East West University

Master's Thesis

Detecting Overdispersion in Count Data: Comparison of Tests

Submitted by
ID-2014-1-82-001

A thesis submitted in fulfilment of the requirements for the degree Master of Science in Applied Statitics

August 2015

Abstract

When data appear more dispersed than expected under a reference model, the situation is termed as overdispersion. In modelling a count variable in terms of some independent predictor variables, theoretically most established and the simplest available reference model is Poisson regression model. For standard Poisson regression model, variance is equal to mean and there is no extra parameter for dispersion. However, in practical scenario, the estimated variance from data often exceeds the mean and the data is considered to be overdispersed. To solve the overdispersion problem, two common alternative approaches are i) fitting a more general parametric distribution ii) having a different form mean variance relationship without fully specifying the distribution. Both approaches include parameters for overdispersion to be estimated from data. However, when there is no overdispersion, Poisson regression model is preferred for its simplicity, interpretability and theoretical basis. Therefore, robust test for detecting the significance of parameter related to overdispersion is important to use before going for alterative to Poisson regression.

In this work, we have investigated tests for detecting overdispersion when Poisson model is used for count data. The tests discussed are derived from partial score and are applicable against negative binomial or more generally mixed Poisson alternatives. These tests do not require fitting alternative models that incorporate overdispersion to check the absence of overdispersion. Only Poisson model is needed to be fitted. Four test statistics are illustrated with their distributional approximations for computing significance level. The test statistics have been analyzed and compared based on the assumptions on deriving the statistics, their limiting distributions and applicability for different number of observation in sample. A simulation study was done to check adequacy of distributional assumption for three of them who follow approximately normal distribution. The study involved generating samples of the statistics and proportion of the time each exceeded the standard normal upper 20%, 10%, 5%, and 1% point were tabulated. From the results, the normality assumption of one of the statistics has been observed to be good for large sample size but less accurate for small size. Another one of the statistics has been found to have almost accurate standard normal distribution even for small sample. Some comparisons and recommendations relating to the applicability and assumptions of the statistics are also presented.

Contents

Chapter 1

Introduction

For count data, the simplest and widely used regression model is the Poisson regression model. Poisson distribution was derived as a limiting case of the binomial by Poisson [1] for rare events. It is a discrete probability distribution function that gives the probability of a number of independent events occurring in a fixed time for a given average rate. In Generalized Linear Model (GLM) framework, it is specified by log-link function, mean variance equality and dispersion parameter equal to one.

Statistical theory provides strong reasons for choosing the Poisson model for count data. However, one main feature of Poisson distribution, the variance being equal to mean, is often not attained in real life data. For the Poisson MLE, it can be shown that consistency requires correct specification of the conditional mean [2]. It does not require that the dependent variable, Y be Poisson distributed. Valid statistical inference using default computed maximum likelihood standard errors and t statistics requires correct specification of both the conditional mean and variance. This also requires equality of conditional variance and mean termed as equidispersion, but not Poisson distribution for Y. Poisson model may fail to model the variance of date, because of unobserved heterogeneity or because of failure of the independence assumption in the usual derivation of the Poisson. If one of such violations of assumptions is present, we no longer can say anything about the distribution of the data. For count data, if the variance of the observations is greater than the mean predicted by the Poisson model, the phenomenon is called Poisson overdispersion or extra-Poisson variation and data is called to be overdispersed.

In the absence of equidispersion, many alternative methods have been suggested for dealing with extra Poisson variation [2–4]. The usual approach is to assume a distribution and proceed with classical testing and estimation procedures. For example, if heterogeneity is gamma distributed, then the count data will be negative binomial distributed. However, negative binomial regression also has limitations. Apart from the higher complexity in estimation, if the distributional assumption is not correct, even if

the variance and mean of the assumed negative binomial distribution are correctly specified, the maximum likelihood estimator is inconsistent. GLM suggests use of inference techniques only based on specification of mean variance relationship. Quasi likelihood approaches with log link function but uses different forms of variance function resulting in varying result in estimation and inference.

If the Poisson assumption still holds acceptably, it is more desirable for its simplicity, strong theoretical basis and having properly matched interpretation of regression coefficient. So it is very necessary to have a systematic way to determine when they apply. Most approaches require to model overdispersion with an extra parameter to a general variance than that attained for Poisson and the significance of that parameter is needed tested before going for such models. Testing with score statistic, likelihood ratio, Pearson statistic, Wald statistic are some options for testing requirement of modelling Poisson overdispersion. Score tests for detecting extra-Poisson variation have been discussed by Collings and Margolin [5], Cameron and Trivedi [6], Dean and Lawless [7]. Collings and Margolin dealt with some special cases of modelling starting with a negative binomial model [5]. They derived score, or $C(\alpha)$ tests and discussed asymptotic distribution theory for the tests. They presented a comparison with Pearson tests in detecting overdispersion. Cameron and Trivedi used the general variance function $Var(Y) = \mu(x) + \lambda\mu(x)^l$ for given l as alternative model for score test [6]. Dean and Lawless suggested tests with mixed Poisson alternative that require only the Poisson model to be fitted [7]. The scenario is different from Wald or likelihood ratio test that require more general models to be fitted.

The main aim of this work is to provide a review on and compare some tests to detect overdispersion in count data. These are formal tests aimed to detect when the variance of the data is signifucantly higher than that estimated using Poisson regression model. Poisson alternative with $Var(y_i|\mathbf{x}_i) = \mu_i + \alpha\mu_i^2$. These test statistics were derived from partial score of more to test the significance of α. A simulation study generating sample distributions of the statistics has been performed and comparison has been made mostly about the adequacy of the suggested approximations of distributional assumptions in literature. These test statistics only need Poisson ML estimates of means, not fitting any alternative distribution.

Chapter 2 provides some necessary backgrounds. A brief on important basics on Generalized Linear Model (GLM) are provided that are necessary in later discussions. Poisson regression model and MLE in GLM framework and Poisson overdispersion problem are reviewed in brief. Also negative binomial regression and its two important versions are revised. Finally, partial score statistic, which the origin of the discussed tests for detecting Poisson overdispersion is derived from negatve binomial log likelihood as a part of background. Chapter 3 presents some forms score tests for detecting Poisson overdispersion. It starts with presenting a test hypothesis with Poisson distribution as null model and mixed Poisson as alternative model. The mixed Poisson differ from Poisson

by a parameter α which is assumed to handle extra Poisson variation. To check the significance of this parameter, four statistics (we have denoted by S_1, S_2, S_a and S_b) and their approximate distributions are discussed which were presented in [7]. All of these are derived from a statistic which can obtain from partial score of negative binomial [5]. Three of the statistics, S_1, S_a and S_b had been shown to converge to standard normal as number of observation $n \to \infty$. In Chapter 4, adequacies of the normality assumptions for S_1, S_a and S_b have been examined by randomly simulating samples of these statistics for varying n given a particular Poisson regression equation. For comparison, similar simulation has been performed on well-established Pearson's Chi square statistic. Results were presented in terms of proportion of time the value of S_1,S_a and S_b each exceeded the standard normal upper 20%, 10%, 5%, and 1% points. Normality assumption S_b has been found to be almost accurate even for small values of n, S_a has shown slower convergence. The result also suggest the approximation for S_1 was not satisfactory. In Chapter 5, some additional comparisons and applicability of the tests have been discussed.

Chapter 2

Background

2.1 Generalized Linear Model

Generalized linear model (GLM) combines most frequently used statistical techniques in a unified framework in order to analyze non-normal data[8]. It was originally introduced by Nelder and Wedderburn [9] to extend linear regression analysis with normal response variable to an exponential family of distribution including normal, binomial, Poisson, gamma, or inverse-Gaussian families of distributions. Subsequent work, however, has extended GLMs to multivariate exponential families (e.g., the multinomial distribution), to some non-exponential families (e.g., the two-parameter negative binomial distribution), and to situations in where the distribution is not specified completely (Quasi-likelihood). GLMs have a general algorithm for estimation for fitting based on likelihood.

2.1.1 Components of GLM

Consider a vector of observations $\mathbf{y} = [y_1, y_2, ...y_n]^T$ to be a realization of a random variable $\mathbf{Y} = [Y_1, Y_2, ...Y_n]^T$ where each Y_i is a one-dimensional response to independent covariates $\mathbf{x}_i = [x_{i1}, \dots\ x_{ip}]^T$. A generalized linear model (GLM) consists of three components [8].

1. **Random component**, $Y_i, 1 \leq i \leq N$, assumed to be independent with probability density or mass function of the form [8]

$$f\left(y_i; \theta_i, \phi\right) = e^{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)} \tag{2.1}$$

   For known $a(.)$, $b(.)$ and $c(.)$, if ϕ, called the dispersion parameter, is known then it becomes an exponential distributions with canonical parameter θ.

2. **Systematic component**, η_i, a linear function of covariates $\mathbf{x}_i$ given by

$$\eta_i = \mathbf{x}_i\boldsymbol{\beta} = \sum_j x_{ij}\beta_j \tag{2.2}$$

where $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_p]^T$ are regression parameters.

3. **Link function,** $g(\mu_i)$, a monotonic function that links it to the linear predictor η_i with the mean μ_i.

$$g(\mu_i) = \eta_i = \sum_j x_{ij}\beta_j \tag{2.3}$$

It is called canonical link function if $\boldsymbol{\theta} = \boldsymbol{\eta}$.

Since the link function invertible, we may write

$$\mu_i = g^{-1}(\eta_i) = g^{-1}\left(\sum x_{ij}\beta_j\right) \tag{2.4}$$

Thus, the GLM may be thought of as a linear model for a transformation of the expected response or as a nonlinear regression model for the response. The inverse link $g^{-1}(.)$ is also called the mean function.

The mean and variance of the random component Y_i can be easily found to be as under(See Section 2.2.2 in [8]).

$$\mu_i = E(Y_i) = b'(\theta_i) \tag{2.5}$$

$$Var(Y_i) = b''(\theta_i)\, a(\phi) \tag{2.6}$$

The variance of Y_i is the product of two factors. One is $b''(\theta)$ that depends on the canonical parameter, therefore on the mean and called the variance function, $V(\mu)$. The other, $a(\phi)$ depending only on ϕ, commonly of the form $a(\phi) = \phi/w$, where w is a known prior weight that varies from observation to observation.

2.1.2 Maximum Likelihood Estimation (MLE) in GLM

Taking log on both side of (2.1) we get the log density for ith observation of Y

$$l_i(\theta_i; y_i, \phi) = \log f(y_i; \theta_i, \phi) = \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \tag{2.7}$$

Summing over these log-densities provides an expression for the log likelihood of n independent observations,

$$l(\mathbf{y}, \boldsymbol{\beta}) = \sum_{i=1}^{n} l_i(y_i; \theta_i, \phi) = \sum_{i=1}^{n}\left[\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right] \tag{2.8}$$

Under the regularity conditions, $\frac{\partial l(\mathbf{y},\boldsymbol{\beta})}{\partial \beta_j} = 0$, $j = 1, 2, \ldots, p$.

Now $g(\mu_i) = \eta_i$, $\mu_i = b'(\theta)$, $\eta_i = \sum_j x_{ij}\beta_j$. So we can employ the chain rule,

$$\frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} \tag{2.9}$$

We can find

$$\frac{\partial l_i}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)} = \frac{y_i - \mu_i}{a(\phi)} \tag{2.10}$$

$$\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i) = \frac{Var(y_i)}{a(\phi)} \tag{2.11}$$

$$\frac{\partial \eta_i}{\partial \beta_j} = x_{ij} \tag{2.12}$$

From (2.7) to (2.10) we can wrie

$$\frac{\partial l_i}{\partial \beta_j} = \frac{y_i - \mu_i}{a(\phi)}\frac{a(\phi)}{Var(y_i)}\frac{\partial \mu_i}{\partial \eta_i}x_{ij} = \frac{(y_i - \mu_i)x_{ij}}{Var(y_i)}\frac{\partial \mu_i}{\partial \eta_i} = 0 \tag{2.13}$$

Then the jth score is given by

$$s_j(\boldsymbol{\beta}) = \frac{\partial l(\mathbf{y},\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n}\left[\frac{(y_i - \mu_i)x_{ij}}{Var(y_i)}\frac{\partial \mu_i}{\partial \eta_i}\right] \tag{2.14}$$

Thus, the MLE of $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_p]^T$ is the solution the system of the score equations given by

$$\mathbf{s}(\boldsymbol{\beta}) = [s_1(\boldsymbol{\beta}), s_2(\boldsymbol{\beta}), \ldots, s_p(\boldsymbol{\beta})]^T = \mathbf{0} \tag{2.15}$$

The estimating equations depend on the density f only through the mean and variance. The estimates are generally not available in closed form, but can be obtained via iterative algorithm. Two popular iterative methods are Newton-Raphson method and Fisher's scoring method. Both take on the same general form and differ only in the variance structure.

2.1.2.1 Newton-Raphson Method

Let $\widehat{\boldsymbol{\beta}}^{(m)}$ be the current estimation to the solution of $\mathbf{s}(\boldsymbol{\beta}) = \mathbf{0}$. Taking a first-order Taylor approximation to the equations, we obtain the new estimation value by Newton-Raphson method

$$\widehat{\boldsymbol{\beta}}^{(m+1)} = \widehat{\boldsymbol{\beta}}^{(m)} - \left[D\mathbf{s}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right)\right]^{-1}\mathbf{s}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right) \tag{2.16}$$

Where D denotes Jacobian of a vector which for score is given by

$$D\mathbf{s}\left(\boldsymbol{\beta}\right)=\left[\begin{array}{ccc}\frac{\partial s_1}{\partial\beta_1} & \cdots & \frac{\partial s_1}{\partial\beta_p}\\ \vdots & \ddots & \vdots\\ \frac{\partial s_p}{\partial\beta_1} & \cdots & \frac{\partial s_p}{\partial\beta_p}\end{array}\right]=\left[\begin{array}{ccc}\frac{\partial^2 l}{\partial\beta_1\partial\beta_1} & \cdots & \frac{\partial^2 l}{\partial\beta_1\partial\beta_p}\\ \vdots & \ddots & \vdots\\ \frac{\partial^2 l}{\partial\beta_p\partial\beta_1} & \cdots & \frac{\partial^2 l}{\partial\beta_p\partial\beta_p}\end{array}\right]=\left[\frac{\partial^2 l\left(\mathbf{y},\boldsymbol{\beta}\right)}{\partial\beta_j\partial\beta_k}\right]_{j,k=1,..,p}$$

Therefore,

$$\widehat{\boldsymbol{\beta}}^{(m+1)}=\widehat{\boldsymbol{\beta}}^{(m)}-\left[\frac{\partial^2 l\left(\mathbf{y},\boldsymbol{\beta}\right)}{\partial\beta_j\partial\beta_k}\right]^{-1}_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{(m)}}\mathbf{s}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right) \tag{2.17}$$

Hence $\left[\frac{\partial^2 l(\mathbf{y},\boldsymbol{\beta})}{\partial\beta_j\partial\beta_k}\right]_{j,k=1,..,p}$ is the Hessian of $l(\mathbf{y},\boldsymbol{\beta})$. When evaluated at $\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}$, it is equal to negative of observed information matrix at mth iteration. Now

$$\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}=\frac{\partial s_r}{\partial\beta_s}=\frac{\partial}{\partial\beta_s}\sum_{i=1}^{n}\left[\frac{\left(y_i-\mu_i\right)x_{ir}}{Var\left(y_i\right)}\frac{\partial\mu_i}{\partial\eta_i}\right]$$

$$=\sum_{i=1}^{n}\left[\left(y_i-\mu_i\right)\frac{\partial}{\partial\beta_s}\left(\frac{x_{ir}}{Var\left(y_i\right)}\frac{\partial\mu_i}{\partial\eta_i}\right)+\frac{x_{ir}}{Var\left(y_i\right)}\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial}{\partial\beta_s}\left(y_i-\mu_i\right)\right]$$

$$\frac{\partial}{\partial\beta_s}\left(y_i-\mu_i\right)=-\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\beta_s}=-\frac{\partial\mu_i}{\partial\eta_i}x_{is}$$

Hence,

$$\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}=\sum_{i=1}^{n}\left[\left(y_i-\mu_i\right)\frac{\partial}{\partial\beta_s}\left(\frac{x_{ir}}{Var\left(y_i\right)}\frac{\partial\mu_i}{\partial\eta_i}\right)-\frac{1}{Var\left(y_i\right)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 x_{ir}x_{is}\right] \tag{2.18}$$

Since $\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}$ depends on $\mathbf{Y}$, in Fisher's scoring method expected information $-E\left(\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}\right)$ used instead. For canonical link $\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}=E\left(\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}\right)$.

2.1.2.2 Fisher's Scoring Method

Now

$$E\left(\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}\right)=E\left(\sum_{i=1}^{n}\left[\left(y_i-\mu_i\right)\frac{\partial}{\partial\beta_s}\left(\frac{x_{ir}}{Var\left(y_i\right)}\frac{\partial\mu_i}{\partial\eta_i}\right)-\frac{1}{Var\left(y_i\right)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 x_{ir}x_{is}\right]\right)$$

$$=-\sum_{i=1}^{n}\left[\frac{1}{Var\left(y_i\right)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 x_{ir}x_{is}\right]=-\sum_{i=1}^{n}W_i x_{ir}x_{is}$$

Let the expected information matrix

$$A(\boldsymbol{\beta}) = \left[-E\left(\frac{\partial^2 l}{\partial\beta_j \partial\beta_k}\right)\right]_{j,k} = X^T X \in \mathbb{R}^{p\times p} \tag{2.19}$$

Where $W = diag(W_{1p}, \ldots, W_n) \in \mathbb{R}^{n\times n}$ and X=[$\mathbf{x_1}$, $\mathbf{x_2}$, $\ldots$,$\mathbf{x_n}$] $\in \mathbb{R}^{n\times p}$. The Fisher scoring method for new estimatiomn is given by

$$\widehat{\boldsymbol{\beta}}^{(m+1)} = \widehat{\boldsymbol{\beta}}^{(m)} + A^{-1}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right)\mathbf{s}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right) \tag{2.20}$$

2.1.2.3 Fisher's Scoring as Iteratively Weighted Least Squares

Fisher's scoring method can be shown as iteratively weighted least squares. (2.20) can be written as

$$A\left(\widehat{\boldsymbol{\beta}}^{(m)}\right)\widehat{\boldsymbol{\beta}}^{(m+1)} = A\left(\widehat{\boldsymbol{\beta}}^{(m)}\right)\widehat{\boldsymbol{\beta}}^{(m)} + s\left(\widehat{\boldsymbol{\beta}}^{(m)}\right) \tag{2.21}$$

$$\implies \left(A\left(\widehat{\boldsymbol{\beta}}^{(m)}\right)\widehat{\boldsymbol{\beta}}^{(m+1)}\right)_j = \sum_{s=1}^{p} A_{js}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right)\widehat{\beta}_s^{(m)} + s_j\left(\widehat{\boldsymbol{\beta}}^{(m)}\right)$$

$$= \sum_{s=1}^{p}\sum_{i=1}^{n} W_i x_{ij} x_{is}\widehat{\beta}_s^{(m)} + \sum_{i=1}^{n}\left[\frac{(y_i-\mu_i)}{Var(y_i)}\frac{\partial\mu_i}{\partial\eta_i}x_{ij}\right]^{(m)}$$

$$= \sum_{s=1}^{p}\sum_{i=1}^{n} W_i^{(m)} x_{ij} x_{is}\widehat{\beta}_s^{(m)} + \sum_{i=1}^{n} W_i^{(m)}(y_i-\mu_i^{(m)})\frac{\partial\eta_i^{(m)}}{\partial\mu_i^{(m)}}x_{ij}$$

$$= \sum_{i=1}^{n} W_i^{(m)} x_{ij}\left[\sum_{s=1}^{p} x_{is}\widehat{\beta}_s^{(m)} + \left(y_i-\mu_i^{(m)}\right))\frac{\partial\eta_i^{(m)}}{\partial\mu_i^{(m)}}\right]$$

$$= \sum_{i=1}^{n} W_i^{(m)} x_{ij}\left[\eta_i^{(m)} + (y_i-\mu_i^{(m)})\frac{\partial\eta_i^{(m)}}{\partial\mu_i^{(m)}}\right]$$

Consider the adjusted dependent variable, Z_i defined by

$$Z_i^{(m)} = \eta_i^{(m)} + (y_i-\mu_i^{(m)})\frac{\partial\eta_i^{(m)}}{\partial\mu_i^{(m)}} \tag{2.22}$$

This gives

$$\left(A\left(\widehat{\boldsymbol{\beta}}^{(m)}\right)\widehat{\boldsymbol{\beta}}^{(m+1)}\right)_j = \sum_{i=1}^{n} W_i^{(m)} x_{ij} Z_i^{(m)} \tag{2.23}$$

Also

$$\left(A\left(\widehat{\boldsymbol{\beta}}^{(m)}\right)\widehat{\boldsymbol{\beta}}^{(m+1)}\right)_j = \sum_{s=1}^{p} A_{js}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right)\widehat{\beta}_s^{(m+1)}$$

$$= \sum_{s=1}^{p}\sum_{i=1}^{n} W_i^{(m)} x_{ij} x_{is}\widehat{\beta}_s^{(m+1)} = \sum_{i=1}^{n} W_i^{(m)} x_{ij}\eta_i^{(m+1)}$$

Hence,

$$\sum_{i=1}^{n} W_i^{(m)} x_{ij} Z_i^{(m)} = \sum_{i=1}^{n} W_i^{(m)} x_{ij} \eta_i^{(m+1)} \tag{2.24}$$

$$\implies X^T W^{(m)} Z^{(m)} = X^T W^{(m)} X \boldsymbol{\beta}^{(m+1)} \tag{2.25}$$

These equations are equivalent to an iteratively weighted least-squares estimation procedure with response Z_i, covariates $\mathbf{x}_1, \ldots, \mathbf{x}_p$ and weight function $W_i = \frac{1}{Var(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$. Since $g(\mu_i) = \eta_i$, it can be noted that Z_i is the linearized form of the link function at y_i, up to the 1^{st} order of Taylor expansion as in

$$g(y_i) \cong g(\mu_i) + (y_i - \mu_i) g'(\mu_i) \tag{2.26}$$

The variance is given by

$$Var\left(Z_i^{(m)}\right) \approx \mathrm{Var}(y_i - \mu_i^{(m)}) \left(\frac{\partial \eta_i^{(m)}}{\partial \mu_i^{(m)}} \right)^2 = \mathrm{Var}(y_i) \left(\frac{\partial \eta_i^{(m)}}{\partial \mu_i^{(m)}} \right)^2 = \left(W_i^{(m)} \right)^{-1} \tag{2.27}$$

if $\eta_i^{(m)}$ and $\mu_i^{(m)}$ are considered fixed and known.

2.1.3 Goodness of Fit in GLM

The goodness of fit of a given generalized linear model can be assessed by scaled deviance and Pearson's chi-square statistic [8]. The scaled deviance is defined to be twice the difference between the maximum achievable log likelihood and the log likelihood at the maximum likelihood estimates of the regression parameters.

The scaled deviance is defined by

$$D^*(\mathbf{Y}, \widehat{\boldsymbol{\mu}}) = 2[l(\mathbf{Y}, \mathbf{Y}) - \ l(\mathbf{Y}, \widehat{\boldsymbol{\mu}})] \tag{2.28}$$

Where $l(\mathbf{Y}, \widehat{\boldsymbol{\mu}})$ is the log likelihood maximized over $\boldsymbol{\mu}$ with $\widehat{\boldsymbol{\mu}}$ being maximum likelihood estimates and $l(\mathbf{Y}, \mathbf{Y})$ is the log likelihood attainable in saturated model with $\widehat{\mu}_i = \hat{Y}_i$. For specific distributions, this can be expressed as.

$$D^*(\mathbf{Y}, \widehat{\boldsymbol{\mu}}) = \frac{D(\mathbf{Y}, \widehat{\boldsymbol{\mu}})}{\phi} = 2[l(\mathbf{Y}, \mathbf{Y}) - \ l(\mathbf{Y}, \widehat{\boldsymbol{\mu}})] \tag{2.29}$$

Where D is the deviance. Another popular measure for is the generalized Pearson X^2 statistic given by

$$X^2 = \sum_{i=1}^{n} \left[\frac{(y_i - \widehat{\mu}_i)^2}{V(\widehat{\mu}_i)} \right] \tag{2.30}$$

Both the deviance and the generalized Pearson X^2 follow χ^2_{n-k} distributions for Normal linear models where k is the number of unkonown parameters. For all other GLMs

general asymptotic results are available [10]. However, for finite number of observations, no theoretical results have been established on whether D^* or X^2 is performing better. The deviance has a general advantage as a measure of discrepancy in that it is additive for nested sets of models if maximum likelihood estimates are used, whereas X^2 in general is not. However, X^2 may sometimes be preferred because of its more direct interpretation.

2.2 Poisson Regression for Count Data

Poisson regression is considered as the basic count model upon which most of the other count models are based. If the discrete random variable, Y is Poisson distributed with parameter μ, then Y has density

$$\Pr[Y = y] = \frac{e^{-\mu}\mu^y}{y!}, y = 0, 1, 2, \ldots \tag{2.31}$$

where $E[Y] = Var[Y] = \mu$. By allowing μ to depend on covariates, Poisson regression model is derived. Let, y_i be the ith observed number of occurrences of an event of interest, and $\mathbf{x}_i$ be the vector of linearly independent covariates that are assumed to determine y_i. A regression model based on this distribution is attained by conditioning the distribution of y_i on a vector of covariates, $\mathbf{x}_i = [x_{i0}, \ldots, x_{ip}]^T$. That is, y_i given $\mathbf{x}_i$ is Poisson distributed with density

$$f(y_i|\mathbf{x}_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \ldots \tag{2.32}$$

From where it can be shown that (See later Section 2.2.1.3),

$$E(y_i|\mathbf{x}_i) = Var(y_i|\mathbf{x}_i) = \mu_i \tag{2.33}$$

Here $\mu_i = \mu(\mathbf{x}_i, \boldsymbol{\beta})$ is a continous function of covariates $\mathbf{x}_i$ and parameters $\boldsymbol{\beta}$. The mean variance equality is one main feature of Poisson distribution. For the Poisson MLE, it can be shown [2] that consistency requires correct specification of the conditional mean. It does not require that the dependent variable Y be Poisson distributed. Valid statistical inference using default computed maximum likelihood standard errors and t statistics requires correct specification of both the conditional mean and variance. In most cases, actual variance exceeds estmated mean and overdispersion occurs for practical data. In those situation, we should be able to detect overdisperon and go for alternatives of Poisson regression.

In Section 2.2.1, Poisson regression model is discussed in GLM framework as a member of exponetial family of ditribution with derivition of important properties and Poisson

MLE. In Section 2.2.2, Poisson overdispersion is defined with a brief discussionoo on its detection and available solutions.

2.2.1 Poisson Model in GLM Framework

Poisson model is a member of an exponential family of distributions. Therefore, methods in the framework of the generalized linear model (GLM) [8] are readily applied. In GLM framework, the random component $y_i \sim Poisson\ (\mu_i)$, $i = 1, 2, \ldots, n$. Let the systematic component, linear in covariates be given by

$$\eta_i = \mathbf{x}_i{}^T \boldsymbol{\beta} = \sum_j x_{ij}\beta_j \tag{2.34}$$

And the link function for Poisson model is log-link expressed as

$$g(\mu_i) = \log \mu_i \tag{2.35}$$

The following subsections we will obtain some important relations for Poisson model in GLM framework.

2.2.1.1 As a Member of Exponential Family of Distribution

We need to compare the log pdf of Y_i to exponential family form as in (2.7). From taking log on both sides of (2.32) we can write

$$l_i = \log f(y_i|\mathbf{x}_i) = y_i \log \mu_i - \mu_i - \log y_i! \tag{2.36}$$

Comparing with (2.7)

$$\theta_i = \log \mu_i$$

$$b(\theta_i) = \mu_i = e^{\theta_i}$$

$$a(\phi) = 1$$

$$c(y_i, \phi) = \log y_i!$$

2.2.1.2 Link Function

The link function,

$$g(\mu_i) = \theta_i = \log \mu_i \ = \eta_i = \mathbf{x}_i{}^T \boldsymbol{\beta} \tag{2.37}$$

Hence, the inverse link

$$g^{-1}(\eta_i) = \mu_i = e^{\eta_i} = e^{\mathbf{x}_i^T \boldsymbol{\beta}} \tag{2.38}$$

2.2.1.3 Conditional Mean and Variance

We can get mean and variance of y_i conditioned on covariates from (2.5) and (2.6) as

$$E\left(y_i|\mathbf{x}_i\right) = b'\left(\theta\right) = e^{\theta_i} = \mu_i = e^{\mathbf{x}_i^T\boldsymbol{\beta}} \tag{2.39}$$

$$Var\left(y_i|\mathbf{x}_i\right) = b''\left(\theta\right)a\left(\phi\right) = e^{\theta_i} = \mu_i = e^{\mathbf{x}_i^T\boldsymbol{\beta}} \tag{2.40}$$

Therefore, mean-variance equality holds.

$$E\left(y_i|\mathbf{x}_i\right) = Var\left(y_i|\mathbf{x}_i\right) = \mu_i = e^{\eta_i} = e^{\mathbf{x}_i^T\boldsymbol{\beta}} \tag{2.41}$$

The two components of variance for Poisson GLM are found to be, variance function

$$V\left(\mu_i\right) = \mu_i \tag{2.42}$$

and dispersion parameter

$$a\left(\phi\right) = \phi = 1 \tag{2.43}$$

2.2.1.4 Maximum Likelihood Estimation in Poisson GLM

As we derived in Section 2.1.2, the MLE of $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_p]^T$ under GLM is the solution the system of the score equations given by

$$\mathbf{s}\left(\boldsymbol{\beta}\right) = \left[s_0\left(\boldsymbol{\beta}\right),\, s_1\left(\boldsymbol{\beta}\right), \ldots, s_p\left(\boldsymbol{\beta}\right)\right]^T = \mathbf{0} \tag{2.44}$$

Where the jth score equation is given by

$$s_j\left(\boldsymbol{\beta}\right) = \frac{\partial l(\mathbf{y}, \boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n}\left[\frac{(y_i - \mu_i)x_{ij}}{Var(y_i)}\ \frac{\partial \mu_i}{\partial \eta_i}\right] = 0 \tag{2.45}$$

For $y_i \sim Poisson\left(\mu_i\right)$, we have found μ_i and $Var(y_i)$ in earlier section. From (2.38), since $\mu_i = e^{\eta_i}$ we get

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial\left(e^{\eta_i}\right)}{\partial \eta_i} = e^{\eta_i} = e^{{\mathbf{x}_i}^T\boldsymbol{\beta}} \tag{2.46}$$

Putting the values of μ_i, $Var(y_i)$and $\frac{\partial \mu_i}{\partial \eta_i}$ as functions covariates we get,

$$s_j\left(\boldsymbol{\beta}\right) = \sum_{i=1}^{n}\left[\frac{(y_i - e^{{\mathbf{x}_i}^T\boldsymbol{\beta}})x_{ij}}{e^{{\mathbf{x}_i}^T\boldsymbol{\beta}}}\ e^{{\mathbf{x}_i}^T\boldsymbol{\beta}}\right] = 0$$

$$\implies \sum_{i=1}^{n}\Big[(y_i - e^{{\mathbf{x}_i}^T\boldsymbol{\beta}})x_{ij}\Big] = 0,\ \ j = 0, 1, 2,\ \ldots,\ p \tag{2.47}$$

These system of the score equations are called the MLE equationa of $\boldsymbol{\beta}$ for Poisson GLM. They can solved by using an iterative algorithms such as Newton-Raphson or

Fisher scoring. With Newton-Raphson method, if $\widehat{\boldsymbol{\beta}}^{(m)}$ be the current estimation to the solution of $\mathbf{s}(\boldsymbol{\beta}) = \mathbf{0}$, then the next approximation

$$\widehat{\boldsymbol{\beta}}^{(m+1)} = \widehat{\boldsymbol{\beta}}^{(m)} - \left[\frac{\partial^2 l}{\partial\beta_j \partial\beta_k}\right]^{-1}_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{(m)}} \mathbf{s}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right) \tag{2.48}$$

Where is $\left[\frac{\partial^2 l}{\partial\beta_j \partial\beta_k}\right]$ the matrix of second derivatives of l evaluated at the current iteration and its element for Poisson. From (2.18) we get

$$\begin{aligned}\frac{\partial^2 l}{\partial\beta_r \partial\beta_s} &= \sum_{i=1}^{n}\left[(y_i - \mu_i)\frac{\partial}{\partial\beta_s}\left(\frac{x_{ir}}{Var(y_i)}\frac{\partial\mu_i}{\partial\eta_i}\right) - \frac{1}{Var(y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 x_{ir}x_{is}\right] \\ &= \sum_{i=1}^{n}\left[(y_i - e^{\mathbf{x}_i{}^T\boldsymbol{\beta}})\frac{\partial}{\partial\beta_s}\left(\frac{x_{ir}}{e^{\mathbf{x}_i{}^T\boldsymbol{\beta}}}e^{\mathbf{x}_i{}^T\boldsymbol{\beta}}\right) - \frac{1}{e^{\mathbf{x}_i{}^T\boldsymbol{\beta}}}\left(e^{\mathbf{x}_i{}^T\boldsymbol{\beta}}\right)^2 x_{ir}x_{is}\right] \\ \Longrightarrow \quad \frac{\partial^2 l}{\partial\beta_r \partial\beta_s} &= -e^{\mathbf{x}_i{}^T\boldsymbol{\beta}} x_{ir}x_{is}\end{aligned} \tag{2.49}$$

Fisher scoring method uses expected information $A(\boldsymbol{\beta}) = \left[-E\left(\frac{\partial^2 l}{\partial\beta_r \partial\beta_s}\right)\right]$ instread. From (2.19), the element of $\mathbf{A}(\boldsymbol{\beta})$

$$A_{jk} = -E\left(\frac{\partial^2 l}{\partial\beta_r \partial\beta_s}\right) = +\sum_{i=1}^{n}\left[\frac{1}{Var(y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 x_{ir}x_{is}\right] = e^{\mathbf{x}_i{}^T\boldsymbol{\beta}} x_{ir}x_{is} \tag{2.50}$$

Therefore, new estimation in this method is given by

$$\widehat{\boldsymbol{\beta}}^{(m+1)} = \widehat{\boldsymbol{\beta}}^{(m)} + \mathbf{A}^{-\mathbf{1}}(\widehat{\boldsymbol{\beta}}^{(m)})\mathbf{s}\left(\widehat{\boldsymbol{\beta}}^{(m)}\right) \tag{2.51}$$

It is easily understandable from (2.48)(2.51) that Newton-Raphson and Fisher scoring method are equivalent in case of Poisson. Finally an iterative weighted least-squares estimation procedure can be formed based on any of the two methods [8].

2.2.2 Overdispersed Poisson Model

When the dispersion of the observations is greater than that predicted by the Poisson model, i.e. $Var(Y) > E(Y)$, the phenomenon is called overdispersion or extra-Poisson variation. A major drawback of Poisson regression is the model restricts the variance of the data to be equal to the mean, conditional on explanatory variables. As we have seen, for Poisson regression model, the variance function equals to mean and the dispersion parameter is fixed to one, that is $Var(y) = \phi.V(\mu(\mathbf{x})) = 1.\mu(\mathbf{x}) = \mu(\mathbf{x}) = E(y)$.- Frequently, however the estimated dispersion parameter, after fitting an otherwise acceptable model, exceeds 1. In such cases, Poisson may underestimate the standard errors and overstate the significance of the regression parameters, and consequently, giving misleading inference about the regression parameters. Overdispersion may result

from neglected or unobserved heterogeneity that is not well captured by the covariates in the conditional mean function. To model overdispersion, many alternatives to Poisson have been suggested. The typical approach to handle overdispersion can be looking for a distribution that fits data. Negative binomial regression model have been approached frequently to model overdispersion. For example, if heterogeneity is gamma distributed, then the count data will be negative binomial distributed. However, even when the variance and mean of the assumed negative binomial distribution are correctly specified, if the distribution is not in fact the negative binomial, the maximum likelihood estimator is inconsistent. Moreover, the resulting compound distribution takes no simple form and approximate methods of estimation are often used [3]. In next section we will have a brief introduction on negative binomial regression and two of its forms.

When the probability distribution for random variable cannot be properly specified with any known formal model, the method of MLE cannot give estimates of the parameters. Wedderburn's Quasi-likelihood model allows the mean-variance relationships to be relaxed by letting variance to be inflated by a proportionality constant ϕ [11]. For moddleing overdispersion, it is suggested that

$$Var\left(y\right) = \phi\mu(\mathbf{x}) \tag{2.52}$$

where ϕ, the dispersion parameter is assumed to be a constant greater than 1. ϕ can be estimated using

$$\hat{\phi} = \frac{X^2}{n-p} = \frac{1}{n-p}\sum_{i=1}^{n}\left[\frac{\left(y_i - \widehat{\mu}_i\right)^2}{\widehat{\mu}_i}\right] \tag{2.53}$$

Where the generalized Pearson X^2 statisitic. Many forms of variance function and dispersion parameter and also their combination have been suggested in literature to model overdispersion in data and detect the significance of overdispersion to take into account. Cameron and Trivedi introduced the general formulation $Var\left(y\right) = \mu(\mathbf{x}) + \lambda\mu(\mathbf{x})^l$ for given l where the distribution for y, may not necessarily be specified [6]. NB-2 variance function given by $Var\left(y\right) = \mu(\mathbf{x}) + \lambda\mu(\mathbf{x})^2$ is a special case with $l = 2$ which is found to be the variance for traditional negative binomial regression. The value $l = 1$ corresponds to the NB-1 that results in variance being linearly proportional to mean, $Var\left(y\right) = \mu\left(\mathbf{x}\right)\left(1+\lambda\right)$ that is equivalent to $Var\left(y\right) = \phi\mu(\mathbf{x})$ in (2.52). Lawless explained the functional form of NB-2 variance function can arise from a mixed Poisson model, which more general than negative binomial distribution if some assumptions are relaxed [12]. Lee and Nelder combined the two approaches and proposed the variance function of the form $Var\left(y\right) = \phi\mu\left(\mathbf{x}\right) + \lambda\mu(\mathbf{x})^2$ [13]. All these approaches use log-link function and same form of mean but uses different forms of variance function resulting in varying result in estimation and inference.

Among the many reasons for overdispersion are an incorrect model, an incorrect distributional specification, incorrect variance functions, positive correlation among the

observations, and so forth. In short, correcting an overdispersion problem, if it exists, requires the appropriate remedy. For example, Poisson-distributed data appear overdispersed relative to a Poisson model with regressors when an important regressor is omitted. If the overdispersion is not significant enough and Poisson assumption is still acceptable, Poisson regression is more desirable for its features. It is very necessary to have a systematic way to determine when they apply. Therefore, in this work, formal overdispersion tests and related statistics for detecting deviation from Poisson assumption have been discussed that call for alternative approaches if overdispersion indecating parameter is found to be significant. These tests will be discussed in Chapter 3. In the next section of background, we are going to get a brief introduction on negative binomial regression, which are alternative to Poisson regression for count data in presence of overdispesion, to understand the tests discussed in this work.

2.3 Negative Binomial Regression

When the variance of Y is appears to increase faster than the Poisson model allows, an option is to fit a parametric model that is more dispersed than the Poisson. A popular choice is the negative binomial model. The negative binomial model is based on an underlying probability distribution function (PDF) similarly as the Poisson model. The traditional negative binomial regression model, termed as NB-2 by Cameron and Trivedi [6], is derived from a Poisson-gamma mixture distribution.

Suppose that $y_i \sim Poisson\left(\nu_i \mu_i\left(\mathbf{x}_i, \boldsymbol{\beta}\right)\right)$. Now if each ν_i is assumed to follow a gamma distribution with $E\left(\nu_i\right) = 1$ and $Var\left(\nu_i\right) = \alpha$, the Marginal distribution of y_i given $\mathbf{x}_i$ has the PDF [3, 12],

$$f\left(y_i, \mu_i, \alpha \mid \mathbf{x}_i\right) = \frac{\Gamma\left(y_i + \alpha^{-1}\right)}{y_i! \Gamma\left(\alpha^{-1}\right)} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}}, y_i = 0,\, 1,\, 2, \ldots \tag{2.54}$$

That is then $y_i \sim NB\left(\mu_i\left(\mathbf{x}_i,\, \boldsymbol{\beta}\right), \alpha\right)$. Here mean $E\left(y_i|\mathbf{x}_i\right) = \mu_i$ and variance $Var\left(y_i|\mathbf{x}_i\right) = \mu_i(1 + \alpha\mu_i)$ which is shown later in Section 2.3.2. Here α can account for inequality of mean and variance as compared to Poisson. If $\alpha \to 0$, than y_i becomes standard Poisson distributed random variable [2, 3]. For NB-2 regression, log-link function is assumed same as Poisson. That is

$$g\left(\mu_i\right) = \log \mu_i \; = \eta_i = \mathbf{x}_i{}^T \boldsymbol{\beta} \tag{2.55}$$

Therefore, from the inverse of link, the relationship between mean and covariates can be written as

$$E\left(y_i|\mathbf{x}_i\right) = \mu_i = g^{-1}\left(\eta_i\right) = e^{\eta_i} = e^{\mathbf{x}_i^T \boldsymbol{\beta}} \tag{2.56}$$

And the variance

$$Var\left(y_i|\mathbf{x}_i\right) = \mu_i \;\left(1 + \alpha\mu_i\right) = e^{\mathbf{x}_i^T \boldsymbol{\beta}} \left(1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}\right) \tag{2.57}$$

The score for NB-2 can be obtained from mean and variance using (2.14),

$$s_j(\boldsymbol{\beta}) = \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \left[\frac{(y_i - e^{\mathbf{x}_i^T \boldsymbol{\beta}} x_{ij}}{\left(1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}\right)} \right] \tag{2.58}$$

Where $j = 0,\ 1, 2, \ldots, p$. Thus, the MLE of $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_p]^T$ is the solution the system of the score equations given by $\mathbf{s}(\boldsymbol{\beta}) = \mathbf{0}$.

NB-2 is not directly derived from the canonical form of negative binomial pdf in (2.54). The canonical negative binomial is deonted as NB-C [3]. In expressing the NB-2 as a GLM, first NB-C must be derived with its canonical link. Derivations related to NB-C is presented in Section 2.3.1. An NB-2 model is created by converting the canonical link and inverse link in the GLM derived directly from NB-C to log-link form (See Section 2.3.2).

It is an important feature of NB-2 that it can be estimated using a standard maximum likelihood function, or it can be estimated as a member of the family of generalized linear models (GLM). A negative binomial model is a GLM only if its heterogeneity parameter, α is entered into the generalized linear models algorithm as a constant. There are very good reasons to prefer the NB-2 parameterization of the negative binomial, primarily because it is suitable as an adjustment for Poisson overdispersion. The NB-C form is not interpretable as a Poisson type model, even though it is the canonical form derived directly from the PDF. There are many other forms. Various forms of negative binomial distibution and regression models can be found in detail in [3].

2.3.1 NB-C GLM

Now let us derive the canonical form of the negative binomial GLM derived directly from the PDF in (2.54) symbolized as NB-C. From the definition of gamma function,

$$\Gamma\left(y_i + \alpha^{-1}\right) = \Gamma\left(\alpha^{-1}\right) \prod_{t=0}^{y_i - 1} \left(t + \alpha^{-1}\right),$$

$$\implies \frac{\Gamma\left(y_i + \alpha^{-1}\right)}{\Gamma\left(\alpha^{-1}\right)} = \prod_{t=0}^{y_i - 1} \left(t + \alpha^{-1}\right) \tag{2.59}$$

Taking log on both sides,

$$\log \frac{\Gamma\left(y_i + \alpha^{-1}\right)}{\Gamma\left(\alpha^{-1}\right)}$$

$$= \sum_{t=0}^{y_i - 1} \log\left(t + \alpha^{-1}\right)$$

$$= \sum_{t=0}^{y_i-1} \log\alpha^{-1}\left(1+\alpha t\right)$$

$$= \sum_{t=0}^{y_i-1} \log\alpha^{-1} \quad + \sum_{t=0}^{y_i-1} \log\left(1+\alpha t\right)$$

$$= y_i \log\alpha^{-1} + \sum_{t=0}^{y_i-1} \log\left(1+\alpha t\right)$$

$$= -y_i \log\alpha + \sum_{t=0}^{y_i-1} \log\left(1+\alpha t\right)$$

Therefore,

$$\log\frac{\Gamma\left(y_i+\alpha^{-1}\right)}{\Gamma\left(\alpha^{-1}\right)} = -y_i\log\alpha + \sum_{t=0}^{y_i-1}\log\left(1+\alpha t\right) \tag{2.60}$$

Taking log on both sides of (2.54) and substituting (2.60) we get the log density for ith observation,

$$l_i = \log f\left(y_i, \mu_i, \alpha \mid \mathbf{x}_i\right)$$

$$= \sum_{t=0}^{y_i-1} \log\left(t+\alpha^{-1}\right) - \log y_i! + y_i\log\alpha + y_i\log\mu_i - y_i\log\left(1+\alpha\mu_i\right) - \alpha^{-1}\log\left(1+\alpha\mu_i\right)$$

$$= y_i\log\mu_i - \left(y_i+\alpha^{-1}\right)\log\left(1+\alpha\mu_i\right) + y_i\log\alpha + \sum_{t=0}^{y_i-1}\log\left(t+\alpha^{-1}\right) - \log y_i!$$

Summimg for n independent obsrevations, the log likelihood,

$$l\left(\mu,\alpha;\mathbf{y}\right) = \sum_{i=1}^{n} l_i\left(\mu_i,\alpha;y_i\right)$$

$$= \sum_{i=1}^{n}\left[y_i\log\alpha\mu_i - \left(y_i+\alpha^{-1}\right)\log\left(1+\alpha\mu_i\right) + \sum_{t=0}^{y_i-1}\log\left(t+\alpha^{-1}\right) - \log y_i!\right]$$

$$= \sum_{i=1}^{n}\left[y_i\log\frac{\alpha\mu_i}{\left(1+\alpha\mu_i\right)} - \alpha^{-1}\log\left(1+\alpha\mu_i\right) + \sum_{t=0}^{y_i-1}\log\left(t+\alpha^{-1}\right) - \log y_i!\right] \tag{2.61}$$

Comparing with exponential form for known α, we get the link function

$$g\left(\mu_i\right) = \theta_i = \log\frac{\alpha\mu_i}{\left(1+\alpha\mu_i\right)} = -\log\left(\frac{1}{\alpha\mu_i}+1\right) \tag{2.62}$$

Hence inverse link

$$g^{-1}\left(\theta_i\right) = \mu_i = \frac{1}{\alpha(e^{-\theta_i}-1)} \tag{2.63}$$

Again

$$b\left(\theta_i\right) = \frac{1}{\alpha}\log\left(1+\alpha\mu_i\right) \tag{2.64}$$

Mean can be $\mu_i = \frac{1}{\alpha(e^{-\theta_i}-1)}$ checked with $b'(\theta_i)$

$$\mu_i = b'(\theta_i) = \frac{\partial b}{\partial \mu_i}.\frac{\partial \mu_i}{\partial \theta_i} = \frac{\alpha}{\alpha(1+\alpha\mu_i)}.\frac{-1\left(-e^{-\theta_i}\right)}{\alpha(e^{-\theta_i}-1)^2}$$

$$= \frac{1}{1+\alpha\mu_i}.\frac{1}{\alpha(e^{-\theta_i}-1)}\left(1+\frac{1}{(e^{-\theta_i}-1)}\right) = \frac{1}{1+\alpha\mu_i}.\ \mu_i\ .\ (1+\alpha\mu_i) = \mu_i$$

Variance function,

$$V(\mu_i) = b''(\theta_i) = \frac{\partial}{\partial \theta_i}(\mu_i) = \frac{\partial}{\partial \theta_i}\left(\frac{1}{\alpha(e^{-\theta_i}-1)}\right)$$

$$= \frac{-1\left(-e^{-\theta_i}\right)}{\alpha(e^{-\theta_i}-1)^2} = \mu_i\ (1+\alpha\mu_i) = {\mu_i}^2 + \alpha\mu_i$$

The dispersion parameter for NB-C is one.

Let the systematic component, linear in covariates is given by

$$\eta_i = {\mathbf{x}_i}^T\boldsymbol{\beta} = \sum_j x_{ij}\beta_j$$

Hence we get,

$$g(\mu_i) = \theta_i = \log\frac{\alpha\mu_i}{(1+\alpha\mu_i)} = -\log\left(\frac{1}{\alpha\mu_i}+1\right) = \eta_i = {\mathbf{x}_i}^T\boldsymbol{\beta} \tag{2.65}$$

$$E(y_i|\mathbf{x}_i) = \mu_i = \frac{1}{\alpha(e^{-\theta_i}-1)} = \frac{1}{\alpha(e^{-\eta_i}-1)} = \frac{1}{\alpha\left(e^{-{\mathbf{x}_i}^T\boldsymbol{\beta}}-1\right)} \tag{2.66}$$

$$Var(y_i|\mathbf{x}_i) = \frac{e^{-\eta_i}}{\alpha(e^{-\eta_i}-1)^2} = \frac{e^{-{\mathbf{x}_i}^T\boldsymbol{\beta}}}{\alpha\left(e^{-{\mathbf{x}_i}^T\boldsymbol{\beta}}-1\right)^2} \tag{2.67}$$

Finally, the log likelihood becomes as a function of covariates,

$$l(\mu,\alpha;\mathbf{y})$$

$$= \sum_{i=1}^{n}\left[y_i \log {\mathbf{x}_i}^T\boldsymbol{\beta} - \alpha^{-1}\log\left(1+\alpha\frac{1}{\alpha\left(e^{-{\mathbf{x}_i}^T\boldsymbol{\beta}}-1\right)}\right) + \sum_{t=0}^{y_i-1}\log\left(t+\alpha^{-1}\right) - \log y_i!\right]$$

$$= \sum_{i=1}^{n}\left[y_i \log {\mathbf{x}_i}^T\boldsymbol{\beta} + \frac{1}{\alpha}\log\left(1-e^{{\mathbf{x}_i}^T\boldsymbol{\beta}}\right) + \sum_{t=0}^{y_i-1}\log\left(t+\alpha^{-1}\right) - \log y_i!\right] \tag{2.68}$$

To get the score equation we obtain,

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial\left(\frac{1}{\alpha\left(e^{-\eta_i}-1\right)}\right)}{\partial \eta_i} = \frac{e^{-\eta_i}}{\alpha(e^{-\eta_i}-1)^2}$$

Hence the jth score is given by

$$s_j(\boldsymbol{\beta}) = \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n}\left[\frac{(y_i-\mu_i)x_{ij}}{Var(y_i)}\frac{\partial \mu_i}{\partial \eta_i}\right]$$

$$= \sum_{i=1}^{n}\left[\frac{(y_i-\mu_i)x_{ij}}{\frac{e^{-\mathbf{x}_i{}^T\boldsymbol{\beta}}}{\alpha\left(e^{-\mathbf{x}_i{}^T\boldsymbol{\beta}}-1\right)^2}}\frac{e^{-\mathbf{x}_i{}^T\boldsymbol{\beta}}}{\alpha\left(e^{-\mathbf{x}_i{}^T\boldsymbol{\beta}}-1\right)^2}\right]$$

$$= \sum_{i=1}^{n}\left[(y_i - \frac{1}{\alpha\left(e^{-\mathbf{x}_i{}^T\boldsymbol{\beta}}-1\right)})x_{ij}\right]$$

$$= \sum_{i=1}^{n}\left[(y_i + \frac{e^{\mathbf{x}_i{}^T\boldsymbol{\beta}}}{\alpha\left(e^{\eta_i}-1\right)})x_{ij}\right]$$

Where $j = 0,\ 1, 2, \ldots, p$. Thus, for NB-C, similarly the MLE of $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_p]^T$ is the solution the system of the score equations given by $\mathbf{s}(\boldsymbol{\beta}) = \mathbf{0}$ that is

$$s_j(\boldsymbol{\beta}) = \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n}\left[(y_i + \frac{e^{\mathbf{x}_i{}^T\boldsymbol{\beta}}}{\alpha\left(e^{\mathbf{x}_i{}^T\boldsymbol{\beta}}-1\right)})x_{ij}\right] = 0, \quad j = 0, 1, 2\ldots, p \tag{2.69}$$

We also get

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{n}\left[\frac{1}{\alpha^2}\left\{\log\left(1-e^{\mathbf{x}_i{}^T\boldsymbol{\beta}}\right) + \sum_{t=0}^{y_i-1}\frac{1}{(t+\alpha^{-1})^2}\right\}\right] \tag{2.70}$$

$$\frac{\partial^2 l}{\partial \alpha \partial \beta_j} = -\sum_{i=1}^{n}\left[\frac{e^{\mathbf{x}_i{}^T\boldsymbol{\beta}}x_{ij}}{\alpha^2\left(e^{\mathbf{x}_i{}^T\boldsymbol{\beta}}-1\right)}\right] \tag{2.71}$$

$$\frac{\partial^2 l}{\partial \beta_k \partial \beta_j} = -\sum_{i=1}^{n}\left[\frac{e^{\mathbf{x}_i{}^T\boldsymbol{\beta}}x_{ij}x_{ik}}{\alpha\left(e^{\mathbf{x}_i{}^T\boldsymbol{\beta}}-1\right)^2}\right] \tag{2.72}$$

The NB-C form is not interpretable as a Poisson type model even though it is the canonical form derived directly from the PDF [3]. Therefore, NB-2 parameterization is preferred as for adjusting Poisson overdispersion.

2.3.2 NB-2 GLM

For NB-2 regression, log-link function is assumed instead of usual canoocal link in (2.65). That is

$$g(\mu_i) = \theta_i = \log\mu_i \ = \eta_i = \mathbf{x}_i{}^T\boldsymbol{\beta} \tag{2.73}$$

Therefore, from the inverse of link, the relationship between mean and covariates can be written as

$$E(y_i|\mathbf{x}_i) = \mu_i = g^{-1}(\theta_i) = e^{\eta_i} = e^{\mathbf{x}_i^T\boldsymbol{\beta}} \tag{2.74}$$

And the variance function stays

$$Var\left(y_i|\mathbf{x}_i\right) = \mu_i \ \left(1+\alpha\mu_i\right) = e^{\mathbf{x}_i^T\boldsymbol{\beta}} \ \left(1+\alpha e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right) \tag{2.75}$$

Similarly,

$$\frac{\partial\mu_i}{\partial\eta_i} = \frac{\partial\left(e^{\eta_i}\right)}{\partial\eta_i} = e^{\eta_i} = e^{\mathbf{x}_i{}^T\boldsymbol{\beta}} \tag{2.76}$$

So log likelihood in (2.65) substitutes the inverse log-link, $\mu_i = e^{\eta_i} = e^{\mathbf{x}_i^T\boldsymbol{\beta}}$ to get NB-2 log likelihood,

$$l\left(\mu,\alpha;\mathbf{y}\right)$$
$$= \sum_{i=1}^{n}\left[y_i \log\frac{\alpha e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{\left(1+\alpha e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)} - \alpha^{-1}\log\left(1+\alpha e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right) + \sum_{t=0}^{y_i-1}\log\left(t+\alpha^{-1}\right) - \log y_i!\right] \tag{2.77}$$

Hence the jth score is given by

$$s_j\left(\boldsymbol{\beta}\right) = \frac{\partial l}{\partial\beta_j} = \sum_{i=1}^{n}\left[\frac{(y_i-\mu_i)x_{ij}}{Var(y_i)}\frac{\partial\mu_i}{\partial\eta_i}\right]$$
$$= \sum_{i=1}^{n}\left[\frac{(y_i - e^{\mathbf{x}_i^T\boldsymbol{\beta}}\,)x_{ij}}{e^{\mathbf{x}_i^T\boldsymbol{\beta}}\ \left(1+\alpha e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)}e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right]$$
$$= \sum_{i=1}^{n}\left[\frac{(y_i - e^{\mathbf{x}_i^T\boldsymbol{\beta}}\,)x_{ij}}{\left(1+\alpha e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)}\right]$$

Where $j = 0,\ 1,2,\ldots,p$. Therefore, the MLE of $\boldsymbol{\beta} = [\beta_0,\ldots,\beta_p]^T$ for NB-2 is the solution the system of the score equations

$$s_j\left(\boldsymbol{\beta}\right) = \frac{\partial l}{\partial\beta_j} = \sum_{i=1}^{n}\left[\frac{(y_i-\mu_i)x_{ij}}{Var(y_i)}\frac{\partial\mu_i}{\partial\eta_i}\right] = \sum_{i=1}^{n}\left[\frac{(y_i - e^{\mathbf{x}_i^T\boldsymbol{\beta}}\,)x_{ij}}{\left(1+\alpha e^{\mathbf{x}_i^T\boldsymbol{\beta}}\right)}\right] = 0, \quad j = 0,1,2\ldots,p \tag{2.78}$$

The NB-2 heterogeneity parameter, α, can model Poisson overdispersion, or extra correlation. However, The above restructured version (2.78) of NB-C which is called the maximum likelihood parameterizations of the NB-2 is an approximation. However, if only regression is of interest, quasi-likelihood approach with log-link and NB-2 variance is found more easy to handle.

2.4 Partial Score Statistics from Negative Binomial Model

In this section, we will derive the partial score statistic, S which the basis for the tests for detecting Poisson overdispersion. From section (2.3) we recall $y_i \sim Poisson\left(\nu_i\mu_i\left(\mathbf{x}_i,\boldsymbol{\beta}\right)\right)$,

where $E(\nu_i) = 1$ and $Var(\nu_i) = \alpha$, has the conditional variance of the form,

$$Var(y_i|\mathbf{x}_i) = \mu_i + \alpha\mu_i^2 \tag{2.79}$$

We also recall the log likelihood of $y_i \sim NB(\mu_i, \alpha)$ as a function of α.

$$l(\alpha) = \sum_{i=1}^{n} \left[y_i \log \alpha\mu_i - \left(y_i + \alpha^{-1}\right) \log\left(1 + \alpha\mu_i\right) + \sum_{t=0}^{y_i-1} \log\ \left(t + \alpha^{-1}\right) - \log y_i! \right] \tag{2.80}$$

It can be shown that, for $\alpha = 0$, (2.80) reduce to Poisson log likelihood as in (2.34). For detecting negative binomial departures, we let $y_i \sim NB(\mu_i, \alpha)$ as our more general alternative model to test against Poisson null model. The data can be said to have overdispersion for large positive values of α. Therefore, the hypothesis testing terms,

$$H_0 :\ \alpha = 0$$

$$Vs.\ \ H_1 :\ \alpha > 0.$$

Here we somewhat follow the steps by Collings and Margolin [5] to determine a partial score test for this problem. Now, the third term of (2.80) can be written as,

$$\sum_{t=0}^{y_i-1} \log\left(t + \alpha^{-1}\right) = \sum_{t=0}^{y_i-1} \log\alpha^{-1}\left(1 + \alpha t\right) \ = \sum_{t=0}^{y_i-1} \log\alpha^{-1} + \sum_{t=0}^{y_i-1} \log\left(1 + \alpha t\right)$$

$$= -y_i \log\alpha + \sum_{t=0}^{y_i-1} \log\left(1 + \alpha t\right)$$

Replacing the value in (2.80),

$$l(\alpha) = \sum_{i=1}^{n} \left[y_i \log \alpha\mu_i - \left(y_i + \alpha^{-1}\right) \log\left(1 + \alpha\mu_i\right) + -y_i \log\alpha + \sum_{t=0}^{y_i-1} \log\left(1 + \alpha t\right) - \log y_i! \right]$$

$$= \sum_{i=1}^{n} \left[y_i \log \mu_i - \left(y_i + \alpha^{-1}\right) \log\left(1 + \alpha\mu_i\right) + \sum_{t=0}^{y_i-1} \log\left(1 + \alpha t\right) - \log y_i! \right]$$

$$= \sum_{i=1}^{n} \left[y_i \log \mu_i - y_i \log\left(1 + \alpha\mu_i\right) - \alpha^{-1} \left\{ \alpha\mu_i - \frac{\left(\alpha\mu_i\right)^2}{2} + \ldots \right\} + \sum_{t=0}^{y_i-1} \log\left(1 + \alpha t\right) - \log y_i! \right]$$

$$= \sum_{i=1}^{n} \left[y_i \log \mu_i - y_i \log\left(1 + \alpha\mu_i\right) - \mu_i - \frac{\alpha\mu_i^2}{2} + \ldots + \sum_{t=0}^{y_i-1} \log\left(1 + \alpha t\right) - \log y_i! \right]$$

From it we can also show,

$$l(0) = \sum_{i=1}^{n} \left[y_i \log \mu_i - \mu_i - \log y_i! \right]$$

If we consider μ_i's known, then according to classical Neyman-Pearson theory, the most powerful test rejects H_0 for large values of $l(\alpha) - l(0)$.

$$l(\alpha) - l(0) = \sum_{i=1}^{n} \sum_{t=0}^{y_i - 1} [\log(1 + \alpha t) - y_i \log(1 + \alpha \mu_i)]$$

Since the test is not independent of α, there does not exist a uniformly most powerful test of H_0 versus H_1, even when the μ_i's are known. Therefore, a locally most powerful test for H_0 versus H_1 can be obtained. Following Ferguson (p. 235 in [14]), the locally most powerful test rejects H_0 for large values of the score $\frac{\partial l}{\partial \alpha}$ at $\alpha = 0$.

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{n} \left[-\frac{y_i \mu_i}{1 + \alpha \mu_i} - \{ -\frac{\mu_i^2}{2} + \frac{\alpha \mu_i^3}{3} - \dots \} + \sum_{t=0}^{y_i - 1} \frac{\mathrm{t}}{1 + \alpha t} \right]$$

$$\left. \frac{\partial l}{\partial \alpha} \right|_{\alpha = 0} = \sum_{i=1}^{n} \left[-y_i \mu_i + \frac{\mu_i^2}{2} + \sum_{t=0}^{y_i - 1} \mathrm{t} \right] = \sum_{i=1}^{n} \left[\frac{\mu_i^2 - 2 y_i \mu_i}{2} - \frac{y_i (y_i - 1)}{2} \right]$$

$$= \sum_{i=1}^{n} \left[\frac{\mu_i^2 - 2 y_i \mu_i + y_i^2}{2} - \frac{y_i}{2} \right] = \frac{1}{2} \sum_{i=1}^{n} \left[(y_i - \mu_i)^2 - y_i \right]$$

We denote it as the partial score statistic,

$$S = \frac{1}{2} \sum_{i=1}^{n} \left[(y_i - \mu_i)^2 - y_i \right] \tag{2.81}$$

In next chapter, we will discuss some test statistics genrated from S and their approxmate distributional assumption to calculate significance of parameter for overdispersion. Chapter 3

Chapter 3

Testing for Detecting Overdispesion in Count Data

To test Poisson overdispersion, we have to question the mean variance equality. The model for null hypothesis is $Y_i \sim Poisson(\mu_i)$ where $E(Y_i|x_i) = Var(Y_i|\mathbf{x}_i) = \mu_i = e^{\mathbf{x}_i^T\boldsymbol{\beta}}$. The natural approach is using a different form variance with an added parameter for the alternative model maintaining the same mean. Some possible functions are mentioned in section 3.2. In this chapter, the tests suggested by Dean and Lawless [7] will be revised. A random effect, or mixed Poisson, model as alternative for testing overdispersion has been conidered here. A mixed Poisson differs from Poisson by a parameter α which is assumed to handle extra Poisson variation and results in a form of variance not equal to mean. To check the significance of this parameter, four statistics (have been denoted by S_1, S_2, S_a and S_b) are presented here. Also their approximate distributions were discussed and compared.

3.1 Modelling Poisson Overdispersion

Let us consider, $Y_i \sim Poisson(\nu_i\mu_i(\mathbf{x}_i, \boldsymbol{\beta}))$ where each ν_i has finite first and second moment and also $E(\nu_i) = 1$ and $Var(\nu_i) = \alpha$. Then Y_i can be expressed to have a mixed Poisson distribution with the conditional mean and variance of Y_i on covariates of the forms [7],

$$E(Y_i|\mathbf{x}_i) = \mu_i \tag{3.1}$$

$$Var(Y_i|\mathbf{x}_i) = \mu_i + \alpha\mu_i^2 \tag{3.2}$$

To test Poisson model against the mixed Poisson alternative with variance as (3.2) that include with overdispersion, Dean and Lawless [7] proposed the hypotheses

$$H_0 : \ \alpha = 0$$

$$Vs.\ \ H_1:\ \alpha > 0.$$

Collings and Margolin also developed their tests againsmt same variance in (3.2) by working from negative binomial models. We already have discussed, if ν_i's are assumed to follow a gamma distribution, then y_i has a negative binomial distribution and this form of variance is also known as NB2 variance. In next section (3.2) are going to discuss some test statistics derived for the mentioned pair of null and alternative models. We will focus on different forms of statistics presented that require only the null model i.e., Poisson model to be fitted.

3.2 Test Statistics and Limiting Distributions

Four statistics for detecting overdispersion that are presented here have originated from partial score statistic for testing overdispersion, S as in (2.81). We have already shown its derivation for negative binomial likelihood in Section 2.4. This statistic was obtained from partial score or partial derivative of log-likelihood of negative binomial by Collings and Margolin for detecting negative binomial deviation from Poisson [5]. Cameron-Trivedi [6], Lee [15] and Dean-Lawless [7] discussed same statistic S for mixed Poisson alternatives. We recalll the partial score statistic

$$S = \frac{1}{2}\sum_{i=1}^{n}\left[(Y_i - \widehat{\mu}_i)^2 - Y_i\right] \tag{3.3}$$

where $\widehat{\mu}_i = \mu_i\left(\mathbf{x}_i, \widehat{\boldsymbol{\beta}}\right)$, with $\widehat{\boldsymbol{\beta}}$, the maximum likelihood estimate of $\boldsymbol{\beta}$ under the null hypothesis, i.e., Poisson model. This statistic, which is given as T by Dean and Lawless [7], generalizes Collings and Margolin's statistic T_B^{**} (See (21) in [5]).

The statistic S_1 (T_1 in [7]) was shown by Dean and Lawless as suitably standardized versions of S. S_2 is shown by Dean and Lawless (T_2 in [7]) as a generalization of both Fisher dispersion statistic for IID Poisson models [16] and the T_B and T_C statistics given by Collings and Margolin for two special cases in [5]. Finally T_a and T_b, the adjusted score statistics presented by Dean and Lawless [7] are also discussed here. We denoted them as S_a and S_b respectively.

3.2.1 S_1

Large values of S indicate overdispersion in data as shown earlier. From standard maximum likelihood large-sample theory, its limiting distribution for $n \to \infty$ can be approximated. Dean and Lawless shows that, when $\alpha = 0$, $S/\sqrt{\frac{1}{2}\sum_{i=1}^{n}\widehat{\mu}_i^2}$ converges in distribution to a standard normal random variable as $n \to \infty$ (See T_1 in Sec 2 of [7]). Therefore, $S/\sqrt{\frac{1}{2}\sum_{i=1}^{n}\widehat{\mu}_i}$ is an asymptotically equivalent standardized statistic.

We denote it S_1.

$$S_1 = \sum_{i=1}^{n} \frac{\left[(y_i - \widehat{\mu}_i)^2 - y_i\right]}{\sqrt{2\sum_{i=1}^{n} \widehat{\mu}_i^2}} \tag{3.4}$$

An R function of for calculating S_1 is given is Appendix A (See A.1.1).

3.2.2 S_2

When n is fixed and the μ_i's $\to \infty$, Collings and Margolin suggested a test of $\alpha = 0$ based on

$$S_2 = \sum_{i=1}^{n} \frac{(Y_i - \widehat{\mu}_i)^2}{\overline{Y}} \tag{3.5}$$

where $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. This statistic is a natural generalization of both the Fisher's dispersion statistic for IID Poisson models [16] and the statistics given by Collings and Margolin for two special cases [5]. An R function of for calculating S_1 is given is Appendix A (See A.1.1). To find limiting distribution of S_2, Dean and Lawless assumed n to be fixed and let $\mu_+ = \sum_i \mu_i$ and μ_i's $\to \infty$; $(i = 1, \ldots, n)$ such that μ_i/μ_+ converges to a positive constant [7]. They let,

$$S_3 = n\sum_{i=1}^{n} \frac{(y_i - \widehat{\mu}_i)^2}{\mu_+} = n\mu_+^{-1}(\mathbf{Y} - \widehat{\boldsymbol{\mu}})^T (\mathbf{Y} - \widehat{\boldsymbol{\mu}}) \tag{3.6}$$

Where $\mathbf{Y} = [Y_1, \ldots, Y_n]^T$ and $\widehat{\boldsymbol{\mu}} = [\widehat{\mu}_1, \ldots, \widehat{\mu}_n]^T$. S_3 has the same limiting distribution as S_2 when $\tau = 0$ since $\mu_+^{-1}\sum_{i=1}^{n} Y_i$ converges in probability to 1. Dean and Lawless showed that

$$\mu_+^{-\frac{1}{2}} (\mathbf{Y} - \widehat{\boldsymbol{\mu}}) \xrightarrow{D} N(\mathbf{0}, V) \tag{3.7}$$

Where

$$V = \mu_+^{-1} W^{\frac{1}{2}} (I - H) W^{\frac{1}{2}} \tag{3.8}$$

$$W = diag(\mu_1, \ldots, \mu_n) \tag{3.9}$$

$$H = W^{\frac{1}{2}} X \left(X^T W X\right)^{-1} X^T W^{\frac{1}{2}} \tag{3.10}$$

V and H asymptotic covariance and leverage matrices for Poisson regression [17, 18]. From (3.7), the asymptotic null distribution of $S_3 = n\mu_+^{-1}(\mathbf{Y} - \widehat{\boldsymbol{\mu}})^T (\mathbf{Y} - \widehat{\boldsymbol{\mu}})$ or S_2 is a linear combination of χ_1^2 independent random variables (See Section 3b in [10]). In particular, S_2 converges to $\sum_{i=1}^{n} \lambda_i U_i$ where the U_i's are χ_1^2 independent random variables and the λ_i's are the n eigenvalues of V [7]. This limiting distrubution gives, as special cases, the limiting distributions of the mentioned statistics obtained by both Fisher and Collings-Margolin.

3.2.3 Adjusted Score Statistic S_a

Dean and Lawless adjusted the statistic $S_1 = \sum_{i=1}^{n} \frac{\left[(Y_i-\widehat{\mu}_i)^2-Y_i\right]}{\sqrt{2\sum_{i=1}^{n}\widehat{\mu}_i^2}}$ for more adequate distributional assumption [7]. From (3.7), it can be show that $E\left\{(Y_i-\widehat{\mu}_i)^2\right\} = (1-h_{ii})\,\mu_i$ where h_{ii} is the ith diagonal element of leverage matrix, H. The leverages account for the fact that the expectation of residual sum of squares $\sum_i^n (Y_i-\widehat{\mu}_i)^2$ is slightly less than $\sum_i^n \mu_i$ after estimation of regression coefficients. We can write $E\left\{(Y_i-\widehat{\mu}_i)^2 - Y_i\right\} = -h_{ii}\mu_i$. Therefore they made first-order adjustment to the numerator of S_1, to obtain

$$S_a = \sum_{i=1}^{n} \frac{\left[(Y_i-\widehat{\mu}_i)^2 - Y_i\right] + \hat{h}_{ii}\widehat{\mu}_i}{\sqrt{2\sum_{i=1}^{n}\widehat{\mu}_i^2}} \tag{3.11}$$

An R function of for calculating S_a is given is Appendix A (See A.1.3). S_a has the approximate distribution $N(0,1)$ and is found to converge to normality considerably faster than S_1[7].

3.2.4 Adjusted Score Statistic S_b

In section 3.2.2, we saw that S_2 has a limiting distribution $\sum_{i=1}^{n}\lambda_i U_i$ where the U_i's are χ_1^2 independent random variables and the λ_i's are the n eigenvalues of V. To use S_2, one can approximate $Pr\{S_2 > t\}$ by inserting estimates $\widehat{\mu}_i$ in equations (3.8)-(3.10) to obtain V. From which the eigenvalues of V, λ_i's are found and probability for a linear combination of χ_1^2 independent random variables can be calculated (method in [19]).

Dean and Lawless presented two-moment $c\chi_d^2$ approximation as follows [7]. The mean and variance of S_2's limiting distribution are, from equations (3.8)-(3.10) as,

$$Asmean(S_2) \;=\; n\;tr(V) \;= nu_+^{-1}\sum_{i=1}^{n}(1-h_{ii})\,\mu_i \tag{3.12}$$

$$Asvar(S_2) \;= 2n^2\;tr(V^TV) \;= 2n^2u_+^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(1-h_{ij})^2\mu_i\mu_j \tag{3.13}$$

Where $tr()$ denotes trace of a matrix. Now keeping the asymptotic mean and variance correct, c and d for $S_2 \sim c\chi_d^2$ can be approximately taken as

$$c = nu_+^{-1}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(1-h_{ij})^2\mu_i\mu_j}{\sum_{i=1}^{n}(1-h_{ii})\,\mu_i} \tag{3.14}$$

$$d = nu_+^{-1}\frac{\{\sum_{i=1}^{n}(1-h_{ii})\,\mu_i\}^2}{\sum_{i=1}^{n}\sum_{j=1}^{n}(1-h_{ij})^2\mu_i\mu_j} \tag{3.15}$$

A simpler approximation that implements Wilson-Hilferty transformation for χ^2 variables (See Chap. 17 of [20]) can be used when it is known that $\hat{d} \geq 10$ approximately [7]. The approximation gives the statistic

$$S_b = \left(4.5\hat{d}\right)^{\frac{1}{2}} \{ \left(\frac{S_2}{\hat{c}\hat{d}} \right)^{\frac{1}{3}} + \frac{2}{9\hat{d}} - 1 \} \tag{3.16}$$

which is found to follow standard normal distribution. An R function of for calculating S_d is given is Appendix A (See A.1.4). S_b is a convenient statistic when $\hat{d}$ is large enough but when $\hat{d}$ is small one can return to the approximation $S_2 \sim c\chi^2_d$ or as the linear combination of χ^2_1 random variables. However, probablities for both of the approximations for S_2 are complicated to calculate.

Chapter 4

Adequacy of the Distributional Assumptions of the Statistics

4.1 A Simulation Study

A simulation study was conducted to examine the normality assumptions of the statistics S_1, S_a, S_b using R. Codes for generating the results in table 4.1 are discussed in Appendix A (See A.2 and A.3). To demonstrate the distributional assumptions of the statistics, we simulated some samples based on a Poisson regression model, $\mu_i = \exp(\beta_0 + \beta_1 x_{1i})\,; (i = 1, \ldots, n)$, for different values of n with letting $[\beta_0, \beta_1] = [2.6, 2]$. The value of covariate x_{1i} was obtained by picking values randomly from the interval (0,1). Here for $n = 20, 30, 50, 100, 200, 300, 500$ were chosen for S_1 and $n = 20, 30, 50, 100, 200$ for S_a, S_b and X^2. For a given value of n the same covariate value was used for simuation of each sample.

4.2 Results

Table 4.1 shows simulation results for S_1, S_a, S_b and for comparison the generalized Pearson goodness-of-fit statistic X^2. 1000 samples were generated for each n and S_1, S_a, S_b and X^2 are calculated. Table 4.1 shows the proportion of the time that S_1 S_a, S_b exceeded the standard normal upper 20%, 10%, 5%, and 1% points (say upper tail probability = 0.20. 0.10, 0.05 and 0.01), and the proportion of the time that X^2 exceeded the corresponding upper percentage points for the χ^2_{n-2} distribution. Histograms from the generated samples of S_1 S_a, S_b for different values n are presented in Figures 4.1-4.3 repectively to get a view on their adequacy with normality assumption.

In this work, we reproduced some results of Dean and Lawless up to some higher values of n. The results show the approximation of S_1 for the range of n considered are not

TABLE 4.1: Estimated Upper Tail Probabilities for S_1, S_a, S_b and X^2 based on 1,000 Samples of the Poisson regression model

		Upper Tail Probability (UTP) from Asymptotic Distribution			
	n	**0.2**	**0.1**	**0.05**	**0.01**
S_1	20	0.104	0.049	0.032	0.01
	30	0.125	0.061	0.037	0.011
	50	0.127	0.072	0.039	0.01
	100	0.142	0.082	0.046	0.011
	200	0.156	0.077	0.04	0.012
	500	0.149	0.076	0.033	0.008
S_a	20	0.144	0.075	0.043	0.012
	30	0.149	0.079	0.041	0.013
	50	0.165	0.094	0.058	0.013
	100	0.191	0.102	0.059	0.011
	200	0.197	0.109	0.051	0.011
S_b	20	0.18	0.096	0.051	0.009
	30	0.186	0.102	0.048	0.016
	50	0.212	0.114	0.068	0.014
	100	0.19	0.088	0.05	0.014
	200	0.193	0.119	0.056	0.011
X^2	20	0.2008	0.1032	0.0478	0.0078
	30	0.209	0.107	0.0466	0.011
	50	0.208	0.1008	0.0478	0.0082
	100	0.2084	0.106	0.0518	0.008
	200	0.2034	0.0924	0.0464	0.0106

that satisfactory. The values are much less from the expected upper tail probablities for standard normal specially for expected upper tail probablities greater than 0.1. From the histograms in Figure 4.1 for S_1, it is easily visible that the generated distributions are positively skewed for all vallues of n; though the skewness is observed to decrease with increasinng n. As Dean and Lawless suggested, S_a is not very good until n becomes fairly large. Closely observing both the results in Table 4.1 and the histograms in FIgure 4.2, we can conclude normality assumptions to be pretty accurate from $n = 100$. S_b shows quicker convergence to standard normal distribution than S_a. The histograms in Figure 4.3 portrays the statistic S_b to come from an normality distribution for any value of n. For both S_b and X^2 the asymptotic approximations were almost accurate even for smaller values of n.

4.3 Discussion

Among the test statistics discussed here, both test statistics S_a and S_b can be recommended for their own advantages. Definitely both are preferred from S_1 for faster

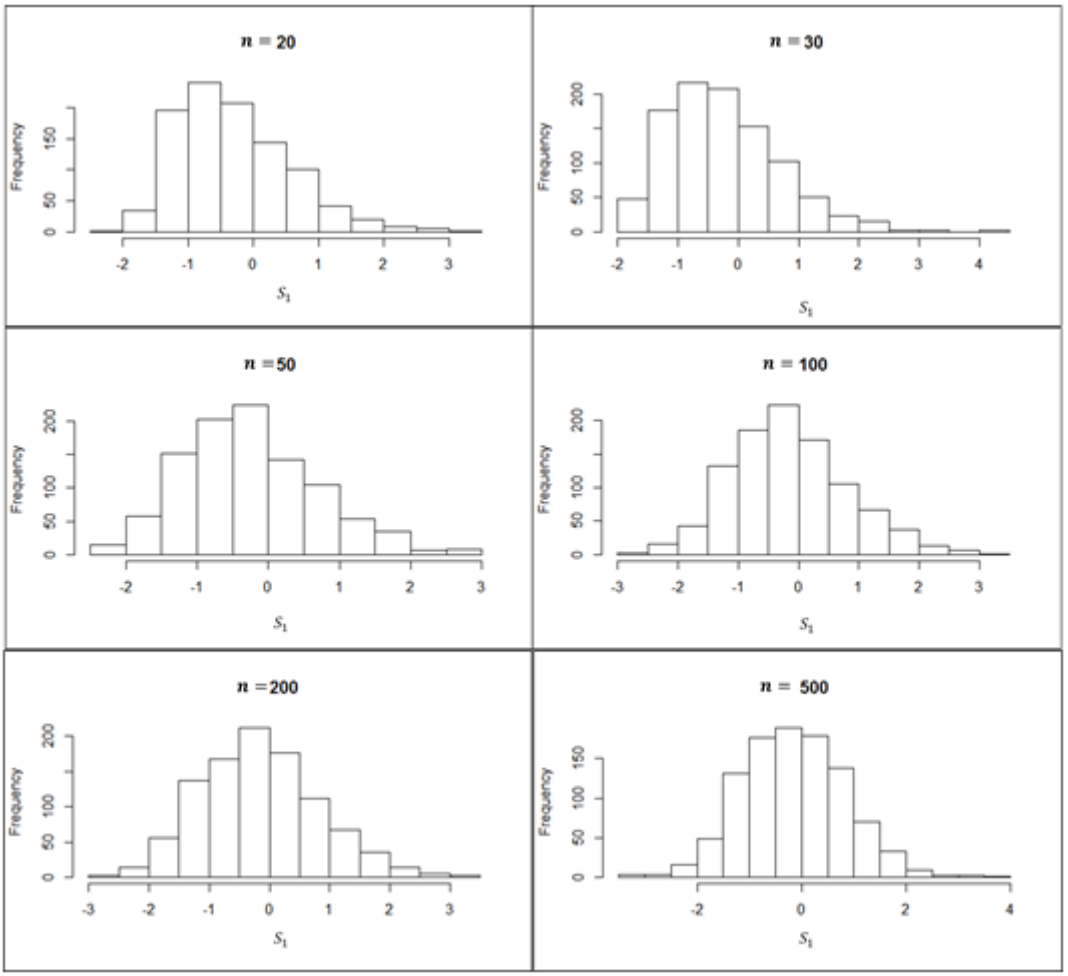

FIGURE 4.1: Histograms of S_1 for $n = 20, 30, 50, 100, 200, 500$ from 1000 samples each.

convergence. In our simulation study, we have not demonstrated the asymptotic distribution of S_2 as linear combination of Chi square random variables. In Figures 4.1-4.3, adequacy of the standard normal asumptions of S_1, S_a and S_b are dispalyed in histograms of the statistics for simualted observations of different sample sizes. repectively. Both S_a and S_b evolved from a score test statistic S for no overdispersion against negative binomial alternatives although they come from different asymptotic cases.

S_a has been derived from the adjustment of S_1. S_1 comes from the asymptotic case of S with the assumption $n \to \infty$. If we compare the results for S_1 and S_a, it is clear that the adjustment from S_1 to S_a definitely has resulted in faster convergence to normality. The normality assumption for S_a was found fairly accurate from around $n = 100$. However, that for S_1 was not satisfactory even for $n = 500$.

On the other hand, S_b originates from S_2 which was derived from the cases when n is fixed and the μ_i's $\to \infty$. Therefore it is expected to have similar accuracy in approximate distribution for varying n. According to results, the normality assumption for S_b is found to be precise for the range of n considered and for this particular Poisson regression

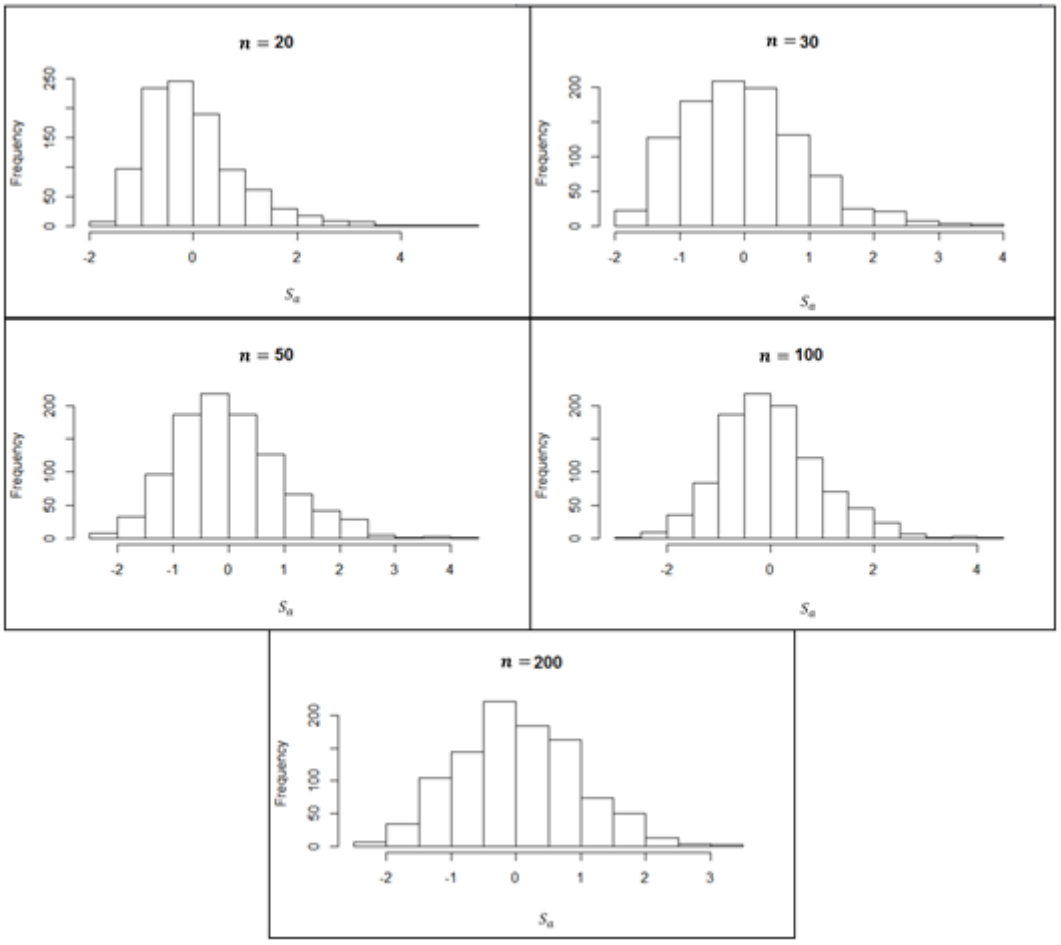

FIGURE 4.2: Histograms of S_a for $n = 20, 30, 50, 100, 200$ from 1000 samples each.

equation. Definitely the data generating regression equation has decided the range of μ_i's. Therefore, it will be interesting to analyze the overall range of μ_i for which the normality assumption for S_b acceptably holds and to compare with other statistics.

From the discussion we can say that S_b shows robustness in approximating standard normal and it is quite insensitive to the value of n. On the other hand the approximation for S_a is preferred for its computational simplicity compared to S_b. It only need the leverage values h_{ii}, which offer useful information from the leverage matrix H [17, 18]. Calcuating S_b requires a little more work, because the $n(n-1)/2$ distinct values h_{ij}, where $(i \neq j)$ are also needed. Also its calculation consists of more number of equations than S_a. As Dean and Lawless said and our results agreed; if $n - p$ is 50 or more, than S_a can be used in conjunction with the normal approximation for computing significance levels. Otherwise, the use of S_b with the normal approximation is recommended. Both S_1 and S_2 are attractive for their computationally simple form. However, calculating the probability for approximate distribution S_2 is tedious as it ir linear combination of chi squares. Fo S_1 the normality assumption was not found acceptable for $n \leq 500$. However, it should be checked for large sized samples.

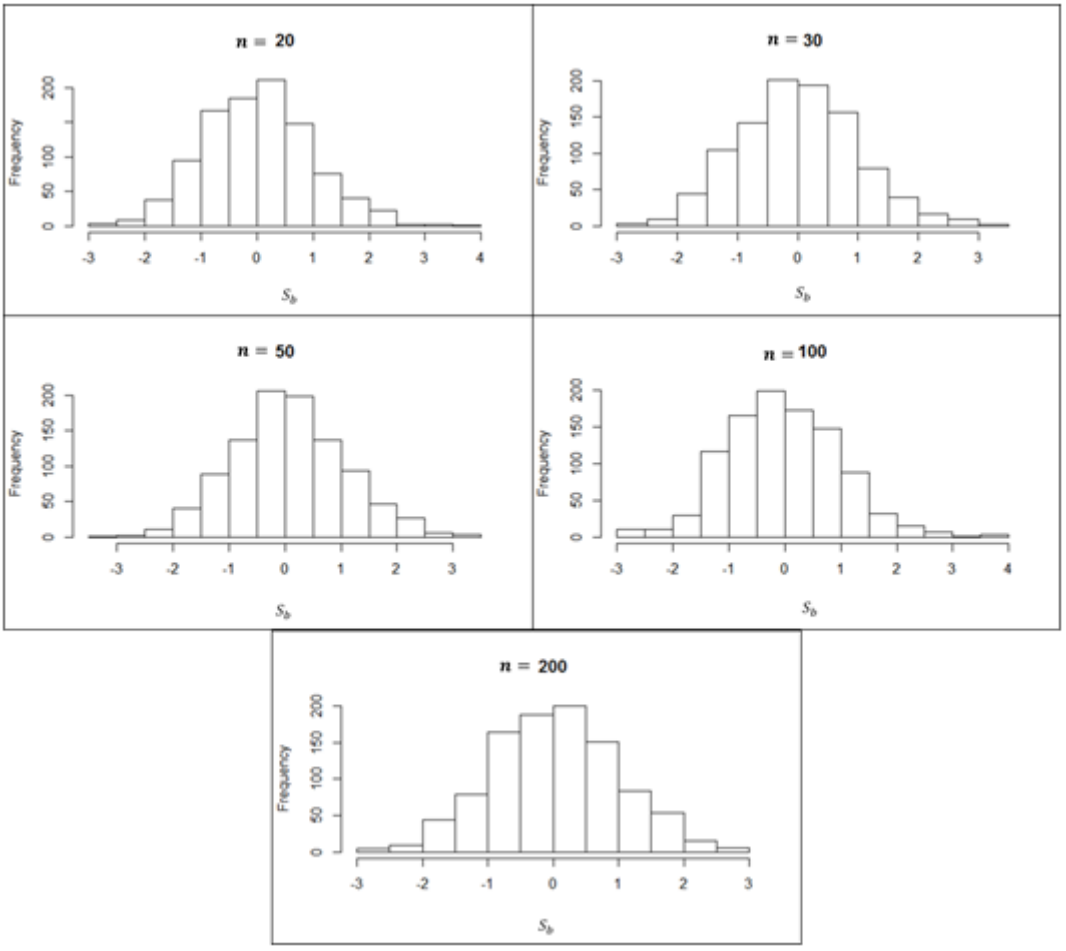

FIGURE 4.3: Histograms of S_1 for $n = 20, 30, 50, 100, 200$ from 1000 samples each.

Though Pearson's goodness of fit statistic X^2 has shown good asymptotic approximations for all values of n and is less complicated than S_a and S_b, it is not particularly formulated for detecting overdispersion as represented by the negative binomial alternatives like the other discussed here. It rather serve for assessing the goodness of the specification for μ_i's in terms of covariates within the Poisson model.

The next chapter gives some further remarks and comparisons on the advantages and applicability of the statistics discussed.

Chapter 5

Additional Comments on the Statistics

Here we present some comparisons and points about the applicability of the discussed tests.

5.1 Advantages Over Likelihood Ratio and Wald Test

Likelihood ratio test and Wald test or the approximate large sample normality of the maximum likelihood estimator $\widehat{\alpha}$ to test $\alpha = 0$, have been suggested in literarture [12, 21]. Such approach requires fitting mixed Poisson or some other more general models that include parameter for incorporating overdispersion. The tests discussed here do not require a more comprehensive model than the Poisson to be fitted and accurate distributional approximations are available for computing significance levels. Also results by Lawless in [12] indicate that unless n or the μ_i's are rather large, the usual χ^2 and normal approximations for the likelihood ratio statistic and $\widehat{\alpha}$, respectively, tend to underestimate substantially the evidence for overdispersion.

5.2 Power comparison of S_a, S_b, X^2 and Deviance Statistic

Both the goodness of fit statistics, Pearson's Chi square, X^2 and deviance D frequently found application in detecting overdispersion. A power comparison for S_a, S_b, X^2 and D for some negative binomial alternatives was performed by Dean and Lawless (Section 5 in [7]). The test hypothesis presented in Sec. 3.1 was used. For $n = 15$ and test size of 0.05, power of test for Poisson regression models similar to the one we considered were calculated for different chosen values of α. According to their findings, S_a and S_b have nearly equal power and that $X^2$and D have nearly equal power. Also S_a and S_b up to a

certain point are found more powerful than X^2 or D. Finally, as α increases, the powers for all of the four test statistics approach 1. The fact that X^2 and D are not particularly intended to test for overdispersion as represented by the negative binomial alternatives discussed here, but rather for assessing the adequacy of the specification for μ_i's within the Poisson model explain why they are somewhat less powerful than S_a and S_b.

5.3 Correct Specification of Mean

The tests for overdispersion discussed here were developed under the assumption that the specification of μ_i in terms of the covariates is correct [7] and all important regressors are present. When regression models are fitted, they should include a careful examination of residuals. The overall lack of fit of a Poisson model can be primarily ascribed either to overdispersion or to misspecification of μ_i. In this regard, Pierce-Schafer [22] and Williams [17] suggested that deviance residuals are preferred for assessing fitted Poisson model. The accurate specification of μ_i before testing for overdispersion should be attempted to be achieved. If there are replicate observations for at least some of the covariate values, both overdispersion and the μ_i specification can be separately checked. If we have independent observations y_{ij}'s are such that

$$E\left(y_{ij} \mid x_i\right) = \mu\left(\mathbf{x}_i, \boldsymbol{\beta}\right) = \mu_i$$

And

$$Var\left(y_{ij} \mid x_i\right) = \mu\left(\mathbf{x}_i, \boldsymbol{\beta}\right) = \mu_i + \alpha\mu_i^2$$

Then the total S_2 statistic of (3.5) can be decomposed as

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\left(Y_{ij}-\widehat{\mu}_i\right)^2}{\overline{Y}_{++}} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\left(Y_{ij}-\overline{Y}_{i+}\right)^2}{\overline{Y}_{++}} + \sum_{i=1}^{k}\frac{n_i\left(\overline{Y}_{i+}-\widehat{\mu}_i\right)^2}{\overline{Y}_{++}},$$

Where $\overline{Y}_{i+} = \sum_{j=1}^{n_i} Y_{ij}/n_i$ and $\overline{Y}_{++} = \sum_{i=1}^{k} n_i\overline{Y}_{i+}/\sum_{i=1}^{k} n_i$. The first term on the right side is the statistic T_C studied by Collings and Margolin (Section 3 in [5]) can be used to test $\alpha = 0$ versus $\alpha > 0$ with no assumption about the parametric form of μ_i. If one is certain of the regression specification it may be preferable to use the entire S_2, particularly if the μ_i's are small. For larger μ_i's and when μ_i specifications are not certain, it is better to use T_C.

5.4 Limiting Distributions for S_a and S_b are Unconditional

The limiting distributions of S_a and S_b were unconditional. That means, they are not conditional on observed values of sufficient statistics for the Poisson model parameters.

In a few special cases exact conditional distributions that are parameter-free and appropriate for S_2 are available[5, 16]. However, when the μ_i's $\to \infty$ with n fixed, conditional and unconditional limiting distributions are the same [7]. So S_b can be suggested for both conditional and unconditional significance levels. For very small sample sizes the exact distribution of S_2 can be considered, but except for the aforementioned cases the distributions either depend on unknown model parameters or are not amenable to calculation.

Chapter 6

Conclusion

The problem of detecting Poisson overdispersion still does not have a generally established solution. Popular approach is to fit negative binomial or some other general model that can accommodate overdispersion without checking whether the parameter indicative to overdispersion is significant or not. However, there is no doubt about the strong theoretical base for Poisson model in count data modelling. Poisson distribution should be the choice in absence of overdispersion; therefore, robust statistical test is needed for that purpose.

In this work, we discussed some test statistics based on score and evaluated them on the basis of adequacy of their distributional assumptions. We have focused on the four test statistics S_1, S_2, S_a and S_b to test mean-variance equality against inequality modelled by NB2 variance. We discussed the derivation and approximate distribution of the test statistics. We presented the ways to compute these statistics and their probabilities from their suggested approximate distribution to calculate test significance. Three of the statistics, S_1, S_a and S_b had been shown to converge to standard normal as number of observation $n \to \infty$. We simulated samples of each of the three statistic from data generated based on a Poisson regression equation that served as the null distribution. A comparison was made checking the adequacies of the suggested approximations of distributional assumptions of S_1, S_a and S_b for varying n. We tabulated the proportion of time the values of S_1, S_a and S_b each exceeded the standard normal upper 20%, 10%, 5%, and 1% points. Normality assumption S_b has been found to be almost accurate even for small values of n, S_a has shown slower convergence but normality assumption was acceptable from $n = 100$. The normality assumption for S_b has shown robustness in terms of variation of n. For values of $n \geq 100$, S_a can be preferred for its less complexity. Another feature of S_a and S_b is need of only fitting the Poisson GLM to estimateμ_i's. Also these statistics are found to be good in terms of power against negative binomial alternatives as compared to the goodness of fit Pearson and deviance statistic [7].

One of the main concern is all these tests are based on the assumption that the specification of μ_i's in terms of the covariates is correct. Even if the data is actuaaly, im may appear overdispersed relative to a Poisson model with regressors when an important regressor is absent. So when applied in practical data, we should take measures to separate the lack of fit due to misspecification of Poisson regression model and that due to not having parameter for overdispersion.

Further assessment of S_1, S_a and S_b are necessary. Since S_b possesses an assumption on μ_i's, further study should be done for checking how sensitive S_b is to variation of μ_i's specially for lower values and compared also with other statistics. These can be done by using other values of β. We have already developed a general frame work of calculating any of these statistics and p-values for the tests. Our next step involves generating data from negative binomial regression for different values of α and assessing how the test statistics perform and how they differ in giving conclusions (say p-value) to the tests. Work is underway in that direction.

Appendix A

Source Codes

A.1 R Functions for Statistics

A.1.1 R Function for calculating S_1

After we obtain $\widehat{\mu}_i = \mu_i\left(x_i,\ \widehat{\beta}\right)$, with $\widehat{\boldsymbol{\beta}}$, the maximum likelihood estimate of $\boldsymbol{\beta}$ under Poisson model, we can calculate S_1 for a given sample by implementing (3.4) with the `ScoreS1` function in following R code. The function returns the value of S_1.

```
#Function for Calculating S1
ScoreS1<-function(X,YY,mu_hat){                    #YY=observed Y vector,
                                                   #mu_hat=estimated means
    dnume<-sqrt(2*sum(mu_hat^2))                   #denumerator
    S1<- sum(((YY-mu_hat)^2-YY)/dnume)
    return(S1)
}
```

A.1.2 R Function for calculating S_2

Similarly we can implement (3.5) to calculate S_2 with the `ScoreS2` function in following R code.

```
#Function for Calculating S2
ScoreS2<-function(X,YY,mu_hat){                    #YY=observed Y vector,
                                                   #mu_hat=estimated means
    S2<-sum((YY- mu_hat)^2)/mean(YY)
    return(S2)
}
```

A.1.3 R Function for calculating S_a

The following R code gives `ScoreSa` function to calculate S_a implementing (3.9-3.11).

```
#Function for Calculating Sa
ScoreSa<-function(X,YY,mu_hat){                  #YY=observed Y vector,
                                                 #mu_hat=estimated means
    nobs<-length(YY)
    WW<-matrix(0,nrow = nobs,ncol = nobs)  #first a n by n zero atrix created
    WW<-diag(mu_hat)                             #W=diag(mu_hat1,...,mu_hatn)
    WWR<-sqrtm(WW)                               #W^(1/2) Matix
    hh<-WWR%*%X%*%solve(t(X)%*%WW%*%X)%*%t(X)%*%WWR      #lev. matrix H
    dnume<-sqrt(2*sum(mu_hat^2))                 #denumerator
    Sa<-sum(((YY-mu_hat)^2-YY+diag(hh)*mu_hat)/dnume)
    return(Sa)
}
```

A.1.4 R Function for calculating S_b

R code for `ScoreSb` function to calculate S_b is given below.

```
#Function for Calculating Sb
ScoreSb<-function(X,YY,mu_hat){                  #YY=observed Y vector,
                                                 #mu_hat=estimated means
    nobs<-length(YY)
    WW<-matrix(0,nrow = nobs,ncol = nobs)       #first a n by n zero matrix created
    WW<-diag(mu_hat)                             #W=diag(mu_hat1,...,mu_hatn)
    WWR<-sqrtm(WW)                               #W^(1/2) Matix
    hh<-WWR%*%X%*%solve(t(X)%*%WW%*%X)%*%t(X)%*%WWR      #leverage matrix H
    V<-(WWR%*%(diag(1,nobs,nobs)-hh)%*%WWR)/sum(mu_hat)  #covariance matrix V
    vv1<-sum(diag(V))                            #trace(V)
    vv2<-sum(diag(t(V)%*%V))                     #trace(V'V)
    CC<-nobs*vv2/vv1                             #c_hat
    DD<-(vv1^2)/vv2                              #d_hat
    Sb<-sqrt(4.5*DD)*((ScoreS2(X,YY,mu_hat)/(CC*DD))^(1/3)+2/(9*DD)-1)
    return(Sb)
}
```

To calculate S_b with (3.16), S_2, $\hat{c}$ and $\widehat{d}$ are required. For calculating $\hat{c}$ and $\hat{d}$ have rewritten (3.14) and (3.15) using (3.12) and (3.13) as

$$c = n\frac{tr(V^TV)}{tr(V)} \tag{A.1}$$

$$d = \frac{\{tr(V)\}^2}{tr\,(V^TV)} \tag{A.2}$$

Therefore asymptotic covariance matrix V is calculated using (3.8) to (3.10). Then V^TV matrix is calculated and traces of V and V^TV are obtained summing the diagonal

elements. Then $\hat{c}$ and $\hat{d}$ are calculated using the above equations. As S_2 is in (3.16), the `ScoreSb` function needs `ScoreS2` function in Section A.1.2 to be run before.

A.2 R Codes for Estimated UTPs of S_1, S_a, $S_b$and X^2

A.2.1 Estimateing UTPs of S_1, S_a and S_b

```
library(MASS)
library(expm)
#Function for Calcualting S1
ScoreS1<-function(X,YY,mu_hat){                #YY=observed Y vector,
                                               #mu_hat=estimated means
    dnume<-sqrt(2*sum(mu_hat^2))               #denumerator
    S1<- sum(((YY-mu_hat)^2-YY)/dnume)
    return(S1)
}
#Function for Calcualting S2
ScoreS2<-function(X,YY,mu_hat){                #YY=observed Y vector,
                                               #mu_hat=estimated means
    S2<-sum((YY- mu_hat)^2)/mean(YY)
    return(S2)
}
#Function for Calcualting Sa
ScoreSa<-function(X,YY,mu_hat){
    nobs<-length(YY)
    WW<-matrix(0,nrow = nobs,ncol = nobs) #first a n by n zero atrix created
    WW<-diag(mu_hat)                           #W=diag(mu_hat1,...,mu_hatn)
    WWR<-sqrtm(WW)                             #W^(1/2) Matix
    hh<-WWR%*%X%*%solve(t(X)%*%WW%*%X)%*%t(X)%*%WWR #leverage matrix H
    dnume<-sqrt(2*sum(mu_hat^2))               #denumerator
    Sa<-sum(((YY - mu_hat)^2 - YY+diag(hh)*mu_hat)/ dnume)
    return(Sa)
}
#Function for Calcualting Sb
ScoreSb<-function(X,YY,mu_hat){
    nobs<-length(YY)
    WW<-matrix(0,nrow = nobs,ncol = nobs)      #first a n by n zero atrix created
    WW<-diag(mu_hat)                           #W= diag(mu_hat1...,mu_hat)
    WWR<-sqrtm(WW)                             #W^(1/2) Matix
    hh<-WWR%*%X%*%solve(t(X)%*%WW%*%X)%*%t(X)%*%WWR      #leverage matrix H
    V<-(WWR%*%(diag(1,nobs,nobs)-hh)%*%WWR)/sum(mu_hat)  #covariance matrix V
    vv1<-sum(diag(V))                          #trace(V)
    vv2<-sum(diag(t(V)%*%V))                   #trace(V'V)
    CC<-nobs*vv2/vv1                           #c_hat
    DD<-(vv1^2)/vv2                            #d_hat
    Sb<-sqrt(4.5*DD)*((ScoreS2(X,YY,mu_hat)/(CC*DD))^(1/3)+2/(9*DD)-1)
    return(Sb)
}
#Function for generating random covariates
genX<-function(nobs = 20,BETA){
    p=length(BETA) - 1
    X<- cbind(1, matrix(runif(nobs * p), ncol = p))
    return(X)
}
#Function for generating samples for a particular statistic function
```

```
genS<-function(nobs = 20,X,BETA,Sfun=ScoreSb){
                                #Sfun: the statistic function,e.g.,ScoreS1, ScoreSa etc.
    p=length(BETA) - 1
    xb <- X %*% BETA
    exb <- exp(xb)
    pyf <-rpois(nobs,exb)
    outpf<- data.frame(cbind(pyf, X[,-1]))
    names(outpf) <- c("pyf", paste("x", 1:p, sep=""))
    poyf <- glm(pyf ~. ,  family=poisson, data=outpf)
    mu_hatf <-predict(poyf, type="response")
    return(Sfun(X,pyf,mu_hatf))
}
#Funtion for genreting Estimated UTPs for N(0,1) distributional assumpion
Eutp<-function(nobs = 20,X,BETA,ss=100,Sfun){
                                            ##ss=no of samples
    zm<-c(rep(NA,ss))
    UTP<-cbind(0.2,.1,.05,.01)
    tot<-c(rep(0,length(UTP)))
    pz<-qnorm(1-UTP)
    for(i in 1:ss ) {
        zm[i] <- genS(nobs,X,BETA,Sfun)
        pz[i]<-pnorm(zm[i],lower.tail = FALSE)
        for (j in 1: length(UTP)){
            if (pz[i]<UTP[j]){
                tot[j]<-tot[j]+1
            }
        }
    }
hist(zm)
e_utp=tot/ss
return(e_utp)
}

#Code for getting estimated UTPs for S1 for given model from 1000 samles
BETA= c(2.6, 3)
n_f=c(20,30,50,100,200,500)
df_S1<-data.frame()
for (i in 1:length(n_f)){
    df_S1<-rbind(df_S1,Eutp(nobs = n_f[i],genX(n_f[i],BETA),BETA,ss=1000,ScoreS1))
}
df1_S1<-cbind(n_f,df_S1)
colnames(df1_S1)<-c("n","tailP_0.20","tailP_0.10","tailP_0.05","tailP_0.01")
df1_S1


#Code for getting estimated UTPs for Sa for given model from 1000 samles
BETA= c(2.6, 3)
n_f=c(20,30,50,100,200)
df_Sa<-data.frame()
for (i in 1:length(n_f)){
    df_Sa<-rbind(df_Sa,Eutp(nobs = n_f[i],genX(n_f[i],BETA),BETA,ss=1000,ScoreSa))
}
df1_Sa<-cbind(n_f,df_Sa)
colnames(df1_Sa)<-c("n","tailP_0.20","tailP_0.10","tailP_0.05","tailP_0.01")
df1_Sa

#Code for getting estimated UTPs for Sb for given model from 1000 samples
BETA= c(2.6, 3)
n_f=c(20,30,50,100,200)
```

```
df_Sb<-data.frame()
for (i in 1:length(n_f)){
    df_Sb<-rbind(df_Sb,Eutp(nobs = n_f[i],genX(n_f[i],BETA),BETA,ss=1000,ScoreSb))
}
df1_Sb<-cbind(n_f,df_Sb)
colnames(df1_Sb)<-c("n","tailP_0.20","tailP_0.10","tailP_0.05","tailP_0.01")
df1_Sb
```

A.2.2 Estimateing UTPs of X^2

```
library(MASS)
tail_prop_P<-function(nobs = 20,M=1,UTP=0.05,ss=100){
        BETA= c(2.6,2)
        p=length(BETA) - 1
        X<- cbind(1, matrix(runif(nobs * p), ncol = p))
        xb <- X %*% BETA
        exb <- exp(xb)
        Pchif=c(rep(NA,ss))
        chi_utp=qchisq(p=1-UTP,df=nobs-length(BETA))
        ex_utp=0
        for(i in 1:ss ) {
                pyf <-rpois(nobs,M*exb)
                outpf<- data.frame(cbind(pyf, X[,-1]))
                names(outpf) <- c("pyf", paste("x", 1:p, sep=""))
                poyf <- glm(pyf ~. ,  family=poisson, data=outpf)
                mu_hatf <-predict(poyf, type="response")
                Pchif[i]<-sum((pyf - mu_hatf)^2/mu_hatf)
                if(Pchif[i]>chi_utp){
                        ex_utp<-ex_utp+1
                }
        }
        e_utp=ex_utp/ss
        return(e_utp)
}
n_f=c(20,30,50,100,200)
tailP_0.20<-c(rep(NA,length(n_f)))
tailP_0.10<-c(rep(NA,length(n_f)))
tailP_0.05<-c(rep(NA,length(n_f)))
tailP_0.01<-c(rep(NA,length(n_f)))
for (i in 1:length(n_f)){
        tailP_0.20[i]<-tail_prop_P(nobs = n_f[i],M=1,UTP=.20,ss=5000)
        tailP_0.10[i]<-tail_prop_P(nobs = n_f[i],M=1,UTP=.10,ss=5000)
        tailP_0.05[i]<-tail_prop_P(nobs = n_f[i],M=1,UTP=.05,ss=5000)
        tailP_0.01[i]<-tail_prop_P(nobs = n_f[i],M=1,UTP=.01,ss=5000)
}
tail_tabp<-data.frame(n_f,tailP_0.20,tailP_0.10,tailP_0.05,tailP_0.01)
```

Bibliography

[1] S. D. Poisson. Recherches sur la probabilité des jugements en matière criminelle et en matière civile. *Paris, Bachelier*, 1837.

[2] A. C. Cameron and P. K. Trivedi. *Regression analysis of count data*. Cambridge university press, 2nd edition, 2013.

[3] J. M. Hilbe. *Negative binomial regression*. Cambridge University Press, 2nd edition, 2011.

[4] J. Hinde and C. G. B. Demétrio. Overdispersion: models and estimation. *Computational Statistics & Data Analysis*, 27(2):151–170, 1998.

[5] B. J. Collings and B. H. Margolin. Testing goodness of fit for the poisson assumption when observations are not identically distributed. *Journal of the American Statistical Association*, 80(390):411–418, 1985.

[6] A. C. Cameron and P. K. Trivedi. Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, 1(1):29–53, 1986.

[7] C. Dean and J. F. Lawless. Tests for detecting overdispersion in poisson regressi;on models. *Journal of the American Statistical Association*, 84(406):467–472, 1989.

[8] P. McCullagh and J. A. Nelder. *Generalized linear models*. Chapman and Hall, 2nd edition, 1983.

[9] J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society. Series A*, 135(3):370–384, 1972.

[10] C. R. Rao. *Linear Statistical Inference and Its Application*. Wiley, 1973.

[11] R. W. M. Wedderburn. Quasi-likelihood functions, generalized linear models, and the gauss—newton method. *Biometrika*, 61(3):439–447, 1974.

[12] J. F. Lawless. Negative binomial and mixed poisson regression. *Canadian Journal of Statistics*, 15(3):209–225, 1987.

[13] Y. Lee and J. A. Nelder. Two ways of modelling overdispersion in non-normal data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(4): 591–598, 2000.

[14] T. S. Ferguson. *Mathematical statistics: A decision theoretic approach.*

[15] L. Lee. Specification test for poisson regression models. *International Economic Review*, pages 689–706, 1986.

[16] R. A. Fisher. The significance of deviations from expectation in a poisson series. *Biometrics*, 6(1):17–24, 1950.

[17] D. A. Williams. Generalized linear model diagnostics using the deviance and single case deletions. *Applied Statistics*, pages 181–191, 1987.

[18] E. L. Frome. The analysis of rates using poisson regression models. *Biometrics*, 39 (3):665–674.

[19] R. W. Farebrother. A remark on algorithms as 106, as 153 and as 155: The distribution of a linear combination of chi2 random variables. *Applied Statistics*, 33:366–369, 1984.

[20] N. L. Johnson and S. Kotz. *Distributions in Statistics: Continuous Univariate Distributions: Vol.: 1-2.* Houghton Mifflin, 1970.

[21] D. R. Brillinger. The Natural variability of vital rates and associated statistics. *Biometrics*, 42:693–734, 1986.

[22] D. A. Pierce and D. W. Schafer. Residuals in generalized linear models. *Journal of the American Statistical Association*, 81(396):977–986, 1986.