Pertanika J. Sci. & Technol. 25 (2): 537 - 544 (2017)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Some New Bivariate Regression Models

Faroughi, P.[1]* and Ismail, N.[2]

[1]*Department of Statistics, Sanandaj Branch, Islamic Azad University, Sanandaj, Iran*

[2]*School of Mathematical Sciences, Faculty of Science and Technology Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*

ABSTRACT

This paper introduces new forms of bivariate generalized Poisson (BGP) and bivariate negative binomial (BNB) regression models which can be fitted to bivariate and correlated count data with covariates. The BGP and BNB regression models can be fitted to bivariate count data with positive, zero or negative correlations. Applications of new BGP and BNB regression models are illustrated on Australian health survey data.

Keywords: Generalized Poisson, negative binomial, bivariate, correlation, overdispersion

INTRODUCTION

Poisson regression is used for modelling count data with covariates. One of the issues with count data is over dispersed. Also, negative binomial (NB) regression can be used to manage over dispersion whilst generalized Poisson (GP) regression may be fitted for under dispersed or over dispersed count data. GP is achieved by the limiting form of a generalized NB distribution (Consul & Jain, 1973). Focusing on previous studies, different forms of GP regressions have been offered for various parameterization of GP regression (Consul & Famoye, 1992; Wang & Famoye, 1997; Famoye, Wulu & Singh, 2004; Zamani & Ismail, 2012; Karimi, Faroughi & Rahim, 2015; Zamani, Faroughi & Ismail, 2015).

When we have bivariate count data, there are some forms of bivariate models which can be fitted to bivariate Poisson (BP) (Campbell, 1934). Kocherlakota and Kocherlakota (1992) using trivariate reduction method. Bivariate generalized Poisson (BGP) distribution which follows the trivariate reduction method was obtained in Famoye and Consul (1995).

BP distribution which applies for different levels of correct as presented by Lakshminarayana Lakshminarayana, Pandit & Rao (1999) where distance is obtained from the product of two Poisson marginals with a multiplicative factor parameter. This study was continued by Famoye (2010a) presented BGP distribution, Famoye (2010b) who introduced BNB regression, and Famoye (2012) who defined BGP regression.

Article history:
Received: 24 February 2016
Accepted: 30 September 2016

E-mail addresses:
pouya.faroughi@gmail.com (Faroughi, P.),
ni@ukm.edu.my (Ismail, N.)
* Corresponding author

ISSN: 0128-7680

MATERIALS AND METHODS

The result of the joint p.m.f. of BP distribution is flexible structure which obtained from the product of two Poisson marginals with a multiplicative factor parameter (Lakshminarayana et al., 1999)

$$\mathrm{P}(y_1, y_2) = e^{-\mu_1-\mu_2}\frac{\mu_1^{y_1}\mu_2^{y_2}}{y_1!\,y_2!}\{1+\alpha[(g_1(y_1)-\bar{g}_1)(g_2(y_2)-\bar{g}_2)]\}$$
$$y_1, y_2 = 0,1,2,..., \qquad \mu_1,\mu_2 > 0 \tag{1}$$

where $g_1(y_1)$ and $g_2(y_2)$ are bounded functions in y_1 and y_2. To confide non-negativity in the value of $\{.\}$ in (1),

$$g_t(y_t) = e^{-y_t}\ , \bar{g}_t = E[g_t(Y_t)] = \mathrm{E}(\mathrm{e}^{-Y_t}), t = 1,2. \tag{2}$$

Suppose Y_{i1} and Y_{i2} $(i = 1,2,...n)$ are count response variables. Following (1)-(2), the joint p.m.f. of BP regression model is

$$\mathrm{P}(y_{i1}, y_{i2}) = \left[\prod_{t=1}^{2} e^{-\mu_{it}}\frac{\mu_{it}^{y_{it}}}{y_{it}!}\right][1+\alpha\prod_{t=1}^{2}(e^{-y_{it}} - e^{-d\mu_{it}})] \tag{3}$$

where $d = 1-e^{-1}$ and α is the correlation parameter. The covariates can be included using log links, $\log(\mu_{it}) = \mathbf{x}_{it}^T\boldsymbol{\beta}_t$, where $\mathbf{x}_{it}$ are vectors of explanatory variables and $\boldsymbol{\beta}_t$ are vectors of regression parameters. The marginal means and variances are $E(Y_{it}) = Var(Y_{it}) = \mu_{it}$, $t = 1,2$, and the covariance is $Cov(Y_{i1}, Y_{i2}) = \alpha\mu_{i1}\mu_{i2}d^2e^{-d(\mu_{i1}+\mu_{i2})}$. When α is zero, it means random variables Y_{i1} and Y_{i2} are not dependent, which they are distributed as a marginal Poisson regression model. If $\alpha > 0$, there are positive and if $\alpha < 0$, there are negative correlations.

The p.m.f. of GP distribution (Consul & Famoye, 1992) is

$$\mathrm{P}(y) = \frac{\theta(\theta+vy)^{y-1}}{y!}e^{-\theta-vy} \tag{4}$$

Where $\theta > 0$, and v called dispersion parameter with $\max(-1, -\frac{\theta}{4}) < v < 1$. The mean is $E(Y) = \mu = \theta(1-v)^{-1}$ and variance is $Var(Y) = \theta(1-v)^{-3}$. The model changes to Poisson when $v = 0$, and manages under-dispersion or over-dispersion when $v < 0$ or $v > 0$. The p.g.f. of GP distribution is $\varphi_Y(u) = E(U^Y) = e^{\theta(t-1)}$ where $t = ue^{v(t-1)}$, and the m.g.f. is

$$M_Y(u) = E(e^{uY}) = e^{\theta(e^t-1)} \tag{5}$$

In (5), $e^t = e^{v(e^t-1)+u}$. If we put $u = -1$ in (5), we gain

$$E(e^{-Y}) = e^{\theta(s-1)} \tag{6}$$

where $\ln s - v(s-1) + 1 = 0$. By differentiating m.g.f. in (5) with respect to u and putting $u = -1$, we obtain

$$\frac{\partial}{\partial u} M_Y(u)\bigg|_{u=-1} = E(Ye^{-Y}) = \frac{\theta}{1-vs} e^{(\theta+v)(s-1)-1} \tag{7}$$

where $\ln s - v(s-1)+1=0$. GP-1 models gained when $\theta_i = (1-v)\mu_i$ in (4) generate

$$\mathrm{P}(y_i) = \frac{(1-v)\mu_i[(1-v)\mu_i + vy_i]^{y_i-1}}{y_i!} e^{-(1-v)\mu_i - vy_i} \tag{8}$$

By using the same method which offered by Lakshminarayana et al. (1999) for obtaining BP distribution, the joint p.m.f. of BGP regression will be equalled

$$\prod_{t=1}^{2} \frac{(1-v_t)\mu_{it}[(1-v_t)\mu_{it} + v_t y_{it}]^{y_{it}-1} e^{-(1-v_t)\mu_{it} - v_t y_{it}}}{y_{it}!}[1+\alpha\prod_{t=1}^{2}(e^{-y_{it}} - c_{it})] \tag{9}$$

From (6), $c_{it} = E(e^{-Y_{it}}) = e^{\mu_{it}(1-v_t)(s_t-1)}$, where $\ln s_t - v_t(s_t-1)+1=0$, $t=1,2$. The mean, variance and covariance for BGP regression are $E(Y_{it}) = \mu_{it}$, $Var(Y_{it}) = \mu_{it}(1-v_t)^{-2}, t=1,2$, and $Cov(Y_{i1},Y_{i2}) = \alpha(c_{i11} - c_{i1}\mu_{i1})(c_{i22} - c_{i2}\mu_{i2})$. From (7), $c_{itt} = E(Y_{it}e^{-Y_{it}}) = \frac{(1-v_t)\mu_{it}}{1-v_t s_t} e^{[(1-v_t)\mu_{it}+v_t](s_t-1)-1}$ where $\ln s_t - v_t(s_t-1)+1=0,\ t=1,2$. When, random variables Y_1 and Y_2 are independent, each is distributed as a marginal GP regression. When $\alpha>0$ and $\alpha<0$, we have positive and negative correlations respectively. BGP regression reduces to BP regression when $v_1=v_2=0$, and handles under- and over dispersion when $v_t<0$ and $v_t>0$, $t=1,2$, respectively.

This article introduces a BGP regression which is based on GP-1 regression exist in (Zamani & Ismail, 2012; Zamani, Faroughi & Ismail, 2016). GP-1 regression is produced by putting v with $\frac{a}{1+a}$ and θ_i with $\frac{\mu_i}{1+a}$ in (4). The mean is for GP-1 regression is $E(Y_i)=\mu_i$ and variance is $Var(Y_i)=\mu_i(1+a)^2$, where a is the dispersion parameter. The model decline to Poisson regression when $a=0$, and handles under- and over dispersion when $a<0$ and $a>0$ respectively. The p.g.f. of GP-1 distribution is $\varphi_{Y_i}(u) = E(U^{Y_i}) = e^{\frac{\mu_i}{1+a}(t-1)}$ where $t = ue^{\frac{a}{1+a}(t-1)}$, so that the m.g.f. is $M_{Y_i}(u) = E(e^{uY_i}) = e^{\frac{\mu_i}{1+a}(e^t-1)}$ where $e^t = e^{\frac{a}{1+a}(e^t-1)+u}$. Therefore, $E(e^{-Y_i}) = e^{\frac{\mu_i}{1+a}(s-1)}$ and $E(Y_i e^{-Y_i}) = \frac{\mu_i}{1-a(s-1)} e^{\frac{\mu_i+a}{1+a}(s-1)-1}$ where $\ln s - \frac{a}{1+a}(s-1)+1=0$.

The joint p.m.f. of BGP regression is

$$\mathrm{P}(y_{i1},y_{i2}) = \prod_{t=1}^{2} \frac{\mu_{it}(\mu_{it}+a_t y_{it})^{y_{it}-1}}{(1+a_t)^{y_{it}} y_{it}!} e^{-\frac{\mu_{it}+a_t y_{it}}{1+a_t}} [1+\alpha\prod_{t=1}^{2}(e^{-y_{it}} - c_{it})] \tag{10}$$

where $c_{it} = e^{\frac{\mu_{it}}{1+a_t}(s_t-1)}$ and $\ln s_t - \frac{a_t}{1+a_t}(s_t-1) + 1 = 0$, $t=1,2$. The mean, variance and covariance for BGP-1 regression are $E(Y_{it}) = \mu_{it}$, $Var(Y_{it}) = \mu_{it}(1+a_t)^2$, $t=1,2$, and $Cov(Y_{i1},Y_{i2}) = \alpha(c_{i11} - c_{i1}\mu_{i1})(c_{i22} - c_{i2}\mu_{i2})$, where $c_{itt} = \frac{\mu_{it}}{1-a_t(s_t-1)} e^{\frac{\mu_{it}+a_t}{1+a_t}(s_t-1)-1}$ and $\ln s_t - \frac{a_t}{1+a_t}(s_t-1) + 1 = 0, t=1,2$. When $\alpha = 0$, random variables Y_1 and Y_2 are independent, each is distributed as a marginal GP-1 regression. When $\alpha > 0$ and $\alpha < 0$, we have positive correlation and negative correlation. BGP-1 regression decline to BP regression when $a_1 = a_2 = 0$, and manages under dispersion and over dispersion when $a_t < 0$ and $a_t > 0$, $t = 1,2$, respectively.

The p.m.f. of NB regression is

$$\mathrm{P}(y_i) = \frac{\Gamma(y_i + v_i)}{y_i!\Gamma(v_i)} \left(\frac{v_i}{v_i + \mu_i} \right)^{v_i} \left(\frac{\mu_i}{v_i + \mu_i} \right)^{y_i} \quad , \quad y_i = 0,1,2,...,n \tag{11}$$

where $v_i^{-1} = a$ is the dispersion parameter. The mean and variance of NB regression are $E(Y_i) = \mu_i$ and $V(Y_i) = \mu_i(1 + v_i^{-1}\mu_i) = \mu_i(1 + a\mu_i)$. NB regression in (11) is also referred as NB-2 regression. NB-2 regression reduces to Poisson regression in the limit as $a \to 0$, and display over dispersion when $a > 0$.

If we replace $v_i = a^{-1}\mu_i$ in p.m.f. (4), NB-1 regression is obtained. The p.m.f. is (Cameron & Trivedi, 2013; Greene, 2008)

$$\mathrm{P}(y_i) = \frac{\Gamma(y_i + a^{-1}\mu_i)}{y_i!\Gamma(a^{-1}\mu_i)} \left(\frac{1}{a^{-1}+1} \right)^{y_i} \left(\frac{a^{-1}}{a^{-1}+1} \right)^{a^{-1}\mu_i} \quad , \quad y_i = 0,1,2,...,n \tag{12}$$

where a is the dispersion parameter. The mean and variance of NB-1 regression are $E(Y_i) = \mu_i$ and $V(Y_i) = \mu_i(1+a)$.

By use the same method offered by Lakshminarayana et al. (1999), BNB regression model can be derived from NB-1 marginals and a mulitiplicative variable parameter. The p.m.f. of BNB regression model is

$$\mathrm{P}(y_1, y_2) = \prod_{t=1}^{2} \frac{\Gamma(y_{it} + a_t^{-1}\mu_{it})}{y_i!\Gamma(a_t^{-1}\mu_{it})} \left(\frac{1}{a_t^{-1}+1} \right)^{y_{it}} \left(\frac{a_t^{-1}}{a_t^{-1}+1} \right)^{a_t^{-1}\mu_{it}} \left[1 + \alpha(e^{-y_{i1}} - c_{i1})(e^{-y_{i2}} - c_{i2}) \right] \tag{13}$$

where $c_{it} = \left[(1-\theta_{it}) / (1-\theta_{it}e^{-1}) \right]^{a_t^{-1}\mu_{it}}$, $\theta_{it} = 1/(a_t^{-1}+1)$, a is the dispersion parameter and α is a multiplicative factor (or correlation) parameter.

From p.m.f. (13), Y_{i1} and Y_{i2} are independent if $\alpha = 0$. When $\alpha < 0$, the correlation between the response variables is negative and when $\alpha > 0$, the correlation between Y_{i1} and

Y_{i2} is positive. If $a \to 0$, BNB regression reduces to BP regression in the limit and if $a > 0$, the variance exceeds the mean and BNB regression allows over dispersion. The correlation between Y_{i1} and Y_{i2} can be defined, and it is equal to

$$\prod_{t=1}^{2} \alpha(1-e^{-1})^2 \sqrt{\mu_{it}(1+a_t^{-1}\mu_{it}^2)} \left[1+(1-e^{-1})a_t^{-1}\mu_{it}^2\right]^{-1-a_t^{-1}\mu_{it}}$$

A two-sided likelihood ratio test (LRT) can be performed to test the dispersion (over- or under dispersion) in BP against BGP alternatives where the hypothesis is $H_0 : a_1 = a_2 = 0$. The LRT is $T = 2(\ln L_1 - \ln L_0)$, where $\ln L_1$ and $\ln L_0$ are the models' log likelihood under their respective hypothesis. Since BP model is nested within BGP model, the statistic is asymptotically distributed as a chi-square with two degrees of freedom.

Likelihood ratio test can be performed to test over dispersion in BP regression against BNB regression. where L_0 and L_1 are the likelihood functions when H_0 and H_1 are true respectively. Since BNB regressions reduce to BP regression in the limit when $a \to 0$, the null hypothesis is $H_0 : a_1 = a_2 = 0$. The LRT statistic is approximately distributed as a probability of 0.25 at point zero, a 0.5 of chi-square with one degree of freedom and a 0.25 of chi-square with two degrees of freedom (Famoye, 2010a).

Akaike Information Criteria (AIC) is defined as $AIC = 2\dim(\theta) - 2\ln(L)$, where $\dim(\theta)$ is the number of parameters and $\ln(L)$ is the log likelihood of the estimated model. The model with the smallest AIC is the best model.

FINDINGS AND DISCUSSION

The health survey Australian data (Cameron, Trivedi, Milne & Piggott, 1988) is used for fitting different types of distributions such as BP, BGP as well as BNB regression models. Cameron and Johansson (1997) applied These data for fitting some univariate models, bivariate generalized negative binomial (BGNB) regression model was defined by Gurmu and Elder (2000). Famoye (2012) defined BGP-2 regression model. The health survey data includes 5190 single-person households for 1977–1978 Australian Health Survey.

In this article, we focus on two possibly dependent and negatively correlated response variables called Y_1, the total number of prescribed medications consumed in two days ago (PRESCRIBED), and Y_2 is the number of non-prescribed medications used in the same period (NON-PRESCRIBED). The mean for prescribed medications is 0.863 and standard deviation for prescribed medications is 1.415, the mean for non-prescribed medications is 0.356 and standard deviation for non-prescribed medications is 0.712 and the correlation between response variables is -0.043. The negative correlation illustrates possible negative dependency between the two variables. Cameron & Trivedi (2013) founded more information on the explanatory variables.

Table 1 shows the estimates and standard errors for BP, BGP, BNB regression models which are fitted jointly to both data.

The LRT for testing BP against BGP regressions is 381.46 and the LRT for testing BP against BNB regressions is 379.74, which is show over dispersion in data sets.

The estimates of correlation parameter for all models are negative. The negative estimates of correlation parameter indicating negative dependence between the two response variables.

The t-ratio for correlation parameter under BP model is 6.66, the t-ratio for correlation parameter under BGP model is 6.94 as well the correlation parameter under BNB model is 6.94 indicating that the two response variables are significantly dependent. Hence, the response variables are better to be fitted under BGP and BNB regression. Based on the AIC the best model is BGP, although the difference between AIC for BGP and BNB is small. BGP and BNB are more suitable than BP.

Table 1
BP, BGP and BNB regression models

Parameter	**BP**		**BGP**		**BNB**	
	est.	**s.e.**	**est.**	**s.e.**	**est.**	**s.e.**
PRESCRIBED						
Intercept	-2.70	0.13	-2.66	0.15	-2.66	0.15
Sex	0.48	0.04	0.55	0.04	0.55	0.04
Age	2.41	0.62	2.33	0.71	2.27	0.71
Agesq	-0.64	0.64	-0.65	0.74	-0.56	0.74
Income	0.00	0.06	-0.01	0.06	0.00	0.06
Levyplus	0.29	0.05	0.27	0.06	0.27	0.06
Freepoor	-0.05	0.12	-0.09	0.14	-0.09	0.14
Freerepa	0.30	0.06	0.28	0.07	0.27	0.07
Illness	0.20	0.01	0.20	0.01	0.20	0.01
Actdays	0.03	0.01	0.03	0.01	0.03	0.00
Hscore	0.02	0.01	0.02	0.01	0.02	0.01
Chcond1	0.77	0.05	0.76	0.05	0.75	0.05
Chcond2	1.01	0.05	1.00	0.06	0.99	0.06
NOPRESCRIBED						
Intercept	-2.03	0.17	-1.95	0.19	-2.02	0.19
Sex	0.27	0.05	0.26	0.06	0.27	0.06
Age	2.86	0.95	2.83	1.05	3.08	1.05
Agesq	-3.90	1.07	-3.89	1.19	-4.19	1.19
Income	0.17	0.08	0.11	0.09	0.13	0.09
Levyplus	-0.03	0.06	-0.05	0.06	-0.03	0.06
Freepoor	0.00	0.12	-0.08	0.14	-0.04	0.14
Freerepa	-0.29	0.09	-0.29	0.10	-0.26	0.10
Illness	0.20	0.02	0.20	0.02	0.20	0.02
Actdays	0.01	0.01	-0.00	0.01	-0.00	0.01
Hscore	0.03	0.01	0.03	0.01	0.03	0.01
Chcond1	0.15	0.06	0.14	0.06	0.13	0.06
Chcond2	0.02	0.08	0.03	0.09	0.03	0.09
a_1, dispersion	-	-	0.18	0.02	0.39	0.04
a_2, dispersion	-	-	0.14	0.02	0.29	0.03
α, correlation	-0.89	0.13	-0.91	0.13	-0.91	0.13
Log likelihood	-9522.59		-9331.86		-9332.744	
AIC	19099.18		18721.72		18723.49	

CONCLUSIONS

This article has introduced new types of BGP and BNB regression models. Because new form of BGP and BNB regression models have transformative mean and variance relationship, therefore could be fitted to bivariate count data with different levels of correlations, and admit over dispersion of the response variables.

New forms of BGP and BNB regression models were fitted to the Australian health survey and shown to have a negative correlation. The best model is BGP regression model based on AIC, for BNB and BP regression models. The estimates of correlation parameter for all models are significantly negative.

REFERENCES

Cameron, A. C., Trivedi, P. K., Milne, F., & Piggott, J. (1988). A microeconometric model of the demand for health care and health insurance in Australia. *The Review of economic studies, 55*(1), 85-106.

Cameron, A. C., & Johansson, P. (1997). Count data regression using series expansions: with applications. *Journal of Applied Econometrics, 12*(3), 203-223.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53). United States of America, USA: Cambridge university press.

Campbell, J. T. (1934). The Poisson correlation action. *Proceedings of the Edinburg Mathematical Society, 4*, 18-26.

Consul, P. C., & Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics, 15*(4), 791-799.

Consul, P. C., & Famoye, F. (1992). Generalized Poisson regression model. *Communications in Statistics-Theory and Methods, 21*(1), 89-109.

Famoye, F., & Consul, P. C. (1995). Bivariate generalized Poisson distribution with some applications. *Metrika, 42*(1), 127-138.

Famoye, F., Wulu, J. T., & Singh, K. P. (2004). On the generalized Poisson regression model with an application to accident data. *Journal of Data Science, 2*(2004), 287-295.

Famoye, F. (2010). A new bivariate generalized Poisson distribution. *Statistica Neerlandica, 64*(1), 112-124.

Famoye, F. (2010). On the bivariate negative binomial regression model. *Journal of Applied Statistics, 37*(6), 969-981.

Famoye, F. (2012). Comparisons of some bivariate regression models. *Journal of Statistical Computation and Simulation, 82*(7), 937-949.

Greene, W. (2008). Functional forms for the negative binomial model for count data. *Economics Letters, 99*(3), 585-590.

Gurmu, S., & Elder, J. (2000). Generalized bivariate count data regression models. *Economics Letters, 68*(1), 31-36.

Karimi, A., Faroughi, P., & Rahim, K. A. (2015). Modeling and Forecasting of International Tourism Demand in ASEAN Countries. *American Journal of Applied Sciences, 12*(7), 479-486.

Kocherlakota, S., & Kocherlakota, K. (1992). Regression in the bivariate Poisson distribution. *Communication Statistics-Theory and Methods*, *30*(5), 815-825.

Lakshminarayana, J., Pandit, S. N. N., & Srinivasa Rao, K. (1999). On a bivariate Poisson distribution. *Communications in Statistics-Theory and Methods, 28*(2), 267-276.

Wang, W., & Famoye, F. (1997). Modeling household fertility decisions with generalized Poisson regression. *Journal of Population Economics, 10*(3), 273-283.

Zamani, H., & Ismail, N. (2012). Functional form for the generalized Poisson regression model. *Communications in Statistics-Theory and Methods, 41*(20), 3666-3675.

Zamani, H., Faroughi, P., & Ismail, N. (2015). Bivariate Poisson-Lindley distribution with application. *Journal of Mathematics and Statistics, 11*(1), 1-6.

Zamani, H., Faroughi, P., & Ismail, N. (2016). Bivariate generalized Poisson regression model: applications on health care data. *Empirical Economics, 51*(4), 1607-1621.