Chapter 1

Introduction

This thesis concerns the analysis of count data which are correlated, and a major component of it will be the use of Poisson models with random effects. The motivating problem arose from a field experiment conducted by NSW Agriculture in which herbicide efficacy on nutgrass populations in cotton crops was studied over three (3) years and the response was measured by counts of nutgrass tubers. It was anticipated that field data such as this could be both spatially and temporally correlated, hence techniques which could appropriately deal with this correlation are investigated. This data set will henceforth be referred to as the herbicide data set.

Independent count data are conveniently interpreted using Poisson models which are analysed as a Generalized Linear Model (GLM). It was expected that the herbicide data were not independent and the residuals from the naive (independent) analysis readily confirmed this (See correlation matrix (2.2) and Figure 2.3). Thus, independence, a major assumption of simple GLMs, was not tenable in this situation.

In recent years there has been a recognition of the increasing requirement of consulting statisticians to model complex non-normal data. For example, in recent literature, the need to model longitudinal data with adjustments for inherent random effects (clustering) has been highlighted.

McDermott and Schukken 1994) reviewed 67 papers from the first 10 volumes of *Preventitive Veterinary Medicine* to examine how the issue of clustering (random effects) was addressed. Of these papers, 31 did not attempt to adjust their analysis to account for the clustering effects, resulting in inefficient or incorrect inferences.

The scientific and economic costs to industry of incorrect or inefficient inferences are a major consideration of this investigation. For example, the cost to the agricultural industry of disgarding a herbicide which has been erroneously shown to be ineffective may be profound in both economic and environmental terms, and may guide the scientific investigation in a potentially less beneficial direction by concentrating resources into developing less effective herbicides.

We shall conclude this chapter with an overview of GLMs and error models. We do this to provide the framework and notation for the extensions to incorporate random effects in GLMs, which is the theme of the thesis.

The analyses of the herbicide data by several techniques are presented in Chapter 2. The disparity of the results amongst the techniques leads to their comparison in Chapters 3 and 4.

In Chapter 3 we shall introduce Poisson models for independent count data and extensions of these model to accomodate correlation. The techniques, among others,

that were used to analyse the Poisson random effects model for the data in Chapter 2 will be discussed and compared algebraically.

Chapter 4 consists of a small simulation study. We use this study to compare the available techniques numerically and to make recommendations under different modelling scenarios.

A variogram of transformed residuals is presented in Chapter 5. We use the variogram as a tool to help guide the selection of appropriate error models.

1.1 Generalized linear models

The evolution of Statistics has been predominately based on the normal distribution. Many of the major developments in interpreting complex data have, consequently, also been based on the assumption of normality. Due to this background, methods for modelling complex normally distributed data, such as those arising from longitudinal studies, are well established in literature. However, this is not always the case when experimental data are distributed in some other form, such as Poisson.

Some data, because of the nature of their measurement scale, can not be accurately modelled by accepted normal distribution techniques. For example, treating counts as a continuous measurement on $[-\infty, \infty]$ can, depending on the distributional range of the counts, lead to predicted values that are negative, which is clearly invalid. Count data are often able to be represented by the Poisson distribution, in which the mean and the variance of the distribution are equal. One method of analysing

such data has been to apply transformations to the response and to then analyse the data on the transformed scale. Existing methodology, which is based on the normal distribution, is then appropriate.

Carroll and Ruppert (1988 (in revisiting Box and Cox, 1964) have shown that for Poisson distributed data, the Box-Cox family of transformations,

$$h(y) = \frac{y^{\frac{1}{2}} - 1}{\frac{1}{2}},$$

where $h()$ represents the transformed data, is capable of providing a constant variance of the mean. However, if such a transformation was successful, the results of the consequential analysis need not be conceptually simple. This is because the interpretations, while valid on the transformed scale, may be difficult to relate to the experimental results on their natural scale.

It is also relevant to consider the three main objectives of transforming data which are to create data on a transformed scale which have;

1. a simple additive mean structure,

2. a constant variance, and

3. can be considered to be normally distributed.

Carroll and Ruppert (1988, p161) state that normal errors and constant variance often cannot be achieved for a data set through a single transformation. This makes it necessary to simultaneously transform in order to provide normal errors, and weight, to ensure constant variance.

However, GLMs, which were introduced by Nelder and Wedderburn (1972), provided statisticians with a means of modelling such data without the need to use inelegant techniques such as variance stabilizing transformations on the data. If the random component of data can be represented by a density from the exponential family (defined below as (1.1)), we may use this alternative procedure which addresses the issues of additivity and error structure separately, while preserving the scale of measurement. Simple GLMs, however, are restricted to independent data.

1.1.1 Definition of generalized linear models

A general algebraic representation of the classical linear model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

In this model, $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)^T$ is the n dimension vector of observed data. The matrix $\mathbf{X}$ is often referred to as the design matrix, and has dimension $n \times p$. Then $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is a p dimension vector of parameters, which needs to be estimated. The product $\mathbf{X}\boldsymbol{\beta}$ represents the systematic effects, while $\boldsymbol{\epsilon}$ represents the random variation within the model.

Generalized linear models are defined by three basic components;

1. Components of $\mathbf{Y}$ are from a distribution of the exponential family with $E(\mathbf{Y}) = \boldsymbol{\mu}$ and $\text{var}(\mathbf{Y}) = a(\phi) \cdot \text{var}(\boldsymbol{\mu})$, where ϕ is a scale parameter and $a()$ is a function of ϕ. This is the random component.

2. $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ is the systematic component.

3. $\boldsymbol{\eta} = g(\boldsymbol{\mu})$ where $g(\cdot)$ is ca led the link function.

The general algebraic repres entation of the generalized linear model in this terminology is

$$g(\mathbf{Y}) = \boldsymbol{\eta} + \boldsymbol{\epsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

1.1.2 Likelihood functions

When using generalized linear models, we assume that the Y components have a distribution which is a member of the exponential family. Following McCullagh and Nelder (1989, p28), we have the definition of the exponential family as follows: If the density can be manipulated into the algebraic form of (1.1)

$$f_Y(y, \theta, \phi) = \exp\{[y\theta - b(\theta)]/a(\phi) + c(y; \phi)\}, \tag{1.1}$$

where ϕ is a dispersion parameter and θ is referred to as the canonical parameter, then the density is said to be one from the exponential family. The log-likelihood function is taken as $log[f_Y(y, \theta, \phi)]$ (also referred to as $l(y, \theta, \phi)$).

Maximum likelihood estimates of parameters from such a density may be written as (see McCullagh and Nelder, 1989, p29)

$$E(Y) = \mu = \frac{d}{d\theta}[b(\theta)]$$

$$\begin{aligned} \text{var } Y) &= a(\phi) \cdot \frac{d^2}{d^2\theta}[b(\theta)] \\ &= a(\phi) \cdot V(\mu) \end{aligned}$$

where $V(\mu)$ is the variance function of μ. Derivation of the likelihood estimates above is given in Appendix B.

A significant benefit to flow from generalized linear models is the working algorithm for fitting the models. The algorithm approximates maximum likelihood estimation with an iteratively reweighted least squares procedure (with accuracy $O(n^{-1})$), and thus can efficiently compute the estimates. This enhances the process of adding and deleting terms in a hierarchical model. A brief description of the fitting algorithm (Appendix A) is provided to show the connection between iteratively reweighted least squares and maximum likelihood methods.

While the GLM framework has the advantage of avoiding the complications of transformations, we cannot overlook correlation when it is present. In the following sections of this chapter we shall include models to account for correlation structure and foreshadow the complications random effects models can introduce.

1.1.3 Correlation in longitudinal and temporal data

In the simplest case of independent data, the estimates of the model's coefficients and their subsequent variances (with subscripts dropped for brevity) at the $(j+1)$th iteration are

$$\hat{\boldsymbol{\beta}}^{(j+1)} = \hat{\boldsymbol{\beta}}^{(j)} + \left(\mathbf{X}^T \left(\frac{d\boldsymbol{\mu}^{(j)}}{d\boldsymbol{\eta}^{(j)}}\right) \frac{1}{Var(\boldsymbol{\mu}^{(j)})} \left(\frac{d\boldsymbol{\mu}^{(j)}}{d\boldsymbol{\eta}^{(j)}}\right)^T \mathbf{X}\right)^{-1} \cdot \left(\mathbf{X}^T \left(\frac{d\boldsymbol{\mu}^{(j)}}{d\boldsymbol{\eta}^{(j)}}\right) \frac{1}{Var(\boldsymbol{\mu}^{(j)})} \left[\mathbf{Y} - \boldsymbol{\mu}^{(j)}\right]\right), \quad (1.2)$$

$$\text{var}\left(\hat{\boldsymbol{\beta}}\right) = \phi\left(\mathbf{X}^T\left(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\eta}}\right)\frac{1}{Var(\boldsymbol{\mu})}\left(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\eta}}\right)^T\mathbf{X}\right)^{-1}, \tag{1.3}$$

where $\text{cov}(y_i, y_k) = 0\ \forall i \neq k$, and so $\mathbf{V}$ (the variance matrix) is diagonal. It is often the case in practice that this independence assumption is not sustainable. However, certain aspects of GLMs, such as the transformation from the additive covariates to the non-linear mean response, and the relationship of the variance to the mean, make it desirable to retain the GLM framework for data analysis. Thus, we consider modelling such data with a link and variance function in a way which can account for the heterogeneity that will arise because some of our observations are inherently more alike than other observations (*i.e.* correlated).

1.2 Models for correlation structure

A model for longitudinal and spatial data must incorporate a component to represent the associations among repeated measurements and field plot position.

Let us consider a linear model in a longitudinal setting. We work first in the linear framework and later extend some of these concepts to generalized linear models. The model will take the form

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \ldots + \beta_p x_{itp} + \epsilon_{it}, \tag{1.4}$$

where p is the number of explanatory variables in the model, i is the subject and t represents sampling time.

If the data were independent ($\text{cov}(\epsilon_{it}, \epsilon_{ik}) = 0\ \forall t \neq k$) the residuals ϵ_{it} should be

distributed as

$$\epsilon_i \sim N(0, \sigma^2 \mathbf{I}), \tag{1.5}$$

where $\mathbf{I}$ is an identity matrix, and $\sigma^2\mathbf{I}$ represents the variance-covariance matrix of the errors, resulting in it being diagonal. However, in longitudinal studies, non-negligible covariances of the residuals (from the same subject at different times) are to be expected, and therefore the variance-covariance matrix will not be diagonal. In the non-independence case the residuals can be represented by

$$\epsilon_i \sim MVN(0, \mathbf{V}_i), \tag{1.6}$$

where $\mathbf{V}_i$ represents the variance-covariance matrix of measurements for subject i. As measurements between subjects are assumed to be independent the residuals can be represented in matrix form as

$$\epsilon \sim MVN(0, \mathbf{V}), \tag{1.7}$$

where $\mathbf{V}$ is block diagonal, with $\mathbf{V}_i$ being its non-zero blocks for each subject.

It is desirable to represent the variance-covariance matrix ($\mathbf{V}$) as a function of a small set of parameters since;

1. there are generally too few degrees of freedom to identify individual elements of $\mathbf{V}$, and

2. the "smooth" parameterised version of $\mathbf{V}$ can have general applicability and need not be specific to a particular data set.

In defining the parametric form of $\mathbf{V}$ we need to consider the sources of error which give rise to the dependence of data. To achieve an adequate model it is necessary to consider the possible sources of variation which could occur in any particular longitudinal or spatial study. Three commonly attributed sources of this variation are;

1. measurement error,

2. serial correlation,

3. and random effects.

It is possible for all three sources, which we shall now describe, to contribute to the variation in a study.

Measurement error occurs when the measurement taken on a response or response variable in the study is not exact.

Serial correlation occurs if the response profile of an observed unit in the study is determined by a time-varying stochastic model for that particular unit. With serial correlation, a unit measured at time t is directly influenced by the prior history of that unit. These previous measurements can be used as predictor variables, and modelled using Auto-regressive Moving Average (ARMA) techniques.

Dependence due to repeated measures on a subject (or plots within the same block) can be accounted for in the *random effects* model. For example, when the profile of an individual in a longitudinal study can be represented by a parametric model and summarized by the coefficients of that model, we estimate the response

for the population by assuming the coefficients for all individuals are samples from a distribution. Correlation can be accounted for by random effects because all the repeated measures for an individual share the same subject effects.

If all three sources of variation were considered to be pertinent to achieving a good working model, the residuals given in (1.5) could be represented in the additive form

$$\epsilon_{it} = \mathbf{z}_{it}^T \mathbf{U}_i + W_i(\tau_{it}) + M_{it}, \tag{1.8}$$

where $\mathbf{z}_{it}$ represents a design vector for the random effects, and $\mathbf{U}_i$ is a vector of coefficients for the random effects. Serial correlation is represented by a stationary Gaussian process, W_i, τ_{it} represents sampling time, and M_{it} denotes the measurement error component.

To distinguish between serial correlation and correlation due to random effects, consider the simple case of two observations made on the same subject at sampling times t_1 and t_2. We express the observations (x_{t_1}, x_{t_2}) as functions of the deterministic effects $(\beta_{t_1}, \beta_{t_2})$, the subject effects (u) and the residuals $(\epsilon_{t_1}, \epsilon_{t_2})$. That is,

$$x_{t_1} = \beta_{t_1} + u + \epsilon_{t_1}, \quad x_{t_2} = \beta_{t_2} + u + \epsilon_{t_2}.$$

The correlation of the errors, $\{(u + \epsilon_{t_1}), (u + \epsilon_{t_2})\}$, may be due to;

1. the random subject effect u,

2. the serial correlation between ϵ_{t_1} and ϵ_{t_2}, or

3. a combination of both 1 and 2.

If the correlation is modelled as being due to only random effects, then we assume that $u \sim N(0, \sigma_u^2)$, ϵ_{t_1}, and ϵ_{t_2} are distributed as $N(0, \sigma_\epsilon^2)$ and that all other covariances are zero. Then;

- $\text{var}(x_{t_1}) = \text{var}(x_{t_2}) = \sigma_u^2 + \sigma_\epsilon^2$,
- $\text{cov}(x_{t_1}, x_{t_2}) = \sigma_u^2$, and
- $\text{corr}(x_{t_1}, x_{t_2}) = \sigma_u^2/(\sigma_u^2 + \sigma_\epsilon^2)$.

When we model the correlation as being due to only serial correlation, we assume that σ_u^2 is negligible and;

- $\text{var}(x_{t_1}) = \text{var}(x_{t_2}) = \sigma_\epsilon^2$,
- $\text{cov}(x_{t_1}, x_{t_2}) = \alpha \cdot \sigma_\epsilon^2$, so that
- $\text{corr}(x_{t_1}, x_{t_2}) = \alpha$.

Modelling the correlation as being due to both serial correlation and random effects, and assuming independence between the two error sources gives;

- $\text{var}(x_{t_1}) = \text{var}(x_{t_2}) = \sigma_u^2 + \sigma_\epsilon^2$,
- $\text{cov}(x_{t_1}, x_{t_2}) = \sigma_u^2 + \alpha \cdot \sigma_\epsilon^2$, then
- $\text{corr}(x_{t_1}, x_{t_2}) = (\sigma_u^2 + \alpha \cdot \sigma_\epsilon^2)/(\sigma_u^2 + \sigma_\epsilon^2)$.

Diggle, Liang and Zeger (1994) expound the advantages of using a random effects model for the covariance by explaining that they can be non-stationary in structure.

For example, let us view the residuals as

$$\epsilon_{it} = \mathbf{z}_{it}^T \mathbf{U}_i, \tag{1.9}$$

where $\mathbf{z}_{it} = (1, \tau_{it})$, so there is a linear time trend and random intercepts and slopes for each individual in the study. If we assume that $\mathbf{U}_i \sim MVN(0, \mathbf{D})$, where $\mathbf{D}$ is a diagonal variance-covariance matrix,

$$\mathbf{D} = \begin{pmatrix} \nu_1^2 & 0 \\ 0 & \nu_2^2 \end{pmatrix}, \tag{1.10}$$

and $\mathbf{U}_i$ represent the random subject effects, then

$$\text{var}(\epsilon_{it}) = \nu_1^2 + \tau_{it}^2 \nu_2^2, \quad \text{and} \tag{1.11}$$

$$\text{cov}(\epsilon_{ij}, \epsilon_{it}) = \nu_1^2 + \tau_{ij}\tau_{it}\nu_2^2. \tag{1.12}$$

Such a model is often appropriate for growth data, which generally consist of a series of response profiles which will characteristically increase in variance through time or with an increase in the mean.

This characteristic can be seen in the simulated data set given in Figure 1.1. In this data set we have simulated growth profiles for four (4) subject units over eight (8) repeated measures. For each of the units, coefficients were simulated from a $MVN(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ distribution where $\boldsymbol{\beta} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$. The simulations were made by obtaining the

Choleski decomposition of $\mathbf{\Sigma}$ (*i.e.* $\mathbf{L}\mathbf{L}^T = \mathbf{\Sigma}$), following which the random coefficients were calculated as

$$\boldsymbol{\beta}_i^* = \boldsymbol{\beta} + \mathbf{L}^T\mathbf{U}_i, \tag{1.13}$$

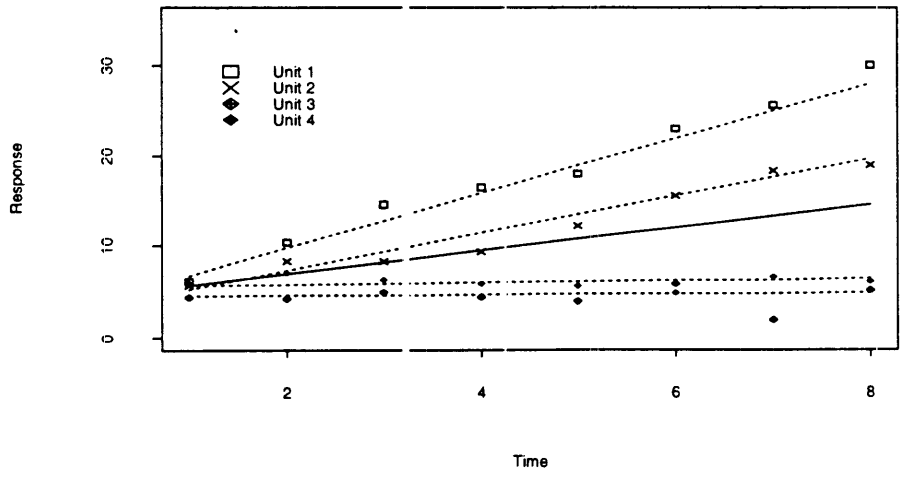
where $\mathbf{U}_i \sim N(0, \mathbf{I}_2)$ and i indicates subject. The simulated response of each subject (i) at time t is then given by

$$y_{it} = \mathbf{x}_{it}^T\boldsymbol{\beta}_i^* + \epsilon_{it}, \tag{1.14}$$

where $\epsilon_{it} \sim N(0,1)$ and $\mathbf{x}_{it}$ is the design vector for subject i at time t. Morgan (1984, p85) outlines this method of simulating multivariate normal data.

From Figure 1.1, we note that the variance of the data increases with the overall mean (represented by the solid line), but the variance of each individual from their own growth profile is constant. The heterogeneity of the data is induced by individuals having random time slopes, and it is therefore reasonable to model such data in a random effects framework.

Figure 1.1: Growth profiles from simulated data.



Our primary objective is to interpret the coefficients when the correlation has been adequately represented. Diggle *et al.* (1994, p137) show an example in the linear model case where the coefficients under the different covariance models have similar interpretations regardless of whether the correlation is modelled as serial correlation or random effects. This may not apply in the nonlinear setting, so choice of correlation structure needs to be considered. In choosing a model we should be guided by the data. In growth curve situations, the random effects model is likely to be appropriate (Diggle *et al.*, 1994, p89).

In recent literature, the random effects model for longitudinal data has taken a dominant role (*e.g.* Stiratelli, Laird, and Ware (1984), Laird and Ware (1982), and Carlin (Editors Gilks, Richardson and Spiegelhalter, 1996, pp 303-319)). Diggle *et al.* (1994, p88) have speculated that this has occurred because although serial correlation is an intuitive feature of longitudinal data in many situations, its contribution to the covariance structure is of a lesser magnitude than that of the random effects and the measurement error. As a result, the inclusion of serial correlation in the covariance model becomes unnecessary and can decrease the model's efficiency.

1.3 Random effects models

In the previous section we established that a common approach to the problem of longitudinal data is to assume that the heterogeneity present is due to some unobserved variables ($\mathbf{U}$), which are collectively known as random effects. We are now

going to introduce some of the difficulties that can arise in random effects modelling.

Representing the model as

$$\Pr(\mathbf{Y} = \mathbf{y}|\mathbf{U} = \mathbf{u}) = g(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}), \tag{1.15}$$

the **marginal distribution** (L) of the observed data will be

$$L = \prod_i \Pr(Y_i = y_i) = \prod_i \int_u g(X\beta + u)dD(u), \tag{1.16}$$

where $D(u)$ is the distribution of the random effects (U). The mixing of the random effect induces the correlation and heterogeneity into our data.

Integrals in this form are very often difficult to evaluate because they are not in closed form, and estimation based upon expressions for the likelihood such as (1.16) would be complex. Also, determining which model provides the most adequate fit to the data through adding and deleting terms can become tedious.

This facet of the problem has been noted by many statisticians in recent years. Stiratelli *et al.* (1984) noted one of the problems with the maximum likelihood approach was that integrals, such as those of (1.16), do not have general closed form expressions, and as such, standard analytical techniques cannot be used. McCullagh and Nelder (1989, p437) use the method of quasi-likelihood estimation in preference to explicit maximum likelihood Their justification of this technique is that it is simpler than maximum likelihood, and assumes only the first two moments. Breslow and Clayton (1993) noted that, for outcomes of counts or proportions, the full maximum likelihood analysis has been used in relatively simple problems of binomial or Poisson

mixtures and in situations where the data was largely independent. However, they state that "to date it has proved intractable for more complicated problems involving irreducibly high-dimensional integrals".

The unobserved random variable which we shall use to incorporate the heterogeneity of the data could be interpreted in several ways. It is commonly convenient to assume that it operates on the same scale as the fixed effects and enters into the model as part of the linear predictor (η). Some alternative ways of representing such a model are;

$$\eta = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{U}, \tag{1.17}$$

$$\eta = (\mathbf{X} + \epsilon)\boldsymbol{\beta}, \quad \text{and} \tag{1.18}$$

$$\eta = \mathbf{X}(\boldsymbol{\beta} + \boldsymbol{\delta}). \tag{1.19}$$

Equation (1.17) is the random effects model, (1.18) is the "errors in X variable" model (measurement error) and (1.19) is the random coefficients model. It may be our personal perspective which leads to how we couch the model, but often the structure of the data, the experimental design, and the sampling scheme will suggest which of these forms is appropriate. In this thesis we shall concentrate on models of the form (1.17) because it is relevant to the herbicide field trial which we consider in Chapter 2.

Future developments in this area may see it unnecessary to have the fixed and random effects on the same scale. Nelder and Lee (1994) propose that the random effect can have a distribution conjugate to that of the fixed effect so that the integrals

in (1.16) are more tractable. Van de Ven and Weber (1995) used the Gamma distribution to account for random effects in Poisson data. In that case, the random effects were nuisance parameters and the objective was to reliably estimate fixed effects and their variances, without concern for the relative contribution of the random effects. Situations such as this, where the extra variation is absorbed into measurement error, are categorized as "overdispers on models".

In this thesis, we shall pursue a detailed examination of several alternative methods for approximating generalized linear models with random effects so that intractable likelihoods such as (1 16) are overcome. These methods fall into the following categories;

1. quasi-likelihood techniques,

2. overdispersion models,

3. generalized estimating equations, and

4. Markov Chain Monte Carlo techniques.

We examine these models in detail because our understanding of them and knowledge of their strengths and shortcomings is important in the model selection process as a whole.

Chapter 2

Comparative analysis of herbicide experiment

In this chapter we shall describe a field experiment in which the effect of 11 herbicides, a cultivation and a control treatment on nutgrass tuber populations in cotton crops was assessed. We illustrate the need for a random effects model when analysing the experiment, and compare the results from different modelling techniques.

2.1 The experimental design

The experimental design consisted of 5 blocks of 13 treatments, which are listed in Table 2.1. These treatments were arranged in 10 columns and 9 rows, with two treatments, control and cultivation, being repeated within each block. Cultivation (treatment 2) was repeated once, and the control (treatment 1) was repeated twice.

The response to the treatments was measured as the number of nutgrass tubers present in a soil core. Measurements were taken yearly over a three year period.

Table 2.1: Table of experimental treatments.

Treatment Number	Treatment Description
1	Control
2	Cultivation
3	MSMA
4	MSMA & cultivation
5	Defoliant glyphosate
6	MSMA & defoliant glyphosate
7	MSMA & cultivation & defoliant glyphosate
8	Glyphosate & defoliant glyphosate
9	Glyphosate & cultivation
10	Norflurazon & cultivation
11	Benfuserate & cultivation
12	Norflurazon & benfuserate & cultivation
13	Norflurazon & glyphosate & defoliant glyphosate

The allocation of treatments within the design can be seen in Figure 2.1. The treatment numbers correspond to those given in Table 2.1, and blanks indicate plots where treatments were missing or altered during the course of the experiment. These plots were not of interest in our case, so for analytical purposes they were ignored.

We note from Figure 2.1 that the design for this experiment is quite poor. For example, the cultivation treatment (2) is found 6 times on border plots, and in consecutive plots in row 9 and column 10. Some treatments, such as treatment 10, which is only represented in rows 1 and 3, are all found in the lower (or upper) rows of the design.

Figure 2.1: Experimental design of nutgrass trial.

										ROW
	13	9	2	2	2		3	1	5	9
7	4	8	5	1		9	1	3	8	8
3	1	3	2	5	4	5	6	9	7	7
9		11	4	3		1	11	2	13	6
11	6	7	1	8	1	13	1	11	2	5
1	12	6	12	13	12	2	10		1	4
2	10	1		9	10	4	12	4	6	3
2	1	13	1	7	1	2	8	1	12	2
5	8		10	6	11	7			10	1
REP 5		REP 4		REP 3		REP 2		REP 1		

Discussions with the researcher involved in this project indicate that this design was meant to be of a row-column nature to account for spatial effects, and hence the positioning of these treatments is not due to chance.

The inadequacies of this design have led to a partial confounding of treatment effects with row effects, which can be seen in Table 2.2. This poor spatial design will limit the success of any model which attempts to account for spatial trends, and as such this limitation should be taken into consideration when determining the success or otherwise of the models we fit to this data.

Table 2.2: Treatment counts by design row.

	Row Number								
	1	2	3	4	5	6	7	8	9
Treatment 1	0	4	1	2	3	1	1	2	1
2	0	2	1	1	1	1	1	0	3
3	0	0	0	0	0	1	2	1	1
4	0	0	2	0	0	1	1	1	0
5	0	0	0	0	0	0	2	1	1
6	1	0	1	1	1	0	1	0	0
7	1	1	0	0	1	0	1	1	0
8	1	1	0	0	1	0	0	2	0
9	0	0	1	0	0	1	1	1	1
10	2	0	2	1	0	0	0	0	0
11	1	0	0	0	2	2	0	0	0
12	0	1	1	3	0	0	0	0	0
13	0	1	0	1	1	1	0	0	1

2.2 Preliminary analysis

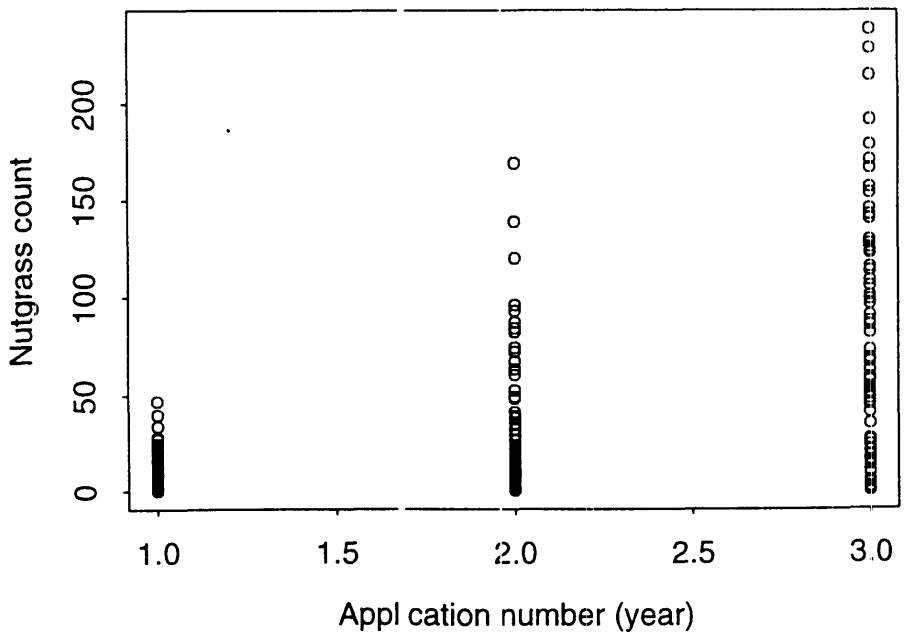
Figure 2.2 illustrates the raw data collected over the 3 year period of the trial. As the data is potentially Poisson (the simplest distributional assumption) it is expected that the variance will be roughly equal to the mean. However, this graph provides some evidence that the data may be overdispersed, as the variance appears to be becoming greater than the mean as the trial progresses. Following from Section 1.3, we would expect that this could occur in a growth profile situation if random effects, such as plot*time effects were present. Therefore, one plausible model in this case is Poisson random effects.

The initial step taken in this analysis was to model the data assuming time independence and fixed replicate effects, although strictly replicates could be viewed as random effects because they represent a random selection of plots. The model used was a GLM with log link and Poisson errors. The significant terms are given below in (2.1).

$$\begin{aligned} \log\,[\mathrm{E(counts)}] \;=\; & \text{treatment*time} + \text{replicate} + \text{replicate:treatment} \\ & + \log\,(\text{soil core weight}), \end{aligned} \tag{2.1}$$

where : indicates an interation term and * indicates that each term and their interaction are fitted. The offset, log (soil core weight), in (2.1) is necessary because cores differ in weight and hence volume. Consequently the cores have different potentials to yield nutgrass tubers, regardless of other effects.

Figure 2.2: Raw data from nutgrass experiment.



The spatial independence assumption of the GLM was examined using contour plots of the Pearson residuals for each year. In the case of the Poisson model the Pearson residuals can be considered to be standardized residuals (Dobson, 1990, p133), and with large count values, such as we have in the herbicide experiment, these can be treated as standard normal variables (Aitkin, Anderson, Francis and Hinde, 1989, p219).

Figure 2.3: Contour plot of residuals from model (2.1), year 1.

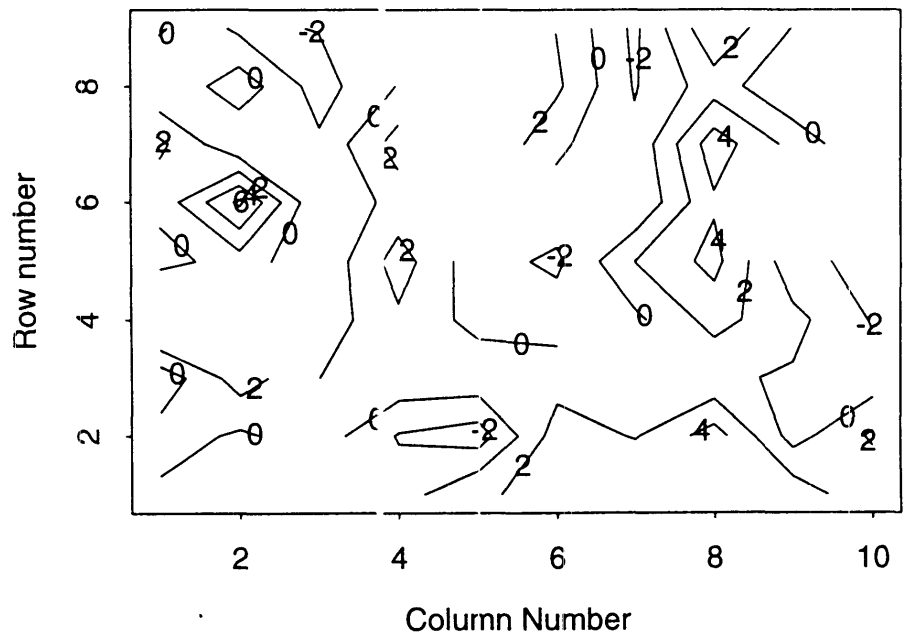


Figure 2.3 is the contour plot of Pearson residuals over the trial site for the first year. From this graph we noted areas in the trial where there are clusters of positive residuals. For example, the right hand side of the trial (columns 7-10) is a region where positive residuals are predominant. Negative residuals are most often located at the edges of the trial. Peaks and troughs, which indicate spatial dependence are evident over the trial site. A normal quantile plot of the Pearson residuals from

the first year indicated that they were not normally distributed and therefore the assumption that the data could be modelled by (2.1) was violated. The contour plots for the following 2 years indicated similar model deficiencies.

The correlation matrix of the Pearson residuals between years from (2.1) was

$$\boldsymbol{\rho} = \begin{pmatrix} 1.00 & \cdot & \cdot \\ 0.16 & 1.00 & \cdot \\ 0.17 & 0.44 & 1.00 \end{pmatrix}. \tag{2.2}$$

This correlation matrix is consistent with the concept of random intercepts and slopes for individual trial plots in the experiment. We will examine this concept in section 2.3.

2.3 Diagnosing random effects from correlation structure

To illustrate the concept of between time correlations and their relationship to random effects, let us consider an example in which the errors, consisting of both random effects (subject intercept and slope only) and measurement error, are given by

$$\epsilon_{it} = \mathbf{z}_{it}^T \mathbf{U}_i + M_{it}. \tag{2.3}$$

In this example $\mathbf{U}_i$ are the random effects for subject i, and they are distributed as $N(0, \mathbf{D})$, where

$$\mathbf{D} = \begin{pmatrix} \nu_1^2 & 0 \\ 0 & \nu_2^2 \end{pmatrix}, \tag{2.4}$$

where ν_1^2 is the random subject intercept variance and ν_2^2 is the corresponding random slope variance. The design vector for the random effects, $\mathbf{z}_{it}$, is $(1, \tau_{it})$. The measurement error is represented by M_{it}. The measurement error is independent of the random effects, and it is distributed as $N(0, \sigma_M^2)$. From this,

$$\text{var}(\epsilon_{it}) = \nu_1^2 + \tau_{it}^2\nu_2^2 + \sigma_M^2, \tag{2.5}$$

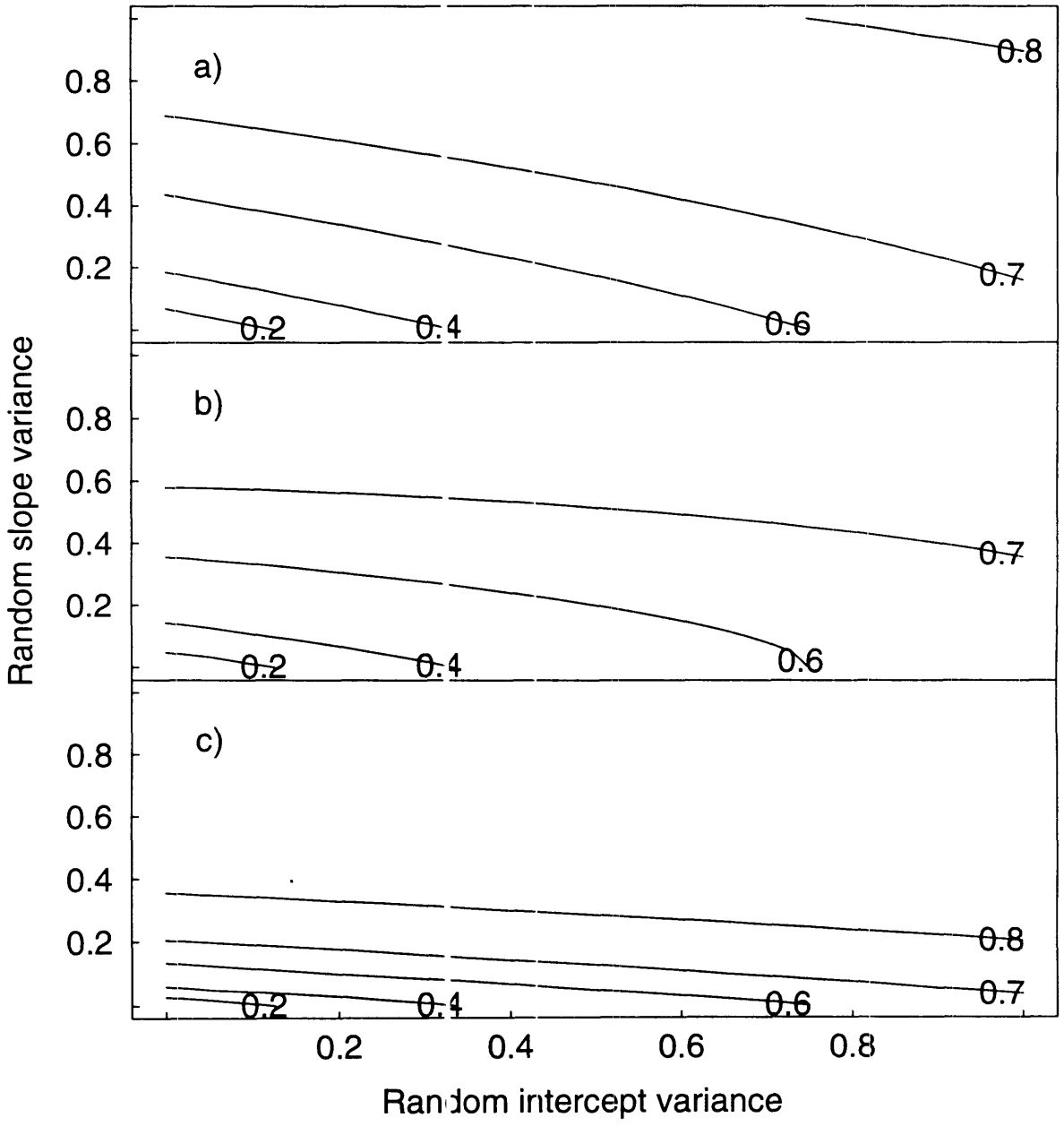
$$\text{cov}(\epsilon_{it}, \epsilon_{ij}) = \nu_1^2 + \tau_{it}\tau_{ij}\nu_2^2, \quad \text{and} \tag{2.6}$$

$$\text{corr}(\epsilon_{it}, \epsilon_{ij}) = \frac{\nu_1^2 + \tau_{it}\tau_{ij}\nu_2^2}{\sqrt{(\nu_1^2 + \tau_{it}^2\nu_2^2 + \sigma_M^2)(\nu_1^2 + \tau_{ij}^2\nu_2^2 + \sigma_M^2)}}. \tag{2.7}$$

Figure 2.4 is an illustration of the correlations, (2.7), expected from the residuals given by (2.3), over a range of values for the variance-covariance, D, of the random effects. Measurement error has been included in the calculations, with $\sigma_M^2 = 0.5$, and the linear time trend, τ_{it} has taken the value of the sampling time, t.

The differences over the range of values in Figure 2.4 a) and b), which are the correlations between sampling times (1,2) and (1,3) are very slight. This feature was seen in (2.2) where the correlations are 0.16 and 0.17 respectively. However, the correlations seen in c) are usually much larger than those of both a) and b). This feature is also evident in (2.2) where the correlation of the Pearson residuals between year 2 and 3 is 0.44, substantially larger than the other between year correlations. Therefore, the same error variances in intercept and slope for a subject yield stronger correlations with increasing time.

Figure 2.4: Contour plot of theoretical correlations amongst residuals. a) Correlation between time 1 & 2, b) Correlation between time 1 & 3 and c) Correlation between time 2 & 3.



Our preliminary analysis has suggested that the spatial (by graphical illustration Figure 2.3) and temporal correlation is of a magnitude that cannot be ignored. We shall now examine some models that include components to account for correlation in a GLM framework.

The three techniques used are available in the Genstat 5.3.2 procedure library (Genstat 5, 1993). They are marginal quasi-likelihood (MQL, Breslow and Clayton, 1993), penalized quasi-like ihood (PQL, Schall,1991), and generalized estimating equations (GEEs, Liang and Zeger, 1986). The PQL and MQL techniques are used when the model is formulated as a generalized linear mixed model (GLMM), and hence the correlation in the data is assumed to be due to the presence of random effects. The GEEs used are an extension of GLMs which attempt to adequately account for correlation without specific reference in the model to the cause of this correlation. The discussion of these methods is deferred to Section 3.2, however, we use these techniques on our herbicide data to illustrate that we have a disparity in the results, and therefore a need to examine each of them, as well as other available techniques, in detail.

2.4 GLMM analysis

The model for both PQL and MQL was formulated as

$$\mathrm{E}(\mathbf{count} \mid \mathbf{U}) = \mu = \exp\left[\mathbf{X}\boldsymbol{\beta} + \mathbf{ZU} + \log(\mathbf{soil\ weight})\right], \tag{2.8}$$

where $\boldsymbol{\beta}$ is a vector of fixed effects consisting of treatment, time and their interaction. The vector $\mathbf{U}$ contains the random effects of row, column, row by column (plot), and the interactions of these with time. That is, for the ith plot at sampling time t

$$\log(\mu_{it}) = \text{treatment}_i * \text{time}_{it} + (\text{row}_i * \text{column}_i)/\text{time}_{it} + \log(\text{soil weight})_{it} \quad (2.9)$$

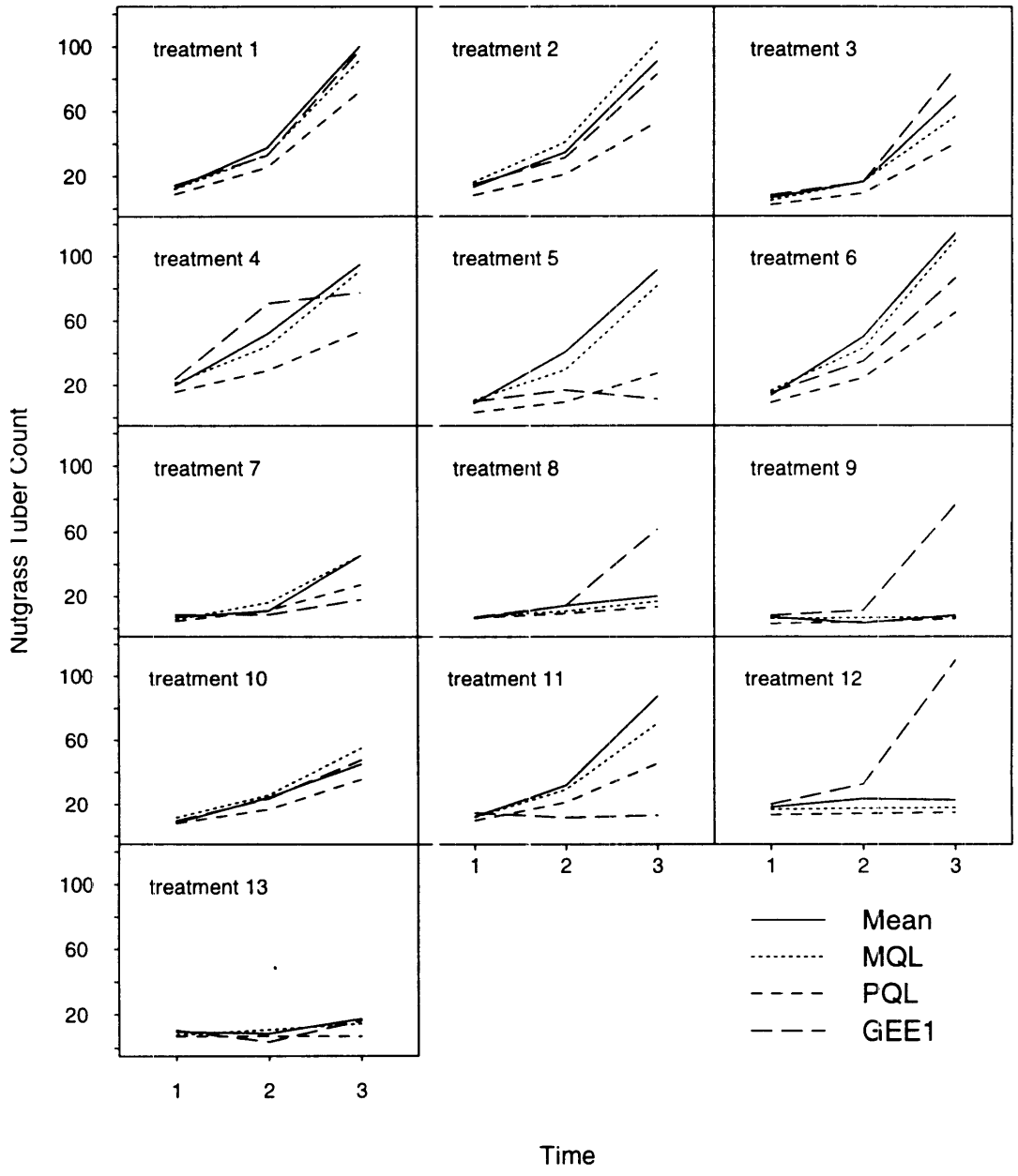
For these two techniques, the population mean is

$$\mathrm{E}(\mathbf{count}) = \exp\left[\mathbf{X}\boldsymbol{\beta} + \log(\overline{\mathbf{soil\ weight}})\right], \quad (2.10)$$

where $\overline{\text{soil weight}}$ represents the mean of the soil weights. These population means are compared with the sample means in Figure 2.5.

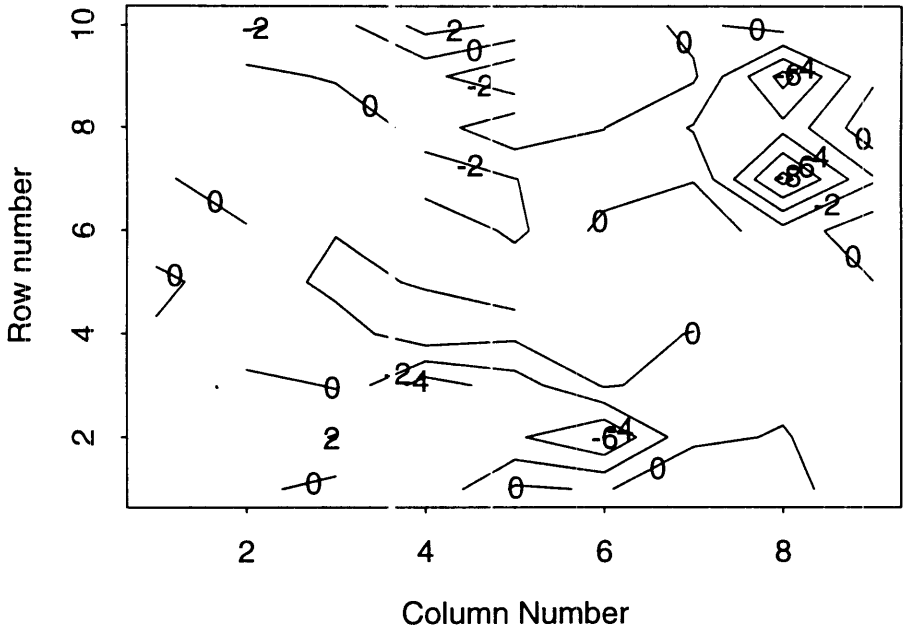
We note from Figure 2.5 that the estimates from the MQL technique tend to be close to the observed mean, estimating above and below this average with roughly equal frequency; the MQL estimate is slightly above the observed mean for treatments 2 and 10, slightly below the observed mean for treatments 1, 3, 4, 5, 6,and 11 and about equal to the observed mean for treatments 7, 8, 9 and 13. This would seemingly indicate that adjustments are being made for random plot effects as some estimated treatment means will be increased compared to the observed mean by the MQL model to account for "poor" field positioning and vice versa.

Figure 2.5: Comparison of estimates for nutgrass model.



The contour plot of the MQL Pearson residuals for year 1 (Figure 2.6) does not contain as many of the characteristic peaks and troughs which would indicate spatial correlations in the modelled residuals as was seen in Figure 2.3. The positive and negative residuals are not following as many evident patterns as they were in the independent model, indicating that the removal of the random plot effects has been moderately successful, when taking into consideration the poor spatial layout in the design. One area still of concern is between columns 7 - 9 and rows 7 - 10 where a large trough occurs, indicating that further modelling of the spatial structure may be warranted. However, we will not be pursuing this further at this stage of the analysis.

Figure 2.6: Contour plot of residuals from MQL model, year 1.

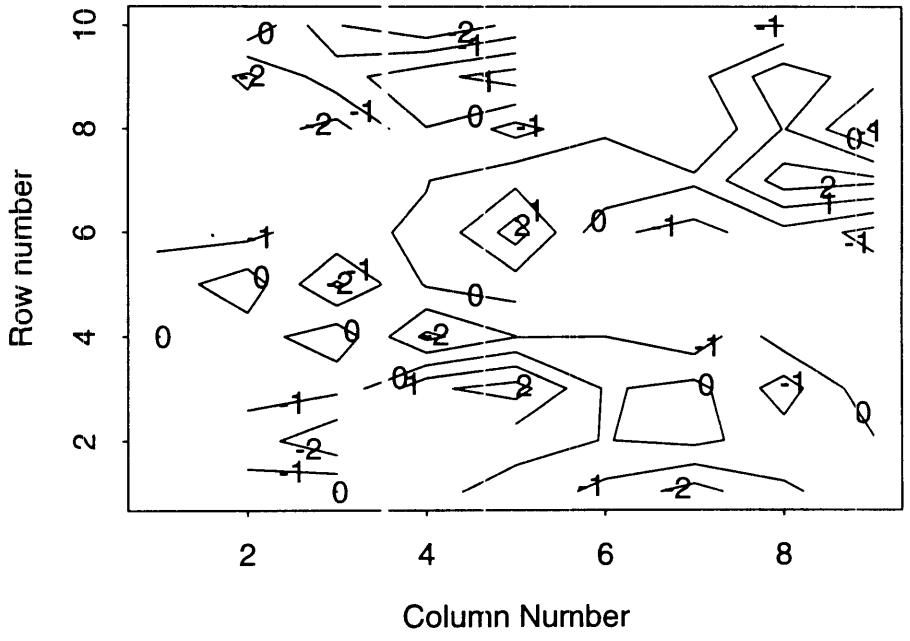


The PQL estimates shown in Figure 2.5 are consistently lower than the observed mean and the population mean estimated using MQL. This possible underestimation of the population mean by PQL techniques is unexpected when adjustments are

being made for random plot effects, as we would expect some plots to be in poor positions, some in better positions and some to be about average. If this were the case then estimates would be adjusted up and down according to field position. This is of concern in the herbicide experiment because underestimation of the nutgrass tuber population will falsely indicate that the herbicides being tested perform more effectively than they actually do in practice.

The contour plot of the PQL Pearson residuals for the first year of the experiment is given in Figure 2.7. This plot indicates that the spatial correlation has not been adequately modelled, as the peaks and troughs are still clearly evident in a number of areas over the field. Results from the following two years, which are not presented, highlighted similar deficiencies in the model.

Figure 2.7: Contour plot of residuals from PQL model, year 1.



To determine if PQL is underestimating the population mean, and when each of

these two techniques (PQL and MQL) is appropriate we shall examine and compare them algebraically in Chapter 3. In Chapter 4 we shall examine their performance in a small simulation study.

2.5 GEE analysis

The model used in the GEE method was

$$\mathrm{E}(\mathbf{count}) = \boldsymbol{\mu} = \exp\left[\mathbf{X}\boldsymbol{\beta} + \log(\mathbf{soil\ weight})\right], \tag{2.11}$$

where the fixed effects vector $\boldsymbol{\beta}$ is equivalent to that of (2.8). That is

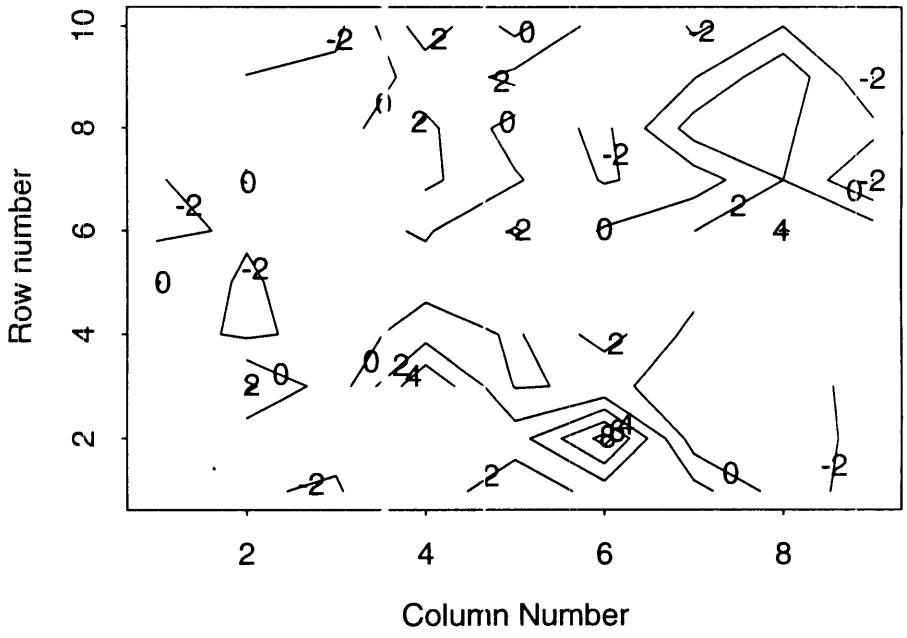
$$\log(\mu_{it}) = \quad \text{treatment}_i * \text{time}_{it} + \log(\text{soil weight})_{it}. \tag{2.12}$$

An unconstrained correlation structure was specified for the between years correlation on plots. This was considered appropriate as with only three repeated measurements, and an expected correlation structure (2.2), the standard correlation structures available in the GEE Genstat procedure (exchangeable, autoregressive or independence) would not be adequate.

The GEE technique appears to be giving reasonable approximations which closely map the MQL estimates and observed means for treatments 1, 2, 3, 4, 6, 10 and 13 (see Figure 2.5). However, for treatments 5, 8, 9, 11 and 12 the technique seems to be yielding spurious results, particularly between years 2 and 3. For example, with treatment 12, the estimate from the GEE method in year three is more than five times greater than the observed mean, and it seems unlikely that this could be a true correction for random plot effects.

The contour plot of the GEE Pearson residuals for the first year (results in following years being similar) are given in Figure 2.8. These contours indicate that spatial correlation is an issue with this model due to the peaks and troughs seen in the field layout of the model residuals. In (2.12), no attempt has been made to account for row and column effects, as only between time correlations on a subject have been considered, hence possible spatial correlations have been ignored. Attempts which were made to account for spatial correlation by treating rows and columns as fixed and then averaging over them were not successful.

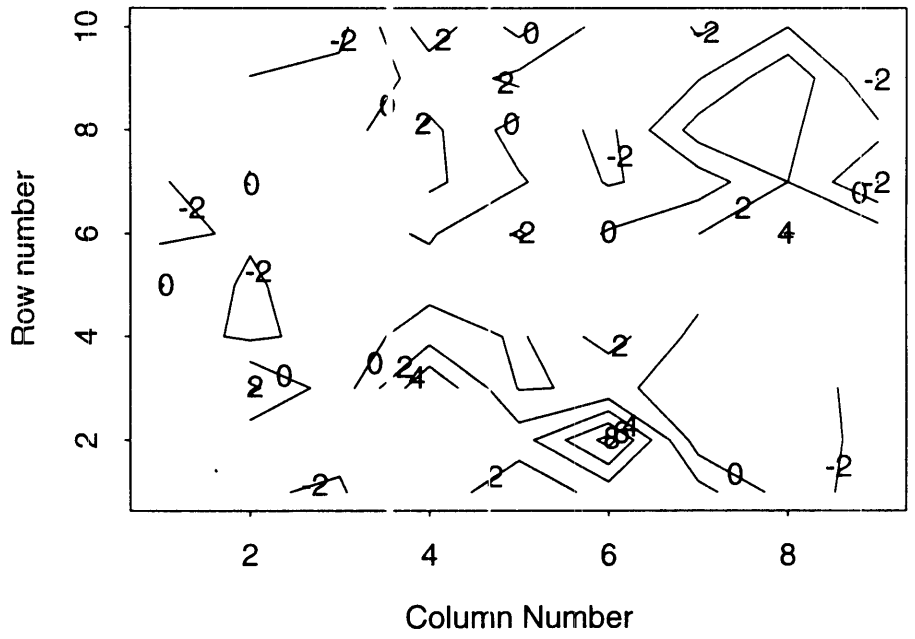
Figure 2.8: Contour plot of residuals from GEE model, year 1.



The model could further be refined to accommodate spatial correlation by using a specially defined correlation type. This capability is provided within the GEE Genstat procedure, and is known as GEECORREL. This, however, would involve a higher level of programming than may be warranted in the analysis of field trials

The contour plot of the GEE Pearson residuals for the first year (results in following years being similar) are given in Figure 2.8. These contours indicate that spatial correlation is an issue with this model due to the peaks and troughs seen in the field layout of the model residuals. In (2.12), no attempt has been made to account for row and column effects, as only between time correlations on a subject have been considered, hence possible spatial correlations have been ignored. Attempts which were made to account for spatial correlation by treating rows and columns as fixed and then averaging over them were not successful.

Figure 2.8: Contour plot of residuals from GEE model, year 1.



The model could further be refined to accommodate spatial correlation by using a specially defined correlation type. This capability is provided within the GEE Genstat procedure, and is known as GEECORREL. This, however, would involve a higher level of programming than may be warranted in the analysis of field trials

when other techniques can provide satisfactory results for data that is both spatially and temporally correlated.

Chapter 3

Applications to count data

3.1 Introduction

Longitudinal studies are often conducted to observe the effect of experimental treatments over time. Repeated measures of counts, which are not Normally distributed, present us with challenges as to how they may be modelled. In the univariate setting, the Poisson model and its associated negative binomial (McCullagh and Nelder, 1989, p198) models are appropriate. In this chapter we consider modelling the repeated measures of counts where the marginal distributions are Poisson and the counts are correlated over time due to random effects. We shall use the term "correlated Poisson" as a name for these types of data.

The probability function for the Poisson distribution is of the form

$$\Pr(Y = y) = \frac{\mu^y \cdot \exp(-\mu)}{y!}. \tag{3.1}$$

If the counts over time were independent, the likelihood would take the form (3.2),

where y_{it} denotes the measurement of subject i at time t, $(i = 1, 2, \ldots, m)$, and $(t = 1, 2, \ldots, n_i)$.

$$\begin{aligned} L(\mathbf{y};\theta) &= \prod_{i=1}^{m}\prod_{t=1}^{n_i} c \cdot \Pr(Y_{it} = y_{it}) \\ &= \prod_{i=1}^{m}\prod_{t=1}^{n_i} \mu_{it}^{y_{it}} \cdot \exp(-\mu_{it}) \qquad (\text{given} \quad c = y_{it}!) \\ &= \prod_{i=1}^{m}\prod_{t=1}^{n_i} \exp(y_{it}\log\mu_{it} - \mu_{it}) \\ &= \exp[\sum_{i=1}^{m}\sum_{t=1}^{n_i} y_{it}\theta_{it} - \sum_{i=1}^{m}\sum_{t=1}^{n_i}\exp(\theta_{it})] \qquad (\text{given} \quad \theta_{it} = \log\mu_{it}). \quad (3.2) \end{aligned}$$

The likelihood given in (3.2) is of the exponential family form given by (1.1) with a dispersion parameter of 1 and the canonical link (logarithm). From (3.2) we can see that the log likelihood has the form

$$l(\mathbf{y};\theta) = \sum_{i=1}^{m}\sum_{t=1}^{n_i} y_{it}\theta_{it} - \sum_{i=1}^{m}\sum_{t=1}^{n_i}\exp(\theta_{it}). \tag{3.3}$$

Standard results (see Appendix A) give the maximum likelihood estimates of the parameters $\hat{\boldsymbol{\beta}}$ in the GLM for the independent Poisson model, given some initial estimates $\boldsymbol{\beta}_0$, and using the link function are

$$\log\mu_{it} = \eta_{it} = \mathbf{x}_{it}^T\boldsymbol{\beta} + \epsilon_{it}, \tag{3.4}$$

and the estimating equations for this are given by (1.3).

As discussed in Section 1.2, in the case of longitudinal studies, it is naive to ignore the possibility of correlation in the data. We shall now introduce some methods of accounting for correlation in Poisson data. The main methods we shall examine relate to random effects models and we shall also consider overdispersion models and generalized estimating equation techniques.

Prior to introducing GLMMs we shall sketch the results for linear mixed models (LMMs). This is done because the main thrust of several of the approximate methods for GLMMs is to linearize the data and then proceed as in the linear model case. Thus we shall be able to compare the results for linear models with their non-linear counterparts.

The LMM can be represented by

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \epsilon, \tag{3.5}$$

where the link is the identity, $\mathbf{U} \sim N(0, \mathbf{D})$ is the distribution of the random effects, $\mathbf{Z}$ is the corresponding design matrix, and $\boldsymbol{\epsilon}$ are the measurement error terms, which are distributed as $N(0, \sigma_\epsilon^2 \mathbf{I})$.

The variance of this LMM can be shown to be

$$\text{var}(\mathbf{Y}) = \mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}^T + \sigma_\epsilon^2 \mathbf{I}_n, \tag{3.6}$$

assuming that random effects and measurement error components are independent.

With the LMM, restricted maximum likelihood (REML) can be used to estimate the variance components ($\mathbf{D}$), then the fixed effects ($\boldsymbol{\beta}$) are estimated via generalized least squares using $\hat{\mathbf{D}}$ to estimate the variance-covariance matrix $\mathbf{V}$:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}. \tag{3.7}$$

The random effects are estimated using best linear unbiased predictors (BLUP)

$$\tilde{\mathbf{U}} = \hat{\mathbf{D}} \mathbf{Z}^T \mathbf{V}^{-1} \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right). \tag{3.8}$$

We now introduce the random effects model for Poisson data in a general form to derive the likelihood and principles for estimation. Let the Poisson random effects model be represented as

$$\log[E(Y_{it}|\mathbf{U}_i)] = \log \mu_{it} = \eta_{it} = \mathbf{x}_{it}^T\boldsymbol{\beta} + \mathbf{z}_{it}^T\mathbf{U}_i. \tag{3.9}$$

In this model $\mathbf{U}_i \sim MVN(0, \mathbf{D})$ are independent realizations from a distribution with density function

$$f(\mathbf{U}_i; \mathbf{D}) = (2\pi)^{-q/2} \mid \mathbf{D} \mid^{-1/2} \exp\left(-\frac{1}{2}\mathbf{U}_i^T\mathbf{D}^{-1}\mathbf{U}_i\right), \tag{3.10}$$

where q is the dimension of the vector $\mathbf{U}_i$. Conditional upon $\mathbf{U}_i$ it is assumed that the responses $Y_{i1}, \ldots, Y_{it}$ are independent Poisson variables, and

$$f(y_{it} \mid \mathbf{U}_i; \mathbf{D}) = \exp\left[y_{it}\theta_{it} - \exp(\theta_{it})\right]/y_{it}!. \tag{3.11}$$

where $\theta_{it} = \log \mu_{it}$. The **marginal likelihood** for the observed data can then be derived by integrating out the random effects

$$\begin{aligned} L &= \prod_{i=1}^{m} \int_{Ru} \prod_{t=1}^{n_i} f(y_{it} \mid \mathbf{U}_i; \mathbf{D}) \cdot f(\mathbf{U}_i; \mathbf{D})\, d\mathbf{U}_i \\ &= \prod_{i=1}^{m} \int_{Ru} \prod_{t=1}^{n_i} \exp\left[y_{it}\theta_{it} - \exp(\theta_{it})\right]/y_{it}! \cdot (2\pi)^{-q/2} \cdot \mid \mathbf{D} \mid^{-1/2} \cdot \exp\left(-\frac{1}{2}\mathbf{U}_i^T\mathbf{D}^{-1}\mathbf{U}_i\right) d\mathbf{U}_i \\ &= c \cdot \prod_{i=1}^{m} \int_{Ru} \prod_{i=1}^{n_i} \left(2\pi^{-q/2} \mid \mathbf{D} \mid^{-1/2}\right) \exp\left[\sum_{t=1}^{n_i} y_{it}\theta_{it} - \sum_{t=1}^{n_i} \exp(\theta_{it}) - \frac{1}{2} \cdot n_i \cdot \mathbf{U}_i^T\mathbf{D}^{-1}\mathbf{U}_i\right] d\mathbf{U}_i, \end{aligned} \tag{3.12}$$

where $c = \prod_{i=1}^{m} \prod_{t=1}^{n_i} y_{it}!^{-1}$. This integral does not have a closed form solution due to $\mathbf{U}_i$ being quadratic in the exponential, and numerical methods are necessary for evaluation.

The classic procedure for estimating parameters is maximum likelihood. In the above case we need to evaluate the integral with respect to $\mathbf{U}$ and then find the estimate of $\hat{\theta}$ such that the likelihood (L) is maximized. Newton-Raphson or Fisher's scoring are the usual algorithms used to find the maximum.

Assuming the random effects ($\mathbf{U}_i$) are independent we have derived the log-likelihood in Appendix D:

$$\begin{aligned} l &= \log L \\ &= \sum_{i=1}^{m} \left[\log \left(\mid \mathbf{D} \mid^{-1/2} \right) + \log \left(\int_{Ru} \exp \left\{ \sum_{t=1}^{n_i} \left[y_{it}\theta_{it} - \exp\left(\theta_{it}\right) \right] - \frac{1}{2} n_i \mathbf{U}_i^T \mathbf{D}^{-1} \mathbf{U}_i \right\} d\mathbf{U}_i \right) \right]. \end{aligned} \tag{3.13}$$

The score equations for the fixed effects ($\boldsymbol{\beta}$) and the variance of the random effects ($\mathbf{D}$), if working with "complete' data, comprising of $(y, \mathbf{U})$, are

$$S_{\beta|y,U} = \sum_{i=1}^{m} \sum_{t=1}^{n_i} [y_{it}\mathbf{x}_{it}^T - \mathbf{x}_{it} \overbrace{\exp\left(\mathbf{x}_{it}^T\boldsymbol{\beta} + \mathbf{z}_{it}^T\mathbf{U}_i\right)}^{E(y_{it}|U_i)}] = 0, \quad \text{and} \tag{3.14}$$

$$S_{D|y,U} = \frac{1}{2}\mathbf{D}^{-1} \sum_{i=1}^{m} E\left(\mathbf{U}_i\mathbf{U}_i^T\right)\mathbf{D}^{-1} - \frac{m}{2}\mathbf{D}^{-1} = 0. \tag{3.15}$$

Direct solution of these score equations is complicated by $\mathbf{U}_i$ being an unobserved variable, and this draws us to the EM algorithm where $\mathbf{U}_i$ can be replaced by its expected value (Diggle *et al.*, 1994). This involves iterating between an E-step to estimate the conditional expectations (*i.e.* (3.16) and (3.17)) in the score equations, and an M-step in which these estimates are used to solve the score equations to gain updated parameter estimates.

The conditional expectations to be solved in the E-step are

$$E\left[E\left(y_{it} \mid \mathbf{U}_i\right)\right] = \exp\left(x_{it}^T\boldsymbol{\beta} + z_{it}^T\mathbf{D}z_{it}/2\right), \quad \text{and} \tag{3.16}$$

$$E\left(\mathbf{U}_i\mathbf{U}_i^T\right) = \int_{R^q} \mathbf{U}_i\mathbf{U}_i^T\left(2\pi\right)^{-q/2} \mid \mathbf{D} \mid^{-1/2} \exp\left(\frac{-1}{2}\mathbf{U}_i^T\mathbf{D}^{-1}\mathbf{U}_i\right) d\mathbf{U}_i \tag{3.17}$$

where q is the dimension of $\mathbf{U}_i$. The integral in (3.17) is of the same dimension as the random effects, $\mathbf{U}_i$, and needs to be evaluated numerically. This has been done successfully in cases where there was a small number of random effects (*e.g.* Hinde, 1982). However, in higher dimensional problems, alternative strategies have often been employed as the numerical integration has become too unwieldy.

The numerical analyses described above are non-standard and do not lend themselves to the use of commonly available procedures. Research during the past decade has been directed towards finding working approximations. In section 3.2 we consider some of the research and the evolution of methods that can be used for analysing correlated Poisson data. Later in the thesis, a simulation study with constructed correlated Poisson variables is used to compare the techniques.

3.2 Methods for analysing correlated Poisson data

In this review we shall compare and contrast 5 methods of approaching the approximation of GLMMs. These methods are generalized estimating equations, marginal quasi-likelihood techniques, penalized quasi-likelihood, the Gibbs sampler and mixing distributions.

3.2.1 Generalized estimating equations

Full likelihood models that arise from (1.17), (1.18) and (1.19) are generally not practicable for routine data analysis. However, GEEs extend the GLM methodology to correlated non-normal data by approximating the likelihood with a quasi-likelihood based on the mean and variance. These procedures are much simpler than full likelihood and amenable to routine use. As far as we are aware GEEs have been applied mainly to longitudinal data in medical statistics and remain relatively unexploited in other contexts.

Liang and Zeger (1986) extended generalized linear models to a longitudinal setting, when regression was required, through the use of estimating equations which took the correlation of the data into account. The joint distribution does not need to be used in this method. However, weak assumptions about it are used to obtain estimating equations that will yield consistent estimates of the regression parameters and their subsequent variances. With this method we let $\mathbf{R}(\boldsymbol{\alpha})$ be a symmetric "working" correlation matrix for the responses which is fully characterized by the vector $\boldsymbol{\alpha}$. We then let

$$\mathbf{V}_i = \mathbf{V}(\mu_i)^{1/2}\mathbf{R}(\boldsymbol{\alpha})\mathbf{V}(\mu_i)^{1/2}a(\phi). \tag{3.18}$$

Equation (3.18) will be equal to $\text{cov}(\mathbf{Y}_i)$ if $\mathbf{R}(\boldsymbol{\alpha})$ is the true correlation matrix of the response variables $\mathbf{Y}_i$. Some common forms of $\mathbf{R}(\boldsymbol{\alpha})$ are given below. Figure 3.1 a) corresponds to an autoregressive structure and b) to a uniform structure.

Figure 3.1: Example correlation structures, a) autoregressive, b) uniform.

$$a) \quad \mathbf{R}(\boldsymbol{\alpha}) = \begin{pmatrix} 1 & \cdot & \cdot \\ \alpha & 1 & \cdot \\ \alpha^2 & \alpha & 1 \end{pmatrix} \qquad b) \quad \mathbf{R}(\boldsymbol{\alpha}) = \begin{pmatrix} 1 & \cdot & \cdot \\ \alpha & 1 & \cdot \\ \alpha & \alpha & 1 \end{pmatrix}$$

The uniform correlation structure is equivalent to a random intercept (for subjects) and measurement error model in the linear model. The errors in such a model are represented by

$$\epsilon_{it} = \mathbf{z}_{it}^T \mathbf{U}_i + M_{it}, \tag{3.19}$$

where M is measurement error, distributed as $N(0, \sigma_m^2)$, $\mathbf{U}$ are the random effect coefficients which are distributed as $N(0, \sigma_u^2)$ and $\mathbf{z}_{it}$ is the corresponding design vector. As measurement error is assumed to be independent of the random subject intercept then

$$\text{var}(\epsilon_i) = \sigma_u^2 + \sigma_m^2, \quad \text{and} \tag{3.20}$$

$$\text{cov}(\epsilon_{it}, \epsilon_{ij}) = \sigma_u^2, \quad \forall t \neq j. \tag{3.21}$$

The correlation within subjects of the errors given by (3.19) is therefore

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_m^2}, \tag{3.22}$$

and this is a uniform correlation structure. Diggle *et al.* (1994) note that the uniform correlation structure is the same as that which results from a split-plot experiment. However, with longitudinal studies, the repeated measurements on the same subjects

means that use of a split-plot model is no longer designed based as the allocation of sampling times cannot be randomized. Analysis via the split-plot ANOVA is therefore model based using the assumption of uniform correlation.

Liang and Zeger (1986) define the generalized estimating equations by

$$\sum_{i=1}^{n} \mathbf{G}_i^T \mathbf{V}_i^{-1} \mathbf{S}_i = 0, \tag{3.23}$$

where $\mathbf{G}_i = \frac{d\mu_i}{d\eta_i}\mathbf{X}_i$ and $\mathbf{S}_i = \mathbf{Y}_i - \mu_i$. If the identity correlation matrix was chosen, (3.23) would reduce to the score equation for the independent GLM. The GEEs differ from the quasi-likelihood method (McCullagh and Nelder, 1989) in that $\mathbf{V}_i$ is a function of the correlation parameters, $\boldsymbol{\alpha}$, in addition to the model parameters, $\boldsymbol{\beta}$.

To compute the score equations we replace both $\boldsymbol{\alpha}$ and ϕ with $n^{1/2}$ consistent estimators $\hat{\boldsymbol{\alpha}}(\mathbf{Y}, \boldsymbol{\beta}, \phi)$ and $\hat{\phi}(\mathbf{Y}, \boldsymbol{\beta})$. The estimate then proposed for the variance of the coefficients $\hat{\boldsymbol{\beta}}$ from (3.23) is

$$\mathbf{V}_{\boldsymbol{\beta}} = \phi \left(\sum_{i=1}^{n} \mathbf{G}_i^T \mathbf{V}_i^{-1} \mathbf{G}_i \right)^{-1} \left(\sum_{i=1}^{n} \mathbf{G}_i^T \mathbf{V}_i^{-1} \mathbf{S}_i \mathbf{S}_i^T \mathbf{V}_i^{-1} \mathbf{G}_i \right) \left(\sum_{i=1}^{n} \mathbf{G}_i^T \mathbf{V}_i^{-1} \mathbf{G}_i \right)^{-1}. \tag{3.24}$$

If we were using true maximum likelihood we would have the identity

$$E\left(\frac{d^2 l}{d\theta^2}\right) + E\left(\frac{dl}{d\theta}\right)^2 = 0, \quad \text{at the maximum likelihood estimate.} \tag{3.25}$$

Then $\left(\sum_{i=1}^{n} \mathbf{G}_i^T \mathbf{V}_i^{-1} \mathbf{S}_i \mathbf{S}_i^T \mathbf{V}^{-1} \mathbf{G}_i\right)$, which is the quasi-likelihood counterpart of $\left(\frac{dl}{d\theta}\right)^2$ would cancel with $\left(\sum_{i=1}^{n} \mathbf{G}_i^T \mathbf{V}_i^{-1} \mathbf{G}_i\right)^{-1}$, the quasi-likelihood counterpart of $-\left(\frac{d^2 l}{d\theta^2}\right)$. Following from this, (3.24) would reduce to the usual parameter variance of the independent GLM. However, as we are using a quasi-likelihood technique we do not expect (3.25) to hold. Royall (1986) recommended variance estimates like

(3.24) for robust estimation of the parameter variance. This is often referred to as the sandwich estimate of the parameter variances ($\mathbf{V}(\hat{\boldsymbol{\beta}})$).

The estimates of the regression coefficients ($\hat{\boldsymbol{\beta}}$) are computed by a modified Fisher scoring technique then $\hat{\phi}$ and $\hat{\alpha}$ are calculated through moment estimation. This algorithm is often referred to as GEE1

GEE2 (Zhao and Prentice, 1990) encompassed a second stage where variance parameters were able also estimated by different regressions, exploiting the independence of the mean and dispersions. These algorithms are computationally demanding and not considered further in this thesis.

3.2.2 Approximations using penalized quasi-likelihood

Schall (1991) produced an algorithm for generalized linear models with random effects which gave the same estimates for binary data as that proposed by Stiratelli *et al.* (1984). Schall's algorithm was an adaption of Fellner's (1986) work on estimation for the LMM and works with the adjusted dependent variable (McCullagh and Nelder, 1989, p40). For the case of Poisson data, following Diggle *et al.* (1994, p174), we can specify the algorithm as follows.

The adjusted dependent variable is represented by

$$\begin{aligned} \xi_{it} &= g(\mu_{it}) + (y_{it} - \mu_{it})\frac{d\eta_{it}}{d\mu_{it}} \\ &= \log \mu_{it} + \frac{(y_{it} - \mu_{it})}{\mu_{it}} \end{aligned} \tag{3.26}$$

and we use

$v_{it} = \text{var}(y_{it} \mid \mathbf{U}_i) = a(\phi)V(\mu_{it}) = \mu_{it}$,

$\mathbf{Q}_i = \text{diag}\left[v_{it}\left(\frac{d\eta_{it}}{d\mu_{it}}\right)^2\right] = \text{diag}[\mu_{it}^{-1}]$, and

$\mathbf{V}_i = \mathbf{Q}_i^{-1} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T$.

The first step of the algorithm is to use the current estimate of $\mathbf{D}$ to obtain updated estimates for the fixed and random effects, $\boldsymbol{\beta}$ and $\mathbf{U}$. To do this, the following equations, (3.27) and (3.28) are solved iteratively;

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{V}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\boldsymbol{\xi}, \tag{3.27}$$

$$\hat{\mathbf{U}}_i = \mathbf{D}\mathbf{Z}_i^T\mathbf{V}_i^{-1}\left(\boldsymbol{\xi}_i - \mathbf{X}_i\boldsymbol{\beta}\right). \tag{3.28}$$

The second step then consists of using the updated estimates of $\boldsymbol{\beta}$ and $\mathbf{U}$ to calculate an updated estimate for $\mathbf{D}$. The equation used for this is

$$\hat{\mathbf{D}}^{(j+1)} = m^{-1}\left\{\sum_{i=1}^{m}\left[\hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^T + \left(\mathbf{Z}_i^T\mathbf{Q}_i^{-1}\mathbf{Z}_i + \mathbf{D}^{(j)^{-1}}\right)^{-1}\right]\right\}. \tag{3.29}$$

These two steps are iterated until convergence of the parameters is obtained.

It can be seen that (3.26) is the linearized form (first order Taylor's expansion), in the Poisson case, of $\log(y)$:

$$\log(y) \simeq \log(\mu) + \frac{(y-\mu)}{\mu}, \tag{3.30}$$

where $\mu = \exp(x\beta + zU)$. An advantage of this linearization scheme is that the random effects have been kept on their original scale, thereby giving their estimates an easily interpreted value.

3.2.3 Approximations using marginal quasi-likelihood

Breslow and Clayton (1993) give an algorithm for the GLMM using MQL. This marginal model is appropriate when the fixed effects are used to determine their effect on the population average rather than on individual profiles.

Using this model (in the Poisson case) we take

$$E(y_{it}) = \mu_{it}^{*} = \exp\left(\mathbf{x}_{it}^{T}\boldsymbol{\beta}\right). \tag{3.31}$$

From Appendix C, (C.11), we see that this does not coincide with the marginal mean given by the model (3.9), so (3.31) is thought of as a crude 1st order approximation to the population mean averaged over infinite subjects, then $\mathbf{D} \approx 0$.

Rather than using the adjusted dependent variable, (as in Schall, 1991), the value of y is linearized using a Taylor's expansion about $\mathbf{U}_i = 0$;

$$\begin{aligned} y_{it} &\approx \exp\left(\mathbf{x}_{it}^{T}\boldsymbol{\beta}\right) + \mathbf{U}_i \left\{ \frac{d}{d\mathbf{U}_i}\left[\exp\left(\mathbf{x}_{it}^{T}\boldsymbol{\beta} + \mathbf{z}_{it}^{T}\mathbf{U}_i\right)\right] |_{\mathbf{U}_i=0} \right\} + \epsilon_{it} \\ &= \exp\left(\mathbf{x}_{it}^{T}\boldsymbol{\beta}\right) + \mathbf{z}_{it}^{T}\mathbf{U}_i \exp\left(\mathbf{x}_{it}^{T}\boldsymbol{\beta}\right) + \epsilon_{it}, \quad \text{and then} \\ &\quad . \end{aligned} \tag{3.32}$$

$$\begin{aligned} \operatorname{var}(y_{it}) &\approx \operatorname{var}\left[\mathbf{z}_{it}^{T}\mathbf{U}_i \exp\left(\mathbf{x}_{it}^{T}\boldsymbol{\beta}\right)\right] + \operatorname{var}(\epsilon_{it}) \\ &= \exp\left(\mathbf{x}_{it}^{T}\boldsymbol{\beta}\right) \mathbf{z}_{it}^{T}\mathbf{D}\mathbf{z}_{it} \exp\left(\mathbf{x}_{it}^{T}\boldsymbol{\beta}\right) + a(\phi_{it})\, V(\mu_{it}) \\ &= \Delta_{it}^{-1}\mathbf{z}_{it}^{T}\mathbf{D}\mathbf{z}_{it}\Delta_{it}^{-1} + v_{it}^{*} \\ &= \Delta_{it}^{-1}\left(\Delta_{it}v_{it}^{*}\Delta_{it} + \mathbf{z}_{it}^{T}\mathbf{D}\mathbf{z}_{it}\right)\Delta_{it}^{-1}, \end{aligned} \tag{3.33}$$

where $\Delta_{it} = \frac{d\eta_{it}}{d\mu_{it}} |_{U_i=0} = \left[\exp\left(\mathbf{x}_{it}^T\boldsymbol{\beta}\right)\right]^{-1}$, and $v_{it}^* = \mu_{it}^* = \exp\left(\mathbf{x}_{it}^T\boldsymbol{\beta}\right)$.

The fixed effects are estimated using the quasi-likelihood equations for dependent outcomes (McCullagh and Nelder, 1989, pp332-336), which in our Poisson model are

$$\mathbf{X}\left(\boldsymbol{\Delta}\boldsymbol{v}\boldsymbol{\Delta} + \mathbf{ZDZ}^T\right)^{-1}\boldsymbol{\Delta}\left(\mathbf{Y} - \boldsymbol{\mu}^*\right) = 0. \tag{3.34}$$

We proceed by allowing $\eta_{it}^* = \mathbf{x}_{it}^T\boldsymbol{\beta}$, and using the dependent variable

$$\xi_{it} = \eta_{it}^* + \Delta_{it}\left(y_{it} - \mu_{it}^*\right), \tag{3.35}$$

and then the inverse weight matrix becomes

$$\mathbf{V}_i = \left(\boldsymbol{\Delta}_i\boldsymbol{v}_i^*\boldsymbol{\Delta}_i + \mathbf{z}_i^T\mathbf{D}\mathbf{z}_i\right). \tag{3.36}$$

The fixed effects of the model are then estimated by iteratively reweighted least squares:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{V}^{*-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\mathbf{V}^{*-1}\boldsymbol{\xi}^*\right). \tag{3.37}$$

We can see that the differences of the approach to that of Schall (1991) are in the linearization technique, and as a result the methods differ only slightly.

The estimates of the variance components, $\hat{\boldsymbol{\alpha}}$, where $\mathbf{D}(\boldsymbol{\alpha}) = 0$, are obtained using pseudo-likelihood and the REML estimating equations (Harville, 1977)

$$-\frac{1}{2}\left[\left(\boldsymbol{\xi}^* - \mathbf{X}\boldsymbol{\beta}\right)^T\mathbf{V}^{*-1}\frac{\partial\mathbf{V}^*}{\partial\alpha_j}\mathbf{V}^{*-1}\left(\boldsymbol{\xi}^* - \mathbf{X}\boldsymbol{\beta}\right) - tr\left(P\frac{\partial\mathbf{V}^*}{\partial\alpha_j}\right)\right] = 0, \tag{3.38}$$

where $P = \mathbf{V}^{*-1} - \mathbf{V}^{*-1}\mathbf{X} \times \left(\mathbf{X}^T\mathbf{V}^{*-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{*-1}$. In this MQL method we iterate between (3.37) and (3.38) until convergence is obtained, and then the BLUP estimates of the random effects, $\mathbf{U}$, are calculated as

$$\tilde{\mathbf{U}} = \mathbf{DZ}^T\mathbf{V}^{*-1}\left(\boldsymbol{\xi}^* - \mathbf{X}\boldsymbol{\beta}\right). \tag{3.39}$$

This method would be reasonable to use if the random effects themselves were of no interest, but were rather seen as a way of adequately accounting for the correlation present in the data. This is because the linearization scheme has placed the estimates of the random effects on a new scale, making interpretation difficult. However, if the random effects were of some interest then the fact that the linearization scheme used in the MQL approach has potentially set them on a different scale could make interpretation a little more difficult.

3.2.4 Approximations using the Gibbs sampler

Zeger and Karim (1991) set the GLMM in a Bayesian setting, using the Gibbs sampler to overcome the difficulties associated with likelihood inferences. The Bayesian formulation is to let $p(\boldsymbol{\beta}, \mathbf{D})$ be the joint prior for these random variables and derive the posterior distribution:

$$f(\boldsymbol{\beta}, \mathbf{D} \mid \mathbf{y}) = \frac{\prod_{i=1}^{m} \int f(\mathbf{y}_i \mid \mathbf{U}_i, \boldsymbol{\beta}) g(\mathbf{U}_i \mid \mathbf{D}) p(\boldsymbol{\beta}, \mathbf{D}) d\mathbf{U}_i}{\int \int \prod_{i=1}^{m} \int f(\mathbf{y}_i \mid \mathbf{U}_i, \boldsymbol{\beta}) g(\mathbf{U}_i \mid \mathbf{D}) p(\boldsymbol{\beta}, \mathbf{D}) d\mathbf{U}_i d\boldsymbol{\beta} d\mathbf{D}}. \tag{3.40}$$

Because the denominator is a normalizing constant, and independent of $\boldsymbol{\beta}$ and $\mathbf{D}$, the posterior mode can be derived using only the numerator. The posterior distribution (3.40), and similarly $f(\mathbf{U}_i \mid \mathbf{y})$ are usually both numerically intractable and so the Gibbs sampler is proposed as a three step method. With this method we seek the joint distribution of $\mathbf{U}_i, \boldsymbol{\beta}$ and $\mathbf{D}$, as well as the posterior distribution (3.40) and $f(\mathbf{U}_i \mid \mathbf{y})$. The Gibbs sampler allows us to achieve this by working with the conditional distributions of $\mathbf{U}_i, \boldsymbol{\beta}$ and $\mathbf{D}$, which are easier to sample from.

The algorithm of Zeger and Karim (1991) iterates between three steps to obtain updated estimates of $(\boldsymbol{\beta}, \mathbf{U}, \mathbf{D})$. This follows from the work of Geman and Geman (1984) which showed that

$$\lim_{N\to\infty} \left(\boldsymbol{\beta}^{(N)}, \mathbf{U}^{(N)}, \mathbf{D}^{(N)}\right) = (\boldsymbol{\beta}, \mathbf{U}, \mathbf{D}),$$

where convergence was obtained exponentially under mild conditions.

The first step is to obtain a simulated value for the fixed effects $(\boldsymbol{\beta})$ from the conditional distribution $[\boldsymbol{\beta} \mid \mathbf{U}^{(\cdot)}, \mathbf{y}]$. By using the current estimates of $\mathbf{U}$, the random effects can be set as an offset to a standard GLM, and current estimates for $\hat{\boldsymbol{\beta}}^{(k)}$, and their variance, $\hat{\mathbf{V}}_{\boldsymbol{\beta}^{(k)}}$, determined. From these values we simulate an updated estimate $\hat{\boldsymbol{\beta}}^{(k+1)}$. It is difficult to know when sample size is large enough to justify sampling a new estimate of $\boldsymbol{\beta}$ directly from the $MVN\left(\hat{\boldsymbol{\beta}}^{(k)}, \hat{\mathbf{V}}_{\boldsymbol{\beta}^{(k)}}\right)$, so a rejection sampling technique is used. This process is described briefly below, and a more detailed account is available in Morgan (1984).

It is known that the true density of the fixed effects $\boldsymbol{\beta}$ is

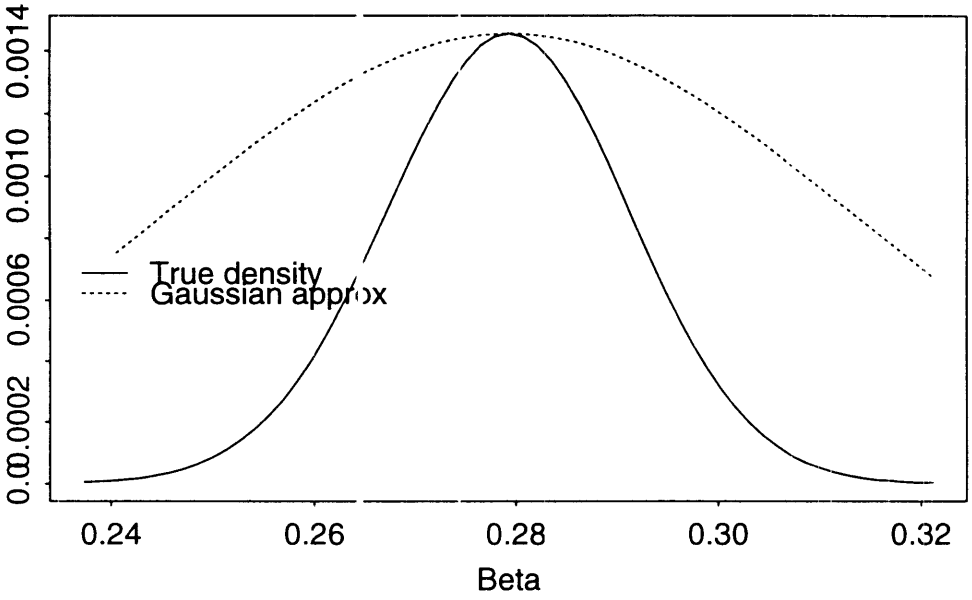
$$f(\boldsymbol{\beta}) \propto \prod_{it} f(y_{it} \mid \mathbf{U}_i), \tag{3.41}$$

and we can represent the multivariate Gaussian approximation to $\boldsymbol{\beta}$ by

$$g(\boldsymbol{\beta}) = (2\pi)^{-p/2} \left|\mathbf{V}_{\boldsymbol{\beta}}\right|^{-1/2} \exp\left[-\frac{1}{2}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)^T \mathbf{V}_{\boldsymbol{\beta}}^{-1} \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right], \tag{3.42}$$

with $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$ being p dimension vectors and $\mathbf{V}_{\boldsymbol{\beta}}$ is a d x d matrix.

Figure 3.2: Illustration of rejection sampling concept.



The rejection sampling algorithm requires a constant, c, to be chosen such that $c \cdot g(\boldsymbol{\beta}) \geq f(\boldsymbol{\beta})$ over the required range, as illustrated in Figure 3.2. An updated estimate $\hat{\boldsymbol{\beta}}^{(k+1)}$ is then produced as follows;

1. generate $\boldsymbol{\beta}^*$ from $g(\hat{\boldsymbol{\beta}}^{(k)})$,

2. calculate the ratio $r = f(\boldsymbol{\beta}^*)/c \cdot g(\boldsymbol{\beta}^*)$, and

3. generate a Uniform(0,1) random variable, u. If $r > u$ accept $\boldsymbol{\beta}^*$ as the updated $\hat{\boldsymbol{\beta}}^{(k+1)}$, otherwise repeat until an acceptable update is found.

From Figure 3.2, it can be seen that the probability of $\boldsymbol{\beta}^*$ being rejected increases with its distance from $\hat{\boldsymbol{\beta}}^{(k)}$. This rejection sampling can be done at little cost and effort computationally.

The second step is to update the estimated variance-covariance matrix for the random effects, $\mathbf{D}$. As $\mathbf{U}_i \sim MVN(0, \mathbf{D})$ the standard non-informative prior (given

by Box and Tiao, 1973) is

$$P(\mathbf{D}) \propto |\mathbf{D}|^{-(q+1)/2}, \tag{3.43}$$

where q is the dimension of $\mathbf{U}_i$. Then $\mathbf{D}^{-1}$ has a posterior distribution which follows a Wishart distribution with $(r - q + 1)$ degrees of freedom and parameters $S^{(k)} = \mathbf{U}^{(k)}\mathbf{U}^{(k)T}$. An update of $\mathbf{D}$ can be obtained by generating a standardized Wishart distribution (Odell and Feiveson, 1966) with the appropriate degrees of freedom, represented here by W^*. Using a Choleski decomposition, $S^{(k)} = \mathbf{H}^{(k)T}\mathbf{H}^{(k)}$, we obtain the updated estimate

$$\mathbf{D}^{(k+1)} = \left(\mathbf{H}^{(k)T}\mathbf{W}^*\mathbf{H}^{(k)}\right)^{-1}. \tag{3.44}$$

The third step is to generate updated random effects ($\mathbf{U}_i^{(k+1)}$). The density of these random effects is given by

$$f(\mathbf{U}_i \mid \boldsymbol{\beta}, \mathbf{D}, \mathbf{y}_i) = \frac{f(\mathbf{y}_i \mid \mathbf{U}_i, \boldsymbol{\beta}) g(\mathbf{U}_i \mid \mathbf{D}) p(\boldsymbol{\beta}, \mathbf{D})}{\int f(\mathbf{y}_i \mid \mathbf{U}_i, \boldsymbol{\beta}) g(\mathbf{U}_i \mid \mathbf{D}) p(\boldsymbol{\beta}, \mathbf{D}) d\mathbf{U}_i}. \tag{3.45}$$

The denominator of (3.45) involves an intractable integral which is avoided by the Gibbs sampler. Only a simulated value from the conditional distribution is needed, and once again this is obtained using rejection sampling. For brevity let the numerator of (3.45) be known as $p(\mathbf{U})$.

We use the adjusted dependent variable given by (3.26), and updated estimates $\hat{\boldsymbol{\beta}}^{(k+1)}$ and $\hat{\mathbf{D}}^{(k+1)}$, to iterate between (3.46) and (3.47) and gain measures of mode and curvature;

$$\hat{\mathbf{U}}_i = \left(\mathbf{Z}_i^T\mathbf{V}_i\mathbf{Z}_i + \hat{\mathbf{D}}\right)^{-1}\mathbf{Z}_i^T\mathbf{V}_i\left(\boldsymbol{\xi}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}\right), \tag{3.46}$$

$$\hat{\boldsymbol{v}}_i = \left(\mathbf{Z}_i^T\mathbf{V}_i\mathbf{Z}_i + \hat{\mathbf{D}}^{-1}\right)^{-1}. \tag{3.47}$$

Rejection sampling is now conducted to select an update on the estimates for the random effects. The procedure is;

1. generate $\mathbf{U}_i^* \sim MVN\left(\hat{\mathbf{U}}, c_2\hat{\boldsymbol{v}}_i\right)$. where c_2 is a constant.

2. calculate $c_{1i} = p\left(\hat{\mathbf{U}}_i\right)/g\left(\hat{\mathbf{U}}_i\right)$ (this ensures equal ordinates at the common mode $\hat{\mathbf{U}}_i$), and

3. generate Unif(0,1), and call it u If the ratio $p(\mathbf{U}_i^*)/c_{1i} \cdot g(\mathbf{U}_i^*) \geq u$ we allow $\mathbf{U}_i^* = \mathbf{U}_i^{(k+1)}$, otherwise the process is repeated.

This three step algorithm is iterated a large number of times until convergence is considered to be satisfactory, so we can be sure of stochastic stability.

3.2.5 Dealing with overdispersion by using mixing distributions

Van de Ven and Weber (1995) created the overdispersion seen in the Poisson random effects model by compounding the Poisson random variable with a gamma random variable. Then the mean of the Poisson and the gamma variance were modelled via log-linear models. By mixing the Poisson distribution with the gamma to account for subject induced overdispersion, the integrals needed to calculate the likelihood are solvable. Lee and Nelder (1996) extended the idea of using conjugate distributions to other members of the exponential family.

The model is defined as

$$Y_t|\gamma_i \sim \quad \text{Poisson}(\gamma_i \mu_{it}), \tag{3.48}$$

where γ_i are independent $G(x, \alpha_i)$. In this model, G, is the distribution of a $\Gamma(\alpha_i, \frac{1}{\alpha_i})$ random variable. With this model we have

$$\begin{aligned} \log(\mu_{it}) &= \mathbf{x}_{it}^T \boldsymbol{\beta}, \qquad \text{and} \\ \log(\alpha_i) &= \mathbf{z}_i^T \boldsymbol{\eta}, \end{aligned} \tag{3.49}$$

where $\boldsymbol{\beta}$ are parameter values for the fixed effects, and $\boldsymbol{\eta}$ relate to the gamma random variable. It can be shown (Appendix E) that under this model the likelihood is given by

$$L = \prod_{i=1}^{m} \left[\frac{\Gamma(\alpha_i + y_{i\cdot})}{\Gamma(\alpha_i)\left(\prod_{t=1}^{n_i} y_{it}!\right)} \left(\frac{\alpha_i}{\alpha_i + \mu_{i\cdot}}\right)^{\alpha_i} \prod_{t=1}^{n_i} \left(\frac{\mu_{it}}{\alpha_i + \mu_{i\cdot}}\right)^{y_{it}} \right], \tag{3.50}$$

where $x_{i\cdot} = \sum_{t=1}^{n_i} x_{it}$.

The solutions for the fixed and random effects are then given by iteratively solving the following equations (3.51) and (3.52), using Fisher's scoring method;

$$\hat{\boldsymbol{\beta}}_{(s+1)} = \hat{\boldsymbol{\beta}}_{(s)} + \left(\sum_{i=1}^{m} \mathbf{X}_i^T \mathbf{A}_i \mathbf{V}_i^{-1} \mathbf{A}_i \mathbf{X}_i \right)^{-1} \sum_{i=1}^{m} \mathbf{X}_i^T \mathbf{A}_i \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i), \tag{3.51}$$

$$\hat{\boldsymbol{\eta}}_{(s+1)} = \hat{\boldsymbol{\eta}}_{(s)} + \left(\mathbf{Z}^T \mathbf{A}_{m+1} \mathbf{B}_{m+1} \mathbf{A}_{m+1} \mathbf{Z} \right)^{-1} \mathbf{Z}^T \mathbf{A}_{m+1} \mathbf{l}_{m+1}. \tag{3.52}$$

In these equations we use

$$\begin{aligned} \mathbf{A}_i &= \text{diag}(\mu_{i1}, \ldots, \mu_{in_i}), \\ \mathbf{V}_i &= \mathbf{A}_i + \boldsymbol{\mu}_i \boldsymbol{\mu}_i^T / \alpha_i, \end{aligned}$$

$$
\begin{aligned}
\mathbf{A}_{m+1} &= \operatorname{diag}(\alpha_1, \ldots, \alpha_m), \\
\mathbf{l}_{m+1} &= \left(\frac{\partial \log L}{\partial \alpha_1} \ldots, \frac{\partial \log L}{\partial \alpha_m}\right)^T, \quad \text{where} \\
\frac{\partial \log L}{\partial \alpha_i} &= \sum_{r=1}^{Y_{i\cdot}} \left(\frac{1}{\alpha_i + r - 1}\right) + \log\left(\frac{\alpha_i}{\alpha_i + \mu_{i\cdot}}\right) + \frac{\mu_{i\cdot} - y_{i\cdot}}{\mu_{i\cdot} + \alpha_i}, \quad \text{and} \\
\mathbf{B}_{m+1} &= \operatorname{diag}(b_{m+1,1}, \ldots, b_{m+1,n}), \quad \text{where} \\
b_{m+1,i} &= \sum_{r=1}^{Y_{i\cdot}} \left[\left(\frac{1}{\alpha_i + r - 1}\right)^2\right] - \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i + 1)}\left(\frac{\mu_{i\cdot}}{\mu_{i\cdot} + \alpha_i}\right).
\end{aligned}
$$

Van de Ven and Weber (1995) also proposed a method which did not require the mixing distribution $G(x, \alpha_i)$ to be gamma. This required use of quasi-likelihood methods. The **extended quas -likelihood** for the subject totals is given by

$$
QL = \sum_{i=1}^{m}\left[-\frac{1}{2}\log(2\pi y_{i\cdot}) - \frac{1}{2}\log\left(1 + \frac{y_{i\cdot}}{\alpha_i}\right) + y_{i\cdot}\log\left(\frac{\mu_{i\cdot}}{y_{i\cdot}}\right) + (y_{i\cdot} + \alpha_i)\log\left(\frac{y_{i\cdot} + \alpha_i}{\mu_{i\cdot} + \alpha_i}\right)\right]. \tag{3.53}
$$

To estimate the parameters under the extended quasi-likelihood model $\mathbf{l}_{m+1}$ and $\mathbf{B}_{m+1}$ in (3.52) are replaced by

$$
\begin{aligned}
l^*_{m+1} &= \operatorname{diag}(\frac{\partial Q}{\partial \alpha_1}, \ldots, \frac{\partial Q}{\partial \alpha_m}) \quad \text{where} \\
\frac{\partial Q}{\partial \alpha_i} &= \log\left(\frac{\alpha_i + y_{i\cdot}}{\alpha_i}\right) + \frac{y_{i\cdot}}{2\alpha_i(\alpha_i + y_{i\cdot})} + \log\left(\frac{\alpha_i}{\alpha_i + \mu_{i\cdot}}\right) + \frac{\mu_{i\cdot} - y_{i\cdot}}{\mu_{i\cdot} + \alpha_i}, \quad \text{and} \\
\mathbf{B}^*_{m+1} &= \operatorname{diag}(b^*_{m+1,1}, \ldots, b^*_{m+1,n}), \quad \text{where} \\
b^*_{m+1,i} &= -\frac{1}{\alpha_i} \times l^*_{m+1} - \frac{\mu_{i\cdot}^2 + \alpha_i y_{i\cdot}}{\alpha_i(\mu_{i\cdot} + \alpha_i)^2} + \frac{y_{i\cdot}}{\alpha_i(\alpha_i + y_{i\cdot})} + \frac{y_{i\cdot}(2\alpha_i + y_{i\cdot})}{2\alpha_i^2(\alpha_i + y_{i\cdot})^2}.
\end{aligned}
$$

In the next chapter we will compare the performance of the models through the use of a small simulation study with some constructed correlated Poisson data.

Chapter 4

Simulation Study

A small simulation study was conducted to examine the performance of GEEs (using the correct correlation structure), the PQL (Schall, 1991), the EQL (Van de Ven and Weber, 1995) and the MQL (Breslow and Clayton, 1993) approximate methods for GLMMs. Restrictions on computational resources did not allow the inclusion of the Gibbs sampler approach in this study, which we had programmed initially using Splus, and an attempt was made to model by using the program BUGS 0.5 (Spiegelhalter, Thomas, Best and Gilks, 1996)

4.1 Outline of algorithm

Sim (1993) provided an algorithm for simulating Poisson data with a fixed covariance matrix which we shall now briefly describe. First, we define the elements of the vector $\mathbf{Y} = (Y_1, \cdots, Y_p)$ as independent Poisson variables with mean λ_i $(i = 1, \cdots, p)$, and

a second vector $\mathbf{W} = (W_1, \cdots, W_p)$ as our required Poisson variates with a fixed covariance matrix, $\mathbf{\Sigma}$, the elements of which are σ_{it} $(i, t = 1, \cdots, p)$. Hence the mean values of W_i are σ_{ii}.

The required vector, $\mathbf{W}$, can be obtained from the independent vector, $\mathbf{Y}$, using the following transformation;

$$\begin{aligned} W_1 &= Y_1 \qquad \text{(as first sample is independent)} \\ W_2 &= \alpha_{21} * Y_1 + Y_2 \\ &\vdots \qquad \vdots \\ W_p &= \alpha_{p1} * Y_1 + \cdots + \alpha_{p,p-1} * Y_{p-1} + Y_p \end{aligned} \tag{4.1}$$

where $\alpha_{it} \in [0, 1]$, $1 \leq i \leq t \leq p$. This new random variable $\alpha * Y$ is defined as:

$$\alpha * Y = \sum_{i=1}^{Y} B_i(\alpha) \tag{4.2}$$

where $\alpha \in [0, 1]$, Y is an independent Poisson variate and $B_i(\alpha) \sim iid$ Bernoulli(α). Therefore it is the sum of Y independent units, with probability α of being retained in the transformation (4.1).

This algorithm can be shown to produce the required Poisson variates (when applied under some constraints) because $\alpha * Y$ is itself a random Poisson variate with a lower mean than Y, $\sigma_{it} = \text{Cov}(W_i, W_t) = \sum_{k=1}^{i-1} \alpha_{ik}\alpha_{tk}\lambda_k + \alpha_{ti}\lambda_i$ and $\sigma_{tt} = \text{Var}(W_t) = \sum_{k=1}^{t-1} \alpha_{tk} + \lambda_t$.

4.2 Implementation of algorithm

The algorithm given by Sim (1993) and implemented using the Genstat procedure of Murison and Harden (Appendix F) was used to simulate correlated Poisson data with exchangeable correlation structure. Sim's algorithm allows any specified correlation structure, however we limit ourselves to the exchangeable structure in this study as it is a structure which we would reasonably expect to see in practice, as it corresponds to a random intercept model, and will serve to test the different models. Several levels of correlation were also simulated, $\rho = 0.2$ and 0.5. These levels were chosen as 0.2 is representative of approximately the minimum level of correlation which would need to be modelled and 0.5 represents a high degree of correlation. For each correlation size, 50 sets of data were simulated and for each parameter the mean and standard deviation of the 50 estimates was calculated for comparison with the known parameter.

The marginal means for 2 treatment groups at 4 times were modelled by

$$\log(\mu) = \beta_0 + \beta_{\text{treatment}} + \beta_{\text{time}} \tag{4.3}$$

where β_0 and $\beta_{\text{treatment}}$ are subject to random effects. The known values of β_0, $\beta_{\text{treatment}}$ and β_{time} were chosen to give the means of 5, 6, 7, 8 for group 1 and 6, 7, 8, 9 for group 2. This simple mean model allows direct focus on the comparison of techniques, unobstructed by treatment nuances.

The PQL (Schall, 1991) and MQL (Breslow and Clayton, 1993) models were fitted using the procedure GLMM which is available in the Genstat 5.3.2 procedure library.

The GEE1 models were fitted using procedures which are also available in the Genstat 5.3.2 procedure library. The extended quasi-likelihood (EQL) technique of Van de Ven and Weber (1995) was fitted using the Splus program provided in Appendix G.

4.3 Simulation results

The results for the parameters of interest (intercept and treatment) are given in Table 4.1 and Table 4.2.

Table 4.1: Simulation results (±se) for exchangeable correlation structure, $\rho = 0.2$

Parameter	Sample size (n)	Actual	GEE1	MQL	EQL	PQL
β_0	5	1.61	1.48 (±0.181)	1.53 (±0.177)	1.53 (±0.177)	1.50 (±0.176)
	10		1.60 (±0.137)	1.60 (±0.134)	1.60 (±0.134)	1.56 (±0.137)
	20		1.59 (±0.092)	1.59 (±0.090)	1.59 (±0.090)	1.56 (±0.098)
$\beta_{\text{treatment}}$	5	0.18	0.24 (±0.235)	0.18 (±0.207)	0.18 (±0.207)	0.18 (±0.208)
	10		0.18 (±0.137)	0.19 (±0.127)	0.19 (±0.127)	0.19 (±0.132)
	20		0.18 (±0.097)	0.18 (±0.094)	0.18 (±0.094)	0.18 (±0.097)

The results from the PQL method of analysis seemingly indicate a small but consistent bias which is not prominent in the other approaches. Considering the

results it can be seen that this bias decreases with both increasing sample size and a decrease in the level of correlat on. The bias of $\beta_{\text{treatment}}$ is also small at low levels of correlation, where mainly β_1 is effected, and this bias increases with correlation levels.

Performances of all methods were better at lower levels of correlation when the sample size was small (10 subjects per group). This can be seen in the results by both an increase in the accuracy of the estimate of the mean at correlation level 0.2 as compared with 0.5, and a decrease in the subsequent standard deviations of these means.

Table 4.2: Simulation results (±se) for exchangeable correlation structure, $\rho = 0.5$

Parameter	Sample size (n)	Actual	GEE1	MQL	EQL	PQL
β_0	5	1.61	1.58 (±0.200)	1.57 (±0.169)	1.57 (±0.171)	1.53 (±0.172)
	10		1.55 (±0.158)	1.56 (±0.145)	1.56 (±0.145)	1.52 (±0.150)
	20		1.58 (±0.097)	1.58 (±0.093)	1.58 (±0.093)	1.55 (±0.098)
$\beta_{\text{treatment}}$	5	0.18	0.18 (±0.268)	0.22 (±0.241)	0.23* (±0.238)	0.23 (±0.249)
	10		0.22 (±0.200)	0.21 (±0.172)	0.21 (±0.172)	0.21 (±0.173)
	20		0.21 (±0.125)	0.19 (±0.118)	0.19 (±0.118)	0.20 (±0.118)

NOTE: * Simulation 26 values failed to compute for EQL estimates. All other run estimates were the same as MQL to 4 decimal places.

Table 4.1 and Table 4.2 illustrate that with larger sample sizes (20 subjects per group) the estimates of the means are more consistent between correlation levels for all four techniques tested.

The bias illustrated in this simulation study is consistent with the findings of Kuk (1995) and Lin and Breslow (1996) who proposed methods to correct for this bias.

4.4 Practical implications

In the results for this simulated exchangeable correlation structure we see that GEE1 and MQL are comparable in both their estimates of the mean and their corresponding standard deviations. This was observed in the case where the working correlation matrix for the GEE1 procedure was correctly specified. Crowder (1995) illustrated that problems may be encountered when the structure of the working matrix is incorrectly specified. In his paper he showed that, over a specific range of correlation levels, estimations of the parameters, α, in the working correlation matrix, have no solution if specified as exchangeable when the true correlation is autoregressive.

In our consulting role we feel that we would use MQL procedures in an analysis where spatial and/or temporal correlation was present in the data, and accounting for this correlation in terms of the model was intuitive or of some interest. The use of MQL in field trials may often involve spatial effects that are due to serial correlation, which would be indicated by field trends, rather than the patchy row and column effects seen in the nutgrass data. In such cases, a more sophisticated model than

the one used in the nutgrass da:a analysis, which could incorporate serial correlation terms to account for trend, woi ld be required.

The primary use we see for GEE1 procedures is in cases when a study involves repeated measurements in time, such as animal experiments. The GEE procedure has the facilities required to cc nstruct complex correlation structures through their specification in GEECORREL. This feature gives GEE greater scope for use in field trials, however, this involves i nderstanding the nature of the correlation, that is how the correlation is induced serial correlation or random effects), and thinking of this correlation in model terms. Whereas GEEs, as discussed in Section 3.2.1, are generally used to account for correlation when the cause or source of the correlation is of little interest other than to refine our parameter estimates.

The trend of increased accuracy in the PQL estimates with increasing sample size, which can be seen in Table 4.1 and Table 4.2, combines with a lower accuracy than the other techniques. This has ed us to conclude that the use of PQL is appropriate only in cases where there is a large sample. We will discuss some possible reasons for the failure of PQL under the conditions of small sample size in Section 4.5.

With these simulated data sets it was observed that the estimates for the fixed effects from the EQL method were equivalent (to four decimal places) to the estimates from the MQL analysis. To help determine whether this was a perculiarity of the simulated data or a real effect, an analysis of the data provided in the paper by Van de Ven and Weber (1995) was carried out using MQL techniques. A comparison of

the estimated fixed coefficients for the model, given in the paper as

$$\log(\mu_{it}) = \sum_{k=1}^{7} x_{itk}\beta_k$$

where $(\beta_1, \cdots, \beta_6)$ relates to the slide number and β_7 indicates a counter status of zero, is given in Table 4.3.

Table 4.3: Comparison of MQL and EQL results for slide example

Method	β_1	β_2	β_3	β_4	β_5	β_6	β_7
EQL	3.189	0.1945	0.0536	-0.0160	-0.1069	-0.1158	-0.3834
MQL	3.177	0.1896	0.0502	0.0053	-0.0930	-0.1202	-0.3712

These results were considered to be sufficiently different and combined with the non convergence of a simulation set (Table 4.2) seemed to rule out the possibility that the methods were equivalent.

In the EQL analysis it was also noted that the convergence of the random effects was heavily dependent on the starting values used in the algorithm. However, despite the failure of the random effects to converge, the estimates for the fixed effects had little difficulty in convergence.

4.5 PQL with small sample sizes

In the case of the linear mixed model, parameters are estimated using REML but there is no real counterpart to REML in the nonlinear setting. In the GLM setting, PQL methods use REML type estimation with the updated dependent variable, given by

(3.26). In section 3.2.2 we also noted that (3.26), in the Poisson case, was equivalent to the first order Taylor's expansion of $\log(\mathbf{y})$ (given by (3.30)).

The PQL algorithm is derived by using the adjusted dependent variable ($\boldsymbol{\xi}$) in a linear mixed model context. As such, the parameter estimates are only reliable for making inferences about data if $\boldsymbol{\xi}$ is a reliable approximation for $\log(\mathbf{y})$.

Using (3.26), we may denoted the adjusted dependent variable by

$$\begin{aligned}
\boldsymbol{\xi} &= g(\boldsymbol{\mu}) + (\mathbf{y} - \boldsymbol{\mu})g'(\boldsymbol{\mu}), \quad \text{and in the Poisson case,} \\
&= \log(\boldsymbol{\mu}) + (\mathbf{y} - \boldsymbol{\mu})\frac{1}{\boldsymbol{\mu}} \\
&= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \frac{\mathbf{y} - \boldsymbol{\mu}}{\boldsymbol{\mu}}, \quad \text{and then,} \\
\operatorname{cov}(\boldsymbol{\xi}) &= \mathbf{Z}\mathbf{D}\mathbf{Z}^T + \sigma^2 I \left(\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}}\right)^2, \quad \text{in the general case, and} \\
&= \mathbf{Z}\mathbf{D}\mathbf{Z}^T + \operatorname{diag}(\boldsymbol{\mu}^{-1}), \quad \text{in the Poisson case.}
\end{aligned} \tag{4.4}$$

For fixed $\mathbf{D}$, updated estimates of $\boldsymbol{\beta}$ and $\mathbf{U}$ are found by solving (3.27) and (3.28). However, an estimate of $\mathbf{D}$ must be used in these equations. We take the score equation for $\mathbf{D}$, (3.15), and its conditional expectation, (3.17), to gain an estimate for $\mathbf{D}$. From these equations we derive

$$\begin{aligned}
\hat{\mathbf{D}} &= m^{-1}\sum_{i=1}^{m} E\left(\mathbf{U}_i\mathbf{U}_i^T \mid \mathbf{y}_i\right) \\
&= m^{-1}\sum_{i=1}^{m} E\left(\mathbf{U}_i \mid \mathbf{y}_i\right) E\left(\mathbf{U}_i \mid \mathbf{y}_i\right)^T + m^{-1}\sum_{i=1}^{m} \operatorname{var}\left(\mathbf{U}_i \mid \mathbf{y}_i\right)
\end{aligned} \tag{4.5}$$

Approximating the distribution $(\mathbf{U}_i \mid \mathbf{y}_i)$ by a Gaussian distribution allows us to replace $E(\mathbf{U}_i \mid \mathbf{y}_i)$ by the mode of the Gaussian counterpart. Using the curvature of

the Gaussian distribution,

$$\text{var}\left(\mathbf{U}_i \mid \mathbf{y}_i\right) = \sum_{i=1}^{m} \left(\mathbf{Z}_i \mathbf{Q}_i^{-1} \mathbf{Z}_i^T + \mathbf{D}^{-1}\right)^{-1} \tag{4.6}$$

where $\mathbf{Q}_i$ is as defined in Section 3.2.2

Diggle *et al.* (1994, pp174-5) suggest that (4.4) and (4.5) are the limitations of PQL. In the case of (4.4), cov($\boldsymbol{\xi}$) is not constant but may be satisfactory if the GLM is not far from Gaussian. This would be the case with large sample sizes of Poisson variates which also have large mean parameters. Such a sample may be reliably approximated by the Normal distribution and $\frac{d\boldsymbol{\eta}}{d\boldsymbol{\mu}} \approx 1$. Equation (4.5) is possibly more vulnerable for data such as the ryegrass experiment since there are small samples per subject. The Gaussian distributions constructed from these samples may not match the data well and the mode calculated with Gaussian assumptions could be dissimilar to the mean it is intended to approximate.

Chapter 5

Error model diagnostics for Poisson GLMMs using the variogram

5.1 Introduction

In this thesis so far, we have discussed models where the nature of correlation could be assumed. In this chapter we examine diagnostics for selecting the correlation structure of a data set.

The variogram, which has its origins in geostatistics, has often been used as a tool for identifying error structures in linear repeated measures analysis (Diggle *et al.*, 1994). Its function is to describe the association amongst repeated measurements, and it can be used when the measurements are not equally spaced in time. In the context of examining error structures, the residuals are used to calculate the variogram.

The definition of a variogram can be represented as

$$\gamma(u) = \frac{1}{2}E\left[(\epsilon_{it} - \epsilon_{ij})^2\right], \tag{5.1}$$

where $u = |t - j|$ represents the lag between measurements on a subject. The variogram is the plot of $\gamma(u)$ against u. The lag may be a time or spatial separation.

The model for residuals containing all three sources of variation (random effects, serial correlation and measurement error) is given by equation (1.8). From this representation we can derive from this (see Appendix H) that the variogram for the model with only a random intercept or no random components at all will be

$$\gamma(u) = \sigma_M^2 + \sigma_{SC}^2\left[1 - \rho(u)\right] \tag{5.2}$$

The term σ_M^2 represents the variance of measurement error, σ_{SC}^2 is the variance of the serial correlation and $\rho(u)$ is the autocorrelation function.

To interpret a variogram for linear data we can use

- $\lim_{u\to\infty}\gamma(u) = \sigma_M^2 + \sigma_{SC}^2 +$ var(other random terms),
- $\gamma(0) = \sigma_M^2$, and the process variance
- $\text{var}(\epsilon_{it}) = \mathbf{z}_{it}^T\mathbf{D}\mathbf{z}_{it} + \sigma_M^2 + \sigma_{SC}^2$

The variogram will indicate which components of dispersion are present and their relative contributions to the model. An example variogram is given in Figure 5.1, where

$$\mathbf{D} = \begin{pmatrix} \nu_1^2 & 0 \\ 0 & \nu_2^2 \end{pmatrix},$$

ν_1^2 is the variance of the random subject intercept and ν_2^2 is the variance of the random subject slope effects.

The sample variogram is calculated from data by firstly calculating the residuals $r_{it} = y_{it} - \bar{y}_t$ (subject i, time t) and then averaging the half squared observed differences (v_{ijk}) for each appropriate lag (u_{ijk}) where

$$v_{ijk} = \frac{1}{2}(r_{ij} - r_{ik})^2, \quad \text{and} \tag{5.3}$$

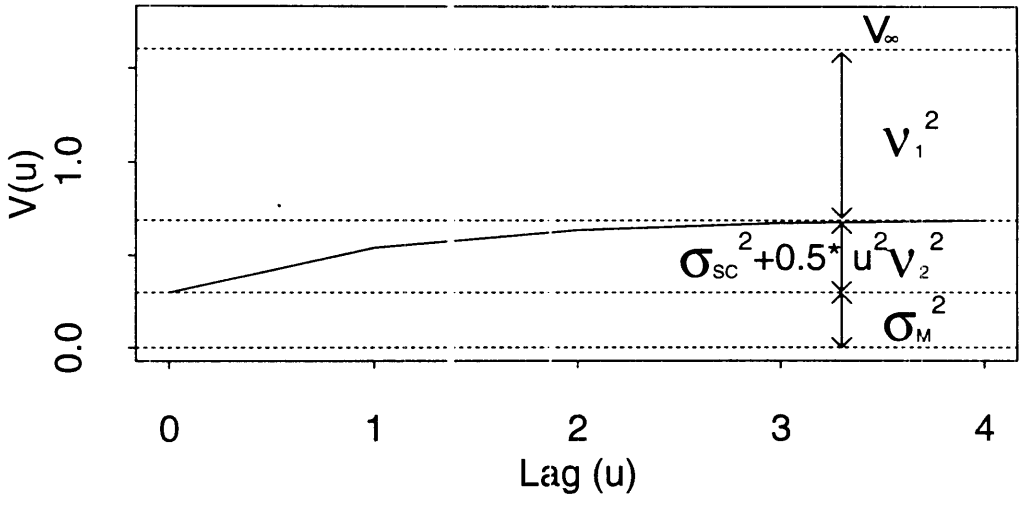
$$u_{ijk} = t_{ij} - t_{ik}. \tag{5.4}$$

The process variance (V_∞) is similarly calculated as the mean of

$$\frac{1}{2}(r_{ij} - r_{lk})^2 \tag{5.5}$$

where $i \neq l$.

Figure 5.1: Example variogram containing 3 error sources.



We are interested in extending the use of the variogram to repeated measures of count data as an aid in determining which error model is appropriate for any given

data set, in particularly whether the use of the random effects model is justified. We shall consider the use of the variogram for identifying error structures for both GLMMs and GEEs.

5.2 The variogram for detecting random effects and serial correlation in correlated Poisson models.

In the linear model, the variogram is calculated from the residuals after fitting the fixed effects at each sampling. In the case of count data, the initial Poisson based model is the naive GLM and we examine the resultant residuals for the error structure.

The adjusted dependent variable in the Poisson case is a linearized form of $\log(y)$ to the first order, that is,

$$\log(y) \simeq \log(\mu) + \frac{(y-\mu)}{\mu}, \tag{5.6}$$

where $\mu = \exp(x\boldsymbol{\beta})$. (McCullagh and Nelder, 1989, p40)

Schall (1991) exploits this linearization by assuming that linear methods such as BLUP and REML are satisfactory. One advantage of the linearization scheme discussed in section 3.2 is that it keeps the random effects on the same scale as the parameters, and if serial correlation were present it would similarly be kept on the same scale.

From (5.6) it can be seen that the "working" residuals $\left\{(y-\mu)\frac{\partial \eta}{\partial \mu}\right\}$ are on the

linearized scale and as such are the residuals which are required for the variogram.

We now show how residuals from the linearized scale can be used to interpret the error components.

Suppose the true underlying structure of the model conforms to

$$\log(y_{it}) = \mathbf{x}_{it}^T\boldsymbol{\beta} + \mathbf{z}_{it}^T\mathbf{U}_i + W_i(\tau_{it}) + M_i. \tag{5.7}$$

The first step is to construct a variogram of the residuals from the fit of a naive Poisson GLM, assuming only one error source and independence of data. In this case $\mu_{it} = \exp(\mathbf{x}_{it}^T\boldsymbol{\beta})$.

The "working" residuals for the Poisson model would be

$$\begin{aligned}\frac{(y_{it} - \hat{\mu}_{it})}{\hat{\mu}_{it}} &= \frac{\exp(\mathbf{x}_{it}^T\hat{\boldsymbol{\beta}})\{\exp[\mathbf{z}_{it}^T\mathbf{U}_i + W_i(\tau_{it}) + M_i] - 1\}}{\exp(\mathbf{x}_{it}^T\hat{\boldsymbol{\beta}})} \\ &= \mathbf{z}_{it}^T\mathbf{U}_i + W_i(\tau_{it}) + M_i + \quad \text{higher order terms.}\end{aligned} \tag{5.8}$$

If the higher order terms are negligible, the variogram of the "working" residuals will be an appropriate estimate of the magnitude of the random intercept and the other error components in a given set of data.

Another residual we could construct is

$$\begin{aligned} \omega_{it} &= \log(y_{it}) - \log(\hat{\mu}_{it}) \\ &= \mathbf{z}_{it}^T \mathbf{U}_i + W_i(\tau_{it}) + M_i \qquad (5.9) \\ &\simeq \frac{(y_{it} - \hat{\mu}_{it})}{\hat{\mu}_{it}} \end{aligned}$$

(5.10)

We note that the residuals ω_{it} given by (5.9) are approximately equal (first order) to the working residuals (5.8). But since the working residuals from the naive Poisson GLM potentially contain the random effects and serial correlation, they may be too inaccurate to be useful in the diagnosis of the error components.

5.3 Variograms from a simulation study

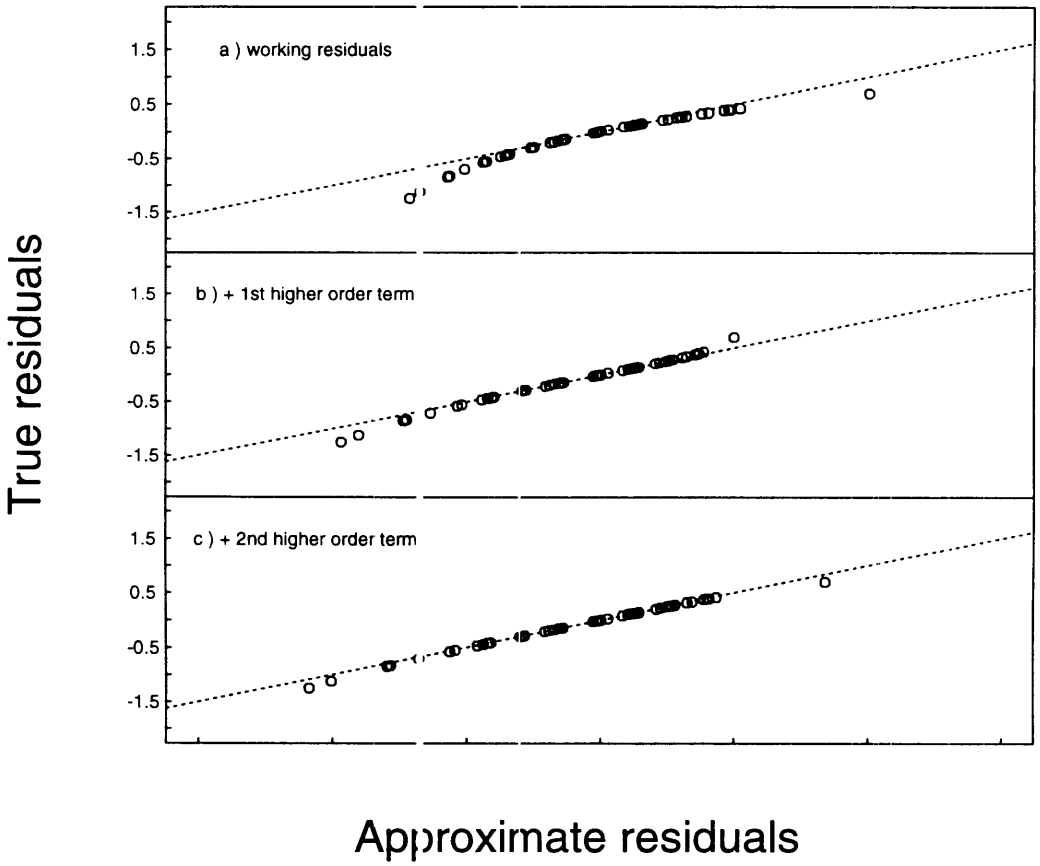
To determine the appropriateness of the variogram of "working" residuals (from the independence model) a small simulation study was conducted. The model simulated contained random effects and measurement error. There was a random intercept and slope which were distributed as $N(0, \mathbf{D})$, with

$$\mathbf{D} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.01 \end{pmatrix},$$

and the measurement error was distributed as $N(0, 0.25)$.

The first stage of the study was to calculate both the working residuals (5.8) and the alternative residuals (5.9) for each of the 10 simulated data sets to determine whether the higher order terms of the Taylor's expansion in (5.8) were negligible.

Figure 5.2: Comparison of two residuals in Simulation study (one set). The dotted line is 1:1.



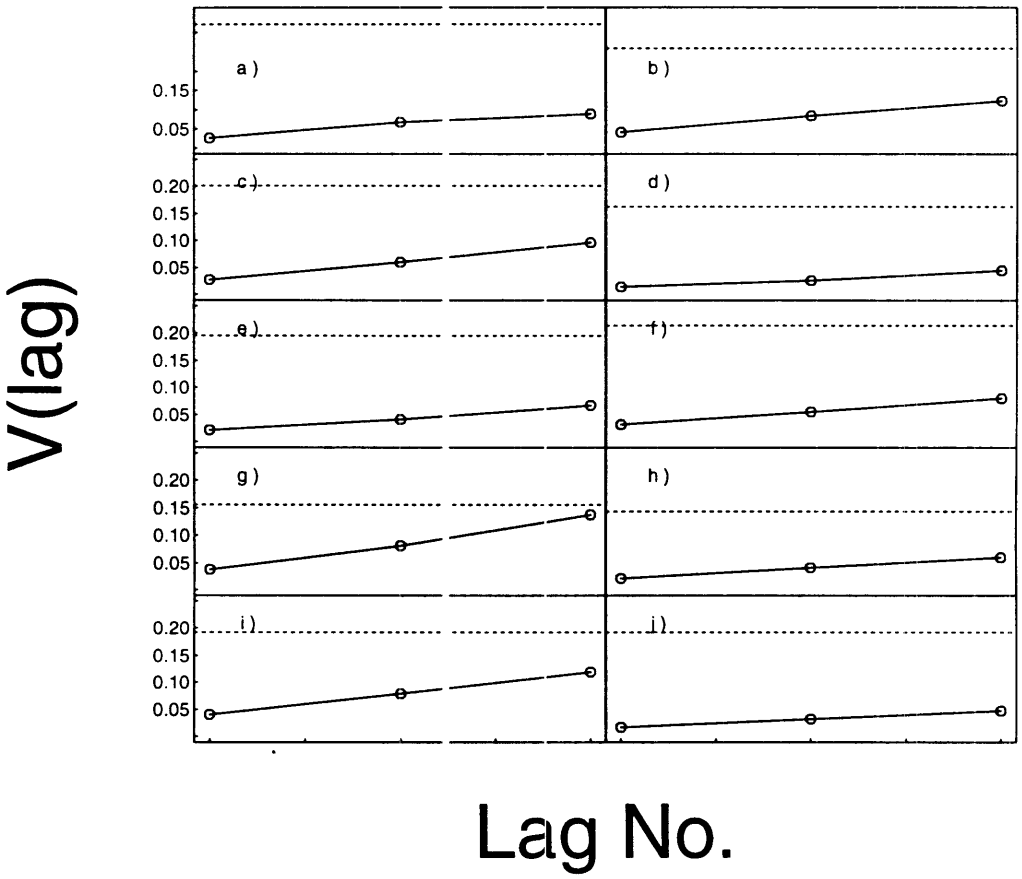
Plots of ω_{it} (termed true residuals) against the first order working residuals (Figure 5.2) given by the naive GLM show that these working residuals may be inappropriate for use in the variogram because the first order linear approximation discards important information. Successive plots where the higher order terms

$$\sum_{p=2,3} \frac{(-1)^{p-1}}{p} \cdot \left(\frac{y_{it} - \mu_{it}}{\mu_{it}} \right)^p$$

have been added to the linear working residuals show an improvement in fit over these

first order residuals. As a resul; we expect that the first order working residuals will not be as informative as the re;iduals (5.9) when used in a variogram.

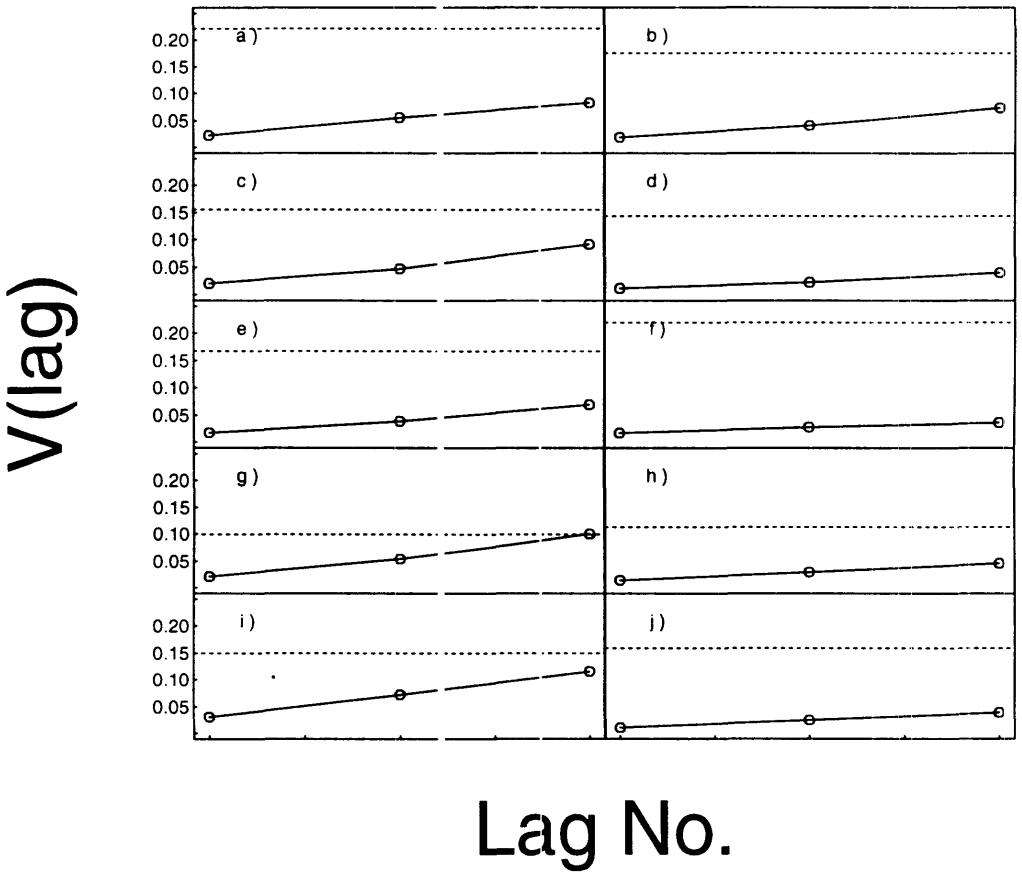
Figure 5.3: Simulatic n study variograms for true residuals (5.9).



The variograms for the residuals (5.9) in Figure 5.3 verify the use of these residuals from a naive Poisson GLM fit to determine appropriate error structures. All the variograms except g) would cle arly indicate at the investigative analysis stage that a random intercept model for the error structure, with possibly other random or serial correlation terms would be appropriate. By comparison with the generic variogram

(Figure 5.1), we deduce that the variance due to random intercepts is not negligable. Variogram g) would falsely indicate that the correlation in the data had been induced by some random effects other than an intercept or an autocorrelation function.

Figure 5.4: Simulation study variograms for working residuals.



The variograms for the linear predictors (Figure 5.4) show that the use of the working residuals to approximate residuals (5.9) from a naive Poisson GLM fit is potentially subject to an inaccuracy which can be misleading when trying to determine the appropriate error structures. Variograms a), b), d), e), f), h), and j) would still

clearly indicate (although often not as well as the variograms for residuals (5.9)) that a random intercept model for the error structure would be an appropriate starting place, due to the large gap between the variogram and the estimate of V_∞, with possible other random effects and/or autocorrelation functions being neccessary. Variogram c) would be inconclusive and variograms g) and i) would falsely indicate that the correlation in the data had not been induced by a random intercept. In practice, c) would probably be treated as a borderline case and examined more thoroughly at the exploratory analysis stage.

In linear modelling the variogram would usually be constructed using the residuals on the original scale (known as "response" or "raw" residuals). To demonstrate why this would be inappropriate in the Poisson case we can draw upon (5.7) and calculate the "raw" residuals from the naive GLM,

$$\begin{aligned} y_{it} - \hat{\mu}_{it} &= \exp(\mathbf{x}_{it}^T\hat{\boldsymbol{\beta}})\left(\exp[\mathbf{z}_{it}^T\mathbf{U}_i + W_i(\tau_{it}) + M_i] - 1\right) \\ &= \exp(\mathbf{x}_{it}^T\hat{\boldsymbol{\beta}})\left[\mathbf{z}_{it}^T\mathbf{U}_i + W_i(\tau_{it}) + M_i + \quad \text{higher order terms}\right]. \quad (5.11) \end{aligned}$$

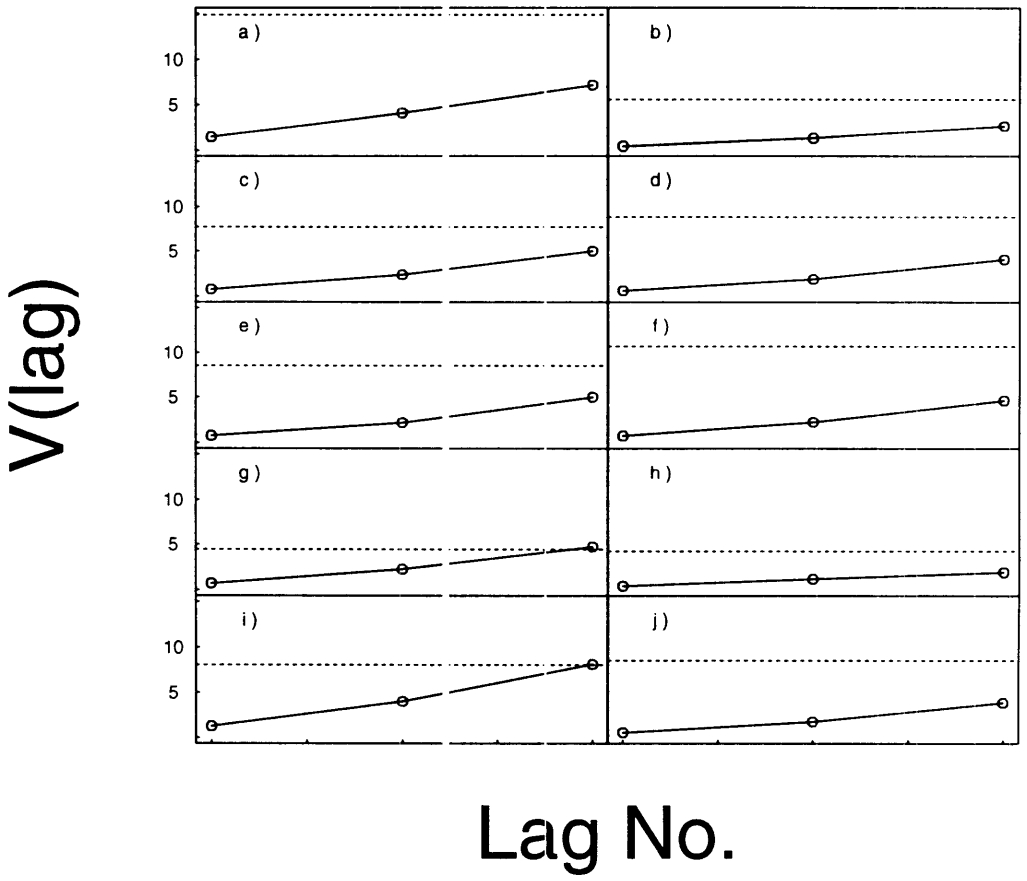
We can see from (5.11) that the magnitudes in the variogram of such "raw" residuals would have been affected by a scale change due to the multiplication of the errors by the mean calculated from the naive GLM. As a result, the variogram of these residuals may not be as simply used to identify the magnitude of the effect each error component has in the data.

This is confirmed by the variograms of the raw residuals given in Figure 5.5. Although these variograms would have basically indicated the same error components,

it was not as visually obvious which model to choose. This can be illustrated by comparing the variograms of data set j). In the variogram constructed from both residuals (5.9) and the working residuals it was quite evident that serial correlation was not a major contributor to the error term, and consequently the random intercept model (and possibly other random effects) would be appropriate. However, in the corresponding variogram yielded by the raw residuals, serial correlation and/or random effects other than an intercept appear to play a more significant role in the error structure of the data.

We also note, using data set i) as an example, that the appropriate error model has become more obscured as we move from residuals (5.9), to (5.8) and (5.11). In Figure 5.3 (residuals ω_{it}) it is clear that a random intercept model is appropriate, with possible other random terms or serial correlation. the same data set with a variogram of the working residuals (Figure 5.4) indicates that the possibility of a random intercept could be investigated as it is a borderline case. With the variogram of raw residuals (Figure 5.5) there is no indication of a random intercept being present in data set i). This decrease in sensitivity through the three residual types can be observed in the majority of the data sets, specifically b), c), d), e), f) (in which the random intercept became more prominent with the working residuals), h), i) and j).

Figure 5.5: Simulation study variograms for raw residuals.



The comparison of these variograms implies that in practice the use of working residuals, (5.8), to determine error structure from an initial independence GLM, in the Poisson case, is likely to direct the analysis appropriately. However, the residuals given by (5.9) are more sensitive to the underlying error structure and easy to construct, therefore making them adoptable in routine practice. The raw residuals (5.11) are too insensitive to the true error structure to be used for this purpose.

Appendix A

Iteratively weighted least squares

Our starting point is the likelihood equation:

$$l = \frac{[y\theta - b(\theta)]}{a(\phi)} + c(y;\phi) \tag{A.1}$$

We require estimates $\hat{\beta}$ which are obtained from starting values by the Newton-Rhapson procedure,

$$\hat{\beta}^{(j+1)} = \hat{\beta}^{(j)} + \frac{\partial l}{\partial \beta^{(j)}} \cdot E\left(-\frac{\partial^2 l}{\partial \beta_k^{(j)} \partial \beta_l^{(j)}}\right)^{-1} \tag{A.2}$$

The chain rule for differentiation:

$$\frac{\partial l}{\partial \beta} = \frac{\partial l}{\partial \theta} \cdot \frac{d\theta}{d\mu} \cdot \frac{d\mu}{d\eta} \cdot \frac{\partial \eta}{\partial \beta}$$

can be used to find the score statistic:

$$\frac{\partial l}{d\beta} = \frac{1}{a(\phi)}\left[\frac{(y-\mu)}{V} \cdot X \cdot \frac{d\mu}{d\eta}\right] \tag{A.3}$$

and the expected value of the second derivative is:

$$-\frac{\partial^2 l}{\partial \beta_k \partial \beta_l} = \frac{1}{a(\phi)} \cdot \left[X^T (\frac{d\mu}{d\eta}) V^{-1} (\frac{d\mu}{d\eta})^T X\right] \tag{A.4}$$

Substituting the values of equations (A.3) and (A.4) into equation (A.2) yields:

$$\hat{\beta}^{(j+1)} = \hat{\beta}^{(j)} + \left[\frac{1}{a(\phi)} \cdot X^T \left(\frac{du}{a\eta}\right) V^{-1} \left(\frac{d\mu}{d\eta}\right)^T X\right]^{-1} \cdot \frac{1}{a(\phi)} \left[\frac{(y-\mu)}{V} \cdot X \cdot \frac{d\mu}{d\eta}\right] \quad \text{(A.5)}$$

Thus estimation can be done by weighted least squares. We calculate the updated dependent variable,

$$z^{(j)} = \eta^{(j)} + (y - \mu^{(j)}) \cdot \frac{d\eta^{(j)}}{d\mu^{(j)}} \quad \text{(A.6)}$$

at the jth iteration, and define weights

$$w^{(j)} = \left(\frac{d\mu^{(j)}}{d\eta^{(j)}}\right)^2 \cdot V^{-1}(\mu^{(j)}) = DV^{-1}D \quad \text{(A.7)}$$

We can then calculate:

$$\hat{\beta}^{(j+1)} = [X^T(DV^{-1}D)X]^{-1}[X^T(DV^{-1}D)z^{(j)}] \quad \text{(A.8)}$$

or equivalently from A.5;

$$\hat{\beta}^{(j+1)} = \hat{\beta}^{(j)} + \frac{X^T(\frac{d\mu^{(j)}}{d\eta^{(j)}})\frac{1}{Var(\mu^{(j)})}(y - \mu^{(j)})}{X^T(\frac{d\mu^{(j)}}{d\eta^{(j)}})\frac{1}{Var(\mu^{(j)})}(\frac{d\mu^{(j)}}{d\eta^{(j)}})^T X} \quad \text{(A.9)}$$

Appendix B

Identities

Using the identity $E(\frac{\partial l}{\partial \theta}) = 0$ and $E(Y) = \mu$ then:

$$\begin{aligned} \frac{\partial l}{\partial \theta} = \quad \{\mu - \tfrac{\partial}{\partial \theta}[b(\theta)]/a(\phi)\} &= 0 \\ \mu - \tfrac{\partial}{\partial \theta}[b(\theta)] &= 0 \end{aligned}$$

So $E(Y) = \mu = \frac{\partial}{\partial \theta}[b(\theta)]$.

Using $E(\frac{\partial^2 l}{\partial \theta^2}) + E(\frac{\partial l}{\partial \theta})^2 = 0$ implies:

$$\begin{aligned} -\frac{\partial^2}{\partial \theta^2}/a(\phi) + E\left[\{\mu - \frac{\partial}{\partial \theta}[b(\theta)]\}/a(\phi)\right]^2 &= 0 \\ -\frac{\partial^2}{\partial \theta^2}/a(\phi) + E\{[\mu - E(Y)]/a(\phi)\}^2 &= 0 \\ -\frac{\partial^2}{\partial \theta^2}/a(\phi) + \frac{Var(Y)}{a(\phi)^2} &= 0 \\ Var(Y) &= a(\phi) \cdot V(\mu) \end{aligned}$$

where $V(\mu)$ is the variance function of μ.

Appendix C

Derivation of $E(\mathbf{y})$ and var$(\mathbf{y})$ in Poisson GLMM

The Poisson glmm is given by:

$$\log(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} \tag{C.10}$$

Now, the expected value of $\mathbf{y}$ under this model is:

$$\begin{aligned}
E(\mathbf{y}) &= E[E(\mathbf{y} \mid \mathbf{U})] \\
&= \int_{-\infty}^{\infty} \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U}) \cdot (2\pi)^{-m/2} \cdot \mid \mathbf{D} \mid^{-1/2} \cdot \exp\left(\frac{-1}{2}\mathbf{U}^T\mathbf{D}^{-1}\mathbf{U}\right) d\mathbf{U} \\
&= (2\pi)^{-m/2} \cdot \mid \mathbf{D} \mid^{-1/2} \cdot \exp(\mathbf{X}\boldsymbol{\beta}) \int_{-\infty}^{\infty} \exp\left(\mathbf{Z}\mathbf{U} - \frac{1}{2}\mathbf{U}^T\mathbf{D}^{-1}\mathbf{U}\right) d\mathbf{U} \\
&= (2\pi)^{-m/2} \cdot \mid \mathbf{D} \mid^{-1/2} \cdot \exp(\mathbf{X}\boldsymbol{\beta}) \cdot \left\{\prod^{m} \left(\frac{\pi}{\mathbf{D}^{-1} \cdot \frac{1}{2}}\right)^{1/2} \cdot \exp\left[\mathbf{Z}\mathbf{D}\mathbf{Z}^T / \left(4 \cdot \frac{1}{2}\right)\right]\right\} \\
&= \exp\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{D}\mathbf{Z}^T/2\right)
\end{aligned} \tag{C.11}$$

The expected value of $\mathbf{y}^2$ is needed:

$$\begin{aligned} E\left(\mathbf{y}^2\right) &= E\left[E\left(\mathbf{y}^2 \mid \mathbf{U}\right)\right] \\ &= \int_{-\infty}^{\infty} \exp\left(2\mathbf{X}\boldsymbol{\beta} + 2\mathbf{ZU}\right) \cdot (2\pi)^{-m/2} \cdot \mid \mathbf{D} \mid^{-1/2} \cdot \exp\left(\frac{-1}{2}\mathbf{U}^T\mathbf{D}^{-1}\mathbf{U}\right) d\mathbf{U} \\ &= (2\pi)^{-m/2} \cdot \mid \mathbf{D} \mid^{-1/2} \cdot \exp\left(2\mathbf{X}\boldsymbol{\beta}\right) \int_{-\infty}^{\infty} \exp\left[-\left(\frac{1}{2}\mathbf{U}^T\mathbf{D}^{-1}\mathbf{U} - 2\mathbf{ZU}\right)\right] d\mathbf{U} \\ &= (2\pi)^{-m/2} \cdot \mid \mathbf{D} \mid^{-1/2} \cdot \exp\left(2\mathbf{X}\boldsymbol{\beta}\right) \cdot (2\pi)^{m/2} \cdot \mid \mathbf{D} \mid^{-1/2} \cdot \exp\left[4\mathbf{ZDZ}^T / \left(4 \cdot \frac{1}{2}\right)\right] \\ &= \exp\left(2\mathbf{X}\boldsymbol{\beta} + 2\mathbf{ZDZ}^T\right) \end{aligned} \tag{C.12}$$

From (C.11) and (C.12) we can calculate;

$$\begin{aligned} \text{var}\left[E\left(\mathbf{y} \mid \mathbf{U}\right)\right] &= E\left[E\left(\mathbf{y}^2 \mid \mathbf{U}\right)\right] - \left\{E\left[E\left(\mathbf{y} \mid \mathbf{U}\right)\right]\right\}^2 \\ &= \exp\left(2\mathbf{X}\boldsymbol{\beta} + 2\mathbf{ZDZ}^T\right) - \exp\left(2\mathbf{X}\boldsymbol{\beta} + \mathbf{ZDZ}^T\right) \\ &= \exp\left(2\mathbf{X}\boldsymbol{\beta}\right)\left[\exp\left(2\mathbf{ZDZ}^T\right) - \exp\left(\mathbf{ZDZ}^T\right)\right] \end{aligned} \tag{C.13}$$

As we are using the Poisson distribution of the GLMM:

$$E\left[\text{var}\left(\mathbf{y} \mid \mathbf{U}\right)\right] = \exp\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{ZU}\right) \tag{C.14}$$

From (C.14) and (C.13) we calculate the variance of $\mathbf{y}$ as:

$$\begin{aligned} \text{var}\left(\mathbf{y}\right) &= E\left[\text{var}\left(\mathbf{y} \mid \mathbf{U}\right)\right] + \text{var}\left[E\left(\mathbf{y} \mid \mathbf{U}\right)\right] \\ &= \exp\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{ZU}\right) + \exp\left(2\mathbf{X}\boldsymbol{\beta}\right)\left[\exp\left(2\mathbf{ZDZ}^T\right) - \exp\left(\mathbf{ZDZ}^T\right)\right] \end{aligned} \tag{C.15}$$

Appendix D

Derivation of log-likelihood and score functions in Poisson GLMM

Working from the marginal likelihood given by (3.12) we can calculate the log-likelihood as:

$$
\begin{aligned}
l &= \log L \\
&= \log \left\{ \prod_{i=1}^{m} \int_{Ru} \exp \left[\sum_{t=1}^{n_i} y_{it} \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) - \sum_{t=1}^{n_i} \exp \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) \right] \right. \\
&\qquad \left. \prod_{t=1}^{n_i} \left[\mid \mathbf{D} \mid^{-1/2} (2\pi)^{-q/2} \exp \left(-\frac{1}{2} \mathbf{U}_i^T \mathbf{D}^{-1} \mathbf{U}_i \right) \right] d\mathbf{U}_i \right\} \\
&= \sum_{i=1}^{m} \log \left\{ \int_{Ru} \exp \left[\sum_{t=1}^{n_i} y_{it} \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) - \sum_{t=1}^{n_i} \exp \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) \right] \right. \\
&\qquad \left. \prod_{t=1}^{n_i} \left[\mid \mathbf{D} \mid^{-1/2} (2\pi)^{-q/2} \exp \left(-\frac{1}{2} \mathbf{U}_i^T \mathbf{D}^{-1} \mathbf{U}_i \right) \right] d\mathbf{U}_i \right\}
\end{aligned}
$$

$$= \sum_{i=1}^{m} \log \left\{ \int_{Ru} \exp \left[\sum_{t=1}^{n_i} y_{it} \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) - \sum_{t=1}^{n_i} \exp \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) \right] f(\mathbf{U}_i; \mathbf{D}) d\mathbf{U}_i \right\} \tag{D.16}$$

From (D.16) we can develop the following score equations:

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{m} \left\{ \frac{\frac{d}{d\boldsymbol{\beta}} \int_{Ru} \exp \left[\sum_{t=1}^{n_i} y_{it} \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) - \exp \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) \right] f(\mathbf{U}_i; \mathbf{D}) d\mathbf{U}_i}{\int_{Ru} \exp \left[\sum_{t=1}^{n_i} y_{it} \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) - \exp \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) \right] f(\mathbf{U}_i; \mathbf{D}) d\mathbf{U}_i} \right\}$$

$$= \sum_{i=1}^{m} \left\{ \frac{\int_{Ru} \frac{d}{d\boldsymbol{\beta}} \left(\exp \left[\sum_{t=1}^{n_i} y_{it} \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) - \exp \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) \right] \right) f(\mathbf{U}_i; \mathbf{D}) d\mathbf{U}_i}{\int_{Ru} \exp \left[\sum_{t=1}^{n_i} y_{it} \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) - \exp \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) \right] f(\mathbf{U}_i; \mathbf{D}) d\mathbf{U}_i} \right\} \tag{D.17}$$

where $f(\mathbf{U}_i; \mathbf{D})$ is defined by (D.24).

Now $\frac{\partial l}{\partial \boldsymbol{\beta}}$ will equal zero if the numerator in (D.17) is equal to zero. To show this we focus only on parts involving $\boldsymbol{\beta}$. Expanding the derivative in the numerator we have

$$\frac{d}{d\boldsymbol{\beta}} \left(\exp \left[\sum_{t=1}^{n_i} y_{it} \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) - \exp \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) \right] \right)$$

$$= \left\{ \frac{d}{d\boldsymbol{\beta}} \left[\sum_{t=1}^{n_i} y_{it} \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) - \exp \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) \right] \right.$$
$$\left. \exp \left[\sum_{t=1}^{n_i} y_{it} \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) - \exp \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) \right] \right\}$$

$$= \left\{ \left[\sum_{t=1}^{n_i} y_{it} \mathbf{x}_{it}^T - \mathbf{x}_{it} \exp \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) \right] \right.$$
$$\left. \exp \left[\sum_{t=1}^{n_i} y_{it} \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) - \exp \left(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{U}_i \right) \right] \right\} \tag{D.18}$$

The term $\exp\left[\sum_{t=1}^{n_i} y_{it}\left(\mathbf{x}_{it}^T\boldsymbol{\beta}+\mathbf{z}_{it}^T\mathbb{U}_i\right)-\exp\left(\mathbf{x}_{it}^T\boldsymbol{\beta}+\mathbf{z}_{it}^T\mathbf{U}_i\right)\right]$ cannot be equal to zero and so $\frac{\partial l}{\partial\boldsymbol{\beta}}=0$ implies

$$S_{\boldsymbol{\beta}|y,U} = \sum_{i=1}^{m}\left\{\sum_{t=1}^{n_i}\mathbf{x}_{it}\left[y_{it}-\exp\left(\mathbf{x}_{it}^T\boldsymbol{\beta}+\mathbf{z}_{it}^T\mathbf{U}_i\right)\right]\right\}=0 \tag{D.19}$$

The score equation for $\mathbf{D}$ are derived using the following identities (see Hogg and Craig (1995)).

$$f_Y(\ ) = \int_{Rx} f_{XY}(x,y)\quad dx \tag{D.20}$$

$$f_{Y|X}(y\mid\ ) = \frac{f_{XY}(x,y)}{f_X(x)} \tag{D.21}$$

$$E[h(Y)\mid\ ] = \int_{Ry} h(Y)f_{Y|X}(y\mid x)\quad dy \tag{D.22}$$

and the assumptions of the model that given $\mathbf{U}_i$, the repeated measurements y_{it} on subject i are independent.

The conditional density of y_{it}, given $\mathbf{U}_i$ is

$$f_{Y|U}(y_{it}\mid\mathbf{U}_i;\boldsymbol{\beta})=C\cdot\exp\left[\sum_{t=1}^{n_i}y_{it}\left(\mathbf{x}_{it}^T\boldsymbol{\beta}+\mathbf{z}_{it}^T\mathbf{U}_i\right)-\exp\left(\mathbf{x}_{it}^T\boldsymbol{\beta}+\mathbf{z}_{it}^T\mathbf{U}_i\right)\right] \tag{D.23}$$

where C is a normalizing constant.

The density of the q dimensional random effects $\mathbf{U}_i$ is

$$f_U(\mathbf{U}_i;\mathbf{D})=(2\pi)^{-q/2}\mid\mathbf{D}\mid^{-1/2}\exp\left(-\frac{1}{2}\mathbf{U}_i^T\mathbf{D}^{-1}\mathbf{U}_i\right), \tag{D.24}$$

We derive

$$\frac{\partial l}{\partial \mathbf{D}} = \sum_{i=1}^{m} \left(\frac{\int_{Ru} n_i \left\{ \frac{d}{dD} \left[\mid \mathbf{D} \mid^{-1/2} \exp\left(-\frac{1}{2} \mathbf{U}_i^T \mathbf{D}^{-1} \mathbf{U}_i \right) \right] \right\} \prod_{t=1}^{n_i} f_{Y|U}(y_{it} \mid \mathbf{U}_i; \boldsymbol{\beta}) (2\pi)^{-q/2} d\mathbf{U}_i}{\int_{Ru} \prod_{t=1}^{n_i} f_{Y|U}(y_{it} \mid \mathbf{U}_i; \boldsymbol{\beta}) f_U(\mathbf{U}_i; \mathbf{D}) d\mathbf{U}_i} \right) \tag{D.25}$$

as C is a constant and hence cancels top and bottom in (D.25).

Now we determine

$$\begin{aligned} \frac{d}{d\mathbf{D}} \left[\mid \mathbf{D} \mid^{-1/2} \exp\left(-\frac{1}{2} \mathbf{U}_i^T \mathbf{D}^{-1} \mathbf{U}_i \right) \right] &= \left\{ \left(\frac{d}{d\mathbf{D}} \mid \mathbf{D} \mid^{-1/2} \right) \left[\exp\left(-\frac{1}{2} \mathbf{U}_i^T \mathbf{D}^{-1} \mathbf{U}_i \right) \right] \right. \\ &\quad \left. + \left[\frac{d}{d\mathbf{D}} \exp\left(-\frac{1}{2} \mathbf{U}_i^T \mathbf{D}^{-1} \mathbf{U}_i \right) \right] \left(\mid \mathbf{D} \mid^{-1/2} \right) \right\} \\ &= \left\{ -\frac{1}{2} \mathbf{D}^{-1} \mid \mathbf{D} \mid^{-1/2} \exp\left(-\frac{1}{2} \mathbf{U}_i^T \mathbf{D}^{-1} \mathbf{U}_i \right) \right. \\ &\quad \left. + \frac{1}{2} \mathbf{D}^{-1} \mathbf{U}_i \mathbf{U}_i^T \mathbf{D}^{-1} \exp\left(-\frac{1}{2} \mathbf{U}_i^T \mathbf{D}^{-1} \mathbf{U}_i \right) \mid \mathbf{D} \mid^{-1/2} \right\} \end{aligned} \tag{D.26}$$

Using (D.26) we obtain

$$\frac{\partial l}{\partial \mathbf{D}} = \sum_{i=1}^{m} \left[\frac{\int_{Ru} n_i \left(\frac{1}{2} \mathbf{D}^{-1} \mathbf{U}_i \mathbf{U}_i^T \mathbf{D}^{-1} - \frac{1}{2} \mathbf{D}^{-1} \right) \prod_{t=1}^{n_i} f_{Y|U}(y_{it} \mid \mathbf{U}_i; \boldsymbol{\beta}) f_U(\mathbf{U}_i; \mathbf{D}) d\mathbf{U}_i}{\int_{Ru} \prod_{t=1}^{n_i} f_{Y|U}(y_{it} \mid \mathbf{U}_i; \boldsymbol{\beta}) f_U(\mathbf{U}_i; \mathbf{D}) d\mathbf{U}_i} \right] \tag{D.27}$$

where the normalizing constants have cancelled each other. The above result uses the assumption of conditional independence of subjects, and using this assumption further the denominator becomes

$$\int_{Ru} f_{Y|U}(y_{i\cdot} \mid \mathbf{U}_i) f_{\mathbf{U}}(\mathbf{U}_i) \quad d\mathbf{U}_i = f_Y(y_{i\cdot}). \tag{D.28}$$

With $\mathbf{U}_i$ integrated out we can include this density in the numerator and obtain

$$\frac{\partial l}{\partial \mathbf{D}} = \sum_{i=1}^{m} \left[\int_{Ru} n_i \left(\frac{1}{2} \mathbf{D}^{-1} \mathbf{U}_i \mathbf{U}_i^T \mathbf{D}^{-1} - \frac{1}{2} \mathbf{D}^{-1} \right) \frac{f_{Y|U}(y_{i\cdot}) f_U(\mathbf{U}_i)}{f_Y(y_{i\cdot})} \quad d\mathbf{U}_i \right]$$

$$= \sum_{i=1}^{m} \left[\int_{Ru} n_i \left(\frac{1}{2}\mathbf{D}^{-1}\mathbf{U}_i\mathbf{U}_i^T\mathbf{D}^{-1} - \frac{1}{2}\mathbf{D}^{-1} \right) f_{U|Y}(\mathbf{U}_i \mid y_{i\cdot}) \quad d\mathbf{U}_i \right]$$

$$= \sum_{i=1}^{m} \left[\frac{n_i}{2} E\left(\mathbf{D}^{-1}\mathbf{U}_i\mathbf{U}_i^T\mathbf{D}^{-1} - \mathbf{D}^{-1} \right) \right], \qquad \text{using (D.22)}$$

$$= \sum_{i=1}^{m} \left[\frac{n_i}{2}\mathbf{D}^{-1} E\left(\mathbf{U}_i\mathbf{U}_i^T \right) \mathbf{D}^{-1} - \frac{n_i}{2}\mathbf{D}^{-1} \right] \tag{D.29}$$

From (D.29) we get the score equation for $\mathbf{D}$

$$S_{D|y,U} = \frac{1}{2}\mathbf{D}^{-1} \sum_{i=1}^{m} E\left(\mathbf{U}_i\mathbf{U}_i^T \right) \mathbf{D}^{-1} - \frac{m}{2}\mathbf{D}^{-1} = 0 \tag{D.30}$$

Appendix E

Deriving the likelihood for Poisson mixed with Gamma distribution

Given

$$f(\gamma_i \mid \alpha_i) = \frac{\exp(-\alpha_i\gamma_i)\alpha_i^{\alpha_i}\gamma_i^{(\alpha_i-1)}}{\Gamma(\alpha_i)}, \quad \text{and}$$

$$f(y_{it} \mid \gamma_i, \alpha_i) = \frac{(\gamma_i\mu_{it})^{y_{it}}\exp(-\gamma_i\mu_{it})}{y_{it}!}$$

then we obtain the likelihood as

$$L = \prod_{i=1}^{m}\int_{R_\gamma}\prod_{t=1}^{n_i} f(\gamma_i \mid \alpha_i) f(y_{it} \mid \gamma_i, \alpha_i)d\gamma_i$$

$$= \prod_{i=1}^{m}\prod_{t=1}^{n_i}\int_0^\infty \left[\frac{\mu_{it}^{y_{it}}\alpha_i^{\alpha_i}}{y_{it}!\Gamma(\alpha_i)}\right]\gamma_i^{y_{it}}\exp[-\gamma_i(\mu_{it}+\alpha_i)]d\gamma_i$$

$$= \prod_{i=1}^{m} \left\{ \frac{(\prod_{t=1}^{n_i} \mu_{it}^{y_{it}})\, n_i \alpha_i^{\alpha_i}}{(\prod_{t=1}^{n_i} y_{it}!)\, n_i \Gamma(\alpha_i)} \prod_{t=1}^{n_i} \int_0^{\infty} \gamma_i^{y_{it}} \exp[-\gamma_i(\mu_{it} + \alpha_i)] d\gamma_i \right\}$$

$$= \prod_{i=1}^{m} \left\{ \left(\frac{\prod_{t=1}^{n_i} \mu_{it}^{y_{it}}}{\prod_{t=1}^{n_i} y_{it}!} \right) \left(\frac{\alpha_i^{\alpha_i}}{\Gamma(\alpha_i)} \right) \left[\frac{\prod_{t=1}^{n_i} \Gamma(\alpha_i + y_{it})}{\prod_{t=1}^{n_i} (\alpha_i + \mu_{it})^{y_{it} + \alpha_i}} \right] \right\}$$

$$= \prod_{i=1}^{m} \left\{ \left[\frac{\Gamma(\alpha_i + \sum_{t=1}^{n_i} y_{it})}{\Gamma(\alpha_i)(\prod_{t=1}^{n_i} y_{it}!)} \right] \left(\frac{\alpha_i}{\alpha_i + \sum_{t=1}^{n_i} \mu_{it}} \right)^{\alpha_i} \prod_{t=1}^{n_i} \left(\frac{\mu_{it}}{n_i \alpha_i + \sum_{t=1}^{n_i} \mu_{it}} \right)^{y_{it}} \right\}$$

$$= \prod_{i=1}^{m} \left[\frac{\Gamma(\alpha_i + y_{i\cdot})}{\Gamma(\alpha_i)\,(\prod_{t=1}^{n_i} y_{it}!)} \left(\frac{\alpha_i}{\alpha_i + \mu_{i\cdot}} \right)^{\alpha_i} \prod_{t=1}^{n_i} \left(\frac{\mu_{it}}{\alpha_i + \mu_{i\cdot}} \right)^{y_{it}} \right]$$

where $x_{i\cdot} = \sum_{t=1}^{n_i} x_{it}$.

Appendix F

Simulation procedure for correlated Poisson data (Exchangeable structure)

This Genstat procedure was written by S. Harden and R. Murison, NSW Agriculture Tamworth, to simulate correlated Poisson data using the algorithm of Sim [29].

```
   PROCEDURE 'POISMEANS'
PARAMETER NAME='RHO',  "Input:  exchangeable correlation"
'BETACONS', "Input:  constant term in L.P. "
'BETATREAT', "Input:  "
'BETATIME', "Input:  "
'SIG','P','LAMBDA','MU';MODE=p;
TYPE='SCALAR','SCALAR','SCALAR','VARIATE','MATRIX','SCALAR',
```

```
'VARIATE','VARIATE'


SCALAR a,b,dum;VALUE=0,0,0

CALC notime=NOBSERVATIONS(BETATIME)+1

VARIATE[NVALUES=notime]LAMBDA

CALC LAMBDA$[1]=EXP(BETACONS+BETATREAT)

FOR I =2...#notime

CALC Ii=I-1

CALC LAMBDA$[I]=EXP(BETACONS+BETATREAT+BETATIME$[Ii])

ENDFOR


CALC P=NOBSERVATIONS(LAMBDA)


VARIATE[NVALUES=P] MU

CALC MU=LAMBDA

DIAGONALMATRIX [ROWS=P] DSig;VALUES=MU

MATRIX [ROWS=P;COLUMNS=P] SIG

CALC SIG = DSig / 2


FOR I = 2...#P

CALC IM1 = I-1

FOR J = 1...IM1
```

```
CALC SIG$[I;J] = RHO*SQRT(4*SIG$[I;I]*SIG$[J;J])
ENDFOR
ENDFOR

CALC SIG = SIG + TRANSPOSE(SIG)

ENDPROCEDURE

OPEN 'poismeans.proc';CHANNEL=2;FILETYPE=backingstore
STORE[CHANNEL=2;SUBFILE=POISMEANS;METHOD=o;PROCEDURE=yes]POISMEANS
CLOSE CHANNEL=2;FILETYPE=backingstore

PROCEDURE 'POISSIM'
PARAMETER NAME='NEWZ', "Output:the simulated Poisson data"
'SIG', "Input:  Matrix with means and covariances"
'P', "Input:Number of repeated measures"
'LAMBDA', "Input:  "
'FLAG', "Output:  1 if successful, 0 if not"
'LV', "Input:  length of each variable"
'RANSED' "Input:  random number seed"
;MODE=p;
TYPE='VARIATE','MATRIX','SCALAR',
```

```
'VARIATE','SCALAR','SCALAR','SCALAR'

CALC FLAG=1

CALC P2p = P*P

MATRIX [ROWS=P;COLUMNS=P;\ALUES=#P2p(1)] alpha

CALC PL = LV*P

MATRIX [ROWS=LV;COLUMNS=P;VALUES=#PL(0)] Z,x

CALC LAMBDA$[1] = SIG$[1;1]

FOR J = 2...#P

CALC alpha$[J;1] = SIG$[J;1]/LAMBDA$[1]

IF alpha$[J;1].LT.0.OR.alpha$[J;1].GT.1

CALC FLAG=0

PRINT !T(ERROR alpha)

EXIT[CONTROL=procedure]

ENDIF

IF J.GT.2

CALC JM1 = J-1
```

```
FOR I = 2...JM1

CALC IM1 = I-1

CALC P1[1...IM1] = alpha$|I;1...IM1]*alpha$[J;1...IM1]*LAMBDA$[1...IM1]

CALC a = 0

FOR K = 1...IM1

CALC a = a + P1[K]

ENDFOR


IF a.LT.SIG$[J;I].AND.SIG$[J;I].LT.(LAMBDA$[I]+a)

CALC alpha$[J;I] = (SIG$[.;I]-a)/LAMBDA$[I]

ELSE

CALC FLAG=0

PRINT !T(ERROR_a)

EXIT[CONTROL=procedure]

ENDIF


ENDFOR "FOR I = 2...JM1"

ENDIF "IF J.GT.2"


CALC JM1 = J-1

CALC P2[1...JM1] = alpha$[J;1...JM1]*LAMBDA$[1...JM1]

CALC b = 0
```

```
FOR K = 1...JM1

CALC b = b + P2[K]

ENDFOR "FOR K=1...JM1"


IF SIG$[J;J].GT.b

CALC LAMBDA$[J] = SIG$[J;J]-b

ELSE

CALC FLAG=0

PRINT !T(ERROR_b)

EXIT[CONTROL=procedure]

ENDIF "IF SIG$[J;J].GT.b"


ENDFOR "FOR J = 2...#p"


GRANDOM[DISTRIBUTION=Poisson;MEAN=LAMBDA$[1];

VARIANCE=LAMBDA$[1];NVALUES=LV;SEED=RANSED]prans


FOR K = 1...LV

CALC x$[K;1]=prans$[K]

ENDFOR "K = 1...LV"

CALC Z=x
```

```
FOR J=2...#P

VARIATE[NVALUES=LV]alx

CALC alx=0

CALC IM=J-1


CALC rsed=RANSED

FOR I=1...IM

FOR K=1...LV

CALC nbnt=x$[K;I]

FOR L=1...nbnt

CALC valp=alpha$[J;I]*(1-alpha$[J;I])


IF valp.GT.0

CALC rsed=rsed+L

GRANDOM[DISTRIBUTION=binomial;MEAN=alpha$[J;I];VARIANCE=valp;

NVALUES=1;SEED=rsed]rbin

CALC alx$[K]=alx$[K]+rbin

ELSE

CALC FLAG=0

PRINT !T(ERROR_variance_binomial)

EXIT[CONTROL=procedure]

ENDIF "About binomial"
```

```
ENDFOR "FOR L=1...nbnt"


ENDFOR "FOR K=1...lv"

ENDFOR "FOR I=1...IM"


CALC prsed2=RANSED+101

GRANDOM[DISTRIBUTION=Poisson;MEAN=LAMBDA$[J];

VARIANCE=LAMBDA$[J];NVALUES=LV;SEED=prsed2]prans


FOR K=1...LV

CALC x$[K;J]=prans$[K]

CALC Z$[K;J]=alx$[K]+x$[K;J]

ENDFOR "K=1...LV"


ENDFOR "J=1...#P"


VARIATE nlen

CALC nlen=P*LV

VARIATE[NVALUES=#nlen]NEWZ

CALC NEWZ$[1...#nlen]=Z$[(1...#LV)#P;#LV(1...#P)]
```

```
ENDPROCEDURE


OPEN 'poisim.proc';CHANNEL=2;FILETYPE=backingstore

STORE[CHANNEL=2;SUBFILE=POISSIM;METHOD=o;PROCEDURE=yes]POISSIM

CLOSE CHANNEL=2;FILETYPE=backingstore


ENDJOB

STOP
```

Appendix G

Splus routine for simulation study using EQL methodology

The following program was written in Splus code to perform the EQL technique on the simulations conducted in Chapter 4. Comments are provided to help link the program to the terminology used in Van de Ven and Weber [32].

```
   #Setting up contrasts.
options(contrasts=c("contr.treatment","contr.poly"))
# Matrix of 0's to put 50 coefficient estimates for each parameter.
sim.betas <- matrix(0,50,5)

# Read in data and declare factors.
sim1.df <- read.table("geese.gendata",header=T)
sim1.df$trt <- as.factor(sim1.df$trt)
```

```
sim1.df$repm <- as.factor(sim1.df$repm)


# Iterate process from data set 1 to 50.
for(w in 1:50){


# Temporary data frame of simulation data set w.
temp.df <- sim1.df[sim1.df$simno==w,]


# len is the number of measurements in total, len2 is the number of subjects.
len <- length(temp.df[,1])
len2 <- length(unique(temp.df$idno))


# Using glm() to get design matrix of fixed effects, XX.
glm.model <- glm(YY repm+trt,family=poisson,data=temp.df,x=T)
XX <- glm.model$x


# smui represents \sum_{j=1}^{n_i} \mu_{ij} = \mu_{i.}.
smui <- rep(0,len)
Yvar <- as.matrix(temp.df$YY)


# Calculate initial parameter values.
halfsies <- rep(0.5,len)
```

```
TT <- mui <- numeric(0)

AA <- matrix(0,len,len)

syi <- rep(0,len)


# syi represents \sum_{j=1}^{n_i} y_{ij} = y_{i.}.

for(i in 1:len2){

syi[temp.df$idno==i] <- sum(temp.df$YY[temp.df$idno==i]) }


# Initial starting values for \mu and \log(\mu).  TT and mui represent T_{ij} and

\mu_{ij}.

for(i in 1:len){

mui[i] <- max(temp.df$YY[i],halfsies[i])

TT[i] <- log(mui[i]) }

mui <- as.matrix(mui)

# AA represents A, a diagonal matrix of \mu's.

diag(AA) <- mui


Xt <- Iwt <- Tt <- list()

# Xt, Iwt and Tt are list forms of matrices X_i, A_i and T_i.

# For extra comments on this loop see main body of program (j loop).

for(i in 1:len2){

a.temp <- AA[temp.df$idno==i,temp.df$idno==i]
```

```
Iwt <- c(Iwt,list(a.temp))
x.temp <- XX[temp.df$idno==i,]
if(length(temp.df$idno[temp.df$idno==i])==1)
x.temp <- t(as.matrix(x.temp))
Xt <- c(Xt,list(x.temp))
t.temp <- TT[temp.df$idno==i]
Tt <- c(Tt,list(t.temp)) }

parts1.beta <- parts2.beta <- list()
# Calculate initial estimates of fixed effects,
```

$\hat{\beta_{(0)}} = (\sum_{i=1}^{m} X_i^T \hat{W}_i^{-1} X_i)^{-1} \sum_{i=1}^{m} X_i^T \hat{W}_i^{-1} T_i$ where $\hat{W}_i^{-1} = A_i$ (Iwt).

```
for(i in 1:len2){
parts1.temp <- t(Xt[[i]])%*%Iwt[[i]]%*%Xt[[i]] # X_i^T A_i X_i
parts2.temp <- t(Xt[[i]])%*%Iwt[[i]]%*%Tt[[i]] # X_i^T A_i T_i
parts1.beta <- c(parts1.beta,list(parts1.temp)) # Make into lists.
parts2.beta <- c(parts2.beta,list(parts2.temp)) }

len3 <- dim(XX)[2] # Number of fixed effect coefficients, p.
p1.beta <- matrix(0,len3,len3) # p x p matrix of 0's.
p2.beta <- matrix(0,len3,1) # p x 1 matrix of 0's.
for(i in 1:len2){
p1.beta <- p1.beta+parts1 beta[[i]] # sum_{i=1}^m X_i^T A_i X_i, a p x p matrix.
```

```
p2.beta <- p2.beta+parts2.beta[[i]] # ∑_{i=1}^m X_i^T A_i T_i, a p x 1 matrix.  }
p1.beta <- solve(p1.beta) # (∑_{i=1}^m X_i^T A_i X_i)^{-1}
B0 <- p1.beta%*%p2.beta # (∑_{i=1}^m X_i^T A_i X_i)^{-1} ∑_{i=1}^m X_i^T A_i T_i, a p x 1 matrix, β̂_(0).
mui <- as.matrix(exp(XX%*%B0)) # New estimates of μ_ij.

for(i in 1:len2){
smui[temp.df$idno==i] <- sum(mui[temp.df$idno==i]) }
smui <- as.matrix(smui)

smu.alp <- sy.alp <- numeric(0)
# In this loop find unique vector of y_i. and μ_i..
for(i in 1:len2){
sy.alp[i] <- syi[temp.df$idno==i][1]
smu.alp[i] <- smui[temp.df$idno==i][1] }

# Initial estimate of η̂_1(0).
# exp(η̂_1(0)) = (∑_{i=1}^m μ̂_i.)^2 / ((∑_{i=1}^m (y_i. - μ̂_i.))^2 - ∑_{i=1}^m μ̂_i.).
eta0 <- (sum(smu.alp))^2 / (sum((sy.alp-smu.alp)^2)-sum(smu.alp))
eta0 <- log(eta0)
eta0 <- as.matrix(eta0)

# Z1 is the design matrix of the random effects.
```

```
Z1 <- matrix(1,len2,1)

# AK and BK represent A_{K+1} and B_{K+1}.

AK <- BK <- matrix(0,len2,len2)

BK.temp <- numeric(0)

bk.1st <- rep(0,len2)

LK <- rep(0,len2)


# Use initial estimates of β and η.

Beta <- B0

eta <- eta0

# Number of iterations to be carried out.

niter <- 10


for(j in 1:niter){

# μ_ij's and α_i's (exp(z_i^T η)) recalculated at the start of each iteration.

mui <- as.matrix(exp(XX%*%Beta))

alphai <- exp(ZZ%*%eta)


# X_i, S_i, A_i and V_i given in list form.

Xt <- St <- AA <- VV <- list()

for(i in 1:len2){

# Forming A_i = diag(μ_i1,.. ,μ_in_i).
```

```
a.temp <- mui[temp.df$idnc==i]

if(length(a.temp)>1){

a.temp <- diag(a.temp)}

AA <- c(AA,list(a.temp))

# Forming design matrix for fixed effects X_i.

x.temp <- XX[temp.df$idno==i,]

if(length(temp.df$idno[temp.df$idno==i])==1)

x.temp <- t(as.matrix(x.temp))

Xt <- c(Xt,list(x.temp))

# Forming S_i = y_i - mu_i.

s.temp <- Yvar[temp.df$idno==i]-mui[temp.df$idno==i]

St <- c(St,list(s.temp)) # Calculate 1/alpha_i.

invalp <- 1 / alphai[temp.df$idno==i][1]

# Calculate V_i = A_i + mu_i mu_i^T / alpha_i.

v.temp <- AA[[i]]+mui[temp.df$idno==i]%*%t(mui[temp.df$idno==i])*invalp

# Let v.temp = V_i^{-1}.

v.temp <- solve(v.temp)

VV <- c(VV,list(v.temp)) }


parts1.beta <- parts2.beta <- list()

for(i in 1:len2){

# Calculate X_i^T A_i V_i^{-1} A_i X_i.
```

```
parts1.temp <- t(Xt[[i]])%*%AA[[i]]%*%VV[[i]]%*%AA[[i]]%*%Xt[[i]]
parts2.temp <- t(Xt[[i]])%*%AA[[i]]%*%VV[[i]]%*%St[[i]] # X_i^T A_i V_i^{-1} S_i.
parts1.beta <- c(parts1.beta,list(parts1.temp)) # Form lists.
parts2.beta <- c(parts2.beta,list(parts2.temp)) }

p1.beta <- matrix(0,len3,len3) # p x p matrix of 0's.
p2.beta <- matrix(0,len3,1) # p x 1 matrix 0's.
for(i in 1:len2){
p1.beta <- p1.beta+parts1.beta[[i]] # sum_{i=1}^m X_i^T A_i V_i^{-1} A_i X_i.
p2.beta <- p2.beta+parts2.beta[[i]] # sum_{i=1}^m X_i^T A_i V_i^{-1} S_i }
p1.beta <- solve(p1.beta) # (sum_{i=1}^m X_i^T A_i V_i^{-1} A_i X_i)^{-1}.
# beta_new = beta_old + (sum_{i=1}^m X_i^T A_i V_i^{-1} A_i X_i)^{-1} (sum_{i=1}^m X_i^T A_i V_i^{-1} S_i).
Beta <- Beta+p1.beta%*%p2.beta

# Recalculate estimates of mu_ij to use in estimating eta_new.
mui <- as.matrix(exp(XX%*%Beta))
for(i in 1:len2){
smui[temp.df$idno==i] <- sum(mui[temp.df$idno==i]) # mu_i.'s.  }

BK.temp <- numeric(0)
bk.1st <- rep(0,len2)
```

```
smu.alp <- sy.alp <- uni.alp <- numeric(0)
for(i in 1:len2){
uni.alp[i] <- alphai[temp.df$idno==i][1] # α_i.
sy.alp[i] <- syi[temp.df$idno==i][1] # y_i..
smu.alp[i] <- smui[temp.df$idno==i][1] # μ_i..  }

# AK is A_{K+1} = diag(α_1,...,α_K).  In this case α_i = α  ∀i.
diag(AK) <- uni.alp

# This loop calculates B^+_{K+1}.  See pp210-211 of paper.
for(i in 1:len2){
bk.1st[i] <- 1 / uni.alp[i]*(
log((uni.alp[i]+sy.alp[i]) / uni.alp[i])+
sy.alp[i] / (2*uni.alp[i]*(uni.alp[i]+sy.alp[i]))+
log(uni.alp[i] / (uni.alp[i]+smu.alp[i]))+
(smu.alp[i]-sy.alp[i]) / (smu.alp[i]+uni.alp[i])
)+
(smu.alp[i]^2+uni.alp[i]*sy.alp[i]) /
(uni.alp[i]*(smu.alp[i]+uni.alp[i])^2)-
sy.alp[i] / (uni.alp[i]*(uni.alp[i]+sy.alp[i]))-
(sy.alp[i]*(2*uni.alp[i]+sy.alp[i])) /
(2*uni.alp[i]^2*(uni.alp[i]+sy.alp[i])^2)
```

```
bk.1st[i] <- -1*bk.1st[i] }

bk.1st <- as.matrix(bk.1st)

BK.temp <- bk.1st

diag(BK) <- BK.temp


# LK is l+K+1.  See pp210-211 of paper.

LK <- rep(0,len2)

for(i in 1:len2){

LK[i] <- log((uni.alp[i]+sy.alp[i]) / uni.alp[i])+

sy.alp[i] / (2*uni.alp[i]*(uni.alp[i]+sy.alp[i]))+

log(uni.alp[i] / (uni.alp[i]+smu.alp[i]))+

(smu.alp[i]-sy.alp[i]) / (2*smu.alp[i]+uni.alp[i]) }

LK <- as.matrix(LK)


# η̂new = η̂old + (Z^T A_{K+1} B+_{K+1} A_{K+1} Z)^{-1} Z^T A_{K+1} l+_{K+1}.

eta <- eta+solve(t(Z1)%*%AK%*%BK%*%AK%*%Z1)%*%t(Z1)%*%AK%*%LK


} #end iterative loop j.


# Making a matrix of estimates for β with each of the 50 data sets.

sim.betas[w,] <- Beta
```

```
} # end simulation loop w.


# Results from the 50 data sets are in the matrix sim.betas.
sim.betas <- round(sim.betas,4)
dimnames(sim.betas) <- list(NULL,c("Constant","repm2","repm3","repm4",
"Treatment"))
```

Appendix H

Derivation of variogram

The formula for the error term containing all three sources of variation used is:

$$\epsilon_{it} = \mathbf{z}_{it}^T \mathbf{U}_i + W_i(\tau_{it}) + M_i \tag{H.31}$$

From (H.31) we can derive

$$\text{var}(\epsilon_{it}) = \mathbf{z}_{it}^T \mathbf{D} \mathbf{z}_{it} + \sigma_{SC}^2 + \sigma_M^2 \tag{H.32}$$

and using the independence property of measurement errors

$$\text{cov}(\epsilon_{it}, \epsilon_{ij}) = \mathbf{z}_{it}^T \mathbf{D} \mathbf{z}_{ij} + \sigma_{SC}^2 \rho(u) \tag{H.33}$$

where u represents the time lag between measurements.

The variogram can then be represented as

$$\frac{1}{2} E\left[(\epsilon_{it} - \epsilon_{ij})^2\right] = \frac{1}{2} E\left(\epsilon_{it}^2 + \epsilon_{ij}^2 - 2\epsilon_{it}\epsilon_{ij}\right)$$

$$= \frac{1}{2}\left[\text{var}(\epsilon_{it}) + \text{var}(\epsilon_{ij}) - 2\text{cov}(\epsilon_{it}, \epsilon_{ij})\right]$$

$$= \frac{1}{2}\left\{\mathbf{z}_{it}^T\mathbf{D}\mathbf{z}_{it} + \mathbf{z}_{ij}^T\mathbf{D}\mathbf{z}_{ij} + 2(\sigma_{SC}^2 + \sigma_M^2) - 2[\mathbf{z}_{it}\mathbf{D}\mathbf{z}_{ij}^T + \sigma_{SC}^2\rho(u)]\right\} \tag{H.34}$$

If the model contains only a random intercept, the variogram then becomes:

$$\gamma(u) = \sigma_M^2 + \sigma_{SC}^2[1 - \rho(u)] \tag{H.35}$$

However, if other random effects, such as slope, are present in the data these effects will also be contained in the variogram.

References

1. Aitkin, M, Anderson, D, Francis. B and Hinde, J (1989). *Statistical Modelling in GLIM.* Oxford University Press. Oxford.

2. Box, G.E.P and Cox, D.R (1964). An analysis of transformations. *Journal of the Royal Statistical Society B*, **26**, 211-252.

3. Box, G.E.P and Tiao, G.C (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley. Reading, MA.

4. Breslow, N.E and Clayton, D.G (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9-25.

5. Carroll, R.J and Ruppert D (1988). *Transformation and weighting in regression.* Chapman and Hall. New York.

6. Crowder, M (1995). On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika*, **82**, 407-410.

7. Diggle, P.J, Liang, K and Zeger, S.L (1994). *Analysis of Longitudinal Data.* Oxford University Press. Oxford

8. Dobson, A.J (1990). *An Introduction to Generalized Linear Models.* Chapman and Hall. London.

9. Fellner, W.H (1986). Robust estimation of variance components. *Technometrics*, **28**, 51-60.

10. Geman, S and Geman, D (1984). Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intellegence* **6**, 721-741.

11. Genstat 5 Committee (1993). *Genstat 5 Release 3 Reference Manual.* Clarendon Press. Oxford.

12. Editors; Gilks, W.R, Richardson, S and Spiegelhalter, D.J (1996). *Markov Chain Monte Carlo in practice.* Chapman and Hall. London.

13. Harville, D.A (1977). Maximum likelihood approaches to variance component estimation and related problems. *Journal of the American Statistical Society*, **72**, 320-340.

14. Hinde, J (1982). *Compound Poisson regression models in GLIM 82: Proceedings of the International Conference on Generalized Linear Models (ed R. Gilchrist), pp109-121.* Springer. Berlin.

15. Hogg, R.V and Craig, A T (1995). *Introduction to Mathematical Statistics.* Prentice-Hall. New Jersey.

16. Kuk, A.Y (1995). Asymptotically unbiased estimation in generalized linear models with random effects. *Journal of the Royal Statistical Society B*, **57**, 395-407.

17. Laird, N.M and Ware, J.H (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963-974.

18. Liang, K and Zeger, S.L (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

19. Lin, X and Breslow, N.E (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Society*, **91**, 1007-1016.

20. McCullagh, P and Nelder, J.A (1989). *Generalized Linear Models.* Chapman and Hall. New York.

21. McDermott, J.J and Schukken, Y.H (1994). A review of methods used to adjust for cluster effects in explanatory epidemiological studies of animal populations. *Preventitive Veterinary Medicine*, **18**, 155-173.

22. Morgan, B.J.T (1984). *Elements of Simulation.* Chapman and Hall. Cambridge.

23. Nelder, J.A and Lee, Y (1994). Double Generalized Linear Models. *A statistical conference for Genstat users*, **Wagga Wagga**, Australia.

24. Lee, Y and Nelder, J.A (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society B*, **58**, 619-78.

25. Nelder, J.A and Wedderburn, R.W.M (1972). Generalized linear models. *Journal of the Royal Statistical Society A*, **135**, 370-84.

26. Odell, P.L and Feiveson, A.H (1966). A numerical procedure to generate a sample covariance matrix *Journal of the American Statistical Society*, **61**, 198-203.

27. Royall, R.M (1986). Model robust inference using maximum likelihood estimators. *International Statistical Review*, **54**, 221-226.

28. Schall, R (1991). Estimation in generalized linear models with random effects. *Biometrika*, **78**, 719-727.

29. Sim, C.H (1993). Generation of Poisson and gamma random vectors with given marginals and covariance matrix. *Journal of Statistical Computing and Simulation*, **47**, 1-10.

30. Spiegelhalter, D, Thomas. A, Best, N and Gilks, W (1996). *BUGS test version 0.5* . MRC Biostatistics Unit, Instititute of Public Health. Cambridge.

31. Stiratelli, R, Laird, N, and Ware. J.H (1984). Random-effects models for serial observations with binary response. *Biometrics*, **40**, 961-971.

32. Van de Ven, R. and Weber, N.C (1995). Log-linear models for mean and dispersion in mixed Poisson regression models. *Australian Journal of Statistics*,

37(2), 205-216.

33. Zeger, S.L and Karim, M.R (1991). Generalized linear models with random effects; A Gibbs sampling approach. *Journal of the American Statistical Association*, **86**, 79-86.

34. Zhao, L.P and Prentice, R.L (1990). Correlated binary regression using a quadratic exponential mo lel. *Biometrika*, **77**, 642-648.