Probability Based Estimator for Estimation of Zero-Inflation Parameter

K. M. Sakthivel
Department of Statistics
Bharathiar University
Coimbatore - 641046
Tamilnadu, India
Email : sakthithebest@gmail.com

C. S. Rajitha
Department of Statistics
Bharathiar University
Coimbatore - 641046
Tamilnadu, India
Email : rajitha.sugun@gmail.com

Abstract— **In field of research, modeling of count data plays significant and inevitable role. In modeling count data, the assumption of symmetric and normality is not fulfilled due to excess numbers of presence of certain values in the data. Most of the time, the frequency of zero is very high hence the data is over dispersed. Since the traditional standard count models such as Poisson and negative binomial distribution will not provide better fit, hence the researchers started using zero-inflated models with an additional parameter called zero inflation parameter to deal with such type of count data. Zero inflation parameter represents the proportion of excess number of zeros in the data which highly influence the accuracy and efficiency of the zero-probability models. In this paper, we have introduced an estimator for zero inflation parameter of zero-inflated negative binomial (ZINB) distribution based on the non zero probabilities and compared the performance of the proposed estimator via a simulation study.**

AMS Subject Classification (2010): 62F10; 62P05

***Keywords*—** Zero-Inflated Models, Zero-Inflation Parameter, Maximum Likelihood Estimation, Zero- Inflated Negative Binomial Distribution, Simulation, Mean Square Error.

I. INTRODUCTION

In recent years researchers are interested in the analysis of count data having excess number of zero counts. The most commonly used model for analyzing count data is the Poisson distribution. But this distribution follows a restrictive property that mean and variance are equal. But in many applications the count data shows over dispersion and excess number of zero counts. Hence researchers utilized negative binomial distribution for modeling the over dispersed count data. And many attempts have been made by many researchers to find good alternatives for modeling the count data having over dispersion and excess number of zero counts. Zero-inflated models showed better performance for modeling count data with excess number of zero counts. Among the zero-inflated models zero-inflated Poisson distribution (Lambert, 1992), zero-inflated negative binomial distribution (Neelon et al., 2010) etc are the frequently used models. But in practice, the count data often shows over dispersion nature, hence zero-inflated negative binomial (ZINB) distribution provides appropriate fitting to the count data. Furthermore if the count data shows severe over dispersion for non zero counts, zero-inflated Poisson (ZIP) model does not provides consistent estimates of the parameters. For testing a zero-inflated Poisson model against a zero-inflated negative binomial model Ridout et al., (2001) provided a score test. Many other zero-inflated and zero modified distributions are also proposed by many researchers. A zero-inflated binomial regression model was proposed by Hall (2000) and he studied the random effects in to zero-inflated Poisson and zero- inflated binomial models. Famoye and Singh (2006) proposed a zero-inflated generalized Poisson regression model for modeling the zero–inflated count data. Further Dietz and Bohning (2000) discussed about a class of zero-modified Poisson models.

Various methods are used in the literature for estimating the parameters of the zero-inflated models. But the standard way followed by the literature is based on maximum likelihood estimation (MLE) method. An alternative way is based on the first moment embed it in a linear exponential family of distributions and there by estimate the parameters by quasi-maximum likelihood method. A ZIP model in a real data context is discussed by Nanjundan *et a*l.,(2009) and they attained the model parameters using moment estimators and compared its performance with MLE via a simulation study. For estimating the parameters of the zero-inflated Poisson distribution Nanjundan and Naika (2012) used asymptotic comparison of MLE and ME methods through study. Further Nanjundan and Naika (2013) discussed the parameter estimation of the power series distribution using method of maximum likelihood and method of moments asymptotically via EM algorithm. Becket et.al also discussed about the parameter estimation of ZIP using MLE and ME in terms of standardized bias and standardized MSE through simulation study. Recently Wagh and Kamalja (2017) proposed a new estimator called probability estimator (PE) for making inferences about the inflation parameter of ZIP distribution and compared the performance of this estimator with MLE and ME through simulation study in terms of mean square error (MSE), standard error (SE) and bias. Also Puig and Valero (2006)

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

introduced a measure called zero inflation index for measuring the zero inflation and over dispersion present in the data.

In this paper we introduced an estimator named as non-zero probability estimator (NZPE) for making inference about the inflation parameter of the zero- inflated negative binomial distribution and for demonstrating the performance of NZPE, a simulation study is carried out. The organization of the paper is as follows. Section 2 provides the structure of zero-inflated models. In section 3 details of the zero-inflated negative binomial model are given. In section 4 provides an overview of the zero inflation index and demonstrate the details of the zero inflation parameter for different values of the zero inflation index. In section 5, we proposed an estimator for making inferences about the inflation parameter of the zero-inflated negative binomial distribution. In section 6 we provide the performance of the new estimator via a simulation study. Sections 7 provide a conclusion to the study.

II. STRUCTURE OF ZERO-INFLATED MODELS

Zero-inflated models are basically mixture models that combine a point mass at zero component with the count component. The structure of zero-inflated models allows for two different processes. First process occurs the participation of values which drives whether the value is positive or zero. The second part occurs the value of the strictly positive counts. The probability mass function of the zero-inflated model can be written as

$$p\left(X = x / \Theta, \omega\right) = \begin{cases} \omega + \left(1-\omega\right) g(0, \Theta) & , if\ x = 0 \\ \left(1-\omega\right) g(x, \Theta) & , if\ x = 1,2,.. \end{cases}$$

where X is the count random variable, $g(x,\Theta)$ is the PMF of X with parameter space Θ and ω represents the proportion of excess zero counts called the inflation parameter and its value lies between 0 and 1.

III. ZERO-INFLATED NEGATIVE BINOMIAL DISTRIBUTION

For analyzing the data with over dispersion nature, lot of studies have been done in the literature. And one major cause of over dispersion is the existence of excess number of zero counts in the data. Hence in different times different probability distribution models were proposed by various researchers for analyzing the zero-inflated data. Among these models zero-inflated models plays a very significant role. The concept of zero inflation was first introduced by Neyman (1939) and Feller (1943). One of the most frequently used model for modeling the count data is the zero-inflated negative binomial distribution. The PMF of the zero-inflated negative binomial distribution (p, r, ω) can be written as

$$p\left(X = x / \lambda, r, \omega\right) = \begin{cases} \omega + \left(1-\omega\right) e^{-\lambda r} & , if\ x = 0 \\ \left(1-\omega\right) \binom{x+r-1}{x} e^{-\lambda r}\left(1-e^{-\lambda}\right)^{x} & , if\ x > 0 \end{cases}$$

$x = 0,1,2,...,\ 0 < \lambda < 1,\ 0 < \omega < 1$ and $p = e^{-\lambda}$. The ZINB distribution reduces to the ZIP distribution in the limit $r \to \infty$.

The mean and variance of the distribution are

$$E(X) = (1-\omega)\lambda$$

$$V(X) = \lambda(1-\omega)(1+\lambda(\omega+r))$$

IV. ZERO INFLATION INDEX

The concept of zero inflation index was first introduced by Puig and Valero (2006) and they defined the zero inflation index of the random variable Y as

$$z_i = 1 + \log(p_0) / mean\,(Y)$$

where p_0 represents the proportion of zero counts and $mean\,(Y)$ denotes the mean of the count random variable Y . If the zero inflation index produce the value zero, then the random variable Y follows negative binomial distribution. Otherwise the random variable follows zero inflated negative binomial distribution.

To find the nature of zero inflation parameter we generate 500 random samples from a zero-inflated negative binomial distribution with $r = 5, p = 0.5$ and $\omega = \omega_0$ where $\omega_0 = 0.1, 0.2, 0....,0.8$ for different sample sizes n=25, 50, 75 and 100. The average zero inflation index of the above mentioned sample sizes across ω_0 is given in table 1.The table shows that for most of the sample sizes, the zero inflation parameter is greater than zero and the true parameter value ω_0 is sometimes overestimated or under estimated.

TABLE I. AVERAGE ZERO INFLATION INDEX OF ZINB FOR DIFFERENT SAMPLE SIZES ACROSS ω

Sample size (n)	**Average zero-inflation index**							
	0.1	***0.2***	***0.3***	***0.4***	***0.5***	***0.6***	***0.7***	***0.8***
25	0.11	0.22	0.32	0.37	0.42	0.44	0.47	0.51
50	0.15	0.25	0.32	0.38	0.42	0.46	0.49	0.56
75	0.15	0.26	0.32	0.39	0.43	0.46	0.49	0.58
100	0.15	0.26	0.33	0.39	0.43	0.47	0.49	0.51

V. NON - ZERO PROBABILITY ESTIMATOR

For estimating the parameters of the zero-inflated negative binomial distribution some studies have been done in the literature (Astuti and Mulyanto, 2016; Nanjundan and Naika, 2013). In this paper we proposed a new estimator for making inferences about the inflation parameter of the zero-inflated negative binomial distribution and the new estimator is named as non-zero probability estimator (NZPE). This estimator doesn't need any of the iterative methods, so that the calculation is very simple for estimating the inflation parameter of ZINB distribution. The new estimator can be obtained as follows.

$$\hat{\omega}_{NZPE} = \frac{P_{ZINB}(Y \neq 0) - P_{NB}(Y \neq 0)}{P_{NB}(Y \neq 0)}$$

Where $P_{ZINB}(Y \neq 0)$ represents the probability of the ZINB distribution at $Y \neq 0$ and $P_{NB}(Y \neq 0)$ denotes the probability of NB distribution at $Y \neq 0$.

VI. Estimation of Inflation Parameter of ZINB Using Simulation Study

Here we conducted a simulation study and showed the performance of the NZPE through mean square error (MSE), standard error (SE) and bias.

We adopted the following algorithm for the simulation study

1. Generate m=500 random sample for each of size n from the $ZINB(r, p, \omega)$ for $r = 5$, $p = 0.5$ and $\omega = \omega_0$ where $\omega_0 = 0.1, 0.2, ...0.8$
2. Estimate $MSE(\hat{\omega}_{NZPE})$ as follows

$$MSE(\hat{\omega}_{NZPE}) = \frac{\sum_{i=1}^{n} (\hat{\omega}^{(i)}_{NZPE} - \omega)^2}{n}$$

3. The sample variance $v(\hat{\omega}_{NZPE})$ are evaluated as follows.

$$v(\hat{\omega}_{NZPE}) = \frac{\sum_{i=1}^{n} (\hat{\omega}^{i}_{NZPE} - \bar{\hat{\omega}}_{NZPE})^2}{n}$$

where $\bar{\hat{\omega}}_{NZPE} = \frac{\sum_{i=1}^{n} \hat{\omega}^{i}_{NZPE}}{n}$

4. The average biases of the estimates for NZPE are calculated as follows.

$$bias(\bar{\hat{\omega}}_{NZPE}) = \frac{\sum_{i=1}^{n} (\hat{\omega}^{(i)}_{NZPE} - \omega)}{n}$$

5. Repeat steps (1)-(4) for m=500 samples and $\omega_0 = 0.1, 0.2, ...0.8$.

A simulation study is carried out for sample sizes n = 25, 50 and 100 and following tables (2-5) shows the results of the study for different sample sizes.

The Fig.1 and Fig 2 shows the performance of NZPE for making inference about the inflation parameter ω of ZINB. The behavior (diagram) of NZPE for sample sizes 25, 50, 75 and 100 shows that as the sample size increases error will be reduced. From table 2 and the diagrams it can be seen that the NZPE performs better as the sample size increases.

TABLE II. MSE, SE, and Bias for ω of ZINB distribution for n=25

Inflation parameter	n = 25		
	MSE	*SE*	*BIAS*
0.1	0.000022	0.000439	-0.000197
0.2	0.000085	0.000361	-0.000380
0.3	0.000190	0.000299	-0.000590
0.4	0.000360	0.000238	-0.000780
0.5	0.000545	0.000210	-0.00990
0.6	0.000787	0.000242	-0.001180
0.7	0.000997	0.000175	-0.001390
0.8	0.001299	0.000091	-0.001570

TABLE III. MSE, SE, and Bias for ω of ZINB distribution for n=50

Inflation parameter	n = 50		
	MSE	*SE*	*BIAS*
0.1	0.000020	0.000435	-0.000200
0.2	0.000080	0.000341	-0.000400
0.3	0.000180	0.000272	-0.000600
0.4	0.000320	0.000228	-0.000800
0.5	0.000500	0.000199	-0.001000
0.6	0.000720	0.000230	-0.001200
0.7	0.000980	0.000161	-0.001400
0.8	0.001280	0.000084	-0.001600

TABLE IV. MSE, SE, and Bias for ω of ZINB distribution for n=75

Inflation parameter	n = 75		
	MSE	*SE*	*BIAS*
0.1	0.000019	0.000430	-0.000201
0.2	0.000078	0.000339	-0.000402
0.3	0.000178	0.000270	-0.000610
0.4	0.000318	0.000226	-0.000820
0.5	0.000497	0.000199	-0.001030
0.6	0.000715	0.000228	-0.001250
0.7	0.000977	0.000160	-0.001420
0.8	0.001275	0.000082	-0.001650

TABLE V. MSE, SE, and Bias for ω of ZINB distribution for n=100

Inflation parameter	n = 100		
	MSE	*SE*	*BIAS*
0.1	0.000015	0.000428	-0.000205
0.2	0.000067	0.000336	-0.000408
0.3	0.000168	0.000265	-0.000617
0.4	0.000308	0.000219	-0.000828
0.5	0.000485	0.000193	-0.001035
0.6	0.000709	0.000221	-0.001259
0.7	0.000968	0.000156	-0.001428
0.8	0.001265	0.000076	-0.001657

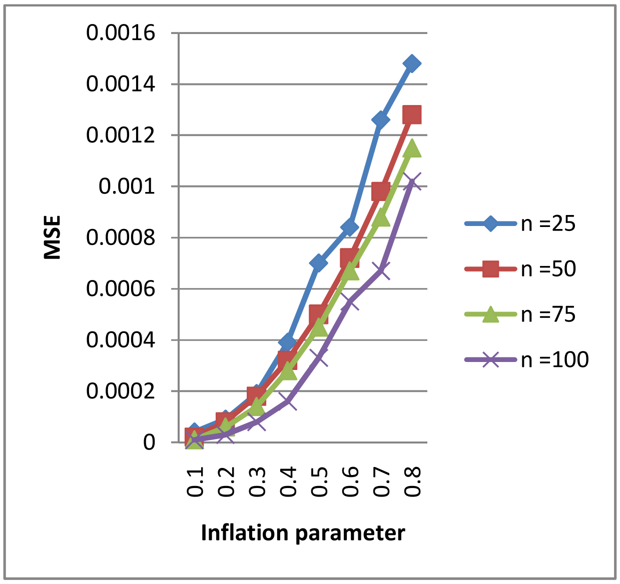

Fig. 1. SE of the NZPE for different sample sizes

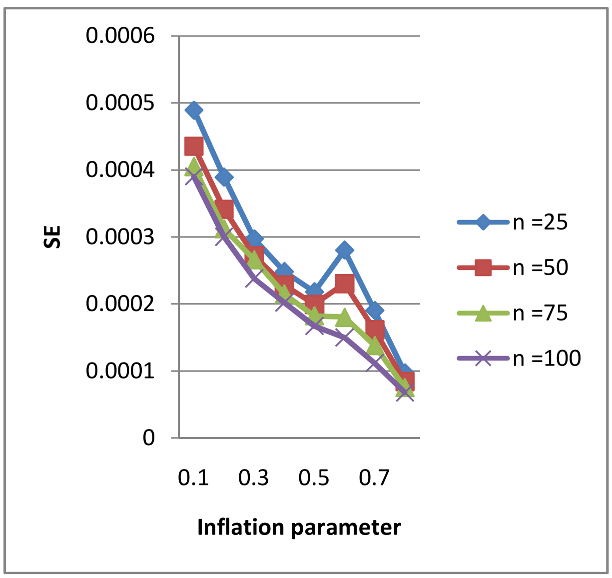

Fig. 2. SE of the NZPE for different sample sizes

VII. Conclusion

In this paper we discussed about some inflated probability distributions and estimation of inflation parameter of ZINB distribution. Instead of using MLE for estimating the parameter of ZINB we proposed a new estimator for making inferences about the inflation parameter named as non-zero probability estimator (NZPE). And performed a simulation study for showing the performance of the proposed NZPE for different sample sizes. The result also shows that for increasing value of inflation rate, the error terms such as MSE, SE and bias is decreasing gradually. Hence NZPE performs better for increasing value of inflation rate and increasing values of sample size.

References

[1] C C. Astuti1, and A D. Mulyanto, **"**Estimation Parameters and Modelling Zero-Inflated Negative Binomial," Cauchy – Jurnal Matematika Murni Dan Aplikasi, vol. 4(3), pp. 115-119, 2016

[2] S. Beckett, J. Jee, T. Ncube, S. Pompilus, Q. Washington, A. Singh, and N. Pal, "Zero-inflated Poisson (ZIP) distribution: parameter estimation and applications to model data from natural calamities," Involve: A Journal of Mathematics, vol. 7(6),pp.751.–767, 2014

[3] P.C. Consul, and G.C. Jain, "On some interesting properties of the generalized Poisson distribution," Biometrical Journal, vol.15,pp. 495–500, 1973b

[4] F.Famoye and W. Wang, Censored generalized Poisson regression model," Computational Statistics and Data Analysis, vol.46, pp.547-560,2004.

[5] W. Feller, "On a general class of contagious distributions," Annals of Mathematical Statistics, 14(4):389–400, 1943.

[6] D.B Hall, "Zero-inflated Poisson and binomial regression with random effects: A case study," Biometrics, vol. 56, pp. 1030- 1039, 2000.

[7] D. Lambert, "Zero-inflated Poisson regression with an application to defects in manufacturing," Technometrics, vol. 34(1),pp.1–14, 1992.

[8] G. Nanjundan, A. Loganathan, and T. R. Naika, "An Empirical Comparison of Maximum Likelihood and Moment Estimators of Parameters in a Zero-Inflated Poisson Model," Mapana Journal of Sciences, vol. 8(2), pp. 59-72, 2009.

[9] G. Nanjundan, and T. R. Naika, "Asymptotic comparison of method of moments estimators and maximum likelihood estimators of parameters in zero-inflated poisson model," AppliedMathematics, vol. 3, pp.610–616, 2012

[10] G. Nanjundan, and T. R. Naika, "Estimation of Parameters in a Zero-Inflated Power Series Model," International Journal of Statistika and Mathematika, vol. 6(3),pp.109-114, 2013.

[11] B. Neelon, A.J. O'Malley, and S.L.T. Normand, "A Bayesian model for repeated measures zero-inflated count data with application to psychiatric outpatient service use," Statistical Modelling, vol. 10, pp. 421-439, 2010.

[12] J. Neyman, "On a new class of contagious distributions applicable in entomology and bacteriology," Annals of Mathematical Statistics, vol. 10(1), pp.35–57, 1939.

[13] M. K. Patil, and D. T. Shirke, "Testing parameter of the power series distribution of a zero-inflated power series model," Statistical Methodology, vol. 4(4), pp. 393–406, 2007.

[14] M. K.Patil, and D. T. Shirke, "Tests for equality of inflation parameters of two zero-inflated power series distributions," Communications in Statistics-Theory and Methods,vol. 40(14), pp. 2539–2553, 2011.

[15] P. Puig, and J.Valero, "Count data distributions: Some characterizations with applications," Journal of American Statistical Association, vol.101(473),pp.332–340, 2006.

[16] M. Ridout, J. Hinde, , and C.G.B. Demetrio, "A score test for a zero-Inflated Poisson regression model against zero-inflated negative binomial alternatives," Biometrics, vol. 57(1), pp. 219-223, 2001 DOI: 10.1111/j.0006-341X.2001.00219.x

[17] Y. S. Wagh, and K. K. Kamalja, "Comparison of methods of estimation for parameters of Generalized Poisson distribution through simulation study,"

Communications in Statistics-Simulation and Computation, vol. 46(5), pp.4098–4112, 2017.

[18] Y. S. Wagh, and K. K. Kamalja, "Zero-inflated models and estimation in zero-inflated Poisson distribution,"Communications in Statistics-Simulation and Computation, vol. 0, pp.1-18, 2017.

[19] E. Dietz and D. Bohning, " On estimation of the Poisson parameter in zero-modified Poisson models," Computational Statistics and Data Analysis, vol. 34, pp.441–459,2000.