Behavior of Residuals Given by Linear Mean-Variance Negative Binomial Models

Mallika Inprom

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Mathematics and Statistics
Prince of Songkla University
2014

Title	Behavior of Residuals Given by Linear Mean-Variance Negative Binomial Models
Author	Miss Mallika Inprom
Major Program	Mathematics and Statistics

Advisor:

.....

(Asst. Prof. Dr. Naratip Jansakul)

Co-Advisor:

.....

(Dr. Klairung Samart)

Examination Committee:

.....Chairperson

(Dr. Sompong Chuysurichay)

.....

(Asst. Prof. Dr. Naratip Jansakul)

.....

(Dr. Klairung Samart)

.....

(Asst. Prof. Dr. Winai Bodhisuwan)

The Graduate School, Prince of Songkla University, has approved this thesis as partial fulfillment of the requirements for the Master of Science Degree in Mathematics and Statistics

.....

(Assoc. Prof. Dr. Teerapol Srichana)

Dean of Graduate School

This is to certify that the work here submitted is the result of the candidate's own investigations. Due acknowledgement has been made of any assistance received.

..... Signature
(Asst. Prof. Dr. Naratip Jansakul)
Major Advisor

..... Signature
(Dr. Klairung Samart)
Major Co-Advisor

..... Signature
(Mallika Inprom)
Candidate

I hereby certify that this work has not been accepted in substance for any degree, and is not being currently submitted in candidature for any degree.

..... Signature

(Mallika Inprom)

Candidate

ชื่อวิทยานิพนธ์ พฤติกรรมของส่วนตกค้างจากตัวแบบทวินามลบที่ความแปรปรวนเป็นฟังก์ชันเชิงเส้นของค่าเฉลี่ย

ผู้เขียน นางสาวมัลลิกา อินพรหม

สาขาวิชา คณิตศาสตร์และสถิติ

ปีการศึกษา 2557

บทคัดย่อ

ตัวแบบทวินามแบบลบ (Negative binomial (NB) model) เป็นตัวแบบที่นิยมใช้วิเคราะห์ข้อมูลแบบนับที่มีความแปรปรวนมากกว่าค่าเฉลี่ย หรือเรียกว่า Overdispersion ฟังก์ชันความน่าจะเป็นของการแจกแจงทวินามแบบลบเกิดจากการสมมติให้ค่าเฉลี่ยของตัวแปรสุ่มที่มีการแจกแจงปัวซงมีการแจกแจงแบบแกมม่า ตัวแบบทวินามแบบลบมีหลายรูปแบบ แต่ที่นิยมใช้มากที่สุดคือ ตัวแบบทวินามแบบลบที่มีความแปรปรวนเป็นฟังก์ชันเชิงเส้นของค่าเฉลี่ย (Linear mean-variance NB model: NB1) และที่มีความแปรปรวนเป็นฟังก์ชันกำลังสองของค่าเฉลี่ย (Quadratic mean-variance NB model: NB2)

วิทยานิพนธ์ฉบับนี้ศึกษากระบวนการพัฒนาค่าประมาณส่วนตกค้างดีเวียนซ์ (Deviance residual) ของตัวแบบ NB1 ซึ่งได้มีการใช้ในงานวิจัยของ Jansakul และ Hinde ในปี ค.ศ. 2004 [13] การตรวจสอบเริ่มต้นโดยการแสดงว่า ตัวแบบ NB1 ไม่ได้เป็นสมาชิกของตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized linear models: glms) วิทยานิพนธ์ฉบับนี้ได้ดำเนินการจำลองข้อมูลเพื่อพิจารณาค่าปรับเปลี่ยนที่เหมาะสมที่ทำให้ฟังก์ชัน log-likelihood ของตัวแบบ NB1 มีค่าสูงสุด เพื่อให้ได้สูตรการคำนวณส่วนตกค้างดีเวียนซ์ประมาณของตัวแบบ NB1 ของ Jansakul และ Hinde จากการศึกษาพบว่า ส่วนตกค้างดีเวียนซ์ประมาณของตัวแบบ NB1 มีคุณลักษณะหรือพฤติกรรมเหมือนกับส่วนตกค้างดีเวียนซ์ของการแจกแจง

ที่เป็นสมาชิกของตัวแบบเชิงเส้นวางนัยทั่วไป

นอกจากศึกษากระบวนการต่างๆ ที่กล่าวมาข้างต้นแล้ว ผู้วิจัยได้พัฒนา hat value หรือ leverage ประมาณของตัวแบบ NB1 เพื่อใช้ในการคำนวณส่วนตกค้างดีเวียนซ์มาตรฐาน (Standardized deviance residual) ประมาณของตัวแบบ NB1 และเพื่อเป็นการยืนยันว่า hat value และส่วนตกค้างดีเวียนซ์มาตรฐานที่เราได้นำเสนอมีความเหมาะสม เราจึงได้ดำเนินการจำลองข้อมูลเพื่อตรวจสอบคุณสมบัติของตัวประมาณค่าทั้งสองอีกครั้ง จากการตรวจสอบพบว่า ในกรณีที่ Overdispersion parameter มีขนาดเล็ก hat value และส่วนตกค้างดีเวียนซ์มาตรฐานประมาณของตัวแบบ NB1 จะมีคุณสมบัติทั่วไปตามที่ปรากฏในเอกสารที่ตรวจสอบ ส่วนในกรณีที่ Overdispersion parameter มีขนาดใหญ่จะต้องมีการดำเนินการศึกษาต่อไป

ผู้วิจัยได้สาธิตการใช้ hat value และส่วนตกค้างดีเวียนซ์มาตรฐานประมาณของตัวแบบ NB1 ที่พัฒนาขึ้นในการตรวจสอบตัวแบบ NB1 ที่พิจารณาแล้วว่าเป็นตัวแบบที่เหมาะสมที่สุดในการอธิบายค่าเฉลี่ยของจำนวนตัวอ่อนที่เจริญจากการเพาะเลี้ยงเนื้อเยื่อของส้มสายพันธุ์ Valencia ดังแสดงในบทความวิจัยของ Jansakul และ Hinde ในปี ค.ศ. 2004 [13] และสามารถสรุปได้ว่าสามารถใช้ส่วนตกค้างดีเวียนซ์มาตรฐานนี้ในการตรวจสอบความแปรปรวนคงที่และการแจกแจงปกติมาตรฐานของค่าคลาดเคลื่อนเชิงสุ่มของตัวแบบ NB1 ได้

Title Behavior of Residuals Given by Linear Mean-Variance Negative Binomial Models

Author Miss Mallika Inprom

Major Program Mathematics and Statistics

Academic Year 2014

ABSTRACT

Negative binomial (NB) models are commonly used in modeling count data with overdispersion. The model can be formed by assuming that the counts come from a Poisson distribution with varying means. Taking the Poisson mean as gamma distributed random variable leads to various mean-variance related NB models. However, the most commonly used models are the linear mean-variance NB (NB1) and the quadratic mean-variance NB (NB2).

This thesis presents a procedure of obtaining an approximated deviance residual of the NB1 model used in the research paper of Jansakul and Hinde in 2004 [13]. The investigation is started by showing that the NB1 model is not a member of generalized linear models (glms), conducting some simulation study to consider an appropriate adjusted value for the mode of the NB1 log-likelihood and applying the value to form the NB1 deviance residual. Our study shows that the approximated NB1 deviance residual behaves as well as the glm deviance residual.

We developed an approximated hat value (or leverage) of the NB1 model and used that to form an estimated NB1 standardized deviance residual. In order to confirm that our proposed hat value and standardized deviance residual are appropriate, we again conducted some simulation study to investigate the properties of both quantities. Our studies show that our approximated hat value and the NB1 standardized deviance residual with a small value of overdispersion parameter follow the general theoretical properties. In the case of the NB1 model with a large value of the overdispersion, a further study is required.

For illustration, we applied our proposed quantities to check the adequacy of the final NB1 model for the mean numbers of embryos of orange variety Valencia tissue culture as presented in Jansakul and Hinde (2004) [13]. It shows that our the NB1 standardized deviance residual can be used to check the constant variance and normality assumption of NB1 model in a population of study.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisors, Asst. Prof. Dr. Naratip Jansakul and Dr. Klairung Samart, for their initial ideas, guidance, constant instruction, kindness and encouragement which enable me to carry out my study successfully.

My special appreciation is expressed to Dr. Sompong Chuysurichay and Asst. Prof. Dr. Winai Bodhisuwan for valuable comments and helpful suggestions.

I wish to thank all my teachers in the Department of Mathematics and Statistics, Prince of Songkla University for sharing their knowledge and support so that I can obtain this Master degree.

I gratefully appreciate my family for their love and encouragement. I would like to thank my friends for their encouragement and suggestions.

Finally, I would like to thank every one, who supported me, but I did not mention above.

Mallika Inprom

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIC	Akaike information criterion
BIC	Bayesian information criterion
df	Degree of freedom
glm	Generalized linear model
IRLS	Iteratively reweighted least square
MSE	Mean square residual
NB	Negative binomial
NB1	Linear mean-variance negative binomial
NB2	Quadratic mean-variance negative binomial
p.d.f	Probability density function
p.m.f	Probability mass function

CHAPTER 1

Introduction

1.1 Background and Motivation

Negative binomial (NB) models are commonly used in modeling count data with overdispersion. The overdispersion is explained by the count variance exceeding the mean and a standard Poisson model is not appropriate. The NB model can be formed by assuming that the counts come from a Poisson distribution with varying means. Taking the Poisson mean as a gamma distributed random variable, it leads to various mean-variance related NB models; see Hinde and Demetrio (1998) [11] and Hilbe (2007) [10] for detail. However, the most commonly used models are the linear mean-variance NB model, denoted by NB1 model and the quadratic mean-variance NB model, represented by NB2 model. The NB2 model is a generalized linear model (glm) when the shape parameter is known; hence the methods of model fitting and model checking are widely available in associated literatures.

Model checking including checking systematic departure, outliers and influential values, is an important part of fitting a model. Quantity mostly used in the model diagnostics is residuals and the associated procedure is known as residual analysis. There are three basic types of residuals in statistical modeling; the Pearson, Anscombe and deviance residuals (McCullagh and Nelder, 1989

[16], Cameron and Trivedi, 1998 [6]). This thesis will mainly concentrate on the deviance residual because the context of statistical modeling with known probability mass (density) functions, the linear predictor coefficients are mainly estimated by using the maximum likelihood method and the model selection along with model checking are performed by using the deviance residuals. For distributions that are member of generalized linear models, an individual deviance residual equals zero when the response and its mean are identical.

The NB1 model is not a member of glms resulting its deviance residual cannot be obtained as simply as a glm, since the individual components of deviance function can be negative (Jansakul and Hinde, 2004 [13]). Jansakul and Hinde (2004) [13] proposed an approximated NB1 deviance residual and used that in checking the adequacy of the final NB1 fitted model for the mean numbers of embryos the orange variety *Valencia* used in a tissue-culture experiment. However, they did not report their exploration.

This thesis will focus on exploring the background of obtaining the approximated NB1 deviance residual given in Jansakul and Hinde (2004) [13], developing an approximated standardized NB1 deviance residual together with its associated exploration.

1.2 Objectives

1. To study a general form of NB1 models and its properties;

2. To explore the properties of NB1 deviance residuals;

3. To develop an approximated standardized deviance residual of the NB1 model.

1.3 Scope and Methodology

1. Study characteristics, theories and properties of generalize linear models and their residuals;

2. Study a general form and properties of NB1 models;

3. Explore how to obtain the approximated NB1 deviance residual presented in Jansakul and Hinde (2004) [13];

4. Develop an approximated NB1 standardized deviance residual.

1.4 Advantages

1. Full fill the knowledge of NB1 models and theirs properties;

2. Understand and be able to fit a NB1 model, investigate the background of obtaining an approximated NB1 deviance residual used in an associated literature;

3. Understand and be able to develop an approximated standardized deviance residual of the NB1 model.

1.5 Thesis Overview

We have described the motivation, objectives and scopes of our study. We now summarized where in this thesis the exploration and investigation are performed.

In Chapter 2, we explore the basic concepts of statistical errors and residuals of linear regression and generalized linear models, followed by three basic types of residuals: Pearson, Anscombe and deviance residuals, used in statistical modeling. This thesis will mainly concentrate on the deviance residual, where we firstly explore the association of the deviance residuals and the log-likelihood function of the four well known glms; normal, binomial, Poisson and gamma distribution.

Chapter 3 firstly presents a general form of the negative binomial model and separates that to two well known forms: the linear mean-variance negative binomial model and the quadratic mean-variance negative binomial model. The NB2 model is a glm when the shape parameter is known, then the model fitting and deviance residual are easily obtained. Secondly, we show that the NB1 model is not a member of glms and go to explore the background of obtaining an approximated NB1 deviance residual used in Jansakul and Hinde (2004) [13]. The investigation is started by considering an appropriate adjusted value for the mode of the NB1 log-likelihood and applying the value to form the NB1 deviance residual. The exploration is form using a simulation study.

In Chapter 4, we develop an approximated hat value of the NB1 model and conducted a simulation study in order to investigate the general properties

of the hat values and apply the value to form an approximated NB1 standardized deviance residual using the NB1 deviance residual in Chapter 3 as a base line. Again a simulation study is conducted to investigate its ideal (standard normal) distribution. The approximated standardized NB1 deviance residual is used to check the adequacy of the final NB1 model for the numbers of embryos of orange variety *Valencia* tissue culture data as an illustration.

This thesis ends with a summary of the main conclusion a long with a suggestion for a further study.

CHAPTER 2

Reviews of Literature

2.1 Basic concepts of statistical errors and residuals

In the context of statistical analysis, in particular, statistical modeling, statistical errors and residuals are very closely related and easily confused. Some think that they are identical. They are actually not the same, though they are related.

For a series of observations y_i, $i = 1, 2, ..., N$ of a univariate distributed random variable, Y with mean μ and variance σ^2, the statistical errors and residuals are defined as follows:

2.1.1 Statistical errors are the difference of observed values and the population mean, i.e. for the i^{th} statistical errors, denoted by ε_i is then

$$\varepsilon_i = y_i - \mu. \tag{2.1}$$

The term ε_i is a random variable having similar distribution of Y with mean 0 and variance σ^2. Since μ is unobservable, hence error ε_i can't be observed.

2.1.2 Residuals are the difference of sample observed values and the sample mean, $\overline{y}$. For a sample of size n, the i^{th} residual, denoted by e_i is defined by

$$e_i = y_i - \overline{y}, \tag{2.2}$$

which is the i^{th} observable estimates of ε_i. e_i is a random variable with mean

of zero (0) and the mean square residual; $\frac{1}{n-1}\sum_{i=1}^{n} e_i^2$, respectively. Based on a statistical theory, the mean square residual is an unbiased estimator of σ^2.

In the context of statistical modeling including linear regression analysis and generalized linear models, the expression of residual is not as simply as shown in (2.2), where the description of those is given here.

2.1.3 Statistical errors and residuals in linear regression analysis

In the case of linear regression analysis, when observed values y_i of a response variable Y_i, $i = 1, 2, ..., N$ is a linear function of p explanatory variables; $x_{1i}, x_{2i}, ..., x_{pi}$ and the constant term, the relationship between y_i and $\boldsymbol{x}_i$ can be defined as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i. \tag{2.3}$$

Here, ε_i is a statistical error which is assumed to be independent and identically distributed. Under the stronger assumption, ε_i is independent and identically normally distributed, denoted by $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$. This gives Y_i are normally distributed with mean $\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}$ and variance σ^2, denoted by $Y_i \sim \mathrm{N}(\mu_i, \sigma^2)$. Again the error term $\varepsilon_i = y_i - \mu_i$ and is unobservable.

For a random sample of observations, $y_1, y_2, ..., y_n$, the parameter estimates, $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p$, for the regression coefficients $\beta_0, \beta_1, \beta_2, ..., \beta_p$ can be obtained using the method of least square or maximum likelihood. The well known expression for $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p\right)^{\mathrm{T}}$ in a matrix form is

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}, \tag{2.4}$$

where $\boldsymbol{X}$ is an $n \times (p+1)$ design matrix with the elements of the first column

are 1 for the constant term and $\boldsymbol{y}$ is the vector of responses. The vector of fitted values, $\hat{\boldsymbol{y}}$ of $\boldsymbol{y}$ is then

$$\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{y}} = \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{H}\boldsymbol{y}, \tag{2.5}$$

where $\boldsymbol{H} = \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}$ is known as **the hat or hessian matrix** with the rank of $p+1$. The main diagonal element, denoted by h_{ii}, $i = 1, 2, ..., n$, where $0 \leq h_{ii} \leq 1$ is call **hat values or leverages**. Hoaglin (1992) [12] presents a simple form of the hat values for the simple linear regression as

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{k=1}^{n}(x_k - \bar{x})^2}. \tag{2.6}$$

The hat values plays important role in the statistical modeling: checking outliers and influential values.

In the linear regression modeling, the i^{th} observed residual, e_i, observable estimates of the ε_i is defined by

$$e_i = y_i - \hat{\mu}_i, \tag{2.7}$$

which is normally distributed with mean 0 and variance $\hat{\sigma}^2(1 - h_{ii})$, where $\hat{\sigma}^2 = \dfrac{1}{n-p-1}\sum_{i=1}^{n} e_i^2$. The residuals play important role in model assumption checking: constant variance and normality in regression analysis.

2.1.4 Statistical errors and residuals in generalized linear models

In the case of generalized linear models, where the distribution of the response variable Y_i is in an exponential family distribution, defined (ignored i) as

$$f(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}, \tag{2.8}$$

with specific functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$. θ and $a(\phi)$ are **the natural or canonical parameter** and **scale or dispersion parameter**, respectively. The mean of Y; μ is obtained through a linear predictor defined by $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$ which is linked by a link function $g(\mu)$ written as

$$g(\mu) = \eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p. \tag{2.9}$$

The estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p$ for $\beta_0, \beta_1, \beta_2, ..., \beta_p$ can be obtained by using the method of maximum likelihood via an iteratively reweighted least square algorithm, (McCullagh and Nelder, 1989 [16] and Aitkin *et al.*, 1990 [2]).

Given a vector of n observations $\boldsymbol{y} = (y_1, y_2, ..., y_n)^{\mathrm{T}}$ of Y from the exponential family distribution, the likelihood function $L = L(\boldsymbol{\theta}, \phi; \boldsymbol{y})$ is

$$L(\boldsymbol{\theta}, \phi; \boldsymbol{y}) = \prod_{i=1}^{n} f(y_i; \theta_i, \phi) = \exp\left[\sum_{i=1}^{n} \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}\right] \tag{2.10}$$

and the corresponding log-likelihood function $\ell = \ell(\boldsymbol{\theta}, \phi; \boldsymbol{y}) = \ln L(\boldsymbol{\theta}, \phi; \boldsymbol{y})$ is

$$\ell = \ell(\boldsymbol{\theta}, \phi; \boldsymbol{y}) = \sum_{i=1}^{n} \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}. \tag{2.11}$$

Some elementary properties of linear exponential families follow the familiar identities

$$\mathrm{E}\left(\frac{\partial \ell}{\partial \theta}\right) = 0 \qquad \text{and} \qquad \mathrm{E}\left(\frac{\partial^2 \ell}{\partial \theta^2}\right) + \mathrm{Var}\left(\frac{\partial \ell}{\partial \theta}\right) = 0. \tag{2.12}$$

Simple calculation shows that the mean and variance of Y_i are

$$\mathrm{E}(Y_i) = b'(\theta_i) = \mu_i \qquad \text{and} \qquad \mathrm{Var}(Y_i) = a(\phi) b''(\theta_i) = a(\phi) V(\mu_i). \tag{2.13}$$

Here $b'(\theta_i)$ and $b''(\theta_i)$ denote the first and second derivatives of $b(\theta_i)$ with respect to θ_i, respectively. The function $b''(\theta_i)$ can be defined as $V(\mu_i)$, because it depends on μ_i through $b'(\theta_i)$, and is called **the variance function** of the model.

In general, the log-likelihood function (2.11) is expressed as a function of the mean-value parameter, $\mathrm{E}(Y) = \mu$. That is $\ell = \ell(\boldsymbol{\mu}; \boldsymbol{y}) = \ell(\boldsymbol{\theta}, \phi; \boldsymbol{y})$.

Since the maximum likelihood estimate $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p\right)^{\mathrm{T}}$ for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, ..., \beta_p)^{\mathrm{T}}$ cannot be obtained directly from the normal equation $(\frac{\partial \ell}{\partial \beta_j} = 0, j = 0, 1, 2, ..., p)$. It needs an iterative scheme, such as the iteratively reweighted least square (IRLS) algorithm, which is defined by

$$\hat{\boldsymbol{\beta}}^{(\mathrm{r}+1)} = \hat{\boldsymbol{\beta}}^{(\mathrm{r})} + \left(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{W}^{(\mathrm{r})} \boldsymbol{X}\right)^{-1} \mathbf{s}\left(\hat{\boldsymbol{\beta}}^{(\mathrm{r})}\right). \tag{2.14}$$

That is the new estimate $\hat{\boldsymbol{\beta}}^{(\mathrm{r}+1)}$ of $\boldsymbol{\beta}$ is obtained from the previous estimate, $\hat{\boldsymbol{\beta}}^{(\mathrm{r})}$. $\boldsymbol{W}$ is a diagonal weight matrix with elements w_{ii} defined by

$$w_{ii} = \frac{1}{\mathrm{Var}(Y_i)\left[g'(\mu_i)\right]^2} \tag{2.15}$$

and $\mathbf{s}(\boldsymbol{\beta}) = \dfrac{\partial \ell}{\partial \beta_j} = \displaystyle\sum_{i=1}^{n} \frac{(y_i - \mu_i) x_{ij}}{\mathrm{Var}(Y_i) g'(\mu_i)}$, $g'(\mu_i) = \dfrac{\partial g(\mu_i)}{\partial \mu_i}$ are evaluated at $\hat{\boldsymbol{\beta}}^{(\mathrm{r})}$, see Aitkin *et al.* (1990) [2] for details. A general form of the hat matrix for the glms is defined by

$$\boldsymbol{H} = \boldsymbol{W}^{\frac{1}{2}} \boldsymbol{X} \left(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{W}^{\frac{1}{2}}, \tag{2.16}$$

where $\boldsymbol{W}$ is a diagonal matrix with the elements w_{ii} evaluated at $\mu_i = \hat{\mu}_i$. In this case, the hat values, h_{ii} are the main diagonal elements of $\boldsymbol{H}$. However, the h_{ii}, in particular, for glms, can be easily obtained using any statistical package.

2.2 Types of Residuals

There are three basic types of residuals in statistical modeling including the Pearson, Anscombe and deviance residuals (McCullagh and Nelder, 1989 [16], Cameron and Trivedi, 1998 [6]). However, ones that are commonly used, in particular, for generalized linear models, where the distribution of response variable Y is in an exponential family, are Pearson residuals and deviance residuals.

2.2.1 Pearson residuals

Following McCullagh and Nelder (1989) [16], the Pearson residual is defined by the raw residual divided by the estimated standard deviation of the variance function. Then the i^{th} Pearson residual is then

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}} \tag{2.17}$$

and the i^{th} standardized Pearson residual, denoted by $\dot{r}_i$, is

$$\dot{r}_i = \frac{r_i}{\sqrt{a\left(\hat{\phi}\right)(1-h_{ii})}}, \tag{2.18}$$

where h_{ii} is the i^{th} main diagonal elements of the hat matrix.

Following we present the i^{th} Pearson residual along with its corresponding standardized Pearson residual for four well known glm responses:

(1) Normal responses: Let $Y_1, Y_2, ..., Y_n$ be a random sample from a normal distribution with mean μ_i and variance σ^2, represented by $Y_i \sim \mathrm{N}(\mu_i, \sigma^2)$, where its probability density function (p.d.f) is

$$f(y_i; \mu_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu_i)^2\right\}, \quad -\infty < y_i < \infty, \ \ \sigma > 0.$$

This can be written in the exponential family distributional form as

$$f(y_i;\mu_i,\sigma^2)=\exp\left\{\frac{y_i\mu_i-\frac{1}{2}\mu_i^2}{\sigma^2}+\left[-\frac{y_i^2}{2\sigma^2}-\frac{1}{2}\ln(2\pi\sigma^2)\right]\right\}, \tag{2.19}$$

with $\mathrm{E}(Y_i)=\mu_i$ and $\mathrm{Var}(Y_i)=\sigma^2$. Then the i^{th} Pearson residual, $r_{\mathrm{N}i}$ and standardized Pearson residual, $\dot{r}_{\mathrm{N}i}$ are

$$r_{\mathrm{N}i}=y_i-\hat{\mu}_i \qquad \text{and} \qquad \dot{r}_{\mathrm{N}i}=\frac{r_{\mathrm{N}i}}{\sqrt{s^2(1-h_{ii})}},$$

where s^2 is the mean square residual (MSE), an unbiased estimator of σ^2.

(2) Binomial responses: Let $Y_1,Y_2,...,Y_n$ be a random sample from a binomial distribution with probability of success p_i and number of trial m_i, denoted by $Y_i\sim\mathrm{Bin}(m_i,p_i)$. Its probability mass function (p.m.f) is defined by

$$f(y_i;m_i,p_i)=\binom{m_i}{y_i}p_i^{y_i}(1-p_i)^{m_i-y_i},\quad y_i=0,1,2,...,m_i,\ \ 0\le p_i\le 1.$$

This can be written in the exponential family distributional form as

$$f(y_i;m_i,p_i)=\exp\left\{\frac{y_i\ln\left(\frac{p_i}{1-p_i}\right)+m_i\ln(1-p_i)}{1}+\ln\binom{m_i}{y_i}\right\}, \tag{2.20}$$

with $\mathrm{E}(Y_i)=\mu_i=m_ip_i$ and $\mathrm{Var}(Y_i)=m_ip_i(1-p_i)$. Since $p_i=\dfrac{\mu_i}{m_i}$, then (2.20) can be written in the form of μ_i as

$$f(y_i;m_i,\mu_i)=\exp\left\{\frac{y_i\ln\left(\frac{\mu_i}{m_i-\mu_i}\right)+m_i\ln\left(\frac{m_i-\mu_i}{m_i}\right)}{1}+\ln\binom{m_i}{y_i}\right\}. \tag{2.21}$$

Since $a\left(\hat{\phi}\right)=1$ and $\mathrm{Var}(Y_i)=\dfrac{\mu_i(m_i-\mu_i)}{m_i}$, then the i^{th} Pearson residual, $r_{\mathrm{Bin}i}$ and standardized Pearson residual, $\dot{r}_{\mathrm{Bin}i}$ are

$$r_{\mathrm{Bin}i}=\frac{\sqrt{m_i}(y_i-\hat{\mu}_i)}{\sqrt{\hat{\mu}_i(m_i-\hat{\mu}_i)}} \qquad \text{and} \qquad \dot{r}_{\mathrm{Bin}i}=\frac{r_{\mathrm{Bin}i}}{\sqrt{(1-h_{ii})}}.$$

(3) Poisson responses: Let $Y_1, Y_2, ..., Y_n$ be a random sample from a Poisson distribution with mean μ_i, denoted by $Y_i \sim \text{Pois}(\mu_i)$. Its p.m.f is defined by

$$f(y_i; \mu_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \qquad y_i = 0, 1, 2, ... \,; \;\; \mu_i > 0.$$

This can be written in the exponential family distributional form as

$$f(y_i; \mu_i) = \exp\left\{\frac{y_i \ln \mu_i - \mu_i}{1} - \ln(y_i!)\right\}, \tag{2.22}$$

with $\text{E}(Y_i) = \mu_i$ and $\text{Var}(Y_i) = \mu_i$. Since $a\left(\hat{\phi}\right) = 1$, then the variance function, $V(\mu_i) = \mu_i$, so the i^{th} Pearson residual, $r_{\text{Pois}i}$ and standardized Pearson residual, $\dot{r}_{\text{Pois}i}$ are

$$r_{\text{Pois}i} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}} \qquad \text{and} \qquad \dot{r}_{\text{Pois}i} = \frac{r_{\text{Pois}i}}{\sqrt{(1 - h_{ii})}}.$$

(4) Gamma responses: Let $Y_1, Y_2, ..., Y_n$ be a random sample from a gamma distribution with shape parameter α and scale parameter β_i, denoted by $Y_i \sim \text{Gamma}(\alpha, \beta_i)$. The p.d.f is defined by

$$f(y_i; \beta_i, \alpha) = \frac{1}{\Gamma(\alpha)\beta_i^{\alpha}} y_i^{\alpha-1} e^{-y_i/\beta_i}, \qquad y_i \geq 0 \;\text{ and }\; \alpha, \beta_i > 0.$$

Here $\Gamma(\alpha)$ is the gamma function defined by $\Gamma(\alpha) = \displaystyle\int_0^{\infty} e^{-t} t^{\alpha-1}\, dt$.

This can be written in the exponential family distributional form as

$$f(y_i; \beta_i, \alpha) = \exp\left\{\frac{-\dfrac{y_i}{\beta_i} - \alpha \ln \beta_i}{1} + (\alpha - 1)\ln y_i - \ln \Gamma(\alpha)\right\}, \tag{2.23}$$

with $\text{E}(Y_i) = \mu_i = \alpha\beta_i$ and $\text{Var}(Y_i) = \alpha\beta_i^2$. Since $\beta_i = \dfrac{\mu_i}{\alpha}$, then (2.23) can be written in the form of μ_i as

$$f(y_i; \mu_i, \alpha) = \exp\left\{\frac{-y_i\mu_i^{-1} - \ln \mu_i + \ln \alpha}{\alpha^{-1}} + (\alpha - 1)\ln y_i - \ln \Gamma(\alpha)\right\}. \tag{2.24}$$

Then the i^{th} Pearson residual, $r_{\text{Gamma}i}$ and standardized Pearson residual, $\dot{r}_{\text{Gamma}i}$ are

$$r_{\text{Gamma}i} = \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \qquad \text{and} \qquad \dot{r}_{\text{Gamma}i} = \frac{r_{\text{Gamma}i}\sqrt{\hat{\alpha}}}{\sqrt{(1-h_{ii})}}, \quad \hat{\alpha} > 0.$$

2.2.2 Anscombe residuals

As mentioned in McCullagh and Nelder (1989) [16], a disadvantage of the Pearson residual is the distribution of the residual for non-Normal distributions is often skewed, and so it may fail to have properties similar to those of a Normal-theory residual. An alternative that can be used for the non-Normal distributed response is the Anscombe residual (Anscombe, 1953 [3]). The Anscombe residual is defined by using a function $A(y)$ in place of y. The function $A(\cdot)$ is chosen to make the distribution of $A(Y)$ as **normal as possible**. For the likelihood functions forming in glms, the function $A(\cdot)$ is given by

$$A(\cdot) = \int \frac{1}{\sqrt[3]{V(\mu)}}\, d\mu. \tag{2.25}$$

A general form of the Anscombe residual is then

$$r_{\text{A}_i} = \frac{A(y_i) - A(\hat{\mu}_i)}{A'(\hat{\mu}_i)\sqrt{V\left(\hat{\mu}_i\right)}}, \tag{2.26}$$

where $A'(\hat{\mu}_i)$ is the first derivative of $A(\mu_i)$ evaluated at $\hat{\mu}_i$.

The i^{th} standardized Anscombe residual $\dot{r}_{\text{A}_i}$ is

$$\dot{r}_{\text{A}_i} = \frac{r_{\text{A}_i}}{\sqrt{a\left(\hat{\phi}\right)(1-h_{ii})}}. \tag{2.27}$$

The Anscombe residual and its standardized form for the commonly used response distributions in the statistical modeling presented in Gill (2000) [9] are as follows.

Let $r_{\text{A\_N}i}, r_{\text{A\_Bin}i}, r_{\text{A\_Pois}i}$ and $r_{\text{A\_Gamma}i}$ denote the i^{th} Anscombe residual of the normal, binomial, Poisson and gamma distribution, respectively.

(1) Normal responses:

$$r_{\text{A\_N}i} = y_i - \hat{\mu}_i \qquad \text{and} \qquad \dot{r}_{\text{A\_N}i} = \frac{r_{\text{A\_N}i}}{\sqrt{s^2(1-h_{ii})}}.$$

(2) Binomial responses:

$$r_{\text{A\_Bin}i} = \sqrt{m_i}[\hat{\mu}_i(1-\hat{\mu}_i)]^{-1/6}\left[B\left(y_i, \frac{2}{3}, \frac{2}{3}\right) - B\left(\hat{\mu}_i, \frac{2}{3}, \frac{2}{3}\right)\right] \quad \text{and}$$

$$\dot{r}_{\text{A\_Bin}i} = \frac{r_{\text{A\_Bin}i}}{\sqrt{(1-h_{ii})}},$$

where $B(z,a,b) = \int_0^z t^{a-1}(1-t)^{b-1}\,dt.$

(3) Poisson responses:

$$r_{\text{A\_Pois}i} = \frac{3}{2}\left(y_i^{2/3}\hat{\mu}_i^{-1/6} - \hat{\mu}_i^{1/2}\right) \qquad \text{and} \qquad \dot{r}_{\text{A\_Pois}i} = \frac{r_{\text{A\_Pois}i}}{\sqrt{(1-h_{ii})}}.$$

(4) Gamma responses with $\mu_i = \alpha\beta_i$:

$$r_{\text{A\_Gamma}i} = 3\left[\left(\frac{y_i}{\hat{\mu}_i}\right)^{1/3} - 1\right] \qquad \text{and} \qquad \dot{r}_{\text{A\_Gamma}i} = \frac{r_{\text{A\_Gamma}i}(\sqrt{\hat{\alpha}})}{\sqrt{(1-h_{ii})}}.$$

2.2.3 Deviance residuals

Using the definition of scale deviance, $D^*(\boldsymbol{\mu};\boldsymbol{y})$ given in Macullagh and Nelder (1989) [16], where

$$\begin{aligned} D^*(\boldsymbol{\mu};\boldsymbol{y}) &= 2\times\{\ell(\boldsymbol{y};\boldsymbol{y}) - \ell(\boldsymbol{\mu};\boldsymbol{y})\} \\ &= 2\times\sum_{i=1}^{n}\{\ell(y_i;y_i) - \ell(\mu_i;y_i)\} \\ &= 2\times\sum_{i=1}^{n} d_i^*, \end{aligned} \tag{2.28}$$

gives the deviance function

$$D(\boldsymbol{\mu};\boldsymbol{y}) = a(\phi) \times D^*(\boldsymbol{\mu};\boldsymbol{y}).$$

A general form of the i^{th} deviance residual, r_{D_i} is defined by

$$\begin{aligned} r_{\text{D}_i} &= \text{sign}(y_i - \hat{\mu}_i)\sqrt{2 \times a\left(\hat{\phi}\right)\{\ell(y_i;y_i) - \ell(\hat{\mu}_i;y_i)\}} \\ &= \text{sign}(y_i - \hat{\mu}_i)\sqrt{d_i}\,, \end{aligned} \tag{2.29}$$

where $\text{sign}(y_i - \hat{\mu}_i)$ is a function that makes r_{D_i} positive when $y_i > \hat{\mu}_i$ and negative when $y_i < \hat{\mu}_i$ (Collett, 1991 [7]). The i^{th} standardized deviance residual, represented by $\dot{r}_{\text{D}_i}$, is

$$\dot{r}_{\text{D}_i} = \frac{r_{\text{D}_i}}{\sqrt{a\left(\hat{\phi}\right)(1 - h_{ii})}}. \tag{2.30}$$

Following we present the i^{th} deviance residual along with its corresponding standardized deviance residual for some well known glms:

(1) Normal responses:

From (2.19), we have the likelihood function defined by

$$L_{\text{N}}(\boldsymbol{\mu}, \sigma^2; \boldsymbol{y}) = \exp\left\{\sum_{i=1}^{n}\left[\frac{y_i\mu_i - \frac{1}{2}\mu_i^2}{\sigma^2} + \left(-\frac{y_i^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)\right)\right]\right\} \tag{2.31}$$

and the log-likelihood function is

$$\ell_{\text{N}} = \ell_{\text{N}}(\boldsymbol{\mu}, \sigma^2; \boldsymbol{y}) = \sum_{i=1}^{n}\left[\frac{y_i\mu_i - \frac{1}{2}\mu_i^2}{\sigma^2} + \left(-\frac{y_i^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)\right)\right]. \tag{2.32}$$

Considering the individual log-likelihood, $\ell_{\text{N}i}(\mu_i, \sigma^2; y_i)$;

$$\ell_{\text{N}i}(\mu_i, \sigma^2; y_i) = \frac{y_i\mu_i - \frac{1}{2}\mu_i^2}{\sigma^2} - \frac{y_i^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2), \tag{2.33}$$

then the i^{th} deviance residual, $r_{\text{D\_N}i}$ and its standardized deviance residual, $\dot{r}_{\text{D\_N}i}$ are

$$r_{\text{D\_N}i} = y_i - \hat{\mu}_i \qquad \text{and} \qquad \dot{r}_{\text{D\_N}i} = \frac{r_{\text{D\_N}i}}{\sqrt{s^2(1-h_{ii})}}.$$

(2) Binomial responses:

From (2.21), we have the likelihood function defined by

$$L_{\text{Bin}}(\boldsymbol{\mu}, \boldsymbol{m}; \boldsymbol{y}) = \exp\left\{\sum_{i=1}^{n}\left[y_i \ln\left(\frac{\mu_i}{m_i - \mu_i}\right) + m_i \ln\left(\frac{m_i - \mu_i}{m_i}\right) + \ln\binom{m_i}{y_i}\right]\right\} \tag{2.34}$$

and the log-likelihood function is

$$\ell_{\text{Bin}} = \ell_{\text{Bin}}(\boldsymbol{\mu}, \boldsymbol{m}; \boldsymbol{y}) = \sum_{i=1}^{n}\left[y_i \ln\left(\frac{\mu_i}{m_i - \mu_i}\right) + m_i \ln\left(\frac{m_i - \mu_i}{m_i}\right) + \ln\binom{m_i}{y_i}\right]. \tag{2.35}$$

The individual log-likelihood; $\ell_{\text{Bin}i}(\mu_i, m_i; y_i)$, is

$$\ell_{\text{Bin}i}(\mu_i, m_i; y_i) = y_i \ln\left(\frac{\mu_i}{m_i - \mu_i}\right) + m_i \ln\left(\frac{m_i - \mu_i}{m_i}\right) + \ln\binom{m_i}{y_i}, \tag{2.36}$$

then the i^{th} deviance residual, $r_{\text{D\_Bin}i}$ and its standardized deviance residual, $\dot{r}_{\text{D\_Bin}i}$ are

$$r_{\text{D\_Bin}i} = \text{sign}(y_i - \hat{\mu}_i)\left[2y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) + 2(m_i - y_i)\ln\left(\frac{m_i - y_i}{m_i - \hat{\mu}_i}\right)\right]^{1/2} \qquad \text{and}$$

$$\dot{r}_{\text{D\_Bin}i} = \frac{r_{\text{D\_Bin}i}}{\sqrt{(1-h_{ii})}}.$$

(3) Poisson responses:

From (2.22), we have the likelihood function defined by

$$L_{\text{Pois}}(\boldsymbol{\mu}; \boldsymbol{y}) = \exp\left\{\sum_{i=1}^{n}\left[y_i \ln \mu_i - \mu_i - \ln(y_i!)\right]\right\} \tag{2.37}$$

and the log-likelihood function is

$$\ell_{\text{Pois}} = \ell_{\text{Pois}}(\boldsymbol{\mu}; \boldsymbol{y}) = \sum_{i=1}^{n} \left[y_i \ln \mu_i - \mu_i - \ln (y_i!) \right]. \tag{2.38}$$

Considering the term of individual log-likelihood, $\ell_{\text{Pois}i}(\mu_i; y_i)$;

$$\ell_{\text{Pois}i}(\mu_i; y_i) = y_i \ln \mu_i - \mu_i - \ln (y_i!), \tag{2.39}$$

then the i^{th} deviance residual, $r_{\text{D\_Pois}i}$ and its standardized deviance residual, $\dot{r}_{\text{D\_Pois}i}$ are

$$r_{\text{D\_Pois}i} = \text{sign}(y_i - \hat{\mu}_i) \left[2y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - 2(y_i - \hat{\mu}_i) \right]^{1/2} \quad \text{and} \quad \dot{r}_{\text{D\_Pois}i} = \frac{r_{\text{D\_Pois}i}}{\sqrt{(1 - h_{ii})}}.$$

(4) Gamma responses:

From (2.24), we have the likelihood function defined by

$$L_{\text{Gamma}}(\boldsymbol{\mu}, \alpha; \boldsymbol{y}) = \exp \left\{ \sum_{i=1}^{n} \left[\frac{-y_i \mu_i^{-1} - \ln \mu_i + \ln \alpha}{\alpha^{-1}} + (\alpha - 1) \ln y_i - \ln \Gamma(\alpha) \right] \right\} \tag{2.40}$$

and the log-likelihood function is

$$\ell_{\text{Gamma}} = \ell_{\text{Gamma}}(\boldsymbol{\mu}, \alpha; \boldsymbol{y}) = \sum_{i=1}^{n} \left[\frac{-y_i \mu_i^{-1} - \ln \mu_i + \ln \alpha}{\alpha^{-1}} + (\alpha - 1) \ln y_i - \ln \Gamma(\alpha) \right]. \tag{2.41}$$

Considering the term of individual log-likelihood, $\ell_{\text{Gamma}i}(\mu_i, \alpha; y_i)$;

$$\ell_{\text{Gamma}i}(\mu_i, \alpha; y_i) = \frac{-y_i \mu_i^{-1} - \ln \mu_i + \ln \alpha}{\alpha^{-1}} + (\alpha - 1) \ln y_i - \ln \Gamma(\alpha), \tag{2.42}$$

then the i^{th} deviance residual, $r_{\text{D\_Gamma}i}$ and its standardized deviance residual, $\dot{r}_{\text{D\_Gamma}i}$ are

$$r_{\text{D\_Gamma}i} = \text{sign}(y_i - \hat{\mu}_i) \left[2 \left(\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right) - 2 \ln \left(\frac{y_i}{\hat{\mu}_i} \right) \right]^{1/2} \quad \text{and}$$

$$\dot{r}_{\text{D\_Gamma}i} = \frac{r_{\text{D\_Gamma}i} \sqrt{\hat{\alpha}}}{\sqrt{(1 - h_{ii})}}.$$

Since in the context of statistical modeling with known probability mass (density) functions, the linear predictor coefficients are mainly estimated by using the maximum likelihood method and the model selection along with model checking are performed by using the deviance residuals, therefore this thesis will mainly explore the association of the deviance residuals and the log-likelihood function in both theoretical derivation and empirical studies.

2.3 Glm log- likelihood functions and their corresponding deviance residuals

The expressions of the glm log-likelihood function presented in (2.11) and the deviance residual (2.29) show that the mode of the log-likelihood function occurs at $\mu_i = y_i$ or the deviance residual equals zero. This is clear from taking $\frac{\partial \ell_i}{\partial \theta_i} = 0$, which gives $y_i = b'(\theta_i) = \mu_i$ and then the maximum likelihood estimate of μ_i is $\hat{\mu}_i = y_i$. This derivation is true for all distribution mentioned above.

Following we explore the mode of the four distributional log-likelihood functions described in Section 2.2 by plotting calculated individual values of log-likelihood at given various values of μ via R (R Core Team, 2014 [18]). For the associated theoretical investigation, see Appendix A.

(1) Normal responses:

We explore the properties by calculating individual values of log-likelihood (2.33), here denoted by, $\ell_{\mathrm{N}y} = \ell_{\mathrm{N}i}$, for given values of $y = 1, 10, 40$ and three different values of $\sigma^2 = 0.5, 2.0, 5.5$ over a grid of $\mu = 0.10, 0.11, 0.12, ..., y$ and $\mu = 0.10, 0.11, 0.12, ..., 2y$. We then plot $\ell_{\mathrm{N}y|\mu}$ against μ for each σ^2 and superimposed by the straight line with coordinates: $\mu = (0.1, y)$ and $\ell_{\mathrm{N}y} = \big(\ell_{\mathrm{N}y|\mu=0.1}, \ell_{\mathrm{N}y|\mu=y}\big)$. The corresponding displayed in Figures 2.1, 2.2 and 2.3 show that the mode of $\ell_{\mathrm{N}y|\mu}$ is found at $\hat{\mu} = y$ implying that the deviance residual, $r_{\mathrm{D\_N}i}$, at the mode is zero.

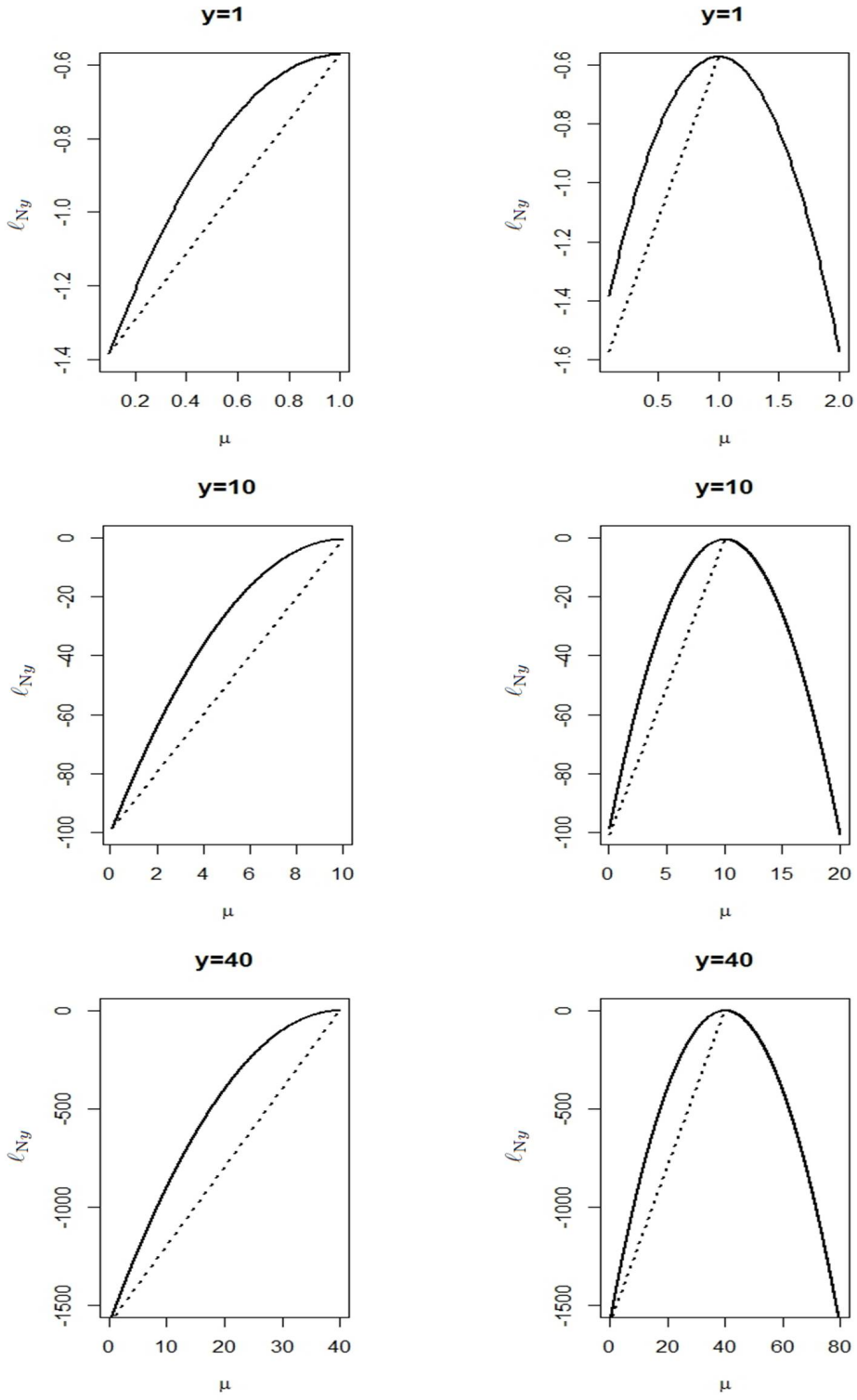


Figure 2.1: Mode of $\ell_{\mathrm{N}y}$ for $\sigma^2 = 0.5$.

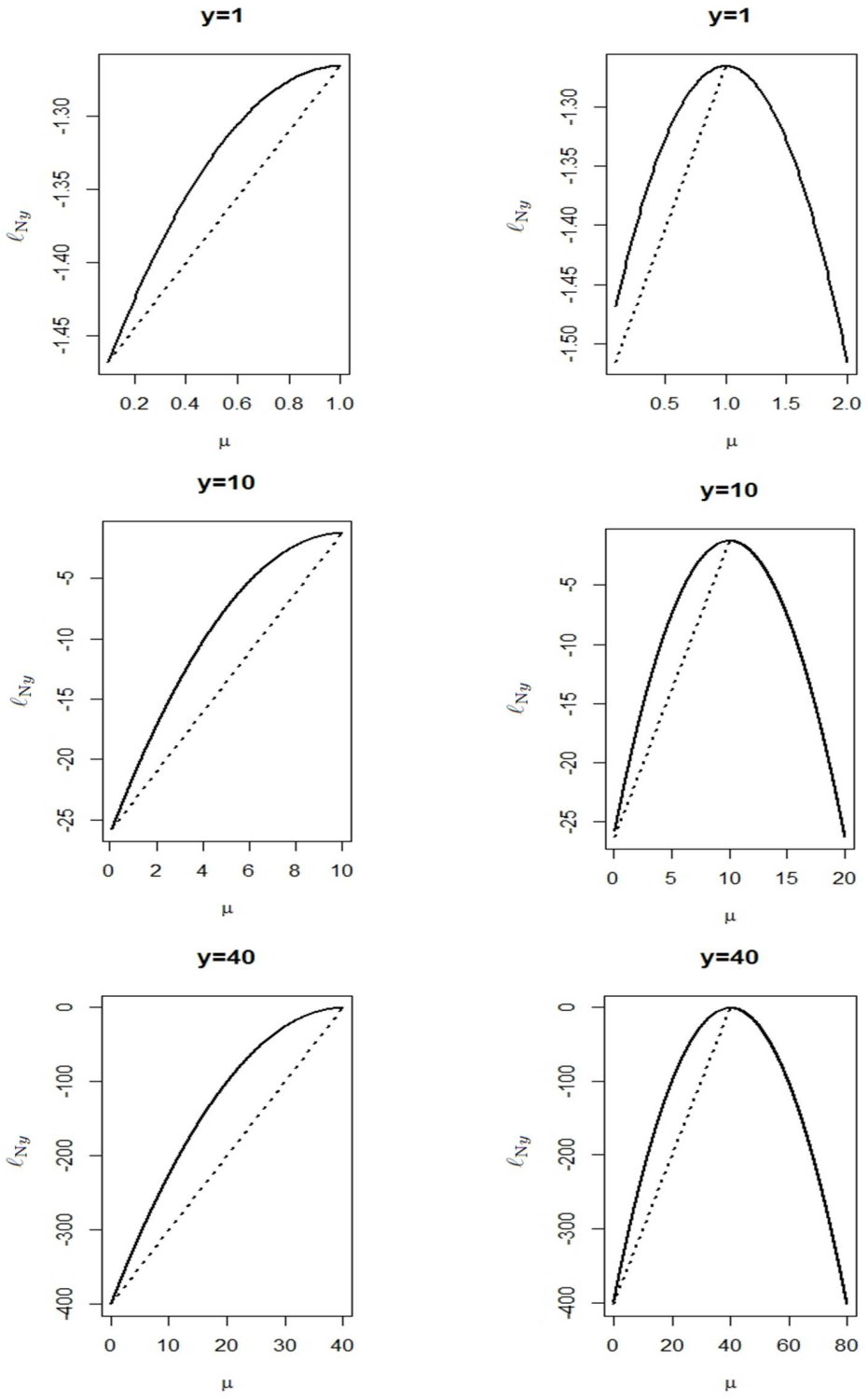


Figure 2.2: Mode of $\ell_{\mathrm{N}y}$ for $\sigma^2 = 2$.

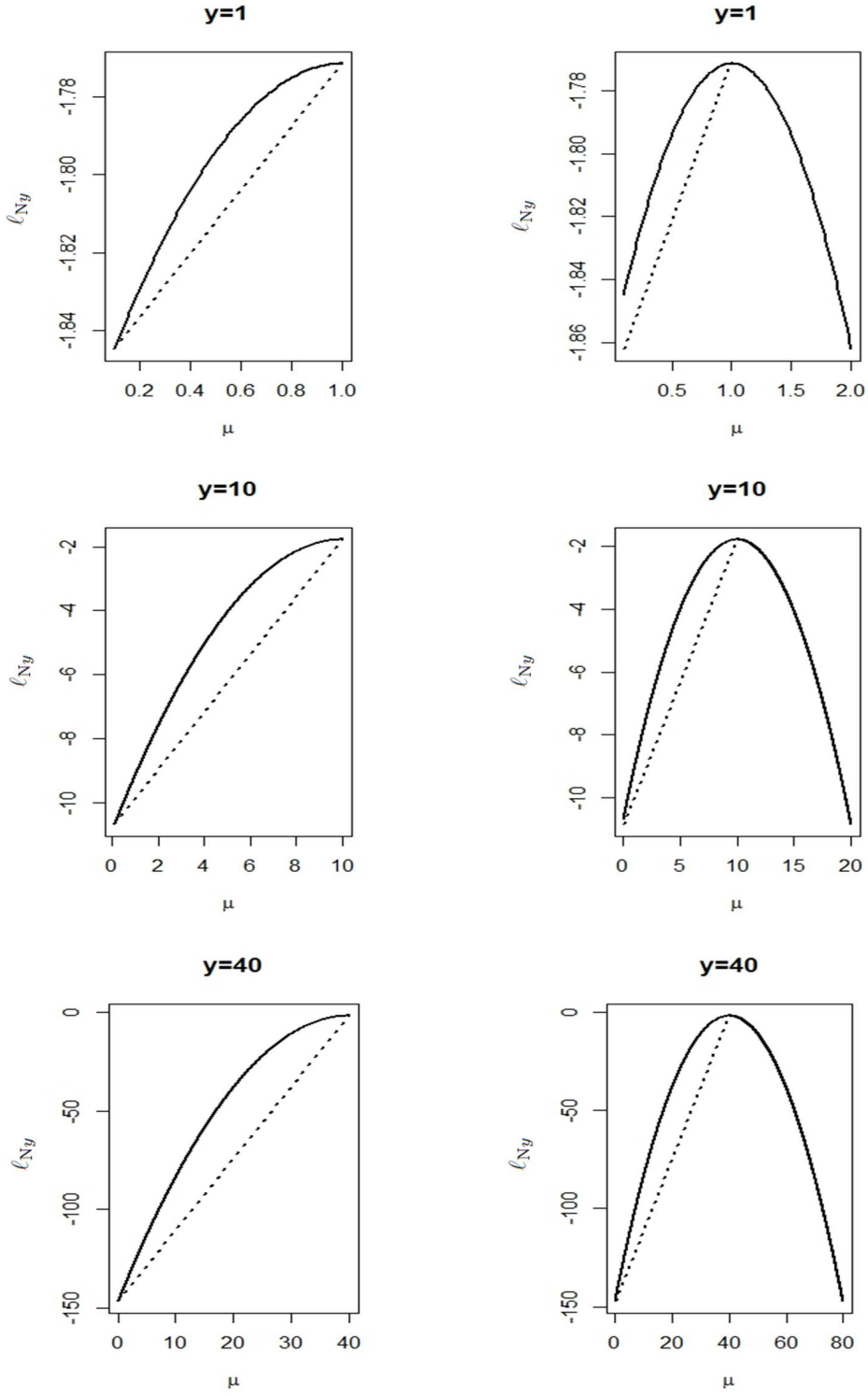


Figure 2.3: Mode of ℓ_{Ny} for $\sigma^2 = 5.5$.

(2) Binomial responses:

We explore the properties of the binomial log-likelihood by calculating individual values of log-likelihood (2.36), denoted by $\ell_{\text{Bin}y} = \ell_{\text{Bin}i}$, for three sets of $y = \{1, 3, 5\}$ for number of trial, $m = 10$; $y = \{3, 5, 10\}$ for $m = 20$ and $y = \{10, 15, 20\}$ for $m = 40$ over a grid of values $\mu = 0.10, 0.11, 0.12, ..., y$ and $\mu = 0.10, 0.11, 0.12, ..., 2y - 1$ (subtracted by 1 to avoid a missing value). We plot $\ell_{\text{Bin}y|\mu}$ against μ for each m, superimposed by the straight line with coordinates: $\mu = (0.1, y)$ and $\ell_{\text{Bin}y} = \left(\ell_{\text{Bin}y|\mu=0.1}, \ell_{\text{Bin}y|\mu=y}\right)$. The corresponding plots presented in Figures 2.4 to 2.6 show that the mode of $\ell_{\text{Bin}y|\mu}$ is found at $\hat{\mu} = y$ implying that the binomial deviance residual, $r_{\text{D\_Bin}i}$, at the mode is zero.

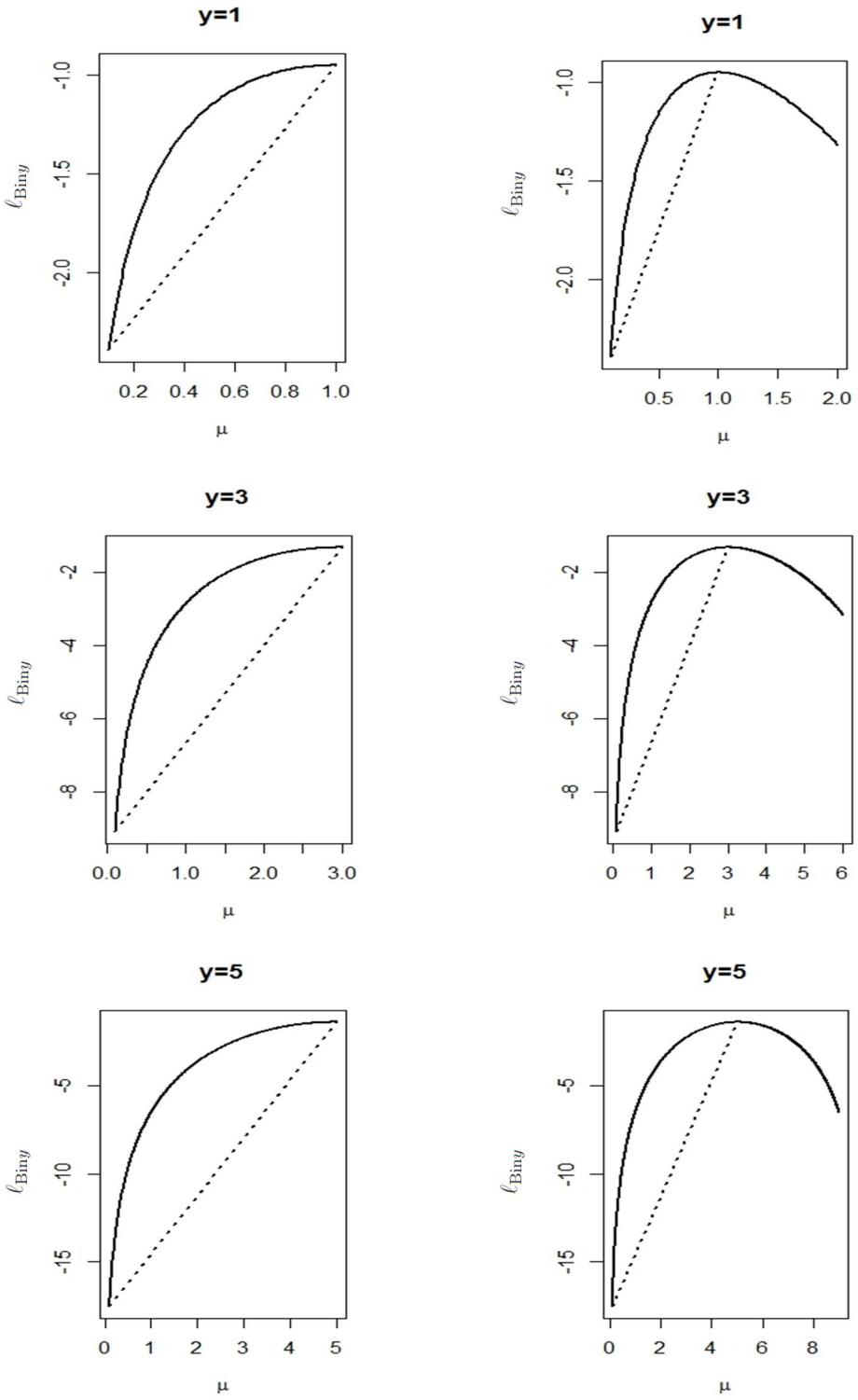


Figure 2.4: Mode of $\ell_{\text{Bin}y}$ for $m = 10$.

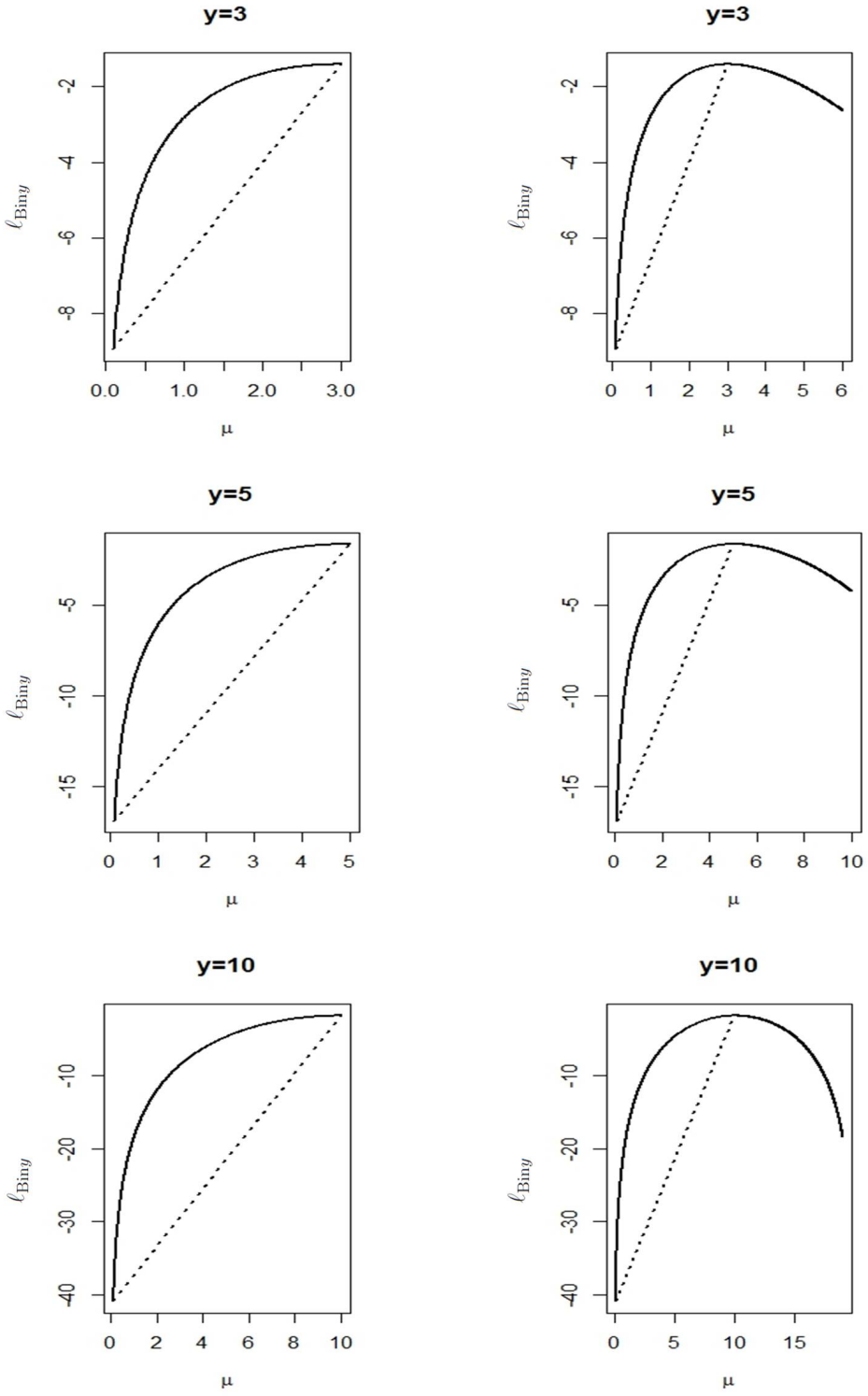


Figure 2.5: Mode of $\ell_{\text{Bin}y}$ for $m = 20$.

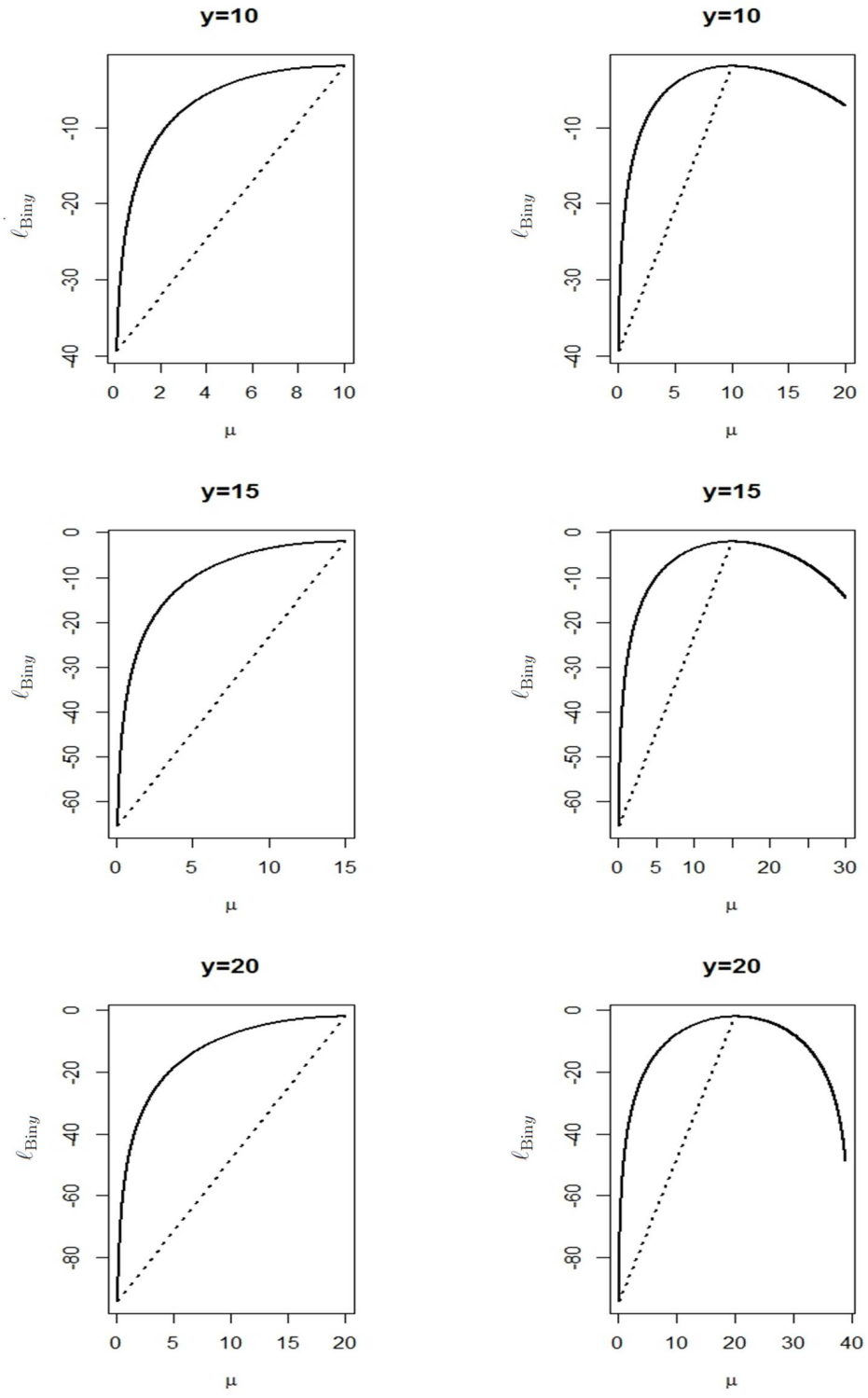


Figure 2.6: Mode of ℓ_{Biny} for $m = 40$.

(3) Poisson responses:

We explore the properties of the Poisson log-likelihood by calculating individual values of log-likelihood (2.39), denoted by $\ell_{\text{Pois}y} = \ell_{\text{Pois}i}$ for $y = 1, 5, 10, 20, 40, 60$ over a grid of values $\mu = 0.10, 0.11, 0.12, ..., y$ and $\mu = 0.10, 0.11, 0.12, ..., 2y$, then plotting $\ell_{\text{Pois}y|\mu}$ against μ, superimposed by the straight lines with coordinates: $\mu = (0.1, y)$ and $\ell_{\text{Pois}y} = \left(\ell_{\text{Pois}y|\mu=0.1}, \ell_{\text{Pois}y|\mu=y}\right)$. The corresponding plots presented in Figure 2.7-2.8 show that the mode of $\ell_{\text{Pois}y|\mu}$ is found at $\hat{\mu} = y$ implying that the Poisson deviance residual, $r_{\text{D\_Pois}i}$, at the mode is zero.

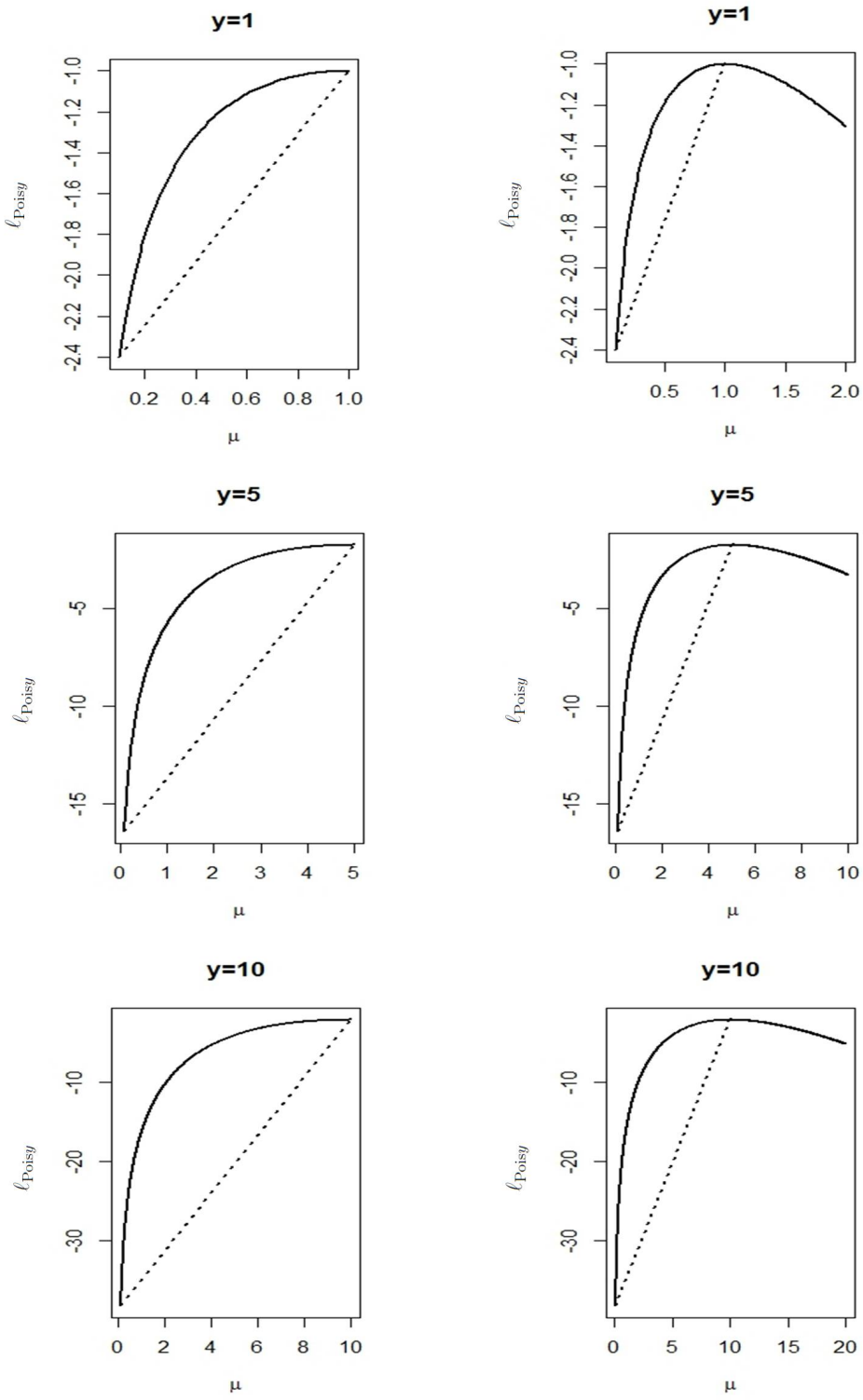


Figure 2.7: Mode of ℓ_{Poisy} at low values of y.

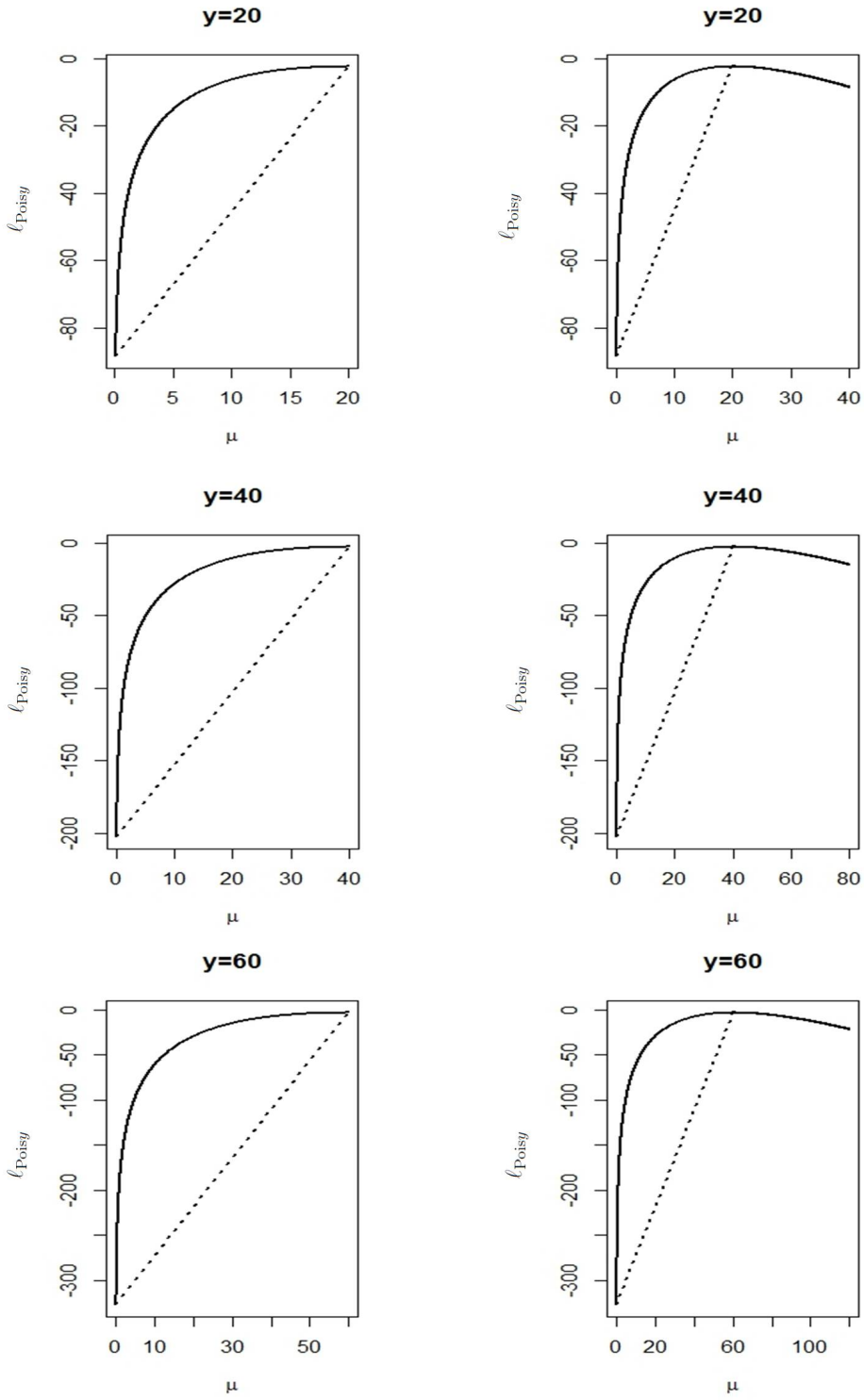


Figure 2.8: Mode of ℓ_{Poisy} at middle to high values of y.

(4) Gamma responses:

The properties of the gamma log-likelihood are explored by calculating individual values of log-likelihood (2.42), denoted by $\ell_{\text{Gamma}y} = \ell_{\text{Gamma}i}$ for $y = 1, 10, 40$ and three different values of $\alpha = 0.5, 1.5, 4.5$ over a grid of values $\mu = \alpha\beta$ $\left(\text{where } \beta = 0.10, 0.11, 0.12, ..., \frac{y}{\alpha} \text{ and } \beta = 0.10, 0.11, 0.12, ..., \frac{2y}{\alpha}\right)$. Then we plot the calculated $\ell_{\text{Gamma}y|\mu}$ against μ for each α, superimposed the straight line with two coordinates: $\mu = (\min(\mu), y)$ and $\ell_{\text{Gamma}y}$ $=(\ell_{\text{Gamma}y|\mu=\min(\mu)}, \ell_{\text{Gamma}y|\mu=y})$. The corresponding plots displayed in Figures 2.9 to 2.11 show that the mode of $\ell_{\text{Gamma}y|\mu}$ is found at $\hat{\mu} = y$ implying that the deviance residual, $r_{\text{D\_Gamma}i}$, at the mode is zero.

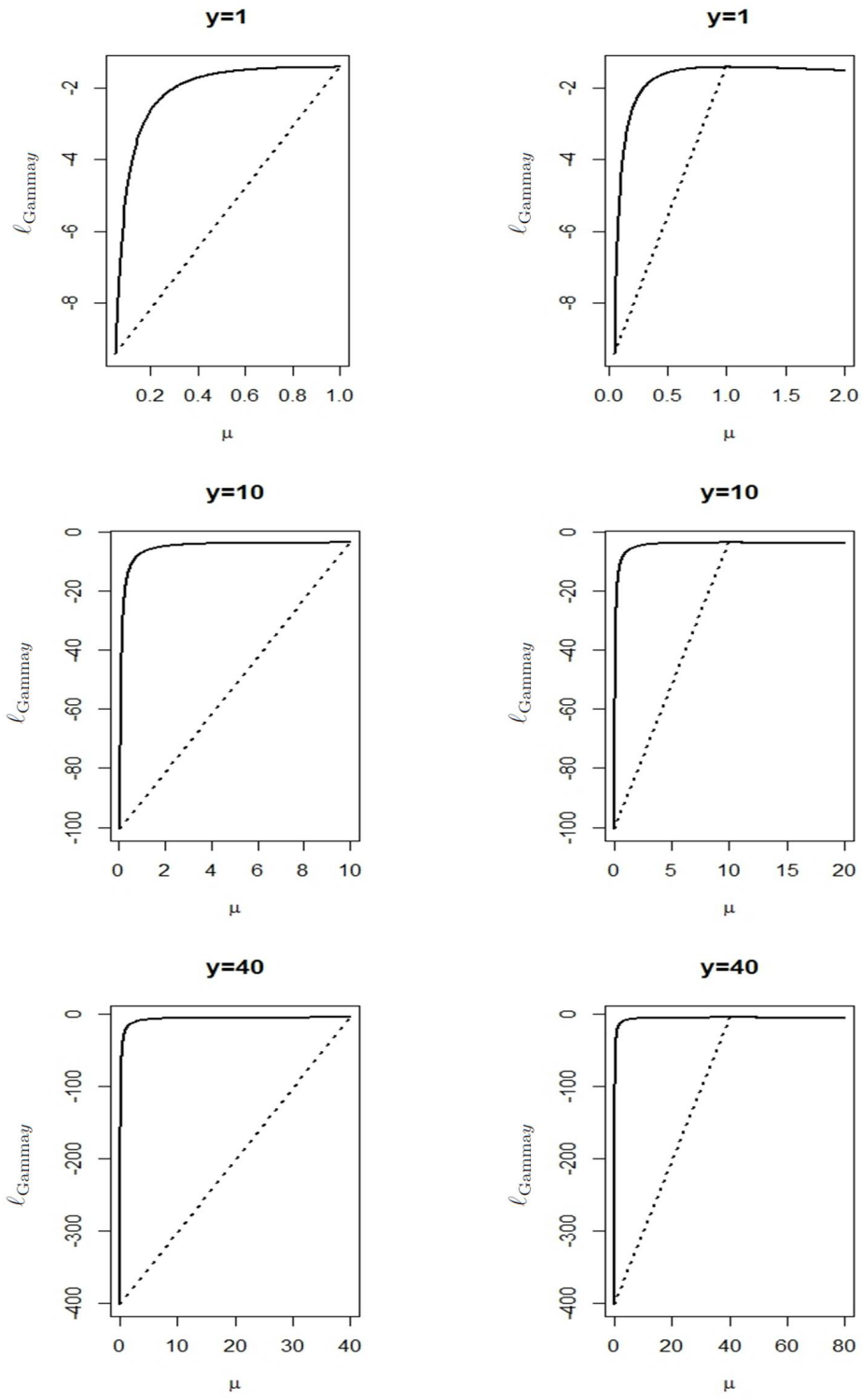


Figure 2.9: Mode of $\ell_{\text{Gamma}y}$ for $\alpha = 0.5$.

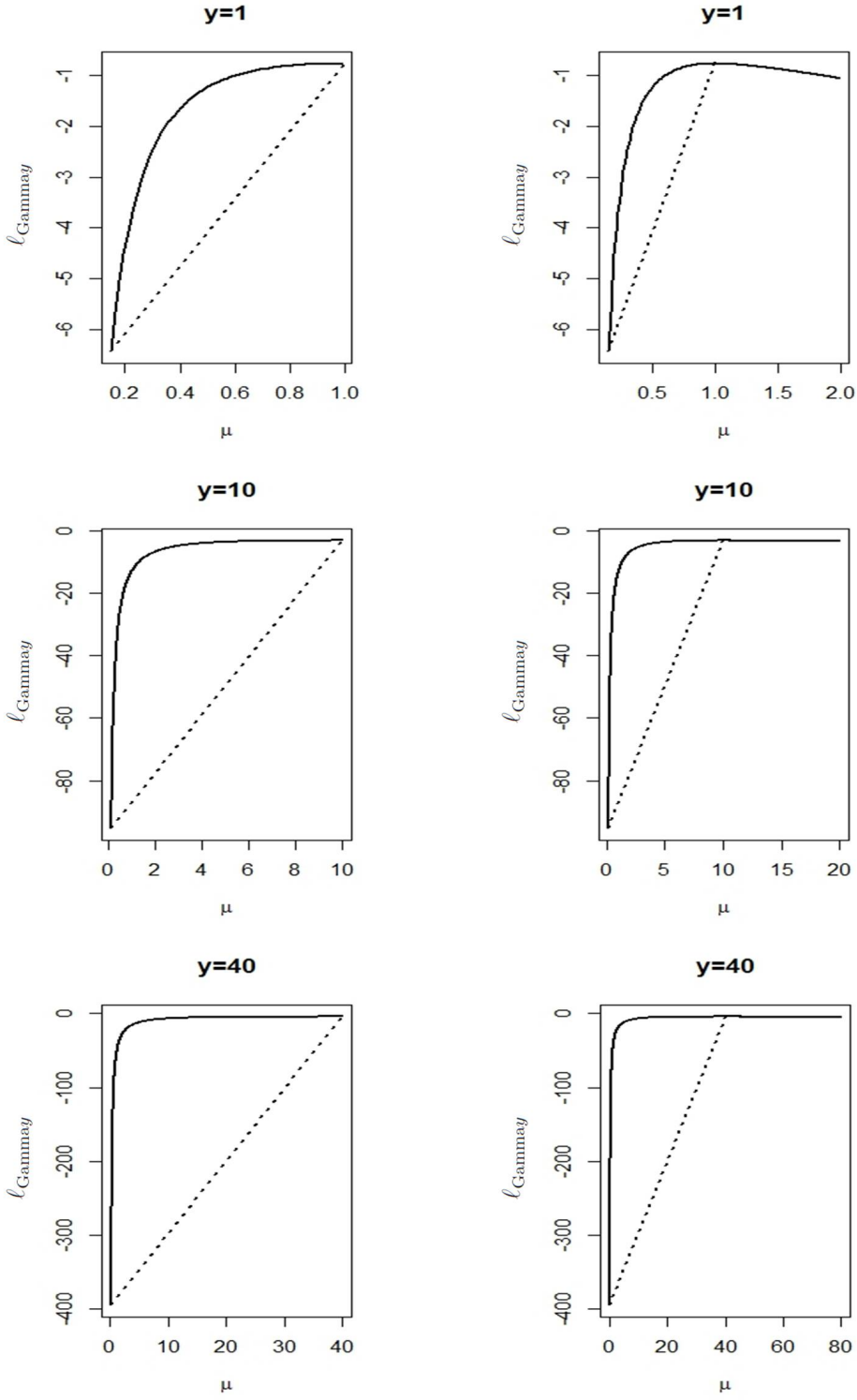


Figure 2.10: Mode of ℓ_{Gammay} for $\alpha = 1.5$.

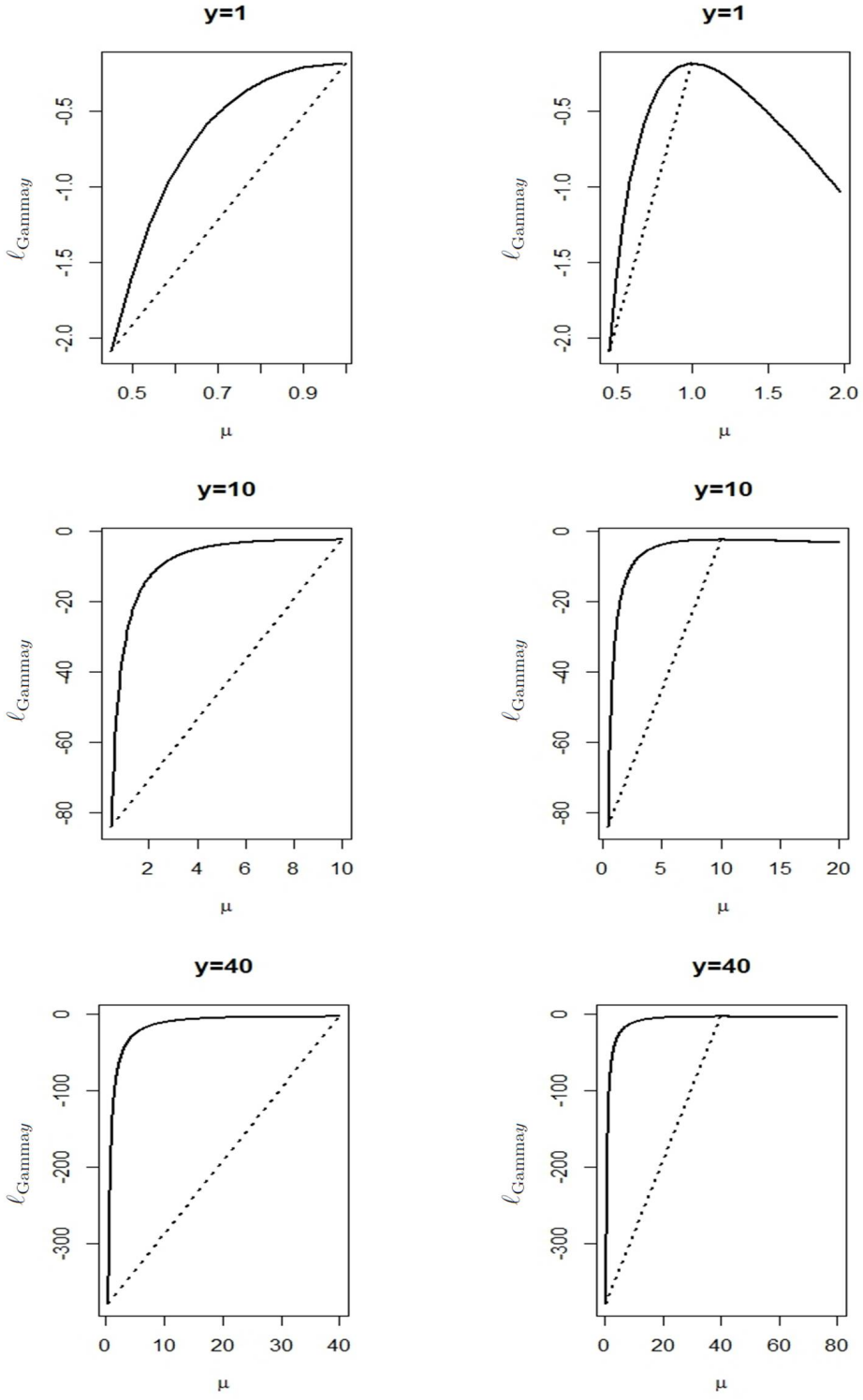


Figure 2.11: Mode of ℓ_{Gammay} for $\alpha = 4.5$.

CHAPTER 3

Deviance Residuals of Linear Mean-Variance Negative Binomial Models

So far we have explored the relationship of the deviance residuals and the log-likelihood of the well known glm responses. In this chapter, we will focus on exploring the background of the deviance residual of a linear mean-variance negative binomial model used in Jansakul and Hinde (2004) [13].

3.1 Negative binomial models

Even though the log linear or Poisson regression provides a standard model for count data, the assumption that the mean and variance are identical, does not hold in all situations. In practice, the phenomenon of the more variation; variance exceeding the mean, commonly occurs and then regression modeling ideas need to be extended to overdispersed count modeling. Various models have been proposed to tackle this situation, the negative binomial (NB) family, in particular, linear and quadratic mean-variance negative binomial models are more commonly used as all important statistical inferences can be carried out more easily and conveniently than the other types (Lawless, 1987 [15] and Jansakul and Hinde, 2004 [13]). Details of the model estimation (both methodologies and applications using SAS statistical software) are given in Hilbe (2007) [10]. Applying these two

NB models to practical data can be found in for example, Byers *et al.* (2003) [5], Jansakul and Hinde (2004) [13] and Elhi *et al.* (2007) [8].

Here we will firstly describe a general form of NB probability mass function and separate that to the two well known distributions followed by their diagnostic quantities used in model checking.

If $Y_i,\ i = 1, 2, ..., n$ are negative binomial distributed counts with mean μ_i and dispersion parameter α, denoted by $Y_i \sim \text{NB}(\mu_i, \alpha)$, a general form of the p.m.f can be defined by

$$f(y_i; \mu_i, \alpha) = \begin{cases} \dfrac{\Gamma(y_i + \alpha^{-1}\mu_i^{1-\tau})}{y_i!\Gamma(\alpha^{-1}\mu_i^{1-\tau})} \dfrac{\alpha^{y_i}\mu_i^{\tau y_i}}{(1+\alpha\mu_i^{\tau})^{y_i+\alpha^{-1}\mu_i^{1-\tau}}} & , y_i = 0, 1, ...\ ;\ \ \alpha > 0 \\ 0 & , \text{otherwise} \end{cases} \tag{3.1}$$

with $\text{E}(Y_i) = \mu_i$ and $\text{Var}(Y_i) = \mu_i(1 + \alpha\mu_i^{\tau})$, where α is assumed to be a constant. The index τ identifies various forms of the NB model, but two well-known models are the linear mean-variance NB model given by taking $\tau = 0$ and the more usual quadratic mean-variance NB model given by taking $\tau = 1$. The former model is commonly denoted by NB1 model, where the latter is represented by NB2 model. As $\alpha \to 0$, the NB model reduces to the Poisson model. In the context of modeling for the mean, the log link is assumed, i.e. $\ln(\mu_i) = \boldsymbol{x}_i^{\text{T}}\boldsymbol{\beta}$, where $\boldsymbol{x}_i^{\text{T}} = (1, x_{i1}, x_{i2}, ..., x_{ip})^{\text{T}}$ is a vector of $n \times (p+1)$ explanatory variables and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, ..., \beta_p)^{\text{T}}$ is a $p + 1$ vector of unknown parameters. The parameter estimates for both NB1 and NB2 models can be easily obtained using a full Newton-Raphson method, as in Jansakul and Hinde (2004) [13] for the NB1 model

and in Lawless (1987) [15] for the NB2 model. Based on the p.m.f of the NB model (3.1), a Newton-Raphson for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$ at the $(m+1)^{\text{th}}$ iteration is; for example,

$$\begin{bmatrix} \boldsymbol{\beta}^{(m+1)} \\ \boldsymbol{\alpha}^{(m+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}^{(m)} \\ \boldsymbol{\alpha}^{(m)} \end{bmatrix} + \left[\boldsymbol{I}^{(m)}\right]^{-1} \boldsymbol{s}^{(m)}. \tag{3.2}$$

Here $\boldsymbol{I}^{(m)}$ is a $(p+1)\times(p+1)$ observed information matrix, which is defined by

$$\boldsymbol{I}^{(m)} = \boldsymbol{I}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \begin{bmatrix} \boldsymbol{I}_{\beta\beta}(\boldsymbol{\beta},\boldsymbol{\alpha}) & \boldsymbol{I}_{\beta\alpha}(\boldsymbol{\beta},\boldsymbol{\alpha}) \\ \boldsymbol{I}_{\alpha\beta}(\boldsymbol{\beta},\boldsymbol{\alpha}) & \boldsymbol{I}_{\alpha\alpha}(\boldsymbol{\beta},\boldsymbol{\alpha}) \end{bmatrix},$$

where $\boldsymbol{I}_{\beta\beta} = -\dfrac{\partial^2\ell}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}$ is the $p\times p$ symmetric matrix, $\boldsymbol{I}_{\alpha\alpha} = -\dfrac{\partial^2\ell}{\partial\boldsymbol{\alpha}^2}$ is a scalar and $\boldsymbol{I}_{\alpha\beta} = \boldsymbol{I}_{\beta\alpha}^{\mathrm{T}} = -\dfrac{\partial^2\ell}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}}$ is a $1\times(p+1)$ matrix. $\boldsymbol{s}^{(m)}$ is the score vector defined by

$$\boldsymbol{s}^{(m)} = \boldsymbol{s}(\beta,\alpha) = \begin{bmatrix} \boldsymbol{s}_\beta(\beta,\alpha) \\ \boldsymbol{s}_\alpha(\beta,\alpha) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial\ell}{\partial\beta} \\ \dfrac{\partial\ell}{\partial\alpha} \end{bmatrix}.$$

$\boldsymbol{s}^{(m)}$ and $\boldsymbol{I}^{(m)}$ are evaluated at $\boldsymbol{\beta} = \boldsymbol{\beta}^{(m)}$ and $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(m)}$. The iteration (3.2) must be carried out until convergence, which can be assessed using a stopping rule such as

$$\left|\boldsymbol{\alpha}^{(m+1)} - \boldsymbol{\alpha}^{(m)}\right| < \epsilon \qquad \text{or} \qquad \left|\ell^{(m+1)} - \ell^{(m)}\right| < \epsilon.$$

However, codes or functions for fitting the NB2 regression model are available in most statistical software; R, SPLUS, STATA, SAS, etc., since it is a generalized linear model, when the shape or dispersion parameter is known (Hinde and Demetrio, 1998 [11]).

Checking if the NB2 model is a glm, we let $Y_i \sim \text{NB2}(\mu_i, \alpha)$, where the p.m.f is specified by

$$f(y_i;\mu_i,\alpha) = \begin{cases} \dfrac{\Gamma(y_i+\alpha^{-1})}{y_i!\Gamma(\alpha^{-1})}\dfrac{\alpha^{y_i}\mu_i^{y_i}}{(1+\alpha\mu_i)^{y_i+\alpha^{-1}}} & ,y_i = 0,1,\ldots;\ \ \alpha>0 \\ 0 & ,\text{otherwise.} \end{cases} \tag{3.3}$$

Checking whether (3.3) is a glm, we rewrite the p.m.f in an exponential form:

$$f(y_i;\mu_i,\alpha) = \exp\left\{y_i \ln\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right) - \alpha^{-1}\ln(1+\alpha\mu_i) + \text{dlg}(y_i,\alpha^{-1}) - \ln(y_i!)\right\}, \tag{3.4}$$

where $\text{dlg}(y_i,\alpha^{-1})$ is represents the difference of two log gamma functions for any a defined as

$$\text{dlg}(y_i,a) = \ln\Gamma(y_i+a) - \ln\Gamma(a) = \begin{cases} \displaystyle\sum_{t=0}^{y_i-1}\ln(a+t) & ,y_i>0 \\ 0 & ,y_i=0. \end{cases} \tag{3.5}$$

Taking $\theta_i = \ln\left(\dfrac{\alpha\mu_i}{1+\alpha\mu_i}\right)$, then $\mu_i = \dfrac{\alpha^{-1}\exp(\theta_i)}{1-\exp(\theta_i)}$ and the distribution of Y_i can be written in the canonical form:

$$f(y_i;\theta_i,\alpha) = \exp\left\{y_i\theta_i - \alpha^{-1}\ln\left(\frac{1}{1-\exp(\theta_i)}\right) + \text{dlg}(y_i,\alpha^{-1}) - \ln(y_i!)\right\}, \tag{3.6}$$

which is in the linear exponential family with the canonical parameter, $\theta_i = \ln\left(\dfrac{\alpha\mu_i}{1+\alpha\mu_i}\right)$, $a(\phi)=1$, $c(y_i,\alpha) = \text{dlg}(y_i,\alpha^{-1}) - \ln(y_i!)$ and $b(\theta_i) = \alpha^{-1}\ln\left(\dfrac{1}{1-\exp(\theta_i)}\right)$. That is NB2 distribution is a glm having $\text{E}(Y_i) = b'(\theta_i) = \mu_i$ and $\text{Var}(Y_i) = a(\phi)b''(\theta_i) = \mu_i(1+\alpha\mu_i)$, where the variance function is $V(\mu_i) = \mu_i(1+\alpha\mu_i)$ and the canonical link function is $g(\mu_i) = \ln\left(\dfrac{\alpha\mu_i}{1+\alpha\mu_i}\right)$. However, in practice the link function

commonly used is $g(\mu_i) = \ln(\mu_i)$. This follows the basic Poisson regression model. The likelihood function for giving a vector of n observations $\boldsymbol{y} = (y_i, y_2, ..., y_n)^{\mathrm{T}}$ is

$$L_{\mathrm{NB2}}(\boldsymbol{\mu}, \alpha; \boldsymbol{y}) = \exp\left\{\sum_{i=1}^{n}\left[y_i \ln\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right) - \alpha^{-1}\ln(1+\alpha\mu_i) + \mathrm{dlg}(y_i, \alpha^{-1}) - \ln(y_i!)\right]\right\} \tag{3.7}$$

and the log-likelihood function is

$$\ell_{\mathrm{NB2}} = \ell_{\mathrm{NB2}}(\boldsymbol{\mu}, \alpha; \boldsymbol{y}) = \sum_{i=1}^{n}\left[y_i \ln\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right) - \alpha^{-1}\ln(1+\alpha\mu_i) + \mathrm{dlg}(y_i, \alpha^{-1}) - \ln(y_i!)\right]. \tag{3.8}$$

Checking whether the NB1 model is a glm, let $Y_i \sim \mathrm{NB1}(\mu_i, \alpha)$, where its p.m.f is specified by

$$f(y_i; \mu_i, \alpha) = \begin{cases} \dfrac{\Gamma(y_i + \alpha^{-1}\mu_i)}{y_i!\Gamma(\alpha^{-1}\mu_i)} \dfrac{\alpha^{y_i}}{(1+\alpha)^{y_i+\alpha^{-1}\mu_i}} & , y_i = 0, 1, ...\,; \ \ \alpha > 0 \\ 0 & , \text{otherwise.} \end{cases} \tag{3.9}$$

Again, we rewrite the p.m.f in an exponential form:

$$f(y_i; \mu_i, \alpha) = \exp\left\{y_i \ln\left(\frac{\alpha}{1+\alpha}\right) - \alpha^{-1}\mu_i \ln(1+\alpha) + \mathrm{dlg}(y_i, \alpha^{-1}\mu_i) - \ln(y_i!)\right\}. \tag{3.10}$$

Since $\theta_i = \ln\left(\dfrac{\alpha}{1+\alpha}\right)$ is not a function of μ_i and $c(y_i, \alpha) = \mathrm{dlg}(y_i, \alpha^{-1}\mu_i) - \ln(y_i!)$ includes μ_i, then NB1 distribution is not a glm. The likelihood function; $L_{\mathrm{NB1}}(\boldsymbol{\mu}, \alpha; \boldsymbol{y})$ and the log-likelihood function; $\ell_{\mathrm{NB1}}(\boldsymbol{\mu}, \alpha; \boldsymbol{y})$ for NB1 are respectively defined by

$$L_{\mathrm{NB1}}(\boldsymbol{\mu}, \alpha; \boldsymbol{y}) = \exp\left\{\sum_{i=1}^{n}\left[y_i \ln\left(\frac{\alpha}{1+\alpha}\right) - \alpha^{-1}\mu_i \ln(1+\alpha) + \mathrm{dlg}(y_i, \alpha^{-1}\mu_i) - \ln(y_i!)\right]\right\} \tag{3.11}$$

and

$$\ell_{\mathrm{NB1}} = \ell_{\mathrm{NB1}}(\boldsymbol{\mu}, \alpha; \boldsymbol{y}) = \sum_{i=1}^{n}\left[y_i \ln\left(\frac{\alpha}{1+\alpha}\right) - \alpha^{-1}\mu_i \ln(1+\alpha) + \mathrm{dlg}(y_i, \alpha^{-1}\mu_i) - \ln(y_i!)\right]. \tag{3.12}$$

Based on the maximum likelihood estimation, we have $\mathrm{E}(Y_i) = \mu_i$ and $\mathrm{Var}(Y_i) = \mu_i(1+\alpha)$.

3.2 Deviance Residuals of NB2 and NB1 Models

This section we present the NB1 and NB2 deviance residuals and report the exploration for the mode of their log-likelihoods. We firstly focus on the NB2 log-likelihood since the NB2 model is a glm and the expression of its observed and standardized deviance residuals are already presented in associated literatures.

3.2.1 NB2 Deviance Residuals

Following the definition of deviance residuals given in Subsection 2.2.3 and using the log-likelihood function in (3.8), the individual observed and standardized deviance residual of NB2 model, denoted by $r_{\mathrm{D\_NB2}i}$ and $\dot{r}_{\mathrm{D\_NB2}i}$, respectively, are

$$r_{\mathrm{D\_NB2}i} = \mathrm{sign}(y_i - \mu_i)\sqrt{2\left[y_i \ln\left(\frac{y_i}{\mu_i}\right) - \left(y_i + \frac{1}{\alpha}\right)\ln\left(\frac{1+\alpha y_i}{1+\alpha\mu_i}\right)\right]} \tag{3.13}$$

and

$$\dot{r}_{\mathrm{D\_NB2}i} = \frac{r_{\mathrm{D\_NB2}i}}{\sqrt{1-h_{ii}}}, \tag{3.14}$$

where h_{ii} is the i^{th} diagonal element of the hat matrix, $\boldsymbol{H} = \boldsymbol{W}^{\frac{1}{2}}\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}^{\frac{1}{2}}$ and $\boldsymbol{W}$ is an $n \times n$ diagonal matrix with the i^{th} diagonal element $w_{ii} = \dfrac{\mu_i}{1+\alpha\mu_i}$. The equality of μ_i and y_i gives $r_{\mathrm{D\_NB2}i} = 0$ and hence $\dot{r}_{\mathrm{D\_NB2}i} = 0$.

3.2.2 The Mode of the NB2 Log-likelihood

To investigate whether the mode of (3.8) occurs at $\hat{\mu}_i = y_i$, we consider the individual log-likelihood component denoted by $\ell_{\mathrm{NB2}y} = \ell_{\mathrm{NB2}i}$, take

the first derivative of $\ell_{\text{NB2}y}$ with respect to μ and equate that to 0 as follows:

$$\ell_{\text{NB2}y} = y \ln\left(\frac{\alpha\mu}{1+\alpha\mu}\right) - \alpha^{-1}\ln(1+\alpha\mu) + \text{dlg}(y, \alpha^{-1}) - \ln(y_i!)\,, \tag{3.15}$$

$$\begin{aligned}
\frac{\partial \ell_{\text{NB2}y}}{\partial \mu} = \frac{y(1+\alpha\mu)}{\alpha\mu}\left[\frac{(1+\alpha\mu)\alpha - \alpha\mu(\alpha)}{(1+\alpha\mu)^2}\right] - \frac{1}{\alpha}\left(\frac{\alpha}{1+\alpha\mu}\right) &= 0 \\
\frac{y}{\mu(1+\alpha\mu)} - \frac{1}{(1+\alpha\mu)} &= 0 \\
\frac{y-\mu}{\mu(1+\alpha\mu)} &= 0 \\
\hat{\mu} &= y.
\end{aligned}$$

Hence the mode of the log-likelihood function is $\hat{\mu} = y$ and not a function of α. This implies that the NB2 deviance residual, $r_{\text{D\_NB2}i}$ in (3.13), is zero at $\hat{\mu} = y$ or the mode of $\ell_{\text{NB2}y}$.

We explore the mode of $\ell_{\text{NB2}y}$ by calculating $\ell_{\text{NB2}y}$ excluding $\ln(y!)$ for $y = 1, 10, 40$ and three different values of $\alpha = 0.5, 1.5, 4.5$ over a grid of values $\mu = 0.10, 0.11, 0.12, ..., y$ and $\mu = 0.10, 0.11, 0.12, ..., 2y$. We plot the calculated $\ell_{\text{NB2}y|\mu}$ against μ for each α, superimposed by the straight line with coordinates: $\mu = (0.1, y)$ and $\ell_{\text{NB2}y} = \left(\ell_{\text{NB2}y|\mu=0.1}, \ell_{\text{NB2}y|\mu=y}\right)$. These plots displayed in Figures 3.1 - 3.3 show that the mode of $\ell_{\text{NB2}y|\mu}$ at the various values of α is found at $\hat{\mu} = y$ as we had in the above proof.

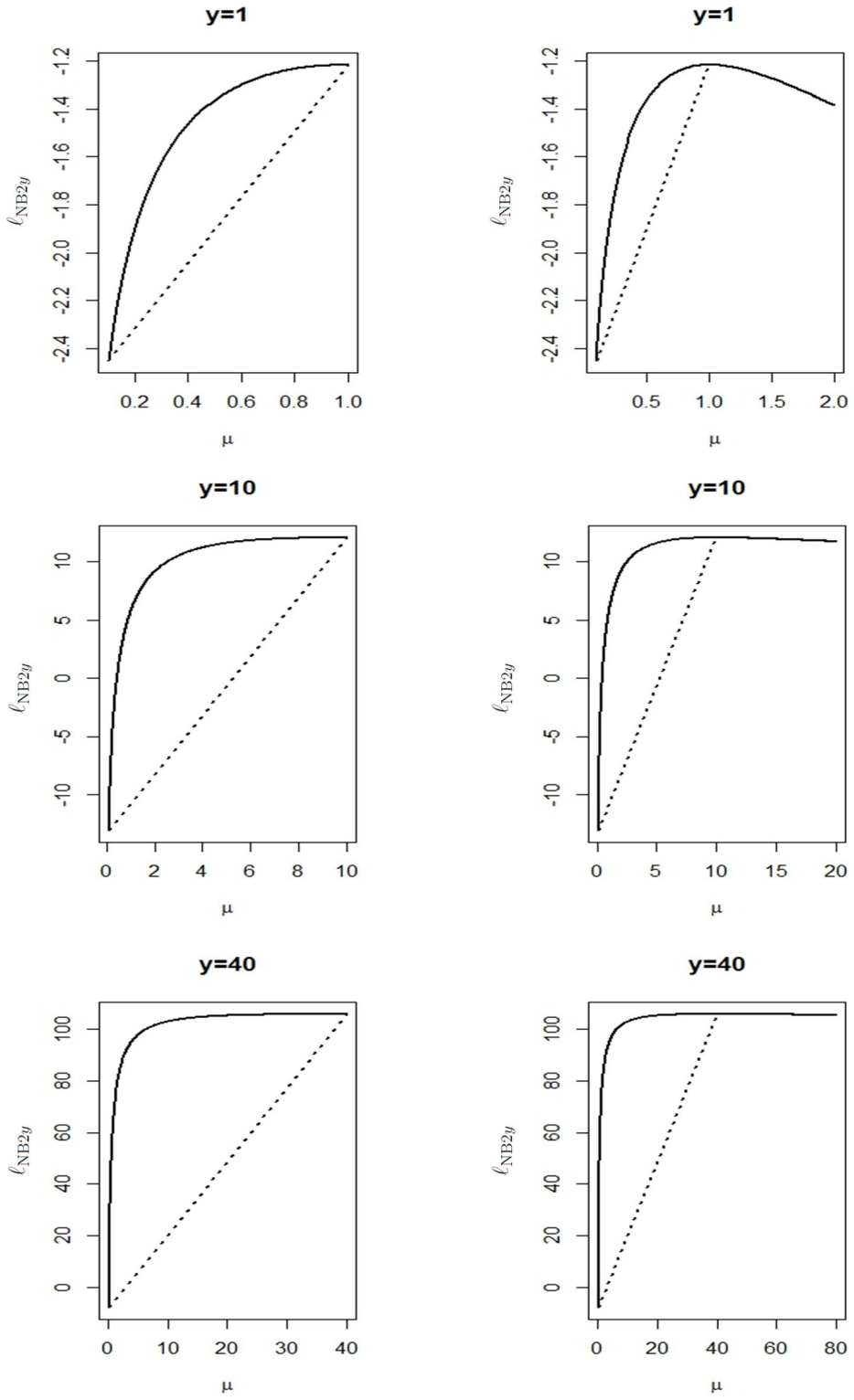


Figure 3.1: Mode of $\ell_{\mathrm{NB2}y}$ for $\alpha = 0.5$.

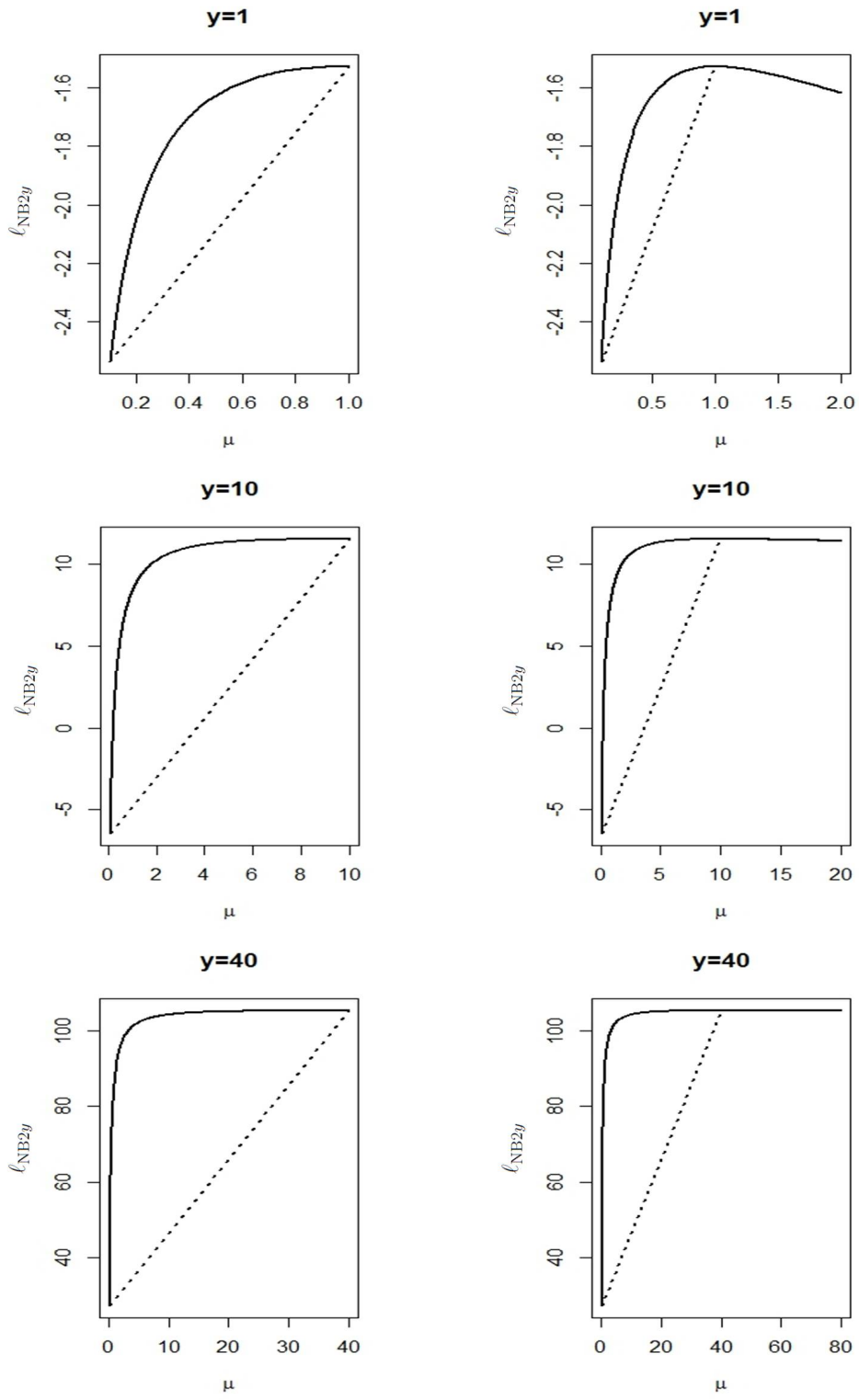


Figure 3.2: Mode of $\ell_{\mathrm{NB}2y}$ for $\alpha = 1.5$.

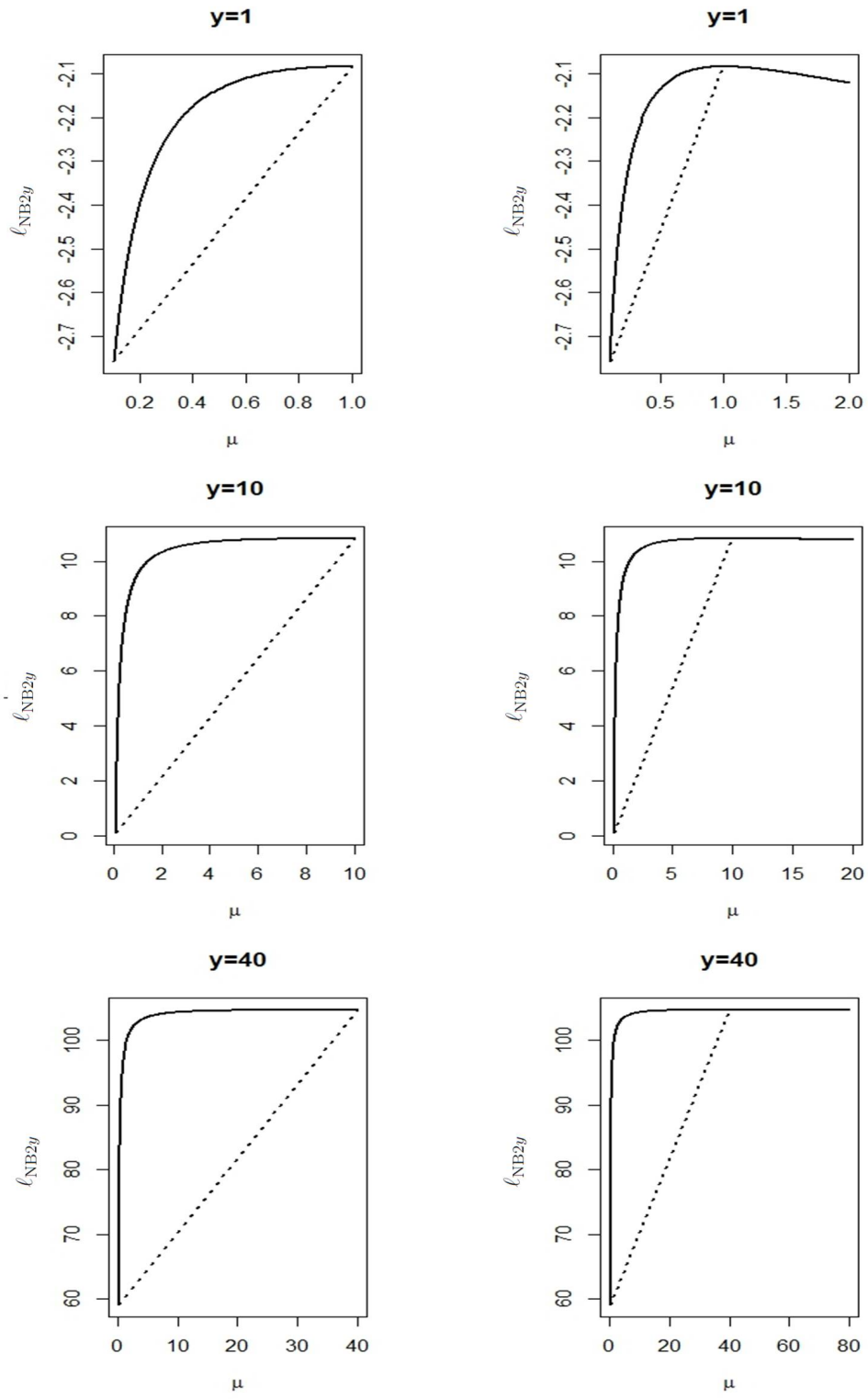


Figure 3.3: Mode of $\ell_{\mathrm{NB}2y}$ for $\alpha = 4.5$.

3.2.3 An NB1 Deviance Residual

Although the NB1 regression model is not a member of glms, its complete p.m.f and hence the log-likelihood function can be defined, see equation (3.12). Based on the usual deviance residuals expression, we have the deviance function:

$$\begin{aligned}
\mathrm{D}(\boldsymbol{\mu}, \boldsymbol{y}) &= 2\left\{\ell_{\mathrm{NB1}}(\boldsymbol{y}, \alpha; \boldsymbol{y}) - \ell_{\mathrm{NB1}}(\boldsymbol{\mu}, \alpha; \boldsymbol{y})\right\} \\
&= 2\sum_{i=1}^{n}\left\{y_i \ln\left(\frac{\alpha}{1+\alpha}\right) - \alpha^{-1} y_i \ln(1+\alpha) + \ln\Gamma(y_i + \alpha^{-1} y_i) - \ln\Gamma(\alpha^{-1} y_i)\right\} \\
&\quad +2\sum_{i=1}^{n}\left\{-\ln(y_i!) - y_i \ln\left(\frac{\alpha}{1+\alpha}\right) + \alpha^{-1}\mu_i \ln(1+\alpha) - \ln\Gamma(y_i + \alpha^{-1}\mu_i)\right\} \\
&\quad +2\sum_{i=1}^{n}\left\{\ln\Gamma(\alpha^{-1}\mu_i) + \ln(y_i!)\right\} \\
&= -2\sum_{i=1}^{n}\left\{\alpha^{-1}(y_i - \mu_i)\ln(1+\alpha) - \ln\Gamma(y_i + \alpha^{-1} y_i) + \ln\Gamma(\alpha^{-1} y_i)\right\} \\
&\quad -2\sum_{i=1}^{n}\left\{\ln\Gamma(y_i + \alpha^{-1}\mu_i) - \ln\Gamma(\alpha^{-1}\mu_i)\right\} \\
&= -2\sum_{i=1}^{n}\left\{(y_i - \mu_i)\frac{\ln(1+\alpha)}{\alpha} - \mathrm{dlg}(y_i, \alpha^{-1} y_i) + \mathrm{dlg}(y_i, \alpha^{-1}\mu_i)\right\}.
\end{aligned}$$

This shows that a simple form of NB1 deviance residual cannot be obtained. Nelder (1991) [17], cited in Jansakul and Hinde (2004) [13], pointed out that the individual log-likelihood of (3.12) does not have the property that its mode occurs at $\mu_i = y_i$ if $y_i \neq 0$. He suggested that the mode of $\ell_{\mathrm{NB1}i}$ can be approximately

evaluated at $y_i + \frac{1}{2}$ and the i^{th} approximated deviance component for y_i;

$$\begin{cases} \dfrac{2\mu_i \ln(1+\alpha)}{\alpha} & , y_i = 0 \\ -2 \times \left\{ (y_i - \mu_i) \dfrac{\ln(1+\alpha)}{\alpha} - \text{dlg}(y_i, \alpha^{-1} y_i) + \text{dlg}(y_i, \alpha^{-1}\mu_i) \right\} & , y_i > 0 \end{cases} \tag{3.16}$$

can be replaced by

$$\begin{cases} \dfrac{2\mu_i \ln(1+\alpha)}{\alpha} & , y_i = 0 \\ -2 \times \left\{ \left(y_i + \dfrac{1}{2} - \mu_i \right) \dfrac{\ln(1+\alpha)}{\alpha} - \text{dlg}\left(y_i, \alpha^{-1} \left(y_i + \dfrac{1}{2} \right) \right) + \text{dlg}(y_i, \alpha^{-1}\mu_i) \right\} & , y_i > 0. \end{cases} \tag{3.17}$$

Jansakul and Hinde (2004) [13] claimed that their exploration (not reported) found that $y_i + \frac{1}{2}$ is not the appropriate approximated mode of $\ell_{\text{NB1}i}$. They suggested values, such as $y_i + k$, where $\frac{\alpha}{2 + \frac{1}{y_i}} < k < \frac{\alpha}{2}$ are likely to be close to giving the mode and for large y_i, $k \approx \frac{\alpha}{2}$ works well.

3.2.4 The Mode of the NB1 Log-likelihood

Permitted by Jansakul and Hinde (2004) [13], we develop the exploration for the approximated mode of $\ell_{\text{NB1}i}$. We started by considering the first derivative of the individual log-likelihood component $\ell_{\text{NB1}i}$ with respect to μ_i,

α and equating that to 0. From the equation (3.12), we use the full definition of

$$\mathrm{dlg}(y_i, a) = \ln\Gamma(y_i + a) - \ln\Gamma(a) = \begin{cases} \sum_{t=0}^{y_i-1} \ln(a+t) & , y_i > 0 \\ 0 & , y_i = 0 \end{cases}$$

we have

$$\ell_{\mathrm{NB1}i} = y_i \ln\left(\frac{\alpha}{1+\alpha}\right) - \alpha^{-1}\mu_i \ln(1+\alpha) + \sum_{t=0}^{y_i-1} \ln\left(\alpha^{-1}\mu_i + t\right) - \ln\left(y_i!\right), \quad (3.18)$$

$$\begin{aligned} \frac{\partial \ell_{\mathrm{NB1}i}}{\partial \mu_i} = -\frac{\ln(1+\alpha)}{\alpha} + \frac{1}{\alpha}\sum_{t=0}^{y_i-1}\left(\alpha^{-1}\mu_i + t\right)^{-1} &= 0 \\ \alpha^{-1}\left[-\ln(1+\alpha) + \sum_{t=0}^{y_i-1}\left(\alpha^{-1}\mu_i + t\right)^{-1}\right] &= 0 \\ \left[-\ln(1+\alpha) + \sum_{t=0}^{y_i-1}\left(\alpha^{-1}\mu_i + t\right)^{-1}\right] &= 0 \end{aligned} \quad (3.19)$$

and

$$\begin{aligned} \frac{\partial \ell_{\mathrm{NB1}i}}{\partial \alpha} = \frac{y_i - \mu_i}{\alpha(1+\alpha)} + \alpha^{-2}\mu_i \ln(1+\alpha) - \alpha^{-2}\mu_i \sum_{t=0}^{y_i-1}\left(\alpha^{-1}\mu_i + t\right)^{-1} &= 0 \\ \alpha^{-2}\left[\frac{y_i - \mu_i}{\alpha^{-1}+1} + \mu_i \ln(1+\alpha) - \mu_i \sum_{t=0}^{y_i-1}\left(\alpha^{-1}\mu_i + t\right)^{-1}\right] &= 0 \\ \left[\frac{y_i - \mu_i}{\alpha^{-1}+1} + \mu_i \ln(1+\alpha) - \mu_i \sum_{t=0}^{y_i-1}\left(\alpha^{-1}\mu_i + t\right)^{-1}\right] &= 0. \end{aligned} \quad (3.20)$$

Considering $\ell_{\mathrm{NB1}i}$ in (3.18) and the proof; (3.19) - (3.20), we have $\hat{\mu}_i = y_i$, if $y_i = 0$. However for $y_i > 0$, the maximum likelihood estimates of μ_i, i.e. the mode of $\ell_{\mathrm{NB1}i}$ is a function of both y_i and α. We explore this property by calculating $\ell_{\mathrm{NB1}y} = \ell_{\mathrm{NB1}i}$ excluding $\ln(y!)$ at $y = 1, 3, 5, 10, 20, 40$ and three different values of $\alpha = 0.5, 1.5, 4.5$ over a grid of values $\mu = 0.10, 0.11, 0.12, ..., 2y$. Then plot

$\ell_{\text{NB1}y}$ against μ for each α, superimposed by two straight lines: the first line is drawn at $\mu = (0.1, y)$ and $\ell_{\text{NB1}y} = \left(\ell_{\text{NB1}y|\mu=0.1}, \ell_{\text{NB1}y|\mu=y}\right)$ and the other is drawn at $\mu = \left(0.1, y + \frac{1}{2}\right)$ and $\ell_{\text{NB1}y} = \left(\ell_{\text{NB1}y|\mu=0.1}, \ell_{\text{NB1}y|\mu=y+\frac{1}{2}}\right)$. The corresponding plots shown in Figures 3.4 to 3.6 indicate that the mode of $\ell_{\text{NB1}y}$ does not occur at either $\mu = y$ or $\mu = y + \frac{1}{2}$ for all studied values of y or α.

In order to search an appropriate value of k (in a function of α) mentioned in Section 3.2.3, we calculate $\ell_{\text{NB1}y} = \ell_{\text{NB1}y}(y+k, \alpha, y)$ excluding $\ln(y!)$ at $y = 1, 3, 5, 10, 20, 40$ and three different values of $\alpha = 0.5, 1.5, 4.5$ over a grid of values $k = 0.05, 0.06, ..., 5$ and plot $\ell_{\text{NB1}y}$ against k for each α. The corresponding plots are shown in Figures 3.7 to 3.9. The plots indicate that the approximation $y + \frac{1}{2}$ suggested by Nelder (1991) [17] is not adequate for all values of y. To investigate this further, we search for k, such that $y+k$ can be used to approximate the maximum likelihood estimate of μ, by calculating $\ell_{\text{NB1}y}$ and $\left|\frac{\partial \ell_{\text{NB1}y}}{\partial \mu}\right|$ for $y = 1, 3, 5, 10, 20, 40$ over a grid of values for $k = \alpha\{0.1, 0.2, 0.3, ..., 10\}$ as shown in the contour plots in Figures 3.10 and 3.11, respectively. Superimposed on these plots are three straight lines of the linear relationships between k and α, for fixed values of y. The plots show that there is no simple form for k, but values such that $\frac{\alpha}{2 + \frac{1}{y}} < k < \frac{\alpha}{2}$ are likely to be close to giving the mode of $\ell_{\text{NB1}y}$ and for large y, $k \approx \frac{\alpha}{2}$ works well as mentioned in Jansakul and Hinde (2004) [13]. Hence their approximated individual NB1 deviance residual is $r_{\text{D\_NB1}i} = \text{sign}(y_i - \hat{\mu}_i)\sqrt{\text{D}_{\text{NB1}i}}$

where

$$\mathrm{D}_{\mathrm{NB1}i} = \begin{cases} \dfrac{2\hat{\mu}_i \ln(1+\hat{\alpha})}{\hat{\alpha}} & , y_i = 0 \\ -2 \times \left\{ \left(y_i + \dfrac{\hat{\alpha}}{2} - \hat{\mu}_i \right) \dfrac{\ln(1+\hat{\alpha})}{\hat{\alpha}} - \mathrm{dlg}\left(y_i, \hat{\alpha}^{-1} \left(y_i + \dfrac{\hat{\alpha}}{2} \right) \right) \right\} & \\ -2 \times \mathrm{dlg}(y_i, \hat{\alpha}^{-1}\hat{\mu}_i) & , y_i > 0. \end{cases} \tag{3.21}$$

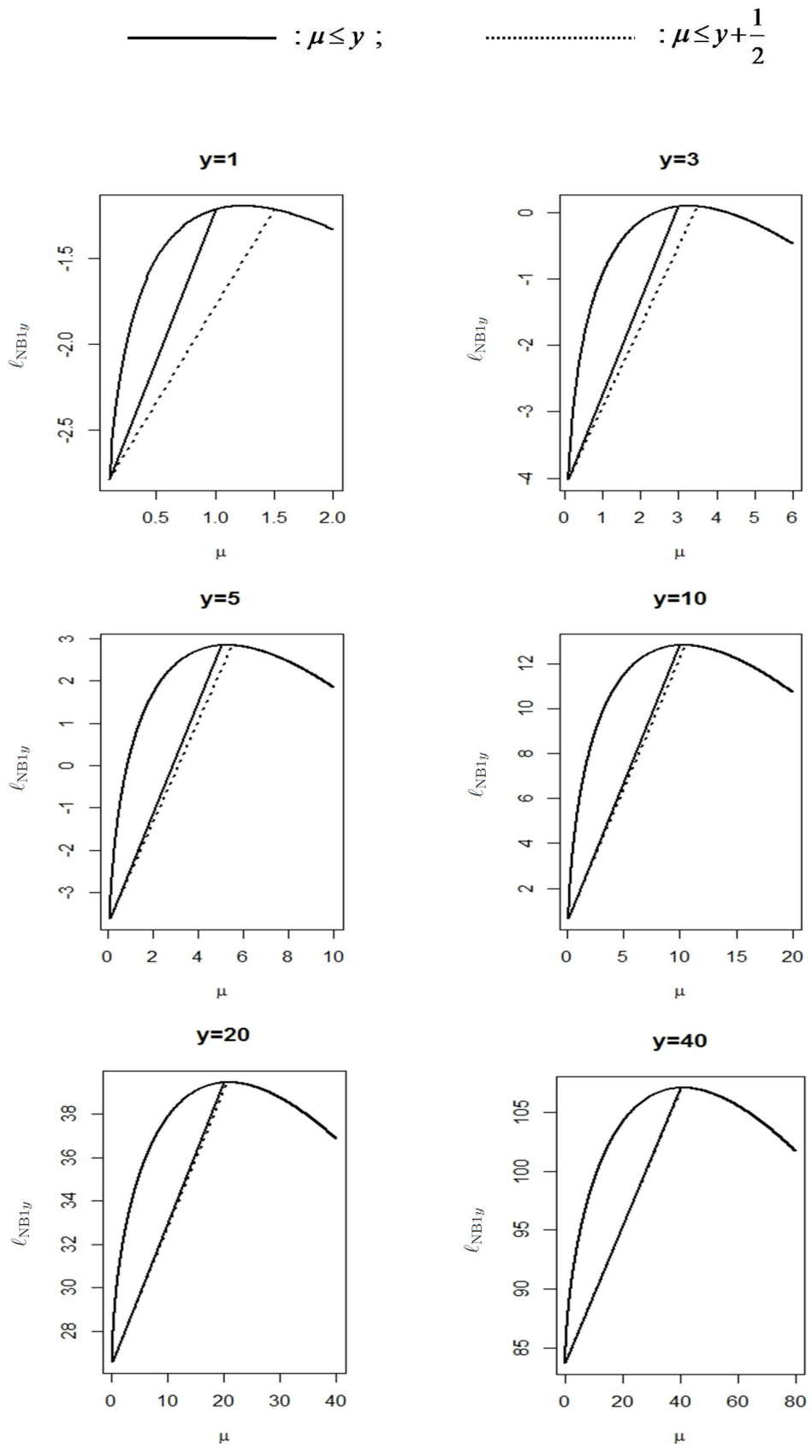


Figure 3.4: Mode of $\ell_{\mathrm{NB1}y}$ for $\alpha = 0.5$.

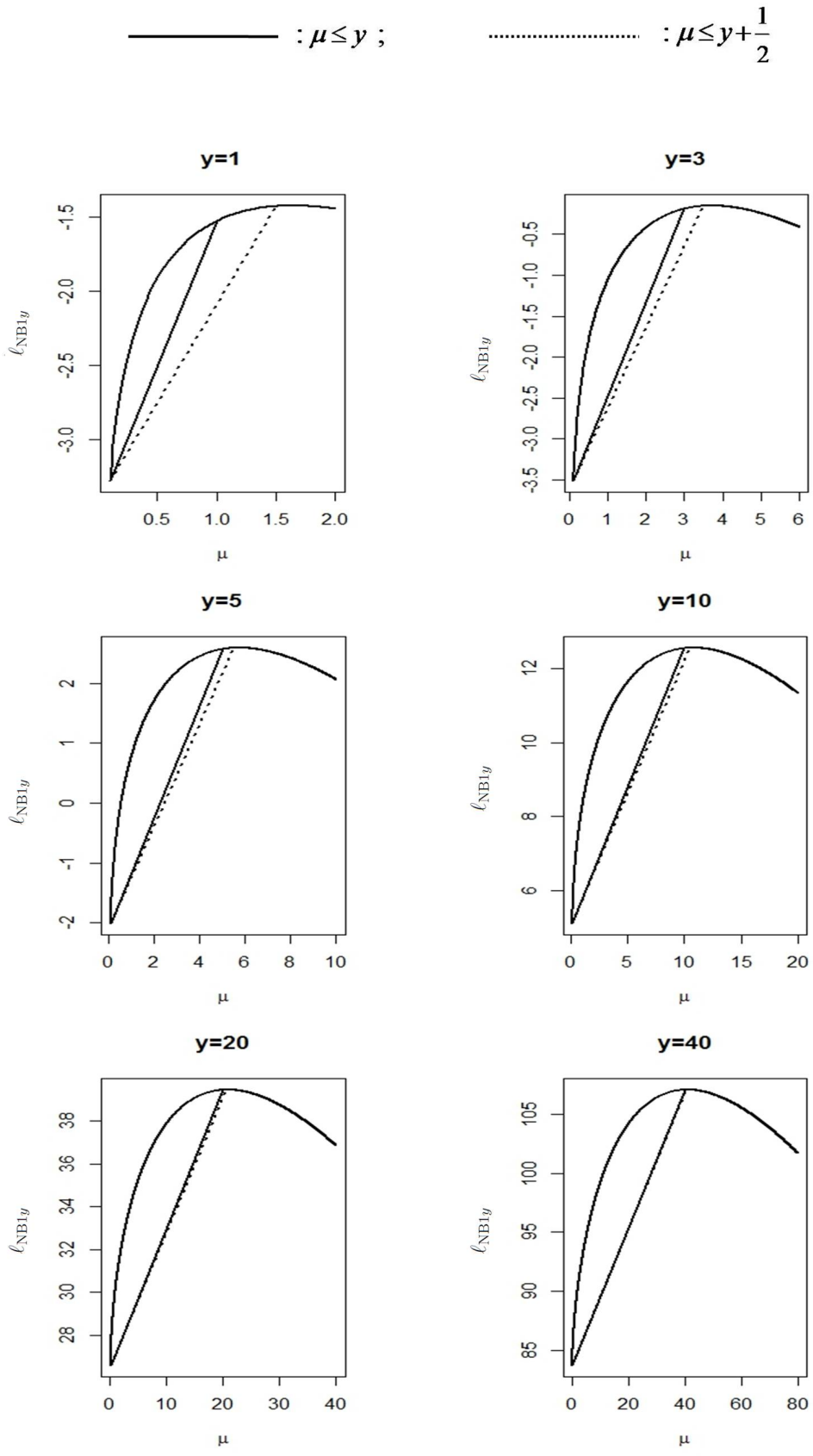


Figure 3.5: Mode of $\ell_{\mathrm{NB1}y}$ for $\alpha = 1.5$.

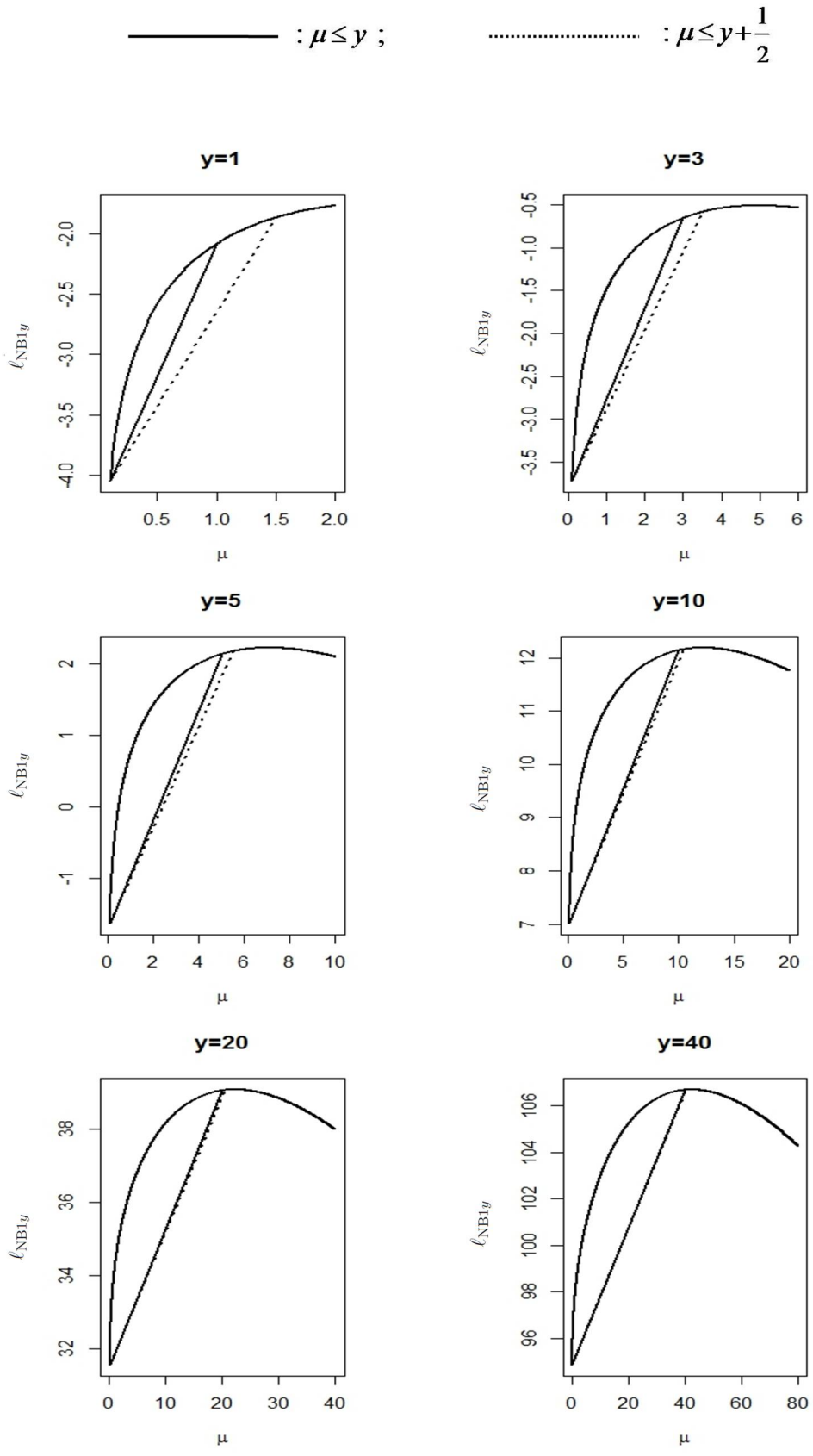


Figure 3.6: Mode of $\ell_{\mathrm{NB1}y}$ for $\alpha = 4.5$.

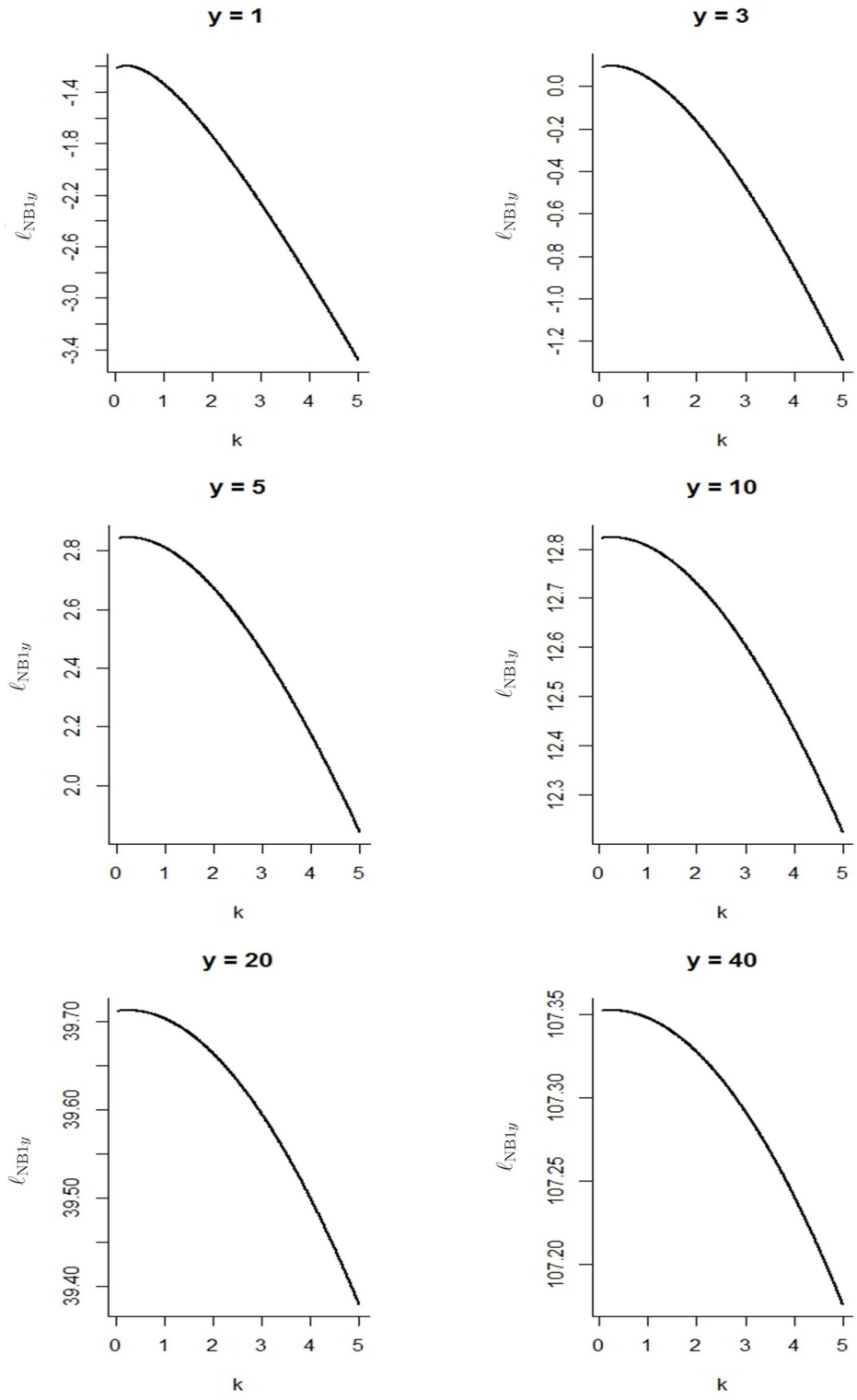


Figure 3.7: Mode of $\ell_{\mathrm{NB1}y}$ for $\alpha = 0.5$.

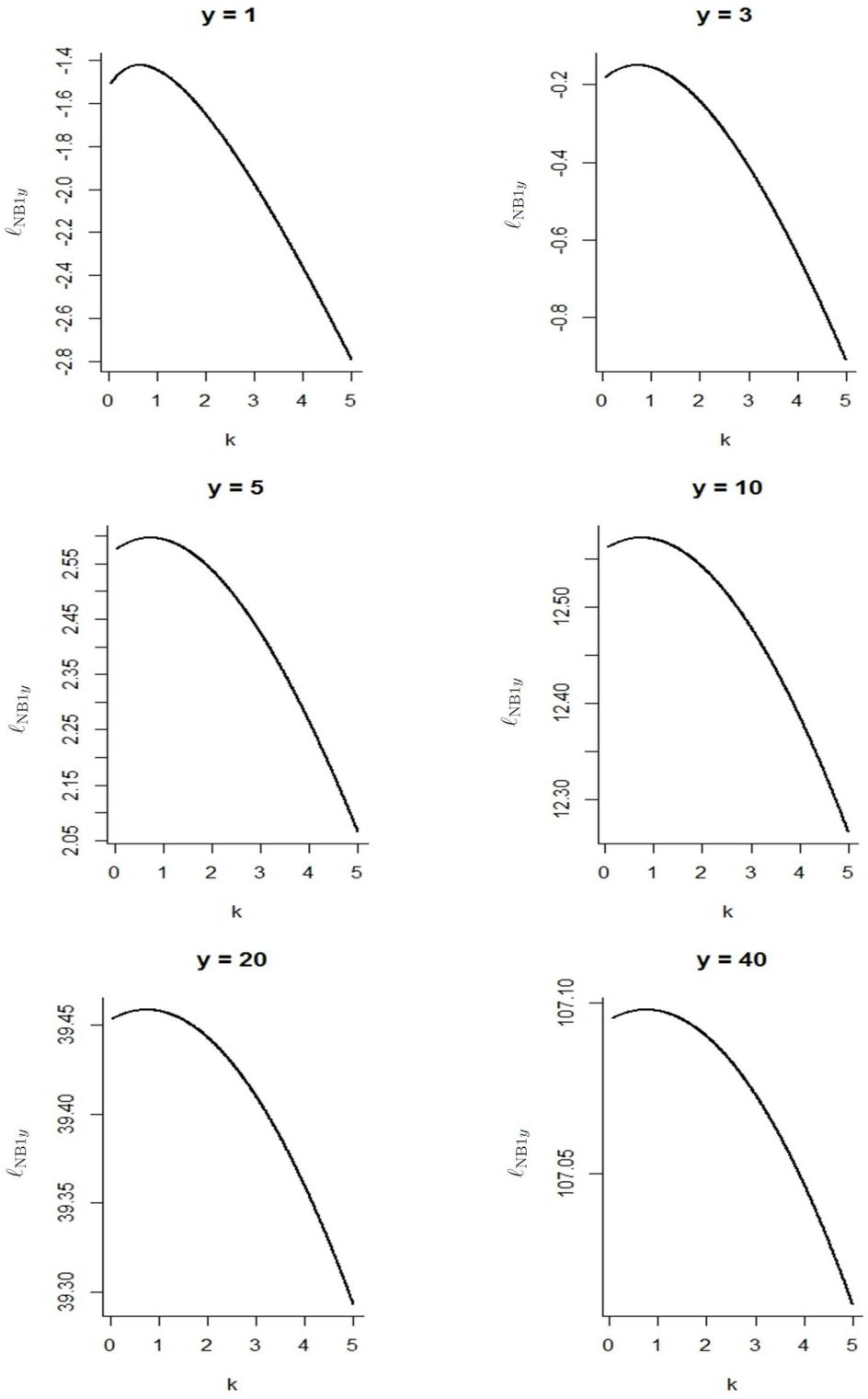


Figure 3.8: Mode of $\ell_{\mathrm{NB1}y}$ for $\alpha = 1.5$.

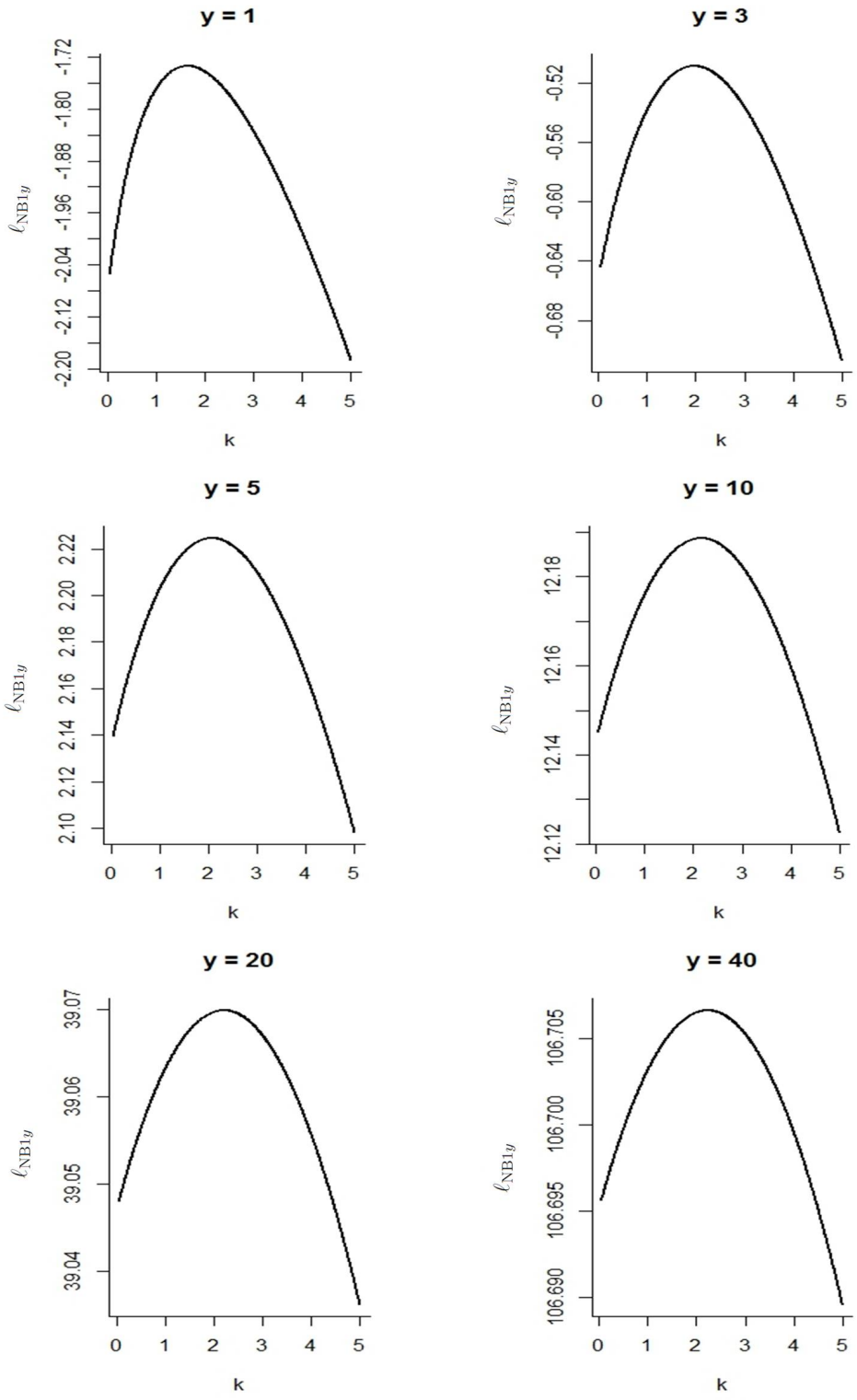


Figure 3.9: Mode of $\ell_{\mathrm{NB1}y}$ for $\alpha = 4.5$.

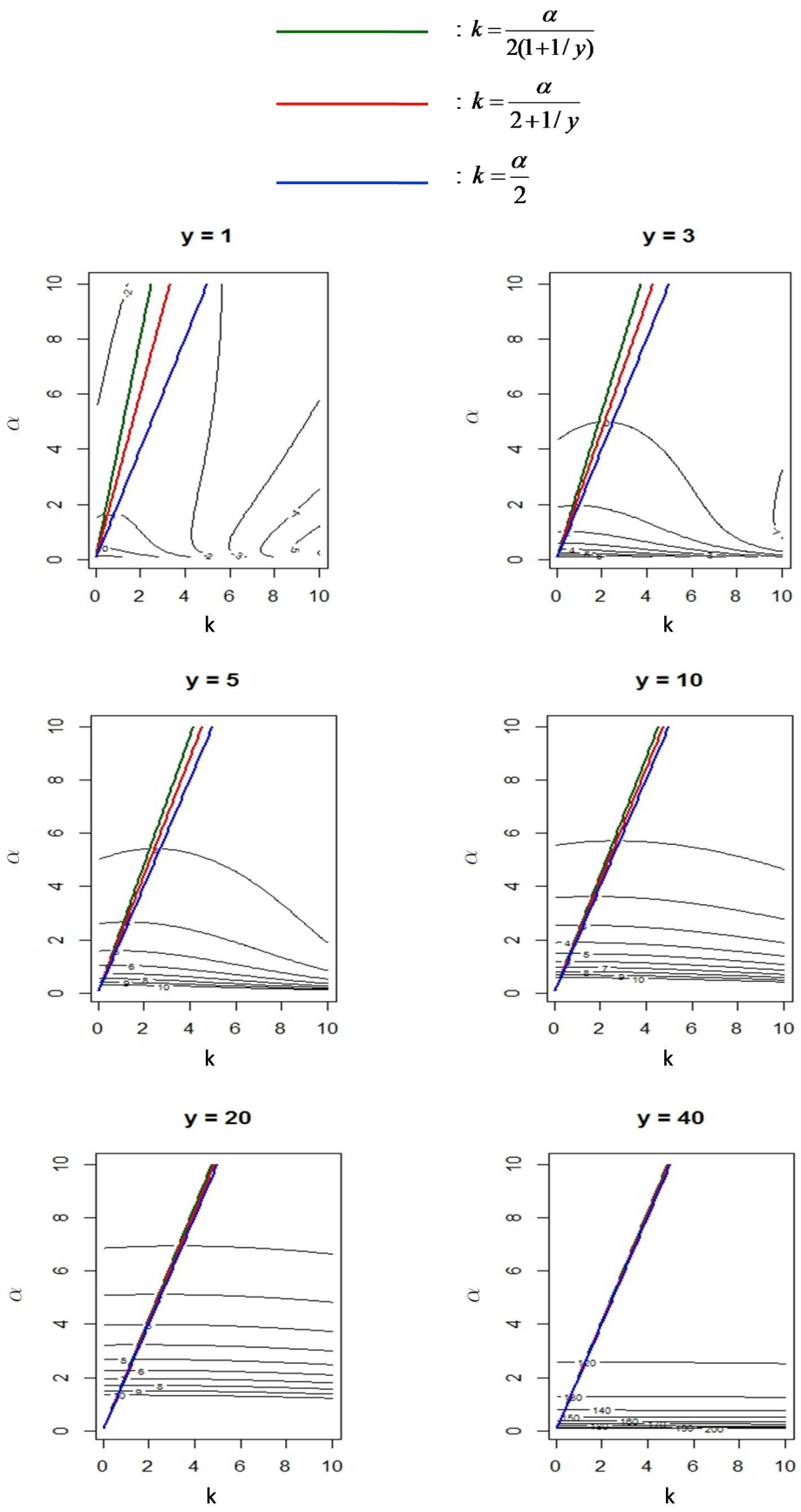


Figure 3.10: Contour plot of $\ell_{\mathrm{NB1}y}$.

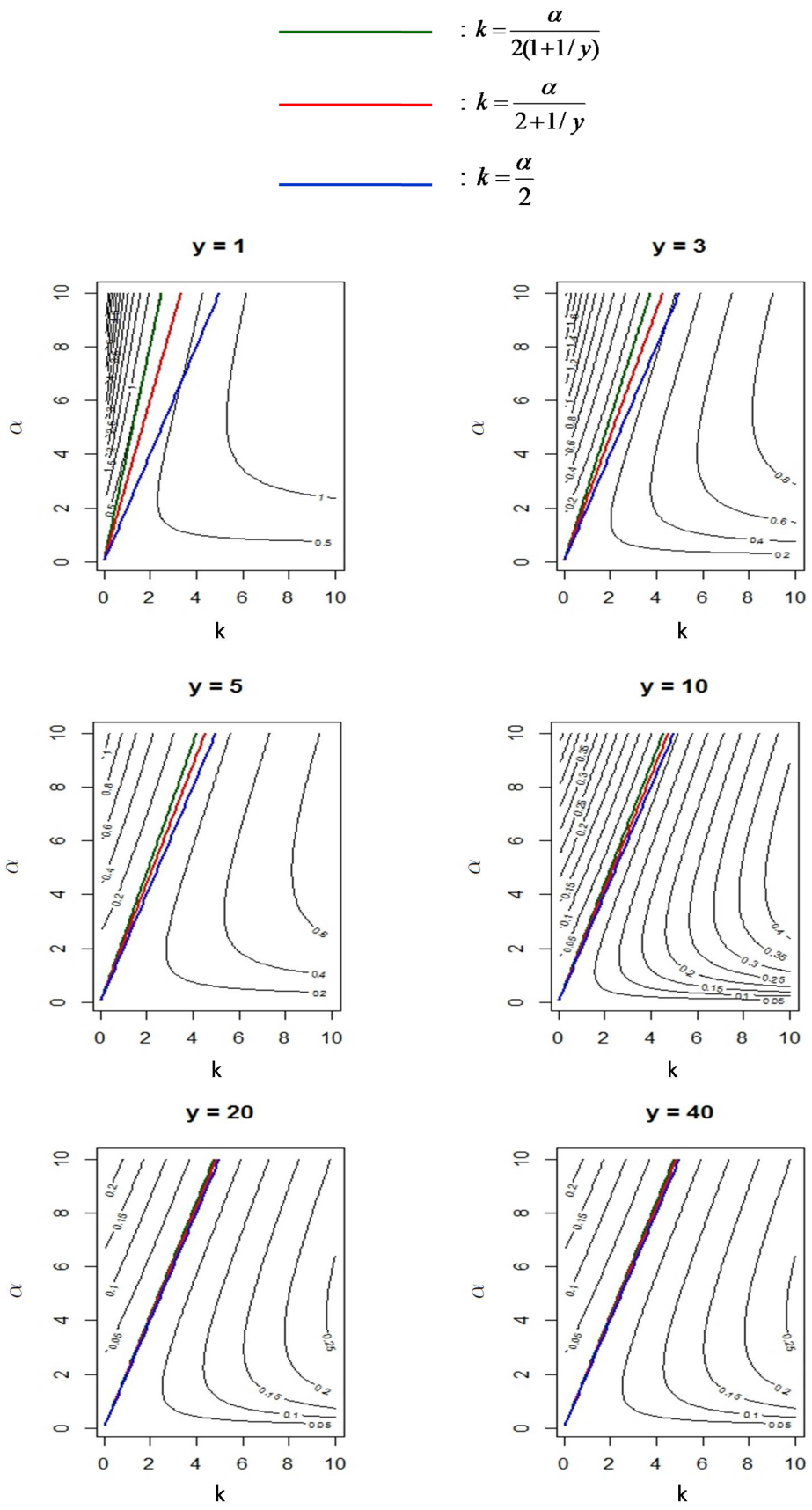


Figure 3.11: Contour plot of $\left|\frac{\partial \ell_{\mathrm{NB1}y}}{\partial \mu}\right|$.

CHAPTER 4

An Approximated NB1 Standardized Deviance Residual

Based on the standardized deviance residuals of NB2 models shown in expression (3.14), we here form a hat matrix; $\boldsymbol{H} = \boldsymbol{W}^{\frac{1}{2}}\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}^{\frac{1}{2}}$, for fitted NB1 models. The key idea is finding the main diagonal elements, w_{ii}, of the weight matrix $\boldsymbol{W}$.

4.1 An Approximated Hat Value of NB1 Models

To obtain an individual hat value of the NB1 model, we use the definition of w_{ii} in equation (2.15) to approximate ones for the NB1 model with mean μ_i, overdispersion α and approximated link function, $g(\mu_i) = \ln(\mu_i)$. That is

$$\begin{aligned} w_{ii} &= \frac{1}{\mathrm{Var}(Y_i)\left[g'(\mu_i)\right]^2} \\ &= \frac{1}{\mu_i(1+\alpha)\left(\frac{1}{\mu_i}\right)^2} \\ &= \frac{\mu_i}{1+\alpha}. \end{aligned} \tag{4.1}$$

Hence, the approximated hat matrix for a fitted NB1 model is $\boldsymbol{H} = \boldsymbol{W}^{\frac{1}{2}}\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}^{\frac{1}{2}}$, where $\boldsymbol{W}$ is an $n \times n$ diagonal matrix with the i^{th} diagonal element $w_{ii} = \dfrac{\mu_i}{1+\alpha}$, evaluated at $\mu_i = \hat{\mu}_i$ and $\alpha = \hat{\alpha}$ and $\boldsymbol{X}$ is an $n \times (p+1)$ design matrix.

If the hat matrix, its approximated hat value (h_{ii}) chooses is correct, the hat value will have the property that $0 \leq h_{ii} \leq 1$ with mean $\frac{p+1}{n}$ and $\sum_{i=1}^{n} h_{ii} = p + 1$, where $\sum_{i=1}^{n} h_{ii}$ is the trace or rank of $\boldsymbol{H}$ for all models with a constant or intercept (Atkinson, 1985 [4]). In order to investigate that the hat value, we conducted a simulation study by using R (R Core Team, 2014 [18]) and following the description in Jansakul and Hinde (2009) [14]. In the experiment, we simulated 1000 sets of NB1 responses based on some simple models with covariates consisting of a two level factor and a continuous variate, for a sample of size $n = 20, 50, 100$ and $\alpha = 1.5, 4.5$. The various models studied are shown in Table 4.1.

Table 4.1:

Working models for investigating the properties of the hat values of the NB1 regression and their statistics given by the experiments.

	Working models		**Statistics of h_{ii} from the experiment (average values)**			
n	$\ln(\boldsymbol{\mu})$	α	**Min**	**Max**	**Mean**	**Rank**
20	$1.75 + 1.05\boldsymbol{x}_1$	1.5	0.083	0.125	0.100	2
		4.5	0.083	0.125	0.100	2
	$-0.45 + 0.75\boldsymbol{x}_2$	1.5	0.052	0.223	0.100	2
		4.5	0.053	0.232	0.100	2

	Working models		Statistics of h_{ii} from the experiment (average values)			
n	$\ln(\boldsymbol{\mu})$	α	**Min**	**Max**	**Mean**	**Rank**
20	$1.25 - 2.45\boldsymbol{x}_1 + 0.85\boldsymbol{x}_2$	1.5	0.064	0.297	0.150	3
		4.5	0.065	0.344	0.150	3
50	$1.75 + 1.05\boldsymbol{x}_1$	1.5	0.033	0.050	0.040	2
		4.5	0.033	0.050	0.040	2
	$-0.45 + 0.75\boldsymbol{x}_2$	1.5	0.021	0.078	0.040	2
		4.5	0.021	0.107	0.040	2
	$1.25 - 2.45\boldsymbol{x}_1 + 0.85\boldsymbol{x}_2$	1.5	0.027	0.106	0.060	3
		4.5	0.027	0.108	0.060	3
100	$1.75 + 1.05\boldsymbol{x}_1$	1.5	0.017	0.025	0.020	2
		4.5	0.017	0.025	0.020	2
	$-0.45 + 0.75\boldsymbol{x}_2$	1.5	0.010	0.062	0.020	2
		4.5	0.010	0.075	0.020	2
	$1.25 - 2.45\boldsymbol{x}_1 + 0.85\boldsymbol{x}_2$	1.5	0.013	0.095	0.030	3
		4.5	0.013	0.100	0.030	3

The explanatory variables are $\boldsymbol{x}_1$, a two level factor with two-fifths of the observations in the first group; $\boldsymbol{x}_2$, a variable with values uniformly distributed on (1,3) rounded to two decimals. With these values $\boldsymbol{\mu}$ varies from $0.70 - 44.70$.

For each set of simulated NB1 responses, we fitted the NB1 models using the same explanatory variables as ones in the working models and then calculated w_{ii} followed by h_{ii} and the rank of the associated $\boldsymbol{H}$. In order to check the properties, we then calculated the average of their minimum, maximum, mean and rank values. Such statistics are presented in Table 4.1. We also draw an index plot of those values (except the rank) for $n = 20, 50, 100$ to see the variation. These plots are shown in Figures 4.1 to 4.3, respectively.

From Table 4.1 and the plots; Figures 4.1 to 4.3, we can see that the properties of our approximated hat values follow the general theoretical properties as mentioned above. Hence they can be used to develop an expression of the NB1 standardized deviance residuals.

4.2 An Approximated Standardized Deviance Residual of NB1 Models

Using the definition of standardized deviance residuals, see equation (2.30) and the approximated NB1 deviance residuals (3.21), an approximated individual NB1 standardized deviance residual, denoted by $\dot{r}_{\text{D\_NB1}i}$ is then

$$\dot{r}_{\text{D\_NB1}i} = \frac{r_{\text{D\_NB1}i}}{\sqrt{1-h_{ii}}}. \tag{4.2}$$

In order to check whether $\dot{\boldsymbol{r}}_{\text{D\_NB1}}$ is normally distributed with mean 0 and variance 1 ($\dot{\boldsymbol{r}}_{\text{D\_NB1}} \sim \text{N}(0,1)$) which corresponds to testing

$$\text{H}_0 : \dot{\boldsymbol{r}}_{\text{D\_NB1}} \text{ has standard normal distribution}$$

against $\quad \text{H}_1 : \dot{\boldsymbol{r}}_{\text{D\_NB1}}$ has other distribution,

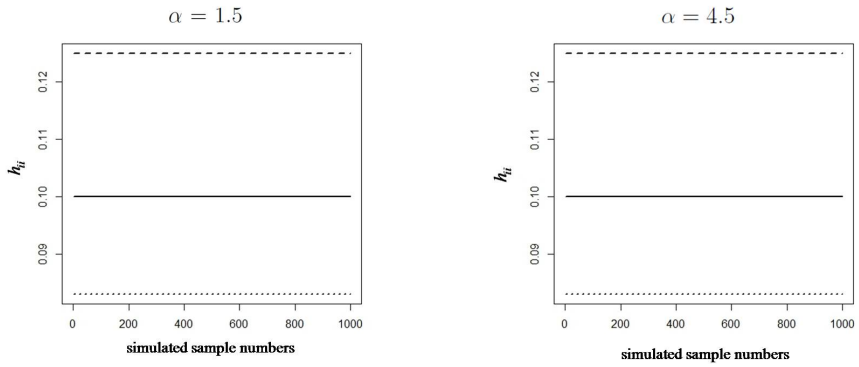
we again calculated 1000 sets (for $n = 20$, 50 and 100) of $\dot{\boldsymbol{r}}_{\mathrm{D\_NB1}}$ by using the working NB1 models in Table 4.1. We then check the agreement with the assumed distribution of $\dot{r}_{\mathrm{D\_NB1}i}$ by using the Kolmogorov-Smirnov test statistics and the Shapiro-Wilk test statistic. From 1000 test statistics, we calculated the percentages of times of rejecting H_0 at the nominal significance level of 0.05. The corresponding values including some descriptive statistics of $\dot{\boldsymbol{r}}_{\mathrm{D\_NB1}}$ are presented in Table 4.2.

..... : minimum values

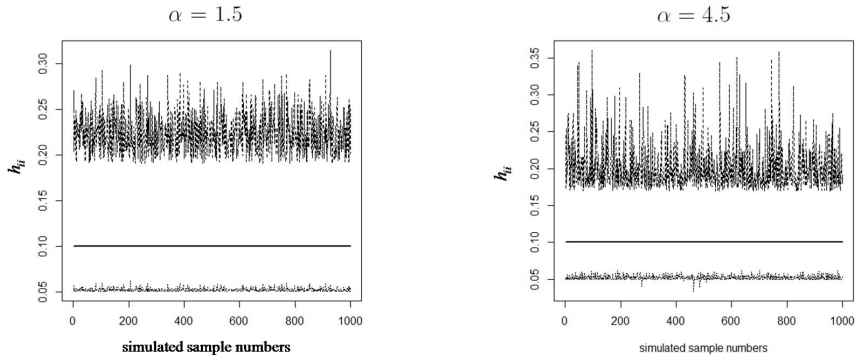
——————— : mean values

- - - - - - - - - - - - : maximum values

(a) $\ln(\boldsymbol{\mu}) = 1.75 + 1.05\boldsymbol{x}_1$



(b) $\ln(\boldsymbol{\mu}) = -0.45 + 0.75\boldsymbol{x}_2$



(c) $\ln(\boldsymbol{\mu}) = 1.25 - 2.45\boldsymbol{x}_1 + 0.85\boldsymbol{x}_2$

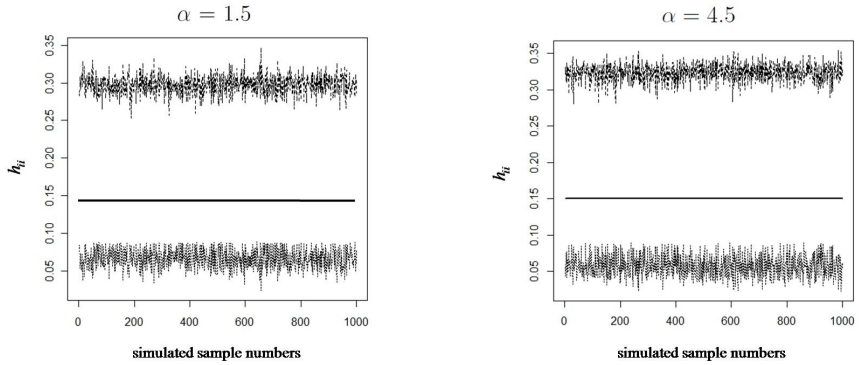


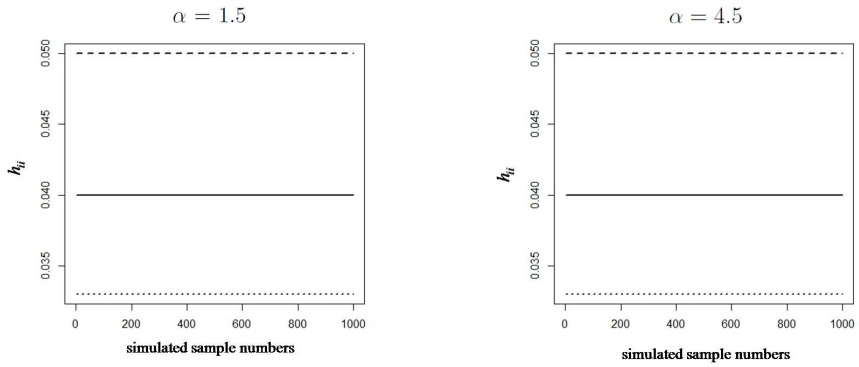
Figure 4.1: Plots of h_{ii} based on 1000 simulations under working NB1 models for $n = 20$.

..... : minimum values

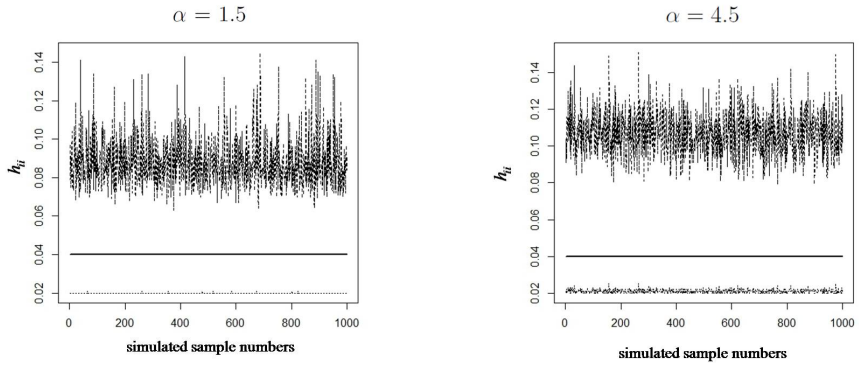
——————— : mean values

- - - - - - - - - - - - : maximum values

(a) $\ln(\boldsymbol{\mu}) = 1.75 + 1.05\boldsymbol{x}_1$



(b) $\ln(\boldsymbol{\mu}) = -0.45 + 0.75\boldsymbol{x}_2$



(c) $\ln(\boldsymbol{\mu}) = 1.25 - 2.45\boldsymbol{x}_1 + 0.85\boldsymbol{x}_2$

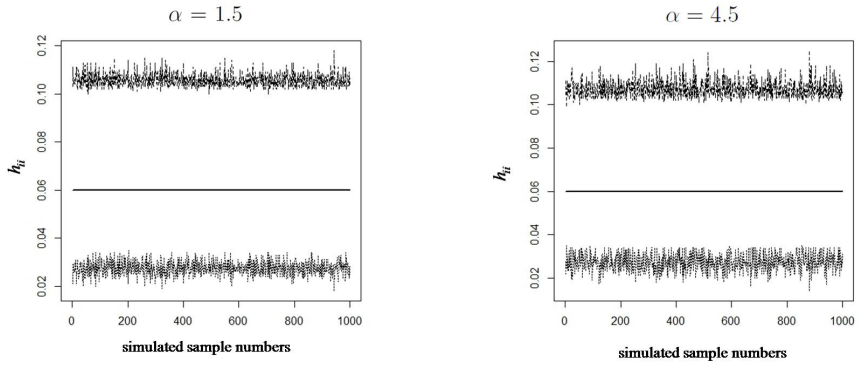


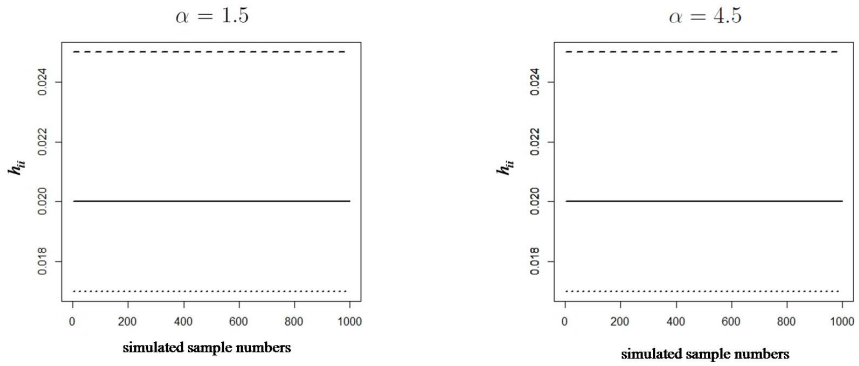
Figure 4.2: Plots of h_{ii} based on 1000 simulations under working NB1 models for $n = 50$.

..... : minimum values

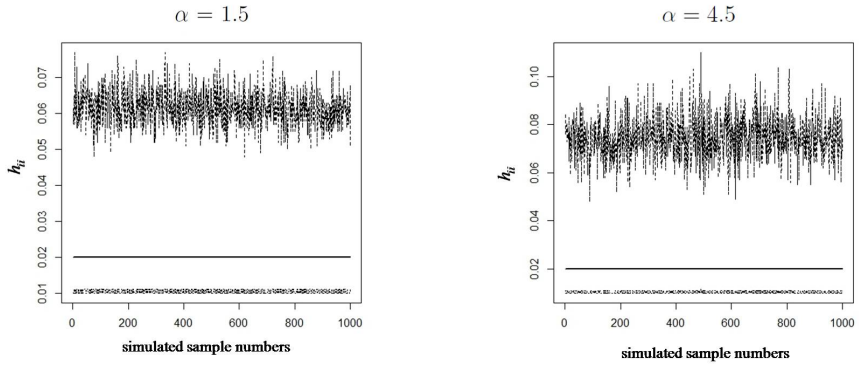
——————— : mean values

- - - - - - - - - - - - : maximum values

(a) $\ln(\boldsymbol{\mu}) = 1.75 + 1.05\boldsymbol{x}_1$



(b) $\ln(\boldsymbol{\mu}) = -0.45 + 0.75\boldsymbol{x}_2$



(c) $\ln(\boldsymbol{\mu}) = 1.25 - 2.45\boldsymbol{x}_1 + 0.85\boldsymbol{x}_2$

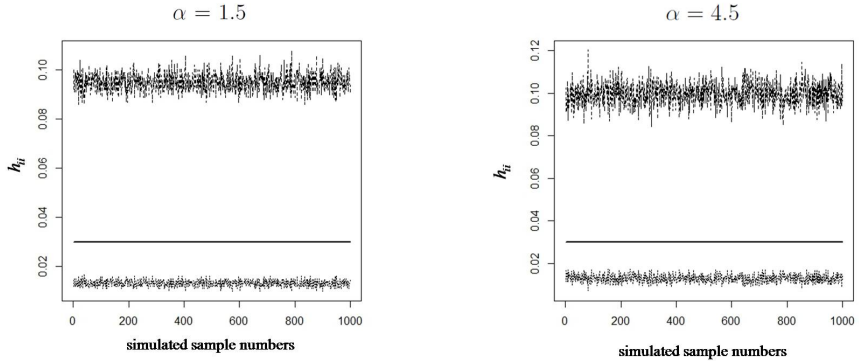


Figure 4.3: Plots of h_{ii} based on 1000 simulations under working NB1 models for $n = 100$.

Table 4.2:

Working models for investigating the distribution of the approximated NB1 standardized deviance residual and their statistics given by the experiments.

			Statistics of $\dot{r}_{\text{D\_NB1}}$ from the experiment (average values)				**Percentages***	
	Working models						**Kolmogorov–**	**Shapiro–**
n	$\ln(\boldsymbol{\mu})$	α	**Min**	**Max**	**Mean**	**Variance**	**Smirnov test**	**Wilk test**
20	$1.75 + 1.05\boldsymbol{x}_1$	1.5	-2.153	2.067	-0.018	1.240	0.8	**6.6**
		4.5	-2.162	2.161	-0.009	1.340	1.7	9.0
	$-0.45 + 0.75\boldsymbol{x}_2$	1.5	-2.149	2.084	-0.095	1.368	2.0	**4.6**
		4.5	-1.929	2.166	**−0.152**	1.450	17.9	27.8
	$1.25 - 2.45\boldsymbol{x}_1 + 0.85\boldsymbol{x}_2$	1.5	-2.108	2.150	-0.063	1.391	1.9	**5.4**
		4.5	-1.987	2.199	-0.080	1.361	4.1	14.5
50	$1.75 + 1.05\boldsymbol{x}_1$	1.5	-2.451	2.348	-0.009	1.112	0.6	9.2
		4.5	-2.271	2.416	-0.003	1.204	4.2	26.0
	$-0.45 + 0.75\boldsymbol{x}_2$	1.5	-2.334	2.387	-0.081	1.248	4.1	8.8
		4.5	-2.051	2.432	**−0.159**	1.321	71.6	69.9
	$1.25 - 2.45\boldsymbol{x}_1 + 0.85\boldsymbol{x}_2$	1.5	-2.261	2.405	-0.061	1.188	1.6	8.6
		4.5	-2.202	2.490	-0.072	1.223	16.3	34.2

			Statistics of $\dot{r}_{\text{D\_NB1}}$ from the experiment (average values)				Percentages*	
	Working models						**Kolmogorov–**	**Shapiro–**
n	$\ln(\boldsymbol{\mu})$	α	**Min**	**Max**	**Mean**	**Variance**	**Smirnov test**	**Wilk test**
100	$1.75 + 1.05\boldsymbol{x}_1$	1.5	-2.647	2.581	-0.007	1.080	**5.6**	17.6
		4.5	-2.348	2.595	0.000	1.144	23.9	72.8
	$-0.45 + 0.75\boldsymbol{x}_2$	1.5	-2.481	2.600	-0.080	1.208	23.0	21.4
		4.5	-2.136	2.637	**−0.148**	1.279	98.8	96.4
	$1.25 - 2.45\boldsymbol{x}_1 + 0.85\boldsymbol{x}_2$	1.5	-2.452	2.642	-0.064	1.141	**8.2**	20.1
		4.5	-2.505	2.698	-0.075	1.214	53.3	61.5

Percentages * : Percentages of rejecting H_0 at the significance level of 0.05.

Our study shows following results :

1. The mean and variance of simulated $\dot{\boldsymbol{r}}_{\text{D\_NB1}}$ is approximately 0 and 1, respectively. However, there are some that deviate from the nominal values larger than 0.1. These are found for the working models with $\alpha = 4.5$ (see the highlight values) as expected.

2. The percentages of rejecting H_0 at the significance level of 0.05 for the Kolmogorov-Smirnov test and the Shapiro-Wilk test statistic are satisfactory for the working models with $\alpha = 1.5$. Based on the advantages of the test; the former test is suitable for a large sample where the other is good for a small sample (Ahad *et al.*, 2011 [1]). Again, see the highlight values shown at the last two columns of Table 4.2.

3. In the case of a large overdispersion parameter, here for example, the working models with $\alpha = 4.5$, the distributional approximation of $\dot{\boldsymbol{r}}_{\text{D\_NB1}}$ is rather unsatisfactory. This may be caused by unadjusted h_{ii}.

4.3 Example

We here illustrate the use of the approximated NB1 standardized deviance residual with the set of orange variety *Valencia* tissue culture data (Tomaz *et al.*,2001 [19] and Jansakul and Hinde, 2004 [13]). In this example, the authors showed that a standard Poisson model is not adequate caused by varying mean and went on to consider various models including NB1 type. We now use this data set to refit the Poisson, NB1 and NB2 model and do model checking for the final NB1 model using our proposed standardized deviance residual.

An orange variety *Valencia* tissue culture data

The set of orange variety *Valencia* tissue culture data (Tomaz *et al.*, 2001 [19] and Jansakul and Hinde, 2004 [13]) shown in Table 4.3 presents the mean and variance of the number of embryos from a *Valencia* orange tissue culture experiment using 3 carbohydrate sources or sugars: maltose, lactose and galactose at dose levels of 18, 37, 75, 110 and 150 μM (3 replicates of each treatment) after experiment approximately four weeks. Most of sample shows overdispersion (variance exceeds the mean). Jansakul and Hinde (2004) [13] used the NB1 and NB2 models to fit this data set and concluded that the NB1 model is

preferable. Here we refitted their suggested final NB1 model and used our proposed standardized deviance residual for model checking.

Table 4.3:

Orange variety *Valencia* tissue culture data: Mean and variance of the number of embryos classified by sugars and dose levels.

		Dose levels (μM)				
Sugar		**18**	**37**	**75**	**110**	**150**
Maltose	Mean	233.000	245.333	369.667	407.000	424.333
	Variance	2368.000	654.333	9952.333	2356.000	506.333
Lactose	Mean	47.333	219.333	239.333	174.333	260.500
	Variance	224.333	1310.333	5854.333	1234.333	2964.500
Galactose	Mean	21.667	14.000	18.333	4.000	75.667
	Variance	185.333	76.000	408.333	13.000	508.333

Denoting $\boldsymbol{\mu}$ for the vector of the mean numbers of embryos and writing $DOSE$ and sugar (S) as factors, we firstly fitted the full interaction Poisson regression model $\ln(\boldsymbol{\mu}) = S^{*}DOSE$. The model gives the residual deviance of 298.04 on 29 degrees of freedom (df), showing strong evidence of overdispersion (the residual deviance is much greater than its df). In order to search for an appropriate dose level for the number of the embryos, we wrote $\boldsymbol{D}$ for a vector of the carbohydrate dose levels and fitted the NB1 and NB2 model with the interaction between sugar and a series of order polynomial over dose levels, where the most complicated model is

$\ln(\hat{\boldsymbol{\mu}}) = S^*(\boldsymbol{D} + \boldsymbol{D}^2 + \boldsymbol{D}^3 + \boldsymbol{D}^4)$. The corresponding essential statistics for studied fitted models including the mean and variance of $\dot{\boldsymbol{r}}_{\mathrm{D}}$ are presented in Table 4.4.

Table 4.4:

Orange variety *Valencia* tissue culture data: Statistics for Poisson, NB1 and NB2 models.

S is a three-level factor for sugar

$\boldsymbol{D}$ is a variate for the dose level: 18, 37, 75, 110 and 150 μM

							$\dot{\boldsymbol{r}}_{\mathrm{D}}$	
Models	$\ln(\boldsymbol{\mu})$	$\boldsymbol{\alpha}$	$-\mathbf{2}\ell$	**df**	**AIC**	**BIC**	**Mean**	**Variance**
Poisson	$S^*(\boldsymbol{D} + \boldsymbol{D}^2 + \boldsymbol{D}^3 + \boldsymbol{D}^4)^\dagger$	0	573.143	29	603.143	629.906	-0.016	9.250
NB1	$S^*(\boldsymbol{D} + \boldsymbol{D}^2 + \boldsymbol{D}^3 + \boldsymbol{D}^4)$	6.331	406.552	28	438.552	467.099	0.040	1.642
	$S^*(\boldsymbol{D} + \boldsymbol{D}^2 + \boldsymbol{D}^3)$	**7.772**	**413.000**	**31**	**439.000**	**462.194**	**0.042**	**1.396**
	$S^*(\boldsymbol{D} + \boldsymbol{D}^2)$	15.360	438.385	34	458.385	476.227	0.059	1.275
	$S^*\boldsymbol{D}$	23.859	457.234	37	471.234	483.724	0.068	1.229
NB2	$S^*(\boldsymbol{D} + \boldsymbol{D}^2 + \boldsymbol{D}^3 + \boldsymbol{D}^4)$	0.060	434.293	28	466.293	494.840	-0.027	1.549
	$S^*(\boldsymbol{D} + \boldsymbol{D}^2 + \boldsymbol{D}^3)$	0.114	451.576	31	477.576	500.771	-0.096	1.488
	$S^*(\boldsymbol{D} + \boldsymbol{D}^2)$	0.244	473.058	34	493.058	510.900	-0.134	1.261
	$S^*\boldsymbol{D}$	0.373	487.576	37	501.576	514.066	-0.144	1.055

$S^*(\boldsymbol{D} + \boldsymbol{D}^2 + \boldsymbol{D}^3 + \boldsymbol{D}^4)^\dagger$ is equivalent to S^*DOSE.

The results in Table 4.4 show that the appropriate model (with smallest value of AIC and BIC of 439.00 and 462.194, respectively) which is consistent with the data is the NB1 model with the cubic function over the dose

levels: $\ln(\hat{\boldsymbol{\mu}}) = S^*(\boldsymbol{D} + \boldsymbol{D}^2 + \boldsymbol{D}^3)$ as presented in Jansakul and Hinde (2004) [13]. Then the fitted model for an orange variety *Valencia* tissue culture data can be defined by

$$\begin{aligned}\ln(\hat{\boldsymbol{\mu}}) \;=\; & 5.864 - 0.303S_1 - 3.834S_2 + 0.314\widetilde{\boldsymbol{D}} - 0.075\widetilde{\boldsymbol{D}}^2 - 0.031\widetilde{\boldsymbol{D}}^3 - 0.920(S_1\widetilde{\boldsymbol{D}}) \\ & (0.072)\quad (0.112)\quad (0.394)\quad (0.129)\quad (0.076)\quad (0.082)\quad (0.235) \\ & -1.127(S_2\widetilde{\boldsymbol{D}}) - 0.563(S_1\widetilde{\boldsymbol{D}}^2) + 0.763(S_2\widetilde{\boldsymbol{D}}^2) + 0.708(S_1\widetilde{\boldsymbol{D}}^3) + 0.560(S_2\widetilde{\boldsymbol{D}}^3) \\ & (0.604)\qquad (0.130)\qquad (0.319)\qquad (0.151)\qquad (0.357)\end{aligned} \tag{4.3}$$

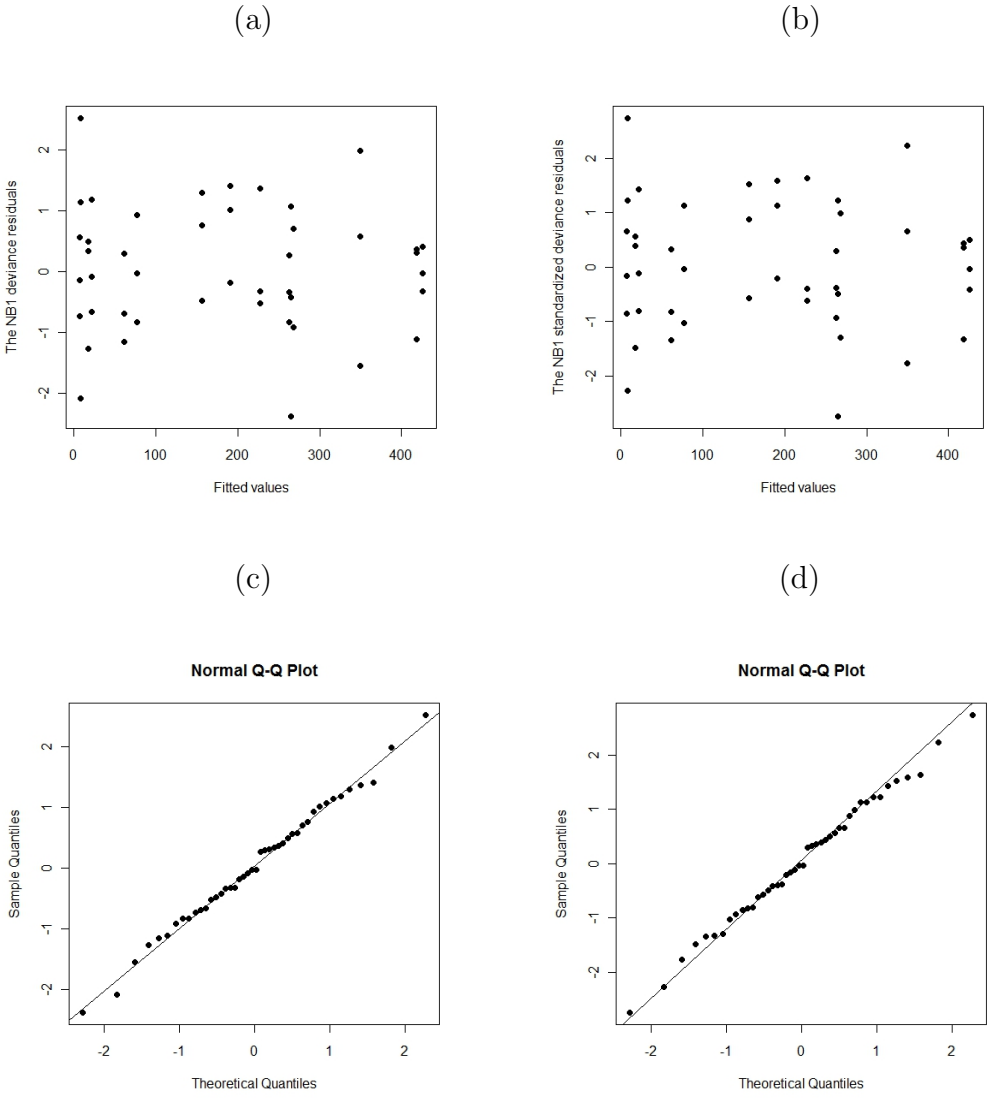
with $\hat{\alpha} = 7.772$.

Here $\widetilde{\boldsymbol{D}}$ is a vector of the standardized dose levels, where $\widetilde{D}_i = \dfrac{D_i - \bar{D}}{\sqrt{\mathrm{Var}(D)}}$; $i = 1, 2, ..., 44$ and $\bar{D} = n^{-1}\sum_{i=1}^{44} D_i$. We transformed $\boldsymbol{D}$ to avoid convergence problems in the maximum likelihood estimation procedure and S denote a three-level factor (maltose, lactose and galactose) for sugars with

$$S_1 = \begin{cases} 1 & ;\text{if } S_1 \text{ is lactose,} \\ 0 & ;\text{otherwise} \end{cases} \qquad \text{and} \qquad S_2 = \begin{cases} 1 & ;\text{if } S_2 \text{ is galactose,} \\ 0 & ;\text{otherwise.} \end{cases}$$

Checking the adequacy of the selected model (4.3) including constant variance and normality assumption of the (standardized) deviance residuals; we used both graphical analysis and test statistics. For the constant variance, the plot of $\boldsymbol{r}_{\text{D\_NB1}}$ and $\dot{\boldsymbol{r}}_{\text{D\_NB1}}$ against the fitted values have no pattern, see Figure 4.4

(a) and (b). Moreover, the Bartlett test statistic of homogeneity of variances of the two quantities on sugars is not significance, see the left column of Table 4.5.



The NB1 deviance residuals The NB1 standardized deviance residuals

Figure 4.4: Model checking for orange variety *Valencia* tissue culture data.

Table 4.5:

Orange variety *Valencia* tissue culture data: Statistics for checking model (4.3).

	Test constant variance (Bartlett test)	Test normal distribution (Shapiro-Wilk test)
$\boldsymbol{r}_{\text{D\_NB1}}$	Bartlett's K-squared = 0.7555, df = 2 and p-value = 0.6854	W = 0.994 and p-value = 0.9981
$\dot{\boldsymbol{r}}_{\text{D\_NB1}}$	Bartlett's K-squared = 0.7343, df = 2 and p-value = 0.6927	W = 0.9945 and p-value = 0.9989

For the normality assumption, the normal q-q plot of $\boldsymbol{r}_{\text{D\_NB1}}$ and $\dot{\boldsymbol{r}}_{\text{D\_NB1}}$ displayed in Figure 4.4 (c) and (d) show a straight line and the Shapiro-Wilk test of normality for both residuals is not significance, see the right column of Table 4.5. Therefore, we can conclude that the two essential assumptions are met. For $\dot{\boldsymbol{r}}_{\text{D\_NB1}}$, the mean and variance presented in the last two columns of Table 4.4 are respectively around 0 and 1, we then can conclude that $\dot{\boldsymbol{r}}_{\text{D\_NB1}} \sim \text{N}(0, 1)$.

CHAPTER 5

Conclusion and Discussion

The main aspects throughout this thesis have been the discussion of studying the behavior of the deviance residual; an important quantity used in statistical model checking. For well known generalize linear models: normal, binomial, Poisson and gamma model, the deviance residual followed by the standardized form has been proposed. This thesis focused on the deviance residual of a linear mean-variance negative binomial model. The NB1 model is a model for overdispersed counts and not a glm. In investigating, we first explored the relationship of the deviance residuals and log-likelihood of the mentioned glms in both theoretical and empirical aspects and found that the mode of glm log-likelihood occurs at the equality of the observed value and its mean, implying that the deviance residual is zero. We then used this idea to form an approximated NB1 deviance residual and its standardized version.

In Chapter 3, we first presented a general form of the negative binomial model and showed that the quadratic mean-variance negative binomial model is a glm when the shape parameter is known, where the association of the deviance residuals and log-likelihood follows the glms' agreement. Even though the NB1 model is not a member of glms, a deviance function can be defined. The function is rather complicated and indicates that the mode of the log-likelihood can be found when the mean is a function of both the observed value and overdispersion

parameter, unless the observed value is zero (Jansakul and Hinde, 2004 [13]). Our investigation under their permission presented the procedure of obtaining their proposed approximated NB1 deviance residual and also found that their approximated deviance residual behaves as well as the glm deviance residual.

Chapter 4 firstly proposed an approximated leverage or hat value (h_{ii}) of the NB1 model. Our conducted simulation studies revealed that the values follow essential theoretical properties. That is $0 \leq h_{ii} \leq 1$, the mean equals the ratio of the number of model parameters and the sample size and $\sum_{i=1}^{n} h_{ii}$ or the rank of the hat matrix equals the number of model parameters. Following the methodology of obtaining a standardized deviance residual for glms with unit variance function, we developed an approximated NB1 standardized using the deviance in Chapter 3 and our proposed hat value. Our investigation using some simulation study showed that the percentages of rejecting null hypothesis (the NB1 standardized deviance residual has standard normal distribution) at the significance level of 0.05 for the Kolmogorov-Smirnov test and the Shapiro-Wilk test statistic are satisfactory for the working models with a small overdispersion parameter. However, in the case of a large overdispersion parameter, the distributional approximation of the NB1 standardized deviance residual is rather unsatisfactory. Moreover, we illustrate the use of the approximated NB1 standardized deviance residual with the set of orange variety *Valencia* tissue culture data. Our study shows that the appropriate model which is consistent with the data is the NB1 log cubic model. From checking the adequacy of the selected model by using the approximated NB1 standardized deviance residuals we found that the plot of

the approximated NB1 standardized deviance residuals against the fitted values has no pattern and the Bartlett test statistic of the NB1 standardized deviance residuals on sugars has p-value greater than 0.05, therefore the approximated NB1 standardized deviance residuals have constant variance. Then, by considering the normal q-q plot of the NB1 standardized deviance residuals it appears to be distributed around the normal quantile-quantile line. Moreover, the Shapiro - Wilk test of normality has p-value much larger than 0.05, and the results from Table 4.4 showed that the mean is 0 and variance is 1, indicating that the approximated NB1 standardized deviance residuals has standard normal distribution.

Finally the approximated NB1 standardized deviance residual developed in this thesis might need some adjustment, in particular, in the case of an NB1 model with a large overdispersion. A suggestion raised here is that the hat values should be adjusted with a half an overdispersion as it was found in obtaining the NB1 deviance residual.

Bibliography

[1] Ahad, N.A., Yin, T.S., Othman, A.R. and Yaacob, C.R. 2011. Sensitivity of Normality Tests to Non-normal Data. *Sains Malaysiana*. 40(6): 637-641.

[2] Aitkin, M., Anderson, D., Francis, B. and Hinde, J. 1990. *Statistical Modelling in GLIM*. Oxford: Clarendon Press.

[3] Anscombe, F.J. 1953. Contribution of discussion paper by H. Hotelling 'New light on the correlation coefficient and its transforms'. *Journal of the Royal Statistical Society*. Series B.15: 229-230.

[4] Atkinson, A. 1985. *Plots,Transformation and Regression. An introduction graphical methods of diagnostic regression analysis*. Oxford: Clarendon Press.

[5] Byers, A.L., Allore, H., Gill, T.M. and Peduzzi, P.N. 2003. Application of negative binomial modeling for discrete outcomes a case study in aging research. *Journal of Clinical Epidemiology*. 56: 559-564.

[6] Cameron, A.C. and Trivedi, P.K. 1998. *Regression Analysis of Count Data*. New York: Cambridge University Press.

[7] Collett, D. 1991. *Modelling Binary Data*. London: Chapman and Hall.

[8] Elhai, J.D., Calhoun, P.S. and Ford, J.D. 2007. Statistical procedures for analyzing mental health services data. *Psychiatry Research*. 160: 129-136.

[9] Gill, J. 2000. *Generalized Linear Models: A Unified Approach*. California: Sage Publications.

[10] Hilbe, J.M. 2007. *Nagative Binomial Regression*. New York: Cambridge University Press.

[11] Hinde, J.P. and Demetrio, C.G.B. 1998. Overdispersion: model and estimation. *Computational Statistics and Data Analysis*. 27: 151-170.

[12] Hoaglin, D.C. 1992. Diagnostics. In *Perspectives on Contemporary Statistics*, Hoaglin, D.C. and Moore, D.S., Editors. 21: 123-129. Mathematical association of America.

[13] Jansakul, N. and Hinde, J.P. 2004. Linear Mean-Variance Negative Binomial Models for Analysis of Orange Tissue-Culture Data. *Songklanakarin J. Sci Technol*, 26(5): 683-696.

[14] Jansakul, N. and Hinde, J.P. 2009. Score tests for extra-zero models in zero-inflated negative binomial models. *Communications in Statistics-Simulation and Computation*, 38: 92-108.

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

[16] McCullagh, P. and Nelder, J.A. 1989. *Generalized Linear Models*, 2nd edition, London: Chapman and Hall.

[17] Nelder, J.A. 1991. Generalized Linear Models with Negative Binomial or Beta-Binomial Error. Department of Mathematics, Imperial College of Science, Technology and Medicine. Unpublished.

[18] R Core Team 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL http://www.r-project.org.

[19] Tomaz, M.L., Mendes, B.M.J., Mourão Filho, F.A., Demétrio, C.G.B., Jansakul, N. and Rodriguez, A.P.M. 2001. Somatic embryogenesis in citrus spp: Carbohydrate stimulation and histodifferentiation. *In Vitro Cellular & Developmental Biology-Plant*, 37: 446-4852.

APPENDIX A

The mode of log-likelihood function

Following we present the mode of four distributional log-likelihood functions described in Section 2.2 by theoretical investigation, that is find the maximum likelihood estimator for μ, denoted by $\hat{\mu}$.

(1) Normal responses :

Let $\ell_{\mathrm{N}y} = \ell_{\mathrm{N}i}$, then the individual log-likelihood function (2.33) is

$$\ell_{\mathrm{N}y} = \frac{y\mu - \frac{1}{2}\mu^2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2).$$

Take the first derivative of $\ell_{\mathrm{N}y}$ with respect to μ and equate that to 0 ;

$$\begin{aligned} \frac{\partial \ell_{\mathrm{N}y}}{\partial \mu} = \frac{y-\mu}{\sigma^2} &= 0 \\ \hat{\mu} &= y. \end{aligned}$$

Hence the mode of the log-likelihood function is $\hat{\mu} = y$.

(2) Binomial responses :

Let $\ell_{\text{Bin}y} = \ell_{\text{Bin}i}$, then the individual log-likelihood function (2.36) is

$$\ell_{\text{Bin}y} = y \ln\left(\frac{\mu}{m-\mu}\right) + m \ln\left(\frac{m-\mu}{m}\right) + \ln\binom{m}{y},$$

where $\mu = mp$.

Take the first derivative of $\ell_{\text{Bin}y}$ with respect to μ and equate that to 0 ;

$$\begin{aligned}
\frac{\partial \ell_{\text{Bin}y}}{\partial \mu} = \frac{y(m-\mu)}{\mu}\left[\frac{(m-\mu)(1)-\mu(-1)}{(m-\mu)^2}\right] + \frac{m^2}{(m-\mu)}\left[\frac{m(-1)-0}{m^2}\right] &= 0 \\
\frac{ym}{\mu(m-\mu)} + \frac{-m}{(m-\mu)} &= 0 \\
\frac{m}{(m-\mu)}\left[\frac{y}{\mu} - 1\right] &= 0 \\
\hat{\mu} &= y.
\end{aligned}$$

Hence the mode of the log-likelihood function is $\hat{\mu} = y$.

(3) Poisson responses :

Let $\ell_{\text{Pois}y} = \ell_{\text{Pois}i}$, then the individual log-likelihood function (2.39) is

$$\ell_{\text{Pois}y} = y \ln \mu - \mu - \ln y!.$$

Take the first derivative of $\ell_{\text{Pois}y}$ with respect to μ and equate that to 0 ;

$$\begin{aligned}
\frac{\partial \ell_{\text{Pois}y}}{\partial \mu} = \frac{y}{\mu} - 1 &= 0 \\
\frac{y}{\mu} &= 1 \\
\hat{\mu} &= y.
\end{aligned}$$

Hence the mode of the log-likelihood function is $\hat{\mu} = y$.

(4) Gamma responses :

Let $\ell_{\text{Gamma}y} = \ell_{\text{Gamma}i}$, then the individual log-likelihood function (2.42) is

$$\ell_{\text{Gamma}y} = \frac{-y\mu^{-1} - \ln\mu + \ln\alpha}{\alpha^{-1}} + (\alpha - 1)\ln y - \ln\Gamma(\alpha),$$

with $\mu = \alpha\beta$.

Take the first derivative of $\ell_{\text{Gamma}y}$ with respect to μ and equate that to 0 ;

$$\begin{aligned}
\frac{\partial \ell_{\text{Gamma}y}}{\partial \mu} = \alpha\left[\frac{0-(-y)}{\mu^2} - \frac{1}{\mu}\right] &= 0 \\
\alpha\left[\frac{(y-\mu)}{\mu^2}\right] &= 0 \\
y - \mu &= 0 \\
\hat{\mu} &= y.
\end{aligned}$$

Hence the mode of the log-likelihood function is $\hat{\mu} = y$.

APPENDIX B

R functions

1. A function for calculating and drawing the graph of individual log-likelihood for normal model.

```
normal.loglink <- function(y,mu,var)   {
   # mu=mean and var=variance
       ll.norm <- ((y*mu)-((mu^2)/2))/(var)
       ll.norm <- ll.norm-(y^2/(2*var))-((log(2*pi*var))*(1/2))
       ll.norm <- ll.norm

# Create a matrix of mean and individual values of normal log-likelihood.
       norm.mat <- cbind(mu,ll.norm)
       colnames(norm.mat) <- c("mean","log-likelihood")
       cat("\n")
       cat("Individual values of the normal log-likelihood function.","\n")
       cat("\n")
       print(norm.mat)
       cat("\n")
```

```
# Find the maximum likelihood estimator for mean.
        x <- which.max(norm.mat[ ,2])        #rank of maximum likelihood
        max.ll <- norm.mat[x, ]
        cat("The mode of the normal log-likelihood function","\n")
        cat("\n")
        print(max.ll)
        cat("\n")

# Scatter plot of mean and normal log-likelihood function.
        plot(mu,ll.norm,type="n",xlab=expression(mu),ylab=expression(l[Ny]))
        lines(mu,ll.norm)
        lines(c(0.1,y),c(min(ll.norm),max(ll.norm)),lty="dotted")

    result <- list(y=y,varince=var)
    result
}
```

2. A function for calculating and drawing the graph of individual log-likelihood for binomial model.

```
binom.loglink <- function(y,mu,m)  {
  # p=probability of success, m=number of trial and mu=mean=m*p
      ll.binom <- y*log(mu/(m-mu))+m*log((m-mu)/m)+log(choose(m,y))
      ll.binom <- ll.binom

# Create a matrix of mean and individual values of binomial log-likelihood.
      binom.mat <- cbind(mu,ll.binom)
      colnames(binom.mat) <- (“mean”,“log-likelihood”)
      cat(“\n”)
      cat(“Individual values of the binomial log-likelihood function.”,“\n”)
      cat(“\n”)
      print(binom.mat)
      cat(“\n”)

# Find the maximum likelihood estimator for mean.
      x <- which.max(binom.mat[ ,2])      # rank of maximum likelihood
      max.ll <- binom.mat[x, ]
      cat(“The mode of the binomial log-likelihood function”,“\n”)
      cat(“\n”)
```

```
        print(max.ll)
        cat("\n")

# Scatter plot of mean and binomial log-likelihood function.
        plot(mu,ll.binom,type="n",xlab=expression(mu),ylab=expression
    (l[Biny]))
        lines(mu,ll.binom)
        lines(c(0.1,y),c(min(ll.binom),max(ll.binom)),lty="dotted")

    result <- list(y=y,m=m)
    result
}
```

3. A function for calculating and drawing the graph of individual log-likelihood for Poisson model.

```
pois.loglink <- function(y,mu)   {
        ll.pois <- y*log(mu)-mu-log(gamma(y+1))       # mu=mean

# Create a matrix of mean and individual values of Poisson log-likelihood.
        pois.mat <- cbind(mu,ll.pois)
        colnames(pois.mat) <- ("mean","log-likelihood")
```

```
        cat("\n")
        cat("Individual values of the Poisson log-likelihood function.","\n")
        cat("\n")
        print(pois.mat)
        cat("\n")

# Find the maximum likelihood estimator for mean.
        x<-which.max(pois.mat[ ,2])        # rank of maximum likelihood
        max.ll<-pois.mat[x, ]
        cat("The mode of the Poisson log-likelihood function","\n")
        cat("\n")
        print(max.ll)
        cat("\n")

# Scatter plot of mean and Poisson log-likelihood function.
        plot(mu,ll.pois,type="n",xlab=expression(mu),ylab=expression
    (l[Poisy]))
        lines(mu,ll.pois)
        lines(c(0.1,y),c(min(ll.pois),max(ll.pois)),lty="dotted")

    result <- list(y=y)
    result
}
```

4. A function for calculating and drawing the graph of individual log-likelihood for gamma model.

```
gamma.loglink <- function(y,beta,al)  {
   # al=shap parameter, beta=scal parameter and mu=mean=al*beta
      mu <- al*beta
      ll.gamma <- (-y*(1/mu))-log(mu)+log(al)
      ll.gamma <- al*ll.gamma+((al-1)*log(y))-log(gamma(al))
      ll.gamma <- ll.gamma
# Create a matrix of mean and individual values of gamma log-likelihood.
      gamma.mat <- cbind(mu,ll.gamma)
      colnames(gamma.mat) <- c("mean","log-likelihood")
      cat("\n")
      cat("Individual values of the gamma log-likelihood function.","\n")
      cat("\n")
      print(gamma.mat)
      cat("\n")
# Find the maximum likelihood estimator for mean.
      x <- which.max(gamma.mat[ ,2])       # rank of maximum likelihood
      max.ll <- gamma.mat[x, ]
```

```
        cat(˝The mode of the gamma log-likelihood function˝,˝\n˝)
        cat(˝\n˝)
        print(max.ll)
        cat(˝\n˝)

# Scatter plot of mean and gamma log-likelihood function.
        plot(mu,ll.gamma,type=˝n˝,xlab=expression(mu),ylab=expression
    (l[gammay]))
        lines(mu,ll.gamma)
        lines(c(min(mu),y),c(min(ll.gamma),max(ll.gamma)),lty=˝dotted˝)

    result<-list(y=y,al=al,beta=beta)
    result
}
```

5. A function for calculating and drawing the graph of individual log-likelihood for NB2 model.

```
nb2.loglink <- function(y,mu,al)   {
    # al=dispersion parameter and mu=mean
        ll.nb2 <- y*log(al*mu/(1+al*mu))-(log(1+al*mu))/al
```

```
        ll.nb2 <- ll.nb2+lgamma(y+1/al)-lgamma(1/al)

        ll.nb2 <- ll.nb2        # this excludes y! term

# Create a matrix of mean and individual values of NB2 log-likelihood.

        nb2.mat <- cbind(mu,ll.nb2)

        colnames(nb2.mat) <- c("mean","log-likelihood")

        cat("\n")

        cat("Individual values of the NB2 log-likelihood function.","\n")

        cat("\n")

        print(nb2.mat)

        cat("\n")

# Find the maximum likelihood estimator for mean.

        x <- which.max(nb2.mat[ ,2])        #rank of maximum likelihood

        max.ll <- nb2.mat[x, ]

        cat("The mode of the NB2 log-likelihood function","\n")

        cat("\n")

        print(max.ll)

        cat("\n")

# Scatter plot of mean and NB2 log-likelihood function.

        plot(mu,ll.nb2,type="n",xlab=expression(mu),ylab=expression
    (l[NB2y]))

        lines(mu,ll.nb2)
```

```
        lines(c(0.1,y),c(min(ll.nb2),max(ll.nb2)),lty="dotted")
    result <- list(y=y,al=al)
    result
}
```

6. A function for calculating and drawing the graph of individual log-likelihood for NB1 model.

6.1 A function for calculating and drawing the graph of individual log-likelihood for NB1 model where $y = 1$.

```
nb1.loglink <- function(al)   {
    # al=dispersion parameter and mu=mean
        y <- 1
        mu <- seq(0.1,2*y,0.01)
        ll.nb1 <- y*log(al/(1+al))-mu*log(1+al)/al
        ll.nb1 <- ll.nb1+lgamma(y+mu/al)-lgamma(mu/al)
        ll.nb1 <- ll.nb1        # this excludes y! term
# Create a matrix of mean and individual values of NB1 log-likelihood.
        nb1.mat <- cbind(mu,ll.nb1)
        colnames(nb1.mat) <- c("mean","log-likelihood")
```

```
        cat("\n")
        cat("Individual values of the NB1 log-likelihood function.","\n")
        cat("\n")
        print(nb1.mat)
        cat("\n")

# Find the maximum likelihood estimator for mean.
        x <- which.max(nb1.mat[ ,2])        #rank of maximum likelihood
        max.ll <- nb1.mat[x, ]
        cat("The mode of the NB1 log-likelihood function","\n")
        cat("\n")
        print(max.ll)
        cat("\n")

# Scatter plot of mean and NB1 log-likelihood function.
        plot(mu,ll.nb1,type="n",xlab=expression(mu),ylab=expression(l[NB1y]),
    main="y = 1")
        lines(mu,ll.nb1)
        lines(c(0.1,y),c(ll.nb1[1],ll.nb1[91]))
        lines(c(0.1,y+(1/2)),c(ll.nb1[1],ll.nb1[141]),lty="dotted")

    result <- list(y=y,al=al)
    result
}
```

6.2 A function for calculating and drawing the graph of individual log-likelihood for NB1 model where $y = 3$.

```
nb1.loglink <- function(al)   {
   # al=dispersion parameter and mu=mean
       y <- 3
       mu <- seq(0.1,2*y,0.01)
       ll.nb1 <- y*log(al/(1+al))-mu*log(1+al)/al
       ll.nb1 <- ll.nb1+lgamma(y+mu/al)-lgamma(mu/al)
       ll.nb1 <- ll.nb1        # this excludes y! term
# Create a matrix of mean and individual values of NB1 log-likelihood.
       nb1.mat <- cbind(mu,ll.nb1)
       colnames(nb1.mat) <- c("mean","log-likelihood")
       cat("\n")
       cat("Individual values of the NB1 log-likelihood function.","\n")
       cat("\n")
       print(nb1.mat)
       cat("\n")
# Find the maximum likelihood estimator for mean.
       x <- which.max(nb1.mat[ ,2])        #rank of maximum likelihood
```

```
        max.ll <- nb1.mat[x, ]
        cat("The mode of the NB1 log-likelihood function","\n")
        cat("\n")
        print(max.ll)
        cat("\n")

# Scatter plot of mean and NB1 log-likelihood function.
        plot(mu,ll.nb1,type="n",xlab=expression(mu),ylab=expression(l[NB1y]),
    main="y = 3")
        lines(mu,ll.nb1)
        lines(c(0.1,y),c(ll.nb1[1],ll.nb1[291]))
        lines(c(0.1,y+(1/2)),c(ll.nb1[1],ll.nb1[341]),lty="dotted")

    result <- list(y=y,al=al)
    result
}
```

6.3 A function for calculating and drawing the graph of individual log-likelihood for NB1 model where $y = 5$.

```
nb1.loglink <- function(al)   {
   # al=dispersion parameter and mu=mean
       y <- 5
       mu <- seq(0.1,2*y,0.01)
       ll.nb1 <- y*log(al/(1+al))-mu*log(1+al)/al
       ll.nb1 <- ll.nb1+lgamma(y+mu/al)-lgamma(mu/al)
       ll.nb1 <- ll.nb1        # this excludes y! term
# Create a matrix of mean and individual values of NB1 log-likelihood.
       nb1.mat <- cbind(mu,ll.nb1)
       colnames(nb1.mat) <- c("mean","log-likelihood")
       cat("\n")
       cat("Individual values of the NB1 log-likelihood function.","\n")
       cat("\n")
       print(nb1.mat)
       cat("\n")
# Find the maximum likelihood estimator for mean.
       x <- which.max(nb1.mat[ ,2])        #rank of maximum likelihood
```

```
        max.ll <- nb1.mat[x, ]
        cat("The mode of the NB1 log-likelihood function","\n")
        cat("\n")
        print(max.ll)
        cat("\n")

# Scatter plot of mean and NB1 log-likelihood function.
        plot(mu,ll.nb1,type="n",xlab=expression(mu),ylab=expression(l[NB1y]),
    main="y = 5")
        lines(mu,ll.nb1)
        lines(c(0.1,y),c(ll.nb1[1],ll.nb1[491]))
        lines(c(0.1,y+(1/2)),c(ll.nb1[1],ll.nb1[541]),lty="dotted")

    result <- list(y=y,al=al)
    result
}
```

6.4 A function for calculating and drawing the graph of individual log-likelihood for NB1 model where $y = 10$.

```
nb1.loglink <- function(al)   {
   # al=dispersion parameter and mu=mean
       y <- 10
       mu <- seq(0.1,2*y,0.01)
       ll.nb1 <- y*log(al/(1+al))-mu*log(1+al)/al
       ll.nb1 <- ll.nb1+lgamma(y+mu/al)-lgamma(mu/al)
       ll.nb1 <- ll.nb1        # this excludes y! term
# Create a matrix of mean and individual values of NB1 log-likelihood.
       nb1.mat <- cbind(mu,ll.nb1)
       colnames(nb1.mat) <- c("mean","log-likelihood")
       cat("\n")
       cat("Individual values of the NB1 log-likelihood function.","\n")
       cat("\n")
       print(nb1.mat)
       cat("\n")
# Find the maximum likelihood estimator for mean.
       x <- which.max(nb1.mat[ ,2])        #rank of maximum likelihood
```

```
        max.ll <- nb1.mat[x, ]
        cat("The mode of the NB1 log-likelihood function","\n")
        cat("\n")
        print(max.ll)
        cat("\n")

# Scatter plot of mean and NB1 log-likelihood function.
        plot(mu,ll.nb1,type="n",xlab=expression(mu),ylab=expression(l[NB1y]),
    main="y = 10")
        lines(mu,ll.nb1)
        lines(c(0.1,y),c(ll.nb1[1],ll.nb1[991]))
        lines(c(0.1,y+(1/2)),c(ll.nb1[1],ll.nb1[1041]),lty="dotted")

    result <- list(y=y,al=al)
    result
}
```

6.5 A function for calculating and drawing the graph of individual log-likelihood for NB1 model where $y = 20$.

```
nb1.loglink <- function(al)   {
   # al=dispersion parameter and mu=mean
       y <- 20
       mu <- seq(0.1,2*y,0.01)
       ll.nb1 <- y*log(al/(1+al))-mu*log(1+al)/al
       ll.nb1 <- ll.nb1+lgamma(y+mu/al)-lgamma(mu/al)
       ll.nb1 <- ll.nb1        # this excludes y! term
# Create a matrix of mean and individual values of NB1 log-likelihood.
       nb1.mat <- cbind(mu,ll.nb1)
       colnames(nb1.mat) <- c("mean","log-likelihood")
       cat("\n")
       cat("Individual values of the NB1 log-likelihood function.","\n")
       cat("\n")
       print(nb1.mat)
       cat("\n")
# Find the maximum likelihood estimator for mean.
       x <- which.max(nb1.mat[ ,2])        #rank of maximum likelihood
```

```
        max.ll <- nb1.mat[x, ]
        cat("The mode of the NB1 log-likelihood function","\n")
        cat("\n")
        print(max.ll)
        cat("\n")

# Scatter plot of mean and NB1 log-likelihood function.
        plot(mu,ll.nb1,type="n",xlab=expression(mu),ylab=expression(l[NB1y]),
    main="y = 20")
        lines(mu,ll.nb1)
        lines(c(0.1,y),c(ll.nb1[1],ll.nb1[1991]))
        lines(c(0.1,y+(1/2)),c(ll.nb1[1],ll.nb1[2041]),lty="dotted")

    result <- list(y=y,al=al)
    result
}
```

6.6 A function for calculating and drawing the graph of individual log-likelihood for NB1 model where $y = 40$.

```
nb1.loglink <- function(al)   {
   # al=dispersion parameter and mu=mean
       y <- 40
       mu <- seq(0.1,2*y,0.01)
       ll.nb1 <- y*log(al/(1+al))-mu*log(1+al)/al
       ll.nb1 <- ll.nb1+lgamma(y+mu/al)-lgamma(mu/al)
       ll.nb1 <- ll.nb1        # this excludes y! term
# Create a matrix of mean and individual values of NB1 log-likelihood.
       nb1.mat <- cbind(mu,ll.nb1)
       colnames(nb1.mat) <- c("mean","log-likelihood")
       cat("\n")
       cat("Individual values of the NB1 log-likelihood function.","\n")
       cat("\n")
       print(nb1.mat)
       cat("\n")
# Find the maximum likelihood estimator for mean.
       x <- which.max(nb1.mat[ ,2])        #rank of maximum likelihood
```

```
        max.ll <- nb1.mat[x, ]
        cat("The mode of the NB1 log-likelihood function","\n")
        cat("\n")
        print(max.ll)
        cat("\n")

# Scatter plot of mean and NB1 log-likelihood function.
        plot(mu,ll.nb1,type="n",xlab=expression(mu),ylab=expression(l[NB1y]),
    main="y = 40")
        lines(mu,ll.nb1)
        lines(c(0.1,y),c(ll.nb1[1],ll.nb1[3991]))
        lines(c(0.1,y+(1/2)),c(ll.nb1[1],ll.nb1[4041]),lty="dotted")

    result <- list(y=y,al=al)
    result
}
```

7. A function for calculating and drawing the graph for search an appropriate value of k.

7.1 A function for for search an appropriate value of k where $y = 1, 3, 5$ and 10.

```
nb1.llink <- function(y,al)   {
    # al=dispersion parameter
       k <- seq(0.05,5,0.01)
       llk.nb1 <- lgamma(y+(y+k)/al)-lgamma((y+k)/al)
       llk.nb1 <- y*log(al/(1+al))-(y+k)*log(1+al)/al
       llk.nb1 <- llk.nb1        # this excludes y! term
# Create a matrix of k and individual values of NB1 log-likelihood.
       nb1.mat <- cbind(k,llk.nb1)
       colnames(nb1.mat) <- c("k","log-likelihood")
       cat("\n")
       cat("Individual values of the NB1 log-likelihood function for each k.",
    "\n")
       cat("\n")
       print(nb1.mat)
       cat("\n")
# Search for k to giving the mode of NB1 log-likelihood.
       x <- which.max(nb1.mat[ ,2])       # rank of maximum log-likelihood
```

```
        max.ll <- nb1.mat[x, ]
        cat("The mode of the NB1 log-likelihood function","\n")
        cat("\n")
        print(max.ll)
        cat("\n")

# Scatter plot of k and the NB1 log-likelihood function.
       plot(k,llk.nb1,type="n",xlab="k",ylab=expression(l[NB1y]),axes=F)
        mi.y <- min(as.integer(llk.nb1))-2
        ma.y <- max(as.integer(llk.nb1))+2
        mi.x <- min(as.integer(k))-1
        ma.x <- max(as.integer(k))+1
        axis(1,at=seq(mi.x,ma.x,by=1))
        axis(2,at=seq(mi.y,ma.y,by=0.2))
        lines(k,llk.nb1,lty=1)

    result <- list(y=y,al=al)
    result
}
```

7.2 A function for for search an appropriate value of k where $y = 20$ and 40.

```
nb1.llink <- function(y,al)  {
    # al=dispersion parameter
       k <- seq(0.05,5,0.01)
       llk.nb1 <- lgamma(y+(y+k)/al)-lgamma((y+k)/al)
       llk.nb1 <- y*log(al/(1+al))-(y+k)*log(1+al)/al
       llk.nb1 <- llk.nb1        # this excludes y! term
# Create a matrix of k and individual values of NB1 log-likelihood.
       nb1.mat <- cbind(k,llk.nb1)
       colnames(nb1.mat) <- c("k","log-likelihood")
       cat("\n")
       cat("Individual values of the NB1 log-likelihood function for each k.",
    "\n")
       cat("\n")
       print(nb1.mat)
       cat("\n")
# Search for k to giving the mode of NB1 log-likelihood.
       x <- which.max(nb1.mat[ ,2])       # rank of maximum log-likelihood
```

```
        max.ll <- nb1.mat[x, ]
        cat("The mode of the NB1 log-likelihood function","\n")
        cat("\n")
        print(max.ll)
        cat("\n")

# Scatter plot of k and the NB1 log-likelihood function.
       plot(k,llk.nb1,type="n",xlab="k",ylab=expression(l[NB1y]),axes=F)
        mi.y <- min(as.integer(llk.nb1))-2
        ma.y <- max(as.integer(llk.nb1))+2
        mi.x <- min(as.integer(k))-1
        ma.x <- max(as.integer(k))+1
        axis(1,at=seq(mi.x,ma.x,by=1))
        axis(2,at=seq(mi.y,ma.y,by=0.05))
        lines(k,llk.nb1,lty=1)

    result <- list(y=y,al=al)
    result
}
```

8. A function for drawing the graph contour plot of $\left|\frac{\partial \ell_{\mathrm{NB1}y}}{\partial \mu}\right|$.

8.1 A function for drawing the graph contour plot of $\left|\frac{\partial \ell_{\mathrm{NB1}y}}{\partial \mu}\right|$ where $y = 1$.

```
nb1.llink <- function(y)  {
    # al=dispersion parameter
       al <- seq(0.1,10,length=100)
       k <- al       # a constant for ℓ(y + k, α; y)
       lf <- outer(k,al,function(k,al)
    -log(1+al)+(digamma(y+(y+k)/al)-digamma((y+k)/al)))
       lf <- abs(lf)

# Contour plot.
       contour(k,al,lf,zlim=c(6,4,2,1,0.5,0.1,01,0),main="y = 1",
    ylab=expression(alpha),xlab="k")
       lines(al/(2*(1+1/y)),al,lty="dotted")
       lines(al/(2+1/y),al,lty="dotdash")
       lines(al/2,al)

    result <- list(y=y,lf=lf)
    result
}
```

8.2 A function for drawing the graph contour plot of $\left|\frac{\partial \ell_{\mathrm{NB1}y}}{\partial \mu}\right|$ where $y = 3$.

```
nb1.llink <- function(y)   {
    # al=dispersion parameter
       al <- seq(0.1,10,length=100)
       k <- al        # a constant for ℓ(y + k, α; y)
       lf <- outer(k,al,function(k,al)
    -log(1+al)+(digamma(y+(y+k)/al)-digamma((y+k)/al)))
       lf <- abs(lf)

# Contour plot.
       contour(k,al,lf,zlim=c(2,1,0.5,0.1,0.05,0),main="y = 3",
    ylab=expression(alpha),xlab="k")
       lines(al/(2*(1+1/y)),al,lty="dotted")
       lines(al/(2+1/y),al,lty="dotdash")
       lines(al/2,al)

    result <- list(y=y,lf=lf)
    result
}
```

8.3 A function for drawing the graph contour plot of $\left|\frac{\partial \ell_{\mathrm{NB1}y}}{\partial \mu}\right|$ where $y = 5$.

```
nb1.llink <- function(y)   {
    # al=dispersion parameter
       al <- seq(0.1,10,length=100)
       k <- al        # a constant for ℓ(y + k, α; y)
       lf <- outer(k,al,function(k,al)
    -log(1+al)+(digamma(y+(y+k)/al)-digamma((y+k)/al)))
       lf <- abs(lf)

# Contour plot.
       contour(k,al,lf,zlim=c(2,1,0.5,0.1,0.05,0),main="y = 5",
    ylab=expression(alpha),xlab="k")
       lines(al/(2*(1+1/y)),al,lty="dotted")
       lines(al/(2+1/y),al,lty="dotdash")
       lines(al/2,al)

    result <- list(y=y,lf=lf)
    result
}
```

8.4 A function for drawing the graph contour plot of $\left|\frac{\partial \ell_{\mathrm{NB1}y}}{\partial \mu}\right|$ where $y = 10$.

```
nb1.llink <- function(y)   {
    # al=dispersion parameter
       al <- seq(0.1,10,length=100)
       k <- al        # a constant for ℓ(y + k, α; y)
       lf <- outer(k,al,function(k,al)
    -log(1+al)+(digamma(y+(y+k)/al)-digamma((y+k)/al)))
       lf <- abs(lf)

# Contour plot.
       contour(k,al,lf,zlim=c(0.5,0.4,0.3,0.2,0.1,0.05,0.03,0.01,0),
    main="y = 10",ylab=expression(alpha),xlab="k")
       lines(al/(2*(1+1/y)),al,lty="dotted")
       lines(al/(2+1/y),al,lty="dotdash")
       lines(al/2,al)

    result <- list(y=y,lf=lf)
    result
}
```

8.5 A function for drawing the graph contour plot of $\left|\frac{\partial \ell_{\mathrm{NB1}y}}{\partial \mu}\right|$ where $y = 20$.

```
nb1.llink <- function(y)   {
    # al=dispersion parameter
       al <- seq(0.1,10,length=100)
       k <- al        # a constant for ℓ(y + k, α; y)
       lf <- outer(k,al,function(k,al)
    -log(1+al)+(digamma(y+(y+k)/al)-digamma((y+k)/al)))
       lf <- abs(lf)

# Contour plot.
       contour(k,al,lf,zlim=c(0.5,0.4,0.3,0.2,0.1,0.05,0.03,0.01,0),
    main="y = 20",ylab=expression(alpha),xlab="k")
       lines(al/(2*(1+1/y)),al,lty="dotted")
       lines(al/(2+1/y),al,lty="dotdash")
       lines(al/2,al)

    result <- list(y=y,lf=lf)
    result
}
```

8.6 A function for drawing the graph contour plot of $\left|\frac{\partial \ell_{\mathrm{NB1}y}}{\partial \mu}\right|$ where $y = 40$.

```
nb1.llink <- function(y)   {
    # al=dispersion parameter
       al <- seq(0.1,10,length=100)
       k <- al        # a constant for ℓ(y + k, α; y)
       lf <- outer(k,al,function(k,al)
    -log(1+al)+(digamma(y+(y+k)/al)-digamma((y+k)/al)))
       lf <- abs(lf)

# Contour plot.
       contour(k,al,lf,zlim=c(0.5,0.4,0.3,0.2,0.1,0.09,0.08,0.07,0.06,0.05,0.03,
    0.01,0.009,0.008,0.007,0.0060,0.004,0),main="y = 40",
    ylab=expression(alpha),xlab="k")
       lines(al/(2*(1+1/y)),al,lty="dotted")
       lines(al/(2+1/y),al,lty="dotdash")
       lines(al/2,al)

    result <- list(y=y,lf=lf)
    result
}
```

9. A function for drawing the graph contour plot of individual log-likelihood for NB1 model.

9.1 A function for drawing the graph contour plot of individual log-likelihood for NB1 model where $y = 1$.

```
nb1.llink <- function(y)   {
    # al=dispersion parameter
       al <- seq(0.1,10,length=100)
       k <- al        # a constant for ℓ(y + k, α; y)
       llk.nb1 <- outer(k,al,function(k,al)
    (lgamma(y+(y+k)/al)-lgamma((y+k)/al)-(y+k)*log(1+al)/al))
    # this excludes y! term

# Contour plot.
       contour(k,al,llk.nb1,zlim=c(2,1,0.9,0.8,0.7,0.5,0.4,0.3,0.1,0,-1,-2,-3,-4,
    -6,-6.05,-6.02,-8,-9),main="y = 1",ylab=expression(alpha),xlab="k")
       lines(al/(2*(1+1/y)),al,lty="dotted")
       lines(al/(2+1/y),al,lty="dotdash")
       lines(al/2,al)

    result <- list(y=y,llk.nb1=llk.nb1)
    result
}
```

9.2 A function for drawing the graph contour plot of individual log-likelihood for NB1 model where $y = 3$.

```
nb1.llink <- function(y)   {
    # al=dispersion parameter
       al <- seq(0.1,10,length=100)
       k <- al        # a constant for ℓ(y + k, α; y)
       llk.nb1 <- outer(k,al,function(k,al)
    (lgamma(y+(y+k)/al)-lgamma((y+k)/al)-(y+k)*log(1+al)/al))
    # this excludes y! term

# Contour plot.
       contour(k,al,llk.nb1,zlim=c(7,5,4,3,0.5,0.1,0,-1,-2),main="y = 3",
    ylab=expression(alpha),xlab="k")
       lines(al/(2*(1+1/y)),al,lty="dotted")
       lines(al/(2+1/y),al,lty="dotdash")
       lines(al/2,al)

    result <- list(y=y,llk.nb1=llk.nb1)
    result
}
```

9.3 A function for drawing the graph contour plot of individual log-likelihood for NB1 model where $y = 5$.

```
nb1.llink <- function(y)   {
    # al=dispersion parameter
       al <- seq(0.1,10,length=100)
       k <- al        # a constant for ℓ(y + k, α; y)
       llk.nb1 <- outer(k,al,function(k,al)
    (lgamma(y+(y+k)/al)-lgamma((y+k)/al)-(y+k)*log(1+al)/al))
    # this excludes y! term

# Contour plot.
       contour(k,al,llk.nb1,zlim=c(10,5,4,3,2,1),main="y = 5",
    ylab=expression(alpha),xlab="k")
       lines(al/(2*(1+1/y)),al,lty="dotted")
       lines(al/(2+1/y),al,lty="dotdash")
       lines(al/2,al)

    result <- list(y=y,llk.nb1=llk.nb1)
    result
}
```

9.4 A function for drawing the graph contour plot of individual log-likelihood for NB1 model where $y = 10$.

```
nb1.llink <- function(y)   {
    # al=dispersion parameter
       al <- seq(0.1,10,length=100)
       k <- al        # a constant for ℓ(y + k, α; y)
       llk.nb1 <- outer(k,al,function(k,al)
    (lgamma(y+(y+k)/al)-lgamma((y+k)/al)-(y+k)*log(1+al)/al))
    # this excludes y! term

# Contour plot.
       contour(k,al,llk.nb1,zlim=c(10,5,4,3,2,1),main="y = 10",
    ylab=expression(alpha),xlab="k")
       lines(al/(2*(1+1/y)),al,lty="dotted")
       lines(al/(2+1/y),al,lty="dotdash")
       lines(al/2,al)

    result <- list(y=y,llk.nb1=llk.nb1)
    result
}
```

9.5 A function for drawing the graph contour plot of individual log-likelihood for NB1 model where $y = 20$.

```
nb1.llink <- function(y)   {
    # al=dispersion parameter
       al <- seq(0.1,10,length=100)
       k <- al        # a constant for ℓ(y + k, α; y)
       llk.nb1 <- outer(k,al,function(k,al)
    (lgamma(y+(y+k)/al)-lgamma((y+k)/al)-(y+k)*log(1+al)/al))
    # this excludes y! term

# Contour plot.
       contour(k,al,llk.nb1,zlim=c(10,5,4,3,2,1),main="y = 20",
    ylab=expression(alpha),xlab="k")
       lines(al/(2*(1+1/y)),al,lty="dotted")
       lines(al/(2+1/y),al,lty="dotdash")
       lines(al/2,al)

    result <- list(y=y,llk.nb1=llk.nb1)
    result
}
```

9.6 A function for drawing the graph contour plot of individual log-likelihood for NB1 model where $y = 40$.

```
nb1.llink <- function(y)   {
    # al=dispersion parameter
       al <- seq(0.1,10,length=100)
       k <- al        # a constant for ℓ(y + k, α; y)
       llk.nb1<- outer(k,al,function(k,al)
    (lgamma(y+(y+k)/al)-lgamma((y+k)/al)-(y+k)*log(1+al)/al))
    # this excludes y! term

# Contour plot.
       contour(k,al,llk.nb1,main="y = 40",ylab=expression(alpha),
    xlab="k")
       lines(al/(2*(1+1/y)),al,lty="dotted")
       lines(al/(2+1/y),al,lty="dotdash")
       lines(al/2,al)

    result <- list(y=y,llk.nb1=llk.nb1)
    result
}
```

10. A function for fitting the NB1 regression.

```
nb1.fit <- function(formula,X.mat)  {
# The method of Newton-Raphson.
      S.E <- 0;      y <- NULL
      y.pois <- glm(formula,family=poisson)
      lamb <- y.pois$fitted.values
      y <- y.pois$y;    beta <- coefficients(y.pois)    # Initial value for beta
      al <- (sum((y-lamb)^2/lamb)/y.pois$df.residual)-1    # Initial value
                                                            for alpha
# Based on the mean deviance.
      al.old <- al;      al.new <- 0
      al.diff <- al.old;      i <- 0
      beta.al <- c(beta,al)

# Calculate -2 x logL for a Poisson model.
      plikf <- -2*sum(y*log(lamb)-lamb-lgamma(y+1))

    while(al.diff > 0.0001)  {
# Calculate digamma and trigamma function.
      trigam <- -(trigamma(y+lamb/al)-trigamma(lamb/al))
```

```
        digam <- digamma(y+lamb/al)-digamma(lamb/al)

# Gradient vectios.

        s1 <- (al)^(-1)*(digam - log(1+al))*lamb

        S1 <- t(X.mat)%*%s1        # for beta

        s2 <- -(al)^(-2)*((lamb-y)/(1+1/(al))-lamb*log(1+al)+lamb*digam)

        S2 <- sum(s2)        # for alpha

# Create minus the 2nd derivative and information matrices.

        c1 <- -((s1)-(al)^(-2)*lamb*trigam*lamb)

        I1 <- (t(X.mat)%*%diag(c1[1:length(c1)]))%*%X.mat     # for beta

        i21 <- -(2*(al)^(-3)*((lamb-y)/(1+1/(al))-lamb*log(1+al)

    +lamb*digam))

        i22 <- -(al)^(-4)*((y-lamb)/((1+1/(al))^2)+al*lamb/(1+1/(al))

    -lamb^2*trigam)

        I2 <- i21+i22;      I2 <- sum(I2)       # for alpha

        i12 <- -((al)^(-3)*lamb*trigam -(al)^(-2)*digam+(al)^(-2)*log(1+al)

    -1/((al)*(1+al)))*lamb

        I12 <- t(X.mat)%*%i12        # for beta, alpha

# Creat a score vector and partitioned information matrix.

        S <- c(S1,S2)

        PI1 <- c(I12,I2)

        PI2 <- rbind(I1,t(I12))
```

```
        PI <- cbind(PI2,PI1)        # completed infomation matrix
        PI.inv <- solve(PI);      S.E <- sqrt(PI.inv)
        beta.al <- beta.al+S%*%PI.inv        # The N-R Method
        beta <- beta.al[1:length(beta.al)-1]
        lamb <- exp(X.mat%*%beta)        # Fitted vales
        al <- beta.al[length(beta.al)]

# Calculate -2 x log-likelihood for a NB1.
        llik <- -2*sum(lgamma(y+1/(al)*lamb)-lgamma(1/(al)*lamb)
    +y*log(al)-(y+lamb/(al))*log(1+al))
        llikf <- llik+2*sum(lgamma(y+1))        # this includes y! term
        AIC <- llikf+2*length(beta.al)

# updated al
        al.new <- al
        al.diff <- abs((al.old - al.new))        # Stopping rule
        al.old <- al.new;      i <- i+1
}
        se <- NULL
    for(i in 1:nrow(S.E))    {
        ii <- S.E[i,i];      se <- c(se,ii)
}
        se.beta <- round(se[1:length(beta)],7)
        len <- length(se.beta)+1
```

```
    se.al <- se[len];      len <- len+1
    zval.beta <- round(beta/se.beta,7)
    pval.beta <- 2*pnorm(abs(zval.beta),lower.tail=F)
    y.sum <- cbind(beta,se.beta,zval.beta,pval.beta)
  colnames(y.sum) <- c("Estimate","Std. Error","z value","Pr(> |z|)")
    name <- dimnames(X.mat)[[2]];      name <- name[2:length(name)]
    rownames(y.sum) <- c("Intercept ",name)
    cat("\n","==== NB1 model ====","\n")
    print(y.sum)
    cat("\n")

    nblik <- round(llikf,7)
    df <- round(length(y)-length(se),0)
    aic <- round(nblik+2*(length(se)),7)
    Alpha <- rbind(al,se.al);      colnames(Alpha) <- c(" ")
    rownames(Alpha) <- c("alpha :","S.E (alpha)")
    cat("\n","==== Overdispersion ====","\n")
    print(Alpha)
    cat("\n")

    LL <- rbind(plikf, nblik, df, aic)
  rownames(LL) <- c("-2 x Pois log-likelihood :","-2 x NB1 log-likelihood :",
"df.residuals :","AIC :")
    colnames(LL) <- c(" ")
```

```
        cat("\n","==== Log-likelihood ====","\n")
        print(LL)
        cat("\n")
    result <- list(beta.nb1=beta,al.nb1=al,fitted.values=lamb)
    result
}
```

11. A function for calculating and drawing the graph of the hat values for the NB1 regression.

11.1 A function for calculating and drawing the graph of the hat values, under working model: $\ln(\boldsymbol{\mu}) = 1.75 + 1.05\boldsymbol{x}_1$.

```
hii.nb1 <- function(al,beta,X.mat)  {
        ys <- 0;      A <- NULL
        mu <- exp(X.mat%*%beta);      n <- length(mu)
        x1 <- X.mat[,2]
        source("nb1fit1.txt",echo=TRUE)
# Simulated envelop.
    for(i in 1:1000)    {
        e01 <- rgamma(n,mu/al,1)
        eij <- e01*(al/mu)
```

```
        ys <- rpois(n,mu*eij)

        nb1.e1 <- nb1.fit(ys~x1,X.mat)

        lamb <- nb1.e1$fitted.values

        al.sim <- nb1.e1$al.nb1

# Calculated the hat values.

        wii.sim <- lamb/(1+al.sim)

        wii.sim <- c(wii.sim)

        W.sim <- diag(wii.sim)

        H.sim <- solve(t(X.mat)%*%W.sim%*%X.mat)

        H.sim <- sqrt(W.sim)%*%X.mat%*%H.sim%*%t(X.mat)
    %*%sqrt(W.sim)

        hii.sim <- diag(H.sim)    # hat values (h_ii)

        rank <- sum(hii.sim)       # rank values of h_ii

        min.hii <- min(hii.sim)    # minimum values of h_ii

        mean.hii <- mean(hii.sim)     # mean values of h_ii

        max.hii <- max(hii.sim)       # maximum values of h_ii

        h <- c(min.hii,max.hii,mean.hii,rank)

        A <- c(A,h)

}

        hii.mat <- matrix(A,ncol=4,byrow=TRUE)

        colnames(hii.mat) <- c("min","max","mean","rank")
```

```
# Calculated the average of minimum, maximum, mean and rank values.
        av.min <- mean(hii.mat[,1])
        av.max <- mean(hii.mat[,2])
        av.mean <- mean(hii.mat[,3])
        av.rank <- mean(hii.mat[,4])
        av.hii <- c(av.min,av.max,av.mean,av.rank)
        av.hii <- matrix(av.hii,ncol=4,byrow=TRUE)
        colnames(av.hii) <- c("min","max","mean","rank")
        av.hii <- round(av.hii,3)

# Plot of h_ii.
        hii.mat1 <- round(hii.mat[,1:3],3)
        number <- seq(1,3000,by=1)
        plot(number,hii.mat1,type="n",xlab="simulated sample numbers",
    ylab=expression(h[ii]),xlim=c(0,1000))
        num <- seq(1,1000,by=1)
        lines(num,hii.mat[,1],lty="dotted")
        lines(num,hii.mat[,2],lty="dashed")
        lines(num,hii.mat[,3])

     result <- list(av.hii=av.hii,n=n)
     result
}
```

11.2 A function for calculating and drawing the graph of the hat values, under working model: $\ln(\boldsymbol{\mu}) = -0.45 + 0.75\boldsymbol{x}_2$.

```
hii.nb1 <- function(al,beta,X.mat)   {
        ys <- 0;      A <- NULL
        mu <- exp(X.mat%*%beta);      n <- length(mu)
        x2 <- X.mat[,2]
        source("nb1fit1.txt",echo=TRUE)

# Simulated envelop.
    for(i in 1:1000)    {
        e01 <- rgamma(n,mu/al,1)
        eij <- e01*(al/mu)
        ys <- rpois(n,mu*eij)
        nb1.e2 <- nb1.fit(ys~x2,X.mat)
        lamb <- nb1.e2$fitted.values
        al.sim <- nb1.e2$al.nb1

# Calculated the hat values.
        wii.sim <- lamb/(1+al.sim)
        wii.sim <- c(wii.sim)
        W.sim<- diag(wii.sim)
```

```
        H.sim<- solve(t(X.mat)%*%W.sim%*%X.mat)
        H.sim<- sqrt(W.sim)%*%X.mat%*%H.sim%*%t(X.mat)
    %*%sqrt(W.sim)
        hii.sim <- diag(H.sim)  # hat values (h_ii)

        rank <- sum(hii.sim)       # rank values of h_ii
        min.hii <- min(hii.sim)    # minimum values of h_ii
        mean.hii <- mean(hii.sim)     # mean values of h_ii
        max.hii <- max(hii.sim)       # maximum values of h_ii
        h <- c(min.hii,max.hii,mean.hii,rank)
        A <- c(A,h)
}

        hii.mat <- matrix(A,ncol=4,byrow=TRUE)
        colnames(hii.mat) <- c("min","max","mean","rank")

# Calculated the average of minimum, maximum, mean and rank values.
        av.min <- mean(hii.mat[,1])
        av.max <- mean(hii.mat[,2])
        av.mean <- mean(hii.mat[,3])
        av.rank <- mean(hii.mat[,4])
        av.hii <- c(av.min,av.max,av.mean,av.rank)
        av.hii <- matrix(av.hii,ncol=4,byrow=TRUE)
        colnames(av.hii) <- c("min","max","mean","rank")
```

```
        av.hii <- round(av.hii,3)

# Plot of h_ii.

        hii.mat1 <- round(hii.mat[,1:3],3)

        number <- seq(1,3000,by=1)

        plot(number,hii.mat1,type="n",xlab="simulated sample numbers",

    ylab=expression(h[ii]),xlim=c(0,1000))

        num <- seq(1,1000,by=1)

        lines(num,hii.mat[,1],lty="dotted")

        lines(num,hii.mat[,2],lty="dashed")

        lines(num,hii.mat[,3])

     result <- list(av.hii=av.hii,n=n)

     result

}
```

11.3 A function for calculating and drawing the graph of the hat values, under working model: $\ln(\boldsymbol{\mu}) = 1.25 - 2.45\boldsymbol{x}_1 + 0.85\boldsymbol{x}_2$.

```
hii.nb1 <- function(al,beta,X.mat)   {
        ys <- 0;      A <- NULL
        mu <- exp(X.mat%*%beta);      n <- length(mu)
        x1 <- X.mat[,2];      x2 <- X.mat[,3]
        source("nb1fit1.txt",echo=TRUE)

# Simulated envelop.
    for(i in 1:1000)    {
        e01 <- rgamma(n,mu/al,1);      eij <- e01*(al/mu)
        ys <- rpois(n,mu*eij)
        nb1.e3 <- nb1.fit(ys~x1+x2,X.mat)
        lamb <- nb1.e3$fitted.values
        al.sim <- nb1.e3$al.nb1

# Calculated the hat values.
        wii.sim <- lamb/(1+al.sim)
        wii.sim <- c(wii.sim)
        W.sim<- diag(wii.sim)
        H.sim<- solve(t(X.mat)%*%W.sim%*%X.mat)
```

```
        H.sim<- sqrt(W.sim)%*%X.mat%*%H.sim%*%t(X.mat)
    %*%sqrt(W.sim)
        hii.sim <- diag(H.sim)    # hat values (h_ii)

        rank <- sum(hii.sim)        # rank values of h_ii
        min.hii <- min(hii.sim)    # minimum values of h_ii
        mean.hii <- mean(hii.sim)     # mean values of h_ii
        max.hii <- max(hii.sim)       # maximum values of h_ii
        h <- c(min.hii,max.hii,mean.hii,rank)
        A <- c(A,h)
}

        hii.mat <- matrix(A,ncol=4,byrow=TRUE)
        colnames(hii.mat) <- c("min","max","mean","rank")

# Calculated the average of minimum, maximum, mean and rank values.
        av.min <- mean(hii.mat[,1])
        av.max <- mean(hii.mat[,2])
        av.mean <- mean(hii.mat[,3])
        av.rank <- mean(hii.mat[,4])
        av.hii <- c(av.min,av.max,av.mean,av.rank)
        av.hii <- matrix(av.hii,ncol=4,byrow=TRUE)
        colnames(av.hii) <- c("min","max","mean","rank")
        av.hii <- round(av.hii,3)
```

```
# Plot of h_ii.
        hii.mat1 <- round(hii.mat[,1:3],3)
        number <- seq(1,3000,by=1)
        plot(number,hii.mat1,type="n",xlab="simulated sample numbers",
    ylab=expression(h[ii]),xlim=c(0,1000))
        num <- seq(1,1000,by=1)
        lines(num,hii.mat[,1],lty="dotted")
        lines(num,hii.mat[,2],lty="dashed")
        lines(num,hii.mat[,3])
     result <- list(av.hii=av.hii,n=n)
     result
}
```

12. A function for calculating the NB1 standardized deviance residual.

12.1 A function for calculating the NB1 standardized deviance residual, under working model: $\ln(\boldsymbol{\mu}) = 1.75 + 1.05\boldsymbol{x}_1$.

```
sd.nb1 <- function(al,beta,X.mat)   {
        ys <- 0;      B <- NULL
        KS <- NULL;      S <- NULL
```

```
        mu <- exp(X.mat%*%beta);      n <- length(mu)
        x1 <- X.mat[,2]
        source("nb1fit1.txt",echo=TRUE)
# Simulated envelop.
    for(i in 1:1000)    {
        e01 <- rgamma(n,mu/al,1)
        eij <- e01*(al/mu)
        ys <- rpois(n,mu*eij)
        nb1.e1 <- nb1.fit(ys~x1,X.mat)
        lamb <- nb1.e1$fitted.values
        al.sim <- nb1.e1$al.nb1
# Calculated the hat values.
        wii.sim <- lamb/(1+al.sim)
        wii.sim <- c(wii.sim)
        W.sim<- diag(wii.sim)
        H.sim<- solve(t(X.mat)%*%W.sim%*%X.mat)
        H.sim<- sqrt(W.sim)%*%X.mat%*%H.sim%*%t(X.mat)
    %*%sqrt(W.sim)
        hii.sim <- diag(H.sim)
# Calculated standardized deviance residual.
        dlg_y <- lgamma(ys+((ys+(al.sim/2))/al.sim))
    -lgamma((ys+(al.sim/2))/al.sim)
```

```
        dlg_lamb <- lgamma(ys+(lamb/al.sim))-lgamma(lamb/al.sim)

        dev <- (ys+(al.sim/2)-lamb)*log(1+al.sim)/al.sim

        dev <-dev.nb1-dlg_y+dlg_lamb

        dev <- dev.nb1*(-2)     # for   y > 0

        dev.y0 <- 2*lamb*log(1+al.sim)/al.sim       # for   y = 0

        dev.nb1 <- ifelse(ys == 0,dev.y0,dev)

        dres.nb1 <- sqrt(dev.nb1)

        dres.nb1 <- ifelse(is.na(dres.nb1),0,dres.nb1)

        dres.nb1 <- ifelse(ys<lamb,-dres.nb1,dres.nb1)

        sd.nb1 <- dres.nb1/sqrt(1-hii.sim) # NB1 standardized deviance
                                                    residuals

        min.sd <- min(sd.nb1)

        max.sd <- max(sd.nb1)

        mean.sd <- mean(sd.nb1)

        var.sd <- var(sd.nb1)

        sd <- c(min.sd,max.sd,mean.sd,var.sd)

        B <- c(B,sd)

# Test N(0,1).

        p.ks <- ks.test(sd.nb1,"pnorm",0,1)        # Kolmogorov-Smirnov

        KS <- c(KS,p.ks$p.value)

        p.sha <- shapiro.test(sd.nb1)       # Shapiro-Wilk

        S <- c(S,p.sha$p.value)

}
```

```
        sd.mat <- matrix(B,ncol=4,byrow=TRUE)
        colnames(sd.mat) <- c("min","max","mean","variance")
# Calculated the average of minimum, maximum, mean and
variance values.
        av.min <- mean(sd.mat[,1])
        av.max <- mean(sd.mat[,2])
        av.mean <- mean(sd.mat[,3])
        av.var <- mean(sd.mat[,4])
        av.sd <- c(av.min,av.max,av.mean,av.var)
        av.sd <- matrix(av.sd,ncol=4,byrow=TRUE)
        colnames(av.sd) <- c("min","max","mean","variance")
        av.sd <- round(av.sd,3)
# Calculated the proportion of times the p-value is less than or equal
the nominal significance size of 0.05.
        KS1 <- KS<=0.05
        S1 <- S<=0.05
        k <- sum(KS1)/1000
        m <- sum(S1)/1000
    result <- list(av.sd=av.sd,n=n,m=m,k=k)
    result
}
```

12.2 A function for calculating the NB1 standardized deviance residual, under working model: $\ln(\boldsymbol{\mu}) = -0.45 + 0.75\boldsymbol{x}_2$.

```
sd.nb1 <- function(al,beta,X.mat)   {
        ys <- 0;      B <- NULL
        KS <- NULL;      S <- NULL
        mu <- exp(X.mat%*%beta);      n <- length(mu)
        x2 <- X.mat[,2]
        source("nb1fit1.txt",echo=TRUE)

# Simulated envelop.
    for(i in 1:1000)    {
        e01 <- rgamma(n,mu/al,1)
        eij <- e01*(al/mu)
        ys <- rpois(n,mu*eij)
        nb1.e2 <- nb1.fit(ys~x2,X.mat)
        lamb <- nb1.e2$fitted.values
        al.sim <- nb1.e2$al.nb1

# Calculated the hat values.
        wii.sim <- lamb/(1+al.sim)
        wii.sim <- c(wii.sim)
```

```
        W.sim<- diag(wii.sim)

        H.sim<- solve(t(X.mat)%*%W.sim%*%X.mat)

        H.sim<- sqrt(W.sim)%*%X.mat%*%H.sim%*%t(X.mat)
    %*%sqrt(W.sim)

        hii.sim <- diag(H.sim)

# Calculated standardized deviance residual.

        dlg_y <- lgamma(ys+((ys+(al.sim/2))/al.sim))
    -lgamma((ys+(al.sim/2))/al.sim)

        dlg_lamb <- lgamma(ys+(lamb/al.sim))-lgamma(lamb/al.sim)

        dev <- (ys+(al.sim/2)-lamb)*log(1+al.sim)/al.sim

        dev <-dev.nb1-dlg_y+dlg_lamb

        dev <- dev.nb1*(-2)     # for   y > 0

        dev.y0 <- 2*lamb*log(1+al.sim)/al.sim       # for   y = 0

        dev.nb1 <- ifelse(ys == 0,dev.y0,dev)

        dres.nb1 <- sqrt(dev.nb1)

        dres.nb1 <- ifelse(is.na(dres.nb1),0,dres.nb1)

        dres.nb1 <- ifelse(ys<lamb,-dres.nb1,dres.nb1)

        sd.nb1 <- dres.nb1/sqrt(1-hii.sim) # NB1 standardized deviance
                                                residuals

        min.sd <- min(sd.nb1)

        max.sd <- max(sd.nb1)

        mean.sd <- mean(sd.nb1)

        var.sd <- var(sd.nb1)
```

```
        sd <- c(min.sd,max.sd,mean.sd,var.sd)

        B <- c(B,sd)

# Test N(0,1).

        p.ks <- ks.test(sd.nb1,"pnorm",0,1)        # Kolmogorov-Smirnov

        KS <- c(KS,p.ks$p.value)

        p.sha <- shapiro.test(sd.nb1)       # Shapiro-Wilk

        S <- c(S,p.sha$p.value)

}

        sd.mat <- matrix(B,ncol=4,byrow=TRUE)

        colnames(sd.mat) <- c("min","max","mean","variance")

# Calculated the average of minimum, maximum, mean and variance values.

        av.min <- mean(sd.mat[,1])

        av.max <- mean(sd.mat[,2])

        av.mean <- mean(sd.mat[,3])

        av.var <- mean(sd.mat[,4])

        av.sd <- c(av.min,av.max,av.mean,av.var)

        av.sd <- matrix(av.sd,ncol=4,byrow=TRUE)

        colnames(av.sd) <- c("min","max","mean","variance")

        av.sd <- round(av.sd,3)

# Calculated the proportion of times the p-value is less than or equal

the nominal significance size of 0.05.

        KS1 <- KS<=0.05
```

```
        S1 <- S<=0.05
        k <- sum(KS1)/1000
        m <- sum(S1)/1000
    result <- list(av.sd=av.sd,n=n,m=m,k=k)
    result
}
```

12.3 A function for calculating the NB1 standardized deviance residual, under working model: $\ln(\boldsymbol{\mu}) = 1.25 - 2.45\boldsymbol{x}_1 + 0.85\boldsymbol{x}_2$.

```
sd.nb1 <- function(al,beta,X.mat)   {
        ys <- 0;      B <- NULL
        KS <- NULL;      S <- NULL
        mu <- exp(X.mat%*%beta);      n <- length(mu)
        x1 <- X.mat[,2];      x2 <- X.mat[,3]
        source("nb1fit1.txt",echo=TRUE)

# Simulated envelop.
    for(i in 1:1000)    {
        e01 <- rgamma(n,mu/al,1)
```

```
        eij <- e01*(al/mu)

        ys <- rpois(n,mu*eij)

        nb1.e3 <- nb1.fit(ys~x1+x2,X.mat)

        lamb <- nb1.e3$fitted.values

        al.sim <- nb1.e3$al.nb1

# Calculated the hat values.

        wii.sim <- lamb/(1+al.sim)

        wii.sim <- c(wii.sim)

        W.sim<- diag(wii.sim)

        H.sim<- solve(t(X.mat)%*%W.sim%*%X.mat)

        H.sim<- sqrt(W.sim)%*%X.mat%*%H.sim%*%t(X.mat)
    %*%sqrt(W.sim)

        hii.sim <- diag(H.sim)

# Calculated standardized deviance residual.

        dlg_y <- lgamma(ys+((ys+(al.sim/2))/al.sim))
    -lgamma((ys+(al.sim/2))/al.sim)

        dev <- (ys+(al.sim/2)-lamb)*log(1+al.sim)/al.sim

        dev <-dev.nb1-dlg_y+dlg_lamb

        dev <- dev.nb1*(-2)     # for   y > 0

        dev.y0 <- 2*lamb*log(1+al.sim)/al.sim       # for   y = 0

        dev.nb1 <- ifelse(ys == 0,dev.y0,dev)

        dres.nb1 <- sqrt(dev.nb1)
```

```
        dres.nb1 <- ifelse(is.na(dres.nb1),0,dres.nb1)
        dres.nb1 <- ifelse(ys<lamb,-dres.nb1,dres.nb1)
        sd.nb1 <- dres.nb1/sqrt(1-hii.sim) # NB1 standardized deviance
                                                residuals
        min.sd <- min(sd.nb1)
        max.sd <- max(sd.nb1)
        mean.sd <- mean(sd.nb1)
        var.sd <- var(sd.nb1)
        sd <- c(min.sd,max.sd,mean.sd,var.sd)
        B <- c(B,sd)
# Test N(0,1).
        p.ks <- ks.test(sd.nb1,"pnorm",0,1)       # Kolmogorov-Smirnov
        KS <- c(KS,p.ks$p.value)
        p.sha <- shapiro.test(sd.nb1)      # Shapiro-Wilk
        S <- c(S,p.sha$p.value)
}
        sd.mat <- matrix(B,ncol=4,byrow=TRUE)
        colnames(sd.mat) <- c("min","max","mean","variance")
# Calculated the average of minimum, maximum, mean and variance values.
        av.min <- mean(sd.mat[,1])
        av.max <- mean(sd.mat[,2])
        av.mean <- mean(sd.mat[,3])
```

```
        av.var <- mean(sd.mat[,4])

        av.sd <- c(av.min,av.max,av.mean,av.var)

        av.sd <- matrix(av.sd,ncol=4,byrow=TRUE)

        colnames(av.sd) <- c("min","max","mean","variance")

        av.sd <- round(av.sd,3)

# Calculated the proportion of times the p-value is less than or equal
the nominal significance size of 0.05.

        KS1 <- KS<=0.05

        S1 <- S<=0.05

        k <- sum(KS1)/1000

        m <- sum(S1)/1000

    result <- list(av.sd=av.sd,n=n,m=m,k=k)

    result

}
```

VITAE

Name Miss Mallika Inprom

Student ID 5510220153

Educational Attainment

Degree	Name of Institution	Year of Graduation
B.Sc. (Mathematics)	Prince of Songkla University	2011

Scholarship Awards during Enrolment

Teaching Assistantship from Faculty of Science, Prince of Songkla University, 2012-2013.

List of Publication and Proceeding

M. Inprom, N. Jansakul and K. Samart, Deviance Residuals of Linear Mean-variance Negative Binomial Models, *The* 19^{th} *Annual Meeting in Mathematics (AMM2014).,* Thammasat University, Bangkok, Thailand, March 20-22, 2014, 71–80.