International Journal of Applied Mathematics and Statistics,
Int. J. Appl. Math. Stat.; Vol. 56; Issue No. 5; Year 2017,
ISSN 0973-1377 (Print), ISSN 0973-7545 (Online)

Comparison of Bivariate Negative Binomial Regression Models for Handling Over dispersion

[1]Y.S.Dewi, [2] Purhadi,[3]Sutikno and **[4]S.W. Purnami**

[1]Department of Mathematics, Faculty of MIPA, University of Jember,
Jl Kalimantan 37 Jember, 68121 Indonesia
Email : [1]yulidewi.fmipa@unej.ac.id

[2,3,4]Department of Statistics, Faculty of MIPA, Sepuluh Nopember Institute of Technology,
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
Email : [2]purhadi@statistika.its.ac.id, [3]sutikno@statistika.its.ac.id , [4]santi_wp@statistika.its.ac.id

ABSTRACT

Some methods have been proposed for dealing with extra Poisson variation when conducting regression analysis of count data. One of them is negative binomial regression model. For bivariate cases, there are some methods for constructing bivariate negative binomial distributions. Two of them are bivariate negative binomial distribution as a mixture Poisson gamma and a result of multiplication of negative binomial marginals by a multiplicative factor. In this paper we will review the bivariate negative binomial regression models based on those distributions by using maximum likelihood estimation (MLE) method, including the parameters estimation and hypothesis testing. We use health care datasets as the application. The bivariate negative binomial models tend to give better performance than the bivariate Poisson models for analyzing the data with over-dispersion. In this work, a model that comes from a result of multiplication of negative binomial marginals by a multiplicative factor has best performance in modeling the health care data.

Keywords–Bivariate negative binomial models, MLE method, estimation, hypothesis testing

Mathematics Subject Classification: 62-02, 62J02

Computing Classification System: G.1.6

1. INTRODUCTION

The existence of over-dispersion can cause standard error underestimated on Poisson regression because there is an equality of the mean and variance assumption on it, so that it will give wrong inferences (Hinde & Dem'etrio 1998). Besides, count data with over-dispersion is common in many fields. Then, those led to development of statistical methods to analyze such data.

Some research themes have been conducted for handling over-dispersion. Dean & Lawless (1989) developed two test statistics to detect the existence of over-dispersion on Poisson regression model. Consul & Famoye (1992) proposed the method to model Poisson data with over-dispersion/under-dispersion, *Generalized Poisson Regression* (GPR), using MLE and moment methods to estimate the parameters. By using salmonella data and Maximum Likelihood Estimation (MLE), based on deviance indicator, Lawless (1987) showed that GPR has the same performance as *Negative Binomial Regression* (NBR) and it is better than Poisson Regressionfor data with over-dispersion. Lord & Park (2008) presented the parameterization of Negative Binomial 1 and 2. Ismail &Zamani (2013) modeled count data with over-dispersion using some methods based on MLE and compared them using AIC and BIC.The themes of the research which related to over-dispersion which is caused by excess zero are discussed by Chananet, *et.al* (2015)andSapuan, *et.al* (2016).

In analysis of univariate count data with over-dispersion, univariate negative binomial from the mixture Poisson Gamma distribution has been commonly used; see Lawless (1987), Gurmu (1991), Ismail &Jemain (2007), Cameron & Trivedi (2013). Furthermore,for bivariate negative binomial,there are some ways to construct the distributions. One of them is bivariatePoisson gamma mixture distribution (BPGMD); see Marshall &Olkin (1990), Munkin& Trivedi (1999), Gurmu& Elder (2000) andWang (2003). Another way isa result of multiplication of negative binomial marginals with a multiplicative factor (BNBD), see Famoye (2010). In this paper, we make comparison of two models ofbivariate negative binomial regression. The first oneis a model derived by mixing Poisson gamma distribution (BPGMR).The second one isa model derived by a result ofmultiplication of negative binomial marginalswith a multiplicative factor distribution (BNBR).

2. GENERATION OF THE DATA

We use the health care data from Deb & Trivedi (1997) in this research. The data consists of 4406 observations and sixteen covariates. The BPGMR and BNBR models are defined by assuming that two response variables have the same covariates. We use MLE method via Nelder Mead Algorithm and generate initial value using runif (-3,3) function for estimating parameters and testing hypotheses.For measuring how well the models fit the data, some goodness of fits of the models are described.

2.1. Bivariate Poisson Gamma Mixture Regression (BPGMR)

Consider random variablesY_{i1} and Y_{i2} (i = 1,...,n) which follow Poissondistributions with means $\lambda\mu_{i1}$ and $\lambda\mu_{i2}$ respectively, where λ is random variable following gamma $\left(\tau^{-1},\tau^{-1}\right)$ distribution, then the random variables Y_{i1} and Y_{i2} follow bivariate negative binomial or Poisson gamma mixture distribution, $\left(y_{i1},y_{i2}\right)\sim \mathrm{BPGM}\left(\mu_{i1},\mu_{i2},\tau\right)$, with joint probability function as follows :

$$f\left(y_{1i},y_{2i}\right)=\frac{\Gamma\left(y_{i1}+y_{i2}+\tau^{-1}\right)}{\Gamma\left(\tau^{-1}\right)\Gamma\left(y_{i1}+1\right)\Gamma\left(y_{i2}+1\right)}\mu_{i1}^{\ y_{i1}}\mu_{i2}^{\ y_{i2}}\tau^{-\tau^{-1}}\left(\mu_{i1}+\mu_{i2}+\tau^{-1}\right)^{-\left(y_{i2}+y_{i1}+\tau^{-1}\right)} \tag{1}$$

where $\tau\geq 0$ is dispersion parameter which is assumed not depending on the covariate. The mean and variance of Y_{i1} and $Y_{i2}$are

$$\begin{aligned} &E\left(y_{i1}\right)=\mu_{i1}=\exp\left(\mathbf{x}_{i1}^{T}\boldsymbol{\beta}_1\right),\quad E\left(y_{i2}\right)=\mu_{i2}=\exp\left(\mathbf{x}_{i2}^{T}\boldsymbol{\beta}_2\right)\\ &\operatorname{var}\left(y_{ij}\right)=\mu_{ij}+\mu_{ij}^{2}\tau,\ \ (j{=}1{,}2)\end{aligned} \tag{2}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are $p_1 \times 1$ and $p_2 \times 1$ vectors of unknown parameters and $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$ are $p_1 \times 1$ and $p_2 \times 1$ vectors of covariates respectively. Moreover, $\operatorname{cov}(y_{i1}, y_{i2}) = \tau\mu_{i1}\mu_{i2}$. The correlation coefficient formula, in term of μ_{ij} can be written as:

$$\operatorname{Corr}(y_{i1}, y_{i2}) = \sqrt{\frac{\mu_{i1}\mu_{i2}\tau^2}{(1+\mu_{i1}\tau)(1+\mu_{i2}\tau)}} \quad (3)$$

The bivariate distribution in equation(1)allows for positivecorrelation between the two random variables (equation (3)). If $\tau$is close to zero,then the two variables are independent and the bivariate Poisson gamma mixture distribution reduced into independent Poissondistribution. In this research, the bivariate negative binomial model is defined by assuming that $\log \mu_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}_j$,where $\mathbf{x}_{i1}^T = \mathbf{x}_{i2}^T = \mathbf{x}_i^T = (x_0, x_1, x_2, \ldots, x_\kappa)$.

Consider the vectors (y_{i1},y_{i2}),i = 1,2,…,$n$are independent where the i^{th} vector has the BPGMR model in equation (1). The log-likelihood function for BPGMR modelis given by

$$l(\boldsymbol{\theta}) = \log L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \tau)$$
$$= \sum_{i=1}^{n} \left\{ \begin{array}{l} \log\left(\dfrac{\Gamma(y_{i1} + y_{i2} + \tau^{-1})}{\Gamma(\tau^{-1})}\right) - \log y_{i1}! - \log y_{i2}! + \tau^{-1} \log \tau^{-1} \\ -\tau^{-1} \log\left(\tau^{-1} + \mu_{i1} + \mu_{i2}\right) + y_{i1} \log\left(\dfrac{\mu_{i1}}{\tau^{-1} + \mu_{i1} + \mu_{i2}}\right) \\ + y_{i2} \log\left(\dfrac{\mu_{i2}}{\tau^{-1} + \mu_{i1} + \mu_{i2}}\right) \end{array} \right\} \quad (4)$$

For a more common parameterization uses $\tau^{-1} = \delta$, and δ is referred to as an index parameter, then the equation (4) can be written as follows

$$l(\boldsymbol{\theta}) = \log L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \delta)$$
$$= \sum_{i=1}^{n} \left\{ \begin{array}{l} \log\left(\dfrac{\Gamma(y_{i1} + y_{i2} + \delta)}{\Gamma(\delta)}\right) - \log y_{i1}! - \log y_{i2}! + \delta \log \delta \\ -\delta \log\left(\delta + \mu_{i1} + \mu_{i2}\right) + y_{i1} \log\left(\dfrac{\mu_{i1}}{\delta + \mu_{i1} + \mu_{i2}}\right) \\ + y_{i2} \log\left(\dfrac{\mu_{i2}}{\delta + \mu_{i1} + \mu_{i2}}\right) \end{array} \right\} \quad (5)$$

The log likelihood in equation (5) is maximized over the parameters β_{jk} for j=1,2; k = 0,1,…,κ and τ. By using the parameter estimates and their standard errors, the asymptotic Wald statistics can be obtained to test the significance of each independent variable and over-dispersion parameter.

2.2.A Result Of Multiplication Of Negative Binomial Marginalsby A Multiplicative Factor Regression (BNBR)

By the same way as Lakshminarayana, *et.al.* (1999), Famoye (2010) proposed an alternative model for bivariate negative binomial regression model with more flexible correlation structure. The correlation between Y_{i1} and Y_{i2} could be positive or negative. He defined a bivariate negative binomial distribution as a result of multiplication of negative binomial marginals by a multiplicative factor (BNBD). The distribution function of BNBD is given as

$$f\left(y_{i1},y_{i2}\right)=\prod_{j=1}^{2}\frac{\Gamma\left(\tau_j^{-1}+y_{ij}\right)}{\Gamma\left(\tau_j^{-1}\right)\Gamma\left(y_{ij}+1\right)}\left(\frac{\mu_{ij}}{\tau_j^{-1}+\mu_{ij}}\right)^{y_{ij}}\left(\frac{\tau_j^{-1}}{\tau_j^{-1}+\mu_{ij}}\right)^{\tau_j^{-1}} \times\left[1+\lambda\left(e^{-y_{i1}}-c_1\right)\left(e^{-y_{i2}}-c_2\right)\right] \qquad (6)$$

where

$c_j=\left[1-\mu_{ij}\left(\tau_j^{-1}+\mu_{ij}\right)^{-1}\right]/\left[1-e^{-1}\mu_{ij}\left(\tau_j^{-1}+\mu_{ij}\right)^{-1}\right]^{\tau_j^{-1}};(i=1,2,...,n,\ j=1,2)$ and y_{i1}, y_{i2} = 0,1,2,….

The marginal distributions of Y_{ij} (i = 1,2,…,n, j = 1,2) are negative binomial with means and variances $E\left(Y_{ij}\right)=\mu_{ij}$ and $\mathrm{var}\left(Y_{ij}\right)=\mu_{ij}\left(1+\tau_j^{-1}\mu_{ij}\right)$ respectively. The correlation between the response variables Y_{i1} and Y_{i2} is given as

$$\mathrm{Corr}\left(Y_{i1},Y_{i2}\right)=\lambda d^2\frac{\sqrt{\left[\mu_{i1}\mu_{i2}\left(1+\tau_1\mu_{i1}\right)\left(1+\tau_2\mu_{i2}\right)\right]}}{\left[1+d\tau_1\mu_{i1}\right]^{1+1/\tau_1}\left[1+d\tau_2\mu_{i2}\right]^{1+1/\tau_2}} \qquad (7)$$
$$i=1,2,...,n$$

where $d=1-e^{-1}$. The dispersion and independence parameters of Y_{ij} are τ_j and λ respectively. The correlation between Y_{i1} and Y_{i2} variables will be positive whenλ >0, negative when λ <0and independent when λ = 0. The existence of over-dispersion is indicated by parameter τ_j.If τ_j>0,there is over-dispersion.The BNBR reduces to BPR model when $\tau_j \rightarrow 0$ and there is no dispersion.

Let Y_{i1} and Y_{i2} be random variables that follow bivariate negative binomial distribution and X_0,X_1,X_2,…, X_κ be covariates.The BNBR model can be written as

$$\left(Y_{i1},Y_{i2}\right)\sim BNB\left(\mu_{i1},\mu_{i2},\tau_1,\tau_2,\lambda\right)$$
$$\mu_{ij}=\exp\left(\mathbf{x}_i^T\boldsymbol{\beta}_j\right);i=1,2,..,n;\ j=1,2. \qquad (8)$$

where $\mathbf{x}_i=\left[1\quad x_{i1}\quad x_{i2}\quad ...\quad x_{i\kappa}\right]^T$ and $\boldsymbol{\beta}_j=\left[\beta_{j0}\quad \beta_{j1}\quad \beta_{j2}\quad ...\quad \beta_{j\kappa}\right]^T$.

The estimation of parameters is conducted by using MLE method. Before we present the likelihood function of BNBR model let’s see the gamma function below.

$$\frac{\Gamma\left(\tau_j^{-1}+y_{ij}\right)}{\Gamma\left(\tau_j^{-1}\right)}=\prod_{l=0}^{y_{ij}-1}\left(\tau_j^{-1}+l\right)\text{ and }\Gamma\left(y_{ij}+1\right)=y_{ij}! \qquad (9)$$

By using the equations (6) and (9), the likelihood function is written as

$$L\left(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\tau_1,\tau_2\right)=\prod_{i=1}^{n}\left\{\prod_{j=1}^{2}\prod_{l=0}^{y_j-1}\frac{\left(\tau_j^{-1}+l\right)}{y_{ij}!}\left(\mu_{ij}\right)^{y_{ij}}\left(\tau_j^{-1}\right)^{\tau_j^{-1}}\left(\tau_j^{-1}+\mu_j\right)^{-\left(y_{ij}+\tau_j^{-1}\right)} \times\left[1+\lambda\left(e^{-y_{i1}}-c_1\right)\left(e^{-y_{i2}}-c_2\right)\right]\right\} \qquad (10)$$

where $\mu_{ij} = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}_j\right); c_j = \left(1 + d\tau_j \mu_{ij}\right)^{-1/\tau_j}$ and $d = 1 - e^{-1}$.

It is the same as BPGMR, we can use the asymptotic Wald statistics to test the significance of each independent variable and over-dispersion parameters.

2.3. Goodness of Fit of the Model

A goodness of fit refers to measuring how well do the observed data corresponding to the fitted model. Some goodness of fit measures connected to BPGMR and BNBR are presented as below

2.3.1. Deviance

The distance between the saturated and fitted models is measured by Deviance (D). The Deviancefor BPGMR model can be written as follows:

$$\begin{aligned} D &= 2\left[\log L(y;y) - \log L(\mu;y)\right] \\ &= 2\sum_{i=1}^{n}\left\{\delta \log \frac{\delta + \widehat{\mu}_{i1} + \widehat{\mu}_{i2}}{\delta + y_{i1} + y_{i2}} + \sum_{j=1}^{2} y_{ij} \log \frac{y_{ij}\left(\delta + \widehat{\mu}_{i1} + \widehat{\mu}_{i2}\right)}{\widehat{\mu}_{ij}\left(\delta + y_{i1} + y_{i2}\right)}\right\} \end{aligned} \quad (11)$$

When the model provides a good fit, then $\log L(\mu;y)$ is expected to be close to $\log L(y;y)$. A small value of the devianceindicates a good model. The Deviance for BNBR is equal to

$$\begin{aligned} D &= 2\left[\log L(y;y) - \log L(\mu;y)\right] \\ &= 2\sum_{i=1}^{n}\left\{\begin{array}{l} \sum_{j=1}^{2}\left[y_{ij}\log\left(\dfrac{y_{ij}\left(\tau_j^{-1} + \widehat{\mu}_{ij}\right)}{\widehat{\mu}_{ij}\left(\tau_j^{-1} + y_{ij}\right)}\right) + \tau_j^{-1}\log\left(\dfrac{\tau_j^{-1} + \widehat{\mu}_{ij}}{\tau_j^{-1} + y_{ij}}\right)\right] \\ + \log\left(\dfrac{1 + \lambda\left(e^{-y_{i1}} - \tilde{c}_1\right)\left(e^{-y_{i2}} - \tilde{c}_2\right)}{1 + \lambda\left(e^{-y_{i1}} - \widehat{c}_1\right)\left(e^{-y_{i2}} - \widehat{c}_2\right)}\right) \end{array}\right\} \end{aligned} \quad (12)$$

where $\tilde{c}_j$ is the value of $c_j$evaluated at $\mu_{ij}= y_{ij}$and $\widehat{c}_j$ is the value of $c_j$evaluated at $\mu_{ij} = \widehat{\mu}_{ij}$. The best model usually has the smallest value of the deviance *D*, among all models. For an adequate model, *D* has an asymptotic chi square distribution with *n-p* degree of freedom, where *p* isthe total number of estimated parametersand *n* is the sample size. Therefore, if the ratio *D*/*df*valueis veryclose to one, the model may be considered as adequate model.

If dispersion parameter in equation (1)closes to zero, BPGMR reduces to independent bivariate Poisson regression model (IBPR). SimilarlyBNBR reduces to BPR modelif the two dispersion parameters close to zero in equation (6). Those are usually called nested models, one of models is simplified model of the other.

2.3.2. Likelihood Ratio Test

One of statistical tests that is used to make comparison of the goodness of fit of two nested models, one of which is a special case of the other is a likelihood ratio test. In this section we will compare

BPGMR vs IBPR and BNBR vs BPRmodels to determine whether BPGMR and BNBR modelsare more suitable than IBPR and BPR models. If dispersion parameter is very close to zero, the BPGMR model reduces to IBPR and the correlation of two response variables can be 0. The hypothesis can be written as

$$H_0: \ \tau=0 \quad (13)$$

From the equation(1) and(4), τ does not allow a negative value.Therefore, the alternative hypothesis in equation(13)must be one sided.The hypothesis testing can be carried out using likelihood ratio test :

$$LR = 2\left(l_1 - l_0\right) \quad (14)$$

where l_1 and l_0 are the log likelihood function when H_0 is false and true respectively.

Since the null hypothesis is on the boundary of parameter space, the asymptotical distribution of LR is as half of χ_0^2 and half of χ_1^2, Self & Liang (1987).

For the BNBR model, the test for dispersion parameter is given as

$$H_0: \ \tau_1=\tau_2=0 \quad (15)$$

If $\tau_j \rightarrow 0$,a result of multiplication of negative binomial marginals by a multiplicative factor reduces to a result of multiplication of Poisson marginals with a multiplicative factor and there is no over-dispersion in the data. When $\tau_j>0$, there will be over-dispersion. Let l_0 be the loglikelihood function when H_0 is true and let l_1 be the loglikelihood function when H_0 is false. The test statistic of the hypothesis (15) is written as

$$LR_0 = 2\left(l_1 - l_0\right) \quad (16)$$

The statistic LR_0 is asymptotically distributed as a random variable which has a probability mass of 0.25 at the point 0, a $0.5\chi_1^2$ and $0.25\chi_2^2$ distribution above 0, (Famoye, 2010).

The statistical hypothesis for independence of Y_{i1} and Y_{i2} isgiven as

$$H_0: \ \lambda=0 \quad (17)$$

Let l_0 be the loglikelihood function when H_0 is true and let l_1 be the loglikelihood function when H_0 is false for hypothesis (17). The likelihood ratio test statistic of the hypothesis (17) is

$$LR_\lambda = 2\left(l_1 - l_0\right) \quad (18)$$

The approximate distribution of statistic LR_λ is chi square with one degree of freedom.

2.3.3. Pseudo R-Squared

The approximation of coefficient of determination for regression models with count dependent variables is pseudo R-squared. One of pseudo R-squared formula is written below :

$$R^2 = \frac{l_{fit} - l_0}{l_{max} - l_0} = 1 - \frac{l_{max} - l_{fit}}{l_{max} - l_0} \qquad (19)$$

where l_{fit} and l_{max} denote the log likelihood in the fitted and saturated models and l_0 denotes the log likelihood in intercept-only model, (Cameron & Trivedi, 2013).

2.3.4. AIC and BIC

The Akaike information criterion (AIC) is the most commonly used as general fit statistic.AIC has formula as follows:

$$\text{AIC} = -2\,l + 2v, \qquad (20)$$

where $l$is the maximized value of the loglikelihood and $v$is the number of estimated parameters. A model with smallest AIC is generally preferred. Another AIC statistic which is used for small sample size is finite sample AIC, usually mentioned as AIC-correctedwhich may be defined as

$$\text{AIC}_c = \text{AIC} + \frac{2v(v+1)}{n\text{-}v\text{-}1} \qquad (21)$$

where v is the number of parameters in the modeland $n$is the number of observations. Note that $\text{AIC} \approx \text{AIC}_c$ for models with large number of observations.

The second measure related to the AIC is Bayesian information criterion (BIC).It is formulated as

$$\text{BIC} = -2\,l + v\log(n), \qquad (22)$$

where l denotes the maximized value of the model log likelihood, $v$is the number of estimated parameters and $n$is the number of observations. The AIC gives a smalleradjustment weight , $2v$ than does the BIC statistic, where $v\log(n)$, is employed for adjusting $-2l$.

3. RESULTS

In this paper, we fit and make comparison of bivariate negative binomial regression models on the data concerning health care data taken from Deb & Trivedi (1997). Deb & Trivedi (1997) analyzed the number of patient visits by using univariate hurdle and finite mixtures negative binomial models. They didn't consider correlation between response variables. The two response variables in this analysis are the number of hospital stays (HOSP) and non physician hospital outpatient visits (OPNP). There are sixteen covariatesfrom health measures in the data include regional, demographic,self-perceived measures of health, the number of chronic diseases and conditions , a measure of disability status , and economic variables. Additional information for the covariates is provided in Deb and Trivedi (1997).

The Mean of HOSP and OPNP are 0.296 and 0.536 , the variance of those are0.557 and 15.054.The response variables are correlated (Pearson correlation is equal to 0.065) indicating that we should use bivariate negative binomial regression model.We suggest the following model to fit the expected number of HOSP and OPNP:

(i) $y_{ij} \overset{\text{ind}}{\sim} \mathrm{BPGM}\left(\mu_{ij},\tau^{-1}\right)$ or $y_{ij} \overset{\text{ind}}{\sim} \mathrm{BNB}\left(\mu_{ij},\tau_j,\lambda\right)$

(ii) $\log\left(\mu_{ij}\right)=\beta_{j0}+\beta_{j1}x_1+...+\beta_{j16}x_{16}$

where $y_{ij}$is the number of defects of the *i*thsample in the*j*th type of visits, for *i*= 1,2,…,4406 and *j* = 1 for HOSP and *j*=2 for OPNP, with $x_1,\ldots,x_{16}$denoting the covariates. The parameter estimates and approximate standard errors of bivariatenegative binomial models are given in Table 1. By assuming there is not dispersion parameter, we get IBPR from BPGMR model and BPR from BNBR model. The asymptotic Wald statistic for testing the significance of over-dispersions shows that the dispersion parameters are significantly different from zero for BPGMR and BNBR models. The asymptotic Wald z statistics indicate that there are some covariates influencesignificantly the model by using significance level 5 %. It appears that the standard errors from BPGMR and BNBR parameter estimates tend to be larger than those from BPR and IBPR models. The value of AIC andDeviance of BPGMR and BNBRmodels tend to be smaller than those of BPR and IBPR models. For this health care data case BNBR is the best model. It has the smallest value of AIC and Deviance. The Deviance/df of BNBR is the closest to one than the other models as indication that it is an adequate model.

Table 1. Parameter Estimates and Hypotheses testingof BPGMR, BNBR, IBPR and BPR models.

Variable	BPGMR	BNBR	IBPR	BPR
(Y1/Hosp)				
Intercept	0.606 (0.169) *	-1.509 (0.495) *	-3.161 (0.626) *	-2.611 (0.363) *
Exclhlth	-5.098 (0.591) *	-0.657 (0.191) *	-0.710 (0.176) *	-0.756 (0.172) *
Poorhlth	-0.549 (0.214) *	0.521 (0.101) *	0.536 (0.070) *	0.477 (0.067) *
Numchron	0.664 (0.124) *	0.256 (0.027) *	0.251 (0.019) *	0.250 (0.018) *
Adldiff	0.328 (0.034) *	0.417 (0.093) *	0.341 (0.072) *	0.437 (0.065) *
Noreast	0.413 (0.114) *	-0.114 (0.105)	-0.008 (0.082)	-0.047 (0.079)
Midwest	0.227 (0.125)	0.061 (0.094)	0.113 (0.073)	0.135 (0.069)
West	0.287 (0.112) *	0.064 (0.105)	0.100 (0.081)	0.039 (0.079)
Age	0.351 (0.129) *	-0.051 (0.061)	0.128 (0.075)	0.059 (0.044)
Black	0.297 (0.073) *	0.126 (0.122)	0.098 (0.095)	0.375 (0.087) *
Male	0.413 (0.146) *	0.240 (0.082) *	0.154 (0.063) *	0.200 (0.061) *
Married	0.502 (0.100) *	-0.081 (0.086)	-0.024 (0.066)	-0.004 (0.063)
School	-0.169 (0.102)	-0.010 (0.011)	0.003 (0.009)	0.001 (0.008)
Faminc	-0.007 (0.012)	-0.004 (0.014)	0.007 (0.010)	0.003 (0.010)
Employed	-0.011 (0.017)	0.037 (0.130)	0.043 (0.109)	0.018 (0.104)
Privins	0.286 (0.156)	0.078 (0.106)	0.215 (0.080) *	0.303 (0.078) *
Medicaid	0.641 (0.123) *	0.056 (0.141)	0.182 (0.102)	0.208 (0.096) *
(Y2/Opnp)				
Intercept	1.856 (0.550) *	-1.152 (0.767)	2.587 (0.412) *	2.659 (0.305) *
Exclhlth	-0.775 (0.184) *	-0.622 (0.233) *	-0.941 (0.136) *	-0.915 (0.134) *
Poorhlth	-0.006 (0.121)	0.270 (0.174)	-0.170 (0.063) *	-0.160 (0.062) *
Numchron	0.188 (0.031) *	0.195 (0.045) *	0.165 (0.015) *	0.163 (0.015) *
Adldiff	0.873 (0.106) *	0.603 (0.144) *	0.681 (0.052) *	0.683 (0.051) *
Noreast	0.287 (0.111) *	0.263 (0.151)	-0.081 (0.060)	-0.076 (0.060)

Midwest	0.544 (0.099) *	0.456 (0.135) *	0.242 (0.051) *	0.249 (0.051) *
West	0.181 (0.121)	0.029 (0.164)	-0.346 (0.072) *	-0.331 (0.071) *
Age	-0.577 (0.070) *	-0.109 (0.098)	-0.594 (0.052) *	-0.600 (0.039) *
Black	1.281 (0.126) *	1.237 (0.178) *	1.114 (0.055) *	1.097 (0.054) *
Male	0.282 (0.089) *	-0.027 (0.121)	0.134 (0.046) *	0.134 (0.045) *
Married	-0.043 (0.092)	0.190 (0.125)	0.053 (0.049)	0.052 (0.048)
School	0.011 (0.011)	-0.016 (0.016)	-0.005 (0.007)	-0.007 (0.006)
Faminc	-0.025 (0.016)	-0.006 (0.023)	0.000 (0.008)	0.000 (0.008)
Employed	-0.039 (0.140)	0.097 (0.187)	-0.280 (0.078) *	-0.272 (0.077) *
Privins	0.805 (0.113) *	0.591 (0.159) *	0.678 (0.065) *	0.683 (0.064) *
Medicaid	0.088 (0.162)	-0.108 (0.219)	-0.017 (0.084)	-0.005 (0.083)
τ	3.874 (0.169) *			
$\tau 1$		1.863 (0.174) *		
$\tau 2$		9.102 (0.464) *		
λ		1.412 (0.201) *		1.768 (0.152) *
Lnlikelihood	-6622.52	-5944.02	-10050.83	-9992.97
AIC	13315.04	11962.03	20169.65	20055.95
Deviance	6189.63	4388.41	16382.35	16280.68
Deviance/df	1.416	1.004	3.747	3.725

Notes : * Significant at 5 % level, standard errors in parentheses.

4. DISCUSSION AND CONCLUSION

The bivariate model that is discussed in this paper takes over-dispersion and correlation into consideration. It can be applied to bivariate response variables cases in which the correlation coefficient is positive for BPGMR and positive/negative for BNBR model. The BPGMR model reduces to IBPR model when $\tau \to 0$ and the BNBR model reduces to BPR model when $\tau_j \to 0$, j = 1,2 . Those are the situations in which there is no over-dispersion in the data.

For the case of health care data, the standard errors of BPGMR and BNBR models tend to be larger than those from IBPR and BPR models. It indicates that IBPR and BPR models tend to underestimate the standard errors.

Based on the goodness of fit of the models, AIC, Deviance and Deviance/df, BNBR model is the best and most adequate model for estimating the parameters of the health care data. The BPGMR and BNBR models tend to perform better than IBPR and BPR models.

5. REFERENCES

Cameron, A.C. &Trivedi, P.K., 2013,*Regression Analysis of CountData*, 2nd edition, Cambridge University Press, United Kingdom.

Chananet,C., Areepong,P. &Sukparungsee, S. , 2015, A Markov Chain Approach for Average Run Length of EWMA and CUSUM Control Chart Based on ZINB Model. *International Journal of Applied Mathematics and Statistics*. Vol. **53,** Issue number 1.

Consul, P. C.&Famoye, F.,1992, GeneralizedPoisson Regression Model,*Communications in Statistics*: Theory and Method, **21** : 89 – 109

Dean, C & Lawless, J.F., 1989, Tests for Detecting Over-dispersion in Poisson Regression Models,*Journal of The American Statistical Association*. Vol **84**, Issue 406,pp.467 – 472.

Deb P &Trivedi,P.K., 1997, Demand For Medical Care by The Elderly : A finite Mixture Approach, J Appl,Econom**12**:313–336

Famoye, F.,2010,On The Bivariate Negative Binomial Regression Model,*Journal of Applied Statistics* Vol. **37,** No. 6 : 969–981.

Gurmu, S.,1991, Test for Detecting Over dispersion in The Positive Poisson Regression Model,*Journal of Business & Economic Statistics*, April, Vol. **9**, No. 2.

Gurmu, S. & Elder, J., 2000, Generalized Bivariate Count Data Regression Models,*Econometrics Letters*, **68**: 31 – 36.

Hinde, J. &Dem'etrio, C.G.B, 1998, Over dispersion: Models and Estimation,*ComputationalStatistics and Data Analysis* **27** : 151-170.

Ismail, N &Jemain, A.A., 2007, Handling Over dispersion with Negative Binomial and Generalized Poisson Regression Models. Casualty Actuarial Society Forum, Winter 2007.

Ismail, N &Zamani, H.,2013, Estimation of Claim Count Data using Negative Binomial, Generalized Poisson, Zero-Inflated Negative Binomial and Zero Inflated Generalized Poisson Regression Models : *Casualty Actuarial Society E- Forum, Spring 2013*.

Lakshminarayana,J., Pandit, S.N.N. &Rao, K.S,1999,On a bivariate Poisson distribution, Commun. Statist. Theory Methods **28**(2), pp. 267–276.

Lawless, J.F.,1987, Negative Binomial and Mixed Poisson Regression, The Canadian Journal of Statistics, Vol **15**, No. 3, 209 – 225.

Lord, D & B.J. Park, 2008, Negative Binomial Regression Models and Estimation Methods, Texas University.

Marshall, A. W., &Olkin, I., 1990, Multivariate distributions generated from mixtures of convolution and multiplication families,*IMS Lecture Notes-Monograph Series*, Vol. **16**, Topics in Statistical Dependence, pp.371-393.

Munkin, M.K. & Trivedi, P.K., 1999, Simulated Maximum Likelihood Estimation of Multivariate Mixed-Poisson Regression Models, with Applications, *Econometrics Journal*, **2**, 29 – 48.

Sapuan, M. S., Razali, A.M., &Zamzuri, Z. H., 2016, Zero Inflated Negative Binomial Model to Overcome Excess Zeros Count in Motorcycles Road Accident, *International Journal of Applied Mathematics and Statistics*, Vol. **55**, Issue number 2.

Self, S.G. & Liang, K., 1987, Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions, *Journal of the American Statistical Association*, Vol. **82**, No. 398, pp. 605 – 610.

Wang, P., 2003, A Bivariate Zero Inflated Negative Binomial Regression Model for Count Data with Excess Zeros, Economics Letter, **78** :373 – 378