HERBST, ALIDA

THE STATISTICAL MODELLING OF GROWTH

PhD UP 1994

THE STATISTICAL MODELLING OF GROWTH

BY

ALIDA HERBST

Submitted in accordance with the requirements for the of Doctor of Philosophy in the subject Mathematical Statistics at the University of Pretoria.

Promotor: S.H.C. du Toit

Date submitted: July 1993

TABLE OF CONTENTS

CHAPTER

SUMMARY

In many fields of application, such as biology, psychology, agriculture, geology, botany, engineering and medicine, experiments are conducted in which a number of responses are repeatedly measured on each of a number of experimental units under differing experimental conditions. *Longitudinal data* which consists of observations that are ordered by time or position in space, for example the height of a child measured annually or monthly over a period of time, is considered.

The study of growth in height is an excellent model for the investigation of other forms of growth and is considered in the practical applications of this thesis. Any progress made in measuring and modelling physical growth will serve as a good basis when attempts are made in future on the more difficult task of describing cognitive, affective or social development. The Richards growth function, a generalisation of nonlinear functions with a flexible point of inflection, often used to describe and compare growth curves, is considered.

The generalised least squares, the maximum likelihood and the asymptotic distribution free frequentist estimation procedures for *linear* and *nonlinear* random parameter models are discussed. Two algorithms namely the Fisher scoring algorithm and the Expected Maximization (EM) algorithm are discussed. The Gauss quadrature numerical integration technique, which usually provide reliable approximations when closed form solutions for integrals are not available, is considered.

The Bayes and Maximum Aposteriori (MAP) estimators are discussed, for linear and nonlinear models. An empirical Bayes method for the estimation of unknown model parameters is applied to an incomparable collection of longitudinal human growth records begun at the Fels institute in 1929, as well as to the Berkeley human growth records (see Tuddenham and Snyder, 1954).

The *nonlinear* fixed and random parameter Richards models with time series deviations ($ARMA(1,1)$), for non-consecutive data, are considered and applied to different datasets. Most of the theory discussed has been implemented in computer programs.

OPSOMMING

In verskeie toepassingsvelde soos byvoorbeeld biologie, sielkunde, landbou, geologie, plantkunde, ingenieurswese en medies word eksperimente uitgevoer wat gebaseer is op metings wat onder verkillende omstandighede waargeneem is vir eksperimentele eenhede. Data wat hoofsaaklik oorweeg is, bestaan uit waarnemings wat georden is in tyd soos byvoorbeeld die lengte van 'n kind wat maandeliks of jaarliks waargeneem is.

Die studie van groei is 'n uitstekende model vir die ondersoek na ander vorme van groei en word hier ondersoek. Enige vooruitgang wat gemaak word in die meting en modellering van fisiese groei kan as 'n goeie basis dien wanneer daar in die toekoms gepoog word om kognitiewe, effektiewe of sosiale ontwikkeling te ontleed. Die Richards groeikrommes, wat 'n veralgemening van nie-lineêre funksies met 'n buigbare punt van infleksie is, is gebruik om groei te beskryf en te vergelyk.

Die algemene kleinste kwadrate, maksimum aanneemlikheid en verdelingsvrye prosedures vir die beraming van die onbekende parameters vir lineêre en nie-linieêre modelle word bespreek. Die Fisher en "Expected Maximization" (EM) algoritmes word ook beskou. Die "Gauss quadrature" numeriese integrasie tegniek, wat goeie benaderings verskaf indien integrale nie 'n geslote oplossing besit nie, word bespreek en toegepas.

Die Bayes and "Maximum Aposteriori" MAP beramers word bespreek, vir lineêre en nie-lineêre modelle en toegepas vir verskeie datastelle. 'n Empiriese Bayes metode vir die beraming van onbekende model parameters, word toegepas op die versameling metings op menslike groei data wat in 1929 by die Fels instituut begin is, sowel as op die Berkeley data (sien Tuddenham and Snyder, 1954).

Die nie-lineêre vaste en stogastiese parameter Richards modelle met tydafwykings ($ARMA(1,1)$) vir nie-opeenvolgende data word beskou en toegepas. Meeste van die teorie bespreek in hierdie studie is geimplementeer in rekenaar programme.

SIGNED STATEMENT

I declare that: "Statistical modelling of Growth" is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

(ALIDA HERBST)

ACKNOWLEDGEMENTS

I am deeply indebted to Professor S.H.C. du Toit for suggesting the topics of this research and for his guidence, constructive comments and encouragement. I am also grateful for the 486 computer of Professor du Toit, that was put at my disposal to use for the simmulation studies and practical applications.

Professor du Toit's unselfish attention during crucial stages of the research programme deserves the highest merit of praise and I sincerely appreciate his invaluable advice in these moments.

I thank Professor Crowther for his permission to make use of the facilities at the Pretoria University and gratefully acknowledge the financial support of the Vista University.

Finally, I should like to acknowledge the unfailing support and patience of my husband and children during the trying period of this research as well as the continuous encouragement of my parents.

Notation

The following notation shall be adopted:

π	: constant, $\pi = 3.14159\cdots$
e	: Euler's constant, $e = 2.71828\cdots$
$exp(x)$	: $e^x, -\infty < x < \infty$
$ln\, x$	: natural logarithm of the real number $x, x \geq 0$
δ_{ij}	: Kronecker's delta (=1 if $i=j$ and =0 if $i \neq j$)
$\boldsymbol{A}:(p \times q)$	: matrix of order $p \times q$
$\mathbf{a}:(p \times 1)$	: column vector of order $p \times 1$
a	: scalar
$\boldsymbol{A}'$	: transpose of $\boldsymbol{A}$
$\mathbf{a}'$	: transpose of $\mathbf{a}$(a row vector)
a_{ij} or $[\boldsymbol{A}]_{ij}$	: the element in the i-th row and j-th column of $\boldsymbol{A}$
a_i or $[\mathbf{a}]_{i1}$	: the i-th element of $\mathbf{a}$
$\boldsymbol{A}^{-1}$	: the inverse of $\boldsymbol{A}$
a^{ij}	: $[\boldsymbol{A}^{-1}]_{ij}$
$\lvert \boldsymbol{A} \rvert$	: determinant of $\boldsymbol{A}$
$tr\, \boldsymbol{A}$	: trace of $\boldsymbol{A}$
$\boldsymbol{D}_a$	: diagonal matrix with diagonal elements $a_{11}, a_{22}, \cdots$
vec[$\boldsymbol{A}$]	: $(pq \times 1)$ vector formed from the q columns of the $p \times q$ matrix $\boldsymbol{A}$
vecs[$\boldsymbol{A}$]	: $(p(p+1)/2 \times 1)$ vector formed from the nonduplicate elements of the $(p \times p)$ symmetric matrix $\boldsymbol{A}$
$\mathbf{0}$	: null matrix, $[\mathbf{0}]_{ij}$
$\boldsymbol{J}_{ij}$	: matrix with all elements equal to zero with the exception of the element in the i-th row and j-th column which is equal to unity
$\boldsymbol{J}_{1j}$	: column vector with all elements equal to zero with the exception of the i-th elements which is equal to unity
$\boldsymbol{A} \otimes \boldsymbol{B}$	: The right direct product or "Kronecker product" of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ defined by:

CHAPTER 1 INTRODUCTION.

Longitudinal data, namely observations that are ordered by time or position in space, can be found in many fields of application, such as biology, psychology, agriculture, botany, engineering and medicine. One such an example, used frequently in this dissertation, is the height of a child measured annually or monthly over a period of time.

Repeated measurements generated by linear or polynomial "growth" curve models enjoy a wide range of applications (e.g., Laird and Ware, 1982; Stiratelli, Laird, and Ware, 1984; Laing and Zeger, 1986; and Zeger, Liang and Albert, 1988). Lindley and Smith (1972) proposed and analyzed a general Bayesian linear model and their theory was applied to the estimation problem for growth curves by Fearn (1975). Instead of considering only the overall growth curve, Fearn assumed that there is a separate growth curve for each individual and that the observations on an individual are independently and normally distributed about the curve for that individual. Rosenberg (1973) considered models similar to Fearn (1975) in a more general context, using maximum likelihood and empirical Bayes techniques.

By fitting polynomial "growth" curves to repeated measurements using any number of multivariate analysis techniques two goals can be achieved (Vonesh & Carter 1992) : (i) changes in an individual's response as time or conditions vary can be described; (ii) mean responses over time among several groups of individuals can be compared. For balanced and complete data, a generalised multivariate analysis of variance model can be used to fit and compare "growth" curves without assuming any special covariance structure on the repeated measurements (Potthoff and Roy, 1964; Grizzle and Allen, 1969). For unbalanced data, random-effects models are often used in conjunction with iterative (Laird and Ware, 1982; Jennrich and Schluchter, 1986) or noniterative (Vonesh and Carter, 1987) techniques to estimate and compare population "growth" curves. These method and others, including those which entail an autoregressive error structure, are discussed by Ware (1985) in an overview of linear models for repeated measures.

In both balanced and unbalanced repeated measures data, nonlinear random-effects models have been utilized in growth studies to describe the mean response function as well as within- and between-subject variability (Scheiner and Beal, 1980; Berkey, 1982; Racine-Poon, 1985; Steimer, Mallet, and Mentrè , 1985). These models are similar to random coefficient growth curve models in that: (i) a nonlinear regression model is assumed for each individual; and, (ii) a stochastic structure is imposed on the parameters of the individual nonlinear response functions. Various procedures have been proposed for estimating the population parameters; chief among them are the NONMEM method of Sheiner and Beal (1979), the GTS (global two-stage) and ITS (iterated two-stage) methods of Steimer, et al. (1984), and the Bayesian-based EM algorithm method of Racine-Poon (1985).

The purpose of this thesis is to investigate various models for the statistical modelling of growth. Since the thesis is meant to be self contained, Sections 2.2, 2.3, 3.2 to 3.6, and 4.2 to 4.4 are primarily reviews of relevant existing results. The study of growth in height is an excellent model for the investigation of other forms of growth and is considered in the practical applications of this thesis. The goals in the analysis of height data are varied (Ramsay, Altman and Bock (1992)): (i) to offer a satisfactory model for individual growth curves; (ii) to describe particular features of growth curves, such as the onset and duration of the pubertal growth spurt; (iii) to define the normal range of variation in growth curves; and, (iv) to identify unusual growth patterns. Any progress made in measuring and modelling physical growth will serve as a good basis when attempts are made in future on the more difficult task of describing cognitive, affective or social development.

In Chapter 2 the Richards growth function, a generalization of nonlinear functions with a flexible point of inflection, often used to describe and compare growth curves, is considered. The change in growth over time can often be described more accurately if a linear combination of two or more Richards functions is used. The properties of multi-component Richards functions are also discussed in this chapter.

There is a continuous controversy between the Frequentist and Bayesian schools of thought about the foundations of statistical inference. Ideas from both sides are useful in thinking about practical problems.

Frequentist estimation procedures such as Maximum likelihood (ML), Marginal Maximum Likelihood (MML) and asymptotic distribution free (ADF) procedures are discussed in Chapter 3 for models with linear or nonlinear response functions. The Fisher scoring procedure, often used when estimating the unknown parameters with maximum likelihood or ADF procedures, is also given. The EM algorithm consisting of an E (expectation) and an M (maximization) step when estimating unknown parameters is also discussed in this chapter. The Gauss quadrature technique often used when integrals cannot be solved in closed form is introduced and an example is given. The random parameter model is extended to incorporate so-called second-stage covariates and examples using data from real life, are given in the Application Section.

In Chapter 4 the procedures for calculating the Bayes and MAP estimators are given. References to the statistical properties of the Bayes estimator may be found in, for example, Lindley, 1971; Lindley and Smith, 1972; Novickand Jackson, 1974 and DeGroot, 1975. Chapter 4 concludes with a practical application section.

In the modelling of human growth it was found that height from birth to maturity can be adequately described by a triple logistic growth function. Multi-component models of which the triple logistic is a special case, is discussed in Chapter 5. Approximate expressions for the first and second order moments of a triple component random parameter Richards model is also given. The exact likelihood function for this model as well as the derivatives of the log-likelihood function is derived. An Empirical Bayes method for the estimation of the unknown parameters using the MML estimation procedure, is also given.

In the Application Section of Chapter 5, use was made of a collection of longitudinal growth records begun at the Fels institute in 1929. This superb collection of data was obtained from Professor Bock at Chicago university, who obtained access to the data through Dr. Alex Roche (see Bock 1989). The Fels data include 158 males and 132 female records from birth to maturity. These records include the child's sex, height of first-degree relatives, especially parents and estimates of the skeletal age of the child based on hand-wrist of knee radiographs. A similar analysis was also done by fitting the triple-logistic model of Bock and Thissen (1976) to male and female data from the Berkeley Guidance study (Tuddenham and Snyder, 1984).

In Chapter 6 the nonlinear fixed and random parameter Richards models with $ARMA(1,1)$ (Auto regressive moving average) deviations, for non-consecutive data is considered followed by a number of applications.

Most of the theory discussed in this study has been implemented in FORTRAN computer programs. Raw datasets are given in Appendix A.

CHAPTER 2 FLEXIBLE GROWTH FUNCTIONS

2.1 INTRODUCTION

The term growth curve is usually applied to the situation in which we have data on the change, over a period of time, of some measurement on each member of a group, or groups, of individuals. Typically a mathematical function is chosen to describe as closely as possible the trajectory of growth over time.

There are many possible growth curves to select from, examples being the polynomial, exponential, hyperbolic, and logistic functions (see e.g., Gallant 1987). The Richards function (Richards, 1959), a generalization of nonlinear functions with a flexible point of inflection, is often used to describe and compare growth curves (Koops 1987). In this chapter the discussion is limited to a general growth function which consists of a combination of so-called Richards growth functions (see e.g. Richards 1959; Du Toit 1979).

The general Richards response function is defined by Du Toit (1979) as follows:

$$f(\mathbf{a},\alpha,t_i) = a\left(1 + s b_{(s)} c^{t_i}\right)^{\alpha} \qquad t_i \geq 0, \qquad i=1,2,\cdots,n; \qquad (2.1.1)$$

where

$$\mathbf{a} = (a,\ b_{(s)}, c)', \qquad s=1,-1. \qquad (2.1.2)$$

The parameter α is a fixed parameter that may be estimated or set to a predetermined value, and $\mathbf{a}$ is a vector of fixed or random parameters.

The single component Richards function is discussed in Section 2.2. Three well known Richards functions, the monomolecular curve (α=1 in (2.1.1)), the logistic curve ($\alpha=-1$ in (2.1.1)) and the Gompertz curve ($|\alpha|\rightarrow\infty$ in (2.1.1)) are discussed in Section 2.3.

In the analysis of the random parameter model with response function (2.1.1), it is difficult to give distributional assumptions for the parameter set **a** (cf. 2.1.2), apart from the fact that the ranges of the random variables are restricted. Box, Davies and Swann (1969, p42) pointed out that before considering the constrained optimization problem in its most general form, the problem of constraints on parameters can be most effectively handled by a transformation of each of the parameters $[\mathbf{a}]_{i1}$, i=1,2,3; to a new set of unbounded parameters. If we allow for the transformed parameter set to be random, the usual assumption of multivariate normality becomes more realistic.

Reparameterization of the original parameters is discussed in Section 2.3. One possible set of transformations is the following:

$$a = x_1 ,$$
$$b_{(1)} = exp\, x_2 ,$$

and

$$c = \frac{1}{1 + exp\, x_3} .$$

In this case, the reparameterized Richards growth curve may be written as

$$f(\mathbf{x},\alpha,t_i) = x_1\left(1 + exp\, x_2(1 + exp\, x_3)^{-t_i}\right)^{\alpha} \qquad t_i \geq 0, \qquad i=1,2,\cdots,n. \tag{2.1.6}$$

All of the functions described in Section 2.2 represent but a single curve whose only change in form may be realized by different values of α. Many growth curves fail to conform satisfactorily with any *one* of the functions and it is subsequently necessary to combine perhaps 2 to 3 functions, and enlarge the number of parameters to be estimated. Illustrative examples of combinations of two or more component Richards functions are given in Section 2.4.

2.2 THE RICHARDS FAMILY OF GROWTH CURVES

A flexible response function, which is similar to the Richards (1959) family of growth curves as well as to the general function given by Gregg, Hossel and Richardson (1964), is defined in (2.1.1) and (2.1.2).

If the response function *increases* monotonically in t_i, then s$=-1$ for $\alpha \geq 0$ and s$=+1$

for $\alpha < 0$. Opposite signs are allocated when the reponse function *decreases* monotonically in t_i. The parameter α is a fixed parameter that is usually pre-assigned while the elements of the vector **a** are fixed or random parameters that has to be estimated.

The following linear inequality constraints are imposed upon the elements of **a** :

$$a \geq 0\,,\quad b_{(1)} \geq 0;\quad 0 \leq b_{(-1)} \leq 1 \text{ and } 0 \leq c \leq 1. \qquad (2.2.1)$$

Note that the constraint : $0 \leq b_{(-1)} \leq 1$ is imposed to avoid the occurence of complex roots for the case s= -1 in the expression $(1 - b_{(-1)}c^{t_i})^{\alpha}$.

Since the parameters in (2.1.1) have definite physical meanings, a curve from this family is preferred to a polynomial curve, which may often be fitted to a set of responses with the same degree of accuracy. The parameter a represents the time asymptotic value of the characteristic which has been measured, the parameter b represents the potential increase (or decrease) in the value of the function $f(\mathbf{a},\alpha,t_i)$ during the course of time t_1 to t_p, and the parameter c characterizes the rate of growth. Apart from the degree of compression, the differing shapes of the curves are due solely to differences in the parameter α. Examples of curves for various values of α are given in Figures 2.2.1 to 2.2.4.

Three well-known growth functions may be obtained as special cases of (2.1.1), these being the monomolecular or simple modified exponential, the logistic or autocatalytic and the Gompertz. A brief discussion of each case will be given and the reader is referred to Richards (1959) for a more detailed discussion.

Monomolecular curve: α=1

$$f(\mathbf{a},1,t_i)= a(1 - b\ e^{-c^{t_i}})\ . \qquad (2.2.2)$$

This function has no point of inflection, and its growth rate declines linearly with increasing function values. This function is sometimes used to present later portions in life history. Examples of monomolecular curves are given in Figures 2.2.1 and 2.2.2.

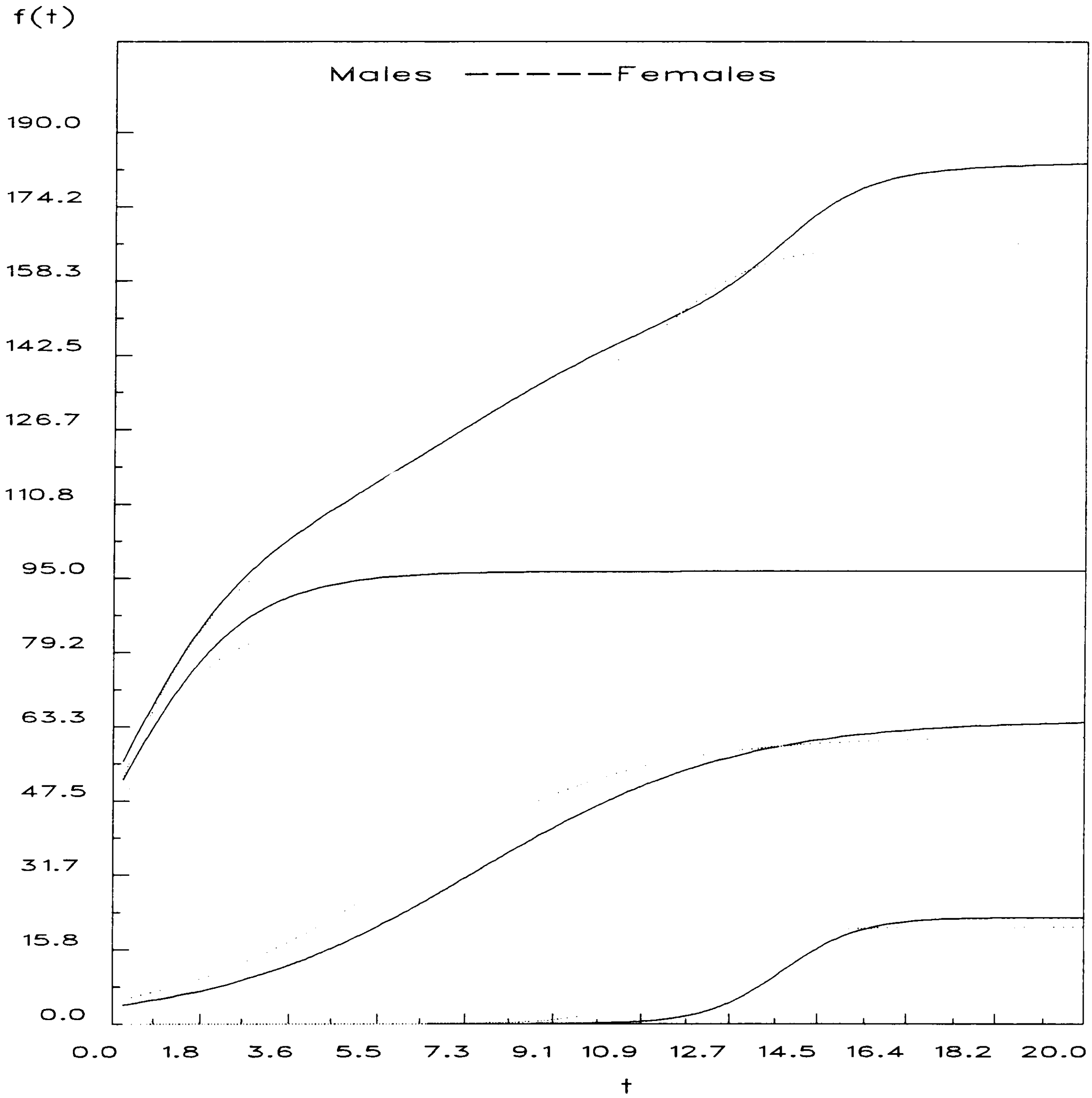

Figure 2.4.1 Graph of triple logistic growth functions:

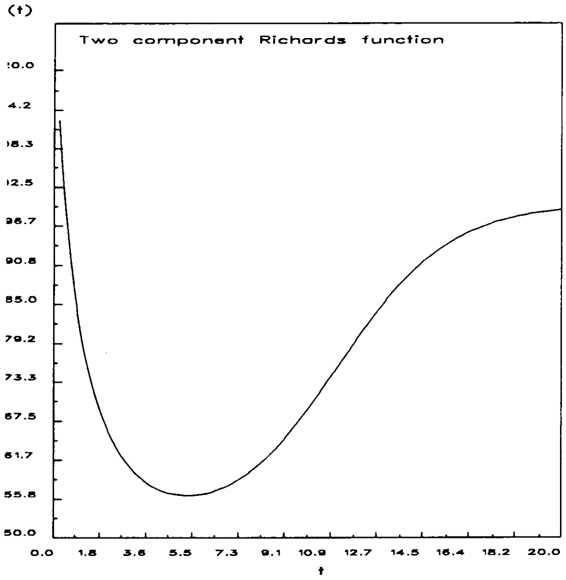

Figure 2.4.2

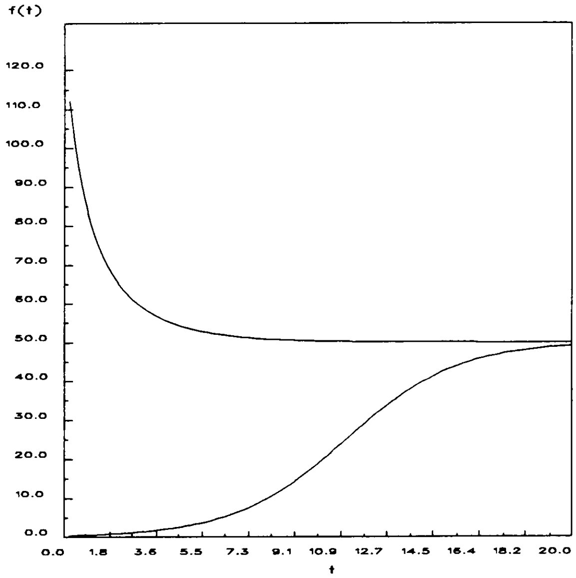

Figure 2.4.3

$$f(\mathbf{x},\alpha,t_i)=50(1-e^{-0.5+0.45})^{-1}+50(1+e^{5+0.45})^{-1}$$

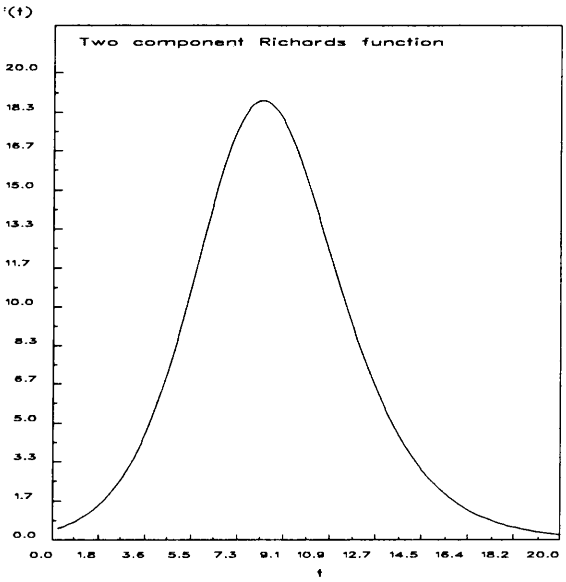

Figure 2.4.4

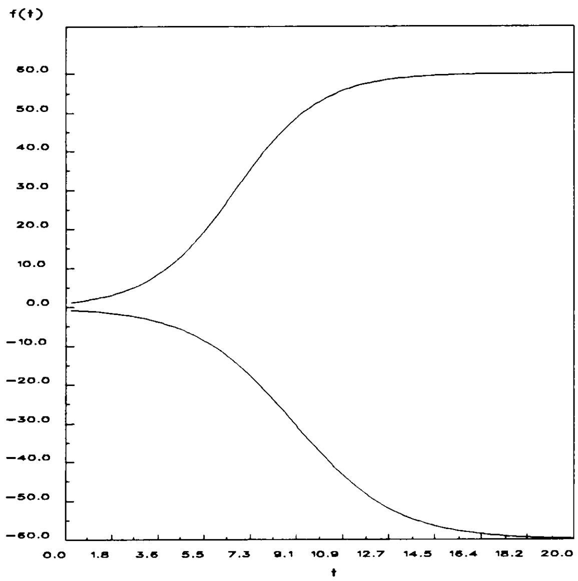

Figure 2.4.5

$$f(\mathbf{x},\alpha,t_i)=60(1+e^{4.0+0.6})^{-1}-60(1+e^{4.5+0.5})^{-1}$$

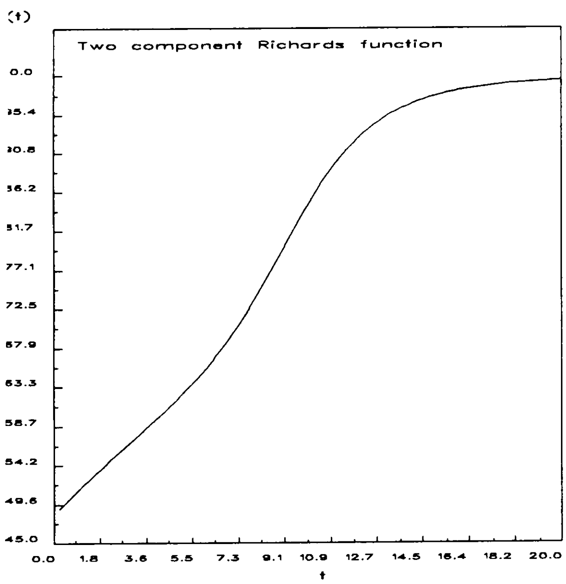

Figure 2.4.6

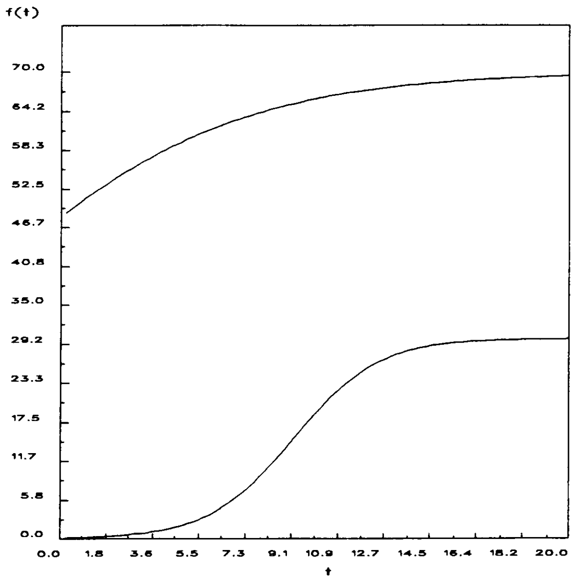

Figure 2.4.7

$$f(\mathbf{x},\alpha,t_i)=30(1+e^{5.5+0.6})^{-1}+70(1+e^{-0.8+0.2})^{-1}$$

2.8 SUMMARY

A multiphasic Richards function, discussed in this chapter, used to describe growth curves provides great insight, for understanding the biology of growth, especially in humans. It is shown that when using reparameterisation it is no longer necessary to impose inequality constraints on the parameters. The assumption of normality for the random parameter set $\mathbf{x}$ (cf. (2.4.1)), is more realistic than the same assumption for the set $\mathbf{a}$ (cf. (2.1.2)).

CHAPTER 3 FREQUENTIST ESTIMATION

3.1 INTRODUCTION

For some time there has been a continuing controversy among statisticians about the foundations of statistical inference, between the Frequentist and Bayesian schools of statistical thought (see Kotz and Johnson (1988)). The mutual criticism of Frequentists and Bayesians has given rise to a lively (and sometimes acrimonious) debate, which has helped to clarify a number of basic statistical issues. One of the central concerns is that of subjective versus objective data evaluation in scientific inference and reporting. Fisher, Neyman and Pearson developed their frequentist theories in a deliberate effort to free statistics from the Bayesian dependence on a prior distribution, and this aspect has continued as the central frequentist objection. Probabilities must rely on frequency-based interpretation (see Porkess (1991)).

The Bayesians pointed out that frequentist analysis involves similar types of specification. There is a choice of model and loss function, both of which must be chosen in the light of previous experience and involve judgements that are likely to vary from one person to another. Bayesians also pointed out that there is the problem of selecting a frame of reference that form the basis of the frequency calculations. For Bayesians the probabilities will always refer to the particular human being, animal, plant insect or object under observation. Ideas from both Frequentist and Bayesian sides are useful in thinking about practical problems, and in this chapter the Frequentist approach is discussed while the Bayesian approach is considered in the next chapter.

The gulf between the two approaches is considerably narrowed in the combination of two facts (see Smith (1984)) which is excellently summed up by Kotz and Johnson (1988): "The first is a basic result of Wald's (frequentist) decision theory to the effect that every admissible procedure is a Bayes solution or a limit of Bayes solutions. Secondly, frequentist tends not to believe in a unique correct approach and may therefore try a number of different solutions corresponding to different optimality principles robustness properties and so on. If this leads to similar conclusions, any one of them can be adopted. Otherwise, a careful examination of the differences may clarify the reason for the discrepancies and point to one as the most appropriate. Lacking such a resolution, one may instead prefer to report a number of different procedures".

To demonstrate the existence of growth phases or cycles in growth curves, it is necessary to have a number of repeated measurements, say n, taken over time on each of N experimental units. In this framework, the repeated measurements, on the i^{th} experimental unit, i=1,2,$\cdots$N, are referred to as the dependent variables, y_{it}, t=1,2,$\cdots$,n. The time points as well as additional variables that may be included in a model to predict the variation in the y_t, are referred to as the independent variables.

The following basic repeated measurements model will be considered in this chapter. Suppose the vector variate $\mathbf{y}_i$ (nx1) represents a set of n measurements, made on an individual, i, and that the individual's responses can be described by the model

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i,\boldsymbol{t}) + \epsilon_i, \qquad i = 1,2,\cdots,N \qquad (3.1.1)$$

Assume $\mathbf{x}_i$, i=1,2,$\cdots$,N are a random sample of an r-dimensional vector $\mathbf{x}$ of stochastic paremeters where $\mathbf{x}$ has a N($\boldsymbol{\theta}$,$\boldsymbol{\Phi}$) distribution. It is further assumed that the error vectors ϵ_i, i=1,2,$\cdots$,N are independently distributed as N($\mathbf{0}$,$\boldsymbol{\Lambda}$) variates, uncorrelated with the $\mathbf{x}_i$.

The majority of work on methods for repeated measures data has focussed on data that can be modelled by an expectation function that is *linear* in its parameters (e.g. Laird and Ware, 1982; Stiratelli, Laird, and Ware, 1984; Liang and Zeger, 1986; and Zeger, Liang, and Albert, 1988).

A model is said to be linear if a typical element f($\mathbf{x}_i$,t) of the nx1 component vector $\mathbf{f}(\mathbf{x}_i,\boldsymbol{t})$ in (3.1.1) can be expressed as a linear combination of the independent variables, that is

$$\mathbf{f}(\mathbf{x}_i,\boldsymbol{t}) = \mathbf{b}_t'\mathbf{x}_i \qquad i = 1,2,\cdots,N. \qquad (3.1.2)$$

An example of a linear model is

$$\mathbf{f}(\mathbf{x}_i,t) = x_1 + x_2 b_t + x_3 b_t^2 + \cdots + x_r b_t^{r-1}$$

where

$$\mathbf{b}_t = (1,\ b_t,\ b_t^2, \cdots, b_t^{r-1})\ . \qquad (3.1.3)$$

More recently considerable attention has been given to research on repeated measures data with *nonlinear* expectation functions (e.g. Berkley 1982; Bates and Watts 1988; Lindstrom and Bates 1990; Bock 1990; Davidian and Gallant 1991). Both the linear and nonlinear models are considered in this chapter.

Model (3.1.1) is referred to as a *nonlinear* model if the parameters in $f(\mathbf{x}_i,t)$ can not be expressed as a linear combination. An example of a nonlinear model is

$$f(\mathbf{x}_i,t) = \frac{x_1}{1 + exp\,(x_2 \cdot x_3^t)} \qquad (3.1.4)$$

It often happens that measurements on N experimental units are not made at the same timepoints and the number of responses, n_i, per individual, i, may differ. For example height measurements of a specific human being may have been obtained at ages 2, 4, 17, 22 and 35 months respectively while for another individual they may have been obtained only at 2, 6, 18 and 32 months respectively. For data with an unequal number of measurements per individual, model (3.1.1) may be adapted as follows:

$$\mathbf{y}_i \quad = \mathbf{f}(\mathbf{x}_i,\mathbf{t}_i) + \epsilon_i, \qquad\qquad i = 1,2,\cdots,N \qquad (3.1.5)$$

where $\mathbf{y}_i$, $\mathbf{f}(\mathbf{x}_i,\mathbf{t}_i)$, ϵ_i are n_ix1 component vectors with typical elements y_{ir}, $f(\mathbf{x}_i,t_{ir})$ and ϵ_{ir} respectively with r=1,2,$\cdots n_i$.

In Section 3.2 the maximum likelihood method of estimating unknown parameters is discussed. The marginal maximum likelihood estimation procedure (MML) which has been dealt with, directly or indirectly, in a number of papers recently (see e.g., Dempster, Rubin & Tsutakawa, 1981; Laird and Ware, 1982; Longford, 1987; Bock, 1989; Bock 1990; Strenio, Weisberg & Bryk, 1983) is discussed in Section 3.3.

The asymptotically distribution free (ADF) procedure for the analysis of covariance structures, is given in Sections 3.4. The Fisher scoring algorithm sometimes used for the frequentist estimation procedure is given Section 3.5.

The iterative computation of unknown parameter estimates consisting of an expectation step (E-step) followed by a maximization step (M-step) is called the EM (Expected Maximization) algorithm and is discussed in Section 3.6. The EM algorithm was explicitly introduced by Hartley (1958) as a procedure for calculating maximum likelihood estimates, given a random sample of size N, from a discrete population. Carter and Myers (1973) proposed the EM algorithm for maximum likelihood estimation from linear combinations of discrete probability functions, using linear combinations of Poisson random variables as an example. The algorithm was suggested a year later by Brown (1974) for computing the maximum-likelihood estimates of expected cell frequencies under an independence model in a two-way table with some missing cells. Chen and Fienberg (1976) suggested the EM algorithm for computing the maximum-likelihood estimates for the special case of cross-classified data with some observations only partially classified. The EM algorithm is a rapid, robust method for obtaining close approximations to the MML esimates and it has a wide range of applications which fall under its umbrella (see Dempster, Laird and Rubin 1977).
For some models, especiallly nonlinear models, it is often not possible to find a closed form solution for the integrals that have to be solved when using the estimation procedures discussed in this chapter. These integrals may be evaluated to a high degree of accuracy by means of the Gauss quadrature numerical integration technique described in Section 3.7. An illustration of the Gauss-quadrature integration involving the Richards growth curve model (see Section 2.6) is also given.

In Section 3.8 applications of the estimation procedures discussed in this chapter are given.

3.2. PARAMETER ESTIMATION USING MAXIMUM LIKELIHOOD.

Linear model

A generalisation of model (3.1.1), in compact matrix notation, for a *linear* function (cf. (3.1.2)), with unequal number of measurements per individual (cf. (3.1.5)) is

$$\mathbf{y}_i = \boldsymbol{B}_i\mathbf{x}_i + \epsilon_i, \qquad i = 1,2,\cdots,\mathrm{N}\ , \tag{3.2.1}$$

where $\mathbf{y}_i$ and ϵ_i are $\mathrm{n}_i \mathrm{x} 1$ component vectors with typical elements y_{it} and ϵ_{it} respectively where t=1,2,3,$\cdots$,n_i. The $\mathrm{n}_i \mathrm{x} r$ matrix, $\boldsymbol{B}_i$, is the design matrix.

Under the assumption that $\mathbf{x}_i \sim \mathrm{N}(\boldsymbol{\theta},\boldsymbol{\Phi})$ and since $\mathbf{y}_i$ is a *linear* combination of $\mathbf{x}_i$ and ϵ_i, it follows that $\mathbf{y}_i \sim \mathrm{N}(\xi_i,\Sigma_i)$, where

$$\xi_i\colon (\mathrm{n}_i \mathrm{x} 1) = \boldsymbol{B}_i\boldsymbol{\theta} \tag{3.2.2}$$

and

$$\Sigma_i\colon (\mathrm{n}_i \mathrm{x} \mathrm{n}_i) = \boldsymbol{B}_i\boldsymbol{\Phi}\boldsymbol{B}_i' + \Lambda_i\ . \tag{3.2.3}$$

The likelihood function of $\mathbf{y}_1$, $\mathbf{y}_2$,$\cdots$,$\mathbf{y}_\mathrm{N}$ is

$$\mathrm{L} = \prod_{i=1}^{\mathrm{N}} \mathrm{f}(\mathbf{y}_i)\ ,$$

where

$$\mathrm{f}(\mathbf{y}_i) = (2\pi)^{-\mathrm{n}_i/2}\,|\Sigma_i|^{-1/2} exp\{-\tfrac{1}{2}\,(\mathbf{y}_i - \xi_i)'\Sigma_i^{-1}(\mathbf{y}_i - \xi)\}\ . \tag{3.2.4}$$

Therefore

$$\mathrm{L} = (2\pi)^{\sum\limits_{i=1} \mathrm{n}_i/2} \prod_{i=1}^{\mathrm{N}} |\Sigma_i|^{-1/2}\ exp\{-\tfrac{1}{2}\sum_{i=1}^{\mathrm{N}} Q_i\}\ , \tag{3.2.5}$$

with

$$Q_i = (\mathbf{y}_i - \xi_i)'\Sigma_i^{-1}(\mathbf{y}_i - \xi_i) .$$

Thus

$$ln\,\mathrm{L} = -\tfrac{1}{2}\sum_{i=1}^{\mathrm{N}} \mathrm{n}_i ln(2\pi) - \tfrac{1}{2}\sum_{i=1}^{\mathrm{N}} ln\,|\Sigma_i| - \tfrac{1}{2}\sum_{i=1}^{\mathrm{N}} Q_i$$

$$= -\tfrac{1}{2}\left(\sum_{i=1}^{\mathrm{N}} \mathrm{n}_i ln(2\pi) + \sum_{i=1}^{\mathrm{N}} ln\,|\Sigma_i| + \sum_{i=1}^{\mathrm{N}} tr(\Sigma_i^{-1}\boldsymbol{G}_i)\right), \qquad (3.2.6)$$

with $\boldsymbol{G}_i$ defined as follows:

$$\boldsymbol{G}_i = (\mathbf{y}_i - \xi_i)(\mathbf{y}_i - \xi_i)' . \qquad (3.2.7)$$

Elements of the kx1 vector of unknown parameters γ are the (rx1) vector $\boldsymbol{\theta}$, the $\frac{1}{2}\mathrm{r}(\mathrm{r}+1)$ nonduplicated elements of $\boldsymbol{\Phi}$ and the m unknown parameters of $\Sigma_i = \Sigma_i(\tau)$ or $\Sigma_i = \Sigma_i(\tau_i)$.

Note that for

(i) Σ_i:$(\mathrm{n}_i \mathrm{x} \mathrm{n}_i) = \Sigma_i(\tau)$, (3.2.8)

τ could be a kx1 vector of unknown parameters common to each covariance matrix Σ_i, i=1,2,$\cdots$,N. For example, suppose that ϵ_{it}, j=1,2,$\cdots$,n_i is generated by an $AR(2)$ process, so that

$$\epsilon_{it} - \alpha_1\epsilon_{i,t-1} - \alpha_2\epsilon_{i,t-2} = \mathrm{u}_{it} , \qquad i=1,2,\cdots,\mathrm{N}$$

where u_{it} is a row of uncorrelated random variables with $\mathrm{E}(\mathrm{u}_{it})=0$ and $\mathrm{var}(\mathrm{u}_{it})=\sigma_u^2$. In this case $\Sigma_i = \Sigma_i(\tau)$, where $\tau=(\sigma_u^2,\alpha_1,\alpha_2)'$,

(ii) $\Sigma_i:(n_i x n_i) \ = \Sigma_i(\tau_i)$, (3.2.9)

where τ_i is a $(k_i x 1)$ vector of unknown parameters, $(k_i+r) < n_i$. Suppose for example that the errors for each experimental unit follows a different $AR(2)$ process and that a polynomial of degree 2 is to be fitted to the data. In this case r=3, k_i=3, $\tau_i=(\sigma^2_{ui},\alpha_{1i},\alpha_{2i})$, i=1,2,$\cdots$,N. If a common set of parameters β_0, β_1, β_2 is to be fitted to the responses of each experimental unit, the total number of unknown parameters is $3(N+1)$ which must be less in number than $\sum_{i=1}^{N} n_i$ to ensure the estimability of the unknown parameters.

Assume that $\Sigma_i = \Sigma_i(\tau)$, where τ is a parameter vector common to $\Sigma_1, \Sigma_2, \cdots, \Sigma_N$, for example $\Sigma_i = \sigma^2 \mathbf{I}_{(n_i x\ n_i)}$, then $\tau = \sigma^2$ and therefore

$$\boldsymbol{\gamma} = (\boldsymbol{\theta},\ \text{vecs}\boldsymbol{\Phi},\ \sigma^2)'. \tag{3.2.10}$$

The maximum likelihood estimate $\hat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ is found by solving the equations $\frac{\partial ln L}{\partial \gamma_i} = 0$, $i=1,2,\cdots,k$ with γ_i a typical element of the kx1 vector $\boldsymbol{\gamma}$.

From the results of Browne and Du Toit (1992) it follows that

$$\frac{\partial ln L}{\partial \gamma_i} = \frac{1}{2}\sum_{i=1}^{N} tr\left(\boldsymbol{P}_i \frac{\partial \Sigma_i}{\partial \gamma_i}\right) + \sum_{i=1}^{N} tr\left(\boldsymbol{R}_i \frac{\partial \xi}{\partial \gamma_i}\right), \tag{3.2.11}$$

where the symmetric matrix $\boldsymbol{P}_i$ and the general matrix $\boldsymbol{R}_i$ are defined by:

$$\boldsymbol{P}_i = \Sigma_i^{-1}(\boldsymbol{G}_i - \Sigma_i)\Sigma_i^{-1} , \tag{3.2.12}$$

$$\boldsymbol{R}_i = (\mathbf{y}_i - \xi_i)'\Sigma_i^{-1}. \tag{3.2.13}$$

Solving (3.2.11) for different values fo $\boldsymbol{\gamma}$ (cf. (3.2.10)) (see Bock 1990) it follows that

$$\hat{\boldsymbol{\theta}} = \left(\sum_{i=1}^{N} (\boldsymbol{B}_i'\hat{\Sigma}_i^{-1}\boldsymbol{B}_i)\right)^{-1}(\ \boldsymbol{B}_i'\hat{\Sigma}_i^{-1}\mathbf{y}_i) , \tag{3.2.14}$$

$$\text{vec}\hat{\boldsymbol{\Phi}} = \left(\sum_{i=1}^{N}(\boldsymbol{B}_i'\boldsymbol{B}_i)\otimes(\boldsymbol{B}_i'\boldsymbol{B}_i)\right)^{-1}\text{vec}\sum_{i=1}^{N}\left(\boldsymbol{B}_i'(\hat{\boldsymbol{G}}_i-\hat{\sigma}^2\boldsymbol{I}_{\text{n}_i})\boldsymbol{B}_i\right., \quad (3.2.15)$$

and

$$\hat{\sigma}^2 = (\sum_{i=1}^{N}\text{n}_i)^{-1}\left(\sum_{i=1}^{N}(\mathbf{y}_i-\boldsymbol{B}_i\hat{\boldsymbol{\theta}})'(\mathbf{y}_i-\boldsymbol{B}_i\hat{\boldsymbol{\theta}})-tr\,(\boldsymbol{B}_i'\boldsymbol{B}_i\hat{\boldsymbol{\Phi}})\right). \quad (3.2.16)$$

Equations (3.2.14) to (3.2.16) can be solved iteratively.

Nonlinear model

In nonlinear models the parameters appear not as separate coefficients of additive terms, but as products, quotients or exponents. Thus the estimators are nonlinear in the observations and in many cases expressable only as an infinite recursion (Bock & Thissen 1980).

Examples of possible nonlinear models for the function in (3.1.1), when $\mathbf{x}_i$ is not a random but a fixed parameter $\boldsymbol{\theta}$ are:

i) $$\text{f}(\boldsymbol{\theta},t_j) = \theta_1 - \theta_2 exp^{\theta_3 t_j} \quad (3.2.17)$$

ii) $$\text{f}(\boldsymbol{\theta},t_j) = \theta_1 + \theta_2\cos(\theta_3 t_j) \quad (3.2.18)$$

Graphical illustrations of examples of these nonlinear functions are given in Figures 3.2.1 and 3.2.2.

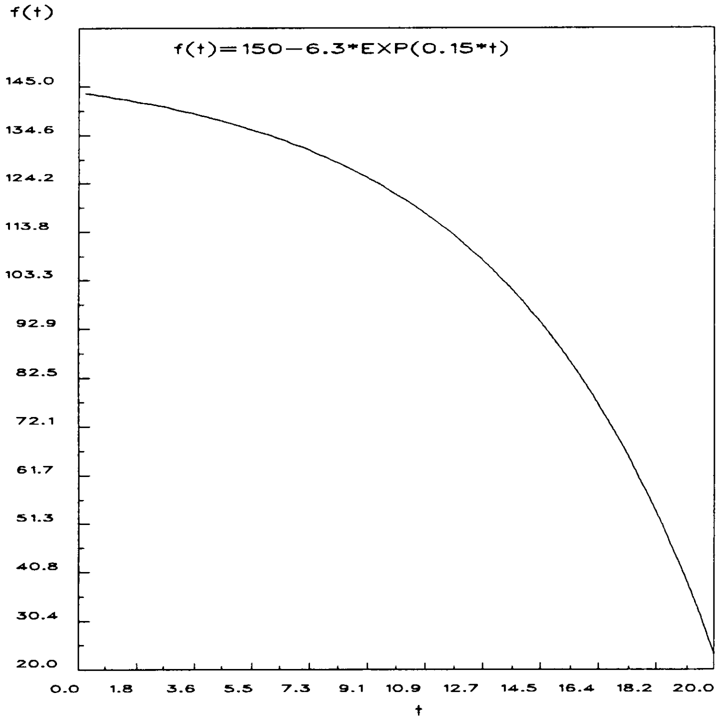

Figure 3.2.1 f(t)=150–6.3 $exp\,0.15t$

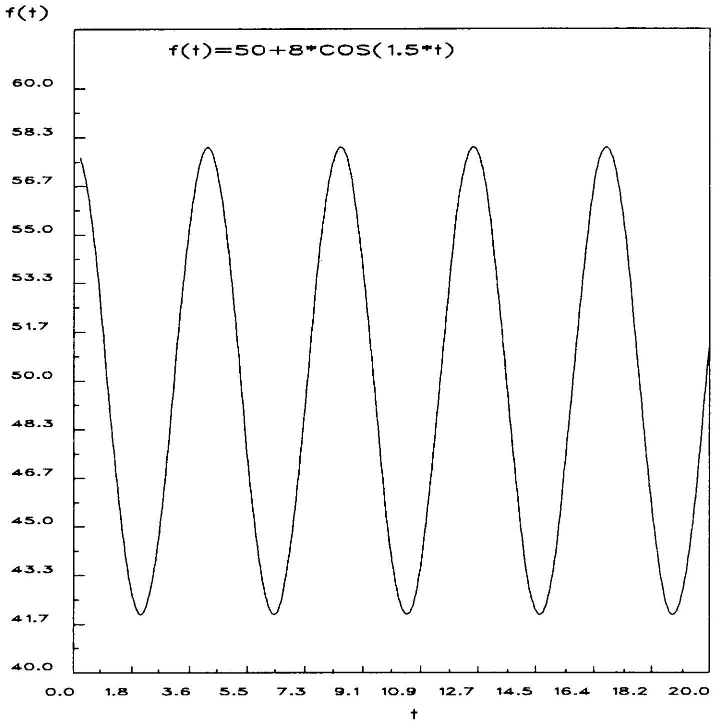

Figure 3.2.2 f(t)=50+8$*cos\,1.5t$

The set of regression equations (cf. (3.1.1)) for the *nonlinear fixed* parameter model with equal number of observations is

$$\mathbf{y}_i = \mathrm{f}(\boldsymbol{\theta},\boldsymbol{t}) + \epsilon_i, \qquad i = 1,2,\cdots,\mathrm{N}\ . \tag{3.2.19}$$

where $\boldsymbol{\theta}$ is an $\mathrm{r}\times 1$ vector of *fixed* but unknown parameters, and the error vectors $\epsilon_1,\epsilon_2,\cdots,\epsilon_\mathrm{N}$ are independently normally distributed, $\mathrm{E}(\epsilon_i)=\mathbf{0}$ and $\mathrm{Cov}(\epsilon_i\epsilon_i')=\boldsymbol{\Lambda}$.

Denote the expected value $\mathrm{f}(\boldsymbol{\theta},\boldsymbol{t})$ of $\mathbf{y}_i$ by $\boldsymbol{\xi}$ then the p.d.f. $\mathrm{f}(\mathbf{y}_i)$ of $\mathbf{y}_i$ under the distributional assumptions given above is

$$\mathrm{f}(\mathbf{y}_i) = (2\pi)^{-\mathrm{n}/2}|\boldsymbol{\Lambda}|^{-1/2}exp\Big\{-\tfrac{1}{2}(\mathbf{y}_i-\boldsymbol{\xi})'\boldsymbol{\Lambda}^{-1}(\mathbf{y}_i-\boldsymbol{\xi})\Big\}\ . \tag{3.2.20}$$

The likelihood function of $\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_\mathrm{N}$ (cf. (3.2.5)) is as follows:

$$\mathrm{L} = (2\pi)^{-\mathrm{Nn}/2}\prod_{i=1}^{\mathrm{N}}|\boldsymbol{\Lambda}|^{-1/2}exp\Big(-\tfrac{1}{2}tr\boldsymbol{\Lambda}^{-1}(\mathbf{y}_i-\boldsymbol{\xi})(\mathbf{y}_i-\boldsymbol{\xi})'\Big), \tag{3.2.21}$$

so that

$$\begin{aligned} ln\,\mathrm{L} &= -\tfrac{\mathrm{Nn}}{2}ln\,(2\pi) - \tfrac{\mathrm{N}}{2}ln\,|\boldsymbol{\Lambda}| - \tfrac{\mathrm{N}}{2}tr\boldsymbol{\Lambda}^{-1}\boldsymbol{G} \\ &= -\tfrac{\mathrm{N}}{2}\Big(\mathrm{n}ln(2\pi) + |\boldsymbol{\Lambda}| + tr(\boldsymbol{\Lambda}^{-1}\boldsymbol{G})\Big) \end{aligned} \tag{3.2.22}$$

where

$$\boldsymbol{G} = \tfrac{\mathrm{N}-1}{\mathrm{N}}\boldsymbol{S} + (\bar{\mathbf{y}}-\boldsymbol{\xi})(\bar{\mathbf{y}}-\boldsymbol{\xi})', \tag{3.2.23}$$

with

$$\bar{\mathbf{y}} = \tfrac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\mathbf{y}_i\ , \tag{3.2.24}$$

and

$$\boldsymbol{S} = \frac{1}{N-1}\sum_{i=1}^{N} (\mathbf{y}_i - \overline{\mathbf{y}})(\mathbf{y}_i - \overline{\mathbf{y}})'. \qquad (3.2.25)$$

Let

$$\begin{aligned} F^* &= -2\,ln\,L \\ &= N\Big(n\,ln\,(2\pi) + ln\,|\,\boldsymbol{\Lambda}\,| + tr\,(\boldsymbol{\Lambda}^{-1}\boldsymbol{G})\Big)\,. \end{aligned}$$

Minimization of F^* with respect to the unknown parameters is equivalent to the minimization of the discrepancy function F in which the constant term $n\,ln\,2\pi$ is omitted, therefore

$$F = N\Big(ln\,|\,\boldsymbol{\Lambda}\,| + tr(\boldsymbol{\Lambda}^{-1}\boldsymbol{G})\Big)\,. \qquad (3.2.26)$$

The minimum $\frac{\partial F}{\partial \gamma} = \mathbf{0}$ in (3.2.26) yields the normal maximum likelihood estimator $\hat{\gamma}$ of the vector of unknown parameters γ where $\gamma'=(\boldsymbol{\theta}',\boldsymbol{\tau}')$. The iterative Fischer scoring algorithm (discussed in Section 3.5) can be used if the model does not yield estimates in closed form. The gradient vector of the discrepancy function and the information matrix required in the optimization algorithm are given by (3.2.27) and (3.2.28) respectively:

$$\frac{1}{2}\frac{\partial F}{\partial \gamma_s} = [\boldsymbol{g}(\gamma)]_s = -N\left(2tr\left\{\boldsymbol{R}\frac{\partial \xi}{\partial \gamma_s}\right\} + tr\left\{\boldsymbol{P}\frac{\partial \boldsymbol{\Lambda}}{\partial \gamma_s}\right\}\right),\ s=1,2,\cdots,k \qquad (3.2.27)$$

with

$$\begin{aligned} \boldsymbol{P} &= \boldsymbol{\Lambda}^{-1}(\boldsymbol{G} - \boldsymbol{\Lambda})\boldsymbol{\Lambda}^{-1}, \\ \boldsymbol{R} &= (\overline{\mathbf{y}} - \xi)'\boldsymbol{\Lambda}^{-1}\,. \end{aligned}$$

An approximation to the Hessian (Browne and Du Toit 1992) is as follows:

$$\tfrac{1}{2}E\Big(\frac{\partial^2 F}{\partial \gamma_r \partial \gamma_s}\Big) \approx [H(\gamma)]_{r,s} = N\left(tr\Big(\frac{\partial \xi'}{\partial \gamma_r}\boldsymbol{\Lambda}^{-1}\frac{\partial \xi}{\partial \gamma_s}\Big) + tr\Big(\tfrac{1}{2}\boldsymbol{\Lambda}^{-1}\frac{\partial \boldsymbol{\Lambda}}{\partial \gamma_r}\,\boldsymbol{\Lambda}^{-1}\,\frac{\partial \boldsymbol{\Lambda}}{\partial \gamma_s}\Big)\right)\,. \qquad (3.2.28)$$

Extension to a q-group model.

It is frequently possible to divide the population from which the sampling units are drawn into q mutually exclusive subpopulations or groups. For example, the human population can be divided into the q=2 groups according to sex or for example into q=3 subgroups 18 to 25 years, 26 to 45 years, and older than 45 years according to the

variable age.
Suppose that

$$\{\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_{N_1}\};\ \{\mathbf{y}_{N_1+1},\mathbf{y}_{N_1+2},\cdots,\mathbf{y}_{N_1+N_2}\};\cdots;\{\mathbf{y}_{N_1+N_2+\cdots+N_{q-1}+1},\cdots,\mathbf{y}_N\}$$

are q independent random samples, each drawn from a $N(\mathbf{f}(\boldsymbol{\theta}_g),\boldsymbol{\Lambda})$ population, $g=1,2,\cdots,N$, and $N=N_1+N_2+\cdots+N_q$.

Note that a common covariance structure across groups is assumed but the means are allowed to vary. Denote the likelihood function for group g by L_g and let $\boldsymbol{\xi}_g=\mathbf{f}(\boldsymbol{\theta}_g)$, then

$$L_g \;=(2\pi)^{-Ngn/2}\prod_{i=m}^{l}|\boldsymbol{\Lambda}|^{-1/2}exp\left(-\tfrac{1}{2}tr\boldsymbol{\Lambda}^{-1}(\mathbf{y}_i-\boldsymbol{\xi}_g)(\mathbf{y}_i-\boldsymbol{\xi}_g)'\right),$$

where $m=N_1+N_2+\cdots N_{g-1}+1$ and $l=N_1+N_2+\cdots+N_g$.

The likelihood function of $\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_N$, denoted by L, is

$$L=\prod_{g=1}^{q}L_g\ .$$

Hence

$$ln\ L=\sum_{g=1}^{q}\left(\frac{Ng}{2}\left(n\ ln2\pi+|\boldsymbol{\Lambda}|+tr\boldsymbol{\Lambda}^{-1}\boldsymbol{G}_g\right)\right) \qquad (3.2.29)$$

where

$$\boldsymbol{G}_g=\sum_{i=m}^{l}(\mathbf{y}_i-\boldsymbol{\xi}_g)(\mathbf{y}_i-\boldsymbol{\xi}_g)'$$

$$=\boldsymbol{S}_g+(\overline{\mathbf{y}}_g-\boldsymbol{\xi}_g)(\overline{\mathbf{y}}_g-\boldsymbol{\xi}_g)', \qquad (3.2.30)$$

with

$$\boldsymbol{S}_g=\frac{1}{N_g}\sum_{i=m}^{l}(\mathbf{y}_i-\overline{\mathbf{y}}_g)(\mathbf{y}_i-\overline{\mathbf{y}}_g)',$$

and

$$\bar{\mathbf{y}}_g = \frac{1}{N_g} \sum_{i=m}^{l} \mathbf{y}_i \,.$$

The extension to a q-group model can be further generalised if it is assumed that n_g repeated measurements are made on the experimental units from group g, $g=1,2,\cdots,q$. To calculate $ln\,L$, n and Λ are replaced in (3.2.29) by n_g and Λ_g respectively. As before, it is assumed that $n_g x n_g$ covariance Λ_g is a function of mx1 parameters, that is, $\Lambda_g = \Lambda_g(\tau)$.

Expressions for the first order derivatives of $F = -2 ln\,L$ with respect to the unknown parameters are

$$\frac{\partial F}{\partial[\boldsymbol{\theta}_g]_s} = -2N_g\, tr\Big(\bar{\mathbf{y}}_g - \xi_g)'\Lambda_g^{-1}\frac{\partial \xi_g}{\partial[\boldsymbol{\theta}_g]_s}\,, \tag{3.2.31}$$

$$\frac{\partial F}{\partial \tau_s} = -N_g \sum_{g=1}^{q} tr\,[\boldsymbol{P}_g \frac{\partial \Lambda_g}{\partial \tau_s}]\,, \tag{3.2.32}$$

where

$$\boldsymbol{P}_g = \Lambda_g^{-1}(\boldsymbol{G}_g - \Lambda_g^{-1})\Lambda_g^{-1},$$

Let

$$\gamma = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{\theta}_g \\ \boldsymbol{\tau} \end{bmatrix}. \tag{3.2.33}$$

An approximation to the Hessian is given by

$$[H(\gamma)]_{r,s} = \sum_{g=1}^{q} N_g\, tr\left(\frac{\partial \xi_g'}{\partial \gamma_r}\Lambda_g^{-1}\frac{\partial \xi_g}{\partial \gamma_s} + \frac{1}{2}\,\Lambda_g^{-1}\frac{\partial \Lambda_g}{\partial \gamma_r}\,\Lambda_g^{-1}\,\frac{\partial \Lambda_g}{\partial \gamma_s}\right)\,. \tag{3.2.34}$$

An example of a fixed parameter two group model will be given in Chapter 6, Example 6.4.2.

For a *nonlinear random* parameter model in (3.1.1), based on an *equal* number of observations (n) per sampling unit (i), let

$$\boldsymbol{\xi} = \mathrm{E}(\mathbf{y}_i) = \mathrm{E}[\mathbf{f}(\mathbf{x}_i,\boldsymbol{t})] \ , \qquad (3.2.35)$$

$$\boldsymbol{\Psi} = \mathrm{Cov}[\mathbf{f}(\mathbf{x}_i,\boldsymbol{t}),\mathbf{f}'(\mathbf{x}_i,\boldsymbol{t})] \ , \qquad (3.2.36)$$

and

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{y}_i,\mathbf{y}_i') = \boldsymbol{\Psi} + \boldsymbol{\Lambda} \ . \qquad i = 1,2,\cdots,\mathrm{N}$$

Let ξ_s ψ_{rs} and σ_{rs} denote typical elements of the mean vector $\boldsymbol{\xi}$, the covariance matrix $\boldsymbol{\Psi}$ and the covariance matrix $\boldsymbol{\Sigma}$ respectively.

The moments can be calculated as follows:

$$\xi_s = \int \mathrm{f}(\mathbf{x}_i,t_s)\,\mathrm{g}(\mathbf{x}_i)d\mathbf{x}_i \ , \qquad (3.2.38)$$

$$\psi_{rs} = \int \mathrm{f}(\mathbf{x}_i,t_r)\cdot \mathrm{f}(\mathbf{x}_i,t_s)\,\mathrm{g}(\mathbf{x}_i)d\mathbf{x}_i \ - \xi_r\xi_s \ , \qquad (3.2.39)$$

$$\sigma_{rs} = \omega_{rs} + \lambda_{rs}, \qquad (3.2.40)$$

where $\mathrm{g}(\mathbf{x}_i)$ denotes the probability density function (pdf) of the r x 1 vector of stochastic parameters $\mathbf{x}_i$.

For the *nonlinear* reponse function $\mathrm{f}(\mathbf{x}_i,t_s)$ a closed form solution of the integrals (3.2.37) and (3.2.38) usually do not exist. These integrals may be evaluated, to a high degree of accuracy, by means of the Gauss quadrature numerical integration technique described in Section 3.7 of this chapter.

Remarks

1) An iterative procedure is required to obtain ML estimates for $\boldsymbol{\theta}$, $\boldsymbol{\Phi}$ and τ. Du

Toit(1991) wrote a modified version of the computer program AUFIT (See du Toit & Browne, 1992) to obtain the maximum likelihood (ML) estimates using (3.2.11) and similar expressions to evaluate the information matrix.

2) It should be noted, however, that the estimation procedure described above, involves the inversion of nxn matrices. For large values of n_i, the iterative procedure may become very time consuming.

3) Bock (1990) suggested an alternative method for obtaining maximum likelihood estimates which we will refer to as MML (marginal maximum likelihood method). This method will be discussed in detail in the next section.

3.3 PARAMETER ESTIMATION USING MARGINAL MAXIMUM LIKELIHOOD.

Consider the random parameter model defined in (3.1.1). It is assumed that $\mathbf{x}_i \sim N(\boldsymbol{\theta},\boldsymbol{\Phi})$, thus the pdf of $\mathbf{x}_i$ is as follows:

$$g(\mathbf{x}_i) = (2\pi)^{-r/2} \,|\boldsymbol{\Phi}|^{-1/2}\, exp -\tfrac{1}{2}\,P \qquad (3.3.1)$$

with

$$P = tr\,\boldsymbol{\Phi}^{-1}(\mathbf{x}_i-\boldsymbol{\theta})(\mathbf{x}_i-\boldsymbol{\theta})'.$$

The unknown parameters in the density function $g(\mathbf{x}_i)$ defined in (3.3.1) are the rx1 vector $\boldsymbol{\theta}$, as well as the $\frac{1}{2}r(r+1)$ nonduplicate elements of $\boldsymbol{\Phi}$.

$$\boldsymbol{\beta}' = (\boldsymbol{\theta}',(\text{vecs}\boldsymbol{\Phi})') \,. \qquad (3.3.2)$$

It further follows from the distributional assumptions given above that the pdf of $\mathbf{y}_i$ given $\mathbf{x}_i$ is:

$$f(\mathbf{y}_i\,|\,\mathbf{x}_i) = (2\pi)^{-n_i/2}\,|\Lambda_i|^{-1/2}\; exp\Big\{-\tfrac{1}{2}\big(\mathbf{y}_i-\mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i)\big)'\Lambda_i^{-1}\big(\mathbf{y}_i-\mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i)\big)\Big\}\,. \qquad (3.3.3)$$

The vector $\boldsymbol{\tau}$ denotes the unknown parameters of the pdf, $f(\mathbf{y}_i\,|\,\mathbf{x}_i)$, (cf. (3.3.3)), where it is assumed that (cf. (3.2.8))

$$\boldsymbol{\Lambda}_i = \boldsymbol{\Lambda}_i(\boldsymbol{\tau}). \tag{3.3.4}$$

From standard results of conditional distributions (e.g. Morrison, 1991) it follows that

$$\mathrm{h}(\mathbf{y}_i) \;=\; \int \mathrm{f}(\mathbf{y}_i \mid \mathbf{x}_i)\mathrm{g}(\mathbf{x}_i)\, d\,\mathbf{x}_i \,,$$

thus the likelihood function L would be written as follows

$$\mathrm{L}(\mathbf{y}_1, \cdots, \mathbf{y}_\mathrm{N}) \;=\; \prod_{i=1}^{\mathrm{N}} \int \mathrm{f}(\mathbf{y}_i \mid \mathbf{x}_i)\mathrm{g}(\mathbf{x}_i)\, d\,\mathbf{x}_i \,. \tag{3.3.5}$$

Proposition 3.3.1

For $\beta_k \in \boldsymbol{\beta}$ (cf. (3.3.2)) an unknown parameter of the function $\mathrm{g}(\mathbf{x}_i)$ (cf. (3.3.1)) it follows that for the likelihood function, L defined in (3.3.5)

$$\begin{aligned}\frac{\partial\, ln\, \mathrm{L}}{\partial\, \beta_k} &= \sum_{i=1}^{\mathrm{N}} \int \left(\frac{\partial\, ln\, \mathrm{g}(\mathbf{x}_i)}{\partial\, \beta_k} \right) \mathrm{p}(\mathbf{x}_i \mid \mathbf{y}_i)\, d\,\mathbf{x}_i \\ &= \sum_{i=1}^{\mathrm{N}} \mathrm{E}_{\mathbf{x}_i \mid \mathbf{y}_i} \left(\frac{\partial\, ln\, \mathrm{g}(\mathbf{x}_i)}{\partial\, \beta_k} \right). \end{aligned} \tag{3.3.6}$$

Similarily for $\tau_k \in \boldsymbol{\tau}$ (cf. (3.2.8) and (3.2.9)), an unknown parameter of the function $\mathrm{f}(\mathbf{y}_i|\mathbf{x}_i)$ (cf. (3.3.3)), the corresponding equation is:

$$\begin{aligned}\frac{\partial\, ln\, \mathrm{L}}{\partial\, \tau_k} &= \sum_{i=1}^{\mathrm{N}} \int \left(\frac{\partial\, ln\, \mathrm{f}(\mathbf{y}_i|\mathbf{x}_i)}{\partial\, \tau_k} \right) \mathrm{p}(\mathbf{x}_i \mid \mathbf{y}_i)\, d\,\mathbf{x}_i \\ &= \sum_{i=1}^{\mathrm{N}} \mathrm{E}_{\mathbf{x}_i \mid \mathbf{y}_i} \left(\frac{\partial\, ln\, \mathrm{f}(\mathbf{y}_i|\mathbf{x}_i)}{\partial\, \tau_k} \right). \end{aligned} \tag{3.3.7}$$

Proof

For the likelihood function, L, defined in (3.3.5) and γ_k an element of the vector of unknown parameters

$$\frac{\partial}{\partial \gamma_k} ln\, L(\mathbf{y}_1, \cdots, \mathbf{y}_N) = \sum_{i=1}^{N} \frac{\int \frac{\partial}{\partial \gamma_k}\Big(f(\mathbf{y}_i \,|\, \mathbf{x}_i) g(\mathbf{x}_i) \Big)\, d\,\mathbf{x}_i}{\int f(\mathbf{y}_i \,|\, \mathbf{x}_i) g(\mathbf{x}_i)\, d\,\mathbf{x}_i} \qquad (3.3.8)$$

For $\gamma_k = \beta_k$ it follows that

$$\frac{\partial}{\partial \beta_k} ln\, L(\mathbf{y}_1, \cdots, \mathbf{y}_N) = \sum_{i=1}^{N} \frac{\int f(\mathbf{y}_i \,|\, \mathbf{x}_i)\, \frac{\partial\, g(\mathbf{x}_i)}{\partial \beta_k} d\,\mathbf{x}_i}{h(\mathbf{y}_i)}\,. \qquad (3.3.9)$$

Subsitution of the following result into (3.3.9)

$$\frac{\partial\, ln\, g(\mathbf{x}_i)}{\partial \beta_k} = \frac{1}{g(\mathbf{x}_i)} . \frac{\partial\, g(\mathbf{x}_i)}{\partial \beta_k}\ ,$$

gives

$$\frac{\partial\, ln\, L}{\partial \beta_k} = \sum_{i=1}^{N} \frac{\int \left(\frac{\partial\, ln\, g(\mathbf{x}_i)}{\partial \beta_k} \right) f(\mathbf{y}_i \,|\, \mathbf{x}_i)\ g(\mathbf{x}_i)\, d\,\mathbf{x}_i}{h(\mathbf{y}_i)}\,. \qquad (3.3.10)$$

Using the definition of the conditional density function

$$p(\mathbf{x}_i \,|\, \mathbf{y}_i) = \frac{f(\mathbf{y}_i \,|\, \mathbf{x}_i)\ g(\mathbf{x}_i)}{h(\mathbf{y}_i)}\,, \qquad (3.3.11)$$

the result in (3.3.5) follows. Similarily the result in (3.3.6) is obtained. □

From (3.3.1)

$$ln\, g(\mathbf{x}_i) = -\tfrac{r}{2} ln\, 2\pi - \tfrac{1}{2} ln\, |\boldsymbol{\Phi}| - \tfrac{1}{2} tr\, \boldsymbol{\Phi}^{-1} (\mathbf{x}_i - \boldsymbol{\theta})(\mathbf{x}_i - \boldsymbol{\theta})'. \qquad (3.3.12)$$

Hence

$$\frac{\partial\, ln\, g(\mathbf{x}_i)}{\partial \boldsymbol{\theta}} = \boldsymbol{\Phi}^{-1} (\mathbf{x}_i - \boldsymbol{\theta}).$$

Therefore (cf. (3.3.9))

$$\frac{\partial\, ln\, \mathrm{L}}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{\mathrm{N}} \boldsymbol{\Phi}^{-1} \mathrm{E}_{\mathbf{x}_i\,|\,\mathbf{y}_i}(\mathbf{x}_i - \hat{\boldsymbol{\theta}}) = \mathbf{0}\ ,$$

and hence

$$\hat{\boldsymbol{\theta}} = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \mathrm{E}(\mathbf{x}_i | \mathbf{y}_i)\ . \tag{3.3.13}$$

Similarily (see Bock 1990) it follows that

$$\hat{\boldsymbol{\Phi}} = \frac{1}{\mathrm{N}} \Bigg(\sum_{i=1}^{\mathrm{N}} \mathrm{E}(\mathbf{x}_i | \mathbf{y}_i) \mathrm{E}(\mathbf{x}_i | \mathbf{y}_i{}') + \mathrm{Cov}(\mathbf{x}_i | \mathbf{y}_i) \Bigg) - \hat{\boldsymbol{\theta}} \hat{\boldsymbol{\theta}}'\ ,$$

which can be alternatively written as

$$\hat{\boldsymbol{\Phi}} = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \Big(\mathrm{Cov}(\mathbf{x}_i | \mathbf{y}_i) + (\mathrm{E}(\mathbf{x}_i | \mathbf{y}_i) - \hat{\boldsymbol{\theta}})(\mathrm{E}(\mathbf{x}_i | \mathbf{y}_i) - \hat{\boldsymbol{\theta}})' \Big)\ . \tag{3.3.14}$$

From (3.3.3) it follows that

$$ln\, \mathrm{f}(\mathbf{y}_i \,|\, \mathbf{x}_i) = -\frac{\mathrm{n}_i}{2}\, ln\, (2\pi) - \frac{1}{2} ln\, |\, \boldsymbol{\Lambda}_i \,| - \frac{1}{2} tr\, \boldsymbol{\Lambda}_i^{-1} \boldsymbol{G}_{\mathbf{y}_i}\ ,$$

$$\boldsymbol{G}_{\mathbf{y}_i} = (\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i, \boldsymbol{t}_i))(\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i, \boldsymbol{t}_i))'$$

and for

$$\boldsymbol{\Lambda}_i = \sigma^2 \boldsymbol{I}_{\mathrm{n}_i}\ ,$$

$$\frac{\partial\, ln\, \mathrm{f}(\mathbf{y}_i | \mathbf{x}_i)}{\partial\, \sigma^2} = \frac{1}{2}\, tr\, \sigma^{-4} \left(\boldsymbol{G}_{\mathbf{y}_i} - \sigma^2 \boldsymbol{I}_{\mathrm{n}_i} \right).$$

Using (3.3.10) it follows that

$$\frac{\partial\, ln\, \mathrm{L}}{\partial\, \sigma^2} = \frac{1}{2} \sum_{i=1}^{\mathrm{N}} \mathrm{E}_{\mathbf{x}_i\,|\,\mathbf{y}_i} \Big(tr\ \sigma^{-4}\ (\hat{\boldsymbol{G}}_{\mathbf{y}_i} - \hat{\sigma}^2 \boldsymbol{I}_{\mathrm{n}_i}) \Big) = 0\ .$$

thus

$$\hat{\sigma}^2 = \Big(\sum_{i=1}^{\mathrm{N}} \mathrm{n}_i \Big)^{-1} \sum_{i=1}^{\mathrm{N}} \mathrm{E}_{\mathbf{x}_i\,|\,\mathbf{y}_i} tr\, \hat{\boldsymbol{G}}_{\mathbf{y}_i} \tag{3.3.15}$$

Remark:

Using the identity $(\boldsymbol{D}+\boldsymbol{E}^{-1})^{-1} = \boldsymbol{E}-\boldsymbol{E}\boldsymbol{D}(\boldsymbol{E}^{-1}+\boldsymbol{D})^{-1}$ (for $\boldsymbol{D}$ and $\boldsymbol{E}$ any two matrices) it can be shown that for the linear model $\mathrm{E}(\mathbf{x}_i|\mathbf{y}_i)$ (cf. (3.3.16)) is equivalent to the Bayes estimator

$$\mathrm{E}(\mathbf{x}_i|\mathbf{y}_i) = (\boldsymbol{\Phi}^{-1}+\boldsymbol{B}_i'\Lambda_i^{-1}\boldsymbol{B}_i)^{-1}(\mathbf{y}_i'\Lambda_i^{-1}\boldsymbol{B}_i+\boldsymbol{\theta}'\boldsymbol{\Phi}^{-1}). \qquad (3.3.21)$$

given in (4.3.15) of Section 4.3 in Chapter 4.

Nonlinear model

To obtain $\mathrm{E}(\mathbf{x}_i|\mathbf{y}_i)$ and $\mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i)$ for the *nonlinear* model the following integral has to be solved (cf. (3.3.9))

$$\begin{aligned}\mathrm{E}(x_\alpha{}^k x_\beta{}^l\,|\,\mathbf{y}_i) &= \int x_\alpha{}^k x_\beta{}^l\ p(\mathbf{x}_i|\mathbf{y}_i)\,d\mathbf{x}_i \\ &= \frac{\int x_\alpha{}^k x_\beta{}^l\ f(\mathbf{y}_i\,|\,\mathbf{x}_i)\ g(\mathbf{x}_i)d\mathbf{x}_i}{\int f(\mathbf{y}_i\,|\,\mathbf{x}_i)\ g(\mathbf{x}_i)d\mathbf{x}_i}, \qquad \begin{array}{l} k,l=0,1,2,\cdots \\ \alpha,\beta=1,2,\cdots,r\ . \end{array} \qquad (3.3.20)\end{aligned}$$

For some *nonlinear* functions it may not be possible to obtain a closed form solution to the integral in (3.3.20), in which case use is made of the Gauss-quadrature numerical integration technique discussed in Section 3.7.

3.4 ASYMPTOTICALLY DISTRIBUTION FREE PROCEDURES.

Browne (1974) considered generalised least squares estimators in the analysis of covariance structures for a case, implied by multivariate normality, where the variances and covariances of the sample covariance matrix are functions of their expected values. Subsequently this condition was discarded (Browne and Greenacre, 1976) and an asymptotically distribution free (ADF) method for the analysis of covariance structures was obtained.

Suppose that $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N$ are independently and identically distributed vector variates and that

$$E(\mathbf{y}_i) = \xi , \quad \text{and} \quad Cov(\mathbf{y}_i, \mathbf{y}_i') = \Sigma . \tag{3.4.1}$$

It is assumed that the n elements of $\xi = \xi(\gamma)$ and the $\frac{1}{2}n(n+1)$ nonduplicated elements of $\Sigma = \Sigma(\gamma)$ are functions of k $(k \leq n + \frac{1}{2}(n+1)$) parameters which are to be estimated. In the general framework the parameters are represented by the (kx1) vector γ.

Let $\bar{\mathbf{y}}$ and S respectively denote the sample mean vector and sample covariance matrix of the $\mathbf{y}_i$, $i=1,2,\cdots,N$, that is

$$\bar{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i ,$$

$$S = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})' . \tag{3.4.2}$$

Let

$$n^* = \frac{1}{2} n(n+3) \tag{3.4.3}$$

and define the $(n^* x 1)$ vectors $\mathbf{u}$ and $\boldsymbol{v}$ (Du Toit, 1979) as

$$\mathbf{u} = \begin{bmatrix} \bar{\mathbf{y}} \\ vecs(S) \end{bmatrix} , \tag{3.4.4}$$

and

$$\boldsymbol{v} = E(\mathbf{u}) = \begin{bmatrix} \xi \\ vecs(\Sigma) \end{bmatrix} . \tag{3.4.5}$$

It is known (e.g. Cramér, 1946, p 365) that the limiting distribution of $\sqrt{N}(\mathbf{u} - \boldsymbol{v})$ is multivariate normal with mean vector $\mathbf{0}$. Suppose that the covariance matrix of $\mathbf{u}$ exists and is nonsingular. Consider the structural model $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{\gamma})$ and assume that $\boldsymbol{v}$ and the first and second order partial derivatives of $\boldsymbol{v}$, with respect to the elements of $\boldsymbol{\gamma}$, are continuous. It is further assumed that the $(n^* x k)$ matrix

$$\Delta = \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\gamma}'} , \qquad (3.4.6)$$

is of full column rank and that $\boldsymbol{\gamma}$ is identified, that is

$$\boldsymbol{v}(\gamma_1) = \boldsymbol{v}(\gamma_2) \Rightarrow \gamma_1 = \gamma_2 . \qquad (3.4.7)$$

Let

$$\boldsymbol{\Omega}: (n^* x n^*) = \mathrm{Cov}(\mathbf{u},\mathbf{u}') \qquad (3.4.8)$$

and

$$\begin{aligned} \bar{\boldsymbol{\Omega}}: (n^* x n^*) &= \mathrm{Cov}(\sqrt{N}\mathbf{u}, \sqrt{N}\mathbf{u}') \\ &= N\boldsymbol{\Omega} \end{aligned} \qquad (3.4.9)$$

Now let $\hat{\gamma}$ be a generalised least square (GLS) estimator of γ obtained by minimizing

$$F_{GLS}(\gamma) = (\mathbf{u} - \boldsymbol{v}(\gamma))'\bar{\boldsymbol{\Omega}}_c^{-1}(\mathbf{u} - \boldsymbol{v}(\gamma)), \qquad (3.4.10)$$

where $\bar{\boldsymbol{\Omega}}_c$ is any consistent estimator of $\bar{\boldsymbol{\Omega}}$, that is

$$\operatorname*{plim}_{N \to \infty} \bar{\boldsymbol{\Omega}}_c = \bar{\boldsymbol{\Omega}} . \qquad (3.4.11)$$

Then (cf. Browne, 1974), the limiting distribution of $\sqrt{N}(\hat{\gamma} - \gamma)$ is multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $(\Delta'(\bar{\boldsymbol{\Omega}}_c)^{-1}\Delta)^{-1}$.

If $\boldsymbol{v} = \boldsymbol{v}(\gamma)$ is true, then the limiting distribution of

$$F^*(\gamma) = N\, F_{GLS}(\gamma) \qquad (3.4.12)$$

is chi-square with $n^* - k$ degrees of freedom.

General expressions for the first and second order derivatives are

$$\frac{\partial F_Q^*}{\partial \gamma_s} = -2\Big(\mathbf{u} - \boldsymbol{v}(\gamma)\Big)' \, \bar{\boldsymbol{\Omega}}_c{}^{-1} \frac{\partial \boldsymbol{v}(\gamma)}{\partial \gamma_s} , \qquad (3.4.13)$$

$$\frac{\partial^2 F_Q^*}{\partial \gamma_r \partial \gamma_s} = 2\left(\frac{\partial [\boldsymbol{v}(\gamma)]'}{\partial \gamma_r} \, \bar{\boldsymbol{\Omega}}_c{}^{-1} \frac{\partial \boldsymbol{v}(\gamma)}{\partial \gamma_s}\right) - 2\left((\mathbf{u} - \boldsymbol{v}(\gamma))' \, (\bar{\boldsymbol{\Omega}}_c)^{-1} \frac{\partial^2 \boldsymbol{v}(\gamma)}{\partial \gamma_r \partial \gamma_s}\right) . \qquad (3.4.14)$$

An approximation to the Hessian matrix is obtained by taking expected values in (3.4.14). Therefore

$$[\mathrm{H}]_{r,s} \simeq 2 \frac{\partial [\boldsymbol{v}(\gamma)]'}{\partial \gamma_r} \, \bar{\boldsymbol{\Omega}}_c{}^{-1} \frac{\partial \boldsymbol{v}(\gamma)}{\partial \gamma_s} , \qquad (3.4.15)$$

hence

$$\mathrm{H} \simeq 2 \, \Delta' \bar{\boldsymbol{\Omega}}_c{}^{-1} \Delta \, . \qquad (3.4.16)$$

Using expressions (3.4.13) and (3.4.16) to calculate the gradient vector and the approximation to the Hessian matrix, the GLS estimator $\hat{\gamma}_{GLS}$ of γ may be obtained by using, for example, the optimization procedure described in Section 3.5.

In the case where $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N$ are not normally distributed, Du Toit (1979) gave general expressions for the calculation of the elements of the weight matrix $\bar{\boldsymbol{\Omega}}_c$ (cf. (3.4.10)). These expressions involve the evaluation of moments of random variables of up to fourth order. Du Toit has shown that the number of 4-th order nonduplicate moments of an n-dimensional stochastic variable is $\binom{n+d-1}{d}$, where $\binom{n}{r} = \frac{n!}{r!(n-r)!}$.

For example, if n=10, the elements of $\boldsymbol{\Omega}$ will consist of 715 nonduplicated 4-th order moments. Computation of $\boldsymbol{\Omega}$ for large values of n can become very tedious and time consuming.

Under the assumption that $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N$ are identically and independently distributed as $N(\xi, \Sigma)$ variates, it can be shown (e.g. Browne (1974), Du Toit (1993)) that

$$\boldsymbol{\Omega}^{-1} = \begin{bmatrix} N\Sigma^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{N}{2}\boldsymbol{K}'\left(\Sigma^{-1} \otimes \Sigma^{-1}\right)\boldsymbol{K} \end{bmatrix}. \tag{3.4.17}$$

where (see e.g. Browne (1974) or McGullogh (1982)) $\boldsymbol{K}$ is an unique $n^2 \mathrm{x} \frac{1}{2}n(n+1)$ matrix such that

$$\operatorname{vec} \boldsymbol{A} = \boldsymbol{K} \operatorname{vecs} \boldsymbol{A} \ , \tag{3.4.18}$$

with $\boldsymbol{A}$ a symmetric $n \mathrm{x} n$ matrix.

Let $\xi = \xi(\gamma)$ and $\Sigma = \Sigma(\gamma)$ where γ is a $k \mathrm{x} 1$ vector of unknown parameters to be estimated. Suppose that $\boldsymbol{V}$ is a consistent estimator of Σ, that is

$$\operatorname*{plim}_{N \to \infty} \ \boldsymbol{V} = \Sigma \tag{3.4.19}$$

It follows that (cf. (3.4.9))

$$\overline{\boldsymbol{\Omega}}_c = \begin{bmatrix} \boldsymbol{V}^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\boldsymbol{K}'\left(\boldsymbol{V}^{-1} \otimes \boldsymbol{V}^{-1}\right)\boldsymbol{K} \end{bmatrix}^{-1}. \tag{3.4.20}$$

is a consistent estimator of $\overline{\boldsymbol{\Omega}}$. (3.4.21)

Using the result (see e.g. Browne (1974))

$$\mathbf{x}'(\boldsymbol{V} \otimes \boldsymbol{W})\mathbf{y} = tr[\boldsymbol{X}\ \boldsymbol{V}\ \boldsymbol{Y}'\boldsymbol{W}'] \ , \tag{3.4.22}$$

where $\mathbf{x}$=vec$\boldsymbol{X}$ and $\mathbf{y}$=vec$\mathbf{Y}$ it follows that

$$\mathrm{F}_{\mathrm{GLS}}(\gamma) = tr\Big(\boldsymbol{V}^{-1}(\bar{\mathbf{y}} - \xi(\gamma))(\mathbf{y} - \xi(\gamma))'\Big) + \tfrac{1}{2}tr\Big(\boldsymbol{V}^{-1}(\boldsymbol{S} - \Sigma(\gamma))\Big)^2 \tag{3.4.23}$$

The first order derivatives of $\mathrm{F}_{\mathrm{GLS}}(\gamma)$ with respect to the elements of the parameter vector γ are

$$\frac{\partial \mathrm{F}_{\mathrm{GLS}}(\gamma)}{\partial \gamma_s} = -2\ tr\Big((\bar{\mathbf{y}} - \xi(\gamma))'\boldsymbol{V}^{-1}\frac{\partial \xi(\gamma)}{\partial \gamma_s}\Big) - tr\Big(\boldsymbol{V}^{-1}(\boldsymbol{S} - \Sigma(\gamma))\boldsymbol{V}^{-1}\frac{\partial \Sigma(\gamma)}{\partial \gamma_s}\Big), \quad s=1,2,\cdots,k \tag{3.4.24}$$

The discrepancy function F under the assumption of multivariate normality (Browne and Du Toit, 1992) is given by

$$\mathrm{F} = ln\,|\Sigma| - ln\,|\boldsymbol{S}^*| + tr\,[\boldsymbol{G}\Sigma^{-1}] - \mathrm{n} \tag{3.4.25}$$

where

$$\boldsymbol{S}^* = \frac{\mathrm{N}-1}{\mathrm{N}}\ \boldsymbol{S}$$

and (cf. (3.2.23))

$$\boldsymbol{G} = \boldsymbol{S}^* + (\bar{\mathbf{y}} - \xi(\hat{\gamma}))(\bar{\mathbf{y}} - \xi(\hat{\gamma}))' \tag{3.4.26}$$

Anderson and Rubin (1956, pp. 145-146) have shown, that in the case of non-normality, a consistent estimate, $\hat{\gamma}$, of the unknown parameter vector γ is obtained as the solution to the equations

$$\frac{\partial \mathrm{F}}{\partial \gamma_s} = 0\ , \qquad s = 1,2,\cdots,k,$$

where

$$\frac{\partial \mathrm{F}}{\partial \gamma_s} = -2\,tr\,(\mathbf{y}-\xi)'\Sigma^{-1}\frac{\partial \xi(\gamma)}{\gamma_s} - \; tr\,\Sigma^{-1}(\boldsymbol{G}-\Sigma)\Sigma^{-1}\frac{\partial \Sigma}{\partial \gamma_s} \;. \tag{3.4.27}$$

Remarks

(1) Du Toit (1979) used a GLS method to obtain the correct standard errors and χ^2-test statistic in the analysis of nonlinear random parameter models. His approach is essentially a two stage procedure. In the first stage the normal maximum likelihood discrepancy function (cf. (3.4.27)) is minimized to obtain a consistent estimator of the unknown parameters. These values are used to calculate a consistent estimate of the weight matrix which involves the calculation of fourth order moments. As the number of parameters r of the nonlinear response function and the number of repeated measurements n increase, the computational labour and storage requirements for the calculation of $\overline{\boldsymbol{\Omega}}_{\mathrm{c}}$ increase.

(2) Since the evaluation of fourth order moments of nonlinear random parameter models involves the solution of multi dimensional integrals (which usually have to be evaluated numerically), the calculation of the weight matrix can be time consuming. For this reason, in the estimation procedures discussed in Chapter 5 the weight matrix $\overline{\boldsymbol{\Omega}}_{\mathrm{c}}{}^{-1}$ (cf. (3.4.10)) under the assumption of multivariate normality is used since this matrix depends only on second order central moments.

(3) In Gauss-Newton type algorithms (cf. Section 3.5) practical experience (Browne and Du Toit, 1992) has shown that the convergence of the algorithm does not appear to be sensitive to whether exact derivatives are used or not.

The approximation used for evaluating $\dfrac{\partial \Sigma}{\partial \gamma_j}$ is

$$\frac{\partial \Sigma}{\partial \gamma_j} \approx \frac{\Sigma(\gamma + \epsilon_j \mathbf{i}_j) - \Sigma(\gamma)}{\epsilon_j}\,, \tag{3.4.28}$$

where $\mathbf{i}_j$ is a column vector with zero elements except for the $j-$th element that is equal to unity and

$$\epsilon_j = \epsilon \max\{1, ln|1+\gamma_j|\} ,$$

where ϵ is a small positive scalar (e.g. $\epsilon = 10^{-5}$) .

Choice of ϵ may affect the estimation of standard errors. The sensitivity to the choice of ϵ will be greatest when the model is highly nonlinear in which case the asymptotic approximations for standard errors will tend to be inaccurate even if exact derivatives are used.

3.5 THE FISHER SCORING ALGORITHM

Consider a twice differentiable continuous discrepancy function $\mathrm{F}(\gamma)$, where γ is a kx1 vector of unknown parameters to be estimated. An estimate $\hat{\gamma}$ of γ is obtained as the solution of the equations

$$\frac{\partial \mathrm{F}(\gamma)}{\partial \gamma_s} = 0 , \qquad s=1,2,\cdots,\mathrm{k}. \tag{3.5.1}$$

A necessary condition for $\mathrm{F}(\hat{\gamma})$ to attain a minimum value is that the Hessian matrix with typical element (cf. (4.3.28)) $\dfrac{\partial^2 \mathrm{F}(\hat{\gamma})}{\partial \gamma_r \partial \gamma_s}$, $r,s=1,2,\cdots,\mathrm{k}$ be positive definite at the point $\gamma=\hat{\gamma}$.

In Sections 3.2 and 3.4 two types of discrepancy functions were considered namely $\mathrm{F}=-2ln\mathrm{L}$ (cf. (3.2.26), where L denotes the likelihood function of a set of random variables, and $\mathrm{F}_{\mathrm{GLS}}= (\mathbf{u}-\mathrm{E}(\mathbf{u}))'\boldsymbol{\Omega}^{-1}(\mathbf{u}-\mathrm{E}(\mathbf{u}))$ (cf. (3.4.10)) where $\boldsymbol{\Omega}$ denotes the covariance matrix of the vector variate $\mathbf{u}$.

Many approaches to the solution of the minimizing equations (3.5.1) have been used including the method of steepest descent (e.g., Everett 1987), the Newton-Raphson method (e.g. Burden and Faires, 1985), the Davidson-Fletcher-Powell approach (Davidson, 1959; Fletcher and Powell 1963) and the Gauss-Newton method (e.g., Lee and Jennrich, 1979; Cudeck Klebe and Henly, 1991). The Gauss-Newton method is robust to poor starting values, usually converges rapidly, and produces consistent standard errors of the estimated parameters.

The optimization method described in this section is given by Browne and Du Toit (1992) and is based on the so-called Fisher scoring algorithm. Fisher scoring algorithms require the gradient vector of the discrepancy function and use $E(\frac{\partial^2 F}{\partial\gamma_r \partial\gamma_s})$ as approximations to the elements of the Hessian matrix. Denote the gradient vector by $\boldsymbol{g}(\gamma)$, and the approximate Hessian matrix by $\boldsymbol{H}(\gamma)$.

Suppose that γ_k is the k-th approximation to the $\hat{\gamma}$ which minimizes $F(\gamma)$. Let

$$\boldsymbol{g}_k = \boldsymbol{g}(\gamma_k), \qquad \boldsymbol{H}_k = \boldsymbol{H}(\gamma_k), \quad \text{and } F_k = F(\gamma_k). \tag{3.5.2}$$

The next approximation is obtained from

$$\hat{\gamma}_{k+1} = \hat{\gamma}_k + \alpha_k \boldsymbol{\delta}_k, \tag{3.5.3}$$

where

$$\hat{\boldsymbol{\delta}}_k = -\boldsymbol{H}_k^{-1}\boldsymbol{g}_k\ , \tag{3.5.4}$$

and α_k is a step size parameter (initially 1), required to give the minimum. If $F_{k+1} >$ F_k then repeat (3.5.4) with $\alpha_k = \alpha_k/2$, until $F_{k+1} \leq F_k$, or until q step halvings has been executed.

Agresti (1990) pointed out that the Fisher scoring method resembles the Newton-Raphson method, the distinction being that Fisher scoring uses the expected value of the second order derivative matrix. In the case of structured means and covariances, the Fisher scoring algorithm may be regarded as a sequence of Gauss-Newton steps with quantities to be fitted as well as the weight matrix changing at each step.

A convenient feature of the Fisher scoring algorithm is that an estimate, $\{\boldsymbol{H}(\hat{\gamma})\}^{-1}$ of the asymptotic covariance matrix of estimators γ is available on convergence as a by-product of the calculations.

Although constrained estimation is not used in this dissertation it may sometimes be necessary to minimize $\mathrm{F}(\gamma)$ subject to r nonlinear constraints of the form

$$\mathbf{c}(\gamma) = \mathbf{0} , \tag{3.5.5}$$

where $\mathbf{c}(\gamma)$ is a continuously differentiable rx1 vector valued function of γ.

Let $\mathbf{c}_k = \mathbf{c}(\gamma_k)$ and $\boldsymbol{L}_k = \boldsymbol{L}(\gamma_k)$. Then the linear Taylor approximation for the constraint function is

$$\mathbf{c}(\gamma) \simeq \mathbf{c}_k + \boldsymbol{L}_k\boldsymbol{\delta} , \tag{3.5.6}$$

where $\boldsymbol{\delta} = \gamma - \gamma_k$ is a kx1 vector. A typical element of the rxk Jacobian matrix $\boldsymbol{L}_k$ is given by

$$[\boldsymbol{L}_k]_{i,j} = \frac{\partial \mathbf{c}_i}{\partial \gamma_j} \mid \gamma = \gamma_k , \tag{3.5.7}$$

where $\mathbf{c}_k = [\mathbf{c}(\gamma)]_i$. Consequently the nonlinear constraints (3.5.5) may be approximated by the linear constraints

$$\boldsymbol{L}_k\boldsymbol{\delta} = -\mathbf{c}_k . \tag{3.5.8}$$

The increment vector $\boldsymbol{\delta}_k$ is obtained (Browne and Du Toit, 1992) as the solution of

$$\begin{bmatrix} \boldsymbol{\delta}_k \\ \boldsymbol{\lambda}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_k + \boldsymbol{L}_k'\boldsymbol{D}_k\boldsymbol{L}_k & \boldsymbol{L}_k' \\ \boldsymbol{L}_k & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} -(\boldsymbol{g}_k + \boldsymbol{L}_k'\boldsymbol{D}_k\mathbf{c}_k) \\ -\mathbf{c}_k \end{bmatrix}, \tag{3.5.9}$$

where $\boldsymbol{\lambda}_k$ is a rx1 vector of Lagrange multipliers and $\boldsymbol{D}_k$ is an arbitrary nonnegative definite matrix. The scaling matrix $\boldsymbol{D}_k$ does not affect the solution and is often chosen to be the null matrix (Gill, Murray and Wright, 1981, Section 5.4). The next approximation γ_{k+1} for $\hat{\gamma}$ is obtained from

$$\gamma_{k+1} = \gamma_k + \alpha_k \delta_k \ ,$$

where α_k is chosen initially as 1 and is halved successively until

$$F_b + 2\sum_{i=1}^{r} |[\lambda_k]_i \ [c_{k+1}]_i| < F_a + 2\sum_{i=1}^{r} |[\lambda_k]_i \ [c_k]_i| \qquad (3.5.10) \ ,$$

where for a specific model

$$F_b = F(\gamma_{k+1}) \qquad (3.5.11)$$

and

$$F_a = F(\gamma_k) \ . \qquad (3.5.12)$$

If no constraints are imposed, all terms involving $\mathbf{c}_k$ and $\boldsymbol{L}_k$ are omitted.

It can happen that the matrix to be inverted in (3.5.9) is singular or near singular. An adaptation of the Jennrich and Sampson (1968) stepwise regression procedure may be used to obtain an appropriate conditional inverse. Their procedure for imposing bounds on the estimates may also be employed.

3.6 THE EM ALGORITHM.

The EM algorithm is directed at finding a value of the unknown parameters which maximizes the function given the observed values. It does so by making essential use of the associated family. Each iteration of the EM algorithm involves two steps which is known (Dempster Laird & Rubin, 1977) as the expectation step (E-step) and the maximization step (M-step).

The EM iterations based on repeated substitutions provide a rapid, robust method for obtaining close approximations to the marginal maximum likelihood estimates discussed in Section 3.3 of this chapter.

Linear model

For the *linear* model initial estimates of $E(\mathbf{x}_i|\mathbf{y}_i)$ (cf. (3.3.17)) and $Cov(\mathbf{x}_i|\mathbf{y}_i)$ (cf. (3.3.18)) may be otained by setting $\sigma^2=1$, $\boldsymbol{\theta}=\mathbf{0}$ and $\boldsymbol{\Phi}=\mathbf{0}$. These values are then used to solve for $\hat{\boldsymbol{\theta}}$ (cf. (3.3.13)), $\hat{\boldsymbol{\Phi}}$ (cf. (3.3.14)) and $\hat{\sigma}^2$ (cf. (3.3.16)). At iteration 2, the estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Phi}}$ and $\hat{\sigma}^2$ replace the initial estimates. The iterative procedure is then repeated until convergence is attained.

Nonlinear model

Consider the *nonlinear* model

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i) + \epsilon_i, \qquad i=1,2,\cdots,N \qquad (3.6.1)$$

where $\mathbf{x}_1,\cdots,\mathbf{x}_N$ are assumed to be independent and identically normally distributed with mean $\boldsymbol{\theta}$ and covariance matrix, $\boldsymbol{\Phi}$. Initial estimates may be obtained by approximating the model using a first order Taylor series expansion.

A first order Taylor series expansion of $\mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i)$ about the mean yields:

$$\mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i) \simeq \mathbf{f}(\boldsymbol{\theta}) + \boldsymbol{J}_i(\mathbf{x}_i-\boldsymbol{\theta}), \qquad i=1,2,\cdots,N \qquad (3.6.2)$$

where the $n_i$xr matrix $\boldsymbol{J}_i$ has typical element

$$[\boldsymbol{J}_i]_{j,k} = \frac{\partial f(\mathbf{x}_i,t_{ij})}{\partial x_{ik}}\Big|_{\mathbf{x}_i=\boldsymbol{\theta}} \qquad j=1,2,\cdots,n_i, \quad k=1,2,\cdots,r \qquad (3.6.3)$$

and $f(\mathbf{x}_i,t_{ij})$ denotes a typical element of the n_ix1 vector $\mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i)$ of nonlinear response functions.

From (3.6.1) and (3.6.2) it follows that

$$\mathbf{y}_i^* \simeq J_i \mathbf{x}_i^* + \epsilon_i, \qquad i=1,2,\cdots,\mathrm{N} \qquad (3.6.4)$$

where

$$\mathbf{y}_i^* = \mathbf{y}_i - \mathbf{f}(\boldsymbol{\theta}, t_i) \qquad (3.6.5)$$

and

$$\mathbf{x}_i^* = \mathbf{x}_i - \boldsymbol{\theta}, \qquad i=1,2,\cdots,\mathrm{N}\ . \qquad (3.6.6)$$

Denote the mean of $\mathbf{x}_i^*$ by $\boldsymbol{\theta}^*$, then from (3.6.6) and the distributional assumptions given for $\mathbf{x}_i$, it follows that $\boldsymbol{\theta}^*=\mathbf{0}$. After each iteration of the EM-algorithm $\mathbf{y}_i^*$ and the "design" matrix J_i (cf. (3.6.3)) are updated using the new estimate $\boldsymbol{\theta}_{k+1}$ of $\boldsymbol{\theta}$, where

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \hat{\boldsymbol{\theta}}^*_{k+1}, \qquad k=0,1,2,\cdots.$$

Prior to the iterative procedure, $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\Phi}}$ are set to $\mathbf{0}$ and I respectively. The number of iterations required for convergence may be drastically reduced if the elements of $\boldsymbol{\theta}_0$ are set to reasonable values, for example the ordinary least squares estimators of the unknown parameters of the fixed parameter model $\mathbf{f}(\boldsymbol{\theta}, t_i)$.

Note that the use of a first order Taylor series expansion to linearise (3.6.2) differs from the way in which it was used by Browne and Du Toit (1991). In their approach, an equal number of measurements were assumed for each experimental unit, furthermore, these measurements were made on the same occasions, say $t_1, t_2, \cdots, t_n$.

It should also be pointed out that the linearisation process described above may be incorporated in a Gauss-Newton optimization procedure.

The theoretical principles given in this section are implemented in the Fortran program EMLIN, and an illustration involving weight measurements on female and male mice is given in Section 3.8.

Remarks

1) It is a good strategy to continue with an optimization procedure which utilizes second order derivative information after, say 50 to 100 EM iterations, as convergence of this process is very slow (Longford 1987) when approaching the solution point. An example of such an optimization procedure is the Fisher-scoring method (see Section 3.5). To use the Fisher-scoring method or any Newton algorithm an estimate of the information matrix for the elements of $\boldsymbol{\theta}$, $\boldsymbol{\Phi}$ and σ^2 is required. In the addendum to this chapter (cf. Section 3.9) the expressions for the information matrix (see Bock, 1990) for the elements of $\boldsymbol{\theta}$, $\boldsymbol{\Phi}$ and σ^2 is given.

2) Bock (1990) has extended the random regressions model to incorporate so-called second stage covariates. The extended model is given by (cf. (3.2.1)) $\mathbf{y}_i = \boldsymbol{B}_i \mathbf{x}_i + \epsilon_i$, however, it is now assumed that

$$E(\mathbf{x}_i) = \boldsymbol{\theta} + \boldsymbol{\Gamma} \mathbf{z}_i \ ,$$

where $\boldsymbol{\Gamma}$ is an r x q matrix of second-stage regression coefficients and $\mathbf{z}_i$ is a q x 1 vector of covariates. Illustrations of the extended model is given in Examples 3.8.4 and 3.8.5 respectively. This model will be discussed in more detail in Section 4.3 of Chapter 4.

3.7 GAUSS-QUADRATURE INTEGRATION TECHNIQUE

When closed form solutions for integrals such as (3.2.37) and (3.2.38) do not exist, they may be evaluated to a high degree of accuracy by means of the Gauss quadrature numerical integration procedure (e.g. Du Toit 1977, and Press, Flannery, Teukolsky & Vetterling 1989).

Consider the integral

$$\int_a^b w(x) f(x) dx, \qquad (3.7.1)$$

where $w(x)$ is called the weight function and [a,b] is any finite or infinite segment of the real line. When other methods for calculating (3.7.1) fail, the integral may be evaluated approximately by means of a linear combination of the values of f(x), so that

$$\int_a^b w(x)f(x)dx \simeq \sum_{i=1}^{q} w_i f(x_i) , \qquad (3.7.2)$$

where the w_i are called coefficients (or weights) and the x_i the points (or nodes) of the function. Formulae of the form (3.7.2) are called quadrature formulae.

For fixed q, the formula (3.7.2) contains 2q parameters w_i and x_i, $i=1,2,\cdots,q$. It is possible to find q points and coefficients which will make (3.7.2) exact whenever f(x) is a polynomial of degree not greater than $2q-1$ provided that:

(i) w(x) is nonnegative in [a,b]
(ii) w(x) is a function for which all moments

$$c_r \equiv \int_a^b w(x)x^r dx , \qquad r=0,1,2,\cdots \qquad (3.7.3)$$

are defined and are finite.

(iii) $c_0 > 0$.

Such formulae are usually called Gaussian quadrature formulae. A systematic introduction to the theory of Gaussian quadrature is given by Krylov (1962) and by Stroud and Secrest (1966).

Suppose that the weight function w(x) is given by

$$w(x) = e^{-x^2} \qquad (3.7.4)$$

and that [a,b] is the real line, then (3.7.2) may be written as

$$\int_{-\infty}^{+\infty} e^{-x^2} f(x)dx \simeq \sum_{i=1}^{q} w_i f(x_i) \ . \tag{3.7.5}$$

Tables are available for the different values of w_i and x_i for $i = 2,3,\cdots, 60$ (See Stroud and Secrest, 1966).
A formula of the type (3.7.2) is said to have degree of exactness m if it is exact whenever f(x) is a polynomial of degree not greater than m. Stroud (1971) considers the following formula

$$\int_{R_k} \cdots \int w(x_1,\cdots,x_k) f(x_1,\cdots,x_k) dx_1 \cdots dx_k \simeq \sum_{i=1}^{N} B_i f(v_{i,1},\cdots,v_{i,k}) \ , \tag{3.7.6}$$

where $w(x_1,\cdots,x_k)$ is a given weight function, the $(v_{i,1},\cdots,v_{i,k})$ are points which lie in the k – dimensional Euclidean space, E_k, $R_k \subset E_k$, and the B_i are called the weights or coefficients of the formula. Tabulations of the abscissas and weights can also be found in standard references such as Abramovitz and Stegun (1965) and Ralston and Rabinowitz (1978). Formula (3.7.5) is said to have degree of exactness d if it is exact for any linear combination of monomials

$$x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_k^{\alpha_k}, \tag{3.7.7}$$

where the α_i are nonnegative integers,

$$0 \le \alpha_1 + \alpha_2 + \cdots + \alpha_k \le d,$$

and there is at least one linear combination of monomials of degree $d+1$ for which (3.7.6) is not exact.

Consider the integral

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x_1^2 - x_2^2} f(x_1,x_2) dx_1 dx_2 \ . \tag{3.7.8}$$

It is possible to construct a product-type Gaussian quadrature formula from (3.7.5) for the above integral as follows:

Since

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x_1^2-x_2^2} f(x_1,x_2)dx_1dx_2 = \int_{-\infty}^{\infty} e^{-x_2^2}\left(\int_{-\infty}^{\infty} e^{-x_1^2} f(x_1,x_2)dx_1\right)dx_2 , \qquad (3.7.9)$$

it follows that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x_1{}^2-x_2{}^2} f(x_1,x_2)dx_1dx_2 = \sum_{j=1}^{n} w_j\left(\sum_{i=1}^{n} w_i\, f(x_i,x_j)\right)$$
$$= \sum_{m=1}^{N^*} B_m\, f(v_{m1},v_{m2}) , \qquad (3.7.10)$$

where

$$N^* = n^2,$$
$$B_m = w_j \cdot w_i ,$$
$$(v_{m1},v_{m2}) = (x_i,x_j) , \quad m=n(i-1)+j \qquad (3.7.11)$$

The weights B_m and points (v_{m1},v_{m2}), $m=1,\cdots,N$ can therefore be constructed from the tabulated w_i and x_i.

Example 3.7.1

For q=4 it follows from the tables in Stroud and Secrest (1966) that

i	w_i	x_i
1	0,081313	1,6506801
2	0,804914	0,5246476
3	0.804914	-0,5246476
4	0,081313	-1,6506801

Let x $\sim$ N(θ,σ^2) and suppose that θ=100 and σ=15 then

$$E(x^2) = \sigma^2 + \theta^2 = (15)^2 + (100)^2 = 10225 .$$

But

$$E(x^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x^2 \, exp[-\tfrac{1}{2} (\tfrac{x-\theta}{\sigma})^2] \, dx . \qquad (3.7.12)$$

Let $u = \frac{x-\theta}{\sqrt{2}\cdot\sigma}$, hence $x = \sqrt{2}\sigma u + \theta$, and $\frac{\partial x}{\partial \theta} = \sqrt{2}\sigma$. It is thus possible to write (3.7.6) as follows:

$$E(x^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \sqrt{2}\sigma \int_{-\infty}^{+\infty} exp \, -\tfrac{1}{2}u^2 \cdot (\sqrt{2}\sigma u + \theta)^2 \, du$$

$$\simeq \frac{1}{\pi} \sum_{i=1}^{4} w_i \left(\sqrt{2}\sigma x_i + \theta\right)^2$$

$$= \tfrac{1}{\pi}[0{,}081313\left(\sqrt{2} \text{ x } 15 \text{ x } 1{,}6506801 + 100\right)^2 + 0{,}804914\left(\sqrt{2} \text{ x } 15 \text{ x } 0{,}5246476 + 100\right)^2 +$$

$$0{,}804914\left(\sqrt{2} \text{ x } 15 \text{ x } -0{,}5246476 + 100\right)^2 + 0{,}081313\left(\sqrt{2} \text{ x } 15 \text{ x } -1{,}6506801 + 100\right)^2$$

$$= 10225 .$$

3.8 PRACTICAL APPLICATIONS

Example 3.8.1

A data set comprising the weights of 42 male mice and 40 female mice were obtained from the Department of Zoology, University of Pretoria, South Africa (see Du Toit, 1979). Over a period of two years a number of female and male striped mice were released in a 36.34m^2 outdoor enclosure with 8 nest boxes and sufficient food and water, where they were allowed to multiply freely. Occurrence of birth was recorded daily and the mice were weighed weekly from the end of the 2^{nd} week until physical maturity was reached. For the male and female mice data 9 and 8 repeated weight measurements were recorded respectively.

The following nonlinear regression model (cf. (2.2.3)) was used to describe change in weight over time:

$$y_i = a(1+bc^{t_i}) + \epsilon_i , \qquad i = 1,2,\cdots,n. \tag{3.8.1}$$

with parameter transformations as follows:

$$\begin{aligned} a &= x_1 , \\ b &= exp\, x_2 , \\ c &= \frac{1}{1+exp\, x_3} , \end{aligned} \tag{3.8.2}$$

and where $\mathbf{x}_i = (x_1,x_2,x_3)' \sim N(\boldsymbol{\theta},\boldsymbol{\Phi})$.

Estimates of the unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$ are

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} 48.26378 \\ 1.72658 \\ -0.31320 \end{bmatrix},$$

$$\hat{\boldsymbol{\Phi}} = \begin{bmatrix} 131.85846 & & \\ 2.61150 & 0.15389 & \\ -2.29458 & 0.01357 & 0.15974 \end{bmatrix}.$$

A researcher may prefer to report his results in terms of the original parameter set $\mathbf{a}=(a,b,c)'$. Estimates of the moments $E(a^k b^l c^m)$, $k,l,m = 0,1,2\cdots$ may be obtained from $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Phi}}$. These estimates are then used to find estimates of $E(\mathbf{a})$ and $Cov(\mathbf{a},\mathbf{a}')$. From (3.8.2) it follows that

$$E(a^k b^l c^m) = E\left(x_1{}^k (exp\, x_2))^l (\frac{1}{1+exp\, x_3})^m \right), \qquad i = 1,2,\cdots,n. \tag{3.8.3}$$

Let $u_1=x_3$, $u_2=x_2$ and $u_3=x_1$ and denote $E(\mathbf{u})$ and $Cov(\mathbf{u},\mathbf{u}')$ by $\boldsymbol{\mu}$ and Σ respectively.

Note that Σ is a permutation of the rows and columns of $\mathbf{\Phi}$. Likewise, μ_1, is the third row of $\boldsymbol{\theta}$, μ_2 the second and μ_3 the first.

The estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ of $\boldsymbol{\mu}$ and Σ are to be employed to obtain estimates for $E(\mathbf{a})$ and $Cov(\mathbf{a},\mathbf{a}')$ respectively. The moments of the parameter set $\mathbf{a} = (a,\ b_{(s)},c)$ can be determined by calculating the triple integral

$$E(k,l,m) = E\left((u_3)^k(e^{u_2})^l(1+e^{u_1})^{-m}\right). \tag{3.8.4}$$

Let

$$h(\mathbf{u}) = u_3{}^k \cdot f(u_1,u_2), \qquad k=1,2,\cdots, \tag{3.8.5}$$

where

$$f(u_1,u_2) = (e^{u_2})^l(1+e^{u_1})^{-m} \qquad l\ ,m = 1,2,\cdots, \tag{3.8.6}$$

is a continuous function which cannot be integrated by any of the usual analytical methods. It is shown that the expected value of $h(\mathbf{u})$ can be expressed as a two dimensional integral which can be evaluated by means of a product-type Gaussian quadrature formulae. From (3.8.5) it follows that

$$E[h(\mathbf{u})] = (2\pi)^{-3/2}|\Sigma|^{-1/2}\int\int\int h(\mathbf{u})\ exp\left(-\tfrac{1}{2}(\mathbf{u}-\boldsymbol{\theta})'\Sigma^{-1}(\mathbf{u}-\boldsymbol{\theta})\right)du_1du_2du_3\ . \tag{3.8.7}$$

Proposition 3.8.1

Using Gauss quadrature the expected value of $h(\mathbf{u})$ for $k = 1$ may be approximated by

$$E[h(\mathbf{u})] \simeq C\cdot\sum_{i=1}\sum_{j=1} w_iw_j(t_{31}x_i+t_{32}x_j)(e^{t_{11}x_i+\mu_1})^l(1+e^{t_{11}x_i+t_{22}x_j+\mu_2})^{-m}. \tag{3.8.8}$$

with t_{ij} i,j=1,2,3 elements of the matrix $\boldsymbol{T}$ where $\Sigma=\boldsymbol{TT}'$.

Proof

Consider the transformation

$$\mathbf{z} = \frac{1}{\sqrt{2}}\boldsymbol{T}^{-1}(\mathbf{u}-\boldsymbol{\mu}) \ , \quad \text{where } \Sigma = \boldsymbol{T}\boldsymbol{T}',$$

then

$$\mathbf{u} = \sqrt{2}\boldsymbol{T}\mathbf{z} + \boldsymbol{\mu}$$

and the Jacobian $|\frac{\partial \mathbf{u}}{\partial \mathbf{z}'}|$ of the transformation is $(2)^{3/2}|\Sigma|^{1/2}$.

It further follows that

$$u_1 = g_1(\mathbf{z}) = t_{11}z_1 + \mu_1 \ ,$$

$$u_2 = g_2(\mathbf{z}) = t_{21}z_1 + t_{22}\, z_2 + \mu_2 \ ,$$

$$u_3 = g_3(\mathbf{z}) = t_{31}z_1 + t_{32}z_2 + t_{33}z_3 + \mu_3 \ . \qquad (3.8.9)$$

Using the transformations in (3.8.9) the integral in (3.8.7) may be expressed as

$$E[\,h(\mathbf{u})] = (\pi)^{-3/2}\int\int\int e^{-z_1^{\,2}-z_2^{\,2}-z_3^{\,2}}\big(g_3(\mathbf{z})\big)^k\Big((e^{g_2(\mathbf{z})})^l(1+e^{g_1(\mathbf{z})})-m\Big)^\alpha dz_1dz_2dz_3 \ . \qquad (3.8.10)$$

From (3.8.9)

$$[g_3(\mathbf{z})]^k = [t_{31}z_1 + t_{32}z_2 + t_{33}z_3 + \mu_3]$$

$$= [a + t_{33}z_3]^k$$

$$= \sum_{j=1}^{k}\binom{n}{j}(t_{33}z_3)^j a^{k-j}$$

where $a = t_{31}z_1 + t_{32}z_2 + \mu_3$.

For k=1 integrating over $g_3(\mathbf{z})$ results in the following double integral, which can be solved using Gauss quadrature numerical integration, discussed above.

$$E[\,h(\mathbf{u})] = C\int_{z_1}\int_{z_2} e^{-z_1^2 - z_2^2}(t_{31}z_1 + t_{32}z_2 + \mu_3)\left(\left(e^{g_1(\mathbf{z})}\right)^l\left(\frac{1}{1+e^{g_2(\mathbf{z})}}\right)^m\right)^\alpha dz_1 dz_2\ , \tag{3.8.11}$$

with

$$C = (\pi)^{-3/2}\ . \tag{3.8.12}$$

□

For the parameterization described in (3.8.2), with $\mathbf{x}_i \sim N(\boldsymbol{\theta},\boldsymbol{\Phi})$, the moments to determine $E(\mathbf{a})$ and $Cov(\mathbf{a},\mathbf{a}')$ for model (3.8.1) were calculated for the 42 male mice with the estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Phi}}$ of $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$ given above. The following is a summary of values for $E(\mathbf{a})$ and $Cov(\mathbf{a},\mathbf{a}')$ for 4, 6, 8, 10, 16 and 20 Gauss quadrature points, respectively.

No. of points	μ_a	μ_b	μ_c
4	48.26378	6.07101	0.57485
6	48.26378	6.07101	0.57485
8	48.26378	6.07101	0.57485
10	48.26378	6.07101	0.57485
16	48.26378	6.07101	0.57485
20	48.26378	6.07101	0.57485

No. of points	σ_a^2	σ_{ab}	σ_b^2	σ_{ac}	σ_{bc}	σ_c^2
4	131.85846	15.85417	6.12865	0.54044	−0.01942	0.00888
6	131.85846	15.85445	6.13165	0.54041	−0.01942	0.00888
8	131.85846	15.85445	6.13165	0.54041	−0.01942	0.00888
10	131.85846	15.85445	6.13165	0.54041	−0.01942	0.00888
16	131.85846	15.85445	6.13165	0.54041	−0.01942	0.00888
20	131.85846	15.85445	6.13165	0.54041	−0.01942	0.00888

Example 3.8.2

The dental measurement dataset (see Appendix A) was first considered by Potthoff and Roy (1964) and later analyzed by Lee and Geissser (1975), Fearn (1975), Rao (1987), Lee (1988 & 1991). Dental measurements were made on 11 girls and 16 boys at ages 8, 10, 12 and 14 years. Each measurement is the distance, in millimeters, from the center of the pituitary to the pterygomaxillary fissure. In Figures 3.8.1 and 3.8.2 the ordinary least squares curves

$$\hat{y}_{it} = \hat{x}_0 + \hat{x}_1 t$$

are plotted for each boy and girl.

Let

$$\boldsymbol{B} = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix} .$$

The model fitted to the combined dental data of the boys and girls is:

$$\mathbf{y}_i = \boldsymbol{B}\mathbf{x}_i + \epsilon_i , \qquad i=1,2,\cdots,27 , \qquad (3.8.13)$$

where it is assumed that the uncorrelated vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{27}$ are distributed as:

$$\mathbf{x}_i \sim \mathrm{N}(\boldsymbol{\theta}_1, \boldsymbol{\Phi}) , \qquad i=1,2,\cdots,11$$

$$\mathbf{x}_i \sim \mathrm{N}(\boldsymbol{\theta}_2, \boldsymbol{\Phi}) . \qquad i=12, 13, \cdots,27$$

It is further assumed that the ϵ_i, $i=1,2,\cdots,N$ are independent and identically distributed as $\mathrm{N}(\mathbf{0},\sigma^2 \boldsymbol{I})$ independent of $\mathbf{x}_1, \cdots, \mathbf{x}_{27}$.

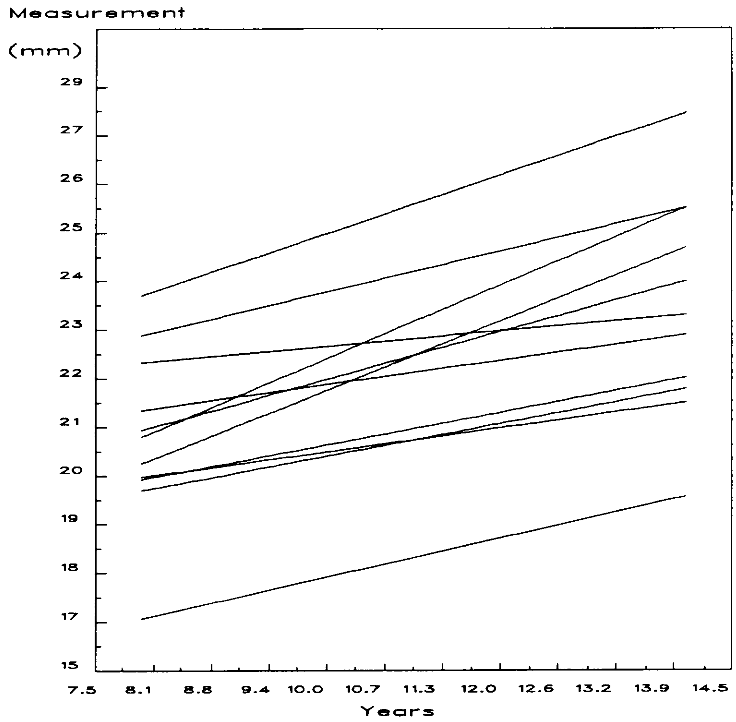

Figure 3.8.1 OLS Regression lines of distance (mm) on years for girls

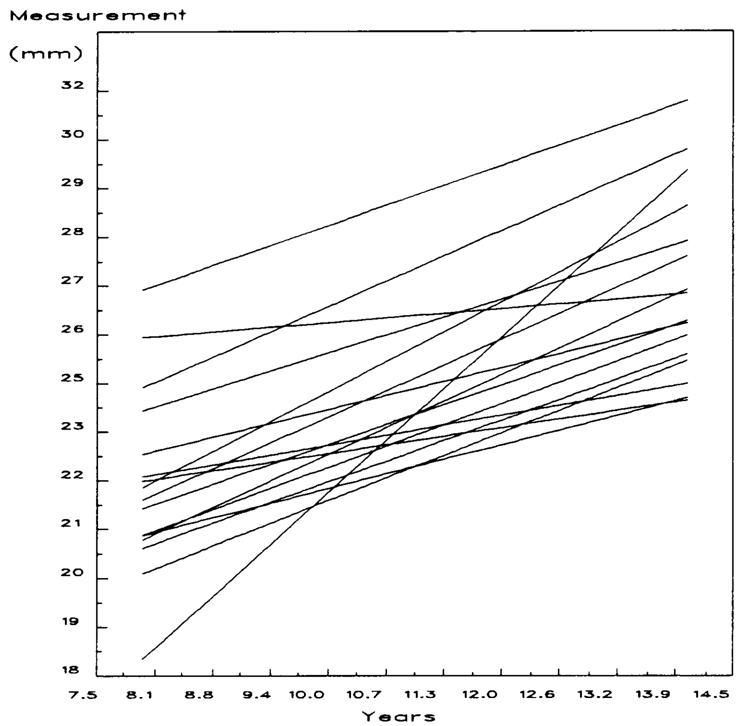

Figure 3.8.2 OLS Regression lines of distance(mm) on years for boys

Estimates of the unknown parameters are as follows:

	θ_{10}	θ_{11}	θ_{20}	θ_{21}	ϕ_{11}	ϕ_{21}	ϕ_{22}	σ^2
Estimate :	16.34063	0.78437	17.37273	0.47955	4.55690	-0.19825	0.02376	1.71621
Std. error:	0.98008	0.08275	1.18202	0.09980	4.67188	0.37906	0.03409	0.33028

In Figure 3.8.3 a graphical representation of the fitted model (3.8.13) is given where the solid line and dotted lines indicate the estimated mean curves for the 16 boys and 11 girls respectively. The filled triangles indicate the sample means of the boys and the unfilled triangles indicate the sample means of the girls at the ages 8, 10, 12 and 14 years respectively. The sample means are as follows:

Age	$\overline{y}_{boys}$	$\overline{y}_{girls}$	*Age*	$\overline{y}_{boys}$	$\overline{y}_{girls}$
8	22.875	21.182	10	23.813	22.227
12	25.719	23.091	14	27.469	24.091

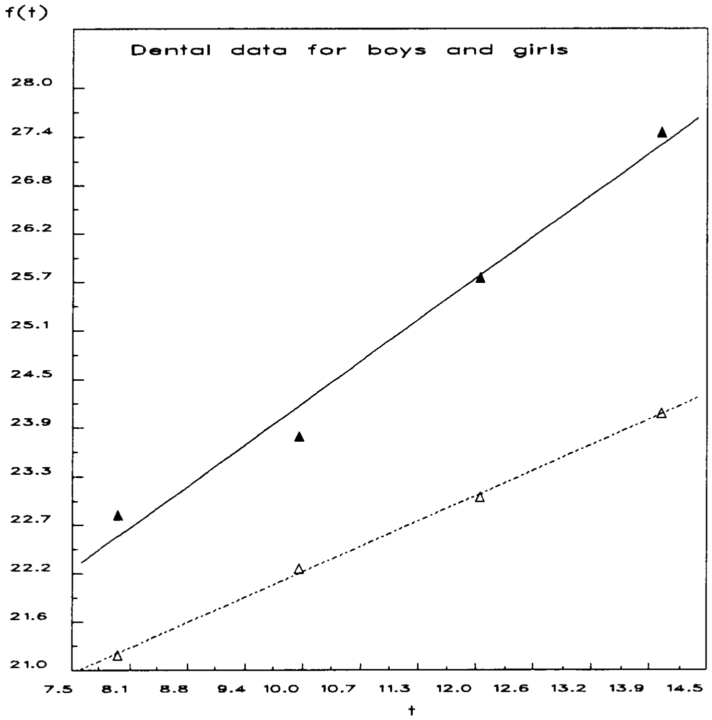

Figure 3.8.3 Fitted models for boys (solid line) and girls (dotted line), as well as sample means for the boys (unfilled triangle) and girls (filled triangle) .

An alternative formulation (cf. remark in Section 3.6) for model (3.8.3) is:

$$\mathbf{x}_i \;= \boldsymbol{\theta} + \boldsymbol{v} z_i + \mathbf{u}_i \;,$$

where it is assumed that $\mathbf{u}_1, \cdots, \mathbf{u}_{27}$ are independent and indentically distributed $N(\mathbf{0},\mathbf{\Phi})$ deviates and where the covariate z represents sex (z=−1 denotes a girl and z=1 denotes a boy) therefore,

$$z_i = -1 \;;\quad i=1,2,\cdots,11 \quad \text{or} \quad z_i = 1; \quad i=12,13,\cdots,27 \;.$$

Estimates of the unknown parameters are as follows:

	θ_1	θ_2	v_1	v_2	ϕ_{11}	ϕ_{21}	ϕ_{22}	σ^2
Estimate:	16.85668	0.63196	-0.51605	0.15241	4.55690	-0.19825	0.02376	1.71621
Std. error:	0.76775	0.06482	0.76775	0.06482	4.67188	0.37906	0.03409	0.33028

For the random parameter model (3.8.1) the Chi-square likelihood ratio test statistic yielded $\chi^2 = 11.30$, with 10 degrees of freedom, and exceedence probability $p=0.335$.

The fit of the fixed parameter model

$$\mathbf{y}_i \;= \boldsymbol{B\theta}_j + \epsilon_i \;, \qquad j=1, \quad i=1,2,\cdots,11$$
$$j=2, \quad i=12,13,\cdots,27 \qquad (3.8.14)$$

resulted in a Chi-square value of, $\chi^2 = 61.73$, with 13 degrees of freedom.

Denote the χ^2-values for the fitted and random parameter models by χ^2_{FIX} and χ^2_{RAN} respectively. From the information given above it follows that

$$\chi^2_{FIX} - \chi^2_{RAN} \;=\; 50.43, \text{ with 3 degrees of freedom and } p=0.000.$$

The highly significant results indicates a random parameter regression model, contradicting the conclusion that may be drawn from the reported standard errors of the estimated ϕ-values. Note that the expressions for the standard errors of maximum likelihood estimators are based on asymptotic theory. Hence, when dealing with the small sample case, the standard errors will often be inaccurate.

Example 3.8.3

The following model was fitted to the 40 female mice data discussed in Example 3.8.1,

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i) + \boldsymbol{\epsilon}_i\ , \qquad i=1,2,\cdots,40 \qquad (3.8.15)$$

where it is assumed that $\mathbf{x}_1$, $\mathbf{x}_2,\cdots,\mathbf{x}_N$ are a random sample of an r-dimensional vector $\mathbf{x}$ which has a $N(\boldsymbol{\theta},\boldsymbol{\Phi})$ distribution, furthermore the error vectors $\boldsymbol{\epsilon}_1,\boldsymbol{\epsilon}_2,\cdots,\boldsymbol{\epsilon}_N$ are assumed to be independent and identically distributed as $N(\mathbf{0},\sigma^2\boldsymbol{I})$ variates, uncorrelated with the $\mathbf{x}_i$.

A typical element of $\mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i)$ is

$$f(\mathbf{x}_i,t_{ij}) = \frac{x_1}{1+exp\,x_2(1+exp\,x_3)^{-t_{ij}}}\ , \qquad t_{ij}=j, \qquad j=1,2,\cdots,8; \quad i=1,2,\cdots,40. \qquad (3.8.16)$$

Using the modified EM-algorithm discussed in Section 3.3 convergence was reached in 36 iterations and the estimates of the unknown parameters are as follows:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} 41.365 \\ 1.427 \\ -0.520 \end{bmatrix}, \quad \hat{\boldsymbol{\Phi}} = \begin{bmatrix} 59.428 & & \\ 1.287 & 0.107 & \\ -0.389 & 0.024 & 0.086 \end{bmatrix} \text{ and } \hat{\sigma}^2 = 2.558\ .$$

The same model was fitted to the male mice data set of 42 observations discussed in

Example 3.8.1 with 9 equally spaced measurements per observation. Convergence was reached in 21 iterations and the estimates of the unknown parameters are as follows:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} 46.043 \\ 1.585 \\ -0.391 \end{bmatrix}, \quad \hat{\boldsymbol{\Phi}} = \begin{bmatrix} 95.884 & & \\ 2.242 & 0.172 & \\ -1.485 & 0.039 & 0.154 \end{bmatrix} \text{ and } \hat{\sigma}^2 = 2.933 \ .$$

Example 3.8.4

The following model, with sex as covariate, was fitted for the male and female mice data (cf. Example 3.8.3):

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i, \boldsymbol{t}_i) + \boldsymbol{\epsilon}_i \ , \qquad i=1,2,\cdots,82 \qquad (3.8.17)$$

where it is assumed that

$$\mathbf{x}_i = \boldsymbol{\theta} + \boldsymbol{v}\, \mathrm{z}_i + \mathbf{u}_i \ , \qquad i=1,2,\cdots,82 \qquad (3.8.18)$$

with $\mathrm{z}_i=1$ for males, $\mathrm{z}_i=-1$ for females.

It is further assumed that $\mathbf{u}_1,\cdots,\mathbf{u}_N$ are independent and identically distributed as $N(\mathbf{0},\boldsymbol{\Phi})$ variates, independent of the uncorrelated error vectors $\boldsymbol{\epsilon}_1,\boldsymbol{\epsilon}_2,\cdots,\boldsymbol{\epsilon}_N$ with $\boldsymbol{\epsilon}_i \sim N(\mathbf{0},\sigma^2\boldsymbol{I}_{n_i})$.

A typical element of $\mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i)$ is

$$f(\mathbf{x}_i,t_{ij}) = \frac{\mathrm{x}_1}{1+exp\,\mathrm{x}_2(1+exp\,\mathrm{x}_3)^{-t_{ij}}} \ , \qquad t_{ij}=j, \quad \begin{array}{ll} j=1,2,\cdots,8; & i=1,2,\cdots,42 \\ j=1,2,\cdots,9; & i=43,\cdots,82. \end{array} \qquad (3.8.19)$$

The EM algorithm using the first order Taylor series expansion method described in Section 3.3 converged in 22 iterations yielding the following estimates of the unknown parameters:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} 43.674 \\ 1.509 \\ -0.450 \end{bmatrix}, \quad \hat{\boldsymbol{\upsilon}} = \begin{bmatrix} 2.355 \\ 0.080 \\ 0.064 \end{bmatrix}, \quad \hat{\boldsymbol{\Phi}} = \begin{bmatrix} 81.218 & & \\ 1.859 & 0.146 & \\ -1.116 & 0.031 & 0.132 \end{bmatrix} \text{ and } \hat{\sigma}^2 = 2.754 .$$

Example 3.8.5

The data used in this example is described by Kanfer and Ackerman (1989). The raw data was obtained directly from Professors Kanfer and Ackerman and is used here with their permission. The data description (see Browne and Du Toit (1992)) is as follows: Each of the 141 subjects was a U.S. Air Force enlisted personnel who carried out a computerized air traffic controller task developed by Kanfer and Ackerman (1989, pp 666-669). They were instructed to accept planes into their hold pattern and land them safely and efficiently on one of four runways (varying in length and compass direction) according to rules governing plane movements and landing requirements. For each subject the success of a series of six 10-minute trials was recorded. The measurement employed was the number of correct landings per trial, yielding six scores. The Armed Services Vocational Aptitude Battery (ASVB) see Wilfgong, 1980 was also administered to each subject. A global measure of cognitive ability, obtained from the sum of scores on the 10 subscales, was used as a covariate in the analysis.

The following model was fitted to the data:

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i, \boldsymbol{t}_i) + \boldsymbol{\epsilon}_i \,, \qquad i=1,2,\cdots,141 \qquad (3.8.20)$$

where it is assumed that the vector of random parameters, $\mathbf{x}_i$, $i=1,2,\cdots,141$ is as follows

$$\mathbf{x}_i = \boldsymbol{\theta} + \boldsymbol{v}\, z_i + \mathbf{u}_i\ , \qquad i=1,2,\cdots,141, \qquad (3.8.21)$$

the covariate, z_i, being a measure of cognitive ability. It is further assumed that $\mathbf{u}_1,\cdots,\mathbf{u}_N$ are independent and identically distributed as $N(\mathbf{0},\boldsymbol{\Phi})$ variates, independent of the uncorrelated error vectors $\epsilon_1,\epsilon_2,\cdots,\epsilon_N$ with $\epsilon_i \sim N(\mathbf{0},\sigma^2 \boldsymbol{I}_{n_i})$. The vectors $\mathbf{x}_1$, $\mathbf{x}_2,\cdots,\mathbf{x}_N$ are therefore assumed to be random samples of an r-dimensional vector $\mathbf{x}$ which has a $N(\boldsymbol{\theta}+\boldsymbol{v}\mathbf{z},\ \boldsymbol{\Phi})$ distribution.

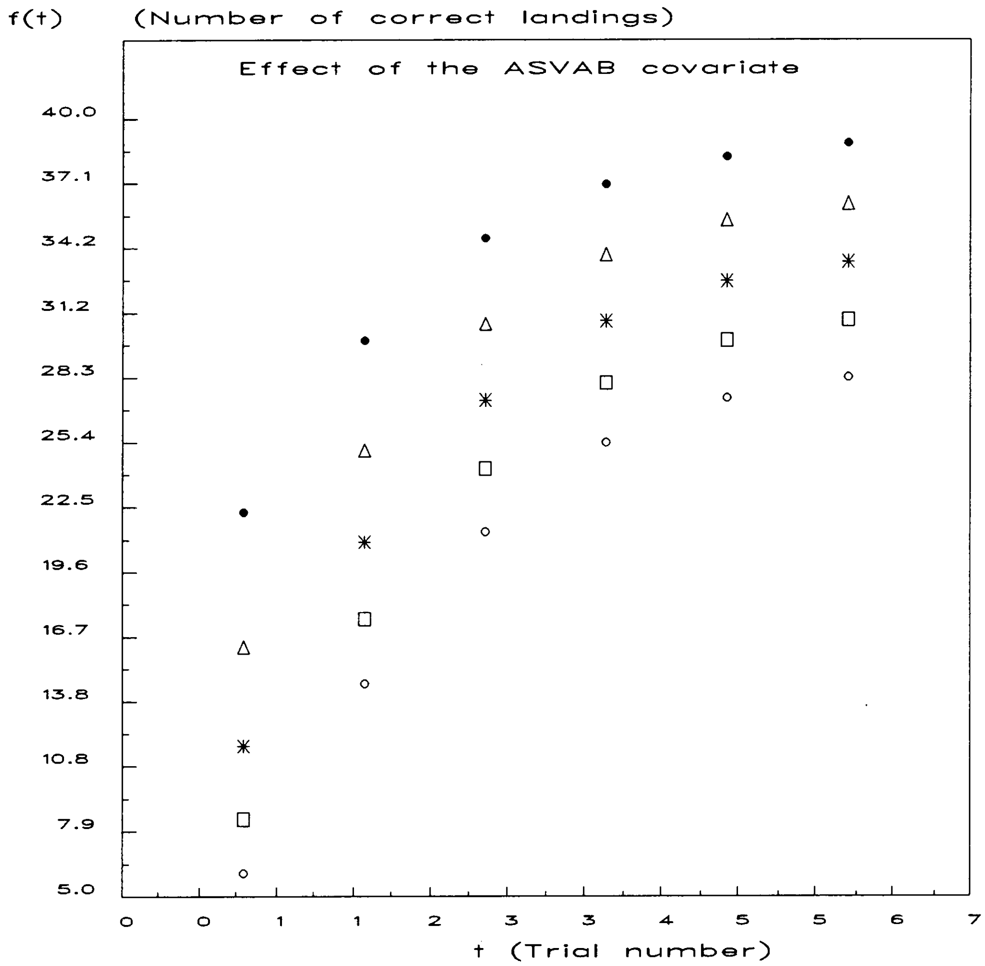

Figure 3.8.4 The effect of the covariate ability **Key to figure:** **o : z = −200** **□ : z = −100** **∗ : z = 0** **Δ : z = 100** **• : z = 200**

A typical element of $\mathbf{f}(\mathbf{x},\boldsymbol{t}_i)$ is assumed to be a Gompertz function,

$$\mathrm{f}(\mathrm{x}_i,t_{ij}) = \mathrm{x}_1 exp\Big(-\mathrm{x}_2 - exp\,(t_{ij}-1)\,\mathrm{x}_3\Big) \qquad t_{ij}=j, \qquad j=1,2,\cdots,6; \tag{3.8.22}$$

The function (3.8.22) employs the same parameterization as that used by Browne and Du Toit (1992).

Using the estimated parameter values reported by Browne and Du Toit as initial values the EM-algorithm described in Section 3.6 converged in 17 iterations yielding the following estimates of the unknown parameters:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} 34.2922 \\ 1.0713 \\ 0.7756 \end{bmatrix}, \; \hat{\boldsymbol{v}} = \begin{bmatrix} 0.0258 \\ -0.0025 \\ -0.0002 \end{bmatrix}, \; \hat{\boldsymbol{\Phi}} = \begin{bmatrix} 84.9402 & & \\ 0.5978 & 0.2006 & \\ -0.4503 & 0.0003 & 0.0955 \end{bmatrix} \text{ and } \hat{\sigma}^2 = 8.1023 \,.$$

In Figure 3.8.4 the effect of the covariate is illustrated for values of z equal to -200, -100, 0, 100 and 200, respectively.

ADDENDUM

3.9 THE INFORMATION MATRIX

The unknown parameters in the density function $\mathrm{g}(\mathbf{x})$ defined in (3.3.2) are the r x 1 vector $\boldsymbol{\theta}$, as well as the $\frac{1}{2}\mathrm{r}(\mathrm{r}+1)$ nonduplicate elements of $\boldsymbol{\Phi}$, thus

$$\boldsymbol{\varrho}' \;=\; (\boldsymbol{\theta}',(\mathrm{vecs}\boldsymbol{\Phi})') \,. \tag{3.9.1}$$

The vector $\boldsymbol{\tau}$ denotes the unknown parameters of $\mathrm{f}(\mathbf{y}_i|\mathbf{x})$ and it is assumed that

$$\boldsymbol{\Lambda}_i \;=\boldsymbol{\Lambda}_i(\boldsymbol{\tau}) \,. \tag{3.9.2}$$

From Section 3.3, Chapter 3 it follows that

$$ln\, g(\mathbf{x}) \;=\; -\tfrac{r}{2}\, ln 2\pi - \tfrac{1}{2} ln\,|\,\boldsymbol{\Phi}\,| - \tfrac{1}{2}\;\; tr\, \boldsymbol{\Phi}^{-1}(\mathbf{x}-\boldsymbol{\Theta})(\mathbf{x}-\boldsymbol{\Theta})'. \tag{3.9.3}$$

$$ln\, f(\mathbf{y}_i\,|\,\mathbf{x}) \;=\; -\tfrac{n_i}{2} ln\, 2\pi - \tfrac{1}{2}|\Lambda_i| - \tfrac{1}{2} tr\, \boldsymbol{\Lambda}_i^{-1}\boldsymbol{G}_{\mathbf{y}_i}\,, \tag{3.9.4}$$

with

$$\boldsymbol{G}_{\mathbf{y}_i} \;=\; (\mathbf{y}_i - \boldsymbol{B}_i\mathbf{x})(\mathbf{y}_i - \boldsymbol{B}_i\mathbf{x})'. \tag{3.9.5}$$

Let

$$\boldsymbol{\Lambda}_i \;=\; \sigma^2 \boldsymbol{I}_{n_i} \tag{3.9.6}$$

It has been shown in Chapter 3 that

$$\frac{\partial\, ln L}{\partial\, \sigma^2} \;=\; \tfrac{1}{2}\sum_{i=1}^{N} E_{\mathbf{x}\,|\,\mathbf{y}_i}\Big(tr\ \sigma^{-4}\ (\hat{\boldsymbol{G}}_{\mathbf{y}_i} - \hat{\sigma}^2 \boldsymbol{I}_{n_i})\Big) = 0\ . \tag{3.9.7}$$

$$\frac{\partial\, ln L}{\partial\, \boldsymbol{\Theta}} \;=\; \sum_{i=1}^{N} \boldsymbol{\Phi}^{-1} E_{\mathbf{x}\,|\,\mathbf{y}_i}(\mathbf{x} - \hat{\boldsymbol{\Theta}}) = \mathbf{0}\,, \tag{3.9.8}$$

and

$$\begin{aligned}
\frac{\partial\, ln\, g(\mathbf{x}_i)}{\partial\, \boldsymbol{\Phi}_{mn}} \;&=\; \tfrac{1}{2}\, tr\, \Big(\boldsymbol{\Phi}^{-1}(\boldsymbol{G}_{\mathbf{x}_i} - \boldsymbol{\Phi})\boldsymbol{\Phi}^{-1}(\boldsymbol{J}_{mn} + (1-\delta_{mn})\boldsymbol{J}_{nm})\Big) \\
&= \frac{(2-\delta_{mn})}{2} tr\Big(\boldsymbol{\Phi}^{-1}(\boldsymbol{G}_{\mathbf{x}_i} - \boldsymbol{\Phi})\boldsymbol{\Phi}^{-1}\boldsymbol{J}_{mn}\Big) \\
&= \frac{(2-\delta_{mn})}{2} tr\,[\boldsymbol{\Phi}^{-1}(\boldsymbol{G}_{\mathbf{x}_i} - \boldsymbol{\Phi})\boldsymbol{\Phi}^{-1}]_{m,n}
\end{aligned} \tag{3.9.9}$$

with

$$\begin{aligned}
\boldsymbol{G}_{\mathbf{x}_i} &= (\mathbf{x} - \boldsymbol{\Theta})(\mathbf{x} - \boldsymbol{\Theta})' \\
&= \Big(\big(\mathbf{x} - E(\mathbf{x}|\mathbf{y}_i)\big) - \big(\boldsymbol{\Theta} - E(\mathbf{x}|\mathbf{y}_i)\big)\Big)\Big(\big(\mathbf{x} - E(\mathbf{x}|\mathbf{y}_i)\big) - \big(\boldsymbol{\Theta} - E(\mathbf{x}|\mathbf{y}_i)\big)\Big)'\,.
\end{aligned}$$

When using any from of a Newton algorithm an estimate of the information matrix for the elements of $\boldsymbol{\theta}$, $\boldsymbol{\Phi}$ and σ^2 is required. The expressions for the information matrix are as follows:

$$\mathfrak{I}(\sigma^2) \;=\; \mathrm{E}\left(\frac{\partial \ln \mathrm{L}}{\partial \sigma^2}\right)^2$$

From (3.9.7)

$$\mathfrak{I}(\sigma^2) \;=\; \sum_{i=1}^{\mathrm{N}} \mathrm{E}\left(\tfrac{1}{2}\ \mathrm{E}_{\mathbf{x}\,|\,\mathbf{y}_i}\left(tr\ \sigma^{-4}\ (\hat{\boldsymbol{G}}_{\mathbf{y}_i} - \hat{\sigma}^2 \boldsymbol{I}_{\mathrm{n}_i})\right)\right)^2$$

with $\boldsymbol{G}_{\mathbf{y}_i}$ given in (3.9.5), thus

$$\begin{aligned}\mathfrak{I}(\sigma^2) &= \tfrac{1}{4}\sigma^{-8}\sum_{i=1}^{\mathrm{N}} \mathrm{E}\left(tr\left(\mathbf{y}_i\mathbf{y}_i' - (\hat{\sigma}^2\boldsymbol{I}_{\mathrm{n}_i} - \boldsymbol{B}_i\mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i)\boldsymbol{B}_i')\right)^2\right)\\ &= \tfrac{1}{2}\sigma^{-8}\sum_{i=1}^{\mathrm{N}} tr\left(\hat{\sigma}^2\boldsymbol{I}_{\mathrm{n}_i} - \boldsymbol{B}_i\mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i)\boldsymbol{B}_i'\right)^2 . \qquad (3.9.10)\end{aligned}$$

From (3.9.9)

$$\frac{\partial \ln \mathrm{g}(\mathbf{x}_i)}{\partial\, \mathrm{vecs}\boldsymbol{\Phi}} \;=\; \tfrac{1}{2}\,\boldsymbol{G}'\mathrm{vec}\left(\ \boldsymbol{\Phi}^{-1}(\boldsymbol{G}_{\mathbf{x}_i} - \boldsymbol{\Phi})\boldsymbol{\Phi}^{-1}\right)\ , \qquad (3.9.11)$$

where $\boldsymbol{G}$ is a $r^2 \mathrm{x} r(r+1)/2$ matrix with the property that vec$\boldsymbol{A}$=$\boldsymbol{G}$vecs$\boldsymbol{A}$ (see for example McCullogh, 1982), where $\boldsymbol{A}$ is a symmetric matrix.

Hence

$$\frac{\partial \ln \mathrm{L}}{\partial\, \mathrm{vecs}\boldsymbol{\Phi}} \;=\; \tfrac{1}{2}\,\boldsymbol{G}'\sum_{i=1}^{\mathrm{N}} \mathrm{vec}\ \boldsymbol{\Phi}^{-1}[-\boldsymbol{\Phi} + \mathrm{E}(\boldsymbol{G}_{\mathbf{x}_i}|\mathbf{y}_i)]\boldsymbol{\Phi}^{-1}\ ,$$

where

$$\begin{aligned}\mathrm{E}(\boldsymbol{G}_{\mathbf{x}_i}|\mathbf{y}_i) &= \mathrm{Cov}(\mathbf{x}|\mathbf{y}_i) + \mathrm{E}(\mathbf{x}_i|\mathbf{y}_i - \boldsymbol{\Theta})(\mathrm{E}(\mathbf{x}_i|\mathbf{y}_i - \boldsymbol{\Theta}))'\\ &= \mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i) + \boldsymbol{w}_i\boldsymbol{w}_i'\ ,\end{aligned}$$

and

$$\boldsymbol{w}_i = \mathrm{E}(\mathbf{x}_i|\mathbf{y}_i) - \boldsymbol{\Theta}. \tag{3.9.12}$$

Therefore

$$\frac{\partial \, ln\, \mathrm{L}}{\partial \,\mathrm{vecs}\boldsymbol{\Phi}} = \tfrac{1}{2}\, \boldsymbol{G}' \sum_{i=1}^{\mathrm{N}} \mathrm{vec}\Big(\boldsymbol{\Phi}^{-1}[-\boldsymbol{\Phi} + \mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i) + \boldsymbol{w}_i\boldsymbol{w}_i{}']\boldsymbol{\Phi}^{-1}\Big)\ . \tag{3.9.13}$$

Using the well known result

$$\mathfrak{I}(\gamma_i,\gamma_j) = -\mathrm{E}\Big(\frac{\partial^2 ln\,\mathrm{L}}{\partial\gamma_i\partial\gamma_j}\Big) = \mathrm{E}\Big(\frac{\partial\, ln\,\mathrm{L}}{\partial\gamma_i}\cdot\frac{\partial\, ln\,\mathrm{L}}{\partial\gamma_j}\Big) \tag{3.9.14}$$

it follows that

$$\begin{aligned}\mathfrak{I}(\boldsymbol{\Phi}) &= -\mathrm{E}\Big(\frac{\partial\, ln\,\mathrm{L}}{\partial\,\mathrm{vecs}\boldsymbol{\Phi}\,\partial\,\mathrm{vecs}'\boldsymbol{\Phi}}\Big)\\ &= \mathrm{E}\Big(\frac{\partial\, ln\,\mathrm{L}}{\partial\,\mathrm{vecs}\boldsymbol{\Phi}}\cdot\frac{\partial\, ln\,\mathrm{L}}{\partial\,\mathrm{vecs}'\boldsymbol{\Phi}}\Big)\\ &= \tfrac{1}{4}\boldsymbol{G}'\sum_{i=1}^{\mathrm{N}}\mathrm{E}\Big(\mathrm{vec}\,\boldsymbol{\Phi}^{-1}[-\boldsymbol{\Phi}+\mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i)+\boldsymbol{w}_i\boldsymbol{w}_i{}']\boldsymbol{\Phi}^{-1}\cdot\\ &\qquad \mathrm{vec}'\,\boldsymbol{\Phi}^{-1}[-\boldsymbol{\Phi}+\mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i)+\boldsymbol{w}_i\boldsymbol{w}_i{}']\boldsymbol{\Phi}^{-1}\Big)\boldsymbol{G}\ .\end{aligned} \tag{3.9.15}$$

Let

$$\boldsymbol{S}_i = \boldsymbol{w}_i\boldsymbol{w}_i{}' \tag{3.9.16}$$

and

$$\boldsymbol{\Omega}_i = \boldsymbol{\Phi} - \mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i) \tag{3.9.17}$$

then using (3.9.15)

$$\mathfrak{I}(\boldsymbol{\Phi}) = \tfrac{1}{4}\boldsymbol{G}'\sum_{i=1}^{\mathrm{N}}\mathrm{E}\Big(\mathrm{vec}\big(\boldsymbol{\Phi}^{-1}(\boldsymbol{S}_i-\boldsymbol{\Omega}_i)\boldsymbol{\Phi}^{-1}\big)\cdot\big(\mathrm{vec}'\,\boldsymbol{\Phi}^{-1}(\boldsymbol{S}_i-\boldsymbol{\Omega}_i)\boldsymbol{\Phi}^{-1}\big)\boldsymbol{G}\Big)\ . \tag{3.9.18}$$

Using the result that

$$\mathrm{vec}\boldsymbol{CAC'} = \left(\boldsymbol{C} \otimes \boldsymbol{C}\right)\mathrm{vec}\boldsymbol{A} \ ,$$

it follows that

$$\mathfrak{J}(\boldsymbol{\Phi}) = \tfrac{1}{4}\boldsymbol{G}'\left(\sum_{i=1}^{N} \boldsymbol{\Phi}^{-1} \otimes \boldsymbol{\Phi}^{-1}\mathrm{E}\left(\mathrm{vec}\,(\boldsymbol{S}_i - \boldsymbol{\Omega}_i)\mathrm{vec}'\,\boldsymbol{\Phi}^{-1}(\boldsymbol{S}_i - \boldsymbol{\Omega}_i)\right)\boldsymbol{\Phi}^{-1} \otimes \boldsymbol{\Phi}^{-1}\right)\boldsymbol{G} \quad (3.9.19)$$

where (see e.g. Magnus and Neudecker(1979))

$$\mathrm{E}\left(\mathrm{vec}\left(\boldsymbol{S}_i - \boldsymbol{\Omega}_i\right)\mathrm{vec}'\,(\boldsymbol{S}_i - \boldsymbol{\Omega}_i)\right) = (\boldsymbol{\Omega}_i \otimes \boldsymbol{\Omega}_i)[\boldsymbol{I}_r{}^2 + \boldsymbol{K}_{r,r}] \ ,$$

where $\boldsymbol{K}_{r,r}$ is the commutation matrix of Magnus and Neudeker (1979) with amongst others, the properties $\mathrm{vec}\boldsymbol{A}=\boldsymbol{K}_{r,r}\mathrm{vec}\boldsymbol{A}'$, $\boldsymbol{K}_{r,r} \cdot \boldsymbol{K}_{r,r}=\boldsymbol{I}_r{}^2$ and $\boldsymbol{K}_{r,r}\boldsymbol{G}=\boldsymbol{G}$.

Note that the maginal expectations (cf. (3.9.12))

$$\mathrm{E}(\boldsymbol{w}_i) = \mathrm{E}(\mathrm{E}(\mathbf{x}|\mathbf{y}) - \boldsymbol{\Theta}) = \boldsymbol{\Theta} - \boldsymbol{\Theta} = \mathbf{0} \quad (3.9.20)$$

and (see Bock 1990)

$$\mathrm{Cov}(\mathbf{x}|\mathbf{y}_i) + \mathrm{E}(\boldsymbol{w}_i\boldsymbol{w}_i') = \boldsymbol{\Phi} \ .$$

$$\mathrm{E}(\boldsymbol{w}_i\boldsymbol{w}_i') = \boldsymbol{\Phi} - \mathrm{Cov}(\mathbf{x}|\mathbf{y}). \quad (3.9.21)$$

Thus

$$\boldsymbol{w}_i \sim \mathrm{N}(\mathbf{0},\boldsymbol{\Omega}),\ \text{where}\ \boldsymbol{\Omega}_i = \boldsymbol{\Phi} - \mathrm{Cov}(\mathbf{x}|\mathbf{y}_i) \ , \quad (3.9.22)$$

and for $\boldsymbol{S}_i$ defined in (3.9.16),

$$\mathrm{E}[\mathrm{vecs}(\boldsymbol{S}_i - \boldsymbol{\Omega}_i)] = \mathbf{0} \ . \quad (3.9.23)$$

From (3.9.8) and (3.9.14) it follows that for $\boldsymbol{w}_i$ defined in (3.9.12) ,

$$\mathfrak{I}(\boldsymbol{\Theta})=\mathrm{E}\Big(\frac{\partial\, ln\, \mathrm{L}}{\partial\, \boldsymbol{\Theta}} \cdot \frac{\partial\, ln\, \mathrm{L}}{\partial\, \boldsymbol{\Theta}'}\Big)$$

$$= \mathrm{E}\ \Big(\sum_{i=1}^{\mathrm{N}} \boldsymbol{\Phi}^{-1}\mathrm{E}_{\mathbf{x}\,|\,\mathbf{y}_i}\big((\mathbf{x}-\boldsymbol{\Theta})(\mathbf{x}-\boldsymbol{\Theta})'\big)\boldsymbol{\Phi}^{-1}\Big)$$

$$= \mathrm{E}\ \Big(\sum_{i=1}^{\mathrm{N}} \boldsymbol{\Phi}^{-1}\boldsymbol{w}_i\, \boldsymbol{w}_i{}'\boldsymbol{\Phi}^{-1}\Big)$$

$$= \boldsymbol{\Phi}^{-1}\Big(\sum_{i=1}^{\mathrm{N}} [\boldsymbol{\Phi} - \mathrm{Cov}(\mathbf{x}|\mathbf{y}_i)]\Big)\ \boldsymbol{\Phi}^{-1}$$

$$= \boldsymbol{\Phi}^{-1}\ \Big(\mathrm{N}\boldsymbol{\Phi} - \sum_{i=1}^{\mathrm{N}} \mathrm{Cov}(\mathbf{x}|\mathbf{y}_i)\Big)\boldsymbol{\Phi}^{-1}. \tag{3.9.24}$$

From (3.9.14) using (3.9.8) and (3.9.13)

$$\mathfrak{I}(\boldsymbol{\Phi},\boldsymbol{\Theta})=\mathrm{E}\Big(\frac{\partial\, ln\, \mathrm{L}}{\partial\, \mathrm{vecs}\boldsymbol{\Phi}} \cdot \frac{\partial\, ln\, \mathrm{L}}{\partial\, \boldsymbol{\Theta}'}\Big)$$

$$= \tfrac{1}{2}\, \boldsymbol{G}'\sum_{i=1}^{\mathrm{N}} \mathrm{E}\Big(\mathrm{vecs}[\boldsymbol{\Phi}^{-1}\big(- \boldsymbol{\Phi} + \mathrm{Cov}(\mathbf{x}|\mathbf{y}_i) + \boldsymbol{w}_i\boldsymbol{w}_i{}'\big)\boldsymbol{\Phi}^{-1}] \cdot \boldsymbol{w}_i{}'\boldsymbol{\Phi}^{-1}\Big)\,,$$

$$= \tfrac{1}{2}\, \boldsymbol{G}'\sum_{i=1}^{\mathrm{N}} \mathrm{E}\Big(\mathrm{vecs}\big(\boldsymbol{\Phi}^{-1}(\boldsymbol{S}_i - \boldsymbol{\Omega}_i)\boldsymbol{\Phi}^{-1}\big) \cdot \boldsymbol{w}_i{}'\boldsymbol{\Phi}^{-1}\Big)$$

$$= \tfrac{1}{2}\, \boldsymbol{G}'\sum_{i=1}^{\mathrm{N}} \boldsymbol{\Phi}^{-1}\mathrm{E}\Big(\mathrm{vecs}(\boldsymbol{S}_i - \boldsymbol{\Omega}_i)\Big)\boldsymbol{\Phi}^{-1}\mathrm{E}(\boldsymbol{w}_i{}')\boldsymbol{\Phi}^{-1}\Big)\,,$$

$$= \quad \mathbf{0}\,,$$

$\boldsymbol{S}_i$ and $\boldsymbol{\Omega}_i$ is defined in (3.9.16) and (3.9.17) respectively.

$$\mathfrak{I}(\sigma^2,\boldsymbol{\Theta}) = \mathrm{E}\Big(\frac{\partial\, ln\, \mathrm{L}}{\partial\, \sigma^2} \cdot \frac{\partial\, ln\, \mathrm{L}}{\partial\, \boldsymbol{\Theta}'}\Big)$$

$$= \sum_{i=1}^{\mathrm{N}} \mathrm{E}\ \Big(\ \tfrac{1}{2}\ \mathrm{E}_{\mathbf{x}\,|\,\mathbf{y}_i}\big(tr\ \sigma^{-4}\ (\mathbf{y}_i - \boldsymbol{B}_i\mathbf{x})(\mathbf{y}_i - \boldsymbol{B}_i\mathbf{x})' - \hat{\sigma}^2\boldsymbol{I}_{\mathrm{n}_i}\big) \cdot \boldsymbol{w}_i{}'\boldsymbol{\Phi}^{-1}\Big)$$

$$= \mathbf{0}\,. \tag{3.9.26}$$

$$\mathfrak{I}(\sigma^2,\boldsymbol{\Phi}) = \mathrm{E}\Big(\frac{\partial\, ln\, \mathrm{L}}{\partial\, \sigma^2} \cdot \frac{\partial\, ln\, \mathrm{L}}{\partial\, \mathrm{vecs}\boldsymbol{\Phi}}\Big)$$

$$= \sum_{i=1}^{N} \mathrm{E}\left(\tfrac{1}{2}\, \mathrm{E}_{\mathbf{x}\,|\,\mathbf{y}_i}\!\left(tr\ \sigma^{-4}\ (\mathbf{y}_i - \boldsymbol{B}_i\mathbf{x})(\mathbf{y}_i - \boldsymbol{B}_i\mathbf{x})' - \sigma^2 \boldsymbol{I}_{\mathrm{n}_i}\right) \cdot \right.$$

$$\left. \tfrac{1}{2}\, \boldsymbol{G}'\ \mathrm{vecs}\left(\boldsymbol{\Phi}^{-1}\left(-\boldsymbol{\Phi} + \mathrm{Cov}(\mathbf{x}|\mathbf{y}_i) + \boldsymbol{w}_i\boldsymbol{w}_i'\right)\boldsymbol{\Phi}^{-1}\right)\right)$$

$$= \tfrac{1}{2}\boldsymbol{G}'\sigma^{-4} \sum_{i=1}^{N} \mathrm{E}\left(tr\,(\mathbf{y}_i - \boldsymbol{B}_i\mathbf{x})(\mathbf{y}_i - \boldsymbol{B}_i\mathbf{x})' - \sigma^2 \boldsymbol{I}_{\mathrm{n}_i}\right) \cdot$$

$$\mathrm{vecs}\left(\boldsymbol{\Phi}^{-1}\left(-\boldsymbol{\Phi} + \mathrm{Cov}(\mathbf{x}|\mathbf{y}_i) + \boldsymbol{w}_i\boldsymbol{w}_i'\right)\boldsymbol{\Phi}^{-1}\right)\Big)$$

$$= \tfrac{1}{2}\boldsymbol{G}'\sigma^{-4} \sum_{i=1}^{N} tr\left(\sigma^2 \boldsymbol{I}_{\mathrm{n}_i} - \boldsymbol{B}_i \mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i)\boldsymbol{B}_i'\right) \cdot$$

$$\mathrm{E}\left(\mathrm{vecs}\left(\boldsymbol{\Phi}^{-1}\left(-\boldsymbol{\Phi} + \mathrm{Cov}(\mathbf{x}|\mathbf{y}_i) + \boldsymbol{w}_i\boldsymbol{w}_i'\right)\boldsymbol{\Phi}^{-1}\right)\right)$$

$$= \tfrac{1}{2}\boldsymbol{G}'\sigma^{-4} \sum_{i=1}^{N} \mathrm{E}\left(\sigma^2\ \mathrm{vecs}\left(\boldsymbol{\Phi}^{-1}\left(-\boldsymbol{\Phi} + \mathrm{Cov}(\mathbf{x}|\mathbf{y}_i) + \boldsymbol{w}_i\boldsymbol{w}_i'\right)\boldsymbol{\Phi}^{-1}\right) - \right.$$

$$\left. - \boldsymbol{B}_i'\boldsymbol{B}_i \mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i) \cdot\ \mathrm{vecs}\left(\boldsymbol{\Phi}^{-1}\left(-\boldsymbol{\Phi} + \mathrm{Cov}(\mathbf{x}|\mathbf{y}_i) + \boldsymbol{w}_i\boldsymbol{w}_i'\right)\boldsymbol{\Phi}^{-1}\right)\right). \tag{3.9.27}$$

3.10 SUMMARY

In this chapter frequentist estimation procedures for *linear* as well as *nonlinear* parameter models were discussed. The procedures include maximum likelihood, marginal maximum likelihood and asymptotically distribution free procedures. To obtain these parameter estimates use often has to be made of an iterative procedure. Two algorithms are discussed in this chapter namely the Fisher scoring and EM algorithm. The Fisher scoring algorithm is often used when using the ML or ADF estimation procedures. In the MML estimation procedure use is often made of the EM algorithm to obtain initial estimates of the unknown parameters. Since Bayes estimators can be obtained as a by-product of the MML procedure, this procedure may be regarded as empirical Bayes.

A Gauss quadrature numerical integration technique which usually provides reliable approximations for integrals with no closed form solution is discussed and an example is given (cf. example 3.8.1) using a Richards model discussed in Chapter 2.

CHAPTER 4 BAYES ESTIMATION.

4.1 INTRODUCTION

In this chapter Maximum aposteriori (MAP), and empirical Bayes estimation procedures for *linear* and *nonlinear* random parameter models are considered.

The term Bayesian approach (see Melnyk (1974)) has its origin in the work on probability theory done during the mid-eighteenth century by the Reverend Thomas Bayes, an English clergyman. It was he who first suggested and proved the result known as Bayes' theorem, a cornerstone of this whole approach to the revision of probability through evidence. The essential difference between the approach to inference discussed in Chapter 3 and the alternate Bayesian approach is that in the latter approach there is one further ingredient in a general mathematical model, namely the Bayes theorem on conditional probability where observational data is combined with personalistic or subjective believes.

The Bayes theorem essential to the Bayesian approach, also known as the principle of inverse probability, states that

$$\mathrm{p}(\mathbf{x}|\mathbf{y}) = \frac{\mathrm{g}(\mathbf{x}) \cdot \mathrm{f}(\mathbf{y}|\mathbf{x})}{\mathrm{h}(\mathbf{y})}, \tag{4.1.1}$$

where $\mathrm{g}(\mathbf{x})$, $\mathrm{f}(\mathbf{y}|\mathbf{x})$ and $\mathrm{h}(\mathbf{y})$ respectively denote the probability density functions of $\mathbf{x}$, $\mathbf{y}|\mathbf{x}$ and $\mathbf{y}$. All of the information about $\mathbf{x}$, that is available in $\mathbf{y}$, is conveyed by the *posterior* probability density $\mathrm{p}(\mathbf{x}|\mathbf{y})$. From a strict Bayesian point of view only the posterior distribution and the posterior expected loss are relevant while frequentists measure the performance of a procedure by its risk function. The principles of Bayesian analysis will be briefly discussed in Section 4.2.

Suppose that $\mathbf{x}$ is a vector of unknown parameters to be estimated given the vector of observations $\mathbf{y}$. The Bayes estimator $\mathbf{x}$ is defined as $\mathrm{E}(\mathbf{x}|\mathbf{y})$, where

$$\mathrm{E}(\mathbf{x}|\mathbf{y}) = \int \mathbf{x}\, \mathrm{p}(\mathbf{x}|\mathbf{y})\, d\,\mathbf{x} \tag{4.1.2}$$

and p($\mathbf{x}|\mathbf{y}$) is the pdf of $\mathbf{x}|\mathbf{y}$. An attractive property of the Bayes estimator derived in Section 4.3 (see e.g. Bock 1989) is that for any amount or quality of data it will always give a plausible estimate of the parameters, provided that realistic prior information is available. In practice, the parameters of the prior distribution is usually unknown and have to be estimated using the available data. In such instances the Bayes estimators are called empirical Bayes estimators, see e.g. (Maritz and Lwin 1989).

The statistical properties of the Bayes estimator have been studied intensively (see, for example, Lindley, 1971; Lindley & Smith, 1972; Novick & Jackson, 1974; DeGroot, 1975). It is optimal among all estimators in having the minimum mean-square error when integrated over the distribution of $\mathbf{x}$. In a sense it is on average closer to the true value of $\mathbf{x}$ than any other estimator. However, for any particular $\mathbf{x}_i$ it is in general a biased estimator of the population distribution of $\mathbf{x}$ but $\hat{\mathbf{x}}|\mathbf{y}_i$ is nevertheless the best estimator of $\mathbf{x}$ for predicting future responses for a sampling unit or individual i.

It is well known (see e.g. Anderson 1986) that if ($\mathbf{y}$,$\mathbf{x}$) has a joint multivariate normal distribution, then the conditional distribution of $\mathbf{x}|\mathbf{y}$ is also multivariate normal. It will be shown in Section 4.3, that in the case of a random parameter linear model the Bayes estimator of the vector of unknown parameters, $\mathbf{x}$, is the mean of a conditional normal distribution, given in equation (3.3.17) of Section 3.3, Chapter 3.

In Section 4.4 the MAP estimator, which is the mode of the posterior distribution of $\mathbf{x}$ given $\mathbf{y}$, is discussed. An advantage of the MAP method of predicting, for instance a childs growth, is that it can utilize the information contained, not just in the current measurement of a child's height but in measures taken over previous months and years, as is typical in clinical applications. In addition this method predicts (or postdicts) growth, that is: it predicts height at any age not just height at maturity.

In Section 4.5 an algorithm used to obtain the MAP estimator is given. Applications of the theory discussed in this chapter are given in Section 4.6.

4.2 PRINCIPLES OF BAYESIAN ANALYSIS

Let $f(\mathbf{y},\mathbf{x})$ denote the joint probability density function (pdf) for a random observation vector $\mathbf{y}$ and a random parameter vector $\mathbf{x}$. According to the usual operations with pdf's,

$$f(\mathbf{y},\mathbf{x}) = f(\mathbf{y}\,|\,\mathbf{x})g(\mathbf{x}) \tag{4.2.1}$$

$$= p(\mathbf{x}\,|\,\mathbf{y})\,h(\mathbf{y}), \tag{4.2.2}$$

where $h(\mathbf{y})$ is the marginal distribution of $\mathbf{y}$ and $g(\mathbf{x})$ is the probability density function of $\mathbf{x}$.

From (4.2.1) and (4.2.2) it follows that

$$p(\mathbf{x}\,|\,\mathbf{y}) = \frac{g(\mathbf{x})f(\mathbf{y}\,|\,\mathbf{x})}{h(\mathbf{y})}, \tag{4.2.3}$$

where from standard results of conditional distributions (e.g. Morrison, 1991) it follows that

$$h(\mathbf{y}) = \int g(\mathbf{x})f(\mathbf{y}\,|\,\mathbf{x})\,d\mathbf{x}\,. \tag{4.2.4}$$

Equation (4.2.3) can be written as follows:

$$\begin{aligned} p(\mathbf{x}\,|\,\mathbf{y}) &\propto g(\mathbf{x})f(\mathbf{y}\,|\,\mathbf{x}) \\ &\propto \text{prior pdf times likelihood function}\,, \end{aligned} \tag{4.2.5}$$

where $\propto$ denotes proportionality, $p(\mathbf{x}\,|\,\mathbf{y})$ is the posterior pdf and is employed in the Bayesian approach to make inferences about parameters. The density function $g(\mathbf{x})$ is known as the prior pdf for the parameter vector $\mathbf{x}$. The likelihood function $f(\mathbf{y}\,|\,\mathbf{x})$ is a function of $\mathbf{x}$. Equation (4.2.5) is a mathematical result in the theory of probability and the basis of the Bayes theorem. It is important to note that $p(\mathbf{x}\,|\,\mathbf{y})$, the posterior pdf, contains the sample information (in the likelihood function $f(\mathbf{y}\,|\,\mathbf{x})$), as well as prior information (in the prior pdf $g(\mathbf{x})$).

4.3 THE BAYES ESTIMATOR

The assumption of an unequal number of measurements (n_i) for each experimental (i) unit is incorporated in the random parameter regressions model by allowing for different design matrices for different individuals. The repeated measurements nonlinear random parameter model is defined as follows:

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i, \boldsymbol{t}_i) + \boldsymbol{\epsilon}_i, \qquad i=1,2,\cdots,N \qquad (4.3.1)$$

where $\mathbf{y}_i$, $\mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i)$, $\boldsymbol{t}_i$ and $\boldsymbol{\epsilon}_i$ are n_ix1 component vectors. Assume that $\mathbf{x}_1,\cdots,\mathbf{x}_N$ are a random sample from an (rx1) vector $\mathbf{x}$ of stochastic parameters where $\mathbf{x}$ has a $N(\boldsymbol{\theta},\boldsymbol{\Phi})$ distribution. It is further assumed that $\boldsymbol{\epsilon}_i \sim N(\mathbf{0},\boldsymbol{\Lambda}_i)$ and $Cov(\mathbf{x}_i,\boldsymbol{\epsilon}_i')=\mathbf{0}$.

The Bayes estimator of $\mathbf{x}_i|\mathbf{y}_i$ (cf. (4.2.3) and (4.2.4)) is

$$E(\mathbf{x}_i\,|\,\mathbf{y}_i) = \frac{\int \mathbf{x}_i\, f(\mathbf{y}_i\,|\,\mathbf{x}_i)\; g(\mathbf{x}_i)\, d\,\mathbf{x}_i}{\int f(\mathbf{y}_i\,|\,\mathbf{x}_i)\; g(\mathbf{x}_i)\, d\,\mathbf{x}_i}\ . \qquad (4.3.2)$$

The density function of $\mathbf{x}_i$, known as the prior distribution of $\mathbf{x}_i$ (cf (3.3.1)) is denoted as $g(\mathbf{x}_i)$, with

$$g(\mathbf{x}_i) = (2\pi)^{-r/2}\,|\boldsymbol{\Phi}|^{-1/2}\; exp -\tfrac{1}{2}\,P \qquad (4.3.3)$$

with

$$P = tr\,\boldsymbol{\Phi}^{-1}(\mathbf{x}_i-\boldsymbol{\theta})(\mathbf{x}_i-\boldsymbol{\theta})'. \qquad (4.3.4)$$

The pdf for the elements of $\mathbf{y}_i$ given the parameter vector $\mathbf{x}_i$ (cf. (3.3.3)) is:

$$f(\mathbf{y}_i\,|\,\mathbf{x}_i) = (2\pi)^{-n_i/2}\,|\boldsymbol{\Lambda}_i|^{-1/2} exp -\tfrac{1}{2}Q_i \qquad (4.3.5)$$

with

$$Q_i = tr\,\boldsymbol{\Lambda}_i^{-1}\Big(\mathbf{y}_i-\mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i)\Big)\Big(\mathbf{y}_i-\mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i)\Big)'. \qquad (4.3.6)$$

The product (cf. (4.3.3) and (4.3.5)) is given by

$$f(\mathbf{y}_i \mid \mathbf{x}_i) \cdot g(\mathbf{x}_i) = (2\pi)^{-\frac{n_i+r}{2}} \mid \mathbf{\Lambda}_i \mid^{-1/2} |\mathbf{\Phi}|^{-1/2} exp -\tfrac{1}{2}(Q_i+P) , \qquad (4.3.7)$$

with P and Q_i defined in (4.3.4) and (4.3.5) respectively.

Linear model

For $\mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i)$ *linear* in the parameters it is possible to write (4.3.1) as follows:

$$\mathbf{y}_i = \boldsymbol{B}_i\mathbf{x}_i + \epsilon_i, \qquad i=1,2,\cdots,N \qquad (4.3.8)$$

where $\boldsymbol{B}_i$ is a $n_i$xr design matrix. From the assumptions for model (4.3.1) it follows that

$$\Sigma_i = \boldsymbol{B}_i\mathbf{\Phi}\boldsymbol{B}_i' + \mathbf{\Lambda}_i , \qquad (4.3.9)$$

If $E(\mathbf{y}_i)$ is denoted by ξ_i, then

$$\xi_i = \boldsymbol{B}_i\boldsymbol{\theta} . \qquad (4.3.10)$$

From the above assumptions,

$$\mathbf{y}_i \sim N(\xi_i,\Sigma_i) .$$

It is possible to write equation (4.3.7) as follows (see Bock 1989) :

$$\int f(\mathbf{y}_i \mid \mathbf{x}_i) g(\mathbf{x}_i)\, d\,\mathbf{x}_i = (2\pi)^{-\frac{n_i}{2}} \mid \mathbf{\Lambda}_i \mid^{-\frac{1}{2}} \mid \boldsymbol{A}|^{-\frac{1}{2}} \mid \mathbf{\Phi} \mid^{-\frac{1}{2}} e^{-\frac{1}{2}K} \cdot e^{\frac{1}{2}(\boldsymbol{A}^{-1}\mathbf{b}')(\boldsymbol{A}^{-1}\mathbf{b}')'} . \qquad (4.3.11)$$

where

$$\boldsymbol{A} = \boldsymbol{B}_i'\Lambda_i^{-1}\boldsymbol{B}_i + \mathbf{\Phi}^{-1} , \qquad (4.3.12)$$

$$\mathbf{b}' = \bar{\mathbf{y}}'\, \Lambda_i^{-1}\boldsymbol{B}_i + \boldsymbol{\theta}'\mathbf{\Phi}^{-1} \qquad (4.3.13)$$

and

$$K \quad = tr\,\Lambda_i^{-1}\mathbf{y}_i\mathbf{y}_i{}' + tr\,\mathbf{\Phi}^{-1}(\boldsymbol{\theta}\boldsymbol{\theta}') \,, \qquad (4.3.14)$$

For the linear regression model it follows from (4.3.11) that the posterior distribution function, $p(\mathbf{x}_i|\mathbf{y}_i)$ is normally distributed with Bayes estimator

$$\hat{\mathbf{x}}_i\,|\,\mathbf{y}_i \quad = \boldsymbol{A}^{-1}\mathbf{b}' = \left(\boldsymbol{B}_i{}'\Lambda_i^{-1}\boldsymbol{B}_i + \mathbf{\Phi}^{-1}\right)^{-1}\left(\mathbf{y}_i{}'\Lambda_i{}^{-1}\boldsymbol{B}_i + \boldsymbol{\theta}'\mathbf{\Phi}^{-1}\right) \qquad (4.3.15)$$

and

$$\mathrm{Cov}(\hat{\mathbf{x}}_i\,|\,\mathbf{y}_i) \quad = \boldsymbol{A}^{-1} = \left(\boldsymbol{B}_i{}'\Lambda_i{}^{-1}\boldsymbol{B}_i + \mathbf{\Phi}^{-1}\right)^{-1}. \qquad (4.3.16)$$

As mentioned in the remark of Section 3.3 in Chapter 3 it is also possible to write $\hat{\mathbf{x}}_i|\mathbf{y}_i$ as follows:

$$\hat{\mathbf{x}}_i|\mathbf{y}_i \quad = [\boldsymbol{B}_i{}'\Lambda_i^{-1}\boldsymbol{B}_i + \mathbf{\Phi}^{-1}]^{-1}\boldsymbol{B}_i{}'\Lambda_i^{-1}[\mathbf{y}_i - \boldsymbol{B}_i\boldsymbol{\theta}_i] + \boldsymbol{\theta}. \qquad (4.3.17)$$

Remark

The random parameter regression model may be extended (see e.g. Bock 1990, and Du Toit 1992) to incorporate so-called second stage covariates as mentioned in Chapter 3 and shown in examples 3.8.4 and 3.8.5 respectively.

The extended model with second-stage covariates is defined as follows (cf. (4.3.8)):

$$\mathbf{y}_i \quad = \boldsymbol{B}_i\mathbf{x}_i + \epsilon_i, \qquad i=1,2,\cdots,N, \qquad (4.3.18)$$

with

$$\mathbf{x}_i \quad = \boldsymbol{\theta} + \boldsymbol{\Upsilon}\mathbf{z}_i + \mathbf{u}_i, \qquad i=1,2,\cdots,N, \qquad (4.3.19)$$

where $\boldsymbol{\theta}$ is an (rx1) vector (the intecept term x_0 being included), $\boldsymbol{\Upsilon}$ a (rxq) matrix of fixed but unknown coefficients, and, $\mathbf{u}_i$ a (rx1) random vector with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Phi}$, and $\mathrm{Cov}(\epsilon_i,\mathbf{u}_i{}')=\mathbf{0}$.

Use of (4.3.18) and (4.3.19) gives

$$\mathbf{y}_i = \boldsymbol{B}_i(\boldsymbol{\theta}+\boldsymbol{\Upsilon}\mathbf{z}_i+\mathbf{u}_i)+\epsilon_i, \qquad i=1,2,\cdots,N, \tag{4.3.20}$$

with

$$E(\mathbf{y}_i) = \boldsymbol{B}_i(\boldsymbol{\theta}+\boldsymbol{\Upsilon}\mathbf{z}_i), \qquad i=1,2,\cdots,N, \tag{4.3.21}$$

and assuming that $\boldsymbol{\Lambda}_i=\sigma^2 \boldsymbol{I}_{n_i}$

$$\mathrm{Cov}(\mathbf{y}_i) = \boldsymbol{B}_i\boldsymbol{\Phi}\boldsymbol{B}_i+\sigma^2\boldsymbol{I}_{n_i} \qquad i=1,2,\cdots,N\ . \tag{4.3.22}$$

On the assumption that $\mathbf{x}_i$ and ϵ_i, $i=1,2,\cdots,N$ are independent random samples from multivariate normal populations, Bock (1990) showed that

$$E(\mathbf{x}_i|\mathbf{y}_i) = \left(\boldsymbol{B}_i{}'\boldsymbol{\Lambda}_i^{-1}\boldsymbol{B}_i + \boldsymbol{\Phi}^{-1}\right)^{-1}\boldsymbol{B}_i{}'\boldsymbol{\Lambda}_i^{-1}\left(\mathbf{y}_i-\boldsymbol{B}_i(\boldsymbol{\theta}+\boldsymbol{\Upsilon}\mathbf{z}_i)\right) + \boldsymbol{\theta}+\boldsymbol{\Upsilon}_i\mathbf{z}_i \tag{4.3.23}$$

and

$$\mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i) = \left(\boldsymbol{B}_i{}'\boldsymbol{\Lambda}^{-1}\boldsymbol{B}_i + \boldsymbol{\Phi}^{-1}\right)^{-1}. \tag{4.3.24}$$

Consider the case $\boldsymbol{\Lambda}_i=\sigma^2\boldsymbol{I}_{n_i}$ and let $\overline{\boldsymbol{\theta}} = \frac{1}{N}\sum_{i=1}^{N} E(\mathbf{x}_i|\mathbf{y}_i)$, it follows that (cf. (3.3.13) to (3.3.16)).

$$\hat{\boldsymbol{\theta}} = \overline{\boldsymbol{\theta}} - \boldsymbol{\Upsilon}\overline{\mathbf{z}}\ , \tag{4.3.25}$$

$$\hat{\boldsymbol{\Upsilon}} = \left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i{}' - N\overline{\mathbf{z}}\,\overline{\mathbf{z}}\,'\right)^{-1}\left(\sum_{i=1}^{N}E(\mathbf{x}_i|\mathbf{y}_i)\mathbf{z}_i{}' - N\overline{\boldsymbol{\theta}}\ \overline{\mathbf{z}}\,'\right), \tag{4.3.26}$$

$$\hat{\boldsymbol{\Phi}} = \frac{1}{N}\Bigg(\sum_{i=1}^{N}E(\mathbf{x}_i|\mathbf{y}_i)E(\mathbf{x}_i|\mathbf{y}_i{}') - N\overline{\boldsymbol{\theta}}\,\overline{\boldsymbol{\theta}}\,' - \left(\sum_{i=1}^{N}E(\mathbf{x}_i|\mathbf{y}_i)\ \mathbf{z}_i{}' - N\overline{\boldsymbol{\theta}}\ \overline{\mathbf{z}}\,'\right)\hat{\boldsymbol{\Upsilon}}{}' - \hat{\boldsymbol{\Upsilon}}\left(\sum_{i=1}^{N}\mathbf{z}_iE(\mathbf{x}_i|\mathbf{y}_i) - N\overline{\mathbf{z}}\,\overline{\boldsymbol{\theta}}\right) + \sum_{i=1}^{N}\mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i) + \hat{\boldsymbol{\Upsilon}}\left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i{}' - N\overline{\mathbf{z}}\ \overline{\mathbf{z}}\,'\right)\hat{\boldsymbol{\Upsilon}}{}'\Bigg) \tag{4.3.27}$$

and

$$\hat{\sigma}^2 = \Big(\sum_{i=1}^{N} \mathrm{n}_i\Big)^{-1}\sum_{i=1}^{N}\Big((\mathbf{y}_i - \boldsymbol{B}_i\mathrm{E}(\mathbf{x}_i|\mathbf{y}_i))'(\mathbf{y}_i - \boldsymbol{B}_i\mathrm{E}(\mathbf{x}_i|\mathbf{y}_i)) + tr\,\boldsymbol{B}_i'\boldsymbol{B}_i\mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i)\Big)\,. \tag{4.3.28}$$

Nonlinear model.

For some nonlinear models a closed form solution for (4.3.2) cannot be obtained. The Gauss-quadrature numerical integration technique discussed in Section 3.7 of Chapter 3 can then be used.

Example 4.3.1

Suppose that the response pattern over time may be adequately described by a Gompertz function:

$$\mathrm{f}(\mathbf{x}_i, t_{ij}) = \mathrm{x}_1\; exp(-\,\mathrm{x}_2\, \mathrm{x}_3^{\,t_{ij}}) \qquad i=1,2,\cdots,\mathrm{N},\; j=1,2,\cdots,\mathrm{n}_i. \tag{4.3.29}$$

Assume further that $\boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0},\sigma^2\boldsymbol{I}_{\mathrm{n}_i})$, that is, $\boldsymbol{\Lambda}_i = \sigma^2\boldsymbol{I}_{\mathrm{n}_i}$.

Then (cf. (4.2.4)

$$\mathrm{h}(\mathbf{y}_i) = \mathrm{k}\int\!\!\int\!\!\int exp\Big(-\frac{1}{2\sigma^2}\sum_{j=1}^{\mathrm{n}_i}\big(\mathrm{y}_{ij} - \mathrm{f}(\mathbf{x}_i,\mathrm{t}_{ij})\big)^2\Big) exp\, -\tfrac{1}{2}(\mathbf{x}_i - \boldsymbol{\theta})'\boldsymbol{\Phi}^{-1}(\mathbf{x}_i - \boldsymbol{\theta})d\mathrm{x}_1 d\mathrm{x}_2 d\mathrm{x}_3, \tag{4.3.30}$$

where

$$\mathrm{k} \;=\; (2\pi\sigma^2)^{\mathrm{n}_i/2}\,(2\pi)^{-3/2}|\boldsymbol{\Phi}|^{-\frac{1}{2}}\,.$$

Consider the following transformation

$$\mathbf{u} \;= \frac{1}{\sqrt{2}}\,\boldsymbol{T}^{-1}(\mathbf{x}_i - \boldsymbol{\theta}), \tag{4.3.31}$$

where

$$\boldsymbol{T}\boldsymbol{T}' \;= \boldsymbol{\Phi}\,. \tag{4.3.32}$$

It follows that

$$\mathbf{x}_i = \sqrt{2}\,\boldsymbol{T}\mathbf{u} + \boldsymbol{\theta} = \mathbf{g}(\mathbf{u}) \ , \tag{4.3.33}$$

and the Jacobian of the transformation is given by $|\sqrt{2}\boldsymbol{T}|$.

From (4.3.30) to (4.3.33) it follows that

$$\mathrm{h}(\mathbf{y}_i) = \mathrm{k}^* \int\int\int exp\, -\mathrm{u}_1^2 - \mathrm{u}_2^2 - \mathrm{u}_3^2\, exp\, -\frac{1}{2\sigma^2}\sum_{j=1}^{\mathrm{n}_i}\left(\mathrm{y}_{ij} - \mathrm{g}_1(\mathbf{u})\, exp\, -\mathrm{g}_2(\mathbf{u})\mathrm{g}_3(\mathbf{u})^{t_{ij}}\right)^2 d\mathrm{u}_1 d\mathrm{u}_2 d\mathrm{u}_3, \tag{4.3.34}$$

where

$$\mathrm{k}^* = (2\pi\sigma^2)^{\mathrm{n}_i/2}\,(\pi)^{-3/2} \ .$$

There is no closed form solution for the integral in (4.3.34) and it may be approximated using the Gauss quadrature procedure (cf. Section 3.7).

$$\mathrm{h}(\mathbf{y}_i^*) \simeq \mathrm{k}^* \sum_{l=1}^{\mathrm{q}}\sum_{m=1}^{\mathrm{q}}\sum_{n=1}^{\mathrm{q}} \mathrm{w}(l)\mathrm{w}(m)\mathrm{w}(n)\cdot \mathrm{Q}(l,m,n) \ , \tag{4.3.35}$$

where

$$\mathrm{Q}(l,m,n) = exp\, -\frac{1}{2\sigma^2}\sum_{j=1}^{\mathrm{n}_i}\left(\mathrm{y}_{ij} - \mathrm{g}_{1l}\cdot exp\, -\mathrm{g}_{2m}\cdot \mathrm{g}_{3n}{}^{t_{ij}}\right)^2,$$

and where

$$\mathrm{g}_{1l} = \sqrt{2}t_{11}\mathrm{z}_l + \theta_1,$$

$$\mathrm{g}_{2m} = \sqrt{2}(t_{21}\mathrm{z}_l + t_{22}\mathrm{z}_m) + \theta_2$$

and

$$\mathrm{g}_{3n} = \sqrt{2}(t_{31}\mathrm{z}_l + t_{32}\mathrm{z}_m + t_{33}\mathrm{z}n) + \theta_3 \ . \tag{4.3.36}$$

From (4.3.35) it follows that, for this example, the Bayes estimator may be

From (4.3.35) it follows that, for this example, the Bayes estimator may be approximated by

$$E_{\mathbf{x}_i|\mathbf{y}_i}(x_\alpha) \simeq \frac{\sum_{l=1}^{q}\sum_{m=1}^{q}\sum_{n=1}^{q} w(l)w(m)w(n)\, x_\alpha\, Q(l,m,n)}{\sum_{l=1}^{q}\sum_{m=1}^{q}\sum_{n=1}^{q} w(l)w(m)w(n)\, Q(l,m,n)} \qquad \alpha = 1,2,\cdots r \tag{4.3.37}$$

The evaluation of (4.3.37) involves triple summations. Therefore the solutions to these integrals may be computionally intensive for q and N large. In Chapter 5 it is shown how the dimensionality of integrals of this form may be reduced when multi-component functions of the type

$$f(\mathbf{x}_i,\boldsymbol{t}_i) = \sum_{i=1}^{k} x_{1k}\, f_k(x_{2k},x_{3k}), \tag{4.3.38}$$

such as the multi-component Richards function (cf. Chapter 2), is considered.

4.4 THE MAXIMUM APOSTERIOR ESTIMATOR.

For the posterior function defined in (4.2.3) it follows that

$$ln\, p(\mathbf{x}_i|\mathbf{y}_i) = ln\, f(\mathbf{y}_i|\, \mathbf{x}_i) + ln\, g(\mathbf{x}_i) - ln\, h(\mathbf{y}_i)\ . \tag{4.4.1}$$

For $g(\mathbf{x}_i)$, $f(\mathbf{y}_i|\, \mathbf{x}_i)$ and $h(\mathbf{y}_i)$ defined in (4.3.3), (4.3.5) and (4.3.7) respectively it follows that

$$\begin{aligned} ln\, p(\mathbf{x}_i|\mathbf{y}_i) &= -\tfrac{n_i}{2}\, ln(2\pi) - \tfrac{m}{2}\, ln(2\pi) - \tfrac{1}{2} ln|\boldsymbol{\Lambda}_i| - \tfrac{1}{2}(\mathbf{y}_i - f(\mathbf{x}_i,\boldsymbol{t}_i))'\boldsymbol{\Lambda}_i^{-1}(\mathbf{y}_i - f(\mathbf{x}_i,\boldsymbol{t}_i)) \\ &\quad - \tfrac{1}{2} ln|\boldsymbol{\Phi}| - \tfrac{1}{2}(\mathbf{x}_i - \boldsymbol{\theta})'\boldsymbol{\Phi}^{-1}(\mathbf{x}_i - \boldsymbol{\theta}) - ln\, h(\mathbf{y}_i) \\ &= K - \tfrac{1}{2}(\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i))'\boldsymbol{\Lambda}_i^{-1}(\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i,\boldsymbol{t}_i)) - \tfrac{1}{2}(\mathbf{x}_i - \boldsymbol{\theta})'\boldsymbol{\Phi}^{-1}(\mathbf{x}_i - \boldsymbol{\theta})\ ; \end{aligned} \tag{4.4.2}$$

where

$$K = -\tfrac{n_i}{2}\, ln(2\pi) - \tfrac{m}{2}\, ln(2\pi) - \tfrac{1}{2} ln|\boldsymbol{\Lambda}_i| - \tfrac{1}{2} ln|\boldsymbol{\Phi}| - ln\, h(\mathbf{y}_i).$$

Assume $\text{Cov}(\boldsymbol{\epsilon}_i,\boldsymbol{\epsilon}_i') = \boldsymbol{\Lambda}_i = \sigma^2 \boldsymbol{I}_{\text{n}_i}$.

The MAP estimator $\hat{\mathbf{x}}_{\text{MAP}}$ is found by solving

$$\frac{\partial ln\, \text{p}(\mathbf{x}_i \,|\, \mathbf{y}_i)}{\partial \mathbf{x}} = \mathbf{0}. \tag{4.4.3}$$

given initial values for σ^2, $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$.

Linear model

For the linear model (cf. (4.3.8)) it follows that (see for example, Bock and Du Toit (1992))

$$\frac{\partial\, ln\, \text{p}(\mathbf{x}_i|\, \mathbf{y}_i)}{\partial \mathbf{x}_i} = -\tfrac{1}{2}\Big(-2\boldsymbol{B}_i{}'\boldsymbol{\Lambda}_i{}^{-1}(\mathbf{y}_i - \boldsymbol{B}_i\mathbf{x}_i)\Big) - \tfrac{1}{2}\Big(2\boldsymbol{\Phi}^{-1}(\mathbf{x}_i - \boldsymbol{\theta})\Big)\;. \tag{4.4.4}$$

Setting (4.4.4) equal to zero yields

$$\hat{\mathbf{x}}_{\text{MAP}} \;=\; \Big(\boldsymbol{B}_i{}'\boldsymbol{\Lambda}_i{}^{-1}\boldsymbol{B}_i + \boldsymbol{\Phi}^{-1}\Big)^{-1}\Big(\mathbf{y}_i{}'\boldsymbol{\Lambda}_i{}^{-1}\boldsymbol{B}_i + \boldsymbol{\theta}'\boldsymbol{\Phi}^{-1}\Big)\;. \tag{4.4.5}$$

Note that for the linear model the MAP estimator is identical to the Bayes estimator $\text{E}(\mathbf{x}_i|\mathbf{y}_i)$ in (4.3.15).

Nonlinear model

Maximization of (4.4.3) with respect to $\mathbf{x}_i$ is equivalent to the minimization of

$$\text{F}(\mathbf{x}_i|\mathbf{y}_i) = -ln\, \text{p}(\mathbf{x}_i|\mathbf{y}_i)\;,$$

with the constant K omitted $(\frac{\partial \text{K}}{\partial \mathbf{x}_i} = \mathbf{0})$ that is

$$\text{F}(\mathbf{x}_i|\mathbf{y}_i) \;=\; \frac{1}{2\sigma^2}\Big(\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i, \boldsymbol{t}_i)\Big)'\Big(\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i, \boldsymbol{t}_i)\Big) + \tfrac{1}{2}(\mathbf{x}_i - \boldsymbol{\theta})'\boldsymbol{\Phi}^{-1}(\mathbf{x}_i - \boldsymbol{\theta})\;. \tag{4.4.6}$$

Hence for $\dfrac{\partial \text{F}(\mathbf{x}_i \,|\, \mathbf{y}_i)}{\partial \mathbf{x}_i} = \mathbf{0}$, it follows that where the $n_i$xr matrix, $\boldsymbol{J}_i$, of first order derivatives has typical element

$$[\boldsymbol{J}_i]_{j,k} = \left\{ \frac{\partial f(\mathbf{x}_i, t_{ij})}{\partial \mathbf{x}_k} \right\}, \qquad j=1,2,\cdots,n_i \quad \text{and } k=1,2,\cdots,r. \tag{4.4.8}$$

4.5 THE MAXIMUM APOSTERIORI ALGORITHM

The maximum aposteriori algorithm is an iterative procedure, making use of the MAP estimator along with the information matrix, $\mathfrak{I}$, given by -E($\boldsymbol{H}$), where

$$\boldsymbol{H} = \left(\frac{\partial^2 ln\, p(\mathbf{x}_i \mid \mathbf{y}_i)}{\partial \mathbf{x}_i' \partial \mathbf{x}_i} \right), \tag{4.5.1}$$

to find the unknown parameter estimates $\boldsymbol{\theta}$, $\boldsymbol{\Phi}$, and σ^2.

Linear model

From (4.4.4) and (4.5.1) it follows that for the linear model

$$\mathfrak{I} = (\boldsymbol{B}_i' \Lambda_i^{-1} \boldsymbol{B}_i + \boldsymbol{\Phi}^{-1}) \tag{4.5.2}$$

In the linear case, the MAP estimators (4.4.5) and (4.5.2) is used iteratively to find the unknown parameter estimates $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\Phi}}$, and σ^2 given in (3.3.13), (3.3.14) and (3.3.16) respectively. Note that in the linear case $E(\mathbf{x}_i|\mathbf{y}_i) = \hat{\mathbf{x}}_{MAP}$ and $Cov(\mathbf{x}_i|\mathbf{y}_i) = \mathfrak{I}^{-1}$, thus for the linear model the maximum aposteriori algorithm is equivalent to the EM algorithm described in Section 3.6.

Nonlinear model

From (4.4.7) it follows that the Hessian matrix (cf. (4.5.1)) for the nonlinear model can be obtained as follows:

$$\boldsymbol{H} = \frac{\partial^2 ln\, p(\mathbf{x}_i \mid \mathbf{y}_i)}{\partial \mathbf{x}_i \partial \mathbf{x}_i'}$$

$$\begin{aligned} \boldsymbol{H} &= \frac{\partial^2 ln\, \mathrm{p}(\mathbf{x}_i \mid \mathbf{y}_i)}{\partial \mathbf{x}_i \partial \mathbf{x}_i{}'} \\ &= \frac{\partial}{\partial \mathbf{x}_i}\left((\mathbf{x}_i - \boldsymbol{\theta})'\boldsymbol{\Phi}^{-1} - \frac{1}{\sigma^2}\,(\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i, \boldsymbol{t}_i))'\boldsymbol{J}_i \right) \\ &= \boldsymbol{\Phi}^{-1} + (\boldsymbol{J}_i{}'\boldsymbol{J}_i)/\sigma^2 - \frac{1}{\sigma^2}\,\boldsymbol{H}^* \,, \end{aligned} \tag{4.5.3}$$

where $\boldsymbol{J}_i$ is given in (4.4.8) and

$$\boldsymbol{H}^* = \left(\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i, \boldsymbol{t}_i)\right)' \frac{\partial^2 \mathbf{f}(\mathbf{x}_i, \boldsymbol{t}_i)}{\partial \mathbf{x}_k \partial \mathbf{x}_l} \qquad k,l = 1,2,\cdots,\mathrm{r} \tag{4.5.4}$$

For the maximum aposteriori algorithm in case of the nonlinear model, $\mathrm{E}(\mathbf{x}_i|\mathbf{y}_i)$ and $\mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i)$ in (3.3.13), (3.3.14) and (3.3.15) is substituted iteratively by (4.4.7) and (4.5.3) respectively.

Substituting $\mathrm{E}(\mathbf{x}_i|\mathbf{y}_i)$ and $\mathrm{Cov}(\mathbf{x}_i|\mathbf{y}_i)$ with equations (4.4.7) and (4.5.3) respectively $\hat{\sigma}^2$ in (3.3.15) can be written as

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{\mathrm{N}} \mathbf{y}'\mathbf{y} + tr\boldsymbol{J}_i{}'\boldsymbol{J}_i[\boldsymbol{H}^{-1}]}{\sum_{i=1}^{\mathrm{N}} \mathrm{n}_i} \,. \tag{4.5.5}$$

The unknown parameter estimates $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\Phi}}$ and $\hat{\sigma}^2$ can also be found iteratively using the Fisher scoring algorithm described in Section 3.5.

4.6 PRACTICAL APPLICATIONS

Example 4.6.1

The data used in this section is reported in Childs, et al (1992), see Appendix A for the raw data. Reaction times were measured monthly on 8 occasions. The covariate, z, is the amount of damage caused to the neural paths in the brains of 9 rats. A negative value indicates damage to those neural paths which are believed to stimulate reaction. A positive value indicates damage to those neural paths believed to suppress reaction, this mechanism being linked to the chemical dopamine. If the neural paths stimulating reaction are damaged, $(z < 0)$, the suppressing paths are dominant and hence reaction suppressing reaction (z<0), is expected to decrease reaction times.

A second degree polynomial was fitted to the reaction times. The model includes the covariate, z, and is as follows:

$$\mathbf{y}_i \quad = \boldsymbol{B}\mathbf{x}_i, + \epsilon_i \qquad\qquad i=1,2,\cdots,N \,. \tag{4.6.1}$$

It is assumed that $\mathbf{x}_i \sim N(\boldsymbol{B}(\boldsymbol{\theta}+\boldsymbol{v}z_i),\boldsymbol{\Phi})$ independent of ϵ_i where $\epsilon_i \sim N(\mathbf{0},\sigma^2\boldsymbol{I})$, and where

$$\boldsymbol{B} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \\ 1 & 6 & 36 \\ 1 & 7 & 49 \\ 1 & 8 & 64 \end{bmatrix},$$

$\boldsymbol{\theta}=(\theta_0,\theta_1,\theta_2)'$, and $\boldsymbol{v}=(v_0,v_1,v_2)'$ (cf. (4.3.9)). The covariate values are respectively z_1= -0.543 , z_2= -0.450, z_3=-0.399, z_4=-0.153, z_5=1.013, z_6=1.019, z_7=1.331, z_8=1.477 and z_9=2.030.

Estimates of the unknown parameters (cf. (4.3.25) to (4.3.28)) are as follows:

$$\hat{\boldsymbol{\theta}} = (\, 396.2306,\ 16.4567,\ -0.8356)',$$

with standard errors 12.9805, 8.1507, and 0.9772 respectively.

The maximum likelihood estimates of $\boldsymbol{v}$ and $\boldsymbol{\Phi}$ are

$$\hat{\boldsymbol{v}} = (\, -1.0753\,,\ -19.3992,\ 2.2427)',$$

and

$$\hat{\mathbf{\Phi}} = \begin{bmatrix} \begin{matrix} 613.0541 \\ (714.04) \end{matrix} & & \\ \begin{matrix} -382.7771 \\ (417.32) \end{matrix} & \begin{matrix} 349.4373 \\ (281.11) \end{matrix} & \\ \begin{matrix} 40.7486 \\ (47.21) \end{matrix} & \begin{matrix} -40.9746 \\ (32.84) \end{matrix} & \begin{matrix} 5.4718 \\ (4.03) \end{matrix} \end{bmatrix},$$

with standard errors indicated in brackets below each estimate.

Finally the maximum likelihood estimate of σ^2 is $\hat{\sigma}^2$=445.3415.

Let

$$\mathrm{E}(\mathbf{y}) = \boldsymbol{B}\,(\boldsymbol{\theta} + \boldsymbol{v}\mathrm{z})$$

and

$$\hat{\mathbf{y}} = \boldsymbol{B}\,(\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{v}}\mathrm{z}_i)\ .$$

A typical element $\hat{\mathrm{y}}_t$ of $\hat{\mathbf{y}}$ is

$$\hat{\mathrm{y}}_t = (\theta_0 + v_0\mathrm{z}) + (\hat{\theta}_1 + \hat{v}_1\mathrm{z})t + (\hat{\theta}_2 + \hat{v}_2\mathrm{z})t^2\ , \qquad t{=}1,\cdots,8. \qquad (4.6.2)$$

Figure 4.6.1 is a graphical representation of $\hat{\mathrm{y}}_t$ for various values of t. Each graph corresponds to a fixed covariate value in the range z=−2.0, −1.5, −1.0, −0.5, 0, 0.5, 1.0, 1.5, 2.0.

It is apparent from Figure 4.6.1 that an increase in reaction times is associated with negative covariate (z) values, whereas a decline in reaction times is associated with positive values of z. In both instances it seems that the effect of the covariate diminishes after a period of eight months.

The Bayes estimates $\mathrm{E}(\mathbf{x}_i|\mathbf{y}_i)$, i=1,2,⋯,9, of the random parameter vector $\mathbf{x}$, are given below ($\mathbf{x}=(\mathrm{x}_0, \mathrm{x}_1, \mathrm{x}_2)'$):

Rat number.	$E(x_0 \mid y_i)$	$E(x_1 \mid y_i)$	$E(x_2 \mid y_i)$
1	398.458	19.774	-1.878
2	419.871	20.188	-2.014
3	369.691	44.511	-2.780
4	395.680	22.745	-1.518
5	422.863	-39.652	6.755
6	376.556	3.426	0.292
7	391.644	-3.079	1.560
8	391.401	-11.750	1.448
9	394.186	-11.354	2.557

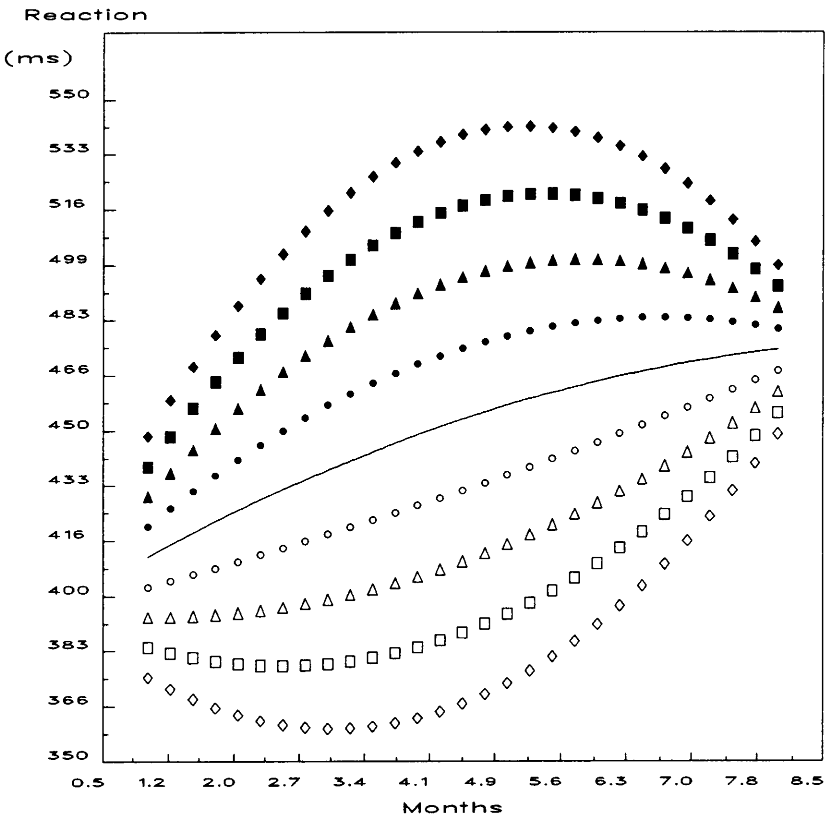

Figure 4.6.1 Reaction times of rats. Key to figure: z = 0 solid line

z =-2.0 : filled diamond z = 2.0 : diamond

z =-1.5 : filled square z = 1.5 : square

z =-1.0 : filled triangle z = 1.0 : triangle

z =-0.5 : filled circle z = 0.5 : circle

Figure 4.6.2 is a graphical representation of the curves fitted to the reaction times of the 9 rats. The solid lines denote curves based on Bayes parameter estimates whereas the dotted lines indicate curves based on OLS parameter estimates.

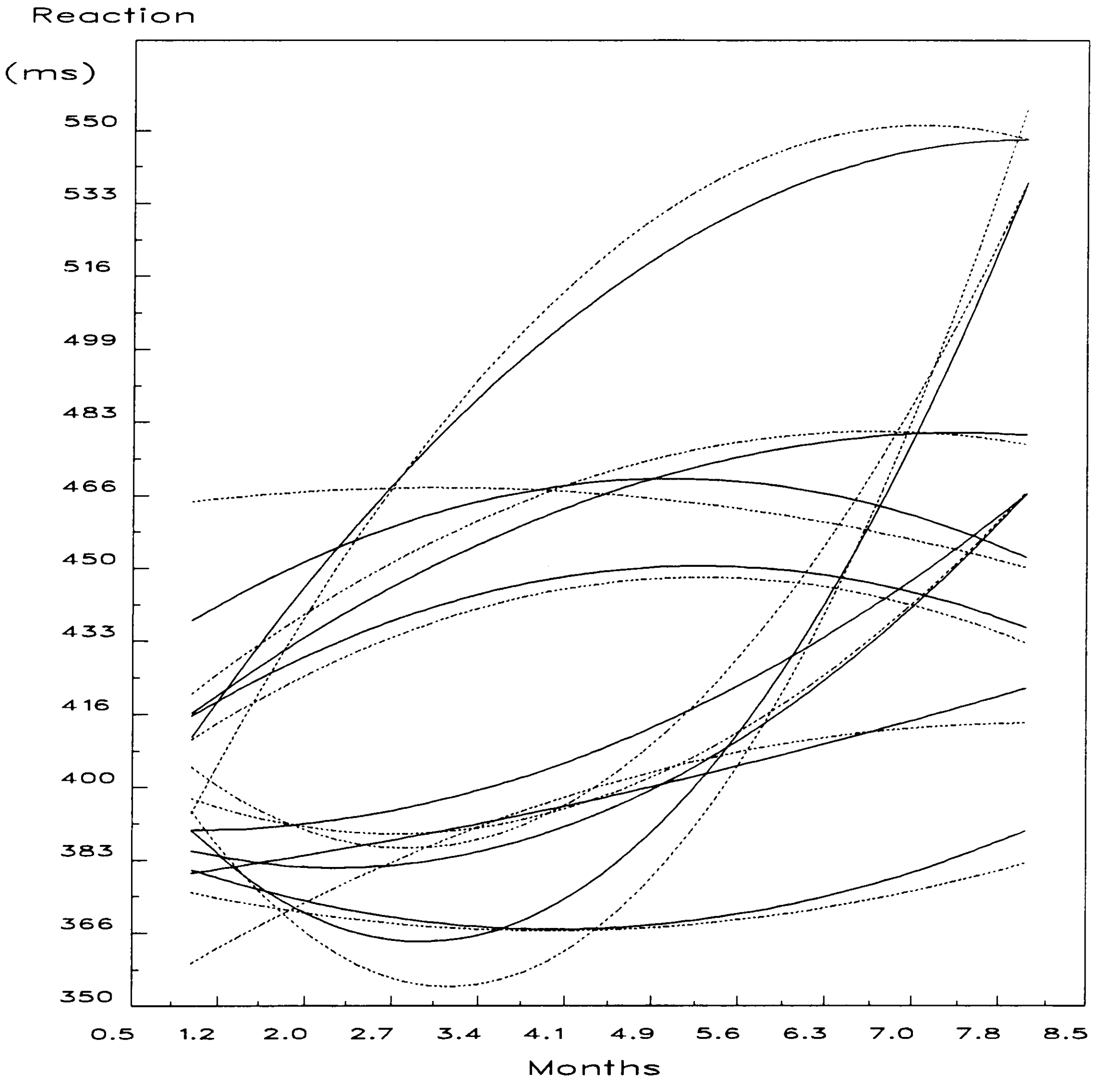

Figure 4.6.2 Curves fitted to the reaction times of 9 rats.

Example 4.6.2

The dataset used in this example was obtained from Dr. Greeff, Department of Agriculture, University of Pretoria (see Appendix A for part of the data). Weights of lambs were recorded weekly over a period of 14 weeks. The covariate, z, is the birth weight of the father ram. The birth weight, weight for each of the first 13 weeks, as well as weight at weaning in the 14^{th} week were recorded for each lamb with an unique identity number. The identity number of the father ram, including his age as well as the identity number of the mother ewe including her age, is known in each case.

A first degree polynomial was fitted to the weights. The model includes the covariate, z, and is as follows:

$$\mathbf{y}_i = \boldsymbol{B}_i\mathbf{x}_i + \boldsymbol{\epsilon}_i, \qquad i=1,2,\cdots,N, \tag{4.6.3}$$

with $\mathbf{x}_i \sim N(\boldsymbol{B}(\boldsymbol{\theta}+\boldsymbol{v}z_i),\boldsymbol{\Phi})$ independent of $\boldsymbol{\epsilon}_i$ and where $\boldsymbol{\epsilon}_i \sim N(\mathbf{0},\sigma^2\boldsymbol{I})$. For an ewe lamb with a father ram of 55 months, $\boldsymbol{B}_i$ is given by:

$$\boldsymbol{B}_i = \begin{bmatrix} 1 & 1 & 1 & 55 \\ 1 & 2 & 1 & 55 \\ 1 & 3 & 1 & 55 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & n_{i\text{-}1} & 1 & 55 \\ 1 & n_i & 1 & 55 \end{bmatrix},$$

In the model considered, $\boldsymbol{\theta}=(\theta_0,\theta_1,\theta_2,\theta_3)'$, and $\boldsymbol{v}=(v_0,v_1,v_2,v_3)'$, and the different values of the covariate, z, for the 21 rams are as follows:

$z_1 = 3.0$	$z_2 = 2.8$	$z_3 = 3.6$
$z_4 = 1.4$	$z_5 = 2.2$	$z_6 = 2.6$
$z_7 = 3.3$	$z_8 = 2.0$	$z_9 = 2.0$
$z_{10} = 2.5$	$z_{11} = 2.5$	$z_{12} = 2.5$
$z_{13} = 4.0$	$z_{14} = 3.0$	$z_{15} = 1.8$
$z_{16} = 3.0$	$z_{17} = 2.2$	$z_{18} = 3.0$
$z_{19} = 2.0$	$z_{20} = 2.6$	$z_{21} = 3.0$

Note that the first column of the design matrix $\boldsymbol{B}_i$ relates to the intercept term x_0, the elements of the second consists of the birth order (x_1) of the lambs, 1 denoting the first born, 2 the second born and so on. The third column of $\boldsymbol{B}_i$ is a dummy variable (x_3) for gender, 1 denoting male and -1 female, the fourth column denotes the age (x_4) of the ram at birth.

Estimates of the unknown parameters (cf. (4.3.25) to (4.3.28)) are as follows:

$$\hat{\boldsymbol{\theta}} = (\ 2.64810,\quad 1.30878,\quad -5.05868,\quad 0.00965)',$$

with standard errors: 0.67582, 0.02595, 1.21871, 0.01296;

$$\hat{\boldsymbol{v}} = (\ -0.21888,\quad -0.02082,\quad 4.35167,\quad -0.00048)',$$

$$\hat{\boldsymbol{\Phi}} = \begin{bmatrix} \underset{(0.00)}{8.117728} & & & \\ \underset{(0.00)}{0.131244} & \underset{(0.00)}{0.011964} & & \\ \underset{(0.00)}{-4.313944} & \underset{(0.00)}{0.091234} & \underset{(0.00)}{26.780573} & \\ \underset{(0.01)}{-0.151547} & \underset{(0.00)}{-0.003276} & \underset{(0.01)}{0.055602} & \underset{(0.00)}{0.002933} \end{bmatrix},$$

with standard errors indicated in brackets below each estimate.

Finally the maximum likelihood estimate of σ^2 is 8.692198.

Let

$$E(\mathbf{y}) = \boldsymbol{B}(\boldsymbol{\theta} + \boldsymbol{v}z)$$

and

$$\hat{\mathbf{y}} = \boldsymbol{B}(\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{v}}z_i) \ .$$

A typical element $\hat{y}_t$ of $\hat{\mathbf{y}}$ is

$$\hat{y}_t = (\theta_0 + v_0 z) + (\hat{\theta}_1 + \hat{v}_1 z)t + \hat{\theta}_2(\text{sex}) + \hat{\theta}_3(\text{age of ram}) \qquad t=1,\cdots,n_i. \quad i=1,2,\cdots,21 \qquad (4.6.4)$$

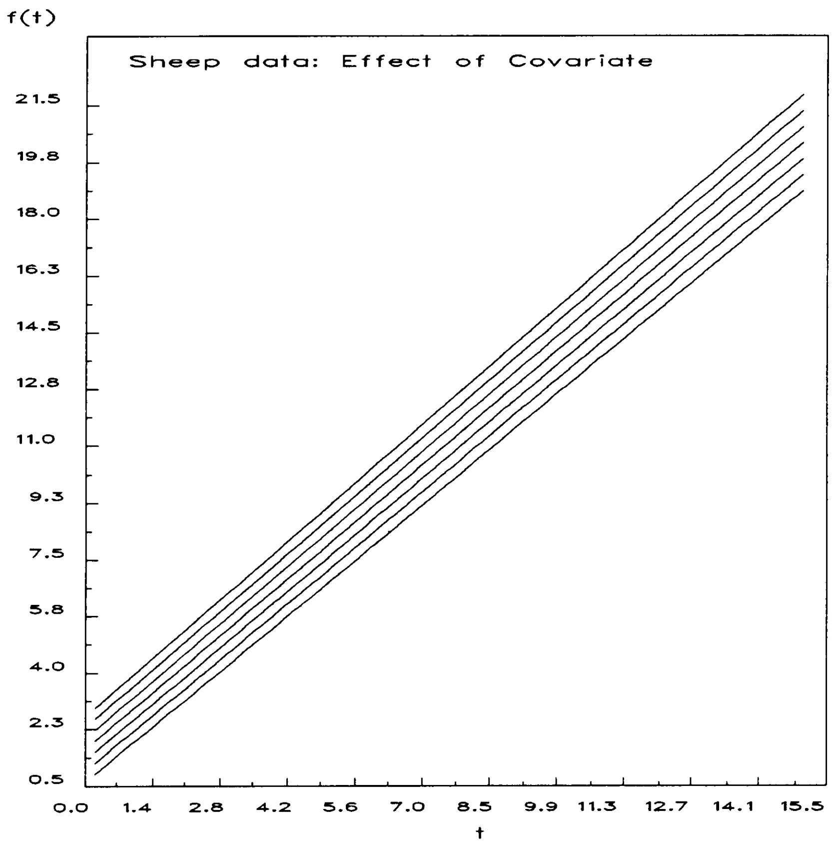

Figure 4.6.3 Effect of covariate. The upper line denotes the line with covariate 1.5 and the bottom line is associated with covariate value 4.5.

Figure 4.6.3 is a graphical representation to indicate the effect of the covariate on the weight of lambs, given that the age of the ram is fixed at 55 months and that the covariate values are 1.5, 2.0, 2.5, 3.0, 3.5, 4.0 and 4.5 respectively. It is apparent from this representation that the birth weight of the father ram has an influence on the weights of the lambs.

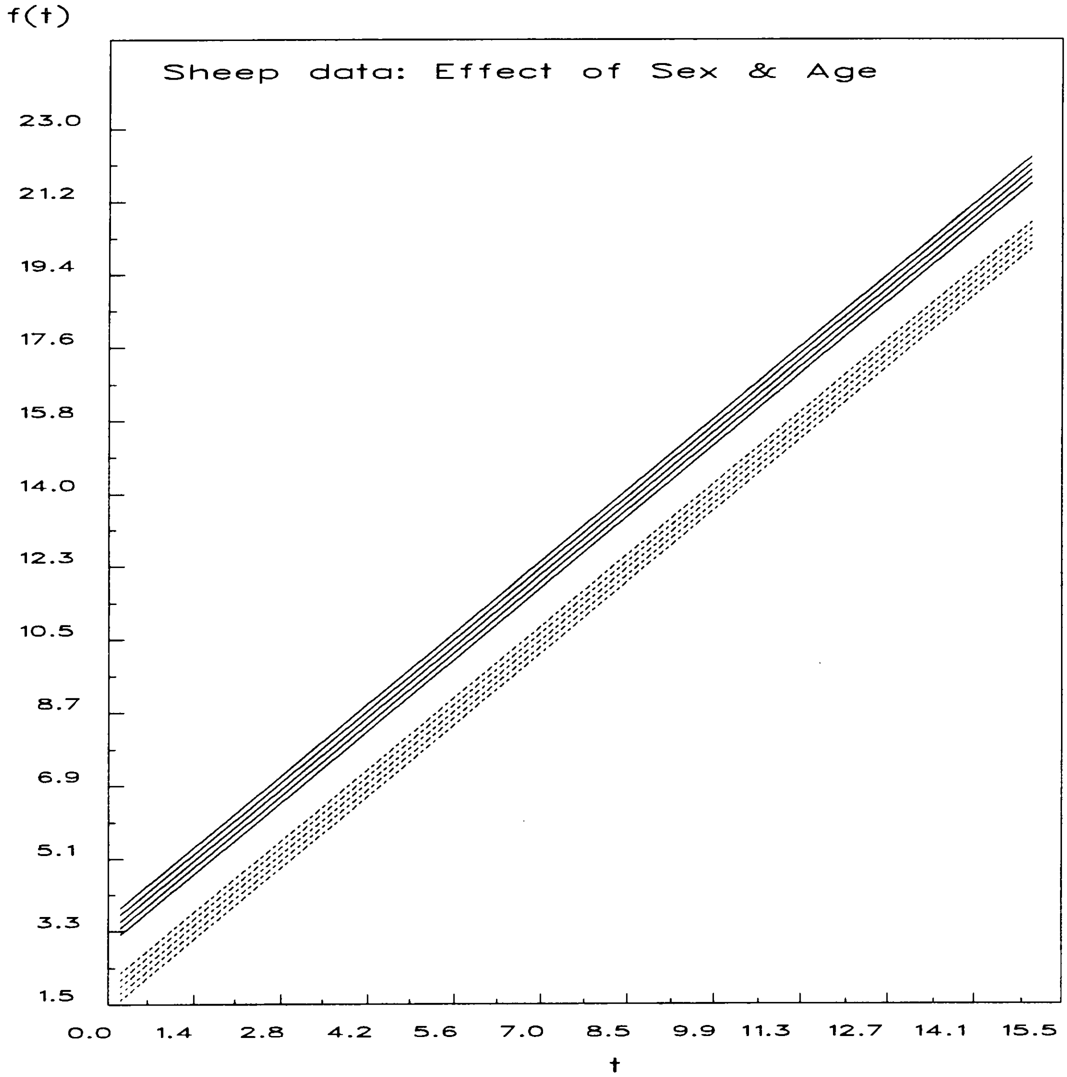

Figure 4.6.4 Effect of age and sex on lambs.

In Figure 4.6.4 the effect of age and sex is illustrated with a fixed covariate value, z=3.0. The solid line indicates the male lambs while the dotted line indicates the female lambs. The different ages that were used are 20, 40, 60, 80 and 100 months, where the upper solid line as well as dotted lines are associated with 100 months for males and females respectively and the bottom solid and dotted lines are associated with a father ram age of 20 months. It seems that sex has a greater influence on the weights of the lambs than the age of the father ram.

The Bayes estimates $E(\mathbf{x}_i|\mathbf{y}_i)$, i =1,2,$\cdots$,21 are given below:

Ram	$E(x_0\|\mathbf{y}_i)$	$E(x_1\|\mathbf{y}_i)$	$E(x_2\|\mathbf{y}_i)$	$E(x_3\|\mathbf{y}_i)$
1	2.00365	1.26975	5.14966	0.00014
2	3.11075	1.29008	7.83652	-0.01598
3	2.68172	1.27350	11.60170	-0.00990
4	2.80059	1.49301	6.40398	-0.01890
5	-1.04168	0.99817	-4.09239	0.08925
6	4.87727	1.24648	10.49291	-0.04016
7	0.22920	1.23377	7.20904	0.04751
8	1.18164	1.19044	2.41839	0.03302
9	-0.78231	1.28181	10.04484	0.06218
10	1.17579	1.21410	8.44762	0.02182
11	2.56605	1.28483	9.58174	-0.00396
12	1.35833	1.39151	3.34767	0.00965
13	2.98541	1.29129	10.37900	-0.01588
14	2.31415	1.33187	11.53216	-0.00388
15	11.44295	1.37180	-8.83541	-0.15265
16	-1.23074	1.10396	10.49478	0.07238
17	-0.07368	1.23312	12.34438	0.04635
18	3.60221	1.17296	12.24061	-0.01938
19	-0.66151	1.14515	0.48296	0.07200
20	1.96677	1.20638	1.64170	0.01636
21	3.06588	1.31534	4.38799	-0.01384

Example 4.6.3

Initial estimates using the MAP estimation procedure described Sections 4.4 and 4.5 were computed for the female mice data set as well as the male data set using model (3.8.3) with response function (3.8.4) in Chapter 3.

When fitting the model to the female mice data, the MAP procedure converged in 41 iterations yielding the following estimates of the unknown parameters:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} 41.182 \\ 1.462 \\ -0.457 \end{bmatrix}, \quad \hat{\boldsymbol{\Phi}} = \begin{bmatrix} 58.688 & & \\ 1.441 & 0.098 & \\ -0.111 & 0.008 & 0.047 \end{bmatrix} \text{ and } \hat{\sigma}^2 = 3.158 \; .$$

For the male mice data, the MAP procedure converged in 23 iterations yielding the following estimates of the unknown parameters.

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} 47.331 \\ 1.683 \\ -0.301 \end{bmatrix}, \quad \hat{\boldsymbol{\Phi}} = \begin{bmatrix} 108.354 & & \\ 2.561 & 0.134 & \\ -1.592 & 0.002 & 0.123 \end{bmatrix} \text{ and } \hat{\sigma}^2 = 3.881 \; .$$

Comparing the results in this example with those of Example 3.8.2 it is seen that the estimates of the unknown parameters are nearly the same.

Example 4.6.4

Model (3.8.17) with sex as covariate for the male and female mice data and $\mathbf{x}_i$ given in (3.8.18) of Example 3.8.4 were used to obtain initial estimates using the MAP estimation procedure. The MAP procedure converged in 53 iterations yielding the following estimates of the unknown parameters:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} 44.424 \\ 1.545 \\ -0.405 \end{bmatrix}, \quad \hat{\boldsymbol{v}} = \begin{bmatrix} 2.823 \\ 0.083 \\ 0.058 \end{bmatrix}, \quad \hat{\boldsymbol{\Phi}} = \begin{bmatrix} 83.579 & & \\ 2.165 & 0.119 & \\ -0.771 & 0.004 & 0.084 \end{bmatrix} \text{ and } \hat{\sigma}^2 = 3.974 \; .$$

Example 4.6.5

Initial estimates were obtained using the MAP procedure for the data described in Example 3.8.5 with the Gompertz function given in equation (3.8.23) and the measure of cognitive ability ($\mathbf{z}_i$) as covariate for the U.S. Air Force enlisted personnel (cf. (3.8.21) and (3.8.22)).

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} 34.609 \\ 1.0811 \\ 0.772 \end{bmatrix}, \quad \hat{\boldsymbol{v}} = \begin{bmatrix} 0.0267 \\ -0.0015 \\ -0.0001 \end{bmatrix}, \quad \hat{\boldsymbol{\Phi}} = \begin{bmatrix} 91.1471 & & \\ 1.1476 & 0.1609 & \\ -0.3423 & -0.0116 & 0.0318 \end{bmatrix} \text{ and } \hat{\sigma}^2 = 11.2872 \; .$$

Comparing the results of the MAP estimation procedure in Examples 4.6.3 and 4.6.4 with the results of the EM-algorithm in Examples 3.8.4 and 3.8.5 respectively it is seen that the estimates of the unknown parameters are nearly the same. The modified EM-algorithm as well as MAP estimation procedure seem to be both very reliable.

4.7 SUMMARY.

The posterior distribution of $\mathbf{x}_i$, given $\mathbf{y}_i$, contains all the information about the $\mathbf{x}_i$ afforded by the observation $\mathbf{y}_i$, and the information is conveyed by the posterior mean and covariance matrix. Correspondingly, all the sample information about the population parameters σ^2, $\boldsymbol{\mu}$ and Σ, is contained in the marginal distribution whose density is expressed in (4.3.7).

The procedures for structural (MML estimation) and nonstructural (MAP estimation) analysis, described in chapters 3 and 4, are based on proven statistical techniques and are excellent to describe and predict the future course of growth. The Bayesian origin of these procedures enables them to make the best use of the observations available, however fragmentary, when fitting individual growth curves.

The procedures discussed in Section 4.3 are implemented in the Fortran program NLIN, and a program MAPESTIM has been written, by S. H. C. du Toit, Department of Statistics, University of Pretoria, to implement the theory discussed in Section 4.4.

CHAPTER 5 ESTIMATION OF THE UNKNOWN PARAMETERS IN MULTI-COMPONENT RICHARDS MODELS.

5.1 INTRODUCTION

In this chapter models are considered to demonstrate the existence of growth phases or cycles in growth curves, using data with frequent measurements in time. Two different approaches for estimating the unknown parameters in nonlinear multi-component Richards models (cf. (2.1.1)) are considered.

The Fels as well as the Berkeley data sets which comprise of large male and female longitudinal growth records are used. R. D. Bock (1989) obtained access through Dr. Alex Roche to the growth records begun at the Fels Institute in 1929. The Fels data include a great number of cases measured from birth to maturity, making it possible to evaluate the accuracy of the growth prediction system by predicting growth at maturity from data recorded at any earlier stage of development. The following information required for the growth prediction system is included in the records: sex of the child, height of first degree relatives, especially parents, and estimates of the skeletal age of the child based on hand-wrist or knee radiographs. The Berkeley data (Tuddenham and Snyder 1954) is similar and some analyses are given in Appendix C.

We started off by fitting a nine parameter multi-component model to the Fels and Berkeley human growth records. We found however that both these data sets were over parameterized using the nine parameter triple-logistic model. This result was verified by an eigenvalue analysis on the correlation matrix of the estimated parameters. For both male and female datasets the smallest eigenvalue was essentially zero. This result is in agreement with the findings of Bock and Thissen (1980), who placed a linear restriction on the parameters thereby reducing the model fitting problem to the estimation of 8 free parameters.

Instead of resorting to the use of linear restrictions, an eight parameter function was employed in the following model:

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i,\boldsymbol{\alpha},\boldsymbol{t}_i) + \boldsymbol{\epsilon}_i, \qquad i = 1,2,\cdots,N. \tag{5.1.1}$$

where y_{ir}, $f(\mathbf{x}_i,\boldsymbol{\alpha},t_{ir})$ and ϵ_{ir}, $r=1,2,\cdots,n_i$, denote typical elements of the n_ix1 component vectors $\mathbf{y}_i$, $\mathbf{f}(\mathbf{x}_i,\boldsymbol{\alpha},\boldsymbol{t}_i)$ and $\boldsymbol{\epsilon}_i$ respectively, $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\alpha_3)'$ with $\boldsymbol{\alpha}$ a set of fixed parameters that may be set to predetermined values, and where $f(\mathbf{x}_i,\boldsymbol{\alpha},t_{ir})$ is defined as follows

$$f(\mathbf{x}_i,\boldsymbol{\alpha},t_r) = x_1[\,1+e^{-x_4 t_{ir}}]^{\alpha_1} + x_2[\,1+ e^{x_5 - x_6 t_{ir}}]^{\alpha_2} + x_3[\,1+ e^{x_7 - x_8 t_{ir}}]^{\alpha_3} + \epsilon_{ir}\,. \tag{5.1.2}$$

It is assumed that $\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_N$ are a random sample from a $N(\boldsymbol{\theta},\boldsymbol{\Phi})$ distribution. The error vectors, $\boldsymbol{\epsilon}_1,\boldsymbol{\epsilon}_2,\cdots,\boldsymbol{\epsilon}_N$ are assumed to be independently distributed as $N(\mathbf{0},\boldsymbol{\Lambda}_i)$ variates, independently distributed form the $\mathbf{x}_i$. It is also assumed that the covariance matrices, $\boldsymbol{\Lambda}_i$, can be expressed in terms of a common set of unknown parameters $\boldsymbol{\Lambda}_i=\boldsymbol{\Lambda}_i(\boldsymbol{\tau})$ (see Chapter 3 Section 3.2). The vector of unknown parameters, $\boldsymbol{\gamma}$, is defined as follows:

$$\boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\theta} \\ \text{vecs } \boldsymbol{\Phi} \\ \boldsymbol{\tau} \end{bmatrix}. \tag{5.1.3}$$

In the first approach, considered in Section 5.2, it is assumed, when estimating the unknown parameters, that n repeated measurements on each of N experimental units were obtained on the same occasions $t_1,t_2,\cdots,t_n$. Since the response functions to be considered in this chapter are nonlinear in the parameters, it follows from (5.1.1) and the distributional assumptions for $\mathbf{x}_i$ and $\boldsymbol{\epsilon}_i$, $i=1,2,\cdots,N$, that $\mathbf{y}_1,\cdots\mathbf{y}_N$ are not normally distributed. However, Anderson and Rubin (1956: 145-146) showed that in such cases consistent estimates of the unknown parameters may still be obtained using normal maximum likelihood, although the standard errors of the estimated parameters and test statistic values may then be incorrect.

Estimates of the unknown parameters, γ, (cf. (5.1.3)) are obtained using the approach described in Section 3.4.

Consider the case $n_1=n_2=\cdots=n_N=n$ and let

$$\xi = E(\mathbf{y}) = E[\mathbf{f}(\mathbf{x}_i,\boldsymbol{\alpha},\boldsymbol{t})] , \tag{5.1.4}$$

$$\Sigma = \mathrm{Cov}[\mathbf{y}_i,\mathbf{y}_i']=\mathrm{Cov}(\mathbf{f}_i,\mathbf{f}_i') + \boldsymbol{\Lambda} , \tag{5.1.5}$$

where $\mathbf{f}=\mathbf{f}(\mathbf{x},\boldsymbol{\alpha},\boldsymbol{t})$.

It is shown in Section 5.2 how to calculate ξ and Σ using Gauss quadrature numerical integration (see Section 3.7). The high dimensional integrals involved are respectively reduced to lower dimensional integrals when calculating the moments of the response function.

Du Toit and Browne (1991) developed a computer program, AUFITPC, to implement the theoretical procedures required to handle the automated fitting of nonstandard models under the assumption of multivariate normality. The user has to supply a subroutine for the evaluation of the mean vector, ξ, and covariance matrix, Σ.

The second approach, namely the MML (cf. Section 3.3) procedure of estimating unknown parameters, is given in Section 5.3. The derivatives of the log likelihood funcion are given in Section 5.4.

The data used in the application in Section 5.5 is the incomparable collection of data contained in the Fels growth study (see e.g. Bock 1973), which is a large sample of longitudinal male and female human growth records begun at the Fels Institute in 1929. In Examples 5.5.1 and 5.5.2 the theory discussed in Section 5.2 is applied to the Fels growth data. In Examples 5.5.3 and 5.5.4 the results of the MML procedure using the similar subsets of Examples 5.5.1 and 5.5.2 of the Fels data, are given.

5.2 FIRST AND SECOND ORDER MOMENTS OF MULTI-COMPONENT RICHARDS MODELS

At timepoint t_j, j=1,2,···,n, the triple-component response function defined in (5.1.2) can conveniently be expressed in the following form:

$$\mathrm{f}(\mathbf{x},\boldsymbol{\alpha},t_j) = [\mathrm{x}_1,\mathrm{x}_2,\mathrm{x}_3]\begin{bmatrix} \mathrm{f}_{1j} \\ \mathrm{f}_{2j} \\ \mathrm{f}_{3j} \end{bmatrix} \qquad (5.2.1)$$

where

$$\mathrm{f}_{1j} = (1 + \mathrm{e}^{-\mathrm{x}_4 t_j})^{\alpha_1} ,$$

$$\mathrm{f}_{2j} = (1 + \mathrm{e}^{\mathrm{x}_5 - \mathrm{x}_6 t_j})^{\alpha_2} ,$$

$$\mathrm{f}_{3j} = (1 + \mathrm{e}^{\mathrm{x}_7 - \mathrm{x}_8 t_j})^{\alpha_3} , \qquad j=1,2,\cdots,\mathrm{n}. \qquad (5.2.2)$$

From (5.2.1) and (5.2.2) it follows that

$$\mathrm{f}(\mathbf{x},\boldsymbol{\alpha},\boldsymbol{t}) = \boldsymbol{F}\boldsymbol{x}_1 \qquad (5.2.3)$$

where $\boldsymbol{F}$ is defined by

$$\boldsymbol{F} = \begin{bmatrix} \mathrm{f}_{11} & \mathrm{f}_{21} & \mathrm{f}_{31} \\ \mathrm{f}_{12} & \mathrm{f}_{22} & \mathrm{f}_{32} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \mathrm{f}_{1\mathrm{n}} & \mathrm{f}_{2\mathrm{n}} & \mathrm{f}_{3\mathrm{n}} \end{bmatrix},$$

and $\boldsymbol{x}_1$ by

$$\boldsymbol{x}_1 = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} . \qquad (5.2.4)$$

The mean is found by determining the expected value of the response function

$$E(y_r) = E[f(\mathbf{x},\boldsymbol{\alpha},t_r)] = E(x_1f_{1r} + x_2f_{2r} + x_3f_{3r}) \ , \qquad j=1,2,\cdots n. \qquad (5.2.5)$$

Proposition 5.2.1

The expected value $\xi_{kr} = E(x_kf_{kr})$, k=2,3 can be expressed as

$$\xi_{kr} \simeq \frac{1}{\pi}\sum_{i=1}^{N_t}\sum_{j=1}^{N_t} w_iw_j(\mu_3+t_{31}x_i + t_{32}x_j)h(x_i,x_j) \ ; \qquad (5.2.6)$$

with

$$\begin{aligned} h(x_i,x_j) &= \Big(1 + exp\big(t_{11}x_i + \mu_1 - (t_{21}x_i + t_{22}x_j + \mu_2\big)t_r\Big)^{\alpha_k} \\ &= \Big(1 + exp\big(\mu_1 - \mu_2t_r + (t_{11} - t_{21}t_r)x_i - t_{22}t_rx_j\big)\Big)^{\alpha_k} \ , \end{aligned} \qquad (5.2.7)$$

and where w_i and x_i in (5.2.6) denote the weights and nodes (See Stroud and Secrest 1966), respectively.

Proof

For k=2 and $\mathbf{x}_1=(x_2,x_5,x_6)'$, it follows that

$$\xi_{2r} = E(x_2f_{2r}) = \int\int\int x_2\big(1 + e^{x_5 - x_6t_r}\big)^{\alpha_2} \cdot \ g(\mathbf{x}_1)dx_2dx_5dx_6 \ , \ r=1,2,\cdots,n, \qquad (5.2.8)$$

where $g(\mathbf{x}_1)$ is the marginal distribution of $\mathbf{x}_1$, with probability density function

$$g(\mathbf{x}_1) = (2\pi)^{-3/2}|\boldsymbol{\Phi}_1|^{-\frac{1}{2}} \, exp\Big(-\tfrac{1}{2}(\mathbf{x}_1-\boldsymbol{\theta}_1)'\boldsymbol{\Phi}_1^{-1}(\mathbf{x}_1-\boldsymbol{\theta}_1)\Big) \ , \tag{5.2.9}$$

with

$$\boldsymbol{\theta}_1=(\theta_2, \theta_5, \theta_6)' \ ,$$

and

$$\boldsymbol{\Phi}_1=\begin{bmatrix} \phi_{22} & \phi_{25} & \phi_{26} \\ \phi_{25} & \phi_{55} & \phi_{56} \\ \phi_{26} & \phi_{56} & \phi_{66} \end{bmatrix}.$$

Transform $\mathbf{x}_1=(x_2,x_5,x_6)'$, to $\mathbf{z}_1=(z_1,z_2,z_3)'$ where $z_1=x_5$, $z_2= x_6$ and $z_3= x_2$, (5.2.10)
with

$$E(\mathbf{z}_1) = \boldsymbol{\mu}_1=\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}=\begin{bmatrix} \theta_5 \\ \theta_6 \\ \theta_2 \end{bmatrix} \tag{5.2.11}$$

and

$$Cov(\mathbf{z}_1,\mathbf{z}_1') = \boldsymbol{\Gamma}_1=\begin{bmatrix} \Gamma_{11} & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{21} & \Gamma_{22} & \Gamma_{23} \\ \Gamma_{31} & \Gamma_{32} & \Gamma_{33} \end{bmatrix}=\begin{bmatrix} \phi_{55} & \phi_{56} & \phi_{52} \\ \phi_{65} & \phi_{66} & \phi_{62} \\ \phi_{25} & \phi_{26} & \phi_{22} \end{bmatrix}. \tag{5.2.12}$$

For $\mathbf{z}_1=\mathbf{x}_1$ (cf. (5.2.10)), the Jacobian of the transformation is equal to 1 and hence (cf. (5.2.8))

$$E(x_2f_{2r}) =\int\int\int z_3\Big(1+e^{z_1-z_2t_r}\Big)^{\alpha_2}\cdot \ g(\mathbf{z}_1)dz_1dz_2dz_3 \ , \qquad r=1,2,\cdots,n; \tag{5.2.13}$$

where

$$g(\mathbf{z}_1) = (2\pi)^{-3/2}|\boldsymbol{\Gamma}_1|^{-1/2}exp\Big(-\tfrac{1}{2}(\mathbf{z}_1-\boldsymbol{\mu}_1)'\boldsymbol{\Gamma}_1^{-1}(\mathbf{z}_1-\boldsymbol{\mu}_1)\Big). \tag{5.2.14}$$

To evaluate (5.2.13), the following transformation is used. Let

$$\mathbf{u}:(3\text{x}1)= \frac{1}{\sqrt{2}}\boldsymbol{T}^{*-1}(\mathbf{z}_1-\boldsymbol{\mu}_1) \;, \tag{5.2.15}$$

where $\boldsymbol{T}^*$ is the Choleski square root of $\boldsymbol{\Gamma}_1$, that is,

$$\boldsymbol{T}^*\boldsymbol{T}^{*\prime} = \boldsymbol{\Gamma}_1 \quad \Rightarrow \quad \boldsymbol{\Gamma}_1^{-1} = (\boldsymbol{T}^{*\prime})^{-1}\boldsymbol{T}^{*-1}. \tag{5.2.16}$$

It follows that

$$\mathbf{u}'\mathbf{u} \;=\; \tfrac{1}{2}(\mathbf{z}_1-\boldsymbol{\mu}_1)'\boldsymbol{\Gamma}_1^{-1}(\mathbf{z}_1-\boldsymbol{\mu}_1), \tag{5.2.17}$$

and

$$\mathbf{z}_1= \sqrt{2}\boldsymbol{T}^*\mathbf{u}+\boldsymbol{\mu}_1 \;. \tag{5.2.18}$$

The Jacobian of the transformation $\left|\frac{\partial \mathbf{z}_1}{\partial \mathbf{u}'}\right|$, is equal to the determinant $|\sqrt{2}\boldsymbol{T}^*|$ where $\boldsymbol{T}^*$ is a 3x3 matrix, so that (cf. (5.2.16))

$$|\sqrt{2}\boldsymbol{T}^*| =2\,|\boldsymbol{\Gamma}_1|^{1/2} \;. \tag{5.2.19}$$

Let

$$\boldsymbol{T} \;=\; \sqrt{2}\boldsymbol{T}^* \;. \tag{5.2.20}$$

Using the above transformation with ξ_{2r}=E(x_2f_{2r}) it follows that (cf. (5.2.13))

$$\xi_{2r}=\text{c}\int\int\int \text{e}^{-\text{u}_1^2-\text{u}_2^2-\text{u}_3^2}\left(\text{t}_{31}\text{u}_1+\text{t}_{32}\text{u}_2+\text{t}_{33}\text{u}_3+\mu_3\right)\cdot \text{g}(\text{u}_1,\text{u}_2)\; d\text{u}_1 d\text{u}_2 d\text{u}_3 \tag{5.2.21}$$

where

$$\text{c}=\pi^{-3/2} \tag{5.2.22}$$

and

$$g(u_1,u_2) = \left(1 + e^{g_1(\mathbf{u}) - g_2(\mathbf{u})t_r}\right)^{\alpha_2}$$

$$= \left(1 + e^{(t_{11}u_1 + \mu_1) - (t_{21}u_1 + t_{22}u_2 + \mu_2)t_r}\right)^{\alpha_2} . \qquad (5.2.23)$$

□

Note that $g(u_1,u_2)$ defined in (5.2.23) can denote any other transformation of the parameters (see Chapter 2), for example

$$g(u_1,u_2) = 1 + e^{g_1(\mathbf{u})}\left(\frac{1}{1 + e^{g_2(\mathbf{u})}}\right). \qquad (5.2.24)$$

The three dimensional integral (5.2.21) can be reduced to a two dimensional integral as follows.

Let $\xi_{1r} = E(x_2 f_{2r}) = I_1 + I_2$,

where

$$I_1 = c \cdot \int e^{-u_3^2} du_3 \times \int\int e^{-u_1^2-u_2^2}\left(t_{31}u_1 + t_{32}u_2 + \mu_3\right) g(u_1,u_2)du_1du_2 \qquad (5.2.25)$$

and

$$I_2 = c \cdot \int e^{-u_3^2}\left(t_{33}u_3\right) du_3 \times \int\int e^{-u_1^2-u_2^2} g(u_1,u_2)du_1du_2 , \qquad (5.2.26)$$

with c defined in (5.2.22).

Note that u_3 is distributed $N(0,\frac{1}{2})$ and therefore $I_2 = 0$ and it follows that

$$\xi_{2r} = \frac{1}{\pi} \int\int e^{-u_1^2-u_2^2} \left(t_{31}u_1 + t_{32}u_2 + \mu_3\right) g(u_1,u_2)du_1du_2 . \qquad (5.2.27)$$

The integral in (5.2.27), has no closed form solution and can be evaluated to a high level of accuracy by making use of the Gauss quadrature numerical integration procedure, outlined in Section 3.7. The result in (5.2.6) follows. Similarily an expression for ξ_{3r} can be obtained.

Proposition 5.2.2

The expected value $\xi_{1r} = \mathrm{E}(\mathrm{x}_1 \mathrm{f}_{1r})$, can be expressed as

$$\xi_{1r} \simeq \frac{1}{\sqrt{\pi}} \sum_{i=1}^{\mathrm{N}_t} w_i (\mu_2 + t_{21} x_i) \mathrm{h}(x_i) \, , \tag{5.2.28}$$

where

$$\mathrm{h}(x_i) = \left(1 + exp - \left(\mu_1 + \mathrm{t}_{11} x_i\right) t_r \right)^{\alpha_1} . \tag{5.2.29}$$

Proof

The computation of ξ_{1r}=$\mathrm{E}(\mathrm{x}_1 \mathrm{f}_{1r})$ is similar to the computation of ξ_{kr}, for k=2,3 but involves the solution of a double and not a triple integral and ξ_{1r} reduces to the expression in (5.2.28). □

From Propositions 5.2.1 and 5.2.2 it is possible to evaluate

$$\xi_r = \xi_{1r} + \xi_{2r} + \xi_{3r} \, , \qquad r=1,2,\cdots,\mathrm{n}. \tag{5.2.30}$$

The covariance for the response function at timepoints t_r and t_s is as follows:

$$\mathrm{Cov}(\mathrm{f}(\mathbf{x},\alpha,t_r),\, \mathrm{f}(\mathbf{x},\alpha,t_s)) = \mathrm{E}\Big(\mathrm{f}(\mathbf{x},\alpha,t_r) \cdot \mathrm{f}(\mathbf{x},\alpha,t_s)\Big) - \mathrm{E}\Big(\mathrm{f}(\mathbf{x},\alpha,t_r)\Big)\mathrm{E}\Big(\mathrm{f}(\mathbf{x},\alpha,t_s)\Big) . \tag{5.2.31}$$

From (5.2.1) to (5.2.4) it follows that

$$\begin{aligned} \mathrm{E}\Big(\mathrm{f}(\mathbf{x},\alpha,t_r) \cdot \mathrm{f}(\mathbf{x},\alpha,t_s)\Big) &= \mathrm{E}\Big((\mathrm{x}_1\mathrm{f}_{1r} + \mathrm{x}_2\mathrm{f}_{2r} + \mathrm{x}_3\mathrm{f}_{3r}) \cdot (\mathrm{x}_1\mathrm{f}_{1s} + \mathrm{x}_2\mathrm{f}_{2s} + \mathrm{x}_3\mathrm{f}_{3s})\Big) \\ &= \mathrm{E}(\mathrm{x}_1^2 \cdot \mathrm{f}_{1r} \cdot \mathrm{f}_{1s}) + \mathrm{E}(\mathrm{x}_2^2 \cdot \mathrm{f}_{2r} \cdot \mathrm{f}_{2s}) + \mathrm{E}(\mathrm{x}_3^2 \cdot \mathrm{f}_{3r} \cdot \mathrm{f}_{3s}) \\ &\quad + \mathrm{E}(\mathrm{x}_1\mathrm{f}_{1r} \cdot \mathrm{x}_2\mathrm{f}_{2s}) + \mathrm{E}(\mathrm{x}_1\mathrm{f}_{1r} \cdot \mathrm{x}_3\mathrm{f}_{3s}) + \mathrm{E}(\mathrm{x}_2\mathrm{f}_{2r} \cdot \mathrm{x}_3\mathrm{f}_{3s}) \\ &\quad + \mathrm{E}(\mathrm{x}_1\mathrm{f}_{1s} \cdot \mathrm{x}_2\mathrm{f}_{2r}) + \mathrm{E}(\mathrm{x}_1\mathrm{f}_{1s} \cdot \mathrm{x}_3\mathrm{f}_{3r}) + \mathrm{E}(\mathrm{x}_2\mathrm{f}_{2s} \cdot \mathrm{x}_3\mathrm{f}_{3r}). \end{aligned} \tag{5.2.32}$$

Proposition 5.2.3

The expression for $E(x_k^2 \cdot f_{kr} \cdot f_{ks})$ for $k = 2$ or 3 reduces to:

$$E(x_k^2 \cdot f_{kr} \cdot f_{ks}) \approx 1/\pi \sum_i \sum_j w_i w_j h(x_i,x_j) \; . \tag{5.2.33}$$

$$h(x_i,x_j) = h_1(x_i,x_j)\Big(h_{2r}(x_i,x_j)\Big)^{\alpha_k}\Big(h_{3s}(x_i,x_j)\Big)^{\alpha_k} \; , \tag{5.2.34}$$

with

$$h_1(x_i,x_j) = t_{33}^2 + \mu_3^2 + t_{31}^2 x_i^2 + 2t_{31}t_{32}x_i x_j + t_{32}^2 x_j^2 + 2t_{31}x_i\mu_3 + 2t_{32}x_j\mu_3 \tag{5.2.35}$$

and

$$h_{2r}(x_i,x_j) = 1 + exp\Big(\mu_1 - \mu_2 t_r + (t_{11} - t_{21}t_r)x_i - t_{22}t_r x_j\Big) \; , \; r{=}s{=}1,2,\cdots,n \tag{5.2.36}$$

Proof

The computational procedure as well as the transformations used when computing $E(x_k^2 \cdot f_{kr} \cdot f_{ks})$ for $k = 2$ and 3 respectively is similar to the procedure discussed in Proposition 5.2.1. □

Proposition 5.2.4

The expression for $E(x_1^2 \cdot f_{1r} \cdot f_{1s})$ reduces to:

$$E(x_1^2 \cdot f_{1r} \cdot f_{1s}) \approx \frac{1}{\sqrt{\pi}} \sum_i w_i h_r(x_i) \; . \tag{5.2.37}$$

$$h_r(x_i) = h_1(x_i) \times \Big(h_{2r}(x_i)\Big)^{\alpha_1}\Big(h_{2s}(x_i)\Big)^{\alpha_1} \; , \tag{5.2.38}$$

with

$$h_1(x_i) = t_{22}^2 + \mu_2^2 + t_{21}^2 x_i^2 + 2t_{21}x_i\mu_2 \; , \tag{5.2.39}$$

$$h_{2r}(x_i) = 1 + exp\text{-}\left(t_{11}x_i + \mu_1\right)t_r \qquad r=1,2,\cdots,n. \tag{5.2.40}$$

Proof

The proof is similar to Proposition 5.2.2. □

Proposition 5.2.5

The expected value of the product,

$$E(x_2f_{2r}\cdot x_3f_{3s}) \approx \pi^{-2}\sum_i^q\sum_j^q\sum_k^q\sum_l^q w_iw_jw_kw_l h^*(x_i,x_j,x_k,x_l) \tag{5.2.41}$$

where w_i and x_i in (5.2.41) denote the weights and nodes (see Stroud and Secrest 1966) respectively.

$$h^*(x_i,x_j,x_k,x_l) = h_1^*(x_i,x_j,x_k,x_l)\left(h_{2r}^*(x_i,x_j)\right)^{\alpha_2}\left(h_{3s}^*(x_i,x_j,x_k,x_l)\right)^{\alpha_3}. \tag{5.2.42}$$

The functions h_1^*, h_{2r}^* and h_{3s}^* are defined as follows

$$h_1^*(x_i,x_j,x_k,x_l)=\Big(\left(s_{51}x_i+s_{52}x_j+s_{53}x_k+s_{54}x_l+\beta_5\right)\left(s_{61}x_i+s_{62}x_j+s_{63}x_k+s_{64}x_l+\beta_6\right) + s_{65}s_{55}\Big). \tag{5.2.43}$$

$$\begin{aligned} h_{2r}^*(x_i,x_j) &= 1 + exp\left(\beta_1 + s_{11}x_i - s_{21}x_it_r - s_{22}x_jt_r - \beta_2t_r\right) \\ &= 1 + exp\left(\beta_1 - \beta_2t_r + (s_{11} - s_{21}t_r)x_i - s_{22}t_rx_j\right). \end{aligned} \tag{5.2.44}$$

$$\begin{aligned} h_{3s}^*(x_i,x_j,x_k,x_l) = 1 + exp\Big(&\beta_3 - \beta_4t_s + (s_{31} - s_{41}t_s)x_i + (s_{32} - s_{42}t_s)x_j \\ &+ (s_{33} - s_{43}t_s)x_k - s_{44}t_sx_l\Big). \end{aligned} \tag{5.2.45}$$

Proof

$$E(x_2f_{2r} \cdot x_3f_{3s}) \;= \int x_2\Big(1+e^{x_5-x_6t_r}\Big)^{\alpha_2} \cdot x_3\Big(1+e^{x_7-x_8t_s}\Big)^{\alpha_3}\Big)\, g(\mathbf{x}^*)\; d\mathbf{x}^* \;, \qquad (5.2.46)$$

where $g(\mathbf{x}^*)$ is the marginal distribution of $\mathbf{x}^*=(x_2,\, x_3,\, x_5,\, x_6,\, x_7,\, x_8)'$ with pdf

$$g(\mathbf{x}^*) = (2\pi)^{-6/2} \,|\, \mathbf{\Phi}^* \,|^{-1/2} exp\Big(-\tfrac{1}{2}(\mathbf{x}^* - \boldsymbol{\mu}^*)'\mathbf{\Phi}^{*-1}(\mathbf{x}^* - \boldsymbol{\mu}^*)\Big)\,, \qquad (5.2.47)$$

$$\boldsymbol{\theta}^* = (\theta_2,\, \theta_3,\, \theta_5,\, \theta_6,\, \theta_7,\, \theta_8)'\;,$$

and

$$\mathbf{\Phi}^* = \begin{bmatrix} \phi_{22} & \phi_{23} & \cdots & \phi_{28} \\ \phi_{32} & \phi_{33} & \cdots & \phi_{38} \\ \vdots & \vdots & & \vdots \\ \phi_{82} & \phi_{83} & \cdots & \phi_{88} \end{bmatrix}. \qquad (5.2.48)$$

Consider the transformation

$$\mathbf{v} \;=\; \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} \;=\; \begin{bmatrix} x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_2 \\ x_3 \end{bmatrix} \qquad (5.2.49)$$

and let

$$E(\mathbf{v}) = \boldsymbol{\beta}, \qquad (5.2.50)$$

$$Cov(\mathbf{v},\mathbf{v}') = \mathbf{\Psi}, \qquad (5.2.51)$$

then

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{bmatrix} = \begin{bmatrix} \theta_5 \\ \theta_6 \\ \theta_7 \\ \theta_8 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

and

$$\mathbf{\Psi} = \begin{bmatrix} \phi_{55} & & & & & \\ \phi_{65} & \phi_{66} & & & & \\ \phi_{75} & \phi_{76} & \phi_{77} & & & \\ \phi_{85} & \phi_{86} & \phi_{87} & \phi_{88} & & \\ \phi_{25} & \phi_{26} & \phi_{27} & \phi_{28} & \phi_{22} & \\ \phi_{35} & \phi_{36} & \phi_{37} & \phi_{38} & \phi_{32} & \phi_{33} \end{bmatrix}. \qquad (5.2.52)$$

For the transformation in (5.2.49) the Jacobian of the transformation is equal to 1 and hence (cf. (5.2.46))

$$\mathrm{E}(\mathrm{x}_2\mathrm{f}_{2r}\cdot \mathrm{x}_3\mathrm{f}_{3s}) = \int \cdots \int \mathrm{v}_5\left(1+\mathrm{e}^{\mathrm{v}_1-\mathrm{v}_2\mathrm{t}_r}\right)^{\alpha_2}\cdot \mathrm{v}_6\left(1+\mathrm{e}^{\mathrm{v}_3-\mathrm{v}_4\mathrm{t}_s}\right)^{\alpha_3}\mathrm{g}(\mathbf{v})d\mathrm{v}_1 d\mathrm{v}_2 d\mathrm{v}_3 d\mathrm{v}_4 d\mathrm{v}_5 d\mathrm{v}_6\ , \quad r,\ s=1,2,\cdots,\mathrm{n}; \qquad (5.2.53)$$

with $\mathrm{g}(\mathbf{v})$ (cf. (5.2.47))

$$\mathrm{g}(\mathbf{v}) = (2\pi)^{-6/2}\,|\,\mathbf{\Psi}\,|^{-1/2} exp\left(-\tfrac{1}{2}(\mathbf{v}-\beta)'\mathbf{\Psi}^{-1}(\mathbf{v}-\beta)\right). \qquad (5.2.54)$$

To evaluate (5.2.53), the following transformation is used.

$$\mathbf{w}:(6\mathrm{x}1) = \frac{1}{\sqrt{2}}S^{*-1}(\mathbf{v}-\beta)\ , \qquad (5.2.55)$$

where S^* is the Choleski square root of $\mathbf{\Psi}$, that is,

$$S^*S^{*\prime} = \mathbf{\Psi} \quad \Rightarrow \quad \mathbf{\Psi}^{-1} = (S^{*\prime})^{-1}S^{*-1}. \qquad (5.2.56)$$

It follows that

$$\mathbf{w}'\mathbf{w} = \tfrac{1}{2}(\mathbf{v}-\boldsymbol{\beta})'\boldsymbol{\Psi}^{-1}(\mathbf{v}-\boldsymbol{\beta})\,, \tag{5.2.57}$$

and that

$$\mathbf{v} = \sqrt{2}S^{*}\mathbf{w}+\boldsymbol{\beta}\ . \tag{5.2.58}$$

The Jacobian of the transformation is

$$J = \left|\frac{\partial \mathbf{v}}{\partial \mathbf{w}'}\right| = (2)^{6/2}\,|\boldsymbol{\Psi}|^{1/2}\ . \tag{5.2.59}$$

It follows from (5.2.53) to (5.2.59) that

$$\begin{aligned} \mathrm{E}(\mathrm{x}_2\mathrm{f}_{2r}\cdot \mathrm{x}_3\mathrm{f}_{3s}) &= \int\cdots\int \mathrm{e}^{-\mathbf{w}'\mathbf{w}}\mathrm{g}_5(\mathbf{w})\Big(1+\mathrm{e}^{\mathrm{g}_1(\mathbf{w})\,-\,\mathrm{g}_2(\mathbf{w})t_r}\Big)^{\alpha_2} \\ &\qquad \cdot \mathrm{g}_6(\mathbf{w})\Big(1+\mathrm{e}^{\mathrm{g}_3(\mathbf{w})\,-\,\mathrm{g}_4(\mathbf{w})t_s}\Big)^{\alpha_3}d\mathbf{w}\ , \end{aligned} \tag{5.2.60}$$

where $\mathrm{g}_i(\mathbf{w}) = \mathrm{v}_i$ with $\mathbf{v}$ given in (5.2.58).

Let

$$\mathrm{E}(\mathrm{x}_2\mathrm{f}_{2r}\cdot \mathrm{x}_3\mathrm{f}_{3s}) = \mathrm{II}_1+\mathrm{II}_2\ ,$$

where

$$\begin{aligned} \mathrm{II}_1 &= \mathrm{c}\cdot\int\int \mathrm{e}^{-\mathrm{w}_5^2-\mathrm{w}_6^2}d\mathrm{w}_5 d\mathrm{w}_6\ \cdot \\ &\qquad \int\int\int\int \mathrm{e}^{-\mathrm{w}_1^2-\mathrm{w}_2^2-\mathrm{w}_3^2-\mathrm{w}_4^2}\cdot \mathrm{b}(\mathbf{w})\cdot \mathrm{h}(\mathrm{w}_1,\mathrm{w}_2,\mathrm{w}_3,\mathrm{w}_4)d\mathrm{w}_1 d\mathrm{w}_2 d\mathrm{w}_3 d\mathrm{w}_4 \end{aligned} \tag{5.2.61}$$

and

$$\begin{aligned} \mathrm{II}_2 &= \mathrm{c}\cdot\int\int \mathrm{e}^{-\mathrm{w}_5^2-\mathrm{w}_6^2}\ \mathrm{d}(\mathbf{w})\ d\mathrm{w}_5 d\mathrm{w}_6\ \cdot \\ &\qquad \int\int\int\int \mathrm{e}^{-\mathrm{w}_1^2-\mathrm{w}_2^2-\mathrm{w}_3^2-\mathrm{w}_4^2}\cdot \mathrm{h}(\mathrm{w}_1,\mathrm{w}_2,\mathrm{w}_3,\mathrm{w}_4)d\mathrm{w}_1 d\mathrm{w}_2 d\mathrm{w}_3 d\mathrm{w}_4 \end{aligned} \tag{5.2.62}$$

with

$c = \pi^{-3}$, and for $s_{ij} = [\sqrt{2}S^*]_{i,j}$

$$b(\mathbf{w}) = (s_{54}w_4 + s_{53}w_3 + s_{52}w_2 + s_{51}w_1 + \beta_5)(s_{64}w_4 + s_{63}w_3 + s_{62}w_2 + s_{61}w_1 + \beta_6), \quad (5.2.63)$$

$$d(\mathbf{w}) = (s_{55}w_5)(s_{64}w_4 + s_{63}w_3 + s_{62}w_2 + s_{61}w_1 + \beta_6) + (s_{66}w_6 + s_{65}w_5)(s_{55}w_5 + s_{54}w_4 + s_{53}w_3 + s_{52}w_2 + s_{51}w_1 + \beta_5) \quad (5.2.64)$$

and

$$h(w_1,w_2,w_3,w_4) = \Big(1 + exp\big(g_1(\mathbf{w}) - g_2(\mathbf{w})t_r\big)\Big)^{\alpha_2}\Big(1 + exp\big(g_3(\mathbf{w}) - g_4(\mathbf{w})t_s\big)\Big)^{\alpha_3} \quad (5.2.65)$$

For w_5 and w_6 each $N(0,\frac{1}{2})$ distributed and independent, it follows that:

$$II_1 = c \cdot \int\int\int\int e^{-w_1^2-w_2^2-w_3^2-w_4^2} \cdot \pi \cdot b(\mathbf{w}) \cdot h(w_1,w_2,w_3,w_4)d\mathbf{w} \quad (5.2.66)$$

and

$$II_2 = c \cdot \int\int\int\int e^{-w_1^2-w_2^2-w_3^2-w_4^2} \cdot (s_{65}s_{55}) \cdot \pi \cdot h(w_1,w_2,w_3,w_4)d\mathbf{w} \quad (5.2.67)$$

The expected value (5.2.60) may be evaluated numerically by making use of the Gauss quadrature integration technique discussed in Section 3.7 of Chapter 3. Hence using (5.2.66) and (5.2.67) the result follows. □

Proposition 5.2.6

The expression for $E(x_1f_{1r} \cdot x_kf_{ks})$ for k=2 or 3 is as follows:

$$E(x_1f_{1r} \cdot x_kf_{ks}) \approx \pi^{-3/2}\sum_i \sum_j \sum_k w_iw_jw_k k(x_i,x_j,x_k) \quad (5.2.68)$$

where w_i and x_i respectively denote the weights and nodes of a q-point Gauss quadrature numerical integration formula.

$$\mathrm{k}(x_i,x_j,x_k) = \mathrm{k}_1(x_i,x_j,x_k)\Big(\mathrm{k}_2(x_i)\Big)^{\alpha_1}\Big(\mathrm{k}_3(x_i,x_j,x_k)\Big)^{\alpha_k} , \quad k{=}2 \text{ or } 3. \tag{5.2.69}$$

The functions k_1, k_{2r} and k_{3s} are defined as follows:

$$\mathrm{k}_1(x_i,x_j,x_k)=\Big(s_{41}x_i+s_{42}x_j+s_{43}x_k+\beta_4\Big)\Big(s_{51}x_i+s_{52}x_j+s_{53}x_k+\beta_5\Big)+s_{54}s_{44} , \tag{5.2.70}$$

$$\mathrm{k}_{2r}(x_i) = 1+exp\Big(\beta_1 t_r + s_{11}t_r x_i\Big) \tag{5.2.71}$$

and

$$\mathrm{k}_{3s}(x_i,x_j,x_k) = 1+exp\Big(\beta_3-\beta_2 t_s+(s_{31}-s_{21}t_s)x_i+(s_{32}-s_{22}t_s)x_j+s_{33}x_k\Big) . \tag{5.2.72}$$

Proof

The calculation of $E(x_1 f_{1r}\cdot x_k f_{ks})$, $k{=}2$ or 3 is similar to the calculation of $E(x_2 f_{2r}\cdot x_3 f_{3s})$ described in Proposition 5.2.5 except that instead of 6 parameters there are only 5 parameters. The transformation (5.2.55) would thus have one dimension less and be as follows

$$\mathbf{w}{:}(5\mathrm{x}1)= \frac{1}{\sqrt{2}}S^{*-1}(\mathbf{v}-\beta) , \tag{5.2.73}$$

where

$$\mathbf{v} = \begin{bmatrix} v_1\\ v_2\\ v_3\\ v_4\\ v_5 \end{bmatrix} = \begin{bmatrix} x_4\\ x_5\\ x_6\\ x_1\\ x_2 \end{bmatrix} \tag{5.2.74}$$

and

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} x_4 \\ x_7 \\ x_8 \\ x_1 \\ x_3 \end{bmatrix} \tag{5.2.75}$$

when calculating $E(x_1 f_{1r} \cdot x_2 f_{2s})$ and $E(x_1 f_{1r} \cdot x_3 f_{3s})$ respectively. □

It has been shown how to calculate $E(x_i f_{ir} \cdot x_j f_{js})$ as well as $E(f_{ir})$ for i,j=1,2,3 and r,s=1,2,···,n. It is thus possible to calculate the covariance matrix, $\mathrm{Cov}(\mathbf{f},\mathbf{f}')$, (cf. (5.2.31)) for an eight parameter multi-component model. It is now possible to compute $\boldsymbol{\xi}$ (cf. (5.1.4)) and Σ (cf. (5.1.5)). The evaluation of the mean vector and covariance matrix, for the eight parameter model, outlined above has been implemented in the Fortran subroutine COVAX8.

Using the iterative generalised least squares approach it is possible to obtain the unknown parameters. The theory discussed above was applied to the Fels data (Examples 5.5.1 and 5.5.2) using the computer program, AUFITPC, developed by Du Toit and Browne (1991), along with the subroutine COVAX8 necessary to run the program.

5.3 THE LIKELIHOOD FUNCTION OF MULTI-COMPONENT RICHARDS MODELS

For model (5.1.1) let $\mathbf{f}(\mathbf{x},\boldsymbol{\alpha},\boldsymbol{t}) = \boldsymbol{F}\boldsymbol{x}_1$ (cf. (5.2.3)) with $\boldsymbol{F}$ and $\boldsymbol{x}_1$ given in (5.2.3) and (5.2.4) respectively and let

$$\boldsymbol{x}_2 = \begin{bmatrix} x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix}, \tag{5.3.1}$$

and partition $\boldsymbol{\Phi}^{-1}$ as follows

$$\boldsymbol{\Phi}^{-1} = \begin{bmatrix} \boldsymbol{\Phi}^{11} & \boldsymbol{\Phi}^{12} \\ \boldsymbol{\Phi}^{21} & \boldsymbol{\Phi}^{22} \end{bmatrix} . \qquad (5.3.2)$$

Under the priori assumption of normality for $\mathbf{x}$ it follows that

$$\mathrm{g}(\mathbf{x}) = (2\pi)^{-\mathrm{r}/2}|\boldsymbol{\Phi}|^{-1/2} exp\Big(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\theta})'\boldsymbol{\Phi}^{-1}(\mathbf{x}-\boldsymbol{\theta})\Big) . \qquad (5.3.3)$$

From the distributional assumptions given in Section 5.1 it follows that:

$$\mathrm{f}(\mathbf{y}\,|\,\mathbf{x}) = (2\pi)^{-\mathrm{n}/2}\,|\,\boldsymbol{\Lambda}\,|^{-1/2} exp\left(-\tfrac{1}{2}(\mathbf{y}-\mathbf{f}(\mathbf{x}))'\boldsymbol{\Lambda}^{-1}(\mathbf{y}-\mathbf{f}(\mathbf{x}))\right). \qquad (5.3.4)$$

Let

$$\boldsymbol{L}'\boldsymbol{L} = \boldsymbol{\Lambda}^{-1} \qquad (5.3.5)$$

and

$$\begin{aligned} \mathbf{y}^* &= \boldsymbol{L}\mathbf{y} , \\ \boldsymbol{F}^* &= \boldsymbol{L}\boldsymbol{F} , \\ \mathbf{q} &= \mathbf{y}^{*\prime}\boldsymbol{F}^*, \\ \boldsymbol{Q} &= \boldsymbol{F}^{*\prime}\boldsymbol{F} . \end{aligned} \qquad (5.3.6)$$

then the conditional p.d.f. (5.3.4) with $\mathbf{x}' = (\boldsymbol{x}_1',\boldsymbol{x}_2')$ and $\boldsymbol{x}_1$ defined in (5.2.3) can be written as follows:

$$\mathrm{f}(\mathbf{y}\,|\,\mathbf{x}) = (2\pi)^{-\mathrm{n}/2}\,|\,\boldsymbol{L}\,|\, exp\Big(-\tfrac{1}{2}\boldsymbol{x}_1\boldsymbol{Q}\boldsymbol{x}_1 + \mathbf{q}'\boldsymbol{x}_1 - \tfrac{1}{2}\mathbf{y}^{*\prime}\mathbf{y}^*\Big) . \qquad (5.3.7)$$

Use of (5.3.7) and (5.3.3) shows that

$$\mathrm{f}(\mathbf{y}\,|\,\mathbf{x})\mathrm{g}(\mathbf{x}) = (2\pi)^{(\mathrm{n}+\mathrm{r})/2}|\boldsymbol{\Phi}|^{-1/2}\,|\prod_{i=1}^{\mathrm{n}} l_{ii}\,|^{-1/2} exp[-\tfrac{1}{2}(\mathbf{y}^{*\prime}\mathbf{y}^* + \boldsymbol{Q}^*) , \qquad (5.3.8)$$

with

$$Q^* \ = -\tfrac{1}{2}\, \boldsymbol{x}_1' \boldsymbol{Q} \boldsymbol{x}_1 + \mathbf{q}'\boldsymbol{x}_1 - \tfrac{1}{2}(\mathbf{x}-\boldsymbol{\theta})'\boldsymbol{\Phi}^{-1}(\mathbf{x}-\boldsymbol{\theta}) \ . \qquad (5.3.9)$$

$\boldsymbol{Q}^*$ as defined in (5.3.9) with

$$\mathbf{s}':(1\,\mathrm{x}\,3) \ = \Big(\mathbf{q}' + \boldsymbol{\theta}'\boldsymbol{\Phi}^{11} - (\boldsymbol{x}_2 - \boldsymbol{\theta}_2)'\boldsymbol{\Phi}^{21}\Big)\,, \qquad (5.3.10)$$

$$\boldsymbol{\Psi}^{-1}:(8\,\mathrm{x}\,8) \ = \ \boldsymbol{Q} \ + \boldsymbol{\Phi}^{11}, \qquad (5.3.11)$$

is given by:

$$\boldsymbol{Q}^* \ = -\tfrac{1}{2} exp\Big(\boldsymbol{\theta}'_1 \boldsymbol{\Phi}^{11}\boldsymbol{\theta}_1 - \tfrac{1}{2}\boldsymbol{x}_1'\boldsymbol{\Psi}^{-1}\boldsymbol{x}_1 + \mathbf{s}'\boldsymbol{x}_1 + \boldsymbol{\theta}_1{}'\boldsymbol{\Phi}^{12}(\boldsymbol{x}_2 - \boldsymbol{\theta}_2) - \tfrac{1}{2}(\boldsymbol{x}_2 - \boldsymbol{\theta}_2)'\boldsymbol{\Phi}^{22}(\boldsymbol{x}_2 - \boldsymbol{\theta}_2)\Big)\,, \qquad (5.3.12)$$

Note that s has the same dimension as $\boldsymbol{x}_1$.

Proposition 5.3.1

$$\mathrm{h}(\mathbf{y}) \ = \ \int \mathrm{f}(\mathbf{y}|\mathbf{x})\mathrm{g}(\mathbf{x})\,d\mathbf{x}$$

$$= \ \mathrm{C}_1 \int \ |\,\boldsymbol{\Psi}\,|^{\frac{1}{2}} exp\Big(\tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} + \boldsymbol{\theta}_1'\boldsymbol{\Phi}^{12}(\boldsymbol{x}_2 - \boldsymbol{\theta}_2)\Big) \cdot exp\Big(-\tfrac{1}{2}\,(\boldsymbol{x}_2 - \boldsymbol{\theta}_2)'\boldsymbol{\Phi}^{22}(\boldsymbol{x}_2 - \boldsymbol{\theta}_2)\Big)d\boldsymbol{x}_2, \qquad (5.3.13)$$

where

$$\mathrm{C}_1 \ = \ (2\pi)^{\frac{-(6+\mathrm{n})}{2}} \ [\prod_{i=1}^{\mathrm{n}} l_{ii}] \ |\,\boldsymbol{\Phi}\,|^{-\frac{1}{2}} \ exp\Big(-\tfrac{1}{2}\mathbf{y}^{*\prime}\mathbf{y}^* - \tfrac{1}{2}\boldsymbol{\theta}_1'\boldsymbol{\Phi}^{11}\boldsymbol{\theta}_1\Big) \qquad (5.3.14)$$

Proof

From (5.3.8), with r = 8 parameters and

$$\mathrm{C}_2 \ = \ (2\pi)^{(\mathrm{n}+8)/2}\,|\prod_{i=1}^{\mathrm{n}} l_{ii}|^{-1/2}|\boldsymbol{\Phi}|^{-\frac{1}{2}} \ exp\Big(-\tfrac{1}{2}\,\mathbf{y}^{*\prime}\mathbf{y}^*\Big)\,, \qquad (5.3.15)$$

it follows that

$$h(\mathbf{y}) \;=\; C_2 \int \left(\int exp\, \boldsymbol{Q}^* \; d\boldsymbol{x}_1 \right) d\boldsymbol{x}_2 \; .$$

Using (5.3.9) h($\mathbf{y}$) can be expressed as

$$h(\mathbf{y}) \;=\; C_2 \; exp\Big(-\tfrac{1}{2}\boldsymbol{\theta}_1'\boldsymbol{\Phi}^{11}\boldsymbol{\theta}_1 \Big) \int \Big(\int exp\,(-\tfrac{1}{2}\boldsymbol{x}_1'\boldsymbol{\Psi}^{-1}\boldsymbol{x}_1 + \mathbf{s}'\boldsymbol{x}_1) \; d\boldsymbol{x}_1 \Big) \cdot \Big(exp\, \Big(\boldsymbol{\theta}_1{}'\boldsymbol{\Phi}^{12}(\boldsymbol{x}_2 - \boldsymbol{\theta}_2) - \tfrac{1}{2}(\boldsymbol{x}_2 - \boldsymbol{\theta}_2)'\boldsymbol{\Phi}^{22}(\boldsymbol{x}_2 - \boldsymbol{\theta}_2) \Big) \Big) d\boldsymbol{x}_2 \; .$$

But

$$\int exp\,(-\tfrac{1}{2}\boldsymbol{x}_1'\boldsymbol{\Psi}^{-1}\boldsymbol{x}_1 + \mathbf{s}'\boldsymbol{x}_1) \; d\boldsymbol{x}_1 \;=\; (2\pi)^{3/2} \,|\, \boldsymbol{\Psi} \,|\,^{1/2} exp\tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} \; . \qquad (5.3.16)$$

The result in (5.3.16) follows from the fact that for

$$\boldsymbol{v} : (p \times 1) \;\sim\; N(\mathbf{0}, \boldsymbol{\Psi})$$

$$M_{\boldsymbol{v}}(\boldsymbol{t}) \;=\; E(\, exp\,(\boldsymbol{t}'\boldsymbol{v})) \;=\; exp\tfrac{1}{2}\boldsymbol{t}'\boldsymbol{\Psi}\boldsymbol{t} \; .$$

□

Theorem 5.3.1

$$h(\mathbf{y}) \;\simeq\; C \sum_{\alpha}^{q^*} w_\alpha \,|\, \boldsymbol{\Psi} \,|\,^{\frac{1}{2}} \; exp\Big(\tfrac{1}{2}\mathbf{s}_\alpha{}'\boldsymbol{\Psi}_\alpha\mathbf{s}_\alpha + \mathbf{u}\boldsymbol{x}_\alpha \Big) \qquad (5.3.17)$$

where

$$\mathbf{u} \;=\; \sqrt{2}\boldsymbol{\theta}_1'\boldsymbol{\Phi}^{12}\boldsymbol{T}1 \qquad (5.3.18)$$

$$C \;=\; C_1 \cdot 2^{6/2} \,|\, \boldsymbol{\Phi}^{22} \,|\,^{-1/2} \, , \qquad (5.3.19)$$

$$\sum_{\alpha}^{q^*} = \sum_{i=1}^{q}\sum_{j=1}^{q} \cdots \sum_{m=1}^{q} , \qquad (5.3.20)$$

$$w_\alpha = w_i w_j w_k w_l w_m , \qquad (5.3.21)$$

$$\boldsymbol{x}_\alpha = (x_i, x_j, x_k, x_l, x_m)' . \qquad (5.3.22)$$

The w_i and x_i respectively denote the associated weights and nodes of a q-point Gauss quadrature numerical integration formula.

Proof

Let

$$\mathbf{z} = \frac{1}{\sqrt{2}} \boldsymbol{T}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\theta}_2), \qquad (5.3.23)$$

where $(\boldsymbol{\Phi}^{22})^{-1} = \boldsymbol{T}'\boldsymbol{T}$, with $\boldsymbol{T}$ a lower triangular matrix.

From (5.3.23) it follows that

$$\boldsymbol{x}_2 = \sqrt{2}\, \boldsymbol{T}\mathbf{z} + \boldsymbol{\theta}_2 , \qquad (5.3.24)$$

and the Jacobian of the transformation is

$$\left|\frac{\partial \boldsymbol{x}_2}{\partial \mathbf{z}'}\right| = 2^{5/2} |\boldsymbol{T}| = 2^{5/2} |\boldsymbol{\Phi}^{22}|^{-1/2} . \qquad (5.3.25)$$

Using (5.3.23) to (5.3.25) it follows that h($\mathbf{y}$) (cf. (5.3.13)) can be written as

$$\mathrm{h}(\mathbf{y}) = \mathrm{C} \int exp\ -\mathbf{z}'\mathbf{z} |\boldsymbol{\Psi}|^{1/2}\ exp\left(\tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} + \sqrt{2}\boldsymbol{\theta}_1'\boldsymbol{\Phi}^{12}\boldsymbol{T}\mathbf{z}\right)d\mathbf{z} . \qquad (5.3.26)$$

Approximating the 5–th dimensional integral

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{z}_1^2 - \mathrm{z}_2^2 - \cdots - \mathrm{z}_5^2} . \mathrm{f}(\mathbf{z}) d\mathbf{z} ,$$

by the Gauss quadrature formula (cf. Section 3.7)

$$\sum \sum \sum \sum \sum w_i \, w_j w_k w_l w_m \, f(x_i, x_j, x_k, x_l, x_m) \,,$$

for α denoting the summation over the nodes x_i, x_j, x_k, x_l, and x_m in the quadrature formula with weights w_i, w_j, w_k, w_l, and w_m the result in (5.3.17) follows. □

Since

$$\frac{\partial^{r+s}}{\partial s_i^r \partial s_j^s} \int exp\left(-\tfrac{1}{2}\boldsymbol{x}_1'\boldsymbol{\Psi}^{-1}\boldsymbol{x}_1 + \mathbf{s}'\boldsymbol{x}_1\right)d\boldsymbol{x}_1 = \int x_i^r \, x_j^s \; exp\left(-\tfrac{1}{2}\boldsymbol{x}_1'\boldsymbol{\Psi}^{-1}\boldsymbol{x}_1 + \mathbf{s}'\boldsymbol{x}_1\right)d\boldsymbol{x}_1 \,.$$

it follows that (cf. (5.3.16))

$$\int x_i^r \, x_j^s \; exp\left(-\tfrac{1}{2}\boldsymbol{x}_1'\boldsymbol{\Psi}^{-1}\boldsymbol{x}_1 + \mathbf{s}'\boldsymbol{x}_1\right)d\boldsymbol{x}_1 = (2\pi)^{3/2} \, | \boldsymbol{\Psi} |^{1/2} \frac{\partial^{r+s}}{\partial s_i^r \partial s_j^s} \; exp \tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} \qquad i,j=1,2,3 \qquad (5.3.27)$$

Consider the next three examples:

i) $$\int \boldsymbol{x}_1 \, exp\left(-\tfrac{1}{2}\boldsymbol{x}_1'\boldsymbol{\Psi}^{-1}\boldsymbol{x}_1 + \mathbf{s}'\boldsymbol{x}_1\right)d\boldsymbol{x}_1$$

$$= (2\pi)^{3/2} \, | \boldsymbol{\Psi} |^{1/2} \left(s_1\Psi_{11} + s_2\Psi_{21} + s_3\Psi_{31}\right) \, exp \, \tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} \,; \qquad (5.3.28)$$

ii) $$\int \boldsymbol{x}_1^2 \, exp\left(-\tfrac{1}{2}\boldsymbol{x}_1'\boldsymbol{\Psi}^{-1}\boldsymbol{x}_1 + \mathbf{s}'\boldsymbol{x}_1\right)d\boldsymbol{x}_1$$

$$= (2\pi)^{3/2} \, | \boldsymbol{\Psi} |^{1/2} \left(\Psi_{11} + (s_1\Psi_{11} + s_2\Psi_{21} + s_3\Psi_{31})^2 \right) \, exp \tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} \,; \qquad (5.3.29)$$

and

iii) $$\int x_1 x_2 \, exp \left(-\tfrac{1}{2}\boldsymbol{x}_1'\boldsymbol{\Psi}^{-1}\boldsymbol{x}_1 + \mathbf{s}'\boldsymbol{x}_1\right)d\boldsymbol{x}_1$$

$$= (2\pi)^{3/2} |\boldsymbol{\Psi}|^{1/2} \Big(\Psi_{21} + (s_1\Psi_{11} + s_2\Psi_{21} + s_3\Psi_{31})(s_2\Psi_{22} + s_1\Psi_{12} + s_3\Psi_{31}) \Big) exp \tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s}. \tag{5.3.30}$$

Approximate expressions for the Bayes estimator $E(\mathbf{x}\,|\,\mathbf{y})$ of the parameter vector $\mathbf{x}$ and for $Cov(\mathbf{x},\mathbf{x}'\,|\,\mathbf{y})$ using (cf. (3.3.20)) can be obtained as follows:

$$E(x_k^r\, x_m^s\,|\,\mathbf{y}) \quad = \frac{\int x_k^r x_m^s\, f(\mathbf{y}\,|\,\mathbf{x})g(\mathbf{x})d\mathbf{x}}{\int f(\mathbf{y}\,|\,\mathbf{x})g(\mathbf{x})d\mathbf{x}} .$$

Case 1 ($k,l\ =\ 1,2,3$):

$$E(x_k x_l\,|\,\mathbf{y}) \ = \frac{\sum_{\alpha=1}^{q^*} w_\alpha \ |\boldsymbol{\Psi}|^{1/2} \left(\Psi_{kl} \ + C_k C_l\right) exp\left(\tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} \ + \mathbf{u}'\mathbf{z}_\alpha\right)}{\sum_{\alpha=1}^{q^*} w_\alpha |\boldsymbol{\Psi}|^{1/2} exp\left(\tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} \ + \mathbf{u}'\mathbf{z}_\alpha\right)} \tag{5.3.31}$$

where, by comparing with equations (5.3.28) to (5.3.30), C_1, C_2 and C_3 are defined as:

$$C_1 \ = s_1\Psi_{11} \ + s_2\Psi_{21} \ + s_3\Psi_{31} \ ,$$

$$C_2 \ = s_2\Psi_{22} \ + s_1\Psi_{12} \ + s_3\Psi_{32} \ ,$$

$$C_3 \ = s_3\Psi_{33} \ + s_1\Psi_{13} \ + s_2\Psi_{23} \ . \tag{5.3.32}$$

Note that

$$C_i \ = \frac{\partial}{\partial s_i}\left(\tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s}\right). \qquad\qquad i=1,2,3$$

Case 2 ($k,l\ =\ 1,2,\cdots,5$):

$$E(x_{k+3}x_{l+3}\,|\,\mathbf{y}) \ = \frac{\sum_{\alpha=1}^{q^*} w_\alpha x_k x_l |\boldsymbol{\Psi}|^{\frac{1}{2}} exp\left(\tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} \ + \mathbf{u}'\mathbf{z}_\alpha\right)}{\sum_{\alpha=1}^{q^*} w_\alpha |\boldsymbol{\Psi}|^{\frac{1}{2}} exp\left(\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} \ + \mathbf{u}'\mathbf{z}_\alpha\right)} . \tag{5.3.33}$$

Case 3 ($k = 1,2,3$; $l = 1,2,\cdots,5$):

$$E(x_k x_{l+3} \mid \mathbf{y}) = \frac{\sum_{\alpha=1}^{q^*} w_\alpha C_k x_l \mid \boldsymbol{\Psi} \mid^{\frac{1}{2}} exp\left(\frac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} + \mathbf{u}'\mathbf{z}_\alpha\right)}{\sum_{\alpha=1}^{q^*} w_\alpha \mid \boldsymbol{\Psi} \mid^{\frac{1}{2}} exp\left(\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} + \mathbf{u}'\mathbf{z}_\alpha\right)} . \qquad (5.3.34)$$

Using equations (5.3.31), (5.3.33) and (5.3.34), the conditional moments $E(\mathbf{x}|\mathbf{y})$ and $Cov(\mathbf{x},\mathbf{x}'|\mathbf{y})$ can be evaluated. For example $Cov(x_1,x_2|\mathbf{y})=E(x_1x_2|\mathbf{y}) - E(x_1|\mathbf{y}) \cdot E(x_2|\mathbf{y})$.

Marginal maximum likelihood estimators (cf. Section 3.3) of the unknown parameters $\boldsymbol{\gamma}$ (cf. (5.1.3)) are obtained by maximizing the marginal likelihood,

$$L = \prod_{i=1}^{N} \int f(\mathbf{y}_i|\mathbf{x})g(\mathbf{x})d\mathbf{x} . \qquad (5.3.35)$$

From 5.3.35 it follows that

$$ln L = \sum_{i=1}^{N} ln\, h(\mathbf{y}_i) \qquad (5.3.36)$$

Each marginal p.d.f $h(\mathbf{y}_i)$ (cf. Theorem 5.3.1) can be numerically evaluated using an appropriate Gauss quadrature formula in which allowance is made for the number of repeated measurements n_i and for the set of time points t_{i1}, t_{i2}, $\cdots$, t_{in_i} on which these measurements were made.

The marginal likelihood equations for the parameters of the prior probability density function, $\boldsymbol{\varrho}$ (cf. (3.3.2)), and measurement error, $\boldsymbol{\tau}$ (cf. (3.3.4)), are obtained by differentiating and equating (5.3.35) to zero. This aspect is dealt with in the following Section.

5.4 DERIVATIVES OF THE LOG LIKELIHOOD FUNCTION.

Using the marginal maximum likelihood method of estimation, estimates of the unknown parameters $\boldsymbol{\varrho}$ (of the prior p.d.f $g(\mathbf{x})$, cf. (3.3.2)) and $\boldsymbol{\tau}$ (of the p.d.f. $f(\mathbf{y}|\mathbf{x})$, cf.(3.3.4)) are obtained by differentiating and equating (5.3.35) to zero as follows (cf. (3.3.10) and (3.3.11))

$$\frac{\partial ln\,\mathrm{L}}{\partial \varrho_k} = \sum_{i=1}^{N} \frac{1}{\mathrm{h}(\mathbf{y}_i)} \left(\frac{\partial \mathrm{h}(\mathbf{y}_i)}{\partial \varrho_k}\right) = \sum_{i=1}^{N} \mathrm{E}_{\mathbf{x}|\mathbf{y}_i} \left(\frac{\partial ln\,\mathrm{g}(\mathbf{x})}{\partial \varrho_k}\right) = 0\ , \quad k=1,2,\cdots,44 \tag{5.4.1}$$

$$\frac{\partial ln\,\mathrm{L}}{\partial \tau_l} = \sum_{i=1}^{N} \frac{1}{\mathrm{h}(\mathbf{y}_i)} \left(\frac{\partial \mathrm{h}(\mathbf{y}_i)}{\partial \tau_l}\right) = \sum_{i=1}^{N} \mathrm{E}_{\mathbf{x}|\mathbf{y}_i} \left(\frac{\partial ln\,\mathrm{f}(\mathbf{y}_i|\mathbf{x})}{\partial \tau_l}\right) = 0, \quad l=1,2,\cdots,\mathrm{m}\ . \tag{5.4.2}$$

Maximization of $ln\,\mathrm{L}$ is equivalent to the minimization of $\mathrm{F} = -\frac{2}{\mathrm{N}} ln\,\mathrm{L}$. Denote the gradient vector by $\boldsymbol{g}(\boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is defined in (5.1.3). It follows from (5.4.1) and Section 3.3 of Chapter 3, that the derivatives of F with respect to the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$, of the prior probability function are as follows:

$$\begin{aligned}\boldsymbol{g}(\boldsymbol{\theta}) &= -\frac{2}{\mathrm{N}} \sum_{i=1}^{N} \mathrm{E}_{\mathbf{x}\,|\,\mathbf{y}_i}\left(\frac{\partial}{\partial \boldsymbol{\theta}} ln\,\mathrm{g}(\mathbf{x})\right) \\ &= -\frac{2}{\mathrm{N}} \sum_{i=1}^{N}\left(\boldsymbol{\Phi}^{-1}\, \mathrm{E}_{\mathbf{x}\,|\,\mathbf{y}_i}(\mathbf{x}-\boldsymbol{\theta})\right).\end{aligned} \tag{5.4.3}$$

Substituting $\hat{\boldsymbol{\theta}}$ defined in (3.3.13) into (5.4.3) it follows that

$$\boldsymbol{g}(\boldsymbol{\theta}) = -2\boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\ . \tag{5.4.4}$$

It also follows from (5.4.1) that

$$\begin{aligned}\boldsymbol{g}(\mathrm{vecs}\boldsymbol{\Phi}) &= -\frac{2}{\mathrm{N}} \sum_{i=1}^{N} \mathrm{E}_{\mathbf{x}\,|\,\mathbf{y}_i}\left(\frac{\partial}{\partial \mathrm{vecs}\boldsymbol{\Phi}} ln\, g(\boldsymbol{x})\right) \\ &= -\frac{2}{\mathrm{N}} \sum_{i=1}^{N} \mathrm{E}_{\mathbf{x}\,|\,\mathbf{y}_i}\left(-\tfrac{1}{2}\boldsymbol{\Phi}^{-1}\Big((\mathbf{x}-\boldsymbol{\theta})(\mathbf{x}-\boldsymbol{\theta})'-\boldsymbol{\Phi}\Big)\boldsymbol{\Phi}^{-1}G'\right) \\ &= \frac{1}{\mathrm{N}}\boldsymbol{\Phi}^{-1} \sum_{i=1}^{N} \mathrm{E}_{\mathbf{x}\,|\,\mathbf{y}_i}\Big(\big((\mathbf{x}-\mathbf{x}|\mathbf{y}_i)(\mathbf{x}-\mathbf{x}|\mathbf{y}_i)+(\mathbf{x}|\mathbf{y}_i-\boldsymbol{\theta})(\mathbf{x}|\mathbf{y}_i-\boldsymbol{\theta})'-\boldsymbol{\Phi}\big)\boldsymbol{\Phi}^{-1}G'\Big)\end{aligned} \tag{5.4.5}$$

Substituting $\hat{\boldsymbol{\Phi}}$ defined in (3.3.14) into (5.4.5) it follows that

$$\boldsymbol{g}(\text{vecs}\boldsymbol{\Phi}) \quad = \boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{\Phi}}-\boldsymbol{\Phi})\boldsymbol{\Phi}^{-1}\boldsymbol{G}'. \tag{5.4.6}$$

If for example $\boldsymbol{\Lambda}_i=\sigma^2\boldsymbol{I}_{\mathrm{n}_i}$, $\tau=\sigma^2$, it follows from (3.3.11) that

$$\boldsymbol{g}(\sigma^2) = \frac{\partial \mathrm{F}}{\partial \sigma^2} = -\tfrac{2}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} \mathrm{E}_{\mathbf{x}\,|\,\mathbf{y}_i}\Big(\frac{\partial}{\partial \sigma^2} ln\ \mathrm{f}(\mathbf{y}_i\,|\,\mathbf{x})\Big). \tag{5.4.7}$$

Denoting $\mathbf{f}(\mathbf{x}_i,\boldsymbol{\alpha},\boldsymbol{t}_i)$ by $\mathbf{f}_i$, it can be shown that

$$\mathrm{f}(\mathbf{y}_i\,|\,\mathbf{x}) \quad = (2\pi)^{\mathrm{n}_i/2}(\sigma^2)^{-\mathrm{n}_i/2} exp\ -\tfrac{1}{2\sigma^2}\Big(\mathbf{y}_i'\mathbf{y}_i+\mathbf{f}_i'\mathbf{f}_i-2\mathbf{y}_i'\mathbf{f}_i\Big)\ . \tag{5.4.8}$$

Therefore

$$ln\mathrm{f}(\mathbf{y}_i\,|\,\mathbf{x}) = \tfrac{\mathrm{n}_i}{2} ln\,(2\pi)-\tfrac{\mathrm{n}_i}{2} ln\,(\sigma^2)-\tfrac{1}{2\sigma^2}\Big(\mathbf{y}_i'\mathbf{y}_i+\mathbf{f}_i'\mathbf{f}_i-2\mathbf{y}_i'\mathbf{f}_i\Big)\ . \tag{5.4.9}$$

So that

$$\frac{\partial ln\,\mathrm{f}(\mathbf{y}_i\,|\,\mathbf{x})}{\partial \sigma^2} = -\frac{\mathrm{n}_i}{2\sigma^2} + \frac{1}{2\sigma^4}\Big(\mathbf{y}_i'\mathbf{y}_i\ + \mathbf{f}_i'\mathbf{f}_i-2\mathbf{y}_i'\mathbf{f}_i\Big)\ . \tag{5.4.10}$$

From (5.4.7) and (5.4.10) it follows that

$$\boldsymbol{g}(\sigma^2) = \frac{1}{\mathrm{N}\sigma^4}\sum_{i=1}^{\mathrm{N}}\bigg(\sigma^2\mathrm{n}_i-\Big(\mathbf{y}_i'\mathbf{y}_i\ +\mathrm{E}_{\mathbf{x}\,|\,\mathbf{y}_i}(\mathbf{f}_i'\mathbf{f}_i-2\mathbf{y}_i'\mathbf{f}_i)\Big)\bigg). \tag{5.4.11}$$

Solving $\boldsymbol{g}(\sigma^2)=0$, yields the maximum likelihood estimate for σ^2,

$$\hat{\sigma}^2 = \frac{1}{\sum \mathrm{n}_i}\sum_{i=1}^{\mathrm{N}}\mathbf{y}_i'\mathbf{y}_i + \sum_{i=1}^{\mathrm{N}}\mathrm{E}_{\mathbf{x}\,|\,\mathbf{y}_i}\Big(\mathbf{f}_i'\mathbf{f}_i-2\mathbf{y}_i'\mathbf{f}_i\Big)\ . \tag{5.4.12}$$

Let

$$\mathrm{h}_1 = (\mathrm{C}_1^2+\Psi_{11})\mathrm{Q}_{11}+(\mathrm{C}_2^2+\Psi_{22})\mathrm{Q}_{22}+(\mathrm{C}_3^2+\Psi_{33})\mathrm{Q}_{33},$$

$$\mathrm{h}_2 = 2(\mathrm{C}_1\mathrm{C}_2+\Psi_{12})\mathrm{Q}_{12}\ +2(\mathrm{C}_1\mathrm{C}_3+\Psi_{13})\mathrm{Q}_{13}+2(\mathrm{C}_2\mathrm{C}_3+\Psi_{23})\mathrm{Q}_{23},$$

and

$$h_3 \;=\; -2(C_1q_1 \;+C_2q_2 \;+C_3q_3) \;, \tag{5.4.13}$$

where Q_{ij} denotes a typical element of the (3x3) matrix $\boldsymbol{Q}$, q_i a typical element of the (3x1) vector $\mathbf{q}$ (cf. (5.3.6)), Ψ_{ij} a typical element of the (3x3) matrix $\boldsymbol{\Psi}$ defined by (5.3.11) and C_i, i=1,2,3 defined by (5.3.32).

Using Theorem 5.3.1 it follows that:

$$\begin{aligned} E_{\mathbf{x}\,|\,\mathbf{y}_i}(\mathbf{f}_i'\mathbf{f}_i - 2\mathbf{y}_i'\mathbf{f}_i) &= E_{\mathbf{x}\,|\,\mathbf{y}_i}\left(\boldsymbol{x}_1'\boldsymbol{F}_i'\boldsymbol{F}_i\boldsymbol{x}_1 - 2\mathbf{y}'\boldsymbol{F}_i\boldsymbol{x}_1\right) \\ &= \frac{\sum_{\alpha}^{q^*} w_\alpha\,|\,\boldsymbol{\Psi}\,|^{\frac{1}{2}}\; exp\left(\frac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} \;+\mathbf{u}'\boldsymbol{x}_\alpha\right)\cdot\left(h_1 + h_2 + h_3\right)}{\sum_{\alpha}^{q^*} w_\alpha\,|\,\boldsymbol{\Psi}\,|^{\frac{1}{2}}\; exp\left(\frac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} \;+\mathbf{u}'\boldsymbol{x}_\alpha\right)} \;, \end{aligned} \tag{5.4.14}$$

with α, w_α and $\boldsymbol{x}_\alpha$ as defined by (5.3.21) and (5.3.22).

Information Matrix

In the optimization algorithm the information matrix is used as an approximation to the Hessian, where (cf. Addendum to Chapter 3)

$$\mathcal{I} \;=\; E\left(\frac{\partial ln\,L}{\partial\gamma_i}\cdot\frac{\partial ln\,L}{\partial\gamma_j}\right).$$

Under the assumption of multivariate normal third and fourth order moments, the expressions for $\mathcal{I}(\boldsymbol{\theta})$, $\mathcal{I}(\boldsymbol{\Phi})$ and $\mathcal{I}(\boldsymbol{\theta},\boldsymbol{\Phi})$ are similar to those given by Bock(1990) where

$$E_{\mathbf{x}\,|\,\mathbf{y}_i}\left((\mathbf{x} - E(\mathbf{x}\,|\,\mathbf{y}_i))(\mathbf{x} - E(\mathbf{x}\,|\,\mathbf{y}_i))'\right) \;=\; \Sigma_{\mathbf{x}\,|\,\mathbf{y}_i}$$

and

$$E(\mathbf{x}\,|\,\mathbf{y}_i) \;=\; \overline{\boldsymbol{\beta}}_i \,.$$

To calculate $\mathfrak{I}(\sigma^2)$, $\mathfrak{I}(\sigma^2,\boldsymbol{\Phi})$ and $\mathfrak{I}(\sigma^2,\boldsymbol{\theta})$ numerical derivatives were used. A FORTRAN program NLIN has been written to obtain estimates of the unknown parameters $\boldsymbol{\theta}$, $\boldsymbol{\Phi}$ and τ.

5.5 PRACTICAL APPLICATIONS.

In Example 5.5.1 the estimation of the unknown parameters for the male and female Fels data using the first approach discussed in Section 5.2 is considered.

The computer program AUFITPC written by Du Toit and Browne (1990) consists of two parts: The nucleus program AUFIT.OBJ which is supplemented by the user supplied subroutines COVAXI.FOR (for the evaluation of covariance structures), CONSTR.FOR (an optional subroutine for the evaluation of equality and/or inequality constraints) as well as SPECIN.FOR and SPCOUT.FOR (optional special input and output subroutines). The user has to supply a subroutine for the evaluation of the mean vector, $\boldsymbol{\xi}$, and covariance matrix, Σ. The subroutine, COVAX8 was developed to calculate the mean vector, $\boldsymbol{\xi}$, and the covariance matrix, Σ, for the three component eight parameter Richards model.

The Cross-Validation Index (CVI) is used in AUFITPC to assess the model fit once the parameters have been estimated (Browne & Cudeck, 1989). The CVI indicates to what extent a fitted moment structure derived from one sample can be expected to fit another sample of the same size from the same population. Assuming that the discrepancy function is correctly specified for the distribution of the data, the expected cross-validation index is obtained from a single sample using

$$\mathrm{CVI} = \hat{\mathrm{F}} + 2\mathrm{q}/\mathrm{N}. \tag{5.5.1}$$

where

$$\hat{\mathrm{F}} = -\tfrac{2}{\mathrm{N}} ln\,\mathrm{L}(\hat{\gamma}) \tag{5.5.2}$$

The point estimate of the CVI is linearly related to the Akaike Information Criterion (Akaike 1973) and will lead to the same conclusions. It can be seen that a penalty for the number, k, of parameters is added to $\hat{\mathrm{F}}$ in (5.5.1). If N becomes very large, however, the effect of this penalty becomes negligible. An interval estimate of the CVI is obtained from a confidence interval on the noncentrality parameter of a chi-squared distribution. The CVI of the saturated model, where no structure is imposed on μ and Σ, is as follows:

$$\text{CVI(Saturated Model)} = p(p+3)/N. \tag{5.5.3}$$

The CVI takes inaccuracy in parameter estimates obtained from a sample of specified size into account and N does appear in (5.5.1).

Example 5.5.1

Use was made of two datasets from the Fels institute, the one being the height measurements (cm.) of 158 males and the other one the corresponding measurements of 132 females. For both the males and females the number of responses per individual differ. A subset of the dataset for both males and females were created to obtain an equal number of observations for all the individuals. Growth measurements made at intervals of length 0.5 years were selected. Measurements made at timepoints 2.491, 2.991, 3.491, 3.991, 4.491, $\cdots$, 16.491 years, regardless of the number of the experimental unit, were selected for both the male and female datasets.

The following 8-parameter triple-logistic function was employed to describe human growth over the period 2 years to 17 years.

$$f(\mathbf{x},\alpha,t_r) = x_1[1+(1+e^{x_4})^{-t_r}]^{\alpha} + x_2[1+e^{x_5}(1+e^{x_6})^{-t_r}]^{\alpha} + x_3[(1+e^{x_7}(1+e^{x_8})^{-t_r}]^{\alpha} \tag{5.5.4}$$

$$t_r = (\ 2.491\ \ 2.991\ \ 3.491\ \ 3.991\ \ 4.491\ \ 4.991\ \ 5.491\ \ 5.991\ \ 6.491\ \ 6.991\ \ 7.491\ \ 7.991\ \ 8.491\ \ 8.991\ \ 9.491\ \ 9.991\ \ 10.491\ \ 10.991\ \ 11.491\ \ 11.991\ \ 12.491\ \ 12.991\ \ 13.491\ \ 13.991\ \ 14.491\ \ 14.991\ \ 15.491\ \ 15.991\ \ 16.491\ )$$

where $\alpha = \alpha_1 = \alpha_2 = \alpha_3 = -1$. (5.5.5)

The vector **x**, of the growth parameters, is assumed to be multivariate normally distributed N($\boldsymbol{\theta}$,$\boldsymbol{\Phi}$) in the population from which the data cases were drawn.

It was first assumed that $\boldsymbol{\Lambda}=\sigma^2\boldsymbol{I}$. Thereafter a better fit was obtained for the same male and female datasets when the residuals were assumed to be generated by an $ARMA(1,1)$ process (see Chapter 6) with unknown parameters α, β and σ^2. Part of these outputs for the male and female datasets are given below. The input of the male dataset along with a description of the input is given in Appendix B.

OUTPUT OF AUFIT PROGRAM:

FELS DATA WITH 29 RESPONSES AT THE SAME TIMEPOINTS FOR 96 MALES

1) MEASURES OF FIT OF THE MODEL FOR MALE DATA
SAMPLE DISCREPANCY FUNCTION VALUE : 9.164 (0.916357E + 01)
EXPECTED CROSS-VALIDATION INDEX
POINT ESTIMATE (MODIFIED AIC) : 9.226
90 PERCENT CONFIDENCE INTERVAL : (8.393 ;10.140)
CVI (MODIFIED AIC) FOR THE SATURATED MODEL: 9.667

2) SAMPLE MEANS:

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
92.872	96.845	100.763	104.386	107.996	111.713	114.998	118.293	121.470	124.600
T11	T12	T13	T14	T15	T16	T17	T18	T19	T20
127.640	130.655	133.461	135.798	138.374	141.532	144.324	146.951	149.787	152.759
T21	T22	T23	T24	T25	T26	T27	T28	T29	
155.994	159.770	163.351	166.841	169.938	173.352	175.819	177.564	178.648	

3) REPRODUCED MEANS :

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
93.156	97.626	101.557	105.175	108.618	111.965	115.254	118.498	121.694	124.829
T11	T12	T13	T14	T15	T16	T17	T18	T19	T20
127.889	130.859	133.727	136.486	139.140	141.707	144.224	146.753	149.390	152.249

T21	T22	T23	T24	T25	T26	T27	T28	T29
155.431	158.961	162.732	166.515	170.043	173.108	175.616	177.582	179.086

4) SAMPLE COVARIANCE MATRIX :

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
T1	10.6														
T2	11.2	12.4													
T3	12.0	13.3	14.7												
T4	12.2	13.6	14.9	15.8											
T5	12.1	13.6	15.0	15.7	16.5										
T6	12.7	14.0	15.4	16.2	16.9	18.2									
T7	12.9	14.2	15.7	16.6	17.3	18.3	19.1								
T8	13.1	14.6	16.1	17.1	17.7	18.7	19.4	20.3							
T9	13.3	14.8	16.4	17.4	18.0	19.1	19.8	20.6	21.5						
T10	13.9	15.5	17.1	18.1	18.9	20.1	20.9	21.6	22.4	23.9					
T11	13.7	15.3	17.0	18.1	18.8	20.0	20.8	21.7	22.5	23.8	24.3				
T12	13.9	15.5	17.3	18.5	19.3	20.5	21.2	22.1	23.0	24.2	24.7	25.6			
T13	14.2	16.1	17.9	19.0	19.8	21.1	21.8	22.8	23.7	25.1	25.7	26.3	27.7		
T14	14.8	16.7	18.6	19.7	20.6	21.9	22.6	23.5	24.5	26.0	26.5	27.2	28.4	30.3	
T15	14.8	16.6	18.6	19.6	20.5	21.8	22.5	23.5	24.5	26.0	26.6	27.2	28.5	30.2	30.9
T16	14.9	16.9	18.9	20.1	21.0	22.3	23.1	24.3	25.3	26.8	27.5	28.2	29.5	30.9	31.3
T17	15.0	17.0	19.1	20.2	21.3	22.6	23.3	24.3	25.5	27.0	27.8	28.4	29.8	31.4	31.7
T18	15.5	17.5	19.6	20.8	21.8	23.2	24.0	25.1	26.2	27.6	28.5	29.2	30.7	32.1	32.6
T19	15.8	17.9	20.3	21.4	22.5	23.8	24.6	25.7	26.9	28.4	29.4	30.2	31.7	33.3	33.9
T20	16.5	18.9	21.4	22.6	23.5	24.7	25.7	27.1	28.3	29.8	30.8	31.7	33.2	35.0	35.6
T21	17.1	19.4	22.0	23.1	23.9	25.2	26.2	27.6	28.8	30.5	31.7	32.5	34.3	35.8	36.7
T22	17.5	19.8	22.6	23.6	24.4	25.7	26.8	28.2	29.5	31.1	32.4	33.4	35.1	36.5	37.5
T23	17.6	20.0	22.7	23.6	24.3	25.6	26.7	28.1	29.6	31.2	32.5	33.5	35.4	36.7	37.8
T24	18.5	20.7	23.6	24.3	25.0	26.6	27.5	29.0	30.6	32.2	33.3	34.5	36.3	37.5	38.9
T25	18.4	20.5	23.1	23.9	24.5	26.1	26.8	28.4	29.9	31.5	32.5	33.7	35.3	36.9	38.0
T26	17.2	19.2	21.6	22.2	23.0	24.6	25.2	26.8	28.2	29.6	30.5	31.7	33.3	34.7	35.6
T27	16.4	18.3	20.5	21.3	22.0	23.7	24.1	25.6	26.9	28.4	29.1	30.4	31.6	33.0	33.6
T28	15.5	17.4	19.6	20.4	21.1	22.5	23.1	24.4	25.5	27.0	27.7	29.0	30.1	31.6	32.2

T29	15.2	17.0	19.1	19.8	20.5	21.8	22.4	23.5	24.6	26.0	26.6	27.9	28.9	30.5	31.0

	T16	T17	T18	T19	T20	T21	T22	T23	T24	T25	T26	T27	T28	T29
T16	32.6													
T17	32.8	33.8												
T18	33.8	34.4	35.9											
T19	34.9	35.7	37.2	39.3										
T20	36.8	37.6	39.4	41.6	45.2									
T21	37.8	38.7	40.8	43.5	47.2	51.5								
T22	38.6	39.4	42.0	45.0	49.3	54.5	59.6							
T23	39.1	39.8	42.4	45.5	50.0	55.7	61.3	64.4						
T24	40.0	40.5	43.5	46.1	50.6	55.9	61.7	64.8	67.5					
T25	39.2	39.6	42.1	44.4	48.4	52.7	57.3	60.5	63.7	62.4				
T26	36.7	37.1	39.3	41.2	44.7	48.0	51.8	54.6	57.8	57.2	54.5			
T27	34.7	34.9	36.7	38.4	41.2	43.3	45.7	47.8	50.9	51.2	49.2	47.2		
T28	33.0	33.4	34.8	36.4	38.7	40.1	41.7	43.3	45.9	46.4	45.2	44.3	43.1	
T29	31.6	32.0	33.1	34.4	36.4	37.0	38.0	39.0	41.4	42.1	41.4	41.2	40.6	39.6

5) REPRODUCED COVARIANCE MATRIX :

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
T1	10.9														
T2	11.2	12.6													
T3	11.6	12.9	14.3												
T4	11.9	13.3	14.5	15.7											
T5	12.1	13.6	14.8	15.8	16.9										
T6	12.3	13.9	15.1	16.1	16.9	17.9									
T7	12.6	14.1	15.3	16.4	17.2	18.0	19.0								
T8	12.8	14.3	15.6	16.6	17.5	18.3	19.1	20.2							
T9	13.0	14.5	15.8	16.9	17.9	18.7	19.5	20.3	21.4						
T10	13.3	14.8	16.1	17.2	18.2	19.1	20.0	20.8	21.6	22.7					
T11	13.5	15.1	16.4	17.6	18.6	19.6	20.4	21.3	22.1	23.0	24.1				
T12	13.8	15.4	16.7	18.0	19.0	20.0	20.9	21.8	22.6	23.5	24.4	25.6			
T13	14.1	15.7	17.1	18.3	19.5	20.5	21.4	22.3	23.2	24.1	25.0	25.9	27.1		

T14	14.3	16.0	17.4	18.7	19.9	20.9	21.9	22.8	23.7	24.6	25.5	26.4	27.4	28.6
T15	14.5	16.2	17.8	19.1	20.3	21.3	22.3	23.3	24.2	25.2	26.1	27.0	28.0	29.0
T16	14.7	16.4	18.0	19.4	20.6	21.7	22.7	23.7	24.7	25.7	26.6	27.6	28.6	29.5
T17	14.8	16.6	18.2	19.6	20.9	22.0	23.1	24.1	25.1	26.1	27.1	28.1	29.1	30.2
T18	14.9	16.7	18.3	19.8	21.1	22.3	23.4	24.5	25.6	26.6	27.6	28.7	29.7	30.8
T19	15.0	16.8	18.4	19.9	21.3	22.6	23.8	25.0	26.1	27.2	28.2	29.3	30.4	31.5
T20	15.3	17.0	18.6	20.1	21.6	22.9	24.2	25.5	26.7	27.9	29.0	30.1	31.2	32.4
T21	15.8	17.4	18.9	20.5	21.9	23.4	24.8	26.1	27.4	28.7	29.9	31.0	32.1	33.3
T22	16.4	17.9	19.4	20.9	22.4	23.9	25.4	26.8	28.2	29.5	30.7	31.9	33.0	34.2
T23	17.0	18.4	19.8	21.3	22.8	24.3	25.8	27.2	28.6	29.9	31.2	32.4	33.5	34.7
T24	17.2	18.7	20.1	21.5	23.0	24.4	25.9	27.3	28.6	29.9	31.2	32.3	33.5	34.6
T25	17.0	18.6	20.0	21.4	22.8	24.1	25.4	26.7	28.0	29.3	30.4	31.6	32.7	33.8
T26	16.4	18.1	19.5	20.9	22.2	23.4	24.6	25.8	26.9	28.1	29.2	30.3	31.4	32.4
T27	15.7	17.5	19.0	20.3	21.5	22.6	23.7	24.7	25.8	26.8	27.9	28.9	30.0	31.0
T28	15.0	16.8	18.4	19.7	20.8	21.8	22.7	23.7	24.6	25.6	26.6	27.6	28.6	29.6
T29	14.3	16.2	17.8	19.1	20.2	21.1	22.0	22.8	23.7	24.6	25.5	26.5	27.4	28.4

	T16	T17	T18	T19	T20	T21	T22	T23	T24	T25	T26	T27	T28	T29	
T15	30.2														
T16	30.6	31.9													
T17	31.2	32.3	33.8												
T18	31.9	33.1	34.4	36.2											
T19	32.7	34.0	35.5	37.2	39.5										
T20	33.6	35.0	36.6	38.6	41.1	44.3									
T21	34.5	36.0	37.7	40.0	43.0	46.6	50.7								
T22	35.4	36.8	38.6	41.1	44.3	48.5	53.1	57.5							
T23	35.9	37.2	39.0	41.3	44.6	49.0	54.1	59.0	62.2						
T24	35.7	37.0	38.6	40.7	43.7	47.8	52.9	58.0	61.7	63.0					
T25	34.9	36.0	37.4	39.2	41.7	45.2	49.7	54.5	58.4	60.2	59.3				
T26	33.5	34.5	35.7	37.1	39.1	41.8	45.4	49.4	53.0	55.2	55.2	53.3			
T27	31.9	32.9	33.9	35.0	36.5	38.4	41.0	44.1	47.2	49.5	50.2	49.3	47.5		
T28	30.5	31.4	32.3	33.1	34.1	35.4	37.1	39.4	41.8	43.9	45.1	45.1	44.3	43.2	
T29	29.3	30.2	30.9	31.6	32.3	33.0	34.1	35.6	37.4	39.3	40.8	41.5	41.6	41.2	40.9

6) ESTIMATES OF THE ELEMENTS OF $\mathbf{\Phi}$

46.26993							
-2.74716	0.18700						
21.31685	-1.26545	33.63064					
0.60121	-0.03624	0.24014	0.01755				
-0.88388	0.05664	-0.30813	-0.01156	0.01858			
-31.02166	1.84258	-17.67083	-0.39811	0.57867	31.73510		
21.34192	-1.26558	8.87019	0.27888	-0.41142	-16.93392	15.2155	
1.78576	-0.09941	0.88299	0.02069	-0.03254	-1.64857	1.39933	0.14553

7) ESTIMATES OF THE ELEMENTS OF $\boldsymbol{\theta}$:

76.74165 0.66951 87.44015 1.80262 -1.09774 21.44730 18.98271 1.06660

8) ESTIMATES OF TIMESERIES PARAMETERS:

$\hat{\sigma}^2$=0.55982 $\hat{\alpha}$=0.42975 $\hat{\beta}$=0.11233 .

FELS DATA WITH 29 RESPONSES AT THE SAME TIMEPOINTS FOR 72 FEMALES

1) MEASURES OF FIT OF THE MODEL

SAMPLE DISCREPANCY FUNCTION VALUE	:10.540	(0.105396E + 02)
EXPECTED CROSS-VALIDATION INDEX		
POINT ESTIMATE (MODIFIED AIC)	:11.845	
90 PERCENT CONFIDENCE INTERVAL	: (10.827 ;12.972)	
CVI (MODIFIED AIC) FOR THE SATURATED MODEL	:12.889	

2) SAMPLE MEANS:

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
91.349	95.418	99.242	102.926	106.483	109.927	113.394	116.654	119.954	122.889
T11	T12	T13	T14	T15	T16	T17	T18	T19	T20
125.900	128.797	131.614	133.878	136.690	139.900	143.419	146.601	149.972	153.449
T21	T22	T23	T24	T25	T26	T27	T28	T29	
156.732	159.517	161.593	162.819	163.836	165.068	165.806	166.300	166.535	

3) REPRODUCED MEANS :

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
91.217	95.457	99.266	102.861	106.352	109.785	113.163	116.472	119.686	122.782
T11	T12	T13	T14	T15	T16	T17	T18	T19	T20
125.742	128.563	131.268	133.907	136.570	139.374	142.430	145.771	149.304	152.816
T21	T22	T23	T24	T25	T26	T27	T28	T29	
156.059	158.844	161.092	162.823	164.118	165.074	165.780	166.307	166.708	

4) SAMPLE COVARIANCE MATRIX :

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
T1	8.9														
T2	9.2	10.6													
T3	9.9	11.3	12.8												
T4	10.4	12.1	13.5	14.8											
T5	11.0	12.7	14.2	15.5	17.0										
T6	10.9	12.7	14.3	15.5	17.0	17.5									
T7	11.4	13.2	14.8	16.0	17.5	17.9	18.8								
T8	11.9	13.9	15.6	17.0	18.5	19.0	19.8	21.2							
T9	12.1	14.0	15.8	17.2	18.9	19.3	20.2	21.6	23.0						
T10	12.2	14.1	16.0	17.4	18.9	19.3	20.0	21.5	22.4	22.7					
T11	12.7	14.6	16.6	17.9	19.6	20.0	20.9	22.5	23.6	23.6	25.1				
T12	12.8	14.8	16.7	18.1	19.8	20.3	21.3	22.7	23.9	24.0	25.3	26.1			
T13	12.5	14.4	16.3	17.7	19.4	19.9	21.0	22.4	23.6	23.7	25.2	25.8	26.0		
T14	13.0	15.1	17.0	18.5	20.4	20.7	21.7	23.2	24.3	24.5	25.9	26.6	26.7	28.2	
T15	13.5	15.4	17.2	18.7	20.5	20.9	22.0	23.5	24.7	24.9	26.4	27.1	27.3	28.6	29.7
T16	13.8	15.6	17.4	19.0	20.9	21.5	22.8	24.3	25.5	25.6	27.3	28.2	28.5	29.8	31.0
T17	14.6	16.3	18.3	19.9	21.8	22.2	23.5	25.1	26.5	26.5	28.3	29.3	29.5	30.8	32.4
T18	14.9	16.4	18.6	20.1	22.1	22.5	23.9	25.3	26.9	26.8	28.5	29.7	29.8	31.2	32.9
T19	14.6	16.2	18.4	20.1	22.1	22.6	24.1	25.5	27.1	26.9	28.6	29.8	30.2	31.6	33.4
T20	15.0	16.7	19.1	21.0	23.1	23.5	25.1	26.6	28.3	27.9	29.6	30.9	31.2	32.7	34.3
T21	15.0	17.0	19.3	21.2	23.1	23.6	25.2	26.6	28.3	28.1	29.7	30.9	31.3	32.7	34.3
T22	13.8	15.9	18.2	20.1	21.8	22.3	23.7	25.0	26.5	26.3	27.8	28.8	29.1	30.4	31.6
T23	12.9	14.9	17.1	18.8	20.2	20.6	22.1	23.4	24.7	24.6	26.0	26.9	27.1	28.2	29.1

T24	11.6	13.8	15.7	17.3	18.7	19.1	20.4	21.5	22.5	22.5	23.7	24.6	24.7	25.8	26.4
T25	11.5	13.6	15.5	17.0	18.3	18.9	20.0	21.0	22.1	21.9	23.0	23.9	23.8	24.9	25.4
T26	10.4	12.3	14.2	15.4	16.5	17.1	18.1	19.1	20.4	20.4	21.4	22.3	22.1	22.7	23.1
T27	10.3	12.1	13.8	14.8	15.8	16.3	17.5	18.5	19.8	19.7	20.8	21.5	21.2	21.6	22.1
T28	9.8	11.6	13.3	14.1	15.1	15.6	16.7	17.7	19.0	18.7	19.9	20.6	20.3	20.6	21.0
T29	9.7	11.5	13.2	14.1	15.0	15.4	16.5	17.5	18.8	18.6	19.7	20.3	20.0	20.2	20.6

	T16	T17	T18	T19	T20	T21	T22	T23	T24	T25	T26	T27	T28	T29
T16	33.6													
T17	34.9	37.6												
T18	35.6	38.6	40.5											
T19	36.3	39.4	41.4	43.3										
T20	37.1	40.2	42.3	44.5	46.8									
T21	36.8	39.3	41.2	43.5	45.9	46.6								
T22	33.6	35.5	37.1	39.1	41.7	43.0	40.9							
T23	30.7	32.1	33.2	34.8	37.3	38.9	37.3	35.4						
T24	27.5	28.4	29.4	30.5	32.5	34.1	33.1	31.8	30.1					
T25	26.1	26.9	27.6	28.4	30.4	32.5	32.0	31.4	30.0	31.5				
T26	23.7	24.5	24.9	25.2	26.8	28.6	28.3	28.3	27.4	28.8	28.2			
T27	22.3	23.3	23.5	23.7	25.3	27.0	26.8	27.0	26.2	27.7	27.3	27.8		
T28	21.3	22.2	22.4	22.4	23.9	25.6	25.4	25.8	25.2	26.9	26.6	27.0	26.6	
T29	20.7	21.6	21.6	21.5	23.0	24.8	24.8	25.2	24.7	26.3	26.4	26.9	26.4	26.7

5) REPRODUCED COVARIANCE MATRIX :

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
T1	9.3														
T2	9.7	11.2													
T3	10.2	11.7	13.2												
T4	10.7	12.3	13.8	15.2											
T5	11.1	12.9	14.4	15.7	17.1										
T6	11.6	13.4	15.0	16.3	17.5	18.8									
T7	11.9	13.9	15.5	16.9	18.1	19.2	20.5								
T8	12.3	14.3	16.0	17.4	18.6	19.8	20.8	22.0							

T9	12.6	14.6	16.3	17.8	19.1	20.3	21.3	22.3	23.4						
T10	12.8	14.9	16.7	18.2	19.5	20.7	21.8	22.7	23.7	24.7					
T11	13.0	15.1	16.9	18.5	19.8	21.1	22.2	23.1	24.0	24.9	25.9				
T12	13.2	15.3	17.1	18.7	20.1	21.4	22.5	23.5	24.4	25.2	26.0	26.9			
T13	13.4	15.5	17.3	18.9	20.3	21.6	22.8	23.8	24.8	25.6	26.4	27.1	28.1		
T14	13.6	15.7	17.5	19.2	20.6	21.9	23.1	24.2	25.2	26.1	26.9	27.6	28.5	29.7	
T15	13.9	16.0	17.9	19.5	21.1	22.4	23.7	24.8	25.8	26.7	27.6	28.4	29.3	30.5	32.2
T16	14.5	16.5	18.4	20.2	21.7	23.2	24.5	25.7	26.7	27.7	28.6	29.5	30.5	31.8	33.7
T17	15.2	17.3	19.2	21.0	22.7	24.2	25.6	26.9	28.0	29.0	30.0	30.9	32.0	33.5	35.7
T18	16.0	18.1	20.1	22.0	23.8	25.4	26.9	28.2	29.4	30.4	31.4	32.4	33.6	35.2	37.5
T19	16.5	18.7	20.8	22.7	24.5	26.2	27.7	29.1	30.3	31.4	32.3	33.3	34.5	36.1	38.5
T20	16.5	18.7	20.9	22.8	24.7	26.3	27.8	29.2	30.4	31.4	32.4	33.4	34.4	35.9	38.1
T21	15.9	18.2	20.3	22.2	23.9	25.5	27.0	28.3	29.4	30.4	31.4	32.2	33.2	34.5	36.3
T22	14.9	17.1	19.1	20.9	22.6	24.1	25.4	26.6	27.7	28.6	29.5	30.3	31.1	32.1	33.5
T23	13.7	15.8	17.7	19.4	20.9	22.3	23.5	24.6	25.6	26.4	27.2	27.9	28.6	29.4	30.5
T24	12.6	14.6	16.3	17.9	19.2	20.5	21.6	22.6	23.5	24.3	25.0	25.6	26.2	26.8	27.5
T25	11.6	13.5	15.1	16.5	17.8	18.9	19.9	20.8	21.7	22.4	23.0	23.6	24.1	24.6	25.1
T26	10.8	12.6	14.2	15.5	16.7	17.7	18.6	19.5	20.2	20.9	21.5	22.1	22.5	22.9	23.2
T27	10.3	12.0	13.5	14.7	15.8	16.8	17.6	18.4	19.2	19.8	20.4	20.9	21.3	21.6	21.9
T28	9.9	11.6	13.0	14.2	15.2	16.1	17.0	17.7	18.4	19.1	19.6	20.1	20.5	20.8	20.9
T29	9.6	11.3	12.6	13.8	14.8	15.7	16.5	17.2	17.9	18.5	19.1	19.5	19.9	20.2	20.3

	T16	T17	T18	T19	T20	T21	T22	T23	T24	T25	T26	T27	T28	T29
T16	36.3													
T17	38.6	42.2												
T18	40.8	44.7	48.6											
T19	41.8	46.0	50.3	53.3										
T20	41.2	45.3	49.7	52.9	53.9									
T21	39.0	42.6	46.6	49.9	51.3	50.3								
T22	35.6	38.5	41.9	44.9	46.7	46.3	44.3							
T23	32.0	34.1	36.8	39.4	41.1	41.5	40.2	38.2						
T24	28.6	30.1	32.1	34.1	35.8	36.5	36.1	34.9	33.3					
T25	25.8	26.8	28.2	29.8	31.3	32.2	32.4	31.9	31.0	30.2				

T26 23.7 24.3 25.3 26.5 27.9 28.9 29.4 29.5 29.2 28.7 28.4

T27 22.1 22.6 23.3 24.2 25.4 26.5 27.2 27.7 27.7 27.6 27.4 27.5

T28 21.1 21.3 21.8 22.6 23.7 24.8 25.7 26.4 26.7 26.8 26.8 26.9 27.1

T29 20.4 20.5 20.9 21.6 22.6 23.7 24.6 25.5 26.0 26.3 26.4 26.5 26.7 27.0

6) ESTIMATES OF THE ELEMENTS OF $\mathbf{\Phi}$:

36.37360							
-2.09097	0.15911						
-15.40616	0.88381	29.13850					
0.92392	-0.04093	-0.64422	0.04108				
0.11233	0.00263	-0.10920	0.00783	0.00622			
-8.96931	0.51314	-1.86263	-0.16557	-0.01292	13.91069		
12.61590	-0.81447	-4.49620	0.28273	0.01564	-5.24238	7.13797	
1.27870	-0.08387	-0.15928	0.02161	0.00325	-0.90309	0.77360	0.10543

7) ESTIMATES OF THE ELEMENTS OF $\boldsymbol{\theta}$:

73.60763 0.92061 75.67389 1.81054 -0.88171 19.16425 15.98011 1.05029

8) ESTIMATES OF TIMESERIES PARAMETERS :

$\hat{\sigma}^2 = 0.52257$ $\qquad \hat{\alpha} = 0.54458$ $\qquad \hat{\beta} = 0.13758$.

Conclusion

The male data CVI point estimate 10.174 was within the 90% confidence interval. The female data the CVI point estimate 12.251 was also within the 90% confidence interval. The eight parameter logistic model seemed to fit well for both the male and female data.

In Example 5.5.2 the same datasets are analysed using the marginal maximum likelihood approach, discussed in Section 5.3, when estimating the unknown parameters.

Example 5.5.2

The theoretical results, described in Section 5.4, were implemented in the FORTRAN program NLIN. The computer program was designed to accommodate the analyses of single, double or triple component Richards models.

The two datasets being subsets of the growth data from the Fels institute, described and analysed in Example 5.5.1 above are considered here to compare the results of the estimation procedure described in Section 5.3 to the estimation procedure described in Section 5.3 and 5.4. Part of the computer output is given below. The analyses reported here were carried out under the assumption that the error terms are generated by an $ARMA(1{,}1)$ process (cf. Chapter 6). The unknown parameters are: $\boldsymbol{\theta}$, $\boldsymbol{\Phi}$, σ^2, α and β, that is a total of 47 parameters.

FELS DATA WITH 29 RESPONSES AT THE SAME TIMEPOINTS FOR 96 MALES

1) ESTIMATES OF THE ELEMENTS OF $\boldsymbol{\theta}$:

1	77.063515
2	0.545378
3	87.226905
4	1.818112
5	−1.114967
6	21.351310
7	18.361727
8	1.002274

2) ESTIMATES OF THE ELEMENTS OF $\boldsymbol{\Phi}$:

1	42.849613					
2	-1.892674	0.114719				
3	15.826220	-0.699014	26.561223			
4	0.725772	-0.034155	0.324034	0.022380		
5	-0.639478	0.033365	-0.238112	-0.013126	0.011331	
6	-25.920165	1.145144	-19.928187	-0.467133	0.387856	44.258444

7	16.407684	-0.722915	5.889351	0.277361	-0.244560	-19.932980	15.666908	
8	1.276104	-0.044769	0.390637	0.018566	-0.016550	-1.881860	1.525804	0.170849

3) ESTIMATES OF TIMESERIES PARAMETERS :

$\hat{\sigma}^2 = 0.551706$ $\hat{\alpha} = 0.42975$ $\hat{\beta} = 0.11233$

4) Sample discrepancy function value is 10.52 with degrees of freedom equal to 417.

FELS DATA WITH 29 RESPONSES AT THE SAME TIMEPOINTS FOR 72 FEMALES

1) ESTIMATES OF THE ELEMENTS OF $\boldsymbol{\theta}$:

1	74.066726
2	0.870408
3	75.748935
4	1.857916
5	−0.863224
6	18.798948
7	15.439837
8	0.987948

2) ESTIMATES OF THE ELEMENTS OF $\boldsymbol{\Phi}$:

1	30.983525							
2	-0.366152	0.059921						
3	-14.660585	0.172064	36.572845					
4	1.036158	0.011083	-0.717127	0.058089				
5	0.351636	0.005968	-0.524290	0.021780	0.017791			
6	-11.950499	0.141286	-0.780095	-0.350767	-0.057996	16.024422		
7	5.395030	-0.187010	0.765326	0.102403	-0.001754	-10.002325	9.979458	
8	0.882658	-0.019883	-0.222350	0.021856	0.011667	-1.316145	1.364687	0.203048

3) ESTIMATES OF TIMESERIES PARAMETERS :

$\hat{\sigma}^2 = 0.475874$ $\hat{\alpha} = 0.54458$ $\hat{\beta} = 0.13758$

4) Sample discrepancy function value is 11.75 with degrees of freedom equal to 417.

CONCLUSION

A comparison of the estimates of the unknown parameters obtained in Example 5.5.2 with those reported in Example 5.5.1 shows that there is a reasonable agreement between them. This remark holds for both the male and female data sets. It can be concluded that the method of analysis described in Section 5.3 and 5.4 is reliable and give accurate results.

Example 5.5.3

A single component nonlinear Richards function was fitted to a dataset containing measurements on mice reported by Williams and Izeman (1981) and analyzed by Rao (1984,1987) and Lee (1988,1991). This dataset, given in Appendix A, consists of the weights of 13 male mice measured on 7 occasions at intervals of 3 days over a period of 21 days, from birth to weaning. The single component Gompertz function (cf. (4.3.3)) used to describe individual growth over time is

$$f(\mathbf{x},\alpha,t) \quad = x_1[1 + exp(x_2 - x_3 t_i)]^{\alpha} \qquad \alpha = -1000 \qquad (5.5.6)$$

In Figure 5.5.1 the fitted single component Gompertz models for mouse numbers 1, 5 and 13, as well as the observed data points for these 3 mice are illustrated.

The Bayes estimates of the parameters x_1, x_2, x_3 of the 13 mice are as follows:

Mouse	$E(x_1\|\mathbf{y}_i)$	$E(x_2\|\mathbf{y}_i)$	$E(x_3\|\mathbf{y}_i)$
1	1.3221	-5.7651	0.16517
2	1.0820	-5.9891	0.15037
3	1.0395	-6.0276	0.14680
4	1.1868	-5.8426	0.17365

5	0.85345	-6.2139	0.12814
6	0.98292	-6.1375	0.12209
7	1.0409	-6.0407	0.13905
8	1.0631	-6.0109	0.14148
9	1.0243	-6.0580	0.14314
10	1.0123	-6.0357	0.15098
11	1.1147	-5.9524	0.15401
12	0.87185	-6.2252	0.11655
13	1.1853	-5.9535	0.15458

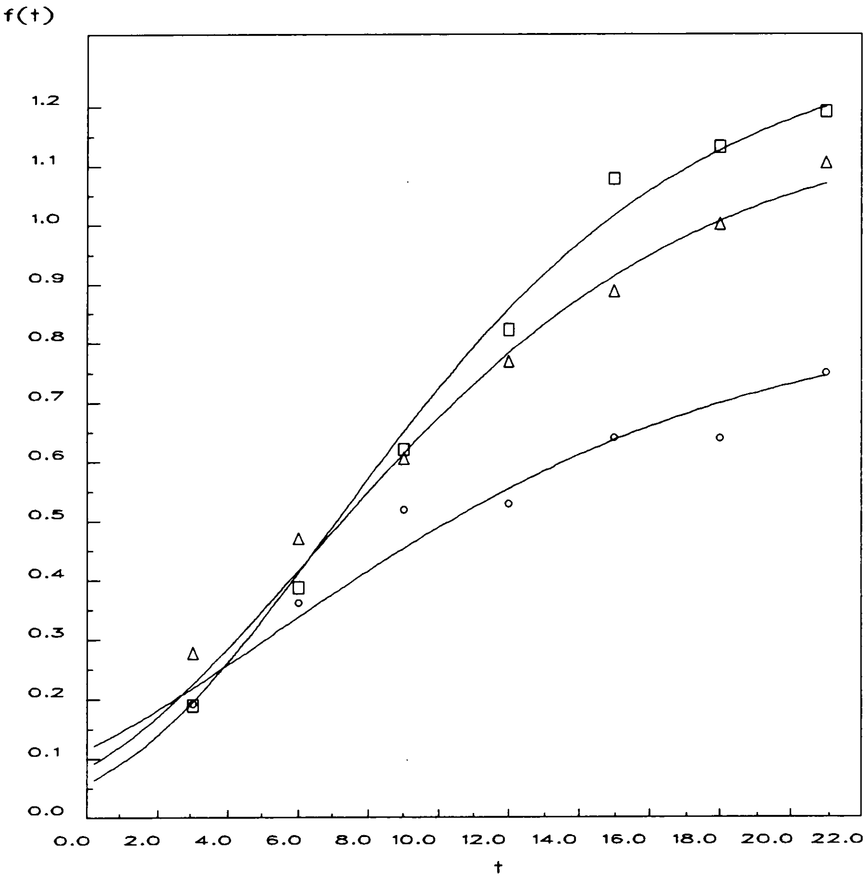

Figure 5.5.1 Graph of 3 fitted Grompertz curves as well as the observed data points.

Observed data points	**Fitted curve**
□	$f(\mathbf{x},\alpha,t) = 1.3221[1 + exp(-5.7651 - 0.16517]^{-1000}$
○	$f(\mathbf{x},\alpha,t) = 0.85345[1 + exp(-6.2139 - 0.12814]^{-1000}$
△	$f(\mathbf{x},\alpha,t) = 1.1853[1 + exp(-5.9353 - 0.15458]^{-1000}$

Example 5.5.4

In this example the **complete** datasets from the Fels growth study, the one being the height measurements (cm.) of N=158 males and the other one the corresponding measurements of N=132 females, were used. In Figures 5.5.2 and 5.5.3 graphical illustrations of the fitted models, for the first male and female in the Fels data set, are given. As in Examples 5.5.1 and 5.5.2 the 8 parameter function was used (cf. (5.5.4)).

Computer Output: Male data

(i) Estimates of unknown parameters:

ESTIMATES OF THE ELEMENTS OF $\boldsymbol{\theta}$:

1	96.43346
2	0.77303
3	64.87808
4	2.79312
5	0.37227
6	22.62562
7	17.40924
8	1.25737

ESTIMATES OF THE ELEMENTS OF $\boldsymbol{\Phi}$:

1	18.62534							
2	-0.31942	0.01131						
3	2.35331	-0.04036	5.41299					
4	0.01936	0.00226	0.00360	0.00177				
5	-0.04493	0.00238	-0.00775	0.00067	0.00093			
6	-4.13284	0.07091	-4.34873	-0.00483	0.01096	15.86118		
7	-0.08920	-0.01526	0.28878	-0.00764	-0.00450	0.13559	0.25168	
8	-0.02130	0.00162	-0.00688	0.00053	0.00040	0.02150	0.00442	0.00597

ESTIMATE OF σ^2

0.61938

(ii) Estimates of parameters given $\mathbf{y}_i$ (height measurements), Cases 1-2:

Case number : 1 $\hat{\boldsymbol{\theta}}|\mathbf{y}_i$

99.002	0.70961	65.341	2.7844	0.36683	20.629	17.071	1.1926

Case number : 2 $\hat{\boldsymbol{\theta}}|\mathbf{y}_i$

90.644	0.91804	64.561	2.8098	0.37890	27.693	17.488	1.2199

(iii) Estimates of covariance matrix given $\mathbf{y}_i$ (height measurements), Cases 1-2:

Case number : 1 $\hat{\boldsymbol{\Phi}}|\mathbf{y}_i$

2.2800	-0.54530E-01	0.14316E-02	-0.54060	0.92608E-02
0.56270	-0.11625E-02	0.23192E-03	0.99701E-03	0.64374E-03
-0.14927E-01	0.39220E-03	0.21533E-02	0.67908E-04	0.10989E-03
-1.6221	0.42200E-01	-0.40509E-02	0.47787E-03	0.11853E-01
1.6319	0.80436	-0.21203E-01	-0.73482E-01	-0.28410E-02
-0.58455E-02	-0.65927	0.39108	0.63752E-01	-0.16694E-02
-0.83934E-02	-0.24793E-03	-0.45898E-03	-0.51118E-01	0.29114E-01
0.22256E-02				

Case number : 2 $\hat{\boldsymbol{\Phi}}|\mathbf{y}_i$

3.0590	-0.17251	0.10301E-01	2.7351	-0.17508
6.1768	-0.41632E-01	0.28228E-02	-0.53437E-01	0.13288E-02
-0.44443E-01	0.26527E-02	-0.58545E-01	0.77368E-03	0.75578E-03
-5.0606	0.30369	-7.7589	0.82512E-01	0.89731E-01
11.331	0.60552	-0.36959E-01	0.72082	-0.10396E-01
-0.97935E-02	-1.1307	0.26128	0.44900E-01	-0.27167E-02
0.50003E-01	-0.76351E-03	-0.71156E-03	-0.82873E-01	0.16845E-01
0.11593E-02				

(iv) Value of $-2 \cdot$log-likelihood function at convergence.

F = 10694.4 Number of parameters are 45.

Computer Output: Female data

(i) Estimates of unknown parameters:

ESTIMATES OF THE ELEMENTS OF $\boldsymbol{\theta}$:

1	84.69261
2	1.09609
3	61.24890
4	2.45085
5	0.41993
6	20.51877
7	14.39602
8	1.23351

ESTIMATES OF THE ELEMENTS OF $\boldsymbol{\Phi}$:

1	11.39391							
2	-0.21002	0.01227						
3	3.57264	-0.06587	6.94299					
4	-0.09546	0.00807	-0.02127	0.00612				
5	-0.03179	0.00212	-0.01578	0.00161	0.00089			
6	-1.76134	0.03247	-1.14412	0.01387	0.00550	9.99570		
7	-0.54427	-0.01105	-0.47681	-0.01174	-0.00204	-1.05599	2.13510	
8	0.01982	-0.00073	-0.06802	-0.00057	-0.00005	-0.15950	0.19986	0.02583

ESTIMATE OF σ^2

0.50918

(ii) Estimates of parameters given $\mathbf{y}_i$ (height measurements), Cases 1-2:

Case number : 1 $\hat{\theta}|\mathbf{y}_i$

85.240	1.1523	60.523	2.4839	0.40686	18.249	13.899	1.2446

Case number : 2 $\hat{\theta}|\mathbf{y}_i$

86.157	0.91738	66.376	2.3643	0.44663	20.468	14.137	1.3184

(iii) Estimates of covariance matrix given $\mathbf{y}_i$ (height measurements), Cases 1-2:

Case number : 1 $\hat{\Phi}|\mathbf{y}_i$

0.18851	-0.23224E-03	0.51814E-04	-0.82558E-01	-0.11722E-02
2.5075	0.60454E-02	0.28108E-04	-0.21827E-02	0.59320E-03
-0.71222E-03	0.11655E-04	-0.14342E-01	0.36024E-04	0.10298E-03
-0.81487E-01	0.11652E-02	-2.1810	-0.39639E-02	0.13171E-01
2.0736	0.22708E-01	-0.88614E-03	0.73445	0.10582E-02
-0.42283E-02	-0.74780	0.81959	0.20630E-02	-0.72426E-04
0.67921E-01	0.10348E-03	-0.39011E-03	-0.69161E-01	0.75876E-01
0.70262E-02				

Case number : 2 $\hat{\Phi}|\mathbf{y}_i$

0.47877	-0.12265E-01	0.70606E-03	-0.86951E-01	-0.70063E-02
1.7182	-0.17514E-02	0.45466E-03	-0.12856E-01	0.78626E-03
-0.42045E-02	0.19861E-03	-0.11677E-01	0.19589E-03	0.14472E-03
-0.31676	0.15758E-01	-1.5566	0.11533E-01	0.14195E-01
1.7925	0.12329	-0.85108E-02	0.76867	-0.51329E-02
-0.59842E-02	-0.89722	1.0820	0.91157E-02	-0.65588E-03
0.69992E-01	-0.39010E-03	-0.51760E-03	-0.80261E-01	0.98780E-01
0.90408E-02				

(iv) Value of $-2 \cdot$log-likelihood function at convergence.

F = 13178.6 Number of parameters are 45.

In Figures 5.4.2 and 5.4.3 the male and female fitted models, obtained through the empirical Bayes analysis described in Section 5.5.3 and 5.5.4, along with the observed data, from the Fels growth data, are given respectively.

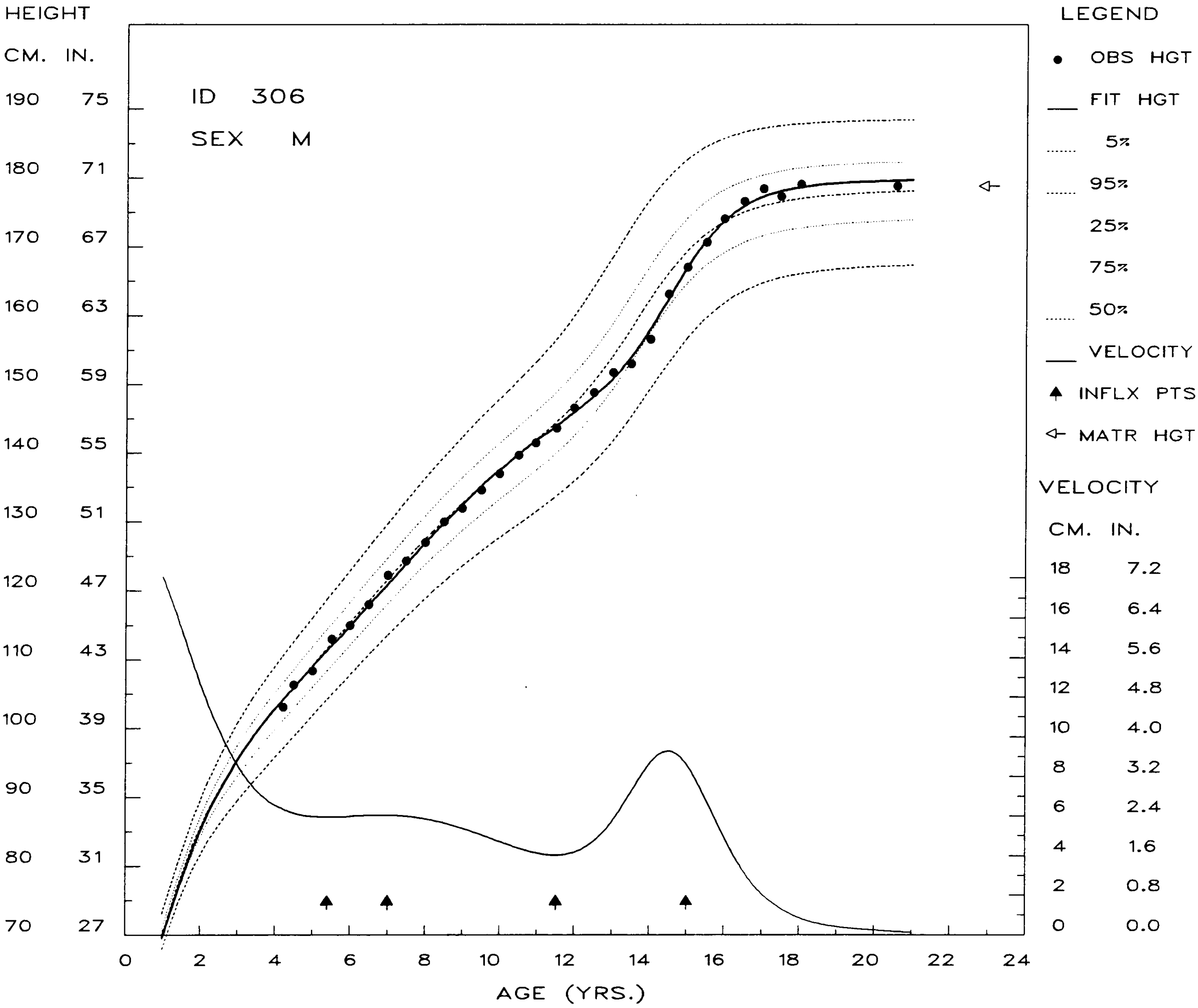

Figure 5.5.2 Raw and fitted data for a male

Additional analyses were also performed on the male and female datasets originating from the Berkeley growth study (Tuddenham and Snyder, 1984). The results of these analyses are reported in Appendix C.

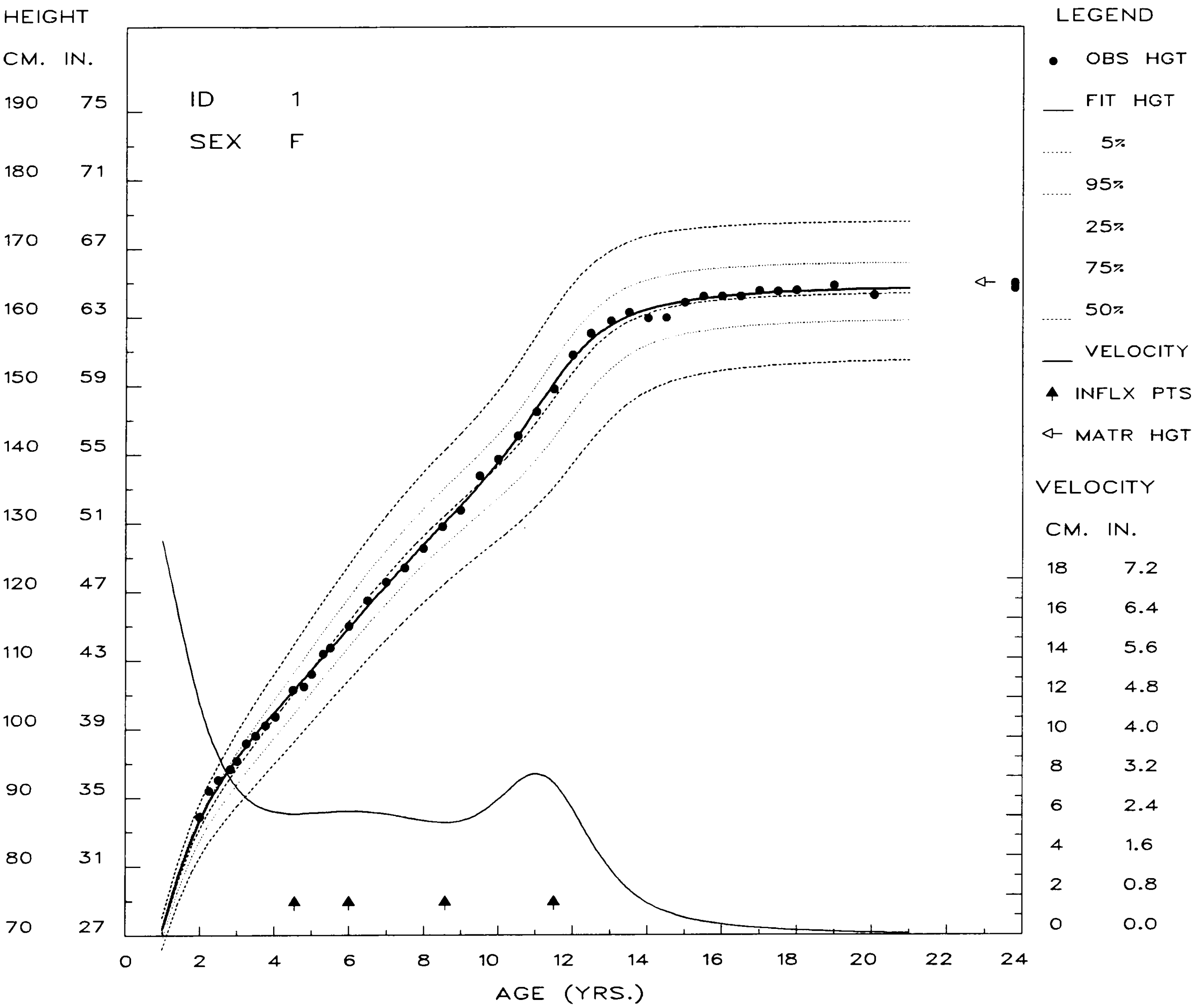

Figure 5.5.3 Raw and fitted data for a female

5.6 SUMMARY.

The multi-component Richards models, used for the description of growth, were considered in this chapter. Two approaches of estimating the unknown parameters were considered and compared. In the first approach of Section 5.2 use were made of the fact that although the function is no longer normally distributed consistent estimates are still obtained (see Anderson and Rubin (1956: 145-146)) using normal likelihood in the approach discussed in Section 3.4.

In the second approach, Sections 5.3 and 5.4, the analysis of the population dispersion of the parameters of multi-component Richards models using maximum marginal likelihood (MML), also known as the empirical Bayes method of estimation, is considered. Approximate expressions for the mean and covariance of the parameters were obtained for a three component Richards model. The Fortran program NLIN was used to apply the theory discussed in this chapter using the Fels and Berkeley growth data.

CHAPTER 6 MULTI-COMPONENT RICHARDS MODELS WITH $ARMA(1,1)$ DEVIATIONS.

6.1 INTRODUCTION.

In longitudinal data research with equally spaced measurements the first order autoregression, ($AR(1)$) model, have long been a popular model for the within subject error structure (Potthoff & Roy, 1964; Chi & Reinsel, 1989). Extensions to discrete time autoregressive-moving average, $ARMA$, processes have also been developed (Rochon & Helms, 1989). In this chapter methods for fitting a nonlinear model, which is commonly used in growth curve studies using an $ARMA(1,1)$ structure for the within subject errors, is presented for the general case of nonconsecutive measurements.

Estimation methods for $ARMA$ models were developed under the assumption that the observations are made at consecutive equally spaced intervals of time. In practice, however, there may often be a substantial number of missing values in the time series for various reasons. With nonconsecutive data, the usual "complete data" estimation procedures need to be generalised and estimation becomes more complicated because the covariance matrix of the observed variables is no longer Toeplitz, a structure that the complete data procedures either explicitly or implicitly make use of. Wincek and Reinsel (1986) gives an explicit procedure for obtaining the exact maximum likelihood estimates of the parameters in a regression-$ARMA$ time series model with possibly nonconsecutive data. The method of Wincek and Reinsel is based on an innovation transformation approach from which an explicit recursive procedure is derived for the efficient calculation of the exact likelihood function and associated derivatives.

In Section 6.2 a nonlinear *fixed* parameter regression model with non-consecutive $ARMA(1,1)$ deviations is considered. It is assumed that each element of the error vector is a weighted linear sum of past values and residual deviations.

The random parameter nonlinear model with $ARMA(1,1)$ deviations is introduced in Section 6.3. The theoretical principles of both Section 6.2 and 6.3 are applied in Section 6.4 by re-analysing the Fels data, as well as including some other examples.

6.2. NONLINEAR FIXED PARAMETER MODELS WITH NON-CONSECUTIVE *ARMA*(1,1) DEVIATIONS.

Suppose that the change in response pattern over time may be represented by the following fixed parameter nonlinear regression model.

$$\mathbf{y}_i = \mathrm{f}(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{t}) + \boldsymbol{\epsilon}_i, \qquad i = 1,2,\cdots,\mathrm{N} \tag{6.2.1}$$

where $\mathbf{y}_i$, $\mathbf{f}(\boldsymbol{\theta},\alpha,\boldsymbol{t})$ and $\boldsymbol{\epsilon}_i$ have typical elements y_{ij}, $\mathrm{f}(\boldsymbol{\theta},\alpha,t_j)$ and ϵ_{ij}, $j=1,2,\cdots,\mathrm{n}$ and where the response function is defined as follows:

$$\mathrm{f}(\boldsymbol{\theta},\boldsymbol{\alpha},t_j) = \theta_1(1 + \mathrm{e}^{-\theta_4 t_j})^{\alpha_1} + \theta_2(1 + \mathrm{e}^{\theta_5 - \theta_6 t_j})^{\alpha_2} + \theta_3(1 + \mathrm{e}^{\theta_7 - \theta_8 t_j})^{\alpha_3} . \tag{6.2.2}$$

Assume that $\boldsymbol{\epsilon}_1,\boldsymbol{\epsilon}_2,\cdots,\boldsymbol{\epsilon}_\mathrm{N}$ are a random sample of the random vector $\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ has a normal distribution with zero mean and covariance matrix $\boldsymbol{\Lambda}$. Assume also that each element ϵ_t of the vector $\boldsymbol{\epsilon}$ is a weighted linear sum of past values and residual deviations u_t:

$$\epsilon_t - \alpha\epsilon_{t-1} = \mathrm{u}_t - \beta\mathrm{u}_{t-1}, \qquad t = \cdots -1,0,1,\cdots \quad ; \tag{6.2.3}$$

which generates the covariance structure $\boldsymbol{\Lambda}$ of an *ARMA*(1,1) process.

The sequence $\{\mathrm{u}_t\}$ consists of Gaussian distributed variates, with $\mathrm{E}(\mathrm{u}_t)=0$ and $\mathrm{E}(\mathrm{u}_t\mathrm{u}_s) = \delta_{ts}\sigma^2$, where $\delta_{ts}=1$, $t=s$ and $\delta_{ts}=0$ otherwise, thus

$$\mathbf{u} \sim \mathrm{N}(\mathbf{0},\sigma^2\boldsymbol{I}). \tag{6.2.4}$$

Proposition 6.2.1

For $t=k+m$, the residual ϵ_{k+m} can be expressed in terms of ϵ_k and a linear combination of $\mathrm{u}_k,\mathrm{u}_{k+1},\cdots,\mathrm{u}_{k+m}$, where

$$\epsilon_{k+m} - \alpha^m \epsilon_k = -\alpha^{m-1}\beta u_k + \alpha^{m-2}\eta u_{k+1} + \cdots + \alpha\eta u_{k+m-2} + \eta u_{k+m-1} + u_{k+m} \tag{6.2.5}$$

and

$$\eta = (\alpha - \beta) \; . \tag{6.2.6}$$

Proof

Assume that the first observation was taken at time $t = k$ and that the next observation was taken at time $t = k + s$; s=1,2,$\cdots$,m. Therefore, (cf. (6.2.3)) at

$$t = k: \qquad \epsilon_k = u_k + x_k \; , \tag{6.2.7}$$

where

$$x_k = (\alpha\epsilon_{k-1} - \beta u_{k-1}) \; . \tag{6.2.8}$$

Similarly for $t = k+1$,

$$\epsilon_{k+1} - \alpha\epsilon_k = u_{k+1} - \beta u_k \; , \tag{6.2.9}$$

and for $t = k+2$,

$$\epsilon_{k+2} - \alpha\epsilon_{k+1} = u_{k+2} - \beta u_{k+1} \; . \tag{6.2.10}$$

Substitution of ϵ_{k+1} defined by (6.2.9) into (6.2.10) gives

$$\epsilon_{k+2} - \alpha^2\epsilon_k = -\alpha\beta u_k + (\alpha - \beta)u_{k+1} + u_{k+2} \; . \tag{6.2.11}$$

At $t = k+3$,

$$\epsilon_{k+3} - \alpha\epsilon_{k+2} = u_{k+3} - \beta u_{k+2} \; . \tag{6.2.12}$$

Substitution of ϵ_{k+2} defined by (6.2.11) into (6.2.12) gives

$$\epsilon_{k+3} - \alpha^3 \epsilon_k = -\alpha^2 \beta u_k + \alpha\eta u_{k+1} + \eta u_{k+2} + u_{k+3} ,$$

Continuation of the above procedure results in (6.2.5). □

Proposition 6.2.2

$$\boldsymbol{\epsilon} = \boldsymbol{T}_{\alpha}^{-1}\left(\boldsymbol{T}_{\beta}\mathbf{u} + \boldsymbol{I}_{n,1} x_{k_1}\right), \tag{6.2.13}$$

with

$$\boldsymbol{\epsilon}:(n \times 1) = (\epsilon_{k_1}, \epsilon_{k_2}, \cdots, \epsilon_{k_n})' , \tag{6.2.14}$$

and

$$\boldsymbol{T}_{\alpha}:(n \times n) = \begin{bmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ -\alpha^{m_2} & 1 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & -\alpha^{m_3} & 1 & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & -\alpha^{m_n} & 1 \end{bmatrix} \tag{6.2.15}$$

The $n^* \times 1$ vector $\mathbf{u}$ is defined by

$$\mathbf{u} = (u_{k_1}, u_{k_1+1}, \cdots, u_{k_2}, u_{k_2+1}, \cdots, u_{k_3}, u_{k_3+1}, \cdots u_{k_n})', \tag{6.2.16}$$

where

$$n^* = 1 + \sum_{i=2}^{n} m_i . \tag{6.2.17}$$

Note that in the case of consecutive data, $m_i = 1$, $i = 2,3,\cdots,n$ and therefore $n^* = n$.

The (nxn*) matrix $\boldsymbol{T}_\beta$ is defined as follows:

$$[\boldsymbol{T}_\beta]_{1,1} = 1; \qquad [\boldsymbol{T}_\beta]_{1,j} = 0; \qquad j=2,3,\cdots,n^*, \tag{6.2.18}$$

$$[\boldsymbol{T}_\beta]_{i,j} = \begin{cases} 0\ ; & j=1,2,\cdots,k_{i-1}-1 \\ -\beta\alpha^{m_i-1}; & j=k_{i-1} \\ \alpha^{m_i-1-l}; & j=k_{i-1}+l;\ l=1,\cdots,m_i-1 \\ 1; & j=k_i \\ 0; & j=k_i+1,\ k_i+2,\cdots,n^* \end{cases} \tag{6.2.19}$$

Proof

Suppose that n repeated measurements were made on occasions $t=t_1,t_2,\cdots,t_n$. For a stationary *ARMA* process, there is no loss in generality if we assume that $t_1=1$. Suppose therefore that the measurements were made on the occasions $t=k_1,k_2,\cdots,k_n$, where

$$k_i = t_i-(t_1-1), \quad i = 1,2,\cdots,n$$

with

$$k_1 = 1,$$

$$k_i = k_{i-1}+m_i \ , \quad i = 2,3,\cdots,n \tag{6.2.20}$$

and

$$m_i = 1,2,3,\cdots$$

We then obtain (cf. (6.2.5))

$$\boldsymbol{T}_\alpha\epsilon = \boldsymbol{T}_\beta\mathbf{u} + \boldsymbol{I}_{n,1}x_{k_1}\ , \tag{6.2.21}$$

the result in (6.2.13) follows from (6.2.21). □

Proposition 6.2.3

Denote the covariance matrix of ϵ by $\mathbf{\Lambda}$, then

$$\mathbf{\Lambda} = \mathrm{Cov}(\epsilon,\epsilon') = \sigma^2 [\boldsymbol{T}_\alpha]^{-1}(\boldsymbol{I}_{\mathrm{n},1}\mathrm{p}^*\boldsymbol{I}'_{\mathrm{n},1} + \boldsymbol{T}_\beta \boldsymbol{T}_\beta{}')[\boldsymbol{T}'_\alpha]^{-1} \tag{6.2.22}$$

with

$$\mathrm{p}^* = \frac{(\alpha-\beta)^2}{(1-\alpha^2)} \, . \tag{6.2.23}$$

Proof

It follows from (6.2.8) that for

$$t = k+1: \quad \mathrm{x}_{k+1} = (\alpha\epsilon_k - \beta \mathrm{u}_k) \, . \tag{6.2.24}$$

Substituting $\epsilon_k = \mathrm{u}_k + \mathrm{x}_k$ (cf. (6.2.7)) in (6.2.24)

$$\mathrm{x}_{k+1} = \alpha \mathrm{x}_k + (\alpha-\beta)\mathrm{u}_k \, , \tag{6.2.25}$$

where x_k and u_k are independent. Let p denote the variance of x_k. For a stationary process using (6.2.4) and (6.2.25) it follows that,

$$\mathrm{p} = \alpha^2\mathrm{p} + (\alpha-\beta)^2\sigma^2 \, ,$$

and therefore

$$\mathrm{p} = \frac{(\alpha-\beta)^2}{(1-\alpha^2)}\sigma^2 = \sigma^2\mathrm{p}^* \tag{6.2.26}$$

For $\mathbf{\Lambda} = \mathrm{Cov}(\epsilon,\epsilon')$ the result in (6.2.22) follows from (6.2.26) and (6.2.13) in Proposition 6.2.2. □

Theorem 6.2.1

Let (cf. (6.2.22))

$$\boldsymbol{A} = \boldsymbol{I}_{n,1} p^* \boldsymbol{I}'_{n,1} + \boldsymbol{T}_\beta \boldsymbol{T}'_\beta, \tag{6.2.27}$$

then for an $ARMA(1,1)$ process sampled at integer points $t_1, t_2, \cdots, t_n$, or equivalently $t = k_i$ where $k_i = t_i - (t_1 - 1)$, $i = 1,2,\cdots,n$; $k_i = k_{i-1} + m_i$; $i = 2,3,\cdots,n$; and $m_i = 1,2,3,\cdots$, $\boldsymbol{A}$ is a symmetric (nxn) tridiagonal matrix of the form:

$$\boldsymbol{A} = \begin{bmatrix} a_1 & b_2 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ b_2 & a_2 & b_3 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & b_3 & a_3 & b_4 & \cdot & \cdot & 0 & 0 \\ 0 & 0 & b_4 & a_4 & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & \cdot & a_{n-1} & b_n \\ 0 & 0 & 0 & 0 & \cdot & \cdot & b_n & a_n \end{bmatrix} \tag{6.2.28}$$

The elements of $\boldsymbol{A}$ are defined as follows:

$$a_1 = 1 + p^*, \tag{6.2.29}$$

with p^* defined by (6.2.23) , and for $i = 2,3,\cdots,n$:

$$a_i = 1 + \beta^2 \alpha^{2(m_i-1)} + \frac{\eta^2}{(1-\alpha^2)}(1 - \alpha^{2(m_i-1)}) \tag{6.2.30}$$

$$b_i = -\beta \alpha^{m_i - 1} \tag{6.2.31}$$

Proof

$$[\boldsymbol{I}_{\mathrm{n},1}\mathrm{p}^*\boldsymbol{I}'_{\mathrm{n},1}]_{i,j} = \mathrm{p}^*, \qquad i=j=1$$
$$= 0\ , \qquad \text{otherwise}$$

Therefore (cf. (6.2.19))

$$a_1 = 1+\mathrm{p}^*$$

Since $[\boldsymbol{T}_\beta\ \boldsymbol{T}_\beta']_{i,j} = \sum_{s=1}^{\mathrm{n}^*}[\boldsymbol{T}_\beta]_{i,s}[\boldsymbol{T}_\beta]_{j,s}$, it follows from (6.2.18) and (6.2.19) that

$$[\boldsymbol{T}_\beta\ \boldsymbol{T}_\beta']_{i,i} = \sum_{s=1}^{\mathrm{k}_{i-1}-1}[\boldsymbol{T}_\beta]^2_{i,s} + \sum_{s=\mathrm{k}_{i-1}}^{\mathrm{k}_i}[\boldsymbol{T}_\beta]^2_{i,s} + \sum_{s=\mathrm{k}_i+1}^{\mathrm{n}^*}[\boldsymbol{T}_\beta]^2_{i,s}$$

$$= 0+\sum_{s=\mathrm{k}_{i-1}}^{\mathrm{k}_i}[\boldsymbol{T}_\beta]^2_{i,s}+0$$

$$= \beta^2\alpha^{2(\mathrm{m}_i-1)}+\eta^2\alpha^{2(\mathrm{m}_i-2)}+\eta^2\alpha^{2(\mathrm{m}_i-3)}+\cdots+\eta^2+1$$

Using the well known result

$$\mathrm{a}+\ \mathrm{ar}+\ \mathrm{ar}^2\ +\ \cdots+\mathrm{ar}^{\mathrm{n}-1} = \frac{\mathrm{a}(1-\mathrm{r}^\mathrm{n})}{1-\mathrm{r}}\ , \qquad (6.2.32)$$

it follows that

$$[\boldsymbol{T}_\beta\ \boldsymbol{T}_\beta']_{i,i} = 1+\beta^2\alpha^{2(\mathrm{m}_i-1)}+\frac{\eta^2}{(1-\alpha^2)}(1-\alpha^{2(\mathrm{m}_i-1)}) \qquad (6.2.33)$$

Similarly, for $i=2,3,\cdots,\mathrm{n}$:

$$[\boldsymbol{T}_\beta\boldsymbol{T}_\beta']_{i,i-1} = 0+[\boldsymbol{T}_\beta]_{i,\mathrm{k}_{i-1}}[\boldsymbol{T}_\beta]_{i,\mathrm{k}_{i-1}}+0,$$

and from (6.2.19):

$$[\boldsymbol{T}_\beta \boldsymbol{T}_\beta']_{i,i-1} = [\boldsymbol{T}_\beta \boldsymbol{T}_\beta']_{i-1,i} = -\beta\alpha^{m_i-1} \tag{6.2.34}$$

$$[\boldsymbol{T}_\beta \boldsymbol{T}_\beta']_{i,j} = 0 \qquad |i-j| \geq 2 \tag{6.2.35}$$

For the elements of the matrix $\boldsymbol{A}$ defined in (6.2.27) and (6.2.28) it follows that $a_i = [\boldsymbol{T}_\beta \boldsymbol{T}_\beta']_{i,i}$ and $b_i = [\boldsymbol{T}_\beta \boldsymbol{T}_\beta']_{i,i-1}$ for $i = 2,3,\cdots$n. The results in (6.2.30) and (6.2.31) follows from (6.2.33) and (6.2.34) respectively. Finally, from (6.2.35) it follows that the remaining elements of $\boldsymbol{A}$ are all zero. □

Proposition 6.2.4

The tridiagonal matrix $\boldsymbol{A}$ defined by (6.2.28) can be written as the product of a lower and upper triangular matrix.

$$\boldsymbol{A} = \boldsymbol{LU}, \tag{6.2.36}$$

where

$$\boldsymbol{L} = \begin{bmatrix} \zeta_1 & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ b_2 & \zeta_2 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & b_3 & \zeta_3 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & 0 & b_4 & \zeta_4 & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & \cdot & \zeta_{n-1} & 0 \\ 0 & 0 & 0 & 0 & \cdot & \cdot & b_n & \zeta_n \end{bmatrix}, \tag{6.2.37}$$

$$U = \begin{bmatrix} 1 & \gamma_2 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & 1 & \gamma_3 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & 0 & 1 & \gamma_4 & \cdot & \cdot & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & \cdot & 1 & \gamma_n \\ 0 & 0 & 0 & 0 & \cdot & \cdot & 0 & 1 \end{bmatrix} , \tag{6.2.38}$$

and where

$$\begin{aligned} \zeta_1 &= a_1 \ , \\ \gamma_k &= b_k/\zeta_{k\text{-}1} \quad ; \qquad k=2,3,\cdots,n \\ \zeta_k &= a_k - b_k\gamma_k \ ; \qquad k=2,3,\cdots,n \end{aligned} \tag{6.2.39}$$

Proof

From (6.2.37) and (6.2.38) it follows that

$$LU = \begin{bmatrix} \zeta_1 & \zeta_1\gamma_2 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ b_2 & (b_2\gamma_2+\zeta_2) & \zeta_2\gamma_3 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & b_3 & (b_3\gamma_3+\zeta_3) & \zeta_3\gamma_4 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & b_4 & (b_4\gamma_4+\zeta_4) & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & b_n & (b_n\gamma_n+\zeta_n) \end{bmatrix}$$

Since $\boldsymbol{A} = \boldsymbol{LU}$, b_i, $i=2,\cdots,$n as defined by (6.2.37) is equal to the b_i defined by (6.2.31). It follows further that

$$\zeta_1 = a_1,$$

and also that

$$\zeta_{k-1}\gamma_k = b_k\ , \qquad k=2,3,\cdots,\text{n}$$

Therefore

$$\gamma_k = b_k/\zeta_{k-1}\ , \qquad k=2,3,\cdots,\text{n}$$

Finally,

$$b_k\gamma_k + \zeta_k = a_k\ , \qquad k=2,3,\cdots,\text{n}$$

and therefore

$$\zeta_k = a_k - b_k\gamma_k\ , \qquad k=2,3,\cdots,\text{n}$$

Note that the elements of $\boldsymbol{L}$ and $\boldsymbol{U}$ may be determined recursively from (6.2.39). □

Proposition 6.2.5

Let $\boldsymbol{A}$: (n x n) be the matrix defined by (6.2.27), and let

$$\mathbf{x}^* = \boldsymbol{A}^{-1}\mathbf{x}, \tag{6.2.40}$$

then $\mathbf{x}^*$ can be determined by operations of order n for a given $\mathbf{x}$.

Proof

From (6.2.40) and (6.2.36) we have

$$\boldsymbol{LU}\mathbf{x}^* = \mathbf{x} \,. \tag{6.2.41}$$

Let

$$\boldsymbol{U}\mathbf{x}^* = \mathbf{v} \,,$$

then (6.2.41) becomes

$$\boldsymbol{L}\mathbf{v} = \mathbf{x} \,. \tag{6.2.42}$$

We first use forward substitution to obtain a solution to $\mathbf{v}$ using (6.2.42) and then backward substitution to solve $\mathbf{x}^*$ from $\boldsymbol{U}\mathbf{x}^* = \mathbf{v}$. Note that the elements of $\boldsymbol{L}$ and $\boldsymbol{U}$ are both obtained during the forward substitution stage. □

Proposition 6.2.6

Let $\boldsymbol{\Lambda}$:(nxn) denote the covariance matrix of an $ARMA$(1,1) process for which the data is non-consecutive, then

$$|\boldsymbol{\Lambda}| = \sigma^{2\mathrm{n}} \prod_{i=1}^{\mathrm{n}} \zeta_i \tag{6.2.43}$$

with ζ_i, i=1,2,$\cdots$,n; defined by (6.2.39).

Proof

From (6.2.22) and (6.2.27) it follows that

$$|\boldsymbol{\Lambda}| = \sigma^{2\mathrm{n}}|\boldsymbol{A}| \,,$$

since (cf. (6.2.15)) $|\boldsymbol{T}_\alpha| = 1$.

But from Proposition 6.2.4,

$$\begin{aligned} |\boldsymbol{A}| &= |\boldsymbol{L}||\boldsymbol{U}| \\ &= \prod_{i=1}^{n} \zeta_i \,. \end{aligned}$$ □

Proposition 6.2.7

Let

$$Q = \tilde{\mathbf{y}}'\Lambda^{-1}\tilde{\mathbf{y}}, \tag{6.2.44}$$

then Q can be calculated in order n operations instead of the usual n^3 operations required for matrix inversion.

Proof

From (6.2.22) and (6.2.27) it follows that

$$\Lambda^{-1} = \sigma^{-2}\Lambda[\boldsymbol{T}_\alpha]'\boldsymbol{A}^{-1}\boldsymbol{T}_\alpha \tag{6.2.45}$$

Let

$$\mathbf{u} = \mathbf{T}_\alpha\,\tilde{\mathbf{y}}\,, \tag{6.2.46}$$

and

$$\mathbf{y}^* = \boldsymbol{A}^{-1}\mathbf{u}, \tag{6.2.47}$$

then

$$Q = \mathbf{u}'\mathbf{y}^*. \tag{6.2.48}$$

From (6.2.46) it follows that **u** may be obtained using simple forward substitution. Use of Proposition 6.2.5 then completes the proof. □

Under the assumptions above and that $\boldsymbol{\epsilon}$ is generated by an $ARMA(1,1)$ process (cf. (6.2.3)) and $\mathbf{u} \sim \mathrm{N}(\mathbf{0},\sigma^2\boldsymbol{I})$ (cf. (6.2.4)), it follows that

$$\mathbf{y} \sim \mathrm{N}(\mathrm{f}(\boldsymbol{\theta}), \boldsymbol{\Lambda}) \tag{6.2.49}$$

were $\boldsymbol{\Lambda}$ is defined by (6.2.22).

If $\{\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_N\}$ is a random sample of $\mathbf{y}$, it follows that the likelihood function, $\mathrm{L}(\boldsymbol{\theta},\boldsymbol{\Lambda})$, of $\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_N$ can be expressed as

$$\mathrm{L}(\boldsymbol{\theta},\boldsymbol{\Lambda}) = (2\pi)^{-\mathrm{Nn}/2}|\boldsymbol{\Lambda}|^{-\mathrm{N}/2} \exp -\tfrac{1}{2}\sum_{i=1}^{\mathrm{N}}\tilde{\mathbf{y}}_i'\boldsymbol{\Lambda}^{-1}\tilde{\mathbf{y}}_i, \tag{6.2.50}$$

where

$$\tilde{\mathbf{y}}_i = \mathbf{y}_i - \mathrm{f}(\boldsymbol{\theta}) \tag{6.2.51}$$

From (6.2.50) using (6.2.45) it follows that the logarithm of the likelihood function is given by

$$\begin{aligned} ln\,\mathrm{L}(\boldsymbol{\theta},\boldsymbol{\Lambda}) &= \tfrac{-\mathrm{nN}}{2}\,ln\,2\pi - \tfrac{\mathrm{N}}{2}ln\,|\boldsymbol{\Lambda}| - \frac{1}{2\sigma^2}\sum_{i=1}^{\mathrm{N}}\tilde{\mathbf{y}}_i'\boldsymbol{T}_\alpha'\boldsymbol{A}^{-1}\boldsymbol{T}_\alpha\tilde{\mathbf{y}}_i\,, \\ &= -\tfrac{1}{2}\sum_{i=1}^{\mathrm{N}}\left(\mathrm{n}\,ln\,(2\pi\sigma^2) + \sum_{i=1}^{\mathrm{n}} ln\,\zeta_i + \mathbf{u}_i'\mathbf{y}_i^*/\sigma^2\right) \end{aligned} \tag{6.2.52}$$

where $\mathbf{u}_i$ and $\mathbf{y}^*$ is given in (6.2.46) and (6.2.47) respectively and use is made of (6.2.43).

Consider a generalization of model (6.2.1) to the model

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\theta},\alpha,\boldsymbol{t}_i) + \boldsymbol{\epsilon}_i, \qquad i = 1,2,\cdots,\mathrm{N}, \tag{6.2.53}$$

where it is supposed that for individual i measurements were taken at time points $\mathrm{t}_{i1},\mathrm{t}_{i2},\cdots,\mathrm{t}_{in_i}$. Let us further suppose that $\mathrm{t}_{i,j+1} - \mathrm{t}_{ij} = \mathrm{m}_{ij}$, $j=2,3,\cdots,\mathrm{n}$. The covariance matrix $\boldsymbol{\Lambda}_i$ of $\boldsymbol{\epsilon}_i$ is obtained from (6.2.22), where m_{ij} replaces m_i in the definitions of $\boldsymbol{T}_\alpha$ and $\boldsymbol{T}_\beta$ (cf. (6.2.15) and (6.2.19)). Also (cf. (6.2.43))

$$|\boldsymbol{\Lambda}_i| = \sigma^{2\mathrm{n}}\prod_{j=1}^{\mathrm{n}_i}\zeta_{ij},$$

where ζ_{ij} denotes the diagonal elements of $\boldsymbol{L}_i$, and $\boldsymbol{L}_i$ corresponds to $\boldsymbol{L}$ (cf. (6.2.37) when m_i is replaced by m_{ij}, so that $\boldsymbol{A}_i = \boldsymbol{L}_i\boldsymbol{U}_i$.

For the general model described in (6.2.53) it follows that the log-likelihood function of $\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_\mathrm{N}$ is given by (cf. (6.2.52))

$$ln\,\mathrm{L}(\boldsymbol{\theta},\boldsymbol{\Lambda}^*) \;=\; -\tfrac{1}{2}\sum_{i=1}^{\mathrm{N}}\left(\mathrm{n}_i ln\,(2\pi\sigma^2) + \sum_{j=1}^{\mathrm{n}_i} ln\,(\zeta_{ij}) + \mathbf{u}_i'\mathbf{y}_i/\sigma^2\right) \qquad (6.2.54)$$

where, $\tilde{\mathbf{y}}_i = \mathbf{y}_i - \mathbf{f}(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{t}_i)$, $\mathbf{u}_i = \boldsymbol{T}_\alpha^*\,\tilde{\mathbf{y}}_i$, and $\mathbf{y}_i^* = \boldsymbol{A}_i^{-1}\mathbf{u}_i$.

Derivatives:

Let γ_s denote a typical element of the vector $\boldsymbol{\gamma}$ of unknown parameters. From (6.2.54) it follows that

$$\frac{\partial\, ln\,\mathrm{L}(\boldsymbol{\theta},\boldsymbol{\Lambda}^*)}{\partial\gamma_s} = -\tfrac{1}{2}\sum_{i=1}^{\mathrm{N}}\left(\frac{\partial}{\partial\gamma_s}(\mathbf{u}_i'\mathbf{y}_i^*/\sigma^2) + \sum_{j=1}^{\mathrm{n}_i}\left(\frac{\partial\zeta_{ij}}{\partial\gamma_s}\cdot\zeta_{ij}^{-1}\right)\right). \qquad (6.2.55)$$

From (6.2.55) it follows that

$$\frac{\partial\, ln\,\mathrm{L}(\boldsymbol{\theta},\boldsymbol{\Lambda}^*)}{\partial\sigma^2} = -\tfrac{1}{2}\sum_{i=1}^{\mathrm{N}}\left(\frac{\mathrm{n}_i}{\sigma^2} - \frac{\mathbf{u}_i'\mathbf{y}_i^*}{\sigma^4}\right) \qquad (6.2.56)$$

and

$$\frac{\partial\, ln\,\mathrm{L}(\boldsymbol{\theta},\boldsymbol{\Lambda}^*)}{\partial\alpha} = -\tfrac{1}{2}\sum_{i=1}^{\mathrm{N}}\left(\left(\sum_{j=1}^{\mathrm{n}_i}\frac{\frac{\partial\zeta_{ij}}{\partial\alpha}}{\zeta_{ij}} + \frac{1}{\sigma^2}\left(\frac{\mathbf{u}_i'\mathbf{y}_i^*}{\partial\alpha}\right)\right)\right) \qquad (6.2.57)$$

where

$$\begin{aligned}\frac{\partial}{\partial\alpha}\,\mathbf{u}_i'\mathbf{y}_i^* &= \frac{\partial}{\partial\alpha}\left(\mathbf{u}_i'\,\mathbf{A}_i^{-1}\mathbf{u}_i\right)\\ &= 2\frac{\partial\mathbf{u}_i'}{\partial\alpha}\mathbf{y}_i^* - \mathbf{u}_i'\left(\boldsymbol{A}_i^{-1}\frac{\partial\boldsymbol{A}_i}{\partial\alpha}\boldsymbol{A}_i^{-1}\right)\mathbf{u}_i\\ &= 2\frac{\partial\mathbf{u}_i'}{\partial\alpha}\mathbf{y}_i^* - \mathbf{y}_i^{*\prime}\left(\frac{\partial\boldsymbol{A}_i}{\partial\alpha}\right)\mathbf{y}_i^*\,. \qquad (6.2.58)\end{aligned}$$

It also follows that

$$\frac{\partial \mathbf{u}_i'}{\partial \alpha} = \left(\frac{\partial}{\partial \alpha}\boldsymbol{T}^*_{\alpha,i}\right)\tilde{\mathbf{y}}\,.$$

Similarly

$$\frac{\partial \, ln\, \mathrm{L}(\boldsymbol{\theta},\Lambda^*)}{\partial \beta} = 0 \; - \mathbf{y}_i^{*\prime}\left(\frac{\partial \boldsymbol{A}_i}{\partial \beta}\right)\mathbf{y}_i^*\,.$$

General remarks

Typical elements for $\boldsymbol{T}^*_\alpha$, $\boldsymbol{A}$, $\frac{\partial}{\partial \alpha}\boldsymbol{T}^*_\alpha$, $\frac{\partial \boldsymbol{A}}{\partial \alpha}$, $\frac{\partial \boldsymbol{A}}{\partial \beta}$, are calculated for $\mathrm{m}_{ij}=1,2,3,\cdots,\mathrm{M}$. These values (e.g. the off-diagonal elements of $\boldsymbol{T}^*_\alpha$, the diagonal elements and the off-diagonal elements of $\boldsymbol{A}$) are respectively stored as (M x 1) vectors. Appropriate elements are selected from these vectors, given the differences in time points for a specific individual. Suppose measurements were made as follows:

Individual 1:

$$\underbrace{20\;\;2}_{4}\underbrace{4\;\;2}_{1}\underbrace{5\;\;2}_{3}\underbrace{8\;\;3}_{2}\underbrace{0\;\;3}_{1}1\;\cdots\;\text{(months)}$$

Individual 2:

$$\underbrace{16\;\;1}_{1}\underbrace{7\;\;2}_{5}\underbrace{2\;\;2}_{1}\underbrace{3\;\;2}_{5}\underbrace{8\;\;3}_{2}\underbrace{0\;\;3}_{7}7\;\cdots\;\text{(months)}$$

At time $t=t_2$ we use m=4 for individual 1 and hence obtain the appropriate elements of $\boldsymbol{A}_i$, these being the 4^{th} element of the vectors containing information about the diagonal and off-diagonal elements of $\boldsymbol{A}$. Since all operations are of order (n), the procedure is computationally efficient since it avoids the actual inversion of matrices of high order.

To obtain $\frac{\partial \boldsymbol{A}}{\partial \gamma_s}$, expressions for $\frac{\partial a_i}{\partial \gamma_s}$ and $\frac{\partial b_i}{\partial \gamma_s}$ are required.

For the sake of completeness, expressions for the typical non-zero elements of $\boldsymbol{A}$ along with the first and second order derivatives are given in the addendum to this chapter in Section 6.5.

6.3. STOCHASTIC PARAMETER NONLINEAR RICHARD MODELS WITH $ARMA(1,1)$ DEVIATIONS.

Consider the following nonlinear regression model

$$\mathbf{y} = \mathbf{f}(\mathbf{x},\boldsymbol{\alpha},\boldsymbol{t}) + \boldsymbol{\epsilon}\ , \qquad i = 1,2,\cdots,N \qquad (6.3.1)$$

where (cf. Section 5.1) $\mathbf{f}(\mathbf{x},\boldsymbol{\alpha},\boldsymbol{t})$ denotes an nx1 vector-valued function with typical element $f(\mathbf{x},\boldsymbol{\alpha},t_j)$; $j=1,2,\cdots,n$ defined by (5.1.2). We assume that $\mathbf{x}$ is distributed as a $N(\boldsymbol{\theta},\boldsymbol{\Phi})$ variate, independent of $\boldsymbol{\epsilon}$ which is assumed to be independently distributed as a $N(\mathbf{0},\boldsymbol{\Lambda})$ random variate.

Assume also that $\boldsymbol{\Lambda}$ is the covariance matrix of an $ARMA(1,1)$ process corresponding to the case where the observed measurements are non-consecutive. Hence (cf. (6.2.40))

$$\boldsymbol{\Lambda}^{-1} = \sigma^{-2}\,[\boldsymbol{T}_\alpha]'\boldsymbol{A}^{-1}\boldsymbol{T}_\alpha \qquad (6.3.2)$$

Denote $\mathbf{f}(\mathbf{x},\boldsymbol{\alpha},\boldsymbol{t})$ by $\mathbf{f}(\mathbf{x})$. Since $f(\mathbf{y}|\mathbf{x})$ (cf. (5.3.4)) may be expressed as

$$f(\mathbf{y}\,|\,\mathbf{x}) = (2\pi)^{-n/2}\,|\,\boldsymbol{\Lambda}\,|^{-1/2} exp\Big(-\tfrac{1}{2}(\mathbf{y}-\mathbf{f}(\mathbf{x}))'\boldsymbol{\Lambda}^{-1}(\mathbf{y}-\mathbf{f}(\mathbf{x}))\Big),$$

it follows from (6.3.2) that

$$f(\mathbf{y}\,|\,\mathbf{x}) = (2\pi)^{-n/2}\,|\,\boldsymbol{A}\,|^{-1/2}(\sigma^2)^{-n/2} exp\Big(-\frac{1}{2\sigma^2}(\tilde{\mathbf{y}}-\tilde{\mathbf{f}}(\mathbf{x}))'\boldsymbol{A}^{-1}(\tilde{\mathbf{y}}-\tilde{\mathbf{f}}(\mathbf{x}))\Big), \qquad (6.3.3)$$

since (cf. (6.2.16)) $|\boldsymbol{T}^*_\alpha| = 1$.

In the expression (6.3.3) $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{f}}(\mathbf{x})$ are respectively defined by

$$\tilde{\mathbf{y}} = \boldsymbol{T}_\alpha\mathbf{y}\ , \qquad (6.3.4)$$

$$\tilde{\mathbf{f}}(\mathbf{x}) = \boldsymbol{T}_\alpha\mathbf{f}(\mathbf{x})\ . \qquad (6.3.5)$$

Proposition 6.3.1

The conditional pdf (6.3.3) may be expressed in the form

$$f(\mathbf{y}\,|\,\mathbf{x}) = (2\pi)^{-n/2}\Big(\prod_{i=1}^{n}\zeta_i\Big)^{-1/2}(\sigma^2)^{-n/2}exp\Big(-\frac{1}{2\sigma^2}(\tilde{\mathbf{y}}'\mathbf{y}^* + Q_f)\Big), \qquad (6.3.6)$$

where ζ_i, $i=1,2,\cdots,n$ are defined by (6.2.34),

$$\mathbf{y}^* = \boldsymbol{A}^{-1}\tilde{\mathbf{y}} \qquad (6.3.7)$$

$$Q_f = \boldsymbol{x}_1'\tilde{\boldsymbol{Q}}\boldsymbol{x}_1 - 2\tilde{\mathbf{q}}'\boldsymbol{x}_1, \qquad (6.3.8)$$

$$\tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{F}}'\boldsymbol{F}^*, \qquad (6.3.9)$$

$$\tilde{\boldsymbol{F}} = \boldsymbol{T}_\alpha\boldsymbol{F}, \qquad (6.3.10)$$

$$\boldsymbol{F}^* = \boldsymbol{A}^{-1}\tilde{\boldsymbol{F}}, \qquad (6.3.11)$$

and

$$\tilde{\mathbf{q}}' = \mathbf{y}^{*\prime}\tilde{\boldsymbol{F}} \qquad (6.3.12)$$

Proof

It was shown in Proposition 6.2.6 that

$$|\Lambda| = \sigma^{2n}\prod_{i=1}^{n}\zeta_i$$

Since (cf. (5.2.3))

$$f(\mathbf{x}) = \boldsymbol{F}\boldsymbol{x}_1, \qquad (6.3.13)$$

where $\boldsymbol{F}$ and $\boldsymbol{x}_1$ are defined in (5.2.3) and (5.2.4) respectively,

$$\mathbf{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix},$$

and the elements of $\boldsymbol{x}_1$ are the asymptotes of the three components, it follows that

$$(\tilde{\mathbf{y}}-\tilde{\mathbf{f}}(\mathbf{x}))'\boldsymbol{A}^{-1}(\tilde{\mathbf{y}}-\tilde{\mathbf{f}}(\mathbf{x})) = \tilde{\mathbf{y}}'\left(\boldsymbol{A}^{-1}\tilde{\mathbf{y}}\right)-2\tilde{\mathbf{y}}'\boldsymbol{A}^{-1}\tilde{\mathbf{f}}(\mathbf{x})+\tilde{\mathbf{f}}'(\mathbf{x})\boldsymbol{A}^{-1}\tilde{\mathbf{f}}(\mathbf{x}) \quad (6.3.14)$$

Finally, substitution of (6.3.5), (6.3.7), and (6.3.13) into (6.3.14) gives

$$\tilde{\mathbf{y}}'\mathbf{y}^* - 2\tilde{\mathbf{y}}'\boldsymbol{A}^{-1}\boldsymbol{T}_\alpha\boldsymbol{F}\boldsymbol{x}_1 + \boldsymbol{x}_1'(\boldsymbol{T}_\alpha\boldsymbol{F})'\boldsymbol{A}^{-1}\boldsymbol{T}_\alpha\boldsymbol{F}\boldsymbol{x}_1,$$

which using (6.3.10), (6.3.11) and (6.3.12) simplifies to

$$\tilde{\mathbf{y}}'\mathbf{y}^* - 2\tilde{\mathbf{y}}'\tilde{\boldsymbol{F}}\boldsymbol{x}_1 + \boldsymbol{x}_1'\tilde{\boldsymbol{F}}'\boldsymbol{F}^*\boldsymbol{x}_1,$$

$$= \tilde{\mathbf{y}}'\mathbf{y}^* - 2\tilde{\mathbf{q}}'\boldsymbol{x}_1 + \boldsymbol{x}_1'\tilde{\boldsymbol{Q}}\boldsymbol{x}_1 \, . \quad (6.3.15)$$

□

From Proposition 6.3.1 it further follows that

$$\mathrm{f}(\mathbf{y}\,|\,\mathbf{x})\mathrm{g}(\mathbf{x}) = (2\pi)^{(\mathrm{n}+\mathrm{r})/2}|\boldsymbol{\Phi}|^{-1/2}\Big(\prod_{i=1}^{\mathrm{n}}\zeta_i\Big)^{-1/2}(\sigma^2)^{-\mathrm{n}/2} exp\Big(-\frac{1}{2\sigma^2}(\tilde{\mathbf{y}}'\mathbf{y}^*+Q^*)\Big), \quad (6.3.16)$$

with

$$Q^* = \mathrm{Q}_f + \sigma^2(\mathbf{x}-\boldsymbol{\theta})'\boldsymbol{\Phi}^{-1}(\mathbf{x}-\boldsymbol{\theta}) \, . \quad (6.3.17)$$

Approximate expressions for the estimators of the parameter vector $\mathbf{x}$ can be obtained (cf. Section 5.3 of Chapter 5) from

$$\mathrm{E}(\mathrm{x}_k^{\mathrm{r}}\,\mathrm{x}_m^{\mathrm{s}}\,|\,\mathbf{y}) = \frac{\int \mathrm{x}_k^{\mathrm{r}}\mathrm{x}_m^{\mathrm{s}}\,\mathrm{f}(\mathbf{y}\,|\,\mathbf{x})\mathrm{g}(\mathbf{x})d\mathbf{x}}{\int \mathrm{f}(\mathbf{y}\,|\,\mathbf{x})\mathrm{g}(\mathbf{x})d\mathbf{x}} \, .$$

Let

$$\mathbf{s}' = \left(\tilde{\mathrm{q}}' + \boldsymbol{\theta}'\boldsymbol{\Phi}^{11} - (\boldsymbol{x}_2-\boldsymbol{\theta}_2)'\boldsymbol{\Phi}^{21}\right), \quad (6.3.18)$$

and

$$\boldsymbol{\Psi}^{-1} = \tilde{\boldsymbol{Q}} + \boldsymbol{\Phi}^{11}, \tag{6.3.19}$$

where $\boldsymbol{\Phi}^{-1}$ is defined in (5.3.2), $\tilde{\boldsymbol{Q}}$ in (6.3.9) and $\tilde{\mathbf{q}}$ in (6.3.12) then:

Case 1 (k,l = 1,2,3):

$$\mathrm{E}(\mathrm{x}_k \mathrm{x}_l \,|\, \mathbf{y}) = \frac{\sum_{\alpha=1}^{\mathrm{q}^*} \mathrm{w}_\alpha \,|\boldsymbol{\Psi}|^{1/2} \left(\Psi_{kl} + \mathrm{C}_k \mathrm{C}_l\right) exp\left(\tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} + \mathbf{u}'\mathbf{z}_\alpha\right)}{\sum_{\alpha=1}^{\mathrm{q}^*} \mathrm{w}_\alpha \,|\boldsymbol{\Psi}|^{1/2} exp\left(\tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} + \mathbf{u}'\mathbf{z}_\alpha\right)} \tag{6.3.20}$$

where,

$$\mathrm{C}_1 = \mathrm{s}_1\Psi_{11} + \mathrm{s}_2\Psi_{21} + \mathrm{s}_3\Psi_{31}\,,$$

$$\mathrm{C}_2 = \mathrm{s}_2\Psi_{22} + \mathrm{s}_1\Psi_{12} + \mathrm{s}_3\Psi_{32}\,,$$

$$\mathrm{C}_3 = \mathrm{s}_3\Psi_{33} + \mathrm{s}_1\Psi_{13} + \mathrm{s}_2\Psi_{23}\,. \tag{6.3.21}$$

Note that

$$\mathrm{C}_i = \frac{\partial}{\partial \mathrm{s}_i}\left(\tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s}\right). \qquad i = 1,2,3$$

Case 2 (k,l = 1,2,$\cdots$,6):

$$\mathrm{E}(\mathrm{x}_{k+3} \mathrm{x}_{l+3} \,|\, \mathbf{y}) = \frac{\sum_{\alpha=1}^{\mathrm{q}^*} \mathrm{w}_\alpha \mathrm{x}_k \mathrm{x}_l \,|\boldsymbol{\Psi}|^{\frac{1}{2}} exp\left(\tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} + \mathbf{u}'\mathbf{z}_\alpha\right)}{\sum_{\alpha=1}^{\mathrm{q}^*} \mathrm{w}_\alpha \,|\boldsymbol{\Psi}|^{\frac{1}{2}} exp\left(\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} + \mathbf{u}'\mathbf{z}_\alpha\right)}\,. \tag{6.3.22}$$

Case 3 (k = 1,2,3 ; l = 1,2,$\cdots$,6):

$$\mathrm{E}(\mathrm{x}_k \mathrm{x}_{l+3} \,|\, \mathbf{y}) = \frac{\sum_{\alpha=1}^{\mathrm{q}^*} \mathrm{w}_\alpha \mathrm{C}_k \mathrm{x}_l \,|\boldsymbol{\Psi}|^{\frac{1}{2}} exp\left(\tfrac{1}{2}\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} + \mathbf{u}'\mathbf{z}_\alpha\right)}{\sum_{\alpha=1}^{\mathrm{q}^*} \mathrm{w}_\alpha \,|\boldsymbol{\Psi}|^{\frac{1}{2}} exp\left(\mathbf{s}'\boldsymbol{\Psi}\mathbf{s} + \mathbf{u}'\mathbf{z}_\alpha\right)}\,. \tag{6.3.23}$$

Using equations (6.3.20), (6.3.22) and (6.3.23), the conditional moments $E(\mathbf{x}|\mathbf{y})$ and $Cov(\mathbf{x},\mathbf{x}'|\mathbf{y})$ can be evaluated.

Maximum likelihood estimators of the unknown parameters $\tau=(\alpha,\beta,\sigma^2)$ and $\beta=(\boldsymbol{\theta},\text{vecs}\boldsymbol{\Phi})$ are obtained by maximizing the marginal likelihood thus

$$\frac{\partial ln\,L}{\partial\beta_k}=\sum_{i=1}^{N} E_{\mathbf{x}|\mathbf{y}_i}\left(\frac{\partial ln\,g(\mathbf{x})}{\partial\gamma_k}\right), \qquad (6.3.24)$$

$$\frac{\partial ln\,L}{\partial\tau_l}=\sum_{i=1}^{N} E_{\mathbf{x}|\mathbf{y}_i}\left(\frac{\partial ln\,f(\mathbf{y}_i|\mathbf{x})}{\partial\tau_l}\right). \qquad (6.3.25)$$

Remarks

1) Since (cf. Proposition 6.2.5), $\boldsymbol{A}^{-1}$ is tridiagonal, we solve for $\mathbf{y}^*$ and $\boldsymbol{F}^*$ in order n operations.

3) Expressions for the derivatives of the log-likelihood function with respect to the *ARMA* parameters are as given in Section 6.2, but preceded by the $E_{\mathbf{x}|\mathbf{y}}$ operator.

6.4. PRACTICAL APPLICATIONS

Example 6.4.1 Analysis of a fixed parameter model of air pollution data with *ARMA*(1,1) deviations.

Air pollution data measured in the vicinity of a coal fired power station in the Transvaal in South Africa were analysed using *ARMA* deviations from a fixed parameter curve. The dependent variable is the amount of air pollution (in parts per billion) measured. In Figure 6.4.1 a graphical illustration of the raw data, recorded over a period of six years, is given. The x-axis denotes the months when the measurements were recorded. Although the period is 70 months there are only 57 observations, resulting in 13 missing observations (see Appendix A).

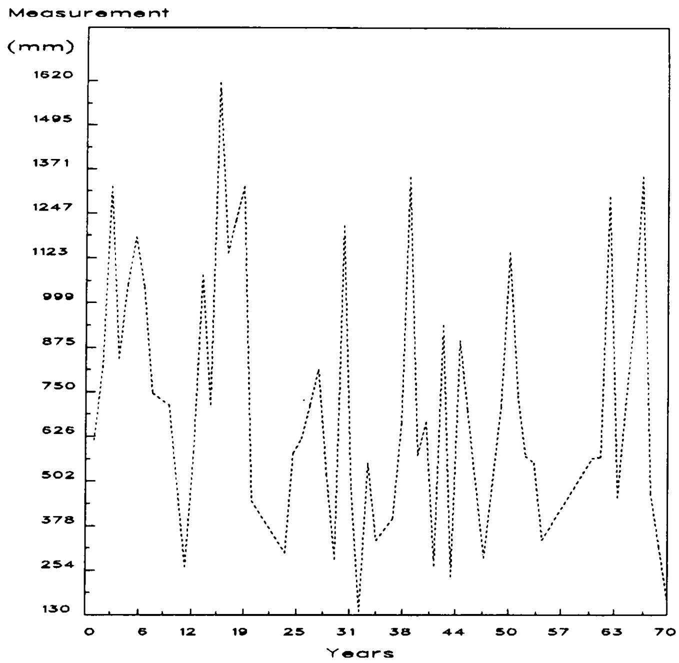

Figure 6.4.1 Raw data of air pollution recorded over a period of six years.

The model that was fitted to the data consists of a trend function, a seasonal function and an error term and is as follows:

$$y_{t_i} = a(1 + b\, r^{t_i}) + K \sin(2\pi \tfrac{1}{P} t_i + \theta) + \epsilon_{t_i} \qquad i = 1,2,\cdots,n \qquad (6.4.1)$$

where $t_1 < t_2 \cdots < t_n$ are integers that represent the n observation times, which are not consecutive or equally spaced. Under the assumption of a non-consecutive $ARMA(1,1)$ model the estimated model for the error terms is

$$\epsilon_t - 0.84\epsilon_{t-1} = u_t - 0.97u_{t-1} \text{ and } \quad \hat{\sigma} = 278.4 \ .$$

It can easily be shown that a straight line fit to the trend results in forecasted air pollution values which eventually become negative and therefore become unrealistic. A parabola fitted to the data results in high positive future predicted values which are also not realistic. It was found that the trend is best described by the nonlinear function

$$f_{trend} = a(1 + br^t) \ , \qquad (6.4.2)$$

the function with the parameter estimates is given and illustrated in Figure 6.4.2.

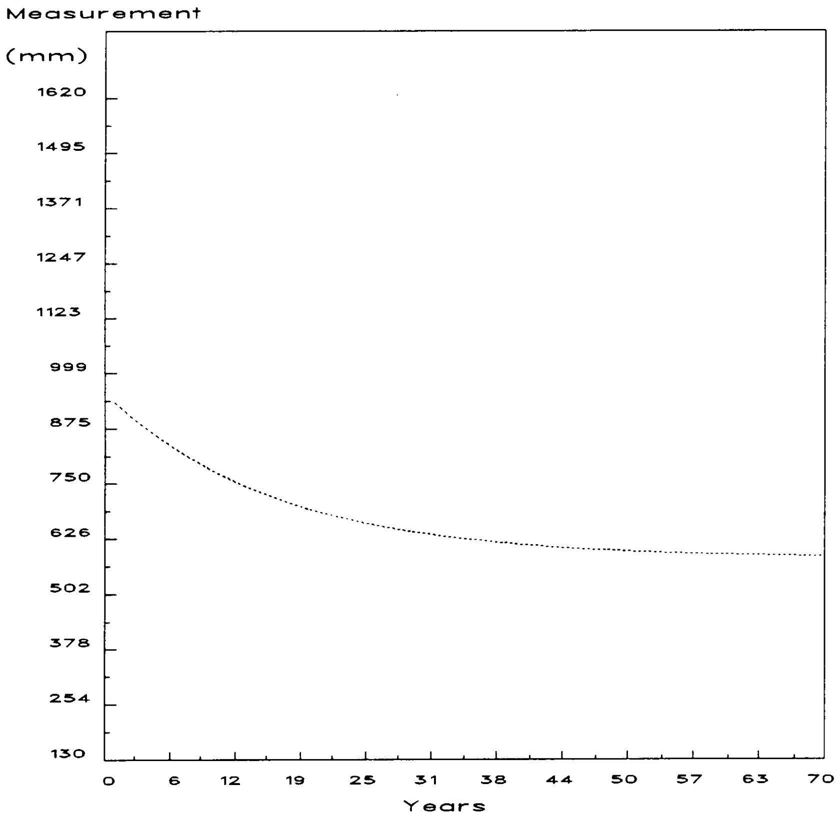

Figure 6.4.2 The nonlinear function describing the trend: $\hat{f}_{\text{trend}} = 580.6(1 + 0.65 \text{ X } 0.94^t)$.

Looking at Figure 6.4.1 it seems that the seasonal effect is decreasing. The amplitude of the cyclical movement is thus changing and decreasing over time. The seasonal function that we fitted is,

$$f_{\text{season}} = \text{Ksin}(2\pi \cdot \frac{1}{\text{period}} t + \theta) \ , \qquad (6.4.3)$$

where

$$\text{K} = \text{c}(1 + \text{de}^t).$$

The parameter estimates obtained for the seasonal function as well as a graphical illustration is given in Figure 6.4.3.

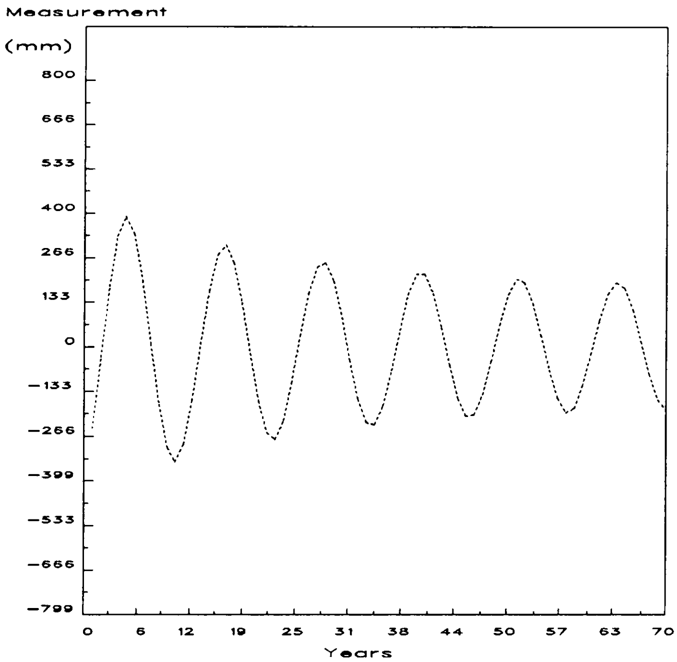

Figure 6.4.3 The fitted seasonal function is $\hat{f}_{season} = \hat{K}\sin(2\pi \cdot \frac{1}{11.82}t + -1.14)$
with $\hat{K} = 173.2(1 + 1.52 \text{ X } 0.96^{t})$

Three reasons why this dataset could not be analysed with the SAS ARIMA procedure using the Box and Jenkins (see Box and Jenkins (1976)) methodology are:

(i) The time points at which measurements were made are not consecutive and equally spaced.

(ii) With the Box and Jenkins methodology polynomial trends of degree k are eliminated using k-th order differencing. The trend seems to follow the pattern of an asymptotic growth curve rather than that of a polynomial.

(iii) The seasonal model of Box and Jenkins does not provide for a trend on the amplitude of the cyclical movement. The amplitude of the seasonal model for this data set decreases in time.

In Figure 6.4.4 a graphical illustration is given of the raw data (dotted line) and the fitted data (straight line).

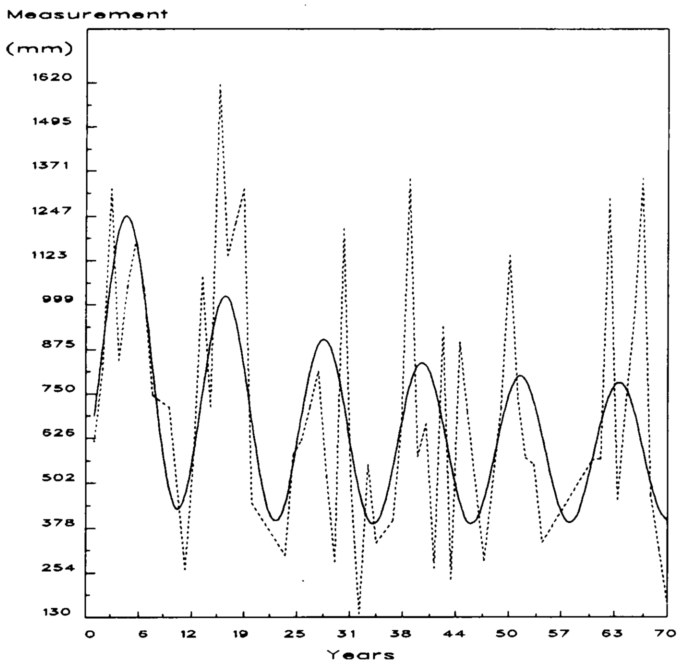

Figure 6.4.4 Raw data as well as fitted model.

Example 6.4.2 Analysis of a Gompertz parameter model with $ARMA$(1,1) deviations.

Using the data set described in Example 3.8.1 of Chapter 3 (see also Du Toit 1979) an illustration of an extension to a 2 group fixed parameter model (cf. Section 3.2 of Chapter 3) is given. The two group model (groups defined as male and female), allows for the number of repeated weight measurements on mice to differ across groups. A Gompertz function was chosen to describe the change in mean weights over time.

$$f(\boldsymbol{\theta}_j,t_{ij})= \theta_{j1}\, exp\, -\theta_{j2}\,\theta_{j3}^{t_{ij}}. \qquad j=1 \text{ males}, \qquad t_{i1}=1,2,\cdots,9$$
$$j=2 \quad \text{females} \quad t_{i2}=1,2,\cdots,8\,. \qquad (6.4.4)$$

A common $ARMA$(1,1) covariance structure with zero initial state variance was fitted. (see Chapter 6). In this example allowance was made for the white noise variances to differ across groups. The model is therefore

$$\mathbf{y}_{ij}= \mathbf{f}(\boldsymbol{\theta}_j,t_{ij})+\epsilon_{ij} \qquad i=1,2,\cdots,N_j\,;\ j=1,2. \qquad (6.4.5)$$

where $\epsilon_{ij} \sim N(\mathbf{0},\Lambda_j)$ and $\Lambda_j=\Lambda_j(\sigma_j^2,\alpha,\beta)$. In Figure 6.4.5 a graphical display of the raw data and the fitted means are given.

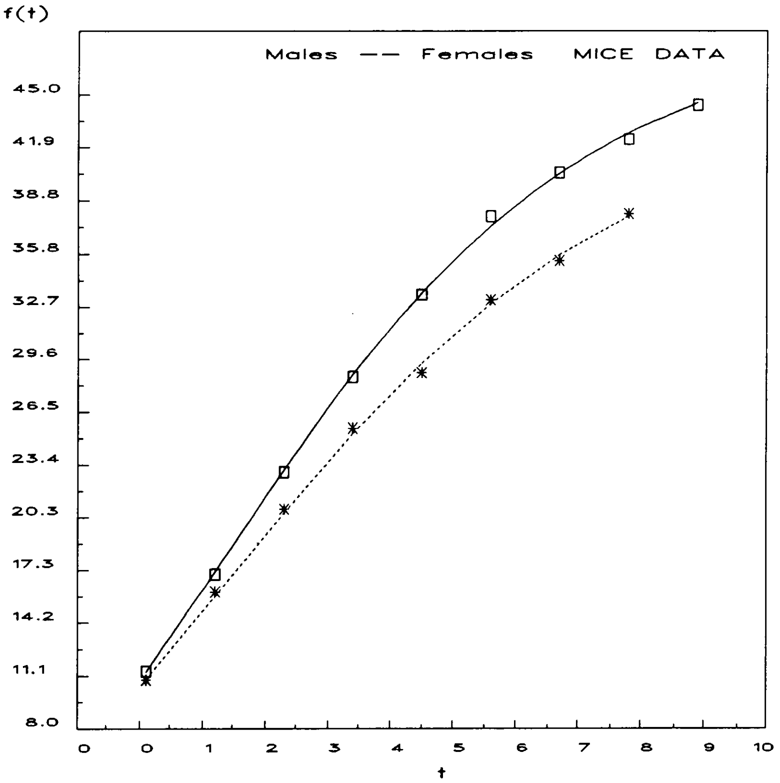

Figure 6.4.5 Mean weights of mice and fitted Gompertz functions.

*** Male mice ☐ Female mice**

A partial listing of the computer output is given below.

MALE GROUP (N_1=42)

1) SAMPLE MEANS :

T1	T2	T3	T4	T5	T6	T7	T8	T9
11.333	17.000	23.000	28.524	33.333	37.905	40.476	42.405	44.429

2) REPRODUCED MEANS :

T1	T2	T3	T4	T5	T6	T7	T8	T9
11.309	17.152	23.126	28.658	33.425	37.327	40.405	42.768	44.549

3) SAMPLE COVARIANCE MATRIX :

	T1	T2	T3	T4	T5	T6	T7	T8	T9
T1	7.032								
T2	7.571	16.952							
T3	5.976	13.405	16.000						
T4	4.944	13.190	16.881	25.011					
T5	5.389	16.881	21.119	30.825	44.413				
T6	1.865	11.000	16.476	26.693	40.722	46.372			
T7	0.746	8.071	12.571	21.155	36.056	41.450	47.107		
T8	-0.563	6.333	9.762	18.669	35.079	42.800	51.188	62.717	
T9	-0.762	4.571	8.381	15.871	31.905	40.588	50.224	62.446	66.959

3) REPRODUCED COVARIANCE MATRIX:

	T1	T2	T3	T4	T5	T6	T7	T8	T9
T1	7.780								
T2	7.835	15.670							
T3	7.767	15.656	23.424						
T4	7.700	15.520	23.342	31.043					
T5	7.633	15.386	23.140	30.896	38.532				
T6	7.567	15.253	22.940	30.628	38.319	45.891			
T7	7.501	15.120	22.741	30.363	37.987	45.615	53.123		
T8	7.436	14.989	22.544	30.100	37.658	45.220	52.784	60.230	
T9	7.372	14.860	22.349	29.839	37.332	44.828	52.327	59.830	67.215

$\hat{\sigma}_1^2$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\theta}_{13}$
7.78045	-0.99134	0.01566	49.41309	2.05512	0.71754

FEMALE GROUP (N_2=40)

1) SAMPLE MEANS :

T1	T2	T3	T4	T5	T6	T7	T8
10.850	15.975	20.825	25.550	28.775	33.025	35.325	38.075

2) REPRODUCED MEANS :

T1	T2	T3	T4	T5	T6	T7	T8
10.880	15.698	20.614	25.240	29.337	32.807	35.648	37.917

3) SAMPLE COVARIANCE MATRIX :

	T1	T2	T3	T4	T5	T6	T7	T8
T1	6.027							
T2	6.296	12.974						
T3	5.699	12.296	17.144					
T4	4.732	10.689	16.571	21.747				
T5	6.016	12.119	20.311	23.574	32.174			
T6	7.204	12.476	19.979	23.686	31.981	36.974		
T7	6.474	10.433	18.457	23.521	31.773	36.067	38.969	
T8	5.736	8.877	16.888	24.109	31.117	36.323	39.651	45.219

4) REPRODUCED COVARIANCE MATRIX :

	T1	T2	T3	T4	T5	T6	T7	T8
T1	5.808							
T2	5.849	11.698						
T3	5.798	11.687	17.486					
T4	5.748	11.586	17.425	23.174				
T5	5.698	11.486	17.274	23.064	28.764			
T6	5.649	11.386	17.125	22.864	28.606	34.257		
T7	5.600	11.288	16.976	22.666	28.358	34.052	39.656	
T8	5.551	11.190	16.829	22.470	28.112	33.757	39.404	44.962

$\hat{\sigma}_2^2$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$	$\hat{\theta}_{23}$
5.80813	-0.99134	0.01566	45.32695	1.92040	0.74307

$\chi^2 = 135.0$ Degrees of freedom=(9+45+8+36) − 10=88

Total number of observations: 82

Example 6.4.3 Analysis of a triple logistic fixed parameter model with *ARMA*(1,1) deviations.

Consider the following fixed parameter growth curve model

$$y_t = \frac{\theta_1}{1+e^{-\theta_2 t}} + \frac{\theta_3}{1+e^{\theta_4-\theta_5 t}} + \frac{\theta_6}{e^{\theta_7-\theta_8 t}} + \epsilon_t, \qquad (6.4.5)$$

where ϵ_t is generated by an $ARMA(1,1)$ process.

We have further assumed that measurements were taken at intervals of length 0.125 years. This assumption implies that the growth measurements were made at 2.000, 2.125, 2.250, 2.375, 2.500, $\cdots$, 24.875 years. The occasions on which growth measurements were made for each individual are subsets of the 184 timepoints given above. On average, there are approximately 40 measurements per individual, so that the timepoints of an individual corresponds to roughly 20% of the possible times at which height measurements may have been taken.

Part of the computer output is given below. Note that the unknown parameters in (6.4.5) are $\boldsymbol{\theta}$ (8x1), σ^2, the AR parameter α and the MA parameter β, therefore a total of 11 parameters.

Computer Ouput: Male data (N=158)

(i) Estimates of unknown parameters

$\hat{\sigma}^2 = 0.40549$

$\hat{\alpha} = 0.99185$

$\hat{\beta} = 0.14934$

$\hat{\theta}_1 = 37.93766$

$\hat{\theta}_2 = 0.77303$

$\hat{\theta}_3 = 115.03313$

$\hat{\theta}_4 = 0.83153$

$\hat{\theta}_5 = 0.37227$

$\hat{\theta}_6 = 28.99544$

$$\hat{\theta}_7 = 11.03194$$
$$\hat{\theta}_8 = 0.81739$$

(ii) Value of $-2 \cdot$log-likelihood function at convergence
F= 11226.283, number of parameters are 11

Computer Ouput: Female data (N=132)

(i) Estimates of unknown parameters

$$\hat{\sigma}^2 = 0.37266$$
$$\hat{\alpha} = 0.99304$$
$$\hat{\beta} = 0.16898$$
$$\hat{\theta}_1 = 42.12805$$
$$\hat{\theta}_2 = 0.27303$$
$$\hat{\theta}_3 = 98.28330$$
$$\hat{\theta}_4 = 0.85954$$
$$\hat{\theta}_5 = 0.37227$$
$$\hat{\theta}_6 = 26.67480$$
$$\hat{\theta}_7 = 9.41146$$
$$\hat{\theta}_8 = 0.82461$$

(ii) Value of $-2 \cdot$log-likelihood function at convergence
F= 14027.201, number of parameters are 11

Example 6.4.4 Analysis of a single logistic stochastic parameter model $AR(1)$ deviations.

The mice data set consisting of 8 weight measurements on each of 40 female mice, and 9 weight measurements on each of 42 male mice (cf. Section 4.7 and 7.5) is analysed. Sex is used as a covariate and it is assumed that the errors are generated by a first order autoregressive process.

The model fitted to the data is as follows

$$\mathbf{y}_i = \mathrm{f}(\mathbf{x},\boldsymbol{t}_i) + \epsilon_i \qquad i=1,2,\cdots,\mathrm{N} \tag{6.4.6}$$

It is assumed that the errors are generated by a first order autoregressive process with parameter α and white noise variance σ_a^2, that is, for any element ϵ_{ij} of ϵ_i,

$$\epsilon_{ij} - \epsilon_{i,j-1} = a_{ij} \tag{6.4.7}$$

where $a_{1j}, \cdots a_{in_i}$ are uncorrelated with mean 0 and variance σ_a^2. Hence $\boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0},\boldsymbol{\Lambda}_i)$ where $\boldsymbol{\Lambda}_i = \boldsymbol{\Lambda}_i(\tau)$ and $\tau = (\sigma_a^2 , \alpha)'$.

The response function $\mathbf{f}(\mathbf{x},\boldsymbol{t}_i)$ has typical element

$$\mathrm{f}(\mathbf{x},t_{ij}) = \frac{\mathrm{x}_1}{1+\exp(\mathrm{x}_2)(1+\exp(\mathrm{x}_3))^{t_{ij}}} \tag{6.4.8}$$

where

$$\mathbf{x} = \boldsymbol{\theta} + \gamma \mathrm{z} + \mathbf{u} .$$

It is assumed that $\mathbf{u}$ has a normal distribution with zero mean and covariance matrix $\boldsymbol{\Phi}$. The covariate z assumes the values 1 and -1 for males and females respectively.

Estimates of the unknown parameters are as follows:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} 45.068 \\ 1.578 \\ -0.411 \end{bmatrix}, \quad \hat{\gamma} = \begin{bmatrix} 2.263 \\ 0.090 \\ 0.090 \end{bmatrix}, \quad \hat{\boldsymbol{\Phi}} = \begin{bmatrix} 83.356 & & \\ 2.123 & 0.082 & \\ -0.873 & -0.018 & 0.077 \end{bmatrix},$$

$\hat{\sigma}_a^2 = 3.875$ and $\hat{\alpha} = 0.466$.

Example 6.4.5 Analysis of a triple logistic stochastic parameter model with $ARMA(1,1)$ deviations.

Consider the random parameter growth curve model (cf. 6.4.6)

$$y_t = \frac{x_1}{1+e^{-x_2 t}} + \frac{x_3}{1+e^{x_4 - x_5 t}} + \frac{x_6}{e^{x_7 - x_8 t}} + \epsilon_t, \qquad (6.4.9)$$

where it is assumed that the 8-dimensional vector $\mathbf{x}$ of stochastic parameters is distributed $N(\boldsymbol{\theta},\boldsymbol{\Phi})$ and that the ϵ_t's are generated by an $ARMA(1,1)$ process with parameters σ^2, α and β. We use the same spacing between time points as in Example 6.4.3.

Note that in contrast with the 45 parameter model considered in Example 5.5.4 of Chapter 5, the total number of unknown parameters assuming $ARMA(1,1)$ deviations becomes 47. Part of the computer output is given below.

Computer Output: Male data

(i) Estimates of unknown parameters

Estimate of $\boldsymbol{\theta}$

1	94.40887
2	0.84438
3	64.54539
4	2.83259
5	0.38581
6	23.34824
7	17.42463
8	1.25866

Estimate of $\boldsymbol{\Phi}$

1	19.99671		
2	-0.32962	0.01169	
3	2.29715	-0.03787	11.99040

4	-0.03873	0.00363	0.00607	0.00229				
5	-0.04147	0.00222	-0.02996	0.00079	0.00084			
6	-3.29680	0.05434	-5.37246	0.00190	0.01756	29.30710		
7	-0.15748	-0.01858	0.69018	-0.00918	-0.00639	-0.00583	0.26684	
8	-0.06057	0.00240	0.05358	0.00084	0.00034	-0.03028	-0.00493	0.00553

Estimate of σ^2	Estimate of α	Estimate of β
0.38336	0.87348	0.36817

(ii) Estimates of parameters given $\mathbf{y}_i$ (height measurements), Cases 1-2

Case number: 1 $\hat{\boldsymbol{\theta}}|\mathbf{y}_i$

97.964	0.75796	65.586	2.8116	0.37818	20.530	17.502	1.2215

Case number: 2 $\hat{\boldsymbol{\theta}}|\mathbf{y}_i$

90.963	0.94033	64.643	2.8627	0.38415	27.250	17.488	1.2212

(iii) Estimates of covariance matrix given $\mathbf{y}_i$ (height measurements), Cases 1-2

Case number: 1 $\hat{\boldsymbol{\Phi}}|\mathbf{y}_i$

3.4027	-0.91331E-01	0.31779E-02	-1.6779	0.33407E-01
3.5691	-0.16484E-01	0.95865E-03	0.25171E-02	0.11219E-02
-0.19931E-01	0.64059E-03	-0.43161E-02	0.23184E-03	0.21613E-03
-1.4437	0.48906E-01	-1.7910	0.11154E-01	0.21019E-01
3.1389	-0.84865E-01	-0.50709E-03	0.34582	-0.10487E-02
-0.98624E-03	-0.23181	0.15451	-0.49354E-02	-0.30969E-04
0.20077E-01	-0.62947E-04	-0.57889E-04	-0.13543E-01	0.90587E-02
0.53866E-03				

Case number: 2 $\hat{\boldsymbol{\Phi}}|\mathbf{y}_i$

1.8976	-0.66342E-01	0.36371E-02	-0.17640	-0.17137E-01
6.0339	-0.12481E-01	0.12265E-02	-0.72778E-02	0.12547E-02
-0.16319E-01	0.77790E-03	-0.28178E-01	0.28724E-03	0.32250E-03
-1.4763	0.72172E-01	-5.4276	0.16137E-01	0.39706E-01
6.5628	-0.10316E-01	-0.26154E-02	0.35216	-0.13540E-02

-0.17476E-02	-0.31281	0.11216	-0.93416E-04	-0.18233E-03
0.22279E-01	-0.91739E-04	-0.11721E-03	-0.20779E-01	0.69259E-02
0.46449E-03				

(iv) Value of $-2 \cdot$log-likelihood function at convergence
F = 10348.8, number of parameters are 47.

Computer Output: Female data

(i) Estimates of unknown parameters

Estimate of $\boldsymbol{\theta}$

1	84.52477
2	1.10812
3	61.24890
4	2.45085
5	0.42264
6	21.14966
7	14.12992
8	1.21111

Estimate of $\boldsymbol{\Phi}$

1	10.67189							
2	-0.20115	0.01291						
3	4.35935	-0.08217	7.56808					
4	-0.11417	0.00927	-0.03797	0.00736				
5	-0.02801	0.00214	-0.01708	0.00175	0.00088			
6	-2.33425	0.04399	-1.15755	0.02466	0.00632	9.49455		
7	-0.52424	-0.01163	-0.59830	-0.01176	-0.00206	-1.22454	3.05061	
8	0.02168	-0.00078	-0.07384	-0.00065	-0.00004	-0.19876	0.29747	0.03638

Estimate of σ^2	Estimate of α	Estimate of β
1.21111	0.42264	0.36280

(ii) Estimates of parameters given $\mathbf{y}_i$ (height measurements), Cases 1-2

Case number: 1 $\hat{\theta}|\mathbf{y}_i$

85.298	1.1439	59.889	2.4664	0.40795	18.925	13.447	1.2084

Case number: 2 $\hat{\theta}|\mathbf{y}_i$

87.239	0.88753	64.694	2.3126	0.43493	21.012	13.240	1.2607

(iii) Estimates of covariance matrix given $\mathbf{y}_i$ (height measurements), Cases 1-2

Case number: 1 $\hat{\Phi}|\mathbf{y}_i$

0.33451	-0.12399E-01	0.37488E-02	-0.13954	-0.39983E-01
2.6252	-0.51453E-02	0.26930E-02	-0.35306E-01	0.24214E-02
-0.31092E-02	0.62935E-03	-0.17768E-01	0.53084E-03	0.20449E-03
-0.14373	0.43769E-01	-2.2358	0.33511E-01	0.17905E-01
2.2040	0.19634	-0.54258E-01	0.92650	-0.39556E-01
-0.11385E-01	-1.0420	1.5392	0.16517E-01	-0.45009E-02
0.81152E-01	-0.32889E-02	-0.97322E-03	-0.91528E-01	0.13707
0.12269E-01				

Case number: 2 $\hat{\Phi}|\mathbf{y}_i$

0.61411	-0.84829E-02	0.73529E-03	-0.16152	0.63299E-02
3.1421	-0.32483E-03	0.56910E-03	0.70144E-02	0.91707E-03
-0.60882E-02	0.14718E-03	-0.16841E-01	0.11360E-03	0.20661E-03
-0.35948	0.20466E-02	-2.7870	-0.63722E-02	0.20228E-01
3.0130	-0.40990E-01	0.35600E-02	1.0707	0.55700E-02
-0.40480E-02	-1.1387	1.6800	-0.21931E-02	0.19849E-03
0.10096	0.42266E-03	-0.41540E-03	-0.10918	0.15981
0.15233E-01				

(iv) Value of $-2\cdot$log-likelihood function at convergence

F=13010.92, number of parameters are 47.

Testing of Hypotheses.

Consider the following restricted alternative hypothesis:

H_a: The triple-logistic stochastic parameter model with $ARMA(1,1)$ deviations is an adequate model for the description of height from birth to maturity (Example 6.4.5).

Further, consider the following null hypotheses.

H_{O1}: The triple-logistic fixed parameter model with $ARMA(1,1)$ deviations is an adequate model for the description of height (Example 6.4.3).

H_{O2}: The triple-logistic stochastic parameter model with uncorrelated errors is an adequate model for the description of height (Example 5.5.4).

Asymptotic test statistics.

Let l_a denote $-2 \cdot$log-likelihood function under H_a and l_O the corresponding value under H_O.

Then $l_O - l_a$ has an asymptotic χ^2-distribution with degrees of freedom (d.f.f.) equal to the number of parameters estimated under H_a – the number estimated under H_O.

For the male dataset (cf. Sections 5, 10 and 12) we have

$\chi_1^2 = l_{O1} - l_a \;=\; 11226.283 - 10348.8 = 877.5$ with d.f.f. $(47-11) = 36$
and
$\chi_2^2 = l_{O2} - l_a \;=\; 10694.4 - 10348.8 = 345.6$ with d.f.f. $(47-45) = 2$

The corresponding values for the female dataset is

$\chi_1^2 = l_{O1} - l_a = 14027.201 - 13010 = 1016.3$ with d.f.f. $(47-11) = 36$
and
$\chi_2^2 = l_{O2} - l_a = 13178.6 - 13010 = 167.68$ with d.f.f. $(47-45) = 2$.

The test statistic values are all highly significant and indicate that the stochastic parameter model with time series deviations is to be preferred to the stochastic parameter model without time series deviations or the fixed parmeter model with time series deviations.

The theoretical results, as implemented in the computer programs, were tested and verified with the use of a large number of simulation studies. In the simulation studies it became evident that one requires at least 6 Gauss-quadrature points per integral to obtain accurate estimates of the unknown parameters. However, good initial approximations can be obtained by using as few as 3 Gauss-quadrature points per integral.

6.5 ADDENDUM – DERIVATIVES OF *ARMA*(1,1) MODEL.

From Theorem 6.2.1 it follows that

$$\boldsymbol{A} = \boldsymbol{I}_{\mathrm{n},1}\mathrm{p}^*\boldsymbol{I}'_{\mathrm{n},1} + \boldsymbol{T}^*_\beta \boldsymbol{T}^{*\,\prime}_\beta, \tag{6.5.1}$$

where $\boldsymbol{A}$ is a symmetric (nxn) tridiagonal matrix of the form:

$$\boldsymbol{A} = \begin{bmatrix} a_1 & b_2 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ b_2 & a_2 & b_3 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & b_3 & a_3 & b_4 & \cdot & \cdot & 0 & 0 \\ 0 & 0 & b_4 & a_4 & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & \cdot & a_{\mathrm{n}-1} & b_{\mathrm{n}} \\ 0 & 0 & 0 & 0 & \cdot & \cdot & b_{\mathrm{n}} & a_{\mathrm{n}} \end{bmatrix} \tag{6.5.2}$$

The elements of $\boldsymbol{A}$ are defined as follows:

$$a_1 = 1 + \mathrm{p}^* = 1 + \frac{(\alpha-\beta)^2}{1-\alpha^2},$$

and for $i=2,3,\cdots,$n:

$$a_i = 1+\beta^2\alpha^{2(\mathrm{m}_i-1)} + \frac{\eta^2}{(1-\alpha^2)}(1-\alpha^{2(\mathrm{m}_i-1)}) \tag{6.5.3}$$

$$b_i = -\beta\alpha^{\mathrm{m}_i-1} \tag{6.5.4}$$

Therefore for p* defined in (6.2.25) it follows that

$$\frac{\partial a_1}{\partial\alpha} = \frac{\partial \mathrm{p}^*}{\partial\alpha} = \frac{2(\alpha\mathrm{p}^* - \beta)}{1-\alpha^2} \tag{6.5.5}$$

$$\frac{\partial a_1}{\partial\beta} = \frac{\partial \mathrm{p}^*}{\partial\beta} = \frac{-2(\alpha - \beta)}{1-\alpha^2} \tag{6.5.6}$$

$$\frac{\partial a_i}{\partial\alpha} = \frac{\partial a_1}{\partial\alpha} + 2(\mathrm{m}_i-1)\alpha^{2\mathrm{m}_i-3}(\beta^2-\mathrm{p}^*) - \alpha^{2(\mathrm{m}_i-1)}\frac{\partial \mathrm{p}^*}{\partial\alpha} \tag{6.5.7}$$

$$\frac{\partial a_i}{\partial\beta} = \frac{\partial a_1}{\partial\beta} + \alpha^{2(\mathrm{m}_i-1)}\left(2\beta - \frac{\partial \mathrm{p}^*}{\partial\alpha}\right) \tag{6.5.8}$$

$$\frac{\partial b_i}{\partial\alpha} = -\beta(\mathrm{m}_i-1)\alpha^{(\mathrm{m}_i-2)} \tag{6.5.9}$$

$$\frac{\partial b_i}{\partial\beta} = -\alpha^{(\mathrm{m}_i-1)} \tag{6.5.10}$$

For $\boldsymbol{T}_\alpha^*$ defined in (6.2.16) and **u** defined in (6.2.41) it follows that

$$\frac{\partial\mathbf{u}}{\partial\alpha} = \frac{\partial\boldsymbol{T}_\alpha^*}{\partial\alpha}\tilde{\mathbf{y}} ;$$

where

$$\left[\frac{\partial\boldsymbol{T}_\alpha^*}{\partial\alpha}\right]_{ij} = \begin{cases} -\mathrm{m}_i\alpha^{\mathrm{m}_i-1}, & i-j=1 \\ 0, & \text{otherwise} \end{cases} \tag{6.5.11}$$

$\frac{\partial\zeta_i}{\partial\gamma}$ is obtained recursively. From (6.2.34) it follows that for example,

$$\gamma_2 = \frac{b_2}{a_1},$$

hence

$$\frac{\partial \gamma_2}{\partial \alpha} = \frac{\partial b_2}{\partial \alpha} a_1^{-1} - b_2 \cdot a_1^{-2} \cdot \frac{\partial a_1}{\partial \alpha} ,$$

and

$$\zeta_2 = a_2 - b_2 \gamma_2 .$$

Therefore

$$\frac{\partial \zeta_2}{\partial \alpha} = \frac{\partial a_2}{\partial \alpha} - \left(\frac{\partial b_2}{\partial \alpha} \gamma_2 + b_2 \cdot \frac{\partial \gamma_2}{\partial \alpha} \right), \text{ and so on.}$$

From (6.2.49) and (6.5.1) it further follows that

$$\frac{\partial^2 ln\, \mathrm{L}(\boldsymbol{\theta},\boldsymbol{\Lambda}^*)}{\partial \gamma_r \partial \gamma_s} = -\tfrac{1}{2} \sum_{i=1}^{\mathrm{N}} \left(\frac{\partial^2}{\partial \gamma_r \partial \gamma_s} (\mathbf{u}_i' \mathbf{y}_i^* / \sigma^2) + \sum_{j=1}^{\mathrm{n}_i} \left(\frac{\partial^2 \zeta_{ij}}{\partial \gamma_r \partial \gamma_s} \cdot \zeta_{ij}^{-1} - \zeta_{ij}^{-2} \frac{\partial \zeta_{ij}}{\partial \gamma_r} \right) \right) . \qquad (6.5.12)$$

From (6.5.4) we have

$$\frac{\partial}{\partial \gamma_s} \mathbf{u}' \mathbf{y}^* = 2 \left(\frac{\partial \boldsymbol{T}_\alpha^*}{\partial \gamma_s} \tilde{\mathbf{y}}\, ' \right) \mathbf{y}^* - \mathbf{y}^{*\prime} \left(\frac{\partial \boldsymbol{A}}{\partial \gamma_s} \right) \mathbf{y}^* . \qquad (6.5.13)$$

Hence

$$\frac{\partial^2}{\partial \gamma_r \partial \gamma_s} [\mathbf{u}' \mathbf{y}^*] = 2 \left(\frac{\partial^2 \boldsymbol{T}_\alpha^*}{\partial \gamma_r \partial \gamma_s} \tilde{\mathbf{y}}\, ' \right) \mathbf{y}^* + 2 \left(\frac{\partial \boldsymbol{T}_\alpha^*}{\partial \gamma_s} \tilde{\mathbf{y}}\, ' \right) \frac{\partial \mathbf{y}^*}{\partial \gamma_r} - 2 \mathbf{y}^{*\prime} \left(\frac{\partial \boldsymbol{A}}{\partial \gamma_s} \right) \frac{\partial \mathbf{y}^*}{\partial \gamma_r}$$
$$+ \mathbf{y}^{*\prime} \left(\frac{\partial^2 \boldsymbol{A}}{\partial \gamma_r \gamma_s} \right) \mathbf{y}^* . \qquad (6.5.14)$$

Since (cf. (6.2.42) $\mathbf{y}^* = \boldsymbol{A}^{-1} \mathbf{u}$, it follows that

$$\frac{\partial \mathbf{y}^*}{\partial \gamma_r} = -\boldsymbol{A}^{-1} \left(\frac{\partial \boldsymbol{A}}{\partial \gamma_r} \right) \boldsymbol{A}^{-1} \mathbf{u} + \boldsymbol{A}^{-1} \left(\frac{\partial \mathbf{u}}{\partial \gamma_r} \right) , \qquad (6.5.15)$$

with

$$\frac{\partial \mathbf{u}}{\partial \gamma_r} = \left(\frac{\partial \boldsymbol{T}_\alpha^*}{\partial \gamma_r} \right) \tilde{\mathbf{y}} .$$

Typical elements of $\dfrac{\partial^2 \boldsymbol{A}}{\partial\gamma_r\gamma_s}$ may, for example, be obtained as follows:

$$\frac{\partial^2 a_1}{\partial\alpha^2} = \frac{\partial}{\partial\alpha}\left(\frac{2\alpha p^* - \beta}{1-\alpha^2}\right) = \frac{2p^* + 2\alpha\dfrac{\partial p^*}{\partial\alpha}}{1-\alpha^2} + \frac{2(2\alpha p^* - \beta)\alpha}{(1-\alpha^2)^2} \qquad (6.5.16)$$

$$\frac{\partial^2 a_1}{\partial\beta^2} = \frac{\partial}{\partial\beta}\left(\frac{-2(\alpha - \beta)}{1-\alpha^2}\right) = \frac{2}{1-\alpha^2} \qquad (6.5.17)$$

$$\frac{\partial^2 a_1}{\partial\beta\partial\alpha} = \frac{\partial}{\partial\alpha}\left(\frac{-2(\alpha-\beta)}{1-\alpha^2}\right) = \frac{4\alpha(\beta-\alpha) - 2(1-\alpha^2)}{1-\alpha^2}\ . \qquad (6.5.18)$$

For $i=2,3,\cdots,n$:

$$\frac{\partial^2 a_i}{\partial\alpha^2} = \frac{\partial}{\partial\alpha}\left(\frac{\partial a_1}{\partial\alpha} + 2(m_i-1)\alpha^{2m_i-3}(\beta^2 - p^*) - \alpha^{2(m_i-1)}\frac{\partial p^*}{\partial\alpha}\right)$$
$$= \left(\frac{\partial^2 a_1}{\partial\alpha^2} + 2(m_i-1)(2m_i-3)\alpha^{2(m_i-2)}(\beta^2 - p^*) - 4(m_i-1)\alpha^{2m_i-3}\frac{\partial p^*}{\partial\alpha} - \alpha^{2(m_i-1)}\frac{\partial^2 p^*}{\partial\alpha^2}\right). \qquad (6.5.19)$$

$$\frac{\partial^2 a_i}{\partial\beta^2} = \frac{\partial^2 a_1}{\partial\beta^2} + \alpha^{2(m_i-1)}\left(2 - \frac{\partial^2 p^*}{\partial\alpha\partial\beta}\right) \qquad (6.5.20)$$

$$\frac{\partial^2 a_i}{\partial\beta\partial\alpha} = \frac{\partial^2 a_1}{\partial\beta\partial\alpha} + 2(m_i-1)\alpha^{2m_i-3}\left(2\beta - \frac{\partial p^*}{\partial\alpha}\right) + \alpha^{2(m_i-1)}\left(\frac{\partial^2 p^*}{\partial\alpha^2}\right). \qquad (6.5.21)$$

$$\frac{\partial^2 b_i}{\partial\alpha^2} = -\beta(m_i-1)(m_i-2)\alpha^{(m_i-3)}. \qquad (6.5.22)$$

$$\frac{\partial^2 b_i}{\partial\beta^2} = 0\ . \qquad (6.5.23)$$

$$\frac{\partial^2 b_i}{\partial\alpha\partial\beta} = -(m_i-1)\alpha^{(m_i-2)}. \qquad (6.5.24)$$

Typical elements of $\dfrac{\partial^2 \boldsymbol{T}_\alpha}{\partial \gamma_r \partial \gamma_s}$ may, for example, be obtained as follows:

$$\frac{[\partial^2 \boldsymbol{T}^*_\alpha]_{ij}}{\partial \alpha^2} = \begin{cases} \mathrm{m}_i(\mathrm{m}_i - 1)\alpha^{\mathrm{m}_i - 2} \,, & i - j = 1 \,; \\ 0 \,, & \text{otherwise} \,. \end{cases} \tag{6.5.25}$$

$$\frac{[\partial^2 \boldsymbol{T}^*_\alpha]_{ij}}{\partial \alpha \partial \beta} = 0 \quad \text{and} \quad \frac{[\partial^2 \boldsymbol{T}^*_\alpha]_{ij}}{\partial \beta^2} = 0 \,.$$

6.6 SUMMARY.

In this chapter the analysis of multi component models with time series deviations is considered. The theoretical results derived are used to estimate parameters in an 8-parameter triple-logistic model. The latter model was used to describe human growth from 2 years to maturity of the Fels data. A similar analysis was done on the Berkeley growth study and is given in Appendix C.

In Section 6.2 a fixed parameter nonlinear model with $ARMA(1,1)$ deviations is discussed. The derivation of the likelihood functions of nonlinear random parameter regression models with non-consecutive $ARMA(1,1)$ deviations is discussed in Section 6.3.

Using the Fels growth data two possible models, namely the *random* parameter model with uncorrelated errors (cf. Example 5.5.4) and the *fixed* parameter model with $ARMA(1,1)$ deviations (cf. Example 6.4.3), were compared with the *random* parameter model with $ARMA(1,1)$ deviations (cf. Example 6.4.5).

CHAPTER 7 SUGGESTIONS FOR FURTHER RESEARCH.

It was not proved in Chapter 3 that the iterative reweighted GLS estimator $\hat{\gamma}_{GLS}$ that minimizes (3.4.23) is identical to $\hat{\gamma}$ that minimizes (3.4.25). This matter requires further research.

The dispersion matrix of the sample means and covariances, $\boldsymbol{\Omega}$, is defined differently for the normal theory generalised least squares (GLS), maximum Wishart likelihood (MWL) and asymptotically distribution free (ADF) estimators, available for use in the analysis of covariance structures, because the assumptions about the distribution of the data are different. Further attention has to be given to the elliptically contoured multivariate class of distributions which has not been considered here.

The computation of fourth order moments for the type of nonlinear random parameter models discussed in this thesis is currently very time consuming. It will be worthwhile to program these, once faster computers become affordable.

In Section 5.3 of Chapter 5 it was shown that in the case of random parameter models, the derivatives of the likelihood function may be expressed in terms of the moments of the posterior probability density function, for example

$$\mathrm{E}(\mathrm{x}_l|\mathbf{y}) = \frac{\int \mathrm{x}_l \, \mathrm{g}(\mathbf{x}) \, \mathrm{f}(\mathbf{y}|\mathbf{x}) d\mathbf{x}}{\int \mathrm{g}(\mathbf{x}) \, \mathrm{f}(\mathbf{y}|\mathbf{x}) d\mathbf{x}} \quad , \qquad l=1,2,\cdots,\mathrm{r} \qquad (7.1)$$

It was further shown (cf. (5.3.31) to (5.3.34)) that the integrals in (7.1) can be evaluated by means of a Gauss quadrature integration formula. Bock (personal communication to Du Toit, June 1993) proposed that Newton-Coates quadrature now substitute the Gauss-Hermite quadrature, and that the points $\mathbf{x}=(\mathrm{x}_1, \mathrm{x}_2,\cdots,\mathrm{x}_\mathrm{r})'$ of the integration formula be obtained from

$$\mathbf{x} = \overline{\boldsymbol{\theta}}_i + (\boldsymbol{D}_{\boldsymbol{\theta}_i}\boldsymbol{T})\mathbf{k} \qquad (7.2)$$

where

$$\mathbf{k} = \begin{bmatrix} \mathrm{k}_1 - 1 \\ \mathrm{k}_2 - 1 \\ \cdot \\ \cdot \\ \cdot \\ \mathrm{k}_r - 1 \end{bmatrix} \tag{7.3}$$

and where k_i, i=1,2,$\cdots$,r are the points of a 3^{m} fractional factorial design.

In the above notation $\boldsymbol{\theta}_i$ is a provisional estimate of $\mathbf{x}$ for subject i from a previous iteration and the elements of $\boldsymbol{D}_{\boldsymbol{\theta}_i}$ are the standard deviations of a previous estimate of Cov($\mathbf{x}$,$\mathbf{x}'|\mathbf{y}_i$). Note further that $\boldsymbol{TT}'=\boldsymbol{P}$, with $\boldsymbol{P}$ an estimate of the correlation matrix of $\mathbf{x}$. An advantage of the proposed method is that for the fractional factorial design, only 243 r-dimensional points are evaluated. The numerical integration procedure described above, is known as adaptive quadrature, since use is made of the Bayes estimates and conditional covariance matrices of each individual.

In Chapter 6 specific attention was given to the nonlinear fixed and random parameter models with $ARMA$(1,1) deviations. It may be interesting to consider higher order $ARMA$ models, for the case of non-consecutive data.

BIBLIOGRAPHY

ABRAMOWITZ, M. and STEGUN, I.A. (1965). *Handbook of mathematical functions.* New York: Dover Publications.

ANDERSON, T.W. and RUBIN, H. (1956). Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,* Vol.V. ed. J. Neyman, 115-150, Berkeley: University of California Press.

ANDERSON, T.W. (1958). *An introduction to Multivariate Statistical Analysis.* New York: John Wiley & Sons.

ANDERSON, T.W. (1975). Maximum likelihood estimation of parameters of autoregressive processes with moving average residuals and other covariance matrices with linear structure. In *The Annals of statistics:* 3,No. 6, 1283-1304.

BATES, D.M. and WATTS, D.G. (1988). *Nonlinear Regression Analysis and Its Applications.* New York: Wiley.

BERKEY, C.S. (1982). Bayesian approach for a nonlinear growth model. In Biometrics 38:953-961.

BOCK, R.D., WAINER, H., PETERSEN, A., THISSEN D., MURRAY, J. and ROCHE, A. (1973). A parameterization for individual human growth curves. In *Human Biology,* 45: 63-80.

BOCK, R.D. and THISSEN, D. (1976). Fitting multi-component models for growth in stature. Proceedings of the 9th International Biometric Conference, 431-442.

BOCK, R.D. and THISSEN, D. (1980). Statistical problems of fitting individual growth curves. In F.E. Johnson, A.F. Roche, & C. Susanne (Eds.), *Human physical growth and maturation: methodologies and factors,* 265-290. New York: Plenum.

BOCK, R.D. and AITKIN, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika,* 46: 443-445.

BOCK, R.D. (1986). Unusual patterns in the Fels data. In A. Demirjian (Ed.), *Human growth: a multidisciplinary review, 69-84.* London: Taylor & Francis.

BOCK, R.D. (1989). Prediction of growth. In L.M. Collins & J.L. Horne (Eds.), *Best methods for the analysis of change: recent advances, unawnsered questions, future directions.* Washington, D.C.: American Psychological Association.

BOCK, R.D. (1989). Measurement of human variation: a two-stage model. In R.D. Bock (AEd.), *Multilevel analysis of educational data,* 319-342. New York: Academic Press.

BOCK, R.D. (1990). *Marginal maximum likelihood estimation for the two-stage random regressions model with second stage covariates.* Unpublished manuscript. University of Chicago.

BOCK, R.D. and DU TOIT, S.H.C.(1991). Multi-component models for the prediction of human growth based on marginal maximum likelihood estimation. *Technical Report,* University of Chicago: Chicago

BOX, M.J., DAVIES, D., SWANN, W.H. (1969). Non linear optimization techniques. *I.C.I. Monograph* No. 5. New York: Oliver & Boyd.

BOX, G.E.P. and JENKINS, G.M. (1976). *Time series analysis: Forecasting and control.* Revised Edition. Oakland, California: Holden-Day, Inc.

BROWN, M.L. (1974). Identification of sources of significance in two-way tables. *Applied Statistics,* 23: 405-413.

BROWNE, M.W. (1974). Generalized least squares estimators in the analysis of covariance structures. In *South African Statistical Journal,* 8, No. 1: 1-24.

BROWNE, M.W. (1975). Supplementary notes STW (H)/2/75. Methods of multivariate analysis. *Department of Statistics & O.R.*, University of South Africa.

BROWNE, M.W. and GREENACRE, M (1976). Program for analysis of covariance structures. Department of Statistics & O.R., University of South Africa. Unpublished report.

BROWNE, M.W. (1982). Covariance structures. In D.M. Hawkins (ed.), *Topics in Applied Multivariate Analysis:* 72-141. Cambridge: Cambridge University Press.

BROWNE, M.W. and DU TOIT, S.H.C. (1987). Automated fitting of nonstandard models, *Report WS-39*, Pretoria: HSRC

BROWNE, M.W. (1991). *STA 401-Matrix methods in statistics.* Supplementary notes. University of South Africa, Pretoria.

BROWNE, M.W. and DU TOIT, S.H.C. (1991). Models for learning data. In L.M. Collins & J.L. Horne (Eds.) *Best methods for the analyses of change: recent advances, unawnsered questions, future directions,* Washington, D. C.: American Psychological Association.

BURT, C. (1937). *The backward child.* New York: Appleton-Century.

CARTER, W.H., Jr and MYERS, R.H. (1973). Maximum likelihood estimation from linear combinations of discrete probability functions. *Journal of American Statistical Association,* 68: 203-206.

CHEECK, D.B. (1968). A new look at growth. Chapter 1 in: *Human Growth. Body Composition, Cell Growth, Energy and Intelligence.* Lea and Febiger, Philadelphia.

CHEN, T. and FIENBERG, S. (1976). The analysis of contingency tables with incompletely classified data, *Biometrics*, 32:133-144.

CHI, E.M. and REINSEL, G.C. (1989). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association,* 84: 452-459.

CHILDS, R.A. MCFARLANE, M. POMMERICH, M.R. WIESEN, C.A. WILLIAMS V.S.L. (1992). Modest multilevel models and methods: some examples. *Research Memorandum Number* 92-1, Chapel Hill, N.C.

CRAMÉR, H. (1946). *Mathematical methods of statistics.* Princeton University Press, 1946.

DANIELS, R.W. (1978). *An Introduction to numerical methods and optimization techniques.* New York: Elsevier North-Holland, Inc.

DAVIDIAN, M. and GALLANT, A.R. (1991). The nonlinear mixed effects model with a smooth random effects density. North Carolina State University.

DE GROOT, M. (1975) Probability and statistics. Reading, MA: Addison-Wesley

DEMPSTER, A.P., LAIRD, N.M. and RUBIN, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, 39: 1-38.

DEMPSTER, A.P., RUBIN, D.B. and TSUTAKAWA, R.K. (1981). Estimation in covariance component models. *Journal of the American Statistical Association,* 76: 341-353.

DU TOIT, S.H.C. (1974). 'n Ondersoek na die doeltreffendheid van kleinste vierkante skatters. *Unpublished Msc. Thesis.* Pretoria: University of Pretoria.

DU TOIT, S.H.C. (1979). The Analysis of growth curves. *Unpublished Phd. Thesis.* Pretoria: University of South Africa.

DU TOIT, S.H.C. and BROWNE, M.W. (1982). Aufit (version 1). A computer program for automated fitting of nonstandard models for means and covariances. HSRC research report WS-27. Pretoria: HSRC.

DU TOIT, S.H.C. (1990). The analysis of repeated measurement time series models. Technical Report (Parts 1 & 2), Department of Statistics, Stanford University.

DU TOIT, S.H.C. (1990). Repeated measurement arma-processes. Technical Report no. 276. Department of Statistics, Stanford University.

DU TOIT, S.H.C. (1991). Multi-component models for the prediction of human growth based on marginal maximum likelihood estimation. Provisional Report, University of Pretoria.

EL LOZY, M. (1978). A critical analysis of the double and triple logistic growth curves. *Annals of Human Biology,* 5: 389-394.

FEARN, T. (1975). A Bayesian approach to growth curves. *Biometrika,* 62: 89-100.

GALLANT, A. R. (1975). Seemingly unrelated nonlinear regressions. *Journal of Econometrics* 3: 35-50.

GALLANT, A. R. (1987). *Nonlinear statistical models.* New York: Wiley.

GALLANT, A. R. and GOEBEL, J. J. (1976). Nonlinear regression with autocorrelated errors. *Journal of the American Statistical Association* 71: 961-967.

GENNINGS, D., CHINCHILLI, V. M. and CARTER, W. H. (1989). Response surface analysis with correlated data: A nonlinear model approach. *Journal of the American Statistical Association* 84: 805-809.

GILL, P.E. and MURRAY, W. (1974). Newton-type methods for unconstrained and lineary constrained optimization. *Math. Programming,* 7: 311-350.

GILL, P.E. and MURRAY, W. (1978). Algorithms for the solution of the nonlinear least squares problem. *Siam j. Numer. Anal* 15: 977-992.

GILL, P.E., MURRAY, W. and WRIGHT, M. (1981). *Practical optimization.* Academic Press, London.

GLASBEY, C. A. (1979). Correlated residuals in nonlinear regression applied to growth data. *Applied Statistics* 28: 251-259.

GOLDSTEIN, H. (1971). Factors influencing the height of 7 year old children-results from the National Child Development Study. In *Human biology,* 43: 92-111.

GRIZZLE, J. E. and ALLEN, D. M. (1969). Analysis of growth and dose-response curves. *Biometrics* 25:357-382.

HARTLEY, H. O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics,* 14: 174-194.

HENLY, S. J. (1991). *Unpublished Phd. thesis.* University of Minnesota, Minneapolis.

JENNRICH, R.I AND LEE, S Y (1979). A Study of algorithms for covariance structure analysis with specific comparisons using factor analysis. *Psychometrika* 44: 99-113.

JENNRICH, R.I AND MOORE, R.H. (1975). Maximum likelihood estimation by means of nonlinear least squares. American Statistical Association. *Proceedings of the statistical computing section.*

JENNRICH, R.I and SCHLUCHTER, M.D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 42: 805-820.

JENNS, R.M., and BAILEY, N. (1937). A mathematical model for studying the growth of a child. In *Human biology,* 9: 556-563.

JOHNSON, N.L. (1949). Systems of frequency curves generated by methods of translation. In *Biometrika,* 36: 149-176.

JOHNSON, L.W. and REISS, R.D. (1982). *Numerical Analysis* 2nd. ed (Reading, Mass.: Addison-Wesley), par 6.5.

JONES, R.H. and ACKERSON, L.M. (1990). Serial correlation in unequally spaced longitudinal data. In *Biometrika* 77, No. 4: 721-731.

KENDALL, M.G. and STUART, A. (1967). *The advanced theory of statistics,* 2, London: Griffin.

KHATRI, C.G. (1966). A note on a MANOVA model applied to problems in growth curves. In *Annals of the Institute of Statistics and Mathematics*, 18: 75-86.

KOOPS. W.J. (1986). Multiphasic growth curve analysis. *Growth* 50: 169-177.

KOOPS, W.J., GROSSMAN, M., and MICHALSKA, E. (1987). Multiphasic growth analysis in mice. *Growth* 51: 372-382.

KOOPS, W.J. (1988). Multiphasic Analysis of Growth Curves in Chickens. *Poultry Science* 67: 33-42.

KOTZ, S. and JOHNSON, N.L. (1988). Encyclopedia of statistical science. Vol. 8: 690-692.

KRYLOV, V.I. (1962). *Approximate calculation of integrals.* New York: Macmillan. (Translated from Russian by Stroud A.H.).

LAIRD, N.M. and WARE, J.H. (1982). Random effects models for longitudinal data. *Biometrics,* 38: 963-974.

LEE, J.C. and GEISSER, S. (1975). Applications of growth curve prediction. *Sankhyā, Series A* 37: 239-257.

LEE, J.C. (1988a). Prediction and estimation of growth curves with special covariance structures. *Journal of the American Statistical Association* 83: 432-440.

LEE, J.C. (1988b). Growth curve model and technological forecasting. In *Statistical Theory and Data Analysis,* Vol, II, K. Matusita (ed.): 369-378. Amsterdam: North-Holland.

LEE, J.C. (1991). Test and model selection for general growth curve models. *Biometrics* 47: 147-159.

LIANG, K. and ZEGER, S.L. (1986). Longitudinal data analysis using generalized linear models. In *Biometrika,* 73: 13-22.

LINDLEY, D.V. (1971) Bayesian statistics: A review Philadelphia: Society for Industrial and Applied Mathematics.

LINDLEY, D.V. and SMITH, A.F.M. (1972) Bayes estimates for linear models, (with discussion). *Journal of Royal Statistical Society,* Series B, 34:1-41.

LINDSTROM M.J. and BATES, D.M. (1990). Nonlinear mixed effects models for repeated measures data. In *Biometrics*, 46: 673-687.

LONGFORD, N.T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. In *Biometrika,* 74: 817-827.

MAGNUS J.R. and NEUDECKER,H. (1988). Matrix differential calculus with applications in statistics and econometrics. New York: Wiley.

MARITZ, J. S. and LWIN, T. (1989). *Empirical Bayes methods.* Chapman and Hall, New York.

McCULLOCH, C.E. (1982). Symmetric matrix derivatives with applicatons. *Journal of the American Statistical Association,* 77, 679-682.

MELNYK, M. (1974). Principles of Applied Statistics. *Pergamon Press inc.* New York.

MERRIL,M. (1931). The relationship of individual growth to average growth. In *Human Biology,* 3: 37-70.

MISLEVY, R.J. (1984). Estimating latent distributions. *Psychometrika,* 49: 359-381.

MUIRHEAD, R.J. (1982). *Aspects of multivariate statistical theory.* New York: John Wiley and Sons, Inc.

MORRISON, D.F. (1991). *Multivariate statistical methods.* New York: McGraw-Hill.

MOSTERD, J. (1991). Stogastiese parameter groeikrommemodelle. Unpublished Msc. Pretoria: University of Pretoria.

MULLER,K.E. and HELMS, R.W. (1985). Repeated meausres in nonlinear models. In *Proceedings of the Tenth Annual SAS Users Group International Conference:* 698-702. Cary, North Carolina: SAS Institute, Inc.

NOVICK, M.R. and JACKSON, P.H. (1974). *Statistical methods for educational and psychological research.* New York: McGraw-Hill.

PALMER, M.J., PHILLIPS, B.F. and SMITH, G.T. (1991). Application of nonlinear models with random coefficients to growth data. *Biometrics,* 47: 623-635.

PEIL, J. and HELWIN, H. (1981). A phenomenologic-mathematical model of growth dynamics. *Biometrica* 23: 41-54.

PORKESS, R. (1991). The Harper Collins Dictionary of Statistics. Harper Collins publishers. New York.

POTTHOFF, R.F. and ROY, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika,* 51: 313-26.

POWELL, M.J.D. (1972). Problems related to unconstrained optimization. In: W Murray (Ed). *Numerical Methods for Unconstrained Optimization.* London Academic

Press.

PREECE, M.A. and BAINES, M.J. (1978). A new family of mathematical models describing the human growth curve, *Annals of Human Biology.* 5: 1-24.

PRESS, S. J. (1982). *Applied multivariate analysis: using Bayesian and frequentist methods of inference.* 2*nd. ed*. Malabar, Florida: R.E. Krieger.

PRESS, W.H., FLANNERY,B.P., TEUKOLSKY,S.A. and VETTERING, W.T. (1988). *Numerical recipes in the Fortran . The Art of Scientific computing.* USA: Cambridge university Press.

RACINE-POON, A. (1985). Bayesian approach to nonlinear random effects models. *Biometrics* 41: 1015-1023.

RALTSON, M.L. and JENNRICH, R.I. (1978). Dud, A derivative-free algorithm for nonlinear least squares. *In Technometrics,* 20: 7-14.

RALTSON, A and RABINOWITZ, P. (1978). *A first course in numerical analysis,* 2nd ed. (New York: McGraw-Hill) par 4.4-4.8.

RAMSAY, J.O., ALTMAN, N. and BOCK, R.D. (1992). Variation in height acceleration in the Fels growth data.

RAO, C.R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. In *Biometrika,* 52: 447-458.

RAO, C.R. (1984). Prediction of future observations in polynomial growth curve models. In *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions,* 512-520. Calcutta: Indian Statistical Institute.

RAO, C.R. (1987). Prediction of future observations in growth curve models. In *Statistical Science* 2: 434-471.

REINSEL, G.C. and WINECK, M.A. (1987). Asymptotic distribution of parameter estimators for nonconsecutively observed time series. In *Biometrika,* 74, No. 1: 115-124.

RICHARDS, F.J. (1959). A flexible growth function for empirical use. In *Journal of Experimental Botany,* 10, No. 29: 290-300.

ROBERTSON, T.B. (1923). *The chemical basis of growth and senescence.* Monographs of Experimental Biology, Philadelphia.

ROCHON, J. and HELMS, R.W. (1989). Maximum likelihood estimation for incomplete repeated-measures experiments under and ARMA covariance structure. *Biometrics* 45: 207-218.

SEARLE, S.R. (1971). *Linear Models.* New york: Wiley.

SHEINER, L.B. and BEAL, S.L. (1979). NONMEM *Users Guide – Part 1: User's Basic Guide.* San Francisco: Division of Clinical Pharmacology, University of California.

SHUMWAY, R.H. (1988). *Applied statistical time series analysis.* Prentice-Hall.

SMITH, A.F.M. (1973) A general Bayesian linear model. *Journal of Royal Statistical Society. B* 35 67-75.

SMITH, A.F.M. (1973) Bayes estimates in one-way and two-way models. *Biometrika* 60, No. 2: 319-329.

SMITH, A.F.M. and SPIEGELHALTER, D.J. (1980). Bayes factors and choice criteria for linear models. *Journal of Royal Statistical Society. B* 42, No.2: 213-220.

SMITH, A.F.M. (1984). *Journal of Royal Statistical Society. A* 147, 245-259. Bayes factors and choice criteria for linear models. *Journal of Royal Statistical Society.* B 42, No.2: 213-220.

STEIMER, J.L., MALLET, A., GOLMARD, J., and BOISIEUX, J. (1984). Alternative approaches to estimation of population pharmacokinetic parameters: Comparison with the nonlinear mixed effects model. *Drug Metabolism Reviews* 15: 265-292.

STIRATELLI, R. LAIRD, N., and WARE, J. H. (1984). Random-effects models for serial observations with binary response. In *Biometrics* 40: 961-971.

STOKER, D.J. et al. (1980). *Inleiding tot Wiskundige Statistiek.* Published by authors. Pretoria.

STRASHEIM, C. (1989). The simultaneous normalization of test scores over various age levels. *Unpublished Msc thesis.* Pretoria: UNISA.

STROUD, A. H. and SECREST, D. (1966). *Gaussian quadrature formulas.* New Jersey: Prentice-Hall.

STUTZLE, W. (1977). Estimation and parameterization of growth curves, Juris Zurich.

TANNER, J.M., WHITEHOUSEA, R.H. and TAKAISHI, M., (1966), Standards from birth to maturity for height, weight, height velocity and weight velocity: *British children 1965. Archives of Diseases in Childhood,* 41, 454-471, 613-635.

THISSEN, D., BOCK R.D., WAINER, H., and ROCHE, A. (1976). Individual growth in stature: a comparison of four growth studies in the U.S.A. *Annals of Human Biology,* 3, 529-542.

THISSEN, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika,* 47: 175-186.

THISSEN, D. and BOCK R.D. (1988). Linear and nonlinear curve fitting. In *dA. von Eye (Ed.), New statistical methods in development research.* Hillsdale, NJ: Erlbaum.

TUDDENHAM, R.D. and SNYDER, M.M. (1954). Physical growth of California boys and girls from birth to eighteen years. *University of California Publications in Child Development,* 1: 183-364.

VONESH, E.F. and CARTER, R.L. (1987). Efficient inference for a random coefficient growth curve model with unbalanced data. *Biometrics* 43: 617-628.

VONESH, E.F. and CARTER, R.L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures. In *Biometrics* 48: 1-17.

WARE, J. H. (1985). Linear models for analysis of longitudinal studies. *American Statistician* 39: 95-101.

WINCEK, M.A. and REINSEL, G.C. (1986). An exact maximum likelihood Estimation procedure for regression ARMA time series models with possibly nonconsecutive data. In *Journal of Royal Statistical Society.* 48, No. 3: 303-313.

ZEGER, S.L., LIANG, K. and ALBERT, P. (1988). Models for longitudinal data: A generalized estimating equation approach. In *Biometrics* 44: 1049-1060.

ZELLNER, A. (1971). *An introduction to Bayesian inference in Econometrics.* New York: Wiley.

APPENDIX A

Table 1. Reaction times measured over 8 months on 9 rats.

Reaction time measured over 8 months

RAT	1	2	3	4	5	6	7	8	*Amount of neural patch laceration*
1	397.4	462.0	417.0	427.8	462.6	453.0	434.2	434.8	-0.543
2	480.6	450.0	455.6	472.4	477.8	468.0	452.4		-0.450
3	406.6	439.6	445.0	528.0	530.6	549.8	571.8	530.8	-0.399
4	424.6	428.4		470.0	469.4	481.8	478.8	480.0	-0.153
5	372.0	379.8	373.2	367.4	399.4	412.8	407.6	606.0	1.013
6	358.6	377.0	386.8	394.8	395.4	426.8	403.8	416.4	1.019
7	403.0	383.2	392.4	389.4	449.2	431.8	437.0		1.331
8	378.6	372.2	359.8	368.2	364.8	386.4	368.0	382.2	1.477
9	380.4	412.2	398.8	389.6	396.6	411.2	441.6	473.2	2.030

Table 2. Dental measurements of 11 girls and 16 boys.

Case no.	**Time**				**Case no.**	**Time**			
	8	10	12	14		8	10	12	14
1	21	20	21.5	23	15	25.5	27.5	26.5	27
2	21	21.5	24	25.5	16	20	23.5	22.5	26
3	20.5	24	24.5	26	17	24.5	25.5	27	28.5
4	23.5	24.5	25	26.5	18	22	22	24.5	26.5
5	21.5	23	22.5	23.5	19	24	21.5	24.5	25.5
6	20	21	21	22.5	20	23	20.5	31	26
7	21.5	22.5	23	25	21	27.5	28	31	31.5
8	23	23	23.5	24	22	23	23	23.5	25
9	20	21	22	21.5	23	21.5	23.5	24	28
10	16.5	19	19	19.5	24	17	24.5	26	29.5
11	24.5	25	28	28	25	22.5	25.5	25.5	26

12	26	25	29	31	26	23	24.5	26	30
13	21.5	22.5	23	26.5	27	22	21.5	23.5	25
14	23	22.5	24	27.5					

* Individuals 1-11 are girls, 12-27 are boys.

Table 3. Weights of 13 male mice measured at successive intervals of 3 days over 21 days from birth to weaning.

	Days						
Mouse	3	6	9	12	15	18	21
1	.190	.388	.621	.823	1.078	1.132	1.191
2	.218	.393	.568	.739	.839	.852	1.004
3	.211	.394	.549	.700	.783	.870	.925
4	.209	.419	.645	.850	1.001	1.026	1.069
5	.193	.362	.520	.530	.641	.640	.751
6	.201	.361	.502	.530	.657	.762	.888
7	.202	.370	.498	.650	.795	.858	.910
8	.190	.350	.510	.666	.819	.879	.929
9	.219	.399	.578	.699	.709	.822	.953
10	.225	.400	.545	.690	.796	.925	.836
11	.224	.381	.577	.756	.869	.939	.999
12	.187	.329	.441	.525	.589	.621	.796
13	.278	.471	.606	.770	.888	1.001	1.105

Table 4. Jawbone length measurements (in mm) of 20 boys.

Individual	Age				Individual	Age			
	8	8.5	9	9.5		8	8.5	9	9.5
1	47.8	48.8	49.0	49.7	11	51.2	51.4	51.6	51.9
2	46.4	47.3	47.7	48.4	12	48.5	49.2	53.0	55.5
3	46.3	46.8	47.8	48.5	13	52.1	52.8	53.7	55.0
4	45.1	45.3	46.1	47.2	14	48.2	48.9	49.3	49.8

5	47.6	48.5	48.9	49.3	15	49.6	50.4	51.2	51.8
6	52.5	53.2	53.3	53.7	16	50.7	51.7	52.7	53.3
7	51.2	53.0	54.3	54.5	17	47.2	47.7	48.4	49.5
8	49.8	50.0	50.3	52.7	18	53.3	54.6	55.1	55.3
9	48.1	50.8	52.3	54.4	19	46.2	47.5	48.1	48.4
10	45.0	47.0	47.3	48.3	20	46.3	47.6	51.3	51.8

Table 5. **Sheep data.**

The abstract of the data obtained from Dr. Greeff at the research laboratory of Irene are given in the following order.

1: Mass of the ewe after birth.
2: Sex of lamb (1=male, 2=female).
3: The identity code of ram.
4: The birth date of the lamb.
5: The birth weight of the lamb.
6-18: The weights of the lambs for the first 13 weeks.
19: The weaning weight at the 14th week.
20: The identity code of the lamb.
21: The identity code of the ewe.

Note that a 0 indicates a missing data point and the first two digits of the identity codes indicates the year of birth.

41	2	790051	810908	2.3	4	5	6	7	8	10	11	11	12	14	15	16	17	18.5	810049	800047
41	1	790051	810909	3.4	5	7	8	10	12	12	14	15	16	17	21	21	22	23.4	810067	800179
43	2	790051	810911	3.0	4	5	7	9	11	12	14	14	16	17	19	20	21	23.4	810074	800049
54	2	790051	810922	2.2	3	5	6	7	8	8	9	10	10	12	13	14	15	15.9	810096	790074
54	1	790051	810922	2.2	3	4	5	6	6	6	7	8	10	9	10	10	11	11.3	810097	790074
54	1	790051	810922	2.2	2	3	3	4	4	4	5	5	6	6	7	7	8	8.5	810098	790074
47	1	790051	811009	3.8	5	6	7	7	8	10	11	12	12	13	14	16	15	16.8	810114	800108
57	1	790101	830912	3.4	6	8	9	11	12	13	15	17	19	0	0	0	0	25.5	830113	800022
57	1	790101	830912	3.4	6	8	9	12	14	15	18	0	0	0	0	0	0	29.9	830114	800022
74	2	790146	820830	2.6	5	7	8	10	12	13	15	17	19	20	21	24	25	24.6	820006	790145
74	2	790146	820830	2.6	5	7	9	10	13	14	16	18	21	22	23	24	25	24.8	820007	790145

64	2	800020	830902	2.5	3	3	4	5	0	5	5	5	5	4	0	0	0	0.0	830014	800105
64	2	800020	830902	2.5	2	2	3	4	5	5	6	0	0	0	0	0	0	0.0	830015	800105
64	2	800020	830902	2.5	2	3	4	5	6	7	7	8	9	12	13	12	13	0.0	830016	800105
64	2	800020	830902	2.5	2	3	5	5	6	6	8	8	7	7	6	0	0	0.0	830017	800105
64	2	800020	830902	2.5	2	2	3	4	5	5	6	5	5	5	5	0	0	0.0	830018	800105
61	2	800020	830904	3.9	5	7	8	10	12	12	15	16	16	17	19	22	21	23.6	830044	745031
61	1	800020	830904	3.9	5	6	7	9	11	12	12	14	15	16	17	20	17	19.4	830045	745031
58	2	800020	830905	3.4	5	7	9	11	12	14	15	16	16	16	17	19	18	20.6	830056	760028

Table 6. Pollution data.

Timepoint	Value(p.p.m)	Timepoint	Value(p.p.m)	Timepoint	Value(p.p.m)
1	616.855	2	818.254	3	1322.637
4	844.011	5	1050.540	6	1180.684
7	1042.918	8	747.600	10	713.190
12	263.027	13	581.766	14	1077.637
15	714.565	16	1615.300	17	1135.947
18	1228.400	19	1321.792	20	448.354
24	301.492	25	581.062	26	620.200
27	715.146	28	812.405	29	521.200
30	284.211	31	1213.387	32	484.604
33	139.339	34	554.069	35	335.701
37	399.280	38	665.000	39	1348.672
40	575.399	41	665.429	42	266.226
43	936.629	44	236.629	45	894.708
46	698.000	48	288.403	50	711.546
51	1138.071	52	740.069	53	571.692
54	553.775	55	337.992	58	565.800
62	567.977	63	1293.680	64	456.973
65	728.662	66	974.878	67	1351.759
68	468.269	69	318.0	70	168.340

APPENDIX B

INPUT FOR MALE DATA IN EXAMPLE 5.5.1, CHAPTER 5

The following statements are used as input for AUFITPC:

PROG AUFIT options;
VARNAMES = variable names;
NCASES= number of cases per group;
GRNAMES=group names;
PARNAMES=parameter names;
INDPAR=information on free, fixed and equality parameters;
INDCON=information on equality and inequality constraints;
PARMVALS=initial values of unknown parameters;
LBOUNDS=lower bounds for parameters;
UBOUNDS=upper bounds for parameters;
TITLE=program title;
SCOV=values of sample covariance (correlation) matrix;
XBAR=values of sample mean vector;

REMARKS

Except for the PROG AUFIT statement, which should appear first in the list of instructions, the other paragraphs may be given in any order.

Free format is used throughout and commas as well as parenthesis may be used freely in each paragraph.

PROG AUFIT options;

Mode of analysis options are:

MODE=MEAN specifies analysis of structured mean vector (matrix) only.

MODE=COV specifies analysis of structured covariance matrix only.

MODE=MEAN+COV specifies analysis of structured mean- and covariance matrices.

MODE=CORR is the default, if MOD=type not specified. This specifies the analysis of a structured correlation matrix.

Type of discrepancy function options are:

DISCREP=ML good approximation to final solution available - use ML (maximum likelihood) discrepancy function throughout.

DISCREP=GLS+ML is the default. Initial approximation for ML obtained by GLS (generalised least squares).

DISCREP=GLS discrepancy function is GLS.

DISCREP=OLS discrepancy function is OLS (ordinary least squares).

Printout options are:

OUTPUT=PART is the default. Prints out all results, except the large sample correlation matrix of the parameters.

OUTPUT=FULL prints out all results.

Maximum number of iterations option is:

MAXITER=n (default is n=30). If n=0, limited printout is provided.

Convergence option is:

CONVERGE=value(default is 0.001).

Maximum number of stephalving option is:

MAXHALVE = n (default is n=3).

(ii) VARNAMES

The VARNAMES paragraph allows AUFIT to generate 3 types of variable names, which , for the sake of convenience, is denoted as scalar, vector and matrix names. Scalar names are names with no subscripts such as LENGTH or HEIGHT. Vector names are names with single numerical subscripts such as IQ1,IQ2,···,IQ12. Matrix names are names with double numerical subscripts such as PSI1,1;

PSI2,1; $\cdots$;PSI10,5.

(iii) NCASES

The NCASES paragraph supplies information on the number of cases per group and has the form:

NCASES=n_1 $n_2 \cdots n_g$;

(iv) GRNAMES

Group names follw the same convention as the variable names.

(v) PARNAMES

The parameter names may also be mixed and may be of scalar, vector or matrix type. Valid expressions are:

(a) PARNAMES=THETA1:THETA20 ALPHA:5*3 PSI TAU RHO
(b) PARNAMES=P1:P20;

(vi) INDPAR

The INDPAR paragraph supplies information on free, fixed and equality parameter. To specify for 10 free parameters the following statements are valid:

(a) INDPAR=1:10;
(b) INDPAR= 1 2 3 4 5 6 7 8 9 10;

If the element of INDPAR corresponding to the parameter in question, is set equal to zero, then this parameter is regarded as fixed in the subsequent analysis.

(vii) INDCON

The INDCON paragraph supplies information on inequality (≤ 0 or ≥ 0) as well as equality constraints.

If value = -1 then constraint is of the form ≤ 0
If value = 0 then constraint is an equality constraint
If value = 1 then constraint is of the form ≥ 0.

(viii) PARMVALS

Initial values for the parameters have to be supplied in this paragraph, or alternatively, be computed via the subroutine SPECIN.

(ix) LBOUNDS (Optional)

This paragraph contains information on the lower bounds for the parameters.

(x) UBOUNDS (Optional)

This paragraph contains information on the upper bounds for the parameters.

(xi) TITLE (Optional)

The program title must be less or equal to 80 alpha-numerical characters.

(xii) SCOV

In this paragraph the n(n+1)/2 distinct values of the sample covariance (or correlation) matrix is supplied. The corresponding statement in the SCOV paragraph is

SCOV = ;

(xiii) XBAR

In this paragraph the sample mean vector is supplied.
The corresponding statement in the XBAR paragraph is

XBAR= ;

The input data file to run AUFIT for the male Fels data in Chapter 5 is as follows:

```
PROG AUFIT  MODE=MEAN+COV DISCREP=ML  MAXITER= 1
        OUTPUT=PART  MAXHALVE=4  CONVERGE=0.001 ;
TITLE    =  MALE HEIGHT DATA ;
VARNAMES =  TIM2:TIM30 ;
NCASES   =  96  ;
INDPAR = 1:47;
PARNAMES =  gphi1:gphi36 gtheta1:gtheta8 gsigma^2  galpha gbeta ;
```

PARMVALS =

46.26993

-2.74716 0.18700

21.31685 -1.26545 33.63064

0.60121 -0.03624 0.24014 0.01755

-0.88388 0.05664 -0.30813 -0.01156 0.01858

-31.02166 1.84258 -17.67083 -0.39811 0.57867 31.73510

21.34192 -1.26558 8.87019 0.27888 -0.41142 -16.93392

15.21554

1.78576 -0.09941 0.88299 0.02069 -0.03254 -1.64857

1.39933 0.14553

76.74165

0.66951

87.44015

1.80262

-1.09774

21.44730

18.98271

1.06660

0.55982 0.42975 0.11233 ;

XBAR =

92.87169 96.84548 100.76302 104.38611 107.99583 111.71251

114.99826 118.29271 121.46979 124.60000 127.63958 130.65521

133.46146 135.79792 138.37396 141.53229 144.32396 146.95104

149.78750 152.75938 155.99427 159.76979 163.35104 166.84063

169.93750 173.35208 175.81875 177.56354 178.64792 ;

SCOV =

10.59180

11.15167 12.42921

11.96544 13.25453 14.67357

12.17951 13.55731 14.87773 15.75449

12.13434 13.55455 14.97821 15.72884 16.50527

12.66988 13.96568 15.43627 16.20097 16.88052 18.18876

12.85515 14.24052 15.73592 16.63878 17.29806 18.30922
19.10763
13.07583 14.59350 16.09359 17.05049 17.66305 18.67710
19.39299 20.25568
13.26667 14.75901 16.36078 17.37125 18.00599 19.11655
19.82223 20.56332 21.50065
13.88554 15.46429 17.09854 18.12825 18.90054 20.09309
20.86710 21.61724 22.39646 23.90792
13.65265 15.27968 16.96593 18.05889 18.81868 20.04341
20.82263 21.66044 22.49026 23.83031 24.33427
13.89219 15.54544 17.34646 18.50484 19.26867 20.48355
21.18618 22.07275 22.97750 24.24917 24.70240 25.56206
14.19643 16.05161 17.89577 18.98596 19.81721 21.11829
21.84903 22.80363 23.73977 25.12458 25.70173 26.27682
27.74216
14.76156 16.67370 18.56903 19.73122 20.64056 21.87857
22.56113 23.46014 24.52994 25.98667 26.51727 27.16803
28.39284 30.26833
14.76672 16.58503 18.56351 19.61712 20.52403 21.76866
22.52581 23.51294 24.51963 25.98156 26.55603 27.19779
28.52139 30.15776 30.86088
14.90309 16.86013 18.87347 20.05192 20.97355 22.34135
23.11535 24.28143 25.31025 26.77417 27.51820 28.15593
29.52072 30.92705 31.27855 32.63760
15.04714 17.04338 19.06503 20.23743 21.25629 22.55786
23.27633 24.34346 25.47760 26.97542 27.82145 28.41284
29.83988 31.38203 31.69437 32.81162 33.76161
15.46451 17.47449 19.64977 20.78524 21.76418 23.19539
23.97016 25.08787 26.16383 27.59698 28.49621 29.23708
30.65624 32.12261 32.64248 33.82929 34.41649 35.93958
15.75540 17.91505 20.25210 21.44543 22.45697 23.76005
24.58267 25.73017 26.90525 28.36510 29.35810 30.24246
31.68087 33.25456 33.90603 34.94759 35.69072 37.15501

39.34151					
16.45381	18.90589	21.37287	22.62927	23.50299	24.67842
25.70695	27.07205	28.32784	29.80052	30.84567	31.72141
33.23468	34.98471	35.64342	36.79423	37.62191	39.35187
41.64585	45.16345				
17.07985	19.37740	22.01592	23.13903	23.93273	25.16056
26.18518	27.59324	28.82852	30.47281	31.68174	32.52881
34.28593	35.79983	36.68959	37.81873	38.66295	40.76034
43.51186	47.16320	51.51791			
17.53082	19.83759	22.60461	23.63217	24.36190	25.67896
26.80141	28.16718	29.45377	31.07573	32.35036	33.40708
35.09644	36.45171	37.49671	38.61660	39.36166	42.01113
45.03243	49.26794	54.53118	59.59544		
17.62983	19.96632	22.70990	23.59111	24.25201	25.63495
26.74077	28.11193	29.55477	31.22427	32.45673	33.52854
35.35259	36.69709	37.83748	39.05283	39.75336	42.44687
45.49876	50.02072	55.71753	61.25113	64.37792	
18.51648	20.69920	23.56381	24.34667	25.00181	26.61881
27.47988	28.96165	30.56300	32.17510	33.30902	34.49057
36.27313	37.53561	38.85158	39.99327	40.53538	43.48699
46.14384	50.60811	55.94805	61.69571	64.77730	67.47116
18.42863	20.49743	23.14491	23.92449	24.51238	26.09510
26.80386	28.36210	29.93238	31.48677	32.45372	33.73522
35.29613	36.89966	37.95702	39.19868	39.60358	42.11652
44.37474	48.36298	52.72818	57.33113	60.53736	63.68921
62.38318					
17.22312	19.16289	21.58047	22.17527	23.02777	24.61082
25.17857	26.82502	28.20428	29.57271	30.50033	31.74015
33.27211	34.67615	35.58979	36.74905	37.09094	39.33745
41.21201	44.65055	48.02723	51.75845	54.57578	57.83903
57.18034	54.53916				
16.36467	18.25668	20.52094	21.28114	21.99110	23.69936
24.14925	25.60764	26.94265	28.38677	29.12624	30.40230

```
     31.58104   33.01108   33.61882   34.66846   34.93684   36.69592
     38.35034   41.16087   43.31495   45.73411   47.82915   50.94309
     51.15805   49.17309   47.19819
     15.52175   17.41684   19.64279   20.38767   21.11438   22.54054
     23.14786   24.40130   25.54400   26.95229   27.69738   29.01555
     30.07995   31.63482   32.20457   32.96264   33.38473   34.80988
     36.42652   38.71498   40.05052   41.71921   43.28082   45.86940
     46.37751   45.24086   44.25829   43.05065
     15.17120   17.00725   19.07755   19.84904   20.47412   21.82261
     22.40440   23.54748   24.60249   26.03677   26.62581   27.85194
     28.87070   30.54197   30.98125   31.59262   31.98031   33.06953
     34.44612   36.40309   36.95095   37.98863   39.04766   41.41545
     42.09622   41.35709   41.15723   40.60102   39.58395 ;
/*
10 10  (NTDIM2 NTDIM4)
-1 -1 -1             (IALPHA)
 2.491   2.991   3.491   3.991   4.491   4.991   5.491   5.991   6.491   6.991
 7.491   7.991   8.491   8.991   9.491   9.991  10.491  10.991  11.491  11.991
12.491  12.991  13.491  13.991  14.491  14.991  15.491  15.991  16.491 (VTIME)
```

Note that NTDIM2 and NTDIM4 denotes the number of Gauss quadrature terms to use when computing the expected values of the response function in the subroutine COVAX8.FOR . The values in the line immediately below the Gauss quadrature terms denote the values for α in the three component Richards function. The last three lines denote the selected timepoints.

APPENDIX C

BERKELEY DATA ANALYSES.

Uncorrelated errors:

(1) BERKELEY GROWTH STUDY MALES N=66

F= –2*LN(L)= N*118.7

MEAN VECTOR $\boldsymbol{\theta}$:

	Estimates	Std.Errors
1	81.127136	0.404335
2	1.387041	0.025976
3	75.716948	0.651938
4	2.250795	0.011638
5	0.360581	0.004199
6	23.977645	0.567817
7	16.316685	0.371770
8	1.192397	0.026986

COVARIANCE MATRIX $\boldsymbol{\Phi}$:

(i) Estimates:

1	19.13012							
2	-0.91678	0.08230						
3	3.50075	-0.16796	38.58706					
4	-0.03139	0.00181	0.07728	0.00216				
5	-0.08185	0.00690	-0.07944	-0.00017	0.00138			
6	-6.74768	0.32350	-16.66603	-0.02270	0.05509	25.76471		
7	2.71789	-0.15336	-0.23887	-0.00627	-0.01217	-6.14615	7.08039	
8	0.19256	-0.00603	0.06527	0.00032	-0.00062	-0.40529	0.51145	0.04567

(ii) Standard Errors

1	1.13727							
2	0.04129	0.00343						
3	1.33442	0.07538	3.01889					
4	0.02255	0.00133	0.03548	0.00052				
5	0.00702	0.00037	0.01129	0.00019	0.00009			
6	1.05570	0.05923	1.66615	0.03129	0.00926	1.85313		
7	0.37905	0.02076	0.63975	0.01171	0.00337	0.48518	0.23950	
8	0.02932	0.00162	0.04887	0.00088	0.00026	0.03843	0.01057	0.00143

CORRELATION MATRIX:

1	1.00000							
2	-0.73064	1.00000						
3	0.12885	-0.09425	1.00000					
4	-0.15451	0.13597	0.26782	1.00000				
5	-0.50453	0.64894	-0.34480	-0.09977	1.00000			
6	-0.30394	0.22216	-0.52857	-0.09627	0.29262	1.00000		
7	0.23353	-0.20090	-0.01445	-0.05069	-0.12336	-0.45505	1.00000	
8	0.20600	-0.09837	0.04916	0.03193	-0.07767	-0.37361	0.89939	1.00000

ERROR VARIANCE :

Estimate	Std.Error
0.55831	0.00000

(2) BERKELEY GROWTH DATA FEMALES N=70

F=-2*LN(L)= N*118.7

MEAN VECTOR $\boldsymbol{\theta}$:

	Estimates	Std.Errors
1	69.339921	0.496921
2	1.708820	0.030954
3	77.608391	0.667789

4	1.771315	0.021686
5	0.377633	0.004707
6	20.455258	0.529785
7	13.611135	0.213778
8	1.179327	0.017888

COVARIANCE MATRIX Φ:

(i) Estimates

1	15.52072							
2	-0.81392	0.11525						
3	-1.52777	0.07999	29.90624					
4	0.09332	-0.00370	-0.19986	0.02452				
5	-0.06426	0.01069	-0.05656	0.00096	0.00213			
6	-8.08989	0.42550	-12.33809	0.03543	0.06137	31.91939		
7	2.24904	-0.22870	-1.73213	0.02143	-0.01730	-7.14624	4.04894	
8	0.17070	-0.00800	-0.23488	0.00574	0.00019	-0.41646	0.27780	0.02819

(ii) Standard Errors

1	1.52831							
2	0.07224	0.00682						
3	1.41039	0.10179	2.78920					
4	0.05009	0.00372	0.07298	0.00373				
5	0.01119	0.00073	0.01486	0.00056	0.00016			
6	1.14908	0.07822	1.48959	0.05910	0.01197	1.69626		
7	0.34234	0.02279	0.44862	0.01795	0.00357	0.33056	0.14210	
8	0.03268	0.00220	0.04290	0.00168	0.00034	0.03334	0.00838	0.00131

CORRELATION MATRIX:

1	1.00000				
2	-0.60855	1.00000			
3	-0.07091	0.04309	1.00000		
4	0.15127	-0.06970	-0.23339	1.00000	
5	-0.35265	0.68104	-0.22360	0.13264	1.00000

6	-0.36346	0.22184	-0.39934	0.04005	0.23487	1.00000		
7	0.28371	-0.33479	-0.15741	0.06802	-0.18594	-0.62861	1.00000	
8	0.25808	-0.14038	-0.25582	0.21841	0.02496	-0.43904	0.82228	1.00000

ERROR VARIANCE :

Estimate	Std.Error
0.32628	0.00000

***ARMA*(1,1) DEVIATIONS:**

BERKELEY GROWTH DATA MALES N=66

F=−2∗LN(L)= N∗69.8

MEAN VECTOR $\boldsymbol{\theta}$:

1	51.11639
2	1.56381
3	117.16589
4	1.13019
5	0.23458
6	16.58997
7	21.75483
8	1.57483

COVARIANCE MATRIX $\boldsymbol{\Phi}$:

1	35.72877							
2	-1.30306	0.13223						
3	-10.64028	0.38699	34.26880					
4	0.66273	-0.01739	-0.19250	0.01975				
5	-0.00191	0.00249	0.00931	-0.00020	0.00025			
6	-3.75982	0.13713	-23.93465	-0.07376	-0.00687	46.64464		
7	2.75895	-0.09739	-0.80850	-0.05143	-0.00005	-0.77480	6.76420	
8	0.25177	0.00417	-0.20384	0.00597	0.00032	0.20272	0.34566	0.03023

CORRELATION MATRIX:

1	1.00000							
2	-0.59951	1.00000						
3	-0.30408	0.18180	1.00000					
4	0.78892	-0.34028	-0.23398	1.00000				
5	-0.02036	0.43605	0.10138	-0.09206	1.00000			
6	-0.09210	0.05522	-0.59866	-0.07685	-0.06407	1.00000		
7	0.17747	-0.10298	-0.05310	0.14070	-0.00125	-0.04362	1.00000	
8	0.24224	0.06592	-0.20025	0.24439	0.11882	0.17070	0.76433	1.00000

$ARMA(1,1)$ PARAMETERS

$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\beta}$
0.10857	0.96603	0.16415

BERKELEY GROWTH DATA FEMALES N=70

F=−2∗LN(L)= N∗75.46

MEAN VECTOR $\boldsymbol{\theta}$:

1	71.51844
2	1.75703
3	73.63314
4	1.93508
5	0.40535
6	21.87592
7	12.86096
8	1.11132

COVARIANCE MATRIX $\boldsymbol{\Phi}$:

1	13.14121			
2	-0.83825	0.10820		
3	1.96016	-0.12502	20.49084	
4	0.00078	0.00952	-0.12700	0.00400

5	-0.03126	0.00995	-0.06195	0.00168	0.00207			
6	-4.77515	0.30574	-7.33915	0.04162	0.03032	18.71622		
7	1.67541	-0.17371	-2.37196	0.00491	-0.00627	-3.64289	2.78060	
8	0.19950	-0.01028	-0.30012	0.00270	0.00082	-0.23981	0.21866	0.02648

CORRELATION MATRIX:

1	1.00000							
2	-0.70297	1.00000						
3	0.11945	-0.08396	1.00000					
4	0.00341	0.45780	-0.44362	1.00000				
5	-0.18945	0.66469	-0.30067	0.58431	1.00000			
6	-0.30448	0.21484	-0.37476	0.15213	0.15399	1.00000		
7	0.27716	-0.31669	-0.31424	0.04658	-0.08259	-0.50497	1.00000	
8	0.33822	-0.19213	-0.40747	0.26246	0.11129	-0.34068	0.80590	1.00000

$ARMA$(1,1) PARAMETERS

$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\beta}$
0.17001	0.99045	0.22576