Power of Overdispersion Tests in Zero-Truncated Negative Binomial Regression Model

Enas Gawdat Yehia

Department of Statistics, Faculty of Commerce, Al-Azhar University, Tafahna Al-Ashraf, Egypt

Email address:
enasyehia538.el@azhar.edu.eg

Abstract: Poisson regression is the most extensively used model for modeling data that are measured as counts. The main characteristic of Poisson regression model is the equidispersion limitation in which the mean and variance of the count variable are the same. However, in many situations the variance of the count variable is greater than the mean which causes overdispersion, and hence, poor fit will be resulted when inference about regression parameters. Alternatively, the negative binomial regression is preferred when overdispersion is present. In addition, in particular cases, the zero counts are not observed in data which is known as zero-truncation. In the presence of overdispersion in zero-truncated count data, the zero-truncated negative binomial (ZTNB) regression model can be used as an alternative to zero-truncated Poisson (ZTP) regression model. In this paper, for testing overdispersion in ZTNB regression model against ZTP regression model, the likelihood ratio test (LRT), score test, and Wald test are proposed. A Monte-Carlo simulation is carried out in order to examine the empirical power for statistics of these tests under different levels of overdispersion and various sample sizes. The simulation results indicate that Wald test is more powerful than the LRT and score test for detecting the overdispersion parameter in ZTNB regression model against ZTP regression model, since it provides the highest statistical power. Thus, the Wald test is preferable for detecting the overdispersion problem in zero-truncated count data.

Keywords: Count Regression, Zero-Truncated Poisson, Overdispersion, Likelihood Ratio Test, Score Test, Wald Test

1. Introduction

The counting data can be defined as the number of occurrences of an event within a fixed period of time, where which this data can take only non-negative discrete numbers. Poisson regression is the most common modeling technique for count data in a wide variety of fields such as biostatistics, agriculture, econometric, epidemiology, psychology, and many others. The standard Poisson regression models have the equidispersion limitation which the mean and variance of counts are equal. However, many count data do not satisfy the equidispersed property, they are either overdispersed (variance is greater than mean) or underdispersed (variance is less than mean) Cameron and Trivedi [4]. The negative binomial regression is an appropriate approach to model overdispersed count data as an alternative to Poisson regression model.

In practice, many studies have been established to analyze overdispersed count data (e.g., Lee et al [15], Zhang et al [28], Hossain et al [11], Ismah et al [12]).

In addition, in many cases, the count data are recorded only over part of the response variable's range, then the data are said to be truncated such as the length of hospital stay, the age of an animal in years, and the number of accidents per worker in a factory. In particular, the zero counts are not observed which is known as zero-truncation, or more generally left-truncation or truncation from below. Also, right- truncation or truncation from above may be arise.

Zero-truncated Poisson (ZTP) distribution is the most widely used to model count data with zero-truncation Gurmu [8], Johnson et al [13]. It is also usually called the positive Poisson distribution, so for positive counts, the fit of truncated Poisson distribution has been preferred Singh [22], Matthews and Appleton [16]. Poisson regression model has been proposed to analyze the truncated count data by Shaw [21]. However, in the presence of overdispersion, poor fit will be resulted and the estimates of regression parameters will be biased and inconsistent Grogger and Carson [7],

Cameron and Trivedi [4]. The most common way to handle overdispersion and truncated data at zero is to employ zero-truncated negative binomial (ZTNB) distribution which is presented by Sampford [20]. Further, the ZTNB regression model is suitable approach for modeling zero-truncated count data in the presence of overdispersion problem.

For testing overdispersion in count data, several tests have been proposed, one can refer to Lee [14], Dean and Lawless [6], Cameron and Trivedi [3], O'Hara Hines [18], Hilbe [10] for more details. According to three categories of nested models, various tests of overdispersion have been proposed in the Poisson model versus more general parametric model (e.g., Gurmu and Trivedi [9], Yang et al [25], Yang et al [26], Zhao et al [29], Molla and Muniswamy [17], Zamani and Ismail [27], Pongsapukdee et al [19]). These categories are based on the estimation of the unrestricted model as in the Wald test Winkelmann [24], the estimation of the restricted model as in the score test Cameron and Trivedi [2], Dean [5], Gurmu and Trivedi [9], and the difference between the restricted and the unrestricted log likelihood values as in the likelihood ratio test (LRT) Vives et al [23], Winkelmann [24].

The aim of this paper is to detect the overdispersion problem in zero-truncated count data based on ZTNB regression model against ZTP regression model. Therefore, the likelihood ratio test (LRT), score test, and Wald test are proposed for testing the overdispersion parameter in ZTNB against ZTP regression models. In addition, the empirical power for statistics of these tests is examined under different cases of overdispersion and sample sizes by simulation study to choose the most powerful test in detecting overdispersion.

The rest of this paper is structured as follows: A brief review of ZTP and ZTNB regression models is provided in section 2. Maximum likelihood estimates for parameters of ZTP and ZTNB are derived in section 3. Testing for overdispersion is discussed in section 4. A simulation study is carried out in section 5 to investigate the empirical power for statistics of LRT, score, and Wald tests under different levels of overdispersion and various sample sizes in ZTNB against ZTP regression models. Finally, some conclusions are summarized in section 6.

2. Zero-Truncated Poisson and Negative Binomial Regression Models

Consider $Y_i, i = 1, 2, \dots, n$ be a count variable which follows by a discrete probability function $Pr(Y_i = y_i)$. Let $k + 1$ are omitted, then the resulting distribution is called left-truncated and its probability function will be denoted by

$$f_k(Y_i) = Pr(Y_i = y_i | Y_i > k) = \frac{Pr(Y_i = y_i)}{Pr(Y_i > k)} = \frac{f(y_i)}{1 - F(k)}, \quad (1)$$

where $f_k(Y_i)$ is the truncated (above k) probability function, $f(y_i) = Pr(Y_i = y_i)$ is the probability function of the random variable Y_i, and $F(k)$ is the distribution function evaluated at k.

The left-truncation or zero-truncation is the most common way of truncation in count models, in which $k = 0$. Since many count data have been analyzed by several generalized distributions of Poisson, Grogger and Carson [7] proposed a Poisson distribution to model left-truncated count data at the value $k = 0$ as follows:

$$Pr(Y_i = y_i | Y_i > 0) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i! \, (1 - F_P(0))}, = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i! \, (1 - e^{-\lambda_i})}, = \frac{\lambda_i^{y_i}}{y_i! \, (e^{\lambda_i} - 1)}, \quad (2)$$

where the value of $Y_i = 0$ is omitted, $\lambda_i = exp(x_i' \beta), i = 1, 2, \dots, n$ is the mean of Poisson distribution, x_i is a $(1 \times k)$ vector of covariates, β is a $(k \times 1)$ vector of parameters, and $F_P(0) = (1 - e^{-\lambda_i})$ is the distribution function at 0.

The conditional mean and variance of Y_i are given respectively by

$$E(Y_i | X_i, Y_i > 0) = \lambda_i [1 - F_P(0)]^{-1}, = \lambda_i [1 - e^{-\lambda_i}]^{-1}. \quad (3)$$

$$Var(Y_i | X_i, Y_i > 0) = E(Y_i | X_i, Y_i > 0) \times [1 - F_P(0)E(Y_i | X_i, Y_i > 0)], = \frac{\lambda_i}{1 - e^{-\lambda_i}}\left[1 - \frac{\lambda_i e^{-\lambda_i}}{1 - e^{-\lambda_i}}\right]. \quad (4)$$

In many applications, since data are often overdispersed, the estimates of regression parameters of truncated Poisson model will be biased and inconsistent Cameron and Trivedi [4]. One way to handle overdispersion is to consider a mixture model with overdispersed distribution, for example, the negative binomial model is a gamma-Poisson mixture model that can be preferred when Poisson mean has a gamma distribution Winkelmann [24], Arrabal et al [1].

In the situation of truncated count data, the left-truncated negative binomial (LTNB) distribution at zero is given by

$$Pr(Y_i = y_i | Y_i > 0) = \frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})} (\alpha \, \lambda_i)^{y_i} \times (1 + \alpha \, \lambda_i)^{-(y_i + \frac{1}{\alpha})} [1 - F_{NB}(0)]^{-1}, \quad (5)$$

$$= \frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})\left[1 - (1 + \alpha \, \lambda_i)^{-\frac{1}{\alpha}}\right]} (\alpha\lambda_i)^{y_i} \times (1 + \alpha\lambda_i)^{-(y_i + \frac{1}{\alpha})}, \quad (6)$$

where $F_{NB}(0) = (1 + \alpha \, \lambda_i)^{-\frac{1}{\alpha}}, \alpha > 0$ is the distribution function at 0.

Then, the conditional mean and variance of Y_i are given respectively as follows:

$$E(Y_i \mid X_i, Y_i > 0) = \lambda_i [1 - F_{NB}(0)]^{-1}, = \lambda_i [1 - (1 + \alpha\lambda_i)^{-\frac{1}{\alpha}}]^{-1}. \tag{7}$$

$Var(Y_i \mid X_i, Y_i > 0) =$

$$\frac{E(Y_i \mid X_i, Y_i > 0) \, [1 - [F_{NB}(0)]^{1+\alpha} \, E(Y_i \mid X_i, Y_i > 0)]}{[F_{NB}(0)]^{\alpha}}, = \frac{\lambda_i}{1 - (1 + \alpha\lambda_i)^{-\frac{1}{\alpha}}} \left[1 + \alpha\lambda_i - \frac{\lambda_i (1 + \alpha\lambda_i)^{-\frac{1}{\alpha}}}{1 - (1 + \alpha\lambda_i)^{-\frac{1}{\alpha}}} \right]^{-1}. \tag{8}$$

3. Maximum Likelihood Estimation of ZTP and ZTNB Regression Models

For the model of ZTP regression as in (2), the maximum likelihood (ML) estimation method can be used to estimate its parameter by taking the partial derivative of the likelihood function with respect to the parameter and setting it equals to zero. The likelihood function is

$$L = \prod_{i=1}^{m} \frac{\lambda_i{}^{y_i}}{y_i! \, (e^{\lambda_i} - 1)}, \tag{9}$$

where m is the number of observations in the truncated sample. Then, the log likelihood function is given by

$$Log \, L = \ell = \sum_{i=1}^{m} [y_i x_i'\beta - Log(y_i!) - Log(e^{\lambda_i} - 1)]. \tag{10}$$

The first partial derivative with respect to β is given as follows:

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{m} x_i \left[y_i - \frac{\lambda_i}{1 - e^{-\lambda_i}} \right]. \tag{11}$$

Therefore, the first order condition for ML is given by maximizing ℓ for β as:

$$\sum_{i=1}^{m} x_i \left[y_i - \frac{\lambda_i}{1 - e^{-\lambda_i}} \right] = 0, \tag{12}$$

An iterative procedure such as the Newton-Raphson or Fisher scoring can be implemented to solve (12) numerically.

The second partial derivatives with respect to β can be obtained, forming the elements of Fisher's information matrix $I(\beta)$, where $I(\beta) = -E\left[\frac{\partial^2 \ell}{\partial \beta \partial \beta'} \right]$,

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta'} = -\sum_{i=1}^{m} \lambda_i x_i x_i' \left[\frac{1 - (1 + \lambda_i) e^{-\lambda_i}}{(1 - e^{-\lambda_i})^2} \right]. \tag{13}$$

Then, the maximum likelihood estimator of β is normally asymptotic distribution with mean β and variance matrix $[I(\beta)]^{-1}$.

In addition, for ZTNB regression model, the maximum likelihood estimation method can be used to obtain the parameters estimates as follows:
From (6), the likelihood function is given by the following form:

$$L = \prod_{i=1}^{m} \left[\frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha}) \, [1 - (1 + \alpha\lambda_i)^{-\frac{1}{\alpha}}]} \times (\alpha\lambda_i)^{y_i} (1 + \alpha\lambda_i)^{-(y_i + \frac{1}{\alpha})} \right]. \tag{14}$$

The log likelihood function is given by

$$Log \, L = \ell = \sum_{i=1}^{m} \left[\sum_{r=0}^{y_i - 1} Log(1 + \alpha r) - Log(y_i!) + y_i x_i'\beta - \left(y_i + \frac{1}{\alpha} \right) Log(1 + \alpha\lambda_i) - Log(1 - (1 + \alpha\lambda_i)^{-\frac{1}{\alpha}}) \right]. \tag{15}$$

Then, the maximum likelihood estimators for the parameters in zero-truncated negative binomial can be obtained by setting the first partial derivatives of log likelihood function, ℓ in (15) with respect to β and α equal to zero as follows:

$$\frac{\partial \ell}{\partial \beta} = \frac{\partial \ell}{\partial \lambda} \frac{\partial \lambda}{\partial \beta} = 0, \sum_{i=1}^{m} x_i \left[\frac{y_i - \lambda_i}{1 + \alpha\lambda_i} - \frac{\lambda_i (1 + \alpha\lambda_i)^{-(1 + \frac{1}{\alpha})}}{1 - (1 + \alpha\lambda_i)^{-\frac{1}{\alpha}}} \right] = 0, \tag{16}$$

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{m} \left[\sum_{r=0}^{y_i - 1} \frac{r}{1 + \alpha r} - \frac{\lambda_i (y_i + \frac{1}{\alpha})}{1 + \alpha\lambda_i} - \frac{\lambda_i (1 + \alpha\lambda_i)^{-\frac{1}{\alpha} - 1}}{\alpha \, [1 - (1 + \alpha\lambda_i)^{-\frac{1}{\alpha}}]} \right] = 0. \tag{17}$$

One can use the Newton-Raphson procedure or Fisher scoring to obtain the solution of (16), and (17) numerically.
The second order derivatives of ℓ with respect to the parameters, β and α are

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta'} = -\sum_{i=1}^{m} \lambda_i x_i x_i' \left[\frac{1 + \alpha y_i}{c^2} - A \right], \tag{18}$$

where

$$A = \frac{\lambda_i c^{-2\left(1 + \frac{1}{\alpha}\right)} + v \left[c^{-\left(1 + \frac{1}{\alpha}\right)} - \lambda_i (1 + \alpha) c^{-2 + \frac{1}{\alpha}} \right]}{v^2},$$

$c = 1 + \alpha\lambda_i$, and $v = 1 - c^{-\frac{1}{\alpha}}$.

$$\frac{\partial^2 \ell}{\partial \alpha^2} = \sum_{i=1}^{m}[\sum_{r=0}^{y_i-1} \frac{-r^2}{(1+\alpha r)^2} + \frac{\lambda_i}{c\alpha^2} + \frac{\lambda_i D}{\alpha^3 c^{2+\frac{1}{\alpha}}(c^{\frac{1}{\alpha}}-1)^2}], \tag{19}$$

where $D = Log[c]^{-c^{\frac{2}{\alpha}}} - Log[\alpha\lambda_i c]^{c^{\frac{2}{\alpha}}} + \alpha c^{\frac{2}{\alpha}}(1+\lambda_i) - \alpha c^{\frac{1}{\alpha}}$.

$$\frac{\partial^2 \ell}{\partial\beta\partial\alpha} = \sum_{i=1}^{m}[\alpha^2\lambda_i c^{\frac{2}{\alpha}}\left(c^{\frac{1}{\alpha}}-1\right) - \alpha^2 y_i c^{\frac{1}{\alpha}}\left(1+c^{\frac{2}{\alpha}}\right) + \alpha c^{\frac{2}{\alpha}}(2\alpha y_i + \lambda_i) - Log[\alpha\lambda_i c]^{c^{\frac{2}{\alpha}}} - Log[c]^{c^{\frac{2}{\alpha}}}\left[\frac{\lambda_i x_i}{\alpha^2 c^{2+\frac{1}{\alpha}}(c^{\frac{1}{\alpha}}-1)^2}\right]. \tag{20}$$

Let $\xi = (\alpha, \beta)'$ *be a* $p+1$ vector of parameters with β having p elements, then, the observed Fisher information matrix can be partitioned as follows:

$$I(\xi) = \begin{bmatrix} I_{\beta\beta}(\xi) & I_{\beta\alpha}(\xi) \\ I_{\alpha\beta}(\xi) & I_{\alpha\alpha}(\xi) \end{bmatrix}, \tag{21}$$

where $I_{\beta\beta}(\xi) = -E[\frac{\partial^2 \ell(\xi)}{\partial\beta\,\partial\beta'}]$ is the $p \times p$ symmetric matrix, $I_{\beta\alpha}(\xi) = I'_{\alpha\beta}(\xi) = -E[\frac{\partial^2 \ell(\xi)}{\partial\beta\,\partial\alpha}]$ is the $p \times 1$ matrix and $I_{\alpha\alpha}(\xi) = -E[\frac{\partial^2 \ell(\xi)}{\partial\alpha^2}]$ is a scalar. Then, when the sample size is large and the usual regularity conditions of maximum likelihood estimation are satisfied, $\hat{\xi}$ is approximately distributed as $N_p(\xi, I^{-1}(\xi))$.

4. Testing for Overdispersion

For detecting the overdispersion parameter in ZTNB regression model versus ZTP regression model, the following hypothesis can be tested:

$$H_0: \alpha = 0 \text{ against } H_1: \alpha > 0, \tag{22}$$

where the null hypothesis is rejected when there is evidence that the overdispersion parameter is significant.

The likelihood ratio test (LRT) is one of possible tests that can be used to carry out the hypothesis in (22) based on the ratio of two log likelihood functions evaluated at the restricted and unrestricted maximum likelihood estimates. The statistic of LRT denoted T_{LR} and is given by

$$T_{LR} = -2[\ell(\hat{\lambda}) - \ell(\hat{\lambda}, \hat{\alpha})], \tag{23}$$

where $\ell(\hat{\lambda})$ and $\ell(\hat{\lambda}, \hat{\alpha})$ are the log likelihood under ZTP and ZTNB respectively, and $\hat{\lambda}$, $\hat{\alpha}$ are the maximum likelihood estimates of λ and α respectively. Under H_0, the T_{LR} statistic has an asymptotical Chi-square distribution with one degree of freedom. If T_{LR} statistic is significant, then the unrestricted model is said to fit the data significantly better than the restricted model.

A score test is an alternative test that can be used for testing the significance of overdispersion parameter due to (22). Let $\xi^* = (0, \beta')'$ be the restricted maximum likelihood estimates of ξ under H_0 true. Then, the score test statistic denoted T_S and is given by

$$T_S = S'(\xi)I_{\alpha\alpha}^{-1}S(\xi)|_{\xi^*}, \tag{24}$$

where $S(\xi) = \frac{\partial\ell(\xi)}{\partial\alpha}$ is the score vector and $I_{\alpha\alpha}^{-1} = I_{\alpha\alpha}(\hat{\xi}) - I_{\alpha\beta}(\hat{\xi})I_{\beta\beta}^{-1}(\hat{\xi})I_{\alpha\beta}(\hat{\xi})$. Then,

$$T_S = \frac{(\sum_{i=1}^{n}[(y_i-\hat{\lambda}_i)^2 - y_i])^2}{2\sum_{i=1}^{n}\hat{\lambda}_i^2}, \tag{25}$$

where $\hat{\lambda}_i$ is the estimated value from the ZTP model. Under H_0, the score statistic has an asymptotical Chi-square distribution with one degree of freedom Cameron and Trivedi [2].

In addition, the statistic T_S in (25) can be equivalently written as follows:

$$T_S = \frac{\sum_{i=1}^{n}[(y_i-\hat{\lambda}_i)^2 - y_i]}{\sqrt{2\sum_{i=1}^{n}\hat{\lambda}_i^2}}, \tag{26}$$

which has an asymptotical standard normal distribution under H_0.

Also, the Wald type t-statistic can be used for testing overdispersion due to (22) defined as the ratio of the estimate of α to its standard error. An advantage of the Wald test over the LRT and score tests is that it only requires estimating the unrestricted model, which reduces the computations. Then, the Wald statistic denoted T_W and is given as:

$$T_W = \frac{\hat{\alpha}}{SE(\hat{\alpha})}, \tag{27}$$

where $\hat{\alpha}$ is the maximum likelihood estimate of α under the ZTNB model, and $SE(\hat{\alpha})$ is the standard error of $\hat{\alpha}$. Under H_0, the T_W statistic has an asymptotical standard normal distribution.

5. Simulation Study

In this section, a simulation study is carried out to compare the empirical power of LRT, score, and Wald tests for testing the overdispersion parameter in a zero-truncated negative binomial model under different situations. The model considered in this study is

$$Y_i \sim \text{ZTNB}(\alpha, \lambda_i), i = 1, 2, \dots, n$$

where $log\,\lambda_i = 2 - 0.8\,x_{1\,i} + x_{2\,i}$.

A set of random numbers is generated from a continuous uniform $[0, 1]$ distribution for the covariate $x_{1\,i}$, and another set is generated from a continuous uniform $[0, 0.5]$ distribution for the covariate $x_{2\,i}$ by the R program with sample sizes are chosen to be $20, 50, 100, and\ 200$ to get the simulated power of tests. For each sample size, the power of tests are computed based on 1000 replications with $\alpha = 0.01, 0.025, 0.05, 0.075, 0.1, and\ 0.25$ by calculating the proportion of times that the test value is greater than the

critical value at the significance level 0.05. The results are   given in Table 1 and displayed in Figure 1.

Table 1. Power of LRT, score, and Wald tests for overdispersion parameter.

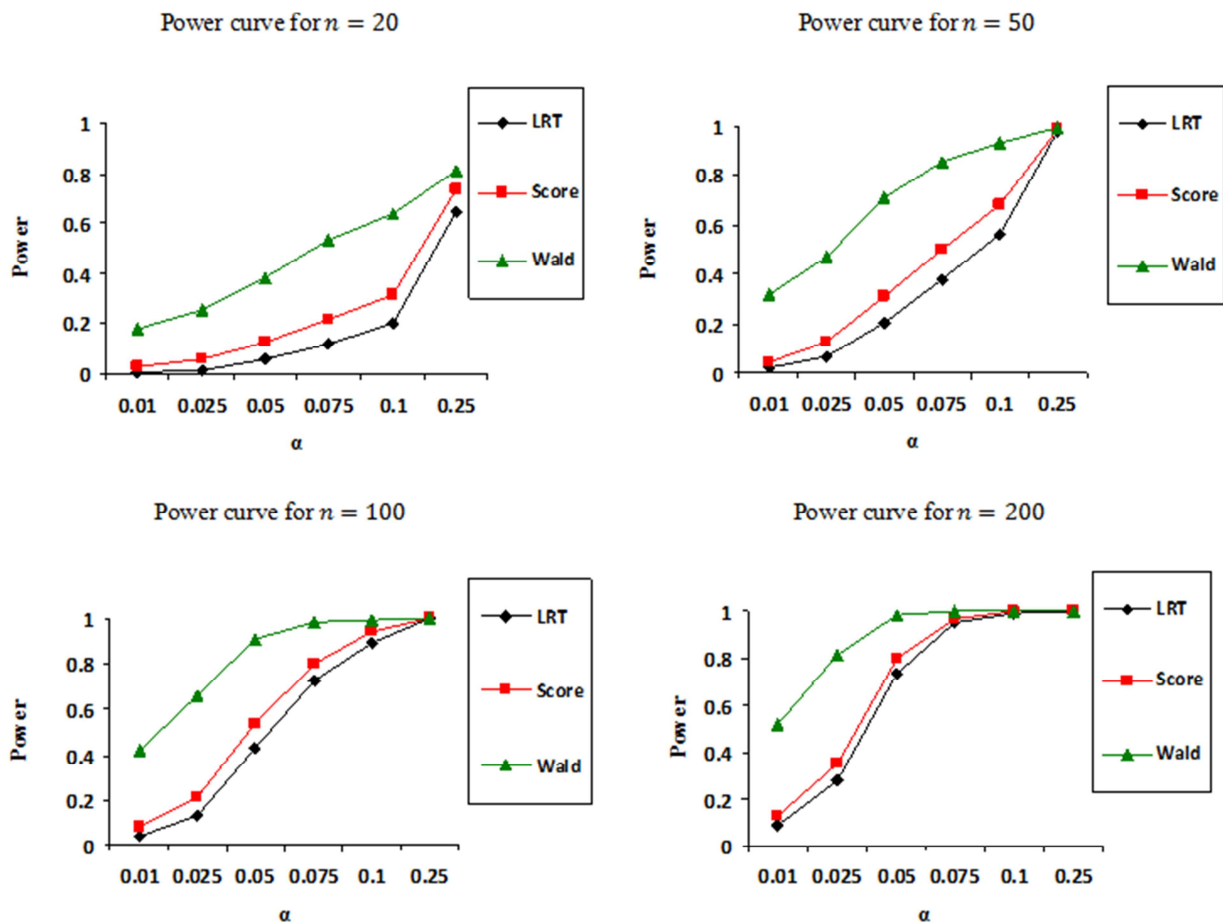| n	Method	Power					
		$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.075$	$\alpha = 0.1$	$\alpha = 0.25$
20	LRT	0.005	0.018	0.059	0.123	0.201	0.650
	Score	0.031	0.057	0.130	0.215	0.313	0.735
	Wald	0.178	0.259	0.380	0.535	0.638	0.814
50	LRT	0.025	0.065	0.203	0.382	0.557	0.977
	Score	0.046	0.127	0.310	0.502	0.679	0.988
	Wald	0.321	0.472	0.714	0.853	0.932	0.994
100	LRT	0.041	0.130	0.427	0.729	0.896	1
	Score	0.081	0.211	0.537	0.803	0.940	1
	Wald	0.420	0.663	0.909	0.985	0.995	1
200	LRT	0.085	0.283	0.734	0.951	0.996	1
	Score	0.125	0.354	0.797	0.971	0.998	1
	Wald	0.521	0.818	0.986	0.999	1	1

Figure 1. The empirical power curve for LRT, score, and Wald tests.

Based on Table 1, it is clear that the power of three tests: LRT, score, and Wald increases for all sample sizes $n = 20, 50, 100, and\ 200$ when α increases. Also, as n increases, the power of these tests increases and the Wald test has the greatest power level uniformly for all cases of α and n as displayed in Figure 1. The score test outperforms LRT test in terms of power for all values of $(n = 20, 50)$, $\alpha < 0.25$ $(n = 100, 200)$, while the LRT is powerful than the score test for all values of α when $n = 200$.

For small and moderate sample size $n = 20, 50$, the power of LRT and score tests initially increases slowly then if $\alpha > 0.1$, the power increases quickly, while the power of Wald test increases steadily as α increases. For large sample size $n = 100, 200$, the power of Wald test increases very quickly and approaches 1 when $\alpha = 0.25$ $(n = 100)$, and $\alpha = 0.1$ (n = 200), while the power of the LRT and score gets to nearly 1 when $\alpha = 0.1$ $(n = 200)$ and approaches 1 when $\alpha = 0.25$ $(n = 100, 200)$.

Moreover, as α and n increase, the difference between ZTP and ZTNB increases for all tests which emphasizes that there is very strong evidence against the fit of the ZTP model to the overdispersed data. Overall, the Wald test dominates

uniformly over score and LRT tests in terms of power for choosing between ZTP and ZTNB in the presence of overdispersion in the data.

6. Conclusion

Overdispersion is often encountered in count regression which leads to poor fit when inference about regression parameters. For testing overdispersion in ZTNB regression model against ZTP regression model, the LRT, score, and Wald tests were proposed. The empirical power for statistics of these tests was assessed under different levels of overdispersion by Monte-Carlo simulation. The results showed that the Wald test provides the highest statistical power. Thus, it is preferable for detecting the overdispersion problem in zero-truncated count data.

On the other hand, the power of overdipersion tests can be examined for right-truncated count data in future work.

References

[1] Arrabal, C. T., Silva, K. P., & Bandeira, L. N. (2014). Zero-truncated negative binomial applied to nonlinear data. JP Journal of Biostatistics, 11 (1): 55-67.

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

[3] Cameron, A. C., & Trivedi, P. K. (1990). Regression-based tests for overdispersion in the Poisson model. Journal of Econometrics, 46, 347-364.

[4] Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data. 2nd edition, Cambridge University Press.

[5] Dean, C. B. (1992). Testing for overdispersion in Poisson and binomial regression models. Journal of the American Statistical Association, 87 (418): 451-457.

[6] Dean, C. B., & Lawless, J. F. (1989). Tests for detecting overdispersion in Poisson regression models. Journal of the American Statistical Association, 84 (406): 467-472.

[7] Grogger, J. T., & Carson, R. T. (1991). Models for truncated counts. Journal of Applied Econometrics, 6, 225-238.

[8] Gurmu, S. (1991). Tests for detecting overdispersion in the positive Poisson regression model. Journal of Business & Economic Statistics, 9 (2): 215-222.

[9] Gurmu, S., & Trivedi, P. K. (1992). Overdispersion tests for truncated Poisson regression models. Journal of Econometrics, 54, 347-370.

[10] Hilbe, J. M. (2011). Negative binomial regression. 2nd edition, Cambridge University Press.

[11] Hossain, Z., Akter, R., Sultana, N., & Kabir, E. (2020). Modelling zero-truncated overdispersed antenatal health care count data of women in Bangladesh. PLoS ONE, 15 (1): 1-10.

[12] Ismah, Kurnia, A., & Sadik, K. (2020). Analysis of overdispersed count data by Poisson model. European Journal of Molecular & Clinical Medicine, 7 (10): 1400-1409.

[13] Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). Univariate discrete distributions. 3rd edition, John Wiley & Sons, Inc., New Jersey.

[14] Lee, L.-F. (1986). Specification test for Poisson regression models. International Economic Review, 27 (3): 689-706.

[15] Lee, J.-H., Han, G., Fulp, W. J., & Giuliano, A. R. (2012). Analysis of overdispersed count data: application to the human papillomavirus infection in men (HIM) study. Epidemiology and Infection, 140, 1087-1094.

[16] Matthews, J. N. S., & Appleton, D. R. (1993). An application of the truncated Poisson distribution to immunogold assay. Biometrics, 49 (2): 617-621.

[17] Molla, D. T., & Muniswamy, B. (2012). Power of tests for overdispersion parameter in negative binomial regression model. IOSR Journal of Mathematics, 1 (4): 29-36.

[18] O'Hara Hines, R. J. (1997). A comparison of tests for overdispersion in generalized linear models. Journal of Statistical Computation and Simulation, 58 (4): 323-342.

[19] Pongsapukdee, V., Khawsittiwong, P., & Yamjaroenkit, M. (2016). Test statistics for dispersion parameter in Poisson regression and generalized Poisson regression models. Silpakorn University Science and Technology Journal, 10 (2): 31-36.

[20] Sampford, M. R. (1955). The truncated negative binomial distribution. Biometrika, 42 (1-2): 58-69.

[21] Shaw, D. (1988). On-site samples' regression: problems of non-negative integers, truncation, and endogenous stratification. Journal of Econometrics, 37 (2): 211-223.

[22] Singh, J. (1978). A characterization of positive Poisson distribution and its application. SIAM Journal on Applied Mathematics, 34 (3): 545-548.

[23] Vives, J., Losilla, J.-M., Rodrigo, M.-F., & Portell, M. (2008). Overdispersion tests in count-data analysis. Psychological Reports, 103, 145-160.

[24] Winkelmann, R. (2008). Econometric analysis of count data. 5th edition, Springer.

[25] Yang, Z., Hardin, J. W., Addy, C. L., & Vuong, Q. H. (2007). Testing approaches for overdispersion in Poisson regression versus the generalized Poisson model. Biometrical Journal, 49 (4): 565-584.

[26] Yang, Z., Hardin, J. W., & Addy, C. L. (2009). A score test for overdispersion in Poisson regression based on the generalized Poisson-2 model. Journal of Statistical Planning and Inference, 139, 1514-1521.

[27] Zamani, H., & Ismail, N. (2013). Score test for testing Poisson regression against generalized Poisson alternatives. AIP Conference Proceedings, 1522, 1204-1212.

[28] Zhang, H., Pounds, S. B., & Tang, L. (2013). Statistical methods for overdispersion in mRNA-Seq count data. The Open Bioinformatics Journal, 7, 34-40.

[29] Zhao, W., Feng, Y., & Li, Z. (2010). Zero-truncated generalized Poisson regression model and its score tests. Journal of East China Normal University (Natural Science), 1, 17-23.