STATISTICAL ANALYSIS OF THERMOLUMINESCENCE EXPERIMENTS FOR SEDIMENTARY DATING

by

W.Chandanie Wijayalatha Perera

B.Sc.(Hons.), University of Colombo, Sri Lanka, 1985

Postgraduate Diploma(Applied Statistics), University of Colombo, Sri Lanka, 1987

M.Sc.(Statistics), Simon Fraser University, Canada, 1992

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department

of

Mathematics and Statistics

SIMON FRASER UNIVERSITY

July 1996

ISBN 0-612-17048-9

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

Statistical Analysis of Thermoluminescence Experiments for Sedimentary Dating

Author:_____
(signature)

(name)

July 17, 1996.

(date)

APPROVAL

Name: W.Chandanie Wijayalatha Perera

Degree: DOCTOR OF PHILOSOPHY

Title of thesis: STATISTICAL ANALYSIS OF THERMOLUMINESCENCE EXPERIMENTS FOR SEDIMENTARY DATING

Examining Committee: Dr. Carl J. Schwarz
Chair

Dr. Richard A. Lockhart
Senior Supervisor

Dr. Charmaine B. Dean

Dr. Richard D. Routledge

Dr. T. Swartz

Dr. Peter Guttorp, University of Washington, Seattle
External Examiner

Date Approved: July 17, 1996

Abstract

Sediment (or other buried material) when heated gently glows with light called thermoluminescence. The amount of light given off depends on the material and on the amount of radiation impinging on the sample while buried. Comparison of the equivalent dose (a known laboratory dose required to produce the same amount of luminescence as the original untreated sample) with historical radiation rates permits estimation of the age (duration of burial) of the sample, a process called thermoluminescence dating.

We study statistical techniques for estimating the equivalent dose from the data collected for thermoluminescence dating. Physical models are used to motivate generalized non-linear models for the data and to justify assumptions about the distribution of errors in these models. Maximum likelihood, quasi-likelihood and least squares estimators are compared by examining their statistical properties. Formulae are provided for the biases and the mean squared errors of these estimators valid in the limit of small measurement errors.

In thermoluminescence studies, data are collected on a single sample at a series of temperatures. Consequently, observations collected at different temperatures are correlated. We propose a generalized estimating equations procedure for estimating the equivalent dose from the correlated data. Large sample asymptotic properties of the proposed estimate are examined and a formula is provided for estimating the error of the estimate. We propose symmetric confidence intervals for the equivalent dose with a t quantile; a formula is provided for the approximate degrees of freedom of the suggested t quantile, valid in the limit of small measurement errors. Finite sample performance of the asymptotic results is examined by Monte Carlo.

Tests based on the empirical distribution function (EDF tests) of the standardized residuals are proposed for testing the distributional assumptions on the random errors in two situations: without assuming the fitted model is correct and assuming the fitted model is correct. We propose a recurrence formula for evaluating the cumulative distribution function of two fitted standardized residuals needed in the proposed EDF tests. Weak convergence properties of the related empirical processes are examined. Finite sample performance of the suggested EDF tests is examined by Monte Carlo.

Acknowledgments

It is my pleasure to express my sincere appreciation and gratitude to my senior supervisor Prof. Richard Lockhart for his kind guidance and invaluable advice that made this work a success. I feel very fortunate to have him as my supervisor.

I am indebted for the help I received from Professors Charmaine Dean, David Eaves, Richard Lockhart, Richard Routledge, Michael Stephens, Tim Swartz and Larry Weldon which is well beyond academic work. I immensely value the knowledge they gave to me throughout my stay in Canada. I also feel fortunate to have known Dr. François Bellavance, Dr. Gary Parker, Dr. Carl Schwarz, Dr. John Spinelli and Dr. Randy Sitter.

I would also like to thank Dr. D. J. Huntley, Dept. of Physics, Simon Fraser University who introduced me to the area of thermoluminescence dating and provided the data used in the examples. He was always available for help whenever I had difficulties understanding the physics related to this work.

I also wish to thank Ms. Maggie Fankboner, Sylvia Holmes, Ana Leon, Dianne Pogue and other department staff. The financial support provided to me through the commonwealth scholarship, by the Department of Mathematics and Statistics and from the research grants received by Prof. Richard Lockhart is greatly appreciated. I would also extend my thanks to my colleagues and friends whose help and friendship made my stay in Canada a valuable experience. Friendly faces of faculty, students and the staff made me feel the Department of Mathematics and Statistics as a family unit and a comfortable place to work. Finally, I would like to acknowledge with thanks the kind, caring support and utmost help of my dear friend and colleague Sarath Banneheka.

Dedication

To my parents and teachers

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Estimating the ages of historic materials is an important problem, needed in many fields such as archaeology, geology and quaternary science. Radio carbon dating has been the most widely used technique for this purpose. Berger *et.al.* [10] reports that radiocarbon dating of Quaternary deposits is limited to organic material younger than about 50000 years old. According to the same source, many of the most important stratigraphic units are much older than this. Recent research in this area has proved that the luminescence phenomena can be utilized for estimating the ages of old samples.

What is 'thermoluminescence'? Upon gentle heating, crystalline or glassy materials begin to glow. This weak but measurable light which is observed before the characteristic 'red hot glow' or incandescence is called 'Thermoluminescence' (TL). Luminescence phenomena have been studied by scientists for several centuries. However, use of luminescence phenomena for measuring exposure to nuclear radiation was not developed until early 1950's. Subsequently, during the 1960's thermoluminescence was developed for archaeological dating.

The application of TL dating to sediments follows from the work of Huntley and Johnson [35]. Wintle and Huntley [60] report the first successful application of TL dating to ocean sediments. In the dating of sediments the event being dated is the last exposure to sunlight.

Thermoluminescence dating of sediments is not limited to ocean sediments. It extends to desert loess, sand dunes, sediments in lakes filled by melting glaciers and even to the

dust incorporated in the ice of a glacier itself. It can also be used to analyze the history of meteorites and lunar material (Aitken [1]). For this reason many fields such as geology, archaeology, quaternary research etc. benefit from its use.

According to Aitken [1], the key concept behind the thermoluminescence dating of sediments could briefly be described as follows. While the sediments are buried in the ground they are exposed to the nuclear radiation emitted by the naturally occurring radioactive materials such as K^{40}, Th^{232}, U^{238} etc. This results in an ionization of electrons of the crystal lattice; these electrons are now free to diffuse around the crystal (see also Divigalpitiya [26]). Due to imperfections of crystalline materials there are negative charge deficit sites that are available to such diffusing electrons. Some of the electrons while diffusing around the crystal get trapped at these sites. When trapped, they remain in these sites, or 'traps' as they are usually called, as long as the temperature is not raised. In the measurement process, heating causes eviction from traps at a temperature characteristic of the type of trap. Some of these evicted electrons reach luminescence centers and in the process of recombination there is emission of light. The amount of light is proportional to the number of trapped electrons. This number depends on the amount of nuclear radiation to which the crystal had been exposed during its burial and on the sensitivity of the crystal to ionizing radiation. Hence, by measuring the sensitivity of the sample to acquiring thermoluminescence, archaeologically acquired thermoluminescence and the radiocarbon content of the surrounding soil it is possible to calculate the age. Sensitivity of the sample is measured by exposing it to a calibrated radio isotope source. Several methods for measuring the radioactive content of the surrounding soil are discussed in the literature. In this study, we focus our attention on estimating the acquired thermoluminescence during sediment burial.

Lack of knowledge about the amount of thermoluminescence at the time of deposition of sediments burdens the problem of estimating the acquired thermoluminescence during burial. The same reason invalidates the direct applicability of the procedures that are already developed for pottery dating which hinge on the assumption that the act of firing zeros the thermoluminescence acquired during geological times. Wintle and Huntley [60] report that the study of the reduction in TL caused by various sunlamp exposures helps

us understand the TL at the time of deposition. Initially the TL is rapidly removed by the sunlamp exposure but for times longer than one hour the TL is reduced much more slowly. Therefore, Wintle and Huntley [60] assume that the natural TL is made up of two components; an easily bleachable component I_d and a residual component I_0, which is the component that cannot be bleached by a laboratory sunlamp exposure. Wintle and Huntley [60] assume this residual component I_0 to be the TL at the time of deposition of the sediment. The easily bleachable component I_d is assumed to be the TL due to the radiation dose since deposition. Each gives rise to a different fraction of the TL at different glow curve temperatures. The total TL $I(T)$ at temperature T is thus given by $I(T) = I_0(T) + I_d(T)$. At a given temperature, only the total TL $I(T)$ is measurable. However, for dating purposes, it is required to measure the component $I_d(T)$. Physicists have developed several techniques to quantify the components $I_0(T)$ and $I_d(T)$ by simply measuring the TL of natural samples and of samples left in the sun (or exposed to a sunlamp in the laboratory). In this study we looked at estimating the equivalent dose from data collected for three experimental techniques: the additive dose method, the partial bleach method and the regeneration method. A brief introduction to these techniques is presented in Chapter 2. This chapter also provides a description of the data. Plausible physically motivated models for thermoluminescence data are also introduced in Chapter 2.

Exposure to sunlight (or to a sunlamp in the laboratory) drives off the trapped electrons and reduces the intensity of the TL signal. This is called 'bleaching' the sample. Hereafter we refer to the data collected on natural samples as 'unbleached data'. The data collected on samples that are given a laboratory bleaching are referred to as 'bleached data'. Chapter 3 describes model fitting for the unbleached data (or the bleached data) collected at a given temperature. This fitting process applies to the additive dose method. We examine four estimation techniques: maximum likelihood, quasi-likelihood, weighted least squares and a slightly modified weighted least squares technique. Formulae for approximate biases and mean squared errors of these estimators are derived for the small σ asymptotic case. We examine the assumptions made in estimating the biases and assess the validity of the derived formulae for σ in the range of real samples by a Monte Carlo study. These results are also

presented in Chapter 3.

In Chapter 4, we compare the four estimation methods by examining their asymptotic properties. In Section 4.1, we examine the large sample asymptotic behavior of the estimators. We show that the maximum likelihood and quasi-likelihood estimators are consistent and generalized least squares and data weighted least squares estimators are generally not. Distributional approximations for maximum likelihood and quasi-likelihood estimators are also provided in Section 4.1. In Section 4.2, we analyze the large sample behavior of our small σ approximation to $\hat{\theta} - \theta$. In this limiting case, we show that the quasi-likelihood estimator and the maximum likelihood estimator are mean squared error consistent. In general, generalized least squares and data weighted least squares estimators were found to have biases that do not vanish even asymptotically. However, the analysis of small σ large n behavior of the estimators reveals that for the parameter of interest in our model, generalized least squares and data weighted least squares also produce asymptotically unbiased estimators. The discussion presented in Section 4.2 is valid for more general response functions than simply the response functions described for the additive dose method, the partial bleach method and the regeneration method. Examining the behavior of our small σ approximations we found that for the parameter of interest in thermoluminescence studies, maximum likelihood and generalized least squares have almost the same bias for σ values and sample sizes used in practice.

In the partial bleach method, the equivalent dose is estimated as the dose corresponding to the intersection of the fitted response curves for the unbleached and bleached data. Chapter 5 describes estimation of the equivalent dose from partial bleach data together with an estimated standard error of the estimate. Formulae are derived for the biases and the mean squared errors of the estimators for the small σ asymptotic case. The validity of the formulae for σ in the range of typical samples is explored by simulation studies. Interval estimation is also discussed for each method. For maximum likelihood, we describe computing profile likelihood intervals and symmetric intervals with Z and t critical values. We also describe confidence intervals based on a transformation of the likelihood ratio statistic with a transformed F critical value. For quasi-likelihood, we describe symmetric

confidence intervals based on a t quantile and also based on inversion of the quasi-score test. The finite sample performance of the suggested confidence intervals is examined by simulation studies. These are described in Chapter 5.

In the regeneration method, the equivalent dose is estimated as the dose shift necessary for the unbleached response curve to match the bleached response curve. Chapter 6 describes estimating the equivalent dose from regeneration data. Here again we examine the four estimators mentioned earlier. For each method we provide formulae for the biases of the estimators. As for the partial bleach method, finite sample performance of the asymptotic theoretical results are examined by a simulation study. These results are presented in Chapter 6.

In thermoluminescence studies data are collected at a series of temperatures. In chapters 3 to 6 we focused on analyzing the data at a given temperature. A series of estimates for the equivalent dose are available from separate analyses at different temperatures. These estimates are then plotted against the temperature. A region over which the equivalent dose does not vary with temperature is identified from this plot. This region is called a 'plateau region'. Since the same samples are used to collect the observations over the plateau region, the resulting observations are correlated (in fact correlations are very high). We propose a generalized estimating equations (GEE) procedure closely related to that of Liang and Zeger [43] to estimate the equivalent dose taking correlation into account. In Section 7.2.2, we examine the large sample behavior of the suggested estimate. We provide a formula for the standard error of the estimate. For thermoluminescence data, the sample sizes are quite small compared to the number of parameters fitted in the models suggested for these data. Therefore, confidence intervals for the equivalent dose based on the large sample asymptotic theory are found to have smaller coverage probabilities than the nominal coverages. In Section 7.2.3, we examine the behavior of the estimate in the limit of small measurement error. For this case, we propose constructing confidence intervals based on an approximate t quantile; a formula based on the Satterthwaite's approximation [53] is provided for the degrees of freedom of the suggested t quantile. The finite sample performance of the suggested theoretical results is examined by a Monte Carlo study. Confidence intervals based on the t

quantile were found to have coverage probabilities closer to their nominal levels than those based on standard normal quantiles. The suggested theory is demonstrated on a real data set.

In Chapter 8, we propose tests based on the empirical distribution function (EDF tests) of the fitted standardized residuals for testing the assumption of normality of random errors. In Section 8.2, we present a test procedure for testing normality which does not require the fitted regression model to be correct. In Section 8.2.6, we prove the weak convergence of the related empirical processes. Section 8.3 offers a test procedure for testing the assumption of normality assuming the fitted regression model is correct. For each case, we show how to compute the test statistic and an approximate p-value for testing the assumption of normality. The finite sample performance of the suggested asymptotic theory is examined by a simulation study. Section 8.6 summarizes the chapter.

Chapter 9 summarizes the conclusions and suggests guidance for analyzing TL data. We end Chapter 9 with proposals for further work.

Chapter 2

Description of the data

A clear understanding of the data is the foundation of a "good" data analysis. Section 2.1 describes the data collected in thermoluminescence studies. The goal of the experiment is to determine the age of the sample by estimating the TL acquired during sediment burial. Section 2.2 describes some widely used techniques for this purpose[1]. These methods require fitting of nonlinear response curves to the data. Some plausible models for the response curves and for the error structure are suggested in Section 2.3. Model fitting is discussed in subsequent chapters.

2.1 Method of data collection

Sedimentary samples are collected from dunes in a manner that avoids exposure to sunlight. (Exposure to sunlight drives off the trapped electrons and zeroes — or decreases — the TL signal.) Further experiments are carried out under subdued orange light.

About 2mm thick of the outer layer is etched away from each surface of the core sample. The sample is then purified by subjecting it to a series of acidification and oxidation steps and washings. Desired sized grains are then separated out from the resulting slurry, and are dried on aluminum discs of about 1cm in diameter. About 50 such subsamples are prepared from each core. After weighing each subsample, some of the subsamples are placed in an

[1]Much of the content of sections 2.1 and 2.2 is based on Aitken [1] and Wintle and Huntley [60].

oven. While heating the sample gently, the photon count and the corresponding temperature is recorded every few minutes (usually 5 to 10 minutes). These photon counts represent the 'natural thermoluminescence'. The other samples are irradiated with different known gamma doses before heating in the oven. The TL signal from these samples is simply called 'thermoluminescence'.

In summary, a measure of the photon counts per unit mass, the corresponding temperature and the amount of added dose are recorded from each subsample. Usually the dose is measured in gray[2].

2.2 Equivalent dose (ED) determination

The plot of TL vs the temperature is called a 'glow curve'. Figure 2.1 illustrates a set of glow curves obtained for a data set (coded as WFP2-7R1) kindly provided to us by D.J. Huntley. In Figure 2.1, each glow curve represents photon counts observed from one subsample over the corresponding series of temperatures. For a fixed temperature, the plot of TL vs added dose is called an 'additive dose curve' or a 'dose response curve'. Figure 2.2 illustrates such an additive dose curve obtained for the unbleached data set (QNL84-2) given in Berger *et. al.* [12]. In Figure 2.2, each dot represents the photon count obtained from one subsample at temperature 300^0C. For young samples the additive dose curve is nearly linear; see Berger [6]. However, for older samples the additive dose curve departs from linear behavior. Cubic or saturating exponential models[3] appear more suitable for the additive dose curves for older samples (See Berger [9], Berger [6], Berger *et. al.* [5]).

The ED is defined as the dose required to produce an amount of thermoluminescence equivalent to that which the sample had acquired during burial. Different techniques are used to estimate the equivalent dose from the additive dose curves. The appropriateness of the method depends on the the age of the core and in particular, the non-linearity of the response curves. Next we briefly describe these methods.

[2]The 'gray', abbreviated Gy, is defined as the dose required to provide 1 joule of energy per kilogram of material.

[3]Saturating exponential models are defined in Section 2.3.

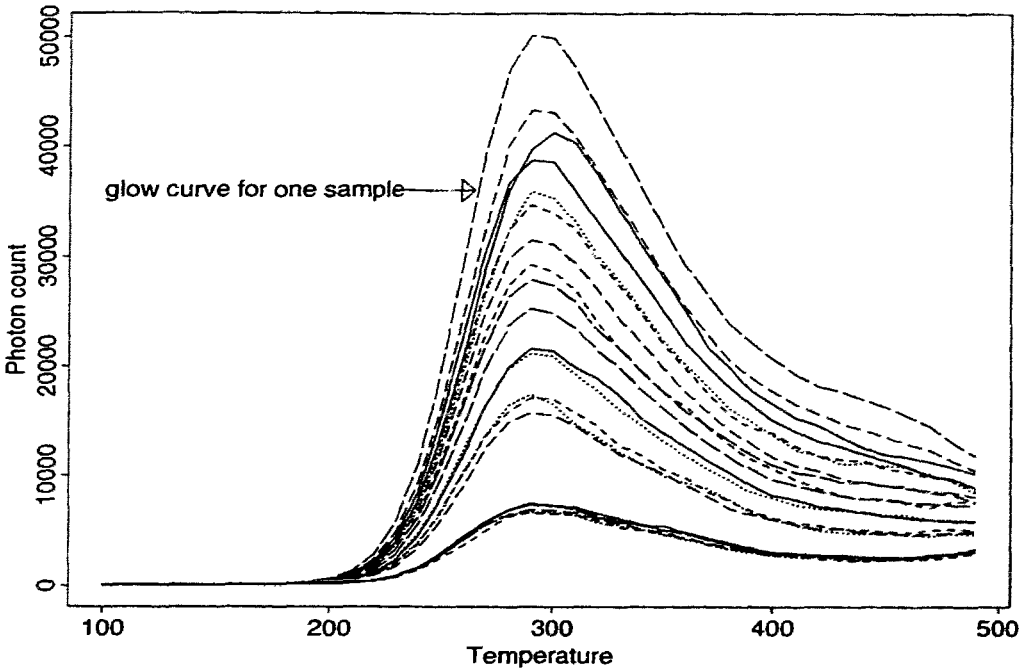

Figure 2.1: Glow curves for the data WFP2-7R1

2.2.1 Additive dose method

In this method, the equivalent dose is estimated as the dose corresponding to the level of TL at the time of deposition of sediments. To read the equivalent dose we need to extrapolate the additive dose curve back to the point where the TL at the deposition would be. Figure 2.2 illustrates this method. This plot was obtained for the unbleached data set of QNL84-2 data given in Berger *et. al.* [12]. The complete QNL84-2 data set was collected for the partial bleach method. Here we use the unbleached data set for illustration of the additive dose method. It is worth noting that for the additive dose method no samples are bleached and the unbleached response curve would be the only response curve available, had the data been collected for the additive dose method. Since it is necessary to extrapolate the response curve to read the equivalent dose, the additive dose method is more appropriate for young samples for which the additive dose curve is nearly linear.

As illustrated in Figure 2.2, one needs to know the TL of the sample at the time of deposition ($Y(0)$) to use this method. In pottery dating, it is quite certain that initially

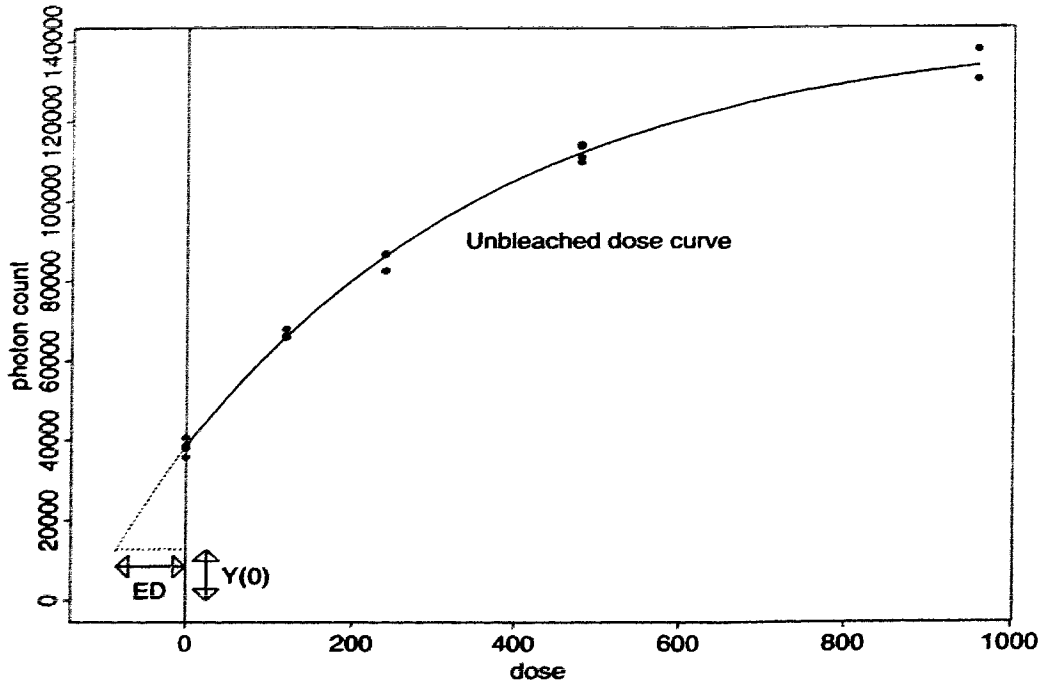

Figure 2.2: Unbleached response curve for the data QNL84-2: additive dose method

the TL is zero (i.e. $Y(0) = 0$) because heating to very high temperatures at the time of firing releases the trapped electrons. However, in sedimentary dating, the amount of TL at the time of deposition is not known. Wintle and Huntley [60] assume that only the easily bleachable component had been removed at the time of deposition. Therefore, the residual TL, I_0, remaining after a long laboratory bleaching is assumed to be the TL had it been measured at the time of deposition of sediments (i.e. take $Y(0) = I_0$). However, if the bleachable TL had not been fully removed at the time of deposition, the equivalent dose thus determined is erroneous. The estimation is done on the assumption that the bleachable TL had been fully removed. This assumption is then tested by the 'plateau test' described next (Aitken [1]).

The plateau test

From separate analyses of data at different temperatures an estimate for the equivalent dose is available for each temperature. These estimated equivalent doses are then plotted against

the temperature. A 'plateau' is the region where we observe that the ED does not change with the temperature. Figure 2.3 illustrates the estimated equivalent dose vs temperature for the data WFP2-7R1 cited in Section 2.2. In this plot, each dot represents the equivalent dose estimated from the data collected at the corresponding temperature. If the bleachable TL was not fully removed at the time of deposition of sediments, a plateau may not be apparent (Aitken [1]).

The 'plateau test' is not merely testing whether or not the bleachable TL had been fully removed at the time of deposition of sediments. Only those traps that have accumulated electrons without leakage can provide reliable information for dating purposes. These traps are identified as the traps corresponding to the 'stable' region of the glow curve. This 'stable' region is recognized as the region corresponding to the 'plateau' on the plot of equivalent dose vs temperature. (Aitken [1],Berger *et. al.* [10], Huntley *et. al* [38]).

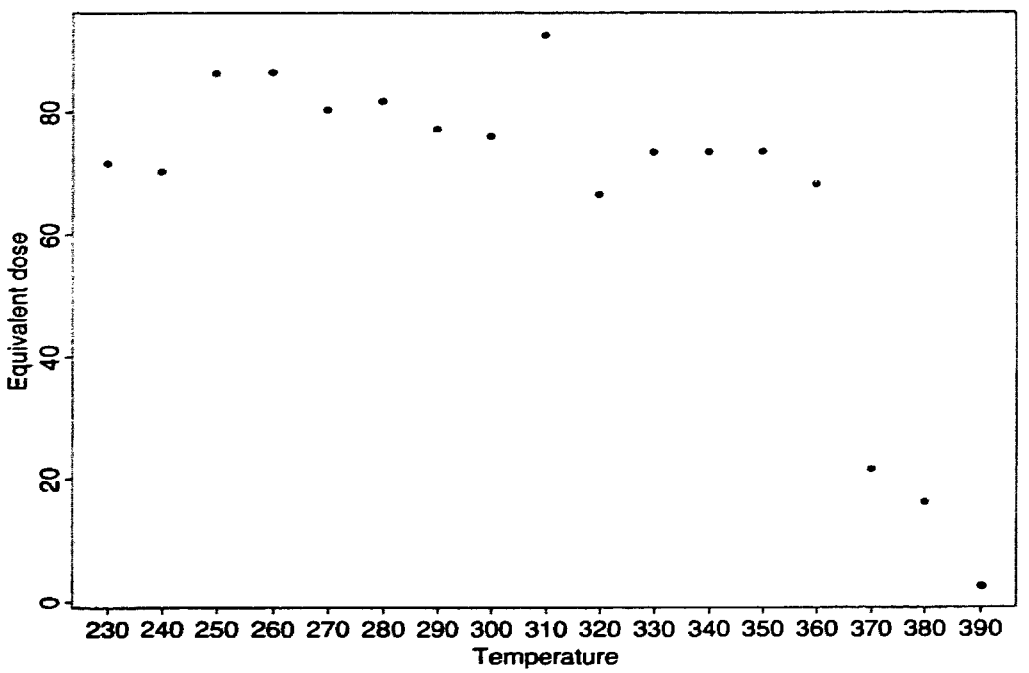

Figure 2.3: Plot of estimated equivalent dose vs temperature : Data WFP2-7R1

The lack of knowledge about the TL at the time of deposition of sediments, makes the additive dose method less suitable for sedimentary dating. The partial bleach method

and the regeneration method that we describe next are two methods that are widely used for sedimentary dating which avoid the need to know the amount of TL at the time of deposition.

2.2.2 Partial bleach method

The partial bleach method or the $R - \Gamma$ method was introduced by Wintle & Huntley in 1980 [60]. Here a portion of the samples is used to determine the additive dose curve as in the additive dose method. This dose curve is called the 'unbleached dose curve'. Remaining samples are subjected to afternoon sunlight for about 40 minutes and are then irradiated with Gamma doses to define the additive dose curve for the bleached portion. This is called the 'bleached dose curve'. Figure 2.4 illustrates the unbleached and bleached dose curves for the data QNL84-2 given in Berger *et. al* [12]. This method is so named because the sample is only partially bleached so that the bleaching does not totally erase the signal. Here we measure the reduction in TL (R) caused by the sunlamp exposure. A plot of R vs the dose (Γ) when extrapolated to $R = 0$ yields the ED on the Γ axis. Equivalently, the ED can be estimated as the dose corresponding to the point of intersection of the bleached and unbleached curves.

As we mentioned earlier, the estimates derived from the additive dose method are erroneous if the bleaching prior to deposition was not complete. The partial bleach method works well even if the bleaching prior to deposition of sediments was incomplete as long as it was more complete than that caused by the short laboratory bleaching (Aitken [1]).

As for the additive dose method, the partial bleach method also requires extrapolation of the additive dose response curves. Therefore, the partial bleach method is also preferred for samples for which the additive dose curves are nearly linear or can be reliably extrapolated.

2.2.3 Regeneration method

The regeneration method is more appropriate for old samples. As in the previous method, a portion of the sample is used to define the unbleached curve and the remaining samples are used to define the bleached curve. Here the bleaching is more vigorous. In the regeneration

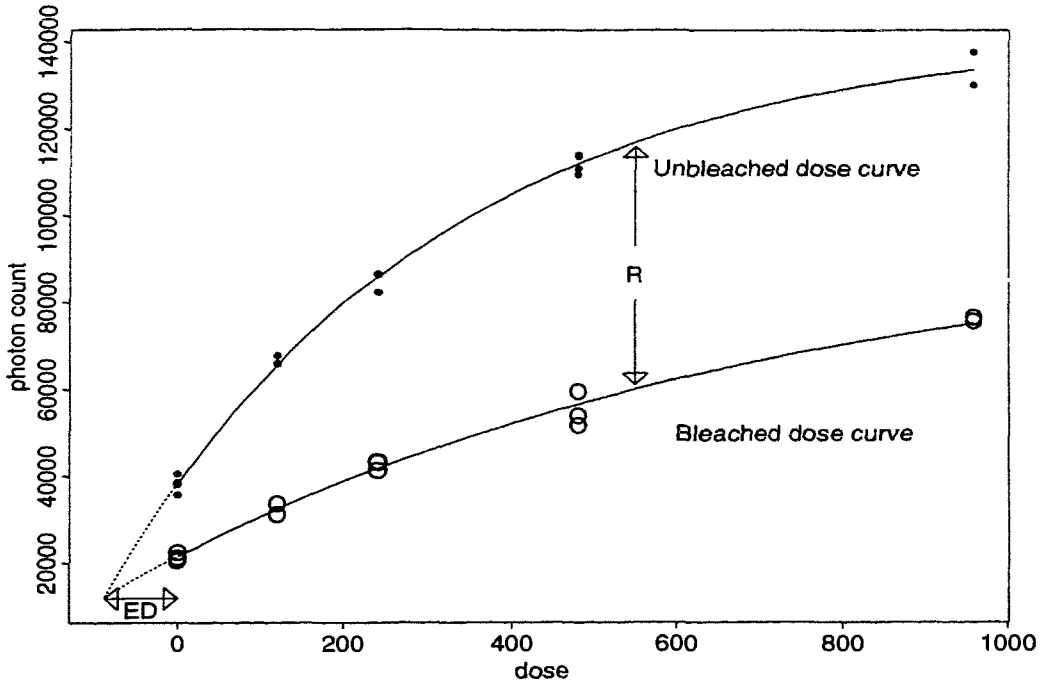

Figure 2.4: Dose response curves for the data QNL84-2

method, for the bleached curve, samples are left in the sun for about 15 hours. This totally erases the TL signal. Then artificial irradiations are administered, thereby regenerating the TL growth characteristic. To avoid the possible effect of sensitivity change due to bleaching a new portion of the sample is used for each data point. The purpose of the bleached curve is to use it as a pattern for the unobservable portion of the unbleached curve. If the bleaching had not caused a sensitivity change it should be possible to match the two curves by a shift along the dose axis (Huntley *et. al.* [38]). The equivalent dose is calculated as the dose shift required for the unbleached curve to match the bleached curve. Figure 2.5 illustrates the unbleached and bleached response curves for the data set SESA1 cited in Huntley *et. al.* [38]. This data were kindly provided to us by D.J. Huntley.

The equivalent dose can also be determined as follows. The dose R required to generate a level equal to the natural TL is determined on the bleached curve. The procedure is then repeated for portions which have been given a known laboratory dose β. The equivalent dose is then estimated as the horizontal intercept of the plot of R vs the additive dose β;

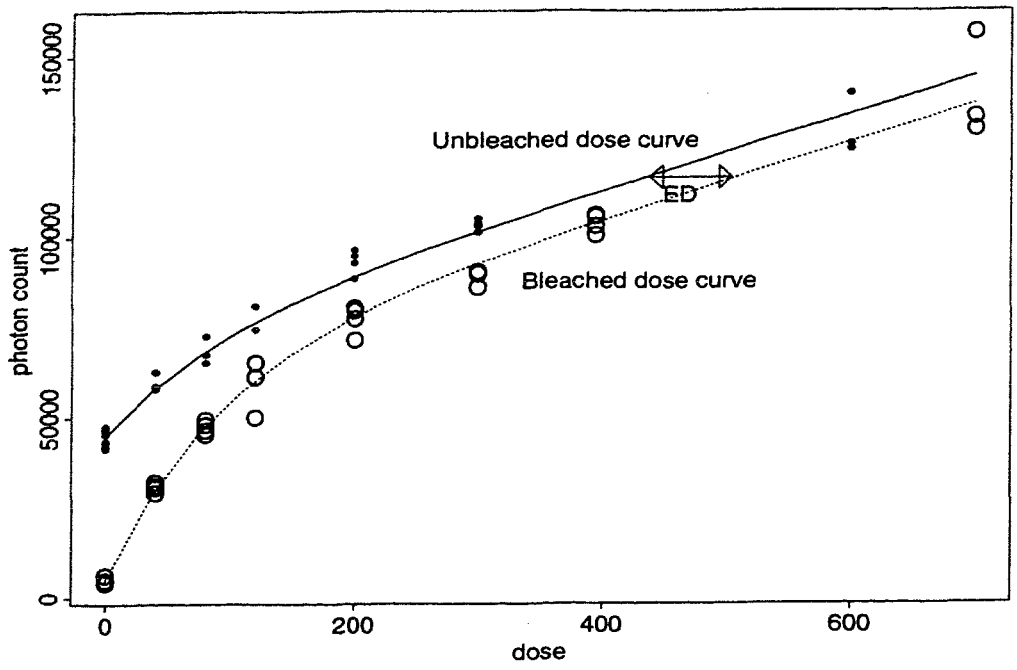

Figure 2.5: Dose response curves for the data SESA1

see Aitken [1].

2.3 Plausible models

According to Huntley *et. al.* [36] (see also Kirkey [42]) only a few particles in a sample produce most of the TL. The grains that glow upon gentle heating in the laboratory (or emit light in the visible range) are called 'emitting grains'. Franklin *et. al.* [30] reports that only about 8% of the grains are emitting grains (or produce the observed TL). In the following, we describe models for the mean photon count emitted by the sample at a given temperature, in response to the dose received (natural dose during burial plus the artificial dose administered).

2.3.1 Models for the mean response

Notation:

We use the subscripts i and j to denote the dose and the temperature of the subsample respectively. We refer to the kth replicate subsample receiving dose i as the ikth subsample. Let

N_{ijk} = the number of emitting grains in the ikth sub sample at temperature j

m_{ik} = the mass of the ikth subsample

and X_{ijkl} = emission from the lth grain of the ikth sub sample at the jth temperature

We make the following assumptions.

1. As in Berger *et. al.* [11], the TL signal is assumed to be made up of the photons emitted by the individual grains in the subsample.
2. The emissions from different grains in the same subsample are independent.
3. All emitting grains are identical and grains that have received the same laboratory dose emit at the same rate.
4. The rate of emission from a given grain is independent of the number of emitting grains in the sample.
5. At any temperature j, $X_{ijkl} \sim Pois(\lambda_{ij})$, where the rate λ_{ij} is a function of the dose x_i and the temperature j.
6. All the emitted photons are detected. (See the remarks below.)
7. The laboratory applied dose, x_i, can be measured without any error.
8. The mass of the sample is measured without any error.

Remarks:

The calculations we make are still valid if we make the weaker assumption that only a fraction of the emitted photons are detected, but all the emitted photons have the same

chance p of being detected. In this case, the N_{ijk} in the given formulae has to be replaced by ND_{ijk}, which is, the number of emitting particles detected from the ikth sample. The $E(ND_{ijk})$ and $V(ND_{ijk})$ can be computed as follows:
Noting that $ND_{ijk} \sim bin(N_{ijk}, p)$, we find

$$\begin{aligned} E(ND_{ijk}) &= E\left[E\left(ND_{ijk} \middle| N_{ijk}\right)\right] \\ &= E(N_{ijk}p) = pE(N_{ijk}), \end{aligned}$$

and

$$\begin{aligned} V(ND_{ijk}) &= V\left[E\left(ND_{ijk} \middle| N_{ijk}\right)\right] + E\left[V(ND_{ijk} \middle| N_{ijk})\right] \\ &= V(pN_{ijk}) + E\left[N_{ijk}p(1-p)\right] \\ &= p^2V(N_{ijk}) + p(1-p)E(N_{ijk}). \end{aligned}$$

In the calculations presented below, we assume that all the emitted particles are detected. (i.e. $p = 1$.)
The total emission from the ikth subsample is

$$E_{ijk} = \sum_{l=1}^{N_{ijk}} X_{ijkl}.$$

Therefore,

$$\begin{aligned} E\left(E_{ijk}\right) &= E\left[E\left[\sum_{l=1}^{N_{ijk}} X_{ijkl} \middle| N_{ijk}\right]\right] \\ &= E\left(N_{ijk}\right) E\left(X_{ijkl}\right) \\ &= E\left(N_{ijk}\right) \lambda_{ij} \end{aligned}$$

and

$$\begin{aligned} V\left(E_{ijk}\right) &= V\left[E\left[\sum_{l=1}^{N_{ijk}} X_{ijkl} \middle| N_{ijk}\right]\right] + E\left[V\left[\sum_{l=1}^{N_{ijk}} X_{ijkl} \middle| N_{ijk}\right]\right] \\ &= V\left(N_{ijk}\right) \lambda_{ij}^2 + E\left(N_{ijk}\right) \lambda_{ij}. \end{aligned}$$

Subject to the assumptions listed earlier, the expected TL can be computed as follows. The emission per unit mass of ikth subsample is

$$TL_{ijk} = E_{ijk}/m_{ik}.$$

Therefore,

$$E\left(TL_{ijk}\right) = E\left(E_{ijk}\right)/m_{ik} = E\left(N_{ijk}\right)\lambda_{ij}/m_{ik} \tag{2.1}$$

and

$$\begin{aligned} V\left(TL_{ijk}\right) &= V\left(E_{ijk}\right)/m_{ik}^2. & (2.2)\\ &= V\left(N_{ijk}\right)\lambda_{ij}^2/m_{ik}^2 + E\left(N_{ijk}\right)\lambda_{ij}/m_{ik}^2. & (2.3) \end{aligned}$$

Without loss of generality we take the available y observations as the photon counts per unit mass. If the data are not normalized by dividing by the mass of the sample then take $m_{ik} = 1$.

We consider several models.

Model 1:

This is the model considered in Berger *et. al.* [11]. They suggest the following. For grains of diameter $< 10\mu m$, $E(N_{ijk})$ is estimated to be of the order $10^3 - 10^4$, λ_{ij} is of the order 10^2 and $V(N_{ijk})$ is approximately of the order $10^2 - 10^4$. So, the term $E\left(N_{ijk}\right)\lambda_{ij}$ is small compared to the term $V\left(N_{ijk}\right)\lambda_{ij}^2$. Ignoring this term,

$$V\left(TL_{ijk}\right) \approx V\left(N_{ijk}\right)\lambda_{ij}^2/m_{ik}^2.$$

Therefore, $\frac{V(TL_{ijk})}{E^2(TL_{ijk})} \approx \frac{V(N_{ijk})}{E^2(N_{ijk})}$ is a constant.

Let $f(x_i,\theta) = E\left(TL_{ijk}\right)$ be the mean response for dose x_i. For notational convenience we denote the observed photon count TL_{ijk} by y_{ijk}. We drop the suffix j when we refer to the observed photon counts taken at a given temperature. A simple model for the observed photon count y_{ik} and the dose received x_i can therefore be described by $y_{ik} = f(x_i,\theta)(1+\sigma\epsilon_{ik})$, where the constant $\sigma = \frac{\sqrt{V(TL_{ijk})}}{E(TL_{ijk})}$ is referred to as the *relative error in a single measurement*; θ is a vector of parameters that we wish to estimate and ϵ_{ik} is the random error in the photon count observed from the ikth sample.

The mean response function $f(x_i,\theta)$ indicates how the expected photon count varies with the applied dose x_i. The plot of $f(x_i,\theta)$ against the applied dose defines the dose response curve.

For sediments younger than 10-20 kilo years (Ka[4]) the dose response curves are nearly linear. Thus, for sediments of this age the deterministic part of the model is described by the relation $f(x_i, \theta) = \alpha + \beta\, x_i$ where $\theta = (\alpha, \beta)$ is the vector of parameters to be estimated and x_i is the applied dose.

For sediments older than $20Ka$, dose response curves are not linear. More precisely these have been described as sub-linear dose response curves. Sub-linear curves show a linear relationship with a positive slope between TL and the applied dose at low dose levels but curvature is observed as the dose level increases. Quadratic models and cubic models have been used (Berger *et. al.* [9], [6], [5]) to describe dose response curves for moderately old samples. Some response curves show a flat TL intensity at very high dose levels. This happens when the traps are filled and no more capturing of electrons is possible. This is described as the saturation point and can be observed for very old samples (greater than $100Ka$) or with young samples that are subjected to very high applied doses. (Berger [9].) Huntley [37] introduces saturating exponential models (below we provide motivation for using saturating exponential models) to describe the response curves for samples approaching saturation. The saturating exponential model is represented by the function $f(x_i, \theta) = \alpha_1 \left\{1 - \exp\left[-\frac{(x_i+\alpha_2)}{\alpha_3}\right]\right\}$, where $\theta = (\alpha_1, \alpha_2, \alpha_3)$ is the vector of parameters we wish to estimate and x_i is the applied dose.

Motivation for using the saturating exponential model

Suppose the probability of filling an empty trap per unit dose of radiation is constant (say μ) . Let $N(d)$ denote the number of empty traps available, when the total acquired dose is d units. Then

$$E\left[N(d+\delta d) - N(d) \middle| N(d)\right] = -\mu N(d)\delta d + o(\delta d).$$

Let $E(N(d)) = M(d)$. Taking expectations we find

$$M(d+\delta d) - M(d) = -\mu M(d)\delta d + o(\delta d).$$

Taking limits as $\delta d \longrightarrow 0$, and solving the resulting differential equation, we find

$$M(d) = \exp(-\mu d + C),$$

[4]The letter 'a', from the French 'an', is used when specifying annual dose, following practice in most TL laboratories. (Aitken [1].)

where C is the constant of integration. Let $M(0)$ denote the total number of available traps. Noting that all the traps were empty when no dose was acquired, we find $\exp(C) = M(0)$. Further, assume that no new traps are formed due to acquired radiation. Then, at saturation all traps are filled and therefore the intensity at saturation, I_{sat}, is proportional to $M(0)$. Thus, $I_{sat} = C_1 M(0)$, where C_1 is some constant. When d units of radiation is acquired the expected number of filled traps is $M(0) - M(d)$. Consequently, the intensity at dose d, can be computed as:

$$\begin{aligned} I_d &= C_1(M(0) - M(d)) \\ &= C_1(M(0) - M(0)\exp(\mu d)) \\ &= I_{sat}\left[1 - \exp(-\mu d)\right]. \end{aligned}$$

The total acquired dose d is made up of the laboratory applied dose x and the dose acquired during burial γ. Thus, $I_d = I_{sat}\left[1 - \exp(-\mu(x + \gamma)\right]$.

The distribution of the random error

Three possible choices for the distribution of the random error are:

1. As in Berger *et. al.* [11] assume the random error ϵ to be normally distributed.
2. The assumption that the standard deviation of the TL count is proportional to its mean suggests a gamma distribution as a plausible model for the random error.
3. Use quasi-likelihood estimating techniques where the distribution of ϵ is unspecified.

In Chapter 8 we propose procedures for testing the assumption of normality of random errors.

We note that photon counts are never negative. Therefore the normal error model is suitable only if $\sigma < 1/3$, roughly.

Model 2:

This is a simple extension of Model 1. Here we use the complete variance function instead

of the approximate function used in Model 1. From equations 2.1 and 2.3 we have

$$\begin{aligned} E(y_{ik}) &= E(N_{ijk})\,\lambda_{ij}/m_{ik} \\ V(y_{ik}) &= E(N_{ijk})\,\lambda_{ij}/m_{ik}^2 + V(N_{ijk})\,\lambda_{ij}^2/m_{ik}^2 \\ &= \delta_1 E(y_{ik}) + \delta_2 E^2(y_{ik}), \end{aligned}$$

where $\delta_1 = \frac{1}{m_{ik}}$ and $\delta_2 = \frac{V(N_{ijk})}{E^2(N_{ijk})}$.

Thus, in this model, we have an additional parameter, δ_2, (notice that δ_1 is measurable) in the variance function. Techniques are available in the literature (see for eg. Davidian and Carroll [22]) on how to estimate such parameters in the variance function.

The distribution of the random error

The gamma model is unsuitable for the distribution of random errors in Model 2. Two possible choices for estimating the parameters are:

1. quasi-likelihood estimating techniques where the distribution of ϵ is unspecified.

2. Assuming random errors ϵ approximately follow a normal distribution, model y_i as $y_i = f(x_i, \theta)(1 + \sigma\epsilon_i)$ with $E(y_i)$ and $V(y_i)$ as defined above.

Model 3:

Let

$$\begin{aligned} m_{ik} &= \text{mass of the } ik\text{th subsample}, \\ M_{ik} &= \text{mass of the emitting grains in the } ik\text{th subsample}(\leq m_{ik}), \\ \text{and } E_{ijk} &= \text{total emission from the } ik\text{-th sample at temperature } j. \end{aligned}$$

Assume that $E_{ijk} \sim Pois(\lambda_{ij} M_{ik})$. Then,

$$\begin{aligned} E(E_{ijk}) &= E[E(E_{ijk}|\, M_{ik})] \\ &= E(\lambda_{ij} M_{ik}) \end{aligned}$$

Assuming $E(M_{ik}) = \zeta m_{ik}$, for some constant ζ, we write $E(E_{ijk}) = \lambda_{ij}\zeta m_{ik}$. The variance of E_{ijk} can be computed as

$$\begin{aligned} V(E_{ijk}) &= V[E(E_{ijk}|M_{ik})] + E[V(E_{ijk}|M_{ik})] \\ &= V(\lambda_{ij}M_{ik}) + E(\lambda_{ij}M_{ik}) \\ &= \lambda_{ij}^2 V(M_{ik}) + \lambda_{ij}\zeta m_{ik}. \end{aligned}$$

Since $TL_{ijk} = \frac{E_{ijk}}{m_{ik}}$, we find $E(TL_{ijk}) = \zeta\lambda_{ij}$ and

$$\begin{aligned} V(TL_{ijk}) &= \frac{1}{m_{ik}^2}V(E_{ijk}) \\ &= \frac{1}{m_{ik}^2}\left\{\lambda_{ij}^2 V(M_{ik}) + \lambda_{ij}\zeta m_{ik}\right\} \\ &= \frac{\lambda_{ij}^2}{m_{ik}^2}V(M_{ik}) + \frac{E(TL_{ijk})}{m_{ik}} \\ &= \frac{1}{\zeta^2 m_{ik}^2}V(M_{ik})E^2(TL_{ijk}) + \frac{1}{m_{ik}}E(TL_{ijk}) \\ &= \delta_1 E(TL_{ijk}) + \delta_2 E^2(TL_{ijk}), \end{aligned}$$

where $\delta_1 = \frac{1}{m_{ik}}$ and $\delta_2 = \frac{V(M_{ik})}{E^2(M_{ik})}$. Thus, we have the additional parameter δ_2 (δ_1 is measurable) to estimate; again these can be estimated as in Model 2.

2.3.2 Correlation structure

According to the models suggested earlier, if $i \neq i'$ or $i = i'$ but $k \neq k'$ then $Cov(TL_{ijk}, TL'_{i'j'k'}) = 0$, since these correspond to observations from different discs. If $i = i'$ and $k = k'$ then $Cov(TL_{ijk}, TL'_{i'j'k'}) = \eta_{jj'}$, which can be computed as

$$\rho_{jj'} = Cov(TL_{ijk}, TL_{ij'k}) = \frac{1}{m_{ik}^2}Cov(E_{ijk}, E_{ij'k}).$$

Now consider

$$\begin{aligned} Cov(E_{ijk}, E_{ij'k}) &= E(E_{ijk}E_{ij'k}) - E(E_{ijk})E(E_{ij'k}) \\ &= E\left[(\sum_{l=1}^{N_{ijk}} X_{lijk})(\sum_{m=1}^{N_{ij'k}} X_{mij'k})\right] - \lambda_{ij}\lambda_{ij'}E(N_{ijk})E(N_{ij'k}) \\ &= E\left[E\left[(\sum_{l=1}^{N_{ijk}} X_{lijk})(\sum_{m=1}^{N_{ij'k}} X_{mij'k})\middle| N_{ijk}, N_{ij'k}\right]\right] - \lambda_{ij}\lambda_{ij'}E(N_{ijk})E(N_{ij'k}) \end{aligned}$$

Since X's are independent, the above can be simplified to get

$$\begin{aligned} Cov(E_{ijk}, E_{ij'k}) &= E\left[N_{ijk}N_{ij'k}E(X_{lijk})E(X_{mij'k})\right] - \lambda_{ij}\lambda_{ij'}E(N_{ijk})E(N_{ij'k}) \\ &= \lambda_{ij}\lambda_{ij'}E(N_{ijk}N_{ij'k}) - \lambda_{ij}\lambda_{ij'}E(N_{ijk})E(N_{ij'k}) \\ &= \lambda_{ij}\lambda_{ij'}Cov\left(N_{ijk}, N_{ij'k}\right). \end{aligned}$$

The correlation $\rho_{jj'}$ between the observations collected on the same disc at different temperatures j and j' can therefore be computed as follows.

Model 1

For Model 1,

$$\begin{aligned} \rho_{jj'} &= \frac{\lambda_{ij}\lambda_{ij'}Cov\left(N_{ijk}, N_{ij'k}\right)}{\sqrt{\lambda_{ij}^2 V(N_{ijk})}\sqrt{\lambda_{ij'}^2 V(N_{ij'k})}} \\ &= Cor(N_{ijk}, N_{ij'k}) \\ &\leq 1. \end{aligned}$$

Remarks

If the same grains emit photons at temperatures j and j' then $N_{ijk} = N_{ij'k'}$. In this case, Model 1 predicts perfect correlation. Further note that for all k, $E(Y_{ik}) = f(x_i, \theta_j)$ and $V(Y_{ik}) = \sigma^2 f^2(x_i, \theta_j)$. Therefore, $E(E_{ijk}) = \lambda_{ij}E(N_{ijk}) = f(x_i, \theta_j)m_{ik}$ and $V(E_{ijk}) = \lambda_{ij}^2 V(N_{ijk}) = \sigma^2 f^2(x_i, \theta_j)m_{ik}^2$. Consequently, $\frac{E(N_{ijk})}{\sqrt{V(N_{ijk})}} = \frac{1}{\sigma}$. Thus, a gamma distribution with shape parameter $\frac{1}{\sigma^2}$ and scale parameter $\beta_{ik} = \sigma^2 \frac{f(x_i,\theta_j)m_{ik}}{\lambda_{ij}}$ is a plausible candidate for the distribution of N_{ijk}. We note that N_{ijk}, the number of emitting grains in the sample, can take only integer values, but if N_{ijk}'s are large, the distribution can reasonably be approximated by a continuous distribution.

Model 2

For Model 2,

$$\begin{aligned} \rho_{jj'} &= \frac{\lambda_{ij}\lambda_{ij'}Cov(N_{ijk}, N_{ij'k})}{\sqrt{\lambda_{ij}^2 V(N_{ijk}) + \lambda_{ij}E(N_{ijk})}\sqrt{\lambda_{ij'}^2 V(N_{ij'k}) + \lambda_{ij'}E(N_{ij'k})}} \\ &= \frac{Cor(N_{ijk}, N_{ij'k})}{\sqrt{\left[1 + \frac{E(N_{ijk})}{\lambda_{ij}V(N_{ijk})}\right]}\sqrt{\left[1 + \frac{E(N_{ij'k})}{\lambda_{ij'}V(N_{ij'k})}\right]}} \end{aligned}$$

Remarks

For Model 2, note that even under the assumption $N_{ijk} = N_{ij'k'}$, $\rho_{jj'} < 1$. Further note that

$$E(E_{ijk}) = \lambda_{ij}E(N_{ijk}) = f(x_i, \theta_j)m_{ik}$$

and

$$V(E_{ijk}) = \lambda_{ij}^2 V(N_{ijk}) + \lambda_{ij}E(N_{ijk}) = \sigma^2 f^2(x_i, \theta_j)m_{ik}^2.$$

Therefore, for $m_{ik}\sigma^2 f(x_i, \theta_j) > 1$,

$$\begin{aligned}\frac{E(N_{ijk})}{\lambda_{ij}V(N_{ijk})} &= \frac{\lambda_{ij}E(N_{ijk})}{\lambda_{ij}^2 V(N_{ijk})} \\ &= \frac{f(x_i, \theta_j)m_{ik}}{[m_{ik}^2\sigma^2 f^2(x_i, \theta_j) - f(x_i, \theta_j)m_{ik}]} \\ &= \frac{1}{[m_{ik}\sigma^2 f(x_i, \theta_j) - 1]}.\end{aligned}$$

Thus,

$$1 + \frac{E(N_{ijk})}{\lambda_{ij}V(N_{ijk})} = \frac{m_{ik}\sigma^2 f(x_i, \theta_j)}{[m_{ik}\sigma^2 f(x_i, \theta_j) - 1]}.$$

Consequently,

$$\rho_{jj'} = \frac{Cor(N_{ijk}, N_{ij'k})}{\left[\sqrt{\frac{m_{ik}\sigma^2 f(x_i,\theta_j)}{[m_{ik}\sigma^2 f(x_i,\theta_j)-1]}}\sqrt{\frac{m_{ik}\sigma^2 f(x_i,\theta_{j'})}{[m_{ik}\sigma^2 f(x_i,\theta_{j'})-1]}}\right]}.$$

Remarks:

1. According to Model 1, if the variability in the number of emitting grains does not depend on the applied dose, then the correlation between photon counts taken on the same disc at two different temperatures does not depend on the laboratory applied dose x_i.

2. According to Model 2, the correlation between the observed photon counts taken on the same disc at two different temperatures depends on the applied dose x_i. Our notation $\rho_{jj'}$ suppressed this dependence.

Chapter 3

Model fitting

In this chapter attention is focused on fitting the saturating exponential model defined in Chapter 2. We intend to present a detailed study of fitting this model for a single data set with two perspectives in mind. First, the results we derive are directly applicable if the additive dose method is used to estimate the equivalent dose. Second, as we see in chapters 5 and 6, the methodology developed for this case can easily be extended to estimate the equivalent dose from the partial bleach method and the regeneration method.

Section 3.1 presents initial estimates for the parameters. In sections 3.2 to 3.4, we derive Maximum Likelihood (ML), Quasi Likelihood (QL) and Generalized Least Squares (GLS) estimates for the parameters. In Section 3.5, we examine another estimation procedure used by physicists. This procedure is similar to generalized least squares, except that it uses observed Y values in place of the expected Y values in the weight function of generalized least squares. We refer to this procedure as Data Weighted Least Squares (DWLS).

As we described in Chapter 2, our model suggests that the variance function of the photon counts is proportional to the squared mean function. Furthermore, photon counts are never negative. Consequently, the gamma distribution is a natural candidate for the distribution of photon counts (McCullagh [49]). In Section 3.6, we obtain maximum likelihood estimates assuming photon counts follow a gamma distribution.

In Section 3.8, we derive formulae for the biases of the estimators assuming the relative

error in a single measurement, $\sigma \left(= \frac{\sqrt{V(Y)}}{E(Y)}\right)$, is small. We examine this assumption in Section 3.10. In Section 3.11, we examine the biases of the estimators from a Monte Carlo study and compare with the results obtained from the derived formulae. In Section 3.12, we demonstrate the theoretical results discussed in this chapter using a real data set. Section 3.13 summarizes the chapter.

3.1 Initial estimates

Kuo [41] has described two methods that yield initial parameter estimates; a graphical method and a quadratic equation method. For completeness we now briefly describe those two methods. We also describe an alternative method that can be used to obtain initial estimates.

3.1.1 Quadratic equation method

The quadratic equation method proceeds as described by the following steps.

1. The average TL count ($\overline{TL}$) is calculated at each applied dose.

2. As a first step, the dependence of TL on the dose (x) is approximated by a quadratic relationship. In other words, the coefficients r_1, r_2, r_3 of the equation

$$\overline{TL} = r_1 + r_2 x + r_3 x^2$$

   are estimated by regressing $\overline{TL}$ on x, x^2.

3. Near zero dose, the two models (exponential and quadratic) have similar behavior. The coefficients of the saturating exponential model are thus estimated by equating the first and second derivatives of the saturating model at zero dose to those of the quadratic model.

The last step is more clearly expressed as follows:

$$r_1 = \overline{TL}_{x=0} = f(0) = \alpha_1 \left[1 - \exp\left(\frac{-\alpha_2}{\alpha_3}\right)\right],$$

$$
\begin{aligned}
r_2 &= \overline{TL}'_{x=0} = f'(0) = \frac{\alpha_1}{\alpha_3}\exp\left(\frac{-\alpha_2}{\alpha_3}\right), \\
\text{and } 2r_3 &= \overline{TL}''_{x=0} = f''(0) = -\frac{\alpha_1}{\alpha_3^2}\exp\left(\frac{-\alpha_2}{\alpha_3}\right)
\end{aligned}
$$

The values for r_1, r_2, r_3 are available from the previous step. Thus, the initial estimates for α_1, α_2 and α_3 can be obtained as follows:

$$
\begin{aligned}
\alpha_1 &= r_1 - \frac{r_2^2}{2r_3}, \\
\alpha_2 &= \frac{r_2}{r_3}\log\left[\frac{r_2^2}{r_2^2 - 2r_1r_3}\right], \\
\text{and } \alpha_3 &= -\frac{r_2}{2r_3}.
\end{aligned}
$$

3.1.2 Graphical method

The graphical method is described by the following steps:

1. The saturating exponential model has the form

$$
f(x,\theta) = E(TL) = \alpha_1\left[1 - \exp\left(\frac{-(x+\alpha_2)}{\alpha_3}\right)\right].
$$

   For large applied doses the thermoluminescence approaches α_1. Thus α_1 is estimated from a plot of $\overline{TL}$ vs dose (x) as the photon count $\hat{\overline{TL}}$ approached at high applied doses.

2. Noting that $(\overline{TL} - \hat{\overline{TL}}) = -\hat{\overline{TL}}\exp\left[\frac{-(x+\alpha_2)}{\alpha_3}\right]$, the parameters α_2 and α_3 are then estimated from a plot of $\ln|(\overline{TL} - \hat{\overline{TL}})|$ vs x.

Next we suggest another method that can be used to find the initial estimates.

3.1.3 Alternative method

The model of interest is $y = f(x,\theta)(1+\sigma\epsilon)$, where $f(x,\theta) = \alpha_1\{1 - \exp[\frac{-(x+\alpha_2)}{\alpha_3}]\}$; here $\theta = (\alpha_1, \alpha_2, \alpha_3)$ is the vector of unknown true parameters and x is the applied dose. Noting that as $x \to \infty$, $f(x,\theta) \to \alpha_1$, we estimate α_1 by the average of the photon counts corresponding to the highest dose. At the low dose values the dose response curve is

approximately linear. Therefore, we estimate α_2 and α_3 by fitting a linear function for the response at low dose levels. Note that $\frac{\partial f}{\partial x} = \frac{\alpha_1}{\alpha_3} \exp\left[\frac{-(x+\alpha_2)}{\alpha_3}\right]$. Thus,

$$\text{slope at dose zero} = f'(0) = \frac{\alpha_1}{\alpha_3} \exp\left(\frac{-\alpha_2}{\alpha_3}\right)$$

and

$$\text{photon count at dose zero} = f(0) = \alpha_1 \left[1 - \exp\left(\frac{-\alpha_2}{\alpha_3}\right)\right].$$

This gives $\alpha_3 = \frac{(\alpha_1 - f(0))}{f'(0)}$ and $\alpha_2 = -\alpha_3 \log\left(\frac{\alpha_3 f'(0)}{\alpha_1}\right)$. We estimate $f(0)$, which is the photon count at zero dose, as the average of the observed counts at zero dose. Slope at dose zero is estimated by $\frac{(y[2]-y[1])}{(x[2]-x[1])}$, where $x[1] = 0$, $x[2]$ = smallest positive dose value, $y[1]$ = average photon count at dose zero, and $y[2]$ = average photon count at dose x[2]. Since we already estimated α_1, the parameters α_2 and α_3 can now be estimated using the above equations.

For all the data sets we analyzed, the initial estimates found from this method served as good starting points for the estimation programs we developed. We did not use the graphical method to obtain starting values since the other two methods are easier to use than the graphical method. For some data sets, the quadratic method failed to provide good starting values while the alternative method did not fail.

For some data sets, the parameter α_1 is much larger (roughly 1000 times in some cases) than the other two parameters. In such cases, estimating the parameters instead of y_i we use a scale multiple of y_i which makes the magnitude of α_1 compatible with the other parameters. For example, when α_1 is 1000 times larger in magnitude then we use $y_i/1000$ instead of y_i and instead of α_1 we take $\alpha_1/1000$. After we find the solution we convert the parameters back to the original scale. This was found to improve the convergence rate of the procedures we describe next.

3.2 Maximum likelihood estimates (ML)

We begin by making the following assumptions:

1. The photon count from one aluminum disc does not affect the photon count observed from another disc.

2. The applied dose x can be measured without any error.

3. The observed photon count y is related to the dose received by the sample x according to the model $y = f(x,\theta)(1+\sigma\epsilon_i)$, where $f(x,\theta) = \alpha_1\left\{1-\exp\left[\frac{-(x+\alpha_2)}{\alpha_3}\right]\right\}$ (Section 2.3).

4. The random errors in the photon counts (ϵ's) follow a normal distribution with zero mean and unit variance.

5. The relative error in a single measurement σ $(= \frac{\sqrt{V(Y)}}{E(Y)})$ is constant.

Under the above assumptions the likelihood function for a sample of n observations is

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma f(x_i,\theta)} \exp\left\{\frac{-1}{2\sigma^2 f^2(x_i,\theta)}\left[y_i - f(x_i,\theta)\right]^2\right\}.$$

The log-likelihood $l(\theta,\sigma)$ apart from a constant is

$$-n\,\log(\sigma) - \sum_{i=1}^{n}\log(f(x_i,\theta)) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\left[y_i - f(x_i,\theta)\right]^2}{\left[f(x_i,\theta)\right]^2}. \tag{3.1}$$

Let $\phi^T = (\alpha_1,\alpha_2,\alpha_3,\sigma) = (\theta^T,\sigma)$. The maximum likelihood estimates solve the system of equations $\frac{\partial l}{\partial \phi} = 0$. Equating $\frac{\partial l}{\partial \sigma}$ to zero, we find

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{\left[y_i - f(x_i,\theta)\right]^2}{\left[f(x_i,\theta)\right]^2}. \tag{3.2}$$

The gradient vector of l with respect to θ is

$$\nabla l \;= \frac{\partial l}{\partial \theta} = \sum_{i=1}^{n}\frac{\partial l}{\partial f(x_i,\theta)}\frac{\partial f(x_i,\theta)}{\partial \theta} = \sum_{i=1}^{n}\frac{\partial l}{\partial f_i}\frac{\partial f_i}{\partial \theta}, \tag{3.3}$$

where

$$\frac{\partial l}{\partial f(x_i,\theta)} = -\sum_{i=1}^{n}\frac{1}{f(x_i,\theta)} + \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{(y_i - f(x_i,\theta))}{\left[f(x_i,\theta)\right]^2} + \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{(y_i - f(x_i,\theta))^2}{\left[f(x_i,\theta)\right]^3},$$

and

$$(\nabla f_i)^T = \left(\frac{\partial f(x_i,\theta)}{\partial \theta}\right)^T = \left(\frac{\partial f_i}{\partial \theta}\right)^T = \left(\frac{\partial f_i}{\partial \alpha_1}, \frac{\partial f_i}{\partial \alpha_2}, \frac{\partial f_i}{\partial \alpha_3}\right).$$

The partial derivatives of f_i with respect to the components of θ are

$$\left.\begin{aligned} \tfrac{\partial f_i}{\partial \alpha_1} &= 1 - \exp\left[\tfrac{-(x_i+\alpha_2)}{\alpha_3}\right] \\ \tfrac{\partial f_i}{\partial \alpha_2} &= \tfrac{\alpha_1}{\alpha_3}\exp\left[\tfrac{-(x_i+\alpha_2)}{\alpha_3}\right] \\ \tfrac{\partial f_i}{\partial \alpha_3} &= \tfrac{-\alpha_1}{\alpha_3^2}(x_i+\alpha_2)\exp\left[\tfrac{-(x_i+\alpha_2)}{\alpha_3}\right] \end{aligned}\right\} \tag{3.4}$$

The solution of likelihood equations:

We use a 2-part iteration (Green [33]). This means,

1. Use the starting values for θ to estimate σ.
2. Holding σ fixed, solve the likelihood equations for θ.
3. Use the new parameter estimates for θ to update σ.
4. Repeat the process until desired convergence.

We describe two algorithms that solve the likelihood equations.

1. An iteratively re-weighted least squares algorithm.
2. The Newton Raphson algorithm

An iteratively re-weighted least squares algorithm:

Green [33] describes an iteratively re-weighted least squares algorithm that solves likelihood equations. First we briefly describe the algorithm in general terms. Then we apply the algorithm to solve the likelihood equations in our problem.

Let $l(\theta, \sigma)$ be the log-likelihood function of an $n-$vector η of predictors. The maximum likelihood estimates solve the likelihood equations $\frac{\partial l}{\partial \theta} = 0$. The standard Newton Raphson method uses the iterative scheme

$$\theta_{(k+1)} = \theta_k - H_k^{-1} \bigtriangledown l_k$$

to find the estimates for θ at the $(k+1)$st iteration; here H_k is the Hessian matrix $(\frac{\partial^2 l}{\partial \theta \theta^T})$ and $\bigtriangledown l_k$ is the gradient vector $(\frac{\partial l}{\partial \theta})$, both evaluated at the parameter estimates from the kth iteration.

Let u be the $n-$vector $\frac{\partial l}{\partial \eta}$ and D be the $n \times p$ matrix $\frac{\partial \eta}{\partial \theta}$. Then, the likelihood equations can be written as $\frac{\partial l}{\partial \theta} = D^T u = 0$. Using the introduced matrix notation, the Newton Raphson iterative scheme can be written as

$$\left(\frac{-\partial^2 l}{\partial \theta \theta^T}\right)(\hat{\theta} - \theta) = \frac{\partial l}{\partial \theta} = D^T u. \tag{3.5}$$

Fisher's scoring technique replaces $\frac{\partial^2 l}{\partial\theta\partial\theta^T}$ by its expectation value. Note that

$$\frac{\partial^2 l}{\partial\theta\partial\theta^T} = \sum_{i=1}^{n} \frac{\partial l}{\partial\eta_i}\frac{\partial^2\eta_i}{\partial\theta\partial\theta^T} + \left(\frac{\partial\eta}{\partial\theta}\right)^T \frac{\partial^2 l}{\partial\eta\partial\eta^T}\left(\frac{\partial\eta}{\partial\theta}\right).$$

From the likelihood equations for the normal error model we find $E\left(\frac{\partial l}{\partial\eta_i}\right) = 0$. Let A denote the matrix $E\left(-\frac{\partial^2 l}{\partial\eta\partial\eta^T}\right)$. Then, it is easy to see that $E\left(-\frac{\partial^2 l}{\partial\theta\partial\theta^T}\right) = D^T A D$. Thus, using Fisher's scoring technique the algorithm 3.5 can be written as

$$(D^T A D)(\hat{\theta} - \theta) = D^T u.$$

Writing $(\hat{\theta} - \theta) = \hat{\beta}$ we get

$$(D^T A D)\hat{\beta} = D^T A A^{-1} u. \tag{3.6}$$

(Assume that D is of full rank p, and that A is positive definite throughout the parameter space.)

The above equations have the form of normal equations for a weighted least squares regression. So $\hat{\beta}$ can be found by minimizing the weighted error sum of squares

$$\left(A^{-1}u - D\hat{\beta}\right)^T A \left(A^{-1}u - D\hat{\beta}\right).$$

Next we apply the above algorithm to solve the likelihood equations of our problem. The log-likelihood is

$$l(\theta,\sigma) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \sum_{i=1}^{n}\log(f(x_i,\theta) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{[y_i - f(x_i,\theta)]^2}{[f(x_i,\theta)]^2}.$$

Writing $f(x_i,\theta)$ as f_i we get

$$\frac{\partial l}{\partial f_i} = \frac{-1}{f_i} + \frac{1}{\sigma^2}\frac{[y_i - f_i]}{f_i^2} + \frac{1}{\sigma^2}\frac{[y_i - f_i]^2}{f_i^3}, \tag{3.7}$$

and

$$\frac{\partial^2 l}{\partial f_i^2} = \frac{1}{f_i^2} - \frac{1}{\sigma^2 f_i^2} - \frac{1}{\sigma^2}\frac{4\,[y_i - f_i]}{f_i^3} - \frac{1}{\sigma^2}\frac{3\,[y_i - f_i]^2}{f_i^4}.$$

For our model, $E\,(y_i - f_i)^2 = E\,(f_i^2\sigma^2\epsilon_i^2) = \sigma^2 f_i^2$. Thus,

$$E\left(\frac{\partial^2 l}{\partial f_i^2}\right) = -\frac{2}{f_i^2} - \frac{1}{\sigma^2 f_i^2} \tag{3.8}$$

and

$$E\left(\frac{\partial l}{\partial f_i}\frac{\partial l}{\partial f_j}\right) = 0, \quad \text{for } i \neq j.$$

Thus, in our case the matrix A is diagonal with the ith diagonal element

$$A_{ii} = \frac{2}{f_i^2} + \frac{1}{\sigma^2 f_i^2} \tag{3.9}$$

and D is the $n \times 3$ matrix with the ith row having entries $\frac{\partial f_i}{\partial \alpha_1}$, $\frac{\partial f_i}{\partial \alpha_2}$ and $\frac{\partial f_i}{\partial \alpha_3}$ (Equations 3.4).

Next we outline an algorithm for solving likelihood equations.

1. Find an initial estimate θ_0 for θ.

2. Evaluate f_i at the starting values.

3. Estimate σ from equation 3.2.

4. Evaluate the matrix $D(= \frac{\partial f}{\partial \theta})$.

5. Compute the $n-$vector $u(= \frac{\partial l}{\partial f_i})$ consisting of the elements given by 3.7.

6. Compute the matrix A that has elements given by 3.9.

7. At the $(k+1)$st iteration, $\hat{\beta}_k$ is estimated from $\hat{\beta}_k = (D_k^T A_k D_k)^{-1} D_k^T u_k$. The subscript k indicates that the corresponding terms are evaluated at the parameter estimates from the kth iteration. (Gaussian elimination can be used to solve $(D_k^T A_k D_k)\hat{\beta}_k = D_k^T u_k$ for $\hat{\beta}_k$, thus avoiding matrix inversion.)

8. At the $(k+1)$st iteration estimate θ from $\hat{\theta}_{(k+1)} = \hat{\theta}_k + \hat{\beta}_k$.

9. Repeat the above procedure until desired convergence, each time replacing θ from its current estimate. In the algorithms we developed, the convergence criteria for θ was taken as when the step size $\beta < 1e-5$, or, when all the components of the p-vector $D^T u$ are less than $1e-5$.

Variance covariance matrix of maximum likelihood estimates:

Let $\phi^T = (\theta^T, \sigma)$. From the large sample theory for maximum likelihood estimators, the asymptotic variance covariance matrix of $\hat{\phi}$ is $I^{-1}(\phi)$, where $I(\phi) = E\left(-\frac{\partial^2 l}{\partial \phi \phi^T}\right)$. An estimate for the variance covariance matrix can be obtained by evaluating $I(\phi)$ at $\hat{\phi}$. In particular, the standard errors of the maximum likelihood estimates can be estimated as the square roots of the diagonal elements of $I^{-1}(\hat{\phi})$. Recall that $E\left(-\frac{\partial^2 l}{\partial \theta \partial \theta^T}\right) = D^T A D$. Now consider

$$\frac{\partial^2 l}{\partial \theta \partial \sigma} = -\frac{2}{\sigma^3}\sum_{i=1}^{n}\left(\frac{y_i - f_i}{f_i}\right)\left(\frac{\nabla f}{f}\right)_i - \frac{2}{\sigma^3}\sum_{i=1}^{n}\left(\frac{y_i - f_i}{f_i}\right)^2\left(\frac{\nabla f}{f}\right)_i.$$

It is easy to see that $E\left[-\frac{\partial^2 l}{\partial \theta \partial \sigma}\right] = \frac{2}{\sigma}\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i$. Since $\frac{\partial^2 l}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}\left(\frac{y_i - f_i}{f_i}\right)^2$, we find $E\left(-\frac{\partial^2 l}{\partial \sigma^2}\right) = \frac{2n}{\sigma^2}$. Thus the matrix $I(\phi)$ is given by

$$I(\phi) = \begin{bmatrix} D^T A D & \frac{2}{\sigma}\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i \\ \frac{2}{\sigma}\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i^T & \frac{2n}{\sigma^2} \end{bmatrix}^{-1}.$$

Using standard matrix inversion results (see Rao [51]) we find that the variance covariance matrix of $\hat{\theta}$ is

$$V(\hat{\theta}) = \left(D^T A D\right)^{-1} + E_1 E_2^{-1} E_1^T,$$

where $E_2 = \frac{2n}{\sigma^2} - \frac{4}{\sigma^2}\left[\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i\right]^T \left(D^T A D\right)^{-1}\left[\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i\right]$ (note that E_2 is a scalar) and $F = \frac{2}{\sigma}\left(D^T A D\right)^{-1}\left(\frac{\nabla f_0}{f_0}\right)_i$. Since $\left(D^T A D\right)^{-1}$ is immediately available from the last step of the algorithm we can easily compute the asymptotic variance covariance matrix of $\hat{\theta}$.

The Newton Raphson algorithm:

Before presenting the algorithm we first compute the matrix of second derivatives $H = \frac{\partial^2 l}{\partial \theta \partial \theta^T}$ that is needed in the Newton Raphson algorithm.

Differentiating equation 3.3 with respect to θ, we get

$$\frac{\partial^2 l}{\partial \theta \partial \theta^T} = \sum_{i=1}^{n}\left[\left(\frac{\partial^2 l}{\partial f_i^2}\right)\left(\frac{\partial f_i}{\partial \theta}\right)\left(\frac{\partial f_i}{\partial \theta}\right)^T + \left(\frac{\partial l}{\partial f_i}\right)\left(\frac{\partial^2 f_i}{\partial \theta \partial \theta^T}\right)\right] \tag{3.10}$$

where

$$\frac{\partial^2 l}{\partial f_i^2} = \sum_{i=1}^{n}\frac{1}{f_i^2} - \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{1}{f_i^2} - \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{(y_i - f_i)}{f_i^3} - \frac{3}{\sigma^2}\sum_{i=1}^{n}\frac{(y_i - f_i)^2}{f_i^4}. \tag{3.11}$$

Note that A_i which is the matrix of second derivatives of f_i with respect to θ is a symmetric matrix of order 3. The elements of $A_i = (a_{lm})_{3\times 3}$ are:

$$
\begin{aligned}
a_{11} &= \frac{\partial^2 f_i}{\partial \alpha_1^2} = 0 \\
a_{12} &= \frac{\partial^2 f_i}{\partial \alpha_1 \partial \alpha_2} = \frac{1}{\alpha_3} \exp\left[\frac{-(x_i+\alpha_2)}{\alpha_3}\right] \\
a_{13} &= \frac{\partial^2 f_i}{\partial \alpha_1 \partial \alpha_3} = \frac{-1}{\alpha_3^2}(x_i+\alpha_2)\exp\left[\frac{-(x_i+\alpha_2)}{\alpha_3}\right] \\
a_{22} &= \frac{\partial^2 f_i}{\partial \alpha_2^2} = \frac{-\alpha_1}{\alpha_3^2} \exp\left[\frac{-(x_i+\alpha_2)}{\alpha_3}\right] \\
a_{23} &= \frac{\partial^2 f_i}{\partial \alpha_2 \partial \alpha_3} = \frac{-\alpha_1}{\alpha_3^2} \exp\left[\frac{-(x_i+\alpha_2)}{\alpha_3}\right] + \frac{\alpha_1}{\alpha_3^3}(x_i+\alpha_2)\exp\left[\frac{-(x_i+\alpha_2)}{\alpha_3}\right] \\
a_{33} &= \frac{\partial^2 f_i}{\partial \alpha_3^2} = \frac{2\alpha_1}{\alpha_3^3}(x_i+\alpha_2)\exp\left[\frac{-(x_i+\alpha_2)}{\alpha_3}\right] - \frac{\alpha_1}{\alpha_3^4}(x_i+\alpha_2)^2\exp\left[\frac{-(x_i+\alpha_2)}{\alpha_3}\right].
\end{aligned}
$$

The Newton Raphson algorithm for solving the likelihood equations proceeds as follows.

1. Find an initial estimate for θ.

2. Estimate σ using equation 3.2.

3. Compute $\bigtriangledown l$ using equation 3.3.

4. Compute H using equation 3.10.

5. At the $(k+1)$st iteration, θ is estimated from

$$\hat{\theta}_{(k+1)} = \hat{\theta}_k - H_k^{-1} \bigtriangledown l_k.$$

6. Re-estimate sigma from equation 3.2.

7. Repeat until the desired convergence criteria are met.

We developed software using the statistical package "S-plus" to implement the above algorithm. (We experimented with the built-in procedure "nlmin" of S-plus but found it sometimes failed to converge.) The convergence criteria were determined as follows:

1. The absolute values of all the components of the gradient vector are all less than a small value ϵ_1 (in our program we used $\epsilon_1 = \times 10^{-5}$) and, the eigenvalues of the Hessian matrix at the solution point are all negative or,

2. The absolute difference of the successive iterative points of all the parameters α_1, α_2 and α_3 are less than a small value ϵ_2 (we used $\epsilon_2 = \times 10^{-5}$) or,

3. A maximum number of iterations(e.g. 200) is reached (this was never the case for the data sets we analyzed).

Variance covariance matrix of maximum likelihood estimates:

Let $\phi^T = (\theta^T, \sigma)$. From the large sample theory for maximum likelihood estimators, the asymptotic variance covariance matrix of $\hat{\phi}$ is $I^{-1}(\phi)$, where $I(\phi) = E\left(-\frac{\partial^2 l}{\partial \phi \phi^T}\right)$. An estimate for the variance covariance matrix can be obtained by evaluating $I(\phi)$ at $\hat{\phi}$. In particular, the standard errors of the maximum likelihood estimates can be estimated as the square roots of the diagonal elements of $I^{-1}(\hat{\phi})$.

3.3 Quasi-likelihood Estimates (QL)

One of the assumptions used in deriving maximum likelihood estimates is that the random errors ϵ's follow a standard normal distribution. This distributional assumption is not required for deriving quasi-likelihood estimates. Quasi-likelihood estimates only require the assumptions made on the first two moments of the distribution of the photon counts. Thus, we assume the model $y = f(x, \theta)(1 + \sigma\,\epsilon)$, with $E(y) = f(x, \theta) = \mu$ and $Var(y) = \sigma^2\, f^2(x, \theta) = \sigma^2\, E^2(y) = V(\mu)$.

The quasi-likelihood Q introduced by Wedderburn [59] and later extended to the multivariate case by McCullaugh [48] is defined as any function of μ satisfying

$$\frac{\partial Q}{\partial \mu} = V^{-1}(\mu)(y - \mu),$$

where V^{-1} is a generalized inverse of V.

The quasi-likelihood estimates for θ are obtained as solutions of the quasi-likelihood equations

$$\frac{\partial Q}{\partial \theta} = \left(\frac{\partial \mu}{\partial \theta^T}\right)^T \frac{\partial Q}{\partial \mu} = 0.$$

For our model these can be written as

$$F^T(\theta)V^{-1}(\theta)\,[y - f] = 0, \tag{3.12}$$

where the $n \times p$ matrix F denotes the derivative of the n-vector f with respect to the p-vector θ.

The Gauss Newton algorithm can be used to solve the quasi-likelihood equations. This proceeds as follows.

If the estimates for θ (say θ_0) are close to the true parameter values, the non linear function $f(\theta)$ can be approximated by the Taylor expansion

$$f(x, \theta) \approx f(x, \theta_0) + (\theta - \theta_0)^T F_0.$$

The derivative vector $F(\theta)$ and the variance covariance matrix $V(\theta)$ can be approximated by F_0 and V_0 which are the derivative vector and the variance covariance matrix evaluated at θ_0.

The set of equations 3.12 can therefore be written as

$$F_0^T V_0^{-1} \left[y - f(\theta_0) - (\theta - \theta_0)^T F_0\right] \approx 0.$$

Let $z_0 = [y - f(\theta_0)]$ and $\delta = (\theta - \theta_0)$. Then, the above system of equations gives

$$F_0^T V_0^{-1} [z_0 - F_0 \delta] \approx 0,$$

which are the same set of equations as

$$F_0^T V_0^{-1} F_0 \delta = F_0^T V_0^{-1} z_0.$$

The above equations have the form of the normal equations for a weighted least squares regression with the dependent variable z, design matrix F, and parameter δ.

Thus, the estimates for δ can be found by minimizing the weighted error sum of squares $R(\delta) = \sum_{i=1}^{n} w_{ik}[z_i - F_i^T\delta]^2$, where

$$
\begin{aligned}
x_i &= \text{dose received by the } i\text{th subsample,} \\
\theta_k &= \text{the value of the parameter vector } (\alpha_1, \alpha_2, \alpha_3) \text{ at the } k\text{th iteration,} \\
f(x_i, \theta_k) &= \alpha_1 \left\{1 - \exp\left(-(x_i + \alpha_2)/\alpha_3\right)\right\}, \\
w_{ik} &= f(x_i, \theta_k)^{-2} \text{ is the weight on the } i\text{th observation at the } k\text{th iteration,} \\
z_i &= (y_i - f(x_i, \theta_0)), \\
\text{and } F_i &= \text{gradient vector of } f(x_i, \theta) \text{ evaluated at the } k\text{th step of the iteration.}
\end{aligned}
$$

We use an iteratively re-weighted least squares algorithm to solve the above equations. The algorithm proceeds as follows.

1. Calculate the weights w_{i0} using the initial parameter estimates. Let W_0 be the diagonal matrix of order n that has w_{i0}'s as diagonal elements.

2. New estimates for δ are given by $\delta_0 = (F_0^T W_0 F_0)^{-1}(F_0^T W_0 z_0)$; here z_0 is the n-vector with entries $y_i - f(x_i, \theta_0)$, $i = 1, \ldots, n$.

3. New estimates for θ are given by $\theta_1 = \theta_0 + \delta_0$.

4. Recalculate the weight matrix W and the $n \times p$ matrix F using the new estimates for θ and use it to compute δ.

5. At the $(r+1)$st step, the estimates for θ are given by $\theta_{(r+1)} = \theta_r + \delta_r$, where $\delta_r = (F_r^T W_r F_r)^{-1}(F_r^T W_r z_r)$; here z_r is the n-vector with entries $y_i - f(x_i, \theta_r), i = 1, \ldots, n$.

6. Repeat the above procedure until the desired convergence. In the algorithms we developed the convergence criteria for θ was taken as when the step size $\delta < 1e-5$.

Remarks:

At the kth step of the iteration, the quasi-likelihood estimate $\hat{\theta}_k$ minimizes

$$S(\theta_k) = \sum_{i=1}^{n} \frac{\left[y_i - f(x_i, \hat{\theta}_k)\right]^2}{f(x_i, \hat{\theta}_{(k-1)})^2}.$$

So, the estimates $\hat{\theta}_k$ solve the system of equations

$$\frac{\partial S}{\partial \theta_k} = \sum_{i=1}^{n} \frac{\left[y_i - f(x_i, \hat{\theta}_k)\right]}{f(x_i, \hat{\theta}_{(k-1)})^2} \bigtriangledown f_{\hat{\theta}_k} = 0,$$

where $\bigtriangledown f_{\hat{\theta}}$ denotes the gradient vector evaluated at $\hat{\theta}$. In the limit, the estimates $\hat{\theta}$ solve the system of equations

$$\frac{\partial S}{\partial \theta} = \sum_{i=1}^{n} \frac{\left[y_i - f(x_i, \hat{\theta})\right]}{f(x_i, \hat{\theta})^2} \bigtriangledown f_{\hat{\theta}} = 0.$$

Estimating σ

We note that the quasi-likelihood estimating equations do not involve σ. So, we can estimate θ without knowledge of σ. We have two choices for estimating σ:

1. For a model with mean function $E(y_i) = \mu_i$ and variance function $V(y_i) = \sigma_i$, Davidian and Carroll [23] suggests estimating σ_i from the estimating equation

$$\sum_{i=1}^{n} \left[\frac{(y_i - \mu_i)^2 - \sigma_i^2}{\sigma_i^2}\right] \frac{\mu_i^2}{\sigma_i^2} = 0. \tag{3.13}$$

   For our problem this gives

$$\sum_{i=1}^{n} \left[\frac{(y_i - f_i)^2 - \sigma^2 f_i^2}{\sigma^2 f_i^2}\right] \frac{f_i^2}{\sigma^2 f_i^2} = 0.$$

   So, we can estimate σ^2 as $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \frac{(y_i - \hat{f}_i)^2}{\hat{f}_i^2}$; here $\hat{f}$ indicates the fitted values, which are $f(x_i, \theta)$ evaluated at $\hat{\theta}$. For our problem this estimate is identical to the maximum likelihood estimate for σ.

2. The parameter σ can be estimated by equating the Pearson χ^2 to its degrees of freedom. This gives

$$\sum_{i=1}^{n} \left[\frac{(y_i - f_i)^2 - \sigma^2 f_i^2}{\sigma^2 f_i^2}\right] - (n - p) = 0.$$

   This suggests the estimate

$$\hat{\sigma}^2 = \frac{1}{(n-p)} \sum_{i=1}^{n} \frac{(y_i - \hat{f}_i)^2}{\hat{f}_i^2}, \tag{3.14}$$

where p is the number of components of θ. We use the latter estimate (Equation 3.14) when we estimate the error of the quasi-likelihood estimate. We refer to this estimate as the unbiased estimate of σ; a small σ justification of the jargon is in Chapter 5.

Variance Covariance matrix of quasi-likelihood estimates:

The asymptotic variance covariance matrix of the quasi-likelihood estimates is $V(\theta) = \sigma^2 \left(F^T W F\right)^{-1}$ for the matrices F and W defined earlier. An estimate for the variance covariance matrix of $\hat{\theta}$ can be obtained by replacing the unknown parameters by their estimates. In particular, the diagonal elements of $V(\hat{\theta})$ provide estimates for the errors of the quasi-likelihood estimates.

3.4 Generalized least squares estimates (GLS)

This is a generalization of the linear generalized least squares procedure. The linear generalized least squares estimate $\hat{\theta}$ is found by minimizing the weighted error sum of squares

$$S(\theta) = [y - f(\theta)]^T V^{-1}(\theta) [y - f(\theta)],$$

where $f(\theta)$ is linear in θ and $V(\theta)$ is the variance covariance matrix of Y. The value of θ minimizing S solves

$$\begin{aligned} 0 &= \frac{\partial S}{\partial \hat{\theta}_k} \\ &= \frac{\partial S}{\partial \theta} = \sum_{i=1}^{n} \frac{\left[y_i - f(x_i, \hat{\theta})\right]}{f(x_i, \theta)^2} \bigtriangledown f_{\hat{\theta}} + \sum_{i=1}^{n} \frac{\left[y_i - f(x_i, \hat{\theta})\right]^2}{f(x_i, \theta)^3} \bigtriangledown f_{\hat{\theta}} = 0. \end{aligned}$$

We used the IMSL subroutine NEQNF to solve these equations starting from an initial estimate $\hat{\theta}$ from quasi-likelihood estimate. This system of equations is different from quasi-likelihood estimating equations. As in quasi-likelihood, the estimating equations for θ does not involve σ. Consequently, we can estimate θ without knowledge of σ. Once we have found the generalized least squares estimates for θ, we can estimate σ from $\hat{\sigma}^2 = \frac{1}{(n-p)} \sum_{i=1}^{n} \frac{(y_i - \hat{f}_i)^2}{\hat{f}_i^2}$.

We will later (Section 4.1) see that GLS estimates are not consistent. Consequently, variance estimates are not relevant. However, asymptotic theory justifying the procedure and producing approximate standard errors is available for a small σ approximation; see Section 3.9.

3.5 Data weighted least squares (DWLS)

We now describe another method cited in Berger and Huntley ([12]). This is similar to the generalized least squares described in the previous section except for the weights used in the minimization. In generalized least squares $[E(y)]^{-2} = [f(x_i, \theta)]^{-2}$ are used as weights. In this method, y_i^{-2} are used as weights. Consequently, the estimates $\hat{\theta}$ minimize the weighted error sum of squares

$$S(\theta) = \sum_{i=1}^{n} \left\{ \frac{[y_i - f(x_i, \theta)]^2}{y_i^2} \right\}.$$

We refer to this method as "Data Weighted Least Squares" and use the abbreviation DWLS. The data weighted least squares estimates $\hat{\theta}$ solve the system of equations

$$\frac{\partial S}{\partial \theta} = \sum_{i=1}^{n} \frac{\left[y_i - f(x_i, \hat{\theta})\right]}{y_i^2} \bigtriangledown f_{\hat{\theta}} = 0,$$

where $\bigtriangledown f_{\hat{\theta}}$ denotes the gradient vector evaluated at $\hat{\theta}$.

The algorithm described for solving quasi-likelihood equations can be used to solve the above equations by modifying the weight matrix accordingly.

Similar to quasi-likelihood estimating equations, the equations that solve for DWLS estimates do not involve σ. Once we have found the DWLS estimates, σ can be estimated as $\hat{\sigma}^2 = \frac{1}{(n-p)} \sum_{i=1}^{n} \frac{(y_i - \hat{f}_i)^2}{\hat{f}_i^2}$.

As for GLS, data weighted least squares estimates are not in general consistent; see Section 4.1. Again, small σ theory is available; see Section 3.9.

3.6 Maximum likelihood estimates for the gamma model

Suppose the photon count y_i has a gamma distribution with mean $\mu = f(x_i, \theta)$ and variance $V(y) = \sigma^2 [f(x_i, \theta)]^2$. Assuming that photon counts from different discs are independent, the likelihood for the sample $y_1, \ldots, y_n$ can be written as

$$L'(\theta, \sigma^2) = \exp \left\{ \frac{\left[\sum_{i=1}^{n} y_i \left(\frac{-1}{f(x_i,\theta)}\right) - \sum_{i=1}^{n} -\log\left(\frac{1}{f(x_i,\theta)}\right)\right]}{\sigma^2} + b(y_1, \ldots, y_n, \sigma^2) \right\}$$

Taking $\eta_i = \frac{-1}{f(x_i,\theta)}$, the log-likelihood can be written as

$$l(\theta, \sigma^2) = \frac{[\sum_{i=1}^n y_i \eta_i + \sum_{i=1}^n \log(-\eta_i)]}{\sigma^2} + b(y_1, \ldots, y_n, \sigma^2).$$

The maximum likelihood estimate $\hat{\theta}$ for θ maximizes the log likelihood function. Since $b(y_1, \ldots, y_n, \sigma^2)$ does not involve θ, this is equivalent to maximizing $L = \sum_{i=1}^n y_i \eta_i + \sum_{i=1}^n \log(-\eta_i)$ with respect to θ. Consequently, the estimates $\hat{\theta}$ solves

$$\frac{\partial L}{\partial \theta_j} = \sum_{i=1}^n \frac{\partial L}{\partial \eta_i} \frac{\partial \eta_i}{\partial \theta_j} = \sum_{i=1}^n \left(y_i - \frac{1}{\eta_i} \right) \left(\frac{\partial \eta_i}{\partial \theta_j} \right) = 0,$$

which is equivalent to solving

$$\sum_{i=1}^n \left[(y_i - f(x_i, \theta)) \left(\frac{1}{(f(x_i, \theta))^2} \frac{\partial f(x_i, \theta)}{\partial \theta_j} \right) \right] = 0; \tag{3.15}$$

here θ_j denotes the jth component of θ. We recall that quasi-likelihood estimates also satisfy the set of equations 3.15. Thus for our problem, the maximum likelihood estimates obtained assuming photon counts follow a gamma distribution are identical to the quasi-likelihood estimates.

3.7 Bias of the least squares estimator for parameters in the standard nonlinear regression model

The standard nonlinear regression model is of the form $y_i = f(x_i, \theta) + \epsilon_i \quad (i = 1, \ldots, n)$, where we assume that the random errors ϵ_i follow a normal distribution with zero mean and constant variance. Several authors have addressed the problem of computing the bias of least squared estimator for θ in this model. (eg. Box [14], Cook [17].). For ease of comparison with our formulae, we present the formula derived by Box [14]:

$$bias = -\frac{\sigma^2}{2} \left[\sum_{i=1}^n (\nabla f)_i (\nabla f)_i^T \right]^{-1} \sum_{i=1}^n (\nabla f)^T \, tr \left\{ \left(\sum_{k=1}^n (\nabla f)_k^T (\nabla f)_k \right)^{-1} H_k \right\},$$

where

$$\begin{aligned} \sigma^2 &= Var(\epsilon_i), \\ f_k &= k\text{th response function}, \\ \nabla f_k &= \text{Gradient vector of the } k\text{th response function}, \end{aligned}$$

and

$$H_k = \text{Matrix of second derivatives of the } k\text{th response function.}$$

3.8 Biases of the estimators

Notation:

Let

$$\begin{aligned}
y_i &= \text{observed photon count from the } i\text{th subsample,} \\
x_i &= \text{dose applied to the } i\text{th subsample,} \\
\theta_0 &= \text{vector of unknown true parameters} = (\alpha_1, \alpha_2, \alpha_3), \\
f(x_i, \theta_0) &= i\text{th response function} = \alpha_1 \left\{ 1 - \exp - \left[\frac{x_i + \alpha_2}{\alpha_3} \right] \right\}, \\
\sigma &= \text{relative error in a single measurement,} \\
\text{and } n &= \text{sample size.}
\end{aligned}$$

For notational convenience we further introduce the following.

- We denote $f(x_i, \hat{\theta})$ and $f(x_i, \theta_0)$ by $f_{\hat{\theta}}$ and f_0 respectively. It is important to note that our notation omits the dependence of f_0 and $f_{\hat{\theta}}$ on i. However, it was decided to use this notation to simplify the work of presenting the lengthy computations that follow.

- We denote the gradient vector and the matrix of second derivatives (i.e. the Hessian matrix) by $\bigtriangledown f$ and H respectively; a subscript is used to indicate the parameter point at which each quantity is evaluated. For example, $(\bigtriangledown f_0)$ indicates the gradient vector evaluated at the true parameter value θ_0 whereas $\bigtriangledown f_{\hat{\theta}}$ indicates the gradient vector evaluated at $\hat{\theta}$. Again note that the quantities $\bigtriangledown f_{\hat{\theta}}$ and H_0 depend on i which we sometimes omit for notational convenience.

Assumptions:

1. The relative error in a single measurement σ is small, so that the terms of order $O(\sigma^3)$ and higher are negligible.

2. For small σ, we approximate $\hat{\theta}$ by $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2$, where C_1 and C_2 depend on $f(x, \theta_0)$ and ϵ but not on σ. In Section 3.10, we examine this approximation by a Monte Carlo study.

First we present some calculations that are common to all four methods of estimation. For $\hat{\theta}$ close to θ_0, the Taylor approximation of $f(x_i, \hat{\theta})$ around θ_0 can be written as

$$\begin{aligned} f_{\hat{\theta}} &\approx f_0 + (\hat{\theta} - \theta_0)^T (\bigtriangledown f_0) + \frac{1}{2}(\hat{\theta} - \theta_0)^T H_0 (\hat{\theta} - \theta_0) \\ &\approx f_0 + C_1^T (\bigtriangledown f_0)\,\sigma + (C_2^T (\bigtriangledown f_0) + \frac{1}{2} C_1^T H_0 C_1)\sigma^2. \end{aligned} \tag{3.16}$$

Therefore, we find

$$\begin{aligned} \frac{1}{f_{\hat{\theta}}} &\approx \frac{1}{f_0 \left[1 + C_1^T \left(\frac{\bigtriangledown f_0}{f_0}\right)\sigma\right]} \\ &\approx \frac{1}{f_0}\left[1 - C_1^T \left(\frac{\bigtriangledown f_0}{f_0}\right)\sigma + C_1^T \left(\frac{\bigtriangledown f_0}{f_0}\right) C_1^T \left(\frac{\bigtriangledown f_0}{f_0}\right)\sigma^2\right], \end{aligned} \tag{3.17}$$

$$\begin{aligned} \frac{1}{f_{\hat{\theta}}^2} &\approx \frac{1}{f_0^2 \left[1 + C_1^T \left(\frac{\bigtriangledown f_0}{f_0}\right)\sigma\right]^2} \\ &\approx \frac{1}{f_0^2}\left[1 - C_1^T \left(\frac{\bigtriangledown f_0}{f_0}\right)\sigma + C_1^T \left(\frac{\bigtriangledown f_0}{f_0}\right) C_1^T \left(\frac{\bigtriangledown f_0}{f_0}\right)\sigma^2\right]^2 \\ &\approx \frac{1}{f_0^2}\left[1 - 2C_1^T \left(\frac{\bigtriangledown f_0}{f_0}\right)\sigma\right], \end{aligned} \tag{3.18}$$

and

$$\frac{1}{f_{\hat{\theta}}^3} \approx \frac{1}{f_0^3}\left[1 - 3C_1^T \left(\frac{\bigtriangledown f_0}{f_0}\right)\sigma\right]. \tag{3.19}$$

Since $y_i = f_0 + f_0\sigma\epsilon_i$, we find

$$(y_i - f_{\hat{\theta}}) \approx \left[f_0\epsilon_i - C_1^T (\bigtriangledown f_0)\right]\sigma - \left[C_2^T (\bigtriangledown f_0) + \frac{1}{2} C_1^T H_0 C_1\right]\sigma^2 \tag{3.20}$$

and

$$\begin{aligned} (y_i - f_{\hat{\theta}})^2 &\approx (f_0\epsilon_i - C_1^T (\bigtriangledown f_0))^2\sigma^2 \\ &= (f_0^2\epsilon_i^2 - 2f_0\epsilon_i C_1^T (\bigtriangledown f_0) + C_1^T (\bigtriangledown f_0)\, C_1^T (\bigtriangledown f_0))\sigma^2. \end{aligned} \tag{3.21}$$

Further note that

$$\begin{aligned} \bigtriangledown f_{\hat{\theta}} &= (\bigtriangledown f_0) + H_0(\hat{\theta} - \theta_0) \\ &= (\bigtriangledown f_0) + H_0 C_1\sigma + H_0 C_2\sigma^2 \end{aligned} \tag{3.22}$$

3.8.1 Maximum likelihood

The maximum likelihood estimates $\hat{\theta}$ solve the system of equations

$$\begin{aligned} 0 &= \left.\frac{\partial l}{\partial \theta}\right|_{\hat{\theta}} \\ \Rightarrow 0 &= \hat{\sigma}^2 \sum_{i=1}^{n} \frac{\nabla \hat{f}}{\hat{f}} - \sum_{i=1}^{n} \frac{[y_i - \hat{f}]}{\hat{f}^2} \nabla \hat{f} - \sum_{i=1}^{n} \frac{[y_i - \hat{f}]^2}{\hat{f}^3} \nabla \hat{f} \\ \Rightarrow 0 &= \frac{1}{n}\left[\sum_{i=1}^{n} \frac{\left[y_i - \hat{f}\right]^2}{\hat{f}^2}\right] \sum_{i=1}^{n} \frac{\nabla \hat{f}}{\hat{f}} - \sum_{i=1}^{n} \frac{[y_i - \hat{f}]}{\hat{f}^2} \nabla \hat{f} - \sum_{i=1}^{n} \frac{[y_i - \hat{f}]^2}{\hat{f}^3} \nabla \hat{f} \end{aligned}$$

Thus,

$$\begin{aligned} 0 &= \frac{1}{n}\left(\sum_{i=1}^{n}\left[f_0^2\epsilon_i^2 - 2f_0\epsilon_i C_1^T(\nabla f_0) + C_1^T(\nabla f_0)C_1^T(\nabla f_0)\right]\sigma^2\frac{1}{f_0^2}\left[1 - 2C_1^T\left(\frac{\nabla f_0}{f_0}\right)\sigma\right]\right) \\ &\times \sum_{i=1}^{n}\left\{\left(\nabla f_0 + H_0C_1\sigma + H_0C_2\sigma^2\right)\frac{1}{f_0}\left[1 - C_1^T\left(\frac{\nabla f_0}{f_0}\right)\sigma + C_1^T\left(\frac{\nabla f_0}{f_0}\right)C_1^T\left(\frac{\nabla f_0}{f_0}\right)\sigma^2\right]\right\} \\ &- \sum_{i=1}^{n}\left[(f_0\epsilon_i - C_1^T(\nabla f_0))\sigma - (C_2^T(\nabla f_0) + \frac{1}{2}C_1^T H_0 C_1)\sigma^2\right]\frac{1}{f_0^2}\left[1 - 2C_1^T\left(\frac{\nabla f_0}{f_0}\right)\sigma\right] \\ &\times \left((\nabla f_0) + H_0C_1\sigma + H_0C_2\sigma^2\right) \\ &- \sum_{i=1}^{n}\left\{[f_0^2\epsilon_i^2 - 2f_0\epsilon_i C_1^T(\nabla f_0) + C_1^T(\nabla f_0)C_1^T(\nabla f_0)]\sigma^2\frac{1}{f_0^3}\left\{1 - 3C_1^T\left(\frac{\nabla f_0}{f_0}\right)\sigma\right\}\right\} \\ &\times((\nabla f_0) + H_0C_1\sigma + H_0C_2\sigma^2). \end{aligned}$$

Equating the coefficients of powers of σ;

$$\begin{aligned} \sigma^1; \quad 0 &= \sum_{i=1}^{n}(f_0\epsilon_i - C_1^T(\nabla f_0))\left(\frac{\nabla f_0}{f_0}\right) \\ \Rightarrow \quad C_1 &= \left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T\right]^{-1}\left[\sum_{i=1}^{n}\epsilon_i\left(\frac{\nabla f_0}{f_0}\right)\right] \qquad (3.23) \\ \sigma^2; \quad 0 &= \frac{1}{n}\left(\sum_{i=1}^{n}\left[\epsilon_i^2 - 2\epsilon_i C_1^T\left(\frac{\nabla f_0}{f_0}\right) + C_1^T\left(\frac{\nabla f_0}{f_0}\right)C_1^T\left(\frac{\nabla f_0}{f_0}\right)\right]\right)\left(\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\right) \\ &- \sum_{i=1}^{n}\left\{\left(f_0\epsilon_i - C_1^T(\nabla f_0)\right)\frac{1}{f_0^2}(H_0C_1 - 2C_1^T\left(\frac{\nabla f_0}{f_0}\right)(\nabla f_0))\right\} \qquad (3.24) \\ &+ \sum_{i=1}^{n}\left\{(C_2^T(\nabla f_0) + \frac{1}{2}C_1^T H_0 C_1)\left(\frac{\nabla f_0}{f_0}\right)\right\} \\ &- \sum_{i=1}^{n}\left\{\left(f_0^2\epsilon_i^2 - 2f_0\epsilon_i C_1^T(\nabla f_0) + C_1^T(\nabla f_0)C_1^T(\nabla f_0)\right)\left(\frac{(\nabla f_0)}{f_0^3}\right)\right\} \end{aligned}$$

$$\Rightarrow \qquad C_2 \;=\; \left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T\right]^{-1} A, \tag{3.25}$$

where

$$\begin{aligned} A \;=\; & -\frac{1}{n}\left(\sum_{i=1}^{n}\left[\epsilon_i^2 - 2\epsilon_i C_1^T\left(\frac{\nabla f_0}{f_0}\right) + C_1^T\left(\frac{\nabla f_0}{f_0}\right)C_1^T\left(\frac{\nabla f_0}{f_0}\right)\right]\right)\left(\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\right) \\ & +\sum_{i=1}^{n}\left(\frac{H_0}{f_0}\right)C_1\epsilon_i - 4\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T C_1\epsilon_i \\ & -\sum_{i=1}^{n} C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{H_0}{f_0}\right)C_1 + 3\sum_{i=1}^{n} C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T C_1\left(\frac{\nabla f_0}{f_0}\right) \\ & -\frac{1}{2}\sum_{i=1}^{n} C_1^T\left(\frac{H_0}{f_0}\right)C_1\left(\frac{\nabla f_0}{f_0}\right) + \sum_{i=1}^{n}\epsilon_i^2\left(\frac{\nabla f_0}{f_0}\right). \end{aligned} \tag{3.26}$$

Since we assume that ϵ_i's have mean zero, $E(C_1) = 0$ (See Equation 3.23.).

Using the results of Appendix 9.3 we find

$$\begin{aligned} E(A) \;=\; & -\left(1-\frac{p}{n}\right)\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right) + \sum_{i=1}^{n}\left(\frac{H_0}{f_0}\right)\Sigma\left(\frac{\nabla f_0}{f_0}\right) \\ & -4\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T\Sigma\left(\frac{\nabla f_0}{f_0}\right) - \sum_{i=1}^{n}\left(\frac{H_0}{f_0}\right)\Sigma\left(\frac{\nabla f_0}{f_0}\right) \\ & +3\sum_{i=1}^{n} tr\left[\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T\Sigma\right]\left(\frac{\nabla f_0}{f_0}\right) \\ & -\frac{1}{2}tr\left[\left(\frac{H_0}{f_0}\right)\Sigma\right]\left(\frac{\nabla f_0}{f_0}\right) + \sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right). \end{aligned}$$

Therefore,

$$E(C_2) = \Sigma\left\{-\sum_{i=1}^{n}(w_{1,i} - \frac{p}{n^2})\left(\frac{\nabla f_0}{f_0}\right) - \frac{1}{2}\sum_{i=1}^{n} w_{2,i}\left(\frac{\nabla f_0}{f_0}\right)\right\},$$

where

$$\begin{aligned} \Sigma \;&=\; \left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)_i\left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1} \;=\; \text{Variance covariance matrix of } C_1 \\ p \;&=\; \text{number of components of } \theta \\ w_{1,i} \;&=\; tr\left[\left(\frac{\nabla f_0}{f_0}\right)_i\left(\frac{\nabla f_0}{f_0}\right)_i^T\Sigma\right] \\ \text{and } w_{2,i} \;&=\; tr\left[\left(\frac{H_0}{f_0}\right)_i\Sigma\right]. \end{aligned}$$

Noting that $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2$, with $E(C_1) = 0$, the bias in $\hat{\theta}$ is computed to order $O(\sigma^3)$ as:

$$\begin{aligned} \text{bias in } \hat{\theta} &= E(\hat{\theta}) - \theta_0 = E(C_2)\sigma^2, & (3.27) \\ &= \Sigma\left\{-\sum_{i=1}^{n}\left(w_{1,i} - \frac{p}{n}\right)\left(\frac{\nabla f_0}{f_0}\right) - \frac{1}{2}\sum_{i=1}^{n} w_{2,i}\left(\frac{\nabla f_0}{f_0}\right)\right\}\sigma^2. & (3.28) \end{aligned}$$

3.8.2 Quasi-likelihood

The quasi-likelihood estimator $\hat{\theta}$ satisfies the equation

$$\begin{aligned} 0 &= \sum_{i=1}^{n} -2\frac{\left[y_i - f(x_i,\hat{\theta})\right]}{f_{\hat{\theta}}^2}\nabla f_{\hat{\theta}} \\ \Rightarrow 0 &= \sum_{i=1}^{n}\left[(f_0\epsilon_i - C_1^T(\nabla f_0)) - (C_2^T(\nabla f_0) + \frac{1}{2}C_1^T H_0 C_1)\sigma\right]\frac{1}{f_0^2}\left[1 - 2C_1^T\left(\frac{\nabla f_0}{f_0}\right)\sigma\right] \\ &\quad \times((\nabla f_0) + H_0C_1\sigma + H_0C_2\sigma^2). \end{aligned}$$

Equating the coefficients of powers of σ;

$$\begin{aligned} \sigma^0; \quad 0 &= \sum_{i=1}^{n}(f_0\epsilon_i - C_1^T(\nabla f_0))\left(\frac{\nabla f_0}{f_0}\right) \\ \Rightarrow \quad C_1 &= \left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T\right]^{-1}\left[\sum_{i=1}^{n}\epsilon_i\left(\frac{\nabla f_0}{f_0}\right)\right] & (3.29) \\ \sigma^1; \quad 0 &= \sum_{i=1}^{n}\left\{(f_0\epsilon_i - C_1^T(\nabla f_0))\frac{1}{f_0^2}\left(H_0C_1 - 2C_1^T\left(\frac{\nabla f_0}{f_0}\right)(\nabla f_0)\right)\right\} \\ &\quad -\sum_{i=1}^{n}\left\{\left(C_2^T(\nabla f_0) + \frac{1}{2}C_1^T H_0 C_1\right)\left(\frac{\nabla f_0}{f_0}\right)\right\} \\ \Rightarrow \quad 0 &= \sum_{i=1}^{n}\epsilon_i\left(\frac{H_0}{f_0}\right)C_1 - 2\sum_{i=1}^{n}\epsilon_i C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right) - \sum_{i=1}^{n}C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{H_0}{f_0}\right)C_1 \\ &\quad +2\sum_{i=1}^{n}C_1^T\left(\frac{\nabla f_0}{f_0}\right)C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right) - \sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T C_2 \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}C_1^T\left(\frac{H_0}{f_0}\right)C_1\left(\frac{\nabla f_0}{f_0}\right) \\ \Rightarrow \quad C_2 &= \left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T\right]^{-1}A, & (3.30) \end{aligned}$$

where

$$A = \sum_{i=1}^{n}\left(\frac{H_0}{f_0}\right)\epsilon_i C_1 - 2\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T C_1\epsilon_i - \sum_{i=1}^{n}C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{H_0}{f_0}\right)C_1$$

$$+2\sum_{i=1}^{n} C_1^T \left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T C_1 \left(\frac{\nabla f_0}{f_0}\right) - \frac{1}{2}\sum_{i=1}^{n} C_1^T \left(\frac{H_0}{f_0}\right) C_1 \left(\frac{\nabla f_0}{f_0}\right)$$

Using the results of Appendix 9.3 we find

$$\begin{aligned} E(A) &= \sum_{i=1}^{n}\left(\frac{H_0}{f_0}\right)\Sigma\left(\frac{\nabla f_0}{f_0}\right) - 2\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T \Sigma\left(\frac{\nabla f_0}{f_0}\right) - \sum_{i=1}^{n}\left(\frac{H_0}{f_0}\right)\Sigma\left(\frac{\nabla f_0}{f_0}\right) \\ &\quad +2\sum_{i=1}^{n} tr\left[\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T \Sigma\right]\left(\frac{\nabla f_0}{f_0}\right) - \frac{1}{2}tr\left[\left(\frac{H_0}{f_0}\right)\Sigma\right]\left(\frac{\nabla f_0}{f_0}\right) \\ &\quad -\frac{1}{2}\sum_{i=1}^{n} w_{2,i}\left(\frac{\nabla f_0}{f_0}\right). \end{aligned}$$

Therefore,

$$E(C_2) = \Sigma\left\{-\frac{1}{2}w_{2,i}\left(\frac{\nabla f_0}{f_0}\right)\right\}. \tag{3.31}$$

As before, $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2$ with $E(C_1) = 0$. Therefore,

$$\text{bias in } \hat{\theta} = \Sigma\left\{-\frac{1}{2}\sum_{i=1}^{n} w_{2,i}\left(\frac{\nabla f_0}{f_0}\right)\right\}\sigma^2.$$

3.8.3 Generalized least squares

The generalized least squares estimator $\hat{\theta}$ minimizes

$$S(\theta) = \sum_{i=1}^{n}\left\{\frac{[y_i - f(x_i,\theta)]^2}{f^2(x_i,\theta)}\right\}.$$

So, $\hat{\theta}$ satisfies the set of equations

$$\begin{aligned} 0 &= \left.\frac{\partial S}{\partial\theta}\right|_{\hat{\theta}} \\ \Rightarrow \quad 0 &= \sum_{i=1}^{n}\left\{\frac{[y_i - f_{\hat{\theta}}]}{f_{\hat{\theta}}^2}\nabla f_{\hat{\theta}}\right\} + \sum_{i=1}^{n}\left\{\frac{[y_i - f_{\hat{\theta}}]^2}{f_{\hat{\theta}}^3}\nabla f_{\hat{\theta}}\right\} \\ \Rightarrow \quad 0 &= \sum_{i=1}^{n}\left[(f_0\epsilon_i - C_1^T(\nabla f_0)) - (C_2^T(\nabla f_0) + \frac{1}{2}C_1^T H_0 C_1)\sigma\right]\frac{1}{f_0^2}\left(1 - 2C_1^T\left(\frac{\nabla f_0}{f_0}\right)\sigma\right) \\ &\quad \times((\nabla f_0) + H_0C_1\sigma + H_0C_2\sigma^2) \\ &\quad +\sum_{i=1}^{n}\left[f_0^2\epsilon_i^2 - 2f_0\epsilon_i C_1^T(\nabla f_0) + C_1^T(\nabla f_0)\,C_1^T(\nabla f_0)\right]\sigma\frac{1}{f_0^3}\left(1 - 3C_1^T\left(\frac{\nabla f_0}{f_0}\right)\sigma\right) \\ &\quad \times((\nabla f_0) + H_0C_1\sigma + H_0C_2\sigma^2) \end{aligned}$$

Equating the coefficients of powers of σ;

$$\begin{aligned}
\sigma^0; \quad 0 &= \sum_{i=1}^{n} \left(f_0\epsilon_i - C_1^T(\nabla f_0)\right)\left(\frac{\nabla f_0}{f_0}\right) \\
\Rightarrow \quad C_1 &= \left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T\right]^{-1}\left[\sum_{i=1}^{n}\epsilon_i\left(\frac{\nabla f_0}{f_0}\right)\right] \qquad (3.32) \\
\sigma^1; \quad 0 &= \sum_{i=1}^{n}\left[\left(f_0\epsilon_i - C_1^T(\nabla f_0)\right)\frac{1}{f_0^2}\left(H_0C_1 - 2C_1^T\left(\frac{\nabla f_0}{f_0}\right)(\nabla f_0)\right)\right] \\
&\quad -\sum_{i=1}^{n}\left[\left(C_2^T(\nabla f_0) + \frac{1}{2}C_1^T H_0 C_1\right)\left(\frac{\nabla f_0}{f_0}\right)\right] \\
&\quad +\sum_{i=1}^{n}\left[\left(f_0^2\epsilon_i^2 - 2f_0\epsilon_i C_1^T(\nabla f_0) + C_1^T(\nabla f_0)\,C_1^T(\nabla f_0)\right)\left(\frac{(\nabla f_0)}{f_0^3}\right)\right] \\
\Rightarrow \quad 0 &= \sum_{i=1}^{n}\epsilon_i\left(\frac{H_0}{f_0}\right)C_1 - 2\sum_{i=1}^{n}\epsilon_i C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right) - \sum_{i=1}^{n}C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{H_0}{f_0}\right)C_1 \\
&\quad +2\sum_{i=1}^{n}C_1^T\left(\frac{\nabla f_0}{f_0}\right)C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right) - \sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T C_2 \\
&\quad -\frac{1}{2}\sum_{i=1}^{n}C_1^T\left(\frac{H_0}{f_0}\right)C_1\left(\frac{\nabla f_0}{f_0}\right) + \sum_{i=1}^{n}\epsilon_i^2\left(\frac{\nabla f_0}{f_0}\right) \\
&\quad -2\sum_{i=1}^{n}\epsilon_i C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right) + \sum_{i=1}^{n}C_1^T\left(\frac{\nabla f_0}{f_0}\right)C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right) \\
\Rightarrow \quad C_2 &= \left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T\right]^{-1} A, \qquad (3.33)
\end{aligned}$$

where

$$\begin{aligned}
A &= \sum_{i=1}^{n}\epsilon_i\left(\frac{H_0}{f_0}\right)C_1 - 4\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T C_1\epsilon_i - \sum_{i=1}^{n}C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{H_0}{f_0}\right)C_1 \\
&\quad +3\sum_{i=1}^{n}C_1^T\left(\frac{\nabla f_0}{f_0}\right)C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right) - \frac{1}{2}\sum_{i=1}^{n}C_1^T\left(\frac{H_0}{f_0}\right)C_1\left(\frac{\nabla f_0}{f_0}\right) \\
&\quad +\sum_{i=1}^{n}\epsilon_i^2\left(\frac{\nabla f_0}{f_0}\right).
\end{aligned}$$

Thus,

$$\begin{aligned}
E(A) &= -4\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T\Sigma\left(\frac{\nabla f_0}{f_0}\right) + 3\sum_{i=1}^{n}w_{1,i}\left(\frac{\nabla f_0}{f_0}\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{n}w_{2,i}\left(\frac{\nabla f_0}{f_0}\right) + \sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right). \qquad (3.34)
\end{aligned}$$

From the result 4 of Appendix 9.3,

$$\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T \Sigma \left(\frac{\nabla f_0}{f_0}\right) = \sum_{i=1}^{n} w_{1,i}\left(\frac{\nabla f_0}{f_0}\right).$$

The expected value of C_2 can therefore be written as

$$E(C_2) = \Sigma\left\{\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right) - \sum_{i=1}^{n} w_{1,i}\left(\frac{\nabla f_0}{f_0}\right) - \frac{1}{2}\sum_{i=1}^{n} w_{2,i}\left(\frac{\nabla f_0}{f_0}\right)\right\}.$$

As before, $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2$ with $E(C_1) = 0$. Therefore,

$$\begin{aligned}\text{bias in } \hat{\theta} &= E(\hat{\theta}) - \theta_0 = E(C_2)\sigma^2 \\ &= \Sigma\left\{\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right) - \sum_{i=1}^{n} w_{1,i}\left(\frac{\nabla f_0}{f_0}\right) - \frac{1}{2}\sum_{i=1}^{n} w_{2,i}\left(\frac{\nabla f_0}{f_0}\right)\right\}\sigma^2. \qquad (3.35)\end{aligned}$$

3.8.4 Data weighted least squares

The data weighted least squares estimator minimizes

$$S(\theta) = \sum_{i=1}^{n}\left\{\frac{[y_i - f(x_i,\theta)]^2}{y_i^2}\right\}.$$

So, $\hat{\theta}$ satisfies the set of equations $\frac{\partial S}{\partial \theta}\Big|_{\hat{\theta}} = 0$, which can be written as

$$\begin{aligned}0 &= \sum_{i=1}^{n}\left\{\frac{[y_i - f(x_i,\hat{\theta})]}{y_i^2}\nabla f_{\hat{\theta}}\right\} \\ \Rightarrow 0 &= \sum_{i=1}^{n}\left[(f_0\epsilon_i - C_1^T(\nabla f_0)) - (C_2^T(\nabla f_0) + \frac{1}{2}C_1^T H_0 C_1)\sigma\right]\frac{1}{f_0^2}(1 - 2\sigma\epsilon_i + 3\sigma^2\epsilon_i^2) \\ &\quad \times((\nabla f_0) + H_0C_1\sigma + H_0C_2\sigma^2).\end{aligned}$$

Equating the coefficients of powers of σ;

$$\begin{aligned}\sigma^0; \quad 0 &= \sum_{i=1}^{n}(f_0\epsilon_i - C_1^T(\nabla f_0))\left(\frac{\nabla f_0}{f_0}\right) \\ \Rightarrow \quad C_1 &= \left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T\right]^{-1}\left[\sum_{i=1}^{n}\epsilon_i\left(\frac{\nabla f_0}{f_0}\right)\right] \qquad (3.36) \\ \sigma^1; \quad 0 &= \sum_{i=1}^{n}(f_0\epsilon_i - C_1^T(\nabla f_0)_i)\frac{1}{f_0^2}(H_0C_1 - 2\epsilon_i(\nabla f_0)_i) \\ &\quad - \sum_{i=1}^{n}(C_2^T(\nabla f_0)_i + \frac{1}{2}C_1^T H_0 C_1)\frac{1}{f_0^2}(\nabla f_0)_i\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \epsilon_i \left(\frac{H_0}{f_0}\right) C_1 - 2\sum_{i=1}^{n} \epsilon_i^2 \left(\frac{\nabla f_0}{f_0}\right) - \sum_{i=1}^{n} C_1^T \left(\frac{\nabla f_0}{f_0}\right)\left(\frac{H_0}{f_0}\right) C_1 \\
&\quad +2\sum_{i=1}^{n} \epsilon_i C_1^T \left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right) - \left[\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T\right] C_2 \\
&\quad -\frac{1}{2}\sum_{i=1}^{n} C_1^T \left(\frac{H_0}{f_0}\right) C_1 \left(\frac{\nabla f_0}{f_0}\right) \qquad (3.37)
\end{aligned}
$$

$$
\Rightarrow \quad C_2 = \left[\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1} A, \qquad (3.38)
$$

where

$$
\begin{aligned}
A &= \sum_{i=1}^{n} \epsilon_i \left(\frac{H_0}{f_0}\right) C_1 - 2\sum_{i=1}^{n} \epsilon_i^2 \left(\frac{\nabla f_0}{f_0}\right) - \sum_{i=1}^{n} C_1^T \left(\frac{\nabla f_0}{f_0}\right)\left(\frac{H_0}{f_0}\right) C_1 \\
&\quad +2\sum_{i=1}^{n} \epsilon_i C_1^T \left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right) - \frac{1}{2}\sum_{i=1}^{n} C_1^T \left(\frac{H_0}{f_0}\right) C_1 \left(\frac{\nabla f_0}{f_0}\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
E(A) &= \sum_{i=1}^{n} \left(\frac{H_0}{f_0}\right) \Sigma \left(\frac{\nabla f_0}{f_0}\right) - 2\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right) - \sum_{i=1}^{n} \left(\frac{H_0}{f_0}\right) \Sigma \left(\frac{\nabla f_0}{f_0}\right) \qquad (3.39) \\
&\quad +2\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T \Sigma \left(\frac{\nabla f_0}{f_0}\right) - \frac{1}{2}\sum_{i=1}^{n} tr\left[\left(\frac{H_0}{f_0}\right)\Sigma\right]\left(\frac{H_0}{f_0}\right).
\end{aligned}
$$

From result 4 of appendix 9.3,

$$
\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T \Sigma \left(\frac{\nabla f_0}{f_0}\right) = \sum_{i=1}^{n} w_{1,i} \left(\frac{\nabla f_0}{f_0}\right).
$$

The expected value of C_2 therefore simplifies to give

$$
E(C_2) = \Sigma \left\{-2\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right) + 2\sum_{i=1}^{n} w_{1,i} \left(\frac{\nabla f_0}{f_0}\right) - \frac{1}{2}\sum_{i=1}^{n} w_{2,i} \left(\frac{\nabla f_0}{f_0}\right)\right\}.
$$

As before,

$$
\text{bias in } \hat{\theta} = \Sigma \left\{-2\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right) + 2\sum_{i=1}^{n} w_{1,i} \left(\frac{\nabla f_0}{f_0}\right) - \frac{1}{2}\sum_{i=1}^{n} w_{2,i} \left(\frac{\nabla f_0}{f_0}\right)\right\} \sigma^2.
$$

3.9 Mean squared errors of the estimators

We begin by proving the following results.

Result 1 *If the random errors ϵ_i's are independent and follow a distribution that is symmetric about zero, then $E(\epsilon_i M C_1 C_1^T) = 0$ for any matrix M with non stochastic elements; the matrix M is assumed to be of the correct order to allow the matrix multiplication.*

Proof: Recall that $C_1 = \Sigma \sum_{i=1}^{n} \epsilon_i \left(\frac{\nabla f_0}{f_0}\right)$. Since Σ is a symmetric matrix,

$$\begin{aligned} E\left(\epsilon_i M C_1 C_1^T\right) &= M\Sigma E\left(\epsilon_i \sum_{j=1}^{n}\sum_{k=1}^{n} \epsilon_j \epsilon_k \left(\frac{\nabla f_0}{f_0}\right)_j \left(\frac{\nabla f_0}{f_0}\right)_k^T\right) \Sigma \\ &= M\Sigma \left(\sum_{j=1}^{n}\sum_{k=1}^{n} E(\epsilon_i \epsilon_j \epsilon_k) \left(\frac{\nabla f_0}{f_0}\right)_j \left(\frac{\nabla f_0}{f_0}\right)_k^T\right) \Sigma. \end{aligned}$$

Since ϵ_i's are independent and $E(\epsilon_i) = E(\epsilon_i^3) = 0$, we find $E\left(\epsilon_i M C_1 C_1^T\right) = 0$.

Result 2 *If the random errors ϵ_i's are independent and follow a distribution that is symmetric about zero, then $E(C_1^T M C_1 C_1^T) = 0$ for any matrix M with non stochastic elements; the matrix M is assumed to be of correct order to allow the matrix multiplication.*

Proof: The proof follows as in the previous result.

Result 3 *If the random errors ϵ_i's are independent and follow a distribution that is symmetric about zero, then for all four estimation methods, vectors C_1 and C_2 are uncorrelated.*

Proof:

We recall that, with an error of approximation $O(\sigma^3)$, each estimator satisfies the approximation $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2$, where

$$C_1 = \left[\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1} \left[\sum_{i=1}^{n} \epsilon_i \left(\frac{\nabla f_0}{f_0}\right)_i\right] = \Sigma \left[\sum_{i=1}^{n} \epsilon_i \left(\frac{\nabla f_0}{f_0}\right)_i\right]$$

and

$$C_2 = \left[\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1} A = \Sigma A,$$

for the matrix A defined in each case. Since $E(C_1) = 0$, we have $Cov(C_1, C_2) = E(C_2 C_1^T)$. Now on substitution for A, it follows that $Cov(C_1, C_2) = 0$.

Mean squared error

For small σ we showed that $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2 + O(\sigma^3)$, where

$$\begin{aligned} C_1 &= \left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)_i\left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1}\left[\sum_{i=1}^{n}\epsilon_i\left(\frac{\nabla f_0}{f_0}\right)_i\right], \\ C_2 &= \left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)_i\left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1} A, \\ \text{and } Cov(C_1, C_2) &= 0. \end{aligned}$$

Since ϵ_i's are independent and $V(\epsilon_i) = 1$, we find $Var(C_1) = \left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)_i\left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1}$. Thus, $Var(\hat{\theta}) = \Sigma\sigma^2 + O(\sigma^4) = \left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)_i\left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1}\sigma^2 + O(\sigma^4)$.

The mean squared error is given by $MSE(\hat{\theta}) = Var(\hat{\theta}) + (bias)^2$. The biases and the variances of the estimators are summarized in the Table 3.1; the resulting mean squared errors are correct to the order $O(\sigma^4)$. In Table 3.1,

Method of Estimation	Bias	$Var(\hat{\theta})$
ML	$\Sigma\left\{-\sum_{i=1}^{n}(w_{1,i} - \frac{p}{n})\left(\frac{\nabla f_0}{f_0}\right)_i - \frac{1}{2}\sum_{i=1}^{n} w_{2,i}\left(\frac{\nabla f_0}{f_0}\right)_i\right\}\sigma^2$	$\Sigma\sigma^2$
QL	$\Sigma\left\{-\frac{1}{2}\sum_{i=1}^{n} w_{2,i}\left(\frac{\nabla f_0}{f_0}\right)_i\right\}\sigma^2$	$\Sigma\sigma^2$
GLS	$\Sigma\left\{\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)_i - \sum_{i=1}^{n} w_{1,i}\left(\frac{\nabla f_0}{f_0}\right)_i - \frac{1}{2}\sum_{i=1}^{n} w_{2,i}\left(\frac{\nabla f_0}{f_0}\right)_i\right\}\sigma^2$	$\Sigma\sigma^2$
DWLS	$\Sigma\left\{-2\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)_i + 2\sum_{i=1}^{n} w_{1,i}\left(\frac{\nabla f_0}{f_0}\right)_i - \frac{1}{2}\sum_{i=1}^{n} w_{2,i}\left(\frac{\nabla f_0}{f_0}\right)_i\right\}\sigma^2$	$\Sigma\sigma^2$

Table 3.1: The biases and the variances of the estimators for single curve fitting

$$\begin{aligned} \Sigma &= \left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)_i\left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1} \\ p &= \text{Number of components of } \theta \\ w_{1,i} &= tr\left[\left(\frac{\nabla f_0}{f_0}\right)_i\left(\frac{\nabla f_0}{f_0}\right)_i^T\Sigma\right] \\ \text{and } w_{2,i} &= tr\left[\left(\frac{H_0}{f_0}\right)_i\Sigma\right]. \end{aligned}$$

3.10 Examination of the assumptions used in estimating the bias

Now we describe a Monte Carlo study that examines the approximation $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2$, in our setting. If this assumption holds, then a plot of each of the components of $\frac{(\hat{\theta}-\theta_0)}{\sigma}$ vs σ should be a straight line with the intercept and slope the corresponding components of C_1 and C_2.

1. We chose σ to be the sequence of values that divide the interval $[0.0001, 0.3]$ into 20 equal parts.

2. We chose the dose levels to be the same as those used for the data set QNL84-2 of Berger and Huntley [12]. This dose values are presented in Table 9.1 of Appendix 9.3 where they are labeled as $P1U$. Let $x_1, \ldots, x_n$ be the chosen dose values; for this data $n = 16$.

3. We set the parameters at $\alpha_1 = 142800.7, \alpha_2 = 122.737$ and $\alpha_3 = 391.9965$. These are the quasi-likelihood estimates obtained for this data set. (The complete data set can be found in Berger and Huntley [12].)

4. We generated a single set of n standard normal random variates[1]. Let this sample be $\epsilon_1, \ldots, \epsilon_n$.

5. For each σ, a sample of y values were then generated according to the relation $y_i = f(x_i, \theta)(1 + \sigma\epsilon_i)$ for $f(x_i, \theta) = \alpha_1 \{1 - \exp[-(x_i + \alpha_2)/\alpha_3]\}$.

6. The procedures described in Sections 3.2 - 3.5 were then followed to estimate the parameters in the model

$$y_i = f(x_i, \theta)(1 + \sigma\epsilon_i) \quad i = 1, \ldots, n.$$

Thus, we have an estimate for θ corresponding to each value of σ.

Figures 3.1 - 3.6 illustrate the plot of the kth component of $(\hat{\theta} - \theta_0)/\sigma$ vs σ, $(k = 1, 2, 3)$.

[1] We used the command "rnorm" in the statistical package "S-plus" to generate the standard normal random variates.

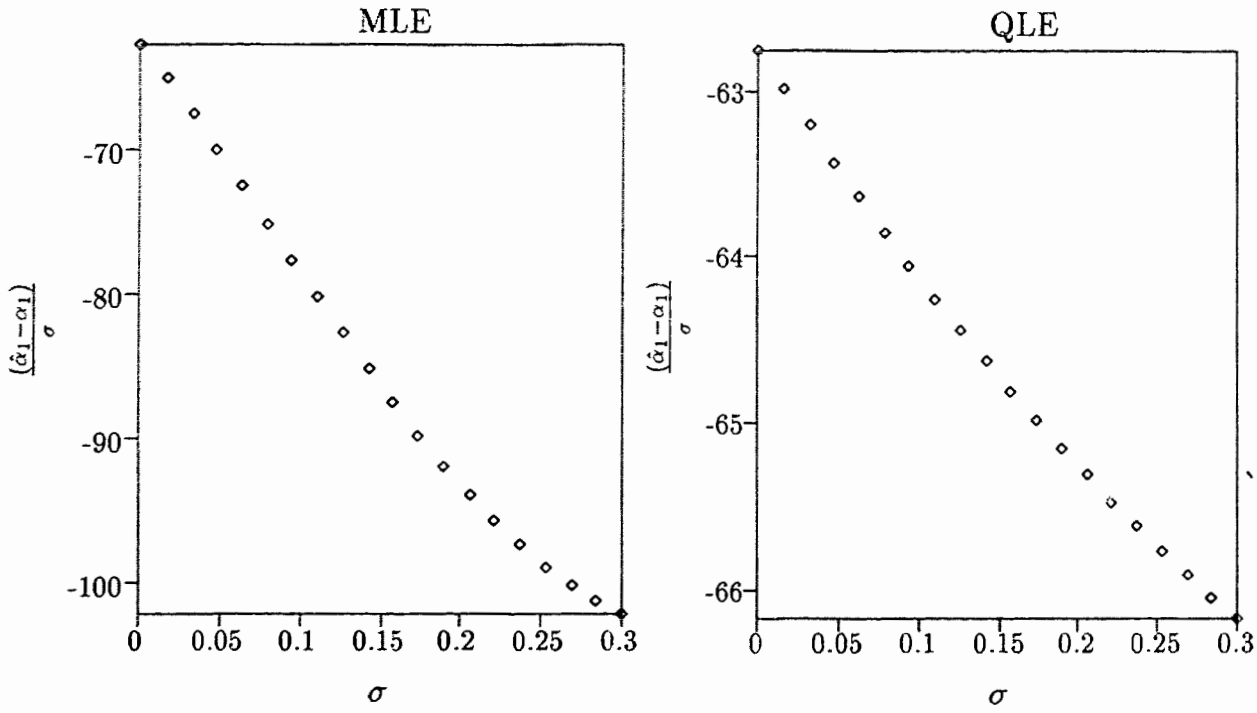

Figure 3.1: Plot of $\frac{(\hat{\alpha}_1-\alpha_1)}{\sigma}$ vs σ: ML and QL

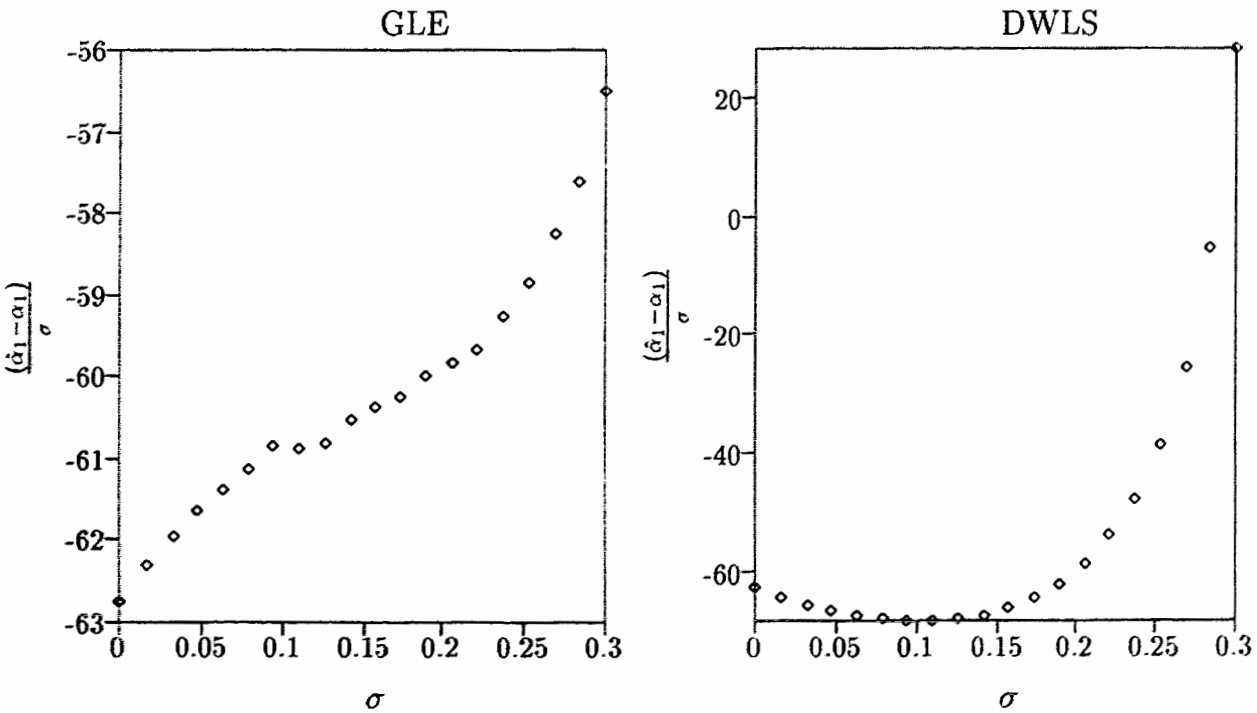

Figure 3.2: Plot of $\frac{(\hat{\alpha}_1-\alpha_1)}{\sigma}$ vs σ: GLS and DWLS

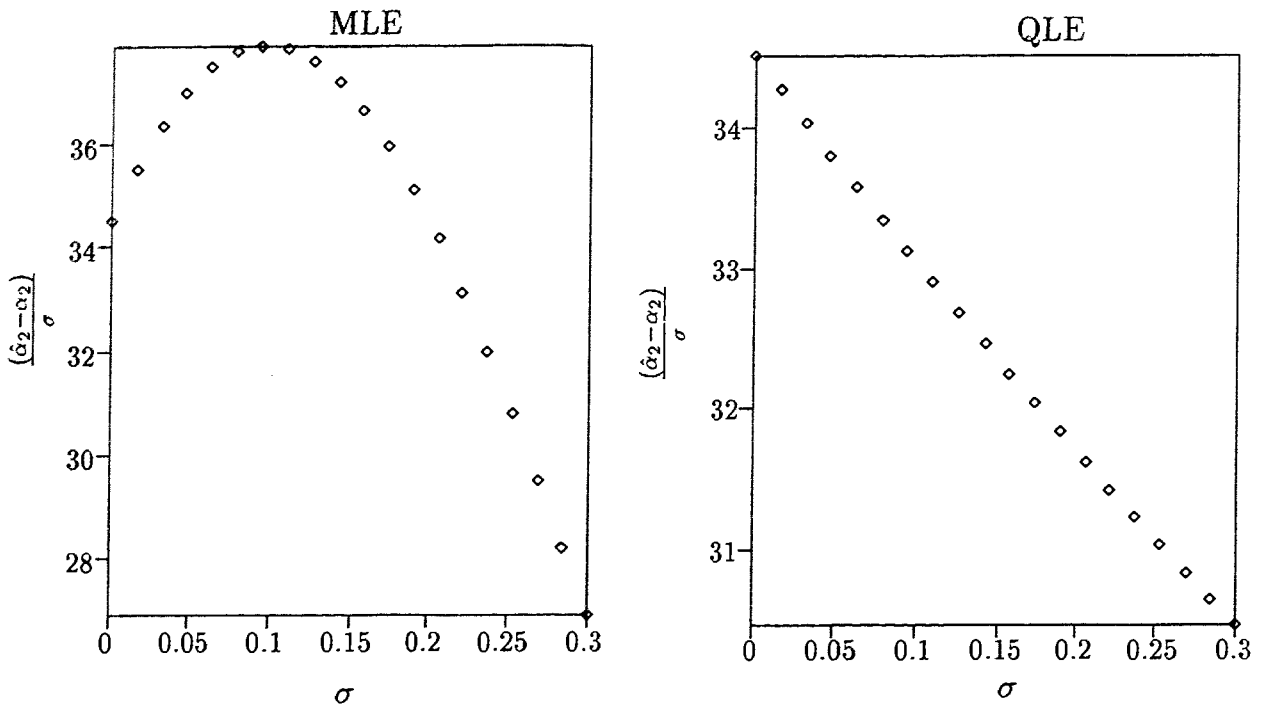

Figure 3.3: Plot of $\frac{(\hat{\alpha}_2-\alpha_2)}{\sigma}$ vs σ: ML and QL

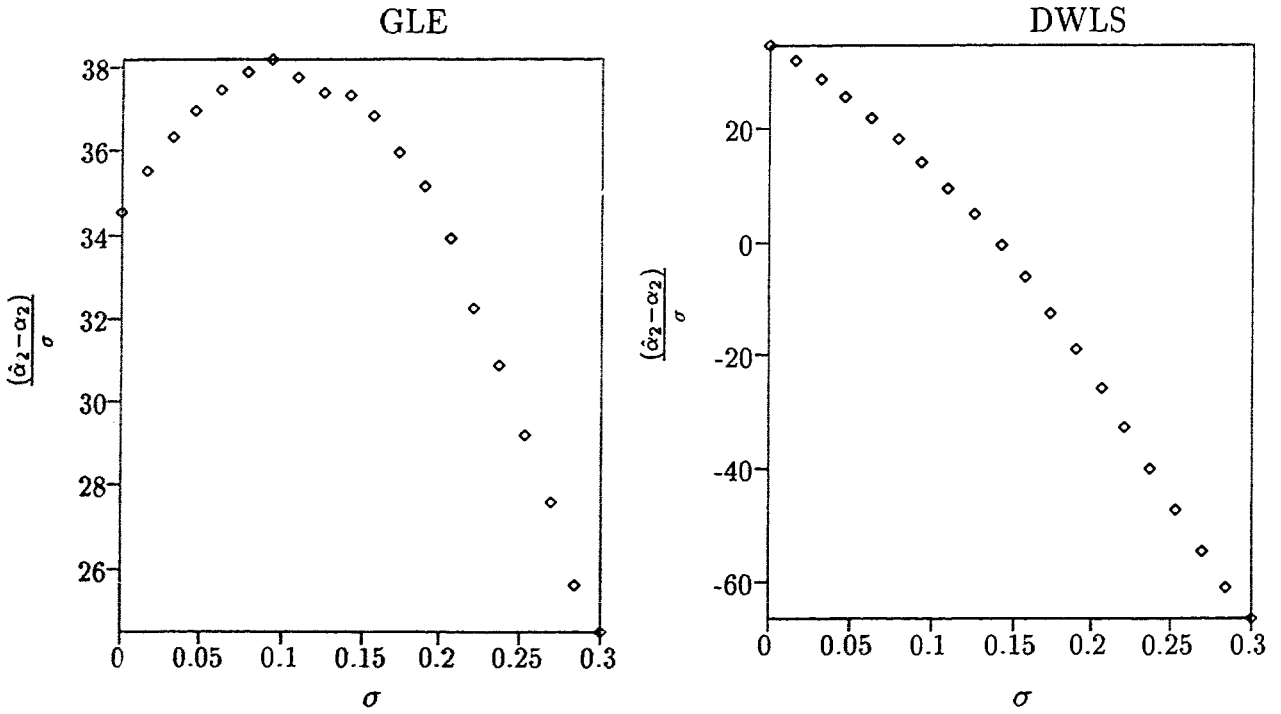

Figure 3.4: Plot of $\frac{(\hat{\alpha}_2-\alpha_2)}{\sigma}$ vs σ: GLS and DWLS

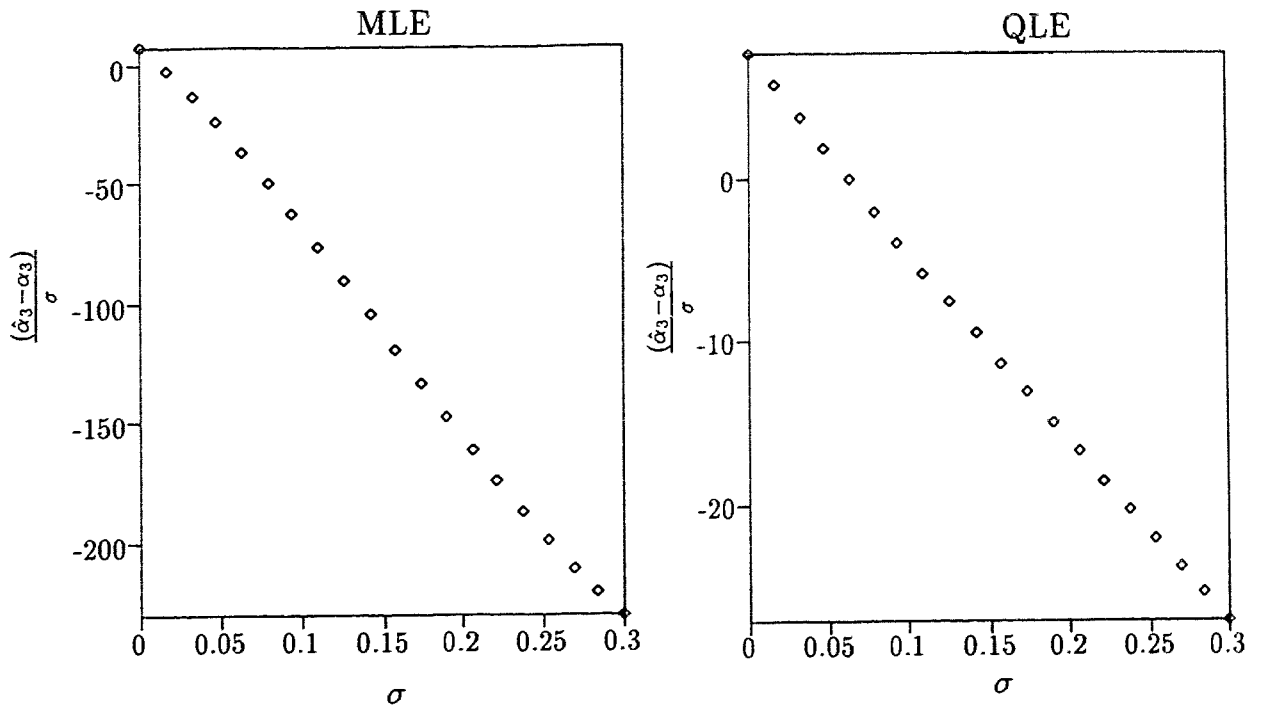

Figure 3.5: Plot of $\frac{(\hat{\alpha}_3-\alpha_3)}{\sigma}$ vs σ: ML and QL

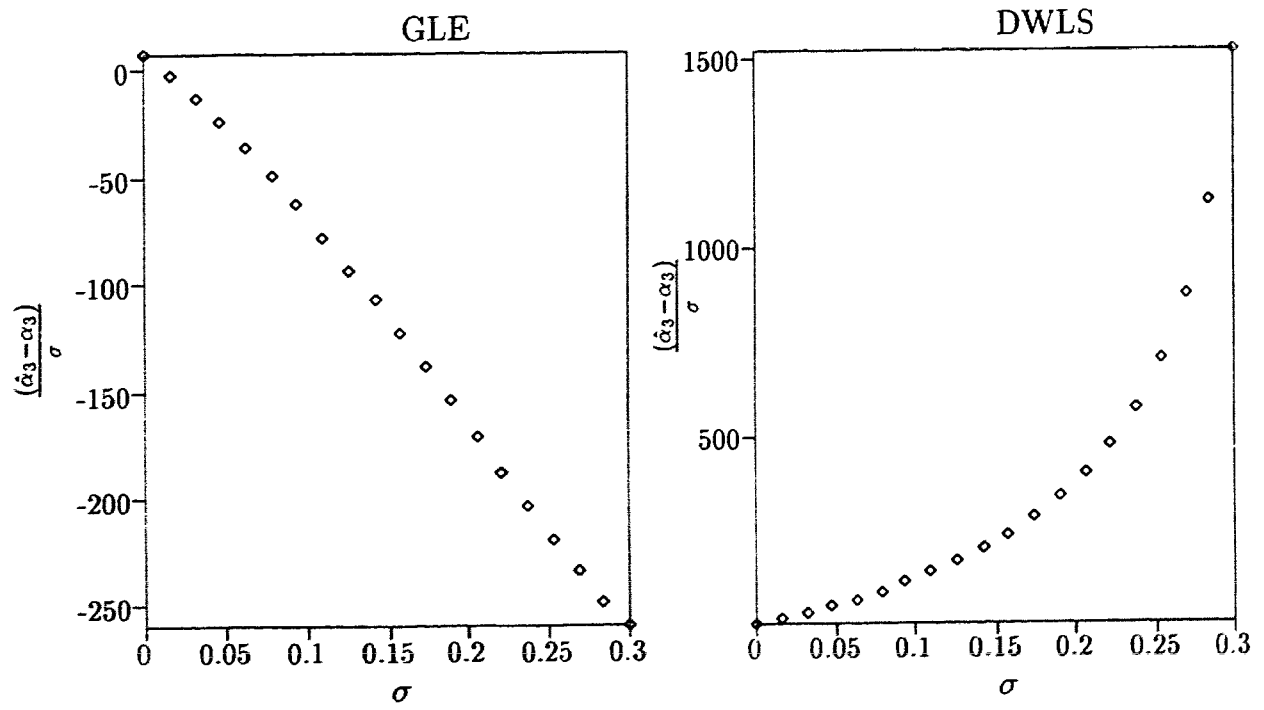

Figure 3.6: Plot of $\frac{(\hat{\alpha}_3-\alpha_3)}{\sigma}$ vs σ: GLS and DWLS

Conclusions:

1. Based on the Figures 3.1 - 3.6, we conclude that when σ is smaller than 0.1, the approximation $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2$ holds for all the estimation methods we studied. This justifies the validity of the above approximation for data collected in thermoluminescence studies.[2] The same figures demonstrate that the approximation does not hold for the methods ML, GLS and DWLS when σ is larger than 0.1.

2. We also computed estimates for the kth components of C_1 and C_2 by regressing the kth component of $(\hat{\theta} - \theta_0)/\sigma$ on σ; $(k = 1, 2, 3)$. Using the known parameter values and the ϵ_i's, the true values of C_1 and C_2 were also computed using the formulae derived earlier. For all four estimation methods the regression estimates were found to agree with the values from the formulae when the regression is based on the values of σ less than 0.1. This justifies the validity of the approximation $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2$ for thermoluminescence data. Using all the points on the plots in the regression, in particular points with $0.1 < \sigma < 0.3$, the ML, GLS and DWLS estimates for C_2 were found to show discrepancies from their true values whereas quasi-likelihood estimates were still found to agree with the true values.

3. We repeated the study by setting the parameters at $\theta = (96428.32, 193.3713, 761.6514)^T$. These are the QL estimates for the bleached data set QNL84-2 (Berger *et. al.* [12]). The dose values were set at the doses used in the same study. These dose levels are presented in Table 9.1 of Appendix 9.3 where they are coded as $P1B$. The results were similar to what we observed for the study reported earlier.

[2] Usually for data collected for thermoluminescence studies σ is around 0.03.

3.11 Assessing the validity of the formulae for estimating the biases

In deriving formulae for the biases of the estimators we assumed σ is small and the quantity $\frac{\hat{\theta}-\theta}{\sigma}$ varies roughly linearly with σ. While the study described in Section 3.10 indicates no evidence for possible violation of this assumption for σ in the range of real data sets, as a further diagnostic, we computed the biases from a simulation study and compared with the biases from the formulae. In contrast to the study we described in the previous section here σ was fixed at values for which we wish to study the biases. The parameter vector $\theta_0 = (\alpha_1, \alpha_2, \alpha_3)$ was set at $\alpha_1 = 142800.7, \alpha_2 = 122.737$ and $\alpha_3 = 391.9965$. The dose values used are presented in Table 9.1 of Appendix 9.3 where they are coded as $P1U$. Using each sample of ϵ_i's a sample of y_i's were generated according to $y_i = f(x_i, \theta_0)(1 + \sigma\epsilon_i)$. Using each of the four methods $\hat{\theta}$ was estimated from each sample. The true value of $C_1 = \left[\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1} \sum_{i=1}^{n} \epsilon_i \left(\frac{\nabla f_0}{f_0}\right)_i$ was computed using the ϵ_i's generated in step 3. The biases from the simulation study were computed as the average of the values for $(\hat{\theta} - \theta_0 - C_1\sigma)$.

Tables 3.2 to 3.4 presents the results of the Monte Carlo study based on $m = 10000$ simulations. In the tables:

B_T = True small σ bias using the formulae derived in Section 3.11

B_1 = the average of the $\hat{\theta} - \theta_0 - C_1\sigma$ values.

Remarks:

It is worth mentioning why we take the average of the differences $(\hat{\theta} - \theta_0 - C_1\sigma)$, but not $(\hat{\theta} - \theta_0)$. Let

$\hat{B}_1$ = the average of the m values of $(\hat{\theta} - \theta_0 - C_1\sigma)$ = the average of $C_2\sigma^2$,

and $\hat{B}_2$ = the average of the m values of $(\hat{\theta} - \theta_0)$ = the average of $C_1\sigma + C_2\sigma^2$.

We note that for large enough m, the average of the m values of $(\hat{\theta} - \theta_0 - C_1\sigma)$ should be close to $E(\hat{\theta} - \theta_0 - C_1\sigma)$. This is the bias of $\hat{\theta}$ since $E(C_1)$ is zero. Theoretically, both

averages $\hat{B}_1$ and $\hat{B}_2$ should provide reasonable estimates for the bias in $\hat{\theta}$. However, the large variability in C_1 produces a large variability in $\hat{B}_2$. In other words, $Var(\hat{B}_2) \geq Var(\hat{B}_1)$. Therefore, we report $\hat{B}_1$ as the bias from the simulation study.

σ	ML		QL		GLS		DWLS	
	B_T	B_1	B_T	B_1	B_T	B_1	B_T	B_1
0.005	4.87	4.80	5.73	5.71	7.77	7.68	1.64	1.77
0.01	19.49	19.65	22.92	22.86	31.09	31.24	6.58	6.54
0.02	77.95	77.25	91.68	91.63	124.36	123.78	26.32	27.71
0.03	175.40	180.01	206.28	209.44	279.82	284.52	59.21	60.70
0.05	487.21	503.50	573.01	590.34	777.28	794.14	164.48	196.62
0.1	1948.85	2089.49	2292.04	2502.18	3109.10	3366.71	657.91	1247.42

Table 3.2: Comparison of exact bias and estimated bias of $\hat{\alpha}_1$

σ	ML		QL		GLS		DWLS	
	B_T	B_1	B_T	B_1	B_T	B_1	B_T	B_1
0.005	0.002	0.002	0.004	0.004	0.002	0.002	0.007	0.007
0.01	0.010	0.010	0.016	0.016	0.010	0.010	0.028	0.027
0.02	0.039	0.039	0.063	0.063	0.039	0.039	0.111	0.112
0.03	0.088	0.091	0.142	0.143	0.088	0.091	0.249	0.249
0.05	0.243	0.260	0.393	0.404	0.243	0.260	0.692	0.710
0.1	0.974	0.994	1.573	1.599	0.974	1.001	2.770	3.118

Table 3.3: Comparison of exact bias and estimated bias of $\hat{\alpha}_2$

σ	ML		QL		GLS		DWLS	
	B_T	B_1	B_T	B_1	B_T	B_1	B_T	B_1
0.005	0.034	0.033	0.041	0.041	0.034	0.034	0.054	0.055
0.01	0.136	0.137	0.163	0.162	0.136	0.137	0.217	0.217
0.02	0.543	0.543	0.652	0.652	0.543	0.543	0.868	0.868
0.03	1.223	1.254	1.370	1.488	1.143	1.254	1.826	1.971
0.05	3.397	3.518	4.074	4.198	3.397	3.516	5.428	5.694
0.1	13.587	14.673	16.294	17.884	13.587	15.283	21.710	27.051

Table 3.4: Comparison of exact bias and estimated bias of $\hat{\alpha}_3$

Conclusions

Based on the study we draw the following conclusions.

1. The biases from the simulation study agree well with the biases from the derived formulae when $\sigma \leq 0.1$.

2. For the parameter α_1, the DWLS estimator appears to have the least bias whereas the GLS estimator appears to have the largest bias.

3. For the parameters α_2 and α_3, the ML and GLS estimators appears to have almost the same bias.

The above results were obtained using a sample of size 16. A detailed study comparing the biases of the four estimators is presented in Chapter 4.

3.12 Worked example

We developed software using the statistical package "S-plus" (and also in FORTRAN) to implement the suggested methodology. Next we demonstrate the theoretical results derived in this chapter using a real data set. The data set used here can be found in Berger *et.*

al. [12], where it is coded as QNL84-2. Parameter estimates for this data set are given in Tables 3.5 and 3.6.

In Tables 3.5 and 3.6, MSE stands for mean squared error. For maximum likelihood we have used the maximum likelihood estimate for σ. For the other three methods we have used the unbiased estimate for σ. The biases in Tables 3.5 and 3.6 are computed using the formulae given in Table 3.1 with parameters replaced by estimates; so, for example, the bias of ML estimates is computed by using ML estimates in Tables 3.5 and 3.6.

Data	para.	Description	Method			
			ML	QL	GLS	DWLS
QNL84-2	α_1	Estimate	142852.8	142800.7	142973.0	142461.8
(Unbleached)		bias	150.29	237.51	323.43	64.82
(n=16)		std. error (se)	4117.15	4557.60	4573.13	4546.62
		$bias/\sqrt{MSE} \times 100\%$	3.46	5.20	7.05	1.43
	α_2	Estimate	123.18	122.74	123.18	121.86
		bias	0.10	0.16	0.10	0.29
		std.error (se)	6.09	6.73	6.76	6.70
		$bias/\sqrt{MSE} \times 100\%$	1.31	2.38	1.48	4.32
	α_3	Estimate	393.07	392.00	393.07	389.92
		bias	1.27	1.69	1.42	2.24
		std.error (se)	27.79	30.68	30.82	30.56
		$bias/\sqrt{MSE} \times 100\%$	4.13	5.50	4.60	7.31
	σ	Estimate	0.029	0.032	0.032	0.032

Table 3.5: Parameter estimates for the QNL84-2 unbleached data set

Data	para.	Description	Method			
			ML	QL	GLS	DWLS
QNL84-2	α_1	Estimate	96144.8	96428.3	96301.7	96725.0
(Bleached)		bias	983.82	1515.38	1483.55	1615.75
(n=13)		std.error (se)	8322.65	9611.64	9511.99	9891.64
		$bias/\sqrt{MSE} \times 100\%$	11.72	15.57	15.41	16.12
	α_2	Estimate	192.55	193.37	192.54	195.18
		bias	0.53	1.02	0.69	1.70
		std.error (se)	16.34	18.75	18.62	19.14
		$bias/\sqrt{MSE} \times 100\%$	3.24	5.43	3.70	8.85
	α_3	Estimate	756.62	761.65	756.60	772.76
		bias	15.30	22.83	19.86	29.50
		std.error (se)	123.42	142.47	140.60	147.50
		$bias/\sqrt{MSE} \times 100\%$	12.30	15.82	13.99	19.61
	σ	Estimate	0.040	0.046	0.046	0.046

Table 3.6: Parameter estimates for the QNL84-2 bleached data set

3.13 Discussion

In this chapter, we discussed fitting the saturating exponential model introduced in Chapter 2. In sections 3.2 to 3.4, we derived Maximum Likelihood (ML), Quasi Likelihood (QL) and Generalized Least squares (GLS) estimates for the parameters. In Section 3.5, we obtained estimates from a procedure that we referred to as Data Weighted Least Squares (DWLS). In Section 3.6, we obtained maximum likelihood estimates assuming photon counts follow a gamma distribution.

In Section 3.8, we derived formulae for the biases of the estimators assuming:

1. the relative error in a single measurement σ is small.

2. when σ is small, $\hat{\theta}$ can be approximated by $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2$.

In Section 3.10, we examined this approximation from a Monte Carlo study. We found that, when σ is small ($\sigma < 0.1$), all four estimators satisfy the approximation $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2$. We also learned that the quasi-likelihood estimates satisfy the assumption for a wider range of σ compared to the estimates from other three procedures. Since data sets collected for thermoluminescence studies have small σ (usually $< 5\%$), we conclude that this approximation holds for all four estimators in our model.

In Section 3.11, we computed the biases from a simulation study and compared with the biases from the derived formulae. We found that the biases from the simulation study agree well with the biases from the formulae when σ is small ($\sigma \leq 0.1$). From the same study we found that, for the parameter α_1, the DWLS estimator has the least bias whereas the generalized least squares has the largest bias. For the parameters α_2 and α_3, the DWLS estimator was found to have the largest bias. For the parameters α_2 and α_3, the maximum likelihood estimator and the generalized least squares estimator were found to have almost the same bias. These results were observed using a sample of size 16. A detailed study on comparing the estimators will be offered in the the next chapter. From the results of the real data set we notice that the relative biases of all parameter estimates are small compared to their standard errors so that the biases are all negligible.

Chapter 4

Comparison of the estimation procedures

In this chapter, we compare the estimators discussed in Chapter 3. Two types of asymptotic results are of interest: behavior of the estimators as the sample size n becomes large and the behavior of the estimators as the relative error in a single measurement σ becomes small. The data sets collected in thermoluminescence studies are frequently small in size. (The partial bleach data sets we analyzed had n in the range 15 -30.) Moreover, the estimate of σ is usually small (in the range 0.02-0.05 in most cases). Thus, for thermoluminescence studies, a comparison of the methods as $\sigma \to 0$ is more appropriate. For completeness, we investigate both asymptotics.

In Section 4.1, we examine the large sample asymptotic behavior of the estimators. We show that the quasi-likelihood estimator and the maximum likelihood estimator are consistent for θ. We also provide distributional approximations for the maximum likelihood and quasi-likelihood estimators. In Section 4.2, we analyze the large sample behavior of our small σ approximation to $\hat{\theta} - \theta$. In this limiting case, we show that the quasi-likelihood estimator and the maximum likelihood estimator are mean squared error consistent. We also show that, in general, the generalized least squares and the data weighted least squares estimators have biases that do not vanish even asymptotically. However, for the parameter of interest in our model, generalized least squares and data weighted least squares were

also found to produce asymptotically unbiased estimators. In Section 4.3, we compare the biases of the estimators in small samples assuming σ is small. For realistic size samples, biases of all four estimators were found to be negligible relative to the standard errors of the estimators. Section 4.4 summarizes the chapter.

4.1 Large n, fixed σ asymptotics

Here we analyze the large sample asymptotic behavior of the estimators when σ is fixed and $n \longrightarrow \infty$. We prove that the maximum likelihood estimator and the quasi-likelihood estimator are consistent and generalized least squares and data weighted least squares are not.

Notation:

Let $f(\theta) = (f_1(\theta), \ldots f_n(\theta))$ be a real valued vector function. Assume that each component $f_i(\theta)$, $i = 1, \ldots n$ is defined on R^k. Let $Df(\theta)$ be the matrix of derivatives of f with respect to the components of θ and $\| f(\theta) \|^2 = \sum_{i=1}^{n} f_i^2(\theta)$ be the Euclidean norm of f.

We use the notations $\lambda_{\min}[A]$ and $\lambda_{\max}[A]$ to denote the smallest and largest eigenvalues of a given matrix A.

We begin by examining the following result which is useful in proving the consistency of the maximum likelihood and the quasi-likelihood estimators.

Result 1 *Let $\Omega \subseteq R^k$ be a bounded open set. Let $\partial\Omega$ denote the boundary of Ω. Let $\overline{\Omega} = \Omega \cup \partial\Omega$ be the closure of Ω.*

Let $f: \overline{\Omega} \longmapsto R^k$ be such that:

1. *f is continuous on $\overline{\Omega}$.*
2. *f is differentiable in Ω.*
3. *$Df(\theta)$ is not singular in Ω.*
4. *$\exists \theta^* \in \Omega$ such that, $\| f(\theta^*) \| < \inf \{ \| f(\theta) \| \mid \theta \in \partial\Omega \}$.*

Then, $\exists \ \tilde{\theta} \in \Omega$ such that, $f(\tilde{\theta}) = 0$.

Proof:

Note that the function g defined by $g(\theta) = \| f(\theta) \|^2 = \sum_{i=1}^{n} f_i^2(\theta)$ is differentiable in Ω. Since f is continuous on $\overline{\Omega}$, g is also continuous on $\overline{\Omega}$. Therefore, $\exists \tilde{\theta}$ in $\overline{\Omega}$ such that $g(\tilde{\theta})$ is a minimum. Condition 4 implies $\tilde{\theta} \in \Omega$. Therefore, $Dg(\tilde{\theta}) = 0$. This implies that $Df(\tilde{\theta})f(\tilde{\theta}) = 0$. But $Df(\tilde{\theta})$ is not singular. Therefore, $f(\tilde{\theta}) = 0$; hence the theorem.

Now we apply the above result to see that the maximum likelihood and quasi-likelihood estimators are consistent. Writing $f_i = f(x_i, \theta)$, the maximum likelihood estimates $(\hat{\theta}, \hat{\sigma})$ solve the set of equations $H_n(\theta)$ defined by

$$-\sum_{i=1}^{n}\left[\frac{\nabla f_i}{f_i}\right] + \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[\frac{y_i - f_i}{f_i^2}\right]\nabla f_i + \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[\frac{y_i - f_i}{f_i^3}\right]\nabla f_i = 0, \tag{4.1}$$

and

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}\left[\frac{y_i - f_i}{f_i^2}\right]\nabla f_i = 0. \tag{4.2}$$

The quasi-likelihood estimator $\hat{\theta}$ solves

$$\sum_{i=1}^{n}\left[\frac{y_i - f_i}{f_i^2}\right]\nabla f_i = 0. \tag{4.3}$$

So both estimators solve a general equation of the form

$$H_n(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_i(Y_i, \theta) = 0. \tag{4.4}$$

Under the assumption that $y_i = f(x_i, \theta)(1 + \sigma\epsilon_i)$ with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = 1$ we have $E\left(h_i(Y_i, \theta)\right) = 0$ and $Var\left(h_i(Y_i, \theta)\right) < \infty$.

Result 2 *Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with mean $E(Y_i) = f(x_i, \theta)$ and $V(Y_i) = \sigma^2 f^2(x_i, \theta)$. Under the following assumptions, the maximum likelihood and quasi-likelihood estimators are consistent for θ.*

Assumptions:

1. *The function $h_i(Y_i, \theta)$ is differentiable in θ for each Y_i.*

2. *As $n \longrightarrow \infty$, $D = \left(\frac{1}{n}\sum_{i=1}^{n} Dh_i(\theta_0)\right)^T \left(\frac{1}{n}\sum_{i=1}^{n} Dh_i(\theta_0)\right) \longrightarrow M$, where M is a positive definite non - random matrix.*

3. *As $n \longrightarrow \infty$, $Var\,(H_n(\theta_0)) = \frac{1}{n}\sum_{i=1}^{n} Var(h_i) \longrightarrow \Sigma(\theta_0)$, where $\Sigma(\theta_0)$ is non-singular.*

4. *$\exists\, T_i$ such that $\frac{1}{n}\sum_{i=1}^{n} T_i$ is $O_p(1)$ and there exists a $\delta > 0$ such that*

$$\| Dh_{ik}(\theta^*) - Dh_{ik}(\theta_0) \| \leq T_i\, |\theta^* - \theta_0|\,, \quad \textit{whenever } |\theta^* - \theta_0| < \delta;$$

here h_{ik} is the kth component of h_i.

Proof

Let θ_0 be the true parameter value. Let L_n be a real number that depends on n such that $\frac{L_n^2}{\sqrt{n}}$ is $o(1)$ (eg. $L_n = log(n)$ or $L_n = n^{1/5}$) and $L_n \longrightarrow \infty$, as $n \longrightarrow \infty$.
Let

$$\Omega = \left\{\theta \;\middle|\; |\theta - \theta_0| < \frac{L_n}{\sqrt{n}}\right\},$$
$$\partial\Omega = \left\{\theta \;\middle|\; |\theta - \theta_0| = \frac{L_n}{\sqrt{n}}\right\},$$

and

$$\overline{\Omega} = \Omega \cup \partial\Omega = \left\{\theta \;\middle|\; |\theta - \theta_0| \leq \frac{L_n}{\sqrt{n}}\right\}.$$

Let $A_n(L_n)$ be the event that

$$\| H_n(\theta_0) \| < \inf\left\{\| H_n(\theta) \| \;\middle|\; |\theta - \theta_0| = \frac{L_n}{\sqrt{n}}\right\}.$$

Take $f = H_n(\theta)$ in Result 1. Then, f is continuous in Ω and Assumption 1 implies f is differentiable in Ω. To apply Result 1, we need to prove that

1. $DH_n(\theta)$ is nonsingular in Ω and

2. $P(A_n(L_n)) \to 1$, as $n \to \infty$, or equivalently, for any given $\epsilon > 0$, a positive integer $n_0(\epsilon)$ exists such that $P(A_n(L_n)) \geq 1 - \epsilon$, for all $n \geq n_0(\epsilon)$.

Proving Condition 1:

We prove that $DH_n(\theta)$ is non-singular in Ω by proving $(DH_n(\theta))^T (DH_n(\theta))$ is uniformly positive definite in Ω. Equivalently, we prove that $x^T (DH_n(\theta))^T (DH_n(\theta)) x$ is positive for arbitrary x such that $x^T x = 1$. Let $Q_n(\theta) = \frac{1}{n} (DH_n(\theta))^T (DH_n(\theta))$ and $Q_n(\theta_0) = \frac{1}{n} (DH_n(\theta_0))^T (DH_n(\theta_0))$. Then, $x^T Q_n(\theta) x = x^T (Q_n(\theta) - Q_n(\theta_0)) x + x^T Q_n(\theta_0) x$. Assumption 2 implies that $x^T Q_n(\theta_0) x \geq \frac{1}{2} \lambda_{\min}[M]$, for all x and for all sufficiently large n.

Now consider

$$\begin{aligned} n(Q_n(\theta) - Q_n(\theta_0)) &= (DH_n(\theta) - DH_n(\theta_0))^T (DH_n(\theta) - DH_n(\theta_0)) \\ &\quad + (DH_n(\theta) - DH_n(\theta_0))^T DH_n(\theta_0) \\ &\quad + (DH_n(\theta_0))^T (DH_n(\theta) - DH_n(\theta_0)). \end{aligned}$$

Let U_j and V_k be the jth row and k th column of the matrix $\frac{1}{\sqrt{n}} (DH_n(\theta) - DH_n(\theta_0))$ respectively. Consider the inner product of U_j and V_k. Using the Cauchy Schwarz inequality,

$$\| U_j V_k \| \leq \| U_j \| \| V_k \| .$$

Now observe that

$$\begin{aligned} \| U_j \| \leq \frac{1}{n} \sum_{i=1}^{n} \| j \text{ th row of } (Dh_i(\theta) - Dh_i(\theta_0)) \| &\leq \frac{1}{n} \sum_{i=1}^{n} T_i |\theta - \theta_0| \\ &= O_p\left(\frac{L_n}{\sqrt{n}}\right). \end{aligned}$$

Thus, the inner product of U_j and V_k is of order $O_p\left(\frac{L_n^2}{n}\right)$. Similarly, since $\frac{1}{\sqrt{n}} DH_n(\theta_0)$ is $O(1)$, we find that the term $(DH_n(\theta_0))^T (DH_n(\theta) - DH_n(\theta_0))$ and its transpose are $O_p\left(\frac{L_n}{\sqrt{n}}\right)$. Recall that $\left(\frac{L_n}{\sqrt{n}}\right)$ is $o(1)$. Consequently, for large enough n, for all x,

$$\begin{aligned} x^T Q_n(\theta) x &\geq \frac{1}{2} \lambda_{\min}[M] + o_p(1) \\ &\geq \frac{1}{4} \lambda_{\min}[M]. \end{aligned}$$

Since M is positive definite $\lambda_{\min}[M] > 0$ and hence $x^T Q_n(\theta) x > 0$. Thus, $\frac{1}{n} (DH_n(\theta))^T (DH_n(\theta))$ is uniformly positive definite in Ω.

Proving Condition 2:

Let θ^* be such that

$$\| H_n(\theta^*) \| = \inf \left\{ \| H_n(\theta) \| \quad | \quad |\theta - \theta_0| = \frac{L_n}{\sqrt{n}} \right\}.$$

Then, $\theta^* = \theta_0 + \frac{L_n}{\sqrt{n}} v$, for some v such that $\| v \|^2 = v^T v = 1$.

Taylor expansion of $H_n(\theta^*)$ around θ_0 gives

$$H_n\left(\theta_0 + \frac{L_n}{\sqrt{n}} v\right) = H_n(\theta_0) + DH_n(\theta_0) \frac{L_n}{\sqrt{n}} v + R_n,$$

where R_n is the remainder term.

Thus,

$$\begin{aligned} \| DH_n(\theta_0) \frac{L_n}{\sqrt{n}} v \| &= \| H_n(\theta_0 + \frac{L_n}{\sqrt{n}} v) - H_n(\theta_0) - R_n \| \\ &\leq \| H_n(\theta_0 + \frac{L_n}{\sqrt{n}} v) \| + \| H_n(\theta_0) \| + \| R_n \| . \end{aligned}$$

Therefore,

$$\| H_n(\theta_0 + \frac{L_n}{\sqrt{n}} v) \| \geq \| DH_n(\theta_0) \frac{L_n}{\sqrt{n}} v \| - \| H_n(\theta_0) \| - \| R_n \| .$$

Use the fact that $\| v \| = 1$ to see

$$\begin{aligned} \| DH_n(\theta_0) \frac{L_n}{\sqrt{n}} v \| &= \sqrt{v^T (DH_n(\theta_0))^T (DH_n(\theta_0)) v} \ \frac{L_n}{\sqrt{n}} \\ &\geq \sqrt{v^T \lambda_{\min} \left[(DH_n(\theta_0))^T (DH_n(\theta_0)) \right] v} \ \frac{L_n}{\sqrt{n}} \\ &= \sqrt{\lambda_{\min} \left[(DH_n(\theta_0))^T (DH_n(\theta_0)) \right]} \ \frac{L_n}{\sqrt{n}}, \\ &= \sqrt{\frac{1}{n} \lambda_{\min} \left[(DH_n(\theta_0))^T (DH_n(\theta_0)) \right]} \ L_n \\ &= \sqrt{\frac{1}{n} \lambda_{\min} \left[\left(\sum_{i=1}^n \frac{1}{\sqrt{n}} Dh_i(\theta_0) \right)^T \left(\sum_{i=1}^n \frac{1}{\sqrt{n}} Dh_i(\theta_0) \right) \right]} \ L_n \\ &= \sqrt{\lambda_{\min} \left[\left(\sum_{i=1}^n \frac{1}{n} Dh_i(\theta_0) \right)^T \left(\sum_{i=1}^n \frac{1}{n} Dh_i(\theta_0) \right) \right]} \ L_n. \end{aligned}$$

Assumption 2 implies that the eigenvalues of $D = (\sum_{i=1}^n \frac{1}{n} Dh_i(\theta_0))^T (\sum_{i=1}^n \frac{1}{n} Dh_i(\theta_0))$ converges to the eigenvalues of M, as $n \to \infty$. So, for large enough n, $\| DH_n(\theta_0) \frac{L_n}{\sqrt{n}} v \| \geq$

$\frac{1}{2}\sqrt{\lambda_{\min}[M]}L_n$. The right hand side of this inequality can be made arbitrarily large by choosing arbitrarily large L_n.

Assumption 3 and $E_{\theta_0}(H_n(\theta_0)) = 0$ imply $H_n(\theta_0) = O_p(1)$. Therefore, for large enough n,

$$\| H_n(\theta_0 + \frac{L_n}{\sqrt{n}}v) \| \geq \frac{1}{2}\sqrt{\lambda_{\min}(M)}L_n - O_p(1) - \| R_n \| . \tag{4.5}$$

Now consider the remainder term

$$R_n = H_n\left(\theta_0 + \frac{L_n}{\sqrt{n}}v\right) - H_n(\theta_0) - DH_n(\theta_0)\frac{L_n}{\sqrt{n}}v.$$

Let R_{nk} and H_{nk} be the $k-th$ components of R_n and H_n respectively. Then,

$$H_{nk}\left(\theta_0 + \frac{L_n}{\sqrt{n}}v\right) = H_{nk}(\theta_0) + DH_{nk}(\theta_*)\frac{L_n}{\sqrt{n}}v,$$

where $\theta_* = h\theta_0 + (1-h)(\theta_0 + \frac{L_n}{\sqrt{n}}v)$, for some $0 \leq h \leq 1$. Thus,

$$\| R_{nk} \| = \| [DH_{nk}(\theta_*) - DH_{nk}(\theta_0)] \frac{L_n}{\sqrt{n}}v \| .$$

Since $v^T v = 1$, we find

$$\begin{aligned} \| R_{nk} \| &\leq \| DH_{nk}(\theta_*) - DH_{nk}(\theta_0) \| \frac{L_n}{\sqrt{n}} \\ &= \| \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} Dh_{ik}(\theta_*) - \sum_{i=1}^{n} Dh_{ik}(\theta_0)\right) \| \frac{L_n}{\sqrt{n}} \end{aligned}$$

Now Assumption 4 implies

$$\| R_{nk} \| \leq \frac{1}{n}\sum_{i=1}^{n} T_i \frac{L_n^2}{\sqrt{n}}.$$

Take $L_n = n^{1/5}$. Then, $\frac{L_n^2}{\sqrt{n}}$ is $o_p(1)$ and from Assumption 4, $\frac{1}{n}\sum_{i=1}^{n} T_i$ is $O_p(1)$. Therefore, we find R_{nk} is $o_p(1)$. For this L_n, Equation 4.5 therefore implies

$$\| H_n(\theta_0 + \frac{L_n}{\sqrt{n}}v) \| \geq \frac{1}{2}\sqrt{\lambda_{\min}(M)}\, n^{1/5} - O_p(1) - o_p(1) \tag{4.6}$$

Since the right hand side of Equation 4.6 is free of v and $\| H_n(\theta_0) \| = O_p(1)$, we find

$$P\left(\inf\left\{\| H_n(\theta_0 + \frac{L_n}{\sqrt{n}}v) \| \middle| \; v^T v = 1\right\} > \| H_n(\theta_0) \|\right) \longrightarrow 1, \text{as } n \longrightarrow \infty. \tag{4.7}$$

Then, Result 1 implies the existence of $\tilde{\theta_n}$ such that $|\tilde{\theta_n} - \theta_0| < \frac{L_n}{\sqrt{n}} = \frac{1}{n^{3/10}}$, and $H_n(\tilde{\theta_n}) = 0$. Thus, we have established the existence of a root of $H_n(\theta)$ in a neighborhood of θ_0.

Proving uniqueness

Suppose there exists more than one root of $H_n(\theta)$ in Ω. Let θ_1 and θ_2 be two such distinct roots. Then, $\| \theta_1 - \theta_2 \| > 0$ and $H_n(\theta_1) = H_n(\theta_2) = 0$. Since $\theta_1, \theta_2 \epsilon \Omega$, $\| \theta_1 - \theta_2 \| < 2\frac{L_n}{\sqrt{n}}$. Taylor expansion of $H_n(\theta_1)$ around θ_2 gives

$$H_n(\theta_1) = H_n(\theta_2) + (DH_n(\theta_1))(\theta_1 - \theta_2) + R_n,$$

where R_n is $o(\| \theta_1 - \theta_2 \|) = o(\frac{L_n}{\sqrt{n}})$.

Consider

$$\begin{aligned} \| (DH_n(\theta_1))(\theta_1 - \theta_2) \|^2 &= (\theta_1 - \theta_2)^T DH_n(\theta_1) (DH_n(\theta_1))^T (\theta_1 - \theta_2) \\ &\geq \lambda_{\min} \left[(DH_n(\theta_1))^T (DH_n(\theta_1)) \right] \| \theta_1 - \theta_2 \|^2, \end{aligned}$$

From the proof of Condition 2, it follows that $DH_n(\theta_1)(DH_n(\theta_1))^T$ is positive definite. Therefore, $\lambda_{\min} > 0$. Since R_n is $o_p(\| \theta_1 - \theta_2 \|)$, with probability tending to 1, $\| R_n \| < \sqrt{\lambda_{\min}} \| \theta_1 - \theta_2 \| /2$.

Now consider

$$\begin{aligned} \| H_n(\theta_1) - H_n(\theta_2) \| &= \| DH_n(\theta_1)(\theta_1 - \theta_2) + R_n \| \\ &\geq \| DH_n(\theta_1)(\theta_1 - \theta_2) \| - \| R_n \| \\ &\geq \sqrt{\lambda_{\min}} \| \theta_1 - \theta_2 \| /2 \end{aligned}$$

Since $\| \theta_1 - \theta_2 \| > 0$, this contradicts $\| H_n(\theta_1) - H_n(\theta_2) \| = 0$. So, there do not exist θ_1, θ_2 in Ω such that $\theta_1 \neq \theta_2$ and $H_n(\theta_1) = H_n(\theta_2) = 0$.

Let $\tilde{\theta_n}$ be the unique root of $H_n(\theta) = 0$. As $|\tilde{\theta_n} - \theta_0| < \frac{L_n}{\sqrt{n}}$, given any $\epsilon > 0$, a positive integer $n_0(\epsilon)$ exists such that, $|\tilde{\theta_n} - \theta_0| < \epsilon$, for all $n \geq n_0(\epsilon)$ (eg. take any integer $n > \frac{1}{\epsilon^{10/3}}$). This implies that $\tilde{\theta_n} \xrightarrow{p} \theta_0$. Thus, under the listed assumptions, we have established $\hat{\theta}_{ML} \xrightarrow{p} \theta_0$, and $\hat{\theta}_{QL} \xrightarrow{p} \theta_0$. Hence, maximum likelihood and quasi-likelihood estimators are consistent for θ_0.

Now we examine the assumptions that guarantee the consistency of maximum likelihood and quasi-likelihood estimators for our model of interest.

Quasi-likelihood

From the quasi-likelihood estimating equations $h_{ik}(\theta) = \frac{(Y_i - f_i)}{f_i^2}(\bigtriangledown f_i)_k$. Thus, for quasi-likelihood $Dh_{ik}(\theta)$ is of the form $Dh_{ik}(\theta) = y_i G_{1i}(\theta) + G_{2i}(\theta)$, where $G_{1i}(\theta)$ and $G_{2i}(\theta)$ are polynomial functions of $f(x_i, \theta)$ and $1/f(x_i, \theta)$. For $Dh_{ik}(\theta)$ to be bounded we therefore need to find δ, K, ϵ such that for all $x \geq 0$, $\quad 0 < \epsilon \leq f(x_i, \theta) \leq K$, whenever $|\theta - \theta_0| < \delta$. For saturating exponential model this is clearly true provided only $f(0, \theta_0) > 0$. Moreover, we must assume the existence of M such that $Var(Y_i) \leq M$ for all i. Note that if $f(x_i, \theta_0)$ is bounded over all $x > 0$ and θ close to θ_0, as is the case for saturating exponentials, then, when the model equation $y_i = f(x_i, \theta_0)(1 + \sigma\epsilon_i)$ hold with iid ϵ_i the variance condition is simply $Var(\epsilon_i) < \infty$.

For the regeneration model that we discuss in Chapter 6, we need the additional condition that $\frac{1}{n}\sum_{i=1}^{n} x_i$ is bounded. This again is clearly satisfied for sensibly chosen designs.

Maximum likelihood

For maximum likelihood $Dh_{ik}(\theta)$ is of the form $Dh_{ik}(\theta) = y_i^2 G_{1i}(\theta) + y_i G_{2i}(\theta) + G_{3i}(\theta)$, where again $G_{1i}(\theta)$, $G_{2i}(\theta)$ and $G_{3i}(\theta)$ are polynomial functions of $f(x_i, \theta)$ and $1/f(x_i, \theta)$. Thus, in this case in addition to the conditions discussed for quasi-likelihood we need the existence of M such that $Var(Y_i^2) < M$. So, if $f(x_i, \theta_0)$ is bounded over all $x > 0$ and θ close to θ_0, then when the model equation $y_i = f(x_i, \theta_0)(1 + \sigma\epsilon_i)$ hold with iid ϵ_i the variance condition is simply $E(\epsilon_i^4) < \infty$.

Generalized least squares and data weighted least squares

For fixed σ, the estimating equations satisfied by the generalized least squares and the data weighted least squares estimators are not unbiased (For GLS it is easy to compute the expected value of the estimating equations; for DWLS this is difficult in general but possible for the gamma model. In both cases the expected value is a multiple of $\sum_{i=1}^{n}\left(\frac{\bigtriangledown f_0}{f_0}\right)$ which is 0.). So, generally the generalized least squares and the data weighted least squares estimates will not be consistent.

4.1.1 Distributional approximations for $\hat{\theta}$

Quasi-likelihood estimate

If $E(Y_i) = f(x_i, \theta)$ the quasi-likelihood estimating equations are unbiased. General large sample considerations will then establish that, in large samples and assuming $Var(Y_i) < \infty$,

$$(\hat{\theta} - \theta_0) \sim MVN\left(0, E\left[-H_n'(\theta)\right]^{-1} Var\left(H_n(\theta)\right) E\left[-H_n'(\theta)\right]^{-1}\right),$$

where $H_n'(\theta)$ is the derivative of $H_n(\theta)$ with respect to θ. It is easy to see that $E[H_n'(\theta)] = \sum_{i=1}^n \left(\frac{\nabla f}{f}\right)_i \left(\frac{\nabla f}{f}\right)_i^T$. Therefore, $E[H_n'(\theta)]^{-1} = \left[\sum_{i=1}^n \left(\frac{\nabla f}{f}\right)_i \left(\frac{\nabla f}{f}\right)_i^T\right]^{-1} = \Sigma$.
Also notice that

$$Var\left(H_n(\theta)\right) = \sum_{i=1}^n \frac{Var(Y_i)}{f_i^2} \left(\frac{\nabla f}{f}\right)_i \left(\frac{\nabla f}{f}\right)_i^T.$$

Thus, we find

$$(\hat{\theta} - \theta_0) \sim MVN\left(0, \Sigma\left[\sum_{i=1}^n \frac{Var(Y_i)}{f_i^2} \left(\frac{\nabla f}{f}\right)_i \left(\frac{\nabla f}{f}\right)_i^T\right] \Sigma\right).$$

For $Var(Y_i) = \sigma^2 f^2(x_i, \theta)$ this simplifies to

$$(\hat{\theta} - \theta_0) \sim MVN\left(0, \sigma^2 \Sigma\right).$$

Maximum likelihood estimate

Recall that maximum likelihood estimates $(\hat{\theta}, \hat{\sigma})$ solve the set of equations $H_n(\theta)$ defined by Equations 4.1 and 4.2. If $E(Y_i) = f(x_i, \theta)$ and $Var(Y_i) = \sigma^2 f^2(x_i, \theta)$ these equations are unbiased. In large samples, assuming $E(Y_i^4) < \infty$ we find

$$\begin{pmatrix} \hat{\theta} - \theta \\ \hat{\sigma} - \sigma \end{pmatrix} \sim MVN\left(0,\ E\left[-H_n'(\theta)\right]^{-1} Var\left(H_n(\theta)\right) E\left[-H_n'(\theta)\right]^{-1}\right),$$

where

$$H_n'(\theta) = \begin{bmatrix} \frac{\partial^2 l}{\partial \theta \partial \theta^T} & \frac{\partial^2 l}{\partial \theta \partial \sigma} \\ \frac{\partial^2 l}{\partial \theta^T \partial \sigma} & \frac{\partial^2 l}{\partial \sigma^2} \end{bmatrix}.$$

As we showed in Chapter 3, $E\left(-\frac{\partial^2 l}{\partial \theta \partial \theta^T}\right) = D^T A D$, where D is the $n \times p$ matrix with (i,j)th entry $\frac{\partial f_i}{\partial \theta_j}$ and A is the diagonal matrix with ith diagonal element $2/f_i^2 + 1/(\sigma^2 f_i^2)$.

So, D^TAD can be written as $\left(\frac{1+2\sigma^2}{\sigma^2}\right)\left(\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i\left(\frac{\nabla f}{f}\right)_i^T\right)$.

Now consider

$$\frac{\partial^2 l}{\partial\theta\partial\sigma} = -\frac{2}{\sigma^3}\sum_{i=1}^{n}\left(\frac{y_i-f_i}{f_i}\right)\left(\frac{\nabla f}{f}\right)_i - \frac{2}{\sigma^3}\sum_{i=1}^{n}\left(\frac{y_i-f_i}{f_i}\right)^2\left(\frac{\nabla f}{f}\right)_i.$$

It is easy to see that $E\left[-\frac{\partial^2 l}{\partial\theta\partial\sigma}\right] = \frac{2}{\sigma}\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i$.

Since $\frac{\partial^2 l}{\partial\sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}\left(\frac{y_i-f_i}{f_i}\right)^2$, we find $E\left(-\frac{\partial^2 l}{\partial\sigma^2}\right) = \frac{2n}{\sigma^2}$.

Now consider

$$Var\left(H_n(\theta)\right) = \begin{bmatrix} Var\left(\frac{\partial l}{\partial\theta}\right) & Cov\left(\frac{\partial l}{\partial\theta},\frac{\partial l}{\partial\sigma}\right) \\ \left(Cov\left(\frac{\partial l}{\partial\theta},\frac{\partial l}{\partial\sigma}\right)\right)^T & Var\left(\frac{\partial l}{\partial\sigma}\right) \end{bmatrix}.$$

The components of $Var\left(H_n(\theta)\right)$ are computed as follows:

$$\begin{aligned} Var\left(\frac{\partial l}{\partial\theta}\right) &= \frac{1}{\sigma^4}\sum_{i=1}^{n} Var\left(\frac{y_i-f_i}{f_i}\right)\left(\frac{\nabla f}{f}\right)_i\left(\frac{\nabla f}{f}\right)_i^T \\ &\quad + \frac{1}{\sigma^4}\sum_{i=1}^{n} Var\left(\frac{y_i-f_i}{f_i}\right)^2\left(\frac{\nabla f}{f}\right)_i\left(\frac{\nabla f}{f}\right)_i^T \\ &\quad + \frac{2}{\sigma^4}\sum_i\sum_j Cov\left[\left(\frac{y_i-f_i}{f_i}\right)\left(\frac{\nabla f}{f}\right)_i, \left(\frac{y_j-f_j}{f_j}\right)^2\left(\frac{\nabla f_0}{f_0}\right)_j\right] \\ &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i\left(\frac{\nabla f}{f}\right)_i^T + \frac{1}{\sigma^4}\sum_{i=1}^{n} Var\left(\frac{y_i-f_i}{f_i}\right)^2\left(\frac{\nabla f}{f}\right)_i\left(\frac{\nabla f}{f}\right)_i^T \\ &\quad + \frac{2}{\sigma^4}\sum_{i=1}^{n} E\left[\left(\frac{y_i-f_i}{f_i}\right)^3\right]\left(\frac{\nabla f}{f}\right)_i\left(\frac{\nabla f}{f}\right)_i^T, \\ Cov\left(\frac{\partial l}{\partial\theta},\frac{\partial l}{\partial\sigma}\right) &= \frac{1}{\sigma^5}\sum_{i=1}^{n} E\left[\left(\frac{y_i-f_i}{f_i}\right)^3\right]\left(\frac{\nabla f}{f}\right)_i + \frac{1}{\sigma^5}\sum_{i=1}^{n} E\left[\left(\frac{y_i-f_i}{f_i}\right)^4\right]\left(\frac{\nabla f}{f}\right)_i \\ &\quad - \frac{1}{\sigma}\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i, \end{aligned}$$

and

$$Var\left(\frac{\partial l}{\partial\sigma}\right) = \frac{1}{\sigma^6}\sum_{i=1}^{n} Var\left[\left(\frac{y_i-f_i}{f_i}\right)^2\right].$$

Now assuming $E\left[(Y_i-f_i)^3\right] = 0$ and $E\left[(Y_i-f_i)^4\right] = 3\sigma^4 f_i^4$ (the third and fourth moments of normals) the components of $Var(H_n(\theta))$ simplify to give

$$Var\left(\frac{\partial l}{\partial\theta}\right) = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i\left(\frac{\nabla f}{f}\right)_i^T + 2\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i\left(\frac{\nabla f}{f}\right)_i^T$$

$$= \left(\frac{1+2\sigma^2}{\sigma^2}\right)\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i\left(\frac{\nabla f}{f}\right)_i^T$$

$$Cov\left(\frac{\partial l}{\partial\theta},\frac{\partial l}{\partial\sigma}\right) = \frac{2}{\sigma}\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i$$

and $Var\left(\frac{\partial l}{\partial\sigma}\right) = \frac{2n}{\sigma^2}$.

So for this case, $E\left(-H_n'(\theta)\right) = Var\left(H_n(\theta)\right)$. Therefore, the variance covariance matrix of $(\hat{\theta},\hat{\sigma})$ reduces to

$$\left(E\left[-H_n'(\theta)\right]\right)^{-1} = \begin{bmatrix} \left(\frac{1+2\sigma^2}{\sigma^2}\right)\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i\left(\frac{\nabla f}{f}\right)_i^T & \frac{2}{\sigma}\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i \\ \frac{2}{\sigma}\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i^T & \frac{2n}{\sigma^2} \end{bmatrix}^{-1}.$$

Thus, the variance covariance matrix of $\hat{\theta}$ is

$$V(\hat{\theta}) = \left[\left(\frac{1+2\sigma^2}{\sigma^2}\right)\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i\left(\frac{\nabla f}{f}\right)_i^T - \frac{2}{n}\left(\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i\right)\left(\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i^T\right)\right]^{-1}.$$

Writing $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i = \overline{\left(\frac{\nabla f_0}{f_0}\right)}$ the above variance can be rewritten as

$$V(\hat{\theta}) = \left[\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i\left(\frac{\nabla f}{f}\right)_i^T + 2\left\{\sum_{i=1}^{n}\left[\left(\frac{\nabla f}{f}\right)_i - \overline{\left(\frac{\nabla f_0}{f_0}\right)}\right]\left[\left(\frac{\nabla f}{f}\right)_i - \overline{\left(\frac{\nabla f_0}{f_0}\right)}\right]^T\right\}\right]^{-1}. \tag{4.8}$$

Remarks

1. Observing that the term $\left\{\sum_{i=1}^{n}\left[\left(\frac{\nabla f}{f}\right)_i - \overline{\left(\frac{\nabla f_0}{f_0}\right)}\right]\left[\left(\frac{\nabla f}{f}\right)_i - \overline{\left(\frac{\nabla f_0}{f_0}\right)}\right]^T\right\}$ is positive definite, we note that the variance of the maximum likelihood estimator is smaller than the variance of the quasi-likelihood estimator, provided the errors have the same first four moments as the normal distribution.

2. If σ is small, ignoring the second term in Equation 4.8, we find that the above variance covariance matrix reduces to the variance covariance matrix derived in Chapter 3, based on small σ asymptotics.

4.2 Small σ large n behavior of the estimators

In this section, we analyze the large sample behavior of our small σ approximations to $\hat{\theta}-\theta_0$. In other words, we let $\sigma \longrightarrow 0$ first and then $n \longrightarrow \infty$. The more general problem in which

$n \longrightarrow \infty$ and $\sigma \longrightarrow 0$ simultaneously is not considered here. Another view of the material which follows is that we analyze the large sample behavior of C_1 and C_2 in our small σ expansions, $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2 + O(\sigma^3)$; see Chapter 3.

Behavior of C_1

Consider first the large sample behavior of C_1. Recall from Chapter 3 that C_1 is the same for all four methods of estimation. If our model has the correct mean structure, i.e., $E(Y_i) = f(x_i, \theta)$ then $E(C_1) = 0$. If in addition, $Var(Y_i) = \sigma^2 f^2(x_i, \theta)$ then $Var(C_1) = \Sigma$, where $\Sigma = \left[\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1}$. A natural and mild assumption is that as $n \longrightarrow \infty$,

$$\left[\frac{1}{n}\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1} \longrightarrow \Sigma_1(\theta),$$

where $\Sigma_1(\theta)$ is non-singular. In this case, neglecting terms of order σ^2, $\sqrt{n}(\hat{\theta} - \theta)$ has approximately a $N\left(0, \sigma^2\Sigma_1(\theta)\right)$ distribution.

Behavior of C_2

For quasi-likelihood we have

$$C_2 = \left[\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1} A, \tag{4.9}$$

where

$$\begin{aligned} A \;=\; & \sum_{i=1}^{n} \left(\frac{H_0}{f_0}\right)_i \epsilon_i C_1 - 2\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{\nabla f_0}{f_0}\right)_i^T C_1\epsilon_i - \sum_{i=1}^{n} C_1^T \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{H_0}{f_0}\right)_i C_1 \\ & + 2\sum_{i=1}^{n} C_1^T \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{\nabla f_0}{f_0}\right)_i^T C_1 \left(\frac{\nabla f_0}{f_0}\right)_i - \frac{1}{2}\sum_{i=1}^{n} C_1^T \left(\frac{H_0}{f_0}\right)_i C_1 \left(\frac{\nabla f_0}{f_0}\right)_i, \end{aligned}$$

and

$$E(C_2) = \Sigma\left\{-\frac{1}{2}\sum_{i=1}^{n} w_{2,i} \left(\frac{\nabla f_0}{f_0}\right)_i\right\}. \tag{4.10}$$

The weights $w_{2,i}$ are each linear combinations of the entries in Σ and so on the order, typically, of $\frac{1}{n}$. The matrix Σ is on the order $O(1/n)$ (see also Halbert [34] (page 15)). Thus, $E(C_2)$ will be $O(\frac{1}{n})$. Examining Equation 4.9 we see that for designs which keep $f(x_i, \theta_0)$ away from zero, a law of large numbers may be expected to apply to C_2 yielding $C_2 = O_p(1/n)$.

In the case of maximum likelihood, C_2 (see equation 3.25) contains the term

$$\Sigma\left\{\sum_{i=1}^{n}\epsilon_i^2\left(\frac{\nabla f_0}{f_0}\right)_i-\left(\sum_{i=1}^{n}\frac{\epsilon_i^2}{n}\right)\left(\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)_i\right)\right\},$$

which is $O_p(1/\sqrt{n})$. The expected value of C_2 for maximum likelihood is

$$E(C_2)=\Sigma\left\{-\sum_{i=1}^{n}(w_{1,i}-\frac{p}{n})\left(\frac{\nabla f_0}{f_0}\right)_i-\frac{1}{2}\sum_{i=1}^{n}w_{2,i}\left(\frac{\nabla f_0}{f_0}\right)_i\right\} \tag{4.11}$$

The weights $w_{1,i}$ are also linear combinations of the entries in Σ and are therefore on the order, typically, of $\frac{1}{n}$. Thus, $E(C_2)$ will be $O(\frac{1}{n})$.

Therefore, for quasi-likelihood and maximum likelihood estimators $E(\hat{\theta}-\theta_0)$ is $O(n^{-1})$, and $Var(\hat{\theta})=Var(\theta_0+C_1\sigma+C_2\sigma^2)=\Sigma\sigma^2+O(\sigma^4)$, which is $O(n^{-1})$. Consequently, for σ small and n large, the maximum likelihood estimator and the quasi-likelihood estimator are mean squared error consistent estimators for θ_0 (see Serfling [56]; technically we have established this only for a limit in which $\sigma\longrightarrow 0$ first.

Turning to generalized least squares and data weighted least squares the situation is somewhat different. In Equations 3.33 and 3.39 we see that for these estimators $E(C_2)$ contains terms such as those above in Equations 4.10 and 4.11 which are $O(1/n)$ but in addition a scalar multiple of the term $\Sigma\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)_i$ which appears to be $O(1)$.

For response functions of the form $f(x,\theta_1,\cdots,\theta_p)=\theta_1 f^*(\theta_2,\cdots,\theta_p)$, where f^* is some function that does not depend on θ_1, it is easy to see that

$$\left(\frac{\nabla f}{f}\right)_i=\begin{pmatrix}\frac{1}{\theta_1}\\ \left(\frac{\nabla f^*}{f^*}\right)_i\end{pmatrix}.$$

and

$$\Sigma^{-1}=\left[\sum_{i=1}^{n}\left(\frac{\nabla f_0}{f_0}\right)_i\left(\frac{\nabla f}{f}\right)_i^T\right]=\begin{bmatrix}\frac{n}{\theta_1^2} & \frac{1}{\theta_1}\sum_{i=1}^{n}\left(\frac{\nabla f^*}{f^*}\right)_i^T\\ \frac{1}{\theta_1}\sum_{i=1}^{n}\left(\frac{\nabla f^*}{f^*}\right)_i & \sum_{i=1}^{n}\left(\frac{\nabla f^*}{f^*}\right)_i\left(\frac{\nabla f^*}{f^*}\right)_i^T\end{bmatrix}.$$

Therefore,

$$\Sigma^{-1}\begin{pmatrix}\theta_1\\ 0\\ \vdots\\ 0\end{pmatrix}=\begin{pmatrix}\frac{n}{\theta_1}\\ \sum_{i=1}^{n}\left(\frac{\nabla f^*}{f^*}\right)_i\end{pmatrix}=\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i.$$

Thus,

$$\Sigma\left[\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right)_i\right] = \begin{pmatrix} \theta_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The crucial term in C_2 giving rise to the $O(1)$ component in $E(C_2)$ is $\Sigma\sum_{i=1}^{n}\epsilon_i^2\left(\frac{\nabla f_0}{f_0}\right)_i$, which will normally be $(\theta_1 \;\; 0\cdots 0)^T + O_p(1/\sqrt{n})$ (for ϵ_i's with four finite moments). Hence, to order σ^2, $\hat{\theta}_1$ is not asymptotically unbiased but all other entries in $\hat{\theta}$ are asymptotically unbiased.

For our model $p = 3$ and $\theta_1 = \alpha_1$. We thus have for generalized least squares the result

$$C_2 = \begin{pmatrix} \alpha_1 \\ 0 \\ 0 \end{pmatrix} + O_p(1/\sqrt{n}),$$

and for data weighted least squares

$$C_2 = \begin{pmatrix} -2\alpha_1 \\ 0 \\ 0 \end{pmatrix} + O_p(1/\sqrt{n}).$$

Hence, to the order σ^2, generalized least squares and data weighted least squares estimators for α_1 are not asymptotically unbiased but for α_2 and α_3 they are.

4.3 Small σ asymptotic behavior of the estimators in finite samples

For reasons we mentioned earlier, a comparison of the estimators in finite samples as $\sigma \to 0$ is of more practical value for thermoluminescence studies. Since for small σ, all four estimators have the same error of estimation, here we compare the biases of the estimators. A comparison of the biases using the derived formulae is not very obvious. Instead, we used the derived formulae to compute the biases of the estimators in some arbitrarily chosen

cases. The results obtained for a selected few cases are given in Tables 4.1-4.4. Similar results were observed for the other cases.

The results given in Tables 4.1 and 4.2 were observed setting the parameters at

$$\alpha_1 = 142800.7, \alpha_2 = 122.737, \alpha_3 = 391.9965 \text{ and } \sigma = 0.029.$$

The dose levels were fixed at (0,120,240,480,960). These are the quasi-likelihood estimates and the dose levels for the unbleached data set 'QNL84-2' given in Berger *et. al.* [12]. The first sample of size 10 was chosen by taking two replicates at each dose level. The rest of the samples were obtained by each time doubling the number of replicates at each dose level.

n	α_1				α_2			
	ML	QL	GLS	DWLS	ML	QL	GLS	DWLS
10	200.1	228.5	284.2	117.0	0.14	0.23	0.14	0.40
20	100.1	114.2	202.1	-61.6	0.07	0.11	0.07	0.20
40	50.0	57.1	161.1	-150.9	0.04	0.06	0.04	0.10
80	25.0	28.6	140.6	-195.5	0.02	0.03	0.02	0.05
160	12.5	14.3	130.3	-217.9	0.01	0.01	0.01	0.02

Table 4.1: Comparison of the biases of $\hat{\alpha}_1$ and $\hat{\alpha}_2$: Example 1

n	α_3			
	ML	QL	GLS	DWLS
10	1.49	1.78	1.49	2.36
20	0.75	0.89	0.75	1.17
40	0.37	0.44	0.37	0.59
80	0.19	0.22	0.19	0.29
160	0.09	0.11	0.09	0.15

Table 4.2: Comparison of the biases of $\hat{\alpha}_3$: Example 1

We repeated the above procedure setting the parameters at

$$\alpha_1 = 212138.3, \alpha_2 = 0.583, \alpha_3 = 5.964, \text{ and } \sigma = 0.029.$$

The dose levels were fixed at (0,1,2,4,8,16). These are the quasi-likelihood estimates and the dose levels for the unbleached data set 'STRB87-1' given in Ancient Thermoluminescence [12]. The results are given in Tables 4.3 and 4.4.

n	α_1				$\alpha_2 \times 10^{-6}$			
	ML	QL	GLS	DWLS	ML	QL	GLS	DWLS
12	109.55	141.04	243.36	-63.61	-1.3	274.2	-1.3	825.2
24	54.77	70.52	210.88	-210.21	-0.7	137.1	-0.7	412.6
48	27.39	35.26	194.65	-283.51	-0.3	68.5	-0.3	206.3
96	13.69	17.62	186.53	-320.16	-0.2	34.3	-0.2	103.2
192	6.85	8.81	182.45	-338.49	-0.1	17.1	-0.1	51.6

Table 4.3: Comparison of the biases of $\hat{\alpha}_1$ and $\hat{\alpha}_2$: Example 2

n	$\alpha_3 \times 10^{-3}$			
	ML	QL	GLS	DWLS
12	5.9	8.4	5.9	13.3
24	2.9	4.2	2.9	6.7
48	1.5	2.1	1.5	3.3
96	0.7	1.0	0.7	1.7
192	0.4	0.5	0.4	0.8

Table 4.4: Comparison of the biases of $\hat{\alpha}_3$: Example 2

Conclusions

Based on the results presented in Tables 4.1 - 4.4, we draw the following conclusions:

1. The biases of the maximum likelihood and the quasi-likelihood estimators for all three parameters converge to zero at a rate $O(1/n)$ as n is increased while σ is fixed.

2. The biases of generalized least squares and data weighted least squares estimators for the parameters α_2 and α_3 also converge to zero at a rate $O(1/n)$, as n is increased while σ is fixed.

3. For the parameters α_2 and α_3, the generalized least squares estimator and the maximum likelihood estimator have almost the same bias. (For the additive dose method α_2 is the parameter of interest.)

4. For the parameter α_1, the absolute values of the biases are in the order $B_{ML} < B_{QL} < B_{GLS}$. Depending on the sample size n, the absolute value of the bias of data weighted least squares estimator could be larger or smaller than the biases of the other estimators.

5. For parameters α_2 and α_3, the biases of the estimators in finite samples are in the order $B_{ML} \approx B_{GLS} \leq B_{QL} \leq B_{DWLS}$.

From the formulae derived in the last chapter, we notice that for all three parameters in the additive dose model, the bias of the data weighted least squares estimator (B_{DWLS}) is related to the biases of the quasi-likelihood estimator (B_{QL}) and the generalized least squares estimator (B_{GLS}) according to, $B_{DWLS} = -2\,B_{GLS} + 3\,B_{QL}$.

4.4 Discussion

In this chapter, we compared maximum likelihood, quasi-likelihood, generalized least squares and data weighted least squares estimators for our model. In Section 4.1, we examined the behavior of the estimators in large samples, for fixed σ. We found that maximum likelihood and quasi-likelihood estimators are consistent while generalized least squares and data weighted least squares estimators are generally not. Least squares estimators in general were found to have biases that do not vanish, even asymptotically. We also examined the distributional approximations for the maximum likelihood and quasi-likelihood estimators.

We found that in large samples both estimators are approximately normally distributed; the approximate asymptotic variance of the maximum likelihood estimator was found to be smaller than that of the quasi-likelihood estimator.

In Section 4.2, we analyzed the large sample behavior of our small σ approximations to $\hat{\theta}$, namely $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2 + O(\sigma^3)$. For quasi-likelihood, the terms C_1 and C_2 were found to be of the order $O_p(1/\sqrt{n})$ and $O_p(1/n)$ respectively. Recall that C_1 is a linear combination of the random errors ϵ and C_2 is a quadratic term in ϵ. For small σ, the quasi-likelihood estimator $\hat{\theta}$ is approximately normally distributed if n is large enough or σ is small enough so that $\sigma/\sqrt{n}$ is small. For maximum likelihood, both C_1 and C_2 were found to be of the order $O_p(1/\sqrt{n})$. Therefore, the maximum likelihood estimator is approximately normally distributed if σ is sufficiently small. For generalized least squares and data weighted least squares, the terms C_1 and C_2 are in general of the order $O_p(1/\sqrt{n})$ and $O_p(1)$ respectively. However, as we clarified in Section 4.2, for our model components of C_1 and C_2 corresponding to α_2 and α_3 were found to be of the order $O_p(1/\sqrt{n})$ and $O_p(1/\sqrt{n})$ respectively. Therefore, for the limiting case of small σ and large n, generalized least squares and data weighted least squares were also found to produce asymptotically unbiased estimators for α_2 and α_3. The discussion presented in Section 4.2 is valid for more general response functions than simply the response functions described for the additive dose method, the partial bleach method and the regeneration method.

In Section 4.3, we used the formulae derived in the limit of small σ to examine the behavior of the biases of the estimators as sample size grows. For sensible designs for the additive dose model we found that:

1. The biases of maximum likelihood and quasi-likelihood estimators for all three parameters converge to zero at a rate $O(1/n)$.

2. The biases of generalized least squares and data weighted least squares estimators for the parameters α_2 and α_3 also converge to zero at a rate $O(1/n)$.

3. Maximum likelihood and generalized least squares estimators for α_2 (the parameter of interest in thermoluminescence studies) and α_3 have almost the same bias, for sample

sizes used in practice.

4. For the parameter α_1, the absolute values of the biases are in the order $B_{ML} < B_{QL} < B_{GLS}$. For the parameter α_1, the absolute value of the bias of data weighted least squares estimator could be larger or smaller than the biases of the other three estimators depending on the sample size n.

In the limit of small σ the biases of all four estimators were found to be negligible relative to their standard errors (see Chapter 3). While all four estimators were found to perform well for σ values and sample sizes used in practice, we favour using the quasi-likelihood estimator since it has the advantage that it does not require any assumptions about the distribution of the data other than those about the first two moments. Also we found that the algorithms that solve quasi-likelihood estimating equations converge faster than those for maximum likelihood and generalized least squares estimators.

Chapter 5

Equivalent dose from partial bleach data

In the partial bleach method introduced in Chapter 2, two data sets are collected from each core; these define the unbleached and bleached dose response curves. These response curves are nonlinear, and can be fitted using the techniques discussed in the previous chapters. The equivalent dose is defined as the dose corresponding to the intersection point of the unbleached and bleached response curves. In this chapter, we discuss estimating the equivalent dose from partial bleach data.

An initial estimate for the equivalent dose is suggested in Section 5.1. We discuss two approaches for estimating the equivalent dose. The first approach fits the unbleached and bleached response curves separately using the techniques described in Chapter 3, and then finds the equivalent dose as the intersection point of the two fitted response curves. We refer to this procedure as the 'two stage approach'. This procedure is described in Section 5.2. For this case, Berger *et.al.*[11] describe the error analysis assuming that a single error factor (that is, a common value of σ) describes both unbleached and bleached response curves. Their method of error analysis and the construction of confidence intervals for the equivalent dose are described in Subsection 5.2.1. In Subsection 5.2.1, we extend these ideas to the case of different error factors for the unbleached and bleached data.

We discuss maximum likelihood, quasi-likelihood, generalized least squares and data

weighted least squares estimators for the equivalent dose. Formulae for the biases of these estimators are derived in Subsection 5.2.2. In Subsection 5.2.3, we examine the bias from a Monte Carlo study. A theoretical justification for the suggested confidence intervals is offered in Subsection 5.2.4. In Subsection 5.2.5, we describe a simulation study that examines the finite sample performance of the asymptotic confidence intervals.

The second method estimates the equivalent dose by fitting the unbleached and bleached curves simultaneously. We refer to this procedure as 'simultaneous curve fitting'. In this setting, the equivalent dose is treated as a parameter that appears explicitly in the estimation procedure. In Section 5.3.1, we discuss obtaining maximum likelihood estimates for the equivalent dose. Under maximum likelihood, we discuss computing profile likelihood intervals and symmetric confidence intervals using Z and t critical values. We also discuss obtaining confidence intervals based on a transformation of the likelihood ratio statistic using a transformed F critical value. Use of the t and F critical values is justified in the limit of small σ. In Section 5.3.7, we discuss computing quasi-likelihood estimates. Under quasi-likelihood, we discuss computing confidence intervals by inverting the quasi-score test. We also discuss computing symmetric confidence intervals based on the quasi-likelihood estimate with a t quantile. In Section 5.3.10, we obtain generalized least squares and data weighted least squares estimates for the equivalent dose. Symmetric confidence intervals based on the least squares estimates are computed with a t critical value. The coverage probabilities of the suggested type confidence intervals are examined by a Monte Carlo study. Section 5.6 summarizes the chapter.

5.1 Initial estimates

Suppose saturating exponential models[1] are suitable for both unbleached and bleached response curves. Let $\theta_1^T = (\alpha_1, \alpha_2, \alpha_3)$, and $\theta_2^T = (\beta_1, \beta_2, \beta_3)$ be the parameter vectors corresponding to these saturating exponential models.

[1]Saturating exponential models were introduced in Chapter 2.

We use the following notation:

$$
\begin{aligned}
y_{1i} &= \text{photon count from the } i\text{th sample of the unbleached data,} \\
y_{2i} &= \text{photon count from the } i\text{th sample of the bleached data,} \\
x_{1i} &= \text{dose applied to the } i\text{th sample of the unbleached data,} \\
x_{2i} &= \text{dose applied to the } i\text{th sample of the bleached data,} \\
-\gamma &= \text{the equivalent dose,} \\
\theta^T &= (\theta_1^T, \theta_2^T) = (\alpha_1, \alpha_2, \alpha_3, \beta_2, \beta_3, \gamma), \\
f_{1i} &= f(x_{1i}, \theta_1) = \alpha_1 \left\{ 1 - \exp\left[-\frac{x_{1i} + \alpha_2}{\alpha_3} \right] \right\}, \\
f_{2i} &= f(x_{2i}, \theta_2) = \beta_1 \left\{ 1 - \exp\left[-\frac{x_{2i} + \beta_2}{\beta_3} \right] \right\}, \\
\text{and } R &= (f_1(x, \theta_1) - f_2(x, \theta_2)).
\end{aligned}
$$

At low dose values, the thermoluminescence vs added dose is roughly linear. Therefore, the difference R between the unbleached and bleached response curves varies roughly linearly with the added dose at low dose values so that approximately, $R = mx + C$, where C and m respectively denote the intercept and the slope of the straight line that describes the relationship. We need to find the dose corresponding to the intersection of the unbleached and bleached response curves. This is the dose corresponding to $R = 0$, which is the absolute value of $-\frac{C}{m}$. We estimate C and m using the average photon counts corresponding to the zero dose and the next smallest dose common to both unbleached and bleached data sets. Thus, an initial estimate for the equivalent dose is given by

$$
-\hat{\gamma}_0 = \left| \frac{(\overline{y}_1(0) - \overline{y}_2(0))}{[(\overline{y}_1(d) - \overline{y}_2(d)) - (\overline{y}_1(0) - \overline{y}_2(0))]} d \right|,
$$

where

$$
\begin{aligned}
\overline{y}_1(0) &= \text{average photon count corresponding to the zero dose for the unbleached data,} \\
\overline{y}_2(0) &= \text{average photon count corresponding to the zero dose for the bleached data,} \\
d &= \text{smallest positive dose value common to both data sets} \\
\overline{y}_1(d) &= \text{average photon count corresponding to the dose } d \text{ for the unbleached data,}
\end{aligned}
$$

and

$\overline{y}_2(d)$ = average photon count corresponding to the dose d for the bleached data.

For the data sets we analyzed (test data from Berger *et. al.* [12]), the initial estimates from the above formula served as good starting values for the algorithm we describe next.

5.2 Estimation from a two stage approach

The equivalent dose γ is a root of the equation $g(x, \theta_1, \theta_2) = f_1(x, \theta_1) - f_2(x, \theta_2) = 0$. The equivalent dose is estimated by $\hat{\gamma}$, which is a root of the equation $g(x, \hat{\theta_1}, \hat{\theta_2}) = f_1(x, \hat{\theta_1}) - f_2(x, \hat{\theta_2}) = 0$ (Berger *et. al.* [11]). First we find the estimates $\hat{\theta_1}$ and $\hat{\theta_2}$ using the procedures described in Chapter 3. Then we solve the nonlinear equation $g = 0$ using the Newton Raphson algorithm described below.

1. Find an initial estimate for the equivalent dose (Section 5.1).

2. At the $(k+1)$st step of the iteration the equivalent dose is estimated as

$$\hat{\gamma}_{k+1} = \hat{\gamma}_k - \frac{g|_{\hat{\gamma}_k}}{\left(\frac{\partial g}{\partial x}\right)\Big|_{\hat{\gamma}_k}},$$

   where

$$g|_{\hat{\gamma}_k} = \hat{\alpha}_1 \left\{1 - \exp\left[-\frac{(\hat{\gamma}_k + \hat{\alpha}_2)}{\hat{\alpha}_3}\right]\right\} - \hat{\beta}_1 \left\{1 - \exp\left[-\frac{(\hat{\gamma}_k + \hat{\beta}_2)}{\hat{\beta}_3}\right]\right\}, \qquad (5.1)$$

   and

$$\frac{\partial g}{\partial x}\Big|_{\hat{\gamma}_k} = \frac{\hat{\alpha}_1}{\hat{\alpha}_3} \exp\left[-\frac{(\hat{\gamma}_k + \hat{\alpha}_2)}{\hat{\alpha}_3}\right] - \frac{\hat{\beta}_1}{\hat{\beta}_3} \exp\left[-\frac{(\hat{\gamma}_k + \hat{\beta}_2)}{\hat{\beta}_3}\right]. \qquad (5.2)$$

3. Iterate until desired convergence. In the software we developed, the stopping criteria was taken as when the absolute difference in the successive iterations is less than 10^{-5}.

An estimate for the error of the estimate

As above, γ and $\hat{\gamma}$ satisfy the respective equations $g(\gamma, \theta_1, \theta_2) = 0$, and $g(\hat{\gamma}, \hat{\theta_1}, \hat{\theta_2}) = 0$. The first order Taylor expansion of $g(\hat{\gamma}, \hat{\theta_1}, \hat{\theta_2})$ around $(\gamma, \theta_1, \theta_2)$ yields

$$g(\hat{\gamma}, \hat{\theta_1}, \hat{\theta_2}) \approx g(\gamma, \theta_1, \theta_2) + \frac{\partial g}{\partial \gamma}(\hat{\gamma} - \gamma) + \left(\frac{\partial g}{\partial \theta_1}\right)^T (\hat{\theta_1} - \theta_1) + \left(\frac{\partial g}{\partial \theta_2}\right)^T (\hat{\theta_2} - \theta_2).$$

In this equation each partial derivative term indicates partial differentiation of g with respect to the estimated parameters and then evaluation of the resulting quantities at the respective true parameter values, and at dose $x = \gamma$. Since $g(\gamma, \theta_1, \theta_2) = 0$, and $g(\hat{\gamma}, \hat{\theta}_1, \hat{\theta}_2) = 0$ the above equation gives

$$(\hat{\gamma} - \gamma) = \left[-\left(\frac{\partial f_1}{\partial \theta_1}\right)^T (\hat{\theta}_1 - \theta_1) + \left(\frac{\partial f_2}{\partial \theta_2}\right)^T (\hat{\theta}_2 - \theta_2) \right] / (\frac{\partial g}{\partial \gamma}). \tag{5.3}$$

Therefore,

$$Var(\hat{\gamma}) = \left[\left(\frac{\partial f_1}{\partial \theta_1}\right)^T Cov(\hat{\theta}_1) \left(\frac{\partial f_1}{\partial \theta_1}\right) + \left(\frac{\partial f_2}{\partial \theta_2}\right)^T Cov(\hat{\theta}_2) \left(\frac{\partial f_2}{\partial \theta_2}\right) \right] / \left[\frac{\partial f_1}{\partial \gamma} - \frac{\partial f_2}{\partial \gamma} \right]^2. \tag{5.4}$$

In equations 5.3 and 5.4, the derivatives of f_1 and f_2 are to be evaluated at $x = \gamma$. Here onwards we use an additional suffix i to indicate when functions f_1 and f_2 are evaluated at dose level x_i. When the functions are evaluated at $x = \gamma$ we suppress this additional suffix. The sample sizes of unbleached and bleached data sets will be denoted by n_1 and n_2 respectively. The error factors for the unbleached and bleached data will be denoted by σ_1 and σ_2 respectively. When the unbleached and bleached response curves correspond to a common error factor (i.e. when $\sigma_1 = \sigma_2$), this common error factor will be denoted by σ.

From the formulae derived in Chapter 3, the variance covariance matrices for $\hat{\theta}_1, \hat{\theta}_2$ are

$$Cov(\hat{\theta}_1) = \sigma_1^2 \left[\sum_{i=1}^{n} \left(\frac{\nabla f_1}{f_1}\right)_i \left(\frac{\nabla f_1}{f_1}\right)_i^T \right]^{-1}, \tag{5.5}$$

and

$$Cov(\hat{\theta}_2) = \sigma_2^2 \left[\sum_{i=1}^{n} \left(\frac{\nabla f_2}{f_2}\right)_i \left(\frac{\nabla f_2}{f_2}\right)_i^T \right]^{-1}. \tag{5.6}$$

The vector ∇f_j, $(j = 1, 2)$ denotes the gradient vector of f_i with respect to the components of θ. In the different error factor case, we estimate σ_1^2 and σ_2^2 by

$$\hat{\sigma_1}^2 = \sum_{j=1}^{n_1} \left[(y_{1j} - \hat{f}_1)/\hat{f}_1 \right]^2 / (n_1 - 3), \tag{5.7}$$

and

$$\hat{\sigma_2}^2 = \sum_{k=1}^{n_2} \left[(y_{2k} - \hat{f}_2)/\hat{f}_2 \right]^2 / (n_2 - 3). \tag{5.8}$$

In Section 5.2.4, we show that in the limit of $\sigma_1, \sigma_2 \longrightarrow 0$, the $\hat{\sigma}_1$ and $\hat{\sigma}_2$ suggested above are approximately unbiased for σ_1 and σ_2 respectively. If the two curves correspond to a common σ we estimate σ from

$$\hat{\sigma}^2 = \left\{ \sum_{i=1}^{n_1} \left[(y_{1i} - \hat{f}_1)/\hat{f}_1 \right]^2 + \sum_{j=1}^{n_2} \left[(y_{2j} - \hat{f}_2)/\hat{f}_2 \right]^2 \right\} /(n_1 + n_2 - 6).$$

In Section 5.2.4, we show that the above $\hat{\sigma}^2$ is an approximately unbiased estimate for σ^2.

To estimate the standard error of $\hat{\gamma}$ we use Equations 5.4 and 5.5 by replacing the unknown parameters by their estimates.

5.2.1 Confidence intervals

Let $\hat{\gamma}$ and $s^2_{\hat{\gamma}}$ be the estimates for γ and $Var(\hat{\gamma})$ discussed in Section 5.2.

Single error factor

In Section 5.2.4, we show that if the response curves correspond to a common error factor σ, in the limiting case as $\sigma \longrightarrow 0$, an approximate $100(1-\alpha)\%$ confidence interval for γ can be constructed by taking, $\hat{\gamma} \mp t_{(n_1+n_2-6),\alpha/2} s_{\hat{\gamma}}$, as lower and upper confidence limits. Here $t_{\nu,\alpha}$ indicates the upper α quantile of a t-distribution with ν degrees of freedom.

Different error factors

Suppose the two response curves correspond to different error factors σ_1 and σ_2. In Section 5.2.4, using Satterthwaite's approximation [53], we show that when $\sigma_1, \sigma_2 \longrightarrow 0$, the distribution of $\frac{\hat{\gamma}-\gamma}{s_{\hat{\gamma}}}$ can be approximated by a t distribution and provide a formula (Equation 5.17) for the degrees of freedom of the approximate t distribution. Therefore, we propose computing confidence intervals with confidence coefficient α by taking $\hat{\gamma} \mp t_{df,\alpha/2} s_{\hat{\gamma}}$ as lower and upper confidence limits. The $t_{df,\alpha}$ indicates the upper α th quantile of a t-distribution with df degrees of freedom.

It is important to note that for the common error case, we compute confidence intervals based on an exact t distribution for the test statistic $\frac{\hat{\gamma}-\gamma}{s_{\hat{\gamma}}}$ valid in the limit of small σ. In

contrast, for the different error factor case, we compute confidence intervals based on an approximate t distribution valid in the limiting case $\sigma_1, \sigma_2 \longrightarrow 0$.

5.2.2 Bias in the estimator for the equivalent dose

Notation:

Let $\theta^T = (\theta_1^T, \theta_2^T)$, $g(\gamma, \theta) = f_1(\gamma, \theta_1) - f_2(\gamma, \theta_2)$, and

$C_1^T = (C_{11}^T, C_{12}^T)$, where

$$\begin{aligned} C_{11} &= \left[\sum_{i=1}^{n_1} \left(\frac{\nabla f_1(x_{1i}, \theta_1)}{f_1(x_{1i}, \theta_1)}\right)\left(\frac{\nabla f_1(x_{1i}, \theta_1)}{f_1(x_{1i}, \theta_1)}\right)^T\right]^{-1} \sum_{i=1}^{n_1}\left[\frac{\nabla f_1(x_{1i}, \theta_1)}{f_1(x_{1i}, \theta_1)}\right] \\ &= \Sigma_1 \sum_{i=1}^{n_1}\left[\frac{\nabla f_1(x_{1i}, \theta_1)}{f_1(x_{1i}, \theta_1)}\right], \end{aligned} \tag{5.9}$$

and

$$\begin{aligned} C_{12} &= \left[\sum_{i=1}^{n_2} \left(\frac{\nabla f_2(x_{2i}, \theta_2)}{f_2(x_{2i}, \theta_2)}\right)\left(\frac{\nabla f_2(x_{2i}, \theta_2)}{f_2(x_{2i}, \theta_2)}\right)^T\right]^{-1} \sum_{j=1}^{n_2}\left[\frac{\nabla f_2(x_{2j}, \theta_2)}{f_2(x_{2j}, \theta_2)}\right] \\ &= \Sigma_2 \sum_{j=1}^{n_2}\left[\frac{\nabla f_2(x_{2j}, \theta_2)}{f_2(x_{2j}, \theta_2)}\right]. \end{aligned} \tag{5.10}$$

Similarly, let C_{21}^T and C_{22}^T be the vectors C_2 introduced in Chapter 3 for the curves defined by $f_1(x_1, \theta_1)$ and $f_2(x_2, \theta_2)$. Let $C_2^T = (C_{21}^T, C_{22}^T)$. For small σ, $\hat{\theta}$ and $\hat{\gamma}$ will have Taylor expansions of the form

$$(\hat{\theta} - \theta) = C_1\sigma + C_2\sigma^2, \tag{5.11}$$

and

$$(\hat{\gamma} - \gamma) = C_3\sigma + C_4\sigma^2. \tag{5.12}$$

Here C_3 and C_4 are scalar quantities that do not depend on σ. We now obtain C_3 and C_4 in terms of C_1 and C_2 so that we can evaluate the small σ behavior of $\hat{\gamma}$.

The equivalent dose γ and the suggested estimate $\hat{\gamma}$ satisfy the equations $g(\gamma, \theta) = 0$ and $g(\hat{\gamma}, \hat{\theta}) = 0$ respectively. Consider the Taylor expansion

$$\begin{aligned} g(\hat{\gamma}, \hat{\theta}) &= g(\gamma, \theta) + (\hat{\gamma} - \gamma)\left(\frac{\partial g}{\partial \gamma}\right) + (\hat{\theta} - \theta)^T\left(\frac{\partial g}{\partial \theta}\right) + \frac{1}{2}(\hat{\gamma} - \gamma)^T\frac{\partial^2 g}{\partial \gamma^2}(\hat{\gamma} - \gamma) \\ &\quad + \frac{1}{2}(\hat{\theta} - \theta)^T\frac{\partial^2 g}{\partial \theta \partial \theta^T}(\hat{\theta} - \theta) + \frac{1}{2}(\hat{\theta} - \theta)^T\frac{\partial^2 g}{\partial \gamma \partial \theta^T}(\hat{\gamma} - \gamma). \end{aligned} \tag{5.13}$$

Equations (5.11) to (5.13) give

$$\begin{aligned} 0 &= (C_3\sigma + C_4\sigma^2)\left(\frac{\partial g}{\partial \gamma}\right) + (C_1\sigma + C_2\sigma^2)^T\left(\frac{\partial g}{\partial \theta}\right) + \frac{1}{2}(C_3\sigma + C_4\sigma^2)^T\frac{\partial^2 g}{\partial \gamma^2}(C_3\sigma + C_4\sigma^2) \\ &\quad + \frac{1}{2}(C_1\sigma + C_2\sigma^2)^T\frac{\partial^2 g}{\partial \theta \partial \theta^T}(C_1\sigma + C_2\sigma^2) + \frac{1}{2}(C_1\sigma + C_2\sigma^2)^T\frac{\partial^2 g}{\partial \gamma \partial \theta^T}(C_3\sigma + C_4\sigma^2). \end{aligned}$$

Equating the coefficients of powers of σ gives, assuming $\left(\frac{\partial g}{\partial \gamma}\right) \neq 0$,

$$C_3 = -\frac{1}{\left(\frac{\partial g}{\partial \gamma}\right)}C_1^T\left(\frac{\partial g}{\partial \theta}\right), \tag{5.14}$$

and

$$\begin{aligned} C_4 &= -\frac{1}{\left(\frac{\partial g}{\partial \gamma}\right)}C_2^T\left(\frac{\partial g}{\partial \theta}\right) - \frac{1}{2\left(\frac{\partial g}{\partial \gamma}\right)^3}\left[C_1^T\left(\frac{\partial g}{\partial \theta}\right)\frac{\partial g}{\partial \theta^T}C_1\right] \\ &\quad + \frac{1}{2\left(\frac{\partial g}{\partial \gamma}\right)^2}\frac{\partial g}{\partial \theta^T}C_1\frac{\partial^2 g}{\partial \gamma \partial \theta^T}C_1 - \frac{1}{2\left(\frac{\partial g}{\partial \gamma}\right)}\left[C_1^T\frac{\partial^2 g}{\partial \theta \partial \theta^T}C_1\right]. \end{aligned} \tag{5.15}$$

Let $B_{\hat{\theta}_1}$ and $B_{\hat{\theta}_2}$ denote the biases of $\hat{\theta}_1$ and $\hat{\theta}_2$ respectively (these will depend on which method of estimation is being discussed). It is easy to see that

$$E\left[C_2^T\left(\frac{\partial g}{\partial \theta}\right)\right]\sigma^2 = \left[\frac{\partial f_1}{\partial \theta_1^T}B_{\hat{\theta}_1} - \frac{\partial f_2}{\partial \theta_2^T}B_{\hat{\theta}_2}\right].$$

Since $C_1^T\left(\frac{\partial g}{\partial \theta}\right)$ is a scalar, we find

$$\begin{aligned} E\left[C_1^T\left(\frac{\partial g}{\partial \theta}\right)\frac{\partial g}{\partial \theta^T}C_1\right] &= E\left[\frac{\partial g}{\partial \theta^T}C_1C_1^T\left(\frac{\partial g}{\partial \theta}\right)\right] \\ &= \frac{\partial g}{\partial \theta^T}E(C_1C_1^T)\left(\frac{\partial g}{\partial \theta}\right) \\ &= \frac{\partial g}{\partial \theta^T}\Sigma\left(\frac{\partial g}{\partial \theta}\right). \end{aligned}$$

where

$$\begin{aligned} \Sigma &= \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}, \\ \Sigma_1 &= \left[\sum_{i=1}^{n_1}\left(\frac{\nabla f_1}{f_1}\right)_i\left(\frac{\nabla f_1}{f_1}\right)_i^T\right]^{-1} \end{aligned}$$

and

$$\Sigma_2 = \left[\sum_{i=1}^{n_1} \left(\frac{\nabla f_2}{f_2}\right)_i \left(\frac{\nabla f_2}{f_2}\right)_i^T\right]^{-1}.$$

Further note that

$$E\left[C_1 \frac{\partial^2 g}{\partial\gamma\partial\theta^T} C_1\right] = E\left[C_1 C_1^T \frac{\partial^2 g}{\partial\gamma\partial\theta}\right] = \Sigma \frac{\partial^2 g}{\partial\gamma\partial\theta},$$

and

$$\begin{aligned} E\left[C_1^T \frac{\partial^2 g}{\partial\theta\partial\theta^T} C_1\right] &= E\left[tr\left(C_1^T \frac{\partial^2 g}{\partial\theta\partial\theta^T} C_1\right)\right] \\ &= E\left[tr\left(\frac{\partial^2 g}{\partial\theta\partial\theta^T} C_1 C_1^T\right)\right] \\ &= tr\left(\frac{\partial^2 g}{\partial\theta\partial\theta^T}\Sigma\right). \end{aligned}$$

Note that $E(C_3) = 0$ and therefore the bias in $\hat{\gamma}$ is

$$\begin{aligned} B_{\hat{\gamma}} &= E(C_4)\,\sigma^2 \\ &= -\frac{1}{\left(\frac{\partial g}{\partial\gamma}\right)}\left[\frac{\partial f_1}{\partial\theta_1^T} B_{\hat{\theta}_1} - \frac{\partial f_2}{\partial\theta_2^T} B_{\hat{\theta}_2}\right] - \frac{\sigma^2}{2\left(\frac{\partial g}{\partial\gamma}\right)^3} \frac{\partial^2 g}{\partial\gamma^2} \frac{\partial g}{\partial\theta^T} \Sigma \left(\frac{\partial g}{\partial\theta}\right) \\ &\quad + \frac{\sigma^2}{2\left(\frac{\partial g}{\partial\gamma}\right)^2} \frac{\partial g}{\partial\theta^T} \Sigma \frac{\partial^2 g}{\partial\gamma\partial\theta} - \frac{\sigma^2}{2\left(\frac{\partial g}{\partial\gamma}\right)} tr\left(\frac{\partial^2 g}{\partial\theta\partial\theta^T}\Sigma\right). \end{aligned} \tag{5.16}$$

To simplify computation of (5.16) we provide explicit formulae for the pieces thereof. The derivatives of g are given by

$$\begin{aligned} \frac{\partial g}{\partial\gamma} &= \left.\frac{\partial f_1}{\partial x}\right|_{x=\gamma} - \left.\frac{\partial f_2}{\partial x}\right|_{x=\gamma} \\ &= \frac{\alpha_1}{\alpha_3}\exp(-(\gamma+\alpha_2)/\alpha_3) - \frac{\beta_1}{\beta_3}\exp(-(\gamma+\beta_2)/\beta_3), \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2 g}{\partial\gamma^2} &= \left.\frac{\partial^2 f_1}{\partial x^2}\right|_{x=\gamma} - \left.\frac{\partial^2 f_2}{\partial x^2}\right|_{x=\gamma} \\ &= -\frac{\alpha_1}{\alpha_3^2}\exp(-(\gamma+\alpha_2)/\alpha_3) + \frac{\beta_1}{\beta_3^2}\exp(-(\gamma+\beta_2)/\beta_3), \end{aligned}$$

and by

$$\left(\frac{\partial g}{\partial \theta^T}\right)_{1\times 6} = \left(\frac{\partial f_1(\gamma,\theta_1)}{\partial \theta_1^T}, \frac{\partial f_2(\gamma,\theta_2)}{\partial \theta_2^T}\right) = (\partial_1^{(1)}, \partial_2^{(1)}, \partial_3^{(1)}, \partial_4^{(1)}, \partial_5^{(1)}, \partial_6^{(1)})$$

where

$$\begin{aligned}
\partial_1^{(1)} &= 1 - \left\{\exp\left[-\left(\frac{\gamma+\alpha_2}{\alpha_3}\right)\right]\right\}, \\
\partial_2^{(1)} &= \frac{\alpha_1}{\alpha_3}\exp\left[-\left(\frac{\gamma+\alpha_2}{\alpha_3}\right)\right], \\
\partial_3^{(1)} &= -\frac{\alpha_1}{\alpha_3}(\gamma+\alpha_2)\exp\left[-\left(\frac{\gamma+\alpha_2}{\alpha_3}\right)\right], \\
\partial_4^{(1)} &= -\left\{1 - \left\{\exp\left[-\left(\frac{\gamma+\beta_2}{\beta_3}\right)\right]\right\}\right\}, \\
\partial_5^{(1)} &= -\frac{\beta_1}{\beta_3}\exp\left[-\left(\frac{\gamma+\beta_2}{\beta_3}\right)\right], \\
\partial_6^{(1)} &= \frac{\beta_1}{\beta_3}(\gamma+\beta_2)\exp\left[-\left(\frac{\gamma+\beta_2}{\beta_3}\right)\right]
\end{aligned}$$

and by

$$\left(\frac{\partial^2 g}{\partial\gamma\partial\theta}\right)_{1\times 6} = (\partial_1^{(2)}, \partial_2^{(2)}, \partial_3^{(2)}, \partial_4^{(2)}, \partial_5^{(2)}, \partial_6^{(2)}),$$

where

$$\begin{aligned}
\partial_1^{(2)} &= \frac{1}{\alpha_3}\exp\left[-\left(\frac{\gamma+\alpha_2}{\alpha_3}\right)\right], \\
\partial_2^{(2)} &= -\frac{\alpha_1}{\alpha_3^2}\exp\left[-\left(\frac{\gamma+\alpha_2}{\alpha_3}\right)\right], \\
\partial_3^{(2)} &= -\frac{\alpha_1}{\alpha_3^2}\exp\left[-\left(\frac{\gamma+\alpha_2}{\alpha_3}\right)\right] + \frac{\alpha_1}{\alpha_3^3}(\gamma+\alpha_2)\exp\left[-\left(\frac{\gamma+\alpha_2}{\alpha_3}\right)\right], \\
\partial_4^{(2)} &= -\frac{1}{\beta_3}\exp\left[-\left(\frac{\gamma+\beta_2}{\beta_3}\right)\right], \\
\partial_5^{(2)} &= \frac{\beta_1}{\beta_3^2}\exp\left[-\left(\frac{\gamma+\beta_2}{\beta_3}\right)\right], \\
\partial_6^{(2)} &= \frac{\beta_1}{\beta_3^2}\exp\left[-\left(\frac{\gamma+\beta_2}{\beta_3}\right)\right] - \frac{\beta_1}{\beta_3^3}(\gamma+\beta_2)\exp\left[-\left(\frac{\gamma+\beta_2}{\beta_3}\right)\right].
\end{aligned}$$

Finally we note

$$tr\left(\frac{\partial^2 g}{\partial\theta\partial\theta^T}\Sigma\right) = tr\,(H_1\Sigma_1) - tr\,(H_2\Sigma_2),$$

where $H_1 = \frac{\partial^2 f_1(\gamma,\theta_1)}{\partial\theta_1\partial\theta_1^T}$ $H_2 = \frac{\partial^2 f_2(\gamma,\theta_2)}{\partial\theta_2\partial\theta_2^T}$ are the Hessian matrices of f_1 and f_2 (see Section 3.2.), evaluated at $x = \gamma$.

The biases $B_{\hat{\theta_1}}$ and $B_{\hat{\theta_2}}$ can be estimated from the formulae given in Table 3.1 of Chapter 3. Thus, an estimate for the bias of $\hat{\gamma}$, $\hat{B}_\gamma$, can be obtained by replacing the unknown parameters in the Equation (5.16) by their estimates.

An approximate $100(1-\alpha)\%$ confidence interval corrected for the bias can therefore be constructed by taking $(\hat{\gamma} - \hat{B}_\gamma) \mp t_{df,\alpha/2}\sqrt{\hat{Var}(\hat{\gamma})}$ as lower and upper confidence limits. For the single error factor case, *df*, which is the degrees of freedom for the t-distribution is $(n_1 + n_2 - 6)$. For the case of different error factors *df* is given by Equation (5.17) of Section (5.2.4).

5.2.3 Examination of the bias from a Monte Carlo study

The formula for the bias derived in Section 5.2.2 is based on the approximation that for small σ, $\hat{\gamma} = \gamma + C_3\sigma + C_4\sigma^2$, for coefficients C_3 and C_4 that do not depend on σ. We examined the validity of the derived formula for σ values in the range of typical TL data sets by a Monte Carlo study. Next we describe the Monte Carlo study.

1. The parameter vectors for the two response curves were chosen so that they intersect at γ. This was achieved by fixing the parameters $\alpha_1, \alpha_2, \alpha_3, \beta_2, \beta_3$ at desired values and then taking, $\beta_1 = \alpha_1 \frac{[1-\exp(-(\gamma+\alpha_2)/\alpha_3)]}{[1-\exp(-(\gamma+\beta_2)/\beta_3)]}$.

2. The parameter values were set at $\alpha_1 = 14.2853, \alpha_2 = 123.182, \alpha_3 = 393.065, \beta_2 = 192.547, \beta_3 = 756.620$ and $\gamma = -87.45$. These are the maximum likelihood estimates for the unbleached data set 'QNL84-2' given in Berger *et. al.* [12].

3. The dose vectors of lengths n_1 and n_2 chosen for the unbleached and bleached curves are presented in Table 9.1 of Appendix 9.3 where they are labeled as $P1U$ and $P1B$.

4. For each study, two sets of random variates of sizes n_1 and n_2 were generated from the standard normal distribution. Let these values be denoted by, ϵ_{ji}, $j = 1, 2$ and $i = 1, \ldots, n_j$.

5. The relative error in a single measurement, σ, was fixed at the values given in Table 5.1.

6. The photon counts, y_{ji}'s, were then simulated using $y_{ji} = f(x_i, \theta_j)(1 + \sigma\epsilon_{ji})$, for $j = 1, 2$ and $i = 1, \ldots, n_j$.

7. The algorithm described in Section 5.2 was used to estimate the equivalent dose γ.

8. From Equation (5.14), the value of C_3 was computed at the true parameter values used in the study.

9. Estimates for the biases were obtained as the averages of the m values for $\hat{\gamma} - \gamma - C_3\sigma$ and $(\hat{\gamma} - \gamma)$. (m is the size of the study.)

The results based on 10000 simulations are presented in Table 5.1, where we use the notation

$$\begin{aligned} B_T &= \text{True bias (Equation 5.16)} \\ B_1 &= \text{average of } \hat{\gamma} - \gamma - C_3\sigma \text{ values} \\ B_2 &= \text{average of } \hat{\gamma} - \gamma \text{ values} \end{aligned}$$

Both B_1 and B_2 provide estimates for the bias of $\hat{\gamma}$. The variability in B_1 is smaller than the variability in B_2. Therefore we only recorded B_1.

Based on the simulation results given in Table 5.1, we conclude that the derived formulae are excellent approximations for small $\sigma(< 0.06)$. For real data sets collected for the partial bleach method, σ is usually around 0.03. Since the sample sizes used in the simulation study are similar to real sample sizes the derived formulae can safely be used to estimate the biases for all four methods. In any case, we observe that the biases are negligible compared to the standard errors.

5.2.4 Theoretical justification for the use of the t-interval

Consider the model $y_i = f(x_i, \theta_0) + f(x_i, \theta_0)\sigma\epsilon_i$, $i = 1, \ldots, n$, where $\epsilon_i \sim N(0, 1)$, and θ_0 is the 3−vector of unknown true parameters.

Data Set	σ	ML		QL		GLS		DWLS	
		B_T	B_1	B_T	B_1	B_T	B_1	B_T	B_1
1	0.01	-0.046	-0.046	-0.049	-0.048	-0.046	-0.045	-0.054	-0.045
2	0.02	-0.182	-0.181	-0.195	-0.195	-0.183	-0.182	-0.217	-0.221
3	0.03	-0.410	-0.429	-0.438	-0.444	-0.412	-0.414	-0.489	-0.508
4	0.04	-0.730	-0.783	-0.778	-0.824	-0.733	-0.784	-0.869	-0.923
5	0.05	-1.140	-1.289	-1.216	-1.329	-1.146	-1.267	-1.358	-1.483
6	0.06	-1.641	-1.779	-1.752	-1.865	-1.650	-1.760	-1.955	-2.687

Table 5.1: Comparison of exact bias and estimated bias of $\hat{\gamma}$

Let $\hat{\theta}$ be an estimator for θ_0 that satisfies the approximation $\hat{\theta}-\theta_0 = C_1\sigma+C_2\sigma^2$ discussed earlier. Let $y, f_0, f_{\hat{\theta}}, \hat{\epsilon}$ denote the $n-$vectors consisting of the elements $y_i, f(x_i,\theta_0), f(x_i,\hat{\theta})$ and $(y_i - f(x_i,\hat{\theta}))/f(x_i,\hat{\theta})$ $(i = 1,\ldots,n)$ respectively.

Theorem 1 *As* $\sigma \to 0$, $\hat{\epsilon}^T\hat{\epsilon} \stackrel{\mathcal{D}}{\Longrightarrow} \chi^2$ *with* $(n-3)$ *degrees of freedom.*

Proof:

Let F_0 be the $n \times 3$ matrix with i th row r_i given by $r_i^T = \bigtriangledown f_0 = \left.\frac{\partial f_i}{\partial\theta}\right|_{\theta=\theta_0}$. The first order Taylor expansion of $f_{\hat{\theta}}$ around f_0 can be written as $f_{\hat{\theta}} \approx f_0 + F_0(\hat{\theta}-\theta_0)$. Then,

$$\begin{aligned} y - f_{\hat{\theta}} &\approx y - f_0 - F_0(\hat{\theta}-\theta_0), \\ \frac{1}{f_{\hat{\theta}}} &\approx \frac{1}{f_0}\left\{1 - \frac{F_0}{f_0}(\hat{\theta}-\theta_0)\right\}, \\ \frac{(y-f_{\hat{\theta}})}{f_{\hat{\theta}}} &\approx \frac{(y-f_0)}{f_0}\left\{1-\frac{F_0}{f_0}(\hat{\theta}-\theta_0)\right\} - \frac{F_0}{f_0}(\hat{\theta}-\theta_0)\left\{1-\frac{F_0}{f_0}(\hat{\theta}-\theta_0)\right\}, \end{aligned}$$

Notice that $\frac{(y-f_{\hat{\theta}})}{f_{\hat{\theta}}} = \sigma\hat{\epsilon}$, and $\frac{(y-f_0)}{f_0} = \sigma\epsilon$. From the results derived in Chapter 3, as $\sigma \longrightarrow 0$, $\hat{\theta} = \theta_0 + C_1\sigma + o(\sigma)$, where $C_1 = \left[\sum_{i=1}^n \left(\frac{\bigtriangledown f_0}{f_0}\right)_i \left(\frac{\bigtriangledown f_0}{f_0}\right)_i^T\right]^{-1} \sum_{i=1}^n \left(\frac{\bigtriangledown f_0}{f_0}\right)_i \epsilon_i = \Sigma\left(\frac{F_0}{f_0}\right)^T\epsilon$, and $\Sigma = \left[\sum_{i=1}^n \left(\frac{\bigtriangledown f_0}{f_0}\right)_i \left(\frac{\bigtriangledown f_0}{f_0}\right)_i^T\right]^{-1} = \left[\left(\frac{F_0}{f_0}\right)^T\left(\frac{F_0}{f_0}\right)\right]^{-1}$.

Taking first order terms in σ,

$$\sigma\hat{\epsilon} = \sigma\epsilon\left\{1-\frac{F_0}{f_0}C_1\sigma\right\} - \frac{F_0}{f_0}C_1\sigma\left\{1-\frac{F_0}{f_0}C_1\sigma\right\}$$

$$= \sigma\epsilon - \sigma\left(\frac{F_0}{f_0}\right)\Sigma\left(\frac{F_0}{f_0}\right)^T\epsilon + o(\sigma)$$

Thus, $\hat{\epsilon} = (I - B)\epsilon + o(\sigma)$, where $B = \left(\frac{F_0}{f_0}\right)\Sigma\left(\frac{F_0}{f_0}\right)^T$.

Note that the matrix B is symmetric and idempotent. (i.e. $B^T = B$ and $BB = B$.)

Therefore, $\hat{\epsilon}^T\hat{\epsilon} \sim \chi^2$ with degrees of freedom $df = tr(I - B) = n - tr(B)$.

Trace of the matrix B is

$$\begin{aligned}
tr(B) &= \sum_{i=1}^{n} B_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{3}\sum_{k=1}^{3}\left(\frac{F_0}{f_0}\right)^T_{ij}\Sigma_{jk}\left(\frac{F_0}{f_0}\right)_{ki} \\
&= \sum_{j=1}^{3}\sum_{k=1}^{3}\Sigma_{jk}\sum_{i=1}^{n}\left(\frac{F_0}{f_0}\right)_{ki}\left(\frac{F_0}{f_0}\right)^T_{ij} \\
&= \sum_{j=1}^{3}\sum_{k=1}^{3}\Sigma_{jk}\left[\left(\frac{F_0}{f_0}\right)\left(\frac{F_0}{f_0}\right)^T\right]_{kj} \\
&= \sum_{j=1}^{3}\left[\Sigma\left(\frac{F_0}{f_0}\right)\left(\frac{F_0}{f_0}\right)^T\right]_{jj} \\
&= tr\left[\Sigma\left(\frac{F_0}{f_0}\right)\left(\frac{F_0}{f_0}\right)^T\right] \\
&= tr\left\{I_{3\times 3}\right\} \\
&= 3.
\end{aligned}$$

Thus, $\hat{\epsilon}^T\hat{\epsilon} = \sum_{i=1}^{n}\left[\frac{(y_i - f(x_i,\hat{\theta}))^2}{f_i^2\sigma^2}\right] = \sum_{i=1}^{n}\left[\frac{w_i(y_i - f(x_i,\hat{\theta}))^2}{\sigma^2}\right]$ has approximately a χ^2 distribution with $(n - 3)$ degrees of freedom.

Theorem 2 *If σ is small, the error sum of squares $\hat{\epsilon}^T\hat{\epsilon}$ and the estimate $\hat{\theta}$ are independent, in the sense that $\lim_{\sigma\longrightarrow 0}\hat{\epsilon}^T\hat{\epsilon}$ and $C_1 = \lim_{\sigma\longrightarrow 0}(\hat{\theta} - \theta)/\sigma$ are independent.*

Proof:

We showed that (see the proof of Theorem 1) $\hat{\theta} - \theta = C_1\sigma + o(\sigma) = A\epsilon + o(\sigma)$, where $A = \Sigma\left(\frac{F_0}{f_0}\right)^T\sigma$, and that $\hat{\epsilon}^T\hat{\epsilon} = \epsilon^T(I - B)\epsilon + o(\sigma)$. Since $\Sigma = \left[\left(\frac{F_0}{f_0}\right)^T\left(\frac{F_0}{f_0}\right)\right]^{-1}$ we find $A(I - B) = \sigma\Sigma\left(\frac{F_0}{f_0}\right)^T - \sigma\left(\frac{F_0}{f_0}\right)^T\left(\frac{F_0}{f_0}\right)\Sigma\left(\frac{F_0}{f_0}\right)^T = 0$. Therefore, using Theorem 4.17 of Graybill (1961) we find that $\epsilon^T(I - B)\epsilon$ and C_1 are independent. Hence the result.

Theorem 3 *Let $-\gamma$ be the equivalent dose and $\hat{\gamma}$ be the quasi-likelihood estimate for γ. Assume that the two response curves have a common σ. Then,*

as $\sigma \to 0$, $\frac{(\hat{\gamma}-\gamma)}{\sqrt{Var(\hat{\gamma})}} \xrightarrow{\mathcal{D}} t_{(n_1+n_2-6)}$.

Proof:

Let $w_{1i} = 1/\hat{f}_{1i}^2$ and $w_{2j} = 1/\hat{f}_{2j}^2$. Applying Theorem 1 for the unbleached data we find that $\frac{\sum_{i=1}^{n} w_{1i}(y_{1i}-\hat{f}_{1i})^2}{\sigma^2}$ has a χ^2 distribution with $(n_1 - 3)$ degrees of freedom. Similarly for the bleached data $\frac{\sum_{j=1}^{n} w_{2j}(y_{2j}-\hat{f}_{2j})^2}{\sigma^2}$ has a χ^2 distribution with $(n_2 - 3)$ degrees of freedom.

The two χ^2 variates are independent. Therefore, $\frac{\sum_{i=1}^{n} w_{1i}(y_{1i}-\hat{f}_{1i})^2}{\sigma^2} + \frac{\sum_{j=1}^{n} w_{2j}(y_{2j}-\hat{f}_{2j})^2}{\sigma^2} = (n_1 + n_2 - 6)\frac{\hat{\sigma}^2}{\sigma^2}$ (where $\hat{\sigma}$ is the quasi-likelihood estimate for σ) has approximately a χ^2 distribution with $(n_1 + n_2 - 6)$ degrees of freedom.

Let $\theta = (\theta_1, \theta_2)$ and $g(\gamma, \theta) = f_1(\gamma, \theta_1) - f_2(\gamma, \theta_2)$. As $\sigma \to 0$, $\hat{\gamma} - \gamma = C_3\sigma + o(\sigma)$, where $C_3 = -\frac{1}{\frac{\partial g}{\partial \gamma}} C_1^T \frac{\partial g}{\partial \theta}$ (see Section 5.2.2.). Since C_1 is a linear combination of the standard normal random variates ϵ_i's (Chapter 3), it follows that $\hat{\gamma}$ is approximately normally distributed with mean γ. In Section 5.2 we showed that

$$Var(\hat{\gamma}) = \left[\left(\frac{\partial f_1}{\partial \theta_1}\right)^T \Sigma_1 \sigma^2 \left(\frac{\partial f_1}{\partial \theta_1}\right) + \left(\frac{\partial f_2}{\partial \theta_2}\right)^T \Sigma_2 \sigma^2 \left(\frac{\partial f_2}{\partial \theta_2}\right)\right] / \left[\frac{\partial f_1}{\partial \gamma} - \frac{\partial f_2}{\partial \gamma}\right]^2.$$

The variance of $\hat{\gamma}$ is estimated by replacing $\theta_1, \theta_2, \sigma$ by their estimated values. Neglecting the terms of $O(\sigma^3)$ from a Taylor expansion

$$\left(\widehat{\frac{\partial f_1}{\partial \theta_1}}\right)^T \hat{\Sigma}_1 \hat{\sigma}^2 \left(\widehat{\frac{\partial f_1}{\partial \theta_1}}\right) \approx \left(\frac{\partial f_1}{\partial \theta_1}\right)^T \Sigma_1 \hat{\sigma}^2 \left(\frac{\partial f_1}{\partial \theta_1}\right),$$

and

$$\left(\widehat{\frac{\partial f_2}{\partial \theta_2}}\right)^T \hat{\Sigma}_2 \hat{\sigma}^2 \left(\widehat{\frac{\partial f_2}{\partial \theta_2}}\right) \approx \left(\frac{\partial f_2}{\partial \theta_2}\right)^T \Sigma_2 \hat{\sigma}^2 \left(\frac{\partial f_2}{\partial \theta_2}\right).$$

Therefore,

$$\begin{aligned}
\frac{(\hat{\gamma}-\gamma)}{\sqrt{\widehat{Var}(\hat{\gamma})}} &= \frac{(\hat{\gamma}-\gamma)\left[\frac{\partial f_1}{\partial \gamma} - \frac{\partial f_2}{\partial \gamma}\right]}{\left[\left(\frac{\partial f_1}{\partial \theta_1}\right)^T \Sigma_1 \hat{\sigma}^2 \left(\frac{\partial f_1}{\partial \theta_1}\right) + \left(\frac{\partial f_2}{\partial \theta_2}\right)^T \Sigma_2 \hat{\sigma}^2 \left(\frac{\partial f_2}{\partial \theta_2}\right)\right]^{1/2}} + o(\sigma) \\
&= \frac{C_3}{\left\{\left[\left(\frac{\partial f_1}{\partial \theta_1}\right)^T \Sigma_1 \left(\frac{\partial f_1}{\partial \theta_1}\right) + \left(\frac{\partial f_2}{\partial \theta_2}\right)^T \Sigma_2 \left(\frac{\partial f_2}{\partial \theta_2}\right)\right] / \left[\frac{\partial f_1}{\partial \gamma} - \frac{\partial f_2}{\partial \gamma}\right]^2\right\}^{1/2} \left(\frac{\hat{\sigma}^2}{\sigma^2}\right)^{1/2}}
\end{aligned}$$

Let $U = \frac{C_3}{\left\{\left[\left(\frac{\partial f_1}{\partial \theta_1}\right)^T \Sigma_1 \left(\frac{\partial f_1}{\partial \theta_1}\right) + \left(\frac{\partial f_2}{\partial \theta_2}\right)^T \Sigma_2 \left(\frac{\partial f_2}{\partial \theta_2}\right)\right] / \left[\frac{\partial f_1}{\partial \gamma} - \frac{\partial f_2}{\partial \gamma}\right]^2\right\}^{1/2}}$ and $V = \left(\frac{\hat{\sigma}^2}{\sigma^2}\right)^{1/2}$. Note that, U, V are approximately distributed as $N(0,1)$ and $\sqrt{(\frac{\chi^2_\nu}{\nu})}$, $(\nu = n_1 + n_2 - 6)$. Furthermore, we showed that U and V are independent. Therefore, as $\sigma \to 0$, $\frac{(\hat{\gamma}-\gamma)}{\sqrt{(Var(\hat{\gamma}))}}$ converges in distribution to Student's t with $n_1 + n_2 - 6$ degrees of freedom.

Remark:

Theorem 3 remains valid if we replace $\hat{\gamma}$ by its maximum likelihood estimate, generalized least squares estimate or the data weighted least squares estimate.

When separate error factors σ_1 and σ_2 are fitted for the two curves we do not get exact t distributions. Instead a Satterthwaite type approximation is available.

Theorem 4 *Assume that the two response curves have different error factors σ_1 and $\sigma_2(\sigma_1 \neq \sigma_2)$. As $\sigma_1, \sigma_2 \to 0$, the distribution of $\frac{(\hat{\gamma}-\gamma)}{\sqrt{Var(\hat{\gamma})}}$ can be approximated by a Student's t distribution with the degrees of freedom 'df' defined as defined in Equation 5.17.*

Let

$$
\begin{aligned}
f_{1,i} &= f(x_i, \theta_1), \\
f_{2,i} &= f(x_i, \theta_2), \\
(\nabla f_1)_i &= \text{Gradient vector of } f_{1,i} \text{ with respect to } \theta_1, \\
(\nabla f_2)_i &= \text{Gradient vector of } f_{2,i} \text{ with respect to } \theta_2, \\
\Sigma_1 &= \left[\sum_{i=1}^{n} \left(\frac{\nabla f_1}{f_1}\right)_i \left(\frac{\nabla f_1}{f_1}\right)_i^T\right]^{-1}, \\
\Sigma_2 &= \left[\sum_{i=1}^{n} \left(\frac{\nabla f_2}{f_2}\right)_i \left(\frac{\nabla f_2}{f_2}\right)_i^T\right]^{-1}, \\
v_1 &= \text{Variance covariance matrix of } \hat{\theta_1} = \Sigma_1 \sigma_1^2, \\
v_2 &= \text{Variance covariance matrix of } \hat{\theta_2} = \Sigma_2 \sigma_2^2,
\end{aligned}
$$

When the response functions and the derivative vectors are evaluated at $x = \gamma$, we drop the suffix i. Let $u_1 = \frac{\nabla f_1^T v_1 \nabla f_1}{\left(\frac{\partial g}{\partial a_6}\right)^2}$, and $u_2 = \frac{\nabla f_2^T v_2 \nabla f_2}{\left(\frac{\partial g}{\partial a_6}\right)^2}$. Then, df is given by

$$
df = \frac{(u_1 + u_2)^2}{\frac{u_1^2}{(n_1-3)} + \frac{u_2^2}{(n_2-3)}}. \tag{5.17}
$$

An estimate for the *df* can be obtained by replacing $\theta_1, \theta_2, \sigma_1, \sigma_2$ by their estimates $\hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}_1, \hat{\sigma}_2$ described before.

Proof:

In the different error factor case, the variance of $\hat{\gamma}$ is

$$Var(\hat{\gamma}) = \left[\left(\frac{\partial f_1}{\partial \theta_1}\right)^T Cov(\hat{\theta}_1)\left(\frac{\partial f_1}{\partial \theta_1}\right) + \left(\frac{\partial f_2}{\partial \theta_2}\right)^T Cov(\hat{\theta}_2)\left(\frac{\partial f_2}{\partial \theta_2}\right)\right] / \left[\frac{\partial f_1}{\partial \gamma} - \frac{\partial f_2}{\partial \gamma}\right]^2,$$

where $Cov(\hat{\theta}_1) = \sigma_1^2\left[\sum_{i=1}^n \left(\frac{\nabla f_1}{f_1}\right)\left(\frac{\nabla f_1}{f_1}\right)^T\right]^{-1}$ and $Cov(\hat{\theta}_2) = \sigma_2^2\left[\sum_{i=1}^n \left(\frac{\nabla f_2}{f_2}\right)\left(\frac{\nabla f_2}{f_2}\right)^T\right]^{-1}$. As for the common error factor case, we can write $\frac{(\hat{\gamma}-\gamma)}{\sqrt{\widehat{Var}(\hat{\gamma})}}$ as U/V where U is a standard normal random variate and $V = \frac{\nabla f_1^T v_1 \nabla f_1}{\left(\frac{\partial g}{\partial a_6}\right)^2}\left(\frac{\hat{\sigma}_1^2}{\sigma_1^2}\right)^{1/2} + \frac{\nabla f_2^T v_2 \nabla f_2}{\left(\frac{\partial g}{\partial a_6}\right)^2}\frac{\nabla f_1^T v_2 \nabla f_1}{\left(\frac{\partial g}{\partial a_6}\right)^2}\left(\frac{\hat{\sigma}_2^2}{\sigma_2^2}\right)^{1/2}$. Also note that in the limit of $\sigma_1, \sigma_2 \longrightarrow 0$, $(n_1-3)\left(\frac{\hat{\sigma}_1^2}{\sigma_1^2}\right)$ and $(n_2-3)\left(\frac{\hat{\sigma}_2^2}{\sigma_2^2}\right)^{1/2}$ are each approximately distributed as χ^2 random variates on degrees of freedom $n_1 - 3$ and $n_2 - 3$ respectively. Furthermore, the two χ^2 random variates are independent. Therefore, the variance of $\hat{\gamma}$ in this case is a complex estimate of variance.[2] As in the case of common error factor, U and V are independent.

Thus, using Satterthwaite's approximation [53], we find that the distribution of the statistic $t = \frac{(\hat{\gamma}-\gamma)}{\sqrt{Var(\hat{\gamma})}}$ can be approximated by a Student's t distribution with the degrees of freedom '*df*' as defined by Equation 5.17.

5.2.5 Finite sample performance of t-intervals

In this section we describe a Monte Carlo study that examines the finite sample performance of the asymptotic theoretical results presented in Subsections 5.2.1. The parameter values were set at $\alpha_1 = 14.2853, \alpha_2 = 123.182, \alpha_3 = 393.065, \beta_2 = 192.547, \beta_3 = 756.620$ and $\gamma = -87.45$. The sample sizes used for the unbleached and bleached data sets are 16 and 13 respectively. The dose values used in the study are presented in Table 9.1 of Appendix 9.3 where they are labeled as data set $P1$. The values chosen for the error factors σ_1 and σ_2 are given in the Table 5.2. In the single error factor case (SEF), $\sigma_1 = \sigma_2 = \sigma$ and this is given

[2]Satterthwaite [54] defines a complex estimate of variance as one which is a linear combination of two or more statistics distributed as Chi square random variates .

in the column indicated by SEF. The DEF stands for the different error factor case and the chosen σ_1 and σ_2 values are given in the column indicated by DEF. Photon counts were simulated according to the model $y_{ji} = f(x_{ij}, \theta_j)(1 + \sigma_j \epsilon_{ji})$, for $j = 1, 2$ and $i = 1, \ldots, n_j$. Procedures described earlier in this Chapter were used to compute a confidence interval for the equivalent dose from each simulated sample. The fraction of times the confidence interval in each case captures the actual equivalent dose is recorded as the observed coverage. The results based on 10000 simulations are presented in Table 5.2.

Study	Nominal	SEF	Observed Coverage		DEF		Observed Coverage	
	Level	σ	case 1.1	case 1.2	σ_1	σ_2	case 2.1	case2.2
1	0.95	0.01	0.9502	0.9502	0.01	0.02	0.9506	0.9503
2	0.95	0.02	0.9481	0.9489	0.01	0.03	0.9515	0.9517
3	0.95	0.03	0.9505	0.9497	0.01	0.05	0.9465	0.9459
4	0.95	0.04	0.9524	0.9538	0.005	0.05	0.9500	0.9495
5	0.95	0.05	0.9509	0.9518	0.04	0.05	0.9534	0.9528

Table 5.2: Coverage probabilities of t intervals for single and different error factor cases

In Table 5.2, we use the following notation:

case 1.1 = the single error factor case, neglecting the bias

case 1.2 = the single error factor case, corrected for the bias

case 2.1 = the different error factor case, neglecting the bias

case 2.2 = the different error factor case, corrected for the bias

Conclusions:

1. The confidence intervals using t critical values were suggested based on the small σ asymptotic theory. The range of σ values chosen for the study well cover the values of σ observed in typical TL studies. The close agreement between the observed coverages and the nominal coverages therefore justifies the use of suggested confidence intervals for the equivalent dose based on the partial bleach method.

2. The agreement of the coverage probabilities in the different error factor case justifies the use of symmetric confidence intervals with an approximate t quantile based on the degrees of freedom suggested in Subsection 5.2.1.

3. The confidence intervals without correcting for the bias also have coverage probabilities in agreement with the nominal coverages. As we already mentioned, the biases are negligible compared to the standard errors of the estimators. Therefore, we did not correct for the bias in the studies described later in this chapter.

We also examined the coverage probabilities of confidence intervals obtained using a single error factor, for data generated using different factors. The results based on 10000 simulations are presented in Table 5.3. The parameter values for the study were set at $\alpha_1 = 14.2853, \alpha_2 = 123.182, \alpha_3 = 393.065, \beta_2 = 192.547, \beta_3 = 756.620$ and $\gamma = -87.45$. The dose values used in the study are presented in Table 9.1 of Appendix 9.3 where they are labeled as data set $P1$. The columns 2 and 3 of Table 5.3 indicate the sample sizes used for the unbleached and bleached data sets respectively. The different error factors used to generate the data are given in the columns 4 and 5. The last two columns indicate the coverage probabilities of confidence intervals ignoring the bias and corrected for the bias respectively.

Conclusions

Based on the results of the study we draw the following conclusions.

1. When the large sample size is associated with the small σ, the coverage probabilities of the resulting confidence intervals are larger than their nominal levels.

2. When the large sample size is associated with the large σ, the coverage probabilities of the resulting confidence intervals are smaller than their nominal levels.

The results of this study establish the importance of testing (for eg. using likelihood ratio test) whether the unbleached and bleached data sets correspond to a common σ or not before further analysis.

Study	Nominal Level	n_1	n_2	DEF		Observed Coverage	
				σ_1	σ_2	case 2.1	case2.2
1	0.95	16	13	0.01	0.02	0.9795	0.9794
2	0.95	16	13	0.01	0.03	0.9851	0.9852
3	0.95	16	13	0.01	0.05	0.9905	0.9904
4	0.95	16	13	0.005	0.05	0.9914	0.9913
6	0.95	16	13	0.02	0.01	0.9155	0.9153
7	0.95	16	13	0.03	0.01	0.9051	0.9044
8	0.95	16	13	0.05	0.01	0.8980	0.8972
9	0.95	16	13	0.05	0.005	0.8918	0.8910

Table 5.3: Coverage probabilities of t intervals using a common error factor for data corresponding to different error factors

5.3 Estimation via simultaneous curve fitting

In Section 5.2, we described the estimation of the equivalent dose as the intersection of the two fitted response curves corresponding to the bleached and unbleached curves. Here we describe another technique for the same purpose. We reparametrize so that γ, the quantity of interest, is one of our parameters in the new setting.

Let the parameter vectors corresponding to the unbleached and bleached curves be $\theta_1 = (\alpha_1, \alpha_2, \alpha_3)^T$ and $\theta_2 = (\beta_1, \beta_2, \beta_3)^T$. Let γ be the dose corresponding to the intersection of the unbleached and bleached response curves. Then, γ satisfies the equation

$$\alpha_1 \left\{1 - \exp\left[-\frac{(\gamma + \alpha_2)}{\alpha_3}\right]\right\} = \beta_1 \left\{1 - \exp\left[-\frac{(\gamma + \beta_2)}{\beta_3}\right]\right\}.$$

This gives

$$\beta_1 = \alpha_1 \frac{\left\{1 - \exp\left[-\frac{(\gamma+\alpha_2)}{\alpha_3}\right]\right\}}{\left\{1 - \exp\left[-\frac{(\gamma+\beta_2)}{\beta_3}\right]\right\}}. \tag{5.18}$$

We eliminate β_1 using the parameters $\alpha_1, \alpha_2, \alpha_3, \beta_2, \beta_3, \gamma$ and fit the two response curves simultaneously treating γ as a parameter that appears explicitly in the new setting. This

avoids the problem of estimating the intersection point after fitting the nonlinear curves as in the previous method described in Section 5.2. This also simplifies the estimation of the error of the estimate. Since this is a reparametrization of the problem, the estimate and the error of the estimate are unchanged.

Next we describe the estimation procedures.

5.3.1 Maximum likelihood estimates

Suppose first the two curves have a common error factor. The likelihood for a sample of $n_3(= n_1 + n_2)$ observations is

$$\begin{aligned} L &= \prod_{i=1}^{n_1} \frac{1}{\sqrt{2}\pi\sigma f_1(x_{1i},\theta_1)} \exp\left\{\frac{-[y_i - f_1(x_{1i},\theta_1)]^2}{2\sigma^2 f_1^2(x_{1i},\theta_1)}\right\} \\ &\quad \times \prod_{j=1}^{n_2} \frac{1}{\sqrt{2}\pi\sigma f_2(x_{2j},\theta_2)} \exp\left\{\frac{-[y_j - f_2(x_{2j},\theta_2)]^2}{2\sigma^2 f_2^2(x_{2j},\theta_1)}\right\}. \end{aligned}$$

where $\theta_1 = (\alpha_1, \alpha_2, \alpha_3)$ and $\theta_2 = (\beta_1, \beta_2, \beta_3)$. The parameter vector of interest[3] is $\theta = (\alpha_1, \alpha_2, \alpha_3, \beta_2, \beta_3, \gamma)^T$.

The log-likelihood for the sample apart from a constant is

$$\begin{aligned} l &= -(n_1 + n_2)\log(\sigma) - \sum_{i=1}^{n_1} \log(f(x_{1i},\theta_1)) - \sum_{i=1}^{n_1} \frac{[y_{1i} - f(x_{1i},\theta_1)]^2}{2\sigma^2 f^2(x_{1i},\theta_1)} \\ &\quad - \sum_{k=1}^{n_2} \log(f(x_{2k},\theta_2)) - \sum_{k=1}^{n_2} \frac{[y_{2k} - f(x_{2k},\theta_2)]^2}{2\sigma^2 f^2(x_{2k},\theta_2)}. \end{aligned} \tag{5.19}$$

The maximum likelihood estimates for the parameters maximize the above log-likelihood function. The algorithm described in 3.2 is modified as follows to obtain the likelihood estimates. As described in Section 3.2, we use a 2-part iteration; see also Green [33]. Equating $\frac{\partial l}{\partial \sigma}$ to zero, we find

$$\hat{\sigma}^2 = \left(\sum_{j=1}^{n_1} [(y_{1j} - f_1)/f_1]^2 + \sum_{k=1}^{n_2} [(y_{2k} - f_2)/f_2]^2\right) /(n_1 + n_2). \tag{5.20}$$

The algorithm that solves for θ is described by the following steps.

1. Find an initial estimate $\hat{\theta}_0$ for θ.

[3]Equation 5.18 expresses β_1 in terms of the components of θ.

2. Evaluate f_1, f_2 at the starting value.

3. Estimate σ using Equation 5.20 at the estimate for θ.

4. Compute the matrix $D = \frac{\partial f}{\partial \theta}$. The $n \times 6$ matrix D consists of the columns $\frac{\partial f_i}{\partial \alpha_1}, \frac{\partial f_i}{\partial \alpha_2}, \frac{\partial f_i}{\partial \alpha_3}, \frac{\partial f_i}{\partial \beta_2}, \frac{\partial f_i}{\partial \beta_3}$ and $\frac{\partial f_i}{\partial \gamma}$ which are given by the equations 5.21 below; here f is the n-vector with entries $f_1(x_i, \theta_1)$ (for $i \leq n_1$) and $f_2(x_i, \theta_2)$ (for $n_1 < i \leq n$).

5. Compute the $n-$vector $u_0 = \frac{\partial l}{\partial f_i}$ (5.22) and the matrix A (5.23).

6. Compute $\hat{\beta}_0 = (D^T A D)^{-1} D^T u_0$.

7. At the $(k+1)$st iteration, estimates for θ are given by $\hat{\theta}_{k+1} = \hat{\theta}_k + \hat{\beta}_k$, where $\hat{\beta}_k = (D_k^T A_k D_k)^{-1} D_k^T u_k$. The subscript k indicates that the respective terms are evaluated using the parameter estimates for θ at the kth iteration. Then update D_k to D_{k+1} by evaluating D given below at θ_{k+1}.

8. Repeat the above procedure until the desired convergence. In the algorithms we developed, the convergence criterion for θ was taken as when the step size $\beta < 10^{-5}$.

The matrix D has the following components:

$$
\begin{aligned}
\frac{\partial f_i}{\partial \alpha_1} &= 1 - \exp\left[\frac{-(x_{1i} + \alpha_2)}{\alpha_3}\right], (\text{for } 1 < i \leq n_1) \\
&= \left\{1 - \exp\left[-\frac{\gamma + \alpha_2}{\alpha_3}\right]\right\}\left\{1 - \exp\left[-\frac{x_{2k} + \beta_2}{\beta_3}\right]\right\}\left\{1 - \exp\left[-\frac{\gamma + \beta_2}{\beta_3}\right]\right\}^{-1} \\
&\quad (\text{for } n_1 < i \leq n_1 + n_2); (k = (i - n_1)) \\
\frac{\partial f_i}{\partial \alpha_2} &= \frac{\alpha_1}{\alpha_3} \exp\left[\frac{-(x_{1i} + \alpha_2)}{\alpha_3}\right], (\text{for } 1 < i \leq n_1) \\
&= \frac{\alpha_1}{\alpha_3} \exp\left[-\frac{\gamma + \alpha_2}{\alpha_3}\right]\left\{1 - \exp\left[-\frac{x_{2k} + \beta_2}{\beta_3}\right]\right\}\left\{1 - \exp\left[-\frac{\gamma + \beta_2}{\beta_3}\right]\right\}^{-1} \\
&\quad (\text{for } n_1 < i \leq n_1 + n_2); (k = (i - n_1)) \\
\frac{\partial f_i}{\partial \alpha_3} &= -\frac{\alpha_1}{\alpha_3^2}(x_{1i} + \alpha_2) \exp\left[\frac{-(x_{1i} + \alpha_2)}{\alpha_3}\right], (\text{for } 1 < i \leq n_1) \\
&= -\frac{\alpha_1}{\alpha_3}(\gamma + \alpha_2) \exp\left[-\frac{\gamma + \alpha_2}{\alpha_3}\right]\left\{1 - \exp\left[-\frac{x_{2k} + \beta_2}{\beta_3}\right]\right\} \\
&\quad \times \left\{1 - \exp\left[-\frac{\gamma + \beta_2}{\beta_3}\right]\right\}^{-1}, (\text{for } n_1 < i \leq n_1 + n_2); (k = (i - n_1))
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial f_i}{\partial \beta_2} &= 0, (\text{for } 1 < i \leq n_1) \\
&= -\frac{\alpha_1}{\beta_3}\left\{1-\exp\left[-\frac{\gamma+\alpha_2}{\alpha_3}\right]\right\}\left\{1-\exp\left[-\frac{x2+\beta_2}{\beta_3}\right]\right\} \\
&\quad \times\exp\left[-\frac{\gamma+\beta_2}{\beta_3}\right]\left\{1-\exp\left[-\frac{\gamma+\beta_2}{\beta_3}\right]\right\}^{-2} \\
&\quad +\frac{\alpha_1}{\beta_3}\left(1-\exp\left[-\frac{\gamma+\alpha_2}{\alpha_3}\right]\right)\exp\left[-\frac{x_{2k}+\beta_2}{\beta_3}\right] \\
&\quad \times\left\{1-\exp\left[-\frac{\gamma+\beta_2}{\beta_3}\right]\right\}^{-1}, (\text{for } n_1 < i \leq n_1+n_2); (k=(i-n_1)) \\
\frac{\partial f_i}{\partial \beta_3} &= 0, (\text{for } 1 < i \leq n_1) \\
&= \frac{\alpha_1}{\beta_3}\left(1-\exp\left[-\frac{\gamma+\alpha_2}{\alpha_3}\right]\right)\left\{1-\exp\left[-\frac{x_{2k}+\beta_2}{\beta_3}\right]\right\} \\
&\quad \times(\gamma+\beta_2)\exp\left[-\frac{\gamma+\beta_2}{\beta_3}\right]\left\{1-\exp\left[-\frac{\gamma+\beta_2}{\beta_3}\right]\right\}^{-2} \\
&\quad -\frac{\alpha_1}{\beta_3}\left\{1-\exp\left[-\frac{\gamma+\alpha_2}{\alpha_3}\right]\right\}(x_{2k}+\beta_2)\exp\left[-\frac{x_{2k}+\beta_2}{\beta_3}\right] \\
&\quad \times\left\{1-\exp\left[-\frac{\gamma+\beta_2}{\beta_3}\right]\right\}^{-1}, (\text{for } n_1 < i \leq n_1+n_2); (k=(i-n_1)) \\
\frac{\partial f_i}{\partial \gamma} &= 0, (\text{for } 1 < i \leq n_1) \\
&= \frac{\alpha_1}{\alpha_3}\exp\left[-\frac{\gamma+\alpha_2}{\alpha_3}\right]\left\{1-\exp\left[-\frac{x_{2k}+\beta_2}{\beta_3}\right]\right\}\left\{1-\exp\left[-\frac{\gamma+\beta_2}{\beta_3}\right]\right\}^{-1} \\
&\quad -\frac{\alpha_1}{\beta_3}\left\{1-\exp\left[-\frac{\gamma+\alpha_2}{\alpha_3}\right]\right\}\left\{1-\exp\left[-\frac{x_{2k}+\beta_2}{\beta_3}\right]\right\}\exp\left[-\frac{\gamma+\beta_2}{\beta_3}\right] \\
&\quad \times\left\{1-\exp\left[-\frac{\gamma+\beta_2}{\beta_3}\right]\right\}^{-2}
\end{aligned}
\tag{5.21}
$$

The n vector $u = \frac{\partial l}{\partial f_i}$ consists of the following elements:

$$
\left.
\begin{aligned}
\frac{\partial l}{\partial f_i} &= \frac{-1}{f_{1i}} + \frac{1}{\sigma^2}\frac{[y_{1i}-f_{1i}]}{f_{1i}^2} + \frac{1}{\sigma^2}\frac{[y_{1i}-f_{1i}]^2}{f_{1i}^3} \\
&\quad \text{for}(1 \leq i \leq n_1) \\
&= \frac{-1}{f_{2k}} + \frac{1}{\sigma^2}\frac{[y_{2k}-f_{2k}]}{f_{2k}^2} + \frac{1}{\sigma^2}\frac{[y_{2k}-f_{2k}]^2}{f_{2k}^3} \\
&\quad \text{for}(n_1 < i \leq n_1+n_2);\ k=(i-n_1)
\end{aligned}
\right\}
\tag{5.22}
$$

Note that $E(y_{1i} - f_{1i}) = 0$, $E(y_{1i} - f_{1i})^2 = \sigma^2 f_{1i}^2$, $E(y_{2k} - f_{2k}) = 0$, and $E(y_{2k} - f_{2k})^2 = \sigma^2 f_{2k}^2$. The matrix A is diagonal with ith diagonal element

$$E\left(-\frac{\partial^2 l}{\partial f_i^2}\right) = \begin{cases} \frac{2}{f_{1i}^2} + \frac{1}{\sigma^2 f_{1i}^2} \\ \text{for}(1 \leq i \leq n_1) \\ \frac{2}{f_{2k}^2} + \frac{1}{\sigma^2 f_{2k}^2} \\ \text{for}(n_1 < i \leq n_1 + n_2);\ k = (i - n_1) \end{cases} \tag{5.23}$$

Let $\phi = (\theta, \sigma)$ and $\hat{\phi} = (\hat{\theta}, \hat{\sigma})$. From the large sample theory for maximum likelihood estimators, the approximate asymptotic variance covariance matrix for $\hat{\phi}$ is given by $E\left(-\frac{\partial^2 l}{\partial \phi \phi^T}\right)^{-1}$.

5.3.2 Profile likelihood intervals for the equivalent dose

Suppose that the parameter vector $\theta = (\alpha_1, \alpha_2, \alpha_3, \beta_2, \beta_3, \gamma)^T$ is partitioned as (Ψ, γ) where Ψ consists of the elements of θ excluding γ.

Let $\hat{\theta} = (\hat{\Psi}, \hat{\gamma})$ be the maximum likelihood estimate for θ and l_{max} be the value of the log-likelihood function (equation 5.19) evaluated at the maximum likelihood estimate $\hat{\theta}$. Let $\hat{l}_\gamma$ be the value of the log-likelihood function evaluated at $(\hat{\Psi}_\gamma, \gamma)$ where $\hat{\Psi}_\gamma$ maximizes the log-likelihood function for fixed γ. Then, the profile likelihood interval for the equivalent dose with confidence coefficient α is

$$\left\{\gamma \middle| 2\left(l_{max} - \hat{l}_\gamma\right) \leq \chi^2_{1,1-\alpha}\right\},$$

where $\chi^2_{1,1-\alpha}$ is the upper α quantile for a chi squared distribution with 1 degree of freedom.

Computing $\hat{\Psi}_\gamma$:

Let D' be the $n \times 5$ matrix consisting of the n vectors $\frac{\partial f_i}{\partial \alpha_1}, \frac{\partial f_i}{\partial \alpha_2}, \frac{\partial f_i}{\partial \alpha_3}, \frac{\partial f_i}{\partial \beta_2}, \frac{\partial f_i}{\partial \beta_3}$. The restricted maximum likelihood estimate $\hat{\Psi}_\gamma$ can be computed by replacing the $n \times 6$ matrix D by the $n \times 5$ matrix D' in the algorithm described in Section 5.3.1. The end points of the profile likelihood interval are the roots of $R(\gamma) = 2\left(l_{max} - \hat{l}_\gamma\right) - \chi^2_{1,1-\alpha} = 0$. The following algorithm is used to find these roots. We have no convincing evidence that the profile, $R(\gamma)$,

must be unimodal but our algorithm never failed to converge for small σ. We suspect, but have not proved, that $R(\gamma)$ is unimodal in the limit of small σ.

1. Evaluate $R(\gamma)$ at $\gamma_1 = \hat{\gamma} - Z_{\alpha/2}s(\hat{\gamma})$. Here $\hat{\gamma}$ is the maximum likelihood estimate for γ and $s(\hat{\gamma})$ is the estimated standard error for $\hat{\gamma}$, which is the square root of the corresponding diagonal element of the inverse of the matrix $E\left(-\frac{\partial^2 l}{\partial\theta\theta^T}\right)$ evaluated at $\hat{\theta}$ and $Z_{\alpha/2}$ is the upper $\alpha/2$ th quantile for a standard normal distribution.

2. If $R(\gamma_1)$ is positive, then use the method of bisection starting with γ_1 and $\hat{\gamma}$ until the absolute value of R_γ is less than a desired small positive number ϵ_1. In the software we developed we use $\epsilon_1 = 10^{-3}$.

3. If $R(\gamma_1)$ is negative, then we keep subtracting $s(\hat{\gamma})$ until $R(\gamma_i)$ is positive at a certain γ_i. Again use the method of bisection starting with γ_i and $\gamma_{(i-1)}(= \gamma_i + s(\hat{\gamma}))$ until the absolute value of R_γ is less than ϵ_1.

4. To find the right end point we start from $\gamma_2 = \hat{\gamma} + Z_{\alpha/2}s(\hat{\gamma})$ and follow similar steps as in the case of left end point but this time adding $s(\hat{\gamma})$ if $R(\gamma_2)$ happens to be negative. (So, $\gamma_{(i-1)} = \gamma_i - s(\hat{\gamma})$ in this case.)

5. Let $\hat{\gamma}_1$ and $\hat{\gamma}_2$ be the roots of $R(\gamma) = 0$ described earlier. Then, (γ_1, γ_2) is a $(1-\alpha)100\%$ profile likelihood interval for the equivalent dose.

5.3.3 Confidence intervals using asymptotic normality

The confidence intervals for γ can be constructed assuming that the maximum likelihood estimator $\hat{\gamma}$ is approximately normally distributed with mean γ and variance the corresponding diagonal element of the inverse of the matrix $E(-\frac{\partial^2 l}{\partial\theta\partial\theta^T})$. The lower and upper confidence limits for an approximate $100(1-\alpha)\%$ confidence interval for γ based on the normal approximation are $\hat{\gamma} \mp Z_{\alpha/2}\sqrt{\widehat{var}(\hat{\gamma})}$, where $Z_{\alpha/2}$ denotes the upper α_2 th quantile of a standard normal distribution. We also computed confidence intervals using a t quantile with $(n_1 + n_2 - 6)$ degrees of freedom instead of the Z quantile. It is important to note that, when computing the confidence intervals based on the t quantile, the error factor σ

was estimated from $\hat{\sigma}^2 = \left\{\sum_{i=1}^{n_1}\left[(y_{1i}-f_1)/f_1\right)^2] + \sum_{j=1}^{n_2}\left[(y_{2j}-f_2)/f_2\right)^2]\right\}/(n_1+n_2-6)$, in contrast to the maximum likelihood estimate for σ used in Z intervals.

5.3.4 Confidence intervals using transformed F critical values

The traditional profile likelihood interval is computed by inverting the test based on the approximation that $2\left(l_{\max} - \hat{l}_\gamma\right)$ is distributed as a chi squared random variate with 1 degree of freedom. Here we suggest computing confidence intervals using approximate transformed F critical values instead of the approximate chi squared critical values.

Notation:

Let the parameter vector in the regression model be partitioned as $\beta = (\beta_1, \beta_2)$, where β_2 is the parameter of interest. Let

$$\begin{aligned} l(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}) &= \text{log likelihood evaluated at the maximum likelihood estimates,} \\ \hat{\beta}_1(\beta_{2,0}) &= \text{maximum likelihood estimate of } \beta_1 \text{ for } \beta_2 \text{ fixed at } \beta_{2,0}, \end{aligned}$$

and

$$\hat{\sigma}(\beta_{2,0}) = \text{maximum likelihood estimate of } \sigma \text{ for } \beta_2 \text{ fixed at } \beta_{2,0}.$$

Let $l(\hat{\beta}_1(\beta_{2,0}), \beta_{2,0}, \hat{\sigma}(\beta_{2,0}))$ denote the log-likelihood evaluated at the restricted maximum likelihood estimates and $LR = 2\left(l(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}) - l(\hat{\beta}_1(\beta_{2,0}), \beta_{2,0}, \hat{\sigma}(\beta_{2,0}))\right)$.

For the linear regression model $Y = X\beta + \epsilon$, it is well known (see, eg., Draper *et. al.* [27], Seber *et. al.* [55], Chatterjee *et. al.* [15]) that the hypothesis $H_0 : \beta_2 = \beta_0$ can be tested based on the fact that $\frac{[S(\beta_1,\beta_{2,0}) - S(\hat{\beta}_1,\hat{\beta}_2)]/p}{S(\hat{\beta}_1,\hat{\beta}_2)/(n-p_f)}$ has a F distribution with numerator and denominator degrees of freedom p and $n - p_f$; here p is the dimension of β_2 while p_f is the dimension of $\beta = (\beta_1, \beta_2)$ and $S(\theta)$ denotes the error sum of squares for the model with parameters β fixed at θ.

For the normal error linear regression model, this is equivalent to tests based on the fact that $2LR$ is distributed as $n \log\left[1 + \frac{F_{p,n-p_f}}{(n-p_f)}\right]$, where F_{ν_1,ν_2} denotes an F distribution on numerator and denominator degrees of freedom ν_1 and ν_2 respectively.

Remarks:

Note that, when $p = 1$, $F_{1,n-p_f} = t^2_{n-p_f}$ and

$$\begin{aligned} 2\ LR &= n \log \left[1 + \frac{t^2_{(n-p_f)}}{(n-p_f)} \right] \\ &\approx \frac{n}{(n-p_f)} t^2_{(n-p_f)}. \end{aligned}$$

So assuming $2LR$ has approximately a χ^2_1 distribution is equivalent to pretending

1. $\frac{n}{(n-p_f)} \approx 1$, and
2. $t_{(n-p_f)} \approx N(0,1)$.

This is true in large samples. However, as we already mentioned thermoluminescence data sets are frequently small in size. Therefore, using F critical values rather than χ^2 critical values, is expected to produce more precise coverage probabilities in small samples.

While the above result does not provide an exact test for our non-linear regression model $Y = f(x_i, \theta)(1 + \sigma\epsilon_i)$, we examined confidence intervals based on

$$2LR = n \log \left[1 + \frac{F_{1,(n-6)}}{(n-6)} \right],$$

where $n(= n_1 + n_2)$ is the sample size. Next we present these results.

5.3.5 Finite sample performance of the confidence intervals

We performed a Monte Carlo study similar to the one described in Section 5.2.5, to examine the coverage probabilities of the suggested confidence intervals. The parameter vector was set at

$$\theta = (14.2853, 123.182, 393.065, 192.547, 756.620, -87.45)^T.$$

The dose vectors used in the study are indicated in Table 5.4 within brackets in the column 'Dose'. Table 9.1 of Appendix 9.3 describes these dose vectors. The samples sizes of the unbleached and bleached data sets are indicated by n_1 and n_2 and $n = n_1 + n_2$. The results based on 10000 simulations are given in the Table 5.4. In the Table 5.4, FCRIT1 and FCRIT2 indicate the coverage probabilities using the approximations described by Equations 5.24

and 5.25:

$$2LR = n\log\left[1+\frac{F_{1,(n-6)}}{(n-6)}\right] \tag{5.24}$$

$$\approx \frac{n}{(n-6)}F_{1,(n-6)}. \tag{5.25}$$

Dose	σ	nominal	Profile	FCRIT1	FCRIT2	Z	t	n_1	n_2
(P1)	0.01	0.95	0.9198	0.9520	0.9620	0.9078	0.9518	16	13
(P1)	0.02	0.95	0.9182	0.9527	0.9622	0.9099	0.9523	16	13
(P2)	0.03	0.95	0.9357	0.9510	0.9562	0.9301	0.9523	26	23
(P3)	0.03	0.95	0.9389	0.9466	0.9487	0.9365	0.9468	50	50
(P3)	0.04	0.95	0.9457	0.9545	0.9571	0.9415	0.9525	50	50
(P4)	0.01	0.95	0.9446	0.9484	0.9497	0.9431	0.9472	100	100
(P4)	0.005	0.95	0.9449	0.9493	0.9504	0.9439	0.9494	100	100
(P5)	0.005	0.95	0.9500	0.9534	0.9543	0.9489	0.9530	125	125

Table 5.4: Coverage probabilities of profile, F, Z and t intervals

Conclusions:

The following conclusions were drawn from the simulation results:

1. When the sample sizes are small, profile likelihood intervals and the Z intervals based on the maximum likelihood estimate were found to have smaller coverage probabilities than the nominal values.

2. The coverage probabilities of profile likelihood intervals based on the maximum likelihood estimates are closer to their nominal values than the corresponding traditional $Z-$intervals based on the large sample theory for maximum likelihood estimates.

3. As the sample size is increased, the coverage probabilities of profile likelihood intervals and the Z intervals approach their nominal values.

4. The coverage probabilities of the confidence intervals based on the transformed F critical value (Equation 5.24) agree with their nominal values even when the sample sizes are small.

5. When the sample sizes are small, the confidence intervals based on the approximation described in Equation 5.25 were found to be conservative. However, the coverages of these confidence intervals were found to converge much faster to the nominal coverages compared to profile likelihood intervals and the Z intervals.

6. The coverage probabilities of t intervals based on the maximum likelihood estimates agree with their nominal coverages even for small samples.

7. The coverage probabilities of the confidence intervals do not appear to be dependent on σ, if σ is small. So, when we simulated large samples we fixed σ at a small value to save computational time.

5.3.6 Robustness of the transformed F test

The confidence intervals based on the likelihood ratio with a transformed F critical value were found to agree well with their nominal coverages, when the observations (Y_i's) are normally distributed. We examined the robustness of this test to departures from normality. The observed photon counts were simulated from a gamma distribution with mean $f(x_i, \theta)$ and variance $\sigma^2 f^2(x_i, \theta)$. The confidence intervals for the equivalent dose γ were computed using the transformed F critical value as described in the Section 5.3.4. The observed coverage probabilities based on 10000 simulations are given in the Table 5.5.

Conclusions:

1. Even with small samples of data from a gamma distribution the coverage probabilities of the intervals based on the transformed F critical values were found to agree with their nominal values.

2. The coverage probabilities of t intervals also agree with their nominal values even when the sample sizes are small.

Dose	σ	nominal	FCRIT	Profile	Z	t	n_1	n_2
(P1)	0.01	0.95	0.9502	0.9178	0.9065	0.9511	16	13
(P1)	0.02	0.95	0.9472	0.9117	0.9018	0.9469	16	13
(P2)	0.029	0.95	0.9535	0.9384	0.9293	0.9522	26	23
(P3)	0.029	0.95	0.9512	0.9442	0.9413	0.9533	50	50
(P4)	0.01	0.95	0.9510	0.9470	0.9463	0.9507	100	100
(P4)	0.005	0.95	0.9492	0.9441	0.9435	0.9486	100	100
(P5)	0.005	0.95	0.9540	0.9502	0.9505	0.9542	125	125

Table 5.5: Coverage probabilities of confidence intervals for gamma distributed data

3. In small samples, the coverage probabilities of Z intervals are lower than their nominal values. As the sample size is increased, the coverage probabilities approach their nominal values.

4. We notice that the shape parameter for the gamma distribution is $1/\sigma^2$. Since we take σ to be small, the shape parameter for the gamma distribution is large. Therefore, the gamma distributions in question are very close to normal distributions.

5.3.7 Quasi-likelihood estimates

The algorithm described in Chapter 3 can be slightly modified to obtain quasi-likelihood estimates for 'simultaneous curve fitting'.

Let y be the n vector of photon counts of unbleached and bleached samples stacked together. Let x be the n vector of corresponding dose values.

Let μ be the vector of mean values defined as

$$\mu = f(x,\theta) = \begin{cases} f_1(x_i,\theta) & \text{for } i = 1,\ldots,n_1 \\ f_2(x_i,\theta) & \text{for } i = (n_1+1),\ldots,n_1+n_2. \end{cases}$$

The variance covariance matrix for y is diagonal with diagonal elements

$$V(i,i) = \begin{cases} \sigma_1^2 f_1^2(x_i,\theta) & \text{for } i = 1,\ldots,n_1 \\ \sigma_2^2 f_2^2(x_i,\theta) & \text{for } i = (n_1+1),\ldots,n_1+n_2. \end{cases}$$

The parameters σ_1 and σ_2 denote the relative error factors for the unbleached and bleached data. We discuss two settings; when the percent error factor does not depend on whether or not the samples had received laboratory bleaching (i.e. $\sigma_1 = \sigma_2$) and when different error factors are more appropriate for the laboratory bleached and unbleached samples (i.e. $\sigma_1 \neq \sigma_2$.)

The quasi-likelihood estimates for θ are obtained as solutions of the quasi-likelihood equations

$$F^T(\theta)V^{-1}(\theta)\,[y - f] = 0. \tag{5.26}$$

For notational convenience, let $s_{1i} = \exp\left[-\frac{(x_i+\alpha_2)}{\alpha_3}\right]$, $s_{2i} = \exp\left[-\frac{(x_i+\beta_2)}{\beta_3}\right]$, $s_3 = \exp\left[-\frac{(\gamma+\alpha_2)}{\alpha_3}\right]$ and $s_4 = \exp\left[-\frac{(\gamma+\beta_2)}{\beta_3}\right]$. The $n \times 6$ matrix F consists of the columns F_j, $j = 1,\ldots,6$ which are defined as follows:

$$F_1 = \begin{cases} (1 - s_{1i}) & \text{for } i = 1,\ldots,n_1 \\ \frac{(1-s_3)(1-s_{2i})}{1-s_4} & \text{for } i = (n_1+1),\ldots,(n_1+n_2) \end{cases}$$

$$F_2 = \begin{cases} \frac{\alpha_1 s_{1i}}{\alpha_3} & \text{for } i = 1,\ldots,n_1 \\ \frac{\alpha_1 s_3 (1-s_{2i})}{\alpha_3 (1-s_4)} & \text{for } i = (n_1+1),\ldots,(n_1+n_2) \end{cases}$$

$$F_3 = \begin{cases} -\frac{\alpha_1 (x+\alpha_2) s_{1i}}{{\alpha_3}^2} & \text{for } i = 1,\ldots,n_1 \\ -\frac{\alpha_1 (\gamma+\alpha_2) s_3 (1-s_{2i})}{{\alpha_3}^2 (1-s_4)} & \text{for } i = (n_1+1),\ldots,(n_1+n_2) \end{cases}$$

$$F_4 = \begin{cases} 0 & \text{for } i = 1,\ldots,n_1 \\ -\frac{\alpha_1 (1-s_3)(1-s_{2i}) s_4}{(1-s_4)^2 \beta_3} + \frac{\alpha_1 (1-s_3) s_{2i}}{(1-s_4)\beta_3} & \text{for } i = (n_1+1),\ldots,(n_1+n_2) \end{cases}$$

$$F_5 = \begin{cases} 0 & \text{for } i = 1,\ldots,n_1 \\ \frac{\alpha_1 (1-s_3)(1-s_{2i})(\gamma+\beta_2) s_4}{(1-s_4)^2 {\beta_3}^2} - \frac{\alpha_1 (1-s_3)(x+\beta_2) s_{2i}}{(1-s_4){\beta_3}^2} & \text{for } i = (n_1+1),\ldots,(n_1+n_2) \end{cases}$$

$$F_6 = \begin{cases} 0 & \text{for } i = 1,\ldots,n_1 \\ \frac{\alpha_1 s_3 (1-s_{2i})}{\alpha_3 (1-s_4)} - \frac{\alpha_1 (1-s_3)(1-s_{2i}) s_4}{(1-s_4)^2 \beta_3} & \text{for } i = (n_1+1),\ldots,(n_1+n_2) \end{cases}$$

The iterative scheme described in Section 3.3 can be used to solve the system of equations 5.26.

5.3.8 Confidence intervals based on the quasi-score test

Let $\phi^T = (\alpha_1, \alpha_2, \alpha_3, \beta_2, \beta_3)$ and $\theta^T = (\phi^T, \gamma)$ be the vector of unknown parameters. Let $U(\theta) = \frac{\partial Q}{\partial \theta}$ (equation 5.26) denote the vector of quasi-score functions.

We computed confidence intervals for the equivalent dose by inverting the score test described in Dean [24].

The procedure is as follows.

1. Fix the equivalent dose at a value γ_0.
2. Partition $U^T(\theta) = [U_1^T(\theta), U_2(\theta)]$, where $U_1^T(\theta) = \frac{\partial Q}{\partial \phi^T}$ and $U_2(\theta) = \frac{\partial Q}{\partial \gamma}$.
3. Compute $\hat{\phi}(\gamma_0)$ that solves $U_1(\phi, \gamma_0) = 0$.
4. Compute $u_0 = U_2(\hat{\phi}(\gamma_0), \gamma_0)$. For our model u_0 is a scalar.
5. Let Σ be the asymptotic variance covariance matrix for the quasi-likelihood estimator for θ (note that Σ depends on σ^2).
6. The parameter σ^2 in the variance covariance matrix is estimated[4] as

$$\hat{\sigma}^2 = \frac{1}{(n_1 + n_2 - 5)} \left[\sum_{i=1}^{n_1} \left(\frac{y_{1i} - \hat{f}_1^r}{\hat{f}_1^r} \right)^2 + \sum_{j=1}^{n_2} \left(\frac{y_{2j} - \hat{f}_2^r}{\hat{f}_2^r} \right)^2 \right].$$

   The f_1^r and f_2^r denote the functions $f_1(x_{1i}, \theta)$ and $f_2(x_{2i}, \theta)$ evaluated at the restricted maximum likelihood estimate $\theta^T = (\hat{\phi}^T(\gamma_0), \gamma_0)$.

7. Partition Σ as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix} \tag{5.27}$$

8. Let σ_0 be the value of Σ_{22}(σ_0 is a scalar for our problem) evaluated at $(\hat{\phi}(\gamma_0), \gamma_0)$.

[4]Note that we estimate σ by equating the Pearson Chi Squared to its degrees of freedom from the restricted model. Using $(n_1 + n_2 - 5)$ as the degrees of freedom could be justified following the same steps as in the Theorem 1. Dean [24] suggests using the pseudo maximum likelihood estimate for σ, which uses $(n_1 + n_2)$.

9. The quasi-score test statistic for testing the hypothesis $H_0 : \gamma = \gamma_0$ is $u_0^2\hat{\sigma}_0$, and is distributed as a chi squared distribution on 1 degree of freedom.

10. Thus, a $(1-\alpha)100\%$ confidence interval for the equivalent dose with confidence coefficient α is

$$\left\{\gamma_0 \,\middle|\, u_0^2\hat{\sigma}_0 \leq \chi^2_{1,\alpha}\right\},$$

where $\chi^2_{1,\alpha}$ is the upper α quantile for a chi squared distribution with 1 degree of freedom.

5.3.9 Finite sample performance of the confidence intervals based on the score test

Now we describe the results of a Monte Carlo study that we performed to examine the finite sample performance of the confidence intervals based on the quasi-score test. We fixed the parameters at

$$\theta = (14.2853, 123.182, 393.065, 192.547, 756.620, -87.45)^T.$$

The dose vectors used in the study are presented in Table 9.1 of Appendix 9.3 where they are coded as indicated in the column 'Dose' of Table 5.6. For each simulated sample, a confidence interval for the equivalent dose was computed by inverting the quasi-score test as described in Section 5.3.8. The results of the study are given in the Table 5.6. In Table 5.6, the column 'Qscore' indicates the observed coverage of the confidence intervals obtained by inverting the quasi-score test; the column t indicates the coverage probability of the t intervals based on the quasi-likelihood estimate.

Conclusions

The following conclusions were drawn from the simulation study:

1. The coverage probabilities of the confidence intervals obtained by inverting the quasi-score test agree with their nominal coverages even for small samples.

2. The coverage probabilities of t intervals based on the quasi-likelihood estimates also agree with their nominal coverages.

Dose	σ	nominal	Qscore	t	n_1	n_2
(P1)	0.01	0.95	0.9528	0.9518	16	13
(P1)	0.02	0.95	0.9547	0.9518	16	13
(P1)	0.02	0.90	0.8974	0.8970	16	13
(P1)	0.02	0.99	0.9936	0.9903	16	13
(P1)	0.029	0.95	0.9572	0.9538	16	13
(P1)	0.029	0.90	0.9043	0.9082	16	13
(P1)	0.029	0.99	0.9948	0.9886	16	13
(P2)	0.029	0.95	0.9478	0.9470	26	23
(P2)	0.029	0.90	0.8964	0.8969	26	23
(P2)	0.029	0.99	0.9923	0.9887	26	23
(P3)	0.029	0.95	0.9481	0.9473	50	50

Table 5.6: Coverage probabilities of quasi-score and t intervals

5.3.10 Generalized least squares and data weighted least squares

The procedures described in the Sections 3.4 and 3.5 could simply be extended to obtain the generalized least squares and the data weighted least squares estimates for parameters in the simultaneous curve fitting.

Generalized least squares:

The generalized least squares estimates minimizes the weighted error sum of squares

$$S(\theta) = \sum_{i=1}^{n_1} \frac{[y_{1i} - f_1(x_{1i}, \theta)]^2}{f_1^2(x_{1i}, \theta)} + \sum_{j=1}^{n_2} \frac{[y_{2j} - f_2(x_{2j}, \theta)]^2}{f_2^2(x_{2j}, \theta)}.$$

Let n be the total number of observations and $y = (y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2})$ be the vector of unbleached data followed by the bleached data. Let x and f respectively denote the vectors of corresponding doses and response function values.

The estimates $\hat{\theta}$ solve the system of equations

$$0 = \frac{\partial S}{\partial \theta} = \sum_{i=1}^{n} \frac{\left[y_i - f(x_i, \hat{\theta}\right]}{f(x_i, \hat{\theta})^2} \bigtriangledown f_{\hat{\theta}} + \sum_{i=1}^{n} \frac{\left[y_i - f(x_i, \hat{\theta})\right]^2}{f(x_i, \hat{\theta})^3} \bigtriangledown f_{\hat{\theta}}.$$

We used the IMSL subroutine 'NEQNF' to solve the above set of equations. When the unbleached and bleached response curves correspond to a common error factor σ, we estimate the parameter σ^2 by its approximately unbiased estimate $\hat{\sigma}^2 = \frac{1}{(n-6)} \sum_{i=1}^{n} \frac{(y_i - \hat{f}_i)^2}{\hat{f}_i^2}$. If the two curves correspond to different error factors σ_1 and σ_2, we estimate them as $\hat{\sigma_1}^2 = \frac{1}{(n_1-3)} \sum_{i=1}^{n_1} \frac{(y_{1i} - \hat{f}_{1i})^2}{\hat{f}_{1i}^2}$ and $\hat{\sigma_2}^2 = \frac{1}{(n_2-3)} \sum_{i=1}^{n_2} \frac{(y_{2i} - \hat{f}_{2i})^2}{\hat{f}_{2i}^2}$ respectively.

Let W be the diagonal matrix with ith entry $w_{ii} = 1/f_i^2$, for $i = 1, \ldots, n$. Let F be the $n \times p$ matrix with (i,j)th entry, the derivative of f_i with respect to the jth component of θ. The standard error of the generalized least squares estimate $\hat{\gamma}$ is estimated as $\hat{\sigma} s$, where s denotes the corresponding diagonal element of the matrix $\left(F^T W F\right)^{-1}$, evaluated at $\hat{\gamma}$. These standard errors are appropriate only in the small σ limit, since otherwise the estimates are not consistent.

Data weighted least squares estimate:

The data weighted least squares estimate for γ minimizes the weighted error sum of squares

$$S(\theta) = \sum_{i=1}^{n} \left\{ \frac{[y_i - f(x_i, \theta)]^2}{y_i^2} \right\}.$$

The iterative scheme described for obtaining quasi-likelihood estimates can be used to obtain the data weighted least squares estimates by replacing the diagonal elements of the weight matrix W by the observed $1/y^2$ values.

The standard error of the estimate $(s_{\hat{\gamma}})$ is estimated as $\hat{\sigma} s$, where s denotes the corresponding diagonal element of the matrix $\left(F^T W F\right)^{-1}$, evaluated at the data weighted least squares estimate. These standard errors are appropriate only in the small σ limit, since otherwise the estimates are not consistent.

Confidence intervals based on least squares estimates:

The confidence intervals for the equivalent dose can be constructed by taking $\hat{\gamma} \mp t_{df,1-\alpha/2} s_{\hat{\gamma}}$ as lower and upper confidence limits. For the single error factor case, the degrees of freedom

Study (dose)	σ	nominal	GLS	DWLS	n_1	n_2
(P1)	0.01	0.95	0.9520	0.9529	16	13
(P1)	0.02	0.95	0.9527	0.9510	16	13
(P2)	0.029	0.95	0.9466	0.9460	26	23
(P2)	0.04	0.95	0.9498	0.9506	26	23
(P2)	0.05	0.95	0.9456	0.9456	26	23

Table 5.7: Coverage probabilities of t intervals based on GLS and DWLS estimates

df is n_1+n_2-6 while for the different error factor case is computed using the Satterthwaite's formula (Equation 5.17). The use of these confidence intervals in large samples is justified by Theorems 3 and 4 of Section 5.2.4.

5.3.11 Finite sample performance of the confidence intervals

We performed a study similar to that which we described for the quasi-likelihood estimates (Section 5.2.5) to examine the performance of the confidence intervals based on the generalized squares and the data weighted least squares. The results of the simulation study are summarized in Table 5.7.

Based on the simulation results, we conclude that the coverage probabilities of the confidence intervals based on the generalized least squares and the data weighted least squares estimates agree with their nominal coverages even for small samples.

5.3.12 Biases of the estimators

Let $\theta = (\alpha_1, \alpha_2.\alpha_3, \beta_2, \beta_3, \gamma)$ be the vector of parameters in the simultaneous curve fitting.

We denote the vector of first derivatives and the matrix of second derivatives of f with respect to the parameter vector θ by $\bigtriangledown f$ and H. In each term, we use the subscripts 1 and 2 to indicate the unbleached and bleached data sets.

The results derived in Chapter 3 can easily be extended to arrive at the formulae for the biases of the estimators given in the Table 5.8.

Method of Estimation	Bias	$Var(\hat{\theta})$
ML	$\Sigma\left\{-W_1+\frac{p}{n}\left(\frac{\nabla f}{f}\right)-\frac{1}{2}W_2\right\}\sigma^2$	$\Sigma\sigma^2$
QL	$\Sigma\left\{-\frac{1}{2}W_2\right\}\sigma^2$	$\Sigma\sigma^2$
GLS	$\Sigma\left\{\left(\frac{\nabla f}{f}\right)-W_1-\frac{1}{2}W_2\right\}\sigma^2$	$\Sigma\sigma^2$
DWLS	$\Sigma\left\{-2\left(\frac{\nabla f}{f}\right)+2\,W_1-\frac{1}{2}W_2\right\}\sigma^2$	$\Sigma\sigma^2$

Table 5.8: Formulae for the biases and the variances of the estimators: partial bleach method

In Table 5.8,

$$\left(\frac{\nabla f}{f}\right) = \sum_{i=1}^{n_1}\left(\frac{\nabla f_1}{f_1}\right)_i + \sum_{j=1}^{n_2}\left(\frac{\nabla f_2}{f_2}\right)_j$$

$$\Sigma = \left[\sum_{i=1}^{n_1}\left(\frac{\nabla f_1}{f_1}\right)_i\left(\frac{\nabla f_1}{f_1}\right)_i^T + \sum_{j=1}^{n_2}\left(\frac{\nabla f_2}{f_2}\right)_i\left(\frac{\nabla f_2}{f_2}\right)_j^T\right]^{-1}$$

$$= \text{Variance covariance matrix of } C_1$$

$$p = \text{Total number of parameters}$$

$$n = n_1 + n_2 = \text{Total number of observations}$$

$$W_1 = \sum_{i=1}^{n_1} w_{11,i}\left(\frac{\nabla f_1}{f_1}\right)_i + \sum_{j=1}^{n_2} w_{12,j}\left(\frac{\nabla f_2}{f_2}\right)_j$$

$$W_2 = \sum_{i=1}^{n_1} w_{21,i}\left(\frac{\nabla f_1}{f_1}\right)_i + \sum_{j=1}^{n_2} w_{22,j}\left(\frac{\nabla f_2}{f_2}\right)_j$$

where $w_{11,i} = tr\left[\left(\frac{\nabla f_1}{f_1}\right)_i\left(\frac{\nabla f_1}{f_1}\right)_i^T\Sigma\right]$

$$w_{12,j} = tr\left[\left(\frac{\nabla f_2}{f_2}\right)_j\left(\frac{\nabla f_2}{f_2}\right)_j^T\Sigma\right]$$

$$w_{21,i} = tr\left[\left(\frac{H_1}{f_1}\right)_i\Sigma\right]$$

$$w_{22,j} = tr\left[\left(\frac{H_2}{f_2}\right)_j\Sigma\right]$$

p = Number of components of θ

n = Total number of observations = $n_1 + n_2$

Remarks:

1. The two stage curve fitting and the simultaneous curve fitting yield the same estimator for the equivalent dose.

2. The biases computed from the formulae given in the Table 5.8 are identical to the biases computed using Equation 5.16 for the simultaneous curve fitting.

5.4 Comparison of the estimators in finite samples

In Chapter 4, we compared maximum likelihood, quasi-likelihood, generalized least squares and data weighted least squares estimators by investigating their behavior in large samples, for fixed σ. We found that maximum likelihood and quasi-likelihood estimators are consistent while generalized least squares and data weighted least squares estimators are generally not. Least squares estimators were found to have biases that do not vanish, even asymptotically. However, we found that except for specific parameters, the component of the bias that does not vanish asymptotically is essentially zero for certain forms of response functions in models with variance function proportional to the square of the mean function. Therefore, for these models generalized least squares and data weighted least squares also provide asymptotically unbiased estimators, except for those specific parameters (see Chapter 4). It is easy to see from the results derived in Chapter 4 that the least squares estimators for γ, the parameter of our interest are asymptotically unbiased.

Next we describe the results of a small study that examines the biases of the estimators in the model for the partial bleach method. For the study described here the parameters were set at $\alpha_1 = 2.121383, \alpha_2 = 0.583, \alpha_3 = 5.964, \beta_2 = 0.68, \beta_3 = 6.67, \gamma = -0.48$ and $\sigma = 0.029$. The dose levels were fixed at, (0,1,2,4,8,16). The procedure is similar to that which we described for the additive dose method in Section 4.3.

Conclusions

Based on the results presented in Tables 5.9 - 5.11, we draw the following conclusions.

n	$\alpha_1 \times 10^{-5}$				$\alpha_2 \times 10^{-6}$			
	ML	QL	GLS	DWLS	ML	QL	GLS	DWLS
12	64.94	141.06	243.36	-63.58	-1.33	274.31	-1.39	825.40
24	32.48	70.52	210.87	-210.17	-0.64	137.16	-0.73	412.90
48	16.33	35.27	194.64	-283.41	-0.41	68.56	-0.42	207.02
96	8.10	17.61	186.50	-320.16	-0.30	34.15	-0.28	102.97
192	4.05	8.80	182.46	-338.57	-0.07	17.01	-0.13	51.10

Table 5.9: Comparison of the biases of $\hat{\alpha}_1$ and $\hat{\alpha}_2$: partial bleach method

n	$\alpha_3 \times 10^{-3}$				$\beta_2 \times 10^{-5}$			
	ML	QL	GLS	DWLS	ML	QL	GLS	DWLS
12	5.88	8.37	5.88	13.35	2.47	35.76	2.47	102.36
24	2.94	4.19	2.94	6.68	1.23	17.89	1.24	51.18
48	1.47	2.09	1.47	3.34	0.62	8.94	0.62	25.59
96	0.73	1.04	0.73	1.67	0.31	4.47	0.31	12.81
192	0.37	0.52	0.37	0.83	0.15	2.24	0.16	6.36

Table 5.10: Comparison of the biases of $\hat{\alpha}_3$ and $\hat{\beta}_2$: partial bleach method

n	$\beta_3 \times 10^{-3}$				$\gamma \times 10^{-3}$			
	ML	QL	GLS	DWLS	ML	QL	GLS	DWLS
12	8.21	11.45	8.21	17.93	-1.85	-2.07	-1.85	-2.51
24	4.10	5.72	4.10	8.97	-0.93	-1.04	-0.93	-1.26
48	2.05	2.86	2.05	4.48	-0.46	-0.52	-0.46	-0.63
96	1.03	1.43	1.03	2.24	-0.23	-0.26	-0.23	-0.31
192	0.51	0.72	0.51	1.12	-0.12	-0.13	-0.12	-0.16

Table 5.11: Comparison of the biases of $\hat{\beta}_3$ and $\hat{\gamma}$: Partial bleach method

1. For fixed σ, as n increases, the biases of maximum likelihood and quasi-likelihood estimators for all six parameters converge to zero at a rate $O(1/n)$.

2. Except for the parameter α_1, the biases of generalized least squares and data weighted least squares estimators also converge to zero at a rate $O(1/n)$, as n is increased while σ is fixed.

3. For all the parameters except α_1 the generalized least squares estimator and the maximum likelihood estimator have almost the same bias.

4. For the parameter α_1, the absolute values of the biases are in the order $B_{ML} < B_{QL} < B_{GLS}$. Depending on the sample size n, the absolute value of the bias of data weighted least squares estimator could be larger or smaller than the biases of the other estimators.

5. For all the parameters except α_1, the absolute values of the biases are in the order $B_{ML} \approx B_{GLS} < B_{QL} < B_{DWLS}$.

Remarks:

1. The conclusions drawn above agree with the theory discussed in Chapter 4.

2. We investigated the above results using different parameter vectors and different dose levels. The results were similar to what we described here.

3. From the formulae derived for the biases of the estimators, it is clear that for all six parameters the bias of data weighted least squares estimator (B_{DWLS}) is related to the biases of quasi-likelihood estimator (B_{QL}) and the generalized least squares estimator (B_{GLS}) according to $B_{DWLS} = 3B_{QL} - 2B_{GLS}$.

4. The results of the study show that for the parameter of interest, γ, maximum likelihood and the generalized least squares estimators have almost the same bias.

5.5 Worked example

Now we demonstrate the theoretical results derived in this chapter using a real data set. The data set used here is presented in Berger *et. al.* [12] where it is coded as 'QNL84-2'. We fitted the model $y = f(x,\theta)(1+\sigma\epsilon)$, using the techniques described earlier in this chapter. Here $f(x,\theta)$ is the saturating exponential model defined by

$$f(x,\theta) = \begin{cases} \alpha_1\left[1-\exp\left(-\frac{(x+\alpha_2)}{\alpha_3}\right)\right], & \text{(for unbleached data),} \\ \beta_1\left[1-\exp\left(-\frac{(x+\beta_2)}{\beta_3}\right)\right], & \text{(for the bleached data);} \end{cases}$$

here $\theta = (\alpha_1,\alpha_2,\alpha_3,\beta_2,\beta_3,\gamma)^T$ and $-\gamma$ denotes the equivalent dose (ED), and

$$\beta_1 = \alpha_1 \frac{\left\{1-\exp\left[-\frac{(\gamma+\alpha_2)}{\alpha_3}\right]\right\}}{\left\{1-\exp\left[-\frac{(\gamma+\beta_2)}{\beta_3}\right]\right\}}.$$

The parameter estimates for this data set assuming a common σ for the two curves are given in Tables 5.12 and 5.13.

	ML			QL		
parameter	estimate	bias	std.error	estimate	bias	std.error
$\alpha_1 \times 10^{-4}$	14.28	0.02	0.49	14.28	0.03	0.55
α_2	123.18	0.12	7.26	122.74	0.24	8.12
α_3	393.07	1.64	33.11	392.0	2.46	37.04
β_2	192.55	0.39	13.97	193.37	0.72	15.80
β_3	756.62	11.20	105.46	761.65	16.21	120.06
ED	87.15	0.55	9.13	86.43	0.72	10.14
σ^2	0.0012			0.0015		

Table 5.12: ML and QL estimates assuming common σ: Data QNL84-2

For the common error factor case 95% confidence intervals for the equivalent dose γ based on the methods discussed earlier in the chapter are given in Table 5.14.

	GLS			DWLS		
parameter	estimate	bias	std.error	estimate	bias	std.error
$\alpha_1 \times 10^{-4}$	14.30	0.05	0.55	14.25	0.09	0.55
α_2	123.18	0.15	8.16	121.86	0.41	8.10
α_3	393.07	2.07	37.20	389.92	3.23	36.94
β_2	192.54	0.49	15.69	195.18	1.19	16.12
β_3	756.59	14.12	118.49	772.76	20.63	124.19
ED	87.16	0.70	10.26	84.98	0.77	9.97
σ^2	0.0015			0.0015		

Table 5.13: GLS and DWLS estimates assuming common σ: Data QNL84-2

Description	Lower bound	Upper bound
Profile likelihood	70.83	108.90
using F critical value (FCRIT1)	68.77	112.72
Z interval based on the ML	69.27	105.04
t interval based on the ML	65.95	108.36
Quasi score interval	67.83	112.10
t interval based on the QL	65.45	107.42
t interval based on the GLS	65.94	108.37
t interval based on DWLS	64.36	105.61

Table 5.14: Confidence intervals for ED assuming common σ: Data QNL84-2

We also analyzed the same data set assuming σ is different for the unbleached and bleached response curves. The parameter estimates for this case are presented in Tables 5.15 and 5.16.

	ML			QL		
parameter	estimate	bias	std.error	estimate	bias	std.error
$\alpha_1 \times 10^{-4}$	14.28	0.01	0.41	14.28	0.02	0.46
α_2	123.18	0.08	6.09	122.74	0.16	6.73
α_3	393.07	1.15	27.77	392.0	1.69	30.68
β_2	192.55	0.53	16.32	193.37	1.02	18.75
β_3	756.62	15.30	123.22	761.65	22.83	142.74
ED	87.20	-0.14	8.27	86.43	-0.17	9.16
σ_1^2	0.0008			0.0010		
σ_2^2	0.0016			0.0021		

Table 5.15: ML and QL estimates assuming different σ: Data QNL84-2

The 95% confidence intervals for the equivalent dose γ assuming unbleached and bleached curves correspond to different error factors are given in Table 5.17.

Remarks

The results of the analysis of this data set indicate that fitting a common relative error σ or fitting two different σ's for the unbleached and bleached curves do not change the parameter estimates; we demonstrate this below. However, as one might expect, the biases and the standard errors of the estimates depend on whether or not the two curves correspond to a common σ.

Data weighted least squares estimating equations do not involve σ. Therefore, the DWLS estimates are unchanged regardless of whether we fit a common σ or two different σ's for the unbleached and bleached curves. For all the other three methods, the estimating equations for the simultaneous curve fitting involve σ. For example, quasi-likelihood estimating

	GLS			DWLS		
parameter	estimate	bias	std.error	estimate	bias	std.error
$\alpha_1 \times 10^{-4}$	14.30	0.03	0.46	14.25	0.01	0.45
α_2	123.18	0.10	6.76	121.86	0.29	6.70
α_3	393.06	1.42	30.82	389.92	2.24	30.55
β_2	192.55	0.69	18.62	195.18	1.70	19.14
β_3	756.63	19.84	140.61	772.76	29.50	147.50
ED	87.15	-0.18	9.26	84.98	-0.17	9.00
σ_1^2	0.0010			0.0010		
σ_2^2	0.0021			0.0022		

Table 5.16: GLS and DWLS estimates assuming different σ: Data QNL84-2

Description	Lower bound	Upper bound
Profile likelihood	72.16	106.73
using F critical value (FCRIT1)	70.21	110.15
Z interval based on the ML	70.99	103.41
t interval based on the ML	70.09	104.31
t interval based on the QL	67.42	105.45
t interval based on the GLS	67.93	106.39
t interval based on DWLS	66.31	103.66

Table 5.17: Confidence intervals for ED assuming different σ: Data QNL84-2

equations for the simultaneous curve fitting for the different σ case are

$$\sum_{i=1}^{n_1} \frac{y_{1i} - f_{1i}}{\sigma_1^2 f_{1i}^2} \bigtriangledown f_{1i} + \sum_{j=1}^{n_2} \frac{y_{2j} - f_{2j}}{\sigma_2^2 f_{2j}^2} \bigtriangledown f_{1i} = 0.$$

To obtain the parameter estimates we use a 2-part iteration similar to what we described for obtaining maximum likelihood estimates. This means, we find the estimates for θ in an iterative fashion, each time upgrading σ_1 and σ_2 using current parameter estimates and solving estimating equations for θ by replacing unknown σ_1 and σ_2 by these estimates. At a glance, it appears that the parameter estimates depend on how we estimate σ_j $(j = 1, 2)$. However, we observed in the example that the estimates did not have such a dependence. Since this observation appeared puzzling, we examined this more carefully. For a fixed σ_1, σ_2 the QL and GLS estimating equations for θ are derivatives of a function (the likelihood for the gamma model or the weighted error sum of squares) which is being optimized. The location of the optimum is invariant under reparametrization of θ. When the curves are fitted separately this optimum clearly does not depend on σ_1, σ_2 (the estimating equations for θ for this case do not involve σ). For QL and GLS, the invariance then guarantees that the same conclusion holds for the simultaneous curve fitting. Thus, the quasi-likelihood and the generalized least squares estimates are unchanged regardless of whether we fit a common σ or not. Furthermore, QL and GLS estimates for θ do not depend on how we estimate σ, in particular, whether we use maximum likelihood or least squares estimates for σ. Turning to maximum likelihood, the situation is somewhat different. As for QL and GLS, the invariance property guarantees that the simultaneous curve fitting and two stage approaches yield the same estimates by the method of maximum likelihood. However, the maximum likelihood equations are coupled with the estimating equations for σ, even for two stage curve fitting. Therefore, the estimates depend on how we estimate σ. In other words, the estimates using maximum likelihood estimates for σ and least squares estimates for σ are not necessarily the same as for the QL and GLS. Furthermore, the estimates for σ using a common error factor are not necessarily the same as those using different error factors. However, for the illustrated example, the maximum likelihood estimates for the two cases, fitting a common σ and fitting two different σ's for unbleached and bleached curves were virtually the same to the degree of accuracy as reported here.

5.6 Discussion

In this chapter, we discussed the estimation of the equivalent dose from partial bleach data. We described two approaches for this purpose based on two different parameterizations of the model. Both approaches yield the same parameter estimates. The first approach fits response curves for the unbleached and bleached data sets separately and estimates the equivalent dose as the intersection of the two fitted curves. For this approach, Berger *et. al.* [11] describes interval estimation assuming a common error factor for the unbleached and bleached data sets. We extended their results to the case of different error factors. The second approach fits the two curves simultaneously treating the equivalent dose γ as a parameter.

For both approaches, we described algorithms for obtaining maximum likelihood, quasi-likelihood, generalized least squares and data weighted least squares estimates. Formulae were derived for the biases of the estimators. We found that if the relative error in a single measurement σ is small, which is usually the case with partial bleach data, the bias is negligible relative to the standard error.

Under maximum likelihood, we described computing profile likelihood intervals and symmetric intervals using z and t quantiles. We also discussed computing profile type confidence intervals using a F critical value. Under quasi-likelihood we discussed computing profile type confidence intervals by inverting the quasi-score test, and symmetric confidence intervals using t quantiles. For generalized least squares and data weighted least squares we described symmetric confidence intervals using t quantiles. The finite sample performance of the suggested confidence intervals were examined by a Monte Carlo study. The following conclusions were drawn from the simulation results:

1. The coverage probabilities of symmetric t intervals based on the maximum likelihood, quasi-likelihood, generalized least squares and the data weighted least squares estimates agree well with their nominal values even when the sample sizes are small.

2. When the sample sizes are small, the coverage probabilities of profile likelihood intervals and z intervals based on the maximum likelihood estimate were found to have

smaller coverage than their nominal values.

3. The coverage probabilities of profile type confidence intervals with a transformed F quantile agree well with their nominal values even in small samples.

4. The coverage probabilities of profile type confidence intervals based on the quasi-score test agree well with their nominal coverages even in small samples.

5. Large sample sizes are not common in TL studies. Based on the simulation results, we recommend using symmetric confidence intervals based on t quantiles, quasi-score intervals and profile likelihood intervals with F critical values as opposed to intervals based on the maximum likelihood estimate with z quantiles and profile likelihood intervals with χ^2 critical values.

6. The t type confidence intervals have the added advantage that they are easier to compute than profile type confidence intervals.

We also examined the robustness of the test with F critical values to the departures from normality of the data. For this study, photon counts were generated from a gamma distribution with mean $f(x_i, \theta)$ and variance $\sigma^2 f^2(x_i, \theta)$. A confidence interval for the equivalent dose was computed from each generated sample using an F critical value. We found that the observed coverages still agree well with the nominal values even if the sample sizes are small. However, for the values of σ occurring in practice, the gamma distributions in question have large shape parameters (shape parameter is $1/\sigma^2$) and so were very close to normal distributions.

From the small study we performed to examine the behavior of the biases of the estimators for fixed σ, as the sample size n becomes large, we found that for all the parameters in the simultaneous curve fitting, the biases of maximum likelihood and quasi-likelihood estimators converge to zero at a rate $O(1/n)$. The same is true for the generalized least squares and data weighted least squares estimators, except for the parameter α_1. For γ, the parameter of interest in thermoluminescence studies, the maximum likelihood estimator and the generalized least squares estimator were found to have almost the same bias.

We end this chapter with a brief description of the computational difficulties encountered in the simulation work. When σ is larger than about 0.04 and the sample sizes are small (less than 50), we found that the convergence of the programs that compute profile type confidence intervals is poor. Since we are dealing with small samples (n around 29) with a non-linear response curve involving about 6 parameters the poor convergence is not surprising. Since the estimates for σ from the real data sets are usually small, the poor convergence for large σ is not viewed as a serious limitation. Due to poor convergence when σ is large (> 0.04) the biases and the coverage probabilities were only examined using samples of size larger than 50.

It may be worthwhile exploring the possibilities for improving the convergence of the programs when σ is large. Perhaps, a transformation of the parameters that removes the parameter effects non-linearity as suggested by Bates and Watts [2] could improve the convergence. In this work, we did not explore these possibilities.

We found that the coverage probabilities of the profile likelihood intervals are lower than their nominal values when the sample sizes are small. Several authors have addressed this problem of narrowing the profile when nuisance parameters are estimated. (See McCullagh and Tibshirani [50], Cox and Reid [20], Fraser and Reid [31]). Further work needs to be done in the area of adjusting the profile likelihood for estimating the nuisance parameters in our problem.

Chapter 6

Equivalent dose from regeneration data

In the regeneration method, a portion of the sample is given a vigorous laboratory bleaching. As in the partial bleach method, this portion is used to define the bleached response curve. The other portion is used to define the unbleached dose response curve. The equivalent dose is defined as the dose shift required for the unbleached curve to match the bleached curve (Aitken [1], Huntley *et. al.* [38]). If such a dose shift does not match the two curves, then the two data sets do not represent the same curve and the equivalent dose is not estimated from such data sets (Huntley *et. al.* [38]). In this chapter, we describe estimating the equivalent dose from regeneration data.

In Section 6.1, we introduce the notation and review the mathematical models for regeneration data. Section 6.2 offers initial estimates for the parameters. The methodology developed for partial bleach data can be applied to regeneration data with slight modifications. Therefore, we do not intend to elaborate on the theory. We describe the modifications that are necessary and offer the simulation results. The maximum likelihood estimates for the parameters are obtained in sections 6.3. Profile likelihood intervals and symmetric confidence intervals based on the maximum likelihood estimate using Z and t critical values are discussed in Section 6.3.1. In Section 6.3, we also discuss confidence intervals using a

transformation of the likelihood ratio statistic with a transformed F critical value. Quasi-likelihood estimates for the model parameters are discussed in Section 6.4. In Section 6.4.1, we discuss symmetric confidence intervals based on the quasi-likelihood estimate with a t critical value and based on inverting the quasi-score test. In Section 6.5, we obtain generalized least squares and data weighted least squares estimates for the equivalent dose. In Section 6.5.1, we discuss symmetric confidence intervals based on the generalized least squares and the data weighted least squares estimates. The finite sample performance of the suggested confidence intervals are examined by a Monte Carlo study. Formulae for the biases of the estimators are provided in Section 6.6.

In Section 6.7, we describe the results of a study that examines the behavior of the biases of estimators for fixed σ, as the sample size n gets bigger. We developed software using the computing language FORTRAN to implement the suggested methodology. In Section 6.8, we demonstrate the suggested theory using a real data set. Section 6.9 summarizes the chapter.

6.1 Mathematical models for regeneration data

For regeneration data, we discuss fitting a model with a saturating exponential plus a linear component. We chose to use this response function for two reasons: first it is commonly used by physicists for regeneration data (physical motivation for using this response function will be described later) and second since the techniques we describe for fitting regeneration data closely follow those for the partial bleach data we get a slightly different illustration of the results we derived for general response functions. The model with saturating exponential plus a linear component is represented by the function

$$f(x,\theta) = \alpha_1 \left\{ 1 - \exp\left[-\frac{x+\alpha_2}{\alpha_3} \right] \right\} + \alpha_4 (x+\alpha_2),$$

where x is the added dose and $\theta = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^T$ is the vector of unknown parameters. The procedures we describe for fitting these models can easily be adapted with slight modifications to fit other forms of response functions.

Motivation for using saturating exponential plus a linear component

Suppose there are two different types of traps approaching saturation at different dose levels as illustrated in Figure 6.1. In Figure 6.1, the dashed lines indicate the response curves, if there were only one type of trap. The solid line indicates the response function we observe from the sample as a result of the two types of traps, one approaching saturation while the other is still at its trap filling stage so that the corresponding response curve is nearly linear. As discussed in Chapter 5, response curves for each type of trap can be approximated by a saturating exponential model. Consequently, the response curve for the photon counts observed from the sample can be well approximated by a saturating plus a linear component.

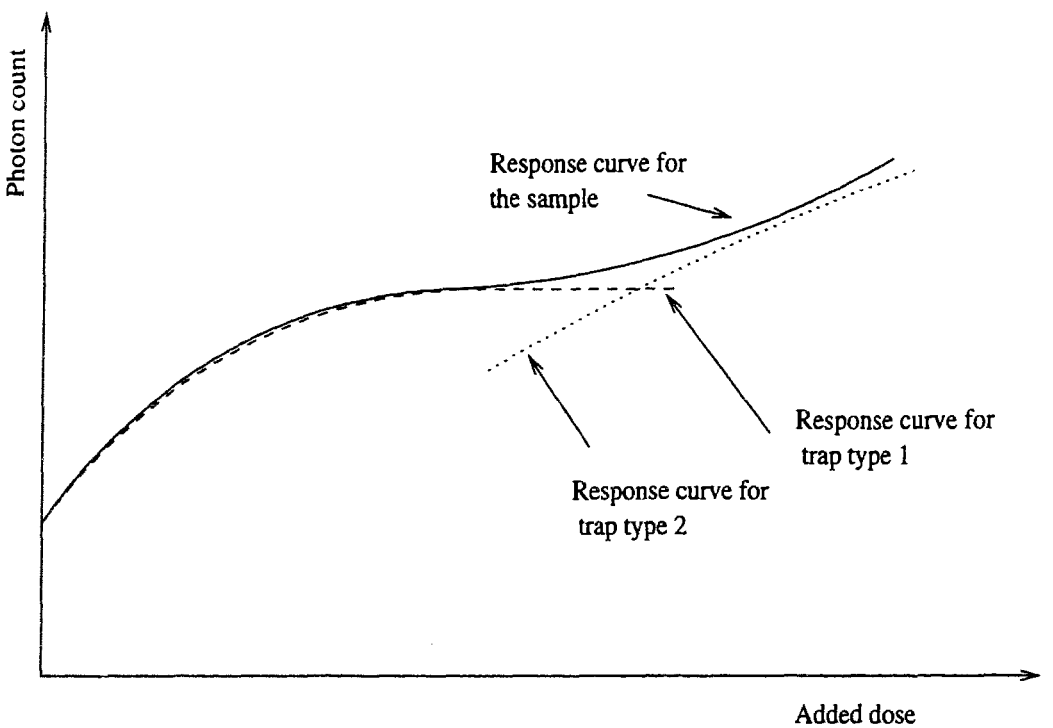

Figure 6.1: Plot of response curves for two different types of traps

Suppose data are collected on n_1 unbleached and n_2 bleached samples. Let x_{1i} and x_{2j} respectively indicate the doses received by the ith unbleached and jth bleached samples. For the unbleached data, we fit the response function

$$f_1(x_{1i}, \theta) = \alpha_1 \left\{ 1 - \exp\left[-\frac{(x_{1i} + \alpha_2)}{\alpha_3} \right] \right\} + \alpha_4 (x_{1i} + \alpha_2). \tag{6.1}$$

For the bleached data, we fit the response function

$$f_2(x_{2j},\theta) = \alpha_1\alpha_5\left\{1-\exp\left[-\frac{(x_{2j}+\alpha_2+\gamma)}{\alpha_3}\right]\right\}+\alpha_5\alpha_4(x_{2j}+\alpha_2+\gamma). \qquad (6.2)$$

The parameter α_5 is called a TL intensity scaling factor. Let $\theta = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \gamma)$ be the vector of unknown true parameters.

Huntley *et. al.* [38] reports that the equivalent dose can only be estimated if the unbleached and the bleached response curves could be matched by a shift along the dose axis. This is only possible if the TL intensity scaling factor, α_5, is unity and the parameter α_3 is the same for both curves. Therefore, we adhere to the following guidance for estimating the equivalent dose from the regeneration data.

1. Fit the response functions described by 6.1 and 6.2 for the unbleached and bleached data sets respectively.

2. Perform a formal hypothesis test to decide whether the intensity scaling factor is significantly different from unity.

3. If the intensity scaling factor is not significantly different from unity, re-fit the data with it fixed at unity.

4. From the re-fitted model, estimate the dose shift γ required for the unbleached response curve to match the bleached response curve. This dose shift γ estimates the equivalent dose.

5. If the intensity scaling factor is significantly different from unity the equivalent dose cannot be estimated from the given data (Huntley *et. al.* [38]).

Remarks:

1. The regeneration method for estimating the equivalent dose assumes that the laboratory bleaching does not cause any sensitivity change of the sample. According to Huntley *et. al.* [38], if the TL intensity (B) obtained from the bleached curve at the zero dose (Figure 6.2) is not equal to the thermoluminescence had it been measured at

the time of deposition it indicates a sensitivity change in the sample. The latter TL, while is unknown, is assumed to be the same as that measured for a modern dune (M) from the same environment. Consequently, if the points B and M do not coincide it is considered as an indication of sensitivity change due to laboratory bleaching. If there is an indication of such sensitivity change, the deduced dose has to be corrected for this. Huntley *et. al.* [38] suggests the correction to be ζ, which is the dose reading from the bleached response curve corresponding to the TL intensity of the modern dune (M). (Figure 6.2).

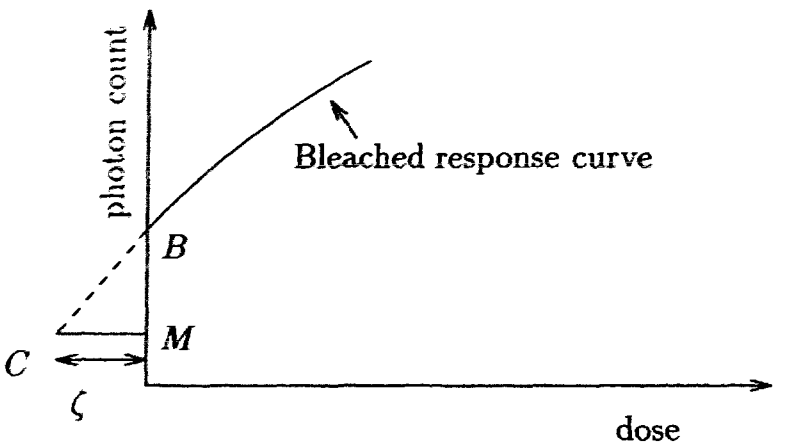

Figure 6.2: Correction for sensitivity change

2. Following the work of Huntley *et. al* [38], the hypothesis that *the unbleached and bleached curves represent the same curve* was examined by testing if the TL intensity scaling factor α_5 is not significantly different from unity. Recall that, when writing the response function for the bleached data we already assumed that α_3 and α_4 are common for the unbleached and bleached response functions. The guidance we followed therefore imposes the restriction that the parameters α_3 and α_4 do not change due to bleaching of the sample. While this assumption allows us to test the hypothesis that the two curves represent the same curve shifted along the dose axis with more power this assumption is not crucial for testing this. In Section 6.3, we describe procedures for testing whether or not the two curves represent the same curve without imposing restrictions on α_3 and α_4.

6.2 Initial estimates

Consider the response function for regeneration data. Note that for large x, f_1 is close to a straight line with slope and intercept:

$$\text{slope} = r_{1.1} = \alpha_4, \tag{6.3}$$

$$\text{intercept} = r_{2.1} = (\alpha_1 + \alpha_2\alpha_4). \tag{6.4}$$

For small x values, f_1 is close to a straight line with the slope and the intercept:

$$\text{slope} = r_{3.1} = \frac{\alpha_1}{\alpha_3} + \alpha_4, \tag{6.5}$$

$$\text{intercept} = r_{4.1} = \frac{\alpha_1\alpha_2}{\alpha_3} + \alpha_4\alpha_2. \tag{6.6}$$

For the bleached data the corresponding terms are:

$$r_{1.2} = \alpha_5\alpha_4, \tag{6.7}$$

$$r_{2.2} = (\alpha_5\alpha_1 + \alpha_5\alpha_4\gamma) + \alpha_5\alpha_4\alpha_2, \tag{6.8}$$

$$r_{3.2} = \frac{\alpha_5\alpha_1}{\alpha_3} + \alpha_5\alpha_4, \tag{6.9}$$

$$\text{and } r_{4.2} = \frac{\alpha_5\alpha_1\alpha_2}{\alpha_3} + \frac{\alpha_5\alpha_1\gamma}{\alpha_3} + \alpha_5\alpha_4\alpha_2 + \alpha_5\alpha_4\gamma. \tag{6.10}$$

Solving the equations 6.3 - 6.10 we get

$$\begin{aligned}
\alpha_1 &= r_{2.1} - \frac{r_{1.1}r_{4.1}}{r_{3.1}}, \\
\alpha_2 &= \frac{r_{4.1}}{r_{3.1}}, \\
\alpha_3 &= \frac{(r_{2.1} - r_{1.1}r_{4.1}/r_{3.1})}{(r_{3.1} - r_{1.1})}, \\
\alpha_4 &= r_{1.1}, \\
\alpha_5 &= \frac{r_{3.2}}{r_{3.1}}, \\
\text{and } \gamma &= \frac{r_{4.2}}{r_{3.2}} - \frac{r_{3.2}}{r_{3.1}}\frac{r_{4.1}}{r_{3.2}}.
\end{aligned}$$

Let

$y_1[1]$ = average photon count corresponding to the zero dose for the

unbleached data,

$x_1[2]$ = smallest positive dose for the unbleached data,

$y_1[2]$ = average photon count corresponding to $x_1[2]$,

$x_1[n_1]$ = largest dose for the unbleached data,

$x_1[(n_1-1)]$ = second largest dose for the unbleached data,

$y_1[n_1]$ = average photon count corresponding to $x_1[n_1]$,

$y_1[(n_1-1)]$ = average photon count corresponding to $x_1[(n_1-1)]$,

$y_2[1]$ = average photon count corresponding to zero dose for the bleached data,

$x_2[2]$ = smallest positive dose for the bleached data,

$y_2[2]$ = average photon count corresponding to $x_2[2]$,

$x_2[n_2]$ = largest dose for the bleached data,

$x_2[(n_2-1)]$ = second largest dose for the bleached data,

$y_2[n_2]$ = average photon count corresponding to $x_2[n_2]$,

and $y_2[(n_2-1)]$ = average photon count corresponding to $x_2[(n_2-1)]$.

We estimate the slopes and the intercepts as follows:

$$\left.\begin{aligned} r_{1.1} &= \tfrac{(y_1[n_1]-y_1[(n_1-1)])}{(x_1[n_1]-x_1[(n_1-1)])}, \\ r_{1.2} &= \tfrac{(y_2[n_2]-y_2[(n_2-1)])}{(x_2[n_2]-x_2[(n_2-1)])}, \\ r_{2.1} &= y_1[n_1] - r_{1.1}\, x_1[n_1], \\ r_{2.2} &= y_2[n_2] - r_{1.2}\, x_2[n_2], \\ r_{3.1} &= \tfrac{(y_1[2]-y_1[1])}{x_1[2]}, \\ r_{3.2} &= \tfrac{(y_2[2]-y_2[1])}{x_2[2]}, \\ r_{4.1} &= y_1[1], \\ \text{and} \quad r_{4.2} &= y_2[1]. \end{aligned}\right\} \qquad (6.11)$$

The initial estimates for the parameters are obtained by replacing the slopes and intercepts by their estimated values using equation 6.11.

6.3 Maximum likelihood estimates

We refer to the model with the parameter α_5 unspecified as the full model. The model with α_5 fixed at unity is called the constrained model. The parameter estimates for the constrained model are referred to as restricted maximum likelihood estimates.

To fit regeneration data, we modify the the algorithm described in Section 5.3.1 as follows.

Maximum likelihood estimates

The columns of the $n \times 6$ matrix D (Section 5.3.1) are replaced by the columns $(\frac{\partial f_i}{\partial \alpha_1}, \frac{\partial f_i}{\partial \alpha_2}, \frac{\partial f_i}{\partial \alpha_3}, \frac{\partial f_i}{\partial \alpha_4}, \frac{\partial f_i}{\partial \alpha_5}, \frac{\partial f_i}{\partial \gamma})$ given below:

$$
\begin{aligned}
\frac{\partial f_i}{\partial \alpha_1} &= 1 - \exp\left[\frac{-(x_{1i} + \alpha_2)}{\alpha_3}\right], \quad (\text{for } 1 < i \leq n_1) \\
&= \alpha_5 \left(1 - \exp\left[\frac{-(x_{2i} + \alpha_2 + \gamma)}{\alpha_3}\right]\right), \\
&\quad (\text{for } n_1 < i \leq n_1 + n_2); (k = (i - n_1)) \\
\frac{\partial f_i}{\partial \alpha_2} &= 1 - \exp\left[\frac{-(x_{1i} + \alpha_2)}{\alpha_3}\right] + \alpha_4 \ (\text{for } 1 < i \leq n_1) \\
&= \alpha_5 \left(1 - \exp\left[\frac{-(x_{2i} + \alpha_2 + \gamma)}{\alpha_3}\right]\right) + \alpha_5 \alpha_4, \\
&\quad (\text{for } n_1 < i \leq n_1 + n_2); (k = (i - n_1)) \\
\frac{\partial f_i}{\partial \alpha_3} &= -\frac{\alpha_1}{\alpha_3^2}(x_{1i} + \alpha_2) \exp\left[\frac{-(x_{1i} + \alpha_2)}{\alpha_3}\right], \quad (\text{for } 1 < i \leq n_1) \\
&= -\frac{\alpha_1 \alpha_5}{\alpha_3^2}(x_{2i} + \alpha_2 + \gamma) \exp\left[\frac{-(x_{2i} + \alpha_2 + \gamma)}{\alpha_3}\right], \\
&\quad (\text{for } n_1 < i \leq n_1 + n_2); (k = (i - n_1)) \\
\frac{\partial f_i}{\partial \alpha_4} &= (x_{1i} + \alpha_2) \ (\text{for } 1 < i \leq n_1) \\
&= \alpha_5 (x_{2j} + \alpha_2), \quad (\text{for } n_1 < i \leq n_1 + n_2); (k = (i - n_1)), \\
\frac{\partial f_i}{\partial \alpha_5} &= 0 \ \ (\text{for } 1 < i \leq n_1) \\
&= \alpha_1 \left(1 - \exp\left[-\frac{(x_{2i} + \alpha_2 + \gamma)}{\alpha_3}\right]\right) + \alpha_4 (x_{2i} + \alpha_2), \\
&\quad (\text{for } n_1 < i \leq n_1 + n_2); (k = (i - n_1)) \\
\frac{\partial f_i}{\partial \gamma} &= 0 \ (\text{for } 1 < i \leq n_1)
\end{aligned}
$$

$$= \frac{\alpha_1 \alpha_5}{\alpha_3} \exp\left[-\frac{(x_{2i} + \alpha_2 + \gamma)}{\alpha_3}\right].$$
$$(\text{for } n_1 < i \leq n_1 + n_2); (k = (i - n_1))$$

The restricted maximum likelihood estimates are obtained by fixing α_5 at unity. To compute the restricted maximum likelihood estimates, we replace the matrix D by the $n \times 5$ matrix $(\frac{\partial f_i}{\partial \alpha_1}, \frac{\partial f_i}{\partial \alpha_2}, \frac{\partial f_i}{\partial \alpha_3}, \frac{\partial f_i}{\partial \alpha_4}, \frac{\partial f_i}{\partial \gamma})$.

The standard errors of the maximum likelihood estimates can be computed as described in in Section 5.3.1.

Tests on the intensity scaling factor

Recall that the first step in estimating the equivalent dose from the regeneration data is to test the hypothesis $H_0 : \alpha_5 = 1$.

The likelihood ratio test

Let the values of the log-likelihood function corresponding to the full model and the restricted model be l_1 and l_2 respectively. Let $\chi^2_{1,\alpha}$ denotes the upper α quantile for a χ^2 distribution on 1 degree of freedom. If $2(l_1 - l_2) > \chi^2_{1,\alpha}$, with confidence coefficient $1 - \alpha$ we conclude that the intensity scaling factor α_5 is significantly different from unity.

The t test

Let $s_{\hat{\alpha}_5}$ be the standard error of $\hat{\alpha}_5$. Let $t = \frac{(\hat{\alpha}_5 - 1)}{s_{\hat{\alpha}_5}}$. If the absolute value of t is greater than $t_{\alpha/2, n_1+n_2-6}$, we conclude that the intensity scaling factor is significantly different from unity. Here $t_{\alpha,\nu}$ denotes the upper αth quantile for a t distribution with ν degrees of freedom.

Worked Example

We applied the suggested methodology on the data set SESA1 (cited in Huntley *et. al.* [38]). The results for the data collected at the temperature 350^0C are as follows. These data were kindly provided to us by D. J. Huntley.

Maximum likelihood

$$\begin{aligned}
\hat{\alpha}_5 &= 0.972, \quad s_{\hat{\alpha}_5} = 0.027 \\
l_1 &= -537.788, \quad l_2 = -538.318 \\
2(l_1 - l_2) &= 1.06 < \chi^2_{1,0.95}; \quad (p - value = 0.30) \\
t &= \left|\frac{(\hat{\alpha}_5 - 1)}{s_{\hat{\alpha}_5}}\right| = 1.053 < 2.00(= t_{0.975,56}); \quad (p - value = 0.15)
\end{aligned}$$

Both tests lead to the conclusion that the data do not provide enough evidence to reject the null hypothesis that the intensity scaling factor is unity. So we re-fit the model by setting the intensity scaling parameter at unity. From the fitted parameters for the restricted model, the maximum likelihood estimate for the equivalent dose = 73.06, with an error of the estimate $s_{\hat{\gamma}} = 3.23$.

Will a dose shift bring the two curves into coincidence?

Now we briefly describe test procedures for testing whether a dose shift would bring the two response curves into coincidence without imposing restrictions on any of the parameters. The procedures described earlier can be slightly modified to fit the models described below.

(a) For the unbleached data, fit the response function

$$f_1(x_{1i}, \theta) = \alpha_1 \left\{1 - \exp\left[-\frac{(x_{1i} + \alpha_2)}{\alpha_3}\right]\right\} + \alpha_4(x_{1i} + \alpha_2). \qquad (6.12)$$

For the bleached data, fit the response function

$$f_2(x_{2j}, \theta) = \beta_1 \left\{1 - \exp\left[-\frac{(x_{2j} + \alpha_2 + \gamma)}{\beta_3}\right]\right\} + \beta_4(x_{2j} + \alpha_2 + \gamma). \qquad (6.13)$$

(b) Test the hypothesis $\alpha_1 = \beta_1, \alpha_3 = \beta_3, \alpha_4 = \beta_4$ using the likelihood ratio test.

Likelihood ratio test

1. Fit the two models described by the equations 6.12 and 6.13 for the unbleached and bleached data respectively. We refer to this model as the full model.

2. Now re-fit the models subject to the restriction, $\alpha_1 = \beta_1, \alpha_3 = \beta_3, \alpha_4 = \beta_4$. We refer to this model as the restricted model; it is the same restricted model described in the previous section.

3. Let l_1 and l_2 be the values of the log-likelihood for the full model and the restricted model respectively.

4. If $2(l_1 - l_2) > \chi^2_{3,\alpha}$ using a significance level α we conclude that the two curves do not represent the same curve.

Worked Example

For the data set SESA1 cited earlier we found that the parameter estimates for the full model are:

$\hat{\alpha}_1 = 78594.0, \hat{\alpha}_2 = 85.72, \hat{\alpha}_3 = 128.19, \hat{\alpha}_4 = 74.97, \hat{\gamma} = -79.51, \hat{\beta}_1 = 54043.80,$

$\hat{\beta}_3 = 74.63$ and $\hat{\beta}_4 = 121.84$.

The standard errors are:

$s_{\hat{\alpha}_1} = 18146.10, s_{\hat{\alpha}_2} = 18.41, s_{\hat{\alpha}_3} = 59.53, s_{\hat{\alpha}_4} = 31.07, s_{\hat{\gamma}} = 18.41, s_{\hat{\beta}_1} = 5251.72,$

$s_{\hat{\beta}_3} = 10.28$ and $s_{\hat{\beta}_4} = 13.10$.

The log-likelihood for the full model is $l_1 = -536.042$.

The parameter estimates for the restricted model are:

$\hat{\alpha}_1 = 62150.1, \hat{\alpha}_2 = 79.6176, \hat{\alpha}_3 = 89.6115, \hat{\alpha}_4 = 105.644, \hat{\gamma} = -73.0640,$

The standard errors are:

$s_{\hat{\alpha}_1} = 4886.85, s_{\hat{\alpha}_2} = 3.41, s_{\hat{\alpha}_3} = 9.66, s_{\hat{\alpha}_4} = 11.12, s_{\hat{\gamma}} = 3.23.$

The log-likelihood for the restricted model is $l_2 = -538.318$.

Note that $2(l_1 - l_2) = 4.552 < \chi^2_{3,0.95}(= 7.81)$, $(\text{p}-value = 0.79)$. Therefore, at a level of significance $\alpha = 0.95$, we conclude that the two curves represent the same curve shifted over the dose axis.

The test that uses the additional information that α_3 and α_4 are common to the unbleached and bleached response curves is more powerful in detecting whether the intensity

scaling factor is unity than the test described here. Since physicists follow the former procedure we do not intend to follow the less powerful test procedure described here to determine whether the two curves follow the same curve; we only examine whether the intensity scaling factor is equal to unity. Similar tests as described in this section can be performed based on the quasi-likelihood estimate or least squares estimates.

6.3.1 Profile likelihood intervals

Recall that, in the regeneration method we estimate the equivalent dose from the restricted model. The full model is only used as a guide to decide whether or not we can estimate the equivalent dose from the given data. The simulation results we present here and elsewhere in this chapter are therefore obtained using the restricted model.

The Monte Carlo study is similar to what we already described for the partial bleach data. Since the response functions we consider for the regeneration data are different from those for the partial bleach data we need to modify the algorithms accordingly. The necessary modifications are similar to what we described for deriving maximum likelihood estimates. For example, to compute profile likelihood intervals, the $n \times 5$ matrix D' of Section 5.3.2 has to be replaced by the $n \times 4$ matrix D' with ith row consisting of the partial derivatives $(\frac{\partial f_i}{\partial \alpha_1}, \frac{\partial f_i}{\partial \alpha_2}, \frac{\partial f_i}{\partial \alpha_3}, \frac{\partial f_i}{\partial \alpha_4})$.

As for the partial bleach data, we examined the finite sample performance of the profile likelihood intervals, symmetric intervals with t and Z quantiles, and the likelihood ratio intervals with transformed F critical values. The parameters were set at

$$\theta = (57486.4, 76.3545, 86.9160, 95.3052, -68.8792)^T.$$

These are the maximum likelihood estimates for the parameters in the full model obtained for the data set SESA1 (collected at the temperature 360^0C) cited in Huntley *et. al.* [38]. The dose vectors used for the study are presented in Table 9.2 of Appendix 9.3 where it is labeled as indicated in the first column of Table 6.1.

The results based on 10000 simulations are given in Table 6.1. In Table 6.1, the column 'PROFILE' indicates the observed coverages for the profile likelihood intervals. The columns

'FCRIT1' and 'FCRIT2' indicate the observed coverages for the confidence intervals using the transformed likelihood ratio statistic LR (Section 5.3.4) with the critical values based on the approximations:

$$2LR = n\log\left[1+\frac{F_{1,(n-5)}}{(n-5)}\right] \tag{6.14}$$

$$\approx \frac{n}{(n-5)}F_{1,(n-5)} \tag{6.15}$$

The columns Z and t indicate the coverage probabilities of the symmetric confidence intervals based on the maximum likelihood estimate with quantiles of standard normal and student t with $n-5$ degrees of freedom respectively.

Dose	σ^2	nominal	PROFILE	FCRIT1	FCRIT2	Z	t	n_1	n_2
(R1)	0.002	0.95	0.9375	0.9507	0.9531	0.9331	0.9485	30	32
(R1)	0.004	0.95	0.9369	0.9475	0.9531	0.9321	0.9482	30	32
(R3)	0.002	0.95	0.9442	0.9502	0.9516	0.9415	0.9494	50	50
(R3)	0.004	0.95	0.9389	0.9454	0.9476	0.9357	0.9455	50	50
(R4)	0.002	0.95	0.9493	0.9527	0.9537	0.9476	0.9493	100	100
(R4)	0.004	0.95	0.9487	0.9514	0.9520	0.9468	0.9510	100	100
(R4)	0.006	0.95	0.9419	0.9456	0.9474	0.9411	0.9456	100	100

Table 6.1: Coverage probabilities of profile, F, Z and t intervals: regeneration data

Conclusions:

Based on the simulation results we recommend using t confidence intervals which are easier to compute and have coverage probabilities in agreement with the nominal coverages. If profile type confidence intervals are preferred, we recommend using a transformed F critical value as oppose to a χ^2 critical value.

6.4 Quasi-likelihood estimates

The algorithm described in Section 5.3.7 can be used to obtain the quasi-likelihood estimates by replacing the matrix F by the matrix D described in Section 6.3.

The quasi-likelihood estimates for the constrained model are obtained by fixing α_5 at unity. As in the case of restricted maximum likelihood estimates, the columns of the $n \times 5$ matrix F of partial derivatives have to be replaced by the columns $(\frac{\partial f_i}{\partial \alpha_1}, \frac{\partial f_i}{\partial \alpha_2}, \frac{\partial f_i}{\partial \alpha_3}, \frac{\partial f_i}{\partial \alpha_4}, \frac{\partial f_i}{\partial \gamma})$ to obtain the restricted quasi-likelihood estimates.

Testing the hypothesis $H_0 : \alpha_5 = 1$:

Let W be the weight matrix, which is diagonal with ith diagonal element equal to the weight on the ith observation, f_i^2. Let $\hat{\alpha}_5$ be the quasi-likelihood estimator for α_5. Let $s_{\hat{\alpha}_5}$ be the square root of the corresponding diagonal element of the matrix $\sigma^2(F^T W F)^{-1}$.

Proceeding as in Section 5.2.4, we find that $\frac{(\hat{\alpha}_5 - \alpha_5)}{s_{\hat{\alpha}_5}}$ has approximately a t distribution with $(n_1 + n_2 - 6)$ degrees of freedom. If the absolute value of $\frac{(\hat{\alpha}_5 - \alpha_5)}{s_{\hat{\alpha}_5}}$ is greater than the corresponding quantile of a t distribution with $(n_1 + n_2 - 6)$ degrees of freedom, we reject the hypothesis that the intensity scaling factor is unity, at the significance level α.

Worked Example continued.

Quasi-likelihood estimates:

We applied the suggested methodology to find the quasi-likelihood estimates for the data set SESA1 (Huntley *et. al.* [38]). The number of observations for the unbleached and bleached data sets are $n_1 = 30$ and $n_2 = 32$. The quasi-likelihood estimates for the model parameters are $\hat{\alpha}_5 = 0.970$, $s_{\hat{\alpha}_5} = 0.028$. Thus, $\left|\frac{(\hat{\alpha}_5 - 1)}{s_{\hat{\alpha}_5}}\right| = 1.062 < t_{56,0.975}(= 2.00)$; $(p - value = 0.15)$. So as in the method of maximum likelihood, we conclude that α_5 is not significantly different from unity. Therefore, we re-fit the data by fixing α_5 at unity. From the fitted parameters of the restricted model, quasi-likelihood estimate for the equivalent dose = 73.17, with a standard error of the estimate $s_{\hat{\gamma}} = 3.39$.

6.4.1 Quasi-score intervals

The procedure described in Chapter 5.3.8 can be used to obtain confidence intervals for the equivalent dose by inverting the quasi-score test. The necessary modifications are similar to those described for the profile likelihood intervals.

Symmetric confidence intervals for the equivalent dose can also be constructed based on the quasi-likelihood estimate using a t critical value. Since the confidence intervals are constructed based on the model with TL intensity scaling factor fixed at unity, for the common error factor case, the degrees of freedom for the t distribution is $(n_1 + n_2 - 5)$. When σ is assumed to be different for the two response curves the corresponding degrees of freedom can be obtained from Satterthwaite's approximation (Equation 5.17).

The results of a simulation study that examines the finite sample performance of these confidence intervals are given in Table 6.2. For this study the parameters were set at

$$\theta = (57486.4, 76.3545, 86.9160, 95.3052, -68.8792)^T.$$

These are the parameters used for examining confidence intervals based on the maximum likelihood estimate. The dose vectors used for the study are presented in Table 9.2 of Appendix 9.3 where they are labeled as indicated in the first column of Table 6.1. In this table, columns 'Qscore' and t indicate the observed coverages for the confidence intervals based on the quasi-score and symmetric t intervals.

Dose	σ^2	nominal	Qscore	t	n1	n2
(R1)	0.002	0.95	0.9448	0.9494	30	32
(R1)	0.004	0.95	0.9429	0.9478	30	32
(R3)	0.002	0.95	0.9467	0.9492	50	50
(R3)	0.004	0.95	0.9407	0.9424	50	50
(R4)	0.002	0.95	0.9493	0.9510	100	100
(R4)	0.004	0.95	0.9495	0.9510	100	100
(R4)	0.006	0.95	0.9449	0.9464	100	100
(R5)	0.002	0.95	0.9475	0.9487	125	125

Table 6.2: Coverage probabilities of quasi-score and t intervals

Conclusions:

Except for the quasi-score intervals in one case, the coverage probabilities of both t intervals and quasi-score intervals agree with the nominal coverages even in small samples. Based on the simulation results we conclude that the small sigma asymptotic results are valid for sigma in the range of real data sets collected for the regeneration method (The estimates for σ^2 in real data sets collected for the regeneration method are usually around 0.004.).

6.5 Generalized least squares and data weighted least squares

We commonly refer to the generalized least squares estimates and the data weighted least squares estimates as least squares estimates. The algorithms described for obtaining least squares estimates for the partial bleach data can be slightly modified to obtain least squares estimates for the regeneration data. Since the necessary modifications are similar to those described for the maximum likelihood estimates we do not elaborate.

Testing the hypothesis $H_0 : \alpha_5 = 1$:

Let $\hat{\alpha}_5$ be any of the least squares estimators for α_5. Let $s_{\hat{\alpha}_5}$ be the estimate for the error of the estimate obtained as described in Section 5.3.10 for the partial bleach data.

As in Section 5.2.4, it can be shown that $\frac{(\hat{\alpha}_5 - \alpha_5)}{s_{\hat{\alpha}_5}}$ has approximately a t distribution with $(n_1 + n_2 - 6)$ degrees of freedom. If the absolute value of $\frac{(\hat{\alpha}_5 - \alpha_5)}{s_{\hat{\alpha}_5}}$ is greater than the corresponding quantile of a t distribution with $(n_1 + n_2 - 6)$ degrees of freedom, we reject the hypothesis that the intensity scaling factor is unity, at the significance level α.

Worked Example continued.

We applied the suggested methodology to obtain the least squares estimates for the data set SESA1 cited in Huntley *et. al.* [38]. The complete data set was kindly provided to us by D.J. Huntley. A summary of the results are given below.

Generalized least squares

Estimate for α_5 is, $\hat{\alpha}_5 = 0.972$, with an error of estimate $s_{\hat{\alpha}_5} = 0.099$. Thus, $\left|\frac{(\hat{\alpha}_5 - 1)}{s_{\hat{\alpha}_5}}\right| =$

$0.287 < t_{56,0.975}(= 2.00)$; $(p - value = 0.39)$. So, as in the method of maximum likelihood, we conclude that α_5 is not significantly different from unity. Therefore, we re-fit the data by fixing α_5 at unity. From the fitted parameters of the restricted model, the generalized least squares estimate for the equivalent dose = 73.06, with an error of the estimate $s_{\hat{\gamma}} = 3.37$.

Data weighted least squares

The estimate for α_5 is, $\hat{\alpha}_5 = 0.967$, with a standard error $s_{\hat{\alpha}_5} = 0.029$. Thus, $\left|\frac{(\hat{\alpha}_5 - 1)}{s_{\hat{\alpha}_5}}\right| = 1.122 < t_{56,0.975}(= 2.00)$; $(p - value = 0.13)$. As before, we conclude that α_5 is not significantly different from unity and re-fit the data by fixing α_5 at unity. From the fitted parameters of the restricted model, data weighted least squares estimate for the equivalent dose = 73.56, with an error of the estimate $s_{\hat{\gamma}} = 3.50$.

6.5.1 Confidence intervals based on least squares estimates

Let $\hat{\gamma}$ be the least squares estimate and $s_{\hat{\gamma}}$ be the standard error of the estimate computed as described under each method. The confidence intervals for the equivalent dose can be constructed by taking $\hat{\gamma} \mp t_{df,1-\alpha/2} s_{\hat{\gamma}}$ as lower and upper confidence limits. For the single error factor case, the degrees of freedom df is $(n_1 + n_2 - 5)$. For regeneration data, we do not have justification for using different error factors. The use of these confidence intervals in large samples can be justified as in Theorems 3 and 4 of Section 5.2.4.

The results of a simulation study that examines the finite sample performance of these confidence intervals are given in Table 6.3. For this study the parameters were set at

$$\theta = (57486.4, 76.3545, 86.9160, 95.3052, -68.8792)^T.$$

These are the parameters used for examining confidence intervals based on the maximum likelihood estimate. The dose vectors used for the study are presented in Table 9.2 of Appendix 9.3 where they are labeled as indicated in the first column of Table 6.3.

Conclusions: Based on the simulation results we conclude that the small sigma asymptotic results for the generalized least squares and data weighted least squares estimates hold for σ values typical of real data sets.

Dose	σ^2	nominal	GLS	DWLS	n_1	n_2
(R1)	0.002	0.95	0.9486	0.9497	30	32
(R1)	0.004	0.95	0.9485	0.9469	30	32
(R3)	0.006	0.95	0.9494	0.9453	50	50
(R3)	0.008	0.95	0.9486	0.9415	50	50
(R3)	0.010	0.95	0.9446	0.9381	50	50

Table 6.3: Coverage probabilities of t intervals based on GLS and DWLS estimates

6.6 Biases of the estimators

6.6.1 Formulae for the biases

Let $\theta = (\alpha_1, \alpha_2.\alpha_3, \alpha_4, \gamma)^T$ denote the vector of unknown true parameters of the restricted model. The equivalent dose is estimated as $\hat{\gamma}$. We denote the vector of first derivatives and the matrix of second derivatives of f with respect to the parameter vector θ by $\bigtriangledown f$ and H. In each term, we use the subscripts 1 and 2 to indicate unbleached and bleached data sets.

The results derived in Chapter 3 can easily be extended to arrive at the following formula for the bias of $\hat{\gamma}$.

Method of Estimation	Bias	$Var(\hat{\theta})$
ML	$\Sigma\left\{-W_1 + \frac{p}{n}\left(\frac{\nabla f}{f}\right) - \frac{1}{2}W_2\right\}\sigma^2$	$\Sigma\sigma^2$
QL	$\Sigma\left\{-\frac{1}{2}W_2\right\}\sigma^2$	$\Sigma\sigma^2$
GLS	$\Sigma\left\{\sum_{i=1}^{n}\left(\frac{\nabla f}{f}\right) - W_1 - \frac{1}{2}W_2\right\}\sigma^2$	$\Sigma\sigma^2$
DWLS	$\Sigma\left\{-2\left(\frac{\nabla f}{f}\right) + 2\,W_1 - \frac{1}{2}W_2\right\}\sigma^2$	$\Sigma\sigma^2$

Table 6.4: The estimators for the bias and the $Var(\hat{\theta})$: regeneration data

In Table 6.4

$$\begin{aligned}
\Sigma &= \left[\sum_{i=1}^{n_1}\left(\frac{\nabla f_1}{f_1}\right)_i\left(\frac{\nabla f_1}{f_1}\right)_i^T + \sum_{j=1}^{n_2}\left(\frac{\nabla f_2}{f_2}\right)_i\left(\frac{\nabla f_2}{f_2}\right)_j^T\right]^{-1}, \\
p &= \text{Total number of parameters,} \\
n &= n_1 + n_2 = \text{Total number of observations,} \\
W_1 &= \sum_{i=1}^{n_1} w_{11}\left(\frac{\nabla f_1}{f_1}\right)_i + \sum_{j=1}^{n_2} w_{12}\left(\frac{\nabla f_2}{f_2}\right)_j \\
\text{and } W_2 &= \sum_{i=1}^{n_1} w_{21}\left(\frac{\nabla f_1}{f_1}\right)_i + \sum_{j=1}^{n_2} w_{22}\left(\frac{\nabla f_2}{f_2}\right)_j, \\
\text{where } w_{11} &= tr\left[\left(\frac{\nabla f_1}{f_1}\right)_i\left(\frac{\nabla f_1}{f_1}\right)_i^T\Sigma\right], \\
w_{12} &= tr\left[\left(\frac{\nabla f_2}{f_2}\right)_j\left(\frac{\nabla f_2}{f_2}\right)_j^T\Sigma\right], \\
w_{21} &= tr\left[\left(\frac{H_1}{f_1}\right)_i\Sigma\right], \\
\text{and } w_{22} &= tr\left[\left(\frac{H_2}{f_2}\right)_j\Sigma\right].
\end{aligned}$$

6.6.2 Biases from a simulation study

Let θ_0 be the vector of unknown true parameters. The formulae for the biases given in Section 6.6 are based on the expansion for small σ, $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2$. We examined this approximation in a simulation study. Since this study is similar to that we described for the partial bleach method we do not provide the details of the procedure. The parameters were fixed[1] at $\alpha_1 = 57486.4, \alpha_2 = 76.3545, \alpha_3 = 86.9160, \alpha_4 = 95.3052$, and $\gamma = -68.8792$. Table 9.2 of the Appendix 2.1 describes the dose vectors used for the study where they are coded as indicated in the column 'Dose' in Tables 6.5 and 6.6. From the simulation study the biases were computed as the averages of $\hat{\gamma} - \gamma - C_1\sigma$ and $(\hat{\gamma} - \gamma)$ values. Here,

$$C_1 = \Sigma\left[\sum_{i=1}^{n_1}\epsilon_{1i}\left(\frac{\nabla f_1}{f_1}\right)_i + \sum_{j=1}^{n_2}\epsilon_{2j}\left(\frac{\nabla f_2}{f_2}\right)_j\right]. \tag{6.16}$$

The biases of the estimators were computed using the formulae given in Table 6.4 and

[1] In our study, we fixed the parameters at the maximum likelihood estimates obtained for the data sets 'WFP2-7R1' of Huntley *et. al.* [38].

compared with the biases from the simulation study. The results based on 25000 simulations are given in Tables 6.5 and 6.6. In the Tables 6.5 and 6.6, B_T, B_1 and B_2 respectively denote the true bias, the average of $\hat{\gamma} - \gamma - C_1\sigma$ values, and average of $\hat{\gamma} - \gamma$ values.

Dose	σ^2	ML			QL			n_1	n_2
		B_T	B_1	B_2	B_T	B_1	B_2		
(R1)	0.004	-0.060	-0.076	-0.076	-0.053	-0.099	-0.096	30	32
(R1)	0.008	-0.120	-0.114	-0.115	-0.106	-0.106	-0.106	30	32
(R2)	0.006	-0.078	-0.086	-0.086	-0.091	-0.100	-0.100	40	40
(R2)	0.008	-0.103	-0.134	-0.134	-0.122	-0.159	-0.159	40	40
(R4)	0.010	-0.032	-0.018	-0.018	-0.073	-0.076	-0.076	100	100

Table 6.5: Results of the Monte Carlo study 6.1(a)

Study	σ^2	GLS			DWLS			n_1	n_2
		B_T	B_1	B_2	B_T	B_1	B_2		
(R1)	0.004	-0.060	-0.078	-0.078	-0.039	-0.062	-0.062	30	32
(R1)	0.008	-0.120	-0.131	-0.131	-0.079	-0.072	-0.072	30	32
(R2)	0.006	-0.078	-0.086	-0.086	-0.119	-0.136	-0.136	40	40
(R2)	0.008	-0.103	-0.139	-0.139	-0.159	-0.227	-0.227	40	40
(R4)	0.010	-0.032	-0.018	-0.017	-0.073	-0.076	-0.076	100	100

Table 6.6: Results of the Monte Carlo study 6.1(b)

Based on the simulation results given in Tables 6.5 and 6.6, we conclude that the derived formulae are valid for small $\sigma(< 0.09)$. For real data sets collected for the regeneration method, σ is usually around 0.08. Since sample sizes used in the simulation study are around the real sample sizes, the derived formulae can safely be used to estimate the biases of all four estimators for real data sets. For σ in the range of real data sets, we notice that the biases are negligible compared to the standard errors.

6.7 Comparison of the estimators in finite samples

Next we describe the results of a small study that examines the biases of the estimators in the model for the regeneration method. For the study described here the parameters were set at $\alpha_1 = 57486.4, \alpha_2 = 76.3545, \alpha_3 = 86.9160, \alpha_4 = 95.3052, \gamma = -68.8792$ and $\sigma = 0.029$. The dose levels were fixed at (0,40,80,120,300,394,700). The procedure is similar to that which we described for the additive dose method in Section 4.3.

n	α_1				α_2			
	ML	QL	GLS	DWLS	ML	QL	GLS	DWLS
14	1130.95	962.02	1590.83	-295.63	0.33	0.44	0.33	0.67
28	655.47	481.00	1025.36	-607.72	0.16	0.22	0.16	0.33
56	282.74	240.50	742.63	-763.78	0.08	0.11	0.08	0.17
112	141.40	120.28	601.31	-841.73	0.04	0.05	0.04	0.08

Table 6.7: Comparison of the biases of $\hat{\alpha}_1$ and $\hat{\alpha}_2$: regeneration method

n	α_3				α_4			
	ML	QL	GLS	DWLS	ML	QL	GLS	DWLS
14	2.16	1.68	2.16	0.70	-1.99	-1.46	-1.22	-1.94
28	1.08	0.84	1.08	0.35	-0.99	-0.73	-0.23	-1.73
56	0.54	0.42	0.54	0.17	-0.50	-0.37	0.27	-1.63
112	0.27	0.21	0.27	0.09	-0.25	-0.18	0.51	-1.58

Table 6.8: Comparison of the biases of $\hat{\alpha}_3$ and $\hat{\alpha}_4$: regeneration method

Conclusions

Based on the results presented in Tables 5.9 - 5.11, we draw the following conclusions.

1. For fixed σ, as n increases, the biases of maximum likelihood and quasi-likelihood estimators for all five parameters converge to zero at a rate $O(1/n)$.

n	γ			
	ML	QL	GLS	DWLS
14	-0.34	-0.44	-0.34	-0.65
28	-0.17	-0.22	-0.17	-0.32
56	-0.08	-0.11	-0.08	-0.16
112	-0.04	-0.05	-0.04	-0.08

Table 6.9: Comparison of the biases of $\hat{\gamma}$: the regeneration method

2. Except for the parameters α_1 and α_4, the biases of generalized least squares and data weighted least squares estimators also converge to zero at a rate $O(1/n)$, as n is increased while σ is fixed.

3. For all the parameters except α_1 and α_4 the generalized least squares estimator and the maximum likelihood estimator have almost the same bias.

4. It is easy to see from the results derived in Chapter 4 that the least squares estimators for γ, the parameter of our interest are asymptotically unbiased.

5. The results we observed here agree with the theoretical results presented in Chapter 4.

6.8 Worked example

We analyzed the data set 'SESA1' (collected at 350^0C) cited in Berger *et. el* [12] using the techniques described earlier in this chapter. Now we present the results for fitting the response functions

$$f(x,\theta) = \alpha_1 \left\{1 - \exp\left[-\frac{(x+\alpha_2)}{\alpha_3}\right]\right\} + \alpha_4(x+\alpha_2)$$

and

$$f(x,\theta) = \alpha_1 \left\{1 - \exp\left[-\frac{(x+\alpha_2+\gamma)}{\alpha_3}\right]\right\} + \alpha_4(x+\alpha_2+\gamma),$$

for the unbleached and bleached data respectively; here $\theta = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \gamma)^T$. The parameter estimates for this data set are given in Tables 6.10 and 6.11.

	ML			QL		
parameter	estimate	bias	std.error	estimate	bias	std.error
$\alpha_1 \times 10^{-3}$	62.15	0.40	4.89	62.53	0.37	5.16
α_2	79.62	0.08	3.41	79.61	0.08	3.58
α_3	89.61	0.69	9.66	89.92	0.58	10.14
α_4	105.64	-0.75	11.12	105.12	-0.58	11.72
ED	73.06	0.08	3.23	73.17	0.08	3.39
σ	0.0667			0.0697		

Table 6.10: ML and QL estimates for the model parameters: Data SESA1

	GLS			DWLS		
parameter	estimate	bias	std.error	estimate	bias	std.error
$\alpha_1 \times 10^{-3}$	62.42	0.74	5.13	62.60	-0.41	5.34
α_2	79.62	0.08	3.57	79.85	0.06	3.68
α_3	89.61	0.75	10.09	90.70	0.23	10.50
α_4	106.12	-0.30	11.67	103.52	-1.16	12.07
ED	73.06	0.08	3.37	73.56	0.06	3.50
σ	0.0694			0.0719		

Table 6.11: GLS and DWLS estimates for the model parameters: Data SESA1

The 95% confidence intervals for the equivalent dose (ED) are given in Table 6.12.

Description	Lower bound	Upper bound
Profile likelihood	66.77	79.82
using F critical value	66.47	80.16
Z interval based on the ML	66.74	79.39
t interval based on the ML	66.32	79.80
Quasi-score interval	66.60	80.24
t interval based on the QL	66.39	79.96
t interval based on the GLS	66.61	80.56
t interval based on DWLS	66.56	80.56

Table 6.12: Confidence intervals for the equivalent dose: Data SESA1

6.9 Discussion

In this chapter we described procedures for estimating the equivalent dose from regeneration data. The techniques developed for partial bleach data were slightly modified to fit the regeneration data. As for the partial bleach data, finite sample performance of the asymptotic results were examined by Monte Carlo simulations. Coverage probabilities of the following confidence intervals were found to agree with the nominal coverages even in small samples:

1. symmetric confidence intervals with t critical values
2. quasi-score intervals
3. profile likelihood intervals with transformed F critical values

When the sample sizes are small, profile likelihood intervals and symmetric confidence intervals with Z critical values were found to have lower coverage probabilities than their nominal values. Based on the simulation results we recommend symmetric confidence intervals with t critical values. These confidence intervals have the added advantage that they are computationally much simpler than profile likelihood intervals.

All four estimators (maximum likelihood, quasi-likelihood, generalized least squares and data weighted least squares) were found to have negligible biases compared to their standard errors, when the relative error in a single measurement σ is small, and in particular for sample sizes and values of σ typical of careful experimental work.

Chapter 7

Equivalent dose from plateau data

In the previous chapters we focused on estimating the equivalent dose from the data collected at a given temperature. In thermoluminescence studies, data are collected at a series of temperatures. The equivalent dose estimated at each temperature is then plotted against the temperature. Figure 2.3 given in section 2.2 demonstrated such a plot. The plateau region, where the estimated equivalent dose does not vary with the temperature, is believed to represent traps which have not been subjected to leakage over the burial time. Therefore, only those traps corresponding to the plateau region can provide reliable information for dating purposes; see Aitken [1]. From separate analyses at each temperature on the plateau, we have several estimates for the equivalent dose corresponding to the plateau. Here we address the problem of combining these estimates to obtain a more precise estimate for the equivalent dose.

We have to face two problems in using the observations at several temperatures. First we need to correctly identify the temperatures belonging to the plateau. From experience with real data sets we understand that it is not clear cut whether or not certain temperatures belong to the plateau. However we do not intend to deal with this problem in this work. The second problem arises from using the same subsamples to collect data at several temperatures. Consequently, the observations collected at temperatures over the plateau are positively correlated. On the other hand over the range of plateau temperatures the same traps are being emptied. This could produce a negative correlation between the photon

counts taken over adjacent temperatures on the plateau. The real data sets we analyzed show that the observations taken on the plateau are highly positively correlated.

In section 7.1, we introduce our notation. Section 7.2 proposes a procedure closely related to that of Liang and Zeger [43] for estimating the equivalent dose from the data corresponding to the plateau region assuming all the response curves correspond to a common σ (relative error in a single measurement). In section 7.2.2, we investigate large sample properties of the suggested estimate and provide an estimate for the approximate error of the estimate. In section 7.2.3, we investigate small σ asymptotic properties of the derived estimate. In section 7.2.4, we propose symmetric confidence intervals for the equivalent dose with a t quantile and provide a formula for an approximate degrees of freedom of the suggested t quantile. Finite sample performance of the proposed asymptotic results are examined by a simulation study which we describe in Section 7.2.5. From the simulation study we find that the coverage probabilities of symmetric confidence intervals using standard normal quantiles are lower than their nominal levels. The t confidence intervals were found to have coverage probabilities close to their nominal confidence coefficients.

As we clarify in section 7.3, if bleaching could cause a change in the variability of the emitting grains relative to the mean number of emitting grains in a sample, the physical model we proposed in Chapter 2 suggests different σ values for the unbleached and bleached response curves. Section 7.3 proposes a procedure for estimating the equivalent dose assuming unbleached and bleached response curves correspond to different σ.

We developed software using the programming language FORTRAN to implement the suggested methodology. In section 7.4, we demonstrate the suggested methodology using a real data set. Section 7.5 summarizes the chapter.

7.1 Introduction to the problem

When analyzing the data at a given temperature, we described estimating the equivalent dose from a simultaneous fitting of unbleached and bleached response curves. In this chapter, we extend this idea and estimate the equivalent dose from a simultaneous fitting of all the curves corresponding to the plateau region. The temperatures corresponding to the plateau

region are assumed to have been correctly identified by some other technique.

We explain the estimation procedure for partial bleach data. To fit regeneration data, the same estimation procedure can be used, by suitably modifying the vectors of response functions and derivatives of response functions with respect to the components of the parameter vector. In simultaneous curve fitting, the equivalent dose (γ) is treated as a parameter common to all the response curves. We notice that, the expected photon count corresponding to the equivalent dose depends on the temperature, but is common to the unbleached and the bleached response curves at a given temperature. We denote the expected photon count corresponding to the equivalent dose at temperature T by δ_T. From a practical point of view, δ_T represents the photon count had the sample been measured at the time of deposition. Consequently, δ_T plotted against the temperature represents a typical glow curve that would have been observed, if the sample was collected at the time of deposition. (Figure 7.1 demonstrates a plot of estimated δ_T against the temperature for the Data WFP2-7R1 cited earlier.) Therefore, we feel that δ_T is a physically meaningful parameter that could perhaps provide insight into the study of thermoluminescence dating. For this reason, we decided to treat δ_T as another parameter in the new setting.

In the partial bleach method, the unbleached and bleached response curves corresponding to a given temperature are represented by the response functions:

$$f_1(x, \alpha_1, \alpha_2, \alpha_3) = \alpha_1 \left[1 - \exp\left(-\frac{(x+\alpha_2)}{\alpha_3}\right)\right]$$

and

$$f_2(x, \beta_1, \beta_2, \beta_3) = \beta_1 \left[1 - \exp\left(-\frac{(x+\beta_2)}{\beta_3}\right)\right].$$

When $x = \gamma$, $f_1 = f_2 = \delta$. Therefore, $\alpha_1 = \frac{\delta}{\left[1-\exp\left(-\frac{(\gamma+\alpha_2)}{\alpha_3}\right)\right]}$, and $\beta_1 = \frac{\delta}{\left[1-\exp\left(-\frac{(\gamma+\beta_2)}{\beta_3}\right)\right]}$. Thus, the parameters α_1 and β_1 can easily be eliminated using the parameters γ and δ.

Figure 7.2 illustrates the response curves corresponding to two temperatures T_1 and T_2. In the simultaneous curve fitting for partial bleach data, $(\delta, \alpha_2, \alpha_3, \beta_2, \beta_3, \gamma)$ is treated as the set of parameters for the response curves at a given temperature[1]. When we refer to the

[1]We observe that if it is possible to regard the response curves corresponding to the plateau region as the same set of unbleached and bleached response curves shifted along the temperature axis then they can be described with fewer parameters than used in our model. However, since physicists believe that (D.J. Huntley, personal communication) this is not the case we did not investigate this possibility.

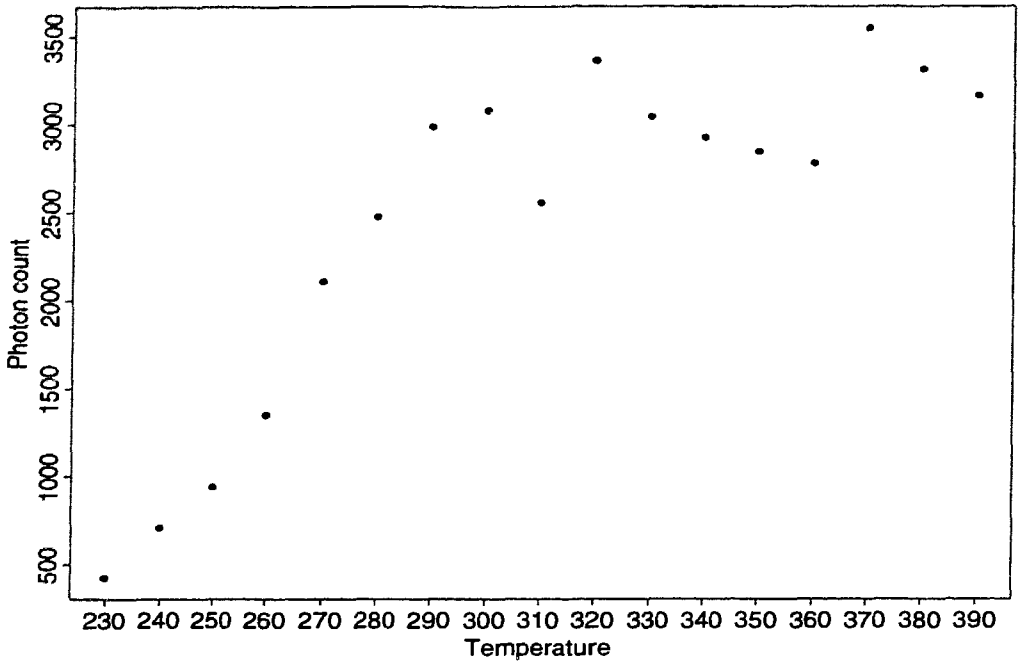

Figure 7.1: Photon count vs temperature at the plateau

response functions at a given temperature we drop the suffix T for notational convenience; however, except for the parameter γ all the other five parameters $(\delta, \alpha_2, \alpha_3, \beta_2, \beta_3)$ depend on the temperature T.

Suppose there are n_0 temperatures on the plateau. Then the parameter vector θ has $(5n_0 + 1)$ components; the equivalent dose common to all the curves and 5 parameters $(\delta, \alpha_2, \alpha_3, \beta_2, \beta_3)$ corresponding to the unbleached and bleached response curves at each temperature. If there are no missing values, there are n_0 measurements from each disc (replicate subsample). Suppose observations are taken on n_1 unbleached and n_2 bleached subsamples. Let $n_3(= n_1 + n_2)$ be the total number of replicate subsamples. Let us denote the photon counts taken on the lth replicate subsample by $y_{l1}, \ldots, y_{ln_0}$. We stack all the observations corresponding to the unbleached data followed by the bleached data to obtain the Y vector with entries $y_{11}, \ldots, y_{1n_0}, \ldots, y_{n_1 1}, \ldots, y_{n_1 n_0}, \ldots, y_{n_3 1}, \ldots, y_{n_3 n_0}$. The vector of dose values is obtained by stacking together the corresponding dose values. We note

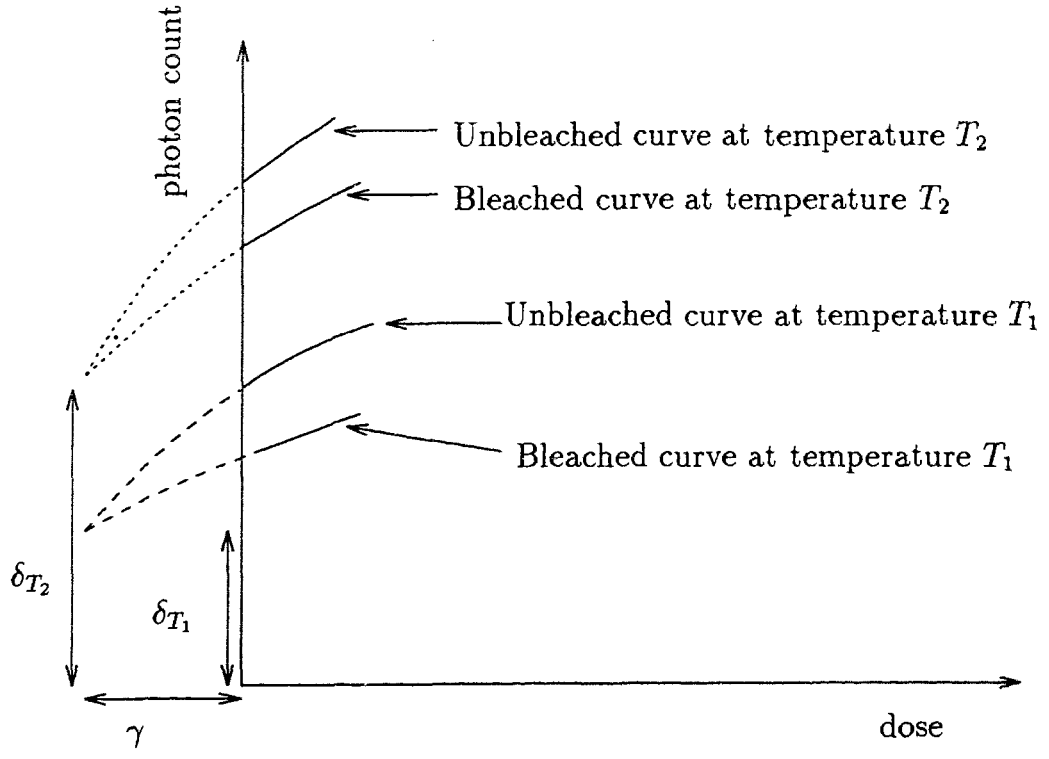

Figure 7.2: Response curves at the plateau

that the variance covariance matrix of Y is block diagonal, since the observations taken on different discs are assumed to be independent. Each block matrix is $n_0 \times n_0$. For notational convenience let $p(= 5n_0 + 1)$ and $n(= n_3 n_0 = (n_1 + n_2)n_0)$ denote the total number of parameters and the total number of observations respectively.

We first discuss estimation assuming that the correlation between any two photon counts taken on the same disc does not depend on the dose received by the sample but may depend on whether or not the sample was subjected to laboratory bleaching. We employ ρ_{ij} (or ρ'_{ij}) to denote the correlations between the observations taken on the same unbleached (or the bleached) disc at temperatures i and j respectively. Suppose Ω_U and Ω_B denote the correlation matrices for vectors of unbleached and bleached photon counts respectively. Under the assumed model, the correlation matrix for Y takes the form

$$\Omega = \begin{bmatrix} \Omega_U & 0 \\ 0 & \Omega_B \end{bmatrix}$$

$$\text{where } \Omega_U = \begin{bmatrix} \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n_0} \\ \vdots & & & \\ \rho_{n_0 1} & & \cdots & \rho_{n_0 n_0} \end{bmatrix} & & \\ & \ddots & \\ & & \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n_0} \\ \vdots & & & \\ \rho_{n_0 1} & & \cdots & \rho_{n_0 n_0} \end{bmatrix} \end{bmatrix}$$

$$\text{and } \Omega_B = \begin{bmatrix} \begin{bmatrix} \rho'_{11} & \rho'_{12} & \cdots & \rho'_{1n_0} \\ \vdots & & & \\ \rho'_{n_0 1} & & \cdots & \rho'_{n_0 n_0} \end{bmatrix} & & \\ & \ddots & \\ & & \begin{bmatrix} \rho'_{11} & \rho'_{12} & \cdots & \rho'_{1n_0} \\ \vdots & & & \\ \rho'_{n_0 1} & & \cdots & \rho'_{n_0 n_0} \end{bmatrix} \end{bmatrix}.$$

Note that $\rho_{ij} = \rho_{ji}$, $\rho'_{ij} = \rho'_{ji}$, and $\rho_{ii} = \rho'_{ii} = 1$, $\forall i, j = 1, \ldots, n_0$. We denote the correlation matrices for the observations taken on the ith unbleached and jth bleached subsample (disc) as Ω_{1i} and Ω_{2j} respectively. The assumption that *correlation between observations taken on the same sample does not depend on the dose received by the sample* leads us to believe that $\Omega_{1i} = \Omega_1$, for $i = 1, \ldots, n_1$ and $\Omega_{2j} = \Omega_2$ for $j = 1, \ldots, n_2$. As we later clarify, this assumption is not crucial for the proposed estimating procedures to be valid.

As in the fitting of data at a single temperature, we assume that standard deviations of photon counts are proportional to their means. We further assume that the constant of proportionality, σ, does not depend on the temperature, nor on whether or not the samples had received laboratory bleaching. Later we show how to extend the results if σ is different for the unbleached and bleached data.

Let $f_1, \ldots, f_n$ denote the mean functions of the observations and D_f denotes the diagonal matrix with $f_1, \ldots, f_n$ along its diagonal. The variance covariance matrix for Y takes the form $V = \sigma^2 D_f \Omega D_f$, where Ω is the correlation matrix for Y. Let θ be the vector of parameters we wish to estimate. Let F be the $n \times p$ matrix of partial derivatives of f with

respect to the components of θ. As described in Chapter 3, quasi-likelihood estimates for θ are found as solutions of the quasi-likelihood estimating equations $F^T W(Y - f) = 0$, where $W = D_f^{-1}(\sigma^2\Omega)^{-1}D_f^{-1}$. Often, the correlation matrix Ω is unknown. Liang and Zeger [43] suggest replacing the unknown correlation matrix Ω by what they call a working correlation matrix R.

First we develop an estimation procedure for our model closely related to that proposed by Liang and Zeger [43] for the linear regression model.

7.2 GEE estimates when unbleached and bleached curves correspond to a common relative error

Let Y_i denote the vector of observations collected from the ith sample with component Y_{ij} denoting the photon count observed from the ith sample at temperature j. In the following, we denote the components of θ corresponding to the jth temperature by θ_j. This means θ_j has entries $\delta_j, \alpha_{2j}, \alpha_{3j}, \beta_{2j}, \beta_{3j}, \gamma$. The entries of θ_j are used to evaluate the response function $E(Y_{ij})$. Let f_i denote the vector of response functions corresponding to the ith disc. Then f_i has entries $E(Y_{ij})$, $j = 1, \ldots, n_0$. The model is:

$$\begin{aligned} E(Y_{ij}) &= f(x_i, \theta_j), \\ Var(Y_{ij}) &= \sigma^2 \left(f(x_i, \theta_j)\right)^2. \end{aligned}$$

Let D_{f_i} denote the $n_0 \times n_0$ diagonal matrix with $f(x_i, \theta_j)$, $j = 1, \ldots, n_0$ along its diagonal. In general the covariance matrix for Y_i can be written as

$$Cov(Y_i) = \sigma^2 D_{f_i} \Omega_i D_{f_i},$$

where Ω_i is the unknown correlation matrix of Y_i.

Motivation for the proposed scheme:

If we assume each Ω_i is known, then the usual quasi-likelihood equations for estimating θ are

$$\sum_{i=1}^{n_3} F_i^T \left(\sigma^2 D_{f_i} \Omega_i D_{f_i}\right)^{-1} (Y_i - f_i) = 0.$$

The above system of equations is equivalent to

$$\sum_{i=1}^{n_2} F_i^T (D_{f_i} \Omega_i D_{f_i})^{-1} (Y_i - f_i) = 0. \tag{7.1}$$

Having motivated the use of Equations 7.1, following Liang and Zeger [43] we plan to estimate θ by replacing the unknown Ω_i's in Equation 7.1 by some correlation matrix R. This suggests estimating $\hat{\theta}$ by solving

$$U(\theta, R) = \sum_{i=1}^{n_3} F_i^T (D_{f_i} R D_{f_i})^{-1} (Y_i - f_i) = 0. \tag{7.2}$$

Typically, unless a specific value for R is assumed (such as the identity matrix in independent equations of Liang and Zeger [43]), Equation 7.2 must be augmented by some procedure for "estimating R". Following Liang and Zeger [43] we refer to R as a "working correlation matrix". In practice, entries of R will often be calculated from residuals to the fit; this requires an iterative process: solving Equation 7.2 to get an estimate for θ, updating R using current estimates for θ, and re-estimating θ etc. In our work, we took R as the sample correlation matrix computed from the Pearson residuals at each iteration, or as a fixed matrix selected on the basis of experience with other data sets.

7.2.1 Solution of generalized estimating equations

Before we propose an algorithm for solving Equations 7.2, we introduce further notation to ease the work of presenting our procedure.

Since observations taken on different discs are assumed to be independent, we can rewrite Equations 7.2 as $U(\theta, R) = F^T (D_f R D_f)^{-1} (Y - f) = 0$, where again, F denotes the $n \times p$ matrix whose (i, j)th element is $\frac{\partial f_i}{\partial \theta_j}$. Here R denotes the working correlation matrix for Y; note the change in notation: R is now $n \times n$ rather than $n_0 \times n_0$. Letting R_U and R_B denote the working correlation matrices for unbleached and bleached data R can be written as

$$R = \begin{bmatrix} R_U & 0 \\ 0 & R_B \end{bmatrix},$$

$$\text{with } R_U = \begin{bmatrix} R_1 & 0 & 0 & \cdots & 0 \\ 0 & R_1 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & \\ 0 & 0 & 0 & \cdots & R_1 \end{bmatrix},$$

$$\text{and } R_B = \begin{bmatrix} R_2 & 0 & 0 & \cdots & 0 \\ 0 & R_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & \\ 0 & 0 & 0 & \cdots & R_2 \end{bmatrix};$$

here R_1 and R_2 denote the $n_0 \times n_0$ working correlation matrices for observations taken on ith unbleached and jth bleached discs. Since we assume that correlation matrices do not depend on the doses received by the samples it is reasonable to take R_1 (or R_2) to be the same for all the unbleached (or bleached) samples.

In this section, we assume that σ is the same for all the response curves. In section 7.3, we suggest estimating equations and present an algorithm for estimating θ when σ is different for unbleached and bleached curves.

Now we present an algorithm for solving Equations 7.2 to obtain an estimate for θ.

1. Find initial estimates for the parameters. We used the parameter estimates obtained for separate curve fit at different temperatures as initial estimates.

2. Compute the fitted values $\hat{f}_1, \ldots, \hat{f}_n$ at the current estimate $\hat{\theta}$.

3. Find a working correlation matrix R. We computed the Pearson residuals $\frac{Y_i - \hat{f}_i}{\hat{f}_i}$ for the unbleached and bleached data, and used the sample correlation matrices computed from Pearson residuals for unbleached data and bleached data as R_1 and R_2 respectively.

4. Let $V^{(k)} = D_{\hat{f}}^{(k)} R^{(k)} D_{\hat{f}}^{(k)}$ be an estimate for the variance covariance matrix for Y evaluated at the current estimate θ_k. Let $W^{(k)}$ denote the inverse of $V^{(k)}$.

   At the $(k+1)$st iteration, θ is estimated from $\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \hat{\zeta}^{(k)}$, where

$$\hat{\zeta}^{(k)} = \left((F^{(k)})^T W^{(k)} F^{(k)} \right)^{-1} (F^{(k)})^T W^{(k)} \left(Y - f^{(k)} \right).$$

The superscript k indicates that each term is evaluated at the parameter estimates from the kth iteration.

5. Iterate until desired convergence. In the software we developed, the convergence criterion was met when the absolute difference in the parameter estimates of successive iterations was less than 10^{-3}.

Following the notation of Liang and Zeger [43], we refer to the estimate obtained as a solution to the set of equations 7.2 as the GEE estimate.

7.2.2 Large sample asymptotics

The proposed estimate $\hat{\theta}$ solves

$$U(\theta, R) = 0.$$

Assume that the working correlation matrix R converges to some fixed correlation matrix $\mho$ (this symbol is "mho"); Crowder [21] shows that there are fairly natural natural examples of working correlation matrices and corresponding true correlation matrices for which this assumption fails. For notational convenience, let α and β be $s \times 1$ $(1 \leq s \leq n_0(n_0-1)/2)$ vectors which fully characterize the elements of R and $\mho$ respectively. Taylor series expansion of $U(\theta, R)$ around $U(\theta, \mho)$ gives

$$U(\theta, R) \approx U(\theta, \mho) + \sum_{i=1}^{s} (\alpha_i - \beta_i) \left[\frac{\partial U(\theta, R)}{\partial \alpha_i}\right]_{R=\mho}.$$

From equation 7.2, we find

$$\left[\frac{\partial U(\theta, R)}{\partial \alpha_i}\right]_{R=\mho} = \sum_{i=1}^{n_3} F_i^T \left(\frac{\partial}{\partial \alpha_i}[D_{f_i} R(\alpha) D_{f_i}]^{-1}\right)_{R=\mho} (Y_i - f_i).$$

Observe that each $\left[\frac{\partial U(\theta,R)}{\partial \alpha_i}\right]_{R=\mho} \partial\alpha_i$ is a sum of n_3 independent mean zero terms (since each is a non-random bounded quantity times $(Y_i - f_i)$ which has mean zero). Hence, under mild regularity conditions $Var\left(\left[\frac{\partial U(\theta,R)}{\partial \alpha_i}\right]_{R=\mho}\right)$ is $O(n_3)$ and $\left[\frac{\partial U(\theta,R)}{\partial \alpha_i}\right]_{R=\mho}$ is $O_p(\sqrt{n_3})$. Since R converges to $\mho$, as $n_3 \longrightarrow \infty$, $(\alpha_i - \beta_i)$ is $o_p(1)$. It follows that, $\sum_{i=1}^{s}(\alpha_i - \beta_i)\left[\frac{\partial U(\theta,R)}{\partial \alpha_i}\right]_{R=\mho} \partial\alpha_i$ is $o_p(\sqrt{n_3})$. Thus, provided R converges as $n_3 \to \infty$ to some deterministic $\mho$,

$$\frac{1}{\sqrt{n_3}} U(\theta, R) = \frac{1}{\sqrt{n_3}} U(\theta, \mho) + o_p(1), \tag{7.3}$$

It is worth noting that we do not have to believe $\mho = corr(Y)$.

Now consider the following:

$$\begin{aligned}\left[U(\hat{\theta},R)-U(\theta,R)\right]-\left[U(\hat{\theta},\mho)-U(\theta,\mho)\right] &\approx \left[\frac{\partial U(\theta,R)}{\partial\alpha_i}\right]_{R=\mho}\partial\theta(\hat{\theta}-\theta)-\frac{\partial U(\theta,\mho)}{\partial\theta}(\hat{\theta}-\theta)\\ &\approx (\hat{\theta}-\theta)\sum_{i=1}^{s}(\alpha_i-\beta_i)\left[\frac{\partial^2 U(\theta,R)}{\partial\theta\partial\alpha_i}\right]_{R=\mho}\end{aligned}$$

Observing that $\sum_{i=1}^{s}(\alpha_i-\beta_i)\left[\frac{\partial^2 U(\theta,R)}{\partial\theta\partial\alpha_i}\right]_{R=\mho}$ is $o_p(n_3)$, $(\hat{\theta}-\theta)$ is $O_p(\frac{1}{\sqrt{n_3}})$, and $(\alpha_i-\beta_i)$ is $o_p(1)$, we find $\left[U(\hat{\theta},R)-U(\theta,R)\right]-\left[U(\hat{\theta},\mho)-U(\theta,\mho)\right]$ is $o_p(\sqrt{n_3})$. Recall that $\hat{\theta}$ solves $U(\theta,R)=0$. Since $U(\theta,R)$ differs from $U(\theta,\mho)$ by a quantity which is $o_p(\sqrt{n_3})$ we conclude that $U(\hat{\theta},\mho)\approx 0$ with an error of approximation which is $o_p(\sqrt{n_3})$.

As usual in estimating equations, under regularity conditions which we have not investigated in detail, $U(\hat{\theta},\mho)=0$ can be expanded about θ to produce

$$U(\hat{\theta},\mho)\approx U(\theta,\mho)+\left[\frac{\partial U(\hat{\theta},\mho)}{\partial\hat{\theta}}\right]_{\hat{\theta}=\theta}(\hat{\theta}-\theta).$$

Since $U(\hat{\theta},\mho)$ is approximately zero (error of approximation is $o_p(\sqrt{n_3})$) we find

$$\left(\hat{\theta}-\theta\right)\approx-\left(\frac{\partial U(\theta,\mho)}{\partial\theta}\right)^{-1}U(\theta,\mho).$$

Next from Equation 7.2, we observe that

$$\left(\frac{\partial U(\theta,\mho)}{\partial\theta}\right)=-\sum_{i=1}^{n_3}F_i^T\left(D_f\mho D_f\right)^{-1}F_i.$$

Hence,

$$\sqrt{n_3}(\hat{\theta}-\theta)=\left(\frac{1}{n_3}\sum_{i=1}^{n_3}F_i^T\left(D_f\mho D_f\right)^{-1}F_i\right)^{-1}\left(\frac{1}{\sqrt{n_3}}\sum_{i=1}^{n_3}U_i(\theta,\mho)\right),$$

where $U_i=F_i^T\left(D_{f_i}\mho D_{f_i}\right)^{-1}(Y_i-f_i)$. For notational convenience we write $W_\mho=\left(D_{f_i}\mho D_{f_i}\right)^{-1}$; notice that $W_\mho$ depends on i. Thus,

$$\sqrt{n_3}(\hat{\theta}-\theta)=\left(\frac{1}{n_3}\sum_{i=1}^{n_3}F_i^TW_\mho F_i\right)^{-1}\left(\frac{1}{\sqrt{n_3}}\sum_{i=1}^{n_3}F_i^TW_\mho(Y_i-f_i)\right). \tag{7.4}$$

For large n_3, the right hand side of Equation 7.4 is approximately normal with mean zero and variance of $\hat{\theta}$ given by

$$Var(\hat{\theta})=\left(\frac{1}{n_3}\sum_{i=1}^{n_3}F_i^TW_\mho F_i\right)^{-1}\left(\frac{1}{n_3^2}\sum_{i=1}^{n_3}F_i^TW_\mho Var(Y_i)W_\mho F_i\right)\left(\frac{1}{n_3}\sum_{i=1}^{n_3}F_i^TW_\mho F_i\right)^{-1}. \tag{7.5}$$

To find an estimate for the approximate variance of $\hat{\theta}$ we have three choices:

1. If we are willing to assume that $\mho$ is close to the true correlation matrix for Y_i, then we can replace $Var(Y_i)$ by $\sigma^2 D_i \mho D_i$ to get $Var(\hat{\theta}) \approx \sigma^2 \left(\sum_{i=1}^{n_3} F_i^T W_\mho F_i\right)^{-1}$. We can estimate the approximate variance of $\hat{\theta}$ by replacing the unknown parameters σ^2 and θ by their estimates. In section 7.2.3, we provide an estimate for σ^2 valid in the limit of small measurement error. Hereafter we refer to this estimate as variance estimate 1. An estimate for σ^2 can also be computed as suggested in Liang and Zeger [43].

2. Assuming all the Y_i's have the same correlation matrix Ω, we can write

$$Var(Y_i) = \sigma^2 D_{f_i} \Omega D_{f_i} \quad = V_\Omega(\text{say}).$$

   Then

$$Var(\hat{\theta}) = \left(\frac{1}{n_3}\sum_{i=1}^{n_3} F_i^T W_\mho F_i\right)^{-1} \left(\frac{1}{n_3^2}\sum_{i=1}^{n_3} F_i^T W_\mho V_\Omega W_\mho F_i\right) \left(\frac{1}{n_3}\sum_{i=1}^{n_3} F_i^T W_\mho F_i\right)^{-1},$$

   and can be estimated by replacing the unknown σ^2 and Ω by their estimates; the correlation matrix Ω can be estimated as the sample correlation matrix calculated from the Pearson residuals to the fit. This estimate does not require $\mho$ to be close to the true correlation matrix for Y_i. Hereafter we refer to this estimate as variance estimate 2.

3. Without assuming that $\mho$ is close to the true correlation matrix for Y_i, and following suggestions of Liang and Zeger [43], we can replace $Var(Y_i)$ in Equation 7.5 by $(Y_i - \hat{f}_i)(Y_i - \hat{f}_i)^T$. As in the previous case, an estimate for the approximate variance can be obtained by replacing the unknown parameter θ by $\hat{\theta}$. Hereafter we refer to this estimate as variance estimate 3.

Confidence intervals with standard normal quantiles

Let $\hat{\theta}$ be a solution of $F^T W(y - f) = 0$ and $Var(\hat{\theta})$ be an approximate variance covariance matrix for $\hat{\theta}$ defined by equation 7.5. Let $\hat{V}(\hat{\theta})$ be an estimate for $V(\hat{\theta})$ obtained from any of the three methods discussed in the previous section. Let $\hat{\gamma}$ be the pth component of $\hat{\theta}$

and $s_{\hat{\gamma}}$ be the square root of the pth diagonal element of $\hat{V}(\hat{\theta})$. Using the results of section 7.2.2, an approximate confidence interval for the equivalent dose can be computed by taking $\hat{\gamma} \mp s_{\hat{\gamma}} z_{\alpha/2}$ as lower and upper confidence limits.

From simulation studies (see Section 7.2.5) we learned that confidence intervals for the equivalent dose using normal quantiles have lower coverage probabilities than corresponding nominal levels. In this regard we notice the following. The response function $f(x, \theta)$ in our problem is non-linear in θ. Moreover, in our model the equivalent dose γ and the relative error in a single measurement σ are the only parameters that are common to all the temperatures on the plateau. The rest of the parameters $(\delta, \alpha_1, \alpha_2, \beta_1, \beta_2)$ are different for different temperatures. Consequently, as the number of plateau temperatures increases the number of parameters also increases. We wish to make inference about the parameter γ and rest of the parameters are considered as nuisance parameters. The appearance of a large number of nuisance parameters in the model makes the effective sample sizes much smaller than the nominal sample sizes. As we mentioned in earlier chapters, for our problem the relative error in a single measurement σ is small. This motivated us to find a better approximation for the distribution of the test statistic in small samples valid in the limit of small measurement errors.

7.2.3 Small σ asymptotics

In this section, we examine small σ asymptotic properties of the estimate suggested in section 7.2.1. We do not necessarily assume that sample sizes are large. However, we need to assume that the working correlation matrix R is chosen in such a way that R when evaluated at $(\hat{\theta} = \theta_0, \sigma = 0)$ is some fixed correlation matrix $\mho$. It is possible to choose such R (but the estimate based on the Pearson residual does not have the desired property). For example one may choose any fixed correlation matrix such as the identity matrix in independent estimating equations proposed by Liang and Zeger [43] as R. Recall that $\mho$ need not be the true correlation matrix Ω; in fact, often Ω is unknown. Liang and Zeger [43] suggests that, the estimates $\hat{\theta}$ are consistent given only that the regression model for $E(Y)$ is correctly specified. It is not required to specify a correct working correlation matrix; this

makes these estimating equations more attractive since often the true correlation matrix is unknown. However, having the working correlation matrix converging to a fixed correlation matrix which is close to the true correlation matrix makes the resulting estimates more efficient. Therefore, one might perhaps choose a working correlation matrix based on prior knowledge about the data or according to some physical model that suggests the correlation structure. In section 7.2.5, we investigate the performance of the suggested procedure for estimating θ by examining the estimates obtained using a few chosen working correlation matrices.

Next, we examine the small σ asymptotic properties of $\hat{\theta}$. As in Chapter 3, we begin by approximating $\hat{\theta}$ using $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2$, for small σ. Then, $C_1 = \frac{\partial\hat{\theta}}{\partial\sigma}\Big|_{\hat{\theta}=\theta_0,\sigma=0}$ and $C_2 = \frac{\partial^2\hat{\theta}}{\partial\sigma^2}\Big|_{\hat{\theta}=\theta_0,\sigma=0}$. The estimates $\hat{\theta}$ solve the system of equations

$$S(\hat{\theta}, \theta_0, \sigma, \hat{\rho}, \hat{\rho}') = (F(\hat{\theta}))^T W_R \left[y - f(\hat{\theta})\right] = 0,$$

where we write $W_R = \left(D_{\hat{f}} R(\hat{\theta}, \sigma, \hat{\rho}, \hat{\rho}') D_{\hat{f}}\right)^{-1}$ for notational convenience. Let $f_1, \ldots, f_n$ denote the response functions evaluated at the true parameter θ_0. Let D_0 denote the diagonal matrix with $f_1, \ldots, f_n$ along its diagonal. Using the introduced notation, the model equations can be written as $Y = f + \sigma D_0\epsilon$. Thus,

$$\frac{\partial S_i}{\partial\sigma} = \frac{\partial}{\partial\sigma}\left\{\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\partial\hat{f}_j}{\partial\hat{\theta}_i}(W_R)_{jk}\left[f_k + \sigma(D_0\epsilon)_k - \hat{f}_k\right]\right\},$$

where $\hat{f}$ denotes $f(\hat{\theta})$, which is $f(\theta)$ evaluated at $\theta = \hat{\theta}$. When $(\hat{\theta} = \theta_0, \sigma = 0)$, the term in the square brackets is zero. Therefore,

$$\begin{aligned}
\frac{\partial S_i}{\partial\sigma}\Big|_{\hat{\theta}=\theta_0,\sigma=0} &= \left[\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\partial f_j}{\partial\hat{\theta}_i}(W_R)_{jk}(D_0\epsilon)_k - \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{p}\frac{\partial f_j}{\partial\hat{\theta}_i}(W_R)_{jk}\frac{\partial f_k}{\partial\hat{\theta}_l}\frac{\partial\hat{\theta}_l}{\partial\sigma}\right]_{\hat{\theta}=\theta_0,\sigma=0} \\
&= \left[\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\partial f_j}{\partial\hat{\theta}_i}(W_R)_{jk}(D_0\epsilon)_k - \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{p}\frac{\partial f_j}{\partial\hat{\theta}_i}(W_R)_{jk}\frac{\partial f_k}{\partial\hat{\theta}_l}(C_1)_l\right]_{\hat{\theta}=\theta_0,\sigma=0}.
\end{aligned}$$

Since $[R]_{\hat{\theta}=\theta_0,\sigma=0} = \mho$, we find $[W_R]_{\hat{\theta}=\theta_0,\sigma=0} = (D_f \mho D_f)^{-1}$ which we denote by $W_\mho$. Thus,

$$\left(\frac{\partial S}{\partial\sigma}\Big|_{\hat{\theta}=\theta_0,\sigma=0}\right)_i = \left(F^T W_\mho D_0\epsilon\right)_i - \left(F^T W_\mho F C_1\right)_i.$$

Therefore,

$$\left(\frac{\partial S}{\partial \sigma}\right)\Big|_{\hat{\theta}=\theta_0,\sigma=0} = \left(F^T W_{\mho} D_0 \epsilon\right) - \left(F^T W_{\mho} F C_1\right).$$

Since $\frac{\partial S}{\partial \sigma} = 0$, we find

$$\begin{aligned} C_1 &= \left(F^T W_{\mho} F\right)^{-1} F^T W_{\mho} D_0 \epsilon \\ &= \frac{1}{\sigma}\left(F^T W_{\mho} F\right)^{-1} F^T W_{\mho}(y - f). \end{aligned}$$

Thus, $E(C_1) = 0$, and $Var(C_1) = \frac{1}{\sigma^2}\left(F^T W_{\mho} F\right)^{-1} F^T W_{\mho} Var(Y) W_{\mho} F \left(F^T W_{\mho} F\right)^{-1}$. Since for small σ, $\hat{\theta} \approx \theta_0 + C_1\sigma + C_2\sigma^2$, we find

$$Var(\hat{\theta}) \approx \left(F^T W_{\mho} F\right)^{-1} F^T W_{\mho} Var(Y) W_{\mho} F \left(F^T W_{\mho} F\right)^{-1}. \tag{7.6}$$

We notice that the variance covariance matrix derived in the limit of small measurement error (Equation 7.6) is identical to that derived for the large sample asymptotic case (Equation 7.5). Also, if the observations are independent, by substituting $\mho = \Omega = I$ we find that this variance estimate reduces to the one derived in Chapter 5.

Remarks: Our working correlation matrix suggested in section 7.2.1 uses

$$\begin{aligned} R_1 &= \frac{1}{n_1}\sum_{l=1}^{n_1}\left[\frac{(Y_l - \hat{f}_l)}{\hat{f}_l}\right]\left[\frac{(Y_l - \hat{f}_l)}{\hat{f}_l}\right]^T. \\ \text{Therefore, } R_1|_{\hat{\theta}=\theta_0,\sigma=0} &= \frac{1}{n_1}\sum_{l=1}^{n_1}\left[\frac{(Y_l - f_l)}{f_l}\right]\left[\frac{(Y_l - f_l)}{f_l}\right]^T. \\ &= \sigma^2\frac{1}{n_1}\sum_{l=1}^{n_1}\epsilon_{1l}\epsilon_{1l}^T \\ &\approx \sigma^2\Omega_1, \text{ only if } n_1 \text{ is large.} \\ \text{Similarly, } R_2|_{\hat{\theta}=\theta_0,\sigma=0} &\approx \sigma^2\Omega_2, \text{ only if } n_2 \text{ is large.} \end{aligned}$$

When n_1 and n_2 are not large enough, $R|_{\hat{\theta}=\theta_0,\sigma=0}$ is quadratic in ϵ and is not a fixed matrix $\mho$ as required by the theory discussed under small σ asymptotics. Since $W_{\mho} = D_0^{-1} R|^{-1}_{\hat{\theta}=\theta_0,\sigma=0} D_0^{-1}$, we find that C_1 is a complicated form in ϵ. Therefore, the results we propose under small σ asymptotics are valid only when n_1 and n_2 are large enough so that $R|_{\hat{\theta}=\theta_0,\sigma=0} \approx \sigma^2\Omega$ or if we use a fixed working correlation matrix $\mho$ at each iteration. Also we note that if we use R in place of $\mho$ when n_1 and n_2 are not large enough for $R|_{\hat{\theta}=\theta_0,\sigma=0}$ to be close to $\sigma^2\Omega$, formula 7.6 underestimates the true variance of $\hat{\theta}$.

Distribution of the test statistic

Let $\hat{\gamma}$ be an estimate for the equivalent dose obtained as the pth component of the GEE estimate $\hat{\theta}$ suggested in section 7.2.2. Let $\hat{V}(\hat{\theta})$ be an estimate for the approximate variance covariance matrix (Equation 7.6) obtained as discussed in steps 1 or 2 of section 7.2.2. Let $s_{\hat{\gamma}}$ be an estimate for the approximate error of $\hat{\gamma}$ obtained as the square root of the pth diagonal element of $\hat{V}(\hat{\theta})$. Further assume that σ^2 is estimated by the estimate $\hat{\sigma}^2$ described below. In this section, we study the distribution of $(\hat{\gamma} - \gamma)/s_{\hat{\gamma}}$.

Estimate for σ:

Let $\hat{f} = (f_1(x_1, \hat{\theta}), \ldots, f_n(x_n, \hat{\theta})^T$ be the vector of fitted values and $D_{\hat{f}}$ be the diagonal matrix with entries of $\hat{f}$ along the diagonal. Let $f_0 = (f_{0,1}(x_1, \theta_0), \ldots, f_{0,n}(x_n, \theta_0)^T$ be the vector of response functions evaluated at the true parameter θ_0 and D_0 be the diagonal matrix with entries of f_0 along the diagonal. Let $\hat{\epsilon} = D_{\hat{f}}^{-1}\left[Y - \hat{f}\right]$ be the vector of fitted standardized residuals. Taylor expansion of $\hat{f}_i$ around θ_0 gives

$$\hat{f}_i = f(x_i, \hat{\theta}) \approx f(x_i, \theta_0) + \left[F(\hat{\theta} - \theta_0)\right]_i, \quad \text{for } i = 1, \ldots, n_3.$$

Thus,

$$\frac{1}{\hat{f}_i} = \frac{1}{f_{0,i}}\left\{1 - \left[\frac{F}{f_0}(\hat{\theta} - \theta_0)\right]_i\right\},$$

and

$$D_{\hat{f}}^{-1} \approx D_0^{-1}\left\{I - D_0^{-1}F(F^T W_{\mathcal{U}} F)^{-1} F^T W_{\mathcal{U}} D_0 \epsilon\sigma\right\} = D_0^{-1} + o(\sigma).$$

Now notice that

$$\begin{aligned}
\hat{\epsilon} &= D_{\hat{f}}^{-1}\left[Y - \hat{f}\right] \\
&= \left(D_0^{-1} + o(\sigma)\right)\left[f_0 + \sigma D_0 \epsilon - f_0 - F(\hat{\theta} - \theta_0)\right] \\
&= \left(D_0^{-1} + o(\sigma)\right)\left\{\sigma D_0 \epsilon - F(F^T W_{\mathcal{U}} F)^{-1} F^T W_{\mathcal{U}} D_0 \epsilon\sigma\right\} \\
&= \sigma\epsilon - D_0^{-1}F(F^T W_{\mathcal{U}} F)^{-1} F^T W_{\mathcal{U}} D_0 \sigma\epsilon + o(\sigma) \\
&= \sigma(I - H)\epsilon + o(\sigma), \quad \text{where } H = D_0^{-1}F(F^T W_{\mathcal{U}} F)^{-1} F^T W_{\mathcal{U}} D_0.
\end{aligned}$$

Observe that

$$\begin{aligned} HH &= D_0^{-1}F(F^TW_{\mathcal{U}}F)^{-1}F^TW_{\mathcal{U}}D_0D_0^{-1}F(F^TW_{\mathcal{U}}F)^{-1}F^TW_{\mathcal{U}}D_0 \\ &= H \end{aligned}$$

Thus, the matrix H is idempotent. The error sum of squares can be written, to order σ^2, as

$$\hat{\epsilon}^T\hat{\epsilon} = \sigma^2\epsilon^T(I-H^T)(I-H)\epsilon.$$

Let $Q = (I-H^T)(I-H)$. Since $\epsilon \sim MVN(0,\Omega)$, we find

$$\begin{aligned} E\left(\hat{\epsilon}^T\hat{\epsilon}\right) &= \sigma^2 E\left[tr\left(\epsilon^TQ\epsilon\right)\right] \\ &= \sigma^2 E\left[tr\left(Q\epsilon\epsilon^T\right)\right] = \sigma^2 tr\left(QE\left[\epsilon\epsilon^T\right]\right) \\ &= \sigma^2 tr\left(Q\Omega\right). \end{aligned}$$

An unbiased estimate for σ^2 is therefore given by $\hat{\sigma}^2 = \frac{\hat{\epsilon}^T\hat{\epsilon}}{tr(Q\Omega)}$, where $Q = (I-H^T)(I-H)$. To estimate σ using the above formula we need Ω, which is often unknown. In applications, we replace Ω by the sample correlation matrix obtained from the Pearson residuals to the fit. In section 7.2.5, we examine the finite sample performance of the suggested results by a simulation study.

Next we show that the test statistic $t = (\hat{\gamma}-\gamma)/s_{\hat{\gamma}}$ is approximately distributed as a t variate and propose a formula to compute an approximate degrees of freedom using Satterthwaite's [53] approximation.

Our expansions described earlier lead to the following.

Result 1 *Let $\hat{\theta}$ denote the GEE estimate described in section 7.1 and $\hat{\epsilon}$ denote the vector of Pearson residuals. Let Ω denote the correlation matrix of the random errors ϵ's. Then, as $\sigma \longrightarrow 0$,*

$$\frac{\hat{\epsilon}^T\hat{\epsilon}}{\sigma^2} \longrightarrow \epsilon^TQ\epsilon \stackrel{\mathcal{D}}{=} \sum_i \lambda_i z_i^2,$$

where z_i are iid $N(0,1)$ random variables and λ_i are the eigenvalues of the matrix $Q\Omega$.

It is easy to see that the eigenvalues of $Q\Omega$ are the same as those of $\Omega^{1/2}Q\Omega^{1/2}$. Since $\Omega^{1/2}Q\Omega^{1/2}$ is symmetric, it follows that eigenvalues of $Q\Omega$ are real.

Theorem 5 *In the limit of small measurement error, (i.e. if σ is small), the error sum of squares $\hat{\epsilon}^T\hat{\epsilon}$, and the estimate $\hat{\theta}$ defined in section 7.1 are independent, in the sense that $\epsilon^T Q\epsilon$ and C_1 are independent.*

Proof: In the small σ asymptotic case, we showed that (section 7.2.3)

$$\begin{aligned}\hat{\theta}-\theta_0 &\approx C_1\sigma = \left(F^T W_{\mho} F\right)^{-1} F^T W_{\mho}\left(\frac{Y-f_0}{\sigma}\right)\sigma \\ &= \left(F^T W_{\mho} F\right)^{-1} F^T W_{\mho} D_0\epsilon.\end{aligned}$$

Since $\epsilon \sim MVN(0,\Omega)$, $\hat{\theta}-\theta_0 \stackrel{\mathcal{D}}{=} \left(F^T W_{\mho} F\right)^{-1} F^T W_{\mho} D_0 A\epsilon_1$, where $AA^T = \Omega$ and $\epsilon_1 \sim MVN(0,I)$. On the other hand the error sum of squares is asymptotic, as $\sigma \longrightarrow 0$, to $\sigma^2\epsilon^T(I-H^T)(I-H)\epsilon = \sigma^2\epsilon_1^T A^T(I-H^T)(I-H)A\epsilon_1$, where $H = D_0^{-1}F(F^T W_{\mho}F)^{-1}F^T W_{\mho}D_0$. If $\mho$ is close to Ω, then $W_{\mho} = D_0^{-1}\Omega^{-1}D_0^{-1}$. Therefore, we find

$$\begin{aligned}BA^T(I-H^T)(I-H)A &= \left(F^T W_{\mho}F\right)^{-1}F^T W_{\mho}D_0AA^T(I-H^T)(I-H)A \\ &= \left(F^T W_{\mho}F\right)^{-1}F^T D_0^{-1}\Omega^{-1}D_0^{-1}D_0\Omega(I-H^T-H+H^TH)A \\ &= \left(F^T W_{\mho}F\right)^{-1}F^T D_0^{-1}(I-H^T-H+H^TH)A.\end{aligned}$$

Now notice that

$$\begin{aligned}\left(F^T W_{\mho}F\right)^{-1}F^T D_0^{-1}H^T A &= \left(F^T W_{\mho}F\right)^{-1}F^T D_0^{-1}D_0 W_{\mho}F\left(F^T W_{\mho}F\right)^{-1}F^T D_0^{-1}A \\ &= \left(F^T W_{\mho}F\right)^{-1}F^T W_{\mho}F\left(F^T W_{\mho}F\right)^{-1}F^T D_0^{-1}A \\ &= \left(F^T W_{\mho}F\right)^{-1}F^T D_0^{-1}A\end{aligned}$$

and

$$\begin{aligned}\left(F^T W_{\mho}F\right)^{-1}F^T D_0^{-1}H^T HA &= \left(F^T W_{\mho}F\right)^{-1}F^T D_0^{-1}D_0^{-1}F\left(F^T W_{\mho}F\right)^{-1}F^T W_{\mho}D_0A \\ &= \left(F^T W_{\mho}F\right)^{-1}F^T D_0^{-1}HA\end{aligned}$$

Thus, $BA^T(I-H^T)(I-H)A = 0$. Therefore, using Theorem 4.17 of Graybill (1961) we find that $\epsilon^T Q\epsilon$ and C_1 are independent.

Theorem 6 *Let γ be the equivalent dose and $\hat{\gamma}$ be the estimate for γ suggested in section 7.1. Let $s_{\hat{\gamma}}$ be an estimate for the error of the estimate obtained as defined in variance estimate 1 of section 7.2.3. If the response curves belonging to the plateau region all correspond to a common σ, then as $\sigma \to 0$, the distribution of the statistic $t = (\hat{\gamma} - \gamma)/s_{\hat{\gamma}}$ can be approximated by a t distribution with degrees of freedom $df = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i^2)}$; here λ_i's are the eigenvalues of the matrix $Q\Omega$ (or the matrix $\Omega^{1/2}Q\Omega^{1/2}$).*

Proof:

Using small σ asymptotics we showed that, $\hat{\theta} \approx \theta_0 + C_1\sigma$, where $C_1 = \frac{\partial \hat{\theta}}{\partial \sigma}\Big|_{\hat{\theta}=\theta_0,\sigma=0} = \left(F^T W_\upsilon F\right)^{-1} F^T W_\upsilon D_0 \epsilon$. Thus, $\hat{\theta} - \theta_0$ is a linear combination of ϵ's. Notice that, $\hat{\gamma} - \gamma = a^T(\hat{\theta} - \theta)$, where a is a p vector with first $(p-1)$ elements equal to zero and the pth element unity. Thus, $(\hat{\gamma} - \gamma)$ is a linear combination of the random errors ϵ's. Since ϵ's follow a multivariate normal distribution with mean zero and covariance matrix Ω, Central limit theorem implies that, $(\hat{\gamma} - \gamma)$ is approximately normally distributed. The estimate $s_{\hat{\gamma}}$ is the square root of the pth diagonal element of the matrix $\sigma^2 \left(F^T W_\upsilon F\right)\Big|_{\hat{\theta}}^{-1}$. Suppose σ^2 is estimated by $\frac{\hat{\epsilon}^T\hat{\epsilon}}{tr(Q\Omega)}$. If $\hat{\theta}$ is close to θ, $(F^T W_\upsilon F)^{-1}\Big|_{\hat{\theta}} \approx (F^T W_\upsilon F)^{-1}$. Therefore, $\frac{\hat{\gamma}-\gamma}{s_{\hat{\gamma}}} = \frac{U}{V}$, where $U = \frac{\hat{\gamma}-\gamma}{\sqrt{\sigma^2(F^T W_\upsilon F)^{-1}}}$ and $V = \frac{\hat{\epsilon}^T\hat{\epsilon}}{\sigma^2 tr(Q\Omega)}$. Then, U is approximately normally distributed with zero mean and unit variance. From Theorem 5, U is independent of V.

From Result 1, it follows that the error sum of squares is a linear combination of independent χ_1^2 random variables. Therefore, V is a complex estimate for the variance as defined by Satterthwaite [53]. Applying Satterthwaite's approximation, we arrive at the formula $df = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i^2)}$, where λ_i's are the eigenvalues of the matrix $Q\Omega$ (or the matrix $\Omega^{1/2}Q\Omega^{1/2}$) for the approximate degrees of freedom. To obtain an estimate for the degrees of freedom we suggest replacing Ω by the sample correlation matrix computed from the Pearson residuals to the fit. In section 7.2.5, we present the results of a simulation study that examines the finite sample performance of the suggested asymptotic results.

7.2.4 Confidence intervals

Let $\hat{\theta}$ be the GEE estimate defined as a solution of $F^T W(y - f) = 0$ and $Var(\hat{\theta})$ be the approximate asymptotic variance covariance matrix given by equation 7.5. Let $\hat{V}(\hat{\theta})$ be an

estimate for the variance covariance matrix for $\hat{\theta}$ obtained as described in step 1 of section 7.2.2. Let $\hat{\gamma}$ be the pth component of $\hat{\theta}$ and $s_{\hat{\gamma}}$ be the square root of the pth diagonal element of $\hat{V}(\hat{\theta})$. Using the results of section 7.2.2, an approximate confidence interval for the equivalent dose can be computed by taking $\hat{\gamma} \mp s_{\hat{\gamma}} t_{\alpha/2,df}$ as lower and upper confidence limits. An approximate degrees of freedom for the corresponding t distribution (df) can be computed as described in Theorem 6.

7.2.5 Simulation results

Now we describe the results of a simulation study that examines the finite sample performance of the suggested asymptotic results. The photon counts for the study were generated according to the model described in section 7.1. The FORTRAN subroutine 'RNMVN' was used to generate the multivariate normal random errors ϵ's. The doses given for the unbleached and bleached discs are presented in Table 9.1 of Appendix 9.3 where they are coded as $P6U$ and $P6B$ respectively. The data were generated assuming correlation does not depend on the dose received by the sample. However, for unbleached and bleached data we used different correlation matrices. Correlation matrices used for the study are presented in Tables 9.3 and 9.4 of Appendix 9.3. These are the sample correlation matrices computed from the Pearson residuals to the fit for the data set (for temperatures in the range $(270 - 320^0C)$ described in Section 7.4. For each generated sample, we computed a confidence interval for the equivalent dose as described in section 7.2.4 by:

1. pretending correlation matrices are known and replacing them by the correlation matrices used to generate the data and using variance estimate 1, defined in section 7.2.2.

2. replacing the unknown correlation matrices by the sample correlation matrices computed from the Pearson residuals and using variance estimate 1.

3. replacing the unknown correlation matrix by an arbitrarily chosen fixed correlation matrix and using variance estimate 2 defined in section 7.2.2; we chose the same fixed correlation matrix for both unbleached and bleached data. In the chosen fixed

correlation matrices all off diagonal elements were set at the correlation coefficients, ρ, given in Table 7.2.

Results using known correlation matrices and Pearson residuals are presented in Table 7.1. Results using fixed correlation matrices are given in Table 7.2. Results given in these tables are based on 10000 simulations. The fraction of confidence intervals that capture the true equivalent dose used to generate the data are recorded as the actual coverage. For both cases, the actual coverages were examined using z quantiles and t quantiles with degrees of freedom estimated as described in section 7.2.3.

Dose	σ	n_0	n_1	n_2	Nominal level	Observed coverage corr. matrix known		Observed coverage corr. matrix unknown	
						t	z	t	z
P1	0.01	2	21	19	0.95	0.9527	0.9435	0.9517	0.9426
P1	0.02	2	21	19	0.95	0.9511	0.9430	0.9508	0.9429
P1	0.03	2	21	19	0.95	0.9500	0.9425	0.9488	0.9406
P1	0.04	2	21	19	0.95	0.9493	0.9419	0.9502	0.9422
P2	0.01	4	21	19	0.95	0.9541	0.9461	0.9455	0.9339
P2	0.02	4	21	19	0.95	0.9539	0.9454	0.9434	0.9325
P2	0.03	4	21	19	0.95	0.9497	0.9440	0.9423	0.9309
P2	0.04	4	21	19	0.95	0.9510	0.9429	0.9407	0.9303
P3	0.01	6	21	19	0.95	0.9518	0.9440	0.9294	0.9148
P3	0.02	6	21	19	0.95	0.9486	0.9407	0.9231	0.9073
P3	0.03	6	21	19	0.95	0.9518	0.9424	0.9257	0.9106
P3	0.04	6	21	19	0.95	0.9515	0.9432	0.9319	0.9178

Table 7.1: Coverage probabilities using correlations calculated from Pearson residuals

We also examined the mean squared errors of the estimators from the simulation study (Tables 7.1 and 7.2) by computing the averages of the $(\hat{\gamma} - \gamma)^2$ values. These are presented in Table 7.3.

Dose	σ	n_0	n_1	n_2	$\rho = 0.99$		$\rho = 0.60$		$\rho = 0.00$	
					t	z	t	z	t	z
P1	0.01	2	21	19	0.9539	0.9455	0.9524	0.9441	0.9528	0.9445
P1	0.02	2	21	19	0.9518	0.9439	0.9517	0.9436	0.9518	0.9439
P1	0.03	2	21	19	0.9519	0.9441	0.9500	0.9417	0.9499	0.9416
P1	0.04	2	21	19	0.9516	0.9442	0.9507	0.9415	0.9504	0.9418
P2	0.01	4	21	19	0.9588	0.9422	0.9459	0.9387	0.9461	0.9382
P2	0.02	4	21	19	0.9537	0.9383	0.9507	0.9437	0.9505	0.9437
P2	0.03	4	21	19	0.9565	0.9397	0.9467	0.9379	0.9459	0.9392
P2	0.04	4	21	19	0.9534	0.9379	0.9501	0.9423	0.9482	0.9425
P3	0.01	6	21	19	0.9394	0.9285	0.9516	0.9452	0.9471	0.9397
P3	0.02	6	21	19	0.9380	0.9254	0.9484	0.9418	0.9488	0.9416
P3	0.03	6	21	19	0.9425	0.9313	0.9512	0.9452	0.9507	0.9448
P3	0.04	6	21	19	0.9414	0.9301	0.9494	0.9436	0.9480	0.9419

Table 7.2: Coverage probabilities using fixed correlation matrices: nominal level =0.95

Conclusions

Based on the simulation results we conclude the following.

1. When the correlation matrix is known, the coverage probabilities of confidence intervals using a t quantile with an approximate degrees of freedom (Section 7.2.3) agree with their nominal levels, even in small samples.

2. Coverage probabilities of confidence intervals with a standard normal quantile are lower than their nominal levels, even if the correlation matrix is assumed to be known.

3. When the correlation matrix is replaced by the sample correlation matrix computed from Pearson residuals, the suggested theory appear to hold with few temperatures on the plateau. However, when we have more than four plateau temperatures the coverages probabilities tend to be lower than the nominal levels. As we mentioned in

Dose	σ	n_0	n_1	n_2	Using true correlations	Pearson residuals	Using fixed correlations		
							$\rho = 0.99$	$\rho = 0.60$	$\rho = 0.00$
P1	0.01	2	21	19	3.79	3.83	4.04	3.98	4.00
P1	0.02	2	21	19	15.44	15.58	16.56	16.12	16.18
P1	0.03	2	21	19	35.76	36.11	38.21	37.34	37.47
P1	0.04	2	21	19	64.44	64.88	68.45	67.81	68.07
P2	0.01	4	21	19	3.07	3.28	4.79	4.71	5.16
P2	0.02	4	21	19	12.46	13.29	19.94	18.36	20.11
P2	0.03	4	21	19	28.05	29.85	43.83	42.92	47.21
P2	0.04	4	21	19	50.52	53.95	78.44	76.81	84.34
P3	0.01	6	21	19	3.24	3.66	5.70	4.11	6.72
P3	0.02	6	21	19	13.54	15.22	23.82	16.99	27.19
P3	0.03	6	21	19	29.59	32.99	51.57	37.38	61.57
P3	0.04	6	21	19	53.55	55.55	92.98	67.71	111.39

Table 7.3: Comparison of mean squared errors

section 7.2.3, when the sample sizes are not large enough variance estimate 1 underestimates the true variance of $\hat{\theta}$ and therefore we may expect the actual coverages to be lower than the nominal levels.

4. When fixed working correlation matrices are used at each iteration, the coverage probabilities of confidence intervals using variance estimate 2 (section 7.2.2) agree well with their nominal levels, even in small samples, regardless of whether the fixed correlation matrix is close to the true correlation matrix or not.

5. Based on the study, we conclude that the mean squared errors of the estimators is minimal when we use Pearson residuals to estimate the correlation matrices rather than using arbitrarily chosen fixed correlation matrices. Therefore, we favour using Pearson

residuals to estimate the correlation matrices. Recall that when we use Pearson residuals to estimate the correlation matrices our formula (Equation 7.6) underestimates the error of the estimate when the sample sizes are small. We hope to pursue further work in this area to find a more accurate estimate for the error of the estimate when using Pearson residuals; this might bring the coverage probabilities of the resulting confidence intervals closer to their nominal values.

6. Comparing mean squared errors of the estimators with the same number of temperatures on the plateau, we conclude that the mean squared errors of the estimators are roughly proportional to σ^2, as we expect based on the small σ limiting results.

7. Independence estimating equations appear to produce confidence intervals with correct coverage probabilities. However, mean squared errors of the estimators appear to grow as the number of plateau temperatures is increased.

7.3 GEE estimates when unbleached and bleached curves correspond to different relative errors

Recall that from the physical model suggested in Chapter 2, $\sigma^2 = \frac{V(N_{ijk})}{E^2(N_{ijk})}$, where N_{ijk} denotes the number of emitting grains of the ikth sample at the jth temperature. Suppose j and j' are two adjacent temperatures belonging to the plateau region. It is reasonable to assume that number of emitting grains at temperature j is close to that of temperature j'. This suggests that first two moments of N_{ijk} are close to those of $N_{ij'k}$ which in turn suggests that σ for different response curves over the plateau temperature are not very different. However, if the bleaching had an effect on the variability of emitting grains relative to the mean number of emitting grains, the unbleached and bleached curves may correspond to different σ. The theory derived so far assumes that all the response curves correspond to a common σ. In this section we suggest the necessary modifications for finding estimates when all the unbleached curves correspond to a common σ but this is different from the common σ for the bleached curves.

7.3.1 Solution of estimating equations

We denote the different σ's for unbleached and bleached curves by σ_1 and σ_2 respectively. In this section, we use the same notation as for the common σ case. However, since we now need to distinguish unbleached and bleached data we use an additional subscript 1 or 2 on each term indicating whether it corresponds to the unbleached data or the bleached data respectively. Let n_1 and n_2 denote the total number of unbleached and bleached samples and Ω_{1i} and Ω_{2j} denote the correlation matrices for the vectors of observations taken on ith unbleached ($i = 1, \ldots, n_1$) and jth bleached ($j = 1, \ldots, n_2$) samples. As for the common σ case, we begin with the estimating equations for the case of known correlation matrices

$$\sum_{i=1}^{n_1} \left\{ F_{1i}^T \left(\sigma_1^2 D_{f_{1i}} \Omega_{1i} D_{f_{1i}} \right)^{-1} (Y_{1i} - f_{1i}) \right\} + \sum_{j=1}^{n_2} \left\{ F_{2j}^T \left(\sigma_2^2 D_{f_{2j}} \Omega_{2j} D_{f_{2j}} \right)^{-1} (Y_{2j} - f_{2j}) \right\} = 0. \tag{7.7}$$

Again, we discuss two procedures for estimating the equivalent dose:

1. We assume that the correlation between observations taken at different temperatures on the plateau does not depend on the dose received by the sample. This suggests that for all the unbleached observations $Cor(Y_i)$ is some unknown correlation matrix which we denote by Ω_U. Note that

$$\Omega_U = E(\epsilon_{1i}\epsilon_{1i}^T) = E\left(\frac{(Y_{1i} - f_{1i})}{\sigma_1 f_{1i}} \right) \left(\frac{(Y_{1i} - f_{1i})}{\sigma_1 f_{1i}} \right)^T, \quad \text{for } i = 1, \ldots, n_1.$$

   Thus, $\sigma_1^2 \Omega_U$ can be estimated as $S_1 = \frac{1}{n_1} \sum_{l=1}^{n_1} \left[\frac{(Y_{1l} - \hat{f}_{1l})}{\hat{f}_{1l}} \right] \left[\frac{(Y_{1l} - \hat{f}_{1l})}{\hat{f}_{1l}} \right]^T$. The subscript l indicates lth unbleached disc. Similarly for the bleached data, $\sigma_2^2 \Omega_B$ can be estimated as $S_2 = \frac{1}{n_2} \sum_{l=1}^{n_2} \left[\frac{(Y_{2l} - \hat{f}_{2l})}{\hat{f}_{2l}} \right] \left[\frac{(Y_{2l} - \hat{f}_{2l})}{\hat{f}_{2l}} \right]^T$. Therefore, we can estimate θ without knowledge of σ_1 and σ_2. We refer to this procedure as Scheme 1. The notion of estimating $\sigma_1^2 \Omega_U$ and $\sigma_2^2 \Omega_B$ using Pearson residuals is similar to that of using Pearson residuals to estimate the correlation matrices in the common σ case; however we note that $\sigma_1^2 \Omega_U$ and $\sigma_2^2 \Omega_B$ are not correlation matrices.

2. Assume that the ratio $\frac{\sigma_2}{\sigma_1}$ is some known fixed quantity λ. In practice, λ can be chosen based on prior knowledge or in our problem using the results for individual curve

fittings. In this case, Equation 7.7 simplifies to give

$$\sum_{i=1}^{n_1}\left\{F_{1i}^T(D_{f_{1i}}\Omega_{1i}D_{f_{1i}})^{-1}(Y_{1i}-f_{1i})\right\}+\sum_{j=1}^{n_2}\left\{F_{2j}^T\left(\lambda^2 D_{f_{2j}}\Omega_{2j}D_{f_{2j}}\right)^{-1}(Y_{2j}-f_{2j})\right\}=0.$$

Now as suggested in Liang and Zeger [43], we can obtain generalized estimating equations by replacing the unknown correlation matrices by some working correlation matrices R_1 and R_2. We do not upgrade λ at each iteration, but once we find the estimates $\hat{\theta}$ we can use the error sum of squares for the unbleached and bleached response curves to estimate λ (See section 7.3.2). Thus, we estimate θ assuming λ is known. However, when we estimate the variance of the suggested estimate we do not necessarily believe that λ is correctly specified. In section 7.3.2, we provide a formula for the standard error of the suggested estimate. The estimating equations we introduce here can be considered as modifications of those using working correlation matrices for the common σ case; as before we have two choices for choosing working correlation matrices: We can use Pearson residuals at each iteration to find suitable working correlation matrices or we can use fixed correlation matrices at each iteration. We refer to this procedure as Scheme 2. (It is not crucial to treat λ as fixed at each iteration. We may estimate λ at each iteration using the error sums of squares for the unbleached and bleached data. In this case, we need to assume that λ converges (as $n \longrightarrow \infty$) to some fixed ratio λ_0; however we do not have to believe that λ_0 is the true ratio $\frac{\sigma_2}{\sigma_1}$.)

Next we describe an algorithm for solving the suggested estimating equations.

1. Find initial estimates for the parameters. We used the parameter estimates obtained for the independent curve fittings at different temperatures as initial estimates.

2. Compute the fitted values $\hat{f}_1,\ldots,\hat{f}_n$ at the current estimate $\hat{\theta}$; note that the first $n_1 \times n_0$ fitted values are computed from the unbleached response function while the remaining $n_2 \times n_0$ fitted values are computed from the bleached response function. Let $\hat{f}$ denote the $n \times 1$ vector whose entries are $\hat{f}_1,\ldots,\hat{f}_n$. Let $D_{\hat{f}}$ denote the $n \times n$ diagonal matrix with entries of $\hat{f}$ along its diagonal.

3. To obtain estimates using Scheme 1, let R be the $n \times n$ block diagonal matrix with the first n_1 blocks equal to S_1 and the last n_2 blocks equal to S_2 for S_1 and S_2 defined earlier. To obtain estimates using Scheme 2, let R be the $n \times n$ block diagonal matrix with the first n_1 blocks equal to the working correlation matrix for the unbleached data (say R_1) and the last n_2 blocks equal to the working correlation matrix for the bleached data (say R_2) with each of the entries multiplied by λ^2 (i.e. last n_2 blocks are equal to $\lambda^2 R_2$).

4. Let $V^{(k)} = D_{\hat{f}}^{(k)} R^{(k)} D_{\hat{f}}^{(k)}$, where we use the superscript k to indicate that each term is evaluated at $\hat{\theta}_k$ (the parameter estimates from the kth iteration). Let $W^{(k)}$ denote the inverse of $V^{(k)}$.

5. Let $F^{(k)}$ denote the $n \times p$ matrix whose (i,j)th element is $\frac{\partial f_i}{\partial \theta_j}$ evaluated at $\hat{\theta}_k$; note that the first $n_1 \times n_0$ rows of F are computed as the derivatives of unbleached response function with respect to the components of θ while the remaining $n_2 \times n_0$ rows are those for the bleached data.

6. At the $(k+1)$st iteration, θ is estimated from $\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \hat{\zeta}^{(k)}$, where

$$\hat{\zeta}^{(k)} = \left((F^{(k)})^T W^{(k)} F^{(k)} \right)^{-1} (F^{(k)})^T W^{(k)} \left(Y - \hat{f}^{(k)} \right).$$

7. Iterate until desired convergence. In the software we developed, the convergence criterion was met when the absolute difference in the parameter estimates of successive iterations was less than 10^{-3}.

7.3.2 Error of the estimate

Estimating σ_1 and σ_2:

Let n_u and n_b denote the total number of unbleached and bleached observations respectively. Let D_1 and D_2 denote the $n_u \times n_u$ and $n_b \times n_b$ diagonal matrices with response functions for unbleached and bleached data as diagonal entries respectively. Let F_1 denote the $n_u \times p$ matrix whose (i,j)th element is $\frac{\partial f_{1i}}{\partial \theta_j}$; similarly let F_2 be the corresponding partial derivative

matrix for the bleached data. Let $W_{\mho_1}$ and $W_{\mho_2}$ be the $n_u \times n_u$ and $n_b \times n_b$ matrices defined by:

$$\begin{aligned} W_{\mho_1} &= (D_1 \mho_1 D_1)^{-1}, \\ \text{and } W_{\mho_2} &= \left(D_2 \lambda^2 \mho_2 D_2\right)^{-1}. \end{aligned}$$

The results for the common σ case (Section 7.2.3) can simply be extended to arrive at the following formulae for estimating σ_1^2 and σ_2^2 :

$$\begin{aligned} \hat{\sigma_1}^2 &= \frac{\hat{\epsilon_1}^T \hat{\epsilon_1}}{tr(Q_1 \Omega_1)}, \\ \text{where } Q_1 &= (I - H_1^T)(I - H_1) \text{ and } H_1 = D_1^{-1} F_1 (F_1^T W_{\mho_1} F_1)^{-1} F_1^T W_{\mho_1} D_1, \\ \text{and } \hat{\sigma_2}^2 &= \frac{\hat{\epsilon_2}^T \hat{\epsilon_2}}{tr(Q_2 \Omega_2)}, \\ \text{where } Q_2 &= (I - H_2^T)(I - H_2) \text{ and } H_2 = D_2^{-1} F_2 (F_2^T W_{\mho_2} F_2)^{-1} F_2^T W_{\mho_2} D_2. \end{aligned}$$

To estimate σ_1^2 and σ_2^2 we need to replace the unknown correlation matrices with their estimates using Pearson residuals.

Estimating λ

Once we estimate σ_1 and σ_2 we can estimate λ as the ratio $\frac{\hat{\sigma}_2}{\hat{\sigma}_1}$.

Variance covariance matrix for $\hat{\theta}$

Assuming

1. $\hat{\sigma}_1^2 R_1 \xrightarrow{p} \sigma_1^2 \mho_1$, where $R_1 = \frac{1}{n_1} \sum_{l=1}^{n_1} \left[\frac{(Y_{1l} - \hat{f}_{1l})}{\hat{f}_{1l}}\right] \left[\frac{(Y_{1l} - \hat{f}_{1l})}{\hat{f}_{1l}}\right]^T$ and $\mho_1$ is some fixed correlation matrix (not necessarily the true correlation matrix Ω_1) and
2. $\hat{\sigma}_2^2 R_2 \xrightarrow{p} \sigma_2^2 \mho_2$, where $R_2 = \frac{2}{n_2} \sum_{l=2}^{n_2} \left[\frac{(Y_{2l} - \hat{f}_{2l})}{\hat{f}_{2l}}\right] \left[\frac{(Y_{2l} - \hat{f}_{2l})}{\hat{f}_{2l}}\right]^T$ and $\mho_2$ is some fixed correlation matrix (not necessarily the true correlation matrix Ω_2)

we can derive the following formula for the variance covariance matrix for $\hat{\theta}$ from Scheme 1:

$$Var(\hat{\theta}) = \left(\frac{1}{n_3} \sum_{i=1}^{n_3} F_i^T W_{\mho} F_i\right)^{-1} \left(\frac{1}{n_3^2} \sum_{i=1}^{n_3} F_i^T W_{\mho} Var(Y_i) W_{\mho} F_i\right) \left(\frac{1}{n_3} \sum_{i=1}^{n_3} F_i^T W_{\mho} F_i\right)^{-1} \quad (7.8)$$

where $W_{\mho} = \begin{cases} D_{f_{1i}}\sigma_1^2\mho_1 D_{f_{1i}}, & \text{for } i = 1,\ldots,n_1 \\ \lambda^2 D_{f_{2j}}\sigma_2^2\mho_2 D_{f_{2j}}, & \text{for } i = n_1+1,\ldots,n_3 \text{ and } j = i-n_1. \end{cases}$

The same result holds for the estimate $\hat{\theta}$ from Scheme 2, under the assumption that the working correlation matrices R_1 and R_2 converge to some fixed correlation matrices $\mho_1$ and $\mho_2$.

We have two choices for estimating $Var(\hat{\theta})$.

1. Letting $V_\Omega = \begin{cases} D_{f_{1i}}\sigma_1^2\Omega_1 D_{f_{1i}}, & \text{for } i = 1,\ldots,n_1 \\ D_{f_{2j}}\sigma_2^2\Omega_2 D_{f_{2j}}, & \text{for } i = n_1+1,\ldots,n_3;\ j = i-n_1 \end{cases}$
we can write

$$Var(\hat{\theta}) = \left(\frac{1}{n_3}\sum_{i=1}^{n_3} F_i^T W_{\mho} F_i\right)^{-1} \left(\frac{1}{n_3^2}\sum_{i=1}^{n_3} F_i^T W_{\mho} V_\Omega W_{\mho} F_i\right) \left(\frac{1}{n_3}\sum_{i=1}^{n_3} F_i^T W_{\mho} F_i\right)^{-1}.$$

We can estimate $Var(\hat{\theta})$ by replacing the unknown $\sigma_1^2, \sigma_2^2, \Omega_1$ and Ω_2 by their estimates. For ease of presentation, we refer to this estimate as variance estimate 4.

2. Following suggestions of Liang and Zeger [43], we can replace $Var(Y_i)$ in Equation 7.5 by $(Y_i - \hat{f}_i)(Y_i - \hat{f}_i)^T$. As in the previous case, an estimate for the approximate variance can be obtained by replacing the unknown parameter θ by $\hat{\theta}$. We refer to this estimate as variance estimate 5.

7.3.3 Confidence intervals

Let $\hat{\gamma}$ be the estimate for the equivalent dose obtained as the pth component of the GEE estimate suggested for the case of different relative errors. Let F be the $n \times p$ matrix whose (i,j)th element is $\frac{\partial f_i}{\partial \theta_j}$. Let $\hat{V}(\hat{\theta})$ be an estimate for the variance of $\hat{\theta}$ obtained as suggested in variance estimate 4 or variance estimate 5. Let $s_{\hat{\gamma}}$ be the square root of the pth diagonal element of $\hat{V}(\hat{\theta})$. Then, an approximate $(1-\alpha)100\%$ confidence interval for the equivalent dose can be constructed by taking $\hat{\gamma} \mp s_{\hat{\gamma}} z_{\alpha/2}$ as lower and upper confidence limits ($z_{\alpha/2}$ denotes the upper $\alpha/2$th quantile for a standard normal distribution).

7.4 Worked Example

We developed software using the programming language FORTRAN to implement the procedures described earlier in this chapter. Now we demonstrate the suggested methodology using a real data set. The data set used for this example (code WFP2-7R1) was kindly provided to us by D.J.Huntley. A plot of the quasi-likelihood estimates for the equivalent dose by fitting the saturating exponential model (see Chapter 2) at different temperatures is illustrated in Figure 7.3. The vertical bars indicate plus or minus one standard error limits. An immediate observation from this plot is that the standard errors of the estimate begin to

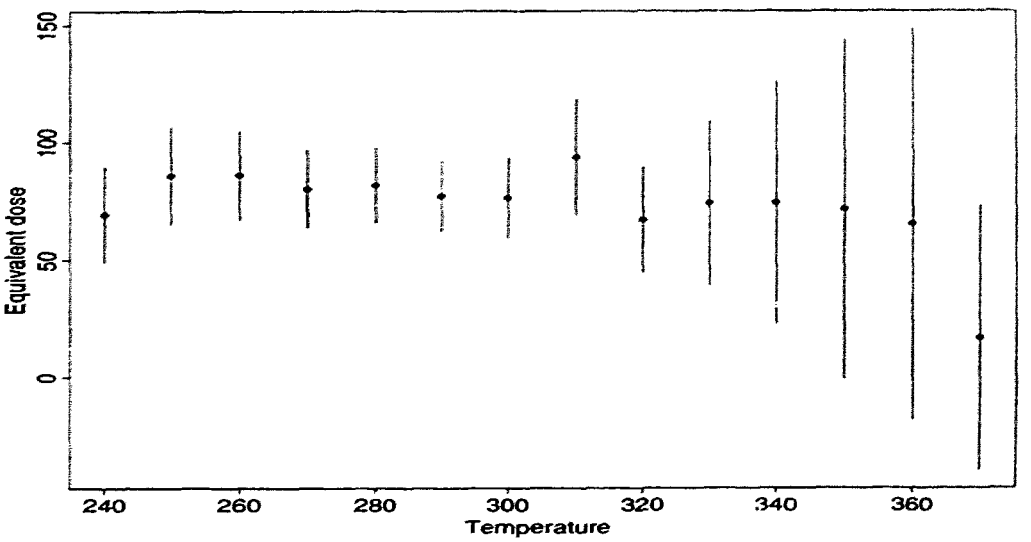

Figure 7.3: Plot of $\hat{\gamma}$ vs temperature with error limits: WFP2-7R1

grow over the temperature region 320^0C to 360^0C. (We have observed similar pattern with a few other real data sets we have analyzed.) At the moment we do not understand why this is so. As we already mentioned, it is not clear cut whether or not certain temperatures belong to the plateau. From the plot of $\hat{\gamma}$ vs temperature (Figure 7.3), we notice a plateau starting around 250^0C. According to Aitken [1], the stable traps usually means traps for which the glow peak occurs at 300^0C or higher. However, for preheated samples the plateau could begin at temperatures lower than 300^0C (Huntley, personal communication). For illustration, we estimated the equivalent dose treating the observations corresponding to

the temperatures given in Tables 7.4 and 7.5 as being belonging to the plateau. The results given in these tables were obtained assuming correlation matrices for the unbleached and bleached data are different. The estimates given in Table 7.4 were obtained using the sample correlation matrices computed from the Pearson residuals to estimate the unknown correlation matrices. The results described in Table 7.5 were obtained as described in Section 7.3. For the common σ case, the reported confidence intervals were obtained using a t quantile with an approximate degrees of freedom as described in Theorem 6. For the different σ case, we used the sum of the approximate degrees of freedom for the error sum of squares for the unbleached and bleached data. Tables 7.4 and 7.5 present the results for fitting a common relative error σ and different σ's for the unbleached and bleached curves respectively. In these tables, the first column (n_0) indicates the number of temperatures on the plateau; the corresponding temperatures are given in column 2. For the different σ case, with 6 observations corresponding to the temperature range 270-320 the program did not converge. For a given type of electron trap, the glow curve is a single peak about 50^0C in width (Aitken [1]). It is possible that at 270^0C and 320^0C we are emptying different type of traps.

Conclusions

The estimates for σ_1 and σ_2 from fitting separate error factors for unbleached and bleached curves indicate that a common error factor is sensible for this data. The parameter estimates obtained by fitting separate error factors and a common error factor are not very different. Another important observation is that in the range $330^0C - 360^0C$, the standard errors of the estimates have gone up even though the common error factor σ has gone down. The large standard errors for the estimates using data collected in the range $330^0C - 360^0C$ makes the estimates from analyses at a single temperature only, less useful in practice (for example for the common σ case, the standard errors of the estimates using only $340^0C, 350^0C$ and 360^0C were 51.48, 72.49 and 83.40). In this temperature range, use of the suggested procedures for obtaining combined estimates has clearly resulted in a gain in precision of the estimate. However, in the range $270^0C - 310^0C$ the gain in precision using more observations in the plateau region is quite small; in fact in some cases the standard error

n_0	Temperatures on the plateau	$-\hat{\gamma}$	Standard Error of $\hat{\gamma}$	An approximate 95% CI		$\hat{\sigma}$
				Lower bound	Upper bound	
1	270	80.21	16.63	46.42	114.01	0.08
2	270-280	79.86	15.79	47.80	111.92	0.08
3	270-290	85.36	15.82	53.28	117.45	0.08
4	270-300	84.14	13.32	56.98	111.30	0.08
5	270-310	80.20	12.23	55.29	105.10	0.07
6	270-320	88.50	13.19	61.67	115.34	0.07
1	330	73.92	34.84	3.12	144.73	0.07
2	330-340	74.24	22.84	27.65	120.82	0.07
3	330-350	75.63	22.07	30.69	120.56	0.07

Table 7.4: Estimate for the equivalent dose ($\hat{\gamma}$) using a common σ: WFP2-7R1 data

n_0	Temp.	$-\hat{\gamma}$	Standard Error of $\hat{\gamma}$	An approximate 95% CI		$\hat{\sigma}_1$	$\hat{\sigma}_2$
				Lower bound	Upper bound		
1	270	80.21	15.51	48.87	111.56	0.08	0.08
2	270-280	83.34	15.42	52.20	114.48	0.07	0.07
3	270-290	81.31	13.78	53.49	109.13	0.07	0.07
4	270-300	82.06	12.00	57.83	106.28	0.07	0.07
5	270-310	71.63	10.03	51.39	91.87	0.07	0.07
1	330	73.92	31.90	9.45	138.39	0.06	0.07
2	330-340	74.40	21.42	31.15	117.66	0.06	0.07
3	330-350	75.89	20.45	34.61	117.16	0.06	0.07

Table 7.5: Estimate for the equivalent dose ($\hat{\gamma}$) using different σ: WFP2-7R1 data

of the resulting estimate may even increase. This is not surprising; since the correlations between the observations are very high, increasing the number of plateau temperatures only provides a little extra information about the value of γ while with each added temperature we have to add more nuisance parameters. Furthermore, when we take a wide range of temperatures as belonging to the plateau, it is possible that observations corresponding to the added temperatures may not relate to the same type of trap. Thus, it is important to note that taking more observations over the plateau may not necessarily improve the quality of the resulting estimate.

7.5 Discussion

In the previous chapters, we discussed estimating the equivalent dose from the data collected at a given temperature. These estimates are plotted against the temperature. The region over which these estimates do not vary with the temperature is believed to represent electron traps that have not been subjected to leakage over the burial time. It is therefore assumed that all estimates for the equivalent dose corresponding to the plateau region estimate the same quantity, which is the equivalent dose corresponding to the stable traps (Aitken [1]). In this chapter, we discussed the problem of combining these several estimates to obtain a more precise estimate for the equivalent dose.

Since data collected at different temperatures are obtained using the same samples, observations collected over the plateau region are correlated (in fact the correlations are very high). We proposed a procedure closely related to that of Liang and Zeger [43] to estimate the equivalent dose from the correlated data. Algorithms were presented for solving the suggested estimating equations for the case where all the response curves correspond to a common σ (relative error in a single measurement) and also for the case where all the unbleached curves correspond to a common σ which is different from the common σ for the bleached response curves. The estimation procedures proposed for estimating the model parameters have to be augmented by some procedure for estimating the unknown correlation matrices. Following the terminology of Liang and Zeger [43], we referred to these as "working correlation matrices". We examined two plausible working correlation matrices:

using working correlation matrices computed from the Pearson residuals at each iteration and using arbitrarily chosen fixed correlation matrices (derived, in practice, from experience with other data sets) as working correlation matrices.

Since different nuisance parameters have to be fitted at each dose level, the number of parameters increases with the addition of each observation on the plateau. In fact, the parameter of interest and the relative error in a single measurement σ (if it is assumed to be common for all the curves) are the only parameters that are common to all the temperatures. Since many nuisance parameters have to be estimated, the effective sample size is much smaller than the nominal sample size. Consequently, confidence intervals using standard normal quantiles were found to have lower coverage probabilities than their nominal levels. As the number of plateau temperatures increases, the discrepancy between observed coverages and nominal levels was found to grow. For thermoluminescence data, sample sizes are quite small (usually under 40 discs). The relative error in a single measurement, σ, is quite small (around 8%). We examined the small σ asymptotic behavior of the suggested estimates and proposed using a t quantile for computing confidence intervals; a formula for the approximate degrees of freedom for the corresponding t distribution was provided using Satterthwaite [53] formula. Finite sample performance of the suggested asymptotic theory was examined by a Monte Carlo study. Confidence intervals with an approximate t quantile based on the estimate from the iterative scheme that uses fixed correlation matrices as working correlation matrices were found to have coverage probabilities closer to their nominal levels. As the number of plateau temperatures increases, confidence intervals based on the estimate from the iterative scheme that uses sample correlation matrices of Pearson residuals as working correlation matrices were found to have lower coverages than their nominal levels, even with the suggested t quantile. As we clarified in section 7.2.3, when the sample sizes are not large enough, the suggested formula underestimates the true error of the estimate from the scheme that uses working correlation matrices computed from Pearson residuals. However, since estimates from the latter scheme were found to have smaller mean squared errors than those using fixed working correlation matrices (see Section 7.2.5), we favour using Pearson residuals to compute working correlation matrices. We hope to pursue

further research in this area to find a better estimate for the error of the estimate from the scheme that uses Pearson residuals.

In section 7.4, we demonstrated the proposed methodology using a real data set. It is worth noting that the increase in the precision of the estimate (i.e. the reduction in the standard error) for taking more observations on the plateau is quite small; in fact in some cases the standard error of the resulting estimate may even increase. Also, it is worth noting that, for some data sets it is not clear cut whether or not certain temperatures belong to the plateau. Addition of an observation which does not belong to the plateau makes the resulting estimate less accurate (more biased). Taking these facts into consideration, we suggest not using observations which are not convincingly obvious as belonging to the plateau.

Chapter 8

Testing the normality of random errors

In this chapter, we describe procedures for testing the normality of random errors. Section 8.1 briefly describes the tests based on the empirical distribution function (EDF) of the residuals for this purpose.

In Section 8.2, we describe a procedure based on the EDF to test the assumption of normality without assuming that the fitted model is correct. We refer to this less restricted model as Model 1. In this model we fit different mean and variance parameters at each dose level. Thus, as we collect more observations, the number of fitted parameters also increases. Usually, in thermoluminescence studies not more than five observations are collected at each dose level. Therefore, the fitted standardized residuals need not be asymptotically normally distributed even if the random errors are. We derive the true distribution of the fitted standardized residuals assuming that the random errors are normally distributed. We test the assumption of normality of the random errors by checking to see if the fitted standardized residuals follow the derived true distribution.

Section 8.3 outlines a procedure for testing the assumption of normality of random errors assuming that the fitted model is correct. We assume that the observed Y_{ij}'s have mean $\mu_i(\theta)$ and variance $\sigma^2\mu_i^2(\theta)$. We refer to this model as Model 2. In Model 2, both mean and variance are functions of a fixed number of unknown parameters. Therefore, we can

find consistent estimates for mean $\mu_i(\theta)$ and variance $V_i(\sigma, \theta)$ by choosing a large enough sample. If the random errors are normally distributed, the fitted standardized residuals are approximately normally distributed. Here we test the assumption of normality of random errors by checking to see if the fitted residuals are approximately normally distributed. Weak convergence of the empirical processes is established in Section 8.2.6. In 8.5, we demonstrate the theoretical results discussed in this chapter using a real data set. Section 8.6 summarizes the chapter.

8.1 EDF tests

We begin by reviewing EDF tests in the context of iid sampling and then extend the discussion to tests of the distributions of residuals. Let $x_1 \leq x_2 \cdots \leq x_n$ denote the order statistics (that is, the observations arranged in the increasing order) from a sample of n values of some variate X with cumulative distribution function (cdf) G. Let $F(x)$ be some hypothesized distribution function of the data. For the moment we take any parameter in F to be known and discuss tests of the simple null hypothesis that $G = F$. The empirical distribution function (EDF) is defined as, $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I[x_i \leq x]$. Here I denotes the indicator function,

$$I_A = \begin{cases} 1, & \text{if the condition } A \text{ holds} \\ 0, & \text{otherwise.} \end{cases}$$

8.1.1 EDF statistics

Any statistic which measures the difference between $F_n(x)$ and $F(x)$ is called an EDF statistic. Here we describe only the Cramér-von Mises statistic and the Anderson-Darling statistic that we used in this work. A discussion of these statistics can be found in Stephens (1986). The Cramér-von Mises statistic is defined as

$$W_n^2 = n \int_{-\infty}^{\infty} \{F_n(x) - F(x)\}^2 \, dF(x).$$

The Anderson-Darling statistic is defined as

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{\{F_n(x) - F(x)\}^2}{[F(x)(1-F(x))]} dF(x).$$

Computing formulae for the above statistics:

Let F be the hypothesized distribution function and $z_i = F(x_i)$ be the probability integral transforms of the x_i's. The following computing formulae for the above statistics are provided in Stephens (1986):

$$\begin{aligned} W_n^2 &= \sum_{i=1}^n \left\{ z_i - \frac{(2i-1)}{2n} \right\}^2 + \frac{1}{12n} \\ A_n^2 &= -n - \frac{1}{n} \sum_{i=1}^n \{(2i-1)\ln z_i + (2n+1-2i)\ln(1-z_i)\}. \end{aligned} \tag{8.1}$$

8.1.2 Computing p-values

Let $W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{I[z_i \leq t] - t\}$. Then, under the null hypothesis, the process $W_n(t)$ converges weakly in $D[0,1]$ to the Brownian Bridge —a Gaussian process W with mean 0 and the covariance function $\rho(s,t) = Cov(W(s), W(t)) = \min(s,t) - st$ (cf. Billingsley [13].)

The limiting distribution of the Cramér-von Mises statistic (W_n^2) is then that of $W^2 = \sum_{i=1}^{\infty} \lambda_i Z_i^2$, where the Z_i's are independent $N(0,1)$ variables and the λ_i's are the eigenvalues of the covariance kernel $\rho(s,t)$, namely, the solutions of the eigenvalue equation $\int_0^1 \rho(s,t) f(t) dt = \lambda f(s)$. For the case at hand (iid sample and simple null hypothesis) these eigenvalues can be found analytically. When we discuss fitted residuals a numerical method is required; this is discussed in Section 8.4.

Let w be the value of the test statistic obtained using the computing formulae (Equation 8.1) provided earlier. An approximate p-value for testing the hypothesis that x_i's follow the hypothesized distribution function $F(x)$ is $P(\sum_{i=1}^{\infty} \lambda_i \chi_i^2 \geq w)$ where χ_i^2's denote a set of independent Chi-squared random variables with 1 degree of freedom and λ_i's are the estimated eigenvalues for $\rho(s,t)$ truncating the sum to a finite number of terms. We have software that uses Imhof's [39] method to compute the above probability. Literature related to the computation of the above probability together with a comparison of the methods can be found in Chen [16].

For the Anderson-Darling statistic the same procedure can be used replacing the covariance kernel $\rho(s,t)$ by $\frac{\rho(s,t)}{\sqrt{s(1-s)t(1-t)}}$.

Now we describe how to apply the EDF tests for our problem. The key problems in applying the EDF tests are,

1. Finding the probability integral transforms to compute the z_i's needed in computing the test statistic (Equation 8.1).

2. Finding the covariance kernel of the appropriate Gaussian process needed in computing the p-value.

In Section 8.2, we describe testing the assumption of normality without assuming that the fitted model is correct. In Section 8.3, we describe testing the assumption of normality assuming that the fitted model is correct. In each case, we describe how to compute the probability integral transforms and the covariance kernel of the appropriate Gaussian process.

8.2 Application of EDF tests: Model 1

We refer to the model with no assumptions about the mean and variance functions as Model 1. This less restricted model is expressed as $Y_{ij} = \mu_i + \sigma_i \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0,1); \quad i = 1,\ldots,k, \quad j = 1,\ldots,n_i$; the suffixes i and j respectively denote the dose level and the replicate. We wish to test the assumption of normality of the random errors.

8.2.1 Computing the test statistic

Let $\hat{\mu}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ be the least squares estimate for μ_i and $\hat{\epsilon'}_{ij} = Y_{ij} - \hat{\mu}_i$. Note that $E(\hat{\epsilon'}_{ij}) = 0$ and $V(\hat{\epsilon'}_{ij}) = \frac{(n_i-1)}{n_i}\sigma_i^2$. We study the standardized fitted residuals $\hat{\epsilon}_{ij} = \frac{(Y_{ij}-\hat{\mu}_i)}{\sqrt{\frac{(n_i-1)}{n_i}}\hat{\sigma}_i}$, where $\hat{\sigma}_i^2 = \sum_{j=1}^{n_i} \frac{(Y_{ij}-\hat{\mu}_i)^2}{(n_i-1)}$. The $\hat{\epsilon}_{ij}$'s replace the x_i's in the discussion of he previous section. Note, however, that the $\hat{\epsilon}_{ij}$'s are not iid. Let $G_{n_i}(\cdot)$ be the true distribution of $\hat{\epsilon}_{ij}$ when the random errors are normally distributed ($G_{n_i}(\cdot)$ depends on n_i but not on μ_i, σ_i or j).

Let $\nu = (n_i - 1)$ and $\hat{\tau}_{ij} = \hat{\epsilon}_{ij}\sqrt{\frac{(\nu-1)}{\nu-\hat{\epsilon}_{ij}^2}}$. According to Beckman *et. al.* [4], the variates $\hat{\tau}_{ij}$ follow the distribution $t_{\nu-1}$, which is the univariate student t distribution with degrees of freedom $\nu - 1$. The probability integral transforms of $\hat{\epsilon}_{ij}$ are therefore given by

$$u_{ij} = G_{n_i}(\hat{\epsilon}_{ij}) = t_{(\nu-1)}\left(\hat{\epsilon}_{ij}\sqrt{\frac{(\nu-1)}{(\nu-\hat{\epsilon}_{ij}^2)}}\right).$$

We notice that in a sample of identically distributed observations, if the sample values are in ascending order, so are the corresponding probability integral transforms. But this does not hold for our case of non identically distributed standardized residuals. Therefore, it is required to arrange the u_{ij}'s in ascending order after computing the probability integral transforms. Let $z_1, \ldots, z_N$ ($N = \sum_{i=1}^{k} n_i$) be the u_{ij}'s arranged in ascending order. The computing formulae (Equation 8.1) given in Section 8.1.1 can then be used to compute the test statistic, with n replaced by N.

8.2.2 An approximate p-value

Let $W_N(s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \{I[u_{ij} \leq s] - s\}$, and

$$\rho_{2,n_i}(s,t) = Cov\left(I\left(G_{n_i}(\hat{\epsilon}_{ij}) \leq s\right), I\left(G_{n_i}(\hat{\epsilon}_{ij}) \leq t\right)\right).$$

In Section 8.2.3 we show that the covariance kernel for the process $W_N(s)$ is

$$\rho(s,t) = \min(s,t) - st + \sum_{i=1}^{k}\sum_{j=1}^{n_i} \frac{n_i(n_i-1)}{N}\rho_{2,n_i}(s,t), \tag{8.2}$$

and show how to compute $\rho_{2,n_i}(s,t)$. The procedure described in Section 8.1.2 can be used to compute a p-value by replacing the λ_i's by the eigenvalues of $\rho(s,t)$.

Next we summarize the test procedure.

1. For each dose i, estimate μ_i and σ_i^2 using $\hat{\mu}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ and $\hat{\sigma}_i^2 = \frac{1}{(n_i-1)}\sum_{j=1}^{n_i}(Y_{ij} - \hat{\mu}_i)^2$.
2. Compute the standardized fitted residuals $\hat{\epsilon}_{ij} = \frac{(Y_{ij}-\hat{\mu}_i)}{\sqrt{\frac{(n_i-1)}{n_i}\hat{\sigma}}}$.
3. Compute the probability integral transforms $G_{n_i}(\hat{\epsilon}_{ij})$.

4. Order the probability integral transforms in ascending order. Let $z_1, \ldots, z_N$ be the ordered probability integral transforms.

5. Compute the Cramér-von Mises statistic

$$W_N^2 = \sum_{i=1}^{N} \left\{ z_i - \frac{(2i-1)}{2N} \right\}^2 + \frac{1}{12N},$$

or, the Anderson-Darling statistic

$$A_N^2 = -N - \frac{1}{N} \sum_{i=1}^{N} \{(2i-1) \ln z_i + (2N+1-2i) \ln(1-z_i)\}.$$

6. Find the eigenvalues of the covariance kernel $\rho(s,t)$ defined by the Equation 8.2. We compute approximations for these eigenvalues as described below:

   (a) Create the matrix Q whose elements are

   $$Q(s,t) = \frac{1}{m}\rho(s,t), \quad \text{for } s,t = \tfrac{1}{(m+1)}, \ldots, \tfrac{m}{(m+1)}.$$

   (Here m is the number of subdivisions of the unit interval. We chose $m = 150$.) For the Anderson-Darling statistic replace $\rho(s,t)$ by $\rho_A(s,t) = \frac{\rho(s,t)}{\sqrt{s(1-s)t(1-t)}}$.

   (b) Compute the eigenvalues $\lambda_1, \ldots, \lambda_m$ of the matrix Q. These eigenvalues provide estimates for the eigenvalues of $\rho(s,t)$. (See Section 8.4.)

7. Compute the p-value based on the asymptotic distribution of the test statistic as $P\left(\sum \lambda_i \chi_i^2 \geq w\right)$ where χ_i^2's denote a set of independent Chi-squared random variables each on 1 degree of freedom and w is the value of the test statistic computed in step 5.

8. Reject (or, do not reject) the null hypothesis if the p-value is less than (or, greater than) the desired significance level α.

8.2.3 Covariance kernel $\rho(s,t)$

In this section we prove that the covariance kernel for the process $W_N(s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \{I[u_{ij} \leq s] - s\}$ is

$$\rho(s,t) = \min(s,t) - st + \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{n_i(n_i-1)}{N} \rho_{2,n_i}(s,t).$$

The process $W_N(s)$ can be rewritten as

$$W_N(s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left[I\left(G_{n_i}(\hat{\epsilon}_{ij}) \leq s\right) - s \right].$$

Hence, $Cov\left(W_N(s), W_N(t)\right)$

$$\begin{aligned}
&= Cov\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left[I\left(G_{n_i}(\hat{\epsilon}_{ij}) \leq s\right) - s \right], \frac{1}{\sqrt{N}} \sum_{i'=1}^{k} \sum_{j'=1}^{n_{i'}} \left[I\left(G_{n_{i'}}(\hat{\epsilon}_{i'j'}) \leq t\right) - t \right] \right) \\
&= \frac{1}{N} Cov\left(\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left[I\left(G_{n_i}(\hat{\epsilon}_{ij}) \leq s\right) \right], \sum_{i'=1}^{k} \sum_{j'=1}^{n_{i'}} \left[I\left(G_{n_{i'}}(\hat{\epsilon}_{i'j'}) \leq t\right) \right] \right) \\
&= \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \sum_{i'=1}^{k} \sum_{j'=1}^{n_{i'}} Cov\left(I\left(G_{n_i}(\hat{\epsilon}_{ij}) \leq s\right), I\left(G_{n_{i'}}(\hat{\epsilon}_{i'j'}) \leq t\right) \right). \qquad (8.3)
\end{aligned}$$

Note that

$$Cov\left(I\left(G_{n_i}(\hat{\epsilon}_{ij}) \leq s\right), I\left(G_{n_{i'}}(\hat{\epsilon}_{i'j'}) \leq t\right) \right) = \begin{cases} 0, & \text{if } i \neq i' \\ & \text{for these correspond to different doses} \\ \rho_1(s,t), & \text{if } i = i', \quad j = j' \\ \rho_{2,n_i}(s,t), & \text{if } i = i', \quad j \neq j' \end{cases}$$

where $\rho_1(s,t) = Cov\left(I\left(G_{n_i}(\hat{\epsilon}_{ij}) \leq s\right), I\left(G_{n_i}(\hat{\epsilon}_{ij}) \leq t\right)\right)$, and
$\rho_{2,n_i}(s,t) = Cov\left(I\left(G_{n_i}(\hat{\epsilon}_{ij}) \leq s\right), I\left(G_{n_i}(\hat{\epsilon}_{ij}) \leq t\right)\right)$. We use the subscript n_i to indicate that $\rho_{2,n_i}(s,t)$ depends on n_i.

Computing $\rho_1(s,t)$ and $\rho_{2,n_i}(s,t)$:

If A and B are two events then $Cov\left(I_A, I_B\right)$ is easily seen to be $P(A \cap B) - P(A)P(B)$. So, $\rho_1(s,t) = \min(s,t) - st$ and

$$\begin{aligned}
\rho_{2,n_i}(s,t) &= P\left(G_{n_i}(\hat{\epsilon}_{ij}) \leq s, G_{n_i}(\hat{\epsilon}_{ij'}) \leq t\right) - st \\
&= P\left(\hat{\epsilon}_{ij} \leq G_{n_i}^{-1}(s), \hat{\epsilon}_{ij'} \leq G_{n_i}^{-1}(t)\right) - st \\
&= G_{(2,n_i)}\left(G_{n_i}^{-1}(s), G_{n_i}^{-1}(t), \rho_{ijj'}\right) - st,
\end{aligned}$$

where $G_{n_i}^{-1}(.)$ denotes the inverse of the true distribution of the standardized residuals and $G_{(2,n_i)}(\cdot,\cdot,\rho_{ijj'})$ is the joint cdf of $\hat{\epsilon}_{ij}$ and $\hat{\epsilon}_{ij'}$. Let $g(\cdot,\cdot)$ denote the joint density function

of $\hat{\epsilon}_{ij}$ and $\hat{\epsilon}_{ij'}$. Noting that the correlation between two fitted standardized residuals is $-1/(n_i-1)$ (see Section 8.2.4), we find $G_{2,n_i}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} g(x,y,-(n_i-1)^{-1})dxdy$. We give an expression for G_{2,n_i} as a univariate integral in Section 8.2.4.

Now from Equation 8.3 the covariance kernel for the Cramér-von Mises statistic is found to be

$$\rho(s,t) = Cov\left(W_N(s), W_N(t)\right) = \min(s,t) - st + \sum_{i=1}^{k}\left\{\frac{n_i(n_i-1)}{N}\rho_{2,n_i}(s,t)\right\}.$$

The covariance kernel for the Anderson-Darling statistic is

$$\rho_A(s,t) = \frac{\rho(s,t)}{\sqrt{s(1-s)t(1-t)}}.$$

To evaluate the above covariance functions we need the inverse distribution function of $\hat{\epsilon}_{ij}$ (i.e. $G_{n_i}^{-1}(\cdot)$) and the joint distribution function of $\hat{\epsilon}_{ij}$ and $\hat{\epsilon}_{ij'}$ (i.e. $G_{2,n_i}(\cdot,\cdot,\rho_{ijj'})$.)
To compute the inverse of the distribution function at y we note that,

$$x = G_{n_i}^{-1}(y) \Longleftrightarrow y \;=\; G_{n_i}(x) = t_{(\nu-1)}\left(x\sqrt{\frac{(\nu-1)}{(\nu-x^2)}}\right).$$

Hence, $t_{(\nu-1)}^{-1}(y) = x\sqrt{\frac{(\nu-1)}{(\nu-x^2)}}$. Thus,

$$G_{n_i}^{-1}(y) = x = \frac{\sqrt{\nu}t_{(\nu-1)}^{-1}(y)}{\sqrt{(\nu-1) + \left[t_{(\nu-1)}^{-1}(y)\right]^2}},$$

where $t_{\nu-1}^{-1}$ denotes the inverse of the student t distribution function with $\nu-1$ degrees of freedom; recall that $\nu = n_i - 1$.

8.2.4 The joint density of the standardized residuals

Ellenberg [29] provides the joint density of the standardized residuals for the linear regression model. First we briefly mention their result and use their result to obtain the joint density of the residuals in our problem.

Consider the general linear regression model, $Y = X\beta + \epsilon$, where β is a k dimensional vector of unknown parameters, and X is fixed and of full rank. Let $M = I_n - X(X^TX)^{-1}X^T$,

where I_n denotes the $n \times n$ identity matrix. Let $\hat{u}_i = Y_i - X_i\hat{\beta}$, where X_i is the ith row of X and $\hat{\beta}$ is the least squares estimate for β. Let $S^2 = \sum_{i=1}^{n} \hat{u}_i{}^2$.

Ellenberg [29] defines the standardized residuals as $\zeta_i = \hat{u}_i/(Sm_{ii}^{1/2}) \quad i = 1, \ldots, n$, where m_{ii} is the ith diagonal element of M. Consider a subset of p residuals. Without loss of generality take these to be the first p residuals. Let M_p be the corresponding $(p \times p)$ principal minor of M. If M_p^{-1} exists, according to Ellenberg [29], the joint density of the p standardized residuals can be written as

$$f(\zeta_1, \zeta_2, \ldots, \zeta_p) = \frac{\Gamma(\nu + p/2)|M_p^{-1}|^{1/2} \prod_{i=1}^{p} m_{ii}^{1/2}}{\pi^{p/2}\Gamma(\nu)} \times \left[1 - \sum_{i=1}^{p}\sum_{j=1}^{p}(m_{ii}m_{jj})^{1/2}m^{ij}\zeta_i\zeta_j\right]^{\nu-1},$$

where m^{ij} are the elements of M_p^{-1}, $\nu = (n-k-p)/2$, and the probability density function is defined over the region

$$\sum_{i=1}^{p}\sum_{j=1}^{p}(m_{ii}m_{jj})^{1/2}m^{ij}\zeta_i\zeta_j \leq 1.$$

Ellenberg [29] uses $S^2 = \sum_{i=1}^{n} \hat{u}_i^2$ which is actually $n-k$ times the usual estimate for σ^2. In this work we defined standardized residuals replacing S by s where $s^2 = S^2/(n-k)$ and use the notation $\hat{\epsilon}_i = \hat{u}_i/(sm_{ii}^{1/2})$.

Note that $Corr(\hat{u}_j, \hat{u}_{j'}) = \frac{m_{jj'}}{\sqrt{m_{jj}m_{j'j'}}} (= \rho$ say). Let m^{ij} be the ijth element of M^{-1}. It is easy to check that $m^{jj}m_{jj} = \frac{1}{1-\rho^2}$ and $m^{jj'}\sqrt{m_{jj}m_{j'j'}} = -\frac{\rho}{(1-\rho^2)}$. Thus, the joint density of ζ_j and $\zeta_{j'}$ can be written as

$$f(u, v; \rho, \nu) = \frac{\nu}{\pi}\frac{1}{\sqrt{(1-\rho^2)}}\left\{1 - \frac{[u^2 - 2\rho uv + v^2]}{(1-\rho^2)}\right\}^{(\nu-1)},$$

and is defined over the region $\sum_{l=1}^{2}\sum_{m=1}^{2}(m_{ii}m_{jj})^{1/2}m^{ij}u_iu_j \leq 1$ which is the region $\frac{u^2-2\rho uv+v^2}{1-\rho^2} \leq 1$, or, in elliptical polar coordinates the region $r^2 \leq 1$. Thus, the support of the joint density of two fitted residuals ζ_{ij}'s (as defined by Ellenberg [29]) is the unit circle.

Next we apply the above result to obtain the joint density of two residuals in our problem. When two residuals come from different temperatures they are independent. For two residuals from the same dose level, we need to compute the correct correlation ρ. But this is simply the correlation between $\epsilon_i - \bar{\epsilon}$ and $\epsilon_j - \bar{\epsilon}$ in a sample of size n_i. This is easily seen

to be $-1/(n_i - 1)$. For notational convenience we now drop the suffix i and compute the joint cdf of two residuals from a sample of size n.

Note that

$$\begin{aligned} P\left(\hat{\epsilon}_j \leq x, \hat{\epsilon}_{j'} \leq y\right) &= P\left(\zeta_j \leq \frac{x}{(n-k)}, \zeta_{j'} \leq \frac{y}{(n-k)}\right) \\ &= F_2(\frac{x}{(n-k)}, \frac{y}{(n-k)}; \rho, \nu) \end{aligned}$$

where $F_2(u, v; \rho, \nu)$ is the joint cdf corresponding to $f(u, v; \rho, v)$. So, we only need to show how to compute the joint distribution function of two fitted residuals as defined in Ellenberg [29]. In the next section, we show how to compute this joint cdf of two residuals for the more general linear model $Y_i = X_i\beta + \sigma\epsilon_i$ discussed in Ellenberg [29]. We are able to do this for general ρ, not just ρ of the form $-\frac{1}{(n-1)}$. We specialize the answers to our case at the end.

8.2.5 The joint distribution function of two fitted residuals

We closely follow the work of Dunnet *et. al.* [28] to evaluate the the integral

$$\int_{-\infty}^{h}\int_{-\infty}^{k} f(u, v; \rho, \nu)dudv,$$

for given values of h and k. First we show the calculations for h and k positive, and extend the results for the negative values of h and k. So, assume h and k are positive unless specified otherwise.

Consider the new coordinate axes

$$\begin{aligned} Y &= v = r\sin\theta, \\ \text{and } X &= \frac{(u - \rho v)}{\sqrt{(1-\rho^2)}} = r\cos\theta. \end{aligned}$$

The Jacobian for this transformation is $r\sqrt{(1-\rho^2)}$. In the new coordinates the joint density can be written as, $g(r, \theta, \rho) = \frac{\nu}{\pi} r\left[1 - r^2\right]^{(\nu-1)}$. Since $\tan\theta = v\frac{\sqrt{(1-\rho^2)}}{(u-\rho v)}$, the lines $\theta =$ constant are straight lines through the origin. The line $U = 0$ can be written as $Y = -\frac{\sqrt{(1-\rho^2)}}{\rho}X$. So, depending on whether ρ is negative or positive the axis U makes an acute

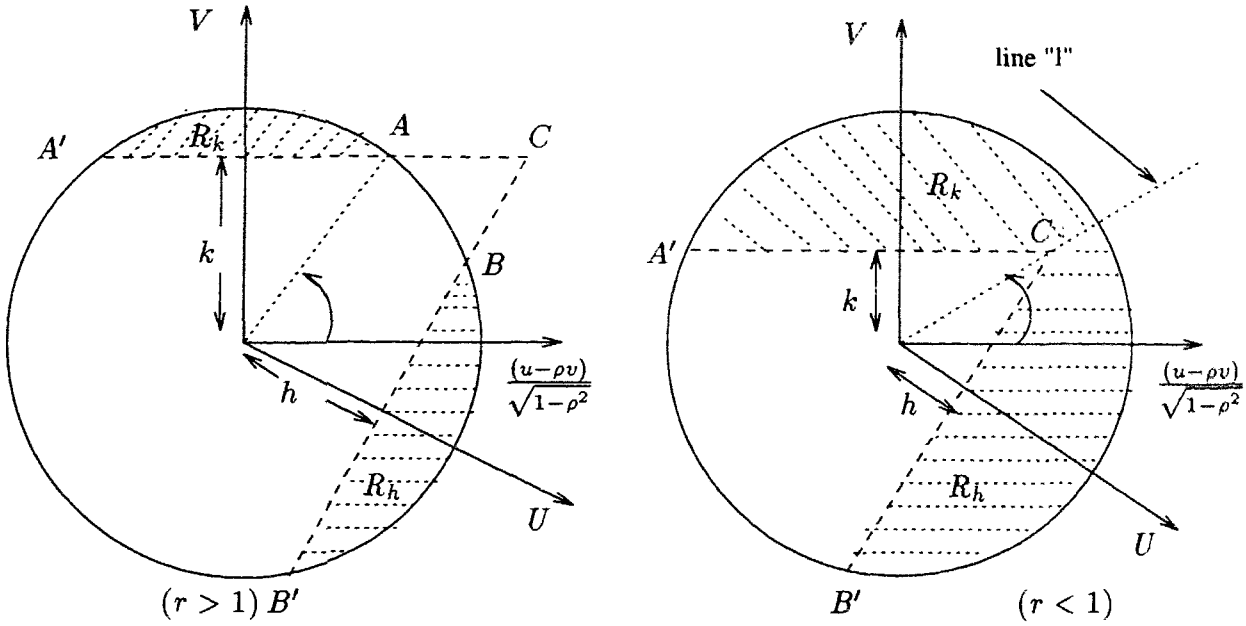

Figure 8.1: The probability integrals for $\rho < 0$

or obtuse angle with the Y axis. The direction of positive U can be identified by considering the coordinates of (h, k) (for ρ negative) and $(-h, k)$ (for ρ positive).

Define $r^2 = \left[\frac{(h-\rho k)}{\sqrt{(1-\rho^2)}}\right]^2 + k^2$. Note that r is the distance from the origin to the point of intersection of the two lines $v = k$ and $u = h$. Thus, the two lines intersect outside or inside the unit circle according as $r > 1$ or < 1. (See Figure 8.1.).

We first evaluate the double integral for the case $r \geq 1$. Then we suggest the suitable modifications and evaluate the integral for the case $r < 1$. Let

$$\begin{aligned} R_k &= \text{The region inside the unit circle outside the line } v = k, \\ R_h &= \text{The region inside the unit circle outside the line } u = h, \\ C &= \text{The point of intersection of } v = k \text{ and } u = h, \\ A, A' &= \text{The points where } v = k \text{ meets the unit circle,} \\ \text{and } B, B' &= \text{The points where } u = h \text{ meets the unit circle.} \end{aligned}$$

Suppose ρ is negative. Then $(h - \rho k)$ and $(k - \rho h)$ are both positive for all positive values of h and k (Figure 8.1), and

$$\int_{-\infty}^{h}\int_{-\infty}^{k} f(u, v; \rho, \nu)dudv = 1 - \int\int_{R_k} g(r, \theta, \rho)drd\theta - \int\int_{R_h} g(r, \theta, \rho)drd\theta.$$

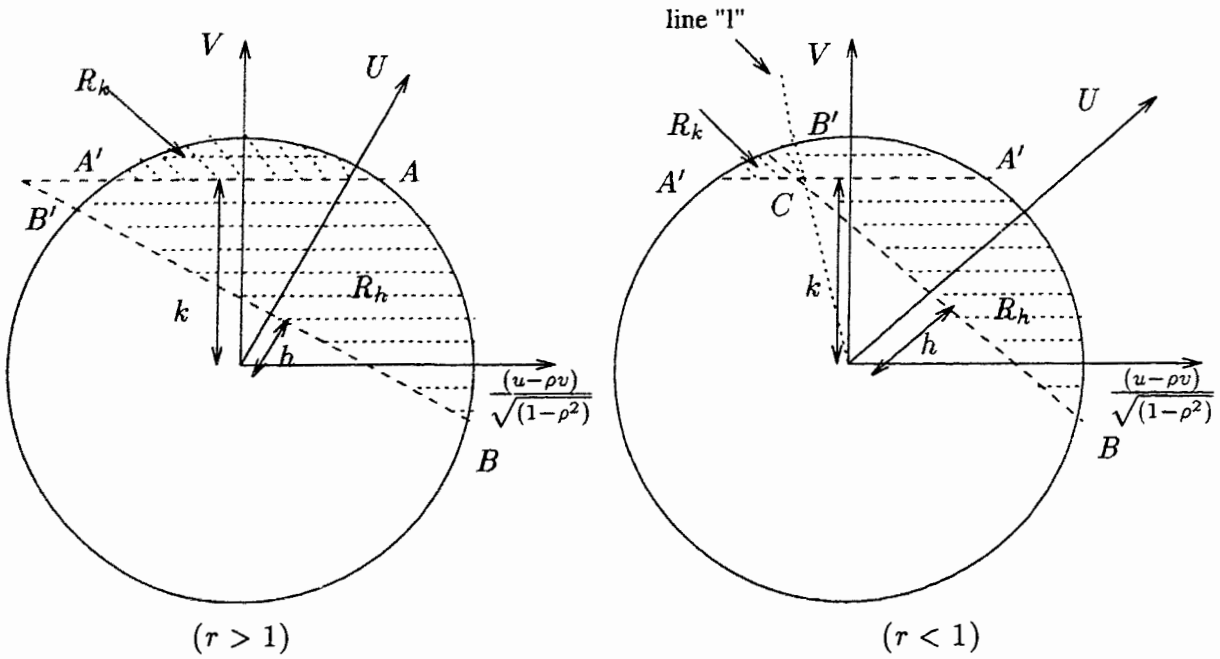

Figure 8.2: The probability integrals for $\rho > 0$

Now suppose ρ is positive. Then if $(h - \rho k) < 0$, then $(k - \rho h) > 0$ (since $\rho h < h < \rho k < k$). But if $(h - \rho k) > 0$, then $(k - \rho h)$ can be either negative or positive. We need to consider these three cases separately.

Case 1: $(h - \rho k) < 0$ (Thus $(k - \rho h) > 0$.) If $r \geq 1$, (Figure 8.2), then

$$\int_{-\infty}^{h}\int_{-\infty}^{k} f(u,v;\rho,\nu)dudv = 1 - \int\int_{R_h} g(r,\theta,\rho)drd\theta.$$

Case 2: $(h - \rho k) > 0$ and $(k - \rho h) < 0$. Then

$$\int_{-\infty}^{h}\int_{-\infty}^{k} f(u,v;\rho,\nu)dudv = 1 - \int\int_{R_k} g(r,\theta,\rho)drd\theta.$$

Case 3: $(h - \rho k) > 0$ and $(k - \rho h) > 0$. Then

$$\int_{-\infty}^{h}\int_{-\infty}^{k} f(u,v;\rho,\nu)dudv = 1 - \int\int_{R_k} g(r,\theta,\rho)drd\theta - \int\int_{R_h} g(r,\theta,\rho)drd\theta.$$

Limits for the integrals over the regions R_k and R_h:

Let θ be the angle measured from the axis perpendicular to the $Y(= V)$ axis at the origin. Let θ_A be the angle corresponding to A. We note that, if $r \geq 1$ (Figure 1), then $\theta_A =$

$\arctan\left(\frac{k}{\sqrt{(1-k^2)}}\right)$ and $\theta_{A'} = \pi - \arctan\left(\frac{k}{\sqrt{(1-k^2)}}\right)$, for $0 < k < 1$. For $k \geq 1$, the integral over R_k is 0, whereas for $k = 0$ it is 0.5.

Note that over the region R_k, θ varies from θ_A to $\theta_{A'}$. For fixed θ, r varies from $k \csc\theta$ to 1.

Now let θ be the angle measured from the axis perpendicular to the U axis at the origin. As before, let θ_B and $\theta_{B'}$ be the angles corresponding to the points B and B' respectively. If $r \geq 1$, then $\theta_B = \arctan\left(\frac{h}{\sqrt{(1-h^2)}}\right)$, and $\theta_{B'} = \pi - \arctan\left(\frac{h}{\sqrt{(1-h^2)}}\right)$, for $0 < h < 1$. Over the region R_h θ varies from θ_B to $\theta_{B'}$. For fixed θ, r varies from $h \csc\theta$ to 1.

Now we describe the modifications required when $r < 1$. When $r < 1$, we slightly modify the definitions of R_k and R_h as follows.

We take the line joining the origin and the point C (the point of intersection of $v = k$ and $u = h$) as a boundary line for the regions. For notational convenience, we refer to this line as line l (See Figure 8.1.) The region R_k is now defined as the region bounded by the curve $r = 1$, the line l, and the line $v = k$. Similarly, the region R_h is defined as the region bounded by the curve $r = 1$, the line l, and the line $u = h$. Accordingly the angles corresponding to the boundary lines for the regions R_k and R_h take the following values.

$$\begin{aligned}
\theta_A &= \arctan\left(\frac{k\sqrt{1-\rho^2}}{(h-\rho k)}\right) \\
\theta_{A'} &= \pi - \arctan\left(\frac{k}{\sqrt{(1-k^2)}}\right) \\
\theta_B &= \arctan\left(\frac{h\sqrt{1-\rho^2}}{(k-\rho h)}\right) \\
\theta_{B'} &= \pi - \arctan\left(\frac{h}{\sqrt{(1-h^2)}}\right).
\end{aligned}$$

When $r < 1$, regardless of the signs of $(h - \rho k)$ and $(k - \rho h)$,

$$\int_{-\infty}^{h}\int_{-\infty}^{k} f(u,v;\rho,\nu)dudv = 1 - \int\int_{R_k} g(r,\theta,\rho)drd\theta - \int\int_{R_h} g(r,\theta,\rho)drd\theta.$$

We notice that, regardless of whether or not the point of intersection is inside or outside

the unit circle,

$$\theta_{A'} = \pi - \arctan\left(\frac{k}{\sqrt{(1-k^2)}}\right),$$
$$\text{and } \theta_{B'} = \pi - \arctan\left(\frac{h}{\sqrt{(1-h^2)}}\right).$$

Considering the symmetry and taking

$$\theta_A = \begin{cases} \arctan\left(\frac{k}{\sqrt{(1-k^2)}}\right), & \text{if } (r > 1), \\ \arctan\left(\frac{k\sqrt{1-\rho^2}}{\sqrt{(h-\rho k)}}\right), & \text{if } (r < 1), \end{cases}$$

$$\theta_B = \begin{cases} \arctan\left(\frac{h}{\sqrt{(1-h^2)}}\right), & \text{if } (r > 1), \\ \arctan\left(\frac{h\sqrt{1-\rho^2}}{\sqrt{(k-\rho h)}}\right), & \text{if } (r < 1), \end{cases}$$

$$\theta_{A'} = \arctan\left(\frac{k}{\sqrt{(1-k^2)}}\right),$$
$$\text{and } \theta_{B'} = \arctan\left(\frac{h}{\sqrt{(1-h^2)}}\right),$$

the integrals over the regions R_k and R_h can be written as

$$\int\int_{R_k} g(r,\theta,\rho)drd\theta = \begin{cases} \int_{\theta_{A'}}^{\pi/2}\int_{k\csc\theta}^{1} g(r,\theta,\rho)drd\theta + \int_{\theta_A}^{\pi/2}\int_{k\csc\theta}^{1} g(r,\theta,\rho)drd\theta \\ \quad (\text{if } (h-\rho k) > 0), \\ \int_{\theta_{A'}}^{\pi/2}\int_{k\csc\theta}^{1} g(r,\theta,\rho)drd\theta - \int_{\theta_A}^{\pi/2}\int_{k\csc\theta}^{1} g(r,\theta,\rho)drd\theta \\ \quad (\text{if } (h-\rho k) < 0). \end{cases}$$

$$\text{and } \int\int_{R_h} g(r,\theta,\rho)drd\theta = \begin{cases} \int_{\theta_{B'}}^{\pi/2}\int_{h\csc\theta}^{1} g(r,\theta,\rho)drd\theta + \int_{\theta_B}^{\pi/2}\int_{h\csc\theta}^{1} g(r,\theta,\rho)drd\theta \\ \quad (\text{if } (k-\rho h) > 0) \\ \int_{\theta_{B'}}^{\pi/2}\int_{h\csc\theta}^{1} g(r,\theta,\rho)drd\theta - \int_{\theta_B}^{\pi/2}\int_{h\csc\theta}^{1} g(r,\theta,\rho)drd\theta \\ \quad (\text{if } (k-\rho h) < 0). \end{cases}$$

Using the notation

$$sgn(a) = \begin{cases} 1, & \text{if } (a > 0) \\ 0, & \text{if } (a = 0) \\ -1, & \text{if } (a < 0), \end{cases}$$

the above integrals can be written as

$$\int\int_{R_k} g(r,\theta,\rho)drd\theta = \int_{\theta_{A'}}^{\pi/2}\int_{k\csc\theta}^{1} g(r,\theta,\rho)drd\theta + sgn(h-\rho k)\int_{\theta_A}^{\pi/2}\int_{k\csc\theta}^{1} g(r,\theta,\rho)drd\theta, \tag{8.4}$$

and

$$\int\int_{R_h} g(r,\theta,\rho)drd\theta = \int_{\theta_{B'}}^{\pi/2}\int_{h\csc\theta}^{1} g(r,\theta,\rho)drd\theta + sgn(k-\rho h)\int_{\theta_B}^{\pi/2}\int_{h\csc\theta}^{1} g(r,\theta,\rho)drd\theta. \tag{8.5}$$

Evaluating the integrals:

We only need to evaluate the double integral over the region R_k, since by interchanging h and k we can obtain the value of the integral over the region R_h. For notational convenience let us denote the lower and upper limits for the variable θ over the region R_k by $C_1(h,k,\rho)$ and $C_2(h,k,\rho)$ respectively. We need to evaluate

$$\int_{C_2(h,k,\rho)}^{\pi/2}\int_{k\csc\theta}^{1} g(r,\theta,\rho)drd\theta + sgn(h-\rho k)\int_{C_1(h,k,\rho)}^{\pi/2}\int_{k\csc\theta}^{1} g(r,\theta,\rho)drd\theta,$$

where $g(r,\theta,\rho) = \frac{\nu}{\pi} r\left[1-r^2\right]^{(\nu-1)}$. Note that $\int_{k\csc\theta}^{1} g(r,\theta,\rho)dr = \frac{1}{2\pi}\left[1-k^2\csc^2\theta\right]^{\nu}$. Let

$$\begin{aligned} Q_\nu(h,k,\rho) &= \frac{1}{2\pi}\int_{C_2(h,k,\rho)}^{\pi/2}\left[1-k^2\csc^2\theta\right]^{\nu} d\theta \\ &\quad + sgn(h-\rho k)\frac{1}{2\pi}\int_{C_1(h,k,\rho)}^{\pi/2}\left[1-k^2\csc^2\theta\right]^{\nu} d\theta. \end{aligned}$$

Then,

$$\begin{aligned} Q_\nu(h,k,\rho) &= \frac{1}{2\pi}\int_{C_2(h,k,\rho)}^{\pi/2}\left[1-k^2\csc^2\theta\right]^{(\nu-1)}\left[1-k^2\csc^2\theta\right] d\theta \\ &\quad + sgn(h-\rho k)\frac{1}{2\pi}\int_{C_1(h,k,\rho)}^{\pi/2}\left[1-k^2\csc^2\theta\right]^{(\nu-1)}\left[1-k^2\csc^2\vartheta\right] d\theta \\ &= Q_{(\nu-1)}(h,k,\rho) + I_\nu(h,k,\rho), \quad \text{say.} \end{aligned} \tag{8.6}$$

To evaluate $I_\nu(h,k,\rho)$ note that

$$\begin{aligned} I_\nu(h,k,\rho) &= \frac{k^2}{2\pi}\int_{C_2(h,k,\rho)}^{\pi/2}\left[1-k^2\csc^2\theta\right]^{(\nu-1)} k^2\csc^2\theta d\theta \\ &\quad + sgn(h-\rho k)\frac{k^2}{2\pi}\int_{C_1(h,k,\rho)}^{\pi/2}\left[1-k^2\csc^2\theta\right]^{(\nu-1)} k^2\csc^2\theta d\theta \end{aligned}$$

$$= \frac{k^2}{2\pi}\left(1-k^2\right)^{(\nu-1)} \int_{C_2(h,k,\rho)} \left[1-\frac{k^2}{(1-k^2)}\cot^2\theta\right]^{(\nu-1)} d(\cot\theta)$$
$$+sgn(h-\rho k)\frac{k^2}{2\pi}\left(1-k^2\right)^{(\nu-1)} \int_{C_1(h,k,\rho)} \left[1-\frac{k^2}{(1-k^2)}\cot^2\theta\right]^{(\nu-1)} d(\cot\theta).$$

Substituting $z = \frac{k^2}{(1-k^2)}\cot^2\theta$, we can write the integral $I_\nu(h,k,\rho)$ as follows:

$$I_\nu(h,k,\rho) = \frac{k^2}{2\pi}\int_{z_2(h,k,\rho)}^{0} \left[1-k^2\csc^2\theta\right]^{(\nu-1)} k^2\csc^2\theta d\theta$$
$$+sgn(h-\rho k)\frac{k^2}{2\pi}\int_{z_1(h,k,\rho)}^{0} \left[1-k^2\csc^2\theta\right]^{(\nu-1)} k^2\csc^2\theta d\theta,$$

where

$$z_1(h,k,\rho) = \begin{cases} 1, & \text{if } (r \geq 1) \\ \frac{(h-\rho k)^2}{(1-k^2)(1-\rho^2)}, & \text{if } (r<1), \end{cases}$$
$$\text{and } z_2(h,k,\rho) = 1.$$

Let $I_x(p,q) = \int_0^x \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1}dy$ be the incomplete beta function. Then $I_\nu(h,k,\rho)$ can be written as

$$I_\nu(h,k,\rho) = -\frac{k}{4\sqrt{\pi}}\left(1-k^2\right)^{(\nu-1/2)} \frac{\Gamma(\nu)}{\Gamma(\nu+1/2)}\left\{1+sgn(h-\rho k)I_z(h,k,\rho)\right\}, \quad (8.7)$$
$$\text{where } z = \begin{cases} 1 & \text{if } (r>1) \\ \frac{(h-\rho k)^2}{(1-k^2)(1-\rho^2)} & \text{if } (r<1). \end{cases}$$

Combining Equations 8.6 and 8.7 we get

$$Q_\nu(h,k,\rho) = Q_{(\nu-1)}(h,k,\rho) - \frac{k}{4\sqrt{\pi}}\left(1-k^2\right)^{(\nu-1/2)} \frac{\Gamma(\nu)}{\Gamma(\nu+1/2)}\left\{1+sgn(h-\rho k)I_z(h,k,\rho)\right\}. \quad (8.8)$$

Using the recurrence formula (Equation 8.8), we can obtain the following expression for $Q_\nu(h,k,\rho)$ $(\nu = (n-3)/2)$:

$$Q_\nu(h,k,\rho) = \begin{cases} Q_0 - \frac{k}{4\sqrt{\pi}}\sum_{j=1}^{\nu}\left(1-k^2\right)^{(j-\frac{1}{2})}\frac{\Gamma(j)}{\Gamma(j+\frac{1}{2})}\left\{1+sgn(h-\rho k)I_z\left[\frac{1}{2},j\right]\right\} \\ \quad \text{(when } n \text{ is odd)} \\ Q_{\frac{1}{2}} - \frac{k}{4\sqrt{\pi}}\sum_{j=1}^{(\nu-\frac{1}{2})}\left(1-k^2\right)^{j}\frac{\Gamma(j+\frac{1}{2})}{\Gamma(j+1)}\left\{1+sgn(h-\rho k)I_z\left[\frac{1}{2},j+\frac{1}{2}\right]\right\} \\ \quad \text{(when } n \text{ is even).} \end{cases} \quad (8.9)$$

The formulae given in Dunnet *et. al.* [28] can be used to evaluate the incomplete beta functions. These are

$$I_x\left[\frac{1}{2}, j+\frac{1}{2}\right] = \frac{2}{\pi}\arctan\sqrt{\left(\frac{x}{1-x}\right)} + \frac{2}{\pi}\sqrt{x(1-x)}\sum_{i=0}^{j-1}\frac{4^i(i!)^2}{(2i+1)!}(1-x)^i,$$

$$\text{and } I_x\left[\frac{1}{2}, j\right] = \sqrt{x}\sum_{i=0}^{(j-1)}\frac{(2i)!}{4^i(i!)^2}(1-x)^i.$$

Computing Q_0 and $Q_{1/2}$:

Recall that

$$Q_\nu(h,k,\rho) = \frac{1}{2\pi}\int_{C_2(h,k,\rho)}^{\pi/2}\left[1-k^2\csc^2\theta\right]^\nu d\theta + sgn(h-\rho k)\frac{1}{2\pi}\int_{C_1(h,k,\rho)}^{\pi/2}\left[1-k^2\csc^2\theta\right]^\nu d\theta.$$

Therefore,

$$Q_0(h,k,\rho) = \frac{1}{2\pi}\left[\pi/2 - C_2(h,k,\rho)\right] + sgn(h-\rho k)\frac{1}{2\pi}\left[\pi/2 - C_1(h,k,\rho)\right],$$

where

$$C_2(h,k,\rho) = \arctan\left(\frac{k}{\sqrt{1-k^2}}\right)$$

and

$$C_1(h,k,\rho) = \begin{cases} \arctan\left(\frac{k}{\sqrt{1-k^2}}\right) & \text{for } (r>1) \\ \arctan\left(\frac{k\sqrt{1-\rho^2}}{(h-\rho k)}\right) & \text{for } (r<1). \end{cases}$$

To compute $Q_{1/2}(h,k,\rho)$ we first compute $\int\left[1-k^2\csc^2\theta\right]^{1/2}d\theta$. We write

$$\int\left[1-k^2\csc^2\theta\right]^{1/2}d\theta = \int\left[1-k^2\csc^2\theta\right]^{-1/2}d\theta - k^2\int\frac{\csc^2\theta}{\sqrt{1-k^2\csc^2\theta}}d\theta.$$

Observe that $\int\left[1-k^2\csc^2\theta\right]^{-1/2}d\theta$

$$= \int\frac{\sin\theta}{\sqrt{\sin^2\theta-k^2}}d\theta$$

$$= \int\frac{1}{\sqrt{(1-k^2)-z^2}}dz, \text{where } z=\sqrt{\sin^2\theta-k^2}.$$

$$= \arcsin\left(\frac{z}{\sqrt{1-k^2}}\right)$$

$$= \arcsin\left(\frac{\sqrt{\sin^2\theta-k^2}}{\sqrt{1-k^2}}\right),$$

and

$$
\begin{aligned}
-k^2 \int \frac{\csc^2\theta}{\sqrt{1-k^2\csc^2\theta}} &= \int \frac{d(\cot\theta)}{\sqrt{1-k^2(1+\cot^2\theta)}} \\
&= k\int \frac{du}{\sqrt{\left(\frac{1-k^2}{k^2}\right)-u^2}}, \quad \text{for } u=\cot\theta. \\
&= k\arcsin\left(\frac{\cot\theta}{\sqrt{\frac{(1-k^2)}{k^2}}}\right).
\end{aligned}
$$

For notational convenience let $\alpha = \arcsin\left(\frac{\sqrt{\sin^2\theta-k^2}}{\sqrt{1-k^2}}\right)$ and $\beta = \arcsin\left(\frac{\cot\theta}{\sqrt{\frac{(1-k^2)}{k^2}}}\right)$. We tabulate some specific values of θ and the corresponding values of α and β which we need in the sequel.

θ	α	β
$\arctan\left(\frac{k}{\sqrt{1-k^2}}\right)$	0	$\frac{\pi}{2}$
$\frac{\pi}{2}$	$\frac{\pi}{2}$	0
$\arctan\left(\frac{k\sqrt{1-\rho^2}}{(h-\rho k)}\right)$	$\arctan\left(\frac{k\sqrt{(1-\rho^2)-[h^2+k^2-2\rho hk]}}{(h-\rho k)}\right)$	$\arctan\left(\frac{(h-\rho k)}{\sqrt{(1-\rho^2)-[h^2+k^2-2\rho hk]}}\right)$

Substituting corresponding limits (see the table), we find

$$
Q_{1/2}(h,k,\rho) = \frac{1}{2\pi}\left\{\frac{\pi}{2} + sgn(h-\rho k)\left(\frac{\pi}{2}-\theta_1\right) - k\left[\frac{\pi}{2} + sgn(h-\rho k)\theta_2\right]\right\},
$$

where

$$
\theta_1 = \begin{cases} 0, & \text{if } (r>1) \\ \arctan\left(\frac{k\sqrt{(1-\rho^2)-[h^2+k^2-2\rho hk]}}{(h-\rho k)}\right) & \text{if } (r<1), \end{cases}
$$

and

$$
\theta_2 = \begin{cases} \frac{\pi}{2} & \text{if } (r>1) \\ \arctan\left(\frac{(h-\rho k)}{\sqrt{(1-\rho^2)-[h^2+k^2-2\rho hk]}}\right) & \text{if } (r<1). \end{cases}
$$

We note that, according to the above formula, when $r>1$ and $(h-\rho k)<0$, the integral over the region R_k is zero. Similarly if $r>1$ and $(k-\rho h)<0$, then the integral over the region R_h is zero. Therefore, regardless of the signs of $(h-\rho k)$ and $(k-\rho h)$ we can

use the formula, $\int_{-\infty}^{h}\int_{-\infty}^{k} f(u,v;\rho,\nu)dudv = 1 - \int\int_{R_k} g(r,\theta,\rho)drd\theta - \int\int_{R_h} g(r,\theta,\rho)drd\theta$ to compute the joint distribution function of u and v at (h,k) for all positive values of h and k.

Computing the distribution function for negative values of h and k:

Recall that $F_2(u,v;\rho,\nu) = \int_{-\infty}^{h}\int_{-\infty}^{k} f(u,v;\rho,\nu)dudv$.

Case 1: $h \geq 0$ and $k < 0$

Let $k_1 = -k$ then $k_1 > 0$. Observe that

$$\begin{aligned} P(U \leq h, V \leq k) &= P(U \leq h, V \leq -k_1) \\ &= P(U \leq h) - P(U \leq h, V > -k_1) \\ &= P(U \leq h, V \leq 1) - P(U \leq h, -V < k_1) \\ &\quad \text{(Since the density is zero when } V > 1.) \\ &= F_2(h,1;\rho,\nu) - F_2(h,k_1,-\rho,\nu). \end{aligned}$$

Case 2: $h < 0$ and $k \geq 0$

Letting $h_1 = -h$, as in Case 1, we find

$$P(U \leq h, V \leq k) = F_2(1,k;\rho,\nu) - F_2(h_1,k;-\rho,\nu).$$

Case 3: $h < 0$ and $k < 0$

Let $h_1 = -h$ and $k_1 = -k$ and write

$$\begin{aligned} P(U \leq h, V \leq k) &= 1 - [P(V \geq k) + P(U \geq h) - P(U \geq h, V \geq k)] \\ &= 1 - P(-V \leq k_1) - P(-U \leq h_1) + P(-U \leq h_1, -V \leq k_1) \\ &= 1 - P(-V \leq k_1, U \leq 1) - P(-u \leq h_1, V \leq 1) \\ &\quad + P(-U \leq h_1, -V \leq k_1) \\ &= 1 - F_2(1,k_1;\rho,\nu) - F_2(h_1,1;-\rho,\nu) + F_2(h_1,k_1;\rho,\nu). \end{aligned}$$

Thus, we can use the formulae already derived for positive h and k to evaluate the cumulative distribution function even if either h or k is negative.

Evaluating the integral when $h = k = 0$.

Now we describe computing the probability integral when $h = k = 0$. We show that the probability integral for this case is identical to the result for the bivariate normal integral. We note that the results for $h = k = 0$ can also be derived from the result we derived earlier as a limiting case for $h \to 0$ and $k \to 0$.

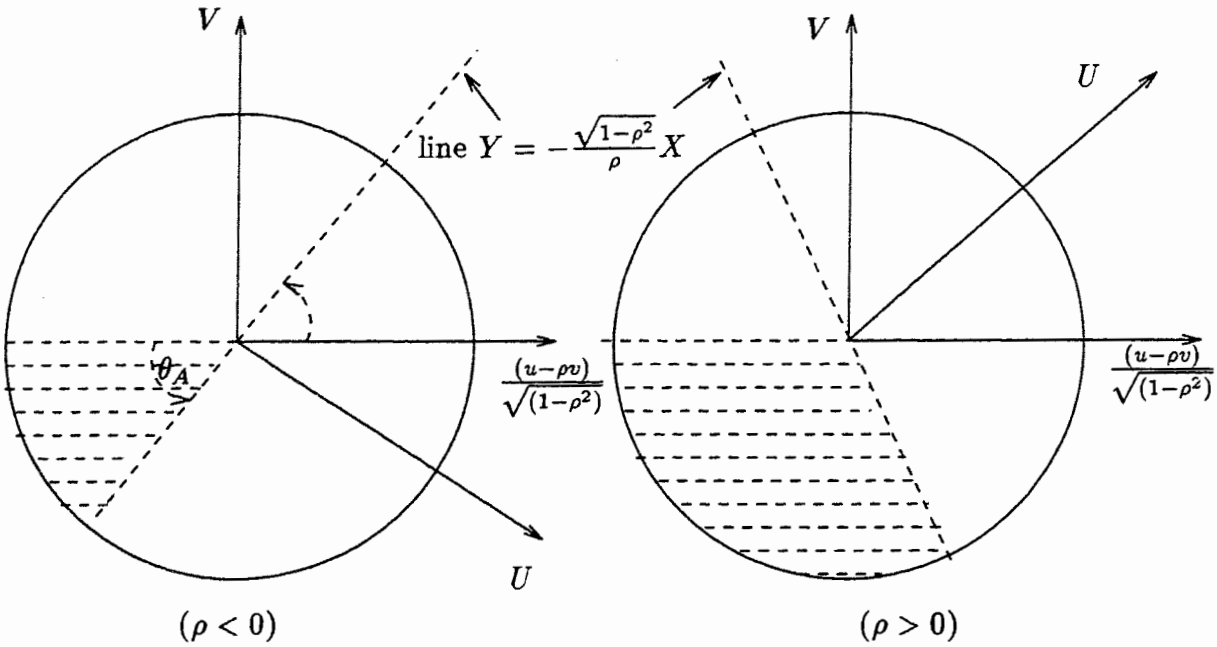

Figure 8.3: The probability integrals for $h = k = 0$

When $h = k = 0$, the required integral $\int_{-\infty}^{h}\int_{-\infty}^{k} f(u,v;\rho,\nu)dudv$ can be computed as follows (See Figure 8.3):

$$\begin{aligned}
\int_{-\infty}^{h}\int_{-\infty}^{k} f(u,v;\rho,\nu)dudv &= \int_0^{\theta_A}\int_0^1 \frac{\nu}{\pi}\left[1-r^2\right]^{(\nu-1)} r dr d\theta \\
&= \frac{1}{2\pi}\int_0^{\theta_A}\left[1-r^2\right]^{\nu} d\theta \\
&= \frac{1}{2\pi}\int_0^{\theta_A} d\theta = \frac{1}{2\pi}\theta_A \\
&= \frac{1}{2\pi}\arctan\left(-\frac{\sqrt{1-\rho^2}}{\rho}\right).
\end{aligned}$$

For ease of reference we now summarize the results derived in this section.

To evaluate the cumulative distribution function, $\int_{-\infty}^{h}\int_{-\infty}^{k} f(u,v;\rho,\nu)dudv$, for positive

values of h and k we proposed the recurrence relation:

$$\int_{-\infty}^{h}\int_{-\infty}^{k} f(u,v;\rho,\nu)dudv = 1 - \int\int_{R_k} g(r,\theta,\rho)drd\theta - \int\int_{R_h} g(r,\theta,\rho)drd\theta,$$

where

$$\int\int_{R_k} g(r,\theta,\rho)drd\theta \;=\; \begin{cases} Q_0 - \frac{k}{4\sqrt{\pi}}\sum_{j=1}^{\nu}(1-k^2)^{(j-\frac{1}{2})}\frac{\Gamma(j)}{\Gamma(j+\frac{1}{2})}\left\{1+sgn(h-\rho k)I_z\left[\frac{1}{2},j\right]\right\} \\ \quad \text{(when } n \text{ is odd)} \\ Q_{\frac{1}{2}} - \frac{k}{4\sqrt{\pi}}\sum_{j=1}^{(\nu-\frac{1}{2})}(1-k^2)^{j}\frac{\Gamma(j+\frac{1}{2})}{\Gamma(j+1)}\left\{1+sgn(h-\rho k)I_z\left[\frac{1}{2},j+\frac{1}{2}\right]\right\} \\ \quad \text{(when } n \text{ is even),} \end{cases}$$

The term Q_0 was found to be

$$Q_0(h,k,\rho) = \frac{1}{2\pi}\left[\pi/2 - C_2(h,k,\rho)\right] + sgn(h-\rho k)\frac{1}{2\pi}\left[\pi/2 - C_1(h,k,\rho)\right],$$

where

$$C_2(h,k,\rho) = \arctan\left(\frac{k}{\sqrt{1-k^2}}\right),$$

and

$$C_1(h,k,\rho) = \begin{cases} \arctan\left(\frac{k}{\sqrt{1-k^2}}\right) & \text{for } (r>1) \\ \arctan\left(\frac{k\sqrt{1-\rho^2}}{(h-\rho k)}\right) & \text{for } (r<1). \end{cases}$$

The term $Q_{1/2}$ was found to be

$$Q_{1/2}(h,k,\rho) = \frac{1}{2\pi}\left\{\frac{\pi}{2} + sgn(h-\rho k)\left(\frac{\pi}{2}-\theta_1\right) - k\left[\frac{\pi}{2} + sgn(h-\rho k)\theta_2\right]\right\}$$

where

$$\theta_1 = \begin{cases} 0, & \text{if } (r>1) \\ \arctan\left(\frac{k\sqrt{(1-\rho^2)-[h^2+k^2-2\rho hk]}}{(h-\rho k)}\right), & \text{if } (r<1) \end{cases}$$

and

$$\theta_2 = \begin{cases} \frac{\pi}{2} & \text{if } (r>1) \\ \arctan\left(\frac{(h-\rho k)}{\sqrt{(1-\rho^2)-[h^2+k^2-2\rho hk]}}\right) & \text{if } (r<1). \end{cases}$$

The double integral over the region R_h can be obtained from that of R_k by interchanging h and k.

For non positive values of h and k we can use the results derived for positive values of h and k as

$$\int_{-\infty}^{h}\int_{-\infty}^{k} f(u,v;\rho,\nu)dudv = \begin{cases} G(h,1,\rho) - G(h,k_1,-\rho), \text{ when } h \geq 0 \text{ and } k < 0 \\ G(1,k,\rho) - G(h_1,k,-\rho), \text{ when } h < 0 \text{ and } k \geq 0 \\ 1 - G(1,k_1,\rho) - G(h_1,1,-\rho) + G(h_1,k_1,\rho), \\ \text{when } h < 0 \text{ and } k < 0 \\ \frac{1}{2\pi}\arctan\left(-\frac{\sqrt{1-\rho^2}}{\rho}\right), \text{ when } h = 0 \text{ and } k = 0. \end{cases}$$

8.2.6 Justification for using the approximate p-value

We provide theoretical justification for using the suggested test procedure by proving weak convergence of the related empirical process. Validity of the suggested asymptotic results in finite samples is justified by a simulation study.

Weak convergence of the empirical process

Let k denotes the number of dose levels and n_i denotes the number of replicates at the ith dose level. Let $N = \sum_{i=1}^{k} n_i$ be the total number of observations. We prove the weak convergence of the process

$$W_N(t) = \frac{1}{\sqrt{N}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\{I\left(G_{n_i}(\hat{\epsilon_{ij}}) \leq t\right) - t\}$$

for the case of equal number of replicates at each dose level (i.e. $n_i = n$ for all i). We fix n and let $k \longrightarrow \infty$. In this case the process

$$W_N(t) = \frac{1}{\sqrt{nk}}\sum_{i=1}^{k}\sum_{j=1}^{n}\{I\left(G_n(\hat{\epsilon_{ij}}) \leq t\right) - t\}$$

can be rewritten as

$$W_N = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} W_{Nj}(t),$$

where each $W_{Nj}(t) = \frac{1}{\sqrt{k}}\sum_{i=1}^{k}\left[I\left(u_{ij} \leq t\right) - t\right]$ and $u_{ij} = G_n(\hat{\epsilon}_{ij})$. For each fixed j, the variables $u_{1j},\ldots,u_{kj}$ are, under the null hypothesis, iid uniform$[0,1]$ variables and so each

W_{Nj} converges weakly in $D[0,1]$ to a Brownian Bridge, that is, a Gaussian process W_j with mean 0 and covariance $\min(s,t) - st$. Therefore, W_{Nj} is tight in $D[0,1]$. This in turn implies that for each j, there is a compact $K_j \subset D[0,1]$ such that $P(W_{Nj} \in K_j) \geq 1 - \frac{\epsilon}{n}$ for any $\epsilon > 0$. Since K_j is compact in $D[0,1]$, $K = K_1 \times \ldots \times K_n$ is compact in $(D[0,1])^n$ and $P((W_{N1}, \ldots, W_{Nn}) \epsilon K) \geq 1 - \epsilon$. Since ϵ is arbitrary it follows that the process $(W_{N1}, \ldots, W_{Nn})$ is tight in $(D[0,1])^n$.

Now consider $0 \leq t_1 < \ldots < t_r \leq 1$. The matrix M_k whose ljth entry is $W_{Nj}(t_l)$ can be written as $\sum_{i=1}^{k} \frac{1}{\sqrt{k}} Q_i$ where the matrices Q_i's are iid and Q_i has ljth entry $I(u_{ij} \leq t_l) - t_l$. Each Q_i has mean 0 and so M_k converges in distribution by the usual Central Limit Theorem, to a Gaussian matrix M with $E(M) = 0$ and

$$\begin{aligned} Cov(M_{lj}, M_{l'j'}) &= Cov(I(u_{ij} \leq t_l), I(u_{ij'} \leq t_{l'})) \\ &= G_2\left(G^{-1}(t_l), G^{-1}(t_{l'}), \rho_{jj'}\right) - t_l t_{l'} \quad \text{(See Section 8.2.3.)}, \end{aligned}$$

where $\rho_{jj'} = 1$ if $j = j'$ and $-\frac{1}{(n-1)}$ if $j \neq j'$.

Thus, $(W_{N1}, \ldots, W_{Nn})$ converges weakly in $(D[0,1])^n$ to a Gaussian process $(W_1, \ldots, W_n)$ with mean 0 and $Cov(W_j(t_l), W_{j'}(t_{l'})) = G_2(G^{-1}(t_l), G^{-1}(t_{l'}), \rho_{jj'}) - t_l t_{l'}$.

Since each W_j is in $C[0,1]$ (each is a Brownian Bridge) it follows that $W_N = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} W_{Nj}$ converges weakly in $D[0,1]$ to $W = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} W_j$ which is a mean 0 Gaussian process with covariance

$$\begin{aligned} \rho(s,t) &= Cov(W(s), W(t)) \\ &= \frac{1}{n} \sum_{j=1}^{n} \sum_{j'=1}^{n} Cov(W_j(s), W_{j'}(t)) \\ &= \min(s,t) - st + \frac{1}{n} \sum_{j \neq j'} \left\{ G_2\left(G^{-1}(s), G^{-1}(t), -\frac{1}{(n-1)}\right) - st \right\}. \end{aligned}$$

Simulation results

Now we describe a Monte Carlo study that we performed to examine the approximate p-value suggested in Section 8.2.2 for testing the normality of the random errors ϵ_{ij} in the model $Y_{ij} = \mu_i + \sigma_i \epsilon_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i$.

In this study we used equal numbers of replicates at each dose level (i.e. $n_i = n, \ \forall i = 1, \ldots, k.$). The chosen numbers of replicates, n, are given in Table 8.1. The mean μ_i and the variance σ_i^2 at each dose level were set at 0 and 1 respectively. In other words the Y_{ij}'s were chosen to be standard normal random variates. We generated 1000 such samples. For each simulated sample, we computed a p-value based on the Cramér-von Mises statistic for testing normality as described in section in 8.2.2. Thus, we have 1000 p-values. If the suggested asymptotic theory holds, these p-values should be approximately uniformly distributed. The validity of the asymptotic theory was tested by checking to see if these p-values are approximately uniformly distributed. We used the Anderson-Darling statistic as a measure of uniformity of p-values. The null hypothesis to be tested is H_0 : the p-values are uniformly distributed on [0,1]. Thus, under H_0 the distribution is completely specified. According to Stephens(1986) it falls under Case 0. We arrange the p-values in ascending order and we note that the probability integral transforms z_i's are exactly the values itself because the distribution of interest is the uniform distribution. The value of the test statistic is computed using the formula for A^2 given in Section 8.1.1. The p-values for the Anderson-Darling test are computed by first estimating the eigenvalues of the covariance kernel for A^2,

$$\rho_A(s,t) = \frac{\min(s,t) - st}{\sqrt{t(1-t)s(1-s)}},$$

and then using these eigenvalues as weights to compute the p-values as described in Section 8.1.2. The results using Cramér-von Mises statistic and Anderson-Darling statistic are given in Tables 8.1 and 8.2 respectively. We also compute the observed levels of nominal 1, 5, and 10% level tests, that is, the fraction of p-values smaller than these nominal levels.

Conclusions

Based on the simulation results, we draw the following conclusions.

1. The Anderson-Darling test does not reject the null hypothesis that the p-values are uniformly distributed in any of the above cases. This justifies the use of the approximate p-values suggested in Section 8.1.2 for testing the normality of random errors without assuming that fitted model is correct.

Data Set	n	k	Anderson test statistic	p-value	Observed level for		
					$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
1	6	30	2.15	0.08	0.088	0.046	0.008
2	5	30	0.42	0.82	0.093	0.057	0.007
3	4	30	0.59	0.65	0.100	0.056	0.011
4	4	10	0.48	0.76	0.108	0.050	0.009
5	3	10	1.60	0.15	0.107	0.049	0.007
6	3	5	1.12	0.30	0.098	0.049	0.008

Table 8.1: Results for testing normality using W^2: Model 1

Data Set	n	k	Anderson test statistic	p-value	Observed level for		
					$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
1	6	30	0.76	0.51	0.010	0.046	0.086
2	5	30	1.77	0.13	0.009	0.056	0.098
3	4	30	1.60	0.15	0.012	0.061	0.104
4	4	10	1.51	0.18	0.011	0.051	0.111
5	3	10	2.24	0.07	0.008	0.055	0.106
6	3	5	1.07	0.32	0.012	0.062	0.100

Table 8.2: Results for testing normality using A^2: Model 1

2. The observed levels from the simulation study agree well with the nominal levels.

3. The suggested asymptotic theory appears to hold for sample sizes as small as 15. The typical sample sizes for thermoluminescence studies are around 30. Therefore, the suggested procedure could safely be used for testing the normality assumption in the model described for thermoluminescence data.

8.2.7 Sensitivity of the tests to departures from normality

In the previous section, we described a simulation study that examines the validity of the approximate $p-$value for testing normality of random errors without assuming that the fitted model is correct. Here we describe a small study that investigates the sensitivity of the suggested tests in detecting departures from normality.

The simulation study is similar to that we described in Section 8.2.6 except that, instead of normal random variates we generated data from different alternative distributions. When generating data from the gamma distribution we fixed the shape parameter at values given in Table 8.6. Since we assume that Y has mean f and variance $\sigma^2 f^2$ we find that σ^2 is the reciprocal of the shape parameter. The scale parameter for the corresponding gamma distribution is $\sigma^2 f^2$. So we only report the shape parameter of the gamma distribution we used to generate the data Y. For each generated sample, we computed an approximate p-value for testing the hypothesis that the random errors are normally distributed. We examined the performance of both the Anderson-Darling statistic and the Cramér-von Mises statistic as test statistics for testing the normality. The powers of the tests were computed as the proportion of samples rejected by each test. The chosen distributions, sample sizes, and the results are given in Table 8.3.

Conclusions:

Based on the simulation results (Table 8.3), we draw the following conclusions.

1. For the alternative distributions considered in the study, the tests appear to detect the departure from normality with reasonably large sample sizes. Since we are fitting different mean and variance parameters at each dose level, the number of fitted parameters is quite large. Therefore, the effective sample size is much smaller than the

Data Set	Distribution	n	k	The power of			
				Cramér-von Mises		Anderson-Darling	
				$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
1	gamma(0.5)	3	10	0.50	0.61	0.56	0.66
2	gamma(2)	3	10	0.14	0.23	0.15	0.24
3	gamma(4)	3	10	0.10	0.16	0.10	0.16
4	gamma(6)	3	10	0.08	0.13	0.08	0.14
5	t_2	3	10	0.06	0.12	0.07	0.13
6	t_5	3	10	0.06	0.10	0.06	0.11
7	t_5	5	10	0.07	0.13	0.06	0.12

Table 8.3: Summary of test results for departures from normality: Model 1

nominal sample sizes used.

2. When the sample sizes are small, the Anderson-Darling test appears to perform better than the Cramér-von Mises test.

3. Since the computational burden is very much the same for both statistics we recommend using the Anderson-Darling statistic — particularly so if the sample sizes are small.

4. Based on the study it appears that the proposed EDF test is more sensitive to skewness than to heavy tails.

8.3 Application of EDF tests: Model 2

Consider the model, $Y_{ij} = \mu_i(\theta) + \sigma\mu_i(\theta)\epsilon_{ij}$, $i = 1, \ldots, k$, and $j = 1, \ldots, n_i$, where the random errors ϵ_{ij}'s have zero mean and unit variance. Here we assume that both mean and variance of Y_{ij}'s are functions of the unknown parameter[1] θ, and that the variance of the

[1] In our problem θ is a vector of parameters.

Y_{ij}'s are proportional to the square of the mean function. We refer to this model as Model 2. We wish to test the assumption of normality of the random errors ϵ_{ij}.

8.3.1 Computing the test statistic

In this model, the number of fitted parameters does not increase with the number of observations as in Model 1. Therefore, we can estimate the mean and variance consistently by choosing a large enough sample. Let $\hat{\theta}$ and $\hat{\sigma}$ be the maximum likelihood estimates for θ and σ respectively. The maximum likelihood estimate for $\mu_i(\theta)$ is $\mu_i(\hat{\theta})$. Let $N = \sum_{i=1}^{k} n_i$ be the total number of observations. We define the standardized fitted residuals as $\hat{\epsilon}_{ij} = \frac{(Y_{ij} - \mu_i(\hat{\theta}))}{\hat{\sigma}\mu_i(\hat{\theta})}$. The $\hat{\epsilon}_{ij}$ are approximately normally distributed, for large enough N. The probability integral transforms of $\hat{\epsilon}_{ij}$'s are therefore given by $u_{ij} = \Phi(\hat{\epsilon}_{ij})$, where $\Phi(\cdot)$ denotes the distribution function of the standard normal distribution. Let $z_1, \ldots, z_N$ be the ordered probability integral transforms, or the u_{ij}'s arranged in ascending order. Formulae (Equation 8.1) provided in Section 8.1.1 can be used to compute the test statistic for these z values.

8.3.2 An approximate p-value

The process $W_N(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \{I[u_{ij} \leq t] - t\}$ can be rewritten as

$$\begin{aligned} W_N(t) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \{I[\Phi(\hat{\epsilon}_{ij}) \leq t] - t\} \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left\{ I\left[\hat{\epsilon}_{ij} \leq \Phi^{-1}(t)\right] - t \right\} \end{aligned}$$

In Section 8.3.3, we show heuristically that the approximate covariance kernel for the Cramér-von Mises statistic is

$$\rho(s,t) = Cov(W_N(s), W_N(t)) = \min(s,t) - st - \frac{1}{N^2} \tau^T(s) I^{-1}(\xi) \tau(t), \qquad (8.10)$$

where

$$\xi^T = (\theta^T, \sigma), \; \tau(s) = \begin{bmatrix} \phi\left[\Phi^{-1}(s)\right] \left[\Phi^{-1}(s) + \frac{1}{\sigma}\right] \sum_{i=1}^{k} \sum_{j=1}^{n_i} \omega_i(\theta) \\ N\phi\left[\Phi^{-1}(s)\right] \left\{\frac{1}{\sigma}\Phi^{-1}(s)\right\} \end{bmatrix},$$

and $I(\xi)$ denotes the average Fisher information per observation. The covariance kernel for the Anderson-Darling statistic is

$$\rho_A(s,t) = \frac{\rho(s,t)}{\sqrt{s(1-s)t(1-t)}}. \tag{8.11}$$

Next we summarize the test procedure.

1. Find the maximum likelihood estimates $\hat{\theta}$ and $\hat{\sigma}$ for the parameters θ and σ. The maximum likelihood estimate for $\mu_i(\theta)$ is then $\mu_i(\hat{\theta})$.

2. Compute the standardized fitted residuals $\hat{\epsilon}_{ij} = \frac{(Y_{ij}-\mu_i(\hat{\theta}))}{\hat{\sigma}\mu_i(\hat{\theta})}$. (As we described earlier, we test the assumption of normality of random errors by testing whether the fitted residuals are approximately normally distributed.)

3. Compute the probability integral transforms $\Phi(\hat{\epsilon}_{ij})$.

4. Order the probability integral transforms in ascending order. Let $z_1, \ldots, z_N$ be the ordered probability integral transforms.

5. Compute the Cramér-von Mises statistic

$$W_N^2 = \sum_{i=1}^{N} \left\{ z_i - \frac{(2i-1)}{2N} \right\}^2 + \frac{1}{12N},$$

   or the Anderson-Darling statistic

$$A_N^2 = -N - \frac{1}{n} \sum_{i=1}^{N} \{(2i-1)\ln z_i + (2N+1-2i)\ln(1-z_i)\}.$$

6. Evaluate the covariance kernel $\rho(s,t)$ given by Equation 8.10 at the estimated parameter values $\hat{\theta}$ and $\hat{\sigma}$. (If the Anderson-Darling statistic is used $\rho(s,t)$ has to be replaced by $\rho_A(s,t)$.)

7. Find the eigenvalues of $\rho(s,t)$. We computed approximations for the eigenvalues as follows:

   (a) Create the matrix Q whose elements are

$$Q(s,t) = \frac{1}{m}\rho(s,t), \quad \text{for} \quad s,t = \frac{1}{(m+1)}, \ldots, \frac{m}{(m+1)}.$$

   (Here m is the number of subdivisions of the unit interval. We chose $m = 150$.)

(b) Compute the eigenvalues $\lambda_1, \ldots, \lambda_m$ of the matrix Q. These eigenvalues provide estimates for the eigenvalues of $\rho(s,t)$. (See Section 8.1.1.)

8. Compute the p-value based on the asymptotic distribution of the test statistic as described in Section 8.1.2.

9. Reject (or do not reject) the null hypothesis if the p-value is less than (or greater than) the desired significance level α.

8.3.3 Covariance kernel $\rho(s,t)$

In this section we use standard expansions to derive an asymptotic Gaussian approximation to the process $W_N(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left\{ I\left[\hat{\epsilon}_{ij} \leq \Phi^{-1}(t)\right] - t \right\}$. Our derivation is heuristic rather than rigorous though we believe that results of Loynes [45] can be used to provide rigorous justification. Note that

$$\begin{aligned} \hat{\epsilon}_{ij} \leq \Phi^{-1}(t) &\Leftrightarrow \sigma\mu_i(\theta)\epsilon_{ij} + \left(\mu_i(\theta) - \mu_i(\hat{\theta})\right) \leq \hat{\sigma}\mu_i(\hat{\theta})\Phi^{-1}(t) \\ &\Leftrightarrow \epsilon_{ij} \leq \frac{\hat{\sigma}\mu_i(\hat{\theta})\Phi^{-1}(t) + \left(\mu_i(\theta) - \mu_i(\hat{\theta})\right)}{\sigma\mu_i(\theta)} \\ &\Leftrightarrow \Phi(\epsilon_{ij}) \leq H_i(t), \end{aligned}$$

where

$$H_i(t) = \Phi\left[\frac{\hat{\sigma}\mu_i(\hat{\theta})\Phi^{-1}(t) + \left(\mu_i(\theta) - \mu_i(\hat{\theta})\right)}{\sigma\mu_i(\theta)}\right].$$

Taylor expansion of $H_i(t)$ around $\hat{\theta} = \theta$ gives

$$H_i(t) = t + \left(\hat{\theta} - \theta\right)^T \psi_i(t) + (\hat{\sigma} - \sigma)\eta_i(t) + \text{negligible terms},$$

where

$$\begin{aligned} \psi_i(t) &= \phi\left[\Phi^{-1}(t)\right]\left\{\frac{1}{\mu_i(\theta)}\frac{\partial\mu_i(\theta)}{\partial\theta}\Phi^{-1}(t) + \frac{1}{\sigma\mu_i(\theta)}\frac{\partial\mu_i(\theta)}{\partial\theta}\right\} \\ &= \phi\left[\Phi^{-1}(t)\right]\left[\Phi^{-1}(t) + \frac{1}{\sigma}\right]\frac{\partial}{\partial\theta}\log(\mu_i(\theta)) \\ &= \phi\left[\Phi^{-1}(t)\right]\left[\Phi^{-1}(t) + \frac{1}{\sigma}\right]\omega_i(\theta), \quad \text{for } \omega_i(\theta) = \tfrac{\partial}{\partial\theta}\log(\mu_i(\theta)) \end{aligned}$$

and $\eta_i(t) = \phi\left[\Phi^{-1}(t)\right]\left\{\frac{1}{\sigma}\Phi^{-1}(t)\right\}$. Therefore,

$$\begin{aligned} W_N(t) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\{I\left[\Phi(\hat{\epsilon}_{ij}) \le t\right] - t\} \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\{I\left[\Phi(\epsilon_{ij}) \le H_i(t)\right] - H_i(t)\} + \frac{1}{\sqrt{N}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[H_i(t) - t\right] \\ &\approx \frac{1}{\sqrt{N}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\{I\left[\Phi(\epsilon_{ij}) \le t\right] - t\} + \frac{1}{\sqrt{N}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\hat{\theta} - \theta\right)^T \psi_i(t) \\ &\quad + \frac{1}{\sqrt{N}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\hat{\sigma} - \sigma)^T \eta_i(t). \end{aligned}$$

In the last step we used the fact that, when $\hat{\theta}$ is close to θ, $H_i(t)$ is close to t. This step can probably be justified rigorously by following Loynes [45] but we have not tried to give precise conditions under which this is possible. Note that the same argument does not hold for the second term since $H_i(t) - t$ is $O_p(N^{-1/2})$.

Letting $\xi^T = (\theta^T, \sigma)$ and $\kappa_i^T(t) = \left(\psi^T(t), \eta_i(t)\right)$ we write

$$\begin{aligned} W_N(t) &\approx \frac{1}{\sqrt{N}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\{I\left[\Phi(\epsilon_{ij}) \le t\right] - t\} + \frac{1}{\sqrt{N}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\hat{\xi} - \xi\right)^T \kappa_i(t) \\ &= U_N(t) + V_N(t) \ \text{(say)}. \end{aligned}$$

Now the covariance kernel of the process $W_N(\cdot)$ can be computed as

$$\begin{aligned} Cov\left(W_N(s), W_N(t)\right) &= Cov\left(U_N(s), U_N(t)\right) + Cov(V_N(s), V_N(t)) \\ &\quad + Cov\left(U_N(s), V_N(t)\right) + Cov\left(U_N(t), V_N(s)\right). \end{aligned}$$

Computing $Cov\left(U_N(s), U_N(t)\right)$:

The process U_N is the standard empirical process of the N iid variates ϵ_{ij}. Hence, $Cov\left(U_N(s), U_N(t)\right) = \min(s,t) - st$. Note that U_N converge, as in Section 8.2.6, to a Brownian Bridge.

Computing $Cov\,(V_N(s), V_N(t))$:

Consider

$$\begin{aligned} Cov\,(V_N(s), V_N(t)) &= Cov\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\hat{\xi}-\xi\right)^T\kappa_i(s), \frac{1}{\sqrt{N}}\sum_{i'=1}^{k}\sum_{j'=1}^{n_{i'}}\left(\hat{\xi}-\xi\right)^T\kappa_{i'}(t)\right) \\ &= \frac{1}{N}\tau^T(s)Cov(\hat{\xi}-\xi)\tau(t), \end{aligned} \tag{8.12}$$

where

$$\begin{aligned} \tau(s) &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\kappa_i(s), \\ &= \begin{bmatrix} \phi\left[\Phi^{-1}(s)\right]\left[\Phi^{-1}(s)+\frac{1}{\sigma}\right]\sum_{i=1}^{k}\sum_{j=1}^{n_i}\omega_i(\theta) \\ N\phi\left[\Phi^{-1}(s)\right]\left\{\frac{1}{\sigma}\Phi^{-1}(s)\right\} \end{bmatrix}. \end{aligned} \tag{8.13}$$

The maximum likelihood estimates $\hat{\xi}$ solves the set of equations $\sum_{i=1}^{k}\sum_{j=1}^{n_i} S_{ij}(\theta,\sigma) = 0$, where $S_{ij}^T(\theta,\sigma) = (\frac{\partial l}{\partial \theta^T}, \frac{\partial l}{\partial \sigma})$ and l is the log-likelihood for the sample. Taylor expansion of $S_{ij}(\hat{\theta},\hat{\sigma})$ around $S_{ij}(\theta,\sigma)$ gives

$$\begin{aligned} 0 &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} S_{ij}(\hat{\theta},\hat{\sigma}) \approx \sum_{i=1}^{k}\sum_{j=1}^{n_i} S_{ij}(\theta,\sigma) + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left.\frac{\partial S_{ij}(\hat{\theta},\hat{\sigma})}{\partial\hat{\theta}}\right|_{(\theta,\sigma)}(\hat{\theta}-\theta) \\ &\quad + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left.\frac{\partial S_{ij}(\hat{\theta},\hat{\sigma})}{\partial\hat{\sigma}}\right|_{(\theta,\sigma)}(\hat{\sigma}-\sigma) \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} S_{ij}(\xi) + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left.\frac{\partial S_{ij}(\hat{\xi})}{\partial\hat{\xi}}\right|_{\xi}(\hat{\xi}-\xi) \end{aligned}$$

Hence,

$$\begin{aligned} (\hat{\xi}-\xi) &\approx -\left[\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left.\frac{\partial S_{ij}(\hat{\xi})}{\partial\hat{\xi}}\right|_{\xi}\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} S_{ij}(\xi)\right] \\ &\approx I^{-1}(\xi)\left[\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} S_{ij}(\xi)\right], \end{aligned}$$

where $I(\xi) = E\left(-\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left.\frac{\partial S_{ij}(\hat{\xi})}{\partial\hat{\xi}}\right|_{\xi}\right)$. Therefore,

$$Cov(\hat{\xi}-\xi) \approx I^{-1}(\xi)Var\left(\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} S_{ij}(\xi)\right)I^{-1}(\xi).$$

As usual in maximum likelihood theory,

$$Var\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} S_{ij}(\xi)\right) = E\left(-\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left.\frac{\partial S_{ij}(\hat{\xi})}{\partial\hat{\xi}}\right|_{\xi}\right) = NI(\xi).$$

Thus, $Cov(\hat{\xi}-\xi) \approx \frac{1}{N}I^{-1}(\xi)$, where $I(\xi)$ is the average Fisher information per observation. It is easy to see that the matrix $I(\xi)$ has the following components:

$$I(\xi) = \begin{bmatrix} (\frac{1}{\sigma^2}+2)\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\omega_i(\theta)\omega_i^T(\theta) & \frac{2}{\sigma}\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\omega_i(\theta) \\ \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{2}{\sigma}\omega_i^T(\theta) & \frac{2}{\sigma^2} \end{bmatrix}.$$

Now from equation 8.12 we find

$$Cov\left(V_N(s), V_N(t)\right) \approx \frac{1}{N^2}\tau^T(s)I^{-1}(\xi)\tau(t).$$

Computing $Cov\left(U_N(s), V_N(t)\right)$**:**

Now consider

$$\begin{aligned} Cov\left(U_N(s), V_N(t)\right) &= Cov\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left\{I\left[\Phi(\epsilon_{ij}) \le s\right] - s\right\}, \frac{1}{\sqrt{N}}\sum_{i'=1}^{k}\sum_{j'=1}^{n_{i'}}(\hat{\xi}-\xi)^T\kappa_{i'}(t)\right) \\ &= \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Cov\left(I\left[\epsilon_{ij} \le \Phi^{-1}(s)\right], (\hat{\xi}-\xi)^T\right)\tau(t) \qquad (8.14) \end{aligned}$$

Recall that

$$\begin{aligned} (\hat{\xi}-\xi) &\approx I^{-1}\left[\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} S_{ij}(\xi)\right] \\ &= I^{-1}\begin{bmatrix} -\frac{1}{N}\sum_{i=1}^{k} n_i\omega_i(\theta) + \frac{1}{N}\frac{1}{\sigma}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\epsilon_{ij}\omega_i(\theta) + \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\epsilon_{ij}^2\omega_i(\theta) \\ -\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{1}{\sigma} + \frac{1}{N}\frac{1}{\sigma}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\epsilon_{ij}^2 \end{bmatrix}, \end{aligned}$$

where $I^{-1} = [I(\xi)]^{-1}$. Observing $E(\hat{\xi}-\xi) = 0$, we write

$$Cov\left(I\left[\epsilon_{ij} \le \Phi^{-1}(s)\right], (\hat{\xi}-\xi)^T\right) = E\left(I\left[\epsilon_{ij} \le \Phi^{-1}(s)\right](\hat{\xi}-\xi)^T\right).$$

To evaluate the above covariance we need the following terms:

$$E\left(I\left[\epsilon_{ij} \le \Phi^{-1}(s)\right]\right) = s.$$

$$
\begin{aligned}
E\left(\epsilon_{ij} I\left[\epsilon_{ij} \leq \Phi^{-1}(s)\right]\right) &= \int_{-\infty}^{\Phi^{-1}(s)} \frac{1}{\sqrt{2\pi}} x e^{-x^2/2} dx, \\
&= -\frac{1}{\sqrt{2\pi}} \exp\left\{\left[\Phi^{-1}(s)\right]^2/2\right\} = -\phi\left[\Phi^{-1}(s)\right]. \\
E\left(\epsilon_{ij}^2 I\left[\epsilon_{ij} \leq \Phi^{-1}(s)\right]\right) &= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} x^2 e^{-x^2/2} dx, \quad \text{where } y = \Phi^{-1}(s) \\
&= -\int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} x^2 \frac{d}{dx} e^{-x^2/2} dx, \\
&= \frac{1}{\sqrt{2\pi}} \left[x e^{-x^2/2}\right]_{y}^{-\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2} dx \\
&= -\frac{1}{\sqrt{2\pi}} y e^{-y^2/2} + \Phi(y) \\
&= -y\phi(y) + \Phi(y) \\
&= -\Phi^{-1}(s)\phi\left[\Phi^{-1}(s)\right] + s.
\end{aligned}
$$

Now it is easy to see using Equation 8.13 that

$$
\begin{aligned}
&Cov\left(I\left[\epsilon_{ij} \leq \Phi^{-1}(s)\right], (\hat{\xi} - \xi)^T\right) \\
&= \begin{bmatrix} -\phi\left[\Phi^{-1}(s)\right] \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \omega_i(\theta) \left(\frac{1}{\sigma} + \Phi^{-1}(s)\right) \\ \\ -\frac{1}{\sigma}\Phi^{-1}(s)\phi\left[\Phi^{-1}(s)\right] \end{bmatrix}^T I^{-1}(\xi) \\
&= -\frac{1}{N}\tau^T(s) I^{-1}(\xi).
\end{aligned}
$$

Now from Equation 8.14 we find $Cov\left(U_N(s), V_N(t)\right) = -\frac{1}{N^2}\tau^T(s) I^{-1}(\xi)\tau(t)$.

Therefore, $Cov(W_N(s), W_N(t)) = \min(s,t) - st - \frac{1}{N^2}\tau^T(s) I^{-1}(\xi)\tau(t)$, where $I(\xi)$ is the average Fisher information per observation. To apply the suggested test procedure we need an estimate for $I(\xi)$. We have two choices for estimating $I(\xi)$.

1. We can estimate $I(\xi)$ by replacing the unknown parameters in the formula for $I(\xi)$ by their maximum likelihood estimates.

2. Let H be Hessian matrix, that is, the matrix of second derivatives of the log likelihood with respect to the parameters (i.e. components of θ and σ). The average Fisher information matrix $I(\xi)$ can be replaced by $-H/N$, where H is evaluated at the maximum likelihood estimates for θ and σ. In this case, we therefore find p-values

from the distribution of $\int_0^1 Y^2(t)dt$ where Y is a mean zero Gaussian process with covariance function $\min(s,t) - st - \frac{1}{N^2}\tau^T(s)\left(\frac{-H}{N}\right)\tau(t)$.

8.3.4 Justification for using the approximate p-value

Now we describe a Monte Carlo study that examines the performance of the approximate p-value suggested in Section 8.3.2 for testing the normality of random errors ϵ_{ij} in the model, $Y_{ij} = \mu_i(\theta) + \sigma\mu_i(\theta)\epsilon_{ij}$, $i = 1,\ldots,k$, $j = 1,\ldots,n_i$. We chose the mean function $\mu_i(\theta) = \alpha_1\left\{1 - \exp\left[-\frac{(x+\alpha_2)}{\alpha_3}\right]\right\}$, where α_1, α_2 and α_3 are the components of θ. This is the mean response function for the unbleached (or bleached) data for the partial bleach method described in chapter 3. We set the parameters at $\alpha_1 = 14.28528, \alpha_2 = 123.1816$ and $\alpha_3 = 393.0665$, which are the maximum likelihood estimates obtained for the unbleached data set QNL84-2 given in Berger et. al. [12]. The vector of dose values used is given in Appendix 9.3 where it is labeled with the code 1A. The values of σ and the sample sizes, n, used are given in Tables 8.4 and 8.5. The standard normal random variates ϵ_{ij} were generated using the IMSL subroutine RNNOA.

For each sample we computed p-values for testing the normality of random errors using the Cramér-von Mises statistic and Anderson-Darling statistic as described in Section 8.3.2. If the suggested asymptotic theory holds these p-values should be uniformly distributed. The results based on 1000 simulations for using Cramér-von Mises statistic and Anderson-Darling statistic are given in Tables 8.4 and 8.5 respectively. For both tests, the uniformity of p-values was tested using the Anderson-Darling statistic. To conserve space in the tables, we denote Cramér-von Mises statistic and Anderson-Darling statistic as W^2 and A^2 respectively. For each set of 1000 p-values we also report the actual levels (fraction of p-values less than the nominal value of α) observed from the study for values of α equal to 0.01, 0.05 and 0.10.

Based on the simulation results we conclude the following.

1. When we use the least squares estimate for σ, the Anderson-Darling test does not reject the null hypothesis that the p-values are uniformly distributed except in one case; including this case the observed coverages were found to agree well with the

n	σ	ML $\hat{\sigma}$					Least squares $\hat{\sigma}$				
		A^2	p-value	Observed level for			A^2	p-value	Observed level for		
		stat		0.01	0.05	0.10	stat		0.01	0.05	0.10
16	0.01	1.05	0.33	0.009	0.058	0.113	0.60	0.65	0.008	0.049	0.098
16	0.02	1.76	0.13	0.010	0.051	0.100	0.94	0.39	0.009	0.043	0.086
16	0.03	4.46	0.00	0.010	0.067	0.116	2.06	0.08	0.008	0.054	0.104
16	0.04	3.10	0.03	0.015	0.066	0.119	1.92	0.10	0.010	0.056	0.101
50	0.01	0.60	0.65	0.011	0.058	0.111	0.45	0.80	0.010	0.056	0.110
50	0.02	0.94	0.39	0.010	0.049	0.100	1.10	0.31	0.009	0.048	0.098
50	0.03	1.10	0.31	0.014	0.060	0.118	0.89	0.42	0.012	0.057	0.111
50	0.04	1.88	0.11	0.008	0.058	0.116	1.79	0.12	0.008	0.054	0.115

Table 8.4: Table of p-values for testing normality using W^2 statistic: Model 2

n	σ	ML $\hat{\sigma}$					Least squares $\hat{\sigma}$				
		A^2	p-value	Observed level for			A^2	p-value	Observed level for		
		stat		0.01	0.05	0.10	stat		0.01	0.05	0.10
16	0.01	0.96	0.38	0.010	0.057	0.107	3.09	0.03	0.007	0.044	0.085
16	0.02	3.02	0.03	0.009	0.050	0.103	1.01	0.35	0.004	0.037	0.080
16	0.03	5.20	0.00	0.009	0.063	0.121	0.49	0.76	0.006	0.044	0.094
16	0.04	4.54	0.00	0.016	0.067	0.118	0.79	0.48	0.010	0.051	0.099
50	0.01	0.56	0.69	0.011	0.058	0.113	0.26	0.96	0.009	0.050	0.104
50	0.02	0.37	0.88	0.009	0.050	0.103	1.00	0.36	0.004	0.037	0.080
50	0.03	0.80	0.48	0.013	0.061	0.112	0.59	0.65	0.010	0.056	0.105
50	0.04	1.15	0.29	0.006	0.060	0.119	1.33	0.22	0.006	0.053	0.108

Table 8.5: Table of p-values for testing normality using A^2 statistic: Model 2

nominal coverages for significance levels $\alpha = 0.01, 0.05, 0.10$. This justifies the use of the approximate p-values using least squares estimate for σ suggested in Section 8.1.2), for testing the normality of random errors.

2. When we use the maximum likelihood estimate for σ, the Anderson-Darling test was found to reject the null hypothesis that the p values are uniformly distributed in cases when the sample sizes are small and σ is large. The Anderson-Darling test is more sensitive to departures in the tails of the distribution. However, for all the cases, the observed coverages for levels $\alpha = 0.01, 0.05, 0.10$ still agree well with the nominal coverages. In goodness of fit problems, often we are interested in the lower tail probabilities. Since, in the lower tails the coverage probabilities agree well, we conclude that we can safely use the approximate p-values using maximum likelihood estimate for σ suggested in Section 8.1.2), for testing the normality of random errors.

3. The suggested asymptotic theory appears to hold for sample sizes as small as 16. The typical sample sizes for thermoluminescence studies are around 30. Therefore, the suggested procedure could safely be used for testing the normality assumption in the model described for thermoluminescence data.

Remarks:

The procedure we described here uses single mean and variance functions for all the data. This is equivalent to assuming all the data correspond to a single response function. For the partial bleach method and the regeneration method two data sets are collected at a given temperature. The observed photon counts for unbleached and bleached data have different mean and variance functions. The procedure described here can easily be extended to situations where the data corresponds to more than one response function. For example, to apply the test procedure for a situation where we have two response curves (such as the partial bleach method or the regeneration method) each corresponding to a common σ, first compute the maximum likelihood estimates for θ as described in Chapters 5 or 6. Then compute the fitted residuals for the unbleached and bleached data sets as described earlier by using the corresponding mean and variance functions. The assumption of normality

of the random errors can then be tested by examining whether the fitted residuals are approximately normally distributed.

8.3.5 Sensitivity of the tests to departures from normality

Now we describe a simulation study that we performed to examine the power of the EDF tests described in Section 8.3 for detecting the departures from normality. The simulation study is similar to that which we described in Section 8.3.4 except that, instead of normal random variates we generated data from different alternative distributions. When generating data from the gamma distribution we fix the shape parameters at values given in Table 8.6. Since the mean of Y is f and variance of Y is $\sigma^2 f^2$, σ^2 is the reciprocal of the shape parameter and the scale parameter for Y is $\sigma^2 f$. So, we only report the shape parameter for the gamma distribution.

For each generated sample, we computed an approximate p-value for testing the hypothesis that the random errors are normally distributed. We examined the performance of both the Anderson-Darling statistic and the Cramér-von Mises statistic as test statistics for testing the normality. The power of the test was computed as the proportion of samples rejected by each test. The chosen distributions, sample sizes, and the results are given in Tables 8.6 and 8.7. The parameter vector $\theta = (\alpha_1, \alpha_2, \alpha_3)$ was fixed at $\alpha_1 = 14.28528, \alpha_2 = 123.1816$ and $\alpha_3 = 393.0665$. The dose vector used for the study is coded as data $1A$ in the Appendix 9.3.

Results based on 1000 simulations are summarized in Tables 8.6 and 8.7 respectively.

Conclusions

Based on the simulation study we draw the following conclusions.

1. The tests appear to detect the departure from normality with reasonably large sample sizes.

2. The Anderson-Darling test appears to be slightly more powerful than the Cramér-von Mises statistic.

Data Set	Distribution	n	The power of W^2			
			for mle $\hat{\sigma}$		for least squares $\hat{\sigma}$	
			$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
1	gamma(0.5)	16	0.37	0.55	0.37	0.55
2	gamma(1)	16	0.20	0.31	0.17	0.30
3	gamma(0.5)	30	0.67	0.84	0.69	0.86
4	gamma(1)	30	0.34	0.50	0.34	0.51
5	gamma(2)	30	0.20	0.31	0.19	0.30
6	gamma(4)	30	0.11	0.19	0.11	0.18

Table 8.6: Results for power studies using W^2: Model 2

3. Comparing the results of Table 8.3 and 8.7 we conclude that the test that assumes fitted model is correct is substantially more powerful than the test that do not assume the fitted model is correct.

8.4 Estimates for the eigenvalues of $\rho(s,t)$

We used the same approach used by Lockhart *et. al.* [44] to find estimates for the eigenvalues of $\rho(s,t)$.

Divide the interval [0,1] into m sub intervals each of length $\frac{1}{(m+1)}$. Then,

$$\begin{aligned} \lambda f(i/(m+1)) &= \int_0^1 \rho(i/(m+1),t)f(t)dt \\ &\approx \frac{1}{m}\sum_{i=1}^{m} \rho(i/(m+1),j/(m+1))f(j/(m+1)), \end{aligned}$$

for sufficiently large m.

Let V be the $m-$vector consisting of the elements $(f(1/(m+1)),\ldots,f(m/(m+1)))$, and Q be the $m \times m$ matrix whose (i,j)th element is $Q_{ij} = \frac{1}{m}\rho(i/(m+1),j/(m+1))$. Then the above set of equations can be written as, $\lambda V = QV$. Thus, finding the eigenvalues of $\rho(s,t)$

Data Set	Distribution	n	The power of A^2			
			for mle $\hat{\sigma}$		for least squares $\hat{\sigma}$	
			$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
1	gamma(0.5)	16	0.38	0.60	0.35	0.56
2	gamma(1)	16	0.18	0.33	0.15	0.29
3	gamma(0.5)	30	0.79	0.92	0.79	0.92
4	gamma(1)	30	0.38	0.61	0.34	0.51
5	gamma(2)	30	0.22	0.34	0.20	0.33
6	gamma(4)	30	0.11	0.19	0.11	0.18

Table 8.7: Results for power studies using A^2: Model 2

is reduced to the discretized problem of finding the eigenvalues of Q. Suppose $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of Q. These eigenvalues are approximations to the true eigenvalues of $\rho(s,t)$. The accuracy of this approximation could be increased by increasing the number of subdivisions m. The results we presented were based on 150 subdivisions.

8.5 Worked examples

In this section, we demonstrate the theoretical results suggested in this chapter using real data sets from the partial bleach method and the regeneration method.

8.5.1 Example from partial bleach data

The data set we used here is presented in Berger *et. al.* [12] where it is coded as QNL84-2. The sample sizes of the unbleached and bleached data sets are respectively $n_1 = 16$ and $n_2 = 13$.

First we describe the results for testing normality of random errors without assuming the fitted model is correct. As we mentioned in Section 8.2, for testing normality without assuming the fitted model is correct, it is necessary to have at least three observations at

each dose level. However for the data set QNL84-2, the dose level 960 of the unbleached data set had only two replicates. (These dose levels are presented in Table 9.2 of Appendix 9.3 where they are coded as P1.) For the bleached data set, each of the dose levels 120 and 960 had only two replicates. So, we had to ignore these dose levels when testing normality without assuming the fitted model is correct. Since different mean parameters are to be fitted for the unbleached and bleached data, we have to consider the dose levels of the unbleached and bleached data sets as different dose levels. Consequently, for testing normality without assuming the fitted model is correct, we only have 23 observations taken at 7 different dose levels. Figure 8.4 illustrates the histogram and the plot of ordered probability integral transforms vs uniform quantiles (probability plot) for the residuals obtained by fitting different mean parameters at each dose level.

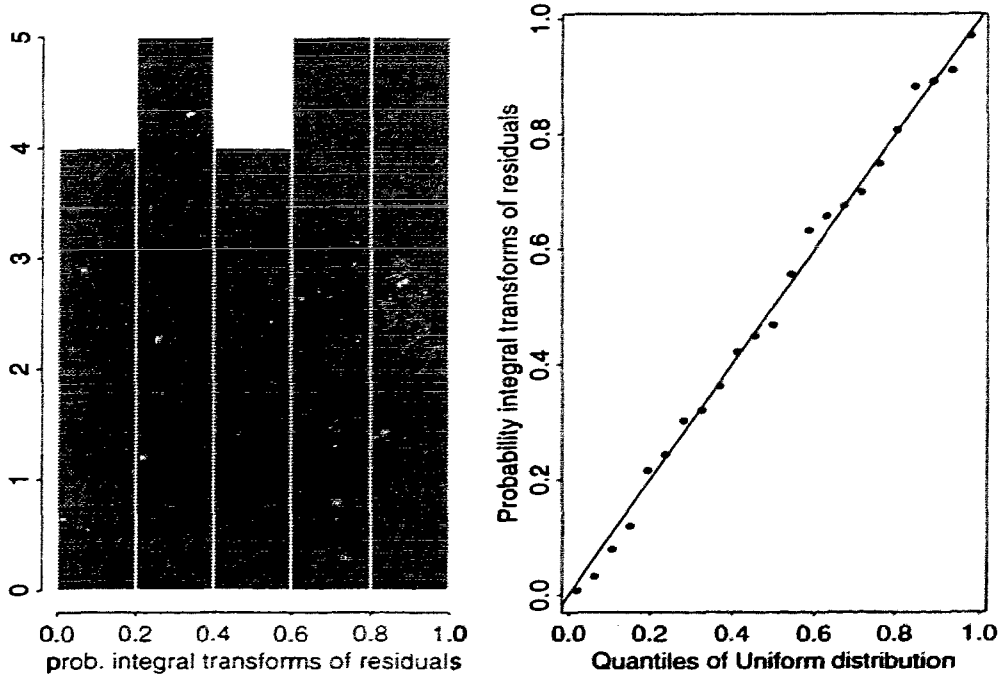

Figure 8.4: Histogram and probability plot of probability integral transforms of residuals: partial bleach method

The results for testing normality of random errors without assuming the fitted model is correct are as follows:

Cramér-von Mises test statistic = 0.014; p-value = 0.98

Anderson-Darling test statistic = 0.145; p-value = 0.95.

Now we describe the results for testing normality assuming the fitted model is correct. For this data set, we plotted the fitted residuals against the applied dose; see Figure 8.5. This

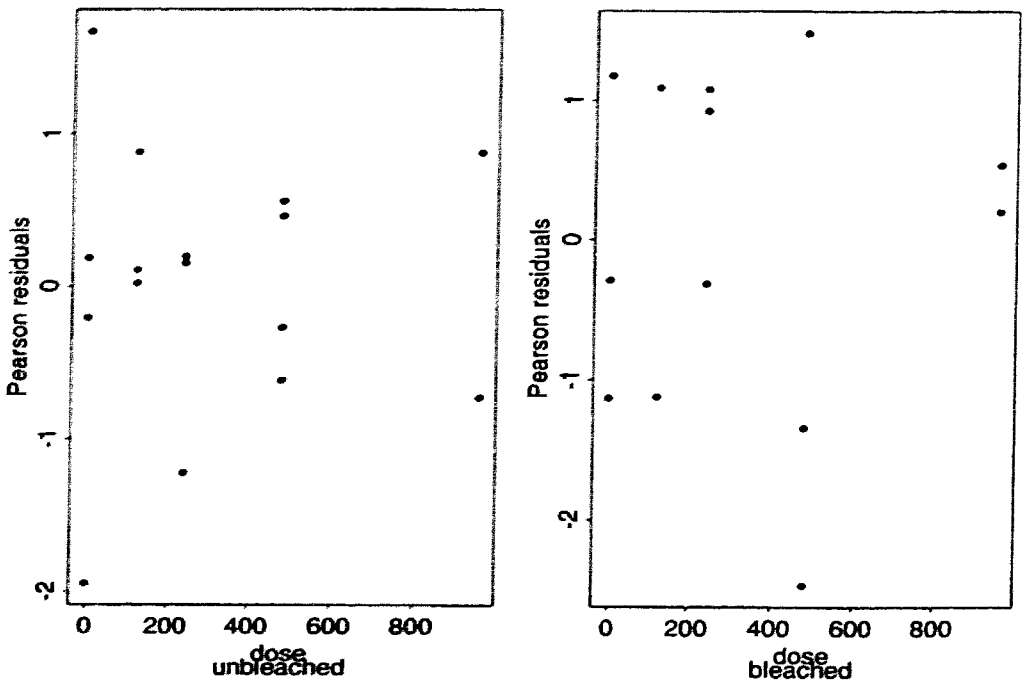

Figure 8.5: Plot of residuals vs applied dose: Data QNL84-2

plot does not indicate any evidence against the hypothesis that the fitted model is correct. For testing normality assuming the fitted model is correct, we have 29 observations. As we presented in Section 5.5, the maximum likelihood estimates for this data set are $\hat{\alpha}_1 = 14.28 \times 10^4, \hat{\alpha}_2 = 123.18, \hat{\alpha}_3 = 393.07, \hat{\beta}_2 = 192.55, \hat{\sigma}^2 = 0.0012$. Summary statistics for these residuals are as follows: N = 29, mean=1.2e-05, median = 0.155, standard deviation=1.018 and the first and the third quartiles = -0.611, 0.878. Figure 8.6 illustrates the histogram and the normal quantile plot for the residuals to the fit from the method of maximum likelihood.

We computed approximate p-values using the maximum likelihood estimate for σ and also using the least squares estimate for σ. The results are as follows:

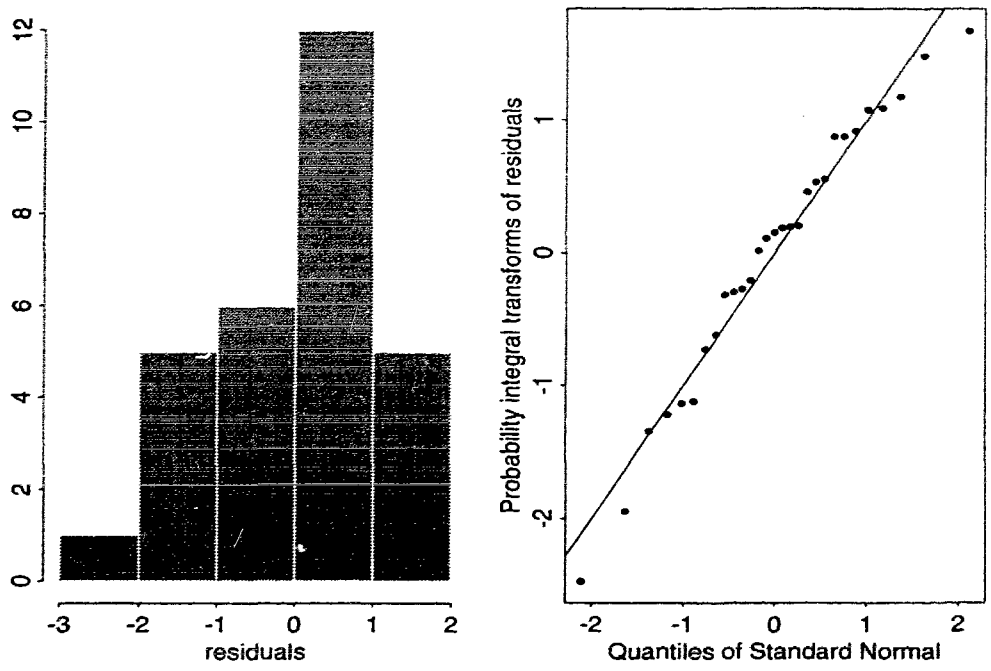

Figure 8.6: Histogram and normal probability plot of residuals: partial bleach method

Using maximum likelihood estimate for σ:

Cramér-von Mises test statistic = 0.046; p-value = 0.58

Anderson-Darling test statistic = 0.307; p-value = 0.58.

Using least squares estimate for σ:

Cramér-von Mises test statistic = 0.051; p-value = 0.50

Anderson-Darling test statistic = 0.344; p-value = 0.48.

Based on the above results we conclude that, we do not have enough evidence to reject the null hypothesis that the random errors in the photon counts are normally distributed.

8.5.2 Example from regeneration data

For this example we used the data SESA1 collected at temperature 360^0C cited in Huntley *et. al* [38]. This data set had 62 observations (30 unbleached and 32 bleached). The unbleached data set had only two replicates at dose level 120. (The dose levels for this data set are presented in Table 9.2 of Appendix 9.3, where they are labeled with the code R1.)

Consequently, we had to ignore this dose level when testing goodness of fit without assuming the fitted model is correct. So we had only 60 observations taken at 14 different dose levels for testing goodness of fit using the procedure proposed in Section 8.2. Figure 8.7 illustrates the histogram and the plot of ordered probability integral transforms vs uniform quantiles (probability plot) for the residuals obtained by fitting different mean parameters at each dose level.

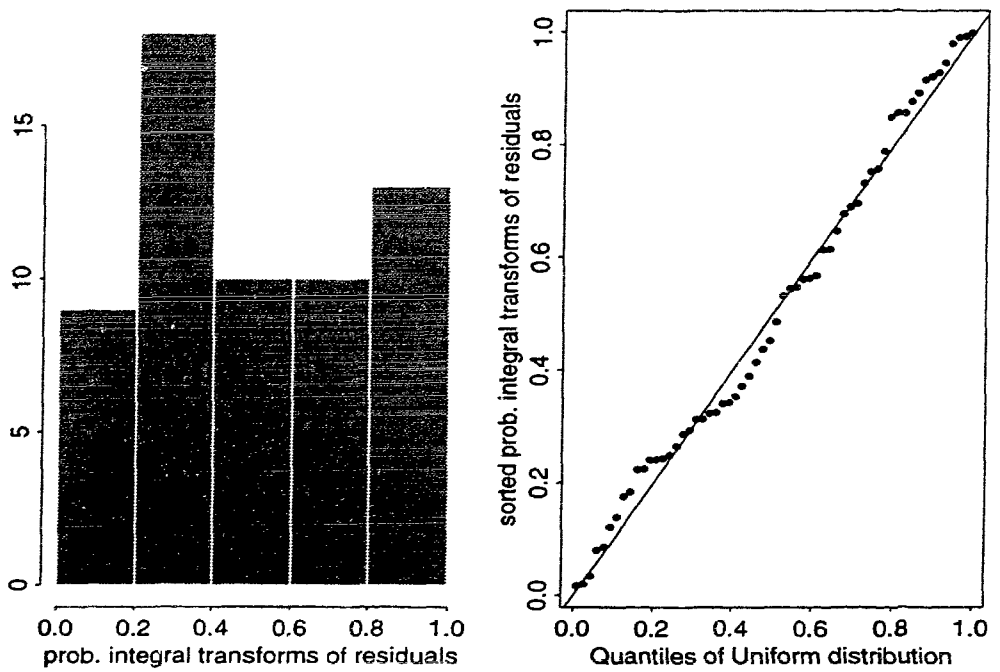

Figure 8.7: Histogram and probability plot of probability integral transforms of residuals

The results for testing normality without assuming the fitted model is correct are as follows:
Cramér-von Mises test statistic = 0.057; p-value = 0.34
Anderson-Darling test statistic = 0.507; p value = 0.24

Now we describe the test results assuming the fitted model is correct. The plot of fitted residuals against the applied dose for this data set is illustrated in Figure 8.8. This plot does not indicate any evidence against the hypothesis that the fitted model is correct. For testing normality assuming the fitted model is correct, we have 62 observations. The maximum

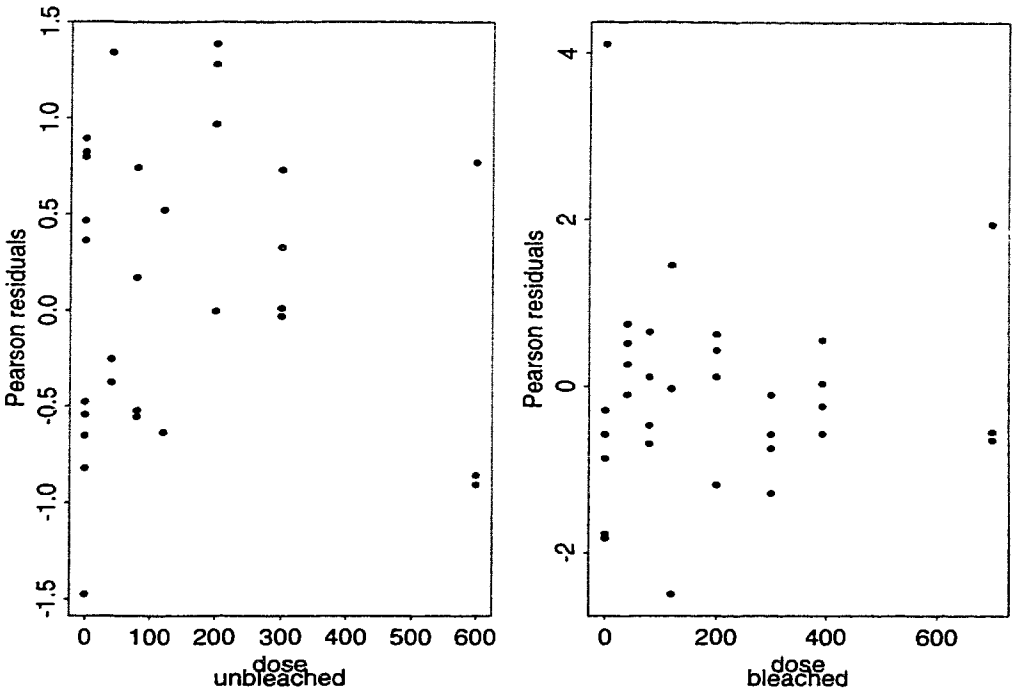

Figure 8.8: Plot of residuals vs dose: Data STRB87-1

likelihood estimates for the parameters are $\hat{\alpha}_1 = 57363.2, \hat{\alpha}_2 = 80.72, \hat{\alpha}_3 = 91.33, \hat{\alpha}_4 = 91.98, \hat{\gamma} = 73.11$ and $\hat{\sigma} = 0.06$. For the goodness of fit test, we used the residuals to the fit (from fitting the restricted model discussed in Chapter 6) from the method of maximum likelihood. Summary statistics for the residuals are as follows: N = 62, Mean=-9.9e-07, Median = -0.0320, standard deviation=1.0081 and the first and the third quartiles = -0.611116, 0.878173. Figure 8.9 illustrates the histogram and the normal probability plot of these residuals.

Again, we computed approximate p-values using the maximum likelihood estimate for σ and also using the least squares estimate for σ. The results are as follows:

Using maximum likelihood estimate for σ:

Cramér-von Mises test statistic = 0.083 ; p-value = 0.19

Anderson-Darling test statistic = 0.618; p value = 0.11.

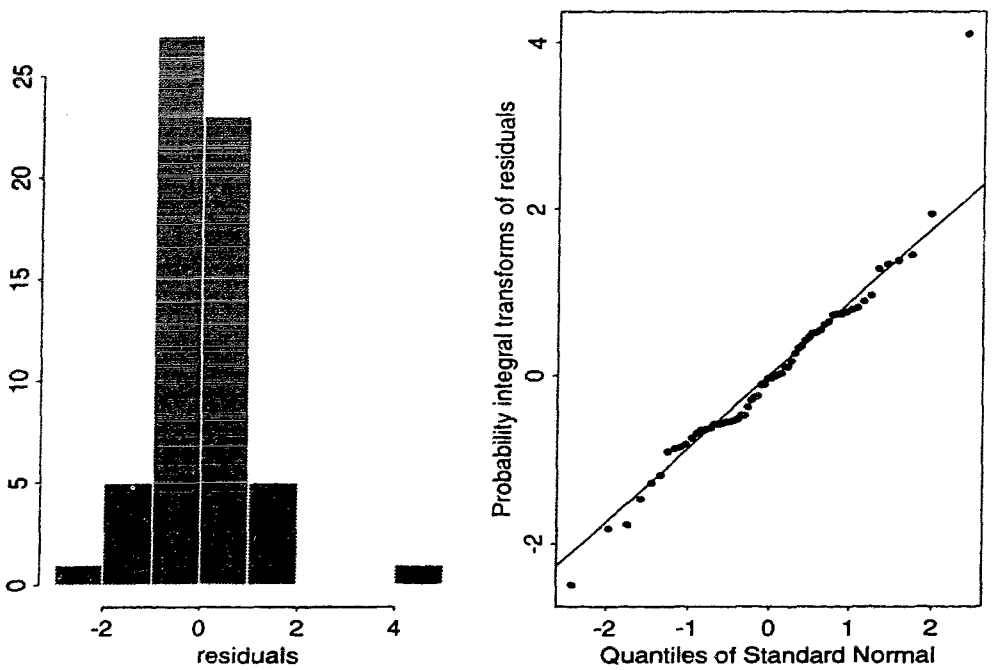

Figure 8.9: Histogram and normal probability plot of residuals: regeneration method

Using least squares estimate for σ:

Cramér-von Mises test statistic = 0.010; p-value = 0.10

Anderson-Darling test statistic = 0.755; p value = 0.05.

From the plots of residuals, we notice that there is an unusual observation in this data set (an observation from a disc receiving no added dose in the bleached data set). Since this might have a large influence on the goodness of fit tests, in particular for the Anderson-Darling test, we also computed approximate p values disregarding this observation. The results are as follows.

Using maximum likelihood estimate for σ:

Cramér-von Mises test statistic = 0.062 ; p-value = 0.36

Anderson-Darling test statistic = 0.407; p value = 0.35.

Using least squares estimate for σ:

Cramér-von Mises test statistic = 0.059; p-value = 0.40

Anderson-Darling test statistic = 0.421; p value = 0.321 .

Based on the above results we conclude that, we do not have enough evidence to reject the null hypothesis that the random errors in the photon counts are normally distributed.

8.6 Discussion

In this chapter we proposed tests based on the empirical distribution function (EDF tests) of the standardized residuals for testing the assumption of normality of random errors. We outlined two test procedures. The first procedure can be used to test the assumption of normality without assuming the fitted model is correct. In this model different mean and variance parameters are fitted at each dose level. Usually for thermoluminescence data, not more than five replicates are available at each dose level. Consequently, mean and variance parameters cannot be estimated consistently. Therefore, the fitted standardized residuals need not be asymptotically normally distributed even if the random errors are. We derived the true distribution and the joint distribution function of two fitted standardized residuals and use these to compute the EDF statistics and corresponding p-values. The assumption of the normality of random errors is tested by checking to see if the fitted standardized residuals follow the derived true distribution.

The second test procedure can be used if the fitted model is assumed to be correct. For each case, we show how to compute the test statistic and an approximate p-value for testing the assumption of normality. The finite sample performance of the suggested asymptotic theory is tested by a Monte Carlo study. The procedure which assumes fitted model is correct was found to be substantially more powerful than the test that do not assume the fitted model is correct.

Weak convergence results for the empirical process of residuals are established rigorously in Section 8.2.6 for the case of Model 1 and heuristically in Section 8.3.2 for the case of Model 2.

Chapter 9

Concluding remarks

In this chapter, we summarize the conclusions of this study and provide some guidance for analyzing data collected in thermoluminescence studies based on the experience of this work. Suggestions are made for further research.

9.1 Summary of work and conclusions

In thermoluminescence studies, the age of the sample is determined by estimating the dose impinging on the sample during its burial period and comparing with the radiation dose rates of the surrounding soil. The radiation dose impinging on the sample is quantified as the equivalent dose, which is a known laboratory dose required to produce the same amount of luminescence as that given off by a natural sample upon gentle heating in the laboratory. Lack of knowledge about the sample at the time of deposition distinguishes sedimentary dating from pottery dating. Unlike pottery dating, not much is known about the sample at the time of deposition. Therefore, part of the luminescence produced upon gentle heating in the laboratory could have been due to emptying of traps that were already filled at the time of deposition. This makes the estimation of the dose acquired during burial more difficult. The partial bleach method and the regeneration method are widely used techniques in sedimentary dating that avoid the necessity to know the amount of thermoluminescence at the time of deposition. In this work, we focused our attention

on the estimation of the equivalent dose from the data collected in thermoluminescence studies. The theoretical results developed in this work have a range of applications outside the framework of thermoluminescence studies.

In Chapter 2, we proposed physical models that motivate generalized non-linear modeling of the TL data. We examined the performance of the traditional estimation procedures (maximum likelihood, quasi likelihood and generalized least squares) for estimating the unknown parameters in these models. We also examined another estimation procedure used by physicists that is closely related to generalized least squares. In this procedure, they use observed data y in place of the expected values $E(y)$ appearing in the weight function. These estimation techniques were compared by examining the statistical properties of these estimators. Large samples are not common in thermoluminescence studies. Usual sample sizes are around 40, but the effective sample sizes are much smaller than this due to the presence of several nuisance parameters (about six nuisance parameters for each model). However, the measurement errors are quite small, the relative error in a single measurement, σ, being around 3% to 8%. Consequently, small sigma asymptotics are more appropriate for thermoluminescence data. However, for completeness we examined both large sample and small σ asymptotics. Examining the large sample properties of the estimators we found that the maximum likelihood and quasi likelihood estimates are consistent while generalized least squares and data weighted least squares are generally not. We also presented distributional approximations to maximum likelihood and generalized least squares estimators. The large sample asymptotic variances of maximum likelihood estimates were found to be smaller than those of quasi likelihood estimates assuming the errors are normally distributed. We derived approximate formulae for the biases and mean squared errors of the suggested estimators valid in the small σ asymptotic case. In the limiting case of small σ and large samples, maximum likelihood and quasi likelihood estimators were found to produce mean squared error consistent estimators. The generalized least squares and data weighted least squares estimators were found to have biases that do not vanish asymptotically. However, for parameters of interest in our model, generalized least squares and data weighted least squares were also found to produce asymptotically unbiased estimators when σ is small.

In realistic size samples, all four estimators were found to have negligible biases compared to their standard errors. Furthermore, formulae derived based on asymptotic results were found to hold for sample sizes and values of σ typical of real data sets. The approximate formulae for the biases and the mean squared errors derived in this work are valid for any response function $f(x, \theta)$, not just those discussed for thermoluminescence data.

In the partial bleach method and the regeneration method two data sets are collected: one set from a portion of the untreated sample and the other set from a portion of the sample left in the sunlight for a desired period of time or subjected to optical bleaching in the laboratory. The data collected on the sample that has been subjected to bleaching are called bleached data whereas the other data are called unbleached data. Berger *et. al.* [11] have shown how to compute quasi likelihood estimates and confidence intervals for the equivalent dose based on the quasi likelihood estimate assuming a single error factor is suitable for both unbleached and bleached data. We extended their results to the case of different error factors using a Satterthwaite type approximate degrees of freedom.

We examined profile likelihood intervals and symmetric confidence intervals based on the maximum likelihood estimate using z and t critical values. We also proposed confidence intervals for the equivalent dose based on a transformation of the likelihood ratio test with a transformed F critical value. Based on the quasi likelihood estimate, we examined quasi score intervals and symmetric confidence intervals using t critical values. Based on the generalized least squares and data weighted least squares estimates, we examined symmetric confidence intervals using t critical values. Finite sample performance of the suggested confidence intervals were examined by a Monte Carlo study. When sample sizes are small ($n < 40$), coverage probabilities of profile likelihood intervals and symmetric confidence intervals with standard normal quantiles were found to have lower coverage probabilities than their nominal levels. The coverage probabilities of quasi score intervals and confidence intervals based on the transformation of the likelihood ratio statistic and symmetric confidence intervals based on the t quantiles were found to agree well with their nominal levels even in small samples. Usually, data sets collected for thermoluminescence studies are small in size ($n < 40$). Therefore, we recommend confidence intervals based on t quantiles as opposed to

profile type confidence intervals. Symmetric confidence intervals have the added advantage that they are computationally much simpler. Of the symmetric confidence intervals we discussed in this work we favour those based on quasi likelihood estimates since they depend on fewer assumptions about the random errors.

In thermoluminescence studies, data are collected on a single sample at a series of temperatures. The equivalent dose is estimated from data collected at each temperature and the estimates are then plotted against the temperature. If the sample is capable of producing a reliable estimate for the equivalent dose it is expected to see a plateau (a region where the estimated equivalent dose does not vary with the temperature). This plateau is believed to represent the *stable traps* (traps that have not been subjected to leakage over the burial time). It is the equivalent dose estimated from the data corresponding to these traps that can provide reliable information for dating purposes; see Aitken [1]. Since the same samples are used to collect the observations over the plateau these observations are correlated (in fact the correlations are very high). We proposed a procedure closely related to that of generalized estimating equations suggested in Liang and Zeger [43] for estimating the equivalent dose from the correlated data. Finite sample performance of the asymptotic theoretical results was examined by a Monte Carlo study. For realistic sample sizes, coverage probabilities of symmetric confidence intervals using standard normal quantiles were found to be lower than their nominal levels. Using small σ asymptotics, we showed that the estimated variance of the estimate for the equivalent dose can be approximated by a sum of independent chi-squared random variables. Hence, we proposed confidence intervals with a t quantile. Using Satterthwaite's [53] approximation, a formula was proposed to compute the approximate degrees of freedom for the appropriate t quantile. The coverage probabilities of confidence intervals with t quantiles were found to be closer to their nominal levels than those with standard normal quantiles.

When the number of temperatures on the plateau is increased up to about four we found that the coverage probabilities begin to drop from their nominal levels. This may be attributed to having to fit a large number of nuisance parameters (five additional nuisance parameters need to be fitted with the addition of each temperature on the plateau) with only

a small number of replicate samples (usually around 40). So we further imposed restrictions on the correlations about the samples at different temperatures which we feel are not too stringent; for example we assumed that the correlation between two observations on the same disc over the region of the plateau depends only on the difference between the temperatures but not on the temperature at which the measurements were taken. This suggests that the correlations along the diagonals at equal distance about the main diagonal are the same. Thus, we have fewer correlation coefficients to estimate. The data were generated with such correlation structures and the coverage probabilities were examined. These restrictions were found to improve the precision of the coverage probabilities of the confidence intervals.

In Chapter 8, we proposed tests based on the empirical distribution function (EDF tests) of the fitted standardized residuals for testing distributional assumptions of the random errors. We proposed two test procedures for this purpose. First we showed how to test the distributional assumptions without relying on the assumption that the fitted model is correct. In this model, different mean and variance functions are fitted at each dose level. Consequently, with the number of dose levels the number of fitted parameters also increases. Therefore, the fitted standardized residuals need not be asymptotically normally distributed even if the random errors are. Using the results of Beckman *et. al.* [3] we derived the true distribution function of the fitted standardized residuals. The joint distribution function of two fitted residuals was derived using the results of Ellenberg [29]. Formulae were derived to compute the joint distribution function of two fitted residuals needed in the proposed EDF tests. The second test procedure for testing the normality of random errors is suitable if the fitted model is correct. For both situations, we showed how to compute the test statistic and an approximate p-value based on the related Gaussian processes. Weak convergence properties of the related empirical processes were examined for both procedures. Finite sample performance of the suggested asymptotic results were examined by a Monte Carlo study. The tests we proposed based on the asymptotic theoretical results were found to work for realistic sample sizes. We also examined the power of the suggested test procedures in detecting the departures from normality. Simulation results show that using a sample of size 30 (10 dose levels with three replicates at each level) the powers of the tests for normality

based on the Cramér-von Mises statistic without assuming the fitted model is correct were around 0.60,0.15,0.10 ($\alpha = 0.05$) for gamma alternatives with shape parameters 0.5,2,4 respectively. For the test that assumes the fitted model is correct the corresponding powers were found to be around 0.84, 0.31 and 0.11.

Using the computing language FORTRAN, software was developed to implement the theoretical results suggested in this work. In each chapter, suggested theoretical results were illustrated using real data sets. Real data sets for illustration were chosen to cover the experimental designs (the additive dose method, partial bleach method and regeneration method) described in this work.

9.2 Some guidance for analyzing data collected in thermoluminescence studies

In this section, we provide some guidance for analyzing data collected in thermoluminescence studies, based on the results of our study.

9.2.1 Estimating the equivalent dose using data at a single temperature

For estimating the parameters using data collected at a single temperature we recommend the method of quasi likelihood for reasons listed below:

1. Quasi likelihood estimating equations are based on fewer assumptions about the unknown distribution of the data.

2. They perform almost as well as the maximum likelihood estimates in terms of their asymptotic statistical properties. Even though the asymptotic variances of quasi likelihood estimates are slightly larger than those of maximum likelihood provided errors are normally distributed, our small σ expansions to $\hat{\theta}$ indicate that both estimators have almost the same variance if σ is small, which certainly is true for thermoluminescence data.

3. Both quasi likelihood and maximum likelihood estimates are consistent. Furthermore, distributional approximations for quasi likelihood estimates are based only on the first two moments of the distribution for the random errors while for maximum likelihood distributional approximations require assumptions on the fist four moments.

4. The quasi likelihood estimates for the parameters in the mean functions are more robust in the sense that they do not depend on how we estimate the unknown relative error factor σ.

5. Programs for solving quasi likelihood estimating equations were found to converge much faster than those for maximum likelihood and generalized least squares.

9.2.2 Confidence intervals for the equivalent dose using data at a single temperature

For obtaining confidence intervals for the equivalent dose using data collected at a single temperature, we recommend using symmetric t intervals with appropriate degrees of freedom as discussed in this work. The advantages of using the suggested intervals are:

1. The symmetric t confidence intervals are computationally much simpler than profile type confidence intervals.

2. The symmetric t intervals were found have coverage probabilities close to their nominal levels in samples sizes of typical thermoluminescence data sets.

It is important to note that the above conclusions were drawn based on our experience with response functions considered in this work and for small measurement errors typical in thermoluminescence data sets. Care needs to be taken when extending these conclusions for cases with large σ; it is possible that the parameter effects of non-linearity of the response functions might then favour using likelihood based confidence intervals. In situations where the likelihood based intervals are preferred, we recommend using transformed F critical values as opposed to χ^2 critical values; the coverage probabilities based on approximate F critical values were found to be closer to their nominal levels than those using χ^2 critical values, in particular when the sample sizes are small.

9.2.3 Utilizing the data collected at several temperatures

The purpose of using the data collected over a series of temperatures is to utilize the extra information delivered by these added observations about the type of traps we are interested in. According to Aitken [1], a single type of trap corresponds to a glow peak of about 50^0C in width. Therefore, taking too wide a range of observations as belonging to the plateau, we increase the risk of confounding the information related to different types of traps. Furthermore, as we learned in Chapter 7, using more data over the plateau may not necessarily produce a more precise estimate. Therefore, we recommend not using observations that are not convincingly obvious as belonging to the plateau.

A few real data sets we analyzed have shown that over certain ranges of plateau temperatures, the standard errors of the estimate for the equivalent dose using separate analyses of single temperatures are quite large so that they are less useful in practice. However, using data at several such temperatures was found to produce estimates with much smaller standard errors. Therefore, the procedures for combining data collected at several such temperatures have certainly resulted a gain in the precision of the estimate.

The procedures we proposed for combining data at several temperatures need to be augmented by procedures for estimating the unknown correlation matrices. We recommend using the sample correlation matrices to estimate the unknown correlation matrices rather than using arbitrarily chosen fixed correlation matrices, unless there is evidence that such chosen fixed correlation matrices are good approximations to unknown true correlation matrices; the mean squared errors of the estimators using the former was found to be smaller than for those using arbitrarily chosen fixed correlation matrices.

9.2.4 Testing model assumptions

Examining the assumptions used in modeling the data is an essential component of any data analysis. In Chapter 8, we proposed two procedures for testing distributional assumptions about the random errors. We favour using the second test procedure where we assume that the fitted model is correct, in particular when we have few replicates at each dose level. Since estimates for σ computed from the residuals to the fit were found to be small for

thermoluminescence data, we have reason to believe that the fitted model is correct. Also, it is important to look at the plots of fitted residuals against the applied dose (see Figures 8.5 and 8.8) to see if there is any indication that the fitted model is not correct. Since data sets for thermoluminescence data are quite small and not many replicates are available at a given dose level, the second test procedure is more powerful in detecting departures from normality. However, we feel that it is always good practice to test the results also by using the procedure which does not assume the fitted model is correct. If this procedure does not reject the hypothesis that the errors are normally distributed but the other does, then this may be an indication that the fitted model is not correct. Based on the simulation results, we recommend using the tests based on the Anderson-Darling test statistic, in particular when the sample sizes are small. Procedures based on the Anderson-Darling statistic were found to be more powerful in detecting departures from normality than those based on the Cramér-von Mises statistic, while computational burden is the same for both procedures.

9.3 Further research

In this section we offer suggestions for further research.

In Chapter 2, we described physical models motivating generalized non-linear models. Under the assumptions listed in Chapter 2, the mean and variance functions of the total emission per unit mass of grains at temperature j from the ikth the sample (i.e. kth replicate sample receiving dose i) can be written as

$$E(TL_{ijk}) = E(N_{ijk})\frac{\lambda_{ij}}{m_{ik}}, \tag{9.1}$$

and

$$V(TL_{ijk}) = V(N_{ijk})\frac{\lambda_{ij}^2}{m_{ik}^2} + E(N_{ik})\frac{\lambda_{ij}}{m_{ik}^2}, \tag{9.2}$$

where N_{ijk} is the number of grains producing thermoluminescence, or emitting grains as they are usually called, in the ikth sample at temperature j. As suggested in Berger *et. al.* [11], ignoring the term $E(N_{ijk})\lambda_{ij}$ comparing with $V(N_{ijk})\lambda_{ij}^2$ (of the order $10^5 - 10^6$ vs $10^6 - 10^8$)) we discussed fitting the model $TL_{ijk} = f(x_{ik}, \theta_j)(1 + \sigma\epsilon_{ijk})$. We referred to

this model as Model 1. A natural extension of this model is therefore described by

$$E(TL_{ijk}) = f(x_{ik}, \theta_j), \tag{9.3}$$

and

$$V(TL_{ijk}) = \delta_1 E(TL_{ijk}) + \delta_2 E^2(TL_{ijk}), \tag{9.4}$$

where $\delta_1 = \frac{1}{m_{ik}}$ and $\delta_2 = \frac{V(N_{ik})}{E^2(N_{ik})}$. The mass of the sample m_{ik} is assumed to be known without error. Thus, δ_1 is known and the variance function for TL_{ijk} has the additional parameter δ_2 that needs to be estimated. In this work, we discussed fitting of Model 1. We referred to this extended model as Model 2. Model 1 is a simplified version of Model 2 that require the additional assumption that the term $E(N_{ijk})\lambda_{ij}$ is negligible compared to $V(N_{ijk})\lambda_{ij}^2$. Methodology for estimating additional parameters in the variance function are available in the literature (eg. pseudo likelihood methods; see Davidian and Carroll [22]). These methods can be used to fit Model 2. The fitting of Model 2 would provide room for assessing the additional assumption that leads to Model 1. Moreover, the estimate for δ_2 obtained by fitting Model 2 gives information about the coefficient of variation of the distribution for the number of emitting grains.

We also proposed a physical model for the total emission treating the mass of the emitting grains in the ikth sample, M_{ik}, as a random subsample of the total mass of the ikth sample m_{ik}. This leads to the model

$$E(TL_{ijk}) = \zeta f(x_{ik}, \theta_j), \tag{9.5}$$

and

$$V(TL_{ijk}) = \delta_1 E(TL_{ijk}) + \delta_2 E^2(TL_{ijk}), \tag{9.6}$$

where $\delta_1 = \frac{1}{m_{ik}}$ is a known quantity, and $\zeta = \frac{E(M_{ik})}{m_{ik}}$ and $\delta_2 = \frac{V(M_{ik})}{E^2(M_{ik})}$ are unknown parameters. We referred to this model as Model 3. In a given sample, only a few particles produce most of the thermoluminescence (See Huntley *et. al.* [36].). The mass of the emitting grains is unknown. An estimate for δ_2 gives information about the coefficient of variation on the emitting mass M_{ik}. In Model 3, we assume that the rate of emission per unit mass of emitting grains at temperature j, λ_{ij}, is related to the dose received by the sample x_i according to $\lambda_{ij} = f(x_i, \theta_j)$, where θ_j is the vector of parameters corresponding

to temperature j and f is the dose response function. If there is a large variation between the mass of emitting grains in different samples, this model might be more appropriate. We would like to investigate including the M_{ik}'s as explicit random effects in the model, noting that given the M_{ik}'s the model would be a non-linear Poisson regression. With data at several temperatures on the plateau, if M_{ik} depended little on temperature we might be able to estimate individual values and improve our fits.

In the partial bleach method described in Chapter 5, a portion of the sample is given an artificial bleaching. Huntley *et. al.* [60] suggest (see also Aitken [1]) that the method works well as long as the artificial bleaching is less severe than that caused by the unknown natural bleaching prior to the deposition (or the burial) of the sediments. It would be useful to investigate if there is any optimal design that would suggest the amount of bleaching (perhaps as a fraction of the thermoluminescence from the unbleached data) that would produce more precise estimates for the equivalent dose. We have not looked at this problem in this work.

In the regeneration method as discussed in this work two response curves are fitted, one for the unbleached data and the other for the bleached data. If a reliable estimate for the equivalent dose can be deduced from the given data then the two curves should be able to match by a shift along the dose axis (Huntley *et. al.* [38]). Huntley *et. al.* [38] suggests estimating the equivalent dose as the dose shift required to match the two curves. Aitken [1] describes estimating the equivalent dose from the bleached curve as the dose corresponding to the thermoluminescence produced by the natural sample (i.e. with no laboratory added dose). Since the dose shift required to match the two curves is the same at any dose level if the regeneration method works both procedures should provide equivalent results. From a statistical point of view the two methods can be distinguished as follows. The method as described in Huntley *et. al.* [38] has the advantage that it allows one to test the assumption that ***unbleached and bleached curves describe the same curve shifted along the dose axis.*** It may also allow one to estimate the model parameters used to describe the curves more precisely. On the other hand, as described in Aitken [1], if only the bleached curve is fitted and all the samples used to estimate the unbleached curve are instead used to estimate the

natural thermoluminescence then one can estimate the natural thermoluminescence more precisely. This would in turn allow to estimate the dose corresponding to the natural thermoluminescence more precisely. From an economic point of view the latter method has the advantage that it is less costly since it is not necessary to provide artificial laboratory doses for samples defining the unbleached curve. Furthermore, the estimation procedure is much simpler in the latter situation. Therefore, it is worth comparing the precisions of the estimates for the two situations: assuming all the unbleached samples are used only to measure the natural thermoluminescence as described in Aitken [1] and assuming unbleached samples are given different laboratory doses and are used to define the unbleached curve as described in Huntley *et. al.* [38]. Using all the samples for the unbleached curve only to estimate the natural thermoluminescence conceals the opportunity to examine whether the bleaching had caused any sensitivity change of the sample; perhaps only a few replicates could be used to define the unbleached curve while more of the unbleached discs could be used to estimate the natural thermoluminescence. This leads to the question that *can we find any optimal design that suggests how we should allocate the samples?*

As we mentioned earlier in this chapter, simulation studies have shown that profile likelihood intervals for the equivalent dose have lower coverage probabilities than their nominal levels when the sample sizes are small. This could possibly be due to the narrowing of the profile when nuisance parameters are estimated. Several authors (McCullagh and Tibshirani [50], Cox and Reid [20], Fraser and Reid [31]) have described adjustments to profile likelihood to account for estimating the nuisance parameters. It would be useful to examine the coverage probabilities of the profile likelihood intervals based on the modified profile likelihood as suggested by these authors.

The discrepancy between the observed coverages and nominal levels of confidence intervals based on the large sample normal approximations could possibly be due to non-linearity of the response functions. Bates and Watts [2] describe two components of non-linearity. The parameter effects non-linearity is that part of non-linearity which can be removed by a transformation of the parameters. The intrinsic non-linearity is that component of non-linearity which cannot be removed. Bates and Watts [2] define *curvature measures of*

non-linearity that quantify this non-linearities of the response functions. It would be useful to investigate the effects of non-linearity and to see if it is possible to reduce the parameter effects non-linearity by a suitable transformation of the parameters.

When investigating the biases and the mean squared errors of the estimators in finite samples we encountered computational difficulties when the relative error in a single measurement is large. A suitable transformation of the parameters might also improve the convergence of the algorithms.

In estimating the equivalent dose using the observations on the plateau, we assumed that the temperatures belonging to the plateau are already identified by some other method. It would be useful to investigate the problem of identifying the temperatures on the plateau. Taking more temperatures on the plateau should reduce the standard error of the estimate. However, if we incorrectly identify a temperature as belonging to the plateau using that temperature might increase the error (bias) of the estimate. The problem of making this trade off would be worth studying.

The observations collected at different temperatures along the plateau were found to be highly correlated. Estimating equations suggested for the parameters in the regression model were therefore augmented by a method for estimating the unknown correlation matrices. We examined two approaches: using Pearson residuals to estimate the correlation matrices and using arbitrarily chosen fixed correlation matrices. Using Pearson residuals was found to produce estimators with smaller mean squared errors than those using arbitrarily chosen fixed correlation matrices. Therefore, we favour using Pearson residuals to estimate the correlation matrices. However, our simulation results presented in Chapter 7 show that the coverage probabilities of the confidence intervals based on the estimates using Pearson residuals are lower than the nominal levels, when more than four temperatures on the plateau are used. As we described in Chapter 7, when the sample sizes are not large enough for the asymptotic theory to hold, our formula (Equation 7.6) underestimates the error of the estimate if we use Pearson residuals. We hope to pursue further work in this area to find a more accurate estimate for the error of the estimate when using Pearson residuals; this might bring the coverage probabilities of the resulting confidence intervals closer to their

nominal values. Also we hope to derive formulae for the biases of the estimators from data using several temperatures, following the same steps as for the single temperature case.

Appendix: Data sets

In Tables 9.1 and 9.2 we present the dose vectors used in the thesis. For ease of reference, we give a label (code number) to each dose vector. Each label has three characters: The first character indicates the method of data collection where we use P for the partial bleach method and R for the regeneration method. The second character is a serial number arbitrarily given to identify the dose vector. The third character indicates the level of bleaching where we use U for the unbleached data and B for the bleached data. When we refer to the complete data set (unbleached and bleached) we omit the last character. For example data set $P1$ refers to using $P1U$ as the unbleached data and $P1B$ as the bleached data. When we have used the same dose levels for the unbleached and bleached data we only report the unbleached dose levels and we indicate this by an equal sign in the first column. The second column indicates the sample size of each dose vector (n_1 for unbleached and n_2 for bleached). The third column indicates the dose levels.

Code	size	Dose values
$P1U$	$n_1 = 16$	0,0,0,0,120,120,120,240,240,240,480,480,480,480,960,960
$P1B$	$n_2 = 13$	0,0,0,120,120,240,240,240,480,480,480,960,960
$P2U$	$n_1 = 26$	0,0,0,0,120,120,120,200,200,200,240,240,240 320,320,320,480,480,480,480,840,880,880,960,960
$P2B$	$n_2 = 23$	0,0,0,120,120,240,240,240,320,320,320,480,480,480,840,840 880,880,880,960,960
$P3U$ $(= P3B)$	$n_1 = 50$	0,0,0,0,100,100,120,120,120,180,180,200,200,220,220,240,240,240 260,260,260,280,280,280,300,300,320,320,320,480,480,480,480,520 520,560,560,560,640,640,660,660,660,840,840,880,880,880,960,960
$P4U$ $(= P4B)$	$n_1 = 100$	0,0,0,0,100,100,120,120,120,180,180,200,200,220,220,240,240,240,260 260,260,280,280,280,300,300,320,320,320,340,340,340,400,400,400 420,420,420,440,440,480,480,480,480,520,520,560,560,560,580,580,580 600,600,600,620,620,620,640,640,660,660,660,700,700,700,720,720 720,740,740,740,760,760,760,780,780,780,800,800,800,820,820,820 840,840,880,880,880,900,900,900,920,920,920,940,940,940,960,960
$P5U$ $(= P5B)$	$n_1 = 125$	0,0,0,0,50,50, 100,100,120, 120,120,140,140,160,160,180, 180,200,200 220,220,240,240, 240,260,260,260,280,280,280,300,300,320,320,320 340,340, 340,360,360,380,380,380,400,400,400,420,420,420,440,440 460,460,460,480,480,480,480,500,500,520,520,540,540,560,560,560,580 580,580,600,600, 600,620,620,620,640,640,660,660,660,680,680,700 700,700,720,720,720,740,740,740,760,760,760,780,780,780,800,800,800 820,820,820,840,840,840,860,860,860,880,880,880,900,900,900 920,920,920,940,940, 940,960,960,960
$P6U$	$n_1 = 21$	0,0,0,0,0,0,200,200,200,400,400,400,600,600,600 1002,1002,1002,1500,1500,1500
$P6B$	$n_2 = 19$	0,0,0,0,0,0,200,200,200,400,400,400,600,600,1002,1002,1500,1500,1500

Table 9.1: Dose values used in the thesis: partial bleach method

Code	Sample size	Dose values
$R1A$	$n_1 = 30$	0,0,0,0,0,0,0,0,0,0,40,40,40,80,80,80,80,120,120,
		200,200,200,200,300,300,300,300,600,600,600
$R1B$	$n_2 = 32$	0,0,0,0,0,0,40,40,40,40,80,80,80,80,120,120,120,200,200,
		200,200,300,300,300,300,394,394,394,394,700,700,700
$R2A$	$n_1 = 40$	0,0,0,0,0,0,40,40,40,40,80,80,80,80,120,120,120,160,160,200,200
$(= R2B)$		200,200,300,300,300,300,394,394,394,394,500,500,600,600,650
		650,700,700,700
$R3A$	$n_1 = 50$	0,0,0,0,0,0,0,0,0,0,40,40,40,80,80,80,80,100,100,100
$(= R3B)$		100,120,120,140,140,140,180,180,180,200,200,200,200,300,300
		300,300,350,350,400,400,400,500,500,500,550,550,600,600,600
$R4A$	$n_1 = 100$	0,0,0,0,0,0,0,0,0,0,40,40,40,60,60,60,80,80,80,80,100,100,100,100
$(= R4B)$		120,120,140,140,140,160,160,160,180,180,180,200,200,200,200
		220,220,220,240,240,240,260,260,260,280,280,280,300,300,300
		300,320,320,320,340,340,340,360,360,360,380,380,380,400,400
		400,420,420,420,440,440,440,460,460,460,480,480,480,500,500
		500,520,520,520,540,540,540,560,560,560,580,580,580,600,600,600
$R5A$	$n_1 = 125$	0,0,0,0,0,0,0,0,0,0,20,20,20,30,30,30,40,40,40,50,50,50,60,60,60,70
$(= R5B)$		70,70,80,80,80,80,90,90,90,100,100,100,110,110,120,120,130,130,140
		140,140,150,150,150,160,160,160,170,170,180,180,180,190,190,200
		200,200,200,220,220,220,240,240,240,260,260,260,280,280,280,300
		300,300,300,320,320,320,340,340,340,360,360,360,380,380,380,400
		400,400,420,420,420,440,440,440,460,460,460,480,480,480,500,500
		500,520,520,520,540,540,540,560,560,560,580,580,580,600,600,600

Table 9.2: Dose values used in the thesis: regeneration method

n_0	Correlation matrix
2	$\begin{bmatrix} 1.00000 & 0.983042 \\ 0.983042 & 1.00000 \end{bmatrix}$
4	$\begin{bmatrix} 1.00000 & 0.983042 & 0.966130 & 0.955667 \\ 0.983042 & 1.00000 & 0.982168 & 0.960905 \\ 0.966130 & 0.982168 & 1.00000 & 0.982078 \\ 0.955667 & 0.960905 & 0.982078 & 1.00000 \end{bmatrix}$
6	$\begin{bmatrix} 1.000000 & 0.983042 & 0.966130 & 0.955667 & 0.887912 & 0.859539 \\ 0.983042 & 1.00000 & 0.982168 & 0.960905 & 0.900910 & 0.880512 \\ 0.966130 & 0.982168 & 1.00000 & 0.982078 & 0.909663 & 0.906823 \\ 0.955667 & 0.960905 & 0.982078 & 1.00000 & 0.942630 & 0.947949 \\ 0.887912 & 0.900910 & 0.909663 & 0.942630 & 1.00000 & 0.966454 \\ 0.859539 & 0.880512 & 0.906823 & 0.947949 & 0.966454 & 1.00000 \end{bmatrix}$

Table 9.3: Correlation matrices for the unbleached data

n_0	Correlation matrix
2	$\begin{bmatrix} 1.00000 & 0.962972 \\ 0.962972 & 1.00000 \end{bmatrix}$
4	$\begin{bmatrix} 1.000000 & 0.962972 & 0.958241 & 0.947413 \\ 0.962972 & 1.00000 & 0.986563 & 0.966392 \\ 0.958241 & 0.986563 & 1.00000 & 0.983300 \\ 0.947413 & 0.966392 & 0.983300 & 1.00000 \end{bmatrix}$
6	$\begin{bmatrix} 1.000000 & 0.962972 & 0.958241 & 0.947413 & 0.926114 & 0.801369 \\ 0.962972 & 1.00000 & 0.986563 & 0.966392 & 0.955948 & 0.866872 \\ 0.958241 & 0.986563 & 1.00000 & 0.983300 & 0.984046 & 0.902601 \\ 0.947413 & 0.966392 & 0.983300 & 1.00000 & 0.977326 & 0.891907 \\ 0.926114 & 0.955948 & 0.984046 & 0.977326 & 1.00000 & 0.913664 \\ 0.801369 & 0.866872 & 0.902601 & 0.891907 & 0.913664 & 1.00000 \end{bmatrix}$

Table 9.4: Correlation matrices for the bleached data

Appendix: Supplement to chapter 3

Consider the model, $y = f(x, \theta_0)(1 + \sigma\epsilon), i = 1, 2, ..., n$, where θ_0 is the vector of unknown true parameter values. Let $\hat{\theta}$ be an estimator for θ_0. Suppose for small σ, $\hat{\theta} = \theta_0 + C_1\sigma + C_2\sigma^2$, where C_1 and C_2 depend on $f(x, \theta_0)$ and ϵ, but not on σ. We employ f_0 and $f_{\hat{\theta}}$ to indicate $f(x, \hat{\theta})$ and $f(x, \theta_0)$ respectively. The gradient vector and the matrix of second derivatives (i.e. the Hessian matrix) evaluated at the true parameter values are respectively denoted by ∇f_0 and H_0.

In chapter 3, we examined four methods of estimation: maximum likelihood (ML), quasi likelihood (QL), generalized least squares (GLS) and data weighted least squares (DWLS). Each method of estimation resulted the same formula for C_1, which is

$$C_1 = \left[\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1} \sum_{i=1}^{n} \epsilon_i \left(\frac{\nabla f_0}{f_0}\right).$$

Under the assumption that the random error ϵ_i are independent mean zero random variables we find $E(C_1) = 0$ and the variance covariance matrix of C_1, Σ, is given by

$$Var(C_1) = \left[\sum_{i=1}^{n} \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1}.$$

With the above notation we prove the following results.

Result 1

$$E\left(C_1^T \left(\frac{\nabla f_0}{f_0}\right) \left(\frac{\nabla f_0}{f_0}\right)^T C_1\right) = E\left[tr\left(C_1^T \left(\frac{\nabla f_0}{f_0}\right) \left(\frac{\nabla f_0}{f_0}\right)^T C_1\right)\right]$$

$$
\begin{aligned}
&= E\left[tr\left(\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T C_1 C_1^T\right)\right] \\
&= tr\left[E\left(\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T C_1 C_1^T\right)\right] \\
&= tr\left[\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T E(C_1 C_1^T)\right] \\
&= tr\left[\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T \Sigma\right].
\end{aligned}
$$

Result 2 *In result 1 replacing the matrix* $\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T$ *by the matrix* $\left(\frac{H_0}{f_0}\right)_i$,

$$
E\left(C_1^T\left(\frac{H_0}{f_0}\right)_i C_1\right) = tr\left[\left(\frac{H_0}{f_0}\right)_i \Sigma\right].
$$

Result 3 *Since the term* $C_1^T\left(\frac{\nabla f_0}{f_0}\right)$ *is a scalar, we find*

$$
\begin{aligned}
E\left(C_1^T\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{H_0}{f_0}\right)_i C_1\right) &= E\left(\left(\frac{H_0}{f_0}\right)_i C_1 C_1^T\left(\frac{\nabla f_0}{f_0}\right)\right) \\
&= \left(\frac{H_0}{f_0}\right)_i E(C_1 C_1^T)\left(\frac{\nabla f_0}{f_0}\right) \\
&= \left(\frac{H_0}{f_0}\right)_i \Sigma\left(\frac{\nabla f_0}{f_0}\right).
\end{aligned}
$$

Result 4 *Since the term* $\left(\frac{\nabla f_0}{f_0}\right)^T \Sigma\left(\frac{\nabla f_0}{f_0}\right)$ *is a scalar, we find*

$$
\begin{aligned}
\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T \Sigma\left(\frac{\nabla f_0}{f_0}\right) &= \left(\frac{\nabla f_0}{f_0}\right)^T \Sigma\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right) \\
&= tr\left[\left(\frac{\nabla f_0}{f_0}\right)^T \Sigma\left(\frac{\nabla f_0}{f_0}\right)\right]\left(\frac{\nabla f_0}{f_0}\right) \\
&= tr\left[\left(\frac{\nabla f_0}{f_0}\right)\left(\frac{\nabla f_0}{f_0}\right)^T \Sigma\right]\left(\frac{\nabla f_0}{f_0}\right).
\end{aligned}
$$

Result 5 *If the random errors are independent and follow a distribution which is symmetric about zero (in particular if* ϵ_i *are iid* $N(0,1)$*), then*

$$
\begin{aligned}
E(C_1\epsilon_i) &= Cov(C_1, \epsilon_i), \\
&= Cov\left(\left[\sum_{i=1}^n \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1} \sum_{k=1}^n \epsilon_k \frac{(\nabla f_0)_k}{(f_0)_k}, \epsilon_i\right) \\
&= \left[\sum_{i=1}^n \left(\frac{\nabla f_0}{f_0}\right)_i \left(\frac{\nabla f_0}{f_0}\right)_i^T\right]^{-1} \left(\frac{\nabla f_0}{f_0}\right)_i.
\end{aligned}
$$

Bibliography

[1] Aitken, M.J. (1985) *Thermoluminescence Dating.* London: Academic Press.

[2] Bates D.M., and Watts, D.G. (1988) *Nonlinear regression analysis and its applications.* New York: John Wiley & Sons Inc.

[3] Becker, R.A, Chambers, J.M., and Wilks, A.R. (1988) *The New S Language.* New Jersey: Bell laboratories, Inc.

[4] Beckman, R. J. and Trussel, H. J. (1974). The distribution of an arbitrary studentized residual and the effects of updating in multiple regression. *J. Amer. Stat. Assoc.*, 69, 199-201.

[5] Berger G.W., Huntley, D. J. and Stipp J. J. (1984). Thermoluminescence studies on a C^{14}-dated marine core. *Can. J. Earth Sci.*, 21, 1145-1150.

[6] Berger, G.W. (1984). Thermoluminescence dating studies of glacial silts from Ontario. *Can. J. Earth Sci.*, 21, 1393-1399.

[7] Berger, G.W. (1985). Thermoluminescence dating studies of rapidly deposited silts from south-central British Columbia. *Can. J. Earth Sci.*, 22, 704-710.

[8] Berger, G.W. (1985). Thermoluminescence dating applied to a thin winter varve of the late glacial South Thompson silt, south-central British Columbia. *Can. J. Earth Sci.*, 22, 1736-1739.

[9] Berger, G.W. (1987). Thermoluminescence dating of the Pleistocene Old Crow tephra and adjacent loess, near Fairbanks, Alaska. *Can J. Earth Sci.*, 24, 1975-1984.

[10] Berger, G.W., Clague, J.J. and Huntley, D.J. (1987). Thermoluminescence dating applied to glaciolacustrine sediments from central British Columbia *Can. J. Earth Sci.*, 24, 425-434.

[11] Berger, G.W., Lockhart, R.A. and Kuo, J. (1987). Regression and error analysis applied to the dose response curves in thermoluminescence dating. *Nucl. Tracks Radiat. Meas.*, 13, 177-184.

[12] Berger, G.W. and Huntley, D.J. (1989). Test data for exponential fits. *Ancient TL*, 7, No. 3, 43-46.

[13] Billingsley, P. (1968). *Convergence of Probability Measures.* NewYork: John Wiley & Sons Inc.

[14] Box, M.J.(1971). Bias in nonlinear estimation. *J. Roy. Statist. Soc.*, B, 33, 171-201.

[15] Chatterjee, S. and Hadi, A.S. (1988). *Sensitivity Analysis in Linear Regression.* NewYork: John Wiley & Sons Inc.

[16] Chen, G. (1991). *Empirical processes based on regression residuals: Theory and applications.* Ph. D. Thesis, Simon Fraser University.

[17] Cook, R.D., Tsai, C.L. and Wei, B.C.(1986). Bias in nonlinear regression. *Biometrika*, 73, 615-623.

[18] Cook, R.D. and Sanford, W. (1990). Confidence curves in nonlinear regression. *J. Amer. Stat. Assoc.*, 85, No. 410, 544-551.

[19] Cook, R.D. and Tsai, C.L.(1990). Diagnostics for assessing the accuracy of normal approximations in exponential family nonlinear models. *J. Amer. Stat. Assoc.*, 85, No. 411, 770-777.

[20] Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *J. Roy. Stat. Assoc.*, B, 49, No 1, 1-39.

[21] Crowder, M. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika*, 82, 407-410.

[22] Davidian, M and Carroll, R.J (1987). Variance Function Estimation. *J. Amer. Stat. Assoc.*, 82, 1079-1091.

[23] Davidian, M and Carroll, R. J.(1988). A note on extended quasilikelihood. J. Roy. Statist. Soc. , B, 50, 74-82.

[24] Dean, C.B. (1991). Estimating equations for mixed poisson models. In *Estimating Functions* Edited by V.P. Godambe, New York: Oxford University Press.

[25] Diggle, P. J., Liang K.Y. and Zeger, S.L. (1994). *Analysis of longitudinal data*, New York: Oxford University Press.

[26] Divigalpitiya, W.M.R. (1978). *Thermoluminescence dating of sediments.* M.Sc. Thesis, Simon Fraser University.

[27] Draper, N.R. and Smith H. (1981). *Applied regression analysis*, 2nd edition, New York: John Wiley & Sons, Inc.

[28] Dunnett, C. W. and Sobel, M. (1954). A bivariate generalization of student's *t*-distribution, with tables for certain special cases. *Biometrika*, 41, 153-169.

[29] Ellenberg, J.H. (1973). The joint distribution of the standardized least squares residuals from a general linear regression. *J. Amer. Stat. Assoc.*, 68, 941-943.

[30] Franklin, A. D. and Hornyak, W. F. (1992). Normalization of inclusion size quartz ¯L data. *Ancient TL*, 10, No. 1, 1-6.

[31] Fraser, D.A.S. and Reid, N. (1989). Adjustments to profile likelihood. *Biometrika*, 76, 477-488.

[32] Graybill, F.A., (1961). *An introduction to linear statistical models.* New York: McGraw-Hill Book Company, Inc.

[33] Green, P.J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *J. Roy. Statist. Soc.*, B, 46, 149-192.

[34] Halbert, W. (1984). *Asymptotic theory for econometricians.* London: Academic Press Inc.

[35] Huntley, D.J., and Johnson, H.P. (1976). Thermoluminescence as a potential means of dating siliceous ocean sediments. *Can. J. Earth Sci.*, 13, 593-596.

[36] Huntley, D.J., and Kirkey, J.J. (1985). The use of an image intensifier to study the TL intensity variability of individual grains. *Ancient TL*, 3, No. 2, 1-4.

[37] Huntley D.J., Berger G.W. and Bowman, S.G.E. (1988). Thermoluminescence responses to alpha and beta irradiations, and age determinations when the high dose response is nonlinear. *Radiat. Effects*, 105, 279-284.

[38] Huntley, D.J., Hutton, J.T. and Prescott, J.R. (1993). The stranded beach-dune sequence of South-East South Australia: A test of thermoluminescence dating, 0-800 ka. *Quaternary Science Reviews.*, 12, 1-20.

[39] Imhof, J.P.(1961). Computing the distribution of the quadratic forms in normal variables. *Biometrika*, 48, 419-426.

[40] Jennrich, R.I. (1969). Asymptotic properties of non-linear least squares estimators. *The Ann. Math. Statist.*, 40, 633-643.

[41] Kuo, J. (1981). *Statistical analysis of the partial bleaching method of thermoluminescence dating of sedimentary rock.* M.Sc. Thesis, Simon Fraser University.

[42] Kirkey, J. J. (1983). *On the reliability of thermoluminescence dating applied to beach dunes.* M.Sc. Thesis, Simon Fraser University.

[43] Liang, K. Y and Zeger, S. L.(1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

[44] Lockhart, R.A., O'Reilly F.J. and Stephens, M.A. (1986). Tests of fit based on normalised spacings. *J. Roy. Statist. Soc.*, B, 48, 344-352.

[45] Loynes, R.M. (1980). The empirical distribution function of residuals from generalized regression. *Ann. Statist.*, 8, 285-298.

[46] Meester, S.G. and Lockhart, R.A. (1988). Testing for normal errors in designs with many blocks. *Biometrika*, 75, 569-575.

[47] Meester, S.G. (1984). *Testing for normally distributed errors in block design experiments.* M.Sc. Thesis, Simon Fraser University.

[48] McCullagh, P. (1983). Quasi-likelihood functions. *Ann. Statist.*, 11, 59-67.

[49] McCullagh, P. and Nelder J. A. (1983). *Generalized Linear Models.* University Press, Cambridge: Chapman and Hall.

[50] McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustments of profile likelihoods. *Roy. Statist. Soc*, B, 52, 325-344.

[51] Rao, C.R. (1967). *Linear Statistical Inference and Its Applications.* New York: John Wiley & Sons Inc.

[52] Ratkowsky, D.A. (1983). *Nonlinear Regression Modelling.* New York: Marcel Dekker Inc.

[53] Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, 2, 110-114.

[54] Satterthwaite, F.E. (1941). *Psychometrika*, 6, 309-316.

[55] Seber, G. A. F. and Wild, C.J. (1989). *Nonlinear Regression.* NewYork: John Wiley & Sons.

[56] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* NewYork: John Wiley & Sons.

[57] Stephens, M. A. (1976). Asymptotic results for goodness of fit statistics with unknown parameters. *Ann. Statist.*, 4, 357-369.

[58] Stephens, M. A.(1986). Chapter 4 in *Goodness of fit techniques.* Edited by D'Agostino, R. B. and Stephens, M. A. New York: Marcel Dekker Inc.

[59] Wedderburn, R.W.M. (1974). Quasilikelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*, 61, 439-447.

[60] Wintle A.G. and Huntley, D.J. (1980). Thermoluminescence dating of ocean sediments. *Can. J. Earth Sci.*, 17, 348-360.