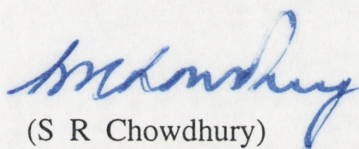# ANALYSIS OF GENERALISED MIXED MODELS

FOR

CATEGORICAL DATA

SADEQUR RAHMAN CHOWDHURY

A thesis submitted for the degree of Doctor of Philosophy
of The Australian National University

June 1998

Unless otherwise stated the research reported in this thesis is my own original work performed at The Australian National University.

(S R Chowdhury)

ACKNOWLEDGEMENTS

ABSTRACT

The development of generalised linear mixed models (GLMM) over the last decade or so has extended generalised linear modelling (GLM) techniques (McCullagh and Nelder, 1989) to incorporate both fixed and random effects in a model. The development has unified the approaches to deal with a wide class of statistical problems such as overdispersion, shrinkage estimation, correlated errors and similar other problems under the common framework of GLMMs. The thesis looks at some theoretical and applied aspects of GLMMs for analysing categorical data.

After introducing the problem and reviewing the literature in the first two chapters it starts with an investigation of some properties of the GLMM estimators. Approximate moments are developed for residual maximum likelihood estimators by utilising the known moment properties of the derivatives of likelihood functions. This development offers an alternative set of estimators for the second order moments of the variance components. The applicability of the development to a number of problems is discussed.

Hierarchical generalised linear models (HGLM), another recent development, allows the distribution of the variance components to be non-normal unlike GLMMs. A comparative study of HGLMs and GLMMs has been undertaken by using a Poisson response variable.

The application of GLMMs to a number of categorical data problems is discussed and specific application strategies are developed. The analysis of matched case control studies with random exposure effects is shown to be more appropriate under the GLMM framework. The analysis of contingency tables with clustered observations is another such application. The methods of analysis with GLMMs have been developed under both Poisson and multinomial sampling assumptions. A general strategy for dealing with overdispersed multi-category response data is also presented.

The applicability of threshold modelling to the analysis of contingency tables with ordered categories is discussed. The strategy has been extended

to cover the situation where not only are the categories ordered but also the observations are clustered.

An approximate method for adjusting the standard errors of the fixed parameters in the absence of a full GLMMs based analysis is presented. This can be applied to any of the above mentioned applications to reduce the risk of misleading inferences in a fixed effect analysis.

The proposed methods are applied to various real datasets and the results from simulation studies are presented where appropriate.

CONTENTS

CHAPTER ONE

INTRODUCTION

1.1 MIXED MODELS

The use of regression models to study the inter relationships among variables is a common phenomenon in socio-economic and biomedical research and the related theory has occupied a substantial part of the statistical literature. In a model the coefficients which relate the dependent variable with the explanatory variables are called parameters. These parameters are often assumed fixed implying that if the whole population could be observed then the actual value of a parameter can be obtained. However in many applications it also makes sense to assume the parameters are random. That means even in the population these values are not fixed and can vary randomly from one to the next realisation of the population. For instance, in some applications the set of parameters in a model may be a sub-set of a bigger set of parameters while an inference is required for the whole set and not for the sub-set in hand. In that case it is more appropriate to assume the parameters in the model are random rather than fixed. Consequently, depending on the nature of the parameters, the models are referred to as fixed effects or random effects. Models which include both fixed and random parameters are called mixed effect models or mixed models.

Although the term 'mixed models' was first introduced by Eisenhart (1947) the use of mixed models in the form of variance components dates back as early as the 1860s. Airy (1861) and Chauvenet (1863) implicitly used variance component models in their work in astronomy. However the real evolution of the linear mixed model was initiated with the work of Fisher

1

(1918, 1925) in genetic modelling and the development of the analysis of variance (ANOVA) technique. This process of evolution continued and received real impetus during the 1950s and the 1960s when more interest was focused on the estimation of variance components, particularly in relation to the construction of a selection index in genetics (Henderson 1950, 1953, 1963; Rao 1952). Since then the theory and application of mixed models have been considerably developed.

Mixed models are found to be a very useful tool to analyse data that are correlated due to single or multi-level clustering or some form of hierarchical structure. Datasets with clustering or hierarchical relationships are very common in practice. For example, data on students from different classes in different schools, patients from different clinics and animals in different litters are naturally correlated at respective levels of grouping. Similar structures are introduced in datasets collected through multistage sample surveys. In longitudinal studies repeated observations on the same individual are another example of correlated data. Mixed models are used very extensively in genetics in assessing the genetic merits of animals. The models can be used to separate fixed effects such as age, sex or breed of animals from random genetic effects. The predictions of random effects can then be used for selecting animals in a breeding program.

1.2 GENERALISED MIXED MODELS

The application of mixed models is not restricted only to the case of continuous response variables with normally distributed errors. During the last couple of decades there has been considerable research in applying

mixed models to discrete data. Following the framework of generalised linear modelling (GLM), as in McCullagh and Nelder (1989), the theory of mixed models has been extended to cover a variety of discrete response variables with non-normal error distributions. The general framework of the extension of GLM to mixed models is called generalised linear mixed models (GLMMs). The generalised framework can be applied to the modelling of either continuous or discrete response data with clustered or hierarchical structure. In epidemiological or biometric studies, the variable of interest is often discrete and in many cases correlated or clustered. In such situations GLMMs have proved to be very useful analysis tools. GLMMs are also found to be useful in accommodating overdispersion often observed in binomial (Williams, 1982) and Poisson (Breslow, 1984) regression models.

However, the use of GLMMs for discrete data is not as widespread as for continuous data for a number of reasons. The method of estimation for discrete data is more complicated than that in the continuous case. The absence of related fully developed asymptotic theory for inference; absence of appropriate modifications required for applications to specific types of discrete data and the unavailability of appropriate software are some of the main reasons preventing the widespread use of GLMMs. In recent years several authors such as Anderson and Aitkin (1985), Breslow (1984), Morton (1988), Stiratelli et al. (1984) and Williams (1982) have investigated the techniques of fitting GLMMs. Influenced by the above work Schall (1991), Breslow and Clayton (1993) and McGilchrist (1994) have developed methods of fitting GLMMs in a more general framework. Although each of these methods use a somewhat different rationale, the methods are substantially in agreement with one another.

1.3 RESEARCH OBJECTIVES

As indicated above the method of estimation in GLMMs is still not completely developed. There is scope for further improvements particularly in relation to the properties of the GLMM estimators and inferential techniques. Also there is a need to analyse the GLMM framework with respect to its applicability to various types of categorical data. The aims of the thesis are: to further investigate the method of estimation in GLMMs; research the properties of the estimators; and analyse and develop the appropriate modifications required for applying the method to some specific types of categorical data which are frequently observed in applied research. Applications to matched case-control studies, analysis of contingency tables with nominal and ordinal categories, and application to multi-category response data are some of these examples.

The research will mainly concentrate on the approach to estimation used by McGilchrist (1994) which uses the best linear unbiased prediction (BLUP) methods. BLUP was first proposed by Henderson (1963, 1973, 1975) to develop approximate maximum likelihood (ML) and residual maximum likelihood (REML) estimators. The approach is based on an approximate linearisation of the model and hence has the potential to apply to wide ranging problems.

1.4 OUTLINES OF CHAPTERS

After setting the background and the broad objective of the research in Chapter 1, an extensive review of the literature related to the estimation of variance components and fitting mixed models is presented in Chapter 2. Because of the difference in the problem of estimation in mixed models for

continuous and discrete response variables the review of literature has been separated accordingly. The first part of Chapter 2 reviews estimation methods associated with a continuous response variable with normal error model, while the second part extends the review to the case where the response variable is discrete with non-normal error distribution.

In Chapter 3, approximate second order moments of GLMM estimators are developed by utilising the connection between REML estimators and known moment properties of the derivatives of the components of penalised likelihood functions. Expressions for estimators of variance components and other parameters are then obtained based on the moments. The expressions provide an alternative strategy for estimating the variance of the dispersion parameters. The applications to specific cases such as generalised mixed models, random component hazard models and threshold models are discussed. The GLMM strategy is then extended to a non-normal random component. This is done for a response variable with Poisson distribution and the random component is assumed to be gamma distributed. The chapter concludes by presenting the results from a simulation study using Poisson-gamma model.

Although the theory has been developed assuming the distribution of the random term is either normal or some other appropriate distribution, no study has so far investigated if the difference in distributional assumptions has any impact on estimates. Poisson-normal and Poisson-gamma distributions are compared and presented in Chapter 4.

Chapter 5 presents a strategy to analyse data from matched case-control studies where exposure effects are random over matched sets. This involves applying GLMMs framework to the conditional likelihood based analysis and deriving expressions for estimation equations. An efficient computing method for dealing with studies involving large numbers of matched sets is also developed. A simulation study is undertaken and the method is applied to a number of real datasets.

In Chapter 6, it is demonstrated how GLMMs can be used to analyse contingency tables where observations are clustered. The method presented in the chapter is developed under Poisson assumptions. The performance of the method is evaluated by using a benchmark dataset called Neighbourhood data (Brier 1980) which has been used by some other authors (Fingleton 1984) to investigate the analysis of clustered contingency tables.

A strategy for analysing a multi-category response variable under GLMMs is developed in Chapter 7. The proposed strategy is applied to a dataset called 'hamsters birth defects study' and the results are compared with existing methods of analysis for multinomial response. The method is then extended to the analysis of contingency tables with clustered observations under product multinomial assumptions. The neighbourhood dataset mentioned above is tested for independence under this approach and compared with the results obtained under the method presented in the previous chapter.

The methods of analysis presented in Chapters 6 and 7 are appropriate for contingency tables with nominal categories. In Chapter 8, a method for analysing contingency tables with ordinal categories is developed based on

threshold models (McCullagh, 1980). It is shown that the method is more efficient than those discussed in Chapters 6 and 7 when the categories are ordinal as the number of parameters to be estimated is fewer due to the utilisation of additional information of the ordered relationship among categories. The method is extended to cover the situation where not only are categories ordinal but also observations are clustered. This is done by incorporating random cluster effects in threshold models. A number of contingency tables with both independent and correlated observations are analysed by using this approach. As the categories of the Neighbourhood dataset are ordinal this is reanalysed here to compare the result under the mixed threshold model with those of other methods that do not utilise the ordered relationship.

In Chapter 9 a method is presented to approximately adjust the standard errors (SEs) of the coefficients in a fixed effect model in the absence of a full mixed model analysis. In most practical applications interest is mainly focused on the estimates of fixed parameters rather than the predictions of random effects. However, if the estimates are obtained by using a fixed effect model ignoring random effects then the main problem appears to be the underestimation of the standard errors (SEs) of the coefficients. The estimates of the coefficients also suffer from bias but to a much lesser extent than the SEs, implying that in most cases the risk of making a wrong inference can be reduced considerably by adjusting only the SEs. This strategy can be useful when an applied researcher is not capable of undertaking a full mixed model analysis because of the absence of appropriate software or expertise. It can also be useful in a situation

where the number of random effects to be predicted is extremely large and it may not be easy to undertake a mixed model analysis.

In Chapter 10, a general discussion of the research is presented. Possible areas of further research are also discussed.

The datasets used in various chapters are presented in Appendix A. DYALOG APL version 7.1 is used for all computing work in the thesis. Appendix B presents the relevant APL programs used in various simulations and analysis undertaken.

CHAPTER TWO

REVIEW : ESTIMATION IN MIXED MODELS

2.1 INTRODUCTION

The estimation of variance components has been a major problem since the beginning of the mixed model analysis. However, not much progress had been made in developing estimating techniques until the 1950s. Previously, the main contribution had come from the development of the analysis of variance (ANOVA) technique by Fisher (1918, 1925). This was further extended, notably by Tippett (1931), Daniels (1939), Winsor and Clarke (1940) and Crump (1946), to the estimation of variance components. The next wave of research on the estimation problem started with the publication of a paper by Henderson (1953) which introduced three different methods for estimating variance components from unbalanced data. Subsequently, the issue has drawn extensive interest and a wide variety of estimation methods have been developed. It is not intended to give an extensive review of the development of the methods of estimating variance components in this chapter. A detailed history of the development can be found in Searle et al. (1992). Khuri and Sahai (1985) also provides a good summary of the more recent work particularly in relation to continuous response variables. A recent paper by Robinson (1991) includes an extensive bibliography.

The history of the development of estimating variance components for discrete response variables on the other hand is not so old. The interest in the use of mixed models for discrete data arose mainly in the 1980s, particularly with the development of generalised modelling techniques (McCullagh and Nelder, 1989). As the focus of the thesis is on discrete

response variables the related estimation issues will be discussed in detail later in the chapter. By contrast the discussion of continuous response variables will include only those methods which are related to the estimation of generalised linear mixed models.

An introduction to the standard notation and general assumptions can be made by defining a linear mixed model

(2.1) $y = X\beta + Zu + e$

where y is an $n \times 1$ vector of observed responses and e is an $n \times 1$ vector of random errors. In the case of a continuous response variable the distribution of e is assumed to be $N(0, \sigma^2 D)$ where D is a known matrix. β is a $p \times 1$ vector of unknown fixed parameters corresponding to the known $n \times p$ matrix X of explanatory variables. The random component Zu can be partitioned conformably into $Z=[Z_1, Z_2, ..., Z_k]$ and $u=[u_1', u_2', ..., u_k']'$ where u_j is a $v_j \times 1$ vector of random effects with incidence matrix Z_j. Each u_j is assumed to be distributed as $N[0, \sigma_j^2 A_j(\rho)]$ where $\rho=(\rho_1, \rho_2, ..., \rho_s)'$, a parameter which describes the covariance structure of the vectors u_j. The inclusion of the covariance parameter ρ is a recent development. If $\sigma_j^2 = \sigma^2 \phi_j$ then

$$(2.2) \qquad A = \begin{bmatrix} \phi_1 A_1(\rho) & 0 & \cdots & 0 \\ 0 & \phi_2 A_2(\rho) & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & \phi_k A_k(\rho) \end{bmatrix}$$

and $Var(u) = \sigma^2 A$.

Therefore, the assumptions in the model are

$$E(\mathbf{u}_j) = \mathbf{0} \text{ and } \text{Var}(\mathbf{u}_j) = \sigma_j^2 \mathbf{A}_j(\rho), \quad \text{for } j=1,2,...,k$$

$$\text{Cov}(\mathbf{u}_j,\mathbf{u}_{j'}) = \mathbf{0}, \text{ for } j \neq j'=1,2,...,k$$

(2.3)

$$E(\mathbf{e}) = \mathbf{0} \text{ and } \text{Var}(\mathbf{e}) = \sigma^2 \mathbf{D}; \text{ Cov}(\mathbf{u}_j,\mathbf{e}) = \mathbf{0}$$

$$E(\mathbf{y}) = \mathbf{X}\beta, \text{ and } E(\mathbf{y}\,|\,\mathbf{u}) = \mathbf{X}\beta + \mathbf{Z}\mathbf{u}$$

$$\text{Var}(\mathbf{y}) = \mathbf{V} = \sigma^2(\mathbf{D} + \mathbf{Z}\mathbf{A}\mathbf{Z}') = \sigma^2 \Sigma$$

2.2 METHODS FOR CONTINUOUS RESPONSE VARIABLES

2.2.1 The Analysis of Variance Method

As mentioned above the ANOVA method was implicitly introduced by Fisher (1925). Later Tippett (1931) used the method explicitly to estimate variance components from balanced data with 2-way cross-classifications. The essence of the method is to calculate the mean squares under the fixed effect model and then equate these to the expected values under mixed or random effect models to derive the estimators of variance components. Under the assumption of normality of the error terms and other assumptions as mentioned in (2.3), the distribution of the estimators are obtained as linear functions of multiples of χ^2-variables. Detailed discussions of the ANOVA method can be found in Searle (1971) and in Searle et al. (1992).

For balanced data, the ANOVA method is still the most widely accepted method. The ANOVA estimators are unbiased for balanced data and have the smallest variance of all estimators that are quadratic functions of observations and are unbiased. Consequently, they are minimum variance quadratic unbiased (MVQU) estimators. Under the assumption of normality the estimators are minimum variance unbiased (MVU). However, the major disadvantage of the method is that the estimates can turn out to be negative

11

which is theoretically not acceptable. One of the suggestions is that the negative estimate is an indication that the variance component is negligible and should be treated as zero. Further discussion of this can be found in Searle (1971).

The performance of the ANOVA method for estimation of variance components from unbalanced data is not as good as for balanced data. The properties of unbiasedness and minimum variance as mentioned above do not hold in the unbalanced case. The seminal paper by Henderson (1953) proposed three modifications of the basic ANOVA method for dealing with unbalanced data. Method 1 is essentially the application of the strategy used for balanced data but the method is not applicable to mixed models and it provides biased estimates for random effect models. Method 2 uses data adjusted for fixed effects and then obtains the estimates of variance components. It does provide unbiased estimates but it is unable to handle any interactions between fixed and random effects. Method 3 uses reduction in sums of squares due to fitting a fixed effect model and various sub-models of it. The variance components are estimated by equating each computed reduction to its expected value under the full model. The estimators under Method 3 are unbiased and the method can be applied to a model with interactions between fixed and random effects. Among the disadvantages of the method is that the estimates are not unique when there is more than one interaction term in the model.

2.2.2 Best Linear Unbiased Prediction (BLUP)

Robinson (1991) provides an extensive review of BLUP and its justification using various other approaches. BLUP is a technique introduced by Henderson

(1963, 1973, 1975) for estimating random effects. The method provides a basis for generating ML and REML estimates particularly for GLMMs. It involves maximising the joint likelihood function, $l=l_1+l_2$ where

$$l_1 = -\frac{1}{2}[n \ln 2\pi\sigma^2 + \ln |\mathbf{D}| + \sigma^{-2}(\mathbf{y}-\mathbf{X}\beta-\mathbf{Zu})'\mathbf{D}^{-1}(\mathbf{y}-\mathbf{X}\beta-\mathbf{Zu})]$$

(2.4)

$$l_2 = -\frac{1}{2}\sum_{j=1}^{k}[v_j \ln(2\pi\sigma_j^2) + \ln |\mathbf{A}_j(\rho)| + \sigma_j^{-2}\mathbf{u}_j'\mathbf{A}^{-1}\mathbf{u}_j]$$

viz. l_1 is the log-likelihood for $\mathbf{y}$ given $\mathbf{u}$ fixed and l_2 is the log-likelihood for $\mathbf{u}$. For normal error models, the estimates $\tilde{\beta}$ and $\tilde{\mathbf{u}}$ are those values which equate the derivatives of l with respect to β and $\mathbf{u}$ to zero. Henderson (1950) derived the simultaneous equations, also referred to as mixed models equations, for calculating BLUP estimates as follows

(2.5)
$$\begin{bmatrix} \mathbf{X'D^{-1}X} & \mathbf{X'D^{-1}Z} \\ \mathbf{Z'D^{-1}X} & \mathbf{Z'D^{-1}Z+A^{-1}} \end{bmatrix} \begin{bmatrix} \tilde{\beta} \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'D^{-1}y} \\ \mathbf{Z'D^{-1}y} \end{bmatrix}.$$

Solving the equations provides explicit expressions for $\tilde{\beta}$ and $\tilde{\mathbf{u}}$

(2.6)
$$\tilde{\beta} = [\mathbf{X'\Sigma^{-1}X}]^{-1}\mathbf{X'\Sigma^{-1}y}$$

$$\tilde{\mathbf{u}} = \mathbf{AZ'\Sigma^{-1}}[\mathbf{y} - \mathbf{X}\beta]$$

An alternative expression for $\tilde{\mathbf{u}}$ is derived by Patterson and Thompson (1971)

(2.7) $$\tilde{\mathbf{u}} = (\mathbf{Z'KZ+A^{-1}})^{-}\mathbf{Z'Ky}$$

where $\mathbf{K} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{X}(\mathbf{X'D^{-1}X})^{-}\mathbf{X'D}^{-1}$ is a symmetric matrix with $\mathbf{X'KX=0}$ implies $\mathbf{KX=0}$.

13

Taking the derivatives with respect to $\tilde{\sigma}_j^2$, $\tilde{\sigma}^2$, ρ_s and equating to zero provides the estimating equations for variance components. Solving the equations and using the estimates $\tilde{\beta}$ and $\tilde{u}$, the BLUP estimators of the variance components are obtained as

(2.8)
$$\tilde{\sigma}^2 = (\mathbf{y} - \mathbf{X}\tilde{\beta} - \mathbf{Z}\tilde{u})'\mathbf{D}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\beta} - \mathbf{Z}\tilde{u})/n$$

$$\tilde{\sigma}_j^2 = \tilde{u}_j'\mathbf{A}_j^{-1}\tilde{u}_j/v_j$$

$$\sum_{j=1}^{k}[r_j^{(s)} - \tilde{\sigma}_j^2\,\tilde{u}_j\mathbf{A}_j^{-1}(\partial\mathbf{A}_j/\partial\rho_s)\mathbf{A}_j^{-1}\tilde{u}_j]\Big|_{\rho=\tilde{\rho}_s} = 0, \ s=1,2,...,S$$

where $r_j^{(s)} = \mathrm{tr}[\mathbf{A}_j^{-1}(\partial\mathbf{A}_j/\partial\rho_s)]$. The equation for ρ_s may not be solvable explicitly.

The above derivation is done without assuming any normality of the joint distribution of $\mathbf{y}$ and $\mathbf{u}$. However, under the condition of normality, BLUP estimators have some good properties.

1. The BLUP estimators $\tilde{\beta}$ and $\tilde{u}$ are identical to ML estimators given the variance-covariance matrix of $\mathbf{y}$.
2. $E(\mathbf{u}\,|\,\tilde{u}) = \tilde{u}$
3. $\mathrm{Var}(\mathbf{u}\,|\,\tilde{u}) = \mathrm{Var}(\mathbf{u}) - \mathrm{Var}(\tilde{u})$

Henderson (1973) also extended BLUP to predict $\mathbf{k}'\beta+\mathbf{m}'\mathbf{u}$ by $\mathbf{k}'\tilde{\beta}+\mathbf{m}'\tilde{u}$.

2.2.3 Maximum Likelihood (ML)

With the advancement of computing technology likelihood based methods for estimating variance components are becoming more and more popular. The applications of ML theory to variance components estimation are discussed in

14

Hartley and Rao (1967), Anderson (1973) and Miller (1973, 1977). A unified review of ML approaches to variance component estimation is provided by Harville (1977). As Harville mentioned, a ML approach to variance estimation has a number of attractive properties. The ML estimators are asymptotically normal and efficient, consistent and are functions of every sufficient statistic. The ML approach does not suffer from the deficiencies of some other methods. For example, the possibility of negative estimates of variance components can easily be removed by including non-negativity constraints in the parameter space. It offers a strategy for simultaneously estimating fixed and random effects and the corresponding variance components. One of the requirements of the ML approach is to assume a parametric form for the distribution of the data vector. Generally, the distribution is assumed to be normal for continuous data.

The ML estimators are derived by maximising the likelihood function with respect of the parameters to be estimated. For the mixed model defined in (2.1) the log-likelihood function of the observation vector $\mathbf{y}$ is given by

$$(2.9) \qquad l = -\frac{1}{2} [n \ln (2\pi) + \ln |\mathbf{V}| + (\mathbf{y} - \mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)]$$

The derivatives of l with respect to β, σ^2, σ_j^2 and ρ are as follows

$$\frac{\partial l}{\partial \beta} = [\mathbf{X}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)],$$

$$\frac{\partial l}{\partial \sigma^2} = (-1/2)[n\sigma^{-2} - \sigma^{-2}(\mathbf{y}-\mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\beta)],$$

(2.10)

$$\frac{\partial l}{\partial \sigma_j^2} = (-1/2) [\text{tr}(\mathbf{V}^{-1}\mathbf{Z}_j\mathbf{A}_j\mathbf{Z}_j') - (\mathbf{y}-\mathbf{X}\beta)'\mathbf{V}^{-1}\mathbf{Z}_j\mathbf{A}_j\mathbf{Z}_j'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\beta)],$$

$$\partial l/\partial \rho_s = (-1/2) [\text{tr}(\mathbf{V}^{-1}\mathbf{Z}\partial\mathbf{A}/\partial\rho_s\mathbf{Z}') - (\mathbf{y}-\mathbf{X}\beta)'\mathbf{V}^{-1}\mathbf{Z}\partial\mathbf{A}/\partial\rho_s\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\beta)],$$

Generally, the ML estimates are obtained by equating the derivatives to zero and solving for the parameters. However, in this case the solutions for β and σ_j^2 may not necessarily be ML estimators because of the boundary restrictions on σ_j^2. If $\mathbf{V}$ is known then the solution for β,

$$(2.11) \qquad \tilde{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y},$$

gives ML estimates because there are no boundary restrictions on β. If the corresponding solutions for other parameters are within the boundary i.e. $\sigma^2 > 0$ and $\sigma_j^2 \geq 0$ then the estimates are ML estimates. If the above conditions are not met then the solutions are not ML. One way of resolving the problem in that situation is by replacing negative values by zero and then recalculating the estimates until the non-negativity conditions are met. The approach is further discussed in Herbach (1959) and Thompson (1962). In most cases the solutions have to be obtained numerically, usually by iteration.

Hartley and Rao (1967) proposed an alternative method by formulating the likelihood function in terms of $\mathbf{H}$ and $\phi = \sigma_j^2/\sigma^2$ where $\mathbf{H}$ is given by

$$(2.12) \qquad \mathbf{V} = \sigma^2\mathbf{H} = \sigma^2[\mathbf{I}+\mathbf{Z}\mathbf{A}\mathbf{Z}']$$

with $\mathbf{A}$ is as defined in (2.2). Using (2.5) in the above derivatives the estimates of β and σ^2 are obtained as

$$(2.13) \qquad \mathbf{X}'\mathbf{H}^{-1}\mathbf{y} = \mathbf{X}'\mathbf{H}^{-1}\mathbf{X}\,\tilde{\beta}$$

$$(2.14) \qquad \sigma^2 = (\mathbf{y} - \mathbf{X}\tilde{\beta})'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\beta})/n$$

and the estimate of ϕ_j is obtained iteratively from

$$(2.15) \qquad \mathrm{tr}(\mathbf{H}^{-1}\mathbf{Z}_j\mathbf{Z}_j') = \sigma^2(\mathbf{y} - \mathbf{X}\tilde{\beta})'\mathbf{H}^{-1}\mathbf{Z}_j\mathbf{Z}_j'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\beta}) .$$

Since (2.12) does not involve ρ, there is no derivative with respect to ρ.

The availability of the large sample, asymptotic dispersion matrix of the estimators is one of the attractive features of the ML estimation. The inverse of the information matrix provides the asymptotic variances of the estimates where the elements of the information matrix are the second order derivatives of the likelihood function.

McGilchrist and Yau (1995) derived an expression for the information matrix I_{ML} for $\tilde{\beta}$, $\tilde{\sigma}^2$, $\tilde{\phi}_j$, $\tilde{\rho}_s$ as

$$
I_{ML} = \begin{bmatrix}
\sigma^{-2}L & 0 & 0 & 0 \\
0 & n/2\sigma^4 & (1/2\sigma^2)[\mathrm{tr}\Sigma^{-1}\partial\Sigma/\partial\phi_j] & (1/2\sigma^2)[\mathrm{tr}\Sigma^{-1}\partial\Sigma/\partial\rho_t] \\
0 & . & (1/2)\mathrm{tr}[\Sigma^{-1}\partial\Sigma/\partial\phi_i\Sigma^{-1}\partial\Sigma/\partial\phi_j] & (1/2)\mathrm{tr}[\Sigma^{-1}\partial\Sigma/\partial\phi_i\Sigma^{-1}\partial\Sigma/\partial\rho_t] \\
0 & . & . & (1/2)\mathrm{tr}[\Sigma^{-1}\partial\Sigma/\partial\rho_s\Sigma^{-1}\partial\Sigma/\partial\rho_t]
\end{bmatrix}
$$

where $L = X'\Sigma^{-1}X$ implying that for a particular value of Σ

(2.16)
$$\tilde{\beta} = L^{-}X'\Sigma^{-1}y$$

$$\tilde{\sigma}^2 = (y-X\tilde{\beta})'\Sigma^{-1}(y-X\tilde{\beta})/n .$$

Given that $\Sigma = D + ZAZ' = D + \sum_{j=1}^{k} \phi_j Z_j A_j(\rho) Z_j'$,

$$\partial\Sigma/\partial\phi_j = Z_j A_j(\rho) Z_j' , \qquad \partial\Sigma/\partial\rho_s = \sum_{j=1}^{k} \phi_j Z_j \partial A_j/\partial\rho_s Z_j' ,$$

$$Q = K(K\Sigma K)^{-}K , \qquad Qy = \Sigma^{-1}(y-X\tilde{\beta}) = D^{-1}(y-X\tilde{\beta}-Z\tilde{u}) ,$$

alternative expressions for variance components and I_{ML} are derived as follows

$$\tilde{\sigma}^2 = y'\Sigma^{-1}(y-X\tilde{\beta})/n = y'D^{-1}(y-X\tilde{\beta}-Z\tilde{u})/n = y'Qy/n,$$

(2.17)

$$\tilde{\sigma}_j^2 = \tilde{u}_j'A_j^{-1}\tilde{u}_j/(v_j-r_j^*)$$

$$\partial l/\partial\rho_s\big|_{\beta=\tilde{\beta}} = -(1/2)\sum_{j=1}^{k}[v_j^{(s)}+r_j^{*(s)}+\sigma_j^{-2}\tilde{u}_j'(\partial A_j^{-1}/\partial\rho_s)\tilde{u}_j]_{\rho=\tilde{\rho}} = 0 \text{ for } \rho_s$$

which may be required to be solved iteratively. The definitions of the terms $v_j^{(s)}$ and $r_j^{*(s)}$ are given below. Let

(2.18)
$$\begin{bmatrix} X'D^{-1}X & X'D^{-1}Z \\ Z'D^{-1}X & Z'D^{-1}Z+A^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} \cdot & \cdot \\ \cdot & T \end{bmatrix}$$

that means T is that part of the inverse corresponding to $Z'D^{-1}Z+A^{-1}$ in the original matrix and $T^*=(Z'D^{-1}Z+A^{-1})^{-1}$. T_{jj}^* is a partition of T^* into blocks conformal to the partition of u.

Then defining the following terms

$$r_j^* = \phi_j^{-1}\text{tr } A_j^{-1}T_{jj}^*, \qquad r_j^{*(s)} = \phi_j^{-1}\text{tr } \partial A_j^{-1}/\partial\rho_s T_{jj}^*,$$

(2.19)

$$r_j^{*(st)} = \phi_j^{-1}\text{tr } \partial A_j^{-1}/\partial\rho_s T_{jj}^* A_j^{-1}\partial A_j^{-1}/\partial\rho_t,$$

$$r_{ij}^* = \text{tr } T_{ij}^*A_j^{-1}T_{ji}^*A_i^{-1}, \qquad r_{ij}^{*(s)} = \text{tr } T_{ij}^* \partial A_j^{-1}/\partial\rho_s T_{ji}^*A_i^{-1},$$

$$r_{ij}^{*(st)} = \text{tr } T_{ij}^* \partial A_j^{-1}/\partial\rho_s T_{ji}^*\partial A_i^{-1}/\partial\phi_t, \qquad v_j = \text{tr } A_j^{-1}A_j,$$

$$v_j^{(s)} = \text{tr } A_j^{-1}\partial A_j/\partial\phi_s, \qquad v_j^{(st)} = \text{tr } \partial A_j/\partial\phi_s\partial A_j^{-1}/\partial\phi_t,$$

the information matrix multiplied by 2 can be expressed as

$$2I_{ML} =$$

$$\begin{bmatrix} 2\sigma^{-2}L & 0 & 0 & 0 \\[2mm] 0 & n/\sigma^4 & \sigma_j^{-2}(v_j-r_j^*) & \sigma_j^{-2}\sum_{j=1}^{k}(v_j^{(t)}+r_j^{*(t)}) \\[2mm] 0 & \cdot & [\phi_i^{-2}(v_i-2r_i^*)\delta_{ij}+\phi_i^{-2}\phi_j^{-2}r_{ij}^*] & [\phi_i^{-1}(v_i^{(t)}+2r_i^{*(t)}-\sum_{j=1}^{k}\phi_i^{-1}\phi_j^{-1}r_{ij}^{*(t)})] \\[2mm] 0 & \cdot & \cdot & \sum_{j=1}^{k}(-v_j^{(st)}+2r_j^{*(st)}+\sum_{m=1}^{k}\phi_j^{-1}\phi_m^{-1}r_{jm}^{*(st)}) \end{bmatrix}$$

2.2.4 Residual Maximum likelihood (REML)

One of the shortcomings of ML estimation of the variance components is that it does not consider the loss in degrees of freedom that results from estimation of the fixed coefficients. The ML estimates, even from simple balanced data, do not coincide with those generated by ANOVA methods. They tend to be negatively biased and the variance is not minimum (Searle et al. 1992). To overcome this drawback the REML method was developed originally by Anderson and Bancroft (1952) and Russell and Bradley (1958) for specific balanced ANOVA models. Thompson (1962) extended it for all balanced ANOVA models, and Patterson and Thompson (1971, 1974) extended it in general form. The essence of REML estimation is to calculate variance components based on residuals calculated after fitting the fixed effects part of the model through ordinary least squares. The estimation can also be viewed as maximising a marginal likelihood. The REML estimators are identical to those resulting from ANOVA methods and likewise bear the property of unbiasedness and minimum variance. An expression for the REML log-likelihood function given by Patterson and Thompson (1971) is

(2.20) $l_{REML} = -(1/2)\,[(n-p)\ln(2\pi\sigma^2) + \ln|K\Sigma K| + \sigma^{-2}y'K(K\Sigma K)^{-}Ky]$

here $\mathbf{K} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{D}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{D}^{-1}$ which satisfies the condition $\mathbf{KX=0}$. $(\mathbf{K\Sigma K})^{-}$ is the generalised inverse of $\mathbf{K\Sigma K}$ and $|\mathbf{K\Sigma K}|$ is the determinant of linearly independent rows and columns of $\mathbf{K\Sigma K}$. The estimators for σ^2 and ϕ are derived by taking the derivatives of l_{REML} as follows:

(2.21)
$$\partial l_{\text{REML}}/\partial\sigma^2 = (-1/2)[(n-v)\sigma^{-2} - \sigma^{-4}\mathbf{y}'\mathbf{Q}\mathbf{y}],$$

$$\partial l_{\text{REML}}/\partial\phi_j = (-1/2)[\text{tr}\mathbf{Q}\partial\Sigma/\partial\phi_j - \sigma^{-2}\mathbf{y}'\mathbf{Q}\partial\Sigma/\partial\phi_j\mathbf{Q}\mathbf{y}],$$

$$\partial l_{\text{REML}}/\partial\rho_s = (-1/2)[\text{tr}\mathbf{Q}\partial\Sigma/\partial\rho_s - \sigma^{-2}\mathbf{y}'\mathbf{Q}\partial\Sigma/\partial\rho_s\mathbf{Q}\mathbf{y}],$$

where $\mathbf{Q} = \mathbf{K}(\mathbf{K\Sigma K})^{-}\mathbf{K}$.

Equating the above derivatives to zero and solving for σ^2 and ϕ gives the REML estimates. Again the equations may have to be resolved iteratively.

Similarly to the ML estimators, the asymptotic dispersion matrix of the parameters can be derived by taking the inverse of the information matrix that is calculated by using the second derivatives of the REML log-likelihood function.

An expression for I_{REML} was derived by McGilchrist and Yau (1995) as follows

$$I_{\text{REML}} = \begin{bmatrix} (n-v)/2\sigma^4 & (1/2\sigma^2)[\text{tr }\mathbf{Q}\partial\Sigma/\partial\phi_j] & (1/2\sigma^2)[\text{tr}\mathbf{Q}\partial\Sigma/\partial\rho_t] \\ \cdot & (1/2)\text{tr}[\mathbf{Q}\partial\Sigma/\partial\phi_i\mathbf{Q}\partial\Sigma/\partial\phi_j] & (1/2)\text{tr}[\mathbf{Q}\partial\Sigma/\partial\phi_i\mathbf{Q}\partial\Sigma/\partial\rho_t] \\ \cdot & \cdot & (1/2)\text{tr}[\mathbf{Q}\partial\Sigma/\partial\rho_s\mathbf{Q}\partial\Sigma/\partial\rho_t] \end{bmatrix}$$

Using $\partial\Sigma/\partial\phi_j = \mathbf{Z}_j\mathbf{A}_j(\rho)\mathbf{Z}_j'$, $\partial\Sigma/\partial\rho_s = \sum_{j=1}^{k}\phi_j\mathbf{Z}_j\partial\mathbf{A}_j/\partial\rho_s\mathbf{Z}_j'$ and the notations used in the previous section, expressions for variance components are derived as:

$$(2.22) \quad \sigma^2_{REML} = (n-v)^{-1}\mathbf{y}'\mathbf{Q}\mathbf{y} \;,$$

$$\partial l_{REML}/\partial\phi_j = -(1/2\phi_j)[v_j - r_j - \sigma_j^{-2}\tilde{\mathbf{u}}_j'\mathbf{A}_j^{-1}\tilde{\mathbf{u}}_j] = 0$$

$$\text{giving} \quad \sigma_j^2 = \tilde{\mathbf{u}}_j'\mathbf{A}_j^{-1}\tilde{\mathbf{u}}_j/(v_j - r_j^*)$$

$$\partial l/\partial\rho_s \big|_{\beta=\tilde{\beta}} = -(1/2)\sum_{j=1}^{k}[v_j^{(s)} + r_j^{(s)} + \sigma_j^{-2}\tilde{\mathbf{u}}_j'(\partial\mathbf{A}_j^{-1}/\partial\rho_s)\tilde{\mathbf{u}}_j]_{\rho=\tilde{\rho}} = 0 \text{ for } \rho_s$$

which may required to be solved iteratively. The information matrix multiplied by 2 is

$$2 I_{REML} =$$

$$\begin{bmatrix} (n-v)/\sigma^4 & \sigma_j^{-2}(v_j - r_j) & \sigma_j^{-2}\sum_{j=1}^{k}(v_j^{(t)} + r_j^{(t)}) \\[2ex] \cdot & [\phi_i^{-2}(v_i - 2r_i)\delta_{ij} + \phi_i^{-2}\phi_j^{-2}r_{ij}] & [\phi_i^{-1}(v_i^{(t)} + 2r_i^{(t)} - \sum_{j=1}^{k}\phi_i^{-1}\phi_j^{-1}r_{ij}^{(t)})] \\[2ex] \cdot & \cdot & \sum_{j=1}^{k}(-v_j^{(st)} + 2r_j^{(st)} + \sum_{m=1}^{k}\phi_j^{-1}\phi_m^{-1}r_{jm}^{(st)}) \end{bmatrix}$$

The expressions for r_j, r_{ij}, $r_j^{(s)}$, $r_j^{(st)}$, $r_{ij}^{(s)}$ and $r_{jm}^{(st)}$ are obtained if $\mathbf{T}^*$ and $\mathbf{T}_{jj}^*$ are replaced by $\mathbf{T}$ and $\mathbf{T}_{jj}$ in the derivations of r_j^*, $r_{ij}^{*(s)}$, $r_j^{*(s)}$, $r_{ij}^{*(st)}$, $r_{ij}^{*(s)}$ and $r_{jm}^{*(st)}$ as presented in the previous section.

The above expression for I_{REML} excludes the variance-covariances for the fixed coefficients. These remain the same as for ML estimates.

Except some specific situations where explicit expressions for ML and REML estimates can be derived (Herbach, 1959; Thompson, 1962), an iterative numerical procedure needs to be employed to obtain the estimates. The procedure turns out to be very computationally intensive as in each iteration the first and second partial derivatives, expected values of the second derivatives and other related quantities are required to be

calculated. Moreover, there are many different numerical algorithms that can be applied and none of these seem to work uniformly well in all applications. Anderson (1973) and Henderson (1973) proposed algorithms for computing ML estimates of ϕ in specific situations. Harville (1977) reviewed various procedures for obtaining efficient solutions in both specific and general situations. The procedures for computing ML and REML estimates of variance components rely mainly on a variation of the methods of steepest ascent (Hartley and Vaughn, 1972), the Newton-Raphson method (Hemmerle and Hartley, 1973; Corbeil and Searle, 1976) and the method of scoring (Jennrich and Sampson, 1976).

2.2.5 Other Methods

There are many other approaches to the estimation of variance components which are not very widely used in practice (Searle et al. 1992). Among these Minimum Norm Quadratic Unbiased Estimation (MINQUE) and Minimum Variance Quadratic Unbiased Estimation (MIVQUE) which were introduced by Rao (1970, 1971a, 1971b, 1972). Under normality, MINQUE and MIVQUE are identical. No distributional assumption is required for MINQUE except the requirement of the existence of the first four moments. The procedure also does not involve any iteration but for solving the linear equations it depends on *a priori* values for variance components and the solutions can be different for different choices of *a priori* values. This is considered to be a disadvantage of the approach. Rao (1979) developed a class of iterated Minimum Norm Quadratic Estimators (MINQE) which do not depend on any *a priori* values. This class also includes ML and REML estimators. For details of these methods of estimation, see Rao and Kleffe (1988).

22

A number of authors proposed Bayesian methods for the estimation of variance components. In the Bayesian framework the posterior distribution of the variance components are obtained based on a prior distribution. Hill (1965, 1967) first considered the Bayesian method for estimating variance components in the one-way random model. A review of the Bayesian methods for variance components can be found in (Khuri and Sahai, 1985). It is noted in the latter paper that the Bayesian methodology was investigated mainly for balanced data and not many papers were published for its application to unbalanced data.

2.3 METHODS FOR DISCRETE RESPONSE VARIABLES

2.3.1 Generalised Linear Mixed Models (GLMMs)

The investigation of the methods of fitting mixed models to a discrete dependant variable has been started rather recently. The development of generalised linear modelling (GLM) techniques by Nelder & Wedderburn (1972) and McCullagh and Nelder (1989) paved the way for extending mixed models to discrete response data. GLM has unified the regression methodology for a variety of discrete and continuous response variables. GLMMs are in fact an extension of GLMs to incorporate both fixed and random effects in the model. Similarly to GLM, in GLMMs the distribution of the error term is not restricted to the normal but can follow any distribution from the exponential family.

The problem of estimation in models with non-normal error is more complex than that of mixed models with normal error. In non-normal error models, the random error is generally assumed to follow a multivariate normal distribution and variance components are estimated based on the marginal

distribution of the response variable. However, a full maximum likelihood analysis based on the joint marginal distribution requires numerical integration techniques for calculation of the log-likelihood by integrating out with respect to the random effect variables. The problem with this marginal likelihood approach is that, except in some simple cases, it has proved intractable to perform the numerical integration operation, particularly for more complicated problems involving high dimensional integrals. A variety of different approaches have been proposed to overcome the problem.

Leonard (1972) gave a Bayesian procedure for estimating the variance of the random effects. The Bayesian approach to binary responses is described in Stiratelli et al. (1984) for modelling the dependence among outcome variables inherent in longitudinal or repeated measures designs. Recent Bayesian techniques overcome the problem of numerical integration by taking repeated samples from the posterior distributions using importance (Raghunathan, 1994) or Gibbs sampling techniques (Besag et al., 1991; Zeger and Karim, 1991).

Williams (1982) proposed an empirical approach for binomial data with extra binomial variation. Breslow (1984) investigated the problem in the context of Poisson-gamma models. Crowder (1985) and Tsutakawa (1988) have investigated log-linear models with random effects for count data. Anderson and Aitkin (1985) and Im and Gionola (1988) used maximum likelihood estimation in logistic and probit models where the random effects are assumed to follow a normal distribution, and the conditional distribution of y is binomial. Other models have been proposed by Williams (1975), Crowder (1978) and Kupper & Haseman (1978).

Methods for estimating random parameters from threshold data are presented by Gianola & Foulley (1982, 1983) and Harville and Mee (1984). These methods, based on Bayesian arguments, maximize the likelihood jointly for both fixed and random effects. Gilmour et al. (1985) proposed an alternative solution for predicting random effects which maximizes the likelihood with respect to the fixed effects while taking expectations over the random effects. The solutions of the resultant equations produce predicted values of the random effects. The Gilmour et al. (1985) approach has elements in common with the EM algorithm of Dempster et al. (1977).

Zeger et al. (1988) introduced the concepts of 'subject specific' (SS) and 'population averaged' (PA) approaches. In SS models, subject-to-subject heterogeneity is explicitly modelled, while in PA models, importance is given to population level inference rather than to any individual. The generalized estimating equation (GEE) approach of Zeger and Liang (1986), Liang and Zeger (1986) and Zeger et al. (1988) is based on the PA approach. This models the marginal expectation, rather than conditional expectation given a cluster-specific effect, thus avoiding the need for numerical integration. As the strategy does not provide any estimates of random effects it is not useful to those situations where interest is on the random effects.

Schall (1991) used a different approach which went directly to a linearisation of the link function applied to the observations rather than approximating the likelihood function. This direct approach enabled Schall to use the relationship among BLUP, ML and REML approaches, developed for normal theory models by Harville (1977), Patterson and Thompson (1971) and

Fellner (1986, 1987), to obtain similar methods for GLMMs. Wolfinger (1993) applied an approximation to Laplace's method for integrals to marginal distributions of non-normal data and extended Schall's approach.

Breslow and Clayton (1993) developed penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL) for approximate inference in GLMMs. Similar approaches have been used by Engel and Keen (1992) and Wolfinger and O'Connell (1993). PQL method was investigated previously by Green (1987) for inference in hierarchical models. Laird (1978) and Stiratelli et al. (1984) also proposed PQL as an approximate Bayes procedure for some GLMMs. The MQL procedure, in fact, originated in Goldstein (1991) as an extension to GLMs of his work on multilevel modelling (Goldstein 1986, 1988). By using informal mathematical arguments, simulations and applying to several data sets, Breslow and Clayton (1993) conclude that PQL is useful in practice for approximate inference on fixed and random effects in the hierarchical model. However, when applied to clustered binomial data the approach underestimates somewhat the variance components and fixed effects, but the bias tends to disappear rapidly for binomial observations with denominators greater than one (Breslow and Clayton 1993). The Solomon and Cox (1992) approach and a Laplace approximation are compared through different expansions of the likelihood function in Breslow and Lin (1995) where a bias correction for PQL is given. The failure to account for the contribution of the estimated variance components when assessing the uncertainty in both fixed and random effects is another limitation of PQL. On the other hand, an important drawback of MQL is its inability to correctly model the heterogeneity in the fixed effects which leads to a misleading correlation in the estimated random effects. Breslow and Clayton

(1993) recommended that MQL should be the preferred method when the marginal relationship between covariables and response is of interest but PQL should be used when estimates of random effects are of interest in the hierarchical model.

McCulloch (1994) proposed a variation of the EM algorithm for computation of ML and REML estimation of variance components and an analog of BLUP for the realized values of the random effects in a class of probit-normal models for binary data. The strategy was extended further in McCulloch (1997) to develop another computation strategy called Monte Carlo Newton-Raphson (MCNR) method.

McGilchrist and Aisbett (1991a) proposed an estimation procedure based on Henderson's BLUP technique (1963, 1973, 1975). They proposed to replace the likelihood function for fixed and random components by the asymptotic likelihood of their ML estimators and the distribution of the random components by a restricted prior. This is essentially a PQL approach.

McGilchrist (1994) further modified the McGilchrist and Aisbett (1991a) approach. The method extended BLUP methods to ML and REML estimation procedures in GLMMs. The approach is still similar in principle to PQL and very much in agreement with McGilchrist and Aisbett (1991a), Schall (1991) and Breslow and Clayton (1993). It uses a slightly different argument for approximate linearisation which provided a rationale for applying the method to a wider class of problems. Applications of the method have been made to Multi-centre clinical trials in McGilchrist and Zhaorong (1990); discordance data in Zhaorong et al. (1992); to threshold models in Zhaorong et al. (1992), Saei (1996) and Saei and McGilchrist (1996); to survival analysis in

McGilchrist and Aisbett (1991b), McGilchrist (1993), McGilchrist and Yau (1996), Yau and McGilchrist (1998). This approach has been investigated further in the current thesis.

Lee and Nelder (1996) has further extended the linearisation approach by introducing a variety of random component model distributions and deriving estimating equations for the parameters of the generalised model. Their approach has been termed the hierarchical likelihood approach.

2.3.2 Estimation Framework in GLMMs

As mentioned above a wide variety of different approaches has been proposed for obtaining ML or REML estimates in GLMMs with non-normal error. A number of authors have proposed penalised quasi-likelihood with some variations. As most of the approaches are very much in agreement, a general framework of GLMM with particular references to McGilchrist (1994) is presented here.

In GLMM, the distribution of the response vector $\mathbf{y}$ given $\mathbf{u}$ depends on an underlying vector quantity η which in turn relates to the explanatory variables through the equation

$$\eta = \mathbf{X}\beta + \mathbf{Z}\mathbf{u}$$

The notation used here is the same as described in section 2.1. The log-likelihood function of $\mathbf{y}$ conditional on fixed $\mathbf{u}$ is

$$l_1 = \ln\ f(\mathbf{y};\beta\,|\,\mathbf{u})$$

where $f(\mathbf{y};\beta\,|\,\mathbf{u})$ is the probability density function of $\mathbf{y}$ conditional on fixed $\mathbf{u}$. As the distribution of the random component vector $\mathbf{u}$ is assumed to be normal the log-likelihood is given by

$$(2.23) \qquad l_2 = -\frac{1}{2} \sum_{j=1}^{k} [v_j \ln(2\pi\sigma_j^2) + \ln |A_j(\rho)| + \sigma_j^{-2} u_j' A^{-1} u_j]$$

Estimators are derived using a penalised likelihood approach. The approach is consistent with the BLUP philosophy which obtains estimators $\tilde{\beta}$, $\tilde{u}$ by maximising the penalised likelihood function $l = l_1 + l_2$. First and second derivatives of the log-likelihood are

$$\partial l/\partial\beta = X'\partial l_1/\partial\eta \quad , \quad \partial l/\partial u = Z'\partial l_1/\partial\eta + \partial l_2/\partial u$$

(2.24)

$$\partial^2 l/\partial\beta\partial\beta' = X'(\partial^2 l_1/\partial\eta\partial\eta')X, \quad \partial^2 l/\partial\beta\partial u' = X'(\partial^2 l_1/\partial\eta\partial\eta')Z$$

$$\partial^2 l/\partial u\partial u' = Z'(\partial^2 l_1/\partial\eta\partial\eta')Z + \partial^2 l_2/\partial u\partial u'$$

Letting $\mathbb{B} = -\partial^2 l_1/\partial\eta\partial\eta'$, $\mathbb{A} = -\partial^2 l_2/\partial u\partial u'$, the values $\tilde{\beta}$ and $\tilde{u}$ which maximise the likelihood function are obtained by using an iterative Newton-Raphson algorithm as follows

$$(2.25) \qquad \begin{bmatrix} \beta_1 \\ u_1 \end{bmatrix} = \begin{bmatrix} \beta_0 \\ u_0 \end{bmatrix} + V^{-1} \begin{bmatrix} X' \\ Z' \end{bmatrix} t + V^{-1} \begin{bmatrix} 0 \\ s \end{bmatrix} , \quad \text{where} \quad t = \partial l_1/\partial\eta \; , \quad s = \partial l_2/\partial u$$

and β_0 , u_0 are initial values which are replaced by β_1 , u_1 after the first iteration and the process continued until convergence is achieved. The variance-covariance matrix V is given by

$$V = - \begin{bmatrix} \partial^2 l/\partial\beta\partial\beta' & \partial^2 l/\partial\beta\partial u' \\ \partial^2 l/\partial u\partial\beta' & \partial^2 l/\partial u\partial u' \end{bmatrix}$$

(2.26)

$$= \begin{bmatrix} X' \\ Z' \end{bmatrix} \mathbb{B} \begin{bmatrix} X & Z \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{A} \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} , \quad V^{-1} = \begin{bmatrix} T_1 & T_2 \\ T_2' & T \end{bmatrix}$$

If V is replaced by $E(V)$ then the iterative procedure becomes the method of scoring.

McGilchrist (1994) argued that the likelihood function l can be approximated by assuming that $\tilde{\beta}$ and $\tilde{u}$ have approximately a joint normal distribution with means β and u and variance matrix V^- as follows

$$(2.27) \qquad l \cong \text{constant} + \frac{1}{2} \begin{bmatrix} \beta - \tilde{\beta} \\ u - \tilde{u} \end{bmatrix}' V \begin{bmatrix} \beta - \tilde{\beta} \\ u - \tilde{u} \end{bmatrix}.$$

In McGilchrist and Aisbett (1991a), the component l_1 of the BLUP procedure is replaced by the log-likelihood of $\tilde{\beta}$ and $\tilde{u}$ based on the normal approximation of its asymptotic distribution. The resulting BLUP log-likelihood is identical with the above quadratic expression. The estimating equations for β and u are also consistent with that of the method of scoring. McGilchrist (1994) concluded that if l is approximately quadratic in β and u then the BLUP estimation may be considered as derived from the very approximate asymptotic distribution of $\tilde{\beta}$ and $\tilde{u}$. Then the joint likelihood of the BLUP procedure is approximated as $l^* = l_1^* + l_2$ with l_2 as defined above and

$$l_1^* = \text{constant} - \frac{1}{2} \begin{bmatrix} \beta - \tilde{\beta} \\ u - \tilde{u} \end{bmatrix}' \begin{bmatrix} X' \\ Z' \end{bmatrix} \mathbb{B} \begin{bmatrix} X & Z \end{bmatrix} \begin{bmatrix} \beta - \tilde{\beta} \\ u - \tilde{u} \end{bmatrix}$$

(2.28)

$$= \text{constant} - (1/2)(y^* - X\beta - Zu)'\mathbb{B}-(y^* - X\beta - Zu)$$

where $y^* = X\tilde{\beta} + Z\tilde{u}$ and $\mathbb{B} = -E(\partial^2 l_1 / \partial \eta \partial \eta')$

2.3.3 BLUP to ML and REML

For given ϕ, both ML and REML estimators of β are identical to the BLUP estimators. But the estimators of the variance components are different.

If $T^* = V_{22}^{-1}$ then the ML estimators are

30

$$(2.29) \quad \sigma^2_{ML} = \mathbf{y}^{*\prime}\mathbb{B}(\mathbf{y}^*\text{-}\mathbf{X}\tilde{\boldsymbol{\beta}}\text{-}\mathbf{Z}\tilde{\mathbf{u}})/n \; ,$$

$$\sigma^2_{j(ML)} = \tilde{\mathbf{u}}'_j\mathbf{A}^{-1}_j\tilde{\mathbf{u}}_j/(v_j\text{-}\mathbf{r}^*_j) \; ,$$

$$\sum_{j=1}^{k}[v_j^{(s)}+r_j^{*(s)}+\sigma_j^{-2}\tilde{\mathbf{u}}'_j(\partial\mathbf{A}^{-1}_j/\partial\rho_s)\tilde{\mathbf{u}}_j] = 0, \; \text{for} \quad \rho_{s(ML)}, \; s=1,2,...,S$$

where $\mathbf{r}^*_j = \phi^{-1}_j \operatorname{tr}(\mathbf{A}^{-1}_j\mathbf{T}^*_{jj})$. The information matrix for the ML estimators of σ^2, ϕ_j and ρ_s are as given in the previous section.

Similarly, the REML estimators of the variance components are derived by using $\mathbf{T}$ instead of $\mathbf{T}^*$;

$$(2.30) \quad \sigma^2_{REML} = \mathbf{y}^{*\prime}\mathbb{B}(\mathbf{y}^*\text{-}\mathbf{X}\tilde{\boldsymbol{\beta}}\text{-}\mathbf{Z}\tilde{\mathbf{u}})/(n\text{-}v) \; ,$$

$$\sigma^2_{j(REML)} = \tilde{\mathbf{u}}'_j\mathbf{A}^{-1}_j\tilde{\mathbf{u}}_j/(v_j\text{-}\mathbf{r}_j) \; ,$$

$$\sum_{j=1}^{k}[v_j^{(s)}+r_j^{(S)}+\sigma_j^{-2}\tilde{\mathbf{u}}'_j(\partial\mathbf{A}^{-1}_j/\partial\rho_s)\tilde{\mathbf{u}}_j] = 0, \; \text{for} \; \rho_{s(REML)}, \; s=1,2,...,S$$

where $\mathbf{r}_j = \phi^{-1}_j \operatorname{tr}(\mathbf{A}^{-1}_j\mathbf{T}_{jj})$. The information matrix for the ML estimators of σ^2, ϕ_j and ρ_s are as given in section 2.2.4.

For the normal error model, $\mathbf{y}^*$ and $\mathbb{B}$ should be replaced in the above expressions by $\mathbf{y}$ and $\mathbf{D}^{-1}$ respectively and for non-normal error model $\sigma^2=1$ implying $\phi_j=\sigma^2_j$.

As mentioned above the solution for the parameters has to be obtained iteratively. The process starts with arbitrary values of ϕ and ρ and obtains the initial estimates of β and $\mathbf{u}$ through the Newton-Raphson algorithm. Then the initial estimates of β and $\mathbf{u}$ are used to calculate the

initial estimates of ϕ and ρ.   The second iteration starts with these initial estimates of ϕ and ρ and then a new set of estimates of β and $\mathbf{u}$ is obtained.   The process is continued until convergence is achieved for β, $\mathbf{u}$, ϕ and ρ.

CHAPTER THREE

MOMENT PROPERTIES FOR ESTIMATORS IN GENERALISED LINEAR MIXED MODELS

3.1 INTRODUCTION

In GLMMs, valid asymptotic theory has been difficult to develop through expansion techniques (see Barndorff-Neilson and Cox, 1994) because of the number of random components. It will usually be true that, as the number of responses increase, so too will the number of random components involved in the corresponding model. Indeed, as the number of observations approach infinity, so too do the number of random components. Because of this difficulty the asympototic properties of the GLMM estimators have not been developed fully.

The aim of this chapter is to develop approximate moments for REML estimators. This is done by linking the distribution of estimators of fixed and random components in the mixed model to the distribution of the derivatives of loglikelihoods and derivatives of logarithm of density functions which have known first and second order moment properties. The linking equations are again based on approximate first order expansions but the resultant theory is further justification for the use of extended REML techniques. In section 3.2 the structure of the inference problem is set out and in sections 3.3, 3.4 and 3.5 estimation techniques are developed. Two alternative strategies for deriving the estimator of the variance components ϕ are discussed and a new method for calculating variance of the estimator $\tilde{\phi}$ is developed under the assumption of approximate normality for $\tilde{u}_j$. An estimator for an additional variance parameter γ and the

corresponding variance of the estimator $\tilde{\gamma}$ are also derived. Final sections consider application to specific classes of problems. One of the applications is to extend the method for a Poisson distributed response variable to the case where the distribution of the random components is non-normal. Finally, results are presented from a simulation study using the Poisson-gamma model.

3.2 THE LINEAR MIXED MODELS

In the previous chapter a basic description of GLMMs and the estimation problem has been presented. However, in this section problems are taken to have a particular structure which is sufficiently general to include GLMMs and dependent failure time models. The ith response is taken to have a distribution dependent on a quantity η_i, for $i=1,2,...,n$. In the structure considered here, regression parameters of interest are contained in the linear combinations

$$(3.1) \qquad \eta_i = x_i'\beta + z_i'u \quad , \quad i = 1,2,...,n$$

where x_i is a v-dimensional vector of risk or regression variables with regression coefficient β fixed across different values of i , while u is a vector of random components and having incidence vector z_i indicating which random components are present in η_i. Letting $\eta'=[\eta_1,\eta_2,...,\eta_n]$ then the linear model becomes $\eta = X\beta + Zu$ as described in 2.3.2. In some cases it is convenient to write $u'=[u_1',u_2',...,u_k']$ and conformal decomposition $Z=[Z_1,Z_2,...,Z_k]$. The u_j are mutually independent and have distributions with parameters ϕ_j , $j=1,2,...,k$ which are distinct one from the other but possibly a common parameter vector ρ . Usually the parameter ϕ_j represents the variance of the random component u_j while the parameter ρ describes the

34

correlation structure of the whole set of random components. For many problems there is no parameter ρ because the dependence structure is adequately explained by the random components, each with their own variance.

The response variables of the experiment may be loosely referred to as the vector $\mathbf{y}$ and these responses have a distribution which depends on η . It is important that, conditional on given random component vector $\mathbf{u}$, the response has a log-likelihood function $l_1(\eta,\gamma\,|\,\mathbf{u})$. The parameter γ is additional to those occurring in η and will often relate to the variance of the response. Not all problems have this extra parameter γ and it, together with ρ , are considered fixed and known in the initial development. The log-likelihood function l_1 is not necessarily the full likelihood function but may be a conditional likelihood, a partial likelihood or some other appropriate likelihood function. For a GLM, the full likelihood function would be appropriate while for multiple failure times the Cox partial likelihood would be used. In what follows we assume that the domain of this likelihood does not depend on η or γ.

3.3 ESTIMATION OF β, $\mathbf{u}$

The estimation procedure has been described in the previous chapter in section 2.3.2 which is briefly reviewed here. The distribution of the random component vector $\mathbf{u}$ is specified by a log-probability density function

$$(3.2) \qquad l_2(\phi,\rho) = \sum_{j=1}^{k} l_{2j}(\phi_j,\rho) \ .$$

Estimators are derived using a penalised likelihood approach where l_2 is a penalty function on random components $\mathbf{u}$ for the log-likelihood function l_1 . The approach chooses estimators $\hat{\beta}$, $\tilde{\mathbf{u}}$ to maximise $l = l_1 + l_2$ similar to the

BLUP principle. This function has been called the hierarchical likelihood by Lee and Nelder (1996). Using the first and second derivatives as presented in section 2.3.2, an iterative Newton-Raphson solution for $\tilde{\beta}$, $\tilde{u}$ of the equations $\partial l / \partial \beta = 0$, $\partial l / \partial u = 0$ begins at initial values β_0, u_0 and finds a first iteration β_1, u_1 given by

$$(3.3) \qquad \begin{bmatrix} \beta_1 - \beta_0 \\ u_1 - u_0 \end{bmatrix} = V^{-1} \begin{bmatrix} X' \\ Z' \end{bmatrix} t + V^{-1} \begin{bmatrix} 0 \\ s \end{bmatrix} \quad , \quad \text{where} \quad t = \partial l_1 / \partial \eta \ , \quad s = \partial l_2 / \partial u$$

and

$$V = - \begin{bmatrix} \partial^2 l / \partial \beta \partial \beta' & \partial^2 l / \partial \beta \partial u' \\ \partial^2 l / \partial u \partial \beta' & \partial^2 l / \partial u \partial u' \end{bmatrix} = \begin{bmatrix} X' \\ Z' \end{bmatrix} \mathbb{B} \begin{bmatrix} X & Z \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{A} \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} ,$$

$$V^{-1} = \begin{bmatrix} T_1 & T_2 \\ T_2' & T \end{bmatrix}$$

where $\mathbb{B} = -\partial^2 l_1 / \partial \eta \partial \eta'$, $\mathbb{A} = -\partial^2 l_2 / \partial u \partial u'$.

The method becomes iterative when the one-step iterations β_1, u_1 replace the initial values for a further iteration and so on. If $\mathbb{B}$ is replaced by $I_{y,u} = E_{y,u}(\mathbb{B})$ and $\mathbb{A}$ by $\mathcal{A} = E_u(\mathbb{A})$ then the Newton-Raphson method becomes the method of scoring. An intermediate position is to use $I_{y|u} = E_{y|u}(\mathbb{B})$ and $\mathbb{A}$ as the appropriate matrices. Note that, for generalised linear models, the matrices $I_{y|u}$ and $\mathbb{B}$ are the same and, if u is normally distributed $\mathbb{A}$ and $\mathcal{A}$ are the same.

To develop the properties of these estimators we consider the equations that would result if we started the iterations from the true values of β, u and considered that $\tilde{\beta}$, $\tilde{u}$ are obtained (at least approximately) as one step iterations much in the same spirit of Pregibon's (1981) one-step residuals. In that case we have approximately

36

$$(3.4)$$
$$\tilde{\beta} = \beta + (T_1 X' + T_2 Z') t + T_2 s$$

$$\tilde{u} = u + (T_2' X' + T Z') t + T s$$

We assume initially below that the intermediate matrices of $I_{y|u}$, $\mathbb{A}$ are used so that V and hence all T matrices are not dependent on y but are possibly dependent on u.

Since we are assuming that the domain of the likelihood function does not involve η we may write

$$E_{y|u}(t) = 0 \quad , \quad Var_{y|u}(t) = -E_{y|u}(\partial^2 l_1 / \partial \eta \partial \eta') = I_{y|u}$$

so that

$$E_{y|u}(\tilde{\beta}) = \beta + T_2 s \quad , \quad E_{y|u}(\tilde{u}) = u + T s$$

and

$$(3.5) \qquad Var_{y|u}\begin{bmatrix} \tilde{\beta} \\ \tilde{u} \end{bmatrix} = V^{-1} \begin{bmatrix} X' \\ Z' \end{bmatrix} I_{y|u} \begin{bmatrix} X & Z \end{bmatrix} V^{-1} = V^{-1} - V^{-1} \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{A} \end{bmatrix} V^{-1}$$

$$= \begin{bmatrix} T_1 - T_2 \mathbb{A} T_2' & T_2 - T_2 \mathbb{A} T \\ T_2' - T \mathbb{A} T_2' & T - T \mathbb{A} T \end{bmatrix}$$

This expression for the variance is exactly true when V is computed from $I_{y|u}$ and $\mathbb{A}$ instead of $I_{y,u}$ and $\mathcal{A}$, otherwise it is an approximation.

For $I_{y,u}$ and $\mathcal{A}$ used to construct V, the matrices T and $\mathcal{A}$ do not depend on y or u. If $E_u(s)=0$ then $E_{y,u}(\tilde{\beta})=\beta$ and using $Var_{y|u}(\tilde{\beta})=T_1-T_2\mathbb{A}T_2'$ as well as $Var_u[E_{y|u}(\tilde{\beta})]=T_2\mathbb{A}T_2'$ then $Var_{y,u}(\tilde{\beta})=T_1$. Thus $\tilde{\beta}$ is an unbiased estimator of β and has variance matrix T_1. We can expect that, if no small subset of observations is very influential, the formation of $\tilde{\beta}$ as a linear combination of components of t will ensure approximate normality of its distribution.

3.4 ESTIMATION OF ϕ VARIANCE COMPONENTS

Interest now centres on appropriate quadratic forms to estimate ϕ_j when these parameters are scalars representing the variance of the random components $\mathbf{u}_j$. Letting $\mathbb{A}_j = -\partial^2 l_2/\partial \mathbf{u}_j \partial \mathbf{u}_j' = -\partial^2 l_{2j}/\partial \mathbf{u}_j \partial \mathbf{u}_j'$, and noting that $-\partial^2 l_2/\partial \mathbf{u}_j \partial \mathbf{u}_r' = 0$ for $j \neq r$ because the $\mathbf{u}_j$ are independent, we have

$$\mathbb{A} = \begin{bmatrix} \mathbb{A}_1 & & & \\ & \mathbb{A}_2 & & \\ & & \cdot & \\ & & & \cdot \\ & & & & \mathbb{A}_k \end{bmatrix} . \quad \text{For some choice of } \mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & & & \\ & \mathbf{P}_2 & & \\ & & \cdot & \\ & & & \cdot \\ & & & & \mathbf{P}_k \end{bmatrix}$$

the quadratic form $\tilde{\mathbf{u}}_j' \mathbf{P}_j \tilde{\mathbf{u}}_j$ will have an expectation depending on

$$(3.6) \qquad E_{y|u}(\tilde{\mathbf{u}}) = \mathbf{u} + \mathbf{Ts} = \mathbf{z} \quad , \quad \text{Var}_{y|u}(\tilde{\mathbf{u}}) = \mathbf{T} - \mathbf{TAT} = \mathbf{R} .$$

Letting $\mathbf{z}' = [\mathbf{z}_1', \mathbf{z}_2', ..., \mathbf{z}_k']$, $\mathbf{R} = [\mathbf{R}_{jr}]$ be partitions of $\mathbf{z}$ and $\mathbf{R}$ conformal to the partitions of $\mathbf{u}$, gives

$$(3.7) \qquad E_{y|u}(\tilde{\mathbf{u}}_j' \mathbf{P}_j \tilde{\mathbf{u}}_j) = \mathbf{z}_j' \mathbf{P}_j \mathbf{z}_j + \text{tr}(\mathbf{P}_j \mathbf{R}_{jj}) .$$

Note that if $E_u(\mathbf{s}) = 0$ then $E_u(\mathbf{z}) = 0$ so that

$$(3.8) \qquad E_{y,u}(\tilde{\mathbf{u}}_j' \mathbf{P}_j \tilde{\mathbf{u}}_j) = \text{tr}[\mathbf{P}_j E_u(\mathbf{z}_j \mathbf{z}_j')] + \text{tr}(\mathbf{P}_j \mathbf{R}_{jj})$$

where it is taken that $\mathbf{P}$ and $\mathbf{R}$ are not functions of $\mathbf{u}$.

If we argue that higher moments of $\tilde{\mathbf{u}}_j' \mathbf{P}_j \tilde{\mathbf{u}}_j$ can be obtained by assuming approximate normality for $\tilde{\mathbf{u}}_j$ then

$$\text{Var}_{y|u}(\tilde{\mathbf{u}}_j' \mathbf{P}_j \tilde{\mathbf{u}}_j) = 4\mathbf{z}_j' \mathbf{P}_j \mathbf{R}_{jj} \mathbf{P}_j \mathbf{z}_j + 2 \, \text{tr}(\mathbf{P}_j \mathbf{R}_{jj})^2$$

$$E_u \text{Var}_{y|u}(\tilde{\mathbf{u}}_j' \mathbf{P}_j \tilde{\mathbf{u}}_j) = 4 \, \text{tr}[\mathbf{P}_j \mathbf{R}_{jj} \mathbf{P}_j E_u(\mathbf{z}_j \mathbf{z}_j')] + 2 \, \text{tr}(\mathbf{P}_j \mathbf{R}_{jj})^2$$

$$\text{Var}_u[E_{y|u}(\tilde{\mathbf{u}}_j' \mathbf{P}_j \tilde{\mathbf{u}}_j)] = 2 \, \text{tr}[\mathbf{P}_j E_u(\mathbf{z}_j \mathbf{z}_j')]^2$$

giving

$$\text{Var}_{y,u}(\tilde{\mathbf{u}}_j' \mathbf{P}_j \tilde{\mathbf{u}}_j) = 4\text{tr}[\mathbf{P}_j \mathbf{R}_{jj} \mathbf{P}_j E_u(\mathbf{z}_j \mathbf{z}_j')] + 2\text{tr}(\mathbf{P}_j \mathbf{R}_{jj})^2 + 2\text{tr}[\mathbf{P}_j E_u(\mathbf{z}_j \mathbf{z}_j')]^2$$

(3.9)

$$= 2 \, \text{tr}[\mathbf{P}_j \mathbf{R}_{jj} + \mathbf{P}_j E_u(\mathbf{z}_j \mathbf{z}_j')]^2$$

An unbiased estimating equation for ϕ_j is constructed by equating $\tilde{\mathbf{u}}_j'\mathbf{P}_j\tilde{\mathbf{u}}_j$ to its expectation $E_{y,u}(\tilde{\mathbf{u}}_j'\mathbf{P}_j\tilde{\mathbf{u}}_j)$ as evaluated above for an appropriate choice of $\mathbf{P}_j$. Often $\mathbf{P}_j$ will be chosen as an identity matrix. Progress beyond this point of development depends on the choice of the model l_2 for the random components. Section 3.6 gives further development when the random component $\mathbf{u}$ is normally distributed.

An alternative approach is to let $\tilde{\mathbf{s}}$ be the value of $\mathbf{s}=\partial l_2/\partial \mathbf{u}$ at $\tilde{\mathbf{u}}$. Thus

$$(3.10) \qquad \tilde{\mathbf{s}} \simeq \mathbf{s} - \mathbb{A}(\tilde{\mathbf{u}}-\mathbf{u}) = (\mathbf{I}-\mathbb{A}\mathbf{T})\mathbf{s} - \mathbb{A}(\mathbf{T}_2'\mathbf{X}'+\mathbf{T}\mathbf{Z}')\mathbf{t}$$

giving

$$E_{y|u}(\tilde{\mathbf{s}}) = (\mathbf{I}-\mathbb{A}\mathbf{T})\mathbf{s} = \mathbf{z}^* \quad ,$$

$$\text{Var}_{y|u}(\tilde{\mathbf{s}}) = \mathbb{A}\mathbf{R}\mathbb{A} \quad , \quad E_u(\mathbf{z}^*\mathbf{z}^{*\prime}) = \mathbb{A}-\mathbb{A}\mathbf{T}\mathbb{A}-\mathbb{A}\mathbf{R}\mathbb{A} \quad .$$

If $\mathbf{z}^{*\prime}=[\mathbf{z}_1^{*\prime},\mathbf{z}_2^{*\prime},...,\mathbf{z}_k^{*\prime}]$, $\tilde{\mathbf{s}}'=[\tilde{\mathbf{s}}_1',\tilde{\mathbf{s}}_2',...,\tilde{\mathbf{s}}_k']$ then

$$E_{y|u}(\tilde{\mathbf{s}}_j'\mathbb{A}_j^{-1}\tilde{\mathbf{s}}_j') = \mathbf{z}_j^{*\prime}\mathbb{A}_j^{-1}\mathbf{z}_j^* + \text{tr}(\mathbf{R}_{jj}\mathbb{A}_j) \quad ,$$

$$E_{y,u}(\tilde{\mathbf{s}}_j'\mathbb{A}_j^{-1}\tilde{\mathbf{s}}_j') = v_j-\text{tr}(\mathbf{T}_{jj}\mathbb{A}_j) \quad .$$

Equating $\tilde{\mathbf{s}}_j'\mathbb{A}_j^{-1}\tilde{\mathbf{s}}_j'$ to $v_j-\text{tr}(\mathbf{T}_{jj}\mathbb{A}_j)$ leads to an unbiased estimating equation for ϕ_j. In a similar manner to previous development we may find

$$(3.11) \qquad \text{Var}_{y,u}(\tilde{\mathbf{s}}_j'\mathbb{A}_j^{-1}\tilde{\mathbf{s}}_j') = 2\ \text{tr}(\mathbf{I}_{v_j}-\mathbf{T}_{jj}\mathbb{A}_j)^2 \quad .$$

3.5 ESTIMATION OF THE γ VARIANCE PARAMETERS

For problems in which γ is a scalar and represents some form of variability of the observations, an estimator can often be constructed by equating $\tilde{\mathbf{t}}'[I_{y|u}]^{-1}\tilde{\mathbf{t}}$ to its expectation, where $\tilde{\mathbf{t}}$ is the value of $\mathbf{t}$ at $\beta=\tilde{\beta}$, $\mathbf{u}=\tilde{\mathbf{u}}$. For

more general types of parameter the equating of $\partial l/\partial \gamma$ to its expectation can be explored. Similarly the equating of $\partial l/\partial \rho$ to its expectation can be used to find an estimating equation for ρ. The remainder of this Chapter considers the simpler problem of there being no parameter ρ. It has been included in the general discussion to indicate how to proceed when such correlational parameters are present in the problem. We now go on to consider the estimation of γ .

An approximate expansion of $\tilde{t}$ about the true value of t at corresponding true β, u is

$$\tilde{t} \simeq t - I_{y|u}[X \ Z]\begin{bmatrix} \tilde{\beta}-\beta \\ \tilde{u}-u \end{bmatrix} .$$

Using the moments of $\tilde{\beta}$, $\tilde{u}$ for given u we have

$$(3.12) \qquad E_{y|u}(\tilde{t}) = -I_{y|u}(XT_2+ZT)s ,$$

$$(3.13) \qquad \mathrm{Var}_{y|u}(\tilde{t}) = I_{y|u} - I_{y|u}[X \ Z]\left(V^{-1}+V^{-1}\begin{bmatrix} 0 & 0 \\ 0 & A \end{bmatrix}V^{-1}\right)\begin{bmatrix} X' \\ Z' \end{bmatrix}I_{y|u} .$$

Using the formula $E_{y|u}(\tilde{t}'I_{y|u}^{-1}\tilde{t}) = E_{y|u}(\tilde{t}')I_{y|u}^{-1}E_{y|u}(\tilde{t}) + \mathrm{tr}\ I_{y|u}^{-1}\mathrm{Var}_{y|u}(\tilde{t})$ and noting that

$$X'I_{y|u}(XT_2+ZT) = 0 \quad , \quad Z'I_{y|u}(XT_2+ZT) = I_{v.}-AT$$

where v_j is the dimension of u_j and $v_.$ is the sum of such dimensions, gives

$$E_{y|u}(\tilde{t}'I_{y|u}^{-1}\tilde{t}) = s'(T-TAT)s + n - v - v_. + \mathrm{tr}(AT)^2 .$$

Since $E_u(s)=0$, $\mathrm{Var}_u(s)=A$ we have

$$(3.14) \qquad E_{y,u}(\tilde{t}'I_{y|u}^{-1}\tilde{t}) = n - (v+v_.) + \mathrm{tr}(AT) .$$

Equating $\tilde{t}'I_{y|u}^{-1}\tilde{t}$ to $n-v-v_.+\mathrm{tr}(AT)$ may be used to find an unbiased estimator of γ.

If $\tilde{t}$ is approximately normally distributed then we may find the variance using standard normal theory as

$$\text{Var}_{y|u}(\tilde{t}'\hat{I}_{y|u}^{-1}\tilde{t}) = 2(n-v-v_*)+ 2\ \text{tr}(\text{A}T)^4+ 4\ (s')[2T\text{A}T-3(T\text{A})^2T+(T\text{A})^3T](s)$$

$$\text{Var}_u E_{y,u}(\tilde{t}'\hat{I}_{y|u}^{-1}\tilde{t}) = 2\ \text{tr}(\text{A}T)^2 - 4\ \text{tr}(\text{A}T)^3 + 2\ \text{tr}(\text{A}T)^4$$

giving

(3.15) $\quad \text{Var}_{y,u}(\tilde{t}'\hat{I}_{y|u}^{-1}\tilde{t}) = 2(n-v-v_*) + 10\ \text{tr}(\text{A}T)^2 - 16\ \text{tr}(\text{A}T)^3 + 8\ \text{tr}(\text{A}T)^4.$

3.6 NORMALLY DISTRIBUTED RANDOM COMPONENTS

If now we specify that $\mathbf{u}_j$ are independent $N(\mathbf{0},\phi_j\mathbf{I})$ where $\mathbf{I}$ is the identity matrix of order v_j then l_2 is given by the sum of

(3.16) $\qquad l_{2j} = -(1/2)[v_j\ln(2\pi\phi_j) + \phi_j^{-1}\mathbf{u}_j'\mathbf{u}_j]$

giving $\mathbf{s}=-\text{A}\mathbf{u}$ with $\text{A}_j=\phi_j^{-1}\mathbf{I}$ so that

$$\mathbf{z} = \mathbf{u}-T\text{A}\mathbf{u} \quad , \quad E_u(\mathbf{z}\mathbf{z}') = \text{A}^{-1}-T-\mathbf{R}.$$

Choosing $\mathbf{P}_j$ as the identity matrix of order v_j gives

$$E_{y,u}(\tilde{\mathbf{u}}_j'\tilde{\mathbf{u}}_j) = \text{tr}[(j,j)\ \text{block of }\text{A}^{-1}-T-\mathbf{R}] + \text{tr}[(j,j)\ \text{block of }\ \mathbf{R}]$$

$$= \phi_j v_j - \text{tr}\ \mathbf{T}_{jj}$$

An unbiased estimating equation for ϕ_j is then

(3.17) $\qquad \tilde{\phi}_j = (\tilde{\mathbf{u}}_j'\tilde{\mathbf{u}}_j + \text{tr}\ \mathbf{T}_{jj})/v_j$

and this estimator is identical to the one derived as the REML estimator by Schall (1991) and by McGilchrist (1994) using somewhat different arguments. The method of derivation is similar to that used by Gilmour et al. (1985).

The variance of $\tilde{\mathbf{u}}_j'\tilde{\mathbf{u}}_j$ may be obtained from the previous expression for $\text{Var}_{y,u}(\tilde{\mathbf{u}}_j'\mathbf{P}_j\tilde{\mathbf{u}}_j)$ by putting $\mathbf{P}_j=\mathbf{I}$ and $E_u(\mathbf{z}_j\mathbf{z}_j') = \phi_j\mathbf{I}-\mathbf{T}_{jj}-\mathbf{R}_{jj}$ giving

41

$$\text{Var}_{y,u}(\tilde{\mathbf{u}}_j'\tilde{\mathbf{u}}_j) = 2 \, \text{tr}[(\phi_j \mathbf{I} - \mathbf{T}_{jj})^2] \quad \text{from which}$$

$$(3.18) \qquad \text{Var } \tilde{\phi}_j = 2 v_j^{-2} \text{tr}[(\phi_j \mathbf{I} - \mathbf{T}_{jj})^2].$$

3.7 APPLICATIONS

For this section, applications are considered that conform to the structure laid out in section 3.3 and have a random component vector which can be partitioned into independent component vectors $\mathbf{u}_j$ distributed as $N(\mathbf{0}, \phi_j \mathbf{I})$. The structure of section 3.2 requires that the outcome variables must have a loglikelihood function which can be written as $l_1(\eta, \gamma \mid \mathbf{u})$, where $\eta = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ and all parameters apart from γ and those describing the distribution of $\mathbf{u}$ enter into the problem through η . The only variation in the method in moving from one such problem to another is in computing the derivatives

$$\mathbf{t} = \partial l_1 / \partial \eta \quad , \quad \mathbb{B} = -\partial^2 l_1 / \partial \eta \, \partial \eta' \ .$$

Preferrably $\mathbb{B}$ should be replaced by its expectation $I_{y \mid u}$ or $I_{y,u}$.

3.7.1 Generalised Linear Mixed Models

Conditional on fixed $\mathbf{u}$ the generalised linear mixed model considers observations Y_i having a distribution which belongs to the exponential family and the loglikelihood is

$$l_1(\eta, \gamma \mid \mathbf{u}) = \sum_{i=1}^{n} \left(\frac{Y_i \theta_i - b(\theta_i)}{a_i(\gamma)} + c(Y_i, \gamma) \right) \ ,$$

$$E_{y \mid u}(Y_i) = \mu_i = b'(\theta_i) \ , \ \text{Var}_{y \mid u} Y_i = \sigma_i^2 = a_i(\gamma) b''(\theta_i).$$

For standard generalised linear models, the variance term can be written $\sigma_i^2 = \gamma q(\mu_i)$. For binomial and Poisson models $\gamma = 1$ but for normal and gamma models there is a separate unknown scalar parameter γ representing extra

42

variability of the observations associated with those distributions distinct from that which is a function of the observation mean.

The inverse link function is $\mu_i = h(\eta_i)$ with $\eta_i = x_i'\beta + z_i'u$. The component of $t = \partial l_i / \partial \eta$ corresponding to η_i is $h'(\eta_i)[Y_i - \mu_i]/\sigma_i^2$ and $I_{y\,|\,u} = I_{y,u}$ is a diagonal matrix with i^{th} term $[h'(\eta_i)]^2/\sigma_i^2$. The variances σ_i^2 may depend on η_i through $\sigma_i^2 = \gamma q(\mu_i) = \gamma q[h(\eta_i)]$. When σ_i^2 depends only on η, then the method of section 3.4 applies. Otherwise the extra parameter γ must be estimated.

The specific equations for estimating β, u may be expressed in terms of $\Delta\beta$, Δu which are the changes in β, u values between successive iterations. Usually the initial value of u can be taken to be a zero vector while initial β may be chosen by standard GLIM techniques. If u is distributed as $N(0, A^{-1})$ then

$$V \begin{bmatrix} \Delta\beta \\ \Delta u \end{bmatrix} = \begin{bmatrix} X' \\ Z' \end{bmatrix} t + \begin{bmatrix} 0 \\ -Au \end{bmatrix} \quad , \quad V = \begin{bmatrix} X' \\ Z' \end{bmatrix} I_{y\,|\,u} \begin{bmatrix} X & Z \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & A \end{bmatrix}$$

with $I_{y\,|\,u} = B = \text{diag}\left([h'(\eta_i)]^2/\sigma_i^2\right)$, $t = \left(h'(\eta_i)[Y_i - \mu_i]/\sigma_i^2\right)$. In the case of the generalised linear mixed model, the parameter σ_i^2 does not cancel from the equations, as it does when random components are not present, so that initial estimates of γ, ϕ_j or, more simply ϕ_j/γ, are required. An initial estimate of unity will often suffice. Following convergence of the iterative estimation of β, u to $\tilde{\beta}$, $\tilde{u}$ for fixed γ, ϕ_j, these latter parameters must then be estimated. For $A = \text{diag}[A_j]$, $A_j = \phi_j^{-1}I$ as in section 3.4, then

(3.19) $\tilde{\phi}_j = (\tilde{u}_j'\tilde{u}_j + \text{tr } T_{jj})/v_j$

43

The estimation of γ has been foreshadowed in section 3.5 as being obtained by equating

$$\tilde{t}' \tilde{I}_{y|u}^{-1} \tilde{t} = \sum_{i=1}^{n} (Y_i - \tilde{\mu}_i)^2 / \gamma q(\tilde{\mu}_i)$$

to its expectation $n - v - v_. + \text{tr}(\mathbf{A}\mathbf{T})$. The estimator of γ is then

$$(3.20) \qquad \tilde{\gamma} = [n - v - v_. + \text{tr}(\mathbf{A}\mathbf{T})]^{-1} \sum_{i=1}^{n} (Y_i - \mu_i)^2 / q(\mu_i).$$

Since we use estimates $\tilde{\phi}_j$ in $\mathbf{A}$ then an alternative expression for $n - v - v_. + \text{tr}(\mathbf{A}\mathbf{T})$ is $n - v - \sum_{j=1}^{k} \tilde{\phi}_j^{-1} \tilde{u}_j' \tilde{u}_j$. This agrees with the expression for REML estimation obtained by Schall.

Note that the formula for the variance of $\tilde{t}' \tilde{I}_{y|u}^{-1} \tilde{t}$, obtained in section 3.5, may be used to find a standard error for the estimator of γ.

3.7.2 Random Component Hazard Models

The treatment of random component hazard models is essentially the same as that given in McGilchrist (1993, 1994) and differs from the application in section 3.5.1 only in the likelihood function l_1 for the observations given random components **u**. Since **u** is again taken to be normally distributed, the estimation of the variances ϕ_j of the random components is identical to the above. There is no parameter γ involved in l_1 which is the Cox (1975) partial likelihood of the failure/censoring times conditional on given random components/frailties. The expressions for l_1 and its derivatives are fully developed in the earlier papers.

44

3.7.3 Threshold Models

Where a response varible Y_i, $i=1,2,...,n$, can take on values $0,1,...,m$, a threshold model for the distribution of Y_i is

$$P(Y_i \leq y) = G(\gamma_y - \eta_i)$$

for some cumulative distribution function $G(.)$ and $\eta_i = x_i'\beta + z_i'u$ representing the mixed linear model. The parameters γ_y are the cut-point (or threshold) parameters which are translated up or down by the regression η_i. We may define $\gamma_{-1} = -\infty$, $\gamma_m = \infty$ so that

$$P(Y_i = y) = G(\gamma_y - \eta_i) - G(\gamma_{y-1} - \eta_i)$$

and if η_i contains a constant term, then we may take $\gamma_0 = 0$ to remove the lack of estimability arising from fact that any change in the regression constant could be compensated by a shift of all the cutpoint parameters by the same amount.

The distribution of $\mathbf{u}$ is specified by l_2 and is often taken to be normal as elsewhere in this paper but l_1 becomes

$$(3.21) \qquad l_1 = \ln \prod_i [G(\gamma_{y_i} - \eta_i) - G(\gamma_{y_i-1} - \eta_i)] \ .$$

In this case the cut-point parameters γ_y form extra parameters of l_1 which may be estimated by equating $\partial l_1 / \partial \gamma$ to zero. However, the vector parameter γ are similar to η in the way it enters the likelihood function and the BLUP equations may be extended to

$$(3.22) \qquad V \begin{bmatrix} \gamma_1 - \gamma_0 \\ \beta_1 - \beta_0 \\ \mathbf{u}_1 - \mathbf{u}_0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & X' \\ 0 & Z' \end{bmatrix} \begin{bmatrix} \partial l_1 / \partial \gamma \\ \partial l_1 / \partial \eta \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ s \end{bmatrix}$$

where $V = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}' \\ \mathbf{0} & \mathbf{Z}' \end{bmatrix} \begin{bmatrix} -\partial^2 l_1/\partial\gamma\partial\gamma' & -\partial^2 l_1/\partial\gamma\partial\eta' \\ -\partial^2 l_1/\partial\eta\partial\gamma' & -\partial^2 l_1/\partial\eta\partial\eta' \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \mathbf{Z} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbb{A} \end{bmatrix}$.

The method then proceeds along the lines of previous sections. Specific examples are given in Saei and McGilchrist (1996) and Saei et al. (1996).

3.7.4 Poisson-Gamma Model

Response variables Y_i , i=1,2,...,n are Poisson distributed with mean $\lambda_i = \exp(\eta_i)$, $\eta_i = x_i'\beta + z_i'u$ for u taken to be conditionally fixed. However, the components of u , viz. U_i are distributed as independent log-gamma variables such that $V_i = \exp(U_i)$ are gamma distributed with mean 1. Thus

$$l_1 = \sum_{i=1}^{n} [-\exp(\eta_i) + Y_i\eta_i - \ln Y_i!] \, ,$$

(3.23)

$$l_2 = \sum_{i=1}^{n} [\phi\ln(\phi) - \ln \Gamma(\phi) - \phi e^{U_i} + \phi U_i]$$

giving

$$\partial l_1/\partial\eta_i = Y_i - \exp(\eta_i) \, , \quad \partial^2 l_1/\partial\eta_i^2 = -\exp(\eta_i) \, ,$$

$$I_{y\,|\,u} = \text{diag}[\exp(\eta_i)] \, , \quad I_{y,u} = \mathbf{D} = \text{diag}[\exp(x_i'\beta)]$$

and

$$\partial l_2/\partial U_i = \phi(1-e^{U_i}) \, , \quad \partial^2 l_2/\partial U_i^2 = -\phi e^{U_i}$$

$$\mathbb{A} = \text{diag}(\phi e^{U_i}) \, , \quad \mathcal{A} = \phi\mathbf{I}$$

The scoring equations for estimating β, u are iterative from initial values β_0, u_0 to first iteration β_1, u_1 given by

(3.24) $V\begin{bmatrix} \beta_1 - \beta_0 \\ u_1 - u_0 \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix}(y - e^{\eta}) + \begin{bmatrix} \mathbf{0} \\ \phi(1 - e^{u}) \end{bmatrix}$,

$$V = \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix} \mathbf{D} \begin{bmatrix} \mathbf{X} & \mathbf{Z} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \phi\mathbf{I} \end{bmatrix}$$

where **y**, η, **u** are vectors formed from Y_i, η_i, U_i. From section 3.4 we have

$$E_{y,u}(\tilde{\mathbf{u}}'\tilde{\mathbf{u}}) = \text{tr}[E_u(\mathbf{zz}')] + \text{tr}(\mathbf{R})$$

where $\mathbf{z} = \mathbf{u}+\mathbf{Ts} = \mathbf{u}+\phi\mathbf{T}(1-e^{\mathbf{u}}) \simeq -(\mathbf{I}-\phi\mathbf{T})(1-e^{\mathbf{u}})$, $\mathbf{R} = \mathbf{T}-\phi\mathbf{T}^2$. This gives

$$E_{y,u}(\tilde{\mathbf{u}}'\tilde{\mathbf{u}}) = v_1/\phi - \text{tr }\mathbf{T} \quad, \quad v_1 = \text{dimension of } \mathbf{u} \text{ so that}$$

(3.25) $\qquad \tilde{\phi}^{-1} = (\tilde{\mathbf{u}}'\tilde{\mathbf{u}} + \text{tr }\mathbf{T})/v_1$.

Since $\mathbf{V}^{-1} = \begin{bmatrix} \mathbf{X'DX} & \mathbf{X'DZ} \\ \mathbf{Z'DX} & \mathbf{D}+\phi\mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} \cdot & \cdot \\ \cdot & \mathbf{T} \end{bmatrix}$ we may find

$$\mathbf{T} = \mathbf{D}_1 + \mathbf{D}_1\mathbf{Z'DX}[\mathbf{X'DX}-\mathbf{X'DZD}_1\mathbf{Z'DX}]^{-1}\mathbf{X'DZD}_1 \quad, \quad \mathbf{D}_1 = (\mathbf{D}+\phi\mathbf{I})^{-1} \quad.$$

We have also (assuming $\tilde{\mathbf{u}}$ approximately normal)

$$\text{Var}_{y,u}(\tilde{\mathbf{u}}'\tilde{\mathbf{u}}) = 2 \text{ tr}[E_u(\mathbf{zz}')+\mathbf{R}]^2 = 2 \text{ tr}[\phi^{-1}\mathbf{I}-\mathbf{T}]^2$$

giving

(3.26) $\qquad \text{Var } \tilde{\phi}^{-1} = 2 \text{ tr}[\phi^{-1}\mathbf{I}-\mathbf{T}]^2/v_1^2$.

If the alternative approach using $\tilde{\mathbf{s}}_j'\mathbf{A}_j^{-1}\tilde{\mathbf{s}}_j'$ had been used, as given in section 3.4, then the above would be altered to using $\exp(\tilde{\mathbf{u}}) - 1$ in place of $\tilde{\mathbf{u}}$ in the estimating equation but the remainder of the working would be the same.

3.8 SIMULATION SUPPORT

A small simulation study is reported here for the Poisson-gamma model. Its purpose is to test out and illustrate the method rather than give exhaustive simulations. Observations Y_{ij}, i=1,2,...,15; j=1,2 are generated according to the model: Given random components U_i, the Y_{ij} are Poisson distributed with parameters $\lambda_{ij} = \exp \eta_{ij}$, $\eta_{ij} = \beta_1+\beta_2 x_{ij}+U_i$. The random components

U_i are distributed according to $V_i = \exp(U_i)$ being gamma distributed with mean 1 and variance ϕ^{-1} as described in section 3.7.4. The x_{ij} are selected randomly as 0 or 1.

For each combination of parameter values β_1, β_2, ϕ^{-1} reported in Table 3.1, 100 data sets , each set containing 30 observations as described above, are generated and the three parameters are estimated. Convergence of estimates was not always obtained from arbitrarily selected initial values but changes to the initial values did produce convergence. We report averages of estimated values and

SE_1 = average of reported standard errors for each parameter estimate,

SE_2 = standard deviations of the estimates obtained from simulations.

Results are reported systematically in Table 3.1.

Table 3.1   Simulation Study for Poisson-gamma model.   True Values, Average Estimates of Parameters are given together with SE$_1$ and SE$_2$

Parameter	True value	Average estimate	SE$_1$	SE$_2$	Parameter	True value	Average estimate	SE$_1$	SE$_2$
Simulation 1					*Simulation* 4				
β_1	1.0	0.92	0.34	0.30	β_1	3.0	2.98	0.32	0.27
β_2	0.5	0.50	0.28	0.31	β_2	1.5	1.50	0.09	0.10
ϕ^{-1}	1.0	1.21	0.35	0.69	ϕ^{-1}	1.0	1.45	0.51	0.70
Simulation 2					*Simulation* 5				
β_1	1.0	0.96	0.25	0.28	β_1	3.0	2.98	0.20	0.17
β_2	0.5	0.48	0.26	0.29	β_2	1.5	1.49	0.09	0.09
ϕ^{-1}	0.5	0.47	0.13	0.31	ϕ^{-1}	0.5	0.53	0.18	0.23
Simulation 3					*Simulation* 6				
β_1	1.0	0.98	0.21	0.22	β_1	3.0	2.97	0.14	0.16
β_2	0.5	0.51	0.24	0.25	β_2	1.5	1.50	0.09	0.09
ϕ^{-1}	0.25	0.23	0.05	0.18	ϕ^{-1}	0.25	0.24	0.08	0.11

From Table 3.1, it is apparent that estimates show no appreciable bias for any of the parameters and there is very good agreement between SE$_1$ and SE$_2$ for all regression parameters.   This good agreement does not carry through for all estimators of ϕ^{-1}.   The method of computing the standard error of the ϕ^{-1} estimator depends on assuming that $\tilde{u}$ is approximately normal - specifically that third and fourth order moments agree with those for a normal distribution and are then functions of first and second order moments.   Clearly there are difficulties in this assumption for smaller values of ϕ^{-1}.   However, the results of the simulations support a high degree of confidence in the estimation process except that the standard error of the estimator of ϕ^{-1} may be understated.

CHAPTER FOUR

HIERARCHICAL GENERALIZED LINEAR MODELS : A COMPARISON BETWEEN POISSON-NORMAL AND POISSON-GAMMA MODELS

4.1 INTRODUCTION

In GLMMs the distribution of random components is often considered to be normal irrespective of the conditional distribution of the response variable y. Recently Lee and Nelder (1996) proposed a class of GLMMs which they referred to as hierarchical generalized linear models (HGLMs). In HGLMs the distribution of random components is not restricted to normal and may come from any arbitrary distribution, often the distribution conjugate to that of y. A generalization of Henderson's joint likelihood, called hierarchical or h-likelihood, is used for estimation in HGLMs. Some examples of HGLMs are Poisson-gamma, binomial-beta and gamma-inverse-gamma models.

In addition to discussing the theoretical aspects, these models were applied to several real datasets and the results were compared with that of GLMMs in Lee and Nelder's paper. It was recommended to decide the distribution of random components based on the nature of the data or the purpose of inference. However, the application was limited to selected datasets and may not be completely generalised. One of the discussants of Lee and Nelder's paper recommended to undertake more detailed comparison to examine whether there is any real gain for using non-normal distribution of the random effects.

The objective of the current chapter is to undertake a simulation based comparative study between GLMMs and HGLMs when the distribution of the

response variable is Poisson conditional on given random components but the random components themselves are allowed to have different distributions. As the conjugate distribution of Poisson is gamma the comparison was undertaken between the Poisson-normal and the Poisson-gamma models. Specifically we looked at the effect of model misspecification, that is, when the Poisson-normal is used for modelling while the dataset is generated by the Poisson-gamma model and vice versa.

4.2 HIERARCHICAL GENERALISED LINEAR MODELS

The structure of HGLMs is basically covered by the general framework of the linear mixed models presented in section 3.2 and will be reviewed only briefly here. In the structure discussed in section 3.2, the response vector $\mathbf{y}$ is assumed to have a distribution dependent on $\eta = \mathbf{X\beta+Zu}$ where $\mathbf{X\beta}$ is the fixed component and $\mathbf{Zu}$ is the random component with $\mathbf{u}$, a vector of random effects of one or more components. As usual β is a vector of fixed coefficients, and $\mathbf{X}$ and $\mathbf{Z}$ are design matrices. The notation and related model assumptions are as discussed in section 3.2.

The conditional log-likelihood for $\mathbf{y}|\mathbf{u}$ has the GLM form as discussed in section 3.7.1.

$$(4.1) \qquad l_1(\eta,\gamma\,|\,\mathbf{u}) = \sum_{i=1}^{n} \left(\frac{Y_i\theta_i - b(\theta_i)}{a_i(\gamma)} + c(Y_i,\gamma) \right) ,$$

where θ_i denotes the canonical parameter and γ is the dispersion parameter. If the conditional distribution of $\mathbf{y}$ given $\mathbf{u}$ is μ then $\eta = g(\mu)$ with appropriate link function $g(.)$. The random component vector $\mathbf{u}$ has a log-probability (density) function $l_2(\phi) = \sum_{j=1}^{k} l_2(\phi_j)$.

51

Then hierarchical likelihood or h-likelihood, h, is defined as $h = l_1 + l_2$ which is the same as the penalised likelihood function l defined in Chapter 3. In GLMMs l_2 is always the log-likelihood of a normal variate while in HGLMs l_2 can take various forms. The distribution of u can have any distribution that is conjugate to the distribution of y where $u=g(v)$ for some strictly monotonic function of v. For example, if Y_i given u is binomial distributed then $v = \exp(u)/\{1+\exp(u)\}$ is considered as beta and the model is called binomial-beta HGLM. Similarly, if the distribution of Y_i given u is gamma then $v = \exp(u)$ is considered as inverse-gamma distributed and the model is called a gamma-inverse-gamma HGLM. When both Y_i and u are normal the model becomes the standard mixed model with normal error and the likelihood converts to Henderson's Joint likelihood.

The Estimation of β, u and ϕ are derived by maximising this penalised likelihood function $h=l$ as discussed in sections 3.3 and 3.4. Lee and Nelder called the estimates maximum h-likelihood estimates (MHLEs) as these are derived from maximising the h-likelihood.

4.2.1 Estimation in Poisson-Gamma and Poisson Normal Models

In both the Poisson-Gamma and the Poisson-normal models, distribution of the response variables Y_i, conditional on given random components u, are assumed to be Poisson distributed with mean $\lambda_i=\exp(\eta_i)$, where $\eta_i=x_i'\beta+z_i'u$. But the difference is that the distribution of the components, u, viz. U_i are considered as $N(0,\phi I)$ in the Poisson-normal model and $G(1,\phi^{-1}I)$ in the Poisson-gamma model.

The details of estimation strategy in the Poisson-gamma model are presented in section 3.7.4 which will not be repeated here.

In the Poisson-normal model the likelihood function l_1 , corresponding differentials $\partial l_1/\partial \eta_i$, $\partial^2 l_1/\partial \eta_i^2$ and information matrices $I_{y|u}$ and $I_{y,u}$ are exactly the same as presented in 3.7.4 for the Poisson-gamma model. But the log-likelihood l_2 is different. The function l_2 and the related expressions are given as follows.

$$l_2 = \text{constant} - (1/2)\sum_{i=1}^{n} \{v_i \ln(2\pi\phi) + \phi^{-1}U_i^2\}$$

giving

$$\partial l_2/\partial U_i = -\phi^{-1}U_i \ , \ \partial^2 l_2/\partial U_i^2 = -\phi^{-1} \ ,$$

(4.2)

$$\mathbf{A} = \text{diag}(\phi^{-1}\mathbf{I}) \ , \ \mathcal{A} = \phi^{-1}\mathbf{I}$$

Consequently the form of the scoring equations are slightly different.

(4.3) $$\mathbf{V}\begin{bmatrix} \beta_1 - \beta_0 \\ \mathbf{u}_1 - \mathbf{u}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix}(y - e^{\eta_0}) + \begin{bmatrix} \mathbf{0} \\ -\phi^{-1}\mathbf{u}_0 \end{bmatrix} \ ,$$

with $$\mathbf{V} = \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix}\mathbf{D}\begin{bmatrix} \mathbf{X} & \mathbf{Z} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \phi^{-1}\mathbf{I} \end{bmatrix} \ .$$

The dispersion parameter ϕ, as opposed to ϕ^{-1} for the Poisson-gamma model, is estimated by

(4.4) $$\tilde{\phi} = (\tilde{\mathbf{u}}'\tilde{\mathbf{u}} + \text{tr } \mathbf{T})/v_1 \ ,$$

where $\mathbf{T} = \mathbf{T} = \mathbf{D}_1 + \mathbf{D}_1\mathbf{Z}'\mathbf{DX}[\mathbf{X}'\mathbf{DX} - \mathbf{X}'\mathbf{DZD}_1\mathbf{Z}'\mathbf{DX}]^{-1}\mathbf{X}'\mathbf{DZD}_1$, which is the same expression as given under section 3.7.4 but the expression for $\mathbf{D}_1$ used in deriving $\mathbf{T}$ is different, $\mathbf{D}_1 = (\mathbf{D} + \phi^{-1}\mathbf{I})^{-1}$. The estimate of ϕ is obtained under REML approach in both the Poisson-normal and Poisson-gamma methods.

The variance of the disperson parameter, $\text{var}(\tilde{\phi})$, is estimated from the information matrix I_{REML} as presented in Chapter 2 and not by the method developed in Chapter 3. For the purpose of comparison, it was considered more appropriate to use a method which is consistent with Lee and Nelder (1996). The same approach is used for the Poisson-gamma estimation method.

A deviance measure is used to compare the goodness-of-fit in individual models. The deviance in the mixed model is calculated by using the expected value of $\mathbf{y} \mid \mathbf{u}$ as follows:

$$(4.5) \qquad \text{Dev } (\mathbf{y};\eta) = 2 \sum \{\mathbf{y} \ln (\mathbf{y}/\tilde{\mu}) - (\mathbf{y} - \tilde{\mu})\}$$

where $\tilde{\mu} = \exp (\tilde{\eta})$.

4.3 SIMULATION

The observations Y_{ij} are generated by assuming the distribution to be Poisson with parameters $\lambda_{ij} = \exp \eta_{ij}$, $\eta_{ij} = \beta_1 + \beta_2 x_{ij} + U_i$, where U_i's are random components distributed as normal in the case of GLMM and, in the case of HGLM, $V_i = \exp(U_i)$ is distributed as gamma with mean 1 and variance ϕ^{-1}.

In each simulation a dataset of 30 observations are generated with 2 observations in each block and consequently 15 effects in the random component $\mathbf{u}$. The x_{ij} are randomly selected as 0 or 1. The simulation was repeated 100 times for each combination of β_1, β_2 and ϕ or ϕ^{-1}. The averages of estimated values are presented with average of reported standard errors for each parameter estimate (SE_1) and standard deviations of the estimates obtained from simulations (SE_2).

4.4 RESULTS

The dispersion parameter of random component is ϕ in the Poisson-normal model and ϕ^{-1} in the Poisson-gamma model. In presenting the results a common notation δ is used for dispersion parameter instead of ϕ or ϕ^{-1}. Table 4.1 presents the results of fitting the Poisson-normal model to the datasets generated by the Poisson-gamma model with a fixed set of β_1 and β_2 values and three different values of δ as shown in the table. Table 4.2 presents similar results from the datasets generated with the same set of δ values but a different set of β_1 and β_2 values.

Tables 4.3 and 4.4 presents the results from similar simulations with reverse procedure to that of Tables 4.1 and 4.2, i.e. the datasets were generated by the Poisson-normal model but the estimates were obtained by using the Poisson-gamma estimation method.

The results show that both methods tend to provide an unbiased estimate of β_2 irrespective of the simulation models. However, the estimates of intercept parameter β_1 tend to be different dependent on simulation models. The model which generates the dataset provides unbiased estimates while the other estimation model produces lower estimates. The reason for this difference in intercept estimate, as explained by Lee and Nelder (1996), is due to the difference in E(Y). Under GLMM, $E(Y) \neq \mu$ but under HGLM, $E(Y) = \mu$.

55

Table 4.1: Comparison of Poisson-Gamma and Poisson-Normal Models using datasets simulated by Poisson-Gamma model.

Simul-ation	Para-meter	True Value	Poisson-Gamma			Poisson-Normal		
			Average estimate	SE_1	SE_2	Average estimate	SE_1	SE_2
1	β_1	1.0	1.01	0.33	0.28	0.66	0.32	0.31
	β_2	0.5	0.50	0.27	0.29	0.50	0.27	0.28
	δ	1.0	1.20	0.44	0.53	0.99	0.41	0.57
2	β_1	1.0	1.03	0.27	0.30	0.83	0.26	0.29
	β_2	0.5	0.48	0.25	0.29	0.50	0.27	0.26
	δ	0.5	0.59	0.24	0.26	0.51	0.22	0.28
3	β_1	1.0	0.99	0.22	0.24	0.89	0.22	0.24
	β_2	0.5	0.49	0.24	0.25	0.50	0.27	0.24
	δ	0.25	0.29	0.14	0.17	0.25	0.14	0.18

Table 4.2: Comparison of Poisson-Gamma and Poisson-Normal Models using datasets simulated by Poisson-Gamma model.

Simul-ation	Para-meter	True value	Poisson-Gamma			Poisson-Normal		
			Average estimate	SE_1	SE_2	Average estimate	SE_1	SE_2
1	β_1	3.0	2.89	0.30	0.25	2.40	0.31	0.28
	β_2	1.5	1.51	0.10	0.10	1.51	0.21	0.10
	δ	1.0	1.35	0.47	0.53	1.37	0.48	0.59
2	β_1	3.0	2.91	0.22	0.21	2.66	0.22	0.22
	β_2	1.5	1.51	0.09	0.10	1.51	0.24	0.10
	δ	0.5	0.67	0.23	0.27	0.68	0.24	0.28
3	β_1	3.0	2.98	0.15	0.15	2.88	0.15	0.15
	β_2	1.5	1.50	0.09	0.09	1.50	0.23	0.09
	δ	0.25	0.28	0.12	0.14	0.25	0.13	0.15

Table 4.3: Comparison of Poisson-Normal and Poisson-Gamma Models using datasets simulated by Poisson-Normal Model

Simul-ation	Para-meter	True value	Poisson-Normal			Poisson-Gamma		
			Average estimate	SE_1	SE_2	Average estimate	SE_1	SE_2
1	β_1	1.0	1.05	0.28	0.33	1.37	0.31	0.34
	β_2	0.5	0.51	0.26	0.22	0.51	0.22	0.22
	δ	1.0	0.94	0.28	0.30	1.37	0.45	0.55
2	β_1	1.0	1.03	0.24	0.25	1.20	0.25	0.26
	β_2	0.5	0.49	0.25	0.24	0.49	0.23	0.25
	δ	0.5	0.47	0.25	0.25	0.54	0.26	0.31
3	β_1	1.0	0.99	0.21	0.23	1.07	0.21	0.23
	β_2	0.5	0.51	0.25	0.26	0.52	0.23	0.27
	δ	0.25	0.24	0.15	0.15	0.26	0.14	0.18

Table 4.4: Comparison of Poisson-Normal and Poisson-Gamma Models using datasets simulated by Poisson-Normal Model

Simul-ation	Para-meter	True value	Poisson-Normal			Poisson-Gamma		
			Average estimate	SE_1	SE_2	Average estimate	SE_1	SE_2
1	β_1	3.0	3.00	0.25	0.25	3.27	0.29	0.25
	β_2	1.5	1.50	0.19	0.08	1.50	0.08	0.08
	δ	1.0	1.04	0.30	0.33	1.35	0.47	0.55
2	β_1	3.0	3.00	0.19	0.20	3.20	0.20	0.19
	β_2	1.5	1.51	0.20	0.08	1.51	0.08	0.09
	δ	0.5	0.50	0.16	0.16	0.51	0.17	0.21
3	β_1	3.0	2.99	0.14	0.14	3.10	0.15	0.15
	β_2	1.5	1.51	0.20	0.08	1.51	0.09	0.08
	δ	0.25	0.25	0.08	0.09	0.26	0.10	0.13

In the case of the main parameter of interest δ, when $\delta = 0.25$ both models give unbiased estimates irrespective of the simulation process but for higher values of δ, the Poisson-gamma estimation procedure tends to overestimate even when the observations are generated by the Poisson-gamma model. The Poisson-normal estimation method performs either better or at least similar to the Poisson-gamma estimation procedure. As expected the Poisson-normal estimation provides better estimates when the dataset is simulated by the Poisson-normal model. Even in the case when observations are generated by the Poisson-gamma model, the Poisson-normal estimation tends to give better estimates than the Poison-gamma method for the smaller set of fixed effect coefficients. While for the larger set of fixed coefficients, the Poisson-normal performs similar to the Poisson-gamma.

Standard errors of the estimates as measured by SE_1 and SE_2 appear to be very similar for β_1 and β_2 but tend to be slightly lower for δ under the Poisson-normal method. SE_2 is an approximate benchmark for SE_1.

Tables 4.5 and 4.6 present the actual results of model fittings for a sample of datasets selected from the above simulation. For each of the true parameter configurations used in the above tables, three datasets are selected randomly and the results of fitting the Poisson-normal and the Poisson-gamma models to each of these datasets are presented. Table 4.5 presents the results when observations are generated by the Poisson-gamma model while Table 4.6 shows the results for observations generated by the Poisson-normal model. The estimates obtained under both models indicate that, even though in most cases average estimates of δ tend to be similar, the individual estimates frequently become different for specific data sets.

58

4.5 DISCUSSION

The chapter compares the relative performance of the Poisson-normal and the Poisson-gamma estimation procedures for datasets simulated by both the Poisson-normal and the Poisson-gamma models. The study indicates that in terms of average of estimates over simulations, the Poisson-normal model performs either better or equivalent to the Poisson-gamma model irrespective of the model used for data simulation. Standard errors of the estimates under the Poisson-gamma method are also not lower than that of the Poisson-normal method. Therefore, in this instance of Poisson response variable, HGLM is not offering any definite improvement over GLMM. The average of estimates are not sensitive to the assumed distributional form of the random effect.

Table 4.5: Comparison of estimates under Poisson-Gamma and Poisson-Normal models for a randomly selected sample of individual datasets generated by Poisson-Gamma model.

True values			Poisson-Gamma				Poisson-Normal			
β_1	β_2	δ	β_1	β_2	δ	Dev	β_1	β_2	δ	Dev
1	0.5	1.00	1.13	0.44	0.44	22.7	0.99	0.45	0.41	23.1
			1.11	0.21	1.22	24.9	0.81	0.25	0.87	27.6
			1.23	0.47	1.58	53.0	0.89	0.43	0.92	23.5
1	0.5	0.50	0.75	0.47	0.25	22.5	0.71	0.46	0.21	23.8
			1.01	0.58	0.67	33.3	0.81	0.61	0.53	18.9
			0.83	0.88	0.54	39.9	0.66	0.89	0.49	27.0
1	0.5	0.25	1.03	0.45	0.25	15.6	0.96	0.44	0.26	14.9
			0.85	0.38	0.12	25.7	0.83	0.35	0.16	23.6
			1.12	0.39	0.38	13.7	1.02	0.41	0.30	16.3
3	1.5	1.00	2.91	1.42	0.57	16.7	2.68	1.43	0.58	07.9
			3.25	1.35	0.70	99.8	2.97	1.35	0.93	20.1
			2.66	1.55	1.55	141.5	2.22	1.51	1.16	20.6
3	1.5	0.50	2.96	1.23	0.22	11.2	2.87	1.23	0.21	11.1
			3.00	1.67	0.44	20.5	2.84	1.68	0.37	20.7
			2.93	1.48	0.67	33.0	2.72	1.48	0.56	16.9
3	1.5	0.25	2.92	1.49	0.31	23.7	2.81	1.49	0.30	23.1
			3.00	1.61	0.14	09.4	2.94	1.62	0.12	10.5
			3.08	1.45	0.17	12.9	3.01	1.45	0.16	13.1

Table 4.6: Comparison of estimates under Poisson-Gamma and Poisson-Normal models for a randomly selected sample of individual datasets generated by Poisson-Normal model.

True Values			Poisson-Normal				Poisson-Gamma			
β_1	β_2	δ	β_1	β_2	δ	Dev	β_1	β_2	δ	Dev
1	0.5	1.00	1.26	0.42	0.79	15.6	1.56	0.42	0.59	22.6
			1.14	0.51	0.47	27.0	1.29	0.53	0.40	29.4
			1.31	0.32	1.01	18.9	1.70	0.34	0.92	50.5
1	0.5	0.50	0.91	0.88	0.53	20.5	1.08	0.88	0.60	19.3
			0.90	0.79	0.48	21.5	1.03	0.92	0.30	62.3
			0.84	0.38	0.53	20.1	0.98	0.49	0.35	77.8
1	0.5	0.25	1.03	0.50	0.40	17.7	1.21	0.43	0.21	40.3
			0.84	0.71	0.21	19.3	0.89	0.73	0.23	18.2
			0.80	0.67	0.34	18.3	0.90	0.69	0.36	18.7
3	1.5	1.00	2.66	1.57	0.57	11.2	2.89	1.58	0.66	57.6
			2.96	1.47	1.11	12.7	3.41	1.47	1.28	18.5
			2.86	1.46	0.90	20.9	3.28	1.45	1.59	84.2
3	1.5	0.50	3.16	1.51	0.33	12.2	3.30	1.51	0.34	15.5
			3.01	1.47	0.58	11.9	3.25	1.46	0.71	34.6
			3.00	1.46	0.51	20.7	3.18	1.47	0.52	21.6
3	1.5	0.25	2.92	1.49	0.15	19.2	2.98	1.49	0.15	19.0
			2.96	1.54	0.22	12.1	3.05	1.54	0.23	12.8
			2.93	1.56	0.52	15.1	3.12	1.56	0.54	16.0

CHAPTER FIVE

MATCHED CASE CONTROL STUDIES WITH RANDOM EXPOSURE EFFECTS

5.1 INTRODUCTION

Analyses of matched case control studies such as Walter (1980), Miettinen (1969, 1970) compared the probabilities of exposure in cases and controls where subjects were either exposed or not exposed. Such probabilities were allowed to vary over the matched sets, thereby considering the possibility that numbers of exposed subjects would be overdispersed from the usual binomial variation and that there could be an association between the results for cases and controls within each matched set.

To a large extent, such analyses have been replaced by a conditional likelihood method when a logistic model can be assumed. The conditional likelihood method for logistic models was given in Cox (1970) and has been applied to matched case control studies in Breslow and Day (1980), Breslow (1982) and others. No overdispersion or correlation between results for the same matched set have as yet been included in this development. It is the intention of the chapter to explore this possibility.

5.2 MODEL AND NOTATION

Let Y_{ij} , j=1,2,...,n_i be the indicator variables for cases and controls in the ith matched set, where $Y_{ij}=1$ if the i,j element is a case and $Y_{ij}=0$ if the element is a control. The number of cases in the ith matched set is denoted by m_i with i=1,2,...,I. Cases and controls are matched for a vector of variables w_i and other potentially important regression variables are

collected into a vector $\mathbf{x}_{ij}$ corresponding to Y_{ij}. The exposure variable for i,j is denoted by Z_{ij}. Let $\pi_{ij} = P(Y_{ij}=1 \mid \mathbf{w}_i, \mathbf{x}_{ij}, Z_{ij})$.

If a logit model applies to π_{ij} then we may express dependence on matching variables, covariables and exposure variable in which the last has a coefficient which varies over the matched sets as

$$(5.1) \qquad \pi_{ij} = \exp(\alpha_i + \mathbf{w}_i'\gamma + \eta_{ij})/[1 + \exp(\alpha_i + \mathbf{w}_i'\gamma + \eta_{ij})]$$

where $\eta_{ij} = \mathbf{x}_{ij}'\beta + Z_{ij}(\tau + U_i)$ and U_i are distributed as independent $N(0,\phi)$. The random component U_i allows variation in the effect of exposure from one matched set to another and thus includes extra variation as well as association of observations within each matched set, reminiscent of the analysis given by Walter (1980).

5.3 Estimation

To estimate the parameters in the model described above, the GLMM technique is applied. As described in section 2.3.2, in GLMMs the response vector $\mathbf{y}$ has a distribution dependent on the vector $\eta = \mathbf{X}\beta + \mathbf{Z}\mathbf{u}$, β is the vector of fixed parameters and $\mathbf{u}$ is the vector of random components distributed as $N[\mathbf{0}, \mathbf{A}]$. The variance matrix $\mathbf{A}$ can depend on vector parameters ϕ and ρ but for the current application $\mathbf{A} = \phi\mathbf{I}$ and the notation will be specialised to this case. For the current application the likelihood function l_1 is taken to be a suitable conditional likelihood and l_2 is, as usual, the likelihood function for the normally distributed random components $\mathbf{u}$. The estimates are obtained by maximising the joint likelihood function $l = l_1 + l_2$ by using the Newton-Raphson iterative method. The estimation equation and the variance matrix $\mathbf{V}$ are as presented in (2.25) and (2.26).

63

5.3.1 Estimation in Matched Case-Control Data

In applying the above estimation technique to the matched case control problem, we let the vector of observations from the ith matched set be

$$\mathbf{y}_i' = [Y_{i1}, Y_{i2}, \ldots, Y_{in_i}] \ , \quad i=1,2,\ldots,I$$

so that the probability distribution of $\mathbf{y}_i$ is proportional to

$$\prod_{j=1}^{n_i} [\pi_{ij}/(1-\pi_{ij})]^{Y_{ij}} \ .$$

From this distribution we construct the conditional likelihood of the observations given that there are m_i cases in the ith matched set. If R_i denotes all possible $\mathbf{y}_i$ for which $\mathbf{1}'\mathbf{y}_i = m_i$ then the conditional probability distribution of $\mathbf{y}_i$ given $\mathbf{1}'\mathbf{y}_i = m_i$ is

$$\left(\prod_{j=1}^{n_i} [\pi_{ij}/(1-\pi_{ij})]^{Y_{ij}} \right) \bigg/ \left(\sum_{k \in R_i} \prod_{r=1}^{n_i} [\pi_{ir}/(1-\pi_{ir})]^{Y_{ikr}^*} \right)$$

where $\mathbf{y}_{ik}' = [Y_{ik1}^*, Y_{ik2}^*, \ldots, Y_{ikn_i}^*]$ is the vector of values in the kth of the possible values of $\mathbf{y}_i$ in R_i. Using the logistic model for π_{ij} given in section 2, the above conditional probability becomes

$$\left(\prod_{j=1}^{n_i} \exp(\alpha_i + \mathbf{w}_i'\gamma + \eta_{ij})Y_{ij} \right) \bigg/ \left(\sum_{k \in R_i} \prod_{r=1}^{n_i} \exp(\alpha_i + \mathbf{w}_i'\gamma + \eta_{ir})Y_{ikr}^* \right)$$

$$= \left(\prod_{j=1}^{n_i} \exp(\eta_{ij}Y_{ij}) \right) \bigg/ \left(\sum_{k \in R_i} \prod_{r=1}^{n_i} \exp(\eta_{ir}Y_{ikr}^*) \right)$$

and the conditional log-likelihood function is

$$(5.2) \qquad l_1 = \sum_{i=1}^{I} \left[\sum_{j=1}^{n_i} Y_{ij}\eta_{ij} - \log a_i \right]$$

where $a_i = \sum_{k \in R_i} \left[\exp \sum_{r} Y_{ikr}^* \eta_{ir} \right]$.

64

Letting $\quad b_{ij} = \sum_{k \in R_i} [Y^*_{ikj} \exp \sum_r Y^*_{ikr} \eta_{ir}] \quad$ and

$$c_{ijj'} = \sum_{k \in R_i} [Y^*_{ikj} Y^*_{ikj'} \exp \sum_r Y^*_{ikr} \eta_{ir}] \, ,$$

the first and second order derivatives may be expressed

$$\partial l_i / \partial \eta_{ij} = Y_{ij} - a_i^{-1} b_{ij} \quad , \quad \partial^2 l_i / \partial \eta_{ij} \partial \eta_{ij'} = a_i^{-2} b_{ij} b_{ij'} - a_i^{-1} c_{ijj'} \, .$$

Other mixed second order derivatives with respect to η_{ij} are zero.

For the special case of one case per matched set, the above expressions simplify to

$$a_i = \sum_r \exp \eta_{ir} \quad , \quad b_{ij} = \exp \eta_{ij} \quad , \quad c_{ijj'} = \delta_{jj'} \exp \eta_{ij}$$

where δ is the usual Kronecker delta.

For two cases per matched set the expressions for a_i, b_{ij} and $c_{ijj'}$ are as follows.

$$a_i = \sum_{j_1 = 1}^{n_i - 1} \sum_{j_2 > j_1}^{n_i} \exp (\eta_{ij_1} + \eta_{ij_2})$$

$$b_{ij} = \sum_{j \neq j'}^{n_i} \exp (\eta_{ij} + \eta_{ij'}) \, , \qquad c_{ijj'} = \delta_{jj'} \exp (\eta_{ij} + \eta_{ij'})$$

The above expressions for three cases per matched set are

$$a_i = \sum_{j_1 = 1}^{n_i - 2} \sum_{j_2 > j_1}^{n_i - 1} \sum_{j_3 > j_2}^{n_i} \exp (\eta_{ij_1} + \eta_{ij_2} + \eta_{ij_3})$$

$$b_{ij} = \sum_{j_1 \neq j}^{n_i} \sum_{j_2 \neq j_1 \neq j}^{n_i} \exp (\eta_{ij} + \eta_{ij_1} + \eta_{ij_2}) \, ,$$

$$c_{ijj'} = \delta_{ii'} \sum_{j_2 \neq j' \neq j}^{n_i} \exp (\eta_{ij} + \eta_{ij'} + \eta_{ij_2})$$

This can be generalized for c_i cases in the ith matched set as follows

$$a_i = \sum_{j_1=1}^{n_i-2} \sum_{j_2>j_1}^{n_i-c_i+2} \ldots \sum_{j_{c_i}>j_{c_i}-1}^{n_i} \exp(\eta_{ij_1} + \eta_{ij_2} + \ldots + \eta_{ij_{c_i}})$$

$$b_{ij} = \sum_{j_1 \neq j}^{n_i} \ldots \sum_{j_{c_i} \neq \ldots \neq j_1 \neq j}^{n_i} \exp(\eta_{ij} + \eta_{ij_1} + \ldots + \eta_{ij_{c_i-1}}) \, ,$$

$$c_{ijj'} = \delta_{ii'} \sum_{j_2 \neq j' \neq j}^{n_i} \ldots \sum_{j_{c_i-2} \neq \ldots j_2 \neq j' \neq j}^{n_i} \exp(\eta_{ij} + \eta_{ij'} + \eta_{ij_2} + \ldots + \eta_{ij_{c_i-2}})$$

5.3.2 Generalisation

The general method of estimation may be applied with a small change in notation and using

$$\eta_i' = [\eta_{i1}, \eta_{i2}, \ldots , \eta_{in_i}] \quad , \quad \eta' = [\eta_1', \eta_2', \ldots , \eta_I']$$
$$X_i' = [x_{i1}, x_{i2}, \ldots , x_{in_i}] \quad , \quad X' = [X_1', X_2', \ldots , X_I']$$
$$z_i' = [Z_{i1}, Z_{i2}, \ldots , Z_{in_i}] \quad , \quad z' = [z_1', z_2', \ldots , z_I']$$

$$Z' = \begin{bmatrix} z_1' & & & \\ & z_2' & & \\ & & \cdot & \\ & & & \cdot \\ & & & & z_I' \end{bmatrix} \quad , \quad u' = [U_1, U_2, \ldots , U_I]$$

The model in section 5.2 is then $\eta = X\beta + z\tau + Zu = [X \ z]\begin{bmatrix} \beta \\ \tau \end{bmatrix} + Zu$ which is consistent with the model given in the general theory above but with $[X \ z]$ in place of X. The BLUP equations are then

$$(5.3) \qquad V \begin{bmatrix} \Delta\beta \\ \Delta\tau \\ \Delta u \end{bmatrix} = \begin{bmatrix} X' \\ z' \\ Z' \end{bmatrix} \partial l_1/\partial\eta - \begin{bmatrix} 0 \\ 0 \\ \phi^{-1}u_0 \end{bmatrix} \, ,$$

66

$$V = \begin{bmatrix} X' \\ z' \\ Z' \end{bmatrix} [-\partial^2 l_1/\partial\eta\,\partial\eta'] \begin{bmatrix} X & z & Z \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \phi^{-1}I \end{bmatrix}$$

and similar changes are made to obtain ML and REML estimators of ϕ.

$$\text{If} \quad V = \begin{bmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{bmatrix} \quad , \quad V^{-1} = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & T \end{bmatrix} \quad \text{then}$$

the ML estimator for ϕ is $\tilde{u}'\tilde{u}/(v-r*)$, $r*=\phi^{-1}\text{tr }V_{33}^{-1}$. If V_{33}^{-1} is replaced by T then REML estimates are obtained.

5.3.3 Efficiency in Computation

Since $\partial^2 l_1/\partial\eta_{ij}\partial\eta_{i'j'}=0$ for $i \neq i'$, further analytic development of the above equations is possible resulting in faster computational method and considerable saving in space requirements. Let $B_i = -\partial^2 l_1/\partial\eta_i\partial\eta_i'$ and

$$C = \sum_{i=1}^{I} X_i'B_iX_i \, , \quad c_i = X_i'B_iz_i \, , \quad C_1 = [c_1, c_2, ..., c_I] \, , \quad d_i = z_i'B_iz_i,$$

$$D = \text{diag } [d_1, d_2, ..., d_I], \quad t = \sum_{i=1}^{I} X_i'\partial l_1/\partial\eta_i, \quad s_i = z_i'\partial l_1/\partial\eta_i,$$

$$s = [s_1, s_2..., s_I]'$$

giving

$$V = \begin{bmatrix} C & C_1 1 & C_1 \\ 1'C_1' & 1'D1 & 1'D \\ C_1' & D1 & D+\phi^{-1}I \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} X' \\ z' \\ Z' \end{bmatrix} \partial l_1/\partial\eta = \begin{bmatrix} t \\ 1's \\ s \end{bmatrix}.$$

Use of the general formula for inversion of partitioned symmetric matrices gives

$$\begin{bmatrix} \mathbf{1'D1} & \mathbf{1'D} \\ \mathbf{D1} & \mathbf{D}+\phi^{-1}\mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{D}_1 \end{bmatrix} + \phi m \begin{bmatrix} 1 \\ -\mathbf{d} \end{bmatrix} \begin{bmatrix} 1 & -\mathbf{d'} \end{bmatrix} \ ,$$

$$\mathbf{D}_1 = (\mathbf{D}+\phi^{-1}\mathbf{I})^{-1}, \quad \mathbf{d} = \mathbf{DD}_1\mathbf{1}, \quad m^{-1} = \mathbf{1'd}.$$

and then

$$\mathbf{V}^{-1} = \begin{bmatrix} 0 & 0 & 0 \\ 0' & 0 & 0' \\ 0 & 0 & \mathbf{D}_1 \end{bmatrix} + \phi m \begin{bmatrix} 0 \\ 1 \\ -\mathbf{d} \end{bmatrix} \begin{bmatrix} 0' & 1 & -\mathbf{d'} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ -m\mathbf{v'} \\ m\mathbf{d}\mathbf{v'}-\mathbf{D}_1\mathbf{C}_1' \end{bmatrix} \mathbf{G}^{-1} \begin{bmatrix} \mathbf{I} & -m\mathbf{v} & m\mathbf{v}\mathbf{d'}-\mathbf{C}_1\mathbf{D}_1 \end{bmatrix}$$

where $\mathbf{v} = \mathbf{C}_1\mathbf{D}_1\mathbf{1}$, $\mathbf{G} = \mathbf{C}-\mathbf{C}_1\mathbf{D}_1\mathbf{C}_1'-\phi^{-1}m\mathbf{v}\mathbf{v'}$. This gives

(5.4) $$\begin{bmatrix} \Delta\beta \\ \Delta\tau \\ \Delta\mathbf{u} \end{bmatrix} = \mathbf{V}^{-1} \begin{bmatrix} \mathbf{t} \\ \mathbf{1's} \\ \mathbf{s}-\phi^{-1}\mathbf{u}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{q} \\ -m m_2 \\ \mathbf{g}+m m_2\mathbf{d}-\mathbf{D}_1\mathbf{C}_1'\mathbf{q} \end{bmatrix}$$

where $\mathbf{g}=\mathbf{D}_1(\mathbf{s}-\phi^{-1}\mathbf{u}_0)$, $m_1=\phi^{-1}\mathbf{1'}(\mathbf{g}+\mathbf{u}_0)$, $\mathbf{q}=\mathbf{G}^{-1}(\mathbf{t}-m m_1\mathbf{v}-\mathbf{C}_1\mathbf{g})$, $m_2=\mathbf{v'q}-\phi m_1$.

To estimate ϕ after convergence of the estimation of β, τ, $\mathbf{u}$ we require

$$\mathrm{tr}\ \mathbf{V}_{22}^{-1} = \mathrm{tr}\ \mathbf{D}_1 \quad \text{and} \quad \mathrm{tr}(\mathbf{I}-\phi^{-1}\mathbf{V}_{22}^{-1})^2 = \mathrm{tr}(\mathbf{I}-\phi^{-1}\mathbf{D}_1)^2 = \mathrm{tr}(\mathbf{DD}_1)^2$$

while $\mathbf{T} = \mathbf{D}_1 + \phi m\mathbf{d}\mathbf{d'} + (m\mathbf{d}\mathbf{v'}-\mathbf{D}_1\mathbf{C}_1')\mathbf{G}^{-1}(m\mathbf{v}\mathbf{d'}-\mathbf{C}_1\mathbf{D}_1)$ giving

$$\mathrm{tr}\ \mathbf{T} = \mathrm{tr}\ \mathbf{D}_1 + \phi m\mathbf{d'd} + m^2\mathbf{d'dv'G}^{-1}\mathbf{v} - 2\mathbf{d'D}_1\mathbf{C}_1'\mathbf{G}^{-1}\mathbf{v} + \mathrm{tr}(\mathbf{D}_1^2\mathbf{C}_1'\mathbf{G}^{-1}\mathbf{C}_1)$$

5.4 SIMULATION SUPPORT

A simulation study was undertaken to investigate the relative performance of the proposed method of analysis with the usual conditional likelihood based analysis. The methods of analysis are referred to as Generalized Linear Mixed model (GLMM) and Generalized Linear Model (GLM) techniques respectively. A number of matched case-control data sets were generated corresponding to different exposure effects and various random effect coefficients. The model included exposure variable only and no other

regression variable was included. Once a number of cases and controls were generated for each matched category then appropriate number of cases and controls were randomly selected for matched analysis. The simulation was undertaken 200 times for each parameter configurations and the analyses were conducted using both GLMM and GLM methods. Tables 5.1 and 5.2 present the comparative results for single cases per matched set for various configurations of exposure and random effects coefficients while Tables 5.3 and 5.4 present similar results for two cases per matched set. The averages of estimated values are compared with the true values in the tables. The standard error SE_1 presented in the tables is the average of the estimated standard errors from each analysis while SE_2 is the standard deviation of estimated values over simulations.

5.4.1 Single Case and Multiple Controls

The comparative estimates from the applications of GLM and GLMM to 30 matched sets of size 5, 1 case and 4 controls, simulated with positive and negative exposure effects are presented in Tables 5.1 and 5.2 respectively. The results shows that the estimates of the exposure coefficient under both methods are very similar and both methods able to approximately reproduce the true values. However, the estimated standard errors SE_1 tend to be under estimated by GLM when compared with SE_2. The estimates of SE_1 under GLMM is either same or very close to SE_2.

Table 5.1: *The Comparison of Estimates Under GLM and GLMM for Datasets with Single Case and Four Controls per Matched Set Simulated with Positive Exposure Effects.*

Simul-ation	Method	Exposure Effect (β)				Dispersion Parameter (ϕ)			
		True Value	Average estimate	SE_1	SE_2	True Value	Average estimate	SE_1	SE_2
1	GLM	1.0	0.99	0.47	0.52	-	-	-	-
	GLMM	1.0	0.98	0.52	0.52	1.0	1.11	1.03	0.80
2	GLM	1.0	1.01	0.45	0.48	-	-	-	-
	GLMM	1.0	1.00	0.47	0.48	0.5	0.58	0.80	0.56
3	GLM	0.5	0.49	0.44	0.48	-	-	-	-
	GLMM	0.5	0.49	0.48	0.48	1.0	0.89	1.11	0.73
4	GLM	0.5	0.51	0.42	0.45	-	-	-	-
	GLMM	0.5	0.51	0.45	0.45	0.5	0.67	0.92	0.65

Table 5.2: *The Comparison of Estimates Under GLM and GLMM for Datasets with Single Case and Four Controls per Matched Set Simulated with Negative Exposure Effects.*

Simul-ation	Method	Exposure Effect (β)				Dispersion Parameter (ϕ)			
		True Value	Average estimate	SE_1	SE_2	True Value	Average estimate	SE_1	SE_2
1	GLM	-1.0	-0.98	0.47	0.52	-	-	-	-
	GLMM	-1.0	-0.98	0.50	0.52	1.0	0.80	0.98	0.62
2	GLM	-1.0	-1.00	0.49	0.51	-	-	-	-
	GLMM	-1.0	-1.00	0.50	0.51	0.5	0.46	0.63	0.56
3	GLM	-0.5	-0.49	0.43	0.45	-	-	-	-
	GLMM	-0.5	-0.49	0.46	0.45	1.0	0.82	0.84	0.53
4	GLM	-0.5	-0.49	0.43	0.45	-	-	-	-
	GLMM	-0.5	-0.49	0.45	0.45	0.5	0.49	0.67	0.50

5.4.2 Multiple Cases and Multiple controls

Tables 5.3 and 5.4 present similar comparative results as Tables 5.1 and 5.2 for 20 matched sets with 2 cases and 3 controls. The results tend to behave

similar to the case of matched set with single case as presented in previous tables. The estimates of SE_1 are lower under GLM in comparison with SE_2 while GLMM is providing estimates of SE_1 closer to SE_2.

Therefore, as expected the simulation results are showing that if GLM is used for analysing matched data in the presence of random exposure effects, then the estimated variation of the estimate of exposure effect would be less than the actual variation.

The estimates of dispersion parameter of random effects which is only available for GLMM are also presented in Tables 5.1 to 5.4. The average estimates and estimated standard errors are reasonable but not as good as fixed exposure effect which is often a general problem with the estimation in GLMM.

Table 5.3: The Comparison of Estimates Under GLM and GLMM for Datasets with Two Cases and Four Controls per Matched Set Simulated with Positive Exposure Effects.

Simul-ation	Method	Exposure Effect (β)				Dispersion Parameter (ϕ)			
		True Value	Average estimate	SE_1	SE_2	True Value	Average estimate	SE_1	SE_2
1	GLM	1.0	0.98	0.43	0.49	-	-	-	-
	GLMM	1.0	0.99	0.48	0.49	1.0	0.89	0.97	0.87
2	GLM	1.0	1.00	0.43	0.46	-	-	-	-
	GLMM	1.0	0.99	0.48	0.47	0.5	0.61	0.91	0.86
3	GLM	0.5	0.51	0.42	0.50	-	-	-	-
	GLMM	0.5	0.50	0.48	0.51	1.0	0.85	0.97	0.75
4	GLM	0.5	0.50	0.42	0.48	-	-	-	-
	GLMM	0.5	0.49	0.45	0.48	0.5	0.46	0.65	0.57

Table 5.4: The Comparison of Estimates Under GLM and GLMM for Datasets with Two Cases and Four Controls per Matched Set Simulated with Negative Exposure Effects.

Simul-ation	Method	Exposure Effect (β)				Dispersion Parameter (ϕ)			
		True Value	Average estimate	SE_1	SE_2	True Value	Average estimate	SE_1	SE_2
1	GLM	-1.0	-0.99	0.43	0.48	-	-	-	-
	GLMM	-1.0	-0.98	0.49	0.48	1.0	0.82	0.93	0.85
2	GLM	-1.0	-1.01	0.43	0.50	-	-	-	-
	GLMM	-1.0	-1.02	0.48	0.51	0.5	0.48	0.67	0.69
3	GLM	-0.5	-0.49	0.41	0.44	-	-	-	-
	GLMM	-0.5	-0.49	0.47	0.45	1.0	0.87	0.97	0.80
4	GLM	-0.5	-0.51	0.42	0.46	-	-	-	-
	GLMM	-0.5	-0.51	0.46	0.46	0.5	0.52	0.74	0.68

5.5 APPLICATIONS

In addition to the simulation study the method was applied to a couple of real datasets. The descriptions of the datasets and the results are presented below.

5.5.1 Los Angeles Study of Endometrial Cancer

Breslow and Day (1980) presented data from a matched case control study reported by Mack et al (1976) which is included in Appendix A (Table A3). The study identified 63 cases of endometrial cancer occurring in a retirement community near Los Angeles, California, USA 1971 to 1975 and each case was matched to four controls who were alive and living in the same retirement community. The variables used for matching were age, marital status and the timing of entering the community. The main objective was to study the effect of exogenous estrogens on the risk of endometrial cancer. In addition to information on exposure to oestrogens, the other information

collected were the history of gall bladder disease, hypertension, obesity, other drugs, etc. Breslow and Day (1980) analysed the dataset by using conditional likelihood with GLM. We analysed the same dataset for some of the exposure variables by including random exposure effects with GLMM.

Three comparisons were made with only one exposure variable in the model each time. The exposure variables included in different models are Estrogen, Gall bladder and Hypertension. As Table 5.5 shows, when the exposure effects of Estrogen and Gall bladder are investigated both GLM and GLMM produce the same estimates of exposure effects and corresponding estimates of standard errors as there are no randomness in exposure effects. The estimates of ϕ in both cases are very small (0.02). However, for Hypertension the estimates and standard errors tend to differ because of the presence of random variation in exposure effect which is reflected in the large estimate of ϕ (1.10). The table also presents the corresponding relative risks and confidence intervals. For Hypertension, the confidence interval is about 10% shorter under GLM than that of GLMM.

Table 5.5: Comparison of Estimates from GLM and GLMM with Matched Data from Los Angeles Study of Endometrial Cancer. Three Comparisons with Different but Single Exposure Variable in the Model Each Time.

Comparison	Risk Factor	Method	Est	SE	ϕ	Rel Risk	95% CI
1	Estrogen	GLM	2.07	0.42	-	7.92	3.48-18.05
		GLMM	2.07	0.42	0.02	7.92	3.48-18.05
2	Gall bladder	GLM	1.31	0.37	-	3.71	1.79-7.65
		GLMM	1.31	0.37	0.02	3.71	1.79-7.65
3	Hypertension	GLM	0.41	0.30	-	1.51	0.83-2.71
		GLMM	0.40	0.34	1.10	1.49	0.76-2.90

Table 5.6 presents the results of another comparison where the exposure variable Hypertension is included in the model with two other regression variables, viz. Gall bladder and Estrogen. The presence of random exposure effect is even more as the estimate of ϕ is 3.09. The differences in estimates and standard errors under GLM and GLMM are prominent not only for the exposure variable but also for the regression variables. A comparison between relative risks or confidence intervals would make it more obvious.

Table 5.6: Comparison of Estimates from GLM and GLMM with Matched Data from Los Angeles Study of Endometrial Cancer. One Exposure Variable with Two Other Regression Variables in the Model.

Method	Risk Factor	Est	SE	ϕ	Rel Risk	95% CI
GLM	Gall	1.28	0.41	-	3.60	1.61-8.03
	Estrogen	2.12	0.45	-	8.33	2.41-20.1
	Hyper	-0.04	0.34	-	0.96	0.49-1.87
GLMM	Gall	1.38	0.37	-	3.97	1.92-8.21
	Estrogen	2.29	0.47	-	9.87	3.93-24.8
	Hyper	-0.08	0.37	3.09	0.92	0.44-1.90

5.5.2 Low Birth Weight Study

Hosmer and Lemeshow (1989) reported a study that was conducted to investigate risk factors associated with giving birth to a low weight baby (less than 2500 grams). Each case, a mother who gave birth to a low weight baby, was matched with three controls selected from the mothers of the same age who gave birth to a normal weight baby. Twenty nine such matched sets were included in the study. We fit GLM with conditional likelihood and GLMM with random exposure effects to this dataset with the case and the first two controls from each set. The dataset is presented in Appendix A (Table A2). The risk factors included in the models are SMOKE (smoked during pregnancy),

UI (presence of uterine irritability), PTD (history of premature delivery), and LWD (mother's weight at last menstrual period is in the first quartile for the study group). Table 5.7 presents comparative results of GLM and GLMM where exposure to SMOKE was considered as random. The estimate of ϕ is 1.89, indicating the presence of considerable random exposure. The difference in estimated exposure (Smoke) effect is relatively small, 0.42 under GLM compared to 0.40 under GLMM, but the difference in standard errors is large, 0.52 compared with 0.62 under GLM and GLMM respectively, which is reflected in the estimates of confidence interval.

Table 5.7: Comparison Between GLM and GLMM Using Data From Low Birth Weight Study.

Method	Risk Factor	Est	SE	ϕ	Rel Risk	95% CI
GLM	PTD	0.37	0.56	-	1.45	0.48- 4.34
	UI	1.92	0.81	-	6.82	1.39-33.37
	LWD	0.30	0.54	-	1.35	0.47- 3.89
	Smoke	0.42	0.52	-	1.52	0.45- 4.22
GLMM	PTD	0.60	0.62	-	1.82	0.54- 6.14
	UI	1.97	0.86	-	7.17	1.33-38.69
	LWD	0.36	0.59	-	1.43	0.45- 4.46
	Smoke	0.40	0.62	1.89	1.49	0.44- 5.23

5.6 DISCUSSION

The possibility of applying GLMM technique to accommodate overdispersion in matched case-control studies is explored in this chapter. The estimation method is discussed and appropriate modifications in GLMM estimation equations are derived for its application to a conditional likelihood analysis as undertaken in matched case-control studies. In addition to providing exact expressions for the quantities required in the estimation process for common situations, a general algorithm is presented for efficient computation which will be particularly useful when the number of

matched sets is large. The validity of the proposed method is analysed by undertaking a simulation study and applying to some real datasets. Our assessment indicates that the method has the capability to improve efficiency in the analysis of matched case control data. Further investigation is required to develop similar exact methods for small samples and examine the implications on study design particularly on sample size determination.

CHAPTER SIX

ANALYSIS OF CONTINGENCY TABLES
WITH CLUSTERED OBSERVATIONS

6.1 INTRODUCTION

The standard analysis of contingency tables or cross-classified categorical data are often performed under the assumption that observations which contribute to the category counts are independent. However, in many practical situations this assumption does not hold as observations which contribute to contingency tables are often clustered or correlated. One common reason is that the method of sampling used in collecting data is not always simple random but rather surveys are often conducted by using more complex multi-stage or cluster sampling designs. As the members of the same primary sampling unit (PSU) or cluster tend to respond similarly, a correlation structure is introduced in the observed data. This can also happen with temporal or spatial data which are gathered at successive points in time or units in maps which are often close enough to be correlated. In that situation a standard analysis for testing independence of category classification by goodness-of-fit would be misleading as the analysis will underestimate the standard errors of the estimates.

Pearson chi-squared and the log-likelihood tests under the assumption of multinomial or product multinomial sampling are often employed for testing independence or goodness-of-fit in contingency tables. Cohen (1976), Altham (1976), Brier (1980), Fellegi (1980), and Rao and Scott (1981) discuss the effects of clustered or correlated data on test statistics for independence or goodness-of-fit in contingency tables. These papers describe methods to

deflate the Pearson's chi-square or likelihood ratio statistics in order to adjust for intra-cluster correlation. These methods apply to two-way tables only and mostly approximate.

Log-linear modelling is another approach to analysing contingency tables which has become more popular since it can easily handle multi-dimensional tables. This paper addresses the problem of how to analyse clustered data in contingency tables using the techniques of GLMM which incorporate both fixed and random cluster effects in log-linear models.

6.2 ANALYSIS OF CONTINGENCY TABLES

6.2.1 Sampling Models

Contingency tables are generally analysed under such sampling models as Poisson, multinomial or product multinomial distributions.

Under the Poisson assumption each cell in a cross classification table is assumed to follow a Poisson process where no a priori knowledge regarding the total number of observations in the table is considered. If f_{ij} are cell counts in a table with I rows and J columns, they are viewed as independent Poisson variables with the expected cell counts m_{ij} then likelihood function for the observed frequencies is given by

$$(6.1) \qquad \mathcal{L}(\mathbf{m},\mathbf{f}) = \prod_{ij} \frac{m_{ij}^{f_{ij}} e^{-m_{ij}}}{f_{ij}!} .$$

Under the multinomial assumption the sample size, i.e. the total number of observations in a table, is treated as fixed and the distribution of the

category counts given the fixed total is considered as multinomial. If f is the total number of observations and w_{ij} is the underlying probability of falling an observation in the i,j^{th} cell then the likelihood function is expressed as

$$(6.2) \qquad \mathcal{L}(w,f) = \frac{f!}{\prod\limits_{ij} f_{ij}!} \prod\limits_{ij} w_{ij}^{f_{ij}} \qquad \text{where} \quad \sum\limits_{ij} w_{ij} = 1$$

When each row in a table is treated as independent with a fixed sample size and the observations in different column categories are distributed as multinomial then the sampling model is called product multinomial. If $f_{i\bullet}$ is the row total for the ith row then the likelihood function can be expressed as

$$(6.3) \qquad \mathcal{L}(w,f) = \prod\limits_{i} \frac{f_{i\bullet}!}{\prod\limits_{j} f_{ij}!} \prod\limits_{j} w_{ij}^{m_{ij}} \qquad \text{where} \quad \sum\limits_{j} w_{ij} = 1$$

Birch (1963), Haberman (1974a), and others have shown that the maximum likelihood estimates (mle) of expected cell counts under the log-linear model are the same under any of the above three sampling assumptions provided a factor for fixed margins is included in the product multinomial model. In the rest of this chapter Poisson model will be adopted as the method of analysis.

6.2.2 Test of Independence

The test of independence in a contingency table is commonly undertaken by using Pearson's goodness-of-fit chi-square (Pearson 1904) or likelihood ratio chi-square (Wilks 1935).

Pearson's chi-square for testing goodness-of-fit, that is to compare observed with expected frequencies, is defined as

$$(6.4) \qquad X^2 = \sum_i \sum_j \frac{(f_{ij} - m_{ij})^2}{m_{ij}}$$

which is asymptotically distributed as χ^2 with (I-1)(J-1) degrees of freedom under the null hypothesis of no association.

The corresponding test based on the likelihood criterion is given by

$$(6.5) \qquad G^2 = -2 \sum_i \sum_j f_{ij} \log \frac{f_{ij}}{m_{ij}}$$

which is also distributed asymptotically as χ^2 with (I-1)(J-1) degrees of freedom under the null hypothesis of independence.

The chi-square approximation used in the above two test statistics depends on the distributional assumptions such as Poisson, multinomial or product multinomial and on the survey sampling scheme. As discussed in the introduction the tests are applicable to the simple random sample case. Any departure from that, for instance in the case of complex sample design involving stratification and clustering, can seriously effect the validity of these tests.

6.2.3 Adjustment for Complex Sample Design

Cohen (1976) proposed a method for adjusting the conventional test statistic X^2 used for independence or goodness of fit. The adjustment procedure was originally designed for a cluster size of two which was later extended by

Altham (1976) to cover the case where cluster size is more than two. The adjusted test statistic is given by

$$(6.6) \qquad X_a^2 = \frac{X^2}{1+a(k-1)} \sim \chi^2_{(I-1)(J-1)}$$

where k is the cluster size and a is some form of measure of positive correlation with $0 \leq a \leq 1$. If there is no intra-cluster correlation then a=0, and $X_a^2 = X^2$. If there is a perfect correlation then $X_a^2 = \frac{X^2}{k}$. Altham (1976) discuss the method for estimating a.

Brier (1980) used the Dirichlet-multinomial distribution to model the distribution of counts in contingency tables generated by cluster sampling. This provided an alternative rationale for Altham's method and also helped extend the method to cover the case of unequal clusters. It is shown by Brier that the asymptotic distribution of conventional test statistics X^2 and G^2 under clustering are $B\chi^2$ with (I-1)(J-1) degrees of freedom and

$$B = \left(\sum_{i=1}^{t} k_i r_i C_i \right) \Big/ \left(\sum_{i=1}^{t} k_i r_i \right)$$

where it is assumed that there are r_1 clusters of size k_1, ..., r_t clusters of size k_t; and

$$1 \leq C_i = (k_i + R)/(1 + R) \leq k_i \;;$$

and R is a structural parameter for clustering effects such that $\rho = 1/(1+R)$, where ρ is the intra-cluster correlation coefficient.

6.2.4 Log-Linear Models

Let us consider the notation for log-linear modelling which will then be extended to GLMM.

Let F_{ij} be the observed frequency in the i,j^{th} cell in a two-way contingency table which is the sum of the frequencies in i,j^{th} cells of k clusters. Then

$$F_{i\bullet} = \sum_j F_{ij} \text{ is the marginal total for the } i^{th} \text{ row;}$$

$$F_{\bullet j} = \sum_i F_{ij} \text{ is the marginal total for the } j^{th} \text{ column;}$$

$$F_{\bullet\bullet} = \sum_i \sum_j F_{ij} = \sum_i F_{i\bullet} = \sum_j F_{\bullet j} \text{ is the overall total.}$$

Similarly,

f_{ijk} is the observed frequency in the i,j^{th} cell of the kth cluster;

$f_{i\bullet k}$ is the marginal total for the ith row in the kth cluster;

$f_{\bullet jk}$ is the marginal total for the jth column in the kth cluster;

$$f_{\bullet\bullet k} = \sum_i \sum_j f_{ijk} = \sum_i f_{i\bullet k} = \sum_j f_{\bullet jk} \text{ is the overall total for the } k^{th}$$

cluster.

Therefore,

$$F_{ij} = \sum_k f_{ijk} \; ; \quad F_{i\bullet} = \sum_k f_{i\bullet k} \; ; \quad F_{\bullet j} = \sum_k f_{\bullet jk} \; ; \text{ and } F_{\bullet\bullet} = \sum_k f_{\bullet\bullet k}.$$

When clustering is ignored the contingency table is analysed using the following log-linear model:

$$(6.7) \qquad \log e_{ij} = \log M + \log a_i + \log b_j + \log (ab)_{ij}$$

$$= \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$$

where e_{ij} is the Poisson mean of F_{ij};

μ is the overall mean;

α_i is the row effect;

β_j is the column effect;

$\alpha\beta_{ij}$ is the interaction effect;

To test for independence or goodness-of-fit in the contingency table, the model is fitted with and without the interaction terms and the likelihood ratio test is applied. In the presence of a cluster effect the test statistic turns out to be larger and leads to the false conclusion of association in the table.

6.3 GENERALIZED LINEAR MIXED MODELS APPROACH

The GLMM approach as outlined in sections 2.3.2 and 3.3 is used to develop an analysis strategy for contingency tables with clustered observations. The analysis strategy is developed under the Poisson assumption and consequently a Poisson-normal GLMM is the basis of the proposed analysis.

6.3.1 Application to Contingency Tables

To account for the cluster level correlation the model can be fitted to cluster level observations as follows:

(6.8) $\qquad \log e_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + u_k + u_{ik} + u_{jk}$

where u_k, u_{ik} and u_{jk} are random coefficients corresponding to overall mean, row and column effects respectively for the kth cluster. The model obtains estimates of common interaction effects after removing the random variation in cluster level main effects.

The above model can be written in a general form as

$$e = f(\eta) \quad \text{where} \quad \eta = X\beta + Zu$$

where

and

$$e' = [e_{111}, e_{121}, ..., e_{IJ1}, ..., e_{11K}, e_{12K}, ..., e_{IJK}],$$

$$\eta' = [\eta_{111}, \eta_{121}, ..., \eta_{IJ1}, ..., \eta_{11K}, \eta_{12K}, ..., \eta_{IJK}] .$$

Let $I \times J = m$, $I \times J \times K = n$ and $1 + (I-1) + (J-1) = P$ then

X is an $n \times m$ and Z is an $n \times P \times K$ design matrices corresponding to the fixed effects and the random effects respectively, Z can be partitioned conformably to $u' = [u_1', u_2', ..., u_p']$ as $Z = [Z_1, Z_2, ..., Z_P]$.

$$\beta' = [\mu, \alpha_1, \alpha_2, .., \alpha_{I-1}, \beta_1, \beta_2, .., \beta_{J-1}, \alpha\beta_{11}, ..., \alpha\beta_{(I-1)(J-1)}]$$

is a P-vector of fixed effects, which will be estimated after removing the cluster level random variation in the main effects.

$$u' = [u_\mu', u_{\alpha_1}', ..., u_{\alpha(I-1)}', u_{\beta_1}', ..., u_{\beta(J-1)}']$$

is an $P \times K$ vector of random effects for P main effects corresponding to K clusters, i.e. each of $u_\mu', u_{\alpha_1}', ..., u_{\beta(J-1)}'$ has K components, for example,

$$u_\mu' = [u_{\mu_1}, u_{\mu_2}, ..., u_{\mu_K}],$$

u is assumed to be distributed as N(0,A) where

$$A = \text{diag} [A_\mu \phi_\mu, A_{\alpha_1} \phi_{\alpha_1}, ..., A_{\alpha(I-1)} \phi_{\alpha(I-1)}, A_{\beta_1} \phi_{\beta_1}, ..., A_{\beta(J-1)} \phi_{\beta(J-1)}]$$

with A_μ, A_{α_1}, etc. known matrices of constants.

The distribution of the components of u can also be assumed to be log-gamma with $V_\mu = \exp(u_\mu)$, $V_{\alpha_1} = \exp(u_{\alpha_1})$, etc. gamma distributed with mean 1. However in the case of a Poisson distributed response variable, a simulation study presented in Chapter 4 indicates that the difference between models

with normal and log-gamma distributed random components is not significant. Therefore, the present application will be derived assuming normally distributed random components.

The distribution of observed frequencies, f_{ijk}, conditional on given random components, $\mathbf{u}$, is assumed to be Poisson with mean $e_{ijk} = \exp(\eta_{ijk})$. Thus the log-likelihood functions for $f|\mathbf{u}$ and $\mathbf{u}$ are

$$(6.9) \qquad l_f|\mathbf{u} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} [-\exp(\eta_{ijk}) + f_{ijk}\eta_{ijk} - \ln f_{ijk}!] \; ,$$

$$(6.10) \qquad l_{u_\mu} = \text{constant} - (1/2) \; \{K \; \ln(2\pi\phi_\mu) + \phi_\mu^{-1}\mathbf{u}_\mu A_\mu^{-1}\mathbf{u}_\mu'\}$$

$$\text{and} \quad l_u = l_{u_\mu} + \ldots\ldots + l_{u_{\beta(J-1)}}$$

giving

$$\partial l_1/\partial \eta_{ijk} = Y_{ijk} - \exp(\eta_{ijk}) \; , \quad \partial^2 l_1/\partial \eta_{ijk}^2 = -\exp(\eta_{ijk}) \; ,$$

$$I_{y|u} = \text{diag}[\exp(\eta_{ijk})] \; , \quad I_{y,u} = \mathbf{B} = \text{diag}[\exp(x_{ijk}'\beta)],$$

here I is the information matrix and

$$\partial l_2/\partial \mathbf{u}_\mu = -\phi_\mu^{-1}A_\mu^{-1}\mathbf{u}_\mu \; , \quad \partial^2 l_2/\partial \mathbf{u}_\mu^2 = -\phi_\mu^{-1}A_\mu^{-1}$$

and so on for $\mathbf{u}_{\alpha 1}$, $\mathbf{u}_{\alpha 2}$, etc.

Therefore, $\partial l_2/\partial \mathbf{u} = -\mathbf{A}^{-1}\mathbf{u}$, $\partial^2 l_2/\partial \mathbf{u}\partial \mathbf{u}' = -\mathbf{A}^{-1}$.

The scoring equations for estimating β, $\mathbf{u}$ are iterative from initial values β_0, $\mathbf{u}_0$ to first iteration β_1, $\mathbf{u}_1$ given by

$$(6.11) \qquad \mathbf{V}\begin{bmatrix} \beta_1 - \beta_0 \\ \mathbf{u}_1 - \mathbf{u}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix}(f - e^{\eta_0}) + \begin{bmatrix} \mathbf{0} \\ -\mathbf{A}^{-1}\mathbf{u}_0 \end{bmatrix} \; ,$$

$$\mathbf{V} = \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix}\mathbf{B}\begin{bmatrix} \mathbf{X} & \mathbf{Z} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{A}^{-1} \end{bmatrix}$$

Once convergence for initial estimates of $\tilde{\beta}$, $\tilde{u}$ for given initial ϕ is achieved, the improved estimate of ϕ is obtained and the iteration process is repeated until the convergence in ϕ is also achieved.

The REML estimate of ϕ_μ, for instance, is given by

(6.12) $\qquad \tilde{\phi}_\mu = (\tilde{u}'_\mu \tilde{u}_\mu + \text{tr } T_\mu)/K$

where K is the number of clusters and

$$
V^{-1} = \begin{bmatrix} X'BX & X'BZ \\ Z'BX & Z'BZ+A^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} V_{11} & V_{12} \\ V'_{12} & V_{22} \end{bmatrix}^{-1} = \begin{bmatrix} T_1 & T_2 \\ T'_2 & T \end{bmatrix} ,
$$

T_μ is the corresponding sub-matrix of T.

Once the model is fitted, the estimate of the standard errors of the fixed and random components are obtained from the information matrix. The Wald test can be used to test the significance of the interaction terms for the hypothesis of independence or goodness-of-fit as follows.

(6.13) $\qquad X_w^2 = [\tilde{\alpha\beta}_{11},...,\tilde{\alpha\beta}_{(I-1)(J-1)}]' T_1^* [\tilde{\alpha\beta}_{11},...,\tilde{\alpha\beta}_{(I-1)(J-1)}]$

where T_1^* is a sub-matrix of T_1 consisting of the appropriate elements.

Under the null hypothesis X_w^2 is distributed as χ^2 with (I-1)(J-1) degrees of freedom.

Alternatively, the modelling can be repeated without the interaction terms and the likelihood ratio type test can be undertaken as suggested by (Lee and Nelder, 1996) as follows. If

$\qquad\qquad H_0 : \beta = \beta_0 \quad$ then

$$(6.14) \qquad G^2 = 2\{l(\tilde{\beta},\phi;\ \mathbf{f},\tilde{\mathbf{u}}) - l(\tilde{\beta}_0,\phi;\ \mathbf{f},\tilde{\mathbf{u}}_{\beta_0})\} + 2\{A(\tilde{\beta}) - A(\beta_0)\}$$

where $A(\tilde{\beta}) = -\frac{1}{2} \log \det (V_{22}|\tilde{\beta})$ and $A(\tilde{\beta}_0) = -\frac{1}{2} \log \det (V_{22}|\tilde{\beta}_0)$. G^2 is approximately distributed as χ^2, for given ϕ.

A 2×2 contingency table can be analysed either by binomial logit or log-linear models and the analysis method for overdispersed data due to clustering is well developed using binomial logit mixed models. The mixed modelling described above for the log-linear (Poisson-normal) model is consistent with binomial logit case and both methods give the same result.

6.3.2 Computation

A usual problem in the analysis of contingency tables using the log-linear model is the number of parameters to be estimated. This problem will potentially increase very rapidly if GLMM is used for the analysis. The number of random effects to be included is directly proportional to the number of clusters contributing to the table. For example, in a 5×5 table with 20 clusters, 25 fixed effects (9 main effects and 16 interaction effects) and 180 (20×9) random effects are required to be estimated. This increasing number of parameters can quickly paralyse the proposed method of analysis. It is shown below how this problem can be removed when applying the technique to large contingency tables with many clusters.

The main problem is to take the inverse of the variance-covariance matrix of very large dimension. However, with slight rearrangement of the random parameters the corresponding variance-covariance matrix, V_{22} , can be

converted to a block diagonal matrix which would make the method of computation considerably simpler.

The vector of random parameters $\mathbf{u}$ can be written as

$$\mathbf{u}' = [\mathbf{u}_1', \mathbf{u}_2', \ldots \mathbf{u}_k', \ldots, \mathbf{u}_K']$$

where $\mathbf{u}_k' = [u_{\mu_k}, u_{\alpha_{1k}}, \ldots, u_{\alpha_{(I-1)k}}, u_{\beta_{1k}}, \ldots, u_{\beta_{(J-1)k}}]$.

Then the following matrices can be expressed in diagonal forms:

$$\mathbf{Z} = \operatorname{diag}[\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_K], \qquad \mathbf{B} = \operatorname{diag}[\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_K],$$

$$\mathbf{V}_{22} = \operatorname{diag}[\mathbb{V}_{11}, \mathbb{V}_{22}, \ldots, \mathbb{V}_{KK}], \quad \mathbf{V}_{12} = \operatorname{diag}[\mathbb{V}_{01}, \mathbb{V}_{02}, \ldots, \mathbb{V}_{0K}].$$

If $\mathbf{X}' = [\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_K']$ then $\mathbb{V}_{00} = \mathbf{X}_k' \mathbf{B}_k \mathbf{X}_k$, $\mathbb{V}_{0k} = \mathbf{X}_k' \mathbf{B}_k \mathbf{Z}_k$

and $\mathbb{V}_{kk} = \mathbf{Z}_k' \mathbf{B}_k \mathbf{Z}_k + \theta^{-1} \mathbf{I}$ where $\theta = [\phi_\mu, \phi_{\alpha 1}, \ldots, \phi_{\alpha(I-1)}, \phi_{\beta 1}, \ldots, \phi_{\beta(J-1)}]$

$$\mathbf{V} = \begin{bmatrix} \mathbb{V}_{00} & \mathbb{V}_{01} & \mathbb{V}_{02} & \cdots & \mathbb{V}_{0K} \\ \mathbb{V}_{01}' & \mathbb{V}_{11} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \mathbb{V}_{01}' & 0 & 0 & \cdots & \mathbb{V}_{KK} \end{bmatrix} \qquad \mathbf{V}^{-1} = \begin{bmatrix} \mathbf{T}_{00} & \mathbf{T}_{01} & \cdots & \mathbf{T}_{0K} \\ \mathbf{T}_{01}' & \mathbf{T}_{11} & \cdots & \mathbf{T}_{1K} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathbf{T}_{0K}' & \mathbf{T}_{1K}' & \cdots & \mathbf{T}_{KK} \end{bmatrix}$$

where

$$\mathbf{T}_{00} = \mathbf{E}^{-1}, \quad \mathbf{T}_{0k} = \mathbf{E}^{-1} \mathbb{V}_{0k} \mathbb{V}_{kk}^{-1},$$

$$\mathbf{T}_{kk} = \mathbb{V}_{kk}^{-1} + \mathbb{V}_{kk}^{-1} \mathbb{V}_{0k}' \mathbf{E}^{-1} \mathbb{V}_{0k} \mathbb{V}_{kk}^{-1}$$

$$\mathbf{T}_{kk'} = \mathbb{V}_{kk}^{-1} \mathbb{V}_{0k}' \mathbf{E}^{-1} \mathbb{V}_{0k'} \mathbb{V}_{kk'}^{-1}$$

and $\mathbf{E} = \mathbb{V}_{00} - [\mathbb{V}_{01} \mathbb{V}_{11}^{-1} \mathbb{V}_{01}' + \mathbb{V}_{02} \mathbb{V}_{22}^{-1} \mathbb{V}_{02}' + \ldots + \mathbb{V}_{0K} \mathbb{V}_{kk}^{-1} \mathbb{V}_{0k}']$

$$= \mathbb{V}_{00} - \sum_{k=1}^{K} \mathbb{V}_{0k} \mathbb{V}_{1k}^{-1} \mathbb{V}_{0k}'$$

$$\tilde{\beta}_1 = \beta_0 + \mathbf{T}_{00} \mathbf{X}'(\mathbf{f} - e^{\eta_0}) + \sum_{k=1}^{K} \mathbf{T}_{0k} \mathbf{Z}_k'(\mathbf{f}_k - e^{\eta_{k0}}),$$

88

$$\tilde{u}_{k1} = u_{k0} + T_{10}\, X'(f - e^{\eta_0}) + \sum_{k=1}^{K} [T_{1k}\, Z'_k(f_k - e^{\eta_{k0}}) - \theta_k^{-1} u_{k0}].$$

The estimates of dispersion parameters are obtained as follows

$$\tilde{\phi}_\mu = (\sum_{k=1}^{K} \tilde{u}^2_{\mu k} + \operatorname{tr} T_\mu)/K,$$

where T_μ is a K×K matrix formed by extracting the first diagonal elements of the sub-matrices of T i.e. T_{11} ,......, T_{KK} .

$$\tilde{\phi}_{\alpha 1} = (\sum_{k=1}^{K} \tilde{u}^2_{\alpha_{1k}} + \operatorname{tr} T_{\alpha 1})/K,$$

where $T_{\alpha 1}$ is a K×K matrix formed by extracting the second diagonal elements of the sub-matrices of T i.e. T_{11} ,......, T_{KK} .

Similarly estimates of ϕ corresponding to other parameters are obtained.

The estimates obtained under the above strategy will not require the handling of any matrix greater than the matrix corresponding to the fixed effect model. Therefore computationally, if a contingency table can be analysed by using an ordinary log-linear model, it should also be able to be analysed under the proposed mixed model scheme for any number of clusters.

6.4 APPLICATION

The method is applied to a benchmark dataset used by Brier (1980) for studying the effect of clustering in contingency tables analysis. The data came from a survey conducted for the U.S. Department of Agriculture to study the levels of satisfaction with housing in the neighbourhood. The survey was conducted in 20 neighbourhoods in Montevideo, Minnesota and five families were selected in each neighbourhood and two questions were asked,

viz. the level of satisfaction with housing in their neighbourhood and the level of satisfaction with their own house. The responses were categorised as unsatisfied, satisfied and very satisfied for each question. The responses from all five selected families were reported for all neighbourhoods except two where only three families were included. Appendix 1 (Table A1) presents the detailed dataset separately for all neighbourhoods and Table 6.1 presents a collapsed version of the dataset for all families.

Table 6.1. Contingency Table of Families Classified by Level of Satisfaction with Housing at Personal Level and Community Level.

	P_1	P_2	P_3
C_1	18 (12.5)	6 (9.75)	0 (1.75)
C_2	28 (30.73)	28 (23.97)	3 (4.30)
C_3	4 (6.77)	5 (5.28)	4 (0.95)

P = Personal Satisfaction,   C = Community Satisfaction

and the subscripts 1 = Unsatisfied, 2 = Satisfied, 3 = Very Satisfied.

The objective of the analysis is to test the hypothesis of whether a family's level of personal satisfaction is independent of its level of community satisfaction.

A conventional chi-square test of independence in the collapsed table, ignoring the existence of clusters and assuming a simple random sample of 96 families, would result in a Pearson χ^2 statistic, $X^2 = 17.89$ and the likelihood ratio χ^2, $G^2 = 15.38$. Both of these values are greater than $\chi^2_{4,0.05} = 9.49$ which indicates there is a significant level of association between categories of personal and community level satisfactions.

Brier (1980) extended a method proposed by Altham (1976) for deflating the χ^2 value for cluster effect. Altham originally derived the deflating factor for equi-size clusters which was extended by Brier for unequal clusters. To apply Brier's method an estimate of intra-cluster correlation is required and, in the absence of such information, a maximum value of 1 for the intra-cluster correlation is suggested to be used. Under this assumption deflated values of X^2 and G^2 are obtained as 3.67 and 3.15 respectively, which are less than $\chi^2_{4,0.05}$ implying no association among the classification of personal and community level satisfactions. Brier (1980) refined this chi-squared statistics further by deriving an estimate of intra-cluster correlation rather than using the maximum possible correlation. That produced the values of X^2 and G^2 of 15.68 and 13.49 respectively, which are again significant compared with $\chi^2_{4,0.05}$. Fingleton (1984) however notes that the conditions under which the method is developed are fairly weak and not highly reliable.

We analyse the same dataset using the method proposed in this paper. The model used to analyse the data is as follows:

$$\mathbf{e} = f(\eta) \quad \text{where } \eta = \mathbf{X}\beta + \mathbf{Zu}$$

and

$\mathbf{e} = [e_{ijk}]$, the Poisson mean of the category counts f_{ijk},

$i=1,2,3$; $j=1,2,3$ and $k=1,2,...,20$;

i and j represent levels of community and personal level satisfactions respectively and k indicates the number of neighbourhoods. We have

$\beta' = [\mu, c_1, c_2, p_1, p_2, cp_{11}, cp_{12}, cp_{21}, cp_{22}]$,

$\mathbf{u}' = [\mathbf{u}'_1,....,\mathbf{u}'_{20}]$,

and $\mathbf{u}'_k = [u_{\mu k}, u_{c1k}, u_{c2k}, u_{p1k}, u_{p2k}]$.

The model is fitted with and without interaction terms and the likelihood ratio statistic is calculated as $G^2 = 11.43$ which is closer to the value obtained by Brier (1980). The calculated $G^2 = 11.43$ lies in between $G^2 = 3.15$ and $G^2 = 13.49$, the likelihood ratio statistics calculated based on Brier's method under perfect intra-cluster correlation and estimated intra-cluster correlation respectively. As the assumption of perfect correlation is a conservative strategy the real G^2 is likely to be higher. On the other hand as Fingleton (1984) mentioned, Brier's method of calculating G^2 based on estimated correlation is dependent on fairly weak conditions and can not be highly relied upon. Thus the actual value of G^2 is more likely to be in between these two values.

6.5 DISCUSSION

A GLMM based framework is proposed for the analysis of contingency tables where observations enter as clusters. The analysis is applicable when Poisson sampling error is assumed. The method has the advantage of accommodating multiple levels of clustering. This is basically an extension of log-linear modelling to incorporate random cluster effects. A disadvantage of the method is that similar to the standard log-linear analysis the numbers of fixed parameters and random components to be dealt with are large since one fixed parameter and one random component are required for each of the main effects. A computing algorithm is developed to ensure that the proposed method does not demand any unmanageable additional computing space. That means from the computing point of view the analysis can be undertaken if the standard log-linear modelling can be carried out. Still the method should be preferred only if there is a

specific reason for conducting the analysis under the Poisson assumption. Otherwise the method proposed in the following chapter assuming product multinomial rather than Poisson is easy to apply because of the reduction in the number of parameters.

CHAPTER SEVEN

MODELING MULTINOMIAL DATA WITH EXTRA VARIATION AND ANALYSIS OF CONTINGENCY TABLES

7.1 INTRODUCTION

Multi-category response variables are common phenomena in social and biomedical research. Multinomial modelling is often adopted to analyse such data. However, in many cases the variances and covariances of the observations are found to be more than the sampling variation assumed by the multinomial model. This syndrome is known as extra variation or overdispersion in the literature. The reason behind such overdispersion is mainly due to clustering of observations which can happen for various reasons as discussed in Chapter 6 in relation to the clustering in contingency tables.

Extra variation in multinomial modelling is often addressed by using quasi-likelihood functions or by using the Dirichlet multinomial model. The multinomial covariance matrix is often multiplied by a scalar parameter to account for extra variation. The Dirichlet-multinomial model is in fact a scalar adjustment of the multinomial covariance matrix used by Brier (1980) for the analysis of contingency tables under cluster sampling. Koehler and Wilson (1986) adopted a similar approach for comparing proportions derived from cluster samples. McCullagh and Nelder (1989) proposed a quasi-likelihood approach to scaled multinomial modelling. Liang and Zeger (1986) proposed a class of generalised estimating equations to address this type of problem. Recently Morel and Koehler (1995) proposed a method based on a one step Gauss Newton estimator which allows flexibility in accommodating

different levels of overdispersion corresponding to different components of variance instead of multiplying the covariance matrix by a single scalar. This method however concentrates on making inference about the population mean rather than about individual respondents.

We propose a method based on the GLMM approach which retains the flexibility of accommodating overdispersions at different levels similar to Morel and Koehler's method but it offers the added advantage of making inference about individual respondents apart from inference about population means. It is also discussed how this method can be used for analysing contingency tables with correlated observations. This offers an alternative to the method discussed in the previous chapter.

7.2 ANALYSIS IGNORING CLUSTERING

Let $\mathbf{y}^{*\prime}=[\mathbf{y}_1^{*\prime},\mathbf{y}_2^{*\prime},....,\mathbf{y}_I^{*\prime}]$ be a vector of counts which consists of a set of sub-vectors which are independent, where $\mathbf{y}_i^{*\prime}=[y_{i1},y_{i2},....,y_{iJ+1}]$, are the category counts in J+1 response categories.

Let $\mathbf{T}^{\prime}=[T_1,T_2,....,T_I]$ be a vector of the totals of I groups

$$\text{where} \quad T_i = \sum_{j=1}^{J+1} y_{ij} \ ,$$

and $\pi^{*\prime}=[\pi_1^{*\prime},\pi_2^{*\prime},....,\pi_I^{*\prime}]$ be a vector of probabilities, where $\pi_i^{*\prime}=[\pi_{i1},\pi_{i2},....,\pi_{iJ+1}]$ be the vector of probabilities for responses in the J+1 categories for each of the T_i observations in group i.

$$\sum_{j=1}^{J+1} \pi_{ij} = 1, \quad \text{and} \quad \pi_{iJ+1} = 1 - \sum_{j=1}^{J} \pi_{ij} \ , \quad E(\mathbf{y}_i^*) = T_i\pi_i^* \ .$$

The likelihood function for product multinomial data can be written as

$$(7.1) \qquad \mathcal{L}(\mathbf{y}) = \prod_{i=1}^{I} T_i! \prod_{j=1}^{J+1} \frac{\pi_{ij}^{y_{ij}}}{y_{ij}!}$$

and the corresponding log-likelihood function excluding the constants can be written as

$$(7.2) \qquad l(\mathbf{y}) = \sum_{i=1}^{I} \sum_{j=1}^{J+1} y_{ij} \ln \pi_{ij} \ .$$

Let us define the link function $\pi_{ij} = g(\eta_{ij})$ with $\eta_{ij} = \mathbf{x}_{ij}\beta$ where $\mathbf{x}_{ij}$ is a vector of covariates and β is a vector of parameters.

Now in dealing with multinomial data since the T_i's are predetermined so is the J+1th category once the first J categories are known.

Therefore dropping the last value of each group we can define:

$$\mathbf{y'} = [\mathbf{y}_1', \mathbf{y}_2', ..., \mathbf{y}_I'], \text{ where } \mathbf{y}_i' = [y_{i1}, y_{i2}, ..., y_{iJ}], \text{ and}$$

$$\pi' = [\pi_1', \pi_2', ..., \pi_I'] \text{ where } \pi_i' = [\pi_{i1}, \pi_{i2}, ..., \pi_{iJ}] \ .$$

Here $\mathbf{y}$ is an m-vector with m=I×J where the first J elements belong to the first group, the second J elements belong to the second group and so on. Similarly π is the corresponding m-vector of probabilities. Then the link function can be written as $\eta = \mathbf{X}\beta$ where $\mathbf{X}$ is an m×p matrix of explanatory variables and β is a p×1 vector of parameters. Then the log likelihood function can be written as

$$(7.3) \qquad l = \mathbf{y'} \ln g(\eta) = \mathbf{y'} \ln g(\mathbf{X}\beta),$$

The estimate of β is obtained by maximizing the likelihood function. This works well when the observations do not show any intra-cluster correlation or any other extra variation. In the presence of such extra variation the

estimate of β is consistent but the estimate of the standard error of β tends to be under estimated. In the following section, we demonstrate how the GLMM technique can be used to estimate of β and the cluster effects.

7.3 GENERALISED LINEAR MIXED MODELS APPROACH

Let us extend the above structure to a situation where the observations come from K clusters and the category counts are recorded at the cluster level for each population. We define, $\mathbf{y}^*$ as a vector of category counts at cluster level such that y_{kij} represents the category counts for the jth category of the ith group in the kth cluster. π^* and $\mathbf{T}$ are the corresponding vectors of probabilities and cluster level population totals respectively. Therefore,

$$\mathbf{y}^{*\prime} = [\mathbf{y}_{11}^{*\prime},...,\mathbf{y}_{1I}^{*\prime},....,\mathbf{y}_{K1}^{*\prime},...,\mathbf{y}_{KI}^{*\prime}] \text{ and } \mathbf{y}_{ki}^{*\prime} = [y_{ki1},...,y_{kiJ+1}]$$

$$\pi^{*\prime} = [\pi_{11}^{*\prime},...,\pi_{1I}^{*\prime},....,\pi_{K1}^{*\prime},...,\pi_{KI}^{*\prime}] \text{ and } \pi_{ki}^{*\prime} = [\pi_{ki1},...,\pi_{kiJ+1}]$$

$$\text{with } \sum_{j=1}^{J+1} \pi_{kij} = 1$$

$$\mathbf{T}^{\prime} = [\mathbf{T}_1^{\prime},\mathbf{T}_2^{\prime},...,\mathbf{T}_K^{\prime}] \quad \text{and} \quad \mathbf{T}_k^{\prime} = [T_{k1},T_{k2},...,T_{ki}] \text{ with } T_{ki} = \sum_{j=1}^{J+1} y_{kij} .$$

As before after removing the last categories we define $\mathbf{y}$ and π corresponding to $\mathbf{y}^*$ and π^*. The dimensions of $\mathbf{y}$ and π is $n \times 1$ where $n = K \times I \times J$.

The link function is now defined as

$$(7.4) \qquad \pi = g(\eta) \text{ and } \eta = \mathbf{X}\beta + \mathbf{Z}\mathbf{u}$$

where

$\mathbf{X}$ is an $n \times p$ matrix of covariates corresponding to the p-vector of fixed coefficients β,

$$\mathbf{u}' = [\mathbf{u}'_1, \mathbf{u}'_2, \ldots, \mathbf{u}'_J] \quad \text{with} \quad \mathbf{u}'_j = [u_{j1}, u_{j2}, \ldots, u_{jK}]$$

a vector of random effects which has J components each with K effects for K clusters and $\mathbf{Z}$ is an $n \times K \times J$ incident matrix indicating the presence or absence of the observations in K clusters for J response categories.

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{A}) \quad \text{where} \quad \mathbf{A} = \text{block diag } [\mathbf{A}_1\phi_1, \mathbf{A}_2\phi_2, \ldots, \mathbf{A}_J\phi_J]$$

where ϕ_j's are dispersion parameters of the random components and $\mathbf{A}_j$ are known matrices.

As usual the log-likelihood function is now defined in two components

$$(7.5) \qquad l = l_1 + l_2 \qquad \text{where}$$

$$l_1 = \ln f(\mathbf{y}; \beta \mid \mathbf{u}) = \ln f(\eta, \phi \mid \mathbf{u}) = \mathbf{y}' \ln g(\eta) \text{ and}$$

$$l_2 = \ln f(\mathbf{u}) = \text{constant} - (1/2) \sum_{j=1}^{J} \{K\ln(2\pi\phi_j) + \phi_j^{-1}\mathbf{u}'_j\mathbf{A}_j^{-1}\mathbf{u}_j\}$$

The estimates of β, $\mathbf{u}$ and ϕ are then obtained by the usual method of maximizing the joint likelihood function l as described in the previous chapters.

7.3.1 Multinomial Logistic Modelling

As a logit link function is commonly used in the case of multinomial data we describe the details of estimation for multinomial logistic modelling in this section.

For multinomial data odds ratios are defined as the ratio of the odds of a particular response category with the odds of the base category. Here the

last category is treated as the base category. The logit link function is given by

$$(7.6) \qquad \pi_{kij} = \frac{\exp(\eta_{kij})}{1 + \sum\limits_{j=1}^{J} \exp(\eta_{kij})}$$

Thus the form of the likelihood function l_1 ignoring irrelevant portions is

$$(7.7) \qquad l_1 = \sum\limits_{k=1}^{K} \sum\limits_{i=1}^{I} \sum\limits_{j=1}^{J+1} y_{kij} \left[\eta_{kij} - \ln \left\{ 1 + \sum\limits_{j=1}^{J} \exp(\eta_{kij}) \right\} \right]$$

and l_2 is as specified earlier.

$$\partial l_1 / \partial \eta_{kij} = y_{kij} - T_{ki} \frac{\exp(\eta_{kij})}{1 + \sum\limits_{j=1}^{J} \exp(\eta_{kij})} = y_{kij} - T_{ki} \pi_{kij}$$

$$\Rightarrow \partial l_1 / \partial \eta = y - \tilde{y} \quad \text{where } \tilde{y} = [\tilde{y}_{ikj}] = T_{ki} \pi_{kij}$$

$$\partial^2 l_1 / \partial \eta_{kij} \, \partial \eta_{k'i'j'} = \pi_{kij}(1 - \pi_{kij}) \quad \text{if } k=k', \ i=i' \text{ and } j=j'$$

$$= - \pi_{kij} \pi_{kij'} \quad \text{if } k=k', \ i=i' \text{ and } j \neq j'$$

$$= 0 \quad \text{otherwise}$$

$$\Rightarrow - \partial^2 l_1 / \partial \eta \partial \eta' = \text{diag}(\pi) - \text{block diag}(\pi\pi') = V(\pi)$$

$$\partial^2 l_2 / \partial u \partial u' = - \left(\text{diag } \phi_j^{-1} A_j^{-1} \right) = - A^{-1}$$

The estimates of β, u and ϕ are obtained by maximizing the joint likelihood function $l = l_1 + l_2$. From initial values β_0, u_0 and given ϕ_0 the estimates $\tilde{\beta}$, $\tilde{u}$ are obtained by using the Newton-Raphson iteration procedure as follows:

$$(7.8) \qquad V \begin{bmatrix} \tilde{\beta} - \beta_0 \\ \tilde{u} - u_0 \end{bmatrix} = \begin{bmatrix} X' \\ Z' \end{bmatrix} (y - \tilde{y}) - \begin{bmatrix} 0 \\ A^{-1} u_0 \end{bmatrix} ,$$

with $V = \begin{bmatrix} X' \\ Z' \end{bmatrix} V(\pi) \begin{bmatrix} X & Z \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & A^{-1} \end{bmatrix}$. If $V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$, $V^{-1} = \begin{bmatrix} \cdot & \cdot \\ \cdot & \mathbb{T} \end{bmatrix}$,

then a new initial value for ϕ_j is $\tilde{u}'_j A_j^{-1} \tilde{u}_j/(K-r_j^*)$, $r_j^* = \phi_j^{-1} \operatorname{tr} V_{22(j)}^{-1}$. Using this estimate of ϕ_j , new estimates for β, u are found and a new value of ϕ_j is estimated as indicated above. The final converged values of $\tilde{\beta}$, $\tilde{u}$, $\tilde{\phi}$ are the approximate maximum likelihood estimates. If V_{22}^{-1} is replaced by $\mathbb{T}$ then REML estimates are obtained.

7.3.2 Multiple levels of Clustering

When there are multiple levels of clustering the above method can be generalised by defining

$u' = [u'_1, u'_2, ..., u'_Q]$, with $u_q = [u_{q1}, u_{q2}, ..., u_{qJ}]$ and $u_{qj} = [u_{qj1}, ..., u_{qjK_q}]$

as a vector of random coefficients with Q levels of clustering. u_{qj} represents a vector of random effects corresponding to the jth response categories for the clustering at q^{th} level. The dimension of u_{qj} will be equal to K_q, the number of clusters at q^{th} level. If the first random component corresponds to K_1 ultimate clusters then u'_{1j} will be of dimension $K_1 \times 1$. If the second component corresponds to an upper level clustering, such as primary sampling units (PSU) in multistage sampling, each of which consists of K_2 ultimate clusters then u'_{2j} will have dimension $K_2 \times 1$ and so on. $Z = [Z_1, Z_2, ..., Z_Q]$ with $Z_q = [Z_{q1}, Z_{q2}, ..., Z_{qJ}]$ is partitioned conformably with the partition of u.

The u_{qj}'s are assumed to be independent with distribution $N(0, A_{qj}\phi_{qj})$ where ϕ_{qj}'s are dispersion parameters of random components and A_{qj}'s are known matrices such that

$\mathbf{A}$ = block diag $[\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_Q]$ and $\mathbf{A}_q$ = block diag $[\mathbf{A}_{q1}\phi_{q1}, ..., \mathbf{A}_{qJ}\phi_{qJ}]$.

In the log likelihood function (7.5) l_1 will remain as before but l_2 will be adjusted as follows

$$(7.9) \qquad l_2 = \ln f(\mathbf{u}) = \text{constant} - (1/2) \sum_{q=1}^{Q} \sum_{j=1}^{J} \{K_{qj}\ln(2\pi\phi_{qj}) + \phi_{qj}^{-1}\mathbf{u}_{qj}'\mathbf{A}_{qj}^{-1}\mathbf{u}_{qj}\}$$

The estimation equation will be the same as defined in (7.8) but with multiple random components and multiple dispersion parameters. The dispersion parameter for qj^{th} component will be calculated as $\phi_{qj} = \tilde{\mathbf{u}}_{qj}'\mathbf{A}_{qj}^{-1}\tilde{\mathbf{u}}_{qj}/(K_{qj}-r_{qj}^*)$, $r_{qj}^* = \phi_{qj}^{-1}$ tr $\mathbf{V}_{22(qj)}^{-1}$ with K_{qj} the number of components of $\mathbf{u}_{qj}$ and $\mathbf{V}_{22(qj)}^{-1}$ is the appropriate sub-matrix of $\mathbf{V}_{22}^{-1}$. If $\mathbf{V}_{22}^{-1}$ is replaced by $\mathbb{T}$ then REML estimates are obtained.

7.4 APPLICATION TO CONTINGENCY TABLES

Contingency tables are often analysed under a multinomial sampling scheme particularly when one or more variables can be thought of as independent variables and the remainder as response variables. The sample size corresponding to each category of independent variable is assumed fixed and the main interest is to test the homogeneity of proportions across independent groups. Multinomial logistic modelling is often employed to obtain the estimates of odds ratios or to test the homogeneity when the response variable has more than two levels. In this section, we describe the analysis of contingency tables using multinomial logistic modelling when the observations are clustered.

Let us consider the case of a two-way contingency table with I rows and J+1 columns. The observations contributing to the table are obtained from K clusters. Category counts are denoted by f_{kij} with expected value e_{kij}. If the columns are considered as response categories and rows are treated as explanatory categories then, taking the last response category as the base category, the multinomial logit for the k,i,j^{th} cell can be written as

$$\log \frac{e_{kij}}{e_{ki,J+1}} = \eta_{kij} \ .$$

If $\mathbf{e}$ and η are vectors of e_{kij}'s and η_{kij}'s then the full mixed effect analysis model can be expressed as

(7.10)   logit $\mathbf{e} = \eta = \mathbf{X}\beta + \mathbf{Z}\mathbf{u}$

where β is a vector of fixed coefficients with dimension J+(I-1)J. It can be decomposed as $\beta' = [\beta^{0\prime}, \beta^{*\prime}]$ where

$$\beta^0 = [\beta^0_1, \ \beta^0_2, \ ..., \ \beta^0_J]'$$

includes an intercept term for each response category except the base category J+1, and

$$\beta^* = [\beta^*_{1,1}, \ \beta^*_{1,2}, \ .., \ \beta^*_{1,J}, \ ....,\ \beta^*_{I-1,1}, \ \beta^*_{I-1,2}, \ ....., \ \beta^*_{I-1,J}]'$$

represents the interaction terms for each combination of the first J columns and I-1 rows. The interaction effect corresponding to the Ith row is confounded with the β^0 parameters.

Let $\mathbf{u}' = (\mathbf{u}'_1, \mathbf{u}'_2, ..., \mathbf{u}'_J)$ and $\mathbf{u}'_j = [u_{j1}, u_{j2}, ..., u_{jK}]$, to allow J random components corresponding to the J response categories with each having K effects corresponding to the K clusters. To allow for clustering at another level a different set of J components are required with number of effects in each

102

components equal to the number of clusters at that level. **X** and **Z** will have to be designed appropriately to conform with the parameter structure of β and **u**. Then the estimate of β, **u** and ϕ can be obtained by following the standard procedure described in the previous section.

To test homogeneity in the contingency table the likelihood ratio test or Wald test can be employed. For the likelihood ratio test the model has to be fitted again without the interaction terms but keeping the random component structure the same as before and test statistics can be constructed as suggested by (Lee and Nelder,1996) as follows.

If $H_0 : \beta^* = 0$ then

$$(7.11) \qquad G^2 = 2\{l(\tilde{\beta}^0, \tilde{\beta}^*, \phi; \mathbf{f}, \tilde{\mathbf{u}}) - l(\tilde{\beta}^0, \phi; \mathbf{f}, \tilde{\mathbf{u}}_{\beta^0})\} + 2\{A(\tilde{\beta}^0, \tilde{\beta}^*) - A(\tilde{\beta}^0)\}$$

where $A(\tilde{\beta}^0, \tilde{\beta}^*) = -\frac{1}{2} \log \det (\mathbf{V}_{22} | \tilde{\beta}^0, \tilde{\beta}^*)$ and $A(\tilde{\beta}^0) = -\frac{1}{2} \log \det (\mathbf{V}_{22} | \tilde{\beta}^0)$.

For given ϕ, G^2 is approximately distributed as χ^2 with degrees of freedom equal to the dimension of β^*.

As the test of homogeneity is equivalent to the test of independence, the multinomial logistic method can be used as an alternative to the method presented in the previous chapter under Poisson (log-linear) modelling for the test of independence. The main advantage of the multinomial logistic approach compared with the Poisson modelling approach is that fewer random terms are required, which is particularly useful when the number of clusters is very large. The number of fixed parameters under the multinomial modelling is also less. However, as discussed in detail by Freeman (1987),

there are situations where log-linear modelling is preferable to multinomial logistic and vice versa.

7.5 EMPIRICAL RESULTS

Two applications of the proposed method are discussed in this section, the first to contingency table analysis and the second to the analysis of general multi-category response data.

7.5.1 Application to Neighbourhood Dataset

The method is applied to the contingency table of the neighbourhood dataset (Brier 1980) as discussed in Chapter 6. Table 6.1 in the previous chapter presents a collapsed version of the dataset and Table A1 in Appendix 1 presents the detailed dataset separately for all neighbourhoods and for all families. This is basically a two-way contingency table, with levels of satisfaction with own home (p's) being columns and the satisfaction with neighbouring housing (c's) being rows. For the current analysis, the satisfaction with own home is treated as the response variable and the satisfaction with neighbouring housing is considered as the independent variable. As the response variable has three levels (p_1, p_2 and p_3) the first level is treated as the reference category while the three row levels (c_1, c_2 and c_3) are taken as the levels of the explanatory variable. Each neighbourhood is considered as a cluster of families and the analysis is undertaken by using cluster level contingency tables with cell frequencies (f_{kij}) where k indicates cluster and i and j are the levels of community and personal satisfaction. The multinomial logistic model used to analyse the data is

$$(7.12) \qquad \mathbf{e} = f(\eta) \qquad \text{where } \eta = \mathbf{X}\beta + \mathbf{Z}\mathbf{u}$$

and

$\mathbf{e} = [e_{kij}]$, is the expected value of f_{kij}

$i=1,2,3; \; j=2,3 \text{ and } k=1,2,\ldots,20;$

$e_{kij} = T_{ki}\pi_{kij} \quad$ with

$$T_{ki} = \sum_{j=1}^{3} f_{kij} = \text{multinomial marginals, and}$$

$$\pi_{kij} = \frac{\exp(\eta_{kij})}{1+\sum\limits_{j=2}^{3} \exp(\eta_{kij})}$$

$\beta' = [p_2, \; p_3, \; cp_{22}, \; cp_{32}, \; cp_{23}, \; cp_{33}],$

$\mathbf{u}' = [\mathbf{u}'_{p2}, \; \mathbf{u}'_{p3}], \quad \text{with} \quad \mathbf{u}'_{pj} = [u_{pj1}, u_{pj2}, \ldots, u_{pj20}].$

Here p_2 and p_3 are two fixed intercept coefficients corresponding to the two levels of the response variable and the remaining four fixed coefficients are interaction or odds ratio parameters. Two random components $\mathbf{u}_{p2}$, $\mathbf{u}_{p3}$ corresponding to two response categories each with twenty levels for twenty communities are included in the model.

$\mathbf{X}$ and $\mathbf{Z}$ are design matrices corresponding to the fixed coefficients β and random components $\mathbf{u}$. $\mathbf{Z}=[\mathbf{Z}_2,\mathbf{Z}_3]$ where $\mathbf{Z}_2$ and $\mathbf{Z}_3$ are design matrices corresponding to the two random components $\mathbf{u}_{p2}$, $\mathbf{u}_{p3}$.

The model was fitted with and without interaction terms and the calculated likelihood ratio statistic λ^2 and Wald statistic W^2 are presented in Table 7.1 in comparison to the test results obtained from other methods.

Table 7.1. Calculated Test Statistics Under Different Methods

Method of Estimation	λ^2	W^2
Multinomial Logistic (ignoring clustering)	15.39	
Log-linear (ignoring clustering)	15.09	
Multinomial Logistic (random cluster effects)	10.32	8.56
Log-linear (random cluster effects)	11.43	11.97
Brier's Dirichlet-multinomial (estimated corr)	13.49	

The results show that, similar to the mixed log-linear model introduced in the previous chapter, the mixed multinomial logistic model is deflates the likelihood ratio statistic, λ^2. The conventional analysis under multinomial and Poisson assumptions produces the values of λ^2 as 15.39 and 15.09 respectively. This are deflated to 10.32 under multinomial and 11.43 under Poisson mixed models. The correction in λ^2 for clustering under these methods is larger than the correction done under Brier's Dirichlet-multinomial model. The value of λ^2 obtained under Brier's method is 13.49. Even after the correction the test result is still significant at 5% level under all methods. However, the p-values are very close to 0.05 under the corrected methods.

7.5.2 Application to Birth Defect Study

Morel and Koehler (1995) present a dataset from a study conducted by Hartsfield *et al.* (1992) on the effects of prenatal exposure to cadmium and zinc on death and malformation rates of hamster fetuses. They used the dataset to compare the estimates obtained from their proposed one-step Gauss-Newton estimator with that of multinomial and scaled multinomial

models. We use the same dataset to show the performance of the proposed method in the case of general multi-category response data.

Table 7.2 presents the dataset where each row within a treatment group represents a litter. The treatment groups are the litters of pregnant hamsters which received various treatments such as 2 mg kg^{-1} of zinc, 2 mg kg^{-1} of cadmium, 3 mg kg^{-1} of cadmium, combination of 2 mg kg^{-1} of zinc and 2 mg kg^{-1} of cadmium and 2 mg kg^{-1} of zinc and 3 mg kg^{-1} of cadmium. In addition two control groups of 10 hamsters in each group received either no treatment or a saline solution only. The two control groups were combined into a single group in the analysis. The outcome of the study, the numbers of dead fetuses, alive fetuses with physical malformations and alive fetuses without physical malformations, are presented in the first, the second and the third columns respectively within each group.

Table 7.2: Data from Hamsters Birth Defect Study

Injected Controls			Untreated Controls			2mg kg^{-1} of zinc			2mg kg^{-1} of cadmium			3mg kg^{-1} of cadmium			2mg kg^{-1} zinc and 2mg kg^{-1} of cad.			2mg kg^{-1} zinc and 3mg kg^{-1} of cad.		
a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
0	0	13	0	0	13	2	0	11	2	5	5	6	5	1	1	1	9	4	1	5
8	0	1	5	1	7	0	0	15	12	0	0	11	0	0	0	0	13	4	7	3
0	0	13	0	0	13	1	0	13	4	5	3	8	5	0	1	1	13	13	5	0
1	0	11	1	0	11	1	1	13	3	8	0	8	0	0	4	4	3	2	4	6
1	0	17	0	0	13	1	0	12	3	8	0	14	0	0	0	1	12	6	5	1
0	0	13	0	0	11	1	0	12	9	4	0	11	1	0	0	0	11	11	0	0
0	0	15	0	0	11	0	0	13	1	12	3	12	0	0	0	3	9	5	6	2
0	0	14	1	2	10	1	0	10	1	9	2	13	5	0	2	2	9	4	2	6
0	0	11	0	1	12	0	0	16	2	4	10	12	0	0	0	4	11	6	3	6
0	0	14	0	0	12	0	1	11	3	0	12	12	0	0	1	5	6	5	3	3

a = dead fetuses, b = alive fetuses with malformations and c = alive fetuses

without malformations

In Morel and Koehler (1995) the dataset was analysed by viewing it as a 3×2 factorial experiment where one factor corresponded to the levels of zinc (control, 2 mg) and the other factor corresponded to the levels of cadmium (control, 2 mg, 3 mg). We analyse the dataset in the same way to make the results comparable.

Let us denote the outcomes in Table 7.2 by y_{kilj}, where k, i, l, j represent litters, levels of zinc, levels of cadmium and the response categories respectively. If $T_k = \sum_j y_{kilj}$ represents the multinomial marginals in litter k then the corresponding probabilities can be defined as $\pi_{kilj} = \dfrac{e_{kilj}}{T_k}$ where e_{kilj} are the expected frequencies. If the number of alive fetuses without any physical malformations is considered as the base category (j=c) then the logits for the alive with physical malformations (j=b) and dead fetuses (j=a) can be defined as $\log \dfrac{\pi_{kilj}}{\pi_{kilc}}$ for j = a and b.

The multinomial logit model is then fitted as

$$(7.13) \qquad e_{kilj} = g(\eta_{kilj}) \qquad \text{with}$$

$$\eta_{kilj} = \log \frac{\pi_{kilj}}{\pi_{kilc}} = \alpha_j + \beta_{ij}x_{ij} + \lambda_{lj}w_{kj} + \gamma_{ilj}xw_{ilj} + u_{kj}$$

where α_j's are intercept parameters that represent the control groups, β_{ij}'s are main effects for zinc, λ_{lj} are main effects for cadmium and γ_{ilj} are interaction parameters for the interactions among zinc and cadmium levels and u_{kj}'s are random effects for the jth response level in the kth litter. x_{ij}, w_{kj} and xw_{ilj} are indicator variables for zinc, cadmium and zinc×cadmium interactions for the jth response category. There are two sets

of parameters corresponding to two outcome levels of interest (j=a,b). In each set there are six fixed parameters - an intercept parameter, a parameter for zinc effect, two parameters for cadmium and two interaction effects. The base levels of zinc and cadmium are confounded with the intercept. There are two random components for two response categories with the number of effects in each component equal to the number of litters. Therefore in each set there are 70 random effects corresponding to 70 litters. The strategy accounts for random litter effects corresponding to each response categories.

Table 7.3: Estimates of Interaction Parameters and The Corresponding Statistics for the Test of Significance.

Parameters	Multi. model	Scaled multi.	Generalised mult. Diagonal	Generalised mult. Unrestrict	GLMM ML	GLMM REML
Logit 1: Dead versus alive without any malformations						
Zinc × Cad2	-2.24 (0.62)	-2.24 (1.09)	-2.24 (1.26)	-2.24 (1.28)	-2.94 (1.05)	-2.99 (1.11)
Zinc × Cad3	-3.78 (1.13)	-3.78 (1.98)	-3.78 (1.79)	-3.78 (1.88)	-4.44 (1.14)	-4.49 (1.46)
Logit 2 : Alive with versus without any malformations						
Zinc × Cad2	-1.90 (0.93)	-1.90 (1.63)	-1.90 (1.31)	-1.90 (1.31)	-2.10 (1.11)	-2.15 (1.51)
Zinc × Cad3	-2.58 (1.37)	-2.58 (2.41)	-2.58 (1.93)	-2.58 (1.93)	-2.28 (1.51)	-2.26 (1.54)
Wald test Logit 1	18.50	6.02	5.52	5.19	13.80	12.51
Logit 2	4.86	1.58	2.47	2.47	3.85	3.77
Logit 1 & 2	27.88	9.08	8.58	7.47	17.59	16.59

The results of fitting the model are presented in Tables 7.3 and 7.4 in comparison to the results of other methods such as scaled multinomial and

generalized multinomial models proposed by Morel and Koehler (1995). Table 7.3 presents the estimates of interaction effects and the corresponding Wald test results. It appears that the methods which only adjust the covariance

Table 7.4: Estimates of Parameters and Standard Errors Under Additive Models

Parameters	Multi. model	Scaled multi.	Generalised mult. Diagonal	Unrestrict	GLMM ML	REML
Logit 1 :Dead versus alive without any malformations						
Intercept	-2.30 (0.21)	-2.30 (0.39)	-2.30 (0.45)	-2.30 (0.45)	-2.47 (0.39)	-2.49 (0.41)
Zinc	-2.43 (0.28)	-2.43 (0.51)	-2.50 (0.54)	-2.20 (0.56)	-2.10 (0.48)	-2.10 (0.50)
Cad 2mg	2.33 (0.29)	2.33 (0.53)	2.29 (0.60)	2.43 (0.61)	2.35 (0.54)	-2.37 (0.57)
Cad 3mg	5.55 (0.36)	5.55 (0.65)	5.67 (0.69)	5.34 (0.71)	5.87 (0.59)	5.91 (0.61)
Logit 2 : Alive with versus without any malformations						
Intercept	-3.76 (0.41)	-3.76 (0.74)	-3.75 (0.59)	-3.79 (0.59)	-3.80 (0.47)	-3.81 (0.48)
Zinc	-1.51 (0.26)	-1.51 (0.47)	-1.51 (0.38)	-1.60 (0.38)	-1.48 (0.43)	-1.47 (0.45)
Cad 2mg	4.00 (0.45)	4.00 (0.81)	3.97 (0.65)	4.10 (0.65)	-3.99 (0.56)	-4.00 (0.58)
Cad 3mg	5.38 (0.50)	5.38 (0.91)	5.46 (0.73)	5.46 (0.73)	5.34 (0.63)	5.35 (0.65)

matrix i.e. scaled and generalised multinomial methods do not adjust the estimates of parameters but only adjust the standard errors. However, ML and REML under GLMM adjust both estimates and standard errors of the fixed coefficients. Consequently, the Wald test results turn out to be substantially different than other methods. The combined Wald statistics for logit 1 and 2 under GLMM reduce to 17.59 (p=0.0015) for ML and 16.59

(p=0.0023) for REML compared with 27.88 (p<0.001) under the ordinary multinomial model and in the range of 7.47 (p=0.113) to 9.08 (p=0.059) under other scaled multinomial models. Similarly, for logit 1 the Wald test under ML is 13.80 (p=0.008) and under REML is 12.51 (p=0.014) which are substantially higher than those of other scaled methods where 6.02 (p=0.049) for scaled multinomial, 5.52 (p=0.063) and 5.19 (p=0.075) under diagonal and unrestricted scaling methods. Table 7.4 shows similar results for main effects when the model is fitted without interaction terms.

7.6 DISCUSSION

A GLMM based strategy for dealing with overdispersed multinomial data has been presented in this chapter. A variety of methods has been proposed in the literature for dealing with such datasets. However, the method proposed offers greater flexibility in analysis. It offers the flexibility of accommodating overdispersion at various levels and making inference at both population and individual or cluster levels. When there are only two response categories the method is consistent with binomial logit mixed model, a recognised method of analysis for overdispersed binomial data.

The method is also shown to be applicable to the analysis of contingency tables with clustered observations under the assumption of multinomial sampling. This can be considered as an alternative to the method discussed in the previous chapter for analysis of clustered contingency tables. The advantage of this method is that the number of parameters to be dealt with is much less than that under a Poisson-normal model.

CHAPTER EIGHT

ANALYSIS OF CONTINGENCY TABLES WITH ORDERED CATEGORIES AND CLUSTERED OBSERVATIONS

8.1 INTRODUCTION

The use of ordinal scales is very widespread in social and health sciences. Attitudes, opinions, severity or stages of various conditions or situations are often measured by using ordinal scales which often result in contingency tables with ordered categories. However, most of the commonly used statistical methods for analysing contingency tables often ignore the ordinal relationship between categories and treat these as nominal categories. The conventional analysis using log-linear modelling or test of independence using Pearson chi-square statistic often treat ordinal variables as nominal variables. Ignoring the additional information of the ordered relationship among categories may lead to less power for detecting an alternative hypothesis. In the analysis of contingency tables ordinal methods of analysis can considerably reduce the number of parameters to be tested for goodness of fit or test of independence compared to that of nominal methods. The methods for analysing ordinal categorical data have been developed considerably over the last couple of decades. The detailed discussion of conventional methods of analysis for ordinal data can be found in the books of Bishop et al. (1975), Gokhale and Kullback (1978), Goodman (1978), Haberman (1974b, 1978, 1979), Fienberg (1980), Gilbert (1981) and Agresti (1984) among others.

In McCullagh (1980) a class of regression models has been developed for analysing ordinal response variables which are known as threshold models. As reflected in the recent literature the threshold models can be considered as the most appropriate method for analysing ordered response variables. The method has subsequently been extended to include both fixed and random effects by Harville and Mee (1984) and Zhaorong et al. (1992) and Saei (1996).

The problem of analysing contingency tables with clustered observations discussed in the previous two chapters is extended here to the case where categories are ordinal. A strategy for analysing such contingency tables with the aid of mixed threshold models is developed.

8.2 LOG-LINEAR MODELS FOR ORDERED CATEGORIES

Agresti (1984) provides a review of the methods used for modelling cross-classified data with ordered categories. The essence of these methods is to replace the interaction terms in conventional log-linear models by lesser number of regression coefficients while assigning a score for the ordinal categories. In a two-way ordinal-ordinal cross-classification, a single regression coefficient is used to replace all interaction terms as explained below.

If m_{ij} is the expected frequency in row i and column j in a two dimensional contingency table with i=1,2,..,I and j=1,2,..,J then a saturated log-linear model can be written as

$$(8.1) \qquad \log m_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} .$$

113

The above model is saturated and includes (I-1)(J-1) interaction parameters and a test for independence will have (I-1)(J-1) degrees of freedom (df). If both row and column categories are ordinal then these interactions terms are replaced by a single regression term as follows

$$(8.2) \qquad \log m_{ij} = \mu + \alpha_i + \beta_j + \gamma(u_i - \bar{u})(v_j - \bar{v})$$

where u_i and v_j are ordered scores for rows and columns respectively. The choice of scores reflect the assumed distances between categories for underlying interval scales. In the absence of any information the scores can be considered as equi-spaced. The advantage of the above model is that the number of parameters to be estimated is much less than the saturated model and the df for goodness of fit is (I-1)(J-1)-1 and the df for testing independence is 1 only.

If the cross-classification categories are ordinal-nominal i.e. columns are ordinal but rows are nominal, say, then an equivalent model can be written as

$$(8.3) \qquad \log m_{ij} = \mu + \alpha_i + \beta_j + \gamma_i(v_j - \bar{v})$$

with (I-1)(J-2) df for goodness of fit and (I-1) df for test of independence.

The models can also be interpreted in terms of log odds or log odds ratios. For the ordinal-nominal model presented above, the log odds for an arbitrary pair of rows i and i' is,

$$(8.4) \qquad \log \frac{m_{ij}}{m_{i'j}} = (\gamma_i - \gamma_{i'}) + (\gamma_i - \gamma_{i'})(v_j - \bar{v})$$

114

that is the log odds is the linear function of the scores with slope as the difference between the slopes for i and i′ rows. Similarly the log odds ratio for an arbitrary pair of rows i and i′ and for an arbitrary pair of columns j<j′ ,

$$(8.5) \qquad \log \frac{m_{ij}/m_{ij'}}{m_{i'j}/m_{i'j'}} = (\gamma_i - \gamma_{i'})(v_j - v_{j'})$$

implying the odds ratio is proportional to the difference between column scores.

For the ordinal-ordinal model, the log odds ratio for an arbitrary pairs of rows i<i′ and an arbitrary pairs of columns j<j′ is given by

$$(8.6) \qquad \log \frac{m_{ij}/m_{ij'}}{m_{i'j}/m_{i'j'}} = \gamma(u_i - u_{i'})(v_j - v_{j'}) \ .$$

Therefore, the odds ratio is proportional to the product of the differences between appropriate row and column scores.

A number of other authors such as Birch (1963), Haberman (1974b), Goodman (1979), Andrich (1979), Duncan and McRae (1979) discuss this type of modelling in various forms.

In the following sections we develop an analysis strategy for ordinal-nominal contingency tables along the line discussed above but by using threshold models. The proposed analysis strategy will not only utilize ordered relationships for interaction effects but will also extend it to main effects for ordered column categories. The strategy will then be

extended to cover the case where observations are clustered by applying GLMM.

8.3 THRESHOLD MODELS

The concept of threshold models is developed by assuming an underlying continuous and maybe unobservable random variable, τ, with a specific distribution. The categories of the observed ordinal variable Y correspond to the various contiguous intervals of the underlying variable, τ. The cut-points of the intervals are called threshold parameters (θ's) which are unknown. If the categories of Y are denoted as 0,1,....,M then the realised value of Y is recorded as Y = m if $\theta_{m-1} < \tau \leq \theta_m$. As the commonly used distributions of τ ranges over $-\infty$ to ∞ the highest and lowest cut-points of θ can be assigned values as $\theta_{-1} = -\infty$ and $\theta_M = \infty$. The threshold model for Y is then

$$P(Y \leq m) = G(\theta_m - \eta)$$

where $\eta = X\beta$ with X the matrix of independent variable values and β a vector of regression coefficients; $G(\theta_m - \eta)$ is a cumulative distribution function. If X includes a column of 1's for an intercept term then the problem of a lack of identifiability arises which can be overcome by setting $\theta_0 = 0$. That leaves the threshold parameters $\theta_1, \theta_2, ..., \theta_{M-1}$ and the regression parameters in η to be estimated.

The mixed model extension of threshold models is done by adding random components in η as $\eta = X\beta + Zu$, where $u' = [u'_1, u'_2, ..., u'_k]$ is a vector of a set of random parameters and $Z = [Z_1, Z_2, ... Z_k]$ is the corresponding incidence matrix. The assumptions related to the distribution of u_j's and the covariance

structure are as usual, that is $\mathbf{u}_j \sim N(0, \phi_j \mathbf{A}_j)$ or $\mathbf{u} \sim N(0, \mathbf{A})$ with $\mathbf{A} =$ block diag $[\mathbf{A}_j \phi_j]$ as the variance-covariance matrix of $\mathbf{u}$.

Depending on the distributional form of the underlying variable τ and hence the cumulative distribution function, G, various threshold models can be defined as follows:

1) If $G(\theta_m - \eta) = \Phi(\theta_m - \eta)$, where Φ is the cumulative distribution function of the standard normal distribution, then the standard threshold model is achieved.

2) If the logistic function is used i.e. $G(\theta_m - \eta) = \dfrac{\exp(\theta_m - \eta)}{1 + \exp(\theta_m - \eta)}$ and then the proportional odds model is obtained. This model will be used mainly in the present chapter for analysis of contingency tables.

3) If $G(\theta_m - \eta) = 1 - \exp[-\exp(\theta_m - \eta)]$, i.e. the extreme minimal value distribution, then the proportional hazards model is obtained.

4) If $G(\theta_m - \eta) = \exp[-\exp(\theta_m - \eta)]$ i.e. then the extreme maximal value distribution is used.

8.4 ESTIMATION IN MIXED THRESHOLD MODELS

The parameters of a mixed threshold model is obtained by following the usual method of maximising the joint likelihood function $l = l_1 + l_2$ for mixed models. The definitions of l_1 and l_2 are same as described in the previous

chapters. However, the form of the conditional likelihood function l_1 is different which is discussed below. Given that

$$P(Y \leq m) = G(\theta_m - \eta),$$

if y is a N×1 vector of ordinal responses with i[th] realisation $Y_i = y_i = m$, where m can be one of 0,1,...,M, then the probability that $Y_i = y_i$ is given by

$$P(Y_i = y_i) = P_i = G(\theta_{y_i} - \eta_i) - G(\theta_{y_i-1} - \eta_i).$$

Then the log-likelihood function for y can be written as

$$l_1 = \sum_{i=1}^{N} \ln [G(\theta_{y_i} - \eta_i) - G(\theta_{y_i-1} - \eta_i)].$$

(8.7)

$$l_2 = -(1/2) \sum_{j=1}^{k} [v_j \ln(2\pi\phi_j) + u_j' A_j^{-1} u_j]$$

and l_2 is the as usual log-likelihood function for k random components under the normal assumption as defined in the previous chapters. The notations are as usual with u_j representing a random component with dimension v_j .

Then the estimates of $\theta' = [\theta_1, \theta_2, ..., \theta_{M-1}]$, β and **u** are obtained by maximising the joint likelihood function $l = l_1 + l_2$ following the usual Newton-Raphson iterative scheme as follows

(8.8)
$$V \begin{bmatrix} \Delta\theta \\ \Delta\beta \\ \Delta u \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & X' \\ 0 & Z' \end{bmatrix} \begin{bmatrix} \partial l_1 / \partial \theta_0 \\ \partial l_1 / \partial \eta_0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ A_0^{-1} u_0 \end{bmatrix}$$

where Δ indicates changes in estimates in each iteration and the subscript 0 indicates the initial starting values and subsequently the values in each iterations. The matrix **V** is defined as follows

$$(8.9) \quad V = \begin{bmatrix} I & 0 \\ 0 & X' \\ 0 & Z' \end{bmatrix} \begin{bmatrix} -\partial^2 l_1/\partial\theta_0\partial\theta_0' & -\partial^2 l_1/\partial\theta_0\partial\eta_0' \\ -\partial^2 l_1/\partial\eta_0\partial\theta_0' & -\partial^2 l_1/\partial\eta_0\partial\eta_0' \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & X & Z \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & A_0^{-1} \end{bmatrix}$$

and the variance covariance matrix is given by

$$V^{-1} = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix}$$

and the derivatives can be expressed by defining

$$(8.10) \quad Q_i = g(\theta_{m_i}-\eta_i)/P_i, \quad Q_i^* = g(\theta_{m_i-1}-\eta_i)/P_i \ ,$$

where $g(.)$ is the derivative of $G(.)$ and θ_{m_i} can be one of the possible cutoff points.

Then the first derivatives can be written as

$$\partial l_1/\partial\eta_i = -(Q_i - Q_i^*)$$

$$\partial l_1/\partial\theta_m = \sum_{y_i=m} Q_i - \sum_{y_i=m+1} Q_i^*$$

The second order derivatives can be expressed by defining

$$D_i = \partial[\ln g(\theta_{y_i}-\eta_i)]/\partial\theta_{y_i}$$

$$(8.11)$$

$$D_i^* = \partial[\ln g(\theta_{y_i-1}-\eta_i)]/\partial\theta_{y_i-1}$$

Note that $D_i=0$ and $D_i^*=0$ when $y_i=0$.

$$\partial^2 l/\partial\theta_m\partial\theta_{m'} = \sum_{y_i=m} Q_i(D_i-Q_i) - \sum_{y_i=m+1} Q_i^* (D_i^*+Q_i^*) \quad , \quad m'=m$$

$$= \sum_{y_i=m} Q_i Q_i^* \ , \quad m'=m-1$$

$$= \sum_{y_i=m+1} Q_i Q_i^* \ , \quad m'=m+1$$

$$= 0 \quad \text{otherwise}$$

119

$$\partial^2 l_1 / \partial \theta_m \partial \eta_i = \sum_{y_i = m} Q_i(Q_i - Q_i^* - D_i) - \sum_{y_i = m+1} Q_i^* (Q_i - Q_i^* - D^*)$$

$$\partial l_1 / \partial \eta_i \partial \eta_{i'} = \sum_{y_i = m} [Q_i D_i - Q_i^* D_i^* - (Q_i - Q_i^*)^2] \quad , \quad i = i'$$

$$= 0 \quad \text{otherwise}$$

Using the derivatives in estimation equation (8.8) the converged values of threshold and regression parameters are obtained through iteration. After convergence the estimate of dispersion parameter is calculated as

$$\phi_{j(ML)} = [u_j' A_j^{-1} u_j + \text{tr } A_j^{-1} T_{jj}^*] / v_j$$

$$\phi_{j(REML)} = [u_j' A_j^{-1} u_j + \text{tr } A_j^{-1} T_{jj}] / v_j$$

where T_{jj} is the appropriate sub-matrix of T_{33} and T_{jj}^* is the appropriate sub-matrix of V_{33}^{-1}. The iteration process is continued until the convergence in ϕ is also obtained.

8.5 THRESHOLD MODELS FOR CONTINGENCY TABLES

Let us consider a IxJ contingency table which may have observations from K clusters. To undertake an analysis by incorporating random effects for the clusters a separate contingency table is formed using observations from each cluster. Then the combined contingency table can be defined as of dimension IxJxK. The J ordinal column categories are considered as the response categories with values 1,2,...,J (0 is excluded for consistency) in the threshold model while the row categories are treated as explanatory variables. In the following description the row categories will be considered as nominal.

8.5.1 Fixed effect Analysis

Let f_{ij} be the frequency count in the ij^{th} cell. Since $\eta_{ij}=\eta_i$ the log-likelihood function can be written as

$$(8.12) \qquad l_1 = \sum_{i=1}^{I} \sum_{j=1}^{J} f_{ij} \ln [G(\theta_j-\eta_i) - G(\theta_{j-1}-\eta_i)].$$

If $\eta=[\eta_1,\eta_2,...,\eta_I]'$ we define $\eta=X\beta$ where X is a design matrix of dimension $I \times I$. Each column of X corresponds with a row effect and one of the row effects is confounded with the intercept. β is a $I \times 1$ vector of coefficients. As a vector of 1's is included in the design matrix we set $\theta_1=0$ and $\theta_0=-\infty$ and $\theta_J=\infty$ with $G(-\infty)=0$ and $G(\infty)=1$. Therefore, the remaining $J-2$ values of θ's are required to be estimated, which can be included in a vector $\theta'=[\theta_2,\theta_3,...,\theta_{J-1}]$. The estimates of θ and β are now obtained by maximising the likelihood. If

$$P_{ij} = G(\theta_j-\eta_i) - G(\theta_{j-1}-\eta_i)$$

$$(8.13) \qquad Q_{ij} = g(\theta_j-\eta_i)/P_{ij}, \qquad Q^*_{ij} = g(\theta_{j-1}-\eta_i)/P_{ij}$$

$$D_{ij} = \partial[\ln g(\theta_j-\eta_i)]/\partial\theta_j$$

$$D^*_{ij} = \partial[\ln g(\theta_{j-1}-\eta_i)]/\partial\theta_{j-1}$$

Note that $D_{iJ}=0$ and $D^*_{i0}=0$ for all i .

Then the derivatives can be expressed as

$$\partial l_1/\partial\beta = -\sum_{j=1}^{J} \sum_{i=1}^{I} f_{ij}(Q_{ij}-Q^*_{ij})x_i$$

$$\partial l_1/\partial\theta_j = \sum_{i=1}^{I} f_{ij} Q_{ij} - \sum_{i=1}^{I} f_{i,j+1}Q^*_{i,j+1}$$

$$\partial^2 l/\partial\theta_j\partial\theta_{j'} = \sum_{i=1}^{I} f_{ij} Q_{ij}(D_{ij}-Q_{ij}) - \sum_{i=1}^{I} f_{i,j+1} Q^*_{i,j+1} (D^*_{i,j+1}+Q^*_{i,j+1}) \;,\; j'=j$$

$$= \sum_{i=1}^{I} f_{ij} \, Q_{ij} \, Q_{ij}^* \quad , \qquad j'=j-1$$

$$= \sum_{i=1}^{I} f_{i,j+1} \, Q_{i,j+1} \, Q_{i,j+1}^* \quad , \qquad j'=j+1$$

$$= 0 \quad \text{otherwise}$$

$\partial^2 l_1 / \partial\theta_j \partial\beta =$

$$\sum_{i=1}^{I} f_{ij} \, Q_{ij} \, (Q_{ij}-Q_{ij}^*-D_{ij}) \, x_i \; - \; \sum_{i=1}^{I} f_{i,j+1} \, Q_{i,j+1}^* \, (Q_{i,j+1}-Q_{i,j+1}^*-D_{i,j+1}^*) \, x_i$$

$$\partial^2 l_1 / \partial\beta \partial\beta' = \sum_{j=1}^{J} \sum_{i=1}^{I} f_{ij} \, [Q_{ij}D_{ij}-Q_{ij}^*D_{ij}^*-(Q_{ij}-Q_{ij}^*)^2] \, x_i x_i'$$

The estimates can then be obtained by using the Newton-Raphson method of iteration as follows

$$(8.14) \qquad \begin{bmatrix} \tilde{\theta} \\ \tilde{\beta} \end{bmatrix} = \begin{bmatrix} \theta_0 \\ \beta_0 \end{bmatrix} + \mathbf{V}^{-1} \begin{bmatrix} \partial l_1/\partial\theta_0 \\ \partial l_1/\partial\beta_0 \end{bmatrix} \quad \text{where } \mathbf{V} = \begin{bmatrix} -\partial^2 l_1/\partial\theta\partial\theta' & -\partial^2 l_1/\partial\theta\partial\beta' \\ -\partial^2 l_1/\partial\theta'\partial\beta & -\partial^2 l_1/\partial\beta\partial\beta' \end{bmatrix} .$$

8.5.2 Mixed Effect Model

Let f_{kij} be the frequency in the jth response category in the ith row of the kth cluster. The joint likelihood function can be written as $l = l_1+l_2$ where, using $\eta_{kij}=\eta_{ki}$,

$$(8.15) \qquad l_1 = \sum_{k=1}^{K} \sum_{i=1}^{I} \sum_{j=1}^{J} \ln \, [G(\theta_j-\eta_{ki}) - G(\theta_{j-1}-\eta_{ki})]$$

and $\eta_{ki} = x_{ki}'\beta+z_{ki}'u \; \Rightarrow \; \eta = X\beta+Zu$

with $\eta' = [\eta_{11},...,\eta_{1I},........,\eta_{K1},...,\eta_{KI}]$, X is a K×I×I design matrix for the fixed effects corresponding to I rows and Z is a K×I×K incidence matrix for the random effects corresponding to K clusters. Z can be designed to

122

accommodate clustering at multi-levels. β and $\mathbf{u}$ are parameter vectors of dimensions I and K respectively.

The random effects in $\mathbf{u}$ is assumed to be distributed as $N(0,\sigma_1^2 I)$. Therefore l_2 is the usual likelihood function of a normal variable.

Let us redefine the following quantities to obtain the derivatives

$$P_{kij} = G(\theta_j - \eta_{ki}) - G(\theta_{j-1} - \eta_{ki})$$

(8.16)
$$Q_{kij} = g(\theta_j - \eta_{ki})/P_{kij}, \quad Q_{kij}^* = g(\theta_{j-1} - \eta_{ki})/P_{kij}$$

$$D_{kij} = \partial[\ln g(\theta_j - \eta_{ki})]/\partial\theta_j$$

$$D_{kij}^* = \partial[\ln g(\theta_{j-1} - \eta_{ki})]/\partial\theta_{j-1}$$

Note that $D_{kiJ}=0$ and $D_{ki0}^*=0$ for all k and i .

Then the derivatives can be expressed as

$$\partial l_1/\partial\beta = -\sum_{k=1}^{K}\sum_{i=1}^{I}\sum_{j=1}^{J} f_{kij}(Q_{kij} - Q_{kij}^*) \, x_{ki}$$

$$\partial l_1/\partial\mathbf{u} = -\sum_{k=1}^{K}\sum_{i=1}^{I}\sum_{j=1}^{J} f_{kij}(Q_{kij} - Q_{kij}^*) \, z_{ki}$$

$$\partial l_1/\partial\theta_j = \sum_{k=1}^{K}\sum_{i=1}^{I} f_{kij} \, Q_{kij} - \sum_{k=1}^{K}\sum_{i=1}^{I} f_{ki,j+1}Q_{ki,j+1}^*$$

$\partial^2 l/\partial\theta_j\partial\theta_{j'} =$

$$\sum_{k=1}^{K}\sum_{i=1}^{I} f_{kij}Q_{kij}(D_{kij} - Q_{kij}) - \sum_{k=1}^{K}\sum_{i=1}^{I} f_{ki,j+1}Q_{ki,j+1}^*(D_{ki,j+1}^* + Q_{ki,j+1}^*), \quad j'=j$$

$$= \sum_{k=1}^{K}\sum_{i=1}^{I} f_{kij} \, Q_{kij} \, Q_{kij}^* \quad , \quad j'=j-1$$

$$= \sum_{k=1}^{K}\sum_{i=1}^{I} f_{i,j+1} \, Q_{i,j+1} \, Q_{ki,j+1}^* \quad , \quad j'=j+1$$

$$= 0 \quad \text{otherwise}$$

$$\partial^2 l_1 / \partial\theta_j \partial\beta = \sum_{k=1}^{K}\sum_{i=1}^{I} f_{kij} Q_{kij} (Q_{kij} - Q^*_{kij} - D_{kij}) x_{ki} -$$

$$\sum_{k=1}^{K}\sum_{i=1}^{I} f_{i,j+1} Q^*_{ki,j+1} (Q_{i,j+1} - Q^*_{ki,j+1} - D^*_{i,j+1}) x_{ki}$$

$$\partial^2 l_1 / \partial\theta_j \partial\mathbf{u} = \sum_{k=1}^{K}\sum_{i=1}^{I} f_{kij} Q_{kij} (Q_{kij} - Q^*_{kij} - D_{kij}) z_{ki} -$$

$$\sum_{k=1}^{K}\sum_{i=1}^{I} f_{i,j+1} Q^*_{ki,j+1} (Q_{i,j+1} - Q^*_{ki,j+1} - D^*_{i,j+1}) z_{ki}$$

$$\partial^2 l_1 / \partial\beta\partial\beta' = \sum_{k=1}^{K}\sum_{i=1}^{I}\sum_{j=1}^{J} f_{kij} [Q_{kij} D_{kij} - Q^*_{kij} D^*_{kij} - (Q_{kij} - Q^*_{kij})^2] \, x_{ki} x'_{ki}$$

$$\partial^2 l_1 / \partial\mathbf{u}\partial\mathbf{u}' = \sum_{k=1}^{K}\sum_{i=1}^{I}\sum_{j=1}^{J} f_{kij} \, [Q_{kij} D_{kij} - Q^*_{kij} D^*_{kij} - (Q_{kij} - Q^*_{kij})^2] \, z_{ki} z'_{ki}$$

The estimates can then be obtained by using the Newton-Raphson method of iteration as follows

$$(8.17) \qquad \begin{bmatrix} \theta \\ \beta \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \theta_0 \\ \beta_0 \\ \mathbf{u}_0 \end{bmatrix} + \mathbf{V}^{-1} \begin{bmatrix} \partial l_1 / \partial\theta_0 \\ \partial l_1 / \partial\beta_0 \\ \partial l_1 / \partial\mathbf{u}_0 \end{bmatrix} - \sigma_1^2 \mathbf{V}^{-1} \begin{bmatrix} 0 \\ 0 \\ \mathbf{u}_0 \end{bmatrix}$$

where $\mathbf{V} = \begin{bmatrix} -\partial^2 l_1/\partial\theta\partial\theta' & -\partial^2 l_1/\partial\theta\partial\beta' & -\partial^2 l_1/\partial\theta\partial\mathbf{u}' \\ -\partial^2 l_1/\partial\beta\partial\theta' & -\partial^2 l_1/\partial\beta\partial\beta' & -\partial^2 l_1/\partial\beta\partial\mathbf{u}' \\ -\partial^2 l_1/\partial\mathbf{u}\partial\theta' & -\partial^2 l_1/\partial\mathbf{u}\partial\beta' & -\partial^2 l_1/\partial\mathbf{u}\partial\mathbf{u}' + \sigma_1^2 I \end{bmatrix}$

8.5.3 Multi-level Clustering

When the observations are clustered at multiple levels say at household level, at block level and also at primary sampling unit (PSU) level, the above development is still applicable. The contingency table has to be

124

formed at the smallest level of clustering and then the **Z** matrix has to be formed such that it reflects the clustering at various levels and the vector of random effects **u** will have more than one term at each level of clustering. The number of effects in each term will depend of the number of clusters at different levels. If there are K_1 households, K_2 blocks and K_3 PSU's then Z will have $K_1+K_2+K_3$ columns and **u** will have three components with K_1, K_2 and K_3 effects respectively. The distribution of **u** will then be multivariate normal. If **u** has M components due to clustering at multiple levels then $\mathbf{u} \sim N(\mathbf{0}, \mathbf{A})$ with $\mathbf{A}$ = block diag $[\mathbf{A}_m\phi_m]$ where $\phi' = [\phi_1, \phi_2, ..., \phi_M]$ is a vector of dispersion parameters. The method of estimation is the same as above with $\sigma_1^2 I$ replaced by **A**.

8.5.4 Test of Independence

The test of independence in a contingency table can then be undertaken either by using the Wald test or the likelihood ratio test as described in the previous chapters. For a likelihood ratio test, initially the model will be fitted by including intercept and row effects and then by excluding the row effects. For a mixed model the likelihood ratio statistic will be calculated as discussed in the previous chapters (see equations 6.14, 7.11).

The test of independence under threshold modelling is more powerful than that under conventional nominal log-linear modelling because of the lesser number of degrees of freedom (df) for χ^2 statistics. If the alternative hypothesis of association is true then with the increase of sample size, χ^2 under nominal modelling would increase at a slower rate than under threshold modelling. Even if the magnitudes of χ^2's under both modelling are similar,

125

threshold modelling would give smaller a p-value because of the smaller df for χ^2.

8.6 APPLICATIONS

The method is initially applied to a number of two-way contingency tables where observations are not clustered and therefore fixed effect threshold models are fitted. The chi-square values for test of independence are compared with that of ordinary log-linear analysis under Poisson assumption and with the method of analysis discussed by Agresti (1984) for ordinal categories. As threshold models only utilise the ordered relationship in the categories of the response variable the comparable model is Agresti's ordinal-nominal model instead of ordinal-ordinal model. In ordinal-nominal model, interaction effects are redefined by introducing a uniform association parameter for each row category. This reduces the number of parameters to be estimated for association which leads to a chi-square test with lesser df. The methods have been applied to three datasets - Dumping Severity, Pain and Spasm and Mental Health datasets. The objective is to show that threshold modelling offers an alternative method for analysis of contingency tables with ordered response categories.

Finally, the method with mixed effects is applied to a dataset where not only are categories ordered but also observations are clustered. The application demonstrates the use of mixed effects threshold model and compares the results with fixed effect model.

126

8.6.1 Fixed Effect Threshold Modelling

As discussed earlier, four different cumulative distribution functions can be used in threshold modelling. As Saei (1996) concludes that the use of different functions does not make any significant difference in estimates, we use proportional odds model i.e. logistic function in all of the following examples. The expressions of the cumulative function and derivations of relevant quantities required for estimation are discussed for the mixed model case as applied later in the Neighbourhood dataset.

Application to Dumping Severity Data

Dumping Severity data reported by Grizzle et al. (1969) which shows the extent of side effects (dumping severity) of four different types of operations for treatment of duodenal ulcer patients. Data are given in Table 8.1. The objective of the analysis is to examine whether there is any association between dumping severity and types of operations. The categories of the response variable, Dumping severity, are ordered (none, slight, moderate). There is also an ordered relationship among types of operations but in this analysis the operation types are treated as nominal categories. Table 8.2 presents the likelihood ratio statistics (G^2) for the test of independence obtained under different methods of analysis. The G^2 value corresponding to the log-linear model has higher degrees of freedom than both ordinal-nominal and threshold models. The test of association is not significant at 5% level under all methods but as expected the p-value for the threshold model is smaller than the other two methods.

Table 8.1: Patients by Dumping Severity and Operation Methods

Operation	Dumping Severity		
	None	Slight	Moderate
A	61	28	7
B	68	23	13
C	58	40	12
D	53	38	16

Table 8.2: Test of Independence for Dumping Severity Data

Method of Analysis	G^2	df	p-value
Log-linear model	10.88	6	0.09
Ordinal-nominal model	6.48	3	0.09
Threshold model	7.31	3	0.06

Application to Pain and Spasm Data

The dataset presented in Table 8.3 is reported by Miller and Landis (1991) from a study comparing two drugs for the relief of pain and spasm. The response variable for the dataset has three-point ordered categories corresponding to two row categories i.e. drug and placebo. Table 8.4 presents the G^2 values under three methods of analysis. The p-value is the smallest for the threshold model (p=0.001) followed by the ordinal-nominal model (p=0.002) and the p-value is the highest (p=0.003) under the log-linear model. The calculated G^2 under the threshold model is highly consistent with that of ordinal-nominal model.

Table 8.3: Severity of Pain and Spasm for Treatment and Control

Treat-ment	Response		
	Worse or no change	Slight Improvement	More Improve-ment or Cure
Drug	24	15	68
Placebo	36	25	45

Table 8.4: Test of Independence for Pain and Spasm Data

Method of Analysis	G^2	df	p-value
Log-linear model	11.58	2	0.003
Ordinal-nominal model	9.18	1	0.002
Threshold model	9.95	1	0.001

Application to Mental Health Data

Srole et al. (1978) presents a contingency table showing the mental health status of a group of offsprings by parent's socio-economic status (Table 8.5). The column variable mental health status has ordered categories which is treated as the response variable in a threshold model. The row variable categories are also ordered but in this analysis the categories are considered as nominal. The objective of the analysis is to examine any significant association between mental health and socio-economic status. Likelihood ratio statistics calculated under different methods are presented in Table 8.6. All three methods of analysis indicating a significant association in the table. However, chi-square values under ordinal-nominal and threshold models are much higher than the log-linear model when df is taken into account.

Table 8.5: Data on Mental Health and Parent's Socio-economic Status

	Mental Health Status			
Parent's Socio-economic Status	Well	Mild Symptom	Moderate Symptom	Impaired
A (high)	64	94	58	46
B	57	94	54	40
C	57	105	65	60
D	72	141	77	94
E	36	97	54	78
F (low)	21	71	54	71

Table 8.6: Test of Independence for Mental Health Data

Method of Analysis	G²	df	p-value
Log-linear model	47.42	15	0.000
Ordinal-nominal model	40.59	5	0.000
Threshold model	39.60	5	0.000

8.6.2 Mixed Effect Threshold Modelling

Application to Neighbourhood data

The method is applied to the Neighbourhood dataset analysed in the previous chapters where the categories are treated as nominal. In the case of multinomial analysis the satisfaction with own house was considered as the response variable. Although the responses i.e. unsatisfied, satisfied, very satisfied have an ordered relationship these were treated as nominal categories in the previous chapters. The threshold modelling will utilise the extra information of the ordered relationship while allowing for the clustering of the observations. That will not only increase the power of the test by reducing the number of fixed parameters to be estimated but it will also reduce the number of random components. To fit a mixed threshold model the overall contingency table will be separated into 20 contingency tables for 20 neighbourhoods each with three rows. If the response variable is denoted as Y with ordered values j=1,2,3 (say) then the model can be fitted as follows.

Let Y_{ki} be the response from a household with i^{th} level of satisfaction with neighbourhood housing, from the kth neighbourhood, where i=1,2,3 and k=1,2,...,20. The level of satisfaction with community housing was

130

considered here as a nominal variable. Then the cumulative probability distribution can be written as

$$P(Y_{ki} \leq j) = G(\theta_j - \eta_{ki}) \quad \text{where} \quad \eta_{ki} = x'_{ki}\beta + z'_{ki}u \ .$$

Here, X is a 60×3 design matrix corresponding to the fixed effects for the levels of row variables i.e. satisfaction with neighbourhood housing, Z is a 60×20 design matrix corresponding to the random effects for 20 neighbourhoods and η is a 60×1 vector. As $\theta_1 = 0$ and $\theta_3 = \infty$ the only threshold parameter to be estimated here is θ_2. Let f_{kij} represent the frequency in k,i,j^{th} cell.

As mentioned above the proportional odds model is used in all applications of threshold modelling. The expressions for cumulative function, derivatives and other quantities for proportional odds model are as follows.

(8.18)
$$G(\theta_{y_{ki}} - \eta_{ki}) = \exp(\theta_{y_{ki}} - \eta_{ki})/[1 + \exp(\theta_{y_{ki}} - \eta_{ki})].$$

$$g(\theta_{y_{ki}} - \eta_{ki}) = \exp(\theta_{y_{ki}} - \eta_{ki})/[1 + \exp(\theta_{y_{ki}} - \eta_{ki})]^2$$

As mentioned above y_{ki} can have ordered values $j = 1,2,3$. We have

$$P_{kij} = \frac{\exp(\theta_j - \eta_{ki})}{[1 + \exp(\theta_j - \eta_{ki})]} - \frac{\exp(\theta_{j-1} - \eta_{ki})}{[1 + \exp(\theta_{j-1} - \eta_{ki})]}$$

$$Q_{kij} = \frac{\exp(\theta_j - \eta_{ki})}{[1 + \exp(\theta_j - \eta_{ki})]^2} P_{kij}^{-1} \ , \quad Q_{kij}^{*} = \frac{\exp(\theta_{j-1} - \eta_{ki})}{[1 + \exp(\theta_{j-1} - \eta_{ki})]^2} P_{kij}^{-1}$$

$$D_{kij} = 1 - \frac{2}{1 + \exp(\theta_j - \eta_{ki})} \ , \quad D_{kij}^{*} = 1 - \frac{2}{1 + \exp(\theta_{j-1} - \eta_{ki})}$$

Using the above quantities, the expressions for the derivatives for the proportional odds model are obtained and parameters are estimated as

131

discussed in the previous section. Table 8.7 presents the parameters estimates and test statistics for the tests of independence. The estimates obtained under the mixed effects model are based on the REML method. The likelihood ratio statistic (G^2) is calculated by fitting the models with and without row effects. In the case of mixed models G^2 is calculated following Lee and Nelder (1996) as discussed in the previous chapters.

Table 8.7: Parameter Estimates and Test of Independence Under Fixed and Mixed Effects Threshold Modelling.

Parameters	Fixed effect Model		Mixed effect Model		
Full Model	Est.	SE	Est.	SE	ϕ
θ_2	2.68	0.42	2.93	0.45	0.78
Int	-1.13	0.47	-1.05	0.53	
β_2	1.19	0.53	1.14	0.56	
β_3	2.47	0.77	2.26	0.80	
Deviance	97.82	-	98.36	-	
Reduced Model					
θ_2	2.46	0.39	2.77	0.42	0.97
Int	-0.08	0.20	-0.52	0.31	
Deviance	109.60	-	110.18	-	
W^2	10.53	-	8.23	-	
p-value	0.005		0.016		
G^2	11.78	-	8.15	-	
p-value	0.003		0.016		

The estimates of parameters and SEs under mixed models are slightly different than those under fixed effect models. However, both the Wald statistic, W^2 and likelihood ratio statistic, G^2 for test of independence are deflated under the mixed effect modelling compared to the fixed effect modelling. Under the fixed effect modelling $W^2 = 10.53$ and $G^2 = 11.78$ which are significant compared with $\chi^2_{2,0.05}$. For the mixed effect modelling $W^2 = 8.23$ and $G^2 = 8.15$. In both cases the tests are significant compared with

$\chi^2_{2,0.05}$. However, the p-values are larger under the mixed model. The result is consistent with the conclusions of the Poisson and multinomial modelling undertaken in the previous two chapters. But because of utilising an ordered relationship the test tends to be more powerful as reflected in the p-values which are much smaller here than those obtained in the previous chapters.

8.7 DISCUSSION

It has been demonstrated in this chapter that the threshold model offers an alternative to the standard analysis of contingency tables when categories are ordered. It can also accommodate the clustering of the observations by including random effects in the model. The chi square statistics calculated for test of independence are found to be consistent with the method discussed by Agresti (1984) for ordinal data. The advantage here is that the number of parameters to be estimated is much less compared with alternative methods. This is particularly convenient in the case of mixed effect modelling where the number of parameters to be dealt with is often large. For the same reason the method also enjoys the benefit of increased power of the test when the alternative hypothesis is true. The proposed method has some computational advantages too. By exploiting the fact that the cells which belong to the same row have the same representation to the design matrix, efficiency in computing can be achieved.

CHAPTER NINE

ADJUSTING STANDARD ERRORS OF THE FIXED COEFFICIENTS FOR RANDOM EFFECTS

9.1 INTRODUCTION

Although there has been considerable development in the theory of GLMMs, the use of such modelling in applied research is still very limited. One of the problems preventing the widespread use of GLMMs, is the absence of appropriate software. In most practical applications the number of random effects tends to increase with the number of observations so that in such situations, a full analysis involving the estimation of random effects and the corresponding dispersion parameters appears to be a daunting task, particularly to researchers of other disciplines. The methods discussed in the previous chapters may turn out to be unmanageable when the number of random effects is very large. The main problem is to deal with sparse matrices of very large dimensions.

In most practical applications, interest is mainly in the fixed effects and the standard errors of the estimates rather than the predictions of random effects or their variances. It is often found that the effect of mixed modelling on the estimates of fixed parameters is relatively small. The effect is mainly on the standard errors of the estimates. Under a mixed model the estimates of the standard errors tend to be larger than that of a fixed effect model. Therefore, any risk of making a misleading inference from a fixed effect model which ignores random effects, can be considerably reduced by adjusting the standard errors of the coefficients obtained under a fixed effect model. The extent of adjustment mainly depends on the

dispersion parameter (ϕ) of random effects. If a rough idea of the dispersion parameter of a random term is available then an approximate adjustment can be made easily without fitting a complete mixed model particularly for the applications discussed in the thesis. Although the adjustment is only approximate it would lead to a more conservative inference.

In the current chapter, we show how the standard errors of the estimates of fixed coefficients, obtained under a fixed effect model, can be adjusted for possible random effects if some knowledge of ϕ is available. It is demonstrated that the elements required for making the adjustment can be easily calculated at the cluster level without being required to deal with the full design matrix for random effects. Only cases, for which there are only one or two random terms in the model, are dealt with here. Simulation results are also presented to indicate the general performance of the proposed method of adjustment.

9.2 ADJUSTMENT

Let us consider the fixed effect model

(9.1) $E(\mathbf{y}) = f(\eta)$, $\eta = \mathbf{X}\beta$

where $\eta = [\eta_1, \eta_2, ..., \eta_n]'$, $\mathbf{X}$ is an nxp matrix of known constants and β is an px1 vector of fixed coefficients. If l is the log likelihood function of the observations and if the parameters are obtained by maximising l then the variance-covariance matrix of the estimated parameters is given by

(9.2) $\Sigma = \mathbf{R}^{-1} = [\mathbf{X}'\mathbf{B}\mathbf{X}]^{-1}$

        with $\mathbf{B} = -\partial^2 l / \partial\eta\partial\eta'$.

135

The corresponding mixed effect model, under the standard notation used in the previous chapters, can be written as

$$(9.3) \qquad E(\mathbf{y}) = f(\boldsymbol{\eta}^*), \quad \boldsymbol{\eta}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$$

where $\mathbf{Z}\mathbf{u}$ is the random component which can be partitioned corresponding to the number of random terms in the model. If $l^* = l_1 + l_2$ is the joint log-likelihood function which is maximised to obtain the parameter estimates under GLMM then the variance-covariance matrix of the estimates is given by

$$(9.4) \qquad \mathbf{V}^{-1} = \begin{bmatrix} \Sigma^* & \Sigma_1 \\ \Sigma_1' & \Sigma_2 \end{bmatrix} \quad \text{with } \mathbf{V} = \begin{bmatrix} \mathbf{R}^* & \mathbf{T}^* \\ \mathbf{T}^{*\prime} & \mathbf{D}^* \end{bmatrix} = \begin{bmatrix} \mathbf{X'B^*X} & \mathbf{X'B^*Z} \\ \mathbf{Z'B^*X} & \mathbf{Z'B^*Z + A^{*-1}} \end{bmatrix}$$

where $\mathbf{B}^* = -\,\partial^2 l / \partial \boldsymbol{\eta}^* \partial \boldsymbol{\eta}^{*\prime}$ and $\mathbf{A}^*$ is a block diagonal matrix $[\mathbf{A}_k^* \phi_k^*]$, with ϕ_k^* the dispersion parameter corresponding to the kth random term. These notations are standard as defined in the earlier chapters.

Now the variance-covariance matrix of the fixed coefficients under the mixed model is Σ^* as obtained from (9.4) which can be expressed in terms of the elements of $\mathbf{V}$ matrix as

$$(9.5) \qquad \Sigma^* = [\mathbf{R}^* - \mathbf{T}^* \mathbf{D}^{*-1} \mathbf{T}^{*\prime}]^{-1} \quad .$$

This variance-covariance matrix is obtained under a full mixed model analysis. Following the expression of the above matrix, the variance-covariance matrix calculated under a fixed effect model can be adjusted approximately for possible overdispersion as follows

$$(9.6) \qquad \Sigma_A = [\mathbf{R} - \mathbf{T}\mathbf{D}^{-1}\mathbf{T}']^{-1},$$

where $\mathbf{T} = \mathbf{X'BZ}$ and $\mathbf{D} = \mathbf{Z'BZ + A}^{-1}$. The approximation basically involves replacing $\mathbf{B}^* = -\,\partial^2 l / \partial \boldsymbol{\eta}^* \partial \boldsymbol{\eta}^{*\prime}$ by $\mathbf{B} = -\,\partial^2 l / \partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'$ ·

To calculate this an approximate value, ϕ, of the vector ϕ^* is required. An approximate value of ϕ can be obtained from a previous analysis of similar data. Even if such a value is not available then any best guess of ϕ can be used so that conservative estimates of the variances of fixed effects can be derived. Once the values of ϕ is determined the values of $\mathbf{T}$ and $\mathbf{D}^{-1}$ are easily calculated at the cluster level without being required to handle the full $\mathbf{Z}$ matrix. In the following sections this is demonstrated specifically in the situations when the adjustment is required for only one or two random terms.

9.2.1 Adjustment for One Random Term

If there is only one random term with J effects and observations are sorted by clusters then $\mathbf{X}$ and $\mathbf{Z}$ matrices can be defined as

$$\mathbf{X} = [\mathbf{x}_1', \mathbf{x}_2', ..., \mathbf{x}_J']' \ ,$$

$$\mathbf{Z} = \text{block diag } [\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_J]$$

where $\mathbf{z}_j$ is a vector of 1's with dimension n_j , the size of cluster j ,

similarly, $\mathbf{B} = \text{diag } [\mathbf{B}_1, \mathbf{B}_2, ..., \mathbf{B}_J]$.

The expression (9.3) can be written as

$$(9.7) \qquad \Sigma_A = [\mathbf{R} - \mathbf{T}\mathbf{D}^{-1}\mathbf{T}']^{-1} = [\mathbf{R} - \sum_{j=1}^{J} \mathbf{t}_j \ d_j^{-1} \ \mathbf{t}_j']^{-1}$$

where
$$\mathbf{D} = \text{diag } [d_1, d_2, ..., d_J],$$

with
$$\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_J],$$

$$d_j = \mathbf{z}_j' \mathbf{B}_j \mathbf{z}_j + \phi^{-1} = \mathbf{1}' \mathbf{B}_j \mathbf{1} + \phi^{-1}$$

$$= \phi^{-1} + \text{ sum of the elements of } \mathbf{B} \text{ in cluster } j \ .$$

As d_j's are scalars

137

$$\mathbf{D}^{-1} = \text{diag } [d_1^{-1}, d_2^{-1}, ..., d_J^{-1}],$$

$$t_j = x_j' B_j z_j = x_j' B_j 1 .$$

The dimension of t_j is $p \times 1$ where p is the number of fixed parameters.

$$t_j = [t_{j1}, t_{j2}, ..., t_{jp}]'$$

$$(9.8) \quad \Sigma_A^{-1} = \begin{bmatrix} R_{11} & R_{12} & . & . & R_{1p} \\ R_{21} & R_{22} & . & . & R_{2p} \\ . & . & . & . & . \\ . & . & . & . & . \\ R_{p1} & R_{p2} & . & . & R_{pp} \end{bmatrix} - \begin{bmatrix} \sum_j t_{j1}^2/d_j & \sum_j t_{j1}t_{j2}/d_j & . & . & \sum_j t_{j1}t_{jp}/d_j \\ \sum_j t_{j1}t_{j2}/d_j & \sum_j t_{j2}^2/d_j & . & . & . \\ . & & & & \\ . & & . & . & . \\ \sum_j t_{j1}t_{jp}/d_j & . & & . & \sum_j t_{jp}^2/d_j \end{bmatrix}$$

In many applications $\mathbf{B}$ is diagonal and in that case the elements of $\mathbf{B}$ can be stored as a vector, $\mathbf{L} = [l_1', l_2', ..., l_J']'$. In such situations $t_j = x_j' l_j$ i.e. the sum of the columns of x_j multiplied by l_j and $t_{jp} = x_{jp}' l_j = \sum_i^{n_j} x_{jpi} l_{ji}$.

Simple Linear Regression

The expressions can be further simplified in the case of simple regression with an intercept (β_1) and a regression coefficient (β_2). The variances of the parameter estimates can be adjusted for one random term as follows.

$$(9.9) \quad \Sigma_A^{-1} = \begin{bmatrix} R_{11} & R_{12} \\ R_{12} & R_{22} \end{bmatrix} - \begin{bmatrix} \sum_j t_{j1}^2/d_j & \sum_j t_{j1}t_{j2}/d_j \\ \sum_j t_{j1}t_{j2}/d_j & \sum_j t_{j2}^2/d_j \end{bmatrix}$$

If Δ is the determinant of Σ_A^{-1} then the exact expressions for the variances can be written as

$$\text{Var }(\beta_1) = \frac{1}{\Delta}\left(R_{22} - \sum_j t_{j2}^2/d_j\right),$$

(9.10)

$$\text{Var }(\beta_2) = \frac{1}{\Delta}\left(R_{11} - \sum_j t_{j1}^2/d_j\right).$$

These expressions can be specified further when $\mathbf{B}$ is a diagonal matrix as follows

$$\text{Var }(\beta_1) = \frac{1}{\Delta}\left(R_{22} - \sum_j \frac{(\sum_i l_{ji})^2}{\phi^{-1} + \sum_i l_{ji}}\right),$$

(9.11)

$$\text{Var }(\beta_2) = \frac{1}{\Delta}\left(R_{11} - \sum_j \frac{(\sum_i x_{ji} l_{ji})^2}{\phi^{-1} + \sum_i l_{ji}}\right),$$

and

$$\Delta = \left[(R_{11} - \sum_j\sum_i l_{ji})(R_{22} - \sum_j \frac{(\sum_i x_{ji} l_{ji})^2}{\phi^{-1} + \sum_i l_{ji}})\right] - \left[R_{12} - \sum_j \frac{(\sum_i l_{ji})(\sum_i x_{ji} l_{ji})}{\phi^{-1} + \sum_i l_{ji}}\right]^2$$

Here x_{j2i} is denoted as x_{ji} and x_{j1i} is a vector of 1's corresponding to the intercept. Therefore, to adjust the variances of fixed coefficients for one random term, the only additional elements required to be calculated are $\sum_i l_{ji}$ and $\sum_i x_{ji} l_{ji}$ for each cluster.

9.2.2 Adjustment for Two Random Terms

When there are two random terms in the model the variance covariance matrix can be written as

(9.12) $\quad \mathbf{V}^{-1} = \begin{bmatrix} \Sigma^* & \Sigma_{01} & \Sigma_{02} \\ \Sigma_{01}' & \Sigma_{11} & \Sigma_{12} \\ \Sigma_{02}' & \Sigma_{12}' & \Sigma_{22} \end{bmatrix} \quad$ with

$$V = \begin{bmatrix} R^* & T^* \\ T^{*\prime} & D^* \end{bmatrix} = \begin{bmatrix} R^* & T_1^* & T_2^* \\ T_1^{*\prime} & D_1^* & D_b^* \\ T_2^{*\prime} & D_b^{*\prime} & D_2^* \end{bmatrix} = \begin{bmatrix} X'B^*X & X'B^*Z_1 & X'B^*Z_2 \\ Z_1'B^*X & Z_1'B^*Z_1+(A_1\phi_1)^{-1} & Z_1'B^*Z_2 \\ Z_2'B^*X & Z_2'B^*Z_1 & Z_2'B^*Z_2+(A_2^*\phi_2^*)^{-1} \end{bmatrix}$$

From the above matrix the variances of the fixed coefficients can be expressed as

(9.13) $\Sigma^* = [R^* - T^*D^{*-1}T^{*\prime}]^{-1}$

$$= [R^* - T_1^*M^*T_1^{*\prime} - T_2^*N^{*\prime}T_1^{*\prime} - T_1^*N^*T_2^{*\prime} - T_2^*E^{*-1}T_2^{*\prime}]^{-1}$$

where $D^{*-1} = \begin{bmatrix} M^* & N^* \\ N^{*\prime} & E^{*-1} \end{bmatrix}$ with

$$M^* = D_1^{*-1} + D_1^{*-1}D_b^*E^{*-1}D_b^{*\prime}D_1^{*-1}$$

$$N^* = - D_1^{*-1}D_b^*E^{-1}$$

$$E^* = D_2^* - D_b^{*\prime}D_1^{*-1}D_b^*$$

Now the equivalent estimates for the fixed effect model can be obtained by replacing R^* by R, using an approximate value ϕ of ϕ^*, and calculating all other elements by using B instead of B^*. The expression can be written as

(9.14) $\Sigma_A = [R - T_1MT_1' - T_2N'T_1' - T_1NT_2' - T_2E^{-1}T_2']^{-1}$

For the purpose of computation let us define

$T_1 = [t_{11}, t_{12}, ..., t_{1J}]$, with $t_{1j} = [t_{1j1}, t_{1j2}, ..., t_{1jp}]'$,

$T_2 = [t_{21}, t_{22}, ..., t_{2Q}]$, with $t_{21q} = [t_{2q1}, t_{2q2}, ..., t_{2qp}]'$,

where J and Q are the number of components in the first and second random terms respectively while the number of fixed effect coefficients is denoted by p.

$$t_{1j} = x_j'B_jz_{1j} = x_j'B_j1 , \qquad t_{2q} = x_q'B_qz_{2q} = x_q'B_q1 ,$$

For nested random component models, components of $\mathbf{x}_j$ not equal to zero are a subset of component of $\mathbf{x}_q$ not equal to zero so that $t_{2q} = \sum_{q \in j} t_{1j}$ and when $\mathbf{B}$ is diagonal $t_{1j} = \mathbf{x}_j' \mathbf{l}_j$ i.e. the sum of the columns of $\mathbf{x}_j$ multiplied by $\mathbf{l}_j$ (the diagonal elements of $\mathbf{B}_j$), with $t_{1jp} = \mathbf{x}_{jp}' \mathbf{l}_j = \sum_{j \in i} x_{jpi} l_{ji}$. Similarly, $t_{2qp} = \mathbf{x}_{qp}' \mathbf{l}_q = \sum_{q \in i} x_{qpi} l_{qi}$. These are basically the sums of columns of $\mathbf{X}$ multiplied by $\mathbf{L} = [\mathbf{l}_1', \mathbf{l}_2', ..., \mathbf{l}_J']'$ for the first and the second set of clustering.

$$d_{1j} = \mathbf{z}_{1j}' \mathbf{B}_j \mathbf{z}_{1j} + \phi_1^{-1} = \mathbf{1}' \mathbf{B}_j \mathbf{1} + \phi_1^{-1}$$

$$= \phi_1^{-1} + \text{sum of the elements of } \mathbf{B} \text{ for cluster } j.$$

As d_{1j}'s are scalars

$$\mathbf{D}_1^{-1} = \text{diag } [d_{11}^{-1}, d_{12}^{-1}, ..., d_{1J}^{-1}] \ .$$

Similarly, $d_{2q} = \phi_2^{-1} + \text{sum of the elements of } \mathbf{B}$ for cluster q,

and $\mathbf{D}_2^{-1} = \text{diag } [d_{21}^{-1}, d_{22}^{-1}, ..., d_{2Q}^{-1}]$,

$$\mathbf{D}_b = \begin{bmatrix} \mathbf{d}_{b1} & 0 & . & 0 \\ 0 & \mathbf{d}_{b2} & . & 0 \\ . & . & . & . \\ 0 & 0 & . & \mathbf{d}_{bQ} \end{bmatrix} \quad \text{with } \mathbf{d}_{bq} = [d_{bq1}, d_{bq2}, ..., d_{bqS}]',$$

and $d_{bqs} = \mathbf{z}_{1j}' \mathbf{B}_j \mathbf{z}_{2qj} = \mathbf{1}' \mathbf{B}_j \mathbf{1}$, $j \in q$, which is basically the sum of the elements of $\mathbf{B}_j$ which correspond to the first stage cluster j and the second stage cluster q.

All the elements required for adjustment are basically calculated at the cluster level and there is no need to deal with the full design matrix for any of the random terms.

9.3 SIMULATION RESULTS

A simulation study was undertaken to demonstrate the performance of the proposed adjustment procedure in relation to mixed model estimates. The proposed method of adjustment is based on the assumption that in most applications the estimates of fixed coefficients obtained under fixed effect model (GLM) do not differ to a significant extent than the estimates obtained under mixed effect model (GLMM). The main difference is between the estimates of standard errors (SEs) and if the SEs under GLM can be adjusted then the constructed confidence intervals would be very similar to those under GLMM. The specific objectives of the simulation was to examine the extent of differences between estimates of coefficients under GLM and GLMM and to compare the adjusted SEs with the estimated SEs under GLMM.

The simulation was conducted using binomial logit model. The data was simulated following McGilchrist and Aisbett (1991a) in which 30 observations of a response variable Y_i with distribution $B(6, \pi_i)$ were generated where π_i is given by

$$\pi_i = \exp \eta_i / (1 + \exp \eta_i), \quad \text{with } \eta' = X\beta + Z_1 u_1,$$

where $X' = \begin{bmatrix} 1 & 1 & . & . & 1 & 1 \\ -14 & -13 & . & . & 14 & 15 \end{bmatrix}$, $Z_1' = \begin{bmatrix} 1 & 1 & & & \\ & 1 & 1 & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & 1 & 1 \end{bmatrix}$,

and $u_1' = [U_{1,1}, U_{1,2}, ..., U_{1,15}]$.

$U_{1,i}$'s are independently and identically distributed $N(0, \sigma_1^2)$ random variables and β is the vector of fixed coefficients with two components.

The simulation was conducted for both one and two random components in the model. When there was only one random component in the model the data were generated as described above but in the case of two random components, Z_1 was replaced by $Z = [Z_1, Z_2]'$ and u_1 was replaced by $u = [u_1', u_2']'$ where Z_1 and u_1 are as defined in the case of one random term and Z_2 and u_2 are defined as follows

$$Z_2' = \begin{bmatrix} 1 & 1 & . & . & 1 & & & & & \\ & & & & 1 & 1 & . & . & 1 & \\ & & & & & & 1 & 1 & . & . & 1 \end{bmatrix}, \qquad u_2' = [U_{2,1}, U_{2,2}, U_{2,3}].$$

Once the datasets were simulated the estimates and SEs were obtained under both GLM and GLMM. The estimates of SEs calculated under GLM were then adjusted by using the proposed method and compared with the SEs calculated under GLMM. In undertaking the adjustment the estimate of the dispersion parameter obtained from GLMM was used.

The results of the simulation are presented in Tables 9.1 and 9.2. The estimates and SEs (SE1) presented in the tables are the average over 200 simulations while SE2 is standard deviation of the estimates from 200 simulations. SE2 can be considered as the expected values of SE1.

In both tables, estimates obtained under GLM and GLMM tend to be broadly similar even though the estimates under GLM often tend to be slightly biased. However, in the case of SEs, the estimates under GLM are severely under estimated which is obvious when compared with SE2. The estimated SEs obtained under the proposed method of adjustment are very much consistent with that of GLMM and also with expected SEs (SE2).

143

Table 9.1: Simulation Results for Binomial Logit Model with One Random Component

Simul-ation	True value	Estimate		SE1			SE2	
		GLM	GLMM	GLM	GLMM	Adjusted	GLM	GLMM
1	$\beta_1=0.20$	0.193	0.206	**0.162**	0.296	0.289	0.269	0.282
	$\beta_2=0.10$	0.084	0.095	**0.020**	0.035	0.034	0.036	0.035
	$\phi=1.0$		0.890					
2	$\beta_1=0.20$	0.195	0.204	**0.164**	0.238	0.234	0.215	0.235
	$\beta_2=0.10$	0.096	0.100	**0.020**	0.028	0.028	0.029	0.028
	$\phi=0.5$		0.446					
3	$\beta_1=-0.2$	-0.186	-0.203	**0.162**	0.296	0.289	0.289	0.304
	$\beta_2=-0.1$	-0.086	-0.097	**0.020**	0.035	0.034	0.035	0.036
	$\phi=1.0$		0.942					
4	$\beta_1=-0.2$	-0.182	-0.194	**0.164**	0.243	0.240	0.230	0.240
	$\beta_2=-0.1$	-0.096	-0.100	**0.020**	0.029	0.029	0.028	0.029
	$\phi=0.5$		0.485					

Table 9.2: Simulation Results for Binomial Logit Model with Two Random Components

Simul-ation	True value	Estimate		SE1			SE2	
		GLM	GLMM	GLM	GLMM	Adjusted	GLM	GLMM
1	$\beta_1=0.20$	0.174	0.192	**0.165**	0.483	0.473	0.458	0.475
	$\beta_2=0.10$	0.085	0.102	**0.020**	0.051	0.048	0.048	0.050
	$\phi_1=1.0$		0.858					
	$\phi_2=0.5$		0.573					
2	$\beta_1=0.20$	0.181	0.187	**0.168**	0.379	0.365	0.326	0.335
	$\beta_2=0.10$	0.106	0.108	**0.021**	0.038	0.037	0.039	0.039
	$\phi_1=0.5$		0.470					
	$\phi_2=0.2$		0.270					
3	$\beta_1=-0.2$	-0.172	-0.193	**0.161**	0.432	0.426	0.401	0.439
	$\beta_2=-0.1$	-0.084	-0.093	**0.019**	0.044	0.043	0.043	0.045
	$\phi_1=1.0$		0.871					
	$\phi_2=0.5$		0.447					
4	$\beta_1=-0.2$	-0.192	-0.212	**0.163**	0.377	0.342	0.314	0.327
	$\beta_2=-0.1$	-0.088	-0.091	**0.020**	0.037	0.036	0.038	0.037
	$\phi_1=0.5$		0.435					
	$\phi_2=0.2$		0.251					

9.4 APPLICATION

The method of adjustment was applied to a dataset, published by Crowder (1978), from an experiment on seed germination. The proportion of seed germinated in each of 21 plates in a 2x2 factorial lay-out of seed variety and root extract is presented in Table 9.3. The dataset is analysed by using GLM and GLMM with plate effects as random to account for overdispersion associated with each plate. The calculated SEs under GLM are then adjusted and compared with the calculated SEs under GLMM. The results are presented in Table 9.4.

Table 9.3: Seed Germination Data. Germination Rate in Each of Twenty One Plots by Seed Type and Root Extract.

Seed =>	O. aegyptiaca 75			O. aegyptiaca 73		
Root Extract	r	n	r/n	r	n	r/n
Bean	10	39	0.26	8	16	0.50
	23	62	0.37	10	30	0.33
	23	81	0.28	8	28	0.29
	26	51	0.51	23	45	0.51
	17	39	0.44	0	4	0.00
Cucumber	5	6	0.83	3	12	0.25
	53	74	0.72	22	41	0.54
	55	72	0.76	15	30	0.50
	32	51	0.63	32	51	0.63
	46	79	0.58	3	7	0.43
	10	13	0.77			

The results show that the differences in the estimates under GLM and GLMM are relatively small compared with the corresponding differences in SEs. GLM did not account for over dispersion and consistently under estimated the SEs. The adjusted SEs are almost the same as those of GLMM. The t-tests undertaken for individual factors indicate that the interaction effect is significant under GLM which is in fact not significant under GLMM. The t-statistics calculated using the estimates as obtained under GLM while using

145

the adjusted SEs are also presented in Table 9.4. The adjusted t-statistics are very much similar to those under GLMM implying that an inference based on the adjusted t-test would be consistent with that from GLMM. Particularly the interaction effect turned out to be not significant after the adjustment which was significant under GLM.

Table 9.4: Analysis Results for Seed Germination Data

Method =>	GLM			GLMM			Adjusted		
Parameter	Est	SE	t-val	Est	SE	t-val	Est	SE	t-val
Constant	-0.558	0.126	-4.43*	-0.542	0.188	-2.88*	-0.558	0.189	-2.95*
Seed	0.146	0.223	0.65	0.078	0.305	0.26	0.146	0.306	0.48
Root	1.318	0.178	7.43*	1.337	0.267	5.01*	1.318	0.268	4.92*
Interact	-0.778	0.306	-2.54*	-0.823	0.426	**-1.93**	-0.778	0.426	**1.82**
ϕ				0.096	0.063				

*Significant at 5% level

Table 9.5: Adjustment with a Range of Values (+50%) of ϕ

Method =>	Adjusted with 0.5ϕ		Adjusted with ϕ		Adjusted with 1.5ϕ	
Parameter	SE	t-val	SE	t-val	SE	t-val
Constant	0.162	-3.43*	0.189	-2.95*	0.216	-2.58*
Seed	0.270	0.54	0.306	0.48	0.342	0.43
Root	0.230	5.73*	0.268	4.92*	0.306	4.31*
Interact	0.375	-2.07	0.426	-1.82	0.478	1.63

*Significant at 5% level

The adjustment presented in Table 9.4 is based on the value of ϕ as estimated under GLMM. However, in practice in the absence of GLMM an estimate of ϕ would not be available and an approximate value of ϕ would be used for adjustment. In Table 9.5, results are presented for adjustments with two other values of ϕ which are in fact ±50% of the actual estimate under GLMM. The results show that in this particular example the inference

146

still remains the same. The parameter of interest, the interaction effect, is still not significant. This is also true for other parameters. That means an adjustment using any value of ϕ within ±50% of the actual value would lead to consistent inference in this particular example. Of course this may not be true in all situations but an adjustment even with a very rough value of ϕ would always reduce the risk of wrong inference particularly in favour of significance.

9.5 DISCUSSION

An approximate method for adjusting the estimated SEs of the coefficients in a fixed effect model for possible over dispersion in a dataset is developed in this chapter. Although the adjustment is not exact it reduces the risk of any misleading inference from an analysis using a fixed effect model. This is particularly useful for analysing datasets for which a proper analysis using GLMM would require dealing with very large number of random effects and may be computationally difficult to handle. It is also useful for analysing small datasets where the data analyst is not familiar with GLMM theory and is not able to undertake proper analysis involving random effects. It is shown that most of the elements required for adjustment can be obtained from a fixed effect model. Many commonly used softwares produce most of these quantities as a by-product of the conventional analysis. The quantities which are not available from the fixed effect modelling are some cluster level totals and the random effects dispersion parameter ϕ. It is shown that for one and two random terms in the model these cluster level quantities can be obtained by simple calculations. This is particularly true when there is only one random term in the model. An approximate value of the dispersion parameter, ϕ, can be obtained from the literature for a

previous analysis of a similar dataset. In those situations where the adjustment will be made because of the large number of random effects to be estimated under a GLMM, an estimate of ϕ can be obtained by using a smaller subset of the dataset. Any approximate value of ϕ would help inflate the estimates of SEs and would lead to a more conservative inference. An exact expression for the adjustment is derived for simple linear regression with only one random term. Similar exact expressions for other cases may not be easy to derive but this is an area which can be investigated further.

CHAPTER TEN

OVERALL DISCUSSION

10.1 OVERVIEW

Some theoretical and application aspects of GLMMs have been researched. The thesis has contributed to the ongoing development of the estimation in GLMMs by investigating the properties of the estimators and developing further arguments for the REML method of estimation. The investigation leads to new expressions for estimating variance of the dispersion parameters ϕ and γ and also a new estimator for γ itself. The potential of the recent development of hierarchical generalised likelihood models (HGLMs) has also been investigated by comparing its performance with GLMM for a Poisson response variable.

The current use of GLMMs is still very limited even though it has enormous potential for analysing categorical data. It has been demonstrated how GLMM can be used in improving efficiency of commonly used statistical procedures. For instance, the conventional methods for analysis of contingency tables are found to be less appropriate in the situations where observations are clustered. Two GLMM based approaches have been proposed and shown to be more efficient than other existing methods when observations are clustered. One of the approaches is developed under Poisson sampling and the other under product multinomial sampling assumptions. By utilising mixed threshold modelling, another method of analysis is proposed for contingency tables with ordinal categories.

A general strategy for dealing with a multinomial response category is also discussed and compared with other available methods. A method for dealing with random exposure effects in matched case-control studies is developed and applied to some real datasets.

In some situations a full GLMM based analysis may not be possible particularly in the absence of appropriate software. A method for undertaking approximate adjustments in the estimated standard errors of the parameters in a fixed effect analysis is also proposed. In the presence of clustering the approximate adjustment will serve as a safeguard from misleading inference.

10.2 PROBLEM AREAS

10.2.1 Estimation of Variance Components

Although estimation in GLMMs has been developed considerably over the recent years, still there is a need for improvement. The ML and REML estimators of the variance component ϕ tend to show large bias in certain situations. In general the ML estimator is negatively biased and it tends to increase rapidly when there are multiple random components in the model. Even though the REML estimator is asymptotically unbiased and in most cases performs better than the ML estimator it also appears to be significantly biased when ϕ is large and also when there are multiple random components. In the applications discussed, the parameters ρ and γ were not required to be dealt with. In other applications where these parameters were also required to be estimated, the performance of the estimators deteriorates further. Also in many cases convergence of the iteration becomes a difficulty as there is no

easily detectable maximum. In some other cases, depending on the starting values, the converged estimates turn out to be considerably different often because of the flatness of the surface corresponding to l_1+l_2.

10.2.2 Asymptotic Theory

The asymptotic properties of the GLMMs estimators have not been developed sufficiently yet. A crucial condition for the asymptotic property to hold is that the number of random effects is fixed but in practice the number of random effects tends to increase with the increase in the number of observations. Although the maximum likelihood principle is used in the estimation, the corresponding likelihood based test statistics are still not available. Lee and Nelder (1996) have proposed test statistics for testing the significance of fixed parameters and the variance components. The likelihood ratio statistic for the fixed parameters appears to work well but there is a problem with the testing of variance components. The two test statistics proposed by Lee and Nelder for testing the parameter ϕ are found to be unsatisfactory in simulation studies. The Wald test based on the information matrix seems to work to some extent when there is only one random component in the model but when the number of random components is two or more the test result always appears to be insignificant.

10.2.3 Model Diagnostics

The absence of enough diagnostic tools for checking the goodness of fit of a GLMM is another drawback. Lee and Nelder (1996) introduced a scaled deviance measure as a goodness-of-fit criterion. However, as the measure uses the distribution $\mathbf{y}|\mathbf{u}$ only it can not be used for checking the random

components. They also introduced a graphical technique called half-normal plots which graphs normal order statistics against ordered residuals. Apart from these there are not many other tools available for checking the goodness-of-fit. In the previous chapters deviance is calculated in some cases assuming **u** is fixed which is similar to the Lee and Nelder approach and can only be used for fixed parameters.

10.2.4 Software

The other issue which is preventing the widespread use of GLMMs in applied research is the absence of appropriate software. What is required is a user friendly software to undertake GLMM based analysis so that applied researchers can benefit from the theoretical development without going into the theoretical detail. The software should also be able to deal with very large numbers of random effects, which is very common in practice. Analytical methods also need to be developed to achieve further efficiency in computing. The straight forward Newton-Raphson estimation algorithm is more likely to break down when the number of random effects becomes large. The efficiency in computing may be achieved better by concentrating on specific applications. For example, different procedures within a software can be developed for different applications such as for contingency tables, survival analysis, threshold modelling, matched analysis and similar other analyses. Diagnostics and hypothesis testing facilities should also be included.

10.3 POTENTIAL RESEARCH PROBLEMS

Obviously there are needs for research in all of the above mentioned issues. However, some specific research issues which have been identified while working on the thesis will be discussed in this section.

10.3.1 Distribution of GLMM Estimators

As mentioned above there is obvious room for further improvement in GLMM estimation. One of the possibilities is to further investigate the distributional properties of the GLMM estimators. An attempt has been made in Chapter 3 to identify some characteristics of GLMM estimators. More investigations along this line are required. A better understanding of the distribution, particularly the second order moments of the estimators, would contribute to more accurate variance-covariance matrix which could increase the efficiency of the estimation methods.

10.3.2 Information Matrix

In computing the variance-covariance matrix V, different convergent methods use one of four possible ways of computing the information matrix, $I = -\partial^2 l_1/\partial\eta\partial\eta'$, or its expectations. These are:

(i) $I_{y,u} = -E_{y,u}(\partial^2 l_1/\partial\eta\partial\eta')$ which does not depend on y, u,

(ii) $I_y|_u = -E_y|_u(\partial^2 l_1/\partial\eta\partial\eta')$ which depends on u but not y,

(iii) $I = -\partial^2 l_1/\partial\eta\partial\eta' = $ sample information matrix depending on y, u,

(iv) $I^* = (I + I_y|_u)/2$ used by AIREML methods (Gilmour et al. 1995).

There needs to be further research on the comparative performances of the above options for both the influence on the estimates and the convergence properties of the algorithms.

For Cases (i) and (ii) suppose that $\mathbf{V}$ is formed from $I_{y,u}$ or $I_y|_u$ so that it is not a stochastic matrix when considered over the distribution of y conditional on fixed $\mathbf{u}$.

Using the notation specified in Chapters 2 and 3, since $E_y|_u(\mathbf{t}) = \mathbf{0}$ we get $E_y|_u(\beta) = \beta - \mathbf{T}_2\mathbf{A}^{-1}\mathbf{u}$ and $E_y|_u(\tilde{\mathbf{u}}) = \mathbf{u} - \mathbf{T}\mathbf{A}^{-1}\mathbf{u} = \mathbf{z}$. Now if

$$\mathbf{V}^{-1}\begin{bmatrix}\mathbf{X}'\\\mathbf{Z}'\end{bmatrix} I_y|_u \begin{bmatrix}\mathbf{X} & \mathbf{Z}\end{bmatrix}\mathbf{V}^{-1} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\begin{bmatrix}0 & 0\\0 & \mathbf{A}^{-1}\end{bmatrix}\mathbf{V}^{-1} = \begin{bmatrix}\mathbf{T}_1 - \mathbf{T}_2\mathbf{A}^{-1}\mathbf{T}_2' & \mathbf{T}_2 - \mathbf{T}_2\mathbf{A}^{-1}\mathbf{T}\\\mathbf{T}_2 - \mathbf{T}\mathbf{A}^{-1}\mathbf{T}_2' & \mathbf{T} - \mathbf{T}\mathbf{A}^{-1}\mathbf{T}\end{bmatrix}$$

then $\text{Var}_y|_u(\beta) = \mathbf{T}_1 - \mathbf{T}_2\mathbf{A}^{-1}\mathbf{T}_2'$, $\text{Var}_y|_u(\tilde{\mathbf{u}}) = \mathbf{T} - \mathbf{T}\mathbf{A}^{-1}\mathbf{T}$.

For the Case (ii) an attempt was made to approximate the elements of $\mathbf{V}^{-1}$ as follows:

If $I_y|_u$ is used to compute $\mathbf{T}$ matrices then these matrices are functions of $\mathbf{u}$ and, if U is a general component of $\mathbf{u}$,

$$\mathbf{T} \simeq \mathbf{T}_0 + \sum_U U\mathbf{T}_0^d , \quad \mathbf{T}_0 = \mathbf{T}_{(u=0)} , \quad \mathbf{T}_0^d = (d\mathbf{T}/dU)_{(u=0)} ,$$

$$\mathbf{T}_1 \simeq \mathbf{T}_{10} + \sum_U U\mathbf{T}_{10}^d , \quad \mathbf{T}_{10} = \mathbf{T}_{1(u=0)} , \quad \mathbf{T}_{10}^d = (d\mathbf{T}_1/dU)_{(u=0)} ,$$

$$\mathbf{T}_2 \simeq \mathbf{T}_{20} + \sum_U U\mathbf{T}_{20}^d , \quad \mathbf{T}_{20} = \mathbf{T}_{2(u=0)} , \quad \mathbf{T}_{20}^d = (d\mathbf{T}_2/dU)_{(u=0)} .$$

and

$$E_{y,u}(\beta) = E_u(\beta - \mathbf{T}_2\mathbf{A}^{-1}\mathbf{u}) = \beta - \sum_U \mathbf{T}_{20}^d\mathbf{A}^{-1}\text{Cov}(U,\mathbf{u}) = \beta - \mathbf{b}_2$$

where $\mathbf{b}_2 = \sum_U$ (column of $\mathbf{T}_{20}^d$ corresponding to U)

$$E_{y,u}(\tilde{\mathbf{u}}) = E_u(\mathbf{u} - \mathbf{T}\mathbf{A}^{-1}\mathbf{u}) = -\sum_U \mathbf{T}_0^d\mathbf{A}^{-1}\text{Cov}(U,\mathbf{u}) = -\mathbf{b}$$

where $\mathbf{b} = -\sum\limits_{U} (\text{column of } \mathbf{T}_0^d \text{ corresponding to } U)$

$\text{Var}_{y,u}(\beta) = E_u(\mathbf{T}_1 - \mathbf{T}_2 \mathbf{A}^{-1} \mathbf{T}_2') = \mathbf{T}_{10} - \mathbf{T}_{20} \mathbf{A}^{-1} \mathbf{T}_{20}' - \sum\limits_{U} \sum\limits_{U*} \mathbf{T}_{20}^d \mathbf{A}^{-1} \mathbf{T}_{20}^{d*'} \text{Cov}(U, U*)$

$\text{Var}_{y,u}(\tilde{\mathbf{u}}) = E_u(\mathbf{T} - \mathbf{T}\mathbf{A}^{-1}\mathbf{T}') = \mathbf{T}_0 - \mathbf{T}_0 \mathbf{A}^{-1} \mathbf{T}_0' - \sum\limits_{U} \sum\limits_{U*} \mathbf{T}_0^d \mathbf{A}^{-1} \mathbf{T}_0^{d*} \text{Cov}(U, U*)$

$\qquad = \mathbf{B}_0 - \sum\limits_{U} \sum\limits_{U*} \mathbf{T}_0^d \mathbf{A}^{-1} \mathbf{T}_0^{d*} \text{Cov}(U, U*)$, where $\mathbf{B}_0 = \mathbf{T}_0 - \mathbf{T}_0 \mathbf{A}^{-1} \mathbf{T}_0'$,

and * corresponds to the derivative with respect to U*. Using

$\mathbf{z} = \mathbf{u} - \mathbf{T}\mathbf{A}^{-1}\mathbf{u} = \mathbf{u} - \mathbf{T}_0 \mathbf{A}^{-1}\mathbf{u} - \sum\limits_{U} U\mathbf{T}_0^d \mathbf{A}^{-1}\mathbf{u}$

we have

$E_u(\mathbf{z}) = -\sum\limits_{U} (\text{column of } \mathbf{T}_0^d \text{ corresponding to } U) = -\mathbf{b}$

and

$E_u(\mathbf{z}\mathbf{z}') = \mathbf{A} - \mathbf{T}_0 - \mathbf{B}_0 + \sum\limits_{U,U*} E_u[UU*\mathbf{T}_0^d \mathbf{A}^{-1}\mathbf{u}\mathbf{u}'\mathbf{A}^{-1}\mathbf{T}_0^{d*}]$

so that

$E_{y,u}(\tilde{\mathbf{u}}_i' \mathbf{A}_i^{-1} \tilde{\mathbf{u}}_i) = \text{tr}(\mathbf{A}_i^{-1}\mathbf{z}_i\mathbf{z}_i') + \text{tr}[\mathbf{A}_i^{-1} E_u(\mathbf{B}_{ii})]$

$\qquad = \theta_i v_i - \text{tr}(\mathbf{A}_i^{-1}\mathbf{T}_{0,ii}) + \text{tr}[\mathbf{A}_i^{-1} \times (i,i) \text{ block of } \mathbf{Q}]$

where $\mathbf{Q} = \sum\limits_{U,U*} \{E_u[UU*\mathbf{T}_0^d \mathbf{A}^{-1}\mathbf{u}\mathbf{u}'\mathbf{A}^{-1}\mathbf{T}_0^{d*}] - \sum\limits_{U}\sum\limits_{U*} \mathbf{T}_0^d \mathbf{A}^{-1}\mathbf{T}_0^{d*}\text{Cov}(U,U*)\}$.

The above strategy of investigation could not be completed and evaluated fully and remains to be followed up further.

10.3.3 Likelihood Ratio Tests

The argument used for developing GLMM estimation strategy in McGilchrist (1994) is that

$\mathbf{y}^* = \mathbf{X}\tilde{\beta} + \mathbf{Z}\tilde{\mathbf{u}} \sim N(\mathbf{X}\beta + \mathbf{Z}\mathbf{u}, \mathbf{B}^{-1})$ approximately

$\Rightarrow \mathbf{y} = \mathbf{X}\tilde{\beta} + \mathbf{Z}\tilde{\mathbf{u}} + \mathbf{e} \sim N(\mathbf{X}\beta, \mathbf{B}^{-1} + \sigma^2 \mathbf{Z}\mathbf{A}\mathbf{Z}')$.

Using this distributional assumption a log-likelihood function can be derived as follows

$l = -(1/2)\{n \ln 2\pi + \ln |\mathbf{B}^{-1} + \sigma^2 \mathbf{Z}\mathbf{A}\mathbf{Z}'| + (\mathbf{y}^* - \mathbf{X}\beta)'(\mathbf{B}^{-1} + \sigma^2 \mathbf{Z}\mathbf{A}\mathbf{Z}')^{-1}(\mathbf{y}^* - \mathbf{X}\beta)\}$

and a likelihood ratio test may be constructed to test $H_0:\phi = 0$ or for some components of β. Some investigations were undertaken on this but could not be resolved completely. Further research can be undertaken along this line to examine whether test statistics can be developed.

10.3.4 Comparison between HGLMs and GLMMs

In Chapter 4 a simulation-based comparison between HGLMs and GLMMs is presented in the case of a Poisson distributed response variable. Further investigation can be undertaken along this line by using response variables with various distributions such as binomial, gamma and inverse-Gaussian. More effective diagnostic tools should be developed to identify the situations where HGLMs may be preferable over GLMMs.

10.3.5 Exact Expressions

Various applications discussed in the previous chapters are dependent on fitting GLMMs and estimating parameters through iteration. This is not always required in conventional analysis with simple situations such as 2×2 contingency tables or case-control studies where exact expression is available for testing independence. That facilitates the use of the techniques by researchers from other disciplines. There may be a possibility for a similar development of such exact or even approximate expressions for corresponding mixed model analysis.

10.3.6 Sampling Weights

In practice correlation or clustering in a dataset is often created due to the sample design used in data collection and in many cases unequal selection probabilities are used. The analysis discussed in the previous

chapters are mainly under the assumption of equal weights to all observations. Specific results can further be derived for the datasets where weights of the observations vary from cluster to cluster or even over observations. The implications for the corresponding finite population inference should also be investigated.

REFERENCES

Agresti, A. (1984). *Analysis of Ordinal Categorical Data.* John Wiley & Sons.

Airy, G.B. (1861). *On the algebraical and numerical theory of errors of observations and the combinations of observations.* MicMillan and Co., Cambridge and London.

Altham, P.M.E. (1976). Discrete variable analysis for individuals grouped into families. *Biometrika* **63**, **2**, 263-9.

Anderson, T.W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Stat.* **1**, 135-141.

Anderson, D.A. and Aitkin, M.A. (1985). Variance components models with binary response: Interviewer variability. *J. Royal Stat. Soc. B* **47**, 203-210.

Anderson, R.L. and Bancroft, P.P. (1952). *Statistical Theory in Rsearch.* McGraw-Hill, New York.

Andrich, D. (1979). A model for contingency tables having an ordered response classifications. *Biometrics*, **35**, 403-15.

Barndorff-Neilson, O.E. and Cox, D.R. (1994). *Inference and Asymptotics.* Chapman and Hall, London.

Besag, J., York, J. and Mollié, A. (1991). Bayesian Image Restoration, with two appilcations in spatial statistics, *Annals of Inst. of Stat. Mathematics* **43**, 1-59.

Birch, M.W. (1963). Maximum likelihood in three-way contingency tables. *J. Royal Stat. Soc. B* **25**, 220-233.

Bishop, Y.M.M. Fienberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis.* Cambridge: MIT Press.

Breslow, N.E. (1982). Covariance Adjustment of Relative-Risk Estimates in Matched Studies, *Biometrics* **38**, 661-672.

Breslow, N.E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics* **33**, 38-44.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalised linear mixed models. *J. of Amer. Stat. Assoc.* **88**, 9-25.

Breslow, N.E. and Day, N.E. (1980). *Statistical Methods in Cancer Research*. Vol. 1. Lyon: International Agency for Research on Cancer.

Breslow, N.E. and Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* **82**, 81-92.

Brier, S.S. (1980). Analysis of contingency tables under cluster sampling. *Biometrika* **67**, **3**, 591-596.

Chauvenet, W. (1863). *A manual of spherical and practical astronomy, Vol. 11: Theory and use of astronomical instruments*. J.B. Lippincott and Co., Philadelphia.

Chowdhury, S.R and McGilchrist, C.A (1997). Hierarchical generalised linear models: a comparison between Poisson-normal and Poisson-gamma models. *J. of Stat. Comp. and Simul.*, Submitted.

Cohen, J.E. (1976). The distribution of the Chi-squared statistic under cluster sampling from contingency tables. *J. of Amer. Stat. Assoc.* **71**, 665-669.

Corbeil, R.R. and Searle, S.R. (1976). Restricted maximum likelihood (REML). estimation of variance components in the mixed model. *Technometrics* **18**, 31-38.

Cox, D.R. (1970). *The Analysis of Binary Data*, London, Methuen.

Cox, D.R. (1972). Regression models and life tables (with discussion). *J. Royal Stat. Soc. B* **34**, 187-220.

Cox, D.R. (1975). Partial likelihood. *Biometrika* **62**, 69-76.

Crowder, M.J. (1978). Beta binomial anova for proportions. *Applied Statistics* **27**, No 1, pp 34-37.

Crowder, M.J. (1985). Gaussian estimation for correlated binary data. *J. Royal Stat. Soc. B* **47**, 229-237.

Crump, S.L. (1946). The estimation of variance components in analysis of variance. *Biometrics Bull.* **2**, 7-11.

Daniels, H.E. (1939). The estimation of components of variance. *J. Royal Stat. Soc. Suppl.* **6**, 186-197.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *J. Royal Stat. Soc. B* **39**, 1-22.

Duncan, O.D. and McRae, J.A., Jr (1979). Multiway contingency analysis with a scaled response or factor. In *Sociological Methodology* (ed. Schuessler, K.F.), 68-85. San Franscisco, Jossey-Bass.

Engel, B. and Keen, A. (1992). A simple approach for the analysis of generalised linear mixed models. *Stat. Neerlandica* **48**, 269-270.

Engel, B. and Keen, B. (1996). An introduction to generalised linear mixed models. IBC96 Invited Papers, 125-135.

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics* **3**, 1-21.

Fellegi, I.P. (1980). Approximate test of independence and goodness of fit based on stratified multistage samples. *J. of Amer. Stat. Assoc.* **75**, 261-268.

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

Fellner, W.H. (1987). Sparcse matrices, and the estimation of variance components. *Communs Stat. Simul* **16**, 439-463.

Fienberg, S (1980). *The Analysis of Cross-classified Categorical Data.* 2nd ed. Cambridge: MIT Press.

Fingleton, B (1984). *Models of Category Counts.* Cambridge, University Press.

Fisher, R.A. (1918). The correlation between relatives on the supposition of Medelian inheritance. *Trans. Royal Soc.* Edinburgh **52**. 399-433.

Fisher, R.A. (1925). *Statistical Methods for Research Workers.* 1st edn. Oliver & Boyd. Edinburgh and London.

Freeman, D.H. (1987). *Applied Categorical Data Analysis.* Marcel Dekker, New York.

Gianola, D. and Foulley, J.L. (1982). Non-linear prediction of latent genetic liability with binary expression: an empirical Bayes approach, in *2nd world congress on genetics applied to livestock production* **7**, 293-303. Madrid: Ministerio de Agricultura, Pesca y Alimentacion.

Gianola, D. and Foulley, J.L. (1983). Sire evaluation for ordered categorical data with a threshold model. *Génétique, Sélection, Évolution* **15**, 201-224.

Gilbert, G.N. (1981). *Modelling Society.* London: Allen & Unwin.

Gilmour, A.R., Anderson, R.D. and Rae, A.L. (1985). The analysis of binomial data by generalised linear mixed model. *Biometrika* **72**, 593-599.

Gilmour, A.R., Thompson, R., Cullis, B.R. (1995). Average Information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51**, 1440-1450.

Gokhale, D.V. and Kullback, S. (1978). *The Information in Contingency Tables*. New York: Marcell-Decker.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika* **73**, 43-56.

Goldstein, H. (1988). Restricted unbiased iterative generalised least-squares estimation. *Biometrika* **76**, 622-623.

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

Goodman, L.A. (1978). *Analysing Qualitative/Categorical Data: Loglinear Analysis and Latent Structure Data*. Cambridge, Mass.

Goodman, L.A. (1979). Simple models for the analysis association in cross-classifications having ordered categories. *J. Amer. Stat. Assn.* **74**, 537-52.

Green, P.J. (1987). Penalized likelihood for general semi-parametric regression models. *Int. Stat. Rev.* **55**, 245-259.

Grizzle, J.E., Starmer, C.F. and Koch, G.G (1969). Analysis of categorical data by linear models. *Biometrics*, **25**, 489-504.

Haberman, S.J. (1974a). *The analysis of frequency data*. Chicago. University of Chicago Press.

Haberman, S.J. (1974b). Loglinear Models for Frequency tables with ordered classifications. *Biometrics* **30**, 589-600.

Haberman, S.J. (1978). *Analysis of Qualitative Data. Vol. 1: Introductory Topic*. New York: Academic.

162

Haberman, S.J. (1979). *Analysis of Qualitative Data. Vol. 2: New Developments*. New York: Academic.

Hartley, H.O. and Rao, J.N.K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, **54**, 93-108.

Hartley, H.O. and Vaughn, W.K. (1972). A computer program for mixed model analysis of variance model based on maximum likelihood. *Statistical Pepers in honour of G. W. Snedecor*, ed. T.A. Bancroft, Ames, Iowa: Iwoa State University Press.

Hartsfield, Jr, J.K., Lee, M., Morel, J.G., and Hilbelink, D.R. (1992). Statistical Analysis of the effect of cadmium and zinc on hamster teratogenesis. *Biochem. Med. Metab. Biol.*, **48**, 159-173.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Stat. Assn.* **72**, 320-340.

Harville, D.A. and Mee, R.W. (1984). A mixed model procedure for analysing ordered categorical data. *Biometrics* **40**, 933-944.

Hemmerle, W.J. and Hartley, H.O. (1973). Computing maximum likelihood estimates for the mixed A.O.V. model using the W transform. *Technometrics* **15**, 819-31.

Henderson, C.R. (1950). Estimation of genetic parameters (Abstract). *Ann. Math. stat.*, **21**, 309-310.

Henderson, C.R. (1953). Estimation of variance and covariance components. *Biometrics* **9**, 226-252.

Henderson, C.R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding*, W.D. Hanson and H.F. Robinson eds. Publication **982**, 141-163, Washington: Nat. Acad. of Sc., Nat. Research Council.

Henderson, C.R. (1973). Sire evaluation and genetic trends. In *Proceedings of the Animal Breeding and Genetics Symposium in honour of Dr Jay Lush,* 10-41, Champaign, Illinois: Amer. Soc. of Animal Science.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423-447.

Herbach, L.H. (1959). Properties of Model II type analysis of variance tests. A: Optimum nature of the F-test for Model II in the balanced case. Ann. Math. Stat. **30**, 939-959.

Hill, B.M. (1965). Inference about variance components in the one-way model, *J. Amer. Stat. Assn.* **60**. 806-825.

Hill, B.M. (1967). Correlated errors in the random model. *J. Amer. Stat. Assn.* **62**, 1387-1400.

Hosmer, D.W. and Lemeshow, S. (1989). *Applied Logistic Regression.* New York: Wiley.

Im, S. and Gianola, D. (1988). Mixed models for binomial data with an application to lamb mortality. *Applied Stat.* **37**, 196-204.

Jennrich, R.I. and Sampson, P.F. (1976). Newton-Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics* **18**, 11-17.

Khuri, A.I. and Sahai, H. (1985). Variance components analysis: a selective literature survey. *Int. Stat. Rev.* **53**, 279-300.

Koehler, K.J. and Wilson, J.R. (1986). Chi-square tests for comparing vectors of proportions for several cluster samples. *Commun. Stat. - Theory Meth.* **15**, 2977-2990.

Kupper, L.L. and Haseman, J.K. (1978). The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics* **34**, 69-76.

Laird, N.M. (1978). Empirical Bayes methods for two-way contingency tables. *Biometrika*, **65**, 581-590.

Laird, N.M. and Louis, T.A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *J. Amer. Stat. Assn.* **82**, 739-757.

Lee, Y. and Nelder, J.A. (1996). Hierarchical generalised linear models. *J. Royal Stat. Soc. B* **58**, 619-678.

Leonard, T. (1972). Bayesian methods for binomial data, *Biometrika* **59**, 581-589.

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

Mack, T.M., Pike, M.C., Henderson, B.E., Pfeffer, R.I., Gerkins, V.R., Arthur, B.S., and Brown, S.E. (1976). Estrogens and endometrial cancer in a retirement community. *New Engl. J. Med.*, 294, 1262-1267.

McCullagh, P. (1980). Regression Models for Ordinal Data. *J. Royal Stat. Soc.* **B**, **42**, 109-142.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. 2nd edition London: Chapman and Hall.

McCulloch, C.E. (1994). Maximum likelihood variance components estimation for Binary data. *J. Amer. Stat. Asson.* **89**, 330-335.

McCulloch, C.E. (1997). Maximum likelihood algorithms for generalised linear mixed models. *J. Amer. Stat. Asson.* **92**, 162-170.

McGilchrist, C.A. (1993). REML estimation for survival models with frailty. *Biometrics* **49**, 221-225.

McGilchrist, C.A. (1994). Estimation in generalised mixed models. *J. Royal Stat. Soc. B* **56**, 61-69.

McGilchrist, C.A. and Aisbett, C.W. (1991a). Restricted BLUP for mixed linear models. *Biom. J.* **33**, 131-141.

McGilchrist, C.A. and Aisbett, C.W. (1991b). Regression with frailty in survival analysis. *Biometrics* **47**, 461-466.

McGilchrist, C.A. and Ragunathan, T.E. (1995). Generalised components of variance models. *Commun. Stat. - Theory Meth.* **24**(2), 349-361.

McGilchrist C.A. and Chowdhury S.R. (1997). Second order moment properties for estimators in generalised linear mixed models. *J. Royal Stat. Soc. B.* Submitted 1997.

McGilchrist, C.A. and Yau, K.K.W. (1995). The derivation of BLUP, ML, REML estimation methods for generalised linear mixed models. *Commun. Stat. - Theory Meth.* **24**(12) 2963-2980.

McGilchrist, C.A. and Yau, K.K.W. (1996). Survival analysis with time dependent frailty using a longitudinal model. *Aust. J. of Stat.*, **38**(1), 53-60.

McGilchrist, C.A. and Zhaorong, J (1990). Multicentre clinical trials and variance components. *Biom. J.* **32**, 545-550.

Miettinen, O.S. (1969). Individual matching with multiple controls in the case of all-or-none responses. *Biometrics* **25**, 339-355.

Miettinen, O.S. (1970). Estimation of relative risk from individually matched series. *Biometrics*, **26**, 75-86.

Miller, J.J. (1973). *Asymptotic properties and computation of maximum likelihood estimates in the mixed model of the analysis of variance.* Technical Report No. 12, Department of Statistics, Stanford University, Stanford, California.

Miller, J.J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Stat.*, **5**, 746-762.

Miller, M.E. and Landis, J.R. (1991). Generalised variance component models for clustered categorical response variables. *Biometrics*, **47**, 33-44.

Morel, J.G. and Koehler, K.J. (1995). A one-step Gauss-Newton Estimator for Modelling Categorical Data wi th Extraneous Variation. *Appl. Stat.* **44**, No 2. pp 187-200.

Morton, R. (1988). Analysis of generalised linear models with nested strata of variation. *Aust. J. of Stat.* **30A**, 215-224.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalised linear models. *J. Royal Stat. Soc.A* **135**, 370-384.

Patterson, H.D. and Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **58**, 545-554.

Patterson, H.D. and Thompson, R. (1974). Maximum likelihood estimation of componenets of variance. *Proc. Eighth Internat. Biom. Conf.*, 197-209.

Pearson, K. (1904). Mathematical contributions to the theory of evolution, XIII. On the theory of contingency and its relation to association and normal correlation. *Draper's company research Memoirs, Biometric Series 1.*

Pregibon, D. (1981). Logistic regression diagnostics. *Ann Statist* **9**, 705-724.

Ragunathan, T.E. (1994). Monte Carlo methods for exploring sensitivity to distributional assumptions in a Bayesian analysis of a series of 2x2 tables. *Statistics in Medicine*, **13**, 1525-1538.

Rao, C.R. (1952). *Advanced Statistical Methods in Biometric Research.* John Wiley & Sons, Inc.

Rao, C.R. (1970). Estimation of heteroscedastic variances in linear models. *J. Amer. Stat. Asson.* **65**, 161-172.

Rao, C.R. (1971a). Estimation of variance and covariance components - MINQUE theory. *J. Mult. Anal.* **1**, 257-275.

Rao, C.R. (1971b). Minimum variance quadratic unbiased estimation of variance components. *J. Mult. Anal.* **1**, 445-456.

Rao, C.R. (1972). Estimation of variance and covariance components in linear models. *J. Amer. Stat. Asson.* **67**, 112-115.

Rao, C.R. (1979). MINQUE theory and its relation to ML and MML estimation of variance components. *Sankhya B* **41**, 138-153.

Rao, C.R. and Kleffe, J. (1988). *Estimation of variance components and applications*. North-Holland, Amsterdam.

Rao, J.N.K. and Scott, A.J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *J. Amer. Stat. Asson.* **76**, 221-230.

Robinson, G.K. (1991). The estimation of random effects. *Stat. Science* **6**, 15-51.

Russell, T.S. and Bradley, R.A. (1958). One-way variances in the two-way classification. *Biometrika* **45**, 111-129.

Saei, A (1996). *Random Component Threshold Models for Ordered Categorical Data*. Ph.D Thesis, The Australian National University.

Saei, A. and McGilchrist, C.A. (1996). Random component threshold models. *J. Ag. Bio. and Env. Stat.* **1**, 288-296.

Saei, A., Ward, J. and McGilchrist, C.A. (1996). Threshold models in methadone program evaluation. *Statistics in Medicine* **15**, 2253-2260.

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

Searle, S.R. (1968). Another look at Henderson's methods of estimating variance components, *Biometrics* **24**, 749-788.

Searle, S.R. (1971). *Linear Models*. Wiley, New York.

Searle, S.R. (1989). Obituary: Charles Roy Henderson 1911-1989. *Biometrics* **45**, 1333-1335.

Searle, S.R., Casella, G. and McCulloch, C.E (1992). *Variance Components*. Wiley, New York.

Solomon, P.J. and Cox, D.R. (1992). Nonlinear components of variance models. *Biometrika* **79**, 1-11.

Speed, T. (1991). Comments on Robinson: Estimation of Random Effects. *Stat. Science* **6**, 42-44.

Srole, L., Langner, Michael, S.T., Kirkpatrick, P., Opler, M.K. and Rennie, T.A.C (1978). *Mental Health in Metropolis: Midtown Manhattan Study*. Rev. ed. New York. NYU Press.

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

Thompson, R. (1980). Maximum likelihood estimation of variance components. *Math. Operforsch. Stat. Ser. Stat.* **11**, 545-561.

Thompson, W.A. Jr (1962). The problem of negative estimates of variance components. *Ann. of Math. Stat.* **33**, 273-289.

Tippett, L.H.C. (1931). *The Methods of Statistics*. 1st edn. Williams and Norgate, London.

Tsutakawa, R.K. (1988). Mixed model for analyzing geographic variability in mortality rates. *J. Amer. Stat. Asson.* **83**, 37-42.

Walter, S.D. (1980). Matched Case-Control Studies With a Variable Number of Controls per Case. *Applied Statistics* 29, No. 2, pp 172-17

Wilks, S.S. (1935). The likelihood test of independence in contingency tables. *Ann. of Math. Stat.*, **6**, 190-196

Williams, D.A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics* **35**, 339-353.

Williams, D.A. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics* **11**, 545-561.

Winsor, C.P. and Clarke, G.L. (1940). Statistical study of variation in the catch of plankton nets. *Sears Foundation J. Marine Res.* **3**, 1-84.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika* **80**, 791-796.

Wolfinger, R. and O'Connell, M. (1993). Generalised linear mixed models: a pseudolikelihood approach. *J. Stat. Comput. Simul.* **48**, 233-244.

Yau, K.K.W. and McGilchrist, C.A. (1998). ML and REML estimation in survival analysis with time dependent correlated frailty. *Statistics in Medicine (to appear)*.

Zeger, S.L. and Karim, M.R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *J. Amer. Stat. Asson.* **86**, 79-86.

Zeger, S.L. and Liang, K.Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121-130.

Zeger, S.L., Liang, K.Y. and Albert, P.S. (1988). Models for longitudinal data: a generalised estimating equation approach. *Biometrics* **44**, 1049-1060.

Zhaorong, J., Matawie, K.M. and McGilchrist, C.A. (1992). Variance components for doscordances. *Math. Biosciences* **110**, 119-124.

Zhaorong, J., McGilchrist, C.A. and Jorgensen, M.A. (1992). Mixed model discrete regression. *Biom. J.* **34**, 691-700.

APPENDIX A

DATASETS

A number of datasets used in various chapters of the thesis. Some of these which are smaller in size are already included within the respective chapters while the bigger size ones are presented in this Appendix. The dataset in Table A1 is used in Chapter 6 and the datasets in Tables A2 and A3 are used in Chapter 5.

Table A1: Responses on Housing Satisfaction of Families in 20 Neighbourhood Clusters in Montevideo, Minnesota.

C_1P_1	C_1P_2	C_1P_3	C_2P_1	C_2P_2	C_2P_3	C_3P_1	C_3P_2	C_3P_3
1	0	0	2	2	0	0	0	0
1	0	0	2	2	0	0	0	0
0	2	0	0	2	0	0	1	0
0	1	0	2	1	0	1	0	0
0	0	0	0	4	0	0	1	0
1	0	0	3	1	0	0	0	0
3	0	0	0	1	0	0	1	0
1	0	0	1	3	0	0	0	0
3	0	0	0	0	0	1	0	1
0	1	0	0	3	1	0	0	0
1	1	0	0	2	0	1	0	0
0	1	0	4	0	0	0	0	0
0	0	0	4	1	0	0	0	0
0	0	0	1	2	0	0	0	2
2	0	0	2	1	0	0	0	0
1	0	0	1	1	0	0	0	0
0	0	0	1	1	1	0	2	0
0	0	0	1	0	1	0	0	1
2	0	0	2	1	0	0	0	0
2	0	0	2	0	0	1	0	0

P = Personal Satisfaction, C = Community Satisfaction
and the subscripts 1 = Unsatisfied, 2 = Satisfied, 3 = Very satisfied.

Source: Brier (1980)

Table A2: Matched Case Control Data from Low Birth Weight Study

Set	Case	Control 1	Control 2	Set	Case	Control 1	Control 2
1	0000	0000	1000	2	1110	0001	1000
3	0000	0000	0000	4	0000	1001	1101
5	1010	1101	1001	6	1111	1000	0000
7	0001	0000	1000	8	1100	0100	0001
9	1000	0000	1000	10	1000	0001	0000
11	1110	0110	1000	12	0100	1100	0000
13	0011	0000	1100	14	1110	1001	1000
15	1000	0000	1010	16	0101	0000	0000
17	1010	0000	0000	18	0000	0000	0000
19	0000	1011	0000	20	0101	1000	0000
21	1001	0100	0000	22	0011	0000	1111
23	1000	1000	1000	24	1110	0000	1000
25	1001	1000	0000	26	0100	0000	0000
27	1010	0111	0000	28	1011	0101	1000
29	1001	0000	0000				

The binary string for each observation is the values of variables in the order SMOKE, PTD, UI and LWD

Source: Hosmer and Lemeshow (1989)

APPENDIX A

TABLE A3: MATCHED DATA FROM THE LOS ANGELES STUDY OF ENDOMETRIAL CANCER USED FOR ILLUSTRATION IN CHAPTER 5

CASE OR CONTROL	AGE	GALL BLADDER DISEASE	HYPER TENSION	OBESITY	ESTROGEN USE	CONJUGATED DOSE	ESTROGEN DURATION (MONTHS)	NON ESTRGN DRUG
CASE	74	NO	NO	YES	YES	3	96+	YES
CONTROL	75	NO	NO	UNK	NO	0	0	NO
CONTROL	74	NO	NO	UNK	NO	0	0	NO
CONTROL	74	NO	NO	UNK	NO	0	0	NO
CONTROL	75	NO	NO	YES	YES	1	48	YES
CASE	67	NO	NO	NO	YES	3	96+	YES
CONTROL	67	NO	NO	NO	YES	3	5	NO
CONTROL	67	NO	YES	YES	NO	0	0	YES
CONTROL	67	NO	NO	NO	YES	2	53	NO
CONTROL	68	NO	NO	NO	YES	2	45	YES
CASE	76	NO	YES	YES	YES	1	9	YES
CONTROL	76	NQ	YES	YES	YES	2	96+	YES
CONTROL	76	NO	YES	NO	YES	1	3	YES
CONTROL	76	NO	YES	YES	YES	2	15	YES
CONTROL	77	NO	NO	NO	YES	1	36	YES
CASE	71	NO	NO	UNK	YES	UNK	96+	NO
CONTROL	70	YES	NO	NO	YES	2	7	YES
CONTROL	70	NO	NO	NO	YES	0	0	YES
CONTROL	71	NO	YES	YES	YES	2	7	YES
CONTROL	70	NO	NO	YES	YES	2	27	YES
CASE	69	YES	NO	YES	YES	2	36	YES
CONTROL	69	NO	YES	NO	YES	1	96+	YES
CONTROL	69	NO	NO	YES	YES	2	1	YES
CONTROL	69	NO	NO	NO	YES	O	O	YES
CONTROL	68	NO	NO	UNK	NO	0	0	NO
CASE	70	NO	YES	YES	YES	2	71	YES
CONTROL	71	NO	NO	NO	NO	0	0	NO
CONTROL	71	NO	YES	YES	YES	3	S	YES
CONTROL	70	NO	NO	YES	NO	0	0	NO
CONTROL	71	NO	NO	UNK	NO	0	0	NO
CASE	65	YES	NO	NO	YES	1	96+	YES
CONTROL	65	NO	NO	UNK	NO	0	0	NO
CONTROL	64	NO	NO	NO	YES	3	91	YES
CONTROL	64	NO	NO	NO	YES	2	96+	YES
CONTROL	65	NO	NO	YES	YES	2	60	NO
CASE	68	YES	YES	YES	YES	1	36	YES
CONTROL	68	NO	YES	UNK	NO	0	0.	YES
CONTROL	68	NO	NO	YES	NO	0	0	YES
CONTROL	68	YES	NO	UNK	YES	O	O	NO

TABLE A3: ENDOMETRIAL CANCER STUDY

CASE OR CONTROL	AGE	GALL BLADDER DISEASE	HYPER TENSION	OBESITY	ESTROGEN USE	CONJUGATED DOSE	ESTROGEN DURATION (MONTHS)	NON ESTRGN DRUG
CONTROL	68	NO	NO	UNK	YES	1	1	YES
CASE	61	NO	NO	UNK	NO	0	0	YES
CONTROL	61	NO	NO	YES	NO	0	0	YES
CONTROL	61	NO	NO	NO	YES	1	24	YES
CONTROL	61	NO	NO	UNK	NO	0	0	YES
CONTROL	60	YES	NO	NO	NO	0	0	NO
CASE	64	NO	NO	YES	YES	1	54	YES
CONTROL	64	NO	NO	UNK	NO	0	0	NO
CONTROL	65	NO	YES	UNK	YES	3	2	YES
CONTROL	64	NO	YES	YES	YES	3	10	YES
CONTROL	65	NO	NO	UNK	NO	0	0	NO
CASE	68	YES	NO	YES	YES	3	96+	YES
CONTROL	69	NO	NO	UNK	NO	0	0	NO
CONTROL	69	NO	NO	YES	NO	0	0	YES
CONTROL	69	YES	YES	YES	YES	0	0	YES
CONTROL	69	YES	NO	YES	YES	1	35	NO
CASE	74	NO	NO	NO	YES	2	96+	YES
CONTROL	74	NO	YES	NO	YES	3	4	YES
CONTROL	73	NO	YES	NO	YES	2	11	YES
CONTROL	74	NO	NO	YES	YES	1	6	YES
CONTROL	74	NO	NO	YES	YES	1	12	NO
CASE	67	YES	NO	YES	YES	0	0	YES
CONTROL	68	NO	YES	NO	YES	0	0	YES
CONTROL	68	NO	YES	YES	YES	3	65	YES
CONTROL	68	NO	NO	UNK	NO	0	0	NO
CONTROL	68	NO	NO	YES	YES	2	96+	YES
CASE	62	YES	NO	NO	YES	1	UNK	YES
CONTROL	62	YES	NO	NO	NO	0	0	NO
CONTROL	63	NO	NO	YES	NO	0	0	NO
CONTROL	63	NO	NO	UNK	NO	0	0	NO
CONTROL	63	NO	NO	YES	YES	2	UNK	NO
CASE	71	YES	NO	YES	YES	2	59	YES
CONTROL	70	NO	YES	YES	NO	0	0	YES
CONTROL	71	NO	NO	YES	YES	UNK	UNK	YES
CONTROL	71	NO	YES	YES	NO	0	0	YES
CONTROL	71	NO	YES	YES	YES	2	84	YES
CASE	83	NO	YES	YES	YES	3	96+	YES
CONTROL	82	NO	NO	YES	NO	0	0	NO
CONTROL	82	NO	YES	NO	YES	3	4	YES
CONTROL	82	NO	YES	NO	NO	0	0	YES
CONTROL	82	NO	NO	UNK	NO	0	0	NO
CASE	70	NO	NO	YES	NO	0	0	YES
CONTROL	70	YES	YES	YES	YES	2	55	YES
CONTROL	70	NO	YES	YES	YES	2	14	YES
CONTROL	70	NO	YES	YES	YES	1	39	YES

TABLE A3: ENDOMETRIAL CANCER STUDY

CASE OR CONTROL	AGE	GALL BLADDER DISEASE	HYPER TENSION	OBESITY	ESTROGEN USE	CONJUGATED DOSE	ESTROGEN DURATION (MONTHS)	NON ESTRGN DRUG
CONTROL	70	NO	YES	YES	NO	0	0	YES
CASE	74	NO	NO	NO	YES	0	0	YES
CONTROL	75	YES	YES	NO	YES	2	6	YES
CONTROL	74	NO	NO	YES	NO	0	0	YES
CONTROL	74	NO	YES	NO	YES	2	46	YES
CONTROL	75	NO	NO	UNK	NO	0	0	NO
CASE	70	NO	NO	UNK	YES	0	0	YES
CONTROL	70	NO	YES	NO	YES	1	96+	YES
CONTROL	70	NO	NO	UNK	NO	0	0	NO
CONTROL	70	NO	NO	UNK	NO	0	0	YES
CONTROL	70	NO	NO	UNK	NO	0	0	NO
CASE	66	NO	YES	YES	YES	3	48	YES
CONTROL	66	NO	NO	UNK	YES	1	96+	YES
CONTROL	66	NO	NO	UNK	NO	0	0	YES
CONTROL	66	NO	NO	YES	NO	0	0	NO
CONTROL	66	NO	YES	YES	YES	1	12	YES
CASE	77	NO	NO	YES	YES	3	4	YES
CONTROL	77	YES	YES	YES	YES	0	0	YES
CONTROL	77	NO	YES	NO	YES	2	24	YES
CONTROL	77	NO	NO	YES	NO	0	0	NO
CONTROL	78	NO	YES	YES	YES	2	9	YES
CASE	66	NO	YES	NO	YES	3	29	YES
CONTROL	67	NO	YES	NO	NO	0	0	YES
CONTROL	66	NO	NO	YES	NO	0	0	YES
CONTROL	67	NO	NO	YES	NO	0	0	YES
CONTROL	69	NO	YES	YES	YES	2	10	YES
CASE	71	NO	YES	YES	YES	1	96+	NO
CONTROL	72	NO	NO	UNK	NO	0	0	NO
CONTROL	72	NO	NO	NO	NO	0	0	YES
CONTROL	71	NO	NO	UNK	NO	0	0	NO
CONTROL	71	NO	YES	YES	NO	0	0	YES
CASE	80	NO	NO	NO	YES	2	UNK	YES
CONTROL	79	NO	NO	UNK	NO	0	0	NO
CONTROL	79	NO	NO	NO	NO	0	0	NO
CONTROL	79	NO	NO	YES	NO	0	0	NO
CONTROL	80	NO	NO	NO	NO	0	0	NO
CASE	64	NO	NO	YES	YES	2	UNK	YES
CONTROL	64	NO	NO	NO	YES	0	0	YES
CONTROL	63	NO	NO	YES	YES	1	60	YES
CONTROL	64	NO	YES	NO	YES	1	6	YES
CONTROL	66	NO	YES	YES	YES	1	UNK	YES
CASE	63	NO	NO	NO	YES	1	60	YES
CONTROL	63	NO	YES	NO	YES	1	96+	YES
CONTROL	65	NO	NO	NO	YES	1	25	NO
CONTROL	65	NO	NO	NO	NO	0	0	YES
CONTROL	64	NO	NO	NO	YES	1	96+	YES

CASE OR CONTROL	AGE	GALL BLADDER DISEASE	HYPER TENSION	OBESITY	ESTROGEN USE	CONJUGATED DOSE	ESTROGEN DURATION (MONTHS)	NON ESTRGN DRUG
CASE	72	YES	NO	YES	NO	0	0	YES
CONTROL	72	NO	YES	UNK	NO	0	0	NO
CONTROL	72	NO	YES	YES	NO	0	0	YES
CONTROL	72	NO	YES	NO	YES	1	48	YES
CONTROL	72	NO	YES	YES	YES	0	0	YES
CASE	57	NO	NO	NO	YES	3	12	NO
CONTROL	57	NO	YES	YES	YES	0	0	YES
CONTROL	58	NO	NO	YES	YES	1	36	YES
CONTROL	57	NO	NO	NO	YES	1	36	NO
CONTROL	57	NO	NO	NO	YES	0	0	NO
CASE	74	YES	NO	YES	NO	0	0	YES
CONTROL	74	NO	NO	YES	NO	0	0	YES
CONTROL	73	NO	NO	YES	YES	2	2	YES
CONTROL	75	NO	NO	YES	NO	0	0	YES
CONTROL	75	NO	NO	UNK	NO	0	0	NO
CASE	62	NO	YES	YES	YES	2	6	YES
CONTROL	62	NO	NO	YES	YES	2	37	YES
CONTROL	62	NO	NO	YES	YES	2	63	YES
CONTROL	63	NO	NO	UNK	NO	0	0	NO
CONTROL	61	YES	YES	YES	YES	3	96+	YES
CASE	73	NO	YES	YES	YES	1	4	YES
CONTROL	72	NO	NO	NO	YES	2	90	YES
CONTROL	73	NO	NO	NO	YES	3	5	YES
CONTROL	73	NO	YES	NO	YES	1	15	YES
CONTROL	73	NO	YES	NO	NO	0	0	NO
CASE	71	NO	YES	YES	YES	1	UNK	YES
CONTROL	71	NO	NO	UNK	NO	0	0	NO
CONTROL	71	NO	NO	NO	NO	0	0	YES
CONTROL	71	NO	NO	NO	NO	0	0	YES
CONTROL	71	NO	YES	NO	YES	UNK	UNK	YES
CASE	64	NO	YES	YES	NO	0	0	YES
CONTROL	65	NO	NO	YES	YES	3	96+	YES
CONTROL	64	NO	NO	YES	YES	3	96+	YES
CONTROL	64	NO	NO	YES	YES	2	36	YES
CONTROL	64	NO	NO	YES	YES	3	96+	NO
CASE	63	NO	NO	NO	YES	UNK	96+	YES
CONTROL	64	NO	NO	YES	NO	0	0	YES
CONTROL	62	NO	NO	YES	NO	0	0	YES
CONTROL	64	YES	NO	NO	YES	1	18	NO
CONTROL	64	NO	YES	YES	YES	3	UNK	YES
CASE	79	YES	YES	YES	YES	1	96+	YES
CONTROL	78	YES	YES	YES	YES	1	96+	YES
CONTROL	79	NO	NO	YES	NO	0	0	YES
CONTROL	79	NO	YES	NO	YES	0	0	YES
CONTROL	78	NO	NO	YES	YES	1	24	YES

TABLE A3: ENDOMETRIAL CANCER STUDY

CASE OR CONTROL	AGE	GALL BLADDER DISEASE	HYPER TENSION	OBESITY	ESTROGEN USE	CONJUGATED DOSE	ESTROGEN DURATION (MONTHS)	NON ESTRGN DRUG
CASE	80	NO	NO	YES	YES	1	15	YES
CONTROL	81	NO	YES	YES	NO	0	0	YES
CONTROL	81	NO	YES	NO	YES	1	18	YES
CONTROL	80	NO	NO	YES	YES	2	74	YES
CONTROL	80	NO	YES	YES	NO	0	0	YES
CASE	82	NO	YES	YES	YES	2	6	YES
CONTROL	82	NO	NO	YES	NO	0	0	YES
CONTROL	81	NO	YES	UNK	NO	0	0	YES
CONTROL	81	NO	YES	YES	YES	1	12	YES
CONTROL	82	NO	YES	YES	YES	2	13	YES
CASE	71	NO	YES	NO	YES	UNK	84	YES
CONTROL	71	NO	YES	YES	NO	0	0	YES
CONTROL	71	YES	NO	YES	NO	0	0	YES
CONTROL	71	YES	NO	YES	YES	1	96+	YES
CONTROL	71	NO	NO	NO	YES	1	30	YES
CASE	83	NO	YES	YES	YES	3	14	YES
CONTROL	83	NO	YES	YES	NO	0	0	YES
CONTROL	83	NO	NO	NO	NO	0	0	YES
CONTROL	83	NO	YES	YES	YES	2	16	YES
CONTROL	83	NO	NO	NO	NO	0	0	YES
CASE	61	NO	YES	NO	YES	3	96+	YES
CONTROL	60	NO	NO	NO	NO	O	O	YES
CONTROL	61	NO	NO	NO	YES	1	24	YES
CONTROL	62	NO	NO	YES	NO	0	0	YES
CONTROL	61	NO	NO	NO	YES	0	0	YES
CASE	71	NO	NO	NO	YES	1	96+	YES
CONTROL	71	NO	NO	YES	NO	0	0	NO
CONTROL	71	NO	YES	YES	NO	O	O	NO
CONTROL	70	NO	NO	NO	NO	0	0	NO
CONTROL	71	NO	YES	YES	YES	1	3	YES
CASE	69	NO	YES	YES	YES	2	40	YES
CONTROL	69	YES	NO	YES	NO	O	O	YES
CONTROL	70	NO	YES	NO	YES	O	O	YES
CONTROL	70	NO	YES	NO	YES	1	32	YES
CONTROL	70	NO	NO	YES	YES	UNK	UNK	YES
CASE	77	NO	NO	YES	YES	3	73	YES
CONTROL	76	NO	YES	NO	YES	0	0	YES
CONTROL	76	NO	YES	YES	YES	0	0	YES
CONTROL	77	YES	YES	YES	YES	0	0	YES
CONTROL	77	NO	YES	NO	NO	0	0	YES
CASE	64	NO	NO	YES	YES	1	37	NO
CONTROL	64	NO	NO	YES	YES	3	6	NO
CONTROL	63	YES	NO	YES	NO	0	0	NO
CONTROL	63	NO	YES	NO	YES	UNK	UNK	YES
CONTROL	63	NO	YES	YES	NO	0	0	YES

TABLE A3: ENDOMETRIAL CANCER STUDY

CASE OR CONTROL	AGE	GALL BLADDER DISEASE	HYPER TENSION	OBESITY	ESTROGEN USE	CONJUGATED DOSE	ESTROGEN DURATION (MONTHS)	NON ESTRGN DRUG
CASE	79	YES	NO	NO	NO	0	0	NO
CONTROL	82	NO	NO	YES	YES	1	UNK	YES
CONTROL	78	NO	NO	NO	NO	0	0	NO
CONTROL	80	NO	NO	YES	NO	0	0	NO
CONTROL	81	NO	NO	NO	NO	0	0	NO
CASE	72	NO	NO	NO	YES	0	0	YES
CONTROL	72	NO	NO	YES	YES	2	57	YES
CONTROL	73	NO	NO	UNK	NO	0	0	NO
CONTROL	73	YES	YES	NO	YES	2	96+	YES
CONTROL	73	NO	NO	NO	NO	0	0	YES
CASE	82	YES	YES	YES	YES	3	96+	YES
CONTROL	81	NO	NO	UNK	NO	0	0	NO
CONTROL	81	NO	NO	YES	NO	0	0	NO
CONTROL	81	NO	NO	YES	YES	0	0	YES
CONTROL	81	NO	YES	YES	NO	0	0	YES
CASE	73	NO	YES	YES	YES	2	60	YES
CONTROL	74	NO	NO	YES	YES	1	1	YES
CONTROL	75	NO	NO	NO	NO	0	0	YES
CONTROL	75	NO	YES	YES	YES	1	96+	YES
CONTROL	74	NO	NO	NO	NO	0	0	NO
CASE	69	NO	NO	UNK	YES	UNK	UNK	YES
CONTROL	68	NO	NO	NO	NO	0	0	YES
CONTROL	68	NO	NO	YES	YES	2	48	YES
CONTROL	68	NO	NO	NO	YES	1	96+	NO
CONTROL	70	NO	NO	NO	NO	0	0	NO
CASE	79	NO	YES	YES	YES	1	67	YES
CONTROL	79	NO	YES	YES	NO	0	0	YES
CONTROL	79	NO	YES	YES	NO	0	0	YES
CONTROL	78	YES	NO	YES	YES	1	UNK	YES
CONTROL	79	NO	NO	YES	NO	0	0	YES
CASE	72	NO	NO	YES	YES	3	60	NO
CONTROL	71	NO	NO	NO	YES	0	0	YES
CONTROL	72	NO	NO	NO	NO	0	0	YES
CONTROL	72	NO	YES	YES	YES	3	96+	YES
CONTROL	71	NO	YES	YES	YES	3	12	YES
CASE	72	NO	YES	YES	YES	1	27	YES
CONTROL	72	NO	YES	YES	YES	1	3	YES
CONTROL	71	NO	NO	UNK	NO	0	0	NO
CONTROL	72	NO	YES	YES	NO	0	0	YES
CONTROL	72	NO	YES	YES	NO	0	0	YES
CASE	65	NO	YES	YES	YES	2	16	YES
CONTROL	67	NO	NO	NO	NO	0	0	NO
CONTROL	67	NO	NO	UNK	NO	0	0	NO
CONTROL	66	NO	NO	YES	NO	O	0	YES
CONTROL	66	NO	NO	NO	YES	2	3	NO

TABLE A3: ENDOMETRIAL CANCER STUDY

CASE OR CONTROL	AGE	GALL BLADDER DISEASE	HYPER TENSION	OBESITY	ESTROGEN USE	CONJUGATED DOSE	ESTROGEN DURATION (MONTHS)	NON ESTRGN DRUG
CASE	67	NO	YES	YES	YES	2	96+	YES
CONTROL	66	NO	NO	YES	YES	2	56	YES
CONTROL	66	NO	NO	YES	NO	0	0	NO
CONTROL	67	NO	NO	YES	YES	1	UNK	YES
CONTROL	67	NO	NO	YES	YES	2	34	YES
CASE	64	YES	NO	YES	YES	3	96+	YES
CONTROL	63	NO	NO	YES	NO	0	0	YES
CONTROL	64	NO	NO	YES	YES	1	4	YES
CONTROL	63	NO	NO	YES	NO	0	0	YES
CONTROL	65	NO	NO	UNK	NO	0	0	NO
CASE	62	NO	NO	UNK	YES	2	36	NO
CONTROL	63	NO	NO	YES	NO	0	0	NO
CONTROL	62	NO	NO	NO	NO	0	0	YES
CONTROL	62	NO	NO	UNK	YES	3	UNK	NO
CONTROL	62	NO	NO	UNK	NO	0	0	NO
CASE	83	YES	YES	UNK	NO	0	0	YES
CONTROL	83	YES	NO	UNK	NO	0	0	NO
CONTROL	82	NO	NO	NO	YES	2	6	YES
CONTROL	83	NO	NO	UNK	NO	0	0	YES
CONTROL	83	YES	NO	UNK	NO	0	0	YES
CASE	81	NO	NO	YES	YES	0	0	YES
CONTROL	79	NO	NO	UNK	NO	0	0	NO
CONTROL	80	NO	NO	YES	NO	0	0	YES
CONTROL	82	NO	NO	YES	NO	0	0	YES
CONTROL	80	NO	NO	NO	NO	0	0	NO
CASE	67	NO	NO	YES	YES	2	96+	YES
CONTROL	66	NO	NO	YES	YES	2	40	YES
CONTROL	68	NO	NO	UNK	NO	0	0	YES
CONTROL	65	NO	NO	ND	NO	0	0	YES
CONTROL	LO	NO	YES	YES	YES	1	96+	YES
CASE	73	YES	YES	YES	YES	1	UNK	YES
CONTROL	72	NO	NO	YES	YES	1	12	YES
CONTROL	71	NO	NO	YES	YES	1	96+	YES
CONTROL	73	YES	NO	NO	YES	2	96+	YES
CONTROL	72	NO	NO	YES	NO	0	0	YES
CASE	67	YES	NO	NO	YES	3	96+	YES
CONTROL	67	YES	NO	YES	YES	2	96+	YES
CONTROL	68	NO	NO	YES	NO	0	0	NO
CONTROL	67	NO	NO	YES	NO	0	0	NO
CON1ROL	67	NO	NO	YES	NO	0	0	YES
CASE	74	NO	YES	YES	YES	2	9	YES
CONTROL	75	NO	NO	NO	NO	0	0	YES
CONTROL	75	NO	NO	UNK	NO	0	0	NO
CONTROL	75	YES	YES	NO	NO	0	0	YES
CONTROL	75	NO	NO	NO	YES	2	41	YES
CASE	68	YES	NO	YES	YES	3	18	YES
CONTROL	69	NO	NO	YES	YES	2	96+	YES
CONTROL	70	NO	NO	UNK	NO	0	0	NO
CONTROL	69	NO	YES	YES	YES	2	92	YES
CONTROL	69	NO	YES	NO	YES	3	59	YES

APPENDIX B

DYALOG APL PROGRAMS

The computations required in the thesis are undertaken by using Dyalog APL Version 7.1. The following table summarises the main programs and related functions used in various chapters.

Table B1: Dyalog ALP Programs used in Various Chapters

Chapter	Description	APL Programs & Sub-programs
Three	To simulate and fit Poisson-gamma model.	SIMPG, POSGAM POISSON
Four	To compare Poisson-normal and Poisson-gamma models.	COMPPGPN, DATAPN DATAPG, NORMALR POSNML, POSGMM
Five	To simulate matched case-control data and fit mixed and fixed effects model. To calculate derivatives. A general program for fitting mixed model to matched case-control data.	SIMMATCH, DEV MATCHDAT, DATA MATCHMIX, GLIM LL0M1, LL0M2, LL0M MCASECONT LIKEDRV, COEFF
Six	To analyse contingency table of Neighbourhood data under Poisson assu--mption and to calculate derivatives. A general space efficient program.	POSCON GLIMP DEVP POSCONSP
Seven	To fit fixed and mixed effect multinomial model and to calculate derivatives and Deviance.	GLIMM MULTMIX LL0MN, DEVM
Eight	To fit fixed and mixed effect Threshold models. To calculate Deviances.	THRES THRESMIX DEVTH, DEVTHM
Nine	To adjust SEs for 1 and 2 random components	ADJSE1 ADJSE2
All	Some general functions used in various chapters.	BINO, BLOCK, BNLEFT DET, FMSE, MEAN DIAG, RND, STD UNIT, VP, VAR

SIMPG
=====

```
     ∇ SIMPG
[1]   ⍝To generate POS-GAM data and fit Pisson-Gamma model
[2]    DATAPG
[3]    POSGAM Y
[4]    →(C=1)/L2
[5]    BETA0←BETA0,1↑BETA
[6]    BETA1←BETA1,BETA[2]
[7]    THETA1←THETA1,THETA
[8]    SETH1←SETH1,SE
[9]    SEBET0←SEBET0,SEBETA[1]
[10]   SEBET1←SEBET1,SEBETA[2]
[11]  L2:'CONVERGED SIMULATION=',⍴THETA1
     ∇
```

POSGAM
======

```
     ∇ POSGAM Y;B;IS;TH;X;Y;L;VV;W;T;I1;I2;J;K
[1]   ⍝Fits Poisson-gamma model. Various initial values are
[2]   ⍝tried with automatically. The first column of Y is the
[3]   ⍝response vector and the remaining columns are design
[4]   ⍝matrix for fixed and random parameters.
[5]    K←(¯1×J)+J←¯0.2
[6]  LBL0:J←J+0.2,0⍴K←0
[7]  LBL1:THETA←0.1×K←K+1
[8]    'Attempt=',K,'Initial U=',J,'Initial TH=',K×0.1
[9]    BETA←2,1,15⍴J,0⍴I1←I2←0
[10] LBL2:L←(Y[;1]-*(0 1↓Y)+.×BETA),30 30⍴(*Y[;2 3]+.×2↑BETA),30 30⍴0
[11]   VV←⊟((⍉0 1↓Y)+.×(0 1↓L)+.×0 1↓Y)+17 17⍴((2⍴0),15⍴÷THETA),17 17⍴0
[12]   BETA←(W←BETA)+VV+.×((⍉0 1↓Y)+.×L[;1])+(2⍴0),(÷THETA)×1-*2↓BETA
[13]   →((0.003≤⌈/2↑|W-BETA)∧8≥I1←I1+1)/LBL2
[14]   THETA←(W←THETA)+((+/2↓BETA*2)++/+/(T←2 2↓VV)×(UNIT 15))÷15
[15]   SE←((2×+/+/(T-THETA×UNIT 15)*2)*0.5)÷15
[16]   →((T←0.02<|THETA-W)∧12≥I2←I2+I1←1)/LBL2
[17]   →((0<T)∧10≥K)/LBL1
[18]   →((0<T)∧0.8≥J)/LBL0
[19]   SEBETA←(VV[1;I2←1],VV[2;2])*0.5,0⍴C←T
[20]   'BETA=',(2↑BETA)
[21]   'SEBETA=',SEBETA
[22]   'Theta= ',THETA,'SETHETA=',SE,T,J,K×0.1
[23]   DEVP BETA
     ∇
```

POISSON
=======

```
     ∇ R←N POISSON L;P;Y
[1]   ⍝920417 Simulates N values of Poisson variable
[2]   ⍝with parameters specified by vector L with N
[3]   ⍝components, one for each simulation
[4]    P←+\*((⍟L)∘.×0,Y)-L∘.+0,+\⍟Y←⍳⌈/L+6×L*0.5
[5]    R←+/P<(0.000001×?N⍴999999)∘.×(1+⍴P)⍴1
     ∇
```

COMPGPN
=======

```
      ∇ COMPGPN
[1]    ⍝Generates data under Poisson-Normal or Poisson-gamma
[2]    ⍝models and fits both Poisson-Normal and Poisson-gamma
[3]    ⍝models. This needs to be run repeatedly for the
[4]    ⍝required number of simulations
[5]     DATAPN
[6]    ⍝DATAPG
[7]    ⍝Use either DATAPN or DATAPG by commenting out one.
[8]     POSNML Y
[9]     POSGMM Y
[10]   →((C+C1)>0)/L2
[11]   BET0←BET0,1↑BET
[12]   BET1←BET1,BET[2]
[13]   THET1←THET1,THET
[14]   S1←S1,SEP
[15]   SEBET0←SEBET0,SEBET[1]
[16]   SEBET1←SEBET1,SEBET[2]
[17]   BETA0←BETA0,1↑BETA
[18]   BETA1←BETA1,BETA[2]
[19]   THETA1←THETA1,THETA
[20]   SE1←SE1,SE
[21]   SEBETA0←SEBETA0,SEBETA[1]
[22]   SEBETA1←SEBETA1,SEBETA[2]
[23] L2:'ACTUAL SIMULATION =',⍴THET1
      ∇
```

DATAPG
======

```
      ∇ DATAPG;ETA;S;U;ETA;BETA;X;A
[1]    ⍝Generates data under Poisson-gamma model
[2]     U←⍟(+/(15,A)⍴-⍟0.0001×?(15×A)⍴9999)÷A←4
[3]     X←(⍉2 30⍴(30⍴1),¯1+?30⍴2),Z←⍉15 30⍴1 1,30⍴0
[4]     BETA←3 1.5,U
[5]     ETA←X+.×BETA
[6]     S←30 POISSON⋆ETA
[7]     Y←S,X
      ∇
```

DATAPN
======

```
      ∇ DATAPN;ETA;S;U;ETA;BETA;X
[1]    ⍝To generate data under Poisson+Normal model
[2]     U←15 NORMALR 0 1
[3]     X←(⍉2 30⍴(30⍴1),¯1+?30⍴2),Z←⍉15 30⍴1 1,30⍴0
[4]     BETA←1 0.5,U
[5]     ETA←X+.×BETA
[6]     S←30 POISSON⋆ETA
[7]     Y←S,X
      ∇
```

POSGMM
======

```
      ∇ POSGMM Y;B;IS;TH;X;Y;L;VV;W;T;I1;I2;CV;J;K
[1]    ⍝Fits Poisson-gamma model. Various initial values are
[2]    ⍝tried with automatically. The first column of Y is the
[3]    ⍝response vector and the remaining columns are design
[4]    ⍝matrix for fixed and random parameters.
[5]    K←(¯1×J)+J←¯0.2
[6]    LBL0:J←J+0.2,0ρK←0
[7]    LBL1:THETA←0.1×K←K+1
[8]    'Attempt=',K,'Initial U=',J,'Initial TH=',K×0.1
[9]    BETA←2,1,15ρJ,0ρI1←I2←0
[10]   LBL2:L←(Y[;1]-*(0 1↓Y)+.×BETA),30 30ρ(*Y[;2 3]+.×2↑BETA),30 30ρ0
[11]   VV←⊞((⍉0 1↓Y)+.×(0 1↓L)+.×0 1↓Y)+17 17ρ((2ρ0),15ρ÷THETA),17 17ρ0
[12]   BETA←(W←BETA)+VV+.×((⍉0 1↓Y)+.×L[;1])+(2ρ0),(÷THETA)×1-*2↓BETA
[13]   →((0.003≤⌈/2↑|W-BETA)∧8≥I1←I1+1)/LBL2
[14]   THETA←(W←THETA)+((+/2↓BETA*2)++/+/(T←2 2↓VV)×(UNIT 15))÷15
[15]   →((T←0.02<|THETA-W)∧12≥I2←I2+I1←1)/LBL2
[16]   →((0<T)∧10≥K)/LBL1
[17]   →((0<T)∧0.8≥J)/LBL0
[18]   T1←+/+/VV[M;M←2↓⍳17]×W∘.=W←⍳15
[19]   T2←+/+/VV[M;M]*2
[20]   B←2×⊞W×(15-2×T1÷THET)+T2×W←THET*¯2
[21]   SEBETA←(VV[1;I2←1],VV[2;2])*0.5,0ρC←T
[22]   SE←B*0.5
[23]   'BETA=',(2↑BETA)
[24]   'SEBETA=',SEBETA
[25]   'SETHETA=',SE
[26]   'Theta= ',THETA,'SETHETA=',SE,T,J,K×0.1
[27]   DEVP BETA
      ∇
```


DEVP
====

```
      ∇ DEVP B;S;F;D;M
[1]    ⍝Calculate deviance for Poisson response
[2]    S←+/Y[;1]-*M←(0 1↓Y)+.×B
[3]    F←Y[;1]+.×(⍟Y[;1]+0.0000000001)-M
[4]    D←2×F-S
[5]    'Deviance='D
      ∇
```

POSNML
======

```
      ∇ POSNML Y;B;IS;TH;X;Y;L;VV;W;T;I1;I2;J;K
[1]    ⍝Fits Poisson-normal model. Various initial values are
[2]    ⍝tried with automatically. The first column of Y is the
[3]    ⍝response vector and the remaining columns are design
[4]    ⍝matrix for fixed and random parameters.
[5]    K←(¯1×J)+J←¯0.2
[6]    LBL0:J←J+0.2,0ρK←0
[7]    LBL1:THET←0.1×K←K+1
[8]     'Attempt=',K,'Initial U=',J,'Initial TH=',K×0.1
[9]    BET←3,1,15ρJ,0ρI1←I2←0
[10]   LBL2:L←(Y[;1]-*(0 1↓Y)+.×BET),30 30ρ(*(0 1↓Y)+.×BET),30 30ρ0
[11]   VV←⊞((⍉0 1↓Y)+.×(0 1↓L)+.×0 1↓Y)+17 17ρ((2ρ0),15ρ÷THET),17 17ρ0
[12]   BET←(W←BET)+VV+.×((⍉0 1↓Y)+.×L[;1])-(2ρ0),(÷THET)×2↓BET
[13]   →((0.002≤⌈/2↑|W-BET)∧8≥I1←I1+1)/LBL2
[14]   THET←(W←THET)+((+/2↑BET*2)++/+/(T←2 2↓VV)×(UNIT 15))÷15
[15]   →((T←0.003<|THET-W)∧10≥I2←I2+I1←1)/LBL2
[16]   →((0<T)∧10≥K)/LBL1
[17]   →((0<T)∧0.6≥J)/LBL0
[18]   T1←+/+/VV[M;M←2+⍳17]×W∘.=W←⍳15
[19]   T2←+/+/VV[M;M]*2
[20]   B←2×⊞W×(15-2×T1÷THET)+T2×W←THET*¯2
[21]   SEBET←(VV[1;I2←1],VV[2;2])*0.5,0ρC1←T
[22]   SEP←B*0.5
[23]   'BET=',(2↑BET)
[24]   'SEBET=',SEBET
[25]   'SETHET=',SEP
[26]   'Thet= ',THET,'SETHET=',SEP,T,J,K×0.1
[27]   DEVP BET
      ∇
```


NORMALR
=======

```
      ∇ R←N NORMALR PA;V
[1]    ⍝82/02/02. RETURNS  N  NORMAL[PA←→MEAN,VARIANCE] VARIATES,
[2]    ⍝USING THE BOX-MULLER, AMSX(1958), METHOD.
[3]    R←(?(2,⌈N÷2)ρ2147483647)÷2147483647
[4]    R←N↑(,1 2∘.○R[2;]×○2)×V,V←(¯2×⍟R[1;])*0.5
[5]    →(∧/PA=0 1)/0
[6]    R←PA[1]+R×PA[2]*0.5
[7]    ⍝
      ∇
```

MATCHMIX
========

```
      ∇ MATCHMIX X;V;N;IC;P;R;L;T;I;J;B;VV;VW;W;M;K;ML;TH
[1]   ⍝Fits mixed models for matched case-control data
[2]   ⍝with random exposure effects. The first column of X
[3]   ⍝indicates case or control and the second column is
[4]   ⍝exposure.
[5]    V←1,(1↑⍴X)÷5
[6]    BETA←1
[7]    P←⍴THETA←0.5,0⍴0
[8]    BETA←BETA,(+/1↓V)⍴0.5,0⍴N←(IC←1)↑⍴X
[9]    ML←2
[10]  ⍝Use LLOM1 for single case and LLOM2 for 2 cases
[11]  LBL1:L←LLOM1(0 1↓X)+.×BETA
[12]   VW←←(I,I)⍴VV,(I,I←⍴VV←(V[1]⍴0),(V[2]⍴÷THETA))⍴0
[13]   VV←⌹((⍉0 1↓X)+.×(0 1↓L)+.×0 1↓X)+VW
[14]   BETA←(B←BETA)+VV+.×((⍉0 1↓X)+.×L[;1])-VW+.×BETA
[15]   →(0.001≤⌈/|B-BETA)/LBL1
[16]   →(ML≠1)/LBL2
[17]   VV←(((J,J)↑VV),(J,B)⍴0),[1]((((B←+/1↓V),J)⍴0),⌹(J,J←V[1])↑⌹VV
[18]  LBL2:T←((2+P),P)⍴¯1+I←J←1
[19]  LBL3:T[1;J]←+/BETA[M←(+/J↑V)+W←⍳V[J+1]]*2
[20]   T[2;J]←+/+/VV[M;M]×W∘.=W
[21]  L4:T[(2+I);J]←T[(2+J);I]←+/+/VV[(K←(+/I↑V)+⍳V[I+1]);M]*2
[22]   →((⍴V)>I←I+1)/L4
[23]   →((⍴V)>I←J←J+1)/LBL3
[24]   TH←T[1;]÷(1↓V)-R←T[2;]÷B←THETA
[25]   THETA←TH+3×TH-B
[26]   →((12≥IC←IC+1)∧C←0.01<⌈/|TH-B)/LBL1
[27]   THETA←TH
[28]   SEP←((2×+/+/(((J,J)↑VV)-THETA×UNIT J←V[1])*2)*0.5)÷V[2]
[29]   L←1 2⍴(⊂'Regression coeff'),(⊂'S.E.')
[30]   B←⌹I←(((⍳J)∘.=⍳J)×VV←(J,J←V[1])↑VV)*0.5
[31]   L,[1](6 RND(J↑BETA),[1.5]S1←+/I)
[32]   L←1 2⍴(⊂'Theta'),(⊂'S.E.')
[33]   B←(UNIT P)×VW←2×⌹(DIAG(W×(1↓V)-2×R))+(2 0↓T)×W∘.×W←THETA*¯2
[34]   L,[1](6 RND THETA),[1.5],SETH←+/I←(|B)*0.5
[35]   'SEP=',SEP
[36]   ⎕←'ZERO/ONE INDICATES CONVERGENCE/NONCONVERGENGE OF THETA EST',C
      ∇
```


GLIM
====

```
      ∇ GLIM X;N;L;VV;W;DEV;B;K;I
[1]   ⍝To undertake a fixed effect analysis ignoring
[2]   ⍝random exposure effect
[3]    BETA←1,0⍴1
[4]   LBL5:L←LLOM(0 1↓X)+.×BETA
[5]    VV←⌹((⍉0 1↓X)+.×(0 1↓L)+.×0 1↓X)
[6]    BETA←(B←BETA)+VV+.×((⍉0 1↓X)+.×L[;1])
[7]    →(0.001≤⌈/|B-BETA)/LBL5
[8]    B←⌹I←VV*0.5
[9]    L←1 2⍴(⊂'Regression coeff'),(⊂'S.E.')
[10]   L,[1](6 RND BETA,[1.5]S←+/I)
      ∇
```

SIMMATCH
========

```
     ∇ SIMMATCH;Q2
[1]    ⍝To generate match dataset and fit models.
[2]    ⍝Q1=no of matched set, Q2 to be taken large
[3]    ⍝enuogh to select required case and control.
[4]    Q←(Q1←30)×Q2←35
[5]   LB:DATA
[6]    5 MATCHDAT 2
[7]    →((1↑⍴V)<Q1)/LB
[8]    MATCHMIX X
[9]    →(C=1)/LB
[10]   BETA1←BETA1,1↑BETA
[11]   SE1←SE1,S1
[12]   THETA1←THETA1,THETA
[13]   SET1←SET1,SEP
[14]   SET2←SET2,SETH
[15]   GLIM X[;1,2]
[16]   BETA0←BETA0,BETA
[17]   SE←SE,S
     ∇
```


MATCHDAT
========

```
     ∇ IK MATCHDAT CS;R;X2;W;W1;I;J
[1]    ⍝To randomly select expoused/not exposed
[2]    ⍝corresponding to cases/controls in each set
[3]    R←(Q1,(Q2))⍴Y
[4]    X2←(Q1,(Q2))⍴X1
[5]    V←(0,IK)⍴I←1
[6]   LO1:W←R[I;]
[7]    W1←X2[I;]
[8]    →(0=+/W)/LO2
[9]    →((IK-CS)>+/W=0)/LO2
[10]   V←V,[1](CS↑(1=W)/W1),(IK-CS)↑(0=W)/W1
[11]  LO2:→(Q1≥I←I+1)/LO1
[12]   I←IK×(IJ←1↑⍴V)
[13]   Y←I⍴((CS⍴1),(IK-CS)⍴0)
[14]   X←⍉(2,I)⍴Y,X←I⍴V
[15]   X←X,(⍉(IJ,I)⍴X[;2])×⍉(IJ,I)⍴(IK⍴1),I⍴0
[16]   X1←I⍴⍉(IK,IJ)⍴⍳IJ
[17]   X1←X1,[2]X[;1,2]
     ∇
```


DATA
====

```
     ∇ DATA;Z1;B;U;W;ETA
[1]    ⍝To generate binary exposure data
[2]    X1←((0.001×?Q⍴999)≤0.4)
[3]    Z1←(Q1,Q)⍴(Q2⍴1),Q⍴0
[4]    B←¯1+(⍉Z1)+.×U←(Q1 NORMALR 0 1)
[5]    PHI←W÷1+W←*ETA←X1×B
[6]    ⍝ ⎕RL←+/⎕TS
[7]    Y←((0.001×?Q⍴999)≤PHI)
     ∇
```

MCASECONT
=========

```
      ∇ MCASECONT X;V;IJ;W;XX;I;N;B;W1;L;W2;T;C;TH;Q;CD;S
[1]    ;W4;W5;G;IC;MV;TR2;VR
[2]   ⍝970422 Matrix X has first column the matched set number with
[3]   ⍝consecutive numbers starting at 1.  Col 2 is case=1, control=0
[4]   ⍝Remaining columns are risk variables, with last column being the
[5]   ⍝treatment/exposure variable. Model fitted takes treat/exp effect
[6]   ⍝as varying randomly over matched sets with variance theta.
[7]    B←0ρV←¯3+1↓ρX
[8]    TH←0.1
[9]    ML←0
[10]   B←B,(N+1)ρ0,0ρN←⌈/X[;I←IJ←1]
[11]  LBL1:C←(V,V)ρT←Vρ(LL←0),,CD←((V+1),0)ρS←⍳¯1+IC←1
[12]  LBL2:W←0 2↓XX←(X[;1]=IC)/[1]X
[13]   L←XX[;2]LIKEDRV((W1←0 ¯1↓W)+.×V↑B)+(W2←,W[;V+1])×B[V+1]+B[V+1+IC]
[14]   C←C+(⍉W1)+.×(1 1↓L)+.×W1
[15]   CD←CD,(⍉W1,W2)+.×(1 1↓L)+.×W2
[16]   T←T+(⍉W1)+.×1↓L[;1]
[17]   S←S,W2+.×1↓L[;1]
[18]   LL←LL+L[1;1]
[19]   →(N≥IC←IC+1)/LBL2
[20]   MV←CD+.×W←÷(W5←,CD[V+1;])+÷TH
[21]   M1←(+/W1+S←(S-(W1←(V+1)↓B)÷TH)×W)÷TH
[22]   G←⌷C-((W1×(Vρ1)∘.×W)+.×⍉W1←¯1 0↓CD)+(W2∘.×W2←¯1↓MV)÷TH×W4←MV[V+1]
[23]   Q←G+.×T-(M1×W2÷W4)+W1+.×S
[24]   B←B+Q,(-W3),S+((((W3←((W2+.×Q)-TH×M1)÷W4)×W5)-Q+.×W1)×W
[25]   →((10≥I←I+1)∧0.001≤⌈/|Q,W3)/LBL1
[26]   TR←(+/W)+0×TR2←+/(W×W5)*2
[27]   →(ML=1)/LBL4
[28]   TR←TR+(TH×TR2÷W4)+(TR2×W2+.×G+.×W2)÷W4*2
[29]   TR←TR+(+/+/((W1×(Vρ1)∘.×W*2)+.×⍉W1)×G)-2×(W1+.×W5×W*2)+.×G+.×W2
[30]  LBL4:TH←(0×W←TH)+(+/(V+1)↓B*2)÷N-TR÷TH
[31]   →((10≥IJ←IJ+I←1)∧0.01<|W-TH)/LBL1
[32]   VR←(G,-W5),[1](-W5),(TH+W2+.×W5←G+.×W2÷W4)÷W4
[33]  ⍝VAR←(TH+W2+.×W5←G+.×W2÷W4)÷W4
[34]   'Estimates and SEs of risk,exposure variable coefficients'
[35]   6 RND((V+1)↑B),[0.5]W←(+/VR×(⍳V+1)∘.=⍳V+1)*0.5
[36]   'Correlation matrix'
[37]   3 RND W+.×VR+.×W←((V+1),V+1)ρ(÷W),((V+1),V+1)ρ0
[38]   'Estimate and standard error of THETA'
[39]   TH,[0.5]TH÷(2×TR2)*0.5
      ∇
```


COEFF
=====

```
      ∇ L←IJ COEFF N;I;W
[1]   ⍝970421 Finds the sum of all products of components of N
[2]   ⍝taken IJ at a time.
[3]    →(1≥L←(IJ≥0)+IJ>0)/0
[4]    I←L←1+0×W←ρN
[5]   LBL1:L←(L,0)+0,L×N[I]
[6]    →(W≥I←I+1)/LBL1
[7]    L←L[IJ+1]
      ∇
```

188

DEV
= = =

```
     ∇ DEV B;S;F;D;M
[1]    ⍝Calculate deviance for Poisson response
[2]    S←+/Y[;1]-*M←(0 1↓Y)+.×B
[3]    F←Y[;1]+.×(⍟Y[;1]+0.0000000001)-M
[4]    D←2×F-S
[5]    'Deviance='D
     ∇
```

POSCON
= = = = = =

```
     ∇ POSCON X;V;N;IC;P;R;L;T;I;J;B;VV;VW;W;M;K;ML;X1
[1]    ⍝The program used to analyse neighbourhood dataset
[2]    ⍝under the proposed mixed model scheme in chapter 6.
[3]    V←9,5⍴20
[4]    X1←((1↑⍴X),V[1]+1)↑X
[5]    BET←¯0.5 0.8 ¯0.5 ¯0.5 ¯0.5,4⍴0.3
[6]    P←⍴THET←((⍴1↓V)⍴0.1),0⍴0.5
[7]    BET←BET,(+/1↓V)⍴0,0⍴N←(IC←1)↑⍴X
[8]  LBL1:L←X[;1]-*(0 1↓X)+.×BET
[9]    L←L,(N,N)⍴(*(0 1↓X)+.×BET),(N,N←1↑⍴X)⍴0
[10]   VW←(I,I)⍴VV,(I,I←⍴VV←(V[1]⍴0),(+/1↓V)⍴÷THET)⍴0
[11]   VV←⊞((⍉0 1↓X)+.×(0 1↓L)+.×0 1↓X)+VW
[12]   BET←(B←BET)+VV+.×((⍉0 1↓X)+.×L[;1])-VW+.×BET
[13]   →(0.0001≤⌈/|B-BET)/LBL1
[14]  LBL2:T←((2+P),P)⍴¯1+J←1
[15]  LBL3:T[1;J]←+/BET[M←(V[1]+J-B)+(B←⍴THET)×⍳V[J+1]]*2
[16]   T[2;J]←+/+/VV[M;M]×W∘.=W←⍳V[J+1]
[17]   →((⍴V)>J←J+1)/LBL3
[18]   THET←(0×B←THET)+1×(T[1;]+T[2;])÷1↓V
[19]   THET←B+2×THET-B
[20]   →((14≥IC←IC+1)∧C1←0.0003<⌈/|THET-B)/LBL1
[21]   THET←B
[22]   SEBETA←+/((J J↑VV)×UNIT J←V[1])*0.5
[23]   'BETA=',V[1]↑BET
[24]   'SEBETA=',SEBETA
[25]   ⍝for likelihood ratio test
[26]   X DEV BET
[27]   DB←(V[2]×⍟THET[1])+(÷THET[1])×IC+.×IC←V[2]↑V[1]↓BET
[28]   DB←-DB+(V[3]×⍟THET[2])+(÷THET[2])×IC+.×IC←V[3]↑(P←V[1]+V[2])↓BET
[29]   DB←-DB+(V[4]×⍟THET[3])+(÷THET[3])×IC+.×IC←V[4]↑(P←P+V[3])↓BET
[30]   DB-⍟DET(J J↑⌹VV)
[31]   ⍝for WALD test
[32]   'WTEST=',(BET[W]+.×(⌹VV[W;W])+.×BET[W←6 7 8 9])
[33]   'THETA=',THET
[34]   'ZERO INDICATES CONVERGENCE',C1
     ∇
```

POSCONSP
========

```
      ∇ POSCONSP X;V;N;IC;P;R;L;T;I;J;B;VV;VW;W;M;K;ML;X1
[1]    ⍝To fit space efficient model. The arrangement of X
[2]    ⍝columns are different here. For random effect design
[3]    ⍝matrix (Z) the columns are rearranged to make it diagonal.
[4]     V←⎕,0ρ⎕←'ENTER NUMBER OF FIXED EFFECTS'
[5]     F←+/V←V,⎕,0ρ⎕←'ENTER NO. OF EFFECTS IN EACH RANDOM TERM'
[6]     X1←((1↑ρX),V[1]+1)↑X
[7]     BET←⎕,0ρ⎕←'ENTER INITIAL VALUES FOR FIXED COEFF'
[8]     P←ρTHET←((ρ1↓V)ρ0.1),0ρ0.5
[9]     BET←BET,(+/1↓V)ρ0,0ρN←(IC←1)↑ρX
[10]  LBL1:L←(N,1)ρL←X[;1]-*(0 1↓X)+.×BET
[11]    L←L,*(0 1↓X)+.×BET
[12]    VW←(I,I)ρVV,(I,I←ρVV←(V[1]ρ0),(+/1↓V)ρ÷THET)ρ0
[13]    VV←(F,F)ρI-I←J←1
[14]  LB11:VV[I;J]←X[;I+1]+.×L[;2]×X[;J+1]
[15]    →((J←J+1)≤F)/LB11
[16]    →((I←I+(J←1))≤V[1])/LB11
[17]  LB13:VV[I;J]←VV[J;I]
[18]    →((I←I+1)≤F)/LB13
[19]    →((J←J+1)<(I←V[1]+1))/LB13
[20]  LB14:VV[I;J]←X[;I+1]+.×L[;2]×X[;J+1]
[21]    →((J←J+1)≤F)/LB14
[22]    J←V[1]+1
[23]    →((I←I+1)≤F)/LB14
[24]    VV←⌹(VV+VW)
[25]    D←FρJ←1
[26]  DD:D[J]←L[;1]+.×X[;J+1]
[27]    →((J←J+1)≤F)/DD
[28]    BET←(B←BET)+VV+.×D-VW+.×BET
[29]    →(0.0001≤⌈/|B-BET)/LBL1
[30]  ⍝remove comment for ML estimate of THET
[31]  ⍝VV←(((J,J)↑VV),(J,B)ρ0),[1](((B←+/1↓V),J)ρ0),⌹(J,J←V[1])↑⌹VV
[32]  LBL2:T←((2+P),P)ρ¯1+J←1
[33]  LBL3:T[1;J]←+/BET[M←(V[1]+J-B)+(B←ρTHET)×⍳V[J+1]]*2
[34]    T[2;J]←+/+/VV[M;M]×W∘.=W←⍳V[J+1]
[35]    →((ρV)>J←J+1)/LBL3
[36]    THET←(0×B←THET)+1×(T[1;]+T[2;])÷1↓V
[37]    THET←B+1×THET-B
[38]    →((14≥IC←IC+1)∧C1←0.0003<⌈/|THET-B)/LBL1
[39]    THET←B
[40]    SEBETA←+/((J J↑VV)×UNIT J←V[1])*0.5
[41]    'BETA=',V[1]↑BET
[42]    'SEBETA=',SEBETA
[43]    X DEV BET
[44]  ⍝For WALD test
[45]    'WTEST=',(BETA[W]+.×(⌹VV[W;W])+.×BETA[W←6 7 8 9])
[46]    'THETA=',THET
[47]    'ZERO INDICATES CONVERGENCE',C1
      ∇
```

190

LIKEDRV
=======

```
     ∇ L←Y LIKEDRV N;A;B;C;J;K;W;M
[1]    ⍝960421 For Y=vector of cases/controls and N=eta, returns
[2]    ⍝matrix L with L[1;1]=loglikelihood, remaining elements of
[3]    ⍝first row/col as derivates of loglikelihood and other rows
[4]    ⍝/cols containing second order derivatives of log likelihood.
[5]    A←(M←+/Y)COEFF N←*N
[6]    C←(W,W)ρB←(W←ρN)ρ0×J←K←1
[7]  LBL1:C[J;J]←B[J]←N[J]×(M-1)COEFF(J≠⍳W)/N
[8]    →(J=1)/LBL3
[9]  LBL2:C[J;K]←C[K;J]←N[J]×N[K]×(M-2)COEFF(1=1-(J=⍳W)+K=⍳W)/N
[10]   →(J>K←K+1)/LBL2
[11] LBL3:→(W≥J←J+K←1)/LBL1
[12]   L←(((+/Y×N)-⍟A),W),[1](W←Y-B÷A),(C÷A)-B∘.×B÷A*2
     ∇
```


LLOM
====

```
     ∇ L←LLOM W;S;B
[1]   ⍝To calculate derivatives in ordinary GLIM
[2]    S←(1↑ρX)÷IK
[3]    R←÷R←(B←S BLOCK IK)+.×(W←*W)
[4]    L←(X[;1]-S),‾1×(B×VP S)-DIAG(S←W×R)
     ∇
```


LLOM1
=====

```
     ∇ L←LLOM1 W;S;B
[1]   ⍝To calculate the derivatives for a single case in a set
[2]    S←(1↑ρX)÷IK
[3]    R←÷R←(B←S BLOCK IK)+.×(W←*W)
[4]    L←(X[;1]-S),‾1×(B×VP S)-DIAG(S←W×R)
     ∇
```


LLOM2
=====

```
     ∇ L←LLOM2 W;S;B;I;J;A;C;T
[1]   ⍝To calculate the derivatives for 2 cases in a set
[2]    S←IJ BLOCK IK
[3]    A←(IJ,IK-1)ρ0ρI←J←1
[4]    T←(IJ,IK)ρW←*W
[5]  LL:A[I;J]←T[I;J]×(+/J↓T[I;])
[6]    →(IK>J←J+1)/LL
[7]    →(IJ≥I←I+J←1)/LL
[8]    A←(I←ρW)ρ⍉(IK,IJ)ρ+/A
[9]    B←W×(Iρ⍉(IK,IJ)ρ+/T)-W
[10]   C←S×(VP W)
[11]   C←(DIAG B)+C×(1-UNIT I)
[12]   L←(X[;1]-A×B),T×C-(T←⍉(I,I)ρA←÷A)×S×B∘.×B
     ∇
```

191

GLIMM
=====

```
      ∇ GLIMM X;N;L;VV;W;DEV;B;K;I;J;V
[1]    ⍝To fit multinomial model. The first column of X
[2]    ⍝is the response and the remaining columns are design
[3]    ⍝matrix for fixed parameters.
[4]     BETA←⎕,0ρ⎕←'ENTER INITIAL VALUES OF FIXED PARAMETERS'
[5]     I←1↓ρNT←⎕,0ρ⎕←'ENTER THE VARIABLE CONTAINING MARGINALS'
[6]    LBL1:L←LLOMN(0 1↓X)+.×BETA
[7]     VV←⌹((⍉0 1↓X)+.×(0 1↓L)+.×0 1↓X)
[8]     BETA←(B←BETA)+VV+.×((⍉0 1↓X)+.×L[;1])
[9]     →(T←(0.001≤⌈/|B-BETA)∧20≥I←I+1)/LBL1
[10]    SEBETA←+/(VV×UNITρBETA)*0.5,0ρC←T
[11]    'BETA=',(8 3⍕BETA),C
[12]    'SEBETA=',9 3⍕SEBETA
[13]    X DEVM BETA
[14]    'WTEST=',(BETA[W]+.×(⌹VV[W;W])+.×BETA[W←3 4])
      ∇
```


DEVM
====

```
      ∇ Y DEVM B;F1;F0;D;SW;T
[1]    ⍝To calculate deviance in MULTINOMIAL model
[2]     SW←(B←((N←1↑ρX)÷R)BLOCK R←2)+.×W←*(0 1↓Y)+.×B
[3]     M←NT×W÷1+SW
[4]     F1←Y[;1]+.×(⍟Y[;1]+0.0000000001)-⍟M+0.0000000001
[5]     N←N÷R
[6]     T←((N,R)ρNT)[;1]
[7]     F0←Y0+.×(⍟0.0000001+Y0←T-+/(N,R)ρY[;1])-⍟0.0000001+D←T-+/(N,R)ρM
[8]     D←2×F1+F0
[9]     'Deviance='D
      ∇
```


GLIMP
=====

```
      ∇ GLIMP Y;L;VV;W;Y;I
[1]    ⍝To fit fixed effect Poisson model, the first column
[2]    ⍝of Y is response and the remaining clumns are explanatory
[3]    ⍝variables
[4]     BG←1,I←1
[5]    LBL2:L←(Y[;1]-*(0 1↓Y)+.×BG),30 30ρ(*Y[;2 3]+.×BG),30 30ρ0
[6]     VV←⌹((⍉0 1↓Y)+.×(0 1↓L)+.×0 1↓Y)
[7]     BG←(W←BG)+VV+.×((⍉0 1↓Y)+.×L[;1])
[8]     →(T←(0.01≤⌈/|2↑W-BG)∧10≥I←I+1)/LBL2
[9]     SEBG←(VV[1;I2←1],VV[2;2])*0.5,0ρC←T
[10]    'BG=',BG
[11]    'SEBG=',SEBG
[12]    DEVP BG
      ∇
```

LLOMN
=====

```
      ∇ L←LLOMN W;SW;B;SD
[1]     SW←(B←((N←1↑ρX)÷R)BLOCK R←2)+.×W←*W
[2]     L←(X[;1]-NT×W÷1+SW),(⍉(N,N)ρNT)×B×(DIAG SD)-VP SD←W÷1+SW
      ∇
```

MULTMIX
=======

```
      ∇ MULTMIX X;V;N;IC;P;R;L;T;I;J;B;VV;VW;W;M;K;C;ML;VT
[1]    ⍝To fit mixed multinomial model. The first column of X
[2]    ⍝is the response and the remaining columns are design
[3]    ⍝matrix for fixed and random parameters.
[4]     V←⎕,0ρ⎕←'ENTER NUMBER OF FIXED COEFFICIENTS'
[5]     V←V,⎕,0ρ⎕←'ENTER NUMBER OF EFFECTS IN EACH RANDOM TERM'
[6]     NT←⎕,0ρ⎕←'ENTER THE VARIABLE CONTAINING MARGINALS'
[7]     BETA←V[1]ρ0.1
[8]     P←ρTHETA←⎕,0ρ⎕←'ENTER INITIAL VARIANCES OF RANDOM VECTORS'
[9]     BETA←BETA,(+/1↓V)ρ0,0ρN←(IC←1)↑ρX
[10]    'INITIAL ESTIMATES OF RANDOM COMPONENTS ARE TAKEN TO BE ZERO'
[11]    ML←⎕,0ρ⎕←'ENTER 1 FOR ML ESTIMATE, OTHERWISE REML GIVEN'
[12]   LBL1:L←LLOMN(0 1↓X)+.×BETA
[13]    VW←(I,I)ρVV,(I,I←ρVV←(V/(0,÷THETA)))ρ0
[14]    VV←(VT,(VT←+/V))ρI-I←J←1
[15]   LB11:VV[;J]←(⍉0 1↓X)+.×(0 1↓L)+.×X[;J+1]
[16]    →((J←J+1)≤VT)/LB11
[17]    VV←⌹(VV+VW)
[18]    D←VTρJ←1
[19]   DD:D[J]←L[;1]+.×X[;J+1]
[20]    →((J←J+1)≤VT)/DD
[21]    BETA←(B←BETA)+VV+.×D-VW+.×BETA
[22]    8 3⍕12↑BETA
[23]    →(0.01≤⌈/V[1]ρ|B-BETA)/LBL1
[24]    →(ML≠1)/LBL2
[25]    VV←(((J,J)↑VV),(J,B)ρ0),[1](((B←+/1↓V),J)ρ0),⌹(J,J←V[1])↑⌹VV
[26]   LBL2:T←((2+P),P)ρ¯1+I←J←1
[27]   LBL3:T[1;J]←+/BETA[M←(+/J↑V)+W←⍳V[J+1]]*2
[28]    T[2;J]←+/+/VV[M;M]×W∘.=W
[29]   L4:T[(2+I);J]←T[(2+J);I]←+/+/VV[(K←(+/I↑V)+⍳V[I+1]);M]*2
[30]    →((ρV)>I←I+1)/L4
[31]    →((ρV)>I←J←J+1)/LBL3
[32]    THETA←T[1;]÷(1↓V)-R←T[2;]÷B←THETA
[33]    →((8≥IC←IC+1)∧C←0.005<⌈/|THETA-B)/LBL1
[34]    SEBETA←+/((J J↑VV)×UNIT J←V[1])*0.5
[35]    'BETA=',V[1]↑BETA
[36]    'SEBETA=',SEBETA
[37]    X DEVM BETA
[38]    'WTEST=',(BETA[W]+.×(⌹VV[W;W])+.×BETA[W←9 10 11 12])
[39]    'WTEST=',(BETA[W]+.×(⌹VV[W;W])+.×BETA[W←9 10])
[40]    'WTEST=',(BETA[W]+.×(⌹VV[W;W])+.×BETA[W←11 12])
[41]    'THETA=',THETA
[42]    ⎕←'ZERO INDICATES CONVERGENCE ','C=',C
      ∇
```

```
      ∇ Y THRES X;X1;C;R;T;E;G;GP;D;DTN;DTP;D1;G1;G1P;I;IC;W;WP;WY
[1]     ;WPY;DT;DB;DP;VT;DB2;DT2;DTB;SEBETA
[2]    ⍝To fit fixed effect threshold model. Y is a vector of
[3]    ⍝responses to be recorded first row then second and so on,
[4]    ⍝X is the transpose of the design matrix for row effects
[5]     R←⎕,0⍴⎕←'ENTER NUMBER OF ROWS'
[6]     C←⎕,0⍴⎕←'ENTER NUMBER OF COLUMNS,C'
[7]     N←R×C×IC←1
[8]     X1←(N,(1+⍴X))⍴X
[9]     BTH←⎕,0⍴⎕←'ENTER INITIAL VALUES FOR C-2 THRESHOLD PARAMETERS'
[10]    BTH←BTH,⎕,0⍴⎕←'ENTER INITIAL VALUES FOR ROW EFFECTS'
[11]   LOOP0:TH←(D←×/⍴TH)⍴TH←(R⍴0),[2](R,C-2)⍴(C-2)↑BTH
[12]    E←D⍴(C-1)/E←X+.×(C-2)↓BTH
[13]   ⍝E←D⍴(C-1)/E←X×(C-2)↓BTH
[14]    G←N⍴(D1←(R,C-1)⍴D←W÷E←1+(W←*TH-E)),[2](R⍴1)
[15]    VT←G-GP←N⍴(R⍴0),[2]D1
[16]    G1←N⍴(D1←(R,C-1)⍴(D÷E)),[2](R⍴0)
[17]    G1P←N⍴(R⍴0),[2]D1
[18]    WY←Y×W←G1÷VT
[19]    WPY←Y×WP←G1P÷VT
[20]    DT←(1↓¯1↓+/⍉(R,C)⍴WY)-2↓+/⍉(R,C)⍴WPY
[21]    DB←+/(⍉X)+.×(R,C)⍴WPY-WY
[22]    D←N⍴(D1←(R,C-1)⍴(1-2×D)),[2]R⍴0
[23]    DP←N⍴(R⍴0),[2]D1
[24]    DB2←(⍉X1)+.×(DIAG(N⍴⍉(R,C)⍴Y×(W×D)+(¯1×WP×DP)-(W-WP)*2))+.×X1
[25]    DT2←(1↓¯1↓+/⍉(R,C)⍴WY×D-W)-2↓+/⍉(R,C)⍴WPY×(WP+DP)
[26]    DTP←1↓¯1↓+/⍉(R,C)⍴WY×WP
[27]    DTP←DTN←2↓+/⍉(R,C)⍴WY×WP
[28]    DTB←(⍉0 1↓0 ¯1↓(R,C)⍴WY×W+(¯1×WP)+¯1×D)+.×X
[29]    DTB←DTB-(⍉0 2↓(R,C)⍴WPY×W+(¯1×WP)+¯1×DP)+.×X
[30]    VT←DIAG DT2
[31]    →((⍴DT)≤I←1)/LBL1
[32]   LOOP1:VT[I;I+1]←DTN[I]
[33]    VT[I+1;I]←DTP[I]
[34]    →((⍴DT)>I←I+1)/LOOP1
[35]   LBL1:VV←¯1×(VT,DTB),[1](⍉DTB),DB2
[36]    VV←⊞VV
[37]    BTH←(B←BTH)+VV+.×(DT,DB)
[38]    BTH←(B×0)+1×BTH
[39]    →((12≥IC←IC+1)∧DP←0.001≤⌈/|B-BTH)/LOOP0
[40]    SEBTH←+/(VV×UNIT⍴BTH)*0.5
[41]    'BTH=',(8 3⍕BTH),W
[42]    'SEB=',8 3⍕SEBTH
[43]    Y DEVTH BTH
[44]    'WTEST=',(BTH[W]+.×(⊟VV[W;W])+.×BTH[W←(C-2)↓⍳⍴BTH])
[45]    ⎕←'ZERO INDICATES CONVERGENCE ',DP
      ∇
```

```
      ∇ Y THRESMIX X;X1;C;R;R1;T;E;K;Z;G;GP;V;D;DTN;DTP;D1;G1;G1P
[1]      ;I;I2;IC;W;WP;WY;WPY;DT;DB;DP;VT;VW;DB2;DT2;DTB;BTHU
[2]    ⍝Contingency tab analysis with mixed effect threshold model.
[3]    ⍝Y is a vector of responses to be recorded first row then
[4]    ⍝second and so on, X is the transpose of the design matrix
[5]    ⍝for row effects
[6]      R1←⎕,0ρ⎕←'ENTER NUMBER OF ROWS'
[7]      C←⎕,0ρ⎕←'ENTER NUMBER OF COLUMNS,C'
[8]      K←⎕,0ρ⎕←'ENTER NUMBER OF CLUSTERS'
[9]      N←C×R←R1×K
[10]     V←1,R1,K
[11]     X2←(X2←(R,V[2])ρX),⍉(K,R)ρ((R÷K)ρ1),Rρ0
[12]     BTHU←(BTH←(+/2↑V)ρ1.4 0.055 ¯0.309),U←V[3]ρ0
[13]     P←ρ(THET←0.78),0ρIC←I2←1
[14]  LOOP0:TH←(D←×/ρTH)ρTH←(Rρ0),[2](R,C-2)ρ(C-2)↑BTH
[15]    ⍝Choose one of the following two lines depending on
[16]    ⍝full model or intercept only model
[17]     E←(Dρ((C-1)×R1)/U)+E←Dρ(C-1)/E←X+.×(C-2)↓BTH
[18]   ⍝E←(Dρ((C-1)×R1)/U)+E←Dρ(C-1)/E←X×(C-2)↓BTH
[19]     G←Nρ(D1←(R,C-1)ρD←W÷E←1+(W←*TH-E)),[2]Rρ1
[20]     GP←Nρ(Rρ0),[2]D1
[21]     G1←Nρ(D1←(R,C-1)ρ(D÷E)),[2](Rρ0)
[22]     G1P←Nρ(Rρ0),[2]D1
[23]     WY←Y×W←G1÷VT←G-GP
[24]     WPY←Y×WP←G1P÷VT
[25]     DT←(1↓¯1↓+/⍉(R,C)ρWY)-2↓+/⍉(R,C)ρWPY
[26]     DB←+/(⍉X2)+.×(R,C)ρWPY-WY
[27]     D←Nρ(D1←(R,C-1)ρ(1-2×D)),[2]Rρ0
[28]     DP←Nρ(Rρ0),[2]D1
[29]     X1←←(N,(1+ρX2))ρX2
[30]     DB2←(⍉X1)+.×(DIAG(Nρ⍉(R,C)ρY×(W×D)+(¯1×WP×DP)-(W-WP)*2))+.×X1
[31]     DT2←(1↓¯1↓+/⍉(R,C)ρWY×D-W)-2↓+/⍉(R,C)ρWPY×(WP+DP)
[32]     DTP←1↓¯1↓+/⍉(R,C)ρW×WPY
[33]     DTN←3↓+/⍉(R,C)ρWY×WP
[34]     DTB←(⍉0 1↓0 ¯1↓(R,C)ρWY×W+(¯1×WP)+¯1×D)+.×X2
[35]     DTB←DTB-(⍉0 2↓(R,C)ρWPY×W+(¯1×WP)+¯1×DP)+.×X2
[36]     VT←DIAG DT2
[37]     →((ρDT)≤1)/LBL1
[38]  LOOP1:VT[I;I+1]←DTN[I]
[39]     VT[I+1;I]←DTP[I]
[40]     →((ρDT)>I←I+1)/LOOP1
[41]  LBL1:VV←¯1×(VT,DTB),[1](⍉DTB),DB2
[42]     VV←⌹VV+VW←(I,I)ρVW,(I,I←ρVW←(V/(0,0,÷THET)))ρ0
[43]     BTHU←(B←BTHU)+VV+.×(DT,DB)-VW+.×BTHU
[44]     BTH←(+/2↑V)↑BTHU
[45]     U←(+/2↑V)↓BTHU
[46]     'BTH=',BTH
[47]     →((8≥I2←I2+1)∧Z←0.01≤⌈/(+/2↑V)↑|B-BTHU)/LOOP0
[48]    ⍝FOR ML
[49]    ⍝VV←(((K,K)↑VV),(K,B)ρ0),[1](((B←ρBTH),K)ρ0),⌹(K,K)↑⌹VV
[50]    ⍝END ML
[51]  LBL2:T←((2+P),P)ρ¯1+J←I2←1
[52]  LBL3:T[1;J]←+/BTHU[M←(+/(J+1)↑V)+W←⍳V[J+2]]*2
[53]     T[2;J]←+/+/VV[M;M]×W∘.=W
[54]     →((ρV)>1+J←J+1)/LBL3
[55]     THET←T[1;]÷(2↓V)-T[2;]÷B←THET
[56]     →((10≥IC←IC+1)∧Z←0.007<⌈/|THET-B)/LOOP0
[57]     SEBTH←+/((J J↑VV)×UNIT J←+/2↑V)*0.5
[58]     'BTH=',(8 3⍕BTH),Z
[59]     'SEB=',8 3⍕SEBTH
```

```
[60]    'WTEST=',(BTH[W]+.×(⊟VV[W;W])+.×BTH[W←3])
[61]    Y DEVTHM BTH
[62]    DB←(V[3]×⊛THET[1])+(÷THET[1])×IC+.×IC←(+/2↑V)↓BTH
[63]    'LKHOOD VALUE=',D+DB←DB+⊛DET(J J↓⊟VV)
[64]    □←'ZERO INDICATES CONVERGENCE ',Z
    ∇
```

DEVTH
=====

```
    ∇ Y DEVTH BTH;M;T;PS;PF;G;E
[1]   ⋀To calculate deviance for fixed effect threshold model
[2]   ⋀Y=response vector, BTH=thres para,row effects
[3]    M←NρC/M←+/T←(R,C)ρY
[4]    PS←+/Y×⊛(Y÷M)+0.00000001
[5]    E←(D←R×C-1)ρ(C-1)/E←X+.×(C-2)↓BTH
[6]    G←(R,C-1)ρG←E÷1+E←*(DρTH←0,(C-2)↑BTH)-E
[7]    PF←+/Y×⊛PF←(Nρ(G,[2]Rρ1)-(Rρ0),[2]G)
[8]    D←2×PS-PF
[9]    'DEV=',D
    ∇
```

DEVTHM
======

```
    ∇ Y DEVTHM BTH;M;T;PS;PF;G;E
[1]   ⋀To calculate deviance for mixed threshold model
[2]    M←NρC/M←+/T←(R,C)ρY
[3]    PS←+/Y×⊛(Y÷M)+0.00000001
[4]    E←(D←R×C-1)ρ(C-1)/(R1/U)+E←X+.×(C-2)↓BTH
[5]    G←(R,C-1)ρG←E÷1+E←*(DρTH←0,(C-2)↑BTH)-E
[6]    PF←+/Y×⊛PF←(Nρ(G,[2]Rρ1)-(Rρ0),[2]G)
[7]    D←2×PS-PF
[8]    'DEV=',D
    ∇
```

196

```
BINO
====

     ∇ BINO NT;P;I
[1]     Y←I←0
[2]   L1:P←(1 BNFR NT(PHI[I←I+1]))
[3]     Y←Y,P
[4]     →(I<120)/L1
[5]     Y←1↓Y
[6]   ⍝X←⍉2 120ρY,X1
     ∇


BLOCK
=====

     ∇ T←N BLOCK B;K;C
[1]     K←1
[2]     T←(B,B)ρ1
[3]   LBL1:T←T,(C,C←K×B)ρ0
[4]     T←T,[1]((C,C)ρ0),(C,C)↑T
[5]     K←K+1×K
[6]     →(N>((1↑ρT)÷B))/LBL1
[7]     T←(K,K←N×B)↑T
     ∇


BNFR
====

     ∇ R←N BNFR NP;PX;B;RN
[1]   ⍝ Binomial[NP←NU,PI] freq table,Random Sample of N≤1000
[2]     →(3=⎕NC'BNLEFT')/L_1
[3]     'BNLEFT'⎕ CY'PRDISC'
[4]   L_1:→(N≤1000)/L_2
[5]     →L_1,N←⎕,0ρ⎕←'SAMPLE SIZE > 1000; ENTER SAMPLE SIZE.'
[6]   L_2:PX←NP[1]BNLEFT NP
[7]     R←+/(¯1+ιρPX)∘.=+/[1](B×+\PX)∘.<?NρB←(⌊/ι0)⌊⌈÷⌊/(PX≠0)/PX
[8]     R←(R>0)/RN←0,ιNP[1]
     ∇


BNLEFT
======

     ∇ R←N BNLEFT NP;NU;X;PI
[1]   L_1:→(N≤NU←NP[1])/L_2
[2]     'TOO MANY TERMS SOUGHT: UP TO WHAT VALUE REALLY? AT MOST NU'
[3]     →L_1,N←⎕
[4]   L_2:→(NU=¯1+ρR←(X!NU)×(PI*X)×(1-PI←NP[2])*NU-X←0,ιN-1)/0
[5]     R←R,0⌈1-+/R
     ∇
```

ADJSE1
======

```
      ∇ ADJSE1 G;D;T;F;FW;V;SE
[1]    ⍝To calculate adjusted SE for single random component
[2]    ⍝G is the second derivatives from fixed effect model.
[3]    ⍝X1 and Z1 are design matrices for fixed and random respectively.
[4]    D←DIAG(+/(⍉Z1)+.×G+.×Z1)+÷THETA
[5]    T←(⍉X1)+.×G+.×Z1
[6]    V←⊞(⊟R)-T←T+.×(⊟D)+.×⍉T
[7]    V←+/V×UNIT(1↓⍴X1)
[8]    'Adjusted SE=',SE←V*0.5
      ∇
```


ADJSE2
======

```
      ∇ ADJSE2 G;D;T1;T2;M;N;E;D1;D2;DB;V;SE
[1]    ⍝To calculate adjusted SE for two random components
[2]    ⍝G is the second derivatives from fixed effect model
[3]    ⍝X1,Z1,Z2 are design matrices for fixed and random respectively.
[4]    D1←DIAG(+/(⍉Z1)+.×G+.×Z1)+÷THETA[1]
[5]    T1←(⍉X1)+.×G+.×Z1
[6]    D2←DIAG(+/(⍉Z2)+.×G+.×Z2)+÷THETA[2]
[7]    T2←(⍉X1)+.×G+.×Z2
[8]    DB←(⍉Z1)+.×G+.×Z2
[9]    E←⊞D2-(⍉DB)+.×F←(⊟D1)+.×DB
[10]   N←-F+.×E
[11]   M←(⊟D1)+F+.×E+.×(⍉F)
[12]   M←(T1+.×M+.×⍉T1)+(T2+.×(⍉N)+.×⍉T1)+(T1+.×N+.×⍉T2)+(T2+.×E+.×⍉T2)
[13]   V←⊞(⊟R)-M
[14]   V←+/V×UNIT(1↓⍴X1)
[15]   'Adjusted SE=',SE←V*0.5
      ∇
```


BLOCK
=====

```
      ∇ T←N BLOCK B;K;C
[1]    K←1
[2]    T←(B,B)⍴1
[3]   LBL1:T←T,(C,C←K×B)⍴0
[4]    T←T,[1]((C,C)⍴0),(C,C)↑T
[5]    K←K+1×K
[6]    →(N>((1↑⍴T)÷B))/LBL1
[7]    T←(K,K←N×B)↑T
      ∇
```

ADJSE1
======


```
     ∇ ADJSE1 G;D;T;F;FW;V;SE
[1]    ⍝To calculate adjusted SE for single random component
[2]    ⍝G is the second derivatives from fixed effect model.
[3]    ⍝X1 and Z1 are design matrices for fixed and random respectively.
[4]    D←DIAG(+/(⍉Z1)+.×G+.×Z1)+÷THETA
[5]    T←(⍉X1)+.×G+.×Z1
[6]    V←⌹(⌹R)-T+T+.×(⌹D)+.×⍉T
[7]    V←+/V×UNIT(1↓ρX1)
[8]    'Adjusted SE=',SE←V*0.5
     ∇
```


ADJSE2
======


```
     ∇ ADJSE2 G;D;T1;T2;M;N;E;D1;D2;DB;V;SE
[1]    ⍝To calculate adjusted SE for two random components
[2]    ⍝G is the second derivatives from fixed effect model
[3]    ⍝X1,Z1,Z2 are design matrices for fixed and random respectively.
[4]    D1←DIAG(+/(⍉Z1)+.×G+.×Z1)+÷THETA[1]
[5]    T1←(⍉X1)+.×G+.×Z1
[6]    D2←DIAG(+/(⍉Z2)+.×G+.×Z2)+÷THETA[2]
[7]    T2←(⍉X1)+.×G+.×Z2
[8]    DB←(⍉Z1)+.×G+.×Z2
[9]    E←⌹D2-(⍉DB)+.×F←(⌹D1)+.×DB
[10]   N←-F+.×E
[11]   M←(⌹D1)+F+.×E+.×(⍉F)
[12]   M←(T1+.×M+.×⍉T1)+(T2+.×(⍉N)+.×⍉T1)+(T1+.×N+.×⍉T2)+(T2+.×E+.×⍉T2)
[13]   V←⌹(⌹R)-M
[14]   V←+/V×UNIT(1↓ρX1)
[15]   'Adjusted SE=',SE←V*0.5
     ∇
```


BLOCK
=====


```
     ∇ T←N BLOCK B;K;C
[1]    K←1
[2]    T←(B,B)ρ1
[3]  LBL1:T←T,(C,C←K×B)ρ0
[4]    T←T,[1]((C,C)ρ0),(C,C)↑T
[5]    K←K+1×K
[6]    →(N>((1↑ρT)÷B))/LBL1
[7]    T←(K,K←N×B)↑T
     ∇
```

```
COR
===

      ∇ X COR Y
[1]     ((MEAN X×Y)-(MEAN X)×(MEAN Y))÷((VAR X)×(VAR Y))*0.5
      ∇


CORM
====

      ∇ I←CORM X;J
[1]     J←1↑ρX
[2]     I←(COVM X)×÷(⍉I)+.×I←(1,J)ρI←(VAR X)*0.5
      ∇


COV
===

      ∇ X COV Y
[1]     (MEAN X×Y)-(MEAN X)×(MEAN Y)
      ∇


COVM
====

      ∇ I←COVM X
[1]     I←X-(ρX)ρMEAN X
[2]     I←(⍉I)+.×I÷(1↑ρX)
      ∇


UNIT
====

      ∇ T←UNIT I
[1]     T←(⍳I)∘.=⍳I
      ∇


VP
==

      ∇ H←VP B;I
[1]     H←((I,1)ρB)+.×((1,I←ρB)ρB)
      ∇


RND
===

      ∇ R←N RND X
[1]    ⍝920310 Rounds X to N decimal places
[2]     R←(10*-N)×⌊0.5+(10*N)×X
      ∇
```

DET
===

```
      ∇ R←DET M;K;S;G;F;P
[1]    ⍝83/08/10. Returns the determinant of the square matrix  M.
[2]    →(0=ρρR←M)/0
[3]    →((=/ρM)∧2=ρρM)/LBL1
[4]    'NO DETERMINANT: SQUARE MATRIX REQUIRED'
[5]    →
[6]   LBL1:K←ι(S←1)↑ρM
[7]   LBL2:S←S×¯1*P≠G←¯1+P+F⍳⌈/F←|M[K;P←(ι0)ρ1↑K]
[8]    M[P,G;]←M[G,P;]
[9]    M[K;]←M[K;]-(M[K←1↓K;P]÷M[P;P])∘.×M[P;]
[10]   →(1<ρK)/LBL2
[11]   R←|×/S,1 1⍉M
[12]  ⍝Determinant of square matrix  M
      ∇
```

DIAG
====

```
      ∇ A←DIAG X
[1]    A←(N,N)ρX,(N,N←ρX)ρ0
      ∇
```

FMSE
====

```
      ∇ FMSE W;M;N;V
[1]    □←'Mean = ',M←(+/[1]W)÷N←1↑ρW
[2]    □←'SE = ',(+/(((⍉W)+.×W←W-(Nρ1)∘.×M)÷¯1+N)×(ιV)∘.=ιV←ρM)*0.5
      ∇
```

MEAN
====

```
      ∇ M←MEAN X
[1]    M←(+/⍉X)÷1↑ρX
      ∇
```

STD
===

```
     ∇ S←STD X
[1]    S←((+/⍉X*2)÷1↑ρX)-(MEAN X)*2
[2]    S←S*0.5
     ∇
```


VAR
===

```
     ∇ S←VAR X
[1]    S←((+/⍉X*2)÷1↑ρX)-(MEAN X)*2
     ∇
```