Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 7, Issue 2 (2014): 139-147

ANALYSIS OF CESAREAN-SECTION BIRTHS- AN APPLICATION OF ZERO-TRUNCATED GENERALIZED POISSON (ZTGP) REGRESSION MODEL

*Talawar A. S. and Sarvade R. M.
Department of Statistics, Karnatak University, Dharwad
E Mail: talwarappu@gmail.com

Received August 12, 2014
Modified October 31, 2014
Accepted November 20, 2014

Abstract

Fitting of ZTGP model is done using C-section data. Data set consists of annual total births, hospital type (private and public) and C-sections. The response variable Y denotes the number of C-sections which do not have any zero values. We first regress the response variable, 'C-sections' against one explanatory variable 'number of births', then we add one more explanatory variable in the form of indicator variable hospital type (public hospital=1, private hospital=0) in the Poisson regression analysis. A measure of goodness of fit of the ZTGP regression model is used on the log-likelihood statistic. The addition of the dispersion parameter α in the ZTGP regression model is justified by testing for the adequacy of the ZTGP model over the ZTP regression model.

Key Words: **c**-Section, Poisson Regression, Truncation, Link Function, Log-Likelihood, Chi Square Test**.**

1. Introduction

In the past when the gynecology was in its sleepy stage and medical science had no sophisticated development, the only way out for the pregnant woman was to give the virginal birth with much pains and pangs. The mid wives used to perform the roles of modern trained nurses and expert obstetricians. Many a time deliveries ended either in the death of a baby or a mother or sometimes both. It was due to some undiagnosed problems of the foetus due to the absence of scanning and advance medical facilities. The cesarean delivery was a dream of the day. The cesarean delivery is not without its bad effects as an abdominal cut is prone to bacterial infection. More than this the mother has to undergo the traumatic physical sufferings. The uterine rupture, injuries to the vital uterine organs add only fuel to the fire. There are psychological factors which may turn the mother into a half prepared mother. Anything unnatural has its disadvantages. In case there is an abnormality, say an abnormal foetal growth, placental problem, sugar and blood-pressure problems, a cesarean delivery is the only answer. An avoidance under an emergency case may result in the death of a mother is of utmost importance.

Cesarean section (C-section) was introduced in clinical practice as a life saving procedure both for the mother and the baby. The use of C-section follows the health care inequity pattern of the world. It has been under used in low income settings and adequate or even some unnecessary usage has been observed in middle and high income settings (Betran et al., (2007), Ronsmans et al., (2006)). Some relations have shown negative relationship between C-section rates and maternal and infant mortality at population level in low income countries (Althabe et al., (2006)). It is observed that a total of 54 countries had C-section rates below 10%, 14 countries had rates between 10% and 15% and 69 countries had C-section rates above 15% (Gibbons et al., (2010)).

For a random variable representing the number of counts, where the count data exhibit unequal sample mean and variance, the sample variance is either smaller (under-dispersion) or larger (over-dispersion) than the sample mean. Various models and associated estimation methods have been proposed to deal with these dispersions. Wang and Famoye, (1997) have developed generalized Poisson regression (GPR) model to study household fertility data set. The same model has been used Wulu et al., (2002) to model injury data. Lambart (1992) explained the zero-inflated Poisson (ZIP) regression models with the application of manufacturing defects and Lee et al., (2001) applied the same model to accommodate the extent of individual exposure. Gupta et al., (1996) fitted zero-adjusted (inflated or depleted) generalized Poisson model with foetal movement data of London times. Lee at el., (2006) developed multi-level Zero Inflated Poisson regression model for the correlated count data with excess zeros. Zhao et al., (2010) used the Zero Truncated Generalized Poisson model to fit data of hospital length of stay.

2. Analysis of Cesarean births:

The chief sources of data for C-section births are UNICEF 2010, www.childbirthconnection.org, www.who.int, www.cdc.gov, www.cpc.unl.edu etc. The analysis of data is carried for both data sets using the χ^2-test of independence of attributes.

		Hospital		Total
		Private	Public	
Births	Normal	412(96.3)	19976(98.6)	20388(98.5)
	C-section	16(3.7)	285(1.4)	301(1.5)
	Total	428(2.1)	20261(97.9)	20689

χ^2=15.87, χ^2_1=3.84 at 5%level of significance

Table 1 (a): Number of Cesarean births in private and public hospitals.

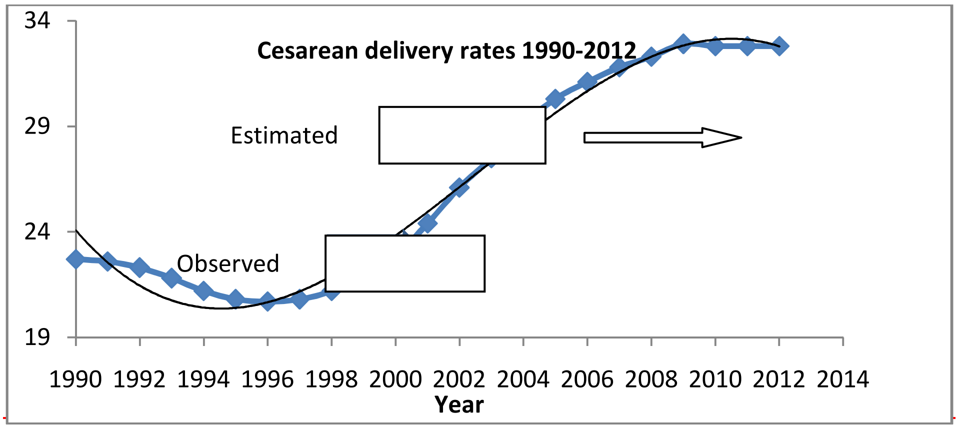
		Hospital		Total
		Private	Public	
Births	Normal	3744(64.37)	10133(87.55)	13877(79.8)
	C-section	2072(35.63)	1441(12.45)	3513(20.2)
	Total	5816(33.44)	11574(66.56)	17390

χ^2=1289.74, χ^2_1=3.84 at 5%level of significance.

Table 1 (b): Number of Cesarean births in private and public hospitals.

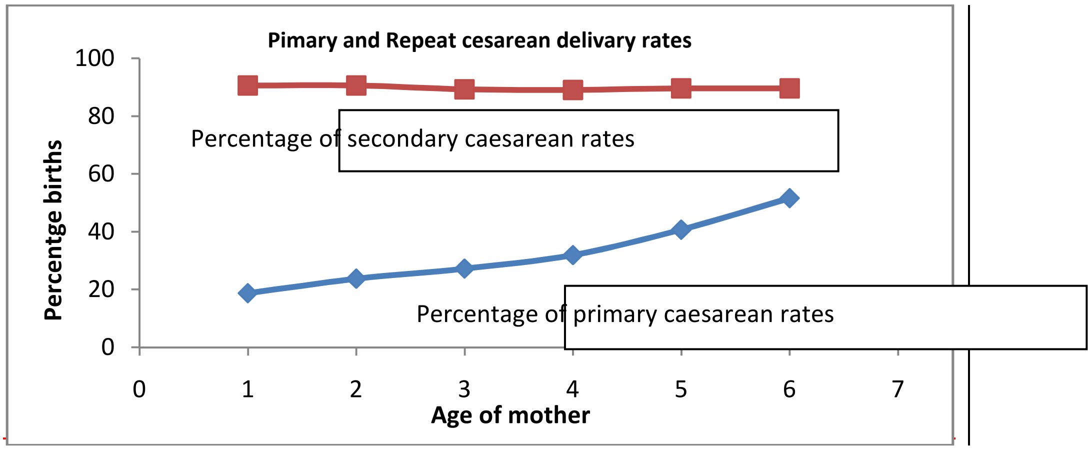
Source: UNICEF 2010

From both data sets, it is observed that χ^2 is significant at 5% level of significance. That is hospital type and births are not independent (Table 1(a) and 1(b)). Table 1(a) shows that out of 428 total births in private hospital, 3.7% are cesarean births, whereas in case of public hospitals, the cesarean births are only 1.4%. We find more number of C-section births (285 out of 20261) in public hospitals compared to only 16 in private hospitals. Table1 (b) shows that nearly 1/5th of births is cesarean births (20.2%). More than 35% cesarean births delivered in private hospitals. It is observed from the table that, the cesarean births are more in private (fee paying) hospitals than the public (non-fee paying) hospitals.



Source: Hamilton et al., (2013)
Fig 1(a): Observed and estimated cesarean births during 1990-2012.

Polynomial equation of degree 3 gives good fit to the percentage c-section and the R-square value is very high ($R^2 = 0.9844$).



Source: Martin et al., (2013)
Fig 1(b): Comparison of primary and secondary (repeat) cesarean delivery rates.

It is observed from the figure that as age increases, the chance of cesarean delivery also increases. 40 and above age group is at high risk.

3. Zero-Truncated Poisson Regression Model

In present Chapter, we propose zero-truncated generalized Poisson model (ZTGP) to cesarean births, because, there is no zero count. Let Y denote the number of cesarean births with probability mass function

$$f(y;\lambda) = \frac{1}{y!}\left(\frac{\lambda}{1+\alpha\lambda}\right)^{y}(1+\alpha y)^{y-1}e^{\left\{-\frac{\lambda(1+\alpha y)}{1+\alpha\lambda}\right\}},\quad \lambda > 0 \qquad (1)$$

where y=0, 1, 2, ... and the parameter α is a dispersion parameter. Mean and variance of this distribution are respectively, λ and $\lambda(1+\alpha\lambda)^2$. If α = 0 (1) reduces to the Poisson distribution. While α>0 is over-dispersion and α<0 is under-dispersion in the generalized Poisson distribution. We often find that in cesarean births, length of hospital stay data etc., no zero counts. Therefore we consider the zero-truncated generalized Poisson distribution which can be expressed as

$$f(y;\lambda|y>0) = \frac{1}{y!\left(e^{\left(\frac{\lambda}{1+\alpha\lambda}\right)}-1\right)}\left(\frac{\lambda}{1+\alpha\lambda}\right)^{y}(1+\alpha y)^{y-1}e^{\left\{-\frac{\alpha\lambda y}{1+\alpha\lambda}\right\}},\quad \lambda > 0 \qquad (2)$$

with y=1, 2, …. .The model (2) is denoted by ZTGP (α, λ). When α=0, the distribution reduces to zero-truncated Poisson (ZTP) model.

Following the generalized linear model approach, we relate parameters λ_i to covariates $x_i \in R^p$ through the log-link function so that

$$\log \lambda_i = x_i^T\beta \quad \text{or} \quad \lambda_i = e^{x_i^T\beta} \qquad (3)$$

Then the model (2) can be written as

$$f(y_i;\lambda_i|y_i>0) = \frac{1}{y_i!\left(e^{\left(\frac{\lambda_i}{1+\alpha\lambda_i}\right)}-1\right)}\left(\frac{\lambda_i}{1+\alpha\lambda_i}\right)^{y_i}(1+\alpha y_i)^{y_i-1}e^{\left\{-\frac{\alpha\lambda_i y_i}{1+\alpha\lambda_i}\right\}}$$

and together (2) and (3) is called the zero-truncated generalized Poisson regression model, whether β is a p-dimension regression coefficient, and $x_i^T = (x_{i1}, x_{i2}, x_{i3}, \ldots x_{ip})$, i=1, 2, 3 . . . n. The log-likelihood function of the ZTGP regression model based on a sample of n independent observations is expressed as

$$\log L = l(\alpha,\beta|y_i>0)$$
$$= \sum_{i=1}^{n}\left[y_i\left(\log\frac{\lambda_i}{(1+\alpha\lambda_i)}\right) + (y_i-1)\log(1+\alpha y_i) - \frac{\alpha\lambda_i y_i}{1+\alpha\lambda_i} - \log y_i! - \log\left(e^{\left(\frac{\lambda_i}{1+\alpha\lambda_i}\right)}-1\right)\right] \qquad (4)$$

The maximum likelihood estimators are obtained by differentiating the log-likelihood function (4) with respect to α and β, we get

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{n}\left[-\frac{y_i\lambda_i}{(1+\alpha\lambda_i)} + \frac{y_i(y_i-1)}{1+\alpha y_i} - \left\{(1+\alpha\lambda_i)y_i\lambda_i - \frac{(\alpha y_i\lambda_i)\lambda_i}{(1+\alpha\lambda_i)^2}\right\} - \frac{e^{\left(\frac{\lambda_i}{(1+\alpha y_i)}\right)}\left(-\frac{\lambda_i}{(1+\alpha\lambda_i)^2}\lambda_i\right)}{e^{\left(\frac{\lambda_i}{(1+\alpha y_i)}\right)}-1}\right]$$

$$= \sum_{i=1}^{n}\left[\frac{y_i(y_i-1)}{1+\alpha y_i} - \frac{y_i\lambda_i}{(1+\alpha\lambda_i)^2}\{(1+\alpha\lambda_i)+1+\alpha\lambda_i-\alpha\lambda_i\} + \frac{\frac{\lambda_i^2}{(1+\alpha\lambda_i)^2}e^{\left(\frac{\lambda_i}{(1+\alpha\lambda_i)}\right)}}{e^{\left(\frac{\lambda_i}{1+\alpha\lambda_i}\right)}-1}\right]$$

$$\frac{\partial l}{\partial\alpha} = \sum_{i=1}^{n}\left[\frac{y_i(y_i-1)}{1+\alpha y_i} - \frac{y_i\lambda_i(2+\alpha\lambda_i)}{(1+\alpha\lambda_i)^2} + \frac{\lambda_i^2\exp\left(\frac{\lambda_i}{(1+\alpha\lambda_i)}\right)}{\left(\exp\left(\frac{\lambda_i}{1+\alpha\lambda_i}\right)-1\right)(1+\alpha\lambda_i)^2}\right] \qquad (5)$$

$$\log\lambda_i = x_i^T\beta \quad \text{and}\ \lambda_i = e^{x_i^T\beta}$$

$$\frac{\partial l}{\partial\beta}\log\lambda_i = x_i \quad \text{and} \quad \frac{\partial l}{\partial\beta}\lambda_i = e^{x_i^T\beta}x_i$$

$$\frac{\partial l}{\partial\beta} = \sum_{i=1}^{n}\left[y_i\left(x_i - \frac{1}{1+\alpha y_i}\alpha e^{x_i^T\beta}x_i\right)\right] - \left\{\frac{(1+\alpha\lambda_i)\alpha y_i e^{x_i^T\beta}x_i - \alpha\lambda_i y_i\alpha e^{x_i^T\beta}x_i}{(1+\alpha\lambda_i)^2}\right\} - \frac{1}{e^{\frac{\lambda_i}{(1+\alpha\lambda_i)}}-1}e^{\frac{\lambda_i}{(1+\alpha\lambda_i)}}\left\{\frac{(1+\alpha\lambda_i)e^{x_i^T\beta}x_i - \alpha\lambda_i\alpha e^{x_i^T\beta}x_i}{(1+\alpha\lambda_i)^2}\right\}$$

$$\frac{\partial l}{\partial\beta} = \sum_{i=1}^{n}\left[y_i x_i - \frac{1}{1+\alpha y_i}\alpha e^{x_i^T\beta}\left\{\frac{1+(1+\alpha\lambda_i)-\alpha\lambda_i}{(1+\alpha\lambda_i)}\right\}x_i - \frac{e^{\frac{\lambda_i}{(1+\alpha\lambda_i)}}}{\left(e^{\frac{\lambda_i}{(1+\alpha\lambda_i)}}-1\right)(1+\alpha\lambda_i)^2}e^{x_i^T\beta}x_i(1+\alpha\lambda_i-\alpha\lambda_i)x_i\right]$$

$$\frac{\partial l}{\partial\beta} = \sum_{i=1}^{n}\left[y_i x_i - \frac{\alpha y_i e^{x_i^T\beta}(2+\alpha\lambda_i)}{(1+\alpha\lambda_i)^2}x_i - \frac{e^{\frac{\lambda_i}{(1+\alpha\lambda_i)}}e^{x_i^T\beta}}{\left(e^{\frac{\lambda_i}{(1+\alpha\lambda_i)}}-1\right)(1+\alpha\lambda_i)^2}x_i\right]$$

$$\frac{\partial l}{\partial\beta} = \sum_{i=1}^{n}\left[\frac{(1+\alpha\lambda_i)^2 y_i - \alpha y_i e^{x_i^T\beta}(2+\alpha\lambda_i)}{(1+\alpha\lambda_i)^2} - \frac{e^{\frac{\lambda_i}{(1+\alpha\lambda_i)}}\lambda_i}{\left(e^{\frac{\lambda_i}{(1+\alpha\lambda_i)}}-1\right)(1+\alpha\lambda_i)^2}\right]x_i$$

$$\frac{\partial l}{\partial\beta} = \sum_{i=1}^{n}\left[\frac{y_i}{(1+\alpha\lambda_i)^2} - \frac{\lambda_i\left(\frac{\lambda_i}{1+\alpha\lambda_i}\right)}{\left(e^{\left(\frac{\lambda_i}{1+\alpha\lambda_i}\right)}-1\right)(1+\alpha\lambda_i)^2}\right]x_i, \qquad (6)$$

4. Application of the model

Fitting of ZTGP model is done using C-section data. Data set consists of annual total births, hospital type (private and public) and C-sections. The response variable Y denotes the number of C-sections which do not have any zero values. There are 301 C-section births vary from hospital to hospital and also by type of hospital. Normally births by C-section are more frequent in private (fee paying) hospitals than public (non-fee paying) hospitals. But in the present data set births by C-section are less frequent in private hospitals as compared to public hospitals (Table 1(a)). Data about total annual births and the number of C-section carried out were considered from the records of 4 private hospitals and 16 public hospitals.

We first regress the response variable, 'C-sections' against one explanatory variable 'number of births', then we add one more explanatory variable in the form of indicator variable hospital type (public hospital=1, private hospital=0) in the Poisson regression analysis. The dispersion index (the ratio of variance to mean) is 4.62, and hence the data exhibit over-dispersion. Here we consider two explanatory variables, x_1 (births annual) and x_2(hospital type). We have considered zero-truncated generalized Poisson regression model because, the data (C-section) contains no zero values. The maximum likelihood estimators of ZTGP (α, λ_i) model are obtained as,

ANOVA					P-Value
	df	SS	MS	F	
Regression	1	7.143	7.143	24.190 ^{s}	0.000
Residual	18	5.315	0.295		
Total	19	12.458			

(a)

	Coefficients	SE	t -test	P-value	Lower 95%	Upper 95%
Intercept	1.800	0.184	9.781^s	0.000	1.413	2.187
Annual births	0.000623	0.000	4.918 ^{s}	0.000	0.000	0.001

(b)

Note: ^{s} indicates significant at 5% level of significance.

Table 2: Analysis of variance (ANOVA) with one explanatory variable

ANOVA					
	df	SS	MS	F	P-Value
Regression	2	10.042	5.021	35.331 ^{s}	0.000
Residual	17	2.416	0.142		
Total	19	12.458			

(a)

	Coefficients	SE	t -test	P-value	Lower 95%	Upper 95%
Intercept	1.172	0.189	6.206^{s}	0.000	0.773	1.570
Annual births	0.000386	0.000	3.769^{s}	0.002	0.000	0.001
Hospital type	1.109	0.245	4.517^{s}	0.000	0.591	1.627

(b)

Note: ^{s} indicates significant at 5% level of significance.

Table 3: Analysis of variance (ANOVA) with two explanatory variables

Therefore from the result we have,
$\log \lambda_i = (1.80+0.000623\text{births}) = 1.80+0.000623x_1$, $\hat{\beta} = (1.80, 0.000623)'$ and $\log \lambda_i = 1.172+0.000386x_1+1.109x_2$, $\hat{\beta} = (1.172, 0.000386, 1.109)'$ and then α is estimated using Newton-Raphson method. Therefore we consider

$$\frac{\partial l}{\partial \alpha} = g(\alpha) \quad \text{and} \quad \frac{\partial^2 l}{\partial \alpha^2} = g'(\alpha),$$

where

$$g(\alpha) = \sum_{i=1}^{n} \left[\frac{y_j(y_i-1)}{1+\alpha y_i} + \frac{y_i\lambda_i(2+\alpha\lambda_i)}{(1+\alpha y_i)^2} + \frac{\lambda_i^2 e^{\left(\frac{\lambda_i}{1+\alpha\lambda_i}\right)}}{\left(e^{\left(\frac{\lambda_i}{1+\alpha\lambda_i}\right)}-1\right)(1+\alpha y_i)^2} \right] \quad \text{from (5)}$$

$$g'(\alpha) = \sum_{i=1}^{n} \left[\frac{y_i\lambda_i^{2}(3+\alpha\lambda_i)}{(1+\alpha\lambda_i)^3} - \frac{y_i^{2}(y_i-1)}{(1+\alpha y_i)^2} - \lambda_i^2 e^{\left(\frac{\lambda_i}{1+\alpha\lambda_i}\right)} \frac{\left(1-\alpha^2\lambda_i^{2}\right)\left(e^{\left(\frac{\lambda_i}{1+\alpha\lambda_i}\right)}-1\right)-\lambda_i}{(1+\alpha y_i)^4\left(e^{\left(\frac{\lambda_i}{1+\alpha\lambda_i}\right)}-1\right)^2} \right] \tag{7}$$

Therefore

$$\alpha_{i+1} = \alpha_i - \left.\frac{g(\alpha)}{g'(\alpha)}\right|_{\alpha=\alpha_i} \tag{8}$$

$\hat{\alpha}$ is estimated, Therefore we get,
$\hat{\alpha}_1 = 0.034$ and $\hat{\alpha}_2 = 0.001242$

5. Goodness of fit test and tests of significance

A measure of goodness of fit of the ZTGP regression model is used on the log-likelihood statistic. The ZTGP regression model reduces to the ZTP regression model when the dispersion parameter α=0. To test for the adequacy of the ZTGP model over the ZTP regression model, we consider the testing of hypothesis H_0: α=0 vs H_1: α≠0. The addition of the dispersion parameter α in the regression model will be justified if H_0 is rejected. To test the null hypothesis H_0, we use the likelihood ratio statistic for the parameter α which is calculated after fitting the ZTGP regression model.

Likelihood Ratio Tests

We wish to test H_0: α =0 against H_1: α ≠0, given $\beta = (\beta_1, \beta_2 \ldots \beta_p)^T$ in the model. Now the likelihood ratio test is

$$\chi^2_\beta = -2\log\left(\frac{L(\hat{\beta})}{L(\hat{\alpha},\hat{\beta})}\right) = -2\log L(\hat{\beta}) - 2\log L(\hat{\alpha},\hat{\beta}) \quad (9)$$

Under H_0, $\chi_\beta{}^2 \sim \chi^2$ on p df

This is also called as the likelihood ratio model chi-square test.

Case 1: When $\hat{\alpha}_1 = 0.034$

$$\chi^2_\beta = -2\log\left(\frac{L(\hat{\beta})}{L(\hat{\alpha},\hat{\beta})}\right) = -2(-67.4658 + 63.1455) = 8.6406$$

Therefore $\chi^2_\beta = 8.6406 > \chi^2_1\ (5\%) = 3.84$

Therefore H_0 is rejected at 5% level of significance.

Thus the addition of dispersion parameter α in the regression model is justified.

Case 2: When $\hat{\alpha}_2 = 0.001242$

$$\chi^2_\beta = -2\log\left(\frac{L(\hat{\beta})}{L(\hat{\alpha},\hat{\beta})}\right) = -2(-53.4546 + 53.1053) = 0.0986$$

Therefore $\chi^2_\beta = 0.0986 < \chi^2_2\ (5\%) = 5.99$, H_0 is accepted at 5% level of significance.

Thus zero-truncated Poisson regression model gives better fit to the number of cesarean cases when two explanatory variables are used in the model.

References

1. Althabe F., Sosa C., Belizan J. M., Gibbons L., Jacquerioz F. and Bergal E. (2006). Cesarean section rates and maternal and neonatal mortality in low, medium and high income countries: an ecological study, Birth, 33(4), p. 270-7.
2. Betran A P. , Merialdi M. , Lauer J A. , Bing-Shun W. , Thomas J. , Van Look P. and Wagher M. (2007). Rates of cesarean section: analysis of global, regional and rational estimates, Paediatr Perinat Epidemiol, 21(2), p. 98-113.
3. Gibbon L., Belizan J. M., Lauer J. A., Betran A. P., Merialdi M. and Althabe F. (2010). The global numbers and costs of additionally needed and unnecessary cesarean section performed per year: overuse as a barrier to universal coverage, World Health Report: Background paper, No. 30, p. 1-19.
4. Gupta, P. L., Gupta, R. C and Tripati, R. C. (1996). Analysis of zero-adjusted count data, Computational Statistics and Data Analysis, 23, p. 207-218.
5. Hamilton B.E., Martin J.A., Ventura S.J., et al. (2013). Births: preliminary data for 2012, National Vital Statistics Reports, Vol. 61, No. 1, Hyattsville, MD: National Center for Health Statistics.
6. Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing, Technometrics, 34, p. 1-14.
7. Lee, A. H., Wang, K., Scott J. A., Yau, K. K. W. and Mclachlan, G. J. (2006). Multi-level zero-inflated Poisson regression modeling of correlated count data with excess zeros, Statistical Methods in Medical Research, 15, p. 47-61.
8. Lee, A.H., Wang, K. and Yau, K.K.W. (2001). Analysis of zero-inflated Poisson data incorporating extent of exposure, Biometrical Journal, 43, p. 963-975.
9. Martin JA, Hamilton BE, Ventura SJ, et al. (2013). Births:final data for 2011, National Vital Statistics Reports, Vol. 61, No.1. Hyattsville, MD: National Center for Health Statistics.

10. Ronasmans C., Holtz S. and Stanton C. (2006). Socio-economic differentials in cesarean rates in developing countries: a retrospective analysis, Laneet, 368 (9546), p. 1516-23.
11. Wang, W. and Famoye, F. (1997). Modelling household fertility decisions with fertilized Poisson regression, Journal of Population Economics, 10, p. 273-283.
12. Wulu, J. T., Singh, K. P. Famoye, F. and McGwin, G. (2002). Regression analysis of count data, Journal of the Indian Society of Agril. Statistics, 55, p. 22-231.
13. Zhao, W. Feng, Y. and Li, Z. (2010). Zero-truncated generalized Poisson regression model and its score tests, Journal of East China Normal University (Natural Science), No. 1, 17-23.