King Saud University

Journal of King Saud University – Science

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

On bivariate Poisson regression models

CrossMark

Fatimah E. AlMuhayfith [a], Abdulhamid A. Alzaid [b], Maha A. Omair [b],*

[a] *Department of Mathematics and Statistics, King Faisal University, Saudi Arabia*
[b] *Department of Statistics and Operations Research, King Saud University, Saudi Arabia*

Abstract In this paper, we consider estimating the parameters of bivariate and zero-inflated bivariate Poisson regression models using the conditional method. This method is compared with the standard method, which uses the joint probability function. Simulations and real applications show that the two methods have almost identical Akaike Information Criteria and parameter estimates, but the conditional method has a much faster execution time than the joint method. We conducted our computations using the R and SAS package. Our results also indicate that the execution time of SAS is faster than that of R.

1. Introduction

Joint modeling of two or more counts data has received a great deal of attention in recent years. Bivariate count models are used in cases where two count variables are correlated and need to be jointly estimated. For example, variables can include the number of emergency and non-emergency visits by a person to the hospital, the number of insurance claims with and without bodily injuries, or the number of voluntary and involuntary job changes.

The bivariate Poisson is the most widely used model for bivariate counts. It was proposed by Holgate (1964) and presented by Johnson and Kotz (1969). The definition of the bivariate Poisson distribution is not unique. Several approaches have been discussed by Kocherlakota and Kocherlakota (1992).

Here, we adopt the trivariate reduction method to construct the distribution (Johnson et al., 1997). Consider three independent Poisson random variables, X_k, $k = 1, 2, 3$, with parameters (means) $\lambda_k > 0$, $k = 1, 2, 3$. Then, the random variables $Y_1 = X_1 + X_3$ and $Y_2 = X_2 + X_3$ are set to follow joint bivariate Poisson distribution JBP[1]$(\lambda_1, \lambda_2, \lambda_3)$. The joint probability mass function (p.m.f) is given by

$$f_{\text{JBP}}(y_1, y_2; \; \lambda_1, \lambda_2, \lambda_3) =$$

$$e^{-(\lambda_1+\lambda_2+\lambda_3)} \frac{\lambda_1^{y_1}}{y_1!} \frac{\lambda_2^{y_2}}{y_2!} \sum_{r=0}^{min(y_1,y_2)} \binom{y_1}{r}\binom{y_2}{r} r! \left(\frac{\lambda_3}{\lambda_1 \lambda_2}\right)^r. \quad (1)$$

It can be easily shown that Y_1 and Y_2 are marginally distributed as Poisson with means $\lambda_1 + \lambda_3$ and $\lambda_2 + \lambda_3$, respectively. The covariance of Y_1 and Y_2 is λ_3, and hence, λ_3 is a measure of dependence between the two random variables. If $\lambda_3 = 0$, then the two variables are independent and the bivariate Poisson distribution reduces to the product of two independent Poisson distributions (referred to as double Poisson distribution "DP"[2]).

* Corresponding author.
  E-mail address: maomair@ksu.edu.sa (M.A. Omair).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

---
[1] JBP: joint bivariate Poisson distribution.
[2] DP: double Poisson distribution.

By means of conditional probability theory, the joint density (1) can be written as the product of a marginal and a conditional distribution. Hence:

$$f(y_1, y_2) = f_{Y_2|Y_1}(y_2|y_1) f_{Y_1}(y_1), \quad (2)$$

or

$$f(y_1, y_2) = f_{Y_1|Y_2}(y_1|y_2) f_{Y_2}(y_2). \quad (3)$$

In general, every decomposition of $f(y_1, y_2)$ will lead to different marginal and conditional distributions.

Returning to Eq. (2), Y_1 is a Poisson distribution with parameter $\mu_1 = \lambda_1 + \lambda_3$ and the conditional distribution of Y_2 given Y_1 is expressed as

$$f_{Y_2|Y_1}(y_2|y_1) = \sum_{r=0}^{min(y_1,y_2)} \binom{y_1}{r} p_1^r (1-p_1)^{y_1-r} \frac{e^{-\lambda_2} \lambda_2^{y_2-r}}{(y_2-r)!}, \quad (4)$$

where $p_1 = \frac{\lambda_3}{\lambda_3 + \lambda_1}$.

The expression in (4) is a convolution of a Poisson variable with parameter λ_2 and a binomial with parameters (y_1, p_1). The conditional mean and variance are given by:

$$E(Y_2|Y_1) = \lambda_2 + p_1 y_1,$$
$$Var(Y_2|Y_1) = \lambda_2 + p_1(1-p_1)y_1.$$

Hence, the joint p.m.f of conditional model 1 "CM1"[3] given in (2):

$$f(y_1, y_2) = f_{Y_2|Y_1}(y_2|y_1) f_{Y_1}(y_1)$$
$$= \sum_{r=0}^{min(y_1,y_2)} \binom{y_1}{r} p_1^r (1-p_1)^{y_1-r} \frac{e^{-\lambda_2} \lambda_2^{y_2-r}}{(y_2-r)!} * \frac{e^{-\mu_1} \mu_1^{y_1}}{y_1!}$$
$$= f_{CM1}(y_1, y_2;\ p_1, \lambda_2, \mu_1). \quad (5)$$

For this model $Cov(y_1, y_2) = p_1 \mu_1$. Hence, the covariance is zero if and only if $p_1 = 0$ (note that $\mu_1 = 0$ correspond to the univariate case).

In the same way, Eq. (3) can be written as:

$$f(y_1, y_2) = f_{Y_1|Y_2}(y_1|y_2) f_{Y_2}(y_2)$$
$$= \sum_{r=0}^{min(y_1,y_2)} \binom{y_2}{r} p_2^r (1-p_2)^{y_2-r} \frac{e^{-\lambda_1} \lambda_1^{y_1-r}}{(y_1-r)!} * \frac{e^{-\mu_2} \mu_2^{y_2}}{y_2!}$$
$$= f_{CM2}(y_1, y_2; p_2, \lambda_1, \mu_2), \quad (6)$$

where $p_2 = \frac{\lambda_3}{\lambda_3 + \lambda_2}$, and $\mu_2 = \lambda_2 + \lambda_3$.

Also, for this model $Cov(y_1, y_2) = p_2 \mu_2$. Hence, the covariance is zero if and only if $p_2 = 0$.

Note that for CM1 and CM2, p_1 and p_2 play the same role as λ_3 in JBP.

The bivariate Poisson model was applied by King (1989) to the annual number of presidential vetoes of social welfare bills and defense bills, by Jung and Winkelmann (1993) to the number of voluntary and involuntary changes, and by Ozuna and Gomez (1994) to the number of trips to different recreational sites. They model the marginal expectation of Y_1 and Y_2 as a log linear function of explanatory variables.

The disadvantage of this particular model is that it does not allow for over/under dispersion (the marginal distributions are Poisson) or negative correlation, and thus lacks generality. For the case of over-dispersed count data, the mixed Poisson models are potentially useful (BermÇdez and Karlis, 2012;

Ghitany et al., 2012; Gurmu and Elder, 2000; Munkin and Trivedi, 1999). There are also some other models that allow for negative correlation (Berkhout and Plug, 2004; Chib and Winkelmann, 2001; Gurmu and Elder, 2008; Karlis and Meligkotsidou, 2007; Van Ophem, 1999).

It is clear that there are one-to-one transformations between the parameters of the three representations of the bivariate Poisson distribution. Hence, according to the maximum likelihood invariance principle, we obtain identical estimates of the parameters no matter which representation has been used. It will be of interest to see whether such an invariance property holds when we have explanatory variables.

In this paper, we compare fitting a bivariate Poisson regression model using both the joint and conditional arguments described above. We also consider the inflated versions of these models to allow for over-dispersion and discuss inferences related to the parameters involved in the models and the execution times, using SAS and R software.

The outline of the paper is as follows. In Section 2, we present the bivariate Poisson regression model and discuss the estimation procedure using both the joint and conditional procedures. In Section 3, we discuss the zero-inflation version of the joint bivariate Poisson regression model and the proposed conditional models. In Section 4, a simulation study is conducted to compare the three models. In Section 5, applications on two data sets from Australia and Saudi Arabia are illustrated using the considered bivariate regression models. Finally, some concluding remarks are provided in Section 6.

2. Bivariate Poisson regression model

Here, we assume that the parameters of the models depend on explanatory variables. In the joint bivariate Poisson regression model "JBPM"[4], $\lambda_k > 0$ with $k = 1$, 2 and 3 can be related to various explanatory variables using the classical exponential link functions. Therefore, the joint bivariate Poisson regression model can take the following form:

$$(Y_{1i}, Y_{2i}) \sim JBP(\lambda_{1i}, \lambda_{2i}, \lambda_{3i})$$
$$log\lambda_{ki} = \mathbf{w}_i^T \boldsymbol{\alpha}_k, \quad k = 1,\ 2,\ 3, \quad (7)$$

where $i = 1, 2, \ldots, n$ denotes the observation number, $\mathbf{w}_i$ denotes a vector of explanatory variables of length l for the i-th observation related to the k-th parameter and $\boldsymbol{\alpha}_k$ is the corresponding vector of regression coefficients.

In the case of the explanatory variables, two aspects should be stressed. First, different covariates can be used to model each parameter (or same covariates) and second, covariates can be introduced to model λ_3 in order to learn more about the influence of the covariates on each pair of variables (or, to facilitate interpretation, no covariates are used to model λ_3).

Gourieroux et al. (1984) derived pseudo maximum likelihood estimation methods. Jung and Winkelmann (1993) and Kocherlakota and Kocherlakota (2001) considered the joint bivariate Poisson regression model using a Newton–Raphson procedure. Karlis and Ntzoufras (2005) constructed an EM algorithm to remedy convergence problems encountered with the Newton Raphson procedure. Ho and Singer (2001) proposed a generalized least squares method for maximizing the log likelihood. Karlis and Tsiamyrtzis (2008) implemented a

---

[3] CM1: Conditional Model 1; CM2: Conditional Model 2.

[4] JBPM: joint bivariate Poisson regression model.

Bayesian inference that does not rely on the MCMC scheme. Tsionas (2001), Ma and Kockelman (2006) and Choe et al. (2012) considered a Bayesian approach based on the MCMC technique to execute some computations.

Here we use the maximum likelihood method for estimating the parameters. Consider independent observations (y_{1i}, y_{2i}) with the i-th vector having the join bivariate Poisson distribution

$$f(y_{1i}, y_{2i}) = e^{-(\lambda_{1i} + \lambda_{2i} + \lambda_{3i})} \varphi(y_{1i}, y_{2i})$$

where

$$\varphi(y_{1i}, y_{2i}) = \sum_{r=0}^{min(y_{1i}, y_{2i})} \frac{\lambda_{3i}^r \lambda_{1i}^{(y_{1i}-r)} \lambda_{2i}^{(y_{2i}-r)}}{r!(y_{1i}-r)!(y_{2i}-r)!}.$$

Then, the corresponding score functions $U_{kj} = \frac{\partial \log L}{\partial \alpha_{kj}}$, $k = 1, 2, 3$; $j = 0, 1, \ldots, l$ are

$$U_{1j} = \sum_{i=1}^{n} \lambda_{1i} w_{ji} \left[\frac{\varphi(y_{1i}-1, y_{2i})}{\varphi(y_{1i}, y_{2i})} - 1 \right],$$

$$U_{2j} = \sum_{i=1}^{n} \lambda_{2i} w_{ji} \left[\frac{\varphi(y_{1i}, y_{2i}-1)}{\varphi(y_{1i}, y_{2i})} - 1 \right],$$

$$U_{3j} = \sum_{i=1}^{n} \lambda_{3i} w_{ji} \left[\frac{\varphi(y_{1i}-1, y_{2i}-1)}{\varphi(y_{1i}, y_{2i})} - 1 \right].$$

For the conditional Poisson regression models, we will only consider the first conditional model (CM1), as the argument for the second conditional model (CM2) is the same. The likelihood of CM1 can be written as the product of two likelihoods; the first one is that of the conditional model and the second is that of the marginal model. Hence, assuming that the regression parameters are different, one can maximize each likelihood separately with respect to the corresponding parameters. For the CM1 we denote the parameters by $\boldsymbol{\beta}_k$, $k = 1, 2, 3$.

For the marginal Poisson regression model, let μ_{1i} be the mean of Y_{1i}, $i = 1, 2, \ldots, n$. Then, using the standard Poisson regression model, we set:

$$\log \mu_{1i} = \beta_{10} + \sum_{j=1}^{l} \beta_{1j} w_{ji}, \quad i = 1, 2, \ldots, n. \tag{8}$$

For the conditional regression model, the conditional density (4) of $Y_{2i}|Y_{1i}$, $i = 1, 2, \ldots, n$ involves the parameters p_{1i} and λ_{2i}. We use the following link functions:

$$\log \lambda_{2i} = \beta_{20} + \sum_{j=1}^{l} \beta_{2j} w_{ji}, \quad i = 1, 2, \ldots, n. \tag{9}$$

$$\text{logit}(p_{1i}) = \beta_{30} + \sum_{j=1}^{l} \beta_{3j} w_{ji}, \quad i = 1, 2, \ldots, n, \tag{10}$$

The score functions $V_{kj} = \frac{\partial \log L}{\partial \beta_{kj}}$, $k = 1, 2, 3$; $j = 0, 1, \ldots, l$ for each regression parameters are summarized as follows:

$$\left. \begin{aligned} V_{1j} &= \sum_{i=1}^{n} [y_{1i} - \mu_{1i}] w_{ji}, \\ V_{2j} &= \sum_{i=1}^{n} \left[-\lambda_{2i} + y_{2i} - \frac{1}{f_{y_{2i}|y_{1i}}(y_{2i}|y_{1i})} \sum_{r=0}^{min(y_{1i}, y_{2i})} r \binom{y_{1i}}{r} p_{1i}^r (1-p_{1i})^{y_{1i}-r} \frac{e^{-\lambda_{2i}} \lambda_{2i}^{y_{2i}-r}}{(y_{2i}-r)!} \right] w_{ji}, \\ V_{3j} &= \sum_{i=1}^{n} \left[-y_{1i} p_{1i} + \frac{1}{f_{y_{2i}|y_{1i}}(y_{2i}|y_{1i})} \sum_{r=0}^{min(y_{1i}, y_{2i})} r \binom{y_{1i}}{r} p_{1i}^r (1-p_{1i})^{y_{1i}-r} \frac{e^{-\lambda_{2i}} \lambda_{2i}^{y_{2i}-r}}{(y_{2i}-r)!} \right] w_{ji} \\ j &= 0, 1, \ldots, l. \end{aligned} \right\}; \tag{11}$$

By equating the score functions in (11) to zero, one can obtain the parameter estimates directly. We note that the likelihood equations are non-linear and do not have a closed-form solution, so they can be solved numerically.

The following theorem gives the relations between the score functions of JBPM and CM1.

Theorem 1

$$V_{1j} = U_{1j} + U_{3j} \tag{12}$$
$$V_{2j} = U_{2j} \tag{13}$$
$$V_{3j} = U_{3j} - U_{1j} \tag{14}$$

Proof: See the Appendix I.

Note that when we set $U_{kj} = 0$ for all $k = 1, 2, 3$ and $j = 0, 1, \ldots, l$ we get $V_{kj} = 0$, for all $k = 1, 2, 3$ and $j = 0, 1, \ldots, l$ and vice versa. This shows that the likelihood invariance property holds when we have explanatory variables.

3. A bivariate zero-inflated model

The zero-inflated model is used when a count data set shows a large proportion of zeros. A bivariate zero-inflated model can be constructed by increasing the probability of the event $(y_1 = 0, y_2 = 0)$ and decreasing the other joint probabilities.

The zero-inflated joint bivariate Poisson model "ZIJBPM"[5] is specified by the probability function:

$$f_{\text{ZIJBPM}}(y_1, y_2) = \begin{cases} \pi + (1-\pi) f_{\text{JBP}}(0, 0; \lambda_1, \lambda_2, \lambda_3) & y_1 = y_2 = 0 \\ (1-\pi) f_{\text{JBP}}(y_1, y_2; \lambda_1, \lambda_2, \lambda_3) & y_1 \text{ or } y_2 \neq 0 \end{cases}, \tag{15}$$

where $f_{\text{JBP}}(y_1, y_2; \lambda_1, \lambda_2, \lambda_3)$ is the joint probability function given in (1) and $0 < \pi < 1$ is the mixing proportion.

This model was proposed by Karlis and Ntzoufras (2003, 2005) in order to allow for over-dispersion of the corresponding marginal distributions. It was applied by Wang et al. (2003) to analyze two types of occupational injuries and by BermÇdez (2009) in the automobile insurance context for a bivariate case.

One can easily see that the marginal distributions are no longer simple Poisson distributions, but are now zero-inflated versions. Note that one may define more complicated models by assuming other kinds of inflations (Karlis and Ntzoufras, 2005). Moreover, one may add covariates to π, implying that inflation depends on external factors.

For estimation of the parameters of ZIJBPM, we use the maximum likelihood procedure as follows. Let (y_{1i}, y_{2i}), $i = 1, 2, \ldots, n$ be a random sample observed from ZIJBPM. The likelihood function for the observed random sample is given by:

$$L(\pi_i, \lambda_{1i}, \lambda_{2i}, \lambda_{3i}; y_{1i}, y_{2i})$$
$$= \prod_{i=1}^{n} \{ (\pi_i + (1-\pi_i) f_{JBP}(0, 0; \lambda_{1i}, \lambda_{2i}, \lambda_{3i}))^{1-a_i} ((1-\pi_i) f_{JBP}$$
$$\times (y_{1i}, y_{2i}; \lambda_{1i}, \lambda_{2i}, \lambda_{3i}))^{a_i} \},$$

where $a_i = 1$ if $(y_{1i}, y_{2i}) \neq (0, 0)$ and $a_i = 0$ otherwise. In order to fit this model, we used the link functions in (7) for λ_{ki}, and for modeling the mixing proportion π_i, we used a logit link function:

$$\text{logit}(\pi_i) = \gamma_0 + \sum_{j=1}^{l} \gamma_j w_{ji}.$$

---

[5] ZIJBPM: Zero-inflated joint bivariate Poisson model.

The maximum likelihood estimates (MLEs) of the parameters can be obtained by solving equations $\frac{\partial \log L}{\partial \alpha_{1j}} = 0$, $\frac{\partial \log L}{\partial \alpha_{2j}} = 0$, $\frac{\partial \log L}{\partial \alpha_{3j}} = 0$ and $\frac{\partial \log L}{\partial \gamma_j} = 0$ for $j = 0, 1, \ldots, l$, simultaneously.

The zero-inflated conditional model "ZICM"[6] is given by:

$$f_{\text{ZICM}}(y_1, y_2) = \begin{cases} \pi + (1 - \pi)f_{\text{CM}}(0, 0; p, \lambda, \mu) & y_1 = y_2 = 0 \\ (1 - \pi)f_{\text{CM}}(y_1, y_2; p, \lambda, \mu) & y_1 \text{ or } y_2 \neq 0 \end{cases} \tag{16}$$

where $f_{\text{CM}}(y_1, y_2)$ is the joint p.m.f of the conditional model given in (5) or (6) and π is the mixing proportion.

For parameter estimations of ZICM1, let (y_{1i}, y_{2i}), $i = 1, 2, \ldots, n$ be a random sample observed from ZICM1. The likelihood function for the observed random sample is given by

$$L(\pi_i, \lambda_{1i}, \lambda_{2i}, \lambda_{3i}; y_{1i}, y_{2i})$$
$$= \prod_{i=1}^{n} \{(\pi_i + (1 - \pi_i)f_{\text{CM1}}(0, 0; p_{1i}, \lambda_{2i}, \mu_{1i}))^{1-a_i}((1 - \pi_i)f_{\text{CM1}}$$
$$\times (y_{1i}, y_{2i}; p_{1i}, \lambda_{2i}, \mu_{1i}))$$

where $a_i = 1$ if $(y_{1i}, y_{2i}) \neq (0, 0)$ and $a_i = 0$ otherwise. In order to fit this model, we used the link functions, as in (8)–(10), for μ_{1i}, λ_{2i} and p_{1i}, respectively, and to model the mixing proportion π_i, we used a logit link function:

$$\text{logit}(\pi_i) = \gamma_0 + \sum_{j=1}^{l} \gamma_j w_{ji}.$$

The MLEs of the parameters can be obtained by solving equations $\frac{\partial \log L}{\partial \beta_{1j}} = 0$, $\frac{\partial \log L}{\partial \beta_{2j}} = 0$, $\frac{\partial LogL}{\partial \beta_{3j}} = 0$ and $\frac{\partial LogL}{\partial \gamma_j} = 0$ for $j = 0, 1, \ldots, l$, simultaneously.

4. Experiment results

We conducted a simulation study to examine the performance of the proposed conditional model compared with that of the joint bivariate Poisson model and the double Poisson model. We used three variables, namely gender, age and income from Health Care Australian data (Cameron et al., 1988) as covariates. For each of the following four cases we generated 1000 samples each of size 5190.

Case 1. Simulation from joint bivariate Poisson model.

We have simulated the data points (y_{1i}, y_{2i}) from the joint bivariate Poisson regression model with $\lambda_{1i}, \lambda_{2i}, \lambda_{3i}$ given by:

$$\lambda_{1i} = \exp(\alpha_{10} + \alpha_{11}w_{1i} + \alpha_{12}w_{2i} + \alpha_{13}w_{3i}),$$
$$\lambda_{2i} = \exp(\alpha_{20} + \alpha_{21}w_{1i} + \alpha_{22}w_{2i} + \alpha_{23}w_{3i}),$$
$$\lambda_{3i} = \exp(\alpha_{30} + \alpha_{31}w_{1i} + \alpha_{32}w_{2i} + \alpha_{33}w_{3i}).$$

We have used the following parameter values:

$\alpha_{10} = -1.9983$, $\alpha_{11} = 0.2268$, $\alpha_{12} = 0.6902$,
$\alpha_{13} = -0.2004$, $\alpha_{20} = -2.1718$, $\alpha_{21} = 0.6436$,
$\alpha_{22} = 3.1092$, $\alpha_{23} = -0.07716$, $\alpha_{30} = -3.0328$,
$\alpha_{31} = 0.1966$, $\alpha_{32} = 2.1272$, $\alpha_{33} = -0.3984$.

Case 2. Simulation from conditional model.

We generated the data points (y_{1i}, y_{2i}) from the conditional regression model 1, assuming the following:

[6] ZICM: Zero-inflated conditional model.

$$\mu_{1i} = \exp(\beta_{10} + \beta_{11}w_{1i} + \beta_{12}w_{2i} + \beta_{13}w_{3i}).$$
$$\lambda_{2i} = \exp(\beta_{20} + \beta_{21}w_{1i} + \beta_{22}w_{2i} + \beta_{23}w_{3i}),$$
$$\text{logit}(p_{1i}) = \beta_{30} + \beta_{31}w_{1i} + \beta_{32}w_{2i} + \beta_{33}w_{3i}.$$

Then,

$$\lambda_{3i} = p_{1i} * \mu_{1i},$$
$$\lambda_{1i} = \mu_{1i} - \lambda_{3i}.$$

We have used the following parameter values:

$\beta_{10} = -1.71473$, $\beta_{11} = 0.21565$, $\beta_{12} = 1.23798$,
$\beta_{13} = -0.27726$, $\beta_{20} = -2.1658$, $\beta_{21} = 0.6316$,
$\beta_{22} = 3.1162$, $\beta_{23} = -0.08287$, $\beta_{30} = -1.2005$,
$\beta_{31} = 0.1642$, $\beta_{32} = 1.4778$, $\beta_{33} = -0.08393$.

Case 3. Simulation from zero-inflated joint bivariate Poisson model.

We have simulated the data points (y_{1i}, y_{2i}) from the zero-inflated joint bivariate Poisson regression model with $\lambda_{1i}, \lambda_{2i}, \lambda_{3i}$ and π_i given by:

$$\lambda_{1i} = \exp(\alpha_{10} + \alpha_{11}w_{1i} + \alpha_{12}w_{2i} + \alpha_{13}w_{3i}),$$
$$\lambda_{2i} = \exp(\alpha_{20} + \alpha_{21}w_{1i} + \alpha_{22}w_{2i} + \alpha_{23}w_{3i}),$$
$$\lambda_{3i} = \exp(\alpha_{30} + \alpha_{31}w_{1i} + \alpha_{32}w_{2i} + \alpha_{33}w_{3i}),$$
$$\text{logit}(\pi_i) = \gamma.$$

We have used the following parameter values:

$\alpha_{10} = -1.4601$, $\alpha_{11} = 0.2223$, $\alpha_{12} = 0.7534$, $\alpha_{13} = -0.1644$,
$\alpha_{20} = -1.6012$, $\alpha_{21} = 0.5709$, $\alpha_{22} = 2.8359$, $\alpha_{23} = -0.03434$,
$\alpha_{30} = -2.7627$, $\alpha_{31} = -0.2254$, $\alpha_{32} = 1.9464$, $\alpha_{33} = -0.5794$,
$\gamma = -0.7853$.

Case 4. Simulation from zero-inflated conditional model.

We generated the data points (y_{1i}, y_{2i}) from the zero-inflated conditional regression model 1, assuming the following:

$$\mu_{1i} = \exp(\beta_{10} + \beta_{11}w_{1i} + \beta_{12}w_{2i} + \beta_{13}w_{3i}).$$
$$\lambda_{2i} = \exp(\beta_{20} + \beta_{21}w_{1i} + \beta_{22}w_{2i} + \beta_{23}w_{3i}),$$
$$\text{logit}(p_{1i}) = \beta_{30} + \beta_{31}w_{1i} + \beta_{32}w_{2i} + \beta_{33}w_{3i}, \text{logit}(\pi_i) = \gamma.$$

Then,

$$\lambda_{3i} = p_{1i} * \mu_{1i},$$

$$\lambda_{1i} = \mu_{1i} - \lambda_{3i}.$$

We have used the following parameter values:

$\beta_{10} = -1.2134$, $\beta_{11} = 0.1149$, $\beta_{12} = 1.034$, $\beta_{13} = -0.257$,
$\beta_{20} = -1.6096$, $\beta_{21} = 0.5437$, $\beta_{22} = 2.8886$, $\beta_{23} = -0.04059$,
$\beta_{30} = -1.2839$, $\beta_{31} = -0.1031$, $\beta_{32} = 0.6529$, $\beta_{33} = -0.2774$,
$\gamma = -0.789$.

Table 1 illustrates the means and standard deviations of the AIC for the fitted models under the four cases. Bias and mean square error (MSE) of the estimated parameters are presented in Table 2. Table 1 shows that the average values of the AIC for the joint bivariate Poisson model are almost identical with those obtained from conditional model 1 irrespective of the mechanism of data generations. On the other hand, when the data are generated from zero-inflated models, the fitted models ZIDPM, ZIJBP and ZICM1 perform better than their non-inflated counter parts, as expected.

Table 1 Mean and standard deviation (between brackets) of the AIC results of fitted models for simulated data.

Model		Case 1 of simulation	Case 2 of simulation	Case 3 of simulation	Case 4 of simulation
Non Inflated Models	Double Poisson Model (DPM)	18,329.47 (173.721)	18,326.61 (172.173)	19,489.94 (195.289)	13,011.69 (262.769)
	Joint bivariate Poisson Model (JBPM)	18,087.63 (172.2957)	18,083.14 (170.814)	19,016.7 (195.479)	12,182.46 (244.162)
	Conditional Model 1 (CM1)	18,081.69 (172.263)	18,077.37 (170.517)	19,009.95 (195.799)	12,179.54 (243.163)
Zero Inflated Models	Zero Inflated Double Poisson Model (ZIDPM)	18,308.27 (170.587)	18,303.82 (169.269)	18,339.16 (189.619)	10,643.36 (220.716)
	Zero Inflated Joint bivariate Poisson Model (ZIJBPM)	18,079.63 (172.29)	18,075.14 (170.814)	18,241.29 (189.959)	10,537.24 (220.085)
	Zero Inflated Conditional Model 1(ZICM1)	18,079.82 (172.29)	18,070.19 (170.819)	18,241.37 (189.955)	10,535.97 (220.005)

To perform official test for the hypothesis that there is no significant difference of the AIC's, we preformed Vuong test. Vuong (1989) set the information criterion in a testing framework in which the null hypothesis is that the two competing models are equally close to the true model. Under the null hypothesis that the models are indistinguishable, the test statistic is asymptotically distributed standard normal. We applied the Vuong test to investigate if there is statistically significant difference between the joint bivariate Poisson model and the conditional model 1 in terms of AICs. We also performed the test to compare the zero inflated models. In all cases the Vuong test confirmed that there is no significant difference between the two models. The smallest p-value obtained was 0.529. Both models always perform better than the double Poisson model.

In general, the simulation study confirmed the expected similarity of the results obtained under joint and conditional bivariate Poisson models. This concludes the invariance property holds when we have explanatory variables as expected from the results of Theorem 1.

5. Applications

5.1. First application

The data analyzed here were originally employed by Cameron et al. (1988) in their analysis of various measures of health-care utilization, using a sample of 5190 single-person households from the 1977–1978 Australian Health Survey. The data were obtained from the Journal of Applied Econometrics 1997 data archive.

Here, we model two possibly jointly dependent variables of health service utilization measures: (1) the number of consultations with doctors during the two-week period prior to the survey (Y_1); and (2) the number of prescribed medicines used in the past 2 days (Y_2).Three variables have been used as covariates, namely gender (1 female, 0 male), age in years divided by 100 (measured as midpoints of age groups) and the annual income in Australian dollars divided by 1000 (measured as midpoint of coded ranges). More details on the data can be found in Cameron et al. (1988).

We have fitted four different models for: joint bivariate Poisson, conditional models 1 and 2, zero-inflated joint bivariate Poisson and zero-inflated conditional Poisson models 1 and 2.

Model (A) covariates: age, gender, income and age * gender with gender as a covariate on the covariance term.
Model (B) covariates: age, gender, income and age * gender with a constant covariance term.
Model (C) covariates: age, gender and income with gender as a covariate on the covariance term.
Model (D) covariates: age, gender and income with a constant covariance term.

For the double Poisson and zero-inflated double Poisson, we fitted models (A) and (C) without the covariance term.

"PROC NLMIXED" and "rootSolve" in SAS and R, respectively, were used in fitting these models. Both packages yielded the same results.

Tables 3 and 4 summarize the model details and the fit of the bivariate models in terms of the AIC (Akaike Information Criteria), number of parameters estimated and the execution time required to fit the models using SAS and R packages. The results show that the execution time of the conditional argument is less than that of the joint model argument using SAS and R.

Table 3 indicates that estimating the two count events jointly is better than estimating the two count events independently.

The model evaluation statistics for zero-inflated bivariate count regression models are reported in Table 4. Since almost 54% of all observations occur when doctor visits and the number of prescribed medicines are zero, a zero-inflated model would be more appropriate. In comparing the model evaluation criterion for Tables 3 and 4, one can conclude that the zero-inflated bivariate models perform better than their counterparts in Table 3.

In comparing the AIC values of conditional models 1 and 2 with the corresponding values of the joint bivariate Poisson, we observed that the results are essentially the same. This is also confirmed using Vuong test. It turns out that all of the estimates are practically identical across all three alternative models. Hence, we can conclude that the proposed conditional model is equivalent to the joint model and computationally simpler.

We found the best model, which includes all significant parameters and has the smallest AIC, is model (C) of the conditional model 2. An analysis of the maximum likelihood parameter estimates derived for this model is provided in Table 5.

Table 2 Bias and MSE of the estimated parameters.

Model	True	α_{10}	α_{11}	α_{12}	α_{13}	α_{20}	α_{21}	α_{22}	α_{23}	α_{30}	α_{31}	α_{32}	α_{33}	γ
JBPM	α	−1.9983	0.2268	0.6902	−0.2004	−2.1718	0.6436	3.1092	−0.07716	−3.0328	0.1966	2.1272	−0.3984	–
	Bias	−0.0047	−1.3E−05	−0.0061	0.0006	0.0019	−0.0028	0.0007	−0.0076	−0.0031	0.0012	−0.01491	−0.0137	–
	MSE	0.0181	0.0083	0.0494	0.0138	0.0068	0.0019	0.0109	0.00289	0.0482	0.0145	0.091591	0.0379	–
CM1	β	β_{10}	β_{11}	β_{12}	β_{13}	β_{20}	β_{21}	β_{22}	β_{23}	β_{30}	β_{31}	β_{32}	β_{33}	γ
		−1.71473	0.21565	1.23798	−0.27726	−2.1658	0.6316	3.1162	−0.08287	−1.2005	0.1642	1.4778	−0.08393	–
	Bias	−0.00202	−0.00402	0.007261	0.000424	0.00138	−0.00262	0.001554	−0.00681	0.002381	0.039676	−0.01463	−0.01341	–
	MSE	0.008372	0.003128	0.016076	0.006666	0.006894	0.001941	0.010943	0.003058	0.091696	0.02406	0.088024	0.00018	–
ZIBPM	α	α_{10}	α_{11}	α_{12}	α_{13}	α_{20}	α_{21}	α_{22}	α_{23}	α_{30}	α_{31}	α_{32}	α_{33}	γ
		−1.4601	0.2223	0.7534	−0.1644	−1.6012	0.5709	2.8359	−0.03434	−2.7627	−0.2254	1.9464	−0.5794	−0.7853
	Bias	0.0051	−3E−05	−0.0189	−0.0167	−0.0025	0.0023	−0.0008	−0.0251	−0.00176	−0.0589	−0.01071	−0.00569	−0.0037
	MSE	0.0206	0.0089	0.0591	0.0131	0.0083	0.0025	0.0135	0.0027	0.072317	0.0489	0.077094	0.084767	0.0027
ZICM1	β	β_{10}	β_{11}	β_{12}	β_{13}	β_{20}	β_{21}	β_{22}	β_{23}	β_{30}	β_{31}	β_{32}	β_{33}	γ
		−1.2134	0.1149	1.034	−0.257	−1.6096	0.5437	2.8886	−0.04058	−1.2839	−0.1031	0.6529	−0.2774	−0.789
	Bias	0.109267	0.146683	−0.11928	−0.00371	−0.00103	0.000364	0.002026	−0.05032	0.068698	−0.02123	0.237688	0.018441	0.002686
	MSE	0.031875	0.036202	0.056103	0.015156	0.021908	0.006418	0.04057	0.00741	0.024504	0.006395	0.091596	0.014975	0.002135

Table 3 Model evaluation statistics of Bivariate Models for Health Care Australia data.

Model		Evaluation criterion		SAS execution time (s)	R execution time (s)
		Par.	AIC		
Double Poisson	A	10	20325.5	0.050	0.1
Model (DPM)	C	8	20482.0	0.048	0.08
Joint bivariate	A	12	19913.0	1.380	19.746
Poisson Model	B	11	19942.0	1.180	18.530
(JBPM)	C	10	20051.0	0.650	17.404
	D	9	20079.0	0.700	15.850
Conditional Model	A	12	19863.5	0.560	7.402
1 (CM1)	B	11	19864.5	0.560	6.639
	C	10	20007.4	0.470	6.738
	D	9	20009.4	0.440	5.694
Conditional Model	A	12	19865.4	0.560	8.114
2 (CM2)	B	11	19878.8	0.590	7.308
	C	10	20003.9	0.513	7.228
	D	9	20022.5	0.500	6.466

Table 4 Model evaluation statistics of Zero-Inflated Bivariate Models for Health Care Australia data.

Model		Evaluation criterion		SAS execution time (s)	R execution time (s)
		Par.	AIC		
Zero Inflated	A	11	19,306	0.56	7.899
Double Poisson	C	9	19,438	0.48	7.032
Model (ZIDPM)					
Zero Inflated Joint	A	13	19,199	1.63	37.533
bivariate Poisson	B	12	19,193	1.46	34.389
Model(ZIJBPM)	C	11	19,334	1.07	28.821
	D	10	19,332	0.73	26.508
Zero Inflated	A	13	19,101	1.40	23.827
Conditional Model	B	12	19,100	1.20	22.252
1(ZICM1)	C	11	19,252	1.06	17.663
	D	10	19250.1	0.64	16.031
Zero Inflated	A	13	19,114	1.62	22.930
Conditional Model	B	12	19,127	1.32	20.948
2(ZICM2)	C	11	19,251.8	1.01	21.722
	D	10	19,270.4	0.71	19.177

5.2. Second application

The data set was obtained from Dallah Hospital at Riyadh for 5000 persons insured directly at the hospital. At Dallah Hospital, there are three types of clinics: consultation clinics with appointments (CONS), open specialist clinics without appointments (OSC) and emergency clinics (AE). Here, we will consider joint modeling of the number of patient visits to the consultation clinics (Y_1) and to the open specialty clinics (Y_2) during the year 2011. The explanatory variables are age and gender. The sample correlation coefficient between the two variables is 0.347. The percentage of zero visits to the CONS and OSC clinics is 39.68% and 32.82%, respectively. Descriptive statistics for these variables are given in Table 6.

We have used different groups of covariates. In model (A), we used age and gender as covariates. Model (B): age, gender and age * gender. Model (C): age, age^2 and gender. Model (D): age, age^2, gender and age * gender. Model (E): age, age^2, gender, age * gender and age^2 * gender. In model (F), we classified

Table 5 Results from fitting the conditional model 2 to the Health Care Australia data.

Model (C)

Results from the fitted conditional model $f_{Y_1|Y_2}(y_1|y_2)$

Parameter	Covariate	Coef.	St.Er.	Pr > |t|
λ_1	Constant	−1.8919	0.1201	< .0001
	Gender	0.2851	0.08408	0.0007
	Age	0.4500	0.1893	0.0175
	Income	−0.2581	0.1097	0.0187
p_2	Constant	−1.4588	0.09368	< .0001
	Gender	−0.5783	0.1244	< .0001

No. Parameter	6
Log-likelihood	−3641.95
AIC	7295.9
BIC	7335.2

Results from the fitted Poisson model $f_{Y_2}(y_2)$

Parameter	Covariate	Coef.	St.Er.	Pr > ChiSq
μ_2	Constant	−1.87209	0.06550	< .0001
	Gender	0.57601	0.03638	< .0001
	Age	2.96270	0.08570	< .0001
	Income	−0.12539	0.05058	0.0132

No. Parameter	4
Log-likelihood	−6349.994
AIC	12707.9879
BIC	12734.2058

Final results from the fitted model $f_{Y_1|Y_2}(y_1|y_2)f_{Y_2}(y_2)$

No. Parameter	10
Log-likelihood	−9991.944
AIC	20,003.8879
BIC	20,069.4058

Table 6 Summary statistics of Dallah data.

Statistic	Number	Mean	St. dev	Median	Minimum	Maximum
CONS	5000	2.8794	4.8740	1	0	57
OSC	5000	4.1154	5.6395	2	0	49
Age	5000	34.1340	23.2540	32	0	102
Gender	5000	0.5082	0.5000	1	0	1

Table 7 Model evaluation statistics of Bivariate Models for Dallah data.

Model		Evaluation criterion		SAS execution time (s)	R execution time (s)
		Par.	AIC		
Double Poisson Model (DPM)	A	6	67,899.88	0.1	0.775
	B	8	67,885.72	0.1	0.679
	C	8	65,676.05	0.1	0.739
	D	10	65,677.3	0.08	0.683
	E	12	65,630.9	0.1	0.729
	F	30	65,108.56	0.15	0.846
Joint bivariate Poisson Model (JBPM)	A	9	66,908	0.89	8.378
	B	12	66,894	2.20	9.568
	C	12	64,758	2.66	11.861
	D	15	64,759	3.83	14.842
	E	18	64,716	5.49	43.014
	F	45	64,189	15.60	51.042
Conditional Model 1 (CM1)	A	9	66,904.89	0.64	3.302
	B	12	66,891.33	0.89	3.99
	C	12	64,713.41	0.97	4.164
	D	15	64,714.25	1.17	4.728
	E	18	64,677.63	2.94	5.25
	F	45	64,190.71	5.76	6.118
Conditional Model 2 (CM2)	A	9	66,882.99	0.85	4.057
	B	12	66,868.39	1.1	5.66
	C	12	64,759.65	1.19	5.98
	D	15	64,759.04	1.39	6.099
	E	18	64,715.27	2.19	9.319
	F	45	64,190.85	6.33	11.3

Table 8 Model evaluation statistics of Zero-Inflated Bivariate Models for Dallah data.

Model		Evaluation criterion		SAS execution time (s)	R execution time (s)
		Par.	AIC		
Zero Inflated Double Poisson Model (ZIDPM)	A	7	64,419	0.29	10.495
	B	9	64,412	0.34	12.199
	C	9	62,908	0.37	13.286
	D	11	62,911	0.50	15.104
	E	13	62,868	0.65	16.151
	F	31	62,473	1.01	19.542
Zero Inflated Joint bivariate Poisson Model (ZIJBPM)	A	10	63,898	1.23	12.757
	B	13	64,067	3.07	30.02
	C	13	62,355	3.54	32.910
	D	16	62,358	5.47	42.003
	E	19	62,513	4.89	45.723
	F	46	61,905	23.22	68.4
Zero Inflated Conditional Model 1 (ZICM1)	A	10	63,896	1.10	6.237
	B	13	64,066	2.83	7.830
	C	13	62,916	2.66	8.084
	D	16	62,921	3.47	9.741
	E	19	62,499	5.83	11.839
	F	46	61,913	17.78	14.483
Zero Inflated Conditional Model 2 (ZICM2)	A	10	63,897	1.29	8.397
	B	13	63,890	2.97	10.302
	C	13	62,358	3.40	10.187
	D	16	62,358	4.63	11.788
	E	19	62,320	5.61	13.902
	F	46	61,913	15.78	14.590

the age into groups (<5, 5–10, 11–15, 16–20, 21–25, 26–30, 31–35, 36–40, 41–45, 46–50, 51–55, 56–60, 61–65, >66) and used gender and age groups as covariates.

The model evaluation statistics for joint and conditional Poisson models 1 and 2 regarding AIC, the number of param- eters estimated and the execution time required to fit the mod- els using SAS and R packages are listed in Table 7, and those for zero-inflated are reported in Table 8. Most of the results of

Table 9 Results from fitting the zero inflated conditional model 1 to the Dallah data.

Parameter	Covariate	Model (A)		
		Coef.	St. Er.	Pr > |t|
μ_1	Constant	0.47620	0.022040	<.0001
	Age	0.01580	0.000382	<.0001
	Gender	0.17770	0.016970	<.0001
λ_2	Constant	0.37220	0.022880	<.0001
	Age	0.02437	0.000364	<.0001
	Gender	0.08910	0.017470	<.0001
p_1	Constant	−2.43820	0.126900	<.0001
	Age	0.00942	0.001765	<.0001
	Gender	0.28040	0.115200	0.0150
Mixing proportion	Constant	−1.69170	0.041470	<.0001
No. Parameter		10		
Log-likelihood		−31,938		
AIC		63,896		
BIC		63,961		

the joint bivariate Poisson model agree with the results of conditional models 1 and 2.

We found the best model, which includes all significant parameters and has the smallest AIC, is model (A) of the zero-inflated conditional model 1. This fitted model is listed in Table 9.

6. Conclusion

In this paper, we used conditional argument to introduce a two-stage procedure for estimating the bivariate Poisson regression model. Our study showed that this method has the same performance as that of using the joint density. The conditional method is simple as it partitions the joint likelihood into two univariate likelihoods. We also applied the two methods to two sets of real data. The conclusion is that the two ways of modeling have the same AIC but the execution times for the two-stage models are shorter.

Acknowledgments

Appendix I. Proof Theorem 1

Proof Eq. (12). $V_{1j} = U_{1j} + U_{3j}$

$$\text{R.H.S} = U_{1j} + U_{3j} = \sum_{i=1}^{n}\left\{\lambda_{1i}w_{ji}\left[\frac{\varphi(y_{1i}-1,y_{2i})}{\varphi(y_{1i},y_{2i})}-1\right] + \lambda_{3i}w_{ji}\left[\frac{\varphi(y_{1i}-1,y_{2i}-1)}{\varphi(y_{1i},y_{2i})}-1\right]\right\}$$

but, from the recurrence relations of the bivariate Poisson distribution (Johnson et al., 1997), we have

$$\frac{f_{\text{JBP}}(y_1-1,y_2-1;\lambda_1,\lambda_2,\lambda_3)}{f_{\text{JBP}}(y_1,y_2;\lambda_1,\lambda_2,\lambda_3)} = \frac{y_1}{\lambda_3} - \frac{\lambda_1}{\lambda_3}\frac{f_{\text{JBP}}(y_1-1,y_2;\lambda_1,\lambda_2,\lambda_3)}{f_{\text{JBP}}(y_1,y_2;\lambda_1,\lambda_2,\lambda_3)}$$

and, this is equivalent to

$$\frac{\varphi(y_{1i}-1,y_{2i}-1)}{\varphi(y_{1i},y_{2i})} = \frac{y_{1i}}{\lambda_{3i}} - \frac{\lambda_{1i}}{\lambda_{3i}}\frac{\varphi(y_{1i}-1,y_{2i})}{\varphi(y_{1i},y_{2i})}.$$

Substituting this result in the R.H.S, we get

$$\text{R.H.S} = U_{1j} + U_{3j} = \sum_{i=1}^{n}\left\{\lambda_{1i}w_{ji}\left[\frac{\varphi(y_{1i}-1,y_{2i})}{\varphi(y_{1i},y_{2i})}-1\right] + \lambda_{3i}w_{ji}\left[\frac{y_{1i}}{\lambda_{3i}} - \frac{\lambda_{1i}}{\lambda_{3i}}\frac{\varphi(y_{1i}-1,y_{2i})}{\varphi(y_{1i},y_{2i})}-1\right]\right\}$$

$$= \sum_{i=1}^{n}[y_{1i} - \lambda_{1i} - \lambda_{3i}]w_{ji} = \sum_{i=1}^{n}[y_{1i} - \mu_{1i}]w_{ji} = V_{1j} = \text{L.H.S}$$

and the proof is complete.

Proof Eq. (13). $V_{2j} = U_{2j}$

$$\text{L.H.S} = V_{2j} = \sum_{i=1}^{n}\left[-\lambda_{2i} + y_{2i} - \frac{1}{f_{Y_{2i}|Y_{1i}}(y_{2i}|y_{1i})}\right]$$
$$\sum_{r=0}^{min(y_{1i},y_{2i})} r\binom{y_{1i}}{r}p_{1i}^r(1-p_{1i})^{y_{1i}-r}\frac{e^{-\lambda_{2i}}\lambda_{2i}^{y_{2i}-r}}{(y_{2i}-r)!}\right]w_{ji}.$$

But, from the conditional density (4) we have

$$f_{Y_2|Y_1}(y_2|y_1) = \sum_{r=0}^{min(y_1,y_2)}\binom{y_1}{r}p_1^r(1-p_1)^{y_1-r}\frac{e^{-\lambda_2}\lambda_2^{y_2-r}}{(y_2-r)!},$$

where

$$p_1 = \frac{\lambda_3}{\lambda_3 + \lambda_1}$$

and $1 - p_1 = \frac{\lambda_1}{\lambda_3 + \lambda_1}$.

Substituting these results in the L.H.S, we get

$$\text{L.H.S} = V_{2j} = \sum_{i=1}^{n}\left[-\lambda_{2i} + \lambda_{2i}\frac{\sum_{r=0}^{min(y_{1i},y_{2i})}\frac{\lambda_{3i}^r\lambda_{1i}^{(y_{1i}-r)}\lambda_{2i}^{(y_{2i}-r-1)}}{r!(y_{1i}-r)!(y_{2i}-r-1)!}}{\sum_{r=0}^{min(y_{1i},y_{2i})}\frac{\lambda_{3i}^r\lambda_{1i}^{(y_{1i}-r)}\lambda_{2i}^{(y_{2i}-r)}}{r!(y_{1i}-r)!(y_{2i}-r)!}}\right]$$

$$w_{ji} = \sum_{i=1}^{n}\lambda_{2i}w_{ji}\left[\frac{\varphi(y_{1i},y_{2i}-1)}{\varphi(y_{1i},y_{2i})}-1\right] = U_{2j} = \text{R.H.S}$$

and the proof is complete.

Proof Eq. (14). $V_{3j} = U_{3j} - U_{1j}$

We have

$$p_{1i} = \frac{\lambda_{3i}}{\lambda_{3i} + \lambda_{1i}}$$

Hence,

$$\log\left(\frac{p_{1i}}{1-p_{1i}}\right) = \log\left(\frac{\lambda_{3i}}{\lambda_{1i}}\right) = \log(\lambda_{3i}) - \log(\lambda_{1i})$$
$$= w_i^T\alpha_3 - w_i^T\alpha_1.$$

On the other hand

$$\log\left(\frac{p_{1i}}{1-p_{1i}}\right) = w_i^T\beta_3$$

Therefore,

$$w_i^T\beta_3 = w_i^T\alpha_3 - w_i^T\alpha_1$$

As the explanatory variables are independent, one gets

$$\beta_{3j} = \alpha_{3j} - \alpha_{1j}, \quad j = 0, 1, \ldots, l.$$

Therefore,

$$\frac{\partial\log L}{\beta_{3j}} = \frac{\partial\log L}{\partial\alpha_{3j}}\cdot\frac{\partial\alpha_{3j}}{\partial\alpha_{3j}} - \frac{\partial\log L}{\partial\alpha_{1j}}\cdot\frac{\partial\alpha_{1j}}{\partial\alpha_{1j}},$$

or equivalently

$$\frac{\partial\log L}{\beta_{3j}} = \frac{\partial\log L}{\partial\alpha_{3j}} - \frac{\partial\log L}{\partial\alpha_{1j}},$$

and hence

$$V_{3j} = U_{3j} - U_{1j}.$$

This completes the proof.

Appendix II

The Appendix II involves the R and SAS language programs that are written in order to obtain the estimates with different models where proposed within the Paper. The programs are listed as follows:

A. R-program for obtaining the estimates using the Joint bivariate Poisson regression model (JBPM).

B. SAS-program for obtaining the estimates using the Joint bivariate Poisson regression model (JBPM).

C. R-program for obtaining the estimates using the conditional regression model 1 (CM1).

D. SAS-program for obtaining the estimates using the conditional regression model 1 (CM1).

A. R code for obtaining the estimates using the Joint bivariate Poisson regression model (JBPM).

```
library(rootsolve)# rootsolve involves command "multiroot"
health < -read.table("C://Users//Administrator//Documents//He
alth Care Aus. DATA//Heath Care Aus. csv.csv",header = TR
UE,sep = ",")   #Download the file in R
n < -5190
h < -matrix(,nrow = n,ncol = 1)
lambda1 < -matrix(,nrow = n,ncol = 1)
lambda2 < -matrix(,nrow = n,ncol = 1)
lambda3 < -matrix(,nrow = n,ncol = 1)
y1 < -matrix(c(health[,13]), ncol = 1,nrow = n)
y2 < -matrix(c(health[,18]),ncol = 1,nrow = n)
x < -matrix(c(rep(1,n),health[,1],health[,2],health[,4]),nrow = n,
ncol = 4)
b1 < -matrix(,nrow = 4,ncol = 1)   #The initial values of the
parameters
b2 < - matrix(,nrow = 4,ncol = 1)   #The initial values of the
parameters
b3 < - matrix(,nrow = 2,ncol = 1)   #The initial values of the
parameters
my < -pmin(y1,y2) #this gives min(y1,y2)
fn < - function (b)
{
lambda1 < - exp(b[1] + b[2]*x[,2] + b[3]*x[,3] + b[4]*x[,4])
lamnda2 < - exp(b[5] + b[6]*x[,2] + b[7]*x[,3] + b[8]*x[,4])
lambda3 < - exp(b[9] + b[10]*x[,2])
for(i in 1:n){
a1 < - lambda1[i]
a2 < - lambda2[i]
a3 < - lambda3 [i]
n1 < -y1[i]
n2 < -y2[i]
w1 < -0.0
w2 < -0.0
for(z in 0:my[i]) {
w1 < -w1 + (a3^z*a1^(n1-z)*a2^(n2-z)/(gamma(z + 1)*gamma
(n1-z + 1)*gamma(n2-z + 1)))
w2 < -w2 + (a3^z*a1^(n1-z)*a2^(n2-z)*z/(gamma(z + 1)*gamma
(n1-z + 1)*gamma(n2-z + 1)))
}
w3 < -w2/w1
h[i] < -w3
}
u10 = sum(-lamnda1 + y1-h)
u11 = sum((-lamnda1 + y1- h)*x[,2])
u12 = sum((-lamnda1 + y1-h)*x[,3])
u13 = sum((-lamnda1 + y1-h)*x[,4])
u20 = sum(-lamnda2 + y2-h)
u21 = sum((-lamnda2 + y2-h)*x[,2])
u22 = sum((-lamnda2 + y2-h)*x[,3])
```

```
u23 = sum((-lamnda2 + y2-h)*x[,4])
u30 = sum(-lamnda3 + h)
u31 = sum((-lamnda3 + h)*x[,2])
c(u10 = u10,u11 = u11,u12 = u12,u13 = u13, u20 = u20,u21 = u21,
u22 = u22,u23 = u23,u30 = u30,u31 = u31)
}
fit < - multiroot(f = fn, start = c(rep(0,10)))
# calculating log-likelihood function
hod < -c()
w < -c()
lamnda1 < -matrix(,nrow = n,ncol = 1)
lamnda2 < -matrix(,nrow = n,ncol = 1)
lamnda3 < -matrix(,nrow = n,ncol = 1)
root < -matrix(c(fit$root),nrow = 5,ncol = 2)
bhat1 < -matrix(c(root[1,1], root[2,1], root[3,1], root[4,1]),
nrow = 4,ncol = 1)
bhat2 < - matrix(c(root [5,1], root [1,2], root [2,2], root [3,2]),
nrow = 4,ncol = 1)
bhat3 < - matrix(c(root [4,2], root [5,2]),nrow = 2,ncol = 1)
lamnda1 < - exp(bhat1[1,1]*x[,1] + bhat1[2,1]*x[,2] + bhat1[3,1]*x
[,3] + bhat1[4,1]*x[,4])
lamnda2 < - exp(bhat2[1,1]*x[,1] + bhat2[2,1]*x[,2] + bhat2[3,1]*x
[,3] + bhat2[4,1]*x[,4])
lamnda3 < - exp(bhat3[1,1]*x[,1] + bhat3[2,1]*x[,2])
for (k in 1:n){
a1 < - lamnda1 [k]
a2 < - lamnda12[k]
a3 < - lamnda13[k]
x0 < -y1[k]
y0 < -y2[k]
xymin < -min(x0,y0)
lambdaratio < -a3/(a1*a2)
i < -0:xymin
sums < - -lgamma(y1[k]-i + 1)-lgamma(i + 1)-lgamma(y2[k]-i + 1)
 + i*log(lambdaratio)
maxsums < -max(sums)
sums < -sums-maxsums
logsummation < -log(sum(exp(sums))) + maxsums
w[k] < - -sum(a1 + a2 + a3) +  y1[k]*log(a1) +  y2[k]*log(a2) +
logsummation
}
hod < -sum(w)
```

B. SAS code for obtaining the estimates using the Joint bivariate Poisson regression model (JBPM).

```
proc nlmixed data = file name;
parameters b01 = 0 b11 = 0 b12 = 0 b13 = 0 b02 = 0 b21 = 0 b22 = 0
b23 = 0 b03 = 0 b31 = 0;
lambda1 = exp(b01 + b11*sex + b12*age + b13*income);
lambda2 = exp(b02 + b21*sex + b22*age + b23*income);
lambda3 = exp(b03 + b31*sex);
f = 0;
do r = 0 to MIN(doctorco,prescrib);
f + exp(-(lambda1 + lambda2 + lambda3))*(lambda1**doctorco)*
(lambda2**prescrib)*((lambda3/(lambda1*lambda2))**r)/
(fact(doctorco-r)*fact(prescrib-r)*fact(r));
end;
ll = log(f);
model prescrib~general(ll);
run;
```

C. R code for obtaining the estimates using the conditional regression model 1 (CM1).

```
p1 < -matrix(,nrow = n,ncol = 1)
h < -matrix(,nrow = n,ncol = 1)
lambda2 < -matrix(,nrow = n,ncol = 1)
mu < -matrix(,nrow = n,ncol = 1)
y1 < -matrix(c(health[,13]), ncol = 1,nrow = n)
y2 < -matrix(c(health[,18]),ncol = 1,nrow = n)
x < -matrix(c(rep(1,n),health[,1],health[,2],health[,4]),nrow = n,
ncol = 4)
b1 < -matrix(,nrow = 4,ncol = 1)
b2 < - matrix(,nrow = 4,ncol = 1)
my < -pmin(y1,y2) #this gives min(y1,y2)
fn < - function (b)
{
lambda2 < - exp(b[1] + b[2]*x[,2] + b[3]*x[,3] + b[4]*x[,4])
mu < - exp(b[5] + b[6]*x[,2] + b[7]*x[,3] + b[8]*x[,4])
p1 < - mu2/ (1 + mu2)
for(i in 1:n){
a1 < -p1[i]
a2 < -lambda2[i]
n1 < -y1[i]
n2 < -y2[i]
w1 < -0.0
w2 < -0.0
for(z in 0:my[i]) {
db < - dbinom(z,n1,a1)
dp < - dpois(n2-z,a2)
w1 < -w1 + db*dp
w2 < -w2 + z*db*dp
}
w3 < -w2/w1
h[i] < -w3
i = i + 1
}
u10 = sum(-lambda2 + y2-h)
u11 = sum((-lambda2 + y2- h)*x[,2])
u12 = sum((-lambda2 + y2-h)*x[,3])
u13 = sum((-lambda2 + y2-h)*x[,4])
u20 = sum(h- y1* p1)
u21 = sum((h - y1*p1)*x[,2])
u22 = sum((h - y1* p1)*x[,3])
u23 = sum((h - y1* p1)*x[,4])
c(u10 = u10,u11 = u11,u12 = u12,u13 = u13, u20 = u20,u21 = u21,
u22 = u22,u23 = u23)
}
fit < - multiroot(f = fn, start = c(rep(0,8)))
#Calculating log likelihood function
root < -matrix(c(fit$root),nrow = 4,ncol = 2)
hod < -c()
w < -c()
bhat1 < -matrix(c(root[,1]),nrow = 4,ncol = 1)
bhat2 < -matrix(c(root[,2]),nrow = 4,ncol = 1)
lambda2 < - exp(x[,]%*%bhat1)
mu < - exp(x[,]%*%bhat2)
p1 < -mu/(1 + mu)
for (k in 1:n) {
a1 < -p1[k]
a2 < -lambda2[k]
x0 < -y1[k]
y0 < -y2[k]
xymin < -min(x0,y0)
i < -0:xymin
sums < -lgamma(y1[k] + 1)-lgamma(i + 1)-lgamma(y1[k]-i + 1)-
lgamma(y2[k]-i + 1) + i*log(a1) + (y1[k]-i)*log(1-a1)
```

```
 + (y2[k]-i)*log(a2)
maxsums < -max(sums)
sums < -sums-maxsums
logsummation < -log(sum(exp(sums))) + maxsums
w[k] < - -a2 + logsummation
}
hod < -sum(w)
```

D. SAS code for obtaining the estimates using the conditional regression model 1 (CM1).

```
proc nlmixed data = file name;
parameters
b10 = 0 b11 = 0 b12 = 0 b13 = 0
eta10 = 0 eta11 = 0 eta12 = 0 eta13 = 0;
lambda2 = exp(b10 + b11*sex + b12*age + b13*income);
p1 = exp(eta10 + eta11*sex + eta12*age + eta13*income)/(1 + exp
(eta10 + eta11*sex + eta12*age + eta13*income));
f = 0;
do r = 0 to MIN(doctorco,prescrib);
f + fact(doctorco)*(p1**r)*((1-p1)**(doctorco-r))*exp(-lambda2)
*(lambda2**(prescrib-r))/
(fact(r)*fact(doctorco-r)*fact(prescrib-r));
end;
ll = log(f);
model prescrib~general(ll);
run;
```

References

Berkhout, P., Plug, E., 2004. A bivariate Poisson count data model using conditional probabilities. Stat. Neerl. 58, 349–364.

Bermúdez, L., 2009. A priori ratemaking using bivariate Poisson regression models. Insurance Math. Econom. 44, 135–141.

Bermúdez, L., Karlis, D., 2012. A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking. Comput. Stat. Data Anal. 56, 3988–3999.

Cameron, A.C., Trivedi, P.K., Milne, F., Piggott, J., 1988. A microeconometric model of the demand for health care and health insurance in Australia. Rev. Econ. Stud. 55, 85–106.

Chib, S., Winkelmann, R., 2001. Markov chain Monte Carlo analysis of correlated count data. J. Bus. Econ. Statist. 19, 428–435.

Choe, H.G., Lim, J.B., Won, Y., Lee, S., Kim, S.W., 2012. Bayesian analysis for the bivariate Poisson regression model: applications to road safety countermeasures. J. Korean Data Inf. Sci. Soc. 23, 851–858.

Ghitany, M.E., Karlis, D., Al-Mutairi, D.K., Al-Awadhi, F.A., 2012. An EM algorithm for multivariate mixed Poisson regression models and its application. Appl. Math. Sci. 6, 6843–6856.

Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: applications to Poisson models. Econometrica 52, 701–720.

Gurmu, S., Elder, J., 2000. Generalized bivariate count data regression models. Econometr. Lett. 68, 31–36.

Gurmu, S., Elder, J., 2008. A bivariate zero-inflated count data regression model with unrestricted correlation. Econom. Lett. 100, 245–248.

Ho, L.L., Singer, J.M., 2001. Generalized least squares methods for bivariate Poisson regression. Commun. Statist. Theory Method 30, 263–277.

Holgate, P., 1964. Estimation for the bivariate Poisson distribution. Biometrika 51, 241–245.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1997. Discrete Multivariate Distributions. John Wiley & Sons, New York.

Johnson, N.L., Kotz, S., 1969. Discrete Distributions. John Wiley & Sons, New York.

Jung, R.C., Winkelmann, R., 1993. Two aspects of labor mobility: a bivariate Poisson regression approach. Empir. Econom. 18, 543–556.

Karlis, D., Meligkotsidou, L., 2007. Finite mixtures of multivariate Poisson distributions with application. J. Statist. Plann. Inference 137, 1942–1960.

Karlis, D., Ntzoufras, I., 2003. Analysis of sports data by using bivariate Poisson models. Statistician 52, 381–393.

Karlis, D., Ntzoufras, I., 2005. Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R. J. Stat. Softw. 14, 1–36.

Karlis, D., Tsiamyrtzis, P., 2008. Exact Bayesian modeling for bivariate Poisson data and extensions. Statist. Comput. 18, 27–40.

King, G., 1989. A seemingly unrelated Poisson regression model. Sociol. Methods Res. 17, 235–255.

Kocherlakota, S., Kocherlakota, K., 1992. Bivariate Discrete Distributions. Marcel Dekker, New York.

Kocherlakota, S., Kocherlakota, K., 2001. Regression in the bivariate Poisson distribution. Commun. Statist. Theory Method 30, 815–825.

Ma, J., Kockelman, K.M., 2006. Bayesian multivariate Poisson regression for models of injury count, by severity. In: Transportation Research Record: Journal of the Transportation Research Board, No. 1950. Transportation Research Board of the National Academies, Washington, DC, pp. 24–34.

Munkin, M.K., Trivedi, P.K., 1999. Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application. The Econometrics Journal 2, 29–48.

Ozuna, T., Gomez, I.A., 1994. Estimating a system of recreation demand functions using a seemingly unrelated poisson regression approach. The Review of Economics and Statistics 76, 356–360.

Tsionas, E.G., 2001. Bayesian multivariate Poisson regression. Commun. Statist Theory Method 30, 243–255.

Van Ophem, H., 1999. A general method to estimate correlated discrete random variables. Econometr. Theory 15, 228–237.

Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.

Wang, K., Lee, A.H., Yau, K.K.W., Carrivick, P.J.W., 2003. A bivariate zero-inflated Poisson regression model to analyze occupational injuries. Accid. Anal. Prev. 35, 625–629.