Power of Tests for Overdispersion Parameter in Negative Binomial Regression Model

Dejen Tesfaw Molla[1], B. Muniswamy [2]

[1]PhD Research Scholar, Department of Statistics, Andhra University, Visakhapatnam, India
[2]Associate Professor, Department of Statistics, Andhra University, Visakhapatnam, India

Abstract : *In this paper we focus on a negative binomial (NB) regression model to take account of overdispersion in Poisson counts. Moreover, we present the power of score test for testing the overdispersion parameter in the negative binomial regression model. The power of the proposed score test was compared with the LRT and Wald test via Monte Carlo simulation technique using SAS 9.2 software. The application of the test was shown using two real datasets such as using numerical illustration and real datasets.*
Keywords- *Count data, Negative binomial regression, Overdispersion, Score test*

I. INTRODUCTION

Poisson regression is one of the most popular techniques for the analysis of count data. Whereas the Poisson regression model may be the foremost candidate, it rarely explains the data due to several important constraints. One important constraint is the mean of the distribution must be equal to the variance. In this case the standard errors, usually estimated by the Maximum Likelihood method, will be biased and the test statistics derived from the models will be incorrect. Therefore, this problem leads an overdispersion. Failure to take overdispersion into account leads to serious underestimation of standard errors and misleading inference for the regression parameters. Consequently, a number of models and associated estimation methods have been proposed for handling overdispersed data. Such models include those based on the negative binomial distributions as well as regression models based on mixtures of Poisson (Lawless[1]; Dean and Lawless[2]; Wang, Puterman, Cockburn and Le[3]) as well as models to deal with overdispersion due to latent heterogeneity such as random-effects models (Ozemen[4]; Lee and Nelder[5]).

A practical and reliable test for overdispersion is important to justify the need for models beyond the standard Poisson regression model. Various tests for overdispersion have been developed, and O'Hara Hines[6] provided a good review on this topic. In the Poisson regression framework, Dean[7] developed a unifying theory and derived score tests for overdispersion with respect to both Poisson and binomial regression models, while the explicit forms of the test statistics are only given in certain special cases. The score statistics developed by Cameron and Trivedi[8] specifically for comparing the Poisson model against the negative binomial model, is a special case of the general score statistics later developed by Dean[7]. It needs to be pointed out that none of the authors cited above verified their work via the Monte Carlo simulation study and didn't compare the power of this test with the existing tests. In this paper, we provide an extensive standard Monte Carlo simulation study to show the power of score test and existing tests for a negative binomial regression models.

Researchers who choose the NB model over the Poisson model should justify the rejection of equi-dispersion in the Poisson model. A likelihood ratio test (LRT) or Wald test can be used, but the score test has the advantage that we only need to fit the Poisson model. The explicit form for a score test was given by Dean[7]. We consider these test statistics and compare the power of this test with the alternative tests in negative binomial regression model. The goal is to help researchers understand the ideas behind the power of test procedures and choose the best tests for count dataset.

The outline of the paper is as follows: In Section 1.1 and 1.2, we introduce the Poisson and negative binomial regression model and its estimation method. Score test and alternative tests for dispersion parameter in the model are discussed in Section 2 and 3, respectively. A simulation study for powers of score test and alternative tests will be presented in Section 4. Section 5 presents an example to illustrate our methodology and some conclusions are given in the last section.

1.1 Poisson regression model

Suppose we have an independent sample of n pairs of observations $\{(y_i, x_i)\}_{i \in 1,2,...,n}$, where y_i denotes the value of an event count outcome variable occurring in a given time or exposure periods with mean parameter μ_i and x_i is the value of the explanatory variables for the ith subject. Assume $y_i \sim Poisson(\mu_i)$, $i = 1, 2, ..., n$. The probability density function of Poisson random variables, Y$_i$, is characterized by

$$P(Y_i = y_i; \mu) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, \qquad y_i = 0, 1, 2, ... \tag{1}$$

with mean and variance,

$$E(y_i) = Var(y_i) = \mu_i.$$

The mean of the response variable μ_i is related with the linear predictor through the so called link function. As μ_i has to be positive, an appropriate choice is the logarithmic function, so that we do not need further restrictions on the parameters β. It is well known from the literature (McCullagh and Nelder[9]) that this is the natural link function for the Poisson distribution.

Let x be $n \times (p + 1)$ matrix of explanatory variables. The relationship between y_i and i^{th} row vect*or* of x, x_i, linked by g(μ_i) is the canonical link function given by:

$$E(y_i) = \mu_i = e^{x_i^T \beta} \tag{2}$$

where, $x_i = \left(x_{io}, x_{i1}, \ldots, x_{ip}\right)^T$ and $\beta = \left(\beta_0, \beta_1, \ldots, \beta_p\right)^T$. The model comprising (1) and (2) is known as the Poisson regression or log-linear model.

The standard estimator for the Poisson model is the maximum likelihood estimator (MLE). To find the ML of (1), we define the likelihood function as follows:

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} = \prod_{i=1}^{n} \frac{e^{-e^{x_i^T \beta}} \left(e^{x_i^T \beta}\right)^{y_i}}{y_i!} \tag{3}$$

Taking log on both sides and we get,

$$\ell(\beta) = \sum_{i=1}^{n} [y_i \ln \mu_i - \mu_i - ln y_i!] = \sum_{i=1}^{n} [y_i x_i^T \beta - e^{x_i^T \beta} - ln y_i!] \tag{4}$$

It can be verified that the first two partial derivatives of the log-likelihood function exists and are given as follows:

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} (y_i - \mu_i) x_{ij} = \sum_{i=1}^{n} (y_i - e^{x_i^T \beta}) x_{ij}$$

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_j \beta_k} = -\sum_{i=1}^{n} \mu_i x_{ij} x_{ik} = -\sum_{i=1}^{n} e^{x_i^T \beta} x_{ij} x_{ik}$$

Hence, equation (3) or (4) is nonlinear in β so that they need to be solved by using an iterative algorithm. The iterative algorithms commonly used are either Newton-Raphson or Fisher scoring. In practice $\hat{\beta}$ is the solution of the estimating equations obtained by differentiating the likelihood in terms of β and solving them to zero. Therefore, β will be obtained by maximizing (4) using a numerical iterative method (McCullagh and Nelder[9]).

1.2 Negative Binomial Models

A random variable Y is called a negative binomial distributed counts with parameter μ_i and α (> 0) if the probability density function is given by:

$$f(y_i; \mu_i, \alpha) = \frac{\Gamma(y_i + 1/\alpha)}{y_i! \, \Gamma(1/\alpha)} (1 + \alpha \mu_i)^{\frac{-1}{\alpha}} \left(1 + \frac{1}{\alpha \mu_i}\right)^{-y_i}, \qquad y_i \geq 0 \tag{5}$$

with mean $E(Y_i) = \mu_i = exp(x_i^T \beta)$, and variance $Var(Y_i) = \mu_i(1 + \alpha \mu_i)$ (McCullagh and Nelder[9]). In (5), the term α plays the role of a dispersion factor and it is a constant. Clearly, when $\alpha \to 0$, the NB distribution reduces to the usual standard Poisson distribution with parameter μ_i. For these model, we assume some specific regression model for the mean, i.e., $\log(\mu_i) = x_i^T \beta$, where x_i is $1 \times p$ row vector of covariates (including an intercept), p is the number of covariates in the model and β is the corresponding $p \times 1$ column vector of unknown regression parameters.

Jansakul, N. and Hinde, J. P.[10] illustrated that covariates can be introduced into a regression model based on the NB distribution via the relationship

$$\log \mu_i = \sum_{r=1}^{p} x_{ij} \beta_j \tag{6}$$

where x_{ij} is the ith observation in the jth covariate, p is the number of covariates in the model, and β_j is the jth regression parameter.

For the NB model, the derivatives of the log-likelihood function ℓ and its first derivatives with respect to β and $\alpha's$ are derived below:

$$\ell(\mu, \alpha; y_i) = \sum_{i=1}^{n} \left\{ -\log y_i! + \log \left(\frac{\Gamma(y_i + 1/\alpha)}{\Gamma(1/\alpha)} \right) - \frac{1}{\alpha} \log(1 + \alpha \mu_i) - y_i \log \left(1 + \frac{1}{\alpha \mu_i}\right) \right\}$$

since

$$\frac{\Gamma(y_i + 1/\alpha)}{y_i! \Gamma(1/\alpha)} = \prod_{k=1}^{y_i}(y_i + 1/\alpha - k) = \alpha^{-y_i}\prod_{k=1}^{y_i}(\alpha y_i - \alpha k + 1),$$

then

$$\ell(\mu, \alpha; y_i) = \sum_{i=1}^{n}\left\{-\log(y_i) + \sum_{k=1}^{y_i}\log(\alpha y_i - \alpha k + 1) - (y_i + 1/\alpha)\log(1 + \alpha\mu_i)\right.$$
$$\left. + y_i \log(\mu_i)\right\} \tag{7}$$

$$\frac{\partial\ell(\mu, \alpha)}{\partial\beta} = \frac{\partial\ell(\mu, \alpha)}{\partial\mu}\frac{\partial\mu}{\partial\beta} = \sum_{i=1}^{n}\frac{(y_i - \mu_i)}{1 + \alpha\mu_i}x_i;$$

$$\frac{\partial\ell(\mu, \alpha)}{\partial\alpha} = \sum_{i=1}^{n}\left\{\sum_{k=1}^{y_i}\frac{y_i - k}{\alpha y_i - \alpha k + 1} + \frac{\log(1 + \alpha\mu_i)}{\alpha^2} - \frac{(y_i + 1/\alpha)\mu_i}{1 + \alpha\mu_i}\right\}.$$

The second-order derivatives of ℓ with respect to the parameters α and β are derived as follows.

$$\frac{\partial^2\ell(\mu, \alpha)}{\partial\beta\partial\beta^T} = -\sum_{i=1}^{n}\left\{\frac{[1 + \alpha y_i]\mu_i}{(1 + \alpha\mu_i)^2}\right\}x_i x_i^{T};$$

$$\frac{\partial^2\ell(\mu, \alpha)}{\partial\alpha^2} = \sum_{i=1}^{n}\left\{\left(\sum_{k=1}^{y_i}\frac{-(y_i - k)^2}{(\alpha y_i - \alpha k + 1)^2}\right) - \frac{2\log(1 + \alpha\mu_i)}{\alpha^3} + \frac{2\mu_i}{\alpha^2(1 + \alpha\mu_i)}\right.$$
$$\left. + \frac{(y_i + 1/\alpha)\mu_i^2}{(1 + \alpha\mu_i)^2}\right\};$$

$$\frac{\partial^2\ell(\mu, \alpha)}{\partial\alpha\partial\beta} = \frac{\partial^2\ell(\mu, \alpha)}{\partial\beta\partial\alpha} = -\sum_{i=1}^{n}\left\{\frac{(y_i - \mu_i)\mu_i}{(1 + \alpha\mu_i)^2}\right\}x_i.$$

The expected information matrix $I(\beta, \alpha)$ can be partitioned as

$$I(\beta, \alpha) = \begin{bmatrix} I_{\beta\beta}(\beta, \alpha) & I_{\beta\alpha}(\beta, \alpha) \\ I_{\alpha\beta}(\beta, \alpha) & I_{\alpha\alpha}(\beta, \alpha) \end{bmatrix}, \tag{8}$$

where, the elements $I_{\beta\beta} = -E\left(\frac{\partial^2\ell(\mu,\alpha)}{\partial\beta\partial\beta^T}\right)$ is the $p \times p$ symmetric matrix, $I_{\beta\alpha} = I_{\alpha\beta}^T = -E\left(\frac{\partial^2\ell(\mu,\alpha)}{\partial\beta\partial\alpha}\right)$ is the $p \times 1$ matrix and $I_{\alpha\alpha} = -E\left(\frac{\partial^2\ell(\mu,\alpha)}{\partial\alpha^2}\right)$ is a scalar.

II.   Score test for overdispersion Parameter

Poisson model is a special case of negative binomial model. The negative binomial regression model reduces to the Poisson regression model when the overdispersion parameter $\alpha \to 0$. To assess the adequacy of the negative binomial model over the Poisson regression model, we can test the hypothesis:

$$H_0: \alpha = 0 \quad \text{versus} \quad H_A: \alpha > 0. \tag{9}$$

This is to test for the significance of the overdispersion parameter α. The presence of the overdispersion parameter α in the NB regression model is justified when the null hypothesis $H_0: \alpha = 0$ is rejected. In order to test the hypothesis in (9) a score test statistic is proposed and compared with the existing ones such as likelihood ratio test and Wald test.

The Newton-Raphson iterative algorithm used the specification of initial values. Our suggestion is setting $\alpha \to 0$ and β using ML estimation obtained from the Poisson regression model. Let $\xi = (\alpha, \beta^T)^T$, under the usual regularity conditions for maximum likelihood estimation, when the sample size is large, $\hat{\xi} \sim N_p(\xi, I^{-1}(\alpha, \beta))$ approximately.

Let $\hat{\xi}_1 = (0, \hat{\beta}^T)^T$ be the restricted maximum likelihood estimates (REML) under the null hypothesis, $\alpha = 0$ or under H_0 true. Then the general score test statistic for testing H_0 can be computed as

$$S_\alpha = S_\alpha^T(\beta, \alpha)I_{\alpha\alpha}^{-1}S_\alpha(\beta, \alpha)\big|_{\hat{\xi}_1} \tag{10}$$

where, $\hat{\beta}$ is the maximum likelihood estimates under the Poisson model and the score vector is given:

$$S_\alpha(\beta, \alpha) = \frac{\partial\ell(\mu, \alpha)}{\partial\alpha}\bigg|_{\alpha=0} = \frac{1}{2}\sum_{i=1}^{n}\{(y_i - \hat{\mu}_i)^2 - y_i\}. \tag{11}$$

From equation (8) the inverse of the Fisher information matrix for overdispersion parameter is given by:

$$I_{\alpha\alpha}{}^{-1} = I_{\alpha\alpha}(\beta, \alpha) - I_{(\beta,\alpha)}^T(\beta, \alpha) I_{(\beta,\beta)}^{-1}(\beta, \alpha) I_{(\beta,\alpha)}(\beta, \alpha)\Big|_{\hat{\xi}_1}$$

$$= \frac{1}{2}\sum_{i=1}^{n}\hat{\mu}_i^2.$$

Therefore, the score statistic is given by

$$S_\alpha = \frac{[\sum_{i=1}^{n}\{(y_i - \hat{\mu}_i)^2 - y_i\}]^2}{2\sum_{i=1}^{n}\hat{\mu}_i^2} \tag{12}$$

where, $\hat{\mu}_i$ is $n \times 1$ vector of estimated value from the Poisson regression model. Under the hypothesis of Poisson model, the limiting distribution of the score statistic is chi-squared with one degree of freedom.

When there is no covariate in the NB model, then the score statistic can be simplified as

$$S_{\alpha 1} = \frac{1}{2n}\left(\frac{\sum_{i=1}^{n}\{(y_i - \hat{\mu}_i)^2 - y_i\}}{\hat{\mu}_i}\right)^2 \tag{13}$$

It is also an asymptotic chi-square distribution with one degree of freedom, χ_1^2.

III.   Alternative tests

Poisson model is a special case of negative binomial model. Within the family of negative binomial models, testing if a Poisson model is adequate corresponding to testing: $H_0: \alpha = 0$ $vs.$ $H_A: \alpha > 0$, where one possible test statistic is the likelihood ratio test (LRT). For a general negative binomial regression model, the LRT for α is given by $\text{LRT}_\alpha = -2\big(\ell(\hat{\mu}) - \ell(\hat{\mu}, \hat{\alpha})\big)$, where $\ell(\hat{\mu})$ and $\ell(\hat{\mu}, \hat{\alpha})$ are, respectively, the maximized log-likelihood under the Poisson regression and NB regression models. Standard asymptotic theory suggests that under H_{01} LRT_α has probability mass of one half at zero and one half – Chi-square distribution with 1 degree of freedom (Cameron and Trivedi[11]). We can also use the asymptotic normal Wald type "t" statistics defined as the ratio of the estimate of α to its standard error.

IV.   Monte Carlo Simulation

A simulation study was conducted to examine the empirical size and power of the proposed score test against the existing test statistics. In order to compare the power of score test with the existing test statistics for testing the overdispersion in a negative binomial model a limited simulation study was carried out under different situations. The model used for simulation study is

$$Y_i \sim NB(\alpha, \mu_i), \quad i = 1, 2, \dots, n; \quad \text{where, } \log(\hat{\mu}_i) = 2.5 - x_i.$$

We first generate a set of random numbers from a uniform distribution in the interval [0, 1] as the values of x_i. To get values of y_i, a random variate is drawn from a negative binomial model with true values of parameters, the value of x_i, and a given α value. Repeating these procedure n times, we get a set of simulated data $\{y_i, x_i, i = 1, 2, \dots n\}$. The values of score statistics are computed by formulas shown in section (2). We take $\alpha = 0.01, 0.025, 0.05, 0.075, 0.1, 0.25$ and 0.5. For each given values of parameters, we do 1000 replications (the values of x_i's are fixed for each replication). Then the proportion of times which rejected the null hypothesis is just the simulated value of power. Here, all the statistics are compared with the χ_α^2 critical value at $\alpha = 0.05$ level. The simulations are performed for samples of size $n = 50, 100, 200, 300$ and 400 to get the simulated powers of test statistics.

For each set of generated data, a negative binomial model is fitted for calculating the score test and the existing tests followed by the powers of the tests. Results from the simulation study are presented in Table- 1.

Table- 1: Empirical power of the Wald, LRT and Score tests for overdispersion parameter based on 1,000 replications generated from the negative binomial regression model.

n	Method	Power						
		$\alpha = 0.01$	0.025	0.05	0.075	0.1	0.25	0.5
50	Wald	0.046	0.054	0.117	0.298	0.492	0.980	1
	LRT	0.061	0.108	0.333	0.556	0.752	0.997	1
	Score	0.088	0.215	0.487	0.704	0.850	0.999	1
100	Wald	0.056	0.097	0.414	0.738	0.912	1	1
	LRT	0.074	0.212	0.602	0.852	0.964	1	1
	Score	0.105	0.305	0.834	0.910	0.982	1	1
200	Wald	0.059	0.275	0.801	0.984	0.998	1	1
	LRT	0.098	0.393	0.871	0.991	1	1	1
	Score	0.144	0.472	0.911	0.995	1	1	1
300	Wald	0.086	0.460	0.951	0.998	1	1	1
	LRT	0.132	0.564	0.970	0.999	1	1	1
	Score	0.172	0.640	0.980	0 .999	1	1	1
400	Wald	0.126	0.622	0.990	1	1	1	1
	LRT	0.177	0.702	0.995	1	1	1	1
	Score	0.219	0.759	0.996	1	1	1	1

The results in Table-1 indicate that the score test holds the nominal level reasonably well. Generally, the power for the three tests increases when α increases, and for large sample, $n = 400$ and $\alpha = 0.075$, the power increases very fast and approaches 1. The behavior of Score, LRT and Wald test statistics indicate different pattern: for small sample $n = 50$, the power increases slowly firstly, and if $\alpha > 0.1$, then the power increase fast, and the power get to nearly 1.0 if $\alpha = 0.5$; while for $n = 400$, the power increases slowly if $\alpha < 0.01$, then it will increase fast and approaches nearly 1.0 quickly at $\alpha = 0.075$. When α is close to 0, there is only small difference between negative binomial and Poisson model, and with α increasing, the difference also increases. Some representative results are shown in Fig. 1 which displays that the plots of power functions of tests with respect to varying values of α. As shown in Fig. 1 as n increase, the powers of tests also increase. Therefore, the proposed score test is more important for general use than the alternative tests because of its high power.
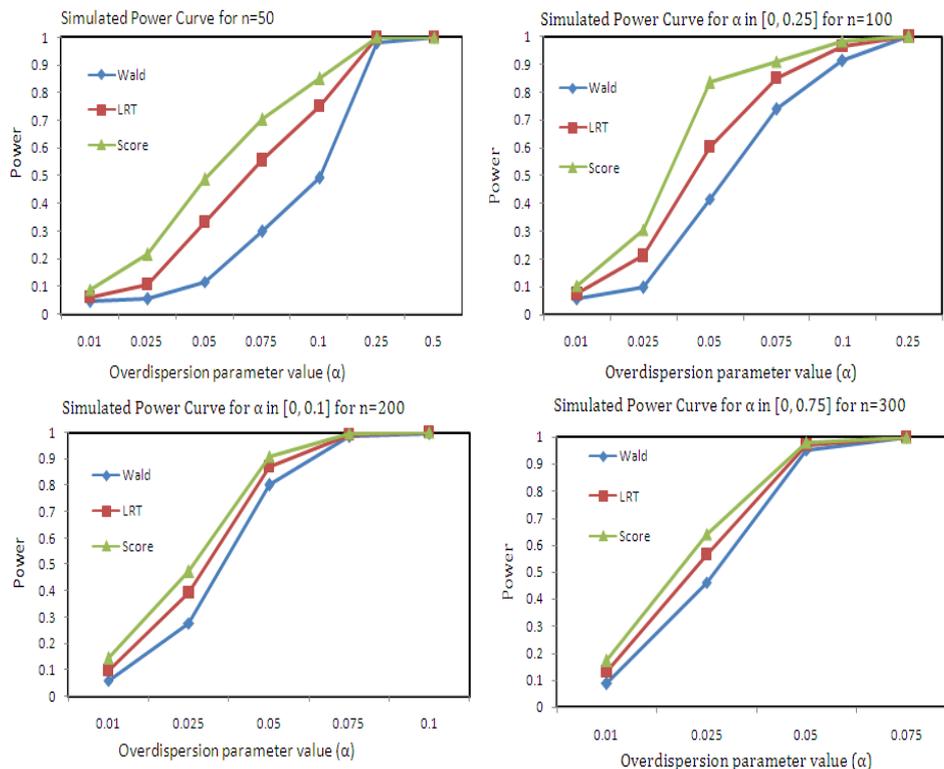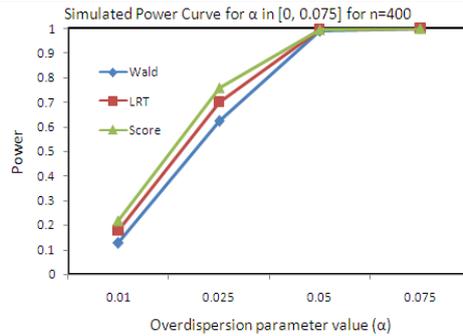
Figure- 1: The Simulated power curve for the Wald, LRT and Score tests.

V.   Examples

1.   Fitting a NB model without covariates

Gupta et al.[12] use the numbers of death notices of women 80 years of age and over, appearing in the London "Times" on each day for three consecutive years to show the variance of the estimates. Here, we also use it to show the proposed score statistics to test overdispersion in Poisson model. The dataset is shown in the Table 2. The Predicted percent for each count (from fitted models) for Poisson and NB model are also presented in Table- 2.

For negative binomial model, the estimated dispersion parameter for $\hat{\alpha} = 0.1009$ with standard error 0.0255 and based on the modeling information, the computed score statistics as $S = 9166.891$ with the corresponding p- value is $p = 0.0000$; from Table- 3, the LRT statistics can be calculated as 2001.398-1990.829 = 10.569 with the p- value as $p = 000575$; and the Wald statistics as 3.96 with its corresponding p-value is $p < 0.0233$. All of the test statistics indicate very strong evidence against the fit of the Poisson model to the data.

Table- 2: Death Notice Data of London Times and predicted percent

Number of Death notices	Observed frequency	Observed Percent (%)	Predicted percent (%)	
			Poisson model	NB model
0	162	14.78	46.3620	47.6953
1	267	24.36	35.6380	30.8972
2	271	24.73	13.6973	11.1110
3	185	16.88	3.5010	2.9347
4	111	10.13	0.6745	0.6337
5	61	5.57	0.1037	0.1186
6	27	2.46	0.0133	0.0199
7	8	0.73	0.0015	0.0030
8	3	0.27	0.0001	0.0004
9	1	0.09	0.0000	0.0000

Table- 3: Model fitting of NB and Poisson models for number of death notices count dataset.

Model	Selection criteria		
	-2ℓ	AIC	BIC
Poisson	2001.398	4004.796	4009.795
Negative binomial	1990.829	3985.657	3995.656

Clearly, from Table- 3, we note that the AIC and BIC values for both Poisson and negative binomial, the negative binomial is better than the Poisson, this can also be seen from the Fig. 2, since the predicted percent for Poisson model is closer to the observed percent for each count.
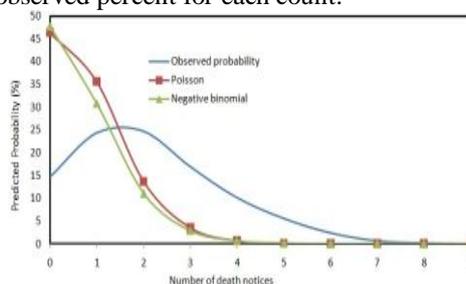
Fig. 2: Compare the Poisson and NB models for Death notice data of London Times

2. Fitting a NB model with covariates

To illustrate our methodology for fitting a NB model, we first consider the 2005 Demographic and Health Survey (DHS) which is obtained from the Central Statistical Agency, Ethiopia. Thus, this study analyzes responses from each of 9210 women (only those who have ever born a child), out of 14070 women of age 15-49 interviewed in 2005 DHS, on the counts of the number of deaths of children aged less than 5 years that the mother has experienced in her lifetime. The response variable of this study, Y_i, is a count, which gives the number of deaths of children aged less than 5 years that each mother has experienced in her lifetime.

The set of children death dataset in Ethiopia is then used to illustrate the score test with covariates in the negative binomial model. The number of child death begins with a value of zero and grows from there. There are 9210 observations in the dataset, and the minimum count is 0 and the maximum count is 7, with mean 0.63 and median 0. The dispersion index (the ratio of variance to mean) is 1.7905. So the data exhibit overdispersion. There are a number of variables in the dataset. Here we select six important dummy explanatory variables, i.e., age of women at first birth (x_1), toilet facility (x_2), source of drinking water (x_3), education status of women (x_4), work status of women (x_5) and type of place of residence (x_6), from the variables and using the negative binomial regression model to fit the data for illustrating our results.

The predicted probability for Poisson and negative binomial model are presented in Table- 4 and Figure- 3. In negative binomial model the estimated dispersion parameter for $\hat{\alpha} = 0.298942$ with standard error 0.028569; and the estimated regression parameters are $\hat{\beta} = $ (-1.9982, -0.1794, -0.0190, 0.01093, -0.4256, -0.1318, 0.2848). Based on the modeling information, the computed score statistics for overdispersion is $S = 183169.60$ with p-value $= 0.0000$; from Table-3.6, the LRT statistics is 275.48-228.91=46.57 with p-value $= 0.0000$; and the Wald statistics as 10.46 and its corresponding p-value is $p < 0.00061$. All of the test statistics indicate very strong evidence against the fit of the Poisson model to the data.

The above analysis doesn't indicate which of the Poisson and negative binomial model will fit the data better. To cheek this, we fitted the maximum likelihood estimate of the parameters and the maximized log-likelihoods for them. The fitted statistics for Poisson and negative binomial models are shown in Table- 5.

Table- 4: The observed and predicted count percent of children who died before age five per   mother by Poisson and NB models

Number of U5CD	Observed frequency	Observed Percent (%)	Predicted percent (%)	
			Poisson model	NB model
0	5897	64.0282	36.7879	41.6896
1	1881	20.4235	36.7879	28.7837
2	793	8.6102	18.3940	13.2487
3	379	4.1151	6.1313	5.0818
4	166	1.8024	1.5328	1.7543
5	60	0.6515	0.3066	0.5652
6	20	0.2172	0.0510	0.1735
7	14	0.1520	0.0073	0.0513

Clearly, from Fig. 2 and the value of AIC, BIC criteria in Table- 5, there is a difference between Poisson and negative binomial model for this dataset. Then we can make a conclusion that the negative binomial model is essentially more appropriate than the Poisson model for the number of less than 5 years children death in Ethiopia dataset.

Table- 5: Model fitting of NB and Poisson models for number of under 5 years children death count dataset.

Model	Selection criteria		
	-2ℓ	AIC	BIC
Poisson	275.48	17203.81	17253.71
Negative binomial	228.91	17004.82	17061.85
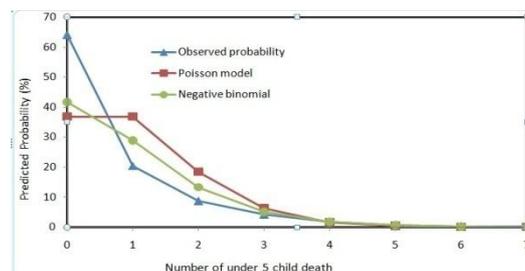
Figure- 3: Graphical comparison of Poisson and NB distribution

VI.   Conclusion

Overdispersion is a common phenomenon in Poisson modeling, and the negative binomial model is also a candidate for overdispersed count data. The score test statistic has an advantage over the LRT and Wald test in that it only requires the parameter estimated under the null hypothesis. This is an important advantage to methodologists utilizing software without negative binomial modeling capabilities. The simulation study indicates the score test is preferable to the LRT and Wald test because of its high power.

Although Poisson and negative binomial may be appropriate for different datasets. In this paper we have seen with two examples, and the AIC, and BIC, test can be used to justify the choice between Poisson and negative binomial, although they can provide some indications.

Our main work has been focused on the negative binomial regression models with testing for overdispersion parameters. However, it seems that it is reasonable to test the coefficient of regression parameters for negative binomial regression model. We will consider it in our future research.

References:

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

[2].   C. Dean and J. F. Lawless, Tests for detecting overdispersion in Poisson regression Models, *Journal of the American Statistical Association. 84, 1989, 467-472.*

[3].   P. Wang, M. L. Puterman, I. Cockburn, and N. Le, Mixed Poisson regression models with covariate dependent rates. *Journal of Biometrics, 52, 1996, 381-400.*

[4].   I. Ozmen, Quasilikelihood/ moment method for generalized and restricted generalized Poisson regression models and its application, Journal of Biometrical, 42, 2000, 303–314.

[5].   Y. Lee, J. A. Nelder, Two ways of modeling overdispersion in non-normal data, Journal of Applied Statistics, 49, 2000, 591–598.

[6].   R. J. O'Hara Hines, A Comparison of Score Tests for Overdispersion in Generalized Linear Models. Journal of Statistical Computation and Simulation. 58, 1997, 323-342.

[7].   C. B. Dean, Testing for overdispersion in Poisson and binomial regression models. Journal of American Statistical Association. 87, 1992, 451–457.

[8].   A. C. Cameron, and P. K. Trivedi, Econometric Models Based on Count Data: Comparisons and Application of Some Estimators and Tests, Journal of Applied. Econometrics, 1, 1986, 29-53.

[9].   P. McCullagh, and J. A. Nelder, *Generalized linear models* (New York, USA, 2nd Ed., Chapman and Hall, 1989).

[10].   N. Jansakul, and J. P. Hinde, Linear mean-variance negative binomial models for analysis of orange tissue-culture data. Journal of Science and Technology, 26(5), 2004, 683-696.

[11].   A. C. Cameron, P.K.Trivedi, Regression Analysis of Count Data, Cambridge University Press, 1998.

[12].   P. L. Gupta, R. C. Gupta, and R. C. Tripathi, 1996, Analysis of zero-adjusted count data, Computational Statistics and Data Analysis, 23, 1996, 207-218.