EVALUATION OF INFERENCE METHODS IN GLMMS FOR ECOLOGICAL MODELING

by

Edward John Reddick

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
December 2010

DALHOUSIE UNIVERSITY

DEPARTMENT OF MATHEMATICS AND STATISTICS

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "EVALUATION OF INFERENCE METHODS IN GLMMS FOR ECOLOGICAL MODELING" by Edward John Reddick in partial fulfillment of the requirements for the degree of Master of Science.

Dated: December 13, 2010

Supervisors:

Dr. Chris Field

Dr. Joanna Flemming

Reader:

Dr. David Hamilton

DALHOUSIE UNIVERSITY

DATE: December 13, 2010

AUTHOR: Edward John Reddick

TITLE: EVALUATION OF INFERENCE METHODS IN GLMMS FOR ECOLOGICAL MODELING

DEPARTMENT OR SCHOOL: Department of Mathematics and Statistics

DEGREE: M.Sc. CONVOCATION: May YEAR: 2011

Signature of Author

To the memory of my mother, Mary Reddick.

Table of Contents

List of Tables

List of Figures

Abstract

Inference in generalized linear mixed models (GLMM) remains a topic of debate. Baayen, Davidson, and Bates (2008) outlines criticism against conventional ways of performing inference for GLMMs. There are various alternatives proposed but little consistency is found on which is the most reasonable. Our focus is on assessing temporal trends for mainly ecological count data. That is, we hope to provide a pragmatic approach to Poisson GLMMs for ecological researchers within the statistical programming environment `R`. To achieve this, we start by providing a description of the selected estimation and inferential procedures. We then complete a large scale simulation to evaluate each of the estimation methods. We implement a power analysis to assess each of the selected inferential procedures. We then go on to apply these procedures to data sampled by The National Parks of Canada. Finally, we conclude by giving a summary of our findings and outlying work for the future. Keywords: GLMM, Poisson, Temporal Trend, P-value, Parametric Bootstrapping

Acknowledgements

I am extremely indebted to all the contributers of my thesis, I relied heavily upon them for guidance, assistance, and financial support. I have been very fortunate to have had the range of support that I have enjoyed. I'm very thankful to my supervisors Dr. Chris Field and Dr. Joanna Flemming whose intuition, expertise, and weekly meetings kept me motivated, interested, and "out of trouble". Furthermore, I am deeply grateful to Dan Kehler who purposed this topic of research, in addition to helping every step of the way.

I acknowledge the financial support of the Department of Mathematics and Statistics of Dalhousie University, School of Graduate Studies and Drs. Chris Field and Joanna Flemming in the form of Graduate Assistantships, and Dan Kehler Monitoring Ecologist of The National Parks of Canada for providing me with a student research assistant position for the duration of my program.

I would also like to thank Deborah Austin, my boss at Parks Canada for the experience of working with challenging statistical problems and for her flexibility and understanding of my student obligations. Thank you to Dr. David Hamilton who was kind enough to read my thesis. I would also like to extend gratitude to Wade Blanchard for advice with respect to this thesis topic.

I'd like to thank my friends and family, among them Joey Mingrone whose computer advice was an invaluable asset. Finally, I cannot express how grateful I am to my parents. I would never have been able to make it this far without their support.

Chapter 1

Introduction

1.1 Motivation

Generalized Linear Mixed Models have become a very powerful and widely used statistical tool. They are often applied to investigate a variety of properties associated with the grouping of data, combined with the situation where the response variable is not continuous. Although Linear models (LM)s are the workhorse of statistics, there are times when they should not be applied. For instance, if one wants to account for variation due to some grouping structure it can be assumed there is a random effect associated with group. This can be achieved by employing a linear mixed model (LMM). Within this context, the random effect or effects are assumed to follow a Gaussian distribution. LMs are also not appropriate if the response is non-normal but can be modeled by a member of the exponential family. The inclusion of a link function connecting the predictor to the response leads to the implementation of a generalized linear model (GLM). Suppose we are dealing with both scenarios; there exists a random effect, and the data is non-normal. A mix between LMMs and GLMs seems to be the most appropriate framework, and hence the need for GLMMs. However, inferential procedures surrounding GLMMs remain a highly debated statistical topic.

GLMMs are often applied in ecological settings. Especially when sampling is done over time on multiple locations and the response is non-normal. Given this context, many researchers are interested in obtaining an estimate of the temporal trend in order to say something about the metric of interest. There are a number of methods that achieve estimation. However, assessing whether or not the trend is statistically significant remains a challenge.

Presently, a variety of statistical methods such as GLIMMIX within SAS give p-values associated with a fixed trend effect. However, such methods use a degree of freedom approximation and are subject to controversy [1]. This debate was further highlighted by the fact that p-values are not included in the summary of `glmer`, a popular R method for GLMMs, which is written and maintained by Doug Bates. Other procedures were also proposed to assess the statistical significance of fixed effects, such as parametric bootstrapping, and Markov chain Monte Carlo algorithms.

The National Parks of Canada have developed ecological monitoring programs to assess the status and trend in ecological integrity using a series of indicators. The programs include numerous measures of ecological integrity in different ecosystems that capture key elements of biological diversity or ecosystem processes. Monitoring measures must be scientifically credible, and provide meaningful results. An important component of these requirements is that the study design have sufficient replication to ensure satisfactory levels of statistical power to detect trend, and to determine status. Most of the measured variables are not normally distributed, and many of the study designs involve repeated sampling of observations over time at a fixed series of sites. This design suggests that a generalized linear mixed model (GLMM) approach would be suitable for analyzing trend and estimating status, where both the sampling site intercepts and slopes are random effects. As noted previously, there is disagreement in the literature over how to assess the statistical significance of these models, and also how to extract appropriate uncertainty estimates for the fixed and random

effects. A related area of uncertainty is how to design a monitoring study that has appropriate statistical power, using a GLMM as the statistical testing framework.

1.2 Literature Review

An understanding of a wide variety of topics was required for the problem at hand. In order to work with GLMMs an understanding of LMMs and GLMs was required. Select sections from several text books were important to achieve this. "Mixed-Effects Models in S and S-plus" by JC Pinheiro and DM Bates [2] was a good reference for LMMs. For a pragmatic approach to GLMs we used "Extending the Linear Model" with R by J. Faraway [3] and for theoretical questions we referred to "Generalized Linear Models" by McCullagh and Nelder [4]. For the background of GLMMs we used "Generalized, Linear, and Mixed Models" by C McCulloch and S Searle [5], "Linear and generalized linear mixed models and their applications" by J Jiang [6], and a draft of the soon to be released "*lme4*: Mixed-effects Modeling with R" by Doug Bates [7].

We came across a promising method that fell within the Bayesian paradigm and was implemented in R. To fill in details we consulted "Bayesian computation with R" by J Albert [8] and "Introducing Monte Carlo Methods with R" by C Robert and C Casella [9]. Also, the writer of this procedure J Hadfield has published a very detailed article outlying the process titled "MCMC Methods for Multi-response Generalized Linear Mixed Models: The *MCMCglmm* R package" [10].

A background in SAS's long standing implementation for GLMMs can be found in an article titled "The Glimmix Procedure" [11]. While R's older procedure's methodology can be found summarized in an article titled "PQL Estimation Biases in Generalized Linear Mixed Models" by J Woncheol [12]. Both of these use degree of freedom approximation for the inference of fixed effects, and have come under scrutiny by Doug Bates [13].

There is also a huge body of information regarding GLMMs found on R-sig-mixed-models. This is an Internet forum interested in issues surrounding mixed effect modeling in R and is primarily interested with the *lme4* package, but often addresses general mixed modeling concerns. There are constant postings from lead statistical researchers and developers. Much of Ben Bolker's work surrounding GLMMs use within ecology was also highly valuable. We also had correspondence with Doug Bates, who gave us recommendations on available methods.

Our review of the literature confirmed that there is still confusion on how to assess the significance of fixed effects for GLMMs. Ecological Researchers constantly face problems surrounding best practices for how certain factors effect the power of various inference methods to assess trend. Such factors include the number of groups sampled from and the number of time points or years of sampling.

1.3 Ecological Data

A real application where GLMMs apply is a Salmon data set from Fundy National Park sampled by Parks Canada and depicted in Figure 1.2. This is count data and there seems to be variability associated with site. Conventionally, this data set would be modeled as a Poisson GLMM with a random intercept and slope effect.

We have analyzed salamander data also collected by Parks Canada (as seen in Figure 1.1) to portray a case where there is reduced random slope variability associated with plot. We include this analysis to demonstrate how the methods contend with such a situation.

The goal of this thesis is to provide ecological researchers with guidelines for GLMMs in deciding such things as number of sites, number of years to sample, and number of repeat observations per site in conjunction with determining which of the readily available procedures should be used for analysis. Chapter 2 will give

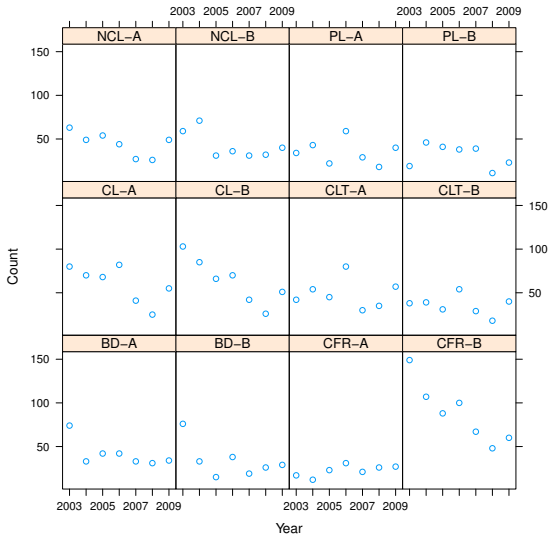

Figure 1.1: Salamander Sampling. The year sample was taken is denoted along the x-axis, Salamander count is on the y-axis. Each window within the figure represents samples taken from a unique sampling site.

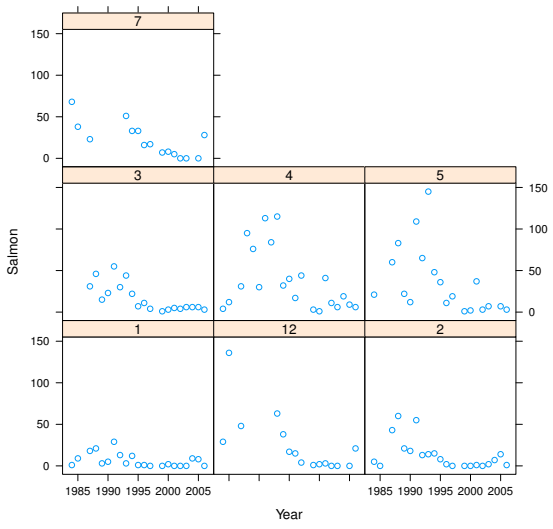

Figure 1.2: Salmon Sampling. The year sample was taken is denoted along the x-axis, Salmon count per plot is along the y-axis. Each window within the figure represents multiple samples taken from unique plots in Fundy National Park.

a description of the methods and inference approaches. Chapter 3 will discuss a simulation study designed to address how well the methods perform with respect to estimation. This chapter will also demonstrate how different factors affect the validity of inference using power analysis. Finally, Chapter 4 will present conclusions of our findings and future work.

Chapter 2

Methodology

2.1 Model

According to [5] in order to specify the model for a GLMM one usually starts with the assumption that the response vectors $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are independent given the random effects, where $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})$ are n_i samples that are observed along with the $p \times n_i$ covariate matrix $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{in_i})$ for the fixed effects $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$. Also, $\mathbf{Z}_i = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, \ldots, \mathbf{z}_{in_i})$ is the $q \times n_i$ covariate matrix for the random effects $\mathbf{u} = (u_1, \ldots, u_q)$. One then defines the conditional distribution of the response vector $\mathbf{y}$ given the random effects $\mathbf{u}$ and further assumes that the conditional density follows a distribution from the exponential family:

$$\begin{aligned} \mathbf{y}_i|\mathbf{u} &\sim \text{indep} f_{Y_i|\mathbf{u}}(\mathbf{y}_i|\mathbf{u}) \\ f_{Y_i|\mathbf{u}}(\mathbf{y}_i|\mathbf{u}) &= \exp\{[\mathbf{y}_i\boldsymbol{\gamma}_i - b(\boldsymbol{\gamma}_i)]/\tau^2 - c(\mathbf{y}_i, \tau)\} \end{aligned} \tag{2.1}$$

We should note that $E(\mathbf{y}_i|\mathbf{u}) = \partial b(\boldsymbol{\gamma}_i)/\partial\boldsymbol{\gamma}_i = \boldsymbol{\mu}_i$ and $V(\mathbf{y}_i|\mathbf{u}) = \frac{\tau^2}{a_{ij}} V(E(\mathbf{y}_i|\mathbf{u}))$ where τ is a dispersion parameter and a_{ij} is a prior weight, usually 1. Following a similar

structure set up for GLM and denoting the link function by $g(\cdot)$:

$$\begin{aligned} E(\mathbf{y}_i|\mathbf{u}) &= \boldsymbol{\mu}_i \\ g(\boldsymbol{\mu}_i) &= \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{z}_i'\mathbf{u}. \end{aligned} \tag{2.2}$$

Within the GLMM context f_U is assumed to be a Gaussian probability density function with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, that is $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$.

2.2 Likelihood Approach

It is very common within classical Statistics to employ maximum likelihood to estimate unknown model parameters. Within a random effects context, in order to write down the likelihood of (2.1) one must integrate over the q-dimensional distribution of the random effects $\mathbf{u}$.

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \int_{\mathbf{u}} \prod_i f_{Y_i|\mathbf{u}}(\mathbf{y}_i|\mathbf{u}) f_U(\mathbf{u}) d\mathbf{u} \tag{2.3}$$

A special case of the the above form is a Poisson GLMM, where there exists an intercept random effect for different groups. Let y_{ij} denote the jth count taken on the ith group. The form of the model is:

$$\begin{aligned} y_{ij}|\mathbf{u} &\sim \text{indep Pois}(\mu_{ij});\ i = 1, 2, ..., N; j = 1, 2, ..., n_i; \\ \log(\mu_{ij}) &= \mathbf{x}_{ij}'\boldsymbol{\beta} + u_i \\ u_i &\sim \mathcal{N}(0,\, \sigma_u^2). \end{aligned} \tag{2.4}$$

We assume that the random effects follow a normal distribution with mean 0 and variance σ_u^2. We employ a log link function and x_{ij}' is simply the covariate associated with the fixed effect for the j^{th} count for the i^{th} group. There is a common random

effect for each of the i groups and through this construct observations from the same group are modeled as correlated. The log likelihood is given by the following equation:

$$\begin{aligned} l &= \log\left(\prod_{i=1}^{N}\int_{-\infty}^{\infty}\prod_{j=1}^{n_i}\frac{\mu_{ij}^{y_{ij}}e^{-\mu_{ij}}}{y_{ij}!}\frac{1}{\sqrt{2\pi\sigma_u^2}}e^{\frac{-1}{2\sigma_u^2}u_i^2}\,du_i\right) \\ &= \mathbf{y}'\mathbf{X}\boldsymbol{\beta} - \sum_{i,j}\log\, y_{ij}! + \sum_{i}\log\int_{-\infty}^{\infty}\exp\left(y_{i.}u_i - \sum_{j}e^{x'_{ij}\boldsymbol{\beta}+u_i}\right)\frac{1}{\sqrt{2\pi\sigma_u^2}}e^{\frac{-1}{2\sigma_u^2}u_i^2}\,du_i. \end{aligned} \tag{2.5}$$

According to [5] Equation (2.5) cannot be simplified further or evaluated in closed form. Many of the methods currently in development for generalized linear mixed models deal with approximating this likelihood. It also becomes increasing more challenging to numerically maximize the likelihood as the dimension of the random effects increases.

Another way to parameterize the model is within the Bayesian context using Markov chain Monte Carlo methods. It is argued that using a diffuse prior for the fixed effects will result in a posterior close to the likelihood.

2.3 Estimation Approaches

In this section we describe a variety of approaches used to estimate unknown GLMM model parameters. We will focus on describing the methods that apply Laplacian approximation, penalized quasi likelihood, and MCMC when preforming our power simulations. This will be discussed in the next chapter.

2.3.1 Maximum Likelihood

2.3.1.1 Laplacian Approximation

In R one of the most popular functions for fitting GLMMs is called `glmer` and is found within the package *lme4* which is written and maintained by Douglas Bates.

Methodology can be found at [7]. By default it uses laplacian approximation within its method to approximate the likelihood. The user is also given the choice to use maximum likelihood (ML) or restricted maximum likelihood (REML). It is well known that ML underestimates the variance of the random effects components [12]. REML was developed as a way to reduce this bias. ML estimates standard deviations of the random effects assuming that the fixed-effect estimates are correct, while REML estimation averages over some of the variability in the fixed-effect parameters. The following derivations are done with respect to ML. `glmer` does calculations with respect to a linear transformation of the random effects rather than the random effect themselves. The reason will become clear after we have shown the motivation behind the transformation $\mathbf{U} = \mathbf{\Lambda}(\boldsymbol{\theta})\mathcal{U}$. The transformation leads to $\mathbf{\Lambda}(\boldsymbol{\theta})$ being the relative covariance factor where $\text{Var}(\mathbf{U}) = \mathbf{\Sigma}(\boldsymbol{\theta}) = \mathbf{\Lambda}(\boldsymbol{\theta})\mathbf{\Lambda}'(\boldsymbol{\theta})$ and $\mathcal{U}(\boldsymbol{\theta}) = \mathbf{Z}\mathbf{\Lambda}(\boldsymbol{\theta})$, where $\mathbf{Z}$ is the design matrix associated with the random effect. For GLMMs, $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}(\theta))$, $\mathcal{U} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q)$. This slight difference allows `glmer` to estimate parameters even when $\mathbf{\Sigma}(\boldsymbol{\theta})$ is singular. The estimates also remain stable as $\mathbf{\Sigma}(\boldsymbol{\theta})$ approaches singularity. In this thesis attention is restricted to the case where the distribution of the observations given the random effects is Poisson. The distribution can be denoted:

$$(\mathbf{Y}|\mathcal{U} = \mathbf{u}) \sim \ pois(\boldsymbol{\mu}_{\mathbf{Y}|\mathcal{U}}), \tag{2.6}$$

where $\boldsymbol{\mu}_{\mathbf{Y}|\mathcal{U}}(\mathbf{u}) = exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{\Lambda}(\theta)\mathbf{u})$, the exponential function being the inverse of the log link function, which is the canonical link function. As previously stated in Equation (2.1), the model is defined in terms of $(\mathbf{Y}|\mathcal{U} = \mathbf{u})$. However, we observe $\mathbf{y}$ not $\mathbf{u}$ so we are forced to focus our attention on derivations surrounding the conditional distribution $(\mathcal{U}|\mathbf{Y} = \mathbf{y})$. The `glmer` fitting procedure is a way of approximating

the likelihood of the parameters given the data,

$$L(\boldsymbol{\theta}, \boldsymbol{\beta}|\mathbf{Y} = \mathbf{y}) = \int_{\mathbf{u}} f_{\mathbf{Y}|\mathcal{U}}(\mathbf{y}|\mathbf{u}) f_{\mathcal{U}}(\mathbf{u}) d\mathbf{u}$$
$$= \int_{\mathbf{u}} \prod_i f_{\mathbf{Y}_i|\mathcal{U}}(\mathbf{y}_i|\mathbf{u}) f_{\mathcal{U}}(\mathbf{u}) d\mathbf{u}. \tag{2.7}$$

The product of distribution functions inside the integral is the unscaled conditional distribution of $\mathcal{U}|\mathbf{Y}$. As stated previously, the likelihood is intractable so in order to approximate it `glmer` starts with the conditional mode:

$$\tilde{u}(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\beta}) = \arg\max_u f_{\mathbf{Y}|\mathcal{U}}(\mathbf{y}|\mathbf{u}) f_{\mathcal{U}}(\mathbf{u}). \tag{2.8}$$

This is done by employing a quadratic approximation to the logarithm of the unscaled conditional distribution of $(\mathcal{U}|\mathbf{Y} = \mathbf{y})$. This can be thought of as a penalized, reweighted residual sum of squares,

$$\tilde{u}(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\beta}) = \arg\min_{\mathbf{u}} \left\| \begin{bmatrix} \mathbf{W}^{1/2}(\boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathcal{U}}(\mathbf{u})) \\ -\mathbf{u} \end{bmatrix} \right\|^2 \tag{2.9}$$

where $\mathbf{W}(\boldsymbol{\mu})$ is the diagonal weight matrix and is defined as the inverse of the variance of $\mathbf{Y}_i|\mathcal{U} = \mathbf{u}$. We are able to rewrite equation (2.8) as equation (2.9) because both $f_{\mathbf{Y}|\mathcal{U}}(\mathbf{y}|\mathbf{u})$ and $f_{\mathcal{U}}(\mathbf{u})$ are exponential functions. We know that exponential functions are monotone increasing and the values inside the functions are negative. Therefore the mode corresponds to when the negative exponentiated value is at its minimum. Parameter estimates for the model are found using a penalized iteratively reweighted least squares algorithm (PIRLS). The parameters are effectively determined for a fixed $\mathbf{W}$ matrix. $\mathbf{W}$ is then updated using the current estimated parameters and the

algorithm repeats. Given a fixed $\mathbf{W}$ we solve

$$\min_{\mathbf{u}} \left\| \begin{bmatrix} \mathbf{W}^{1/2}(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathcal{U}}(\mathbf{u})) \\ -\mathbf{u} \end{bmatrix} \right\|^2 . \tag{2.10}$$

This is essentially solving a nonlinear least squares problem, solving

$$\mathbf{P}(\mathbf{UMWMU}' + \mathbf{I}_q)\mathbf{P}'\boldsymbol{\delta}_{\mathbf{u}} = \mathbf{UMW}(\mathbf{y} - \boldsymbol{\mu}) - \mathbf{u} \tag{2.11}$$

where $\boldsymbol{\delta}_{\mathbf{u}}$ is the update and, $\mathbf{M} = \frac{d\boldsymbol{\mu}}{d\boldsymbol{\eta}} = \mathrm{Var}\,(\mathbf{Y}|\mathcal{U}) = \mathbf{W}^{-1}$ is the diagonal Jacobian matrix. $\mathbf{P}$ is a $q \times q$ fill-reducing permutation. Re-ordering the columns of $\mathbf{Z}\boldsymbol{\Lambda}(\boldsymbol{\theta})$ does not effect the theory of GLMMs however it does effect the time and storage needed to evaluate the matrix $\mathbf{L}$, which will be discussed shortly. For the Poisson case with no overdispersion $\mathrm{Var}\,(\mathbf{Y}|\mathcal{U}) = \mathrm{E}\,(\mathbf{Y}|\mathcal{U}) = \boldsymbol{\mu}_{\mathbf{Y}|\mathcal{U}}(\mathbf{u})$. Convergence occurs when $\boldsymbol{\delta}_{\mathbf{u}}$ is sufficiently close to 0. Once convergence occurs the Cholesky factor, $\mathbf{L}$ used to evaluate the update is

$$\mathbf{LL}' = \mathbf{P}(\mathbf{UMU}')\mathbf{P}'. \tag{2.12}$$

$\mathbf{L}$ is calculated in order to make it easier to solve for the conditional mode. It is a lower triangular matrix which makes otherwise intractable calculations manageable. The ability to compute $\mathbf{L}$ is imperative to making `glmer` a feasible and robust method. The likelihood is proportional to the density of the distribution $\mathcal{N}(\tilde{\mathbf{u}}, \mathbf{LL}')$. Employing the laplace approximation method, is equivalent to fitting a Gaussian at the mode of the function of interest. Written on the deviance scale

$$d(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y}) = d_g(\mathbf{y}, \boldsymbol{\mu}(\tilde{\mathbf{u}})) + \|\tilde{\mathbf{u}}\|^2 + \log\,(|\mathbf{L}|^2), \tag{2.13}$$

where $d_g(\mathbf{y}, \boldsymbol{\mu}(\tilde{\mathbf{u}}))$ is the same deviance found in a GLM i.e. (the same model without random effects for $\mathbf{y}$ and $\boldsymbol{\mu}$). Finally, the above calculations allow us to have estimates of the unknown GLMM model parameters.

2.3.1.2 Adaptive Gauss-Hermite Quadrature Approximation

As an alternative to laplacian approximation one can also employ Adaptive Gauss-Hermite Quadrature (GHQ) approximation. While laplacian and GHQ are both methods to approximate the likelihood. GHQ approximates the likelihood by choosing optimal subsets at which to evaluate the integrand. Adaptive GHQ uses information gained from an initial parameterization to increase precision. While adaptive GHQ is said to be more precise than laplacian approximation there is a sacrifice with respect to speed. Adaptive GHQ gets exponentially slower as one increases the dimension of random effects; so much so that the procedure is not feasible for 2 or 3 random effects [14].

2.3.2 Penalized Quasi Likelihood

Another method used for estimation of the parameters of a GLMM is penalized quasi likelihood. A method for implementation can be found with the R package *MASS*. Following [15] and [12] and in keeping with previously stated notation one can estimate the parameters in the following manner. First, denote the integrated quasi-likelihood $L(\boldsymbol{\beta}, \boldsymbol{\theta})$ by:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^q |\boldsymbol{\Sigma}(\boldsymbol{\theta})|}} \int \exp\Bigg[-\frac{1}{2\tau} \sum_{i=1}^{N} \sum_{j=1}^{n_i} d_{ij}(y_{ij}, \mu_{ij}) - \frac{1}{2} \mathbf{u}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{u} \Bigg] d\mathbf{u} \tag{2.14}$$

where we define $d_{ij}(y, \mu) = -2a_{ij} \int_y^\mu \frac{y_{ij} - c}{v(c)} dc$ to be the deviance measure of fit. Also, $v(\cdot)$ is the variance function coming from an exponential family. We also simplify the

variance of the random effects by assuming they are orthogonal. As noted before, one cannot integrate the true likelihood in closed form, the same is true for the quasi likelihood. One can then use the Laplace approximation for the equation:

$$\mathbf{PQL}(\boldsymbol{\beta}, \mathbf{u}) = -\frac{1}{2\tau}\sum_{i=1}^{N}\sum_{j=1}^{n_i} d_{ij}(y_{ij}, \mu_{ij}) - \frac{1}{2}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{u} \tag{2.15}$$

which we can denote as the penalized quasi likelihood. If [15] Equation (2.15) is replaced with its quadratic expansion at $\hat{\mathbf{u}} = \arg\min\mathbf{PQL}(\boldsymbol{\beta}, \mathbf{u})$ for fixed $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ and defining $\hat{\boldsymbol{\beta}} = \arg\min\mathbf{PQL}(\boldsymbol{\beta}, \hat{\mathbf{u}})$. The standard restricted maximum likelihood estimating equation for $\boldsymbol{\theta}$ using the updating term y_{ij}^* is determined by letting $y_{ij}^* = g(y_{ij}) = g(\hat{\mu_{ij}}) + (y_{ij} - \hat{\mu_{ij}})g'(\hat{\mu_{ij}})$ be the two term Taylor expansion of $g(y_{ij})$. Finally, we can write the model in the following form:

$$Y^* = X_i\hat{\boldsymbol{\beta}} + Z_i\hat{\mathbf{u}} + \epsilon_i \tag{2.16}$$

where $\epsilon_i \sim \mathrm{N}(\mathbf{0}, W_i^{-1})$ and W_i is a diagonal matrix with elements $w_{ij} = [V(\mu_{ij})(g'(\mu_{ij}))^2]^{-1}$. As given in [15], the implementation of the above can be done by looping through the following steps.

1) Given starting values for $\boldsymbol{\theta}$ and $\mathbf{u}$ one can estimate $\boldsymbol{\beta}$ by solving the following normal equation

$$\sum_{i=1}^{N} X_i^T V_i^{-1} X_i\boldsymbol{\beta} = \sum_{i=1}^{N} X_i^T V_i^{-1} Y_i^*, \tag{2.17}$$

where $V_i = W_i^{-1} + Z_i D Z_i^T$.

2) We estimate $\mathbf{u}$ with the following equation

$$\mathbf{u} = \sum_{i=1}^{N} D Z_i^T V_i^{-1}(Y_i^* - X_i\hat{\boldsymbol{\beta}}). \tag{2.18}$$

3) We denote $\boldsymbol{\theta}_s$ as the Restricted Maximum Likelihood (REML) estimator of $\boldsymbol{\theta}$ and define the estimator as

$$\boldsymbol{\theta}_s = \frac{\Sigma_{n\in Q_s}\hat{\mathbf{u}}_n^2}{\Sigma_{n\in Q_s}(1-t_{nn})}, \text{ for } s = 1,\ldots,c, \tag{2.19}$$

where

$$Q_s = n : \sum_{i=1}^{s-1} q_i < n \leq \sum_{i=1}^{s} q_i, S = W - WX(X^TWX)^{-1}X^TW, \tag{2.20}$$

$$X^T = (X_1^T,\cdots,X_N^T),\ Z^T = (Z_1^T,\cdots,Z_N^T),\ W = \text{diag}(W_1^T,\cdots,W_N^T), \tag{2.21}$$

and t_{nn} is the nth diagonal term of $T = (I + Z^TSZD)^{-1}$.

4) We then update Y_i^* with each iteration through these steps. When convergence is reached the final estimators are known as the PQL estimators. The covariance matrix of $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ are given by the equations

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \left[\sum_{i=1}^{N} X_i^TV_i^{-1}X_i\right]^{-1},\ \text{Cov}(\hat{\boldsymbol{\theta}}) = H^{-1}. \tag{2.22}$$

Denoting the H components as

$$h_{st} = \frac{1}{2}\sum_{i\in Q_s}\sum_{i\in Q_t}\left(Z_{(i)}^TPZ_{(j)}\right)^2, \tag{2.23}$$

denoting $Z_{(i)}$ as the ith row vector of Z, $V^{-1} = \text{diag}(V_1^{-1},\cdots,V_N^{-1})$ and

$$P = V^{-1} - V^{-1}X\text{Cov}(\hat{\boldsymbol{\beta}})X^TV^{-1}. \tag{2.24}$$

Despite `glmer` and `glmmPQL` being similar in that they both use laplacian approximation to reach parameter estimates, they are actually quite different. `glmmPQL` approximates a quasi likelihood not a true likelihood. That is, only the expectation

and variance of the distribution of the response variable given the random effects needs to be specified, and not its form. Another key difference is that `glmer` applies a transformation to the (co)variance matrix $\Sigma(\boldsymbol{\theta})$ which is one of the reasons why `glmer` is a more robust estimation procedure. We'll now move onto a method that is philosophically much different than the previous two.

2.3.3 Markov Chain Monte Carlo

The R package *MCMCglmm* relies heavily on a Bayesian paradigm and Markov chain Monte Carlo (MCMC) methods. This thesis is primarily written from a classical statistics standpoint, details such as choosing the prior and burn-in period were chosen pragmatically to mimic classical statistics as closely as possible. Before explaining the package's implementation we give a brief explanation of MCMC [8].

Markov Chain Monte Carlo methods are often used to summarize the posterior distribution. The MCMC sampling scheme sets up an irreducible, aperiodic Markov chain for which the stationary distribution equals the posterior distribution of interest. Two of the most popular sampling schemes are known as the Metropolis-Hastings algorithm and the Gibbs sampler. We are effectively interested in the density of the posterior denoted $p(\boldsymbol{\theta}|\mathbf{y})$, which is the probability of the parameters given the data. This is conceptually different than a classical or frequentist perspective, where one cannot treat the parameter of interest as a random variable. The Metropolis-Hastings algorithm starts with an interest in sampling from this posterior and supposes one has an initial value of the parameter $\theta^{(0)}$ and specifies a rule for simulating the t^{th} value in the sequence $\theta^{(t)}$ given the $(t-1)^{th}$ value in the sequence. The rule is made up of a proposal density which simulates a candidate value $\theta^{(*)}$ and the probability of acceptance of $\theta^{(*)}$ as the next value in the sequence which is known as an acceptance probability and will be temporarily be denoted P. A brief description of the algorithm is as follows:

1) Simulate $\theta^{(*)}$ from the proposal density $p(\theta^{(*)}|\theta^{(t-1)}, \mathbf{y})$

2) Calculate the ratio

$$R = \frac{p(\theta^{(*)}|\mathbf{y})p(\theta^{(t-1)}|\theta^{(*)}, \mathbf{y})}{p(\theta^{(t-1)}|\mathbf{y})p(\theta^{(*)}|\theta^{(t-1)}, \mathbf{y})} \tag{2.25}$$

3) Compute the acceptance probability as $P = \min(R, 1)$

4) Let $\theta^{(t)} = \theta^{(*)}$ with probability P otherwise $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$.

The sequence of simulated draws $\theta^{(1)}, \cdots, \theta^{(t)}$ will converge to a random variable that is distributed according to the posterior $p(\theta|\mathbf{y})$ [16]. Note that, the Metropolis-Hastings algorithm becomes more difficult to set up as the dimensions of $\boldsymbol{\theta}$ increase because one must specify a proposal density that is multi-dimensional.

The Gibbs sampler is also very popular especially when the parameter of interest has high dimension. Consider the case where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$ and we cannot sample directly from the joint density $p(\boldsymbol{\theta}|\text{data})$. However we do have access to the conditional distributions:

$$p(\theta_1|\theta_2, \cdots, \theta_p, \text{data}) \tag{2.26}$$

$$p(\theta_2|\theta_1, \theta_3, \cdots, \theta_p, \text{data}) \tag{2.27}$$

$$\vdots \tag{2.28}$$

$$p(\theta_p|\theta_1, \cdots, \theta_{p-1}, \text{data}) \tag{2.29}$$

By simulating the individual parameters from a set of p conditional distributions and by the property set forth by [17] we now have a way in which to recover a joint distribution from its conditionals. Consider the 2 variable case, where we do not have access to the joint distribution of two random variables (X, Y) or to the marginal

distributions. The joint probability using the product rule can be characterized by

$$f(x, y) = f_{X|Y}(x|y)f(y). \tag{2.30}$$

We follow the algorithm outlined by [18] and [9] for the 2 stage Gibbs sampler. Given a starting value $X_0 = x_0$, for $t = 1, 2, \cdots$ generate

$$1)\ Y_t \sim f_{Y|X}(.|x_{t-1}); \tag{2.31}$$

$$2)\ X_t \sim f_{X|Y}(.|y_t). \tag{2.32}$$

Simulating one value from each of these distributions is known as one cycle of Gibbs sampling. This algorithm has effectively created a Markov chain for the marginal distribution of Y. Now that the conditional distributions have been specified fully and we can sample from the marginal distribution it is straightforward to use the product rule to obtain a sample from the joint distribution $f(x, y)$. There exists a rich literature describing MCMC methods in much more depth; one reference is [19]. The rest of this subsection will follow [10] closely in order to describe the implementation of the R package `MCMCglmm`. `MCMCglmm` is an alternative approach for fitting a GLMM model. The resulting posteriors are the distributions of the model parameters. Philosophically this is distinctly different than working within a classical framework where the parameters of interest are not represented as random variables. `MCMCglmm` begins with the probability of the i^{th} data point and can be denoted as:

$$f(y_i|l_i) \tag{2.33}$$

where f is the probability distribution function of the response variables y and l_i are the so called latent variables. l_i is similar to the inversed link function predictor found in a traditional GLMM. However, it allows for a multiplicative dispersion parameter

by including e_i. Equation (2.33) could be written as:

$$f(y_i|\lambda = g^{-1}(l_i)) \tag{2.34}$$

where g is the so called link function. In a Bayesian context we can write the latent variables in vector form as:

$$\mathbf{l} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{2.35}$$

where $\mathbf{X}$ and $\mathbf{Z}$ are design matrices for the fixed and random effects respectively and $\mathbf{e}$ is a vector of residuals. We consider the situation when y follows a Poisson distribution, in this case $\mathbf{e}$ is designed to account for overdispersion. The effects are assumed to come from a multivariate normal distribution:

$$\begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim \left(\mathcal{N} \begin{bmatrix} \boldsymbol{\beta}_0 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{B} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix} \right) \tag{2.36}$$

where $\boldsymbol{\beta}_0$ is a prior vector for the means $\boldsymbol{\beta}$ with prior (co)variance matrix $\mathbf{B}$. $\mathbf{G}$ and $\mathbf{R}$ are the expected (co)variance for $\mathbf{u}$ and $\mathbf{e}$ respectively. That is, one must make an assumption on the form of the (co)variance matrices. Note that the fixed effects, random effects, and residuals are assumed to be independent. The variance structure for the random effects within `MCMCglmm` can be quite general. Also, the parametrization of $\mathbf{G}$ and $\mathbf{R}$ is interchangeable ie, they are parameterized in the same manner. $\mathbf{G}$'s form is as follows:

$$\mathbf{G} = (\mathbf{V_1} \otimes \mathbf{A_1}) \oplus (\mathbf{V_2} \otimes \mathbf{A_2}) \oplus \cdots \tag{2.37}$$

where $\mathbf{A}$ are design-like matrices that imply the structure of the individual random effects. $\oplus$ is the direct sum operation defined as:

$$\mathbf{A} \oplus \mathbf{B} = \begin{bmatrix} \mathbf{A} & 0 \\ 0 & \mathbf{B} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & b_{11} & \cdots & b_{1q} \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & b_{p1} & \cdots & b_{pq} \end{bmatrix} \tag{2.38}$$

and $\otimes$ denotes the Kronecker product defined as:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}. \tag{2.39}$$

In order to obtain estimates of the parameters one can use MCMC. Starting with a prior each iteration updates the variance structure of $\mathbf{G}$ and $\mathbf{R}$ followed by updating the location vector $\boldsymbol{\theta} = [\boldsymbol{\beta}', u']'$. It is then able to update the latent variable $\mathbf{l}$.

`MCMCglmm` updates the latent distribution $\mathbf{l}$ in the following way. The conditional probability distribution of $\mathbf{l}$ is denoted:

$$P(l_i|\mathbf{y}, \boldsymbol{\theta}, \mathbf{R}, \mathbf{G}) \propto f(y_i|l_i) f_N(e_i|r_i R_{/i}^{-1} e_{/i}, r_i - r_i R_{/i}^{-1} r_i') \tag{2.40}$$

where f_N is a multivariate distribution. Also, vectors and matrices that have $/i$ as a subscript denote that rows and/or columns associated with i^{th} observation have been removed. The second term on the right of Equation (2.40) is the probability distribution of the linear predictor residual, conditional on the residuals associated with data points other than i. Therefore, this conditional distribution only involves

other residuals which are expected to show residual covariation that is set in the design of $\mathbf{R}$. Due to the above, `MCMCglmm` updates the latent distribution in blocks, where the block is defined as groups of residuals which are expected to be correlated. The j blocks that have non-zero residual covariance are defined as:

$$P(\mathbf{l}_j|\mathbf{y}, \boldsymbol{\theta}, \mathbf{R}, \mathbf{G}) \propto \prod_{i \in j} p_i(y_i|l_i) f_N(e_j|\mathbf{0}, \mathbf{R}_j) \tag{2.41}$$

which simplifies to:

$$P(\mathbf{l}_j|\mathbf{y}, \boldsymbol{\theta}, \mathbf{R}, \mathbf{G}) \propto p_i(y_i|l_j) f_N(e_j|\mathbf{0}, \mathbf{R}_j). \tag{2.42}$$

$\mathbf{l}$ is updated using either a Metropolis-Hastings update or the slice sampling method purposed by [20]. During the burn-in phase `MCMCglmm` uses adaptive methods to find an optimal multivariate normal proposal distribution on each iteration of l_j which has a covariance matrix mM, where m is chosen so the proportion of successful jumps is optimal with a rate of .44 when $\mathbf{l}_j$ is scalar and .23 when $\mathbf{l}_j$ has a high dimension [21]. M is the average posterior (co)variance of $\mathbf{l}_j$.

`MCMCglmm` updates the location vector $\boldsymbol{\theta} = [\boldsymbol{\beta}', u']'$ by following a method purposed by [22] that employs Gibbs samples from $\boldsymbol{\theta}$ as a complete block by solving the sparse linear system:

$$\tilde{\boldsymbol{\theta}} = \mathbf{C}^{-1}\mathbf{W}'\mathbf{R}^{-1}(\mathbf{l} - \mathbf{W}\boldsymbol{\theta}_\star - \mathbf{e}_\star) \tag{2.43}$$

defining $\mathbf{C}$ as the mixed model coefficient matrix:

$$\mathbf{C} = \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \begin{bmatrix} \mathbf{B}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{bmatrix} \tag{2.44}$$

and $\mathbf{W} = [\mathbf{X}, \mathbf{Z}]$. $\boldsymbol{\theta}_\star$ and $\mathbf{e}_\star$ are random draws from the distributions:

$$\boldsymbol{\theta}_\star \sim \; N\Bigg(\begin{bmatrix} \boldsymbol{\beta}_0 \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} \end{bmatrix} \Bigg) \tag{2.45}$$

$$\mathbf{e}_\star \sim \; N(\mathbf{W}\boldsymbol{\theta}_\star, \mathbf{R}). \tag{2.46}$$

Obtaining the inverse of $\mathbf{G}$ is generally not hard because of the form of Equation (2.37) which results in:

$$\mathbf{G}^{-1} = (\mathbf{V_1^{-1}} \otimes \mathbf{A_1}^{-1}) \oplus (\mathbf{V_2}^{-1} \otimes \mathbf{A_2}^{-1}) \oplus \cdots \tag{2.47}$$

One would also have to get the inverse of $\mathbf{V}_i$ which is not hard because each V_i is generally of low dimension. $\mathbf{A}_i$ is quite often the identity matrix and therefore does not need to be inverted. $\boldsymbol{\theta}_\star$ and $\mathbf{e}_\star$ are inserted into Equation (2.43) and the inverse of $\mathbf{C}$ is solved using Cholesky factorization [23]. Finally $\tilde{\boldsymbol{\theta}} + \boldsymbol{\theta}_\star$ gives a realization from the probability distribution:

$$P(\boldsymbol{\theta}|\mathbf{l}, \mathbf{W}, \mathbf{R}, \mathbf{G}). \tag{2.48}$$

`MCMCglmm` must also update the variance structures denoted $\mathbf{G}$ and $\mathbf{R}$ using Gibbs Sampling when dealing with conjugate priors. The sum of squares matrix or the first parameter of the inverse Wishart distribution corresponding with variance components has the structure:

$$\mathbf{S} = \mathbf{U}'\mathbf{A}^{-1}\mathbf{U} \tag{2.49}$$

where $\mathbf{U}$ is the matrix of random effects, with each column corresponding to the appropriate row/column of $\mathbf{V}$ which in turn corresponds to the appropriate row/column

of the design-like matrix $\mathbf{A}$. $\mathbf{V}$ can be sampled from an inverse Wishart distribution:

$$\mathbf{V} \sim \ IW((\mathbf{S}_p + \mathbf{S})^{-1}, n_p + n), \tag{2.50}$$

where n is the number of rows in $\mathbf{U}$, and $\mathbf{S}_p$ and n_p are the prior sum of squares and prior degrees of freedom respectively.

The previous explanation is quite general, so to illustrate the procedure we consider our previous example with a Poisson response, a fixed intercept and slope effect, and intercept and slope random effects for each of the three groups. Employing `MCMCglmm` to fit this GLMM structured model is implemented in the following way. We'll start with the design matrices for the fixed and random effects

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 10 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 10 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 10 \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 10 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 & 10 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 & 0 & 10 \end{bmatrix}. \tag{2.51}$$

Next, starting values for the initial variance structures of $\mathbf{G}$ and $\mathbf{R}$ are chosen by assuming

$$\mathbf{G} = \mathbf{V}_{1,1} \otimes \mathbf{A}_{1,1} \oplus \mathbf{V}_{1,2} \otimes \mathbf{A}_{1,2} = \sigma_{\beta_0}\mathbf{I}_3 \oplus \sigma_{\beta_1}\mathbf{I}_3 \tag{2.52}$$

$$\mathbf{R} = \mathbf{V}_2 \otimes \mathbf{A}_2 = \sigma_e \mathbf{I}_{30}, \tag{2.53}$$

where the initial value for $\mathbf{V}_{1,1}$,$\mathbf{V}_{1,2}$, and $\mathbf{V}_2$ follow an inverse Wishart distribution with specified scale matrix, and degree of belief parameter found in the prior specification for variance components. And where S_p and n_p are the prior sum of squares and prior degrees of freedom, respectively. With these temporary or working variance structures one can update the location vector $\boldsymbol{\theta} = [\boldsymbol{\beta}', u']'$. Inputting the working values into Equations (2.44), (2.45), and (2.46), and finally (2.43) and following the steps outlined above effectively updates the location vector. A common prior for the fixed effects $\boldsymbol{\beta}$ is a multivariate normal distribution with a mean of $\boldsymbol{\beta}_0 = \mathbf{0}$ and (co)variance matrix

$$\mathbf{B} = \begin{bmatrix} 10^8 & 0 \\ 0 & 10^8 \end{bmatrix}. \tag{2.54}$$

We are now able to update the latent variable $\mathbf{l}$ and thus have estimated values for the residual vector $\mathbf{e}$. The starting value for $\mathbf{e}$ is $\mathbf{0}$. Using the updated values from the first iteration one can now update the variance structure of $\mathbf{G}$ and $\mathbf{R}$ using Equations (2.49) and (2.50). Repeating these steps and saving the resulting parameter estimates after the burn-in phase gives the desired posterior sample. In summary the `MCMCglmm` procedure updates the linear predictor using the Metropolis-Hasting algorithm or slice sampling then Gibbs samples $\boldsymbol{\theta}$, and finishes by Gibbs sampling (co)variance components sequentially [10].

2.4 Inference Approaches

Once each of the models has been parameterized, inference methods are employed in order to draw conclusions regarding the importance of included variables. As stated

previously, this thesis is concerned with the significance of a temporal trend. Therefore, the proceeding inference methods will focus on the significance of the fixed effect associated with year or some other time related predictor, and not with inferential procedures associated with random effects. Also, different inferential approaches are appropriate for different estimation approaches, Table 2.1 demonstrates these combinations with corresponding *R* packages.

We focused on evaluating *R* procedures for GLMMs. `glmmPQL` is a quasi likelihood

Table 2.1: Estimation and Inference

Inference Methods		Estimation Methods		
		PQL	Laplacian	MCMC
Asymptotic	Wald Test	`glmmPQL`	`glmer`	
	DF Approximation			
Resampling	Parametric Bootstrap (LRT)		`glmer`	
	Credible Interval			`MCMCglmm`

procedure which uses a Wald test to carry out inference on β_1. `glmer` can approximate the likelihood using Laplacian Approximation or Gaussian Quadrature. We focused on Laplacian approximation. Inference within `glmer` can be carried out using a Wald test. Currently, there are no complex degree of freedom approximation techniques available in *R* as there are in *SAS*. We used parametric bootstrapping primarily as an inferential procedure, bootstrapping the likelihood ratio between nested models. But, we can also use bootstrapping as an estimation approach once we have estimated model parameters using other procedures; in which case, we bootstrap under the full model. In Section 3.2 we evaluate the version of bootstrapping that estimates model parameters. In Section 3.3 we evaluate the version of parametric bootstrapping that is an inferential procedure. `MCMCglmm` is a Bayesian procedure that obtains posterior distributions of the model parameters. Inference for `MCMCglmm` is carried out using credible intervals.

2.4.1 Asymptotic Inference

Many statistical estimators have asymptotic properties that make inference straightforward. However, GLMM is not one of these areas and it is a heavily debated topic as to how to do inference for GLMMs. Often, to test the significance of a single fixed factor one will compare two models, one containing the factor of interest, and the other without. The comparison is normally done using a likelihood ratio test (LRT). That is, -2 multiplied by the log likelihood ratio and compared to a χ^2 distribution with the degree of freedom (df) bring the difference of the full model df minus the nested df. However, when testing a fixed effect within a GLMM context employing a (LRT) is unsuitable, for small to moderate sample sizes the ratio is not well approximated by a χ^2 distribution. This may depend on the number of groups, total number of samples, and total number of parameters [2]. Therefore, we will not focus on using this method of inference. Despite the uncertainty, methods of asymptotic inference are available [15]. Running a power simulation will give us a better insight into how useful these methods of inference can be under a variety of scenarios.

2.4.1.1 Wald test

The Wald test is a parametric statistical test which approximates the LRT. It uses both the maximum likelihood estimate for the parameter of interest $\hat{\theta}$ and the standard error $\text{se}(\hat{\theta})$. Where the standard error is calculated using the Laplace approximation to the deviance in place of the standard error of the residuals. In order to investigate a possible difference between $\hat{\theta}$ and θ_0, the parameter of interest under the null hypothesis. Often $\theta_0 = 0$ represents no relationship between the predictor variable and the response. The test statistic can be denoted by Z_{Wald} where

$$Z_{\text{Wald}} = \frac{\hat{\theta} - \theta_0}{\text{se}(\hat{\theta})}, \tag{2.55}$$

follows (approximately) a standard normal distribution. Inference within `glmer` employs the Wald test above.

Inference for `glmmPQL` is done a little differently. The test statistic can be denoted by T_{Wald} where

$$T_{\text{Wald}} = \frac{\hat{\theta} - \theta_0}{\text{se}(\hat{\theta})}, \tag{2.56}$$

and for our purposes will be compared to a T distribution with the degrees of freedom equaling $N - q - (p - 1)$. N is the number of observations, q is the number of random effect groups, and p is the amount of fixed effects parameters. The choice to having two different reference distributions was motivated by the package outputs. Inference for PQL is a somewhat naive approach, as a Wald test requires the maximum likelihood estimates and PQL does not provide the true ML estimates.

The difference between the above approaches will be small while sample size is not too small because all the T distribution will be very close to standard normal [24].

2.4.1.2 Degree of Freedom Approximation

Some statistical software packages use degree of freedom approximation when doing inference for GLMMs. This refers to denominator degrees of freedom within an F statistic. It is known that the degrees of freedom must lie somewhere between 1 and $N - 1$ and there have been a wide variety of methods proposed to approximate this value. The simplest approximation is the minimum number of degrees of freedom contributed by random effects that affect the term being tested, but range in complexity to methods that adjust the standard error such as the Kenward and Roger, and Satterthwaite approximations [25] and [26]. The approximation has come under scrutiny by notable statistical researchers such as Doug Bates, who has argued against the assumption "the reference distribution for these F statistics should be an

F distribution with a known numerator degrees of freedom but a variable denominator degrees of freedom" Bates goes onto argue "we can answer the question of how to calculate a p-value by coming up with a formula to assign different denominator degrees of freedom for each test. The denominator doesn't change. Why should the degrees of freedom for the denominator change?" [13]. We have not included the complex degree of freedom approximation methods within our power analysis. We have restricted our attention to methods that can be applied in R.

2.4.2 Resampling Approaches

2.4.2.1 Parametric Bootstrapping

The concept of the bootstrap was first introduced by [27], to estimate the distribution of test statistics. This of course is very useful when the asymptotic properties of a test statistic are unknown. Suppose θ follows an unknown distribution, it is a member of a parametric family, and we have an estimate of θ denoted $\hat{\theta}$. We would like the so-called sampling distribution $f_\theta(\cdot|\theta)$ in order to do inference. However, since this isn't feasible we apply a bootstrapping method by generating observations from $f_y(\cdot|\hat{\theta})$ followed by computing $\hat{\theta}^*$. We then go on to approximate $f_{\hat{\theta}}(\cdot|\theta)$ by $f_{\hat{\theta}^*}(\cdot|\theta)$. This allows us to have a sample of different summaries such as the mean and standard error. For the purposes of this thesis we will implement parametric bootstrap in the following way:

1) Fit data to the full and null models, that is, a model that does and does not include β_1 denoted $\hat{\beta}_{\text{full}}$ and $\hat{\beta}_{\text{null}}$, respectively. Then, compute the test statistic which for our purposes will be the deviance, denoted $\hat{D}$.

2) Generate data set using $\hat{\beta}_{\text{null}}$, fit the full and null model to the simulated data and record the test statistic of interest, $\hat{D}^*$.

3) Repeat step 2 a sufficient amount of times to obtain a distribution of the test

statistic under the null hypothesis.

4) Finally, we will compute p-values by comparing $\hat{D}$ to the distribution of $\hat{D}^*$.

We have effectively used the bootstrap distribution of $\hat{D}^*$ as an approximation to the sampling distribution of $\hat{D}$.

2.4.2.2 Markov Chain Monte Carlo Generalized Linear Mixed Models

One of the nice properties of working within the Bayesian paradigm is that once a sample from the posterior is obtained it is relatively easy to make inference on the resulting model. Oftentimes, this is done using highest posterior density or credible intervals. These intervals are similar but not equivalent to classical statistics confidence intervals. The package *MCMCglmm* calculates credible intervals in the following manner. For the parameter of interest, the interval is constructed from the empirical cumulative distribution function of the sample as the shortest interval for which the difference in the empirical cumulative distribution function values of the endpoints is the nominal probability. This procedure makes the assumption that the resulting posterior is not severely multi-modal. For the purposes of this paper, if the credible interval includes 0 one would conclude that the predictor of interest has no effect on the response. Also, if the interval does not include 0 then the predictor does have some effect on the response.

Now that the estimation and inference procedures have been described in detail, we proceed to assessing them in the following section using large scale simulations.

Chapter 3

Simulation Approach to Evaluating Estimation and Inference Methods

3.1 Introduction

The problem of selecting a best approach for estimation and inference within GLMMs is very complex. The problem is too hard to solve analytically so we are forced to implement a simulation study. The focus of this chapter is on obtaining a variety of properties associated with the selected methods. This is achieved by the implementation of a simulation study.

Section 3.2 is devoted to the accuracy of the estimation procedures. Given just one possible set of parameters (otherwise known as unique scenarios) we employ all of the estimation methods and see how close the estimates are to the true values. Section 3.2 is also primarily concerned with the finite sampling distribution. We select the most stable method `glmer` and run a large scale simulation in order to mimic the behavior of the finite sampling distribution. We do this for two reasons. The first is that we want to compare the other estimation procedures to the sampling distribution while the second is to determine under what conditions the sampling distribution follows a normal distribution. This is relevant for constructing confidence intervals. In Section 3.3 we switch our focus to inferential procedures. Here, we are interested in which of

the procedures provide the best power associated with detecting a temporal trend for a variety of conditions. Finally, in the fourth section we apply each of the estimation and inferential procedures to 2 data sets provided by The National Parks of Canada.

3.2 Estimation and sampling Distribution

In this section we evaluate the selected methods precision with respect to different estimation approaches. In other words, we want to see how close the estimated values are to the true parameter values. Distinct parameter settings let the number of groups and time points vary. The goal here was to see the effect different sampling sizes have on estimation.

	β_0	σ_{β_0}	β_1	σ_{β_1}	N_g	N_t	N_r
Original	$\log_e(10)$	0.5	0.1	0.05	10	10	1
Reduced Years	$\log_e(10)$	0.5	0.1	0.05	10	5	1
Reduced Groups	$\log_e(10)$	0.5	0.1	0.05	5	10	1
Reduced Both	$\log_e(10)$	0.5	0.1	0.05	5	5	1

Table 3.1: Estimation Scenarios. This table depicts combinations of parameters or unique scenarios we will be sampling from in Sections 3.1 and 3.2. N_g, N_t, and N_r denoting number of sampling groups, number of sampled time points, and number of repetitions per time point per group respectively.

All computation was completed using *R* version 2.12.0, *MASS* version 7.3-7, *lme4* version 0.999375-35, and *MCMCglmm* version 2.06. Given a set of parameters, we simulated 1000 Poisson GLMM data sets and applied each of the 4 estimation methods. For parametric bootstrapping we used 200 bootstrap samples. In this section bootstrapping is used as an estimation method, bootstrapping from the full model. However, to use parametric bootstrapping as an inferential procedure we sampled from the null distribution. For `MCMCglmm` we stayed as close to default input values as possible, with 13000 iterations 3000 of which were a burn-in period. The prior for the random effect (co)variance matrix is an inverse Wishart distribution with 1

and .002 scale and degree of freedom parameters, respectively. We set the additive over-dispersion parameter to be very close to 0 to mimic traditional GLMM models. For `glmmPQL` and `glmer` we saved the resulting parameter estimates. However, since both parametric bootstrap and `MCMCglmm` give us distributions for each parameter for each simulation we saved a sequence of quantiles. For each simulated data set, the re-sampling procedures will produce a bootstrapped or posterior distribution. We save quantiles of interest as a way to summarize the distributions. In order to compare parametric bootstrap and `MCMCglmm` to the other methods we used the 50^{th} percentile or the median of the associated distributions. We will give a fuller explanation of the sequence of quantiles in latter half of this section.

We simulated using the parameter values given in Table 3.1, where each row is a unique scenario. For instance, the first scenario labeled original (given in the first row of Table 3.1) has 10 groups, each with one observation taken at each of 10 time points resulting in a sample size of 100. The parameter values were $\beta_0 = \log_e(10)$, $\beta_1 = 0.1$, $\sigma_{\beta_0} = 0.5$, and $\sigma_{\beta_1} = 0.05$ with N_g, N_t, and N_r denoting number of sampling groups, number of sampled time points, and number of repeated samples per time point per group, respectively. An example of one simulated data set from the first scenario is shown in Figure 3.1. The intercept and slope for each group does vary, for example if considered independently, group 4 and 10 will have larger slope values than groups 2 and 9. 1000 data sets resembling Figure 3.1 were simulated, the density of the estimated values are shown in Figure 3.2. Considering the first scenario, we find that all methods perform similarly and do a relatively good job at estimating the fixed effects β_0 and β_1. However, for the random effect standard deviations this is not the case. Estimating σ_{β_0} using `glmer`, `glmmPQL` and parametric bootstrapping result in similar estimates, which is not surprising. These estimates are also slightly shifted to the left resulting in an underestimation of σ_{β_0}. `MCMCglmm` is more severely shifted to the left than than the other procedures. Estimates of σ_{β_1} exhibit similar patterns

as that of σ_{β_0} except that `MCMCglmm` is severely under estimating σ_{β_1}. However, σ_{β_1} is small, and variances close to 0 can be hard to estimate particularly for MCMC (personal correspondence D. Bates).

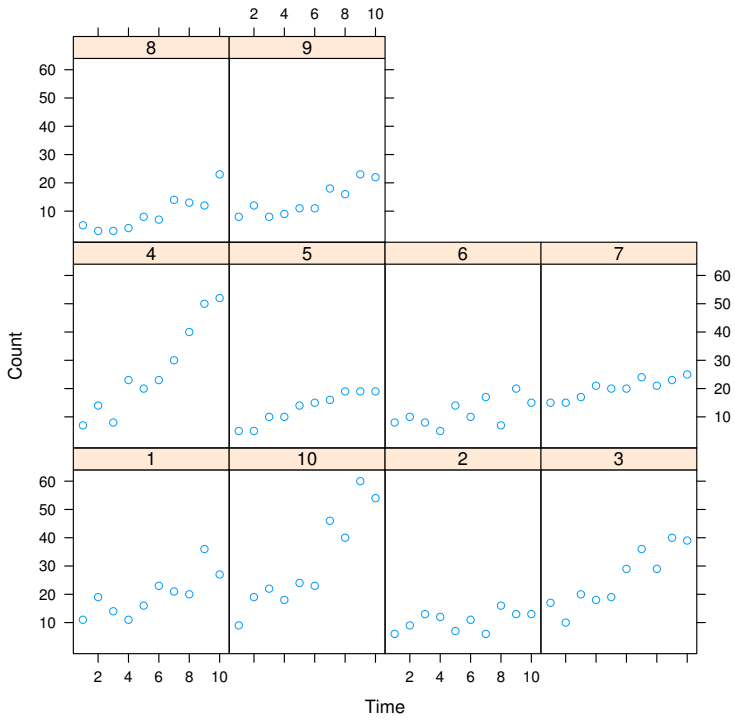

Figure 3.1: Simulated GLMM Data Set with a Poisson response. The parameters are set to $\beta_0 = \log_e(10)$, $\beta_1 = 0.1$, $\sigma_{\beta_0} = 0.5$, and $\sigma_{\beta_1} = 0.05$. $\beta_1 > 0$ signifying an overall increasing trend. Each of the windows within the figure are unique sample groups, where each group has uncorrelated random intercept and slope effects associated with it.

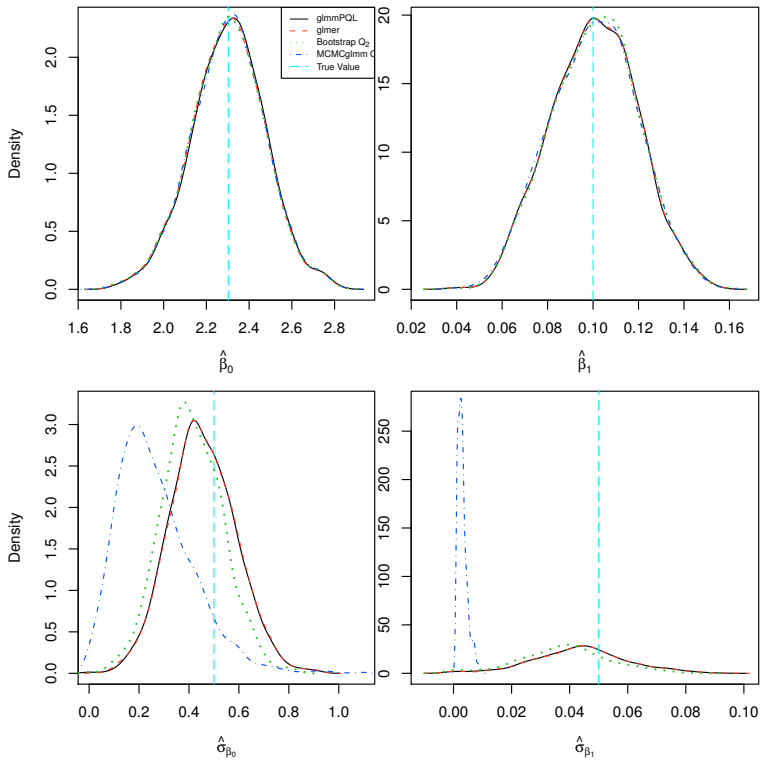

Figure 3.2: Density estimates for each of the methods resulting from 1000 simulations under the same parameter values as Figure 3.1. The methods are distinguished using different types of lines denoted in the legend. The parameter of interest is expressed in each panel's x-axis. For bootstrapping and `MCMCglmm` the 50^{th} quantile of the parameters are presented.

Other scenarios were included under a similar simulation framework (1000 data sets were simulated for each scenario). The scenarios have the same parameters as the first unique scenario but reduced the number of time points sampled to 5 for the first additional scenario. For the second, we held number of time points at 10 and reduced the number of sample groups to 5. Finally for the third, we reduced both the number of groups and time points to 5. These scenarios are also shown in Table 3.1. These simulation results provide a sense of how the procedures behave under less satisfactory sampling conditions. Reducing the number of sample time points to 5 resulted in β_0 being slightly underestimated for each of the methods. Estimates for β_1 remain on target. $\hat{\sigma}_{\beta_0}$ remains underestimated for all of the methods. `MCMCglmm` grossly under

estimates $\hat{\sigma}_{\beta_1}$, with a peak at 0. `glmer`, `glmmPQL`, and bootstrapping under estimate σ_{β_1}. Additional simulations from the parameter settings labeled reduced years and reduced groups (not included as a figure) exhibit similar patterns but with reduced accuracy and precision. Estimates of σ_{β_1} get flatter but with peaks forming at 0 for each of the methods. Reducing the number of sample groups and time points results in estimates for the fixed effects remaining stable while the estimates for the random effect standard deviation generally deteriorate.

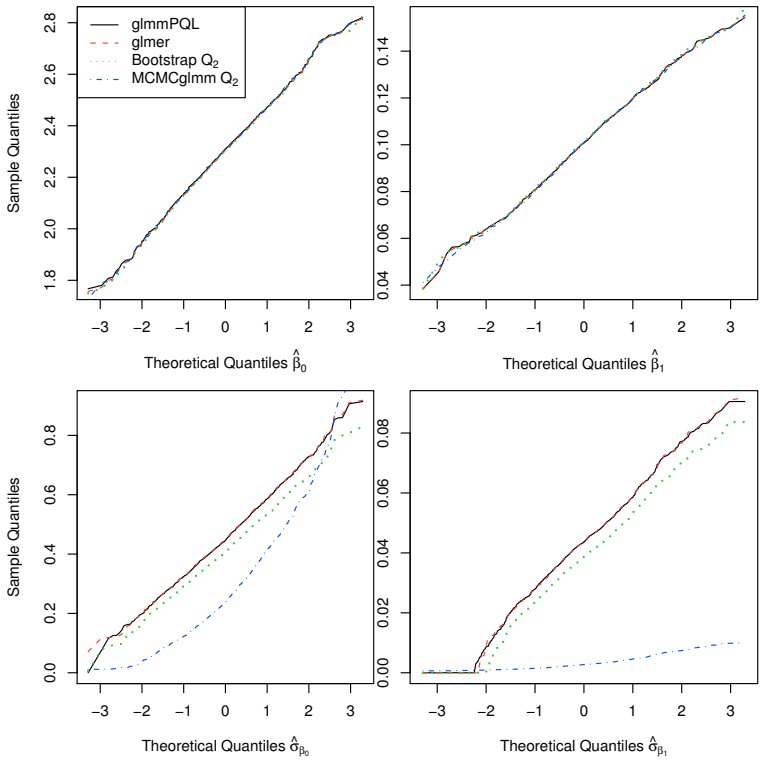

Figure 3.3: Quantile Normal plots. The methods are distinguished using different types of lines denoted in the legend. The parameter of interest is expressed in each panel's x-axis. For bootstrapping and `MCMCglmm` the 50^{th} quantile of the parameters are presented. Curves were plotted with respect to the scale of `glmmPQL`. The data was simulated using the first unique scenario described in Table 3.1.

We are also curious if these estimates obtained from simulating from the first unique scenario follow a normal distribution, particularly in the interest of constructing confidence intervals. If the parameter estimates do indeed follow a normal distribution then one can construct confidence intervals in the conventional manner using the equation $\hat{\theta} \pm Z_{\alpha/2}{}^{\mathrm{SE}(\hat{\theta})}/_{\sqrt{n}}$. Wald tests need a specific distribution to perform adequately. Figure 3.3 contains normal plots for each of the methods with respect to parameters of interest for the first scenario. Normally these types of graphs are plotted with one line and if the line lies on $x = y$ this provides us with evidence that the distribution is normally distributed. However, if the distributions are linearly related, the points in the plot will approximately lie on a straight line. Therefore, curves that are roughly straight will provide evidence that the estimates are normally distributed. Referring to Figure 3.3 we can see that the plots for the fixed effects β_0 and β_1 are all straight and therefore are distributed normally. However, the same cannot be said about the random effect standard deviation estimates. For `glmer`, `glmmPQL` and bootstrapping the plots of σ_{β_0} are approximately straight, indicating that they likely follow a normal distribution. However, σ_{β_0} for `MCMCglmm` is not straight therefore it isn't likely that it follows a normal distribution. Considering σ_{β_1}, we see that the curves are not straight and therefore appear non-normal.

Also note, we chose to employ `glmer` within parametric bootstrapping because it is more stable. For the parameters we have selected, it would make little difference, as `glmer` and `glmmPQL` seem to be performing identically with respect to estimation. Also, `glmmPQL` does not provide a likelihood upon which we base our test statistic.

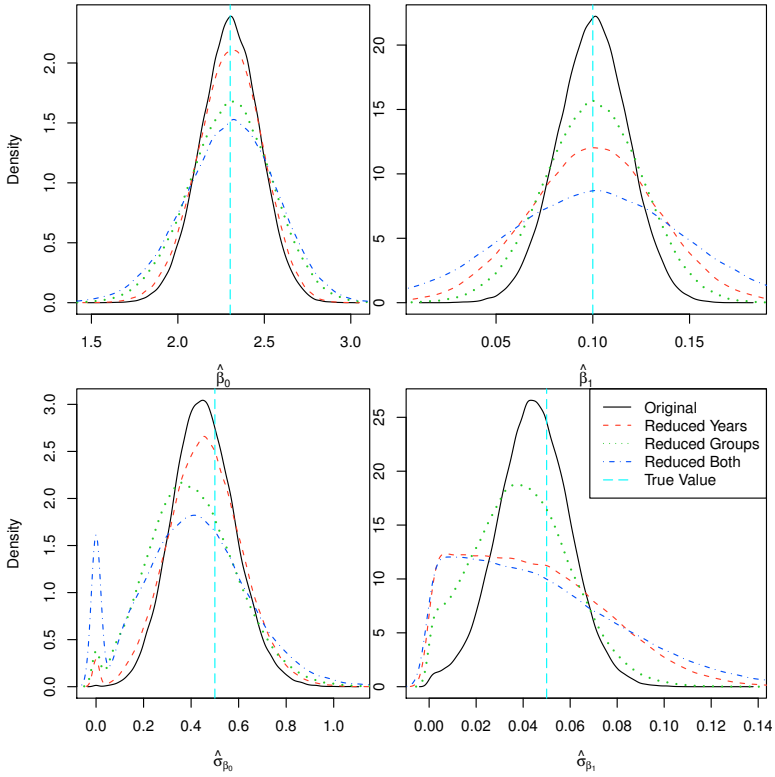

Figure 3.4: Density of finite sampling distributions simulated under 4 sets of parameter values. Sample size parameter values and the true parameter of interest are distinguished using different line types which are indicated in the legend. The parameter of interest is denoted in the x-axis of each panel. Estimates were obtained from 100, 000 simulations applying `glmer`.

In the second half of this section we summarize the results of a large scale simulation designed to get a sense of the properties of the finite-sample distribution of the model parameters. We conduct this simulation to see how close the parameter distributions obtained from the re-sampling approaches are to the sampling distribution, in addition to wanting to know if the sampling distribution follows a normal distribution. If so, we construct confidence intervals in the conventional way. We construct the finite sampling distribution by applying our most stable estimation procedure (`glmer`) to an large number of simulations (100, 000) and save the parameter estimates.

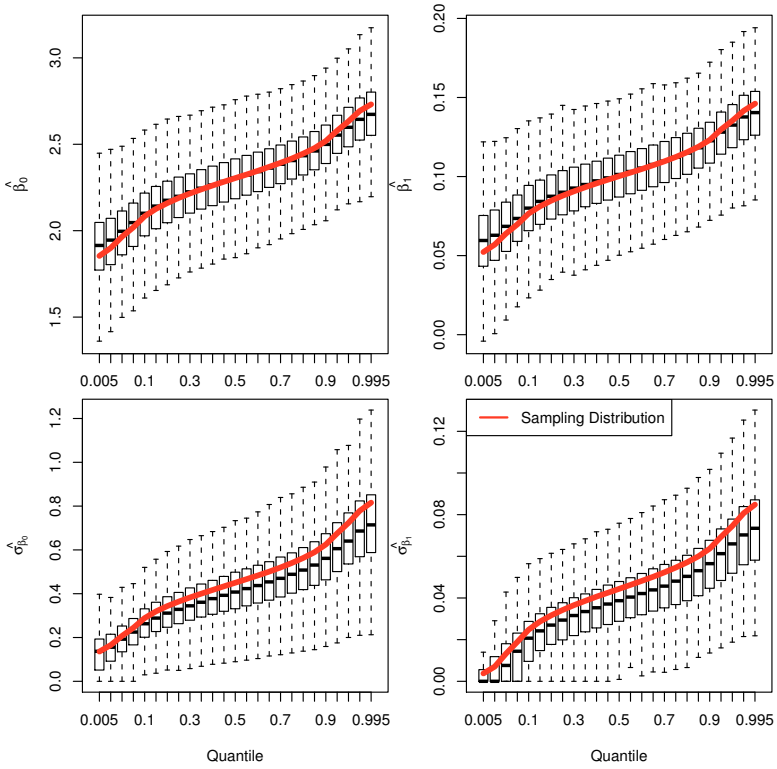

Figure 3.5: Bootstrapped Quantile Distribution. The figure contains box plots for bootstrapped quantile estimates. The quantiles of interest are denoted in the x-axis of each panel. The parameter of interest is expressed in the y-axis of each panel. We have also included quantiles of the discrete sampling distribution as a reference frame for the bootstrapped estimates as is implied in the legend. The data was simulated under the first scenario in Table 3.1.

Using the same parameter values shown in Table 3.1, we simulate $100,000$ data sets and estimate model parameters using `glmer`. The resulting density plots are shown in Figure 3.4. The first set of parameter values (which we'll refer to as the original parameter values) resulted in sampling distributions that look normally distributed, with the random effects being slightly underestimated as expected [28]. However, when we reduce group size and number of time points, the fixed effects estimates become more variable around the true value. The distribution of the random effect standard deviation estimates act very strangely: $\hat{\sigma}_{\beta_0}$ develops a local maximum close to 0, while that of $\hat{\sigma_{\beta_1}}$ becomes skewed with additional probability mass

near 0. This is not surprising as the sampling distribution of variance estimates is in general strongly asymmetric [7]. As mentioned in the previous section, parametric bootstrapping and `MCMCglmm` both effectively estimate a distribution for the parameters of interest. We used 200 data bootstrapped samples. We simulated 1000 data sets under the first unique scenario and saved quantile estimates for Bootstrapping and `MCMCglmm`. The quantiles of interest are the .5%, 1%, 2.5%, 5%, 10%, $\cdots$, 95%, 97.5%, 99%, 99.5% quantiles. The resulting plots are shown in Figures 3.5 and 3.6, respectively. The quantile estimate curves of the sampling distribution are also included in these figures. We see when comparing the bootstrapped quantiles to the sampling distribution that the fixed effects are estimated adequately. However, bootstrapping tends to underestimate both $\hat{\sigma}_{\beta_0}$ and $\hat{\sigma}_{\beta_1}$. `MCMCglmm` also approximates the sampling distribution for the fixed effects adequately. However, it usually underestimates $\hat{\sigma}_{\beta_0}$ and performs horribly for $\hat{\sigma}_{\beta_1}$ especially in the tails. The sampling distribution's quantile estimates are well above the spread of the $\hat{\sigma}_{\beta_1}$ posterior distribution produced by `MCMCglmm`. Given the aforementioned estimation results it seems best to employ `glmer` or `glmmPQL` for estimation in practice, with a slight preference towards `glmer`, as it is approximating a true likelihood and not a quasi likelihood. Furthermore, under this balanced design the discrete sampling distribution for the fixed effects appears to be distributed normally so it is reasonable to construct confidence intervals in the conventional manner.

In this section we have observed a few properties of the estimation procedures. For balanced designs with parameter values given in Table 3.1 the asymptotic estimation procedures perform satisfactory for the fixed effects. Investigation of the finite sampling distributions reveals that constructing confidence intervals for the fixed effects is appropriate. All methods underestimate σ_{β_0} and σ_{β_1} especially `MCMCglmm`.

The next section will proceed to evaluate different methods of inference associated with these estimation procedures.

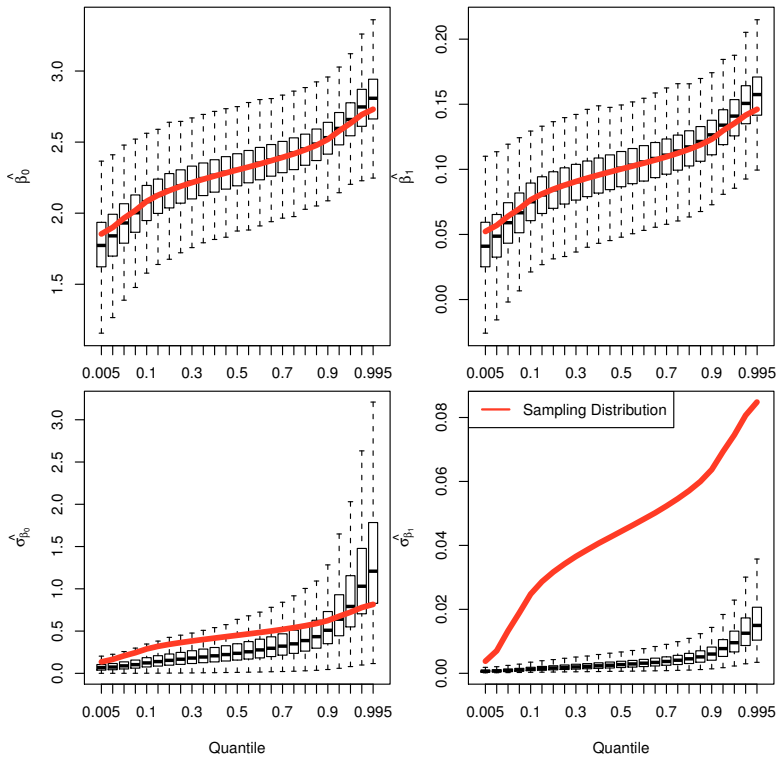

Figure 3.6: MCMC Quantile Distribution. The figure contains box plots for posterior quantile values. The quantiles of interest are denoted in the x-axis of each panel. The parameter of interest is expressed in the y-axis of each panel. We have also included quantiles of the discrete sampling distribution as a reference frame for the posterior quantile values. The data was simulated under the first scenario in Table 3.1.

3.3 Power Analysis Approach

Statistical power is derived from the decisions one can make when performing a statistical test. In standard hypothesis testing, there are two types of error coined Type I and Type II. Type I error occurs if the the null hypothesis (H_0) is rejected when it is true. Type II error occurs if the the null hypothesis is not rejected although the alternative hypothesis (H_1) is true. These decisions are shown in Table 3.2. Statistical power is the probability of rejecting H_0 given H_0 is false [29]. For our purposes, H_0 will be that the time variable has no overall effect on the response. Specifically, H_0 will assume that $\beta_1 = 0$. Furthermore, we will simulate GLMM data

with a Poisson response in which a fixed and random slope is included. An example of such a simulation is shown in Figure 3.1. Knowing that H_0 is false we calculate the probability of rejecting the null hypothesis for each of the inferential procedures using a large scale simulation.

	H_0 True	H_1 True
Fail to Reject H_0	✓	Type II Error
Reject H_0	Type I Error	✓

Table 3.2: Decision Table.

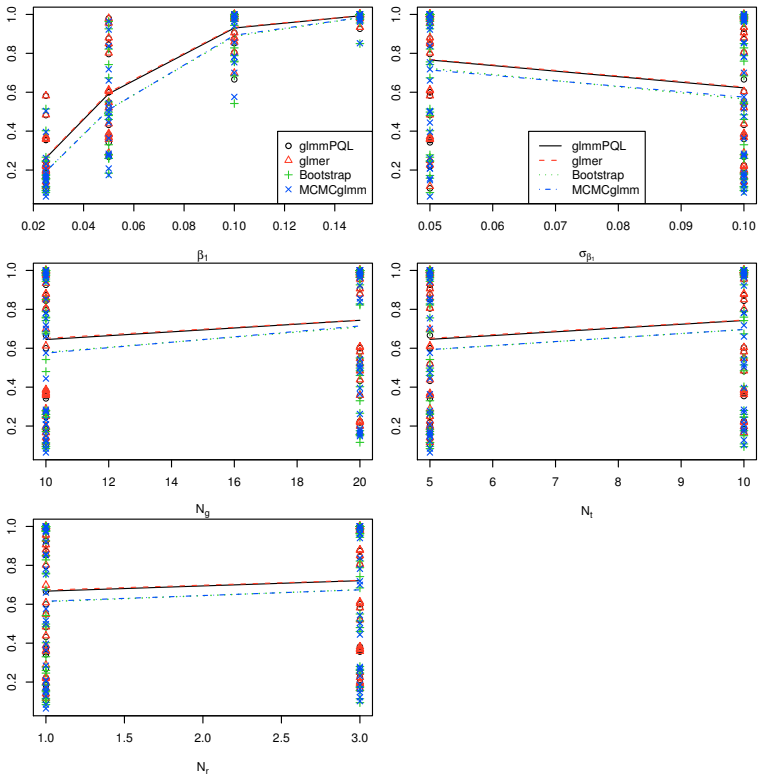

Figure 3.7: Power Figure. Each window plots factor levels along the x-axis and power along the y-axis for each corresponding inferential procedure. Each procedure is denoted by distinguishing symbols and colors as seen in the left legend. The mean of each procedure's power for each level are depicted using distinguishing lines and colors as is denoted in the right legend.

Power analysis is useful in calculating minimal sample size parameters required to accept the outcome of inferential procedures. In our case, factors that effect the

sampling framework are the number of groups (N_g), number of time points (N_t), and repeat samples within a group for each time point (N_r). Power analysis also gives us a way to calculate a lower bound for the effect size that experimental design is likely to detect. In particular, for a fixed sample size we'll have an idea of how small the magnitude of β_1 can be in order for us to detect a trend. But, this is further complicated by σ_{β_1}, as the coefficient of variation increases ($c_v = \frac{\sigma_{\beta_1}}{\beta_1}$) we expect the procedure's ability to detect a trend to deteriorate.

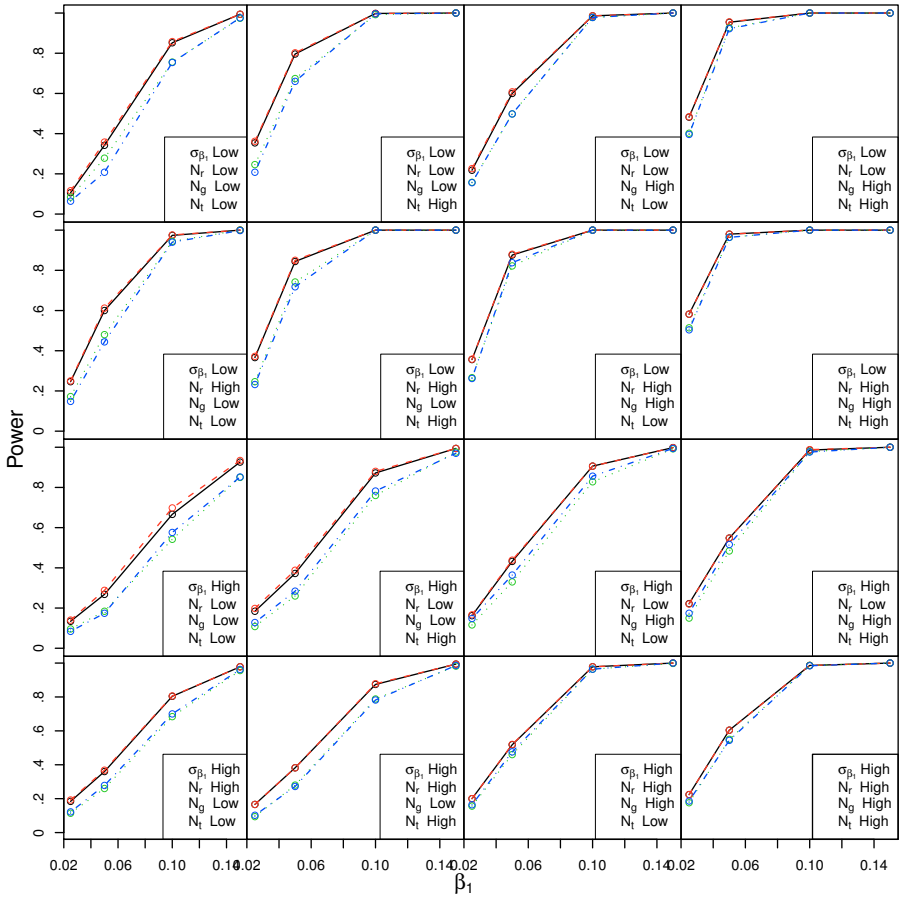

Figure 3.8: β_1 Power Figure. This figure depicts the dynamics of statistical power with respect to β_1 holding all other factors constant for each of the inferential procedures. Each window correspond to one of the 16 unique combinations of σ_{β_1}, N_g, N_t, and N_r used in the power simulation. The windows are ordered from left to right along the rows which correspond to the first 16 rows of Table 3.3. To illustrate, the window in the second row third column has the same power, σ_{β_1}, N_g, N_t, and N_r values as the seventh row of Table 3.3. Values for β_1 are denoted along the x-axis with statistical power on the y-axis. Corresponding inferential procedure are distinguished with different line colors. Black, red, yellow, and blue are associated with `glmmPQL`, `glmer`, Parametric Bootstrapping, and `MCMCglmm`, respectively.

We implemented the power simulation in the following manner. We begin by setting the values of β_1, σ_{β_1}, N_g, N_t, and N_r. We'll refer to these different combinations as unique scenarios. For all scenarios β_0 and σ_{β_0} are set to $\log_e(10)$ and .5, respectively. We keep these constant as to not add more dimensions to our simulation. There are already 64 unique scenarios, if we had let β_0 and σ_{β_0} vary even with just

two levels each there would have been upwards of 256 scenarios. Since we are also simulating 500 times for each scenario it would have not been feasible for the scope of this thesis. We are also not primarily concerned with β_0 and σ_{β_0} as they are nuisance parameters, we fix their values while recognizing that different values of β_0 and σ_{β_0} could affect our power calculations.

Given a unique scenario we simulate 500 data sets and for each data set and apply the estimation and inferential procedures. For `glmmPQL` and `glmer` we declare they have detected a trend if their associated p-value is less than .05. In other words, our testing level (α) is set to .05. For parametric bootstrapping we use 200 bootstrapped simulations and declare that parametric bootstrapping has detected a trend if the p-value $<$ .05. We then carry out `MCMCglmm`, calculating a 95% highest density interval for β_1. If 0 is not included in the interval we declare that β_1 contributes significantly to the model.

We do this for each of the simulated data sets obtaining a power value for the initial unique scenario by computing the number of times we detected a trend and divide by the number of iterations.

Selecting the levels for the unique scenarios was done pragmatically. We used values that were likely to be estimated in ecological GLMMs. We designed the experiment to include 5 factors, each having 2 levels (a high and a low level) except β_1 which has 4. It was hypothesized that the magnitude of β_1 was the most important factor. Each of the 64 unique scenario power simulation took approximately 60 hours of computational time using the cluster provided by the Mathematics and Statistics Department of Dalhousie University. Leading to approximately 3840 hours of computation. Of course, we simulated in parallel to reduce the overall computation time considerably.

The results from the power simulation are shown in Table 3.3. Combinations of parameter values are located on the left of the vertical line while the associated power

for each of the inferential methods is on the right. Another perspective of the power results is shown in Figure 3.7. This figure depicts graphically how change in factor level effects power for each of the inferential procedures.

For a perspective of how β_1 effects statistical power while holding all other factors constant we construct Figure 3.8. We generally see that power increases as we increase the levels of β_1, which is not surprising. The lowest statistical power for all of the inferential procedures are given when $N_g = 10$, $N_t = 5$, $N_r = 1$, and $\sigma_{\beta_1} = .10$ as seen on rows 9, 25, 41, and 57 of Table 3.3, and in the third row first column in Figure 3.8.

In order to get a sense of what factors contribute to each inferential procedure's power we implemented a standard analysis of variance (ANOVA) on the results of the power simulations. We included all 2 factor interactions. We transformed the statistical power (our response) using the following the equation logit(power-.01) thus releasing the restriction that the response must be on the interval $[0, 1]$. Analysis of the Wald test obtained using `glmmPQL` revealed that all the main effects are indeed significant ($\alpha = 0.05$), with three two level interactions contributing significantly to the response. The highest levels of variation in ANOVA come from the factors: β_1, σ_{β_1}, N_g, N_y, and N_r, in descending order. Furthermore, higher levels of β_1, N_r, N_t, N_g lead to higher power. However, increasing the level of σ_{β_1} decreases our ability to detect a trend within `glmmPQL`. Considering the significant interactions we observe a positive two level interaction between N_r and N_t, and between β_1 and N_g, implying an added compound effect when moving to higher levels of these factors. However, we observe a negative two level interaction effect between β_1 and σ_{β_1} and between σ_{β_1} and N_t. That is, power is not as high as it would be considering only the main effects given the increase in power associated with β_1 and decrease associated with σ_{β_1}, the same goes for σ_{β_1} and N_t.

The same analysis as above was carried out with our response being the power of the Wald test obtained using `glmer` and found similar results. All the main effects were significant along with the two level interactions between N_r and N_t, β_1 and N_g, σ_{β_1} and N_t, and between β_1 and σ_{β_1}. The interactions that were positive above remained positive and the negative interaction remained negative. These similar results are not surprising given both procedures use the same inferential method.

We'll now consider the power of parametric bootstrapping. As seen in Table 3.3 there are two variations of parametric bootstrapping. Their distinction lies in the formulation of the null distribution. As we have stated previously, we obtain the test statistic as the difference in the log likelihoods of the full and null models. We construct the distribution of the the test statistic under the null hypothesis by simulating under the null model. The first variation of parametric bootstrapping (referred to as Parametric Bootstrapping) defines the null model as having only β_1 set to 0. Within the second variation (which we'll refer to as modified parametric bootstrapping) β_1 and σ_{β_1} are both set to zero. For a fair comparison to the other inferential procedures which are term-wise, we'll focus on the first variation of parametric bootstrapping. However, the problem with this procedure is that we bootstrap from a model with a random effect and no associated fixed effect which appears to be unconventional. But, as seen in Table 3.3 power increases with σ_{β_1} for modified parametric bootstrapping and remains quite high when β_1 is on its lowest level. More discussion with follow on this topic.

Continuing the description of the procedures, the test statistic is always at least zero because the log likelihood of the full model is always at least as big as the log likelihood of the null model. We parametrize the full and null models using maximum likelihood not restricted maximum likelihood within `glmer`. We are comparing

N_g	N_t	N_r	β_1	σ_{β_1}	glmmPQL	glmer	P BS	MCMCglmm	Mod P BS
10	5	1	0.050	0.05	0.34	0.36	0.28	0.21	0.52
10	10	1	0.050	0.05	0.80	0.80	0.67	0.66	0.99
20	5	1	0.050	0.05	0.60	0.61	0.50	0.50	0.84
20	10	1	0.050	0.05	0.95	0.96	0.93	0.92	1.00
10	5	3	0.050	0.05	0.60	0.61	0.48	0.44	0.90
10	10	3	0.050	0.05	0.84	0.85	0.74	0.72	1.00
20	5	3	0.050	0.05	0.88	0.88	0.82	0.84	0.99
20	10	3	0.050	0.05	0.98	0.98	0.96	0.96	1.00
10	5	1	0.050	0.10	0.27	0.29	0.18	0.17	0.79
10	10	1	0.050	0.10	0.37	0.39	0.26	0.28	1.00
20	5	1	0.050	0.10	0.43	0.44	0.33	0.36	0.96
20	10	1	0.050	0.10	0.55	0.55	0.48	0.52	1.00
10	5	3	0.050	0.10	0.36	0.37	0.26	0.28	0.97
10	10	3	0.050	0.10	0.38	0.38	0.28	0.27	1.00
20	5	3	0.050	0.10	0.52	0.52	0.46	0.48	1.00
20	10	3	0.050	0.10	0.60	0.60	0.55	0.54	1.00
10	5	1	0.150	0.05	0.99	0.99	0.97	0.98	1.00
10	10	1	0.150	0.05	1.00	1.00	1.00	1.00	1.00
20	5	1	0.150	0.05	1.00	1.00	1.00	1.00	1.00
20	10	1	0.150	0.05	1.00	1.00	1.00	1.00	1.00
10	5	3	0.150	0.05	1.00	1.00	1.00	1.00	1.00
10	10	3	0.150	0.05	1.00	1.00	1.00	1.00	1.00
20	5	3	0.150	0.05	1.00	1.00	1.00	1.00	1.00
20	10	3	0.150	0.05	1.00	1.00	1.00	1.00	1.00
10	5	1	0.150	0.10	0.93	0.93	0.85	0.85	1.00
10	10	1	0.150	0.10	0.99	0.99	0.98	0.97	1.00
20	5	1	0.150	0.10	1.00	1.00	0.99	0.99	1.00
20	10	1	0.150	0.10	1.00	1.00	1.00	1.00	1.00
10	5	3	0.150	0.10	0.98	0.98	0.96	0.96	1.00
10	10	3	0.150	0.10	0.99	0.99	0.98	0.99	1.00
20	5	3	0.150	0.10	1.00	1.00	1.00	1.00	1.00
20	10	3	0.150	0.10	1.00	1.00	1.00	1.00	1.00
10	5	1	0.100	0.05	0.85	0.86	0.76	0.75	0.93
10	10	1	0.100	0.05	1.00	1.00	0.99	1.00	1.00
20	5	1	0.100	0.05	0.99	0.99	0.98	0.98	1.00
20	10	1	0.100	0.05	1.00	1.00	1.00	1.00	1.00
10	5	3	0.100	0.05	0.97	0.98	0.94	0.94	1.00
10	10	3	0.100	0.05	1.00	1.00	1.00	1.00	1.00
20	5	3	0.100	0.05	1.00	1.00	1.00	1.00	1.00
20	10	3	0.100	0.05	1.00	1.00	1.00	1.00	1.00
10	5	1	0.100	0.10	0.67	0.70	0.54	0.58	0.97
10	10	1	0.100	0.10	0.87	0.88	0.76	0.78	1.00
20	5	1	0.100	0.10	0.91	0.91	0.83	0.86	1.00
20	10	1	0.100	0.10	0.99	0.99	0.98	0.98	1.00
10	5	3	0.100	0.10	0.80	0.81	0.68	0.70	1.00
10	10	3	0.100	0.10	0.87	0.88	0.79	0.78	1.00
20	5	3	0.100	0.10	0.98	0.98	0.96	0.96	1.00
20	10	3	0.100	0.10	0.99	0.99	0.99	0.98	1.00
10	5	1	0.025	0.05	0.11	0.12	0.08	0.06	0.31
10	10	1	0.025	0.05	0.35	0.36	0.25	0.21	0.93
20	5	1	0.025	0.05	0.22	0.23	0.16	0.16	0.51
20	10	1	0.025	0.05	0.48	0.48	0.40	0.40	1.00
10	5	3	0.025	0.05	0.25	0.25	0.17	0.15	0.65
10	10	3	0.025	0.05	0.37	0.37	0.25	0.23	1.00
20	5	3	0.025	0.05	0.36	0.36	0.27	0.26	0.86
20	10	3	0.025	0.05	0.58	0.58	0.51	0.50	1.00
10	5	1	0.025	0.10	0.13	0.14	0.10	0.08	0.69
10	10	1	0.025	0.10	0.18	0.20	0.11	0.13	1.00
20	5	1	0.025	0.10	0.16	0.17	0.12	0.15	0.91
20	10	1	0.025	0.10	0.22	0.22	0.15	0.17	1.00
10	5	3	0.025	0.10	0.18	0.19	0.11	0.12	0.97
10	10	3	0.025	0.10	0.17	0.17	0.09	0.10	1.00
20	5	3	0.025	0.10	0.20	0.20	0.16	0.16	1.00
20	10	3	0.025	0.10	0.22	0.22	0.18	0.19	1.00

Table 3.3: Power Simulation Results. This table contains the results from our power simulation. The factors of interest are expressed on the left of the table. Resulting in 64 combinations or unique scenarios. The corresponding power of each unique scenario are expressed on the right of the table for each of the inferential methods. P BS and Mod P BS are abbreviations for parametric bootstrapping and modified parametric bootstrapping. N_g, N_t, and N_r denote number of groups, number of time points or years, and number of repetitions per time points per group, respectively.

likelihoods so using ML is more appropriate [2]. Finally, we calculate the p-value as the density to the right of the test statistic of the bootstrapped distribution.

A few more words on the distinction of the variation of parametric bootstrapping. The issue arises when we have to say what we mean by the statement that the slope is 0. There are two possible interpretations. The first interpretation associated with Parametric bootstrapping is that $\beta_1 = 0$. The second interpretation associated with modified parametric bootstrapping is that $\beta_1 = 0$ and $\sigma_{\beta_1} = 0$. The disadvantage with the first approach as that we go against convention and parametrize a model with a random effect without the associated fixed effect. The disadvantage of the second approach is that we would have to simulate data in the following manner: as β_1 moves towards 0 then σ_{β_1} should also approach 0.

To clarify, in Table 3.3 one can see a lot of high power values for modified parametric bootstrapping. In order to shed some light on the issue we simulated from four unique scenarios under both the above null hypotheses. Each scenario was simulated 250 times performing modified parametric bootstrapping and obtaining a p-value upon each iteration. Under the null hypothesis with $\beta_1 = 0$ and $\sigma_{\beta_1} = 0$ we found that the p-values were approaching a uniform distribution on interval $[0, 1]$. Consequentially, the test would reject H_0 with respect to the selected critical region given H_0 is true. On the contrary, if we implement modified parametric bootstrapping to data that has been simulated under the other null hypothesis, that is β_1 and $\sigma_{\beta_1} \neq 0$ than the p-values are not uniform. We decided for a fair comparison to the other inferential procedures we would focus on parametric bootstrapping, that is when we exclude only the fixed effect β_1 from the null model. Similarly as above, we applied ANOVA to see what factors effect the power of parametric bootstrapping. In this instance, all the main effects are declared significant. We also observed the exact same significant two factor interactions as `glmmPQL` and `glmer` which are between N_r and N_t, β_1 and N_g, σ_{β_1} and N_t, and between β_1 and σ_{β_1}. The interactions that were

positive above remained positive and the negative interaction remained negative. Finally, we analyze `MCMCglmm` as above and observe similar results obtained for `glmmPQL` and `glmer`. We should also note that although power is interpreted differently in a Bayesian context, power does remain a useful measure. It symbolizes how much the unique scenarios can be expected to refine our prior beliefs. For `MCMCglmm` power, all the main effects are significant. Also, there is a significant positive interaction between N_r and N_t, and a negative interaction between β_1 and σ_{β_1}, and between σ_{β_1} and N_t.

We have seen that increasing β_1, N_t, N_g, and N_r contributes to higher power for all of the methods. However, for all inferential procedures increasing σ_{β_1} results in decreased power, thus a decreased probability of detecting an existing trend. Often, researchers have some degree of control over N_g N_t and maybe even N_r. Therefore, sampling procedures which have been designed to have a higher number of N_g or N_t are more likely to detect a trend. According to the ANOVA, N_g had the highest amount of variation out of the factors a researchers can control. The significant interactions in all the above analyses do appear reasonable. For example, increasing both N_r and N_t does increase the sample size. This leads to a higher compounded power than if both N_r and N_t were effecting power independently. We'll now compare each of the inferential procedures, in conjunction with making recommendations on the design of sampling that would maximize our ability to detect a trend given different the procedures.

3.3.1 Comparison of Inferential Procedures

With respect to power, `glmer` outperforms all of the selected inferential procedures with `glmmPQL` slightly behind it. Parametric bootstrapping and `MCMCglmm` performed the worst. The resampling approaches generally obtain about .10 less statistical power than the asymptotic approaches for moderate values of β_1. There was no crossover

of power between unique scenarios and procedures except for between `MCMCglmm` and Parametric Bootstrapping. In other words for the other methods, if one procedure outperformed another it did so for all unique scenarios. `MCMCglmm` and parametric bootstrapping outperformed the other exactly 24 times while obtaining the same power 16 times. This is all clear upon viewing Table 3.3. With respect to the design of the simulated data, the asymptotic inferential procedures perform better than the re-sampling approaches. However, we expect that as we move away from balanced designs and include outliers that asymptotic procedures will not perform as well as the re-sampling approaches.

3.3.2 Design Implications

While there is no formal standard on what specifies an adequately powered test, a good rule of thumb is for a test to have power of at least .80. Given we specify a critical value of .05, ($\alpha = .05$) we would have a four to one ratio between the probability of Type II and Type I error. That is, the risk of a false negative would be four times as large as a false positive. In order words, if the power is too low it is not beneficial to implement an experiment.

Researchers are obviously subject to designing their study based upon unknown β_1 and σ_{β_1}. Considering all of the inferential procedures, we see that if $c_v < 1$ (equivalently $\beta_1 > \sigma_{\beta_1}$) all procedures have relatively adequate power with only a few values for `MCMCglmm` dipping below .80. However, when $c_v > 1$ this ceases to be the case. In this situation, researchers must have high levels N_g, N_t, and N_r to ensure adequate power. When the magnitude of β_1 is half the size of σ_{β_1} even having high levels of N_g, N_t, and N_r is not enough to ensure power of .80. If $c_v = 1$ we have adequate power values provided not all of N_g, N_t, and N_r are low. When β_1 is as low as .025 all inferential procedures will be hard pressed to detect a trend.

3.4 Analysis of Real Application

In this section we apply each of the methods to two data sets sampled by The National Parks of Canada. The Salamander data set (as seen in Figure 1.1) serves as an example in which the random slope effect does not appear to be high. We include this analysis to illustrate how the methods behave when there is little slope variability associated with the grouping structure. Since the response is a count, it is conventional to model the response with a Poisson distribution. We include fixed intercept and slope parameters associated with year. We'll assume the random intercept and slope effects associated with group are independent of each other.

Table 3.4: Salamander Estimation. This table contains GLMM estimates obtained using the procedures outlined in the left column for the Salamander data set shown in Figure 1.1.

Parameter	Estimates		95% HD Intervals	
	`glmmPQL`	`glmer`	`MCMCglmm`	PBS
β_0	4.147	4.088	(3.734, 4.426)	(3.753, 4.357)
β_1	-0.09901	-0.08806	(-0.129, -0.03959)	(-0.3869, 0.2184)
σ_{β_0}	0.3283	0.5005	(0.09902, 0.6814)	(0.262, 0.6924)
σ_{β_1}	2.684e-06	0.05709	(0.0007945, 0.01058)	(0.2943, 0.7146)

Table 3.5: Salamander Inference. This table contains the results of the inferential procedures applied to the salamander data set.

method	p-value/sig
`glmmPQL`	3.446e-07
`glmer`	2.107e-06
P BS	≈ 0.005
`MCMCglmm`	sig value $< .05$

Employing each of the estimation methods results in the parameter estimates shown in Table 3.4. Estimates for β_0 and β_1 supplied by `glmmPQL` and `glmer` are similar for the fixed effects. 95% highest density intervals of the posterior distribution obtained from `MCMCglmm` for the fixed effects are also shown. The intervals include the estimated values obtained from `glmmPQL` and `glmer`. We are more interested in how

the estimating procedures cope with little variability associated with the grouping of the data. We see that `glmmPQL` produces estimates of σ_{β_1} close to 0. `glmer` attributes a higher standard deviation of the random effects associated with group. The posterior distribution for the standard deviations is quite close to the boundary. Given the low estimate `glmmPQL` provides for σ_{β_1} it may be reasonable to apply `glmmPQL` without a random slope effect. Inferential procedures for β_1 were applied and their results are shown in Table 3.5. Parametric bootstrapping for estimation and inference are not the same. For estimation we bootstrap under the full distribution and record the parameter estimates. However, for inference we bootstrap under the null distribution and record the deviance. Given lower variability with respect to the grouping structure, the estimation and inferential procedures seem to perform adequately.

We have also included salmon data collected by Parks Canada to illustrate estimation and inference methods. Salmon were sampled using closed site multiple pass electro fishing. A section of a stream called a plot was closed off by a net, preventing salmon from exiting or entering. An electro fishing device was run through the area to temporarily stun the fish. Unconscious fish were collected and measured and the process was repeated. Stunned fish were not returned to the stream until sampling was completed. This was done yearly on multiple plots, returning to the same plot every year. For simplicity, we will not consider the sampling sweeping structure and take the response variable to be the sum of salmon caught per plot. This data set is shown in Figure 1.2. In contrast to the Salamander data the Salmon data exhibits more variability in the slope of each plot. We'll model these data using a GLMM, assuming that the response given the random effects follows a Poisson distribution. The predictor is related to the mean response via a log link function. We include intercept and slope fixed effects, denoted β_0 and β_1. We also assume that the intercept and slope random effects are independent and are normally distributed.

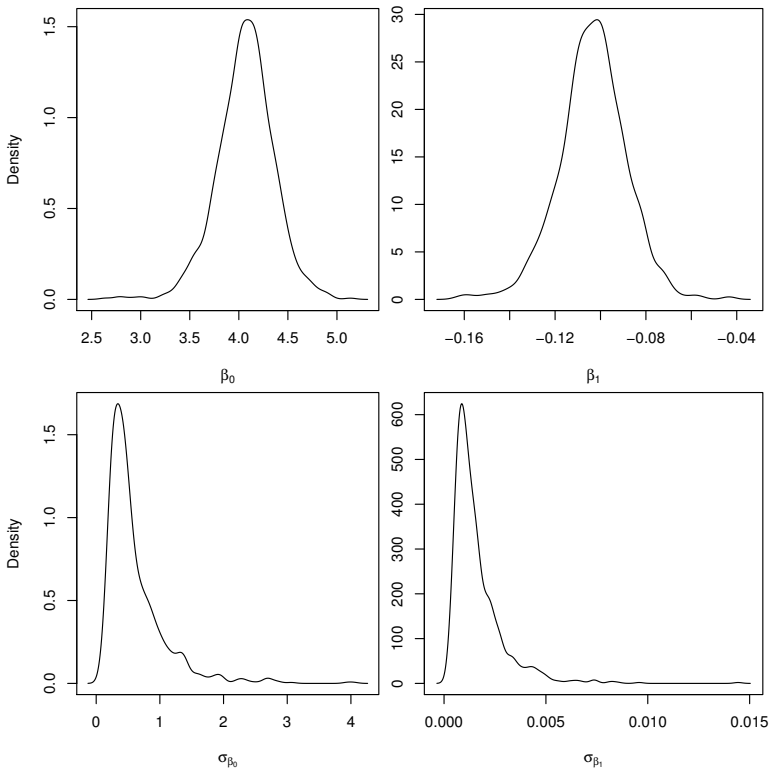

Figure 3.9: Salmon Posterior Distributions. Each window exhibits posterior densities for the Salmon model produced by `MCMCglmm`. Parameters of interest are denoted in the x-axis and density along the y-axis.

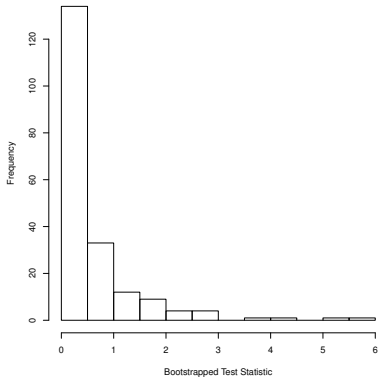

Figure 3.10: Salmon Bootstrapped Test Statistic Histogram. This figure exhibits the distribution of the test statistic derived under the null model using parametric bootstrapping.

Table 3.6: Salmon Estimation. This table contains GLMM estimates obtained using the procedures outlined in the left column for the Salmon data set shown in Figure 1.2.

Parameter	Estimates		95% HD Intervals	
	`glmmPQL`	`glmer`	`MCMCglmm`	PBS
β_0	4.053	4.057	(3.424, 4.581)	(-0.7475, 0.7171)
β_1	-0.09593	-0.1019	(-0.1337, -0.07479)	(-0.1541, -0.06099)
σ_{β_0}	0.4771	0.5816	(0.1009, 1.436)	(0, 1.019)
σ_{β_1}	0.01072	0.02534	(0.0002807, 0.004149)	(0, 0.05449)

Table 3.7: Salmon Inference. This table contains the results of the inferential procedures applied to the Salmon data set.

method	p-value/sig
`glmmPQL`	1.003e-10
`glmer`	1.842e-23
P BS	≈ 0
`MCMCglmm`	sig value $< .05$

Employing each of the methods results in the estimates seen in Table 3.6. The estimates for the fixed effects using `glmmPQL` and `glmer` are quite similar. Both estimate $\hat{\beta}_1 < 0$ indicating that the salmon population is declining each year. A 95% highest density interval of the posterior of β_1 produced by `MCMCglmm` lies within only negative values thus agreeing with the estimates from the other methods. This model also explains variability in the intercept and slope by associating it with different plots. Estimates of σ_{β_0} are high, implying that the initial populations of each plot do vary. Estimates of σ_{β_1} are lower than σ_{β_0} but are not small enough to rule out a random slope effect. Therefore, slope does vary within each plot but in general exhibit a decreasing pattern as is established by a significant negative fixed slope effect estimate.

We are also interested whether the parameter β_1 contributes significantly to the model. This would allow us to say something meaningful about the trend of the Salmon population. Inferential procedure results are shown in Table 3.7. The 3 classical approaches agree with each other. That is, the p-values associated with β_1 are quite small. The histogram of the bootstrapped test statistic under the null

hypothesis is shown in Figure 3.10. The Bayesian approach also agrees with the classical approaches in that the the posterior of β_1 consists of negative values with high density close to `glmmPQL` and `glmer` estimates. A graph of the posterior distribution for all of the parameters is shown in Figure 3.9, and we can say that β_1 contributes to our understanding of salmon populations.

Chapter 4

Conclusion

4.1 Conclusion

Based upon a variety of large scale simulations of a Poisson GLMM with random intercept and slope effects we have observed multiple properties of the selected estimation and inferential procedures. In the long run, `glmmPQL` and `glmer` perform similarly. All methods underestimate the standard deviation of the random effects, especially `MCMCglmm`. This problem is more severe when we are sampling from less than 10 groups, or for less than 10 time points.

Upon generating a discrete sampling distribution we notice that the estimates of the fixed effects are distributed normally, provided there are enough groups and time points. However, the distribution of the random components behaves strangely. This indicates that asymptotic normal confidence intervals for the fixed effects can be appropriate but should not be constructed for the random components. Our primary focus in this thesis was to investigate inferential procedures associated with β_1. We have seen that the Wald test corresponding to `glmer` gives the highest statistical power, with the Wald test associated with `glmmPQL` performing slightly worse. The 95% credible intervals provided by `MCMCglmm` and p-values obtained by parametric bootstrapping produce marginally worse statistical power than the asymptotic approaches. The highest statistical power was observed when β_1, N_g, N_t, and N_r were

high and σ_{β_1} were low. N_g was the most important of the factors that researches normally are able to control. Generally, all the inferential procedures will have difficulty detecting a trend if $\beta_1 < .05$ especially when σ_{β_1} is large. However, increasing the levels of N_g, N_t, and N_r will increase the probability of observing a significant trend. The results of the power simulation would recommend having as many sampling groups and time points as feasible.

4.2 Discussion

It may seem surprising that the asymptotic approaches perform better than the re-sampling approaches. We think this may be a result of the design of our simulated data. We simulated data that was balanced and without outliers. As we move away from such designs the re-sampling approaches may start to out perform the asymptotic approaches.

Within parametric bootstrapping, we had a choice of how to specify the null distribution which we simulate from. We could have let $\beta_1 = 0$ and $\sigma_{\beta_1} \neq 0$ or $\beta_1 = 0$ and $\sigma_{\beta_1} = 0$ (our focus went into the first interpretation but did some analysis corresponding to the second approach as seen in the furthest column to the right in Table 3.3). In order to test the second variation of parametric bootstrapping more fairly we would have had to change how we set up the factors within the power simulation. This would entail reducing σ_{β_1} alongside β_1. We felt it a fairer comparison to the other procedures to implement parametric bootstrapping in the first way. As the second variation would effectively test for both β_1 and σ_{β_1}.

4.3 Future Work

We accomplished what we intended in our investigation of GLMM estimation and inferential procedures. Nevertheless, given the scope of this thesis, there remain a

number of properties that we are interested in. In the future we would like to take our focus away from a Poisson response variable. Although count data are an important part of ecological modeling, we would investigate cases where the response variable follows other members of the exponential family, such as binomial or multinomial. We could also further the analysis by including more levels of each of the contributing factors. Furthermore, we're interested in how unbalanced designs and the inclusion of outliers change power for each of the inferential procedures. One would expect parametric bootstrapping and `MCMCglmm` to become more reliable than conventional inferential procedures in these circumstances. We could also design our simulated data to exhibit a variety of correlation structures. For instance, we could include correlation between the two random effects or include unequal correlation between repetitions of the same group of the same year and the same group of a different year. Likewise, we could design the simulation program to include over and under dispersion. Once we have investigated these properties, we could design a general GLMM method within *R* for Parametric Bootstrapping, as Parametric Bootstrapping can be a challenging procedure for researchers with limited computational background. In closing, many of the problems facing GLMM inference remain an open ended problem that will take considerable time and effort to solve.

Bibliography

[1] R. H. Baayen, D. J. Davidson, and D. M. Bates. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412, 2008.

[2] J. C. Pinheiro and D. M. Bates. *Mixed-effects models in S and S-PLUS*. Springer Verlag, 2009.

[3] J. J. Faraway. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC Press, 2006.

[4] P. McCullagh and J. A. Nelder. Generalized linear models. 1989. *Boca Raton: Chapman &Hall*.

[5] C. E. McCulloch and S. R. Searle. *Generalized, linear, and mixed models*. Wiley-Interscience, 2001.

[6] J. Jiang. *Linear and generalized linear mixed models and their applications*. Springer Verlag, 2007.

[7] D. Bates. lme4: Mixed-effects modeling with r. *URL http://lme4. r-forge. r-project. org/book*, 2010.

[8] J. Albert. *Bayesian computation with R*. Springer Verlag, 2009.

[9] C. P. Robert and G. Casella. *Introducing Monte Carlo Methods with R*. Springer Verlag, 2010.

[10] J. D. Hadfield and L. E. B. Kruuk. Mcmc methods for multi-response generalised linear mixed models: The mcmcglmm r package. *Journal of Statistical Software*, 33(2):1–22, 2010.

[11] *The GLIMMIX Procedure*, 2006.

[12] W. Jang and J. Lim. Pql estimation biases in generalized linear mixed models. Technical report, 2006.

[13] Doug Bates. [r] lmer, p-values and all that. https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html, May 2006.

[14] Benjamin M Bolker, Mollie E Brooks, Connie J Clark, Shane W Geange, John R Poulsen, M Henry H Stevens, and Jada-Simone S White. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol. (Amst.)*, 24(3):127–35, 2009.

[15] N. E. Breslow and D. G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, 1993.

[16] M. K. Cowles and B. P. Carlin. Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434), 1996.

[17] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236, 1974.

[18] G. Casella and E. I. George. Explaining the gibbs sampler. *American Statistician*, 46(3):167–174, 1992.

[19] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Verlag, 2004.

[20] P. Damlen, J. Wakefield, and S. Walker. Gibbs sampling for bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):331–344, 1999.

[21] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. Bayesian data analysis. texts in statistical science series, 2004.

[22] L. A. García-Cortés and D. Sorensen. Alternative implementations of monte carlo em algorithms for likelihood inferences. *Genetics Selection Evolution*, 33(4):443–452, 2001.

[23] T. A. Davis. *Direct methods for sparse linear systems*. Society for Industrial Mathematics, 2006.

[24] F. E. Harrell. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer Verlag, 2001.

[25] M. G. Kenward and J. H. Roger. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53(3):983–997, 1997.

[26] D. W. Gaylor and F. N. Hopper. Estimating the degrees of freedom for linear combinations of mean squares by satterthwaite's formula. *Technometrics*, 11(4):691–706, 1969.

[27] B. Efron. Bootstrap methods: another look at the jackknife. *The annals of statistics*, 7(1):1–26, 1979.

[28] X. Lin and N. E. Breslow. Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association*, 91(435), 1996.

[29] D. C. Montgomery and G. C. Runger. *APPLIED STATISTICS AND PROBABILITY FOR ENGINEERS, With CD*. Wiley-India, 2007.