Count Regression Models with an Application to Zoological Data Containing Structural Zeros

İlknur Özmen[1] and Felix Famoye[2]

[1]*Baskent University and* [2]*Central Michigan University*

Abstract: Recently, count regression models have been used to model over-dispersed and zero-inflated count response variable that is affected by one or more covariates. Generalized Poisson (GP) and negative binomial (NB) regression models have been suggested to deal with over-dispersion. Zero-inflated count regression models such as the zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB) and zero-inflated generalized Poisson (ZIGP) regression models have been used to handle count data with many zeros. The aim of this study is to model the number of C. caretta hatchlings dying from exposure to the sun. We present an evaluation framework to the suitability of applying the Poisson, NB, GP, ZIP and ZIGP to zoological data set where the count data may exhibit evidence of many zeros and over-dispersion. Estimation of the model parameters using the method of maximum likelihood (ML) is provided. Based on the score test and the goodness of fit measure for zoological data, the GP regression model performs better than other count regression models.

Key words: Count regression models, maximum likelihood, over-dispersion, zero-inflation, zoological data.

1. Introduction

Poisson regression is a standard model for analysis of count data. While the Poisson regression model may be the foremost candidate, it rarely explains the data due to several important constraints. One important constraint in the Poisson regression model is that the mean of the distribution must be equal to the variance. If this assumption is not valid, then the standard errors, usually estimated by the ML method, will be biased and the test statistics derived from the models will be incorrect. When the sample variance is larger (or smaller) than the sample mean, the data is said to exhibit over-dispersion (or under-dispersion). In overcoming the problem of over-dispersion, several researchers (Lawless, 1987; Famoye, 1993) employed the NB and GP regression models instead of the Poisson regression model. In these regression models, the estimates of the regression parameters are obtained by incorporating a dispersion parameter.

A feature of many count data sets is the presence of many zero observations relative to the Poisson assumption. This feature may be accounted for by over-dispersion in the data set. Over-dispersion has the tendency to increase the proportion of zeros and whenever there are many zeros relative to Poisson assumption, the NB and GP regression models tend to improve the fit of the data.

If there are many zero counts in the data, two states may be assumed to better reflect the situation. One of the states is the structural zero (or zero count) state where the only counts are zeros. The other state is the sampling zero state where the counts could be zeros or values greater than zero. Famoye and Singh (2006) illustrated these states with the number of accidents that adult drivers aged 65-70 had during the last five years. Those adults who did not drive in the last five years would have zero accident and they belong to the structural zero state. Those who drove and did not have any accidents belong to the sampling zero state. The probability of the structural zero state and the mean number of the event counts in the sampling zero state may depend on the covariates. Sometimes this probability and mean are unrelated while at other times the probability may assume a simple function of the mean.

In recent years, there has been considerable interest in using the ZIP model to fit count data in order to allow for the presence of too many zeros. ZIP regression models were considered as a mixture of a zero point mass and a Poisson distribution and were first used to study soldering defects on print wiring boards (Lambert, 1992). Lambert points out that the probability of a perfect state (i.e., zero defect state) and the mean of the imperfect state (i.e., non-zero defect state) depends on the covariates. On the other hand, to account for over-dispersion (or under-dispersion) in the Poisson part, generalizations of the model are possible. These include ZINB regression model for over-dispersion situation and ZIGP regression model for over- or under-dispersion situation. Heilbron (1994) proposed the use of ZINB regression models to assess the covariate effect on high-risk heterosexual behavior. Gupta *et al.* (1996) proposed the use of the zero-adjusted generalized Poisson model to analyze over-dispersed fetal movement data and the death notice of data of London times. Famoye and Singh (2006) extended their work (Gupta *et al.*, 1996) to a more general situation where the count dependent variable is affected by some covariates. Famoye and Singh (2006) noted cases where the ZIP regression models were inadequate and the ZINB regression model could not be fitted to an observed data set. This realization motivated them to develop a ZIGP regression model for modeling over-dispersed count data with too many zeros. For illustration, they applied ZIGP regression model to domestic violence data.

In this paper, the Poisson, NB, GP, ZIP and ZIGP regression models will be

used to model the zoological data set which is described in section 2. In section 3, we present a review of the count regression models, estimation of model parameters, goodness of fit test and score test. The results of applying the count regression models to model the number of C. caretta hatchlings dying from exposure to the sun are presented in section 4. The paper ends with conclusion and discussion in section 5.

2. Description of Zoological Data

The zoological data were collected from field studies realized in 1991-1993 on Dalyan Beach in Turkey (Canbolat, 1997). The responses $y_i, (i = 1, \ldots, 72)$ are the number of C. caretta hatchlings dying from exposure to the sun. The reason for the deaths is that they hesitate to walk. Thus, their walk takes a long time thereby exposing them to too much sun light and this causes death. While the number of C. caretta hatchlings emerging from the nests for the years 1991-1993 are, respectively, 4804, 4377 and 4704, the deaths of C. caretta hatchlings due to sun during the same periods are 48, 51 and 59. The y_i's ranged in size from 0 to 23. There are three qualitative factors: Area (A1-A6), distance from the sea (D1-D4) and year (1991-1993). Dalyan Beach is constructed from three main areas: the strait, the lake and the small beach. Approximately 80% of the nesting in Dalyan Beach occurs in the strait. Thus, the strait is divided into four equal one-kilometer pieces, (A1: 0-1 kilometer of the strait, A2: 1-2 kilometers of the strait, A3: 2-3 kilometers of the strait, A4: 3-4 kilometers of the strait) and has been evaluated separately from the lake (A5) and the small beach (A6). The distances of the nests from the sea are D1 (0-10 meters), D2 (10-20 meters), D3 (20-30 meters) and D4 (≥ 30 meters). For the results in this paper, binary indicator covariates were used to represent main effects (five for areas, three for distance from the sea and two for year), and log-linear specification $\mu_i = \mu(\mathbf{x}_i) = \exp(\sum_{j=1}^{11} x_{ij}\beta_j), i = 1, \ldots, 72$, was employed. The data are summarized for each level of these three factors in Table 1.

The zoological data include both structural zeros and sampling zeros. If no C. caretta hatchlings emerge from the nests, the number of C. caretta hatchlings dying from exposure to the sun is automatically zero. If C. caretta hatchlings emerge from the nests, the number of C. caretta hatchlings dying from exposure to the sun may be zero or greater than zero. The zeros from the first state occur with probability p_i (C. caretta hatchlings did not emerge from the nests in 12 cells of Table 1). The second state occurs with probability $1 - p_i$ (no C. caretta hatchlings die from exposure to the sun in 29 cells of Table 1).

Table 1: The number of C. caretta hatchlings dying from exposure to the sun on Dalyan Beach in Turkey

	1991				1992				1993			
Area	D1	D2	D3	D4	D1	D2	D3	D4	D1	D	D3	D4
A1	0	17	1	7	0	2	5	0*	1	0	0	0
A2	2	7	2	2	0	1	8	0	23	9	7	3
A3	0	0	0	0	0	3	1	0*	3	1	4	0*
A4	0	0	0	0	0	1	0	0	0*	0	0	0*
A5	8	1	0*	0*	4	3	0*	0*	0	0	0	0*
A6	0	0	1	0*	0	1	0	22	0*	0	2	6

*The zeros denote structural zeros. The others denote sampling zeros.

3. Regression Models and Parameter Estimation

The Poisson regression model is given by

$$p(\mu_i, y_i) = \frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!}, \quad y_i = 0, 1, \dots \tag{3.1}$$

where $\mu_i = \mu_i(\mathbf{x}_i) = \exp(\sum_{j=1}^{p} x_{ij}\beta_j)$, $i = 1, 2, \dots, n$, $\mathbf{x}_i = (x_{i1} = 1, x_{i2}, \dots, x_{ip})$ is the i-th row of covariate matrix $\mathbf{X}$, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ are unknown p-dimensional column vector of parameters. In the Poisson regression model, the mean of the distribution is equal to the variance, i.e. $E(Y_i|\mathbf{x}_i) = Var(Y_i|\mathbf{x}_i) = \mu_i$. Details of the Poisson regression model are given in Frome *et al.* (1973) and Frome (1983).

The Poisson regression model is usually restrictive for count data, leading to alternative models like the NB and GP regression models. One way this restriction manifests itself is that in many applications a Poisson density predicts the probability of a zero count to be considerably less than is actually observed in the data. This is termed excess zeros problem, as there are more zeros in the data than the Poisson predicts. The second and more obvious way that the Poisson is deficient is that for a count data, the variance usually exceeds the mean, a feature called over-dispersion. If there is significant over-dispersion in the distribution of the count, the estimates from the Poisson regression model will be consistent, but inefficient. The standard errors in the Poisson regression model will be biased downward. This situation could lead the investigator to make incorrect statistical inferences about the significance of the covariates. The NB and GP regression models provide an alternative to the Poisson regression model. The NB regression model has been used to deal with only over-dispersion

(Lawless, 1987). The GP regression model has been used to deal with over- and under-dispersion (Famoye, 1993). Therefore, a statistical test of over-dispersion is highly desirable after fitting a Poisson regression model.

The standard form of the NB distribution used in regression applications specifies that $\mu_i = \mu_i(\mathbf{x}_i) = \exp(\sum_{i=1}^{p} x_{ij}\beta_j)$. The standard form includes the dispersion parameter α and the conditional variance function, which is quadratic in the mean. The NB regression model with the mean $E(Y_i|\mathbf{x}_i) = \mu_i$ and variance $Var(Y_i|\mathbf{x}_i) = \mu_i(1 + \alpha\mu_i)$ is given by Lawless (1987) as

$$p(\mu_i, \alpha, y_i) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i!\Gamma(\alpha^{-1})} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \left(\frac{1}{1 + \alpha\mu_i}\right)^{\alpha^{-1}}, \quad y_i = 0, 1, \ldots \quad (3.2)$$

where $\Gamma(\cdot)$ denotes the gamma function and the dispersion parameter α is unknown. In the limit as α goes to 0, (3.2) yields the Poisson regression model. When $\alpha > 0$, there is over-dispersion.

The GP regression model provides an alternative to Poisson regression model for over- and under-dispersion (Famoye, 1993). It is a good competitor to the NB regression model when the count data is over-dispersed. This model can be written as

$$p(\mu_i, \alpha, y_i) = \left(\frac{\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \frac{(1 + \alpha y_i)^{y_i-1}}{y_i!} \exp\left(\frac{-\mu_i(1 + \alpha y_i)}{1 + \alpha\mu_i}\right), \quad (3.3)$$

$y_i = 0, 1, \ldots$. The mean of Y_i is given by $E(Y_i|\mathbf{x}_i) = \mu_i$ and the variance of Y_i is given by $Var(Y_i|\mathbf{x}_i) = \mu_i(1 + \alpha\mu_i)^2$. In model (3.3), α is called dispersion parameter. When $\alpha = 0$, the GP regression model (3.3) reduces to the Poisson regression model (3.1) and this is a case of equi-dispersion. When $\alpha > 0$ (or when $\alpha < 0$), the GP regression model represents count data with over-dispersion (or under-dispersion).

Sometimes there are too many zeros in the count dependent variable than are predicted by the Poisson regression model, resulting in an overall poor fit of the model to the data. Zero-inflated count (ZIP, ZINB and ZIGP) regression models address this problem of excess zeros. One cause of over-dispersion that is expressed above may be the presence of more zeros than expected for the Poisson regression model. The NB, GP, ZINB and ZIGP regression models can be appropriate for the over-dispersion situation. For count data with more zeros than expected, several models have been proposed, for example the "hurdle model" (Mullahy, 1986), the "ZIP regression model" (Lambert, 1992), the "two-part model" (Heilbron, 1994), the "semi-parametric model" (Gurmu, 1997). Details of these models are also given in Ridout *et al.* (1998). The ZINB regression model has been proposed by Heilbron (1994) and Ridout *et al.* (2001) and the ZIGP regression model has been proposed by Famoye and Singh (2006).

If Y_i are independent random variables having a zero-inflated count distribution, the zeros are assumed to arise in two ways corresponding to distinct underlying states. The first state occurs with probability p_i and produces only zeros, while the second state occurs with probability $(1 - p_i)$ and leads to the Poisson, NB and GP count with mean μ_i. In general, the zeros from the first state are called structural zeros and those from the second state are called sampling zeros.

Consider discrete nonnegative random variable Y_i with a zero-inflated count distribution, where p_i and μ_i denote respectively the proportion of zeros and the mean in the Poisson, NB or GP distribution. The distribution of Y_i is given as

$$
\begin{aligned}
Pr(Y_i = y_i) &= p_i + (1 - p_i)Pr(K_i = 0) & y_i = 0 \\
&= (1 - p_i)Pr(K_i = y_i), & y_i > 0
\end{aligned}
\tag{3.4}
$$

The overall probability of zero count is a combination of probabilities of zeros from each state, weighted by the probability of being in that state, i.e. $p_i + (1 - p_i)Pr(K_i = 0)$, where $Pr(K_i = 0)$ is a Poisson, NB or GP probability with zero event that occurs by chance. On the other hand, the probability of positive counts is given by $(1 - p_i)Pr(K_i = y_i)$ where $Pr(K_i = y_i)$ is the Poisson, NB or GP probability with positive counts. Hence, by combining $p_i + (1 - p_i)Pr(K_i = 0)$ and $(1 - p_i)Pr(K_i = y_i)$, the zero-inflated count regression model can be expressed as in (3.4). In (3.4), $0 < p_i < 1$ so those extra zeros in the data are explicitly modeled. For positive value of p_i, it represents zero-inflated distribution. When p_i is allowed to be negative, it represents zero-deflated distribution. However, zero-deflation rarely occurs in practice. The ZIP, ZINB and ZIGP regression models are summarized in Table 2.

Table 2: Probability functions, expected value and variance of zero-inflated count regression models

Models	$Pr(K_i = y_i)$	$E(Y_i)$	$Var(Y_i)$
ZIP	$\frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!}$	$(1 - p_i)\mu_i$	$E(Y_i)(1 + p_i\mu_i))$
ZINB	$\frac{\Gamma(y_i + \alpha^{-1})}{y_i!\Gamma(\alpha^{-1})}\left(\frac{\alpha\mu_i}{A}\right)^{y_i}\left(\frac{1}{A}\right)^{\alpha^{-1}}$	$(1 - p_i)\mu_i$	$E(Y_i)(A + p_i\mu_i)$
ZIGP	$\left(\frac{\mu_i}{A}\right)^{y_i}\frac{(1 + \alpha y_i)^{y_i - 1}}{y_i!}\exp\left(\frac{-\mu_i(1 + \alpha y_i)}{A}\right)$	$(1 - p_i)\mu_i$	$E(Y_i)(A^2 + p_i\mu_i)$

$A = (1 + \alpha\mu_i)$ for short.

Lambert (1992) proposed that be formulated as a logit transformation such that $\text{logit}(p_i) = \log(p_1/(1 - p_i)) = \sum_{j=1}^{m} z_{ij}\delta_j$, where $\mathbf{z}_i = (z_{i1} = 1, z_{i2}, \ldots, z_{im})$

is the i-th row of covariate matrix $\mathbf{Z}$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$ are unknown m-dimensional column vector of parameters. The mean $\mu_i = \mu_i(\mathbf{x}_i)$ satisfies a log-linear relationship with covariates such that $\log(\mu_i) = \sum_{j=1}^{p} x_{ij}\beta_j$, where $\mathbf{x}_i = (x_{i1} = 1, x_{i2}, \ldots, x_{ip})$ is the i-th row of covariate matrix $\mathbf{X}$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ are unknown p-dimensional column vector of parameters. Both the nonnegative functions p_i and μ_i are linear functions of some covariates. The covariates affecting p_i and μ_i may or may not be the same. In the case of dissimilar covariates, the p_i and μ_i of the ZIP, ZINB and ZIGP regression models can be expressed as

$$
\begin{aligned}
\mu_i &= \exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right) \\
p_i &= \frac{\exp(\sum_{j=1}^{m} z_{ij}\delta_j)}{1 + \exp(\sum_{j=1}^{m} z_{ij}\delta_j)}
\end{aligned}
\tag{3.5}
$$

In the case of similar covariates, affecting both p_i and μ_i, the number of the parameters can be reduced by treating p_i as a function of μ_i (see Lambert (1992)).

In the count regression model, the response (the dependent variable y_i), namely, the number of C. caretta hatchlings dying from exposure to the sun, is a nonnegative integer and has a Poisson, or NB or GP distribution. Parameters in the count regression models are estimated by ML method. The ML method starts from the construction of the log-likelihood functions (L). Using the method of ML, the parameter estimates in Poisson, NB and GP regression models are, respectively, given by Frome (1983), Lawless (1987) and Famoye (1993). They discuss the use of algorithms for solving the system of the ML equations. For the zero-inflated count regression models, the inflation is modeled through $\omega_i = p_i/(1-p_i) = \exp(\sum_{j=1}^{m} z_{ij}\delta_j)$. Since we did not consider any covariate as z_{ij}, we used only a column of 1's for z_{i1}. Hence $\omega_i = \exp(z_{i1}\delta) = \exp(\delta)$. We used the SPLUS nonlinear optimization function "nlminb" to obtain the ML estimates of the parameters.

The goodness of fit of the count regression models for model selection can be based on the log-likelihood value or deviance statistic. We use the deviance statistic to measure the goodness of fit of the regression models. The deviance, based on a likelihood ratio statistic, is a measure of the goodness of fit. This is defined as $D = -[L(\hat{\mu}_i; y_i) - L(y_i; y_i)]$. It has been shown that the deviance can be approximated by a chi-square distribution with $n - k$ degrees of freedom (n is number of observations or cells and k is number of estimated parameters) when $\hat{\mu}_i$ is large. The deviance can be used to assess the adequacy of various regression models. The regression model with the smallest value of the deviance, among the regression models considered, is usually taken as the best model for fitting the

given data. The log-likelihood functions and deviances for the count regression models are available from the first author.

The NB and GP regression models reduce to the Poisson regression model when $\alpha = 0$. To assess the significance of the dispersion parameter, we test the hypothesis $H_0 : \ \alpha = 0$ against $H_1 : \ \alpha \neq 0$. Whenever H_0 is rejected, it is recommended to use the NB or GP regression model in place of the Poisson regression model. To carry out the test, one may use the asymptotically normal Wald type "t" statistic defined as the ratio of the estimate of α to its standard error.

A score test is used to test whether there are too many zeros for the Poisson and the generalized Poisson models to adequately fit the data. The reader is referred to Famoye and Singh (2006) for a discussion of the score test for zero inflation with respect to ZIGP regression model.

4. Results

To understand how the different count regression models fit the zoological data, we examine the fit of various regression models to the number of C. caretta hatchlings dying from exposure to the sun. The results of using the Poisson, NB, GP, ZIP and ZIGP regression models are given in Table 3.

First, we consider the Poisson regression model. Based on the deviance in Table 3, the Poisson regression model does not provide an adequate fit to the zoological data. The observed proportion of zeros is 56.9% for the zoological data, but the Poisson regression model predicts a proportion of zeros as 28.6%, which is an under-estimation of the observed proportion of zeros. In such a situation, it would be appropriate to estimate the ZIP regression model. Is the ZIP regression model statistically preferred over the Poisson regression model? We apply the score test to check whether the ZIP regression model is a significant improvement over the Poisson regression model. To test for zero inflation, the value of the score statistic is calculated as 100.68. This value is significant at 0.05 level when compared to the $\chi^2_{0.05;1} = 3.841$. That is, the value of the score statistic provides evidence that many zeros are observed for the Poisson distribution. Although the ZIP regression model does better in predicting the zero proportion (estimated proportion of zeros is 57.0%), the zero inflation parameter (δ) associated with z_{i1} is not significant for the ZIP regression model in Table 3. The Poisson and ZIP regression coefficients are quite different in magnitude; the standard errors for the ZIP regression coefficients tend to be larger than those for the Poisson regression coefficients.

We consider the NB and GP regression models which include a dispersion parameter. The test of $\alpha = 0$ by using the asymptotic Wald statistic showed that α is significantly different from zero for the NB and GP regression models in Table

3. The Poisson regression model is not appropriate for zoological data since we reject the hypothesis $H_0 : \alpha = 0$. The deviances for the Poisson, ZIP, NB and GP regression models are, respectively, 321.08, 250.96, 58.83 and 53.68, which also indicate that modeling over-dispersed data using NB and GP regression models are better than the Poisson and ZIP regression models. In Table 3, there is a significant negative relationship between the A4 area (3-4 kilometer of the strait) and the number of C. caretta hatchlings dying from exposure to the sun in the NB and GP regression models. Thus, the deaths from exposure to the sun of C. caretta hatchlings are low in the farthest zone of the strait area.

5. Conclusion and Discussion

The application addressed in this paper involves the estimation of Poisson, NB, GP, ZIP and ZIGP regression models to predict the number of C. caretta hatchlings dying from exposure to the sun. Since count data frequently exhibit over-dispersion in addition to possible zero inflation, an obvious methodology is to use a model that can accommodate over-dispersion and zero-inflation.

We also consider the ZINB and ZIGP regression models in terms of both zero inflation and over-dispersion situation. The ZINB and ZIGP regression models are alternates to the ZIP regression model when there is a situation of zero inflation. The ZINB regression model did not converge in fitting the zoological data. Lambert (1992), Famoye and Singh (2006) also observed similar problem in fitting ZINB regression model to observed data sets. The ZIGP regression model is a competitor to the ZINB regression model when there is both over-dispersion and zero inflation. For this reason, we apply the ZIGP regression model for modeling over-dispersed zoological data with many zeros. We apply a score test that tests whether the ZIGP regression model is a significant improvement over the GP regression model. The value of the score statistic is 0.63. This value is not significant at 0.05 level when compared to the tabulated chi-square distribution with one degree of freedom. Based on this result, the ZIGP regression model provides an adequate (but not better than GP) fit to the data. In Table 3, the zero inflation parameter (δ) associated with z_{i1} is not significant for the ZIGP regression model.

Although the zoological data has about 56.9% observed proportion of zeros, our results in section 4 showed that the ZIP regression model is not appropriate for fitting it. However, the ZIGP model provides a similar fit as the GP model. It appears the 56.9% of zeros does not constitute zero-inflation when one considers a regression model that incorporates dispersion parameter. Thus, over-dispersion in the zoological data can be a result of unobserved heterogeneity. Based on the findings shown in the previous section, the NB and GP regression models seem to perform better than the Poisson and ZIP regression models. The deviances

of the NB and GP regression models that incorporate dispersion parameter are very close; a slight preference might be given to the GP regression model which has the smallest deviance.

Table 3: Results of fitting Poisson, NB, GP, ZIP and ZIGP regression models

Variable	Poisson	NB	GP	ZIP	ZIGP
Intercept	0.9561	1.0202	2.0529	1.9939	2.3405
	0.2507*	0.7869	1.6332	0.2662*	1.0974*
A2	0.6624	0.6745	0.5932	0.1145	0.0074
	0.2143*	0.7971	1.6164	0.2187	1.0444
A3	−1.0116	−1.0178	−1.1567	−1.0002	−1.4919
	0.3371*	0.8561	1.5919	0.3934*	1.1679
A4	−3.4965	−3.6372	−4.0047	−3.6715	−4.1585
	1.0150*	1.2579*	1.6938*	1.0657*	1.4431*
A5	−0.7239	−0.8738	−1.6485	−0.6433	−1.6201
	0.3046*	0.8300	1.4160	0.3584	1.0358
A6	−0.0308	−0.0384	0.9315	0.1055	0.0097
	0.2481	0.8713	2.6098	0.2765	1.3853
D2	0.1151	0.1390	−0.3354	−0.5055	−0.5907
	0.2148	0.7085	1.0802	0.2469*	0.8534
D3	−0.2796	−0.5063	−1.1837	−0.9005	−1.3293
	0.2380	0.7139	1.2123	0.2749*	0.9244
D4	−0.0247	−0.2760	−1.7356	−0.1984	−1.2478
	0.2222	0.8172	1.7812	0.2698	1.2079
1992	0.0606	0.3864	0.2840	0.3207	0.6312
	0.2011	0.6226	1.3373	0.2255	0.7767
1993	0.2063	0.1028	−0.2755	0.2268	0.2338
	0.1944	0.6793	1.5324	0.2212	0.9168
δ	NA	NA	NA	−0.095	−0.8218
				0.2728	0.6675
Estimated dispersion parameter (α)	NA	2.6311	0.8682	NA	0.4000
		0.7100*	0.2144*		0.2048
Estimated proportion of zeros (%)	28.6	56.5	56.1	57.0	57.2
Log-likelihood	−203.37	−116.91	−117.45	−146.98	−116.75
Deviance	312.08	58.83	53.68	250.96	59.41
Degree of freedom	61	60	60	60	59

*means significant at 0.05 level. Standard errors of estimates are presented on the next rows.

References

Canbolat, A. F. (1997). Population biology of the loggerhead sea turtle Caretta caretta (LINNAEUS. 1758) in Dalyan and Patara. Ph.D. Thesis. University of Hacettepe. Turkey.

Famoye, F. (1993). Restricted generalized Poisson regression model. *Communications in Statistics: Theory and Methods* **22**, 1335-1354.

Famoye, F. and Singh, K. P. (2006). Zero-inflated generalized Poisson regression model with and application to domestic violence data. *Journal of Data Science* **4**, 117-130.

Frome, E. L., Kutner, M. H. and Beauchamp, J. J. (1973). Regression analysis of Poisson distributed data. *Journal of American Statistical Association* **68**, 935-940.

Frome, E. L. (1983). The analysis of rates using Poisson regression models. *Biometrics* **39**, 665-674.

Gupta, P. L., Gupta, R. C. and Tripathi, R. C. (1996). Analysis of zero-adjusted count data. *Computational Statistics and Data Analysis* **23**, 207-218.

Gurmu, S. (1997). Semi-parametric estimation of hurdle regression models with an application to Medicaid utilization. *Journal of Applied Econometrics* **12**, 225-242.

Heilbron, D. C. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal* **36**, 531-547.

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

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* **15**, 209-225.

Mullahy, J. (1986). Specification and testing of some modified count models. *Journal of Econometrics* **33**, 341-365.

Ridout, M. Dometrio, C. G. B. and Hinde, J. (1998). Models for count data with many zeros. Invited paper presented at the Nineteenth International Biometric Conference. Cape Town. South Africa. 179-190.

Ridout, M. Hinde, J. and Dometrio, C. G. B. (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics* **57**, 219-223.

İlknur Ozmen
Department of Statistics and Computer Sciences
Baskent University
Baglica Campus
Ankara, 06530, TURKEY
ilknur@baskent.edu.tr

Felix Famoye
Department of Mathematics
Central Michigan University
Mount Pleasant, MI 48859, USA
felix.famoye@cmich.edu