UNIVERSITY OF NAIROBI

COLLEGE OF BIOLOGICAL AND PHYSICAL SCIENCES

SCHOOL OF MATHEMATICS

REVIEW OF RANDOM EFFECTS ESTIMATION WITH APPLICATION TO THE GROWTH OF ACACIA SENEGAL TREES

BY

MACHARIA WA WAROTHE

A project submitted in partial fulfillment for a degree of Master of Science in Statistics in the school of Mathematics.

August 2009

NAIROBI-KENYA

Declaration

I the undersigned declare that this project is my original work and to the best of my knowledge has not been submitted for the award of degree in any other University.

MACHARIA WA WAROTHE

*Reg.No.*I56/71034/2007

Sign : ..... *Date*...17/08/2009.....

Declaration By Supervisors

This project has been submitted for examination with my approval as supervisor.

Supervisor

Prof. M.M. Manene
School of Mathematics
University of Nairobi

Sign: .....

Date.....17/8/2009.....

Date: August 13, 2009

Supervisor

Dr. Rosemary Nguti
School of Mathematics
University of Nairobi

Sign:.....

Date.....17/8/2009.....

Contents

Acknowledgement

I greetfully aknowledge support from a number of people who has greatly influenced the content and flow of this research work.

My deepest gratitude goes to Prof. M.Manene, Dr.R. Nguti and Dr. T. Achia for their tireless supervision throuhout this project. Without their effort it could has been very difficult for me to finish within the stipulated time frame. Their availability at all times and their guidance in various topics has been so inspirng.

Deep appreciation also goes to school of mathematics teaching staff who in their able hands, I have acquired Knowledge, Skills and Values through constructive, brief and critical approach at all levels.

Finally, my classmates, Kariuki, Nyacihi and Kyalo for their constructive discussion across many issues that helped me patch up ideas into a well laid down research.

Abstract

Estimation of random effects (variance components) is a method often used in population genetics and applied to areas such as animal and plant bleeding and growth. Scientists nowadays feel lost if confronted with huge set of different random effects estimation methods. This is especially because there exists no uniformly best method hence deciding which method should be used is difficult to take. This paper gives an overview of maximum likelihood and restricted maximum likelihood methods of estimating random effects applied to random coefficient models and demostrates them by applying them in determining variability in the growth of Acacia Senegal trees. We can say that both methods gives similar results with the dataset used. However random effects estimated using maximum likelihood method are slightly less than those estimated using restricted maximum likelihood methods. Random intercept and slope model is more appropriate to use in determination of variability in growth of Acacia Senegal trees than random intercept model.

Keywords: Maximum Likelihood, Restricted Maximum Likelihood, Random Effects, Random Intercept Model, Random Intercept and Slope Model.

Chapter 1

Introduction

In longitudinal studies, individuals are measured repeatedly through time. Observations are taken in two or more occasions. Longitudinal data can be collected either prospectively ,following subjects over time, or retrospectively, by extracting multiple measurements on each person from historical situations. The defining feature of a longitudinal dataset is repeated observations on individual, enabling direct study of change. Longitudinal data require special statistical methods because the set of observations on one individual tends to be intercorrelated. Most longitudinal analysis are based on a regression model. The natural experimental unit is not the individual measurement, but sequence of measurements on an individual. Replicatioon here means number of subjects.

There are different approaches that can be adopted to longitudinal data analysis with repeated measurements. A simpler and often effective strategy is to reduce the repeated values into one or two summaries and analyse each summary variable as a function of covariates.

In practice longitudinal data are highly unbalanced in the sense that not equal number of measurements are available for all subjects and/or mearsurements sre not taken at fixed time points. As such Seber (1984)and Taylor (1987) observed that many longitudinal datasets cannot be analyzed using multivariate regression techniques. A natural alternative arises from observing that subject-specific longitudinal profiles cannot often be approximated by linear regression functions. One hereby summarises the vector of repeated measurements for each subject by a vector of relatively small number of estimated subject-specific regression coefficients and in the second stage multivariate regression techniques are used to relate estimates to known covariates.

In general linear mixed effect model,inferences are based on marginal distribution of response.It

assumes that the vector of repeated measurements on each subject follows a linear regression model where some of the regression parameters are population specific, that is same for all subjects, wheres others are subject specific. Subject specific parameters are assumed to be random and are called random effects.

Mixed models may contain interaction between fixed and random effects. An interaction between a random subject effect and a fixed effect associated with an explanatory variable might be used to allow the coefficient of the variable to be different for each subject. This is termed as random coefficient model. For this model Random intercept model is a linear mixed effect model where the only subject specific effects are intercepts. In random intercept model its assumed that all variability in subject specific slopes can be ascribed to treatment differences hence can be obtained by ommitting the random slopes.The subject specific profiles are then assumed to be linear with subject specific intercepts but with the same slopes within each treatment group. We have two covariance mtrices.That is, the random effect covariance matrix D which is a scalar and residual covariance matrices. The implied covariance assumes constant variance over time as well as equal correlation between any two measurements from the same subject. For highly unbalanced data with many repeated measurements per subject, one usually assumes that random effect can account for most of variation in the data assume that the remaining error components have a very simple covariance structure.

Various methods for estimating variance components exist. Among them are maximum generalized least square estimation (GLS), maximum likelihood (ML) estimation and restricted maximum likelihood (REML) estimation .ML and REML are based on maximising likelihood functions corresponding to statistical model underlying experimental design. The REML is based on maximising the portion of likelihood that is invariant to the fixed effects.

ML is a general method for estimating parameters from realisation of random variables. MLE of parameter is the value for which the parameter attains maximum. ML has a long celebrated history going back to Fisher (1922). However it had not been used in mixed model analysis until Hartley and Rao(1967) because estimation of variance component in linear mixed effect model was not easy to handle computationally. Its easier to work with log-likelihood.The vector of first order partial derivatives of log-likelihood is called the scoring vector. A standard approach to maximum likelihood estimation is to find all the roots of the scoring vector and to explore the behaviour of log-likelihood on the boundary of the parameter space and at the point where the

scoring vector is not defined. These are the only possible locations of MLE. Typically if the score vector has roots, then the root is a MLE. Because of computational nature of complex log-likelihood equations, the variance components in general are estimated recursively using either the Newton-Raphson or Fisher scoring method. MLE often produces biased estimators of variance components. With MLE one has to estimate all the parameters involved. Patterson and Thompson (1971) introduced REML as a technique for estimating variance components in a random effects or linear mixed effect model. REML can estimate the parameter of main interest without having to deal with nuisance parameters. REML is defined through a transformation matrix although it does not depend on this transformation matrix as shown by Harville (1974). Hence choice of transformation matrix is not unique and one does not want the estimator to depend on transformation matrix. The idea behind REML estimation is to consider likelihood of linear combination of responses that do not depend on mean parameters. Restricted log-likelihood is a function of variance component only. Thus REML is a method of estimating variance components. However once the REML estimator of variance components is found other parameters are usually estimated the same way as the MLE.

In order to properly specify the model, the covariance parameters are necessary though they are not the parameters of interest for drawing conclusions. Doing a good job in choosing a covariance model improves the efficiecy of the fixed effect estimates and allows for more accurate confidence intervals and hypothesis tests. This is done by selecting a covariance matrix that is close to or include the actual covariance matrix as a special case and vector parameter has few elements as possible.

This paper discusses ML and REML methods of estimation of random effect(variance components) for random coefficient models and apply them in determination of variability in the growth of Acacia Senegal trees.

1.1 Background Information on Growth of Acacia Senegal Trees

Growth and yield models are used to forecast the development of the forest resources in forest management and planning.Long term forecasts about development of forests are needed for decision making in forest policy. The growth and yield model selected should have valid biological

and ecological basis inorder to produce reasonable growth predictions even when applied outside the range of data they are developed from. The input of the growth models must be consistent and compatible with the data available. In forest management planning evaluation of alternative management schedule is an essential part. Growth and yield models should be able to predict reliably the effect of various treatments on development of managed stands. In order to have a reliable model, data from forest inventories as well as from purpose designed experiments are needed. Designed experiments carried out provides information about the effects of various treatments on growth and yield, which is important in developing the model structure.

Acacia Senegal tree grows in arid and semi-arid lands where soils are shallow,rocky with low organic matter, low moisture content and low water holding capacity. Gum arabic is the sap of the Acacia Senegal tree, and some other African species of Acacia, occurring as an exudate from the trunks and branches. It is normally collected by hand when dried, when it resembles a hard, amber-like resin normally referred to as 'tears'. Gum arabic powder is widely used in the food industry, as an emulsifier, thickener, flavour encapsulator and thickening agent. Grieve's classic 'A Modern Herbal': 'Gum Acacia is a demulcent and serves by the viscidity of its solution to cover and sheathe inflamed surfaces.' Mucilage of Acacia is a nearly transparent, colorless or scarcely yellowish, viscid liquid, having a faint, rather agreeable odour and an insipid taste.' 'It is employed as a soothing agent in inflammatory conditions of the respiratory, digestive and urinary tract, and is useful in diarrhoea and dysentery. 'Gum Acacia is highly nutritious. During the time of the gum harvest, the Moors of the desert are said to live almost entirely on it, and it has been proved that 6 oz. is sufficient to support an adult for twenty-four hours. It is related that the Bushman Hottentots have been known in times of scarcity to support themselves on it for days together. In many cases of disease, it is considered that a solution of Gum Arabic may for a time constitute the exclusive drink and food of the patient.'King's 1898 Dispensatory: 'Gum arabic is nutritive and demulcent, and exerts a soothing influence upon irritated or inflamed mucous tissues, by shielding them from the influence of deleterious agents, atmospheric air, etc. On this account it has been used in diarrhoea and dysentery, to remove tenesmus and painful stools, in catarrh, cough, hoarseness, gonorrhoea, ardor urinae, etc.' It exerts a soothing influence upon all the surfaces with which it comes in contact.'In dispensing, Mucilage of Acacia is used for suspending insoluble powders in mixtures, for emulsifying oils and other liquids which are not miscible with water, and as an ingredient of many cough linctures. 'Equal parts of pulverized alum

and gum arabic form a good preparation to check hemorrhages from small cuts, wounds, etc. Externally, the application of its solution to burns and scalds has proved serviceable, repeating it until a complete coating is secured.' 'Mucilage of acacia is soothing to burns and scalds of the mouth and alimentary canal, and may be used as a demulcent after poisoning by irritant and corrosive poisons'.

There has been a wide spread over-exploitation of plant resources leading to an alarming rate of diminishing these tree species. Due to its economic importance, there has been a felt need to sustain and improve its existence. An experiment was carried out on seedlings to study the performance due to the effects of silvicultural treatments to different soil types.

In this paper, we determine variability in growth of acacia senegal trees grown in a glass house simulating arid and semi-arid conditions.The variability in growth to be identified was in the initial individual tree heights and rate of growth.

1.2 Literature Review

1.2.1 Estimation of Random Effects

Variability mathematically is determined by estimating the variance components using the appropriate method. Rasch, D. and Masata, 0. (2006) observed that there exists no best method of variance components estimation. Various methods exists. These incudes, least square methods, maximum likelihood (ML) method, restricted maximum likelihood method (REML), minimum norm quadratic unbiased (MINQUE) method, bayes methods among others.

Herbach (1959) derived a real maximum likelihood(solution restricted to parameter space). Anderson and Bancroft (1952) introduced a restricted maximum likelihood method. This method uses a translation invariant restricted likelihood function depending on the variance components to be estimated only and not on fixed effects.This restricted likelihood function is a function of sufficient statistics for the variance components.REML is derived with respect to variance components under restriction that the solution are non-negative.

In linear mixed effect models for repeated measures analysis, its prefered to estimate the variance components and variance-covariance according to REML criterion which compensate for estimation of fixed effect parameters when estimating the random effects. Harville (1977) suggested that REML may be used even when the data are not normally distributed since REML can be

viewed as iterated MINQUE and theory of MINQUE does not depend on any assumption of normality. Boos, D.D. and Gumpertz, M L. (2001) showed why REML estimation reduce the bias of variance and covariance parameter estimates in small samples.

Random intercept models and Random intercept and slope models, both of which are examples of random coefficient models, for longitudinal data are contained in linear mixed effect models. Cheng Hsiao and Pesaras, M.H. (2004) categorised the random coefficient models into two. That is, stationary random coefficient model where random variables have constant mean and variance or covariance and non-stationary random coefficient model where coefficient vectors are not regarded as having constant mean and variances or covariances. Random intercept model is an example of stationary random coefficient model whereas random intercept and slope model is an example of non-stationary random coefficient model. Bates. D.M. (1993) noted that computational for ML or REML estimation of parameters in linear mixed effect models are greatly enhanced by expressing the variance-covariance matrix of random effects at each level in terms of square root of inverse of relative variance matrix. Using this formulation and matrix decomposition, the profiled log-likelihood or profiled restricted log-likelihood can be gotten and calculated. These expression give an indication of suitable starting values of variance-covariance parameters.

Growth is a function of time, and for trees, growth is not infinite. Growth is considerably influenced by natural and environmental conditions and illegal or unauthorized activities by Man. As such growth mostly inhibits variability. Thus its always in order to identify sources of variabilty at every stage of growth. Different treatment such as micro-nutrients treatments play a major role in tree growth. When all other disturbing variations are removed, one can test whether after application of nutrients, the initial variability in the growth affect the final size of the tree. Different micronutrients plays different roles in the growth of trees. Its always acceptable to select micro-nutrients that gives higher yields and speed the rate of growth. For acacia senegal best nutrients are those that give higher gum arabic yields.

1.3 Study Objectives

1. To give an overview of maximum likelihood and restricted likelihood estimation methods for random effects in random coefficient models.

2. To compare the results the two method gives when applied to the analysis of random effects in the growth of Acacia Senegal.

3. To determine variability in the growth of Acacia Senegal trees.

1.4 Statement of problem

Estimation of random effects are needed whenever interest is in prediction of subjects specific evolution though in practice one is primarily interested in estimating fixed effects parameters. Covariance parameters are necessary in order to properly specify the model.

Chapter 2

Methodology

Two random coefficient models, that is, random intercept and random intercept and slope models were discused and applied to the growth of Acacia Senegal trees. Maximum likelihood and restricted maximum likelihood as some of the random effect estimation method were reviewed. The design, data management and analysis were as follows.

2.1 Design

To carry out the experiment split-split plot design was used. This comprisedof two main plots(media) namely soil and vermiculite, three subplots (sites) namely Kimorok(KK),Kapkun(KN) and Solit(ST)and thirteen sub-sub plots(micro-nutrients treatments). The treatments (micro-nutrients) used included Boron, Zinc, Manganese, Molybdenum, Copper and Iron with each having either low(-),normal(n) or high(+) concentrations. A control consisting of all the micro-nutrients with normal concentrations was also cooporated. Table 2.1 show the treatment structure.

A total of 156 seedlings were randomized to either vermiculite or soil. The experiment was replicated twice. It was carried out in a glasshouse to simulate the arid and semi-arid climate conditions. The micro-nutrients were applied after every two weeks. Measurements were taken for the heights and diameter at ground level(dgl) over time. This was a repeated measures experiment because the same seedlings were measured at each time. Eight seedlings dried within the study period (four on soil medium) and were not used in analysis. This yielded unbalanced design.

Table 2.1: Table of Treatment structure

Treatment No.	Copper	Iron	Manganese	Zinc	Boron	Molybdenum
1	n	n	n	n	n	n
2	-	n	n	n	n	n
3	+	n	n	n	n	n
4	n	-	n	n	n	n
5	n	+	n	n	n	n
6	n	n	-	n	n	n
7	n	n	+	n	n	n
8	n	n	n	-	n	n
9	n	n	n	+	n	n
10	n	n	n	n	-	n
11	n	n	n	n	+	n
12	n	n	n	n	n	-
13	n	n	n	n	n	+

2.2 Data management

74 seedlings on soil medium were used for analysis. The four that dried were excluded from the model. The heights only were considered and not the diameter. Trees are assumed to have a growth that takes a sigmoid curve. The initial heights were considered from the day the treatments were applied inorder to establish the the variability at time t=0. The height were then measured with time for two years. This was done to establish the rate of growth which was used to determine the variabilty of rate of growth. The data was meant to be used for prediction over a period of 25years in intervals of 1 year hence constituting equally spaced time points. Soil medium was chosen so as to incorporate the site characteristics excluding the vermiculite from the analysis. The random intercept model and random intercept and slope model model were fitted to the available data observed over a period of two years and predictions were made. Wiihelm lotshhert and Gerhard Beese, (1983) observed that Acacia Senegal tree grows up to a maximum 15 metres tall. This constitute the carrying capacity of the model . A data set was generated using R program.

The sivilcultural treatments were coded as they appear on Table 2.1 in design section. The

sites;Kapkun(KN),Solit(ST) and Kimorok(KK) were coded 1,2 and 3 respectively. The fields in the data set height(see appendix) include;

Treatment- refers to the thirteen micro-nutrients treatments

Site- refers to the site from which the seeds were extracted from with their corresponding soil sample

Height- refers to the height measurement of each tree seedling in millimetres from the initial day of treatment application

Day- refers to the time measurements used in the analysis

Tree- refers to the identification of each tree seedling in the experiment.

This data set was then used in determination of variation both at intial time and later during growth analysis.

2.3 Method of analysis

The method of analysis was based on the method proposed by Goldstein(1995) which is a multilevel form of representation as shown in chapter 3. The steps for analyzing the data were as follows:

For random intercept model given by

$$y_{ij} = b_{0i} + b_{1i}t_{ij} + \varepsilon_{ij}$$

where

$b_{0i} = \beta_0 + \upsilon_{0i}$

$b_{1i} = \beta_1$

the variaton of the mean (b_{0i}) of intercepts (initial height of trees) D, around population mean $(b_{0i} \sim \mathcal{N}(0, D))$ and variation of observations around the subject specific mean, $(Y_{ij} \sim \mathcal{N}(0, \Sigma_i))$ were evaluated. The variance D shows the variation between individual tree height at t=0. Σ_i show the measurement error.

The total variance was then calculated which is given by $d_{11} + \sigma^2$, since random intercept model is assumed to be comprising of two constant variance components.That is, variation due to intercepts and the measurement error. Then using the values of D and σ^2 the intraclass correlation or reliability coefficient given by

$$\rho_i = \frac{d_{11}}{\sigma^2 + d_{11}}$$

was calcurated. The fitting of the model was done using R program and estimation of the variance components was done using restricted maximum likelihood estimation techniques.
For random intercept and slope model given by

$$y_{ij} = b_{0i} + b_{1i}t_{ij} + \varepsilon_{ij}$$

where

$$b_{0i} = \beta_0 + \upsilon_{0i}$$

$$b_{1i} = \beta_1 + \upsilon_{1i}$$

the variation of the variation of the mean around subject specific mean, $(Y_{ij}/b_{0i} \sim \mathcal{N}(0, \Sigma_i)$, and covariance matrix D of random effects were evaluated. In this case D consists of four variance components. That is d_{11} which is variance of intercepts (spread around average height of the tres at t=o), d_{22} which represents variation of slope (spread around the rate of growth across all trees) and d_{12} which is correlation between d_{11} and d_{22}(degree to which individual tree height at t=o and rate of growth co-vary). From this covariance matrix which is given by

$$V_i = Z_i D Z_i' + \Sigma_i$$

and the corresponding correlation structure were obtained.

Chapter 3

Models

Random coefficient Model

3.0.1 Linear effect mixed model

The general linear mixed effect model for longitudinal data is given by

$$y_{ij} = Z'_{ij}\alpha_i + X'_{ij}\beta + \varepsilon_{ij} \tag{3.1}$$

where

y_{ij} is observed data for subject i at time j

Z_{ij} and X_{ij} are $n_i \times q$ and $n_i \times p$ known covariates matrices

β is p-dimensional $p \times 1$ vector of unknown population parameters (fixed effects)

α_i is q-dimensional $(q \times 1)$ vector of uknown individual effects

ε_{ij} is the error term

$\alpha_i \sim \mathcal{N}(0, D)$

That is α_i is assumed to be normally distributed with mean 0 and covariance matrix D.

$\varepsilon_{ij} \sim \mathcal{N}(0, \Sigma_i)$

$i \ldots N$

D is a $q \times q$ covariance matrix with i, j element $d_{ij} = d_{ji}$

Σ_i is $(n_i \times n_i)$ covariance matrix which depends on i only through its dimension n_i. That is, the set of unknown parametersin Σ_i will not depend on i.

It follows from equation 3.1 that conditional on the random effect α_i, y_{ij} is normally distributed with mean $X_i\beta + Z_i\alpha_i$ and with covariance matrix Σ_i. .

let $f(y_{ij})/\alpha_i$ and $f(\alpha_i)$ be the corresponding density functions, then the marginal density function of y_{ij} is then given by

$$f(y_{ij}) = \int f(y_{ij}/\alpha_i) f(\alpha_i) d\alpha_i$$

which is the density function of n_i-dimensional normal distribution with mean vector $X_i\beta$and with covariance matrix

$$V_i = Z_i D Z_i' + \Sigma_i$$

which show the dependance of mean structure and covariance structure on the covariates X_i and Z_i respectvely. The conditional regression function of 3.1 takes the form

$$E(y_{ij/\alpha_i}) = Z_{ij}'\alpha_i + X_{ij}'\beta \tag{3.2}$$

$Z_{ij}'\alpha_i$ comprises the random effects portion of the model and

$X_{ij}'\beta$ comprises the fixed effect portion

A linear mixed effect model where some of parameters are normally distributed random variables is called random coefficient model.

3.0.2 Derivation of random coefficient model

From the linear mixed effect model, suppose that $q = p$ and $Z_{it} = X_{ij}$.In this case the linear mixed effect model reduces to random coefficient model of the form

Hence

$$\begin{aligned} E(y_{ij}/\alpha_i) &= X_{ij}'(\alpha_i + \beta) \\ &= X_{ij}\beta_i \end{aligned} \tag{3.3}$$

where β_i given by $\alpha_i + \beta$ are random vectors with mean β.

hence

$$y_{ij} = X_{ij}\beta_i + \varepsilon_{ij}$$

In the model y_{ij} is a $t \times 1$ vector representing i^{th} individual's t repeated measurements on response variable, X_{ij} is a $t \times p$ matrix of t corresponding reapeated measurements on p explanatory variables, β_i is $p \times 1$ vector of individual specific regression coefficients corresponding to p explanatory variables and ε_{ij} is $t \times 1$ vector of random errors.

So from this model we can see that each individual tree has a different relationship between its outcome measuerments, height and time variable defined by the p regression parameters of β_i.

For the purpose of inference we need to make the following assumptions.

A1. $\beta_i \sim \mathcal{MVN}(\beta, \Sigma_i)$

That is β_i comes from multivariate normal distribution of dimension p with mean vector β, and variance-covariance matrix Σ_i. $\beta_i's$ are indepedent across individual trees. A2. $\varepsilon_{ij} \sim \mathcal{MVN}(0, \sigma_i I_t)$

That is ε_{ij} $j = 1, ..., t$ comes from a univariate normal population with mean 0 and variance σ^2. All of $\varepsilon'_{ij}s$ are indepedent of each other. This is referred to as the conditional indepedence assumption.

The random coefficient model can be interpreted as a two stage sampling model. In the first stage, one draws the ith subject from the population that yields a vector of parameters β_i. From the population this vector has mean

$$E(\beta_i) = \beta$$

and variance

$$var(\beta_i) = D$$

At the second stage, one draws T_i observations for the ith individual, conditional on having observed β_i.

The mean of observations is given by

$$E(y_{ij}/\beta_i) = X_{ij}\beta_i$$

and the variance is given as

$$Var(y_{ij}/\beta_i) = \sigma_i$$

Putting this two stages together yields,

$$\begin{aligned} E(y_{ij}) &= X_{ij}E\beta_i \\ &= X_{ij}\beta \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} Var(y_{ij}) &= E(Var(y_{ij}/\beta_i)) + Var(E(y_{ij}/\beta_i)) \\ &= \Sigma_i + Var(Z_{ij}\beta_i) \\ &= \Sigma_i + X'_{ij}DZ_{ij} \\ &= V_i \end{aligned} \tag{3.5}$$

However certain variation of two stage interpretation of the random coefficients model lead to other forms of random effect models in equation 3.3. To illustrate, we may take the columns of Z_{ij} to be a strict subset of column of X_{ij}. This is equivalent to assumming that certain components of β_i associated with Z_{ij} are stochastic wheres other components are associated with X_i (but not Z_{ij}) are non stochastic.

By convention we assume that the mean of random effects α_i is known and equal to zero.Thus we arbsorb the additional terms into the $X_{ij}\beta$ portion of the model. Thus its customary to include those explanatory variables in Z_{ij} covariate matrix as part of X_i covariate matrix.

In our case we have used Z_{ij} and X_{ij} interchangably. In random coefficient model the same model holds for all individuals, but the parameters of the model do not have fixed values. Instead for each individual the value of each parameter is regarded as a random observation from its own distribution.

Random coefficient models are of different types. We have used in our case random intercept model and random intercept and slope modelel to model the covariance matrix for the growth of acacia senegal trees under varying conditions.

3.0.3 Random intercept model (RI)

Random intercept model is a linear mixed effect model where the only subject-specific effect is intercept. The random effect covariance matrix D is a scalar and Z_i are of the form 1_{ni}

(n_i-(dimensional vector of ones), with as many 1's as the number of trees used in height determinations ($i = 1...39$). Residual variance matrices are of the form $\Sigma_i = \sigma^2 I_{ni}$ (conditional indepedence). D is variance term which indicates how much spread there is around the initial height(intercepts). A simple extension of regression model for our data is given by

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \varepsilon_{ij} \tag{3.6}$$

To allow for the influence of each individual (tree) on it's repeated outcomes the model becomes

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \upsilon_{0i} + \varepsilon_{ij} \tag{3.7}$$

where

β_0 is intercepts or initial average height of all the trees (population initial height)

β_1 is linear change across time or average rate of growth (population rate of growth)

y_{ij} is outcome variable or height of i^{th} tree measured at time j

t_{ij} is indepedent time variable

υ_{0i} is the influence of individual (tree), i on it's repeated observations. If an individual (tree) have no influence on their repeated outcome, $\upsilon_{0i} = 0$. However its more likely that individual (trees) will have positive or negative influences on their repeated measures and so υ_{0i} will deviate from zero.

To better reflect how this model characterises as individuals influence on their observations, its helpful to represent the model in hierachial or multilevel form as observed by Goldstein (1995). In our case we have represented the models in this manner.

The model is partitioned into the following

Within subject (or level-1) model given by

$$y_{ij} = b_{0i} + b_{1i} t_{ij} + \varepsilon_{ij} \tag{3.8}$$

and between subject(or level-2) model

$$b_{0i} = \beta_0 + \upsilon_{0i} \tag{3.9}$$

$$b_{1i} = \beta_1$$

where

b_{0i} is intercept or initial height of i^{th} individual (tree)

b_{1i} is the individual (tree) rate of growth.

The other parameters remains as stated above. Level-1 indicates that individual (tree) response at time j is influenced by its initial level b_{0i} and time trend or rate of growth (slope) b_{1i}

The level-2 model indicates that individual tree initial level is determined by population initial level β_0 plus a unique contribution for that individual tree υ_{0i}. Thus each individual (tree) has their distinct initial level. Conversely, individual (random deviation from mean slope) rate of growth are assumed to be the same. All are equal to population rate of growth (slope) β_1 . That is, each individual trend line is parallel to population trend (average rate of growth) determined by β_0,initial population height and β_1, population rate of growth . The difference between each individual trend's and population trend is υ_{0i}, which is constant across time.

The hierachial representation shows that just as within subjects (level-1) covariates can be included in the model to explain variation in level-1 outcomes (y_{ij}), between subjects (level-2) covariates can be included to explain variation in level-2 outcomes (subjects intercept $b_{0i}$and slope b_{1i}).

Random intercept model implies a compound symmetry assumption for variances and covariances of longitudinal data. That is both variances and covariances are assummed to be same across time namely

$$Var(y_{ij}) = \sigma_\upsilon^2 + \sigma^2$$

$$cov(y_{ij}y_{ij'}) = \sigma_\upsilon^2$$

where $\upsilon_{0i} \sim \mathcal{N}(0, \sigma_\upsilon^2)$

$$\rho_i = \frac{\sigma_\upsilon^2}{\sigma^2 + \sigma_\upsilon^2}$$

between any two measurements from the same subject. This covariance structure is called compound symmetry.

Let $\sigma_\upsilon^2 = d_{11}$. In general in random intercept models, observations Y_{ij} are modelled as having two sources of variation. One is variation of the means b_{0i},which is zero around their population mean

$$b_{0i} \sim \mathcal{N}(0, D)$$

and second is the variation of observations around the subject-specific mean

$$Y_{ij}/b_{0i} \sim \mathcal{N}(0, \Sigma_i)$$

Observations Y_{ij} vary around a different value b_{0i} for each individual. These values are the intercepts of the line each individual responses vary around, since the true rate of change are zero. The mean of $Y_{ij} \ldots Y_{ij'}$ is an estimate $\hat{b}_{0i}$ of the unkown subject-specific intercept b_{0i}. The sets of intercepts (initial heights) $b_{0i}, i = 1, \ldots, n$ are a sample from the population . The RI assumes observations bouncing around a flat line $Y_{ij} = b_{0i} + \varepsilon_{ij}$.

Random intercept model implies a compound symmetry assumption for variances and covariances of longitudinal data. That is both variances and covariances are assummed to be same across time. Observations have constant variance $d_{11} + \sigma^2$ and constant d_{11}. That is

$$\begin{aligned} Var(Y_{ij}) &= Var(b_{0i} + \varepsilon_{ij})) \\ &= Var(b_{0i}) + Var(\varepsilon_{ij}) \\ &= d_{11} + \sigma^2 \end{aligned} \tag{3.10}$$

where d_{11} is the individual variation from initial mean (height of all the trees) and $\sigma^2 = \sigma^2 In_i$ We assumes that all variability in rate of growth of individuals can be ascribed to treatment differences hence can be obtained by omittimng random rate of growth (slopes). Implied marginal covariance structure ($\Sigma_i = \sigma^2 I_{ni}$) is given by

$$\begin{aligned} V_i &= (1)d_{11}(1) + \sigma^2) \\ &= d_{11} + \sigma^2 \end{aligned} \tag{3.11}$$

Hence the implied covariance matrix is compound symmetry.

The covariance between Y_{ij} and Y_{il} simplifies to

$$
\begin{aligned}
Cov(Y_{ij}, Y_{ij'}) &= Cov(b_{0i} + \varepsilon_{ij}, b_{0i} + \varepsilon_{ij'}) \\
&= Cov(b_{0i}, b_{0i}) + Cov(b_{0i}, \varepsilon_{ij'}) + Cov(\varepsilon_{ij}, b_{0i}) + Cov(\varepsilon_{ij}, \varepsilon_{ij'}) \\
&= Cov(\varepsilon_{ij'}, \varepsilon_{ij'}) \\
&= Var(\varepsilon_{ij'}) \\
&= d_{11} \qquad (3.12)
\end{aligned}
$$

. Under this model the observations have constant variance and constant covariance. Thus this model implies compound symetry covariance structure.

Expressing the covariance as a corelation yields the intraclass correlation which is ratio of the individual variance, d_{11} to total variance $\sigma^2 + d_{11}$. The implied covariance structure assumes constant variance $\sigma^2 + d_{11}$ over time as well as equal correlation.

The intraclass correlation thus is given by

$$\rho_i = \frac{d_{11}}{d_{11} + \sigma^2}$$

Which is correlation between height for two individuals in the same subject.ρ_i is large when inter individual variability d_{11} is large in comparison to intraindividual(tree), σ^2, variability. This coefficient represents degree of association of data within individual and specifically indicates the proportion of variance in the data attribitable to individuals tree. Random intercept models that includes autocorrelated erros provides variance-covariance structure that is general than compound symetry. So its Random intercept model with indepedent errors that implies compound symmetry.

3.0.4 Random intercept and slope model (RIAS)

It is unlikely that rate of change (growth) at a given time point is the same for all individuals . Its more likely that individual differ in their time trends. Not everyone change at the same time. Compound symetry assumption is usually untenable for most longitudinal data. In general measurements at points close in time tends to be more highly correlated than measurements further separated in time. In many studies subjects are more similar at baseline and they grow

at different rates across time. Thus its natural to expect that variability will increase with time. Hence a more realistic model is to allow both initial heights and time trend to vary by individuals tree

. The RIAS model generarizes RI model to observations Y_{ij} falling around subject specific lines $b_{0i} + b_{1i}$ with unkown subject specific intercepts b_{0i} and subject-specific slopes b_{1i}. Thus within-subject (level 1) model remains the same as in RI. That is

$$y_{ij} = b_{0i} + b_{1i}t_{ij} + \varepsilon_{ij}$$

and

level-2 of random intercept model is augmented as

$$b_{0i} = \beta_0 + \upsilon_{0i}$$

$$b_{1i} = \beta_1 + \upsilon_{1i}$$

where

β_0 is overall population intercept (average height of all trees at initial time).

β_1 is overall population slope (mean growth rate across all tree).

$\upsilon_{0i} \sim \mathcal{N}(0, \sigma^2_{\upsilon_0})$ is random deviation from the mean (height for tree i)

$\upsilon_{1i} \sim \mathcal{N}(0, \sigma^2_{\upsilon_1})$ is rate of growth random deviation from the mean (growth rate across time for tree i)

$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ is the residual error term.

$cov(\upsilon_{0i}, \upsilon_{1i}) = \sigma_{\upsilon_0\upsilon_1}$ With two random individual specific effects, the population distribution of intercept and slope deviations is assummed to be bivariate normal $\sim \mathcal{N}(0, D)$ with random effects variance-covariance matrix as

$$D = \begin{pmatrix} \sigma_{\upsilon_0} & \sigma_{\upsilon_0\upsilon_1} \\ \sigma_{\upsilon_0\upsilon_1} & \sigma_{\upsilon_1} \end{pmatrix} \tag{3.13}$$

We make assumption that both υ_{0i} and υ_{1i} are indepedent from residual error but we cannot assume that they are indepedent because a simple shift in the time variable will change the correlation of the two random effects.

This model can be thought of as a individual trend or change model since it represents measurements of Y as a function of time both at individual (υ_{0i} and υ_{1i})and population (β_0 and

β_1)levels.

The intercept parameters indicates starting point and the slopes indicates degree of change over time. The population intercept and slope parameters represent overall (population) trend, while individual parameters express how subjects deviate from population trend. Because slopes varies for individuals, this model allows the possibility that some individuals do not change across time while others can exhibit dramatic change.

The population trend is average across individuals and variance terms indicates how much heterogenity there is in the population. Variance term σ_{υ_0} indicates how much spread there is around the population intercept and σ_{υ_1} represent the spread in slopes. If each individual deviation from the population trend is only due to random error, these variance terms will approach zero. Alternatively as each individual deviation from population is non random, but characterized by individual trend parameters υ_{0i} and υ_{1i} as being non zero, these variance terms will increase from zero. Covariance term $\sigma_{\upsilon_0\upsilon_1}$ represent degree to which the individual intercept and slope parameters co-vary. Positive covariance term would suggests that individuals with intial higher values have greater positive slopes while negative covariance would suggest the opposite.

In RIAS there is a different intercept and slope pair $b_i = (b_{0i}, b'_{1i})$ for each subject. The b_i, $i = 1, \ldots, n$ form a sample from the population of possoble (intercepts, slopes) pairs. As there are two random effects we assume bivariate normal distribution. The mean $\alpha = (\alpha_1, \alpha_2\prime)$ of $b'_i s$ is the population mean of intercept-slope pair. The population mean time trend of the observations at time t_{ij} is the line

$$E[Y_{ij/\alpha}] = \alpha_1 + \alpha_2 t_{ij}$$

Let $\sigma_{\upsilon_0} = d_{11}$, $\sigma_{\upsilon_1} = d_{22}$, $\sigma_{\upsilon_1\upsilon_1} = d_{12} = d_{21}$

$$D = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix} \tag{3.14}$$

The RIAS model has four covariance parameters. The variance σ^2 is the error variance of the observations Y_{ij} around subject-specific line $b_{1i} + b_{2i}t_{ij}$.The 2×2 matrix D has three unique parameters $d_{11}, d_{12} = d_{21}$and d_{22}. The parameter d_{11} is the variance of initial heights (intercepts) b_{0i} in the population. Similarly d_{22} is the population variance of the rate of growth (slopes) b_{1i} and d_{12} is the covariance of the slopes and intercepts.

Unconditionally, that is, ignoring the individual-specific intercept and slope parameters b_i, and thinking in terms of the parameters σ^2 and D, the marginal variance of Y_{ij} is given by

$$Var(Y_{ij}/\sigma^2, D) = \begin{pmatrix} 1 & t_{ij} \end{pmatrix} \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix} \begin{pmatrix} 1 \\ t_{ij} \end{pmatrix} + \sigma^2$$

Thus

$$Var(Y_{ij}/\sigma^2, D) = \sigma^2 + d_{11} + 2t_{ij}d_{12} + t_{ij}^2 d_{22} \tag{3.15}$$

and the covariance between Y_{ij} and $Y_{ij'}$ is

$$Cov(Y_{ij}, Y_{ij'}) = d_{11} + (t_{ij} + t_{ij'})d_{12} + t_{ij}t_{ij'}d_{22}$$

The marginal variance $Var(Y_{ij}/\beta_0, \sigma^2, D)$ is quadratic in time. The coefficient of t_{ij}^2 is d_{22}, the population variance of slopes; the coefficient of $2t_{ij}$ is d_{12}, the population covariance between thge intercepts and slopes; and the constant portion of the marginal variance $d_{11} + \sigma^2$ is the sum of the intercepts plus the residual variance.

Random intercept and slope covariance structure is heteroscedastic where the diagonal element of D increases with time with positive curvature d_{22}. It also assumes that correlation between reapeated measures changes with time. If data is collected over along period of time, the marginal variance explodes which is usually implausible. Over a short period, there is little problem.

3.0.5 Model Assumptions

1. The within errors are indepedent and identically normally distributed with mean zero and variance σ^2 and are indepedent of random effects.

2. The random effects are normally distributed with mean zero and covariance matrix D (not depending on subject) and are indepedent of different groups.

3. The variances of random effects in random intercept and random intercept and slope models are constant and non-negative.

Chapter 4

Methods of Estimation

4.1 Maximum likelihood estimation Method

The maximum likelihood estimator of θ is the value of $\hat{\theta}$ that maximises the likelihood function for all values of θ in parametric space.

Under mixed effect model the normal distribution of y has the joint probability density function

$$f(y) = \frac{1}{(2\pi)^{n/2}V^{1/2}} exp\{-\frac{1}{2}(y - X\beta)'V^{-1}(y - X\beta)\} \tag{4.1}$$

where n is dimension of y.

In deriving this methods of estimation we take $y_{ij} = y_i$.

For random coefficient models, with assumption that $\{y_i\}$ are normally distributed, the log-likelihood of a single subject is

$$\ell_i(\beta, \tau) = -\frac{1}{2}(T_i ln(2\pi) + lndetV_i(\tau) + (y_i - X_i\beta)'V_i^{-1}(\tau)^{-1}(y_i - X_i\beta)) \tag{4.2}$$

The log-likelihood of entire dataset is given by

$$L(\beta, \tau) = \sum_{i=1}^{N} \ell_i(\beta, \tau)$$

The values of β and τ that maximises $L(\beta, \tau)$ are maximum likelihood estimators (MLEs) denoted by β_{MLE} and τ_{MLE}.

The score vector is the vector of derivatives of log-likelihood taken with respect to the parameters. Typically, if the score has roots, then the root is a maximum likelihood estimator. To compute the score vector, first take the derivative with respect to the parameter and find the root. That is

$$\begin{aligned}\frac{\partial}{\partial\beta}L(\beta,\tau) &= \sum_{i=1}^{N}\frac{\partial}{\partial\beta}\ell_i(\beta,\tau)\\ &= -\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial\beta(y-X\beta)'}V_i^{-1}(\tau)^{-1}(y_i-X\beta))\\ &= \sum_{i=1}^{N}V_i^{-1}(\tau)^{-1}(y_i-X\beta) \qquad (4.3)\end{aligned}$$

Setting this score vector equal to zero yields

$$\sum_{i=1}^{N}X_i'V_i(\tau)^{-1}y_i-\sum_{i=1}^{N}X_i'V_iX_i\beta=0$$

This yields

$$(\beta_{GLS},\tau)=(\sum_{i=1}^{N}(X_i'V_i(\tau)^{-1}X_i)^{-1}\sum_{i=1}^{N}X_iV_i(\tau)^{-1}y_i \qquad (4.4)$$

Thus for fixed covariance parameter τ, MLE and GLS are same.

From equation 4.2 and 4.4 substituting expression for GLS in log-likelihood equation yields concentrated or profiled log-likelihood given by,

$$\begin{aligned}L(\beta_{GLS},\tau) &= -\frac{1}{2}\sum_{i=1}^{N}(T_iln(2\pi)+lndetV_i(\tau)+(y_i-X_i\beta_{GLS})'V_{-i}(\tau)^{-1}(y_i-X_i\beta_{GLS})\\ &= -\frac{1}{2}\sum_{i=1}^{N}(T_iln(2\pi)+lndetV_i(\tau)+(errorSS)_i(\tau) \qquad (4.5)\end{aligned}$$

Which is a function of τ.

where

$$(errorSS)_i(\tau)=(y_i-X_i\beta_{GLS})'V^{-i}(\tau)^{-1}(y_i-X_i\beta_{GLS})$$

We maximimises the log-likelihood as a function of τ. In only few cases can we obtain closed form expression for maximising variance components.

For random intercept model the variance components are given as $\tau=(\sigma^2,d_{11})$. Hence

$$V_i=var(y_{ij})=d_{11}J_i+\sigma^2I_i$$

Thus

$$\begin{aligned} lndetV_i &= lndet(d_{11}J_i + \sigma^2 I_i) \\ &= T_i ln\sigma^2 + ln(1 + \frac{T_i d_{11}}{\sigma^2}) \end{aligned} \tag{4.6}$$

From this the concentrated log-likelihood equation becomes

$$\begin{aligned} L(\beta_{GLS}, d_{11}, \sigma^2) &= -\frac{1}{2}\sum_{i=1}^{N}(T_i ln(2\pi) \\ &+ T_i ln\sigma^2 + ln(1 + \frac{T_i d_{11}}{\sigma^2})\frac{1}{\sigma^2}(y_i - X_i\beta_{GLS})'(I_i - \frac{d_{11}}{T_i d_{11} + \sigma^2}J_i)(y_i - X_i\beta_{GLS}) \end{aligned}$$

Removing parts that do not depend on variance components yields

$$L(\beta_{GLS}, d_{11}, \sigma^2) = -\frac{1}{2}\sum_{i=1}^{N}(T_i ln\sigma^2 + ln(1 + \frac{T_{i_{11}}}{\sigma^2})\frac{1}{\sigma^2}(y_i - X_i\beta_{GLS})'(I_i - \frac{d_{11}}{T_i d_{11} + \sigma^2}J_i)(y_i - X_i\beta_{GLS} \tag{4.7}$$

This likelihood can be maximized over (σ^2, d_{11}) using iterative methods.

For random intercept and slope model $V_i(\tau)$ represent all the variance components. That is, variance of intercepts d_{11}, variance of slopes d_{22} and covariance of intercepts and slopes d_{12}. These also are gotten by maximization of likelihood. All these estimates were obtained for the dataset on growth of Acacia Senegal trees considered.

4.1.1 Restricted maximum likelihood estimation (REML)

REML is likelihood based estimation procedure.Harville observed that its not specific to a particular design matrix as are most analysis of variance estimators. REML result in unbiased estimator of variance components for many balanced designs. The idea behind REML estimation is to consider likelihood of linear combinations of responses that do not depend on mean parameters. The responses (y) are assumed to be normally distributed with $E(y) = X\beta$ and variance-covariance matrix $V(\tau)$

4.1.2 REML calculations

Indepedence of residuals and least-squares estimators

Assume that y has a multivariate normal distribution with mean $X\beta$ and variance-covariance matrix V, where X has dimension $N \times p$ with rank p. V depends on variance-covariance parameter τ.

We use the matrix $Q = I - X(X'X)^{-1}X'$. Because Q is idempotent and has rank $N - p$, we canfind an $N \times (N - p)$ matrix A such that

$$AA' = Q$$

and

$$AA' = I_N$$

We also need $G = V_{-1}X(X'V_{-1}X)^{-1}$, an $N \times p$ matrix.Note that the GLS estimator of β is given by $G'y = b_{GLS}$.

With these two matrices, define the transformation matrix $H(A : G)$, an $N * N$ matrix. Consider the transformed variables

$$H'y = \begin{pmatrix} A'y \\ G'Y \end{pmatrix} = \begin{pmatrix} A'y \\ b_{GLS} \end{pmatrix}$$

Basic calculation show that

$$A'y \sim \mathcal{N}(0, A'VA)$$

and

$$G'y = b_{gls} \sim \mathcal{N}(\beta, (X'V_{-1}X)-1)$$

in which $z \sim \mathcal{N}(\mu, V)$ denotes that a vector z has a multivariate normal distribution with mean μ and variance V. Further, we have that $A'y$ and b_GLS are indepedent. This is due to normality and the zero covariance matrix.

$$
\begin{aligned}
Cov(A'y, b_{gls}) &= E(A'yy'G) \\
&= A'VG \\
&= A'X(X'V^{-1}X - 1 \qquad (4.8)
\end{aligned}
$$

We have $A'X = 0$ because $A'X = (A'A)A'X = A'QX$ and $QX = 0$

Derivation of Restricted likelihoods

To develop the REML we first check the rank of transformation matrx H. Thus with H as above we have

$$
\begin{aligned}
det(H^2) &= det(H'H) \\
&= det[[\begin{pmatrix} A' \\ G' \end{pmatrix}][AG]] \\
&= det(A'A)det(G'G - G'A(A'A)^{-1}A'G) \\
&= det(G'G - G'QG) \\
&= det(G'X(X'X)^{-1}X'G) \\
&= det((X'X)^{-1}) \qquad (4.9)
\end{aligned}
$$

Using $G'X = I$. Thus, the transformation H is nonsingular if and only if $X'X$ is nonsingular. No information is lost by considering the transformation $H'y$.

Based on probability density function of $A'y$ the restricted likelihood is developed. We writen $f_{G'y}$ to denote the probablity density function of the random vector $G'y$, evaluated at the (vector) point z with mean parameter β. Because probability density function integrate to one, we have the relation

$$
\begin{aligned}
1 &= \int f_{g'y}(z,\beta)dz \\
&= \int \frac{1}{(2\pi)^{p/2}det(X'V^{-1}X)-1/2} exp(\frac{-1}{2}(z-\beta)')X'V-1X()z-\beta dz \\
&= \int \frac{1}{(2\pi)^{p/2}det(X'V^{-1}X)-1/2} exp(\frac{-1}{2}(z-\beta)')X'V-1X()z-\beta d\beta \\
&= \int f_{g'y}(z,\beta)d\beta \qquad (4.10)
\end{aligned}
$$

for each z with a change of variables.

Because of indepedence of $A'y$ and $G'y = b_{GLS}$, we have

$$f_{H'y} = f_{A'y}f_{G'y}$$

where $f_{H'y}, f_{A'y}$ ands $f_{G'y}$ are the density functions of random vectors of $f_{H'y}, f_{A'y}$ and $f_{G'y}$ respectively.

Let y be a potential realisation of the random vector. Thus, the probability density function of $A'y$ is

$$
\begin{aligned}
f_{A'y}(A'y) &= \int f_{A'y}(A'y)f_{G'y}(G'y,\beta)d\beta \\
&= \int f_{H'y}(H'y,\beta)d\beta \\
&= \int det(H)^{-1}f_y(y,\beta)d\beta \qquad (4.11)
\end{aligned}
$$

using change of variables.

Let b_{GLS} be the realization of β_{GLS} using y. Then, from a standard equality from analysis of variance,

$$(y-X\beta)'V^{-1}(y-X\beta) = (y-Xb_{GLS})'V^{-1}(y^-Xb_{GLS}) + (b_{GLS}-\beta)'X'V^{-1}X(b_{GLS}-\beta)$$

With this equality, the probability density function f_y can be expressed as

$$\begin{aligned} f_y(y,\beta) &= \frac{1}{(2\pi)^{N/2}det(V^{1/2}}exp(-\frac{-1}{2}(y-X\beta)')V^{-1}(y-X\beta) \\ &= \frac{1}{(2\pi)^{N/2}det(V^{1/2}}exp(-\frac{-1}{2}(y-Xb_{GLS})')V^{-1}(y-Xb_{GLS})) \\ &\quad exp(-\frac{1}{2}(b_{GLS}-\beta)'X'V^{-1}X(b_{GLS}-\beta)) \\ &= \frac{(2\pi)^{p/2}det(X'V^{-1}X)-1/2}{(2\pi)^{N/2}det(V^{1/2}}exp(-\frac{-1}{2}(y-Xb_{GLS})'V^{-1}(y-X_bGLS))f_{G'y}(b_{GLS},\beta) \end{aligned}$$

Thus

$$\begin{aligned} f_{A'y}(A'y) &= \frac{(2\pi)^{p/2}det(X'V^{-1}X)-1/2}{(2\pi)^{N/2}det(V^{1/2}}det(H)^{-1}exp(-\frac{-1}{2}(y-Xb_{GLS})'V^{-1}(y-Xb_{GLS})) \\ &\quad \int f_{G'y}(b_{GLS},\beta)d\beta \\ &= (2\pi)^{-(N-p)/2}det(V)^{-1/2}det(X'X)^{1/2}det(X'V^{-1}X)^{-1/2} \\ &\quad exp(-\frac{-1}{2}(y-Xb_{GLS})')V^{-1}(y-Xb_{GLS}) \end{aligned} \tag{4.12}$$

Taking logariths and dropping terms that do not involves variance components τ yields,

$$\begin{aligned} L_{REML}(b_{GLS}(\tau),\tau) &= -\frac{1}{2}[lndet(V(\tau))+lndet(X'V(\tau)^{-1}X)+(y-Xb_{GLS}(\tau))')V^{-1}(y-Xb_{GLS}(\tau)] \\ &= -\frac{1}{2}[lndet(V(\tau))+lndet(X'V(\tau)^{-1}X)+(errorSS)(\tau) \end{aligned} \tag{4.13}$$

which is restricted log-likelihood function.

where

$$(errorSS)(\tau) = (y-X\beta_{GLS}(\tau))')V^{-1}(y-X\beta_{GLS}(\tau)$$

4.1.3 Application of REML

Defining a projection matrix as

$$Q = I - X(X'X^{-1}X)'$$

and considering linear combination Qy, then we find that Qy has mean zero and variance-covariance matrix that do not depend on parameter β. Derivatin of REML function is shown in

appendix 2. Rank of Q is $N-p$ thus we lose some information by considering this transformation of data hence the use of descriptor restricted maximum likelihood.There is some information about τ in vector β_{GLS} that we are not using.The REML as derived in above is given by

$$L_{REML}(\beta_{GLS}(\tau),\tau) = -\frac{1}{2}[lndet(V(\tau))+lndet(X'V(\tau)^{-1}X)+(y-X\beta_{GLS}(\tau))')V^{-1}(\tau)(y-X\beta_{GLS}(\tau)] \tag{4.14}$$

where

$$(ErrorSS)(\tau) = (y-X\beta_{GLS}(\tau))')V^{-1}(\tau)(y-X\beta_{GLS}(\tau)$$

Thus in random coefficient model the likelihood for a single subject is given by

$$\begin{aligned} L_{REML}(\beta_{GLS}(\tau),\tau) &= -\frac{1}{2}[lndet(V_i(\tau)) + lndet(X_i'V_i(\tau)^{-1}X_i) + \\ &\quad (y-X_i\beta_{GLS}(\tau))')V_i^{-1}(\tau)(y_i-X_i\beta_{GLS}(\tau)] \end{aligned} \tag{4.15}$$

The log-likelihood of entire dataset is given by

$$L(\beta,\tau) = \sum_{i=1}^{N} \ell_i(\beta,\tau)$$

The values of β and τ that maximises $L(\beta,\tau)$ are restricted maximum likelihood estimators (RMLEs) denoted by β_{REML} and τ_{REML}.

The score vector is the vector of derivatives of log-likelihood taken with respect to the parameters. Typically, if the score has roots, then the root is a maximum likelihood estimator. To compute the score vector, first take the derivative with respect to the parameter and find the root. That is

$$\begin{aligned} \frac{\partial}{\partial\beta}L(\beta,\tau) &= \sum_{i=1}^{N}\frac{\partial}{\partial\beta}\ell_i(\beta,\tau) \\ &= -\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial\beta(y-X\beta)'}V_i^{-1}(\tau)^{-1}(y_i-X\beta)) \\ &= \sum_{i=1}^{N}V_i^{-1}(\tau)^{-1}(y_i-X\beta) \end{aligned} \tag{4.16}$$

Setting this score vector equal to zero yields

$$\sum_{i=1}^{N}X_i'V_i(\tau)^{-1}y_i - \sum_{i=1}^{N}X_i'V_iX_i\beta = 0$$

This yields

$$(\beta_{GLS}, \tau) = (\sum_{i=1}^{N} (X_i' V_i(\tau)^{-1} X_i)^{-1} \sum_{i=1}^{N} X_i V_i(\tau)^{-1} y_i \tag{4.17}$$

Thus for fixed covariance parameter τ, REML and GLS are same.

From equation 4.15 and 4.17 substituting expression for GLS in log-likelihood equation yields concentrated or profiled log-likelihood given by.

$$\begin{aligned} L(\beta_{GLS}, \tau) &= -\frac{1}{2}\sum_{i=1}^{N}(T_i ln(2\pi) + lndetV_i(\tau) + lndet(X_i' V_i(\tau)^{-1} X_i) \\ &+ (y_i - X_i\beta_{GLS})' V^{-1}(\tau)^{-1}(y_i - X_i\beta_{GLS})) \\ &= -\frac{1}{2}\sum_{i=1}^{N}(T_i ln(2\pi) + lndetV_i(\tau) + lndet(X_i' V_i(\tau)^{-1} X_i) + (errorSS)_i(\tau)) \end{aligned}$$

Which is a function of τ.

We maximimises the log-likelihood as a function of τ. In only few cases can we obtain closed form expression for maximising variance components.

Thus from equation above the REML for random intercept model becomes

$$\begin{aligned} L_{REML}(\beta_{GLS}(\tau), \tau) &= -\frac{1}{2}[lndet(T_i ln\sigma^2 + ln(1 + \frac{T_i d_{11}}{\sigma^2}) + lndet(X_i'(I_i - \frac{d_{11}}{T_i d_{11}}) + \sigma^2 J_i) X_i) \\ &+ (y - X_i\beta_{GLS}(\tau))')(I_i - \frac{d_{11}}{T_i d_{11}} + \sigma^2 J_i)(y_i - X_i\beta_{GLS}(\tau)] \end{aligned}$$

where $\sigma^2 = \sigma^2 I_{ni}$ This likelihood can be maximized over (σ^2, d_{11}) using iterative methods to obtain (σ^2) and d_{11}.

$V_i(\tau)$ also represent all the random effects for random intercept and slope model. That is, variance of initial heights (intercepts) d_{11}, variance of rate of growth (slopes), d_{22} and covariance of intercepts and slopes d_{12}. These also are gotten by maximization of likelihood using iterative methods. These estimates were for the dataset on growth of Acacia Senegal trees considered were obtained.

4.1.4 Matrix inversion

To simplify our equation and provide a better intuition for expression we site a formula for inverting V_i. V_i has dimension $T_i \times T_i$.

$$\begin{aligned} V_i^{-1} &= (\Sigma_i + Z_i D Z_i')^{-1} \\ &= \Sigma_i^{-1} - \Sigma_i^{-1} Z_i (D^{-1} + Z_i' \Sigma_i^{-1} Z_i)^{-1} Z_i' \Sigma_i^{-1} \end{aligned} \tag{4.18}$$

The last equation is easier to compute than the left hand side when temporal covariance Σ_i has an easily computable inverse and dimension q is small relative to T_i. Moreover because the matrix

$$D^{-1} + Z_i' \Sigma_i^{-1} Z_i)$$

is only a $q \times q$ matrix its easier to invert V_i, a $T_i \times T_i$ matrix. In case of no serial correlation $\Sigma_i = \sigma^2 I_{ni}$ hence equation 4.35 reduces to

$$\begin{aligned} V_i^{-1} &= (\sigma^2 I_{ni} + Z_i D Z_i')^{-1} \\ &= \frac{1}{\sigma^2}(I_{ni} - Z_i(\sigma^2 D^{-1} + Z_i' \Sigma_i^{-1} Z_i)^{-1} Z_i') \end{aligned} \tag{4.19}$$

For random intercept model we thus have

$$\begin{aligned} V_i^{-1} &= (\sigma^2 I_{ni} + D Z_i Z_i')^{-1} \\ &= \frac{1}{\sigma^2}(I_{ni} - \frac{D}{T_i D + \sigma^2} J_i) \\ &= \frac{1}{\sigma^2}(I_{ni} - \frac{\zeta_i J_i}{T_i}) \end{aligned} \tag{4.20}$$

where

$\zeta_i = \frac{T_i D}{T_i D + \sigma^2}$ For random intercept and slope model, V_i take the simple form

$$V_i = D + \sigma^2 (Z_i' Z_i)^{-1}$$

hence the weihgt take the form

$$W_{iGLS} = (D + \sigma^2 (Z_i' Z_i)^{-1})^{-1}$$

This shows that subjects with large values of $Z_i' Z_i)$ have greater effect on β_{GLS} than subjects with small values.

4.1.5 Derivation of β_{GLS}

Consider a Linear Model

$$Y = X\beta + \varepsilon$$

and

$$Var(\varepsilon) = \sigma^2 V$$

where V is $N \times N$ positive definite (p.d) matrix. Since V is positive definite it follows from diagonability of positive definite matrices that there exist an $N \times N$ matrix K with R(K)=N such that $V = KK'$. Let$Z = K^{-1}y, B = K^{-1}X, \eta = K^{-1}\varepsilon$
Since $r(X) = r \leq p$ it follows that $r(B) = r$ and $E(\eta) = 0$.

$$\begin{aligned} Var(\eta) &= K^{-1}\sigma^2 V K^{-1} \\ &= \sigma^2 K^{-1} K K' K^{-1} \\ &= \sigma^2 I_N \end{aligned} \tag{4.21}$$

Minimizing

$$\eta'\eta = (Z - B\beta)'(Z - B\beta)$$

with respect to β we obtain the generarized least square(GLS) solution to normal equations as

$$\begin{aligned} \eta'\eta &= \varepsilon' V^{-1} \varepsilon \\ &= (y - X\beta)' V^{-1} (y - X\beta) \\ &= y'y - 2\beta' X' V^{-1} y + \beta' X' V^{-1} X\beta \end{aligned} \tag{4.22}$$

Differentiating with respect to β and setting equal to zero, That is

$$\frac{\partial \eta'\eta}{\partial \beta} = -2X'V^{-1}y + 2X'V^{-1}X\beta = 0$$

Thus we get

$$\beta = (X'V^{-1}X)^{-1} X V^{-1} y \tag{4.23}$$

If $r = p$ we have full rank model hence $X'V^{-1}X^{-1}$ exists and the solution vector is unique geralized least squares(GLS) of β given by

$$\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1} X V^{-1} y \tag{4.24}$$

Variance-covariance matrix of $\hat{\beta}$ is given by

$$\begin{aligned} &= Var[(X'V^{-1}X)^{-1}X'V^{-1}y] \\ &= \sigma^2(X'V^{-1}X)^{-1}X'V^{-1}VV^{-1}X(X'V^{-1}X)^{-1} \\ &= \sigma^2(X'V^{-1}X)^{-1} \end{aligned} \tag{4.25}$$

longitudinal data. The weighted least square of β is given by

$$\hat{\beta}_W = (X'WX)^{-1}X'Wy \tag{4.26}$$

ght matrix. If $W = V^{-1}$ this yields OLS estimate of β. While setting he most efficient estimator.

ficient model shown above the vector y_i has mean $X_i\beta$ and variance

$$\begin{aligned} V_i(\tau) &= Z_iDZ_i' + \Sigma_i \\ &= V_i \end{aligned} \tag{4.27}$$

n coefficient model,the GLS estimator of $\hat{\beta}_{GLS}$ is given by

$$\hat{\beta}_{GLS} = \sum_{i=1}^{n_i}(X_i'V_i^{-1}X_i)^{-1}\sum_{i=1}^{n_i}X_iV_i^{-1}y_i \tag{4.28}$$

zation Algorithm

components are estimated recursively. This can be done using either the Fisher scoring method. We in this case use Newton-Raphson algorithm ive its one of the most widely used optimisation procedures. It uses first score function (the gradient of log-likelihood function) around a current oduce the next estimate $\theta^{(w+1)}$. Each iteration requires the calculation of s derivative, the hessian matrix of log-likelihood. Newton-Raphson converge

$$L = L(\beta_{GLS}(\tau), \tau)$$

method

$$\tau_{NEW} = \tau_{OLD} - \{(\frac{\partial^2 L}{\partial_\tau \partial_{\tau'}})^{-1}\frac{\partial L}{\partial_\tau}\}\mid_{\tau = \tau_{old}}$$

The matrix

$$-\frac{\partial^2 L}{\partial \tau \partial \tau'}$$

is called sample information. Any iterative algorithm requires initial values for parameters. Because we can express both profiled log-likelihood and profiled log restricted likelihood as a function of only τ parameters we only need to formulate starting values for τ when performing iterative optimization for Linear mixed effect model.

Chapter 5

Analysis

5.1 Estimates for Random Intercept (RI) Model

Hierachial model was assumed in which random effects are used to describe the covariance structure in the data whereas all the remaining variability is assumed to be purely measurement error. The corresponding hierachial model is given by ommiting subject specific slopes thereby assuming all individual profiles have equal slopes.

Table 5.2 shows the ML and REML variance estimates for random intercept model with their respective log-likelihood estimates for the data on growth of Acacia Senegal trees. In both the

Table 5.1: Table of ML and REML estimates of random effects for Random intercept model

Effect	Parameter	ML	REML
Covariance of b_i			
$\text{Var}(b_{0i})$	d_{11}	0.2248	0.2511
Residual Variance $\text{Var}(\varepsilon_{ij})$	σ^2	0.1153	0.1155
REML Loglikelihood		-336.5372	-350.6117
Intraclass correlation		0.6611	0.6851

estimates of variances for intial height (intercept) is more than the measurement eroor variances. Thus its not difficult to detect the heterogeneity in random effects thus b_{0i} may reflect the correct distribution shape. However the REML estimates of d_{11} and σ^2 are slightly higher than those estimated using ML.

Intraclass correlation

Observations made on the same subject share the same random effects $b_0 i$ hence they are correlated.The covariance between observations on the same subject is d_{11} corresponding to correlation of

$$\rho_i = \frac{d_{11}}{d_{11} + \sigma^2}$$

This gives intraclass correlation which can be interpretated as correlation between observations within a subject or proportion of variability of response that is due to heterogeneity namong subjects.

The intraclass correlation obtained from ML estimates is given by

$$\begin{aligned} \rho_{ML} &= \frac{0.2248}{0.2248 + 0.1152} \\ &= 0.6611 \end{aligned} \tag{5.1}$$

Similarly the intraclass correlation obtained from REML estimates is given by

$$\begin{aligned} \rho_{ML} &= \frac{0.2511}{0.2511 + 0.1155} \\ &= 0.6851 \end{aligned} \tag{5.2}$$

The intraclass correlation obtained from REML estimates are slightly higher than those obtained from ML estimates. This shows that 66.11% from ML and 68.51% of total variance in acacia Senegal tree growth was due to between repeated observations in the subject differences. That is, correlation between height of trees for two randomly selected trees from the same subject from ML and REML estimates were 0.6611% and 0.6851% respectively.Intraclass correlation obtained from REML estimates are slightly higher than those estimated from ML estimates.

Covariance and Correlation Stuctures

Random intercept models assumes that all variability in subject specific slopes are ascribed to to treatment differences hence random slopes are ommitted. The implied marginal covariance structure is given by

$$\varepsilon_{ij} = \sigma^2 I_{ni}$$

and

$$\begin{aligned} cov(Y_{ij}, Y_{ij'}) &= (1)d_{11}(1) + \sigma^2 I_{ni} \\ &= d_{11} + \sigma^2 I_{ni} \end{aligned} \tag{5.3}$$

Thus the implied covariance matrix is given by

$$V_i = \begin{pmatrix} d_{11} + \sigma^2 & d_{11} & \dots & d_{11} \\ d_{11} & d_{11} + \sigma^2 & \dots & d_{11} \\ \vdots & & & \\ d_{11} & d_{11} & \dots & d_{11} + \sigma^2 \end{pmatrix}$$

The covariance matrix obtained from ML and REML estimates are

$$\begin{pmatrix} 0.6611 & 0.2247 & \dots & 0.2247 \\ 0.2247 & 0.6611 & \dots & 0.2247 \\ \vdots & & & \\ 0.2247 & 0.2247 & \dots & 0.6611 \end{pmatrix}$$

and

$$\begin{pmatrix} 0.66851 & 0.2247 & \dots & 0.2247 \\ 0.2511 & 0.6851 & \dots & 0.2511 \\ \vdots & & & \\ 0.2511 & 0.2511 & \dots & 0.6851 \end{pmatrix}$$

respectively. Thus random intercept implies an equivariance or equicorrelation matrix. The corresponding covariance matrix with constant variance and constant correlation is compound symetry. Positive estimates for d_{11} of random effects estimated by the two methos indicates that the assumptions of positive correlation between the repeated measurements was valid for the dataset.

5.2 Estimates for Random Intercept and Slope (RIAS) Model

RIAS generarizes RI model to observations falling around subject specific lines with unknown subject specific intercepts (b_{0i}) and subject specific slopes(b_{1i}). There is different sntercept and

slope pair.

Table 5.2 shows the ML and REML estimates of random effects for RIAS model and their respective loglikelihood estimates.

Table 5.2: Table of REML estimates of random effects for Random intercept and slope model

Effect	Parameter	ML	REML
Covariance of b_i			
Var b_{0i}	d_{11}	2.0245	2.1030
Var b_{1i}	d_{22}	6.7417e-06	6.9337e-06
Cov (b_{0i}, b_{1i})	d_{12}	-3.4875e+02	-3.5881e+02
Residual Variance			
Var(ε_{ij})	σ^2	0.06516	0.06515
REML Loglikelihood		-176.4778	-189.428

From the table the resulting 2×2 estimates of random effects covariance matrix D are given by,

$$D_{ML} = \begin{pmatrix} 2.0245 & -0.003487 \\ -0.003487 & 0.000006742 \end{pmatrix}$$

for ML estimates and

$$D_{REML} = \begin{pmatrix} 2.1030 & -0.003588 \\ -0.003588 & 0.000006944 \end{pmatrix}$$

for REML estimates.

D has three unique parameters. That is d_{11} which is variance for random intercept, d_{22} which is variance of random slope and d_{12} which is covariance of slopes and intercepts. The estimates of measurement error are almost equal for the two methods. From the D matrix the ML estimates for variance of random intercepts(intial heights) and variance of random slopes (rate of growth) are slightly less than those of REML but are comparable. The ML estimates for correlation between intercepts and slopes (-0.944) is also slightly less than that of REML estimates

(−0.939). This indicates a strong negative correlation of the intial height(intercepts) and rate of growth(slopes).

Since the variance of the slopes is small compared to variance of intercepts, its natural to explore whether a random intercept model is adequate.

Estimated Variance Function

From the D matrix the variance function is predicted by

$$Var(Y_{ij}/\sigma^2, D) = \begin{pmatrix} 1 & t_{ij} \end{pmatrix} \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix} \begin{pmatrix} 1 \\ t_{ij} \end{pmatrix} + \sigma^2$$

Thus

$$Var(Y_{ij}/\sigma^2, D) = d_{22}{t_{ij}}^2 + 2d_{12}t_{ij} + d_{11} + \sigma^2$$

This gives the variance functions estimated by ML and REML to be

$$Var(Y_{ij}(t_{ij})) = 0.000006742{t_{ij}}^2 - 0.006974t_{ij} + 2.08966$$

and

$$Var(Y_{ij}(t_{ij})) = 0.000006944{t_{ij}}^2 - 0.007176t_{ij} + 2.16815$$

respectively.

From both, we have positive estimates of random slopes. This suggests presence of positive curvature in variance function. Thus the obtained variance functions are compatible with proposed hierachial models hence random effect models that naturally arose from two stage approach described in chapter 3 implies an appropriate marginal model.

Covariance and Correlation Matrices

Random effects results from variation between and within subjects hence impact on variance and covariance of the observations. The estimate for marginal covariance matrices obtained by combining the estimate for D with estimate for σI_{ni} is assumed to be of the form

$$V_i = Z_i D Z_i' + \sigma I_{ni}$$

D is as described above.

The following are portions of correlation matrices estimated by ML and REML respectively.

$$\begin{pmatrix} 1.0000000 & 0.9773966 & 0.9738081 & 0.9693334 & 0.9649574 & 0.9633820 & 0.9519445 \\ 0.9773966 & 1.0000000 & 0.9699753 & 0.9654874 & 0.9611015 & 0.9595231 & 0.9480703 \\ 0.9738081 & 0.9699753 & 1.0000000 & 0.9618671 & 0.9574751 & 0.9558949 & 0.9444349 \\ 0.9693334 & 0.9654874 & 0.9618671 & 1.0000000 & 0.9529664 & 0.9513851 & 0.9399228 \\ 0.9649574 & 0.9611015 & 0.9574751 & 0.9529664 & 1.0000000 & 0.9469860 & 0.9355268 \\ 0.9633820 & 0.9595231 & 0.9558949 & 0.9513851 & 0.9469860 & 1.0000000 & 0.9339472 \\ 0.9519445 & 0.9480703 & 0.9444349 & 0.9399228 & 0.9355268 & 0.9339472 & 1.0000000 \end{pmatrix}$$

for ML and

$$\begin{pmatrix} 1.0000000 & 0.9809179 & 0.9782164 & 0.9749571 & 0.9718695 & 0.9707792 & 0.9631519 \\ 0.9809179 & 1.0000000 & 0.9752230 & 0.9719510 & 0.9688531 & 0.9677596 & 0.9601136 \\ 0.9782164 & 0.9752230 & 1.0000000 & 0.9692184 & 0.9661133 & 0.9650176 & 0.9573591 \\ 0.9749571 & 0.9719510 & 0.9692184 & 1.0000000 & 0.9628164 & 0.9617185 & 0.9540490 \\ 0.9718695 & 0.9688531 & 0.9661133 & 0.9628164 & 1.0000000 & 0.9586002 & 0.9509231 \\ 0.9707792 & 0.9677596 & 0.9650176 & 0.9617185 & 0.9586002 & 1.0000000 & 0.9498212 \\ 0.9509231 & 0.9498212 & 1.0000000 & 0.9305383 & 0.9122976 & 0.8790089 & 0.8101871 \end{pmatrix}$$

for REML.

Both methods gives non constant variances that decreases with time. However the variances and covariances estimated using REML are slightely higher than those estimated using ML.

The correlation are seen to decrease with increasing lag. As time increases the correlation between the observations at two adjacenttime points decreases.

RIAS covariance matrix is heteroscedastic where diagonal element of Vi increases quadratically with time with positive curvature. This covariance structure also assumes that correlation between repeated measurements changes with time. This covariance structure is obtained when additional random effect b_{1i} associated with time is assumed. Thus our estimated variances are consistent with this.

Likelihood Ratio TEst (LRT)

A likelihood ratio test can be derived for comparing models with different covariance structure and to test whether variance function is significantly different from constant. More specifically the hypothesis of interest is

$$Ho : d_{12} = d_{22} = 0$$

From table 5.1 and 5.2 the LRT obtained from comparing maximised ML and REML loglikelihood values can be obtained. The observed values for the test statistic obtained by ML method equals

$$\begin{aligned} -2ln\lambda_N &= -2(-336.5372 + 176.4478) \\ &= 320.1788 \end{aligned} \tag{5.4}$$

Similarly the observed values for the test statistics obtained from REML method is

$$\begin{aligned} -2ln\lambda_N &= -2(-350.6117 + 1189.4729) \\ &= 322.2776 \end{aligned} \tag{5.5}$$

The test statistics obtained by ML are slightely less than those obtained by REML. We compare this with chi-square distribution on 2 degree of freedom. The critical values at 5% level of significant is 3.84. Thus for the two methods there is a strong evidence of between subject differences.The two test statistics are significant when compared to a chi-square distribution with 2 degrees of freedom ($p < 0.001$). Hence the random intercept and slope model is more consistent with the Acacia Senegal dataset.

Chapter 6

Conclusions

Maximum likelihood estimation and Restricted maximum likelihood estimation both have same merits of being based on likelihood principle which leads to useful properties such as consistency , asymptotic normality and efficieny. One expect results from ML and REML estimation to differ more as number of parameters of fixed effects increases.

The random effects for both random intercept and random intercept and slope models were comparable for the two estimating methods.

Both intial heights(intercepts) and rate of growth(slope) were strongly negative correlated. The covariance between them were negative hence the individual tree with initial higher values have greater negative slopes. The covariance structure implied by random intercept and random intercept and slope models is more consistent with typical longitudinal data.

Positive estimate of d_{11} random effects estimated by the two methods indicates that assumptions of positive correlation between repeated measurements was valid for the dataset.// The estimated variance functions were compatible with proposed hierachial model assumed hence random effects that naturally arose from two stage approach described by the dataset used implies an appropriate model.

Error variance and covariance structure play an important role in the shape of distribution of random effects b_i. Since error variability σ^2 was small compared to random effect variability b_i reflects the correct distributional shape of the random effects.

Under hierachial interpretation of model, ti may be of scientific interest to test for the need of some of the random effects. Here we test

$$H_o : d_{12} = d_{22} = 0$$

All observed values for $-2ln\lambda_N$ (log-likelihood ratio test)are larger than 3.84 hence there was strong evidence of between subject differences.The $p < 0.0001$ hence covariance structure should not be simplified deleting random effects from the model.

Thus in all ML and REml compared quite well for the dataset on growth of Acacia Senegal trees. Further work acn be done by using the same dataset to compare other estimating methods such as Generalized Estimating Equation(GEE) and Minimum Norm Quadratic unbiased Equation(MINQUE). Other hierachial models could be considered.

6.1 Appendix

6.1.1 R code used in analysis

```
prodata<-read.csv("datapro.csv")
attach(prodata)
prodata
prodataA<-prodata[!is.na(prodataHEIGHT),]
prodataA
library(nlme)
fitproMLS<-lme(HEIGHT~TIME+TREATMENT+SITES, random=~1+TIME|
SUBJECT,data=prodataA,control=list(msVerbose=TRUE),method="ML")
fitproMLI<-lme(HEIGHT~TIME+TREATMENT+SITES, random=~1|SUBJECT,
data=prodataA,control=list(msVerbose=TRUE),method="ML")
Zi1<-c(rep(1,21))
MLI<-1.422844055 ^2
MLS<-0.002596475^2
MLR<-0.255255121^2
DmatML<-matrix(c(2.024485,-0.003487494,-0.003487494,0.000006741682),nrow=2)
covML<-0.002596475*1.422844055*(-0.944)
VarcovMLS<-Zi3%*%DmatREML%*%ZT3
VarSigmaML<-VarcovMLS+( 0.06515077*diag(21))
```

```
corrML<-cor(VarSigmaML)
############################################################
fitproREML<-lme(HEIGHT~TIME+TREATMENT+SITES, random=~1+TIME|
SUBJECT, data=prodataA,control=list(msVerbose=TRUE))
fitproREML<-lme(HEIGHT~TIME+TREATMENT+SITES, random=~1|
SUBJECT, data=prodataA,control=list(msVerbose=TRUE))
RED11<-1.450178869^2
RERES<-0.255246492^2
RECOVA<-1.450178869*0.002635098*(-0.939)
RED22<-0.002635098^2
DmatREML<-matrix(c(2.103019,-0.00358826,-0.00358826,0.00000694374),nrow=2)
ZT1<-t(Zi3)
ZT3<-t(Zi3)
VarcovREML<-Zi3%*%DmatREML%*%ZT3
VarSigmaREML<-VarcovREML+( 0.06515077*diag(21))
corrREML<-cor(VarSigmaREML)
anova(fitproREML2,fitproREML1)
anova(fitproML1,fitproMLS))
```

6.1.2 References

1. Bates, D.M. and Pienheiro, J.C. (2000). Mixed effects models in S and S-Plus. Newyork: Springer

2. Boos,D.D., and Gumpertz, M.L (2001).Comparisons of gee, minque,ml and reml Estimating Equations for Normally distributed Data. The American Statistician, 55,125-130.

3. Davdian, M. and Zhang,D. (2001) Linear Mixed Model with Flexible Distribution of Random Effects for Longitudinal Data. Biometrics, 57: pp795-802.

4. Hsiao, C. and Pesaron, M.H. (2004) Random Coefficient Panel Data Models. University of Cambridge.

5. Laird, N.M. and Ware, J.H. (1982). Random Models for Longitudinal Data. Biometrics,

38: pp 963-974. Marchenko, Y. (2006) Estimating Variance Components in Stata. The Stata Journal, 6: 1-21.

6. Masata, O. and Rasch, D. (2006) Methods of Variance Components Estimation. Biometric, 51: 227-235.

7. Meradith, M.P. and S.V. Stehman. 1991. Repeated measures experiments in forestry; focus on analysis of response curves. CJFR, 21: pp 957-965.

8. Nokoe K. Sagary, 2003. Biomathematical modelling of stand growth: Importance of mathematical modelling of biological and biomedical processes. pp 121-124.

9. Ware, J.H. (1985). Linear Models for analysis of Longitudinal Studies. Journal of The America Statistical Association, 39: 95-101.