2001

Goodness-of-fit, score test, zero-inflation and over-dispersion in generalized linear models.

Dianliang. Deng
University of Windsor

Goodness of Fit, Score Test, Zero-Inflation and Over-Dispersion in Generalized Linear Models

Dianliang Deng

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

Windsor, Ontario, Canada
2001

0-612-62314-9

938513

Abstract

Data in the form of counts and proportions arise in many fields such as public health, toxicology, epidemiology, sociology, psychology, engineering, agriculture and so on. One frequently encountered problem in these data is that simple models such as the Poisson and the binomial models fail to explain the variation that exists in these data. Often, data exhibit extra-dispersion (over or under dispersion). Another complication in data in the form of counts and proportions is that data are sometimes too sparse. That is, smaller values have greater tendency to occur. In the Poisson case counts that occur are generally small and in the binomial case the binomial denominators are often small. So, we need valid procedures to detect departures from the simple models.

Traditionally, goodness of fit of generalized linear models is assessed by using the Pearson χ^2 statistic or the likelihood ratio χ^2 statistic. These goodness of fit tests do not perform well for sparse data. In this thesis we develop goodness of fit tests of the generalized linear model with non-canonical links for data that are extensive but sparse. We derive approximations to the first three moments of the deviance statistic. A supplementary estimating equation is proposed from which the modified deviance statistic is obtained. Applications of the modified deviance statistic to binomial and Poisson data are shown. A simulation study is conducted to compare the behavior, in terms of size and power, of the modified deviance statistic and the modified Pearson statistic developed earlier by Farrington (1996). Three sets of data with different degreees of sparseness and different link functions are analyzed. The simulation results and examples indicate that both the modified Pearson statistic and the modified deviance statistic perform well in terms of holding nominal levels. However, the modified deviance statistic shows much better power properties for the range of parameters investigated under the alternative hypothesis. Theses results also answer a question posed by Farrington (1996) and extend results of McCullagh (1986) for Poisson log-linear models.

In some instances a score or a $C(\alpha)$ statistic performs well. In this thesis we also develop a score test statistic to assess goodness of fit of the generalized linear model for data that are extensive but sparse. The performance of this statistic is then compared with the modified Pearson statistic. Results of simulation show that both the modified score test statistic developed in our paper and the modified Pearson statistic developed by Farrington (1996) maintain nominal levels. However the modified score test has some edge over the modified Pearson statistic in terms of power.

In practice, sometimes, discrete data contain excess zeros that can not be explained by a simple model. In this thesis we develop score tests for testing zero-inflation in generalized linear models. These score tests are then applied to binomial models and Poisson models and their performances are evaluated. A limited simulation study shows that the score tests reasonably maintain the nominal levels. The power of the tests for detecting zero-inflation increases very slowly for Poisson mean μ or binomial parameter p. For large values of μ and p power increases very fast and approaches 1.0 even for moderate zero-inflation.

A discrete generalized linear model (Poisson or binomial) may fail to fit a set of data having a lot of zeros either because of zero-inflation only, because of over-dispersion only, or because there is zero-inflation as well as over-dispersion in the data. In this thesis we obtain score tests (i) for zero-inflation in presence of over-dispersion, (ii) for over-dispersion in presence of zero-inflation, and (iii) simultaneously for testing for zero-inflation and over-dispersion. For Poisson and binomial data these score tests are compared with those obtained from the zero-inflated negative binomial model and the zero-inflated beta-binomial model. Some simulations are performed for Poisson data to study type I error properties of the tests. In general the score tests developed here hold nominal levels reasonably well. The data sets are analyzed to illustrate model section procedure by the score tests.

Acknowledgements

I wish to express my deep appreciation to my supervisor Dr. S. R. Paul for his guiding me into this statistics field. During the past several years of study, Dr. Paul gave me continuous guidance and encouragement, intellectual stimulation and financial support. More importantly, his spirit for academic research will stimulate me in my future career. In addition, I would like to express my thanks to professors Dr. K. Y. Fung, Dr. M. Hlynka, Dr. C. S. Wong and Dr. D. S. Tracy for their offering me encouragement and help during the study. I am also very grateful to Dr. C. B. Dean of Simon Fraser University and Dr. J. J. H. Ciborowski of Department of Biological Science for their helpful suggestions and comments.

I would like to thank the Department of Mathematics and Statistics for providing financial support in terms of graduate teaching assistantships, Natural Sciences and Engineering Research Council of Canada for Postgraduate Scholarship (B) and Student Support Branch, Ministry of Education and Training of Ontario for Ontario Graduate Scholarship in Science and Technology throughout my graduate studies.

Finally, I would like to express my deep thanks to my wife Qiong, son Yilun for their love and support during my studies and to my parents for their guidance and support throughout my life.

Contents

List of Tables

Chapter 1

Introduction

Generalized linear models are extensively used to analyze data in the form of counts and proportions. These data arise in public health, toxicology, epidemiology, sociology, psychology, engineering and agriculture. Data that arise in practice often exhibit extra-dispersion (over or under dispersion). Also, sometimes data contain too many zeros. Simple models such as the binomial and the Poisson models fail to explain these kinds of variations in the data. Another complication in data in the form of counts and proportions is that data are too sparse. That is, smaller values have greater tendency to occur. In the Poisson case counts that occur are generally small and in the binomial case the binomial denominators are often small. In this thesis we study procedures for testing goodness of fit of the generalized linear model.

Traditionally, goodness of fit of generalized linear models is assessed by using the Pearson χ^2 statistic or the likelihood ratio χ^2 statistic. These goodness of fit tests do not perform well for sparse data. Koehler and Larntz (1980) and Koehler (1986) make modifications of the Pearson and the likelihood ratio statistics by using moments of their unconditional distributions. McCullagh (1985) gives approximations to the first three moments of the unconditional and the conditional distributions of Pearson χ^2-statistic for canonical exponential family regression models. Farrington (1996) extends the results of McCullagh (1985) and obtains approximations to the first three moments of the conditional and the unconditional distributions of this statistic for non-canonical generalized linear models. He develops a modified Pearson statistic by using a supplementary unbiased estimating equation. He also shows that the modified Pearson statistic is asymptotically independent of the regression parameter estimates in generalized linear models. Farrington (1996) conducts a goodness of fit test of generalized linear models to sparse data by using a standardized modified Pearson statistic using the approximations to the conditional mean and conditional variance of the modified Pearson statistic. In this thesis we develop procedures for testing goodness of fit of generalized linear models with non-canonical links to sparse data against over-dispersion.

Johnson, Kotz and Kemp (1992), Lambert (1992) and Broek (1995) discuss situations in which there are many extra zeros in the count data. Broek (1995) develops a score test statistic for testing zero-inflation in a Poisson distribution. In this thesis we develop procedures for testing for zero-inflation in generalized linear models. A discrete generalized linear model (Poisson or binomial) may fail to fit a set of data having a lot of zeros either because of zero-inflation only, because of over-dispersion only, or because there is zero-inflation as well as over-dispersion in the data. In this thesis, we also develop procedures for testing (i) for zero-inflation in the presence of over-dispersion, (ii) for over-dispersion in the presence of zero-inflation, and (iii) simultaneously for zero-inflation and over-dispersion.

In the sequel, this thesis will proceed as follows. In Chapter 2, we provide some preliminaries as background. Theory for $C(\alpha)$ tests and score tests are given. Tensor notation and cumulants of unconditional and conditional distributions of test statistics are stated. Results concerning the modified Pearson statistic derived by Farrington (1996) are also reviewed.

In Chapter 3, we propose a supplementary estimating equation to derive a modified deviance statistic. We then obtain approximations to the first three moments of the conditional and the unconditional distributions of the modified deviance statistic. The results obtained are applied to binomial and Poisson data. A performance study is conducted to compare the modified Pearson statistic derived by Farrington and the modified deviance statistic obtained in this chapter.

The issues of goodness of fit in generalized linear models are further discussed in Chapter 4. Although the performance of the modified deviance statistic is better than that of the modified Pearson statistic, we obtain a score test statistic to assess goodness of fit in generalized linear models. This statistic has a very simple form and is closely related to the modified Pearson statistic given by Farrington (1996). Some simulations are conducted.

In Chapter 5 the emphasis is focused on the issues of zero-inflation that may occur in discrete data. We first present a zero-inflated generalized linear model. We then develop score test statistics to test for zero-inflation in generalized linear models. Score tests for testing for zero-inflation in binomial data and Poisson data are obtained as special cases. A small simulation is conducted and an example is given.

In Chapter 6 we consider a zero-inflated over-dispersed generalized linear model. Using this model we obtain score tests (i) for zero-inflation in the presence of over-dispersion, (ii) for over-dispersion in the presence of zero-inflation, and (iii) simultaneously for zero-inflation and over-dispersion. For Poisson and binomial data these score test statistics are compared with those from the zero-inflated negative binomial model and the zero-inflated beta-binomial model. Properties of the test statistics are further investigated using simulations. Two examples are given for illustrative purposes.

In the last chapter, the findings of the thesis are summarized and some ideas for future research regarding goodness of fit tests in generalized linear model are developed and discussed.

Chapter 2

Some preliminaries and review of current literature

2.1 The $C(\alpha)$ test and the score test

Let $y = (y_1, ..., y_n)$ be the observations from a vector of random variables $Y = (Y_1, Y_2, ..., Y_n)$ with distribution function $f(y; \theta, \phi)$ where $\theta = (\theta_1, ..., \theta_p)'$ are parameters of interest and $\phi = (\phi_1, ..., \phi_s)'$ are nuisance parameters. Let $L(\theta, \phi; y)$ be the likelihood function of the data y and l be the log-likelihood function of the data. The partial derivatives evaluated at $\theta = \theta_0 = (\theta_{10}, ..., \theta_{p0})'$ are

$$\psi = \frac{\partial l}{\partial \theta}|_{\theta=\theta_0} = \left[\frac{\partial l}{\partial \theta_1}, ..., \frac{\partial l}{\partial \theta_p}\right]' |_{\theta=\theta_0},$$

and

$$\gamma = \frac{\partial l}{\partial \phi}|_{\theta=\theta_0} = \left[\frac{\partial l}{\partial \phi_1}, ..., \frac{\partial l}{\partial \phi_s}\right]' |_{\theta=\theta_0}.$$

Cramer (1946) has shown that under the null hypothesis and mild regularity conditions, $(\frac{\partial l}{\partial \theta}, \frac{\partial l}{\partial \phi})$ follows multivariate a normal distribution asymptotically with mean vector 0 and variance covariance matrix $I^{-1}(\theta, \phi)$, where

$$I(\theta, \phi) = \begin{bmatrix} I_{\theta\theta} & I_{\theta\phi} \\ I'_{\theta\phi} & I_{\phi\phi} \end{bmatrix},$$

is the Fisher information with elements

$$I_{\theta\theta} = E\left(\frac{\partial^2 l}{\partial\theta\partial\theta'}|_{\theta=\theta_0}\right),$$

$$I_{\theta\phi} = E\left(\frac{\partial^2 l}{\partial\theta\partial\phi'}|_{\theta=\theta_0}\right),$$

and

$$I_{\theta\theta} = E\left(\frac{\partial^2 l}{\partial\theta\partial\theta'}|_{\theta=\theta_0}\right).$$

Define $S = \frac{\partial l}{\partial\theta} - B\frac{\partial l}{\partial\phi}$, where B is the partial regression coefficient matrix obtained by regressing $\frac{\partial l}{\partial\theta}$ on $\frac{\partial l}{\partial\phi}$. From Bartlett (1953), $B = I_{\theta\phi}I_{\phi\phi}^{-1}$ and the dispersion matrix of S is $I_{\theta\theta\cdot\phi} = I_{\theta\theta} - I_{\theta\phi}I_{\phi\phi}^{-1}I_{\phi\theta}$.

Thus S is multivariate normal with mean vector 0 and variance-covariance matrix $I_{\theta\theta\cdot\phi}$, i.e.

$$S \sim MN(0, I_{\theta\theta\cdot\phi}).$$

Hence, following Neyman (1959), $S'I_{\theta\theta\cdot\phi}^{-1}S \sim \chi^2_{(p)}$. But the above expression depends on the nuisance parameters $\phi = (\phi_1, ..., \phi_s)'$, which makes the statistic unsuitable to use for testing the null hypothesis. Moran (1970) suggests that we replace the unknown nuisance parameter ϕ by its $\sqrt{n}$-consistent estimator(n=number of observations used in estimating the parameters), obtained from the data. Let $\tilde{\phi} = (\tilde{\phi}_1, ..., \tilde{\phi}_s)'$ be some $\sqrt{n}$-consistent estimator of the parameter $\phi = (\phi_1, ..., \phi_s)'$. Hence following Neyman (1959),

$$\chi^2_{C(\alpha)} = \tilde{S}'\tilde{I}_{\theta\theta\cdot\phi}^{-1}\tilde{S}$$

is asymptotically distributed as χ^2 distribution with p degrees of freedom. Note that if we replace the nuisance parameter ϕ by its maximum likelihood estimate (MLE) $\hat{\phi}$, then the score function S reduces to ψ. The $C(\alpha)$ statistic then reduces to $\hat{\psi}'\hat{I}_{\theta\theta\cdot\phi}^{-1}\hat{\psi}$ which is referred to as a score test (Rao, 1947). The score test is asymptotically equivalent to the likelihood ratio test (Moran, 1970; Cox and Hinkley, 1974).

The score test (Rao, 1947) is a special case of the more general $C(\alpha)$ test (Neyman, 1959) in which the nuisance parameters are replaced by maximum likelihood estimates which are $\sqrt{n}$-consistent estimates. The score test is particularly appealing as we have only to study the distribution of the test statistic under the null hypothesis. Potential drawbacks to the use of the likelihood ratio and Wald tests include the fact that both require estimates of the parameters under alternative hypotheses and often they do not maintain nominal level. Advantages of the score or the $C(\alpha)$ class of tests are: (i) it often maintains, at least approximately, a preassigned level of significance (Bartoo and Puri, 1967), (ii) it requires estimates of the parameters only under the null hypothesis, and (iii) it often produces a statistic which is simple to calculate. These tests are robust in the sense that their optimality remains true whatever the form of the distribution assumed for the data under the alternative hypothesis - a property called robustness of optimality by Neyman and Scott (1966). For more discussion on the choice of $C(\alpha)$ tests see Breslow (1990).

2.2 Tensor notation and cumulants

McCullagh (1987) points out that with appropriate choice of notation many multivariate statistical calculations can be made simpler and more transparent than the corresponding univariate calculation. This simplicity can be achieved through systematic use of index notation and special arrays called tensors.

2.2.1 Notational conventions and cumulants

Let X be a p-dimensional random variable with components $X^1, ..., X^p$, and let a_i, a_{ij} and a_{ijk} be a vector, matrix and array respectively. We introduce convenient index notation,

$$a_i X^i, a_{ij} X^i X^j, \text{ and } a_{ijk} X^i X^j X^k$$

and so on. These notations mean the implied summation over any index repeated once as a superscript and once as a subscript. That is,

$$a_i X^i = \sum_i a_i X^i, \qquad \text{(linear combination)}$$

$$a_{ij} X^i X^j = \sum_{i,j} a_{ij} X^i X^j (= X'AX) \qquad \text{(quadratic form)}$$

and

$$a_{ijk} X^i X^j X^k = \sum_{i,j,k} a_{ijk} X^i X^j X^k. \qquad \text{(cubic form)}$$

Note that the range of summation is not stated explicitly but is implied by the positions of the superscripts and subscripts. Further, let Y be a vector random variable with components $Y^1, ..., Y^q$, each of which is linear in X. We may write

$$Y^r = a_i^r X^i,$$

where $r = 1, ..., q$ is known as a free index. Similarly, we may write

$$Y^r = a_{ij}^r X^i X^j.$$

Now, we discuss the notations of moments and cumulants of X. We define the moments of X about the origin

$$\kappa^i = E(X^i),$$

$$\kappa^{ij} = E(X^i X^j),$$

$$\kappa^{ijk} = E(X^i X^j X^k)$$

and so on. We assume that the above moments are finite.

Consider the infinite series

$$M_X(\xi) = 1 + \xi_i \kappa^i + \xi_i \xi_j \kappa^{ij}/2! + \xi_i \xi_j \xi_k \kappa^{ijk}/3! + \xi_i \xi_j \xi_k \xi_l \kappa^{ijkl}/4! + ...$$

which we assume to be convergent for all $|\xi|$ sufficiently small.

The sum may be written as

$$M_X(\xi) = E\{\exp(\xi_i X^i)\}$$

and the moments are just the partial derivatives of $M_X(\xi)$ evaluated at $\xi = 0$.

The cumulants are most easily defined via their generating function,

$$K_X(\xi) = \log M_X(\xi),$$

which has an expression

$$\begin{aligned} K_X(\xi) &= \xi_i\kappa^i + \xi_i\xi_j\kappa^{i,j}/2! + \xi_i\xi_j\xi_k\kappa^{i,j,k}/3! + \\ &\quad \xi_i\xi_j\xi_k\xi_l\kappa^{i,j,k,l}/4! + \ldots \quad . \end{aligned}$$

This expression implicitly defines all the cumulants, here denoted by $\kappa^i, \kappa^{i,j}, \kappa^{i,j,k}$ and so on, in terms of the corresponding moments.

Now, we consider the relations between moments and cumulants. Writing $M_X(\xi) = \exp\{K_X(\xi)\}$ and expanding it, we have

$$\begin{aligned} 1 + \xi_i\kappa_i &+ \xi_i\xi_j(\kappa^{i,j}/2! + \kappa^i\kappa^j/2!) \\ &+ \xi_i\xi_j\xi_k\kappa^{i,j,k}/3! + \xi_i\xi_j\xi_k\xi_l\kappa^{i,j,k,l}/4! + \ldots \\ &+ \xi_i\xi_j\xi_k\kappa^i\kappa^{j,k}/2! + \xi_i\xi_j\xi_k\xi_l\{\kappa^i\kappa^{j,k,l}/6 + \kappa^{i,j}\kappa^{k,l}/8\} + \ldots \\ &+ \xi_i\xi_j\xi_k\kappa^i\kappa^j\kappa^k/3! + \xi_i\xi_j\xi_k\xi_l\kappa^i\kappa^j\kappa^{k,l}/4 + \ldots \\ &\quad \xi_i\xi_j\xi_k\xi_l\kappa^i\kappa^j\kappa^k\kappa^l/4! + \ldots \quad . \end{aligned}$$

By comparing the corresponding terms of ξ, we obtain the following expressions for moments in terms of cumulants

$$\begin{aligned} \kappa^{ij} &= \kappa^{i,j} + \kappa^i\kappa^j, \\ \kappa^{ijk} &= \kappa^{i,j,k} + (\kappa^i\kappa^{j,k} + \kappa^j\kappa^{i,k} + \kappa^k\kappa^{i,j}) + \kappa^i\kappa^j\kappa^k \end{aligned}$$

$$= \kappa^{i,j,k} + \kappa^i \kappa^{j,k}[3] + \kappa^i \kappa^j \kappa^k,$$

$$\kappa^{ijkl} = \kappa^{i,j,kl} + \kappa^i \kappa^{j,k,l}[4] + \kappa^{i,j} \kappa^{k,l}[3] + \kappa^i \kappa^j \kappa^{k,l} + \kappa^i \kappa^j \kappa^k \kappa^l.$$

Similarly, the reverse formula giving cumulants in terms of moments can be obtained by expanding $\log M_X(\xi)$ and combining the corresponding terms

$$\kappa^{i,j} = \kappa^{ij} - \kappa^i \kappa^j,$$

$$\kappa^{i,j,k} = \kappa^{ijk} - \kappa^i \kappa^{jk}[3] + 2\kappa^i \kappa^j \kappa^k,$$

$$\kappa^{i,j,k,l} = \kappa^{ijkl} - \kappa^i \kappa^{jkl}[4] - \kappa^{ij} \kappa^{kl}[3] + 2\kappa^i \kappa^j \kappa^{kl}[6] - 6\kappa^i \kappa^j \kappa^k \kappa^l.$$

2.2.2 Edgeworth series

First we discuss the concept of Hermite tensors (McCullagh, 1984, 1987). The components of the Hermite tensors $h_r(x,\lambda), h_{rs}(x,\lambda), h_{rst}(x,\lambda), \ldots$ are polynomials in x whose degree is the same as the order of the tensor. Ordinary Hermite tensors, $h_r(x,\delta), h_{rs}(x,\delta), \ldots$ are obtained by successive partial differentiation of the standard p-variate normal density

$$\phi(x,\delta) = (2\pi)^{-\frac{1}{2}p} \exp(-\frac{1}{2} x^r x^s \delta_{rs}).$$

Let

$$D_1 = \{d_r\}, D_2 = \{d_r d_s\}, d_3 = \{d_r d_s d_t\}, \ldots,$$

$$H_1(x,\delta) = \{h_r(x,\delta)\}, H_2(x,\delta) = \{h_{rs}(x,\delta)\}, \ldots,$$

where the differential operator $d_r = \partial/\partial x^r$ is a covariant tensor. Using the above notations, we have that $D_j \phi(x,\delta) = (-1)^j H_j(x,\delta)\phi(x)$. The first four ordinary Hermite tensors are

$$h_r(x,\delta) = \delta_{rs} x^s = x'_r, \qquad h_{rs}(x,\delta) = x'_r x'_s - \delta_{rs},$$

$$h_{rst}(x,\delta) = x'_r x'_s x'_t - x'_r \delta_{st}[3], \quad h_{rstu}(x,\delta) = x'_r x'_s x'_t x'_u - x'_r x' \delta_{tu}[6] + \delta_{rs}\delta_{tu}[3].$$

Generalized Hermite tensors, formed by differentiation of $\phi(x,\lambda)$, are obtained by replacing δ_{rs} with λ_{rs} in $\phi(x,\delta)$. Thus,

$$h_r(x,\lambda) = \lambda_{rs} x^s = x'_r, \quad h_{rs}(x,\lambda) = x'_r x'_s - \lambda_{rs} = \lambda_{rt}\lambda_{su}(x^t x^u - \kappa^{t,u})$$

and so on, where $\kappa^{r,s}$ is the inverse of λ_{rs}.

Next the application of Hermite tensors to Edgeworth series is given. Let X be a random variable with zero mean and cumulants $\kappa^{r,s}$, $\kappa^{r,s,t}/\sqrt{n}$, $\kappa^{r,s,t,u}/n, \ldots$, where n is a sample size. The Edgeworth expansion with two correction terms may be written as

$$\begin{aligned}\phi(x,\lambda)[1 &+ \kappa^{r,s,t}h_{rst}(x,\lambda)/(6\sqrt{n}) \\ &+ \{3\kappa^{r,s,t,u}h_{rstu}(x,\lambda) + \kappa^{r,s,t}\kappa^{u,v,w}h_{r\ldots w}(x,\lambda)\}/(72n)] + O(n^{-3/2}).\end{aligned}$$

Now we consider the conditional cumulants. Suppose that $X_{(1)}, X_{(2)}$ is a partition of X into components of dimensions $p-q$ and q respectively. For convenience, let the indices $r, s, t, \ldots$ range from 1 to p while $i, j, k, \ldots$ range only over the components of $X_{(2)}$. To compute the conditional cumulants of $X_{(1)}$ a linear transformation to uncorrelated variables Y is introduced. That is, $Y^r = X^r - \beta_j^r X^j, Y_j = X_j$. From McCullagh (1984), the cumulants of Y are

$$\begin{aligned}
\kappa_1^r &= 0, \kappa_1^{r,s} = \kappa^{r,s} - \beta_i^r\beta_j^s\kappa^{i,j}, \kappa_1^{i,r} = 0, \\
\kappa_1^{r,s,t} &= \kappa^{r,s,t} - \beta_i^r\kappa^{i,s,t}[3] + \beta_i^r\beta_j^s\kappa^{i,j,t}[3] - \beta_i^r\beta_j^s\beta_k^t\kappa^{i,j,k}, \\
\kappa_1^{i,r,s} &= \kappa^{i,r,s} - \beta_j^r\kappa^{i,j,s}[2] + \beta_j^r\beta_k^s\kappa^{i,j,k}, \kappa_1^{i,j,r} = \kappa^{i,j,r} - \beta_k^r\kappa^{i,j,k}, \\
\kappa_1^{r,s,t,u} &= \kappa^{r,s,t,u} - \beta_i^r\kappa^{i,s,t,u}[4] + \beta_i^r\beta_j^s\kappa^{i,j,t,u}[6] - \beta_i^r\beta_j^s\beta_k^t\kappa^{i,j,k,u} + \beta_i^r\beta_j^s\beta_k^t\beta_l^u\kappa^{i,j,k,l}, \\
\kappa_1^{i,r,s,t} &= \kappa^{i,r,s,t} - \beta_j^r\kappa^{i,j,s,t}[3] + \beta_j^r\beta_k^s\kappa^{i,j,k,t}[3] - \beta_j^r\beta_k^s\beta_l^t\kappa^{i,j,k,l}, \\
\kappa_1^{i,j,r,s} &= \kappa^{i,j,r,s} - \beta_k^r\kappa^{i,j,k,s}[2] + \beta_k^r\beta_l^s\kappa^{i,j,k,l}, \kappa_1^{i,j,k,r} = \kappa^{i,j,k,r} - \beta_l^r\kappa^{i,j,k,l}.
\end{aligned}$$

Therefore, to the third order of approximation, the first four conditional cumulants of $X_{(1)}$ given $X_{(2)}$ are

$$\begin{aligned}
E(X^r|X_{(2)}) &= \kappa^{r,i}h_i + \kappa_1^{r,i,j}h_{ij}/(2\sqrt{n}) + \{\kappa_1^{r,i,j}\kappa^{k,l,m}(h_{i\ldots m} - h_{ij}h_{klm}) \\
&+ 2\kappa_1^{r,i,j,k}h_{ijk}\}/(12n), \\
\mathrm{cov}(X^i, X^j|X_{(2)}) &= \kappa_1^{r,s} + \kappa_1^{r,s,i}h_i/\sqrt{n} + \{6\kappa_1^{r,s,i,j}h_{ij}
\end{aligned}$$

$$\begin{aligned} &+ \; 2\kappa_1^{r,s,i}\kappa^{j,k,l}(h_{ijkl} - h_i h_{jkl}) \\ &+ \; 3\kappa_1^{r,i,j}\kappa_1^{s,k,l}(h_{ijkl} - h_{ij}h_{kl})\}/(12n), \end{aligned}$$

$$\begin{aligned} \mathrm{cum}(X^r, X^s, X^t | X_{(2)}) &= \kappa_1^{r,s,t}/\sqrt{n} + \{2\kappa_1^{r,s,t,i}h_i \\ &+ \; \kappa_1^{r,s,i}\kappa_1^{t,j,k}(h_{ijk} - h_i h_{jk})[3]\}/(2n) \end{aligned}$$

and

$$\mathrm{cum}(X^r, X^s, X^t, X^u | X_{(2)}) = \kappa^{r,s,t,u}/n + \kappa_1^{r,s,i}\kappa_1^{t,u,j}(h_{ij} - h_i h_j)[3]/n.$$

Further if X is the standardized sum of n independent random variables, the cumulants can be written as $0, \kappa^{\alpha,\beta}, n^{-1/2}\kappa\alpha, \beta, \gamma, n^{-1}\kappa^{\alpha,\beta,\gamma,\delta}$ and so on. Suppose for simplicity that the components $X_{(1)}$ and $X_{(2)}$ are uncorrelated so that $\kappa^{r,i} = 0$. Then the asymptotic expansion of conditional cumulants of $X_{(1)}$ given $X_{(2)}$ have the following form up to terms of order $O(n^{-1})$.

$$\begin{aligned} E(X^r | X_{(2)}) &= \kappa^r + n^{-1/2}\kappa^{r,i,j}h_{ij}/2 + n^{-1}\{\kappa^{r,i,j}\kappa^{k,l,m}h_{ij,klm}/(3!2!) \\ &+ \; \kappa^{r,i,j,k}h_{ijk}/3!\}, \\ \mathrm{cov}(X^r, X^s | X_{(2)}) &= \kappa^{r,s} + n^{-1/2}\kappa^{r,s,i}h_i + n^{-1}\{\kappa^{r,s,i,j}h_{ij}/2! \\ &+ \; \kappa^{r,s,i}\kappa^{j,k,l}h_{i,jkl}/3! + \kappa^{r,i,j}\kappa^{s,k,l}h_{ij,kl}/(2!2!)\}, \\ \mathrm{cum}(X^r, X^s, X^t | X_{(2)}) &= n^{-1/2}\kappa^{r,s,t} + n^{-1}\{\kappa^{r,s,t,i}h_i + \kappa^{r,s,i}\kappa^{t,j,k}[3]h_{i,jk}/2!\} \end{aligned}$$

and

$$\mathrm{cum}(X^r, X^s, X^t, X^u | X_{(2)}) = n^{-1}\{\kappa^{r,s,t,u} + \kappa^{r,s,i}\kappa^{t,u,j}[3]h_{i,j}\}.$$

2.3 Modified Pearson statistic in generalized linear models

Let $Y_i, i = 1, ..., n,$ denote independent random variables from an exponential family distribution with mean μ_i and variance V_i. Also assume a generalized linear regression model with the inverse link function

$$h^{-1}(\mu) = \eta = X\beta, \tag{2.3.1}$$

where X is an $n \times p$ model matrix and β is a vector of p regression parameters. Maximum likelihood estimates of the regression parameters $\beta_1, ..., \beta_p$ are obtained as solutions of the p quasi-likelihood estimating equations $g_r(\hat{\beta}) = 0, r = 1, ..., p$, where

$$g_r(\beta) = \sum_{i=1}^{n} \frac{y_i - \mu_i}{V_i} \frac{\partial \mu_i}{\partial \beta_r}. \tag{2.3.2}$$

By embedding the model in a wider family with variance ϕV_i, goodness of fit is tested by assessing departure from the value $\phi = 1$. As we point out in the Introduction, in order to derive the approximations to the first three moments of the unconditional and conditional distributions of the Pearson statistic for non-canonical exponential family regresion models, Farrington (1996) proposed a supplementary unbiased estimation function

$$g_q(\beta, \phi) = \sum_{i=1}^{n} a_i(y_i - \mu_i) + \sum_{i=1}^{n} \left\{ \frac{(y_i - \mu_i)^2}{V_i} - \phi \right\}, \tag{2.3.3}$$

where $q = p+1$ and the a_i are functions of μ_i depending on β but not on ϕ, defining a family of first-order correction terms to the Pearson statistic. Then the dispersion parameter is estimated by using the equation $g_q(\hat{\beta}, \hat{\phi}) = 0$. From the above equation, the modified Pearson statistic can be obtained as follows:

$$X_*^2 = n\hat{\phi} = \sum_{i=1}^{n} \hat{a}_i(y_i - \hat{\mu}_i) + \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{V}_i}.$$

Note that the choice $a_i = 0$ yields the usual Pearson statistic X^2.

Farrington (1996) derived approximations to the first three moments of the unconditional and the conditional distribution of the modified Pearson statistic. Furthermore, by choosing $a_i = -V_i'/V_i$, the modified Pearson statistic depends only weakly on β, given $\hat{\beta}$. Also this choice of a_i minimizes the unconditional variance of the modified Pearson statistic and also greatly simplifies the expressions of the moments of X_*^2.

Theorem 2.3.1 *Let* $h_i' = \frac{dh(\eta_i)}{d\eta_i}, h_i'' = \frac{d^2h(\eta_i)}{d\eta_i^2}, V_i' = \frac{dV_i(\mu_i)}{d\mu_i}, V_i'' = \frac{d^2V_i(\mu_i)}{d\mu_i^2}, W = \text{diag}(h'^2/V_i), Q = (Q_{ij}) = X(X^TWX)^{-1}X^T$. *Also let* $\kappa_{mi}, m = 3,4,5,6,$ *be the m-order central cumulants of* y_i *(see Section 2.2). Further, let* $a = (a_i), d = (d_i) = (a_i + V_i'/V_i), V = \text{diag}(V_i), H = \text{diag}(h_i')$ *and* $\gamma = V^{-1}HQHd - a$. *Then,*

$$
\begin{aligned}
E(X_*^2) &= n - p + \frac{1}{2}\sum_{i,j=1}^{n}\left(a_i + \frac{V_i'}{V_i}\right)\frac{h_j''}{V_j}h_i'h_j'Q_{ij}Q_{jj} \\
&\quad -\frac{1}{2}\sum_{i=1}^{n}\left\{\frac{V_i''}{V_i}h_i'^2 + \left(a_i + \frac{V_i'}{V_i}\right)h_i''\right\}Q_{ii} + O(n^{-1}), \\
\text{var}(X_*^2) &= n(\bar\rho_4 - \bar\rho_3^2 + 2) + \sum_{i=1}^{n} V_i\left(a_i + \frac{V_i'}{V_i}\right)^2 \\
&\quad - \sum_{i,j=1}^{n}\left(a_i + \frac{V_i'}{V_i}\right)\left(a_j + \frac{V_j'}{V_j}\right)h_i'h_j'Q_{ij} + O(1)
\end{aligned}
$$

and

$$
\begin{aligned}
\kappa_3(X_*^2) &= n(\bar\rho_6 + 12\bar\rho_4 + 10\bar\rho_3^2 + 8) \\
&\quad - \left(3\sum_{i=1}^{n}\gamma_i\frac{\kappa_{5i}}{V_i^2} - 3\sum_{i=1}^{n}\gamma_i^2\frac{\kappa_{4i}}{V_i} + 18\sum_{i=1}^{n}\gamma_i\frac{\kappa_{3i}}{V_i} + \sum_{i=1}^{n}\gamma_i^3\kappa_{3i}\right) + O(1),
\end{aligned}
$$

where $\bar\rho_4 = n^{-1}\sum \kappa_{4i}/V_i^2, \bar\rho_3^2 = n^{-1}\sum\kappa_{3i}^2/V_i^3$ *and* $\bar\rho_6 = n^{-1}\sum\kappa_{6i}/V_i^3$. *In particular, when* $a_i = -V_i'/V_i, \gamma_i = \kappa_{3i}/V_i^2$ *so*

$$
\begin{aligned}
E(X_*^2) &= n - p - \frac{1}{2}\sum_{i=1}^{n}\frac{V_i''}{V_i}h_i'^2 + O(n^{-1}), \\
\text{var}(X_*^2) &= n(\bar\rho_4 - \bar\rho_3^2 + 2) + O(1)
\end{aligned}
$$

and

$$
\kappa_3(X_*^2) = n(\bar\rho_6 - 3\bar\rho_{35} + 3\bar\rho_{34}^2 + 12\bar\rho_4 - 8\bar\rho_3^2 - \bar\rho_3^4 + 8) + O(1),
$$

where $\bar\rho_{35} = n^{-1}\sum\kappa_{3i}\kappa_{5i}/V_i^4, \bar\rho_{34}^2 = n^{-1}\sum\kappa_{3i}^2\kappa_{4i}/V_i^5$ *and* $\bar\rho_3^4 = n^{-1}\sum\kappa_{3i}^4/V_i^6$.

Theorem 2.3.2 *By using the same notations as in Theorem 2.3.1, the first three conditional moments are*

$$
E(X_*^2|\hat\beta) = n - p - \frac{1}{2}\sum_{i=1}^{n}\left\{\frac{\hat V_i''}{\hat V_i}\hat h_i'^2 + \left(\hat a_i + \frac{\hat V_i'}{\hat V_i}\right)\left(2\hat h_i'' - \frac{\hat V_i'}{\hat V_i}\hat h_i'^2\right)\right\}\hat Q_{ii}
$$

$$
\begin{aligned}
& + \frac{1}{2}\sum_{i,j=1}^{n}\left(\hat{a}_i + \frac{\hat{V}_i'}{\hat{V}_i}\right)\frac{1}{\hat{V}_j}\left(2\hat{h}_j'' - \frac{\hat{V}_j'}{\hat{V}_j}\hat{h}_j'^2\right)\hat{h}_i'\hat{h}_j'\hat{Q}_{ij}\hat{Q}_{jj} + O(n^{-1/2}), \\
\operatorname{var}(X_*^2|\hat{\beta}) &= n(\hat{\rho}_4 - \hat{\rho}_3^2 + 2) + \sum_{i=1}^{n}\hat{V}_i\left(\hat{a}_i + \frac{\hat{V}_i'}{\hat{V}_i}\right)^2 \\
& - \sum_{i,j=1}^{n}\left(\hat{a}_i + \frac{\hat{V}_i'}{\hat{V}_i}\right)\left(\hat{a}_j + \frac{\hat{V}_j'}{\hat{V}_j}\right)\hat{h}_i'\hat{h}_j'\hat{Q}_{ij} + O(n^{-1/2})
\end{aligned}
$$

and

$$
\begin{aligned}
\kappa_3(X_*^2|\hat{\beta}) &= n(\hat{\rho}_6 + 12\hat{\rho}_4 + 10\hat{\rho}_3^2 + 8) \\
& - \left(3\sum_{i=1}^{n}\hat{\gamma}_i\frac{\hat{\kappa}_{5i}}{\hat{V}_i^2} - 3\sum_{i=1}^{n}\hat{\gamma}_i^2\frac{\hat{\kappa}_{4i}}{\hat{V}_i} + 18\sum_{i=1}^{n}\hat{\gamma}_i\frac{\hat{\kappa}_{3i}}{\hat{V}_i} + \sum_{i=1}^{n}\hat{\gamma}_i^3\hat{\kappa}_{3i}\right) + O(n^{-1/2}),
\end{aligned}
$$

where a circumflex above the variates denotes evaluation at $\beta = \hat{\beta}$. *In particular, when* $a_i = -V_i'/V_i$ *the approximate conditional moments of* X_*^2 *are*

$$
\begin{aligned}
E(X_*^2|\hat{\beta}) &= n - p - \frac{1}{2}\sum_{i=1}^{n}\frac{V_i''}{V_i}h_i'^2, \\
\operatorname{var}(X_*^2|\hat{\beta}) &= n(\bar{\rho}_4 - \bar{\rho}_3^2 + 2), \\
\kappa_3(X_*^2|\hat{\beta}) &= n(\bar{\rho}_6 - 3\bar{\rho}_{35} + 3\bar{\rho}_{34}^2 + 12\bar{\rho}_4 - 8\bar{\rho}_3^2 - \bar{\rho}_3^4 + 8).
\end{aligned}
$$

Farrington (1996) points out that for generalized linear models both the Pearson statistic X^2 and the modified Pearson statistic X_*^2 are asymptotically independent of $\hat{\beta}$. In addition, given the maximum likelihood estimate $\hat{\beta}$, X_*^2 depends only weakly on β. This supports the use of the conditional distribution of X_*^2 given $\hat{\beta}$ for assessing the goodness of fit even when the link function is not canonical. The approximate standardized quantity

$$Z = \{X_*^2 - E(X_*^2|\hat{\beta})\}/\operatorname{var}(X_*^2|\hat{\beta})^{1/2}$$

may be used to calculate upper tail probabilities $P(Z \geq z|\hat{\beta})$, either directly by reference to the standard normal distribution, or, as suggested by McCullagh (1986), by using the Edgeworth approximation

$$P(Z \geq z) \approx 1 - \Phi(z) + (z^2 - 1)\rho_3\phi(z)/6$$

where ρ_3 is the approximate standardized conditional third moment.

The simulation results and examples in Farrington (1996) show that the standardized modified Pearson statistic performs very well in terms of empirical level.

Chapter 3

Goodness of fit of generalized linear models to sparse data

3.1 Introduction

Traditionally, goodness of fit in contingency tables is tested using either the Pearson chi-square statistic or the likelihood ratio chi-square statistic. Asymptotic properties of these statistics are studied based on the assumption that the expected cell frequencies become large. Recent interest, however, has been to obtain modifications of these statistics for sparse contingency tables using higher-order moment approximations (see, for example, Koehler and Larntz (1980) and Koehler (1986)). These authors derive modifications using moments of the unconditional distribution of the Pearson chi-square statistic and the likelihood ratio statistic.

McCullagh (1986) argues that it is the conditional distribution of the statistic and not its marginal distribution that is relevant for assessing goodness of fit. He obtains conditional distributions of the Pearson chi-square statistic and the likelihood ratio chi-square statistic for discrete data for the case where the data are extensive but sparse. McCullagh (1985) obtains approximations to the first three moments of the unconditional and the conditional distributions of the Pearson chi-square statistic for canonical exponential family regression models. Farrington (1996) extends the results of McCullagh (1985) to models with non-canonical links using an estimating-equations approach following Moore (1986).

The motivation of this chapter comes from a question by Farrington (1996, p360) as to whether similar methods could be applied to the deviance statistic to extend the results of McCullagh (1986). McCullagh and Nelder (1989, p36) also indicate that further work on the asymptotic distribution of the deviance statistic remains to be done. We derive approximations to the first three moments of the unconditional and the conditional distributions of the deviance statistic for assessing goodness of fit of generalized linear models with non-canonical links. As in McCullagh (1986) and Farrington (1996) we consider the asymptotic limit in which the data are extensive but sparse, and also as in Farrington (1996) we consider a supplementary estimating equation for the dispersion parameter. The results derived in this chapter are extensions of the results derived by McCullagh (1986) for assessing goodness of fit of the deviance statistic for discrete data. Applications of the method for assessing goodness of fit of the binomial and the Poisson models to sparse data are shown. A small-scale simulation study of the performance of the modified Pearson statistic of Farrington (1996) and the modified deviance statistic developed in this chapter is reported. Three sets of data with different degrees of sparseness and using different link functions are also analyzed.

3.2 Estimating equations and goodness of fit

As in Section 2.3, let $Y_i, i = 1, ..., n$, denote independent random variables from an exponential family distribution with mean μ_i and variance V_i. Consider a generalized linear regression model with the inverse link function

$$h^{-1}(\mu) = \eta = X\beta, \tag{3.2.1}$$

where X is an $n \times p$ model matrix and β is a vector of p regression parameters. Maximum likelihood estimates of the regression parameters $\beta_1, ..., \beta_p$ are obtained as solutions of the p

quasi-likelihood estimating equations $g_r(\hat{\beta}) = 0, r = 1, ..., p$, where

$$g_r(\beta) = \sum_{i=1}^{n} \frac{y_i - \mu_i}{V_i} \frac{\partial \mu_i}{\partial \beta_r}.$$

The usual deviance statistic is $D = \sum_{i=1}^{n} \hat{d}_i = \sum_{i=1}^{n} d_i(y_i, \hat{\mu}_i)$, where $d_i(y_i, \mu_i) = 2\int_{\mu_i}^{y_i}(y_i - t)/V_i(t)dt$. Proceeding now by analogy with Farrington (1996), we define a family of modified deviance statistics

$$D^* = \sum_{i=1}^{n}(y_i - \hat{\mu}_i)\hat{a}_i + \sum_{i=1}^{n} \hat{d}_i/\hat{\kappa}_1^{(i)},$$

in which $\hat{\kappa}_1^{(i)} = \kappa_1^{(i)}(\hat{\mu}_i)$ estimates $E(d_i) = \kappa_1^{(i)}(\mu_i)$, and $\hat{a}_i = a_i(\hat{\mu}_i)$ where $\{a_i(\cdot) : i = 1, ..., n\}$ is a set of functions to be specified. The modified deviance D^* may alternatively be expressed in terms of the solution of a supplementary estimating equation $g_q(\hat{\beta}, \hat{\phi}) = 0$, where

$$g_q(\beta, \phi) = \sum_{i=1}^{n}(y_i - \mu_i)a_i(\mu_i) + \sum_{i=1}^{n}\left\{\frac{d_i(y_i, \mu_i)}{\kappa_1^{(i)}(\mu_i)} - \phi\right\} \tag{3.2.2}$$

and $q = p+1$. In (3.2.2), ϕ represents a notional dispersion parameter $E\{d_i(y_i, \mu_i)\}/\kappa_1^{(i)}(\mu_i)$ whose value is 1 under the model; and $D^* = n\hat{\phi}$. Note that the notional dispersion parameter ϕ can be written as $\phi = 1 + c\psi$ where c is a known constant. Therefore $\phi = 1$ is equivalent to $\psi = 0$.

With the particular choice $a_i(\mu_i) \equiv 0$ for all i, D^* reduces to $\sum_{i=1}^{n} \hat{d}_i/\hat{\kappa}_1^{(i)}$, which we denote by D_1^*. Replacement of $\kappa_1^{(i)}$ by 1 in the definition of D_1^* would yield the usual deviance statistic D.

3.3 Moments

Unconditional and conditional moments of the modified deviance statistic D^* are given in Theorems 3.3.1 and 3.3.2. The proofs which follow similar steps as Farrington (1996) and which use the conditional moments formulae of McCullagh (1984, 1987) are given in Appendix A.

Theorem 3.3.1 *Let* $h'_i = \frac{dh(\eta_i)}{d\eta_i}$, $W = \text{diag}(h_i'^2/V_i)$, $Q = (Q_{ij}) = X(X^TWX)^{-1}X^T$, $\gamma_i = \sum_{j=1}^n V_i^{-1}h'_iQ_{ij}h'_ja_j$, $\kappa_2^{(i)} = E(d_i-\kappa_1^{(i)})^2$, $\kappa_3^{(i)} = E(d_i-\kappa_1^{(i)})^3$, $\kappa_{11}^{(i)} = E\{(d_i-\kappa_1^{(i)})(y_i-\mu_i)\}$, $\kappa_{12}^{(i)} = E\{(d_i-\kappa_1^{(i)})(y_i-\mu_i)^2\}$, $\kappa_{21}^{(i)} = E\{(d_i-\kappa_1^{(i)})^2(y_i-\mu_i)\}$ *and* $\kappa_{03}^{(i)} = E(y_i-\mu_i)^3$. *Then,*

$$
\begin{aligned}
E(D^*) &= nE\hat{\phi} = n - \sum_{i=1}^n \frac{1}{V_i\kappa_1^{(i)}}h_i'^2Q_{ii} + \sum_{i=1}^n \frac{1}{2}(\gamma_i - a_i)h_i''Q_{ii} + O(n^{-1}), \\
\text{var}(D^*) &= n^2\,\text{var}(\hat{\phi}) = \sum_{i=1}^n \left\{ a_i^2V_i - a_i\gamma_iV_i + 2\frac{\kappa_{11}^{(i)}}{\kappa_1^{(i)}}(a_i-\gamma_i) + \frac{\kappa_2^{(i)}}{(\kappa_1^{(i)})^2} \right\} + O(1),
\end{aligned}
$$

and

$$
\begin{aligned}
\kappa_3(D^*) &= n^3\kappa_3(\hat{\phi}) \\
&= \sum_{i=1}^n \left\{ (a_i-\gamma_i)^3\kappa_{03}^{(i)} + 3(a_i-\gamma_i)^2\frac{\kappa_{12}^{(i)}}{\kappa_1^{(i)}} + 3(a_i-\gamma_i)\frac{\kappa_{21}^{(i)}}{(\kappa_1^{(i)})^2} + \frac{\kappa_3^{(i)}}{(\kappa_1^{(i)})^3} \right\} + O(1).
\end{aligned}
$$

Theorem 3.3.2 *Let* $\mathbf{1} = (1, ..., 1)$, $\kappa_{11} = (\frac{\kappa_{11}^{(1)}h'_1}{V_1\kappa_1^{(1)}}, ..., \frac{\kappa_{11}^{(n)}h'_n}{V_1\kappa_1^{(n)}})^T$, $\kappa_{11}^1 = \text{diag}(\sum_{j=1}^n \frac{\kappa_{11}^{(j)}h'_jQ_{ji}}{\kappa_1^{(j)}V_j})$, $\kappa_{12} = \text{diag}(\frac{\kappa_{12}^{(i)}h_i'^2}{\kappa_1^{(i)}V_i^2})$, $\kappa_{12}^1 = \text{diag}(\frac{\kappa_{12}^{(i)}h'_i}{\kappa_1^{(i)}V_i})$, $\kappa_{21} = \text{diag}(\frac{\kappa_{21}^{(i)}h'_i}{(\kappa_1^{(i)})^2V_i})$, $\kappa_{03}^1 = \text{diag}(\frac{\kappa_{03}^{(i)}h'_i}{V_i})$, $\kappa_{03}^2 = \text{diag}(\frac{\kappa_{03}^{(i)}h_i'^2}{V_i^2})$, $\kappa_{03}^3 = \text{diag}(\frac{\kappa_{03}^{(i)}h_i'^3}{V_i^3})$, $(a-\gamma)^m = \text{diag}\{(a_i-\gamma_i)^m\}$. *Then,*

$$
\begin{aligned}
E(D^*|\hat{\beta}) &= nE(\hat{\phi}|\hat{\beta}) = \hat{E}(D^*) \\
&\quad -\frac{1}{2}\mathbf{1}^TX^T\{\hat{\kappa_{12}} + (\hat{a}-\hat{\gamma})\hat{\kappa}_{03}^2 - \hat{\kappa}_{11}^1\hat{\kappa}_{03}^3\}X(X^T\hat{W}X)^{-1}\mathbf{1} + O(n^{-\frac{1}{2}}), \\
\text{var}(D^*|\hat{\beta}) &= n^2\,\text{var}(\hat{\phi}|\hat{\beta}) \\
&= \hat{\text{var}}(D^*) - \hat{\kappa}_{11}^TX(X^T\hat{W}X)^{-1}X^T\hat{\kappa}_{11} + O(n^{-\frac{1}{2}}), \\
\hat{\kappa}_3(D^*|\hat{\beta}) &= n^3\kappa_3(\hat{\phi}|\hat{\beta}) \\
&= \hat{\kappa}_3(D^*) - 3\hat{\kappa}_{11}^TX(X^T\hat{W}X)^{-1}X^T\{\hat{\kappa}_{21} + \hat{\kappa}_{03}^1(\hat{a}-\hat{\gamma})^2 + 2\hat{\kappa}_{12}^1(\hat{a}-\hat{\gamma})\}\mathbf{1} \\
&\quad +3\hat{\kappa}_{11}^TX(X^T\hat{W}X)^{-1}X^T\{\hat{\kappa}_{12} + \hat{\kappa}_{03}^2(\hat{a}-\hat{\gamma})\}X(X^T\hat{W}X)^{-1}X^T\hat{\kappa}_{11} \\
&\quad -\mathbf{1}^T\hat{\kappa}_{03}^3\hat{\kappa}_{11}^1\hat{\kappa}_{11}^1\hat{\kappa}_{11}^1\mathbf{1} + O(n^{-\frac{1}{2}}).
\end{aligned}
$$

For $a_i = 0$, *we have that*

$$
E(D_1^*|\hat{\beta}) = nE(\hat{\phi}|\hat{\beta})
$$

$$
\begin{aligned}
&= \hat{E}(D_1^*) - \frac{1}{2}\mathbf{1}^T X^T(\hat{\kappa}_{12} - \hat{\kappa}_{11}^1\hat{\kappa}_{03}^3)X(X^T\hat{W}X)^{-1}\mathbf{1} + O(n^{-\frac{1}{2}}), \\
\mathrm{var}(D_1^*|\hat{\beta}) &= n^2\,\mathrm{var}(\hat{\phi}|\hat{\beta}) \\
&= \hat{\mathrm{var}}(D_1^*) - \hat{\kappa}_{11}^T X(X^T\hat{W}X)^{-1}X^T\hat{\kappa}_{11} + O(n^{-\frac{1}{2}}), \\
\kappa_3(D_1^*|\hat{\beta}) &= n^3\kappa_3(\hat{\phi}|\hat{\beta}) \\
&= \hat{\kappa}_3(D_1^*) - 3\hat{\kappa}_{11}^T X(X^T\hat{W}X)^{-1}X^T\hat{\kappa}_{21}\mathbf{1} \\
&\quad +3\hat{\kappa}_{11}^T X(X^T\hat{W}X)^{-1}X^T\hat{\kappa}_{12}X(X^T\hat{W}X)^{-1}X^T\hat{\kappa}_{11} \\
&\quad -\mathbf{1}^T\hat{\kappa}_{03}^3\hat{\kappa}_{11}^1\hat{\kappa}_{11}^1\hat{\kappa}_{11}^1\mathbf{1} + O(n^{-\frac{1}{2}}).
\end{aligned}
$$

Remark: Note that in D^*, if we replace a_i by 0 and $E(d_i)$ by 1, we obtain D. Then, following the derivation of Theorem 3.3.2 we obtain

$$
\begin{aligned}
E(D|\hat{\beta}) &= \hat{\kappa}_1^{(\cdot)} - \frac{1}{2}\mathbf{1}^T X^T\hat{\Sigma}X(X^T\hat{W}X)^{-1}\mathbf{1} + O(n^{-\frac{1}{2}}), \\
\mathrm{var}(D|\hat{\beta}) &= \hat{\kappa}_2^{(\cdot)} - \hat{\kappa}_{11}^T X(X^T\hat{W}X)^{-1}X^T\hat{\kappa}_{11} + O(n^{-\frac{1}{2}}), \\
\kappa_3(D|\hat{\beta}) &= \hat{\kappa}_3^{(\cdot)} - 3\hat{\kappa}_{11}^T X(X^T\hat{W}X)^{-1}X^T\hat{\kappa}_{21}\mathbf{1} \\
&\quad +3\hat{\kappa}_{11}^T X(X^T\hat{W}X)^{-1}X^T\hat{\kappa}_{12}X(X^T\hat{W}X)^{-1}X^T\hat{\kappa}_{11} \\
&\quad -\mathbf{1}^T\hat{\kappa}_{03}^3\hat{\kappa}_{11}^1\hat{\kappa}_{11}^1\hat{\kappa}_{11}^1\mathbf{1} + O(n^{-\frac{1}{2}}),
\end{aligned}
$$

where $\Sigma = \mathrm{diag}\{\kappa_{12}^{(i)}h_i'^2/V_i^2 - (\sum_j \kappa_{11}^{(j)}h_j'Q_{ji}/V_j)\kappa_{03}^{(i)}h_i'^3/V_i^3\}$, $\kappa_{11} = (\kappa_{11}^{(1)}h_1'/V_1, ..., \kappa_{11}^{(n)}h_n'/V_n)^T$, $\kappa_{21} = \mathrm{diag}(\kappa_{21}^{(i)}h_i'/V_i)$, $\kappa_{12} = \mathrm{diag}(\kappa_{12}^{(i)}h_i'^2/V_i^2)$, $\kappa_{11}^1 = \mathrm{diag}(\sum_j \kappa_{11}^{(j)}h_j'Q_{ji}/V_j)$, $\kappa_1^{(\cdot)} = \sum_{i=1}^n \kappa_1^{(i)}$, $\kappa_2^{(\cdot)} = \sum_{i=1}^n \kappa_2^{(i)}$, and $\kappa_3^{(\cdot)} = \sum_{i=1}^n \kappa_3^{(i)}$.

For the log-linear model, $W = \mathrm{diag}(V_i) = V, Q = X(X^TVX)^{-1}X^T, \kappa_{11} = (\kappa_{11}^{(1)}, ..., \kappa_{11}^{(n)})^T$, $\kappa_{21} = \mathrm{diag}(\kappa_{21}^{(i)}), \kappa_{12} = \mathrm{diag}(\kappa_{12}^{(i)}), \Sigma = \mathrm{diag}\{\kappa_{12}^{(i)} - (\sum_j \kappa_{11}^{(j)}Q_{ji})\kappa_{03}^{(i)}\}$ and $\kappa_{11}^1 = \mathrm{diag}(\sum_j \kappa_{11}^{(j)}Q_{ji})$. The expressions for the conditional variance and the conditional third moment of D obtained above agree with those given by McCullagh (1986). However, for the conditional mean of D we find an error in McCullagh (1986) in that $\Sigma = \mathrm{diag}\{k_{12}^{(i)} - (\sum_j \kappa_{11}^{(j)}Q_{ji})\kappa_{03}^{(i)}\}$ instead of $\Sigma = \mathrm{diag}(\kappa_{12}^{(i)} - \kappa_{11}^{(i)}\kappa_{03}^{(i)}/\kappa_{02}^{(i)})$ given by McCullagh (1986). Dr. McCullagh, through personal communication, has acknowledged this.

We now need to choose values for a_i. This, possibly, could be done in a number of ways. We, however, choose $a_i = -\kappa_{11}^{(i)}/(V_i\kappa_1^{(i)})$ by minimizing $\text{var}(D^*)$. This can be seen by writing $\text{var}(D^*)$ as

$$\begin{aligned}\text{var}(D^*) &= \alpha^T\{I - X^*(X^{*T}X^*)^{-1}X^{*T}\}\alpha - \alpha^{1T}\{I - X^*(X^{*T}X^*)^{-1}X^{*T}\}\alpha^1 \\ &\quad + \sum_{i=1}^{n}\frac{\kappa_2^{(i)}}{(\kappa_1^{(i)})^2} + O(1),\end{aligned} \tag{3.3.1}$$

where $\alpha_i = V_i^{\frac{1}{2}}(a_i + \frac{\kappa_{11}^{(i)}}{\kappa_1^{(i)}V_i}), \alpha_i^1 = \frac{\kappa_{11}^{(i)}}{\kappa_1^{(i)}V_i}$ and $X^* = \text{diag}(h_i'V_i^{-\frac{1}{2}})X$. The first term of the right hand side in equation (3.3.1) is the residual sum of squares from the regression of α on X^*. The choice $a_i = -\kappa_{11}^{(i)}/(V_i\kappa_1^{(i)})$ makes this 0. This choice of a_i coincides with the choice of a_i used in the modified Pearson statistic obtained by Farrington (1996). Also, with this choice of a_i, we obtain

$$E(g_r g_q) = 0, \tag{3.3.2}$$

for $r = 1, ..., p$, which we show in what follows. It can be seen that

$$\begin{aligned}E(g_r g_q) &= E\left[\left(\sum_{i=1}^{n}\frac{y_i-\mu_i}{V_i}\frac{\partial\mu_i}{\partial\beta_i}\right)\left\{\sum_{j=1}^{n}a_j(y_j-\mu_j) + \sum_{j=1}^{n}(\frac{d_j}{\kappa_1^{(j)}} - \phi)\right\}\right] \\ &= E\left[\sum_{i=1}^{n}\left\{a_i\frac{(y_i-\mu_i)^2}{V_i}\frac{\partial\mu_i}{\partial\beta_i} + \frac{y_i-\mu_i}{V_i}\frac{\partial\mu_i}{\partial\beta_i}\frac{d_i - \phi\kappa_1^{(i)}}{\kappa_1^{(i)}}\right\}\right] \\ &= E\left[\sum_{i=1}^{n}\left\{a_i\frac{(y_i-\mu_i)^2}{V_i}\frac{\partial\mu_i}{\partial\beta_i} + \frac{(y_i-\mu_i)d_i}{V_i\kappa_1^{(i)}}\frac{\partial\mu_i}{\partial\beta_i}\right\}\right] \\ &= \sum_{i=1}^{n}\left(\frac{-\kappa_{11}^{(i)}}{V_i\kappa_1^{(i)}}\frac{\partial\mu_i}{\partial\beta_i} + \frac{\kappa_{11}^{(i)}}{V_i\kappa_1^{(i)}}\frac{\partial\mu_i}{\partial\beta_i}\right) = 0.\end{aligned}$$

This shows that asymptotically, $g_r, r = 1, ..., p$ and g_q are orthogonal (Cox and Reid, 1987) and hence $\hat{\beta}$ and $\hat{\phi}$ are asymptotically uncorrelated. Results of an empirical study, not reported here, confirm this. For $a_i = 0$, $E(g_r g_q) \neq 0$ for $r = 1, ..., p$. However, using a similar expansion of g_q as Farrington (1996) and Firth (1987), it can be shown that for $a_i = 0$, D_1^* depends weakly on β, given $\hat{\beta}$.

3.4 Application to binomial and Poisson data

3.4.1 Binomial data

Let $Y_i, i = 1, ..., n$, denote independent random variables from a binomial distribution with parameters m_i and π_i,

$$\pi_i = \exp(X_i\beta)/\{1 + \exp(X_i\beta)\},$$

where $X_i\beta = X_{i1}\beta_1 + ... + X_{ip}\beta_p, X_1, ..., X_p$ are p explanatory variables, $\beta_1, ..., \beta_p$ are p regression parameters. Then, $\mu_i = m_i\pi_i, V_i = m_i\pi_i(1-\pi_i), \partial\mu_i/\partial\beta_r = m_i\pi_i(1-\pi_i)X_{ir}$ and hence the estimating equations to obtain $\hat{\beta}_1, ..., \hat{\beta}_p$ are

$$g_r(\beta) = \sum_{i=1}^{n}(y_i - \mu_i)X_{ir} = 0 \quad (r = 1, ..., p).$$

To calculate $D^*, E(D^*|\hat{\beta}), \text{var}(D^*|\hat{\beta})$ and $\kappa_3(D^*|\hat{\beta})$ we need expected values of a number of quantities such as $\kappa_1^{(i)}, k_{11}^{(i)}$ etc (see Theorems 3.3.1 and 3.3.2). Closed form expressions for these quantities do not exist. Approximations using a Taylor expansion of d_i do not work well. So, we need to calculate them directly. However, direct calculations of these expected values do not pose any computational difficulty. For example,

$$\begin{aligned} d_i &= d_i(y_i) = \int_{\mu_i}^{y_i} \frac{2(y_i - t)}{V_i(t)} dt \\ &= 2[y_i \log(y_i/\mu_i) + (m_i - y_i)\log\{(m_i - y_i)/(m_i - \mu_i)\}]. \end{aligned}$$

Then

$$\kappa_1^{(i)} = E(d_i) = \sum_{j=0}^{m_i} d_i(j)p_{ij}$$

and

$$\kappa_{11}^{(i)} = E\{d_i(y_i - \mu_i)\} = \sum_{j=0}^{m_i}(j - \mu_i)d_i(j)p_{ij},$$

where $p_{ij} = C_j^{m_i} p_i^j (1-p_i)^{m_i - j}$. Similarly, $\kappa_2^{(i)}, \kappa_3^{(i)}, \kappa_{12}^{(i)}$ and $\kappa_{21}^{(i)}$ can be calculated.

3.4.2 Poisson data

For Poisson data $Y_i, i = 1, ..., n$,

$$\begin{aligned} \mu_i &= \exp(X_i\beta), \\ g_r(\beta) &= \sum_{i=1}^{n}(y_i - \mu_i)X_{ir} = 0, \qquad r = 1, ..., p, \\ d_i(y_i) &= 2\{y_i \log(y_i/\mu_i) - (y_i - \mu_i)\}. \end{aligned}$$

For large μ, approximate expressions for the moments $\kappa_1^{(i)}, \kappa_2^{(i)}, \kappa_3^{(i)}, \kappa_{11}^{(i)}, \kappa_{12}^{(i)}$ and $\kappa_{21}^{(i)}$ given by McCullagh (1986) can be used. For small μ these approximations will not work well. Again, these moments can easily be obtained by direct summation.

3.5 Simulation

A limited simulation study was conducted to compare the empirical sizes of the modified Pearson statistic X_*^2 (see Farrington, 1996) and the modified deviance statistic D^* derived in the present chapter. In the simulation study we also included the Pearson statistic X^2 and the deviance statistics D and D_1^*. Simulations have been conducted for the binomial model with $p = 2$ and a single continuous covariate chosen to induce very strong regression effects under both logistic and complementary log-log link functions, for sample sizes varying from $n = 10$ to $n = 200$ and binomial denominators $m = 5$ and 10. The results for the logistic and the complementary log-log link functions are very similar. In Table 3.1 we present the results for only the logistic model. The empirical sizes for X^2, D, X_*^2, D_1^* and D^* can be calculated by referring the standardized quantities, such as

$$Z = \{D^* - E(D^*|\hat{\beta})\}/\{\text{var}(D^*|\hat{\beta})\}^{\frac{1}{2}}$$

to the standard normal distribution or by using the procedure based on the Edgeworth approximation suggested by McCullagh (1986). We, however, used the former procedure. The statistics X^2, D, D_1^* all, in general, underestimate the levels, whereas both X_*^2 and D^*

produce empirical levels close to the nominal, although they show some inflated levels for small α and small n.

The simulations were extended to compare power of the statistics X^2, D, X_*^2, D_1^* and D^*. For power comparison we simulated data from the beta-binomial distribution with mean $m\pi$ and variance $m\pi(1-\pi)\{1+(m-1)\psi\}$ where $\pi = \exp(X\beta)/\{1+\exp(X\beta)\}$ and ψ is the over-dispersion parameter. Note that for binomial case $\phi = 1+(m-1)\psi$. Simulations were conducted for $m = 5, 10$, the sample size $n = 10, 20, 50$ and the over-dispersion parameter ψ=0.05, 0.10, 0.15, 0.20 for α =0.01, 0.05, 0.10. The behavior of all the statistics in terms of power is similar for all values of α. The power results are presented in Table 3.2-Table 3.5. The results show that D is more powerful than X^2. The statistics X_*^2 and D^* are more powerful than the other three statistics. This is not surprising as the other three statistics are conservative, whereas the corrected Pearson and the corrected deviance statistics maintain correct levels. However, the modified deviance statistic D^* was found to be uniformly most powerful for the values of ψ considered in our study.

Table 3.1: Empirical sizes of statistics X^2, D, D_1^*, X_*^2 and D^* for nominal levels α = .01, .05, .10, $m = 5, 10$ and $n = 10, 20, 50, 100, 200$ based on 10,000 replications.

nominal level α	binomial parameter m	sample size n	Pearson statistic X^2	Deviance statistic D	Deviance statistic D_1^*	modified Pearson statistic X_*^2	modified deviance statistic D_*
		10	0.0106	0.0123	0.0119	0.0184	0.0238
		20	0.0074	0.0105	0.0104	0.0126	0.0171
	5	50	0.0050	0.0101	0.0103	0.0109	0.0137
		100	0.0049	0.0105	0.0109	0.0119	0.0114
		200	0.0046	0.0100	0.0103	0.0114	0.0117
0.01		10	0.0149	0.0131	0.0133	0.0174	0.0247
		20	0.0100	0.0117	0.0123	0.0127	0.0191
	10	50	0.0076	0.0098	0.0101	0.0107	0.0143
		100	0.0073	0.0106	0.0104	0.0105	0.0124
		200	0.0072	0.0097	0.0106	0.0108	0.0110
		10	0.0290	0.0319	0.0314	0.0453	0.0639
		20	0.0281	0.0389	0.0390	0.0457	0.0561
	5	50	0.0268	0.0448	0.0447	0.0488	0.0484
		100	0.0319	0.0514	0.0514	0.0527	0.0534
		200	0.0303	0.0488	0.0484	0.0493	0.0527
0.05		10	0.0372	0.0331	0.0329	0.0437	0.0614
		20	0.0281	0.0339	0.0338	0.0402	0.0503
	10	50	0.0381	0.0445	0.0445	0.0493	0.0532
		100	0.0409	0.0508	0.0510	0.0520	0.0547
		200	0.0390	0.0472	0.0476	0.0486	0.0484
		10	0.0549	0.0631	0.0610	0.0883	0.1057
		20	0.0583	0.0793	0.0786	0.0912	0.0946
	5	50	0.0632	0.0932	0.0921	0.0944	0.0929
		100	0.0692	0.1007	0.1007	0.1003	0.0988
		200	0.0641	0.0990	0.0990	0.0998	0.1005
0.10		10	0.0653	0.0536	0.0533	0.0806	0.0951
		20	0.0683	0.0686	0.0681	0.0827	0.0859
	10	50	0.0787	0.0908	0.0909	0.0965	0.0961
		100	0.0863	0.1005	0.1005	0.1033	0.1027
		200	0.0815	0.0993	0.1001	0.0974	0.0996

Table 3.2: Empirical power of statistics X^2, D, D_1^*, X_*^2 and D^* for dispersion parameter $\psi = 0.05$, nominal levels $\alpha = .01, .05, .10, m = 5, 10$ and $n = 10, 20, 50$ based on 10,000 replications.

nominal level α	binomial parameter m	sample size n	Pearson statistic X^2	Deviance statistic D	Deviance statistic D_1^*	modified Pearson statistic X_*^2	modified deviance statistic D_*
		10	0.0373	0.0385	0.0382	0.0555	0.0717
	5	20	0.0427	0.0528	0.0520	0.0619	0.0789
		50	0.0669	0.0953	0.0949	0.1038	0.1220
0.01		10	0.1131	0.1010	0.1014	0.1302	0.1560
	10	20	0.1671	0.1711	0.1711	0.1924	0.2284
		50	0.3599	0.3811	0.3813	0.3991	0.4333
		10	0.0831	0.0916	0.0912	0.1136	0.1505
	5	20	0.1013	0.1194	0.1196	0.1353	0.1726
		50	0.1653	0.2064	0.2048	0.2174	0.2555
0.05		10	0.1994	0.1843	0.1855	0.2207	0.2668
	10	20	0.2907	0.2959	0.2968	0.3152	0.3706
		50	0.5399	0.5545	0.5538	0.5708	0.6067
		10	0.1280	0.1382	0.1362	0.1679	0.2126
	5	20	0.1545	0.1818	0.1785	0.2017	0.2452
		50	0.2430	0.2933	0.2897	0.3050	0.3479
0.10		10	0.2604	0.2438	0.2440	0.2825	0.3381
	10	20	0.3781	0.3795	0.3796	0.4048	0.4600
		50	0.6314	0.6463	0.6469	0.6604	0.6950

Table 3.3: Empirical power of statistics X^2, D, D_1^*, X_*^2 and D^* for dispersion parameter $\psi = 0.10$, nominal levels $\alpha = .01, .05, .10, m = 5, 10$ and $n = 10, 20, 50$ based on 10,000 replications.

nominal level α	binomial parameter m	sample size n	Pearson statistic X^2	Deviance statistic D	Deviance statistic D_1^*	modified Pearson statistic X_*^2	modified deviance statistic D_*
		10	0.0824	0.0918	0.0894	0.1161	0.1438
	5	20	0.1288	0.1581	0.1577	0.1764	0.2098
		50	0.2839	0.3498	0.3469	0.3671	0.4026
0.01		10	0.2924	0.2710	0.2715	0.3177	0.3534
	10	20	0.4933	0.4972	0.4981	0.5291	0.5657
		50	0.8543	0.8690	0.8677	0.8770	0.8947
		10	0.1603	0.1735	0.1675	0.1986	0.2502
	5	20	0.2475	0.2783	0.2773	0.3016	0.3563
		50	0.4722	0.5322	0.5296	0.5514	0.5921
0.05		10	0.4071	0.3944	0.3921	0.4347	0.4931
	10	20	0.6372	0.6448	0.6435	0.6617	0.7123
		50	0.9337	0.9408	0.9401	0.9437	0.9530
		10	0.2163	0.2314	0.2253	0.2645	0.3349
	5	20	0.3266	0.3629	0.3592	0.3867	0.4501
		50	0.5806	0.6310	0.6296	0.6484	0.6905
0.10		10	0.4841	0.4687	0.4670	0.5067	0.5696
	10	20	0.7098	0.7150	0.7152	0.7301	0.7752
		50	0.9583	0.9617	0.9614	0.9651	0.9720

Table 3.4: Empirical power of statistics X^2, D, D_1^*, X_*^2 and D^* for dispersion parameter $\psi = 0.15$, nominal levels $\alpha = .01, .05, .10, m = 5, 10$ and $n = 10, 20, 50$ based on 10,000 replications.

nominal level α	binomial parameter m	sample size n	Pearson statistic X^2	Deviance statistic D	Deviance statistic D_1^*	modified Pearson statistic X_*^2	modified deviance statistic D_*
		10	0.1527	0.1661	0.1598	0.2050	0.2454
	5	20	0.2561	0.3018	0.2989	0.3269	0.3770
		50	0.5879	0.6585	0.6556	0.6766	0.7122
0.01		10	0.4825	0.4700	0.4678	0.5125	0.5545
	10	20	0.7576	0.7706	0.7698	0.7846	0.8170
		50	0.9825	0.9861	0.9863	0.9858	0.9893
		10	0.2704	0.2852	0.2771	0.3295	0.3909
	5	20	0.4159	0.4626	0.4586	0.4866	0.5529
		50	0.7607	0.8078	0.8060	0.8135	0.8455
0.05		10	0.6104	0.5964	0.5952	0.6328	0.6868
	10	20	0.8535	0.8594	0.8580	0.8700	0.8928
		50	0.9942	0.9948	0.9947	0.9958	0.9967
		10	0.3403	0.3658	0.3547	0.3985	0.4778
	5	20	0.5120	0.5569	0.5533	0.5758	0.6456
		50	0.8347	0.8662	0.8652	0.8712	0.8948
0.10		10	0.6740	0.6655	0.6641	0.6944	0.7520
	10	20	0.8924	0.8958	0.8947	0.9044	0.9263
		50	0.9972	0.9977	0.9976	0.9976	0.9983

Table 3.5: Empirical power of statistics X^2, D, D_1^*, X_*^2 and D^* for dispersion parameter $\psi = 0.20$, nominal levels $\alpha = .01, .05, .10, m = 5, 10$ and $n = 10, 20, 50$ based on 10,000 replications.

nominal level α	binomial parameter m	sample size n	Pearson statistic X^2	Deviance statistic D	Deviance statistic D_1^*	modified Pearson statistic X_*^2	modified deviance statistic D_*
		10	0.2343	0.2521	0.2435	0.2919	0.3428
	5	20	0.4284	0.4881	0.4814	0.5131	0.5666
		50	0.8226	0.8668	0.8642	0.8771	0.8978
0.01		10	0.6216	0.6143	0.6118	0.6512	0.6921
	10	20	0.8929	0.9014	0.9002	0.9079	0.9268
		50	0.9988	0.9989	0.9989	0.9988	0.9991
		10	0.3632	0.3874	0.3763	0.4290	0.4984
	5	20	0.5993	0.6435	0.6388	0.6651	0.7197
		50	0.9205	0.9447	0.9431	0.9480	0.9594
0.05		10	0.7271	0.7283	0.7222	0.7505	0.7956
	10	20	0.9443	0.9474	0.9475	0.9506	0.9632
		50	0.9998	0.9996	0.9996	0.9998	0.9997
		10	0.4418	0.4706	0.4555	0.5055	0.5841
	5	20	0.6862	0.7230	0.7167	0.7371	0.7953
		50	0.9535	0.9665	0.9659	0.9671	0.9749
0.10		10	0.7791	0.7797	0.7749	0.7992	0.8438
	10	20	0.9605	0.9648	0.9638	0.9670	0.9779
		50	0.9999	0.9997	0.9998	1.0000	0.9999

3.6 Examples

3.6.1 Incidence of Hepatitis *A* in Bulgaria

Farrington (1996) presents an analysis of data on incidence of hepatitis *A* in Bulgaria by age (see Table D.1), given by Keiding (1991). The data are sparse with 19 out of 83 annual age groups contributing non-zero denominators of 5 or less. Using the numbers of seronegatives as response variables with binomial errors and the log link $\log \pi_a = -\lambda a$ we obtain $X^2 = 94.59$ on 82 degrees of freedom and $X^2_* = 107.58$ with conditional expected value 82.10, conditional variance 133.47 and the conditional standardized third moment 0.6305. These are almost identical to those obtained by Farrington (1996). Further, we obtain $D = 97.28$ on 82 degrees of freedom and $D^* = 104.53$ with conditional expected value 81.46, conditional variance 95.30 and the conditional standardized third moment -0.1675. The p-values of X^2 and D on 82 degrees of freedom based on a χ^2 distribution are 0.1614 and 0.1194. The p-values of X^2_* and D^* based on standardized normal distribution are 0.0274 and 0.0175. The p-values of X^2_* and D^* based on the Edgeworth approximation are 0.0559 and 0.0114. Whereas both X^2 and D show that the fit of the model is good, the tests based on X^2_* and D^* indicate evidence against the model. Also, note that the p-values of D^* based on the standardized normal approximation and based on the Edgeworth approximation are close, where as the p-value of X^2_* based on the Edgeworth approximation is almost twice the p-value based on the standardized normal approximation.

3.6.2 Mosquito transmission of Yellow Fever

Farrington (1996) analyzes data on transovarial transmission of yellow fever virus in mosquito populations. The data, originally given by Walter *et.al.* (1980)(see Table D.2), refer to an adult population of *Aedes aegypti* which was infected with yellow fever produced a progeny population which was hatched and reared to adults, separated by sex and grouped in pools of variable size for virus assay. The data consist of 63 triplets (s_i, m_i, r_i) cross-classified by two binary covariates, virus strain and larval development interval. For each triplet, s_i is the pool size, m_i is the number of pools of this size which were assayed and r_i is the number of pools found to be positive. The quantity of interest is the probability of transmission in individual mosquitos. For each observation let λ_i denote this transmission probability and π_i the probability that a pool is positive. Under suitable independence assumptions, $\pi_i = 1-(1-\lambda_i)^{s_i}$ so that given covariates x_i and a linear model $\log\{-\log(1-\lambda_i)\} = x_i^T\beta$ we have $\log\{-\log(1-\pi_i)\} = \log s_i + x_i^T\beta$. The model is fitted by regarding the r_i as binomial with denominator m_i, using the complementary log-log link function with offsets $\log s_i$. In this case

$$h_i' = m_i(1-\pi_i)\log(1-\pi_i).$$

In this instance the data display extreme sparseness, with $m_i > 1$ for only nine of the 63 observations. The main effects model with both binary covariates gives $X^2 = 78.83, D = 72.70$ on 60 degrees of freedom, $X_*^2 = 64.70$ with conditional expectation 61.3, conditional variance 13.4 and the conditional standardized third moment 0.8588, and $D^* = 65.89$ with conditional expectation 61.23, conditional variance 13.58 and the conditional standardized third moment -0.3416. The p-values for X^2 and D on 60 degrees of freedom are 0.052 and 0.126 and those for X_*^2 and D^* based on standardized normal distribution are 0.298 and 0.206 respectively. The p-values of X_*^2 and D^* based on the Edgeworth approximation are 0.303 and 0.193. The p-values for D^* and X_*^2 are much larger than those of D and X^2. Both statistics D^* and X_*^2 show similar evidence of fit of the model to the data. In this example the p-values of D^* based on the standardized normal approximation and based on the Edgeworth approximation are almost identical.

3.6.3 Multiple tumour recurrence data for patients with bladder cancer

Andrews and Herzberg (1985) provide multiple tumour recurrence data for patients with bladder cancer obtained in a randomized clinical trial. Briefly, data (see Table D.3) are obtained for each of three groups of patients and the data consist, among others, of the number of recurrences r_i experienced by each of 118 patients, the number of tumours t_i present initially at the time of randomization in the trial and the diameter of the largest of those d_i. For more detailed description of the data see Andrews and Herzberg (1985). By regarding the number of recurrences r_i as Poisson with log-link function for the mean number of recurrences we fitted a quadratic model relating r_i to t_i and d_i to data for each group separately and to the combined data of all the three groups. Results of all data sets showed highly significant mis-fit of the model, indicating presence of strong over-dispersion. For illustrative purpose only we give results for data of group 3. The data in this example also show extreme sparseness. We obtained $X^2 = 60.29$, $D = 57.60$ on 32 degrees of freedom, $X_*^2 = 57.02$ with conditional expectation 32.0 and conditional variance 64.0 and $D^* = 54.13$ with conditional expectation 24.80 and conditional variance 76.01. The p-values for X^2 and D based on a χ^2 distribution with 32 degrees of freedom are 0.0018 and 0.0036 and those for X_*^2 and D^* based on standardized normal distribution are 0.00176 and 0.000768 respectively. Strong evidence of departure from the generalized linear model is shown by all the statistics. However, this evidence is much stronger when we use the result based on the statistic D^*.

Note that one common feature of these three examples is that the p-value for D^* is always smaller than that of X_*^2, indicating that the former may be more powerful than the latter. This is in line with the conclusion from simulations in Section 3.5.

Chapter 4

Score test for goodness of fit of generalized linear models to sparse data

4.1 Introduction

As we mention in Chapter 3, for the discrete data that are extensive but sparse McCullagh (1985) obtains a goodness of fit test statistic using approximations to the first three moments of the conditional distribution of the Pearson chi-square statistic for canonical exponential family regression models. Farrington (1996) extends the results of McCullagh (1985) to obtain the first three moments of the conditional distribution of a modified Pearson chi-square statistic for models with non-canonical links. The goodness of fit test is conducted using the fact that the distribution of the standardized Pearson chi-square statistic is asymptotically standard normal.

Farrington (1996) develops a modified Pearson statistic using an estimating-equations approach following Moore (1986). In this chapter we derive a score test using the estimating functions used by Farrington (1996). We show that the score test statistic can be derived in only a few steps compared to the modified Pearson chi-square statistic. A modified score test is also obtained. Both the score test and the modified score test are closely related to the test based on the modified Pearson statistic developed by Farrington (1996).

Applications of the test for assessing goodness of fit of binomial and the Poisson models to sparse data are shown. For Poisson log-linear models the modified score test statistic is identical to the standardized modified Pearson statistic. For binomial data we conduct a small-scale simulation study to compare the performance of the score tests with the standardized modified Pearson statistic of Farrington (1996). Two sets of data with different degrees of sparseness and using different link functions are also analyzed.

4.2 The standardized modified Pearson statistic

We use the same notations as in Section 2.3. Note that by taking $a_i = -V_i'/V_i$, asymptotically, the first two conditional moments of X_*^2 to the order $n^{-3/2}$ are

$$\begin{aligned} E(X_*^2|\hat{\beta}) &= n - p - \frac{1}{2}\sum_{i=1}^{n}\frac{\hat{V}_i''}{\hat{V}_i}\hat{h}_i'^2, \\ \mathrm{var}(X_*^2|\hat{\beta}) &= n(\hat{\rho}_4 - \hat{\rho}_3^2 + 2), \end{aligned}$$

Now, note that

$$\begin{aligned} \kappa_{3i} &= E(y_i - \mu_i)^3 = V_i'V_i, \\ \kappa_{4i} &= E(y_i - \mu_i)^4 - 3(E(y_i - \mu_i)^2)^2 = 3V_i^2 + V_i^2V_i'' + V_i'^2V_i - 3V^2 = V_i^2V_i'' + V_i'^2V_i. \end{aligned}$$

So,

$$\begin{aligned} n(\rho_4 - \rho_3^2 + 2) &= nn^{-1}\sum_{i=1}^{n}(\kappa_{4i}/V_i^2 - \kappa_{3i}^2/V_i^3 + 2) \\ &= \sum_{i=1}^{n}[(V_i^2V_i'' + V_i'^2V_i)/V_i^2 - (V_i'V_i)^2/V_i^3 + 2] = \sum_{i=1}^{n}(V_i'' + 2), \end{aligned}$$

and thus,

$$\mathrm{var}(X_*^2|\hat{\beta}) = \sum_{i=1}^{n}(\hat{V}_i'' + 2).$$

The standardized quantity

$$Z_1 = \frac{X_*^2 - (n - p - \frac{1}{2}\sum_{i=1}^{n}\frac{\hat{V}_i''}{\hat{V}_i}\hat{h}_i'^2)}{(\sum_{i=1}^{n}(\hat{V}_i'' + 2))^{1/2}}$$

has an approximate $N(0,1)$ distribution.

4.3 The score test for assessing goodness of fit of the generlaized linear model

Paul and Islam (1995) develop a score test for testing homogeneity of binomial proportions with over/under dispersion based on the quasi likelihood score functions. Paul and Banerjee (1998) develop score tests for interaction and main effects, in unbalanced two-way layout of counts involving two fixed factors, when data are extra dispersed, also, based on the quasi likelihood score function. We follow similar steps here.

The quasi likelihood score functions in our case are $g_r(\beta), r = 1, ..., p$ given in equation (2.3.2) and $g_q(\beta, \phi)$ given in equation (2.3.3). Now, define

$T = g_q(\beta, \phi)|_{\phi=1}, E_\beta = (E(-\frac{\partial g_r}{\partial \beta_s})|_{\phi=1}), E_{\beta\phi} = (E(g_r g_q|_{\phi=1}))$ and $E_\phi = E[(g_q|_{\phi=1})^2]$.

Note that T and E_ϕ are scalars, $E_{\beta\phi}$ is a column vector of dimension p and E_β is a matrix of dimensions $p \times p$. The score test statistic, then, is

$$Z_2 = \hat{T}/\sqrt{\hat{\text{var}}(T)},$$

where, $\hat{\text{var}}(T) = (\hat{E}_\phi - \hat{E}_{\beta\phi}\hat{E}_\beta^{-1}\hat{E}'_{\beta\phi})$. The quantities $\hat{T}, \hat{E}_\beta, \hat{E}_{\beta\phi}$ and $\hat{E}_\phi$ are $T, E_\beta, E_{\beta\phi}$ and E_ϕ respectively evaluated at the maximum quasi likelihood estimate $\hat{\beta}$ under the null hypothesis. By the Central Limit Theorem the distribution of the score statistic Z_2, then, is asymptotically standard normal.

Now, let E_β^{rs} be the (r, s)th term of the matrix E_β and $E_{\beta\phi}^r$ be the rth term of the vector $E_{\beta\phi}$. Then,

$$\begin{aligned}
T &= g_q(\beta, \phi)|_{\phi=1} = \sum_{i=1}^{n} \left\{ -\frac{V_i'}{V_i}(y_i - \mu_i) + \frac{(y_i - \mu_i)^2}{V_i} - 1 \right\}, \\
E_\beta^{rs} &= E(-\frac{\partial g_r}{\partial \beta_s})|_{\phi=1} = \sum_{i=1}^{n} \frac{1}{V_i} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial \mu_i}{\partial \beta_s} = (X'WX)_{rs}, \\
E_{\beta\phi}^r &= E(g_r g_q)|_{\phi=1} = 0, \\
E_\phi &= E\left[(g_q|_{\phi=1})^2\right] = \sum_{i=1}^{n} \left(\frac{E(y_i - \mu_i)^4}{V_i^2} + (-\frac{V_i'}{V_i})^2 V_i - 2\frac{V_i'}{V_i} V_i' - 1 \right)
\end{aligned}$$

$$= \sum_{i=1}^{n}\left(\frac{E(y_i-\mu_i)^4}{V_i^2} - \frac{V_i'^2}{V_i} - 1\right) = \sum_{i=1}^{n}(2+V_i'').$$

Thus,

$$\mathrm{var}(T) = E_\phi - E_{\beta\phi}E_\beta^{-1}E'_{\beta\phi} = \sum_{i=1}^{n}(2+V_i'')$$

Therefore, the score test statistic Z_2 is

$$Z_2 \;=\; \hat{T}/\left(\sum_{i=1}^{n}(2+\hat{V}_i'')\right)^{1/2},$$

where $\hat{T} = \sum_{i=1}^{n}\left\{-\frac{\hat{V}_i'}{\hat{V}_i}(y_i-\hat{\mu}_i) + \frac{(y_i-\hat{\mu}_i)^2}{\hat{V}_i} - 1\right\}$.

Now, following McCullagh and Nelder (1989) and Dean and Lawless (1989), we obtain $E(\hat{T}) = -p$, as $n \to \infty$ and $\mu \to \infty$. The proof is given in the Appendix. So, a modified score test is

$$Z_3 = (\hat{T}+p)/[\sum_{i=1}^{n}(\hat{V}_i''+2)]^{1/2}.$$

It is interesting to note that $\hat{T} = X_*^2 - n$. Thus,

$$Z_2 = (X_*^2 - n)/[\sum_{i=1}^{n}(\hat{V}_i''+2)]^{1/2}$$

and

$$Z_3 = [X_*^2 - (n-p)]/[\sum_{i=1}^{n}(\hat{V}_i''+2)]^{1/2}.$$

We can see that the standardized modified Pearson statistic Z_1, the score test statistic Z_2 and the modified score test statistic Z_3 are closely related. The statistic Z_3 makes a correction over the statistic Z_2 for the number of regression parameters estimated. The statistic Z_1 is related to Z_3 as

$$Z_1 = Z_3 + \frac{(\frac{1}{2}\sum_{i=1}^{n}\frac{\hat{V}_i''}{\hat{V}_i}\hat{h}_i'^2)}{(\sum_{i=1}^{n}(\hat{V}_i''+2))^{1/2}}.$$

Note that the formulae for the statistics Z_2 and Z_3 do not depend on the link function, whereas the formula for Z_1 may depend on the link function as shown in next section.

4.4 Applications to binomial and Poisson data

4.4.1 Binomial data

Let $Y_i, i = 1, ..., n$, denote independent random variables from a binomial distribution with parameters m_i and π_i, where π_i depends on the link function h_i. Now, $\mu_i = m_i\pi_i, V_i = m_i\pi_i(1-\pi_i), V_i' = (1-2\pi_i)$ and $V_i'' = -2/m_i$. Then the statistic Z_1 can be written as

$$Z_1 = \frac{\sum_{i=1}^n \left\{\frac{(y_i - m_i\hat{\pi}_i)^2 + \hat{\pi}_i(y_i - m_i\hat{\pi}_i) - y_i(1-\hat{\pi}_i)}{m_i\hat{\pi}_i(1-\hat{\pi}_i)} + \frac{p}{n} - \frac{\hat{h}_i'^2}{m_i^2\hat{\pi}_i(1-\hat{\pi}_i)}\right\}}{\left(\sum_{i=1}^n 2(m_i-1)/m_i\right)^{1/2}},$$

where $h_i' = m_i\pi_i(1-\pi_i) = V_i$ for logistic link and $h_i' = -m_i(1-\pi_i)\log(1-\pi_i)$ for complementary log-log link.

The statistics Z_2 and Z_3 can be written as

$$Z_2 = \frac{\sum_{i=1}^n \left\{\frac{(y_i - m_i\hat{\pi}_i)^2 + \hat{\pi}_i(y_i - m_i\hat{\pi}_i) - y_i(1-\hat{\pi}_i)}{m_i\hat{\pi}_i(1-\hat{\pi}_i)}\right\}}{\left(\sum_{i=1}^n 2(m_i-1)/m_i\right)^{1/2}}$$

and

$$Z_3 = \frac{\sum_{i=1}^n \left\{\frac{(y_i - m_i\hat{\pi}_i)^2 + \hat{\pi}_i(y_i - m_i\hat{\pi}_i) - y_i(1-\hat{\pi}_i)}{m_i\hat{\pi}_i(1-\hat{\pi}_i)} + \frac{p}{n}\right\}}{\left(\sum_{i=1}^n 2(m_i-1)/m_i\right)^{1/2}}.$$

It ie obvious that $Z_3 \geq Z_1$.

4.4.2 Poisson data

For Poisson data, $V_i = \mu_i, V_i' = 1$ and $V_i'' = 0$. This makes the statistic Z_1 link independent and equal to Z_3. Thus,

$$Z_1 = Z_3 = \frac{1}{\sqrt{2n}} \sum_{i=1}^n \left\{\frac{(y_i - \hat{\mu}_i)^2 - y_i}{\hat{\mu}_i} + \frac{p}{n}\right\}$$

and

$$Z_2 = \frac{1}{\sqrt{2n}} \sum_{i=1}^n \left\{\frac{(y_i - \hat{\mu}_i)^2 - y_i}{\hat{\mu}_i}\right\}.$$

Dean (1992) gives an adjusted statistic:

$$P'_c = \frac{1}{\sqrt{2n}} \sum_{i=1}^{n} \left\{ \frac{(y_i - \hat{\mu}_i)^2 - y_i + \hat{h}_{ii}\hat{\mu}_i}{\hat{\mu}_i} \right\}$$

where $\hat{h}_{ii} = h_{ii}(\hat{\beta})$ and $h_{ii}(\beta)$ is the ith diagonal element of the matrix $H = W^{1/2}X(X^TWX)^{-1} X^TW^{1/2}$, with $W = \text{diag}(\mu_1, , ..., \mu_n)$. Note that if there is no covariate, $p = 1$, $h_{ii} = 1/n$ and thus Z_3 and P'_c are identical. Also note that H is idempotent matrix and $tr(H) = p$. If $\mu_i, i = 1, ..., n$ are close, $h_{ii} \simeq p/n$. Therefore Z_3 is close to P'_c.

4.5 Simulation

A limited simulation study is conducted to compare the empirical size and power of the statistics Z_1, Z_2 and Z_3. Following Farrington (1996) simulations are conducted for binomial model with $p = 2$ and a single continuous covariate chosen to induce very strong regression effects under both logistic and complementary log-log-link functions, for sample sizes varying from $n = 10$ to $n = 100$ and binomial denominators $m = 5$ and $m = 10$. The empirical size and power for all statistics are calculated by referring to the asymptotic standard normal distribution.

Results of the simulation on empirical size for nominal level $\alpha = .05, .10$ are presented in Table 4.1 for both the logistic and complementary log-log-link functions. Results in Table 4.1 show that all the three statistics hold nominal level well in the situations investigated.

An empirical power study is conducted for nominal level $\alpha = 0.05, 0.10$, over-dispersion parameter $\phi = 0.05, 0.10$ and for both the logistic and complementary log-log-link functions. In Table 4.2 we present results only for $\alpha = 0.05$ and the logistic function. Results for other situations investigated are similar. Results in Table 4.2 show that the modified score test has some edge over the modified Pearson statistic in terms of power.

Table 4.1: Empirical sizes of statistics Z_1, Z_2 and Z_3 for nominal levels $\alpha = .05, .10$, $m = 5, 10$ and $n = 10, 20, 50, 100$ based on 10,000 replications.

nominal level α	binomial parameter m	sample size n	modified Pearson statistic Z_1	score test statistic Z_2	modified score test statistic Z_3
		a) Logistic Link			
		10	0.0466	0.0560	0.0522
		20	0.0475	0.0549	0.0508
	5	50	0.0499	0.0512	0.0502
		100	0.0486	0.0505	0.0483
0.05		10	0.0435	0.0553	0.0474
		20	0.0481	0.0541	0.0498
	10	50	0.0531	0.0551	0.0531
		100	0.0523	0.0542	0.0529
		10	0.0902	0.1336	0.0909
		20	0.0964	0.1142	0.0978
	5	50	0.0994	0.1090	0.0992
		100	0.0993	0.1013	0.1000
0.10		10	0.0815	0.1351	0.0836
		20	0.0950	0.1173	0.0958
	10	50	0.1031	0.1082	0.1028
		100	0.1008	0.1052	0.1015
		b) Complementary Log-log Link			
		10	0.0505	0.0631	0.0556
		20	0.0504	0.0585	0.0520
	5	50	0.0516	0.0535	0.0534
		100	0.0515	0.0518	0.0517
0.05		10	0.0469	0.0565	0.0506
		20	0.0475	0.0544	0.0478
	10	50	0.0489	0.0515	0.0492
		100	0.0513	0.0526	0.0516
		10	0.0940	0.1473	0.0972
		20	0.1012	0.1187	0.1027
	5	50	0.1057	0.1118	0.1047
		100	0.1032	0.1061	0.1038
0.10		10	0.0803	0.1378	0.0807
		20	0.0967	0.1184	0.0966
	10	50	0.0983	0.1085	0.0984
		100	0.1005	0.1043	0.1008

Table 4.2: Empirical powers of statistics Z_1, Z_2 and Z_3 for $\phi = 0.10$, nominal levels $\alpha = .05, .10, m = 5, 10$ and $n = 10, 20, 50, 100$ based on 10,000 replications.

nominal level α	binomial parameter m	sample size n	modified Pearson statistic Z_1	score test statistic Z_2	modified score test statistic Z_3
		a) Logistic Link			
		10	0.1986	0.1441	0.2179
		20	0.3016	0.2375	0.3202
	5	50	0.5533	0.4970	0.5669
		100	0.8067	0.7774	0.8135
0.05		10	0.4347	0.3459	0.4476
		20	0.6683	0.5924	0.6762
	10	50	0.9448	0.9273	0.9455
		100	0.9991	0.9983	0.9991
		10	0.2645	0.2076	0.2887
		20	0.3867	0.3133	0.4050
	5	50	0.6539	0.5961	0.6682
		100	0.8705	0.8455	0.8760
0.10		10	0.5067	0.4093	0.5182
		20	0.7371	0.6689	0.7447
	10	50	0.9638	0.9520	0.9650
		100	0.9997	0.9994	0.9997
		b) Complementary Log-log Link			
		10	0.2120	0.1454	0.2331
		20	0.3030	0.2344	0.3223
	5	50	0.5579	0.5011	0.5717
		100	0.8085	0.7808	0.8151
0.05		10	0.4466	0.3534	0.4556
		20	0.6814	0.6091	0.6894
	10	50	0.9469	0.9310	0.9492
		100	0.9983	0.9975	0.9985
		10	0.2904	0.2161	0.3017
		20	0.3897	0.3138	0.4099
	5	50	0.6535	0.5998	0.6675
		100	0.8685	0.8454	0.8742
0.10		10	0.5146	0.4216	0.5267
		20	0.7484	0.6819	0.7553
	10	50	0.9662	0.9549	0.9675
		100	0.9992	0.9989	0.9992

4.6 Examples

We also consider the data in the first two examples of Chapter 3.

Example 1. This data set originally given by Keiding (1991) and later analyzed by Farrington (1996) is on incidence of hepatitis A in Bulgaria by age. The data are sparse with 19 out of 83 annual age groups contributing non-zero denominators of 5 or less. By using the numbers of seronegatives as response variables with binomial errors and the log link $\log \pi_a = -\lambda a$ we obtain $Z_1 = 2.205$, $Z_2 = 2.127$ and $Z_3 = 2.2142$ with p-values 0.0275, 0.0334 and 0.0268 respectively.

Example 2. This second data set originally from Walter *et. al.*(1980) and analyzed by Farrington (1996) is on transovarial transmission of yellow fever virus in mosquito populations. An adult population of *Aedes aegypti* infected with yellow fever produced a progeny population which was hatched and reared to adults, separated by sex and grouped in pools of variable size for virus assay. The data consist of 63 triples (s_i, m_i, r_i) crossed-classified by two binary covariates, virus strain and larval development interval. For each triple s_i is the pool size, m_i is the number of pools of this size which were assayed and r_i is the number of pools found to be positive. The quantity of interest is the probability of transmission in individual mosquitos. For each observation, let λ_i denote this transmission probability and π_i the probability that a pool is positive. Under suitable independence assumptions, $\pi_i = 1 - (1 - \lambda_i)^{s_i}$ so that given covariates x_i and a linear model $\log\{-\log(1-\lambda)\} = x_i^T\beta$ we have $\log\{-\log(1 - \pi_i)\} = \log s_i + x_i^T\beta$. The model is fitted by regarding the r_i as binomial with denominator m_i, using the complementary log-log-link function with offsets $\log s_i$. In this example the data display extreme sparseness, with $m_i > 1$ for only 9 of 63 observations. For these data we obtain $Z_1 = 1.041$, $Z_2 = 0.464$ and $Z_3 = 1.283$ with p-values 0.298, 0.643 and 0.199 respectively.

One common feature of these two examples is that the p-values for Z_3 is always smaller than that of Z_1. Since both statistics hold nominal level well the above result indicates that the modified score test statistic Z_3 may be more powerful than the standardized modified Pearson statistic Z_1.

Chapter 5

Score tests for zero-inflation in generalized linear models

5.1 Introduction

When analyzing discrete data under a Poisson or binomial assumption, sometimes many more zeros are observed than expected. These data are then analyzed as a mixture model (Farewell and Sprott 1988; Lambert 1992; Broek 1995; Mullahy 1997). One of the mixing populations being a population in which only zeros are observed, while the other population is the one in which counts from a discrete distribution are observed. For further interpretation of such a mixing distribution see Farewell and Sprott (1988) and Lambert (1992).

Broek (1995) obtains a score test for zero inflation in a Poisson distribution. In the present chapter we obtain score tests for zero inflation in generalized linear models. We then obtain score tests for zero inflation in the Poisson model and in the binomial model as special cases. The score test for zero inflation in a Poisson model obtained in this chapter is identical to that obtained by Broek (1995).

In Section 5.2 we develop the score tests for zero-inflation in the generalized linear model. The score test statistics for zero-inflation in the Poisson model and the score test statistics for zero-inflation in the binomial model obtained as special cases of the score tests for zero-inflation in the generalized linear model are also given in this section. Results of a small simulation study are reported in Section 5.3. Section 5.4 follows with an illustrative example and a brief discussion.

5.2 The zero-inflated generalized linear model and score tests for zero-inflation

5.2.1 The zero-inflated generalized linear model

Consider the natural exponential family distribution with probability density function

$$f(y;\theta) = \exp\{a(\theta)y - g(\theta) + c(y)\},$$

where y represents the response variable and θ is an unknown parameter on which the distribution of y depends. This family includes both the Poisson distribution and the binomial distribution. Note that zero-inflation can occur only in the discrete data. So, here we deal with only the discrete exponential family. Then, the zero-inflated exponential family is,

$$\begin{aligned} P(Y=0) &= \omega + (1-\omega)f(0;\theta), \\ P(Y=y) &= (1-\omega)f(y;\theta) \qquad (y>0). \end{aligned} \tag{5.2.1}$$

Note that it is possible to take ω less than zero, provided that

$$\omega \geq -\frac{f(0;\theta)}{1-f(0;\theta)}$$

with equality for left truncation. Further, (i) if $\omega > 0, P(Y=0) > f(0;\theta)$ implying that there exist too many zeros (zero inflation) and (ii) if $\omega < 0, P(Y=0) < f(0;\theta)$ implying that there exist too few zeros (zero deflation). For this zero-inflated or zero-deflated exponential family, the mean and variance of Y are

$$E(Y) = (1-\omega)\mu(\theta) = (1-\omega)(a'(\theta))^{-1}g'(\theta)$$

and

$$\begin{aligned} \text{var}(Y) &= (1-\omega)\sigma^2(\theta) + \omega(1-\omega)\mu^2(\theta) \\ &= (1-\omega)(a'(\theta))^{-2}\{g''(\theta) - a''(\theta)(a'(\theta))^{-1}g'(\theta) + \omega(g'(\theta))^2\}. \end{aligned}$$

For $\omega = 0$, we obtain the mean and variance of the natural exponential family distribution

$$E(Y) = \mu(\theta) = (a'(\theta))^{-1}g'(\theta)$$

and

$$\text{var}(Y) = \sigma^2(\theta) = (a'(\theta))^{-2}\{g''(\theta) - a''(\theta)(a'(\theta))^{-1}g'(\theta)\}.$$

5.2.2 Score test for zero-inflation in the generalized linear model

Let $Y_i, i = 1, ..., n$, be a sample of independent observations from (5.2.1) with θ_i a function of a $p \times 1$ vector of covariates X_i and a vector of regression parameters β; that is, $\theta_i = \theta_i(X_i; \beta), i = 1, ..., n$. From (5.2.1), the likelihood can be written as

$$L(\omega, \theta; y) = \prod_{i=1}^{n}\{(\omega + (1-\omega)f(0;\theta_i))I_{\{y_i=0\}} + (1-\omega)f(y_i;\theta)I_{\{y_i>0\}}\}.$$

Thus, the log likelihood is

$$l(\omega, \theta; y) = \sum_{i=1}^{n}\{\log(\omega + (1-\omega)f(0;\theta_i))I_{\{y_i=0\}} + \log((1-\omega)f(y_i;\theta))I_{\{y_i>0\}}\}.$$

Now, for convenience, let $\gamma = \frac{\omega}{1-\omega}$. Note $\gamma = 0 \Leftrightarrow \omega = 0$. Then, the log-likelihood, in terms of the parameters γ and θ_i is

$$\begin{aligned} l(\gamma, \theta; y) &= \sum_{i=1}^{n} l_i(\gamma, \theta_i; y_i) \\ &= \sum_{i=1}^{n}\{-\log(1+\gamma) + I_{\{y_i=0\}}\log(\gamma + f(0;\theta)) + I_{\{y_i>0\}}\log f(y_i;\theta)\} \\ &= \sum_{i=1}^{n}\{-\log(1+\gamma) + I_{\{y_i=0\}}\log(\gamma + \exp(-g(\theta_i) + c(0))) \\ &\quad + I_{\{y_i>0\}}(a(\theta_i)y_i - g(\theta_i) + c(y_i))\}. \end{aligned}$$

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

$$\Psi = \frac{\partial l}{\partial \gamma}|_{\gamma=0} = \sum_{i=1}^{n}\{-\frac{1}{1+\gamma} + I_{\{y_i=0\}}\frac{1}{\gamma + f(0;\theta_i)}\}|_{\gamma=0} = \sum_{i=1}^{n}(\frac{I_{\{y_i=0\}}}{f(0;\theta_i)} - 1).$$

Partition $I(\beta,\gamma)$ as follows

$$I(\beta,\gamma) = \begin{bmatrix} I_{\beta\beta} & I_{\beta\gamma} \\ I'_{\beta\gamma} & I_{\gamma\gamma} \end{bmatrix},$$

where $I_{\beta\beta}, I_{\beta\gamma}$ and $I_{\gamma\gamma}$ are $p\times p, p\times 1$ and 1×1 matrices, respectively. Note that

$$\begin{aligned}
\frac{\partial l_i}{\partial \theta_i} &= I_{\{y_i=0\}}\frac{f(0;\theta_i)(-g_i')}{\gamma+f(0;\theta_i)} + I_{\{y_i>0\}}\{a_i'y_i - g_i'\}, \\
\frac{\partial^2 l_i}{\partial \theta_i^2} &= I_{\{y_i=0\}}\{\frac{f(0;\theta_i)(-g_i')^2}{\gamma+f(0;\theta_i)} - \frac{(f(0;\theta_i))^2(-g_i')^2}{(\gamma+f(0;\theta_i))^2} + \frac{f(0;\theta_i)(-g_i'')}{\gamma+f(0;\theta_i)}\} \\
&\quad + I_{\{y_i>0\}}\{a_i''y_i - g_i''\}, \\
\frac{\partial^2 l_i}{\partial \theta_i \partial \gamma} &= I_{\{y_i=0\}}\frac{-f(0;\theta_i)(-g_i')}{(\gamma+f(0;\theta_i))^2}, \\
\frac{\partial l_i}{\partial \gamma} &= -\frac{1}{1+\gamma} + I_{\{y_i=0\}}\frac{1}{\gamma+f(0;\theta_i)}
\end{aligned}$$

and

$$\frac{\partial^2 l_i}{\partial \gamma^2} = \frac{1}{(1+\gamma)^2} - I_{\{y_i=0\}}\frac{1}{(\gamma+f(0;\theta_i))^2}.$$

Let U be an $n\times p$ matrix with ir-element $\frac{\partial\theta_i}{\partial\beta_r}$, $\mathbf{1}$ be an $n\times 1$ unit vector, $h_i = h_i(\theta_i) = \log a_i'(\theta_i)$ and W_1 and W_2 be diagonal matrices with ith diagonal elements

$$\begin{aligned}
W_{1i} &= E\{-\frac{\partial^2 l_i}{\partial \theta_i^2}\}|_{\gamma=0} = g_i'' f(0;\theta_i) + (-a_i''Ey_i + g_i''(1-f(0;\theta_i)) \\
&= g_i'' - h_i'g_i'
\end{aligned}$$

and

$$W_{2i} = E\{-\frac{\partial^2 l_i}{\partial\theta_i\partial\gamma}\}|_{\gamma=0} = -g_i'.$$

Then,

$$\begin{aligned}
I_{\beta\beta} &= U^T W_1 U, \\
I_{\beta\gamma} &= U^T W_2 \mathbf{1}
\end{aligned}$$

and

$$I_{\gamma\gamma} = \sum_{i=0}^{n} E\{-\frac{\partial^2 l_i}{\partial\gamma^2}\}|_{\gamma=0} = \sum_{i=1}^{n}(-1 + \frac{f(0;\theta_i)}{(f(0;\theta_i))^2} = \sum_{i=1}^{n}(\frac{1}{f(0;\theta_i)} - 1).$$

Thus, the asymptotic variance of Ψ is

$$V = I_{\gamma\gamma} - I'_{\beta\gamma} I_{\beta\beta}^{-1} I_{\beta\gamma} = I_{\gamma\gamma} - \mathbf{1}^T W_2 U (U^T W_1 U)^{-1} U^T W_2 \mathbf{1}.$$

The score test statistic for testing that $\gamma = 0$ is thus

$$S = \hat{\Psi}/\hat{V}^{1/2},$$

where $\hat{\Psi} = \Psi(\hat{\theta}_1, ..., \hat{\theta}_n)$, $\hat{V} = V(\hat{\theta}_1, ..., \hat{\theta}_n)$ and $\hat{\theta}_i$ is the maximum likelihood estimate of θ_i when $\gamma = 0$. The statistic S^2, asymptotically, as $n \to \infty$, has a $\chi^2(1)$ distribution.

5.2.3 Score tests for zero-inflation in Poisson data

For the Poisson distributed data we have $\theta = \log \mu, a(\theta) = \theta, g(\theta) = e^{\theta}, h(\theta) = 0$. Then, the score test statistic reduces to

$$S_1 = \frac{\sum_{i=1}^{n}(\frac{I_{\{y_i=0\}}}{e^{-\hat{\mu}_i}} - 1)}{\sqrt{\sum_{i=1}^{n}(\frac{1}{e^{-\hat{\mu}_i}} - 1) - \hat{\mu}^T U (U^T \operatorname{diag}(\hat{\mu}_i) U)^{-1} U^T \hat{\mu}}},$$

where $\hat{\mu}_i = e^{\hat{\theta}_i} = e^{\theta_i(X_i, \hat{\beta})}$ and $\hat{\beta}$ is the maximum likelihood estimate of the parameter β. For $\theta_i = X_i\beta$ the statistic S_1 is the same as the score test statistic given in equation (3) of Broek (1995). All other special cases in Broek (1995) also follow from S_1.

5.2.4 Score tests for zero-inflation in binomial data

For the binomial distribution: $\theta = \log(p/(1-p)), a(\theta) = \theta, g(\theta) = m\log(1+e^{\theta}), h(\theta) = 0$. Then,

$$f(0;\theta_i) = \exp(-m_i \log(1 + e^{\theta_i})) = (1 - p_i)^{m_i},$$

$$S_i(\theta_i) = (\frac{I_{\{y_i=0\}}}{(1-p_i)^{m_i}} - 1),$$

$$\begin{aligned} W_{1i} &= g_i'' - h_i' g_i' = \frac{m_i e^{\theta_i}}{(1+e^{\theta_i})^2} = m_i p_i (1-p_i) = V_i, \\ W_{2i} &= -g_i' = -\frac{m_i e^{\theta_i}}{1+e^{\theta_i}} = -m_i p_i = -\mu_i, \\ I_{\beta\beta} &= U^T W_1 U = U^T \operatorname{diag}(V_i) U, \\ I_{\beta\gamma} &= U^T \mu \end{aligned}$$

and

$$I_{\gamma\gamma} = \sum_{i=1}^{n} \left(\frac{1}{(1-p_i)^{m_i}} - 1 \right).$$

Thus, the score test statistic for testing zero-inflation in binomial data is

$$S_2 = \frac{\sum_{i=1}^n \left(\frac{I_{\{y_i=0\}}}{(1-\hat{p}_i)^{m_i}} - 1 \right)}{\sqrt{\sum_{i=1}^n \left(\frac{1}{(1-\hat{p}_i)^{m_i}} - 1 \right) - \hat{\mu}^T U (U^T \hat{V} U)^{-1} U^T \hat{\mu}}},$$

where $\hat{p}_i = e^{\hat{\theta}_i}/(1+e^{\hat{\theta}_i}) = e^{\theta_i(X_i,\hat{\beta})}/(1+e^{\theta_i(X_i,\hat{\beta})})$, $\hat{\mu} = (\hat{\mu}_1, ..., \hat{\mu}_n)^T$ and $\hat{\beta}$ is the maximum likelihood estimate of the parameter β. Similarly, if there exists no covariate, then $p_i = p$ and thus the score test statistic for testing whether the binomial distribution fits the number of zeros well is

$$S_2 = \frac{\sum_{i=1}^n \left(\frac{I_{\{y_i=0\}}}{(1-\hat{p})^{m_i}} - 1 \right)}{\sqrt{\sum_{i=1}^n \left(\frac{1}{(1-\hat{p})^{m_i}} - 1 \right) - \frac{(\sum m_i \hat{p})^2}{\sum m_i \hat{p}(1-\hat{p})}}} = \frac{\sum_{i=1}^n \left(\frac{I_{\{y_i=0\}}}{(1-\hat{p})^{m_i}} - 1 \right)}{\sqrt{\sum_{i=1}^n \left(\frac{1}{(1-\hat{p})^{m_i}} - 1 - \frac{m_i \hat{p}}{1-\hat{p}} \right)}},$$

where $\hat{p} = \sum y_i / \sum m_i$. Further, for $m = m_i$,

$$S_2 = \frac{n_0 - n\hat{p_0}}{\sqrt{n\hat{p_0}(1-\hat{p_0}) - \hat{p_0}^2 \frac{nm\overline{y}}{m-\overline{y}}}},$$

where $\hat{p_0} = (1-\hat{p})^m, \overline{y} = \sum y_i/n$ and $\hat{p} = \overline{y}/m$.

5.3 Simulation

A limited simulation study is conducted to examine the empirical size and power of the score test statistics for testing zero-inflation. Two sets of simulations are conducted. One for testing zero-inflation in Poisson(μ), for values of μ is given in Table 5.1 and the other for testing zero inflation in binomial(10, p), for values of p is given in Table 5.2. Covariates are not considered in either study. Samples of size $n = 20, 50, 100$ are taken from either the zero-inflated Poisson(μ) distribution or from the zero-inflated binomial(10, p) distribution for the inflation parameter $\omega = 0$ (for size), 0.05, 0.1, 0.2, 0.3, 0.4. Each experiment for size or power is based on 10,000 replications.

The results in Table 5.1 and Table 5.2 indicate that the score tests hold the nominal level reasonably well, although in some instances these show some conservative behavior. The power of the tests for detecting zero-inflation increases very slowly for low μ and p. In particular, for small $n (n = 20)$ and very small p ($p = 0.01$) the binomial score test shows no power for detecting zero-inflation. For larger values of μ (for example $\mu = 2$) and p (for example $p = 0.2$) power increases very fast and approaches 1.0 for $\omega = 0.4$.

Table 5.1: Power (%) of score test statistic S_1^2 of equation (2.2) with no covariates when data are simulated from Poisson (μ) : $\alpha = 0.05$; based on 10,000 replications.

		ω					
n	μ	0	.05	.10	.20	.30	.40
20		4.51	3.92	4.27	5.08	6.53	8.70
50	0.5	4.56	4.63	5.53	8.11	12.18	18.05
100		5.10	5.24	6.56	10.49	18.54	27.66
20		4.83	5.33	5.68	9.09	14.23	20.25
50	1.0	4.91	5.23	7.47	16.83	30.25	44.27
100		5.17	6.26	11.74	31.41	57.06	76.66
20		4.66	5.68	8.10	16.24	28.04	41.07
50	1.5	5.09	7.18	13.98	36.01	61.84	80.61
100		4.88	9.52	24.36	65.26	91.33	98.27
20		4.08	7.36	12.80	30.52	50.10	67.00
50	2.0	4.91	10.42	25.66	64.03	88.96	97.49
100		4.74	16.59	45.08	90.88	99.49	100.0

Table 5.2: Power (%) of score test statistic S_2^2 of equation (2.3) with no covariates when data are simulated from binomial $(10, p)$: $\alpha = .05$; based on 10,000 replications.

		ω					
n	p	0	.05	.10	.20	.30	.40
20		3.94	4.15	3.92	4.03	4.01	3.51
50	.01	4.35	4.80	4.91	5.58	6.10	6.33
100		4.59	5.12	5.69	6.49	8.37	9.63
20		4.34	5.00	5.23	5.99	6.71	7.10
50	.02	3.32	3.67	4.40	5.54	6.90	8.64
100		3.71	4.39	5.17	6.95	8.95	11.77
20		3.87	3.82	4.28	5.39	7.52	10.27
50	.05	4.98	5.40	6.15	8.22	11.97	16.39
100		5.17	5.38	6.76	10.65	19.12	28.10
20		4.82	5.47	5.86	8.89	14.57	21.63
50	.10	5.28	5.39	8.21	18.36	33.29	49.32
100		5.13	6.43	12.16	32.84	57.29	77.52
20		3.91	7.70	15.40	35.82	56.75	73.38
50	.20	5.02	12.87	31.38	72.46	93.32	98.86
100		4.79	19.84	54.13	95.02	99.87	100.0

5.4 An example and a discussion

Broek (1995) analyzes a set of data consisting of counts involving a large number of zeros. In this paper we analyze a set of data given originally by Berry (1987) as counts and later analyzed by Farewell and Sprott (1988) as proportions. As an illustrative example for the use of the score test for zero-inflation in binomial data we also treat these data as proportions. The data pertain to twelve patients who experience frequent premature ventricular contractions (PVCs) and are administered a drug with antiarrhythmic properties. One-minute EKG recordings are taken before and after drug administration. The PVCs are counted on both recordings. The data are presented in Table D.4. Note that the observations occur as paired data (x_i, y_i), which are the predrug and postdrug count, respectively, for the ith patient. Assume that x_i is a Poisson variate with mean λ_i and that for patients who are not cured y_i is independently Poisson with mean $\beta\lambda_i$. In order to eliminate the "incidental" nuisance parameters λ_i, one for each uncured subject, Farewell and Sprott (1988) use the conditional distribution of y_i given $m_i = x_i + y_i$, which is

$$Pr(y_i|m_i = x_i + y_i) = \binom{m_i}{y_i} p^{y_i}(1-p)^{m_i - y_i}, y_i = 0, 1, ..., m_i,$$

where $p = \beta\lambda_i/(\lambda_i + \beta\lambda_i) = \beta/(1+\beta)$. Letting ω be the probability of cure, which implies directly that $y_i = 0$, Farewell and Sprott (1988) use the distribution of y_i, conditional on m_i as

$$Pr(y_i = 0|m_i) = \omega + (1-\omega)(1-p)^{m_i}$$

and

$$Pr(y_i|m_i) = (1-\omega)\binom{m_i}{y_i} p^{y_i}(1-p)^{m_i - y_i}, y_i = 1, ..., m_i.$$

This is the zero-inflated binomial model. Based on this model we obtain the value of the score test statistic $S_2 = 30.513$ indicating very strongly against the fit of the binomial model to the data.

The failure of the binomial model to fit the data well may also be due to some other reason such as the presence of over-dispersion in the data. So, we next test for the presence of over-dispersion. For this we use a score test

$$S_B = \frac{\sum_{i=1}^{n}\{[\hat{p}(1-\hat{p})]^{-1}[(y_i - m_i\hat{p})^2 + \hat{p}(y_i - m_i\hat{p}) - y_i(1-\hat{p})]\}}{\{2\sum_{i=1}^{n} m_i(m_i-1)\}^{1/2}}$$

given in Dean (1992, p.455). The statistic S_B has asymptotically a standard normal distribution. The value of S_B for the data in Table D.4 is 15.378 indicating very strongly against the binomial model.

The above analyses do not indicate which of the zero-inflated binomial model and the over-dispersed binomial model will fit the data better. To check this we fitted the binomial model, the zero-inflated binomial model as given above and the beta-binomial model, which is an over-dispersed binomial model, with probability parameter π and dispersion parameter ϕ having probability function

$$Pr(y_i|m_i) = \binom{m_i}{y_i} \frac{\prod_{r=0}^{y_i-1}(\pi + r\phi)\prod_{r=0}^{m_i-y_i-1}(1-\pi+r\phi)}{\prod_{r=0}^{m_i-1}(1+r\phi)}.$$

Let $\hat{l}_0$, $\hat{l}_1$ and $\hat{l}_2$ be the maximized log-likelihood under the binomial, the zero-inflated binomial and the beta-binomial model respectively. Then for the data in Table D.4 we obtain the maximum likelihood estimates of the parameters and the maximized log-likelihoods for the three models as: $\hat{p} = 0.125, \hat{l}_0 = -40.69034; \hat{p} = 0.38614, \hat{\omega} = 0.57549, \hat{l}_1 = -18.87348; \hat{\pi} = 0.12496, \hat{\phi} = 0.53169$ and $\hat{l}_2 = -19.17807$. The log-likelihood under the zero-inflated binomial model is by far the largest, indicating that the zero-inflated binomial model is the best for these data.

Note that we have dealt with goodness of fit tests of a generalized linear model (Poisson or binomial) against zero-inflation. In practice, discrete data pertaining to a zero-inflated generalized linear model may also be over-dispersed. The tests developed here are then not appropriate to test for zero-inflation. Further extensive research is necessary to find appropriate procedures to (i)test for zero-inflation in presence of over-dispersion, (ii) test for over-dispersion in presence of zero-inflation, and (iii) test for both over-dispersion and zero-inflation. The third test is necessary when zero-inflation confounds with over-dispersion and vice-versa. We will deal with these issues in Chapter 6.

Chapter 6

Generalized linear model, zero-inflation and over-dispersion

6.1 Introduction

Broek (1995) obtains a score test tc test whether the number of zeros is too large for a Poisson distribution to fit the data well. In Chapter 5 we generalize this to test goodness of fit of generalized linear (Poisson or binomial) models. However, as we point out in Chapter 5, a generalized linear model (Poisson or binomial) may fail to fit a set of data having a large number of zeros purely because of presence of zero-inflation in the data or because there is zero-inflation as well as over-dispersion in the data. This chapter is concerned with analyzing such data. For this we consider a zero-inflated over-dispersed generalized linear model. The zero-inflated over-dispersed generalized linear model is a mixture model. One of the mixing populations is a population in which only zeros are observed, while the other population is the one in which counts from an over-dispersed generalized linear model are observed.

The over-dispersed generalized linear model considered here is of the form considered by Cox (1983) and Dean (1992). We then obtain score tests (i) for zero-inflation in presence of over-dispersion, (ii) for over-dispersion in presence of zero-inflation, and (iii) simultaneously for testing for zero-inflation and over-dispersion. For Poisson and binomial data these score tests are compared with those obtained from the zero-inflated negative binomial model and the zero-inflated beta-binomial model. Some simulations are performed for Poisson data to study level properties of the tests and two data sets are analyzed.

In Section 6.2 the zero-inflated over-dispersed generalized linear model is introduced. Score tests for selecting a model from the class of zero-inflated over-dispersed generalized linear models are developed in Section 6.3. Results for Poisson data based on the zero-inflated over-dispersed generalized linear model with log-link and the zero-inflated negative binomial model are obtained and compared in Section 6.4 and results for binomial data based on the zero-inflated over-dispersed generalized linear model with logit link and the zero-inflated beta-binomial model are obtained and compared in Section 6.5. Results of some simulation experiments are presented in Section 6.6 and two illustrative examples to choose an appropriate model are presented in Section 6.7.

6.2 The zero-inflated over-dispersed generalized linear model

Consider the natural exponential family distribution with probability density function

$$f(y;\theta) = \exp\{a(\theta)y - g(\theta) + c(y)\} \tag{6.2.1}$$

where y represents the response variable and θ is an unknown parameter on which the distribution of y depends. This family includes both the Poisson and binomial distribution.

Departure from the generalized linear model (Poisson or binomial) may be because of having a lot of zeros in the data or because the data are extra-dispersed or because of the presence of zero-inflation as well as over-dispersion in the data.

The exponential family distribution with zero-inflation has probability density

$$f_1(y;\theta) = \begin{cases} \omega + (1-\omega)f(0;\theta) & \text{if } y = 0 \\ (1-\omega)f(y;\theta) & \text{if } y > 0 \end{cases} \tag{6.2.2}$$

where ω is the zero-inflation(deflation) parameter which can take negative values provided

$$\omega \geq -\frac{f(0;\theta)}{1-f(0;\theta)}$$

with equality for left truncation. Note that a zero-inflated model will have $\omega > 0$ and a zero-deflated model will have $\omega < 0$ (see Section 5.2).

Now, suppose that for given θ^*, y has the exponential family model with probability density function

$$f(y;\theta^*) = \exp\{a(\theta^*)y - g(\theta^*) + c(y)\}$$

where θ^* is continuous independent random variate with

$$E(\theta^*) = \theta(x;\beta), \quad \text{var}(\theta^*) = \tau b(\theta) > 0, \quad \alpha_r = E(\theta^* - \theta)^r,$$

where β is the $p \times 1$ vector of regression parameters and τ is the over-dispersed parameter. Then following Cox (1983), Chesher (1984) and Dean (1992) the probability function of the over-dispersed exponential family model is:

$$f_2(y;\theta,\tau) = f(y;\theta)\left\{1 + \sum_{r=2}^{\infty}\frac{\alpha_r}{r!}D_r(y;\theta)\right\} \tag{6.2.3}$$

where

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

Further, for small τ, we assume that $\alpha_r = o(\tau)$ for $r \geq 3$ and

$$\begin{aligned} f_2(y;\theta,\tau) &= f(y;\theta)\{1 + \frac{\alpha_2}{2!}D_2(y,\theta)\} \\ &= f(y;\theta)\{1 + \frac{\tau}{2}b(\theta)D_2(y,\theta)\}. \end{aligned}$$

The zero-inflated over-dispersed exponential family model then can be written as

$$f_3(y;\theta,\tau,\omega) = \begin{cases} \omega + (1-\omega)f_2(0;\theta,\tau) & \text{if } y = 0 \\ (1-\omega)f_2(y;\theta,\tau) & \text{if } y > 0. \end{cases} \tag{6.2.4}$$

Obviously, the above model is a generalization of models (6.2.1), (6.2.2) and (6.2.3).

6.3 Model selection in the zero-inflated over-dispersed generalized linear model

For discrete data in the form of counts or proportions one of the following discrete generalized linear models may fit the data: (i) a generalized linear (a Poisson or a binomial) model; (ii) a zero-inflated generalized linear model; (iii) a over-dispersed generalized linear model; (iv) zero-inflated over-dispersed generalized linear model. Using the over-dispersed exponential family model (6.2.3) Dean (1992) develops score tests to detect over-dispersion in the generalized linear model. She then obtains score tests to detect over-dispersion in Poisson and binomial data separately as special cases of the results she obtains for the generalized linear model. Broek (1995) obtains a score test to test whether the number of zeros is too large for a Poisson distribution to fit the data well. Using the zero-inflated generalized linear model (6.2.2) score tests to detect zero-inflation in generalized linear model are derived and score tests for zero-inflation in Poisson and binomial data separately as special cases of the results are obtained in Chapter 5. For the Poisson data the results obtained in Chapter 5 are identical to those obtained by Broek (1995).

In this section, we derive the score test statistics for selection of a model in the zero-inflated over-dispersed generalized linear model. Specifically we derive (i) score tests for over-dispersion in presence of zero-inflation, (ii) score tests for zero-inflation in presence of over-dispersion,and (iii) score tests simultaneously for zero-inflation and over-dispersion.

Let $Y_i, i = 1, ..., n$, be a sample of independent observations from (6.2.4) with θ_i a function of $p \times 1$ vector of covariates X_i and a vector of regression parameters β; that is, $\theta_i = \theta_i(X_i; \beta), i = 1, ..., n$. From (6.2.4) the likelihood function is

$$L(\omega, \tau, \theta; y) = \prod_{i=1}^{n} \{(\omega + (1-\omega) f_2(0; \theta, \tau)) I_{\{y_i=0\}} + (1-\omega) f_2(y_i; \theta, \tau) I_{\{y_i>0\}}\}.$$

Writing $\gamma = \omega/(1-\omega)$ the log likelihood $l = l(\gamma, \tau, \theta; y)$ can be written as

$$\begin{aligned}
l(\gamma,\tau,\theta;y) &= \sum_{i=1}^{n} l_i(\gamma,\tau,\theta_i;y_i) \\
&= \sum_{i=1}^{n}\{-\log(1+\gamma) + I_{\{y_i=0\}}\log(\gamma + f_2(0;\theta,\tau)) + I_{\{y_i>0\}}\log f_2(y_i;\theta,\tau)\} \\
&= \sum_{i=1}^{n}\{-\log(1+\gamma) + I_{\{y_i=0\}}\log(\gamma + f(0;\theta_i)\{1+\sum_{r=2}^{\infty}\frac{\alpha_r}{r!}D_r(0;\theta_i)\}) \\
&\quad + I_{\{y_i>0\}}(a(\theta_i)y_i - g(\theta_i) + c(y_i) + \log\{1+\sum_{r=2}^{\infty}\frac{\alpha_r}{r!}D_r(y_i;\theta_i)\})\}.
\end{aligned}$$

Now, define the parameter vector $\delta = (\beta', \gamma, \tau)'$. Partition $\delta = (\delta_1', \delta_2')'$. Suppose we want to test $H_0 : \delta_2 = 0$ against $H_A : \delta_2 > 0$. The dimension of the parameter vector δ_2 will depend on the null hypothesis to be tested. For example, for testing $H_0 : \tau = 0, \delta_1 = (\beta', \gamma)', \delta_2 = \tau$ and the dimension of δ_2 is 1. Similarly, for testing $H_0 : (\tau, \gamma) = (0,0), \delta_1 = \beta, \delta_2 = (\tau, \gamma)$ and the dimension of δ_2 is 2. Further, define the likelihood score $S = \frac{\partial l}{\partial \delta_2}|_{\delta_2=0}$ and the expected mixed second partial derivative matrices, $I_{11} = E(-\frac{\partial^2 l}{\partial\delta_1\partial\delta_1'}|_{\delta_2=0}), I_{12} = E(-\frac{\partial^2 l}{\partial\delta_1\partial\delta_2'}|_{\delta_2=0})$ and $I_{22} = E(-\frac{\partial^2 l}{\partial\delta_2\partial\delta_2'}|_{\delta_2=0})$. Then, under some conditions for the application of the central limit theorem to score components and the regularity conditions of maximum likelihood estimates, the score test statistic for testing $H_0 : \delta_2 = 0$ is

$$T = \hat{S}'(\hat{I}_{22} - \hat{I}_{12}'\hat{I}_{11}^{-1}\hat{I}_{12})^{-1}\hat{S},$$

which, asymptotically, has a chi-square distribution with d degrees of freedom , where d is dimension of δ_2, $\hat{S} = S(\hat{\delta}_1), \hat{I}_{11} = I_{11}(\hat{\delta}_1), \hat{I}_{12} = I_{12}(\hat{\delta}_1)$, $\hat{I}_{22} = I_{22}(\hat{\delta}_1)$ and $\hat{\delta}_1$ is the maximum likelihood estimate of δ_1 under the null hypothesis.

We now give the score test statistics for the three null hypotheses $H_0 : \tau = 0, H_0 : \gamma = 0$ and $H_0 : (\tau, \gamma) = (0,0)$ in Theorem 6.3.1, Theorem 6.3.2 and Theorem 6.3.3 respectively. The derivations are given in Appendix A. In what follows, the dependence on θ_i of the functions $\mu(\theta_i), \sigma^2(\theta_i), a(\theta_i), b(\theta_i), g(\theta_i)$ and $D_r(\theta_i)$ will be suppressed for simplicity of notation. For convenience, we replace the $f_2(0;\theta_i), a(\theta_i), g(\theta_i), b(\theta_i)$ and $D_2(y_i;\theta_i)$ with f_0, a, g, b and D_2, respectively, in the following derivation.

Theorem 6.3.1 *Let* $\mathbf{1}$ *be an* $n\times 1$ *unit vector,* U *an* $n\times p$ *matrix with ir-element* $\frac{\partial\theta_i}{\partial\beta_r}$*,* W_1^τ, W_2^τ *and* W_3^τ *diagonal matrices with ith elements* $W_{1i}^\tau = g'' - a''Ey_i - \frac{f_0\gamma}{(\gamma+f_0)(1+\gamma)}g'^2 - \frac{\gamma}{1+\gamma}g''$*,* $W_{2i}^\tau = -\frac{f_0}{(\gamma+f_0)(1+\gamma)}g'$ *and* $W_{3i}^\tau = [\frac{1}{2}g'bD_2\frac{f_0\gamma}{(\gamma+f_0)(1+\gamma)} + \frac{1}{2}(bD_2)'\frac{\gamma}{1+\gamma}]|_{y_i=0} - \frac{1}{2}E[(bD_2)']$ *respectively. Further, let* $S_\tau = \sum_{i=1}^n[-\frac{\gamma I_{\{y_i=0\}}}{\gamma+f_0} + \frac{1}{2}bD_2]$*,* $I_{\tau\tau}^\tau = \sum_{i=1}^n[\frac{1}{4}E(bD_2)^2 - \frac{(\gamma^2+2f_0\gamma)}{(\gamma+f_0)(1+\gamma)}(\frac{1}{2}bD_2)^2|_{y_i=0}]$*,* $I_{\gamma\gamma}^\tau = \sum_{i=1}^n \frac{1-f_0}{(1+\gamma)^2(\gamma+f_0)}|_{y_i=0}$ *and* $I_{\gamma\tau}^\tau = \sum_{i=1}^n \frac{\frac{1}{2}bD_2 f_0}{(\gamma+f_0)(1+\gamma)}|_{y_i=0}$*. Then the score test statistic for over-dispersion in the over-dispersed zero-inflated generalized linear model is*

$$T_1 = \hat{S}_\tau^2/\hat{V}_\tau$$

with $\hat{V}_\tau = V_\tau(\hat{\theta}_1, ..., \hat{\theta}_n; \hat{\gamma})$*, which has an asymptotic* χ^2 *distribution with one degree of freedom, where*

$$\begin{aligned} V_\tau &= I_{\tau\tau} - \mathbf{1}^T W_3 U(U^T W_1 U)^{-1} U^T W_3 \mathbf{1} - \\ &\quad (I_{\gamma\tau} - (\mathbf{1}^T W_3 U)(U^T W_1 U)^{-1}(U^T W_2 \mathbf{1}))^2 (I_{\gamma\gamma} - \mathbf{1}^T W_2 U(U^T W_1 U)^{-1} U^T W_2 \mathbf{1})^{-1}, \end{aligned}$$

and $\hat{\gamma}, \hat{\theta}_i, i = 1, 2, ..., n$ *are the maximum likelihood estimates of* $\gamma, \theta_i, i = 1, 2, ..., n$ *respectively under the null hypothesis* $H_0 : \tau = 0$*.*

Theorem 6.3.2 *Let* $\mathbf{1}$ *be an* $n \times 1$ *unit vector,* U *an* $n \times p$ *matrix with ir-element* $\frac{\partial\theta_i}{\partial\beta_r}$*,* W_1^γ, W_2^γ *and* W_3^γ *diagonal matrices with ith elements* $W_{1i}^\gamma = g'' - a''Ey_i - E\{\frac{\partial^2}{\partial\theta_i^2}\log(1 + \sum_{r=2}^\infty \frac{\alpha_r D_r}{r!})\}$*,* $W_{2i}^\gamma = -E\{\frac{\partial^2}{\partial\theta_i\partial\tau}\log(1+\sum_{r=2}^\infty \frac{\alpha_r D_r}{r!})\}$*,* $W_{3i}^\gamma = -g' + \{\frac{\partial}{\partial\theta_i}\log(1+\sum_{r=2}^\infty \frac{\alpha_r D_r}{r!})\}|_{y_i=0}$ *respectively. Further, let* $S_\gamma = \sum_{i=1}^n(\frac{I_{\{y_i=0\}}}{f_2(0;\theta_i,\tau)} - 1)$*,* $I_{\tau\tau}^\gamma = \sum_{i=1}^n E\{-\frac{\partial^2}{\partial\tau^2}\log(1+\sum_{r=2}^\infty \frac{\alpha_r D_r}{r!})\}$*,* $I_{\tau\gamma}^\gamma = \sum_{i=1}^n \frac{\partial}{\partial\tau}\log(1 + \sum_{r=2}^\infty \frac{\alpha_r D_r}{r!})|_{y_i=0}$ *and* $I_{\gamma\gamma}^\gamma = \sum_{i=1}^n(1/f_2(0;\theta_i,\tau) - 1)$*. Then the score test statistic for zero-inflation in the over-dispersed zero-inflated generalized linear model is*

$$T_2 = \hat{S}_\gamma^2/\hat{V}_\gamma$$

with $\hat{V}_\gamma = V_\gamma(\hat{\theta}_1, ..., \hat{\theta}_n; \hat{\tau})$*, which has an asymptotic* χ^2 *distribution with one degree of freedom, where*

$$\begin{aligned} V_\gamma &= I_{\gamma\gamma} - \mathbf{1}^T W_3 U(U^T W_1 U)^{-1} U^T W_3 \mathbf{1} - \\ &\quad (I_{\tau\gamma} - (\mathbf{1}^T W_3 U)(U^T W_1 U)^{-1}(U^T W_2 \mathbf{1}))^2 (I_{\tau\tau} - \mathbf{1}^T W_2 U(U^T W_1 U)^{-1} U^T W_2 \mathbf{1})^{-1}, \end{aligned}$$

and $\hat{\tau}, \hat{\theta}_i, i = 1, 2, ..., n$ are the maximum likelihood estimates of $\tau, \theta_i, i = 1, 2, ..., n$ respectively under the null hypothesis $H_0 : \gamma = 0$.

Theorem 6.3.3 *Let $\mathbf{1}$ be an $n \times 1$ unit vector, U an $n \times p$ matrix with ir-element $\frac{\partial \theta_i}{\partial \beta_r}$, W_1, W_2 and W_3 diagonal matrices with ith elements $W_{1i} = g_i'' - a_i'' E y_i, W_{2i} = -g', W_{3i} = \frac{1}{2} b_i [g_i' \{(h_i')^2 - h_i''\} - 2 g_i'' h_i' + g_i''']$ respectively. Further, let $S_1 = \sum_{i=1}^n (\frac{I_{\{y_i=0\}}}{f(0;\theta_i)} - 1), S_2 = \sum_{i=1}^n \frac{1}{2} b_i (a_i')^2 \{(y_i - \mu_i)^2 - (a_i')^{-2} (g_i'' - a_i'' y_i)\}, I_{\gamma\gamma} = \sum_{i=1}^n (1/f(0;\theta_i) - 1), I_{\gamma\tau} = \sum_{i=1}^n \frac{1}{2} b_i [(g_i')^2 - g_i'']$ and $I_{\tau\tau} = \sum_{i=1}^n \{\frac{1}{4} b_i^2 [g_i' \{5 h_i' h_i'' - 3(h_i')^3 - h_i'''\} + 2(h_i' g_i' - g_i'')^2 + g_i'' \{6(h_i')^2 - 4h_i''\} - 4 h_i''' h_i' + g_i'''']\}$. Then the score test statistic for zero-inflation and over-dispersion in the over-dispersed zero-inflated generalized linear model is*

$$T_3 = \frac{\hat{V}_{22} \hat{S}_1^2 + \hat{V}_{11} \hat{S}_2^2 - 2 \hat{V}_{12} \hat{S}_1 \hat{S}_2}{\hat{V}_{22} \hat{V}_{11} - \hat{V}_{12}^2},$$

which has an asymptotic χ^2 distribution with two degrees of freedom, where $V_{11} = I_{\gamma\gamma} - \mathbf{1}^T W_2 U (U^T W_1 U)^{-1} U^T W_2 \mathbf{1}, V_{12} = I_{\gamma\tau} - \mathbf{1}^T W_2 U (U^T W_1 U)^{-1} U^T W_3 \mathbf{1}, V_{22} = I_{\tau\tau} - \mathbf{1}^T W_3 U (U^T W_1 U)^{-1} U^T W_3 \mathbf{1}, \hat{S}_1 = S_1(\hat{\theta}), \hat{S}_2 = S_2(\hat{\theta}), \hat{V}_{11} = V_{11}(\hat{\theta}), \hat{V}_{22} = V_{22}(\hat{\theta})$ and $\hat{V}_{12} = V_{12}(\hat{\theta})$, $\hat{\theta}$ is the maximum likelihood estimate of the vector of parameters $\theta = (\theta_1, ..., \theta_n)$ under the null hypothesis $H_0 : (\gamma, \tau) = 0$.

6.4 Score test for Poisson data

For zero-inflated over-dispersed Poisson data, we obtain two sets of score tests. One of these sets is based on the results given in Theorems 6.3.1 - 6.3.3 using the over-dispersed zero-inflated generalized linear model with $\theta = \log \mu = X\beta$, where X is a $n \times p$ matrix of covariates and β is a $p \times 1$ vector of regression parameters, $a(\theta) = \theta, g(\theta) = e^\theta$ and $h(\theta) = 0$. The other set is based on the zero-inflated negative binomial model. Let Y follow a negative binomial distribution with mean m and dispersion parameter c, denoted by $Y \sim NB(m, c)$.

Then, the probability function of Y is

$$P(Y=y) = \frac{\Gamma(y+c^{-1})}{y!\Gamma(c^{-1})}\left(\frac{c\mu}{1+c\mu}\right)^{y}\left(\frac{1}{1+c\mu}\right)^{c^{-1}}$$

for $y = 0, 1, ..., 0 < \mu < \infty, 0 < c < \infty$. Here $E(Y) = \mu$ and $\text{var}(Y) = \mu + c\mu^2$. Then the zero-inflated negative binomial model has the following probability function

$$\begin{aligned} P(Y=0) &= \omega + (1-\omega)(\frac{1}{1+c\mu})^{c^{-1}}, \\ P(Y=y) &= (1-\omega)\frac{\Gamma(y+c^{-1})}{y!\Gamma(c^{-1})}(\frac{c\mu}{1+c\mu})^{y}(\frac{1}{1+c\mu})^{c^{-1}}, \text{ for } y > 0. \end{aligned}$$

6.4.1 Testing for over-dispersion

Here the null hypothesis to be tested is $H_0 : \tau = 0$. Using the results in Theorem 6.3.1, the score test statistic for testing for over-dispersion is

$$Z_1 = \frac{(\sum_{i=1}^{n} \frac{1}{2}((y_i - \hat{\mu}_i)^2 - \hat{\mu}_i) - \frac{\hat{\gamma}(\hat{\mu}_i^2 - \hat{\mu}_i) I_{\{y_i=0\}}}{2(\hat{\gamma}+e^{-\mu_i})})^2}{\hat{V}_\tau},$$

where $\hat{\mu}_i = \exp(\sum_{j=1} X_{ij}\hat{\beta}_j)$ and $\hat{\beta}_j$ and $\hat{\gamma}$ are the maximum likelihood estimates of β_j and γ under the null hypothesis and $\hat{V}_\tau = V_\tau(\hat{\mu}, \hat{\gamma})$ with

$$\begin{aligned} V_\tau &= I_{\tau\tau}^{\tau} - \mathbf{1}^T W_3^\tau X(X^T W_1^\tau X)^{-1} X^T W_3^\tau \mathbf{1} - (I_{\gamma\tau}^{\tau} - \mathbf{1}^T W_3^\tau X(X^T W_1^\tau X)^{-1} X^T W_2^\tau \mathbf{1})^2 \\ &\quad /(I_{\gamma\gamma}^{\tau} - \mathbf{1}^T W_2^\tau X(X^T W_1^\tau X)^{-1} X^T W_2^\tau \mathbf{1}), \\ I_{\gamma\gamma}^{\tau} &= \sum_{i=1}^{n} \frac{e^{\mu_i} - 1}{(1+\gamma)^2(1+\gamma e^{\mu_i})}, \\ I_{\gamma\tau}^{\tau} &= \sum_{i=1}^{n} \frac{\mu_i^2 - \mu_i}{2(1+\gamma)(1+\gamma e^{\mu_i})}, \\ I_{\tau\tau}^{\tau} &= \sum_{i=1}^{n} \left[\frac{2\mu_i^2 + \mu_i}{4(1+\gamma)} - \frac{\gamma(\mu_i^2 - \mu_i)^2}{(1+\gamma)(1+\gamma e^{\mu_i})}\right], \\ W_{1i}^{\tau} &= \frac{\mu_i}{1+\gamma} - \frac{\gamma\mu_i^2}{(1+\gamma)(1+\gamma e^{\mu_i})}, \\ W_{2i}^{\tau} &= -\frac{\mu_i}{(1+\gamma)(1+\gamma e^{\mu_i})} \end{aligned}$$

and

$$W_{3i}^{T} = \frac{\mu_i}{2(1+\gamma)} + \frac{1}{2}\frac{\gamma\mu_i(\mu_i^2-\mu_i)}{(1+\gamma)(1+\gamma e^{\mu_i})}$$

The maximum likelihood estimates $\hat{\beta}_j$ and $\hat{\gamma}$ are obtained by solving the estimating equations:

$$\sum_{i=1}^{n}\left(\frac{-1}{1+\gamma}+\frac{1}{\gamma+e^{-\mu_i}}\right)=0,$$

$$\sum_{i=1}^{n}\left(I_{\{y_i=0\}}\frac{\gamma\mu_i}{\gamma+e^{-\mu_i}}+(y_i-\mu_i)\right)X_{ij}=0, \quad \text{for } j=1,2,...,p. \tag{6.3.1}$$

Using the zero-inflated negative binomial model the null hypothesis to be tested is H_0 : $c=0$. The corresponding score test statistic is

$$Z_2 = \frac{\left\{\sum_{i=1}^{n}[\frac{1}{2}((y_i-\hat{\mu}_i)^2-y_i)-I_{\{y_i=0\}}\frac{\hat{\mu}_i^2\hat{\gamma}}{2(\hat{\gamma}+e^{-\hat{\mu}_i})}]\right\}^2}{\hat{V}_c}$$

where $\hat{\mu}_i=\exp(\sum_{j=1}^{p}X_{ij}\hat{\beta}_j)$ and $\hat{\beta}_j$ and $\hat{\gamma}$ are the maximum likelihood estimates of β_j and γ in the zero-inflated Poisson model and $\hat{V}_c=V_c(\hat{\mu},\hat{\gamma})$ with

$$\begin{aligned}
V_c &= I_{cc}-\mathbf{1}^TW_3X(X^TW_1X)^{-1}X^TW_3\mathbf{1} - \\
&\quad (I_{c\gamma}-\mathbf{1}^TW_3X(X^TW_1X)^{-1}X^TW_2\mathbf{1})^2(I_{\gamma\gamma}-\mathbf{1}^TW_2X(X^TW_1X)^{-1}X^TW_2\mathbf{1})^{-1}, \\
I_{\gamma\gamma} &= \sum_{i=1}^{n}\frac{e^{\mu_i}-1}{(1+\gamma)^2(1+\gamma e^{\mu_i})}, \\
I_{\gamma c} &= \sum_{i=1}^{n}\frac{\mu_i^2}{2(1+\gamma)(1+\gamma e^{\mu_i})}, \\
I_{cc} &= \sum_{i=1}^{n}\left\{\frac{\mu_i^2}{2(1+\gamma)}-\frac{\mu_i^4\gamma}{4(1+\gamma)(\gamma e^{\mu_i}+1)}\right\}, \\
W_{1i} &= \frac{\mu_i}{1+\gamma}-\frac{\mu_i^2\gamma}{(1+\gamma)(\gamma e^{\mu_i}+1)}, \\
W_{2i} &= \frac{-\mu_i}{(1+\gamma)(\gamma e^{\mu_i}+1)}
\end{aligned}$$

and

$$W_{3i} = \frac{\mu_i^3\gamma}{2(1+\gamma)(\gamma e^{\mu_i}+1)}.$$

Note that the estimating equations in the zero-inflated negative binomial model are same as those in the zero-inflated over-dispersed Poisson model. So the maximum likelihood estimates $\hat{\beta}_j$ and $\hat{\gamma}$ used in the statistic Z_2 are the same as those used in Z_1.

We now show that for testing over-dispersion the score test statistic (Z_1) is identical to the score test statistic (Z_2) provided that the model contains a constant, that is, $X_{i1} = 1$ for $i = 1, ..., n$.

From the estimating equation for $j = 1$ in (6.3.1) we have $\sum_{i=1}^n \hat{\mu}_i = \sum_{i=1}^n \{-I_{\{y_i=0\}} \frac{\hat{\gamma}\hat{\mu}_i}{\hat{\gamma}+e^{-\hat{\mu}_i}} + y_i\}$. Using this in the nominator of Z_1, we have

$$\sum_{i=1}^n \frac{1}{2}((y_i - \hat{\mu}_i)^2 - \hat{\mu}_i) - \frac{\hat{\gamma}(\hat{\mu}_i^2 - \hat{\mu}_i)I_{\{y_i=0\}}}{2(\hat{\gamma} + e^{-\hat{\mu}_i})} = \sum_{i=1}^n \frac{1}{2}((y_i - \hat{\mu}_i)^2 - y_i) - \frac{\hat{\mu}_i^2\hat{\gamma}I_{\{y_i=0\}}}{2(\hat{\gamma} + e^{-\hat{\mu}_i})}.$$

Thus, the quantity in the nominator of Z_1 is identical to that in Z_2. Further we note that $W_1^\tau = W_1, W_2^\tau = W_2, W_3^\tau = W_3 + \frac{1}{2}W_1, I_{\gamma\gamma}^\tau = I_{\gamma\gamma}, I_{\gamma\tau}^\tau = I_{\gamma c} + \frac{1}{2}\sum W_{2i}$ and $I_{\tau\tau}^\tau = I_{cc} + \frac{1}{4}\sum W_{1i} + \sum W_{3i}$. Therefore,

$$\begin{aligned}
V_\tau &= I_{\tau\tau}^\tau - \mathbf{1}^T W_3^\tau X(X^T W_1^\tau X)^{-1} X^T W_3^\tau \mathbf{1} - \\
&\quad (I_{\gamma\tau}^\tau - \mathbf{1}^T W_3^\tau X(X^T W_1^\tau X)^{-1} X^T W_2^\tau \mathbf{1})^2 (I_{\gamma\gamma}^\tau - \mathbf{1}^T W_2^\tau X(X^T W_1^\tau X)^{-1} X^T W_2^\tau \mathbf{1})^{-1} \\
&= I_{cc} + \frac{1}{4}\sum W_{1i} + \sum W_{3i} - \mathbf{1}^T(W_3 + \frac{1}{2}W_1)X(X^T W_1 X)^{-1} X^T (W_3 + \frac{1}{2}W_1)\mathbf{1} - \\
&\quad (I_{\gamma c} + \frac{1}{2}\sum W_{2i} - \mathbf{1}^T(W_3 + \frac{1}{2}W_1)X(X^T W_1 X)^{-1} X^T W_2 \mathbf{1})^2 \\
&\quad /(I_{\gamma\gamma} - \mathbf{1}^T W_2 X(X^T W_1 X)^{-1} X^T W_2 \mathbf{1}) \\
&= I_{cc} + \frac{1}{4}\sum W_{1i} + \sum W_{3i} \\
&\quad - \mathbf{1}^T W_3 X(X^T W_1 X)^{-1} X^T W_3 \mathbf{1} - \frac{1}{2}\mathbf{1}^T W_1 X(X^T W_1 X)^{-1} X^T W_3 \mathbf{1} - \\
&\quad \frac{1}{2}\mathbf{1}^T W_3 X(X^T W_1 X)^{-1} X^T W_1 \mathbf{1} - \frac{1}{4}\mathbf{1}^T W_1 X(X^T W_1 X)^{-1} X^T W_1 \mathbf{1} \\
&\quad (I_{\gamma c} + \frac{1}{2}\sum W_{2i} - \frac{1}{2}\mathbf{1}^T W_1 X(X^T W_1 X)^{-1} X^T W_2 \mathbf{1} \\
&\quad - \mathbf{1}^T W_3 X(X^T W_1 X)^{-1} X^T W_2 \mathbf{1})^2 / (I_{\gamma\gamma} - \mathbf{1}^T W_2 X(X^T W_1 X)^{-1} X^T W_2 \mathbf{1}) \\
&= I_{cc} + \frac{1}{4}\sum W_{1i} + \sum W_{3i} - \mathbf{1}^T W_3 X(X^T W_1 X)^{-1} X^T W_3 \mathbf{1}
\end{aligned}$$

$$
\begin{aligned}
& -\frac{1}{2}\sum W_{3i} - \frac{1}{2}\sum W_{3i} - \frac{1}{4}\sum W_{1i} \\
& (I_{\gamma c} + \frac{1}{2}\sum W_{2i} - \frac{1}{2}\sum W_{2i} - \mathbf{1}^T W_3 X (X^T W_1 X)^{-1} X^T W_2 \mathbf{1})^2 \\
& /(I_{\gamma\gamma} - \mathbf{1}^T W_2 X (X^T W_1 X)^{-1} X^T W_2 \mathbf{1}) \\
= \;& I_{cc} - \mathbf{1}^T W_3 X (X^T W_1 X)^{-1} X^T W_3 \mathbf{1} - \\
& (I_{c\gamma} - \mathbf{1}^T W_3 X (X^T W_1 X)^{-1} X^T W_2 \mathbf{1})^2 (I_{\gamma\gamma} - \mathbf{1}^T W_2 X (X^T W_1 X)^{-1} X^T W_2 \mathbf{1})^{-1} \\
= \;& V_c.
\end{aligned}
$$

Thus, the score test statistic Z_2 derived using the specific over-dispersion model, namely, the negative binomial model, is the same as the score test statistic (Z_1) derived using the general over-dispersed Poisson model.

6.4.2 Testing for zero-inflation

In this subsection, for ease of computation, we assume $\alpha_r = o(\tau)$, for $r = 3, ..., \infty$ and $f_2(y;\theta,\tau) = f(y;\theta)(1+\frac{\tau}{2}b(\theta)D_2(y;\theta))$. Then, by using Theorem 6.3.2, the score test statistic for testing the hypothesis $H_0 : \gamma = 0$ in the zero-inflated over-dispersed Poisson model is

$$
Z_3 = (\sum_{i=1}^{n} \frac{I_{\{y_i=0\}} e^{\hat{\mu}_i}}{1+\frac{1}{2}\hat{\tau}((y_i-\hat{\mu}_i)^2-\hat{\mu}_i)} - 1)^2/\hat{V}_\gamma,
$$

where $\hat{\mu}_i = \exp(\sum_{j=1} X_{ij}\hat{\beta}_j)$ and $\hat{\beta}_j$ and $\hat{\tau}$ are the maximum likelihood estimates of β_j and τ under the null hypothesis and $\hat{V}_\gamma = V_\gamma(\hat{\mu}, \hat{\tau})$ with

$$
\begin{aligned}
V_\gamma \;=\;& I^\gamma_{\gamma\gamma} - \mathbf{1}^T W^\gamma_3 U (U^T W^\gamma_1 U)^{-1} U^T W^\gamma_3 \mathbf{1} - \\
& (I^\gamma_{\tau\gamma} - \mathbf{1}^T W^\gamma_3 U (U^T W^\gamma_1 U)^{-1} U^T W^\gamma_2 \mathbf{1})^2 (I_{\tau\tau} - \mathbf{1}^T W^\gamma_2 U (U^T W^\gamma_1 U)^{-1} U^T W^\gamma_2 \mathbf{1})^{-1}, \\
I^\gamma_{\tau\tau} \;=\;& \frac{1}{4}\sum_{i=1}^{n}(2\mu_i^2 + \mu_i), \\
I^\gamma_{\gamma\tau} \;=\;& \frac{1}{2}\sum_{i=1}^{n}(\mu_i^2 - \mu_i), \\
I^\gamma_{\gamma\gamma} \;=\;& \sum_{i=1}^{n}(\frac{e^{\mu_i}}{1+\frac{1}{2}\tau(\mu_i^2-\mu_i)} - 1),
\end{aligned}
$$

$$
\begin{aligned}
W_{1i}^{\gamma} &= \mu_i - \frac{1}{2}\tau(2\mu_i^2 - \mu_i), \\
W_{2i}^{\gamma} &= \frac{1}{2}\mu_i - \frac{1}{2}\tau\mu_i^2
\end{aligned}
$$

and

$$
W_{3i}^{\gamma} = -\mu_i + \frac{1}{2}\tau(2\mu_i^2 - \mu_i)
$$

The maximum likelihood estimates $\hat{\beta}$ and $\hat{\tau}$ are obtained by solving the estimating equations:

$$
\begin{aligned}
\sum_{i=1}^{n}\left(\frac{\frac{1}{2}\tau[(y_i-\mu_i)^2-\mu_i]}{1+\frac{1}{2}\tau[(y_i-\mu_i)^2-\mu_i]}\right) &= 0, \\
\sum_{i=1}^{n}\left((y_i-\mu_i)+\frac{\frac{1}{2}\tau(2(y_i-\mu_i)(-\mu_i)-\mu_i)}{1+\frac{1}{2}\tau((y_i-\mu_i)^2-\mu_i)}\right)X_{ij} &= 0 \text{ for } j=1,2,\dots,p.
\end{aligned}
$$

Further, using the zero-inflated negative binomial model, we obtain the score test statistic for testing the hypothesis $H_0 : \gamma = 0$ as:

$$
Z_4 = \frac{(\sum_{i=1}^{n} I_{\{y_i=0\}}(1+\hat{c}\hat{\mu}_i)^{\hat{c}^{-1}}-1)^2}{V_\gamma}
$$

where $\hat{\mu}_i = \exp(\sum_{j=1} X_{ij}\hat{\beta}_j)$ and $\hat{\beta}_j$ and $\hat{c}$ are the maximum likelihood estimates of β_j and c under the negative binomial model and $\hat{V}_\gamma = V_\gamma(\hat{\mu}, \hat{c})$ with

$$
\begin{aligned}
V_\gamma &= I_{\gamma\gamma} - \mathbf{1}^T W_3 U(U^T W_1 U)^{-1} U^T W_3 \mathbf{1} - \\
&\quad (I_{\gamma c} - \mathbf{1}^T W_3 U(U^T W_1 U)^{-1} U^T W_2 \mathbf{1})^2 (I_{cc} - \mathbf{1}^T W_2 U(U^T W_1 U)^{-1} U^T W_2 \mathbf{1})^{-1}, \\
I_{\gamma\gamma} &= \sum_{i=1}^{n}[(1+c\mu_i)^{c^{-1}} - 1], \\
I_{\gamma c} &= \sum_{i=1}^{n}[\frac{1}{c^2}\log(1+c\mu_i) - \frac{\mu_i}{c(1+c\mu_i)}], \\
I_{cc} &= \sum_{i=1}^{n}\{\sum_{l=1}^{\infty}\frac{(l-1)^2 P(Y_i \geq l)}{(1+(l-1)c)^2} - \frac{\mu_i^3}{(1+c\mu_i)^2} \\
&\quad - \frac{2}{c^3}\log(1+c\mu_i) - \frac{2\mu_i}{c^2(1+c\mu_i)} - \frac{\mu_i^2}{c(1+c\mu_i)^2}\}, \\
W_{1i} &= \frac{\mu_i}{(1+c\mu_i)}, \\
W_{2i} &= 0
\end{aligned}
$$

and

$$W_{3i} = -\frac{\mu_i}{(1+c\mu_i)},$$

where $Y_i \sim NB(\mu_i, c)$ for $i = 1, 2, ..., n$. Note that the maximum likelihood estimates $\hat{\beta}$ and $\hat{c}$ are obtained by the following estimating equations:

$$\sum_{i=1}^{n}\sum_{l=1}^{y_i}\left(\frac{l-1}{1+(l-1)c}\right) - \sum_{i=1}^{n}\left(\frac{y_i\mu_i}{1+c\mu_i} - \frac{1}{c^2}\log(1+c\mu_i) + \frac{\mu_i}{c(1+c\mu_i)}\right) = 0,$$

$$\sum_{i=1}^{n}\left(\frac{y_i-\mu_i}{1+c\mu_i}\right)X_{ij} = 0 \text{ for } j = 1, 2, ..., p.$$

Theoretically, the two test statistics are not same because the exact expressions of $I_{\tau\tau}$, $I_{\gamma\tau}$, $I_{\gamma\gamma}$, W_{1i}, W_{2i} and W_{3i} in the zero-inflated generalized linear model can not be obtained. Note that the statistic Z_4 has a simple form and we show by simulations in Section 6.6 it holds level even when the over-dispersion model is not negative binomial. So, we recommend its use in practice for testing for zero-inflation in the presence of over-dispersion.

6.4.3 Testing for over-dispersion and zero-inflation

Using the results in Theorem 6.3.3 the score test statistic for testing the hypothesis H_0 : $(\gamma, \tau) = 0$ is :

$$Z_5 = \frac{(\sum(\frac{I_{\{y_i=0\}}}{e^{-\hat{\mu}_i}} - 1 - \frac{1}{2}(y_i - \hat{\mu}_i)^2 + \frac{1}{2}\hat{\mu}_i))^2}{(\sum e^{\hat{\mu}_i} - 1 - \hat{\mu}_i - \frac{1}{2}\hat{\mu}_i^2)} + \frac{(\sum((y_i-\hat{\mu}_i)^2 - \hat{\mu}_i))^2}{2\sum \hat{\mu}_i^2},$$

where $\hat{\mu}_i$ is the maximum likelihood estimate of the parameter μ_i under the null hypothesis $(\gamma, \tau) = 0$.

Further, using the zero-inflated negative binomial model, the score test statistic for testing the hypothesis $H_0 : (\gamma, c) = 0$ is:

$$Z_6 = \frac{(\sum(\frac{I_{\{y_i=0\}}}{e^{-\hat{\mu}_i}} - 1 - \frac{1}{2}(y_i - \hat{\mu}_i)^2 + \frac{1}{2}y_i))^2}{(\sum e^{\hat{\mu}_i} - 1 - \hat{\mu}_i - \frac{1}{2}\hat{\mu}_i^2)} + \frac{(\sum((y_i-\hat{\mu}_i)^2 - y_i))^2}{2\sum \hat{\mu}_i^2},$$

where $\hat{\mu_i}$ is the maximum likelihood estimate of the parameter μ_i under the null hypothesis $(\gamma, c) = 0$.

Under the null hypothesis of either $(\gamma, \tau) = 0$ or $(\gamma, c) = 0$ we have the same estimating equations for the μ_i's. Using these estimating equations we can see that $\sum \hat{\mu}_i = \sum y_i$. Thus the two statistics Z_5 and Z_6 are identical.

Also from the above, it is easily seen that the test statistic includes two terms: the second term is the test statistic to test over-dispersion (see Dean (1992)) and the first term is related to a statistic to test zero-inflation. There seems to be some compound effect between zero-inflation and over-dispersion.

6.5 Score test for binomial data

For zero-inflated over-dispersed binomial data, we also obtain two sets of score tests. One of these sets is based on the results given in Theorems 6.3.1 - 6.3.3 using the over-dispersed zero-inflated generalized linear model with $\theta = \pi, a(\theta) = \log\{\pi/(1-\pi)\}, g(\theta) = -m\log(1-\theta), b(\theta) = \theta(1-\theta)$ and $h(\theta) = -\log\theta - \log(1-\theta)$. The other set is based on the zero-inflated beta-binomial model. Note that Y follows a beta-binomial distribution with mean π and dispersion parameter θ, denoted by $Y \sim BB(m, \pi, \theta)$ if Y has the following probability function

$$P(Y = y) = \binom{m}{y} \frac{\prod_{r=0}^{y-1}(\pi + r\theta)\prod_{r=0}^{m-y-1}(1-\pi+r\theta)}{\prod_{r=0}^{m-1}(1+r\theta)},$$

for $y = 0, 1, ..., m, 0 \le \pi \le 1$ and $\theta \ge max[-\pi/(m-1), -(1-\pi)/(m-1)]$ (Prentice, 1986). The mean and variance of Y are $m\pi$ and $m\pi(1-\pi)[1+(m-1)\phi]$, where $\phi = \theta/(1+\theta)$. Then, the zero-inflated beta-binomial model follows the following probability function

$$\begin{aligned} P(Y=0) &= \omega + (1-\omega)\frac{\prod_{r=0}^{m-1}(1+r\theta-\pi)}{\prod_{r=0}^{m-1}(1+r\theta)}, \\ P(Y=y) &= (1-\omega)\binom{m}{y}\frac{\prod_{r=0}^{y-1}(\pi+r\theta)\prod_{r=0}^{m-y-1}(1-\pi+r\theta)}{\prod_{r=0}^{m-1}(1+r\theta)} \text{ for } y = 1, ..., m. \end{aligned}$$

6.5.1 Testing for over-dispersion

Here the null hypothesis to be tested is $H_0 : \tau = 0$. Using the results in Theorem 6.3.1 the score test statistic for testing for over-dispersion is:

$$Z_7 = \frac{\left\{\sum_{i=1}^{n}\left[\frac{-I_{\{y_i=0\}}\hat{\gamma}\hat{\pi}_i m_i(m_i-1)}{2(\hat{\gamma}+(1-\hat{\pi}_i)^{m_i})(1-\hat{\pi}_i)} + \frac{((y_i-m_i\hat{\pi}_i)^2+(y_i-m_i\hat{\pi}_i)\hat{\pi}_i-y_i(1-\hat{\pi}_i))}{2\hat{\pi}_i(1-\hat{\pi}_i)}\right]\right\}^2}{\hat{V}_\tau},$$

where $\hat{\pi}_i = \exp(\sum X_{ij}\hat{\beta}_j)/(1+\exp(\sum X_{ij}\hat{\beta}_j))$ and $\hat{\beta}_j$ and $\hat{\gamma}$ are the maximum likelihood estimates of β_j and γ under the null hypothesis and $\hat{V}_\tau = V_\tau(\hat{\pi},\hat{\gamma})$ with

$$\begin{aligned}
V_\tau &= V_\tau(\pi,\gamma) = I^{\tau}_{\tau\tau} - \mathbf{1}^T W_3^{\tau} U(U^T W_1^{\tau} U)^{-1} U^T W_3^{\tau}\mathbf{1} - \\
&\quad (I^{\tau}_{\gamma\tau} - \mathbf{1}^T W_3^{\tau} U(U^T W_1^{\tau} U)^{-1} U^T W_2^{\tau}\mathbf{1})^2 (I^{\tau}_{\gamma\gamma} - \mathbf{1}^T W_2^{\tau} U(U^T W_1^{\tau} U)^{-1} U^T W_2^{\tau}\mathbf{1})^{-1},\\
I^{\tau}_{\gamma\gamma} &= \sum_{i=1}^{n}\frac{1-(1-\pi_i)^{m_i}}{(1+\gamma)^2(\gamma+(1-\pi_i)^{m_i})},\\
I^{\tau}_{\gamma\tau} &= \sum_{i=1}^{n}\left(\frac{(1-\pi_i)^{m_i-1}\pi_i m_i(m_i-1)}{2(1+\gamma)(\gamma+(1-\pi_i)^{m_i})}\right),\\
I^{\tau}_{\tau\tau} &= \sum_{i=1}^{n}\left(\frac{m_i(m_i-1)}{2(1+\gamma)} - \frac{m_i^2(m_i-1)^2}{4}\frac{\gamma\pi_i^2(1-\pi_i)^{m_i-2}}{(1+\gamma)(\gamma+(1-\pi_i)^{m_i})}\right),\\
W^{\tau}_{1i} &= \frac{m_i}{\pi_i(1-\pi_i)(1+\gamma)} - \frac{\gamma(1-\pi_i)^{m_i-2}m_i^2}{(1+\gamma)(\gamma+(1-\pi_i)^{m_i})},\\
W^{\tau}_{2i} &= -\frac{m_i\gamma(1-\pi_i)^{m_i-1}}{(1+\gamma)(\gamma+(1-\pi_i)^{m_i})}
\end{aligned}$$

and

$$W^{\tau}_{3i} = \frac{m_i(m_i-1)}{2}\frac{\pi_i m_i(1-\pi_i)^{m_i-2}\gamma}{(1+\gamma)(\gamma+(1-\pi_i)^{m_i})}.$$

The maximum likelihood estimates $\hat{\beta}_j$ and $\hat{\gamma}$ are obtained by solving the estimating equations:

$$\sum_{i=1}^{n}\left(\frac{-1}{1+\gamma} + \frac{1}{\gamma+(1-\pi_i)^{m_i}}\right) = 0,$$

$$\sum_{i=1}^{n}\left(I_{\{y_i=0\}}\frac{\gamma m_i\pi_i}{(\gamma+(1-\pi_i)^{m_i})} + (y_i - m_i\pi_i)\right)X_{ij} = 0, \quad \text{for } j = 1,2,...,p.$$

Using the zero-inflated beta-binomial model the null hypothesis to be tested is $H_0 : \theta = 0$. The corresponding score test statistic is

$$Z_8 = \frac{\{\sum_{i=1}^{n}\frac{-I_{\{y_i=0\}}\hat{\gamma}\hat{\pi}_i m_i(m_i-1)}{2(\hat{\gamma}+(1-\hat{\pi}_i)^{m_i})(1-\hat{\pi}_i)} + \frac{1}{2\hat{\pi}_i(1-\hat{\pi}_i)}[(y_i-m_i\hat{\pi}_i)^2 + \hat{\pi}_i(y_i-m_i\hat{\pi}_i) - y_i(1-\hat{\pi}_i)]\}^2}{V_\theta(\hat{\pi},\hat{\gamma})},$$

where $\pi = (\pi_1, ..., \pi_n)$ and $\hat{\pi}_i, i = 1, ..., n$, and $\hat{\gamma}$ are the maximum likelihood estimates of parameters $\pi_i, i = 1, ..., n$, and γ under the null hypothesis $\theta = 0$.

$$\begin{aligned}
V_\theta(\pi, \gamma) &= I^\theta_{\theta\theta} - \mathbf{1}^T W^\theta_3 U(U^T W^\theta_1 U)^{-1} U^T W^\theta_3 \mathbf{1} \\
&- (I^\theta_{\gamma\theta} - \mathbf{1}^T W^\theta_3 U(U^T W^\theta_1 U)^{-1} U^T W^\theta_2 \mathbf{1})^2 / \\
&\quad (I^\theta_{\gamma\gamma} - \mathbf{1}^T W^\theta_2 U(U^T W^\theta_1 U)^{-1} U^T W^\theta_2 \mathbf{1}), \\
I^\theta_{\gamma\gamma} &= = \sum_{i=1}^n \frac{1-(1-\pi_i)^{m_i}}{(1+\gamma)^2(\gamma+(1-\pi_i)^{m_i})}, \\
I^\theta_{\gamma\theta} &= \sum_{i=1}^n \frac{(1-\pi_i)^{m_i-1}\pi_i}{(1+\gamma)(\gamma+(1-\pi_i)^{m_i})} \frac{m_i(m_i-1)}{2}, \\
I^\theta_{\theta\theta} &= \sum_{i=1}^n \left(\frac{m_i(m_i-1)}{2(1+\gamma)} - \frac{m_i^2(m_i-1)^2}{4} \frac{\gamma(1-\pi_i)^{m_i-2}\pi_i^2}{(1+\gamma)(\gamma+(1-\pi_i)^{m_i})} \right), \\
W^\theta_{1i} &= \frac{m_i}{\pi_i(1-\pi_i)(1+\gamma)} - \frac{m_i^2\gamma(1-\pi_i)^{m_i-2}}{(1+\gamma)(\gamma+(1-\pi_i)^{m_i})}, \\
W^\theta_{2i} &= \frac{-m_i\gamma(1-\pi_i)^{m_i-1}}{(1+\gamma)(\gamma+(1-\pi_i)^{m_i})}
\end{aligned}$$

and

$$W^\theta_{3i} = \frac{m_i(m_i-1)\pi_i m_i \gamma(1-\pi_i)^{m_i-2}}{2(1+\gamma)(\gamma+(1-\pi_i)^{m_i})}.$$

The estimating equations for zero-inflated beta-binomial model are identical to those for zero-inflated generalized linear model. Note that $I^\tau_{\gamma\gamma}, I^\tau_{\gamma\tau}, I^\tau_{\tau\tau}, W^\tau_{1i}, W^\tau_{2i}$ and W^τ_{3i} are equal to $I^\theta_{\gamma\gamma}, I^\theta_{\gamma\theta}, I^\theta_{\theta\theta}, W^\theta_{1i}, W^\theta_{2i}$ and W^θ_{3i} respectively. So for testing over-dispersion the score test statistic Z_7 is identical to the score test statistic Z_8.

6.5.2 Testing for zero-inflation

Similar to Section 6.4.2, we assume that $\alpha_r = o(\tau)$, for $r = 3, ..., \infty$. By using Theorem 6.3.2, the test statistic for testing the hypothesis $H_0 : \gamma = 0$ in the zero-inflated over-dispersed binomial model is

$$Z_9 = \frac{\left\{\sum_{i=1}^n \left[I_{\{y_i=0\}}(1-\hat{\pi}_i)^{-m_i}(1+\frac{\hat{\tau}m_i(m_i-1)\hat{\pi}_i}{2(1-\hat{\pi}_i)})^{-1} - 1 \right] \right\}^2}{V_\gamma(\hat{\pi}, \hat{\tau})},$$

where $\hat{\pi}_i = \exp(\sum X_{ij}\hat{\beta}_j)/(1+\exp(\sum X_{ij}\hat{\beta}_j))$ and $\hat{\beta}_j$ and $\hat{\tau}$ are the maximum likelihood estimate of β_j and τ under the null hypothesis and $\hat{V}_\gamma = V_\gamma(\hat{\pi},\hat{\tau})$ with

$$\begin{aligned}
V_\gamma &= V_\gamma(p,\tau) = I^\gamma_{\gamma\gamma} - \mathbf{1}^T W^\gamma_3 U(U^T W^\gamma_1 U)^{-1} U^T W^\gamma_3 \mathbf{1} - \\
&\quad (I^\gamma_{\tau\gamma} - \mathbf{1}^T W^\gamma_3 U(U^T W^\gamma_1 U)^{-1} U^T W^\gamma_2 \mathbf{1})^2 (I_{\tau\tau} - \mathbf{1}^T W^\gamma_2 U(U^T W^\gamma_1 U)^{-1} U^T W^\gamma_2 \mathbf{1})^{-1}, \\
I^\gamma_{\tau\tau} &= \sum_{i=1}^n \frac{m_i(m_i-1)}{2}, \\
I^\gamma_{\gamma\tau} &= \sum_{i=1}^n \frac{\pi_i m_i(m_i-1)}{2(1-\pi_i)}\left(1+\frac{\tau\pi_i m_i(m_i-1)}{2(1-\pi_i)}\right)^{-1}, \\
I^\gamma_{\gamma\gamma} &= \sum_{i=1}^n \left(\frac{(1-\pi_i)^{-m_i}}{1+\frac{\tau\pi_i m_i(m_i-1)}{2(1-\pi_i)}} - 1\right), \\
W^\gamma_{1i} &= \frac{m_i}{\pi_i(1-\pi_i)} - \frac{\tau m_i(m_i-1)}{\pi_i(1-\pi_i)}, \\
W^\gamma_{2i} &= \frac{\tau(1-2\pi_i)m_i(m_i-1)}{2\pi_i(1-\pi_i)}
\end{aligned}$$

and

$$W^\gamma_{3i} = -\frac{m_i}{1-\pi_i} + \frac{\tau m_i(m_i-1)}{2(1-\pi_i)^2}$$

The maximum likelihood estimates β_j and τ are obtained by solving the estimating equations:

$$\sum_{i=1}^n \frac{\frac{\tau}{2\pi_i(1-\pi_i)}[(y_i - m_i\pi_i)^2 + \pi_i(y_i-\pi_i) - y_i(1-\pi_i))]}{1+\frac{\tau}{2\pi_i(1-\pi_i)}[(y_i - m_i\pi_i)^2 + \pi_i(y_i-\pi_i) - y_i(1-\pi_i)]} = 0,$$

$$\sum_{i=1}^n \left(\frac{\tau(1-m_i)(y_i - m_i\pi_i) + \frac{\tau}{2}(\frac{2\pi_i - 1}{\pi_i(1-\pi_i)})[(y_i - m_i\pi_i)^2 + \pi_i(y_i-\pi_i) - y_i(1-\pi_i)]}{1+\frac{\tau}{2\pi_i(1-\pi_i)}[(y_i - m_i\pi_i)^2 + \pi_i(y_i-\pi_i) - y_i(1-\pi_i)]}\right) X_{ij}$$

$$+\sum_{i=1}^n (y_i - m_i\pi_i)X_{ij} = 0 \text{ for } j = 1,2,...p.$$

Next, the test statistics for testing the hypothesis $H_0 : \gamma = 0$ in the zero-inflated beta-binomial model is

$$Z_{10} = \frac{\left\{\sum_{i=1}^n \left[\frac{\prod_{r=0}^{m_i-1}(1+r\hat{\theta})}{\prod_{r=0}^{m_i-1}(1+r\hat{\theta}-\hat{\pi}_i)} - 1\right]\right\}^2}{V_\gamma(\hat{\pi},\hat{\theta})},$$

where $\hat{\pi}_i = \exp(\sum X_{ij}\hat{\beta}_j)/(1+\exp(\sum X_{ij}\hat{\beta}_j))$ and $\hat{\beta}_j$ and $\hat{\theta}$ are the maximum likelihood estimate of β_j and θ under the null hypothesis and $\hat{V}_\gamma = V_\gamma(\hat{\pi}, \hat{\theta})$ with

$$
\begin{aligned}
V_\gamma &= V_\gamma(\pi,\theta) \\
&= I^\gamma_{\gamma\gamma} - \mathbf{1}^T W_3^\gamma U(U^T W_1^\gamma U)^{-1} U^T W_3^\gamma \mathbf{1} - \\
&\quad (I^\gamma_{\gamma\theta} - \mathbf{1}^T W_3^\gamma U(U^T W_1^\gamma U)^{-1} U^T W_2^\gamma \mathbf{1})^2 (I^\gamma_{\theta\theta} - \mathbf{1}^T W_2^\gamma U(U^T W_1^\gamma U)^{-1} U^T W_2^\gamma \mathbf{1})^{-1}, \\
I^\gamma_{\gamma\gamma} &= \sum_{i=1}^{n} \left(\frac{\prod_{r=0}^{m_i-1}(1+r\theta)}{\prod_{r=0}^{m_i-1}(1+r\theta-\pi_i)} - 1 \right), \\
I^\gamma_{\gamma\theta} &= \sum_{i=1}^{n} \sum_{r=1}^{m_i-1} \left(\frac{r}{1+r\theta-\pi_i} - \frac{r}{1+r\theta} \right), \\
I^\gamma_{\theta\theta} &= \sum_{i=1}^{n} \left[\sum_{r=0}^{m_i-1} \frac{-r^2 P(Y_i > r)}{(\pi_i + r\theta)} + \sum_{r=0}^{m_i-1} \frac{-r^2 P(Y_i < m_i - r)}{(1+r\theta-\pi_i)} + \sum_{r=1}^{m_i-1} \frac{r^2}{(1+r\theta)} \right], \\
W^\gamma_{1i} &= \left(\sum_{r=0}^{m_i-1} \frac{P(Y_i > r)}{(\pi_i + r\theta)^2} + \sum_{r=0}^{m_i-1} \frac{P(Y_i < m_i - r)}{(1+r\theta-\pi_i)^2} \right), \\
W^\gamma_{2i} &= \left(\sum_{r=0}^{m_i-1} \frac{rP(Y_i > r)}{(\pi_i + r\theta)^2} - \sum_{r=0}^{m_i-1} \frac{rP(Y_i < m_i - r)}{(1+r\theta-\pi_i)^2} \right)
\end{aligned}
$$

and

$$
W^\gamma_{3i} = \sum_{r=0}^{m_i-1} \frac{-1}{(1+r\theta-\pi_i)},
$$

where $Y_i \sim BB(m_i, \pi_i, \theta)$ for $i = 1, 2, ..., n$. The maximum likelihood estimating equations for θ and β_j are

$$
\sum_{i=1}^{n} \left(\sum_{r=1}^{y_i-1} \frac{r}{\pi_i + r\theta} + \sum_{r=1}^{m_i-y_i-1} \frac{r}{1+r\theta-\pi_i} - \sum_{r=1}^{m_i-1} \frac{r}{1+r\theta} \right) = 0,
$$

$$
\sum_{i=1}^{n} \left(\sum_{r=0}^{y_i-1} \frac{1}{\pi_i + r\theta} + \sum_{r=0}^{m_i-y_i-1} \frac{-1}{1+r\theta-\pi_i} \right) \pi_i(1-\pi_i)X_{ij} = 0 \text{ for } j = 1, 2, ..., p
$$

Also, two test statistics Z_9 and Z_{10} are not equal.

6.5.3 Testing for over-dispersion and zero-inflation

Now,the score test statistic for testing the hypothesis $H_0 : (\gamma, \tau) = 0$ is

$$
Z_{11} = \frac{\hat{V}_{TT}\hat{S}^2 + \hat{V}_{SS}\hat{T}^2 - 2\hat{V}_{ST}\hat{S}\hat{T}}{\hat{V}_{TT}\hat{V}_{SS} - \hat{V}^2_{ST}}
$$

where

$$\begin{aligned}
S(\pi) &= \sum_{i=1}^{n}(I_{\{y_i=0\}}(1-\pi_i)^{-m_i}-1),\\
T(p) &= \sum_{i=1}^{n}\frac{(y_i-m_i\pi_i)^2+\pi_i(y_i-m_i\pi_i)-y_i(1-\pi_i)}{2\pi_i(1-\pi_i)},\\
V_{SS}(p) &= \sum_{i-1}^{n}((1-\pi_i)^{-m_i}-1),\\
V_{TT}(p) &= \sum_{i=1}^{n}\frac{1}{2}m_i(m_i-1)-\mathbf{1}^T W_3 U(U^T W_1 U)^{-1}U^T W_3\mathbf{1},\\
V_{ST}(p) &= \sum_{i=1}^{n}\frac{\pi_i m_i(m_i-1)}{2(1-\pi_i)}
\end{aligned}$$

with $W_{1i}=\frac{m_i}{\pi_i(1-\pi_i)}$, $W_{3i}=-\frac{m_i}{(1-\pi_i)}$, $\hat{S}=S(\hat{\pi}), \hat{T}=T(\hat{\pi}), \hat{V}_{SS}=V_{SS}(\hat{\pi}), \hat{V}_{TT}=V_{TT}(\hat{\pi})$ and $\hat{V}_{ST}=V_{ST}(\hat{\pi})$, where $\hat{\pi}$ is the maximum likelihood estimate of the parameter $\pi=(\pi_1,...,\pi_n)$ under the null hypothesis $(\gamma,\tau)=0$.

By using the zero-inflated beta-binomial model, the score test statistic obtained for testing the hypothesis $H_0:(\gamma,\theta)=0$ is the same as the statistic Z_{11}.

6.6 Simulations

In this section we conduct some simulations to study the performance, in terms of empirical levels, of score test statistics Z_1, Z_4 and Z_5 for Poisson data. Each simulation experiment was based on 10,000 replications.

For testing for over-dispersion in the presence of zero-inflation, data are simulated from a zero-inflated Poisson distribution with zero-inflation parameter ω=0.001, 0.01 and 0.10, Poisson mean μ= 0.5, 1.0, 1.5, 2.0 and 2.5, and sample size n= 20, 50, 100, 200 and 300. However, we only present the simulation results for $\mu=1.0$ and $\mu=2.0$ in Table 6.1 as conclusions for other values of μ are similar. The results show that when n is small, the empirical levels of the statistic Z_1 show some liberal behavior and for large sample sizes it maintains the nominal level well. We also note that the performance of the test statistic Z_1

is the same for all values of the zero-inflated parameter.

For testing for zero-inflation in the presence of over-dispersion, data are simulated from a over-dispersed Poisson distribution with over-dispersed parameter $c=$ 0.01, 0.05, 0.10 and 0.50, Poisson mean μ = 0.5, 1.0, 1.5, 2.0 and 2.5, and sample size n =10, 20, 50 and 100 for test statistics Z_4. Data are first simulated by negative binomial distributions by using the above parameters. Also, the simulation results are given only for $\mu = 1.0$ and $\mu = 2.0$ in Table 6.2. From the results in Table 6.2, except for $c = 0.01$, when sample size is small the test statistic Z_4 is a little conservative. For large n and other values of c, Z_4 maintain the nominal levels well. Similarly, the performance of Z_4 does not depend on the value of over-dispersed parameter. Then, data are simulated from a lognormal mixture Poisson distribution by using the same parameters. The results presented in Table 6.3 are similar to those in Table 6.2 except in some case. For example, for $c = 0.50$ and $n = 100$ the test shows conservative behavior. This fact shows that except for very highly over-dispersed data the statistic Z_4 may be reasonably robust. That is, the test may be applicable when data come from some other over-dispersed generalized linear models.

For simultaneously testing for zero-inflation and over-dispersion, data are simulated from a Poisson distribution with Poisson mean $\mu=$ 0.5, 1.0, 1.5, 2.0, 2.5, 3.0 and 3.5 and sample size $n=$ 10, 20, 50 and 100. The results are presented in the Table 6.4. The results show that the score test statistic Z_5 for testing zero-inflation and over-dispersion maintains the nominal level.

Table 6.1: Empirical levels of score test statistic Z_1 for testing over-dispersion with no covariates when data are simulated from zero-inflated Poisson $P(\mu)$ based on 10,000 replications.

Poisson μ	α	n	$\omega = 0.001$	$\omega = 0.010$	$\omega = 0.100$
		20	0.0411	0.0179	0.0211
		50	0.0135	0.0174	0.0144
	0.01	100	0.0120	0.0117	0.0114
		200	0.0108	0.0118	0.0118
		300	0.0118	0.0109	0.0108
		20	0.0530	0.0293	0.0319
		50	0.0335	0.0372	0.0328
1.0	0.05	100	0.0391	0.0378	0.0400
		200	0.0439	0.0429	0.0437
		300	0.0485	0.0470	0.0449
		20	0.0676	0.0444	0.0440
		50	0.0691	0.0697	0.0637
	0.10	100	0.0794	0.0814	0.0792
		200	0.0898	0.0878	0.0869
		300	0.0974	0.0939	0.0913
		20	0.0172	0.0437	0.0252
		50	0.0198	0.0187	0.0142
	0.01	100	0.0100	0.0102	0.0103
		200	0.0088	0.0100	0.0101
		300	0.0093	0.0103	0.0108
		20	0.0359	0.0606	0.0445
		50	0.0485	0.0479	0.0420
2.0	0.05	100	0.0416	0.0446	0.0437
		200	0.0456	0.0465	0.0458
		300	0.0485	0.0497	0.0509
		20	0.0696	0.0909	0.0758
		50	0.0909	0.0895	0.0859
	0.10	100	0.0885	0.0920	0.0906
		200	0.0959	0.0975	0.0955
		300	0.0970	0.0970	0.1010

Table 6.2: Empirical levels of score test statistic Z_4 for testing zero-inflation with no covariates when data are simulated from negative binomial $NB(\mu, c)$ based on 10,000 replications.

Mean μ	α	n	$c = 0.01$	$c = 0.05$	$c = 0.10$	c=0.50
1.0	0.01	10	0.0019	0.0303	0.0291	0.0083
		20	0.0099	0.0207	0.0188	0.0067
		50	0.0134	0.0095	0.0091	0.0086
		100	0.0129	0.0250	0.0092	0.0090
	0.05	10	0.0144	0.0759	0.0632	0.0345
		20	0.0450	0.0720	0.0607	0.0398
		50	0.0455	0.0391	0.0444	0.0433
		100	0.0469	0.0608	0.0448	0.0489
	0.10	10	0.0505	0.1048	0.0925	0.0786
		20	0.0863	0.1172	0.1073	0.0804
		50	0.0924	0.0812	0.0856	0.0902
		100	0.0933	0.1032	0.0927	0.0910
2.0	0.01	10	0.0128	0.0082	0.0078	0.0081
		20	0.0103	0.0097	0.0078	0.0075
		50	0.0281	0.0083	0.0089	0.0094
		100	0.1223	0.0096	0.0076	0.0111
	0.05	10	0.0466	0.0338	0.0378	0.0399
		20	0.0437	0.0487	0.0480	0.0442
		50	0.0725	0.0462	0.0476	0.0466
		100	0.1973	0.0469	0.0474	0.0487
	0.10	10	0.0869	0.0764	0.0787	0.0879
		20	0.0966	0.0970	0.0951	0.0921
		50	0.1214	0.0960	0.0978	0.0957
		100	0.2519	0.0936	0.0928	0.0995

Table 6.3: Empirical levels of score test statistic Z_4 for testing zero-inflation with no covariates when data are simulated from Lognormal mixture Poisson $LMP(\mu, c)$ based on 10,000 replications.

Mean μ	α	n	$c = 0.01$	$c = 0.05$	$c = 0.10$	c=0.50
1.0	0.01	10	0.0015	0.0188	0.0240	0.0181
		20	0.0076	0.0213	0.0182	0.0100
		50	0.0113	0.0119	0.0108	0.0117
		100	0.0115	0.0487	0.0110	0.0200
	0.05	10	0.0137	0.0623	0.0568	0.0400
		20	0.0434	0.0659	0.0639	0.0445
		50	0.0417	0.0454	0.0474	0.0496
		100	0.0491	0.0864	0.0487	0.0661
	0.10	10	0.0513	0.1058	0.0886	0.0717
		20	0.0852	0.1184	0.1081	0.0865
		50	0.0855	0.0908	0.0934	0.0999
		100	0.0963	0.1339	0.0973	0.1195
2.0	0.01	10	0.0141	0.0075	0.0071	0.0083
		20	0.0111	0.0093	0.0084	0.0136
		50	0.0266	0.0097	0.0095	0.0209
		100	0.1191	0.0092	0.0094	0.0281
	0.05	10	0.0469	0.0335	0.0352	0.0411
		20	0.0516	0.0475	0.0463	0.0549
		50	0.0687	0.0456	0.0506	0.0717
		100	0.1902	0.0479	0.0521	0.0903
	0.10	10	0.0869	0.0769	0.0797	0.0921
		20	0.0995	0.1002	0.0976	0.1077
		50	0.1170	0.0970	0.1011	0.1276
		100	0.2451	0.0983	0.1017	0.1537

Table 6.4: Empirical levels of score test statistic Z_5 for testing both the zero-inflation and over-dispersion with no covariates when data are simulated from Poisson (μ) based on 10,000 replications.

n	Poisson mean μ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
10	0.5	0.0057	0.0289	0.0504
20	0.5	0.0130	0.0373	0.0640
50	0.5	0.0202	0.0469	0.0796
100	0.5	0.0194	0.0503	0.0856
200	0.5	0.0193	0.0515	0.0916
10	1.0	0.0108	0.0329	0.0667
20	1.0	0.0145	0.0401	0.0705
50	1.0	0.0167	0.0487	0.0895
100	1.0	0.0150	0.0508	0.0936
200	1.0	0.0128	0.0503	0.0946
10	1.5	0.0124	0.0349	0.0647
20	1.5	0.0130	0.0419	0.0752
50	1.5	0.0151	0.0484	0.0930
100	1.5	0.0152	0.0523	0.0966
200	1.5	0.0124	0.0475	0.0958
10	2.0	0.0123	0.0312	0.0559
20	2.0	0.0131	0.0399	0.0746
50	2.0	0.0148	0.0476	0.0924
100	2.0	0.0111	0.0487	0.0963
200	2.0	0.0098	0.0469	0.0972
10	2.5	0.0123	0.0334	0.0562
20	2.5	0.0121	0.0376	0.0680
50	2.5	0.0126	0.0469	0.0930
100	2.5	0.0123	0.0464	0.0957
200	2.5	0.0117	0.0478	0.0961
10	3.0	0.0142	0.0364	0.0590
20	3.0	0.0139	0.0387	0.0719
50	3.0	0.0123	0.0437	0.0899
100	3.0	0.0113	0.0464	0.0956
200	3.0	0.0100	0.0479	0.0930
10	3.5	0.0177	0.0413	0.0679
20	3.5	0.0163	0.0443	0.0734
50	3.5	0.0148	0.0469	0.0857
100	3.5	0.0122	0.0461	0.0941
200	3.5	0.0102	0.0470	0.0936

6.7 Two examples

Now we consider an application of the zero-inflated over-dispersed model to the Poisson and the binomial data.

Example 1: The data set are from a prospective study of dental status of school-children from Bohning, Dietz and Schlattmann (1999). The children were all 7 years of age at the beginning of the study. Dental status was measured by the decayed, missing and filled teeth (DMFT) index. Only the eight deciduous molars were considered, which implies that the smallest possible value of the DMFT index is 0 and the largest is 8. The prospective study was for a period of two years. The DMFT index was calculated at the beginning and at the end of the study. The DMFT index data at the beginning of the study are given in Table D.5. We now fit the Poisson model, over-dispersed Poisson model (negative binomial model), zero-inflated Poisson model and zero-inflated negative binomial model to the data. The results of the model fitting are given in Table 6.5. The maximized log-likelihood values indicate that the zero-inflated over-dispersed Poisson model, namely, the zero-inflated negative binomial model fits the data far better than other models.

We now use the score tests to select an appropriate model for the data. The values of the score test statistic for the goodness of a model (such as the Poisson model) against another model (such as the negative binomial model) are given in Table 6.6. The tests overwhelmingly rejects the fit of all other models in favor of the zero-inflated negative binomial model.

Example 2: We reconsider the data given in Table D.4. This data set refers to PVC counts for twelve patients one minute after administering a drug with anti arrhythmic properties (Berry, 1987). The results of the model fitting are given in Table 6.7 and the values of the score test statistic are given in Table 6.8. The maximized log-likelihood values indicate that the zero-inflated beta-binomial model fits the data slightly better than the zero-inflated binomial model and fits the data far better than other models. The values of the score test statistics and their corresponding p-values indicate that either the zero-inflated binomial or the beta-binomial model adequately fits the data. However, the value of the score test statistic for testing the fit of the zero-inflated binomial model against the zero-inflated beta-binomial model is smaller than that for testing the fit of the beta-binomial model against the zero-inflated beta-binomial model. Therefore, the zero-inflated binomial model is the model of choice for the data.

Table 6.5: Results of model fitting of the DMFT index data

Model	Estimates of parameters			log-likelihood
	μ	c	ω	
Poisson	3.3237			-1998.884
zero-inflated Poisson	4.1731		0.2035	-1761.197
negative binomial	3.3237	0.5000		-1833.862
zero-inflated over-dispersed Poisson	4.1378	0.0530	0.1967	-1756.803

Table 6.6: Results of the goodness of fit tests(score tests) for the DMFT index data, P: Poisson, NB: negative binomial, ZIP: zero-inflated Poisson, ZINB: zero-inflated negative binomial

Test	Score statistic	p-value
P vs NB	394.4432	0.0000
P vs ZIP	847.1954	0.0000
ZIP vs ZINB	7.8656	0.0053
NB vs ZINB	120.3656	0.0000
P vs ZINB	890.1677	0.0000

Table 6.7: Results of model fitting for the PVC counts data

Model	Estimates of parameters			log-likelihood
	p/π	ϕ	ω	
Binomial	0.12500			-40.6903
zero-inflated Binomial	0.38614		0.57549	-18.87347
Beta-binomial	0.12496	0.53169		-19.17807
zero- inflated Beta-binomial	0.33563	0.09169	0.55799	-18.02478

Table 6.8: Results of the goodness of fit tests(score tests) for the PVC counts data, B: binomial, BB: beta-binomial, ZIB: zero-inflated binomial, ZIBB: zero-inflated beta-binomial

Test	Score statistic	p-value
B vs BB	236.4964	0.0000
B vs ZIB	931.0414	0.0000
ZIB vs ZIBB	1.288335	0.2564
BB vs ZIBB	176.3775	0.0000
B vs ZIBB	990.3721	0.0000

Chapter 7

Discussion and future research topics

In this chapter we summarize conclusions of this thesis and discuss some further research topics.

It is well known that it is very important to assess goodness of fit of the given model because the data often exhibit variations greater than what is predicted by a simple model. Usually, the Pearson chi-square statistic or the likelihood ratio chi-square is used to test the goodness of fit of the generalized linear model when the expected cell frequencies are large. As we pointed out in Chapter 3, these two test statistics do not perform well for the case in which the data are extensive but sparse. Therefore some authors consider using the modifications of these two test statistics. In Chapter 3 and Chapter 4 we derived two modified test statistics: the modified deviance test statistic and the modified score test statistic. Comparing with Farrington (1996) and McCullagh (1986), the proposed modifications to the deviance statistic and Pearson statistic are simple to implement and have definite advantages for assessing the goodness of fit of generalized linear models when the data are sparse. The optimality properties of the modified statistics allow them to be used with both canonical and non-canonical models and, as confirmed in the beta-binomial simulations, enhance their powers to detect departures from the null distributions and maintain the nominal levels. The simulations reported in this thesis confirm the accuracy of the approximation for moderate sample sizes. Inevitably, however, the normal approximations to the tail probabilities may sometimes fail in small samples. Now we try to compare our proposed two test statistics, namely, the modified deviance statistic and the modified score statistic. We note that the modified score test statistic has a simple form and it needs parameter estimates only under the null hypothesis. Also it is a little more powerful than Modified Pearson statistic of Farrington (1996). But from simulation results in Table 3.3 and Table 4.2, we find that the modified deviance statistic is much more powerful than the modified score statistic. In fact simulations indicate that the modified deviance statistic is most powerful among all of our test statistics. Although, simulations show that the modified deviance statistic is most powerful, a theoretical proof of this result is not available. It would be interesting to see if a theoretical proof could be obtained. Another question is whether we can obtain a more powerful test statistic than the modified deviance statistic. Note that only approximations to the first two moments of conditional distributions of the Pearson chi-square statistic and the deviance statistic are used in order to get the standardized modified Pearson chi-square statistic and the standardized modified deviance statistic. Also, we only use the Edgeworth expansion with one correction term to get the tail probabilities:

$$P(Z \geq z) \doteq 1 - \Phi(z) + (z^2 - 1)\rho_3 \phi(z)/6, \tag{7.1}$$

where $\Phi(z)$ and $\phi(z)$ are standard normal distribution function and density function, ρ_3 is the approximate standardized conditional third moment. Naturally, we would like to ask if we can use the Edgeworth expansion with more correction terms to get the tail probabilities or p-values. That is, can we use the improved Edgeworth approximation of the modified Pearson statistic or the modified deviance statistic as a goodness-of-fit test statistic? The problem then is to obtain the p-values by using the formula similar to (7.1) or following Edgeworth expansion with more correction terms

$$\begin{aligned} \phi(x,\lambda)[1 &+ \kappa^{r,s,t} h_{rst}(x,\lambda)/(6\sqrt{n}) \\ &+ \{3\kappa^{r,s,t,u} h_{rstu}(x,\lambda) + \kappa^{r,s,t}\kappa^{u,v,w} h_{r\ldots w}(x,\lambda)\}/(72n)] + O(n^{-3/2}). \end{aligned} \tag{7.2}$$

The difficulty lies in obtaining a formula for the approximate conditional fourth cumulant of the modified Pearson statistic or the modified deviance statistic. Following the McCullagh (1984, 1987) approximations to the fourth unconditional cumulants $\kappa_4(D^*)$ and $\kappa_4(X_*^2)$ are obtained using the following unconditional cumulants

$$
\begin{aligned}
\kappa_1^r &= 0, \kappa_1^{r,s} = \kappa^{r,s} - \beta_i^r\beta_j^s\kappa^{i,j}, \kappa_1^{i,r} = 0,\\
\kappa_1^{r,s,t} &= \kappa^{r,s,t} - \beta_i^r\kappa^{i,s,t}[3] + \beta_i^r\beta_j^s\kappa^{i,j,t}[3] - \beta_i^r\beta_j^s\beta_k^t\kappa^{i,j,k},\\
\kappa_1^{i,r,s} &= \kappa^{i,r,s} - \beta_j^r\kappa^{i,j,s}[2] + \beta_j^r\beta_k^s\kappa^{i,j,k}, \kappa_1^{i,j,r} = \kappa^{i,j,r} - \beta_k^r\kappa^{i,j,k},\\
\kappa_1^{r,s,t,u} &= \kappa^{r,s,t,u} - \beta_i^r\kappa^{i,s,t,u}[4] + \beta_i^r\beta_j^s\kappa^{i,j,t,u}[6] - \beta_i^r\beta_j^s\beta_k^t\kappa^{i,j,k,u} + \beta_i^r\beta_j^s\beta_k^t\beta_l^u\kappa^{i,j,k,l},\\
\kappa_1^{i,r,s,t} &= \kappa^{i,r,s,t} - \beta_j^r\kappa^{i,j,s,t}[3] + \beta_j^r\beta_k^s\kappa^{i,j,k,t}[3] - \beta_j^r\beta_k^s\beta_l^t\kappa^{i,j,k,l},\\
\kappa_1^{i,j,r,s} &= \kappa^{i,j,r,s} - \beta_k^r\kappa^{i,j,k,s}[2] + \beta_k^r\beta_l^s\kappa^{i,j,k,l}, \kappa_1^{i,j,k,r} = \kappa^{i,j,k,r} - \beta_l^r\kappa^{i,j,k,l}.
\end{aligned}
$$

It is, however, not clear how to construct an appropriate formula for $\kappa_4(D^*)$ and $\kappa_4(X_*^2)$. Once the formula for $\kappa_4(D^*)$ or $\kappa_4(X_*^2)$ is obtained, we can obtain approximate conditional fourth cumulants using the following formulae

$$
\mathrm{cum}(X^r, X^s, X^t, X^u|X_{(2)}) = \kappa^{r,s,t,u}/n + \kappa_1^{r,s,i}\kappa_1^{t,u,j}(h_{ij} - h_i h_j)[3]/n
$$

or

$$
\mathrm{cum}(X^r, X^s, X^t, X^u|X_{(2)}) = n^{-1}\{\kappa^{r,s,t,u} + \kappa^{r,s,i}\kappa^{t,u,j}[3]h_{i,j}\}.
$$

Naturally, these look quite involved.

Next we discuss the issues of over-dispersion and zero-inflation. The data often exhibit these variation. The most common test method for testing over-dispersion is the score test method (See Dean 1992, Dean and Lawless 1989). As we pointed out before, the score test has many advantages. Therefore by using the score test method, we obtained the score test statistics for testing for zero-inflation in generalized linear models, testing for over-dispersion in the presence of zero-inflation, testing for zero-inflation in the presence of over-dispersion and, simultaneously testing for zero-inflation and over-dispersion in the generalized linear models. Also, we directly obtained the score test statistics for the specific models such as the negative binomial model and the beta-binomial model. In general, the test statistics obtained in this thesis maintain nominal levels and can be used to test for zero-inflation and/or over-dispersion in generalized linear models. Also we use the zero-inflated Poisson model, zero-inflated binomial model and zero-inflated over-dispersed Poisson and binomial models (zero-inflated negative binomial and zero-inflated beta-binomial models) to fit to some real life data. From the results of our examples, if there are too many zeros in the observed values the zero-inflated Poisson model obviously is better than over-dispersed Poisson (negative binomial) model and, the zero-inflated binomial model is better than the over-dispersed binomial (beta-binomial) model(see examples in Chapter 5 and Chapter 6). Further, the zero-inflated over-dispersed generalized linear model is better than the zero-inflated generalized linear model and over-dispersed generalized linear model(See examples in Chapter 6).

We note that in the over-dispersed generalized linear model the approximation to the likelihood function is used to obtain the score test statistics (see Cox 1983, Dean 1992). In Dean (1992), the approximation to the likelihood function is as follows:

$$f_2(y;\theta,\tau) = f(y;\theta)\left\{1+\sum_{r=2}^{\infty}\frac{\alpha_r}{r!}D_r(y;\theta)\right\}$$

where

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

Note that the over-dispersed generalized linear models are constructed by extending the natural exponential family with probability density function

$$f(Y;\theta) = \exp\{a(\theta)Y - g(\theta) + c(Y)\}$$

and replacing the parameter θ by the continuous random variable θ^* with finite mean and variance

$$E(\theta^*) = \theta(x;\beta), \ \ \mathrm{var}(\theta^*) = \tau b(\theta) > 0$$

and also assuming that

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

Theoretically, we can obtain only the first two terms in the Taylor expansion of likelihood function because we do not know the expression of α_r for $r \geq 3$. When testing for over-dispersion in the zero-inflated over-dispersed generalized linear model, an approximation to the likelihood function using only the first two terms is enough since the over-dispersion parameter is set to be zero. But when testing the zero-inflation in the over-dispersed generalized linear model the other terms in the Taylor expansion have effect on the score functions. Therefore, when we apply the approximation to likelihood function to obtain the score test statistic, we can not obtain the exact expression. Naturally, we ask if some good approximation to the likelihood function can be obtained.

In Chapter 4, we obtain a modified score test statistic for assessing goodness of fit. Also from the simulation results in Chapter 6, score test statistics do not maintain the nominal level well for small sample sizes. Therefore we should consider the modifications to score test statistics obtained in Chapter 6 for small sample sizes.

Another interesting question is how to get the power approximations to the goodness-of-fit tests. Drost *et al.*(1989) discuss the asymptotic power properties of the multinomial tests of fit. The asymptotic error bounds for the distributions of the goodness-of-fit tests are obtained by using the class of multinomial goodness-of-fit statistics introduced by Cressie and Read (1984). It would be of interest to investigate if similar approximations to the power of the goodness-of-fit test statistics in the present thesis can be obtained. We also are interested in assessing the goodness of fit of the generalized linear model with incomplete data or covariate measurement errors.

Table D.1 : Hepatitis A in Bulgaria

Age	Hepatitis A virus positive	Total	Age	Hepatitis A virus positive	Total
1	3	16	44	5	5
2	3	15	45	7	7
3	3	16	46	9	9
4	4	13	47	9	9
5	7	12	48	22	22
6	4	15	49	6	7
7	3	12	50	10	10
8	4	11	51	6	6
9	7	10	52	13	14
10	8	15	53	8	8
11	2	7	54	7	7
12	3	7	55	13	13
13	2	11	56	11	11
14	0	1	57	8	8
15	5	16	58	8	8
16	13	41	59	9	10
17	1	2	60	13	16
18	3	6	61	5	5
19	15	32	62	5	6
20	22	37	63	5	5
21	15	24	64	5	5
22	7	10	65	10	10
23	8	10	66	8	8
24	7	11	67	4	4
25	12	15	68	5	5
26	5	10	69	4	5
27	10	13	70	8	8
28	15	19	71	0	0
29	9	12	72	9	9
30	9	9	73	1	1
31	9	14	74	4	4
32	8	10	75	7	7
33	9	11	76	6	6
34	8	9	77	2	2
35	9	14	78	3	3
36	13	14	79	2	2
37	6	7	80	4	4
38	15	16	81	1	1
39	11	13	82	1	1
40	6	8	83	2	2
41	8	8	84	0	0
42	13	14	85	0	0
43	7	10	86	1	1

Table D.2 : Pools of progeny *Aedes aegypti* (Santo Domingo Strain) assayed for yellow fever virus

Pools (m, n_m, x_m) by larval development interval			
Virus strain A		Virus strain H	
larval development interval, 6-10 days			
(5, 1,0)	(100,22,1)	(41, 1,0)	(109, 1,0)
(7, 1,0)	(105, 1,0)	(61, 1,0)	(126, 1,0)
(47, 1,0)	(106, 1,0)	(68, 1,0)	(133, 1,1)
(48, 1,0)	(123, 1,0)	(70, 1,0)	(150, 1,1)
(51, 1,0)	(132, 1,0)	(74, 1,0)	(151, 1,0)
(65, 1,0)	(133, 1,0)	(80, 1,0)	(160, 1,0)
(76, 1,0)	(159, 1,0)	(90, 1,0)	(170, 1,1)
(83, 1,0)	(250, 1,0)	(91, 1,1)	(172, 1,0)
	(287, 1,0)	(92, 1,0)	(182, 1,1)
		(94, 1,1)	(187, 1,0)
		(105,24,1)	(194, 1,0)
		(100, 1,0)	(200,14,2)
		(106, 1,1)	(203, 1,0)
			(226, 1,1)
Larval development interval, 11-15 days			
(80, 3,1)	(116, 1,0)	(59, 1,1)	(115, 7,7)
(100,12,3)	(123, 1,0)	(82, 1,0)	(155, 2,2)
(103, 1,0)	(150, 1,1)	(120, 1,0)	(178, 1,0)
(111, 2,1)	(152, 1,0)	(136, 1,0)	(200, 4,2)
(115, 1,0)		(148, 1,0)	(203, 1,0)

Table D.3 : Multiple tumour recurrence data for patients with bladder cancer

Patient number	No. of recurrences	Initial Number	Initial Size	Patient number	No. of recurrences	Initial Number	Initial Size
1	0	1	3	20	2	8	3
2	0	1	1	21	0	1	1
3	1	8	1	22	4	1	1
4	0	1	2	23	7	6	1
5	0	1	1	24	4	3	1
6	0	1	1	25	0	3	2
7	1	2	6	26	0	1	1
8	5	5	3	27	2	1	1
9	0	5	1	28	0	1	1
10	1	1	3	29	5	6	1
11	1	5	1	30	0	1	2
12	2	1	1	31	1	1	4
13	0	1	1	32	0	1	4
14	0	1	3	33	0	3	3
15	0	1	5	34	0	1	1
16	0	1	1	35	3	4	1
17	3	1	1	36	0	3	4
18	1	1	1	37	1	2	1
19	1	2	1	38	0	1	3

Table D.4 : The PVC counts for twelve patients one minute after administrating a drug with antiarrhythmic properties

Patient number	PVCs per minute		
	Predrug(x_i)	Postdrug(y_i)	Total(m_i)
1	6	5	11
2	9	2	11
3	17	0	17
4	22	0	22
5	7	2	9
6	5	1	6
7	5	0	5
8	14	0	14
9	9	0	9
10	7	0	7
11	9	13	22
12	51	0	51

Table D.5 : The counts of the decayed, missing and filled teeth index at the beginning of the study

DMFT	Frequency	percent
0	172	21.6
1	73	9.2
2	96	12.0
3	80	10.0
4	95	11.9
5	83	10.4
6	85	10.7
7	65	8.2
8	48	6.0

Appendix A

Proofs of Theorems

A.1 Proof of Theorem 3.3.1

For convenience and to save space we use the notation of Farrington (1996). Let θ denote the true parameter value $(\beta_1, ..., \beta_p, 1)^T$ and $g = (g_1, ..., g_p, g_q)^T$. Then, by Taylor expansion of $g_r, r = 1, ..., p, q$, about θ and following steps similar to Farrington (1996) we obtain

$$\begin{aligned} E\hat{\phi} &= 1 + \frac{1}{n}\{\Delta_{21}\Delta_{11}^{-1}E(-\frac{1}{2}c^1 + \frac{1}{n}E_{11}\Delta_{11}^{-1}g^1) \\ &\quad -E(-\frac{1}{2}c^2 + \frac{1}{n}E_{21}\Delta_{11}^{-1}g^1)\} + O(n^{-2}), \end{aligned} \tag{A.1}$$

$$\text{var}(\hat{\phi}) = n^{-2}E(\Delta_{21}\Delta_{11}^{-1}\Delta_{21}\Delta_{11}^{-1} - 2\Delta_{21}\Delta_{11}^{-1}g^2 + g^2g^2) + O(n^{-2}),$$

$$\kappa_3(\hat{\phi}) = -n^3E\{(\Delta_{21}\Delta_{11}^{-1}g^1)^3 - (\Delta_{21}\Delta_{11}^{-1}g^1)^2g^2 + \Delta_{21}\Delta_{11}^{-1}g^1(g^2)^2 - (g^2)^3\} + O(\frac{1}{n^3}).$$

The terms $\Delta_{11}, \Delta_{21}, c^1, c^2, E_{11}, E_{21}, g^1$ are defined in Farrington (1996). The term g^2 is g_q defined in Section 2. Define the a_i' as $a_i' = \frac{da_i(\mu_i)}{d\mu_i}$. Now, using tensor notation and omitting summation, we have

$$\begin{aligned} \Delta_{21}\Delta_{11}^{-1} &= -\frac{1}{n}a_i\frac{\partial\mu_i}{\partial\beta_r}(-nb_{rs}) = a_ib_{rs}\frac{\partial\mu_i}{\partial\beta_r}, \\ E(c^1) &= E(c_1, ..., c_p), \\ E(c_r) &= E(Z_1^Tv_rZ_1) \end{aligned} \tag{A.2}$$

$$
\begin{aligned}
&= E\{-n^{-\frac{1}{2}} g^T \Delta^{-1} v_r \Delta^{-1}(-n^{-\frac{1}{2}} g)\} \\
&= n^{-1} E\{g_s(-n b_{ss'}) v_{rs't'}(-n b_{t't}) g_t\} = n E(b_{ss'} b_{tt'} v_{rs't'} g_s g_t) \\
&= b_{ss'} b_{tt'} \frac{1}{V_i}\left(2\frac{V_i'}{V_i} - 3\frac{h_i''}{h_i'^2}\right) \frac{\partial\mu_i \partial\mu_i \partial\mu_i}{\partial\beta_r \partial\beta_{s'} \partial\beta_{t'}} \frac{1}{V_j} \frac{\partial\mu_j \partial\mu_j}{\partial\beta_s \partial\beta_t} \\
&= \frac{1}{V_i}\left(2\frac{V_i'}{V_i} - 3\frac{h_i''}{h_i'^2}\right) h_i'^2 Q_{ii} \frac{\partial\mu_i}{\partial\beta_r}, \qquad \text{(A.3)}
\end{aligned}
$$

$$
\begin{aligned}
E(E_{11}\Delta_{11}^{-1} g^1) &= \frac{1}{n} E\{e_{rs}(-n b_{ss'}) g_{s'}\} \\
&= -nE\left\{\left(-\frac{y_i - \mu_i}{V_i} \frac{\partial\mu_i}{\partial\beta_r} \frac{\partial\mu_i}{\partial\beta_s} g_{s'}\right) b_{ss'} \left(\frac{V_i'}{V_i} - \frac{h_i''}{h_i'^2}\right)\right\} \\
&= n b_{ss'} \frac{1}{V_i}\left(\frac{V_i'}{V_i} - \frac{h_i''}{h_i'^2}\right) \frac{\partial\mu_i}{\partial\beta_r} \frac{\partial\mu_i}{\partial\beta_s} \frac{\partial\mu_i}{\partial\beta_{s'}} \\
&= \frac{n}{V_i}\left(\frac{V_i'}{V_i} - \frac{h_i''}{h_i'^2}\right) h_i'^2 Q_{ii} \frac{\partial\mu_i}{\partial\beta_r}, \qquad \text{(A.4)}
\end{aligned}
$$

$$
\begin{aligned}
E(c^2) &= E(Z_1^T v_q Z_1) = \frac{1}{n} E(g_1^T \Delta_{11}^{-1} v_q^1 \Delta_{11}^{-1} g_1) \\
&= n^{-1} E\{g_r(-n b_{rr'}) v_{qr's'}(-n b_{s's}) g_s\} \\
&= -\left\{2\left(a_i' - \frac{1}{V_i \kappa_1^{(i)}}\right) + a_i \frac{h_i''}{h_i'^2}\right\} h_i'^2 Q_{ii} \qquad \text{(A.5)}
\end{aligned}
$$

and

$$
\begin{aligned}
E(E_{21}\Delta_{11}^{-1} g^1) &= \frac{1}{n} E\{e_{21,r}(-n b_{rs}) g_s\} \\
&= -b_{rs} E\left\{\left(a_i' - \frac{2}{V_i \kappa_1^{(i)}}\right)(y_i - \mu_i) \frac{\partial\mu_i}{\partial\beta_r} g_s\right\} = -\left(a_i' - \frac{2}{V_i \kappa_1^{(i)}}\right) h_i'^2 Q_{ii}. \qquad \text{(A.6)}
\end{aligned}
$$

Then, using (A.2)-(A.6) in (A.1), after simplification, we obtain

$$
E\hat{\phi} = 1 - \frac{1}{n V_i \kappa_1^{(i)}} h_i'^2 Q_{ii} + \frac{1}{2n}(\gamma_i - a_i) h_i'' Q_{ii} + O\left(\frac{1}{n^2}\right). \qquad \text{(A.7)}
$$

Similarly it can be shown that

$$
\begin{aligned}
\operatorname{var}(\hat{\phi}) &= \frac{1}{n^2}\left\{a_i^2 V_i - a_i \gamma_i V_i + 2\frac{\kappa_{11}^{(i)}}{\kappa_1^{(i)}}(a_i - \gamma_i) + \frac{\kappa_2^{(i)}}{(\kappa_1^{(i)})^2}\right\} + O\left(\frac{1}{n^2}\right). \qquad \text{(A.8)} \\
\kappa_3(\hat{\phi}) &= \frac{1}{n^3}\left\{(a_i - \gamma_i)^3 \kappa_{03}^{(i)} + 3(a_i - \gamma_i)^2 \frac{\kappa_{12}^{(i)}}{\kappa_1^{(i)}} + 3(a_i - \gamma_i)\frac{\kappa_{21}^{(i)}}{(\kappa_1^{(i)})^2} + \frac{\kappa_3^{(i)}}{(\kappa_1^{(i)})^3}\right\} + O\left(\frac{1}{n^3}\right). \qquad \text{(A.9)}
\end{aligned}
$$

The results in Theorem 3.3.1 follow from equations (A.7)-(A.9).

A.2 Proof of Theorem 3.3.2

Using the conditional cumulant formulae in McCullagh (1987, p159), we have

$$E(\hat\phi|\hat\beta) = \hat{E}(\hat\phi) + \hat\kappa^r \hat{h}_r + (\hat\kappa^{d,r,s} - \hat\eta_t \hat\kappa^{r,s,t})\frac{\hat{h}_{rs}}{2\sqrt{n}} + O(n^{-\frac{3}{2}}), \tag{A.10}$$

$$\mathrm{var}(\hat\phi|\hat\beta) = \hat{\mathrm{var}}(\hat\phi) - \hat\eta_r\hat\eta_s\hat\kappa^{r,s} + (\hat\kappa^{d,d,r} - \hat\eta_s\hat\kappa^{d,r,s}[2] + \hat\eta_s\hat\eta_t\hat\kappa^{r,s,t})\frac{\hat{h}_r}{\sqrt{n}} + O(n^{-\frac{5}{2}}),$$

$$\kappa_3(\hat\phi|\hat\beta) = (\hat\kappa^{d,d,d} - \hat\eta_r\hat\kappa^{d,d,r}[3] + \hat\eta_r\hat\eta_s\hat\kappa^{d,r,s} - \hat\eta_r\hat\eta_s\hat\eta_t\hat\kappa^{r,s,t})\frac{1}{\sqrt{n}} + O(n^{-\frac{7}{2}}),$$

where

$$h_r(x,\lambda) = \lambda_{r,s}(x^s - \lambda^s) = \lambda_{r,s}(\sqrt{n}\hat\beta - \sqrt{n}E\hat\beta), \qquad h_{rs}(x,\lambda) = h_r h_s - \lambda_{r,s},$$

$$\eta_t = \kappa^s(\kappa^{s,t})^{-1}, \qquad \kappa^r = \mathrm{cov}(\hat\phi, \hat\beta_r),$$

$$\kappa^{r,s} = \mathrm{cov}(\hat\beta_r, \hat\beta_s), \qquad \kappa^{r,s,t} = \sqrt{n}\,\mathrm{cum}(\hat\beta_r, \hat\beta_s, \hat\beta_t),$$

$$\kappa^{d,d,r} = \sqrt{n}\,\mathrm{cum}(\hat\phi, \hat\phi, \hat\beta_r), \qquad \kappa^{d,r,s} = \sqrt{n}\,\mathrm{cum}(\hat\phi, \hat\beta_r, \hat\beta_s),$$

$$\kappa^{d,d,d} = \sqrt{n}\,\mathrm{cum}(\hat\phi, \hat\phi, \hat\phi) = \sqrt{n}\kappa_3(\hat\phi).$$

From Farrington (1996), we have that for $r = 1, ..., p$,

$$h_r(x,\lambda) = \frac{1}{\sqrt{n}} g_r + O_p(n^{-\frac{1}{2}}) \tag{A.11}$$

and

$$h_{rs}(x,\lambda) = \frac{1}{n} g_r g_s + \delta_{rs} + O_p(n^{-\frac{1}{2}}). \tag{A.12}$$

Further, after detailed calculations, we have

$$\kappa^r = \frac{1}{n}\frac{\kappa_{11}^{(i)}}{\kappa_1^{(i)} V_i}\frac{\partial\mu_i}{\partial\beta_t} b_{tr} + O(\frac{1}{n^2}), \tag{A.13}$$

$$\beta_r = \frac{1}{n}\frac{\kappa_{11}^{(i)}}{\kappa_1^{(i)} V_i}\frac{\partial\mu_i}{\partial\beta_r} + O(\frac{1}{n}), \tag{A.14}$$

$$\kappa^{r,s,t} = \sqrt{n} b_{rr'} b_{ss'} b_{tt'} \frac{\partial\mu_i}{\partial\beta_{r'}}\frac{\partial\mu_i}{\partial\beta_{s'}}\frac{\partial\mu_i}{\partial\beta_{t'}}\frac{\kappa_{03}^{(i)}}{V_i^3} + O(n^{-\frac{3}{2}}), \tag{A.15}$$

$$\kappa^{d,r,s} = n^{-\frac{1}{2}}\left\{\frac{\kappa_{12}^{(i)}}{\kappa_1^{(i)} V_i^2} + (a_i - \gamma_i)\frac{\kappa_{03}^{(i)}}{V_i^2}\right\}\frac{\partial\mu_i}{\partial\beta_{r'}}\frac{\partial\mu_i}{\partial\beta_{s'}} b_{rr'} b_{ss'} + O(n^{-\frac{3}{2}}). \tag{A.16}$$

Then, using (A.11)-(A.16) in (A.10), and by simplification, we obtain

$$E(\hat{\phi}|\hat{\beta}) = \hat{E}(\hat{\phi}) - \frac{1}{2n}\mathbf{1}^T X^T\{\hat{\kappa}_{12} + (\hat{a}-\hat{\gamma})\hat{\kappa}_{03}^2 - \hat{\kappa}_{11}^1\hat{\kappa}_{03}^3\}X(X^T\hat{W}X)^{-1}\mathbf{1} + O(n^{-\frac{3}{2}}), \tag{A.17}$$

Following similar calculations it can be shown that

$$\mathrm{var}(\hat{\phi}|\hat{\beta}) = \hat{\mathrm{var}}(\hat{\phi}) - \frac{1}{n^2}\hat{\kappa}_{11}^T X(X^T\hat{W}X)^{-1}X^T\hat{\kappa}_{11} + O(n^{-\frac{5}{2}}), \tag{A.18}$$

and

$$\begin{aligned}\kappa_3(\hat{\phi}|\hat{\beta}) &= \hat{\kappa}_3(\hat{\phi}) - 3n^{-3}\hat{\kappa}_{11}^T X(X^T\hat{W}X)^{-1}X^T\{\hat{\kappa}_{21} + \hat{\kappa}_{03}^1(\hat{a}-\hat{\gamma})^2 + 2\hat{\kappa}_{12}^1(\hat{a}-\hat{\gamma})\}\mathbf{1}\\ &\quad +3n^{-3}\hat{\kappa}_{11}^T X(X^T\hat{W}X)^{-1}X^T\{\hat{\kappa}_{12} + (\hat{a}-\hat{\gamma})\hat{\kappa}_{03}^2\}X(X^T\hat{W}X)^{-1}X^T\hat{\kappa}_{11}\\ &\quad -n^{-3}\mathbf{1}^T\hat{\kappa}_{03}^3\hat{\kappa}_{11}^1\hat{\kappa}_{11}^1\hat{\kappa}_{11}^1\mathbf{1} + O(n^{-\frac{7}{2}}).\end{aligned} \tag{A.19}$$

The results in Theorem 3.3.2 follow from (A.17)-(A.19).

A.3 The expected value of $\hat{T}$

Using the Taylor expansion (McCullagh and Nelder 1983, Appendix C), we have,

$$g_r(\hat{\beta}) - g_r(\beta) = \frac{\partial g_r}{\partial \beta}(\hat{\beta}-\beta) + O_p(1) = E(\frac{\partial g_r}{\partial \beta})(\hat{\beta}-\beta) + O_p(1).$$

Now, $g_r(\hat{\beta}) = 0$ and

$$E(\frac{\partial g_r}{\partial \beta_s}) = -\sum_{i=1}^{n}\frac{1}{V_i}\frac{\partial \mu_i}{\partial \beta_r}\frac{\partial \mu_i}{\partial \beta_s} = -(X^TWX)_{rs}.$$

Thus,

$$(\hat{\beta}-\beta) = (X^TWX)^{-1}(X^T\,\mathrm{diag}(h_i'/V_i)(y-\mu) + O_p(n^{-1}).$$

Further, since

$$(\hat{\mu}-\mu) = \frac{\partial \mu}{\partial \beta}(\hat{\beta}-\beta) + O_p(1),$$

we have

$$\frac{\hat{\mu}-\mu}{V^{1/2}} = W^{1/2}X(X^TWX)^{-1}X^TW^{1/2}\left(\frac{y-\mu}{V^{1/2}}\right) + O_p(1) = H\left(\frac{y-\mu}{V^{1/2}}\right) + O_p(1)$$

and

$$\frac{y-\hat{\mu}}{V^{1/2}} = (I-H)\frac{y-\mu}{V^{1/2}} + O_p(1).$$

Note that $(y-\mu)/V^{1/2} \sim N(O,I)$ as $\mu \to \infty$ and as $n \to \infty$ $\hat{V} \to V$ in probability. Thus, for sufficiently large n, and as $\mu \to \infty$,

$$\begin{aligned} E\left(\sum_{i=1}^{n}\frac{(y_i-\hat{\mu})^2}{\hat{V}_i}\right) &\approx E\left[(\frac{y-\mu}{V^{1/2}})^T(I-H)(\frac{y-\mu}{V^{1/2}})\right] \\ &= tr(I-H) = tr(I) - tr(W^{1/2}X(X^TWX)^{-1}X^TW^{1/2}) \\ &= n - tr((X^TWX)^{-1}(X^TWX)) = n - tr(I_p) = n-p. \end{aligned}$$

and

$$E\left(\sum_{i=1}^{n}-\frac{\hat{V}_i'}{\hat{V}_i}(y_i-\hat{\mu}_i)\right) \approx E\left(\sum_{i=1}^{n}\sum_{j=1}^{n}-\frac{V_i'}{V_i^{1/2}}(1-h_{ij})\frac{(y_j-\mu_j)}{V_j^{1/2}}\right) = 0$$

Thus,

$$\begin{aligned} E(\hat{T}) &= E\left(\sum_{i=1}^{n}\frac{(y_i-\hat{\mu})^2}{\hat{V}_i}\right) - E\left(\sum_{i=1}^{n}\frac{\hat{V}_i'}{\hat{V}_i}(y_i-\hat{\mu}_i)\right) - n \\ &= (n-p)-n = -p. \end{aligned}$$

A.4 Proofs of Theorem 6.3.1- Theorem 6.3.3

Partition $I(\beta,\tau,\gamma)$ as

$$I(\beta,\tau,\gamma) = \begin{bmatrix} I_{\beta\beta} & I_{\beta\tau} & I_{\beta\gamma} \\ I_{\tau\beta} & I_{\tau\tau} & I_{\tau\gamma} \\ I_{\gamma\beta} & I_{\gamma\tau} & I_{\gamma\gamma} \end{bmatrix}$$

where $I_{\beta\beta}, I_{\gamma\gamma}, I_{\beta\gamma}, I_{\gamma\gamma}, I_{\gamma\tau}$ and $I_{\beta\tau}$ are $p\times p, 1\times 1, 1\times p, 1\times 1, 1\times 1$ and $1\times p$ matrices respectively and have the different meanings for different models. To obtain the elements of the matrix $I(\beta, \tau, \gamma)$, we first give the first and second partial derivatives of log likelihood function $l_i(\gamma, \tau, \theta_i; y_i)$ with respect to γ, τ, θ_i as follows.

$$
\begin{aligned}
\frac{\partial l_i}{\partial \gamma} &= \frac{-1}{1+\gamma} + I_{\{y_i=0\}}\frac{1}{\gamma+f_0},\\
\frac{\partial l_i}{\partial \tau} &= I_{\{y_i=0\}}\frac{f_0}{\gamma+f_0}\frac{\frac{\partial}{\partial\tau}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}} + I_{\{y_i>0\}}\frac{\frac{\partial}{\partial\tau}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}},\\
\frac{\partial l_i}{\partial \theta_i} &= I_{\{y_i=0\}}\frac{f_0}{\gamma+f_0}\left[-g' + \frac{\frac{\partial}{\partial\theta_i}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}}\right]\\
&+ I_{\{y_i>0\}}[a'y_i - g'] + I_{\{y_i>0\}}\frac{\frac{\partial}{\partial\theta_i}\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}},\\
\frac{\partial^2 l_i}{\partial \gamma^2} &= \frac{1}{(1+\gamma)^2} - I_{\{y_i=0\}}\frac{1}{(\gamma+f_0)^2},\\
\frac{\partial^2 l_i}{\partial \theta_i^2} &= I_{\{y_i=0\}}\frac{-1}{(\gamma+f_0)^2}\left[f_0(-g') + f_0\frac{\frac{\partial}{\partial\theta_i}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}}\right]^2 + I_{\{y_i=0\}}\frac{f_0}{\gamma+f_0}\\
&\left[(-g' + \frac{\frac{\partial}{\partial\theta_i}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}})^2 - g'' + \frac{\frac{\partial^2}{\partial\theta_i^2}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}} + \frac{-(\frac{\partial}{\partial\theta_i}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}))^2}{(1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})^2}\right]\\
&+I_{\{y_i>0\}}[a''y_i - g''] + I_{\{y_i>0\}}\frac{\frac{\partial^2}{\partial\theta_i^2}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}} + I_{\{y_i>0\}}\frac{-(\frac{\partial}{\partial\theta_i}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}))^2}{(1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})^2},\\
\frac{\partial^2 l_i}{\partial \tau^2} &= I_{\{y_i=0\}}\frac{-f_0^2}{(\gamma+f_0)^2}\left(\frac{\frac{\partial}{\partial\tau}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}}\right)^2 + I_{\{y_i=0\}}\frac{f_0}{\gamma+f_0}\frac{\frac{\partial^2}{\partial\tau^2}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}}\\
&+I_{\{y_i>0\}}\frac{(\frac{\partial}{\partial\tau}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}))^2}{(1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})^2} + I_{\{y_i>0\}}\frac{\frac{\partial^2}{\partial\tau^2}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}},\\
\frac{\partial^2 l_i}{\partial \tau\partial\gamma} &= -I_{\{y_i=0\}}\frac{f_0}{(\gamma+f_0)^2}\frac{\frac{\partial}{\partial\tau}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}},\\
\frac{\partial^2 l_i}{\partial \theta_i\partial\gamma} &= -I_{\{y_i=0\}}\frac{f_0}{(\gamma+f_0)^2}\left(-g' + \frac{\frac{\partial}{\partial\theta_i}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}}\right),\\
\frac{\partial^2 l_i}{\partial \theta_i\partial\tau} &= I_{\{y_i=0\}}\frac{f_0}{\gamma+f_0}\left[-g' + \frac{\frac{\partial}{\partial\theta_i}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}}\right]\frac{\frac{\partial}{\partial\tau}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}}\\
&+I_{\{y_i=0\}}\frac{f_0}{\gamma+f_0}\left[\frac{\frac{\partial^2}{\partial\theta_i\partial\tau}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}} + \frac{-\frac{\partial}{\partial\tau}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})\frac{\partial}{\partial\theta_i}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{(1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})^2}\right]\\
&+I_{\{y_i=0\}}\frac{-f_0^2}{(\gamma+f_0)^2}\left[-g' + \frac{\frac{\partial}{\partial\theta_i}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}}\right]\frac{\frac{\partial}{\partial\tau}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}}
\end{aligned}
$$

$$+I_{\{y_i>0\}}\left[\frac{\frac{\partial^2}{\partial\theta_i\partial\tau}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!}}+\frac{-\frac{\partial}{\partial\tau}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})\frac{\partial}{\partial\theta_i}(\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})}{(1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})^2}\right].$$

A.4.1 The test for $\tau = 0$

Now,let U be an $n \times p$ matrix with ir-element $\frac{\partial\theta_i}{\partial\beta_r}$, **1** an $n \times 1$ unit vector and let W_1^τ, W_2^τ and W_3^τ be diagonal matrices with ith elements W_{1i}^τ, W_{2i}^τ and W_{3i}^τ. From the above partial derivatives of log-likelihood, we have that

$$\begin{aligned}
W_{1i}^\tau &= E\{-\frac{\partial^2 l_i}{\partial\theta_i^2}\}|_{\tau=0} = g'' - a''Ey_i - \frac{f_0\gamma}{(\gamma+f_0)(1+\gamma)}g'^2 - \frac{\gamma}{1+\gamma}g'', \\
W_{2i}^\tau &= E\{-\frac{\partial^2 l_i}{\partial\theta_i\partial\gamma}\}|_{\tau=0} = -\frac{f_0}{(\gamma+f_0)(1+\gamma)}g', \\
W_{3i}^\tau &= E\{-\frac{\partial^2 l_i}{\partial\theta_i\partial\tau}\}|_{\tau=0} = [\frac{1}{2}g'bD_2\frac{f_0\gamma}{(\gamma+f_0)(1+\gamma)} \\
&+ \frac{1}{2}(bD_2)'\frac{\gamma}{1+\gamma}]|_{y_i=0} - \frac{1}{2}E[(bD_2)'].
\end{aligned}$$

Further,

$$\begin{aligned}
I_{\tau\tau}^\tau &= \sum[\frac{1}{4}E(bD_2)^2 - \frac{(\gamma^2+2f_0\gamma)}{(\gamma+f_0)(1+\gamma)}(\frac{1}{2}bD_2)^2|_{y_i=0}], \\
I_{\gamma\gamma}^\tau &= \sum\frac{1-f_0}{(1+\gamma)^2(\gamma+f_0)}|_{y_i=0}, \\
I_{\gamma\tau}^\tau &= \sum\frac{\frac{1}{2}bD_2 f_0}{(\gamma+f_0)(1+\gamma)}|_{y_i=0}.
\end{aligned}$$

Hence, we have

$$S_i^\tau(\theta_i,\gamma) = \frac{\partial l_i}{\partial\tau}|_{\tau=0} = -\frac{\gamma I_{\{y_i=0\}}}{\gamma+f_0} + \frac{1}{2}bD_2.$$

Now,

$$I_\tau(\underline{\beta},\tau,\gamma) = \begin{bmatrix} U^T W_1^\tau U & U^T W_2^\tau \mathbf{1} & U^T W_3^\tau \mathbf{1} \\ \mathbf{1}^T W_2^\tau U & I_{\gamma\gamma}^\tau & I_{\gamma\tau}^\tau \\ \mathbf{1}^T W_3^\tau U & I_{\tau\gamma}^\tau & I_{\tau\tau}^\tau \end{bmatrix}$$

The asymptotic variance of $S_\tau = \sum_{i=1}^n(\frac{1}{2}bD_2 - \frac{\gamma I_{\{y_i=0\}}}{\gamma+f_0}) = \sum_{i=1}^n S_i^\tau(\hat\theta_i, \hat\gamma)$ is then

$$V_\tau = I_{\tau\tau}^\tau - (\mathbf{1}^T W_3^\tau U, I_{\tau\gamma}^\tau)\begin{pmatrix} U^T W_1^\tau U & U^T W_2^\tau \mathbf{1} \\ \mathbf{1}^T W_2^\tau U & I_{\gamma\gamma}^\tau \end{pmatrix}^{-1}\begin{pmatrix} U^T W_3^\tau \mathbf{1} \\ I_{\gamma\tau}^\tau \end{pmatrix},$$

where $\hat\theta_i$ and $\hat\gamma$ are the maximum likelihood estimates of θ_i and γ when $\tau = 0$.

Similarly, by the simplification, we obtain the expression of asymptotic variance of S_τ. Then a standardized test statistic for testing that $\tau = 0$ is thus

$$S_\tau^2/\hat V_\tau$$

with $\hat V_\tau = V_\tau(\hat\theta_1, ..., \hat\theta_n; \hat\gamma)$. Under mild regularity conditions, this score test statistic, asymptotically, as $n \to \infty$, has a $\chi^2(1)$ distribution.

A.4.2 The tests for $\gamma = 0$

Similarly let W_1^γ, W_2^γ and W_3^γ be diagonal matrices with ith diagonal elements W_{1i}^γ, W_{2i}^γ and W_{3i}^γ. Then,

$$\begin{aligned} W_{1i}^\gamma &= E\{-\frac{\partial^2 l_i}{\partial\theta_i^2}\}|_{\gamma=0} = g'' - a''Ey_i - E\{\frac{\partial^2}{\partial\theta_i^2}\log(1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})\}, \\ W_{2i}^\gamma &= E\{-\frac{\partial^2 l_i}{\partial\theta_i\partial\tau}\}|_{\gamma=0} = -E\{\frac{\partial^2}{\partial\theta_i\partial\tau}\log(1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})\}, \\ W_{3i}^\gamma &= E\{-\frac{\partial^2 l_i}{\partial\theta_i\partial\gamma}\}|_{\gamma=0} = -g' + \{\frac{\partial}{\partial\theta_i}\log(1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})\}|_{y_i=0}. \end{aligned}$$

Further,

$$\begin{aligned} I_{\tau\tau}^\gamma &= \sum E\{-\frac{\partial^2}{\partial\tau^2}\log(1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})\}, \\ I_{\tau\gamma}^\gamma &= \sum\frac{\partial}{\partial\tau}\log(1+\sum_{r=2}^{\infty}\frac{\alpha_r D_r}{r!})|_{y_i=0}, \\ I_{\gamma\gamma}^\gamma &= \sum(1/f_0 - 1). \end{aligned}$$

Also, we have

$$S_i^\gamma(\theta_i, \tau) = \frac{\partial l_i}{\partial\gamma}|_{\gamma=0} \;=\; (-\frac{1}{1+\gamma} + I_{\{y_i=0\}}\frac{1}{\gamma+f_M(0)})|_{\gamma=0} = (\frac{I_{\{y_i=0\}}}{f_M(0)} - 1).$$

Note that

$$I_\gamma(\beta,\tau,\gamma) = \begin{bmatrix} U^T W_1^\gamma U & U^T W_2^\gamma \mathbf{1} & U^T W_3^\gamma \mathbf{1} \\ \mathbf{1}^T W_2^\gamma U & I_{\tau\tau}^\gamma & I_{\tau\gamma}^\gamma \\ \mathbf{1}^T W_3^\gamma U & I_{\gamma\tau}^\gamma & I_{\gamma\gamma}^\gamma \end{bmatrix}$$

So, the asymptotic variance of $S_\gamma = \sum_{i=1}^n (\frac{I_{\{y_i=0\}}}{f_0} - 1) = \sum_{i=1}^n S_i^\gamma(\hat{\theta}_i, \hat{\tau})$ is then

$$V_\gamma = I_{\gamma\gamma}^\gamma - (\mathbf{1}^T W_3^\gamma U, I_{\gamma\tau}^\gamma) \begin{pmatrix} U^T W_1^\gamma U & U^T W_2^\gamma \mathbf{1} \\ \mathbf{1}^T W_2^\gamma U & I_{\tau\tau}^\gamma \end{pmatrix}^{-1} \begin{pmatrix} U^T W_3^\gamma \mathbf{1} \\ I_{\tau\gamma}^\gamma \end{pmatrix},$$

where $\hat{\theta}_i$ and $\hat{\tau}$ are the maximum likelihood estimates of θ_i and τ when $\gamma = 0$.

Now, by the simplification, we obtain the asymptotic variance of S_γ. Therefore a standardized test statistic for testing that $\gamma = 0$ is thus

$$S_\gamma^2 / \hat{V}_\gamma$$

with $\hat{V}_\gamma = V_\gamma(\hat{\theta}_1, ..., \hat{\theta}_n; \hat{\tau})$. Under mild regularity conditions, this score test statistic, asymptotically, as $n \to \infty$, has a $\chi^2(1)$ distribution.

A.4.3 Test for $(\gamma, \tau) = 0$

Now, we give the score test statistic for testing both zero-inflation and over-dispersion in the generalized linear model. Similarly, let W_1, W_2 and W_3 be diagonal matrices with ith diagonal elements W_{1i}, W_{2i} and W_{3i}. Then,

$$\begin{aligned} W_{1i} &= E\{-\frac{\partial^2 l_i}{\partial \theta_i^2}\}|_{(\gamma,\tau)=0} = g_i'' - a_i'' E y_i, \\ W_{2i} &= E\{-\frac{\partial^2 l_i}{\partial \theta_i \partial \gamma}\}|_{(\gamma,\tau)=0} = -g', \\ W_{3i} &= E\{-\frac{\partial^2 l_i}{\partial \theta_i \partial \tau}\}|_{(\gamma,\tau)=0} = -\frac{1}{2} E(\frac{d(bD_2)}{d\theta_i}) \\ &= \frac{1}{2} b_i [g_i'\{(h_i')^2 - h_i''\} - 2g_i'' h_i' + g_i''']. \end{aligned}$$

Further,

$$I_{\gamma\gamma} = \sum_{i=1}^n E\{-\frac{\partial^2 l_i}{\partial \gamma^2}|_{(\gamma,\tau)=0} = \sum_{i=1}^n (1/f(0;\theta_i) - 1),$$

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

Now, we note that the score functions have the following forms:

$$
\begin{aligned}
S_i(\theta_i) = \frac{\partial l_i}{\partial\gamma}|_{(\gamma,\tau)=0} &= (-\frac{1}{1+\gamma} + I_{\{y_i=0\}}\frac{1}{\gamma+f_0})|_{(\gamma,\tau)=0} = (\frac{I_{\{y_i=0\}}}{f_0} - 1), \\
T_i(\hat{\theta}_i) = \frac{\partial l_i}{\partial\gamma}|_{(\gamma,\tau)=0} &= \frac{1}{2} b_i(a_i')^2\{(y_i - \mu_i)^2 - (a_i')^{-2}(g_i'' - a_i''y_i)\}
\end{aligned}
$$

So,

$$
I(\underline{\beta},\gamma,\tau) = \begin{bmatrix} U^TW_1U & U^TW_2\mathbf{1} & U^TW_3\mathbf{1} \\ \mathbf{1}W_2U & I_{\gamma\gamma} & I_{\gamma\tau} \\ \mathbf{1}W_3U & I_{\tau\gamma} & I_{\tau\tau} \end{bmatrix}
$$

and thus, the asymptotic covariance matrix of $(\sum_{i=1}^n S_i(\theta_i), \sum_{i=1}^n T_i(\theta_i)) \equiv (S,T)$ is then

$$
\begin{aligned}
\Sigma &= \begin{pmatrix} I_{\gamma\gamma} & I_{\gamma\tau} \\ I_{\tau\gamma} & I_{\tau\tau} \end{pmatrix} - \begin{pmatrix} \mathbf{1}^TW_2Y \\ \mathbf{1}^TW_3U \end{pmatrix} (U^TW_1U)^{-1}(U^TW_2\mathbf{1} \quad U^TW_3\mathbf{1}) \\
&= \begin{pmatrix} I_{\gamma\gamma} - \mathbf{1}^TW_2U(U^TW_1U)^{-1}U^TW_2\mathbf{1} & I_{\gamma\tau} - \mathbf{1}^TW_2U(U^TW_1U)^{-1}U^TW_3\mathbf{1} \\ I_{\tau\gamma} - \mathbf{1}^TW_3U(U^TW_1U)^{-1}U^TW_2\mathbf{1} & I_{\tau\tau} - \mathbf{1}^TW_3U(U^TW_1U)^{-1}U^TW_3\mathbf{1} \end{pmatrix} \\
&= \begin{pmatrix} V_{SS} & V_{ST} \\ V_{TS} & V_{TT} \end{pmatrix}
\end{aligned}
$$

and a standardized test statistic for testing that $(\gamma,\tau) = 0$ is thus

$$
X^2 = (\hat{S},\hat{T})\hat{\Sigma}^{-1}(\hat{S},\hat{T})^T = \frac{\hat{V}_{TT}\hat{S}^2 + \hat{V}_{SS}\hat{T}^2 - 2\hat{V}_{ST}\hat{S}\hat{T}}{\hat{V}_{TT}\hat{V}_{SS} - \hat{V}_{ST}^2}
$$

where $\hat{S} = S(\hat{\theta}), \hat{T} = S(\hat{\theta}), \hat{V}_{SS} = V_{SS}(\hat{\theta}), \hat{V}_{TT} = V_{TT}(\hat{\theta})$ and $\hat{V}_{ST} = V_{ST}(\hat{\theta})$. $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimate of vector of parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)$ under the null hypothesis $H_0 : (\gamma,\tau) = 0$. The asymptotic distribution of this test statistic is the χ^2-distribution with two degrees of freedom. The large value of X^2 provides evidence against the hypothesis.

Bibliography

[1] Andrews, D. F. and Herzberg, A. M. (1985) *Data.* New York: Springer-Verlag.

[2] Bartlett, M. S. (1953) Approximate confidence intervals. *Biometrika*, **40**, 12-19.

[3] Bartoo, J. B. and Puri, P. S. (1967) On optimal asymptotic tasks of composite statistical hypothesis. *The Annals of Mathematical Statistics*, **38**, 1845-1852.

[4] Berry, D. A. (1987). Logarithmic transformations in ANOVA. *Biometrics*, **43**, 439-456.

[5] Böhning, D., Dietz, E. and Schlattmann, P. (1999) The zero-inflated Poisson model and the decayed, missing and filled index in dental epidemiology. *Journal of the Royal Statistical Society*, A, **162**, 195-209.

[6] Breslow, N. (1984) Extra-Poisson variation in log-linear models. *Applied Statistics*, **33**, 38-44.

[7] Breslow, N. (1990). Tests of hypotheses in over-dispersed Poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association,* **85**, 565-571.

[8] Brillinger, D. R. (1986) The natural variability of vital rates and associated statistics (with discussion).*Biometrics*, **42**, 693-734.

[9] Broek, V. J. (1995). A score test for zero inflation in a Poisson distribution. *Biometrics*, **51**, 738-743.

[10] Campbell, M. J., Machin, D. and D'Arcangues, C (1991) Coping with extra-Poisson variability in the analysis of factors influencing vaginal ring expulsions. *Statistics in Medicine.*, **10**, 241-251.

[11] Chesher, A. (1984) Testing for neglected heterogeneity. *Econometrica*, **52**, 865-872.

[12] Cox, D. R. (1983) Some remarks on overdispersion. *Biometrika*, **70**, 269-274.

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

[14] Cox, D. R. and Reid, N. (1987) Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society*, B, **49**, 1-39.

[15] Cramer, H. (1946) *Mathematical Methods of Statistics*, Princeton University Press.

[16] Cressie, N. and Read, T. R. (1984) Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society*, B, **46**, 440-464.

[17] Crowder, M. J. (1978) Beta-binomial ANOVA for proportions, *Applied Statistics*, **27**, 34-37.

[18] Crowder, M. J. (1985) Gaussian estimation for correlated binomial data. *Journal of the Royal Statistical Society*, ser. B, **47**, 229-237.

[19] Dean, C. B. (1992). Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association*, **87**, 451-457.

[20] Dean, C. and Lawless, J. F. (1989). Tests for detecting over-dispersion in Poisson regression models. *Journal of the American Statistical Association*, **84**, 467-472.

[21] Drost, F. C., Kallenberg, W. C. M., Moore, D. S. and Oosterhoff, J. (1989) Power approximations to multinomial tests of fits. *Journal of the American Statistical Association*, **84**, 130-141.

[22] Farewell, V. T. and Sprott, D. A. (1988). The use of a mixture model in the analysis of count data. *Biometrics*, **44**, 1191-1194.

[23] Farrington, C. P. (1996) On assessing goodness of fit of generalized linear models to sparse data. *Journal of the Royal Statistical Society*, B, **58**, 349-360.

[24] Firth, D. (1987) Discussion of 'parameter orthogonality and approximate conditional inference' by D.R. Cox and N. Reid. *Journal of the Royal Statistical Society*, B, **49**, 22-23.

[25] Johnson, N. L., Kotz, S. and Kemp, A. W. (1992) *Univariate Discrete Distributions*, second edition. New York: John Wiley & Sons, Inc.

[26] Jorgensen, B. (1987) Exponential dispersion model. *Journal of the Royal Statistical Society*, B, **49**, 127-162.

[27] Keiding, N. (1991) Age-specific incidence and prevalence: a statistical perspective (with discussion). *Journal of the Royal Statistical Society*, A, **154**, 371-412.

[28] Koehler, K. J. (1986) Goodness-of-fit tests for log-linear models in sparse contingency tables. *Journal of the American Statistical Association*, **81**, 483-493.

[29] Koehler, K. J. and Larntz, K. (1980) An empirical investigation of goodness-of-fit statistics for sparse multinomials. *Journal of the American Statistical Association*, **75**, 336-344.

[30] Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics* **34**, 1-14.

[31] Lawless, J. F.(1987a) Negative binomial regression models. *Canadian Journal of Statistics*, **15**, 209-226.

[32] Lawless, J. F. (1987b) Regression methods for Poisson process data. *Journal of the American Statistical Association*, **82**, 808-815.

[33] Lindsay, B. G. and Roeder, K. (1992) Residual diagnostics for mixture models. *Journal of the American Statistical Association*, **87**, 785-794.

[34] McCullagh, P. (1984) Tensor notation and cumulants of polynomials. *Biometrika*. **71**, 461-476.

[35] McCullagh, P. (1985) On the asymptotic distribution of Pearson's statistic in linear exponential family models.*International Statistical Review*, **53**, 61-67.

[36] McCullagh, P. (1986) The conditional distribution of goodness-of-fit statistics for discrete data. *Journal of the American Statistical Association*, **81**, 104-107.

[37] McCullagh, P. (1987) *Tensor Methods in Statistics*. London: Chapman and Hall.

[38] McCullagh, P. and Nelder, J. (1983). *Generalized Linear Models,* 1st ed. London: Chapman and Hall.

[39] McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*. 2nd ed. London: Chapman and Hall.

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

[41] Moran, P. A. P. (1970) On asymptotically optimal tests of composite hypotheses. *Biometrika*, **57**, 45-75.

[42] Mullahy, J. (1997). Heterogeneity, excess zeros, and the structure of count data models. *Journal of Applied Econometrics* **12**, 337-350.

[43] Neyman, J. (1959). Optimal asymptotic tests for composite hypothesis. In *Probability and Statistics: The Harold Cramer Volume,* U. Grenander (ed). New York: Wiley.

[44] Neyman, J. and Scott, E. L. (1966) On the use of $C(\alpha)$ optimal tests of composite hypotheses. *Bulletin de l'Institut International de Statistique* , **41**, 477-497.

[45] Paul, S. R. and Banerjee. T. (1998). Analysis of two-way layout of count data involving multiple counts in each cell. *Journal of the American Statistical Association,* **93**, 1419-1429.

[46] Paul, S. R. and Islam, A. S. (1995). Analysis of proportions in the presence of over/under-dispersion. *Biometrics* **51**, 1400-1410.

[47] Prentice, G. L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association,* **81**, 167-182.

[48] Rao, C. R. (1947). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, **44**. 50-57.

[49] Walter, S. D., Hildreth, S. W. and Beaty, B. J. (1980) Estimation of infection rates in populations of organisms using pools of variable size. *American Journal of Epidemiology,* **112**, 124-128.

Vita Auctoris

Name:	Dianliang Deng
Date of Birth:	1962
Place of Birth:	Gongzhuling, China
Educations:	B. S. in Mathematics, 1982, Jilin University, China
	M. S. in Probability, 1985, Jilin University, China
	M. S. in Statistics, 1997, University of Windsor, Canada
	Ph. D. in Statistics, 2001, University of Windsor, Canada