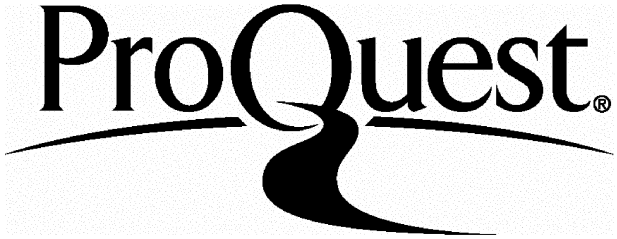
Hierarchical Bayesian models for linear and non-linear animal growth curves

A thesis presented for the degree of Doctor of Philosophy
in the University of London
by

Julie A. Gilg

Department of Statistical Science
University College London
1-19 Torrington Place, London, WC1E 6BT

2000



Abstract

There are many possible ways to analyse repeated measures such as animal growth data. Recent developments in computational methods mean that the natural approach of modelling the growth of each animal with a parametric curve with the parameters allowed to vary randomly between animals is now practically as well as theoretically feasible. The basic model structure is one level for individuals, one for the population and a third for prior beliefs. This means that the individuals are modelled as being a sample from some population, as indeed they are.

We have used Markov chain Monte Carlo methods to fit such models to data for pigs and for cats. For one data set the growth was only recorded over a short time and was approximately linear. For this example we were able to use Gibbs sampling. Over longer time periods animal growth is generally non-linear. We discuss some of the commonly used growth functions for fitting such data. When using these non-linear functions at the first stage of our models we used random walk Metropolis algorithms in order to fit the models.

We also include an analysis of some data which included measurements of various body components made after slaughter as well as series of live weights. For this data we were able to use a more sophisticated model which used diphasic functions at the first phase. These functions comprised of the sum of two phases each of which represented a separate group of body components. This approach provided information on the development, or changes in the form of the animals over time, as well as on the overall growth.

Acknowledgements

Firstly, and most importantly, I would like to thank my supervisor Tom Fearn for his guidance and support throughout this work.

Also, my thanks go to what was Dalgety for their financial support. The project group comprising members from the former components of Dalgety was very helpful, in partcular I would like to thank Mary Garratt, Steve Jagger, Pieter Knap, Sue Leeke and Hein van der Steen.

Finally, thanks to the other students at UCL for their help, comments and friendship.

Contents

List of Figures

List of Tables

Chapter 1

Animal Growth

1.1 Outline of thesis

This first chapter discusses animal growth. In particular, possible methods of analysing repeated measurements data, commonly used growth functions and the possibility of drawing biological interpretations from the fitted parameters of the relatively complex multiphase growth functions.

The second chapter outlines Bayesian hierarchical models which can be used for modelling animal growth. The chapter goes on to introduce and discuss Markov chain Monte Carlo methods which can be used to fit and draw inferences from these hierarchical models.

Chapter 3 illustrates the modelling process for the relatively simple case where the growth is approximately linear over the time period for which we have data. In these cases all of the posterior conditional distributions are standard distributions and Gibbs sampling is straightforward. The chapter also compares the results from our analyses with those from using other methods.

Chapter 4 illustrates the fitting of some non-linear growth data. Some of the conditional distributions are no longer standard and a form of Metropolis-Hastings algorithm was used for the sampling for these parameters. This process was made easier because there was a large amount of data for all individuals.

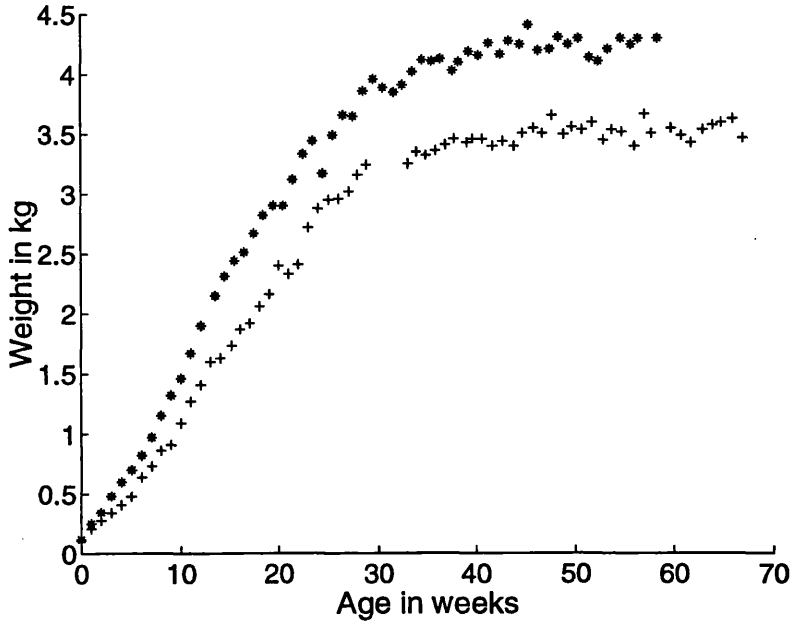


Figure 1.1: Growth of two male cats

The analysis of a more complex data set is described in Chapter 5. As well as a series of live weight measurements for each animal there were also measurements of various body components made after the animals were slaughtered. By using a multiphasic growth function we were able to include some of this extra information in our models. The situation was further complicated by the data including differing amounts of data on different animals, in some cases very few measurements, and the lack of any animals with measurements to, or beyond, maturity.

This work was funded by a CASE Studentship and in Chapter 6 we discuss the value and practicality of using these methods within the sponsoring company.

1.2 Animal growth studies and their aims

There is a lot of interest in the growth of animals and studies aiming to further our knowledge are frequently conducted. In this thesis we have used repeated measurements data. For each animal, the measurements of body weight, or another quantity, form a growth curve showing the pattern of growth over time. We have used data for pigs and for cats in this thesis. Figure 1.1 shows the

growth of two of the cats from the data used in Chapter 4.

Some of the reasons why we may be interested in this type of data are given below:

- to compare the growth of two or more groups (e.g. diets),
- to study the biological processes involved in growth,
- to monitor whether or not an individual's growth is 'normal',
- to quantify the amount of variability between individuals,
- to investigate genetic variability.

1.3 Possible methods of analysis

There are a number of possible ways to analyse repeated measures data. The simplest is perhaps to fit growth functions to the data in order to assist in describing it or to provide smoothed plots of growth. Other methods include single and multiple t-tests. For example, we could use a two-sample t-test to compare two groups at a certain time (for example three months of age). An immediate problem with this is that if measurements have not been taken at the same age for all animals we cannot proceed. A further objection is that all of the data is ignored except for one particular snapshot of it. Aswell as being wasteful, this raises the point that the choice of which age to use will affect the result. To avoid this we may repeat the t-test at every time point. This again assumes that we have measurements at standard times for all individuals which is often not the case. Further, we have a problem of multiple comparisons and we should also note that for such tests there is generally dependence between the test statistics. For example, tests at three and at four weeks are clearly not independent.

In a landmark paper, Wishart (1938) introduced the two-stage model. The first stage was to fit curves to the growth data for each individual. The second was to analyse the fitted parameters as if they were the raw data. This approach has since been improved upon by combining the two stages into a single

model. To do this we assume that the subject-specific parameters are random variables from a distribution whose mean and variance are also estimated by the model. Lindley and Smith (1972) added a third stage by including priors for the second stage parameters. Fearn (1975) applied Lindley and Smith's model to linear growth data. Various approximations were used in order to estimate the posterior distributions.

In their review of the mathematical models used in the first 50 years of the journal Growth, Zeger and Harlow (1987) stated that random effects models (such as those described above) have important applications to growth data. However, they noted that work was needed in order to be able to use such models with the typically, non-linear, functions commonly used to model growth.

Markov chain Monte Carlo (MCMC) methodology means that fully Bayesian analyses are now possible for linear and non-linear growth and without the need for restrictive assumptions or approximations. Gelfand et al (1990) include an analysis of linear growth data using the Gibbs sampler. Wakefield et al (1994) used the Gibbs sampler to reanalyse the data from Fearn. They also used the generalized ratio-of-uniforms technique in an analysis of non-linear pharmacokinetic data.

In this thesis we illustrate the use of MCMC methods to fit Bayesian hierarchical (population) models to both linear and non-linear growth data. One of the benefits of population modelling is that information about the individuals is combined to give information about the population. Also, the combined information for all individuals tells us more about each individual than its data alone. One advantage of this is that individuals with sparse data can be included in the model and make a potentially useful contribution to it. The first step in fitting these models is to choose an appropriate growth function to use at our first stage to model the individual level growth. The remainder of this chapter is therefore dedicated to the subject of growth functions.

1.4 Growth functions

Over the years many growth functions have been proposed. The aims of those working in this field have varied between looking for some kind of biological laws of growth to simply wanting to be able to make useful empirical summaries. Over time the emphasis has shifted towards using the functions as tools rather than as representing laws of growth. This shift has been because of the failure to find the desired fundamental law.

Growth curves are used in other fields as well as for animal growth. For example, plant growth (Hunt, 1982) and population growth (Solomon, 1976). There is quite a large overlap between the functions used in the various fields.

The level of complexity, or number of parameters, required in a growth function will depend on the specific data we wish to analyse. For data over only a short period of time the growth is often well approximated by straight lines. When the data is for a somewhat longer period we may find that the growth is only mildly curvilinear. In these cases low order polynomials may be appropriate.

When we have measurements over a longer time period growth generally follows a roughly sigmoid shape. In these cases a 'specialized' growth function, rather than a polynomial, is usually the best choice.

A large number of possible functions have been suggested. These generally have a lower asymptote of zero and an upper aymptote which represents the mature weight. We note here that some species grow continuously throughout their life. In these cases a function with no upper asymptote is needed. For many species the approximate mature weight is reached and then very slow growth continues, for example, in humans weight tends to increase with age. In these cases a function with an upper asymptote will usually be adequate. This is especially so for the common cases when the data does not extend to, or far beyond, maturity.

1.4.1 Single phase functions

Many of the proposed functions are single phase functions. By this we mean that if we plot the rate of growth against time we find a single peak, in other words, the growth rate increases until it reaches its maximum at the point of inflection and then decreases until zero when mature weight has been reached. Many three or four parameter functions of this form have been suggested. The most commonly used of these are outlined below. In each case, w represents the weight and t the time or age. Figures 1.2 to 1.4 illustrate the three-parameter functions outlined here. The solid lines show the weights and the dotted lines show the rates of growth. In each case we have chosen the parameters so that the maximum weight is 200kg, the point of inflection is at 100 days and the maximum growth rate is 1kg/day.

Figures 1.5 to 1.8 show these functions on the same axes enabling easier comparisons between them. For Figure 1.5 the inital weights, final weights and maximum rate of growth are the same for all three functions. For Figure 1.7 the point of inflection and the initial and final weights are the same for all of the functions. In each case the accompanying figure shows the rates of growth for each function.

Gompertz

The Gompertz function can be written as:

$$w = ae^{-e^{(-k(\frac{t}{c}-1))}}.$$

The three parameters must all be greater than zero and have the following interpretations:

- a maximum weight,
- c point of inflection,
- k 'rate of growth' (dimensionless).

The weight at the point of inflection is $0.368a$ or $\frac{a}{e}$. The maximum growth rate is at this point and is given by $\frac{ak}{ec}$.

Logistic

The Logistic function can be written as:

$$w = \frac{a}{2}(1 + \tanh(k(\frac{t}{c} - 1))).$$

The three parameters must all be greater than zero and have the same 'meanings' as for the Gompertz. The weight at the point of inflection is $\frac{a}{2}$. The maximum growth rate is at this point and is given by $\frac{ak}{2c}$.

von Bertalanffy

The von Bertalanffy function can be written as:

$$w = a(1 - de^{-ft})^3.$$

The three parameters must again be greater than zero. Here a is the maximum weight but the d and f parameters are not so easy to interpret. The point of inflection is at time $\frac{\ln(3d)}{f}$ and weight $\frac{8a}{27}$ or approximately $0.3a$. The maximum growth rate is given by $\frac{4af}{9}$.

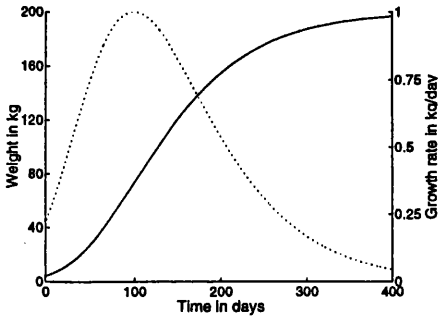


Figure 1.2: The Gompertz function

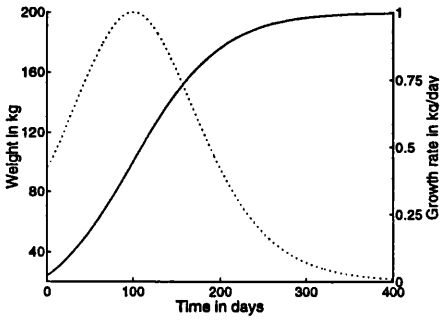


Figure 1.3: The Logistic function

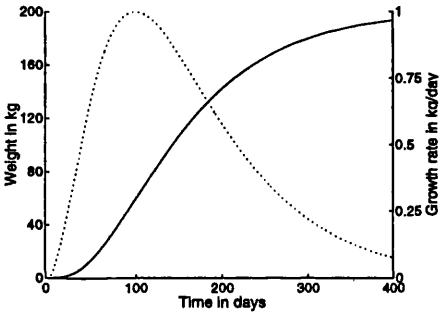


Figure 1.4: The von Bertalanffy function

Richards

For each of the above functions the point of inflection is fixed at a certain percentage of the mature weight. This obviously limits their flexibility and also makes it less likely that they will be applicable to a wide range of species.

Richards introduced a fourth parameter in order to allow the point of inflection to be flexible. Venus and Causton (1979) give the Richards function in the following form:

$$w = a(1 \pm e^{(b-kt)})^{-\frac{1}{d}}$$

where, $a, k > 0$, $-1 \leq d < \infty$ and $d \neq 0$, b may be any real number.

From the plus or minus option we use plus when d is positive and minus when d is negative. If d is -1 then there is no point of inflection. For values of

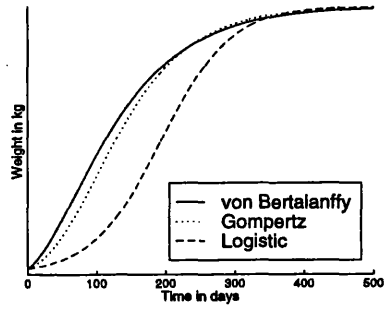


Figure 1.5: The 3 functions with fixed maximum growth rate

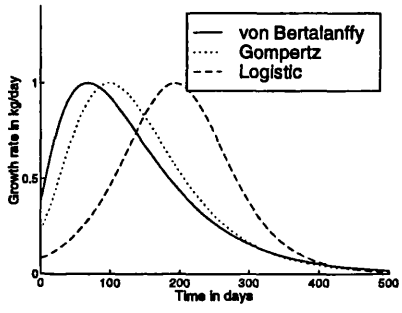


Figure 1.6: Growth rates for the functions in Figure 1.5

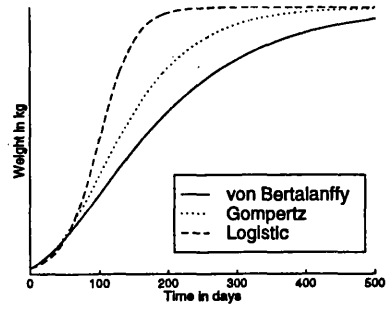


Figure 1.7: The 3 functions with fixed point of inflection

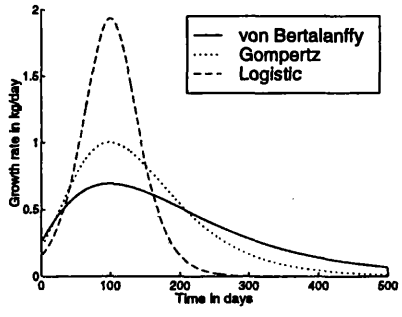


Figure 1.8: Growth rates for the functions in Figure 1.7

d greater than -1, as d increases the point of inflection gets later (as expressed as a percentage of mature weight).

We note here that if, for a particular species, none of the available three parameter functions had an appropriately placed point of inflection but we still wished to use only three parameters it would be possible to use the Richards function but with one parameter fixed in order to give the required position of the point of inflection.

1.4.2 Multiphase functions

Single phase functions may be adequate summaries of growth in many cases, especially when we have relatively few measurements for each animal. However, in some cases more complex functions may be required. A number of studies have found that when weight gain or growth rate is plotted against time it often has more than one peak. For example, in humans three peaks have been distinguished and generally accepted. The largest peak is within the first year. The second peak is at about 7 years of age, sometimes called the mid-growth spurt. The third peak is at about age 12 for girls and age 14 for boys. This third peak is known as the pubertal or adolescence growth spurt and seems to be unique to primates (the highest order of mammals) (Tanner, 1962). Because of these multiple peaks Koops (1989) proposes the use of multiphase rather than single phase functions.

As well as looking at the growth of the whole body, many studies have considered the growth of separate body components. These are known as allometric studies and can relate the growth of the various body components to the growth of other components and also to the body as a whole. These studies have shown that different parts of the body grow/mature at different ages. The order in which components mature is generally related to their importance to the functioning of the body. This order is generally nervous system, bone, muscle and finally fat. For example, Walstra (1980) studied the growth and carcass composition of Dutch Landrace pigs and found this to be the case. These differences in

the relative growth rates of the various body components result in development, i.e. changes in the form of the animal over time.

Different body components maturing at different times fits in with there being several peaks in the growth rate curve. An obvious suggestion is that each peak broadly represents a separate component or a group of components which have similar patterns of growth. A multiphase rather than a single phase function will be a better fit to the data when we have more than one peak in growth rate. If we can justify any biological interpretations of the phases then such information on the growth of the various components would be very useful.

Before Koops there was very little work done on multiphasic growth. He suggests a number of reasons for this:

- the idea that 'growth laws' are the basis for single phase growth functions,
- in many cases a few parameters are adequate to summarize the growth,
- phases are almost undetectable if measurements are not taken frequently and over a long enough period of time,
- the added complexity of using functions with more parameters.

The specific multiphase function suggested by Koops was a summation of n logistic functions where n is the number of phases. The logistic function was chosen because many data sets had shown growth rate to be in symmetrical bell-shaped phases. The logistic function is appropriate in this case. We note that any other suitable function could be used in its place if, for example, the phases were not symmetrical or the logistic was generally found to be a poor fit to the data.

The simplest multiphasic function is the diphasic. This is the sum of two single phase functions and so has six parameters (unless you use the four-parameter Richards function as the two single phase components). Figure 1.9 shows a diphasic-logistic function (with the two phases shown separately as well as the totals and with the growth rates shown as dotted lines). The parameters used

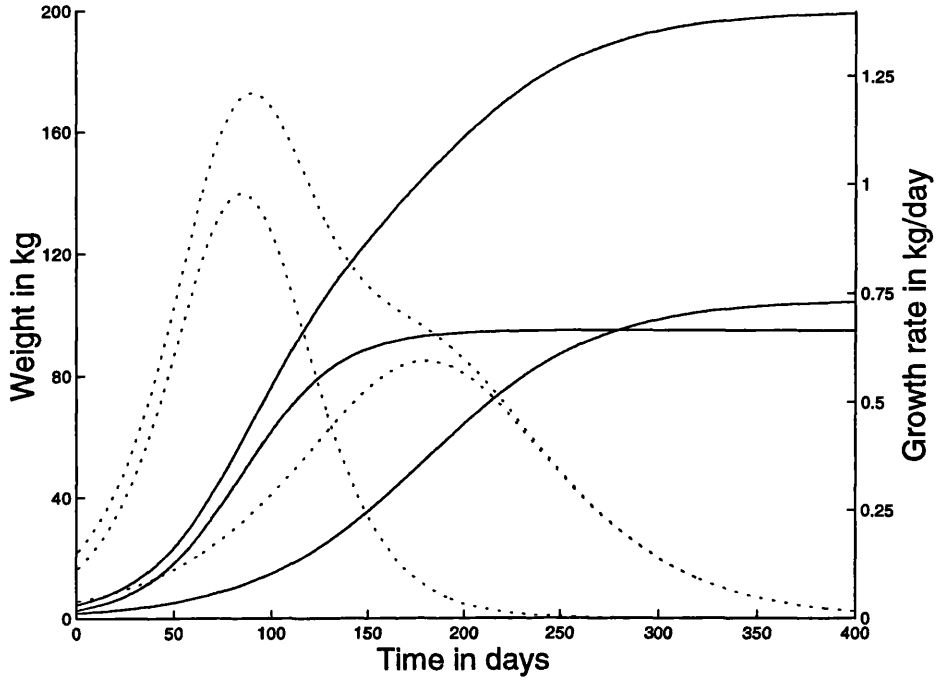


Figure 1.9: Diphasic function (two logistics)

here were (95kg, 1.75, 85days,105kg, 2.04, 180days). Therefore, the function was given by:

$$w = \frac{95}{2}(1 + \tanh(1.75(\frac{t}{85} - 1)))$$

$$+\frac{105}{2}(1 + \tanh(2.04(\frac{t}{180} - 1)))$$

We would expect the diphasic to be a better fit to observed data than a three or four parameter function because of the extra parameters. This does not necessarily make it a better choice. For example, we should not overfit the data when a simpler function is adequate. Nevertheless, when we have a reasonable amount of data the diphasic function may well be a sensible choice.

We note that care must be taken when attempting biological interpretations of the parameters/phases. This is especially the case when we have no data on separate body components with which to back up these interpretations. For example, just because a diphasic function is a good fit to some growth data it does not necessarily mean that the two phases correspond to two actual (groups of) components. The next section discusses ways of measuring the growth of

components and also ways of relating these components to phases of growth.

1.4.3 Attempts at justifying biological interpretations

Unfortunately, measurements of body components are not as easy to obtain as those of the whole body. The most commonly used method of taking such measurements is to do chemical analyses of the body after slaughter. Using this method we can only obtain one set of measurements for each animal rather than several measurements over time. By slaughtering the animals at a range of different ages we can get an idea of the growth of the body components over time. However, this data is not as useful as serial measurements on each animal. We also note here that these chemical analyses do not give such accurate results as a simple weighing of the whole animal.

One way of obtaining serial measurements is to use a non-invasive technique, for example, dual X-ray absorptiometry. Such an approach also means that the animals do not need to be killed. Munday et al (1994) used this method in a study of the body composition of domestic cats. Unfortunately they did not make serial measurements on each animal but instead measured cats at ages ranging from 8 weeks to 10 years. These data confirmed the expected order of growth of components (bone, muscle and finally fat).

A few studies have used data obtained from chemical analyses to investigate whether we can interpret the separate phases of a multiphasic function as representing different body components. Koops and Grossman (1991) and Kwakkel et al (1993) have done this for pigs and for pullets (young hens) respectively. Their methods were to fit diphasic-logistic functions to the average body weights of a group of animals and then separately to fit single phase logistic functions to the components data. At each age the component value used is the average value for the animals slaughtered at that age. Finally, they considered whether the fitted parameters obtained from these two approaches were in general agreement.

Koops and Grossman used two data sets for their work on pigs. For the first of these their overall measure was not whole body weight but total dry

matter (which is roughly total body weight minus water content). They fitted a diphasic function to total dry matter and single phases to each of the two components, fat and fat-free dry matter. There was general agreement between the parameters obtained.

For their second data set the overall measure was carcass side weight. Data was available for five components: offal, muscle, bone, skin and fat. These were grouped by age at maximum gain into: i) offal, muscle and bone and ii) skin and fat. The agreement between parameters in this case was good for the asymptotic weights but less good for the age at maximum gain and the rate of growth parameters. This data was somewhat limited by only including eight measurement points, only one of which was at greater than about half of mature weight, making any conclusions rather tenuous. (The other data was better as it had 19 measurement points). Chapter 5 of this thesis outlines our attempts to investigate the same questions addressed by these papers by using a more sophisticated analysis.

1.4.4 Choosing a growth function

Having rejected the idea that any one growth function represents a fundamental law of growth (at least amongst the growth functions presently available) we are faced with the choice of which function to use for any particular analysis.

Walstra (1980) stated that 'the choice between empirical models is hampered because models have hardly been mutually compared, especially in the case of pigs, on actual growth data.' Since that time, Zullinger et al (1984) tested the logistic, von Bertalanffy and Gompertz functions on data for 331 mammalian species. They found that each of the three functions gave the best fit for some of the species. However, they wished to choose only one function for all of these species in order to enable comparisons to be made between their fitted parameters. They found the Gompertz function to be most suitable for this. Since then (and before) the Gompertz function has often been used to model mammalian growth. For example, Begall (1997) used it for data on Zambian

common mole-rats. The von Bertalanffy function has often been used to model the growth of fish, for example by Misra (1980).

The choice of function is often based on the location of the point of inflection. Differences in this location are believed to reflect differences in the time of onset of puberty (Brody, 1945) . The point of inflection is generally reached at about 30% of mature weight in higher animals. Humans appear to be unique in having a long juvenile period and having the point of inflection at about 60% of mature weight.

In order to choose suitable functions we first need to consider the general requirements of a good growth function:

- it should fit well at all parts of the curve (for which we have data),
- it should not have too many parameters for the amount of data available (i.e. it should not overfit the data),
- it is useful if the parameters have a biological interpretation (although care is needed to ensure that the interpretations are supported by the data),
- it should be easy to fit (this is less important now due to more sophisticated fitting methods being available than in the past).

Commonly used criteria to decide whether a function is suitable or which function fits best are:

- small residuals,
- absence of systematic deviations (trends in residuals),
- realistic parameter estimates.

Unfortunately, results from these criteria do not always agree, for example, functions with a good fit can give poor estimates, especially for mature weight. This is not too much of a problem because the mature weight parameter is often an asymptote (much) larger than any of the observed data. In these cases, any biological interpretations of such estimates are unreliable in any case.

If there is little to choose between the fits of two or more functions then, ideally, rather than simply using one of the functions, we should consider the robustness of any inferences to this choice. For example, if the Gompertz function was chosen rather than the logistic would this substantially change the conclusions from the analysis? If it would then we need either theoretical or empirical justification for our choice of function in order to support our conclusions.

Chapter 2

Hierarchical Models and MCMC Methods

2.1 Introduction to hierarchical models

Growth models fit naturally into a hierarchical framework. At the first stage we have our chosen growth function for which each individual has their own parameters, θ_i. We then model the θ_i as a sample from some population distribution. This distribution forms our second stage. The θ_i are not observable quantities but we can use the data, y, to estimate them and aspects of their population distribution. The parameters defining the population distribution are again random variables for which we assign prior distributions based on further parameters known as hyperparameters.

Non-hierarchical models are not ideal in many cases because if they have few parameters they often cannot fit large data sets accurately. With more parameters they can overfit the data. Hierarchical models avoid these problems because they have enough parameters to fit the data well but the use of population distributions structures some dependence into the parameters thus avoiding overfitting. Also, the hierarchical approach allows us to include prior information in our models.

2.2 Introduction to the Bayesian approach

Lengthy debates about the relative merits of Bayesian and frequentist methodologies are outside the scope of this thesis. However, we note that for relatively simple problems, frequentist and Bayesian methods often give similar results. An important advantage of the Bayesian approach is that it is relatively easy to extend Bayesian models in order to increase their complexity.

A quick introduction to the Bayesian approach follows. For more information see, for example, Lee (1989) (introductory), Gelman et al (1995) (more comprehensive) or O'Hagan (1994) (encyclopaedic).

A Bayesian model allows us to make probability statements about parameters and unobserved data, ϕ, conditional on the observed data, y, and any observed covariates, z.

Given the prior distribution for ϕ and the probability density function of y conditional on ϕ we obtain the joint probability distribution,

$$p(\phi, y) = p(\phi)p(y \mid \phi).$$

Using Bayes' rule we then obtain,

$$p(\phi \mid y) = \frac{p(\phi, y)}{p(y)} = \frac{p(\phi)p(y \mid \phi)}{p(y)}.$$

Since $p(y)$ does not depend on ϕ and can be considered a constant for fixed y, we can omit it giving,

$$p(\phi \mid y) \propto p(\phi)p(y \mid \phi).$$

This is the joint posterior density over all model parameters. We can obtain the marginal posterior distributions for any parameter of interest by integrating the joint distribution over the remaining parameters. However, the form of $p(y \mid \phi)$ and $p(\phi)$ and the typically high dimensions mean that the integrations involved are not straightforward. Fortunately, there is an alternative approach which is to use Markov chain Monte Carlo methods to simulate samples from the joint posterior distribution. From the samples obtained we can then obtain the marginal distributions by simply taking the sampled values for the particular parameter and ignoring the sampled values of the other parameters.

Prior beliefs

As we have seen above, Bayesian inference involves specifying prior beliefs about our model parameters. By combining these prior beliefs with the observed data we obtain the posterior distribution, in other words, the probability distribution of our parameters, ϕ, given the data y and our prior beliefs. The posterior distribution is a compromise between the prior information and the observed data. As the amount of data increases the posterior distribution is increasingly dominated by the data at the expense of the prior beliefs.

Non-informative prior distributions

It is sometimes possible to use prior ignorance or improper priors (see, for example, Gelman et al, 1995) in our models. We may wish to do this to remove the subjectivity involved in the choice of prior and to 'let the data speak for themselves'. However, improper priors are not properly defined distributions. This leads to questions about how sensible their use is. Also, various problems can arise when they are used. These include the following points.

In some cases using improper priors will lead to improper posterior distributions. For example, Hobert and Casella (1996) consider the effects of using improper priors for the variance components of hierarchical linear mixed models. The conjugate structure of the prior specification allows the Gibbs conditionals to be obtained and the Gibbs sampler used without any need to establish that the posterior is proper. They give an example where a posterior is improper due to having an infinite amount of mass near $\sigma^2 = 0$ where σ^2 is the second-stage variance parameter. (The same thing occurs for the models used in this thesis for which we use a second-stage covariance matrix, Σ, instead of the variance σ^2. See page 34.) The authors then discuss how this example shows that the Gibbs sampler output relating to an improper posterior may appear perfectly reasonable. Thus, the user may proceed to draw inferences from the model without realising that the posterior is not, in fact, a proper distribution. They also give a theorem which may be used to check whether the posteriors resulting from

improper priors are proper. Gelfand and Sahu (1999) also discuss the use of improper priors as well as parameter identifiability (and the relationships between these two topics). They give results for checking whether a posterior reulting from a (partly) vague prior specification is proper in the context of generalized linear models.

Often there are many possible improper prior distributions which could be used. For many problems a density that is vague in a given parameterization will not be so for another. (In some cases Jeffreys' priors may be used, these are invariant to changes in the parameterization). Dawid et al (1973) describe a number of routine statistical problems for which improper priors lead to marginal posterior distributions which have an unBayesian property (a lack of consistency). They call this the marginalization paradox and note that it could not occur if proper priors were employed.

Improper priors can also cause problems with Bayesian model selection (see Section 2.9 of this thesis). Further, we are unlikely to have complete prior ignorance about our parameters so why 'pretend' that we do.

However, improper priors can be convenient in situtations where we have a lot of data. This is because as the amount of data increases the likelihood increasingly dominates the prior. Therefore, when we have a large amount of data the prior has little effect on the posterior inference. In these situations we may be able to use weak or improper priors in order to save the effort of specifying more informative ones. In these cases we are not purposely specifying ignorance but using it for convenience since our choice of prior will have little effect anyway.

2.3 The structure of our models

We outline here the most straightforward model structure we have used. Deviations from this basic structure are described later where applicable.

2.3.1 First stage

We have generally worked with univariate weight measurements, y_{ij}, where $i = 1, ..., I$ represents the individual and $j = 1, ..., n_i$ represents the measurement number on that individual. So, for each individual we have a vector of measurements, y_i, and a vector of times at which these measurements were taken, x_i. For each individual, i, each of these vectors, y_i and x_i, has n_i elements.

Let θ_i denote the individual-specific parameters for our chosen growth function. The number of elements, p, in the θ_i depends on the choice of growth function.

The first stage of our model is given by,

$$p_1(y_i \mid \theta_i, x_i, \Lambda) = N_{n_i}(g_1(\theta_i, x_i), h_1(\theta_i, x_i, \Lambda))$$

where $N_{n_i}(a, B)$ represents a multivariate normal distribution with dimension n_i, mean a and covariance matrix, B.

See Section 2.7 for a discussion on the possibility of using multivariate Student's t distributions instead of Normal distributions.

$g_1(\theta_i, x_i)$ is the vector of body weights given by our growth function for parameters θ_i and times x_i. Λ is a vector of variance parameters. h_1 is an n_i by n_i non-singular matrix. The simplest covariance structure is to assume common and uncorrelated variances, so, $\Lambda = \tau$ and $h_1(\theta_i, x_i, \Lambda) = \tau^{-1} I_{n_i \times n_i}$ (τ is known as the precision and is equal to the inverse of the variance).

We might expect the errors to be serially correlated because observations made closely together may have a positive association. For example, if an animal is very light one day it is also likely to be underwieght, compared to the model, the next day. As the data becomes more widely spaced in time (as our data is) this should become less apparent.

Glasbey (1979) gives five reasons why there will be residuals when fitting animal growth curves. These are:

- variations in gut fill between weighings,
- seasonal variations and changes in diet,

- illness,
- errors in measuring procedures,
- choice of wrong parametric form of curve.

Of course, several of these may be happening at the same time. The second, third and fifth of these will probably result in correlated residuals. We may wish to model this correlation, for example if it is due to illness or seasonal variation. Alternatively, it may be that apparent correlation is due to lack of fit of the curve (for example, when a line has been used instead of a curve or the wrong curve has been used).

The use of individual level parameters induces correlations at the marginal level among measurements on individuals. It may be that this induced correlation structure is sufficient to adequately model the data. It is possible to incorporate further correlation structure in the model specification, however, as discussed in Davidian and Giltinan (1995) there are problems with doing this. The results may be difficult to interpret. The amount of data available may not be sufficient. Attempts by the model to include the correlations may interfere with fitting the mean function. There may be problems with model identifiability since the correlation between observations on an individual (at the marginal level) depend on both the individual level and the between-individual covariances. Davidian and Giltinan therefore recommend that unless there are strong reasons to use them and a large amount of data per individual then models explicitly including within-individual correlation are probably not a good idea. The simpler model may be a misspecification but the loss of efficiency involved is likely to be small compared to the potential instability of the more complicated model. For example, Spiegelhalter et al (1995, Examples Vol 2, Birats example) found such models to have very poor convergence properties. Because of the points discussed above we have only used the simple structure throughout this thesis.

2.3.2 Second stage

We now model the θ_i vectors as being exchangeable conditional on μ and Σ,

$$p_2(\theta_i \mid \mu, \Sigma) = N_p(\mu, \Sigma).$$

Again, instead of the Normal distribution we can use a multivariate Student's t distribution with some degrees of freedom. See Section 2.7 for a discussion of this.

Here, μ denotes a p-dimensional vector of second stage mean parameters and Σ represents the $p \times p$ covariance matrix. A common refinement is to have different mean parameters for different groups, for example, one μ for males and one for females or one for each diet group. It is also possible to use different covariance matrices for different groups. In the interests of parsimony we have not done this here.

Exchangeability

We use the definition of exchangeable random variables given in O'Hagan (1994). For a more detailed discussion of exchangeability see that volume. Random variables, $\theta_1, \theta_2, ...$ are said to be exchangeable if for all $k = 1, 2, 3, ...$ and all $1 \leq i_1 < i_2 < ... < i_k$,

$$P(\theta_1 \leq t_1, \theta_2 \leq t_2, ..., \theta_k \leq t_k) = P(\theta_{i_1} \leq t_1, \theta_{i_2} \leq t_2, ..., \theta_{i_k} \leq t_k),$$

i.e. the joint distribution of the first k θ_is is the same as for any other k θ_is. Also, all θ_is have the same marginal distribution (the case $k = 1$).

IID random variables are clearly exchangeable. Further, if conditional on $Z = z$ the θ_is are iid with a common distribution function G_Z and Z has a distribution, $f(Z)$, then it is straightforward to show that the θ_i are exchangeable. De Finetti's theorem can be (informally) generalized to show that if $\theta_1, \theta_2, ...$ are exchangeable then they can be represented as conditionally iid. This is the converse of the above result.

When our prior beliefs are such that we have the same beliefs about any θ_i as about any other, and the same beliefs about any two groups of the same number

of θ_is then we assume that the θ_is are exchangeable random variables. This judgement in turn justifies modelling the θ_i as iid with an unknown distribution function.

We now relate the above theory to the growth curve setting. We have individuals represented by θ_i vectors. Given the second stage of our model, the common distribution function of the θ_is, we have the same beliefs about any two sets of the same number of θ_is. Conditional on the next level of our models our prior beliefs are the same for any two θ_is. These judgements mean that we can assume (conditional) exchangeability and can therefore model the θ_i as iid with unknown distribution function.

Different prior beliefs between the θ_i may be incorporated by having different second stage distributions. For instance, if some individuals are male and some are female we may use two different μ parameters.

2.3.3 Third stage

Finally, we specify prior distributions for μ, Λ and Σ. We denote these priors by $p_3(\mu, \Lambda, \Sigma \mid \Delta)$ where Δ is a vector of known hyperparameters. The separate parts of this prior definition are outlined below.

When using the simple variance structure described above (where $\Lambda = \tau$), we define a gamma prior for τ,

$$\tau \sim Ga(\frac{\nu_0}{2}, \frac{\tau_0 \nu_0}{2}).$$

Throughout, we have written Gamma distributions as $Ga(\alpha, \beta)$ using the notation of Gelman et al (1995). The mean is $\frac{\alpha}{\beta}$ and the variance $\frac{\alpha}{\beta^2}$. Therefore, for the prior given above, the mean is $\frac{1}{\tau_0}$ and the variance is $\frac{2}{\nu_0 \tau_0^2}$.

By our choices of ν_0 and τ_0 it is possible to specify a wide range of prior beliefs. Setting ν_0 and τ_0 to zero results in an improper prior for τ, but, provided the prior for Σ is proper, we still get a proper posterior.

We use a p-dimensional Normal prior for μ,

$$\mu \sim N_p(\eta, C),$$

where η and C are known hyperparameters. When there is little prior information we take C to be diagonal with large elements. We must be careful here since setting the diagonal elements of C to ∞ does not always give a proper posterior. We can however use a diagonal matrix with large elements so that it has little influence.

For the second stage covariance matrix, Σ, we use an inverse Wishart distribution,

$$\Sigma^{-1} \sim W_p((\rho R)^{-1}, \rho).$$

The Wishart distribution was used because the conditional posterior distribution for Σ^{-1} is also Wishart and sampling from it is relatively straightforward.

The mean of this distribution is R^{-1} and we choose the value of R as an approximate prior estimate of Σ. However, we note here that although the prior expectation of Σ^{-1} is R^{-1}, the expectation of Σ is not R and depends on the choice of ρ. Nevertheless, we argue here that R is a prior estimate of Σ in the sense that the full conditional for Σ^{-1} (see for example, Section 3.3.4) includes both an estimate of Σ (from the summation term $\sum_{i=1}^{I}(\theta_i - \mu)(\theta_i - \mu)'$) and R (weighted by I and ρ respectively).

We should not use an improper prior because the data do not rule out the possibility that all of the θ_i are equal to μ, in other words that there is no variability between the individual level parameters. The likelihood is proportional to

$$|\Sigma|^{-\frac{I}{2}} \exp(-\frac{1}{2}\mathrm{tr}\{\sum_{i=1}^{I}(\theta_i - \mu)(\theta_i - \mu)'\Sigma^{-1}\}).$$

As $|\Sigma|$ tends to zero, the first term of the likelihood tends to infinity. This will generally be 'cancelled out' by the second term which tends to zero since $\exp(-f)$ tends to zero as f tends to infinity. However, this will not happen if $\theta_i = \mu$ for all i. When this is the case the second term will be $\exp(0) = 1$. Therefore, unless we have a suitable prior the posterior will have a singularity at this point in the parameter space.

The prior we use is proportional to

$$|\Sigma^{-1}|^{\frac{\rho-p-1}{2}} \exp(-\frac{1}{2}\mathrm{tr}\{R^{-1}\Sigma^{-1}\}).$$

As $|\Sigma|$ tends to zero, the elements of Σ^{-1} tend to infinity and $\text{tr}(R^{-1}\Sigma^{-1})$ also tends to infinity. (R^{-1} is a fixed value, not zero). Therefore, $\exp(-\frac{1}{2}\text{tr}\{R^{-1}\Sigma^{-1}\})$ tends to zero thus 'cancelling out' the term tending to infinity in the likelihood. (Also, the $|\Sigma^{-1}|^{\frac{\rho-p-1}{2}}$ term may tend to zero depending on the value of ρ). Therefore, when using such a prior we do not have the singularity in the posterior parameter space. For our prior to be defined (and proper) we must have ρ greater than or equal to p. Taking ρ equal to p gives the least informative proper prior distribution. We also note here that with a proper but weak prior the posterior may have a significant shoulder (O'Hagan, 1985) which may lead to misleading inferences.

2.3.4 Fitting the models

Inferences from the model come from the posterior distribution which is proportional to $p_1 \times p_2 \times p_3$. We may be most interested in components of the θ_i, i.e. in individuals, or in the second stage parameters, i.e. population characteristics. Unfortunately, the integrals needed are not typically available in closed form. To get around this problem we need to use numerical or simulation techniques. We often have non-linear first stages and a large number of individuals/parameters. This is often also the case when hierarchical models are used for pharmacokinetic data as by Bennett (1996). He outlines possible numerical integration and Laplace methods which could be used for these models but concludes that these methods become impractical as the dimensionality of the posterior distribution increases.

Simulation techniques in the form of Markov chain Monte Carlo (MCMC) appear to be the easiest way to get reliable results. We must remember however that care must be taken to ensure that these methods are used properly.

2.4 Markov chain Monte Carlo methods

2.4.1 The ideas behind MCMC methods

Our aim is to generate a sample from the joint posterior distribution $p(\phi \mid y)$ but we cannot do it directly. Instead we can construct a Markov chain whose equilibrium distribution is $p(\phi \mid y)$. A number of suitable Markov chains can be constructed. Having chosen such a chain we run it until we have approximate convergence to the required distribution. We may then use the simulated values, after convergence, as if they were a sample from $p(\phi \mid y)$. These values can then be used to summarize any features of $p(\phi \mid y)$ which we are interested in.

For simulating from Bayesian posterior distributions the Metropolis-Hastings family of algorithms have been found to work well. We now discuss these algorithms and other possible approaches.

2.4.2 The Gibbs sampler

The Gibbs sampler was introduced by Geman and Geman (1984) as an algorithm to generate joint and therefore marginal distributions when we have full conditional distributions which can be sampled from straightforwardly and efficiently.

Full conditionals will always be available for hierarchical Bayesian models constructed as a sequence of conditional probability distributions. They will be standard distributions when our first stage growth funtions are linear and we take conjugate priors and hyperpriors. In these cases the full conditionals will be easy to sample from.

We divide ϕ into m scalar components or subvectors/matrices. It is often useful to group together the parameters relating to an individual or to a covariance matrix for example. This reduces the number of conditional distributions we have to sample from while increasing their dimension. Grouping, and hence sampling, highly correlated components together may help to avoid problems of slow convergence.

The first step is to take arbitrary starting values, $\phi_1^{(0)}, ..., \phi_m^{(0)}$.

We then sample $\phi_1^{(1)}$ from $p(\phi_1 \mid \phi_2^{(0)}, ..., \phi_m^{(0)})$,

and then, $\phi_2^{(1)}$ from $p(\phi_2 \mid \phi_1^{(1)}, \phi_3^{(0)}, ..., \phi_m^{(0)})$,

and so on

until, $\phi_m^{(1)}$ from $p(\phi_m \mid \phi_1^{(1)}, ..., \phi_{m-1}^{(1)})$.

This completes one iteration of the sampler. If we wish we may choose the updating order randomly at each iteration. After t repetitions of this process we have $(\phi_1^{(t)}, ..., \phi_m^{(t)})$. If t is large enough then $(\phi_1^{(t)}, ..., \phi_m^{(t)})$ can be considered to be a simulated observation from the joint distribution of ϕ and $\phi_s^{(t)}$ a simulated observation from the marginal distribution of ϕ_s. (Geman and Geman).

If all of the full conditionals are standard distributions then the sampling is straightforward. However, when our first-stage growth functions are non-linear the full conditionals for the first-stage, θ_i, parameters will not be standard distributions. In these cases we use a sampling procedure which uses Gibbs sampling for those parameters which have standard full conditionals and then a more complicated algorithm, within the general sampling, for the θ_i parameters.

We now discuss some possible methods of sampling the θ_i parameters starting with the rejection sampling and the ratio-of-uniforms methods. A third possibility is to use some form of Metropolis-Hastings algorithm. This algorithm and some of its special cases are outlined in the following subsection.

Rejection sampling

Samples are drawn from a density proportional to G, where G is an envelope function of our required unnormalized density, g (so $G(Y) \geq g(Y) \; \forall \; Y$). We then accept point Y with probability $\frac{g(Y)}{G(Y)}$. The accepted points then form an independent sample from g. (Ripley, 1987). Each decision between accepting and rejecting a point involves evaluating $g(Y)$ and $G(Y)$. Evaluating $g(Y)$ is typically computationally expensive. Also, many rejections may occur before we get an acceptance. These facts together mean that the method can be slow. Therefore, it is important to reduce the number of rejections by making the envelope

G close to g and to improve computational efficiency by making it easy to sample from and evaluate. It is possible to reduce the required number of evaluations of g by using squeezing functions $a(Y)$ and $b(Y)$, where $a(Y) \geq g(Y) \geq b(Y)$ for all Y and a and b are cheaper to evaluate than g. The accept/reject test then becomes:

sample a $U(0,1)$ random variable U;
if $U > \frac{a(Y)}{G(Y)}$ reject Y;
  else if $U \leq \frac{b(Y)}{G(Y)}$ accept Y;
    else if $U \leq \frac{g(Y)}{G(Y)}$ accept Y.

The first two tests allow a decision to be made without having to evaluate the computationally expensive g.

Generalized ratio-of-uniforms method

This method draws samples from an unnormalized univariate density, g, by generating bivariate points in the region, C:

$$C = \{(u,v) : 0 < u^{r+1} \leq g\left(\frac{v}{u^r}\right)\}, \text{ with } r > 0.$$

The ratio-of-uniforms, $\frac{v}{u^r}$, has distribution $\frac{g}{\int g}$. Further details and ways of improving the efficiency of this method are given in Wakefield et al (1991). A successful strategy is to contain C within a rectangle where the vertices are found by maximisations and minimisations of the function, g, raised to various functions of r. The only restriction on this method is that these maxima/minima exist, log-concavity of g is not required.

This method can be computationally inefficient since several maximisations or minimisations are required to obtain each sampled value, and also, there may be a large number of rejections. However, using a recommended strategy from Wakefield et al (1991), Wakefield et al (1994) found typical acceptance probabilities to be about 0.8. Wakefield et al (1991) also give a multivariate version of the basic method.

Adaptive rejection sampling

For the previous methods there is a need to find a tight envelope function or region in order to reduce the number of rejections. Gilks and Wild (1992) give a method for obtaining such an envelope function when we have log-concave univariate densities.

Gilks (1995) states that multivariate generalizations of this are possible but had not been implemented at that time. The amount of computation involved was thought to be of about order m^5, where m is the number of dimensions. If this is indeed the case then this method would only appear to be useful in low dimensions.

2.4.3 The Metropolis-Hastings algorithm

The basic idea here is to sample a potential new value, or proposal, from a chosen proposal distribution. We then decide whether or not to accept this value as our next iteration by calculating an acceptance probability.

Firstly we choose a starting point, θ^0,

then for t=1,2,...,

sample a value θ^* from the proposal (or jumping) distribution at time t,

$$J_t(\theta^* \mid \theta^{t-1}),$$

calculate, $$a = \frac{p(\theta^*|y)/J_t(\theta^*|\theta^{t-1})}{p(\theta^{t-1}|y)/J_t(\theta^{t-1}|\theta^*)},$$

(which is a ratio of importance ratios),

$$\text{set } \theta^t = \begin{cases} \theta^* & \text{with probability } \min(a,1), \\ \theta^{t-1} & \text{otherwise.} \end{cases}$$

Further details and references are given in Gilks et al (1995). Any sensible proposal distribution will give samples from the required distribution eventually. Section 2.5 considers the choice of proposal.

We note here that the Gibbs sampler is a special case of the Metropolis-Hastings where every proposed value is accepted.

2.4.4 The Metropolis algorithm

This is a special case of the Metropolis-Hastings algorithm where the proposal distribution is symmetric,

$$J_t(\theta^* \mid \theta^{t-1}) = J_t(\theta^{t-1} \mid \theta^*) \text{ for all t and } *.$$

For example, a multivariate normal distribution with constant covariance and mean the present value (θ^{t-1}).

The acceptance probability now simplifies to min(a,1) where,

$$a = \frac{p(\theta^* \mid y)}{p(\theta^{t-1} \mid y)}.$$

Random-walk Metropolis is a special case of this for which,

$$J(\theta^* \mid \theta^{t-1}) = J(\mid \theta^* - \theta^{t-1} \mid).$$

2.4.5 The independence sampler

This sampler is another special case of the Metropolis-Hastings algorithm. Here, the proposal does not depend on the present value, θ^{t-1}. For example, a multivariate normal with a fixed mean. The acceptance probability is given by min(a,1) where,

$$a = \frac{p(\theta^* \mid y)/J_t(\theta^*)}{p(\theta^{t-1} \mid y)/J_t(\theta^{t-1})}.$$

For this sampler to work well the fixed mean used should be a good estimate of the actual mean of the posterior, for example, the mode or the MLE. Also, the proposal should be a good approximation to the target distribution. It is safest to use a proposal with heavier tails than the target distribution. To illustrate why this is the case, suppose we have a proposal with lighter tails than the target distribution and that our current point, θ^{t-1}, is in the tails of the target distribution. Most proposed points will not be in the tails so $\frac{p(\theta^{t-1}|y)}{J_t(\theta^{t-1})}$ will be (possibly much) larger than $\frac{p(\theta^*|y)}{J_t(\theta^*)}$ leading to low acceptance probabilities. Therefore there is the possibility of long spells where the sampler is stuck in the tails of the target distribution. We can avoid this problem by using a proposal

with heavier tails than the target distribution. The downside of doing this is that we will have a lower overall rate of accepting proposed points.

2.5 Choice of algorithm and proposal distribution for sampling the θ_i parameters

2.5.1 Choice of sampling methods

Because of their relative ease of implementation and their computational efficiency we will use some form of Metropolis-Hastings algorithm to sample the first stage θ_i parameters. In order to implement this sampling we need to consider the exact choice of type of algorithm and also the form of the proposal distribution. For Metropolis-Hastings to work well some effort is needed to find suitable proposals. This can be difficult in some cases.

2.5.2 Points to consider

At least in theory, any (sensible) proposal distribution would give samples from our target distribution after enough iterations. However, in practice some fine tuning is needed. How well the chain mixes and therefore the rate of convergence will depend on how similar the proposal distribution is to the target distribution. When choosing a proposal distribution the questions we have to address are:

- what family of distributions to use?
- where to centre the distribution?
- what scale and shape to use?
- how many repetitions of the Metropolis-Hastings sampling to do within each overall iteration?

When considering the family of distributions to use, our requirements are that the proposal should be relatively easy to sample from and to evaluate, in other words:

- for any θ_i^{t-1}, we can easily sample from $J_t(\theta_i^* \mid \theta_i^{t-1})$,
- calculation of the acceptance probabilities is straightforward.

Multivariate Normal distributions are often used as they meet the above criteria. If necessary, the parameters may be transformed so that a Normal proposal is a sensible approximation to the target distribution. Except for the independence sampler, the distribution is often centred at the present value. We shall discuss the choice of scale and shape later.

Two or more repetitions of the Metropolis-Hastings sampling are often done for each overall iteration. This is because not all of the proposed values are accepted.

In order to consider the scale and shape of the proposal distribution we can list these further properties which a good proposal should have:

- jumps should not be rejected too often,
- the jumps should go a reasonable distance.

Essentially, these points have to be traded-off against one another. If the jumps are too small they will not be rejected very often but the chain will move too slowly. Conversely, if the jumps are too big they will be rejected too often and the chain will waste a lot of time standing still. We need to find a sensible compromise between these two extremes.

Gelman et al (1995) give some useful results which have been obtained for Normal distributions. These results were found to hold for many examples. Suppose that the posterior distribution of $\theta_i = (\theta_{i1}, ..., \theta_{id})$ is d-dimensional Normal with known covariance matrix, M, and that we use the Metropolis algorithm with:

$$J_t(\theta_i^* \mid \theta_i^{t-1}) = N(\theta_i^* \mid \theta_i^{t-1}, cM).$$

In other words, our proposal distribution has the same shape as the target distribution. The most efficient choice of scale was found to be $c \approx \frac{2.4^2}{d}$. For the multivariate Normal this choice of c was found to give acceptance rates

from around 0.44 for 1 dimension declining to about 0.23 for more than about 5 dimensions.

Gelman et al (1996) give and discuss a theoretical result supporting the above findings. Further detail of the theory involved is given in Roberts et al (1997). They obtained a weak convergence result giving the asymptotically optimal choice of scaling. Simulations for low dimension cases showed that the limiting results were appropriate approximations for as few as six dimensions. The optimal choice of scaling is related to the asymptotically optimal acceptance rate of the algorithm which is 0.23 for large dimensions. This fact can be used to try to maximise efficiency by tuning algorithms so that the acceptance rates are about 0.23, or higher for lower dimensions.

In practice, aswell as choosing the scale, we also have the not insignificant problem of estimating the covariance matrix, M. Some ways of doing this are detailed in the following section. It seems likely that when we only have an estimate of M (as in nearly all practical applications) and/or the posterior distribution is not Normal then the acceptance rates will be lower and the scale may have to be reduced, from the $\frac{2.4^2}{d}$ suggested above, in order to increase the number of acceptances.

2.5.3 Specific methods suggested in the literature

For two data sets Bennett (1996) compared various possible methods of sampling the individual level parameters, θ_i. Three of these methods are described below. Before describing them we include a brief reminder of our first stage parameters. The first stage of our model is

$$p_1(y_i \mid \theta_i, x_i, \tau) = N_{n_i}(g_1(\theta_i, x_i), \tau^{-1} I_{n_i \times n_i}).$$

Therefore the log likelihood for the i-th individual is given by:

$$\log l_i(\theta_i, \tau) = -\frac{\tau}{2} \sum_{j=1}^{n_i} [y_{ij} - g_1(\theta_i, x_{ij})]^2.$$

Each of the sampling methods described below uses the inverse information

matrix, $\hat{\Omega}_i$, which is given by:

$$\hat{\Omega}_i = -\left[\frac{\delta^2 \log l_i(\theta_i, \tau)}{\delta\theta_i \delta\theta_i'}\right]^{-1},$$

and is evaluated at the present value of τ and at the maximum likelihood estimate or some other estimate of θ_i. Using the $\hat{\Omega}_i$ matrix is only appropriate when we have enough data for the $\hat{\theta}_i$ and hence the $\hat{\Omega}_i$ to be calculated for each individual. We now describe three of the methods used by Bennett.

Random-walk Metropolis

Covariance matrix given by a constant c times the inverse information matrix, $\hat{\Omega}_i$, evaluated at the maximum likelihood estimate or some other estimate. Various different values of c were tried. The proposal was centred at the present value.

Independence Metropolis-Hastings

Covariance $\hat{\Omega}_i$, centred at maximum likelihood estimates.

MLE/prior Metropolis-Hastings

The likelihood function, $l_i(\theta_i, \tau)$ can be approximated for fixed τ by a multivariate Normal distribution for θ_i : $N(\hat{\theta}_i, \hat{\Omega}_i)$. For Normal second-stage distributions the full conditional $[\theta_i \mid .]$ is then approximately multivariate Normal with mean $\hat{\theta}_i - \hat{\Omega}_i(\hat{\Omega}_i + \Sigma)^{-1}(\hat{\theta}_i - \mu)$ and covariance $(\hat{\Omega}_i^{-1} + \Sigma^{-1})^{-1}$ where μ, Σ, τ and $\hat{\Omega}_i^{-1}$ (which depends on τ) take their present values.

Out of these methods the first was found to work best. Presumably, the other two methods might be improved if we allowed the estimated covariance matrices to be scaled by an appropriate factor, c, as well.

For another example, Bennett again used random-walk Metropolis but found that some of the elements of the $\hat{\Omega}_i$ matrices were very large. This was due to poor behaviour of the 6-dimensional likelihood surface. In this case the $\hat{\Omega}_i$ matrices were transformed to asymptotic correlation matrices (rather than covariance

matrices). This retained the main features of the likelihood surface. The value of c (0.005) was chosen to give suitable acceptance rates.

For a further example Bennett did a pilot run of 1250 iterations for each individual. For this pilot run each element of the parameter vector was sampled separately using random-walk Metropolis with arbitrary variance of 0.25. From this pilot run the first 250 iterations were discarded. The remaining 1000 iterations for each individual were used to calculate the (7-dimensional) sample covariance matrices which were then used as the covariance matrices for the main run.

2.5.4 Using mixtures of algorithms

Especially when 'bad' initial values are used, convergence is often slow due to chains becoming 'stuck'. One way to avoid this is to use a mixture of algorithms (Tierney, 1994). For example, a program for the analysis of population pharmacokinetic data (POPKAN, see Bennett, 1996) uses the following algorithm:

i) 5000 iterations using random-walk Metropolis with arbitrary fixed covariance matrix,

ii) main run, each with probability $\frac{1}{2}$,

  a) Random-walk Metropolis with covariance matrix given by initial 5000 iterations,

  b) Metropolis-Hastings with centre at the present value of μ and covariance the present value of Σ (sampling from the prior for θ_i).

2.6 Practical matters

2.6.1 Convergence diagnostics

One of the main difficulties with MCMC analyses is the problem of assessing convergence. We need to ensure that we run the simulation long enough so that

the sequence has converged. Otherwise the sample will not be representative of our required distribution.

Various methods have been proposed for assessing convergence. Unfortunately, none of these is foolproof (Cowles and Carlin, 1995). The best approach appears to be to use some convergence diagnostics and also to inspect plots and/or summary statistics of the sampled values. By doing this we can make a fairly confident assessment of convergence whilst bearing in mind that we cannot be certain that convergence has been achieved. The CODA program (Best et al, 1995) is a collection of S-functions which can be used for this process. We now outline two of the most popular convergence diagnostics.

Gelman and Rubin (1992)

The idea behind this method is to take 2 or more independently simulated sequences with widely varying starting values and then to compare the variation between and within the simulated sequences. If the sequences have mixed then the 'between' variation and the 'within' variation should be roughly equal. Each scalar quantity of interest is monitored separately and any all-positive parameters are transformed by taking logs. For each quantity we discard the first half of the iterations and then compute the between and within sequence variances for the remaining iterations. These enable us to calculate the potential scale reduction, $\sqrt{\hat{R}}$. This is a measure of the factor by which the scale of the current distribution for the quantity might be reduced if the simulations were continued in the limit n tends to infinity. This quantity reduces to 1 as n tends to infinity. If it has a large value then we should increase the number of iterations. Gelman et al (1995) consider that for most examples values under 1.2 can be considered to be suitably small. They note that for expensive datasets and final analyses smaller values, such as less then 1.1, might be sensible.

We note here that having sequences with widely varying starting values may not be suitable for those models where the MCMC approach is found to work poorly unless 'good' starting values are used. For example, for models where

convergence is very slow it will be more efficient to have one long run rather than several shorter ones.

Raftery and Lewis (1992)

This is a method for single chains. Its aims are to try and detect convergence to the stationary distribution and also to estimate how long the chain should be run for in order to give the required accuracy in the estimates of the quantities of interest. To use this method you must specify three quantities. These are the quantile of interest, for example 0.025, the accuracy with which you wish to estimate this quantity, for example $\pm$ 0.0125, and the probability of the estimate being within this level of accuracy, for example 0.95. Given a pilot run of at least N_{min} (depending on the values chosen, 600 for those suggested above) the method estimates how many iterations should be discarded and how many iterations should be done in order to achieve the required accuracy. The method also recommends a value k for thinning the chain by keeping only the kth iterations if there are high correlations between successive iterations.

This method generally works well. However, Wakefield (1993) found that it can be unreliable for non-well-behaved posteriors or bad parameterizations. He gave an example of this which we briefly describe here as it is also of interest on the question of parameterization (see below). He considered two possible parameterizations one of which had high posterior correlations between two parameters, for the other the correlation was much lower. In each case 10 runs of 1000 iterations were done and Raftery and Lewis statistics calculated for each repetition. The average recommended number of iterations for the 'good' parameterization was about half that for the other parameterization. The unreliability of the recommendations was suggested by the large amount of variability between values for the 10 repeats for the 'bad' parameterization (between 5803 and 29744). In contrast, for the other parameterization all of the recommendations were between 4289 and 6165 iterations.

2.6.2 Other practical points

Burn-in

Burn-in is the name given to the early iterations before convergence has been attained. These iterations must be discarded before making any inferences from the sample.

Reparameterization

As discussed in Hills and Smith (1992) the parameterizations used in the prior and in the likelihood will affect the accuracy and efficiency of Markov chain sampling methods. When sampling each component separately using Gibbs sampling it is desirable to have low correlations between the parameters as this generally improves mixing/convergence. For a simple example, Hills and Smith showed that really high correlations greatly slow down the convergence of the Gibbs sampler. By reducing this correlation, even only to 0.8, convergence was much faster with even bad starting values quickly forgotten. Using a suitable parameterization is one way to reduce the correlations. For example, when fitting straight lines it is common to centre the age values by subtracting the mean age from each value. Doing this means that the intercept is the weight at the mean age rather than at age/time zero. This parameter may be less correlated with the slope parameter. For non-linear models the situation is less clear cut. Ross (1990) gives some suggestions on stable parameterizations. By stable he means that the parameters should represent contrasting features of the data and so be less likely to be correlated. When using Metropolis-Hastings algorithms an alternative to reparameterization is to use a well chosen proposal distribution which has the same correlation structure as the target distribution. Unfortunately, this is not always straightforward.

Gelfand et al (1995,1996) propose the use of hierarchical centering reparameterizations in order to reduce correlations and so improve the efficiency of the MCMC sampling. They consider normal linear mixed effects models and non-normal generalized linear mixed models. For example, for a two-stage normal

linear model,

$$Y_i \mid \eta_i \sim N(X_i\eta_i, \tau I_{n_i}),$$

$$\eta_i \mid \mu \sim N(\mu, D),$$

$(\mu, \eta_1, ..., \eta_n)$ is the centered parameterization whereas $(\mu, \alpha_1, ..., \alpha_n)$ with $\alpha_i = \eta_i - \mu$ is the uncentered parameterization. (The α_i have mean zero and the μ term needs to be incorporated into the first-stage of the hierarchy).

Gelfand et al show using analytical arguments, simulations and examples that the centering is likely to improve efficieny in cases where the variance of the random effects is large relative to the first-stage or error variance. This is likely to be the case for models where random effects are used. Our basic model structure (see earlier in this chapter) is defined in a centered form.

This centering is not intended to reduce correlations between the parameters for an individual. However, these parameters are often sampled together as vectors thus reducing the problems caused by high correlations. Also, it is manageable to use transformations within these relatively low dimensional blocks in order to reduce the correlations.

Number of chains

There is ongoing discussion about whether it is better to run one very long chain or several shorter ones. In some situations one very long chain can fail to show a lack of convergence which would have become apparent if several chains had been run with an overdispersed distribution of starting values. On the other hand, the creation of such a distribution can be time consuming. The decision of whether this extra work is necessary or not will depend on the particular application.

Starting values

The choice of starting values is not crucial as the chain will move away from them as it converges. However, by taking some care over the choice we can

reduce the time taken to convergence. In some cases convergence can be very slow if 'good' starting values are not used.

Blocking

When generating highly correlated components individually we may find that movement and therefore convergence is painfully slow. Blocking these components together into a vector or matrix may improve mixing. However, the downside of this is that generating from multivariate distributions may be harder. For instance, for the Metropolis-Hastings algorithm, the acceptance rate often decreases as the dimension increases.

Updating order

The order in which we update, or sample from, the various components may be fixed or random. Also, we do not need to update all components at each iteration. One possibility is to update only one component, i, at each iteration choosing it with some probability, $s(i)$. Zeger and Karim (1991) suggest that mixing may by improved by taking high probabilities, $s(i)$, for highly correlated components. We note here that if the $s(i)$ depend on the present state of the chain then the formula for the acceptance probabilities must be modified.

Thinning

If storage space is limited and a very long run is required due to high correlations between successive iterations then we may wish to thin the chain by only saving every kth iteration ($k > 1$).

2.7 Heavy tailed distributions

It is possible to replace normal distributions in our models with heavier tailed distributions. The most commonly used of these are Student's-t distributions and they are principally used to make inferences more robust. By using a Student-t

distribution at the first stage of our models we can make our models more robust to data outliers. With a Student-t distribution at the second stage the model becomes more robust to the presence of outlying individuals.

In classical analyses we often wish to detect outliers in order to omit them from the data. Bayesian models using Student-t distributions are better because they can automatically accomodate outliers and decrease the contribution they make to the posterior distribution.

Wakefield et al (1994) use a Student-t distribution at the second stage of one of their examples. They wished to guard against outlying individuals unduly influencing the population mean inferences. They used scale mixtres of normals to implement this. Gibbs sampling could still be used and the introduction of an extra parameter for each individual provided a diagnostic for second stage outliers. There is no need to remove these outliers as might be done in classical analysis since their influence on the model is automatically downweighted. However, in order to compare the inferences produced, Wakefield et al did remove the two outlying individuals they found and reanalysed the reduced data (using a normal model). As expected they found that the parameter estimates from the Student-t model lay between those for the normal model with all the data and those for the normal model with the reduced data set.

One representation of a multivariate t distribution is as a multivariate normal with unknown variance. This is the situation we have in our models since the variances (or variance matrices) are themselves unknown parameters of the models. However, this does not make the models robust in any real sense unless they have different variance parameters for different individuals/observations (which ours did not).

We illustrate this with a simple example, we have a single measurement, y, on a number of individuals (indexed by i). In the common variance case we may model these measurements as follows,

$$y_i \sim N(\mu, \sigma^2), \sigma^2 \sim IG(...).$$

With this structure, the y_i have a t-distribution marginally but not indepen-

dently. The estimate of μ would simply be $\bar{y}$.

By using different variances for the different individuals we introduce robustness to the model. We now have,

$$y_i \sim N(\mu, \sigma_i^2), \sigma_i^2 \sim IG(...).$$

The estimate of μ will now be a weighted average of the y_i. Outlying y_i will have large σ_i^2. The effect of this outlying y_i on the estimate of μ will therefore be reduced since the weighting factors decrease as σ_i^2 increases.

To summarize, although we do have t-distributions in a sense, this does not make our (common variance) models robust to outliers.

Outliers in the data are an example of conflict between sources of information. The outlier(s) conflict with the remaining observations. We can also have a conflict between the data and our prior beliefs. Our posterior beliefs are a compromise between the prior and the likelihood. When the difference between the two is large compared to both standard deviations a normal model still gives posterior estimates which are compromises. By using a heaver tailed prior distribution the prior beliefs will be rejected in favour of the data. when the prior beliefs are more reliable we may wish to compromise with the data or we could even reject the data in favour of the prior. In general, the information source with the lighter tailed distribution will be preferred to that with the heavier tailed distribution. Therefore, we have a choice of how such conflicts should be resolved. O'Hagan (1988 and 1995) discusses these issues in more detail.

2.8 Bayesian model checking

Perhaps the simplest approach to model checking is to inspect the residuals. However, as discussed in Pettit (1986) there is no single definition of residuals in Bayesian analyses. We can estimate the residuals by calculating the observed minus the fitted values where the fitted values are obtained from some estimates of the model parameters. The problem then arises of which estimates to use, for example, the mean or the mode of the posterior distribution of the parameters.

Also, this is not necessarily the best method because for non-linear functions the expectation of the whole function is generally not the same as the function of the expectations. In the spirit of using all of the available information, Chaloner and Brant (1988) and Chaloner (1991) discuss the use of the whole posterior distribution of the error or residual terms. They propose the use of such distributions to define outliers and to calculate posterior probabilities of observations being outliers. Also, they advocate the use of residual plots showing both the point estimate and 95% highest posterior density intervals of the residuals. We can easily use these ideas since it is straightforward to calculate the residuals at each iteration of the MCMC process thus obtaining the required posterior distributions.

Many other approaches to model checking are possible, Hodges (1998) gives references to many of these and then proposes a new approach which takes full advantage of the structure of hierarchical models and provides diagnostics for all parts of the models. The basic idea is to add artificial 'cases' to the data corresponding to the higher levels of the hierarchy. This enables the model to be expressed in the form of ordinary linear models.

2.9 Bayesian model selection

Given a collection of models (which may all be adequate) the question often arises as to which is best. Classical methods of answering this question include using likelihhod ratio tests. However, these involve approximations which may not be accurate and they are only applicable for comparing nested models. This restriction does not apply to Bayesian methods, of which a common approach is the use of Bayes factors.

Inference from our models proceeds from the posterior $f(\theta \mid Y)$. However, since different models will have different parameters, θ, we cannot use this posterior for model choice. Instead, we can use $f(Y)$, a density which can be compared with the actual observations. This distribution is the marginal distribution for

the data Y. For model i with parameters θ_i it is given by,

$$f_i(Y) = \int f_i(Y \mid \theta_i) f_i(\theta_i) d\theta_i.$$

The Bayes factor for model 1 against model 2 is then $\frac{f_1(Y)}{f_2(Y)}$. We may specify our prior probability that model 1 is correct, giving prior odds on model 1 of $\frac{p}{1-p}$. Then the posterior odds on model 1 are given by the prior odds multiplied by the Bayes factor. When the Bayes factor is greater than 1 the probability for model 1 is increased from our prior beliefs. This happens, generally, when model 1 fits the data better than model 2.

Unfortunately, there can be problems with the use of Bayes factors, in particular when we have improper priors (since $f(Y)$ will also be improper). We cannot then assess when a Bayes factor is large since any multiple of it can be arbitrarily obtained since $c.f(\theta)$ has the same prior information as $f(\theta)$ when $f(\theta)$ is improper. Two ways of getting around these problems are to use the device of imaginary observations and to use partial Bayes factors. See O'Hagan (1994) for further details of the problems which can arise and also of these possible solutions.

The Bayes factor uses $f(Y)$ which we can call the prior predictive density, it is also possible to use other predictive densities. For example, cross-validation predictive densities, $f(y_r \mid Y_{(r)})$, where $Y_{(r)}$ represents all of the data, Y, except y_r. These are usually proper densities even if $f(Y)$ is not. The quantity

$$f(y_{r,obs} \mid Y_{(r),obs})$$

is the conditional predictive ordinate (CPO) (Pettit and Young, 1990). Small values of the CPO suggest that the value $y_{r,obs}$ does not support the model. We may compare two or more models by plotting the CPO values against the observation number, r, for each model on the same axes. This will show us which models do better by seeing which tend to have the highest values. We will also be able to see which models are similar and whether some points are poorly fit under all models and so on. We may also do a global comparison of two models using the pseudo-Bayes factor. This is the product over all of the

observations of the CPO ratio,

$$C_r = \frac{f(y_{r,obs} \mid Y_{(r),obs}, M_1)}{f(y_{r,obs} \mid Y_{(r),obs}, M_2)}.$$

Gelfand (1995) discusses the computational issues involved and gives a (growth curve) example using these ideas. Also, see Gelfand et al (1992) for further discussion and another (growth curve) example.

A further point worth noting here is that when our main purpose is not model selection but, for example, estimating parameters, then selecting just one model may not be the best approach. This is because model uncertainty is ignored and therefore the uncertainty involved in the estimates of the parameters of interest will be underestimated. See Raftery (1995) for more information. He also discusses methods of computing marginal likelihoods in order to be able to calculate Bayes factors.

2.10 Bayesian non-parametrics

It is possible to use a non-parametric approach at the second-stage of hierarchical models. Suitable methods were proposed by Escobar and West (1992). They modelled the distribution of the individual level parameters as arising from the class of distributions given by the Dirichlet process. Because of the flexibility afforded by this approach multimodality or skewness of the population distribution can be handled by the models. Wakefield and Walker (1994) extended this approach to the nonlinear hierarchy. The downside of the extra flexibility is that the implementation of MCMC methods becomes more complex. We have not used this approach here.

2.11 Classical approaches

There are a number of possible classical approaches to nonlinear population modelling. These include two-stage methods and methods based on linearization. For two-stage methods there needs to be enough data for each individual

to be able to estimate the parameters for each individual separately. These estimates are then used to make inferences about the population parameters. If we have normality at the second-stage then it is possible to incorporate the uncertainty involved in the estimation of the individual level parameters. Iterative procedures can be used to obtain estimates from the model, convergence to sensible estimates occurs for most cases but can be slow. The two-stage approach was used in a growth curve context by Berkey (1982).

Linearization methods approximate the model with a form which is additive in the random effects and individual errors. Estimation methods similar to those for the linear case can then be used. Of course, this depends on the validity of the approximation. This type of approach was used by Lindstrom and Bates (1990) and by the NONMEM package (Beal and Sheiner, 1989) which was developed for pharmacokinetic models. For further references and a comparison of the NONMEM package with the Bayesian approach using MCMC see Bennett (1995).

Other classical approaches which have been proposed include non-parametric and semi-parametric methods. Davidian and Giltinan (1995) give details of these and also further details on the methods discussed above.

Another possible approach is that of multilevel modelling (Goldstein, 1995) and the MLwiN software (Goldstein et al, 1998). MLwiN can be used to fit a variety of multilevel models including repeated measures models and multilevel logit, loglinear and time series models. An iterative generalized least squares (IGLS) algorithm is used and gives consistent estimates of the model parameters, and maximum likelihood estimates when normal assumptions are met.

Goldstein et al (1994) illustrate the use of one of the previous versions of this package (ML3, Prosser et al, 1991) for a repeated measures model with autocorrelated level one residuals. They assume multivariate normality and use the IGLS algorithm. The authors noted that care was needed choosing the starting values for the iterative procedure and that for moderate sized data sets there could be numerical convergence problems due to the relative flatness of

the likelihood surface.

Goldstein (1991) discusses non-linear multilevel models and proposes a procedure using linearization. In particular, loglinear models are studied. The estimation proceeds by first linearizing the nonlinear function and then using a standard procedure for the now, linear multilevel model. In this case, the IGLS algorithm was again used. The linear first-order terms of the Taylor expansion were used and details are given of how the package may be used to consider the adequacy of this first-order approximation. The author noted that further work was needed to consider the properties of the estimates of the random parameters. The current method used a weight matrix which was based upon assumptions of multivariate normality.

Some statistical packages can now be used to fit mixed models. For example, S-Plus (Version 4.5, 1998) can fit both linear and non-linear mixed effects models. Structured covariance matrices can be used. Either maximum likelihood or restricted maximum likelihood is used. Derivatives to be used in the optimization may be supplied by the user or the program can use numerical derivatives.

The Bayesian method using MCMC is better than these classical approaches in the sense that no numerical or analytical approximations are needed. Inaccuracies can arise through incorrect assessment of convergence or through Monte Carlo variability. We have some control over these through our choice of the number of iterations to be done. The other downside is the increased computing time over the classical approaches. The other benefits of the MCMC approach are discussed elsewhere in this thesis, for example the ability to include prior information and the ease with which confidence intervals can be obtained for any complicated (non-linear) function of the parameter(s) we are interested in.

Chapter 3

Linear Growth

3.1 Introduction

This chapter illustrates the use of Bayesian hierarchical models for data which was collected on pigs over the approximately linear portion of growth. For this example, all of the full conditionals are standard distributions and we have used Gibbs sampling to obtain results from the models. For the purposes of comparison we also include some analyses of this data using other, more traditional, methods.

3.2 The data

There are 512 pigs which were weighed at 5 times from about 30kg to about 95kg live weight. For this weight range the growth is approximately linear so we have used straight lines instead of curved growth functions. The pigs were at different (and unknown) ages at the start of the trial. The weighing interval was usually, but not always, 14 days. The pigs were housed in 64 pens in groups of 8. Individual weights were not available, only averages for each pen, therefore, we used pens as our individual units.

Half of the pigs were gilts (female) and the other half were boars (male). For each sex two different diets were used, a control and a new diet. The new diet is

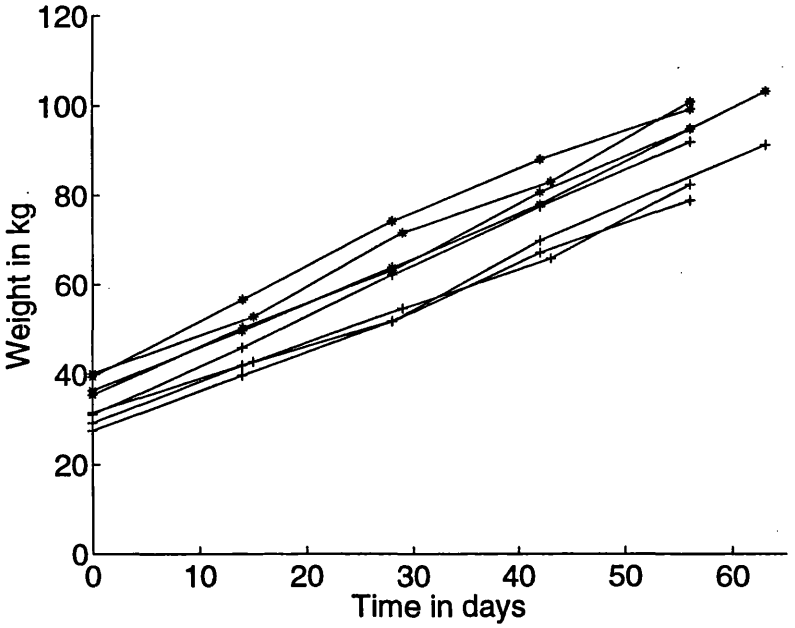


Figure 3.1: The light (+) and heavy (*) pens for boars fed the control diet

cheaper and we are interested in whether or not it is as effective as the control diet. The new diet is different between the two sexes.

For each of the four diet × sex groups there were 16 pens. These 16 pens consisted of four replicates of four weight groups:

1. Heavy
2. Medium heavy
3. Medium light
4. Light

At the start of the trial the weights ranged from about 25 to 41kg. When each batch of pigs started the trial they were allocated to the weight groups simply by dividing them into four groups based on their present weight. The purpose of using these groups was to ensure that the distribution of starting weights was roughly the same for each diet group.

As an illustration of some of the data, Figure 3.1 shows the measurements for the light and heavy pens for boars fed the control diet.

3.3 The models

Firstly, we analysed the data for each weight group separately using the same model in each case. For each weight group there were 16 pens, these were made up of four replicates of each of the four diet × sex groups. After these separate analyses we also discuss a model which includes all of the weight groups together.

3.3.1 First level

$$p_1(y_i \mid \theta_i, x_i, \tau) = N_5(g(\theta_i, x_i), \tau^{-1} I_{5\times 5})$$

where y_i are the vectors of the five weights for each of pens $i = 1 : I(I = 16)$, x_i are vectors of the times these measurements were made. The θ_i vectors consist of the intercept and slope parameters for each pen.

$$g(\theta_i, x_i) = \theta_{i1} + \theta_{i2} x_i = X_i \theta_i$$

where X_i is a 5 by 2 matrix with first column made up of ones and second column equal to x_i, the vector of measurement times.

3.3.2 Second level

We used different means but the same covariance matrix for the θ_i parameters for the four sex × diet combinations:

$$p_2(\theta_i \mid \mu_k, \Sigma) = N_2(\mu_k, \Sigma)$$

where $k = 1, 2, 3, 4$ represents the four sex × diet combinations.

3.3.3 Third level/priors

The structure of the third level is as outlined in Section 2.3.3. ν_0 has been set to zero giving an improper prior for τ. We have also used improper priors for the μ parameters, by making all elements of C^{-1} zero. The prior for Σ^{-1} is the

least informative inverse Wishart distribution that we can use, $\rho = 2 =$ number of parameters in the θ_i.

The pigs chosen to start the trial weighed from about 25kg to about 41kg, a range of 16kg. They were then divided into four weight groups. Therefore, we would expect a rough range of 4kg in the initial weights, or intercepts, within each weight group. We use this value to make a very rough estimate of the standard deviation as 2kg and therefore, the variance as 4kg^2. For an approximately central starting weight of 33kg this variance equates to a coefficient of variation of about 6%.

Consultation with Steve Jagger of Dalgety Feed Limited, where the trial was done, revealed that they generally use the assumption that the coefficient of variation for the slopes, or growth rates, is about 5%. For a typical slope of 1kg/day this corresponds to a variance of 0.0025.

We don't have a strong prior belief about the correlation between the slope and intercept parameters. This is because we might expect pigs which are larger at the start to grow quickly during the trial because they are fast growing pigs. On the other hand, remembering that the ages at the start are unknown, they may be heavier because they are slightly older. Alternatively, we might expect some element of regression to the mean. Therefore, we used zero as our prior estimate of the correlation.

To summarize the above, R, the approximate prior estimate of Σ is a matrix with diagonal elements 4 and 0.0025 and zero off-diagonal elements.

Subsection 3.6 considers the effect of our choice of R.

3.3.4 The full conditionals

Because we are using linear growth functions in this example, the full conditionals are all standard distributions and we may use Gibbs sampling for all of the parameters.

$$[\theta_i \mid y, \mu, \Sigma^{-1}, \tau, \theta_j, j \neq i] = N[\theta_i \mid D_i(\tau X_i' y_i + \Sigma^{-1}\mu_k), D_i],$$

where $D_i^{-1} = \tau X_i' X_i + \Sigma^{-1}$ and the appropriate μ_k parameter is used for each i.

$$[\mu_k \mid y, \theta, \Sigma^{-1}, \tau, \{\mu_m, m \neq k\}] = N[\mu_k \mid V(I_k \Sigma^{-1}\bar{\theta} + C^{-1}\eta), V],$$

for $k = 1 : 4$ where $\bar{\theta}$ represents the mean θ for the relevant pens and I_k is the number of relevant pens (4). The matrix V is defined by $V^{-1} = I_k\Sigma^{-1} + C^{-1}$.

$$[\Sigma^{-1} \mid y, \theta, \mu, \tau] = W[\Sigma^{-1} \mid \{\sum_{i=1}^{I}(\theta_i - \mu_k)(\theta_i - \mu_k)' + \rho R\}^{-1}, I + \rho],$$

where k is the appropriate value for each i.

$$[\tau \mid y, \theta, \mu, \Sigma^{-1}] = Ga[\tau \mid \frac{1}{2}(\nu_0 + n), \frac{1}{2}\{\sum_{i=1}^{I}(y_i - X_i\theta_i)'(y_i - X_i\theta_i) + \nu_0\tau_0\}],$$

where n is the total number of measurements. The full conditionals for the μ_k and for τ simplify somewhat because of our choice of priors, $C^{-1} = \nu_0 = 0$.

3.4 Results from the Gibbs sampling

3.4.1 Convergence

We sampled from the conditional distributions in the order given above and ran five sequences with different starting values. The values used were widely dispersed about the approximate expected posterior distribution. This procedure was repeated for each of the four weight groups.

Gelman and Rubin (1992) statistics were calculated for each of the relevant parameters or scalar components. They were found to reach values close to 1 very quickly. For each weight group five sequences each of 350 iterations were done. After 200 iterations all of the Gelman and Rubin statistics were less than 1.12. The sequences were also inspected graphically to check for any signs of a lack of convergence. Since the Gelman and Rubin method discards the first

half of the sequences before beginning the calculations we discarded the first 100 iterations (half of 200) from each sequence leaving 1250 iterations in total for each weight group.

3.4.2 Results

Figures 3.2 to 3.5 show boxplots of the sampled values of the μ parameters, i.e. the second stage means, for each weight group. The boxplots have lines at the median and the upper and lower quartiles. The rest of the data is shown by 'whiskers'. Outliers are shown by crosses and defined for the boxplots as values beyond the length of the 'whiskers'. This length is defined to be 1.5 times the interquartile range.

For each figure the slope and intercept parameters are shown separately. The four columns in each plot represent the four sex $\times$ diet combinations in the following order:

Female	New
Female	Control
Male	New
Male	Control

No clear pattern emerges from these boxplots.

Our interest centres on the differences, if any, in the population growth rates, or slopes, between the diets. Therefore, for each weight group and sex we obtained a sample of the population slope parameter for the control diet minus the population slope parameter for the new diet. In each case this was done by calculating the 1250 values of $slope_{(control)} - slope_{(new)}$. Figures 3.6 to 3.9 show histograms of these samples. If there was a large difference in the slopes between the diets then zero would fall in the tails of these histograms. The sampled values would be mostly positive if the control diet has larger slopes (growth rates) then the new diet and vice versa. Table 3.1 gives the mean difference in the population slope parameters in each case and also the standard

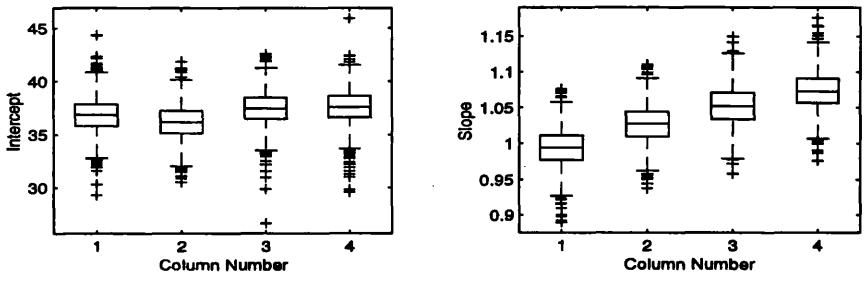


Figure 3.2: Boxplots of the μ for weight group 1

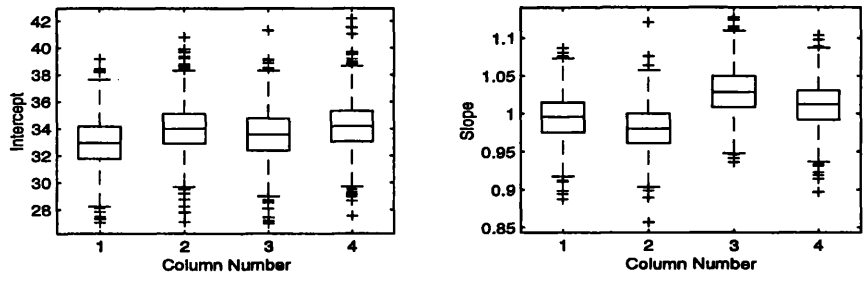


Figure 3.3: Boxplots of the μ for weight group 2

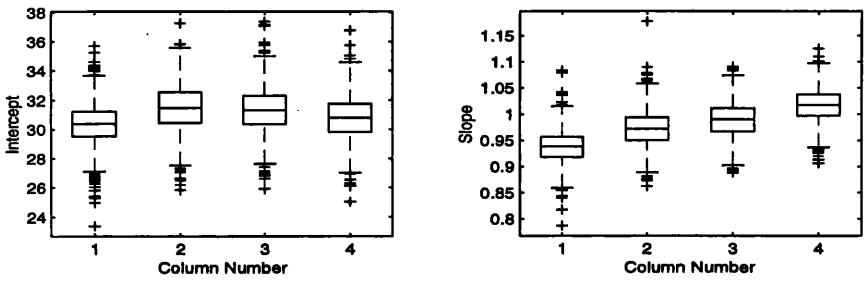


Figure 3.4: Boxplots of the μ for weight group 3

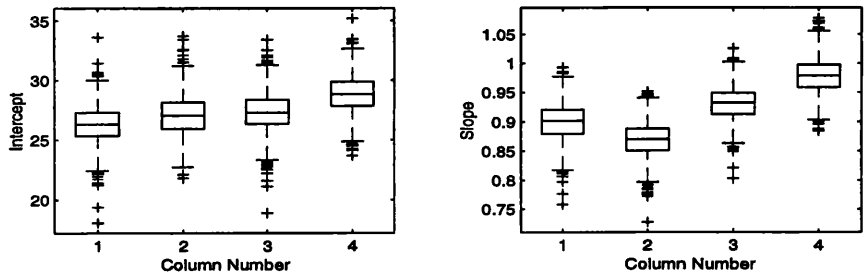


Figure 3.5: Boxplots of the μ for weight group 4

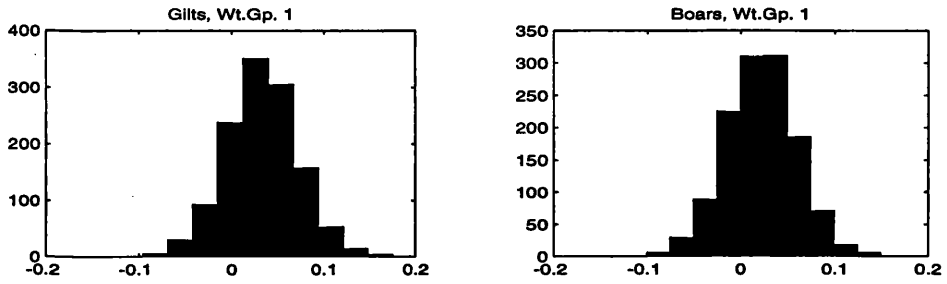


Figure 3.6: Histograms of the samples of the difference in population mean slopes (control diet - new diet) for weight group 1

deviation of the differences. Three of the eight means are negative and they are all close to zero providing little evidence that the new diet is worse than the control.

Figure 3.10 shows the fitted lines for the first two gilts in weight group 1 which were fed the new diet.

Figures 3.11 to 3.14 show the standardised residuals for each of the four weight groups. There is some evidence of curvature in the residuals. This curvature is in the direction we would expect given that we know that the growth will form an S-shaped curve over a longer time period.

Tables 3.2 to 3.5 give medians and 90% sample intervals for τ^{-1} and the elements of Σ for each weight group. We note here that the 90% intervals given here and throughout were found by omitting the lowest 5% and the highest 5% of the sample (equal tails intervals).

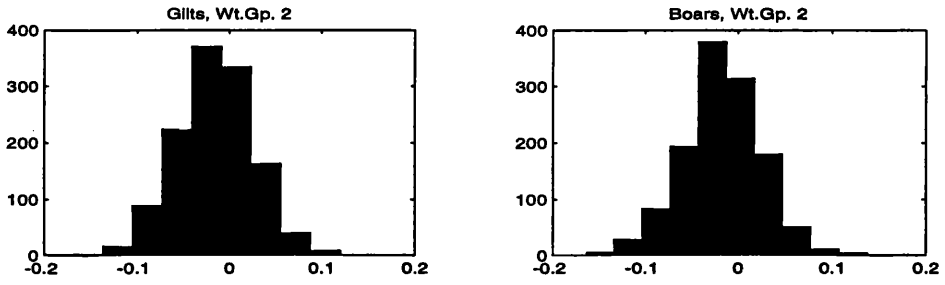


Figure 3.7: Histograms of the samples of the difference in population mean slopes (control diet - new diet) for weight group 2

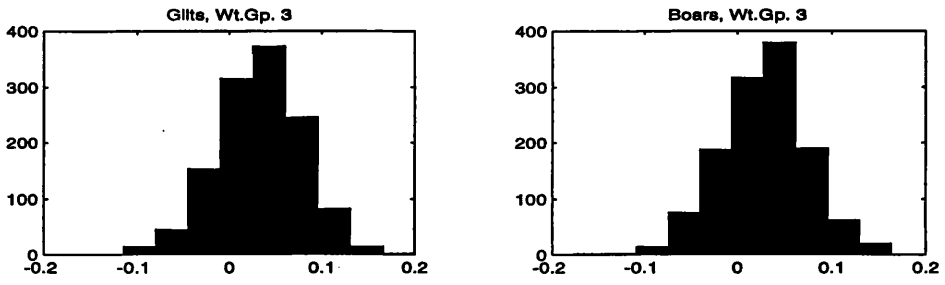


Figure 3.8: Histograms of the samples of the difference in population mean slopes (control diet - new diet) for weight group 3

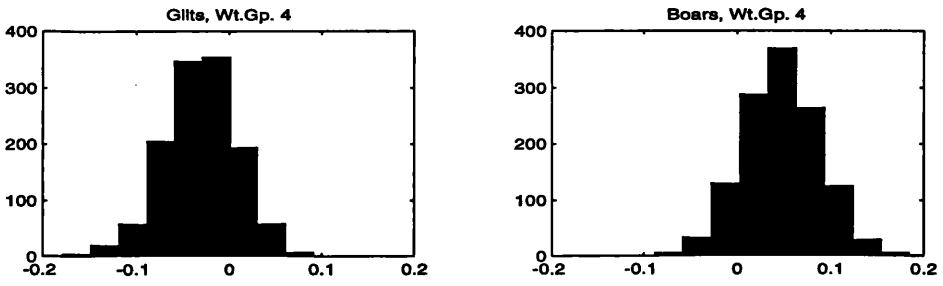


Figure 3.9: Histograms of the samples of the difference in population mean slopes (control diet - new diet) for weight group 4

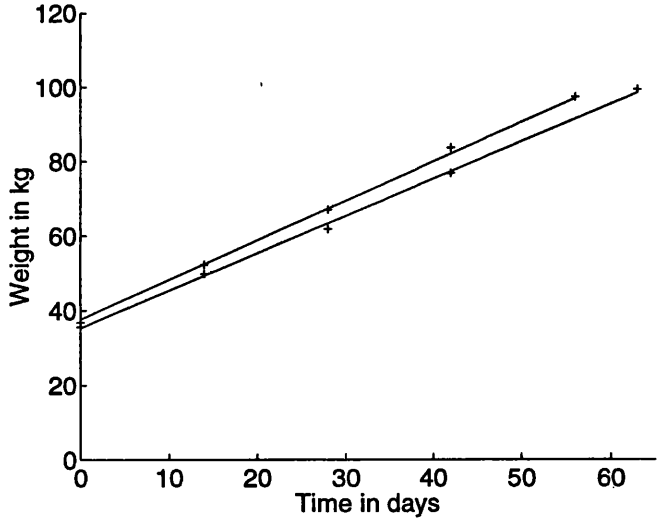


Figure 3.10: Observed and fitted values for two pigs in weight group 1

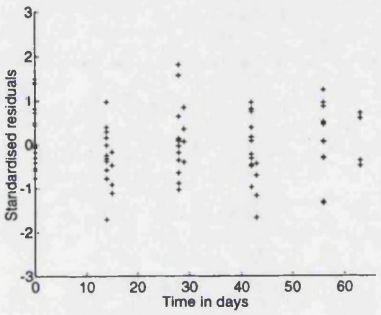


Figure 3.11: Standardised residuals for weight group 1

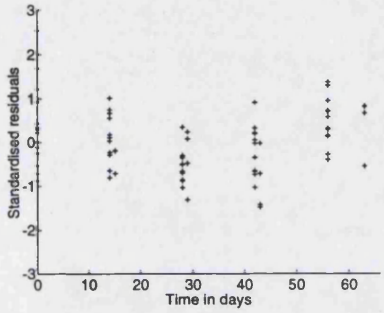


Figure 3.12: Standardised residuals for weight group 2

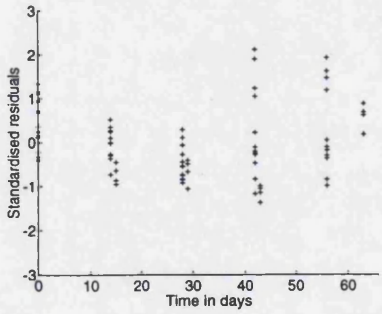


Figure 3.13: Standardised residuals for weight group 3

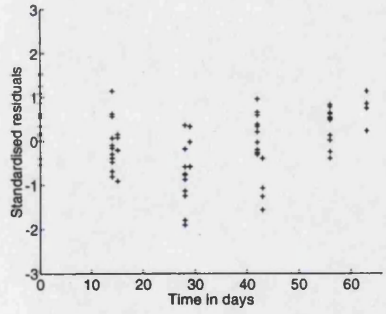


Figure 3.14: Standardised residuals for weight group 4

Weight Group	Sex	Mean	Std.Dev.
1	G	.033	.039
1	B	.022	.037
2	G	-.015	.041
2	B	-.018	.042
3	G	.034	.046
3	B	.028	.046
4	G	-.031	.040
4	B	.047	.041

Table 3.1: Means and standard deviations of the samples of the difference in population mean slopes (control diet - new diet)

Weight Group	Median	90%	Interval
1	3.59	(2.46,	5.24)
2	3.04	(2.08,	4.62)
3	5.26	(3.86,	7.54)
4	3.90	(2.75,	5.67)

Table 3.2: Medians and 90% intervals for τ^{-1}

Weight Group	Median	90%	Interval
1	6.7	(3.1,	15.8)
2	9.2	(4.6,	20.4)
3	4.5	(1.9,	10.8)
4	6.9	(3.1,	17.3)

Table 3.3: Medians and 90% intervals for $\Sigma(1,1)$

Weight Group	Median	90% Interval
1	.037	(-.011, .112)
2	.085	(.016, .204)
3	.040	(-.002, .118)
4	.050	(-.004, .133)

Table 3.4: Medians and 90% intervals for $\Sigma(1,2)$

Weight Group	Median	90% Interval
1	.0016	(.0006, .0037)
2	.0026	(.0012, .0054)
3	.0021	(.0008, .0050)
4	.0019	(.0008, .0048)

Table 3.5: Medians and 90% intervals for $\Sigma(2,2)$

3.5 All-inclusive model

As well as comparing the diets separately for each weight group we also fitted an all-inclusive model allowing an overall comparison between the diets to be made. For this model there is one second level mean, μ_k, for each weight group. Any differences between the diets or between the sexes are accounted for by including ρ and δ parameters as shown below (for one weight group, k):

Diet	Gilts	Boars
Control	μ_k	$\mu_k + \rho$
New	$\mu_k + \delta$	$\mu_k + \rho + \delta$

The models including each weight group separately made few assumptions but had many parameters. The all-inclusive model, on the other hand, has fewer parameters but more assumptions. This simpler model has more chance of finding a small difference between the diets because it pools all of the information together and has fewer parameters.

Non-informative Normal priors were used for ρ and δ with the other priors the same as before. Gibbs sampling was again used. Five cycles of 350 iterations with different starting values were done. For each cycle the first 100 iterations were discarded as all values of the Gelman and Rubin statistic were suitably small by this time.

3.5.1 Results from the all-inclusive model

Figure 3.15 shows histograms of the sampled values of the slope and intercept components of the δ and ρ parameters. We can see that zero is close to the centre of the distributions for both components of δ meaning that we have very little evidence of a difference between the diets. In each case the median is negative, this equates to the parameters being lower for the new diet. The distributions for the components of ρ are further from zero, especially for the slope component. For these parameters the medians are positive, equating to larger values for boars than gilts. Table 3.6 gives the medians, 90% ranges and

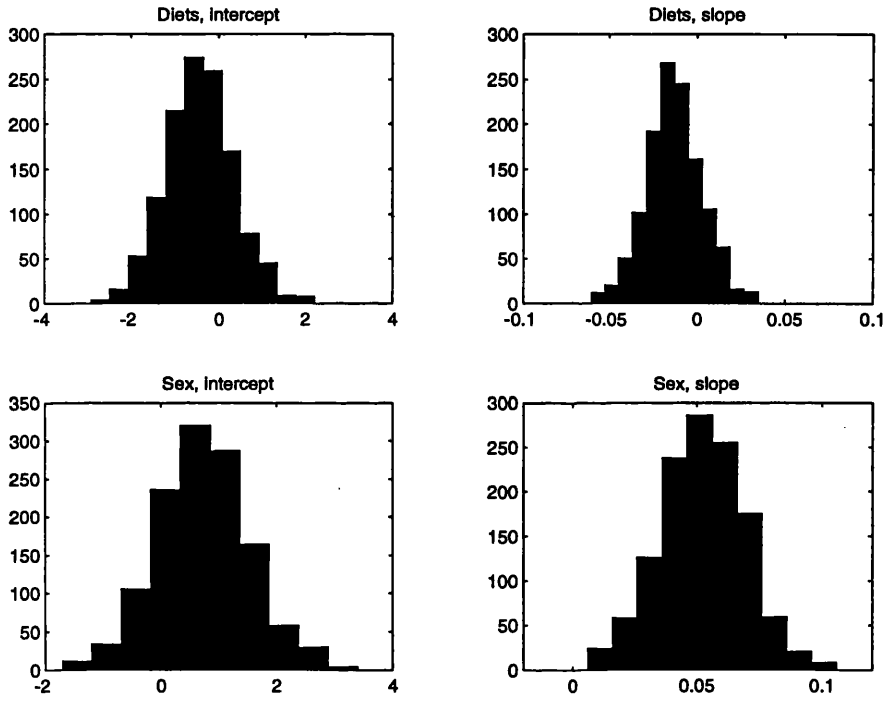


Figure 3.15: Sampled values of the δ (diet) and ρ (sex) parameters

Element	Median	90% Interval	p(element)< 0
δ_{int}	-.43	(-1.71, 0.92)	0.72
δ_{slo}	-.013	(-0.040, 0.014)	0.80
ρ_{int}	.71	(-0.51, 2.06)	0.17
ρ_{slo}	.052	(0.024, 0.080)	0.00

Table 3.6: Medians, 90% intervals and posterior probabilities of parameters being negative for δ and ρ

posterior probabilities of the parameters being negative for the four elements of these vectors. The main hypothesis we are interested in is whether the slopes, or growth rates, are lower for the new diet, i.e. is $\delta_{slo} < 0$. From Table 3.6 we see that the posterior probability of this is 0.8.

Because the new diet is cheaper the users of the feed may be willing to accept slightly lower growth rates because overall they would still save money. An advantage of our approach is that because we have a sample of values from the posterior distribution of δ_{slo} it is straightforward to find the posterior probability that $\delta_{slo} < *$ where $*$ is a cut off point beyond which the benefits of cheaper food are outweighed by the costs of slower growth.

Out of the four elements the one for which we have the strongest posterior belief that the element is non-zero is the slope component of ρ. For this parameter the median is about 0.05 suggesting that the mean slope, or growth rate, for boars is about 0.05kg/day, or roughly 5%, higher than that for gilts. Figure 3.16 shows boxplots of the sampled values for the slope and intercept components of the μ_k parameters. The four columns represent the four weight groups. The plots show the expected pattern over weight groups.

The median of the sampled values of τ was 3.5 with 90% range (3.0, 4.2). Table 3.7 gives a summary of the sampled values of the elements of Σ. The off-diagonal element relates to a correlation of 0.73.

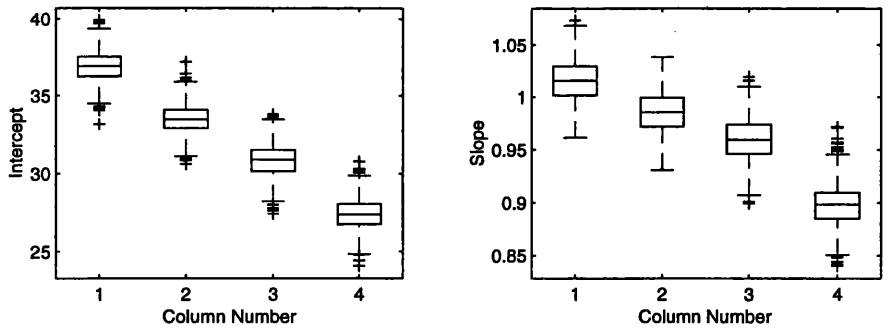


Figure 3.16: Sampled values of the μ_k parameters

Element	Median	90%	Interval
(1,1)	5.0	(3.4,	7.6)
(1,2)	0.065	(0.044,	0.097)
(2,2)	0.0016	(0.0010,	0.0025)

Table 3.7: Medians and 90% intervals for Σ

3.6 Sensitivity analysis and checking assumptions

The use of the Normal distribution at the first stage looks fine as the residuals are generally small and Normal probability plots show no evidence of non-normality. The fact that the weight measurements are averages over eight pigs rather than measurements on individuals lends support to the use of the Normal distribution.

We also need to consider the effect of our choice of R on the results. In particular, does using different values have a substantial effect on the value of δ_{slo} (which is the parameter we are most interested in as it represents the difference in 'mean' slopes or growth rates between the diets).

Therefore, the analysis was repeated with a second choice of R. The diagonal elements were 9 and .0009 with off-diagonal element .045. These values roughly equate to coefficients of variaton of 9% (higher than before) and 3% (lower then before) and a correlation of .5 (zero before). From this analysis the median of the sampled δ_{slo} values was -0.013 with 90% interval (-0.040, 0.011). These values are similar to those in Table 3.6. We also repeated the whole analysis

Element	Median	90%	Interval
(1,1)	5.8	(4.1,	8.6)
(1,2)	0.062	(0.043,	0.092)
(2,2)	0.0012	(0.0007,	0.0019)

Table 3.8: Medians and 90% intervals for Σ for the analysis with an alternative choice of R

with the original choice of R in order to see how much variability there was between repetitions of the whole MCMC sampling procedure. (Rather than the variability between repeated sequences with different starting points which is measured by the Gelman and Rubin statistics). For this analysis the median of the sampled δ_{slo} values was -0.014 with 90% interval (-0.040, 0.016). These values are similar both to those above and to those in Table 3.6. It does not appear that the alternative choice of R had any effect on the estimate of δ_{slo}.

We can also consider the effect of R on our estimate of Σ. The repeated analysis with the original choice of R gave virtually identically medians and intervals for the elements of Σ to those from the original analysis (given in Table 3.7). For the second choice of R the estimates of the diagonal elements were raised or lowered somewhat depending on the change in the relevant element of R as we would expect. There was little change in the values for the off-diagonal element (the estimates now relate to a correlation of 0.74 instead of the 0.73 obtained before). Table 3.8 gives these values.

3.7 Other methods of analysis

Two further analyses of this data were done by Mary Garratt of PIC International Group PLC. Some of the results from these analyses are presented here. These methods, particularly the first, are commonly used to analyse this type of data.

Diet	Start Wt.(kg)	Per.1	Per.2	Per.3	Per.4	Final Wt.(kg)
Control	33.1	0.938	0.964	1.010	1.045	90.2
New	33.0	0.879	0.937	1.033	1.058	89.4
(s.e.)	(0.26)	(0.014)	(0.021)	(0.016)	(0.020)	(0.58)

Table 3.9: Diet means (and standard errors) for analysis A. The middle four columns show the weight gain/pen/day (kg) for each of the four time periods

3.7.1 A: Separate analysis for each time period

For each time period the average daily weight gain was calculated for each diet over all weight groups and both sexes. The mean start and final weights were also calculated. Table 3.9 gives these values and also the standard errors. The difference between the diets is significant only for the first time period for which the mean weight gain was lower for the new diet than for the control.

From Table 3.9 we can also see that the average weight gains increase over time suggesting deviations from linearity. However, when plots of the data were inspected any such patterns appeared much less notable than suggested by the average gain figures. For five of the sixty-four pens there was some evidence of curvature in the form of increased growth rates in the final time period.

3.7.2 B: Individual lines for each pen

Straight lines were fitted individually to the 5 data points for each pen. The fitted intercept and slope parameters were then compared between diets as if they were the raw data. This was done over all weight groups and sexes. The results from this analysis are given in Table 3.10. The difference between the diets was not significant for either the slopes or the intercepts.

We note here that this is a standard analysis which would have been done within the company. It is very crude as it ignores the weight groups and sexes. It is of course possible to do a more sophisticated analysis of variance (ANOVA) including these variables.

Diet	Intercept	Slope
Control	32.6	0.990
New	32.0	0.978
(s.e.)	(0.26)	(0.0086)

Table 3.10: Diet means (and standard errors) for analysis B

3.8 Discussion

This chapter has described the use of Gibbs sampling to fit hierarchical models for linear data consisting of only five measurements per individual. Two alternative formulations of the model are also detailed. Also discussed are two more traditional methods of analysis.

For this data none of the three methods show a very strong difference between the two diets. In general we would expect the Bayesian hierarchical approach to be more powerful than the other methods.

Of the more traditional analyses the first (A) is the most straightforward. The second (B) is a simple hierarchical analysis with two stages carried out one after the other (we first estimate the intercepts and slopes for each individual separately and then analyse these as if they were raw data). Our Bayesian hierarchical model is better than this because all of the parameters are estimated simultaneously. Therefore, the estimation of the individual level parameters is affected by the population level parameters which in turn are affected by all of the individuals. In other words, the process borrows strength from all of the individuals when estimating the parameters for each individual. This means that we can still obtain sensible estimates of the individual level parameters even for individuals with very little data and also that estimates for 'unusual' individuals may be pulled towards the estimates for the other individuals.

Other advantages of the hierarchical approach are that there is no problem with non-standard weighing times, missing values or some individuals having more measurements than others. For this data there are some differences in the weighing times between the replicates. Methods A and B could still be used

because these differences were relatively slight and because the values used in Table 3.9 are weight gains per day rather than absolute weights (thus allowing differences in the number of days between weighings to be accounted for). Large differences in the weighing times or unequal numbers of measurements over individuals would present more serious problems. In the later case this is because any average weight gains or estimated parameters of individually fitted lines would be based of different numbers of observations and would therefore have different variances.

Other advantages of our hierarchical models are that they can estimate variation both within and between animals, we can include prior information, we get estimates for individual animals and that individuals with very little data available can be included and make a useful contribution to the model.

We now consider the other side of the coin, the advantages of methods A and B over the hierarchical appraoch. The most obvious advantage is their ease of application compared with the greater amount of time and thought required by hierarchical modelling. Depending on the importance of the problem being studied, the simpler methods may be perfectly adequate. However, it is worth noting here that the time spent setting up a hierarchial model is not time wasted. This is because it gives us a greater understanding of the data. Implementing the Gibbs sampler can take some time, and understanding, but is by no means prohibitively time consuming.

Finally, it is worth noting a further advantage of method A over the other two methods. Because it makes separate comparisons at each time point it can find differences which only occur at parts of trial. When using methods which fit straight lines to the data for individuals such differences should also be exposed, as patterns in the residuals, but may be overlooked. However, we should also remember that method A uses multiple tests and so the significance levels should be suitably altered.

Chapter 4

Non-linear Growth

4.1 Introduction

This chapter illustrates hierarchical modelling for some non-linear data. This is 'good' data since there are frequent weight measurements over a long period of time. The Gompertz function was found to be suitable for modelling the individual's growth. A random-walk Metropolis algorithm worked well for sampling the individual level parameters of the Gompertz function.

4.2 The data

There are weekly weight measurements for 10 male cats from birth to about 60 weeks of age. About a third of the measurements are after the cats have reached maturity. For the first 24 weeks the cats were 'on trial', being fed a new diet. After this time they changed to an 'old' diet. This was because data analysis within the company was done over the roughly linear period of growth of 8 to 24 weeks and the measurements after this were ignored. After 24 weeks there tends to be more variability in the measurements from one week to the next. This increased variability may be due, in part, to the cats drinking behaviour. Cats do not drink very often and therefore when they do it may have a large effect on weight. It is possible that this behaviour changes as they get older.

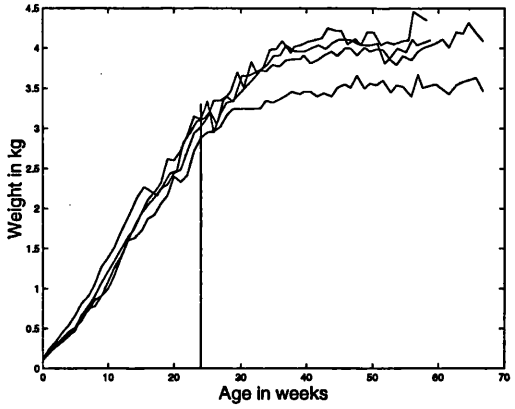


Figure 4.1: Some of the data

Figure 4.1 shows the growth of four of the cats. The vertical line shows the end of the 'on trial' period.

4.3 Choice of growth function to use

We have a relatively large amount of data for each animal making it feasible to fit growth functions with more than one phase such as the diphasic. However, unlike the pig (protein and lipid), the cat has only one major body component (protein/muscle/lean) (Munday et al, 1994). This suggests that a single phase function may be a suitably good fit to our data. Therefore, each of the single phase functions with three parameters detailed in Chapter 1 was fitted to each cat using maximum likelihood estimation. The total residual sums of squares obtained were 7.0 for the Gompertz, 8.2 for the von Bertalanffy and 9.2 for the Logistic. As well as comparing the residual sums of squares we also looked for evidence of systematic departures from each of the functions. Figure 4.2 shows the residuals from each of the functions. All of the cats are included in each plot. All three plots show a systematic lack of fit. The Gompertz function is the best of the three. There is a strong pattern at early ages for the Logistic and von Bertalanffy functions. This suggests that the point of inflection for these functions is incorrectly placed for this data. Therefore, the Gompertz function was used in our models.

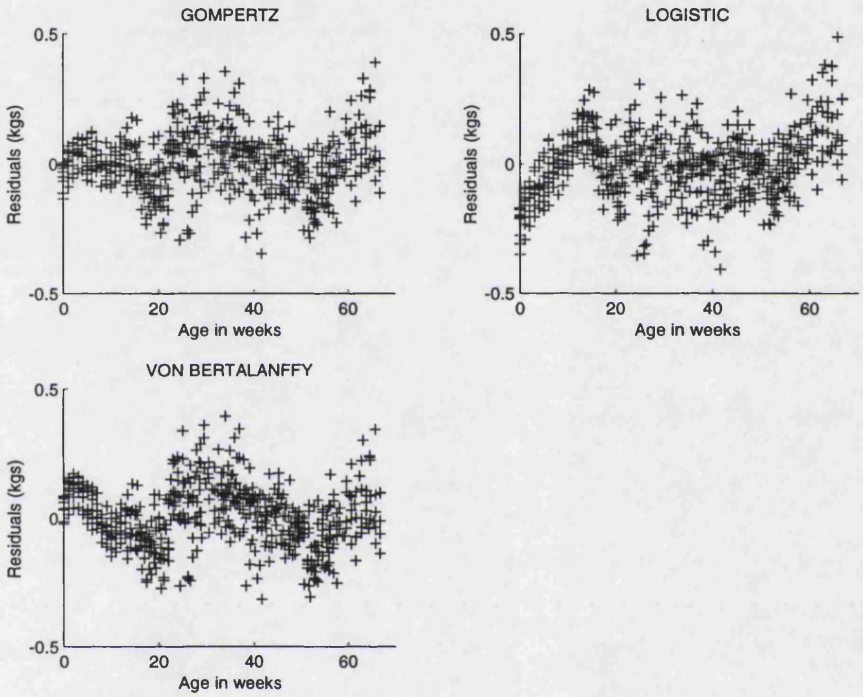


Figure 4.2: Residuals plots for three single phase functions

4.3.1 Choice of parameterization of the Gompertz function

In order to improve the behaviour of the MCMC sampling we would ideally like the three parameters making up the Gompertz function to have only small correlations between them (in other words, to be stable parameters). Also, because we will use a Normal distribution as our proposal distribution (see Section 4.5) we want the likelihood, which dominates the conditional distributions, to be approximately Normal. We would also like the parameters to have straightforward biological interpretations if possible.

Firstly we considered the parameterization given in Chapter 1. Because the k or θ_{i2} parameter must be greater than zero, and takes values close to zero, we have adapted the function slightly by replacing θ_{i2} with $\exp(\theta_{i2})$. Therefore, we have,

$$g(\theta_i, x_i) = \theta_{i1} \exp[-\exp\{-\exp(\theta_{i2})(\frac{x_i}{\theta_{i3}} - 1)\}]$$

We then considered two other possible parameterizations and compared these three possibilities in terms of our requirements (see the first paragraph of this section). Because θ_{i1} and θ_{i3} have straightforward biological interpretations (mature weight and age at point of inflection respectively) we kept these two parameters the same and looked at reparameterizing the 'rate' parameter, θ_{i2}. For the parameterization given above the actual maximum rate of growth is a function of all of the components of θ_i and is given by $\frac{\theta_{i1} exp(\theta_{i2}-1)}{\theta_{i3}}$.

One alternative is for θ_{i2} to represent the maximum rate of growth. We also considered using the log of the maximum rate of growth. The first of these gives,

$$g(\theta_i, x_i) = \theta_{i1} \exp[-\exp\{-\frac{\theta_{i2} \exp(1)}{\theta_{i1}}(x_i - \theta_{i3})\}]$$

The second gives,

$$g(\theta_i, x_i) = \theta_{i1} \exp[-\exp\{-\frac{\exp(\theta_{i2} + 1)}{\theta_{i1}}(x_i - \theta_{i3})\}]$$

For each of these parameterizations we plotted contours of the residual sums of squares for combinations of θ_{i2} and θ_{i1} or θ_{i3} (having fixed the other parameter

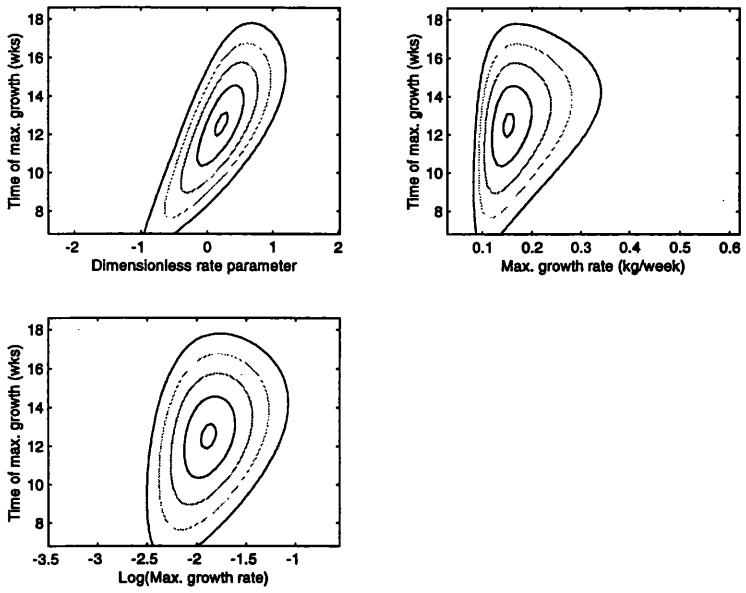


Figure 4.3: Contour plots of the RSS for θ_{i3} and the three suggested θ_{i2}s for i=1

to its maximum likelihood estimate). We also obtained the plot for θ_{i1} versus θ_{i3}. Examples of these plots (for $i = 1$, i.e. the first cat) are given in Figures 4.3 and 4.4. In each case the five contours represent 1, 2.5, 5, 8 and 12. The pattern is similar for all i.

We rejected the possibility of using the maximum growth rate parameterization because of the suggestions of a lack of normality seen in the contour plots. There is little to choose between the other two possibilities. Estimates of the correlations between the parameters were obtained for each individual from the $\hat{\Omega}_i$ matrices (defined in Section 2.5.3). The correlations between the log(maximum rate) parameter and the mature weight parameter were all between -0.62 and -0.44. All of the correlations between the dimensionless rate parameter and the mature weight parameter were between -0.34 and -0.07. Similarly, the correlations with the point of inflection parameter were between -0.11 and 0.19 and between 0.41 and 0.70 for the two parameterizations. Partly because of the low correlations with the point of inflection and also because of the ease of biological interpretation we decided to use the log(maximum rate) parameterization.

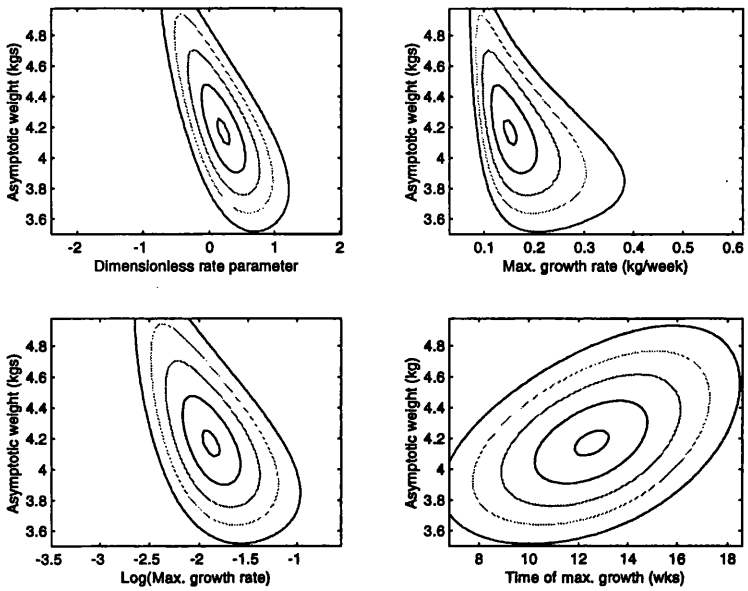


Figure 4.4: Contour plots of the RSS for θ_{i1} and the three suggested θ_{i2}s (also for θ_{i1} and θ_{i3}) for i=1

4.4 The model

4.4.1 First level

There are $I = 10$ individuals and we have vectors, y_i, of n_i measurements taken at ages x_i. The largest number of measurements on an individual animal is 65. The smallest number is 58.

$$p_1(y_i \mid \theta_i, x_i, \Lambda) = N_{n_i}(g(\theta_i, x_i), \Lambda),$$

where g is the Gompertz function in the following form:

$$g(\theta_i, x_i) = \theta_{i1} \exp[-\exp\{-\frac{\exp(\theta_{i2}+1)}{\theta_{i1}}(x_i - \theta_{i3})\}].$$

The three components of the θ_i vectors have the following interpretations:

θ_{i1}	mature weight (kg),
$\exp(\theta_{i2})$	maximum growth rate (kg/week),
θ_{i3}	age at point of inflection (weeks),

Λ is a diagonal matrix with elements:

$$\begin{aligned} &\tau_{trial}^{-1} && \text{for } x_i \leq 24, \\ &\tau_{offtrial}^{-1} && \text{for } x_i > 24. \end{aligned}$$

This allows the precision to be lower after the trial has ended. (See Section 4.2).

4.4.2 Second level

We model the θ_i vectors as being a sample from a Normal distribution with mean μ and covariance matrix Σ,

$p_2(\theta_i \mid \mu, \Sigma) = N_3(\mu, \Sigma)$.

We include a reminder here that the mean is the same for all of the cats because all of the 10 cats/θ_i are exchangeable since they were all male and were all fed the same diet.

In order to consider the suitability of the Normal distribution at this stage we found the maximum likelihood estimates, $\hat{\theta}_i$, and did Normal probability plots for each of the three elements of the vectors. These plots provided no evidence that the Normal distribution was not suitable.

4.4.3 Third level/priors

The structure of the third level is as outlined in Section 2.3.3 except that we now have two first-stage precision parameters. For each of these we use an (improper) Ga(0,0) prior. (i.e. setting ν_0 to zero). The prior for Σ^{-1} is the least informative inverse Wishart distribution that we can use ($\rho = 3$ = number of parameters in the Gompertz function). In order to obtain values for R, the approximate prior estimate of Σ, we estimated ranges for each of the three parameters which we thought nearly all of this type of cat would fall inside (after consideration of other data).

These ranges were:

mature weight	3 to 5.5kg,
maximum growth rate	0.1 to 0.2 kg/week,
point of inflection	9 to 15 weeks.

The range for the maximum growth rate equates to a range of (-2.3, -1.6) for the θ_{i2} parameter. (By taking logs).

We then set these ranges to 4 standard deviations giving diagonal elements of R as follows: $[.39, .03, 2.25]$. We might expect there to be a positive correlation between the mature weight and the maximum growth rate with faster growing cats growing to larger mature weights. On the other hand they may grow quickly but over a shorter period of time and cats with a lower maximum growth rate may end up being the same size or even larger at maturity. We have little substantive knowledge about this correlation or for those correlation terms involving the point of inflection and therefore we have set the off diagonal elements of R to zero.

The prior for μ had mean, $\eta = [4.25, -1.9, 12]$, and a diagonal covariance matrix, C, with elements given by $100 \times R$. The values chosen for η are the midpoints of the ranges used to estimate R and are typical fitted parameters for a normal male cat. The value of -1.9 equates to a maximum growth rate of 0.15kg/week since $\exp(-1.9) = 0.15$. The large value used for C, the covariance matrix, means that this prior provides little information to our model.

4.4.4 Full conditional distributions

For all except the θ_i parameters the full conditionals are standard distributions. These are given below.

$$[\mu \mid y, \theta, \Sigma^{-1}, \Lambda] = N[\mu \mid V(I\Sigma^{-1}\bar{\theta} + C^{-1}\eta), V],$$

where $\bar{\theta}$ represents the mean θ and the matrix V is defined by $V^{-1} = I\Sigma^{-1} + C^{-1}$.

$$[\Sigma^{-1} \mid y, \theta, \mu, \Lambda] = W[\Sigma^{-1} \mid \{\sum_{i=1}^{I}(\theta_i - \mu)(\theta_i - \mu)' + \rho R\}^{-1}, I + \rho],$$

$$[\tau_{trial} \mid y, \theta, \mu, \Sigma^{-1}, \tau_{offtrial}]$$

$$= Ga[\tau_{trial} \mid \frac{1}{2}(\nu_0 + n_{trial}), \frac{1}{2}\{\sum_{i=1}^{I} RSS_i(trial) + \nu_0\tau_0\}],$$

where n_{trial} is the total number of measurements made on trial ($24 \times I$) and the $RSS_i(trial)$s represent the residual sums of squares for the present values of the θ_i but only summing over the trial period. We have used $\nu_0 = 0$ so the distributions for τ_{trial} and $\tau_{offtrial}$ simplify somewhat.

$$[\tau_{offtrial} \mid y, \theta, \mu, \Sigma^{-1}, \tau_{trial}]$$

$$= Ga[\tau_{offtrial} \mid \frac{1}{2}(\nu_0 + n_{offtrial}), \frac{1}{2}\{\sum_{i=1}^{I} RSS_i(offtrial) + \nu_0\tau_0\}],$$

where $n_{offtrial}$ is the total number of measurements made off trial and the $RSS_i(offtrial)$s represent the residual sums of squares for the parts of the fitted curves after the trial has ended.

The full conditionals for the θ_i vectors are proportional to the likelihood multiplied by the prior:

$$\exp(-\frac{\tau_{trial}}{2}\sum_{j=1}^{24}(y_{ij} - g(\theta_i, x_{ij}))^2 - \frac{\tau_{offtrial}}{2}\sum_{j=25}^{n_i}(y_{ij} - g(\theta_i, x_{ij}))^2)$$

$$\times \exp(-\frac{1}{2}(\theta_i - \mu)'\Sigma^{-1}(\theta_i - \mu)).$$

4.5 MCMC methods

Gibbs sampling was used for all parameters except the first stage ones, θ_i. For these, a random walk Metropolis algorithm was used.

This involved calculating the inverse information matrices, $\hat{\Omega}_i$, which we evaluated at the maximum likelihood estimates. We have a relatively large

amount of data for each individual enabling good estimates of these quantities to be made.

Suitable acceptance rates were obtained without multiplying the $\hat{\Omega}_i$ by scaling factors. Therefore, for each i, the proposal was a multivariate Normal distribution with mean the present value of θ_i and covariance matrix $\hat{\Omega}_i$.

4.6 Effectiveness of the MCMC methods

Three repeats comprising 1000 iterations each were done using different starting values for the individual rate and time of point of inflection parameters. The three pairs of values used were (.3, 1.0), (.45, 1.2) and (.6, 1.4) representing early slow growth, central moderately paced growth and late fast growth. These values are well spread about the expected posterior distributions. Because of the relatively large amount of data when the animals are mature the mature weight parameters are easier to estimate than the other two parameters. Therefore we have used the final weight of each animal as the starting values for each of the 3 repeats. The starting values used for all the other parameters (i.e. the 2nd stage ones and the τ's) are also the same for each of the three repeats.

The overall acceptance rate was appropriate (0.25) and the individual acceptance rates were all between 0.24 and 0.27. One Metropolis step was done for each individual at each overall iteration. When new values were not accepted the previously sampled values were retained.

Figure 4.5 shows the first 500 iterations for each of the three repetitions for the time of point of inflection parameter for cat 4. The speed of convergence seen in this plot is typical of that for the other cats and for the other parameters.

Gelman and Rubin's statistic was calculated for each of the 41 parameters (as scalars and taking logs of positive valued parameters) at multiples of 100 iterations. After 500 iterations all of the $\sqrt{\hat{R}}$ values were less than 1.1. Consideration of plots of the iterations also suggested that convergence was attained relatively quickly.

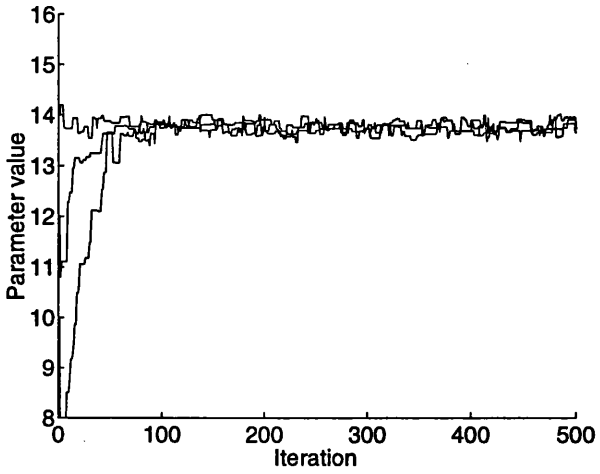


Figure 4.5: The first 3 × 500 iterations for the θ_{43} parameter

We therefore discarded 250 iterations from each repeat leaving $3 \times 750 = 2250$ as our posterior sample. (Since at 500 iterations the $\sqrt{\hat{R}}$ values were all less then 1.1, and these values are calculated after omitting the first halves of the sequences, i.e. the first 250.) Figure 4.6 shows the final 750 (retained) iterations for each of the three repetitions for the log(max. growth rate) parameter for cat 6. The sampling after burn-in can be seen more clearly here than in Figure 4.5.

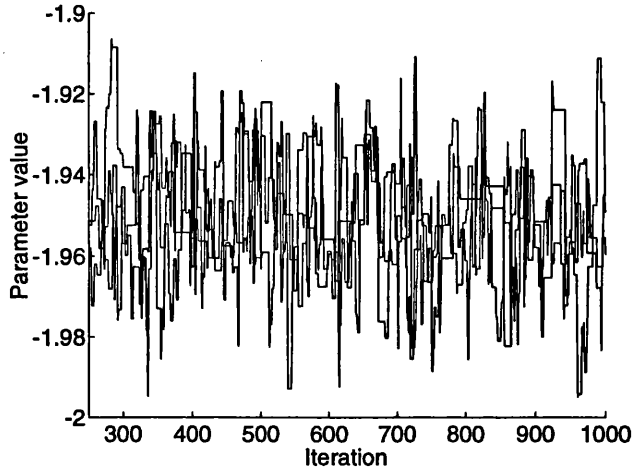


Figure 4.6: The final 3 × 750 iterations for the θ_{62} parameter

4.7 Results

The posterior medians and 90% ranges for the elements of μ are given below:

4.13 KGs	(3.86, 4.41)
-1.91	(-1.97, -1.84) i.e. 0.148 KG/wk (0.139, 0.159)
12.0 weeks	(11.2, 12.8)

The posterior medians of the elements of Σ are given below. These values correspond to correlations of 0.14 between the mature weight and rate parameters, 0.33 between the mature weight and time of point of inflection parameters and -0.08 between the rate and time of point of inflection parameters.

$$\begin{pmatrix} 0.2544 & 0.0084 & 0.2288 \\ & 0.0143 & -0.0131 \\ & & 1.9419 \end{pmatrix}$$

Table 4.1 gives 90% intervals for the elements of Σ. We note that the intervals include zero for each of the correlation terms.

The medians and 90% ranges for the precision parameters when expressed as standard deviations (in grams) are given below:

on trial: 75 (69, 81)

off trial: 126 (119,134)

As expected, the standard deviation was considerably smaller for the on trial period.

Element	90% interval
(1,1)	(0.136 0.596)
(2,2)	(0.007 0.035)
(3,3)	(0.971 4.650)
(1,2)	(-0.028 0.055)
(1,3)	(-0.124 0.896)
(2,3)	(-0.136 0.082)

Table 4.1: 90% intervals for the elements of Σ

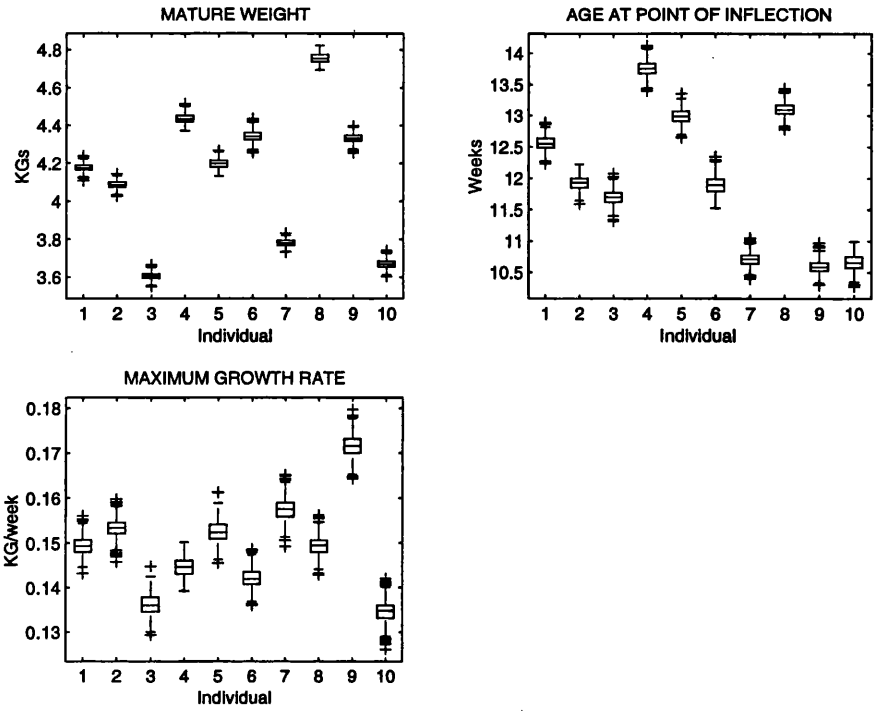


Figure 4.7: Boxplots of the parameters for the individuals

Figure 4.7 shows boxplots of the retained iterations for the individual parameters for the mature weights and point of inflection parameters and also for the maximum growth rates ($\exp(\theta_{i2})$'s). Each column relates to one cat.

The median of each component of the θ_i vectors was used to calculate fitted values for each cat. Figure 4.8 shows these values along with the observed data for four of the cats. The total residual sum of squares for all cats was 7.2 (slightly higher than that from the maximum likelihood estimators). Figure 4.9 shows the standardised residuals for all of the cats and Figure 4.10 shows all of the fitted curves on one graph.

We also used the whole posterior distribution of the residuals as discussed in Section 2.8. The residuals were calculated at each iteration of the sampling procedure. The medians and 95% highest posterior density intervals of these residuals for the retained iterations were plotted for each time point for each individual. These plots are shown for two of the cats in Figures 4.11 and 4.12. In these plots, the vertical lines represent the intervals and the horizontal lines represent two standard deviations (estimated from the medians of the preci-

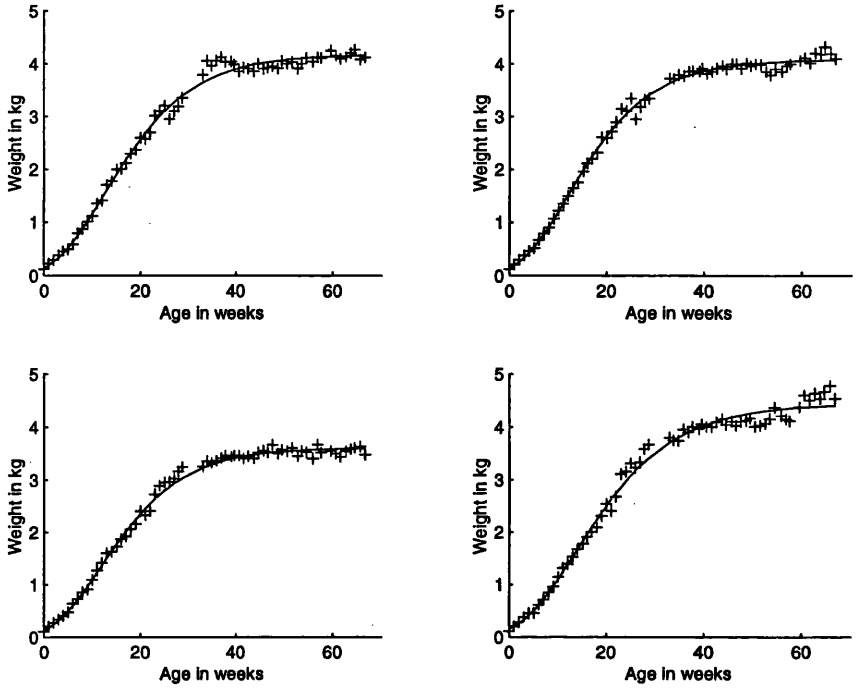


Figure 4.8: Observed and fitted plots for i=1:4 (reading horizontally first i.e. 1st row is $i = 1$ then $i = 2$ and so on)

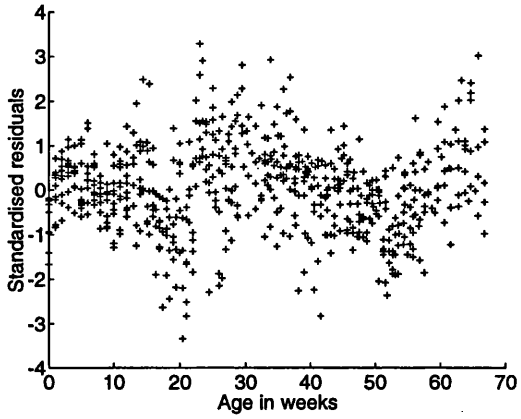


Figure 4.9: Standardised residuals against age for all of the cats

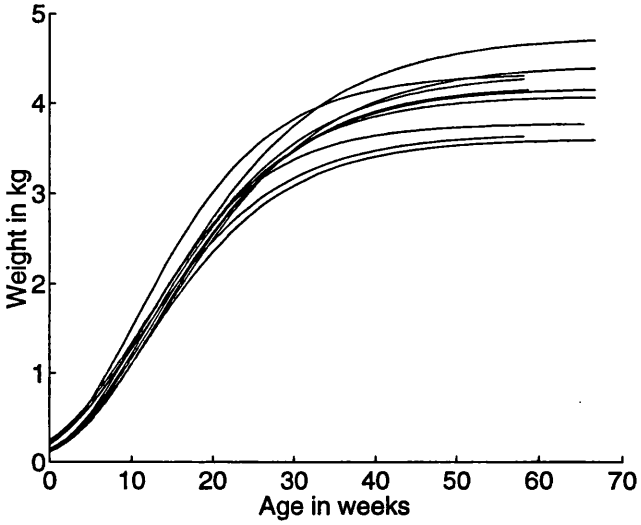


Figure 4.10: The fitted curves

sion parameters, τ_{trial} and $\tau_{offtrial}$). Because we have two different precision parameters these lines showing two standard deviations are different for the two different parts of the data (giving the step pattern to the plots). The plots show that the ranges for the residuals are fairly narrow and also that the values are generally small.

Normal probability plots and also the lack of obvious outliers suggested that our use of normal distributions was reasonable, both for the data and for the first-stage prior. The normal probability plots for the elements of the θ_i vectors are shown in Figure 4.13. This type of model may be sensitive to the assumption of normality. However, normality looks reasonable for our models and so we have used it here and elsewhere in this thesis. We note here that Student-t distributions can (and should) be used in many cases.

Finally, we compared the maximum likelihood estimates with our estimates. They were very similar, particularly so for the mature weight parameters. This is not surprising given the large amount of data for each individual.

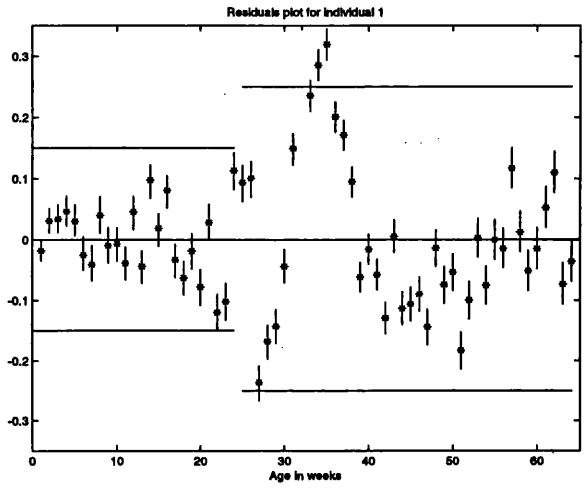


Figure 4.11: Individual 1

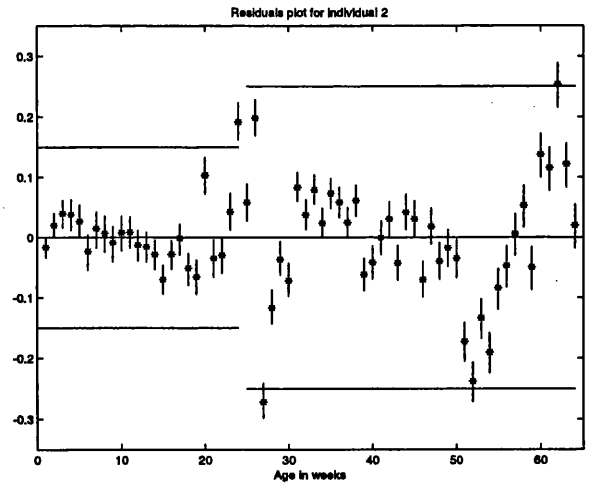


Figure 4.12: Individual 2

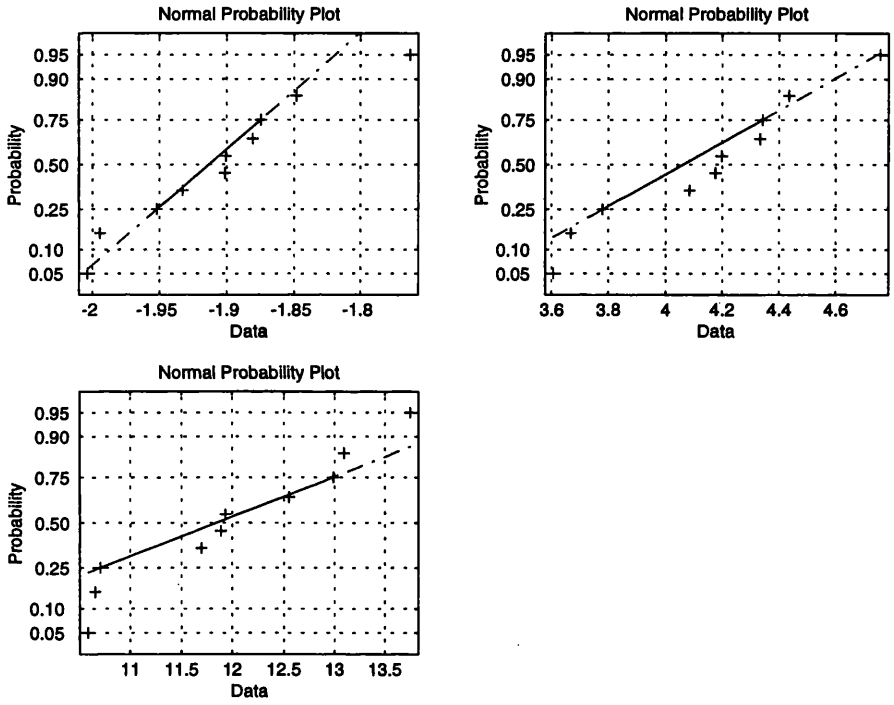


Figure 4.13: Normal probability plots for the elements of the θ_i

4.8 Sensitivity to choice of priors

4.8.1 Sensitivity to choice of R

The analysis was repeated with a different choice of R. The diagonal elements were the same as before but the off-diagonal elements relating to the covariance between the mature weight and maximum rate parameters were changed from zero to 0.054. This relates to a positive correlation of 0.5 between the two parameters.

This second analysis gave almost exactly the same estimates and intervals as before for the elements of μ, and for the τ parameters.

The medians and intervals for the first level, θ_i, parameters are also very similar between the two analyses.

The medians for the elements of Σ are given below and the 90% intervals are in Table 4.2.

$$\begin{pmatrix} 0.2512 & 0.0250 & 0.2338 \\ & 0.0145 & -0.0149 \\ & & 1.9218 \end{pmatrix}$$

Element	90% interval
(1,1)	(0.128 0.585)
(2,2)	(0.008 0.035)
(3,3)	(0.963 4.668)
(1,2)	(-0.004 0.079)
(1,3)	(-0.106 0.830)
(2,3)	(-0.143 0.091)

Table 4.2: 90% intervals for the elements of Σ for the second analysis

The estimates of the elements of Σ given in the matrix are similar to those obtained in the previous analysis except for the covariance term whose prior estimate we changed. The posterior estimate is now 0.0250 as opposed to 0.0084

previously. This relates to a correlation of 0.41 compared with the 0.14 obtained before.

We have seen that for this model the choice of R does have an effect on the estimate of Σ but has little effect on the other parameters. If we are particularly interested in the estimate of Σ then we need to be sure that our choice of R is an accurate representation of our prior beliefs, preferably obtained after discussions with experts in the relevant field.

4.8.2 Using more informative priors

Previously the priors used were weak. However, we did obtain genuine prior information and can make better use of this information by making the priors more informative. Also in this section, we remove some of the data to investigate whether with less data but stronger priors we can obtain answers as good as those obtained previously.

We did five further analyses. Firstly we kept the original data but made the priors stronger. We did this by increasing the value of ρ from 3 to 8. The new prior gives much narrower (and more realistic) ranges of likely values for the elements of Σ. We also decreased C from $100R$ to R. We continued to use improper priors throughout for τ_{trial} and $\tau_{offtrial}$ since their estimates were based on a relatively large amount of data (even when half the animals were removed from the data).

The effect of using the stronger prior was to make the highest posterior density (HPD) regions for the diagonal elements of Σ slightly narrower. The medians moved slightly nearer to the prior estimates. There was little change in the medians and intervals for the elements of μ. These medians and HPD regions are shown in Figures 4.14 and 4.15. For each plot, and for each column, the line represents the 95% HPD region and the * the median. The first column relates to our original model, with all 10 cats and the weak prior. The second column relates to the new analysis with stronger priors and the same data.

The third columns of the plots relate to an analysis for which we randomly

removed half of the cats and returned to our original weak priors. The HPD regions for the diagonal Σ elements are now much wider and the medians have moved slightly towards the prior estimates. The intervals for the elements of μ alos became much wider and the medians moved away from the prior estimates (because the five individuals remaining in the data were further from the prior beliefs than the five removed inidividuals).

For the analysis whose results are shown in the fourth columns of the Figures we again used the reduced data set but made the priors stronger. We used the same priors as in the second analysis ($\rho = 8, C = R$). The medians of the diagonal elements of Σ remained similar but the intervals became much narrower as we would expect (still somewhat wider than for the original analysis though). For the elements of μ the medians moved towards the prior estimate and the intervals became narrower (again, as we would expect when making the prior stronger).

We then did two further analyses maintaining the value of ρ at 8 but making C progressively smaller ($\frac{1}{4}R$ and then $\frac{1}{16}R$) therefore increasing the strength of our belief in η, the prior estimate of μ. There was little change in the estimates and intervals for the elements of Σ. This is as we would expect since we only changed the prior for μ. For μ we observed the same pattern as from the third to the fourth analyses. With each increase in the strength of the prior the medians moved towards the prior estimate and the intervals became narrower. The intervals for the final analysis were narrower than for the original analysis. The increased strength in the prior has more than cancelled out the decrease in the amount of data. We have similar sized intervals although we have used less data. Of course, since the analysis with reduced data 'relies more heavily' on the prior it will be more sensitive to the choice of prior parameters.

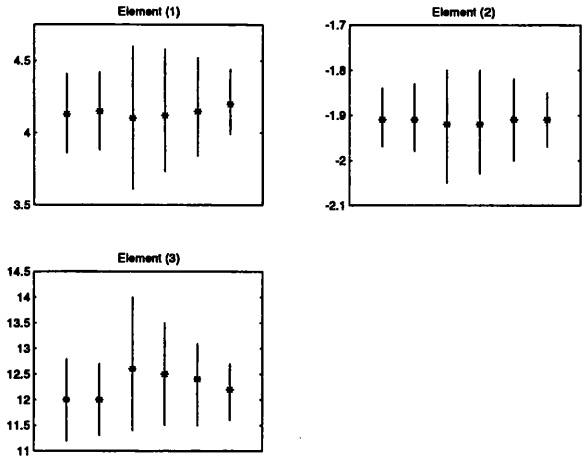


Figure 4.14: Medians and 95% highest posterior density regions for the elements of μ

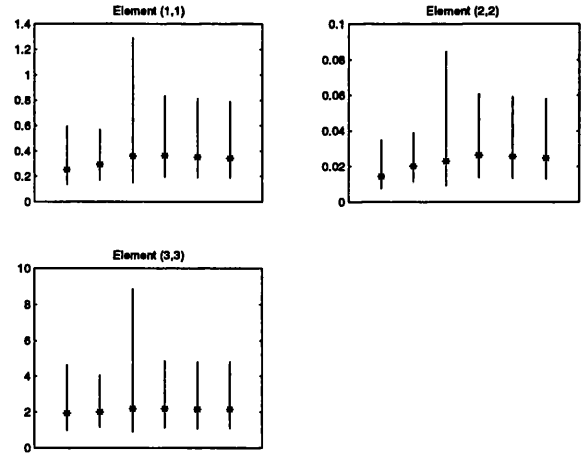


Figure 4.15: Medians and 95% highest posterior density regions for the diagonal elements of Σ

We have not illustrated the effects of changing the model on the off-diagonal elements of Σ but the main effect was that the HPD regions again became narrower when using the stronger prior.

The estimates of the θ_i vectors were very similar between the different analyses. The estimates of τ_{trial} and $\tau_{offtrial}$ were different between the analyses with different numbers of individuals included. Within these groups of analyses the estimates were very similar between the analyses with different priors.

We emphasize here that when making the priors stronger the model becomes more sensitive to the choice of prior parameters. In other words, the model becomes less robust to changes in the values of the prior parameters used.

4.9 Discussion

The residuals were generally small. However, they did have some pattern, in particular a discontinuity at about 24 weeks of age perhaps due to the change in diet. It would be possible to improve the model to take account of this but this has not been done here.

The random-walk Metropolis algorithm worked well for this data since we could obtain good estimates of the posterior distributions to use as our proposal distributions. The model could easily be extended to allow, for example, a comparison of two groups of animals fed different diets.

Chapter 5

Non-linear Growth with Information on Body Components

5.1 Introduction

As discussed in Chapter 1 we are interested not only in the overall growth of animals but also in the growth of the various components which make up the animal. Models including body components are generally of much greater interest to those studying the biological processes of growth than models which only include the overall growth.

Here we have data including measurements made after slaughter on various body components. This enables us to split the overall body into two groups of components modelling each one with a single phase function. The overall body weight is then given by the sum of these two phases (a diphasic function).

Therefore, we have a non-linear first stage and again use a random-walk Metropolis algorithm to sample the first stage parameters. This process is not as straightforward as in the previous chapter because as well as the diphasic function having more parameters than a single phase one we also have fewer measurements per animal and in some cases very few. This causes problems

with finding maximum likelihood estimates and hence also the $\hat{\Omega}_i$ matrices.

5.2 The data

The data used here were collected by van Lunen (1994) and analysed in his thesis. A brief summary of the reasons he was interested in the data follows. It was thought that previously accepted theories on growth and the nutrient requirements of pigs were not accurate for newly developed pig genotypes. These genotypes had been shown to have daily live weight gains of the order of 1kg/day with nitrogen deposition rates (related to protein or lean/muscle growth, protein deposition rate is approximately 6.25 times nitrogen deposition rate) of about 30g/day. This compares to typical values of .8kg/day and 21.5g/day respectively for 'conventional' pigs. Because of these differences it was thought likely that growth patterns and body compositions would be different for these pigs and that they would have a higher daily requirement for protein and energy. Also, in 'conventional' pigs, boars (male) were thought to have higher potential for lean growth than gilts (female) or castrates. No work had been done on whether this sex difference also existed for these new genotypes. Using more advanced statistical methods than van Lunen we may be able to shed further light on these questions and can compare our results with his findings. The work in this chapter is also useful as an illustration of how this type of data may be analysed in order to investigate other biological matters.

The data incudes 60 pigs of a genotype with potential for fast lean growth. They were made up of 30 gilts and 30 boars. One of the boars was omitted because it died whilst on the trial. The pigs began the trial when they were about 10kg in weight (about 4 weeks old - unfortunately the exact ages are not available but are thought to be the same to within $\pm$ 3 days). Their weights were measured roughly weekly. One pig of each sex was slaughtered at 10kg intervals. This process began at 10kg live weight which means that there is only one live weight value for the first two pigs to be slaughtered and similarly small

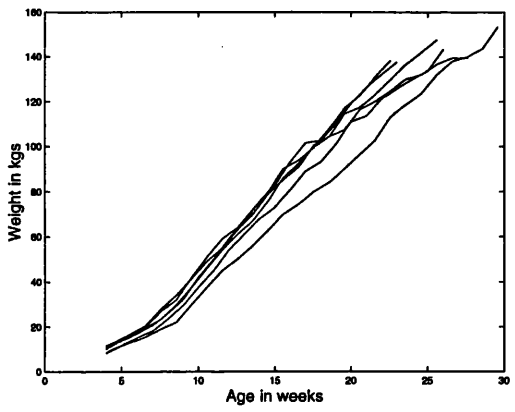


Figure 5.1: Growth for the 6 pigs with most data

numbers of measurements for the other pigs slaughtered early in the trial. The final pig to be slaughtered had reached 154kg after 25.5wks on the trial (i.e. at approximately 29.5 weeks old). This is some way off the expected mature weight of pigs. Nevertheless, pigs are generally slaughtered for their meat before this weight is reached meaning that the data is relevant for practical purposes. However, it does make the fitting of growth curves harder as we have little information on when the curves 'level off' as mature weight is neared. Figure 5.1 shows the observed growth for the 6 pigs for which there is most data. Chemical and other analyses allowed the measurement of various body components after slaughter. The main components are water, lean/muscle/protein, fat/lipid and bone.

Interest centres on the relative amounts of lean and lipid (fat) in an animal and the interplay between these over the growth period. In order to investigate this we have split the slaughter data into two components, non-fat and fat. It is fortuitous that we wish to split the data in this way since the growth of fat has been found to occur later than the other components and it is therefore appropriate to model it using a separate phase (see later).

Figure 5.2 shows the non-fat (*) and fat (+) values for all of the pigs plotted together. For each of non-fat and fat there is only one measurement for each animal. The non-fat values rise rapidly from a relatively early age. This is as we expect since this genotype of pig has the potential for fast lean growth. In

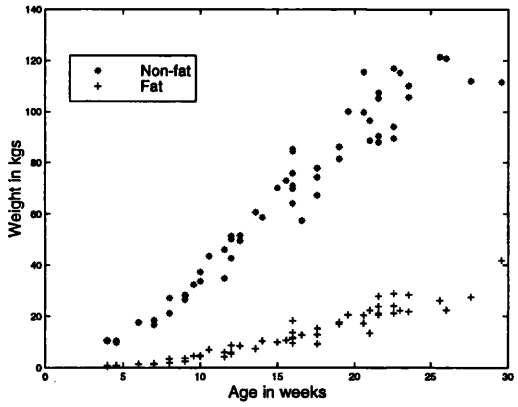


Figure 5.2: Components data

contrast, the fat values show a slower increase becoming more rapid towards the end of the data. Little information is available on the amount of fat in a mature pig of this type since, for economic reasons, the animals are generally slaughtered before 'too much' fat is put on.

By modelling these two components as separate phases of growth we should then be able to investigate 'fat-adjusted' growth by consideration of the first phase.

5.2.1 Feed intake

The pigs were fed *ad libitum* and the weekly food intake values were available for each pig.

Whittemore et al (1995) give the following equation:

$$E = 52P + 53L + 0.44W^{0.75},$$

where E is energy intake in MJ per day, P is daily increase in protein and L is daily increase in lipid. The third term on the right hand side represents the daily maintenance requirement at weight W. The value of 0.44 used in the maintenance requirement was estimated from experimental work mainly done during the 1980s. Because the pigs in our data show fast lean growth they include a lot of protein/lean. This means that their metabolism and therefore maintenance requirements are generally higher than for previous genotypes. Therefore we

have replaced the 0.44 in the formula with 0.6:

$$E = 52P + 53L + 0.6W^{0.75}.$$

This change was made after discussion with Pieter Knap, an expert on pigs from PIC International Group PLC. Whittemore et al developed the equation above in order to predict required feed intakes, but we will use it in reverse to make rough estimates of the likely growth rates when defining our prior distributions.

It would also be possible to incorporate the feed intake data into a model in a more fundamental way. This has not been done here.

5.2.2 Details of the estimation of the fat values

The pigs were slaughtered a few days after their final live weight measurement was taken. During this time weight generally falls mainly due to decreases in gut contents and in water content. This was found to be the case for this data, most of the live weights on the day of slaughter were lower than the final measurements on the trial. Very little of this weight loss was thought to be due to a loss of fat. Therefore, we use the estimate of fat content made after slaughter as our estimate of fat content at the time of the last measurement on the trial. (Assuming also that there was little, if any, growth of fat during the intervening period).

These estimates of the fat content were adjusted to include leaf fat. This is a type of fat which is not included in the measurement process after slaughter (because the body parts containing the leaf fat are removed before the chemical procedure takes place). Leaf fat typically makes up between 6 and 9% of total fat and our estimates of fat content were increased by 7% to allow for this extra fat.

The non-fat values were obtained by subtracting the estimate of fat content from the final live weight on the trial.

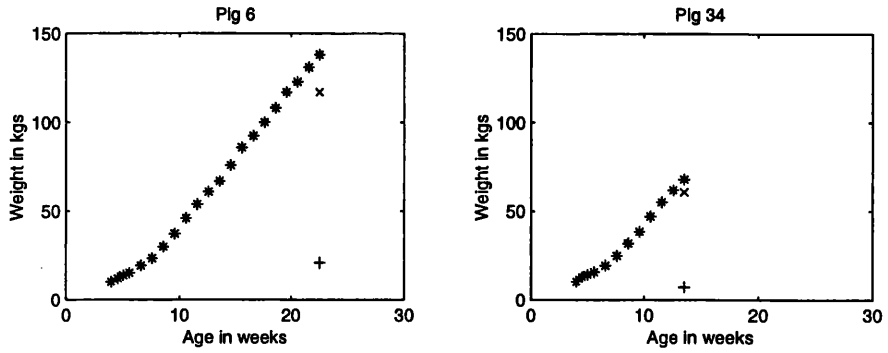


Figure 5.3: The data for two of the pigs

5.3 The data and notation used in our model

For each individual, i, we have live weights, y_i, at times (approximate ages), x_i. Each of these is a vector with n_i elements. We also have the fat value, y_i^* at time x_i^* (which is the final element of x_i). The individuals $i = 1 : I$ are roughly in order of age at slaughter, the pigs with low i values have the most data. Figure 5.3 shows the data for two of the pigs. The fat values (+) and non-fat values obtained by subtraction (x) are shown here along with the live weights (*).

5.4 Choice of growth function

We use a diphasic function in order to model the two components in the slaughter data as two phases of growth. The question is, which single phase function to use as the basis of the diphasic? A single Gompertz function is commonly used to model growth for pigs as the point of inflection is believed to be at about the right time. The Gompertz function has also been used to model the growth of various body components. For example, Whittemore et al (1988) used it to model protein growth, finding it to be more satisfactory than simpler models (linear and quadratic polynomials). We used the Gompertz here. Because of the large number of parameters for each individual and the relative lack of data for many individuals there will be high correlations between the parameters for

each individual whichever parameterization we choose. Therefore, we used the same parameterization as in Chapter 4 because of the ease of interpreting the parameters in biological terms.

The logistic was also tried resulting in fairly good fits to the data but the results were not thought to be biologically reasonable. Because of the lack of data at later ages the estimated mature weights are highly dependent on the fitted position of the point of inflection which will be roughly the same for the logistic and the Gompertz (for data over these age ranges). This means that the Gompertz will give estimated mature weights for each component about 0.35 higher than the logistic (because the point of inflection is at 0.368 of mature weight whereas for the logistic it is at half of mature weight). The mature weights from the diphasic-logistic function were thought to be unfeasibly low. However, we must bear in mind that we do not know how heavy these actual pigs would have become - only that the values were very low compared with other pigs which have been kept to later ages.

5.5 The model

5.5.1 First level

The live weights, y_i, are modelled by the diphasic-Gompertz function.

$$p_1(y_i \mid \theta_i, x_i, \tau_1) = N_{n_i}(g(\theta_i, x_i), \tau_1^{-1} I_{n_i \times n_i}),$$

where g is the sum of two Gompertz functions:

$$g_1(\theta_i, x_i) = \theta_{i1} \exp[-\exp\{-\frac{\exp(\theta_{i2}+1)}{\theta_{i1}}(x_i - \theta_{i3})\}],$$

$$g_2(\theta_i, x_i) = \theta_{i4} \exp[-\exp\{-\frac{\exp(\theta_{i5}+1)}{\theta_{i4}}(x_i - \theta_{i6})\}].$$

The biological interpretations of the parameters are as follows:

θ_{i1}, θ_{i4} mature weight of that phase (kgs),

θ_{i2}, θ_{i5} rate of growth parameters,

θ_{i3}, θ_{i6} times of points of inflection (weeks).

The maximum growth rates (in kgs/week) are given by $\exp(\theta_{i2})$ and $\exp(\theta_{i5})$. We used the simple variance specification of constant variances because although we might expect the variability to increase with age/size we did have quite large residuals at early ages (see later).

The fat values are modelled as follows:

$$p_{12}(y_i^* \mid \theta_i, x_i^*, \tau_2) = N_1(g_2(\theta_i, x_i^*), g_2(\theta_i, x_i^*)^2 \tau_2^{-1}),$$

where g_2 is as given above with the vector x_is replaced by the scalar x_i^*s. We expect the variance to increase with the mean, in particular because the fat values are very small for early ages, and have therefore used a power relationship between them. We used power 2 so we have variance proportional to mean squared. This approximately corresponds to a log-normal error specification. It is possible to include the power to be used as a parameter in the model. Wakefield et al (1995) illustrate this approach. We initially tried this here but found the MCMC sampling to perform badly.

The g_1 component of the total diphasic function therefore represents the non-fat component of the body. The non-fat values are not explicitly modelled except as the difference between the live weights and the fat values.

5.5.2 Second level

We use different means but the same covariance matrix for the θ_i parameters for the two sexes:

$$p_2(\theta_i \mid \mu_k, \Sigma) = N_6(\mu_k, \Sigma),$$

where $k = 1, 2$ represents the two sexes (1=males, 2=females).

5.5.3 Third level/priors

We use the general structure outlined in Chapter 2 for the priors.

We obtained rough estimates of the components of η and of R from consideration of the literature on the growth of pigs and then adapted the values because of the expected differences between these pigs and 'conventional' pigs. This process was again discussed with Pieter Knap.

We estimated the mature weights to be 150kg and 60kg for the non-fat and fat components respectively.

In their analysis of fat and fat-free dry matter (see Chapter 1) Koops and Grossman (1991) estimated the points of inflection as 17 weeks and 26 weeks for non-fat and fat respectively. We have subtracted three weeks from each of these values because of the expected fast/early growth of these pigs. Therefore, the values used were 14 and 23 weeks of age.

Van Lunen (1994) defined fast growing pigs as those having a potential protein depostion rate above 170g/day. Water content is thought to increase with protein and to be about 4 times the protein content. Between them these components make up the bulk of the non-fat component. Therefore, we estimated the maximum daily gain of the non-fat component by multiplying the 170g by 5 and then adding a bit more to allow for the growth of bone and also to allow for the fact that the protein depostition may be above 170g/day. This gave us our estimate of 1kg/day or 7kg/week.

For our estimate of the maximum growth rate of the fat component we used the feed intake information and the energy intake equation given in Section 5.2.1. The feed used for the pigs in our dataset had an energy content of 14.5MJ/kg. The time of maximum lipid gain is thought to be after 20 weeks of age for which times the feed intakes were fairly constant. Taking 23 weeks as a possible time of maximum lipid gain, the feed intakes were about 3.2kg. Therefore, the energy intake was about 46MJ/day. At 23 weeks the pigs weighed approximately 130kg making the maintenance requirement 23MJ/day. This leaves 23MJ/day for growth. We estimated the rate of protein growth at 23 weeks as 0.125kg/day. Using this estimate in the energy intake formula the rate of lipid growth comes out at 0.31kg/day.

When converted into the appropriate units these values give the following values: $\eta = [150, 1.95, 14, 60, 0.775, 23]$.

The C matrix (see below) has quite large elements and so the values used for η do not have a great effect on the analysis. They are also used as the starting values for the θ_i (see below).

The prior structure is more complicated than that outlined in Chapter 2 since we have two first stage precisions instead of one. For each of these we use a Gamma prior as suggested in Chapter 2 but with different parameters for the live weight and for the slaughter precision. For τ_1 we used $\nu_0 = 0$, i.e. an improper prior.

For τ_2 we used a more informative prior as there was a danger that all of the fat values could be fitted exactly by the model giving a residual sum of squares of zero. If this was the case and the prior was improper then the posterior for τ_2 would have infinite variance ($Ga(a, b)$ with $b = 0$).

The values used were $\nu_0 = 32$ and $\tau_0 = 0.0025$. These values relate to a Gamma prior with mean 400 and standard deviation 100. The standard deviation of the fat values from the fitted values is given by $\tau_2^{-\frac{1}{2}} g_2(\theta_i, x_i^*)$. Therefore, for a pig slaughtered at about the middle of the trial and having $g_2(\theta_i, x_i^*) = 20$ (in other words, about 20kg of fat), a τ_2 value of 400 relates to a standard deviation of 1 kg.

For R we used a diagonal matrix with elements:

$$\begin{pmatrix} 225 & & & & & \\ & 0.010 & & & & \\ & & 1.56 & & & \\ & & & 156 & & \\ & & & & 0.017 & \\ & & & & & 6.25 \end{pmatrix}$$

This corresponds to the following 2 standard deviation ranges on either side of the means for the θ_i values:

θ_{i1}	30kg,	θ_{i4}	25kg,
$\exp(\theta_{i2})$	1.4kg/week,	$\exp(\theta_{i5})$	0.56kg/week,
θ_{i3}	2.5weeks,	θ_{i6}	5weeks.

These values were discussed with Pieter Knap and the ranges were thought to be sensible. We used a diagonal matrix because we did not have strong prior beliefs about any of the off-diagonal, correlation, terms. As demonstrated in Chapter 4, the posterior estimates of the off-diagonal elements of Σ can be sensitive to the choice of correlation terms used in the priors and it might be wise to put more thought into the values used if we have a particular interest in the posterior estimate of Σ.

The value used for C was $25 \times R$, a fairly vague specification.

5.6 Full conditionals and 'graph' of the model

Figure 5.4 shows a 'graph' of the model and the full conditionals are given below.

$$[\mu_1 \mid y, \theta, \Sigma^{-1}, \tau_1, \tau_2, \mu_2] = N[\mu_1 \mid V(I_1 \Sigma^{-1} \bar{\theta} + C^{-1}\eta), V],$$

where $\bar{\theta}$ is the mean of the θ_i for sex 1 (males) and I_1 is the number of males (29). The matrix V is defined by $V^{-1} = I_1\Sigma^{-1} + C^{-1}$.

(Similarly for μ_2).

$$[\Sigma^{-1} \mid y, \theta, \mu_1, \mu_2, \tau_1, \tau_2]$$

$$= W[\Sigma^{-1} \mid \{\sum_{males} (\theta_i - \mu_1)(\theta_i - \mu_1)' + \sum_{females} (\theta_i - \mu_2)(\theta_i - \mu_2)' + \rho R\}^{-1}, I + \rho],$$

$$[\tau_1 \mid y, \theta, \mu_1, \mu_2, \Sigma^{-1}, \tau_2]$$

$$= Ga[\tau_1 \mid \frac{1}{2}n, \frac{1}{2}\{\sum_{i=1}^{I} RSS_i(liveweights)\}],$$

where n is the total number of live weight measurements and the RSS_i are the residual sums of squares for these live weights when using the present values of the θ_i to give fitted values.

$$[\tau_2 \mid y, \theta, \mu_1, \mu_2, \Sigma^{-1}, \tau_1]$$

$$= Ga[\tau_2 \mid \frac{1}{2}(\nu_0 + I), \frac{1}{2}\{\sum_{i=1}^{I} \frac{(y_i^* - g_2(\theta_i, x_i^*))^2}{g_2(\theta_i, x_i^*)^2} + \nu_0\tau_0\}].$$

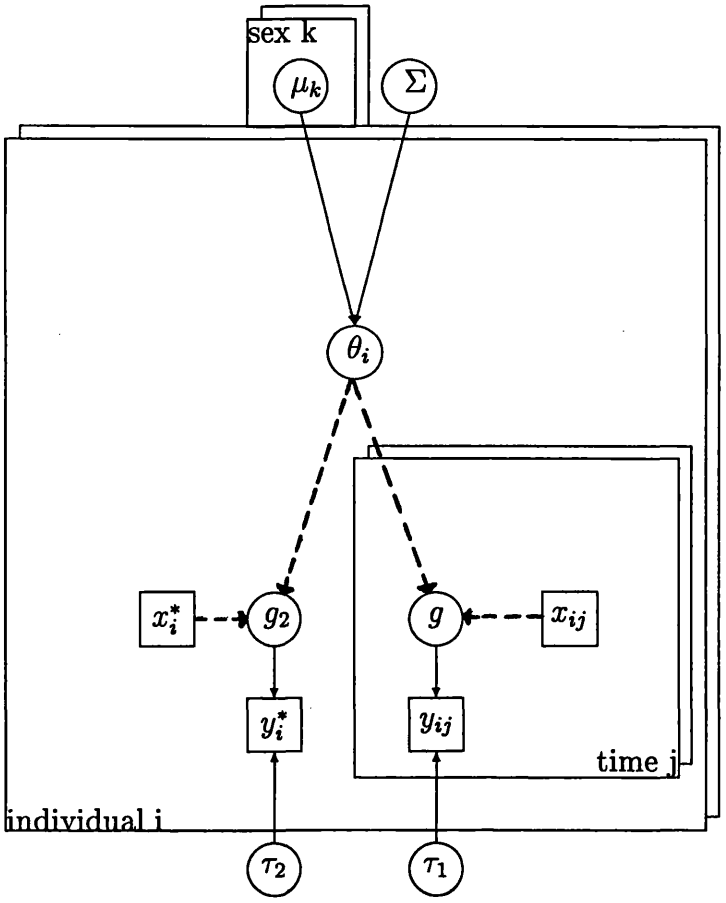


Figure 5.4: The model

The full conditionals for the θ_i vectors are proportional to the likelihood multiplied by the prior (using the appropriate μ_k value depending on the sex of individual i):

$$\exp\left(-\frac{\tau_1}{2}\sum_{j=1}^{n_i}(y_{ij}-g(\theta_i,x_{ij}))^2-\frac{\tau_2}{2}\frac{(y_i^*-g_2(\theta_i,x_i^*))^2}{g_2(\theta_i,x_i^*)^2}\right)$$

$$\times\left(\frac{\tau_2}{g_2(\theta_i,x_i^*)^2}\right)^{\frac{1}{2}}$$

$$\times\exp(-\frac{1}{2}(\theta_i-\mu_k)'\Sigma^{-1}(\theta_i-\mu_k)).$$

5.7 MCMC methods

We used Gibbs sampling for all of the parameters except the θ_i which we sampled as complete vectors using a random-walk Metropolis algorithm.

The present values were used as the means of the proposal distributions. Because of the relative paucity of data for some pigs and consequent problems with estimating the maximum likelihood estimators and therefore also the information matrices, it was not sensible to use the information matrices as the covariance matrices in our proposals. Instead we did some preliminary runs in order to obtain better estimates of the covariance matrices of the posterior or target distributions for each individual. The initial run of 750 iterations used different covariance matrices depending on how much data there was for each individual. For the 39 pigs with least data the matrix used was proportional to Σ, the second stage covariance matrix. The scaling factors were proportional to $\frac{1}{\sqrt{\tau_2}}$ so that as the precision increased, the size of the proposed steps decreased. This meant that for the early iterations, when the θ_i parameters give poor fitted values and hence low precisions, the proposed steps will be quite large and the scheme will quickly move to parameters giving better fits. As this happens the precision increases and the size of the proposed steps decreases and the proposed values of the θ_i do not change so much thus continuing to give good fitted values. The scaling factors were related to the amount of data for an individual, the less data the larger the scaling factor. These scaling factors (and the square root above) were chosen to give appropriate acceptance rates.

For the other 20 pigs the covariance matrix used in the proposal was $2.5 \times \hat{\Omega}_1$ where $\hat{\Omega}_1$ is an estimate of the inverse information matrix for the pig with most data. The expression for $\hat{\Omega}_1$ includes the precision parameters so again the size of the proposed steps decreases as the precision increases.

The initial 750 iterations were then used to calculate covariance matrices for each of the θ_i vectors. These covariance matrices (again scaled appropriately but no longer dependent on the precision parameters since they were relatively constant by now) were then used in our proposals for a further 2000 iterations

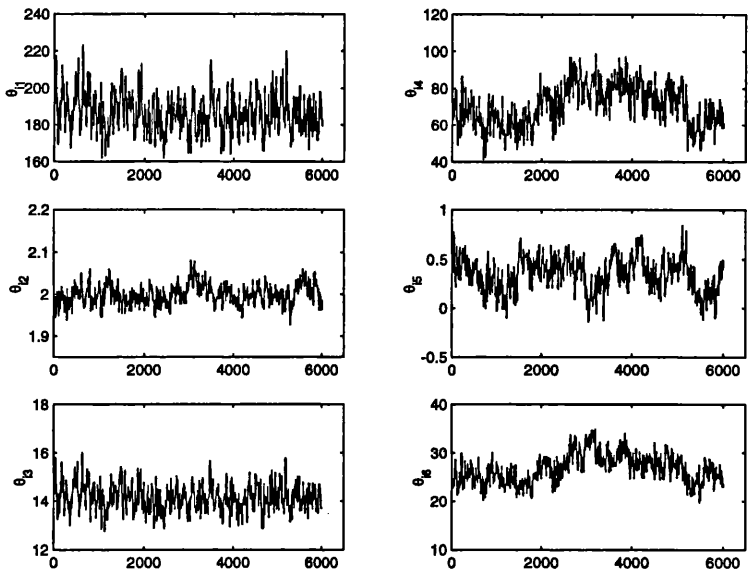


Figure 5.5: The iterations for individual 10

for which the starting values were the final values from the previous run. The covariance matrices were then calculated for these 2000 iterations and used in our proposal distributions for a further run of 2000 iterations. The covariance matrices were again calculated and used for our final run of 6000 iterations (again scaled appropriately and using the final values from the previous run as our starting values).

At each stage of this process the covariance matrices used in the proposals were better estimates of the covariance matrices of the target/posterior distributions. For the final run of 6000 iterations the scaling factors were 0.6 for all individuals.

The average acceptance rate for the final run was 0.23. The acceptance rates varied between 0.13 and 0.31 between individuals but there was no pattern in this variation. Figure 5.5 shows the final 6000 iterations for the components of θ for one individual ($i = 10$).

The original starting values were:

θ_i η for all i,

μ_1 η,

μ_2 η,

Σ R,

τ_1 0.002,

τ_2 5.

The sampled values moved away from the starting values during the initial runs and therefore the final run was used for our inferences without removing any part of it as burn-in. Inspection of plots of the iterations for the various parameters suggested that convergence had been satisfactorily attained.

5.8 Results

The medians of the sampled values of each element of the θ_i vectors were found and used to calculate fitted growth curves for each individual. Figure 5.6 shows observed and fitted plots for the four pigs with most data. The fitted values appear to fit the observed data well.

However, inspection of the residuals in Figure 5.7 reveals a strong pattern at the early ages. No differences were apparent between the sexes in this pattern when they were plotted separately. The model tends to overestimate the live weights at about 6 to 10 weeks of age. The observed weights do not increase very much in the first few weeks on the trial. It may be that the pigs were moved from another location before they started the trial. In this case, the transportation and the acclimatization to the new surroundings may have slowed the growth process down somewhat. Similarly, some biological process or veterinary procedure may have caused the weights to be low between 6 and 10 weeks old. Alternatively, it might be explained by the presence of an earlier phase of growth accounting for part of the nervous system and bone. We could include such a phase in our model by assuming that this phase of growth was

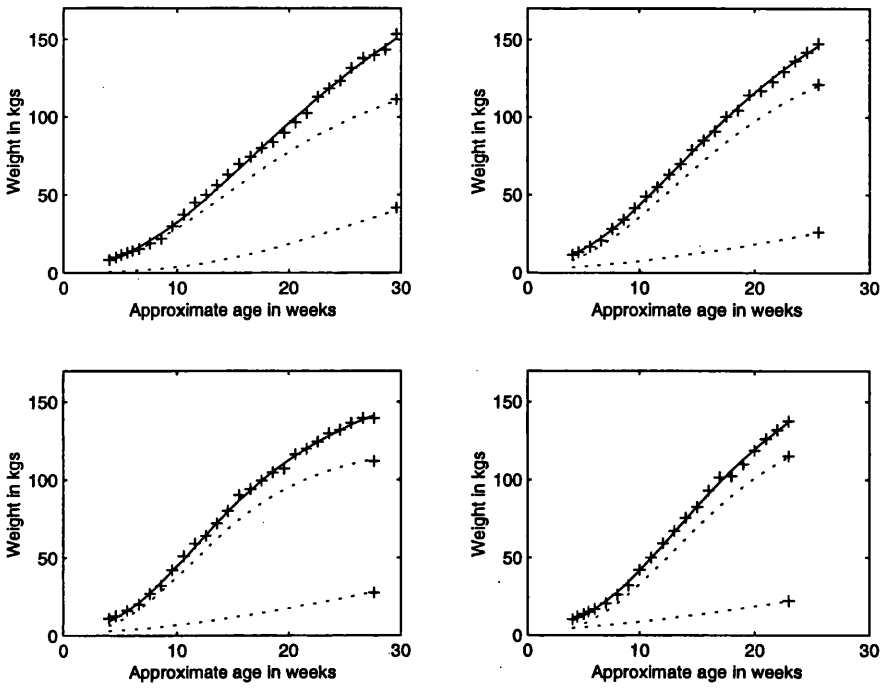


Figure 5.6: Observed and fitted plots

complete before the start of this trial and adding an extra parameter to the diphasic function (so the first phase becomes $g_1(\theta_i, x_i) + \theta_{i7}$). It is also possible that the lack of fit is because the Gompertz function is not exactly right for the phases for this data. The position of its point of inflection is at an essentially arbitrary proportion of mature weight. It is entirely feasible that this proportion is not exactly correct for this data. It is theoretically possible to use a Richards function and thus allow the point of inflection to be flexible. Unfortunately, in practice this would introduce very high correlations between parameters. This has been found to be a problem with the Richards function even for simpler (single phase) cases and with data extending over longer periods (Davies and Ku, 1977).

Figure 5.8 shows the residuals from the slaughter data. They appear to be randomly scattered above and below zero and there is a pattern of increasing magnitude with increasing age suggesting that we were right to make the precision dependent on age at slaughter.

We investigated whether our chosen power relationship between the mean

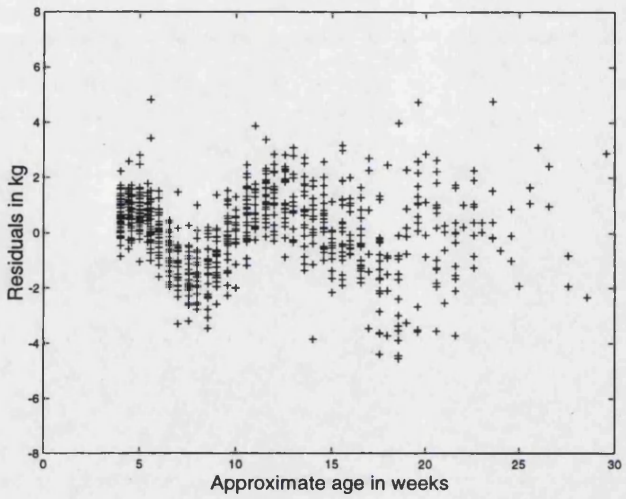


Figure 5.7: The residuals for the live weight data for all individuals

and variance (variance proportional to mean squared) was appropriate by plotting the residuals divided by the standard deviations,

$$\frac{\tau_2^{\frac{1}{2}}(y_i^* - g_2(\theta_i, x_i^*))}{g_2(\theta_i, x_i^*)}.$$

Figure 5.9 shows this plot. Table 5.1 gives the posterior medians and 95% intervals for the μ_k parameters. Figures 5.10 to 5.11 show the diphasic functions obtained from the medians of the elements of the μ vector for each sex. The solid lines represent the two phases of growth and the overall growth (given by the sum of the two phases). The dotted lines represent the growth rates of the two phases and overall (and relate to the right hand y-axes).

By summing the two mature weight parameters for each sex we obtain an estimate of the population mean mature weight for each sex. These values are 235kg for boars and 220kg for gilts. As part of his analysis, van Lunen (1994) estimated the mature weights by fitting a single phase Gompertz equation to all of the data for each sex using maximum likelihood. This method gave estimates of 192kg for both sexes. Our mature weights are higher and more in line with values reported in the literature. Also, we detected a difference between the sexes which the simpler method did not. The sex difference in our results is that we would expect the boars to be on average about 15kg heavier overall.

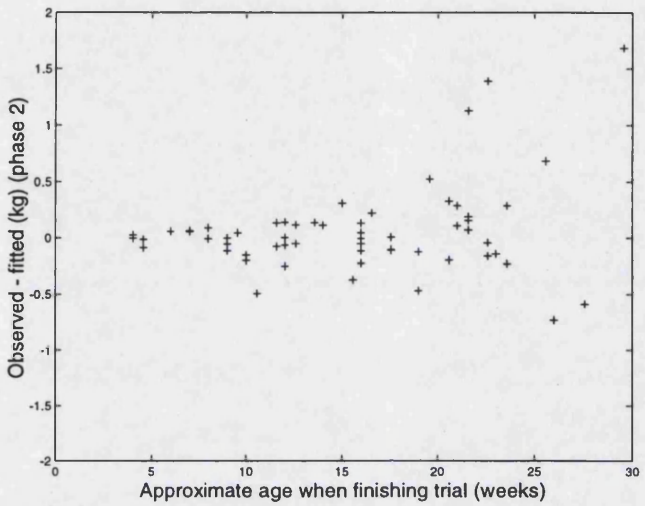


Figure 5.8: Residuals from the slaughter data

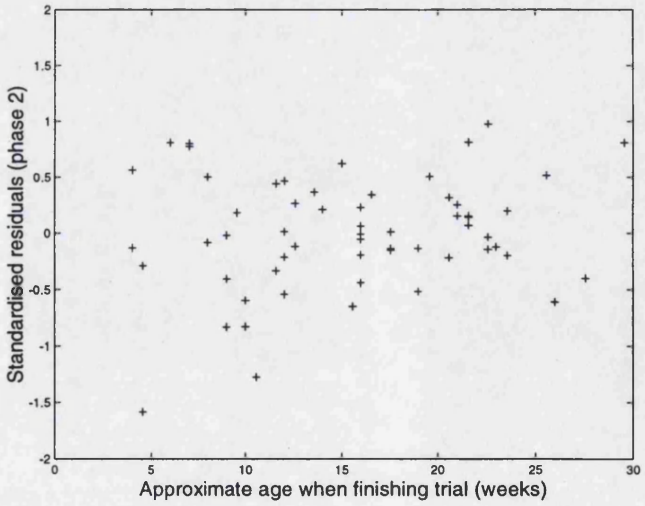


Figure 5.9: Standardised residuals from the slaughter data

		Male		Female	
Phase	Parameter	Median	95% interval	Median	95% interval
1	Mature wt.(kgs)	164	(155, 174)	138	(130, 148)
1	'rate'	1.93	(1.88, 1.97)	1.80	(1.76, 1.84)
1	Pt.of inf.(wks)	14.1	(13.5, 14.7)	13.2	(12.6, 13.9)
2	Mature wt.(kgs)	71	(56, 86)	82	(71, 100)
2	'rate'	0.32	(0.13, 0.45)	0.58	(0.45, 0.73)
2	Pt.of inf.(wks)	27.0	(23.7, 31.6)	27.1	(24.6, 30.3)

Table 5.1: Posterior medians and 95% intervals for the μ parameters

When looking at the components separately the boars tend to be leaner with about 26kg more lean content and about 11kg less lipid/fat than the gilts. These values can be used to answer some of the questions raised in Section 5.2. Our estimates suggest that boars do have a higher potential for lean growth than gilts for this genotype as well as for 'conventional' pigs. Both phases are centred a bit earlier for gilts than boars.

It is useful to convert the medians of the μ elements into estimates of the 'mean' maximum daily growth rates for each phase for each sex. This was done by taking the exponential of the 'rate' elements of the μ vectors and then dividing by seven to give daily not weekly growth rates. The values obtained (and 95% intervals) were 0.98kg/day (0.94, 1.02) and 0.20kg/day (0.16, 0.22) for the boars and 0.86kg/day (0.83, 0.90) and 0.26kg/day (0.22, 0.30) for the gilts. We can see from Figures 5.10 and 5.11 that the overall maximum daily gains occur close to the centres of the first phase and are about 1.13 kg/day for boars and 1.03 kg/day for gilts. We note here that we should not place too much faith in the estimates, particularly for the second phase, because of the lack of data at later ages and consequent model uncertainty as regards the choice of growth function.

Table 5.2 gives the medians and 95% intervals for τ_1, τ_2 and the elements of Σ. The estimate of the precision parameter for the live growth data corresponds to a standard deviation of 1.49kg (95% range 1.41, 1.57).

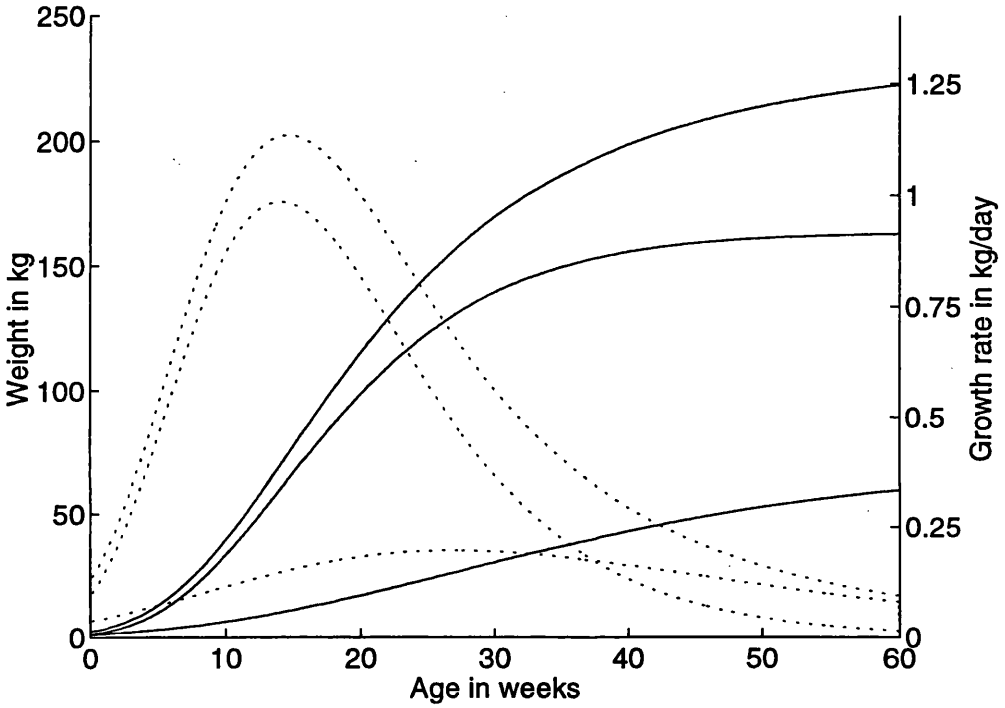


Figure 5.10: The diphasic function given by the medians of the elements of μ for the male pigs

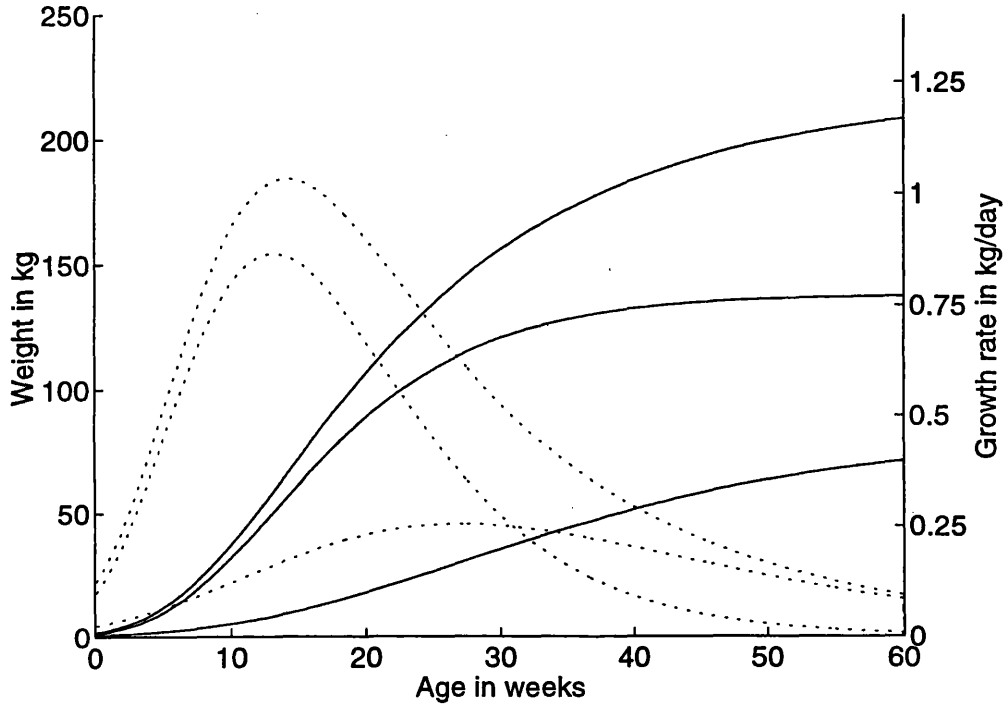


Figure 5.11: The diphasic function given by the medians of the elements of μ for the female pigs

Parameter	Median	95% interval
τ_1	0.452	(0.405,0.505)
τ_2	368	(219,594)
$\Sigma(1,1)$	254	(144,477)
$\Sigma(2,2)$	0.009	(0.006,0.015)
$\Sigma(3,3)$	1.58	(0.92,2.83)
$\Sigma(4,4)$	76	(42,151)
$\Sigma(5,5)$	0.027	(0.011,0.076)
$\Sigma(6,6)$	3.53	(1.74,7.16)
$\Sigma(1,2)$	-0.00	(-0.56, 0.68)
$\Sigma(1,3)$	12.45	(5.31, 27.56)
$\Sigma(1,4)$	-9.01	(-99.87, 66.56)
$\Sigma(1,5)$	1.16	(-0.11, 2.92)
$\Sigma(1,6)$	-3.60	(-27.30, 14.25)
$\Sigma(2,3)$	-0.06	(-0.11, -0.02)
$\Sigma(2,4)$	-0.07	(-0.53, 0.31)
$\Sigma(2,5)$	-0.01	(-0.02,-0.00)
$\Sigma(2,6)$	0.01	(-0.10, 0.14)
$\Sigma(3,4)$	1.52	(-4.66, 7.89)
$\Sigma(3,5)$	0.12	(0.03, 0.26)
$\Sigma(3,6)$	-0.31	(-1.76, 1.11)
$\Sigma(4,5)$	0.39	(-0.23, 1.66)
$\Sigma(4,6)$	3.33	(-3.81,15.63)
$\Sigma(5,6)$	-0.07	(-0.36, 0.12)

Table 5.2: Posterior medians and intervals for the τ and Σ parameters

5.9 Using just live growth data

We also fitted the same model but only including the live weights in order to see how similar the results were. Specifically, to see how similar the two separate phases were to those obtained when including the slaughter data.

When fitting this simpler model the total residual sum of squares was slightly lower than for the previous model. However, the pattern observed in the residuals in Figure 5.7 was still present.

There were quite large differences between the estimated parameters from this model and the one including the slaughter data. We conclude that quite large changes in the fitted parameters will give fitted curves which are almost as good a fit to the live growth data. Therefore, when not including the slaughter data in the model any biological interpretations of the two phases obtained would be very unreliable. On the other hand, our full model, including the slaughter data, does not give the diphasic curves which are the best fit to the live growth data, but instead gives almost as good a fit and has phases which can be reliably linked to body components.

5.10 Discussion

Markov chain Monte Carlo methods

The sampling needed to fit the models in this chapter was less straightforward than that done in the previous two chapters. This was because of the increased number of parameters in our growth function and the relatively small amount of data for all individuals (and some in particular). These factors combined to cause large correlations between the 6 parameters for each individual making sampling and accurate estimation of the parameters harder. Therefore, more effort was required to find suitable estimates of the covariance matrices to use in our proposal distributions. Also, longer runs were needed because of the correlations between parameters.

Benefits of our modelling approach

The use of sophisticated statistical models allowed us to make more use of the available data than other methods could. In particular, we were able to detect a difference in lean growth between the sexes. The models are also more comprehensive than other approaches. For example, they give us information on the amount of variability between individuals (ignored by methods which simply average over individuals).

Further work

Often, back fat measurements are recorded as well as live weights. These are relatively crude measurements of the thickness of fat at certain points on the animal's back. Equations have been developed to relate these measurements to the total amount of body fat. These estimates are less accurate than the chemical measurements made after slaughter but they have the advantage that repeated measurements may be made on each animal. It would be possible to adapt our model to include such measurements possibly increasing the accuracy of the estimation of the 'fat' phase.

Summary

This chapter has shown how complex hierarchical models can be fitted to animal growth data, including components measurements, using Markov chain Monte Carlo methods. At the present time these types of models and the MCMC means of fitting them appear to be the 'best' way of modelling such data. Best is used here to mean that we have powerful and flexible models which can be adapted (for example, to include more data or to better model the error structure), we do not have to use large numbers of restrictive assumptions, we can include all of the data in one comprehensive model (even for animals with little data), we can include prior knowledge in our model and we automatically obtain intervals as well as point values for all estimates from the model. Against all of these advantages we have the effort of specifying prior knowledge and of finding appropriate sampling schemes and the computational burden of the sampling.

Chapter 6

Discussion

6.1 Practicalities of the sponsoring company using these methods

This work was funded by a CASE Studentship, the company involved being interested in the growth of various species. This thesis has used data from trials carried out on pigs and cats to illustrate the value of hierarchical models. We now consider how practical it would be to use these methods within the company.

We have illustrated the use of Bayesian hierarchical models for several datasets of increasing complexity. In Chapter 3, the data used was for only a small part of the growth process and the observed data could be fitted by straight lines. This meant that the Markov chain Monte Carlo procedure was straightforward as all of the full conditionals were standard distributions and Gibbs sampling could be used with little difficulty. Using such hierarchical models provides many benefits as discussed at the end of Chapter 3. Against these we have the extra effort required to set up suitable models and to do the Markov chain Monte Carlo sampling. Because of the relative ease of Gibbs sampling, the programs we have used could be used by a non-expert to analyse other data. However, a certain level of competence would be required in order to fully understand the models and to properly assess convergence and interpret the output. Nevertheless, this should be feasible, certainly for someone with a reasonable amount of statistical

knowledge and some guidance/training initially.

Many of the same points apply to the analysis of non-linear growth data discussed in Chapter 4. The situation is further complicated here by the need for a Metropolis sampling step for the first stage, θ_i, parameters. However, if the Gompertz function was again used and the data had a similar number of measurements (and over a similar age range) we would expect the sampling scheme we used to work reasonably well. In fact, the data in Chapter 4 includes more than enough measurements at maturity and we should still obtain good estimates of the θ_i and hence the $\hat{\Omega}_i$ and the sampling should still work well with fewer measurements at maturity. The program could also easily be adapted to allow different population mean vectors for different diets to allow diet comparisons to be made.

Some further work would be required to allow the use of a different growth function at the first stage. This would not be terribly demanding for an expert and would involve some alterations to the program and also the relevant differentiations in order to obtain the $\hat{\Omega}_i$ matrices for the new function. Given these changes we would expect the sampling to work well, assuming that the chosen growth function was a good fit to the data and also that we had a suitable amount of data for each individual.

The hierarchical modelling approach provides benefits in the relatively straightforward analyses discussed above. However, it really comes into its own for the more complicated data in Chapter 5 and enables us to address difficult questions about the growth of different body components.

Unfortunately, the sampling procedure was not as straightforward for this data. We note here that this was more because of the relative lack of data at later ages than because of the more complicated model structure. Calculating the covariance matrices to use in our proposal distributions from early iterations worked reasonably well. However, care is needed with this approach to ensure that the proposals used initially lead to enough moves of a reasonable size being accepted.

For these reasons and also if refinements to the model are required, for example to include back fat measurements as discussed at the end of Chapter 5, we would recommend that expert assistance would be required for further analyses. After such collaboration, and for 'standard' model structures for which programs would be provided, analyses could probably be done, with care, within the company, but only by the 'trained' person/people.

6.2 BUGS and WinBUGS

An alternative to using the programs we have developed is the BUGS (Spiegelhalter et al, 1995) (or WinBUGS) program. These programs should work well for linear growth data when Gibbs sampling may be used throughout.

The programs now also include Metropolis algorithms, in BUGS Version 0.6 this is implemented using a simple (univariate) histogram-based proposal distribution (Ritter and Tanner, 1992) and any parameters which require this sampling must have bounded range. Unfortunately there is presently a bug in the multivariate normal sampler for BUGS Version 0.6 and so we were unable to try using this program for our models (as they include multivariate normal distributions).

It is planned that WinBUGS will eventually be able to use Metropolis sampling for any distribution for which it is needed. However, at the present time only a univariate Normal proposal density for sampling defined on the whole real line has been implemented. This univariate sampling can make convergence slow when the parameters are highly related. The same problem may occur for the sampling generally since each node is simulated in turn.

We used WinBUGS 1.2 to fit the Gompertz model from Chapter 4 of this thesis. The program uses an adaptive Metropolis algorithm for the θ_i vectors. For the first 4000 iterations the algorithm adapts so that suitable acceptance rates are achieved. This means that the first 4000 iterations must be discarded.

The program was easy to use and entering the model was straightforward

using a Doodle. Obviously, some time was required to become familiar with the program. We also note here that although (or because) the program is easy to use it would be easy to make mistakes because of the complex nature of MCMC sampling. A certain amount of expert knowledge is required in order to use the program properly. The educational version of the program was not sufficient for our model and so the (presently freely distributed) key for unrestricted use was required.

A single run of 6000 iterations with 'good' initial values was done. This consisted of 4000 from the adaptive phase to be discarded and a further 2000 from which to obtain results. This was about the same number of retained iterations as for our MATLAB analysis. Because of the adaptive procedure, the WinBUGS analysis required more iterations (6000 instead of our 3 repetitions of 1000). However, the two programs took almost the same amount of time to do these different numbers of iterations on the same machine. It should be possible to speed up the MATLAB programs by compiling them. This was not necessary for our analyses due to the relatively short run times.

The WinBUGS acceptance rates were suitable after about 1500 iterations. Nevertheless the adaptive procedure had to complete the whole 4000 iterations as that is how the procedure is defined. The final 2000 iterations were used to obtain parameter estimates and ranges. The results from the two programs were very similar. There was no evidence of a lack of convergence. There were autocorrelations up to about lag 20 for the elements of the θ_i. This aspect of the sampling was poorer than for our MATLAB sampling. This was because WinBUGS samples these elements univariately rather than as complete vectors for each individual.

WinBUGS has the advantage that the user does not have to write their own programs. On the other hand, for some models, we would expect slow convergence. For example, those in our Chapter 5 where the correlations between elements of the θ_i for an individual are much higher than for the Chapter 4 model. Also, for some models, WinBUGS will not be suitable. For example, because of

greater complexity or because of the distributions used. In these cases specially written programs will be needed.

We can recommend the use of WinBUGS 1.2 for the hierarchical models discussed in this thesis (with the caveats expressed above). More use would have been made of it in the present work if it had become available sooner.

Appendix A

Sampling methods

Throughout this work the Markov chain Monte Carlo sampling was implemented in MATLAB (The MathWorks, 1997).

For sampling from (multivariate) Normal distributions we used transformations of standard Normals generated by a built-in MATLAB function.

For Gamma distributions a function built into the MATLAB Statistics Toolbox was used.

For Wishart distributions we used the algorithm of Odell and Feiveson (1966) (which in turn used the MATLAB functions for sampling from Normal and Uniform distributions).

We checked that the MATLAB functions produced approximately the required distributions by using Kolmogorov-Smirnov tests of the empirical cdf based on large samples.

Bibliography

Beal, S.L. and Sheiner, L.B. (1989) *NONMEM user guide.* NONMEM Project Group, University of California, San Franciso.

Bennett, J.E. (1996) PhD thesis: *Bayesian analysis of population pharmacokinetic models.* Imperial College of Science, Technology and Medicine, London.

Berkey, C.S. (1982) Bayesian Approach for a Nonlinear Growth Model. *Biometrics*, **38**, 953-961.

Best, N.G., Cowles, M.K. and Vines, S.K. (1995) *CODA Manual version 0.30.* MRC Biostatistics Unit, Cambridge, UK.

Begall, S. (1997) The Application of the Gompertz Model To Describe Body Growth. *Growth, Development and Aging*, **61** 61-67.

Brody, S. (1945) *Bioenergetics and growth.* Rheinhold Publ. Corp., New York.

Chaloner, K. (1991) Bayesian residual analysis in the presence of censoring. *Biometrika*, **78**,3,637-44.

Chaloner, K. and Brant, R. (1998) A Bayesian approach to outlier detection and residual analysis. *Biometrika*, **75**,651-9.

Cowles, M.K. and Carlin, B.P. (1996) Markov chain Monte Carlo convergence diagnostics: a comparitive review. *J. Amer. Statist. Assoc.*, **91**, 883-904.

Davidian, M. and Giltinan, D.M. (1995) *Nonlinear Models for Repeated Measurement Data.* Chapman and Hall, London.

Davies, O.L. and Ku, J.Y. (1977) A re-examination of the fitting of the Richards growth function. *Biometrics*, **33**, 546-547.

Dawid, A.P., Stone, M. and Zidek, J.V. (1973) Marginalization paradoxes in Bayesian and structural inferences (with discussion). *J. R. Statist. Soc. B* **35**,189-233.

Escobar, M.D. and West, M. (1992) Computing Bayesian non-parametric hierarchical models. ISDS Discussion paper 92-A20. Duke University.

Fearn, T. (1975) A Bayesian approach to growth curves. *Biometrika*, **62**, 89-100.

Gelfand, A.E. (1995) Model determination using sampling-based methods. In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson and D. J. Spiegelhalter), pp. 144-161. London: Chapman and Hall.

Gelfand, A.E., Dey, D.K. and Chang, H. (1992) Model determination using predictive distributions with implementation via sampling-based methods. In *Bayesian Statistics 4*, (eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith), pp 147-167. Oxford: Oxford University Press.

Gelfand, A.E., Hills, S.E., Racine-Poon, A. and Smith, A.F.M. (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Am. Staist. Ass.*, **85**, 972-985.

Gelfand, A.E., Sahu, S.K. and Carlin, B.P. (1995a) Efficient parametrisations for normal linear mixed models. *Biometrika*, **82**,2, 479-88.

Gelfand, A.E., Sahu, S.K. and Carlin, B.P. (1996) Efficient parametrizations for generalized linear mixed models (with discussion). In *Bayesian Statistics 5*, (eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith). Oxford: Clarendon Press.

Gelfand, A.E. and Sahu, S.K. (1999) Identifiability, improper priors and Gibbs sampling for generalized linear models. *J. Amer. Statist. Assoc.*, **94**,No.445,247-253.

Gelman, A. and Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences (with discussion). *Statist. Sci.*, **7**,457-511.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995) *Bayesian Data Analysis.* London: Chapman and Hall.

Gelman, A., Roberts, G.O. and Gilks, W.R. (1996) Efficient Metropolis Jumping Rules. In *Bayesian Statistics 5*, (eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith). Oxford: Oxford University Press.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattn. Anal. Mach. Intel.*, **6**, 721-741.

Gilks, W.R. (1995) Full conditional distributions. In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson and D. J. Spiegelhalter), pp. 75-88. London: Chapman and Hall.

Gilks, W.R. and Wild,P. (1992) Adaptive rejection sampling for Gibbs sampling. *Appl. Statist*, **41**, 337-348.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1995) Introducing Markov chain Monte Carlo. In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson and D. J. Spiegelhalter), pp. 1-19. London: Chapman and Hall.

Glasbey, C.A. (1979) Correlated residuals in non-linear regression applied to growth data. *Appl.Statist.*, **28**, No 3, 251-259.

Goldstein, H. (1991) Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, **78**, 45-51.

Goldstein, H. (1995) *Multilevel Statistical Models.* Second Edition. Edward Arnold: London.

Goldstein, H., Healy, M.J.R. and Rasbash, J. (1994) Multilevel time series models with applications to repeated measures data. *Statistics in Medicine*, **13**,1643-1655.

Goldstein, H., Rasbash, J., Plewis, I., Draper, D. et. al. (1998) A user's guide to MLwiN. London, Institute of Education.

Hills, S.E. and Smith, A.F.M. (1992) Parameterization issues in Bayesian inference (with discussion). In *Bayesian Statistics 4*, (eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith), pp 227-246. Oxford: Oxford University Press.

Hobert, J.P. and Casella, G. (1996) The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *J. Amer. Statist. Assoc.* **91**,No.436,1461-1470.

Hodges, J.S. (1988) Some algebra and geometry for hierarchical models, applied to diagnostics. *J. R. Statist. Soc. B* **60**,3,497-536.

Hunt, R. (1982) *Plant Growth Curves.* Edward Arnold, London.

Koops, W.J. (1989) *Multiphasic analysis of growth.* Doctoral thesis, Department of Animal Breeding, Wageningen Agricultural University. P.O. Box 338, 6700 AH Wageningen, The Netherlands.

Koops, W.J. and Grossman, M. (1991) Applications of a Multiphasic Growth Function to Body Composition in Pigs. *Journal of Animal Science*, **69**, 3265-3273.

Kwakkel, R.P., Ducro, B.J. and Koops, W.J. (1993) Multiphasic Analysis of Growth of the Body and its Chemical Components in White Leghorn Pullets. *Poultry Science*, **72**, 1421-1432.

Lee, P.M. (1989) *Bayesian Statistics: An Introduction.* Oxford: Oxford University Press.

Lindley, D.V. and Smith, A.F.M. (1972) Bayes estimates for the linear model (with discussion). *J. R. Statist. Soc. B*, **34**, 1-41.

Lindstrom, L. and Bates, D.M. (1990) Nonlinear mixed-effects models for repeated measures data. *Biometrics*, **46**, 673-687.

MathSoft (1998) S-Plus 4.5, MathSoft, Inc.

MathWorks (1997) MATLAB, Version 5. The MathWorks, Inc.

Misra, R.K. (1980) Statistical comparisons of several growth curves of the von Bertalanffy type. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**, 920-6.

Munday, H.S., Earle, K.E. and Anderson, P. (1994) Changes in the Body Composition of the Domestic Shorthaired Cat During Growth and Development. *J.Nutrition*, **124**, 2622S-2623S.

Odell, P.L. and Fieveson, A.H. (1966) A numerical procedure to generate a sample covariance matrix. *J. Am. Statist. Ass.*, **61**, 198-203.

O'Hagan, A. (1985) Shoulders in Hierarchical Models. In *Bayesian Statistics 2*, (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith). Elsevier Science Publishers B.V. (North-Holland).

O'Hagan, A. (1988) Modelling with Heavy Tails. In *Bayesian Statistics 3*, (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith). Oxford University Press.

O'Hagan, A. (1994) *Kendall's Advanced Theory of Statistics, Volume 2B, Bayesian Inference.* Edward Arnold.

Pettit, L.I. (1986) Diagnostics in Bayesian model choice. *The Statistican*, **35**,183-190.

Pettit, L.I. and Young, K.D.S. (1990) Measuring the effect of observations on Bayes factors. *Biometrika*, **77**, 455-466.

Prosser, R., Rasbash, J. and Goldstein, H. (1991) *ML3: Software for Three-level Analysis. Users Guide*, Institute of Education, University of London.

Raftery, A.E. (1995) Hypothesis testing and model selection. In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson and D. J. Spiegelhalter), pp. 163-187. London: Chapman and Hall.

Raftery, A.E. and Lewis, S. (1992) How many iterations in the Gibbs Sampler? In *Bayesian Statistics 4*, (eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith). Oxford: Oxford University Press.

Ripley, B.D. (1987) *Stochastic Simulation.* New York: Wiley.

Ritter, C. and Tanner, M.A. (1992) Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler. *Journal of the American Statistical Association*, Vol. **87**, No. 419, 861-877.

Roberts, G.O., Gelman, A. and Gilks, W.R. (1997) Weak convergence and optimal scaling of the random walk Metropolis algorithm. *The Annals of Applied Probability*, **7**, 110-120.

Ross, G.S.J. (1990) *Nonlinear Estimation.* New York: Springer-Verlag.

Solomon, M.E. (1976) *Population Dynamics.* 2nd Edition. Studies in Biology No.18. Edward Arnold, London.

Spiegelhalter, D.J., Thomas, A., Best, N.G. and Gilks, W.R. (1995) *BUGS: Bayesian inference Using Gibbs Sampling*, Version 0.50. MRC Biostatistics Unit, Cambridge.

Spiegelhalter, D.J., Thomas, A., Best, N.G. and Gilks, W.R. (1995) *BUGS: Examples*, Version 0.50 Volume 2. MRC Biostatistics Unit, Cambridge.

Tanner, J.M. (1962) *Growth at Adolescence.* 2nd Edition. Blackwell Scientific Publications, Oxford and Edinburgh.

Tierney, L. (1994) Markov chains for exploring posterior distributions. *Ann. Statist.*, **22**, 1701-1762.

Van Lunen, T.A. (1994) *A study of the growth and nutrient requirements of highly selected pigs.* PhD thesis, University of Nottingham.

Venus, J.C. and Causton, D.R. (1979) Confidence limits for Richards functions. *Journal of Applied Ecology*, **16**, 939-49.

Walstra, P. (1980) *Growth and carcass composition from birth to maturity in relation to feeding level and sex in Dutch Landrace pigs.* No. 80-4. Communications Agric. Univ., Wageningen, Netherlands.

Wakefield, J.C. (1993) Contribution to discussion on 'The Gibbs sampler and other Markov chain Monte Carlo methods'. *J. Roy. Statist. Soc. B*, **55**, 3-102.

Wakefield, J.C., Gelfand, A.E. and Smith, A.F.M. (1991) Efficient generation of random variates via the ratio-of-uniforms method. *Statist. Comput.*, **1**, 129-133.

Wakefield, J.C. and Walker, S. (1994) Population models with a non-parametric random coefficient distribution. Technical report. Dept. of Mathematics, Imperial College.

Wakefield, J.C., Smith, A.F.M., Racine-Poon, A. and Gelfand, A.E. (1994) Bayesian Analysis of Linear and Non-linear Population Models by using the Gibbs Sampler. *Applied Statistician*, **43**, No. **1**, 201-221.

Whittemore, C.T. (1995) Modelling the Requirement of the Young Growing Pig for Dietary Protein. *Agricultural Systems*, **47**, 415-425.

Whittemore, C.T., Kerr, J.C. and Cameron, N.D. (1995) An Approach to Prediction of Feed Intake in Growing Pigs Using Simple Body Measurements. *Agricultural Systems*, **47**, 235-244.

Wishart, J. (1938) Growth Rate Determinations in Nutrition Studies with the Bacon Pig and their Analysis. *Biometrika*, **30**, 16-28.

Zeger, S. and Harlow, S. (1987) Mathematical models from laws of growth to tools for biologic analysis: Fifty years of *Growth*. *Growth*, **51**, 1-21.

Zeger, S. and Karim, M.R. (1991) Generalized Linear Models With Random Effects; A Gibbs Sampling Approach. *J. Am. Statist. Ass.*, **86**, 79-86.

Zullinger, E.M., Ricklefs, R.E., Redford, K.H. and Mace, G.M. (1984) Fitting sigmoidal equations to mammalian growth curves. *J. Mamm.*, 65(**4**), 607-636.