**Applied Mathematics, 2016, 7, 1589-1598**
Published Online August 2016 in SciRes. http://www.scirp.org/journal/am
http://dx.doi.org/10.4236/am.2016.714137

# Zero Truncated Bivariate Poisson Model: Marginal-Conditional Modeling Approach with an Application to Traffic Accident Data

**Rafiqul I. Chowdhury[1], M. Ataharul Islam[2]**

[1]ISRT, University of Dhaka, Dhaka, Bangladesh
[2]Department of Applied Statistics, East West University, Dhaka, Bangladesh
Email: chowdhuryri@yahoo.com, mataharul@yahoo.com, ataharul@ewubd.edu

Received 25 June 2016; accepted 22 August 2016; published 25 August 2016



## Abstract

**A new covariate dependent zero-truncated bivariate Poisson model is proposed in this paper employing generalized linear model. A marginal-conditional approach is used to show the bivariate model. The proposed model with estimation procedure and tests for goodness-of-fit and under (or over) dispersion are shown and applied to road safety data. Two correlated outcome variables considered in this study are number of cars involved in an accident and number of casualties for given number of cars.**

## Keywords

**Bivariate Poisson, Conditional Model, Generalized Linear Model, Marginal Model, Road Safety Data, Zero-Truncated**

## 1. Introduction

The count data analysis occupies an important role in applied statistics in various fields. When the observed outcomes are count and the desire is to estimate the covariate effects on outcomes, covariate dependent Bivariate Poisson (BVP) model is a tool of natural choice. It is expected that the observed outcomes on the same subject are be correlated. This type of data arises in many fields, for example, traffic accidents, health sciences, economics, social sciences, environmental studies among others. A typical example of such dependence arises in the number of traffic accidents and the number of injuries or fatalities during a specified period. However, in some situations outcomes may be truncated as zero values of counts may not be observed or may be missing for one or both of the outcomes. For example, in a sample drawn from hospital admission records, frequencies of

**How to cite this paper:** Chowdhury, R.I. and Islam, M.A. (2016) Zero Truncated Bivariate Poisson Model: Marginal-Conditional Modeling Approach with an Application to Traffic Accident Data. *Applied Mathematics*, **7**, 1589-1598.
http://dx.doi.org/10.4236/am.2016.714137

zero accidents and length of stay are not available. Another example is the case where the data on number of traffic accidents and related injuries or fatalities and related risk factors are collected from records and, naturally, zero counts are not available. As an example, road safety data from data.gov.uk website provides detailed information about the conditions of personal injury road accidents in Great Britain including the types of vehicles involved and the consequential casualties on public roads along with other background information. Only those accidents that involve personal injury reported to the police using the accident reporting form are recorded. Damage-only accidents, with no human casualties or accidents on private roads or car parks, are not included generating zero-truncated count data. To investigate the effect of risk factors on this type of outcomes, zero-truncated BVP regression is the appropriate model.

Campbell [1] introduced BVP distribution. Various assumptions have been used to develop BVP distribution. The most comprehensive one has been proposed by Kocherlakota and Kocherlakota [2]. Leiter and Hamdan [3] suggested bivariate probability models applicable to traffic accidents and fatalities. A similar problem was addressed by Cacoullos and Papageorgiou [4]. Several other attempts were made to define and study the BVP distribution [5]-[9]. Jung and Winkelmann [10] showed bivariate Poisson form using a trivariate reduction method allowing for correlation between the variables, which is considered as a nuisance parameter. This bivariate Poisson regression is used by others [11] [12]. Islam and Chowdhury [13] suggested covariate dependent BVP model using generalized linear modeling approach based on Leiter and Hamdan [3] bivariate probability models. They used marginal and conditional models to obtain BVP model.

Studies on the covariate dependent zero-truncated BVP model are scarce. Different techniques of the parameter estimation of BVP distribution are presented in [14]-[16]. A unified treatment of three types of zero-truncated BVP discrete distribution based on probability generating function is shown elsewhere [17]. Properties of BVP distribution truncated from below at an arbitrary point were studied by others [18] [19]. At this backdrop, we proposed a zero-truncated covariate dependent BVP model based on the work of Islam and Chowdhury [13]. The exposition of the following sections of the paper is as follows. Firstly in Section 2, we present briefly the marginal, conditional and BVP distribution for two outcomes without zero truncation as shown in [13]. In Section 3, we have shown the zero-truncated marginal and conditional Poisson distribution and obtained the joint model for both outcomes zero-truncated. The estimation and the related procedures are also shown. In Section 4, applications of the proposed models are illustrated using road safety data for both outcomes zero-truncated published by the Department for Transport, United Kingdom. Finally, concluding remarks can be found in Section 5.

## 2. Poisson Distribution without Zero Truncation

In this section bivariate Poisson model without zero truncation is shown. For simplicity, we shall follow the notations used in [13]. Let $Y_1$ be the number of accidents at a specific location in a given interval that has a Poisson distribution with density

$$g_1(y_1) = P(Y_1 = y_1) = \frac{e^{-\lambda_1}\lambda_1^{y_1}}{y_1!}, y_1 = 0,1,\cdots \tag{1}$$

and the corresponding link function is

$$\ln\lambda_1 = x'\beta_1, \text{ where } x' = (1, x_1, \cdots, x_p), \beta_1' = (\beta_{10}, \beta_{11}, \cdots, \beta_{1p}).$$

If $Y_{2i}$'s are assumed to be mutually independent, then the conditional distribution of $Y_2 = Y_{21} + \cdots + Y_{2y_1}$, the total number of fatalities recorded among the $Y_1$ accidents occurring in the jt-h time interval is Poisson with parameter $\lambda_2 y_1$. Then we can show that

$$g_2(y_2 \mid y_1) = \frac{e^{-\lambda_2 y_1}(\lambda_2 y_1)^{y_2}}{y_2!}, y_2 = 0,1,\cdots \tag{2}$$

and the corresponding link function is

$$\ln\lambda_2 = x'\beta_2, \text{ where } x' = (1, x_1, \cdots, x_p), \beta_2' = (\beta_{20}, \beta_{21}, \cdots, \beta_{2p}).$$

Then following [13] the joint distribution of number of accidents and number of fatalities can be shown as

$$g(y_1, y_2) = g_2(y_2 \mid y_1)\cdot g_1(y_1) = e^{-\lambda_1}\lambda_1^{y_1}e^{-\lambda_2 y_1}(\lambda_2 y_1)^{y_2} \big/ (y_1!\, y_2!). \tag{3}$$

## 3. Zero-Truncated Poisson Distribution

The probability of $Y_1 = 0$ is $e^{-\lambda_1}$, using Equation (1). Hence $Y_1$ is observed conditional on $Y_1 > 0$. Thus, we have the conditional probability mass function

$$P(Y_1 = y_1 \mid Y_1 > 0) = \frac{P(Y_1 = y_1)}{P(Y_1 > 0)} = \frac{P(Y_1 = y_1)}{1 - P(Y_1 = 0)}. \tag{4}$$

Now, using Equation (1) the zero-truncated Poisson probability mass function for $Y_1 \mid Y_1 > 0$ is

$$g_1^*(y_1) = P(Y_1 = y_1 \mid Y_1 > 0) = \frac{e^{-\lambda_1}\lambda_1^{y_1}}{y_1!} \times \frac{1}{\left(1 - e^{-\lambda_1}\right)} = \frac{\lambda_1^{y_1}}{y_1!\left(e^{\lambda_1} - 1\right)}. \tag{5}$$

Then the exponential form of the mass function is

$$g_1^*(y_1) = \exp\left[y_1 \ln \lambda_1 - \ln(y_1!) - \ln\left(e^{\lambda_1} - 1\right)\right]. \tag{6}$$

The mean and variance can be shown as

$$\mu_{Y_1} = E[Y_1 \mid Y_1 > 0] = \frac{\lambda_1 e^{\lambda_1}}{e^{\lambda_1} - 1} \quad \text{and} \quad \sigma_{Y_1}^2 = Var[Y_1 \mid Y_1 > 0] = \frac{\lambda_1 e^{\lambda_1}}{e^{\lambda_1} - 1}\left[1 - \frac{\lambda_1}{e^{\lambda_1} - 1}\right]. \tag{7}$$

Similarly, the zero-truncated conditional distribution of $Y_2 \mid y_1, Y_2 > 0$ is

$$P(Y_2 = y_2 \mid y_1, Y_2 > 0) = \frac{P(Y_2 = y_2 \mid y_1)}{P(Y_2 > 0 / y_1)} = \frac{P(Y_2 = y_2 \mid y_1)}{1 - P(Y_2 = 0 / y_1)}. \tag{8}$$

Then the zero-truncated conditional Poisson distribution is

$$g_2^*(Y_2 = y_2 \mid y_1, Y_2 > 0) = \frac{e^{-\lambda_2 y_1}(\lambda_2 y_1)^{y_2}}{y_2!} \times \frac{1}{\left(1 - e^{-\lambda_2 y_1}\right)} = \frac{(\lambda_2 y_1)^{y_2}}{y_2!\left(e^{\lambda_2 y_1} - 1\right)}. \tag{9}$$

The exponential form of Equation (9) can be shown as

$$g_2^*(Y_2 = y_2 \mid y_1, Y_2 > 0) = \exp\left[y_2 \ln \lambda_2 + y_2 \ln(y_1) - \ln(y_2!) - \ln\left(e^{\lambda_2 y_1} - 1\right)\right]. \tag{10}$$

Then the mean and variance are

$$\begin{aligned} \mu_{Y_2|Y_1} &= E[Y_2 \mid Y_1, Y_2 > 0] = \frac{\lambda_2 y_1 e^{\lambda_2 y_1}}{e^{\lambda_2 y_1} - 1} \text{ and} \\ \sigma_{Y_2|Y_1}^2 &= Var[Y_2 \mid Y_1, Y_2 > 0] = \frac{\lambda_2 y_1 e^{\lambda_2 y_1}}{e^{\lambda_2 y_1} - 1}\left[1 - \frac{\lambda_2 y_1}{e^{\lambda_2 y_1} - 1}\right]. \end{aligned} \tag{11}$$

### 3.1. Zero Truncated Bivariate Poisson (ZTBVP) Model

Now using the marginal and conditional distribution for zero truncation derived above the joint distribution of ZTBVP can be obtained as follows

$$g^*(y_1, y_2) = g_2^*(y_2 \mid y_1)\, g_1^*(y_1) = \frac{(\lambda_2 y_1)^{y_2}(\lambda_1)^{y_1}}{y_1!\,y_2!\left(e^{\lambda_2 y_1} - 1\right)\left(e^{\lambda_1} - 1\right)} \tag{12}$$

The ZTBVP expression in Equation (12) can be expressed in bivariate exponential form as

$$\begin{aligned} g^*(y_1, y_2) = \exp\Big[ & y_1 \ln \lambda_1 - \ln(y_1!) - \ln\left(e^{\lambda_1} - 1\right) + y_2 \ln \lambda_2 + y_2 \ln(y_1) \\ & - \ln(y_2!) - \ln\left(e^{\lambda_2 y_1} - 1\right)\Big], \end{aligned} \tag{13}$$

where the link functions are $\ln \lambda_1 = X'\beta_1$ and $\ln \lambda_2 = X'\beta_2$.

The log-likelihood function is

$$\ln L=\sum_{i=1}^{n}\Big[y_{1i}\left(x'_{ij}\beta_1\right)-\ln\left(y_{1i}!\right)-\ln\left(\mathrm{e}^{\mathrm{e}^{x'_{ij}\beta_1}}-1\right)+y_{2i}\left(x'_{ij}\beta_2\right)+y_{2i}\ln\left(y_{1i}\right)$$
$$-\ln\left(y_{2i}!\right)-\ln\left(\mathrm{e}^{y_{1i}\mathrm{e}^{x'_{ij}\beta_2}}-1\right)\Big]. \tag{14}$$

The estimating equations are

$$\frac{\partial\ln L}{\partial\beta_{1j}}=\sum_{i=1}^{n}\left[y_{1i}-\frac{\mathrm{e}^{x'_{ij}\beta_1}\mathrm{e}^{\mathrm{e}^{x'_{ij}\beta_1}}}{\mathrm{e}^{\mathrm{e}^{x'_{ij}\beta_1}}-1}\right]x_{ij}=0,\ j=0,1,\cdots,p; \tag{15}$$

and

$$\frac{\partial\ln L}{\partial\beta_{2j}}=\sum_{i=1}^{n}\left[y_{2i}-\frac{y_{1i}\mathrm{e}^{x'_{ij}\beta_2}\mathrm{e}^{y_{1i}\mathrm{e}^{x'_{ij}\beta_2}}}{\mathrm{e}^{y_{1i}\mathrm{e}^{x'_{ij}\beta_2}}-1}\right]x_{ij}=0,\ j=0,1,\cdots,p. \tag{16}$$

Then the score vector is

$$U\left(\beta_j\right)=\left[\frac{\partial\ln L}{\partial\beta_{1j}},\frac{\partial\ln L}{\partial\beta_{2j}}\right] \tag{17}$$

The second derivatives are:

$$\frac{\partial^2\ln L}{\partial\beta_{1j}\partial\beta_{1j'}}=-\sum_{i=1}^{n}\left[\frac{e^{x'_{ij}\beta_1}e^{e^{x'_{ij}\beta_1}}}{\left(\mathrm{e}^{\mathrm{e}^{x'_{ij}\beta_1}}-1\right)}+\frac{\left(\mathrm{e}^{x'_{ij}\beta_1}\right)^2\mathrm{e}^{\mathrm{e}^{x'_{ij}\beta_1}}}{\left(\mathrm{e}^{\mathrm{e}^{x'_{ij}\beta_1}}-1\right)}-\frac{\left(\mathrm{e}^{x'_{ij}\beta_1}\right)^2\left(\mathrm{e}^{\mathrm{e}^{x'_{ij}\beta_1}}\right)^2}{\left(\mathrm{e}^{\mathrm{e}^{x'_{ij}\beta_1}}-1\right)^2}\right]x_{ij}x_{ij'},\ j,j'=0,1,\cdots,p; \tag{18}$$

$$\frac{\partial^2\ln L}{\partial\beta_{2j}\partial\beta_{2j'}}=-\sum_{i=1}^{n}\left[\frac{y_{1i}\mathrm{e}^{x'_{ij}\beta_2}\mathrm{e}^{y_{1}\mathrm{e}^{x'_{ij}\beta_2}}}{\left(\mathrm{e}^{y_{1i}\mathrm{e}^{x'_{ij}\beta_1}}-1\right)}+\frac{y_{1i}^2\left(\mathrm{e}^{x'_{ij}\beta_2}\right)^2\mathrm{e}^{y_1\mathrm{e}^{x'_{ij}\beta_2}}}{\left(\mathrm{e}^{y_{1i}\mathrm{e}^{x'_{ij}\beta_1}}-1\right)}-\frac{y_{1i}^2\left(\mathrm{e}^{x'_{ij}\beta_2}\right)^2\left(\mathrm{e}^{y_1\mathrm{e}^{x'_{ij}\beta_2}}\right)^2}{\left(\mathrm{e}^{y_{1i}\mathrm{e}^{x'_{ij}\beta_1}}-1\right)}\right]x_{ij}x_{ij'},\ j,j'=0,1,\cdots,p. \tag{19}$$

The observed information matrix is

$$I_o\left(\beta_{jj'}\right)=\begin{bmatrix}\left(\dfrac{\partial^2\ln L}{\partial\beta_{1j}\partial\beta_{1j'}}\right)_{(p+1)(p+1)} & \mathbf{0}_{(p+1)(p+1)}\\ \mathbf{0}_{(p+1)(p+1)} & \left(\dfrac{\partial^2\ln L}{\partial\beta_{2j}\partial\beta_{2j'}}\right)_{(p+1)(p+1)}\end{bmatrix} \tag{20}$$

and the approximate variance-covariance matrix for $\hat{\beta}'=\left(\hat{\beta}'_1,\hat{\beta}'_2\right)$ is $\widehat{Var}\left(\hat{\beta}\right)=I_o^{-1}\left(\beta_{jj'}\right)$. The estimates of the regression parameters vectors $\beta_1$ and $\beta_2$ can be obtained iteratively by using Newton-Raphson method as follows

$$\hat{\beta}_t=\hat{\beta}_{t-1}+I_0^{-1}\left(\hat{\beta}_{t-1}\right)U\left(\hat{\beta}_{t-1}\right), \tag{21}$$

where $\hat{\beta}_t$ denotes the estimate at t-th iteration.

### 3.2. Test for Significance of Parameters

We can use the likelihood ratio tests for testing $H_0:\beta_1=0$ and $\beta_2=0$ model fit using full model and reduced model. The test statistic is asymptotically chi-square as follows

$$\chi^2_{2p}=-2\left[\text{log likelihood}\left(\text{reduced model}\right)-\text{log likelihood}\left(\text{full model}\right)\right]. \tag{22}$$

For independence, we can test the equality of zero-truncated bivariate models under independence. The independence model can be shown as $g^{**}(y_1, y_2) = g_1^*(y_1) \times g_2^*(y_2)$.

### 3.3. Deviance and Goodness of Fit

The deviance measures the difference in log-likelihood based on observed and fitted values. Let $\hat{\mu}_{Y_{1i}}$ and $\hat{\mu}_{Y_{2i}|Y_{1i}}$ are the estimates of $\mu_{Y_{1i}}$ and $\mu_{Y_{2i}|Y_{1i}}$ under the model of interest as shown before (Section 3.1) and $y_{1i}$ and $y_{2i}$ are the observed values under the saturated model. The deviance for zero-truncated bivariate Poisson, $D = 2\sum_{i=1}^{n}\left[l(y_i, y_i) - l(\mu_i; y_i)\right]$, where $l(.,.)$ represents log-likelihood functions, as follows:

$$\begin{aligned} l(y_i, y_i) = {} & y_{1i}\ln(y_{1i}) - \ln\left(e^{y_{1i}} - 1\right) - \ln(y_{1i}!) + y_{2i}\ln\left(\frac{y_{2i}}{y_{1i}}\right) \\ & + y_{2i}\ln(y_{1i}) - \ln\left(e^{y_{2i}} - 1\right) - \ln(y_{2i}!) \end{aligned} \tag{23}$$

and

$$\begin{aligned} l(\mu_i; y_i) = {} & y_{1i}\ln\left(\hat{\mu}_{Y_{1i}}\right) - \ln\left(e^{\hat{\mu}_{Y_{1i}}} - 1\right) - \ln(y_{1i}!) + y_{2i}\ln\left(\frac{\hat{\mu}_{Y_{2i}|Y_{1i}}}{\hat{\mu}_{Y_{1i}}}\right) \\ & + y_{2i}\ln(y_{1i}) - \ln\left(e^{\left(\frac{\hat{\mu}_{Y_{2i}|Y_{1i}}}{\hat{\mu}_{Y_{1i}}}\right)y_{1i}} - 1\right) - \ln(y_{2i}!). \end{aligned} \tag{24}$$

After some algebra we get the deviance as

$$\begin{aligned} D = 2\sum_{i=1}^{n}\Bigg[ & y_{1i}\ln\left(y_{1i}/\hat{\mu}_{Y_{1i}}\right) - \ln\left(\frac{e^{y_{1i}} - 1}{e^{\hat{\mu}_{Y_{1i}}} - 1}\right) + y_{2i}\ln\left(y_{2i}/\hat{\mu}_{Y_{2i}|Y_{1i}}\right) \\ & - y_{2i}\ln\left(y_{1i}/\hat{\mu}_{Y_{1i}}\right) - \ln\left(\frac{e^{y_{2i}} - 1}{e^{\frac{\hat{\mu}_{Y_{2i}|Y_{1i}}}{\hat{\mu}_{Y_{1i}}}y_{1i}} - 1}\right)\Bigg]. \end{aligned} \tag{25}$$

We can use following test for goodness-of-fit proposed by Islam and Chowdhury (2015).

$$T_1 = \sum_{y_1=0}^{k_1}\begin{pmatrix} \bar{y}_1 & - & \hat{\mu}_{y_1} \\ \bar{y}_2 \mid y_1 & - & \hat{\mu}_{y_2|y_1} \end{pmatrix}' \begin{pmatrix} \sigma_{Y_1}^2 & 0 \\ 0 & \sigma_{Y_2|Y_1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \bar{y}_1 & - & \hat{\mu}_{y_1} \\ \bar{y}_2 \mid y_1 & - & \hat{\mu}_{y_2|y_1} \end{pmatrix} \tag{26}$$

where, $\hat{\mu}_{y_1}, \hat{\mu}_{y_2|y_1}$ are estimates of $\mu_{Y_1}$ and $\mu_{Y_2|Y_1}$, $\hat{\sigma}_{Y_1}^2$ and $\hat{\sigma}_{Y_2|Y_1}^2$ are estimates of $\sigma_{Y_1}^2$ and $\sigma_{Y_2|Y_1}^2$ as defined in Equations (7) and (11), respectively. $T_1$ is distributed asymptotically as $\chi_{2g}^2$ where $g$ is the number of groups of observed values, $y_1, \cdots, y_{1g}$.

### 3.4. Test for Over or Underdispersion

The presence of overdispersion or underdispersion may influence the standard error of parameter estimates, hence, the significance level of the estimates. Test for the goodness of fit as shown in Equation (26) is modified to test the overdispersion or underdispersion. The method of moments estimator suggested by [20] is used to estimate the dispersion parameter, $\phi_r$, as shown below

$$\hat{\phi}_r = \frac{1}{n-p}\sum_{i=1}^{n}\left[\frac{\left(y_{ri} - \hat{\mu}_{ri}\right)^2}{V\left(\hat{\mu}_{ri}\right)}\right] = \frac{\chi_{r,n-p}^2}{n-p}, \quad r = 1, 2, \text{ where } V\left(\hat{\mu}_{ri}\right) = \hat{\mu}_{ri}. \tag{27}$$

Using the mean, variance and correction factor as shown in [21] for truncated marginal and conditional Poisson models for $y_{ri} = k_{r1}, k_{r1} + 1, \cdots$, we can define $\hat{\mu}_{ri} = \hat{\lambda}_{ri} + \hat{\delta}_{ri}$ and $\hat{V}\left(\hat{\mu}_{ri}\right) = \hat{\lambda}_{ri} - \hat{\delta}_{ri}\left(\hat{\mu}_{ri} - 1\right)$ where

$\hat{\delta}_{ri} = \hat{\mu}_{ri} - \hat{\lambda}_{ri} = \hat{\lambda}_{ri} \cdot \hat{\alpha}\left(k_r - 1, \hat{\lambda}_{ri}\right)$, $\hat{\alpha}\left(k_1, \lambda_{ri}\right) = \dfrac{h\left(k_1, \lambda_{ri}\right)}{1 - H\left(k_1, \lambda_{ri}\right)}$, $h\left(k_1, \lambda_{ri}\right) = P\left(Y_{ri} = k_r\right)$, $H\left(k_1, \lambda_{ri}\right) = P\left(Y_{ri} \le k_r\right)$

and then using these values we can estimate $\phi_r$.

Then the test for dispersion $T_2$ for zero-truncated bivariate Poisson regression model is:

$$T_2 = \sum_{y_1=0}^{k_1} \begin{pmatrix} \bar{y}_1 & - & \hat{\mu}_{y_1} \\ \bar{y}_2 \mid y_1 & - & \hat{\mu}_{y_2|y_1} \end{pmatrix}' \begin{pmatrix} \phi_1 \hat{\sigma}^2_{Y_1} & 0 \\ 0 & \phi_2 \hat{\sigma}^2_{Y_2|Y_1} \end{pmatrix}^{-1} \begin{pmatrix} \bar{y}_1 & - & \hat{\mu}_{y_1} \\ \bar{y}_2 \mid y_1 & - & \hat{\mu}_{y_2|y_1} \end{pmatrix} \tag{28}$$

where, $\hat{\mu}_{y_1}, \hat{\mu}_{y_2|y_1}, \hat{\sigma}^2_{Y_1}$ and $\hat{\sigma}^2_{Y_2|Y_1}$ are estimates of expected values and variances as defined in Equations (7) and (11) and $\phi_1$ and $\phi_2$ are dispersion parameters for $Y_1$ and $Y_2$, respectively. $T_2$, is also, distributed asymptotically as $\chi^2_{2g}$ where $g$ is the number of groups of observed values, $y_1, \cdots, y_{1g}$.

## 4. Application

The models proposed in the paper are illustrated using the road safety data published by Department for Transport, United Kingdom. This data set is publicly available for download from UK givernment website (http://data.gov.uk/dataset/road-accidents-safety-data). The data set includes information about the conditions of personal injury road accidents in Great Britain and the consequential casualties on public roads. Background information about vehicle types, location, road conditions, drivers demographics are also available among others. A total of 1,494,275 accident records were in the data set spanning from 2005 to 2013. We have selected a random sample 14005 accident records approximately 1 percent of all accident records. The outcome variables considered are total number of vehicles involved in the accident ($Y_1$) and the number of casualties ($Y_2$). Due to small frequencies, values five or more were coded as five for both outcomes. Risk factors are sex of the driver (0 = female; 1 = male), area (0 = urban; 1 = rural), two dummy variables for accident severity (fatal severity = 1, else 0; serious severity = 1, else = 0; slight severity is the reference category), light condition (daylight = 1; others = 0) and eight dummy variables for year 2006 to year 2013, where year 2005 is considered as reference category.

The average number of vehicles involved in accident and casualties are 1.83 and 1.37, with standard deviations 0.75 and 0.92, respectively. Table 1 displays the bivariate distribution of the number of vehicles and number of casualties. It is evident that 59 percent of the accidents involved two cars, 30 percent single car, and eight percent three cars. The number of casualties was one in three-fourth of the cases and two in one out of six cases. Descriptive statistics of the number of vehicles involved in accidents and number of casualties by risk factors are presented in Table 2. The mean number of vehicles with fatal injuries was 1.94 compared to 1.70 and 1.85 with serious and slight injuries. The mean number of casualties was 2.15 for fatal cases which appears to be much higher than that of serious and slight injuries. There is not much variation in mean number of vehicles and casualties by sex of driver and area. Although the number of vehicles involved in the accident is higher during daylight, number of casualties appear to be higher during other times. The number of vehicles involved in accidents decreased steadily during the study period, but mean number of cars involved in accidents and casualties remained almost similar.

**Table 1.** Number of vehicles involved in the accident ($Y_1$) and number of casualties ($Y_2$).

| Number of Vehicles ($Y_1$) | Number of Casualties ($Y_2$). | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5+ | Total |
| 1 | 3721 | 379 | 3 | 39 | 11 | 4225 |
| 2 | 6091 | 1561 | 75 | 122 | 89 | 8304 |
| 3 | 681 | 286 | 441 | 44 | 37 | 1182 |
| 4 | 93 | 64 | 134 | 22 | 13 | 225 |
| 5+ | 31 | 12 | 33 | 8 | 8 | 69 |
| Total | 10617 | 2302 | 693 | 235 | 158 | 14005 |

We observe that both numbers of vehicles involved in accidents and number of casualties are heavily underdispersed as displayed in **Table 4**. In **Table 3**, the estimates of the parameters are displayed along with standard errors and p-values for both original models as well as for adjustments made for underdispersion. Summary measures of goodness of fit for all the models are summarized in **Table 4**. The proposed full model of ZTBVP (**Table 3**) shows a negative association between fatal and serious severity and number of cars involved in accidents, while there is a positive association (p-value < 0.01) between the number of cars involved in an accident and light condition (daytime driving). The number of cars involved in accidents appears to be negatively associated in years 2008-2010 and 2012 as compared to that of 2005. However, the conditional model for the number of casualties given the number of cars involved in an accidents reveals that male drivers compared to females, rural areas compared to urban and daytime compared to night have lower risks. On the other hand, fatal severity and serious severity are positively associated with the number of casualties for given number of accidents compared to light severity. It is also evident that compared to the reference year, 2005, the number of casualties is negatively associated with the years 2012 and 2013. This indicates a significant reduction in the number of casualties for given number of accidents in recent years as compared to that of 2005.

**Table 2.** Descriptive statistics of the number of vehicles involved in the accident and the number of casualties by risk factors.

| | N | Number of Vehicles | | Number of Casualties | |
|---|---|---|---|---|---|
| Variables | | Mean | SD | Mean | SD |
| Sex of Driver | | | | | |
| Male | 9948 | 1.83 | 0.78 | 1.37 | 0.98 |
| Female | 4057 | 1.85 | 0.66 | 1.38 | 0.76 |
| Accident Severity | | | | | |
| Fatal | 173 | 1.94 | 2.63 | 2.15 | 4.01 |
| Serious | 1913 | 1.70 | 0.74 | 1.45 | 0.92 |
| Slight | 11919 | 1.85 | 0.68 | 1.35 | 0.79 |
| Area | | | | | |
| Urban | 5213 | 1.85 | 0.90 | 1.49 | 1.17 |
| Rural | 8792 | 1.82 | 0.64 | 1.30 | 0.72 |
| Light Condition | | | | | |
| Daylight | 10347 | 1.87 | 0.75 | 1.35 | 0.90 |
| Others | 3658 | 1.73 | 0.73 | 1.42 | 0.96 |
| Years | | | | | |
| 2005 | 1855 | 1.86 | 0.73 | 1.39 | 0.79 |
| 2006 | 1768 | 1.86 | 0.72 | 1.37 | 0.81 |
| 2007 | 1727 | 1.84 | 0.70 | 1.38 | 0.99 |
| 2008 | 1608 | 1.80 | 0.73 | 1.37 | 0.83 |
| 2009 | 1567 | 1.83 | 0.71 | 1.39 | 0.82 |
| 2010 | 1489 | 1.81 | 0.63 | 1.38 | 0.78 |
| 2011 | 1368 | 1.86 | 1.10 | 1.40 | 1.57 |
| 2012 | 1357 | 1.82 | 0.68 | 1.32 | 0.73 |
| 2013 | 1266 | 1.83 | 0.67 | 1.31 | 0.75 |

**Table 3.** Parameter estimates of zero truncated BVP model.

| Variables | Estimate | S.E. | p-value | $\sqrt{\varphi_r V(\hat{\beta})}$ | p-value |
|---|---|---|---|---|---|
| Y1:Constant | 0.280 | 0.034 | 0.000 | 0.017 | 0.000 |
| Sex of Driver | −0.017 | 0.019 | 0.355 | 0.009 | 0.066 |
| Area | −0.030 | 0.018 | 0.091 | 0.009 | 0.001 |
| Fatal severity | −0.101 | 0.082 | 0.218 | 0.041 | 0.014 |
| Serious severity | −0.166 | 0.027 | 0.000 | 0.014 | 0.000 |
| Light Condition | 0.140 | 0.021 | 0.000 | 0.010 | 0.000 |
| Year 2006 | −0.001 | 0.033 | 0.980 | 0.017 | 0.959 |
| Year 2007 | −0.014 | 0.034 | 0.666 | 0.017 | 0.390 |
| Year 2008 | −0.060 | 0.035 | 0.083 | 0.017 | 0.001 |
| Year 2009 | −0.034 | 0.035 | 0.320 | 0.017 | 0.047 |
| Year 2010 | −0.047 | 0.035 | 0.187 | 0.018 | 0.009 |
| Year 2011 | −0.021 | 0.036 | 0.565 | 0.018 | 0.252 |
| Year 2012 | −0.042 | 0.036 | 0.248 | 0.018 | 0.021 |
| Year 2013 | −0.023 | 0.037 | 0.526 | 0.018 | 0.207 |
| Y2:Constant | −0.637 | 0.049 | 0.000 | 0.029 | 0.000 |
| Sex of Driver | −0.058 | 0.029 | 0.049 | 0.018 | 0.001 |
| Area | −0.375 | 0.027 | 0.000 | 0.016 | 0.000 |
| Fatal severity | 0.654 | 0.080 | 0.000 | 0.048 | 0.000 |
| Serious severity | 0.266 | 0.036 | 0.000 | 0.022 | 0.000 |
| Light Condition | −0.231 | 0.029 | 0.000 | 0.018 | 0.000 |
| Year 2006 | −0.042 | 0.051 | 0.415 | 0.031 | 0.175 |
| Year 2007 | −0.051 | 0.052 | 0.326 | 0.031 | 0.102 |
| Year 2008 | −0.034 | 0.053 | 0.519 | 0.032 | 0.283 |
| Year 2009 | 0.029 | 0.052 | 0.579 | 0.031 | 0.356 |
| Year 2010 | 0.017 | 0.054 | 0.748 | 0.032 | 0.593 |
| Year 2011 | −0.030 | 0.055 | 0.590 | 0.033 | 0.370 |
| Year 2012 | −0.151 | 0.058 | 0.009 | 0.035 | 0.000 |
| Year 2013 | −0.186 | 0.060 | 0.002 | 0.036 | 0.000 |

The summary results of estimation and tests of different models (proposed model based on marginal-conditional approach and both marginal models) are presented in **Table 4**. Both the full model and the reduced model under null hypothesis are considered. Both the models indicate that the full models are statistically significant. It is noteworthy that both the outcome variables number of vehicles involved in accidents and number of casualties are substantially underdispersed and adjustments were made accordingly for underdispersion in **Table 3**. Based on AIC, BIC and deviance we observe that the proposed full model using marginal-conditional approach provides the best fit. The goodness of fit test using the test statistic, $T_1$, indicates good fit marginally (p-value = 0.064) for the proposed model. The test for under dispersion reveals the presence of significant deviation from equidispersion in both the variables as observed from $T_2$ (p-value < 0.001). Adjustments are made for underdispersion and the results are shown in **Table 3** (last two columns).

**Table 4.** Test statistics results for reduced and full models of ZTBVP.

| Model Statistics | Reduced Model | Full Model |
|---|---|---|
| Marginal/Conditional | | |
| Log likelihood | −26708.6 | −26453.01 |
| AIC | 53421.1 | 52962.02 |
| BIC | 53433.7 | 52922.61 |
| Deviance | 10593.89 | 10465.07 |
| $T_1$(D.F, p-value) | 17.45(10, 0.065) | 17.48(10, 0.064) |
| $T_2$(D.F, p-value) | 68.45(10, 0.000) | 69.35(10, 0.000) |
| $\phi_1$ | 0.255 | 0.252 |
| $\phi_2$ | 0.377 | 0.361 |
| LR $\chi^2$ Reduced vs. Full Model (D. F, p-value) | | 511.1(26, 0.000) |
| Marginal/Marginal | | |
| Log likelihood | −27235.59 | −26999.44 |
| AIC | 54475.20 | 54054.90 |
| BIC | 54490.28 | 54266.21 |
| Deviance | 11584.13 | 11322.42 |
| $T_1$(D.F, p-value) | 18.48(10, 0.048) | 19.01(10, 0.040) |
| $T_2$(D.F, p-value) | 71.21(10, 0.000) | 73.56(10, 0.000) |
| $\phi_1$ | 0.255 | 0.252 |
| $\phi_2$ | 0.372 | 0.363 |
| LR $\chi^2$ Reduced vs. Full Model (D. F, p-value) | | 1563.7(26, 0.000) |

## 5. Conclusion

A zero-truncated bivariate generalized linear model for count data is proposed in this paper. This model is based on the bivariate model using marginal-conditional models proposed by Islam and Chowdhury (2015) for count data. Covariate dependent bivariate generalized linear model is shown, and canonical link functions are used to estimate the parameters of the Poisson distribution. The usefulness of the proposed model is demonstrated using road safety data published by Department for Transport, United Kingdom. The proposed ZTBVP model can easily accommodate a varying number of covariates for two outcomes. The joint distribution degenerates into a marginal and conditional distribution that makes estimation problem easier.

## Acknowledgements

We acknowledge gratefully that the study is supported by the HEQEP sub-project 3293, University Grants Commission of Bangladesh and the World Bank. This data set was obtained from Police reported road accident statistics (STATS19) Department for Transport (http://data.gov.uk/dataset/road-accidents-safety-data).

## References

[1] Campbell, J.T. (1934) The Poisson Correlation Function. *Proceedings of the Edinburgh Mathematical Society*, **2**, 18-26. http://dx.doi.org/10.1017/S0013091500024135

[2] Kocherlakota, S. and Kocherlakota, K. (1992) Bivariate Discrete Distributions. Marcel Dekker, New York.

[3] Leiter, R.E. and Hamdan, M.A. (1973) Some Bivariate Probability Models Applicable to Traffic Accidents and Fatalities. *International Statistical Review*, **41**, 87-100. http://dx.doi.org/10.2307/1402790

[4] Cacoullos, T. and Papageorgiou, H. (1980) On Some Bivariate Probability Models Applicable to Traffic Accidents and Fatalities. *International Statistical Review*, **48**, 345-356. http://dx.doi.org/10.2307/1402946

[5] Consul, P.C. and Jain, G.C. (1973) A Generalization of the Poisson Distribution. *Technometrics*, **15**, 791-799. http://dx.doi.org/10.1080/00401706.1973.10489112

[6] Consul P.C. (1994) Some Bivariate Families of Lagrangian Probability Distributions. *Communications in Statistics—Theory and Methods*, **23**, 2895-2906. http://dx.doi.org/10.1080/03610929408831423

[7] Consul, P.C. (1989) Generalized Poisson Distributions: Properties and Applications. Marcel Dekker, New York.

[8] Consul, P.C. and Shoukri, M.M. (1985) The Generalized Poisson Distribution When the Sample Mean Is Larger than the Sample Variance. *Communications in Statistics—Simulation and Computation*, **14**, 1533-1547. http://dx.doi.org/10.1080/03610918508812463

[9] Holgate, P. (1964) Estimation for the Bivariate Poisson Distribution. *Biometrika*, **51**, 241-245. http://dx.doi.org/10.1093/biomet/51.1-2.241

[10] Jung, R. and Winkelmann, R. (1993) Two Aspects of Labor Mobility: A Bivariate Poisson Regression Approach. *Empirical Economics*, **18**, 543-556.

[11] Karlis, D. and Ntzoufras, I. (2003) Analysis of Sports Data by Using Bivariate Poisson Models. *Journal of the Royal Statistical Society Series D* (*The Statistician*), **52**, 381-393. http://dx.doi.org/10.1111/1467-9884.00366

[12] Karlis, D. and Ntzoufras, I. (2010) Bivariate Poisson and Diagonal Inflated Bivariate Poisson Regression Models in R. *Journal of Statistical Software*, **14**, 1-36.

[13] Islam, M.A. and Chowdhury, R.I. (2015) A Bivariate Poisson Models with Covariate Dependence. *Bulletin of Calcutta Mathematical Society*, **107**, 11-20.

[14] Hamdan, M.A. (1972) Estimation in the Truncated Bivariate Poisson Distribution. *Technometrics*, **14**, 37-45. http://dx.doi.org/10.1080/00401706.1972.10488881

[15] Dahiya, R.C. (1977) Estimation in a Truncated Bivariate Poisson Distribution. *Communications in Statistics—Theory and Methods*, **6**, 113-120. http://dx.doi.org/10.1080/03610927708827476

[16] Charambids, Ch.A. (1984) Minimum Variance Unbiased Estimation for Zero Class Truncated Bivariate Poisson and Logarithmic Series Distributions. *Metrika*, **31**, 115-123. http://dx.doi.org/10.1007/BF01915193

[17] Piperigou, V.E. and Papageorgiou, H. (2003) On truncated Bivariate Discrete Distributions: A Unified Treatment. *Metrika*, **58**, 221-233. http://dx.doi.org/10.1007/s001840200239

[18] Patil, S.A., Patel, D.I. and Kovner, J.L. (1977) On Bivariate Truncated Poisson Distribution. *Journal of Statistical Computation and Simulation*, **6**, 49-66. http://dx.doi.org/10.1080/00949657708810167

[19] Deshmukh, S.R. and Kasture, M.S. (2002) Bivariate Distribution with Truncated Poisson Marginal Distributions. *Communications in Statistics—Theory and Methods*, **31**, 527-534. http://dx.doi.org/10.1081/STA-120003132

[20] McCullagh, P. and Nelder, J.A. (1989) Generalized Linear Models. 2nd Edition, Chapman and Hall/CRC, Washington, DC. http://dx.doi.org/10.1007/978-1-4899-3242-6

[21] Gurmu, S. and Trivedi, P.K. (1992) Overdispersion Tests for Truncated Poisson Regression Models. *Journal of Econometrics*, **54**, 347-370. http://dx.doi.org/10.1016/0304-4076(92)90113-6