bpglm: R package for Bivariate Poisson GLM with Covariates

R. I. Chowdhury & M. A. Islam

2019-05-29

##

... Draft ...

1. Introduction

There is a growing interest in analyzing bivariate count outcomes with covariate dependence. The bivariate Poisson models have emerged to address a wide range of applications in various fields where paired count data are correlated. Leiter and Hamadan [1] suggested bivariate probability models applicable to traffic accidents and fatalities. The bivariate Poisson distribution has been proposed using various assumptions. Among those, the most comprehensive one has been developed by Kocherlakota and Kocherlakota [2]. Islam and Chowdhury [3], Chowdhury and Islam [4] and Islam and Chowdhury [5] developed untruncated and zero-truncated and right truncated bivariate Poisson model for covariate dependence based on the extended generalized linear model. The package (bpglm) is developed to fit those models. This package could also be used to fit model for multivariate outcomes and theory is presented in the book on *Analysis of repeated measures data* [6]. One can fit full model with varying number of covariates and likelihood ratio test could be performed for model selection. Also, Voung test [7] is available for model comparison. Full text of these papers can be downloaded from https://www.researchgate.net/profile/Rafiqul_Chowdhury3.

2. The Bivariate Poisson-Poisson Model

Un-truncated

The number of occurrences of the first event Y_1 in a given interval follow Poisson distribution with parameter λ_1 and the occurece of the second event, Y_2, for given Y_1, is also Poisson with parameter, $\lambda_1 y_1$. Details along with the link functions can be found in Islam and Chowdhury [3]. The joint pdf of Y_1 and Y_2 is:

$$g(y_1, y_2) = \frac{e^{-\lambda_1}\lambda_1^{y_1} e^{-\lambda_2 y_1}(\lambda_2 y_1)^{y_2}}{y_1! y_2!},\ y_1 = 0, 1, ...;\ y_2 = 0, 1, ...;\ \lambda_1,\ \lambda_2 > 0.$$

Zero-truncated

In some situations zero values of count outcomes may not be observed that generates zero truncated count outcomes. For details please see [4]. The joint distribution of ZTBVP is:

$$g^*(y_1, y_2) = \frac{\lambda_1^{y_1}(\lambda_2 y_1)^{y_2}}{y_1! y_2! (e^{-\lambda_2 y_1} - 1)(e^{-\lambda_1} - 1)},\ y_1 = 0, 1, ...;\ y_2 = 0, 1, ...;\ \lambda_1,\ \lambda_2 > 0.$$

Right-truncated

The joint distribution of RTBVP developed by Islam and Chowdhury [5] is:

$$g(y_1, y_2) = c_1 c_2 e^{-\lambda_1}\lambda_1^{y_1} e^{-\lambda_2 y_1}(\lambda_2 y_1)^{y_2}/(y_1! y_2!).$$

3. The package

This package is on github, and one can install using following code. Load the 'bpglm' library asusual.

```
## install.packages("devtools")
## devtools::install_github("chowdhuryri/bpglm")
```

The main function to fit a model is *bpglm* with following arguments:

function (Y1Y2, X1 = NULL, X2 = NULL, mxit = 150, icob = NULL, mtype = 1, ppy = 1, presc = 1e+05, phi = NULL)

Where, Y1Y2 is a data frmae with two outcomes, X1 is covariates without constant term. Maximum number of default itarion is set to mxit=150. Starting values for beta's icob by default is NULL, but could be supplied incase of convergence problem.

4. Example Data Set One

Fisrt data set comes from Health and Retirement Study [8]. First two columns are two outcomes (r10conde = number of conditions and utiliza10 = number of health care services utilizations). help(exdata) would display the data and variables descriptions or str(exdata) would provide details. Following code chunks load the library and displays first few rows and create a bivariate table between two comes.

```
library(bpglm)
str(exdata)
```

```
## 'data.frame':    5567 obs. of  6 variables:
##  $ r10conde : int  1 3 0 3 4 3 2 2 2 2 ...
##  $ utiliza10: int  2 1 0 1 0 1 0 1 0 0 ...
##  $ Gender   : int  1 0 1 1 1 1 0 0 0 0 ...
##  $ Age      : int  74 72 71 70 72 74 68 75 73 73 ...
##  $ Hispanic : int  0 0 0 0 0 0 0 0 1 1 ...
##  $ Veteran  : int  0 0 0 1 0 1 0 0 0 0 ...
```

```
head(exdata)
```

```
##   r10conde utiliza10 Gender Age Hispanic Veteran
## 1        1         2      1  74        0       0
## 2        3         1      0  72        0       0
## 3        0         0      1  71        0       0
## 4        3         1      1  70        0       1
## 5        4         0      1  72        0       0
## 6        3         1      1  74        0       1
```

```
table(exdata[,1:2])
```

```
##         utiliza10
## r10conde   0   1   2   3   4
##        0 301   0   0   0   0
##        1 396 359 122  32  16
##        2 597 655 191  63  13
##        3 595 535 191  50  14
##        4 390 279 122  31  17
##        5 206 113  57  14   6
##        6 106  64  22   7   3
```

Fitting Un-truncated model

We want to fit the bivariate Poisson model with constant only. The main is bpglm() with minimum two arguments to fit the model.

Reduced model (constant only model)

Following r code fit the un-truncated model.

```
mod1<-bpglm(exdata[,1:2],mtype=2)
```

```
## ----- Bi-variate Poisson Regression ------
## ----- Constant only model.....No covariates -----
## Iteration =  1
## Log Likelihood =  -26941.5
## Iteration =  2
## Log Likelihood =  -21320.13
## Iteration =  3
## Log Likelihood =  -16968.53
## Iteration =  4
## Log Likelihood =  -16709.2
## Iteration =  5
## Log Likelihood =  -16707.66
## Iteration =  6
## Log Likelihood =  -16707.66
##
## -------- BIVP DEPENDENT (Marginal/Conditional) MODEL--------------------
##
## Log Likelihood = -16707.66
## AIC = 33419.33
## AICC = 33419.33
## BIC = 33432.58
## Deviance = 11926.93
## Phi1 =  0.7791603 Phi2 =  1.029702
##
## Iteration =  6
## Parameter Estimates
## -----------------------------------------------------------------------
##    Var.Names     Coeff      s.e   t.value p.value  Adj.s.e Adj.p.value
##  Y1:Constant  0.968676 0.008257 117.30977       0 0.007289           0
##  Y2:Constant -1.231112 0.015282 -80.56025       0 0.015507           0
## -----------------------------------------------------------------------
##
## Overdispersion: Z-test for y1 & y2 =
##                        Z_value p_value
## Z(Y1) Marginal 2-tail 26.74441 0.00000
## Z(Y2) Marginal 2-tail -1.94359 0.05195
##
## Good-ness-of-fit (T1) Overdispersion (T2):
##                      ChiSquare DF   p_value
## T1 (GOF)              8.396505 12 0.7534283
## T2(Dispersion Test) 10.759581 12 0.5496273
##
## Pearson Chisquare GOF using predicted probability
##    Chi.square D.F p.value
```

```
##  1.097286e-28  33       1
##
## [1] "Function Converged....."
```

The first argument of *bpglm* function is a data frame with only two outcomes which is first two columns in exdata and mtype=2 for untruncated model. The ouput shows the loglikelihood value for each itaration and the function converged after six itarations. Next it shows the detail model statistics (eg., AIC, BIC, etc.). Parameter estimates table shows the coefficients, standard error, t-value, p-value adjusted standard error and adjusted p-values.

Z-test for overdispersion is univariate test. T1 and T2 are multivariate test for good-ness-of-fit and overdispersion. Pearson Ch-square is also good-ness-of-fit test based on predicted observed outcomes.

Full model (constant & covariates)

Following r code fit the un-truncated model with four covariates.

```r
mod2<-bpglm(exdata[,1:2],exdata[,3:6],exdata[,3:6],mtype=2)
```

```
## ----- Bi-variate Poisson Regression -----
## ----- Model with covariates --------------------
## Iteration =  1
## Log Likelihood =  -35415.73
## Iteration =  2
## Log Likelihood =  -20745.24
## Iteration =  3
## Log Likelihood =  -17129.88
## Iteration =  4
## Log Likelihood =  -16609.21
## Iteration =  5
## Log Likelihood =  -16586.16
## Iteration =  6
## Log Likelihood =  -16586.07
## Iteration =  7
## Log Likelihood =  -16586.07
##
## -------- BIVP DEPENDENT (Marginal/Conditional) MODEL--------------------
##
## Log Likelihood = -16586.07
## AIC = 33192.13
## AICC = 33192.21
## BIC = 33258.38
## Deviance = 11683.74
## Phi1 =  0.7750906 Phi2 =  1.047914
##
## Iteration =  7
## Parameter Estimates
## ----------------------------------------------------------------------
##    Var.Names     Coeff      s.e   t.value  p.value  Adj.s.e Adj.p.value
##  Y1:Constant -0.054824 0.195344 -0.280651 0.778988 0.171979    0.749906
##       Gender -0.055173 0.021441 -2.573266 0.010100 0.018876    0.003482
##          Age  0.014074 0.002648  5.315410 0.000000 0.002331    0.000000
##     Hispanic  0.006742 0.028811  0.234021 0.814977 0.025365    0.790392
##      Veteran  0.048471 0.025023  1.937009 0.052795 0.022030    0.027836
##  Y2:Constant  0.264682 0.362681  0.729794 0.465547 0.371268    0.475929
##       Gender  0.345119 0.038546  8.953429 0.000000 0.039459    0.000000
```

```
##          Age -0.022722 0.004931 -4.608066 0.000004 0.005048    0.000007
##     Hispanic -0.174334 0.058160 -2.997495 0.002734 0.059537    0.003424
##      Veteran  0.093330 0.042272  2.207823 0.027297 0.043273    0.031067
## ----------------------------------------------------------------------
##
## Overdispersion: Z-test for y1 & y2 =
##                        Z_value p_value
## Z(Y1) Marginal 2-tail 26.74441 0.00000
## Z(Y2) Marginal 2-tail -1.94359 0.05195
##
## Good-ness-of-fit (T1) Overdispersion (T2):
##                     ChiSquare DF   p_value
## T1 (GOF)             8.227958 12 0.7670721
## T2(Dispersion Test) 10.588882 12 0.5644405
##
## Pearson Chisquare GOF using predicted probability
##  Chi.square D.F   p.value
##    24.61843  25 0.4839127
##
## [1] "Function Converged....."
```

The log likelihood value of full model is lower than reduced (constant only) mode. All statistics (AIC,BIC, etc.) are also lower for full model. Hence full model should used. We can use likelihood ratio test between reduced and full model. Following code does that.

```r
ChiRF(mod1,mod2)
```

```
## Likelihood Ratio Test: Reduced Model vs. Full Model
## ----------------------------------------------------------
##  ChiSquare DF       p_value
##  243.19178  8 4.7746315e-48
## ----------------------------------------------------------
```

We can also use Voung test for model comprison as follows:

```r
vountest(mod1, mod2)
```

```
## Positive mi>0 =  2474
## Negative mi<0 =  3093
```

```
##                         Z_value p  q      p_value
## V1 Un adj. one-tail -7.071175 2 10 7.681365e-13
## V1 Adj. one-tail    -5.064984 2 10 2.042171e-07
```

Negative z-value suggests mod2 (full model) is better while positive z-value favours reduced model. Both likelihood ratio test and Voung test suggest that full model should be used.

All results are stored in mod1 which is an R object. Following codes shows what statistics are stored in mod1 object and how to extract results and send to a CSV file which could be open in MS Excel for further manipulation.

```r
names(mod2)
```

```
##  [1] "coeff"      "ithetainv"  "ith"        "utheta"     "logLik"
##  [6] "N"          "nvar1"      "nvar2"      "nvar"       "AIC"
## [11] "BIC"        "AICC"       "Lambdas"    "Phi1"       "Phi2"
## [16] "mtype"      "convg"      "Control"    "GOF.Chi"    "Disp.Ztest"
## [21] "T1"         "Deviance"   "nparam"     "y1"         "y2"
## [26] "logliky1"   "logliky2"
```

```r
mod2$coeff
```

```
##      Var.Names     Coeff      s.e   t.value  p.value  Adj.s.e Adj.p.value
## 1  Y1:Constant -0.054824 0.195344 -0.280651 0.778988 0.171979    0.749906
## 2       Gender -0.055173 0.021441 -2.573266 0.010100 0.018876    0.003482
## 3          Age  0.014074 0.002648  5.315410 0.000000 0.002331    0.000000
## 4     Hispanic  0.006742 0.028811  0.234021 0.814977 0.025365    0.790392
## 5      Veteran  0.048471 0.025023  1.937009 0.052795 0.022030    0.027836
## 6  Y2:Constant  0.264682 0.362681  0.729794 0.465547 0.371268    0.475929
## 7       Gender  0.345119 0.038546  8.953429 0.000000 0.039459    0.000000
## 8          Age -0.022722 0.004931 -4.608066 0.000004 0.005048    0.000007
## 9     Hispanic -0.174334 0.058160 -2.997495 0.002734 0.059537    0.003424
## 10     Veteran  0.093330 0.042272  2.207823 0.027297 0.043273    0.031067
```

Example Data Set Two

The second data set is on road safety published by Department for Transport, United Kingdom. This data set is publicly available for download from http://data.gov.uk/dataset/road-accidents-safety-data. The data comprises the information about the conditions of personal injury road accidents in Great Britain and the consequential casualties on public roads.

```r
library(bpglm)
str(ukdata)
```

```
## 'data.frame':    14005 obs. of  7 variables:
##  $ NV       : num  2 2 2 1 2 2 2 1 2 1 ...
##  $ NCAUS    : num  2 1 1 1 1 1 1 2 1 1 ...
##  $ SexDriver: int  1 1 1 1 0 1 0 1 1 1 ...
##  $ Area     : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ SevFatal : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ SevSerius: int  0 0 0 0 0 1 0 0 0 0 ...
##  $ LightCon : int  0 1 0 1 1 1 1 1 1 0 ...
```

```r
head(ukdata)
```

```
##   NV NCAUS SexDriver Area SevFatal SevSerius LightCon
## 1  2     2         1    1        0         0        0
## 2  2     1         1    1        0         0        1
## 3  2     1         1    1        0         0        0
## 4  1     1         1    1        0         0        1
## 5  2     1         0    1        0         0        1
## 6  2     1         1    1        0         1        1
```

```r
table(ukdata[,1:2])
```

```
##    NCAUS
## NV     1    2    3    4
##   1 3721  379   75   50
##   2 6091 1561  441  211
##   3  681  286  134   81
##   4  124   76   43   51
```

Fitting zero-truncated model

We want to fit the bivariate zero-truncated Poisson model with constant only.

Reduced model (constant only model)

Following r code fit the the reduced model. Output is omited.

```
mod3<-bpglm(ukdata[,1:2],mtype=4)
```

Full model (constant & covariates)

Following r code fit the fullmodel with four covariates.

```
mod4<-bpglm(ukdata[,1:2],ukdata[,3:7],ukdata[,3:7],mtype=4)
```

```
## ----- Bi-variate Poisson Regression -----
## ----- Model with covariates --------------------
## Iteration =  1
## Log Likelihood =  -32210.95
## Iteration =  2
## Log Likelihood =  -27101.57
## Iteration =  3
## Log Likelihood =  -26175.98
## Iteration =  4
## Log Likelihood =  -26105.39
## Iteration =  5
## Log Likelihood =  -26104.59
## Iteration =  6
## Log Likelihood =  -26104.59
##
## -------- ZTBIVP DEPENDENT (Marginal/Conditional) MODEL---------
##
## Log Likelihood = -26104.59
## AIC = 52233.17
## AICC = 52233.21
## BIC = 52323.74
## Deviance = 9780.859
## Phi1 =  0.494567 Phi2 =  4.570604
##
## Iteration =  6
## Parameter Estimates
## ----------------------------------------------------------------------
##    Var.Names     Coeff      s.e    t.value  p.value  Adj.s.e Adj.p.value
##  Y1:Constant  0.243605 0.025718   9.472145 0.000000 0.018086    0.000000
##    SexDriver -0.016441 0.018882  -0.870722 0.383921 0.013279    0.215688
##         Area -0.024700 0.017813  -1.386600 0.165586 0.012527    0.048665
##     SevFatal -0.119248 0.083567  -1.426978 0.153608 0.058769    0.042466
##    SevSerius -0.169469 0.027325  -6.202018 0.000000 0.019216    0.000000
##     LightCon  0.143166 0.020631   6.939264 0.000000 0.014509    0.000000
##  Y2:Constant -0.701748 0.037220 -18.854174 0.000000 0.079572    0.000000
##    SexDriver -0.061683 0.029831  -2.067738 0.038683 0.063776    0.333469
##         Area -0.374765 0.027508 -13.623839 0.000000 0.058809    0.000000
##     SevFatal  0.623478 0.083737   7.445679 0.000000 0.179021    0.000498
##    SevSerius  0.251723 0.036869   6.827414 0.000000 0.078823    0.001409
##     LightCon -0.227105 0.029961  -7.580069 0.000000 0.064053    0.000393
## ----------------------------------------------------------------------
##
## Overdispersion: Z-test for y1 & y2 =
##                        Z_value p_value
```

```
## Z(Y1) Marginal 2-tail 121.3687       0
## Z(Y2) Marginal 2-tail  87.0016       0
##
## Good-ness-of-fit (T1) Overdispersion (T2):
##                     ChiSquare DF      p_value
## T1 (GOF)             7.015762  8 0.5349330749
## T2(Dispersion Test) 29.396459  8 0.0002700956
##
## Pearson Chisquare GOF using predicted probability
##  Chi.square D.F      p.value
##    26.77738   4 2.204889e-05
##
## [1] "Function Converged....."
```

```r
ChiRF(mod3,mod4)
```

```
## Likelihood Ratio Test: Reduced Model vs. Full Model
## ----------------------------------------------------------
##  ChiSquare DF       p_value
##  456.10527 10 1.0413078e-91
## ----------------------------------------------------------
```

We can also use Voung test for model comprison as follows:

```r
vountest(mod3, mod4)
```

```
## Positive mi>0 =  5987
## Negative mi<0 =  8018
```

```
##                         Z_value p  q      p_value
## V1 Un adj. one-tail -10.131408 2 12 2.004187e-24
## V1 Adj. one-tail     -8.010707 2 12 5.702541e-16
```

```r
names(mod4)
```

```
##  [1] "coeff"      "ithetainv"  "ith"        "utheta"     "logLik"
##  [6] "N"          "nvar1"      "nvar2"      "nvar"       "AIC"
## [11] "BIC"        "AICC"       "Lambdas"    "Phi1"       "Phi2"
## [16] "mtype"      "convg"      "Control"    "GOF.Chi"    "Disp.Ztest"
## [21] "T1"         "Deviance"   "nparam"     "y1"         "y2"
## [26] "logliky1"   "logliky2"
```

```r
mod4$coeff
```

```
##      Var.Names     Coeff      s.e    t.value  p.value  Adj.s.e Adj.p.value
## 1  Y1:Constant  0.243605 0.025718   9.472145 0.000000 0.018086    0.000000
## 2    SexDriver -0.016441 0.018882  -0.870722 0.383921 0.013279    0.215688
## 3         Area -0.024700 0.017813  -1.386600 0.165586 0.012527    0.048665
## 4     SevFatal -0.119248 0.083567  -1.426978 0.153608 0.058769    0.042466
## 5    SevSerius -0.169469 0.027325  -6.202018 0.000000 0.019216    0.000000
## 6     LightCon  0.143166 0.020631   6.939264 0.000000 0.014509    0.000000
## 7  Y2:Constant -0.701748 0.037220 -18.854174 0.000000 0.079572    0.000000
## 8    SexDriver -0.061683 0.029831  -2.067738 0.038683 0.063776    0.333469
## 9         Area -0.374765 0.027508 -13.623839 0.000000 0.058809    0.000000
## 10    SevFatal  0.623478 0.083737   7.445679 0.000000 0.179021    0.000498
## 11   SevSerius  0.251723 0.036869   6.827414 0.000000 0.078823    0.001409
## 12    LightCon -0.227105 0.029961  -7.580069 0.000000 0.064053    0.000393
```

Fitting right-truncated model

We will use data set one to fit this model too. The bivariate right-truncated Poisson model with constant only.

Reduced model (constant only model)

Following r code fit the the reduced model.

```r
mod5<-bpglm(exdata[,1:2],mtype=6)
```

```
## ----- Bi-variate Poisson Regression ------
## ----- Constant only model.....No covariates -----
## Iteration =  1
## Log Likelihood =  -25569.18
## Iteration =  2
## Log Likelihood =  -22291.33
## Iteration =  3
## Log Likelihood =  -18754.28
## Iteration =  4
## Log Likelihood =  -17606.32
## Iteration =  5
## Log Likelihood =  -16933.79
## Iteration =  6
## Log Likelihood =  -16686.79
## Iteration =  7
## Log Likelihood =  -16619.4
## Iteration =  8
## Log Likelihood =  -16602.39
## Iteration =  9
## Log Likelihood =  -16597.9
## Iteration =  10
## Log Likelihood =  -16596.66
## Iteration =  11
## Log Likelihood =  -16596.31
## Iteration =  12
## Log Likelihood =  -16596.21
## Iteration =  13
## Log Likelihood =  -16596.18
##
## ------- RTBIVP DEPENDENT (Marginal/Conditional)----
##
## Log Likelihood = -16596.18
## AIC = 33196.37
## AICC = 33196.37
## BIC = 33209.62
## Deviance = 111304.1
## Phi1 =  0.7530277 Phi2 =  0.889884
##
## Iteration =  13
## Parameter Estimates
## ------------------------------------------------------------------------
##    Var.Names     Coeff      s.e   t.value p.value  Adj.s.e Adj.p.value
##  Y1:Constant  1.008300 0.008884 113.49940       0 0.007709           0
##  Y2:Constant -1.098888 0.015122 -72.66663       0 0.014265           0
```

```
## ----------------------------------------------------------------------
##
## Overdispersion: Z-test for y1 & y2 =
##                        Z_value p_value
## Z(Y1) Marginal 2-tail 26.74441 0.00000
## Z(Y2) Marginal 2-tail -1.94359 0.05195
##
## Good-ness-of-fit (T1) Overdispersion (T2):
##                     ChiSquare DF   p_value
## T1 (GOF)             11.88721 12 0.4547802
## T2(Dispersion Test)  13.48915 12 0.3345111
##
## Pearson Chisquare GOF using predicted probability
##    Chi.square D.F p.value
##  1.424473e-28  33       1
##
## [1] "Function Converged....."
```

Full model (constant & covariates)

Following r code fit the fullmodel with five covariates. Here, we will use coefficients from model 2 (mod2) as initial value for convergence.

```
mod6<-bpglm(exdata[,1:2],exdata[,3:6],exdata[,3:6],icob=as.matrix(mod2$coeff[,2],ncol=1),mtype=6)
```

```
## ----- Bi-variate Poisson Regression -----
## ----- Model with covariates --------------------
## Iteration =  1
## Log Likelihood =  -16452.54
## Iteration =  2
## Log Likelihood =  -16459.41
## Iteration =  3
## Log Likelihood =  -16477.78
## Iteration =  4
## Log Likelihood =  -16487.89
## Iteration =  5
## Log Likelihood =  -16493.44
## Iteration =  6
## Log Likelihood =  -16496.56
## Iteration =  7
## Log Likelihood =  -16498.33
## Iteration =  8
## Log Likelihood =  -16499.35
## Iteration =  9
## Log Likelihood =  -16499.94
## Iteration =  10
## Log Likelihood =  -16500.28
## Iteration =  11
## Log Likelihood =  -16500.48
## Iteration =  12
## Log Likelihood =  -16500.6
## Iteration =  13
## Log Likelihood =  -16500.67
## Iteration =  14
## Log Likelihood =  -16500.71
```

```
## Iteration =  15
## Log Likelihood =  -16500.73
## Iteration =  16
## Log Likelihood =  -16500.74
## Iteration =  17
## Log Likelihood =  -16500.75
## Iteration =  18
## Log Likelihood =  -16500.76
## Iteration =  19
## Log Likelihood =  -16500.76
##
## ------- RTBIVP DEPENDENT (Marginal/Conditional)----
##
## Log Likelihood = -16500.76
## AIC = 33021.52
## AICC = 33021.59
## BIC = 33087.76
## Deviance = 102009.8
## Phi1 =  0.7494381 Phi2 =  0.9265475
##
## Iteration =  19
## Parameter Estimates
## ---------------------------------------------------------------------
##    Var.Names     Coeff      s.e   t.value  p.value  Adj.s.e Adj.p.value
##  Y1:Constant -0.179938 0.210687 -0.854054 0.393112 0.182392    0.323908
##       Gender -0.063351 0.022914 -2.764698 0.005716 0.019837    0.001413
##          Age  0.016348 0.002859  5.718809 0.000000 0.002475    0.000000
##     Hispanic  0.007772 0.031027  0.250492 0.802216 0.026860    0.772323
##      Veteran  0.055587 0.026830  2.071838 0.038327 0.023227    0.016733
##  Y2:Constant  0.888045 0.359161  2.472556 0.013445 0.345718    0.010234
##       Gender  0.453267 0.038142 11.883616 0.000000 0.036715    0.000000
##          Age -0.030014 0.004885 -6.144210 0.000000 0.004702    0.000000
##     Hispanic -0.248631 0.057688 -4.309945 0.000017 0.055529    0.000008
##      Veteran  0.219564 0.041788  5.254163 0.000000 0.040224    0.000000
## ---------------------------------------------------------------------
##
## Overdispersion: Z-test for y1 & y2 =
##                        Z_value p_value
## Z(Y1) Marginal 2-tail 26.74441 0.00000
## Z(Y2) Marginal 2-tail -1.94359 0.05195
##
## Good-ness-of-fit (T1) Overdispersion (T2):
##                      ChiSquare DF   p_value
## T1 (GOF)              11.80488 12 0.4614746
## T2(Dispersion Test)  13.42573 12 0.3388705
##
## Pearson Chisquare GOF using predicted probability
##  Chi.square D.F     p.value
##    52.02632  25 0.001189605
##
## [1] "Function Converged....."
```

```
ChiRF(mod5,mod6)
```

```
## Likelihood Ratio Test: Reduced Model vs. Full Model
```

```
## ----------------------------------------------------------
##  ChiSquare DF      p_value
##  190.85116  8 5.3921605e-37
## ----------------------------------------------------------
```

We can also use Voung test for model comprison as follows:

```r
vountest(mod5, mod6)
```

```
## Positive mi>0 =  2602
## Negative mi<0 =  2965
```

```
##                        Z_value p  q      p_value
## V1 Un adj. one-tail -3.914127 2 10 0.0000453659
## V1 Adj. one-tail    -2.499084 2 10 0.0062257373
```

```r
names(mod6)
```

```
##  [1] "coeff"      "ithetainv"  "ith"        "utheta"     "logLik"
##  [6] "N"          "nvar1"      "nvar2"      "nvar"       "AIC"
## [11] "BIC"        "AICC"       "Lambdas"    "Phi1"       "Phi2"
## [16] "mtype"      "convg"      "Control"    "GOF.Chi"    "Disp.Ztest"
## [21] "T1"         "Deviance"   "nparam"     "y1"         "y2"
## [26] "logliky1"   "logliky2"
```

```r
mod6$coeff
```

```
##      Var.Names     Coeff      s.e   t.value  p.value  Adj.s.e Adj.p.value
## 1  Y1:Constant -0.179938 0.210687 -0.854054 0.393112 0.182392    0.323908
## 2       Gender -0.063351 0.022914 -2.764698 0.005716 0.019837    0.001413
## 3          Age  0.016348 0.002859  5.718809 0.000000 0.002475    0.000000
## 4     Hispanic  0.007772 0.031027  0.250492 0.802216 0.026860    0.772323
## 5      Veteran  0.055587 0.026830  2.071838 0.038327 0.023227    0.016733
## 6  Y2:Constant  0.888045 0.359161  2.472556 0.013445 0.345718    0.010234
## 7       Gender  0.453267 0.038142 11.883616 0.000000 0.036715    0.000000
## 8          Age -0.030014 0.004885 -6.144210 0.000000 0.004702    0.000000
## 9     Hispanic -0.248631 0.057688 -4.309945 0.000017 0.055529    0.000008
## 10     Veteran  0.219564 0.041788  5.254163 0.000000 0.040224    0.000000
```

We can also use Voung test for model comprison between untruncated and truncated models as follows:

```r
vountest(mod2, mod6)
```

```
## Positive mi>0 =  2626
## Negative mi<0 =  2941
```

```
##                        Z_value  p  q      p_value
## V1 Un adj. one-tail -6.864799 10 10 3.329255e-12
## V1 Adj. one-tail    -6.864799 10 10 3.329255e-12
```

It seems truncated model is a better one. We can also compare other measures as follows:

```r
cbind(AIC=mod2$AIC[1],BIC=mod2$BIC[1],Loglik=mod2$logLik[1])
```

```
##           AIC      BIC    Loglik
## [1,] 33192.13 33258.38 -16586.07
```

```r
cbind(AIC=mod6$AIC[1],BIC=mod6$BIC[1],Loglik=mod6$logLik[1])
```

```
##           AIC      BIC    Loglik
## [1,] 33021.52 33087.76 -16500.76
```

All measures shows truncated model is better than untruncated one for this data set.

Acknowledgement The authors acknowledge gratefully to the HRS (Health and Retirement Study) which is sponsored by the National Institute of Aging (Grant Number NIA U01AG09740) and conducted by the University of Michigan.The second data set was obtained from "Police reported road accident statistics (STATS19) − Department for Transport" (http://data.gov.uk/dataset/road-accidents-safety-data). We acknowledge gratefully that the study is supported by the HEQEP sub-project 3293, University Grants Commission of Bangladesh and the World Bank.

References:

1. Leiter RE and Hamdan MA. Some Bivariate Probability Models Applicable to Traffic Accidents and Fatalities. *International Statistical Review* (1973) 41(1): 87-100.

2. Kocherlakota S and Kocherlakota K. *Bivariate Discrete Distributions.* Marcel Dekker. 1992.

3. Islam MA and Chowdhury RI. A Bivariate Poisson Models with Covariate Dependence. *Bulletin of Calcutta Mathematical Society* (2015) 107: 11-20.

4. Chowdhury RI and Islam MA. A Bivariate Poisson Models with Covariate Dependence. *Applied Mathematics* (2016) 7: 1589-1598.

5. Islam, M.A., Chowdhury, R.I. (2017). A generalized right truncated bivariate Poisson regression model with applications to health data. *PLOS ONE* (2017) 12(6): e0178153.

6. Islam, M.A., Chowdhury, R.I. *Analysis of Repeated Measures Data.* Springer, Singapore. 2017.

7. Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* (1989) 57: 307-333.

8. Health and Retirement Study, (Wave 10) Public Use Dataset. Produced and Distributed by the University of Michigan with Funding from the National Institute on Aging (Grant Number NIA U01AG09740). Ann Arbor, MI.