STATISTICAL INFERENCE
ON BINOMIAL REGRESSION MODELS
IN THE PRESENCE OF OVER-DISPERSION

by

Wimali Prasangika Lorensu Hewa

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE

School of Mathematics and Statistics
Ottawa-Carleton Institute for Mathematics and Statistics

at

CARLETON UNIVERSITY

Ottawa, Ontario

August 2008

© Copyright by Wimali Prasangika Lorensu Hewa, 2008

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-44131-2

Our file Notre référence

ISBN: 978-0-494-44131-2

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Overdispersion is the presence of greater variability in a data set than one would expect based on a given simple statistical model. When the observed variance is higher than the anticipated variance of an assumed theoretical model, then it may be an indication of overdispersion. When there is such a suspicion, the natural reaction would be to test whether there are valid reasons to assume the same. In many cases, the overdispersion parameter is incorporated to the model with an introduction of random effects. Throughout this thesis, it is assumed that the primary data are coming from a binomial distribution and hence the testing is for extra-binomial variation.

For testing overdispersion, instead of a score test statistic derived from likelihood functions, a quasi-score test statistic is developed based on quasi-likelihood functions. A simulation study is carried out to explore the performance of the quasi-score test for testing overdispersion in binomial regression.

Subsequent to any evidence of an overdispersion in the data set, this parameter together with the regression parameters are estimated using quasi-likelihood methods. The marginal moments used in the quasi-likelihood functions are obtained using two different approaches. Simulations are carried out to study the finite-sample properties of the estimators under correctly specified and misspecified random effects. Thereby, bias, root mean squared error, and coverage probability and mean length of 95% confidence intervals are obtained.

Acknowledgments

I would like to express my sincere gratitude to my supervisor Dr. Sanjoy Sinha for his guidance in making this thesis a reality. The knowledge, support, continuous interest and effort given to me by Dr. Sinha was paramount to the completion of this thesis.

I wish to thank Dr. Avi Singh from Statistics Canada for his helpful comments and suggestions on my thesis.

My deepest thanks to the School of Mathematics and Statistics at Carleton University for providing me the opportunity to continue my studies in Statistics. It has been a pleasure to study at Carleton University under the support and guidance of its knowledgeable faculty and staff members. I specially appreciate Ms. Valerie Daley for her valuable advice and support during my graduate studies.

To my husband Sudarsha Hewa, thank you for your encouragement and support in every possible way from the beginning to completion of this thesis.

For My Parents:

Mr. Wimal Nanayakkara
&
Mrs. Sriyanee Nanayakkara

Table of Contents

Abstract	i
Acknowledgments	ii
Dedication	iii
Table of Contents	iv
List of Tables	vi
Abbreviations	vii
1 Introduction	1
2 Generalized Linear Models	7
2.1 Introduction	7
2.2 Structure of the Model	9
2.3 Likelihood Function	11
3 Modelling Overdispersion	14
3.1 Introduction	14
3.2 Structure of the Model for Overdispersion	16
3.2.1 Conditional Distribution	16
3.3 Aspects of the Marginal Distribution	17
3.3.1 Marginal Mean and Variance	17
3.3.2 Approaches for Approximating Marginal Mean and Variance	18
3.3.2.1 Small Sigma Approximation Method	18
3.3.2.2 Simulation Method	19
4 Testing for Overdispersion	21
4.1 Introduction	21
4.2 Parametric Approach	22
4.2.1 Score Test for τ	22
4.2.2 Testing for Overdispersion in Extra-Binomial Data	24
4.3 Semi-parametric Approach	25
4.3.1 Quasi-Score Test for τ	27
4.3.2 Testing for Overdispersion in Extra-Binomial Data	30
4.4 Misspecified Random Effects	35
4.5 Power of the Tests	36

4.6	Empirical Significance Levels	37
5	Parameter Estimation and Related Topics	38
5.1	Introduction	38
5.2	Quasi-Likelihood Estimating Equations	39
5.3	Parameter Estimation using Newton-Raphson Iterative Method . . .	41
5.4	Variance-Covariance of the Estimators	42
5.5	Confidence Intervals of the Parameters and their Mean Lengths . . .	43
5.6	Coverage Probability	43
5.7	Approximation of Marginal Moments	44
5.7.1	Small Sigma Approximation Method	44
5.7.2	Simulation Method	47
6	Simulation Study	50
6.1	Testing for Overdispersion	51
6.2	Parameter Estimation	53
6.2.1	Bias	53
6.2.2	Root Mean Squared Error	57
6.2.3	Coverage Probability and Mean Length	60
7	Conclusion and Discussion	67
A	Derivation of the Score Test Statistic	70
A.1	General Format	70
A.2	Logit-Normal/Binomial	83
B	Derivatives of the Quasi-Score Functions	89
C	R Programming Code	94
	Bibliography	116

List of Tables

6.1	Empirical level and power of the quasi-score test for both correctly specified and misspecified random effects distributions.	52
6.2	Biases of the estimators when the random effects are correctly specified.	54
6.3	Biases of the estimators when the random effects are misspecified - gamma (0.5, 1).	55
6.4	Biases of the estimators when the random effects are misspecified - gamma (1, 1).	56
6.5	Biases of the estimators when the random effects are misspecified - gamma (2, 1).	56
6.6	Root MSEs of the estimators when the random effects are correctly specified.	57
6.7	Root MSEs of the estimators when the random effects are misspecified - gamma (0.5, 1).	58
6.8	Root MSEs of the estimators when the random effects are misspecified - gamma (1, 1).	59
6.9	Root MSEs of the estimators when the random effects are misspecified - gamma (2, 1).	59
6.10	Coverage probabilities of the parameters when the random effects are correctly specified.	60
6.11	Mean lengths of the confidence intervals when the random effects are correctly specified.	61
6.12	Coverage probabilities of the parameters when the random effects are misspecified - gamma (0.5, 1).	62
6.13	Mean lengths of the confidence intervals when the random effects are misspecified - gamma (0.5, 1).	62
6.14	Coverage probabilities of the parameters when the random effects are misspecified - gamma (1, 1).	63
6.15	Mean lengths of the confidence intervals when the random effects are misspecified - gamma (1, 1).	63
6.16	Coverage probabilities of the parameters when the random effects are misspecified - gamma (2, 1).	64
6.17	Mean lengths of the confidence intervals when the random effects are misspecified - gamma (2, 1).	64

Abbreviations

GEE	Generalized estimating equations
GLMM	Generalized linear mixed models
GLM	Generalized linear models
LM	Linear models
LMM	Linear mixed models
ML	Maximum likelihood
MSE	Mean squared error
NR	Newton-Raphson
QL	Quasi-likelihood
QS	Quasi-score
SSA	Small sigma approximation

Chapter 1

Introduction

Analysis of data via a single-parameter family of distributions implies in particular that the variance is determined by the mean. Familiar examples are Poisson, binomial and exponential distributions. A very common practical complication is the presence of over-dispersion, leading to failure of the variance-mean relation (Cox 1983)[3]. In fact, some would prefer to consider that overdispersion is the norm in practice and nominal dispersion the exception. This is an issue that has been recognized by data analysts for many years. The presence of over-dispersion can arise in many ways. One of the reasons being the clustering in the population. For example, families, households, cities and litters are some of the common instances where clustering of population occurs.

Data which come in the form of counts often known to display greater variability than that can be attributed to the Poisson or binomial sampling variation assumed by

the model. In such a situation, this overdispersion is referred to as extra-poisson or extra-binomial variation respectively (Moore 1986)[15]. For this unexplained residual variation, one can assume a source of extra-poisson or extra-binomial random variation. Crowder (1985, 1978)[5][4] and Williams (1982)[26] discuss the analysis of data in the form of proportions when extra-binomial variation is present. It maybe worthwhile to use a model that allows for a possible extra-binomial variation if one is interested mainly on inference concerning regression parameters and if overdispersion occurs on a routinely basis. Williams (1982)[26] and others have described a simple method of modifying standard logistic-linear analysis, to accommodate extra-binomial variation introducing a continuous independent random variate with finite mean and variance.

One can consider incorporating the overdispersion parameter to the model through a mixed model. In this thesis, the conditional distribution of the observations (given the random effects) is considered to be binomial. Once a larger overdispersed family of models is constructed accommodating overdispersion, the resulting model can be referred to as the “mixed binomial model”. The generalized linear mixed models (GLMMs) are also useful for accommodating the overdispersion often observed among non-normally distributed responses. These GLMMs are derived from the well-known generalized linear models (GLM) (McCullagh and Nelder 1989)[12] by adding random effects to the linear predictor. In many cases, it is assumed that the random effects are normally distributed. One may also consider non-normal distributions for the

random effects. In this thesis, we consider analyzing a cluster as a whole without going into detail of each individual observation within a cluster. Observations within a particular cluster may be correlated but the cluster as a whole is considered to be independent from each other. The conditional observations are assumed to be distributed non-normally (e.g., binomial) and independently. In addition, the random effects are also assumed to be *independently* distributed. The use of mixed models is vast and is expanding so rapidly. These mixed models are useful in analyzing data coming from many different sources. The areas of application include biological and medical research, animal and human genetics, and small area estimation.

When testing for extra-binomial variation, one shall first construct a larger overdispersed family of models and then test for a reduction to the simple model of interest. Several methods have been proposed to test for overdispersion, including the score test. With these models, one may fit the binomial model as a first step in the model building process and then test for overdispersion. Dean (1992)[6] developed a unifying theory for many of the score tests that are currently in use. It is a more generalized version of the currently available overdispersion tests. It is an extended version of the tests for extra-binomial variation developed by Tarone (1979)[24] and Prentice (1986)[17] considering the single sample problem to a general regression setup. These tests suggested by Dean (1992)[6] are claimed to be more powerful than other existing methods based on quasi-likelihood (QL), where only the first two moments are specified. In this thesis, we compare the exact score test with other tests based on

semi-parametric QL methods.

Although there are two sets of parameters, regression and overdispersion, our interest is in testing for overdispersion. In this instance, the regression parameters naturally act as nuisance parameters. When developing the test statistic, we need to adjust for these nuisance parameters. Details on QS tests under nuisance parameters are given by Boos (1992)[1].

A complete maximum likelihood (ML) analysis based on the joint marginal likelihood of the responses can be used to estimate the regression parameters and overdispersion parameter in mixed models. It requires numerical integration techniques for the calculation of the log-likelihood function, score functions, and information matrix. However, the calculation of the numerical integration is tedious when the random effects are multi-dimensional. To avoid such computational difficulty, alternative methods were suggested by many authors. For example, a QL function can be used to estimate the parameters of the fixed and random effects numerically. The idea of QL was developed by Wedderburn (1974)[25] and it greatly widened the scope of GLM by allowing a weaker assumption about the random component which requires only a model for the mean of the data and the relationship between the mean and the variance. Although this method uses minimal information, in many cases it is known to retain full or nearly full efficiency compared to ML. The basic idea behind the QL is to derive a likelihood-like quantity whose construction requires only a few assumptions. As an advantage, with this minimal input, the method is robust to

misspecification of finer details of the model.

When dealing with extra-binomial variation, the model specification is assumed to be conditional on the random effects. Although the form of the conditional distribution is assumed known, no exact distribution is given for the marginal responses. So it is important to derive the aspects of the marginal distribution in order to identify what has been assumed for the observed data. When calculating the marginal mean and variance, one needs to find the probability of successes in the overdispersed model. Dean(1992)[6] used a Taylor series approximation method to derive the marginal mean and variance of the observations. These approximations are restricted to small values of the overdispersion parameter and therefore identified as small sigma approximations (SSA). Another approach is based on simulation, where a sufficiently large number of random effects are generated from the random effects distribution to approximate the marginal moments. Details of these simulation methods are given in Jiang (1998)[7] and Sutradhar and Farrell (2004)[22]. If the probability of successes in the overdispersed model itself is random and distributed independently, then it provides an opportunity to derive the exact mean and variance of the same. Which in turn provide values of the marginal mean and variance of the observations.

In this thesis, we focus on testing for overdispersion under both correctly specified and misspecified random effects distributions, and estimation of regression parameters and the overdispersion parameter in extra-binomial regression. The thesis is organized as follows. In Chapter 2, we briefly review the GLM to get a basic idea about the

structure of the model without considering the overdispersion (or any random effects). Chapter 3 describes the modelling of overdispersion using mixed models. These mixed (binomial) models are used to derive the aspects of the marginal distribution of the observations in order to comprehend what has been assumed for the observed data. Details on different approaches for deriving the marginal mean and variance are also given in this chapter.

Chapter 4 describes both score and quasi-score tests for testing overdispersion. Estimation of regression parameters and overdispersion parameter using QL functions is described in Chapter 5.

In Chapter 6, a set of simulations are carried out to study the empirical powers of the tests for overdispersion and the empirical significance levels under both correctly specified and misspecified random effects. We also study the finite-sample properties of the QL estimators of model parameters in the simulations. Chapter 7 concludes the thesis with some discussion and directions for further research.

Chapter 2

Generalized Linear Models

2.1 Introduction

The general linear regression model for normal data was one of the greatest developments of Statistics in the early 20th century. In this pre-computer age, there was much effort devoted to experimental designs that would not only be statistically efficient, but also be easily analyzed by hand computation. Both Fisher and Yates were pioneers of this work, working at the Rothamsted agricultural research station. The effort of computation was relaxed with the innovation of computers. As a result, statistical research moved on to the extension of these methods to a wider class of statistical problems, in particular, modelling for non-normal data. The use of non-normal data is common in biostatistics in which the observations are often binary, that is, there can only be two possibilities. For example, a patient is diagnosed with

a certain illness or not. Numerical procedures were developed for each and individual situation and computer programs for model fitting were specific to that particular situation. In the meantime, Nelder and Wedderburn (1972)[16] were working towards developing a wide class of models that could all be fitted by a common numerical algorithm, the generalized linear models (GLMs). Even though virtually all the pieces had previously existed, these authors were the first to describe how a collection of seemingly disparate statistical techniques could be unified.

The essence of this generalization has two arguments: one, that observations are not necessarily assumed to be normally distributed; and two, that the mean is not necessarily taken as a linear combination of parameters but some function of the mean is (McCulloch, C. E., and Searle, S. R. 2001[14]). The GLMs represent a class of fixed effects regression models for several types of dependent variables (e.g., continuous, counts, proportions). McCullagh and Nelder (1989)[12] describe these in great detail. An important characteristic of GLMs is that they assume independent (or at least uncorrelated) observations. This assumption of independence is a characteristic of the linear models (LM) of classical regression analysis, and is carried over without any modifications to the wider class of GLMs.

2.2 Structure of the Model

McCulloch (2000)[13] discussed various aspects of GLMs for modelling data. One being the decisions involved in building a GLM.

The decisions are:

1. What is the distribution of the data?
2. What function of the mean will be modeled as linear in the predictors?
3. What will the predictors be?

The observations are assumed to be independent measurements from a distribution with a density from the exponential family or similar to the exponential family:

$$Y_i \sim \text{indep. } f_{Y_i}(y_i).$$

Consider a natural exponential family with the probability density function given in the form:

$$f_{Y_i}(y_i; \theta_i) = \exp\{a_i(\theta_i)y_i - g_i(\theta_i) + c_i(y_i)\}, \quad (2.1)$$

where y_i represents the realization of the response variable and θ_i is an unknown parameter on which the distribution of y_i depends. Also, a_i , g_i and c_i are some known functions of θ_i . The model in (2.1) is in the canonical form when $a_i(\theta_i) = \theta_i$ and then θ_i is the canonical parameter.

We focus on the commonly used binomial regression model which is also referred to as the *logistic linear model*. Let $Y_1, Y_2, \dots, Y_n$ be a sample of n independent obser-

variations from the binomial distribution and let θ_i be a function of a $p \times 1$ covariates $\mathbf{x}_i$ and regression parameters β : $\theta_i = \theta_i(\mathbf{x}_i; \beta)$, $i = 1, 2, \dots, n$.

For the binomial density, we have

$$f_{Y_i}(y_i; m_i, p_i) = \binom{m_i}{y_i} p_i^{y_i} (1 - p_i)^{m_i - y_i}, \quad (2.2)$$

where m_i is the number of trials, p_i is the probability of successes, and $y_i = 0, 1, \dots, m_i$; $0 \leq p_i \leq 1$; $i = 1, 2, \dots, n$.

It is easy to show that the binomial distribution is a member of the exponential family. Equation (2.2) can be rewritten in the following form:

$$\begin{aligned} f_{Y_i}(y_i; m_i, p_i) &= \binom{m_i}{y_i} \left(\frac{p_i}{1 - p_i} \right)^{y_i} (1 - p_i)^{m_i} \\ &= \exp \left\{ y_i \log \left(\frac{p_i}{1 - p_i} \right) + m_i \log(1 - p_i) + \log \binom{m_i}{y_i} \right\}. \end{aligned} \quad (2.3)$$

Assuming $\theta_i = \log \left(\frac{p_i}{1 - p_i} \right)$, (2.3) can be expressed in the form of the exponential family given in (2.1) with

$$a_i(\theta_i) = \theta_i = \log \left(\frac{p_i}{1 - p_i} \right),$$

$$g_i(\theta_i) = m_i \log(1 + e^{\theta_i}), \text{ and}$$

$$c_i(y_i) = \log \binom{m_i}{y_i}.$$

One would typically wish to relate the parameters of a distribution to a set of predictors. It can be achieved by modeling a transformation of the mean of the observations, which would be some function of θ_i , as a LM in the predictors:

$$E(Y_i) = \mu_i = m_i p_i \text{ and } g(\mu_i) = \mathbf{x}_i^t \beta,$$

where $g(\cdot)$ is a known function, which is called the *link function*. In this logistic example, $g(\mu_i) = \log\left(\frac{\mu_i}{m_i - \mu_i}\right)$. The final decision that needs to be taken is about the predictors. In practical situations, one has to decide on what predictors to include in the model and in what form.

2.3 Likelihood Function

Assume that the i^{th} response variable Y_i follows a distribution in the exponential family (2.1). Let $l_i(\theta_i; y_i) = \log f_{Y_i}(y_i; \theta_i)$ be the contribution to the log-likelihood function from the i^{th} observation, so that

$$l_i(\theta_i; y_i) = \log f_{Y_i}(y_i, \theta_i) = a_i(\theta_i)y_i - g_i(\theta_i) + c_i(y_i). \quad (2.4)$$

The mean and variance of Y_i can be derived from the two well-known properties:

$$E\left(\frac{\partial l_i}{\partial \theta_i}\right) = 0, \quad (2.5)$$

$$E\left(\frac{\partial^2 l_i}{\partial \theta_i^2}\right) + E\left(\frac{\partial l_i}{\partial \theta_i}\right)^2 = 0. \quad (2.6)$$

Using equation (2.4), we have

$$\frac{\partial l_i}{\partial \theta_i} = a'_i(\theta_i)y_i - g'_i(\theta_i), \text{ and} \quad (2.7)$$

$$\frac{\partial^2 l_i}{\partial \theta_i^2} = a''_i(\theta_i)y_i - g''_i(\theta_i), \quad (2.8)$$

where prime denotes the differentiation with respect to θ_i . Now using equations (2.5) and (2.7), we have

$$0 = E\left(\frac{\partial l_i}{\partial \theta_i}\right) = E\{a'_i(\theta_i)Y_i - g'_i(\theta_i)\}.$$

Then the mean of Y_i is obtained in the following form:

$$E(Y_i) = \mu_i(\theta_i) = \frac{g'_i(\theta_i)}{a'_i(\theta_i)}. \quad (2.9)$$

Also, from (2.7),

$$\begin{aligned} E\left(\frac{\partial l_i}{\partial \theta_i}\right)^2 &= E\{a'_i(\theta_i)y_i - g'_i(\theta_i)\}^2 \\ &= \{a'_i(\theta_i)\}^2 E\left\{Y_i - \frac{g'_i(\theta_i)}{a'_i(\theta_i)}\right\}^2 \\ &= \{a'_i(\theta_i)\}^2 E\{Y_i - E(Y_i)\}^2 \\ &= \{a'_i(\theta_i)\}^2 V(Y_i), \end{aligned} \quad (2.10)$$

and from (2.8),

$$\begin{aligned} E\left(\frac{\partial^2 l_i}{\partial \theta_i^2}\right) &= E\{a''_i(\theta_i)y_i - g''_i(\theta_i)\} \\ &= a''_i(\theta_i)E(Y_i) - g''_i(\theta_i). \end{aligned} \quad (2.11)$$

Using (2.6), (2.10) and (2.11), we have

$$0 = a''_i(\theta_i)E(Y_i) - g''_i(\theta_i) + \{a'_i(\theta_i)\}^2 V(Y_i).$$

So the variance of Y_i is obtained in the following form:

$$V(Y_i) = \frac{g''_i(\theta_i) - a''_i(\theta_i)E(Y_i)}{\{a'_i(\theta_i)\}^2}. \quad (2.12)$$

It is clear that, the variance is a function of the mean of the observations.

The regression parameter β can be estimated by the iterative reweighted least squares (IRWLS) method as described in McCullagh and Nelder (1989)[12]. Here we use the *R* function “*glm*” for the maximum likelihood (ML) estimates of the regression coefficients in a binomial regression model.

Chapter 3

Modelling Overdispersion

3.1 Introduction

What is overdispersion? Overdispersion is the presence of greater variability (statistical dispersion) in a data set than one would expect from a simple statistical model. When the observed or the empirical variance is higher than the variance of an assumed theoretical model, then it may be an evidence for overdispersion. Typically, this reflects a lack of independence or heterogeneity among individuals within a given cluster. Here the clusters may consist of families, households, cities or litters. This overdispersion is generally caused by positive correlation between the responses within a cluster or by an excess variation between response probabilities or counts. In such a situation, a theoretically defensible alternative model that captures the overdispersion absent from the theoretical model has to be recognized, thereby providing a

better fit to the observed data.

It is necessary to find a suitable method for modelling overdispersion to the data if there is evidence that the data are overdispersed. Modelling of the overdispersion parameter can be performed through an introduction of a random effect to the generalized linear model (GLM). There are several ways of introducing random effects to a GLM, one being the use of a generalized linear mixed model (GLMM). These GLMMs are useful in accommodating overdispersion to the model by incorporating random effects. GLMMs are a natural extension of both linear mixed models (LMMs) and GLMs.

Recall that for a GLM, we consider the form:

$$\boldsymbol{\theta} = \boldsymbol{\theta}(\mathbf{x}; \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}. \quad (3.1)$$

In GLMM, the random effects are incorporated by enlarging the model as

$$\boldsymbol{\theta}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}, \quad (3.2)$$

where $\mathbf{Z}$ is defined as (Sutradhar 2004)[21]:

$$\mathbf{Z} = \sqrt{\tau}\mathbf{U}. \quad (3.3)$$

Here $\mathbf{Z}$ is considered as a random vector with $E(\mathbf{Z}) = \mathbf{0}$ and $V(\mathbf{Z}) = \tau\mathbf{b}(\boldsymbol{\theta}) > 0$, $\mathbf{U}$ is a vector of independent random variates and τ is the overdispersion parameter. $\mathbf{b}(\boldsymbol{\theta})$ depends on the distribution of $\mathbf{U}$.

3.2 Structure of the Model for Overdispersion

3.2.1 Conditional Distribution

To specify the model, we start with the conditional distribution of the observations given the random effect. Suppose conditional on the random effects $Z_1, \dots, Z_n$, the observations $Y_1, \dots, Y_n$ are independent and follow the binomial distribution: $Y_i|Z_i \sim \text{indep. Bin}(m_i, p_i^*)$, where m_i is the number of trials, $p_i^* = p_i^*(z_i)$ is the probability of successes, Z_i are independent continuous random variables with $E(Z_i) = 0$ and $V(Z_i) = \tau b_i(\theta_i) < \infty$.

According to Williams (1982)[26] and others, the extra-binomial variation is of the form:

$$\theta_i^* = \log \left(\frac{p_i^*}{1 - p_i^*} \right) = \mathbf{x}_i^t \boldsymbol{\beta} + z_i = \theta_i + z_i, \quad (3.4)$$

where $\theta_i = \mathbf{x}_i^t \boldsymbol{\beta}$, $\mathbf{x}_i$ is a vector of covariates for the i^{th} observation.

To completely specify the overdispersion model, we consider a distribution for the random effects Z_i :

$$Z_i = \sqrt{\tau} U_i, \quad (3.5)$$

where $U_i \sim \text{indep. N}(0, 1)$. Under this assumption, $E(Z_i) = 0$, $V(Z_i) = \tau$ and $V(U_i) = 1$.

3.3 Aspects of the Marginal Distribution

Although the form of the conditional distribution is known, the marginal distribution of the responses is usually unknown. In some cases, instead of specifying the marginal distribution, we consider finding only the first few marginal moments of the distribution. For example, in quasi-likelihood (QL) approach, we only require the first two marginal moments of the distribution.

3.3.1 Marginal Mean and Variance

For the conditional binomial distribution introduced in section 3.2.1, the marginal mean is:

$$E(Y_i) = E\{E(Y_i|Z_i)\} = E(m_i p_i^*) = m_i E(p_i^*), \quad (3.6)$$

and the marginal variance is:

$$\begin{aligned} V(Y_i) &= E\{V(Y_i|Z_i)\} + V\{E(Y_i|Z_i)\} \\ &= E\{m_i p_i^* (1 - p_i^*)\} + V\{m_i p_i^*\} \\ &= m_i E\{p_i^* (1 - p_i^*)\} + m_i^2 V(p_i^*) \\ &= m_i E(p_i^*) - m_i V(p_i^*) - m_i \{E(p_i^*)\}^2 + m_i^2 V(p_i^*) \\ &= m_i(m_i - 1)V(p_i^*) + m_i E(p_i^*)\{1 - E(p_i^*)\}. \end{aligned} \quad (3.7)$$

Note that here we need to calculate the mean and variance of p_i^* to find the marginal mean and variance of Y_i .

3.3.2 Approaches for Approximating Marginal Mean and Variance

There are a number of ways to derive the marginal mean and variance. In this thesis, two such methods are taken into consideration. The derivation of marginal aspects depends on how the mean and variance of p_i^* are calculated. The details on different approaches are given in the following sections.

3.3.2.1 Small Sigma Approximation Method

The first approach uses the Taylor series approximation to calculate the probability of successes in the overdispersed model. Dean (1992)[6] employed this method when approximating the marginal mean and variance of the observations for small values of the overdispersion parameter.

Using a first order Taylor series expansion about mean $E(Z_i) = 0$, we have

$$p_i^* \approx p_i^* \Big|_{z_i=0} + \frac{dp_i^*}{dz_i} \Big|_{z_i=0} (z_i - 0). \quad (3.8)$$

Here,

$$\begin{aligned} \frac{dp_i^*}{dz_i} \Big|_{z_i=0} &= \frac{d}{dz_i} \left(\frac{\exp(\mathbf{x}_i^t \boldsymbol{\beta} + z_i)}{1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta} + z_i)} \right) \Big|_{z_i=0} \\ &= \frac{\exp(\mathbf{x}_i^t \boldsymbol{\beta} + z_i)}{\{1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta} + z_i)\}^2} \Big|_{z_i=0} = \frac{\exp(\mathbf{x}_i^t \boldsymbol{\beta})}{\{1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta})\}^2} \\ &= p_i(1 - p_i). \end{aligned} \quad (3.9)$$

Hence from (3.8), we have

$$p_i^* \approx p_i + p_i(1 - p_i)z_i. \quad (3.10)$$

So that, $E(p_i^*) \approx p_i$ and $V(p_i^*) \approx p_i^2(1 - p_i)^2\tau$.

Using (3.10), the approximate marginal mean and variance of Y_i can be obtained as,

$$E(Y_i) = m_i E(p_i^*) \approx m_i p_i, \quad (3.11)$$

and

$$V(Y_i) \approx m_i p_i (1 - p_i) [1 + \tau(m_i - 1)p_i(1 - p_i)] \quad \text{for small } \tau. \quad (3.12)$$

3.3.2.2 Simulation Method

As a different approach, a simulation method can be employed to calculate the mean of p_i^* . Simulation is done by generating a sufficiently large number of random effects from its corresponding distribution. Details of these simulation methods are given in Jiang (1998)[7] and Sutradhar and Farrell (2004)[22]. The expectation of p_i^* can be approximated from

$$E(p_i^*) = \int p_i^* f(z_i) dz_i \doteq \frac{1}{S} \sum_{s=1}^S p_i^{*(s)}, \quad (3.13)$$

where

$$p_i^{*(s)} = \frac{\exp(\mathbf{x}_i^t \boldsymbol{\beta} + z_i^{(s)})}{1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta} + z_i^{(s)})} = \frac{\exp(\mathbf{x}_i^t \boldsymbol{\beta} + \sqrt{\tau} u_i^{(s)})}{1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta} + \sqrt{\tau} u_i^{(s)})} \quad (3.14)$$

with $u_i^{(s)}$ being the s^{th} ($s = 1, \dots, S$) realized value of U_i , generated from its corresponding distribution. One should consider a large number for S (here we choose $S = 2000$) in order to obtain good approximations to the marginal moments.

Using (3.6) and (3.13), the marginal mean of the observations can be approximated

from:

$$E(Y_i) = m_i E(p_i^*) \doteq \frac{m_i}{S} \sum_{s=1}^S p_i^{*(s)}. \quad (3.15)$$

Similarly, using (3.7) and simulations, the marginal variance of the observations can be approximated from:

$$\begin{aligned} V(Y_i) &= m_i E\{p_i^*(1 - p_i^*)\} + m_i^2 V(p_i^*) \\ &= m_i \int p_i^*(1 - p_i^*) f(z_i) dz_i + m_i^2 \int \{p_i^* - E(p_i^*)\}^2 f(z_i) dz_i \\ &\doteq \frac{m_i}{S} \sum_{s=1}^S p_i^{*(s)}(1 - p_i^{*(s)}) + m_i^2 \sum_{s=1}^S \frac{\{p_i^{*(s)} - E(p_i^*)\}^2}{S - 1}. \end{aligned} \quad (3.16)$$

In the next chapter, we discuss testing for overdispersion using both score and quasi-score tests.

Chapter 4

Testing for Overdispersion

4.1 Introduction

Recall the extra-binomial regression model (3.4) introduced in Chapter 3. We are interested in testing $H_0 : \tau = 0$ vs $H_A : \tau > 0$. Here we describe both score and quasi-score (QS) tests for testing H_0 . Both tests are developed after adjusting for the nuisance parameters. Details on QS tests under nuisance parameters are given in Boos (1992)[1].

4.2 Parametric Approach

4.2.1 Score Test for τ

Dean (1992)[6] developed a score test for testing overdispersion under the assumption of a small τ . A brief review of this approach is given in this section. The details of derivations are given in Appendix A.

Consider the density function of a larger overdispersed natural exponential family:

$$f_{Y_i}(y_i; m_i, p_i^*) = f(y_i | \theta_i^*) = \exp \{ a_i(\theta_i^*) y_i - g_i(\theta_i^*) + c_i(y_i) \}. \quad (4.1)$$

The likelihood score function (Dean (1992)[6]) for testing the hypothesis, $\tau = 0$ is, $\sum_{i=1}^n \frac{\partial l_i}{\partial \tau} \Big|_{\tau=0, \beta=\hat{\beta}}$, with l_i being the contribution to the log-likelihood function from the i^{th} observation. For the exponential family (4.1), we have

$$\sum_{i=1}^n \frac{\partial l_i}{\partial \tau} \Big|_{\tau=0, \beta=\hat{\beta}} = \sum_{i=1}^n \frac{b_i(\hat{\theta}_i)}{2} [a'_i(\hat{\theta}_i)]^2 \left\{ (y_i - \mu_i(\hat{\theta}_i))^2 - (a'_i(\hat{\theta}_i))^{-2} (g''_i(\hat{\theta}_i) - a''_i(\hat{\theta}_i) y_i) \right\},$$

where $(\hat{\theta}_i)$ is the ML estimate of θ_i when $\tau = 0$ and the function $b_i(\theta_i)$ depends on the distribution assigned for the random effects.

The variance of the score function is, $\hat{V} = V \Big|_{\tau=0, \beta=\hat{\beta}} = \sqrt{I_{\tau\tau} - \mathbf{I}_{\beta\tau}^t \mathbf{I}_{\beta\beta}^{-1} \mathbf{I}_{\beta\tau}}$. Here, I represents the Fisher Information matrix evaluated at $\tau = 0$ and $\beta = \hat{\beta}$. Let $I = I(\beta, \tau)$ be partitioned as,

$$I = \begin{pmatrix} I_{\beta\beta} & \mathbf{I}_{\beta\tau} \\ \mathbf{I}_{\beta\tau}^t & I_{\tau\tau} \end{pmatrix},$$

with

$$I_{\tau\tau} = \sum_{i=1}^n \frac{b_i^2}{4} [g'_i \{5h'_i h''_i - 3(h'_i)^3 - h'''_i\} + 2\{h'_i g'_i - g''_i\}^2 \\ + g''_i \{6(h'_i)^2 - 4h''_i\} - 4g'''_i h'_i + g''''_i],$$

$$I_{\beta\beta} = U^t W_1 U, \text{ and}$$

$$I_{\beta\tau} = U^t W_2 \mathbf{1},$$

where $I_{\beta\beta}$ and $I_{\tau\tau}$ are $p \times p$ and 1×1 matrices, respectively, $h_i = h_i(\theta_i) = \log a'_i(\theta_i)$, U is an $n \times p$ matrix with $(ir)^{th}$ element $\frac{\partial \theta_i}{\partial \beta_r}$, $\mathbf{1}$ is an $n \times 1$ unit vector, and W_1 and W_2 are diagonal matrices. The i^{th} diagonal element of W_1 is $W_{1i} = g''_i - g'_i h'_i$ and for W_2 , it is $W_{2i} = \frac{b_i}{2} \{g'_i [(h'_i)^2 - h''_i] - 2g''_i h'_i + g'''_i\}$. The dependence on θ_i of the functions $b_i(\theta_i)$, $g_i(\theta_i)$ and $h_i(\theta_i)$ are suppressed for simplicity of notation.

Therefore, a standardized score test statistic for testing the null hypothesis $\tau = 0$ is

$$T = \frac{\sum_{i=1}^n \frac{\partial l_i}{\partial \tau} \Big|_{\tau=0, \beta=\hat{\beta}}}{\hat{V}}. \quad (4.2)$$

Refer to Appendix A for more details on these derivations. Large values of T provide evidence against the null hypothesis of $\tau = 0$. Under mild regularity conditions, the test statistic given in (4.2) is asymptotically standard normal as $n \rightarrow \infty$.

4.2.2 Testing for Overdispersion in Extra-Binomial Data

Here we derive the score statistic for testing overdispersion in an extra-binomial regression model. The overdispersed binomial density function is

$$f_{Y_i}(y_i; m_i, p_i^*) = \exp \left\{ y_i \log \left(\frac{p_i^*}{1 - p_i^*} \right) + m_i \log(1 - p_i^*) + \log \binom{m_i}{y_i} \right\},$$

where

$$\theta_i^* = \log \left(\frac{p_i^*}{1 - p_i^*} \right) = \mathbf{x}_i^t \boldsymbol{\beta} + z_i = \theta_i + z_i,$$

with z_i being *iid* variates with $E(Z_i) = 0$ and $V(Z_i) = \tau$. The interest here is to test the hypothesis $\tau = 0$. Details of the derivations of this score test are given in Appendix A.2.

The numerator of the score test statistic for testing the hypothesis $\tau = 0$ is,

$$\begin{aligned} \sum_{i=1}^n \frac{\partial l_i}{\partial \tau} \Big|_{\tau=0, \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} &= \sum_{i=1}^n \frac{b_i}{2} \{ (y_i - E(Y_i))^2 - V(Y_i) \} \\ &= \sum_{i=1}^n \frac{b_i}{2} \{ (y_i - m_i p_i)^2 - m_i p_i (1 - p_i) \}, \end{aligned}$$

where $E(Y_i)$ and $V(Y_i)$ are the mean and variance functions of the original binomial distribution. The function b_i depends on the distribution assigned for the random effects. If the random effects have a Normal distribution with mean 0 and variance τ , then $b_i = 1$, since $V(\theta_i^*) = \tau b_i = V(Z_i) = \tau$, where θ_i^* is the parameter of the overdispersed exponential family of distributions (in this case, binomial).

The terms in the variance function for the score are:

$$I_{\tau\tau} = \sum_{i=1}^n \frac{1}{4} \{2[m_i p_i (1 - p_i)]^2 + m_i p_i (1 - p_i) (1 - 6p_i + 6p_i^2)\},$$

$$I_{\beta\beta} = X^t W_1 X, \text{ and}$$

$$I_{\beta\tau} = X^t W_2 \mathbf{1},$$

where W_1 and W_2 are diagonal matrices with diagonal elements $W_{1i} = m_i p_i (1 - p_i)$ and $W_{2i} = \frac{1}{2} m_i p_i (1 - p_i) (1 - 2p_i)$, respectively.

4.3 Semi-parametric Approach

As an alternative to the likelihood score test, here we introduce a semi-parametric quasi-score (QS) test, which is based on the marginal mean and marginal variance-covariance matrix of the response vector $\mathbf{Y}$.

To define the QS function for (β, τ) , let $\gamma = (\beta^t, \tau)^t$, then the QS function for β can be expressed in the form:

$$K_{\beta}(\gamma) = \left(\frac{\partial E(\mathbf{Y})}{\partial \beta} \right) [V(\mathbf{Y})]^{-1} \{\mathbf{Y} - E(\mathbf{Y})\}, \quad (4.3)$$

where $V(\mathbf{Y})$ is the marginal variance-covariance matrix of $\mathbf{Y}$ and is considered as the working covariance matrix under $H_0 : \tau = 0$.

By the assumed independence of the observations, we can write (4.3) as

$$K_{\beta}(\gamma) = \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \{y_i - E(Y_i)\}. \quad (4.4)$$

For τ , following Sutradhar and Farrell (2004)[22] we consider a QS function in the form:

$$K_\tau(\gamma) = \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \{ [y_i - E(Y_i)]^2 - V(Y_i) \}. \quad (4.5)$$

The working variance under $\tau = 0$ is

$$\begin{aligned} V \{ [y_i - E(Y_i)]^2 \} &= E \{ [y_i - E(Y_i)]^4 \} - E \{ [y_i - E(Y_i)]^2 \}^2 \\ &= E \{ [Y_i - E(Y_i)]^4 \} - [V(Y_i)]^2. \end{aligned} \quad (4.6)$$

Under $\tau = 0$, $E(Y_i) = m_i p_i$ and $V(Y_i) = m_i p_i (1 - p_i)$. Hence the working variance under $\tau = 0$ is,

$$V \{ [y_i - E(Y_i)]^2 \} = E \{ [Y_i - m_i p_i]^4 \} - [m_i p_i (1 - p_i)]^2. \quad (4.7)$$

Here,

$$E[Y_i - m_i p_i]^4 = E(Y_i^4) - 4(m_i p_i)E(Y_i^3) + 6(m_i p_i)^2 E(Y_i^2) - 4(m_i p_i)^3 E(Y_i) + (m_i p_i)^4 \quad (4.8)$$

Therefore, (4.7) becomes,

$$\begin{aligned} V \{ [y_i - E(Y_i)]^2 \} &= E(Y_i^4) - 4(m_i p_i)E(Y_i^3) + 6(m_i p_i)^2 E(Y_i^2) - 4(m_i p_i)^3 E(Y_i) \\ &\quad + (m_i p_i)^4 - [m_i p_i (1 - p_i)]^2. \end{aligned} \quad (4.9)$$

The first four moments of the observations are calculated under $\tau = 0$ in Appendix A equations (A.50), (A.51), (A.52) and (A.53), respectively. Substituting these in equation (4.9), we have

$$V \{ [y_i - E(Y_i)]^2 \} = 2[m_i p_i (1 - p_i)]^2 + m_i p_i (1 - p_i) (1 - 6p_i + 6p_i^2). \quad (4.10)$$

4.3.1 Quasi-Score Test for τ

It is clear that the QS functions in (4.4) and (4.5) are based on the marginal mean and variance of the observations. Under the null hypothesis $\tau = 0$, the observations follow a simple binomial distribution and β can be estimated easily by the ML method. We use the *R* function “*glm*” for the ML estimates of β .

Following Boos (1992)[1], the QS function of the parameter τ adjusted for the nuisance parameters β is defined as

$$K^*(\tau|\beta) = K_\tau(\gamma) - J_{\tau\beta} J_{\beta\beta}^{-1} K_\beta(\gamma), \quad (4.11)$$

where

$$J_{\beta\beta} = E \left[-\frac{\partial K_\beta(\gamma)}{\partial \beta} \right] = \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \left(\frac{\partial E(Y_i)}{\partial \beta} \right)^t \text{ and}$$

$$J_{\tau\beta} = E \left[-\frac{\partial K_\tau(\gamma)}{\partial \beta} \right] = \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \left(\frac{\partial V(Y_i)}{\partial \beta} \right).$$

Refer to Appendix B for details on the above derivations.

Under general conditions and the central limit theorem, the function $K^*(\tau|\beta)$ under the null hypothesis is asymptotically normally distributed with mean 0 and variance function of $K^*(\tau|\beta)$. That is,

$$K^*(\tau|\beta) \Big|_{\tau=0} \sim N(0, V[K^*(\tau|\beta)]).$$

For testing $H_0 : \tau = 0$, we consider a quasi-score statistic in the form:

$$Q_0 = \frac{K^*(\tau|\beta)}{\sqrt{V[K^*(\tau|\beta)]}} \Big|_{\tau=0, \beta=\hat{\beta}}. \quad (4.12)$$

Under the null hypothesis $\tau = 0$, Q_0 in (4.12) asymptotically follows a standard normal distribution.

To find $V[K^*(\tau|\beta)]$, first note that the functions $\mathbf{K}_\beta(\gamma)$ and $K_\tau(\gamma)$ in equation (4.11) are in linear form with coefficients $-J_{\tau\beta}J_{\beta\beta}^{-1}$ and 1 respectively. Then using a sandwich type variance formula to approximate $V[K^*(\tau|\beta)]$, we get

$$\hat{V}[K^*(\tau|\beta)] = \begin{pmatrix} -J_{\tau\beta}J_{\beta\beta}^{-1} & 1 \end{pmatrix} V[\mathbf{K}(\gamma)] \begin{pmatrix} -(J_{\beta\beta}^{-1})^t J_{\tau\beta}^t \\ 1 \end{pmatrix},$$

where $V[\mathbf{K}(\gamma)]$ is the variance-covariance matrix of

$\mathbf{K}(\gamma) = (\mathbf{K}_1(\gamma), \mathbf{K}_2(\gamma), \dots, \mathbf{K}_n(\gamma))$ at $\tau = 0$. Define $\mathbf{K}(\gamma) = [\mathbf{K}_\beta^t(\gamma), K_\tau(\gamma)]^t$ as the joint QS function for β and τ .

The QS test statistic in (4.12) is derived under $\tau = 0$. Following McCullagh and Nelder 1989[12], under $\tau = 0$, $V[\mathbf{K}(\gamma)]$ is the negative expected value of the first derivative of the QS function. The variance-covariance matrix of $\mathbf{K}(\gamma)$ is then given by

$$\begin{aligned} V[\mathbf{K}(\gamma)] &= E \left[-\frac{\partial \mathbf{K}(\gamma)}{\partial \gamma} \right] \\ &= \begin{pmatrix} E \left[-\frac{\partial \mathbf{K}_\beta(\gamma)}{\partial \beta} \right] & E \left[-\frac{\partial \mathbf{K}_\beta(\gamma)}{\partial \tau} \right] \\ E \left[-\frac{\partial K_\tau(\gamma)}{\partial \beta} \right] & E \left[-\frac{\partial K_\tau(\gamma)}{\partial \tau} \right] \end{pmatrix} \\ &= \begin{pmatrix} J_{\beta\beta} & J_{\beta\tau} \\ J_{\tau\beta} & J_{\tau\tau} \end{pmatrix} \end{aligned} \tag{4.13}$$

and thus

$$\begin{aligned}\widehat{V}[K^*(\tau|\beta)] &= \begin{pmatrix} -J_{\tau\beta}J_{\beta\beta}^{-1} & 1 \end{pmatrix} \begin{pmatrix} J_{\beta\beta} & J_{\beta\tau} \\ J_{\tau\beta} & J_{\tau\tau} \end{pmatrix} \begin{pmatrix} -(J_{\beta\beta}^{-1})^t J_{\tau\beta}^t \\ 1 \end{pmatrix} \\ &= J_{\tau\tau} - J_{\tau\beta}J_{\beta\beta}^{-1}J_{\beta\tau},\end{aligned}\tag{4.14}$$

where

$$\begin{aligned}J_{\tau\tau} &= \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right)^2 V \{ [y_i - E(Y_i)]^2 \}^{-1}, \\ J_{\tau\beta} &= \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \left(\frac{\partial V(Y_i)}{\partial \beta} \right), \\ J_{\beta\beta} &= \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \left(\frac{\partial E(Y_i)}{\partial \beta} \right)^t, \quad \text{and} \\ J_{\beta\tau} &= \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \left(\frac{\partial E(Y_i)}{\partial \tau} \right).\end{aligned}\tag{4.15}$$

Refer to Appendix B for details on the above derivations. All the terms in the above are evaluated under $\tau = 0$.

The $K_\beta(\gamma)$ term in the numerator of Q_0 becomes zero when evaluated at $\beta = \hat{\beta}$ giving

$$Q_0 = \frac{K_\tau(\gamma)}{\sqrt{J_{\tau\tau} - J_{\tau\beta}J_{\beta\beta}^{-1}J_{\beta\tau}}}.$$

All the terms in the above equation are evaluated under the null hypothesis of $\tau = 0$.

4.3.2 Testing for Overdispersion in Extra-Binomial Data

Here we illustrate the computational issues of the QS test using the example of extra-binomial model introduced earlier.

The mean of p_i^* is

$$E(p_i^*) = \int p_i^* f(z_i) dz_i. \quad (4.16)$$

In section 3.3.1, the marginal mean and variance of the binomial responses were obtained as:

$$\begin{aligned} E(Y_i) &= m_i E(p_i^*) = m_i \int p_i^* f(z_i) dz_i \quad \text{and} \\ V(Y_i) &= m_i E(p_i^*) - [m_i E(p_i^*)]^2 + m_i(m_i - 1) E(p_i^{*2}) \\ &= m_i \int p_i^* f(z_i) dz_i - \left[m_i \int p_i^* f(z_i) dz_i \right]^2 + m_i(m_i - 1) \int p_i^{*2} f(z_i) dz_i. \end{aligned}$$

Under $H_0 : \tau = 0$, $p_i^* = p_i$, and so we have

$$E(Y_i) = m_i p_i \quad \text{and}$$

$$V(Y_i) = m_i p_i (1 - p_i).$$

Now, for the QL function for β , we have

$$\begin{aligned} \frac{\partial E(Y_i)}{\partial \beta} &= \frac{\partial}{\partial \beta} \left(m_i \int p_i^* f(z_i) dz_i \right) = m_i \int \frac{\partial}{\partial \beta} \left(\frac{\exp(x_i^t \beta + z_i)}{1 + \exp(x_i^t \beta + z_i)} \right) f(z_i) dz_i \\ &= m_i \int p_i^* (1 - p_i^*) \mathbf{x}_i f(z_i) dz_i. \end{aligned} \quad (4.17)$$

Under $H_0 : \tau = 0$, $p_i^* = p_i$, and so we have

$$\frac{\partial E(Y_i)}{\partial \beta} = m_i p_i (1 - p_i) \mathbf{x}_i. \quad (4.18)$$

Similarly, for the QL function for τ , we have

$$\begin{aligned}
 \left(\frac{\partial V(Y_i)}{\partial \tau} \right) &= \frac{\partial}{\partial \tau} \left\{ m_i \int p_i^* f(u_i) du_i - \left[m_i \int p_i^* f(u_i) du_i \right]^2 \right\} \\
 &\quad + \frac{\partial}{\partial \tau} \left\{ m_i(m_i - 1) \int p_i^{*2} f(u_i) du_i \right\} \\
 &= m_i \int \frac{\partial p_i^*}{\partial \tau} f(u_i) du_i - 2m_i^2 \left[\int p_i^* f(u_i) du_i \right] \int \frac{\partial p_i^*}{\partial \tau} f(u_i) du_i \\
 &\quad + m_i(m_i - 1) \int 2p_i^* \frac{\partial p_i^*}{\partial \tau} f(u_i) du_i,
 \end{aligned} \tag{4.19}$$

where

$$\frac{\partial p_i^*}{\partial \tau} = \frac{\partial}{\partial \tau} \left(\frac{\exp(x_i^t \beta + \sqrt{\tau} u_i)}{1 + \exp(x_i^t \beta + \sqrt{\tau} u_i)} \right) = \left(\frac{1}{2\sqrt{\tau}} \right) p_i^* (1 - p_i^*) u_i. \tag{4.20}$$

Therefore, equation (4.19) becomes

$$\begin{aligned}
 \left(\frac{\partial V(Y_i)}{\partial \tau} \right) &= \frac{(m_i/2)}{\sqrt{\tau}} \int p_i^* (1 - p_i^*) u_i f(u_i) du_i \\
 &\quad - \frac{m_i^2}{\sqrt{\tau}} \left[\int p_i^* f(u_i) du_i \right] \int p_i^* (1 - p_i^*) u_i f(u_i) du_i \\
 &\quad + \frac{m_i(m_i - 1)}{\sqrt{\tau}} \int p_i^{*2} (1 - p_i^*) u_i f(u_i) du_i.
 \end{aligned} \tag{4.21}$$

Note that we cannot evaluate the term (4.21) under $\tau = 0$, since $\sqrt{\tau}$ is in the denominator. We use L'Hospital's Rule to find the limiting value of $\left(\frac{\partial V(Y_i)}{\partial \tau} \right)$ when $\tau = 0$.

If f and g are differentiable in a neighbourhood of $x = c$, and $f(c) = g(c) = 0$, then,

$$\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \lim_{x \rightarrow c} \frac{f'(x)}{g'(x)}$$

provided the limit of the right exists.

Since both the numerator and the denominator of equation (4.21) tends to zero as $\tau \rightarrow 0$, the first derivatives of both numerator and denominator with respect to τ are obtained separately. The derivative of the denominator of (4.21) is

$$\frac{\partial \sqrt{\tau}}{\partial \tau} = \frac{1}{2\sqrt{\tau}}.$$

The numerator of (4.21) has the derivative:

$$\begin{aligned} \frac{\partial A}{\partial \tau} &= \frac{(m_i/2)}{2\sqrt{\tau}} \int p_i^*(1-p_i^*)(1-2p_i^*)u_i^2 f(u_i) du_i \\ &\quad - \frac{m_i^2}{2\sqrt{\tau}} \left[\int p_i^*(1-p_i^*)u_i f(u_i) du_i \right] \int p_i^*(1-p_i^*)u_i f(u_i) du_i \\ &\quad - \frac{m_i^2}{2\sqrt{\tau}} \left[\int p_i^* f(u_i) du_i \right] \int p_i^*(1-p_i^*)(1-2p_i^*)u_i^2 f(u_i) du_i \\ &\quad + \frac{m_i(m_i-1)}{2\sqrt{\tau}} \int p_i^{*2}(1-p_i^*)(2-3p_i^*)u_i^2 f(u_i) du_i. \end{aligned} \quad (4.22)$$

We can ignore the term $\frac{1}{2\sqrt{\tau}}$ as it is common in the derivatives of both numerator and denominator.

Now letting $\tau = 0$ and thus $p_i^* = p_i$,

$$\begin{aligned} E(U_i) &= 0 = \int u_i f(u_i) du_i, \quad \text{and} \\ V(U_i) &= 1 = \int u_i^2 f(u_i) du_i, \end{aligned}$$

the equation (4.22) becomes

$$\begin{aligned} \left. \frac{\partial A}{\partial \tau} \right|_{\tau=0} &= \frac{m_i}{2} p_i(1-p_i)(1-2p_i) - m_i^2 p_i^2(1-p_i)(1-2p_i) \\ &\quad + m_i(m_i-1)p_i^2(1-p_i)(2-3p_i) \\ &= \left(\frac{m_i}{2} \right) (1-6p_i+6p_i^2) p_i(1-p_i) + [m_i p_i(1-p_i)]^2. \end{aligned} \quad (4.23)$$

Therefore, using the L'Hospital's Rule, equation (4.19) has the limiting value:

$$\left(\frac{\partial V(Y_i)}{\partial \tau}\right) = \left(\frac{m_i}{2}\right) (1 - 6p_i + 6p_i^2) p_i(1 - p_i) + [m_i p_i(1 - p_i)]^2. \quad (4.24)$$

The working variance under the null hypothesis is (as given in (4.10))

$$V\{[Y_i - E(Y_i)]^2\} = 2[m_i p_i(1 - p_i)]^2 + m_i p_i(1 - p_i)(1 - 6p_i + 6p_i^2). \quad (4.25)$$

Recall the QS statistic Q_0 for testing $\tau = 0$:

$$Q_0 = \frac{K_\tau(\gamma)}{\sqrt{J_{\tau\tau} - J_{\tau\beta} J_{\beta\beta}^{-1} J_{\beta\tau}}}.$$

To calculate the terms in the denominator of Q_0 , we need

$$\begin{aligned} J_{\tau\tau} &= \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau}\right)^2 V\{[y_i - E(Y_i)]^2\}^{-1}, \\ J_{\tau\beta} &= \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau}\right) V\{[y_i - E(Y_i)]^2\}^{-1} \left(\frac{\partial V(Y_i)}{\partial \beta}\right), \\ J_{\beta\beta} &= \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta}\right) [V(Y_i)]^{-1} \left(\frac{\partial E(Y_i)}{\partial \beta}\right)^t, \quad \text{and} \\ J_{\beta\tau} &= \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta}\right) [V(Y_i)]^{-1} \left(\frac{\partial E(Y_i)}{\partial \tau}\right), \end{aligned}$$

where

$$\begin{aligned} \frac{\partial V(Y_i)}{\partial \beta} &= \frac{\partial}{\partial \beta} \left\{ m_i \int p_i^* f(u_i) du_i - \left[m_i \int p_i^* f(u_i) du_i \right]^2 \right\} \\ &\quad + \frac{\partial}{\partial \beta} \left\{ m_i(m_i - 1) \int p_i^{*2} f(u_i) du_i \right\} \\ &= m_i \int \frac{\partial p_i^*}{\partial \beta} f(u_i) du_i - 2m_i^2 \left[\int p_i^* f(u_i) du_i \right] \int \frac{\partial p_i^*}{\partial \beta} f(u_i) du_i \\ &\quad + m_i(m_i - 1) \int 2p_i^* \frac{\partial p_i^*}{\partial \beta} f(u_i) du_i, \end{aligned} \quad (4.26)$$

and $\frac{\partial p_i^*}{\partial \beta} = p_i^*(1 - p_i^*)\mathbf{x}_i$. Therefore (4.26) becomes

$$\begin{aligned} \frac{\partial V(Y_i)}{\partial \beta} &= m_i \int p_i^*(1 - p_i^*)\mathbf{x}_i f(u_i) du_i - 2m_i^2 \left[\int p_i^* f(u_i) du_i \right] \int p_i^*(1 - p_i^*)\mathbf{x}_i f(u_i) du_i \\ &\quad + m_i(m_i - 1) \int 2p_i^{*2}(1 - p_i^*)\mathbf{x}_i f(u_i) du_i. \end{aligned} \quad (4.27)$$

Under $H_0 : \tau = 0$, $p_i^* = p_i$, and so we have

$$\frac{\partial V(Y_i)}{\partial \beta} = m_i p_i (1 - p_i) (1 - 2p_i) \mathbf{x}_i. \quad (4.28)$$

Also we need

$$\frac{\partial E(Y_i)}{\partial \tau} = \frac{\partial}{\partial \tau} \left(m_i \int p_i^* f(z_i) dz_i \right) = \left(\frac{m_i}{2\sqrt{\tau}} \right) \int p_i^*(1 - p_i^*) u_i f(z_i) dz_i. \quad (4.29)$$

Here also we cannot evaluate the term (4.29) under $\tau = 0$, since $\sqrt{\tau}$ is in the denominator. Using L'Hospital's Rule, the limiting value is obtained as

$$\frac{\partial E(Y_i)}{\partial \tau} = \frac{m_i}{2} p_i (1 - p_i) (1 - 2p_i). \quad (4.30)$$

Now, the terms in the variance function are:

$$\begin{aligned} J_{\tau\tau} &= \sum_{i=1}^n \frac{1}{4} \{ 2[m_i p_i (1 - p_i)]^2 + m_i p_i (1 - p_i) (1 - 6p_i + 6p_i^2) \}, \\ J_{\tau\beta} &= \sum_{i=1}^n \frac{1}{2} m_i p_i (1 - p_i) (1 - 2p_i) \mathbf{x}_i^t, \\ J_{\beta\beta} &= \sum_{i=1}^n \mathbf{x}_i m_i p_i (1 - p_i) \mathbf{x}_i^t, \quad \text{and} \\ J_{\beta\tau} &= \sum_{i=1}^n \frac{1}{2} m_i p_i (1 - p_i) (1 - 2p_i) \mathbf{x}_i = J_{\tau\beta}^t. \end{aligned}$$

The numerator of Q_0 under $\tau = 0$ is

$$\begin{aligned} K_\tau(\gamma) &= \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \{ [y_i - E(Y_i)]^2 - V(Y_i) \} \\ &= \sum_{i=1}^n \frac{1}{2} \{ [y_i - m_i p_i]^2 - m_i p_i (1 - p_i) \}. \end{aligned}$$

Note that the terms used in the score test statistic and QS test statistic under the null hypothesis of $\tau = 0$ are identical (see section 4.2.2 for the score test statistic). So for testing overdispersion in extra-binomial data, it is evident that the score test developed by Dean (1992)[6] is identical to the QS test statistic that we consider here. The derivation of the QS test is attractive here in that it requires only the first few moments of the marginal distribution of the responses.

4.4 Misspecified Random Effects

In many cases, we assume a normal distribution for the random effects. In practice, this normality assumption for the random effects may be violated in some situations. So we study the robustness properties of the QS tests under misspecified non-normal random effects.

We consider a mixed model in the form:

$$\log \left(\frac{p_i^*}{1 - p_i^*} \right) = x_i^t \beta + z_i,$$

where the random effects z_i follow a non-normal distribution:

$$Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right) \quad (4.31)$$

with $a_i \sim \text{gamma}(\lambda, 1)$. Since $E(a_i) = \lambda$ and $V(a_i) = \lambda$, clearly, $E(Z_i) = 0$ and $V(Z_i) = \tau$. For this particular case, the mean and variance of the random effects are the same as that of the normally distributed random effects.

A simulation study is done based on these misspecified random effects in Chapter 6.

4.5 Power of the Tests

We study the power of the QS test for testing the overdispersion parameter τ .

The power of a test is defined as

$$\begin{aligned} \text{power} &= 1 - \beta = 1 - \text{Prob}(\text{Type II error}) \\ &= 1 - \text{Prob}(\text{accepting the null hypothesis when the alternative is true}) \\ &= \text{probability of rejecting the null hypothesis when the alternative is true.} \end{aligned}$$

To compute the empirical powers, we ran a simulation study in which P-values were calculated using the asymptotic normality property of the QS statistic. The P-value is calculated from $P[Q \geq Q_0]$, where Q is $N(0, 1)$, and Q_0 is value of the test statistic.

The empirical power was obtained as the proportion of times the null hypothesis was rejected when H_A was true.

4.6 Empirical Significance Levels

The significance level α of a test is the “maximum probability of rejecting the null hypothesis when it is true”.

To calculate the empirical significance level, simulations were carried out using data with no overdispersion. The empirical level was obtained as the proportion of times H_0 was rejected, when, in fact, H_0 was true. We study the empirical levels of QS tests for different sample sizes.

Chapter 5

Parameter Estimation and Related Topics

5.1 Introduction

In the context of models accommodating overdispersion, there are two main sets of parameters, namely, regression parameters and overdispersion parameter, which need to be estimated. When a larger overdispersed family of models are constructed, the overdispersion is accommodated to the model through the random effects. A complete maximum likelihood (ML) analysis based on the joint marginal likelihood of the responses can be used to estimate both regression parameters and overdispersion parameter. To define a likelihood, one has to specify the form of the marginal distribution of the observations. Although the conditional density of the observations (given the random effects) is known, the calculation of the marginal likelihood function is tedious and it requires numerical integration over the distribution of the random effects. In addition, it will require numerical integration techniques for the

calculation of score functions and information matrix. To avoid such computational difficulties, alternative methods were suggested by many authors. For example, the quasi-likelihood (QL) function can be used to estimate the parameters of the fixed and random effects numerically. The idea of QL was developed by Wedderburn (1974)[25] and later examined extensively by McCullagh and Nelder (1989)[12].

5.2 Quasi-Likelihood Estimating Equations

As illustrated by Wedderburn (1974)[25] and McCullagh and Nelder (1989)[12], the QL estimating equations for the regression parameters β in a GLM, may be written in the form $K(\beta)|_{\beta=\hat{\beta}} = 0$, where

$$K(\beta) = D^T V^{-1} (Y - \mu) \quad (5.1)$$

is called the *quasi-score* (QS) function. In this expression the components of D , of order $n \times p$, are $D_{ir} = \frac{\partial \mu_i}{\partial \beta_r}$, the derivatives of $\mu(\beta)$ with respect to the parameters.

The covariance matrix of $K(\beta)$, which is also the negative expected value of $\frac{\partial K(\beta)}{\partial \beta}$, is obtained from

$$J_{\beta\beta} = D^T V^{-1} D. \quad (5.2)$$

For QL functions, this matrix plays the same role as the Fisher information for ordinary likelihood functions. The expected value of $\frac{\partial K(\beta)}{\partial \beta}$ is used to avoid the heavy computational burden. When estimating the parameters using an iterative algorithm,

the inverse of the information matrix given in (5.2) is used instead of employing the inverse of the actual derivative matrix, just by replacing it with its expected value. This method is called the *method of scoring* (Searle, Casella and McCulloch (1992)[18]).

As discussed before, the QS function for a parameter θ has the form.

$$K_{\theta}(\gamma) = \sum_{i=1}^n \left(\frac{\partial E[h(Y_i)]}{\partial \theta} \right) \{V[h(Y_i)]\}^{-1} \{h(y_i) - E[h(Y_i)]\}, \quad (5.3)$$

where $h(Y_i)$ is any function of the observations. For β , we choose $h(Y_i) = Y_i$ and hence

$$K_{\beta}(\gamma) = \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \{y_i - E(Y_i)\}. \quad (5.4)$$

For estimating τ , we choose $h(Y_i) = [y_i - E(Y_i)]^2$ and so the QS function for τ is

$$K_{\tau}(\gamma) = \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{[y_i - E(Y_i)]^2\}^{-1} \{[y_i - E(Y_i)]^2 - V(Y_i)\}. \quad (5.5)$$

Here a working variance is used to replace the true $V[Y_i - E(Y_i)]^2$ and to avoid any computational difficulties. If the true covariance matrix is to be used, the fourth moment of the marginal distribution need to be computed. Although the marginal mean and variance of the observations are known, the form the marginal distribution is unknown. So we consider replacing the true $V[Y_i - E(Y_i)]^2$ by a “working” variance under $\tau = 0$.

5.3 Parameter Estimation using Newton-Raphson Iterative Method

The NR iterative process is used to estimate the regression parameters and overdispersion parameter. Let $\mathbf{K}(\gamma) = [K_\beta^t(\gamma), K_\tau(\gamma)]^t$ be the joint QS function for estimating $(\beta^t, \tau)^t$. The NR iterative equation for γ can be obtained in the form:

$$\gamma^{(k+1)} = \gamma^{(k)} + \left[J_\gamma \Big|_{\gamma=\gamma^{(k)}} \right]^{-1} \mathbf{K}(\gamma) \Big|_{\gamma=\gamma^{(k)}} \quad (5.6)$$

for $k = 0, 1, 2, \dots$, where

$$J_\gamma = \begin{pmatrix} E \left[-\frac{\partial K_\beta(\gamma)}{\partial \beta} \right] & E \left[-\frac{\partial K_\beta(\gamma)}{\partial \tau} \right] \\ E \left[-\frac{\partial K_\tau(\gamma)}{\partial \beta} \right] & E \left[-\frac{\partial K_\tau(\gamma)}{\partial \tau} \right] \end{pmatrix} = \begin{pmatrix} J_{\beta\beta} & J_{\beta\tau} \\ J_{\tau\beta} & J_{\tau\tau} \end{pmatrix}. \quad (5.7)$$

The NR algorithm can be described as follows:

1. Set $k = 0$ and choose initial values $\gamma^{(0)}$.
2. Calculate QS function $\mathbf{K}(\gamma)$ and J_γ evaluated at $\gamma^{(k)}$ to compute $\gamma^{(k+1)}$.
3. The process is terminated when the largest of the relative changes in the values of the parameter estimates are sufficiently small, otherwise, set $k = k + 1$ and continue step 2.
4. At convergence, declare $\gamma^{(k+1)}$ to be the estimates of the parameters.

5.4 Variance-Covariance of the Estimators

We use a sandwich-type estimator for approximating the variance-covariance matrix of the estimators of the regression parameters and the overdispersion parameter.

These variance estimators were used to find the confidence intervals on β and τ .

The sandwich type estimator of $V(\hat{\gamma})$ is obtained from

$$\begin{aligned} V(\hat{\gamma}) &\approx [-\mathbf{K}'(\gamma)]^{-1} V\{\mathbf{K}(\gamma)\} \{[-\mathbf{K}'(\gamma)]^{-1}\}^t \Big|_{\gamma=\hat{\gamma}} \\ &\approx [E[-\mathbf{K}'(\gamma)]]^{-1} V\{\mathbf{K}(\gamma)\} \{[E[-\mathbf{K}'(\gamma)]]^{-1}\}^t \Big|_{\gamma=\hat{\gamma}} \text{ (by scoring method)} \\ &= J_{\gamma}^{-1} V\{\mathbf{K}(\gamma)\} \{J_{\gamma}^{-1}\}^t \Big|_{\gamma=\hat{\gamma}} \end{aligned} \quad (5.8)$$

where J_{γ} is given in (5.7), and

$$V\{\mathbf{K}(\gamma)\} = E\{\mathbf{K}(\gamma)\mathbf{K}^t(\gamma)\},$$

since $E\{\mathbf{K}(\gamma)\} = 0$.

We approximate $V\{\mathbf{K}(\gamma)\}$ as

$$\begin{aligned} \hat{V}\{\mathbf{K}(\gamma)\} &= E\{\mathbf{K}(\gamma)\mathbf{K}^t(\gamma)\} \Big|_{\gamma=\hat{\gamma}} \\ &\approx \sum_{i=1}^n \mathbf{K}_i(\gamma)\mathbf{K}_i^t(\gamma) \Big|_{\gamma=\hat{\gamma}}. \end{aligned} \quad (5.9)$$

Therefore, the approximate variance-covariance matrix of $\hat{\gamma}$ is

$$V(\hat{\gamma}) \approx J_{\gamma}^{-1} \left[\sum_{i=1}^n \mathbf{K}_i(\gamma)\mathbf{K}_i^t(\gamma) \right] \{J_{\gamma}^{-1}\}^t \Big|_{\gamma=\hat{\gamma}}. \quad (5.10)$$

5.5 Confidence Intervals of the Parameters and their Mean Lengths

Assuming asymptotic normality of the estimators, a $100(1 - \alpha)\%$ confidence interval for the regression parameter β_s is obtained from

$$\hat{\beta}_s \pm S.E.(\hat{\beta}_s)Z_{\alpha/2}, \quad (5.11)$$

where $s = 0, 1, 2, \dots, p$, $\hat{\beta}_s$ is the estimated value of the regression parameter β_s and $Z_{\alpha/2}$ is the upper $100(\alpha/2)^{th}$ percentile point in a standard normal distribution.

The length of this confidence interval is obtained as

$$\text{Length} = 2S.E.(\hat{\beta}_s)Z_{\alpha/2}. \quad (5.12)$$

Similarly, a $100(1 - \alpha)\%$ confidence interval for the overdispersion parameter τ is obtained from

$$\hat{\tau} \pm S.E.(\hat{\tau})Z_{\alpha/2}. \quad (5.13)$$

5.6 Coverage Probability

Once the confidence intervals for the parameters are computed, the empirical *coverage probability* can be obtained from a simulation study.

Let the total number of simulations be k . Then we find the confidence intervals of the desired parameter for each simulation. Next we count the number of instances where the chosen value of the parameter is within this calculated confidence interval.

Let this count be denoted by $\tilde{k}$. Then,

$$\text{Coverage Probability} = \frac{\tilde{k}}{k}. \quad (5.14)$$

We find the empirical coverage probabilities of the confidence intervals for all regression and overdispersion parameters in the simulation study.

5.7 Approximation of Marginal Moments

Two different approaches were described in Chapter 3 section 3.3.2 for computing the marginal mean and variance of the observations. Using these two methods, parameters can be estimated iteratively. Details are given in the following two sections.

5.7.1 Small Sigma Approximation Method

In the SSA method, approximate mean of the marginal responses, $E(Y_i) \approx m_i p_i$, depends only on the regression parameter β . But the approximate variance of the marginal responses, $V(Y_i) \approx m_i p_i (1 - p_i) [1 + \tau(m_i - 1)p_i(1 - p_i)]$, depends on both regression parameter β and overdispersion parameter τ .

The QS function for β is

$$K_{\beta}(\gamma) = \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \{y_i - E(Y_i)\}, \quad (5.15)$$

where

$$\frac{\partial E(Y_i)}{\partial \beta} \approx \frac{\partial m_i p_i}{\partial \beta} = m_i \frac{\partial p_i}{\partial \beta} \quad (5.16)$$

with

$$\begin{aligned}\frac{\partial p_i}{\partial \beta} &= \frac{\partial}{\partial \beta} \left(\frac{\exp\{\mathbf{x}_i^t \beta\}}{1 + \exp\{\mathbf{x}_i^t \beta\}} \right) = \left(\frac{\exp\{\mathbf{x}_i^t \beta\}}{[1 + \exp\{\mathbf{x}_i^t \beta\}]^2} \right) \mathbf{x}_i \\ &= p_i(1 - p_i)\mathbf{x}_i.\end{aligned}\tag{5.17}$$

Hence

$$\frac{\partial E(Y_i)}{\partial \beta} \approx m_i p_i(1 - p_i)\mathbf{x}_i.\tag{5.18}$$

For estimating τ , we use the QS function:

$$K_\tau(\gamma) = \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \{ [y_i - E(Y_i)]^2 - V(Y_i) \}.\tag{5.19}$$

As before, we use a working variance for $V[y_i - E(Y_i)]^2$:

$$V \{ [y_i - E(Y_i)]^2 \} = 2[m_i p_i(1 - p_i)]^2 + m_i p_i(1 - p_i)(1 - 6p_i + 6p_i^2).\tag{5.20}$$

Now,

$$\begin{aligned}\left(\frac{\partial V(Y_i)}{\partial \tau} \right) &\approx \frac{\partial}{\partial \tau} \{ m_i p_i(1 - p_i)[1 + \tau(m_i - 1)p_i(1 - p_i)] \} \\ &= m_i(m_i - 1)p_i^2(1 - p_i)^2.\end{aligned}\tag{5.21}$$

The QS function for τ can be obtained by substituting this in equation (5.19).

We estimate β and τ simultaneously using the NR method described earlier. For

this, we need the following:

$$\begin{aligned}
J_{\beta\beta} &\approx \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \left(\frac{\partial E(Y_i)}{\partial \beta} \right)^t, \\
J_{\tau\tau} &\approx \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right)^2 V \{ [y_i - E(Y_i)]^2 \}^{-1}, \\
J_{\beta\tau} &\approx \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \left(\frac{\partial V(Y_i)}{\partial \tau} \right), \quad \text{and} \\
J_{\tau\beta} &\approx \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \left(\frac{\partial E(Y_i)}{\partial \beta} \right). \tag{5.22}
\end{aligned}$$

We refer to Appendix B for details on the above derivations.

The SSA method gives $\frac{\partial E(Y_i)}{\partial \tau} \approx 0$, and

$$\begin{aligned}
\frac{\partial V(Y_i)}{\partial \beta} &\approx \frac{\partial}{\partial \beta} m_i p_i (1 - p_i) [1 + \tau(m_i - 1)p_i(1 - p_i)] \\
&= m_i(1 - 2p_i) \frac{\partial p_i}{\partial \beta} + \tau(m_i - 1)2p_i(1 - p_i)(1 - 2p_i) \frac{\partial p_i}{\partial \beta} \\
&= m_i(1 - 2p_i)p_i(1 - p_i)\mathbf{x}_i[1 + 2\tau(m_i - 1)p_i(1 - p_i)]. \tag{5.23}
\end{aligned}$$

The variance of the QL estimators of β and τ can be obtained from equation (5.10).

5.7.2 Simulation Method

In the simulation method, mean of the marginal responses is approximated as $E(Y_i) \doteq \frac{m_i}{S} \sum_{s=1}^S p_i^{*(s)}$, where $p_i^{*(s)} = \frac{\exp(\mathbf{x}_i^t \boldsymbol{\beta} + \sqrt{\tau} u_i^{(s)})}{1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta} + \sqrt{\tau} u_i^{(s)})}$, and $u_i^{(s)}$ are generated from $N(0, 1)$. Similarly, the variance of the marginal responses is obtained from $V(Y_i) \doteq \frac{m_i}{S} \sum_{s=1}^S p_i^{*(s)}(1 - p_i^{*(s)}) + m_i^2 \sum_{s=1}^S \frac{\{p_i^{*(s)} - E(p_i^*)\}^2}{S - 1}$.

The QS function for the regression parameters $\boldsymbol{\beta}$ is

$$\mathbf{K}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \boldsymbol{\beta}} \right) [V(Y_i)]^{-1} \{y_i - E(Y_i)\}, \quad (5.24)$$

where

$$\begin{aligned} \frac{\partial E(Y_i)}{\partial \boldsymbol{\beta}} &\doteq \frac{\partial}{\partial \boldsymbol{\beta}} \left(m_i \sum_{s=1}^S \frac{p_i^{*(s)}}{S} \right) = \left(\frac{m_i}{S} \right) \frac{\partial}{\partial \boldsymbol{\beta}} \left(\sum_{s=1}^S \frac{\exp(\mathbf{x}_i^t \boldsymbol{\beta} + \sqrt{\tau} u_i^{(s)})}{1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta} + \sqrt{\tau} u_i^{(s)})} \right) \\ &= \frac{m_i}{S} \left\{ \sum_{s=1}^S \frac{\exp(\mathbf{x}_i^t \boldsymbol{\beta} + \sqrt{\tau} u_i^{(s)})}{[1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta} + \sqrt{\tau} u_i^{(s)})]^2} \right\} \mathbf{x}_i \\ &= \frac{m_i}{S} \left\{ \sum_{s=1}^S p_i^{*(s)}(1 - p_i^{*(s)}) \right\} \mathbf{x}_i. \end{aligned} \quad (5.25)$$

The QS function for τ is

$$K_{\tau}(\boldsymbol{\gamma}) = \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{[y_i - E(Y_i)]^2\}^{-1} \{[y_i - E(Y_i)]^2 - V(Y_i)\}, \quad (5.26)$$

where $V \{[y_i - E(Y_i)]^2\}$ is taken to be the working variance under $\tau = 0$. Similar to the SSA method,

$$V \{[y_i - E(Y_i)]^2\} = 2[m_i p_i(1 - p_i)]^2 + m_i p_i(1 - p_i)(1 - 6p_i + 6p_i^2). \quad (5.27)$$

For $K_\tau(\gamma)$ in (5.26), we have

$$\begin{aligned} \left(\frac{\partial V(Y_i)}{\partial \tau} \right) &\doteq \frac{\partial}{\partial \tau} \left\{ \frac{m_i}{S} \sum_{s=1}^S p_i^{*(s)} (1 - p_i^{*(s)}) + m_i^2 \sum_{s=1}^S \frac{\{p_i^{*(s)} - E(p_i^*)\}^2}{S-1} \right\} \\ &= \frac{m_i}{S} \sum_{s=1}^S (1 - 2p_i^{*(s)}) \frac{\partial p_i^{*(s)}}{\partial \tau} \\ &\quad + \frac{m_i^2}{(S-1)} \sum_{s=1}^S 2 [p_i^{*(s)} - E(p_i^*)] \frac{\partial [p_i^{*(s)} - E(p_i^*)]}{\partial \tau}, \end{aligned} \quad (5.28)$$

where

$$\begin{aligned} \frac{\partial p_i^{*(s)}}{\partial \tau} &= \frac{\partial}{\partial \tau} \left(\frac{\exp(x_i^t \beta + \sqrt{\tau} u_i^{(s)})}{1 + \exp(x_i^t \beta + \sqrt{\tau} u_i^{(s)})} \right) = \left(\frac{u_i^{(s)}}{2\sqrt{\tau}} \right) \frac{\exp(x_i^t \beta + \sqrt{\tau} u_i^{(s)})}{[1 + \exp(x_i^t \beta + \sqrt{\tau} u_i^{(s)})]^2} \\ &= \left(\frac{1}{2\sqrt{\tau}} \right) p_i^{*(s)} (1 - p_i^{*(s)}) u_i^{(s)}, \end{aligned} \quad (5.29)$$

and $p_i^{*(s)} - E(p_i^*) \doteq p_i^{*(s)} - \frac{1}{S} \sum_{s=1}^S p_i^{*(s)}$. Also,

$$\frac{\partial [p_i^{*(s)} - E(p_i^*)]}{\partial \tau} \doteq \frac{1}{2\sqrt{\tau}} \left\{ p_i^{*(s)} (1 - p_i^{*(s)}) u_i^{(s)} - \frac{1}{S} \sum_{s=1}^S p_i^{*(s)} (1 - p_i^{*(s)}) u_i^{(s)} \right\}. \quad (5.30)$$

The parameters are estimated using the NR iterative process introduced in section 5.3. Similar to the SSA method, the terms $J_{\beta\beta}$, $J_{\tau\tau}$, $J_{\beta\tau}$ and $J_{\tau\beta}$ given in equation set (5.22) have to be determined. For this, we have

$$\begin{aligned} \frac{\partial E(Y_i)}{\partial \tau} &= \frac{\partial}{\partial \tau} m_i E(p_i^*) \doteq m_i \frac{\partial}{\partial \tau} \left\{ \frac{1}{S} \left(\sum_{s=1}^S p_i^{*(s)} \right) \right\} \\ &= \frac{m_i}{S} \left(\sum_{s=1}^S \frac{\partial p_i^{*(s)}}{\partial \tau} \right). \end{aligned} \quad (5.31)$$

Since $\frac{\partial p_i^{*(s)}}{\partial \tau} = \left(\frac{1}{2\sqrt{\tau}} \right) p_i^{*(s)} (1 - p_i^{*(s)}) u_i^{(s)}$,

$$\frac{\partial E(Y_i)}{\partial \tau} \doteq \frac{m_i}{2\sqrt{\tau}} \sum_{s=1}^S \frac{p_i^{*(s)} (1 - p_i^{*(s)}) u_i^{(s)}}{S}. \quad (5.32)$$

Next,

$$\begin{aligned}
 \frac{\partial V(Y_i)}{\partial \beta} &\doteq \frac{\partial}{\partial \beta} \left\{ \frac{m_i}{S} \sum_{s=1}^S p_i^{*(s)} (1 - p_i^{*(s)}) + m_i^2 \sum_{s=1}^S \frac{\{p_i^{*(s)} - E(p_i^*)\}^2}{S-1} \right\} \\
 &= \frac{m_i}{S} \sum_{s=1}^S (1 - 2p_i^{*(s)}) \frac{\partial p_i^{*(s)}}{\partial \beta} \\
 &\quad + \frac{2m_i^2}{(S-1)} \sum_{s=1}^S [p_i^{*(s)} - E(p_i^*)] \left[\frac{\partial p_i^{*(s)}}{\partial \beta} - \frac{\partial E(p_i^*)}{\partial \beta} \right] \quad (5.33)
 \end{aligned}$$

where $\frac{\partial p_i^{*(s)}}{\partial \beta} = p_i^{*(s)} (1 - p_i^{*(s)}) \mathbf{x}_i$ and $\frac{\partial E(p_i^*)}{\partial \beta} \doteq \frac{1}{S} \sum_{s=1}^S p_i^{*(s)} (1 - p_i^{*(s)}) \mathbf{x}_i$.

Chapter 6

Simulation Study

In this chapter, we study the finite-sample properties of the quasi-likelihood (QL) methods for estimation and hypothesis testing discussed earlier. We ran a series of simulations and investigated the empirical level of significance and power of the quasi-score (QS) test for testing overdispersion in binomial regression. We investigated the biases and mean squared errors of the estimators of regression and overdispersion parameters for both correctly specified and misspecified random effects distributions. We also study the empirical coverage probabilities of the confidence intervals for both regression and overdispersion parameters in the simulations.

In the simulations, the data were generated from the mixed binomial model:

$$\begin{aligned} y_i | z_i &\sim \text{ind. Binom}(m_i, p_i^*), \\ \text{logit}(p_i^*) &= \beta_0 + \beta_1 x_i + z_i, \quad i = 1, \dots, n, \\ z_i &\sim \text{ind. } N(0, \tau). \end{aligned}$$

We chose $m_i = m$, for simplicity. The values of the regression parameters were fixed at $(\beta_0, \beta_1) = (-0.5, 0.5)$. The sample sizes were chosen as $m = 10$ and $n = (50, 100)$. We chose $\tau = 0$ when investigating the empirical level of significance of the QS test for testing overdispersion in the binomial regression model. Empirical power of the test is calculated for $\tau = (0.01, 0.025, 0.05, 0.1, 0.25, 0.5)$.

6.1 Testing for Overdispersion

For each combination of n and τ , we generate $k = 5000$ sets of data from the given mixed binomial model. For each set of data, we apply the QS test for testing $H_0 : \tau = 0$ vs $H_1 : \tau > 0$. We report the empirical level and power of this test under both correctly specified random effects (that is, $Z_i \sim \text{ind. } N(0, \tau)$) and misspecified random effects ($Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{ind. gamma}(\lambda, 1)$).

Table 6.1 presents empirical level and power of the QS test for both correctly specified and misspecified random effects distributions.

Table 6.1: Empirical level and power of the quasi-score test for both correctly specified and misspecified random effects distributions.

Empirical Level and Power					
n	τ	$Z_i \sim N(0, \tau)$	$Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right); a_i \sim \text{gamma}(\lambda, 1)$		
			$\lambda = 0.5$	$\lambda = 1.0$	$\lambda = 2.0$
50	0.000	0.0406	0.0406	0.0406	0.0406
	0.010	0.0496	0.0506	0.0492	0.0512
	0.025	0.0646	0.0746	0.0710	0.0722
	0.050	0.1028	0.1218	0.1204	0.1178
	0.100	0.2138	0.2520	0.2502	0.2420
	0.250	0.6176	0.5922	0.6278	0.6314
	0.500	0.9374	0.8760	0.9160	0.9264
100	0.000	0.0470	0.0470	0.0470	0.0470
	0.010	0.0564	0.0640	0.0656	0.0614
	0.025	0.0836	0.1006	0.1000	0.0908
	0.050	0.1672	0.1826	0.1860	0.1756
	0.100	0.3650	0.3826	0.3998	0.3922
	0.250	0.8798	0.8274	0.8608	0.8826
	0.500	0.9972	0.9842	0.9930	0.9968

It is apparent from Table 6.1 that the power increases as τ increases, as expected. Also, the power increases as the number of clusters (n) increases from 50 to 100. The empirical significance level from the QS test is close to 0.05 and gets closer to it as n increases. When $\tau = (0.25, 0.50)$ the power increases as λ increases when the random effects are misspecified. It is clear that the power of the QS test is not much affected by the misspecified random effects. This is due to the fact that, the mean and variance of the random effects (obtained from the gamma distribution) are still the same as that of the normally distributed random effects. As a result, the test is not quite affected by this misspecification.

6.2 Parameter Estimation

The model parameters were estimated for τ is 0, 0.1, 0.25, 0.5 and 0.8, and sample sizes $n = (50, 100)$. For each combination of n and τ , we generate $k = 2000$ sets of data from the given mixed binomial model. For each set of data, we apply the two approaches for computing the marginal mean and variance of the observations:

- i) SSA method, and
- ii) Simulation method,

for estimating the regression and overdispersion parameters. We report the biases and root MSEs of the estimators, and coverage probabilities and mean lengths of confidence intervals for model parameters under both correctly specified random effects (that is, $Z_i \sim \text{ind. } N(0, \tau)$) and misspecified random effects ($Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{ind. gamma}(\lambda, 1)$).

6.2.1 Bias

The bias of an estimator $\hat{\theta}$ of a parameter θ is the difference between the expected value of $\hat{\theta}$ and θ :

$$\begin{aligned} \text{Bias}(\hat{\theta}) &= E(\hat{\theta}) - \theta \\ &\cong \left(\frac{1}{k} \sum_{k=1}^k \hat{\theta}^{(k)} \right) - \theta, \end{aligned}$$

where $\hat{\theta}^{(k)}$ is the estimator obtained for θ in the k^{th} data set or simulation. Table 6.2 presents biases of the estimators for correctly specified random effects.

Table 6.2: Biases of the estimators when the random effects are correctly specified.

Random Effects: $Z_i \sim N(0, \tau)$							
Bias							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	-0.00039	0.00055	-0.01634	0.00115	-0.00107	0.03789
	0.10	0.00635	-0.00952	-0.02757	-0.00211	-0.00124	-0.00281
	0.25	0.02113	-0.02449	-0.04940	-0.00197	-0.00212	-0.02104
	0.50	0.04135	-0.04483	-0.11264	-0.00348	-0.00138	-0.02969
	0.80	0.06278	-0.06773	-0.22243	-0.00428	-0.00292	-0.03463
100	0.00	0.00044	-0.00097	-0.01003	0.00151	-0.00201	0.02906
	0.10	0.00719	-0.01157	-0.01437	-0.00264	-0.00178	-0.00261
	0.25	0.02273	-0.02707	-0.03307	-0.00234	-0.00238	-0.00810
	0.50	0.04547	-0.04745	-0.09723	-0.00105	-0.00155	-0.01336
	0.80	0.06883	-0.06802	-0.20711	0.00031	-0.00001	-0.01604

It is apparent from Table 6.2 that biases of the estimators for all parameters β_0, β_1 and τ calculated under the SSA method is larger than the simulation method when overdispersion is present. If data do not contain extra variation, then the SSA method provides estimates with lower biases. Furthermore, the bias of the estimator of τ from the SSA method increases at a higher rate than the simulation method when τ increases.

Tables 6.3 - 6.5 present biases of the estimators for misspecified random effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$, where $a_i \sim \text{gamma}(\lambda, 1)$ with $\lambda = (0.5, 1, 2)$. We observe that even when the random effects are misspecified, the estimates of the overdispersion

parameter from the simulation method has lower biases than that of the SSA method. Furthermore, similar to the correct specification of random effects, the bias of the estimator of τ from the SSA method increases at a higher rate than the simulation method when τ increases. The biases of estimators of τ from the SSA method decrease as the shape parameter λ increases. We do not observe this pattern in the simulation method, as the biases of estimators of τ do not show a pattern when λ increases.

Table 6.3: Biases of the estimators when the random effects are misspecified - gamma (0.5, 1).

Random Effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{gamma}(\lambda, 1)$							
Bias: when $\lambda = 0.5$							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	-0.00039	0.00055	-0.01634	0.00115	-0.00107	0.03789
	0.10	0.00117	-0.01124	-0.02000	-0.00854	-0.00221	0.00719
	0.25	0.00295	-0.02552	-0.04330	-0.02191	-0.00210	-0.00518
	0.50	-0.00122	-0.04411	-0.12526	-0.04868	-0.00149	-0.02750
	0.80	-0.00870	-0.05927	-0.26566	-0.08060	0.00211	-0.07033
100	0.00	0.00044	-0.00097	-0.01003	0.00151	-0.00201	0.02906
	0.10	0.00411	-0.00949	-0.00961	-0.00632	0.00098	0.00403
	0.25	0.00292	-0.02337	-0.03951	-0.02224	0.00102	-0.00955
	0.50	0.00125	-0.04141	-0.12809	-0.04572	0.00198	-0.04090
	0.80	-0.00700	-0.05833	-0.27253	-0.07801	0.00365	-0.09549

Table 6.4: Biases of the estimators when the random effects are misspecified - gamma (1, 1).

Random Effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{gamma}(\lambda, 1)$							
Bias: when $\lambda = 1$							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	-0.00039	0.00055	-0.01634	0.00115	-0.00107	0.03789
	0.10	0.00339	-0.00774	-0.01639	-0.00651	0.00201	0.01015
	0.25	0.00992	-0.02248	-0.02969	-0.01620	0.00296	0.00965
	0.50	0.00984	-0.04431	-0.09615	-0.04048	0.00195	0.01098
	0.80	0.00798	-0.06273	-0.21598	-0.06846	0.00449	0.00420
100	0.00	0.00044	-0.00097	-0.01003	0.00151	-0.00201	0.02906
	0.10	0.00720	-0.00703	-0.00705	-0.00388	0.00376	0.00648
	0.25	0.01207	-0.02264	-0.02746	-0.01475	0.00314	0.00340
	0.50	0.01367	-0.04305	-0.09738	-0.03709	0.00382	-0.00123
	0.80	0.01279	-0.06380	-0.22153	-0.06399	0.00391	-0.01749

Table 6.5: Biases of the estimators when the random effects are misspecified - gamma (2, 1).

Random Effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{gamma}(\lambda, 1)$							
Bias: when $\lambda = 2$							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	-0.00039	0.00055	-0.01634	0.00115	-0.00107	0.03789
	0.10	0.00599	-0.00764	-0.02009	-0.00335	0.00154	0.00533
	0.25	0.01176	-0.02249	-0.03688	-0.01348	0.00174	-0.00243
	0.50	0.02038	-0.04303	-0.09453	-0.02877	0.00309	0.00573
	0.80	0.02540	-0.06327	-0.20161	-0.04988	0.00519	0.01825
100	0.00	0.00044	-0.00097	-0.01003	0.00151	-0.00201	0.02906
	0.10	0.00733	-0.00965	-0.01206	-0.00284	0.00046	0.00010
	0.25	0.01516	-0.02452	-0.02747	-0.01111	0.00111	0.00136
	0.50	0.02334	-0.04691	-0.08631	-0.02735	0.00090	0.01065
	0.80	0.02944	-0.06860	-0.19527	-0.04789	0.00163	0.02064

6.2.2 Root Mean Squared Error

The mean squared error (MSE) of an estimator $\hat{\theta}$ of a parameter θ is given by

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 \\ &\cong \frac{1}{k} \sum_{k=1}^k \left(\hat{\theta}^{(k)} - \theta \right)^2, \end{aligned}$$

where $\hat{\theta}^{(k)}$ is the estimator obtained for θ in the k^{th} data set or simulation. Table 6.6 presents root MSEs of the estimators for correctly specified random effects.

Table 6.6: Root MSEs of the estimators when the random effects are correctly specified.

Random Effects: $Z_i \sim N(0, \tau)$							
Root Mean Squared Error							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	0.65171	0.65378	0.41867	0.65046	0.65263	0.41341
	0.10	0.67677	0.62120	0.42748	0.68435	0.62808	0.42789
	0.25	0.71504	0.58644	0.45947	0.73467	0.60293	0.47396
	0.50	0.79564	0.56527	0.54057	0.83372	0.59151	0.59940
	0.80	0.90954	0.59472	0.65803	0.96629	0.62098	0.80192
100	0.00	0.64909	0.64723	0.41247	0.64822	0.64648	0.41035
	0.10	0.66993	0.61844	0.42460	0.67805	0.62584	0.42579
	0.25	0.70628	0.58265	0.45702	0.72695	0.60007	0.46942
	0.50	0.78236	0.55978	0.53944	0.82065	0.58604	0.59243
	0.80	0.89392	0.59053	0.65767	0.94963	0.61604	0.79121

It is apparent from Table 6.6 that SSA method provides lower root MSEs than the simulation method for all cases, except when $\tau = 0$. It is clear from the results that root MSEs decrease as the number of clusters increase from 50 to 100 in both SSA and simulation methods. When τ increases, the root MSEs corresponding to all

the estimators are also increasing in both methods.

Tables 6.7 - 6.9 present the root MSEs of the estimators for misspecified random effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$, where $a_i \sim \text{gamma}(\lambda, 1)$ with $\lambda = (0.5, 1, 2)$. Even when the random effects are misspecified, SSA method provides lower root MSEs than the simulation method for all cases. Furthermore, root MSEs decrease as the number of clusters increase from 50 to 100 in both SSA and simulation methods. When $n = 50$, the root MSEs corresponding to estimators of τ increase as λ increases from 0.5 to 1, but decrease as λ increases from 1 to 2. When $n = 100$ the root MSEs corresponding to estimators of τ increase as λ increases.

Table 6.7: Root MSEs of the estimators when the random effects are misspecified - gamma (0.5, 1).

Random Effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{gamma}(\lambda, 1)$							
Root Mean Squared Error: when $\lambda = 0.5$							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	0.65171	0.65378	0.41867	0.65046	0.65263	0.41341
	0.10	0.68099	0.62424	0.42926	0.68870	0.63106	0.43079
	0.25	0.72925	0.59136	0.46670	0.74964	0.60816	0.48983
	0.50	0.82970	0.57003	0.54751	0.86929	0.59524	0.62829
	0.80	0.96502	0.60100	0.65347	1.02551	0.62575	0.81859
100	0.00	0.64909	0.64723	0.41247	0.64822	0.64648	0.41035
	0.10	0.67395	0.62001	0.42071	0.68241	0.62822	0.42298
	0.25	0.72339	0.58518	0.45273	0.74407	0.60278	0.46740
	0.50	0.81960	0.56224	0.52728	0.85869	0.58748	0.58268
	0.80	0.95561	0.59215	0.62768	1.01432	0.61572	0.74586

Table 6.8: Root MSEs of the estimators when the random effects are misspecified - gamma (1, 1).

Random Effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{gamma}(\lambda, 1)$							
Root Mean Squared Error: when $\lambda = 1$							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	0.65171	0.65378	0.41867	0.65046	0.65263	0.41341
	0.10	0.67984	0.62693	0.43023	0.68787	0.63445	0.43237
	0.25	0.72357	0.59242	0.47031	0.74524	0.61101	0.49448
	0.50	0.81890	0.56882	0.55915	0.86168	0.59668	0.64683
	0.80	0.95054	0.59726	0.67467	1.01545	0.62468	0.86116
100	0.00	0.64909	0.64723	0.41247	0.64822	0.64648	0.41035
	0.10	0.67344	0.61974	0.42115	0.68240	0.62807	0.42326
	0.25	0.71831	0.58410	0.45571	0.74035	0.60239	0.47224
	0.50	0.81151	0.55839	0.54001	0.85390	0.58524	0.60529
	0.80	0.94204	0.58452	0.65222	1.00613	0.60949	0.79971

Table 6.9: Root MSEs of the estimators when the random effects are misspecified - gamma (2, 1).

Random Effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{gamma}(\lambda, 1)$							
Root Mean Squared Error: when $\lambda = 2$							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	0.65171	0.65378	0.41867	0.65046	0.65263	0.41341
	0.10	0.67719	0.62287	0.42744	0.68437	0.62981	0.42993
	0.25	0.72108	0.58957	0.46685	0.74162	0.60671	0.48730
	0.50	0.80987	0.56392	0.55375	0.85073	0.59105	0.63195
	0.80	0.93700	0.59260	0.67537	0.99993	0.61996	0.85922
100	0.00	0.64909	0.64723	0.41247	0.64822	0.64648	0.41035
	0.10	0.67202	0.62046	0.42274	0.68016	0.62816	0.42459
	0.25	0.71417	0.58606	0.45928	0.73569	0.60419	0.47447
	0.50	0.80201	0.56302	0.54768	0.84418	0.59078	0.61315
	0.80	0.92787	0.59224	0.66752	0.99216	0.61912	0.82615

6.2.3 Coverage Probability and Mean Length

Table 6.10 and 6.11 present the coverage probabilities and mean lengths of 95% confidence intervals corresponding to the model parameters for correctly specified random effects, respectively.

Table 6.10: Coverage probabilities of the parameters when the random effects are correctly specified.

Random Effects: $Z_i \sim N(0, \tau)$							
Coverage Probability							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	0.93500	0.92500	0.89850	0.93550	0.92900	0.99750
	0.10	0.93450	0.92750	0.89000	0.93800	0.93450	0.99900
	0.25	0.93700	0.92350	0.86750	0.94450	0.94000	0.88200
	0.50	0.92300	0.91600	0.81800	0.94000	0.93700	0.88700
	0.80	0.91850	0.89900	0.71050	0.94100	0.94500	0.89100
100	0.00	0.94450	0.93700	0.92150	0.94950	0.94300	0.99500
	0.10	0.94300	0.95250	0.91350	0.94900	0.95650	0.99350
	0.25	0.94200	0.93250	0.89500	0.94600	0.94950	0.91700
	0.50	0.91300	0.90750	0.80800	0.94000	0.94750	0.92150
	0.80	0.88850	0.88100	0.59150	0.94300	0.95050	0.91950

It is apparent from Table 6.10 that simulation method provides better coverage probabilities for all the parameters. In fact, the simulation method maintains a reasonable level of coverage probability for τ while they are very low under the SSA method specially when τ is increasing.

Table 6.11: Mean lengths of the confidence intervals when the random effects are correctly specified.

Random Effects: $Z_i \sim N(0, \tau)$							
Mean Length							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	0.37168	0.35068	0.35473	0.37950	0.35604	0.37141
	0.10	0.40376	0.37435	0.41569	0.41697	0.38609	0.45318
	0.25	0.44606	0.40953	0.49525	0.47288	0.43402	0.59451
	0.50	0.50150	0.45839	0.60043	0.55658	0.50750	0.84728
	0.80	0.55220	0.50269	0.69503	0.64386	0.58419	1.16653
100	0.00	0.26105	0.25685	0.25370	0.26576	0.25992	0.26144
	0.10	0.28497	0.27823	0.29830	0.29398	0.28648	0.32386
	0.25	0.31467	0.30606	0.35537	0.33401	0.32431	0.42931
	0.50	0.35245	0.34009	0.42740	0.39142	0.37674	0.60796
	0.80	0.38731	0.37177	0.49034	0.45198	0.43249	0.83080

It is clear from Table 6.11 that mean lengths of the confidence intervals computed using the simulation method are wider than that of the SSA method. Furthermore, in both methods the mean lengths of the confidence intervals for β_0, β_1 and τ get wider as τ increases. In SSA method, even if the mean lengths of the confidence intervals for τ are increasing as τ increases, the corresponding coverage probabilities decrease as τ increases. In simulation method, when n is increasing from 50 to 100, the coverage probabilities for τ are increasing even when the mean lengths are decreasing.

Tables 6.12, 6.14 and 6.16 present the coverage probabilities of 95% confidence intervals corresponding to the model parameters for misspecified random effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$, where $a_i \sim \text{gamma}(\lambda, 1)$ with $\lambda = (0.5, 1, 2)$. Similarly, Tables 6.13, 6.15 and 6.17 present the mean lengths of 95% confidence intervals corresponding to the

model parameters for misspecified random effects.

Table 6.12: Coverage probabilities of the parameters when the random effects are misspecified - gamma (0.5, 1).

Random Effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{gamma}(\lambda, 1)$							
Coverage Probability: when $\lambda = 0.5$							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	0.93500	0.92500	0.89850	0.93550	0.92900	0.99750
	0.10	0.94350	0.91700	0.88000	0.94850	0.92050	0.99800
	0.25	0.94600	0.91250	0.85100	0.94100	0.92650	0.86500
	0.50	0.94100	0.90550	0.78750	0.92050	0.92750	0.85150
	0.80	0.94000	0.90050	0.67900	0.89900	0.92550	0.83700
100	0.00	0.94450	0.93700	0.92150	0.94950	0.94300	0.99500
	0.10	0.94350	0.93450	0.90800	0.94250	0.93900	0.99350
	0.25	0.94200	0.92950	0.86600	0.93350	0.93800	0.88700
	0.50	0.94000	0.91850	0.74350	0.90850	0.94100	0.86750
	0.80	0.93800	0.89600	0.50250	0.87100	0.94750	0.83850

Table 6.13: Mean lengths of the confidence intervals when the random effects are misspecified - gamma (0.5, 1).

Random Effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{gamma}(\lambda, 1)$							
Mean Length: when $\lambda = 0.5$							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	0.37168	0.35068	0.35473	0.37950	0.35604	0.37141
	0.10	0.40579	0.37748	0.44210	0.41555	0.38936	0.48992
	0.25	0.44564	0.41243	0.55653	0.46169	0.43715	0.69293
	0.50	0.49371	0.45531	0.68780	0.52450	0.50115	1.01023
	0.80	0.53556	0.49368	0.79168	0.58611	0.56405	1.37042
100	0.00	0.26105	0.25685	0.25370	0.26576	0.25992	0.26144
	0.10	0.28646	0.27935	0.31572	0.29241	0.28760	0.34727
	0.25	0.31426	0.30491	0.39377	0.32414	0.32174	0.48573
	0.50	0.34831	0.33720	0.48174	0.36726	0.36818	0.69968
	0.80	0.37839	0.36592	0.55236	0.40995	0.41409	0.93934

Table 6.14: Coverage probabilities of the parameters when the random effects are misspecified - gamma (1, 1).

Random Effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{gamma}(\lambda, 1)$							
Coverage Probability: when $\lambda = 1$							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	0.93500	0.92500	0.89850	0.93550	0.92900	0.99750
	0.10	0.94000	0.92450	0.89150	0.94450	0.92850	0.99600
	0.25	0.93900	0.92400	0.87550	0.94300	0.93000	0.89200
	0.50	0.93950	0.90950	0.81900	0.92700	0.93350	0.88650
	0.80	0.93550	0.89500	0.74750	0.90500	0.93500	0.88700
100	0.00	0.94450	0.93700	0.92150	0.94950	0.94300	0.99500
	0.10	0.94050	0.93150	0.91050	0.94350	0.93650	0.99150
	0.25	0.93750	0.93000	0.89200	0.94100	0.93800	0.91000
	0.50	0.94750	0.90600	0.81550	0.92950	0.93950	0.91150
	0.80	0.94450	0.88750	0.62350	0.90950	0.94550	0.90800

Table 6.15: Mean lengths of the confidence intervals when the random effects are misspecified - gamma (1, 1).

Random Effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{gamma}(\lambda, 1)$							
Mean Length: when $\lambda = 1$							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	0.37168	0.35068	0.35473	0.37950	0.35604	0.37141
	0.10	0.40727	0.38118	0.43253	0.41871	0.39361	0.48011
	0.25	0.45041	0.41732	0.54586	0.47027	0.44326	0.68541
	0.50	0.50284	0.46480	0.67538	0.54270	0.51544	1.01208
	0.80	0.54969	0.50621	0.77770	0.61746	0.58733	1.38529
100	0.00	0.26105	0.25685	0.25370	0.26576	0.25992	0.26144
	0.10	0.28700	0.27954	0.31241	0.29410	0.28811	0.34369
	0.25	0.31662	0.30678	0.38663	0.32954	0.32503	0.47915
	0.50	0.35383	0.34153	0.47372	0.37989	0.37683	0.69877
	0.80	0.38681	0.37303	0.54222	0.43083	0.42945	0.95195

Table 6.16: Coverage probabilities of the parameters when the random effects are misspecified - gamma (2, 1).

Random Effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{gamma}(\lambda, 1)$							
Coverage Probability: when $\lambda = 2$							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	0.93500	0.92500	0.89850	0.93550	0.92900	0.99750
	0.10	0.94400	0.93400	0.89000	0.94450	0.94050	0.99800
	0.25	0.94200	0.92450	0.88400	0.94650	0.92850	0.89700
	0.50	0.94050	0.90600	0.84100	0.94100	0.93550	0.89400
	0.80	0.94100	0.90050	0.75000	0.91950	0.93350	0.89350
100	0.00	0.94450	0.93700	0.92150	0.94950	0.94300	0.99500
	0.10	0.94300	0.92500	0.91600	0.94000	0.93250	0.99250
	0.25	0.95000	0.92000	0.89550	0.94350	0.93200	0.91500
	0.50	0.93800	0.89600	0.83550	0.93700	0.93450	0.92400
	0.80	0.93700	0.87050	0.64850	0.92700	0.93750	0.93150

Table 6.17: Mean lengths of the confidence intervals when the random effects are misspecified - gamma (2, 1).

Random Effects: $Z_i = \sqrt{\tau} \left(\frac{a_i - \lambda}{\sqrt{\lambda}} \right)$ with $a_i \sim \text{gamma}(\lambda, 1)$							
Mean Length: when $\lambda = 2$							
		SSA			Simulation		
n	τ	β_0	β_1	τ	β_0	β_1	τ
50	0.00	0.37168	0.35068	0.35473	0.37950	0.35604	0.37141
	0.10	0.40644	0.38018	0.42467	0.41887	0.39248	0.46834
	0.25	0.44940	0.41619	0.52495	0.47174	0.44235	0.64733
	0.50	0.50487	0.46355	0.64924	0.55052	0.51569	0.95478
	0.80	0.55485	0.50732	0.75137	0.63296	0.59305	1.33001
100	0.00	0.26105	0.25685	0.25370	0.26576	0.25992	0.26144
	0.10	0.28566	0.27808	0.30478	0.29314	0.28620	0.33242
	0.25	0.31634	0.30579	0.37564	0.33102	0.32417	0.46112
	0.50	0.35555	0.34253	0.46049	0.38609	0.37970	0.67690
	0.80	0.39077	0.37530	0.53009	0.44334	0.43640	0.93904

Coverage probabilities from the simulation method are lower for τ in the case of misspecified random effects than that for correctly specified random effects, except when the shape parameter of the gamma distribution is 2. The SSA method provides larger coverage probabilities when the random effects have a gamma distribution with shape parameter (1, 2) than that for a normal distribution. The simulation method maintains a reasonable level of coverage probability for τ even when the random effects are misspecified. Similar to the situation where random effects are correctly specified, in SSA method, the coverage probabilities for τ decrease as τ increases. In the simulation method, when n increases from 50 to 100, the coverage probabilities for τ increase.

As a summary of the results obtained from the simulation study, we note that the power of the QS test increases when τ increases as well as when the number of clusters (n) increases from 50 to 100. The empirical significance level from this test is close to 0.05 and gets closer to it as n increases. It was also clear that the power of the QS test is not quite affected by the misspecified random effects considered here.

When estimating the model parameters, the simulation method gives estimators with smaller biases than the SSA method, especially for the overdispersion parameter. A similar pattern can be observed in the case of misspecified random effects. Furthermore, the bias of the estimator of the overdispersion parameter from the SSA method increases at a higher rate when the value of τ increases.

When the root MSEs are compared, the SSA method always give lower values

whenever overdispersion is present. Also, the root MSEs decrease as the number of clusters increase in both SSA and simulation methods, for correctly specified and misspecified random effects.

The simulation method provides better coverage probabilities for the overdispersion parameter, but it also has wider mean lengths. In the SSA method, the coverage probabilities decrease as τ increases.

Chapter 7

Conclusion and Discussion

A study carried out by Dean (1992)[6] for testing overdispersion in binomial regression models comprised the basis of the research carried out in this thesis. The purpose was to supplement this paper with a numerical contribution and to explore further by investigating the possibility of using a semi-parametric approach for testing overdispersion as an alternative to the parametric approach employed by Dean (1992)[6]. Robustness of the test is investigated even when the distributional assumptions of the random effects are misspecified. Also, subsequent to any evidence of an extra variation in the data set, the overdispersion parameter together with the regression parameters are estimated.

We demonstrated that testing for extra-binomial variation using the score test of Dean (1992)[6] is identical to the quasi-score (QS) test statistic considered in this thesis under the null hypothesis of $\tau = 0$. As it gives the same power as the score test,

one may consider using the test statistic based on the QS approach as an alternative to using the score test when testing for overdispersion in a particular data set. The power of the QS test increases when τ increases as well as when the number of clusters (n) increases. The empirical significance level from this test is close to 0.05 and gets closer to it as n increases.

Even when the random effects are distributed differently from a symmetric distribution, the power of the QS test is not affected by the misspecification. This is due to the fact that, the mean and variance of the random effects (obtained using a gamma distribution) are still the same as that of the normally distributed random effects. As a result, this test is not quite as affected from this misspecification.

When estimating the model parameters, the simulation method appears to give estimators with smaller biases than the small sigma approximation (SSA) method, specially for the overdispersion parameter. We also observe similar pattern in the case of misspecified random effects. Furthermore, the bias of the estimator of the overdispersion parameter from the SSA method increases at a higher rate when the value of τ increases. This is due to the fact that the approximation method is developed under small τ , and hence the name “small sigma approximation”. But when the root mean squared errors (MSEs) are compared, the SSA method always give lower values whenever overdispersion is present. Also, the root MSEs decrease as the number of clusters increase in both SSA and simulation methods, for correctly and misspecified random effects. Similarly, the simulation method provides better

coverage probabilities for the overdispersion parameter, but it also has wider mean lengths.

In this thesis, we have studied different methods for inference in binomial regression. As a future research, we may wish to extend these methods for analyzing overdispersed Poisson or count data.

Appendix A

Derivation of the Score Test Statistic

A.1 General Format

The density function of a larger overdispersed natural exponential family is

$$f_{Y_i}(y_i; m_i, p_i^*) = f(y_i | \theta_i^*) = \exp \{a_i(\theta_i^*)y_i - g_i(\theta_i^*) + c_i(y_i)\}. \quad (\text{A.1})$$

According to Williams (1982)[26] and others the extra-binomial variation of a logistic model is of the form:

$$\theta_i^* = \log \left(\frac{p_i^*}{1 - p_i^*} \right) = \mathbf{x}_i^t \boldsymbol{\beta} + z_i = \theta_i + z_i, \quad (\text{A.2})$$

where $\mathbf{x}_i$ is a vector of covariates for the i^{th} observation and $\boldsymbol{\beta}$ is a vector of fixed effect regression parameters. Z_i 's are *iid* random variates with $E(Z_i) = 0$ and $V(Z_i) = \tau b_i(\theta_i) > 0$. θ_i^* 's are continuous independent random variates with finite mean and variance, $E(\theta_i^*) = \theta_i(\mathbf{x}_i; \boldsymbol{\beta})$ and $V(\theta_i^*) = \tau b_i(\theta_i) > 0$.

The probability density function of Y_i in mixed models specified as above is

$$f_M(Y_i) = E_*\{f(Y_i; \theta_i^*)\},$$

where E_* denotes expectation over the distribution of θ_i^* .

Following the work of Cox (1983)[3] and Chesher (1984)[2], Dean (1992)[6] expanded $f(Y_i; \theta_i^*)$ in a Taylor series about θ_i and took the expectation to obtain

$$\begin{aligned} f_M(Y_i) &= E_*\{f(Y_i; \theta_i^*)\} \\ &= E_* \left\{ f(Y_i; \theta_i) + \sum_{r=2}^{\infty} \frac{\partial^{(r)}}{\partial \theta_i^{*(r)}} f(Y_i; \theta_i^*) \Big|_{\theta_i^*=\theta_i} \frac{(\theta_i^* - \theta_i)^r}{r!} \right\} \\ &= f(Y_i; \theta_i) + \sum_{r=2}^{\infty} \frac{E(\theta_i^* - \theta_i)^r}{r!} \frac{\partial^{(r)}}{\partial \theta_i^{*(r)}} f(Y_i; \theta_i^*) \Big|_{\theta_i^*=\theta_i} \\ &= f(Y_i; \theta_i) \left\{ 1 + \sum_{r=2}^{\infty} \frac{\alpha_r}{r!} D_r(Y_i; \theta_i) \right\}, \end{aligned} \quad (\text{A.3})$$

where it is assumed that

$$E\{(\theta_i^* - \theta_i)^r\} = \alpha_r; \quad \alpha_r = o(\tau), \quad r \geq 3, \quad (\text{A.4})$$

and

$$D_r(Y_i; \theta_i) = \left\{ \frac{\partial^{(r)}}{\partial \theta_i^{*(r)}} f(Y_i; \theta_i^*) \Big|_{\theta_i^*=\theta_i} \right\} \{f(Y_i; \theta_i)\}^{-1}. \quad (\text{A.5})$$

So the contribution to the log-likelihood function from the i^{th} observation is

$$\begin{aligned} l_i &= \log f_M(Y_i) \\ &= \log \left\{ f(Y_i; \theta_i) \left[1 + \sum_{r=2}^{\infty} \frac{\alpha_r}{r!} D_r(Y_i; \theta_i) \right] \right\} \\ &= \log f(Y_i; \theta_i) + \log \left\{ 1 + \sum_{r=2}^{\infty} \frac{\alpha_r}{r!} D_r(Y_i; \theta_i) \right\}. \end{aligned} \quad (\text{A.6})$$

For $\sum_{r=2}^{\infty} \frac{\alpha_r}{r!} D_r(Y_i; \theta_i)$ it is sufficient to consider only the $D_2(Y_i; \theta_i)$ term as all the other terms are 0 as $\alpha_r = o(\tau)$ for $r \geq 3$. That is

$$\begin{aligned} \sum_{r=2}^{\infty} \frac{\alpha_r}{r!} D_r(Y_i; \theta_i) &= \frac{\alpha_2}{2!} D_2(Y_i; \theta_i) + \frac{\alpha_3}{3!} D_3(Y_i; \theta_i) + \frac{\alpha_4}{4!} D_4(Y_i; \theta_i) + \dots \\ &= \frac{\alpha_2}{2!} D_2(Y_i; \theta_i) + 0 + 0 + \dots; \quad \text{as } \alpha_r = o(\tau) \text{ for } r \geq 3. \\ &= \frac{\alpha_2}{2!} D_2(Y_i; \theta_i). \end{aligned} \tag{A.7}$$

From (A.4),

$$\begin{aligned} r = 1 &\Rightarrow \alpha_1 = E\{(\theta_i^* - \theta_i)^1\} = 0 \quad \text{and} \\ r = 2 &\Rightarrow \alpha_2 = E\{(\theta_i^* - \theta_i)^2\} = E\{[\theta_i^* - E(\theta_i^*)]^2\} = V(\theta_i^*) = \tau b_i(\theta_i). \end{aligned} \tag{A.8}$$

Using (A.1),

$$\begin{aligned} r = 1 &\Rightarrow \frac{\partial^{(1)}}{\partial \theta_i^{*(1)}} f(Y_i; \theta_i^*) = \exp\{a_i(\theta_i^*)y_i - g_i(\theta_i^*) + c_i(y_i)\} \{a'_i(\theta_i^*)y_i - g'_i(\theta_i^*)\} \\ &= f(Y_i; \theta_i^*) \{a'_i(\theta_i^*)y_i - g'_i(\theta_i^*)\} \quad \text{and} \\ r = 2 &\Rightarrow \frac{\partial^{(2)}}{\partial \theta_i^{*(2)}} f(Y_i; \theta_i^*) = \frac{\partial}{\partial \theta_i^*} \{f(Y_i; \theta_i^*)[a'_i(\theta_i^*)y_i - g'_i(\theta_i^*)]\} \\ &= f(Y_i; \theta_i^*) \{a''_i(\theta_i^*)y_i - g''_i(\theta_i^*)\} \\ &\quad + \{a'_i(\theta_i^*)y_i - g'_i(\theta_i^*)\}^2 f(Y_i; \theta_i^*). \end{aligned} \tag{A.9}$$

Therefore by (A.9),

$$\begin{aligned}
 D_2(Y_i; \theta_i) &= \left\{ \frac{\partial^{(2)}}{\partial \theta_i^{*(2)}} f(Y_i; \theta_i^*) \Big|_{\theta_i^* = \theta_i} \right\} \{f(Y_i; \theta_i)\}^{-1} \\
 &= f(Y_i; \theta_i^*) \{a_i''(\theta_i^*) y_i - g_i''(\theta_i^*)\} \Big|_{\theta_i^* = \theta_i} \{f(Y_i; \theta_i)\}^{-1} \\
 &\quad + \{a_i'(\theta_i^*) y_i - g_i'(\theta_i^*)\}^2 f(Y_i; \theta_i^*) \Big|_{\theta_i^* = \theta_i} \{f(Y_i; \theta_i)\}^{-1} \\
 &= \{a_i''(\theta_i) y_i - g_i''(\theta_i)\} + \{a_i'(\theta_i) y_i - g_i'(\theta_i)\}^2. \tag{A.10}
 \end{aligned}$$

Then

$$\begin{aligned}
 \sum_{r=2}^{\infty} \frac{\alpha_r}{r!} D_r(Y_i; \theta_i) &= \frac{\alpha_2}{2!} D_2(Y_i; \theta_i) \\
 &= \frac{\tau b_i(\theta_i)}{2!} \left\{ [a_i''(\theta_i) y_i - g_i''(\theta_i)] + [a_i'(\theta_i) y_i - g_i'(\theta_i)]^2 \right\}. \tag{A.11}
 \end{aligned}$$

Therefore the contribution to the log-likelihood function from the i^{th} observation is

$$\begin{aligned}
 l_i &= \log f(Y_i; \theta_i) + \log \left\{ 1 + \sum_{r=2}^{\infty} \frac{\alpha_r}{r!} D_r(Y_i; \theta_i) \right\} \\
 &= \log f(Y_i; \theta_i) + \log \left\{ 1 + \frac{\tau b_i(\theta_i)}{2!} \left\{ [a_i''(\theta_i) y_i - g_i''(\theta_i)] + [a_i'(\theta_i) y_i - g_i'(\theta_i)]^2 \right\} \right\}. \tag{A.12}
 \end{aligned}$$

Assumptions are made on conditions for application of the central limit theorem to score components and the regularity conditions of maximum likelihood (ML) theory.

The score test for testing the hypothesis $\tau = 0$ is based on

$$\sum_{i=1}^n \frac{\partial l_i}{\partial \tau} \Big|_{\tau=0} = \sum_{i=1}^n T_i(\hat{\theta}_i), \tag{A.13}$$

where $(\hat{\theta}_i)$ is the ML estimate of θ_i when $\tau = 0$. The parameter θ_i is a function of the covariates and the regression parameters. The estimators of the regression

parameters are computed using the “*glm*” function in *R* programming software and thereby estimator of θ_i are computed.

To find the format of $T_i(\theta_i)$, consider

$$\begin{aligned} \frac{\partial l_i}{\partial \tau} &= \frac{\partial}{\partial \tau} \log f(Y_i; \theta_i) \\ &\quad + \frac{\partial}{\partial \tau} \log \left\{ 1 + \frac{\tau b_i(\theta_i)}{2!} \{a_i''(\theta_i)y_i - g_i''(\theta_i)\} + \{a_i'(\theta_i)y_i - g_i'(\theta_i)\}^2 \right\} \\ &= 0 + \left\{ \frac{\frac{b_i(\theta_i)}{2!} \{[a_i''(\theta_i)y_i - g_i''(\theta_i)] + [a_i'(\theta_i)y_i - g_i'(\theta_i)]^2\}}{1 + \frac{\tau b_i(\theta_i)}{2!} \{[a_i''(\theta_i)y_i - g_i''(\theta_i)] + [a_i'(\theta_i)y_i - g_i'(\theta_i)]^2\}} \right\} \end{aligned} \quad (\text{A.14})$$

$$\left. \frac{\partial l_i}{\partial \tau} \right|_{\tau=0} = \frac{b_i(\theta_i)}{2!} \{[a_i''(\theta_i)y_i - g_i''(\theta_i)] + [a_i'(\theta_i)y_i - g_i'(\theta_i)]^2\}. \quad (\text{A.15})$$

In Chapter 2, for a GLM $E(Y_i) = \mu_i(\theta_i) = \frac{g_i'(\theta_i)}{a_i'(\theta_i)}$ (refer (2.9)). Therefore

$$\left. \frac{\partial l_i}{\partial \tau} \right|_{\tau=0} = \frac{b_i(\theta_i)}{2!} [a_i'(\theta_i)]^2 \left\{ [a_i'(\theta_i)]^{-2} [a_i''(\theta_i)y_i - g_i''(\theta_i)] + \{y_i - [a_i'(\theta_i)]^{-1} g_i'(\theta_i)\}^2 \right\}.$$

From this point the dependence on θ_i of the functions $\mu_i(\theta_i)$, $a_i(\theta_i)$, $b_i(\theta_i)$ and $g_i(\theta_i)$ will be suppressed for simplicity of notation. Then

$$\begin{aligned} \left. \frac{\partial l_i}{\partial \tau} \right|_{\tau=0} &= \frac{b_i}{2} (a_i')^2 \left\{ (a_i')^{-2} (a_i'' y_i - g_i'') + [y_i - (a_i')^{-1} g_i']^2 \right\} \\ &= \frac{b_i}{2} (a_i')^2 \left\{ (a_i')^{-2} (a_i'' y_i - g_i'') + (y_i - \mu_i)^2 \right\} \\ &= \frac{b_i}{2} (a_i')^2 \left\{ (y_i - \mu_i)^2 - (a_i')^{-2} (g_i'' - a_i'' y_i) \right\}. \end{aligned} \quad (\text{A.16})$$

Therefore

$$T_i(\theta_i) = \frac{b_i}{2} (a_i')^2 \left\{ (y_i - \mu_i)^2 - (a_i')^{-2} (g_i'' - a_i'' y_i) \right\}. \quad (\text{A.17})$$

Note that, $T_i(\hat{\theta}_i)$ compares two estimates of $V(Y_i)$, one based on the sample variance and the other derived from the form of the variance (2.12) in the base model (2.1).

To obtain an asymptotic variance for (A.13) it is necessary to find the elements of the Fisher Information matrix. The first four moments of Y_i , evaluated at $\tau = 0$ has to be calculated. Denote this matrix by $I = I(\beta, \tau)$ and let I be partitioned as

$$I(\beta, \tau) = \begin{pmatrix} I_{\beta\beta} & I_{\beta\tau} \\ I_{\beta\tau}^t & I_{\tau\tau} \end{pmatrix}, \quad (\text{A.18})$$

where $I_{\beta\beta}$ and $I_{\tau\tau}$ are $p \times p$ and 1×1 matrices respectively. Let $h_i = h_i(\theta_i) = \log a'_i(\theta_i)$, U be an $n \times p$ matrix with $(ir)^{th}$ - element $\frac{\partial \theta_i}{\partial \beta_r}$, $\mathbf{1}$ be an $n \times 1$ unit vector, and W_1 and W_2 be diagonal matrices. The i^{th} diagonal element of W_1 would be

$$W_{1i} = E \left\{ -\frac{\partial^2 l_i}{\partial \theta_i^2} \right\} \Big|_{\tau=0}. \quad (\text{A.19})$$

Using (A.12) and (A.17) we know

$$l_i = \log f(Y_i; \theta_i) + \log \{1 + \tau T_i(\theta_i)\}. \quad (\text{A.20})$$

Therefore

$$\begin{aligned} \frac{\partial l_i}{\partial \theta_i} &= \frac{f(Y_i; \theta_i) \{a'_i(\theta_i)y_i - g'_i(\theta_i)\}}{f(Y_i; \theta_i)} + \frac{\tau T'_i(\theta_i)}{1 + \tau T_i(\theta_i)} \\ &= \{a'_i y_i - g'_i\} + \frac{\tau T'_i(\theta_i)}{1 + \tau T_i(\theta_i)}. \end{aligned} \quad (\text{A.21})$$

Using (A.21),

$$\begin{aligned} \frac{\partial^2 l_i}{\partial \theta_i^2} &= \{a''_i y_i - g''_i\} + \frac{\{1 + \tau T_i(\theta_i)\} \tau T''_i(\theta_i) - \tau T'_i(\theta_i) \tau T'_i(\theta_i)}{\{1 + \tau T_i(\theta_i)\}^2} \\ \frac{\partial^2 l_i}{\partial \theta_i^2} \Big|_{\tau=0} &= a''_i y_i - g''_i. \end{aligned} \quad (\text{A.22})$$

Therefore

$$\begin{aligned}
 W_{1i} &= E \left\{ -\frac{\partial^2 l_i}{\partial \theta_i^2} \right\} \Big|_{\tau=0} = E(-a_i'' Y_i + g_i'') = g_i'' - a_i'' E(Y_i) \\
 &= g_i'' - a_i'' \frac{g_i'}{a_i'}; \quad \text{where } E(Y_i) = \frac{g_i'}{a_i'} \\
 &= g_i'' - g_i' \frac{a_i''}{a_i'} \\
 &= g_i'' - g_i' h_i'; \quad \text{where } h_i = \log a_i' \Rightarrow h_i' = \frac{a_i''}{a_i'}.
 \end{aligned} \tag{A.23}$$

Similarly, the i^{th} diagonal element of W_2 would be

$$W_{2i} = E \left\{ -\frac{\partial^2 l_i}{\partial \theta_i \partial \tau} \right\} \Big|_{\tau=0}. \tag{A.24}$$

Then

$$\begin{aligned}
 \frac{\partial^2 l_i}{\partial \theta_i \partial \tau} &= \frac{\partial}{\partial \tau} \left\{ (a_i' y_i - g_i') + \frac{\tau T_i'(\theta_i)}{1 + \tau T_i(\theta_i)} \right\} \\
 &= 0 + \frac{\{1 + \tau T_i(\theta_i)\} T_i'(\theta_i) - \tau T_i'(\theta_i) T_i(\theta_i)}{\{1 + \tau T_i(\theta_i)\}^2} \\
 &= \frac{T_i'(\theta_i)}{\{1 + \tau T_i(\theta_i)\}^2} \\
 \frac{\partial^2 l_i}{\partial \theta_i \partial \tau} \Big|_{\tau=0} &= T_i'(\theta_i).
 \end{aligned} \tag{A.25}$$

To find $T_i'(\theta_i)$ and hence $-T_i'(\theta_i)$ consider

$$\begin{aligned}
 T_i'(\theta_i) &= \frac{\partial}{\partial \theta_i} \left\{ \frac{b_i}{2} (a_i')^2 \{ (y_i - \mu_i)^2 - (a_i')^{-2} (g_i'' - a_i'' y_i) \} \right\} \\
 &= \frac{\partial}{\partial \theta_i} \left\{ \frac{b_i}{2} \{ (a_i'' y_i - g_i'') + (a_i' y_i - g_i')^2 \} \right\}; \quad \text{where } \mu_i = \frac{g_i'}{a_i'} \\
 -T_i'(\theta_i) &= \frac{b_i}{2} \{ (g_i''' - a_i''' y_i) - 2(a_i' y_i - g_i')(a_i'' y_i - g_i'') \} \\
 &\quad + \frac{b_i'}{2} \{ (g_i'' - a_i'' y_i) - (a_i' y_i - g_i')^2 \}.
 \end{aligned} \tag{A.26}$$

Therefore

$$\begin{aligned}
E[-T'_i(\theta_i)] &= \frac{b_i}{2} \{g_i''' - a_i''' E(Y_i) + 2[a_i'' g_i' E(Y_i) - g_i' g_i'' - a_i' a_i'' E(Y_i^2) + a_i' g_i'' E(Y_i)]\} \\
&\quad + \frac{b_i'}{2} \{g_i'' - a_i'' E(Y_i) - a_i'^2 E(Y_i^2) + 2a_i' g_i' E(Y_i) - g_i'^2\} \\
&= \frac{b_i}{2} \left[g_i''' - a_i''' \frac{g_i'}{a_i'} \right] \\
&\quad + b_i \left\{ a_i'' g_i' \frac{g_i'}{a_i'} - g_i' g_i'' - a_i' a_i'' \left[\frac{g_i'' - a_i'' \frac{g_i'}{a_i'}}{\{a_i'\}^2} + \left(\frac{g_i'}{a_i'} \right)^2 \right] + a_i' g_i'' \frac{g_i'}{a_i'} \right\} \\
&\quad + \frac{b_i'}{2} \left\{ g_i'' - a_i'' \frac{g_i'}{a_i'} - a_i'^2 \left[\frac{g_i'' - a_i'' \frac{g_i'}{a_i'}}{\{a_i'\}^2} + \left(\frac{g_i'}{a_i'} \right)^2 \right] + 2a_i' g_i' \frac{g_i'}{a_i'} - g_i'^2 \right\} \\
&= \frac{b_i}{2} \left[g_i''' - a_i''' \frac{g_i'}{a_i'} \right] + b_i \left\{ -\frac{a_i'' g_i''}{a_i'} + \left(\frac{a_i''}{a_i'} \right)^2 g_i' \right\} + 0. \tag{A.27}
\end{aligned}$$

When computing (A.27), $E(Y_i) = \frac{g_i'}{a_i'}$ and $V(Y_i) = \frac{\{g_i'' - a_i'' \frac{g_i'}{a_i'}\}}{(a_i')^2}$. Let $h_i = \log a_i' \Rightarrow h_i' = \frac{a_i''}{a_i'} \Rightarrow h_i'' = \left(\frac{a_i'''}{a_i'} \right) - \left(\frac{a_i''}{a_i'} \right)^2 = \left(\frac{a_i'''}{a_i'} \right) - (h_i')^2$. Then

$$\begin{aligned}
E[-T'_i(\theta_i)] &= \frac{b_i}{2} \{g_i''' - g_i' [h_i'' + (h_i')^2]\} + b_i \{-g_i'' h_i' + (h_i')^2 g_i'\} \\
&= \frac{b_i}{2} \{g_i' [(h_i')^2 - h_i''] - 2g_i'' h_i' + g_i'''\} . \tag{A.28}
\end{aligned}$$

Therefore

$$W_{2i} = E[-T'_i(\theta_i)] = \frac{b_i}{2} \{g_i' [(h_i')^2 - h_i''] - 2g_i'' h_i' + g_i'''\} . \tag{A.29}$$

Derivations of the elements of Fisher Information matrix given in (A.18):

The term $I_{\tau\tau}$:

$$I_{\tau\tau} = \sum_{i=1}^n E \left\{ -\frac{\partial^2 l_i}{\partial \tau^2} \right\} \Big|_{\tau=0} . \tag{A.30}$$

Consider

$$\begin{aligned}\frac{\partial l_i}{\partial \tau} &= \frac{\partial}{\partial \tau} \{ \log f(Y_i; \theta_i) + \log[1 + \tau T_i(\theta_i)] \} = \frac{T_i(\theta_i)}{1 + \tau T_i(\theta_i)} \\ \frac{\partial^2 l_i}{\partial \tau^2} &= \frac{\partial}{\partial \tau} \left\{ \frac{T_i(\theta_i)}{1 + \tau T_i(\theta_i)} \right\} \\ \frac{\partial^2 l_i}{\partial \tau^2} \Big|_{\tau=0} &= -[T_i(\theta_i)]^2 \\ -\frac{\partial^2 l_i}{\partial \tau^2} \Big|_{\tau=0} &= [T_i(\theta_i)]^2 = \frac{b_i^2}{4} (a_i')^4 \{ (y_i - \mu_i)^2 - (a_i')^{-2} (g_i'' - a_i'' y_i) \}^2\end{aligned}\quad (\text{A.31})$$

$$\begin{aligned}E \left\{ -\frac{\partial^2 l_i}{\partial \tau^2} \right\} \Big|_{\tau=0} &= \frac{b_i^2}{4} [g_i' \{ 5h_i' h_i'' - 3(h_i')^3 - h_i''' \} + 2\{ h_i' g_i' - g_i'' \}^2 \\ &\quad + g_i'' \{ 6(h_i')^2 - 4h_i'' \} - 4g_i''' h_i' + g_i''''].\end{aligned}\quad (\text{A.32})$$

The equation (A.32) is taken from Dean's (1992)[6] paper. Then

$$\begin{aligned}I_{\tau\tau} &= \sum_{i=1}^n E \left\{ -\frac{\partial^2 l_i}{\partial \tau^2} \right\} \Big|_{\tau=0} \\ &= \sum_{i=1}^n \frac{b_i^2}{4} [g_i' \{ 5h_i' h_i'' - 3(h_i')^3 - h_i''' \} + 2\{ h_i' g_i' - g_i'' \}^2 \\ &\quad + g_i'' \{ 6(h_i')^2 - 4h_i'' \} - 4g_i''' h_i' + g_i''''].\end{aligned}\quad (\text{A.33})$$

The term $I_{\beta\beta}$:

$$I_{\beta\beta} = \sum_{i=1}^n E \left\{ -\frac{\partial^2 l_i}{\partial \beta \partial \beta^t} \right\} \Big|_{\tau=0}.\quad (\text{A.34})$$

Consider

$$\begin{aligned}\frac{\partial l_i}{\partial \beta} &= \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \beta} \\ &= \left\{ a_i' y_i - g_i' + \frac{\tau T_i'(\theta_i)}{1 + \tau T_i(\theta_i)} \right\} \frac{\partial \theta_i}{\partial \beta}.\end{aligned}\quad (\text{A.35})$$

Then

$$\begin{aligned}
\frac{\partial^2 l_i}{\partial \beta \partial \beta^t} &= \frac{\partial}{\partial \beta} \left\{ \left[a'_i y_i - g'_i + \frac{\tau T'_i(\theta_i)}{1 + \tau T_i(\theta_i)} \right] \frac{\partial \theta_i}{\partial \beta^t} \right\} \\
&= \frac{\partial}{\partial \beta} \left\{ a'_i y_i - g'_i + \frac{\tau T'_i(\theta_i)}{1 + \tau T_i(\theta_i)} \right\} \frac{\partial \theta_i}{\partial \beta^t} \\
&\quad + \left\{ a'_i y_i - g'_i + \frac{\tau T'_i(\theta_i)}{1 + \tau T_i(\theta_i)} \right\} \frac{\partial^2 \theta_i}{\partial \beta \partial \beta^t} \\
&= \frac{\partial}{\partial \theta_i} \left\{ a'_i y_i - g'_i + \frac{\tau T'_i(\theta_i)}{1 + \tau T_i(\theta_i)} \right\} \frac{\partial \theta_i}{\partial \beta} \frac{\partial \theta_i}{\partial \beta^t} \\
&\quad + \left\{ a'_i y_i - g'_i + \frac{\tau T'_i(\theta_i)}{1 + \tau T_i(\theta_i)} \right\} \frac{\partial^2 \theta_i}{\partial \beta \partial \beta^t} \\
&= \left\{ a''_i y_i - g''_i + \frac{\{1 + \tau T_i(\theta_i)\} \tau T''_i(\theta_i) + \tau T'_i(\theta_i) \tau T'_i(\theta_i)}{\{1 + \tau T_i(\theta_i)\}^2} \right\} \frac{\partial \theta_i}{\partial \beta} \frac{\partial \theta_i}{\partial \beta^t} \\
&\quad + \left\{ a'_i y_i - g'_i + \frac{\tau T'_i(\theta_i)}{1 + \tau T_i(\theta_i)} \right\} \frac{\partial^2 \theta_i}{\partial \beta \partial \beta^t} \\
\left. \frac{\partial^2 l_i}{\partial \beta \partial \beta^t} \right|_{\tau=0} &= \frac{\partial \theta_i}{\partial \beta} \{a''_i y_i - g''_i\} \frac{\partial \theta_i}{\partial \beta^t} + \{a'_i y_i - g'_i\} \frac{\partial^2 \theta_i}{\partial \beta \partial \beta^t}. \tag{A.36}
\end{aligned}$$

Therefore

$$\begin{aligned}
I_{\beta\beta} &= \sum_{i=1}^n E \left\{ -\frac{\partial^2 l_i}{\partial \beta \partial \beta^t} \right\} \Big|_{\tau=0} \\
&= \sum_{i=1}^n E \left\{ -\frac{\partial \theta_i}{\partial \beta} \{a_i'' Y_i - g_i''\} \frac{\partial \theta_i}{\partial \beta^t} - \{a_i' Y_i - g_i'\} \frac{\partial^2 \theta_i}{\partial \beta \partial \beta^t} \right\} \\
&= \sum_{i=1}^n \left\{ \frac{\partial \theta_i}{\partial \beta} \{g_i'' - a_i'' E(Y_i)\} \frac{\partial \theta_i}{\partial \beta^t} + \{g_i' - a_i' E(Y_i)\} \frac{\partial^2 \theta_i}{\partial \beta \partial \beta^t} \right\} \\
&= \sum_{i=1}^n \left\{ \frac{\partial \theta_i}{\partial \beta} \left(g_i'' - a_i'' \frac{g_i'}{a_i'} \right) \frac{\partial \theta_i}{\partial \beta^t} + \left(g_i' - a_i' \frac{g_i'}{a_i'} \right) \frac{\partial^2 \theta_i}{\partial \beta \partial \beta^t} \right\}; \quad \text{since } E(Y_i) = \frac{g_i'}{a_i'} \\
&= \sum_{i=1}^n \left\{ \frac{\partial \theta_i}{\partial \beta} (g_i'' - g_i' h_i') \frac{\partial \theta_i}{\partial \beta^t} + 0 \right\}; \quad \text{since } h_i' = \frac{a_i''}{a_i'} \\
&= \sum_{i=1}^n \left\{ \frac{\partial \theta_i}{\partial \beta} W_{1i} \frac{\partial \theta_i}{\partial \beta^t} \right\}; \quad \text{since } W_{1i} = (g_i'' - g_i' h_i') \\
&= U^t W_1 U.
\end{aligned} \tag{A.37}$$

The term $I_{\beta\tau}$:

$$I_{\beta\tau} = \sum_{i=1}^n E \left\{ -\frac{\partial^2 l_i}{\partial \beta \partial \tau} \right\} \Big|_{\tau=0}. \tag{A.38}$$

Consider

$$\begin{aligned}
\frac{\partial^2 l_i}{\partial \beta \partial \tau} &= \frac{\partial}{\partial \tau} \left(\frac{\partial l_i}{\partial \beta} \right) = \frac{\partial}{\partial \tau} \left\{ \left[a_i' y_i - g_i' + \frac{\tau T_i'(\theta_i)}{1 + \tau T_i(\theta_i)} \right] \frac{\partial \theta_i}{\partial \beta} \right\} \\
&= \left[a_i' y_i - g_i' + \frac{\tau T_i'(\theta_i)}{1 + \tau T_i(\theta_i)} \right] \frac{\partial^2 \theta_i}{\partial \beta \partial \tau} + \frac{\partial}{\partial \tau} \left[a_i' y_i - g_i' + \frac{\tau T_i'(\theta_i)}{1 + \tau T_i(\theta_i)} \right] \frac{\partial \theta_i}{\partial \beta} \\
&= \left[a_i' y_i - g_i' + \frac{\tau T_i'(\theta_i)}{1 + \tau T_i(\theta_i)} \right] \frac{\partial^2 \theta_i}{\partial \beta \partial \tau} \\
&\quad + \left\{ \frac{[1 + \tau T_i(\theta_i)] T_i'(\theta_i) + \tau T_i'(\theta_i) T_i(\theta_i)}{[1 + \tau T_i(\theta_i)]^2} \right\} \frac{\partial \theta_i}{\partial \beta} \\
\frac{\partial^2 l_i}{\partial \beta \partial \tau} \Big|_{\tau=0} &= (a_i' y_i - g_i') \frac{\partial^2 \theta_i}{\partial \beta \partial \tau} + T_i'(\theta_i) \frac{\partial \theta_i}{\partial \beta}.
\end{aligned} \tag{A.39}$$

Therefore

$$\begin{aligned}
I_{\beta\tau} &= \sum_{i=1}^n E \left\{ -\frac{\partial^2 l_i}{\partial \beta \partial \tau} \right\} \Big|_{\tau=0} \\
&= \sum_{i=1}^n E \left\{ -(a'_i Y_i - g'_i) \frac{\partial^2 \theta_i}{\partial \beta \partial \tau} - T'_i(\theta_i) \frac{\partial \theta_i}{\partial \beta} \right\} \\
&= \sum_{i=1}^n E \left\{ (g'_i - a'_i Y_i) \frac{\partial^2 \theta_i}{\partial \beta \partial \tau} - T'_i(\theta_i) \frac{\partial \theta_i}{\partial \beta} \right\} \\
&= \sum_{i=1}^n \left\{ [g'_i - a'_i E(Y_i)] \frac{\partial^2 \theta_i}{\partial \beta \partial \tau} + \frac{\partial \theta_i}{\partial \beta} E[-T'_i(\theta_i)] \right\} \\
&= \sum_{i=1}^n \left\{ \left[g'_i - a'_i \frac{g'_i}{a'_i} \right] \frac{\partial^2 \theta_i}{\partial \beta \partial \tau} + \frac{\partial \theta_i}{\partial \beta} E[-T'_i(\theta_i)] \right\}; \quad \text{since } E(Y_i) = \frac{g'_i}{a'_i} \\
&= \sum_{i=1}^n \left\{ \frac{\partial \theta_i}{\partial \beta} W_{2i} \right\}; \quad \text{since } W_{2i} = E[-T'_i(\theta_i)] \\
&= U^t W_2 \mathbf{1}.
\end{aligned} \tag{A.40}$$

When developing the score statistic for testing overdispersion, the regression parameters act as nuisance parameters and therefore an adjustment has to be done. Therefore let

$$T = \frac{\left(\sum_{i=1}^n \frac{\partial l_i}{\partial \tau} \right) - I_{\tau\beta} I_{\beta\beta}^{-1} \left(\sum_{i=1}^n \frac{\partial l_i}{\partial \beta} \right)}{\sqrt{V \left\{ \left(\sum_{i=1}^n \frac{\partial l_i}{\partial \tau} \right) - I_{\tau\beta} I_{\beta\beta}^{-1} \left(\sum_{i=1}^n \frac{\partial l_i}{\partial \beta} \right) \right\}}} \Big|_{\tau=0, \beta=\hat{\beta}}. \tag{A.41}$$

The $\sum_{i=1}^n \frac{\partial l_i}{\partial \beta}$ term in the numerator of the score test statistic becomes zero when evaluated at $\beta = \hat{\beta}$. Next, the form of the variance of the score function needs to be determined. The functions $\sum_{i=1}^n \frac{\partial l_i}{\partial \beta}$ and $\sum_{i=1}^n \frac{\partial l_i}{\partial \tau}$ are in the linear form with coefficients $-I_{\tau\beta} I_{\beta\beta}^{-1}$ and 1 respectively. Let $V(\text{Score}) = V \left\{ \left(\sum_{i=1}^n \frac{\partial l_i}{\partial \tau} \right) - I_{\tau\beta} I_{\beta\beta}^{-1} \left(\sum_{i=1}^n \frac{\partial l_i}{\partial \beta} \right) \right\}$

and $\gamma = (\beta^t, \tau)^t$. Using the sandwich type variance formula

$$\begin{aligned}
 V(\text{Score}) &= \begin{pmatrix} -I_{\tau\beta}I_{\beta\beta}^{-1} & 1 \end{pmatrix} V\left(\frac{\partial l}{\partial \gamma}\right) \begin{pmatrix} -I_{\tau\beta}I_{\beta\beta}^{-1} \\ 1 \end{pmatrix} \\
 &= \begin{pmatrix} -I_{\tau\beta}I_{\beta\beta}^{-1} & 1 \end{pmatrix} E\left(-\frac{\partial^2 l}{\partial \gamma^2}\right) \begin{pmatrix} -I_{\tau\beta}I_{\beta\beta}^{-1} \\ 1 \end{pmatrix} \\
 &= \begin{pmatrix} -I_{\tau\beta}I_{\beta\beta}^{-1} & 1 \end{pmatrix} I(\beta, \tau) \begin{pmatrix} -I_{\tau\beta}I_{\beta\beta}^{-1} \\ 1 \end{pmatrix} \\
 &= I_{\tau\tau} - I_{\beta\tau}^t I_{\beta\beta}^{-1} I_{\beta\tau}
 \end{aligned}$$

The numerator in the score test statistic is asymptotically normally distributed when evaluated at $\tau = 0$. Therefore by the central limit theorem

$$\sum_{i=1}^n \frac{\partial l_i}{\partial \tau} \Big|_{\tau=0, \beta=\hat{\beta}} \sim N(0, V(\text{Score})).$$

Thereby

$$T = \frac{\sum_{i=1}^n \frac{\partial l_i}{\partial \tau}}{\sqrt{I_{\tau\tau} - I_{\beta\tau}^t I_{\beta\beta}^{-1} I_{\beta\tau}}} \Big|_{\tau=0, \beta=\hat{\beta}} \sim N(0, 1).$$

A.2 Logit-Normal/Binomial

It is necessary to obtain the formats of the functions a_i , g_i and h_i which are used in equations $T_i(\theta_i)$, W_{1i} , W_{2i} , $I_{\beta\beta}$, $I_{\beta\tau}$ and $I_{\tau\tau}$. The functions a_i , g_i and h_i are from the binomial distribution which is a member of the exponential family.

The format of the binomial distribution:

$$f_{Y_i}(y_i; m_i, p_i) = f(y_i; \theta_i) = \exp \left\{ y_i \theta_i - m_i \log(1 + e^{\theta_i}) + \log \binom{m_i}{y_i} \right\}.$$

It is in the format given in (2.1) where

$$a_i(\theta_i) = \theta_i = \log \left(\frac{p_i}{1-p_i} \right),$$

$$g_i(\theta_i) = m_i \log(1 + e^{\theta_i}), \text{ and}$$

$$c_i(y_i) = \log \binom{m_i}{y_i}.$$

In this case θ_i is the natural parameter of the binomial distribution as $a_i(\theta_i) = \theta_i$. The first, second, third and fourth derivatives of the above functions need to be calculated in order to calculate $T_i(\theta_i)$, W_{1i} , W_{2i} , $I_{\beta\beta}$, $I_{\beta\tau}$ and $I_{\tau\tau}$.

Therefore

$$a_i = \theta_i \Rightarrow a'_i = 1 \Rightarrow a''_i = 0$$

$$h_i = \log a'_i = \log 1 = 0$$

$$g_i = m_i \log(1 + e^{\theta_i})$$

$$g'_i = m_i \left(\frac{e^{\theta_i}}{1 + e^{\theta_i}} \right) = m_i p_i; \quad \text{as } \theta_i = \log \left(\frac{p_i}{1 - p_i} \right) \Rightarrow p_i = \frac{e^{\theta_i}}{1 + e^{\theta_i}}$$

$$\mu_i = \frac{g'_i}{a'_i} = m_i p_i$$

$$g''_i = m_i \left[\frac{(1 + e^{\theta_i})e^{\theta_i} - e^{2\theta_i}}{(1 + e^{\theta_i})^2} \right] = m_i \frac{e^{\theta_i}}{(1 + e^{\theta_i})^2} = m_i p_i(1 - p_i)$$

$$\begin{aligned} g'''_i &= m_i \left[\frac{(1 + e^{\theta_i})^2 e^{\theta_i} - e^{\theta_i} 2(1 + e^{\theta_i})e^{\theta_i}}{(1 + e^{\theta_i})^4} \right] = m_i \left(\frac{e^{\theta_i}}{1 + e^{\theta_i}} \right) \left[\frac{1 - e^{\theta_i}}{(1 + e^{\theta_i})^2} \right] \\ &= m_i p_i \left[1 - \left(\frac{p_i}{1 - p_i} \right) \right] (1 - p_i)^2 = m_i p_i(1 - p_i)(1 - 2p_i) \end{aligned}$$

$$\begin{aligned} g''''_i &= m_i \left[\frac{(1 + e^{\theta_i})^3(e^{\theta_i} - 2e^{2\theta_i}) - e^{\theta_i}(1 - e^{\theta_i})3(1 + e^{\theta_i})^2 e^{\theta_i}}{(1 + e^{\theta_i})^6} \right] \\ &= m_i \left(\frac{e^{\theta_i}}{1 + e^{\theta_i}} \right) \left[\frac{1 - 4e^{\theta_i} + e^{2\theta_i}}{(1 + e^{\theta_i})^3} \right] \\ &= m_i p_i(1 - p_i)^3 \left[1 - 4 \left(\frac{p_i}{1 - p_i} \right) + \left(\frac{p_i}{1 - p_i} \right)^2 \right] = m_i p_i(1 - p_i)(1 - 6p_i + 6p_i^2) \end{aligned}$$

$$\theta_i = \mathbf{x}_i^t \boldsymbol{\beta} \Rightarrow \frac{\partial \theta_i}{\partial \beta_r} = x_{ir}; \quad \text{where } i = 1, \dots, n; r = 1, \dots, p. \quad (\text{A.42})$$

Then using (A.17),

$$\begin{aligned}
 T_i(\theta_i) &= \frac{b_i}{2} (a'_i)^2 \{ (y_i - \mu_i)^2 - (a'_i)^{-2} (g''_i - a''_i y_i) \} \\
 &= \frac{b_i}{2} (1)^2 \{ (y_i - m_i p_i)^2 - (1)^{-2} (m_i p_i (1 - p_i) - 0) \} \\
 &= \frac{b_i}{2} \{ (y_i - m_i p_i)^2 - m_i p_i (1 - p_i) \} \\
 &= \frac{b_i}{2} \{ (y_i - E(Y_i))^2 - V(Y_i) \}, \tag{A.43}
 \end{aligned}$$

where $E(Y_i) = m_i p_i$ and $V(Y_i) = m_i p_i (1 - p_i)$ are the mean and variance functions of the original binomial distribution. The function b_i depends on the distribution assigned for the random effects.

Next from (A.23),

$$W_{1i} = g''_i - g'_i h'_i = m_i p_i (1 - p_i) - (m_i p_i \times 0) = m_i p_i (1 - p_i). \tag{A.44}$$

Similarly from (A.29),

$$\begin{aligned}
 W_{2i} &= \frac{b_i}{2} \{ g'_i [(h'_i)^2 - h''_i] - 2g''_i h'_i + g'''_i \} \\
 &= \frac{b_i}{2} \{ m_i p_i [0] - 2m_i p_i (1 - p_i) \times 0 + m_i p_i (1 - p_i) (1 - 2p_i) \} \\
 &= \frac{b_i}{2} m_i p_i (1 - p_i) (1 - 2p_i). \tag{A.45}
 \end{aligned}$$

As the next step, items in the asymptotic variance function need to be calculated, starting with the $I_{\tau\tau}$ function computed under $\tau = 0$. Following (A.31),

$$E \left\{ -\frac{\partial^2 l_i}{\partial \tau^2} \right\} \bigg|_{\tau=0} = E \left\{ \frac{b_i^2}{4} (a'_i)^4 \{ (y_i - \mu_i)^2 - (a'_i)^{-2} (g''_i - a''_i y_i) \}^2 \right\}. \tag{A.46}$$

Following (A.42),

$$\begin{aligned}
E \left\{ -\frac{\partial^2 l_i}{\partial \tau^2} \right\} \Big|_{\tau=0} &= \frac{b_i^2}{4} E \{ (y_i - \mu_i)^2 - m_i p_i (1 - p_i) \}^2; \quad \text{here } \mu_i = m_i p_i \\
&= \frac{b_i^2}{4} E \{ (y_i - m_i p_i)^4 - 2(y_i - m_i p_i)^2 m_i p_i (1 - p_i) + [m_i p_i (1 - p_i)]^2 \} \\
&= \frac{b_i^2}{4} E \{ y^4 - 4m_i p_i y_i^3 + 6m_i^2 p_i^2 y_i^2 - 3m_i^3 p_i^3 y_i + m_i^4 p_i^4 \\
&\quad - 2m_i p_i (1 - p_i) y_i^2 + 4m_i^2 p_i^2 (1 - p_i) y_i - 4m_i^3 p_i^3 (1 - p_i) \\
&\quad + [m_i p_i (1 - p_i)]^2 \} \\
&= \frac{b_i^2}{4} E \{ y^4 - 4m_i p_i y_i^3 + 2m_i p_i y_i^2 (3m_i p_i - 1 + p_i) \\
&\quad - 4m_i^2 p_i^2 y_i (m_i p_i - 1 + p_i) + m_i^4 p_i^4 - 2m_i^3 p_i^3 (1 - p_i) \\
&\quad + [m_i p_i (1 - p_i)]^2 \} \\
&= \frac{b_i^2}{4} \{ E(Y_i^4) - 4m_i p_i E(Y_i^3) + 2m_i p_i E(Y_i^2)(3m_i p_i - 1 + p_i) \\
&\quad - 4m_i^2 p_i^2 E(Y_i)(m_i p_i - 1 + p_i) + m_i^2 p_i^2 (m_i p_i - 1 + p_i)^2 \}. \quad (\text{A.47})
\end{aligned}$$

As it is apparent, the first four moments of the observations under $\tau = 0$ are needed to compute the above function. When $\tau = 0$ the observations follow a binomial distribution. The moment generating function for data coming from a binomial(m, p) distribution is

$$M_X(t) = [pe^t + (1 - p)]^m. \quad (\text{A.48})$$

The moments are calculated using

$$E(Y^n) = \frac{d^n}{dt^n} M_X(t) \Big|_{t=0}. \quad (\text{A.49})$$

Using a computer software the first four moments of the observations are computed as follows:

$$\begin{aligned}\frac{d^1}{dt^1}M_X(t) &= mpe^t(pe^t - p + 1)^{m-1} \\ E(Y^1) &= \left. \frac{d^1}{dt^1}M_X(t) \right|_{t=0} = mp,\end{aligned}\tag{A.50}$$

$$\begin{aligned}\frac{d^2}{dt^2}M_X(t) &= \frac{d}{dt} \{ mpe^t(pe^t - p + 1)^{m-1} \} \\ &= e^t[m^2p^2e^t + mp(1-p)](pe^t - p + 1)^{m-2} \\ E(Y^2) &= \left. \frac{d^2}{dt^2}M_X(t) \right|_{t=0} = m^2p^2 + mp(1-p),\end{aligned}\tag{A.51}$$

$$\begin{aligned}\frac{d^3}{dt^3}M_X(t) &= \frac{d}{dt} \{ e^t[m^2p^2e^t + mp(1-p)](pe^t - p + 1)^{m-2} \} \\ &= e^t[m^3p^3e^{2t} + mp^2e^t(1-p)(3m-1) + mp(1-p)^2](pe^t - p + 1)^{m-3} \\ E(Y^3) &= \left. \frac{d^3}{dt^3}M_X(t) \right|_{t=0} = mp[m^2p^2 + 3mp(1-p) + (p-1)(2p-1)],\end{aligned}\tag{A.52}$$

$$\begin{aligned}\frac{d^4}{dt^4}M_X(t) &= \frac{d}{dt} \{ e^t[m^3p^3e^{2t} + mp^2e^t(1-p)(3m-1) + mp(1-p)^2](pe^t - p + 1)^{m-3} \} \\ &= e^t\{m^4p^4e^{3t} - mp^3e^{2t}[6m^2(p-1) + 4m(1-p) + p-1] \\ &\quad + mp^2e^t[7m(p-1)^2 - 4p^2 + 8p - 4] - mp[p^3 - 3p^2 + 3p - 1]\} \\ &\quad \times (pe^t - p + 1)^{m-4} \\ E(Y^4) &= \left. \frac{d^4}{dt^4}M_X(t) \right|_{t=0} \\ &= mp[m^3p^3 + 6m^2p^2(1-p) + mp(p-1)(11p-7) - 6p^3 + 12p^2 - 7p + 1].\end{aligned}\tag{A.53}$$

Substituting these in equation (A.47) and using a computer software

$$\begin{aligned}
 E \left\{ -\frac{\partial^2 l_i}{\partial \tau^2} \right\} \Big|_{\tau=0} &= \frac{b_i^2}{4} \{ -m_i p_i [3m_i^3 p_i^3 + 6m_i^2 p_i^2 (1 - p_i) - 3m_i p_i (p_i - 1)^2 \\
 &\quad + 6p_i^3 - 12p_i^2 + 7p_i - 1] \\
 &\quad + m_i^2 p_i^2 [3m_i^2 p_i^2 + 6m_i p_i (1 - p_i) - (p_i - 1)^2] \} \\
 &= \frac{b_i^2}{4} \{ 2[m_i p_i (1 - p_i)]^2 + m_i p_i (1 - p_i) (1 - 6p_i + 6p_i^2) \}. \quad (\text{A.54})
 \end{aligned}$$

Therefore

$$\begin{aligned}
 I_{\tau\tau} &= \sum_{i=1}^n E \left\{ -\frac{\partial^2 l_i}{\partial \tau^2} \right\} \Big|_{\tau=0} \\
 &= \sum_{i=1}^n \frac{b_i^2}{4} \{ 2[m_i p_i (1 - p_i)]^2 + m_i p_i (1 - p_i) (1 - 6p_i + 6p_i^2) \}. \quad (\text{A.55})
 \end{aligned}$$

For the two items $I_{\beta\beta}$ and $I_{\beta\tau}$, they require the term U . As it was stated earlier, U is a $n \times p$ matrix with $(ir)^{th}$ element $\frac{\partial \theta_i}{\partial \beta_r}$ where, $i = 1, \dots, n$; $r = 1, \dots, p$. Since $U_{ir} = \frac{\partial \theta_i}{\partial \beta_r} = x_{ir}$, U becomes X which is a $n \times p$ matrix of x_{ir} terms. As the format of the U matrix, W_{1i} and W_{2i} are known, $I_{\beta\beta}$ and $I_{\beta\tau}$ can be easily derived.

All the necessary elements required to build the score statistic are now derived for a set of data coming from a binomial distribution. The only term still not specified is b_i , which is dependent on the distribution of the random effect. If the random effects are coming from a normal distribution with mean 0 and variance τ , then $b_i = 1$ since $V(\theta_i^*) = \tau b_i = V(Z_i) = \tau$, where θ_i^* is the parameter of the overdispersed exponential family of distributions (in this case, binomial).

Appendix B

Derivatives of the Quasi-Score Functions

The quasi-score (QS) functions considered in this thesis depend on both regression and overdispersion parameters. Therefore let $\gamma = (\beta^t, \tau)^t$. The format of the QS function corresponding to the regression parameters β when the observations are independent:

$$K_{\beta}(\gamma) = \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \{y_i - E(Y_i)\}.$$

The format of the QS function corresponding to the overdispersion parameter τ when the observations are independent:

$$K_{\tau}(\gamma) = \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{[y_i - E(Y_i)]^2\}^{-1} \{[y_i - E(Y_i)]^2 - V(Y_i)\}.$$

Consider the negative expected value $E \left[-\frac{\partial K_{\beta}(\gamma)}{\partial \beta} \right]$, and let $J_{\beta\beta}$ be:

$$\begin{aligned}
 J_{\beta\beta} &= E \left[-\frac{\partial K_{\beta}(\gamma)}{\partial \beta} \right] \\
 &= E \left\{ -\frac{\partial}{\partial \beta} \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \{y_i - E(Y_i)\} \right\} \\
 &= E \left\{ -\sum_{i=1}^n \frac{\partial}{\partial \beta} \left[\left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \right] [y_i - E(Y_i)] \right\} \\
 &\quad + E \left\{ -\sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \frac{\partial}{\partial \beta} [y_i - E(Y_i)] \right\} \\
 &= -\sum_{i=1}^n \frac{\partial}{\partial \beta} \left[\left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \right] E[y_i - E(Y_i)] \\
 &\quad + \sum_{i=1}^n \frac{\partial E(Y_i)}{\partial \beta} [V(Y_i)]^{-1} \left(\frac{\partial E(Y_i)}{\partial \beta} \right)^t \\
 &= \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \left(\frac{\partial E(Y_i)}{\partial \beta} \right)^t. \tag{B.1}
 \end{aligned}$$

When deriving equation (B.1), $\frac{\partial y_i}{\partial \beta} = 0$ and $E[y_i - E(Y_i)] = 0$.

Consider the negative expected value $E \left[-\frac{\partial K_\tau(\gamma)}{\partial \tau} \right]$, and let $J_{\tau\tau}$ be:

$$\begin{aligned}
 J_{\tau\tau} &= E \left[-\frac{\partial K_\tau(\gamma)}{\partial \tau} \right] \\
 &= E \left\{ -\frac{\partial}{\partial \tau} \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \{ [y_i - E(Y_i)]^2 - V(Y_i) \} \right\} \\
 &= E \left\{ -\sum_{i=1}^n \frac{\partial}{\partial \tau} \left[\left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \right] \{ [y_i - E(Y_i)]^2 - V(Y_i) \} \right\} \\
 &\quad + E \left\{ -\sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \frac{\partial}{\partial \tau} \{ [y_i - E(Y_i)]^2 - V(Y_i) \} \right\} \\
 &= -\sum_{i=1}^n \frac{\partial}{\partial \tau} \left[\left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \right] E \{ [y_i - E(Y_i)]^2 - V(Y_i) \} \\
 &\quad - \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} E \left\{ 2[y_i - E(Y_i)] \frac{\partial}{\partial \tau} [y_i - E(Y_i)] - \frac{\partial V(Y_i)}{\partial \tau} \right\} \\
 &= \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} E \left[2[y_i - E(Y_i)] \frac{\partial E(Y_i)}{\partial \tau} + \frac{\partial V(Y_i)}{\partial \tau} \right] \\
 &= \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right)^2 V \{ [y_i - E(Y_i)]^2 \}^{-1}. \tag{B.2}
 \end{aligned}$$

When deriving the above equation, $E \{ [y_i - E(Y_i)]^2 - V(Y_i) \} = 0$, $\frac{\partial y_i}{\partial \tau} = 0$ and $E[y_i - E(Y_i)] = 0$.

Consider the negative expected value $E \left[-\frac{\partial \mathbf{K}_\beta(\gamma)}{\partial \tau} \right]$, and let $\mathbf{J}_{\beta\tau}$ be:

$$\begin{aligned}
 \mathbf{J}_{\beta\tau} &= E \left[-\frac{\partial \mathbf{K}_\beta(\gamma)}{\partial \tau} \right] \\
 &= E \left\{ -\frac{\partial}{\partial \tau} \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \{y_i - E(Y_i)\} \right\} \\
 &= E \left\{ -\sum_{i=1}^n \frac{\partial}{\partial \tau} \left[\left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \right] [y_i - E(Y_i)] \right\} \\
 &\quad + E \left\{ -\sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \frac{\partial}{\partial \tau} [y_i - E(Y_i)] \right\} \\
 &= -\sum_{i=1}^n \frac{\partial}{\partial \tau} \left[\left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \right] E[y_i - E(Y_i)] \\
 &\quad + E \left\{ \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \left(\frac{\partial E(Y_i)}{\partial \tau} \right) \right\} \\
 &= \sum_{i=1}^n \left(\frac{\partial E(Y_i)}{\partial \beta} \right) [V(Y_i)]^{-1} \left(\frac{\partial E(Y_i)}{\partial \tau} \right). \tag{B.3}
 \end{aligned}$$

When deriving the above equation, $E[y_i - E(Y_i)] = 0$ and $\frac{\partial y_i}{\partial \tau} = 0$.

Consider the negative expected value $E \left[-\frac{\partial K_\tau(\gamma)}{\partial \beta} \right]$, and let $J_{\tau\beta}$ be:

$$\begin{aligned}
 J_{\tau\beta} &= E \left[-\frac{\partial K_\tau(\gamma)}{\partial \beta} \right] \\
 &= E \left\{ -\frac{\partial}{\partial \beta} \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \{ [y_i - E(Y_i)]^2 - V(Y_i) \} \right\} \\
 &= E \left\{ -\sum_{i=1}^n \frac{\partial}{\partial \beta} \left\{ \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \right\} \{ [y_i - E(Y_i)]^2 - V(Y_i) \} \right\} \\
 &\quad + E \left\{ -\sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \frac{\partial}{\partial \beta} \{ [y_i - E(Y_i)]^2 - V(Y_i) \} \right\} \\
 &= -\sum_{i=1}^n \frac{\partial}{\partial \beta} \left\{ \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \right\} E \{ [y_i - E(Y_i)]^2 - V(Y_i) \} \\
 &\quad - E \left\{ \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \frac{\partial}{\partial \beta} [[y_i - E(Y_i)]^2 - V(Y_i)] \right\} \\
 &= \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} E \left[2[y_i - E(Y_i)] \frac{\partial E(Y_i)}{\partial \beta} + \frac{\partial V(Y_i)}{\partial \beta} \right] \\
 &= \sum_{i=1}^n \left(\frac{\partial V(Y_i)}{\partial \tau} \right) V \{ [y_i - E(Y_i)]^2 \}^{-1} \left(\frac{\partial V(Y_i)}{\partial \beta} \right). \tag{B.4}
 \end{aligned}$$

When deriving the above equation, $E \{ [y_i - E(Y_i)]^2 - V(Y_i) \} = 0$, $\frac{\partial y_i}{\partial \beta} = 0$ and $E[y_i - E(Y_i)] = 0$.

Appendix C

R Programming Code

```
## Clear any R Objects present in memory.
rm(list=ls())
```

```
#####
#####
##                                ##
#####
#####
```

```
#####
##                                ##
#####
```

```
execute.runs <- function(tau, n, tau.i)
{
  set.seed(030)
```

```
#####
## Actual Values
```

```
## tau    <- ? # Taken from the arguments to this method
## n      <- ? # Taken from the arguments to this method
## tau.i  <- ? # Taken from the arguments to this method
```

```

## S      <- ? # Defined Globally
## no.sim <- ? # Defined Globally

beta      <- c(-.5,.5)
iter      <- 500
m         <- 10

x         <- rnorm(n)
xx        <- cbind(rep(1,n), x)
nom.alpha <- 0.05

## Initial Values
beta.i    <- c(-.5,.5)
gamma.i   <- c(beta.i[1], beta.i[2], tau.i)

#####

output.qf <- NULL
output.sim <- NULL

c.s.s.qf  <- NULL
c.s.s.sim1 <- NULL

for (k in 1:no.sim) {
  ## To have some indication of which trial we are
  ## running during large runs we do the following
  ## print:
  if ( (k %% 100) == 0 ) {
    cat("Running trial #: ",k,"\n")
  }

  y <- y.gen(beta, tau, x, xx, m, n)
  u.s <- rnorm(n=S, mean=0, sd=1)

  if (enable.estimation) {
    quasi.out <- quasi.fit(gamma.i=gamma.i,
                          x=x, xx=xx, y=y, m=m,
                          n=n, iter=iter)
    output.qf <- rbind(quasi.out[nrow(
      quasi.out),], output.qf)

    simulation.out <- simulation.method(
      gamma.i=gamma.i, x=x,
      xx=xx, m=m, n=n, y=y,

```

```

        u.s=u.s, S=S, iter=iter)
    output.sim <- rbind(simulation.out[nrow(
        simulation.out),],
        output.sim)
}

if (enable.power) {
    c.s.s.qf[k] <- Calc.Score.Stat.qf(x=x,
        xx=xx,y=y,m=m, n=n)
    c.s.s.sim1[k] <- Calc.Score.Stat.sim1(
        x=x,xx=xx,y=y,u.s=u.s,
        m=m,n=n,S=S)
}
}

if (enable.estimation) {
    dimnames(output.qf) <- list(NULL,c("Beta0",
        "Beta1", "Tau", "Beta0 SE",
        "Beta1 SE", "Tau SE"))
    dimnames(output.sim) <- list(NULL,c("Beta0",
        "Beta1", "Tau", "Beta0 SE",
        "Beta1 SE", "Tau SE"))

    finaltab <- get.finaltab(output.qf=output.qf,
        output.sim=output.sim,
        nom.alpha=nom.alpha,
        beta=beta, tau=tau)
    cat("\n\n")
    print(round(finaltab, digits=5))
}

if (enable.power) {
    powertab <- get.powertab(c.s.s.qf=c.s.s.qf,
        c.s.s.sim1=c.s.s.sim1,
        alpha=nom.alpha)
    cat("\n\n")
    print(round(powertab, digits=5))
}

if (enable.estimation & enable.power) {
    return(rbind(finaltab, c(powertab[1],
        powertab[2],
        0, 0, 0, 0)))
}

```

```

    } else if (enable.estimation) {
      return(finaltab)
    } else if (enable.power) {
      return(powertab)
    }
  }
}

```

```

#####
#####
#####

```

```

#####
##                                POWER TESTING FUNCTIONS                                ##
#####

```

```

###-----###
## Testing for overdispersion using a Score Test Statistic - ##
## App                                                         ##
###-----###

```

```

Calc.Score.Stat.qf <- function(x,xx,y,m,n)
{

```

```

  beta.p <- get.beta.p(x, m, n, y)

```

```

  stat.num <- 0

```

```

  l.tau <- 0

```

```

  w.1.i <- 0

```

```

  w.2.i <- 0

```

```

  w.1 <- matrix(0, ncol=n, nrow=n)

```

```

  w.2 <- matrix(0, ncol=n, nrow=n)

```

```

  unit.vec <- c(rep(1,n))

```

```

  eta.p <- c(xx %*% beta.p)

```

```

  p.i.p <- exp(eta.p) / (1 + exp(eta.p))

```

```

  mu.i.p <- m * p.i.p

```

```

  y.minus.mu.p <- y - mu.i.p

```

```

  for (j in 1:n) {

```

```

    term1.p <- p.i.p[j] * (1 - p.i.p[j])

```

```

    term2.p <- m * term1.p

```

```

    term3.p <- 1 - (6 * p.i.p[j]) + (6 * p.i.p[j]^2)

```

```

    stat.num.0 <- 0.5 * (y.minus.mu.p[j]^2 - term2.p)

```

```

stat.num <- stat.num + stat.num.0

I.tau.0 <- 0.25 * ((2 * term2.p^2) + (term2.p *
term3.p))
I.tau <- I.tau + I.tau.0

w.1.i[j] <- term2.p
w.2.i[j] <- 0.5 * term2.p * (1 - 2 * p.i.p[j])
}

w.1 <- diag(w.1.i)
w.2 <- diag(w.2.i)
term.v <- t(unit.vec) %*% w.2 %*% xx %*% solve(t(xx) %*%
w.1 %*% xx) %*% t(xx) %*% w.2 %*%
unit.vec

stat.denom <- sqrt(I.tau - term.v)

chi.cal.qf <- stat.num / stat.denom
p.value.qf <- pnorm(chi.cal.qf, 0, 1, lower.tail=FALSE)

return(p.value.qf)
}

#####
## Testing for overdispersion using a Quasi-Score Test      ##
## Statistic - Sim                                           ##
#####
Calc.Score.Stat.sim1 <- function(x,xx,y,u,s,m,n,S)
{
  beta.p      <- get.beta.p(x, m, n, y)
  phil.p      <- 0
  phi2.p      <- 0
  d.phil.d.beta <- 0
  d.phil.d.tau <- 0
  d.phi2.d.beta <- 0
  d.phi2.d.tau <- 0
  phil.p.vec  <- NULL
  phi2.p.vec  <- NULL
  var.cov.matrix <- 0

  eta.p <- c(xx %*% beta.p)
  num.p <- exp(eta.p)

```

```

pi.vec.p <- num.p / (1 + num.p)

mu.i.p <- m * pi.vec.p
y.minus.mu.p <- y - mu.i.p
var.yi.p <- m * pi.vec.p * (1 - pi.vec.p)

for(j in 1:n) {
  term1.p <- 1 - (2 * pi.vec.p[j])
  term2.p <- pi.vec.p[j] - 1
  work.var <- (mu.i.p[j] * (2 * mu.i.p[j] *
    term2.p^2 - (6 * pi.vec.p[j]^3) +
    (12 * pi.vec.p[j]^2) -
    (7 * pi.vec.p[j]) + 1))

  d.mu.d.beta.p <- var.yi.p[j] * xx[j,]
  d.mu.d.tau.p <- 0.5 * var.yi.p[j] * term1.p
  d.var.d.tau.p <- ((var.yi.p[j] * (0.5 - (3 *
    pi.vec.p[j]) + (3 *
    pi.vec.p[j]^2))) + var.yi.p[j]^2)
  d.var.d.beta.p <- term1.p * d.mu.d.beta.p

  phil.p.vec <- rbind(phil.p.vec, (d.mu.d.beta.p *
    (1 / var.yi.p[j]) *
    y.minus.mu.p[j]))

  phi2.p.vec[j] <- (d.var.d.tau.p * (
    y.minus.mu.p[j]^2 -
    var.yi.p[j] ) / work.var)

  d.phil.d.beta.0 <- ((1 / var.yi.p[j]) * (
    d.mu.d.beta.p %*%
    t(d.mu.d.beta.p)))
  d.phil.d.beta <- d.phil.d.beta + d.phil.d.beta.0

  d.phil.d.tau.0 <- d.mu.d.beta.p * (d.mu.d.tau.p /
    var.yi.p[j])
  d.phil.d.tau <- d.phil.d.tau + d.phil.d.tau.0

  d.phi2.d.beta.0 <- ((d.var.d.tau.p / work.var) *
    d.var.d.beta.p)
  d.phi2.d.beta <- d.phi2.d.beta + d.phi2.d.beta.0

  d.phi2.d.tau.0 <- d.var.d.tau.p^2 / work.var
  d.phi2.d.tau <- d.phi2.d.tau + d.phi2.d.tau.0

```

```

    }
    phi2.p <- sum(phi2.p.vec)

    inv.d.phil.d.beta <- solve(d.phil.d.beta)
    J21.inv.J11 <- t(d.phi2.d.beta) %*% inv.d.phil.d.beta

    var.qscore <- d.phi2.d.tau - (J21.inv.J11 %*%
                                d.phil.d.tau)

    chi.cal.sim <- phi2.p / sqrt(var.qscore)
    p.value.sim1 <- pnorm(chi.cal.sim, 0, 1,
                          lower.tail=FALSE)

    return(p.value.sim1)
}

#####
## Calculation of Power
#####
power.test <- function(p.value, alpha)
{
    power.count <- 0
    power.val <- 0

    for (i in 1:length(p.value)) {
        power.val <- ifelse((p.value[i] < alpha), 1, 0)
        power.count <- power.count + power.val
    }
    power.count/length(p.value)
}

#####
#####
#####

```

```
#####
##          PARAMETER ESTIMATION FUNCTIONS          ##
#####
```

```
###-----###
## Estimation using Quasi-Score functions - Approximation ##
## method                                           ##
###-----###
```

```
quasi.fit <- function(gamma.i, x, xx, y, m=10, n=50, iter=5)
{
```

```
  output.i <- NULL
  output    <- c(gamma.i[1], gamma.i[2], gamma.i[3], 0,0,0)
  output.i <- rbind(output.i, output)
```

```
  isConverge <- FALSE
```

```
  invalidTauFound <- FALSE
  forcedTauValue  <- 0.01
```

```
  beta.p <- get.beta.p(x, m, n, y)
  p      <- exp(xx %*% beta.p) / (1 + exp(xx %*% beta.p))
```

```
  for (i in 1:iter) {
    beta.i    <- gamma.i[1:2]
    tau.i     <- gamma.i[3]
    eta      <- xx %*% beta.i
    p.i      <- exp(eta) / (1 + exp(eta))
    mu.i     <- m * p.i
    y.minus.mu <- y - mu.i
    term1    <- m * (m-1)
    term2    <- p.i * (1 - p.i)
    term3    <- m * term2
```

```
    phi1          <- 0
    phi2          <- 0
    d.phi1.d.beta <- 0
    d.phi1.d.tau  <- 0
    d.phi2.d.beta <- 0
    d.phi2.d.tau  <- 0
    phi.gamma.sq  <- 0
```

```
    for(j in 1:n) {
      var.yi <- term3[j] * (1 + (tau.i *
                                (m-1) * term2[j]))
```

```

work.var <- (m * p[j] * (2 * m * p[j] *
  (p[j] - 1)^2 - (6 *
    p[j]^3) + (12 * p[j]^2) -
    (7 * p[j]) + 1))

d.mu.d.beta <- term3[j] * xx[j,]
d.mu.d.tau <- 0
d.var.d.tau <- term1 * term2[j]^2
d.var.d.beta <- ((1 - (2 * p.i[j])) *
  d.mu.d.beta * (1 + (2 *
    tau.i * (m-1) *
    term2[j])))

phil.0 <- d.mu.d.beta * (1/var.yi) *
  y.minus.mu[j]
phil <- phil + phil.0

phi2.0 <- d.var.d.tau * (1/work.var) *
  (y.minus.mu[j]^2 -
    var.yi)
phi2 <- phi2 + phi2.0

d.phil.d.beta.0 <- ((d.mu.d.beta %*%
  t(d.mu.d.beta)) / var.yi)
d.phil.d.beta <- d.phil.d.beta +
  d.phil.d.beta.0

d.phil.d.tau <- 0

d.phi2.d.tau.0 <- d.var.d.tau^2 /
  work.var
d.phi2.d.tau <- d.phi2.d.tau +
  d.phi2.d.tau.0

d.phi2.d.beta.0 <- d.var.d.tau * (1 /
  work.var) * d.var.d.beta
d.phi2.d.beta <- d.phi2.d.beta +
  d.phi2.d.beta.0

phi.gamma.0 <- c(phil.0, phi2.0)

phi.gamma.sq.0 <- phi.gamma.0 %*%
  t(phi.gamma.0)
phi.gamma.sq <- phi.gamma.sq +

```

```

        phi.gamma.sq.0
    }

    phi.gamma <- c(phi1, phi2)

    d.phi.d.gamma <- rbind(cbind(d.phi1.d.beta,
                                d.phi1.d.tau), cbind(
                                t(d.phi2.d.beta), d.phi2.d.tau))

    inv.d.phi.d.gamma <- solve(d.phi.d.gamma)

    gamma.i <- gamma.i + (inv.d.phi.d.gamma %*%
                        phi.gamma)

## TAU ADJUSTMENT
if ((invalidTauFound) | (tau.i < 0)) {
    tau.i <- forcedTauValue
    invalidTauFound <- TRUE
}

var.cov.gamma <- inv.d.phi.d.gamma %*%
                phi.gamma.sq %*%
                t(inv.d.phi.d.gamma)

output <- c(gamma.i[1], gamma.i[2], gamma.i[3],
            sqrt(var.cov.gamma[1,1]),
            sqrt(var.cov.gamma[2,2]),
            sqrt(var.cov.gamma[3,3]))
output.i <- rbind(output.i, output)

## If we are at the 3rd iteration or later then
## check to see if we already have convergence
## if we do then, break out of this iterative
## process for efficiency.
if(i > 2) {
    ## Check (i)th and (i-1)th row for conv.
    conv.cond.1 <- sum(output.i[i, 1:3] -
                      output.i[i, 1:3])
    conv.cond.2 <- sum(output.i[i-1, 1:3] -
                      output.i[i-2, 1:3])
    if (abs(conv.cond.1 + conv.cond.2) <=
        0.00001) {
        ## Assume conv. at this point...
        isConverge <- TRUE
    }
}

```



```

        break
    }
}

output.i <- round(output.i, digits=5)
dimnames(output.i) <- list(NULL, c("Beta0", "Beta1",
    "Tau", "Beta0 SE", "Beta1 SE", "Tau SE"))

if (!isConverge) {
    cat(
">>>> WARNING: We had qf a run without convergence\n\n");
    print(round(output.i, digits=5))
}

return(output.i)
}

###-----###
## Estimation using Quasi-Score functions - Simulation Method ##
###-----###
simulation.method <- function(gamma.i, x, xx, y, u.s, m=10,
    n=60, S=200, iter=50)
{
    output.i <- NULL
    output <- c(gamma.i[1], gamma.i[2], gamma.i[3], 0,0,0)
    output.i <- rbind(output.i, output)

    isConverge <- FALSE

    invalidTauFound <- FALSE
    forcedTauValue <- 0.01

    beta.p <- get.beta.p(x, m, n, y)
    p <- exp(xx %*% beta.p) / (1 + exp(xx %*% beta.p))

    for (i in 1:iter) {
        beta.i <- gamma.i[1:2]
        tau.i <- gamma.i[3]
        sqrt.tau.i <- sqrt(tau.i)
        eta <- xx %*% beta.i

        phil <- 0

```

```

phi2          <- 0
d.phi1.d.beta <- 0
d.phi2.d.tau  <- 0
d.phi1.d.tau  <- 0
d.phi2.d.beta <- 0
phi.gamma.sq  <- 0

for(j in 1:n) {
  num <- exp(eta[j] + (sqrt.tau.i * u.s))
  pi.vec <- num / (1 + num)
  mean.pi.vec <- mean(pi.vec)

  term1 <- pi.vec * (1 - pi.vec)
  mean.term1 <- mean(term1)
  term2 <- term1 * u.s
  mean.term2 <- mean(term2)

  mu.i <- m * mean.pi.vec
  var.yi <- m * mean.term1 + m^2 *
    var(pi.vec)
  y.minus.mu <- y[j] - mu.i
  work.var <- (m * p[j] * (2 * m * p[j] *
    (p[j] - 1)^2 - (6 *
    p[j]^3) + (12 * p[j]^2) -
    (7 * p[j]) + 1))

  d.mu.d.beta <- m * mean.term1 * xx[j,]
  d.mu.d.tau <- (m / (2 * sqrt.tau.i)) *
    mean.term2
  d.var.d.tau <- (((m / (2 * sqrt.tau.i)) *
    mean((1 - 2 * pi.vec) *
    term2)) + ((m^2 / ((S -
    1) * sqrt.tau.i)) * sum((
    pi.vec - mean.pi.vec) *
    (term2 - mean.term2))))
  d.var.d.beta <- (((m * mean((1 - 2 *
    pi.vec) * term1)) +
    (((2 * m^2) / (S - 1)) *
    sum((pi.vec -
    mean.pi.vec) * (term1 -
    mean.term1)))) * xx[j,])

  phi1.0 <- d.mu.d.beta * (1 / var.yi) *
    y.minus.mu

```

```

    phi1 <- phi1 + phi1.0

    phi2.0 <- d.var.d.tau * (1 / work.var) *
              (y.minus.mu^2 - var.yi)
    phi2 <- phi2 + phi2.0

    d.phi1.d.beta.0 <- (1 / var.yi) *
                      (d.mu.d.beta %*%
                       t(d.mu.d.beta))
    d.phi1.d.beta <- d.phi1.d.beta +
                    d.phi1.d.beta.0

    d.phi2.d.tau.0 <- d.var.d.tau^2 /
                     work.var
    d.phi2.d.tau <- d.phi2.d.tau +
                  d.phi2.d.tau.0

    d.phi1.d.tau.0 <- d.mu.d.beta * (1 /
                                     var.yi) * d.mu.d.tau
    d.phi1.d.tau <- d.phi1.d.tau +
                  d.phi1.d.tau.0

    d.phi2.d.beta.0 <- d.var.d.tau * (1 /
                                     work.var) * d.var.d.beta
    d.phi2.d.beta <- d.phi2.d.beta +
                  d.phi2.d.beta.0

    phi.gamma.0 <- c(phi1.0, phi2.0)

    phi.gamma.sq.0 <- phi.gamma.0 %*%
                    t(phi.gamma.0)
    phi.gamma.sq <- phi.gamma.sq +
                  phi.gamma.sq.0
  }

  phi.gamma <- c(phi1, phi2)

  d.phi.d.gamma <- rbind(cbind(d.phi1.d.beta,
                              d.phi1.d.tau), cbind(
                              t(d.phi2.d.beta), d.phi2.d.tau))

  inv.d.phi.d.gamma <- solve(d.phi.d.gamma)

```

```

gamma.i <- gamma.i + (inv.d.phi.d.gamma %*%
  phi.gamma)

## TAU ADJUSTMENT
if ((invalidTauFound) | (gamma.i[3] < 0)) {
  gamma.i[3] <- forcedTauValue
  invalidTauFound <- TRUE
}

var.cov.gamma <- inv.d.phi.d.gamma %*%
  phi.gamma.sq %*%
  t(inv.d.phi.d.gamma)

output <- c(gamma.i[1], gamma.i[2], gamma.i[3],
  sqrt(var.cov.gamma[1,1]),
  sqrt(var.cov.gamma[2,2]),
  sqrt(var.cov.gamma[3,3]))
output.i <- rbind(output.i, output)

## If we are at the 3rd iteration or later then
## check to see if we already have convergence if
## we do then, break out of this iterative
## process for efficiency.
if(i > 2) {
  ## Check (i)th and (i-1)th row for conv.
  conv.cond.1 <- sum(output.i[i, 1:3] -
    output.i[i, 1:3])
  conv.cond.2 <- sum(output.i[i-1, 1:3] -
    output.i[i-2, 1:3])
  if (abs(conv.cond.1 + conv.cond.2) <=
    0.00001) {
    ## Assume conv. at this point...
    isConverge <- TRUE
    break
  }
}
}

output.i <- round(output.i, digits=5)
dimnames(output.i) <- list(NULL, c("Beta0", "Beta1",
  "Tau", "Beta0 SE", "Beta1 SE", "Tau SE"))
if (! isConverge) {
  cat(
">>>> WARNING: We had a sim run without convergence\n\n");

```

```

        print(round(output.i, digits=5))
    }

    return(output.i)
}

```

```

#####
#####
#####

```

```

#####
##                               INPUT GENERATION FUNCTIONS                               ##
#####

```

```

###-----###
## Generation of Data - Logit/Normal - Binomial      ##
###-----###
y.gen.normal <- function(beta, tau, x, xx, m, n)
{
    common <- exp(c(xx %*% beta) + rnorm(n, mean=0,
        sd=sqrt(tau)))
    p <- common/(1+common)
    y <- rbinom(n, size=m, prob=p)
    return(y)
}

```

```

###-----###
## Generation of Data - Binomial                      ##
###-----###
y.gen.binomial <- function(beta, tau, x, xx, m, n)
{
    common <- exp(c(xx %*% beta))
    p <- common/(1+common)
    y <- rbinom(n, size=m, prob=p)
    return(y)
}

```

```

#####
## Generation of Data (Model Misspecification) - Gamma      ##
#####
y.gen.gamma <- function(beta, tau, x, xx, m, n)
{
  lambda <- y.gen.gamma.lambda
  a      <- rgamma(n, lambda, 1, 1)
  z      <- sqrt(tau) * ((a - lambda) / sqrt(lambda))
  common <- exp(c(xx %*% beta) + z)
  p <- common/(1+common)
  y <- rbinom(n, size=m, prob=p)
  return(y)
}

#####

#####

#####
##                               OUTPUT/FORMATTING/SUPPORT FUNCTIONS      ##
#####

#####
## Calculation of Intervals      ##
#####
get.intervals <- function(output, z.alpha.by.2)
{
  output.int <- NULL
  for(i in 1:nrow(output)) {
    llim.b0 <- output[i,1] - (output[i,4] *
                             z.alpha.by.2)
    ulim.b0 <- output[i,1] + (output[i,4] *
                             z.alpha.by.2)
    llim.b1 <- output[i,2] - (output[i,5] *
                             z.alpha.by.2)
    ulim.b1 <- output[i,2] + (output[i,5] *
                             z.alpha.by.2)
    llim.tau <- output[i,3] - (output[i,6] *
                              z.alpha.by.2)
    ulim.tau <- output[i,3] + (output[i,6] *
                              z.alpha.by.2)
  }
}

```

```

        output.int.row <- c(output[i,1:3], llim.b0,
                           ulim.b0, llim.b1, ulim.b1,
                           llim.tau, ulim.tau)
        output.int <- rbind(output.int, output.int.row)
    }
    output.int <- round(output.int, digits=5)
    dimnames(output.int) = list(NULL, c("Beta0", "Beta1",
        "Tau", "Beta0 LL", "Beta0 UL",
        "Beta1 LL", "Beta1 UL", "Tau LL",
        "Tau UL"))
    return(output.int)
}

```

```

#####
## Calculation of Coverage Probability      ##
#####
cov.prob <- function(dat, comp.val, ll.col, ul.col)
{
    boundcount <- 0
    for (i in 1:nrow(dat)) {
        tmp <- boundcount + 1
        boundcount <- ifelse (((comp.val <= dat[i,
            ul.col]) & (comp.val >= dat[i,
            ll.col])), tmp, boundcount)
    }
    boundcount/nrow(dat)
}

```

```

#####
## Beta for the power calculation          ##
#####
get.beta.p <- function(x, m, n, y)
{
    logistic.reg <- data.frame(x, no.of.trials <- rep(m,n), y)
    n.minus.y <- logistic.reg$no.of.trials -
        logistic.reg$y
    logistic.reg$Ymat <- cbind(logistic.reg$y, n.minus.y)
    log.reg <- glm(Ymat ~ x, family = binomial,
        data = logistic.reg)
    beta.p <- summary(log.reg)$coef[,1]
    return(beta.p)
}

```

```

}

####-----###
## Output Generation as a table containing the calculated ##
## results from the quasi.fit and simulation outputs passed ##
## in as arguments ##
####-----###
get.finaltab <- function(output.qf, output.sim, nom.alpha,
                          beta, tau)
{
  alpha.by.2 <- nom.alpha / 2
  z.alpha.by.2 <- qnorm(1-alpha.by.2)

  output.qf <- get.intervals(output.qf, z.alpha.by.2)

  cov.prob.beta0.qf <- cov.prob(dat=output.qf,
                                comp.val=beta[1], ll.col=4, ul.col=5)
  cov.prob.beta1.qf <- cov.prob(dat=output.qf,
                                comp.val=beta[2], ll.col=6, ul.col=7)
  cov.prob.tau.qf <- cov.prob(dat=output.qf,
                              comp.val=tau, ll.col=8, ul.col=9)

  coverage.probability.qf <- rbind(NULL,
                                    c(cov.prob.beta0.qf, cov.prob.beta1.qf,
                                      cov.prob.tau.qf))
  dimnames(coverage.probability.qf) <- list(c(""),
                                             c("Beta0", "Beta1", "Tau"))

  avg.length.beta0.qf <- mean(output.qf[,5] - output.qf[,4])
  avg.length.beta1.qf <- mean(output.qf[,7] - output.qf[,6])
  avg.length.tau.qf <- mean(output.qf[,9] - output.qf[,8])

  avg.length.qf <- rbind(NULL, c(avg.length.beta0.qf,
                                  avg.length.beta1.qf, avg.length.tau.qf))
  dimnames(avg.length.qf) <- list(c(""),
                                   c("Beta0", "Beta1", "Tau"))

  average.qf <- colMeans(output.qf[,1:3])

  bias.qf <- (average.qf - c(beta[1], beta[2], tau))
  rel.bias.qf <- (bias.qf * 100) / c(beta[1], beta[2], tau)

  mse.qf <- colMeans((output.qf[,1:3] -

```

```

      c(beta[1], beta[2], tau))^2)
root.mse.qf <- sqrt(mse.qf)

output.sim <- get.intervals(output.sim, z.alpha.by.2)

cov.prob.beta0.sim <- cov.prob(dat=output.sim,
                               comp.val=beta[1], ll.col=4, ul.col=5)
cov.prob.beta1.sim <- cov.prob(dat=output.sim,
                               comp.val=beta[2], ll.col=6, ul.col=7)
cov.prob.tau.sim <- cov.prob(dat=output.sim,
                              comp.val=tau, ll.col=8, ul.col=9)

coverage.probability.sim <- rbind(NULL,
                                   c(cov.prob.beta0.sim, cov.prob.beta1.sim,
                                     cov.prob.tau.sim))
dimnames(coverage.probability.sim) <- list(c(""),
                                           c("Beta0", "Beta1", "Tau"))

avg.length.beta0.sim <- mean(output.sim[,5] -
                             output.sim[,4])
avg.length.beta1.sim <- mean(output.sim[,7] -
                             output.sim[,6])
avg.length.tau.sim <- mean(output.sim[,9] -
                           output.sim[,8])

avg.length.sim <- rbind(NULL, c(avg.length.beta0.sim,
                                avg.length.beta1.sim, avg.length.tau.sim))
dimnames(avg.length.sim) <- list(c(""),
                                 c("Beta0", "Beta1", "Tau"))

average.sim <- colMeans(output.sim[,1:3])

bias.sim <- (average.sim - c(beta[1], beta[2], tau))
rel.bias.sim <- (bias.sim * 100) / c(beta[1], beta[2], tau)

mse.sim <- colMeans((output.sim[,1:3] -
                     c(beta[1], beta[2], tau))^2)
root.mse.sim <- sqrt(mse.sim)

#####
## Output
finaltab <- rbind(NULL, c(average.qf, average.sim ))
finaltab <- rbind(finaltab, c(bias.qf, bias.sim ))

```

```

finaltab <- rbind(finaltab , c(rel.bias.qf, rel.bias.sim ))
finaltab <- rbind(finaltab , c(root.mse.qf, root.mse.sim ))
finaltab <- rbind(finaltab , c(coverage.probability.qf,
                               coverage.probability.sim ))
finaltab <- rbind(finaltab , c(avg.length.qf,
                               avg.length.sim ))

dimnames(finaltab) <- list(c("Average", "Bias", "R.Bias",
                             "R.MSE", "CovProb", "AvgLength"),
                           c("Beta0", "Beta1", "Tau", "Beta0",
                             "Beta1", "Tau"))

return(finaltab)
}

#####
## Output Generation as a table containing the calculated ##
## powers ##
#####
get.powertab <- function(c.s.s.qf, c.s.s.sim, c.s.s.sim1, alpha)
{
  power.qf <- power.test(p.value=c.s.s.qf, alpha=alpha)
  power.sim1 <- power.test(p.value=c.s.s.sim1, alpha=alpha)

  powertab <- rbind(NULL, c(power.qf, power.sim1))
  dimnames(powertab) <- list(c("Power"),
                             c("Score", "Q.Score_Sim"))

  return(powertab)
}

#####
##                                END FUNCTIONS                                ##
#####

```

```
#####  
#####  
##                                MAIN VARIABLES                                ##  
#####  
#####  
  
### USED THROUGHOUT TO INDICATE ENABLING/DISABLING OF  
### ESTIMATION AND POWER RELATED OPERATIONS...  
enable.estimation <- TRUE  
enable.power      <- TRUE  
  
no.sim            <- 5000 ## Number of simulations  
S                 <- 2000  
  
### There are 3 y.gen methods defined uncomment just one for use.  
y.gen             <- y.gen.normal  
#y.gen            <- y.gen.gamma  
#y.gen            <- y.gen.binomial  
  
### lambda value used within the y.gen.gamma method.  
y.gen.gamma.lambda <- 2  
  
tau.list          <- c(0.01, 0.025, 0.05, 0.1, 0.25, 0.5)  
#tau.list         <- c(0.1, 0.25, 0.5, 0.8)  
tau.i.list        <- tau.list  
# tau.list        <- c(0)  
# tau.i.list      <- c(0.2)  
  
n.list            <- c(50, 100)  
  
#####  
#####
```


Bibliography

- [1] BOOS, D. D. On generalized score tests. *The American Statistician* 46 (1992), 327–333.
- [2] CHESHER, A. Testing for neglected heterogeneity. *Econometrica* 52 (1984), 865–872.
- [3] COX, D. R. Some remarks on overdispersion. *Biometrika* 70 (1983), 269–274.
- [4] CROWDER, M. J. Beta-binomial anova for proportions. *Applied Statistics* 27 (1978), 34–37.
- [5] CROWDER, M. J. Gaussian estimation for correlated binomial data. *Journal of the Royal Statistical Society* 47 (1985), 229–237.
- [6] DEAN, C. B. Testing for overdispersion in poisson and binomial regression models. *Journal of the American Statistical Association* 87 (1992), 451–457.
- [7] JIANG, J. Consistent estimators in generalized linear mixed models. *Journal of the American Statistical Association* 93 (1998), 720–729.

-
- [8] JIANG, J. *Linear and Generalized Linear Mixed Models and Their Applications*. Springer Science+Business Media, LLC, Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA, 2007.
- [9] JOWAHEER, V., AND SUTRADHAR, B. C. Analysing longitudinal count data with overdispersion. *Biometrika* 89 (2002), 389–399.
- [10] KALBFLEISCH, J. D., AND PRENTICE, R. L. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, Inc., Wiley Series in Probability and Statistics, 1980.
- [11] MCCULLAGH, P. Quasi-likelihood functions. *The Annals of Statistics* 11 (1983), 59–67.
- [12] MCCULLAGH, P., AND NELDER, J. A. *Generalized Linear Models*, second edition ed. Chapman & Hall, London, 1989.
- [13] MCCULLOCH, C. E. Generalized linear models. *Journal of the American Statistical Association* 95 (2000), 1320–1324.
- [14] MCCULLOCH, C. E., AND SEARLE, S. R. *Generalized, Linear, and Mixed Models*. John Wiley & Sons, Inc., Wiley Series in Probability and Statistics, 605 Third Avenue, New York, NY, USA, 10158-0012, 2001.
- [15] MOORE, D. F. Asymptotic properties of moment estimators for overdispersed counts and proportions. *Biometrika* 73 (1986), 583–588.

- [16] NELDER, J. A., AND WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society* 135 (1972), 370–384.
- [17] PRENTICE, R. L. Binary regression using an extended beta-binomial distribution, with discussion of correlation included by covariate measurement errors. *Journal of the American Statistical Association* 81 (1986), 321–327.
- [18] SEARLE, S. R., CASELLA, G., AND MCCULLOCH, C. E. *Variance Components*. John Wiley & Sons, Inc., Wiley Series in Probability and Statistics, 111 River Street, Hoboken, NJ, 07030, USA, 1992.
- [19] SINHA, S. K. Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association* 99 (2004), 451–460.
- [20] STIRATELLI, R., LAIRD, N., AND WARE, J. Random effects models for serial observations with binary responses. *Biometrics* 40 (1984), 961–971.
- [21] SUTRADHAR, B. C. On exact quasiliikelihood inference in generalized linear mixed models. *Sankhya: The Indian Journal of Statistics* 66 (2004), 263–291.
- [22] SUTRADHAR, B. C., AND FARRELL, P. J. Analyzing multivariate longitudinal binary data: A generalized estimating equations approach. *The Canadian Journal of Statistics* 32 (2004), 39–55.

-
- [23] SUTRADHAR, B. C., AND RAO, R. P. On joint estimation of regression and overdispersion parameters in generalized linear models for longitudinal data. *Journal Of Multivariate Analysis* 56 (1996), 90–119.
- [24] TARONE, R. E. Testing the goodness of fit of the binomial distribution. *Biometrika* 66 (1979), 585–590.
- [25] WEDDERBURN, R. W. Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika* 61 (1974), 439–447.
- [26] WILLIAMS, D. A. Extra-binomial variation in logistic linear models. *Applied Statistics* 31 (1982), 144–148.