GENERALIZED QUASI-LIKELIHOOD VERSUS HIERARCHICAL LIKELIHOOD INFERENCES IN GENERALIZED LINEAR MIXED MODELS FOR COUNT DATA

MD. RAFIQUL I CHOWDHURY

Your file Votre référence
ISBN: 978-0-494-33412-6
Our file Notre référence
ISBN: 978-0-494-33412-6

Generalized Quasi-likelihood versus Hierarchical Likelihood Inferences in Generalized Linear Mixed Models for Count Data

by

©Md. Rafiqul I Chowdhury

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirement for the Degree of Master of Science in Statistics

Department of Mathematics and Statistics
Memorial University of Newfoundland

August, 2007

St. John's Newfoundland Canada

Abstract

Inferences in generalized linear mixed models (GLMMs) which includes count and binary data as special cases are extremely important. As it is proven to be difficult to obtain consistent and efficient estimates of the parameters (regression effects and variance of the random effects) of such models, there is a vast growing literature dealing with this important estimation problem. Among them, the method of moments (MM), Penalized quasilikelihood (PQL) and Hierarchical likelihood (HL) approaches are more familiar. It is however known that the MM approach always produces consistént estimates, whereas the PQL approach may not provide consistent estimates for all the parameters of the model. A recently proposed generalized quasilikelihood (GQL) approach has proven to be better in the sense of consistency and efficiency as compared to the MM and other improved MM (IMM) procedures. There does not, however, exist any comparative study between the GQL and the HL approaches. In this thesis, we have made a comparison between these two approaches mainly through an extensive simulation study involving the Poisson-normal mixed model. It is found that the HL approach may not produce consistent estimates for the regression effects specially when the variance of the random effects is large. In contrast, the GQL approach is found to always produce consistent estimates for all parameters of the model. These two estimation methodologies are also illustrated by analyzing a data set on the health care utilization in St. John's, Canada .

Acknowledgements

To give just an acknowledgment to Professor B.C. Sutradhar would be an understatement, for his role and plethora of guidance given to me during my M.Sc. program and towards the completion of my thesis. With this, I would like to express my deep heart–felt gratitude to my supervisor, Prof. B.C. Sutradhar for his supervision, advice, encouragement, enduring patience and constant support. He has always provided clear explanations when I was stuck, constantly driving the research with new and innovative research methodologies. It was undoubtedly a great privilege for me to work under his supervision for which I shall forever be grateful.

I would also like to express my heartiest gratitude to my co-supervisor, Dr. Noel Cadigan of Department of Fisheries and Oceans, Northwest Atlantic Fisheries Center for his valuable comments and helpful suggestions in completing this thesis.

I want to sincerely acknowledge the financial support provided by the School of Graduate Studies, Department of Mathematics and Statistics, Prof. B.C. Sutradhar, and Dr. Noel Cadigan in the form of Graduate Assistantship & Teaching Assistantships. Further, I wish to thank the Department, the lovely and helpful staff at our General Office for providing a very friendly ambience, nice conversations and the necessary facilities, especially the stationary material for which they have been very generous to me.

I also thank Dr. JC Loredo–Osti and Dr. Alwell Oyet of Department of Mathematics & Statistics, Memorial University of Newfoundland for taking the time to serve as examiners of this thesis. Their constructive comments and suggestions served to improve the quality of the thesis.

It is my great pleasure to thank all my friends and well–wishers who directly or indirectly encouraged and helped me in the M.Sc. program and contributed to this dissertation.

Last and most importantly, I am very much grateful to my parents, my brother and sisters and specially my wife, for their strong belief in my capabilities and for all of their eternal love, support and encouragement during my stay far away from them.

I'm ever grateful to my parents for what I'm now...

Contents

List of Tables

Chapter 1

Introduction

1.1 Background of the Problem

Analyzing clustered count data is an important problem in economics and biomedical studies, among others. For example, in health economics studies, one may be interested to estimate the effects of certain covariates such as gender and education level on the number of visits to the physician paid by different members of a family. Here, the number of visits to the physician is a count response variable. Similarly, in biomedical studies, the number of weekly asthma attacks on a member of a family can be treated as a count response variable. Interest of the study may be to estimate the effects of different covariates such as gender, smoking habit and mother's smoking habit on the number of weekly asthma attacks.

In both of these examples, the count responses (number of visits to the physician or number of weekly asthma attacks) within a family are correlated. This correlation arises from the shared common unobserved family effects. It is important to determine the effect of the covariates on the responses after taking such familial correlations into account.

There are many studies on the estimation of the regression effects as well as the variation of the family effects. These studies are done in a generalized linear mixed model (GLMM) set-up which accommodates count as well as binary response data.

In general, whether the responses are count or binary, the GLMMs for such responses are derived from the well-known generalized linear model (GLM) (McCullagh and Nelder, 1989) by adding random effects to the linear predictor. A random variable Y_{ij} for the jth member (j=1,..., n_i) of the ith family (i=1,..., K), with exponential density

$$f(y_{ij}) = exp[\{y_{ij}\,\eta_{ij} - a(\eta_{ij})\} + b(y_{ij})] \tag{1.1}$$

follows a GLM when η_{ij} has a linear form, namely, $\eta_{ij} = x'_{ij}\beta$ where x_{ij} is the vector of covariates and β is the vector of regression parameters. In (1.1), a(.) and b(.) are known functions. Note that the exponential form (1.1) contains the Poisson distribution for count response y_{ij}. As far as the GLMM is concerned, it is developed by adding random effects, say γ_i to the linear predictor η_{ij}, where γ_i is independently and identically distributed with mean 0 and variance σ^2, i.e. $\gamma_i \overset{iid}{\sim} (0, \sigma^2)$. Thus in the GLMM set-up, the count response y_{ij} follows (1.1) with $a(\eta_{ij}) = exp\,(\eta_{ij})$, where $\eta_{ij} = x'_{ij}\beta + \sigma\, z_i\, \gamma_i^*$, $\gamma_i^* = \frac{\gamma_i}{\sigma} \overset{iid}{\sim} (0, 1)$, and z_i is a random effects related known covariate for the ith cluster.

Under the GLMMs setup, it has proven to be difficult to obtain consistent and efficient estimators for the regression parameters and the variance of the random effects. A full or exact likelihood analysis is complicated as it requires a complex integration over the distribution of the random effects. This integration problem compels one to avoid the exact likelihood estimation method, even though it is known that maximum likelihood estimators will be fully efficient (optimal). To overcome this computational problem, many authors, over the last decade, have used best linear unbiased predictor (BLUP) analogue analytical methods, where random effects are treated as fixed effects (Henderson 1953) and estimated as such. The regression and variance parameter of the GLMMs are then estimated, based on the estimates of the so-called random effects. For example, we refer to Schall (1991), Breslow and Clayton (1993), Breslow and Lin (1995), Kuk (1995), Lin and Breslow (1996), and Lee and Nelder (1996, 2001).

Breslow and Clayton (1993), by using these BLUP analogue approaches, proposed two closely related approximate methods, namely the penalized quasi-likelihood (PQL) method and the marginal quasi-likelihood (MQL) method for inferences in GLMMs. The PQL generally produces biased estimates, especially for the variance of the random effects. The amount of bias can be considerably large when the true variance of the random effects is large and the cluster size is small (see Sutradhar and Qu, 1998).

To remove biases in the estimates, Kuk (1995) and Lin and Breslow (1996), among others, provided asymptotic bias corrections both for the regression and the variance component estimates. Breslow and Lin (1995), in the context of binary GLMMs with a single component of dispersion, provided a correction factor for the estimator of the variance of the random effects derived from a Laplace approximations (Solomon and Cox, 1992) as well as PQL. They also provided a first order correction term for the regression coefficients estimated by PQL (see also Goldstein and Rasbash, 1996, for improvements). The bias correction in PQL estimators due to Breslow and Lin (1995) appears to improve the asymptotic performance of the uncorrected quantities only when the true variance component is small; more specifically, when it is less than or equal to 0.25.

Following the generalized estimating equation approach of Zeger et al. (1988), Breslow and Clayton (1993), as mentioned above, also used the MQL method to estimate the regression effects of GLMMs. The application of the estimating equation approach for the regression parameters requires the first and second order marginal moments of the responses. The exact first and second order moments of the responses under the GLMMs are, however, typically not available. Breslow and Clayton (1993) used an approximate mean vector and "a working covariance" matrix as in Zeger et al.

(1988) to construct the estimating equations for the regression parameters (see also Waclawiw and Liang, 1993). Similar to PQL approach, this "working" covariance-based MQL approach also produces biased estimates for the regression effects (see also Rodriguez and Goldman, 1995), especially for the intercept parameter (see Sutradhar and Qu, 1998). The standard errors of the MQL estimators are, in general, larger than the corresponding PQL estimators for all the regression parameters. Most importantly, both the PQL and MQL approaches produces highly biased estimates for the variance of the random effects, with the MQL approach being worse between the two.

As opposed to the approximate MQL approach (Zeger et al., 1988; Breslow and Clayton, 1993), there also exists an exact MQL approach (Sutradhar and Rao, 2001) which exploits the correct covariance structure in constructing the estimating equations. Note that the exact MQL approach proposed by Sutradhar and Rao (2001) is however developed only for small values of the variance of the random effects. Recently, this MQL approach has been improved by Sutradhar (2004) where the covariance matrix needed for the construction of the estimating equation has been computed for any small or large values of the variance of the random effects. This approach has been referred to as exact quasi-likelihood, or generalized quasi-likelihood (GQL), approach.

As opposed to the PQL approach of Breslow and Clayton (1993), Jiang (1998) proposed the traditional method of moments (MM) for the estimation of regression effects and variance component, where unconditional first and second order moments are computed by using a simulation approach for numerical integration. However, the MM approach does not yield efficient estimates for the parameters of the mixed model, in particular for the variance components of the model. Jiang and Zhang (2001) have attempted to improve the efficiencies of the MM based estimators of Jiang (1998). Sutradhar (2004) has however shown that the improved method of moments (IMM) by Jiang and Zhang (2001) can also be highly inefficient compared to the generalized quasi-likelihood (GQL) approach constructed by taking the familial correlation into account correctly.

Similar to the PQL approach by Breslow and Clayton (1993), Lee and Nelder (1996) proposed a hierarchical likelihood (HL) approach. Lee and Nelder (1996) used the hierarchical likelihood function to estimate the regression effects as well as random effects considering them as fixed effects and then by using these estimates to obtained the estimate of the variance component (see also Lee and Nelder, 2001). Even though many authors have examined the performance of the PQL approach, there does not appear to be any comparative studies with the HL approach of Lee and Nelder (1996). Since the PQL approach has consistency problems, and IMM is inefficient compared to the GQL approach, it remains to be seen how the HL approach performs compared to the GQL approach.

1.2 Objective of The Thesis

The discussion in the previous section indicates that the PQL approach suffers from an inefficiency problem especially for estimating the variance component. It is also clear that the GQL approach (Sutradhar, 2004) performs better among existing competing approaches such as the IMM approach of Jiang and Zhang (2001). However the GQL approach was not compared by Sutradhar (2004) with the existing HL approach which is another widely used competitive technique. These reasons motivated us to conduct a comparative study between the GQL and the HL approaches in the context of clustered count data analysis, which is an important application of GLMM's.

More specifically, in Chapter 2, we describe the Poisson mixed model and review its basic properties. We derive the exact likelihood function and note its complexity for the estimation of the parameters. We also provide comments on the PQL and IMM approaches. In Chapter 3, we discuss the GQL and HL approaches and provide the necessary estimating equations for the regression effects as well as variance component of the random effects. In Chapter 4 we conduct an extensive simulation study to examine the relative performance of the GQL and HL approaches. In Chapter 5 we provide an illustrative example of the methodologies that we have explained in Chapter 3. Finally, we conclude the thesis with some remarks in Chapter 6.

Chapter 2

Poisson-Normal Mixed Models

Even though the Poisson-Normal mixed model is well studied in the literature, we also review this model here for the convenience of describing the GQL and HL approaches in the next chapter. As far as the basic properties of this model are concerned, we provide all possible moments up to order four. The product moments of the third and fourth orders are required useful for the construction of the GQL approach.

In this chapter, we also discuss the exact likelihood properties of this model for the purpose of estimation. Some comments about the moments based, such as IMM estimation, and approximate likelihood based approaches such as PQL estimation, are also given.

2.1 Poisson Mixed Models

Suppose that y_{ij} is the count response variable for the jth (j=1,....,n_i) individual in the ith (i=1,...,K) cluster (e.g. family). Also suppose that x_{ij} is the p×1 vector of fixed covariates and β is the effect of x_{ij} on y_{ij}. Note that as n_i members belongs to the ith cluster, the count response from these members of the ith cluster will be correlated. This is because they are likely to share a common familial/cluster effect, say γ_i. Here the responses conditional on γ_i are independent.

Suppose that conditional on γ_i, Y_{ij} has a Poisson distribution with density function

$$f(y_{ij} \mid \gamma_i) = \frac{\mu_{ij}^{* \, y_{ij}} \, e^{-\mu_{ij}^*}}{y_{ij}!} \tag{2.1}$$

where $E\,(y_{ij} \mid \gamma_i) = \mu_{ij}^* = exp\,(x'_{ij}\beta + z_i\gamma_i)$. Here z_i is a random effects related known covariate for the ith cluster.

Since responses in a given cluster are conditionally independent, the conditional likelihood of these responses is

$$L_i\,(\beta, \gamma_i) = \prod_{j=1}^{n_i} f(y_{ij}|\gamma_i) \tag{2.2}$$

Under the assumption that $\gamma_i \overset{iid}{\sim} (0, \sigma^2)$, the unconditional likelihood function is

$$\begin{aligned} L\,(\beta, \sigma^2) &= \prod_{i=1}^{K} \int L_i\,(\beta, \gamma_i) \\ &= \prod_{i=1}^{K} \prod_{j=1}^{n_i} \int f(y_{ij}|\gamma_i)\,\phi(\gamma_i)\,d\gamma_i \end{aligned} \tag{2.3}$$

where $\phi(\gamma_i)$ is the probability density for the random effects γ_i.

In general, it is assumed that γ_i in (2.1) follows the normal distribution. For example, we can see Breslow and Clayton (1993), Jiang (1998) and Sutradhar (2004). The unconditional likelihood in (2.3) then takes the form

$$L\,(\beta, \sigma^2) = \prod_{i=1}^{K} \prod_{j=1}^{n_i} \int f(y_{ij}|\gamma_i)\,\phi^*(\gamma_i)\,d\gamma_i \tag{2.4}$$

with $\phi^*(\gamma_i)$ as the normal density with mean 0 and variance σ^2. The likelihood estimation of β and σ^2 will be discussed later in section 2.2. For now we concentrate on the basic moment properties of the mixed model (2.4).

2.1.1 Basic Properties of the Model: Moments up to Order Four

For the Poisson-Normal mixed model discussed in the last section we can derive the marginal and product moments up to order four as in the following lemmas.

Lemma 2.1. *For j=1,...,n_i, the first order marginal moment of Y_{ij} is*

$$E\,(Y_{ij}) = exp\,(x'_{ij}\beta + \frac{1}{2}\, z_i^2\, \sigma^2) = m_{ij} \tag{2.5}$$

Lemma 2.2. *For $j \neq k$, j,k=1,...,n_i, the second order marginal and product moments of Y_{ij} and Y_{ik} are*

$$E\,(Y_{ij}^2) = m_{ij} + m_{ij}^2\, e^{z_i^2\, \sigma^2} \tag{2.6}$$

$$E\,(Y_{ij}\, Y_{ik}) = m_{ij}\, m_{ik}\, e^{z_i^2\, \sigma^2} \tag{2.7}$$

Lemma 2.3. *For $j \neq k \neq l$, j,k,l=1,...,n_i, the third order marginal and product moments of Y_{ij}, Y_{ik} and Y_{il} are*

$$E\,(Y_{ij}^3) = m_{ij} + 3m_{ij}^2\, e^{z_i^2\, \sigma^2} + m_{ij}^3\, e^{3z_i^2\, \sigma^2} \tag{2.8}$$

$$E\,(Y_{ij}^2\, Y_{ik}) = m_{ij}\, m_{ik}\, e^{z_i^2\, \sigma^2} + m_{ij}^2\, m_{ik}\, e^{3z_i^2\, \sigma^2} \tag{2.9}$$

$$E\,(Y_{ij}\, Y_{ik}^2) = m_{ij}\, m_{ik}\, e^{z_i^2\, \sigma^2} + m_{ij}\, m_{ik}^2\, e^{3z_i^2\, \sigma^2} \tag{2.10}$$

$$E\,(Y_{ij}\, Y_{ik}\, Y_{il}) = m_{ij}\, m_{ik}\, m_{il}\, e^{3z_i^2\, \sigma^2} \tag{2.11}$$

Lemma 2.4. *For $j \neq k \neq l \neq m$, $j,k,l,m=1,\ldots,n_i$, the fourth order marginal and product moments of Y_{ij}, Y_{ik}, Y_{il} and Y_{im} are given by*

$$E\,(Y_{ij}^4) = m_{ij} + 7m_{ij}^2\, e^{z_i^2\,\sigma^2} + 6m_{ij}^3\, e^{3z_i^2\,\sigma^2} + m_{ij}^4\, e^{6z_i^2\,\sigma^2} \tag{2.12}$$

$$E\,(Y_{ij}^3\, Y_{ik}) = m_{ij}\, m_{ik}\, e^{z_i^2\,\sigma^2} + 3\, m_{ij}^2\, m_{ik}\, e^{3z_i^2\,\sigma^2} + m_{ij}^3\, m_{ik}\, e^{6z_i^2\,\sigma^2} \tag{2.13}$$

$$E\,(Y_{ij}\, Y_{ik}^3) = m_{ij}\, m_{ik}\, e^{z_i^2\,\sigma^2} + 3\, m_{ij}\, m_{ik}^2\, e^{3z_i^2\,\sigma^2} + m_{ij}\, m_{ik}^3\, e^{6z_i^2\,\sigma^2} \tag{2.14}$$

$$E(Y_{ij}^2\, Y_{ik}^2) = m_{ij}m_{ik}e^{z_i^2\,\sigma^2} + m_{ij}^2 m_{ik}e^{3z_i^2\,\sigma^2} + m_{ij}m_{ik}^2 e^{3z_i^2\,\sigma^2} + m_{ij}^2 m_{ik}^2 e^{6z_i^2\,\sigma^2} \tag{2.15}$$

$$E\,(Y_{ij}^2\, Y_{ik}\, Y_{il}) = m_{ij}\, m_{ik}\, m_{il}\, e^{3z_i^2\,\sigma^2} + m_{ij}^2\, m_{ik}\, m_{il}\, e^{6z_i^2\,\sigma^2} \tag{2.16}$$

$$E\,(Y_{ij}\, Y_{ik}^2\, Y_{il}) = m_{ij}\, m_{ik}\, m_{il}\, e^{3z_i^2\,\sigma^2} + m_{ij}\, m_{ik}^2\, m_{il}\, e^{6z_i^2\,\sigma^2} \tag{2.17}$$

$$E\,(Y_{ij}\, Y_{ik}\, Y_{il}^2) = m_{ij}\, m_{ik}\, m_{il}\, e^{3z_i^2\,\sigma^2} + m_{ij}\, m_{ik}\, m_{il}^2\, e^{6z_i^2\,\sigma^2} \tag{2.18}$$

$$E\,(Y_{ij}\, Y_{ik}\, Y_{il}\, Y_{im}) = m_{ij}\, m_{ik}\, m_{il}\, m_{im}\, e^{6z_i^2\,\sigma^2} \tag{2.19}$$

Proof. Since Y_{ij} conditionally follows the Poisson distribution with mean μ_{ij}^* we can write the first four conditional moments of this distribution as:

$$E\,(Y_{ij} \mid \gamma_i) = \mu_{ij}^* = exp\,(x_{ij}'\beta + z_i\,\gamma_i) \tag{2.20}$$

$$E\,(Y_{ij}^2 \mid \gamma_i) = \mu_{ij}^* + {\mu_{ij}^*}^2 \tag{2.21}$$

$$E\,(Y_{ij}^3 \mid \gamma_i) = \mu_{ij}^* + 3{\mu_{ij}^*}^2 + {\mu_{ij}^*}^3 \tag{2.22}$$

$$E\,(Y_{ij}^4 \mid \gamma_i) = \mu_{ij}^* + 7{\mu_{ij}^*}^2 + 6{\mu_{ij}^*}^3 + {\mu_{ij}^*}^4 \tag{2.23}$$

Because the random effects, $\gamma_i \overset{iid}{\sim} N(0,\sigma^2)$ and z_i are known, we can write

$$\begin{aligned} E\,(Y_{ij}) &= E_{\gamma_i}\, E\,(Y_{ij} \mid \gamma_i) = E_{\gamma_i}\,(\mu_{ij}^*) \\ &= exp\,(x_{ij}'\beta + \frac{1}{2}\, z_i^2\,\sigma^2) = m_{ij}, \end{aligned}$$

which is the marginal moment of first order as in Lemma 2.1.

Now by exploiting the conditional second order moments (2.21) one obtains the unconditional second order marginal moments as

$$\begin{aligned} E\,(Y_{ij}^2) &= E_{\gamma_i}\,E\,(Y_{ij}^2 \mid \gamma_i) \;=\; E_{\gamma_i}\,(\,\mu_{ij}^* + \mu_{ij}^{*\,2}) \\ &= m_{ij} + m_{ij}^2\,e^{z_i^2\;\sigma^2}. \end{aligned}$$

Similarly, the unconditional product moments of order 2 given in (2.7), can be computed as

$$\begin{aligned} E\,(Y_{ij}\,Y_{ik}) &= E_{\gamma_i}\,E\,(Y_{ij}\;Y_{ik} \mid \gamma_i) = E_{\gamma_i}\;\{E\,(Y_{ij} \mid \gamma_i)\;E\,(Y_{ik} \mid \gamma_i)\} \\ &= E_{\gamma_i}(\,\mu_{ij}^*\,\mu_{ik}^*) = \;m_{ij}\,m_{ik}\,e^{z_i^2\,\sigma^2}. \end{aligned}$$

To derive all third order moments in Lemma 2.3, we first compute the marginal third order moments as

$$\begin{aligned} E\,(Y_{ij}^3) &= \;E_{\gamma_i}\,E\,(Y_{ij}^3 \mid \gamma_i) \;=\; E_{\gamma_i}\,(\mu_{ij}^* + \;3\mu_{ij}^{*\,2} + \mu_{ij}^{*\,3}) \\ &= \;m_{ij} + \,3m_{ij}^2\,e^{z_i^2\;\sigma^2} + \,m_{ij}^3\,e^{3z_i^2\;\sigma^2}, \end{aligned}$$

which is (2.8). Next, the product moments of order 3 are computed as

$$\begin{aligned} E\,(Y_{ij}^2\,Y_{ik}) &= \;E_{\gamma_i}\,E\;(Y_{ij}^2\,Y_{ik} \mid \gamma_i) = \;E_{\gamma_i}\,\{E\,(Y_{ij}^2 \mid \gamma_i)\;E\,(Y_{ik} \mid \gamma_i)\} \\ &= E_{\gamma_i}\{\,(\mu_{ij}^*\; + \mu_{ij}^{*\,2})\mu_{ik}^*\} \\ &= \;m_{ij}\,m_{ik}\,e^{z_i^2\,\sigma^2} + \;m_{ij}^2\,m_{ik}\,e^{3z_i^2\,\sigma^2}, \end{aligned}$$

$$E\,(Y_{ij}\,Y_{ik}^2) = \;m_{ij}\,m_{ik}\,e^{z_i^2\,\sigma^2} + m_{ij}\;m_{ik}^2\,e^{3z_i^2\,\sigma^2},$$

and

$$\begin{aligned} E\,(Y_{ij}\,Y_{ik}\,Y_{il}) &= \;E_{\gamma_i}\;E\,(Y_{ij}\,Y_{ik}\,Y_{il} \mid \gamma_i) \\ &= \;E_{\gamma_i}\,\{E\,(Y_{ij} \mid \gamma_i)\;E\,(Y_{ik} \mid \gamma_i)\;E\,(Y_{il} \mid \gamma_i)\} \\ &= E_{\gamma_i}(\,\mu_{ij}^*\;\mu_{ik}^*\,\mu_{il}^*) \\ &= \;m_{ij}\,m_{ik}\,m_{il}\,e^{3z_i^2\,\sigma^2}. \end{aligned}$$

By using computations similar to those of Lemma 2.3, one may obtain the results of Lemma 2.4 for the fourth order moments. The formulas for the marginal as well as product moments of order four are

$$\begin{aligned} E(Y_{ij}^4) &= E_{\gamma_i} E(Y_{ij}^4 \mid \gamma_i) = E_{\gamma_i}(\mu_{ij}^* + 7\mu_{ij}^{*\,2} + 6\mu_{ij}^{*\,3} + \mu_{ij}^{*\,4}) \\ &= m_{ij} + 7m_{ij}^2 e^{z_i^2\,\sigma^2} + 6m_{ij}^3 e^{3z_i^2\,\sigma^2} + m_{ij}^4 e^{6z_i^2\,\sigma^2}, \end{aligned}$$

$$\begin{aligned} E(Y_{ij}^3 Y_{ik}) &= E_{\gamma_i} E(Y_{ij}^3 Y_{ik} \mid \gamma_i) = E_{\gamma_i}\{E(Y_{ij}^3 \mid \gamma_i)\, E(Y_{ik} \mid \gamma_i)\} \\ &= E_{\gamma_i}\{(\mu_{ij}^* + 3\mu_{ij}^{*\,2} + \mu_{ij}^{*\,3})\mu_{ik}^*\} \\ &= m_{ij}\, m_{ik}\, e^{z_i^2\,\sigma^2} + 3\, m_{ij}^2\, m_{ik}\, e^{3z_i^2\,\sigma^2} + m_{ij}^3\, m_{ik}\, e^{6z_i^2\,\sigma^2}, \end{aligned}$$

$$E(Y_{ij}\, Y_{ik}^3) = m_{ij}\, m_{ik}\, e^{z_i^2\,\sigma^2} + 3\, m_{ij}\, m_{ik}^2\, e^{3z_i^2\,\sigma^2} + m_{ij}\, m_{ik}^3\, e^{6z_i^2\,\sigma^2},$$

$$\begin{aligned} E(Y_{ij}^2 Y_{ik}^2) &= E_{\gamma_i} E(Y_{ij}^2 Y_{ik}^2 \mid \gamma_i) = E_{\gamma_i}\{E(Y_{ij}^2 \mid \gamma_i)\, E(Y_{ik}^2 \mid \gamma_i)\} \\ &= E_{\gamma_i}\{(\mu_{ij}^* + \mu_{ij}^{*\,2})\,(\mu_{ik}^* + \mu_{ik}^{*\,2})\} \\ &= m_{ij}m_{ik}e^{z_i^2\,\sigma^2} + m_{ij}^2 m_{ik}e^{3z_i^2\,\sigma^2} + m_{ij}m_{ik}^2 e^{3z_i^2\,\sigma^2} + m_{ij}^2 m_{ik}^2 e^{6z_i^2\,\sigma^2}, \end{aligned}$$

$$\begin{aligned} E(Y_{ij}^2\, Y_{ik}\, Y_{il}) &= E_{\gamma_i}\, E(Y_{ij}^2\, Y_{ik}\, Y_{il} \mid \gamma_i) \\ &= E_{\gamma_i}\{E(Y_{ij}^2 \mid \gamma_i)\, E(Y_{ik} \mid \gamma_i)\, E(Y_{il} \mid \gamma_i)\} \\ &= E_{\gamma_i}\{(\mu_{ij}^* + \mu_{ij}^{*\,2})\, \mu_{ik}^*\, \mu_{il}^*\} \\ &= m_{ij}\, m_{ik}\, m_{il}\, e^{3z_i^2\,\sigma^2} + m_{ij}^2\, m_{ik}\, m_{il}\, e^{6z_i^2\,\sigma^2}, \end{aligned}$$

$$E(Y_{ij}\, Y_{ik}^2\, Y_{il}) = m_{ij}\, m_{ik}\, m_{il}\, e^{3z_i^2\,\sigma^2} + m_{ij}\, m_{ik}^2\, m_{il}\, e^{6z_i^2\,\sigma^2},$$

$$E(Y_{ij}\, Y_{ik}\, Y_{il}^2) = m_{ij}\, m_{ik}\, m_{il}\, e^{3z_i^2\,\sigma^2} + m_{ij}\, m_{ik}\, m_{il}^2\, e^{6z_i^2\,\sigma^2},$$

and

$$\begin{aligned} E\,(Y_{ij}\,Y_{ik}\,Y_{il}\,Y_{im}) &= E_{\gamma_i}\,E\,(Y_{ij}\,Y_{ik}\,Y_{il}\,Y_{im} \mid \gamma_i) \\ &= E_{\gamma_i}(\,\mu^*_{ij}\,\mu^*_{ik}\,\mu^*_{il}\,\mu^*_{im}) \\ &= m_{ij}\,m_{ik}\,m_{il}\,m_{im}\,e^{6z_i^2\,\sigma^2}. \end{aligned}$$

The above results from Lemma 2.1 to 2.4 may be used to construct the mean, variance, covariance and other corrected moments up to order 4. In particular, the mean and the variance of a count response under the Poisson-normal mixed model are given by

$$E\,(Y_{ij}) = exp\,(x'_{ij}\beta\, +\, \frac{1}{2}\,z_i^2\,\sigma^2)\, = m_{ij}, \tag{2.24}$$

and

$$\begin{aligned} V\,(Y_{ij}) &= E\,(Y_{ij}^2) -\,\{E\,(Y_{ij})\}^2 \\ &= m_{ij} + m_{ij}^2\,(e^{z_i^2\,\sigma^2}\, - 1), \end{aligned} \tag{2.25}$$

respectively. By similar calculation as in (2.25), the covariance between the jth and kth (j$\neq$ k) individuals in the ith cluster is

$$\begin{aligned} Cov\,(Y_{ij},\,Y_{ik}) &= E\,(Y_{ij}\,Y_{ik})\,\, -\, E\,(Y_{ij})\,E\,(Y_{ik}) \\ &= m_{ij}\,m_{ik}\,(e^{z_i^2\,\sigma^2}\, - 1). \end{aligned} \tag{2.26}$$

The mean (2.24), variance (2.25) and covariance (2.26) are all functions of β and σ^2. Furthermore, it follows from (2.25) that as σ^2 increases the variance of the data increases exponentially. Consequently, the σ^2 parameter is referred to as the over-dispersion parameter. The primary objective of the analysis is to estimate the β and σ^2 parameters consistently and efficiently by using the available familial data.

As mentioned earlier, in order to derive consistent and efficient estimates for β and σ^2, different estimation techniques have been used in the literature. The moments up to order four that we have presented above may be exploited for such estimation. Two of the moments based estimation techniques, such as IMM and GQL will be discussed in this thesis. The IMM approach is briefly discussed in Section 2.3.1, and the GQL and the HL approaches will be given in the next chapter.

For non-moments based estimation approaches, in the following section, we examine the complexity involved in the exact likelihood approach. We then discuss the PQL approximation in Section 2.3.2 and HL approximation in the next chapter.

2.2 Complexity of Exact Likelihood Estimation for Poisson Mixed Models

In generalized linear mixed model, the exact likelihood function can be written as

$$L = \int_{-\infty}^{\infty} \prod_{i=1}^{K} \prod_{j=1}^{ni} f(y_{ij}|\gamma_i)\, \phi(\gamma_i)\, d\gamma_i. \tag{2.27}$$

Since $f(y_{ij}|\gamma_i)$ is given by (2.1), and under normality for the random effects, the density $\phi(\gamma_i)$ is the same as $\phi^*(\gamma_i)$,

$$\phi^*(\gamma_i) = \frac{1}{\sqrt{2\Pi\sigma^2}}\, exp\,(-\frac{\gamma_i^2}{2\sigma^2}) \tag{2.28}$$

The conditional Poisson model (2.1) can be written as

$$f\,(y_i|\gamma_i^*) = \frac{1}{\prod_{j=1}^{ni} y_{ij}!}\, exp\left(\sum_{j=1}^{ni} (x'_{ij}\beta + \sigma z_i \gamma_i^*)\, y_{ij} \;-\; exp(x'_{ij}\beta + \sigma z_i \gamma_i^*\,)\right) \tag{2.29}$$

where $\gamma_i^* = \frac{\gamma_i}{\sigma} \overset{iid}{\sim} N(0,1)$ as $\gamma_i \overset{iid}{\sim} N(0,\sigma^2)$.

Then by using (2.28) the likelihood function in (2.27) is equivalent to

$$L(\beta,\sigma^2) \propto \int_{-\infty}^{\infty} \prod_{i=1}^{K} f\left(y_i|\gamma_i^*\right) \phi^*(\gamma_i^*)\, d\gamma_i^*. \tag{2.30}$$

But, unfortunately, the integral in (2.30) does not have an analytic solution. One may however compute this likelihood function numerically. For example, Fahrmeir and Tutz (1994) and Jiang (1998) suggested to simulate γ_{iw}^* (w=1,2...,M) for a huge M say M = 10,000, from the standard normal distribution and compute the likelihood as

$$L(\beta,\sigma^2) \simeq \frac{1}{M}\sum_{w=1}^{M}\left[\prod_{i=1}^{K} f\left(y_i|\gamma_{iw}^*\right)\right] \tag{2.31}$$

Note however that when likelihood estimating equations for β and σ^2 are written following (2.31), their solutions will still be numerically cumbersome. As a remedy, there exists several approximations to this likelihood approach, such as PQL, HL and GQL. Since these approaches are relatively simpler than the above exact likelihood approach, we concentrate on such simpler techniques in the present thesis.

2.2.1 Effects of Ignoring Random Effects $\{\gamma_i\}$ in β Estimation

When we ignore random effects $\{\gamma_i\}$, observations in a cluster become independent. One, consequently, can use the simpler version of (2.30) to obtain the likelihood estimates for the regression effects β. This is because, when random effects $\{\gamma_i\}$ consequently $\sigma^2 = 0$, the mixed model (2.30) reduces to the fixed model. The likelihood function for this well known fixed effects model is then given by

$$L = \prod_{i=1}^{K} \prod_{j=1}^{n_i} f(y_{ij})$$
$$= \prod_{i=1}^{K} \prod_{j=1}^{n_i} \frac{\mu_{ij}{}^{y_{ij}} \, e^{-\mu_{ij}}}{y_{ij}!} \tag{2.32}$$

where $\mu_{ij} = exp(x'_{ij}\beta)$.

This computational simplicity for the independence case raises an issue to examine the effects of various values of σ^2 in the estimation of β when $\sigma^2 = 0$ is used in estimation. In this subsection we examine this issue through a simulation study.

For the purpose of computation of the maximum likelihood estimate of β when $\sigma^2 = 0$ by (2.32), we first write the likelihood estimating equation as

$$\frac{\partial logL}{\partial \beta} = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (y_{ij}x_{ij} - \mu_{ij}x_{ij}) = 0, \tag{2.33}$$

which can be written in matrix and vector notation as

$$\sum_{i=1}^{K} X_i' \, (y_i - \mu_i) = 0, \tag{2.34}$$

where $y_i = [y_{i1}, y_{i2}, ..., y_{in_i}]'$ and $\mu_i = E[Y_i] = [\mu_{i1}, \mu_{i2}, ..., \mu_{in_i}]'$, with

$$\mu_{ij} = E(Y_{ij}) = exp(x'_{ij}\beta), \tag{2.35}$$

and

$$X_i = \begin{bmatrix} x_{i11} & x_{i12} & \cdots & x_{i1p} \\ x_{i21} & x_{i22} & \cdots & x_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{in_i1} & x_{in_i2} & \cdots & x_{in_ip} \end{bmatrix}_{n_i \times p} .$$

The estimating equation (2.34) may be easily solved by using the iterative algorithm

$$\hat{\beta}_{(r+1)} = \hat{\beta}_{(r)} + \left[\left(\sum_{i=1}^{K} X_i' A_i X_i' \right)^{-1} \sum_{i=1}^{K} X_i' \left(y_i - \mu_i \right) \right]_{\beta=\hat{\beta}_{(r)}}, \qquad (2.36)$$

where the quantity within the square bracket is evaluated at $\beta = \hat{\beta}_{(r)}$, r being the rth iteration and

$$A_i = \begin{bmatrix} m_{i1} & 0 & \cdots & 0 \\ 0 & m_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_{in_i} \end{bmatrix}_{n_i \times n_i}.$$

We remark that the likelihood estimating equation (2.34) constructed under the independence assumption is in fact the same as the traditional moment estimating equation for β (see Section 2.3.1). Furthermore, the likelihood equation (2.34) is also the same as the well known quasilikelihood (QL) estimating equation. This equivalence will be clear from the next chapter where the GQL estimating equation is developed for the regression effects β in the presence of σ^2.

2.2.2 Regression Estimation when Random Effects $\{\gamma_i\}$ are Ignored: A Simulation Study

In the simulation study, we generate the data by using the Poisson-normal mixed model (2.1) with

$$log\, E\left(Y_{ij}|\gamma_i\right) = \beta_1\, x_{ij1} + \beta_2\, x_{ij2} + \sigma\, \gamma_i^* \qquad (2.37)$$

where $\gamma_i^* \stackrel{iid}{\sim} N(0,1)$, but estimate β_1 and β_2 under the assumption that $\sigma^2 = 0$, i.e, $\gamma_i^* = 0$ for all i = 1, ..., K which in practice means that γ_i has been ignored. We choose the number of cluster, K=100, for convenience. We also consider small or large cluster sizes such as $n_i = 2$ and 4 for all i=1,2,...,K, where n_i is the cluster size.

To use (2.37), we consider the true values of the regression parameters as $\beta_1 =$ 1.0 and $\beta_2 = 1.0$, and select two covariates as follows

$$x_{ij1} = \begin{cases} 1 & \text{for } j = 1, 2, \ldots, n_i/2; \quad i = 1, 2, \ldots, K/2 \\ 0 & \text{for } j = n_i/2 + 1, \ldots, n_i; \quad i = 1, 2, \ldots, K/2 \\ 1 & \text{for } j = 1, \ldots, n_i; \quad i = K/2 + 1, \ldots, K \end{cases}$$

$$x_{ij2} = \begin{cases} 1 & \text{for } j = 1, 2, \ldots, n_i/2; \quad i = 1, 2, \ldots, K/2 \\ 2 & \text{for } j = n_i/2 + 1, \ldots, n_i; \quad i = 1, 2, \ldots, K/2 \\ 0 & \text{for } j = 1, 2, \ldots, n_i/2; \quad i = K/2 + 1, \ldots, K \\ 1 & \text{for } j = n_i/2 + 1, \ldots, n_i; \quad i = K/2 + 1, \ldots, K \end{cases}$$

We choose various small and large values of σ^2, namely, $\sigma^2 = 0.2$. 0.6, 1.0, 1.5, 2.0 and 2.5. Next, under each of 1000 simulations, we use the responses as generated above, as well as the covariates x_{ij1} and x_{ij2}, to obtain the likelihood estimates of β_1 and β_2 by using (2.36). The simulated means (SM) of the the likelihood estimates of β_1 and β_2 along with their simulated standard errors (SSE) are reported in Table 2.1. We also have computed the simulated relative bias (SRB) for the estimates defined by

$$\text{SRB} = \frac{|\text{SM} - \text{True parameter value}|}{\text{SSE}} \times 100.$$

Table 2.1: *The SM, SSE and SRB of the likelihood estimates of* β_1 *and* β_2 *when random effects are ignored in estimation but the data were generated for selected values of* $\sigma^2 > 0$*; K=100; n = 2 and 4; true values of the regression parameters:* $\beta_1 = 1.0$ *and* $\beta_2 = 1.0$*, 1000 simulations.*

Variance of the random effects, σ^2	**Statistic**	**Cluster size 2** β_1	**Cluster size 2** β_2	**Cluster size 4** β_1	**Cluster size 4** β_2
0.20	SM	1.0636	1.0608	1.0625	1.0619
	SSE	0.0387	0.0238	0.0272	0.0170
	SRB	164	255	230	364
0.60	SM	1.1607	1.1816	1.1623	1.1823
	SSE	0.0353	0.0216	0.0244	0.0151
	SRB	455	841	665	1207
1.00	SM	1.2512	1.3012	1.2530	1.3001
	SSE	0.0311	0.0188	0.0219	0.0131
	SRB	808	1602	1155	2291
1.50	SM	1.3516	1.4605	1.3534	1.4604
	SSE	0.0290	0.0184	0.0202	0.0127
	SRB	1212	2503	1750	3625
2.00	SM	1.4085	1.6771	1.4087	1.6778
	SSE	0.0311	0.0283	0.0233	0.0201
	SRB	1314	2393	1754	3372
2.50	SM	1.5369	1.7416	1.5382	1.7416
	SSE	0.0227	0.0131	0.0157	0.0095
	SRB	2365	5661	3428	7806

It is clear from Table 2.1 that when σ^2 is close to zero, the likelihood estimates of β_1 and β_2 appears to be very close to the true parameter values. This is expected as the estimation is done by using the 'working' independence assumption, that is, σ^2 =0. As the value of σ^2 however increases the estimates of β_1 and β_2 appear to deviate more and more from the true parameter values. For example, when $\sigma^2 = 0.2$ and n = 2, in estimating the true parameters $\beta_1 = \beta_2 = 1.0$, the independence assumption based likelihood estimates are found to be $\hat{\beta}_1 = 1.0636$ with standard error 0.0387 and percentage relative bias 164, and $\hat{\beta}_2 = 1.0608$ with standard error 0.0238 and percentage relative bias 255. For a large value of σ^2, such as $\sigma^2 = 2.5$, the estimates are found to be $\hat{\beta}_1 = 1.5369$ with standard error 0.0227 and percentages of relative bias 2365, and $\hat{\beta}_2 = 1.7416$ with standard error 0.0131 with percentages of relative bias 5661. It is clear that $\sigma^2 = 2.5$, β_1 and β_2 are estimated with huge biases. Thus, the independence assumption based likelihood approach fails to estimate the parameters adequately. When cluster size increases, the estimates appear to become much more biased. For example, when n = 4 for the same $\sigma^2 = 2.5$, the percentage relative biases of $\hat{\beta}_1$ and $\hat{\beta}_2$ are 3428 and 7806, whereas for n = 2, they were 2365 and 5661 respectively.

The simulation demonstrate clearly that there is a detrimental effect of ignoring the presence of the random effects even when one is only interested to estimate the main regression effects. This definitely motivates one to estimate the regression effects by taking the random effects into account i.e. removing the assumption $\sigma^2 = 0$. In the next section we discuss the simultaneous estimation of these parameters, namely, β and σ^2.

2.3 Approximation to Likelihood Inference

The existing leading approximations to the (complex) likelihood approach are as follows:

(1) Improved method of moments (IMM).

(2) Penalized quasi-likelihood (PQL) approach.

(3) Hierarchical likelihood (HL) approach.

(4) Generalized quasi-likelihood (GQL) approach.

The advantages and disadvantages of the PQL and the IMM approaches are discussed extensively in the literature. For further discussion on the PQL approach we refer to Breslow and Clayton (1993), Breslow and Lin (1995) and Sutradhar and Qu (1998), among others. For additional discussion on the IMM approach we refer to Jiang and Zhang (2001) and Sutradhar (2004), among others. Nevertheless, for the sake of completeness we provide the estimation formulas under these two approaches in the following subsections. The other two approaches (GQL and HL) will be considered in Chapter 3.

2.3.1 Improved Method of Moments (IMM)

Jiang and Zhang (2001) presented the improved method of moments (IMM) as an improvement over the method of moments (MM) discussed by Jiang (1998). Since, conditional on the random effects, y_{ij}^2; j=1,...,n_i, and $y_{ij}y_{ik}$; j<k =2,...,n_i, may be shown as sufficient statistics for the parameters of the model. Jiang and Zhang (2001) used them and wrote a base statistic

$$S = \sum_{i=1}^{N} \sum_{j=1}^{n_i} y_{ij}^2 + \sum_{i=1}^{N} \sum_{j<k}^{n_i} y_{ij} y_{ik} \tag{2.38}$$

for the estimation of $\theta = (\beta, \sigma)'$, where, N = $\sum_{i=1}^{K} n_i$ and compute its expectation given by $\mu(\theta) = E(S)$. For U = $\partial\mu'(\theta)/\partial\theta$ and V = Cov(S), these authors then solved the estimating equation

$$B\,[\,S\; - \;\mu(\theta)] = 0 \tag{2.39}$$

where B = $U'V^{-1}$. Note that as the computation of V matrix is complicated, Jiang and Zhang (2001) proposed a simple matrix which is free from higher order moments such as B = $B_0 = diag(I_1, 1'_K)$ and solved the estimating equation

$$B_0\,[\,S\; - \;\mu(\theta)] = 0, \tag{2.40}$$

for β and σ. Sutradhar (2004) has however demonstrated that the use of B_0 (free from higher order moments) actually may lead to inefficient estimates.

2.3.2 Penalized Quasi-Likelihood (PQL)

Breslow and Clayton (1993) used the Penalized quasi-likelihood (PQL) estimation approach as an approximation to the exact likelihood approach. The PQL approach can be summarized by the following two steps;

Step 1 : By assuming σ^2 known, the regression effects β and the random effects γ_i (pretended to be fixed effects) are jointly estimated by maximizing a penalized quasi-likelihood function. To be specific, the penalized quasi-likelihood function is given by

$$ql(\beta, \sigma^2, \tilde{\gamma}) = -\frac{1}{2} \sum_{i=1}^{K} log \left(1 + \sigma^2 \sum_{j=1}^{n_i} exp(x'_{ij}\beta + \tilde{\gamma}_i)\right) - \sum_{i=1}^{K} h(\tilde{\gamma}_i) \qquad (2.41)$$

where $\tilde{\gamma} = (\tilde{\gamma}_1, ..., \tilde{\gamma}_i, ..., \tilde{\gamma}_K)'$ with $\tilde{\gamma}_i$ as the posterior mode of γ_i computed from $\partial h(\gamma_i)/\partial\gamma_i = 0$, $h(\gamma_i)$ being given by;

$$h(\gamma_i) = -\sum_{j=1}^{n_i} y_{ij}(x'_{ij}\beta + \gamma_i) + \sum_{j=1}^{n_i} exp(x'_{ij}\beta + \gamma_i) + \frac{\gamma_i^2}{2\sigma^2}. \qquad (2.42)$$

The maximization of the PQL function (2.42) with respect to β and γ_i is achieved by solving the estimating equations

$$g_1(\beta, \gamma) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \{y_{ij} - exp(x'_{ij}\beta + \gamma_i)\} \, x_{ij} = 0 \qquad (2.43)$$

and

$$g_2(\beta, \gamma_i, \sigma^2) = \sum_{j=1}^{n_i} \{y_{ij} - exp(x'_{ij}\beta + \gamma_i)\} - \frac{\gamma_i}{\sigma^2} = 0, \qquad (2.44)$$

where $g_1(.)$ and $g_2(.)$ are obtained by taking the derivatives of the (2.42) with respect to β and γ_i respectively. Let $\hat{\beta}$ denotes the estimate of β and $\hat{\gamma}_i$ denotes the estimate of γ_i.

Step 2 : Then by using $\hat{\beta}$ and $\hat{\gamma}_i$ obtained from Step 1 they constructed a profile quasi-likelihood function in the form of a working normal likelihood function with correct mean and covariance and obtained the restricted maximum-likelihood estimate of σ^2. The profile quasi-likelihood based score equation for σ^2 is given by

$$\begin{aligned} g_3(\hat{\beta}, \sigma^2, \hat{\gamma}_i) &= \frac{\partial ql(\hat{\beta}, \sigma^2, \hat{\gamma})}{\partial \sigma^2} \\ &= \sum_{i=1}^{K} \hat{\gamma_i}^2 - \sigma^4 \sum_{i=1}^{K} \frac{\sum_{j=1}^{n_i} exp(x'_{ij}\hat{\beta} + \hat{\gamma}_i)}{1 + \sigma^2 \sum_{j=1}^{n_i} exp(x'_{ij}\hat{\beta} + \hat{\gamma}_i)} = 0 \end{aligned} \quad (2.45)$$

The estimate of σ^2 obtained from above equation will be denoted by $\hat{\sigma^2}$

It has however been demonstrated by Sutradhar and Qu (1998) [see also Jiang (1998)] that the PQL method can be inconsistent, specially for the variance component of the random effects σ^2, when cluster size is small.

Note that as the IMM approach has major problems with regard to efficiency (Sutradhar, 2004) and the PQL approach has serious drawbacks with regard to the consistency (Sutradhar and Qu, 1998; Jiang, 1998) of the estimate of the variance parameter of the random effects, these approaches will not be studied any more in this thesis.

Chapter 3

Generalized Quasi-Likelihood and Hierarchical Likelihood Inferences

It is clear from the discussion in the last chapter that the PQL approach may not yield consistent estimate for the variance component, σ^2, of the random effects, whereas the IMM approach may yield consistent but inefficient estimate for this parameter. Recall that there also exists generalized quasi-likelihood (GQL) [see Sutradhar (2004)] and hierarchical likelihood (HL) [see Lee and Nelder (1996)] approaches for the estimation of both β and σ^2. Also recall that the GQL approach has been compared with the IMM approach by Sutradhar (2004) where it was shown that GQL approach is not only consistent but also more efficient than the IMM approach. In view of this, there does not arise any necessity to compare the GQL and PQL approaches, mainly because of the fact that the PQL approach may yield inconsistent estimate for σ^2 [Sutradhar and Qu (2001), Jiang (1998)] whereas GQL always produces consistent estimate for this parameter. Note however that the relative performance of the GQL approach as compared to the HL approach has not yet been studied. The purpose of this chapter is to examine the relative performance of these two, i.e. GQL and HL approaches with regard to both consistency and efficiency.

3.1 Generalized Quasi-likelihood (GQL) Inference

In this section we provide a review on the construction of the GQL estimating equations for the parameters of the Poisson-Normal mixed model.

3.1.1 GQL Estimating Equations

To construct the GQL estimating equation for the regression effects, β, we first consider a distance vector $y_i - m_i$ (i=1,2,...,K) such that $E(Y_i - m_i) = 0$. Here, $y_i = [y_{i1}, y_{i2}, ..., y_{in_i}]'$ and $m_i = E[Y_i] = [m_{i1}, m_{i2}, ..., m_{in_i}]'$, with

$$m_{ij} = E(Y_{ij}) = exp(x'_{ij}\beta + \frac{1}{2}z_i^2\sigma^2), \tag{3.1}$$

by Lemma 2.1. Now, for known σ^2, the GQL approach then solves the GQL estimating equation

$$\sum_{i=1}^{K} \frac{\partial m'_i}{\partial \beta} \Sigma_i^{-1} (y_i - m_i) = 0, \tag{3.2}$$

[Sutradhar (2004)], where Σ_i is the covariance matrix of y_i. Note that this GQL estimating equation is an extension of the quasilikelihood (QL) estimating equations proposed by Wedderburn (1979) for the independence case, which is derived by exploiting only the mean and the variance functions. Further note that the formula for the elements of this Σ_i matrix are given by (2.25) and (2.26). After some algebras, the matrix of derivatives in (3.2) may be simplified as

$$\partial m'_i / \partial \beta = X'_i A_i,$$

where

$$X_i = \begin{bmatrix} x_{i11} & x_{i12} & \cdots & x_{i1p} \\ x_{i21} & x_{i22} & \cdots & x_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{in_i1} & x_{in_i2} & \cdots & x_{in_ip} \end{bmatrix}_{n_i \times p} \qquad A_i = \begin{bmatrix} m_{i1} & 0 & \cdots & 0 \\ 0 & m_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_{in_i} \end{bmatrix}_{n_i \times n_i}$$

Note that the GQL estimating equation (3.2) may be solved by using the iterative algorithm

$$\hat{\beta}_{(r+1)} = \hat{\beta}_{(r)} + \left[\left(\sum_{i=1}^{K} \frac{\partial m_i'}{\partial \beta} \Sigma_i^{-1} \frac{\partial m_i}{\partial \beta'} \right)^{-1} \sum_{i=1}^{K} \frac{\partial m_i'}{\partial \beta} \Sigma_i^{-1} (y_i - m_i) \right]_{\beta = \hat{\beta}_{(r)}} , \qquad (3.3)$$

where the quantity within the square bracket is evaluated at $\beta = \hat{\beta}_{(r)}$, r being the rth iteration.

Note that the GQL iterative equation (3.3) produces both consistent and efficient estimates. To be specific, since $E(Y_i) = m_i$, it produces consistent estimates because it is constructed based on the unbiased estimating functions $y_i - m_i$; for all i = 1,2,..., K. Furthermore, as the weight matrix Σ_i in (3.3) is actually the true covariance matrix of y_i, the GQL approach produces efficient estimate for β.

By similar computation as in the GQL estimating equation for β, we may also construct a GQL estimating equation for σ^2. To be specific, we now use the squares and the products of the observations and take their differences from their corresponding means to constitute a distance vector. Let $w_i = [y_{i1}^2, ..., y_{in_i}^2, y_{i1}y_{i2}, ..., y_{in_{i-1}}y_{in_i}]'$ be the $\frac{n_i(n_i+1)}{2} \times 1$ vector of squares and products of the responses from the ith cluster. Also, let $\lambda_i = \text{E}(W_i) = [\lambda_{i11}, ..., \lambda_{in_in_i}, \lambda_{i12}, ..., \lambda_{in_{i-1}n_i}]'$ and $\Omega_i = \text{Cov}\ (W_i)$ be the expectation and covariance of w_i. Now, following (3.2), one may write the GQL estimating equation for σ^2 as

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \sigma^2} \Omega_i^{-1} (w_i - \lambda_i) = 0. \tag{3.4}$$

Note that the GQL estimating equation (3.4) may be solved by using the iterative equation

$$\hat{\sigma^2}_{(r+1)} = \hat{\sigma^2}_{(r)} + \left[\left(\sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \sigma^2} \Omega_i^{-1} \frac{\partial \lambda_i}{\partial \sigma^2} \right)^{-1} \sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \sigma^2} \Omega_i^{-1} (w_i - \lambda_i) \right]_{\sigma^2 = \hat{\sigma^2}_{(r)}}, \tag{3.5}$$

where the quantity within the square bracket is evaluated at $\sigma^2 = \hat{\sigma^2}_{(r)}$, r being the rth iteration.

In (3.5), the derivative $\partial \lambda_i / \partial \sigma^2$ is computed from the following formulas:

$$\frac{\partial \lambda_{ijj}}{\partial \sigma^2} = \frac{1}{2} m_{ij} z_i^2 + 2 m_{ij}^2 \, e^{z_i^2 \sigma^2} \, z_i^2 \tag{3.6}$$

$$\frac{\partial \lambda_{ijk}}{\partial \sigma^2} = 2 m_{ij} m_{ik} \, e^{z_i^2 \sigma^2} \, z_i^2. \tag{3.7}$$

Note that all the elements of the third and fourth order moments matrix Ω_i may be computed following Lemmas 2.2, 2.3 and 2.4. For convenience, we show below how to compute some of them. For example,

$$\begin{aligned} V(Y_{ij}^2) &= E(Y_{ij}^4) - \{E(Y_{ij}^2)\}^2 \\ &= m_{ij} + m_{ij}^2 (7 e^{z_i^2 \sigma^2} - 1) + 2 m_{ij}^3 \, e^{z_i^2 \sigma^2} (3 e^{2 z_i^2 \sigma^2} - 1) \\ &\quad + m_{ij}^4 \, e^{2 z_i^2 \sigma^2} (e^{4 z_i^2 \sigma^2} - 1), \end{aligned} \tag{3.8}$$

$$\begin{aligned} Cov(Y_{ij}^2, Y_{ik}^2) &= E(Y_{ij}^2 Y_{ik}^2) - E(Y_{ij}^2) \, E(Y_{ik}^2) \\ &= m_{ij} \, m_{ik} (e^{z_i^2 \sigma^2} - 1) + m_{ij}^2 \, m_{ik} \, e^{z_i^2 \sigma^2} (e^{2 z_i^2 \sigma^2} - 1) \end{aligned}$$

$$+ \; m_{ij}^2 \, m_{ik}^2 \, e^{2z_i^2 \sigma^2} \, (e^{4z_i^2 \sigma^2} - 1)^2. \tag{3.9}$$

$$\begin{aligned} V\,(Y_{ij},\, Y_{ik}) &= E\,(Y_{ij}^2 Y_{ik}^2) \; - \; \{E\,(Y_{ij}\, Y_{ik})\}^2 \\ &= m_{ij}\, m_{ik}\, e^{z_i^2 \sigma^2} + m_{ij}^2\, m_{ik}\, e^{3z_i^2 \sigma^2} + m_{ij}\, m_{ik}^2\, e^{3z_i^2 \sigma^2} \\ &\quad + m_{ij}^2\, m_{ik}^2\, e^{3z_i^2 \sigma^2}\, (e^{4z_i^2 \sigma^2} - 1) \end{aligned} \tag{3.10}$$

$$\begin{aligned} Cov\,(Y_{ij}^2,\, Y_{ik}\, Y_{il}) &= E\,(Y_{ij}^2\, Y_{ik}\, Y_{ik}) - \; E\,(Y_{ij}^2)\, E\,(Y_{ik})\, E\,(Y_{il}) \\ &= m_{ij}\, m_{ik}\, m_{il}\, e^{z_i^2 \sigma^2}\, (e^{2z_i^2 \sigma^2} - 1) \\ &\quad + m_{ij}^2\, m_{ik}\, m_{il}\, e^{2z_i^2 \sigma^2}\, (e^{4z_i^2 \sigma^2} - 1) \end{aligned} \tag{3.11}$$

$$\begin{aligned} Cov\,(Y_{ij}\, Y_{ik},\, Y_{il}\, Y_{im}) &= E\,(Y_{ij}\, Y_{ik}\, Y_{ik}\, Y_{im}) - \; E\,(Y_{ij}^2)\, E\,(Y_{ik})\, E\,(Y_{il})\, E\,(Y_{im}) \\ &= m_{ij}\, m_{ik}\, m_{il}\, m_{im}\, e^{2z_i^2 \sigma^2}\, (e^{4z_i^2 \sigma^2} - 1). \end{aligned} \tag{3.12}$$

Note that both (3.2) and (3.4) have to be solved simultaneously for β and σ^2. Let $\hat{\beta}_{GQL}$ and $\hat{\sigma^2}_{GQL}$ be the solutions of (3.2) and (3.4) respectively.

3.2 Hierarchical likelihood (HL) Inference

Lee and Nelder (1996) used the Hierarchical likelihood (HL) approach for estimation of the parameters in GLMM. In this approach, similar to that of the PQL approach, they jointly estimate β and γ_i. They however estimate these quantities by maximizing the hierarchical likelihood function, whereas in the PQL approach Breslow and Clayton (1993) estimated these quantities by maximizing a penalized quasi-likelihood function. The estimate of σ^2 by HL approach appears to be quite different than the PQL approach. To be specific, in the HL approach, an adjusted profile hierarchical likelihood function is constructed and maximized with respect to σ^2. Then by using the estimates of β and γ_i, they obtained the maximum adjusted profile h-likelihood estimate of σ^2.

We now provide a brief review of the HL approach as follows.

3.2.1 HL Estimating Equations

The hierarchical likelihood or h-likelihood, denoted by **h**, is defined as

$$\mathbf{h} = l(y_{ij}|\gamma_i) + l(\gamma_i; \sigma^2), \tag{3.13}$$

where

$$l(y_{ij}|\gamma_i) = logL(y_{ij}|\gamma_i)$$

$$= -\sum_{i=1}^{K}\sum_{j=1}^{n_i} \mu_{ij}^* + \sum_{i=1}^{N}\sum_{j=1}^{n_i} y_{ij}\, log(\mu_{ij}^*) - \sum_{i=1}^{N}\sum_{j=1}^{n_i} log\,(y_{ij}!), \tag{3.14}$$

and

$$l(\gamma_i; \sigma^2) = -\frac{K}{2} log(2\pi) - \frac{K}{2} log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{K} \gamma_i^2 \tag{3.15}$$

under the normality assumption for γ_i; that is, under the assumption that $\gamma_i \overset{iid}{\sim} N(0, \sigma^2)$.

In (3.14),

$$\mu_{ij}^* = E(y_{ij}|\gamma_i) = exp(x_{ij}'\beta + z_i\gamma_i). \tag{3.16}$$

Now for known γ_i, the HL estimating equation for β may be written as

$$\frac{\partial h}{\partial \beta} = \sum_{i=1}^{K} X_i'\,(y_i - \mu_i^*) = 0, \tag{3.17}$$

where, X_i and y_i are the same as in the GQL approach, whereas $\mu_i^* = [\mu_{i1}^*, \mu_{i2}^*, ..., \mu_{in_i}^*]'$ with μ_{ij}^* as in (3.16).

Note that the HL estimating equation (3.17) may be solved by using the following iterative equation

$$\hat{\beta}_{(r+1)} = \hat{\beta}_{(r)} + \left[\left(\sum_{i=1}^{K} X_i' W_i X_i \right)^{-1} \sum_{i=1}^{K} X_i' (y_i - \mu_i^*) \right]_{\beta = \hat{\beta}_{(r)}} , \tag{3.18}$$

where, $W_i = diag\,(\mu_{i1}^*, \mu_{i2}^*, ..., \mu_{in_i}^*)$ the quantity within the square bracket is evaluated at $\beta = \hat{\beta}_{(r)}$, r being the rth iteration.

Next, for known β and σ^2 the estimating equation for γ_i in the HL approach is given by

$$\frac{\partial h}{\partial \gamma_i} = \sum_{j=1}^{n_i} (y_{ij} - \mu_{ij}^*)\, z_i - \frac{\gamma_i}{\sigma^2} = 0. \tag{3.19}$$

Note that the HL estimating equation (3.19) may be solved by using the iterative equation

$$\hat{\gamma}_{i(r+1)} = \hat{\gamma}_{i(r)} + \left[\left(\sum_{j=1}^{n_i} \mu_{ij}^* z_i^2 + \frac{1}{\sigma^2} \right)^{-1} \left\{ \sum_{j=1}^{n_i} (y_{ij} - \mu_{ij}^*)\, z_i - \frac{\gamma_i}{\sigma^2} \right\} \right]_{\gamma_i = \hat{\gamma}_{i(r)}} , \tag{3.20}$$

where the quantity within the square bracket is evaluated at $\gamma_i = \hat{\gamma}_{i(r)}$, r being the rth iteration.

For the estimation of σ^2, Lee and Nelder (1996) exploited the general adjusted profile h-likelihood given by

$$h_A = h + \frac{1}{2} log\,\{det\,(2\pi\,\varphi\, H^{-1})\} \tag{3.21}$$

under the GLMM set-up. Since $\varphi = 1$ for the Poisson model, the general adjusted profile h-likelihood in (3.21) reduces to

$$h_A = h + \frac{1}{2} log\,(2\pi) - \frac{1}{2} log\,\{det(H)\}, \tag{3.22}$$

for the Poisson-normal mixed model. In both (3.21) and (3.22) the H matrix is defined as

$$H = \begin{bmatrix} X'WX & X'WZ \\ Z'WX & Z'WZ + U \end{bmatrix}_{(p+K)\times(p+K)}$$

where

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jp} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_K 1} & x_{n_K 2} & \cdots & x_{n_K p} \end{bmatrix}_{\sum n_i \times p} \qquad W = \begin{bmatrix} \mu^*_{11} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \mu^*_{1n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \mu^*_{Kn_K} \end{bmatrix}_{\sum n_i \times \sum n_i}$$

$$Z = \begin{bmatrix} 1_{n_1} & 0 & \cdots & 0 \\ 0 & 1_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1_{n_K} \end{bmatrix}_{\sum n_i \times K} \qquad U = \frac{1}{\sigma^2} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{K \times K}$$

Now maximization of (3.22) with regard to σ^2 is achieved by using the iterative equation given by

$$\hat{\sigma}^2_{(r+1)} = \hat{\sigma}^2_{(r)} + \left[\left(\frac{\partial^2 h_A}{\partial \sigma^4} \right)^{-1} \frac{\partial h_A}{\partial \sigma^2} \right]_{(r)}, \tag{3.23}$$

where the square bracket $[\ ]_{(r)}$ indicates that the quantity in $[\ ]$ is evaluated at $\sigma^2 = \hat{\sigma}^2_{(r)}$, r being the rth iteration.

For the purpose of computing $\frac{\partial h_A}{\partial \sigma^2}$ and $\frac{\partial^2 h_A}{\partial \sigma^4}$ in (3.23), we now give their formulas in details. That is,

$$\frac{\partial h_A}{\partial \sigma^2} = \frac{\partial h}{\partial \sigma^2} - \frac{1}{2} \frac{\partial log\{det(H)\}}{\partial \sigma^2}, \tag{3.24}$$

where

$$\frac{\partial h}{\partial \sigma^2} = -\frac{K}{2\sigma^2} + \frac{\sum_{i=1}^{K} \gamma_i^2}{2\sigma^4}$$

and

$$\frac{\partial log\{det(H)\}}{\partial \sigma^2} = trace\left(H^{-1}\frac{\partial H}{\partial \sigma^2}\right) = trace\left(D\frac{\partial U}{\partial \sigma^2}\right),$$

with D $= [(Z'WZ + U) - Z'WXX'WXX'WZ]^{-1}$ as the bottom diagonal matrix of H^{-1} with appropriate dimension. It then follows from equation (3.24) that

$$\frac{\partial h_A}{\partial \sigma^2} = -\frac{K}{2\sigma^2} + \frac{\sum_{i=1}^{K} \gamma_i^2}{2\sigma^4} + \frac{tr(D)}{2\sigma^4}. \tag{3.25}$$

Now, by taking a further derivative of (3.25) or (3.24) with respect to σ^2, we obtain:

$$\frac{\partial^2 h_A}{\partial \sigma^4} = \frac{\partial^2 h}{\partial \sigma^4} - \frac{1}{2}\frac{\partial^2 log\{det(H)\}}{\partial \sigma^4}, \tag{3.26}$$

where

$$\frac{\partial^2 h}{\partial \sigma^4} = \frac{K}{2\sigma^4} - \frac{\sum_{i=1}^{K} \gamma_i^2}{\sigma^6} \tag{3.27}$$

and

$$\frac{\partial^2 log\{det(H)\}}{\partial \sigma^4} = trace\left\{H^{-1}\left(\frac{\partial^2 H}{\partial \sigma^4}\right)\right\} - trace\left\{H^{-1}\left(\frac{\partial H}{\partial \sigma^2}\right).H^{-1}\left(\frac{\partial H}{\partial \sigma^2}\right)\right\}$$

$$= trace\left\{D\left(\frac{\partial^2 U}{\partial \sigma^4}\right)\right\} - trace\left\{D\left(\frac{\partial U}{\partial \sigma^2}\right).D\left(\frac{\partial U}{\partial \sigma^2}\right)\right\}$$

$$= \frac{2\,tr(D)}{\sigma^6} - \frac{tr(DD)}{\sigma^8}. \tag{3.28}$$

By using (3.27) and (3.28) in (3.26) one obtains

$$\frac{\partial^2 h_A}{\partial \sigma^4} = \frac{K}{2\sigma^4} - \frac{\sum_{i=1}^K \gamma_i^2}{\sigma^6} - \frac{tr(D)}{\sigma^6} + \frac{tr(DD)}{2\sigma^8}. \tag{3.29}$$

Note that all the estimating equations (3.18), (3.20) and (3.23) have to be solved simultaneously for β, γ_i and σ^2. Let $\hat{\beta}_{HL}$ and $\hat{\gamma}_{iHL}$ be the solution of (3.18) and (3.20) respectively. Similarly, we denote the estimate of σ^2 obtained from (3.23) by $\hat{\sigma^2}_{HL}$.

In the next chapter we conduct a simulation study to compare the performance of the GQL estimates of β and σ^2 obtained from (3.3) and (3.5) with those of the HL estimates obtained from equations (3.18), and (3.23).

Chapter 4

GQL vs HL Estimation: A Simulation Study

Recall from Chapter 2 that when count responses of the member of a family are treated to be independent, the likelihood estimate of β obtained from (2.36) performs very poorly especially when σ^2 is large. As a remedy, the GQL and HL approaches described in Chapter 3 in order to estimate both β and σ^2 parameters. The purpose of this chapter is to compare the performances of $\hat{\beta}_{HL}$ and $\hat{\sigma^2}_{HL}$ obtained from (3.18) and (3.23) with those of $\hat{\beta}_{GQL}$ and $\hat{\sigma^2}_{GQL}$ obtained from equations (3.3) and (3.5), through a simulation study. Note that as mentioned in Chapter 2, apart from the GQL and HL approaches, there also exists other alternative approaches, namely PQL and IMM, for the estimation of β and σ^2. These later approaches are not included here for comparison as they have been found to have serious consistency and efficiency problems mainly for the estimation of the variance parameter (Sutradhar and Qu, 1998) of the random effects.

4.1 Simulation Design

Note that in Chapter 2 we carried out a simulation study to examine the effects of ignoring σ^2 on the likelihood estimation for the regression parameter $\beta = (\beta_1, \beta_2)'$. It was found that there is a large negative bias when σ^2 is ignored when estimating β. For convenience, here we consider the same simulation design as in Chapter 2 but we examine the performances of the GQL and HL approaches in estimating both β and σ^2. Thus, the count data in a given simulation are generated following (2.37) with the same $\beta_1 = \beta_2 = 1.0$ and the same covariates x_{ij1} and x_{ij2}. For the selection of σ^2, we however consider various small and large values, namely,

$$\sigma^2 = 0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.0 \text{ and } 1.2.$$

Note that even though much more larger σ^2 such as $\sigma^2 = 1.5$, 2.0 and 2.5 were considered in Chapter 2, here we have chosen the values of σ^2 up to 1.2. We remark here that the values of $\sigma^2 = 1.0$ and 1.2 are themselves quite large. This is because under the Poisson mixed model, the variance of the response y_{ij} is given by (2.25), namely, $V(Y_{ij}) = m_{ij} + m_{ij}^2 (e^{z_i^2 \sigma^2} - 1)$, which become quite large even if $\sigma^2 = 1.0$ or 1.2. By the same token, some authors such as, Breslow and Lin (1995, P.90) were able to estimate this variance parameter σ^2 consistently where σ^2 ranges up to 0.5.

4.2 GQL and HL Estimation

Recall that under the Poisson mixed model, the GQL estimate of β is obtained by solving the estimating equation

$$\sum_{i=1}^{K} \frac{\partial m_i'}{\partial \beta} \Sigma_i^{-1} (y_i - m_i) = 0,$$

[see equation (3.2)] whereas the variance of the random effects is estimated by solving the estimating equation

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \sigma^2} \Omega_i^{-1} (w_i - \lambda_i) = 0,$$

[see equation (3.4)]. Note that both of these equations are known to be unbiased. This is because $\mathrm{E}(Y_i - m_i) = 0$ as well as $\mathrm{E}(W_i) = \lambda_i$. Moreover, the weight matrices Σ_i and Ω_i in the above estimating equations are the true covariance matrices of Y_i and W_i respectively. This principle of using the true covariance matrix as the weight matrix makes the GQL estimating equations efficient. This is already known from Sutradhar (2004) that the GQL estimates of β and σ^2 are consistent and efficient as compared the MM and IMM estimates. These estimates were however not compared with the so-called HL estimates.

Now to shed some light on the properties of the HL estimators for β and σ^2, we recall for (3.17) that the HL estimate of β is obtained by solving

$$\frac{\partial h}{\partial \beta} = \sum_{i=1}^{K} X_i' (y_i - \mu_i^*) = 0.$$

Note that this equation is unbiased conditional on γ_i. Thus it is not surprising that weight matrix is the identity matrix in such an equation. This is because conditional on γ_i, the clustered responses are independent. Nevertheless, the estimation effect of γ_i will be reflected on the behavior of the HL estimate of β, γ_i being estimated by solving

$$\frac{\partial h}{\partial \gamma_i} = \sum_{j=1}^{n_i} (y_{ij} - \mu_{ij}^*) z_i - \frac{\gamma_i}{\sigma^2} = 0,$$

[see equation (3.19)]. In fact the estimation effects of γ_i on β estimation may be unsatisfactory at times specially when σ^2 is large. This is because, as σ^2 increases, one may not be able to estimate γ_i consistently. An intuitional justification for this inconsistency becomes clear when conditional on γ_i, an expectation is taken over the responses in the estimating equation for γ_i. This operation produces γ_i on the average as 0, even though γ_i can be quite large when σ^2 is large. The purpose of the simulation study is to examine the behavior of this γ_i estimation based approach. Further note that the PQL approach of Breslow and Clayton (1993) also estimates γ_i in a similar fashion that leads to biased and hence inconsistent estimates for the parameters, specially for σ^2 (Sutradhar and Qu, 1998), when the true σ^2 is large.

Turning back to the HL estimation of the remaining parameter, that is of σ^2, we recall its estimating equation from (3.25) given by

$$\frac{\partial h_A}{\partial \sigma^2} = -\frac{K}{2\sigma^2} + \frac{\sum_{i=1}^{K} \gamma_i^2}{2\sigma^4} + \frac{tr(D)}{2\sigma^4} = 0.$$

This equation reveal that the estimate of σ^2 also highly depends on the estimation of γ_i. Since γ_i, on the average, is estimated to be zero or small quantities, this equation also shows that σ^2 as a function of γ_i may not be consistently estimated, specially when the true σ^2 is large.

4.3 Relative Performance of GQL and HL Approaches

In this section we examine the performances of the GQL and HL estimation approaches in estimating the regression effects (β) as well as the variance of the random effects (σ^2) of the model through the simulation study. Note that the HL approach requires the estimation of γ_i for the estimation of the β and σ^2, whereas the GQL approach does not at all require such estimation of random effects. This makes the GQL approach simpler. Moreover, as indicated earlier, the use of the estimates of γ_i for the estimation of β and σ^2 under the HL approach may in fact be counter productive in the sense of consistency. Nevertheless, we had to compute the estimates of γ_i under the HL approach under each simulation. But there are too many estimates to report as i goes from 1 to K = 100.

Table 4.1: *Comparison of simulated mean values, standard errors (SSE), and relative bias (SRB) of the regression estimates and estimates of variance of random effects by the GQL and HL approaches for selected values of* σ^2*: K=100; n=2; true values of the regression parameters:* $\beta_1 = 1.0$ *and* $\beta_2 = 1.0$

Cluster size	σ^2	**Method**	**Quantity**	**Value** β_1	β_2	$\hat{\sigma}^2$
2	0.20	GQL	SM	1.0101	1.0105	0.19464
			SSE	0.0429	0.0292	0.03485
			SRB	24	36	15
		HL	SM	1.0402	1.0312	0.18771
			SSE	0.0412	0.0262	0.02524
			SRB	98	119	49
	0.40	GQL	SM	1.0151	1.0177	0.38080
			SSE	0.0458	0.0335	0.05454
			SRB	33	52	35
		HL	SM	1.0777	1.0671	0.37277
			SSE	0.0399	0.0265	0.03490
			SRB	195	253	78
	0.60	GQL	SM	1.0156	1.0202	0.57406
			SSE	0.0518	0.0389	0.07232
			SRB	30	52	36
		HL	SM	1.1275	1.1157	0.56804
			SSE	0.0379	0.0330	0.04994
			SRB	336	351	64

..... Continued

Table 4.1: (Continued)

Cluster size	σ^2	Method	Quantity	Value β_1	β_2	$\hat{\sigma}^2$
2	0.80	GQL	SM	1.0167	1.0211	0.76685
			SSE	0.0577	0.0434	0.08797
			SRB	29	49	38
		HL	SM	1.2498	1.2320	0.86248
			SSE	0.0467	0.0613	0.09693
			SRB	535	378	64
	1.00	GQL	SM	1.0136	1.0184	0.96705
			SSE	0.0603	0.0503	0.10138
			SRB	23	37	33
		HL	SM	1.5517	1.5168	1.86070
			SSE	0.0860	0.1038	0.39500
			SRB	642	498	218
	1.20	GQL	SM	1.0151	1.0167	1.1580
			SSE	0.0648	0.0532	0.10970
			SRB	23	31	38
		HL	SM	2.0532	1.9972	4.94560
			SSE	0.1144	0.1306	0.91610
			SRB	921	764	409

Now to compute the GQL and HL estimates of β and σ^2 we have chosen to use 1000 simulations. In each simulation, the count responses were first generated as described in Section 4.1. These responses along with the covariates explained in the same section, Section 4.1 [see also Section 2.2.2], are then used in the GQL and HL estimating equations for β and σ^2 provided in the last section [see also Section 3.1 for the GQL and Section 3.2 for HL estimating equations]. To be specific, $\hat{\beta}_{GQL}$

and $\hat{\sigma^2}_{GQL}$ under each simulation were obtained by using the iterative equations (3.3) and (3.5), respectively. Similarly, under each simulation, $\hat{\beta}_{HL}$ and $\hat{\sigma^2}_{HL}$ were obtained by using the iterative equations (3.18) and (3.23) respectively. Based on 1000 simulations, the average of these estimates, that is, the simulated means (SM), along with their simulated standard errors (SSE) and simulated percentage relative biases (SRB), were computed and reported in Tables 4.1, 4.2 and 4.3, for cluster size n = 2, 4 and 6, respectively.

For cluster size 2, it is clear from Table 4.1 that the GQL estimators for β_1 and β_2 are almost unbiased, whereas the HL estimators appear to have large biases, especially when σ^2 is large such as $\sigma^2 > 0.6$. As far as their standard errors are concerned, the HL estimates of β_1 and β_2 appears to have smaller standard errors than the corresponding GQL estimators. This result clearly shows that for large σ^2 the HL estimator converges to the wrong value with smaller standard error, and hence the HL estimator may not be consistent. With regard to the GQL estimators, as they are almost unbiased, and they are actually consistent even though their SSE are slightly larger. These results motivated us to display the percentage relative biases for both GQL and HL estimates. These SRBs clearly demonstrate that HL estimates have very large relative biases as compared to those of GQL estimates.

We now interpret some of the specific results from Table 4.1. To do this, we choose a moderately small value of $\sigma^2 = 0.4$ and a large value of $\sigma^2 = 1.2$ and examine the estimates of β_1 and β_2. It is seen that when $\sigma^2 = 0.4$, the GQL estimates of β_1 and β_2 have SRB 33 and 52, whereas the HL estimates have SRB 195 and 253, respectively. Thus, the HL estimates clearly exhibit large biases even if σ^2 is small, such as $\sigma^2 =$ 0.4. When the SRBs for $\sigma^2 = 1.2$ are considered, it is found that the GQL estimates of β_1 and β_2 have SRBs 23 and 31, whereas the HL approach produces estimates with 921 and 764 SRBs. These and other results from Table 4.1 clearly show that the GQL estimators performs uniformly better than the HL estimators in estimating the main parameters, i.e., the regression effects of the model.

Table 4.2: *Comparison of simulated mean values, standard errors (SSE), and relative bias (SRB) of the regression estimates and estimates of variance of random effects by the GQL and HL approaches for selected values of σ^2: K=100; n=4; true values of the regression parameters: $\beta_1 = 1.0$ and $\beta_2 = 1.0$*

Cluster size	σ^2	**Method**	**Quantity**	**Value**		
				β_1	β_2	$\hat{\sigma}^2$
4	0.20	GQL	SM	1.0073	1.0084	0.20012
			SSE	0.0341	0.0243	0.03557
			SRB	21	35	0.33
		HL	SM	1.0408	1.0344	0.19816
			SSE	0.0302	0.0196	0.01870
			SRB	135	176	10
	0.40	GQL	SM	1.0109	1.0122	0.39234
			SSE	0.0405	0.0305	0.06464
			SRB	27	40	12
		HL	SM	1.0878	1.0760	0.40431
			SSE	0.0293	0.0185	0.05466
			SRB	300	411	8
	0.60	GQL	SM	1.0142	1.0136	0.58426
			SSE	0.0454	0.0357	0.06864
			SRB	31	38	23
		HL	SM	1.1380	1.1197	0.61907
			SSE	0.0286	0.0183	0.04224
			SRB	483	654	45

..... Continued

Table 4.2: (Continued)

Cluster size	σ^2	Method	Quantity	Value β_1	β_2	$\hat{\sigma}^2$
4	0.80	GQL	SM	1.0125	1.0108	0.78193
			SSE	0.0483	0.0399	0.07726
			SRB	26	27	23
		HL	SM	1.1661	1.1430	0.82808
			SSE	0.0277	0.0205	0.05278
			SRB	600	698	53
	1.00	GQL	SM	1.0092	1.0086	0.98110
			SSE	0.0533	0.0438	0.08812
			SRB	17	20	21
		HL	SM	1.2435	1.2147	1.09220
			SSE	0.0256	0.0156	0.06490
			SRB	951	1376	142
	1.20	GQL	SM	1.0079	1.0097	1.17190
			SSE	0.0523	0.0442	0.08980
			SRB	15	22	31
		HL	SM	1.2894	1.2594	1.34880
			SSE	0.0249	0.0164	0.07270
			SRB	1162	1582	205

For the estimation of σ^2, the GQL and HL approaches appear to perform almost the same when the true value of σ^2 is small such as $\sigma^2 < 1.0$. For $\sigma^2 = 1.0$ and 1.2, the HL approach becomes highly biased, whereas the GQL approach appears to be only slightly biased. For example, for the true $\sigma^2 = 0.4$, the GQL approach produced $\hat{\sigma}^2_{GQL} = 0.3808$ with standard error 0.0545 and percentage relative bias 35, whereas the HL approach produced $\hat{\sigma}^2_{HL} = 0.3728$ with standard error 0.0349 and percentage relative bias 78. Furthermore, when the true σ^2 is large such as $\sigma^2 = 1.2$, the GQL approach produced $\hat{\sigma}^2_{GQL} = 1.158$ with standard error 0.1097 and percentage relative bias 38, whereas the HL approach produced $\hat{\sigma}^2_{HL} = 4.9456$ with standard error 0.9161 and percentage relative bias 409. These results clearly show that the GQL estimator for σ^2 is almost unbiased irrespective of the true value (small or large) of σ^2, whereas the HL estimator is almost unbiased for small σ^2 but highly biased for large σ^2. These results about the performances of the GQL and HL estimates of σ^2 are also verified through the comparison of SRBs. For example, when $\sigma^2 = 0.4$, the SRBs of the GQL estimate is found to be 35, whereas the SRBs produced by the HL approach is found to be 78.

When the simulation results for Tables 4.2 and 4.3 for n = 4 and 6, are compared with those of Table 4.1, it appears that the GQL estimators of β_1 and β_2 continue to perform better with lower SRBs, when the cluster size increases. The HL approach however shows poor performances with higher SRBs when cluster size increases. Thus the GQL approach uniformly performed better than the HL approach in estimating regression effects even if the cluster size was small, Note that in practice in familial studies cluster sizes will usually be small.

With regard to the estimation of σ^2, the result of all three Tables 4.1, 4.2 and 4.3 show that when n increases both GQL and HL approaches produce better estimates of σ^2, but the GQL approach always remains better than the HL approach in terms of SRB.

Table 4.3: *Comparison of simulated mean values, standard errors (SSE), and relative bias (SRB) of the regression estimates and estimates of variance of random effects by the GQL and HL approaches for selected values of* σ^2*: K=100; n=6; true values of the regression parameters:* $\beta_1 = 1.0$ *and* $\beta_2 = 1.0$

Cluster				**Value**		
size	σ^2	**Method**	**Quantity**	β_1	β_2	$\hat{\sigma}^2$
6	0.40	GQL	SM	1.0112	1.0115	0.39263
			SSE	0.0351	0.0283	0.05388
			SRB	32	41	14
		HL	SM	1.09580	1.08310	0.41699
			SSE	0.02400	0.01520	0.02610
			SRB	399	547	65
	0.80	GQL	SM	1.0081	1.0080	0.79385
			SSE	0.0460	0.0375	0.14485
			SRB	18	21	4
		HL	SM	1.15830	1.13660	0.84985
			SSE	0.02700	0.02070	0.04668
			SRB	586	660	107
	1.20	GQL	SM	1.0049	1.0053	1.18060
			SSE	0.0491	0.0415	0.08470
			SRB	10	13	23
		HL	SM	0.62217	0.64976	1.83400
			SSE	0.11312	0.10386	0.32340
			SRB	334	337	196

In the next chapter, we will provide an illustration of the relative performance of the GQL and HL approaches for the Health Care Utilization Data.

Chapter 5

A Numerical Illustration: Health Care Utilization Data Analysis

In Chapter 4 we discussed the relative performances of the GQL and HL approaches under the Poisson-normal mixed model setup through an extensive simulation study. In this chapter, we provide a numerical illustration of these two estimation methodologies by analyzing the Health Care Utilization Data collected by the General Hospital, St. John's, Canada. We must however caution that the GQL estimates should be recommended for any practical use. This is because it was demonstrated in the last chapter through a simulation study that overall, the GQL approach performs better in estimating all parameters when compared to the HL approach.

5.1 Health Care Utilization Data

Consider the health care utilization data collected by the Department of Community Medicine, Health Science Center (General Hospital) in St. John's, Canada in 1985. This data set comprises information from K = 48 families. Of these families, 36 are of size 4 ($n_i = 4$, i = 1,....,36), and the remaining 12 are of size 3 ($n_i = 3$, i = 37,....,48).

Table 5.1: *Summary Statistics of Physician Visits by Different Covariates in the Health Care Utilization Data for Year 1985.*

		Number of Visits					
Covariates	**Level**	**0**	**1**	**2**	**3-5**	**≥6**	**Total**
Gender	Male	28	22	18	16	12	96
	Female	11	5	15	21	32	84
Chronic Diseases	No	26	20	15	16	11	88
	Yes	13	7	18	21	33	92
Education Level	< High School	17	5	11	10	15	58
	≥ High School	22	22	22	27	29	122
Age	20-30	23	17	14	15	15	84
	31-40	1	1	3	3	3	11
	41-50	4	4	5	12	8	33
	51-65	10	5	8	5	13	41
	66-85	1	0	3	2	5	11

Each of the family members was asked about the number of visits they made to a physician during 1985. Their gender, the number of chronic diseases in 1985 they have been suffering from, education level and age were recorded. In fact, these families were followed for 6 years up to 1990, but in the present application we will deal with the 1985 data only. Note that in the present setup the responses, i.e., the number of visits paid by each member, are counts. Further, as n_i (3 or 4) members belong to the same ith (i = 1,....,48) family, it is likely that the responses of the family members are correlated. In this chapter we will take these correlations into account and examine the effects of the four associated covariates on the number of visits to the physician by the members of the family.

5.1.1 Exploratory Analysis

To have a feel for the relationship between the number of physician visits by a family member and his/her covariates, we have computed some summary statistics as shown in Table 5.1.

It is seen from Table 5.1 that, in general, more males appear to visit a physician a smaller number of times, while a large number of females visit a physician at least 3 times. As expected, we see that an individual with chronic diseases visits a physician more often. Physician visits for individuals with a higher level of education seems to be evenly distributed, i.e. individuals are just as likely to visit a physician once as 3-5 times. For those with lower level of eduction, they appear to either not visit a physician, or visit a large number of times. With regard to the relationship between number of visits and age, we have temporarily made 5 age groups and observed that some of the individuals in the 20-30 age group have visited a physician a large number of times. As expected, a large number of individuals did not visit a physician at all. For older age groups, there was a tendency for an individual to see a physician more often.

It is also seen that irrespective of the covariates, an individual on the average has visited his/her physician 3.92 times with variance 22.66.

5.2 GQL and HL Based Analysis of the Data

Note that although the summary statistics shed some light on the relationship between the number of physician visits and the four covariates, we wish to fit an appropriate model to these data and make a valid confirmatory analysis. For this we note that the responses are counts, which may be treated as a Poisson variable. However, it was found that the average number of physician visits for an individual was 3.92, with variance 22.66. This indicates that there is overdispersion in the data. This is not surprising, as the variance component of the random family effects may cause this. Since the responses are collected from members of the same family, they will be correlated which again may be measured through the variance component of the family effects. In terms of the notation of this thesis, these family effects are considered as γ_i for the ith family and it is assumed that these random family effects follow univariate normal distribution with mean 0 and variance σ^2.

We denote the covariates gender, chronic disease status (CD), education level (EL) and age by x_1, x_2, x_3 and x_4 respectively. To be specific, we define these four covariates for the jth (j = 1,..., n_i) member of the ith (i = 1,..., K = 48) family as

$$x_{ij1} = \begin{cases} 0 & \text{female} \\ 1 & \text{male} \end{cases} \qquad x_{ij2} = \begin{cases} 0 & \text{without chronic diseases} \\ 1 & \text{with chronic diseases} \end{cases}$$

$$x_{ij3} = \begin{cases} 0 & \text{less than high school} \\ 1 & \text{high school or above} \end{cases} \qquad x_{ij4} = \text{ exact age of the individual}$$

The purpose of this section is to compute the regression effects $\beta = (\beta_1, ..., \beta_4)'$ of the four covariates on the number of physician visits, taking the overdispersion parameter σ^2 into account. We do this computation by using both the GQL and HL approaches described in Chapter 3.

The GQL estimates of β and σ^2 were obtained by solving the iterative equations (3.3) and (3.5) respectively. Similarly, the HL estimates of β and σ^2 were obtained by solving the iterative equations (3.18) and (3.23) respectively. These estimates along with their standard errors are displayed in Table 5.2.

Table 5.2: *The GQL and HL Estimates for the Health Care Utilization Data for Year 1985.*

Method	Quantity	Effects of the Covariates				Variance
		Gender($\hat{\beta}_1$)	CD($\hat{\beta}_2$)	EL($\hat{\beta}_3$)	Age($\hat{\beta}_4$)	$\hat{\sigma}^2$
GQL	**Value**	-0.742	0.671	0.493	0.012	1.543
	SE	0.109	0.149	0.157	0.0038	1.566
HL	**Value**	-0.693	0.689	0.633	0.016	0.187
	SE	0.080	0.088	0.067	0.0017	0.020

The results in Table 5.2 show that except for the CD covariate, the HL approach produces different estimates for the remaining 3 covariates compared to the GQL approach. Also, similar to the simulation results, the HL approach appears to have smaller standard errors. Thus, the HL approach may have produced unreliable estimates for the covariates except for the CD covariate.

With regard to the estimation of σ^2, the HL approach produces quite different estimate than the GQL approach. But, as the simulation study indicated that the GQL approach always produces consistent estimates also for the σ^2 parameter, we take the $\hat{\sigma}^2_{GQL} = 1.543$ as a reliable estimate. Note that to verify the reliability of the GQL estimate for σ^2, we have further estimated this parameter by using the method of moments (MM) discussed by Jiang (1998). To be specific, the σ^2 is estimated by solving

$$[\text{S - E (S)}] = 0$$

where

$$S = \sum_{i=1}^{N} \sum_{j=1}^{n_i} y_{ij}^2 + \sum_{i=1}^{N} \sum_{j<k}^{n_i} y_{ij} y_{ik} \tag{5.1}$$

and

$$E(S) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} [\, m_{ij} + m_{ij}^2 \, e^{\sigma^2}] + \sum_{i=1}^{N} \sum_{j<k}^{n_i} m_{ij} \, m_{ik} \, e^{\sigma^2} \tag{5.2}$$

by Lemma 2.2 with $m_{ij} = exp\,(x_{ij}'\beta + \frac{1}{2}\,\sigma^2)$ by (3.1). It is known that this method of moments produces consistent estimate for σ^2 parameter. For the health care utilization data, this method of moments produces $\hat{\sigma}_{MM}^2 = 1.558$. This is quite close to the GQL estimate verifying the reliability of the GQL estimate.

We now interpret the effects of the covariates using the GQL estimates. Thus, $\hat{\sigma}_{GQL}^2 = 1.543$ indicates that the data contain large overdispersion. This is also in agrement with the results reported in Section 5.1.1 under Exploratory Analysis, where it was shown that an individual visits the physician 3.92 times on the average with a very large variance 22.66.

Furthermore, the negative value of $\hat{\beta}_{1(GQL)}$, namely $\hat{\beta}_{1,(GQL)}$ = - 0.742, indicates that females made more visits to the physician as compared to males. Next, $\hat{\beta}_{2(GQL)} = 0.671$ and $\hat{\beta}_{4,(GQL)} = 0.012$ suggest that the individuals having some chronic diseases or individuals who are older pay more visits to the physician, as expected. The effect of the education level on the health condition, however, appears to be intriguing. This is because $\hat{\beta}_{3(GQL)} = 0.493$ suggests that highly educated individuals have more visits compared to individuals with a lower level of education. One of the reasons for this type of behavior of this covariate may be that individuals with a higher level of education are more concerned about their health condition compared to individuals with a lower level of education.

Chapter 6

Concluding Remarks

In this thesis we have considered a Poisson-normal mixed model which is an important special case of the well-known generalized linear mixed model. In this problem, it is of interest to estimate the regression effects and variance of the random effects, consistently and efficiently. A great deal of discussion has taken place over the last two decades on the relative performance of some of the widely used estimation methods such as MM (Jiang, 1998), IMM (Jiang and Zhang, 2001), PQL (Breslow and Clayton, 1995) and GQL (Sutradhar, 2004) approaches. But none of these procedures were compared with the existing HL (Lee and Nelder, 1996) approach, even though this later approach appears to be quite familiar. Since the GQL approach was found to be better than MM, IMM and PQL approaches, in this thesis, we have examined the relative performance of this well behaved GQL approach with the HL approach.

For the comparison between the HL and GQL approaches, we have first simplified all related estimating equations under these two approaches. We then conducted an extensive simulation study to examine the relative performances of these procedures in estimating both regression effects and variance of the random effects. Note that the HL approach requires 3 estimating equations including the estimation of the random effects, whereas the GQL approach requires only two estimating equations where it is not needed to estimate the random effects.

The simulation study was conducted for three different cluster sizes and various values of σ^2 (variance of the random effects), small and large. It was found that as the value of σ^2 increases, the HL approach starts to produce highly biased estimates for the regression effects. The GQL approach was however found to be producing almost unbiased estimates for the regression effects, irrespective of the magnitude of σ^2. As far as the estimation of the variance parameter σ^2 is concerned, the GQL approach was also found to be uniformly better than the HL approach. In this case, in contrary to the regression estimation, the HL approach was found to perform better even though it trails to the GQL approach.

Hence, the GQL approach is definitely better than the HL approach in estimating all parameters of the model. When other studies mentioned above are taken into consideration, the GQL approach appears to be the best so far among the MM, IMM, PQL and HL approaches. We therefore recommend the use of the GQL approach in practice irrespective of the magnitude of the overdispersion in the familial/cluster count data.

In light of the present thesis, it may be of interest to make a comparative study between the GQL and HL approaches in estimating the parameters of the binary mixed model. This is however beyond the scope of this thesis. One may further consider a much more wider familial longitudinal model for count or binary data. An extension of the present GQL approach to such a familial-longitudinal model would be an interesting and challenging problem. We hope to study this model in the future.

Bibliography

[1] Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.*, **88**, 9–25.

[2] Breslow, N.E. and Lin, X. (1995). Bias correction in generalized linear models with single component of dispersion. *Biometrika*, **82**, 81–92.

[3] Fahrmeir, L. and Tutz, G. (1994). Multivariate Statistical Modelling Based on Generalized Linear Models. Springer - Verlag, New York.

[4] Goldstein, H. and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *J. Roy. Statist. Soc. Ser. A*, **159**, 505–513.

[5] Henderson, C. R. (1953). Estimation of variance and covariacne components. *Biometrices*, **9**, 226–252.

[6] Jiang, J. (1998). Consistent estimators in generalized linear mixed models. *J. Amer. Statist. Assoc.*, **93**, 720–729.

[7] Jiang, J. and Zhang, W. (2001). Robust estimation in generalized linear mixed models. *Biometrika*, **88**, 753–765.

[8] Kuk, A. C. (1995). Asymptotically unbiased estimation in generalized linear models with random effects. *J. Roy. Statist. Soc. Ser. B*, **57**, 395–407.

[9] Lee, Y. and Nelder, J.A. (1996). Hierarchical generalized linear models. *J. Roy. Statist. Soc. Ser. B*, **58**, 619–678.

[10] Lee, Y. and Nelder, J.A. (1996). Hierarchical generalized linear models: A sythesis of generalized linear models, random-effect models and structured dispersions. *Biometrika*, **88**, 987–1006.

[11] Lin, X. and Breslow, N.E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *J. Amer. Statist. Assoc.*, **91**, 1007–1016.

[12] McCullagh, P. and Nelder, J.A. (1989). Generalized Linear Models. Chapman and Hall, London.

[13] Rodriguez, G. and Goldman, N. (1995). An assesment of estimation procedures for multilevel models with binary responses. *J. Roy. Statist. Soc. Ser. A*, **158**, 73–89.

[14] Schall, R. (1991). Estimation of generalized linear models with random effects. *Biometrika*, **78**, 719–727.

[15] Solomon, P. J. and Cox, D. R. (1992). Nonlinear components of variance models. *Biometrika*, **79**, 1–11.

[16] Sutradhar, B.C. and Qu, Z. (1998). On approximation likelihood inference in Poisson mixed Model. *Canad. J. Statist.*, **26**, 169–186.

[17] Sutradhar, B.C. and Rao, R.P. (2001). On marginal quasi-likelihood inference in generalized linear mixed models. *J. Multivariate Anal.*, **76**, 1–34.

[18] Sutradhar, B.C. (2004). On exact quasilikelihood inference in generalized linear mixed models. *The Ind. J. of Statist.*, **66**, 263–291.

[19] Waclawiw, M.A. and Liang, K.Y. (1993). Prediction of random effects in the generalized linear model. *J. Amer. Statist. Assoc.*, **88**, 171–178.

[20] Wedderburn, R. (1979). Quasi-likelihood functions, generalized linear models, the Gauss-Newton method. *Biometrika*, **61**, 439–447.

[21] Zeger, S.L. (1988). A Regression Model for Time Series of Counts. *Biometrika*, **75**, 621–629.