



Original Article

Zero inflated negative binomial-generalized exponential distribution and its applications

Sirinapa Aryuyuen, Winai Bodhisuwan*, and Thidaporn Supapakorn

*Department of Statistics, Faculty of Science,
Kasetsart University, Chatuchak, Bangkok, 10900 Thailand.*

Received 3 December 2013; Accepted 16 April 2014

Abstract

In this paper, we propose a new zero inflated distribution, namely, the zero inflated negative binomial-generalized exponential (ZINB-GE) distribution. The new distribution is used for count data with extra zeros and is an alternative for data analysis with over-dispersed count data. Some characteristics of the distribution are given, such as mean, variance, skewness, and kurtosis. Parameter estimation of the ZINB-GE distribution uses maximum likelihood estimation (MLE) method. Simulated and observed data are employed to examine this distribution. The results show that the MLE method seems to have high-efficiency for large sample sizes. Moreover, the mean square error of parameter estimation is increased when the zero proportion is higher. For the real data sets, this new zero inflated distribution provides a better fit than the zero inflated Poisson and zero inflated negative binomial distributions.

Keywords: overdispersion, zero inflated distribution, count data with extra zeros, random variate generation, MLE

1. Introduction

Poisson distribution provides a standard model for the analysis of count data with the assumption of equal mean and variance. However, in practice, count data often shows overdispersion, the variance is greater than the mean. In practice, negative binomial (NB) distribution was introduced to solve this problem, and it has become increasingly popular as a more flexible alternative to fit models. The NB distribution is a better fit for over-dispersed count data which is not necessarily heavy-tailed (Wang, 2011).

For over-dispersed count data, some mixed NB distributions offer a better fit when compared with the Poisson and NB distributions. Such findings are related to those of Gomez-Déniz *et al.* (2008), Zamani and Ismail (2010), Wang (2011), and Pudprommarat *et al.* (2012). Recently, a negative binomial-generalized exponential (NB-GE) distribution was

employed to fit count data. It was obtained by mixing the NB distribution with a generalized exponential (GE) distribution. Also, the NB-GE distribution is more appropriate to fit count data with overdispersion and heavy-tailed datasets than the Poisson and NB distributions (Aryuyuen and Bodhisuwan, 2013).

Moreover, in practice, one frequent demonstration of overdispersion is a zero count incidence. Count data with extra zeros occur in many fields, such as public health, epidemiology, medicine, sociology, engineering and agriculture. Standard discrete distribution may fail to fit such data either because of zero inflation or over/underdispersion. Now, there is increased interest in a zero inflated distribution to account for extra zeros in data (Xie *et al.*, 2009). Count data with extra zeros create problems of violating basic assumptions implicit in the standard distribution. Failure to account for extra zeros may result in biased parameter estimates and misleading inference. Because the zero inflated distributions usually provide better statistical fit, some researchers, e.g., Lambert (1992), Greene (1994), Hall (2000), Famoye and Singh (2003), Bodhisuwan (2011) proposed these distributions.

* Corresponding author.
Email address: fsciwnb@ku.ac.th

As mentioned above, the mixed distribution defines one of the most important ways to obtain a new probability distribution in applied probability and operational research (Gomez-Deniz *et al.*, 2008). For the purpose, we are looking for a new zero inflated distribution which is a more flexible alternative to fit count data with excess zeros.

The rest of this article is organized as follows. We propose the new zero inflated distribution that is a zero inflated negative binomial-generalized exponential (ZINB-GE) distribution. Some characteristics, graphs of probability mass function (pmf) and a random variate generation of ZINB-GE distribution are introduced in Section 2. In Section 3 the maximum likelihood estimation (MLE) method is handled to estimate the parameters of ZINB-GE distribution. Numerical examples are provided in Section 4, including simulated data which are utilized to examine the efficiency of MLE method, and real data sets that are used to evaluate the performance of the proposed distribution. Finally, some conclusions are presented in Section 5.

2. New zero inflated negative binomial distribution

The zero inflated (ZI) distribution can be used to fit count data with extra zeros, and assumes that the observed data are the result of a two-part process; a process that generates structural zeros and a process that generates random counts. The model of ZI can be summarized as follows:

$$P(X = x | \phi) = \phi \omega_0(x) + (1 - \phi) f(x; \theta), \quad (1)$$

where X is the count variable, ϕ is the extra proportion of zeros, $f(x; \theta)$ is the pmf of X with the parameter θ , and $\omega_0(x) = 1$ if $x = 0$; otherwise, $\omega_0(x) = 0$.

2.1 The ZINB-GE distribution

The ZINB-GE distribution is a new mixture distribution, combining Bernoulli and NB-GE distributions. First, we introduce a definition and some characteristics of NB-GE distribution as follows.

Definition 1 Let Y be a random variable of the NB-GE distribution. Random variable Y has the NB distribution with the parameters r and $p = \exp(-\lambda)$, where λ is distributed as the GE distribution with the positive parameters α and β , i.e.,

$y | \lambda \sim \text{NB}(r, p = \exp(-\lambda))$ and $\lambda \sim \text{GE}(\alpha, \beta)$. The pmf of Y is given by

$$f(y) = \binom{r+y-1}{y} \sum_{j=0}^y \binom{y}{j} (-1)^j M_{(r+j)}, \quad y = 0, 1, 2, \dots, \quad (2)$$

$$M_{(u)} = \frac{\Gamma(\alpha+1)\Gamma(1+u/\beta)}{\Gamma(\alpha+u/\beta+1)} \text{ for } r, \alpha, \beta > 0 \quad (3)$$

where $\Gamma(\cdot)$ is a gamma function denoted as $\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy$, for $t > 0$, and

Next, the mean and variance of Y are respectively given by

$$E(Y) = r(\delta_{(1)} - 1) \text{ and}$$

$$\text{Var}(Y) = \sigma_Y^2 = r(r+1)\delta_{(2)} - r(r\delta_{(1)} + 1)\delta_{(1)},$$

where $\delta_{(u)}$ is

$$\delta_{(u)} = \frac{\Gamma(\alpha+1)\Gamma(1-u/\beta)}{\Gamma(\alpha-u/\beta+1)}. \quad (4)$$

Now, by using the expression (1) and Definition 1, we obtain the closed formulas for the pmf and some characteristics of ZINB-GE as follows.

Definition 2 Let X be a random variable of ZINB-GE distribution with the parameters r, α, β and ϕ , denoted as $X \sim \text{ZINB-GE}(r, \alpha, \beta, \phi)$. The pmf of X is given by

$$f(x) = \begin{cases} \phi + (1-\phi)M_{(r)} & \text{if } x = 0, \\ (1-\phi) \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j M_{(r+j)} & \text{if } x = 1, 2, \dots \end{cases} \quad (5)$$

where $0 < \phi < 1$, $r, \alpha, \beta > 0$, and $M_{(u)}$ is defined in (3).

The graphs of ZINB-GE's pmf distribution with specified parameters r, α, β and ϕ , offered in Figure 1, shows that the distribution will be flat when parameter ϕ is increasing. However, the distribution will be flat when the parameter β is decreasing. In addition, the pmf of ZINB-GE distribution can take different shapes when values of α differ.

Theorem 1 If $X \sim \text{ZINB-GE}(\phi, r, \alpha, \beta)$ then some characteristics of X are as follows.

(a) The mean and variance of X are given by

$$E(X) = (1-\phi)(r\delta_{(1)} - r) \text{ and } \text{Var}(X) = \sigma_X^2, \quad (6)$$

where $\delta_{(u)}$ is defined in (4), and

$$\sigma_X = \left[r \left[(r+1)\delta_{(2)} - (2r+1)\delta_{(1)} + r \right] (1-\phi) - \left[r(\delta_{(1)} - 1)(1-\phi) \right]^2 \right]^{1/2}.$$

(b) The skewness and kurtosis of X are

$$Sk(X) =$$

$$\left\{ r \left[(r+1)(r+2)\delta_{(3)} - 3(r+1)^2\delta_{(2)} + (3r^2 + 3r + 1)\delta_{(1)} - r^2 \right] (1-\phi) - 3r^2(\delta_{(1)} - 1) \right. \\ \left. \times \left[(r+1)\delta_{(2)} - (2r+1)\delta_{(1)} + r \right] (1-\phi)^2 + 2 \left[r(\delta_{(1)} - 1)(1-\phi) \right]^3 \right\} / \sigma_X^3, \quad (7)$$

$$Ku(X) =$$

$$\left\{ r \left[(r+1)(r+2)(r+3)\delta_{(4)} - 2(r+1)(2r^2 + 7r + 6)\delta_{(3)} + (r+1) \right. \right. \\ \left. \times (6r^2 + 12r + 7)\delta_{(2)} - (2r+1)(2r^2 + 2r + 1)\delta_{(1)} + r^3 \right] (1-\phi) - 4r^2(\delta_{(1)} - 1) \\ \times \left[(r+1)(r+2)\delta_{(3)} - 3(r+1)^2\delta_{(2)} + (3r^2 + 3r + 1)\delta_{(1)} - r^2 \right] (1-\phi)^2 + 6r^3(\delta_{(1)} - 1)^2 \\ \times \left[(r+1)\delta_{(2)} - (2r+1)\delta_{(1)} + r \right] (1-\phi)^3 - 3 \left[r(\delta_{(1)} - 1)(1-\phi) \right]^4 \right\} / \sigma_X^4. \quad (8)$$

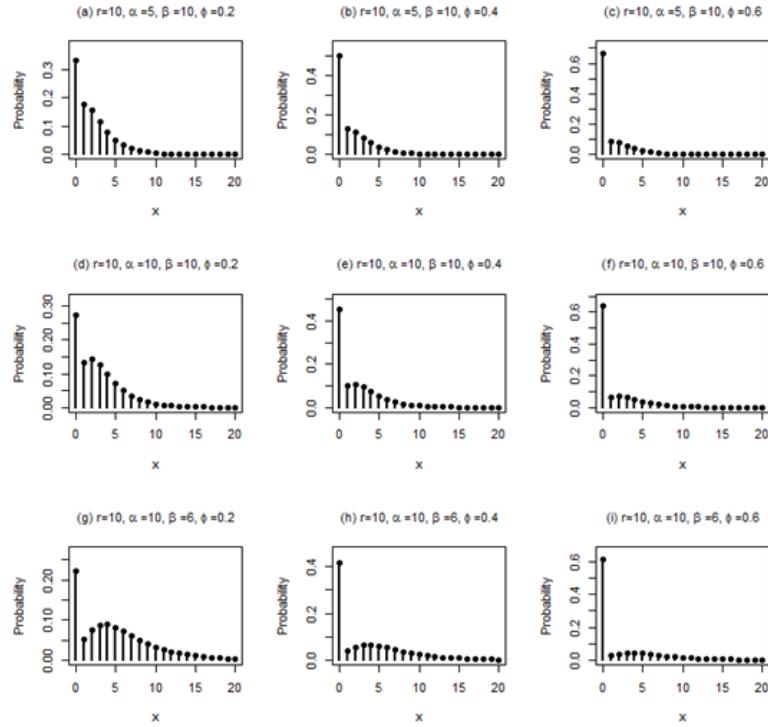


Figure 1. The pmf of ZINB-GE distribution with specified parameters

The proof of Theorem 1 is in the Appendix.

2.2 Random variate generation of ZINB-GE distribution

To generate a random variable X from the ZINB-GE (ϕ, r, α, β) , one can use the following algorithm:

- 1) Generate U from the uniform distribution, $U(0,1)$.
- 2) Set $\lambda = -\frac{1}{\beta} \log(1-U^{1/\alpha})$ from the GE distribution, $\lambda \sim \text{GE}(\alpha, \beta)$.
- 3) Generate Y from the $\text{NB}(r, p = \exp(-\lambda))$ distribution.
- 4) Generate U^* from the uniform distribution, $U(0,1)$.
- 5) If $U^* > \phi$, then set $X = Y$; otherwise, $X = 0$.

3. Parameter estimation of ZINB-GE distribution

In this section, the parameter estimation of ZINB-GE distribution via MLE procedure is provided. The likelihood function of ZINB-GE (ϕ, r, α, β) is given by

$$L(\phi, r, \alpha, \beta) = \prod_{i=1}^n \left[I_{(x_i=0)} (\phi + (1-\phi)M_{(r)}) + I_{(x_i>0)} \left((1-\phi) \binom{r+x_i-1}{x_i} \sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j M_{(r+j)} \right) \right],$$

then we can write the log-likelihood of the ZINB-GE (ϕ, r, α, β) as

$$\mathcal{L} = \log L(\phi, r, \alpha, \beta)$$

$$\begin{aligned} &= \sum_{i=1}^n \left[I_{(x_i=0)} \log \left(\phi + (1-\phi)M_{(r)} \right) \right. \\ &+ I_{(x_i>0)} \left(\log(1-\phi) + \log \Gamma(r+x_i) - \log \Gamma(r) - \log \Gamma(x_i+1) \right. \\ &+ \log \Gamma(\alpha+1) + \log \sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \frac{\Gamma(1+(r+j)/\beta)}{\Gamma(\alpha+(r+j)/\beta+1)} \left. \right] \end{aligned}.$$

By differentiating the log-likelihood function of ZINB-GE distribution, partial derivatives of the log likelihood function with respect to ϕ , r , α and β are given by

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \phi} &= \sum_{i=1}^n \left[I_{(x_i=0)} \left(\frac{\Gamma(\alpha+r/\beta+1) - \Gamma(\alpha+1)\Gamma(1+r/\beta)}{\phi\Gamma(\alpha+r/\beta+1) + (1-\phi)\Gamma(\alpha+1)\Gamma(1+r/\beta)} \right) - I_{(x_i>0)} \left(\frac{1}{1-\phi} \right) \right] \\ \frac{\partial \mathcal{L}}{\partial r} &= \sum_{i=1}^n \left[I_{(x_i=0)} \left(\frac{(1-\phi)\Gamma(\alpha+1)}{\phi + (1-\phi)M_{(r)}} \frac{\partial}{\partial r} \frac{\Gamma(1+r/\beta)}{\Gamma(\alpha+r/\beta+1)} \right) \right. \\ &\quad \left. + I_{(x_i>0)} \left(\psi(r+x_i) - \psi(r) + \frac{1}{\zeta(r+j, \alpha, \beta)} \frac{\partial}{\partial r} \zeta(r+j, \alpha, \beta) \right) \right], \end{aligned}$$

where $\psi(s) = \frac{\Gamma'(s)}{\Gamma(s)}$, and

$$\zeta(r+j, \alpha, \beta) = \sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \frac{\Gamma(1+(r+j)/\beta)}{\Gamma(\alpha+(r+j)/\beta+1)}.$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \alpha} &= \sum_{i=1}^n \left[I_{(x_i=0)} \left(\frac{(1-\phi)\Gamma(1+r/\beta)}{\phi+(1-\phi)M_{(r)}} \frac{\partial}{\partial \alpha} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+r/\beta+1)} \right) \right. \\ &\quad \left. + I_{(x_i>0)} \left(\psi(\alpha+1) + \frac{1}{\zeta(r+j, \alpha, \beta)} \frac{\partial}{\partial \alpha} \zeta(r+j, \alpha, \beta) \right) \right], \\ \frac{\partial \mathcal{L}}{\partial \beta} &= \sum_{i=1}^n \left[I_{(x_i=0)} \left(\frac{(1-\phi)\Gamma(\alpha+1)}{\phi+(1-\phi)M_{(r)}} \frac{\partial}{\partial \beta} \frac{\Gamma(1+r/\beta)}{\Gamma(\alpha+r/\beta+1)} \right) \right. \\ &\quad \left. + I_{(x_i>0)} \left(\frac{1}{\zeta(r+j, \alpha, \beta)} \frac{\partial}{\partial \beta} \zeta(r+j, \alpha, \beta) \right) \right]. \end{aligned}$$

The MLE solutions of the parameter estimates can be obtained by using numerical optimization with the *nlm* function in the R program (R Core Team, 2012). The R code of parameter estimation of ZINB-GE distribution using the MLE method is given in the Appendix.

4. Numerical study

This section presents the efficiency of the MLE method for parameter estimation of ZINB-GE distribution by using simulated data. In addition, we illustrate the application study of ZINB-GE distribution compared to the zero inflated Poisson (ZIP) and zero inflated negative binomial (ZINB) distributions.

4.1 Simulation study

In illustrating the simulation study, the sample data generated from the ZINB-GE distribution with specified parameters ($r = 10$, $\alpha = 10$, $\beta = 10$) and four values for proportion of zero ($\phi = 0.2, 0.4, 0.6, 0.8$) for the sample sizes (n) as 50, 100 and 200, respectively. In each situation, the parameters are estimated from 500 replications. Hence the maximum likelihood estimators may be biased. The biased value is a difference value between the estimator and true parameter values. The sample average of the estimated parameter, bias, variance, standard deviation (SD), and mean squared error (MSE) are computed by the formulas: $\hat{\theta}_{av} = \frac{1}{500} \sum_{t=1}^{500} \hat{\theta}_t$, $\text{Bias}(\hat{\theta}_{av}) = \hat{\theta}_{av} - \theta$, $\text{Var}(\hat{\theta}_{av}) = \frac{1}{500-1} \sum_{t=1}^{500} (\hat{\theta}_t - \hat{\theta}_{av})^2$, $\text{SD}(\hat{\theta}_{av}) = \sqrt{\text{Var}(\hat{\theta}_{av})}$ and $\text{MSE}(\hat{\theta}_{av}) = \text{Var}(\hat{\theta}_{av}) + \text{Bias}^2(\hat{\theta}_{av})$.

Table 1 illustrates statistic values from the results of studies. From these results, MLE seems to have high-efficiency when the sample size is large. The efficiency of MLE seems to be poor for small sample sizes. In addition, the MSE of parameter estimation is increasing when the zero proportion is higher as in Figure 2, which displays the graphs of MSE for parameter estimates of ZINB-GE distribution.

4.2 Application study

We used two data sets for this part of the analysis to illustrate the applications of the ZINB-GE distribution. The first data set has the number of hospital stays by United States residents aged 66 and over (see Flynn, 2009), which is shown in Table 2. This data has 80.37% of zeros and the sample index of dispersion is 1.882. In addition, Table 3 shows the number of units of consumers good purchased by households over 26 weeks (see Lindsey, 1995). This data set has 80.60% of zeros and the sample index of dispersion is 3.337. These examples have sample indices of dispersion bigger than 1 and a high percentage of zeros.

For model selection in this study, we use the criteria of AIC (Akaike information criterion) and BIC (Bayesian information criterion), the goodness of fit test (chi-squared test: χ^2 -test) is used to compare between observed and expected values of data. From the results in Table 2-3, we found that the AIC and BIC values for ZINB-GE distribution are the smallest when compared with existing models. Thus, the ZINB-GE distribution can be chosen as the best model. Also, based on the *p*-values of chi-squared test, the proposed distribution is appropriate to fit the data unlike the ZIP and ZINB distributions. These three criteria indicate that the ZINB-GE distribution is the best fit, while the ZIP and ZINB distributions are very poor fits.

5. Conclusions

This work proposes the new zero inflated distribution, which is called the zero inflated negative binomial-general-

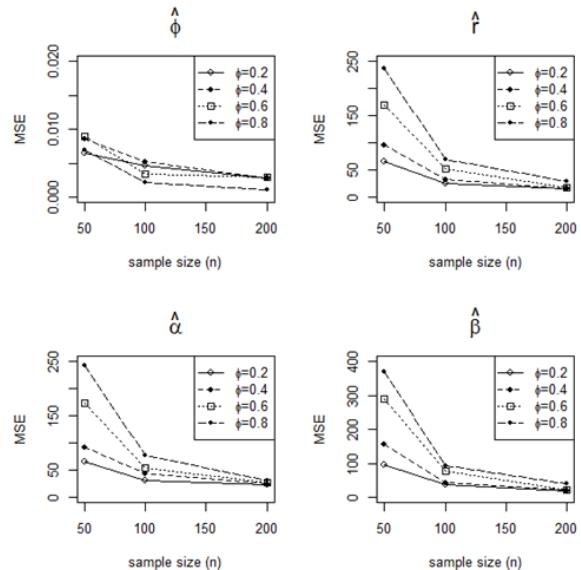


Figure 2. The MSE of parameter estimates of ZINB-GE distribution

Table 1. The parameter estimates of ZINB-GE distribution with different parameter ϕ

Parameter		$n = 50$		$n = 100$		$n = 200$	
		Estimate (SD)	MSE	Estimate (SD)	MSE	Estimate (SD)	MSE
$\phi=0.2$	ϕ	0.19 (0.08)	0.0065	0.19 (0.07)	0.0046	0.18 (0.05)	0.0027
	r	12.60 (7.62)	64.83	10.24 (5.04)	25.51	9.19 (3.87)	15.61
	α	10.82 (8.06)	65.64	9.31 (5.54)	31.13	8.21 (4.47)	23.19
	β	12.72 (9.31)	94.05	10.26 (5.98)	35.85	8.82 (4.01)	17.45
$\phi=0.4$	ϕ	0.39 (0.09)	0.0086	0.38 (0.07)	0.0052	0.39 (0.05)	0.0028
	r	14.81 (8.49)	95.28	11.07 (5.55)	32.01	9.66 (4.01)	16.23
	α	12.59 (9.30)	93.13	9.43 (6.67)	44.75	8.96 (4.85)	24.59
	β	15.52 (11.20)	155.97	10.98 (6.58)	44.24	9.57 (4.57)	21.09
$\phi=0.6$	ϕ	0.58 (0.09)	0.0090	0.59 (0.06)	0.0034	0.59 (0.05)	0.0029
	r	17.14 (10.85)	168.63	12.52 (6.75)	51.92	10.31 (4.07)	16.62
	α	14.65 (12.33)	173.65	11.28 (7.26)	54.39	9.72 (5.15)	26.55
	β	18.29 (14.87)	289.80	12.70 (8.43)	78.29	10.28 (4.73)	22.44
$\phi=0.8$	ϕ	0.78 (0.08)	0.0070	0.79 (0.05)	0.0022	0.80 (0.03)	0.0011
	r	19.84 (11.86)	237.38	13.75 (7.37)	68.40	11.07 (5.58)	28.49
	α	17.25 (13.92)	243.34	12.25 (13.74)	77.55	10.77 (5.58)	31.75
	β	20.25 (16.29)	370.38	13.74 (8.92)	93.50	11.17 (6.21)	39.87

ized exponential distribution. In particular, the closed form and some characteristics of the proposed distribution are introduced. Parameter estimation is also implemented by using MLE, and the usefulness of the ZINB-GE distribution is illustrated by real and generated data. Based on the results, the proposed distribution is the best fit while the ZIP and ZINB distributions are very poor fits for count data with extra zeros. In conclusion, the ZINB-GE distribution is a flexible alternative for analysis of count data characterized by extra zeros. Also, the MLE method seems to have high-efficiency for large sample sizes, and the MSE of parameter estimation increases when the zero proportion is higher.

Acknowledgements

The authors wish to gratefully acknowledge to the referee of this paper who helped to clarify and improve the presentation. The first author would like to thank the Research Professional Development Project under the Science Achievement Scholarship of Thailand (SAST) for funding.

References

Aryuyuen, S. and Bodhisuwan, W. 2013. The negative binomial-generalized exponential (NB-GE) distribution. *Applied Mathematical Sciences*. 7, 1093-1105.

Table 2. Observed and expected frequencies of the number of hospital stays of United States residents aged 66 and over

Number of hospital stays	Observed value	Expected value by fitting distribution		
		ZIP	ZINB	ZINB-GE
0	3541	1816.5	3541.1	3541.1
1	599	1609.5	533.4	601.0
2	176	712.9	217.7	168.7
3	48	210.6	68.7	56.4
4	20	46.7	18.9	21.6
5	12	8.4	4.4	9.7
6	5	1.3	1.3	4.4
7	1	0.1	0.5	1.8
8	4	0.0	0.0	1.3
Parameter estimates		$\hat{\phi} = 0.6659$	$\hat{\phi} = 0.6040$	$\hat{\phi} = 0.1645$
		$\hat{\lambda} = 0.8859$	$\hat{r} = 3.9683$	$\hat{r} = 2.2040$
			$\hat{p} = 0.8415$	$\hat{\alpha} = 1.0331$
				$\hat{\beta} = 7.3573$
AIC		6122	6078	6022
BIC		6135	6097	6048
χ^2		2831.93	58.31	3.07
Degree of freedom		3	2	2
p-value		<0.0001	<0.0001	0.2155

Table 3. Observed and expected frequencies of the number of units of consumers good purchased by households over 26 weeks

Units of consumers good	Households/ Observed value	Expected value by fitting distribution		
		ZIP	ZINB	ZINB-GE
0	1612	93.6	1612.0	1613.4
1	164	286.8	81.4	175.6
2	71	439.0	89.0	74.6
3	47	447.8	76.0	40.4
4	28	342.6	55.6	24.8
5	17	209.8	36.6	16.8
6	12	107.0	22.2	12.0
7	12	46.8	12.8	9.0
8	5	18.0	7.0	7.0
9	7	6.0	3.8	5.8
10-11	9	2.4	3.0	8.8
12-17	10	0.2	0.6	6.4
18-27	6	0.0	0.0	5.4
Parameter estimates		$\hat{\phi} = 0.7965$	$\hat{\phi} = 0.7837$	$\hat{\phi} = 0.3861$
		$\hat{\lambda} = 3.0609$	$\hat{r} = 4.9719$	$\hat{r} = 1.7860$
			$\hat{p} = 0.6332$	$\hat{\alpha} = 0.4325$
				$\hat{\beta} = 1.6496$
AIC		3893	3638	3477
BIC		3904	3655	3499
χ^2		26000.20	209.81	6.35
Degree of freedom		7	6	8
p-value		<0.0001	<0.0001	0.6080

Bodhisuwan, W. 2011. Zero inflated Waring distribution and its application. Proceedings of the 37th Congress on Science and Technoloby of Thailand, Thailand.

Famoye, F. and Singh, K.P. 2003. On inflated generalized Poisson regression models. *Advances and Applications in Statistics*. 3, 145-158.

Flynn, M. 2009. More flexible GLMs zero-inflated models and hybrid models. *Casualty Actuarial Society E-Forum*, Winter, U.S.A., pp. 148-224.

Gomez-Déniz E., Sarabia, J.M. and Calderín-Ojeda, E. 2008. Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications. *Insurance Mathematics and Economics*, 42, 39-49.

Greene, W.H. 1994. Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Working paper No. 94-10, Department of Econometrics, Stern School of Business, New York University, New York, U.S.A.

Hall, D.B. 2000. Zero-inated Poisson and binomial regression with random effects: a case study. *Biometrics*. 56, 1030-1039.

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

Lindsey, J. K. 1995. *Modelling Frequency and Count Data*. Oxford science publications, Clarendon Press, UK., p. 146.

Pudprommarat, C., Bodhisuwan, W. and Zeephongsekul, P. 2012. A new mixed negative binomial distribution. *Journal of Applied Sciences*. 17, 1853-1858.

R Core Team, 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Wang, Z. 2011. One mixed negative binomial distribution with application. *Statistical Planning and Inference*. 141, 1153-1160.

Xie, F., Wei B. and Lin, J. 2009. Score tests for zero-inflated generalized Poisson mixed regression models. *Computational Statistics and Data Analysis*. 53, 3478-3489.

Zamani, H. and Ismail, N. 2010. Negative binomial-Lindley distribution and its application. *Journal of Mathematics and Statistics*. 1, 4-9.

Appendix

Proof of Theorem 1:

From Definition 1, let $Y \sim \text{NB-GE}(r, \alpha, \beta)$ be a random variable of the NB-GE distribution and some expected values of Y are

$$\begin{aligned} E(Y^2) &= (r^2 + r)\delta_{(2)} - (2r^2 + r)\delta_{(1)} + r^2, \\ E(Y^3) &= (r^3 + 3r^2 + 2r)\delta_{(3)} - (3r^3 + 6r^2 + 3r)\delta_{(2)} + (3r^3 + 3r^2 + r)\delta_{(1)} - r^3, \\ E(Y^4) &= (r^4 + 6r^3 + 11r^2 + 6r)\delta_{(4)} - (4r^4 + 18r^3 + 26r^2 + 12r)\delta_{(3)} \\ &\quad + (6r^4 + 18r^3 + 19r^2 + 7r)\delta_2 - (4r^4 + 6r^3 + 4r^2 + r)\delta_1 + r^4. \end{aligned}$$

If $X \sim \text{ZINB-GE}(\phi, r, \alpha, \beta)$, then the expected value of X is given by

$$E(X) = \sum_{x=0}^{\infty} x f(x) = (0) \left[\phi + (1-\phi)M_{(r)} \right] + (1-\phi) \sum_{x=1}^{\infty} x \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j M_{(r+j)}.$$

From $E(Y) = \sum_{y=0}^{\infty} y f(y) = \sum_{y=1}^{\infty} y \binom{r+y-1}{y} \sum_{j=0}^y \binom{y}{j} (-1)^j M_{(r+j)} = r(\delta_{(1)} - 1)$, we have

$$E(X) = (1-\phi)E(Y) = (1-\phi) \left[r(\delta_{(1)} - 1) \right].$$

Consequently, we obtained the expected values of X as follows

$$\begin{aligned} E(X^2) &= (1-\phi) \sum_{x=1}^{\infty} x^2 \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j M_{(r+j)} = (1-\phi)E(Y^2), \\ E(X^3) &= (1-\phi) \sum_{x=1}^{\infty} x^3 \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j M_{(r+j)} = (1-\phi)E(Y^3), \\ E(X^4) &= (1-\phi) \sum_{x=1}^{\infty} x^4 \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j M_{(r+j)} = (1-\phi)E(Y^4). \end{aligned}$$

From $\text{Var}(X) = E(X^2) - (E(X))^2$, the variance of X is

$$\text{Var}(X) = r \left[(r+1)\delta_{(2)} - (2r+1)\delta_{(1)} + r \right] (1-\phi) - \left[r(\delta_{(1)} - 1)(1-\phi) \right]^2.$$

The skewness and kurtosis of X are

$$\begin{aligned} Sk(X) &= [E(X^3) - 3E(X)E(X^2) + 2(E(X))^3] / \sigma_X^3, \\ Ku(X) &= [E(X^4) - 4E(X)E(X^3) + 6(E(X))^2 E(X^2) - 3(E(X))^4] / \sigma_X^4. \end{aligned}$$

Now, by using the expression as below, the skewness and kurtosis of X can be written in (7) and (8), respectively.

The R code of the parameter estimation of the ZINB-GE distribution with the MLE method:

```

mlogl<-function(theta,x){
  fzinbge<-function(theta,x){
    mm<-length(x);
    k<-numeric(mm);
    zinbge<-function(theta,x){
      if(x==0){
        p<-(-log(theta[4]+((1-theta[4])*((gamma(theta[2]+1)
          *gamma(1+theta[1]/theta[3]))/gamma(theta[2]+theta[1]
          /theta[3]+1))));}
      }else
      if(x>0){
        pp1<-(gamma(theta[2]+1)*(gamma(1+theta[1]/theta[3])))
          /(gamma(theta[2]+theta[1]/theta[3]+1));
        for(j in 1:x){
          p1<-((factorial(x)/(factorial(j)*factorial(x-j)))
            *(-1)^j*((gamma(theta[2]+1)*gamma(1+(theta[1]+j)
              /theta[3]))/gamma(theta[2]+(theta[1]+j)/theta[3]+1));
          pp1<-pp1+p1;
        }
        p<-(-log(1-theta[4]))-log(factorial(theta[1]+x-1))
          +log(factorial(theta[1]-1))+log(factorial(x))-log(pp1);
      }
    }p}
  for(i in 1 : length(x)){
    k[i]<-zinbge(theta,x[i])}k}
  sum(fzinbge(theta,x))}

theta.start<-c(start_r,start_alpha,start_beta,start_phi)
out<-nlm(mlogl,theta.start,x = x)
r_MLE<-out$estimate[1]
alpha_MLE<-out$estimate[2]
beta_MLE<-out$estimate[3]
phi_MLE<-out$estimate[4]

```