Jour. Ind. Soc. Ag. Statistics
55(2), 2002 : 220-231

Regression Analysis of Count Data

J.T. Wulu, Jr., K.P. Singh[1], F. Famoye[2], T.N. Thomas[3] and G. McGwin[4]
Department of Health and Human Services
Bureau of Primary Health Care, Bethesda, MD 20814, USA
(Received : October, 2001)

SUMMARY

Three nonlinear count models, Poisson Regression (PR), Negative Binomial Regression (NBR), and Generalized Poisson Regression (GPR) are used for assessing the effects of risk factors on agricultural injuries from farm injury data. A sample of 1,322 respondents who participated in the farm safety/injury baseline survey in nine rural counties in Alabama and Mississippi, aged 18 years and older are considered for analysis. The dispersion parameter estimates and their standard errors for GPR models were consistently smaller than that of NBR models. Estimated dispersion parameters in the NBR and GPR models were positive and significantly different from zero. Estimated goodness-of-fit measures showed that GPR models outperformed the NBR and PR models.

Key words: Count data, Poisson, Parameter estimation, Generalized linear models.

1. Introduction

In many epidemiological studies where relationship between exposure and an outcome is being studied, response or dependent variable is often quantified by a count generated process in which number of incidents is due to a rare or chance event. Often that rare or chance event obeys principle of randomness, thus providing basis for application of poisson models. However, principle for complete randomness, providing the poisson distribution may be an excellent idea, but it is not very practical for all situations. An assumption of poisson process is that counts in one time interval must be independent of counts in other

1 Department of Biostatistics, School of Public Health, University of North Texas Health Science Center, Fort Worth, TX 76109-22699, USA

2 Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859, USA

3 Department of Epidemiology, School of Public Health, Harvard University, MA 02115, USA

4 Center for Injury Sciences, University of Alabama at Birmingham, AL 35294, USA

time intervals. In populations where events are very rare, poisson distribution is highly right skewed and as mean of events rises, distribution increasingly resembles the normal. Approximation of empirical count data (which are assumed to be poisson) by normal distribution often fails to account for skewness in the data.

A number of studies including Breslow [2], Consul and Famoye [3], Cox [4], Efron [5], Hinde [7], Lawless [8], Manton *et al.* [10], McCullagh and Nelder [11], Singh and Famoye [13], and Stein and Juritz [14] have suggested various models to deal with over-dispersed or extra-poisson variation data. Approaches and models for analysing over-dispersed poisson data and poisson rates include generalized linear models given in McCullagh and Nelder [11] and by Nelder and Wedderburn [12]; asymmetric maximum likelihood methods proposed by Efron and Thisted [6]; methods using double-exponential families suggested by Efron [5]; and Bayesian over-dispersed models and quasi-likelihood methods recommended by Albert and Pepple [1] and Lu and Morris [9]. The PR model has been found very useful for analysis of count data in which discrete response variable follows poisson distribution, but in the event such a variable is observed to be over-, or under-dispersed, it is appropriate to analyze the data using Generalized Poisson Regression (GPR) models.

The additional parameter in the GPR model provides useful information on dispersion in response count variable. As a result, the GPR model shows statistical advantages over standard poisson regression, negative binomial regression (NBR), generalized negative binomial regression, and generalized linear models in the event of fitting count data that may be over-, under-, or equi-dispersed. The reader is referred to Wang and Famoye [15] for further details. Hence, major aim of this research is to examine hypothesized statistical advantages of GPR models over PR or NBR models, through parameter estimates comparison, as it is applied to farming injury data.

2. *Poisson Regression Model*

Suppose, Y is a discrete random variable having independent response values $y_1, y_2, \ldots, y_n$ which follow a poisson distribution. Then the poisson regression model of Y given x_i may be defined as follows

$$P(y_i;\beta)=\frac{[\mu\{c_i(x_i,\beta)\}]^{y_i}\exp[-\mu\{c_i(x_i,\beta)\}]}{y_i!} \tag{2.1}$$

where $i = 1, 2, \ldots, n$; $y_i = 0, 1, 2, \ldots$; $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ is a p-dimensional vector of explanatory variables, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a p-dimensional vector of unknown parameters, and c_i denotes some measure of exposure.

Within the framework of PR model, equality constraint is observed between conditional mean, $E(Y/x_i)$, and conditional variance, $V(Y/x_i)$, of dependent variable for each observation. In practical applications, this assumption is often untrue since variance can either be larger or smaller than mean. If variance is not equal to mean, estimates in PR models are still consistent but are inefficient, which leads to invalidation of inference based on estimated standard errors.

3. Negative Binomial Regression (NBR) Model

Let $Y_1, Y_2, \ldots, Y_n$ be a set of n independent random variables where Y_i follows a NBD with mean m_i and dispersion parameter k, denoted by $Y_i \sim NB(m_i, k)$. Then

$$P(Y_i = y_i) = \frac{\Gamma(k + y_i)}{y_i!\,\Gamma(k)} \left(\frac{k}{m_i + k}\right)^k \left(\frac{m_i}{m_i + k}\right)^{y_i} \tag{3.1}$$

where $y_i = 0, 1, 2, \ldots$ and $m_i, k > 0$. The probability function (3.1) is as follows

$$P(Y_i = y_i) = \frac{(k + y_i - 1)!\,k^k m_i^{y_i}}{y_i!(k - 1)!(k + m_i)^{k + y_i}} \tag{3.2}$$

As $k \to \infty$, NB converges in distribution to poisson

$$P(Y_i = y_i) = \frac{\exp(-m_i) m_i^{yi}}{y_i!} \tag{3.3}$$

Let $\alpha = \frac{1}{k}$. When $\alpha = 0$, there is equi-dispersion; and when $\alpha > 0$, there is over-dispersion.

4. Generalized Poisson Regression (GPR) Model

GPR model is a natural extension of poisson regression model based on GPD model. Let Y be a generalized poisson random variable depending on x_i. Observe that expected value of GPD model is given by

$$\mu = \frac{\theta}{(1 - \lambda)} = \theta\varphi \tag{4.1}$$

where $$\varphi = \frac{1}{(1 - \lambda)}$$

Then, probability function of Y given x_i may be defined as follows

$$P(Y = y/x_i) = \frac{\mu[\mu + (\varphi - 1)y]^{y-1} \exp[-\{\mu + (\varphi - 1)y\}/\varphi]}{y!}, \text{ for } y = 0, 1, 2, \ldots$$

$$= 0, \text{ for } y > m \text{ when } \varphi < 1 \quad (4.2)$$

or otherwise where $\mu = \mu(\mathbf{x}) > 0$ and $\varphi \geq \max\left(-\frac{1}{2}, 1 - \frac{\mu}{4}\right)$ denotes the square root of the index of dispersion and μ is the largest positive integer for which $\mu + m(\varphi - 1) > 0$, when $\mu < 1$. When $\varphi = 1$, GPR model converges to poisson regression model; when $\varphi > 1$, GPR model represents count data with over-dispersion; and when $\frac{1}{2} < \varphi < 1$, GPR model represents count data with under-dispersion provided $\mu > 2$.

The probability function of restricted GPR model Y_i given x_i is defined by $f_i(y_i, \mu_i\ \alpha) = \left[\frac{\mu_i}{1 + \alpha\mu_i}\right]^{y_i} \left[\frac{(1 + \alpha y_i)^{y_i - 1}}{y_i!}\right] \exp\left[-\{\mu_i\ (1 + \alpha y_i)\}/(1 + \alpha\mu_i)\right]$ (4.5)

where $y_i = 0, 1, 2, \ldots, \alpha$ is the dispersion parameter

when $\alpha > 0$, $V(Y_i/\mathbf{x}_i) > E\ (Y_i/\mathbf{x}_i)$ [i.e., over-dispersion]

when $\alpha < 0$, $V\ (Y_i/\mathbf{x}_i) < E\ (Y_i/\mathbf{x}_i)$ [i.e., under-dispersion]

when $\alpha = 0$, $V\ (Y_i/\mathbf{x}_i) = E\ (Y_i/\mathbf{x}_i)$ [i.e., equi-dispersion]

and $\mu_i = \mu_i(\mathbf{x}_i) = e^{\mathbf{x}_i'\beta_i}$ where $\mathbf{x}_i$ is a (k − 1) dimensional vector or explanatory variables and β_i is a k-dimensional vector of regression parameters. The expected value of Y_i for any given x_i is defined by

$$E\ (Y_i/\mathbf{x}_i) = \mu_i \quad (4.6)$$

Variance of Y_i for any given x_i is defined by

$$V\ (Y_i/\mathbf{x}_i) = \mu_i(1 + \mu_i)^2 \quad (4.7)$$

In this study, parameter estimates for PR, NBR, and GPR models were constructed using SAS software package and Fortran programming. Additionally, standard error, Wald t-statistics, dispersion parameters, Pearson's and generalized chi-square, deviance, and log-likelihood estimates, and the number of iterations were generated for comparisons.

5. Application

5.1 Description of farm injury data structure

Study participants were obtained from farming injury data that was collected as a joint effort that began in 1994 by the University of Alabama at

Birmingham and Tennessee State University. A set of exclusion criteria for the Farm Safety/Injury baseline data was used. Records or observations from the analysis were excluded if they were missing information for any of the following: age, sex, race, marital status, educational level, information on safe use chemical training, information on farm safety training, county/state location of farming, acreage/size of farm, work other than farming, use a tractor, distance farm away from nearest medical facility, and ever had an agricultural injury. Table 1 shows the definitions for additional variables that were used in this analysis. Thus, final study population consisted of 1,322 respondents aged 18 years and older of which 96% were at least 25 years old; 91% were male; 52% were white (non-Hispanic); 72% were married; 73% had at least a high school education; 75% described their farming income as fair; 25% reported income from farming was good to excellent; 35% were within 10 miles of a medical facility; 45% had crop liability insurance; 22% were smokers; and 16% were classified as heavy drinkers.

Table 1. Farm safety/injury variable definitions

Variable	Definition
nf_inj	number of agricultural injuries sustained by each respondent in 1993
	Independent/Predictors or Covariates (coded as 1 if true, 0 = otherwise except the covariate f_inc)
f_life	respondents who live on a farm
f_own	respondents who own and work on a farm
f_szm	respondents who work on farm of size 200 to under 1000 acres
f_szl	respondents who work on farm of size 1000 acres or more
age2	respondents who are at least 25 years and under 45 years of age
age3	respondents who are at least 45 years and under 65 years of age
age4	respondents who are at least 65 years of age
sex1	respondents who are male
race1	respondents who are White (non-Hispanic)
mar1	respondents who are single (never married)
mar2	respondents who are married
educ2	respondents who have completed high school or GED
educ3	respondents who have completed post high school/higher education
inc1	respondents who describe income from farming as excellent/good
inc2	respondents who describe income from farming as fair

Variable	Definition
inc3	respondents who describe income from farming as poor
f_inc	respondents gross farming income in 1993 (1 = under \$ 10K, 2 = \$ 10K to under \$ 50K, 3 = \$ 50K to under \$ 100K, 4 = \$ 100K or more)
f_exp2	respondents who have farmed at least 10 years to under 20 years
f_exp3	respondents who have farmed at least 20 years
satisfy	respondents who are at least satisfied with farming
s_belt	respondents who always or usually wear seat belts
fe_sbelt	respondents who drive equipment on public roads with seatbelts
med_cond	respondents who may have any of following medical conditions: amputation, arthritis, asthma, back problem, depression, diabetes, emotional problem, heart disease, high blood pressure, poor hearing, paralysis, poor sight, problem moving, shortness of breath, stroke, weakness
acc_med	respondents whose farm is under 10 miles from a medical facility
insure	respondents who have crop or liability insurance
govt_a	respondents who receive government farm/land payments in 1993
heav_dr	respondents who drink at least 5 drinks/wk (wine, beer, liquor)
smoke	respondents who are current cigarette smokers
crops_f	respondents who spend at least 50% of time in field crops farming
livest-f	respondents who spend at least 50% of time in livestock farming
mult_f	respondents who spend at least 50% of time doing aquaculture, forestry/timber, fruit/vegetable or other types of farming
c_expo	respondents who mix or apply farm chemicals
c_train	respondents who have training about the safe chemicals use (CU)
c_trhr1	respondents who have no hours of safety training on CU
c_trhr2	respondents who have below 1 hour CU training
f_train	respondents who have farm safety training (not CU)
f_trhr1	respondents who have no hours of farm safety training (not CU)

Variable	Definition
f_trhr2	respondents who have below 1 hour farm safety training (not CU)
f_trhr4	respondents who have 4 to 8 hours farm safety training (not CU)
weekend	respondents who spend at least 1 hour on Saturday/Sunday farming
mhwf_ot	respondents who spend more than 10 hr/wk during any month of 1993 in farm work
pctmfot	respondents who spend more than 10% of farming time in any month of 1993 doing machinery operation
pctlfot	respondents who spend more than 10% of farming time in any month of 1993 doing livestock care taking
pctpfot	respondents who spend at least 1% of farming time in any month of 1993 applying pesticide/herbicide

5.2 Results

Tables 2 through 4 examine possible determinants of agricultural injuries. The application of PR, NBR, and GPR models and methods identified significant determinants on the number of injuries reported by the study population. Table 2 suggests race, gender, educational level, and marital status are positively associated with agricultural injuries. Table 3 shows that there is a significant inverse relationship between seat belt use and farm injury. Respondents who reported consistent use of seat belts were less likely to sustain farm injuries, and those with some form of medical conditions appear to have positive significant impact on the frequency of farm injuries. Table 4 shows that respondents with more hours (i.e. 4 to 8) of farm safety training demonstrate significant reduction in the number of injuries reported.

The following results were obtained from GPD analysis of the frequency of farming injuries: For values 0, 1, 2, 3, 4, 5, and 6 representing the number of farming injuries, observed frequencies were 1051, 168, 61, 15, 10, 8, and 8, respectively. Expected GPD frequencies were 1053, 171, 56, 23, 11, 6, and 3, respectively. The sample mean and sample variance were 0.35 and 0.78, respectively. Results for parameter estimates from GPD analysis showed that the number of iterations is 2, ML estimate θ is 0.229, and ML estimate of λ is 0.347. The chi-square test for the goodness of fit of the GPD, performed after pooling the last three cells (due to jumps in x-variate and small frequencies), yielded a chi-square statistic of 14.78 with 4 degrees of freedom and a corresponding p-value of 0.005; hence, using GPD to predict the frequency of farming injuries is shown to be consistent and supportive for the application of GPR models.

Table 2. Demographic determinants (age, gender, race, marital status, educational level, income level) of farming injuries: poisson, negative binomial, generalized poisson regression models

Variable	Poisson Regression			Negative Binomial Regression			Generalized Poisson Regression		
	Estimate	SE	Wald t-statistics	Estimate	SE	Wald t-statistics	Estimate	SE	Wald t-statistics
Intercept	–4.08	0.60	–6.79*	–4.04	0.77	–5.23*	–4.02	0.77	–5.18*
age2	0.79	0.46	1.69	0.77	0.54	1.41	0.77	0.54	1.42
age3	0.73	0.47	1.55	0.69	0.55	1.23	0.67	0.55	1.22
age4	1.11	0.48	2.32*	0.97	0.57	1.69	0.94	0.57	1.63
sex1	0.63	0.25	2.46*	0.80	0.33	2.38*	0.81	0.33	2.44*
race1	0.70	0.12	5.66*	0.72	0.17	4.12*	0.72	0.17	4.09*
mar1	0.68	0.24	2.74*	0.56	0.32	1.71	0.54	0.32	1.66
mar2	0.61	0.21	2.91*	0.56	0.27	2.03*	0.56	0.27	2.03*
educ2	–0.13	0.14	–0.89	–0.17	0.20	–0.87	–0.18	0.20	–0.91
educ3	0.58	0.14	4.13*	0.50	0.20	2.48*	0.48	0.20	2.38*
inc1	0.00	0.12	0.02	0.01	0.19	0.06	0.02	0.20	0.10
inc2	0.10	0.12	0.86	0.08	0.19	0.42	0.08	0.20	0.40
inc3	0.27	0.11	2.48*	0.24	0.17	1.43	0.24	0.17	1.36
f_inc	0.08	0.03	2.17*	0.08	0.05	1.34	0.07	0.06	1.27
Dispersion parameter α				2.97	0.36	8.17*	1.24	0.14	8.71*
Pearson's Chi-Square	2987.49						3057.82		
Generalized Chi-Square				1437.63			1407.34		
Deviance	1621.88			732.31			701.25		
Log-Likelihood	–1129.00			–959.88			–959.14		
Number of Iterations	4			8			6		

* Significant at 0.05 level.
SE : Standard Error

Note: See Table 1 for variable definitions.

Table 3. Health and social welfare determinants (experience, satisfaction, seatbelt use, medical conditions, access to health care, insurance, government assistance, and alcohol use, smoking status) of farming injuries: poisson, negative binomial, generalized poisson regression models

Variable	Poisson Regression			Negative Binomial Regression			Generalized Poisson Regression		
	Estimate	SE	Wald t-statistics	Estimate	SE	Wald t-statistics	Estimate	SE	Wald t-statistics
intercept	−1.76	0.23	−7.49*	−1.72	0.32	−5.39*	−1.72	0.32	−5.34*
f_exp2	0.37	0.22	1.65	0.34	0.30	1.12	0.34	0.30	1.12
f_exp3	0.59	0.21	2.82*	0.56	0.28	1.99*	0.56	0.28	1.98*
satisfy	−0.09	0.10	−0.91	−0.05	0.16	−0.31	−0.04	0.17	−0.24
s_belt	−0.86	0.25	−3.42*	−0.95	0.32	−2.93	−0.97	0.32	−3.01*
fe_sbelt	0.69	0.24	2.83*	0.73	0.31	2.35*	0.74	0.31	2.39*
med_cond	0.45	0.09	4.76*	0.49	0.14	3.44*	0.50	0.14	3.43*
acc_med	−0.05	0.09	−0.59	−0.09	0.14	−0.65	−0.10	0.15	−0.71
insure	0.23	0.10	2.20*	0.18	0.16	1.12	0.16	0.16	1.01
govt_a	0.23	0.10	2.21*	0.26	0.16	1.58	0.27	0.17	1.57
heav_dr	0.58	0.11	5.19*	0.61	0.19	3.23*	0.63	0.20	3.09*
smoke	−0.20	0.12	−1.67	−0.25	0.18	−1.32	−0.26	0.19	−1.36
Dispersion parameter α				3.31	0.39	8.44*	1.38	0.15	9.04*
Pearson's Chi-Square		3421.39						3381.96	
Generalized Chi-Square					1512.07			1450.68	
Deviance		1688.73			727.29			690.03	
Log-Likelihood		−1162.43			−974.14			−972.75	
Number of Iterations		2			8			6	

* Significant at 0.05 level.
SE : Standard Error

Note: See Table 1 for variable definitions.

Table 4. Farm-type and precautionary determinants (types of farming, chemical exposure, chemical use and farm safety training) of farming injuries: poisson, negative binomial, generalized poisson regression models

Variable	Poisson Regression			Negative Binomial Regression			Generalized Poisson Regression		
	Estimate	SE	Wald t-statistics	Estimate	SE	Wald t-statistics	Estimate	SE	Wald t-statistics
intercept	−1.37	0.17	−7.95*	−1.42	0.24	−5.76*	−1.44	0.25	−5.75*
crops_f	−0.31	0.15	−2.08*	−0.29	0.21	−1.34	−0.29	0.22	−1.31
livest_f	0.30	0.15	2.00*	0.37	0.21	1.71	0.38	0.21	1.77
mult_f	0.03	0.17	0.19	0.00	0.26	0.02	−0.00	0.27	−0.01
c_expo	0.04	0.10	0.46	0.07	0.16	0.45	0.08	0.17	0.46
c_train	0.64	0.13	4.67*	0.62	0.21	2.87*	0.61	0.22	2.76*
c_trhr1	−0.01	0.14	−0.08	0.05	0.23	0.22	0.07	0.25	0.31
c_trhr2	−0.52	0.27	−1.88	−0.56	0.41	−1.38	−0.58	0.41	−1.39
f_train	−0.04	0.12	−0.32	−0.03	0.20	−0.17	−0.03	0.21	−0.15
f_trhr1	0.23	0.15	1.47	0.21	0.25	0.87	0.21	0.26	0.82
f_trhr2	−0.07	0.23	−0.34	−0.00	0.36	−0.01	0.01	0.38	0.04
f_trhr4	−0.46	0.21	−2.10*	−0.43	0.30	−1.42	−0.43	0.30	−1.40
Dispersion parameter α				3.52	0.41	8.55*	1.47	0.16	9.15*
Pearson's Chi-Square	3474.45						3439.36		
Generalized Chi-Square				1497.26			1428.23		
Deviance	1725.32			726.04			685.98		
Log-Likelihood	−1180.72			−983.25			−981.76		
Number of Iterations	2			8			6		

* Significant at 0.05 level.
SE : Standard Error

Note: See Table 1 for variable definitions.

6. Discussion

Estimated dispersion parameter in each NBR and GPR model is positive and significantly different from zero. The implication is that further investigation of results of NBR and GPR rather than PR is needed, since conditional variance of response variable given selected explanatory variables is significantly greater than the associated conditional mean (an indication of over-dispersion). It is observed from the results that dispersion parameter estimates and their standard errors for GPR are consistently smaller than that of NBR; also, t statistics for the dispersion parameter under GPR models indicate higher significant values than that under NBR models. Furthermore, goodness-of-fit measures for PR, NBR, and GPR models shown in the tables indicate that GPR models outperform NBR and PR models.

GPR models obtained through use of ML and MM estimation procedures are quite similar in terms of parameter estimates and standard errors. Goodness-of-fit measures were not significantly different. The ML models showed higher log-likelihood than the MM models. ML and MM models indicated the same number of iterations. In summary, our results demonstrate that GPR models have statistical advantage over the PR and NBR models and are suitable for fitting various types of dispersed count data. It is observed that in the situation of equi-dispersion, estimated standard errors for the poisson regression model are over-estimated or larger than that of GPR or NBR.

Limitations in GPR modeling include the following: (i) Studies on how to obtain prediction interval for the single observation that follows a PR or GPR model are needed. (ii) On the conditional inference of mixture distributions for count data, there is a need to explore with various combinations of parameters through simulation to determine the information loss incurring for negative binomial and generalized poisson distributions. (iii) Investigations of GPR models with small data samples ($n < 30$) and effects of sample size in fitting GPR models to data need to be conducted.

REFERENCES

[1] Albert, J. and Pepple, P. (1989). A Bayesian approach to some over-dispersion models. *Canadian Journal of Statistics*, **17**, 333-344.

[2] Breslow, N. (1990). Tests of hypotheses in over-dispersed poisson regression and other quasi-likelihood models. *Jour. Amer. Statist. Assoc.*, **85**, 565-571.

[3] Consul, P.C. and Famoye, F. (1992). Generalized poisson regression model. *Comm. Stat. – Theory and Methods*, **21**, 89-109.

[4] Cox, D.R. (1983). Some remarks on over-dispersion. *Biometrika*, **70**, 269-274.

[5] Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Jour. Amer. Statist. Assoc.*, **8**, 709-721.

[6] Efron, B. and Thisted, R. (1976). Estimating the number of unseen species : how many words did Shakespeare know? *Biometrika*, **63**, 435-448.

[7] Hinde, J. (1982). Compound poisson regression models. In : *Glim 82, Proceedings of International Conference on Generalized Linear Models.* R Gilchrist, Springer, Berlin, 109-121.

[8] Lawless, J. (1987). Negative binomial and mixed poisson regression. *Canadian Journal of Statistics*, **15**, 209-225.

[9] Lu, W. and Morris, C. (1994). Estimation in the generalized linear empirical Bayes model using the extended quasi-likelihood. *Comm. Stat. – Theory and Methods (Part A)*, **23**, 661-688.

[10] Manton, K., Woodbury, M. and Stallard, E. (1981). A variance components approach to categorical data models with heterogeneous cell population: analysis of spatial gradients in lung cancer mortality rates in North Carolina counties. *Biometrics*, **37**, 259-269.

[11] McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models.* 2nd ed., Chapman and Hall, London.

[12] Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *J. Roy. Statist. Soc.*, **A135**, 370-384.

[13] Singh, K. and Famoye, F. (1993). Analysis of rates using a generalized poisson regression model. *Biometrical Journal*, **35**, 917-923.

[14] Stein, G. and Juritz, J. (1988). Linear models with an inverse Gaussian-poisson error distribution. *Comm. Stat. – Theory and Methods*, **17**, 557-571.

[15] Wang, W. and Famoye, F. (1997). Modeling household fertility with generalized poisson regression. *Journal of Population Economics*, **10**, 1-11.