
$\widehat{\theta}$	3607.389	6912.7018	878.78057
l_w	−73.92301	−75.13367	−59.69894
l_b	−73.52035	−75.16857	−59.7505
Δ_s	0.58939	−0.35085	−0.83045

Table 5.2: Two-parameter Weibull and Burr XII fit to subsets of Electric Motor data, Nelson (1990)

results for stress $150^\circ C$ since complete censoring observed there renders this subset unusable in the present context. We can see quite a variation in the parameter estimates across stress levels and whilst $\widehat{\theta}$ for x_i =220$^\circ C$ and 170$^\circ C$ seem appropriate, the estimate of $\widehat{\theta}$ for x_i =190$^\circ C$ may be considered to be a little high (purely on the basis of the observed failure times); values for $\widehat{\phi}$ seem appropriate at all stress levels. Clearly, in terms of maximised log-likelihood, the Weibull model is a better fit than the Burr XII at the two highest stress settings, while the Burr XII is favoured to the Weibull at $x_i = 170^\circ C$, and this is endorsed by the corresponding values of Δ_s, the censored form of which is given by Watkins (2001).

Moving on to fit a Weibull ALT model to the electric motor data, we obtain the parameter estimates $\widehat{B} = 2.9930846, \widehat{\alpha}_w = 16.874116$ and $\widehat{\beta}_w = -0.048093$; and hence $\widehat{\phi}_1 = 15680.38, \widehat{\phi}_2 = 5992.75, \widehat{\phi}_3 = 2290.31$ and $\widehat{\phi}_4 = 541.12$, with associated log-likelihood, $l_W = -217.15$. We see the estimates of the $\widehat{\phi}_i$ are in keeping with the data values at each stress level, while the single estimate $\widehat{B}$ here is in the vicinity of the separate estimates in Table 5.2. Fitting a Burr XII ALT model to the data set, we obtain the parameter estimates $a = 40231077, \tau = 2.9930846, \alpha_b = 22.724406$ and $\beta_b = -0.048093$ with associated log-likelihood $l_B = -217.15$. Hence, estimates of the scale parameters are $(\theta_1, \theta_2, \theta_3, \theta_4) = (5446348.81, 2081493.03, 795507.85, 187952.77)$. As with example 2.1.1.3 (electrode data), the value of a was not actually converged upon. There, l_B was seen to increase with a, while the scale parameters at each stress level tended to be overestimated..As such, the Burr XII ALT parameter values stated here are not maximum likelihood estimates; consequently we omit the carets over the letters. However, we see that the value of τ is in keeping with the separate estimates observed in Table 5.2. We also note that the ALT estimate of τ is the same as the ALT value of $\widehat{B}$, to seven decimal places, while $\beta_b = \widehat{\beta}_w$, to six decimal places and $l_W = l_B$, to two decimal places. In Chapter four, we went on to obtain a negative value of Δ_s for the electrode data set, indicating that the Weibull ALT model was a superior fit than the Burr XII ALT model. Naturally, in light of this evidence we would expect to reach the same conclusion with the electric motor data set here. We shall consider this further below, when we accomodate censoring in our derivation of Δ_s.

5.7 Practical Implications

In this section we shall consider elements of the inverse of the EFI matrix for the Burr XII ALT model. We look to compare simulated results with their theoretical counterparts for increasing censoring time c, and for varying sample sizes N_1. Again, we shall take the parameter values $a = 2$, $\tau = 3$, $\alpha_b = 17.60139$, $\beta_b = -0.056282$ and $(x_1, x_2, x_3) = (150, 170, 190)$. Our concern here lies solely with those elements lying on the diagonal of the inverse of the EFI matrix since, asymptotically, these are the variances of the maximum likelihood estimates of the model parameters. Now, we can obtain the theoretical approximations to the variances (or, more specifically in our case, the standard deviations) in much the same way as in section 3.2.1. That is, we can formulate the Type I EFI matrix, extracting the sample size, N_1, as a factor. For instance, with the parameter values above, together with a stopping time of $c = 1500$, we find the inverse of the EFI matrix as

$$\frac{1}{N_1}\begin{bmatrix} 131.619 & & & \\ -33.4333 & 13.9033 & & \\ 44.5396 & -22.2606 & 98.775 & \\ -0.0724 & 0.07442 & -0.4708 & 0.0024 \end{bmatrix}.$$

Setting $N_1 = 30$, for example, and calculating the square root of the diagonal elements of this matrix, we have the first diagonal element (the approximation to the standard deviation of $\widehat{a}$) as 2.09459, which is plotted as the first figure on the continuous line in the Figure 5.28 below.

We can see from the Figures 5.28 to 5.43 that for smaller sample sizes there is a considerable discrepancy between the simulated ($\times$) and theoretical (continuous line) standard deviations, while the percentages of valid cases (in the usual sense) are represented by the broken line. For comparison, we also present corresponding results for the complete case from Chapter three, see Table 3.2, where the theoretical complete result is represented by a single point ($\cdot$), the percentage of valid cases is represented by a triangle ($\blacktriangle$) and the simulated standard deviation of the parameter estimate is represented by a diamond ($\blacklozenge$). We see in Figure 5.28 that for the variable a, there is some disagreement between the simulated complete result from Chapter three, and the simulated results observed here for large c. However, this difference decreases as the sample size increases. Particularly marked for the variable a, is the difference between the simulated complete result ($\blacklozenge$) and the theoretical complete result ($\cdot$) for small sample sizes. We see this difference decreases as the sample size increases, to such an extent that for $N_1 \geq 1000$, Figures 5.30 and 5.31, we find it difficult to distinguish between the two results. For the other three variables we see that the difference between these results is negligible, even for the smallest sample size considered, see, for example, Figures 5.32, 5.36 and 5.40. Note also, for $N_1 = 1000$ and 3000, 100% of cases considered were valid, for all values of c considered. Consequently, percentages have been omitted from Figures 5.30 and 5.31. From an experimental point of view, it was observed

that all items had failed for c approximately equal to 13000. Therefore, with no further information being obtainable for higher values of c, we terminate the figures at this point. A characteristic feature of the figures, particularly noticeable for $\widehat{\tau}$ - Figures 5.32 to 5.35 - is a marked 'kink' for $c \backsimeq 3500$. Since, for larger values of N_1, simulated values exhibit the same pattern, it seems that this 'temporary-flattening-out' of the standard deviations, is due to some combination of the choice of parameter values and failure time. Nevertheless, the overall shape, with standard deviations decreasing for increasing stopping time, is in keeping with intuition.

For $\widehat{a}$, we see a large disparity between simulated and theoretical standard deviations for $N_1 = 30$, see Figure 5.28, although this difference diminishes as the sample size increases, such that, at $N_1 = 1000$, Figure 5.30, the agreement between the two sets of results is actually quite good. For the other three model parameters, the disagreement between simulated and theoretical values is less pronounced than for $\widehat{a}$, at smaller sample sizes. At $N_1 = 1000$, there is almost complete conformity between the two sets of results for $\widehat{\tau}$, see Figure 5.34, while for $\widehat{\alpha}_b$ and $\widehat{\beta}_b$, the agreement is good even at the smallest sample size considered, see Figures 5.36 and 5.40. In each figure, we can clearly see the Type I theoretical results tending to their complete counterparts, as c gets larger, as expected. In addition, as the sample size increases, the standard deviation of a given parameter estimate for a given c, decreases, also in keeping with intuition. Naturally, the longer the experiment has to run, the more failures are observed and more information is known about the parameter estimates, thereby reducing the standard deviation in those estimates.

Of course, this is only one example, the results of which extend to sixteen figures. Naturally, with the inclusion of the stopping time c as another parameter, to consider as many simulations as in Chapter three would require hundreds of figures. What we must bear in mind, however, is that with censoring, theoretical results tie-in asymptotically with theoretical results from the complete case; whereas with the original complete results, there is no way of making any further theoretical validations. Consequently, we feel that, while complete results need to be rigorously endorsed through simulations, censored results require fewer examples to illustrate the point. As such, we limit ourselves to consideration of this single example.

5.8 Limiting agreement with complete data

The previous section shows, for a particular example, good agreement between simulations and Type I expected results. The figures also suggested that, as $c \to \infty$, the Type I expected results tended to their complete counterparts. In this section we look to prove this limiting agreement from a theoretical standpoint. Firstly, we shall establish some groundwork necessary for our proofs. We then move on to consider our particular application and examine the fundamental expressions that are the basis for all further expectations. We observed above that we can write the expectations of all second derivatives of the Burr XII

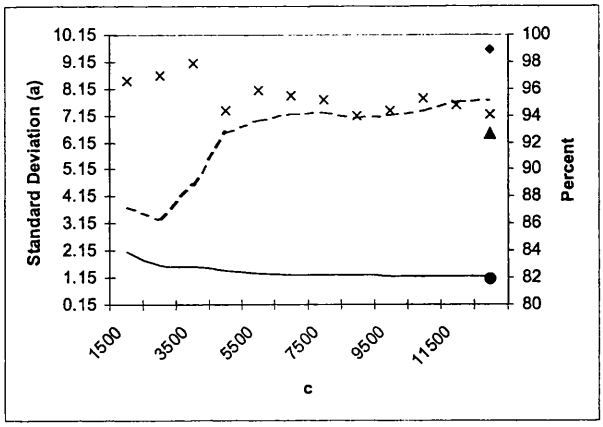

Figure 5.28: Theoretical and simulated standard deviation of $\widehat{a}$ versus c, for $N_1 = 30$.

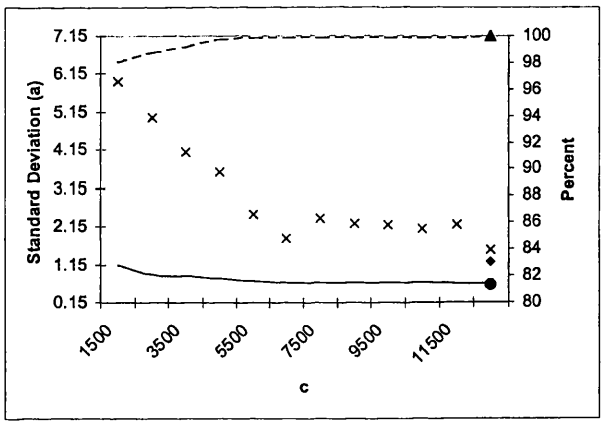

Figure 5.29: Theoretical and simulated standard deviation of $\widehat{a}$ versus c, for $N_1 = 100$.

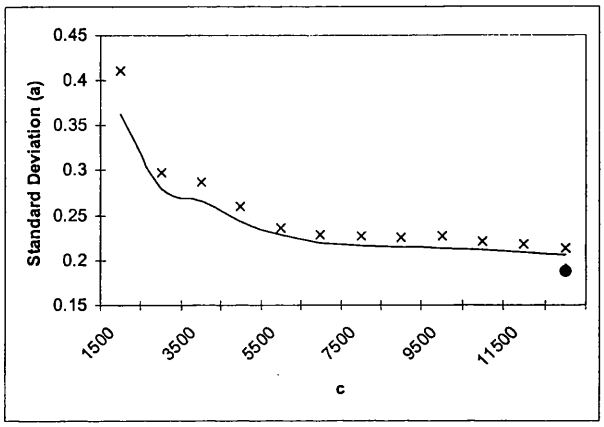

Figure 5.30: Theoretical and simulated standard deviation of $\widehat{a}$ versus c, for $N_1 = 1000$.

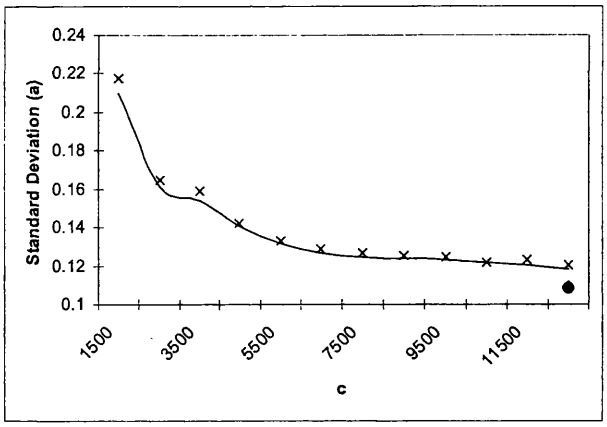

Figure 5.31: Theoretical and simulated standard deviation of $\widehat{a}$ versus c, for $N_1 = 3000$.

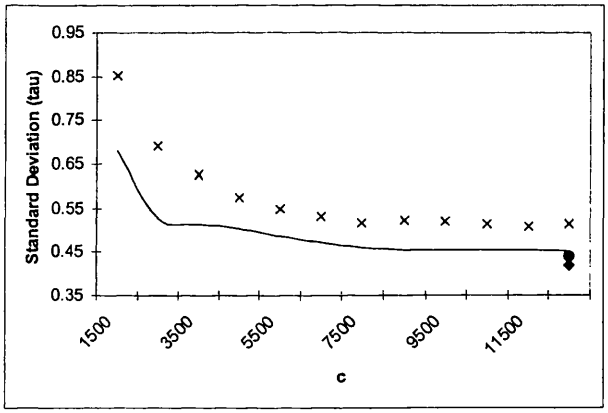

Figure 5.32: Theoretical and simulated standard deviation of $\widehat{\tau}$ versus c, for $N_1 = 30$.

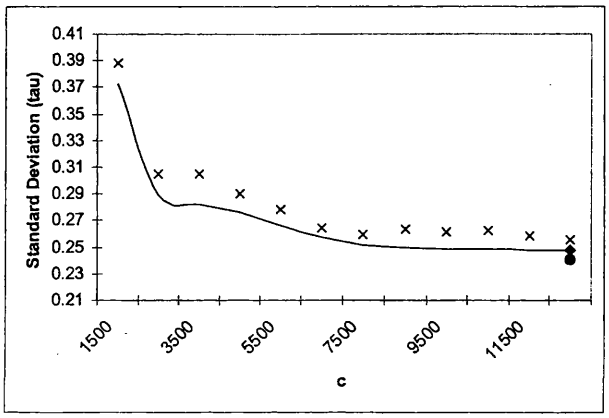

Figure 5.33: Theoretical and simulated standard deviation of $\widehat{\tau}$ versus c, for $N_1 = 100$.

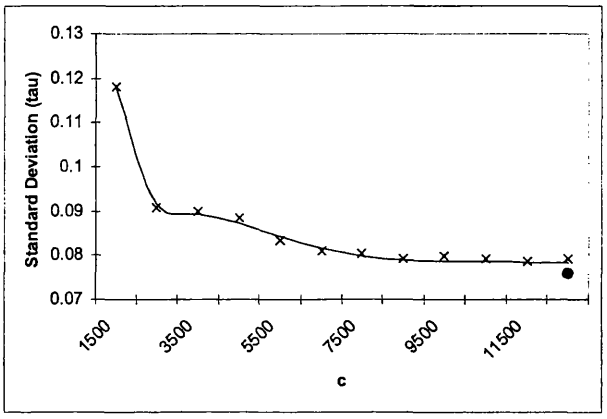

Figure 5.34: Theoretical and simulated standard deviation of $\widehat{\tau}$ versus c, for $N_1 = 1000$.

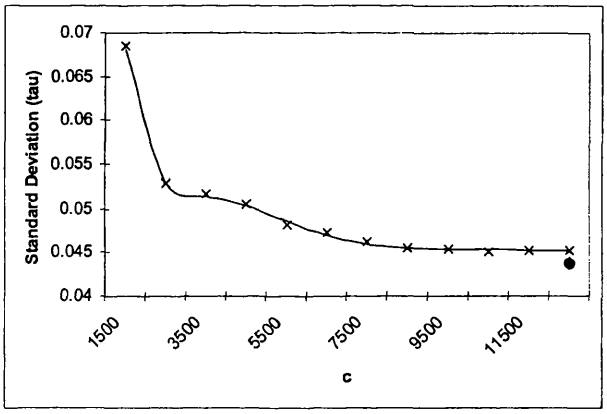

Figure 5.35: Theoretical and simulated standard deviation of $\widehat{\tau}$ versus c, for $N_1 = 3000$.

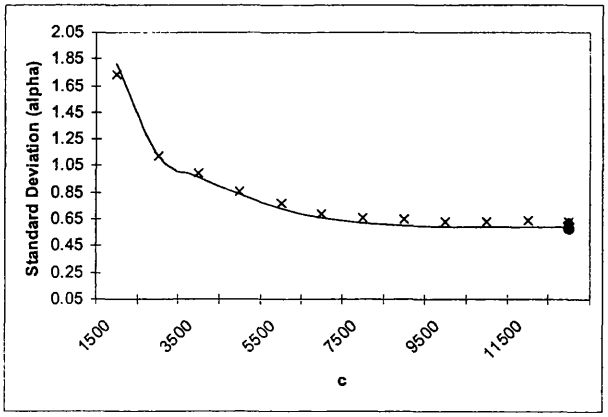

Figure 5.36: Theoretical and simulated standard deviation of $\widehat{\alpha}_b$ versus c, for $N_1 = 30$.

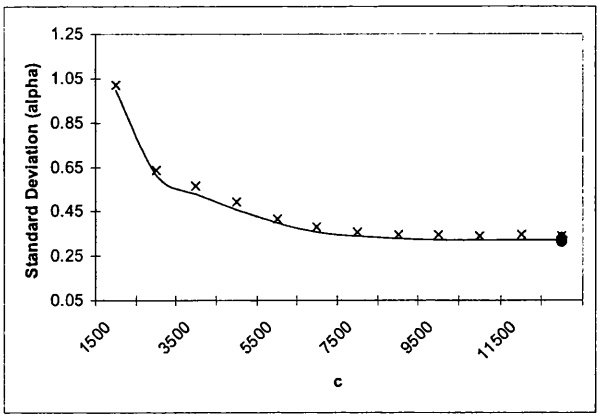

Figure 5.37: Theoretical and simulated standard deviation of $\widehat{\alpha}_b$ versus c, for $N_1 = 100$.

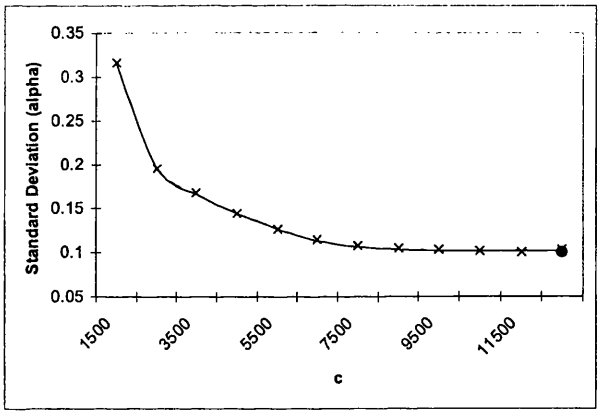

Figure 5.38: Theoretical and simulated standard deviation of $\widehat{\alpha}_b$ versus c, for $N_1 = 1000$.

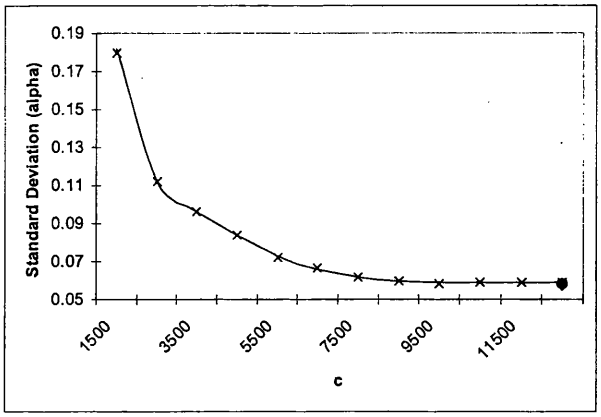

Figure 5.39: Theoretical and simulated standard deviation of $\widehat{\alpha}_b$ versus c, for $N_1 = 3000$.

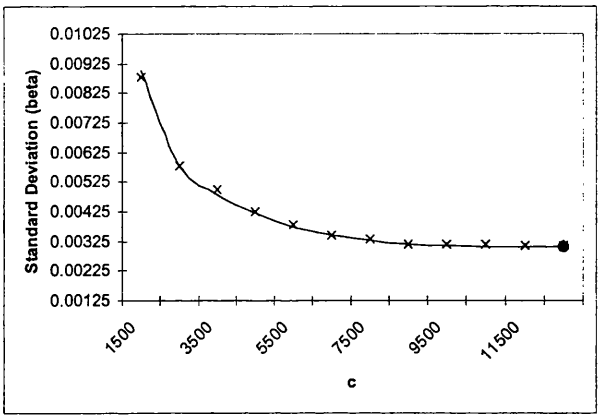

Figure 5.40: Theoretical and simulated standard deviation of $\widehat{\beta}_b$ versus c, for $N_1 = 30$.

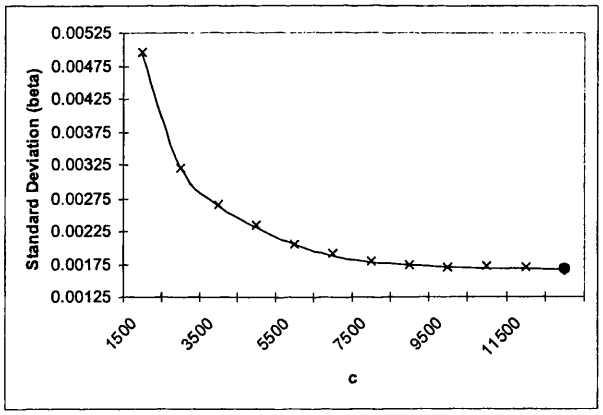

Figure 5.41: Theoretical and simulated standard deviation of $\widehat{\beta}_b$ versus c, for $N_1 = 100$.

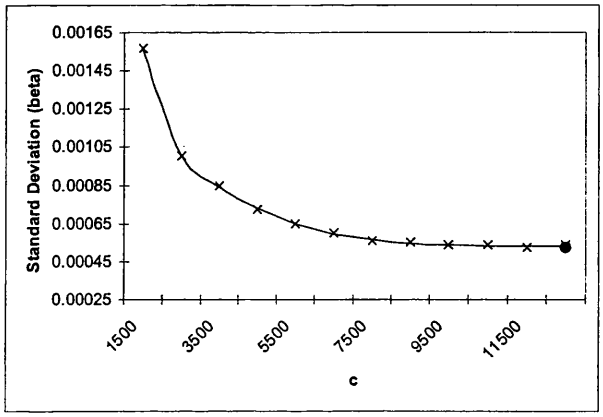

Figure 5.42: Theoretical and simulated standard deviation of $\widehat{\beta}_b$ versus c, for $N_1 = 1000$.

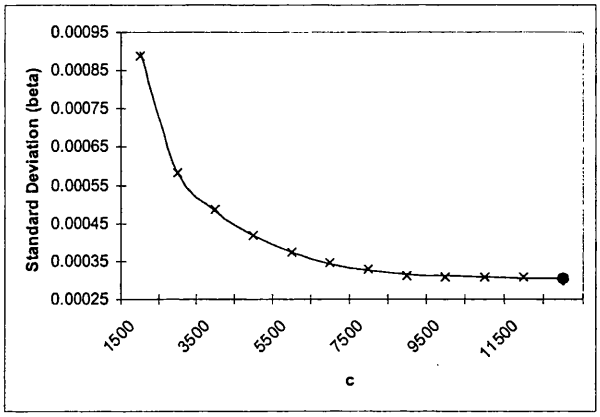

Figure 5.43: Theoretical and simulated standard deviation of $\widehat{\beta}_b$ versus c, for $N_1 = 3000$.

ALT model using only the terms

$$\begin{aligned} \kappa_0 &= (1-z)^{-a-1}, \\ \kappa_1 &= f_1(1, a+1, z) = \frac{-q_{c,a}}{za}, \\ \kappa_2 &= f_2(1, a+1, z), \end{aligned}$$

and

$$\kappa_3 = f_3(1, a+1, z),$$

where $z = -\omega = -c^\tau$, and the subscript i - relating to the stress factor - is omitted here for the sake of brevity. Therefore we shall examine the limiting properties of these terms first. Clearly, as $c \to \infty$,

$$\kappa_0 \to 0, \tag{5.58}$$

while, since $q_{c,a} \to 1$ in the same limit, we clearly have

$$\omega\kappa_1 \to \frac{1}{a}. \tag{5.59}$$

Consideration of κ_2 and κ_3 will require results from complex analysis; in particular the consideration of poles, residues and contour integration.

Consideration of κ_2 Let us begin by examining $\kappa_2 = F_{3,2}(\{1, 1, a+1\}, \{2, 2\}, z)$, a particular example of the form $F_{3,2}(\{t, t, u\}, \{t+1, t+1\}, z)$. From paragraph 15.3.2 in Abramowitz and Stegun (1972), we can express

$$\frac{\Gamma(t)^2 \Gamma(u)}{\Gamma(t+1)^2} F_{3,2}(\{t, t, u\}, \{t+1, t+1\}, z)$$

as

$$\begin{aligned}&\tfrac{1}{2\pi i}\int_{-i\infty}^{i\infty}\tfrac{\Gamma(t+s)\Gamma(t+s)\Gamma(u+s)}{\Gamma(t+1+s)\Gamma(t+1+s)}\Gamma(-s)(-z)^s\,ds\\&=\tfrac{1}{2\pi i}\int_{-i\infty}^{i\infty}\tfrac{\Gamma(u+s)}{(t+s)(t+s)}\Gamma(-s)(-z)^s\,ds\\&=\tfrac{1}{2\pi i}\int_{-i\infty}^{i\infty}\tfrac{\Gamma(u+s)}{(t+s)^2}\Gamma(-s)(-z)^s\,ds\end{aligned}\tag{5.60}$$

Here, $\frac{1}{(t+s)^2}$ has a *pole of order two*, (double pole), at $s=-t$, $\Gamma(u+s)$ has a *pole of order one*, (single pole), at $s=-u,-u-1,-u-2,\cdots$ and, if $t-u$ is not an integer, then all poles are simple; see Bak and Newman (1982, p97). Bak and Newman (1982, p107) also outlines the manner in which we can obtain the *residue* of a function, $f(z)$ say, with a pole of order k at z_0; this residue is given by

$$\frac{1}{(k-1)!}\frac{d^{k-1}}{dz^{k-1}}\left[(z-z_0)^k f(z)\right]\Big|_{z=z_0}.$$

The merit of calculating the residue comes from the application of Cauchy's Residue Theorem, which, in essence, states that for a function $f(z)$,

$$\int_C f(z)\,dz = 2\pi i \times \text{ sum of the residues of any poles within } C,$$

where C is a closed curve that fails to intersect any of the poles; see Bak and Newman (1982, p110) for further discussion. Our procedure is as follows:

1) Include the required integral as part of a contour integral.
2) Identify the poles inside the contour
3) Evaluate the residuals at the poles
4) Obtain the integral as the sum of residues.

We can best illustrate a suitable choice for a contour C for (5.60), graphically. We first write (5.60) as

$$\lim_{w\to\infty}\frac{1}{2\pi i}\int_{-iw}^{iw}\frac{\Gamma(u+s)}{(t+s)^2}\Gamma(-s)(-z)^s\,ds$$

and this is represented in the first part of Figure 5.44 and, slightly adjusted to avoid the origin, in the second part. The integral is part of the entire contour C shown in Figure 5.45. So,

$$\int_{iw}^{-iw}=\int_C-\int_{-iw}^{-iw-w-\frac{1}{2}}-\int_{-iw-w-\frac{1}{2}}^{iw-w-\frac{1}{2}}-\int_{iw-w-\frac{1}{2}}^{iw}.$$

For this particular function, the final three integrals disappear in the limit $w\to\infty$, leaving

$$\int_{-i\infty}^{i\infty}=\lim_{w\to\infty}-\int_C=-\lim_{w\to\infty}\int_C;$$

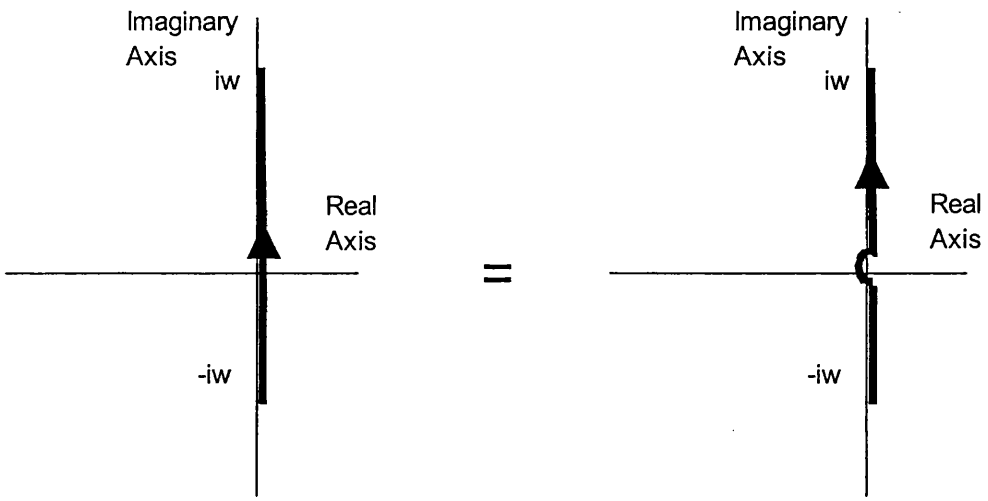

Figure 5.44: Representation of a contour integral in the complex plane. In the second figure, we illustrate how the path of integration avoids the origin, thereby failing to intersect any of the poles.

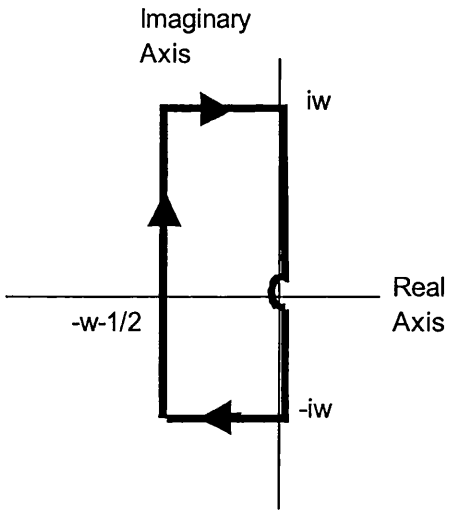

Figure 5.45: Contour integral in the complex plane.

see Copson (1935, p254). However, our concern lies with an expression of the form

$$\frac{\Gamma(t+s)\Gamma(t+s)\Gamma(u+s)}{\Gamma(t+1+s)\Gamma(t+1+s)}\Gamma(-s)(-z)^s,$$

see (5.60), and we see the poles of this expression, within C, are at

$$\left.\begin{array}{c} t+s \\ u+s \\ t+1+s \end{array}\right\} = 0,-1,-2,\cdots,$$

from which,

$$\frac{\Gamma(t+s)\Gamma(t+s)\Gamma(u+s)}{\Gamma(t+1+s)\Gamma(t+1+s)}\Gamma(-s)(-z)^s$$

has simple poles at $s=-t$ and $s=-u,-u-1,-u-2,\cdots$ and double poles at $s=-t-1,-t-2,\cdots$. Therefore,

$$\int_{-i\infty}^{i\infty} = -2\pi i \times \text{sum of the residues at the poles within } C.$$

Having identified the poles in the contour we can now move on to evaluate the residues at the poles. Considering the final form of the integrand in (5.60), the residue at $s=-t$ is given by

$$\begin{aligned} & \frac{d}{ds}\left[(t+s)^2\frac{\Gamma(t+s)^2\Gamma(u+s)}{\Gamma(t+1+s)^2}\Gamma(-s)(-z)^s\right]_{s=-t} \\ = & \frac{d}{ds}\left[\Gamma(u+s)\Gamma(-s)(-z)^s\right]_{s=-t} \end{aligned}$$

$$\begin{aligned} = & \Gamma'(u+s)\Gamma(-s)(-z)^s+\Gamma(u+s)\left(-\Gamma'(-s)\right)(-z)^s+\Gamma(u+s)\Gamma(-s)(-z)^s\ln(-z) \\ = & \Gamma'(u+s)\Gamma(-s)(-z)^s-\Gamma(u+s)\Gamma'(-s)(-z)^s+\Gamma(u+s)\Gamma(-s)(-z)^s\ln(-z) \end{aligned}$$

At the pole $s=-t$, this becomes,

$$\begin{aligned} & \Gamma'(u-t)\Gamma(t)(-z)^{-t}-\Gamma(u-t)\Gamma'(t)(-z)^{-t}+\Gamma(u-t)\Gamma(t)(-z)^{-t}\ln(-z) \\ & (-z)^{-t}\Gamma(u-t)\left\{\frac{\Gamma'(u-t)}{\Gamma(u-t)}\Gamma(t)-\Gamma'(t)+\Gamma(t)\ln(-z)\right\} \\ = & (-z)^{-t}\Gamma(u-t)\left\{\psi(u-t)\Gamma(t)-\Gamma'(t)+\Gamma(t)\ln(-z)\right\} \\ = & (-z)^{-t}\Gamma(u-t)\Gamma(t)\left\{\psi(u-t)-\psi(t)+\ln(-z)\right\}. \end{aligned}$$

At the pole $s=-u-w$, the residue of $\Gamma(u+s)$ is

$$\frac{(-1)^w}{w!};$$

see Abramowitz and Stegun (1972,p255). Therefore, the residue at $s = -u - w$ is given by

$$\frac{(-1)^w}{w!}\frac{\Gamma(u+w)}{(t-u-w)^2}(-z)^{-u-w},$$

and it follows that the sum of all residues is then

$$(-z)^{-t}\Gamma(u-t)\Gamma(t)\{\psi(u-t)-\psi(t)+\ln(-z)\}+(-z)^{-u}\sum_{w=0}^{\infty}\frac{(-1)^w}{w!}\frac{\Gamma(u+w)}{(t-u-w)^2}(-z)^{-w}.$$

In κ_2, $t = 1$ and $u = a+1$. Also, as $z \to -\infty$, the summation in this expression disappears. Therefore, we can write (5.60) as,

$$\begin{aligned}\Gamma(a+1)\kappa_2 &= (-z)^{-1}\Gamma(a)\Gamma(1)\{\psi(a)-\psi(1)+\ln(-z)\}\\ (-z)a\kappa_2 &= \Gamma(1)\{\psi(a)-\psi(1)+\ln(-z)\}\\ (-z)a\kappa_2-\ln(-z) &= \psi(a)-\psi(1)=\psi(a)+\gamma\end{aligned}$$

and so, as $z \to -\infty$, - and hence, as $c \to \infty$ - we have

$$\omega a\kappa_2 - \ln\omega \to \psi(a)+\gamma \tag{5.61}$$

and we now have the result in a form involving terms that we recognise from the complete case of Chapter three.

Consideration of κ_3 Finally, we have $\kappa_3 = F_{4,3}(\{1,1,1,a+1\},\{2,2,2\},z)$, which can be expressed as

$$\begin{aligned}\frac{\Gamma(t)^3\Gamma(u)}{\Gamma(t+1)^3}\kappa_3 &= \frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{\{\Gamma(t+s)\}^3\Gamma(u+s)}{\{\Gamma(t+1+s)\}^3}\Gamma(-s)(-z)^s\,ds \qquad (5.62)\\ &= \frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{\Gamma(u+s)}{(t+s)^3}\Gamma(-s)(-z)^s\,ds\end{aligned}$$

Now, by analogy with κ_2, we note that the sum of the residues at the poles $s = -u-w$ will be zero in the limit $z \to -\infty$. Therefore we shall concentrate on the poles of $(t+s)^{-3}$ and shall see that consideration of this term alone is sufficient to establish the required result. So, within C, $(t+s)^{-3}$ has a triple pole at $s = -t$. The residue is then given by

$$\frac{1}{2!}\frac{d^2}{ds^2}\left[\frac{(s+t)^3}{(s+t)^3}\Gamma(u+s)\Gamma(-s)(-z)^s\right]_{s=-t} \tag{5.63}$$

The first derivative from (5.63) is given by

$$\Gamma'(u+s)\Gamma(-s)(-z)^s-\Gamma(u+s)\Gamma'(-s)(-z)^s+\Gamma(u+s)\Gamma(-s)(-z)^s\ln(-z),$$

while the second derivative from (5.63) is

$$\begin{aligned}
&\Gamma''(u+s)\Gamma(-s)(-z)^s - \Gamma'(-s)\Gamma'(u+s)(-z)^s + \Gamma'(u+s)\Gamma(-s)(-z)^s\ln(-z) \\
&- \left[\Gamma'(u+s)\Gamma'(-s)(-z)^s - \Gamma(u+s)\Gamma''(-s)(-z)^s + \Gamma(u+s)\Gamma'(-s)(-z)^s\ln(-z)\right] \\
&+\ln(-z)\left[\begin{array}{c}\Gamma'(u+s)\Gamma(-s)(-z)^s - \Gamma(u+s)\Gamma'(-s)(-z)^s \\ +\Gamma(u+s)\Gamma(-s)(-z)^s\ln(-z)\end{array}\right].
\end{aligned}$$

The residue then becomes,

$$\begin{aligned}
&\frac{1}{2}(-z)^s\Gamma(u+s)\left[\frac{\Gamma''(u+s)}{\Gamma(u+s)}\Gamma(-s) - \Gamma'(-s)\frac{\Gamma'(u+s)}{\Gamma(u+s)} + \frac{\Gamma'(u+s)}{\Gamma(u+s)}\Gamma(-s)\ln(-z)\right. \\
&-\frac{\Gamma'(u+s)}{\Gamma(u+s)}\Gamma'(-s) + \Gamma''(-s) - \Gamma'(-s)\ln(-z) + \ln(-z)\frac{\Gamma'(u+s)}{\Gamma(u+s)}\Gamma(-s) \\
&\left.-\ln(-z)\Gamma'(-s) + \ln(-z)^2\Gamma(-s)\right] \\
= &\frac{1}{2}(-z)^s\Gamma(u+s)\left[\frac{\Gamma''(u+s)}{\Gamma(u+s)}\Gamma(-s) - \Gamma'(-s)\psi(u+s) + \psi(u+s)\Gamma(-s)\ln(-z)\right. \\
&-\psi(u+s)\Gamma'(-s) + \Gamma''(-s) - \Gamma'(-s)\ln(-z) + \ln(-z)\psi(u+s)\Gamma(-s) \\
&\left.-\ln(-z)\Gamma'(-s) + \ln(-z)^2\Gamma(-s)\right] \\
= &\frac{1}{2}(-z)^s\Gamma(u+s)\left[\psi'(u+s)\Gamma(-s) + \Gamma(-s)\psi(u+s)^2 - \Gamma'(-s)\psi(u+s)\right. \\
&+\Gamma(-s)\psi(u+s)\ln(-z) - \psi(u+s)\Gamma'(-s) + \Gamma''(-s) - \Gamma'(-s)\ln(-z) \\
&\left.+\ln(-z)\psi(u+s)\Gamma(-s) - \ln(-z)\Gamma'(-s) + \ln(-z)^2\Gamma(-s)\right].
\end{aligned}$$

We wish to evaluate this at $s=-t$ and then set $t=1$, $u=a+1$. So, we have

$$\begin{aligned}
&\frac{1}{2}(-z)^{-1}\Gamma(a)\left[\psi'(a)\Gamma(1) + \Gamma(1)\psi(a)^2 - \Gamma'(1)\psi(a) + \Gamma(1)\psi(a)\ln(-z)\right. \quad (5.64) \\
&-\psi(a)\Gamma'(1) + \Gamma''(1) - \Gamma'(1)\ln(-z) + \ln(-z)\psi(a)\Gamma(1) - \ln(-z)\Gamma'(1) \\
&\left.+\ln(-z)^2\Gamma(1)\right].
\end{aligned}$$

Now, from section 1.4.1

$$\Gamma(1)=1, \Gamma'(1)=-\gamma, \Gamma''(1)=\gamma^2+\frac{\pi^2}{6}.$$

Therefore, using (5.64) to represent the integral (5.62) in terms of the sum of residues - evaluated at $t=1$ and $u=a+1$ - we have

$$\begin{aligned}
\Gamma(a+1)\kappa_3 = &\frac{1}{2}(-z)^{-1}\Gamma(a)\left[\psi'(a) + \psi(a)^2 + \psi(a)\gamma + \psi(a)\ln(-z) + \gamma\psi(a)\right. \\
&\left.+\gamma^2 + \frac{\pi^2}{6} + \ln(-z)[\gamma+\psi(a)] + \ln(-z)\gamma + \ln(-z)^2\right]
\end{aligned}$$

so that

$$\begin{aligned}(-z)\,a\kappa_3 - \frac{\ln(-z)^2}{2} &= \frac{1}{2}\left[\psi'(a) + \psi(a)^2 + \gamma\psi(a) + \psi(a)\ln(-z) + \gamma\psi(a) + \gamma^2 + \frac{\pi^2}{6}\right.\\ &\quad \left.+\gamma\ln(-z) + \ln(-z)\,\psi(a) + \ln(-z)\,\gamma\right]\\ &= \frac{1}{2}\left[\psi'(a) + \psi(a)^2 + 2\gamma\psi(a) + \gamma^2 + 2\ln(-z)\{\gamma + \psi(a)\} + \frac{\pi^2}{6}\right]\\ &= \frac{1}{2}\left[\psi'(a) + \{\psi(a) + \gamma\}^2 + \frac{\pi^2}{6}\right] + \ln(-z)\{\gamma + \psi(a)\},\end{aligned}$$

or, alternatively,

$$\omega a\kappa_3 - \frac{(\ln\omega)^2}{2} = \frac{1}{2}\left[\psi'(a) + \{\psi(a) + \gamma\}^2 + \frac{\pi^2}{6}\right] + \ln\omega\{\gamma + \psi(a)\}. \tag{5.65}$$

We now have κ_3 in a manageable form; involving terms that we recognise from the complete case of Chapter three. We shall see in the discussion that follows, how we use these limiting results in the expectations particular to our work.

An examination of the terms at (5.53), in conjunction with (5.34), implies that we need to find the following results,

$$\begin{aligned}&E_a[\ln(1+Y^\tau)], E_a[\ln Y], E_{a+1}[Y^\tau \ln Y], E_{a+1}[Y^\tau], E_{a+1}\left[Y^\tau(\ln Y)^2\right],\\ &E_{a+2}\left[Y^{2\tau}(\ln Y)^2\right], E_{a+2}\left[Y^{2\tau}\ln Y\right], E_{a+2}\left[Y^{2\tau}\right].\end{aligned}$$

We shall examine each of these terms in turn without the suffixes. The complete results, determined from expressions in Watkins (1997), are restated for ease of comparison.

E $[\ln(1+Y^\tau)]$ From (3.2), the complete result is given by,

$$E[\ln(1+Y^\tau)] = \frac{1}{a},$$

while, from (5.7), the Type I censored result is

$$E[\ln(1+Y^\tau)] = \frac{1}{a} - \frac{\ln(1+c^\tau)}{(1+c^\tau)^a - 1}.$$

Clearly, as $c \to \infty$, $E[\ln(1+Y^\tau)] \to \frac{1}{a}$.

E $[\ln Y]$ The complete result is given by

$$E[\ln Y] = -\left[\frac{\gamma + \psi(a)}{\tau}\right],$$

see (3.6); while the corresponding Type I censored result is

$$\begin{aligned} E[\ln Y] &= \frac{1}{\tau}\left[\ln\omega - \frac{f_2(1,a+1,-\omega)}{f_1(1,a+1,-\omega)}\right] \\ &= \frac{1}{\tau}\left[\ln\omega - \frac{f_2(1,a+1,-\omega)}{q_{c,a}/a\omega}\right], \end{aligned}$$

from (5.29). Now, as $c \to \infty$, $\frac{q_{c,a}}{a\omega} \to \frac{1}{a\omega}$. Therefore,

$$E[\ln Y] \to \frac{1}{\tau}\left[\ln\omega - a\omega f_2(1,a+1,-\omega)\right]$$

But, from (5.61),

$$a\omega f_2(1,a+1,-\omega) - \ln\omega \to \psi(a) + \gamma$$

So,

$$E[\ln Y] \to -\left[\frac{\gamma+\psi(a)}{\tau}\right]$$

as required.

$\mathbf{E}[Y^\tau \ln Y]$ From (3.7), we have the complete result as

$$E[Y^\tau \ln Y] = -\left(\frac{\psi(a-1)+\gamma-1}{\tau(a-1)}\right),$$

while the analogous result for Type I censored data, given by (5.14), is

$$E[Y^\tau \ln Y] = \frac{a\omega^2}{2\tau q_{c,a}}\left[\frac{(2\ln\omega-1)f_1(2,a+1,-\omega)}{2} - \frac{(a+1)\omega}{9}f_2(3,a+2,-\omega)\right].$$

Now, using (5.26) and (5.28) with a replaced by $a-1$, we have the following expression, (where we use the notation $\kappa_j^{(a-1)}$ $(0 \le j \le 3)$ to indicate that the κ_j should be taken with a replaced by $a-1$.),

$$\frac{a\omega^2}{2\tau q_{c,a}}\left[\frac{(2\ln\omega-1)}{2}\frac{2}{a\omega}\left(\kappa_1^{(a-1)} - \kappa_0^{(a-1)}\right) - \frac{(a+1)\omega}{9}\frac{9\left[2\kappa_2^{(a-1)} - 3\kappa_1^{(a-1)} + \kappa_0^{(a-1)}\right]}{a(a+1)\omega^2}\right];$$

using (5.58), (5.59) and (5.61), and simplifying, we have the limiting result

$$\begin{aligned} E[Y^\tau \ln Y] &\to \frac{a\omega^2}{2\tau q_{c,a}}\left[\frac{(2\ln\omega-1)2}{2a\omega\omega(a-1)} - \frac{(a+1)\omega}{9}\frac{9\left[2\frac{\psi(a-1)+\gamma+\ln\omega}{\omega(a-1)} - \frac{3}{\omega(a-1)}\right]}{a(a+1)\omega^2}\right] \\ &= -\frac{(1+\omega)^a(-1+\gamma+\psi(a-1))}{(a-1)\tau(-1+(1+\omega)^a)} \\ &= -\left(\frac{\psi(a-1)+\gamma-1}{\tau(a-1)}\right) \end{aligned}$$

as required.

$\mathbf{E}[Y^{\tau}]$ From (3.4), we have the complete result as,

$$E[Y^{\tau}] = \frac{1}{a-1};$$

while, from (5.10) and subsequently (5.26), we have the Type I censored result,

$$\begin{aligned} E[Y^{\tau}] &= \frac{a\omega}{q_{c,a}} \frac{c^{\tau}}{2} f_1(2, a+1, -\omega) \\ &= \frac{a\omega^2}{q_{c,a}2} \left\{ \frac{2}{a\omega} \left[\kappa_1^{(a-1)} - \kappa_0^{(a-1)} \right] \right\} \\ &= \frac{a\omega^2}{q_{c,a}2} \left\{ \frac{2}{a\omega} \left[\frac{q_{c,a-1}}{\omega(a-1)} - (1+\omega)^{-a} \right] \right\} \end{aligned}$$

and this approaches

$$\frac{1}{a-1}$$

as $c \to \infty$.

$\mathbf{E}[Y^{2\tau}]$ The result (3.5) gives us the complete result as

$$\begin{aligned} E[Y^{2\tau}] &= \frac{\Gamma(3)\Gamma(a-2)}{\Gamma(a)} \\ &= \frac{2}{(a-1)(a-2)}, \end{aligned}$$

with the Type I censored result, from (5.10), being represented by

$$\begin{aligned} E[Y^{2\tau}] &= \frac{a\omega}{q_{c,a}} f_1(3, a+1, -\omega) \frac{c^{2\tau}}{3} \\ &= \frac{a\omega^3}{3q_{c,a}} f_1(3, a+1, -\omega). \end{aligned}$$

Now, using (5.22), we then have,

$$\begin{aligned} E[Y^{2\tau}] &= \frac{a\omega^3}{3q_{c,a}} \{f_1(2, a+1, -\omega) - f_1(2, a, -\omega)\} \frac{3}{(-\omega)2} \\ &= \frac{a\omega^3}{3q_{c,a}} \left\{ \frac{2}{a\omega} \left(\kappa_1^{(a-1)} - \kappa_0^{(a-1)} \right) - \frac{2}{(a-1)\omega} \left(\kappa_1^{(a-2)} - \kappa_0^{(a-2)} \right) \right\} \left(\frac{-3}{2\omega} \right) \\ &= \frac{a\omega^3}{3q_{c,a}} \left\{ \frac{2}{a\omega} \left(\frac{q_{c,a-1}}{\omega(a-1)} - p_{c,a} \right) - \frac{2}{(a-1)\omega} \left(\frac{q_{c,a-2}}{(a-2)\omega} - p_{c,a-1} \right) \right\} \left(\frac{-3}{2\omega} \right) \end{aligned}$$

from (5.26). As $c \to \infty$, we have

$$E\left[Y^{2\tau}\right] \to \frac{a\omega^3}{3}\left(\frac{-3}{2\omega}\right)\left\{\frac{2}{a(a-1)\omega^2} - \frac{2}{(a-1)(a-2)\omega^2}\right\}$$

and, on simplification, this limit reduces to

$$\frac{2}{(a-1)(a-2)},$$

as required.

$\mathbf{E}\left[Y^{2\tau}\ln Y\right]$ Multiplying (3.15) by $\frac{a+2}{a}$ and then replacing a by $a-2$ gives the complete result as

$$E\left[Y^{2\tau}\ln Y\right] = \frac{3-2\gamma-2\psi(a-2)}{\tau(a-1)(a-2)}.$$

It is straightforward to show that

$$\frac{a-1}{a}E_a\left[Y^{2\tau}\ln Y\right] = E_{a-1}\left[Y^{\tau}\ln Y\right] - \frac{a-1}{a}E_a\left[Y^{\tau}\ln Y\right],$$

from which

$$E_a\left[Y^{2\tau}\ln Y\right] = \frac{a}{a-1}\left\{E_{a-1}\left[Y^{\tau}\ln Y\right] - \frac{a-1}{a}E_a\left[Y^{\tau}\ln Y\right]\right\}.$$

We have seen earlier in this section that

$$E_a\left[Y^{\tau}\ln Y\right] \to \left(\frac{1-\psi(a-1)-\gamma}{\tau(a-1)}\right)$$

as $c \to \infty$. Substituting in we have

$$\begin{aligned}&\frac{a}{a-1}\left\{\frac{1-\psi(a-2)-\gamma}{\tau(a-2)} - \frac{a-1}{a}\frac{1-\psi(a-1)-\gamma}{\tau(a-1)}\right\}\\ =\ &\frac{3-2\gamma-2\psi(a-2)}{\tau(a-1)(a-2)}\end{aligned}$$

on exploiting the recurrence relation

$$\psi(a+1) = \psi(a) + \frac{1}{a}, \tag{5.66}$$

see result 6.3.5 in Abramowitz and Stegun (1972); this is the required result.

$\mathbf{E}\left[Y^{\tau}(\ln Y)^2\right]$ From (3.8), we have the result for complete data as

$$E\left[Y^{\tau}(\ln Y)^2\right] = \frac{1}{\tau^2(a-1)}\left[\begin{array}{c}\left\{\frac{\pi^2}{6}+\gamma^2-2\gamma\right\} + 2(\gamma-1)\psi(a-1) + \{\psi(a-1)\}^2\\ +\psi'(a-1)\end{array}\right].$$

From (5.15), the Type I censored result is

$$E\left[Y^{\tau}(\ln Y)^2\right] = \frac{a\omega^2}{\tau^2 q_{c,a}} \left[\begin{array}{c} \frac{1}{8}\left\{1+(2\ln\omega-1)^2\right\} f_1(2,a+1,-\omega) \\ -\frac{1}{54}(a+1)\,\omega\,(6\ln\omega-5)\, f_2(3,a+2,-\omega) \\ +\frac{1}{192}(a+1)(a+2)\,\omega^2 f_3(4,a+3,-\omega) \end{array} \right].$$

Now, $f_1(2,a+1,-\omega)$ can be obtained from (5.26) with $a+2$ replaced by $a+1$, while $f_2(3,a+2,-\omega)$ can be obtained from (5.28) with $a+3$ replaced by $a+2$. From (5.21),

$$\begin{aligned} f_3(4,a+3,-\omega) &= [f_3(3,a+2,-\omega) - f_2(3,a+2,-\omega)]\left(-\frac{4^3}{(-\omega)\,3^2\,(a+2)}\right), \\ f_3(3,a+2,-\omega) &= [f_3(2,a+1,-\omega) - f_2(2,a+1,-\omega)]\left(-\frac{3^3}{(-\omega)\,2^2\,(a+1)}\right), \\ f_3(2,a+1,-\omega) &= [f_3(1,a,-\omega) - f_2(1,a,-\omega)]\left(-\frac{2^3}{(-\omega)\,a}\right) \end{aligned}$$

where, from (5.24), (5.25), (5.26), (5.27) and (5.28) respectively,

$$\begin{aligned} f_2(1,a,-\omega) &= \kappa_2^{(a-1)}, \\ f_3(1,a,-\omega) &= \kappa_3^{(a-1)}, \\ f_1(2,a+1,-\omega) &= \frac{2}{a\omega}\left[\kappa_1^{(a-1)} - \kappa_0^{(a-1)}\right], \\ f_2(2,a+1,-\omega) &= \frac{4\left[\kappa_2^{(a-1)} - \kappa_1^{(a-1)}\right]}{a\omega}, \\ f_2(3,a+2,-\omega) &= \frac{9\left[2\kappa_2^{(a-1)} - 3\kappa_1^{(a-1)} + \kappa_0^{(a-1)}\right]}{a(a+1)\,\omega^2} \end{aligned}$$

Substituting in and simplifying, (in a routine sense),

$$E\left[Y^{\tau}(\ln Y)^2\right] = \frac{\omega}{q_{c,a}\tau^2} \left\{ \begin{array}{c} 2\kappa_3^{(a-1)} - 2\kappa_2^{(a-1)}(1+\ln\omega) \\ +\ln\omega\left(-\kappa_0^{(a-1)}\ln\omega + \kappa_1^{(a-1)}(2+\ln\omega)\right) \end{array} \right\}.$$

We know that as $\omega \to \infty$,

$$q_{c,a} = 1 - (1+\omega)^{-a} \to 1.$$

We also have the other limiting results given by (5.58), (5.59), (5.61) and (5.65). Substituting these results in and cancelling terms, we obtain

$$E\left[Y^{\tau}(\ln Y)^2\right] \to \frac{1}{\tau^2(a-1)}\left\{\frac{\pi^2}{6} + \gamma^2 - 2\gamma + \psi'(a-1) + [\psi(a-1)]^2 + 2(\gamma-1)\,\psi(a-1)\right\},$$

as required.

$\mathbf{E}\left[Y^{2\tau}(\ln Y)^2\right]$ From (3.9), we have the complete result

$$E\left[Y^{2\tau}(\ln Y)^2\right] = \frac{1}{\tau^2(a-1)(a-2)}\left[\begin{array}{c}\left\{\frac{\pi^2}{3}+2\gamma^2-6\gamma+2\right\}-2(3-2\gamma)\psi(a-2)\\ +2\left\{[\psi(a-2)]^2+\psi'(a-2)\right\}\end{array}\right].$$

Now, it is also straightforward to show that we can write

$$\frac{a-2}{a}E_a\left[Y^{2\tau}(\ln Y)^2\right] = \frac{a-2}{a-1}E_{a-1}\left[Y^{\tau}(\ln Y)^2\right] - \frac{a-2}{a}E_a\left[Y^{\tau}(\ln Y)^2\right],$$

from which

$$E_a\left[Y^{2\tau}(\ln Y)^2\right] = \frac{a}{a-2}\left\{\frac{a-2}{a-1}E_{a-1}\left[Y^{\tau}(\ln Y)^2\right] - \frac{a-2}{a}E_a\left[Y^{\tau}(\ln Y)^2\right]\right\}.$$

Now, we have already shown

$$E_a\left[Y^{\tau}(\ln Y)^2\right] \to \frac{1}{\tau^2(a-1)}\left\{\frac{\pi^2}{6}+\gamma^2-2\gamma+\psi'(a-1)+[\psi(a-1)]^2+2(\gamma-1)\psi(a-1)\right\}.$$

Therefore,

$$E_a\left[Y^{2\tau}(\ln Y)^2\right] \to \frac{a}{a-2}\left\{\begin{array}{c}\frac{a-2}{a-1}\frac{1}{\tau^2(a-2)}\left\{\begin{array}{c}\frac{\pi^2}{6}+\gamma^2-2\gamma+\psi'(a-2)+[\psi(a-2)]^2\\ +2(\gamma-1)\psi(a-2)\end{array}\right\}\\ -\frac{a-2}{a}\frac{1}{\tau^2(a-1)}\left\{\begin{array}{c}\frac{\pi^2}{6}+\gamma^2-2\gamma+\psi'(a-1)+[\psi(a-1)]^2\\ +2(\gamma-1)\psi(a-1)\end{array}\right\}\end{array}\right\}$$

To get this expression into the correct form, we exploit (5.66) and result 6.4.6 in Abramowitz and Stegun (1972),

$$\psi^{(n)}(a+1) = \psi^{(n)}(a) + (-1)^n n! a^{-n-1}.$$

Making the appropriate substitutions and simplifying in a routine way we get the required result.

Since we have proven that all the above expressions tend to their complete counterparts when $c \to \infty$, it follows that all derived terms, see sections 5.4 and 5.5 above, also tend to their complete counterparts. This is because the only difference between the Type I and complete versions is the inclusion of a term of the form $\frac{q_{c,i,a+j}}{q_{c,i,a}}$ (for $j = 1$ or 2), which clearly tends to unity as c tends to infinity.

5.9 Δ for Censored Data

In Chapter four we determined Δ_d at (4.14) for an accelerated data set, when the data had completely failed. We now extend this result to accomodate censored data. Naturally, many of the expressions involved require only a minor adaptation of the complete case. Results here are based on the reparameterisation of a. Under this reparameterisation, and

analogous to (4.4), the Burr XII log-likelihood becomes

$$\begin{aligned} l_B &= S_n \ln\tau + S_n \ln\left(\ln\left(-\ln\omega\right)\right) + \left(\tau-1\right) S_e - \tau S_n \left[\ln\psi + \tau^{-1}\ln\left(\ln\left(-\ln\omega\right)\right)\right] \\ &\quad -\tau\beta_b S_x - \left(1+\ln\left(-\ln\omega\right)\right) t_*^d\left(\psi,\omega\right) - \ln\left(-\ln\omega\right) t_*^c\left(\psi,\omega\right), \end{aligned} \tag{5.67}$$

where $t_*^d\left(\psi,\omega\right)$ is defined at (4.5) and we now define

$$t_*^c\left(\psi,\omega\right) = \sum_{i=1}^{k}\sum_{j=1}^{m_i} \ln\left\{1+\left(\frac{\frac{c_{ij}}{\exp\left(\beta_b x_i\right)}}{\psi\left[\ln\left(-\ln\omega\right)\right]^{1/\tau}}\right)^{\tau}\right\}.$$

As in Chapter four, we are interested in the behaviour of the score function for (5.67) at $\tau = \widehat{B}, \psi = \exp\left(\widehat{\alpha}_w\right), \beta_b = \widehat{\beta}_w$ and as $\omega \to 0$. It is convenient to define

$$\begin{aligned} s_{v,w}\left(t,\beta_b\right) &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_i^w \left(\frac{d_{ij}}{\exp\left(\beta_b x_i\right)}\right)^t \left(\ln\left\{\frac{d_{ij}}{\exp\left(\beta_b x_i\right)}\right\}\right)^v \\ &\quad + \sum_{i=1}^{k}\sum_{j=1}^{m_i} x_i^w \left(\frac{c_{ij}}{\exp\left(\beta_b x_i\right)}\right)^t \left(\ln\left\{\frac{c_{ij}}{\exp\left(\beta_b x_i\right)}\right\}\right)^v, \end{aligned}$$

and we also make use of the notation $s_{v,w}^d\left(t\right)$ at (4.6), and $t_{v,w}^d\left(\psi,\omega\right)$ at (4.7). Finally, we now also define

$$t_{v,w}^c\left(\psi,\omega\right) = \sum_{i=1}^{k}\sum_{j=1}^{m_i} x_i^w \left[\frac{\left(\frac{\left(\frac{c_{ij}}{\exp\left(\beta_b x_i\right)}\right)}{\psi\left[\ln\left(-\ln\omega\right)\right]^{1/\tau}}\right)^{\tau}\left(\ln\left\{\frac{\left(\frac{c_{ij}}{\exp\left(\beta_b x_i\right)}\right)}{\psi\left[\ln\left(-\ln\omega\right)\right]^{1/\tau}}\right\}\right)^{v}}{1+\left(\frac{\left(\frac{c_{ij}}{\exp\left(\beta_b x_i\right)}\right)}{\psi\left[\ln\left(-\ln\omega\right)\right]^{1/\tau}}\right)^{\tau}}\right].$$

First Derivatives

Taking first derivatives of (5.67) we have,

$$\begin{aligned} \frac{\partial l_B}{\partial\psi} &= -\tau S_n \psi^{-1} + \left(1+\ln\left(-\ln\omega\right)\right)\tau\psi^{-1} t_{0,0}^d\left(\psi,\omega\right) + \ln\left(-\ln\omega\right)\tau\psi^{-1} t_{0,0}^c\left(\psi,\omega\right), \\ \frac{\partial l_B}{\partial\tau} &= S_n\tau^{-1} + S_e - S_n - S_n\ln\psi - \beta_b S_x - \left(1+\ln\left(-\ln\omega\right)\right)\left\{\begin{array}{c}\tau^{-1}\ln\left(\ln\left(-\ln\omega\right)\right)\times \\ t_{0,0}^d\left(\psi,\omega\right) + t_{1,0}^d\left(\psi,\omega\right)\end{array}\right\} \\ &\quad -\ln\left(-\ln\omega\right)\left\{\tau^{-1}\ln\left(\ln\left(-\ln\omega\right)\right) t_{0,0}^c\left(\psi,\omega\right) + t_{1,0}^c\left(\psi,\omega\right)\right\}, \\ \frac{\partial l_B}{\partial\beta_b} &= -\tau S_x + \left(1+\ln\left(-\ln\omega\right)\right)\tau t_{0,1}^d\left(\psi,\omega\right) + \ln\left(-\ln\omega\right)\tau t_{0,1}^c\left(\psi,\omega\right), \end{aligned}$$

and

$$\frac{\partial l_B}{\partial\omega} = -\frac{1}{\omega\ln\omega}\left\{t_*^d\left(\psi,\omega\right) + t_*^c\left(\psi,\omega\right)\right\} + \frac{\left\{1+\ln\left(-\ln\omega\right)\right\} t_{0,0}^d\left(\psi,\omega\right)}{\omega\ln\omega\ln\left(-\ln\omega\right)} + \frac{\ln\left(-\ln\omega\right) t_{0,0}^c\left(\psi,\omega\right)}{\omega\ln\omega\ln\left(-\ln\omega\right)},$$

which are clear extensions of those first derivatives expressed in section 4.2. As before, we now make the substitution

$$\lambda = \psi^{\tau}\left[\ln\left(-\ln\omega\right)\right],$$

so that, with ψ and τ finite, $\omega \to 0$ implies $\lambda \to \infty$. The first derivatives then become

$$\begin{aligned}
\frac{\partial l_B}{\partial \psi} &= -\tau S_n \psi^{-1} + \left(1+\lambda\psi^{-\tau}\right)\tau\psi^{-1}t^d_{0,0}(\lambda) + \lambda\psi^{-\tau}\tau\psi^{-1}t^c_{0,0}(\lambda) \\
\frac{\partial l_B}{\partial \beta_b} &= -\tau S_x + \left(1+\lambda\psi^{-\tau}\right)\tau t^d_{0,1}(\lambda) + \lambda\psi^{-\tau}\tau t^c_{0,1}(\lambda), \\
\frac{\partial l_B}{\partial \tau} &= S_n\tau^{-1} + S_e - S_n\ln\psi - \beta_b S_x - \left(1+\lambda\psi^{-\tau}\right)\left\{\begin{array}{c}\tau^{-1}\ln\left(\lambda\psi^{-\tau}\right)t^d_{0,0}(\lambda) \\ +t^d_{1,0}(\lambda)\end{array}\right\}, \\
&\quad -\lambda\psi^{-\tau}\left\{\tau^{-1}\ln\left(\lambda\psi^{-\tau}\right)t^c_{0,0}(\lambda) + t^c_{1,0}(\lambda)\right\},
\end{aligned}$$

and

$$\frac{\partial l_B}{\partial \omega} = \frac{1}{\omega\left(-\log\omega\right)}\left\{t^d_*(\lambda) + t^c_*(\lambda) - \frac{1+\lambda\psi^{-\tau}}{\lambda\psi^{-\tau}}t^d_{0,0}(\lambda) - t^c_{0,0}(\lambda)\right\}; \tag{5.68}$$

where we now write the $t^p_{v,w}$ and t^p_* with a λ notation, simply to indicate the role of the parameter in the summations. In order to consider the behaviour of the derivatives in the appropriate limits, we now re-state remarks and lemmas from Chapter four, altered slightly to accomodate censoring.

Lemma 1. As $\lambda \to \infty$, we have

$$\lambda\left\{t^d_*(\lambda) + t^c_*(\lambda)\right\} = s_{0,0}(\tau,\beta_b) - \frac{1}{2\lambda}s_{0,0}(2\tau,\beta_b) + O\left(\lambda^{-2}\right).$$

Remark 1. From Lemma 1, we have, as $\lambda \to \infty$

$$\lambda\left\{t^d_*(\lambda) + t^c_*(\lambda)\right\} = s_{0,0}(\tau,\beta_b) + O\left(\lambda^{-1}\right).$$

Lemma 2. As $\lambda \to \infty$, we have

$$\lambda\left\{t^d_{1,0}(\lambda) + t^c_{1,0}(\lambda)\right\} = s_{1,0}(\tau,\beta_b) - \tau^{-1}\left(\ln\lambda\right)s_{0,0}(\tau,\beta_b) + O\left(\lambda^{-1}\right).$$

Remark 2. Adapting the argument in Lemma 2, we have, as $\lambda \to \infty$,

$$t^q_{1,0}(\lambda) = O\left(\lambda^{-1}\right),$$

for $q = d, c$.

Lemma 3. We have, as $\lambda \to \infty$,

$$\lambda\left\{t^d_{0,0}(\lambda) + t^c_{0,0}(\lambda)\right\} = s_{0,0}(\tau,\beta_b) - \lambda^{-1}s_{0,0}(2\tau,\beta_b) + O\left(\lambda^{-2}\right).$$

We shall also require results analogous to (4.8) and (4.9), now extended to the censored

environment. The first, as given by Watkins (1994), for example, is

$$\{\exp(\widehat{\alpha}_w)\}^{-\widehat{B}} s_{0,0}\left(\widehat{B},\widehat{\beta}_w\right) = S_n,$$

while we also have

$$S_n\left(\frac{s_{1,0}\left(\widehat{B},\widehat{\beta}_w\right)}{s_{0,0}\left(\widehat{B},\widehat{\beta}_w\right)}\right) = S_n\widehat{B}^{-1} + s_{1,0}(0,0) - \widehat{\beta}_w S_x. \tag{5.69}$$

Behaviour of $\frac{\partial l_B}{\partial \psi}$

$$\frac{\partial l_B}{\partial \psi} = \tau\psi^{-1}\left[\lambda\psi^{-\tau}\left\{t^d_{0,0}(\lambda) + t^c_{0,0}(\lambda)\right\} + t^d_{0,0}(\lambda) - S_n\right];$$

adapting Lemma 3 gives

$$\lambda\left\{t^d_{0,0}(\lambda) + t^c_{0,0}(\lambda)\right\} = s_{0,0}(\tau,\beta_b) + O\left(\lambda^{-1}\right),$$

while

$$t^d_{0,0}(\lambda) = \lambda^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\frac{\left(\frac{d_{ij}}{\exp(\beta_b x_i)}\right)^{\tau}}{1+\frac{\left(\frac{d_{ij}}{\exp(\beta_b x_i)}\right)^{\tau}}{\lambda}}\right] = O\left(\lambda^{-1}\right).$$

Therefore,

$$\frac{\partial l_B}{\partial \psi} = \tau\psi^{-1}\left[\psi^{-\tau}s_{0,0}(\tau,\beta_b) - S_n + O\left(\lambda^{-1}\right)\right].$$

Now, since

$$\psi^{-\tau}s_{0,0}(\tau,\beta_b)\big|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = \{\exp(\widehat{\alpha}_w)\}^{-\widehat{B}} s_{0,0}\left(\widehat{B},\widehat{\beta}_w\right) = S_n \tag{5.70}$$

we have the result

$$\lim_{\lambda\to\infty}\frac{\partial l_B}{\partial\psi}\bigg|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = \lim_{\omega\to 0}\frac{\partial l_B}{\partial\psi}\bigg|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = 0.$$

Behaviour of $\frac{\partial l_B}{\partial \beta_b}$

$$\frac{\partial l_B}{\partial \beta_b} = \tau\left[\lambda\psi^{-\tau}\left\{t^d_{0,1}(\lambda) + t^c_{0,1}(\lambda)\right\} + t^d_{0,1}(\lambda) - S_x\right].$$

Here,

$$t^d_{0,1}(\lambda) = O\left(\lambda^{-1}\right)$$

and

$$\lambda\left\{t^d_{0,1}(\lambda) + t^c_{0,1}(\lambda)\right\} = s_{0,1}(\tau,\beta_b) + O\left(\lambda^{-1}\right),$$

adapting Lemma 3; while, adapting (5.70), and writing in full, we have

$$\begin{aligned}\psi^{-\tau}s_{0,1}(\tau,\beta_b) &= \psi^{-\tau}\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{d_{ij}}{\exp(\beta_b x_i)}\right)^{\tau}x_i+\sum_{i=1}^{k}\sum_{j=1}^{m_i}\left(\frac{c_{ij}}{\exp(\beta_b x_i)}\right)^{\tau}x_i\right]\\ &= \sum_{i=1}^{k}n_i x_i = S_x,\end{aligned}$$

in the limit $\tau=\widehat{B}$, $\psi=\exp(\widehat{\alpha}_w)$ and $\beta_b=\widehat{\beta}_w$. Thus, we also have the result

$$\lim_{\lambda\to\infty}\left.\frac{\partial l_B}{\partial\beta_b}\right|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = \lim_{\omega\to 0}\left.\frac{\partial l_B}{\partial\beta_b}\right|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = 0.$$

Behaviour of $\frac{\partial l_B}{\partial\tau}$ For $\frac{\partial l_B}{\partial\tau}$, we have

$$\begin{aligned}\frac{\partial l_B}{\partial\tau} &= S_n\tau^{-1}-S_n\ln\psi+s^d_{1,0}(0,0)-\beta_b S_x-\left(1+\lambda\psi^{-\tau}\right)t^d_{1,0}(\lambda)-\lambda\psi^{-\tau}t^c_{1,0}(\lambda)\\ &\quad-\left(1+\lambda\psi^{-\tau}\right)\left(\tau^{-1}\ln\lambda-\ln\psi\right)\left\{t^d_{0,0}(\lambda)+t^c_{0,0}(\lambda)\right\}\\ &= S_n\tau^{-1}-S_n\ln\psi+s^d_{1,0}(0,0)-\beta_b S_x-\psi^{-\tau}\left(\tau^{-1}\ln\lambda-\ln\psi\right)\left\{s_{0,0}(\tau,\beta_b)\right.\\ &\quad\left.+O\left(\lambda^{-1}\right)\right\}-\psi^{-\tau}\left\{s_{1,0}(\tau,\beta_b)-\tau^{-1}\ln\lambda s_{0,0}(\tau,\beta_b)+O\left(\lambda^{-1}\right)\right\}\\ &\quad-\tau\ln\left(\lambda\psi^{-\tau}\right)O\left(\lambda^{-1}\right)+O\left(\lambda^{-1}\right),\end{aligned}$$

using Remark 2, Lemma 2 and Lemma 3. Simplification yields

$$\begin{aligned}\frac{\partial l_B}{\partial\tau} &= S_n\tau^{-1}+s^d_{1,0}(0,0)-\beta_b S_x-\psi^{-\tau}s_{1,0}(\tau,\beta_b)+\ln\psi\left\{\psi^{-\tau}s_{0,0}(\tau,\beta_b)-S_n\right\}\\ &\quad+O\left(\lambda^{-1}\right)\\ &= S_n\tau^{-1}+s^d_{1,0}(0,0)-\beta_b S_x-\psi^{-\tau}s_{1,0}(\tau,\beta_b)+O\left(\lambda^{-1}\right)\end{aligned}\tag{5.71}$$

using (5.70) for the term in brackets. In order to show that this expression is zero in the limit, we use the result at (5.69), together with the fact that we can write (5.70) as

$$\left.\psi^{-\tau}\right|_{\psi=\exp(\widehat{\alpha}_w),\tau=\widehat{B}} = \frac{S_n}{s_{0,0}\left(\widehat{B},\widehat{\beta}_w\right)}.$$

Substituting into (5.71) gives the result

$$\lim_{\lambda\to\infty}\left.\frac{\partial l_B}{\partial\tau}\right|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = \lim_{\omega\to 0}\left.\frac{\partial l_B}{\partial\tau}\right|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w} = 0.$$

Behaviour of $\frac{\partial l_B}{\partial \omega}$ Finally, we turn our attention to $\frac{\partial l_B}{\partial \omega}$. From (5.68), we see the behaviour of this derivative is determined by the expression

$$\begin{aligned} & t_*^d(\lambda) + t_*^c(\lambda) - \frac{1+\lambda\psi^{-\tau}}{\lambda\psi^{-\tau}} t_{0,0}^d(\lambda) - t_{0,0}^c(\lambda) \\ = & \frac{\lambda\psi^{-\tau}\left\{t_*^d(\lambda) + t_*^c(\lambda)\right\} - \left(1+\lambda\psi^{-\tau}\right) t_{0,0}^d(\lambda) - \lambda\psi^{-\tau} t_{0,0}^c(\lambda)}{\lambda\psi^{-\tau}} \\ = & \lambda^{-1}\left[\lambda\left\{t_*^d(\lambda) + t_*^c(\lambda)\right\} - \lambda\left\{t_{0,0}^d(\lambda) + t_{0,0}^c(\lambda)\right\} - \psi^{\tau} t_{0,0}^d(\lambda)\right] \end{aligned}$$

as $\lambda \to \infty$. Using Lemma 1 to deal with $\left\{t_*^d(\lambda) + t_*^c(\lambda)\right\}$ and Lemma 3 for the other terms, this expression becomes

$$\begin{aligned} & \lambda^{-1}\left[\begin{array}{c} s_{0,0}(\tau,\beta_b) - \frac{1}{2\lambda} s_{0,0}(2\tau,\beta_b) + O\left(\lambda^{-2}\right) - \left\{s_{0,0}(\tau,\beta_b) - \frac{1}{\lambda} s_{0,0}(2\tau,\beta_b)\right. \\ \left. + O\left(\lambda^{-2}\right)\right\} - \frac{\psi^{\tau}}{\lambda}\left\{s_{0,0}^d(\tau,\beta_b) + O\left(\lambda^{-1}\right)\right\} \end{array}\right] \\ = & \lambda^{-2}\left[\frac{1}{2} s_{0,0}(2\tau,\beta_b) - \psi^{\tau} s_{0,0}^d(\tau,\beta_b)\right] + O\left(\lambda^{-3}\right). \end{aligned}$$

So, as $\lambda \to \infty$, the sign of

$$\left.\frac{\partial l_B}{\partial \omega}\right|_{\tau=\widehat{B},\psi=\exp(\widehat{\alpha}_w),\beta_b=\widehat{\beta}_w}$$

is determined by

$$\Delta_{dc} = \frac{s_{0,0}\left(2\widehat{B},\widehat{\beta}_w\right)}{2} - \frac{s_{0,0}\left(\widehat{B},\widehat{\beta}_w\right) s_{0,0}^d\left(\widehat{B},\widehat{\beta}_w\right)}{S_n}$$

where we now have

$$\begin{aligned} s_{0,0}^d\left(\widehat{B},\widehat{\beta}_w\right) &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{d_{ij}}{\exp\left(\widehat{\beta}_w x_i\right)}\right)^{\widehat{B}}, \\ s_{0,0}\left(r,\widehat{\beta}_w\right) &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{d_{ij}}{\exp\left(\widehat{\beta}_w x_i\right)}\right)^{r} + \sum_{i=1}^{k}\sum_{j=1}^{m_i}\left(\frac{c_{ij}}{\exp\left(\widehat{\beta}_w x_i\right)}\right)^{r}. \end{aligned}$$

As for the complete result of Chapter four, we can rescale the value of Δ_{dc} by dividing each d_{ij} by $\exp(\widehat{\alpha}_w)$, giving the censored extension of (4.15) as

$$\begin{aligned} \Delta_{sc} = & \frac{1}{2}\left\{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{d_{ij}}{\exp\left(\widehat{\alpha}_w + \widehat{\beta}_w x_i\right)}\right)^{2\widehat{B}} + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{c_{ij}}{\exp\left(\widehat{\alpha}_w + \widehat{\beta}_w x_i\right)}\right)^{2\widehat{B}}\right\} \\ & - \frac{1}{S_n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{d_{ij}}{\exp\left(\widehat{\alpha}_w + \widehat{\beta}_w x_i\right)}\right)^{\widehat{B}} \times \end{aligned}$$

$$\left\{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{d_{ij}}{\exp\left(\widehat{\alpha}_w+\widehat{\beta}_w x_i\right)}\right)^{\widehat{B}}+\sum_{i=1}^{k}\sum_{j=1}^{m_i}\left(\frac{c_{ij}}{\exp\left(\widehat{\alpha}+\widehat{\beta}_w x_i\right)}\right)^{\widehat{B}}\right\} \tag{5.72}$$

Example 5.1.1.1 [Electric Motor Data] revisited

Section 5.6 showed that, for the electric motor data, in terms of maximised log-likelihoods, the Weibull ALT model was preferred over the Burr XII ALT model. For this data set, the value of (5.72) was,

$$\Delta_{sc} = -2.233668,$$

while

$$\Delta_{dc} = -1.651 \times 10^{44}.$$

This confirms our belief that the Weibull ALT model is a better fit to the data set than the Burr XII ALT model. This completes our discussion of Δ for this chapter.

5.9.1 Summary

We began this chapter with a discussion on the general principles of censoring, establishing our notation. We then considered the Burr XII two-parameter and three-parameter models, determining the expected Fisher information matrices and checking, for each model, that the expectations of the score functions were zero. We proceeded to extend the Burr XII distribution to an accelerated framework, again checking that expectations of score functions were zero. For this final model, we also examined the asymptotic properties of expectations of the terms that contributed to the second derivatives, firstly from a numerical standpoint and then theoretically; we also showed that theoretical expectations matched with complete counterparts as the stopping time c tended to infinity. For completeness we then established the Weibull ALT model for Type I censoring, omitting its EFI matrix. We then proceeded to fit Weibull and Burr XII non-ALT and ALT models to a published data set and examine the MLEs. Finally, we determined an expression for Δ when the data set was censored.

Chapter 6

Some Aspects of The Design of ALT Experiments

6.1 Aims and Scope

Our work so far has been primarily concerned with establishing and validating theoretical results. In this chapter, we shall consider some practical implications of our work, with the ultimate aim of providing some guidelines to practitioners. In a real life scenario, an engineer will typically wish to make inferences on the running time at design stress, or some percentile of life at the design stress; for instance, estimating the 10^{th} percentile of failure time, B_{10}, of specimens at design stress. The aim then would be to minimise the mean square error - or, if the bias is sufficiently small, the standard deviation - of the estimate of B_{10}.

Nelson (1990) first considers this for complete data sets, using the principle of least squares analysis; a further extension to Type I censored data is based on maximum likelihood methods. Kielpinski and Nelson (1975) consider this topic for data from Normal and Lognormal distributions, while Meeker and Nelson (1975) analyse data from Weibull and Extreme value distributions. Since we have obtained the Burr XII EFI matrices for complete and Type I censored data, we use maximum likelihood theory as the basis of our investigations.

The first part of this chapter develops the necessary theory when the data is correctly assumed to follow a Burr XII ALT distribution, and, additionally, the life-stress relationship is log-linear, see (1.12). We then consider $\widehat{B}_{10}$ for the real-life examples previously considered, and then assess the penalty for replacing true parameter values with their maximum likelihood counterparts; this will be particularly relevant when considering published data sets, where we do not know the true parameter values.

We then move on to consider real life applications. We begin by looking at complete data without design; all items are tested to failure at the design stress x_d, corresponding to normal operating conditions. We then introduce censoring and examine the effect on the

mean square error (or standard deviation) of $\widehat{B}_{10}$, by stopping the experiment at an earlier time. Consequently, we are interested in the relationship between c and the mean square error (or standard deviation) of $\widehat{B}_{10}$. Acceleration is then considered for completely failed data, and we are interested in the impact on the mean square error (or standard deviation) of $\widehat{B}_{10}$ by altering the value, number and distribution of stress levels and the sample allocation to these stress levels. Finally, we examine censoring in the accelerated framework, for which we consider the same factors as in the complete case, but, of course, we are now concerned with the role of the stopping time c. For each of these four key areas, we present a series of simulations and highlight their degree of agreement with theoretical results. When the agreement is good, we can carry out a wide range of theoretical investigations and observe some immediate results. In particular, we can alter the stress level settings and examine the impact upon the mean square error (or standard deviation) of $\widehat{B}_{10}$. Some investigations will be purely theoretical; in these instances, time constraints mean that we do not assess agreement with simulated results or determine the bias in $\widehat{B}_{10}$. However, an array of simulations will be considered for a suitable portion of the parameter space, for which we discuss agreement with the theoretical mean square error (or standard deviation) of $\widehat{B}_{10}$, and comment on the observed bias in $\widehat{B}_{10}$.

6.2 B_{10} for a Burr XII ALT

Suppose we wish to find the $100p^{th}$ percentile of the distribution of lifetimes at the design stress x_d say; throughout, we set $p = 0.1$. The details and principles for other values of p are similar. From (2.9), we have

$$\begin{aligned} B_{10}(a, \tau, \alpha_b, \beta_b) &= \left[(1-p)^{-\frac{1}{a}} - 1\right]^{\frac{1}{\tau}} \exp(\alpha_b + \beta_b x_d) \\ &= \left[0.9^{-\frac{1}{a}} - 1\right]^{\frac{1}{\tau}} \exp(\alpha_b + \beta_b x_d) \\ &= \rho^{\frac{1}{\tau}} \exp(\alpha_b + \beta_b x_d), \end{aligned} \tag{6.1}$$

where

$$\rho = 0.9^{-\frac{1}{a}} - 1.$$

Now, when running simulations, we know the true parameter values and can also obtain a set of maximum likelihood estimates for each of, say, 10000 replications. Consequently, in this case we can look at the average bias in $\widehat{B}_{10}$; we can also find the simulated standard deviation of $\widehat{B}_{10}$, and obtain a theoretical approximation to its true standard deviation, which we detail below. However, in most practical circumstances, we do not know the true parameter values, but estimate them from the data. Therefore, we use the approximation $\widehat{B}_{10}$, where we use the maximum likelihood estimates of the parameters in (6.1), and then determine an estimate of the theoretical standard deviation of $\widehat{B}_{10}$. In order to make suitable progress here, we first linearise the right hand side of (6.1), and then employ asymptotic

theory; see Cox and Hinkley (1974). We therefore expand $\widehat{B}_{10}$ in a Taylor Series expansion; the first order approximation is given by

$$\widehat{B}_{10} \simeq B_{10} + B'_{10,\pi}\left(\widehat{\underline{\pi}} - \underline{\pi}\right); \tag{6.2}$$

see, for example, Binmore (1983). In our case,

$$\underline{\pi} = (a, \tau, \alpha_b, \beta_b)'$$

and

$$B_{10,\pi} = \left(\frac{\partial B_{10}}{\partial a}, \frac{\partial B_{10}}{\partial \tau}, \frac{\partial B_{10}}{\partial \alpha_b}, \frac{\partial B_{10}}{\partial \beta_b}\right)'.$$

Since

$$\frac{d\rho}{da} = \frac{(\rho+1)\ln 0.9}{a^2},$$

we have

$$\frac{\partial B_{10}}{\partial a} = \frac{\exp\left(\alpha_b + \beta_b x_d\right)\rho^{-1+\frac{1}{\tau}}(\rho+1)\ln 0.9}{a^2\tau}, \tag{6.3}$$

$$\frac{\partial B_{10}}{\partial \tau} = \frac{-\exp\left(\alpha_b + \beta_b x_d\right)\rho^{\frac{1}{\tau}}\ln\rho}{\tau^2}, \tag{6.4}$$

$$\frac{\partial B_{10}}{\partial \alpha_b} = \exp\left(\alpha_b + \beta_b x_d\right)\rho^{\frac{1}{\tau}} \tag{6.5}$$

and

$$\frac{\partial B_{10}}{\partial \beta_b} = \exp\left(\alpha_b + \beta_b x_d\right)\rho^{\frac{1}{\tau}} x_d, \tag{6.6}$$

and, so, upon simplifying (6.2), we see that the first order approximation is given by

$$\begin{aligned} &\widehat{a}\tfrac{\partial B_{10}}{\partial a} + \widehat{\tau}\tfrac{\partial B_{10}}{\partial \tau} + \widehat{\alpha}_b\tfrac{\partial B_{10}}{\partial \alpha_b} + \widehat{\beta}_b\tfrac{\partial B_{10}}{\partial \beta_b} \\ &+\tfrac{\exp(\alpha_b+\beta_b x_d)\rho^{\frac{1}{\tau}}}{a\tau\left(-1+0.9^{\frac{1}{a}}\right)}\left\{-a\tau\left(-1+0.9^{\frac{1}{a}}\right)(-1+\alpha_b+\beta_b x_d)\right. \\ &\left.+a\left(-1+0.9^{\frac{1}{a}}\right)\ln\rho + \ln 0.9\right\}. \end{aligned} \tag{6.7}$$

Classical theory now states that an MLE $\widehat{\pi}$ asymptotically follows a Normal distribution with mean given by the true parameter value and variance given by the inverse of the EFI matrix, (Cox and Hinkley 1974), and much of our previous work supports this. Thus, we approximate the asymptotic distribution of $\widehat{B}_{10}$ by that of the right hand side of (6.2), and the asymptotic distribution of this linear function is $N\left(B_{10}, V_{10}\right)$,where

$$V_{10} = B'_{10,\pi} I^{-1} B_{10,\pi}, \tag{6.8}$$

and I is the EFI matrix for $\underline{\pi}$; all quantities are evaluated at the true parameter values. From this, the bias in $\widehat{B}_{10}$ is expected to be negligible, and its asymptotic theoretical standard

deviation is $\sqrt{V_{10}}$. Naturally, the form of the EFI matrix will change, depending on whether we are using complete, Type I, Type II or any other form of censored data, but, in general, we need to calculate V_{10}; this quantity can be evaluated at the true parameter values when considering simulations; while, in practice, we use the maximum likelihood estimates in place of the true values. The expression for V_{10} includes parameters we are not in a position to change - a, τ and even α_b and β_b directly - and quantities that we can change, such as k, $x_1, \cdots x_k$ and $n_1, \cdots, n_k$. We see - particularly from simulation experiments in Chapters three and five - that the inverse of the EFI matrix has a factor of n_1^{-1}. Consequently, there is little value in considering changing the overall sample size, since we know immediately what impact this will have upon the estimate of the variance of $\widehat{B}_{10}$.

6.2.1 Examples

Example 2.1.1.1 [Carbon Fibre Rod Data] revisited

Substituting the Burr XII ALT parameter estimates - see section 2.3.3 - into (6.1), and taking the design length to be $x_d = 0.5mm$, we calculate

$$\widehat{B}_{10} = \left[0.9^{\frac{-1}{1.8993}} - 1\right]^{\frac{1}{6.1234}} \exp\left(1.41096 - 0.011 \times 0.5\right) = 2.5542.$$

Therefore, we would expect 10% of items to have failed at a stress of 2.5542 GPa at the design length of 0.5mm. In order to find the theoretical variance of $\widehat{B}_{10}$, we need to substitute the true parameter values into the vector $B_{10,\pi}$ and the EFI matrix for complete data, given by (3.21). Substituting these into (6.8), would then give us a first order approximation to the theoretical variance of $\widehat{B}_{10}$. However, as mentioned previously, in practice we do not know the true parameter values. Consequently we use the next best thing, the maximum likelihood estimates. This gives us a first order approximation to the theoretical variance of $\widehat{B}_{10}$ as

$$var\left(\widehat{B}_{10}\right) = 0.006609,$$

and hence,

$$sd\left(\widehat{B}_{10}\right) = 0.081301.$$

Naturally, we can then consider confidence intervals around the true value of B_{10}.

Example 2.1.1.2 [Aluminium Coupon Data] revisited

Using the Burr XII ALT parameter estimates - again, found in section 2.3.3 - in (6.1), and taking the assumed design stress to be $x_d = 1.1$ psi/cycle, we calculate

$$\widehat{B}_{10} = \left[0.9^{\frac{-1}{1.8200143}} - 1\right]^{\frac{1}{7.0197622}} \exp\left(7.6785289 - 2.361454 \times 1.1\right) = 107.683.$$

So, we would expect 10% of items to have failed after 107.683 cycles at the design stress of 1.1 psi/cycle. Substituting the maximum likelihood estimates, stress levels and corresponding sample sizes into (6.3) to (6.6) and (3.21), and then into (6.8) gives the first order approximation to the theoretical variance of $\widehat{B}_{10}$ as,

$$var\left(\widehat{B}_{10}\right) = 27.5917,$$

and hence

$$sd\left(\widehat{B}_{10}\right) = 5.25278.$$

Of course, since we do not know the true parameter values, we have no way of telling whether this approximation to the standard deviation - which is based only on a first order Taylor series approximation - is close enough to that which would be obtained using the true parameter values. However, through simulations, where we know true values, we can compare the two. This should then give us some indication as to how appropriate it is to replace the true values with the maximum likelihood estimates.

Simulation 1 This simulation is based on section 3.2, and uses the parameter values $a = 2$, $\tau = 5$, $\alpha_b = 17.60139$, $\beta_b = -0.056282$, $k = 3$ with $(x_1, x_2, x_3) = (150, 170, 190)$, $n_1 = n_2 = n_3 = 100$ and $x_d = 130$. The value of B_{10}, using the true parameter values in (6.1) is 16339.1, while the average of the $\widehat{B}_{10}$, over the 10000 sets of simulations, is 16372.6. The standard deviation of the 10000 simulated $\widehat{B}_{10}$ is 827.2358, while the first order Taylor series approximation to the theoretical standard deviation of $\widehat{B}_{10}$, found by evaluating (6.8) at the true parameter values, is 825.75. Consequently, the mean square error of $\widehat{B}_{10}$ is estimated by

$$827.2358^2 + (16372.6 - 16339.1)^2 = 684319.068 + 1122.25 = 685441.318,$$

and the percentage difference between the estimated mean square error and the theoretical variance of $\widehat{B}_{10}$ is 0.52%. Clearly, with such a small bias, relative to the variance, we could use the theoretical standard deviation of $\widehat{B}_{10}$ as an alternative to the mean square error. The skewness of $\widehat{B}_{10}$ was found to be 0.148, while a Normal Q-Q plot of the 10000 simulated values of $\widehat{B}_{10}$ is shown in Figure 6.1; the K-S test statistic is 1.151, with a significance level of 0.141; this confirms that we can regard $\widehat{B}_{10}$ as Normally distributed for this choice of parameter values and sample size.

Simulation 2 This simulation uses parameter values previously used in section 3.3, where we now choose $a = \tau = 2$; other parameter values are $\alpha_b = 5.5$, $\beta_b = -0.01$, $k = 4$ with $(x_1, x_2, x_3, x_4) = (10, 80, 150, 240)$ and $n_1 = n_2 = n_3 = n_4 = 25$; we also choose $x_d = 5$. The value of B_{10}, using the true parameter values in (6.1) is 54.1344, while the average of the 10000 estimates, $\widehat{B}_{10}$, from simulations, is 54.8008. The first order Taylor series approximation to the theoretical standard deviation of $\widehat{B}_{10}$ was found to be 8.89687, while the standard deviation of the 10000 simulated $\widehat{B}_{10}$ is 8.8272. Consequently, the mean square

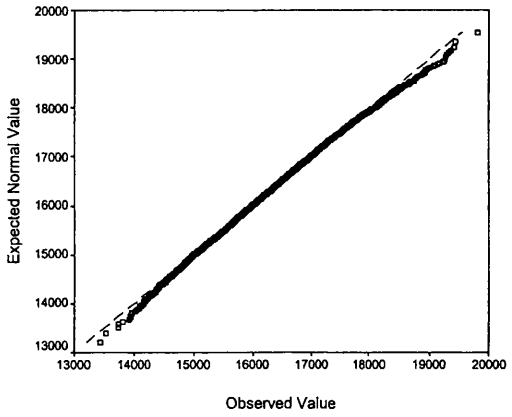

Figure 6.1: Simulation 1: Normal Q-Q plot of $\widehat{B}_{10}$, for $a = 2$, $\tau = 5$, $\alpha_b = 17.60139$, $\beta_b = -0.056282$, $(x_1, x_2, x_3) = (150, 170, 190)$, $n_1 = 100$ and $x_d = 130$.

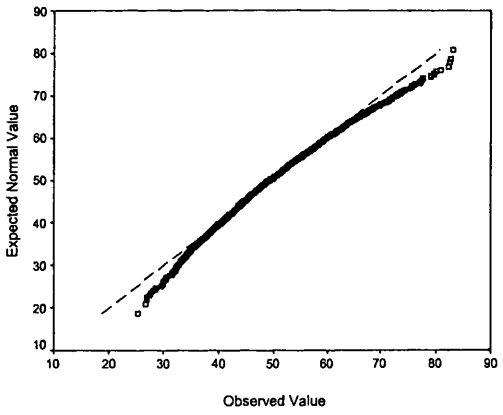

Figure 6.2: Simulation 2: Normal Q-Q plot of $\widehat{B}_{10}$, for $a = 2$, $\tau = 2$, $\alpha_b = 5.5$, $\beta_b = -0.01$, $(x_1, x_2, x_3, x_4) = (10, 80, 150, 240)$, $n_1 = 25$ and $x_d = 5$.

error of $\widehat{B}_{10}$ is estimated by

$$8.8272^2 + (54.8008 - 54.1344)^2 = 78.3635,$$

and the percentage difference between the theoretical variance of $\widehat{B}_{10}$ and the estimated mean square error is 0.9%. The skewness of $\widehat{B}_{10}$ is 0.354, while a Normal Q-Q plot of the 10000 simulated values of $\widehat{B}_{10}$ is shown in Figure 6.2; the K-S test statistic is 2.885, which corresponds to a significance level of < 0.001, indicating that we cannot regard $\widehat{B}_{10}$ as being Normally distributed. This is possibly due to the small sample size; for comparison, we recall section 4.3, where we studied the behaviour of Δ_s, that the smaller the sample size the less Normal the distribution of Δ_s.

Simulation 1 suggests that we can replace the true parameter values with their maximum

likelihood estimates, at least for sample sizes as small as 300. Then, we observed the average $\widehat{B}_{10}$ to be sufficiently close to the true value of B_{10} so that the resulting bias is negligible and we can study the standard deviation of $\widehat{B}_{10}$ instead of its mean square error. We also see that the distribution of $\widehat{B}_{10}$ is less Normal for smaller sample sizes, which is to be expected. Now, the total sample size in the first simulation was 300, and there we observed a very small bias, which, in turn, resulted in the theoretical variance of $\widehat{B}_{10}$ being very close to its estimated mean square error; we also saw some evidence of Normality in the distribution of $\widehat{B}_{10}$. Consequently, we shall use this value as the total sample size when next discussing design optimisation.

6.3 Design Optimisation: Non-ALT Complete

6.3.1 Theoretical Considerations

In this section, we consider completely failed data, where all items have been tested at the design stress. Since we do not have several stress levels, the design stress - that is, normal operating conditions - is effectively equivalent to the value of the scale parameter, θ. To determine the first order approximation to the theoretical standard deviation of $\widehat{B}_{10}$, we use the EFI matrix of the three-parameter Burr XII distribution, given by (3.1). By comparison with (6.1), it is clear that, for the three-parameter model, we have

$$B_{10} = \left[0.9^{-\frac{1}{a}} - 1\right]^{\frac{1}{\tau}} \theta = \rho^{\frac{1}{\tau}}\theta, \tag{6.9}$$

so that, with $\theta = \exp(\alpha_b + \beta_b x_d)$, we obtain the expression for the accelerated case in (6.1). From (6.9), and using (6.2), we obtain the first order Taylor series expansion of $\widehat{B}_{10}$ as

$$\begin{aligned} &\widehat{a}\frac{\theta\rho^{-1+\tau^{-1}}(1+\rho)\ln 0.9}{a^2\tau} \\ &-\widehat{\tau}\frac{\theta\rho^{\tau^{-1}}\ln\rho}{\tau^2} \\ &+\widehat{\theta}\rho^{\tau^{-1}} \\ &+\frac{\rho^{\tau^{-1}}\theta\left\{a\left(-1+0.9^{a^{-1}}\right)\ln\rho+\ln 0.9\right\}}{a\tau\left(-1+0.9^{a^{-1}}\right)}, \end{aligned} \tag{6.10}$$

upon simplifying. Using the true parameter values in the EFI matrix (3.1), we can then obtain a first order approximation to the theoretical standard deviation of $\widehat{B}_{10}$ from (6.8), where, in this case, the vector $B_{10,\pi}$ is $\left(\frac{\partial B_{10}}{\partial a}, \frac{\partial B_{10}}{\partial \tau}, \frac{\partial B_{10}}{\partial \theta}\right)'$; see (6.10). We choose the set of parameter values, $a = 2, \tau = 4, \theta = 29282.05$ and $N = 300$ and, for the sake of argument, we assume that failure data in this experiment, as with all experiments that follow in this chapter, are measured in hours. The value of θ, is to ensure consistency with results that follow later, for alternative experimental designs. We regard this value of θ as appropriate

to running items at a temperature of 130 degrees Celsius. This gives the experiment a more practical feel and should facilitate comparisons between the four design strategies considered in this chapter. However, the value of θ is merely an indication of the average failure time of a single experimental item, not of the length of the experiment, which, of course, is governed by the longest time to failure of all items on test.

For this set of parameter values, we find the value of B_{10}, from (6.9), as 14121.663 hours. So, we would expect 10% of items to have failed by this time at this operating temperature. Using (6.10), together with (3.1), we find the first order approximation to the theoretical standard deviation of $\widehat{B}_{10}$, as 545.248 hours, and this sets the benchmark for further investigations below. To validate this result we carry out a simulation using the parameter values above. As usual, our results are based on 10000 replications. We found the mean of $\widehat{B}_{10}$ to be 14126.99, with associated standard deviation 547.9822. From this, we can easily show that the relative margin of error between the theoretical and simulated mean of $\widehat{B}_{10}$ is 0.03%, while the corresponding result for the standard deviations is 0.5%. Finally, the relative margin of error between the mean square error of $\widehat{B}_{10}$ and its theoretical variance is found to be 1.01%. With such small percentage errors, we are clearly more confident that the theoretical expressions obtained above are valid. However, as previously noted, the experiment is not complete until all items have failed and, from the same simulation, we observe the average of the longest failure times to be 63919.11 hours, or about $7\frac{1}{4}$ years.

6.3.2 Simulations

In this section, we validate some theoretical results with simulations. We consider various pairs for the shape parameters, and let a take the values $2, 4$ and 6, and τ take the values 2 and 5; we also assume a single value of $\theta = 100$ and choose five total sample sizes, $N = 100, 300, 600, 1500$ and 3000. We first note that, for each combination of parameter values considered, the bias in $\widehat{B}_{10}$ is less than 1%, and this decreased as the sample size increased, as expected. As with simulations at the end of Chapter three, we tabulate values of $\sqrt{N}\times$ standard deviation of $\widehat{B}_{10}$. Simulated values are based on 10000 replications, and these results, together with their theoretical counterparts are presented in Table 6.1. We see the simulated values are close to their theoretical counterparts, even for the smallest sample size considered.

6.4 Design Optimisation: Non-ALT Type I Censoring

6.4.1 Theoretical Considerations

This section introduces the stopping time c, and we consider the same set of parameter values as in the complete case above, but now examine the impact on the theoretical standard deviation of $\widehat{B}_{10}$ by varying the stopping time of the experiment. We use the EFI matrix for the three-parameter Burr XII distribution with Type I censoring - expectations contributing

		Simulated Value					Theoretical
a	τ	$N = 100$	$N = 300$	$N = 600$	$N = 1500$	$N = 3000$	Value
2	2	30.232	31.285	31.020	31.538	31.093	31.129
2	5	29.185	29.618	29.877	30.165	29.788	29.853
4	2	21.028	21.018	21.313	21.664	21.526	21.363
4	5	24.297	25.148	25.460	25.476	25.186	25.341
6	2	16.836	16.936	17.161	17.247	17.443	17.267
6	5	22.531	22.502	22.981	23.106	23.302	23.162

Table 6.1: Theoretical and simulated values of $\sqrt{N} \times$ standard deviations of $\widehat{B}_{10}$ for various a, τ and sample size N.

to this matrix can be found in sections 5.3 and 5.4 - together with the coefficients of the MLEs in (6.10); both evaluated at the true parameter values. In this example, the vector $B_{10,\pi}$ is

$$\begin{pmatrix} \frac{\partial B_{10}}{\partial a} \\ \frac{\partial B_{10}}{\partial \tau} \\ \frac{\partial B_{10}}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \frac{1}{2^2 4}\left\{29282.05\left(-1+0.9^{-1/2}\right)^{-1+1/4} 0.9^{-1/2} \ln 0.9\right\} \\ -\frac{1}{4^2}\left\{29282.05\left(-1+0.9^{-1/2}\right)^{1/4} \ln\left(-1+0.9^{-1/2}\right)\right\} \\ \left(0.9^{-1/2}-1\right)^{1/4} \end{pmatrix}$$

$$= \begin{pmatrix} -1812.11 \\ 2574.61 \\ 0.482263 \end{pmatrix},$$

from (6.10), while the lower triangle of the non-ALT Type I censored EFI matrix is

$$I = \begin{pmatrix} 5.5044 & & \\ -8.8262 & 15.6549 & \\ -0.001475 & 0.002373 & 3.9577 \times 10^{-7} \end{pmatrix},$$

evaluated at $c = 13000$. Calculating (6.8), then gives us the first order Taylor series approximation to the theoretical standard deviation of $\widehat{B}_{10}$ as 1009.35 hours, and this is the first point plotted on the continuous line in Figure 6.3. We can clearly see how quickly the standard deviation of $\widehat{B}_{10}$ drops as the stopping time increases. For example, if we are prepared to accept a 5% increase in the standard deviation of $\widehat{B}_{10}$ over the complete value of 545.248 - indicated by the single point in Figure 6.3 - then Figure 6.3 implies that we could stop the experiment after 20000 hours, which amounts to a saving of some 44000 hours, or around five years. Consequently, our complete experiment, which would take over seven years to run, would be reduced by approximately five years.

6.4.2 Simulations

In this section, we again validate some theoretical results with simulations. We use the same set of parameter values as in the complete case above. With a scale parameter value

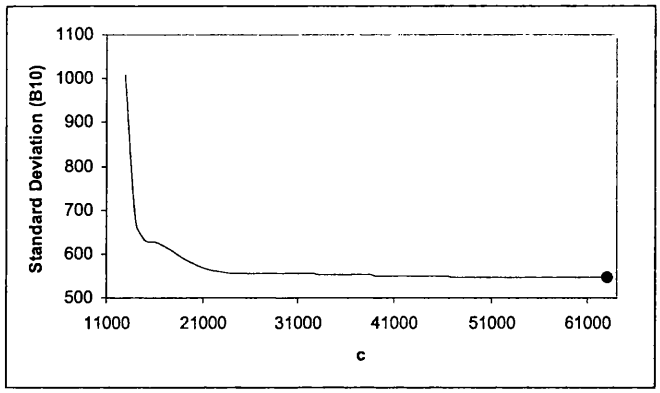

Figure 6.3: Theoretical standard deviation of $\widehat{B}_{10}$ versus the stopping time c. The corresponding complete result is indicated by the point $(\cdot)$.

of $\theta = 100$, we let the stopping time c take the values $50, 80$ and 500, with this final value effectively representing the complete case, and hence enabling us to compare results with the complete case of the previous section. Results are presented in Table 6.2. We see there is consistently good agreement between simulated and theoretical results, for all sets of parameter values and sample sizes considered. We also see that, for a given a, τ and N, on the whole, the standard deviation of $\widehat{B}_{10}$ decreases as c increases, with results for $c = 500$ being very close to their counterparts for the complete case in Table 6.1, on allowing for minor differences due to simulation variation. We also noted that the relative bias in $\widehat{B}_{10}$ was less than 2% for every combination of parameter values and sample size considered. As with the standard deviation of $\widehat{B}_{10}$, the bias was, on the whole, seen to decrease as the stopping time rose and, independently, as the sample size increased.

6.5 Design Optimisation: ALT Complete

In this section we discuss complete data sets using an accelerated framework. We use the accelerated Burr XII EFI matrix for complete data, derived in section 3.1.3, while, in section 6.2 above, we determined $B_{10,\pi}$ as the vector of coefficients of the MLEs for the accelerated model, appropriate here. We examine the impact upon the theoretical standard deviation of $\widehat{B}_{10}$ by making modifications to some changeable quantities, as discussed above. We begin with a theoretical investigation into the standard deviation of $\widehat{B}_{10}$, and then present a series of simulations, assessing their agreement with theoretical counterparts.

6.5.1 Theoretical Considerations

In this section we consider the impact upon the theoretical standard deviation of $\widehat{B}_{10}$ by changing the values of the stress levels. We use three stress levels, firstly looking at the

			Simulated Value					Theoretical
a	τ	c	$N = 100$	$N = 300$	$N = 600$	$N = 1500$	$N = 3000$	Value
2	2	50	29.550	33.326	32.376	31.721	32.890	32.495
		80	30.993	31.687	31.513	31.787	31.781	31.686
		500	30.310	31.416	30.916	31.245	31.134	31.129
2	5	50	81.368	86.237	89.151	87.346	90.565	89.476
		80	30.037	30.352	30.541	30.806	30.801	30.708
		500	29.016	29.416	29.567	29.997	29.614	29.859
4	2	50	19.937	22.485	21.844	20.489	20.488	21.925
		80	21.374	22.411	21.733	21.594	21.508	21.852
		500	21.164	21.214	21.165	21.546	21.427	21.363
4	5	50	26.071	26.921	27.233	27.678	27.766	29.713
		80	25.045	24.963	24.985	25.850	25.686	26.000
		500	24.165	25.246	25.374	25.469	25.274	25.341
6	2	50	16.153	17.557	17.404	16.546	17.979	17.763
		80	15.979	16.871	17.018	17.628	17.295	17.572
		500	16.724	16.843	17.210	17.267	17.389	17.267
6	5	50	24.452	24.324	26.425	25.677	26.496	26.969
		80	22.123	22.064	23.038	23.517	23.423	23.797
		500	22.498	22.516	22.847	22.998	23.216	23.162

Table 6.2: Theoretical and simulated values of $\sqrt{N} \times$ standard deviations of $\widehat{B}_{10}$ for various a, τ, sample sizes, and stopping times.

effect of changing the value of the middle stress and then changing the end stresses. We begin by considering the intuitive outcome of this experiment. In an ideal scenario we would test all items at the design stress and wait until all have failed; effectively a non-accelerated experiment, as discussed in section 6.3. Thus, intuition says that the best results will be obtained when the ALT case is as close to the non-ALT case as is possible.

Changing the middle stress level

We first consider the scenario where $k = 3$, with the upper and lower stresses fixed, and with equal numbers of items tested at each of the three levels. To maintain consistency, we again set $a = 2$ and $\tau = 4$, while $x_1 = 150, x_3 = 190$ and α_b, β_b, x_d and the sample sizes are as defined in Simulation 1 above. This ties in with results from section 6.3 above, where we saw the true value of B_{10} for the non-ALT complete case was $14121.663\cdots$. Substituting the true parameter values here, into (6.1), we again determine B_{10} to be 14121.66, as anticipated.

We now consider what happens to the theoretical standard deviation of $\widehat{B}_{10}$ when x_2 ranges from 151 to 189, see Figure 6.4. We see the smallest standard deviation of $\widehat{B}_{10}$ is obtained when the middle stress level is as small as possible; here the standard deviation is 754.1856, which is 38% larger than the value obtained for the non-ALT complete case; recall that this value was 545.248. This seems to be suggesting that we should take just two stress levels. We consider this scenario next.

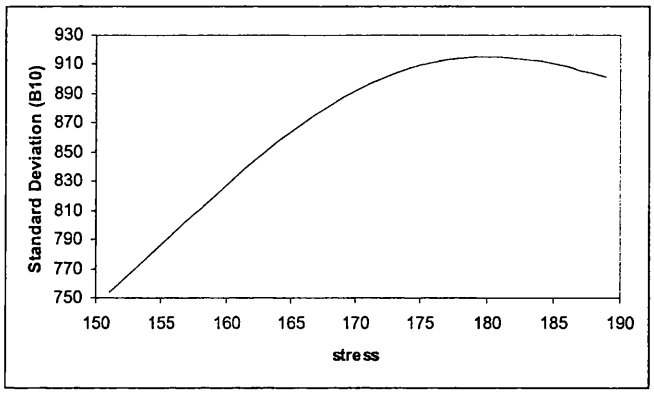

Figure 6.4: Theoretical standard deviation of $\widehat{B}_{10}$ when changing the middle stress level.

Two stress levels

We have $k = 2$, with stress levels $(x_1, x_2) = (150, 190)$, and now consider a range of values for n_1, with the remaining $300 - n_1$ items being tested at x_2. Now, in section 3.1.3.3, we derived all the terms necessary to formulate the EFI matrix for an ALT Burr XII model with completely failed data. In section 3.2, we then saw how it was possible to simplify this matrix for equal sample sizes across the stress levels. However, here, as in general, $n_1 \neq n_2$, so we must consider the terms S_x and S_{xx} in full. For example, when $n_1 = 130$, $n_2 = 300 - 130 = 170$ and

$$S_x = \sum_{i=1}^{k} x_i n_i = (150 \times 130) + (190 \times 170) = 51800,$$

while

$$S_{xx} = \sum_{i=1}^{k} x_i^2 n_i = \left(150^2 \times 130\right) + \left(190^2 \times 170\right) = 9062000.$$

Of course, for the complete case, S_n is simply the sum of the number of items and will be N, regardless of the allocation of these items over the stress levels. We can then proceed to formulate the EFI matrix for the ALT Burr XII model with completely failed data, by substituting these values for S_x, S_{xx} and S_n, together with the true parameter values above, into the expectations of second derivatives, determined in section 3.1.3.3. It is then just a case of evaluating (6.8), with the elements of $B_{10,\pi}$, previously derived in section 6.2 above.

From Figure 6.5, the lowest theoretical standard deviation of $\widehat{B}_{10}$ is 739.3053 which is obtained when $n_1 = 225$. We see, that by varying the allocation of the sample size across the stresses, we can reduce the standard deviation of $\widehat{B}_{10}$ from that which was observed for constant sample sizes above; there the figure was 754.1856. However, this figure of 739.3053 is still significantly higher than our benchmark figure of 545.248, observed for the non-ALT,

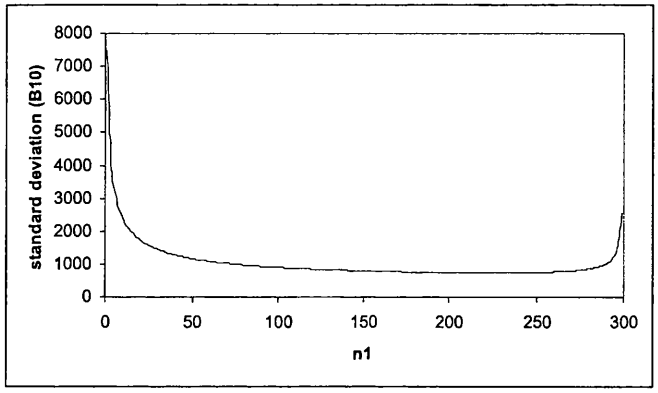

Figure 6.5: Theoretical standard deviation of $\widehat{B}_{10}$ for sample size $(n_1, 300 - n_1)$.

complete case.

So far we have seen that, to minimise the theoretical standard deviation of $\widehat{B}_{10}$, we should have no items on the middle of our three stress levels, effectively suggesting that we should consider only two stresses. For an overall sample size of 300 we saw that we should allocate our two stress levels $(x_1, x_2) = (150, 190)$ with sample sizes $(n_1, n_2) = (225, 75)$. That is, about $\frac{3}{4}$ of the total sample size should be placed at the lowest stress level. We now look to see to what extent this finding is true for other sample sizes; the results are presented in Table 6.3. It certainly seems the case that we should place $\frac{3}{4}$ of the total sample size at the first stress level and the remaining $\frac{1}{4}$ at the final stress level; irrespective of the sample size. Of course, to validate such a result in a general and theoretical sense requires a more detailed study of the full algebraic form of (6.8). Although this result is readily obtainable through the use of appropriate computer software, its simplified form is by no means easily manipulated. Consequently, it is felt that a full investigation is beyond the scope of the current discussion.

Changing the end stress level We next assume that we have an overall sample size of 300, which we shall allocate as (225,75). We look to see the effect on the theoretical standard deviation of $\widehat{B}_{10}$ of varying the second stress level. Firstly, consider the scenario where the first stress level varies between the design stress $x_d = 130$ and 150, that is, $(x_1, x_2) = (x_1, 190)$, where $131 \leq x_1 \leq 150$; see Figure 6.6. Here, the smallest standard deviation of $\widehat{B}_{10}$ is 573.4771, which occurs when $x_1 = 131$ and this corresponds to only a 5% increase over the non-ALT complete figure of 545.248. So, we can reduce the standard deviation further by letting x_1 get smaller and smaller. Conversely, we next consider $(x_1, x_2) = (131, x_2)$, where $151 \leq x_2 \leq 190$; see Figure 6.7. Here, the smallest standard deviation of $\widehat{B}_{10}$ is 573.4771, which occurs when $x_2 = 190$. It seems then, that for this allocation of items, the minimum standard deviation of $\widehat{B}_{10}$ is achieved when x_1 is as small as possible and x_2 is as

N	Most Effective (n_1, n_2)	Standard Deviation of $\widehat{B}_{10}$	$\frac{n_1}{N}$
10	(7, 3)	4063.97	0.7
20	(15, 5)	2863.62	0.75
30	(22, 8)	2338.9	0.73
40	(30, 10)	2024.67	0.75
50	(37, 13)	1811.21	0.74
60	(45, 15)	1653.14	0.75
100	(75, 25)	1280.51	0.75
200	(150, 50)	905.46	0.75
300	(225, 75)	739.305	0.75
500	(375, 125)	572.663	0.75
1000	(750, 250)	404.934	0.75

Table 6.3: Distribution of sample sizes that results in the minimum standard deviation of $\widehat{B}_{10}$ for various total sample sizes.

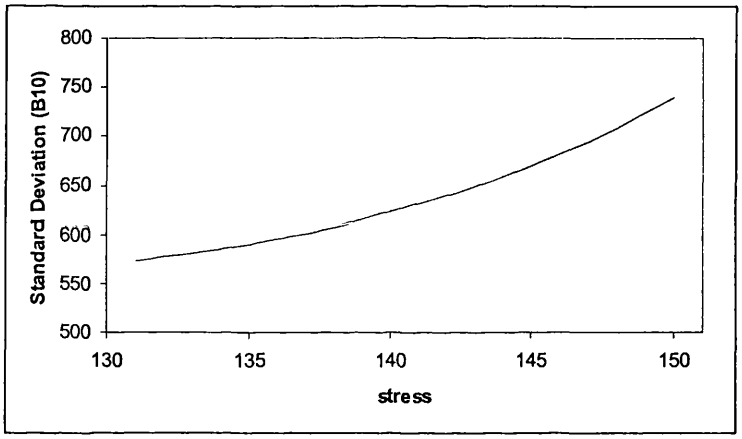

Figure 6.6: Theoretical standard deviation of $\widehat{B}_{10}$ for $(x_1, 190)$.

large as possible. Let us now return to the study of the allocation of the sample items; we thus take $(x_1, x_2) = (130, 190)$ and let $(n_1, n_2) = (n_1, 300 - n_1)$, where $1 \leq n_1 \leq 299$, with results being shown in Figure 6.8. Here, the smallest standard deviation of $\widehat{B}_{10}$ is 545.5030 which occurs when $n_1 = 299$, and, since this experiment is essentially testing practically all items at the lower stress of 130, the resulting standard deviation is correspondingly close to our benchmark figure of 545.248. We see we have reduced the standard deviation of $\widehat{B}_{10}$ even further by having as large a sample size as possible on stress level x_1. It appears then, that the proportionate allocations observed above - where we concluded that three quarters of the items should be tested at the lower stress level - were influenced by the distance between the stress levels and the design stress. For example, a similar chart for $(x_1, x_2) = (140, 190)$ is shown in Figure 6.9. Here, the smallest theoretical standard deviation of $\widehat{B}_{10}$ is 614.0558 which occurs when $n_1 = 257$, and this constitutes a 12% increase over the minimum figure of 545.248.

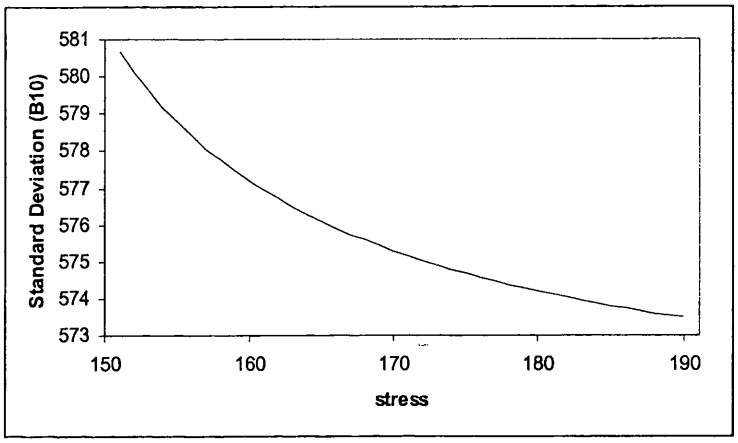

Figure 6.7: Theoretical standard deviation of $\widehat{B}_{10}$ for $(131, x_2)$.

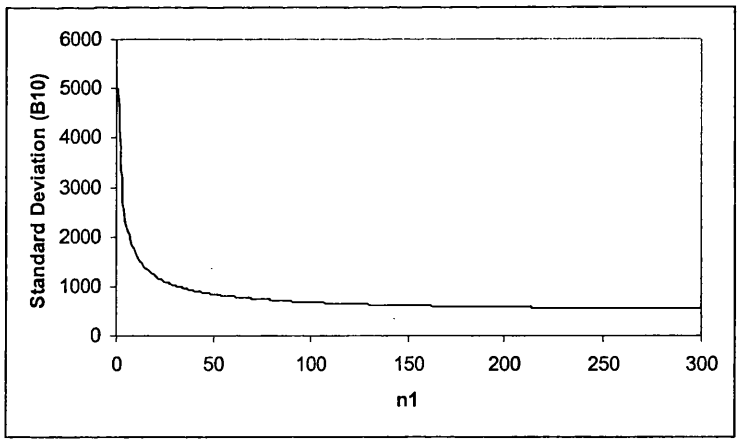

Figure 6.8: Theoretical standard deviation of $\widehat{B}_{10}$ for sample size $(n_1, 300 - n_1)$ and stresses $(130, 190)$.

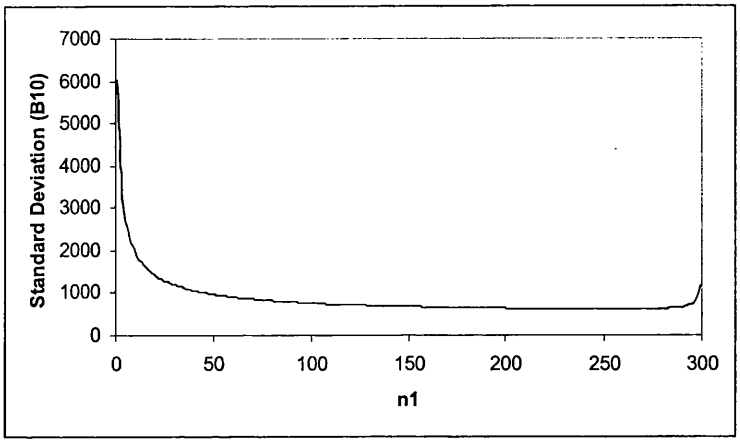

Figure 6.9: Theoretical standard deviation of $\widehat{B}_{10}$ for sample sizes $(n_1, 300 - n_1)$ and stresses $(140, 190)$.

In summary then, our theoretical results appear to suggest that we should use one stress level, which should be as close to the design stress as possible, which concurs with the intuitive argument given at the beginning of this section. Of course, these theoretical investigations are only concerned with the standard deviation of $\widehat{B}_{10}$, ignoring the timing element. In an ideal world, we would test all items at the design stress; however, in practice, we are often restricted by time, and, as noted at the outset of this discussion, are obliged to consider and analyse censored and/or accelerated data.

6.5.2 Simulations

The work above is from a theoretical viewpoint only, and we should seek some reassurance that this theory is in agreement with practice, as represented by simulations. We therefore consider a wide range of possible values for parameters, stress levels and their sample allocations. As before, we consider the following parameter values

$$\alpha_b = 17.60139, \beta_b = -0.056282, x_d = 130.$$

Table 6.4 presents simulated - $s(\cdot)$- and theoretical - $t(\cdot)$ - standard deviations of $\widehat{B}_{10}$ for various sample sizes and the allocation of those sample sizes across stresses, for $a = 2, 4, 6$ and $\tau = 4, 7$. Here, we have three stress levels, $(x_1, x_2, x_3) = (150, 160, 190)$. We see that for a given overall sample size, the standard deviation of $\widehat{B}_{10}$ is reduced by placing more items at lower stress levels. Additionally, the standard deviation is reduced for a larger overall sample size, as expected.

The same pattern is observed in Tables 6.5 and 6.6, for four stress levels. Contrasting corresponding points in the two tables, we see that the standard deviations are lower in the first than in the second. This ties in with the theory above where we observed that the

a	2		4		6	
τ	4	7	4	7	4	7
$s\,(100,100,400)$	712.05	557.67	535.71	447.75	468.36	487.04
$t\,(100,100,400)$	703.27	549.31	531.09	447.44	460.45	405.34
$s\,(200,200,200)$	580.27	455.45	588.65	381.11	401.23	377.19
$t\,(200,200,200)$	584.81	456.78	449.07	378.34	392.17	345.23
$s\,(500,500,500)$	373.11	291.05	286.09	239.12	248.95	217.21
$t\,(500,500,500)$	369.86	288.89	284.01	239.28	248.03	218.34
$s\,(1000,1000,1000)$	265.52	205.67	200.13	168.32	173.05	154.87
$t\,(1000,1000,1000)$	261.53	204.27	200.83	169.19	175.38	154.39
$s\,(500,500,2000)$	310.88	246.41	239.33	199.94	206.29	180.88
$t\,(500,500,2000)$	214.51	245.66	237.51	200.10	205.92	181.27

Table 6.4: Simulated (s) and theoretical (t) standard deviations of $\widehat{B}_{10}$ for various sample sizes and parameter values, for $(x_1, x_2, x_3) = (150, 160, 190)$.

higher the values of the stress levels - and therefore the further we are from the design stress - the higher the standard deviation of $\widehat{B}_{10}$. We also note that the relative bias in $\widehat{B}_{10}$ was less than 0.5% for every combination of parameter values and sample sizes considered.

Clearly, our investigations so far have centred on one fundamental experiment, and, particularly from the theoretical angle, only three - and subsequently two - stress levels. Naturally, we could extend this work to accommodate four, five or even more stress levels and all permutations of sample sizes. However, we have already seen that for best results - that is a small as possible standard deviation of $\widehat{B}_{10}$ - the values of the stress levels should be as close to the design stress as possible. The extension to include additional stress levels would naturally present us with the same conclusions. The obvious, and seemingly most fruitful, area of investigation is when we employ censoring, and are therefore in a position to examine potential trade-offs between stopping time and standard deviation. We look at this next.

6.6 Design Optimisation: ALT Type I Censoring

Naturally, a censored data set will provide us with less information regarding B_{10} than a completely failed one. The issue we deal with here is the extent to which we are penalised statistically by introducing censoring So, as with the non-ALT case in section 6.4 above, we are interested in finding a trade-off between the stopping time and the theoretical standard deviation of $\widehat{B}_{10}$. We would imagine, for instance, that for a given stopping time, we could reduce the standard deviation of $\widehat{B}_{10}$ by increasing the values of the stress levels - thereby inducing earlier failure - or increasing the sample size in order to give a better representation of the underlying population. However, we may also, for example, expect to see a favouring of lower stress levels as the stopping time increases: with more information being obtained from the data set, we would be happy to set lower values for the stress levels, expecting

a		2		4		6	
τ		4	7	4	7	4	7
s	$\begin{pmatrix}150,150,\\600,600\end{pmatrix}$	490.27	381.76	371.44	307.69	334.05	281.45
t	$\begin{pmatrix}150,150,\\600,600\end{pmatrix}$	491.28	383.73	368.44	310.41	318.45	280.34
s	$\begin{pmatrix}300,300,\\1200,1200\end{pmatrix}$	349.65	271.10	260.27	217.19	227.29	198.21
t	$\begin{pmatrix}300,300,\\1200,1200\end{pmatrix}$	347.39	271.34	260.53	219.49	225.18	198.23
s	$\begin{pmatrix}375,375,\\375,375\end{pmatrix}$	406.06	324.66	313.95	260.21	269.21	313.95
t	$\begin{pmatrix}375,375,\\375,375\end{pmatrix}$	407.78	318.51	310.16	261.31	269.76	237.47
s	$\begin{pmatrix}750,750,\\750,750\end{pmatrix}$	287.64	225.48	217.94	182.53	190.26	168.00
t	$\begin{pmatrix}750,750,\\750,750\end{pmatrix}$	288.35	225.22	219.32	184.77	190.75	167.92

Table 6.5: Simulated (s) and theoretical (t) standard deviations of $\widehat{B}_{10}$ for various sample sizes and parameter values, for $(x_1, x_2, x_3, x_4) = (150, 160, 170, 190)$.

a		2		4		6	
τ		4	7	4	7	4	7
s	$\begin{pmatrix}150,150,\\600,600\end{pmatrix}$	592.67	469.48	444.94	374.75	386.30	334.55
t	$\begin{pmatrix}150,150,\\600,600\end{pmatrix}$	596.83	466.18	443.01	373.23	381.09	335.48
s	$\begin{pmatrix}300,300,\\1200,1200\end{pmatrix}$	423.42	331.22	313.24	263.28	269.56	238.87
t	$\begin{pmatrix}300,300,\\1200,1200\end{pmatrix}$	422.02	329.64	313.25	263.91	269.47	237.22
s	$\begin{pmatrix}375,375,\\375,375\end{pmatrix}$	442.84	342.23	334.33	283.78	293.49	255.71
t	$\begin{pmatrix}375,375,\\375,375\end{pmatrix}$	443.48	346.40	334.98	282.22	290.45	255.69
s	$\begin{pmatrix}750,750,\\750,750\end{pmatrix}$	311.16	246.32	238.35	199.28	202.92	179.46
t	$\begin{pmatrix}750,750,\\750,750\end{pmatrix}$	313.59	244.94	236.86	199.56	205.38	180.80

Table 6.6: Simulated (s) and theoretical (t) standard deviations of $\widehat{B}_{10}$ for various sample sizes and parameter values, for $(x_1, x_2, x_3, x_4) = (150, 170, 180, 190)$.

these conditions to be more indicative of those seen at the design stress.

As with complete data, we begin with a theoretical exploration, considering a single example in which we vary the allocation of the overall sample size, N, and change the values of the stress levels, examining their impact upon the theoretical standard deviation of $\widehat{B}_{10}$. We then consider a range of compromises: we look to see what sort of percentage increase can be obtained between the theoretical standard deviation of $\widehat{B}_{10}$ for a certain stopping time c and that obtained for the equivalent completely failed data set. We then present a series of simulations for three and four stress levels, and various values of the shape parameters a and τ. As before, the aim of these simulations is to provide us with increased confidence that the theoretical expressions we use are sound and are providing us with correct results.

Of course, the scope of these practical investigations is rather narrow; real life experiments are many and varied, and we cannot hope to cover the wide range of possible parameter combinations that are used in the real world. Rather, the principle followed here is to give guidance to a practitioner on how to formulate useful results for a given set of parameter values, prior to running an experiment.

6.6.1 Theoretical Considerations

Here, we use the same vector $B_{10,\pi}$ of parameter coefficients as for the ALT complete case above, with elements given by (6.3) to (6.6). Expectations of second derivatives for the ALT Burr XII Type I censored model are presented in section 5.5; from these, we can then formulate the corresponding EFI matrix, and from (6.8), the theoretical standard deviation of $\widehat{B}_{10}$.

As before, we start by considering the parameter values $a = 2$, $\tau = 4$, $\alpha_b = 17.60139$, $\beta_b = -0.056282$ and $N = 300$, equally allocated across the three stress levels $(x_1, x_2, x_3) = (150, 160, 190)$. Again, the design stress is $x_d = 130$ hours, and we let the stopping time c range from 1000 to 12000 hours; see Figure 6.10. This also shows the corresponding theoretical standard deviation of $\widehat{B}_{10}$ for completely failed data, and this is 827.039 (see Figure 6.4 at $x_2 = 160$), which, we note, is 51% larger than non-ALT complete result of 545.248. The upper limit of c was chosen to be practically equivalent to a completely failed data set.

From Figure 6.10, it is apparent that we should be in a position to make a compromise between the stopping time and the theoretical standard deviation of $\widehat{B}_{10}$. We see that from around $c = 6000$ onwards, there is very little change in the theoretical standard deviation of $\widehat{B}_{10}$. In fact, it transpires that there is only an 9% difference between the standard deviation observed at $c = 6000$ (where this quantity is equal to 904.642) compared to that observed at $c = 12000$ (= 829.565); this difference drops to 4% at $c = 7000$. Since these values are based on a real experiment which was measured in hours, this corresponds to a reduction in running time of around 5000 hours (or about 7 months), from the complete experiment of 12000 hours. Put into a more practical context, if this experiment was to run

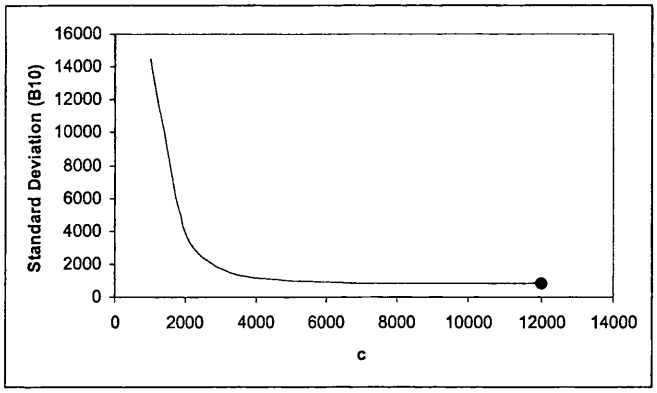

Figure 6.10: Theoretical standard deviation of $\widehat{B}_{10}$ for increasing values of the stopping time, c and with stresses $(150, 160, 190)$. The corresponding complete result is given by the single point $(\cdot)$.

to completion it would take, around 16 months (approximately 12000 hours.) If, instead, we ran the experiment for about 9 months (7000 hours), there would be an increase of only 4% in the standard deviation of $\widehat{B}_{10}$.

Trade-Offs

We next consider the case where we have the same parameter values as in the previous example, but with the minor modification to the stress levels, $(x_1, x_2, x_3) = (150, 170, 190)$; see Figure 6.11.

As in the previous example, we see how the theoretical standard deviations for the censored model tend to the corresponding complete result as c increases. Now, the standard deviation at $c = 8000$ is 923.467, which is only a 3.5% increase over the complete figure of 892.104 (represented by the single point), which, in turn, is 63% larger than the benchmark figure of 545.248 hours obtained for the non-ALT complete case, in section 6.3. We now look to see the extent to which we can obtain a similar theoretical standard deviation to that obtained at $c = 8000$, by shortening the running time of the experiment and by changing the values of the stress levels and the allocation of the sample across stress levels, while maintaining an overall sample size of $N = 300$. A sample of results is presented in Table 6.7.

These examples illustrate how we can get approximately equivalent standard deviations, by making minor adjustments to the values of the stress levels and the allocation of the sample sizes at those levels. As a further illustration, we use the same parameter values, with stress levels $(150, 170, 190)$ and an equally distributed, total sample size of $N = 300$, and show how the percentage error between the censored result and the complete result

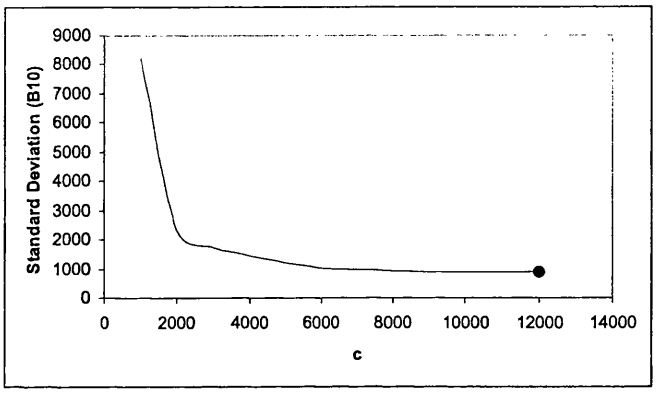

Figure 6.11: Theoretical standard deviation of $\widehat{B}_{10}$ for increasing values of the stopping time, c and with stresses $(150, 170, 190)$.

(x_1, x_2, x_3)	(n_1, n_2, n_3)	c	Standard Deviation $\left(\widehat{B}_{10}\right)$
$(155, 157, 190)$	$(150, 50, 100)$	5000	925.826
$(155, 157, 190)$	$(200, 40, 60)$	5000	926.617
$(155, 160, 190)$	$(110, 60, 130)$	6000	929.223
$(155, 160, 190)$	$(150, 20, 130)$	5500	927.191
$(155, 160, 210)$	$(140, 10, 150)$	4500	923.848

Table 6.7: Variations in stopping time, values of stress levels and the allocation of the total sample size across those levels, with similar theoretical standard deviations of $\widehat{B}_{10}$.

decreases for increasing values of c; we have

c	5000	7000	9000	11000	13000	15000	17000	19000	21000
% error	35.69	7.69	1.77	0.56	0.22	0.09	0.04	0.02	0.01

So, depending on the acceptable tolerance level, we can then choose the appropriate stopping time for the experiment.

These results also show, as with the results for complete data sets, that the ideal scenario - from a statistical perspective - is to test as many items as possible at one stress level, and that this stress level should be as close to the design stress as possible. Thus, if an engineer says, "I have 100 items, how should I distribute them to minimize the theoretical standard deviation of $\widehat{B}_{10}$?", then one answer is to, "Pick the stress level that equates to a time that is as long as you can wait, and test all items at that level." Of course, the practical difficulty with this particular scenario arises when we need to interpolate back to the design stress, as we are faced with fitting a scale-stress relationship on the basis of information obtained at one stress level. For (1.12), we need at least two stresses; in practice, we might use three to provide some evidence that this relationship is a sensible one.

From the above, we can see that when c is small, we should place as many test items as possible at the highest stress setting, since these items are most likely to fail first and therefore the data will contain a greater proportion of failures. As c increases, it seems sensible to assume that more items should be placed at lower stresses, since we can now expect to observe more failures at a stress closer to the optimum or design stress. To illustrate this point, consider Figure 6.12, which shows how the theoretical standard deviation of $\widehat{B}_{10}$ varies for the two separate sample allocations $(200, 50, 50)$ and $(50, 50, 200)$; $a, \tau, \alpha_b\beta_b, x_d$ and p are unchanged from the previous example and $(x_1, x_2, x_3) = (150, 170, 190)$.

Note, up to a stopping time of about 2100 hours we prefer the sample allocation $(50, 50, 200)$, whereas after this time we prefer the alternative sample allocation, which agrees with the assumptions we made in the previous paragraph. The interesting point to note is that we can obtain more information about $\widehat{B}_{10}$ using the allocation $(200, 50, 50)$ as opposed to $(50, 50, 200)$, when the stopping time is still relatively small; indeed still less than the average failure time at the middle stress level, determined as

$$\exp(17.60139 - 0.056282 \times 170) = 3082.$$

Now, for example, consider alternative values for the middle stress level, with sample allocation $(50, 200, 50)$; see Figure 6.13.Yet again, we see trade-offs as c increases. For small c, we should choose a large value for the middle stress level; as we obtain more and more information from the data, so we can reduce the value of the stress level nearer to the desirable design stress.

To asses the theoretical validity of these results, we now consider the usual set of parameter values together with a sample allocation of $(n_1, n_2, n_3) = (200, 50, 50)$, and stress levels

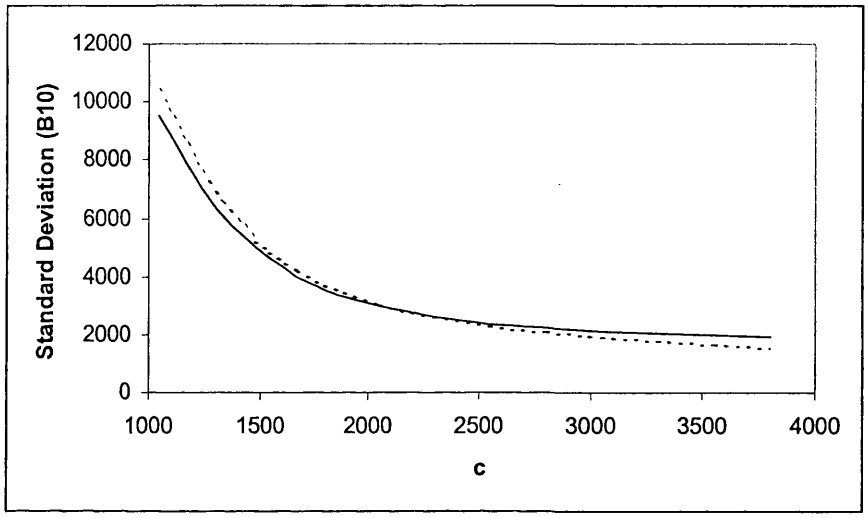

Figure 6.12: Theoretical standard deviation of $\widehat{B}_{10}$ for increasing values of c, under two sample distributions, (n_1, n_2, n_3): $(200, 50, 50)$ is represented by the broken line while $(50, 50, 200)$ is represented by the continuous line.

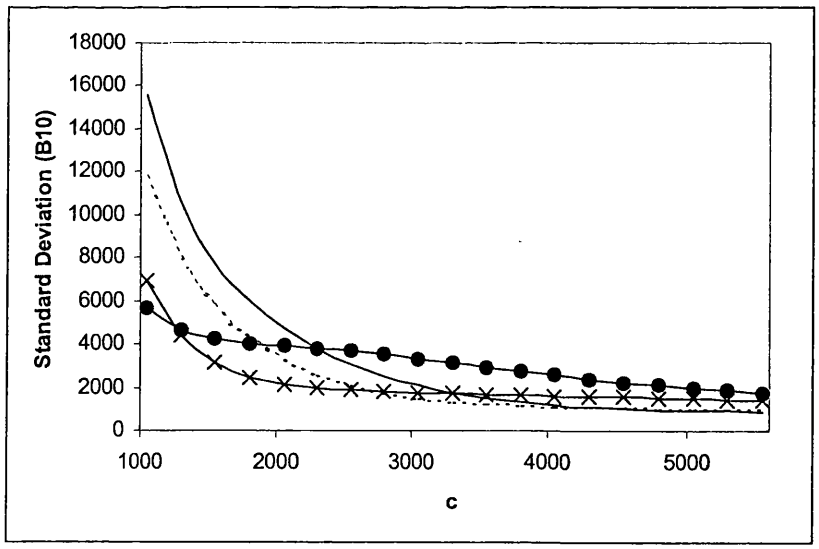

Figure 6.13: Theoretical standard deviation of $\widehat{B}_{10}$ for increasing values of c, when $x_1 = 150, x_3 = 190$ and x_2 takes on the four values, 155 (continuous line), 160 (broken line), 170 ($\times$) and 180 ($\cdot$).

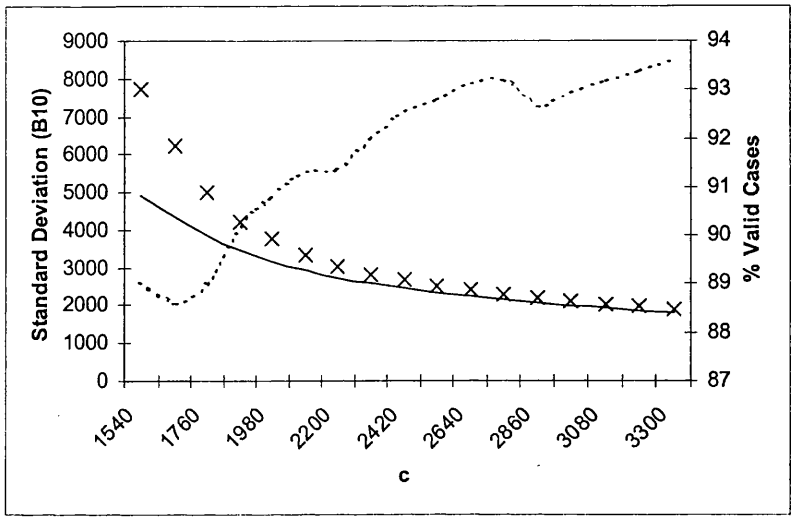

Figure 6.14: Theoretical and observed standard deviations of $\widehat{B}_{10}$ for increasing stopping time c. Simulated values are represented by $\times$, while the correpsonding theoretical values are indicated by the continuous line. The broken line shows the percentage of valid cases (out of 10000) that contributed to the simulated result.

$(x_1, x_2, x_3) = (150, 170, 190)$; we let c vary from 1540 to 3300 and see where the simulated and theoretical standard deviations of $\widehat{B}_{10}$ begin to disagree. The results are presented in Figure 6.14. We see, that for small c, there is a marked discrepancy between the simulated and theoretical standard deviations. It seems that, while the theory is endeavouring to accomodate for the fact that the standard deviation of $\widehat{B}_{10}$ will increase with the amount of censoring, due to the nature of the algorithm - which is a minor adaptation of that outlined in Chapter three for complete data sets - its ability to converge on $\widehat{a}$ is limited for small c. Consequently, the theory and simulations do not always match up well for small values of c. Nevertheless, depending on the nature of the experiment, at the crucial points - where we are interested in potential trade-offs between censoring time and standard deviation of $\widehat{B}_{10}$ - there is some agreement between the simulated values and their theoretical counterparts. For example, consider the experiment illustrated by Figure 6.12, where the intersection of the two lines occurs at approximately $c = 2100$. From Figure 6.14, we see that the agreement between simulated and theoretical values is actually rather good at $c = 2100$. As such, there is clearly considerable value behind the detailed theoretical investigations summarised here.

As a further consideration, we can also compare the Type I censored ALT case with the Type I censored non-ALT case. For example, consider the usual set of parameter values ($a = 2, \tau = 4, \alpha_b = 17.60139$ and $\beta_b = -0.056282$) with three stress levels $(x_1, x_2, x_3) = (150, 170, 190)$, and a sample allocation $(n_1, n_2, n_3) = (150, 50, 100)$. When $c = 4900$, the theoretical standard deviation of $\widehat{B}_{10}$ is found to be 1008.24. However, we saw in section 6.4, that for the comparable non-ALT case, we obtained a theoretical standard deviation

of 1009.35 when $c = 13000, N = 300$ and all items were tested at the design temperature of $130°C$. We see that, through the introduction of acceleration, we are able to obtain the same information about $\widehat{B}_{10}$ in almost a third of the time. Further, alternative sample allocations and stress level settings, in conjunction with the stopping time, can also result in a theoretical standard deviation of around 1008.24, as we have seen in a similar experiment in Table 6.7.

6.6.2 Simulations

As with complete data, we now provide some validation of the theoretical results through simulation experiments. Using parameter values $a = 2, \tau = 4, \alpha_b = 17.60139, \beta_b = -0.056282$, and with c taking the four values $1015, 2570, 5800$ and 30000, we obtain Figures 6.15 and 6.16. The conclusions reached for complete data are endorsed here: when the sample size on the lower stresses increases, the standard deviation of $\widehat{B}_{10}$ decreases. As expected, we also see the standard deviation of $\widehat{B}_{10}$ decrease for larger values of c. Note, for the sample allocation $(n_1, n_2, n_3) = (100, 100, 400)$, we have no observed values for $c = 1015$ in Figure 6.15. This is due to difficulties in converging on estimates for parameter values in what are heavily censored data sets. Two reference points have been included in each figure - in the scale of the original data - to confirm that lower standard deviations are observed for stress levels $(x_1, x_2, x_3) = (150, 160, 190)$ than for $(x_1, x_2, x_3) = (150, 180, 190)$, as suggested by the theory above. Charts for four stress levels, at three different stress level settings, with $(n_1, n_2, n_3, n_4) = (375, 375, 375, 375)$, are also presented in Figures 6.17, 6.18 and 6.19. Again, we have omitted results for $c = 1015$, due to the disparity between simulated and theoretical results: inclusion would result in uninterpretable charts. On the whole, we observe good agreement between simulated and theoretical results, thereby providing considerable support for the theoretical investigations summarised above. We also note that, for all plotted points, the relative bias in $\widehat{B}_{10}$ is at most 5%, with, perhaps unsurprisingly, the larger biases observed with small c. On average, over all plotted points, the relative bias was less than 1%. Consequently, the theoretical standard deviation of $\widehat{B}_{10}$ - or, more specifically, the first order Taylor series approximation to the theoretical standard deviation of $\widehat{B}_{10}$ - was seen to be very close to the mean square error of $\widehat{B}_{10}$, thereby allowing us to use the theoretical standard deviation in place of the mean square error.

6.7 Summary

We have seen how making adjustments to the values of the stress levels and the sample allocation can reduce the standard deviation of $\widehat{B}_{10}$. We have also seen how, for a given set of parameter values, we can determine the shortest stopping time, c, for a given tolerance level; that is, the relative difference between the theoretical standard deviation of $\widehat{B}_{10}$ for the (non-) ALT censored data set and that which would be observed for a (non-)ALT completely failed data set. In addition, when considering acceleration, we observed, for smaller c,

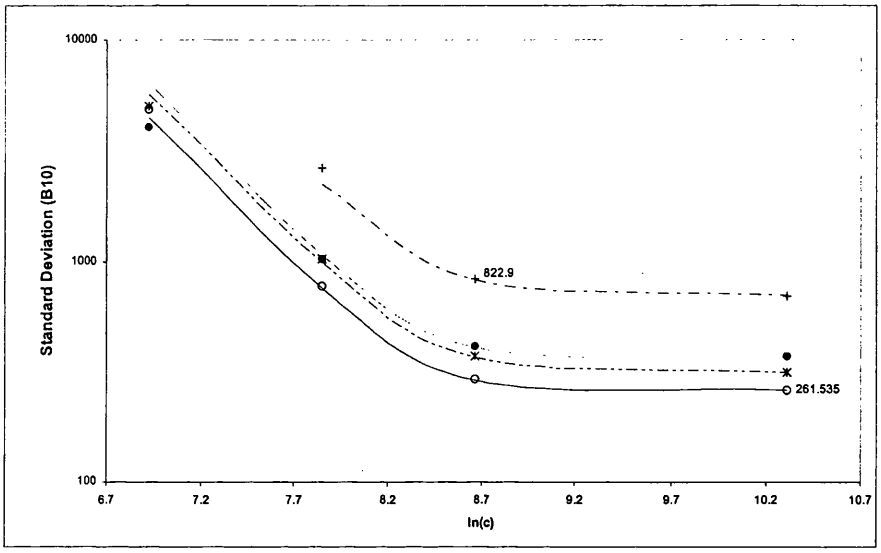

Figure 6.15: Theoretical and observed standard deviations of $\widehat{B}_{10}$ for increasing values of c, with stress levels $(x_1, x_2, x_3) = (150, 160, 190)$. For sample size distributions, theoretical results are represented by lines, while simulated results are indicated by: $+\,(100, 100, 400)$, $\bullet\,(500, 500, 500)$, $*\,(500, 500, 2000)$, $\circ\,(1000, 1000, 1000)$.

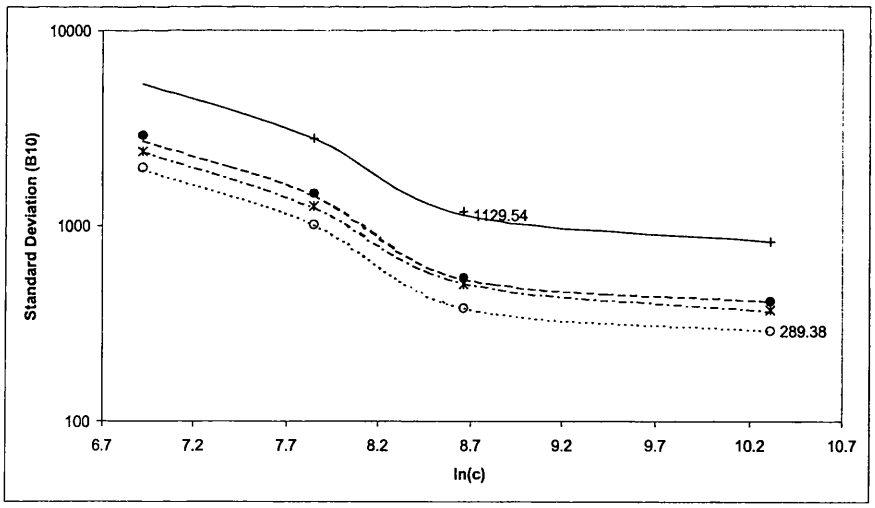

Figure 6.16: Theoretical and observed standard deviations of $\widehat{B}_{10}$ for increasing values of c, with stress levels $(x_1, x_2, x_3) = (150, 180, 190)$. For sample size distributions, theoretical results are represented by lines, while simulated results are indicated by: $+\,(100, 100, 400)$, $\bullet\,(500, 500, 500)$, $*\,(500, 500, 2000)$, $\circ\,(1000, 1000, 1000)$.

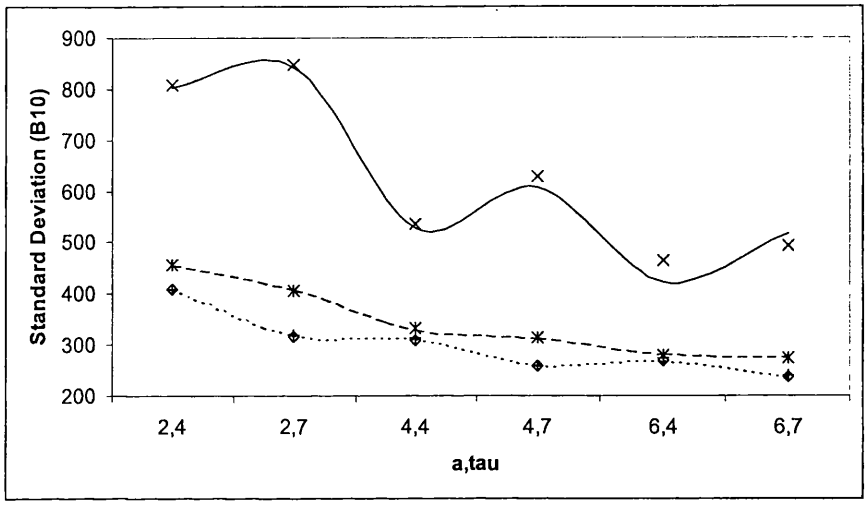

Figure 6.17: Theoretical and obsreved standard deviations of $\widehat{B}_{10}$ for various a, τ, with stress levels $(x_1, x_2, x_3, x_4) = (150, 160, 170, 190)$. For differing values of c, theoretical results are represented by lines, while simulated results are indicated by: $\times\,(c = 2570)$, $*\,(c = 5800)$, $\diamond\,(c = 30000)$.

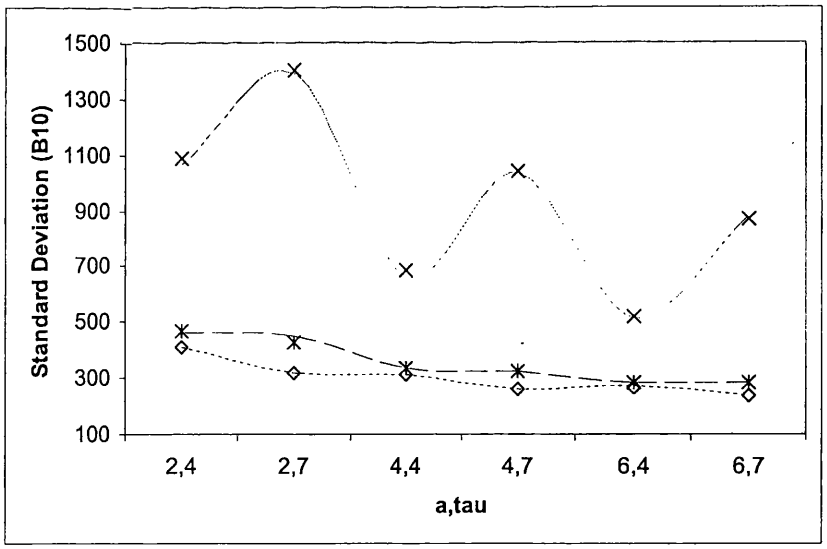

Figure 6.18: Theoretical and obsreved standard deviations of $\widehat{B}_{10}$ for various a, τ, with stress levels $(x_1, x_2, x_3, x_4) = (150, 160, 180, 190)$. For differing values of c, theoretical results are represented by lines, while simulated results are indicated by: $\times\,(c = 2570)$, $*\,(c = 5800)$, $\diamond\,(c = 30000)$.

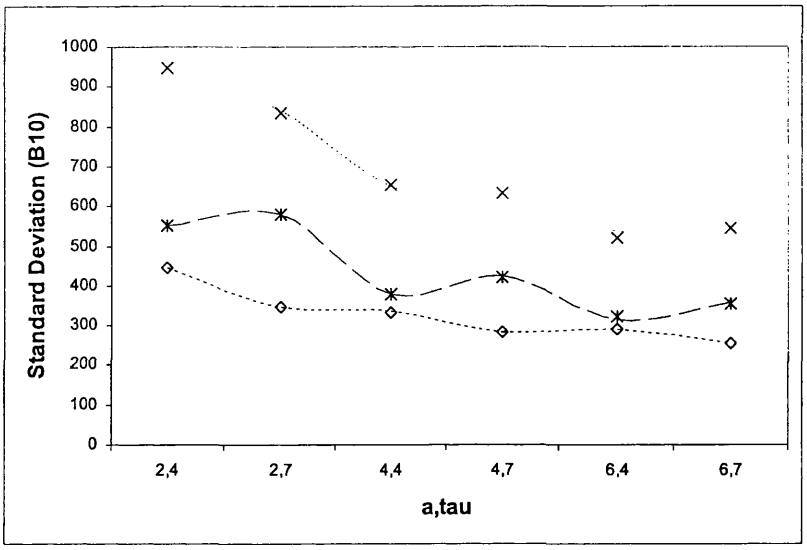

Figure 6.19: Theoretical and observed standard deviations of $\widehat{B}_{10}$ for various a, τ, with stress levels $(x_1, x_2, x_3, x_4) = (150, 170, 180, 190)$. For differing values of c, theoretical results are represented by lines, while simulated results are indicated by: $\times\,(c = 2570)$, $*\,(c = 5800)$, $\diamond\,(c = 30000)$.

a desire to allocate more items to the highest stress levels, gradually favouring a higher sample allocation on the lowest level as c increased. When looking to endorse our theoretical results through simulations, we observe good agreement between the corresponding standard deviations of $\widehat{B}_{10}$ for practically all scenarios considered. However, as c decreased, the large degree of censoring in the data set often made it difficult for the algorithm - see Appendix A - to converge on a sensible value of a. Consequently, a large disparity was observed between simulated and theoretical results for small c.

The principal aim of this chapter is to formulate the theory necessary to draw conclusions about some percentile of the estimated running time of an experiment at its design stress; in particular, we were interested in the standard deviation of $\widehat{B}_{10}$. Consideration of completely failed, non-accelerated data is clearly of limited practical value; it does, however, provide a useful benchmark, in terms of minimum standard deviation of $\widehat{B}_{10}$, both for any given experimental design, and for comparing alternative design scenarios. The introduction of Type I censoring illustrated potential compromises between the stopping time, c, and the standard deviation of $\widehat{B}_{10}$, and this sets the scene for a more rigorous investigation when we introduced acceleration.

As with the non-accelerated case, consideration of a complete ALT experiment may seem to be of limited practical use. However, with the introduction of acceleration, and a suitable choice of stress levels, a completely failed data set will offer itself for analysis sooner than without acceleration. Naturally, the theory consistently favours a design where there is no acceleration, placing all items at the design stress and waiting for complete failure. However,

in practice, we know that we cannot always wait for all items to fail, either at the design stress or at some higher stress level, and, so, we consider censoring. At the same time, to fit the log-linear scale-stress relationship, (1.12), we need at least two stress levels; practical considerations mean that three levels are often used, see, for example, the carbon fibre rod data (example 2.1.1.1), the aluminium coupon data (example 2.1.1.2) and the electrode data (example 2.1.1.3). The most fruitful area of investigation, therefore, concerns acceleration and Type I censoring. In particular, we observe several instances of trade-offs between the stopping time and the theoretical standard deviation of $\widehat{B}_{10}$, and between competing design scenarios; that is, combinations of stress level settings, sample allocations and stopping times. In particular, we present an example in which a number of different design scenarios returned approximately the same theoretical standard deviation of $\widehat{B}_{10}$. We also witnessed the great advantage of coupling acceleration with Type I censoring, with one particular ALT experiment offering the same information on $\widehat{B}_{10}$, in less than a third of the time of its non-ALT Type I counterpart.

However, we must always bear in mind that these conclusions are based on a relatively small number of sets of parameter values and, as such, cannot be claimed to cover every eventuality. Nevertheless, the theory behind the results and the method of approach is general, and can be easily and effectively adapted and employed by practitioners wishing to carry out experiments pertaining to their own circumstances. In addition, we stress this chapter only provides an overview of some of the investigations that are possible and is, therefore, by no means an exhaustive account on the subject. Rather, it is offered as a basis for future discussion and research.

Chapter 7

Summary and Conclusions

In this final chapter, we provide an overview of our work and present our conclusions. We begin by summarising each chapter in turn and then move on to discuss the aims of our work and the extent to which these were met. We then present an overall conclusion and finish by considering further areas of investigation.

7.1 Chapter Summary

Chapter One

In Chapter one we outlined the basic reliability distributions that would be the focus of our work, namely the Weibull and Burr XII, and discussed the link between them. We also gave brief details of additional lifetime distributions and non-parametric approaches to accelerated lifetime modelling. Censoring was then discussed from a Type I viewpoint. Having detailed any standard mathematical functions that would be necessary in our work, we then considered the various ways in which we could link the scale parameters in our models to the stress factor. Due, in large, to its versatility, we opted for the log-linear link. Finally, some computational issues were considered, paying particular attention to the lengthy time taken to run simulations.

Chapter Two

Chapter two was concerned with fitting the relevant reliability distributions to data sets. Throughout the chapter, three distinct data sets were considered: carbon fibre rods, consisting of four stress levels, aluminium coupons, consisting of three stress levels and electrodes in oil, also consisting of three stress levels. We began by fitting the Weibull distribution to subsets of the three data sets and then moved on to fit an accelerated Weibull model to the each complete data set. Next, the three-parameter Burr XII distribution was fitted to each data subset and comparisons were made between these results - in particular between maximised log-likelihoods - and those obtained for the Weibull distribution. We

then presented a new area of research by fitting an accelerated Burr XII model to each of three data sets, and, as with the non-ALT case, made comparisons between these results and those obtained using the accelerated Weibull model. When fitting the Burr XII models - both non-ALT and ALT - to some data (sub)sets, we witnessed one of the shape and scale parameters tending to increase without limit. In these circumstances, the Weibull model was seen to provide a superior fit - in terms of maximised log-likelihood - than the Burr XII model. This prompted further investigations, described in Chapter four.

Chapter Three

The primary aim of Chapter three was to establish the complete-data, expected Fisher information matrices (EFI matrices) for the lifetime models considered in Chapter two. The non-ALT and ALT Weibull distributions were first referenced, before moving on to the two and three-parameter Burr XII distributions. We then derived previously unobtained results for the EFI matrix for the Burr XII ALT model. Being new research, we sought to validate the results for this final model through extensive simulations, where we saw the agreement between simulated and theoretical results improve with increasing sample size, as expected.

Chapter Four

Chapter four essentially carried on where Chapter two left off. Having seen that occasionally the Weibull is preferred to the Burr XII, we wished to find the conditions that dictate which model would provide the superior fit. We began by furthering previous work on non-accelerated models, deriving a form of 'discriminating function' (termed Δ) that enabled us to determine which of the Weibull ALT or Burr XII ALT models would provide the superior fit to a given data set based solely on the Weibull ALT parameter estimates. A positive Δ implied a preference for the Burr XII model, while a negative Δ implied a preference for the Weibull model. The values of Δ found for each of the three data sets studied in Chapter two, concurred with the log-likelihoods observed in that chapter. We then sought the general circumstances under which one model would be preferred to the other. That is, whether a given set of parameter values would dictate which model would provide the superior fit. We found that when the data set originated from a Weibull distribution, the distribution of Δ_s - a scaled form of Δ - became more Normal with increasing sample size (in agreement with the Central Limit Theorem), but also more centred around zero, with a sizeable standard deviation, thus limiting our ability to choose between the models. On the other hand, when the data came from a Burr XII distribution, the distribution of Δ_s, although becoming more Normal with increasing sample size, was never actually observed to be significantly Normal for any of the sample sizes considered. Even so, we saw the mean of Δ_s increase and the standard deviation (as a percentage of the mean) decrease in the same limit. Consequently, as the sample size rose, we could be more confident that we

would observe a positive Δ_s, but due to the consistent lack of Normality - even for sample sizes as large as 15000 - further inferential progress was severely limited. At this point, we drew a line under our theoretical investigation into the behaviour of Δ.

Chapter Five

The extension of previous work to a Type I censored environment was the focus of Chapter five. We began by looking at results for a two-parameter Burr XII distribution, deriving the terms that contributed to the EFI matrix. We then went on to formulate the EFI matrix for the three-parameter Burr XII distribution and the accelerated Burr XII model. Throughout, we ensured the expectations of the score functions were zero. Our attention then turned to the practical implications: the extent to which the standard deviations of the parameter estimates in the Burr XII ALT model were affected by the sample size. For all four parameters, we saw the simulated values decreasing in agreement with theoretical values as the sample size increased. We then proved that the theoretical expectations for the Burr XII ALT model tended to their complete counterparts when the stopping time tended to infinity. This required a brief foray into the subject of complex analysis, and in particular, contour integration, poles and residues. Finally, we derived Δ for censored data.

Chapter Six

In this chapter, we presented an exploration of some of the practical applications of our work. In particular, we focussed on a quantity called $\widehat{B}_{10}$, the estimate of the 10^{th} percentile of life at the design stress. We were interested in finding a combination of stress levels and sample allocations that minimised the approximation to the theoretical standard deviation of $\widehat{B}_{10}$, when the underlying life distribution was Burr XII. We began by looking at complete data without acceleration. From a purely statistical perspective, this is the ideal scenario, where all items are tested at their normal operating conditions until they all fail. We then introduced Type I censoring for non-accelerated data sets, thereby enabling us to assess the trade-off between the stopping time and the theoretical standard deviation of $\widehat{B}_{10}$. We then considered completely failed accelerated data, concluding that the most effective set-up required us to place all test items on one stress level as close to the design stress as possible - effectively an experiment without acceleration. This situation is clearly idealistic, having the inevitable drawback of having to wait until all items have failed, and is therefore of limited use in a world often dictated by severe time constraints. Consequently, we turned our attention to the Type I censored case. Using a theoretical approach, we selected a single experimental set-up, and examined potential compromises between the stopping time c and the theoretical standard deviation of $\widehat{B}_{10}$. We were able to assess points at which we started to prefer one sample allocation or stress level setting over another. We were also able to draw conclusions as to what time we should stop the experiment, depending on what was deemed to be an acceptable tolerance level, that is the difference between the theoretical standard

deviation of $\widehat{B}_{10}$ at the stopping time and that observed for an analogous completely failed ALT experiment. For both the complete and Type I investigations, a series of simulations were presented to ensure that the theoretical conclusions reached were valid. Naturally, for small c, there was some disparity between simulated and theoretical results; a consequence of the algorithm used. It is possible that there are alternative algorithms and methods of converging on parameter estimates that may provide better agreement at these early stopping times, but we found that in the large majority of cases, and for moderate to large c, agreement between simulations and theory was good. In a similar vein, we remark here that the algorithm used to obtain MLE of parameters in the ALT Burr XII model, was less robust for values of a less than two. Again, alternative algorithms, particularly those adopting a more rigorous numerical approach to finding MLE, may prove to be more satisfactory. However, we found our algorithm to be effective in the majority of cases considered.

Since investigations in Chapter six were largely based on one experimental design, to discuss numerical conclusions would be of limited use. However, the intrinsic value of the chapter comes from providing a practitioner with a guide to carrying out detailed theoretical explorations, before embarking on the time consuming business of running the experiment.

7.2 Conclusions

With the Weibull model being particularly popular in the field of accelerated lifetime modelling, the principle aim of this thesis was to consider a lifetime distribution more parameter rich than the Weibull, and make an assessment on its ability to provide a superior fit than the Weibull, for given failure data and a given censoring regime. Due to the limiting relationship that exists between the Weibull and Burr XII distributions, and the well established asymptotic results for maximum likelihood estimates, this model and approach seemed a sensible choice. Ultimately, we wanted to establish the variation - or some estimate of it - in the expected time to failure of some percentile of life at some specified design stress using the Burr XII model, and this became the focus of Chapter six. In order to investigate this, we were required to formulate the expected Fisher information matrices for the Burr XII model with complete (Chapter three) and Type I (Chapter five) censoring; the inverses of the matrices providing us with asymptotic estimates of the standard deviations of the maximum likelihood estimates of the model parameters. Having previously observed a useful discriminating function that enables one to determine whether the Burr XII distribution will provide a better fit to a given data set than the Weibull, without actually going to the trouble of fitting the Burr XII - see Watkins (2001) - we were interested in extending this result to an accelerated framework. Chapter four discussed this and also explored the extent to which a given data set generated from the Weibull, or alternatively the Burr XII, distribution determined which of the two models would ultimately provide the better fit.

Throughout, it was our intention to assess the extent to which asymptotic results held

for small sample sizes, when considering the inverses of the expected Fisher information matrices and the role of censoring. In order to go some way towards discussing this problem, numerous simulations were carried out and, in addition, theoretical results were obtained. For various sample sizes, we were then able to observe how close to the true, asymptotic results, simulated values were.

Since accelerated life testing is widely applicable in the real-world, an important aim of this thesis was to apply the theoretical results to circumstances that could well exist in real-life. By considering a wide range of possible scenarios and illustrating how to employ the theory, we have provided not only a guide to practitioners wishing to consider their own particular experimental circumstances, but also a useful basis for further investigation.

7.3 Areas For Future Research

Naturally, one of the first things to consider, is a variation in the life-stress relationship. We chose a log-linear approach for its adaptability and ease of use. As discussed in Chapter one, the Arrhenius model is also very popular in the field, and an investigation into the relative merits of an array of models could prove very constructive. In the same vein, we opted for the Burr XII as an extension to the popularly used Weibull distribution; alternative lifetime distributions could well prove to be even more fruitful in terms of quality of fit to data sets, robustness and widespread applicability. Step-stress testing could also be examined. In this case, items are placed at a certain stress level and, after some specified time, if there are any items still running, the stress level is increased. Clearly, this process would introduce many more variables, for example, the size and number of steps, the time between steps and the extent to which these values are kept constant across stress levels. We could also investigate the extent to which the log-likelihoods observed for the Burr XII models are regarded as significantly different from the log-likelihoods observed for the equivalent Weibull model.

The application of censoring also has much room for further investigation. We observed in Chapter five, that there are several ways of carrying out Type I and Type II censoring, each of which will have an impact on the manner in which we develop the theory. While we chose to consider the case where we have a common starting time and a common stopping time for all test items, it is also feasible to have unique starting and stopping times for every stress level. Equivalently, for Type II censoring, we could choose to stop the experiment once, say, ten items have failed at a specified stress level, or wait until ten items have failed from all items on test, although this final option is perhaps not as practically viable as others mentioned in Chapter five. Naturally, to derive theoretical results for Type II censoring would require the consideration of order statistics, and even conditional expectations of order statistics.

From the design perspective, we could investigate the standard deviation of $\widehat{B}_{10}$, using the Weibull distribution as a model for life. We could then see how this compares to the Burr XII distribution, and what circumstances, if any, result in the Weibull being the better

choice of model. In Chapter six, we saw that, theoretically, we should place all items to run at the design stress and wait for complete failure, an unsurprising result. However, we also noted that, due to time constraints in the real world, we need to employ censoring and use at least three stress levels to make interpolations from higher stress levels to the design stress, more accurate. Since true parameter values are rarely known in the real world, more extensive simulations could be carried out to determine the extent to which we can make a trade off between having as few stress levels as possible - and as close to the design stress as possible - while at the same time having enough stress levels - and enough failures at those stress levels - to maximise the information on the scale parameters α_b and β_b; this in turn, through interpolation, would then serve to further reduce the standard deviation of $\widehat{B}_{10}$.

From a programming perspective, we could investigate alternative algorithms for converging on parameter estimates. In particular, we would be interested in both the robustness of the algorithm, and also in the extent to which convergence depends on starting values.

Bibliography

Abramowitz, M. and Stegun, I.A. (1972), *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables.* John Wiley and Sons.

Anderson, D.R., Sweeney, D.J. and Williams, T.A. (1996), *Statistics for Business and Economics.* West Publishing Company.

Bak, J. and Newman, D. (1982), *Complex Analysis.* Springer-Verlag, New York.

Binmore, K.G. (1983), *Calculus - L.S.E. Mathematics Series.* Cambridge University Press.

Birnbaum, Z.W. and Saunders, S.C. (1969), A New Family of Life Distributions. *Journal of Applied Probability* , **6** : 319-327.

Burr, I.W. (1942), Cumulative Frequency Functions. *Annals of Mathematical Statisitcs*, **13** : 215-232.

Copson, E.T. (1935), *An Introduction to the Functions of a Complex Variable.* Oxford.

Cox, D.R. and Hinkley, D.V. (1974), *Theoretical Statistics.* Chapman and Hall.

Crowder, M.J., Kimber, A.C., Smith, R.L. and Sweeting, T.J. (1991), *Statistical Analysis of Reliability Data* . Chapman and Hall.

Crowder, M.J. and Kimber, A.C. (1997), A Score Test for the Multivariate Burr and Other Weibull Mixture Distributions. *Scandinavian Journal of Statistics*, **24** : 419-432.

Farnum, N.R. and Booth, P. (1997), Uniqueness of Maximum Likelihood Estimators of the 2-Parameter Weibull Distribution. *IEEE Transactions on Reliability* , **46** : 523-525.

Hogg, R.V. and Klugman, S.A. (1984), *Loss Distributions.* John Wiley and Sons, New York.

Johnson, N.L. and Kotz, S. (1970), *Distributions in Statistics: Continuous Univariate Distributions - 1* . Houghton Mifflin.

Kalbfleisch, J.D. and Prentice, W.J. (1980), *The Statistical Analysis of Failure Time Data.* Wiley, New York.

Kielpinski, T.J. and Nelson, W. (1975), Optimum Censored Accelerated Life-Tests for Normal and Lognormal Life Distributions. *IEEE Transtactions on Reliability*, **24** : 310-320.

Lawless, J.F. (1982), *Statistical Models and Methods for Lifetime Data.* John Wiley & Sons.

Meeker, W.Q. and Nelson, W. (1975), Optimum Accelerated Life-Tests for Weibull and Extreme Value Distributions. *IEEE Transactions on Reliability*, **24** : 321-332.

Meeker, W.Q. and Escobar, L.A. (1993), A Review of Recent Research and Current Issues in Accelerated Testing. *International Statistical Review* , **61** : 147-168.

Nelson, W. (1990), *Accelerated Testing, Statistical Models, Test Plans and Data Analysis.* John Wiley and Sons.

Newton, D.W. (1991), Some Pitfalls in Reliability Data Analysis [with Discussion]. *Reliability Engineering and System Safety* , **34** : 7-21.

Owen, W.J. and Padgett, W.J. (2000), A Birnbaum-Saunders Accelerated Life Model. *IEEE Transactions on Reliability*, **49** : 224-229.

Patel, J.K., Kapadia, C.H. and Owen, D.B. (1976), *Handbook of Statistical Distributions* . Marcel Dekker, Inc.

Qiu, P. and Tsokos, C.P. (2000), Accelerated Life Testing Model Building with Box-Cox Transformations. *Sankhya* , **62** : 223-235.

Richards, D.O. and McDonald, J.B. (1987), A General Methodology for Determining Distributional Forms with Applications in Reliability. *Journal of Statistical Planning and Inference*, **16** : 365-376.

Tadikamalla, P.R. (1980), A Look at the Burr and Related Distributions. *International Statistical Review*, **48** : 337-355.

Watkins, A.J. (1991), On the Analysis of Accelerated Life Testing Experiments. *IEEE Transactions on Reliability* , **40** : 98-101.

Watkins, A.J. (1994), Review: Likelihood Method For Fitting Weibull Log-Linear Models To Accelerated Life-Test Data. *IEEE Transactions On Reliability*, **43** : 361-365

Watkins, A.J. (1997), A Note on Expected Fisher Information For the Burr XII Distribution. *Microelectonics Reliability* , **37** : 1849-1852.

Watkins, A.J. (1998a), A SAS/IML Algorithm for Maximum Likelihood Estimation in the Three Parameter Burr XII Distribution. *Proceedings of SEUGI 16, Praha, Czech Republic.*

Watkins, A.J. (1998b), On Expectations Associated with Maximum Likelihood Estimation in the Weibull Distribution. *Journal of the Italian Statistical Society*, **7** : 15-26.

Watkins, A.J. (1999), An Algorithm for Maximum Likelihood Estimation in the Three Parameter Burr XII Distribution. *Computational Statistics & Data Analysis*, **32** : 19-27.

Watkins, A.J. (2001), On the Likelihood Functions for the Three Parameter Burr XII Distribution. *International Journal of Reliability, Quality and Safery Engineering*, **8** : 173-188.

Watkins, A.J. and Johnson, R. (1999), Using SAS/IML to Fit Embedded Models to Accelerated Life Test Data. *Proceedings of SEUGI 17, Den Haag, The Netherlands.*

Watkins, A.J. and Johnson, R. (2002), Two Result on a Sub-Class of Hypergeometric Functions with Applications in Reliability Analysis. *International Journal of Pure and Applied Mathematics*, **3** : 71-89.

Weber, J.E. (1982), *Mathematical Analysis, Business and Economic Applications*, Harper and Row, New York.

Appendix A : Algorithms

In this appendix we give details of the SAS IML algorithms used to determine the maximum likelihood estimates of parameters from the various life models considered. Within the code, modules are defined by 'start <module_name> $\cdots$ finish <module_name>'; the module then being called from elsewhere by the command 'run <module_name>'. Throughout, comments shall be inserted and italicised.

Weibull

This algorithm determines the maximum likelihood estimates of parameters in a non-accelerated Weibull distribution, see section 2.1.1.

```
proc iml;

start weibmle2;
 n = nrow(wdata);
 m = n-sum(ind);
 lnx = log(wdata);
 se = sum(lnx#(1-ind));
```

The starting value of the shape parameter b is set to 1

```
 b = 1.0;
```

We iterate on b to obtain its MLE

```
 do iter = 1 to 15;
  s0 = sum(exp(b*lnx));
  s1 = sum(lnx#exp(b*lnx));
  s2 = sum(lnx#lnx#exp(b*lnx));
  ratio = s1/s0;
  pl = m*log(b)+(b-1)*se-m*log(s0);
  plb = m/b +se - m*ratio;
  plbb = - m/(b**2) - m*(s2/s0-ratio**2);
  b = b - plb/plbb;
```

```
  print b;
 end;
```

We can now find the MLE of the scale phi and the maximised log-likelihood.

```
 phi = exp(log(s0/m)/b);
 lw = m*log(m*b/s0)+(b-1)*se-m;
 print b phi lw;
finish weibmle2;
```

Main Program. Here we consider the first subset of the aluminium coupon data, that is where stress = 2.1 psi/cycle.

```
do;
data={3.7, 7.46, 8.44, 8.86, 9.6, 10.00, 10.18, 10.85, 11.08, 11.34,
12.00,12.22,12.52,12.69,12.93,13.13,13.55,14.19,14.50,14.78,
15.02,15.22,15.4,15.78,16.04,16.42,17.5,17.68,17.92,18.81,
18.95,19.4,21.0,22.68};
```

The indexing is established to flag failed items.

```
ind={0};
ind=repeat(ind,34,1);
 wdata = data;
 run weibmle2;
end;
quit;
```

Weibull_ALT

Here, we determine the maximum likelihood estimates for parameters from a Weibull ALT model; see section 2.1.2.

```
proc iml;

start weib;
do i=1 to 20;
sn=sum(n);
sx=sum(n#x);
se=sum(log(fmat)#indf);
fmat1=indf#exp(b*log(fmat));
cmat1=indc#exp(b*log(cmat));
```

```
s00=sum((fmat1[,+] + cmat1[,+])#exp(-beta*b*x));
fmat2=fmat1#log(fmat);
cmat2=cmat1#log(cmat);
s01=sum((fmat2[,+] + cmat2[,+])#exp(-beta*b*x));
fmat3=fmat2#log(fmat);
cmat3=cmat2#log(cmat);
s02=sum((fmat3[,+] + cmat3[,+])#exp(-beta*b*x));
s10=sum((fmat1[,+] + cmat1[,+])#exp(-beta*b*x)#x);
s11=sum((fmat2[,+] + cmat2[,+])#exp(-beta*b*x)#x);
s20=sum((fmat1[,+] + cmat1[,+])#exp(-beta*b*x)#(x##2));
lb=sn/b + se - sn*(s01 - beta*s10)/s00 - beta*sx;
lbeta=sn*b*s10/s00 - b*sx;
lbb=-sn/(b**2) - sn*(s02 - 2*beta*s11 + (beta**2)*s20)/s00
    + sn*((s01 - beta*s10)**2)/((s00)**2);
lbbeta=sn*(b*s11 + s10 - beta*b*s20)/s00
    - sn*b*(s01 - beta*s10)*s10/((s00)**2) - sx;
lbetbet=sn*(b**2)*(((s10)**2)/((s00)**2) - s20/s00);
m=j(2);
m[1,1]=lbb; m[1,2]=lbbeta; m[2,1]=lbbeta; m[2,2]=lbetbet;
m=-m; m=inv(m);
vec={0,0}; vec[1]=lb; vec[2]=lbeta;
data={0,0}; data[1]=b; data[2]=beta;
res=data + m*vec;
b=res[1]; beta=res[2];
print b beta;
print lb lbeta;
end;
finish weib;
```

Main program. The aluminium coupon data is included explicitly.

```
do;
 fmat={3.7 7.46 8.44 8.86 9.6 10.00 10.18 10.85 11.08 11.34
12.00 12.22 12.52 12.69 12.93 13.13 13.55 14.19 14.50 14.78
15.02 15.22 15.4 15.78 16.04 16.42 17.5 17.68 17.92 18.81
18.95 19.4 21.0 22.68,
 2.33 2.76 3.12 3.21 3.35 3.38 3.42 3.5 3.51 3.56
 3.6 3.66 3.7 3.74 3.79 3.82 3.95 4.0 4.04 4.08
 4.14 4.16 4.23 4.32 4.33 4.39 4.45 4.56 4.64 4.7
 4.74 4.86 4.9 5.17,
```

```
 0.7 0.97 1.03 1.05 1.08 1.09 1.13 1.14 1.2 1.21
 1.24 1.24 1.28 1.3 1.31 1.31 1.32 1.34 1.34 1.36
 1.38 1.39 1.42 1.42 1.44 1.46 1.49 1.52 1.57 1.57
 1.62 1.64 1.68 1.96};
```

Indexing is set up, with 1 indicating a failure

```
 indf={1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1,
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1,
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1};
n={34,34,34};
```

The program was written to accomodate potential censoring. As such we include a customary sample of censored items, which, through the indexing, we then ignore for the purposes of this example. Nevertheless, sample allocations on these 'censored items' still need to be specified.

```
cmat={1,1,1};
indc={0,0,0};
m={1,1,1};
```

Define stress levels.

```
x={2.1,2.6,3.1};
xf=repeat(x,1,34);
xc=x;
```

Set starting values for shape parameter b and scale parameter beta.

```
b=5.0;
beta=-2.37495;
```

Run module 'weib' to determine MLEs of b and beta.

```
run weib;
```

Finally determine the MLE of the scale parameter alpha and the maximised log-likelihood.

```
alpha=log(s00/sn)/b;
l=sn*log(b)+(b-1)*se-(alpha*sn+beta*sx)*b-exp(-alpha*b)*s00;
```

The estimates of the scales at each stress level are then found.

```
theta=exp(alpha + beta*x);
print b beta alpha theta;
print l;
end;
quit;
```

Burr3

Determines the maximum likelihood estimates of parameters from a Burr XII non-ALT distribution for complete / censored data; see section 2.2.

```
proc iml;
```

Fit the usual two parameter Weibull distribution

```
start weibmle2;
  n = nrow(wdata);
  m = n-sum(ind);
  lnx = log(wdata);
  se = sum(lnx#(1-ind));
```

Starting value of shape parameter is 1

```
  beta = 1.0;
  do iter = 1 to 15;
    s0 = sum(exp(beta*lnx));
    s1 = sum(lnx#exp(beta*lnx));
    s2 = sum(lnx#lnx#exp(beta*lnx));
    ratio = s1/s0;
    pl  = m*log(beta)+(beta-1)*se-m*log(s0);
    plb = m/beta +se - m*ratio;
    plbb = - m/(beta**2) - m*(s2/s0-ratio**2);
    beta = beta - plb/plbb;
 print beta;
  end;
  theta = exp(log(s0/m)/beta);
  lw    = m*log(m*beta/s0)+(beta-1)*se-m;
```

Print shape, scale and log-likelihood

```
  print beta theta lw;
finish weibmle2;
```

Fit a two parameter Burr XII distribution
shape1 is the shape parameter tau, shape2 is the shape parameter a

```
start b2mle;
  n    = nrow(bdata);
  m    = n-sum(ind);
  lnx  = log(bdata);
  se   = sum(lnx#(1-ind));
```

Iterate on tau

```
  do iter = 1 to 15;
    term = exp(shape1*lnx);
    sfstar = sum(log(term+1)#(1-ind));
    sf111  = sum(term#lnx/(term+1)#(1-ind));
    sf122  = sum(term#lnx#lnx/((term+1)#(term+1))#(1-ind));
    scstar = sum(log(term+1)#ind);
    sc111  = sum(term#lnx/(term+1)#ind);
    sc122  = sum(term#lnx#lnx/((term+1)#(term+1))#ind);
    sstar = sfstar + scstar;
    s111  = sf111 + sc111;
    s122  = sf122 + sc122;
    pl = m*log(shape1) + (shape1-1)*se - sfstar - m*log(sstar);
    pl1 = m/shape1 + se - sf111 - m*s111/sstar;
    pl11 = -m/(shape1**2) - sf122 - m*(s122/sstar-(s111/sstar)**2);
    shape1 = shape1 - pl1/pl11;
    print shape1 pl1 pl11;
  end;
  shape2 = m/sstar;
  lb = m*log(shape1*shape2)+(shape1-1)*se-(shape2+1)*sfstar-shape2*scstar;
```

Print tau, a and maximised log-likelihood

```
  print shape1 shape2 lb;
finish b2mle;
```

Fit a three parameter Burr XII distribution

```
start b3eval;
  n    = nrow(bdata);
  m    = n-sum(ind);
  lnx  = log(bdata/phi);
  se   = sum(log(bdata)#(1-ind));
  term = exp(shape1*lnx);
  termp1 = term + 1;
  tdstar = sum(log(term+1)#(1-ind));
  td101  = sum(term/termp1#(1-ind));
  td102  = sum(term/(termp1#termp1)#(1-ind));
  td111  = sum(term#lnx/(term+1)#(1-ind));
  td112  = sum(term#lnx/(termp1#termp1)#(1-ind));
  td122  = sum(term#lnx#lnx/((term+1)#(term+1))#(1-ind));
  tcstar = sum(log(term+1)#ind);
  tc101  = sum(term/termp1#ind);
  tc102  = sum(term/(termp1#termp1)#ind);
  tc111  = sum(term#lnx/(term+1)#ind);
  tc112  = sum(term#lnx/(termp1#termp1)#ind);
  tc122  = sum(term#lnx#lnx/((term+1)#(term+1))#ind);
  lb = m*log(shape1*shape2) - m*shape1*log(phi)
        + (shape1-1)*se - (shape2+1)*tdstar-shape2*tcstar;
  print shape1 shape2 phi lb;
  lb1 = m/shape1 - m*log(phi) + se - (shape2+1)*td111 - shape2*tc111;
  lb2 = m/shape2 - (tdstar+tcstar);
  lbp = shape1*((shape2+1)*td101+shape2*tc101-m)/phi;
  print lb1 lb2 lbp;
  lb11 = -m/(shape1**2) - (shape2+1)*td122 - shape2*tc122;
  lb12 = -(td111+tc111);
  lb1p = ((shape2+1)*(td101+shape1*td112)
            + shape2*(tc101+shape1*tc112) - m)/phi;
  lb22 = -m/(shape2**2);
  lb2p = shape1*(td101+tc101)/phi;
  lbpp = shape1*(m-(shape2+1)*td101-shape2*tc101
          - shape1*((shape2+1)*td102+shape2*tc102))/(phi**2);
  rmat = {0.0 0.0 0.0, 0.0 0.0 0.0, 0.0 0.0 0.0};
```

Set up the full matrix of second derivatives

```
  rmat[1,1] = lb22; rmat[1,2] = lb12; rmat[1,3] = lb2p;
```

```
  rmat[2,1] = lb12; rmat[2,2] = lb11; rmat[2,3] = lb1p;
  rmat[3,1] = lb2p; rmat[3,2] = lb1p; rmat[3,3] = lbpp;

  rmat = inv(rmat);
  rvec = {0.0, 0.0, 0.0};
  rvec[1] = lb2;  rvec[2] = lb1;  rvec[3] = lbp;
  result = rmat*rvec;
```

Print first and second derivatives to check we have MLEs

```
  print lb1 lb2 lbp;
  print lb11 lb22 lbpp;
finish b3eval;
```

Main Program - Carbon Fibre Rod Data for stress=1mm included

```
do;
```

Flag failed items with a zero

```
ind={0};
ind=repeat(ind,57,1);
 data= {2.247,2.640,2.842,2.908,3.099,3.126,3.245,3.328,3.355,3.383,3.572,
        3.581,3.681,3.726,3.727,3.728,3.783,3.785,3.786,3.896,3.912,3.964,
        4.050,4.063,4.082,4.111,4.118,4.141,4.216,4.251,4.262,4.326,4.402,
        4.457,4.466,4.519,4.542,4.555,4.614,4.632,4.634,4.636,4.678,4.698,
        4.738,4.832,4.924,5.043,5.099,5.134,5.359,5.473,5.571,5.684,5.721,
        5.998,6.060};
 wdata = data;
```

Fit a two parameter Weibull distribution

```
 run weibmle2;
 thetah = theta;
```

Scale the data by the MLE of theta

```
 wdata = data/thetah;
 run weibmle2;
```

Set starting value of phi and tau to be 1 and beta respectively

```
 phi=1.0;
 shape1=beta;
 do it=1 to 50;
```

Scale the 'Weibull-scaled' data by phi

```
   bdata = wdata/phi;
```

Fit a two parameter Burr XII distribution

```
   run b2mle;
```

Take the original 'Weibull-scaled' data and fit a three parameter Burr XII distribution

```
   bdata = wdata;
   run b3eval;
```

Iterate once on phi

```
   phi=phi-result[3];
   print phi;
 end;
```

Find phi for raw data set

```
 phi = phi*thetah;
 print phi;
```

As a check, fit a three parameter Burr XII model to the raw data set

```
 bdata = data;
 run b3eval;
 end;
quit;
```

Burr_ALT/Burrcens

Algorithm to find MLEs for parameters in a Burr XII ALT model for complete / censored data; see sections 2.3.2 and 5.5 respectively.

```
proc iml;

start buraccel;
start weibmle2;
```

See code for Weibull above for details of this module.

```
finish weibmle2;
```

The following module fits a Weibull ALT model.

```
start weibmlec;
 n = nrow(wdata);
 m = n-sum(ind);
 lnx = log(wdata);
 se = sum(lnx#(1-ind));
 sx = sum(stress#(1-ind));
 do iter = 1 to 15;
  term = exp(b*lnx)#exp(-b*beta*stress);
  s00 = sum(term);
  s01 = sum(lnx#term);
  s02 = sum(lnx#lnx#term);
  s10 = sum(stress#term);
  s11 = sum(stress#lnx#term);
  s12 = sum(stress#lnx#lnx#term);
  s20 = sum(stress#stress#term);
  s21 = sum(stress#stress#lnx#term);
  s22 = sum(stress#stress#lnx#lnx#term);
  pl = m*log(b)+(b-1)*se-m*log(s00)-b*beta*sx;
  print pl;
  pl1 = m/b + se - m*(s01-beta*s10)/s00 - beta*sx;
  pl2 = m*b*s10/s00 - b*sx;
  pl11= -m/(b**2) - m*(s02-2*beta*s11+s20*beta**2)/s00 +
  m*(s01-beta*s10)**2/s00**2;
  pl12 = m*(b*s11+s10-b*beta*s20)/s00 -
  m*b*(s01-beta*s10)*s10/s00**2 - sx;
  pl22 = m*(b**2)*(s10**2/s00**2-s20/s00);
  det = pl11*pl22-pl12**2;
  corsh = (pl12*pl2-pl22*pl1)/det;
  corsl = (pl12*pl1-pl11*pl2)/det;
  b = b + corsh;
  beta = beta + corsl;
  print b beta;
 end;
 alpha = log(s00/m)/b;
 print alpha;
```

```
finish weibmlec;

start b2mle;
```

See code for Burr3 above for details of this module.

```
finish b2mle;

start b3eval;
```

See code for Burr3 above for details of this module.

```
finish b3eval;
```

The next module fits the full Burr XII ALT model.

```
start burr3acc;
 n = nrow(wdata);
 sn = n-sum(ind);
 lnx = log(wdata);
 se = sum(lnx#(1-ind));
 sx = sum(stress#(1-ind));
```

Define the starting values for alpha and beta

```
 alpha = log(phi);
 beta = 0;
```

Iterate on a, tau, alpha and beta.

```
 do iter = 1 to 50;
  work = lnx - alpha - beta*stress;
  termf = exp(t*work);
  termfp1 = termf+1;
  sfstar = sum((1-ind)#log(termfp1));
  scstar = sum(ind#log(termfp1));

  sf0101 = sum(termf/termfp1#(1-ind));
  sf0111 = sum(termf#work/termfp1#(1-ind));
  sf0121 = sum(termf#work#work/termfp1#(1-ind));
  sf0202 = sum(termf#termf/(termfp1#termfp1)#(1-ind));
  sf0212 = sum(termf#termf#work/(termfp1#termfp1)#(1-ind));
  sf0222 = sum(termf#termf#work#work/(termfp1#termfp1)#(1-ind));
```

```
sf1101 = sum(stress#termf/termfp1#(1-ind));
sf1111 = sum(stress#termf#work/termfp1#(1-ind));
sf1202 = sum(stress#termf#termf/(termfp1#termfp1)#(1-ind));
sf1212 = sum(stress#termf#termf#work/(termfp1#termfp1)#(1-ind));

sf2101 = sum(stress#stress#termf/termfp1#(1-ind));
sf2202 = sum(stress#stress#termf#termf/(termfp1#termfp1)#(1-ind));

sc0101 = sum(termf/termfp1#ind);
sc0111 = sum(termf#work/termfp1#ind);
sc0121 = sum(termf#work#work/termfp1#ind);
sc0202 = sum(termf#termf/(termfp1#termfp1)#ind);
sc0212 = sum(termf#termf#work/(termfp1#termfp1)#ind);
sc0222 = sum(termf#termf#work#work/(termfp1#termfp1)#ind);

sc1101 = sum(stress#termf/termfp1#ind);
sc1111 = sum(stress#termf#work/termfp1#ind);
sc1202 = sum(stress#termf#termf/(termfp1#termfp1)#ind);
sc1212 = sum(stress#termf#termf#work/(termfp1#termfp1)#ind);

sc2101 = sum(stress#stress#termf/termfp1#ind);
sc2202 = sum(stress#stress#termf#termf/(termfp1#termfp1)#ind);
l = sn*log(a*t) - alpha*t*sn - beta*t*sx + (t-1)*se - (a+1)*sfstar
    - a*scstar;
print l;
```

Define the vector of first derivatives.

```
deriv[1] = sn/a - (sfstar + scstar);
deriv[2] = sn/t - alpha*sn - beta*sx + se - (a+1)*sf0111 - a*sc0111;
deriv[3] = -sn*t + (a+1)*t*sf0101 + t*a*sc0101;
deriv[4] = -sx*t + (a+1)*t*sf1101 + t*a*sc1101;
```

Define the matrix of second derivatives.

```
hess[1,1] = -sn/(a*a);
hess[1,2] = -(sf0111 + sc0111);
hess[1,3] = t*(sf0101 + sc0101);
hess[1,4] = t*(sf1101 + sc1101);

hess[2,2] = -sn/(t*t) - (a+1)*(sf0121-sf0222) - a*(sc0121-sc0222);
```

```
  hess[2,3] = -sn + (a+1)*(sf0101+t*(sf0111-sf0212)) + a*(sc0101
              +t*(sc0111-sc0212));
  hess[2,4] = -sx + (a+1)*(sf1101+t*(sf1111-sf1212)) + a*(sc1101
              +t*(sc1111-sc1212));

  hess[3,3] = (a+1)*t*t*(sf0202-sf0101) + t*t*a*(sc0202-sc0101);
  hess[3,4] = (a+1)*t*t*(sf1202-sf1101) + t*t*a*(sc1202-sc1101);

  hess[4,4] = (a+1)*t*t*(sf2202-sf2101) + t*t*a*(sc2202-sc2101);

  hess[2,1] = hess[1,2];
  hess[3,1] = hess[1,3];
  hess[4,1] = hess[1,4];
  hess[3,2] = hess[2,3];
  hess[4,2] = hess[2,4];
  hess[4,3] = hess[3,4];

  corr = inv(hess)*deriv;
```

Update the values of a, tau, alpha and beta.

```
  a = a - corr[1];
  t = t - corr[2];
  alpha = alpha - corr[3];
  beta = beta - corr[4];
  print a t alpha beta deriv;
 end;
finish burr3acc;

do;
 wdata = data;
```

Fit a Weibull ALT model to the data set

```
 run weibmlec;
 print b beta;
 x1= slevel;
 the1=exp(alpha+beta*x1);
 alpha1=alpha; beta1=beta;
 par={1,1,1};
 hess={1 1 1, 1 1 1, 1 1 1};
 deriv = par;
```

```
corr = par;
```

Rescale the raw data by Weibull scale MLEs

```
data1 = data#exp(-alpha-beta*stress);
wdata=data1;
run weibmle2;
```

Iterative scaling of 'Weibull-scaled' data, by phi, and fit a Burr XII model to entire data set
Set starting value for phi to be 1

```
phi=1.0;
```

Fit a Burr XII three parameter model to the entire data set

```
do it=1 to 15;
 bdata = wdata/phi;
 run b2mle;
 bdata = wdata;
 run b3eval;
 hess[1,1] = lb11;
 hess[1,2] = lb12; hess[2,1] = hess[1,2];
 hess[1,3] = lb1p; hess[3,1] = hess[1,3];
 hess[2,2] = lb22;
 hess[2,3] = lb2p; hess[3,2] = hess[2,3];
 hess[3,3] = lbpp;
 deriv[1] = lb1;
 deriv[2] = lb2;
 deriv[3] = lbp;
 corr = inv(hess)*deriv;
 phi = phi - corr[3];
end;
par = {1, 1, 1, 1};
hess = {1 1 1 1, 1 1 1 1, 1 1 1 1, 1 1 1 1};
deriv = par;
corr = par;

run burr3acc;
the2=exp(alpha+beta*x1);
finthe=the1#the2;
```

Determine alpha and beta for raw data set

```
 finalp=alpha+alpha1; finbet=beta+beta1;
 alpha=finalp; beta=finbet;
 wdata=data;
```

Finally, we determine the Burr XII ALT maximised log-likelihood for the raw data set

```
 n = nrow(wdata);
 sn = n-sum(ind);
 lnx = log(wdata);
 se = sum(lnx#(1-ind));
 sx = sum(stress#(1-ind));
 work = lnx - alpha - beta*stress;
 termf = exp(t*work);
 termfp1 = termf+1;
 sfstar = sum((1-ind)#log(termfp1));
 scstar = sum(ind#log(termfp1));
 l = sn*log(a*t) - alpha*t*sn - beta*t*sx + (t-1)*se - (a+1)*sfstar
     - a*scstar;
 print l;
 print a t finalp finbet finthe x1;
end;
finish buraccel;
```

Main Program. Aluminium coupon data included.

```
do;
slevel={2.1,2.6,3.1};
data={3.7, 7.46, 8.44, 8.86, 9.6, 10.00, 10.18, 10.85, 11.08, 11.34,
 12.00,12.22,12.52,12.69,12.93,13.13,13.55,14.19,14.50,14.78,
15.02,15.22,15.4,15.78,16.04,16.42,17.5,17.68,17.92,18.81,
18.95,19.4,21.0,22.68,
 2.33 ,2.76, 3.12, 3.21, 3.35, 3.38, 3.42, 3.5, 3.51, 3.56,
 3.6, 3.66, 3.7, 3.74, 3.79, 3.82, 3.95, 4.0, 4.04, 4.08,
 4.14, 4.16, 4.23, 4.32, 4.33, 4.39, 4.45, 4.56, 4.64, 4.7,
 4.74, 4.86, 4.9, 5.17,
 0.7, 0.97, 1.03, 1.05, 1.08, 1.09, 1.13, 1.14, 1.2, 1.21,
 1.24, 1.24, 1.28, 1.3, 1.31, 1.31, 1.32, 1.34, 1.34, 1.36,
 1.38, 1.39, 1.42, 1.42, 1.44, 1.46, 1.49, 1.52, 1.57, 1.57,
```

```
 1.62, 1.64, 1.68, 1.96};
 tot=nrow(data);
 ind=repeat(0,tot);
 stress={2.1,2.1,2.1,2.1,2.1,2.1,2.1,2.1,2.1,2.1,
 2.1,2.1,2.1,2.1,2.1,2.1,2.1,2.1,2.1,2.1,
 2.1,2.1,2.1,2.1,2.1,2.1,2.1,2.1,2.1,2.1,
 2.1,2.1,2.1,2.1,
 2.6,2.6,2.6,2.6,2.6,2.6,2.6,2.6,2.6,2.6,
 2.6,2.6,2.6,2.6,2.6,2.6,2.6,2.6,2.6,2.6,
 2.6,2.6,2.6,2.6,2.6,2.6,2.6,2.6,2.6,2.6,
 2.6,2.6,2.6,2.6,
 3.1,3.1,3.1,3.1,3.1,3.1,3.1,3.1,3.1,3.1,
 3.1,3.1,3.1,3.1,3.1,3.1,3.1,3.1,3.1,3.1,
 3.1,3.1,3.1,3.1,3.1,3.1,3.1,3.1,3.1,3.1,
 3.1,3.1,3.1,3.1};
```

Set starting values for the Weibull ALT module.

```
 b=5.0; beta=-2.3; alpha=7.8;
```

Fit a Burr XII ALT model

```
 run buraccel;
end;
quit;
```

Appendix B : Neighbouring Hypergeometric Functions

Here, we seek to prove the recurrence relations between 'neighbouring' generalised hypergeometric functions. These functions were seen to be necessary when looking at censored data sets; see section 5.3.2.

Firstly, we have,

$$\begin{aligned}
f_q(a,b,z) - f_{q-1}(a,b,z) &= \sum_{i=0}^{\infty} \left[\frac{(a)_i}{(a+1)_i}\right]^{q-1} \left[\frac{(a)_i}{(a+1)_i} - 1\right] (b)_i \frac{z^i}{i!} \\
&= -\frac{1}{a} \sum_{i=0}^{\infty} i \left[\frac{(a)_i}{(a+1)_i}\right]^{q} (b)_i \frac{z^i}{i!} \\
&= -\frac{z}{a} \sum_{i=1}^{\infty} \left[\frac{(a)_i}{(a+1)_i}\right]^{q} (b)_i \frac{z^{i-1}}{(i-1)!} \\
&= -\frac{z}{a} \sum_{i=0}^{\infty} \left[\frac{(a)_{i+1}}{(a+1)_{i+1}}\right]^{q} (b)_{i+1} \frac{z^i}{i!} \\
&= -\frac{za^q b}{a(a+1)^q} \sum_{i=0}^{\infty} \left[\frac{(a+1)_i}{(a+2)_i}\right]^{q} (b+1)_i \frac{z^i}{i!} \\
&= -\frac{za^{q-1}b}{(a+1)^q} \times f_q(a+1, b+1, z),
\end{aligned}$$

as required. For the second recurrence relation, we have

$$\begin{aligned}
f_q(a,b+1,z) - f_q(a,b,z) &= \sum_{i=0}^{\infty} \left[\frac{(a)_i}{(a+1)_i}\right]^{q} [(b+1)_i - (b)_i] \frac{z^i}{i!} \\
&= \frac{1}{b} \sum_{i=0}^{\infty} i \left[\frac{(a)_i}{(a+1)_i}\right]^{q} (b)_i \frac{z^i}{i!} \\
&= \frac{z}{b} \sum_{i=1}^{\infty} \left[\frac{(a)_i}{(a+1)_i}\right]^{q} (b)_i \frac{z^{i-1}}{(i-1)!} \\
&= \frac{z}{b} \sum_{i=0}^{\infty} \left[\frac{(a)_{i+1}}{(a+1)_{i+1}}\right] (b)_{i+1} \frac{z^i}{i!}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{za^q}{(a+1)^q} \sum_{i=0}^{\infty} \left[\frac{(a+1)_i}{(a+2)_i}\right]^q (b+1)_i \frac{z^i}{i!} \\
&= \frac{za^q}{(a+1)^q} \times f_q\left(a+1, b+1, z\right),
\end{aligned}
$$

as required.