2011

Some Inference Problems in Clustered (Longitudinal) Count Data with Over-dispersion

Kazi Azad
University of Windsor

Some Inference Problems in Clustered (Longitudinal) Count Data With Over-Dispersion

by

Kazi Mahbubur Azad

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Mathematics and Statistics
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada
2009

Some Inference Problems in Clustered (Longitudinal) Count Data With Over-Dispersion

by
Kazi Mahbubur Azad

APPROVED BY:

Dr. Grace Y. Yi, External Examiner
University of Waterloo

Dr. Y. Aneja
Odette School of Business.

Dr. Myron Hlynka
Department of Mathematics & Statistics.

Dr. A. Hussein
Department of Mathematics & Statistics.

Dr. Sudhir Paul, Advisor
Department of Mathematics & Statistics.

Dr. Aaron Fisk, Chair of Defense
Faculty of Graduate Studies and Research

Date: August 14, 2009

Author's Declaration of Originality

I hereby certify that I am the sole author of this dissertation and that no part of this dissertation has been published or submitted for publication.

I certify that, to the best of my knowledge, my dissertation does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my dissertation, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my dissertation and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my dissertation, including any final revisions, as approved by my dissertation committee and the Graduate Studies office, and that this dissertation has not been submitted for a higher degree to any other University or Institution.

Abstract

Clustered (includes longitudinal) count data arise in many bio-statistical practices in which a number of repeated responses are observed over time from a number of individuals. One important problem that arises in practice is to test homogeneity within clusters (individuals) and between clusters (individuals). As data within clusters are observations of repeated responses, the count data may be correlated and/or over-dispersed. Jacqmin-Gadda and Commenges (1995) derive a score test statistic H_S by assuming a random intercept model within the framework of the generalized linear mixed model by obtaining exact variance of the likelihood score under the null hypothesis of homogeneity and a score test statistic H_T using the generalized estimating equation (GEE) approach (Liang and Zeger, 1986; Zeger and Liang, 1986). They further show that the two tests are identical when the covariance matrix assumed in the GEE approach is that of the random-effects model. In each of these cases they deal with (a) the situation in which the dispersion parameter ϕ is assumed to be known and (b) the situation in which the dispersion parameter ϕ is assumed to be unknown. The second situation, however, is more realistic as ϕ will be unknown in practice. For over-dispersed count data with unknown over-dispersion parameter we use the score test procedure of Rao (1947) and derive three tests by assuming a random intercept model within the framework of (i) the over-dispersed generalized linear model (ii) the negative binomial model, and (iii) the double extended quasi likelihood model (Lee and Nelder, 2001). All these three statistics are much simpler than the statistic obtained from the statistic H_S derived by Jacqmin-Gadda and Commenges (1995) under the framework of the over-dispersed generalized linear mixed effects model. The second statistic takes the over-dispersion more directly into the model and therefore is expected to do well when the model assumptions are satisfied and the other statistics are expected to be robust. Simulations show superior level property of the statistics derived under the negative binomial and double extended quasi-likelihood model assumptions. Further, two score tests have been developed to test for over-dispersion in the generalized linear mixed model. The four score tests of homogeneity and the two score tests for detecting over-dispersion are applied to two real life data examples. A plan for future study is given.

Dedication

Dedicated to my parents, wife and children.

Acknowledgements

First of all thanks to Almighty Allah for His all blessings that He has bestowed.

I am very much grateful to my supervisor Dr. Sudhir R. Paul for his innovative ideas and continuous help, advice, encouragement and guidance throughout the course of this dissertation. I am also thankful to him for his generous financial support from his NSERC grant throughout my study period.

I am thankful to my committee members Dr. M. Hlynka, Dr. A. Hussein and Dr. Y. Aneja for their review, comments and suggestions which improved this dissertation. I would also like to express my gratitude to my external examiner Dr. Grace Y. Yi, for her valuable suggestions and constructive criticisms.

Thanks to the Department of Mathematics and Statistics for giving me the financial support through the graduate assistantships. I am also thankful to the Faculty of Graduate Studies of the University of Windsor for providing me financial support throughout my study. Thanks are also due to the Ontario Ministry of Training and Colleges for the Ontario Graduate Scholarship (OGS). Many thanks to my fellow graduate students. I am indebted to my wife and children for their patience.

Contents

List of Tables

Chapter 1

Introduction

Clustered (includes longitudinal) count data arise in many bio-statistical practices in which a number of repeated responses and a set of covariates are observed (may be over time) from a number of individuals. For example, in health care utilization, the number of visits to the physician by a number of independent individuals may be recorded over a period of several years. Also, information on covariates, for example, gender, number of chronic conditions, educational level, age etc. may be recorded for each individual. A similar example of clustered count data is given by Gadda and Commenges (1995). As compared to cross-sectional studies where we collect data on the individuals on one occasion only, clustered or longitudinal study provides repeated measurements on the same subject (may be over time), thus enhancing the study of assessing within-subject changes in the response variable.

An important problem, in these situations, is to test homogeneity of the repeated observations within clusters (individuals) and also between clusters. Jacqmin-Gadda and Commenges (1995) develop a score test of homogeneity H_S by assuming a random intercept model within the framework of the generalized linear mixed model for clustered responses by obtaining the exact variance of the likelihood score under the null hypothesis of homogeneity and a score test statistic H_T using the generalized estimating equation (GEE) approach (Liang and Zeger, 1986) . They further show that the two tests are identical when the covariance matrix assumed in the GEE approach is that of the random intercept model. The random intercept model not only reflects the natural heterogeneity across individuals or clusters, but also accounts for within individual correlation. In each of these cases they deal with (a) the situation in which the dispersion parameter ϕ is assumed to be known and (b) the situation in which the dispersion parameter ϕ is assumed to be unknown. Note that Jacqmin-Gadda and Commenges (1995) deal with the generalized linear mixed model which is applicable to discrete or continuous response variable as long as the distribution belongs to the exponential family.

In this dissertation we deal with homogeneity testing for clustered (longitudinal) count data with over-dispersion. As pointed out earlier, Jacqmin-Gadda and Commenges (1995) deal with the situation in which the dispersion parameter ϕ is known as well as the situation in which the dispersion parameter ϕ is unknown. However, the second situation is more realistic as ϕ will be unknown in practice. Over-dispersion is a common feature in longitudinal or clustered discrete data, that is, data show more variation than is accounted for by the common discrete distributions (Poisson, binomial). For example, in many biomedical applications count data have variability that far exceeds that predicted by the Poisson distribution.

For the situation in which the dispersion parameter is unknown we first obtain a specific formula H_S for count data. We then use the score test procedure of Rao (1947) and derive three score tests by assuming a random intercept model within the framework of (i) the over-dispersed generalized linear model, (ii) the negative binomial model and (iii) the double extended quasi likelihood model (Lee and Nelder, 2001). All these three statistics are much simpler than the statistic H_S. The second of the latter three statistics takes over-dispersion more directly into the model and therefore is expected to do well when the model assumptions are satisfied and the other statistics are expected to be robust.

We begin Chapter 2 by reviewing generalized linear models (GLM), procedures for the maximum likelihood estimation of the parameters and extension of GLM to clustered or longitudinal data. A brief discussion is given for the generalized linear mixed effects model in general and generalized linear mixed effects model for count data in particular. The negative binomial, quasi-likelihood, extended quasi-likelihood and double extended quasi-likelihood for the estimation of the regression parameters are also discussed in this chapter. Further, a review of the empirical Bayes estimation is given.

In Chapter 3 we first obtain a specific formula H_S for count data and then we obtain the three score tests discussed above by using the score test procedure of Rao (1947). Simulations to compare level and power of the four statistics are performed. Some examples are also given.

In Chapter 4 we develop two score tests for over-dispersion in the generalized linear mixed effects model. One of these is based on the over-dispersed generalized linear mixed effects model of Cox (1983) and the other is based on the negative binomial mixed effects model. Some simulations are conducted.

A summary and some concluding remarks are given in Chapter 5 and a plan for future study is given in Chapter 6.

Chapter 2

A Review of Current Literature

2.1 Generalized linear models (GLM)

In this section we review the GLM using McCullagh and Nelder (1989).

The GLM extends ordinary regression models to include non-normal response distributions. Three components specify a generalized linear model

a) a random component identifies the response variable Y and its probability distribution,

b) a systematic component specifies explanatory variables used in a linear predictor function

   and

c) a link function specifies the function $E(Y)$ that the model equates to the systematic component.

2.1.1 Components of generalized linear models

The random component of a GLM consists of a response variable Y with independent observations $(y_1, y_2, \ldots, y_n)$ from a distribution in the natural exponential family. This family has probability density function or mass function of the form

$$f(y_i;\theta_i,\phi) = \exp\left[\phi^{-1}\{y_i\theta_i - b(\theta_i)\} + C(y_i;\phi)\right], \tag{2.1}$$

where ϕ is constant. The constant ϕ can be assumed to be known or a parameter to be estimated. This is said to be in canonical form with canonical or natural parameter θ. The value of the parameter θ_i may vary for $i = 1, 2, \ldots, n$ depending on values of the explanatory variables.

The systematic component relates a vector $(\eta_1, \eta_2, \ldots, \eta_n)$ to the explanatory variables through a linear model. Let x_{ij} denote the value of predictor j $(j = 1, 2, \ldots, p)$ for subject i. Then

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}.$$

This linear combination of explanatory variables is called the linear predictor. Usually, we take $x_{i1} = 1$ for all i. Then the coefficient β_1 is the intercept of the model.

The link function connects the random and the systematic components of the model. Let $\mu_i = E(Y_i),\ i = 1, 2, \ldots, n$. The model links μ_i to η_i by $\eta_i = g(\mu_i)$, where the link function g is a monotonic, differentiable function. Thus, g links $E(Y_i)$ to the explanatory variables through the formula

$$g(\mu_i) = \sum_{j=1}^{p} \beta_j x_{ij}, \quad i = 1, 2, \ldots, n.$$

The link function $g(\mu) = \mu$, called the identity link, has $\eta_i = \mu_i$. It specifies a linear model for the mean itself. The link function that transforms the mean to the natural parameter is called the canonical link. For this, $g(\mu_i) = \theta_i$ and $\theta_i = \sum_{j=1}^{p} \beta_j x_{ij}$.

The mean and the variance of Y are $E(Y) = b'(\theta)$ and $\text{Var}(Y) = \phi b''(\theta)$ respectively. Many known distributions such as the normal, Poisson and the binomial belong to the exponential family. For the normal distribution $\phi = \sigma^2$, for the Poisson distribution $\phi = 1$ and for the binomial distribution $\phi = 1/n$, where n is the binomial index. For over-dispersed count data, ϕ can be considered as an over-dispersion parameter to be estimated from the data.

2.1.2 Estimation of the parameters of the GLM

The log-likelihood in terms of the canonical parameter θ is

$$l(\theta, \phi; y) = \sum_{i=1}^{n} \left[\phi^{-1}\{y_i\theta_i - b(\theta_i)\} + C(y_i; \phi)\right].$$

We want to estimate $\beta_1, \beta_2, \ldots, \beta_p$. After detailed derivation it can be shown that

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \left[\frac{y_i - b'(\theta_i)}{\phi} \frac{1}{b''(\theta_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}\right], \quad j = 1, 2, \ldots, p.$$

Recall that $\text{Var}(Y_i) = \phi b''(\theta_i) = V_i$, say. Thus

$$\begin{aligned} u_j = \frac{\partial l}{\partial \beta_j} &= \sum_{i=1}^{n} \frac{y_i - b'(\theta_i)}{V_i} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \\ &= \sum_{i=1}^{n} w_i (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i} x_{ij}, \end{aligned}$$

where $w_i = \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 V_i^{-1}$ and $\mu_i = E(Y_i) = b'(\theta_i)$. This is a convenient form to solve. Thus we solve the p equations

$$u_j = 0, \; j = 1, 2, \ldots p,$$

simultaneously for β_j. The equations are non-linear in nature in β_j, so we must solve them iteratively. Two methods, the Newton-Raphson and the Fisher scoring methods, described below are available to solve these equations.

2.1.3 Newton-Raphson method

Let us denote $u = (u_1(\beta), u_2(\beta), \ldots, u_p(\beta))' = \left(\frac{\partial l}{\partial \beta_1}, \frac{\partial l}{\partial \beta_2}, \ldots, \frac{\partial l}{\partial \beta_p}\right)'$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$. Thus $\frac{\partial u}{\partial \beta} = \left\{\frac{\partial^2 l}{\partial \beta_j \partial \beta_k}\right\}$ is the matrix of second derivatives of l. If $\beta^{(t)}$ is the estimate of β at the tth iteration, then the estimate of β at the *(t+1)*th iteration is

$$\beta^{(t+1)} = \beta^{(t)} - \left(\frac{\partial^2 l}{\partial \beta_j \partial \beta_k}\right)^{-1} u^{(t)}, \quad j, k = 1, 2, \ldots, p,$$

for $t = 0, 1, 2, \ldots$. The $p \times p$ matrix $\left(\frac{\partial^2 l}{\partial \beta_j \partial \beta_k}\right)^{-1}$ on the right hand side of the above equation is also to be evaluated at the tth iteration.

2.1.4 Fisher scoring method

A better method which often simplifies the expression is to replace $\frac{\partial u}{\partial \beta} = \left\{\frac{\partial^2 l}{\partial \beta_j \partial \beta_k}\right\}$ by its expected value. This is called Fisher's Scoring method. We know that

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{V_i} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}.$$

Then,

$$\begin{aligned} E\left(\frac{\partial^2 l}{\partial \beta_j \partial \beta_k}\right) &= E\left[\frac{\partial}{\partial \beta_k} \sum_{i=1}^{n} \frac{y_i - \mu_i}{V_i} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}\right] \\ &= -\sum_{i=1}^{n} w_i x_{ij} x_{ik}, \end{aligned}$$

where w_i is defined above. Now, denote

$$I = -E\left(\frac{\partial^2 l}{\partial \beta_j \partial \beta_k}\right)_{p \times p} = \sum_{i=1}^{n} w_i x_{ij} x_{ik} = (X'WX)_{p \times p},$$

where

$$X = \begin{pmatrix} x_{11} & x_{21} & \dots & x_{p1} \\ x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \dots & x_{pn} \end{pmatrix} \text{ and } W = \begin{bmatrix} w_1 & & & 0 \\ & w_2 & & \\ & & \ddots & \\ 0 & & & w_n \end{bmatrix}.$$

Note that W is an $n \times n$ diagonal matrix and X is a $n \times p$ covariate matrix. Again

$$u_j = \sum_{i=1}^{n} w_i(y_i - \mu_i)\frac{\partial \eta_i}{\partial \mu_i}\ x_{ij}.$$

Thus,

$$u = X'W(Y - \mu)\frac{\partial \eta}{\partial \mu}.$$

Therefore, by Fisher's scoring method, we have

$$\begin{aligned} \beta^{(t+1)} &= \beta^{(t)} + I^{-1}u^{(t)} \\ \Rightarrow\ I^{(t)}\beta^{(t+1)} &= I^{(t)}\beta^{(t)} + u^{(t)} \\ \Rightarrow\ (X'WX)\beta^{(t+1)} &= (X'WX)\beta^{(t)} + X'W(Y - \mu)\frac{\partial \eta}{\partial \mu} \\ &= X'W\left[X\beta + (Y - \mu)\frac{\partial \eta}{\partial \mu}\right] \\ &= X'WZ \\ \Rightarrow\ \beta^{(t+1)} &= \left[(X'WX)^{-1}X'WZ\right]^{(t)}, \end{aligned}$$

where $Z = X\beta + (Y - \mu)\frac{\partial \eta}{\partial \mu}$. This is weighted least squares estimate of β obtained by regressing Z on X with weight matrix W. This is also called the Iteratively Re-weighted Least Squares (IRLS) method.

2.1.5 Estimation of the parameters of a GLM for counts

The log-likelihood function for Poisson data is given by

$$l(\mu, \phi; y) = \sum_{i=1}^{n} \left[\phi^{-1}\{y_i \log \mu_i - \mu_i\} - C(y_i)\right].$$

Assuming a log-linear model $\eta_i = \log \mu_i = X'\beta$, we get $\partial\eta_i/\partial\mu_i = \mu_i^{-1}, w_i = \phi^{-1}\mu_i, u_j = \phi^{-1}\sum_{i=1}^{n}\mu_i\frac{y_i-\mu_i}{\mu_i}x_{ji}$ and $I = \phi^{-1}\sum_{i=1}^{n}\mu_i x_{ji}x_{ki} = \phi^{-1}(X'WX)$, where $W = \text{diag}(\mu_i)$. Therefore, following the procedure in Section 2.1.4, the Fisher scoring equation for obtaining the maximum likelihood estimates of the parameters β with $\phi = 1$ is

$$\beta^{(t+1)} = \left[(X'WX)^{-1}X'WZ\right]^{(t)}, \quad t = 1, 2, 3, \ldots,$$

where $Z = X\beta + \frac{y-\mu}{\mu}$.

2.2 Generalized linear models for clustered data

Let Y_{ij} be the response variable for the j^{th} observation in the i^{th} group, $j = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, k$, with $N = \sum_{i=1}^{k} n_i$. The probability density function of Y_{ij} is a member of the exponential family defined as follows

$$f(Y_{ij}; \theta_{ij}, \phi) = \exp\{\phi^{-1}[Y_{ij}\ \theta_{ij} - b(\theta_{ij})] + C(Y_{ij}, \phi)\}. \quad (2.2)$$

If θ_{ij} is a linear combination of a vector of explanatory variables, then (2.2) specifies a generalized linear model (GLM), where θ_{ij} is the canonical parameter, $\theta_{ij} = \eta_{ij} = g(\mu_{ij})$ is the link function and ϕ is the dispersion parameter. The mean and variance of Y_{ij} are $\mu_{ij} = E(Y_{ij}) = b'(\theta_{ij})$ and $\sigma_{ij}^2 = \text{var}(Y_{ij}) = \phi\ b''(\theta_{ij})$, where $'$ denotes differentiation with respect to θ.

2.2.1 Data layout

In the clustered or longitudinal data setting we assume that the responses are observed repeatedly or followed over time for each of k individuals. The subjects may not have the same number of repeated observations or the repeated observations may not be taken at a common set of occasions. Thus, we assume that there are n_i repeated measurements of the response on the i^{th} subject. The layout of the response count data along with the p covariates can be represented as in Table 2.1.

Table 2.1: Representation of Clustered or Longitudinal Data

	Response				Covariates					
Subject	1	2	$\dots$	n_i	X		1	2	$\dots$	n_i
1	y_{11}	y_{12}	$\dots$	y_{1n_i}	X_{1j1}	:	x_{111}	x_{121}	$\dots$	x_{1n_i1}
					X_{1j2}	:	x_{112}	x_{122}	$\dots$	x_{1n_i2}
					$\vdots$		$\vdots$			$\vdots$
					X_{1jp}	:	x_{11p}	x_{12p}	$\dots$	x_{1n_ip}
2	y_{21}	y_{22}	$\dots$	y_{2n_i}	X_{2j1}	:	x_{211}	x_{221}	$\dots$	x_{2n_i1}
					X_{2j2}	:	x_{212}	x_{222}	$\dots$	x_{2n_i2}
					$\vdots$		$\vdots$			$\vdots$
					X_{2jp}	:	x_{21p}	x_{22p}	$\dots$	x_{2n_ip}
$\vdots$	$\vdots$			$\vdots$	$\vdots$		$\vdots$			$\vdots$
k	y_{k1}	y_{k2}	$\dots$	y_{kn_i}	X_{kj1}	:	x_{k11}	x_{k21}	$\dots$	x_{kn_i1}
					X_{kj2}	:	x_{k12}	x_{k22}	$\dots$	x_{kn_i2}
					$\vdots$		$\vdots$			$\vdots$
					X_{kjp}	:	x_{k1p}	x_{k2p}	$\dots$	x_{kn_ip}

Note that $Y_{ij} = g(\sum_k X_{ijk}\beta_k),\ k = 1, 2, \dots, p$.

2.3 Generalized linear mixed effects models (GLMM)

The GLMM is reviewed here following Fitzmaurice, Laird and Ware (2004). Conditional on the random effects, we assume that the responses for any particular individual are independent observations from a distribution belonging to the exponential family (e.g., the Poisson distribution if Y_{ij} is a count or the Bernoulli distribution if Y_{ij} is binary). We can formulate the generalized linear mixed effects model (GLMM) by the following specifications

1. The conditional distribution of each Y_{ij}, given a $q \times 1$ vector of random effects α_i, is assumed to be a member of the exponential family (2.2) with $\text{var}(Y_{ij}|\alpha_i) = \phi\ v\{E(Y_{ij}|\alpha_i)\}$, where $v(.)$ is a known variance function, a function of the conditional mean $E(Y_{ij}|\alpha_i)$. Also given the random effects α_i, the Y_{ij}'s are assumed to be independent of each other (conditionally independent).

2. We assume that the conditional mean of Y_{ij} depends on both fixed and random effects through the following linear predictor

$$\begin{aligned} \theta_{ij} = \eta_{ij} &= \mathbf{X}_{ij}^T\beta + Z_{ij}\ \alpha_i \\ &= g(\mu_{ij}) \end{aligned} \tag{2.3}$$

   for some known function $g(.)$, where β denotes a $p \times 1$ vector of fixed effects with its associated design vector $\mathbf{X}_{ij}$ and α_i is the random subject/cluster effect with the associated covariates Z_{ij}.

3. The random effects follow some probability distribution. In practice, we assume that α_i have a multivariate normal distribution with zero mean and $q \times q$ covariance matrix D. The random effects α_i are assumed to be independent of the covariates $\mathbf{X}_{ij}$.

The three components of a generalized linear mixed model given above completely specify the joint distribution of Y_{ij}. In what follows we explain the above specifications by a model for count data with over-dispersion.

2.3.1 Generalized linear mixed model for counts

Suppose that Y_{ij} is a count. An example of a generalized linear mixed model for Y_{ij} (see Fitzmaurice et. al., 2004) follows

a. Conditional on a vector of random effects α_i, the Y_{ij} are independent and have a over-dispersed Poisson distribution with $\text{Var}(Y_{ij}|\alpha_i) = \phi E(Y_{ij}|\alpha_i)$.

b. The conditional mean of Y_{ij} depends on fixed and random effects via the following linear predictor

$$\log\{E(Y_{ij}|\alpha_i)\} = \eta_{ij} = \mathbf{X}_{ij}^T\beta + Z_{ij}\alpha_i.$$

That is, the conditional mean of Y_{ij} is related to the linear predictor by a log link function.

c. The random effects are assumed to follow a multivariate normal distribution with zero mean and a $q \times q$ covariance matrix D.

Note that the random effects α_i vary from cluster (individual) to cluster representing natural heterogeneity among the individuals.

A special case of the generalized linear mixed effects model is the random intercept model in which $Z_{ij} = 1$ for all i and j in which case

$$\theta_{ij} = \mathbf{X}_{ij}^T\beta + \alpha_i,$$

and

$$\log\{E(Y_{ij}|\alpha_i)\} = \mathbf{X}_{ij}^T\beta + \alpha_i.$$

It can be shown that (see also Carrasco and Jover, 2005)

$$\text{Var}(Y_{ij}) = \phi \exp(\mathbf{X}_{ij}^T\beta + \sigma_\alpha^2/2) + \exp(2\mathbf{X}_{ij}^T\beta)\ e^{\sigma_\alpha^2}(e^{\sigma_\alpha^2} - 1)$$

and

$$\text{Cov}(Y_{ij}, Y_{ik}) = \exp(\mathbf{X}_{ij}^T\beta + \mathbf{X}_{ik}^T\beta + \sigma_\alpha^2)\ (e^{\sigma_\alpha^2} - 1),$$

where σ_α^2 is the variance of the random effects. The intra-cluster correlation then is

$$\text{Corr}(Y_{ij}, Y_{ik}) = \frac{\exp(\mathbf{X}_{ij}^T\beta + \mathbf{X}_{ik}^T\beta + \sigma_\alpha^2)\ (e^{\sigma_\alpha^2} - 1)}{\left[\phi \exp(\mathbf{X}_{ij}^T\beta + \sigma_\alpha^2/2) + \exp(2\mathbf{X}_{ij}^T\beta)\ e^{\sigma_\alpha^2}(e^{\sigma_\alpha^2} - 1)\right]}.$$

2.4 The negative binomial regression model

Let Y be the response variable, which is a count, x be a $p \times 1$ vector of explanatory variables and β be a $p \times 1$ vector of regression parameters. The Poisson-gamma relationship produces the negative binomial distribution which is described below. In the absence of covariates, let $Y|\lambda \sim P(\lambda)$ and $\lambda \sim gamma(\alpha, \beta)$, that is,

$$g(\lambda) = \frac{1}{\Gamma(\alpha)\ \beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta}.$$

Then the unconditional distribution of Y is given by

$$\begin{aligned} Pr(Y = y) &= \frac{(\alpha + y - 1)!}{y!\ \Gamma(\alpha)} \left(\frac{1}{\beta+1}\right)^\alpha \left(1 - \frac{1}{\beta+1}\right)^y, \quad y = 0, 1, 2, \ldots \\ &= \frac{\Gamma(\alpha + y)}{y!\ \Gamma(\alpha)} \left(\frac{1}{\beta+1}\right)^\alpha \left(\frac{\beta}{\beta+1}\right)^y \\ &= \frac{\Gamma(y + c^{-1})}{y!\ \Gamma(c^{-1})} \left(\frac{1}{1+c\mu}\right)^{c^{-1}} \left(\frac{c\mu}{1+c\mu}\right)^y, \end{aligned} \tag{2.4}$$

where $E(Y) = \alpha\beta = \mu$, $\text{Var}(Y) = \alpha\beta + \alpha\beta^2 = \mu(1 + c\mu)$ and $c = 1/\alpha$. This is the negative binomial distribution, denoted by $NB(\mu, c)$, with mean μ and over-dispersion parameter c (see Paul and Placket, 1978). Taking into consideration the covariates x, the mean and variance of Y can be written as

$$\begin{aligned} E(Y|x) &= \mu(x) \\ \text{and} \quad \mathrm{Var}(Y|x) &= \mu(x) + c\mu(x)^2 \end{aligned}$$

with $\mu(x) = x\beta$.

Thus, equation (2.4), with covariates present in the model, can be written as

$$Pr(Y = y|x) = \frac{\Gamma(y + c^{-1})}{y!\ \Gamma c^{-1}} \left(\frac{1}{1 + c\mu(x)}\right)^{c^{-1}} \left(\frac{c\mu(x)}{1 + c\mu(x))}\right)^{y}, \quad y = 0, 1, 2, \ldots \qquad (2.5)$$

Maximum likelihood estimation of the parameters of the model (2.5) is discussed in Lawless (1987). For data given in Section 2.2.1 we discuss maximum likelihood estimation in what follows.

We deal with the log-linear model $\log(\mu_{ij}) = X_{ij}^T\beta$. Then $Y_{ij} \sim NB(\mu_{ij}, c)$, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$, where Y_{ij}'s are independent with $\mu_{ij} = \exp(X_{ij}^T\beta)$. The likelihood function is given by

$$L(\beta, c) = \prod_{i=1}^{k}\prod_{j=1}^{n_i} \frac{\Gamma(y_{ij} + c^{-1})}{\Gamma(c^{-1})} \left(\frac{1}{1 + c\mu_{ij}}\right)^{c^{-1}} \left(\frac{c\mu_{ij}}{1 + c\mu_{ij}}\right)^{y_{ij}}.$$

Noting that for any $c > 0$,

$$\Gamma(y + c^{-1}) = (y + c^{-1} - 1)(y + c^{-1} - 2)\ldots(c^{-1} + 1)c^{-1}\Gamma(c^{-1}),$$

we obtain $\Gamma(y + c^{-1})/\Gamma(c^{-1}) = c^{-1}(c^{-1} + 1)\ldots(c^{-1} + y - 1)$. The log-likelihood function can then be written as

$$\begin{aligned}
l(\beta, c) &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}\log\left(\frac{1+cl}{c}\right) + y_{ij}\log(c\mu_{ij}) - (y_{ij}+c^{-1})\log(1+c\mu_{ij})\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}\log(1+cl) - \sum_{l=0}^{y_{ij}-1}\log c + y_{ij}\log c + y_{ij}\log(\mu_{ij})\right. \\
&- \left.(y_{ij}+c^{-1})\log(1+c\mu_{ij})\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}\log(1+cl) + y_{ij}\log(\mu_{ij}) - (y_{ij}+c^{-1})\log(1+c\mu_{ij})\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}\log(1+cl) + y_{ij}X_{ij}^T\beta - (y_{ij}+c^{-1})\log\left(1+ce^{X_{ij}^T\beta}\right)\right].
\end{aligned}$$

If $y_{ij} < 1$ then $\sum_{l=0}^{y_{ij}-1}$ is zero. The first and second derivatives of l, with respect to $\beta_s,\ s = 1, 2, \ldots, p$, and c are

$$\begin{aligned}
\frac{\partial l}{\partial \beta_s} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[y_{ij}x_{ijs} - (y_{ij}+c^{-1})\left(\frac{c\mu_{ij}x_{ijs}}{1+c\mu_{ij}}\right)\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[y_{ij}x_{ijs} - (1+cy_{ij})\left(\frac{\mu_{ij}x_{ijs}}{1+c\mu_{ij}}\right)\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{x_{ijs}(y_{ij}-\mu_{ij})}{1+c\mu_{ij}}, \quad s = 1, 2, \ldots, p, \qquad (2.6)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial l}{\partial c} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}\left(\frac{l}{1+cl}\right) - (y_{ij}+c^{-1})\left(\frac{\mu_{ij}}{1+c\mu_{ij}}\right) + c^{-2}\log(1+c\mu_{ij})\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}\left(\frac{l}{1+cl}\right) + c^{-2}\log(1+c\mu_{ij}) - (y_{ij}+c^{-1})\left(\frac{\mu_{ij}}{1+c\mu_{ij}}\right)\right],
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 l}{\partial\beta_s\partial\beta_r} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ijs}\frac{(1+c\mu_{ij})(-\mu_{ij}x_{ijr})-(y_{ij}-\mu_{ij})c\mu_{ij}x_{ijr}}{(1+c\mu_{ij})^2} \\
&= -\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{(1+cy_{ij})\mu_{ij}x_{ijs}x_{ijr}}{(1+c\mu_{ij})^2} \quad r,s=1,2,\ldots,p, \\
\frac{\partial^2 l}{\partial\beta_s\partial c} &= -\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{(y_{ij}-\mu_{ij})\mu_{ij}x_{ijs}}{(1+c\mu_{ij})^2}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2 l}{\partial c^2} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1} -\frac{l^2}{(1+cl)^2} + c^{-2}\frac{\mu_{ij}}{1+c\mu_{ij}} + \log(1+c\mu_{ij})(-2)c^{-3} \right.\\
&\qquad \left. - \mu_{ij}\frac{(1+c\mu_{ij})(-1)c^{-2}-(y_{ij}+c^{-1})\mu_{ij}}{(1+c\mu_{ij})^2}\right] \\
&= -\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}\left(\frac{l}{1+cl}\right)^2 + 2c^{-3}\log(1+c\mu_{ij}) - c^{-2}\frac{\mu_{ij}}{1+c\mu_{ij}} \right.\\
&\qquad \left. - \mu_{ij}\frac{c^{-2}(1+c\mu_{ij})+\mu_{ij}(y_{ij}+c^{-1})}{(1+c\mu_{ij})^2}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= -\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}\left(\frac{l}{1+cl}\right)^2 + 2c^{-3}\log(1+c\mu_{ij}) - \mu_{ij}\left\{\frac{2c^{-2}(1+c\mu_{ij})+\mu_{ij}(y_{ij}+c^{-1})}{(1+c\mu_{ij})^2}\right\}\right] \\
&= -\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}\left(\frac{l}{1+cl}\right)^2 + 2c^{-3}\log(1+c\mu_{ij}) - \frac{2c^{-2}\mu_{ij}}{1+c\mu_{ij}} - \frac{\mu_{ij}^2(y_{ij}+c^{-1})}{(1+c\mu_{ij})^2}\right].
\end{aligned}
$$

2.4.1 Fisher information matrix

The Fisher information matrix $I(\beta, c)$ is obtained by taking expectations of minus the second derivatives which are given below.

$$
\begin{aligned}
E\left(-\frac{\partial^2 l}{\partial\beta_s\partial\beta_r}\right) &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\mu_{ij}(1+c\mu_{ij})}{(1+c\mu_{ij})^2}x_{ijs}x_{ijr} \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\mu_{ij}}{1+c\mu_{ij}}x_{ijs}x_{ijr}, \quad s,r=1,2,\ldots,p
\end{aligned}
$$

and

$$E\left(-\frac{\partial^2 l}{\partial\beta_s\partial c}\right) = 0, \quad s = 1, 2, \ldots, p.$$

To get the expression for $E\left(-\partial^2 l/\partial c^2\right)$, we redefine $\tau = c^{-1}$, so the probability function can be written as

$$\begin{aligned}
Pr(Y = y) &= \frac{\Gamma(y+\tau)}{y!\Gamma(\tau)}\left(\frac{1}{1+\mu/\tau}\right)^{\tau}\left(\frac{\mu/\tau}{1+\mu/\tau}\right)^{y} \\
&= \frac{\Gamma(y+\tau)}{y!\Gamma(\tau)}\left(\frac{\tau}{\mu+\tau}\right)^{\tau}\left(\frac{\mu}{\mu+\tau}\right)^{y} \\
&= \frac{\Gamma(y+\tau)}{y!\Gamma(\tau)}\,\tau^{\tau}\mu^{y}\left(\frac{1}{\mu+\tau}\right)^{y+\tau}, \quad y = 0, 1, 2, \ldots.
\end{aligned}$$

Now, since $\Gamma(y+\tau)/\Gamma(\tau) = \tau(\tau+1)\ldots(\tau+y-1)$, then $\log\left(\frac{\Gamma(y+\tau)}{\Gamma(\tau)}\right) = \sum_{l=0}^{y-1}\log(\tau+l)$. The log-likelihood function for β and τ can then be written as

$$l(\beta,\tau) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}\log(\tau+l) + \tau\log\tau - (y_{ij}+\tau)\log(\mu_{ij}+\tau) + y_{ij}\log\mu_{ij} - \log(y_{ij}!)\right].$$

The partial derivatives of $l(\beta,\tau)$ with respect to τ are

$$\begin{aligned}
\frac{\partial l}{\partial\tau} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}\left(\frac{1}{\tau+l}\right) + \frac{\tau}{\tau} + \log\tau - \left\{\frac{y_{ij}+\tau}{\mu_{ij}+\tau} + \log(\mu_{ij}+\tau)\right\}\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}(\tau+l)^{-1} + 1 + \log\tau - \left\{\frac{y_{ij}+\tau}{\mu_{ij}+\tau} + \log(\mu_{ij}+\tau)\right\}\right]
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 l}{\partial\tau^2} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[-\sum_{l=0}^{y_{ij}-1}(\tau+l)^{-2} + \frac{1}{\tau} - \left\{\frac{(\tau+\mu_{ij}) - (y_{ij}+\tau)}{(\tau+\mu_{ij})^2} + \frac{1}{\tau+\mu_{ij}}\right\}\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[-\sum_{l=0}^{y_{ij}-1}(\tau+l)^{-2} + \frac{1}{\tau} - \frac{\mu_{ij}-y_{ij}}{(\tau+\mu_{ij})^2} - \frac{1}{\tau+\mu_{ij}}\right] \\
&= -\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}(\tau+l)^{-2} - \frac{\mu_{ij}}{\tau(\mu_{ij}+\tau)} - \frac{y_{ij}-\mu_{ij}}{(\mu_{ij}+\tau)^2}\right].
\end{aligned}$$

Thus

$$E\left(-\frac{\partial^2 l}{\partial \tau^2}\right) \;=\; \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[E\sum_{l=0}^{y_{ij}-1}(\tau+l)^{-2} - \frac{\mu_{ij}}{\tau(\tau+\mu_{ij})}\right].$$

Now, by noting that $E(-\partial^2 l/\partial c^2) = c^{-4}E(-\partial^2 l/\partial \tau^2)$, the *(ij)*th term of this expectation is equal to

$$c^{-4}\left(\sum_{l=0}^{\infty}(c^{-1}+l)^{-2}Pr(Y_{ij}\geq l) - \frac{c\mu_{ij}}{\mu_{ij}+c^{-1}}\right).$$

Therefore,

$$E\left(-\frac{\partial^2 l}{\partial c^2}\right) = c^{-4}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{\infty}\left(\frac{c}{1+cl}\right)^2 Pr(Y_{ij}\geq l) - \frac{c^2\mu_{ij}}{1+c\mu_{ij}}\right].$$

Following Fisher (1941) and Collings(1981), this equation can be simplified as

$$E\left(-\frac{\partial^2 l}{\partial c^2}\right) = c^{-4}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=0}^{\infty}\frac{l!(cq_{ij})^{l+1}}{(l+1)d_l} - \frac{c^2\mu_{ij}}{1+c\mu_{ij}}\right]$$

$$= \; c^{-4}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=1}^{\infty}\frac{l!(cq_{ij})^{l+1}}{(l+1)d_l} + \frac{c^2\mu_{ij}}{1+c\mu_{ij}} - \frac{c^2\mu_{ij}}{1+c\mu_{ij}}\right].$$

Thus,

$$E\left(-\frac{\partial^2 l}{\partial c^2}\right) = c^{-4}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=1}^{\infty}\frac{l!(cq_{ij})^{l+1}}{(l+1)d_l}\right],$$

where $q_{ij} = c\mu_{ij}/(1+c\mu_{ij})$ and $d_l = \prod_{j=1}^{l}(1+jc)$.

2.4.2 Fisher scoring method for the estimation of β and the dispersion parameter c

Define the function $\eta_{ij} = \log \mu_{ij}$. Then, $\partial\eta_{ij}/\partial\mu_{ij} = 1/\mu_{ij}$. Also let $V_{ij} = Var(Y_{ij}) = \mu_{ij} + c\mu_{ij}^2$. Further we define

$$\begin{aligned} w_{ij} &= \left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 V_{ij}^{-1} \\ &= \mu_{ij}^2 . \frac{1}{\mu_{ij}(1+c\mu_{ij})} \\ &= \frac{\mu_{ij}}{1+c\mu_{ij}}. \end{aligned}$$

Then the score equation (2.6) for $\beta_s,\ s = 1, 2, \ldots, p$, can be written as

$$\begin{aligned} u_s &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} w_{ij}(y_{ij}-\mu_{ij})\frac{\partial\eta_{ij}}{\partial\mu_{ij}} x_{ijs} \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(\frac{\mu_{ij}}{1+c\mu_{ij}}\right)(y_{ij}-\mu_{ij})\frac{1}{\mu_{ij}} x_{ijs} \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(\frac{\mu_{ij}}{1+c\mu_{ij}}\right)\left(\frac{y_{ij}-\mu_{ij}}{\mu_{ij}}\right) x_{ijs} \end{aligned}$$

and the Fisher information matrix for β is

$$I = \left[E\left(-\frac{\partial^2 l}{\partial\beta_s\partial\beta_r}\right)\right] = \left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} w_{ij}x_{ijs}x_{ijr}\right], \quad s, r = 1, 2, \ldots, p.$$

Now, define $u = (u_1, u_2, \ldots, u_p)'$, $Y = (y_{11}, \ldots, y_{1n_1}, \ldots, y_{k1}, \ldots, y_{kn_k})'$, $\mu = (\mu_{11}, \ldots, \mu_{1n_1}, \ldots, \mu_{k1}, \ldots, \mu_{kn_k})'$ and $N = \sum_{i=1}^{k} n_i$. Further, let X be a $N \times p$ matrix with elements $x_{ijs},\ i = 1, 2, \ldots, k,\ j = 1, 2, \ldots, n_i,\ s = 1, 2, \ldots, p$ and W is a $N \times N$ diagonal matrix with elements w_{ij}. Then the score equations in vector notation can be written as

$$u = X^T W \left(\frac{Y-\mu}{\mu}\right)$$

and the Fisher information matrix can be written as

$$I = (X^T W X).$$

The Fisher Scoring equations for solving for the regression parameters β become

$$\begin{aligned}
I^{(t)}\beta^{(t+1)} &= I^{(t)}\beta^{(t)} + u^{(t)} \\
\Rightarrow \quad \left(X^T W X\right)\beta^{(t+1)} &= \left(X^T W X\right)\beta^{(t)} + u^{(t)} \\
&= \left(X^T W X\right)\beta^{t} + X^T W \left(\frac{Y-\mu}{\mu}\right) \\
&= X^T W \left[X\beta + \frac{Y-\mu}{\mu}\right] \\
&= X^T W Z.
\end{aligned}$$

Thus

$$\beta^{(t+1)} = \left[\left(X^T W X\right)^{-1}\left(X^T W Z\right)\right]^{(t)}, \quad t = 0, 1, 2, \ldots \tag{2.7}$$

with $Z = X\beta + \frac{Y-\mu}{\mu}$.

The Fisher scoring equation to solve for c is

$$c^{(t+1)} = c^{(t)} + I_c^{-1} v^{(t)}, \quad t = 0, 1, 2, \ldots \tag{2.8}$$

where $I_c = E\left(-\partial^2 l/\partial c^2\right)$ and $v = \partial l/\partial c$ as defined before.

Maximum likelihood estimates of β and c are obtained by iterating between equations (2.7) and (2.8) after putting in initial values.

2.5 Cumulants of the negative binomial distribution

The moment generating function (mgf) of the distribution (2.4) is given by

$$M_Y(t) = \left(1 + c\mu - c\mu e^t\right)^{-c^{-1}}.$$

Then the cumulant generating function is $\psi(t) = -c^{-1}\log(1+c\mu-c\mu e^t)$. The cumulants can be derived directly from

$$K_i = \left.\frac{\partial^{(i)}\psi(t)}{\partial t^{(i)}}\right|_{t=0}, \quad i = 1, 2, \ldots$$

which are as follows

$$\begin{aligned} K_1 &= \mu. \\ K_2 &= \mu + c\mu^2. \\ K_3 &= \mu + 3c\mu^2 + 2c^2\mu^3. \\ K_4 &= \mu + 7c\mu^2 + 12c^2\mu^3 + 6c^3\mu^4. \end{aligned}$$

2.6 Quasi-likelihood (QL) and the extended quasi-likelihood (EQL)

In many applications the full distributional assumptions of the GLM cannot be justified. To avoid the full distributional assumptions, Wedderburn (1974) proposes a quasi-likelihood (QL) model which is based on the knowledge of the first two moments of the random variable Y. The quasi-likelihood for a single data point y is defined as

$$Q(y;\mu) = \int_y^{\mu} \frac{(y-t)}{\phi V(t)} dt,$$

where ϕ is a known constant or a parameter to be estimated and $V(t)$ is a variance function. In the framework of the generalized linear model ϕ can be considered as an over-dispersion parameter and a moment estimate of ϕ can be obtained. Note that the variance of the random variable Y is assumed to be $\text{Var}(Y) = \phi V(\mu)$ and note further that a maximum quasi-likelihood estimate for the dispersion parameter ϕ cannot be obtained as the quasi-likelihood is used only for the estimation of the β parameters.

In many real life applications a variance function of the form $\text{Var}(Y) = \phi V(\mu)$ is not suitable. For example, the negative binomial variance is $\text{Var}(Y) = \mu(1 + c\mu)$, where μ is the mean and c is the over-dispersion parameter. For this variance function a quasi-likelihood can be defined with $\phi = 1$ and $V(\mu) = \mu(1 + c\mu)$. However, such a quasi-likelihood does not facilitate estimation of the over-dispersion parameter c. In this situation, for the joint estimation of the mean and dispersion parameters, Nelder and Pregibon (1987) and Godambe and Thompson (1989) suggest an Extended Quasi-likelihood (EQL). The EQL is given by

$$Q^{+}(y; \mu, \phi) = \left[-\frac{1}{2} \ln[2\pi\phi V(y)] + Q(y; \mu) \right].$$

The second term on the right hand side of the above equation is the QL for y and the first term is the normalizing factor; thus making $\exp(Q^{+})$ resemble a log-likelihood. The EQL Q^{+} can then be used to estimate the mean (regression) parameters and the over-dispersion parameter. The advantage of using Q^{+} for estimating the over-dispersion parameter is that it can be robust to the Maximum Likelihood (ML) estimate as the full distributional assumptions are not required, yet Q^{+} behaves like a log-likelihood.

2.6.1 Double extended quasi-likelihood (DEQL)

In generalized linear models, the variance function characterizes the family of distributions. Thus, a quasi-generalized linear model is characterized by the first two moments, specified by $(V(.), g(.))$. Quasi-likelihood allows inferences for mean (regression) parameters for models having arbitrary variance functions. For estimation of dispersion parameters for such models, Nelder and Pregibon (1987) propose the extended quasi-likelihood (EQL), alternatively defined as

$$-2Q^{+} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} [d_{ij}/\phi + \log 2\pi\phi V(y_{ij})],$$

where d_{ij} is the deviance component given by

$$d_{ij} = 2\int_{\mu_{ij}}^{y_{ij}} \frac{(y_{ij} - s)}{V(s)} ds.$$

Lee and Nelder (2001) introduce double-extended quasi-likelihood (DEQL) also for the joint estimation of the mean and the dispersion parameters. The DEQL methodology requires an EQL for Y_{ij} given some random effect α_i and an EQL for α_i from a conjugate distribution given some mean parameter μ and dispersion parameter ϕ. The DEQL is then obtained by combining the two EQL's. The random effect α_i's or some transformed variables s_i are then replaced by their maximum likelihood estimates resulting in a profile DEQL.

To form the DEQL we first define the following Hierarchical Generalized Linear Models (HGLM) (see Lee and Nelder (2001)):

i) $y_{ij}|\alpha_i \sim$ GLM with $E(y_{ij}|\alpha_i) = \alpha_i = \mu_{0ij}$ and $\text{var}(y_{ij}|\alpha_i) = \phi V_0(\mu_{0ij}) = \mu_{0ij}$ with $\phi = 1$, and

ii) $\alpha_i \sim$ GLM with $\theta(m) = \ln(m), b(\theta(m)) = e^{\theta(m)}, E(\alpha_i) = b'(\theta(m)) = e^{\theta(m)} = m, \text{var}(\alpha_i) = \alpha V_1(m)$ with $\alpha = cm$ and $V_1(m) = b''(\theta(m)) = m.$

If we define the deviance components of $y_{ij}|\alpha_i$ by

$$d_{0ij} = 2\int_{\mu_{ij}}^{y_{ij}} \frac{(y_{ij} - s)}{V_0(s)} ds,$$

and the deviance components of α_i by

$$d_{1ij} = 2\int_{\alpha_i}^{\psi} \frac{(\psi - s)}{V_1(s)} ds,$$

the double extended quasi-likelihood can be formulated as $Q^{++} = Q_0^{+}\{\theta(\mu_0), \phi; y|\alpha\} +$

$Q_1^+(\lambda; v_1)$, where

$$-2Q_0^+\{\theta(\mu_0), \phi; y|\alpha\} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[d_{0ij}/\phi + \log\{2\pi\phi V_0(y_{ij})\}\right],$$

and

$$-2Q_1^+(\lambda; v_1) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[d_{1ij}/\lambda + \log\{2\pi\lambda V_1(\psi_1)\} - 2\log\left\{\left|\frac{d\theta(\alpha_i)}{dv_{1i}}\right|\right\}\right].$$

Double extended quasi-likelihood allows not only the extension of models to those with an arbitrary variance function $V_0(\mu_0)$ for $y|\alpha$ with no corresponding generalized model family of distributions, such as the over-dispersed Poisson or binomial, but also provides for the formulation of a quasi-conjugate distribution, characterized entirely by the variance function $V_1(.)$.

Following Paul and Saha (2007), using a modified Stirling approximation, recommended by Lee and Nelder (2001), which is given by

$$\log\Gamma(z) \simeq \left(z - \frac{1}{2}\right)\log(z) + \frac{1}{2}\log(2\pi) - z + \frac{1}{12z},$$

the profile DEQL for count data can be computed as

$$\begin{aligned} p_v^*(Q^{++}) &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[y_{ij}\ln(m) - \left(y_{ij} + \frac{1}{c}\right)\ln(1+cm) + \left(y_{ij} + \frac{1}{c} - \frac{1}{2}\right)\ln(1+cy_{ij})\right. \\ &+ \left.\frac{c}{12(1+cy_{ij})} - \frac{c}{12} - \left(y_{ij} + \frac{1}{2}\right)\ln(y_{ij}) - \frac{1}{12(y_{ij}+1)} - \frac{1}{2}\ln(2\pi)\right]. \quad (2.9) \end{aligned}$$

For over-dispersed count data this profile DEQL is the same as the negative binomial log-likelihood with the factorials replaced by the modified Stirling approximations. Inference for the mean (regression) parameter(s) and the dispersion parameter c can then be made on $p_v^*(Q^{++})$. For more details see Saha (2004) and Paul and Saha (2007).

2.6.2 Estimation of regression and dispersion parameters of the DEQL

Taking into consideration the covariates x, the mean and variance of Y are

$$\begin{aligned} E(Y_{ij}|x) &= \mu_{ij}(x) \\ \text{and} \quad \mathrm{Var}(Y_{ij}|x) &= \mu_{ij}(x) + c\mu_{ij}^2(x). \end{aligned}$$

Using the log link, that is, $\log(\mu_{ij}) = X_{ij}^T\beta$, equation (2.9) can be written as

$$\begin{aligned} p_v^*(Q^{++}) &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\Bigg[y_{ij}X_{ij}^T\beta - \left(y_{ij}+\frac{1}{c}\right)\ln(1+ce^{X_{ij}^T\beta}) + \left(y_{ij}+\frac{1}{c}-\frac{1}{2}\right)\ln(1+cy_{ij}) \\ &+ \frac{c}{12(1+cy_{ij})} - \frac{c}{12} - \left(y_{ij}+\frac{1}{2}\right)\ln(y_{ij}) - \frac{1}{12(y_{ij}+1)} - \frac{1}{2}\ln(2\pi)\Bigg]. \qquad (2.10) \end{aligned}$$

As in Section 2.4.1, the first and second derivatives of $p_v^*(Q^{++})$ with respect to β and c are given by

$$\begin{aligned} \frac{\partial p_v^*(Q^{++})}{\partial \beta_s} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{(y_{ij}-\mu_{ij})}{1+c\mu_{ij}}\, x_{ijs}, \quad s = 1,2,\ldots,p, \\ \frac{\partial^2 p_v^*(Q^{++})}{\partial \beta_s \partial \beta_r} &= -\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{(1+cy_{ij})\mu_{ij}}{(1+c\mu_{ij})^2}\, x_{ijs}x_{ijr}, \quad r,s = 1,2,\ldots,p \end{aligned}$$

and

$$\frac{\partial^2 p_v^*(Q^{++})}{\partial \beta_s \partial c} = -\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{(y_{ij}-\mu_{ij})\mu_{ij}}{(1+c\mu_{ij})^2}\, x_{ijs}, \quad s = 1,2,\ldots,p.$$

To obtain the element of the information matrix pertaining to the dispersion parameter c we redefine $\tau = 1/c$. Therefore, the double extended quasi log-likelihood function of equation (2.10) in terms of τ becomes

$$\begin{aligned} p_v^*(Q^{++}) &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[y_{ij}X_{ij}^T\beta - (y_{ij}+\tau)\ln(\tau + e^{X_{ij}^T\beta}) + \left(y_{ij}+\tau-\frac{1}{2}\right)\ln(y_{ij}+\tau)\right. \\ &+ \left.\frac{1}{12(y_{ij}+\tau)} + \frac{1}{2}\ln(\tau) - \frac{1}{12\tau} - (y_{ij}+\frac{1}{2})\ln(y_{ij}) - \frac{1}{12(y_{ij}+1)} - \frac{1}{2}\ln(2\pi)\right]. \end{aligned} \tag{2.11}$$

Differentiating equation (2.11) with respect to τ, we obtain

$$\begin{aligned} \frac{\partial p_v^*(Q^{++})}{\partial \tau} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[1 + \ln\left(\frac{y_{ij}+\tau}{\mu_{ij}+\tau}\right) - \frac{1}{2(y_{ij}+\tau)} - \frac{1}{12(y_{ij}+\tau)^2} + \frac{1}{2\tau} + \frac{1}{12\tau^2}\right. \\ &- \left.\frac{y_{ij}+\tau}{\mu_{ij}+\tau}\right] \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2 p_v^*(Q^{++})}{\partial \tau^2} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\frac{1}{y_{ij}+\tau} + \frac{1}{2(y_{ij}+\tau)^2} + \frac{1}{6(y_{ij}+\tau)^3} - \frac{1}{2\tau^2} - \frac{1}{6\tau^3} + \frac{y_{ij}-\mu_{ij}}{(\mu_{ij}+\tau)^2}\right. \\ &- \left.\frac{1}{\mu_{ij}+\tau}\right]. \end{aligned}$$

The Fisher information matrix can then be obtained by taking expectations of minus the second derivatives. The elements of the Fisher information matrix for the regression and the dispersion parameters are

$$\begin{aligned} E\left(-\frac{\partial^2 p_v^*(Q^{++})}{\partial \beta_s \partial \beta_r}\right) &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\mu_{ij}}{1+c\mu_{ij}}x_{ijs}x_{ijr}, \quad r,s = 1,2,\ldots,p, \\ E\left(-\frac{\partial^2 p_v^*(Q^{++})}{\partial \beta_s \partial c}\right) &= 0, \quad s = 1,2,\ldots,p \end{aligned}$$

and

$$\begin{aligned}
E\left(-\frac{\partial^2 p_v^*(Q^{++})}{\partial\tau^2}\right) &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[E\left\{\frac{1}{\mu_{ij}+\tau}+\frac{1}{2\tau^2}+\frac{1}{6\tau^3}-\frac{1}{y_{ij}+\tau}-\frac{1}{2(y_{ij}+\tau)^2}\right.\right.\\
&- \left.\left.\frac{1}{6(y_{ij}+\tau)^3}\right\}-\frac{E(y_{ij}-\mu_{ij})}{(\mu_{ij}+\tau)^2}\right]\\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\frac{1}{\mu_{ij}+\tau}+\frac{1}{2\tau^2}+\frac{1}{6\tau^3}\right]-E\left[\frac{1}{y_{ij}+\tau}+\frac{1}{2(y_{ij}+\tau)^2}\right.\\
&+ \left.\frac{1}{6(y_{ij}+\tau)^3}\right].
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\frac{1}{\mu_{ij}+\tau}+\frac{1}{2\tau^2}+\frac{1}{6\tau^3}\right]\\
&- \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{y_{ij}=0}^{\infty}\left(\frac{1}{y_{ij}+\tau}+\frac{1}{2(y_{ij}+\tau)^2}+\frac{1}{6(y_{ij}+\tau)^3}\right)\times f(y_{ij};\mu_{ij},c)\right],
\end{aligned}$$

where $f(y_{ij};\mu_{ij},c)$ is the probability function of the negative binomial distribution. The regression parameter β is estimated by the Fisher scoring equation given in Section 2.4.2. However, to estimate the dispersion parameter c, the Fisher scoring equation is $c^{t+1}=c^t+I^{-1}\left(\frac{\partial p_v^*(Q^{++})}{\partial c}\right)$, where $I=E\left(-\frac{\partial^2 p_v^*(Q^{++})}{\partial c^2}\right)=c^{-4}E\left(-\frac{\partial^2 p_v^*(Q^{++})}{\partial \tau^2}\right)$. Here also the two equations must be solved simultaneously to get the maximum likelihood estimates of the parameters γ and c.

2.7 $\sqrt{k}$ consistent estimators

Let $\{\hat{\theta}_k\}$, $k=1,2,\ldots$, be a sequence of estimators. If the quantity $|\hat{\theta}_k-\theta|\sqrt{k}$ remains bounded in probability as $k\to\infty$, then the sequence of estimates $\hat{\theta}_k$ is called $\sqrt{k}$ consistent estimators (see Lehman, 1999).

Theorem: Let $\hat{\theta}_k$ be a sequence of estimates of θ, and $\text{var}(\hat{\theta}_k)=O\left(\frac{1}{k}\right)$. Then this sequence of estimates is $\sqrt{k}$-consistent.

Proof: By using Chebyshev's inequality, for a given $\epsilon > 0$,

$$P\left(|\hat{\theta}_k - \theta|\sqrt{k} < \epsilon\right) \geq 1 - \frac{\text{var}(\hat{\theta}_k)k}{\epsilon^2}.$$

Let $\hat{\theta}_k$ be the sequence of maximum likelihood estimates (MLE), then by the asymptotic properties of MLE, $\hat{\theta}_k$ is distributed as normal with mean θ and variance $\frac{1}{kI(\theta)}$, where $I(\theta)$ is the Fisher information matrix defined as $I(\theta) = E\left[\frac{\partial}{\partial\theta}\log f(y|\theta)\right]^2$ and the probability density function $f(y|\theta)$ comes from the natural exponential family given by (2.1). Therefore, as $k \to \infty$, $\text{var}(\hat{\theta}_k)$ tends to zero, that is, $\text{var}(\hat{\theta}_k)$ is $O(k^{-1})$. Thus, MLE is $\sqrt{k}$-consistent. The method of moment estimators are also $\sqrt{k}$-consistent estimates (Moore, 1986).

2.8 Empirical Bayes estimation of a parameter θ

Suppose $L(y|\theta)$ is the likelihood for a parameter θ for observations $y = (y_1, y_2, \ldots, y_n)$ from a distribution $f(y|\theta)$. Let $p(\theta|\nu)$ be the prior probability density function of the parameter of interest θ given a hyper-parameter ν. Then the likelihood of the data y is a function of ν which can be written as

$$L(y|\nu) = \int L(y|\theta)p(\theta|\nu)d\theta. \tag{2.12}$$

In the empirical Bayes approach the parameter ν is estimated by maximizing (2.12). Then, the prior distribution of θ is taken as $p(\theta|\hat{\nu})$ and inference about the parameter θ is based on its posterior distribution. The said posterior density is proportional to

$$L(y|\theta)p(\theta|\hat{\nu}) \tag{2.13}$$

and the posterior density of θ given $\hat{\nu}$ is

$$p(\theta|y,\hat{\nu}) = \frac{L(y|\theta)p(\theta|\hat{\nu})}{\int L(y|\theta)p(\theta|\hat{\nu})d\theta}. \tag{2.14}$$

The empirical Bayes estimate of θ can then be taken either as the posterior mode which is obtained by differentiating (2.14) with respect to θ and equating the differential to zero or by taking the posterior mean which is the expected value of $E(\theta|y,\hat{\nu})$ which is

$$E(\theta|y,\hat{\nu}) = \frac{\int \theta L(y|\theta)p(\theta|\hat{\nu})d\theta}{\int L(y|\theta)p(\theta|\hat{\nu})d\theta}. \tag{2.15}$$

In general the posterior mean is difficult to calculate and we need numerical methods such as the Markov chain Monte Carlo (MCMC) method.

Chapter 3

Score Test of Homogeneity for Over-Dispersed Clustered Count Data

Clustered count data arise in many bio-statistical practices in which a number of repeated count responses are observed on a number of individuals. The repeated observations may also represent counts over time from a number of individuals. One important problem that arises in practice is to test homogeneity within clusters (individuals) and between clusters (individuals). As data within clusters are observations of repeated responses, the count data may be correlated and/or over-dispersed. Jacqmin-Gadda and Commenges (1995) derive a score test statistic H_S by assuming a random intercept model within the framework of the generalized linear mixed model by obtaining the exact variance of the likelihood score under the null hypothesis and a score test statistic H_T using the generalized estimating equation (GEE) approach. They further show that the two tests are identical when the covariance matrix assumed in the GEE approach is that of the random-effects model. In each of these cases they dealt with (a) the situation in which the dispersion parameter ϕ is assumed to be known and (b) the situation in which the dispersion parameter ϕ is assumed to be unknown. The second situation, however, is more realistic as ϕ will be unknown in practice.

In this chapter we first obtain a score test of homogeneity for over-dispersed count data with unknown over-dispersion parameter using the score test results of Jacqmin-

Gadda and Commenges (1995). We then use the score test procedure of Rao (1947) and derive three tests by assuming a random intercept model within the framework of (i) the over-dispersed generalized linear model (ii) the negative binomial model, and (iii) the double extended quasi likelihood model (Lee and Nelder (2001). All these three statistics are much simpler than the statistic H_S derived by Jacqmin-Gadda and Commenges (1995) under the framework of the over-dispersed generalized linear model. The second statistic takes over-dispersion more directly into the model and therefore is expected to do well when the model assumptions are satisfied, and the other statistics are expected to be robust. Simulations show superior level property of the statistics derived under the negative binomial and double extended quasi-likelihood model assumptions. Two data sets are analyzed and a discussion is given.

3.1 The score test obtained from Jacqmin-Gadda and Commenges (1995)

Let Y_{ij} denote the j^{th} response in group i, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k,$. Conditionally on a $q \times 1$ vector of random effects α_i, Y_{ij}'s are independently distributed from an over-dispersed exponential family

$$f(Y_{ij}; \theta_{ij}, \phi) = \exp\{\phi^{-1}[Y_{ij}\theta_{ij} - b(\theta_{ij})] + C(Y_{ij}, \phi)\}, \tag{3.1}$$

with mean $\mu_{ij} = E(Y_{ij}|\alpha_i) = b'(\theta_{ij})$, variance $\sigma_{ij}^2 = \text{var}(Y_{ij}|\alpha_i) = \phi\ b''(\theta_{ij})$, where $'$ denotes differentiation with respect to θ and ϕ is the over-dispersion parameter. The mixed-effects model considered by Jacqmin-Gadda and Commenges (1995) is

$$g(\mu_{ij}) = \theta_{ij} = \eta_{ij} = \mathbf{X}_{ij}^T\beta + Z_{ij}\alpha_i, \tag{3.2}$$

for some known function $g(.)$, where $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ denotes a $p \times 1$ vector of fixed effects with its associated design vector $\mathbf{X}_{ij}$ and α_i is the scalar random subject/cluster effect with the associated covariate Z_{ij}. Since we want to test homogeneity across and within groups we consider the random intercept model in which $Z_{ij} = 1$ for all i, j.

The parameter α_i can be written as $\alpha_i = \alpha + D^{1/2} v_i$, where the v_i's are independently and identically distributed with unspecified distribution F with zero mean and unit variance. Therefore, α_i's are iid with mean α and variance D. Our interest is to test $H_0 : D = 0$ against the alternative $H_A : D > 0$. Note that for count data models this is equivalent to testing homogeneity across groups as well as testing homogeneity within groups as the intra-cluster or within group correlation coefficient is

$$\rho = \frac{\exp(\mathbf{X}_{ij}^T\beta + \mathbf{X}_{ik}^T\beta + D)(e^D - 1)}{\left[\phi \exp(\mathbf{X}_{ij}^T\beta + D/2) + \exp(2\mathbf{X}_{ij}^T\beta)e^D(e^D - 1)\right]}.$$

The log-likelihood for group i is

$$l_i(\beta, \alpha, \phi, D) = \log \int \prod_{j=1}^{n_i} f_{ij}(Y_{ij}; \beta, \alpha + D^{1/2} v_i, \phi)\ f(v_i) dv_i. \tag{3.3}$$

The score statistic for testing the hypothesis of homogeneity when the parameters are known is

$$S(\beta, \alpha, \phi) = \sum_{i=1}^{k} \frac{\partial l_i(\beta, \alpha, \phi, D)}{\partial D},$$

which needs to be obtained at $D = 0$. Thus

$$\begin{aligned} \frac{\partial l_i}{\partial D} &= \left[\int \prod_{j=1}^{n_i} f_{ij}(Y_{ij}; \beta, \alpha + D^{1/2} v_i, \phi)\ f(v_i) dv_i\right]^{-1} \\ &\quad \times \frac{\partial}{\partial D} \int \prod_{j=1}^{n_i} f_{ij}(Y_{ij}; \beta, \alpha + D^{1/2} v_i, \phi)\ f(v_i) dv_i. \end{aligned} \tag{3.4}$$

Zhu and Zhang (2006) derived a more general score test statistic than (3.4) for testing homogeneity in mixed effects models.

Following Zhu and Zhang (2006) the second part of equation (3.4) can be written as

$$\left(\frac{\partial\sqrt{D}}{\partial D}\right)\ \frac{\partial}{\partial\sqrt{D}}\int\prod_{j=1}^{n_i} f_{ij}(Y_{ij};\beta,\alpha+D^{1/2}v_i,\phi)\ f(v_i)dv_i.$$

Since $\frac{\partial}{\partial D}\sqrt{D}=\frac{1}{2\sqrt{D}}$, in the limit as $D\to 0^+$ the above equation becomes

$$\frac{1}{2}\lim_{D\to 0^+}\frac{\partial}{\partial\sqrt{D}}\frac{\int\prod_{j=1}^{n_i} f_{ij}(Y_{ij};\beta,\alpha+D^{1/2}v_i,\phi)\ f(v_i)dv_i}{\sqrt{D}}.$$

Using L'Hôpital's rule we get

$$\begin{aligned}
&\frac{1}{2}\lim_{D\to 0^+}\frac{\frac{\partial^2}{\partial^2\sqrt{D}}\int\prod_{j=1}^{n_i} f_{ij}(Y_{ij};\beta,\alpha+D^{1/2}v_i,\phi)\ f(v_i)dv_i}{\frac{\partial}{\partial\sqrt{D}}\sqrt{D}}\\
=\ &\frac{1}{2}\lim_{D\to 0^+}\int\frac{\partial^2}{\partial^2\sqrt{D}}\prod_{j=1}^{n_i} f_{ij}(Y_{ij};\beta,\alpha+D^{1/2}v_i,\phi)\ f(v_i)dv_i. \qquad (3.5)
\end{aligned}$$

Using the following argument

$$\frac{\partial^2 f}{\partial\theta^2}=f\left[\frac{\partial^2\log f}{\partial\theta^2}+\left(\frac{\partial\log f}{\partial\theta}\right)^2\right],$$

for a density function f with a vector of parameters θ equation (3.5) can be written as

$$\frac{1}{2}\lim_{D\to 0^+}\int\prod_{j=1}^{n_i} f_{ij}(Y_{ij};\beta,\alpha+D^{1/2}v_i,\phi)\left[\frac{\partial^2}{\partial^2\sqrt{D}}\log\prod_{j=1}^{n_i} f_{ij}(Y_{ij})+\left\{\frac{\partial}{\partial\sqrt{D}}\log\prod_{j=1}^{n_i} f_{ij}(Y_{ij})\right\}^2\right]. \qquad (3.6)$$

From (3.1) and (3.2) with $Z_{ij}=1$ for all i,j

$$\begin{aligned}
\log f_{ij}(\beta,\alpha,\phi) &= \phi^{-1}\left\{Y_{ij}\theta_{ij}-b(\theta_{ij})\right\}+C(Y_{ij},\phi)\\
&= \phi^{-1}\left\{Y_{ij}\left(\mathbf{X}_{ij}^T\beta+\alpha_i\right)-b\left(\mathbf{X}_{ij}^T\beta+\alpha_i\right)\right\}+C(Y_{ij},\phi)\\
&= \phi^{-1}\left\{Y_{ij}\left(\mathbf{X}_{ij}^T\beta+\alpha+\sqrt{D}v_i\right)-b\left(\mathbf{X}_{ij}^T\beta+\alpha+\sqrt{D}v_i\right)\right\}\\
&\quad + C(Y_{ij},\phi).
\end{aligned}$$

Now,

$$\begin{aligned}\frac{\partial}{\partial\sqrt{D}}\log\prod_{j=1}^{n_i} f_{ij}(Y_{ij}) &= \sum_{j=1}^{n_i}\frac{\partial}{\partial\sqrt{D}}\log f_{ij}(Y_{ij}) \\ &= \phi^{-1}v_i\sum_{j=1}^{n_i}(Y_{ij}-b'(\theta_{ij}))\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial^2\sqrt{D}}\log\prod_{j=1}^{n_i} f_{ij}(Y_{ij}) &= \sum_{j=1}^{n_i}\frac{\partial}{\partial\sqrt{D}}\left\{\frac{\partial}{\partial\sqrt{D}}\log f_{ij}(Y_{ij})\right\} \\ &= -\phi^{-1}v_i^2\sum_{j=1}^{n_i}b''(\theta_{ij}).\end{aligned}$$

Therefore, equation (3.6) becomes

$$\frac{1}{2}\lim_{D\to 0^+}\int\prod_{j=1}^{n_i} f_{ij}(Y_{ij})\left[-\phi^{-1}v_i^2\sum_{j=1}^{n_i}b''(\theta_{ij})+\phi^{-2}v_i^2\left\{\sum_{j=1}^{n_i}(Y_{ij}-b'(\theta_{ij}))\right\}^2\right] f(v_i)dv_i.$$

Finally at $D=0$ we have from (3.4)

$$\begin{aligned}\frac{\partial l_i(0)}{\partial D} &= \frac{1}{2}\left[-\phi^{-1}\sum_{j=1}^{n_i}b''(\theta_{ij})\int v_i^2 f(v_i)dv_i+\phi^{-2}\left\{\sum_{j=1}^{n_i}(Y_{ij}-b'(\theta_{ij}))\right\}^2\int v_i^2 f(v_i)dv_i\right] \\ &= \frac{1}{2}\phi^{-2}\left[\left\{\sum_{j=1}^{n_i}(Y_{ij}-b'(\theta_{ij}))\right\}^2-\phi\sum_{j=1}^{n_i}b''(\theta_{ij}))\right].\end{aligned}$$

Under null hypothesis, for Poisson count data $\theta_{ij}=\log\mu_{ij}=\mathbf{X}_{ij}^T\beta+\alpha,\ b(\theta_{ij})=\exp\left(\mathbf{X}_{ij}^T\beta+\alpha\right)$, $\mu_{ij}=b'(\theta_{ij})=\exp\left(\mathbf{X}_{ij}^T\beta+\alpha\right)$ and $\sigma_{ij}^2=\phi b''(\theta_{ij})=\phi\mu_{ij}$.

Substituting the above quantities we have the score statistics as

$$S(\beta,\alpha,\phi)=\frac{1}{2}\phi^{-2}\sum_{i=1}^{k}\left\{\left[\sum_{j=1}^{n_i}(Y_{ij}-\mu_{ij})\right]^2-\phi\sum_{j=1}^{n_i}\mu_{ij}\right\}. \tag{3.7}$$

Further, Jacqmin-Gadda and Commenges (1995) decompose the above score statistic into two terms as

$$S(\beta,\alpha,\phi) = \frac{1}{2}\phi^{-2}[S_1(\beta,\alpha,\phi) + S_2(\beta,\alpha,\phi)],$$

where

$$S_1(\beta,\alpha,\phi) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}(Y_{ij} - \mu_{ij})(Y_{ij'} - \mu_{ij'})$$

and

$$S_2(\beta,\alpha,\phi) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[(Y_{ij} - \mu_{ij})^2 - \phi\mu_{ij}\right].$$

Jacqmin-Gadda and Commenges (1995) show that the asymptotic variance of the score statistic $S(\beta,\alpha,\phi)$ can be obtained as $I_S = I_{S_1} + I_{S_2}$, where

$$I_{S_1} = \frac{1}{2\phi^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\sum_{1\le j'\neq j\le n_i}\mu_{ij}\mu_{ij'}$$

and

$$I_{S_2} = I + J_{S_2}I_{\gamma\gamma}^{-1}J_{S_2}^{T} - 2K_{S_2}I_{\gamma\gamma}^{-1}J_{S_2}^{T},$$

where

$$\begin{aligned} I \;&=\; \frac{1}{4\phi^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\phi\mu_{ij} + 2\mu_{ij}^2) + \frac{1}{4N^2\phi^2}\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\mu_{ij}\right)^2 \times \left[\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\phi\mu_{ij} + 3\mu_{ij}^2}{\mu_{ij}^2}\right) + N\right] \\ &\quad - \frac{1}{2N\phi^2}\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\mu_{ij}\right) \times \left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\phi\mu_{ij} + 3\mu_{ij}^2}{\mu_{ij}}\right), \end{aligned}$$

$$J_{S_2} = K_{S_2} \;=\; \frac{1}{4\phi^3}\left\{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\mu_{ij}\mathbf{W_{ij}^T} - \left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\mu_{ij}\right)\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\mathbf{W_{ij}^T}}{N}\right)\right\}$$

is a $1 \times (p+1)$ vector,

$$I_{\gamma\gamma} \;=\; \frac{1}{4\phi^5}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\mu_{ij}\mathbf{W}_{ij}\mathbf{W}_{ij}^{T}$$

is a $(p+1) \times (p+1)$ matrix, $N = \sum_{i=1}^{k} n_i$, $\gamma^T = [\alpha, \beta^T]$, $\mathbf{W}_{ij}^T = [\mathbf{1}, \mathbf{X}_{ij}^T]$, where $\mathbf{1}$ is an $N \times 1$ vector of 1's. Note that I_{S_2} now simplifies to $I_{S_2} = I - J_{S_2} I_{\gamma\gamma}^{-1} J_{S_2}^T$.

Let $\hat{\gamma}$ and $\hat{\phi}$ be some $\sqrt{k}$ consistent estimates of γ and ϕ respectively under the null hypothesis. Further, let $S(\hat{\alpha}, \hat{\beta}, \hat{\phi})$ and $\hat{I}_S$ be the estimate of $S(\alpha, \beta, \phi)$ and I_S respectively, after replacing γ and ϕ by $\hat{\gamma}$ and $\hat{\phi}$. Then, following Jacqmin-Gadda and Commenges (1995) the statistic

$$H_S = S^2(\hat{\beta}, \hat{\alpha}, \hat{\phi}) / \hat{I}_S \tag{3.8}$$

has, asymptotically, as $k \to \infty$, a chi-square distribution with one degree of freedom.

Now the mle of γ can be obtained iteratively by Fisher's scoring method from the following equation with $\phi = 1$

$$\gamma^{(t+1)} = \left[\left(\mathbf{W} \mathbf{Q} \mathbf{W}^{\mathbf{T}} \right)^{-1} \mathbf{W} \mathbf{Q} Z \right]^{(t)}, \quad t = 1, 2, 3, \ldots,$$

where $\mathbf{Q} = \text{diag}(\hat{\mu})$ is an $N \times N$ matrix and $Z = \mathbf{W}^T \hat{\gamma}^{(t)} + \frac{Y - \hat{\mu}}{\hat{\mu}}$, $t = 1, 2, 3, \ldots$ is an $N \times 1$ vector. Jacqmin-Gadda and Commenges (1995) suggest using the moment estimator $\hat{\phi} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_{ij})^2 / ((N - p) \hat{\mu}_{ij})$ of ϕ or the consistent estimator $\phi^* = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_{ij})^2 / (N \hat{\mu}_{ij})$. In our simulations in Section 3.5 we use the moment estimator as it has a degree of freedom correction which is expected to give a better estimate of ϕ.

3.2 Score test of homogeneity in the generalized linear mixed effects model using the procedure of Rao (1947)

The score statistic that we obtain and denoted by $S_R(\beta, \alpha, \phi)$ is the same as (3.7). In what follows we derive the score test by following Rao (1947) in which we obtain the asymptotic variance of the score statistic by following the procedure of Cox and Hinkley (1974).

Now, the asymptotic variance, as $k \to \infty$, of $S_R(\beta, \alpha, \phi)$ under H_0 is

$$I = I_{DD} - I_{D\gamma} I_{\gamma\gamma}^{-1} I_{D\gamma}^T,$$

where $\gamma^T = [\alpha, \beta^T]$,

$$I_{DD} = \sum_{i=1}^{k} E\left[\left.\frac{\partial l_i}{\partial D}\right|_{D=0}\right]^2, \quad I_{\gamma\gamma} = \sum_{i=1}^{k} E\left[\left(\left.\frac{\partial l_i}{\partial \gamma}\right|_{D=0}\right)\left(\left.\frac{\partial l_i}{\partial \gamma}\right|_{D=0}\right)^T\right]$$

and

$$I_{D\gamma} = \sum_{i=1}^{k} E\left[\left(\left.\frac{\partial l_i}{\partial D}\right|_{D=0}\right)\left(\left.\frac{\partial l_i}{\partial \gamma}\right|_{D=0}\right)^T\right].$$

By defining $U_{ij} = Y_{ij} - \mu_{ij}$, the ith term of (3.7) can be written as

$$\left.\frac{\partial l_i}{\partial D}\right|_{D=0} = \frac{1}{2}\phi^{-2}\left[\left(\sum_j U_{ij}\right)^2 - \sum_j \sigma_{ij}^2\right].$$

Thus

$$E\left(\left.\frac{\partial l_i}{\partial D}\right|_{D=0}\right)^2 = \frac{1}{4\phi^4} E\left(U_i^2 - \sum_j \sigma_{ij}^2\right)^2,$$

where $U_i = \sum_{j=1}^{n_i} U_{ij}$. Now since $E(U_{ij}) = 0$, we have

$$\begin{aligned}
E(U_i^2) &= E\left[\left(\sum_{j=1}^{n_i} U_{ij}\right)^2\right] = E\left(\sum_{j=1}^{n_i} U_{ij}^2 + \sum_{j=1}^{n_i}\sum_{j' \neq j}^{n_i} U_{ij}U_{ij'}\right) \\
&= E\left(\sum_{j=1}^{n_i} U_{ij}^2\right) = E\left[\sum_{j=1}^{n_i} (Y_{ij} - \mu_{ij})^2\right] \\
&= \sum_{j=1}^{n_i} E\left(Y_{ij} - \mu_{ij}\right)^2 \\
&= \sum_{j=1}^{n_i} \sigma_{ij}^2.
\end{aligned}$$

Therefore,

$$\begin{aligned} E\left(\left.\frac{\partial l_i}{\partial D}\right|_{D=0}\right)^2 &= \frac{1}{4\phi^4}E\left(U_i^2 - E(U_i^2)\right)^2 = \frac{1}{4\phi^4}\text{var}(U_i^2) \\ &= \frac{1}{4\phi^4}\left[E(U_i^4) - [E(U_i^2)]^2\right] \\ &= \frac{1}{4\phi^4}\left(\mu_4 - \mu_2^2\right), \end{aligned}$$

where μ_2 and μ_4 are the second and fourth central moments of Y_{ij}, respectively which can be expressed as a function of the second and fourth cumulants K_2 and K_4 of Y_{ij} (Kendall and Stuart, 1977)

$$\mu_4 = K_4 + 3K_2^2, \quad \mu_2 = K_2 = \sigma^2.$$

After simplification we get

$$I_{DD} = \frac{1}{4\phi^4}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(K_4(\theta_{ij}) + 2\sigma_{ij}^4\right),$$

where $K_4(\theta_{ij}) = \phi^3 b^{(iv)}(\theta_{ij})$ and $b^{(iv)}(\theta_{ij})$ is the fourth derivative of $b(\theta_{ij})$ with respect to θ_{ij}.

Further, using similar derivations we obtain

$$\begin{aligned} I_{\gamma\gamma} &= \phi^{-2}\sum_{i=1}^{k} E\left[\sum_{j=1}^{n_i} U_{ij}^2 \mathbf{W}_{ij}\mathbf{W}_{ij}^T\right] \\ &= \phi^{-2}\sum_{i=1}^{k}\sum_{j=1}^{n_i} E(U_{ij}^2)\mathbf{W}_{ij}\mathbf{W}_{ij}^T \\ &= \phi^{-2}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \sigma_{ij}^2\mathbf{W}_{ij}\mathbf{W}_{ij}^T \end{aligned}$$

and

$$\begin{aligned} I_{D\gamma} &= \frac{1}{2\phi^3}\sum_{i=1}^{k} E\left[\left\{\left(\sum_j U_{ij}\right)^2 - \sum_j \sigma_{ij}^2\right\}\left\{\sum_j U_{ij}\mathbf{W}_{ij}^T\right\}\right] \\ &= \frac{1}{2\phi^3}\sum_{i=1}^{k} E\left[\left\{\sum_j U_{ij}^2 + \sum_j\sum_{j'\neq j} U_{ij}U_{ij'} - \sum_j \sigma_{ij}^2\right\}\left\{\sum_j U_{ij}\mathbf{W}_{ij}^T\right\}\right] \\ &= \frac{1}{2\phi^3}\sum_{i=1}^{k} E\left[\sum_j U_{ij}^3\mathbf{W}_{ij}^T + \sum_j\sum_{j'\neq j} U_{ij}^2U_{ij'}\mathbf{W}_{ij}^T - \sum_j U_{ij}\sigma_{ij}^2\mathbf{W}_{ij}^T\right] \\ &= \frac{1}{2\phi^3}\sum_{i=1}^{k}\sum_{j=1}^{n_i} E(U_{ij}^3)\mathbf{W}_{ij}^T. \end{aligned}$$

As $E(U_{ij}) = E(U_{ij'}) = 0$, then

$$\begin{aligned} I_{D\gamma} &= \frac{1}{2\phi^3}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \mu_3(\theta_{ij})\mathbf{W}_{ij}^T \\ &= \frac{1}{2\phi^3}\sum_{i=1}^{k}\sum_{j=1}^{n_i} K_3(\theta_{ij})\mathbf{W}_{ij}^T, \end{aligned}$$

where $K_3(\theta_{ij}) = \phi^2 b'''(\theta_{ij})$ is the third cumulant of Y_{ij} and $b^{'''}(\theta_{ij})$ is the third derivative of $b(\theta_{ij})$ with respect to θ_{ij}. Note that in $I_{DD}, I_{D\gamma}$ and $I_{\gamma\gamma}$ we need the cumulants $K_2(\theta_{ij}),\ K_3(\theta_{ij})$ and $K_4(\theta_{ij})$. For over-dispersed count data these are $K_2(\theta_{ij}) = \sigma_{ij}^2 = \phi\mu_{ij},\ K_3(\theta_{ij}) = \phi^2\mu_{ij}$ and $K_4(\theta_{ij}) = \phi^3\mu_{ij}$.

Then, following Rao (1947) the score test of homogeneity for over-dispersed count data with γ and ϕ in $S, I_{DD}, I_{D\gamma}$ and $I_{\gamma\gamma}$ being replaced by their maximum likelihood estimates is $H_{SC} = S_R^2(\beta,\alpha,\phi)/[I_{DD} - I_{D\gamma}I_{\gamma\gamma}^{-1}I_{D\gamma}^T]$. Using the moment results given above, after simplification and replacement of $\beta,\ \alpha$ and ϕ by their maximum likelihood estimates $\hat{\beta},\ \hat{\alpha}$ and $\hat{\phi}$ the approximate score test statistic is

$$H_{SC} = S_R^2(\hat{\beta},\hat{\alpha},\hat{\phi})/[\hat{I}_{DD} - \hat{I}_{D\gamma}\hat{I}_{\gamma\gamma}^{-1}\hat{I}_{D\gamma}^T], \tag{3.9}$$

which, asymptotically, as $k \to \infty$, has a chi-square distribution with one degree of freedom, where

$$\begin{aligned} S_R(\hat{\beta}, \hat{\alpha}, \hat{\phi}) &= \frac{1}{2\hat{\phi}^2} \sum_{i=1}^{k} \left\{ \left[\sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_{ij}) \right]^2 - \hat{\phi} \sum_{j=1}^{n_i} \hat{\mu}_{ij} \right\}, \\ \hat{I}_{DD} &= \frac{1}{4\hat{\phi}^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\hat{\phi}\hat{\mu}_{ij} + 2\hat{\mu}_{ij}^2 \right), \\ \hat{I}_{\gamma\gamma} &= \frac{1}{\hat{\phi}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \hat{\mu}_{ij} \mathbf{W}_{ij} \mathbf{W}_{ij}^T, \end{aligned}$$

and

$$\hat{I}_{D\gamma} = \frac{1}{2\hat{\phi}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \hat{\mu}_{ij} \mathbf{W}_{ij}^T.$$

The quantities $\hat{\mu}_{ij}$ and $\mathbf{W}_{ij}$ are the same as in Section 3.1. Now the mle of γ is as in Section 3.1. However, the mle of ϕ is not obtainable from the likelihood. So we use the degree of freedom corrected method of moment estimator $\hat{\phi}$ given in Section 3.1.

Note that the asymptotic variance of S_R is computationally much simpler than that of S, and hence the computation of the score test statistic H_{SC}. In our simulation in Section 3.5 we have seen that the performance of the test statistic H_{SC} is almost identical to that of H_S in maintaining level and power.

We need to mention here that the above variance components of the test statistic H_{SC} are special cases of the variance of the global score test statistic χ^2_G of Lin (1997) for testing global variance components of the GLMM.

3.3 The score test based on the negative binomial distribution (NBD)

Here we consider a negative binomial mixed effects model and as in Section 3.1 our purpose is to develop a score test of homogeneity between and within groups for over-dispersed count data. Let Y_{ij} be the response variable for the jth observation in group

$i,\ j = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, k$, from the negative binomial distribution, denoted by $NB(\mu_{ij}, c_i)$ and given by

$$f(y_{ij}; \mu_{ij}, c_i) = \frac{\Gamma(y_{ij} + c_i^{-1})}{y_{ij}!\ \Gamma(c_i^{-1})} \left(\frac{1}{1 + c_i\mu_{ij}(\mathbf{x})}\right)^{c_i^{-1}} \left(\frac{c_i\mu_{ij}(\mathbf{x})}{1 + c_i\mu_{ij}(\mathbf{x}))}\right)^{y_{ij}}, \tag{3.10}$$

where $\log(\mu_{ij}) = X_{ij}^T\beta + Z_{ij}\alpha_i$ is the mixed effects model for the mean response with $\alpha_i = \alpha + D^{1/2}v_i$, c_i is the dispersion parameter for group i and X_{ij} is a $p \times 1$ vector of time independent covariates. The distribution of v_i is the same as specified in Section 3.1. Again, as in Section 3.1, since we want to test homogeneity across and within groups we consider the random intercept model in which $Z_{ij} = 1$ for all i, j. As in Section 3.1 our interest is to test $H_0 : D = 0$ against the alternative $H_A : D > 0$ which is equivalent to testing homogeneity across groups as well as testing homogeneity within groups as the intra-cluster or within group correlation coefficient (assuming common over-dispersion parameter c over all groups or individuals) is (see also Carrasco and Jover, 2005)

$$\rho = \frac{\exp(\mathbf{X}_{ij}^T\beta + \mathbf{X}_{ik}^T\beta + D)(e^D - 1)}{\left[\exp(\mathbf{X}_{ij}^T\beta + D/2) + \exp(2\mathbf{X}_{ij}^T\beta + D)(ce^D + e^D - 1)\right]}.$$

The ith term in the log-likelihood of the negative binomial distribution can be written as,

$$\begin{aligned} l_i(\beta, \alpha, c) &= \log f_i(y_{ij}; \alpha + D^{1/2}v_i, \beta, c) \\ &= \sum_{j=1}^{n_i} \left[\sum_{l=0}^{y_{ij}-1} \log(1 + cl) + y_{ij}(X_{ij}^T\beta + (\alpha + D^{1/2}v_i)) \right. \\ &\quad \left. -(y_{ij} + c^{-1}) \log\left(1 + ce^{X_{ij}^T\beta + (\alpha + D^{1/2}v_i)}\right)\right]. \end{aligned} \tag{3.11}$$

To obtain the score function, we follow the procedure of Section 3.1. Now

$$\begin{aligned}\frac{\partial}{\partial\sqrt{D}}\log\prod_{j=1}^{n_i} f_{ij}(y_{ij};\beta,\alpha,c) &= \sum_{j=1}^{n_i}\frac{\partial}{\partial\sqrt{D}}\log f_{ij}(y_{ij};\beta,\alpha,c)\\ &= v_i\sum_{j=1}^{n_i}\frac{(y_{ij}-\mu_{ij})}{1+cy_{ij}}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial^2\sqrt{D}}\log\prod_{j=1}^{n_i} f_{ij}(y_{ij};\beta,\alpha,c) &= \sum_{j=1}^{n_i}\frac{\partial}{\partial\sqrt{D}}\left\{\frac{\partial}{\partial\sqrt{D}}\log f_{ij}(y_{ij};\beta,\alpha,c)\right\}\\ &= -v_i^2\sum_{j=1}^{n_i}\frac{\mu_{ij}(1+cy_{ij})}{(1+c\mu_{ij})^2}.\end{aligned}$$

Therefore, at $D=0$ the score statistic becomes

$$\begin{aligned}S_N(\beta,\alpha,c) &= \sum_{i=1}^{k}\frac{\partial l_i(0)}{\partial D}\\ &= \frac{1}{2}\sum_{i=1}^{k}\left\{\left[\sum_{j=1}^{n_i}\frac{(y_{ij}-\mu_{ij})}{(1+c\mu_{ij})}\right]^2-\sum_{j=1}^{n_i}\frac{\mu_{ij}(1+cy_{ij})}{(1+c\mu_{ij})^2}\right\}.\end{aligned} \tag{3.12}$$

Then, the score test statistic for testing $H_0: D=0$ for the known nuisance parameters γ and c is

$$H_{NB} = S_N^2(\beta,\alpha,c)/(I_{DD}-AB^{-1}A^T), \tag{3.13}$$

where

$$I_{DD}=\sum_{i=1}^{k}E\left[\left.\frac{\partial l_i}{\partial D}\right|_{D=0}\right]^2 \text{ is a scalar and } \quad A=\begin{pmatrix}A_1 & A_2\end{pmatrix},\quad B=\begin{pmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{pmatrix},$$

$$A_1 = \sum_{i=1}^{k} E\left(\frac{-\partial^2 l_i}{\partial D \partial \gamma}\bigg|_{D=0}\right) = \sum_{i=1}^{k} E\left[\left(\frac{\partial l_i}{\partial D}\bigg|_{D=0}\right)\left(\frac{\partial l_i}{\partial \gamma}\bigg|_{D=0}\right)^T\right] \text{ is a } 1 \times (p+1) \text{ vector,}$$

$$A_2 = \sum_{i=1}^{k} E\left(\frac{-\partial^2 l_i}{\partial D \partial c}\bigg|_{D=0}\right) \text{ is a scalar,}$$

$$B_{11} = \sum_{i=1}^{k} E\left(\frac{-\partial^2 l_i}{\partial \gamma_s \partial \gamma_r}\bigg|_{D=0}\right) = \sum_{i=1}^{k} E\left[\left(\frac{\partial l_i}{\partial \gamma}\bigg|_{D=0}\right)\left(\frac{\partial l_i}{\partial \gamma}\bigg|_{D=0}\right)^T\right]$$
is a $(p+1) \times (p+1)$ matrix,

$$B_{12} = B_{21} = \sum_{i=1}^{k} E\left(\frac{-\partial^2 l_i}{\partial \gamma \partial c}\bigg|_{D=0}\right) \text{ is a } (p+1) \times 1 \text{ vector}$$

and

$$B_{22} = \sum_{i=1}^{k} E\left(\frac{-\partial^2 l_i}{\partial c^2}\bigg|_{D=0}\right) \text{ is a scalar.}$$

3.3.1 Computation of the variance of the score statistic

Now we need to evaluate the variance of S defined as $\text{Var}(S) = I_{DD} - AB^{-1}A^T$. The ith summand of I_{DD} can be written as

$$E\left(\frac{\partial l_i}{\partial D}\bigg|_{D=0}\right)^2 = \frac{1}{4}E\left\{\left[\sum_{j=1}^{n_i}\frac{y_{ij}-\mu_{ij}}{1+c\mu_{ij}}\right]^2 - \sum_{j=1}^{n_i}\frac{\mu_{ij}(1+cY_{ij})}{(1+c\mu_{ij})^2}\right\}^2$$
$$= \frac{1}{4}E(a_i-b_i)^2 = \frac{1}{4}E(a_i^2) - \frac{1}{2}E(a_i b_i) + \frac{1}{4}E(b_i^2),$$

where

$$a_i = \left[\sum_{j=1}^{n_i}\frac{y_{ij}-\mu_{ij}}{1+c\mu_{ij}}\right]^2$$

and

$$b_i = \sum_{j=1}^{n_i}\frac{\mu_{ij}(1+cy_{ij})}{(1+c\mu_{ij})^2}.$$

To derive quantities such as $E(a_i^2)$, we need some basic moment results from the $NB(\mu_{ij}, c)$ distribution. Let $U = \frac{Y-\mu}{(1+c\mu)}$. Then, following Section 2.5 it can be shown that the first four cumulants of U are

$$\begin{aligned} K_1 &= 0, \\ K_2 &= \frac{\mu}{1+c\mu}, \\ K_3 &= \frac{\mu+2c\mu^2}{(1+c\mu)^2} \end{aligned}$$

and

$$K_4 = \frac{\mu+6c\mu^2+6c^2\mu^3}{(1+c\mu)^3}.$$

Applying these results we obtain

$$E(a_i^2) = \sum_{j=1}^{n_i}\left[\frac{\mu_{ij}+6c\mu_{ij}^2+6c^2\mu_{ij}^3}{(1+c\mu_{ij})^3} + 3\left(\frac{\mu_{ij}}{1+c\mu_{ij}}\right)^2\right].$$

Now

$$\begin{aligned} a_i b_i &= \left[\sum_{j=1}^{n_i}\frac{(y_{ij}-\mu_{ij})^2}{(1+c\mu_{ij})^2} + \sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\frac{(y_{ij}-\mu_{ij})(y_{ij'}-\mu_{ij'})}{(1+c\mu_{ij})(1+c\mu_{ij'})}\right]\left[\sum_{j=1}^{n_i}\frac{\mu_{ij}(1+cy_{ij})}{(1+c\mu_{ij})^2}\right] \\ &= \sum_{j=1}^{n_i}\mu_{ij}\frac{(y_{ij}-\mu_{ij})^2(1+cy_{ij})}{(1+c\mu_{ij})^4} + \sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\mu_{ij}\frac{(1+cy_{ij})(y_{ij}-\mu_{ij})(y_{ij'}-\mu_{ij'})}{(1+c\mu_{ij})^3(1+c\mu_{ij'})}. \end{aligned}$$

Then

$$\begin{aligned} E(a_i b_i) &= \sum_{j=1}^{n_i}\mu_{ij}\frac{E[(y_{ij}-\mu_{ij})^2(1+cy_{ij})]}{(1+c\mu_{ij})^4} + \sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\mu_{ij}\frac{E[(1+cy_{ij})(y_{ij}-\mu_{ij})]E(y_{ij'}-\mu_{ij'})}{(1+c\mu_{ij})^3(1+c\mu_{ij'})} \\ &= \sum_{j=1}^{n_i}\mu_{ij}\frac{E[(y_{ij}-\mu_{ij})^2(1+cy_{ij})]}{(1+c\mu_{ij})^4} + 0. \end{aligned}$$

Further,

$$\begin{aligned} E[(y_{ij}-\mu_{ij})^2(1+cy_{ij})] &= E[(y_{ij}-\mu_{ij})^2 + cy_{ij}(y_{ij}^2+\mu_{ij}^2-2y_{ij}\mu_{ij})] \\ &= E(y_{ij}-\mu_{ij})^2 + cE(y_{ij}^3) + c\mu_{ij}^3 - 2cE(y_{ij}^2)\mu_{ij} \end{aligned}$$

$$\begin{aligned} &= \sigma_{ij}^2 + cE(y_{ij}^3) + c\mu_{ij}^3 - 2c(\mu_{ij}^2+\mu_{ij}+c\mu_{ij}^2) \\ &= \mu_{ij} + c\mu_{ij}^2 + c\mu_{ij}^3 + c\mu_{ij} + c\mu_{ij}^3 + 3c^2\mu_{ij}^2 + 2c^3\mu_{ij}^3 + 3c\mu_{ij}^2 \\ &\quad + 3c^2\mu_{ij}^3 - 2c\mu_{ij}^3 - 2c\mu_{ij}^2 - 2c^2\mu_{ij}^3 \\ &= \mu_{ij} + c\mu_{ij} + 2c\mu_{ij}^2 + 3c^2\mu_{ij}^2 + 2c^3\mu_{ij}^3 + c^2\mu_{ij}^3. \end{aligned}$$

Thus

$$\begin{aligned} E(a_ib_i) &= \sum_{j=1}^{n_i} \mu_{ij}\frac{\mu_{ij}+c\mu_{ij}+2c\mu_{ij}^2+3c^2\mu_{ij}^2+2c^3\mu_{ij}^3+c^2\mu_{ij}^3}{(1+c\mu_{ij})^4} \\ &= \sum_{j=1}^{n_i}\left(\frac{\mu_{ij}^2+c\mu_{ij}^2+2c\mu_{ij}^3+3c^2\mu_{ij}^3+2c^3\mu_{ij}^4+c^2\mu_{ij}^4}{(1+c\mu_{ij})^4}\right). \end{aligned}$$

Again

$$\begin{aligned} b_i^2 &= \left(\sum_{j=1}^{n_i}\frac{\mu_{ij}(1+cy_{ij})}{(1+c\mu_{ij})^2}\right)^2 \\ &= \sum_{j=1}^{n_i}\frac{\mu_{ij}^2(1+cy_{ij})^2}{(1+c\mu_{ij})^4} + \sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\frac{\mu_{ij}(1+cy_{ij})\mu_{ij'}(1+cy_{ij'})}{(1+c\mu_{ij})^2(1+c\mu_{ij'})^2}. \end{aligned}$$

$$E(b_i^2) = \sum_{j=1}^{n_i}\mu_{ij}^2\frac{E(1+cy_{ij})^2}{(1+c\mu_{ij})^4} + \sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\frac{\mu_{ij}\mu_{ij'}E(1+cy_{ij})E(1+cy_{ij'})}{(1+c\mu_{ij})^2(1+c\mu_{ij'})^2}.$$

Now,

$$\begin{aligned}
E(1+cy_{ij})^2 &= E(1+c^2y_{ij}^2+2cy_{ij}) \\
&= 1+c^2E(y_{ij}^2)+2c\mu_{ij} \\
&= 1+c^2(\mu_{ij}^2+\mu_{ij}+c\mu_{ij}^2)+2c\mu_{ij} \\
&= 1+c^2\mu_{ij}^2+c^2\mu_{ij}+c^3\mu_{ij}^2+2c\mu_{ij}.
\end{aligned}$$

So,

$$\begin{aligned}
E(b_i^2) &= \sum_{j=1}^{n_i}\mu_{ij}^2\frac{(1+2c\mu_{ij}+c^2\mu_{ij}^2+c^2\mu_{ij}+c^3\mu_{ij}^2)}{(1+c\mu_{ij})^4}+\sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\frac{\mu_{ij}\mu_{ij'}(1+c\mu_{ij})(1+c\mu_{ij'})}{(1+c\mu_{ij})^2(1+c\mu_{ij'})^2} \\
&= \sum_{j=1}^{n_i}\frac{(\mu_{ij}^2+2c\mu_{ij}^3+c^2\mu_{ij}^3+c^2\mu_{ij}^4+c^3\mu_{ij}^4)}{(1+c\mu_{ij})^4}+\sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\frac{\mu_{ij}\mu_{ij'}}{(1+c\mu_{ij})(1+c\mu_{ij'})}.
\end{aligned}$$

Finally

$$\begin{aligned}
E\left(\frac{\partial l_i}{\partial D}\right)^2 &= \frac{1}{4}\left[E(a_i^2)-2E(a_ib_i)+E(b_i^2)\right] \\
&= \frac{1}{4}\left[\sum_{j=1}^{n_i}\frac{(\mu_{ij}+6c\mu_{ij}^2+6c^2\mu_{ij}^3)}{(1+c\mu_{ij})^3}+3\sum_{j=1}^{n_i}\left(\frac{\mu_{ij}}{1+c\mu_{ij}}\right)^2-\right. \\
&\quad 2\sum_{j=1}^{n_i}\frac{(\mu_{ij}^2+c\mu_{ij}^2+2c\mu_{ij}^3+3c^2\mu_{ij}^3+2c^3\mu_{ij}^4+c^2\mu_{ij}^4)}{(1+c\mu_{ij})^4}+ \\
&\quad \left.\sum_{j=1}^{n_i}\frac{(\mu_{ij}^2+2c\mu_{ij}^3+c^2\mu_{ij}^3+c^2\mu_{ij}^4+c^3\mu_{ij}^4)}{(1+c\mu_{ij})^4}+\sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\frac{\mu_{ij}\mu_{ij'}}{(1+c\mu_{ij})(1+c\mu_{ij'})}\right].
\end{aligned}$$

Hence

$$\begin{aligned}
I_{DD} &= \frac{1}{4}\sum_{i=1}^{k}\left[\sum_{j=1}^{n_i}\frac{(\mu_{ij}+6c\mu_{ij}^2+6c^2\mu_{ij}^3)}{(1+c\mu_{ij})^3}+3\sum_{j=1}^{n_i}\left(\frac{\mu_{ij}}{1+c\mu_{ij}}\right)^2-\right. \\
&\quad 2\sum_{j=1}^{n_i}\frac{(\mu_{ij}^2+c\mu_{ij}^2+2c\mu_{ij}^3+3c^2\mu_{ij}^3+2c^3\mu_{ij}^4+c^2\mu_{ij}^4)}{(1+c\mu_{ij})^4}+ \\
&\quad \left.\sum_{j=1}^{n_i}\frac{(\mu_{ij}^2+2c\mu_{ij}^3+c^2\mu_{ij}^3+c^2\mu_{ij}^4+c^3\mu_{ij}^4)}{(1+c\mu_{ij})^4}+\sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\frac{\mu_{ij}\mu_{ij'}}{(1+c\mu_{ij})(1+c\mu_{ij'})}\right].
\end{aligned}$$

We still need to evaluate the elements of A and B. Now,

$$\begin{aligned}
\left.\frac{\partial l_i(\beta,\alpha,c)}{\partial\gamma}\right|_{D=0} &= \sum_{j=1}^{n_i}\left[y_{ij}W_{ij}^T - \frac{(1+cy_{ij})}{c}\frac{c\mu_{ij}W_{ij}^T}{(1+c\mu_{ij})}\right] \\
&= \sum_{j=1}^{n_i}\left[y_{ij} - \left(\frac{1+cy_{ij}}{1+c\mu_{ij}}\right)\mu_{ij}\right]W_{ij}^T \\
&= \sum_{j=1}^{n_i}\frac{y_{ij}-\mu_{ij}}{1+c\mu_{ij}}W_{ij}^T.
\end{aligned}$$

Then,

$$\begin{aligned}
E\left[\left(\left.\frac{\partial l_i}{\partial D}\right|_{D=0}\right)\left(\left.\frac{\partial l_i}{\partial\gamma}\right|_{D=0}\right)\right] &= \frac{1}{2}E\left[\left\{\sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\frac{(y_{ij}-\mu_{ij})(y_{ij'}-\mu_{ij'})}{(1+c\mu_{ij})(1+c\mu_{ij'})} + \sum_{j=1}^{n_i}\frac{(y_{ij}-\mu_{ij})^2}{(1+c\mu_{ij})^2}\right.\right. \\
&\quad \left.\left. - \sum_{j=1}^{n_i}\frac{\mu_{ij}(1+cy_{ij})}{(1+c\mu_{ij})^2}\right\}\left\{\frac{(y_{ij}-\mu_{ij})}{(1+c\mu_{ij})}W_{ij}^T\right\}\right]
\end{aligned}$$

$$\begin{aligned}
&= E\left[\sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\frac{(y_{ij}-\mu_{ij})^2(y_{ij'}-\mu_{ij'})}{(1+c\mu_{ij})^2(1+c\mu_{ij'})} + \sum_{j=1}^{n_i}\frac{(y_{ij}-\mu_{ij})^3}{(1+c\mu_{ij})^3}\right. \\
&- \left.\sum_{j=1}^{n_i}\mu_{ij}\frac{(y_{ij}-\mu_{ij})(1+cy_{ij})}{(1+c\mu_{ij})^3}\right]W_{ij}^T \\
&= \left[\sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\frac{E(y_{ij}-\mu_{ij})^2E(y_{ij'}-\mu_{ij'})}{(1+c\mu_{ij})^2(1+c\mu_{ij'})} + \sum_{j=1}^{n_i}\frac{E(y_{ij}-\mu_{ij})^3}{(1+c\mu_{ij})^3}\right. \\
&\quad \left.-\sum_{j=1}^{n_i}\mu_{ij}\frac{E\{(y_{ij}-\mu_{ij})(1+cy_{ij})\}}{(1+c\mu_{ij})^3}\right]W_{ij}^T \\
&= \sum_{j=1}^{n_i}\left[\frac{E(y_{ij}-\mu_{ij})^3}{(1+c\mu_{ij})^3} - \mu_{ij}\frac{E\{(y_{ij}-\mu_{ij})(1+cy_{ij})\}}{(1+c\mu_{ij})^3}\right]W_{ij}^T.
\end{aligned}$$

Now

$$\begin{aligned} E[(1+cy_{ij})(y_{ij}-\mu_{ij})] &= E\left[(y_{ij}-\mu_{ij})+cy_{ij}^2-cy_{ij}\mu_{ij}\right] \\ &= cE(y_{ij}^2)-c\mu_{ij}^2 \\ &= c(\mu_{ij}^2+\sigma_{ij}^2)-c\mu_{ij}^2 \\ &= c\sigma_{ij}^2 \end{aligned}$$

and

$$E(y_{ij}-\mu_{ij})^3 = \mu_{ij}+3c\mu_{ij}^2+2c^2\mu_{ij}^3.$$

Then after simplification we obtain

$$E\left[\left(\left.\frac{\partial l_i}{\partial D}\right|_{D=0}\right)\left(\left.\frac{\partial l_i}{\partial \gamma}\right|_{D=0}\right)^T\right] = \frac{1}{2}\sum_{j=1}^{n_i}\frac{\mu_{ij}}{(1+c\mu_{ij})}W_{ij}^T$$

and hence we obtain

$$A_1 = \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\mu_{ij}}{(1+c\mu_{ij})}W_{ij}^T.$$

Similar calculations show that

$$A_2 = \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\mu_{ij}^2}{(1+c\mu_{ij})^2},$$

$$B_{11} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\mu_{ij}}{(1+c\mu_{ij})}W_{ij}W_{ij}^T$$

and

$$B_{12} = B_{21} = 0.$$

We now obtain B_{22}. The partial derivative of of the log likelihood function of the negative binomial with respect to c is given by

$$\frac{\partial l_i}{\partial c} = \sum_{j=1}^{n_i}\left[\sum_{l=0}^{y_{ij}-1}\left(\frac{l}{1+cl}\right) + c^{-2}\log(1+c\mu_{ij}) - (y_{ij}+c^{-1})\left(\frac{\mu_{ij}}{1+c\mu_{ij}}\right)\right].$$

Then, applying the formulation described in Section 2.4.1 for calculating the Fisher Information, we obtain

$$B_{22} = c^{-4}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{l=1}^{\infty}\frac{l!(cq_{ij})^{l+1}}{(l+1)d_l}\right],$$

where $q_{ij} = c\mu_{ij}/(1+c\mu_{ij})$ and $d_l = \prod_{j=1}^{l}(1+jc)$. The parameters γ and c in H_N given in equation (3.13) are replaced by their maximum likelihood estimates, obtained from the negative binomial regression model under the null hypothesis (see also Lawless (1987)). The score test statistic H_N then reduces to $H_N = S_N^2(\beta,\alpha,c)/\left[I_{DD} - (A_1B_{11}^{-1}A_1^T + A_2^2B_{22}^{-1})\right]$.

Now the mle of γ can be estimated iteratively by Fisher's scoring method from the following equation

$$\gamma^{(t+1)} = \left[\left(\mathbf{WQW^T}\right)^{-1}\mathbf{WQ}Z\right]^{(t)}, \quad t = 1,2,3,\ldots,$$

where $\mathbf{Q} = \text{diag}\left(\frac{\hat{\mu}}{1+\hat{c}\hat{\mu}}\right)$ is an $N \times N$ matrix and $Z = \mathbf{W}^T\hat{\gamma}^{(t)} + \frac{Y-\hat{\mu}}{\hat{\mu}}$, $t = 1,2,3,\ldots$ is an $N \times 1$ vector. Fisher's scoring equation to estimate c is given by $c^{(t+1)} = c^{(t)} + \left(B_{22}^{-1}\left(\frac{\partial l}{\partial c}\right)\right)^{(t)}$, where l is the log-likelihood function given by (3.11) with $D = 0$. Note that these two equations must be solved simultaneously to get the maximum likelihood estimates of the parameters γ and c under the null hypothesis.

3.4 The score test based on the quasi-likelihood (DEQL)

For the joint estimation of the mean and the dispersion parameters, Nelder and Pregibon (1987) suggest using an extended quasi-likelihood (EQL), which assumes only the first two moments of the response variable. Lee and Nelder (2001) introduce double-extended quasi-likelihood (DEQL) also for the joint estimation of the mean and the dispersion parameters. The DEQL methodology requires an EQL for Y_{ij} given some random effect α_i and an EQL for α_i from a conjugate distribution given some mean parameter μ and dispersion parameter c. The DEQL is then obtained by combining the two EQL's. The random effect α_i's or some transformed variables s_i are then replaced by their maximum likelihood estimates resulting in a profile DEQL. For over-dispersed count data this profile DEQL is the same as the negative binomial log-likelihood with the factorials replaced by the usual Stirling approximations (Lee and Nelder, 2001, Result 5, p.996). They argue, however, that the Stirling approximation may not be good for small z, so for nonnormal random effects, they suggest using the modified Stirling approximation

$$\ln \Gamma(z) \simeq \left(z - \frac{1}{2}\right) + \frac{1}{2}\ln(2\pi) - z + \frac{1}{12z}.$$

Using the modified Stirling approximation given above, the log link $\log(\mu_{ij}) = X_{ij}^T\beta + Z_{ij}\alpha_i$, with $\alpha_i = \alpha + D^{1/2}v_i$, where α_i is the random effect defined earlier. Following

Paul and Saha (2007), the ith term of the profile DEQL is

$$
\begin{aligned}
DEQ_i(y_{ij};\alpha + D^{1/2}v_i, \beta, c) \;=\; & \sum_{j=1}^{n_i} \Big[y_{ij}(X_{ij}^T\beta + Z_{ij}(\alpha + D^{1/2}v_i)) \\
- & \left(y_{ij} + \frac{1}{c}\right) \log(1 + ce^{X_{ij}^T\beta + Z_{ij}(\alpha + D^{1/2}v_i)}) \\
+ & \left(y_{ij} + \frac{1}{c} - \frac{1}{2}\right) \ln(1 + cy_{ij}) + \frac{c}{12(1 + cy_{ij})} - \frac{c}{12} \\
- & \left(y_{ij} + \frac{1}{2}\right) \ln(y_{ij}) - \frac{1}{12(y_{ij} + 1)} - \frac{1}{2}\ln(2\pi)\Big] .
\end{aligned} \tag{3.14}
$$

Then following the procedure in Section 3.1 the score statistic becomes

$$
S_Q(\beta, \alpha, c) \;=\; \frac{1}{2}\sum_{i=1}^{k} \left\{ \left[\sum_{j=1}^{n_i} \frac{(y_{ij} - \mu_{ij})}{(1 + c\mu_{ij})} Z_{ij} \right]^2 - \sum_{j=1}^{n_i} \frac{\mu_{ij}(1 + cy_{ij})}{(1 + c\mu_{ij})^2} Z_{ij}^2 \right\}. \tag{3.15}
$$

Assuming the random intercept model, that is $Z_{ij} = 1$ for all i and j, equation (3.15) becomes

$$
S_Q(\beta, \alpha, c) \;=\; \frac{1}{2}\sum_{i=1}^{k} \left\{ \left[\sum_{j=1}^{n_i} \frac{(y_{ij} - \mu_{ij})}{(1 + c\mu_{ij})} \right]^2 - \sum_{j=1}^{n_i} \frac{\mu_{ij}(1 + cy_{ij})}{(1 + c\mu_{ij})^2} \right\}. \tag{3.16}
$$

The statistic

$$
H_{QL} = S_Q^2(\hat{\beta}, \hat{\alpha}, \hat{c})/(I_{DD} - AB^{-1}A^T), \tag{3.17}
$$

has, asymptotically, a chi-square distribution with one degree of freedom. The definitions of the mixed partial derivatives and their corresponding expected values of I_{DD} and the elements of A and B are exactly the same as those in Section 3.3 with l_i replaced by DEQ_i. The variance $I = I_{DD} - AB^{-1}A^T$ of the double extended quasi-likelihood score function S_Q is computed according to the procedure in Section 3.3. We find that all the variance components in I are exactly the same as those of the S_N except for the component B_{22} which we obtain in what follows.

To obtain B_{22} we define $\tau = 1/c$. Then

$$B_{22} = \sum_{i=1}^{k} E\left(-\partial^2 l_i/\partial c^2\right) = c^{-4} \sum_{i=1}^{k} E\left(-\partial^2 l_i/\partial \tau^2\right),$$

where

$$\begin{aligned}
\sum_{i=1}^{k} E\left(-\frac{\partial^2 DEQ_i}{\partial \tau^2}\bigg|_{D=0}\right) &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[E\left\{\frac{1}{\mu_{ij}+\tau} + \frac{1}{2\tau^2} + \frac{1}{6\tau^3} - \frac{1}{y_{ij}+\tau}\right.\right. \\
&- \left.\left.\frac{1}{2(y_{ij}+\tau)^2} - \frac{1}{6(y_{ij}+\tau)^3}\right\} - \frac{E(y_{ij}-\mu_{ij})}{(\mu_{ij}+\tau)^2}\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\frac{1}{\mu_{ij}+\tau} + \frac{1}{2\tau^2} + \frac{1}{6\tau^3}\right] \\
&\quad -E\left[\frac{1}{y_{ij}+\tau} + \frac{1}{2(y_{ij}+\tau)^2} + \frac{1}{6(y_{ij}+\tau)^3}\right]
\end{aligned}$$

$$\begin{aligned}
= &\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\frac{1}{\mu_{ij}+\tau} + \frac{1}{2\tau^2} + \frac{1}{6\tau^3}\right] \\
&-\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\sum_{y_{ij}=0}^{\infty}\left(\frac{1}{y_{ij}+\tau} + \frac{1}{2(y_{ij}+\tau)^2} + \frac{1}{6(y_{ij}+\tau)^3}\right) \times f(y_{ij};\mu_{ij},c)\right],
\end{aligned}$$

where $f(y_{ij};\mu_{ij},c)$ is the probability function of the negative binomial distribution given by (3.10) with $D = 0$. The regression parameter γ is estimated by the Fisher scoring equation given in Section 2.4.2. However, to estimate the dispersion parameter c the Fisher scoring equation is $c^{(t+1)} = c^{(t)} + \left(B_{22}^{-1}\left(\frac{\partial l}{\partial c}\right)\right)^{(t)}$, $t = 1, 2, 3, \ldots$, where l is the log-likelihood function given by (3.14) with $D = 0$. Here also the two equations must be solved simultaneously to get the maximum likelihood estimates of the parameters γ and c.

The variances of the score functions S, S_R, S_N and S_Q are all different. The variance of S_R is the simplest to calculate and the statistics H_{SC}, H_{NB} and H_{QL} are all computationally simpler than the statistic H_S.

3.5 Simulation study

In this section we conduct a simulation study to compare, in terms of size and power, the four score test statistics H_S, H_{SC}, H_{NB} and H_{QL}. For studying the properties of the statistics in terms of empirical size we generate count data from Poisson, negative binomial and lognormal-Poisson mixture distributions under the hypothesis of homogeneity. We assume random effect is the intercept ($Z_{ij} = 1$).

Two sets of data are simulated for each distribution of the response variable assuming homogeneous and heterogeneous inner group sizes ($= n_i$) with different number of groups/individuals (k) according to the variance (D) of the distribution of the group-specific random effect of the response variable and for different values of the over-dispersion parameter c. The samples are comprised of $k = 10, 20, 50, 100$ individuals/groups with $n_i = 5$ observations in the homogeneous group and n_i distributed uniformly between 5 and 20 in the heterogeneous group. The values of the over-dispersion parameter c considered are $0.10, 0.22, 0.40, 0.67, 0.91$ and 1.25. We generate 10,000 samples from each experiment in computing the nominal levels and power. The following log-linear model for the response variable is assumed (see Jacqmin-Gadda and Commenges (1995))

$$\log(\mu_{ij}) = 0.8x_{1ij} + 0.5x_{2i} - 0.5, \tag{3.18}$$

for $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_i$. The variable x_1 is subject-specific and x_2 is group-specific and are simulated according to a standard normal distribution.

For drawing samples and for estimating the maximum likelihood estimates of the regression and dispersion parameters of interest under the null hypothesis, different R functions are applied. To simulate correlated data, we add a group-specific random intercept in the model for the response variable which is $\alpha_i = \alpha + D^{1/2}v_i$, where v_i is standard normal. Therefore, the random effects are normally distributed with mean α and variance D.

Our first objective is to compare the estimated Type I error of the four tests. Table

3.1 displays the computed nominal levels for the four tests when data are generated from the Poisson distribution according to the variance of the distribution of the group-specific random effect under the hypothesis of homogeneity, that is, $D = 0$. None of the statistics maintain level, although the level property, in general, improves as the sample size increases, except for the statistics H_{QL}. However, H_{NB} performs the best in maintaining level. The performance of the statistic H_{QL} is the worst, severely underestimating the nominal levels. There does not seem to be any difference in level properties of all the statistics between cases when the groups have homogeneous sample size and the case when the sample sizes are not equal.

Tables 3.2 and 3.3 display the estimated Type I error for the four tests when data are simulated from the negative binomial distribution under the hypothesis of homogeneity with common over-dispersion parameter c. From the results in Tables 3.2 and 3.3 it is evident that both the statistics based on the generalized linear mixed model (GLMM) framework perform the worst, severely overestimating the level as the number of groups increases. This property is even worst when the number of groups as well as the over-dispersion are large. The other two statistics H_{NB} and H_{QL}, in general, show conservative behavior, although as the number of groups increases the estimated levels become closer to the nominal levels. Level properties of both of these statistics are similar.

We then generated count data with over-dispersion from the Log-normal-Poisson mixture model under the hypothesis of homogeneity with common over-dispersion parameter c. The mean and variances of the log-normal distribution assumed were $m = \log(\mu_{ij}) - \frac{1}{2}\log(1+c)$ and $\sigma^2 = \log(1+c)$ as used by Paul and Banerjee (1998), where $\log(\mu_{ij})$ is given by (3.18) and c is the common over-dispersion parameter. Tables 3.4 and 3.5 display the estimated Type I error of the four tests. From Tables 3.4 and 3.5 we see that the overall properties of all of the four statistics remain almost the same as those shown in Tables 3.2 and 3.3 when data were simulated from the negative binomial distribution, except when k and c are large ($k \geq 50$ and $c \geq .67$) in which case both the statistics H_{NB} and H_{QL} show liberal behavior.

Our second objective is to compare power of the four tests when data are generated from heterogeneous count data models. Here we first consider data from the heterogeneous negative binomial model. As for size here also we consider nominal levels $0.10, 0.05$ and 0.01. We computed power for $k = 20$ and 50 with common $c = .10, .40$ for $D = 0, .05, .10, .15, .20$. Table 3.6 presents computed power of the four tests.

From the results in Table 3.6 we first discuss results for $k = 20$ and $c = .10$. As the value of D increases, the power increases for all the statistics and they all show similar power. Note that at these values of k and c all the four statistics have similar level property (and they all reasonably hold level) and they also have similar power. This indicates that if all the four statistics have similar level, they will have similar power.

Now we consider $k = 20$ and $c = .40$. As the statistics H_S and H_{SC} produce highly inflated Type I error, their powers are also overestimated. Power properties of the other two statistics H_{NB} and H_{QL} are similar. However, the power of H_{NB} and H_{QL} increases faster as D increases. For example, at $\alpha = .01$, the estimated level of H_S and H_{SC} is .06 and for H_{NB} and H_{QL} is .007. The corresponding empirical power for the four statistics H_S, H_{SC}, H_{NB} and H_{QL} for $D = .2$ are .70, .69, .60 and .60, respectively. Note that empirical level 0.007 for H_{NB} and H_{QL} is close to the nominal level of $\alpha = .01$, whereas the empirical levels for both H_S and H_{SC} are about six times the nominal level.

Next we consider $k = 50$. Generally, as c increases power increases and also as the number of observations in each group increases power increases. The general properties of H_S and H_{SC} for $k = 50$ and $c = .10$ and .40 are similar to those for $k = 20$ and $c = .40$. As D increases ($D > .1$) power of all of the statistics become almost identical, although empirical levels of the statistics H_{NB} and H_{QL} are close to the nominal and those of H_S and H_{SC} are highly inflated.

The power study was extended for the situation in which data are generated from heterogeneous log-normal-Poisson mixture distribution. The results are given in Table 3.7. The overall finding from the results in Table 3.7 seem to be similar to those in Table

3.6 when data are generated from the heterogeneous negative binomial model.

In summary, as k, c and n_i increase, the power increases for all the statistics. The statistics H_S and H_{SC}, in general, show highly inflated level properties. The statistics H_{NB} and H_{QL} show some conservative level properties, however, as the values of c and k increase, empirical levels become closer to the nominal. The power of the statistics H_S and H_{SC} are, in general, larger than those of H_{NB} and H_{QL} which is expected. What is interesting is that as D increases ($D > .1$), the power of all of the statistics become almost identical, although empirical levels of the statistics H_{NB} and H_{QL} are close to the nominal and those of H_S and H_{SC} are highly inflated. The power of both the statistics H_{NB} and H_{QL} are very similar in all the cases studied.

We extended this simulation study of the properties of the four statistics in terms of empirical size and power to situations where the over-dispersion parameter c is not the same for all groups. For this we generated data from the heterogeneous negative binomial and Log-normal-Poisson mixture distributions with heterogeneous over-dispersion parameter $c(.10 \leq c \leq 1.0)$. The results for size and power are given in Tables 3.8 and 3.9. The overall conclusion of the level and power properties of the four statistics remain the same as those for homogenous c.

The level and power properties of all the statistics, in general, remain similar irrespective of which mechanism of over-dispersion is used to generate count data. This also seems to be true irrespective of whether the over-dispersion parameter c is varying or constant.

3.6 Examples

In this section we analyze two real data sets. The first example is the epileptic seizures count data from a clinical trial of an anti-epileptic drug obtained from Table 2 (page 664) in Thall and Vail (1990), also discussed and analyzed by Fitzmaurice, Laird and Ware (2004). The second example represents the counts of new skin cancers per year taken from the Skin Cancer Prevention Study, a randomized, double-blind, placebo-controlled clinical trial from Greenberg, Baron, Stukel, Stevens, Mandel, Spencer, Elias, Lowe, Nierenberg, Bayrd, Vance, Freeman, Clendenning and Kwan (1990). We first test whether over-dispersion exists in these data sets and then test for homogeneity of within and between groups.

3.6.1 Example 1: Clinical trial of an anti-epileptic drug

The data are from a placebo-controlled clinical trial of 59 epileptic patients. Patients with partial seizures were enrolled in a randomized clinical trial of the anti-epileptic drug, progabide. Participants in the study were randomized to either progabide or a placebo, as an adjuvant to the standard anti-epileptic chemotherapy. Progabide is an anti-epileptic drug whose primary mechanism of action is to enhance gamma-aminobutyric acid (GABA) content; GABA is the primary inhibitory neurotransmitter in the brain. Prior to receiving treatment, baseline data on the number of epileptic seizures during the preceding 8-week interval were recorded. Counts of epileptic seizures during 2-week intervals before each of four successive post-randomization clinic visits were recorded. The data are given in Table 3.10. The covariates recorded are:

Patient ID, Treatment (0 = Placebo, 1 = Progabide), Age, Baseline 8 week seizure count (Time = 0), First 2 week seizure count (Time = 2), Second 2 week seizure count (Time = 4), Third 2 week seizure count (Time = 6), Fourth 2 week seizure count (Time = 8). These data show clear over-dispersion as can be seen from the estimated mean and variances of the seizure counts given in Table 3.11.

In Chapter 4 we developed two score tests for over-dispersion in generalized linear mixed effects model. The two score tests are denoted by T and T_c (for details see Chapter 4). To obtain the statistic T for testing the presence of over-dispersion in the data given in Table 3.10 we first fitted the following model (see also Fitzmaurice *et al.*, 2004), assuming data distributed as Poisson,

$$\log(\mu_{ij}) = \beta_1 + \beta_2 Treatment + \beta_3 Age + \beta_4 Time + \beta_5 Treatment * Time. \quad (3.19)$$

The maximum likelihood estimates obtained from this model fit are (standard errors in parentheses) $\hat{\beta}_1 = 3.760$ (0.075), $\hat{\beta}_2 = 0.023$ (0.045), $\hat{\beta}_3 = -0.0202$ (0.0024), $\hat{\beta}_4 = -0.187$ (0.009) and $\hat{\beta}_5 = -0.026$ (0.013).

To obtain the statistic T_c we fitted the following random intercept model to the same data, assuming data distributed as Poisson,

$$\log(\mu_{ij}) = \beta_1 + \beta_2 Treatment + \beta_3 Age + \beta_4 Time + \beta_5 Treatment * Time + \alpha_i, \quad (3.20)$$

where α_i's are normal with mean zero and variance D. The maximum likelihood estimates of the regression parameters and the variance component D obtained from the fit of the model are (standard errors in parentheses) $\hat{\beta}_1 = 3.324$ (0.471), $\hat{\beta}_2 = -0.016$ (0.229), $\hat{\beta}_3 = -0.016$ (0.015), $\hat{\beta}_4 = -0.375$ (0.018), $\hat{\beta}_5 = -0.026$ (0.005) and $\hat{D} = 0.594$ (0.073). Note that here we need the estimates of the random effects α_i which are given in Table 3.12.

The values of the score test statistics T and T_c are then obtained as 22.47 and 18.07 respectively. This shows severe over-dispersion, in agreement with the preliminary analysis based on the results in Table 3.11, indicating that data analysis should take account of the over-dispersion present.

We now test for homogeneity in the seizures count data. For this we fitted the following model assuming that the data come from a negative binomial distribution $NB(\mu_{ij}, c)$,

$$\log(\mu_{ij}) = \beta_1 + \beta_2 Treatment + \beta_3 Age + \beta_4 Time + \beta_5 Treatment * Time. \quad (3.21)$$

The maximum likelihood estimates of the parameters obtained from the fit of this model are (standard errors in parentheses) $\hat{\beta}_1 = 3.66$ (0.289), $\hat{\beta}_2 = -0.019$ (0.211), $\hat{\beta}_3 = -0.020$ (0.009), $\hat{\beta}_4 = -0.150$ (0.032), $\hat{\beta}_5 = -0.022$ (0.044) and $\hat{c} = 1.036$ (0.0899).

The values of the test statistics are as follows $H_S = 26.09$, $H_{SC} = 23.44$, $H_{NB} = 24.81$ and $H_{QL} = 25.03$ which all show high heterogeneity.

3.6.2 Example 2: The skin cancer prevention study

These data are from the Skin Cancer Prevention Study of Greenberg et al. (1990). This was a randomized, double-blind, placebo-controlled clinical trial of beta-carotene to prevent non-melanoma skin cancer in high risk subjects. A total of 1805 subjects were randomized to either placebo or 50mg of beta-carotene per day for 5 years. Subjects were examined once a year and biopsied if cancer was suspected to determine the number of new skin cancers occurring since the last exam. The data are given in Greenberg *et al.* (1990).

In this example the response variable Y is a count representing the number of new skin cancers per year. The explanatory variables are: treatment coded as 1 for beta-carotene and 0 for placebo, years of follow-up, gender coded as 1 for male and 0 for female, skin cancer type coded as 1 for burns and 0 otherwise, exposure representing count of the number of previous skin cancers and age in years. The study has variable number of repeated observations for each individual (n_i between 1 and 5). The complete data on 1683 subjects comprising a total of 7081 measurements are given in Fitzmaurice *et al.* (2004) which we analyze here for the presence of over-dispersion and for testing homogeneity within and between subjects.

The data set shows over-dispersion as can be seen from Table 3.13. The model

$$\begin{aligned}\log(\mu_{ij}) &= \beta_1 + \beta_2 Year + \beta_3 Treatment + \beta_4 Treatment \times Year + \beta_5 Age + \beta_6 Skin \\ &\quad + \beta_7 Gender + \beta_8 Exposure \qquad (3.22)\end{aligned}$$

has been fitted to obtain the value of the test statistic T assuming data are distributed as Poisson. The maximum likelihood estimates of the parameters of this model are (standard errors in parentheses) $\hat{\beta}_1 = -3.464$ (0.192), $\hat{\beta}_2 = -0.008$ (0.025), $\hat{\beta}_3 = 0.009$ (0.104), $\hat{\beta}_4 = 0.039$ (0.034), $\hat{\beta}_5 = 0.016$ (0.003), $\hat{\beta}_6 = 0.136$ (0.046), $\hat{\beta}_7 = 0.596$ (0.060) and $\hat{\beta}_8 = 0.135$ (0.003) producing a value of $T = 54.84$.

The random intercept model

$$\begin{aligned}\log(\mu_{ij}) &= \beta_1 + \beta_2 Year_{ij} + \beta_3 Treatment + \beta_4 Treatment \times Year + \beta_5 Age + \beta_6 Skin \\ &\quad + \beta_7 Gender + \beta_8 Exposure + \alpha_i, \qquad (3.23)\end{aligned}$$

where α_i is normal with mean zero and variance D, was fitted to obtain the value of the test statistic T_c. The maximum likelihood estimates of the parameters are (standard errors in parentheses) $\hat{\beta}_1 = -4.588$ (0.294), $\hat{\beta}_2 = 0.003$ (0.023), $\hat{\beta}_3 = 0.0204$ (0.112), $\hat{\beta}_4 = 0.036$ (0.029), $\hat{\beta}_5 = 0.019$ (0.004), $\hat{\beta}_6 = 0.339$ (0.079), $\hat{\beta}_7 = 0.660$ (0.089), $\hat{\beta}_8 = 0.178$ (0.011) and $\hat{D} = 1.236$ (0.0387). The estimates of the 1683 random effects α_i are given in Tables 3.14, 3.15 and 3.16. The value of the test statistic T_c is obtained as 37.13.

For testing homogeneity in these data we fitted the following model, assuming that data come from a negative binomial distribution $NB(\mu_{ij}, c)$,

$$\begin{aligned}\log(\mu_{ij}) &= \beta_1 + \beta_2 Year_{ij} + \beta_3 Treatment + \beta_4 Treatment \times Year + \beta_5 Age + \beta_6 Skin \\ &\quad + \beta_7 Gender + \beta_8 Exposure. \qquad (3.24)\end{aligned}$$

The maximum likelihood estimates obtained are (standard errors in parentheses) $\hat{\beta}_1 = -3.743$ (0.254), $\hat{\beta}_2 = 0.0027$ (0.0344), $\hat{\beta}_3 = 0.0022$ (0.145), $\hat{\beta}_4 = 0.029$ (0.048), $\hat{\beta}_5 = 0.017$ (0.003), $\hat{\beta}_6 = 0.267$ (0.064), $\hat{\beta}_7 = 0.613$ (0.076), $\hat{\beta}_8 = 0.169$ (0.007) and $\hat{c} = 2.245$ (0.1538).

The values of the test statistics are $H_S = 66.50, H_{SC} = 63.44, H_{NB} = 11.10$ and $H_{QL} = 12.08$. As in example 1, these values show significant heterogeneity within and

Table 3.1: Comparison of performances of four score tests in respect of Type I error when data are simulated from Poisson distribution according to the variance of the distribution of the group-specific random effect under H_0, that is, $D = 0$. The nominal levels of significance considered are 10%, 5% and 1%. Two sample structures are followed for k = 10,20,50 and 100 to simulate the data: $n_i = 5$ (Homogeneous Group) and n_i uniformly distributed between 5 and 20 (Heterogeneous Group).

		Homogeneous Group Sizes				Heterogeneous Group Sizes			
k	α	H_S	H_{SC}	H_{NB}	H_{QL}	H_S	H_{SC}	H_{NB}	H_{QL}
10	.10	2.7	2.3	4.6	1.9	2.8	2.7	3.5	1.9
	.05	1.4	1.2	2.3	1.1	1.7	1.6	1.9	1.1
	.01	0.4	0.3	0.6	0.7	0.5	0.5	0.5	0.4
20	.10	4.1	3.5	5.3	1.8	4.5	4.3	5.4	2.0
	.05	2.2	2.0	3.2	1.1	2.5	2.4	2.9	0.9
	.01	0.5	0.5	0.7	0.3	0.8	0.8	0.9	0.3
50	.10	5.4	4.8	6.9	1.5	6.1	5.7	6.4	1.7
	.05	2.9	2.4	3.5	0.6	2.8	2.6	3.6	0.7
	.01	0.8	0.7	0.9	0.01	0.8	0.7	0.9	0.2
100	.10	6.5	5.9	7.8	1.1	6.8	6.6	7.4	1.0
	.05	3.2	2.9	3.8	0.5	3.8	3.6	3.6	0.5
	.01	0.9	0.7	0.8	0.05	1.2	1.1	1.2	0.0

Note: k is the number of groups (individuals), and n_i is the number of observations per group (individual).

across the individuals.

Table 3.2: Estimated Type I error of four tests when data are generated from negative binomial distribution under the hypothesis of homogeneity. Levels considered are $10\%, 5\%$ and 1%.

			$n_i = 5$				$5 \leq n_i \leq 20$			
k	c	α	H_S	H_{SC}	H_{NB}	H_{QL}	H_S	H_{SC}	H_{NB}	H_{QL}
10	.10	.10	3.5	3.2	4.2	2.9	4.6	4.4	3.5	2.9
		.05	2.1	1.8	2.4	1.9	2.6	2.5	2.1	1.7
		.01	0.8	0.6	0.7	1.1	1.0	0.9	0.6	0.4
	.22	.10	4.4	4.1	4.2	4.4	6.3	6.1	3.7	3.9
		.05	2.6	2.4	2.2	2.3	3.9	3.7	2.0	2.4
		.01	0.8	0.7	0.7	1.7	1.6	1.5	0.6	0.6
	.40	.10	5.4	5.0	3.6	6.4	7.7	7.5	4.1	4.1
		.05	3.3	3.0	2.0	2.6	5.2	5.0	2.2	2.3
		.01	1.3	1.2	0.5	0.6	2.4	2.3	0.7	0.7
	.67	.10	7.5	7.2	3.5	3.6	9.9	9.6	4.2	4.3
		.05	4.4	4.1	2.0	2.1	6.4	6.2	2.3	2.3
		.01	1.9	1.8	0.7	0.8	3.2	3.1	0.8	0.5
	.91	.10	7.8	7.7	2.9	2.8	12.0	11.8	3.9	4.3
		.05	4.9	4.9	2.4	2.2	8.5	8.3	2.1	2.3
		.01	2.2	2.1	0.4	0.5	4.2	4.1	0.8	0.9
	1.25	.10	9.1	8.9	2.8	3.2	13.2	13.1	3.1	3.6
		.05	6.1	6.1	2.3	2.4	9.2	9.0	2.0	2.1
		.01	2.7	2.9	0.3	0.5	4.7	4.8	0.6	0.6
20	.10	.10	6.1	5.6	5.5	3.3	7.4	7.2	4.9	3.6
		.05	3.7	3.2	2.9	1.8	4.3	4.1	2.8	2.0
		.01	1.3	1.1	0.8	0.6	1.7	1.5	0.8	0.6
	.22	.10	9.6	8.7	5.2	5.0	11.3	10.9	5.4	5.4
		.05	5.9	5.2	2.7	3.2	7.4	7.1	2.9	3.1
		.01	2.4	2.2	0.8	1.5	3.4	3.2	0.8	1.1
	.40	.10	13.9	13.1	5.2	7.3	16.2	15.8	5.5	5.5
		.05	9.2	8.4	2.9	4.8	11.2	10.9	2.6	2.7
		.01	4.1	3.7	0.8	2.7	6.0	5.7	0.7	0.7
	.67	.10	17.8	17.1	4.8	5.5	22.1	21.7	5.0	5.4
		.05	12.7	11.9	2.6	2.8	16.4	16.0	2.7	2.8
		.01	6.4	5.9	0.7	0.7	8.9	8.7	0.8	1.0
	.91	.10	19.9	19.3	4.5	6.9	25.8	25.6	5.2	5.8
		.05	14.1	13.5	2.6	4.7	19.2	18.9	2.8	3.3
		.10	7.3	7.0	0.8	2.3	11.1	10.9	0.9	1.0
	1.25	.10	23.6	23.0	4.4	5.0	29.0	28.8	5.2	6.1
		.05	17.2	16.5	2.4	2.7	22.1	21.9	2.7	3.4
		.01	9.1	8.9	0.7	0.8	12.8	12.6	0.9	1.1

Note: k is the number of groups (individuals) and n_i is the number of observations per group (individual)

Table 3.3: Estimated Type I error of four tests when data are generated from negative binomial distribution under the hypothesis of homogeneity. Levels considered are $10\%, 5\%$ and 1%.

			$n_i = 5$				$5 \leq n_i \leq 20$			
k	c	α	H_S	H_{SC}	H_{NB}	H_{QL}	H_S	H_{SC}	H_{NB}	H_{QL}
50	.10	.10	12.7	11.6	6.9	4.4	13.8	13.3	6.7	6.6
		.05	8.0	7.0	3.5	2.4	8.7	8.3	3.8	3.4
		.01	3.3	2.9	0.7	0.6	3.5	3.3	1.2	0.9
	.22	.10	21.2	20.2	7.0	6.7	23.5	22.9	7.0	7.2
		.05	14.5	13.5	3.7	3.7	16.2	15.7	3.7	3.7
		.01	6.8	6.1	0.9	1.1	8.2	7.8	1.1	1.2
	.40	.10	32.8	31.3	6.7	7.8	36.1	35.6	7.4	7.5
		.05	24.4	23.1	3.7	4.8	27.2	26.5	3.9	4.0
		.01	13.8	12.7	0.9	2.0	15.4	15.0	1.0	1.0
	.67	.10	44.0	42.6	6.6	8.0	48.4	48.0	7.1	7.3
		.05	34.6	33.3	3.6	4.5	39.3	38.7	4.0	4.2
		.01	21.9	20.6	1.0	1.8	25.1	24.5	1.1	1.4
	.91	.10	51.7	50.4	6.6	8.7	55.1	54.6	6.9	7.5
		.05	42.3	40.9	3.7	5.6	46.3	45.8	3.7	4.2
		.01	28.1	26.5	1.1	2.7	31.3	30.7	1.1	1.4
	1.25	.10	56.9	55.9	6.2	8.4	62.4	62.0	7.0	8.3
		.05	47.5	46.3	3.4	4.9	53.7	53.2	4.2	5.0
		.01	32.6	31.6	0.9	1.7	38.4	37.9	1.0	1.4
100	.10	.10	20.1	18.9	8.0	5.8	20.9	20.3	7.8	7.3
		.05	13.0	11.6	4.2	2.9	14.0	13.6	4.3	3.8
		.01	5.5	4.7	1.0	0.6	5.9	5.7	1.2	0.9
	.22	.10	36.8	35.0	7.4	7.3	38.7	38.0	7.7	7.7
		.05	27.3	25.5	4.0	4.0	29.1	28.3	4.1	4.0
		.01	14.4	13.0	0.9	0.9	15.6	15.1	1.1	1.1
	.40	.10	55.7	54.1	7.7	8.3	58.7	58.2	8.2	8.4
		.05	46.0	44.2	4.2	4.7	48.8	48.2	4.5	4.6
		.01	29.1	27.0	1.2	1.6	32.5	31.8	1.4	1.4
	.67	.10	71.6	70.6	7.8	8.5	74.0	73.6	7.8	8.9
		.05	62.8	61.1	4.3	4.7	65.1	64.5	4.5	5.0
		.01	45.7	43.8	1.0	1.5	48.9	48.2	1.2	1.5
	.91	.10	78.3	77.5	7.7	9.2	82.7	82.3	7.6	9.0
		.05	70.9	70.0	4.1	5.3	75.0	74.6	4.2	4.8
		.01	55.1	53.2	1.2	1.9	59.4	58.6	1.1	1.3
	1.25	.10	84.4	83.7	7.2	10.0	87.1	86.9	7.7	11.0
		.05	78.1	77.0	4.1	5.8	81.2	80.9	4.5	6.3
		.01	63.7	62.3	1.1	1.9	68.3	67.7	1.4	2.0

Note: k is the number of groups (individuals) and n_i is the number of observations per group (individual)

Table 3.4: Estimated Type I error of four tests when data are generated from the Lognormal-Poisson mixture distribution under the hypothesis of homogeneity. Levels considered are $10\%, 5\%$ and 1%.

			$n_i = 5$				$5 \leq n_i \leq 20$			
k	c	α	H_S	H_{SC}	H_{NB}	H_{QL}	H_S	H_{SC}	H_{NB}	H_{QL}
10	.10	.10	3.6	3.3	3.9	3.9	4.6	4.5	4.1	4.5
		.05	2.0	1.7	2.2	2.0	2.8	2.7	2.4	2.8
		.01	0.8	0.7	0.7	0.7	1.0	0.9	0.9	0.9
	.22	.10	4.2	3.8	3.8	4.1	5.5	5.4	3.5	3.6
		.05	2.5	2.1	2.1	2.4	3.4	3.2	2.0	2.0
		.01	0.9	0.7	0.7	0.8	1.4	1.3	0.6	0.7
	.40	.10	5.6	5.2	4.3	4.4	7.6	7.5	4.2	4.4
		.05	3.5	3.2	2.3	2.4	5.0	4.9	2.4	2.5
		.01	1.4	1.2	0.7	0.5	2.2	2.2	0.8	0.8
	.67	.10	6.8	6.4	3.9	4.2	9.1	9.0	4.8	5.0
		.05	4.3	4.0	2.1	2.4	6.5	6.3	2.8	2.8
		.01	1.7	1.6	0.6	0.7	3.3	3.2	0.9	0.8
	.91	.10	7.9	7.8	4.6	4.6	10.6	10.4	5.2	5.3
		.05	4.5	5.0	2.6	2.9	7.3	7.2	3.0	3.2
		.01	2.1	2.0	0.5	0.6	3.6	3.6	1.1	1.2
	1.25	.10	8.2	7.9	4.4	4.4	12.5	12.4	5.7	5.8
		.05	5.3	5.2	2.3	2.2	8.7	8.6	3.5	3.6
		.01	2.5	2.5	1.0	1.1	4.7	4.7	1.5	1.5
20	.10	.10	6.7	6.1	5.9	5.9	7.6	7.4	4.8	4.8
		.05	3.9	3.4	3.2	3.5	4.8	4.5	2.7	2.8
		.01	1.4	1.2	0.8	1.2	1.9	1.8	0.7	0.7
	.22	.10	9.2	8.5	5.4	5.7	11.6	11.2	5.4	5.5
		.05	5.8	5.2	3.1	3.3	7.5	7.3	2.9	3.0
		.01	2.3	2.1	0.9	0.9	3.3	3.2	0.8	0.8
	.40	.10	12.9	12.1	5.9	6.3	16.5	16.2	5.5	5.4
		.05	8.8	8.2	3.2	3.4	11.2	11.0	3.3	3.2
		.01	4.0	3.5	0.8	0.9	5.6	5.4	1.1	1.1
	.67	.10	16.6	15.9	6.7	6.9	20.8	20.4	7.1	7.1
		.05	11.7	11.1	4.0	4.0	15.5	15.1	4.2	4.3
		.01	5.8	5.4	1.4	1.5	8.5	8.3	1.5	1.5
	.91	.10	18.7	18.2	7.5	8.1	23.2	23.0	8.1	8.4
		.05	13.5	13.1	4.4	4.8	17.3	17.2	4.9	5.2
		.01	6.9	6.7	1.4	1.5	10.2	10.0	2.1	2.1
	1.25	.10	20.5	20.0	8.6	9.2	26.4	26.2	10.0	10.9
		.05	14.8	14.4	5.1	5.5	20.0	19.7	6.2	6.8
		.01	7.9	7.9	2.0	2.1	12.1	11.9	2.5	2.8

Note: k is the number of groups (individuals) and n_i is the number of observations per group (individual)

Table 3.5: Estimated Type I error of four tests when data are generated from the Log-normal-Poisson mixture distribution under the hypothesis of homogeneity. Levels considered are 10%, 5% and 1%.

			$n_i = 5$				$5 \leq n_i \leq 20$			
k	c	α	H_S	H_{SC}	H_{NB}	H_{QL}	H_S	H_{SC}	H_{NB}	H_{QL}
50	.10	.10	12.2	11.2	6.7	6.8	14.2	13.8	6.8	7.1
		.05	7.4	6.6	3.5	3.3	8.9	8.7	4.0	4.3
		.01	2.9	2.5	0.8	0.6	3.6	3.4	1.1	1.2
	.22	.10	20.8	19.7	7.4	7.7	22.9	22.4	7.4	7.5
		.05	14.5	13.4	4.0	4.1	16.2	15.7	4.1	4.1
		.01	7.2	6.3	1.0	1.1	7.7	7.3	1.2	1.2
	.40	.10	30.9	29.6	8.6	8.7	35.6	35.2	8.6	8.7
		.05	22.9	21.5	4.9	4.9	27.3	26.7	5.1	5.2
		.01	12.9	12.1	1.5	1.5	15.8	15.3	1.8	1.8
	.67	.10	40.8	39.6	10.7	11.0	46.1	45.5	11.4	11.7
		.05	32.2	30.9	6.6	6.8	37.7	37.2	6.9	7.1
		.01	20.1	18.7	2.4	2.5	24.3	23.7	2.2	2.3
	.91	.10	44.3	43.0	13.6	14.0	52.1	51.7	14.4	14.9
		.05	35.8	34.6	8.7	8.9	43.4	42.9	8.9	9.1
		.01	24.0	22.7	3.4	3.4	29.0	28.4	3.3	3.4
	1.25	.10	50.4	49.1	17.8	18.7	56.5	56.1	18.9	20.2
		.05	41.5	40.3	12.2	12.6	47.8	47.4	12.5	13.3
		.01	27.9	27.2	5.6	5.9	33.7	33.3	5.8	6.0
100	.10	.10	20.4	18.9	8.2	8.7	21.1	20.6	8.0	8.3
		.05	13.3	12.0	4.5	4.9	13.9	13.6	4.2	4.4
		.01	5.8	4.8	1.2	1.2	5.8	5.5	1.2	1.1
	.22	.10	35.4	33.9	8.6	8.7	38.9	38.3	8.5	8.5
		.05	26.3	24.7	4.4	4.5	29.4	28.5	4.8	4.9
		.01	14.2	12.7	1.2	1.1	15.5	14.8	1.4	1.4
	.40	.10	52.6	51.3	10.9	10.9	56.1	55.7	10.1	10.3
		.05	43.4	41.4	6.2	6.3	47.0	46.3	6.0	6.1
		.01	27.4	25.3	1.8	1.8	30.7	29.8	1.9	2.0
	.67	.10	66.4	65.1	15.7	16.3	71.7	71.1	15.3	15.8
		.05	57.5	56.0	9.8	10.2	63.5	62.7	9.6	9.9
		.01	42.1	40.2	3.7	3.8	47.0	46.2	3.5	3.6
	.91	.10	71.9	70.6	21.1	21.6	77.7	77.4	21.6	22.7
		.05	63.5	62.1	14.2	14.6	69.7	69.2	14.3	15.3
		.01	48.7	46.8	6.5	6.7	54.2	53.5	5.8	6.4
	1.25	.10	75.9	74.8	28.2	30.3	82.0	81.8	28.9	31.8
		.05	68.6	67.3	20.3	21.9	75.3	74.9	20.3	22.6
		.01	53.6	52.3	10.5	11.1	61.2	60.7	9.9	11.2

Note: k is the number of groups (individuals) and n_i is the number of observations per group (individual)

Table 3.6: Power (in percent) of the four tests when data are generated from negative binomial distribution with $c = .10, .40$ and $k = 20, 50$. Levels considered are $10\%, 5\%$ and 1%.

				$n_i = 5$				$5 \le n_i \le 20$			
k	c	D	α	H_S	H_{SC}	H_{NB}	H_{QL}	H_S	H_{SC}	H_{NB}	H_{QL}
20	.10	0	.10	6.1	5.6	5.5	3.3	7.4	7.2	4.9	3.6
			.05	3.7	3.2	2.9	1.8	4.3	4.1	2.8	2.0
			.01	1.3	1.1	0.8	0.6	1.7	1.5	0.8	0.6
		.05	.10	16.6	15.6	14.7	16.3	41.6	41.1	38.0	41.7
			.05	11.5	10.3	9.4	10.4	33.5	32.8	29.2	32.7
			.01	5.6	4.9	3.4	4.1	21.6	20.9	17.2	19.6
		.10	.10	28.9	27.4	27.9	30.8	67.6	67.0	64.3	68.6
			.05	21.7	20.1	20.0	22.5	60.2	59.7	56.0	60.4
			.01	11.9	10.7	8.7	10.3	46.7	45.8	41.4	45.6
		.15	.10	40.1	38.4	39.3	43.0	82.0	81.6	80.6	83.9
			.05	31.8	30.0	29.5	32.5	76.5	76.0	74.6	78.4
			.01	19.7	17.9	15.7	18.0	66.1	65.3	62.0	66.4
		.20	.10	50.7	49.2	49.6	53.1	89.6	89.4	89.1	91.1
			.05	42.0	40.1	39.4	43.1	86.1	85.7	85.1	87.5
			.01	28.4	26.5	23.7	26.7	77.2	76.6	75.7	78.4
	.40	0	.10	13.9	13.1	5.2	7.3	16.2	15.8	5.5	5.5
			.05	9.2	8.4	2.9	4.8	11.2	10.9	2.6	2.7
			.01	4.1	3.7	0.8	2.7	6.0	5.7	0.7	0.7
		.05	.10	22.2	21.1	10.7	11.2	43.4	43.0	27.2	27.5
			.05	16.2	15.2	6.5	6.9	35.4	34.8	19.4	19.6
			.01	8.4	7.8	2.5	2.6	23.2	22.6	10.2	10.3
		.10	.10	30.5	29.5	18.3	19.1	64.0	63.5	50.7	50.7
			.05	23.8	22.5	12.1	12.8	56.3	55.8	41.9	41.8
			.01	14.0	12.9	4.9	5.1	43.0	42.2	27.7	27.7
		.15	.10	40.6	39.0	27.8	28.7	77.0	76.6	68.7	68.8
			.05	32.2	30.7	19.4	20.3	70.6	70.1	60.2	60.2
			.01	20.2	18.9	9.6	10.2	58.5	57.8	45.8	45.6
		.20	.10	47.6	46.2	35.4	36.1	84.8	84.5	79.3	79.4
			.05	39.1	37.6	26.7	27.4	79.8	79.3	72.8	72.7
			.01	26.4	24.9	14.3	14.6	69.9	69.2	59.7	59.4
50	.10	0	.10	12.7	11.6	6.9	4.4	13.8	13.3	6.7	6.6
			.05	8.0	7.0	3.5	2.4	8.7	8.3	3.8	3.4
			.01	3.3	2.9	0.7	0.6	3.5	3.3	1.2	0.9
		.05	.10	37.8	36.1	27.1	30.3	77.6	77.1	68.9	71.7
			.05	28.8	26.7	18.5	21.2	70.6	70.0	59.5	62.4
			.01	16.0	14.3	7.6	9.1	54.9	53.9	41.7	44.8
		.10	.10	62.8	61.1	51.4	54.9	95.9	95.7	94.3	94.9
			.05	53.2	51.1	40.4	43.7	94.2	93.9	90.9	92.0
			.01	36.4	33.9	22.9	25.4	87.5	87.0	82.2	83.7
		.15	.10	79.2	77.8	71.2	73.2	99.4	99.4	99.1	99.3
			.05	71.6	69.6	60.8	63.4	98.9	98.9	98.3	98.5
			.01	56.5	54.0	42.2	44.4	96.8	96.6	95.3	95.9
		.20	.10	89.3	88.5	84.4	85.5	99.8	99.8	99.8	99.8
			.05	83.9	82.8	76.9	78.2	99.8	99.7	99.7	99.7
			.01	71.9	69.7	60.4	61.9	99.1	99.1	98.9	98.9
	.40	0	.10	32.8	31.3	6.7	7.8	36.1	35.6	7.4	7.5
			.05	24.4	23.1	3.7	4.8	27.2	26.5	3.9	4.0
			.01	13.8	12.7	0.9	2.0	15.4	15.0	1.0	1.0
		.05	.10	53.1	51.6	20.1	20.4	81.1	80.8	53.2	53.5
			.05	43.6	41.8	13.1	13.1	74.4	73.7	42.8	43.2
			.01	28.5	26.6	5.1	5.2	59.3	58.4	25.9	26.2
		.10	.10	68.4	67.1	37.6	38.0	94.8	94.7	85.4	85.7
			.05	60.1	58.3	27.5	28.1	92.2	91.9	79.2	79.3
			.01	44.1	41.8	14.3	14.4	85.3	84.8	65.2	65.1
		.15	.10	78.9	77.8	53.5	54.3	98.8	98.8	96.1	96.0
			.05	71.5	69.9	43.1	43.8	97.9	97.9	93.8	93.7
			.01	56.4	54.5	26.3	26.9	95.4	95.1	87.0	87.1
		.20	.10	86.7	85.8	68.5	68.9	99.7	99.6	98.9	98.9
			.05	80.9	79.8	58.9	59.2	99.2	99.2	98.0	98.0
			.01	69.1	66.8	40.5	40.8	98.2	98.1	95.3	95.3

Note: k is the number of groups (individuals) and n_i is the number of observations per group (individual)

Table 3.7: Power (in percent) of the four tests when data are generated from Log-normal-Poisson mixture distribution with $c = .10, .40$ and $k = 20, 50$. Levels considered are 10%, 5% and 1%.

				$n_i = 5$				$5 \le n_i \le 20$			
k	c	D	α	H_S	H_{SC}	H_{NB}	H_{QL}	H_S	H_{SC}	H_{NB}	H_{QL}
20	.10	0	.10	6.7	6.1	5.9	5.9	7.6	7.4	4.8	4.8
			.05	3.9	3.4	3.2	3.5	4.8	4.5	2.7	2.8
			.01	1.4	1.2	0.8	1.2	1.9	1.8	0.7	0.7
		.05	.10	16.5	15.3	16.6	16.9	41.2	40.6	38.9	33.9
			.05	11.0	10.0	10.8	10.9	33.1	32.4	30.4	25.8
			.01	5.3	4.5	4.2	4.3	21.1	20.5	17.3	14.4
		.10	.10	29.0	27.4	28.1	21.7	65.9	65.2	66.2	60.9
			.05	21.1	19.7	19.9	15.1	58.4	57.7	57.6	52.5
			.01	11.7	10.5	9.8	7.4	44.9	44.0	42.7	38.3
		.15	.10	39.6	38.0	41.8	34.0	81.7	81.4	82.1	78.9
			.05	31.3	29.5	31.6	25.6	76.2	75.7	76.1	72.5
			.01	19.4	17.7	17.0	13.8	65.4	64.7	63.6	59.8
		.20	.10	50.5	48.8	51.9	45.6	89.2	88.9	90.1	88.2
			.05	41.8	40.1	41.7	36.2	85.2	84.7	85.9	83.6
			.01	28.8	26.6	26.5	22.5	77.1	76.6	76.7	74.0
	.40	0	.10	12.9	12.1	5.9	6.3	16.5	16.2	5.5	5.4
			.05	8.8	8.2	3.2	3.4	11.2	11.0	3.3	3.2
			.01	4.0	3.5	0.8	0.9	5.6	5.4	1.1	1.1
		.05	.10	21.0	20.3	12.3	13.6	42.1	41.6	28.4	28.8
			.05	15.2	14.4	7.4	8.9	34.5	33.9	21.1	21.6
			.01	8.3	7.5	3.1	4.4	22.5	21.9	11.3	11.8
		.10	.10	30.7	29.3	21.0	22.1	63.3	63.0	52.7	53.1
			.05	23.7	22.5	14.5	16.1	55.7	55.1	43.4	44.0
			.01	14.0	13.1	6.7	8.4	42.1	41.3	29.6	30.3
		.15	.10	39.5	38.1	29.6	31.4	77.6	77.3	70.8	70.9
			.05	31.1	29.6	21.0	23.1	71.4	70.9	63.4	63.8
			.01	19.4	18.1	10.8	13.0	59.3	58.7	48.6	49.1
		.20	.10	47.9	46.7	38.8	40.5	85.4	85.1	81.5	81.4
			.05	39.7	37.8	29.6	32.0	80.4	80.0	75.5	75.1
			.01	26.6	25.3	16.6	19.4	70.6	70.1	63.1	63.3
50	.10	0	.10	12.2	11.2	6.7	6.8	14.2	13.8	6.8	7.1
			.05	7.4	6.6	3.5	3.3	8.9	8.7	4.0	4.3
			.01	2.9	2.5	0.8	0.6	3.6	3.4	1.1	1.2
		.05	.10	38.6	36.9	28.9	23.8	77.9	77.3	70.0	67.6
			.05	29.6	27.6	19.9	16.0	70.6	70.0	61.0	59.0
			.01	16.6	14.9	8.8	6.7	55.4	54.5	43.0	41.1
		.10	.10	63.1	61.4	53.0	49.0	96.2	96.0	94.3	93.9
			.05	53.5	50.8	41.5	38.3	93.8	93.6	91.4	90.8
			.01	36.5	34.3	24.3	22.4	87.8	87.4	82.8	82.1
		.15	.10	79.3	78.1	71.9	69.9	99.3	99.3	99.0	99.1
			.05	72.2	70.2	62.9	60.6	99.0	99.0	98.6	98.5
			.01	56.3	53.7	43.3	41.5	97.2	97.1	95.9	95.8
		.20	.10	88.3	87.4	83.9	83.0	99.9	99.9	99.9	99.8
			.05	83.2	82.0	77.4	76.4	99.7	99.7	99.7	99.7
			.01	71.3	69.0	61.1	60.3	99.2	99.2	99.1	99.1
	.40	0	.10	30.9	29.6	8.6	8.7	35.6	35.2	8.6	8.7
			.05	22.9	21.5	4.9	4.9	27.3	26.7	5.1	5.2
			.01	12.9	12.1	1.5	1.5	15.8	15.3	1.8	1.8
		.05	.10	51.0	49.4	23.5	24.3	49.5	79.0	56.8	57.3
			.05	41.4	39.4	15.5	16.2	72.2	71.8	46.8	47.2
			.01	27.0	25.2	6.4	7.0	58.3	57.5	30.4	30.7
		.10	.10	67.0	65.2	42.6	43.0	94.5	94.4	87.6	87.7
			.05	58.3	56.4	31.8	32.1	91.9	91.7	81.7	81.9
			.01	42.6	40.7	16.6	17.0	85.5	85.1	68.7	69.0
		.15	.10	78.7	77.7	59.0	59.3	98.7	98.6	96.6	96.7
			.05	71.7	70.2	49.1	50.0	97.7	97.6	94.4	94.5
			.01	56.8	54.5	31.5	31.6	94.9	94.8	88.6	88.8
		.20	.10	86.3	85.5	72.2	72.6	99.6	99.6	99.1	99.2
			.05	81.1	79.8	63.0	63.4	99.3	99.3	98.4	98.5
			.01	68.5	66.1	43.6	44.0	98.2	98.1	96.3	96.4

Note: k is the number of groups (individuals) and n_i is the number of observations per group (individual)

Table 3.8: Power (in percent) of the four tests when data are generated from heterogeneous negative binomial distribution with $.1 \leq c \leq 1.0$. Levels considered are $10\%, 5\%$ and 1%.

			$n_i = 5$				$5 \leq n_i \leq 20$			
k	D	α	H_S	H_{SC}	H_{NB}	H_{QL}	H_S	H_{SC}	H_{NB}	H_{QL}
20	0	.10	16.2	15.4	6.2	6.4	19.2	18.9	5.4	5.6
		.05	11.7	10.8	3.4	3.4	13.7	13.1	3.0	3.1
		.01	5.6	5.3	1.1	1.2	7.0	6.8	1.0	1.1
	.05	.10	23.0	22.0	10.6	10.9	43.0	42.3	24.8	25.3
		.05	16.8	15.6	6.3	6.5	35.1	34.7	18.0	18.5
		.01	9.2	8.4	2.3	2.3	23.0	22.3	9.2	9.5
	.10	.10	31.2	30.0	17.2	17.7	62.4	62.0	47.2	47.9
		.05	23.6	22.3	11.0	11.3	54.9	54.4	38.2	38.7
		.01	13.5	12.5	4.7	4.9	42.2	41.3	24.6	25.1
	.15	.10	40.2	38.7	25.0	25.6	74.5	74.3	63.6	64.1
		.05	32.0	30.5	17.0	17.5	67.9	67.5	55.0	55.7
		.01	20.0	18.6	8.4	8.6	56.0	55.3	40.0	40.8
	.20	.10	47.1	45.8	33.5	34.3	83.3	83.1	76.7	77.4
		.05	38.4	37.2	24.3	25.0	78.1	77.7	69.2	70.0
		.01	26.5	24.9	13.0	13.5	67.3	66.7	55.5	56.0
50	0	.10	38.6	37.5	7.2	7.4	44.0	43.4	7.8	8.1
		.05	30.1	28.6	4.1	4.3	34.6	34.1	4.4	4.7
		.01	17.8	16.4	1.1	1.2	21.9	21.3	1.2	1.2
	.05	.10	55.7	54.6	20.3	20.9	79.8	79.4	49.6	50.3
		.05	46.9	45.4	14.1	14.5	72.9	72.5	39.1	39.7
		.01	31.8	29.9	5.6	5.9	59.1	58.4	23.4	23.9
	.10	.10	69.4	68.0	34.9	36.0	94.8	94.6	81.5	82.0
		.05	60.5	58.8	26.1	27.1	91.8	91.5	73.9	74.5
		.01	44.8	42.9	13.4	13.8	84.0	83.4	58.6	59.5
	.15	.10	78.0	76.9	49.6	50.7	98.3	98.2	93.7	93.8
		.05	70.6	69.3	39.1	40.1	97.3	97.3	90.5	90.8
		.01	56.5	54.2	23.0	23.9	93.8	93.5	82.1	82.7
	.20	.10	86.1	85.1	64.3	65.7	99.5	99.5	98.5	98.5
		.05	79.7	78.4	54.0	55.5	99.0	99.0	97.2	97.3
		.01	66.7	64.8	35.3	36.6	97.7	97.6	93.4	93.6

Note: k is the number of groups (individuals) and n_i is the number of observations per group (individual)

Table 3.9: Power (in percent) of the four tests when data are generated from heterogeneous Log-normal-Poisson mixture distribution with $.1 \leq c \leq 1.0$. Levels considered are $10\%, 5\%$ and 1%.

			$n_i = 5$				$5 \leq n_i \leq 20$			
k	D	α	H_S	H_{SC}	H_{NB}	H_{QL}	H_S	H_{SC}	H_{NB}	H_{QL}
20	0	.10	14.3	13.6	6.6	6.5	17.7	17.3	6.9	7.0
		.05	9.7	9.1	3.7	3.5	12.6	12.3	4.1	4.2
		.01	4.8	4.3	1.0	1.1	6.7	6.5	1.5	1.5
	.05	.10	22.3	20.9	12.6	12.5	40.6	40.1	27.0	27.3
		.05	15.9	14.9	7.9	7.7	32.4	31.9	20.0	20.1
		.01	8.9	8.2	2.9	2.8	21.5	21.2	10.1	10.3
	.10	.10	31.1	30.2	21.1	21.2	61.7	61.3	51.6	51.8
		.05	24.3	23.5	14.2	14.4	54.2	53.5	43.0	43.4
		.01	14.6	13.5	6.9	6.9	40.5	39.9	29.2	29.6
	.15	.10	38.8	37.3	28.2	28.4	73.8	73.4	67.3	67.6
		.05	31.4	30.0	20.6	20.5	67.4	67.0	59.4	59.9
		.01	18.9	17.8	10.4	10.6	55.6	54.8	45.6	45.9
	.20	.10	47.1	45.7	36.4	36.7	82.6	82.4	79.1	79.4
		.05	38.3	37.1	26.9	27.2	77.4	77.0	72.4	72.7
		.01	25.5	24.2	14.9	15.1	67.4	66.8	59.2	59.6
50	0	.10	35.7	34.5	10.6	10.6	40.0	39.7	11.5	11.6
		.05	27.2	25.4	6.5	6.5	31.7	31.4	7.0	7.1
		.01	16.0	15.0	2.4	2.4	20.0	19.7	3.0	3.1
	.05	.10	52.4	50.9	24.8	25.2	79.2	78.8	56.2	56.7
		.05	43.8	42.1	16.6	17.0	71.8	71.3	46.0	46.5
		.01	29.3	27.6	7.7	7.7	58.8	58.0	28.8	29.3
	.10	.10	67.8	66.4	42.8	43.2	93.7	93.5	85.5	85.9
		.05	58.7	56.7	32.1	32.5	90.5	90.3	79.4	79.8
		.01	43.4	41.4	17.1	17.4	82.9	82.4	66.1	66.6
	.15	.10	78.1	77.1	58.1	58.7	98.5	98.5	96.5	96.7
		.05	70.9	69.4	47.3	48.0	97.6	97.6	94.2	94.4
		.01	57.0	54.7	30.6	31.1	94.4	94.1	87.2	87.6
	.20	.10	84.6	83.8	70.4	70.9	99.6	99.6	99.1	99.0
		.05	78.4	76.8	60.7	61.2	99.1	99.1	98.4	98.4
		.01	65.3	63.3	43.0	43.7	97.9	97.9	95.4	95.5

Note: k is the number of groups (individuals) and n_i is the number of observations per group (individual)

Table 3.10: Epileptic seizures counts for 59 epileptics obtained from a placebo-controlled clinical trial of an anti-epileptic drug

Patient ID	Treatment	Age	Baseline	First 2 week	Second 2 week	Third 2 week	Fourth 2 week
1	0	31	11	5	3	3	3
2	0	30	11	3	5	3	3
3	0	25	6	2	4	0	5
4	0	36	8	4	4	1	4
5	0	22	66	7	18	9	21
6	0	29	27	5	2	8	7
7	0	31	12	6	4	0	2
8	0	36	52	40	20	23	12
9	0	37	23	5	6	6	5
10	0	28	10	14	13	6	0
11	0	36	52	26	12	6	22
12	0	24	33	12	6	8	5
13	0	28	18	4	4	6	2
14	0	36	42	7	9	12	14
15	0	26	87	16	24	10	9
16	0	26	50	11	0	0	5
17	0	28	18	0	0	3	3
18	0	31	111	37	29	28	29
19	0	32	18	3	5	2	5
20	0	21	20	3	0	6	7
21	0	29	12	3	4	3	4
22	0	21	9	3	4	3	4
23	0	32	17	2	3	3	5
24	0	25	28	8	12	2	8
25	0	30	55	18	24	76	25
26	0	40	9	2	1	2	1
27	0	19	10	3	1	4	2
28	0	22	47	13	15	13	12
29	1	18	76	11	14	9	8
30	1	32	38	8	7	9	4
31	1	20	19	0	4	3	0
32	1	20	10	3	6	1	3
33	1	18	19	2	6	7	4
34	1	24	24	4	3	1	3
35	1	30	31	22	17	19	16
36	1	35	14	5	4	7	4
37	1	57	11	2	4	0	4
38	1	20	67	3	7	7	7
39	1	22	41	4	18	2	5
40	1	28	7	2	1	1	0
41	1	23	22	0	2	4	0
42	1	40	13	5	4	0	3
43	1	43	46	11	14	25	15
44	1	21	36	10	5	3	8
45	1	35	38	19	7	6	7
46	1	25	7	1	1	2	4
47	1	26	36	6	10	8	8
48	1	25	11	2	1	0	0
49	1	22	151	102	65	72	63
50	1	32	22	4	3	2	4
51	1	25	42	8	6	5	7
52	1	35	32	1	3	1	5
53	1	21	56	18	11	28	13
54	1	41	24	6	3	4	0
55	1	32	16	3	5	4	3
56	1	26	22	1	23	19	8
57	1	21	25	2	3	0	1
58	1	36	13	0	0	0	0
59	1	37	12	1	4	3	2

Table 3.11: Estimated mean and variances (in parentheses) of seizure counts for the placebo and progabide groups

Group	Baseline	Week 2	Week 4	Week 6	Week 8
Placebo	30.79	9.36	8.29	8.79	8.00
	(681.21)	(102.82)	(66.58)	(215.21)	(57.91)
Progabide	31.65	8.58	8.42	8.13	6.74
	(783.44)	(332.70)	(140.66)	(192.93)	(126.56)

Table 3.12: Estimates of the random patient effects for the epileptic seizures count data

-0.5741	-0.5889	-0.9982	-0.6561	0.8113	0.0283	-0.6109	1.2244	0.0680
-0.1127	1.0044	0.2117	-0.3355	0.6655	1.0613	0.2727	-0.6553	1.6127
-0.3032	-0.3875	-0.5683	-0.7968	-0.3921	0.1309	1.4293	-0.8858	-0.9480
0.6215	0.8029	0.4433	-0.6268	-0.7366	-0.3052	-0.2921	0.8733	-0.1536
-0.2736	0.5755	0.3471	-1.2280	-0.5145	-0.3653	1.1311	0.2125	0.6421
-1.0283	0.3802	-1.0840	2.2146	-0.1714	0.3648	0.0482	0.9151	0.0174
-0.2856	0.4502	-0.4507	-0.9865	-0.5255				

Table 3.13: Estimated mean and variance (in parentheses) of skin cancer counts for the two treatment groups

Treatment Group	Year 1	Year 2	Year 3	Year 4	Year 5
Placebo	0.271	0.240	0.247	0.233	0.272
	(0.762)	(0.477)	(0.607)	(0.611)	(0.716)
Beta Carotene	0.298	0.261	0.286	0.315	0.298
	(0.646)	(0.457)	(1.117)	(1.263)	(0.803)

Table 3.14: Estimates of the random patient effects for the skin cancer data

-0.2739	-0.3564	0.6570	1.0631	-0.3273	1.0587	-0.4872	0.1870	0.3588
-0.1708	0.1418	-0.5120	0.6904	-0.2849	0.5818	-0.5441	-0.2144	-0.2356
-0.4964	0.4217	1.2110	-0.2423	0.7035	-0.5089	-0.4804	-0.2263	0.5117
-0.2013	-0.1923	0.0698	0.8892	-0.2312	-0.2468	-0.2076	-0.4478	0.5204
-0.0265	-0.1665	0.2991	-0.4110	0.6423	-0.3991	0.7190	-0.1139	0.2035
-0.2628	-0.4751	-0.2814	0.3359	-0.2111	-0.4255	-0.2569	2.2347	-0.4449
-0.3253	-0.2946	-0.4112	0.6270	-0.3257	1.7871	-0.3297	-0.3350	0.3487
-0.3845	-0.2488	1.0704	0.7222	-0.2423	-0.4222	0.9001	-0.2319	-0.3591
-0.4632	-0.3538	0.7100	0.6033	-0.2044	1.5790	-0.9995	-0.3793	-0.2044
-0.3303	0.4558	1.3072	-0.2350	0.1134	0.4994	-0.2681	-0.2206	-0.2531
0.0555	-0.3412	-0.1351	1.3314	-0.5507	-0.2479	-0.2699	-0.9253	-0.6938
0.5735	0.4954	0.2331	-0.2611	-0.2860	-0.4596	-0.2791	-0.3166	-0.2854
-0.2393	1.5577	-0.1923	-0.1667	1.3684	1.2930	0.3270	-0.4425	-0.2384
-0.3073	-0.3182	-0.2955	0.4301	1.0023	-0.2733	0.7784	-1.8394	-0.2573
-0.2569	1.0634	-0.5906	-0.2384	-0.3227	0.2528	0.1863	-0.3270	-0.2180
0.7679	0.2987	-0.3789	0.1770	-0.5411	-0.4993	0.2528	-0.3361	-0.3397
-0.3739	-0.5276	-0.2211	0.8195	-0.4624	-0.1161	-0.3253	-0.2082	1.1091
0.1878	-0.2534	-0.4588	1.2095	0.3623	-0.2312	-0.2247	-0.2855	0.7510
0.3721	-0.4302	-0.2277	-0.1953	-0.1921	1.3724	1.0686	-0.4282	-0.3481
1.6767	0.3490	-0.0027	1.3611	0.3310	-0.3036	-0.2925	-0.2855	1.0874
-0.1775	-0.2534	-0.3946	-0.2315	0.3299	-0.2731	-0.2486	1.0414	-0.5744
-0.2982	0.2531	-0.4223	0.1924	0.1440	1.7466	-0.3190	-0.3946	0.7251
-0.4478	0.4301	0.3402	0.4837	-0.0241	0.4133	-0.1068	-0.3208	-0.4031
2.1554	-0.3950	-0.2791	-0.3121	-0.2573	1.0885	0.2393	0.8080	0.5396
-0.2808	0.6244	0.2848	-0.3641	0.8609	0.2903	-0.4865	-1.6559	0.3044
-0.2479	-0.4334	-0.5216	-0.6752	-0.4424	-0.7139	-0.4217	-0.4171	-0.3880
-0.4770	-1.6162	-0.0047	-0.4107	-0.3121	0.2348	-0.2865	-0.3033	1.2329
-0.4278	-0.5016	-0.3925	-0.3077	-0.1308	-0.5054	1.3378	0.4794	-0.5949
-0.4286	0.3918	0.3134	-0.3303	0.3524	0.5616	-0.4655	0.2576	0.5685
-0.3350	-0.4409	-0.3292	-0.5242	-0.1834	-0.3564	-0.3320	-0.1682	-0.3822
-0.4416	-0.2144	-0.4008	-0.2628	-0.3629	-0.2460	-0.2420	-0.3046	-0.3249
-0.2875	-0.3738	-0.7979	-0.5774	-0.3860	-0.3845	0.0130	1.2574	-0.3121
-0.1749	0.3224	0.7566	-0.3246	-0.8726	-0.1749	-1.8050	-0.2277	0.2435
-0.3446	-0.2333	-0.2015	-0.3793	-0.3346	-0.2700	1.5688	-0.3109	1.7111
-0.4081	-0.2176	-0.2566	-0.2013	0.5209	-1.1047	1.5045	-0.2248	-0.3443
-0.5178	-0.4166	-0.3765	-0.2481	-0.2420	-0.2143	0.7735	-0.0582	0.4467
-0.3166	-0.2894	1.5915	-0.3135	-0.3565	-0.3589	-0.3282	-0.3900	-0.1862
1.4463	0.3521	-0.2230	-0.4168	-0.5411	-0.1605	-0.4225	-0.0948	-0.2244
-0.4958	-0.3543	-0.3351	0.9180	-0.3318	-0.2247	-0.3258	1.0517	-0.2252
0.4415	0.4339	-0.3540	-0.2390	-0.3158	0.2903	-0.2440	-0.2821	-0.3056
-0.3249	-0.2194	0.8718	-0.2178	-0.1952	0.6345	-0.1956	-0.1615	-0.2466
-0.2317	0.4088	-0.3646	-0.4331	-0.2321	-0.1614	-0.3790	-0.2179	1.2664
0.2202	-0.1720	-0.1983	-0.3077	-0.4227	-0.2472	1.1224	-0.1228	0.6306
1.8646	-0.3899	-0.3037	-0.3258	0.3922	-0.1614	-0.3745	0.2384	-0.2863
-0.4005	-0.2740	-0.3255	-0.3351	-0.5021	0.5194	-0.3375	-0.1864	1.1010
-0.1720	0.4422	0.2941	-0.3003	-0.1349	-0.2654	-0.2240	-0.3074	0.4755
-0.5309	0.5828	0.3473	-0.3214	-0.2320	0.5643	1.0727	1.2665	0.8568
0.9189	-0.3314	-0.2247	-0.2754	1.5633	-0.1553	-0.4238	2.0987	0.7603
-0.1721	-0.5021	1.5795	-0.5051	-0.2728	-0.7041	1.6108	-0.2015	-0.7594
-1.9591	0.7746	1.0644	-0.2214	0.9678	-0.3709	0.2903	-0.4395	-0.2894
1.5200	-0.4655	-0.6215	1.5588	-0.7280	-0.0981	-0.2582	0.9200	1.6445
-0.6130	-0.4350	1.2926	0.3916	1.9082	-0.2546	-0.3860	0.5925	-0.5087
-0.3850	1.0303	1.5632	0.7812	-0.3445	-0.4749	-0.6410	-0.2318	-0.2849
-0.3605	-0.2356	-0.2794	-0.3718	-0.6193	-0.3367	-0.3166	-0.1957	1.4642
2.6853	0.3428	-0.2620	1.9788	-0.3200	0.7259	0.6188	-0.1101	0.7886
-0.3969	0.6494	0.4384	-0.1738	0.5331	1.7004	0.5375	0.7307	-0.0500
0.9450	-0.1754	1.8445	0.7196	0.3697	-0.3838	0.5828	0.8945	-0.1060
0.9199	-0.1456	-0.4183	-1.0985	1.3945	-0.3319	1.2531	-0.9199	-0.3253
-0.3080	-0.1244	1.0578	-0.6830	1.1519	1.3960	-0.4107	0.1141	-0.4426
0.2610	0.8565	2.2265	1.2868	0.4298	0.0641	1.0923	1.1821	-0.1896
0.6174	-1.0745	-0.1253	-0.8338	-0.3303	1.8262	0.3001	-0.3164	-0.3952
-0.4748	-0.9093	1.0201	-0.2286	-0.6981	-0.0653	-0.6620	-0.4171	1.2203
-0.4162	-0.8717	-0.2737	-0.0051	-0.5447	0.9303	-0.0990	0.2348	0.1685
-0.1192	-0.3257	0.3174	-0.1303	0.5002	-0.4654	1.4525	-0.4111	-1.1766
1.5720	0.9339	-0.3637	0.2438	0.3870	0.5159	-0.4059	0.2790	1.0303
1.5438	-0.4444	1.4303	0.3854	-1.7313	-0.3445	0.2247	0.7842	0.9912
1.5764	1.3946	-0.3077	-0.1380	1.8640	-0.4274	1.6888	-0.2029	-0.3591
-0.3689	0.2852	2.0541	-0.5466	-0.2875	0.3954	0.2846	-0.0279	0.2846
-0.5619	0.2442	0.4619	1.6765	0.5571	0.1871	1.4720	0.5189	0.4052

Table 3.15: Estimates of the random patient effects for the skin cancer data

-0.4804	0.6894	-1.5417	-0.4062	0.3568	-0.7024	0.7034	-0.3477	1.7568
-0.9340	-0.9990	-0.3928	1.1246	0.8226	1.3658	-0.2184	0.7290	2.0784
-0.1689	0.4716	-0.4363	1.8167	0.3438	-0.2536	0.7874	0.8423	1.5733
0.5999	1.1341	-0.2885	0.1442	-0.4931	1.8956	-0.5154	-0.9524	0.6274
1.2025	-0.6099	1.2772	-0.2814	-0.2761	-0.1920	-0.3902	1.1874	-0.5992
2.2305	1.2485	-0.2660	-0.3609	1.8550	0.0881	1.6640	-0.1751	-0.1897
-0.1262	-0.2212	-0.2653	-0.2356	-0.1660	0.3480	-0.2132	-0.4729	0.8415
-0.3977	-0.2172	-0.3357	-0.2882	-0.1286	-0.5146	-0.1477	-0.2621	0.3026
-0.3902	-0.2252	-0.1191	1.3042	-0.1972	-0.4354	1.0900	-0.3145	1.6674
1.2654	-0.4256	0.8066	1.6954	2.4975	0.5614	1.2494	-0.3894	1.2923
-0.4166	0.8473	0.7627	-0.1284	0.5451	1.4410	-0.2345	0.1480	-0.3850
-0.3074	-0.3122	-0.2244	0.9331	-0.3842	0.5422	-0.2285	0.5903	1.0117
-0.2678	0.6232	1.9265	-0.1767	-0.3060	0.4381	-0.2338	-0.2447	-0.8598
-0.4924	0.7784	0.3571	-0.8026	-0.3493	-0.3542	0.6164	-0.2417	-1.0820
0.1181	-0.4065	-0.1685	-0.6522	-0.1617	0.9015	0.2886	0.2455	0.7891
-0.3845	1.8920	0.7311	0.4719	0.2804	1.6802	-0.4569	1.5958	-0.5307
-0.7865	1.8741	-0.4483	-0.3825	1.5181	-0.6696	0.3832	-0.3303	0.4088
-0.4710	3.2535	-0.6053	-0.3397	0.8640	-0.0603	0.8262	-0.3166	-0.1923
0.8520	0.3571	-0.5271	1.1992	-0.3281	0.3402	-0.0932	-1.2577	0.6291
-0.7058	-0.5541	-0.3350	-0.6981	0.4133	0.1280	-0.4453	0.4116	0.1527
-0.1754	-0.2211	0.8075	1.4850	0.6364	-0.4452	-0.6124	-0.2897	1.2759
-0.3077	-0.4253	-0.4474	-0.3144	-0.5472	1.0607	1.4299	-0.2913	-0.5411
0.5253	0.9129	-0.4368	-0.3641	0.6586	-0.0885	2.5784	0.0776	-0.2520
-0.4939	0.8986	0.5427	-0.1696	1.7756	-0.2855	-0.3587	-0.3596	-0.2313
0.1082	-0.5242	-0.1921	-0.5608	0.7105	-0.2534	-0.1972	-0.3493	-0.3589
-0.4368	2.4633	-0.3730	-0.2700	0.3912	-0.3303	-0.4532	-0.1233	0.7129
-0.3372	0.4669	1.0378	-0.3489	-0.4008	0.7959	-0.2424	-0.2181	0.0040
1.2559	-0.5380	-0.5544	-0.4865	0.0426	1.3152	0.5362	-0.5879	2.0881
0.8126	-0.4746	-0.3255	-0.4330	1.2071	0.6857	0.5493	-0.2582	1.2084
0.1521	0.7242	0.5040	0.6771	1.7064	0.0122	1.3537	-0.4391	-0.3031
-0.2082	-0.3033	0.0122	0.2713	0.0130	-0.4279	0.9217	1.1858	1.1667
1.0647	0.3805	-0.3793	-0.4227	-0.3494	1.4410	0.3260	1.9844	0.8089
-0.5472	0.0641	0.4196	1.4142	-0.2572	-0.4117	-0.1195	0.3256	-0.4009
1.1920	-0.3583	-0.3795	-0.4269	-0.4683	-0.3769	-0.3565	-0.4336	2.0455
-0.6517	0.3883	-0.1921	-0.1494	0.6549	0.1474	0.2111	-0.3074	-0.4386
-0.3304	-0.3952	1.5403	-1.2143	-0.3013	0.2104	0.6462	-0.1015	0.4635
0.1928	1.1664	-0.2146	2.5345	-0.3121	-0.1776	-0.3303	-0.2316	-0.6161
0.3345	1.1834	-0.3445	0.6664	0.7789	-0.3350	0.4149	-0.3798	0.2052
-0.4453	-0.2262	-0.3789	-0.3209	1.0301	0.5353	1.0541	-0.5477	0.9283
1.7544	-0.3519	-0.2917	0.6773	-0.2931	-0.2853	-0.2658	0.3349	-0.1775
-0.3850	-0.1835	0.3788	-0.4059	1.0764	-0.2113	-0.4626	2.5158	1.0652
-0.1233	0.3532	-0.3691	0.1577	-0.2652	0.1584	-0.0659	1.3256	-0.1764
-0.3200	-0.4282	-0.3746	0.4513	1.2468	-0.3033	1.0788	-0.4036	-0.4391
-0.1749	-0.2777	0.0782	-0.3211	-0.1982	0.5191	-0.5539	-0.2087	0.5234
-0.4743	1.6101	-0.3211	0.4948	-0.3350	1.6616	-0.3493	0.6671	-0.2430
1.3085	-0.3068	-0.3797	1.0430	-0.1468	0.3490	0.5801	1.3861	-0.3591
0.6904	0.2542	-0.5242	-0.5051	-0.4226	0.8845	-0.2423	-0.3331	1.0852
-0.2114	-0.4003	-0.2162	-0.2109	0.8597	-0.4860	0.3117	-0.4687	-0.2229
-0.5307	-0.2731	0.3710	-0.5604	-0.1667	-0.4657	-0.3048	-0.3530	-0.1923
-0.1777	-0.2659	-0.1832	0.2542	-0.3211	-0.1542	-0.1832	0.4630	2.6129
-0.1958	-0.3238	2.0929	0.2103	0.3791	-0.3029	-0.3950	-0.1861	1.8529
-0.3969	0.5269	-0.1982	1.4817	0.2110	-0.2114	-0.4777	-0.4029	-0.5472
-0.2699	-0.2575	-0.3701	0.3483	0.7575	-0.2763	0.9858	-0.4112	1.5424
-0.3208	-0.2406	-0.3999	-0.6161	-0.3901	2.1161	-0.5474	0.4931	-0.2348
-0.3350	0.9405	-0.2229	-0.4645	-0.5245	-0.5021	-0.3999	1.1697	0.4048
-0.5084	0.5263	-0.4483	-0.2670	-0.3891	-0.3316	-0.4683	-0.3026	1.7740
-0.2885	-0.5512	1.1926	-0.3303	-0.4167	-0.3166	-0.3591	-0.3158	-0.0394
1.3411	-0.3978	0.5818	-0.3036	-0.3397	-0.2531	0.3623	0.5423	-0.4162
1.1136	-0.4717	0.2903	-0.3901	-0.2620	-0.3538	-0.3493	-0.2109	-0.2079
-0.4444	-0.2015	-0.4444	-0.2874	-0.1803	1.4191	-0.0607	-0.3092	-0.4391
-0.2315	-0.3696	-0.4425	-0.0757	0.3127	-0.4226	0.9258	-0.3841	0.4156
0.9243	2.1077	-0.2777	1.6520	-0.4101	-0.1588	-0.3175	0.9597	-0.2690
1.0484	-0.5544	-0.2423	-0.3864	0.7077	0.7196	-0.2534	1.3076	0.1936
0.3306	-0.3266	1.8073	0.6459	-0.4892	0.6057	-0.0633	2.4430	-0.3386
0.0118	0.2251	-0.2175	0.7911	-0.1749	-0.5528	0.8267	0.3270	-0.4280
-0.2209	-0.2894	1.9566	-0.2109	-0.3397	-0.5089	2.0875	-0.3978	0.0782
-0.4624	-0.3253	-0.2014	-0.4222	-0.2468	-0.2214	-0.1861	2.0478	-0.2209
-0.2211	-0.1953	-0.7104	-1.0790	-0.3437	-0.2180	-0.2109	1.2703	1.5684
1.5683	-0.2989	2.8070	-0.4853	-0.2248	-0.3841	-0.3841	0.6775	-0.2049

Table 3.16: Estimates of the random patient effects for the skin cancer data

-0.5541	-0.2946	-0.2658	0.0388	-0.2476	1.8067	-0.2905	-0.2778	-0.4282
1.1749	-0.2087	-0.5116	-0.2623	-0.3255	-0.2987	-0.2286	-0.1746	1.6168
-0.1305	-0.3013	0.3870	0.6608	-0.5949	1.1185	-0.2287	-0.3845	-0.2905
0.9122	0.9252	0.8650	-0.1907	-0.3589	-0.2386	-0.5242	1.5820	-0.4453
-0.1088	-0.5769	-0.3547	-0.4685	-0.4334	-0.3841	2.3115	-0.1926	1.1711
-0.1696	-0.2405	0.4039	1.6864	1.2415	-0.2906	-0.2943	-0.3274	-0.4059
1.2664	2.1429	-0.2444	-0.3581	-0.2944	-0.1990	1.5599	-0.2829	0.6803
1.0845	-0.1669	-0.2778	0.6523	-0.2582	0.6674	-0.2821	-0.1925	-0.4055
-0.4395	-0.2467	0.9180	-0.2324	-0.2286	-0.3258	0.7812	-0.2821	-0.3225
-0.2506	-0.2420	-0.4081	0.3524	1.4995	-0.3133	1.1244	-0.2639	-0.1747
1.6299	-0.2780	-0.3643	-0.4359	-0.0927	-0.4053	0.5122	-0.4107	-0.2386
-0.4195	-0.0247	-0.3227	0.9633	-0.3449	-0.4739	-0.2534	1.3176	1.7635
1.3708	0.6877	0.5512	-0.2714	-0.6538	-0.2243	-0.2299	0.4997	-0.1893
-0.1959	-0.1551	0.3795	-0.7984	-0.3739	1.0361	-0.5218	-0.4227	-0.5813
-0.4005	-0.3301	-0.1223	-0.3104	-0.4426	-0.5509	-0.3547	-0.1685	0.7728
1.9205	2.0111	-0.4201	-0.3589	-0.2660	-0.3952	-0.3495	0.3782	0.6527
0.2058	0.2394	-0.3646	-0.5548	-0.3665	0.2108	-0.3798	-0.6242	-0.2145
0.3733	-0.2697	-0.2700	0.6758	-0.2580	0.4740	1.3915	-0.2821	-0.4688
-0.4871	-0.3165	0.9328	0.3695	0.3563	-0.3498	-0.3122	0.9511	0.4459
-0.3646	-0.5509	-0.2045	2.6204	-0.1563	-0.1840	1.3818	-0.2467	0.3432
0.4553	-0.2495	-0.2099	-0.3236	-0.2360	-0.3036	-0.1779	1.9586	-0.3387
-0.0540	0.3704	-0.0467	0.5129	-0.2212	0.4946	1.3734	0.1183	-0.0391
-0.1474	0.5993	1.4327	0.4299	-0.2145	-0.1513	-0.1959	-0.3615	-0.2004
-0.1162	-0.2902	-0.3161	0.7227	-0.2949	0.8205	-0.3093	0.3836	-0.2542
-0.1587	0.5503	-0.1441	-0.1694	-0.2660	-0.4745	-0.2227	-0.2580	-0.2536
-0.2227	-0.4061	-0.3119	-0.3795	-0.4716	-0.1756	-0.2018	-0.5784	-0.0662
-0.2283	-0.2082	-0.2931	0.5796	-0.2048	-0.3825	-0.4841	0.3303	-0.2211
-0.1692	-0.3142	-0.2087	-0.2660	-0.2580	-0.2335	0.3986	-0.2322	0.5951
-0.1551	0.3524	-0.2860	-0.3304	-0.2218	-0.3643	1.7575	-0.2181	-0.4391
-0.3429	-0.4369	0.4541	0.8258	2.5534	-0.2580	-0.4484	0.3701	0.6948
-0.3598	-0.2990	-0.2181	-0.2731	0.3881	-0.2118	-0.4221	-0.2283	1.8717
-0.2602	0.2848	-0.2894	0.2300	1.4019	-0.3744	1.5497	-0.2506	0.8449
-0.3745	-0.2477	-0.1667	-0.2085	-0.2394	1.0500	0.3071	-0.3693	0.7144
-0.4283	-0.4277	-0.2595	-0.3212	-0.4062	-0.3209	-0.3449	-0.3791	-0.2352
-0.2106	-0.5481	-0.4961	-0.2082	-0.2254	-0.1692	-0.2247	0.5116	-0.2032
-0.5251	-0.1488	0.8367	-0.1163	0.6732	-0.1640	1.1920	-0.3420	-0.7809
-0.1338	0.7099	0.2153	-0.2463	-0.1634	0.7079	-0.2184	-0.4140	-0.2143
0.4904	-0.2426	-0.1749	-0.5546	0.8479	0.7327	1.5108	-0.1828	2.6167
-0.1480	0.5901	1.0283	-0.1484	-0.1743	1.4104	-0.2132	-0.2165	0.8107
-0.3747	0.6269	-0.1456	-0.1979	-0.1978	0.1079	1.6414	-0.2042	-0.2821
-0.1661	1.1746	-0.1430	-0.2761	-0.4068	-0.1978	-0.2784	-0.2704	-0.2975
1.4298	-0.2561	-0.2035	0.6676	-0.1723	-0.1504	0.4836	0.7436	-0.2072
-0.2074	-0.2225	-0.1453	-0.1232	-1.1996	1.1109	-0.1274	-0.1850	-0.2380
-0.1454	-0.1317	-0.2216	-0.2137	-0.1995	-0.2175	-0.1263	-0.1115	-0.1642
-0.2697	-0.1253	-0.2548	1.4584	1.9987	-0.1828	-0.5086	1.4286	-0.2240
1.0065	-0.5260	-0.2723	-0.2452	-0.3523	0.4904	1.3928	0.7719	-0.1942
0.5804	-0.1453	-0.2558	-0.1202	-0.2932	-0.2010	2.4072	-0.1695	-0.2204
1.0801	-0.1716	0.6685	-0.1941	-0.3669	-0.3598	-0.0409	-0.1882	-0.1661
-0.2275	2.2664	-0.4087	1.0484	0.8685	-0.2004	-0.2310	0.6198	-0.1383

Chapter 4

Test for Presence of Over-Dispersion

In Chapter 3 we developed score tests for homogeneity between and within groups for over-dispersed clustered count data. However, in practice, in some situations, the data may not be over-dispersed. In this chapter we develop two score tests for over-dispersion in generalized linear mixed effects model. One of these is based on the over-dispersed generalized linear mixed effects model of Cox (1983) and the other is based on the negative binomial mixed effects model.

4.1 Test based on the over-dispersed generalized linear model (OGLM)

The probability density function that we assume for Y_{ij} is the exponential family

$$f(Y_{ij}; \theta_{ij}) = \exp\{[Y_{ij}\theta_{ij} - g(\theta_{ij})] + C(Y_{ij})\}, \tag{4.1}$$

with the parameters as defined in Section 3.1. To construct the extended family, let the density of Y_{ij} given θ^*_{ij} be $f(Y_{ij}|\theta^*_{ij})$ as given by (4.1), where the θ^*_{ij}'s are continuous independently distributed with finite mean

$$E(\theta^*_{ij}) = \theta_{ij}(x^T_{ij}; \beta)$$

and variance

$$\text{var}(\theta^*_{ij}) = \tau b(\theta_{ij}) > 0.$$

We assume that

$$E\{(\theta^*_{ij} - \theta_{ij})^r\} = \alpha_r; \quad \alpha_r = o(\tau), \quad r \geq 3.$$

In the limit as $\tau \to 0$, this model reduces to natural exponential family.

The probability density function of Y_{ij} in the mixed model is $f_M(Y_{ij}) = E_*\{f(Y_{ij}; \theta^*_{ij})\}$, where E_* denotes expectations over the distribution of θ^*_{ij}. By expanding $f(Y_{ij}; \theta^*_{ij})$ in a Taylor series about θ_{ij} (see Cox (1983) and Chesher (1984)) and taking expectations we obtain

$$f_M(Y_{ij}) = f(Y_{ij}; \theta_{ij}) \left\{ 1 + \sum_{r=2}^{\infty} \frac{\alpha_r}{r!} D_r(Y_{ij}; \theta_{ij}) \right\},$$

where

$$D_r(Y_{ij}; \theta_{ij}) = \left\{ \frac{\partial^{(r)}}{\partial \theta^{*(r)}_{ij}} f(Y_{ij}; \theta^*_{ij})|_{\theta^*_{ij} = \theta_{ij}} \right\} \{f(Y_{ij}; \theta_{ij})\}^{-1}.$$

Further, for small τ and $r = 2$ we have

$$\begin{aligned} f_2(Y_{ij}; \theta_{ij}, \tau) &= f(Y_{ij}; \theta_{ij})\{1 + \frac{\alpha_2}{2!} D_2(Y_{ij}; \theta_{ij})\} \\ &= f(Y_{ij}; \theta_{ij})\{1 + \frac{1}{2}\tau b(\theta_{ij}) D_2(Y_{ij}; \theta_{ij})\}. \end{aligned} \tag{4.2}$$

For tractability, in most practical applications the above form given by (4.2) is used as an approximation to the generalized linear mixed model (see Cox, 1983; Dean, 1992 and Deng and Paul, 2005).

Now, the contribution to the log-likelihood function for the ith group is

$$l_i = \sum_{j=1}^{n_i} \log f_{ij}(Y_{ij}; \theta_{ij}) + \sum_{j=1}^{n_i} \log \left\{ 1 + \frac{1}{2}\tau b(\theta_{ij}) D_2(Y_{ij}; \theta_{ij}) + \sum_{r=3}^{\infty} \frac{\alpha_r}{r!} D_r(Y_{ij}; \theta_{ij}) \right\}. \tag{4.3}$$

Now

$$\begin{aligned} D_2(Y_{ij};\theta_{ij}) &= \frac{1}{f(Y_{ij};\theta_{ij})}\frac{\partial^2}{\partial\theta_{ij}^{*2}}f(Y_{ij};\theta_{ij}^*)|_{\theta_{ij}^*=\theta_{ij}} \\ &= \left(\frac{\partial\log f_{ij}(Y_{ij};\theta_{ij}^*)}{\partial\theta_{ij}^*}|_{\theta_{ij}^*=\theta_{ij}}\right)^2+\frac{\partial^2\log f_{ij}(Y_{ij};\theta_{ij}^*)}{\partial\theta_{ij}^{*2}}|_{\theta_{ij}^*=\theta_{ij}}. \end{aligned}$$

Further,

$$\begin{aligned} \log f_{ij}(Y_{ij};\theta_{ij}^*) &= Y_{ij}\theta_{ij}^* - g(\theta_{ij}^*) + C(Y_{ij}), \\ \frac{\partial\log f_{ij}(Y_{ij};\theta_{ij}^*)}{\partial\theta_{ij}^*} &= Y_{ij} - g'(\theta_{ij}^*) \end{aligned}$$

and

$$\frac{\partial^2\log f_{ij}(Y_{ij};\theta_{ij}^*)}{\partial\theta_{ij}^{*2}} = -g''(\theta_{ij}^*).$$

Therefore

$$D_2(Y_{ij};\theta_{ij}) = (Y_{ij} - g'(\theta_{ij}))^2 - g''(\theta_{ij}).$$

Then, the log-likelihood l_i given in (4.3) becomes

$$\begin{aligned} l_i &= \sum_{j=1}^{n_i}\log f_{ij}(Y_{ij};\theta_{ij}) + \frac{\tau}{2}\sum_{j=1}^{n_i}b(\theta_{ij})\left\{(Y_{ij}-g'(\theta_{ij}))^2 - g''(\theta_{ij})\right\} + O(\tau) \\ &= \sum_{j=1}^{n_i}\{[Y_{ij}\theta_{ij} - g(\theta_{ij})] + C(Y_{ij})\} + \frac{\tau}{2}\sum_{j=1}^{n_i}b(\theta_{ij})\left\{(Y_{ij}-g'(\theta_{ij}))^2 - g''(\theta_{ij})\right\} + O(\tau), \end{aligned} \tag{4.4}$$

where $O(\tau)$ is a higher order function of τ $(\tau^2, \tau^3, etc.)$

Suppose again that with independent responses Y_{ij} we have a $p \times 1$ vector of covariates $X_{ij},\ j = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, k$. Now, we assume the log-linear model, $\theta_{ij} = \log\mu_{ij} = X_{ij}^T\beta$ and $\theta_{ij}^* = X_{ij}^T\beta + \alpha_i$, where the α_i's are iid random variables with $E(\alpha_i) = 0,\ \mathrm{Var}(\alpha_i) = \tau < \infty$. Thus, $E(\theta_{ij}^*) = X_{ij}^T\beta,\ \mathrm{Var}(\theta_{ij}^*) = \tau$ and $b(\theta_{ij}) = 1$. Then it can be shown that the score function for testing over-dispersion, that is, for testing

$H_0 : \tau = 0$ against the alternative $H_A : \tau > 0$ is given by

$$S(\beta) = \sum_{i=1}^{k} \frac{\partial l_i}{\partial \tau}\bigg|_{\tau=0} = \frac{1}{2}\sum_{i=1}^{k}\left\{\sum_{j=1}^{n_i}(Y_{ij}-\mu_{ij})^2 - \sum_{j=1}^{n_i}\mu_{ij}\right\}.$$

The asymptotic variance of the statistic S can be obtained as (Cox and Hinkley, 1974)

$$I = I_{\tau\tau} - I_{\tau\beta} I_{\beta\beta}^{-1} I_{\tau\beta}^{T},$$

where

$$I_{\tau\tau} = \sum_{i=1}^{k} E\left[\frac{\partial l_i}{\partial \tau}\bigg|_{\tau=0}\right]^2, \quad I_{\beta\beta} = \sum_{i=1}^{k} E\left[\left(\frac{\partial l_i}{\partial \beta}\bigg|_{\tau=0}\right)\left(\frac{\partial l_i}{\partial \beta}\bigg|_{\tau=0}\right)^T\right],$$

and

$$I_{\tau\beta} = \sum_{i=1}^{k} E\left[\left(\frac{\partial l_i}{\partial \tau}\bigg|_{\tau=0}\right)\left(\frac{\partial l_i}{\partial \beta}\bigg|_{\tau=0}\right)^T\right].$$

Evaluating the expected values in $I_{\tau\tau}, I_{\beta\beta}$ and $I_{\tau\beta}$, simplifying and replacing the parameter β by its maximum likelihood estimate $\hat{\beta}$ under the null hypothesis, the score test statistic for testing over-dispersion in clustered count data is obtained as

$$T = S^2(\hat{\beta})/[\hat{I}_{\tau\tau} - \hat{I}_{\tau\beta}\hat{I}_{\beta\beta}^{-1}\hat{I}_{\tau\beta}^{T}],$$

where

$$\begin{aligned} S(\hat{\beta}) &= = \frac{1}{2}\sum_{i=1}^{k}\left\{\sum_{j=1}^{n_i}(Y_{ij}-\hat{\mu}_{ij})^2 - \sum_{j=1}^{n_i}\hat{\mu}_{ij}\right\}, \\ \hat{I}_{\tau\tau} &= \frac{1}{4}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\hat{\mu}_{ij} + 2\hat{\mu}_{ij}^2\right), \\ \hat{I}_{\beta\beta} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\hat{\mu}_{ij}X_{ij}X_{ij}^T \end{aligned}$$

and

$$\hat{I}_{\tau\beta} = \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\hat{\mu}_{ij}X_{ij}^T,$$

where $\hat{\mu}_{ij} = e^{X_{ij}^T \hat{\beta}}$ and $\hat{\beta}$ is the maximum likelihood estimate of β, under the Poisson regression model, which can be obtained by Fisher scoring method. Asymptotically, as $k \to \infty$, T has a chi-square distribution with one degree of freedom.

4.2 Test based on the negative binomial model

Now we assume that the data $Y_{ij},\ j = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, k$ come from the negative binomial model

$$f(y_{ij}; \mu_{ij}, c_i) = \frac{\Gamma(y_{ij} + c_i^{-1})}{y_{ij}!\ \Gamma c_i^{-1}} \left(\frac{1}{1 + c_i \mu_{ij}(\mathbf{x})} \right)^{c_i^{-1}} \left(\frac{c_i \mu_{ij}(\mathbf{x})}{1 + c_i \mu_{ij}(\mathbf{x}))} \right)^{y_{ij}}, \tag{4.5}$$

given by equation (3.10). We assume a common over-dispersion parameter c, that is, $c_1 = c_2 = \ldots = c_k = c$. Further, we assume the mixed effects model

$$\theta_{ij} = \log(\mu_{ij}) = X_{ij}^T \beta + \alpha_i, \tag{4.6}$$

where β is a vector of p unknown regression parameters and α_i's are iid random variables having a normal distribution with mean zero and variance D. The log-likelihood function for the ith group is given by

$$l_i(\beta, c) = \log \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}; \beta, c|\alpha_i) f(\alpha_i) d\alpha_i, \tag{4.7}$$

where

$$\log f_{ij}(y_{ij}; \beta, c|\alpha_i) = \left[\sum_{l=0}^{y_{ij}-1} \log(1 + cl) + y_{ij}(X_{ij}^T \beta + \alpha_i) - (y_{ij} + c^{-1}) \log \left(1 + c e^{X_{ij}^T \beta + \alpha_i} \right) \right]. \tag{4.8}$$

Our purpose is to test $H_0 : c = 0$ against the alternative $H_A : c > 0$. To obtain the score function we need to integrate out α_i from (4.7). However, in practice, it is difficult to carry out the integration. So instead we use (4.8) to obtain the log-likelihood and develop the score test for given α_i. That is, for the development of the score test we consider β to be a nuisance parameter and α_i to be known. We deal with the issue of α_i being random later in this chapter. The resulting log-likelihood of β and c for given α_i is

$$\begin{aligned} l &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} \log f_{ij}(y_{ij};\beta, c|\alpha_i) \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left[\sum_{l=0}^{y_{ij}-1} \log(1+cl) + y_{ij}(X_{ij}^T\beta + \alpha_i) - (y_{ij}+c^{-1})\log\left(1+ce^{X_{ij}^T\beta+\alpha_i}\right) \right]. \end{aligned} \tag{4.9}$$

Then the score function for testing $H_0 : c = 0$ is obtained as (see also Collings and Margolin, 1985)

$$\begin{aligned} S_c &= \left.\frac{\partial l}{\partial c}\right|_{c=0} \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left.\frac{\partial \log}{\partial c} f_{ij}(y_{ij};\beta, c|\alpha_i)\right|_{c=0} \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left\{ (\mu_{ij}^2 - 2\mu_{ij}y_{ij})/2 + \sum_{l=0}^{y_{ij}-1} l \right\} \end{aligned}$$

$$\begin{aligned} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left\{ \frac{y_{ij}(y_{ij}-1)}{2} + (\mu_{ij}^2 - 2\mu_{ij}y_{ij})/2 \right\} \\ &= \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \{(y_{ij}-\mu_{ij})^2 - y_{ij}\} \\ &= \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \{(y_{ij}-e^{X_{ij}^T\beta+\alpha_i})^2 - y_{ij}\}. \end{aligned} \tag{4.10}$$

So, the score test statistic for testing $H_0 : c = 0$ is $T_c = S_c^2/I$, where

$$\begin{aligned} I &= I_{cc} - I_{c\beta} I_{\beta\beta}^{-1} I_{c\beta}^T, \\ I_{cc} &= E\left[\left. \frac{\partial l}{\partial c} \right|_{c=0} \right]^2, \end{aligned}$$

$$I_{c\beta} = E\left(\left. \frac{-\partial^2 l}{\partial c \partial \beta} \right|_{c=0} \right) \quad \text{is a } 1 \times p \text{ vector}$$

and

$$I_{\beta\beta} = E\left(\left. \frac{-\partial^2 l}{\partial \beta_s \partial \beta_r} \right|_{c=0} \right) \quad \text{is a } p \times p \text{ matrix.}$$

Now

$$\begin{aligned} E\left(\frac{\partial}{\partial c} \sum_{j=1}^{n_i} \log f_{ij} \right)^2 &= \frac{1}{4} E\left[\sum_{j=1}^{n_i} (y_{ij} - \mu_{ij})^2 - \sum_{j=1}^{n_i} y_{ij} \right]^2 \\ &= \frac{1}{4} E\left[\left\{ \sum_{j=1}^{n_i} (y_{ij} - \mu_{ij})^2 \right\}^2 + \left(\sum_{j=1}^{n_i} y_{ij} \right)^2 - 2 \sum_{j=1}^{n_i} (y_{ij} - \mu_{ij})^2 y_{ij} \right] \\ &= \frac{1}{4} E\left[\sum_{j=1}^{n_i} (y_{ij} - \mu_{ij})^4 + \sum_{j=1}^{n_i} \sum_{j \neq j'}^{n_i} (y_{ij} - \mu_{ij})^2 (y_{ij'} - \mu_{ij'})^2 + \sum_{j=1}^{n_i} y_{ij}^2 \right. \\ &+ \left. \sum_{j=1}^{n_i} \sum_{j \neq j'}^{n_i} y_{ij} y_{ij'} - 2 \sum_{j=1}^{n_i} (y_{ij} - \mu_{ij})^2 y_{ij} \right], \end{aligned}$$

$$E\left(\frac{\partial^2}{\partial c \partial \beta} \sum_{j=1}^{n_i} \log f_{ij} \right) = E \sum_{j=1}^{n_i} (y_{ij} - \mu_{ij})(-\mu_{ij} X_{ij}^T),$$

and

$$E\left(\frac{\partial^2}{\partial \beta_s \partial \beta_r} \sum_{j=1}^{n_i} \log f_{ij} \right) = -E \sum_{j=1}^{n_i} \mu_{ij} X_{ij} X_{ij}^T$$

After detailed calculations using the first four moments of the Poisson distribution in the above expressions it can be shown that

$$I_{cc} = \frac{1}{2}\sum_{i=1}^{k}\left(\sum_{j=1}^{n_i}\mu_{ij}^2 + \sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\mu_{ij}\mu_{ij'}\right),$$

$$I_{c\beta} = 0$$

and

$$I_{\beta\beta} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\mu_{ij}X_{ij}X_{ij}^T.$$

Then, the score test statistic T_c can be written as $T_c = \hat{S}_c^2/\hat{I}_{cc}$, where

$$\hat{I}_{cc} = \frac{1}{2}\sum_{i=1}^{k}\left(\sum_{j=1}^{n_i}\hat{\mu}_{ij}^2 + \sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\hat{\mu}_{ij}\hat{\mu}_{ij'}\right), \ \hat{\mu}_{ij} = e^{X_{ij}^T\hat{\beta}+\alpha_i}$$

and $\hat{\beta}$ is the maximum likelihood estimator of β under the null hypothesis. Asymptotically, as $k \to \infty$, T_c has a chi-square distribution with one degree of freedom.

Note that the above results are based on α_i being known. However, since α_i's are random effects these should have been integrated out of (4.7). As indicated earlier such integration is difficult to carry out. So we replace these by their estimates. One way of obtaining estimates of the random effects is through using an empirical Bayes procedure (see Collet, 2003). The maximum likelihood estimates of β and the empirical Bayes estimates of α_i under the null hypothesis are given in what follows.

4.2.1 Estimation of the parameter β under the null hypothesis

Note that the mixed effects model (4.6) can be written as

$$\log(\mu_{ij}) = X_{ij}^T\beta + \sqrt{D}v_i, \tag{4.11}$$

where v_i has a standard normal distribution. Now define $\eta_{ij} = X_{ij}^T\beta$ for the linear component of the model obtained from the fixed effects, then (4.11) becomes $\log(\mu_{ij}) =$

$\eta_{ij} + \sqrt{D}v_i$. The kernel of the likelihood for β, D and $v_i,\ i = 1, 2, \ldots, k$, for Poisson data is given by

$$
\begin{aligned}
L &= L(\beta, D, v_1, v_2, \ldots, v_k) \\
&= \prod_{i=1}^{k}\prod_{j=1}^{n_i} \left[\exp\{Y_{ij}\log(\mu_{ij}) - \mu_{ij}\}\right], \\
&= \prod_{i=1}^{k}\prod_{j=1}^{n_i} \left[\exp\{Y_{ij}(\eta_{ij} + \sqrt{D}v_i) - \exp(\eta_{ij} + \sqrt{D}v_i)\}\right]. \qquad (4.12)
\end{aligned}
$$

Further, since v_i is a random variable it needs to be integrated out. Then the likelihood function for β and D can be written as

$$
L(\beta, D) = \prod_{i=1}^{k}\int_{-\infty}^{\infty}\prod_{j=1}^{n_i}\left[\exp\{Y_{ij}(\eta_{ij} + \sqrt{D}v_i) - \exp(\eta_{ij} + \sqrt{D}v_i)\}\right]\frac{\exp(-v_i^2/2)}{\sqrt{2\pi}}dv_i. \qquad (4.13)
$$

The likelihood function (4.13) has $(p+1)$ unknown parameters $\beta_1, \ldots, \beta_p$ and D. Maximum likelihood estimates of the parameters β and D are obtained by maximizing (4.13). The integration in (4.13) is difficult to carry out. However, this can be evaluated approximately by using Gauss-Hermite formula for numerical integration. Therefore, the marginal likelihood function (4.13) becomes

$$
\pi^{-N/2}\prod_{i=1}^{k}\prod_{j=1}^{n_i}\sum_{r=1}^{m} w_r \left[\exp\{Y_{ij}(\eta_{ij} + \sqrt{D}s_r\sqrt{2}) - \exp(\eta_{ij} + \sqrt{D}s_r\sqrt{2})\}\right], \qquad (4.14)
$$

where $w_1, w_2, \ldots, w_m$ are the weights with

$$
w_r = \frac{2^{m-1}m!\sqrt{\pi}}{m^2[H_{m-1}(s_r)]^2},
$$

m is the number of quadrature points and $s_1, s_2, \ldots, s_m$ are the roots of the Hermite polynomial $H_m(s)$ given by

$$
H_m(s) = (-1)^m e^{s^2/2}\frac{d^m}{ds^m}e^{-s^2/2}.
$$

The evaluation points s_r (abscissas) and weights w_r are given in Table 25.10 of Abramowitz and Stegun (1972).

The values $\hat{\beta}$ and $\hat{D}$, which maximizes (4.14), or its logarithm, can then be determined numerically. The computer package SAS procedure GLIMMIX or R function glmmML can be used to evaluate equation (4.14).

4.2.2 Estimation of the random effects α_i

From equation (4.13) the joint posterior density of $v_1, \ldots, v_k$, given $\hat{\beta}$ and $\hat{D}$, the maximum likelihood estimates of β and D obtained in Section 4.2.1, is proportional to

$$\prod_{i=1}^{k}\prod_{j=1}^{n_i}\left[\exp\{Y_{ij}(\hat{\eta}_{ij}+\sqrt{\hat{D}}v_i)-\exp(\hat{\eta}_{ij}+\sqrt{\hat{D}}v_i)\}\right]\frac{\exp(-v_i^2/2)}{\sqrt{2\pi}}, \tag{4.15}$$

where $\hat{\eta}_{ij} = X_{ij}^T\hat{\beta}$.

Now, the log of the ith term of (4.15) is given by

$$l_i(\hat{\beta},\hat{D},v_i) = \text{Constant} + \sum_{j=1}^{n_i}\left[Y_{ij}(\hat{\eta}_{ij}+\sqrt{\hat{D}}v_i)-\exp(\hat{\eta}_{ij}+\sqrt{\hat{D}}v_i)\right]-\frac{v_i^2}{2}. \tag{4.16}$$

The empirical Bayes estimate $\hat{v}_i$ of v_i is obtained by solving $\frac{\partial l_i(\hat{\beta},\hat{D},v_i)}{\partial v_i} = 0$. This is equivalent to obtaining $\hat{v}_i$ by solving

$$\sqrt{\hat{D}}\sum_{j=1}^{n_i}\exp(\hat{\eta}_{ij}+\sqrt{\hat{D}}v_i)+\hat{v}_i = \sqrt{\hat{D}}\sum_{j=1}^{n_i}Y_{ij}.$$

This non-linear equation is to be solved by using a numerical method. The empirical Bayes estimate of α_i then is $\hat{\alpha}_i = \sqrt{\hat{D}}\hat{v}_i$.

It is interesting to note that we obtain the same test statistic as given above for testing $c = 0$ obtained by using the double extended quasi-likelihood given in (3.14).

4.3 Simulation study

In this section we conduct a simulation study to compare level properties of the test statistics T and T_c. Simulations for power properties of these statistics will be conducted in a future study. To estimate the type I error rate of the test statistic T, samples have been generated from the Poisson log-linear model with

$$\log(\mu_{ij}) = 0.8x_{1ij} + 0.5x_{2i} - 0.5, \tag{4.17}$$

for $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_i$. The variable x_1 is subject-specific and x_2 is group-specific and is simulated from the standard normal distribution. Similarly, to estimate the type I error of the test statistic T_c we obtain samples from the Poisson log-linear model with

$$\log(\mu_{ij}) = -0.5 + 0.8x_{1ij} + 0.5x_{2i} + \alpha_i, \tag{4.18}$$

where α_i's are normal with mean zero and variance D, and x_1 and x_2 are the same as above.

Two sets of data are simulated from each model assuming homogeneous and heterogeneous inner group sizes ($= n_i$) with different number of groups/subjects (k). The samples comprised $k = 10, 20, 50, 100$ groups/subjects with $n_i = 5$ observations in the homogeneous group and n_i distributed uniformly between 5 and 20 in the heterogeneous group. The value of the variance D of the random effects considered was $D = 1$, although from our experience other values of D produce similar empirical type I error rates which is expected. Each simulation experiment was based on 10,000 replicated samples. Results of the estimated type I error of the two tests are given in Table 4.1. The results in Table 4.1 show better level performance of the statistic T than its counterpart T_c.

Table 4.1: Estimated Type I error of the tests T and T_c when data are generated from the Poisson distribution. Levels considered are 10%, 5% and 1%

k	α	$n_i = 5$		$5 \leq n_i \leq 20$	
		T	T_c	T	T_c
10	0.10	3.5	4.1	3.3	3.4
	0.05	1.8	3.8	1.9	3.1
	0.01	0.7	3.3	0.6	2.5
20	0.10	4.9	6.2	4.4	10.2
	0.05	2.7	4.5	2.5	10.0
	0.01	0.7	2.9	0.8	9.9
50	0.10	6.3	17.6	6.0	0.6
	0.05	3.1	17.5	3.4	0.6
	0.01	0.7	17.5	1.0	0.6
100	0.10	6.7	3.2	6.9	4.9
	0.05	3.7	3.1	3.8	4.1
	0.01	0.8	3.1	1.0	3.3

Chapter 5

Summary and Conclusions

We derived four score tests for testing homogeneity between and within individuals for clustered (longitudinal) count data with over-dispersed. Two of these tests, namely, H_S and H_{SC} are based on the over-dispersed generalized linear mixed effects model. The remaining two tests, H_{NB} and H_{QL} are based on specific over-dispersion models, namely the negative binomial mixed effects model and the double extended quasi-likelihood mixed effects model. We also developed two score tests for testing over-dispersion in clustered count data. One of these was developed using a generalized over-dispersed mixed effects model given by Cox (1983) and the other was developed using the negative binomial mixed effects model.

The statistics H_S and H_{SC}, in general, show highly inflated level properties, except when the number of groups k and the over-dispersion parameter c are both small, in which case the level property of all four statistics are similar. The statistics H_{NB} and H_{QL} show some conservative level properties, however, as the values of c and k increase, empirical levels become closer to the nominal. The power of the statistics H_S and H_{SC} are, in general, larger than those of H_{NB} and H_{QL} which is expected. What is interesting is that as D increases ($D > .1$), the power of all of the statistics become almost identical, although empirical levels of the statistics H_{NB} and H_{QL} are close to the nominal and those of H_S and H_{SC} are highly inflated. The power of both the statistics H_{NB} and H_{QL} are similar in all the cases studied. The statistic H_{NB} is simpler to calculate.

The level and power properties of all the statistics, in general, remain similar irrespective of which mechanism of over-dispersion is used to generate count data. This also seems to be true irrespective of whether the over-dispersion parameter c is varying or constant.

For testing homogeneity between and within individuals for clustered (longitudinal) count data with over-dispersion, our recommendation, then, is to use either H_{NB} or H_{QL}, although computationally H_{NB} is easier, so it might be preferable in that sense. For testing the presence of over-dispersion in clustered count data we prefer the statistic T as it is simpler to calculate and it holds level, in general, closer to the nominal level than the statistic T_c.

Chapter 6

Future Study: Maximum Likelihood and Bayesian Estimation

In Chapter 3 we developed score tests for homogeneity between and within groups for over-dispersed count data, presented simulation results and showed some examples. In the examples, we saw that the null hypotheses of homogeneity have been rejected which may be the case in most practical situations. Therefore, estimation of the parameters under the alternative hypothesis, that is, with heterogeneous data, is needed. In Chapter 3 and Chapter 4 we dealt with the generalized linear mixed effects model and the negative binomial mixed effects model and used maximum likelihood estimates and empirical Bayes estimates of the parameters under certain null hypotheses.

The purpose of this chapter is to provide an outline for future study of estimation of the parameters of the generalized linear mixed effects model with over-dispersion. We consider maximum likelihood and Bayesian procedures for the estimation of the regression parameters of the generalized linear mixed effects model for count data with over-dispersion and the negative binomial mixed effects model.

6.1 Estimation of the parameters of the generalized linear mixed effects model

6.1.1 Maximum likelihood estimation

Let Y_{ij} denote the j^{th} response in group i, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$, from an over-dispersed exponential family distribution

$$f(y_{ij}; \theta_{ij}, \phi) = \exp\{\phi^{-1}[y_{ij}\theta_{ij} - b(\theta_{ij})] + C(y_{ij}, \phi)\}, \tag{6.1}$$

with mean $\mu_{ij} = E(Y_{ij}|\alpha_i) = b'(\theta_{ij})$, variance $\sigma_{ij}^2 = \text{var}(Y_{ij}|\alpha_i) = \phi\ b''(\theta_{ij})$, where $'$ denotes differentiation with respect to θ and ϕ is the over-dispersion parameter. The mixed-effects model considered in Chapter 3 is

$$g(\mu_{ij}) = \theta_{ij} = \eta_{ij} = \mathbf{X}_{ij}^T\beta + Z_{ij}^T\alpha_i, \tag{6.2}$$

where $\mathbf{X}_{ij}$ is a $p \times 1$ vector of explanatory variables, β is a $p \times 1$ vector of fixed effects, Z_{ij} is a $q \times 1$ vector of covariates and α_i is a $q \times 1$ vector of random cluster/subject effects. We further assume that the random effects α_i have a multivariate normal distribution with a zero mean vector and $q \times q$ covariance matrix $\mathbf{D}$.

Note that in this dissertation we are interested in the random intercept model only. So the rest of this chapter will deal with the situation in which α_i is a scalar random effect and $Z_{ij} = 1$ for all i and j. Extension to the general case in which α_i is a vector is straightforward.

We assume that the random effect α_i has a normal distribution with mean zero and variance D. Then the likelihood function for the parameters β, D and ϕ is given by

$$
\begin{aligned}
L(\beta, D, \phi; Y) &\propto \prod_{i=1}^{k} \int \prod_{j=1}^{n_i} f(Y_{ij}|\alpha_i) \times D^{-1/2} \exp\left(-\frac{\alpha_i^2}{2D}\right) d\alpha_i \\
&\propto \prod_{i=1}^{k} \int \prod_{j=1}^{n_i} \exp\{\phi^{-1}[Y_{ij}\theta_{ij} - b(\theta_{ij})]\} D^{-1/2} \exp\left(-\frac{\alpha_i^2}{2D}\right) d\alpha_i.
\end{aligned}
\tag{6.3}
$$

The integral above is analytically intractable except in the case of the normal linear model. Maximum likelihood estimates $\hat{\beta}, \hat{D}$ and $\hat{\phi}$ of the parameters β, D and ϕ can be obtained by maximizing

$$
\pi^{-N/2} \prod_{i=1}^{k} \prod_{j=1}^{n_i} \sum_{r=1}^{m} w_r \exp\left[\phi^{-1}\left\{Y_{ij}(X_{ij}^T\beta + D^{1/2}s_r\sqrt{2}) - b(X_{ij}^T\beta + D^{1/2}s_r\sqrt{2})\right\}\right], \tag{6.4}
$$

numerically, where $w_1, w_2, \ldots, w_m$ are the weights with

$$
w_r = \frac{2^{m-1}m!\sqrt{\pi}}{m^2[H_{m-1}(s_r)]^2},
$$

m is the number of quadrature points and $s_1, s_2, \ldots, s_m$ are the roots of the Hermite polynomial

$$
H_m(s) = (-1)^m e^{s^2/2} \frac{d^m}{ds^m} e^{-s^2/2}.
$$

The evaluation points s_r (abscissas) and the weights w_r are given in Table 25.10 of Abramowitz and Stegun (1972). The computer package SAS procedure GLIMMIX or R function glmmML can be used to evaluate (6.4).

By modifying expression (6.4) for count data with $\theta_{ij} = \log \mu_{ij} = X_{ij}^T\beta + \alpha_i$, the maximum likelihood estimates $\hat{\beta}, \hat{D}$ and $\hat{\phi}$ of the parameters β, D and ϕ are obtained by maximizing

$$\pi^{-N/2}\prod_{i=1}^{k}\prod_{j=1}^{n_i}\sum_{r=1}^{m} w_r \exp\left[\phi^{-1}\left\{Y_{ij}(X_{ij}^T\beta + D^{1/2}s_r\sqrt{2}) - \exp(X_{ij}^T\beta + D^{1/2}s_r\sqrt{2})\right\}\right], \quad (6.5)$$

numerically. The definitions of w_r, s_r and $H_m(s)$ are the same as above.

For some data sets, maximum likelihood estimation of the parameters using the above procedure poses difficulty (convergence). However, any of the Markov chain Monte carlo (MCMC) methods, for example, the Gibbs sampling approach to the GLM requires only a minor extension to accommodate the introduction of the random effects (see Clayton, 1996). In this case, the amount of computation depends only linearly upon the total number of parameters and a Bayesian formulation of the model is required which is given below.

6.1.2 Bayesian estimation

Specification of the prior distributions

To complete the Bayesian formulation of the model (6.2) with scalar α_i, we need to specify prior distributions for β, D and ϕ. Without having any particular prior information, the usual choice of a prior distribution for β is to take a noninformative diffuse prior on β that is uniform and independent of D. However, this non-informative prior may produce an improper posterior distribution which is undesirable. For a full discussion see Gelfand and Ghosh (2000). It is a usual practice in Bayesian analysis to take a normal prior for β. So, following Gelfand and Ghosh (2000) we assume a prior distribution for β as $N(\beta_0, \Sigma_\beta)$, where β_0 is a $p \times 1$ vector and Σ_β is a $p \times p$ variance-covariance matrix of β. The hyper-parameters β_0 and Σ_β are assumed to be known from prior knowledge.

We now need to set up prior distributions for the two variance parameters ϕ and D. First we discuss prior specification for the parameter ϕ. For this one may choose a prior as $p(\phi) \propto \phi^{-1}$. However, this can make the resulting posterior distribution improper.

So, we define $\psi = \log(\phi)$. Then by using the Jeffreys' invariance principle (see Gelman, Carlin, Stern and Rubin, 2004) for a noninformative prior we obtain the prior density for ψ as

$$p(\psi) = p(\phi) \; \left| \frac{d\phi}{d\psi} \right| \propto \frac{1}{\phi} \times \phi = 1.$$

That is, we choose a uniform prior for ψ on $[0, 1]$. An alternative is to take a conjugate prior for ϕ (Natarajan and Kass, 2000) as

$$p(\phi) = \frac{\nu_c^{\nu_s}}{\Gamma(\nu_s)} \phi^{-(\nu_s+1)} e^{-\nu_c/\phi}.$$

However, the former specification is much simpler so we adopt that here. Similarly, we define $\delta = \log(D)$ and take a uniform prior distribution for δ on $[0, 1]$.

Note that as we discussed earlier we can take an inverse gamma prior for each of the parameters ϕ and D with known parameters. This will make the posterior distribution complicated. In a future study we will deal with this situation.

In the situation in which β_0 and Σ_β are not known, one can take a uniform prior distribution for each component of β_0 and an inverse Wishart distribution

$$p(\Sigma_\beta) \propto \exp\left[-\frac{1}{2} tr(S\Sigma_\beta^{-1}) \right] |\Sigma_\beta|^{-\omega/2}$$

as a prior distribution for Σ_β, where S is a known $p \times p$ scale matrix and ω is a known parameter representing degrees of freedom of the Wishart distribution.

Bayesian formulation

With the prior distributions of β, δ and ψ specified above and β_0 and Σ_β assumed to be known the joint posterior distribution of β, δ and ψ is

$$
\begin{aligned}
f(\beta,\delta,\psi|Y) &= \frac{\prod_{i=1}^{k}\int\prod_{j=1}^{n_i} f(Y_{ij}|\alpha_i,\beta,\psi)p(\beta|\Sigma_\beta)p(\alpha_i|\delta)p(\delta)p(\psi)\ d\alpha_i}{\int\prod_{i=1}^{k}\int\prod_{j=1}^{n_i} f(Y_{ij}|\alpha_i,\beta,\psi)p(\beta|\Sigma_\beta)p(\alpha_i|\delta)p(\delta)p(\psi)\ d\alpha_i\ d\beta\ d\delta\ d\psi} \\
&= C\times|\Sigma_\beta|^{-p/2}\exp[-\frac{1}{2}(\beta-\beta_0)^T\Sigma_\beta^{-1}(\beta-\beta_0)] \\
&\times \prod_{i=1}^{k}\prod_{j=1}^{n_i}\sum_{r=1}^{m} w_r \exp\left[e^{-\psi}\left\{Y_{ij}(X_{ij}^T\beta+e^{\delta/2}s_r\sqrt{2})-b(X_{ij}^T\beta+e^{\delta/2}s_r\sqrt{2})\right\}\right],
\end{aligned}
\tag{6.6}
$$

where w_r and s_r are as specified in Section 6.1.1 and

$$
\begin{aligned}
C &= \int|\Sigma_\beta|^{-p/2}\exp[-\frac{1}{2}(\beta-\beta_0)^T\Sigma_\beta^{-1}(\beta-\beta_0)] \\
&\times \prod_{i=1}^{k}\prod_{j=1}^{n_i}\sum_{r=1}^{m} w_r \exp\left[e^{-\psi}\left\{Y_{ij}(X_{ij}^T\beta+e^{\delta/2}s_r\sqrt{2})-b(X_{ij}^T\beta+e^{\delta/2}s_r\sqrt{2})\right\}\right] d\beta\ d\delta\ d\psi
\end{aligned}
$$

Note that the integration is over all the parameters $\beta_1, \beta_2, \ldots, \beta_p, \delta$ and ψ. For count data the above expressions can be modified by replacing $b(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2})$ by $\exp(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2})$.

Now our interest is to obtain the marginal posterior distributions for β, δ and ψ which are given below.

$$
\begin{aligned}
f(\beta|Y) &\propto |\Sigma_\beta|^{-p/2}\exp[-\frac{1}{2}(\beta-\beta_0)^T\Sigma_\beta^{-1}(\beta-\beta_0)] \\
&\times \int\prod_{i=1}^{k}\prod_{j=1}^{n_i}\sum_{r=1}^{m} w_r \exp\left[e^{-\psi}\left\{Y_{ij}(X_{ij}^T\beta+e^{\delta/2}s_r\sqrt{2})-\exp(X_{ij}^T\beta+e^{\delta/2}s_r\sqrt{2})\right\}\right] d\psi\ d\delta,
\end{aligned}
\tag{6.7}
$$

$$
\begin{aligned}
f(\delta|Y) \quad \propto \quad & \int |\Sigma_\beta|^{-p/2} \exp[-\frac{1}{2}(\beta-\beta_0)^T \Sigma_\beta^{-1}(\beta-\beta_0)] \\
\times \quad & \prod_{i=1}^{k}\prod_{j=1}^{n_i}\sum_{r=1}^{m} w_r \exp\left[e^{-\psi}\left\{Y_{ij}(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2}) - \exp(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2})\right\}\right] d\beta d\psi
\end{aligned}
\tag{6.8}
$$

and

$$
\begin{aligned}
f(\psi|Y) \quad \propto \quad & \int |\Sigma_\beta|^{-p/2} \exp[-\frac{1}{2}(\beta-\beta_0)^T \Sigma_\beta^{-1}(\beta-\beta_0)] \\
\times \quad & \prod_{i=1}^{k}\prod_{j=1}^{n_i}\sum_{r=1}^{m} w_r \exp\left[e^{-\psi}\left\{Y_{ij}(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2}) - \exp(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2})\right\}\right] d\beta d\delta.
\end{aligned}
\tag{6.9}
$$

Posterior means for β, δ and ψ can be obtained from (6.7), (6.8) and (6.9) respectively. However, evaluation of the integrals involved are computationally difficult. To overcome this a Markov chain Monte carlo (MCMC) method such as the Gibbs sampler can be used which we describe below.

Gibbs sampler

The Gibbs sampler is a MCMC method for estimating the marginal posterior distributions. It is a special case of the single component Metropolis-Hastings algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller, 1953) and consists entirely in sampling from full conditional distributions. Let us consider three variables X, Y and Z. Suppose that the conditional distribution of each variable, given the others, has a simple form and the joint distribution is more complicated. We denote the conditional distributions by $[X|Y,Z], [Y|X,Z]$ and $[Z|X,Y]$ and the joint distribution by $[X,Y,Z]$. The joint distribution must be positive over its entire domain. Then by using the method of Gibbs sampler we can generate random values from $[X,Y,Z]$. We review the method here from Zeger and Karim (1991).

With the arbitrary starting values $X^{(0)}, Y^{(0)}, Z^{(0)}$, we draw $X^{(1)}$ from $[X|Y^{(0)}, Z^{(0)}]$, then we draw $Y^{(1)}$ from $[Y|X^{(1)}, Z^{(0)}]$, and finally, complete the first iteration by drawing $Z^{(1)}$ from $[Z|X^{(1)}, Y^{(1)}]$. After a large number of M iterations, we obtain $(X^{(M)}, Y^{(M)}, Z^{(M)})$. This process is repeated a large number of times, say N, obtaining N values from the marginal distributions of X, Y and Z. Geman and Geman (1984) show that under mild conditions, the joint distribution of $(X^{(M)}, Y^{(M)}, Z^{(M)})$ converges at an exponential rate to $[X, Y, Z]$ as $M \to \infty$. The mean of the marginal posterior distribution of, for example, X is the arithmetic mean of the N values obtained above.

Now to use the Gibbs sampler we need to find the conditional $f(\beta|\delta, \psi, Y)$, $f(\delta|\beta, \psi, Y)$ and $f(\psi|\beta, \delta, Y)$. These are obtained as given below.

$$\begin{aligned} f(\beta|\delta, \psi, Y) &\propto |\Sigma_\beta|^{-p/2} \exp[-\frac{1}{2}(\beta - \beta_0)^T \Sigma_\beta^{-1}(\beta - \beta_0)] \\ &\times \prod_{i=1}^{k} \prod_{j=1}^{n_i} \sum_{r=1}^{m} w_r \exp\left[e^{-\psi}\left\{Y_{ij}(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2}) - b(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2})\right\}\right]. \end{aligned}$$

$$f(\delta|\beta, \psi, Y) \propto \prod_{i=1}^{k} \prod_{j=1}^{n_i} \sum_{r=1}^{m} w_r \exp\left[e^{-\psi}\left\{Y_{ij}(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2}) - b(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2})\right\}\right].$$

$$f(\psi|\beta, \delta, Y) \propto \prod_{i=1}^{k} \prod_{j=1}^{n_i} \sum_{r=1}^{m} w_r \exp\left[e^{-\psi}\left\{Y_{ij}(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2}) - b(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2})\right\}\right].$$

Assuming an over-dispersed Poisson log-linear random intercept model $\log(\mu_{ij}) = X_{ij}^T\beta + \alpha_i$, the above conditional distributions become

$$\begin{aligned} f(\beta|\delta, \psi, Y) &\propto |\Sigma_\beta|^{-p/2} \exp[-\frac{1}{2}(\beta - \beta_0)^T \Sigma_\beta^{-1}(\beta - \beta_0)] \\ &\times \prod_{i=1}^{k} \prod_{j=1}^{n_i} \sum_{r=1}^{m} w_r \exp\left[e^{-\psi}\left\{Y_{ij}(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2}) - \exp(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2})\right\}\right]. \end{aligned}$$

$$f(\delta|\beta,\psi,Y) \propto \prod_{i=1}^{k}\prod_{j=1}^{n_i}\sum_{r=1}^{m} w_r \exp\left[e^{-\psi}\left\{Y_{ij}(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2}) - \exp(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2})\right\}\right].$$

$$f(\psi|\beta,\delta,Y) \propto \prod_{i=1}^{k}\prod_{j=1}^{n_i}\sum_{r=1}^{m} w_r \exp\left[e^{-\psi}\left\{Y_{ij}(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2}) - \exp(X_{ij}^T\beta + e^{\delta/2}s_r\sqrt{2})\right\}\right].$$

6.2 Estimation of the parameters of the negative binomial mixed effects model

6.2.1 Maximum likelihood estimation

Let Y_{ij} be the response variable for the jth observation in group i, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$, from the negative binomial distribution, denoted by $NB(\mu_{ij}, c_i)$ and given by

$$f(y_{ij};\mu_{ij},c_i) = \frac{\Gamma(y_{ij}+c_i^{-1})}{y_{ij}!\ \Gamma c_i^{-1}}\left(\frac{1}{1+c_i\mu_{ij}(\mathbf{x})}\right)^{c_i^{-1}}\left(\frac{c_i\mu_{ij}(\mathbf{x})}{1+c_i\mu_{ij}(\mathbf{x}))}\right)^{y_{ij}}, \qquad (6.10)$$

where $\log(\mu_{ij}) = X_{ij}^T\beta + \alpha_i$ is the random intercept model for the mean response.

The likelihood function for the parameters β, c and D, assuming common overdispersion parameter c over all groups or individuals, is given by

$$\begin{aligned} L(\beta,c,D;Y) &= \prod_{i=1}^{k}\int\prod_{j=1}^{n_i}\frac{\Gamma(y_{ij}+c^{-1})}{\Gamma(c^{-1})}\left(\frac{1}{1+c\mu_{ij}}\right)^{c^{-1}}\left(\frac{c\mu_{ij}}{1+c\mu_{ij}}\right)^{y_{ij}} \\ &\times \frac{D^{-1/2}}{\sqrt{2\pi}}\exp\left(-\frac{\alpha_i^2}{2D}\right)d\alpha_i \\ &= \prod_{i=1}^{k}\int\prod_{j=1}^{n_i}\frac{\Gamma(y_{ij}+c^{-1})}{\Gamma(c^{-1})}\left(\frac{1}{1+ce^{X_{ij}^T\beta+\sqrt{D}v_i}}\right)^{c^{-1}}\left(\frac{ce^{X_{ij}^T\beta+\sqrt{D}v_i}}{1+ce^{X_{ij}^T\beta+\sqrt{D}v_i}}\right)^{y_{ij}} \\ &\times \frac{\exp(-v_i^2/2)}{\sqrt{2\pi}}. \end{aligned} \qquad (6.11)$$

As before, the above integral is analytically intractable. Therefore, by using the Gauss-Hermite procedure for numerical integration, maximum likelihood estimates $\hat{\beta}, \hat{c}$ and $\hat{D}$

of the parameters β, c and D can be obtained by maximizing

$$\pi^{-N/2}\prod_{i=1}^{k}\prod_{j=1}^{n_i}\sum_{r=1}^{m} w_r \frac{\Gamma(y_{ij}+c^{-1})}{\Gamma(c^{-1})}\left(\frac{1}{1+ce^{X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}\right)^{c^{-1}}\left(\frac{ce^{X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}{1+ce^{X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}\right)^{y_{ij}} \tag{6.12}$$

or its logarithm, where w_r and s_r are as given before.

6.2.2 Bayesian estimation

To obtain the joint posterior distribution of the parameters β, c and D, we need to specify the prior distributions of the parameters. As before we assign a $N(\beta_0, \Sigma_\beta)$ prior for β. The hyper-parameters β_0 and Σ_β are assumed to be known. Moreover, we choose a uniform prior for $\psi = \log(c)$ on $[0, 1]$ and a uniform prior for $\delta = \log(D)$ on $[0, 1]$.

Therefore, the joint posterior distribution of β, ψ and δ become

$$\begin{aligned} f(\beta,\psi,\delta,Y) &\propto |\Sigma_\beta|^{-p/2}\exp[-\frac{1}{2}(\beta-\beta_0)^T\Sigma_\beta^{-1}(\beta-\beta_0)] \\ &\times \prod_{i=1}^{k}\prod_{j=1}^{n_i}\prod_{l=1}^{Y_{ij}-1}(1+le^{-\psi})\sum_{r=1}^{m} w_r\left(\frac{1}{1+e^{\psi+X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}\right)^{e^{-\psi}} \\ &\times \left(\frac{e^{\psi+X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}{1+e^{\psi+X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}\right)^{Y_{ij}}. \end{aligned}$$

The conditionals for β, ψ and δ are as given below.

$$\begin{aligned} f(\beta|\psi,\delta,Y) &\propto |\Sigma_\beta|^{-p/2}\exp[-\frac{1}{2}(\beta-\beta_0)^T\Sigma_\beta^{-1}(\beta-\beta_0)] \\ &\times \prod_{i=1}^{k}\prod_{j=1}^{n_i}\prod_{l=1}^{Y_{ij}-1}(1+le^{-\psi})\sum_{r=1}^{m} w_r\left(\frac{1}{1+e^{\psi+X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}\right)^{e^{-\psi}} \\ &\times \left(\frac{e^{\psi+X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}{1+e^{\psi+X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}\right)^{Y_{ij}}, \end{aligned}$$

$$
\begin{aligned}
f(\psi|\beta,\delta,Y) &\propto \prod_{i=1}^{k}\prod_{j=1}^{n_i}\prod_{l=1}^{Y_{ij}-1}(1+le^{-\psi})\sum_{r=1}^{m} w_r\left(\frac{1}{1+e^{\psi+X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}\right)^{e^{-\psi}} \\
&\times \left(\frac{e^{\psi+X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}{1+e^{\psi+X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}\right)^{Y_{ij}}
\end{aligned}
$$

and

$$
\begin{aligned}
f(\delta|\beta,\psi,Y) &\propto \prod_{i=1}^{k}\prod_{j=1}^{n_i}\prod_{l=1}^{Y_{ij}-1}(1+le^{-\psi})\sum_{r=1}^{m} w_r\left(\frac{1}{1+e^{\psi+X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}\right)^{e^{-\psi}} \\
&\times \left(\frac{e^{\psi+X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}{1+e^{\psi+X_{ij}^T\beta+\sqrt{D}s_r\sqrt{2}}}\right)^{Y_{ij}}.
\end{aligned}
$$

The marginal posterior means of β, ψ and δ are obtained using the Gibbs sampler procedure described in Section 6.1.2.

Work is continuing to compare the efficiency of the maximum likelihood estimates of the regression parameters for the two models considered, namely the generalized linear mixed effects model and the negative binomial mixed effects model with those obtained by Bayesian procedures. Moreover, as mentioned in Chapter 4 the power study of the statistics T and T_c for testing the presence of over-dispersion is ongoing.

Bibliography

[1] Abramowitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions*, New York: Dover Publications, Inc.

[2] Carrasco, J.L. and Jover, L. (2005). Concordance Correlation Coefficient Aplied to Discrete Data. *Statistics in Medicine*, 24, 4021-4034.

[3] Chesher, A. (1984). Testing for Neglected Heterogeneity. *Econometrica*, 52, 865-872.

[4] Clayton, D.G. (1996). Generalized Linear Mixed Models. Chapter 16 in *Markov Chain Monte Carlo in Practice*. Chapman and Hall.

[5] Collet, D. (2003). *Modelling Binary Data*. Second Edition, Chapman and Hall.

[6] Collings, B.J. (1981). The negative binomial distribution: an alternative to the Poisson. *Ph.D. Thesis*, The University of North Carolina at Chapel Hill.

[7] Collings, J., and Margolin, H. (1985). Testing Goodness of Fit for the Poisson Assumption when Observations are not Identically Distributed. *Journal of the American Statistical Association*, 80, 411-418.

[8] Cox, D.R., and Hinkley, D.V. (1974). *Theoretical Statistics*. Chapman and Hall.

[9] Cox, D. R. (1983). Some Remarks on Overdispersion. *Biometrika*, 70, 269-274.

[10] Dean, C.B. (1992). Testing for Over-dispersion in Poisson and Binomial Regression Models. *Journal of the American Statistical Association*, 87, 451-457.

[11] Deng, D. and Paul, S. R. (2005). Score Tests for Zero-Inflation and Over-dispersion in Generalized Linear Models. *Statistica Sinica*, 15, 257-276.

[12] Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2004). *Applied Longitudinal Analysis*. Wiley-Interscience.

[13] Gelfand, A. and Ghosh, M. (2000). Generalized Linear Models: A Bayesian View. Chapter 1 in *Generalized Linear Models: A Bayesian Perspective*. Marcel Dekker.

[14] Gelman, A., Carlin J.B., Stern, H.S. and Rubin, D.B. (2004). *Bayesian Data Analysis*. Chapman and Hall.

[15] Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

[16] Godambe, V.P. and Thompson, M.E. (1989). An Extension of Quasi-likelihood Estimation. *Journal of Statistical Planning and Inference*, 22, 137-152.

[17] Greenberg, E.R., Baron, J.A., Stukel, T.A., Stevens, M.M., Mandel, J.S., Spencer, S.K., Elias, P.M., Lowe, N., Nierenberg, D.W., Bayrd, G., Vance, J.C., Freeman, D.H., Clendenning, W.E., Kwan, T. and the Skin Cancer Prevention Study Group (1990). A Clinical Trial of Beta Carotene to Prevent Basal-cell and Squamous-cell Cancers of the Skin. *New England Journal of Medicine*, 323, 789-795.

[18] Jacqmin-Gadda, H., and Commenges, D. (1995). Tests of Homogeneity for Generalized Linear Models. *Journal of the American Statistical Association*, 90, 1237-1246.

[19] Kass, R.E. and Nataranjan, R. (2006). A Default Conjugate Prior for Variance Components for Generalized Linear Mixed Models. *Bayesian Analysis*, 1, 535-542.

[20] Kendall, M. and Stuart, A. (1977). *The Advanced Theory of Statistics*, Volume 1, 4th edition.

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

[22] Lawless, J.F., (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, 15, 209-225.

[23] Lee, Y., and Nelder, J., (1996). Hierarchical Generalized Linear Models. *Journal of the Royal Statistical Society*, B 58, 619-678.

[24] Lee, Y., and Nelder, J., (2001). Hierarchical generalized linear models: A synthesis of generalized linear models, random-effect models and structure dispersions. *Biometrika*, 88, 987-1006.

[25] Lehman, E.L. (1999). *Elements of Large-Sample Theory*. Springer.

[26] Leppik, I., Dreifuss, F.E., Porter, R., Bowman, T., Santilli, N., Jacobs, M., Crosby, C., Cloyd, J., Stackman, J., Sutula, T.P., Graves, N., Welty, T., Vickery, T., Bundage, R., Gates, J., Gummit, R. and Gutierres, A. (1987). A controlled study of progabide in partial seizure: Methodology and results. *Neurology*, 37, 963-968.

[27] Liang, K.Y., (1987). A Locally Most Powerful Test for Homogeneity with many strata. *Biometrika*, 74, 259-264.

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

[29] Lin, X. (1997). Variance Component Testing in Generalilsed Linear Models with Random Effects. *Biometrika*, 84, 309-326.

[30] McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models* (2nd edition). Chapman and Hall.

[31] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N. Teller, A.H. and Teller, E. (1953). Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21, 1087-1092.

[32] Moore, D.F. (1986). Asymptotic properties of moments estimators for over-dispersed counts and proportions. *Biometrika*, 73, 583-588.

[33] Nataranjan, R. and Kass, R.E. (2000). Reference Bayesian Methods for Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 95, 227-236.

[34] Nelder, J.A. and Pregibon, D. (1987). An Extended Quasi-Likelihood Function. *Biometrika*, 74, 221-232.

[35] Paul, S.R. and Banerjee, T. (1998). Analysis of Two-way Layout of Count Data Involving Multiple Counts in Each Cell. *Journal of the American Statistical Association*, 93, 1419-1429.

[36] Paul, S.R. and Placket, R.L. (1978). Inference sensitivity for Poisson mixtures. *Biometrika*, 65, 591-602.

[37] Paul, S.R. and Saha, K., (2007). The generalized linear model and extensions: a review and some biological and environmental applications. *Environmetrics*, 18, 421-443.

[38] Rao, C.R. (1947). Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Application to Problems of Estimation. *Proceedings of Cambridge Philosophical Society*, 44, 50-57.

[39] Saha, K. (2004). On the Estimation of Dispersion Parameters for Data in the Form of Count and Proportions. *Ph.D. Thesis*, University of Windsor, Windsor, Canada.

[40] Thall, P.F., and Vail, S.C. (1990). Some Covariance Models for Longitudinal Count Data with Over-dispersion. *Biometrics*, 46, 657-671.

[41] Ware, J.H. (1985). Linear Models for the analysis of Longitudinal Studies. *The American Statistician*, 39, 95-101.

[42] Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61, 3, 439.

[43] Zeger, S.L. and Liang, K.Y. (1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, 42, 121-130.

[44] Zeger, S.L. and Karim, M.R. (1991). Generalized Linear Models with Random Effects. *Journal of the American Statistical Association*, 86, 79-86.

[45] Zhu, H. and Zhang, H. (2006). Generalized Score Test of Homogeneity for Mixed Effects Models. *Annals of Statistics*, 34, 1545-1569.

Vita Auctoris

The author was born in 1971 in Feni, Bangladesh. He obtained his higher secondary degree from Chowmuhani College, Noakhali, Bangladesh in 1988. From there he went to the University of Dhaka, Bangladesh where he obtained a B.Sc. and an M.Sc. in Statistics in 1991 and 1992 respectively. He has also earned an M.Sc. degree in Statistics from the University of British Columbia, Vancouver, Canada in 2003. He is currently a candidate for the Doctor of Philosophy in Statistics at the University of Windsor and will graduate in Fall 2009.