Two new bivariate zero-inflated generalized Poisson distributions with a flexible correlation structure

Chi Zhang*, Guoliang Tian, Xifen Huang

Department of Statistics and Actuarial Science, The University of Hong Kong, P. R. China

Abstract   To model correlated bivariate count data with extra zero observations, this paper proposes two new bivariate *zero-inflated generalized Poisson* (ZIGP) distributions by incorporating a multiplicative factor (or dependency parameter) λ, named as Type I and Type II bivariate ZIGP_λ distributions, respectively. The proposed distributions possess a flexible correlation structure and can be used to fit either positively or negatively correlated and either over- or under-dispersed count data, comparing to the existing models that can only fit positively correlated count data with over-dispersion. The two marginal distributions of Type I bivariate ZIGP_λ share a common parameter of zero inflation while the two marginal distributions of Type II bivariate ZIGP_λ have their own parameters of zero inflation, resulting in a much wider range of applications. The important distributional properties are explored and some useful statistical inference methods including maximum likelihood estimations of parameters, standard errors estimation, bootstrap confidence intervals and related testing hypotheses are developed for the two distributions. A real data are thoroughly analyzed by using the proposed distributions and statistical methods. Several simulation studies are conducted to evaluate the performance of the proposed methods.

Keywords   Bivariate generalized Poisson distribution, Expectation–maximization algorithm, Newton–Raphson algorithm, Over-dispersion, Zero-inflated generalized Poisson distribution.

AMS 2010 subject classifications 62Exx, 60E05

DOI: 10.19139/soic.v3i2.133

1. Introduction

As the simplest distribution for modeling count data, the Poisson distribution possesses an exceptional property of equi-dispersion, i.e., its mean and variance are identical. When the observed variance is larger than the theoretical variance, over-dispersion has occurred and it is a very common feature in applications because in practice populations are frequently heterogeneous. In such situation, a useful alternative such as the negative-binomial model with an additional free parameter may provide a better fit. Conversely, under-dispersion means that there is less variation in the data than predicted. To model count data with both over-dispersion and/or under-dispersion, the *generalized Poisson* (GP) distribution can be used and its *probability mass function* (pmf) with two parameters is defined by ([7], p.5)

$$\text{GP}(x|\theta, \alpha) = \frac{\theta^x(1 + \alpha x)^{x-1}}{x!} \exp[-\theta(1 + \alpha x)], \qquad x = 0, 1, 2, \ldots,$$

---

*Correspondence to: Chi Zhang (Email: zhangchi@hku.hk). Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, P. R. China.

in which $\theta > 0$ and the domain of α is $\max(-\theta^{-1}, -m^{-1}) \leqslant \alpha < \theta^{-1}$, where $m \ (\geqslant 4)$ is the largest positive integer for which $1 + \alpha m > 0$ when $\alpha < 0$. And $\Pr(X = x) = 0$ for $x > m$ when $\alpha < 0$. It is denoted by $X \sim \text{GP}(\theta, \alpha)$. Its mean and variance are given by

$$E(X) = \frac{\theta}{1 - \alpha\theta} \quad \text{and} \quad \text{Var}(X) = \frac{\theta}{(1 - \alpha\theta)^3}.$$

Especially, when $\alpha = 0$, $\text{GP}(\theta, \alpha)$ reduces to the Poisson distribution with mean θ.

To model two-dimensional correlated count data, the bivariate Poisson, bivariate negative-binomial and bivariate generalized Poisson models have been developed. By using the method of trivariate reduction, Holgate [15] first proposed the bivariate Poisson distribution and Famoye & Consul [9] then extended it to the more general bivariate generalized Poisson distribution. A major disadvantage for the method of trivariate reduction is that so-generated bivariate distributions can be used to model count data with positive correlation only.

However, two-dimensional count data with negative correlation are often encountered in practice. For example, Table 1 below gives a paired count data of twelve patients who experienced frequent premature ventricular contractions (PVCs) and were administered a drug with antiarrhythmic properties. One-minute EKG recordings were taken before and after the drug administration as shown in the second and third rows of Table 1. To describe the data, we calculated the sample means and variances as $\bar{x} = 13.4167, s_x^2 = 166.9924, \bar{y} = 1.9167, s_y^2 = 14.4470$ and the sample correlation coefficient as $r = -0.2305847$. Both X and Y are obviously over-dispersed and they are negatively correlated. Thus, the aforementioned distributions are not appropriate to fit this data set.

Table 1
The PVC counts for twelve patients one minute after administering a drug with antiarrhythmic properties (Berry, [4])

Patient number	1	2	3	4	5	6	7	8	9	10	11	12
Predrug(X)	6	9	17	22	7	5	5	14	9	7	9	51
Postdrug(Y)	5	2	0	0	2	1	0	0	0	0	13	0

Lakshminarayana *et al.* [20] first proposed a new type of bivariate Poisson distribution, whose joint pmf was expressed as a product of two Poisson margins with a multiplicative factor. Subsequently, a new bivariate negative-binomial regression model and a new bivariate GP distribution are extended in a similar way by Famoye [12, 13]. This kind of expression not only ensures the marginal distribution as the traditional distribution (e.g., Poisson, generalized Poisson, negative-binomial) but also admits positive or negative correlation. The general form of the joint bivariate pmf of (X, Y) is

$$p(x, y) = p_1(x)p_2(y)\Big(1 + \lambda\big\{g_1(x) - E[g_1(X)]\big\}\big\{g_2(y) - E[g_2(Y)]\big\}\Big),$$

where $g_1(x)$ and $g_2(y)$ are two bounded functions, and λ is called the multiplicative factor or dependency parameter. It is easy to verify that $p_1(x)$ and $p_2(y)$ are two marginal pmfs.

When there is a larger frequency of (0,0) observations in the bivariate count data, the zero-inflation phenomenon arises. For example, there are seven zero values of Y out of a total of 12 observations in Table 1. The multivariate zero-inflated Poisson models were developed by using the mixture technique [21] and by using the stochastic representation technique [22]. Their models can only fit over-dispersed data with positive correlation. Krishna & Tukaram [19] provided a bivariate power series distribution that utilized the multiplicative factor form and incorporated inflation at zero. Bivariate zero-inflated Poisson distribution were then discussed as a special case of power series distributions and the associated properties were explored. The univariate marginally follows zero-inflated Poisson distribution which is still over-dispersed, and the correlation can have either sign. All these zero-inflated models cannot comprehensively address the correlation and dispersion issues.

The aim of this paper is to develop two new bivariate *zero-inflated generalized Poisson* (ZIGP) distributions by incorporating a multiplicative factor (or dependency parameter) λ, named as Type I and Type II bivariate

ZIGP_λ distributions, respectively. The two distributions have the following features: (i) Each marginal follows a univariate ZIGP distribution, which is a mixture of the degenerate distribution with all mass at $(0,0)$ point and a GP distribution. A ZIGP distribution can model zero-inflated count data with both over-dispersion and/or under-dispersion. (ii) They possess a flexible correlation structure and can be used to fit either positively or negatively correlated zero-inflated count data, comparing to the existing models that can only fit positively correlated zero-inflated count data. (iii) The two marginal distributions of Type I bivariate ZIGP_λ share a common parameter of zero inflation while the two marginal distributions of Type II bivariate ZIGP_λ have their own parameters of zero inflation, resulting in a much wider range of applications.

The rest of the paper is organized as follows. In Section 2, we propose a new bivariate ZIGP distribution indexed by a multiplicative factor via a mixture of a Bernoulli variable with the bivariate GP distribution of Famoye [13], provide some important distributional properties, maximum likelihood estimations of parameters, standard errors estimation, bootstrap confidence intervals and related testing hypotheses. In Section 3, the Type II bivariate zero-inflated generalized Poisson distribution with a multiplicative factor is also developed. The proposed distributions and statistical methods are applied to analyze the Australian health care utilization data in Section 4. In Section 5, several simulation studies are conducted to evaluate the performance of the proposed methods. Section 6 presents a discussion. Some technical details are put in the Appendix.

2. Type I bivariate zero-inflated generalized Poisson distribution with a multiplicative factor

Famoye [13] introduced a new bivariate GP distribution indexed by an unknown real number λ called the multiplicative factor. Its joint pmf $\Pr(X_1 = x_1, X_2 = x_2)$ is defined by

$$\left\{ \prod_{i=1}^{2} \frac{\theta_i^{x_i}(1 + \alpha_i x_i)^{x_i - 1}}{x_i!} \exp\left[-\theta_i(1 + \alpha_i x_i)\right] \right\} \left[1 + \lambda(e^{-x_1} - c_1)(e^{-x_2} - c_2)\right], \tag{1}$$

for $x_1, x_2 = 0, 1, 2, \ldots$, where $c_i = E(e^{-X_i}) = \exp[\theta_i(s_i - 1)]$ and s_i is a function of (θ_i, α_i) satisfying

$$\log(s_i) - \theta_i \alpha_i(s_i - 1) + 1 = 0, \qquad i = 1, 2.$$

We denote this new bivariate GP distribution by $(X_1, X_2)^\top \sim \text{BGP}_\lambda(\theta_1, \theta_2, \alpha_1, \alpha_2)$, where λ satisfies $|\lambda| \leqslant 1/[(1 - c_1)(1 - c_2)]$. It is easy to verify that $X_i \sim \text{GP}(\theta_i, \alpha_i)$ for $i = 1, 2$, which are irrelevant to the multiplicative factor λ. When $\alpha_1 = \alpha_2 = 0$, this bivariate GP distribution reduces to the bivariate Poisson distribution with the multiplicative factor $\lambda \in \mathbb{R}$ developed by Lakshminarayana *et al.* [20].

When there is a larger frequency of $(0,0)$ observations in the bivariate count data, the issue of the over-dispersion may arise. To model such extra zero points in the data, we propose a new bivariate ZIGP distribution indexed by a multiplicative factor $\lambda \in \mathbb{R}$ via a mixture of a Bernoulli variable with the bivariate GP distribution (1).

Let $Z \sim \text{Bernoulli}(1 - \phi)$, $(X_1, X_2)^\top \sim \text{BGP}_\lambda(\theta_1, \theta_2, \alpha_1, \alpha_2)$ and $Z \perp\!\!\!\perp (X_1, X_2)^\top$. A discrete random vector $(Y_1, Y_2)^\top$ defined by

$$(Y_1, Y_2)^\top = Z(X_1, X_2)^\top, \tag{2}$$

is said to follow the Type I bivariate ZIGP distribution indexed by a multiplicative factor λ, denoted by $(Y_1, Y_2)^\top \sim \text{ZIGP}_\lambda^{(\text{I})}(\phi; \theta_1, \theta_2, \alpha_1, \alpha_2)$, where $\phi \in [0, 1)$ and c_i $(i = 1, 2)$ are defined in (1), the range of λ is to be discussed in Section 2.1. It is easy to derive the joint pmf $\Pr(Y_1 = y_1, Y_2 = y_2)$, which is given by

$$\begin{cases} \phi + (1 - \phi)e^{-(\theta_1 + \theta_2)}[1 + \lambda(1 - c_1)(1 - c_2)], & \text{if } y_1 = y_2 = 0, \\[3mm] (1 - \phi)\left[\prod_{i=1}^{2} \text{GP}(y_i | \theta_i, \alpha_i)\right]\left[1 + \lambda(e^{-y_1} - c_1)(e^{-y_2} - c_2)\right], & \text{otherwise.} \end{cases}$$

Let $(Y_1, Y_2)^\top \sim \text{ZIGP}_\lambda^{(\text{I})}(\phi; \theta_1, \theta_2, \alpha_1, \alpha_2)$, we discuss the following three special cases:

(a)   When $\phi = 0$, we have $(Y_1, Y_2)^\top \overset{\mathrm{d}}{=} (X_1, X_2)^\top \sim \mathrm{BGP}_\lambda(\theta_1, \theta_2, \alpha_1, \alpha_2)$ defined by (1);

(b)   When $\alpha_1 = \alpha_2 = 0$, it reduces to the Type I bivariate ZIP distribution indexed by the multiplicative factor λ, denoted by $(Y_1, Y_2)^\top \sim \mathrm{ZIP}_\lambda^{(\mathrm{I})}(\phi; \theta_1, \theta_2)$. In the circumstances, $s_i = \mathrm{e}^{-1}$ and $c_i = \exp[\theta_i(\mathrm{e}^{-1} - 1)]$ for $i = 1, 2$.

(c)   When $\alpha_1 = \alpha_2 = \lambda = 0$, it reduces to the Type I bivariate ZIP distribution, denoted by $(Y_1, Y_2)^\top \sim \mathrm{ZIP}^{(\mathrm{I})}(\phi; \theta_1, \theta_2)$, which is a special case of the Type I multivariate ZIP distribution recently developed by Liu & Tian [22].

2.1. Marginal distributions, moments and correlation

Let $(Y_1, Y_2)^\top \sim \mathrm{ZIGP}_\lambda^{(\mathrm{I})}(\phi; \theta_1, \theta_2, \alpha_1, \alpha_2)$. From (2), we have $Y_i = ZX_i$ following an univariate zero-inflated generalized Poisson distribution, denoted by

$$Y_i \sim \mathrm{ZIGP}(\phi; \theta_i, \alpha_i), \qquad i = 1, 2. \tag{3}$$

The pmf, expectation and variance of Y_i are given by

$$\Pr(Y_i = y_i) = \begin{cases} \phi + (1 - \phi)\mathrm{e}^{-\theta_i}, & \text{if } y_i = 0, \\[2mm] (1 - \phi)\,\mathrm{GP}(y_i | \theta_i, \alpha_i), & \text{if } y_i > 0, \end{cases}$$

$$E(Y_i) = (1 - \phi)\frac{\theta_i}{1 - \alpha_i\theta_i} \,\hat{=}\, (1 - \phi)\mu_i,$$

$$\mathrm{Var}(Y_i) = (1 - \phi)\left[\frac{\theta_i}{(1 - \alpha_i\theta_i)^3} + \phi\mu_i^2\right],$$

respectively. It is shown that Y_1 and Y_2 can be either over- or under-dispersed.

The covariance between Y_1 and Y_2 is

$$\mathrm{Cov}(Y_1, Y_2) = (1 - \phi)[\phi\mu_1\mu_2 + \lambda(c_{11} - c_1\mu_1)(c_{22} - c_2\mu_2)],$$

where $c_{ii} = E(Y_i\mathrm{e}^{-Y_i}) = \theta_i(1 - \alpha_i\theta_i s_i)^{-1}\exp[\theta_i(1 + \alpha_i)(s_i - 1) - 1]$ for $i = 1, 2$. From (2), we can see that the dependency structure between Y_1 and Y_2 hinges on the common Bernoulli variable Z and the correlation between X_1 and X_2. In fact, the correlation coefficient of Y_1 and Y_2 is

$$\rho_1 = \mathrm{Corr}(Y_1, Y_2) = \frac{\phi\mu_1\mu_2 + \lambda(c_{11} - c_1\mu_1)(c_{22} - c_2\mu_2)}{\sqrt{[\theta_1/(1 - \alpha_1\theta_1)^3 + \phi\mu_1^2]\,[\theta_2/(1 - \alpha_2\theta_2)^3 + \phi\mu_2^2]}}, \tag{4}$$

which could be positive, zero or negative depending on the values of parameters ϕ and λ. If $\phi\mu_1\mu_2 + \lambda(c_{11} - c_1\mu_1)(c_{22} - c_2\mu_2) > (=, <)\, 0$, the correlation is positive (zero, negative). The joint pmf (1) should be nonnegative and the correlation coefficient (4) should satisfy $|\rho_1| \leqslant 1$, thus the lower bound $\lambda_{\mathrm{L},1}$ and upper bound $\lambda_{\mathrm{U},1}$ of the parameter λ are given by

$$\lambda_{\mathrm{L},1} = \max\left(-\frac{1}{(1 - c_1)(1 - c_2)}, \ \min(A_1, B_1)\right) \quad \text{and} \quad \lambda_{\mathrm{U},1} = \max(A_1, B_1),$$

respectively, where

$$A_1 = \frac{\sqrt{[\theta_1/(1 - \alpha_1\theta_1)^3 + \phi\mu_1^2]\,[\theta_2/(1 - \alpha_2\theta_2)^3 + \phi\mu_2^2]} - \phi\mu_1\mu_2}{(c_{11} - c_1\mu_1)(c_{22} - c_2\mu_2)} \quad \text{and}$$

$$B_1 = \frac{-\sqrt{[\theta_1/(1 - \alpha_1\theta_1)^3 + \phi\mu_1^2]\,[\theta_2/(1 - \alpha_2\theta_2)^3 + \phi\mu_2^2]} - \phi\mu_1\mu_2}{(c_{11} - c_1\mu_1)(c_{22} - c_2\mu_2)}.$$

Figure 1 plots the possible range of the correlation coefficient ρ_1 against the range of λ $\left(\lambda \in \left[\lambda_{L,1}, \lambda_{U,1}\right]\right)$ for various combinations of $(\phi, \theta_1, \theta_2, \alpha_1, \alpha_2)$.

Especially, when $\alpha_1 = \alpha_2 = 0$, (4) becomes

$$\text{Corr}(Y_1, Y_2) = \frac{\phi\theta_1\theta_2 + \lambda\theta_1\theta_2 e^{(\theta_1+\theta_2)(e^{-1}-1)}(e^{-1}-1)^2}{\sqrt{(\theta_1 + \phi\theta_1^2)(\theta_2 + \phi\theta_2^2)}}. \tag{5}$$
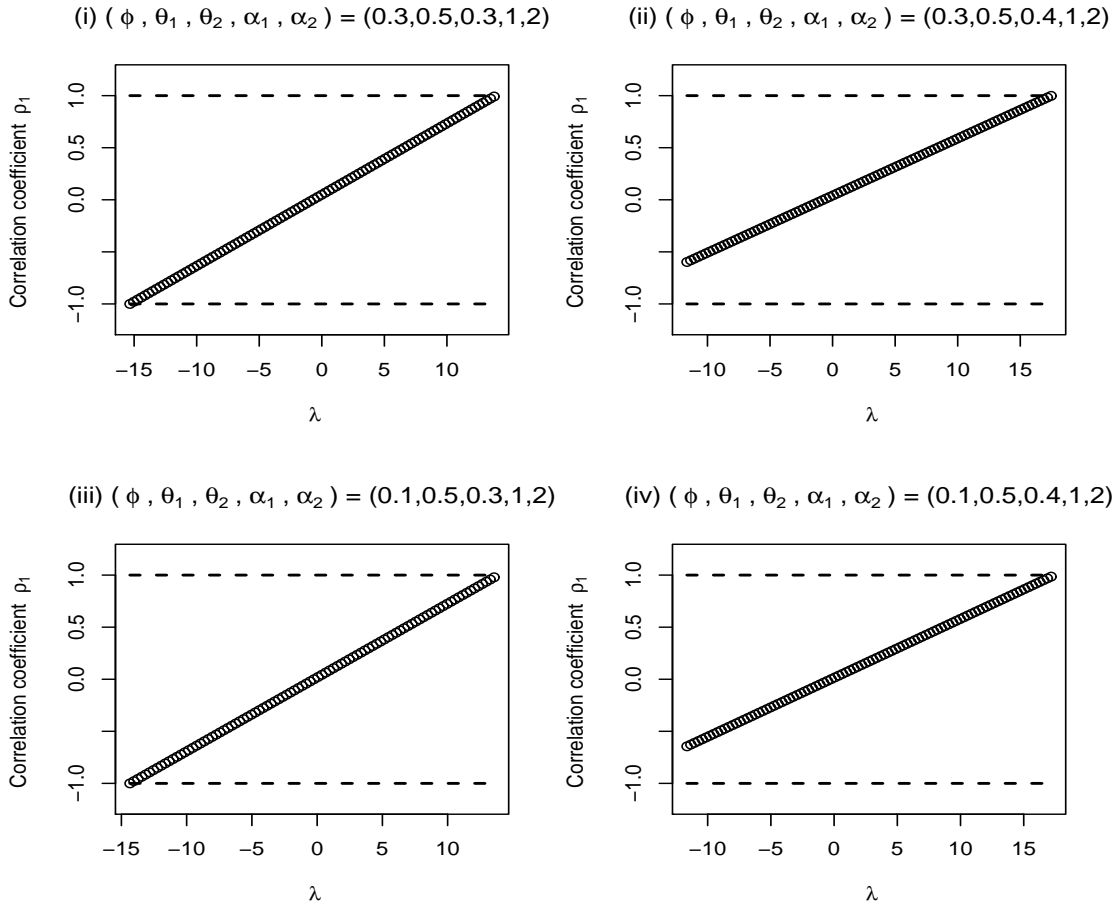
(i) (ϕ , θ_1 , θ_2 , α_1 , α_2) = (0.3,0.5,0.3,1,2)   (ii) (ϕ , θ_1 , θ_2 , α_1 , α_2) = (0.3,0.5,0.4,1,2)

(iii) (ϕ , θ_1 , θ_2 , α_1 , α_2) = (0.1,0.5,0.3,1,2)   (iv) (ϕ , θ_1 , θ_2 , α_1 , α_2) = (0.1,0.5,0.4,1,2)

Figure 1. Range of correlation coefficient ρ_1 against the range of λ for various combinations of $(\phi, \theta_1, \theta_2, \alpha_1, \alpha_2)$ in Type I bivariate ZIGP_λ distribution. (i) $\lambda \in (-15.33, 13.77)$, $\rho_1 \in (-1.00, 0.99)$; (ii) $\lambda \in (-11.63, 17.47)$, $\rho_1 \in (-0.60, 1.00)$; (iii) $\lambda \in (-14.34, 13.56)$, $\rho_1 \in (-1.00, 0.98)$; (iv) $\lambda \in (-11.63, 17.17)$, $\rho_1 \in (-0.64, 0.99)$.

2.2. *Generation of i.i.d. random vectors*

The *stochastic representation* (SR) specified by (2) can be used to generate i.i.d. samples of $(Y_1, Y_2)^\top$ from $\text{ZIGP}_\lambda^{(I)}(\phi; \theta_1, \theta_2, \alpha_1, \alpha_2)$. Since $Z \sim \text{Bernoulli}(1-\phi)$, the key step is how to generate i.i.d. samples of $(X_1, X_2)^\top$ from $\text{BGP}_\lambda(\theta_1, \theta_2, \alpha_1, \alpha_2)$. Note that $X_1 \sim \text{GP}(\theta_1, \alpha_1)$ and the conditional pmf $\Pr(X_2 = x_2 | X_1 = x_1)$ of $X_2 | X_1$ is

$$\text{GP}(x_2 | \theta_2, \alpha_2)\left[1 + \lambda(e^{-x_1} - c_1)(e^{-x_2} - c_2)\right], \quad x_2 = 0, 1, 2, \ldots. \tag{6}$$

Thus, we suggest adopting the conditional sampling method by first generating $X_1 = x_1$ from $\text{GP}(\theta_1, \alpha_1)$, and then generating $X_2 = x_2$ from (6). The built-in R function

```
sample(x, size, replace = FALSE, prob = NULL)
```
can facilitate the simulation.

2.3. *Maximum likelihood estimations of parameters*

Let $(Y_{1j}, Y_{2j})^\top \overset{\text{iid}}{\sim} \text{ZIGP}_\lambda^{(\text{I})}(\phi; \theta_1, \theta_2, \alpha_1, \alpha_2)$ for $j = 1, \ldots, n$, and $Y_{\text{obs}} = \{(y_{1j}, y_{2j})^\top : j = 1, \ldots, n\}$ denote the observation data. Let $\mathbb{J} = \{j : (y_{1j}, y_{2j})^\top = \mathbf{0}, \, j = 1, \ldots, n\}$ and m_0 denote the number of elements in $\mathbb{J}$. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \alpha_1, \alpha_2, \lambda)^\top$, then the observed-data likelihood function is

$$
\begin{aligned}
L(\phi, \boldsymbol{\theta}|Y_{\text{obs}}) &= \left\{ \phi + (1-\phi)\mathrm{e}^{-(\theta_1+\theta_2)}[1 + \lambda(1-c_1)(1-c_2)] \right\}^{m_0} \\
&\quad \times \prod_{j \notin \mathbb{J}} \left\{ (1-\phi) \left[\prod_{i=1}^2 \text{GP}(y_{ij}|\theta_i, \alpha_i) \right] [1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(\mathrm{e}^{-y_{2j}} - c_2)] \right\},
\end{aligned}
$$

To obtain the MLEs of the parameters, we employ the EM algorithm. A latent variable Z is introduced to split m_0 into Z and $m_0 - Z$, so that the conditional predictive distribution of Z given Y_{obs} and $(\phi, \boldsymbol{\theta})$ is

$$
Z|(Y_{\text{obs}}, \phi, \boldsymbol{\theta}) \sim \text{Binomial}\left(m_0, \ \frac{\phi}{\phi + (1-\phi)\mathrm{e}^{-(\theta_1+\theta_2)}[1 + \lambda(1-c_1)(1-c_2)]} \right). \tag{7}
$$

On the other hand, the complete-data likelihood function is proportional to

$$
\begin{aligned}
L(\phi, \boldsymbol{\theta}|Y_{\text{com}}) &\propto \phi^z(1-\phi)^{m_0-z}\mathrm{e}^{-(m_0-z)(\theta_1+\theta_2)}[1 + \lambda(1-c_1)(1-c_2)]^{m_0-z}(1-\phi)^{n-m_0} \\
&\quad \times \prod_{j \notin \mathbb{J}} \left\{ \left[\prod_{i=1}^2 \text{GP}(y_{ij}|\theta_i, \alpha_i) \right] [1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(\mathrm{e}^{-y_{2j}} - c_2)] \right\} \\
&= L_1(\phi|Y_{\text{com}}) \cdot L_2(\boldsymbol{\theta}|Y_{\text{com}}), \tag{8}
\end{aligned}
$$

where $L_1(\phi|Y_{\text{com}}) = \phi^z(1-\phi)^{n-z}$ only involves ϕ and

$$
\begin{aligned}
L_2(\boldsymbol{\theta}|Y_{\text{com}}) &= \mathrm{e}^{-(m_0-z)(\theta_1+\theta_2)}[1 + \lambda(1-c_1)(1-c_2)]^{m_0-z} \\
&\quad \times \prod_{j \notin \mathbb{J}} \left\{ \left[\prod_{i=1}^2 \text{GP}(y_{ij}|\theta_i, \alpha_i) \right] [1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(\mathrm{e}^{-y_{2j}} - c_2)] \right\} \\
&= \mathrm{e}^{(\theta_1+\theta_2)z}[1 + \lambda(1-c_1)(1-c_2)]^{-z} \\
&\quad \times \prod_{j=1}^n \left\{ \left[\prod_{i=1}^2 \text{GP}(y_{ij}|\theta_i, \alpha_i) \right] [1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(\mathrm{e}^{-y_{2j}} - c_2)] \right\}
\end{aligned}
$$

only involves $\boldsymbol{\theta}$. Obviously, the complete-data MLE of ϕ is $\hat{\phi} = z/n$. The logarithm of $L_2(\boldsymbol{\theta}|Y_{\text{com}})$ is given by

$$
\begin{aligned}
\ell_2(\boldsymbol{\theta}|Y_{\text{com}}) &= z(\theta_1 + \theta_2) + N_1 \log\theta_1 + N_2 \log\theta_2 + \sum_{j=1}^n (y_{1j} - 1)\log(1 + \alpha_1 y_{1j}) \\
&\quad + \sum_{j=1}^n (y_{2j} - 1)\log(1 + \alpha_2 y_{2j}) - \sum_{j=1}^n \theta_1(1 + \alpha_1 y_{1j}) - \sum_{j=1}^n \theta_2(1 + \alpha_2 y_{2j}) \\
&\quad - z\log[1 + \lambda(1-c_1)(1-c_2)] + \sum_{j=1}^n \log[1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(\mathrm{e}^{-y_{2j}} - c_2)], \tag{9}
\end{aligned}
$$

where $N_i = \sum_{j=1}^{n} y_{ij}$ for $i = 1, 2$. Since the complete-data MLEs of $\boldsymbol{\theta}$ are not available in closed form, we adopt the Newton–Raphson algorithm to calculate the complete-data MLEs of $\boldsymbol{\theta}$. The score vector and the Hessian matrix

$$\nabla \ell_2(\boldsymbol{\theta}|Y_{\mathrm{com}}) = \frac{\partial \ell_2(\boldsymbol{\theta}|Y_{\mathrm{com}})}{\partial \boldsymbol{\theta}} \quad \text{and} \quad \nabla^2 \ell_2(\boldsymbol{\theta}|Y_{\mathrm{com}}) = \frac{\partial^2 \ell_2(\boldsymbol{\theta}|Y_{\mathrm{com}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \tag{10}$$

are derived in Appendix A. Thus, the M-step is to calculate

$$\phi^{(t+1)} = \frac{z^{(t)}}{n} \quad \text{and} \quad \boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} + [-\nabla^2 \ell_2(\boldsymbol{\theta}^{(t)}|Y_{\mathrm{com}})]^{-1} \nabla \ell_2(\boldsymbol{\theta}^{(t)}|Y_{\mathrm{com}}), \tag{11}$$

The E-step is to replace $z^{(t)}$ in (11) by the conditional expectation

$$E(Z|Y_{\mathrm{obs}}, \phi^{(t)}, \boldsymbol{\theta}^{(t)}) = \frac{m_0 \phi^{(t)}}{\phi^{(t)} + (1 - \phi^{(t)}) \mathrm{e}^{-(\theta_1^{(t)} + \theta_2^{(t)})} [1 + \lambda^{(t)}(1 - c_1^{(t)})(1 - c_2^{(t)})]}, \tag{12}$$

where $c_i^{(t)} = \exp[\theta_i^{(t)}(s_i^{(t)} - 1)]$ and $s_i^{(t)}$ satisfies $\log[s_i^{(t)}] - \theta_i^{(t)} \alpha_i^{(t)}(s_i^{(t)} - 1) + 1 = 0$ for $i = 1, 2$.

2.4. Estimation of standard errors

In this subsection, we are interested in deriving the standard deviations of the MLEs $\hat{\phi}$ and $\hat{\boldsymbol{\theta}}$ of the parameters. Let $\ell(\phi, \boldsymbol{\theta}|Y_{\mathrm{com}}) = \log L(\phi, \boldsymbol{\theta}|Y_{\mathrm{com}})$ denote the logarithm of the complete-data likelihood function (8). Then, its first and second derivatives are given by

$$\nabla \ell(\phi, \boldsymbol{\theta}|Y_{\mathrm{com}}) = \left(\nabla \ell_1(\phi|Y_{\mathrm{com}}), \nabla^\top \ell_2(\boldsymbol{\theta}|Y_{\mathrm{com}})\right)^\top,$$

$$\nabla^2 \ell(\phi, \boldsymbol{\theta}|Y_{\mathrm{com}}) = \begin{pmatrix} \nabla^2 \ell_1(\phi|Y_{\mathrm{com}}) & \mathbf{0}_5^\top \\ \mathbf{0}_5 & \nabla^2 \ell_2(\boldsymbol{\theta}|Y_{\mathrm{com}}) \end{pmatrix},$$

where

$$\nabla \ell_1(\phi|Y_{\mathrm{com}}) = \frac{z - n\phi}{\phi(1 - \phi)}, \qquad \nabla^2 \ell_1(\phi|Y_{\mathrm{com}}) = -\frac{z}{\phi^2} - \frac{n - z}{(1 - \phi)^2},$$

$\nabla \ell_2(\boldsymbol{\theta}|Y_{\mathrm{com}})$ and $\nabla^2 \ell_2(\boldsymbol{\theta}|Y_{\mathrm{com}})$ are given by (10).

According to Louis [23], the observed information $\mathbf{I}(\hat{\phi}, \hat{\boldsymbol{\theta}}|Y_{\mathrm{obs}})$ can be expressed as

$$\left\{ E\{-\nabla^2 \ell(\phi, \boldsymbol{\theta}|Y_{\mathrm{com}})|Y_{\mathrm{obs}}, \phi, \boldsymbol{\theta}\} - E\{[\nabla \ell(\phi, \boldsymbol{\theta}|Y_{\mathrm{com}})]^{\otimes 2}|Y_{\mathrm{obs}}, \phi, \boldsymbol{\theta}\} \right\}\Big|_{(\phi, \boldsymbol{\theta}) = (\hat{\phi}, \hat{\boldsymbol{\theta}})}, \tag{13}$$

where $x^{\otimes 2} = x x^\top$. The computation of the conditional expectations in (13) involves the conditional predictive distribution (7). We have

$$E(Z|Y_{\mathrm{obs}}, \phi, \boldsymbol{\theta}) = m_0 p \quad \text{and} \quad E(Z^2|Y_{\mathrm{obs}}, \phi, \boldsymbol{\theta}) = m_0 p(1 - p) + m_0^2 p^2,$$

where

$$p \hat{=} \frac{\phi}{\phi + (1 - \phi)\mathrm{e}^{-(\theta_1 + \theta_2)}[1 + \lambda(1 - c_1)(1 - c_2)]}.$$

The estimated standard errors are the square roots of the diagonal elements of the inverse observed information matrix $\mathbf{I}^{-1}(\hat{\phi}, \hat{\boldsymbol{\theta}}|Y_{\mathrm{obs}})$. Finally, we can use these estimated standard errors to construct the $(1 - \alpha)100\%$ asymptotic Wald *confidence interval* (CIs) for the parameters $(\phi, \boldsymbol{\theta})$.

Alternatively, when the observed information matrix is too complicated, we may use the square roots of the diagonal elements of the inverse complete information matrix to approximate the estimated standard errors. The complete information matrix is defined by

$$\mathbf{I}(\hat{\phi}, \hat{\boldsymbol{\theta}}|Y_{\mathrm{com}}) = E\left\{-\nabla^2 \ell(\phi, \boldsymbol{\theta}|Y_{\mathrm{com}})|Y_{\mathrm{obs}}, \phi, \boldsymbol{\theta}\right\}\Big|_{(\phi, \boldsymbol{\theta}) = (\hat{\phi}, \hat{\boldsymbol{\theta}})}. \tag{14}$$

2.5. Bootstrap confidence intervals

The Wald CIs are reliable only for large sample sizes and the lower/upper bound may exceed the natural bound of parameters. For example, the Wald CI of ϕ may fall outside the unit interval $(0, 1]$.

The bootstrap method is useful in calculating the bootstrap CIs, which are reliable for small to moderate sample sizes. For an arbitrary function of $(\phi, \boldsymbol{\theta})$, say $\vartheta = h(\phi, \boldsymbol{\theta})$, the bootstrap method can be used to obtain the bootstrap CI of ϑ. Let $(\hat{\phi}, \hat{\boldsymbol{\theta}})$ denote the MLEs of $(\phi, \boldsymbol{\theta})$ obtained by the EM algorithm embedded with the Newton–Raphson algorithm specified by (11)–(12), then $\hat{\vartheta} = h(\hat{\phi}, \hat{\boldsymbol{\theta}})$ is the MLE of ϑ. Based on the obtained MLEs $(\hat{\phi}, \hat{\boldsymbol{\theta}}) = (\hat{\phi}, \hat{\theta}_1, \hat{\theta}_2, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\lambda})$, we can generate

$$(Y_{1j}^*, Y_{2j}^*)^\top \overset{\text{iid}}{\sim} \text{ZIGP}_{\hat{\lambda}}^{(\text{I})}(\hat{\phi}; \hat{\theta}_1, \hat{\theta}_2, \hat{\alpha}_1, \hat{\alpha}_2)$$

for $j = 1, \ldots, n$ using the conditional sampling technique presented in Section 2.2. Based on the bootstrap sample $Y_{\text{obs}}^* = \left\{ (y_{1j}^*, y_{2j}^*)^\top \right\}_{j=1}^n$, we first compute the MLEs $(\hat{\phi}^*, \hat{\boldsymbol{\theta}}^*)$ and then obtain a bootstrap replication $\hat{\vartheta}^* = h(\hat{\phi}^*, \hat{\boldsymbol{\theta}}^*)$. Independently repeating the above process G times, we obtain G bootstrap samples $\{Y_{\text{obs}}^*(g)\}_{g=1}^G$ and G bootstrap replications $\{\hat{\vartheta}_g^*\}_{g=1}^G$. The standard error of $\hat{\vartheta}$, $\text{se}(\hat{\vartheta})$, can be estimated by the sample standard deviation of the G replications, i.e.,

$$\widehat{\text{se}}(\hat{\vartheta}) = \left\{ \frac{1}{G-1} \sum_{g=1}^G [\hat{\vartheta}_g^* - (\hat{\vartheta}_1^* + \cdots + \hat{\vartheta}_G^*)/G]^2 \right\}^{1/2}. \tag{15}$$

If $\{\hat{\vartheta}_g^*\}_{g=1}^G$ is approximately normally distributed, the $(1 - \alpha)100\%$ simple bootstrap CI for ϑ is

$$\left[\hat{\vartheta} - z_{\alpha/2} \cdot \widehat{\text{se}}(\hat{\vartheta}), \ \hat{\vartheta} + z_{\alpha/2} \cdot \widehat{\text{se}}(\hat{\vartheta}) \right]. \tag{16}$$

Alternatively, if $\{\hat{\vartheta}_g^*\}_{g=1}^G$ is non-normally distributed, the $(1 - \alpha)100\%$ bootstrap percentile CI of ϑ is given by

$$\left[\hat{\vartheta}_{\text{L}}, \ \hat{\vartheta}_{\text{U}} \right], \tag{17}$$

where $\hat{\vartheta}_{\text{L}}$ and $\hat{\vartheta}_{\text{U}}$ are the $100(\alpha/2)$ and $100(1 - \alpha/2)$ percentiles of $\{\hat{\vartheta}_g^*\}_{g=1}^G$, respectively.

Note that the CI in (16) is based on the normality assumption, while (17) provide accurate confidence intervals without making the normality assumption. However, (17) does not adjust the confidence interval to account for skewness in the underlying population or other errors that can result in where $\hat{\vartheta}$ is not the sample mean. Thus, we consider the bootstrap percentile-t CI, which can adjust such errors. Based on the generated G bootstrap samples $\{Y_{\text{obs}}^*(g)\}_{g=1}^G$, we compute $t_g^* = (\hat{\vartheta}_g^* - \hat{\vartheta})/\widehat{\text{se}}_g^*$ for $g = 1, \ldots, G$, where $\widehat{\text{se}}_g^*$ is the estimated standard error of the $\hat{\vartheta}_g^*$ in the g-th bootstrap sample $Y_{\text{obs}}^*(g)$. The $(1 - \alpha)100\%$ bootstrap percentile-t CI for ϑ is given by [8]

$$\left[\hat{\vartheta} - \hat{t}_{1-\alpha/2} \cdot \widehat{\text{se}}(\hat{\vartheta}), \ \hat{\vartheta} - \hat{t}_{\alpha/2} \cdot \widehat{\text{se}}(\hat{\vartheta}) \right], \tag{18}$$

where $\hat{t}_{\alpha/2}$ and $\hat{t}_{1-\alpha/2}$ are the $100(\alpha/2)$ and $100(1 - \alpha/2)$ percentiles of $\{t_g^*\}_{g=1}^G$, respectively.

2.6. Testing hypotheses

2.6.1. Likelihood ratio test for zero inflation. Suppose we want to test the null hypothesis

$$H_0 \colon \phi = 0 \quad \text{against} \quad H_1 \colon \phi > 0. \tag{19}$$

Under H_0, the *likelihood ratio test* (LRT) statistic

$$T_1 = -2\{\ell(0, \hat{\boldsymbol{\theta}}_{H_0} | Y_{\text{obs}}) - \ell(\hat{\phi}, \hat{\boldsymbol{\theta}} | Y_{\text{obs}})\}, \tag{20}$$

where $(0, \hat{\boldsymbol{\theta}}_{H_0})$ denote the constrained MLEs of $(\phi, \boldsymbol{\theta})$ under H_0 and $(\hat{\phi}, \hat{\boldsymbol{\theta}})$ denote the unconstrained MLEs of $(\phi, \boldsymbol{\theta})$, which are obtained by the algorithm specified in (11)–(12). Since the null hypothesis in (19) corresponds to ϕ being on the boundary of the parameter space and the appropriate null distribution is a 50:50 mixture of $\chi^2(0)$ (i.e., the degenerate distribution with all mass at zero) and $\chi^2(1)$, see Self & Liang [25] and Feng & McCulloch [14]. Hence, the p-value ([16], p.78; [17], p.225) is

$$p_{\mathrm{v1}} = \Pr(T_1 > t_1 | H_0) = \frac{1}{2} \Pr(\chi^2(1) > t_1). \tag{21}$$

2.6.2. LRT for independency parameter. If $\lambda = 0$, the Type I bivariate ZIGP random vector reduces to the bivariate random vector that is the product of a common zero-inflated Bernoulli variable Z and a random vector with two independent GP random components. Correlation between Y_1 and Y_2 in (4) becomes

$$\mathrm{Corr}(Y_1, Y_2) = \frac{\phi \mu_1 \mu_2}{\sqrt{\left[\frac{\theta_1}{(1 - \alpha_1 \theta_1)^3} + \phi \mu_1^2\right]\left[\frac{\theta_2}{(1 - \alpha_2 \theta_2)^3} + \phi \mu_2^2\right]}},$$

which must be non-negative. Therefore, we would like to test

$$H_0\colon \lambda = 0 \quad \text{against} \quad H_1\colon \lambda \neq 0. \tag{22}$$

The corresponding LRT statistic is given by

$$T_2 = -2\{\ell(\hat{\phi}_{H_0}, \hat{\theta}_{1,H_0}, \hat{\theta}_{2,H_0}, \hat{\alpha}_{1,H_0}, \hat{\alpha}_{2,H_0}, 0 | Y_{\mathrm{obs}}) - \ell(\hat{\phi}, \hat{\boldsymbol{\theta}} | Y_{\mathrm{obs}})\}. \tag{23}$$

Under H_0, T_2 approximately follows the $\chi^2(1)$ distribution. The corresponding p-value is

$$p_{\mathrm{v2}} = \Pr(T_2 > t_2 | H_0) = \Pr(\chi^2(1) > t_2). \tag{24}$$

2.6.3. LRT of Type I bivariate ZIGP$_\lambda$ against Type I bivariate ZIP$_\lambda$. We have mentioned that when the parameters $\alpha_1 = \alpha_2 = 0$, the Type I bivariate ZIGP distribution indexed by the multiplicative factor λ reduces to the Type I bivariate ZIP distribution indexed by the multiplicative factor λ. We test the following null hypothesis

$$H_0\colon \alpha_1 = \alpha_2 = 0 \quad \text{against} \quad H_1\colon H_0 \text{ is not true.} \tag{25}$$

To assess the adequacy of the Type I bivariate ZIGP over the Type I bivariate ZIP. The LRT statistic is

$$T_3 = -2\{\ell(\hat{\phi}_{H_0}, \hat{\theta}_{1,H_0}, \hat{\theta}_{2,H_0}, 0, 0, \hat{\lambda}_{H_0} | Y_{\mathrm{obs}}) - \ell(\hat{\phi}, \hat{\boldsymbol{\theta}} | Y_{\mathrm{obs}})\}. \tag{26}$$

Note that under H_0, the MLEs of parameters in $\mathrm{ZIP}_\lambda^{(\mathrm{I})}(\phi; \theta_1, \theta_2)$ can be obtained through an EM algorithm embedded with a Newton–Raphson algorithm in a similar way as that for the Type I bivariate ZIGP distribution. Under H_0, the test statistic T_3 follows $\chi^2(2)$, thus the corresponding p-value is

$$p_{\mathrm{v3}} = \Pr(T_3 > t_3 | H_0) = \Pr(\chi^2(2) > t_3). \tag{27}$$

3. Type II bivariate zero-inflated generalized Poisson distribution with a multiplicative factor

According to (3), two marginal distributions of the Type I bivariate ZIGP distribution with a multiplicative factor are zero-inflated GP and are forced into sharing a common parameter of zero inflation. This limitation evokes us to extend the Type I bivariate ZIGP to a Type II bivariate ZIGP distribution with a multiplicative factor in which two different parameters of zero inflation are allowed.

Let $\{Z_i\}_{i=1}^2 \overset{\mathrm{ind}}{\sim} \mathrm{Bernoulli}\,(1 - \phi_i)$, $(X_1, X_2)^\top \sim \mathrm{BGP}_\lambda(\theta_1, \theta_2, \alpha_1, \alpha_2)$, and $(Z_1, Z_2, (X_1, X_2)^\top)$ are mutually independent. A discrete random vector $(Y_1, Y_2)^\top$ defined by

$$(Y_1, Y_2)^\top = (Z_1 X_1, Z_2 X_2)^\top \tag{28}$$

is said to follow the Type II bivariate ZIGP distribution indexed by a multiplicative factor λ, denoted by $(Y_1, Y_2)^\top \sim$ $\text{ZIGP}_\lambda^{(\text{II})}(\phi_1, \phi_2; \theta_1, \theta_2, \alpha_1, \alpha_2)$. In Appendix C, the joint pmf $\Pr(Y_1 = y_1, Y_2 = y_2)$ will be derived as

$$
\begin{cases}
\phi_1\phi_2 + \phi_1(1-\phi_2)\mathrm{e}^{-\theta_2} + (1-\phi_1)\phi_2\mathrm{e}^{-\theta_1} \\
\quad + (1-\phi_1)(1-\phi_2)\mathrm{e}^{-(\theta_1+\theta_2)}[1+\lambda(1-c_1)(1-c_2)], & \text{if} \quad y_1 = 0, y_2 = 0, \\[2mm]
(1-\phi_2)\text{GP}\left(y_2|\theta_2,\alpha_2\right) \cdot \left\{ \phi_1 + (1-\phi_1)\mathrm{e}^{-\theta_1}[1+\lambda(1-c_1)(\mathrm{e}^{-y_2}-c_2)] \right\}, & \text{if} \quad y_1 = 0, y_2 > 0, \\[2mm]
(1-\phi_1)\text{GP}\left(y_1|\theta_1,\alpha_1\right) \cdot \left\{ \phi_2 + (1-\phi_2)\mathrm{e}^{-\theta_2}[1+\lambda(\mathrm{e}^{-y_1}-c_1)(1-c_2)] \right\}, & \text{if} \quad y_1 > 0, y_2 = 0, \\[2mm]
(1-\phi_1)(1-\phi_2)\text{GP}\left(y_1|\theta_1,\alpha_1\right)\text{GP}\left(y_2|\theta_2,\alpha_2\right) \cdot [1+\lambda(\mathrm{e}^{-y_1}-c_1)(\mathrm{e}^{-y_2}-c_2)], & \text{if} \quad y_1 > 0, y_2 > 0.
\end{cases}
\tag{29}
$$

In particular, $\phi_1 = 0$ indicates that only the Y_2 marginally follows a zero-inflated GP distribution. Similarly, $\phi_2 = 0$ indicates that only the Y_1 marginally follows a zero-inflated GP distribution. Similar to Section 2.2, we can generate i.i.d. random vectors from (28) by using the conditional sampling technique.

3.1. Marginal distributions, moments and correlation

From (28), we have $Y_i = Z_i X_i$, so that $Y_i \sim \text{ZIGP}(\phi_i; \theta_i, \alpha_i)$ for $i = 1, 2$. Thus, Y_1 and Y_2 have different zero inflation parameters ϕ_1 and ϕ_2. The moments Y_i $(i = 1, 2)$ are given by

$$
\begin{cases}
E(Y_i) &= (1-\phi_i)\dfrac{\theta_i}{1-\alpha_i\theta_i} \,\hat{=}\, (1-\phi_i)\mu_i, \\[3mm]
\text{Var}(Y_i) &= (1-\phi_i)\left[\dfrac{\theta_i}{(1-\alpha_i\theta_i)^3} + \phi_i\mu_i^2 \right], \\[3mm]
\text{Cov}(Y_1, Y_2) &= (1-\phi_1)(1-\phi_2)\lambda(c_{11} - c_1\mu_1)(c_{22} - c_2\mu_2),
\end{cases}
$$

where $c_{ii} = E(Y_i\mathrm{e}^{-Y_i}) = \theta_i(1-\alpha_i\theta_i s_i)^{-1}\exp[\theta_i(1+\alpha_i)(s_i - 1) - 1]$ for $i = 1, 2$. Thus, the correlation coefficient between Y_1 and Y_2 is

$$
\rho_2 = \text{Corr}(Y_1, Y_2) = \frac{\sqrt{(1-\phi_1)(1-\phi_2)}\lambda(c_{11} - c_1\mu_1)(c_{22} - c_2\mu_2)}{\sqrt{[\theta_1/(1-\alpha_1\theta_1)^3 + \phi_1\mu_1^2][\theta_2/(1-\alpha_2\theta_2)^3 + \phi_2\mu_2^2]}}.
\tag{30}
$$

Similarly, the range of the parameter λ, $\left[\lambda_{\text{L},2}, \lambda_{\text{U},2}\right]$, is given by

$$
\lambda_{\text{L},2} = \max\left(-\frac{1}{(1-c_1)(1-c_2)}, \; \min(A_2, B_2) \right) \quad \text{and} \quad \lambda_{\text{U},2} = \max(A_2, B_2),
$$

respectively, where

$$
\begin{aligned}
A_2 &= \frac{\sqrt{[\theta_1/(1-\alpha_1\theta_1)^3 + \phi\mu_1^2][\theta_2/(1-\alpha_2\theta_2)^3 + \phi\mu_2^2]}}{\sqrt{(1-\phi_1)(1-\phi_2)}(c_{11} - c_1\mu_1)(c_{22} - c_2\mu_2)} \quad \text{and} \\[3mm]
B_2 &= \frac{-\sqrt{[\theta_1/(1-\alpha_1\theta_1)^3 + \phi\mu_1^2][\theta_2/(1-\alpha_2\theta_2)^3 + \phi\mu_2^2]}}{\sqrt{(1-\phi_1)(1-\phi_2)}(c_{11} - c_1\mu_1)(c_{22} - c_2\mu_2)}.
\end{aligned}
$$

When $\phi_1 = 0$ or $\phi_2 = 0$, the corresponding correlation coefficient becomes

$$
\text{Corr}(Y_1, Y_2) = \frac{\sqrt{1 - \phi_2} \lambda (c_{11} - c_1 \mu_1)(c_{22} - c_2 \mu_2)}{\sqrt{[\theta_1/(1 - \alpha_1 \theta_1)^3] [\theta_2/(1 - \alpha_2 \theta_2)^3 + \phi_2 \mu_2^2]}} \quad \text{or}
$$
$$
\text{Corr}(Y_1, Y_2) = \frac{\sqrt{1 - \phi_1} \lambda (c_{11} - c_1 \mu_1)(c_{22} - c_2 \mu_2)}{\sqrt{[\theta_1/(1 - \alpha_1 \theta_1)^3 + \phi_1 \mu_1^2] [\theta_2/(1 - \alpha_2 \theta_2)^3]}} .
$$

(31)

From both (30) and (31), it can be seen that the correlation coefficient between Y_1 and Y_2 could be either positive or negative.

Figure 2 plots the possible range of the correlation coefficient ρ_2 against the range of λ for four various combinations of (ϕ_1, ϕ_2) with $(\theta_1, \theta_2, \alpha_1, \alpha_2) = (0, 6, 0.2, 1, 2)$.

Figure 2. Range of correlation coefficient ρ_2 against the range of λ for four various combinations of (ϕ_1, ϕ_2) with $(\theta_1, \theta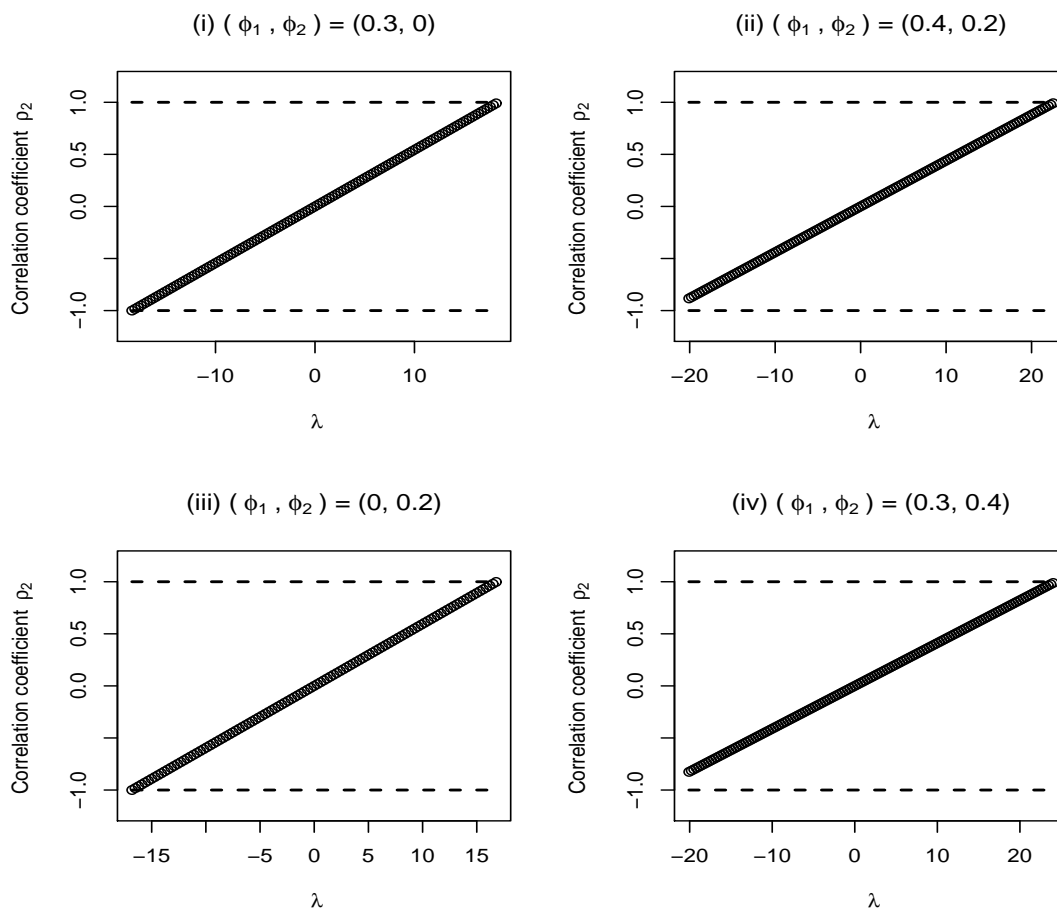_2, \alpha_1, \alpha_2) = (0, 6, 0.2, 1, 2)$ in Type II bivariate ZIGP$_\lambda$ distribution. (i) $\lambda \in (-18.39, 18.39)$, $\rho_2 \in (-1.00, 0.99)$; (ii) $\lambda \in (-20.07, 22.53)$, $\rho_2 \in (-0.88, 0.99)$; (iii) $\lambda \in (-16.81, 16.79)$, $\rho_2 \in (-1.00, 1.00)$; (iv) $\lambda \in (-20.07, 24.03)$, $\rho_2 \in (-0.83, 0.99)$.

3.2. MLEs of parameters

Let $(Y_{1j}, Y_{2j})^\top \overset{\text{iid}}{\sim} \text{ZIGP}_\lambda^{(\text{II})}(\phi_1, \phi_2; \theta_1, \theta_2, \alpha_1, \alpha_2)$ for $j = 1, \ldots, n$ and $Y_{\text{obs}} = \{(y_{1j}, y_{2j})^\top : j = 1, \ldots, n\}$. Define

$$
\begin{aligned}
\mathbb{J}_0 &= \{j : y_{1j} = 0, \ y_{2j} = 0, \ j = 1, \ldots, n\}, \quad m_0 = \#\{\mathbb{J}_0\}, \\
\mathbb{J}_1 &= \{j : y_{1j} = 0, \ y_{2j} > 0, \ j = 1, \ldots, n\}, \quad m_1 = \#\{\mathbb{J}_1\}, \\
\mathbb{J}_2 &= \{j : y_{1j} > 0, \ y_{2j} = 0, \ j = 1, \ldots, n\}, \quad m_2 = \#\{\mathbb{J}_2\}, \\
\mathbb{J}_3 &= \{j : y_{1j} > 0, \ y_{2j} > 0, \ j = 1, \ldots, n\}, \quad m_3 = \#\{\mathbb{J}_3\} = n - m_0 - m_1 - m_2.
\end{aligned}
$$

Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \alpha_1, \alpha_2, \lambda)^\top$, the observed-data likelihood function is

$$
\begin{aligned}
L(\phi_1, \phi_2, \boldsymbol{\theta} | Y_{\text{obs}}) &= \left\{ \phi_1\phi_2 + \phi_1(1 - \phi_2)e^{-\theta_2} + (1 - \phi_1)\phi_2 e^{-\theta_1} \right. \\
&\quad \left. + (1 - \phi_1)(1 - \phi_2)e^{-(\theta_1 + \theta_2)}[1 + \lambda(1 - c_1)(1 - c_2)] \right\}^{m_0} \\
&\quad \times \prod_{j \in \mathbb{J}_1} (1 - \phi_2)\text{GP}(y_{2j} | \theta_2, \alpha_2) \left\{ \phi_1 + (1 - \phi_1)e^{-\theta_1}[1 + \lambda(1 - c_1)(e^{-y_{2j}} - c_2)] \right\} \\
&\quad \times \prod_{j \in \mathbb{J}_2} (1 - \phi_1)\text{GP}(y_{1j} | \theta_1, \alpha_1) \left\{ \phi_2 + (1 - \phi_2)e^{-\theta_2}[1 + \lambda(e^{-y_{1j}} - c_1)(1 - c_2)] \right\} \\
&\quad \times \prod_{j \in \mathbb{J}_3} (1 - \phi_1)(1 - \phi_2)\text{GP}(y_{1j} | \theta_1, \alpha_1)\text{GP}(y_{2j} | \theta_2, \alpha_2)[1 + \lambda(e^{-y_{1j}} - c_1)(e^{-y_{2j}} - c_2)].
\end{aligned}
$$

We adopt the EM algorithm to obtain the MLEs of parameters. First we augment Y_{obs} with latent variables $\mathbf{u} = (U_0, U_1, U_2, U_3)^\top$ that split m_0 into $U_0 + U_1 + U_2 + U_3$ with $U_3 = m_0 - U_0 - U_1 - U_2$, W_j that splits 1 into W_j plus $1 - W_j$ for $j \in \mathbb{J}_1$, Z_j that splits 1 into Z_j plus $1 - Z_j$ for $j \in \mathbb{J}_2$. The conditional predictive distributions of these latent variables are given by

$$
\begin{aligned}
\mathbf{u} | (Y_{\text{obs}}, \phi_1, \phi_2, \boldsymbol{\theta}) &\sim \text{Multinomial}\left(m_0; \ \frac{f_0}{f}, \frac{f_1}{f}, \frac{f_2}{f}, \frac{f_3}{f} \right), \\
W_j | (Y_{\text{obs}}, \phi_1, \phi_2, \boldsymbol{\theta}) &\overset{\text{iid}}{\sim} \text{Bernoulli}\left(\frac{\phi_1}{\phi_1 + (1 - \phi_1)e^{-\theta_1}[1 + \lambda(1 - c_1)(e^{-y_{2j}} - c_2)]} \right), \ j \in \mathbb{J}_1, \\
Z_j | (Y_{\text{obs}}, \phi_1, \phi_2, \boldsymbol{\theta}) &\overset{\text{iid}}{\sim} \text{Bernoulli}\left(\frac{\phi_2}{\phi_2 + (1 - \phi_2)e^{-\theta_2}[1 + \lambda(e^{-y_{1j}} - c_1)(1 - c_2)]} \right), \ j \in \mathbb{J}_2,
\end{aligned}
$$

where

$$
\begin{aligned}
f_0 &= \phi_1\phi_2, \quad f_1 = \phi_1(1 - \phi_2)e^{-\theta_2}, \quad f_2 = (1 - \phi_1)\phi_2 e^{-\theta_1}, \\
f_3 &= (1 - \phi_1)(1 - \phi_2)e^{-(\theta_1 + \theta_2)}[1 + \lambda(1 - c_1)(1 - c_2)], \\
f &= f_0 + f_1 + f_2 + f_3.
\end{aligned}
$$

The complete data are denoted by $Y_{\text{com}} = \{Y_{\text{obs}}, u_0, u_1, u_2, u_3, \{w_j\}_{j \in \mathbb{J}_1}, \{z_j\}_{j \in \mathbb{J}_2}\}$. Thus, the complete-data likelihood function is proportional to

$$
\begin{aligned}
L(\phi_1, \phi_2, \boldsymbol{\theta} | Y_{\text{com}}) \quad \propto \quad & (\phi_1 \phi_2)^{u_0} \left[\phi_1 (1 - \phi_2) e^{-\theta_2}\right]^{u_1} \left[(1 - \phi_1) \phi_2 e^{-\theta_1}\right]^{u_2} \\
& \times \left\{ (1 - \phi_1)(1 - \phi_2) e^{-(\theta_1 + \theta_2)} [1 + \lambda(1 - c_1)(1 - c_2)] \right\}^{u_3} \\
& \times \prod_{j \in \mathbb{J}_1} (1 - \phi_2) \text{GP}(y_{2j} | \theta_2, \alpha_2) \phi_1^{w_j} \left\{ (1 - \phi_1) e^{-\theta_1} [1 + \lambda(1 - c_1)(e^{-y_{2j}} - c_2)] \right\}^{1 - w_j} \\
& \times \prod_{j \in \mathbb{J}_2} (1 - \phi_1) \text{GP}(y_{1j} | \theta_1, \alpha_1) \phi_2^{z_j} \left\{ (1 - \phi_2) e^{-\theta_2} [1 + \lambda(e^{-y_{1j}} - c_1)(1 - c_2)] \right\}^{1 - z_j} \\
& \times \prod_{j \in \mathbb{J}_3} (1 - \phi_1)(1 - \phi_2) \text{GP}(y_{1j} | \theta_1, \alpha_1) \text{GP}(y_{2j} | \theta_2, \alpha_2) [1 + \lambda(e^{-y_{1j}} - c_1)(e^{-y_{2j}} - c_2)] \\
= \quad & L_1(\phi_1, \phi_2 | Y_{\text{com}}) \times L_2^*(\boldsymbol{\theta} | Y_{\text{com}}),
\end{aligned}
$$

where

$$
\begin{aligned}
L_1(\phi_1, \phi_2 | Y_{\text{com}}) \quad = \quad & (\phi_1 \phi_2)^{u_0} \left[\phi_1 (1 - \phi_2)\right]^{u_1} \left[(1 - \phi_1) \phi_2\right]^{u_2} \left[(1 - \phi_1)(1 - \phi_2)\right]^{u_3} \\
& \times (1 - \phi_2)^{m_1} \phi_1^{N_w} (1 - \phi_1)^{m_1 - N_w} (1 - \phi_1)^{m_2} \phi_2^{N_z} (1 - \phi_2)^{m_2 - N_z} \\
& \times \left[(1 - \phi_1)(1 - \phi_2)\right]^{n - m_0 - m_1 - m_2}
\end{aligned}
$$

only containing (ϕ_1, ϕ_2), $N_w = \sum_{j \in \mathbb{J}_1} w_j$, $N_z = \sum_{j \in \mathbb{J}_2} z_j$, and

$$
\begin{aligned}
L_2^*(\boldsymbol{\theta} | Y_{\text{com}}) \quad = \quad & e^{-\theta_2 u_1 - \theta_1 u_2 - (\theta_1 + \theta_2) u_3} [1 + \lambda(1 - c_1)(1 - c_2)]^{u_3} \\
& \times \prod_{j \in \mathbb{J}_1} \text{GP}(y_{2j} | \theta_2, \alpha_2) \left\{ e^{-\theta_1} [1 + \lambda(1 - c_1)(e^{-y_{2j}} - c_2)] \right\}^{1 - w_j} \\
& \times \prod_{j \in \mathbb{J}_2} \text{GP}(y_{1j} | \theta_1, \alpha_1) \left\{ e^{-\theta_2} [1 + \lambda(e^{-y_{1j}} - c_1)(1 - c_2)] \right\}^{1 - z_j} \\
& \times \prod_{j \in \mathbb{J}_3} \text{GP}(y_{1j} | \theta_1, \alpha_1) \text{GP}(y_{2j} | \theta_2, \alpha_2) [1 + \lambda(e^{-y_{1j}} - c_1)(e^{-y_{2j}} - c_2)]
\end{aligned}
$$

only involving $\boldsymbol{\theta}$. Since

$$
\begin{aligned}
\ell_1(\phi_1, \phi_2 | Y_{\text{com}}) \quad = \quad & \log L_1(\phi_1, \phi_2 | Y_{\text{com}}) \\
= \quad & (u_0 + u_1 + N_w) \log \phi_1 + (u_2 + u_3 + n - m_0 - N_w) \log(1 - \phi_1) \\
& + (u_0 + u_2 + N_z) \log \phi_2 + (u_1 + u_3 + n - m_0 - N_z) \log(1 - \phi_2),
\end{aligned}
$$

the complete-data MLEs of ϕ_1 and ϕ_2 are given by

$$
\hat{\phi}_1 = \frac{u_0 + u_1 + N_w}{n + u_+ - m_0} \quad \text{and} \quad \hat{\phi}_2 = \frac{u_0 + u_2 + N_z}{n + u_+ - m_0},
$$

where $u_+ = \sum_{i=0}^{3} u_i$. On the other hand,

$$
\begin{aligned}
\ell_2^*(\boldsymbol{\theta}|Y_{\text{com}}) &= \log L_2^*(\boldsymbol{\theta}|Y_{\text{com}}) \\
&= -(n - u_0 - u_1 - N_w)\theta_1 - (n - u_0 - u_2 - N_z)\theta_2 + N_1 \log \theta_1 + N_2 \log \theta_2 \\
&\quad - N_1 \theta_1 \alpha_1 - N_2 \theta_2 \alpha_2 + \sum_{j=1}^{n} (y_{1j} - 1)\log(1 + \alpha_1 y_{1j}) + \sum_{j=1}^{n} (y_{2j} - 1)\log(1 + \alpha_2 y_{2j}) \\
&\quad + \sum_{j=1}^{n} \log\left[1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(\mathrm{e}^{-y_{2j}} - c_2)\right] - (u_0 + u_1 + u_2)\log[1 + \lambda(1 - c_1)(1 - c_2)] \\
&\quad - \sum_{j \in \mathbb{J}_1} w_j \log\left[1 + \lambda(1 - c_1)(\mathrm{e}^{-y_{2j}} - c_2)\right] - \sum_{j \in \mathbb{J}_2} z_j \log\left[1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(1 - c_2)\right], \quad (32)
\end{aligned}
$$

where $N_1 = \sum_{j=1}^{n} y_{1j}$ and $N_2 = \sum_{j=1}^{n} y_{2j}$. Since the complete-data MLEs of $\boldsymbol{\theta}$ are not available in closed form, we adopt the Newton–Raphson algorithm to calculate the complete-data MLEs of $\boldsymbol{\theta}$. The score vector and the Hessian matrix

$$
\nabla \ell_2^*(\boldsymbol{\theta}|Y_{\text{com}}) = \frac{\partial \ell_2^*(\boldsymbol{\theta}|Y_{\text{com}})}{\partial \boldsymbol{\theta}} \quad \text{and} \quad \nabla^2 \ell_2^*(\boldsymbol{\theta}|Y_{\text{com}}) = \frac{\partial^2 \ell_2^*(\boldsymbol{\theta}|Y_{\text{com}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \tag{33}
$$

are derived in Appendix B. The M-step is to update

$$
\begin{aligned}
\phi_1^{(t+1)} &= \frac{u_0^{(t)} + u_1^{(t)} + N_w^{(t)}}{n + u_+^{(t)} - m_0}, \qquad \phi_2^{(t+1)} = \frac{u_0^{(t)} + u_2^{(t)} + N_z^{(t)}}{n + u_+^{(t)} - m_0}, \\
\boldsymbol{\theta}^{(t+1)} &= \boldsymbol{\theta}^{(t)} + [-\nabla^2 \ell_2^*(\boldsymbol{\theta}^{(t)}|Y_{\text{com}})]^{-1} \nabla \ell_2^*(\boldsymbol{\theta}^{(t)}|Y_{\text{com}}),
\end{aligned} \tag{34}
$$

where $N_w^{(t)} = \sum_{j \in \mathbb{J}_1} w_j^{(t)}$, $N_z^{(t)} = \sum_{j \in \mathbb{J}_2} z_j^{(t)}$ and $u_+^{(t)} = \sum_{i=0}^{3} u_i^{(t)}$. The E-step is to replace (u_0, u_1, u_2, u_3), $\{w_j\}_{j \in \mathbb{J}_1}$ and $\{z_j\}_{j \in \mathbb{J}_2}$ in (34) by their conditional expectations

$$
\begin{aligned}
E(U_i|Y_{\text{obs}}, \phi_1, \phi_2, \boldsymbol{\theta}) &= \frac{m_0 f_i}{f}, \quad i = 0, 1, 2, 3, \\
E(W_j|Y_{\text{obs}}, \phi_1, \phi_2, \boldsymbol{\theta}) &= \frac{\phi_1}{\phi_1 + (1 - \phi_1)\mathrm{e}^{-\theta_1}[1 + \lambda(1 - c_1)(\mathrm{e}^{-y_{2j}} - c_2)]}, \\
E(Z_j|Y_{\text{obs}}, \phi_1, \phi_2, \boldsymbol{\theta}) &= \frac{\phi_2}{\phi_2 + (1 - \phi_2)\mathrm{e}^{-\theta_2}[1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(1 - c_2)]}.
\end{aligned} \tag{35}
$$

Note that if the initial value of ϕ_1 is set to be $\phi_1^{(0)} = 0$, the above algorithm yields the estimates of parameters, where that only Y_2 is marginally inflated; if the initial value of ϕ_2 is set to be $\phi_2^{(0)} = 0$, it yields the estimates of parameters, where only Y_1 is marginally inflated.

The estimation of standard errors can be obtained via Louis's method, which is very similar to that stated in Section 2.4. The bootstrap method presented in Section 2.5 can be used to construct bootstrap CIs of parameters for the current situation.

3.3. Testing hypotheses

3.3.1. LRT for zero inflation. First, we want to test whether there exists zero inflation in both marginal distributions. Thus, we consider the following null and alternative hypotheses:

$$
H_0: (\phi_1, \phi_2) = (0, 0) \quad \text{against} \quad H_1: (\phi_1, \phi_2) \neq (0, 0). \tag{36}
$$

Under H_0, the LRT statistic is given by

$$T_4 = -2\{\ell(0, 0, \hat{\boldsymbol{\theta}}_{H_0}|Y_{\text{obs}}) - \ell(\hat{\phi}_1, \hat{\phi}_2, \hat{\boldsymbol{\theta}}|Y_{\text{obs}})\}. \tag{37}$$

The test statistic T_4 does not follow the chi-squared distribution with two degrees of freedom because when H_0 is true, the parameter values are located at the vertex boundary of the bounded parameter space. The reference distribution for T_4 is a mixture of $(1/4)\chi^2(0)$ (i.e., a constant at zero), $(1/2)\chi^2(1)$ and $(1/4)\chi^2(2)$, see Chernoff [6] and Self & Liang [25].

Subsequently, if the null hypothesis in (36) is rejected, we consider the marginal zero inflation. We test $\phi_i = 0$ for $i = 1, 2$, respectively. The hypotheses are

$$H_0: \phi_1 = 0 \qquad \text{against} \qquad H_1: \phi_1 > 0; \quad \text{and} \tag{38}$$

$$H_0: \phi_2 = 0 \qquad \text{against} \qquad H_1: \phi_2 > 0, \tag{39}$$

respectively. Under H_0, the LRT statistics are given by

$$T_5 = -2\{\ell(0, \hat{\phi}_{2,H_0}, \hat{\boldsymbol{\theta}}_{H_0}|Y_{\text{obs}}) - \ell(\hat{\phi}_1, \hat{\phi}_2, \hat{\boldsymbol{\theta}}|Y_{\text{obs}})\}; \quad \text{and} \tag{40}$$

$$T_6 = -2\{\ell(\hat{\phi}_{1,H_0}, 0, \hat{\boldsymbol{\theta}}_{H_0}|Y_{\text{obs}}) - \ell(\hat{\phi}_1, \hat{\phi}_2, \hat{\boldsymbol{\theta}}|Y_{\text{obs}})\}, \tag{41}$$

respectively. The null distributions of test statistics T_5 and T_6 are both the equal mixture of $\chi^2(0)$ and $\chi^2(1)$.

3.3.2. LRT for independency parameter. From (30), we know that if $\lambda = 0$, then $\text{Corr}(Y_1, Y_2) = 0$. Thus, we consider to test

$$H_0: \lambda = 0 \quad \text{against} \quad H_1: \lambda \neq 0. \tag{42}$$

Under H_0, the LRT statistic

$$T_7 = -2\{\ell(\hat{\phi}_{1,H_0}, \hat{\phi}_{2,H_0}, \hat{\theta}_{1,H_0}, \hat{\theta}_{2,H_0}, \hat{\alpha}_{1,H_0}, \hat{\alpha}_{2,H_0}, 0|Y_{\text{obs}}) - \ell(\hat{\phi}_1, \hat{\phi}_2, \hat{\boldsymbol{\theta}}|Y_{\text{obs}})\} \tag{43}$$

approximately follows $\chi^2(1)$ distribution. The corresponding p-value is

$$p_{\text{v7}} = \Pr(T_7 > t_7|H_0) = \Pr(\chi^2(1) > t_7). \tag{44}$$

4. Applications in Australian health care utilization data

Cameron & Trivedi [5] reported data concerning the demand for health care in Australia which refers to the Australian Health Survey for 1977–1978. Let Y_1 denote the number of consultations with a doctor or a specialist and Y_2 denote the total number of prescribed medications used in past two days. The data are given in Table 2.

Before the data analysis, we first examine the descriptive statistics for the data: $\bar{y}_1 = 0.3017341$, $\bar{y}_2 = 0.8626204$; $s_1^2 = 0.6370176$, $s_2^2 = 2.0032855$; the sample correlation coefficient $r = 0.3077787$. We have observed that both marginal count data exhibit over-dispersion and very high frequencies of zero observations since most of the observed frequencies fall in the $(0,0)$ category. The positive sample correlation coefficient r indicates that there is a positive correlation between Y_1 and Y_2. Based on these observations, we first try to fit the bivariate count data by the Type I bivariate ZIGP model with a multiplicative factor in Section 2.

4.1. Statistical inferences based on Type I bivariate ZIGP$_\lambda$

Suppose $(Y_{1j}, Y_{2j})^\top \overset{\text{iid}}{\sim} \text{ZIGP}_\lambda^{(\text{I})}(\phi; \theta_1, \theta_2, \alpha_1, \alpha_2)$ for $j = 1, \ldots, n$ $(n = 5190)$. To find the MLEs of $(\phi, \theta_1, \theta_2, \alpha_1, \alpha_2, \lambda)$, we choose $(\phi^{(0)}, \theta_1^{(0)}, \theta_2^{(0)}, \alpha_1^{(0)}, \alpha_2^{(0)}, \lambda^{(0)}) = (0.2, 0.3, 0.8, 1, 0.5, 2)$ as their initial values. The MLEs of $(\phi, \theta_1, \theta_2, \alpha_1, \alpha_2, \lambda)$ converged to $(\hat{\phi}, \hat{\theta}_1, \hat{\theta}_2, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\lambda})$ as shown in the second column of Table 3 in

750 iterations for the algorithm specified in (11)–(12). From (4), the estimated correlation coefficient is

$$\widehat{\mathrm{Corr}}(Y_1, Y_2) = 0.2370878,$$

which is slightly different from the sample correlation coefficient $r = 0.3077787$. The estimated standard errors of the MLEs $(\hat\phi, \hat\theta_1, \hat\theta_2, \hat\alpha_1, \hat\alpha_2, \hat\lambda)$ based on the observed information matrix [23] are given in the third column and 95% asymptotic Wald CIs are listed in the last column of Table 3.

    Besides, we adopt the bootstrap method to compute the bootstrap CIs for these parameters. Using the obtained MLEs, we generate $G = 6,000$ bootstrap samples and the three types of bootstrap CIs (16)–(18) are summarized in Table 4.

Table 2

Cross tabulation of the health care utilization data in the Australian Health Survey for 1977–1978 (Cameron and Trivedi, [5])

$Y_1 \backslash Y_2$	0	1	2	3	4	5	6	7	8	Total
0	2789	726	307	171	76	32	16	15	9	4141
1	224	212	149	85	50	35	13	5	9	782
2	49	34	38	11	23	7	5	3	4	174
3	8	10	6	2	1	1	2	0	0	30
4	8	8	2	2	3	1	0	0	0	24
5	3	3	2	0	1	0	0	0	0	9
6	2	0	1	3	1	2	2	0	1	12
7	1	0	3	2	1	2	1	0	2	12
8	1	1	1	0	1	0	1	0	0	5
9	0	0	0	0	0	0	0	0	1	1
Total	3085	994	509	276	157	80	40	23	26	5190

Table 3

MLEs and CIs of parameters for the Australian Health Survey data (Type I)

Parameter	MLE	std	95% Wald CI
ϕ	0.229847	0.036141	[0.159013, 0.300682]
θ_1	0.297364	0.017737	[0.262599, 0.332128]
θ_2	0.756942	0.051924	[0.655172, 0.858712]
α_1	0.826190	0.081361	[0.666725, 0.985654]
α_2	0.427029	0.048782	[0.331419, 0.522640]
λ	1.900207	0.200254	[1.507718, 2.292697]

std: Calculated as the square roots of the diagonal elements of $\mathbf{I}^{-1}(\hat\phi, \hat{\boldsymbol\theta}|Y_{\mathrm{obs}})$, cf. (13).

    Suppose that we want to test the null hypothesis $H_0\colon \phi = 0$ against the alternative hypothesis $H_1\colon \phi > 0$. According to (20), we calculate the value of the LRT statistic which is given by $t_1 = 20.2776$, then from (21), we have the corresponding $p_{\mathrm{v1}} = 3.3491 \times 10^{-6} < \alpha = 0.05$. Thus, the null hypothesis should be rejected at 5% significance level.

    If we want to test the null hypothesis $H_0\colon \lambda = 0$ against the alternative hypothesis $H_1\colon \lambda \neq 0$. According to (23), the value of the LRT statistic is $t_2 = 88.9215$ and from (24) the p-value is $p_{\mathrm{v2}} < 0.0001$. Thus, we should reject the null hypothesis at the significance level of 5%.

    If we want to test the null hypothesis $H_0\colon \alpha_1 = \alpha_2 = 0$. According to (26), the value of the LRT statistic is $t_3 = 1615.4960$ and from (27), we have $p_{\mathrm{v3}} \ll 0.0001$. Thus, the null hypothesis is rejected at the significance level of 5%.

Table 4
Three 95% boostrap CIs for the Australian Health Survey data (Type I)

Parameter	Mean	std$^{\mathrm{B}}$	Simple bootstrap CI	Bootstrap percentile CI	Bootstrap percentile-t CI
ϕ	0.22852	0.03689	[0.15621, 0.30082]	[0.15040, 0.29502]	[0.14502, 0.29188]
θ_1	0.29760	0.01837	[0.26159, 0.33361]	[0.26185, 0.33388]	[0.26117, 0.33428]
θ_2	0.75837	0.05130	[0.65783, 0.85891]	[0.65576, 0.85822]	[0.65773, 0.86019]
α_1	0.82894	0.08753	[0.65738, 1.00050]	[0.66688, 1.01528]	[0.67324, 1.02453]
α_2	0.42857	0.04831	[0.33388, 0.52325]	[0.34173, 0.53167]	[0.34626, 0.53721]
λ	1.90180	0.20800	[1.49412, 2.30948]	1.49575, 2.31837]	[1.50518, 2.31860]

std$^{\mathrm{B}}$: The sample standard deviation based on the bootstrap samples, cf. (15).

4.2. *Marginal analysis*

The sample correlation coefficient $r = 0.3077787$, indicating that there is a positive correlation between Y_1 and Y_2. Thus, any two independent univariate distributions cannot be used to model the health care utilization data. Besides, since both marginal count data are over-dispersed according to their marginal sample means and variances, the bivariate Poisson distribution constructed by the trivariate reduction method is also not appropriate due to the property of equi-dispersion.

Note that the over-dispersion may result from the excess (0,0) points in the observations, the Type I multivariate ZIP distribution proposed by Liu & Tian [22] could be considered. If we fit the data by the Type I bivariate ZIP distribution, denoted by $(Y_1, Y_2)^\top \sim \mathrm{ZIP}^{(\mathrm{I})}(\phi; \theta_1, \theta_2)$, the MLEs of three parameters are $\hat{\phi} = 0.483007$, $\hat{\theta}_1 = 0.583633$, $\hat{\theta}_2 = 1.668533$. The estimated correlation coefficient is

$$\widehat{\mathrm{Corr}}(Y_1, Y_2) = \sqrt{\frac{\hat{\theta}_1 \hat{\theta}_2}{(\hat{\theta}_1 + 1/\hat{\phi})(\hat{\theta}_2 + 1/\hat{\phi})}} = 0.3132676,$$

which is very close to the sample correlation coefficient $r = 0.3077787$. From the definition of $(Y_1, Y_2)^\top \sim \mathrm{ZIP}^{(\mathrm{I})}(\phi; \theta_1, \theta_2)$, we have $Y_i \sim \mathrm{ZIP}(\phi; \theta_i)$ for $i = 1, 2$, indicating that the marginal distributions must share a common zero inflation parameter ϕ. On the other hand, based on the marginal counts, the univariate ZIP distribution of Y_1 and Y_2 are estimated to be $\mathrm{ZIP}(\hat{\phi}_1^{\mathrm{M}} = 0.650393; \hat{\theta}_1^{\mathrm{M}} = 0.863068)$ and $\mathrm{ZIP}(\hat{\phi}_2^{\mathrm{M}} = 0.510188; \hat{\theta}_2^{\mathrm{M}} = 1.761464)$, respectively. It is clear that the difference between $\hat{\phi}_1^{\mathrm{M}}$ and $\hat{\phi}_2^{\mathrm{M}}$ is slight.

An alternative to the above $\mathrm{ZIP}^{(\mathrm{I})}(\phi; \theta_1, \theta_2)$ is the Type I bivariate ZIP distribution indexed by the multiplicative factor λ, i.e., $\mathrm{ZIP}_\lambda^{(\mathrm{I})}(\phi; \theta_1, \theta_2)$, which is reduced from the proposed Type I bivariate ZIGP distribution with the same multiplicative factor λ by setting $\alpha_1 = \alpha_2 = 0$. The corresponding MLEs of the four parameters are given by $\hat{\phi} = 0.489964, \hat{\theta}_1 = 0.588041, \hat{\theta}_2 = 1.685382, \hat{\lambda} = -0.477044$. From (5), the estimated correlation coefficient is $\widehat{\mathrm{Corr}}(Y_1, Y_2) = 0.2886614$, which is close to the sample correlation coefficient $r = 0.3077787$. Note that the distribution $\mathrm{ZIP}_\lambda^{(\mathrm{I})}(\phi; \theta_1, \theta_2)$ has the same marginal distributions $\mathrm{ZIP}(\phi; \theta_i)$ for $i = 1, 2$ as $\mathrm{ZIP}^{(\mathrm{I})}(\phi; \theta_1, \theta_2)$.

For the proposed Type I bivariate ZIGP distribution with the multiplicative factor λ, according to (3), we have $Y_i \sim \mathrm{ZIGP}(\phi; \theta_i, \alpha_i)$ for $i = 1, 2$. That is, the two marginal distributions should share a common zero inflation parameter ϕ. However, based on the marginal counts, the univariate ZIGP distribution of Y_1 and Y_2 are estimated to be

$$Y_1 \sim \mathrm{ZIGP}(\phi_1^{\mathrm{M}} = 0; \theta_1^{\mathrm{M}} = 0.221530, \alpha_1^{\mathrm{M}} = 1.199886) = \mathrm{GP}(\theta_1^{\mathrm{M}} = 0.221530, \alpha_1^{\mathrm{M}} = 1.199886), \qquad (45)$$

and

$$Y_2 \sim \mathrm{ZIGP}(\phi_2^{\mathrm{M}} = 0.311382; \theta_2^{\mathrm{M}} = 0.889408, \alpha_2^{\mathrm{M}} = 0.326211). \qquad (46)$$

Since there is a larger difference between ϕ_1^{M} and ϕ_2^{M}, from the viewpoint of marginal analysis, Type I bivariate ZIGP distribution with the multiplicative factor λ is not appropriate to fit the data.

4.3. Statistical inferences based on Type II bivariate ZIGP$_\lambda$

Since the marginal analysis indicates that only Y_2 follows a ZIGP distribution, we suppose $(Y_{1j}, Y_{2j})^\top \overset{\text{iid}}{\sim}$ ZIGP$_\lambda^{(\text{II})}(0, \phi_2; \theta_1, \theta_2, \alpha_1, \alpha_2)$ for $j = 1, \ldots, n$ $(n = 5190)$. To find the MLEs of $(\phi_1, \phi_2, \theta_1, \theta_2, \alpha_1, \alpha_2, \lambda)$, we choose $(\phi_1^{(0)}, \phi_2^{(0)}, \theta_1^{(0)}, \theta_2^{(0)}, \alpha_1^{(0)}, \alpha_2^{(0)}, \lambda^{(0)}) = (0, 0.3, 0.2, 0.9, 1.2, 0.3, 4)$ as their initial values. The MLEs of $(\phi_2, \theta_1, \theta_2, \alpha_1, \alpha_2, \lambda)$ converged to $(\hat{\phi}_2, \hat{\theta}_1, \hat{\theta}_2, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\lambda})$ as shown in the second column of Table 5 in 8317 iterations for the algorithm specified in (34)–(35). From the first formula of (31), the estimated correlation coefficient is

$$\widehat{\text{Corr}}(Y_1, Y_2) = 0.2678956,$$

which has an improvement compared with that in Type I bivariate ZIGP$_\lambda$ distribution. The estimated standard errors of the MLEs $(\hat{\phi}_2, \hat{\theta}_1, \hat{\theta}_2, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\lambda})$ based on the complete information matrix are given in the third column and 95% Wald CIs are listed in the last column of Table 5. The two 95% bootstrap CIs of $(\phi_2, \theta_1, \theta_2, \alpha_1, \alpha_2, \lambda)$ produced by $G = 1,000$ bootstrap replications are presented in the last two columns of Table 6.

Table 5
MLEs and CIs of parameters for the Australian Health Survey data (Type II)

Parameter	MLE	std	95% Wald CI
ϕ_2	0.335757	0.006555	[0.322909, 0.348605]
θ_1	0.222826	0.006838	[0.209424, 0.236228]
θ_2	0.933804	0.019452	[0.895678, 0.971930]
α_1	1.193268	0.079058	[1.038318, 1.348218]
α_2	0.298603	0.016168	[0.266915, 0.330291]
λ	4.266230	0.056552	[4.155389, 4.377070]

std: Calculated as the square roots of the diagonal elements of $\mathbf{I}^{-1}(\hat{\phi}_2, \hat{\boldsymbol{\theta}}|Y_{\text{com}})$, cf. (14).

Table 6
Two 95% bootstrap CIs for the Australian Health Survey data (Type II)

Parameter	Mean	std$^{\text{B}}$	Simple bootstrap CI	Bootstrap percentile CI
ϕ_2	0.360594	0.028119	[0.305381, 0.415606]	[0.281119, 0.399288]
θ_1	0.238961	0.007039	[0.225165, 0.252757]	[0.226017, 0.252277]
θ_2	0.966234	0.061423	[0.845844, 1.086623]	[0.812727, 1.062516]
α_1	1.233189	0.071703	[1.092652, 1.373726]	[1.114951, 1.389566]
α_2	0.282438	0.038001	[0.207957, 0.356920]	[0.222765, 0.376424]
λ	3.684713	0.605455	[2.498021, 4.871404]	[3.045801, 4.095513]

std$^{\text{B}}$: The sample standard deviation based on the bootstrap samples, cf. (15).

Suppose that we want to test the null hypothesis $H_0\colon \phi_2 = 0$ against the alternative hypothesis $H_1\colon \phi_2 > 0$. According to (41), we calculate the value of the LRT statistic which is given by $t_6 = 50.9026$, the corresponding p-value is $4.8533 \times 10^{-13} \ll \alpha = 0.05$. Thus, the null hypothesis should be rejected at 5% level of significance.

If we want to test the null hypothesis $H_0\colon \lambda = 0$ against the alternative hypothesis $H_1\colon \lambda \neq 0$. According to (43), the value of the LRT statistic is $t_7 = 585.2057$ and from (44) we have the corresponding p-value is less than 0.0001. Thus, we should reject the null hypothesis.

The marginal analysis based on the Type II bivariate ZIGP$_\lambda$ is the same as that based on the Type I bivariate ZIGP$_\lambda$, which is still given by (45) and (46).

4.4. Model comparison

We adopt the *Akaike information criterion* (AIC; [2]) and *Bayesian information criterion* (BIC; [24]) to compare six different models:

(a)   Karlis & Ntzoufras [18] considered two regression models to fit the data; i.e., *bivariate Poisson* (BP) constructed by the method of trivariate reduction based on three independent Poisson(θ_i) variables for $i = 0, 1, 2$, and

(b)   *diagonal inflated bivariate Poisson* (DIBP), which led to a model with only $(0, 0)$ inflation but without $(1,1)$ inflation in the current case. For the purpose of comparison, we concentrate our attention upon the situation of without covariates.

(c)   Liu & Tian [22] recently proposed the Type I multivariate ZIP distribution to fit the data in which Y_1 and Y_2 are correlated just through a common zero inflation factor.

(d)   Type I bivariate ZIP$_\lambda$; i.e., ZIP$_\lambda^{(\mathrm{I})}(\phi; \theta_1, \theta_2)$, see the third paragraph of Section 4.2.

(e)   Type I bivariate ZIGP$_\lambda$, see Section 4.1.

(f)   Type II bivariate ZIGP$_\lambda$, see Section 4.3.

The estimates of parameters for first four models are reported in the second column of Table 7. The values of AIC and BIC based on the Total Likelihood Function are summarized in the last two columns of Table 7. From Table 7, we can see that the Type II bivariate ZIGP$_\lambda$ is the best model in terms of the log-likelihood, AIC and BIC. It also coincides with the result from the marginal analysis.

Table 7
Comparison of six different models via AIC and BIC

Model	Parameter estimation	Log-likelihood	Criterion AIC	Criterion BIC
(a) BP	$\hat{\theta}_0 = 0.125601, \hat{\theta}_1 = 0.176134,$ $\hat{\theta}_2 = 0.737020$	-11268.36	22542.71	22562.38
(b) DIBP	$\hat{\phi} = 0.4763110, \hat{\theta}_0 = 0.074470,$ $\hat{\theta}_1 = 0.501701, \hat{\theta}_2 = 1.572730$	-10260.96	20529.92	20556.14
(c) Type I bivaraite ZIP	$\hat{\phi} = 0.483007, \hat{\theta}_1 = 0.583633,$ $\hat{\theta}_2 = 1.668533$	-10279.91	20565.82	20585.48
(d) Type I bivaraite ZIP$_\lambda$	$\hat{\phi} = 0.489964, \hat{\theta}_1 = 0.588041,$ $\hat{\theta}_2 = 1.685382, \hat{\lambda} = -0.477044$	-10707.34	21422.69	21448.90
(e) Type I bivariate ZIGP$_\lambda$		-9899.595	19811.19	19850.52
(f) Type II bivariate ZIGP$_\lambda$		-9884.282	**19780.56**	**19819.89**

BP and DIBP: See Karlis & Ntzoufras [18].
Type I bivariate ZIP: See Liu & Tian [22].
Type I bivariate ZIP$_\lambda$: $=$ ZIP$_\lambda^{(\mathrm{I})}(\phi; \theta_1, \theta_2)$, see the third paragraph of Section 4.2.

On the other hand, we can compare the performances of these models via the estimated mean, variance and correlation coefficient which are reported in Table 8.

From Table 8, we can see that both the Type I and Type II bivariate ZIGP$_\lambda$ distributions perform better at adjusting the over-dispersion of the data, while the Type I bivariate ZIP and Type I bivariate ZIP$_\lambda$ have better estimates on

the correlation coefficient. The key point lies on that the GP distribution in the $ZIGP_\lambda$ has one more parameter than the Poisson distribution, thus it can better accommodate the degree of over-dispersion and zero-inflation in the data. Moreover, the GP distribution can also model under-dispersed data depending on the values of the two parameters. Besides, the addition of the multiplicative factor λ in the pmf (1) can further improve the model fitting so that it can fit either positively or negatively correlated count data.

Table 8
Comparison of estimated mean, variance and correlation coefficient for four different models

	Y_1		Y_2		Correlation
Model	Mean	Variance	Mean	Variance	coefficient
Sample	0.3017341	0.6370176	0.8626204	2.0032855	0.3077787
Type I bivariate ZIP	0.3017342	0.3867927	0.8626199	1.5578166	0.3132676
Type I bivaraite ZIP_λ	0.2999224	0.3863357	0.8596060	1.5694480	0.2886614
Type I bivariate $ZIGP_\lambda$	0.3036044	0.5610839	0.8613954	2.1021857	0.2370878
Type II bivariate $ZIGP_\lambda$	0.3035323	0.5632274	0.8601007	2.0277335	0.2678956

5. Simulation studies

To evaluate the performance of the proposed methods for the two types of bivariate $ZIGP_\lambda$ distributions, we first investigate the accuracy of point estimators and interval estimators. Next, we focus on the type I error rates and powers of the LRT in the Type I and Type II bivariate $ZIGP_\lambda$ distributions, respectively.

5.1. Accuracy of point estimators and interval estimators

In this subsection, we assess the accuracy of the point estimators and interval estimators in the two types of bivariate $ZIGP_\lambda$ distributions. We consider one case for each type. For the Type I bivariate $ZIGP_\lambda$ distribution, the true values of the parameters $(\phi; \theta_1, \theta_2, \alpha_1, \alpha_2, \lambda)$ are set to be $(0.4; 0.5, 0.3, 0.6, 0.2, 2)$. Based on (4), the estimated correlation coefficient between the two random variables Y_1 and Y_2 is 0.276215; i.e., they are positively correlated. For the Type II bivariate $ZIGP_\lambda$ distribution, we consider the situation with zero inflation only in one margin, and the true values of the parameters $(\phi_1, \phi_2; \theta_1, \theta_2, \alpha_1, \alpha_2, \lambda)$ are set to be $(0, 0.4; 1, 0.6, 0.3, 0.4, -2)$. Based on the first formula of (31), the estimated correlation coefficient between Y_1 and Y_2 is -0.1513673, indicating negative correlation.

Sample size is set to be $n = 1,000$. In the first case, we generate $(Y_{1j}, Y_{2j})^\top \overset{\text{iid}}{\sim} ZIGP_\lambda^{(I)}(\phi; \theta_1, \theta_2, \alpha_1, \alpha_2)$ for $j = 1, \ldots, n$, then calculate the MLEs via the EM algorithm embedded with the Newton–Raphson algorithm specified by (11)–(12) and 95% bootstrap CIs with $G = 5,000$ bootstrap replications. In the second case, we generate $(Y_{1j}, Y_{2j})^\top \overset{\text{iid}}{\sim} ZIGP_\lambda^{(II)}(\phi_1, \phi_2; \theta_1, \theta_2, \alpha_1, \alpha_2)$ for $j = 1, \ldots, n$ based on the given parameter configuration and calculate the MLEs via the EM algorithm embedded with the Newton–Raphson algorithm specified by (34)–(35) and 95% bootstrap CIs with $G = 5,000$ bootstrap replications. Next, we independently repeat each process for 1000 times, the resultant mean of the MLEs (denoted by MLE), the mean squared errors of the estimates (denoted by MSE), the average width of confidence intervals for parameters (denoted by Width) and the average coverage proportion of bootstrap CIs (denoted by CP) are reported in Table 9, respectively.

5.2. Performance of the LRT in Type I bivariate $ZIGP_\lambda$

In this subsection, we investigate the performance of the LRT for testing hypotheses (19) and (22) for the Type I bivariate $ZIGP_\lambda$ distribution. We conduct extensive simulations to observe the changes of type I error rates and powers against the sample sizes which are set to be $n = 500(50)950$.

5.2.1. LRT for testing H_0: $\phi = 0$.   To estimate the type I error rates (with H_0: $\phi = 0$) and powers (with H_1: $\phi > 0$) for various sample sizes, where the values of ϕ in H_1 are chosen to be 0.01 and 0.05. Given a combination of parameters $(n, \phi, \theta_1 = 1, \theta_2 = 0.6, \alpha_1 = 0.3, \alpha_2 = 0.4, \lambda = 2)$, we generate

$$(Y_{11}^{(l)}, Y_{21}^{(l)})^\top, \ldots, (Y_{1n}^{(l)}, Y_{2n}^{(l)})^\top \overset{\text{iid}}{\sim} \text{ZIGP}_\lambda^{(\text{I})}(\phi; \theta_1, \theta_2, \alpha_1, \alpha_2)$$

for $l = 1, \ldots, L$ ($L = 5,000$). For each group of samples $\{Y_{1j}^{(l)}, Y_{2j}^{(l)})^\top\}_{j=1}^n$, we conduct the testing hypothesis. Let r_1 denote the number of rejecting the null hypothesis H_0: $\phi = 0$ by the test statistic T_1 given by (20). Then the actual significance level can be estimated by r_1/L with $\phi = 0$ and the power of the test statistic T_1 can be estimated by r_1/L with $\phi > 0$.

The empirical levels/powers of the LRT statistic T_1 are summarized in Table 10. Figure 3 plots the type I error rates and powers of the LRT in testing H_0: $\phi = 0$ against H_1: $\phi > 0$ with two different values of $\phi > 0$ for various sample sizes.

Table 9
MLEs and bootstrap confidence intervals of parameters

	Parameter	ϕ	θ_1	θ_2	α_1	α_2	λ
	True value	0.4	0.5	0.3	0.6	0.2	2.0
Type I	MLE	0.382	0.517	0.308	0.630	0.217	2.028
bivariate	MSE	−0.018210	0.016535	0.007816	0.030112	0.017301	0.027906
ZIGP$_\lambda$	Width	0.538	0.519	0.296	1.049	0.795	3.854
	CP	0.937	0.963	0.962	0.943	0.947	0.943
	Parameter	ϕ_2	θ_1	θ_2	α_1	α_2	λ
	True value	0.4	1.0	0.6	0.3	0.4	−2.0
Type II	MLE	0.381	1.001	0.615	0.299	0.425	−2.046
bivariate	MSE	−0.019835	0.001094	0.014582	−0.000573	0.025374	−0.046284
ZIGP$_\lambda$	Width	0.501	0.152	0.588	0.111	0.794	1.600
	CP	0.941	0.949	0.959	0.950	0.945	0.952

NOTE: MLE is the mean of the 1000 point estimates via the EM algorithm embedded with the Newton–Raphson algorithm (11)–(12) and (34)–(35), respectively; MSE is equal to the sum of the variance and the squared bias of the estimator; Width and CP are the average width and coverage proportion of 1000 bootstrap confidence intervals.

Table 10
Empirical levels/powers of the LRT statistic T_1 based on $L = 5,000$ replications

Sample size (n)	Empirical level	Empirical power ϕ		Sample size (n)	Empirical level	Empirical power ϕ	
		0.01	0.05			0.01	0.05
500	0.048	0.064	0.194	750	0.046	0.078	0.236
550	0.058	0.068	0.186	800	0.046	0.076	0.232
600	0.046	0.068	0.196	850	0.062	0.084	0.246
650	0.060	0.066	0.210	900	0.042	0.084	0.238
700	0.056	0.076	0.212	950	0.056	0.088	0.250

5.2.2. LRT for testing H_0: $\lambda = 0$.   To estimate the type I error rates (with H_0: $\lambda = 0$) and powers (with H_1: $\lambda \neq 0$) for various sample sizes, where the values of λ in H_1 are chosen to be -1, 1, 2. Given a combination

of parameters $(n, \phi = 0.2, \theta_1 = 1, \theta_2 = 0.6, \alpha_1 = 0.3, \alpha_2 = 0.4, \lambda)$, we generate

$$(Y_{11}^{(l)}, Y_{21}^{(l)})^\top, \ldots, (Y_{1n}^{(l)}, Y_{2n}^{(l)})^\top \overset{\text{iid}}{\sim} \text{ZIGP}_\lambda^{(\text{I})}(\phi; \theta_1, \theta_2, \alpha_1, \alpha_2)$$

for $l = 1, \ldots, L$ $(L = 5{,}000)$. For each group of samples $\{Y_{1j}^{(l)}, Y_{2j}^{(l)})^\top\}_{j=1}^n$, we conduct the testing hypothesis. Let r_2 denote the number of rejecting the null hypothesis $H_0\colon \lambda = 0$ by the test statistic T_2 given by (23). Then the actual significance level can be estimated by r_2/L with $\lambda = 0$ and the power of the test statistic T_2 can be estimated by r_2/L with $\lambda \neq 0$.

The empirical levels/powers of the LRT statistic T_2 are summarized in Table 11 and Figure 4 shows the changes of the type I error rates and powers of the LRT in testing $H_0\colon \lambda = 0$ against $H_1 : \lambda \neq 0$ with three different values of $\lambda \neq 0$ for various sample sizes.
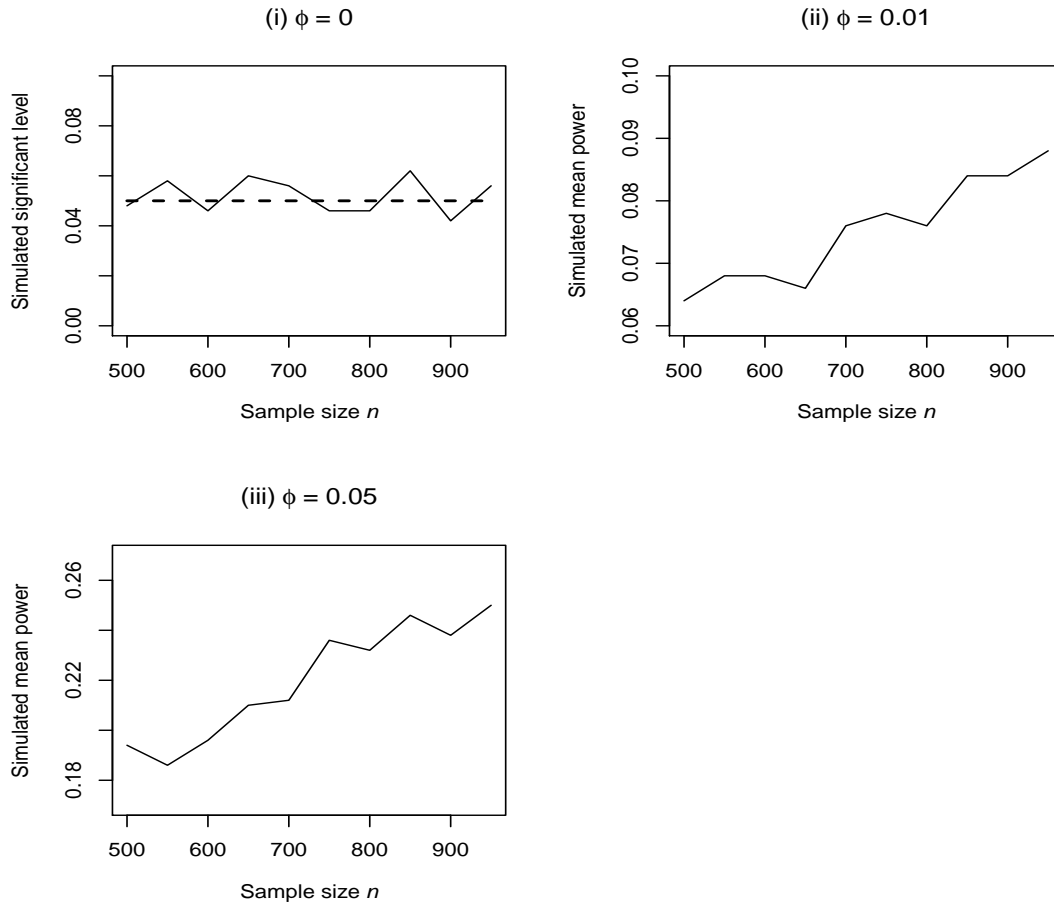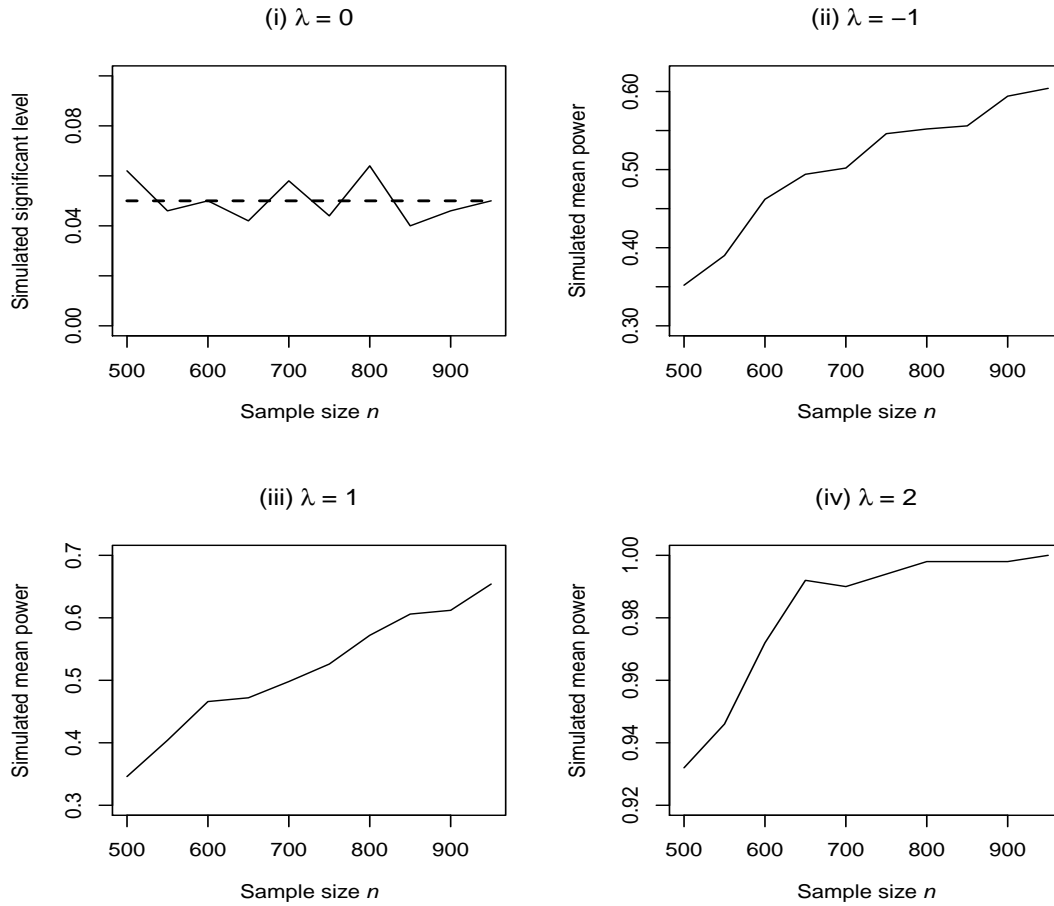

Figure 3. (i) The type I error rates for testing $H_0\colon \phi = 0$ against $H_1\colon \phi > 0$ in the Type I bivariate ZIGP$_\lambda$ distribution and the dashed line is set as the predetermined significance level of $\alpha = 0.05$; (ii) the powers when $\phi = 0.01$ in H_1; (iii) the powers when $\phi = 0.05$ in H_1.

5.3. Performance of the LRT in Type II bivariate ZIGP$_\lambda$

In this subsection, we investigate the performance of the LRT for the testing hypotheses (36) and (39) for the Type II bivariate ZIGP$_\lambda$ distribution. We conduct extensive simulations to observe the changes of type I error rates and powers of the LRT. The sample sizes are set to be $n = 500(50)950$.

Table 11
Empirical levels/powers of the LRT statistic T_2 based on $L = 5,000$ replications

Sample size (n)	Empirical level	Empirical power λ			Sample size (n)	Empirical level	Empirical power λ		
		-1	1	2			-1	1	2
500	0.062	0.352	0.346	0.932	750	0.044	0.546	0.526	0.994
550	0.046	0.390	0.404	0.946	800	0.064	0.552	0.572	0.998
600	0.050	0.462	0.466	0.972	850	0.040	0.556	0.606	0.998
650	0.042	0.494	0.472	0.992	900	0.046	0.594	0.612	0.998
700	0.058	0.502	0.498	0.990	950	0.050	0.604	0.654	1

Figure 4. (i) The type I error rates for testing H_0: $\lambda = 0$ against H_1: $\lambda \neq 0$ in the Type I bivariate ZIGP$_\lambda$ distribution and the dashed line is set as the predetermined significance level of $\alpha = 0.05$; (ii) the powers when $\lambda = -1$ in H_1; (iii) the powers when $\lambda = 1$ in H_1; (iv) the powers when $\lambda = 2$ in H_1.

5.3.1. LRT for testing H_0: $\phi_1 = \phi_2 = 0$. For simplicity, we only estimate the type I error rates (with H_0: $\phi_1 = \phi_2 = 0$) of the LRT for various sample sizes. Given a combination of parameters ($n, \phi_1 = 0, \phi_2 = 0, \theta_1 = 1, \theta_2 = 0.6, \alpha_1 = 0.3, \alpha_2 = 0.4, \lambda = -2$), we generate

$$(Y_{11}^{(l)}, Y_{21}^{(l)})^\top, \ldots, (Y_{1n}^{(l)}, Y_{2n}^{(l)})^\top \overset{\text{iid}}{\sim} \text{ZIGP}_\lambda^{(\text{II})}(0, 0; \theta_1, \theta_2, \alpha_1, \alpha_2)$$

for $l = 1, \ldots, L$ ($L = 1,000$). For each group of samples $\{Y_{1j}^{(l)}, Y_{2j}^{(l)})^\top\}_{j=1}^n$, we conduct the testing hypothesis. Let r_4 denote the number of rejecting the null hypothesis H_0: $(\phi_1, \phi_2) = (0, 0)$ by the test statistic T_4 given by (37). Then the actual significance level can be estimated by r_4/L with $(\phi_1, \phi_2) = (0, 0)$.

The empirical levels of the LRT statistic T_4 are summarized in Table 12 and Figure 5 plots the type I error rates of the LRT in testing H_0: $(\phi_1, \phi_2) = (0, 0)$ against H_1: $(\phi_1, \phi_2) \neq (0, 0)$ for different sample sizes.

Table 12
Empirical levels of the LRT statistic T_4 based on $L = 1,000$ replications

Sample size (n)	500	500	600	650	700	750	800	850	900	950
Empirical level	0.050	0.048	0.056	0.046	0.050	0.054	0.052	0.044	0.048	0.062
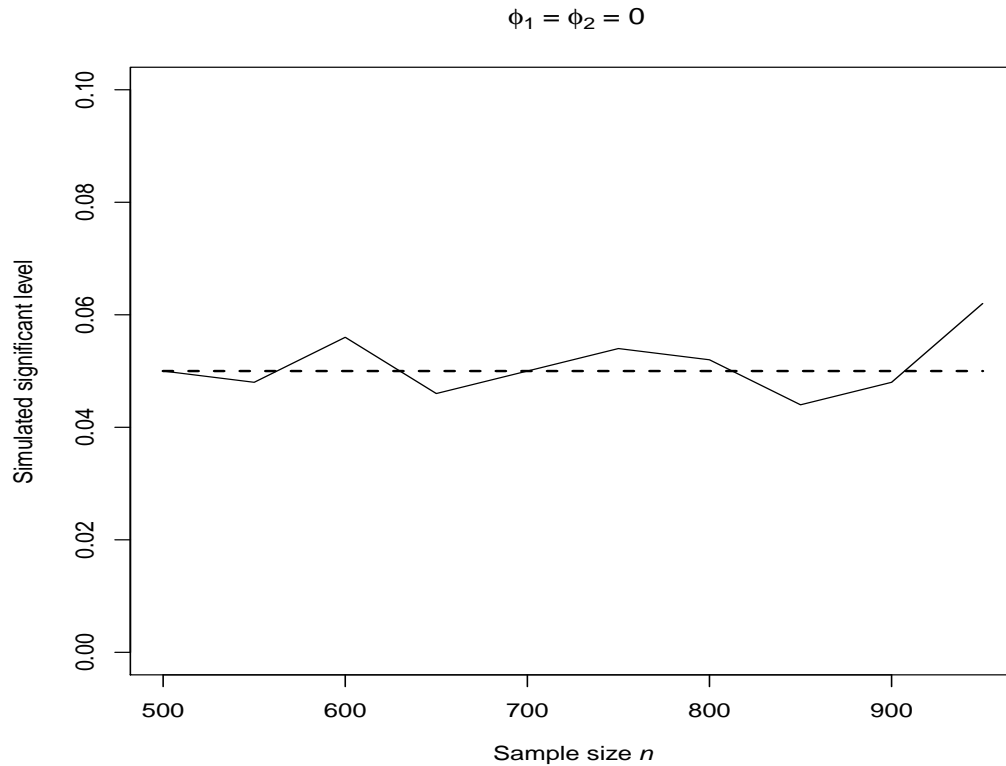
Figure 5. The type I error rates for testing H_0: $(\phi_1, \phi_2) = (0, 0)$ against H_1: $(\phi_1, \phi_2) \neq (0, 0)$ in the Type II bivariate ZIGP$_\lambda$ distribution and the dashed line is set as the predetermined significance level of $\alpha = 0.05$.

5.3.2. LRT for testing H_0: $\phi_2 = 0$.   We consider the case with zero inflation only in Y_2 margin and the true values of the parameters $(\phi_1, \phi_2; \theta_1, \theta_2, \alpha_1, \alpha_2, \lambda)$ are set to be $(0, \phi_2; 1, 0.6, 0.3, 0.4, -2)$. To estimate the type I error rates (with H_0: $\phi_2 = 0$) and powers (with H_1: $\phi_2 > 0$) for various sample sizes, where the values of ϕ_2 in H_1 are chosen to be 0.01, 0.05, 0.10. Based on the given parameters $(n, \phi_1 = 0, \phi_2, \theta_1 = 1, \theta_2 = 0.6, \alpha_1 = 0.3, \alpha_2 = 0.4, \lambda = -2)$, we generate

$$(Y_{11}^{(l)}, Y_{21}^{(l)})^\top, \ldots, (Y_{1n}^{(l)}, Y_{2n}^{(l)})^\top \overset{\text{iid}}{\sim} \text{ZIGP}_\lambda^{(\text{II})}(0, \phi_2; \theta_1, \theta_2, \alpha_1, \alpha_2)$$

for $l = 1, \ldots, L$ ($L = 1,000$). For each group of samples $\{Y_{1j}^{(l)}, Y_{2j}^{(l)})^\top\}_{j=1}^n$, we conduct the testing hypothesis. Let r_6 denote the number of rejecting the null hypothesis H_0: $\phi_2 = 0$ by the test statistic T_6 given by (41). Then

the actual significance level can be estimated by r_6/L with $\phi_2 = 0$ and the power of the test statistic T_6 can be estimated by r_6/L with $\phi_2 > 0$.

The empirical levels/powers of the LRT statistic T_6 are summarized in Table 13 and Figure 6 portrays the trend of the type I error rates and powers of the LRT in testing H_0: $\phi_2 = 0$ against $H_1 : \phi_2 > 0$ with three different values of $\phi_2 > 0$ for various sample sizes.

From Figures 3–6, we can see that the lines for empirical levels of these tests all fluctuate near the line of $\alpha = 0.05$, indicating that they perform well in controlling the type I error rates around the pre-chosen nominal level. The curves for empirical powers under different situations are non-decreasing in general although including some downs, which still shows that the tests tend to be more powerful as the sample size n becomes larger.

Table 13
Empirical levels/powers of the LRT statistic T_6 based on $L = 1,000$ replications

Sample size (n)	Empirical level	Empirical power ϕ_2			Sample size (n)	Empirical level	Empirical power ϕ_2		
		0.01	0.05	0.10			0.01	0.05	0.10
500	0.052	0.046	0.092	0.134	750	0.046	0.066	0.110	0.154
550	0.046	0.054	0.092	0.138	800	0.052	0.070	0.112	0.158
600	0.056	0.058	0.096	0.140	850	0.050	0.070	0.114	0.168
650	0.050	0.056	0.106	0.140	900	0.042	0.074	0.116	0.188
700	0.048	0.060	0.106	0.146	950	0.048	0.076	0.120	0.200

6. Discussion

This paper extended the new bivariate generalized Poisson distribution [13] to zero-inflated situation that can model bivariate count data with zero-inflated marginal distributions. An important characteristic is that it can model negatively correlated data while most existing models cannot. It retains the flexibility of GP distribution in allowing either over- or under-dispersion and accepts either positive or negative correlation from data. A common parameter or two different parameters on zero inflation of the margins are both considered, including only one marginal zero inflation in the model as a special case. However, due to the limitation that it can only apply to the bivariate case, zero-inflated multivariate GP model are desired on the basis of the new multivariate GP distribution [1] in which the correlation can also be positive, zero or negative.

Another problem on the function in the multiplicative factor are all chosen to be $g(y) = e^{-y}$ when setting distribution, see [20, 12, 13, 1]. While no evidence is provided to show it outperformed others. Actually, the constructed structure in the expression of probability mass function represents a group of distributions. In the future, we may extend the proposed distribution to parametric model by considering covariates and semi-parametric random effects model ([3]).
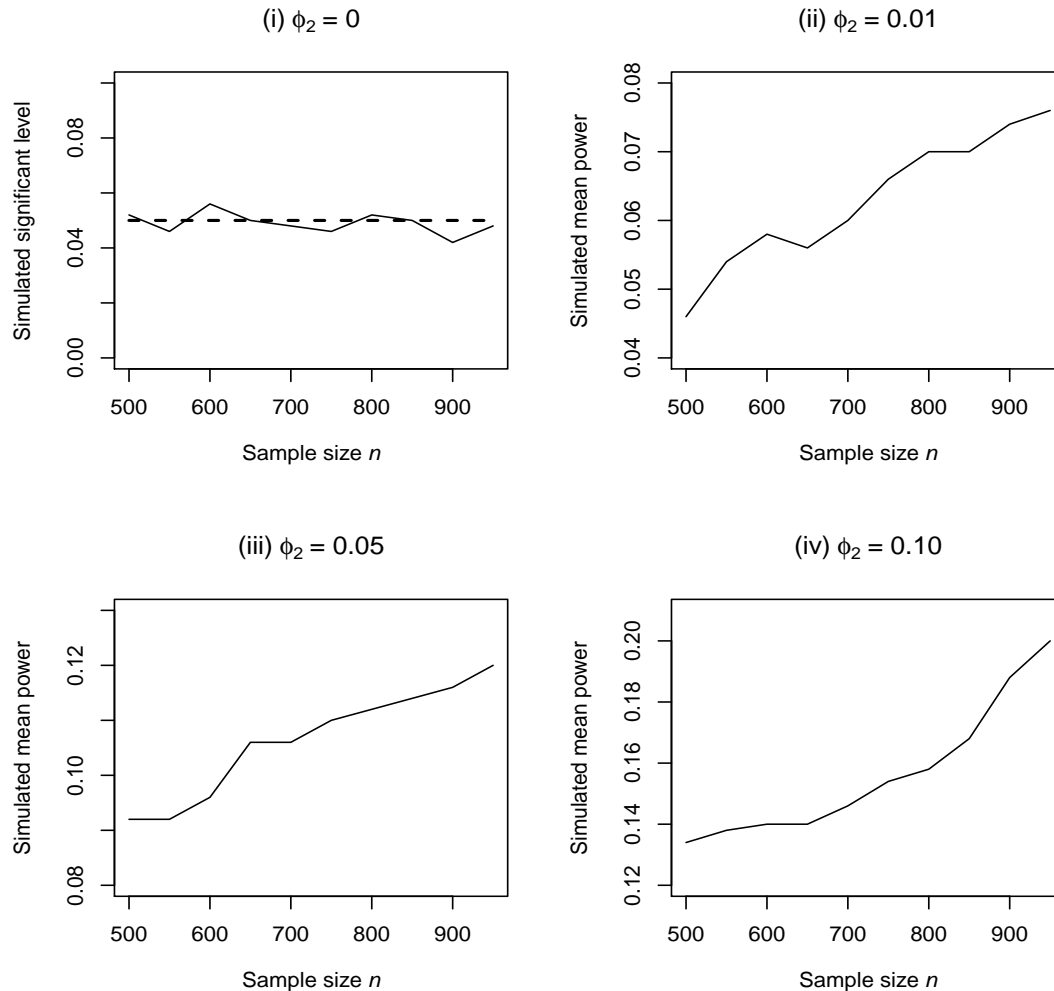
Figure 6. (i) The type I error rates for testing H_0: $\phi_2 = 0$ against H_1: $\phi_2 > 0$ in the Type II bivariate ZIGP$_\lambda$ distribution and the dashed line is set as the predetermined significance level of $\alpha = 0.05$; (ii) the powers when $\phi_2 = 0.01$ in H_1; (iii) the powers when $\phi_2 = 0.05$ in H_1; (iv) the powers when $\phi_2 = 0.10$ in H_1.

Appendix A: Derivation of (10)

Since $c_i = \exp[\theta_i(s_i - 1)]$ and $\log s_i - \alpha_i \theta_i(s_i - 1) + 1 = 0$ for $i = 1, 2$, the related partial derivatives are given by

$$\frac{\partial c_i}{\partial \theta_i} = (s_i - 1)(1 - \alpha_i \theta_i s_i)^{-1} c_i,$$

$$\frac{\partial c_i}{\partial \alpha_i} = \theta_i^2 s_i(s_i - 1)(1 - \alpha_i \theta_i s_i)^{-1} c_i,$$

$$\frac{\partial^2 c_i}{\partial \theta_i^2} = \left\{ (s_i - 1)^2(1 - \alpha_i \theta_i s_i)^{-2} + \alpha_i s_i(s_i - 1)(1 - \alpha_i \theta_i s_i)^{-2}[2 + \alpha_i \theta_i(s_i - 1)(1 - \alpha_i \theta_i s_i)^{-1}] \right\} c_i,$$

$$\frac{\partial^2 c_i}{\partial \alpha_i^2} = \left\{ [s_i(s_i - 1)\theta_i^2(1 - \alpha_i \theta_i s_i)^{-1}]^2 + \theta_i^3 s_i(s_i - 1)(1 - \alpha_i \theta_i s_i)^{-2}[2s_i + (s_i - 1)(1 - \alpha_i \theta_i s_i)^{-1}] \right\} c_i,$$

$$\frac{\partial^2 c_i}{\partial \theta_i \partial \alpha_i} = \left\{ s_i[(s_i - 1)\theta_i(1 - \alpha_i \theta_i s_i)^{-1}]^2 + \theta_i s_i(s_i - 1)(1 - \alpha_i \theta_i s_i)^{-2}[2 + \alpha_i \theta_i(s_i - 1)(1 - \alpha_i \theta_i s_i)^{-1}] \right\} c_i.$$

First we set

$$A_1 = \frac{\lambda(1 - c_2)}{1 + \lambda(1 - c_1)(1 - c_2)}, \qquad A_{2j} = \frac{\lambda(\mathrm{e}^{-y_{2j}} - c_2)}{1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(\mathrm{e}^{-y_{2j}} - c_2)},$$

$$B_1 = \frac{\lambda(1 - c_1)}{1 + \lambda(1 - c_1)(1 - c_2)}, \qquad B_{2j} = \frac{\lambda(\mathrm{e}^{-y_{1j}} - c_1)}{1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(\mathrm{e}^{-y_{2j}} - c_2)},$$

$$C_1 = \frac{(1 - c_1)(1 - c_2)}{1 + \lambda(1 - c_1)(1 - c_2)}, \qquad C_{2j} = \frac{(\mathrm{e}^{-y_{1j}} - c_1)(\mathrm{e}^{-y_{2j}} - c_2)}{1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(\mathrm{e}^{-y_{2j}} - c_2)},$$

$$D_1 = \frac{\lambda}{[1 + \lambda(1 - c_1)(1 - c_2)]^2}, \qquad D_{2j} = \frac{\lambda}{[1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(\mathrm{e}^{-y_{2j}} - c_2)]^2},$$

$$E_1 = \frac{1 - c_2}{[1 + \lambda(1 - c_1)(1 - c_2)]^2}, \qquad E_{2j} = \frac{(\mathrm{e}^{-y_{2j}} - c_2)}{[1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(\mathrm{e}^{-y_{2j}} - c_2)]^2},$$

$$F_1 = \frac{1 - c_1}{[1 + \lambda(1 - c_1)(1 - c_2)]^2}, \qquad F_{2j} = \frac{(\mathrm{e}^{-y_{1j}} - c_1)}{[1 + \lambda(\mathrm{e}^{-y_{1j}} - c_1)(\mathrm{e}^{-y_{2j}} - c_2)]^2}.$$

The first and second partial derivatives in (10) are given by the following expressions:

$$\frac{\partial \ell_2}{\partial \theta_1} = z + \frac{N_1}{\theta_1} - n - N_1 \alpha_1 + \left(zA_1 - \sum_{j=1}^{n} A_{2j} \right) \frac{\partial c_1}{\partial \theta_1},$$

$$\frac{\partial \ell_2}{\partial \theta_2} = z + \frac{N_2}{\theta_2} - n - N_2 \alpha_2 + \left(zB_1 - \sum_{j=1}^{n} B_{2j} \right) \frac{\partial c_2}{\partial \theta_2},$$

$$\frac{\partial \ell_2}{\partial \alpha_1} = \sum_{j=1}^{n} \frac{y_{1j}(y_{1j} - 1)}{1 + \alpha_1 y_{1j}} - N_1 \theta_1 + \left(zA_1 - \sum_{j=1}^{n} A_{2j} \right) \frac{\partial c_1}{\partial \alpha_1},$$

$$\frac{\partial \ell_2}{\partial \alpha_2} = \sum_{j=1}^{n} \frac{y_{2j}(y_{2j} - 1)}{1 + \alpha_2 y_{2j}} - N_2 \theta_2 + \left(zB_1 - \sum_{j=1}^{n} B_{2j} \right) \frac{\partial c_2}{\partial \alpha_2},$$

$$\frac{\partial \ell_2}{\partial \lambda} = -zC_1 + \sum_{j=1}^{n} C_{2j},$$

and

$$\frac{\partial^2 \ell_2}{\partial \theta_1^2} = -\frac{N_1}{\theta_1^2} + \left(zA_1^2 - \sum_{j=1}^{n} A_{2j}^2\right)\left(\frac{\partial c_1}{\partial \theta_1}\right)^2 + \left(zA_1 - \sum_{j=1}^{n} A_{2j}\right)\frac{\partial^2 c_1}{\partial \theta_1^2},$$

$$\frac{\partial^2 \ell_2}{\partial \theta_2^2} = -\frac{N_2}{\theta_2^2} + \left(zB_1^2 - \sum_{j=1}^{n} B_{2j}^2\right)\left(\frac{\partial c_2}{\partial \theta_2}\right)^2 + \left(zB_1 - \sum_{j=1}^{n} B_{2j}\right)\frac{\partial^2 c_2}{\partial \theta_2^2},$$

$$\frac{\partial^2 \ell_2}{\partial \alpha_1^2} = -\sum_{j=1}^{n} \frac{y_{1j}^2(y_{1j}-1)}{(1+\alpha_1 y_{1j})^2} + \left(zA_1^2 - \sum_{j=1}^{n} A_{2j}^2\right)\left(\frac{\partial c_1}{\partial \alpha_1}\right)^2 + \left(zA_1 - \sum_{j=1}^{n} A_{2j}\right)\frac{\partial^2 c_1}{\partial \alpha_1^2},$$

$$\frac{\partial^2 \ell_2}{\partial \alpha_2^2} = -\sum_{j=1}^{n} \frac{y_{2j}^2(y_{2j}-1)}{(1+\alpha_2 y_{2j})^2} + \left(zB_1^2 - \sum_{j=1}^{n} B_{2j}^2\right)\left(\frac{\partial c_2}{\partial \alpha_2}\right)^2 + \left(zB_1 - \sum_{j=1}^{n} B_{2j}\right)\frac{\partial^2 c_2}{\partial \alpha_2^2},$$

$$\frac{\partial^2 \ell_2}{\partial \lambda^2} = zC_1^2 - \sum_{j=1}^{n} C_{2j}^2,$$

$$\frac{\partial^2 \ell_2}{\partial \theta_1 \partial \theta_2} = -\left(zD_1 - \sum_{j=1}^{n} D_{2j}\right)\frac{\partial c_1}{\partial \theta_1}\frac{\partial c_2}{\partial \theta_2},$$

$$\frac{\partial^2 \ell_2}{\partial \theta_1 \partial \alpha_1} = -N_1 + \left(zA_1^2 - \sum_{j=1}^{n} A_{2j}^2\right)\frac{\partial c_1}{\partial \theta_1}\frac{\partial c_1}{\partial \alpha_1} + \left(zA_1 - \sum_{j=1}^{n} A_{2j}\right)\frac{\partial^2 c_1}{\partial \theta_1 \partial \alpha_1},$$

$$\frac{\partial^2 \ell_2}{\partial \theta_1 \partial \alpha_2} = -\left(zD_1 - \sum_{j=1}^{n} D_{2j}\right)\frac{\partial c_1}{\partial \theta_1}\frac{\partial c_2}{\partial \alpha_2},$$

$$\frac{\partial^2 \ell_2}{\partial \theta_1 \partial \lambda} = \left(zE_1 - \sum_{j=1}^{n} E_{2j}\right)\frac{\partial c_1}{\partial \theta_1},$$

$$\frac{\partial^2 \ell_2}{\partial \theta_2 \partial \alpha_1} = -\left(zD_1 - \sum_{j=1}^{n} D_{2j}\right)\frac{\partial c_1}{\partial \alpha_1}\frac{\partial c_2}{\partial \theta_2},$$

$$\frac{\partial^2 \ell_2}{\partial \theta_2 \partial \alpha_2} = -N_2 + \left(zB_1^2 - \sum_{j=1}^{n} B_{2j}^2\right)\left(\frac{\partial c_2}{\partial \theta_2}\right)^2 + \left(zB_1 - \sum_{j=1}^{n} B_{2j}\right)\frac{\partial^2 c_2}{\partial \theta_2^2},$$

$$\frac{\partial^2 \ell_2}{\partial \theta_2 \partial \lambda} = \left(zF_1 - \sum_{j=1}^{n} F_{2j}\right)\frac{\partial c_2}{\partial \theta_2},$$

$$\frac{\partial^2 \ell_2}{\partial \alpha_1 \partial \alpha_2} = -\left(zD_1 - \sum_{j=1}^{n} D_{2j}\right)\frac{\partial c_1}{\partial \alpha_1}\frac{\partial c_2}{\partial \alpha_2},$$

$$\frac{\partial^2 \ell_2}{\partial \alpha_1 \partial \lambda} = \left(zE_1 - \sum_{j=1}^{n} E_{2j} \right) \frac{\partial c_1}{\partial \alpha_1},$$

$$\frac{\partial^2 \ell_2}{\partial \alpha_2 \partial \lambda} = \left(zF_1 - \sum_{j=1}^{n} F_{2j} \right) \frac{\partial c_2}{\partial \alpha_2},$$

where $\ell_2 \,\hat{=}\, \ell_2(\boldsymbol{\theta}|Y_{\text{com}})$ which is given in (9).

Appendix B: Derivation of (33)

First we set

$$G_{1j} = \frac{\lambda(e^{-y_{2j}} - c_2)}{1 + \lambda(e^{-y_{1j}} - c_1)(e^{-y_{2j}} - c_2)}, \qquad G_2 = \frac{\lambda(1 - c_2)}{1 + \lambda(1 - c_1)(1 - c_2)},$$

$$G_{3j} = \frac{\lambda(e^{-y_{2j}} - c_2)}{1 + \lambda(1 - c_1)(e^{-y_{2j}} - c_2)}, \qquad G_{4j} = \frac{\lambda(1 - c_2)}{1 + \lambda(e^{-y_{1j}} - c_1)(1 - c_2)},$$

$$H_{1j} = \frac{\lambda(e^{-y_{1j}} - c_1)}{1 + \lambda(e^{-y_{1j}} - c_1)(e^{-y_{2j}} - c_2)}, \qquad H_2 = \frac{\lambda(1 - c_1)}{1 + \lambda(1 - c_1)(1 - c_2)},$$

$$H_{3j} = \frac{\lambda(1 - c_1)}{1 + \lambda(1 - c_1)(e^{-y_{2j}} - c_2)}, \qquad H_{4j} = \frac{\lambda(e^{-y_{1j}} - c_1)}{1 + \lambda(e^{-y_{1j}} - c_1)(1 - c_2)},$$

$$I_{1j} = \frac{(e^{-y_{1j}} - c_1)(e^{-y_{2j}} - c_2)}{1 + \lambda(e^{-y_{1j}} - c_1)(e^{-y_{2j}} - c_2)}, \qquad I_2 = \frac{(1 - c_1)(1 - c_2)}{1 + \lambda(1 - c_1)(1 - c_2)},$$

$$I_{3j} = \frac{(1 - c_1)(e^{-y_{2j}} - c_2)}{1 + \lambda(1 - c_1)(e^{-y_{2j}} - c_2)}, \qquad I_{4j} = \frac{(e^{-y_{1j}} - c_1)(1 - c_2)}{1 + \lambda(e^{-y_{1j}} - c_1)(1 - c_2)},$$

$$J_{1j} = \frac{\lambda}{[1 + \lambda(e^{-y_{1j}} - c_1)(e^{-y_{2j}} - c_2)]^2}, \qquad J_2 = \frac{\lambda}{[1 + \lambda(1 - c_1)(1 - c_2)]^2},$$

$$J_{3j} = \frac{\lambda}{[1 + \lambda(1 - c_1)(e^{-y_{2j}} - c_2)]^2}, \qquad J_{4j} = \frac{\lambda}{[1 + \lambda(e^{-y_{1j}} - c_1)(1 - c_2)]^2},$$

$$K_{1j} = \frac{(e^{-y_{2j}} - c_2)}{[1 + \lambda(e^{-y_{1j}} - c_1)(e^{-y_{2j}} - c_2)]^2}, \qquad K_2 = \frac{(1 - c_2)}{[1 + \lambda(1 - c_1)(1 - c_2)]^2},$$

$$K_{3j} = \frac{(e^{-y_{2j}} - c_2)}{[1 + \lambda(1 - c_1)(e^{-y_{2j}} - c_2)]^2}, \qquad K_{4j} = \frac{(1 - c_2)}{[1 + \lambda(e^{-y_{1j}} - c_1)(1 - c_2)]^2},$$

$$L_{1j} = \frac{(e^{-y_{1j}} - c_1)}{[1 + \lambda(e^{-y_{1j}} - c_1)(e^{-y_{2j}} - c_2)]^2}, \qquad L_2 = \frac{(1 - c_1)}{[1 + \lambda(1 - c_1)(1 - c_2)]^2},$$

$$L_{3j} = \frac{(1 - c_1)}{[1 + \lambda(1 - c_1)(e^{-y_{2j}} - c_2)]^2}, \qquad L_{4j} = \frac{(e^{-y_{1j}} - c_1)}{[1 + \lambda(e^{-y_{1j}} - c_1)(1 - c_2)]^2}.$$

The first derivatives in (33) are given by

$$\frac{\partial \ell_2^*}{\partial \theta_1} = -(n - u_0 - u_1 - N_w) + \frac{N_1}{\theta_1} - N_1 \alpha_1$$

$$- \left[\sum_{j=1}^{n} G_{1j} - (u_0 + u_1 + u_2)G_2 - \sum_{j \in \mathbb{J}_1} w_j G_{3j} - \sum_{j \in \mathbb{J}_2} z_j G_{4j} \right] \frac{\partial c_1}{\partial \theta_1},$$

$$\frac{\partial \ell_2^*}{\partial \theta_2} = -(n - u_0 - u_2 - N_z) + \frac{N_2}{\theta_2} - N_2 \alpha_2$$

$$- \left[\sum_{j=1}^{n} H_{1j} - (u_0 + u_1 + u_2) H_2 - \sum_{j \in \mathbb{J}_1} w_j H_{3j} - \sum_{j \in \mathbb{J}_2} z_j H_{4j} \right] \frac{\partial c_2}{\partial \theta_2},$$

$$\frac{\partial \ell_2^*}{\partial \alpha_1} = -N_1 \theta_1 + \sum_{j=1}^{n} \frac{y_{1j}(y_{1j} - 1)}{1 + \alpha_1 y_{1j}} - \left[\sum_{j=1}^{n} G_{1j} - (u_0 + u_1 + u_2) G_2 - \sum_{j \in \mathbb{J}_1} w_j G_{3j} - \sum_{j \in \mathbb{J}_2} z_j G_{4j} \right] \frac{\partial c_1}{\partial \alpha_1},$$

$$\frac{\partial \ell_2^*}{\partial \alpha_2} = -N_2 \theta_2 + \sum_{j=1}^{n} \frac{y_{2j}(y_{2j} - 1)}{1 + \alpha_2 y_{2j}} - \left[\sum_{j=1}^{n} H_{1j} - (u_0 + u_1 + u_2) H_2 - \sum_{j \in \mathbb{J}_1} w_j H_{3j} - \sum_{j \in \mathbb{J}_2} z_j H_{4j} \right] \frac{\partial c_2}{\partial \alpha_2},$$

$$\frac{\partial \ell_2^*}{\partial \lambda} = \sum_{j=1}^{n} I_{1j} - (u_0 + u_1 + u_2) I_2 - \sum_{j \in \mathbb{J}_1} w_j I_{3j} - \sum_{j \in \mathbb{J}_2} z_j I_{4j}.$$

The second derivatives in (33) are given by

$$\frac{\partial^2 \ell_2^*}{\partial \theta_1^2} = -\frac{N_1}{\theta_1^2} - \left[\sum_{j=1}^{n} G_{1j}^2 - (u_0 + u_1 + u_2) G_2^2 - \sum_{j \in \mathbb{J}_1} w_j G_{3j}^2 - \sum_{j \in \mathbb{J}_2} z_j G_{4j}^2 \right] \left(\frac{\partial c_1}{\partial \theta_1} \right)^2$$

$$- \left[\sum_{j=1}^{n} G_{1j} - (u_0 + u_1 + u_2) G_2 - \sum_{j \in \mathbb{J}_1} w_j G_{3j} - \sum_{j \in \mathbb{J}_2} z_j G_{4j} \right] \frac{\partial^2 c_1}{\partial \theta_1^2},$$

$$\frac{\partial^2 \ell_2^*}{\partial \theta_2^2} = -\frac{N_2}{\theta_2^2} - \left[\sum_{j=1}^{n} H_{1j}^2 - (u_0 + u_1 + u_2) H_2^2 - \sum_{j \in \mathbb{J}_1} w_j H_{3j}^2 - \sum_{j \in \mathbb{J}_2} z_j H_{4j}^2 \right] \left(\frac{\partial c_2}{\partial \theta_2} \right)^2$$

$$- \left[\sum_{j=1}^{n} H_{1j} - (u_0 + u_1 + u_2) H_2 - \sum_{j \in \mathbb{J}_1} w_j H_{3j} - \sum_{j \in \mathbb{J}_2} z_j H_{4j} \right] \frac{\partial^2 c_2}{\partial \theta_2^2},$$

$$\frac{\partial^2 \ell_2^*}{\partial \alpha_1^2} = -\sum_{j=1}^{n} \frac{y_{1j}^2(y_{1j} - 1)}{(1 + \alpha_1 y_{1j})^2} - \left[\sum_{j=1}^{n} G_{1j}^2 - (u_0 + u_1 + u_2) G_2^2 - \sum_{j \in \mathbb{J}_1} w_j G_{3j}^2 - \sum_{j \in \mathbb{J}_2} z_j G_{4j}^2 \right] \left(\frac{\partial c_1}{\partial \alpha_1} \right)^2$$

$$- \left[\sum_{j=1}^{n} G_{1j} - (u_0 + u_1 + u_2) G_2 - \sum_{j \in \mathbb{J}_1} w_j G_{3j} - \sum_{j \in \mathbb{J}_2} z_j G_{4j} \right] \frac{\partial^2 c_1}{\partial \alpha_1^2},$$

$$\frac{\partial^2 \ell_2^*}{\partial \alpha_2^2} = -\sum_{j=1}^{n} \frac{y_{2j}^2(y_{2j} - 1)}{(1 + \alpha_2 y_{2j})^2} - \left[\sum_{j=1}^{n} H_{1j}^2 - (u_0 + u_1 + u_2) H_2^2 - \sum_{j \in \mathbb{J}_1} w_j H_{3j}^2 - \sum_{j \in \mathbb{J}_2} z_j H_{4j}^2 \right] \left(\frac{\partial c_2}{\partial \alpha_2} \right)^2$$

$$- \left[\sum_{j=1}^{n} H_{1j} - (u_0 + u_1 + u_2) H_2 - \sum_{j \in \mathbb{J}_1} w_j H_{3j} - \sum_{j \in \mathbb{J}_2} z_j H_{4j} \right] \frac{\partial^2 c_2}{\partial \alpha_2^2},$$

$$\frac{\partial^2 \ell_2^*}{\partial \lambda^2} = -\sum_{j=1}^n I_{1j}^2 + (u_0 + u_1 + u_2)I_2^2 + \sum_{j \in \mathbb{J}_1} w_j I_{3j}^2 + \sum_{j \in \mathbb{J}_2} z_j I_{4j}^2,$$

$$\frac{\partial^2 \ell_2^*}{\partial \theta_1 \partial \theta_2} = \left[\sum_{j=1}^n J_{1j} - (u_0 + u_1 + u_2)J_2 - \sum_{j \in \mathbb{J}_1} w_j J_{3j} - \sum_{j \in \mathbb{J}_2} z_j J_{4j} \right] \frac{\partial c_1}{\partial \theta_1} \frac{\partial c_2}{\partial \theta_2},$$

$$\frac{\partial^2 \ell_2^*}{\partial \theta_1 \partial \alpha_1} = -N_1 - \left[\sum_{j=1}^n G_{1j}^2 - (u_0 + u_1 + u_2)G_2^2 - \sum_{j \in \mathbb{J}_1} w_j G_{3j}^2 - \sum_{j \in \mathbb{J}_2} z_j G_{4j}^2 \right] \frac{\partial c_1}{\partial \theta_1} \frac{\partial c_1}{\partial \alpha_1}$$
$$- \left[\sum_{j=1}^n G_{1j} - (u_0 + u_1 + u_2)G_2 - \sum_{j \in \mathbb{J}_1} w_j G_{3j} - \sum_{j \in \mathbb{J}_2} z_j G_{4j} \right] \frac{\partial^2 c_1}{\partial \theta_1 \partial \alpha_1},$$

$$\frac{\partial^2 \ell_2^*}{\partial \theta_1 \partial \alpha_2} = \left[\sum_{j=1}^n J_{1j} - (u_0 + u_1 + u_2)J_2 - \sum_{j \in \mathbb{J}_1} w_j J_{3j} - \sum_{j \in \mathbb{J}_2} z_j J_{4j} \right] \frac{\partial c_1}{\partial \theta_1} \frac{\partial c_2}{\partial \alpha_2},$$

$$\frac{\partial^2 \ell_2^*}{\partial \theta_1 \partial \lambda} = -\left[\sum_{j=1}^n K_{1j} - (u_0 + u_1 + u_2)K_2 - \sum_{j \in \mathbb{J}_1} w_j K_{3j} - \sum_{j \in \mathbb{J}_2} z_j K_{4j} \right] \frac{\partial c_1}{\partial \theta_1},$$

$$\frac{\partial^2 \ell_2^*}{\partial \theta_2 \partial \alpha_1} = \left[\sum_{j=1}^n J_{1j} - (u_0 + u_1 + u_2)J_2 - \sum_{j \in \mathbb{J}_1} w_j J_{3j} - \sum_{j \in \mathbb{J}_2} z_j J_{4j} \right] \frac{\partial c_2}{\partial \theta_2} \frac{\partial c_1}{\partial \alpha_1},$$

$$\frac{\partial^2 \ell_2^*}{\partial \theta_2 \partial \alpha_2} = -N_2 - \left[\sum_{j=1}^n H_{1j}^2 - (u_0 + u_1 + u_2)H_2^2 - \sum_{j \in \mathbb{J}_1} w_j H_{3j}^2 - \sum_{j \in \mathbb{J}_2} z_j H_{4j}^2 \right] \frac{\partial c_2}{\partial \theta_2} \frac{\partial c_2}{\partial \alpha_2}$$
$$- \left[\sum_{j=1}^n H_{1j} - (u_0 + u_1 + u_2)H_2 - \sum_{j \in \mathbb{J}_1} w_j H_{3j} - \sum_{j \in \mathbb{J}_2} z_j H_{4j} \right] \frac{\partial^2 c_2}{\partial \theta_2 \partial \alpha_2},$$

$$\frac{\partial^2 \ell_2^*}{\partial \theta_2 \partial \lambda} = -\left[\sum_{j=1}^n L_{1j} - (u_0 + u_1 + u_2)L_2 - \sum_{j \in \mathbb{J}_1} w_j L_{3j} - \sum_{j \in \mathbb{J}_2} z_j L_{4j} \right] \frac{\partial c_2}{\partial \theta_2},$$

$$\frac{\partial^2 \ell_2^*}{\partial \alpha_1 \partial \alpha_2} = \left[\sum_{j=1}^n J_{1j} - (u_0 + u_1 + u_2)J_2 - \sum_{j \in \mathbb{J}_1} w_j J_{3j} - \sum_{j \in \mathbb{J}_2} z_j J_{4j} \right] \frac{\partial c_1}{\partial \alpha_1} \frac{\partial c_2}{\partial \alpha_2},$$

$$\frac{\partial^2 \ell_2^*}{\partial \alpha_1 \partial \lambda} = -\left[\sum_{j=1}^n K_{1j} - (u_0 + u_1 + u_2)K_2 - \sum_{j \in \mathbb{J}_1} w_j K_{3j} - \sum_{j \in \mathbb{J}_2} z_j K_{4j} \right] \frac{\partial c_1}{\partial \alpha_1},$$

$$\frac{\partial^2 \ell_2^*}{\partial \alpha_2 \partial \lambda} = -\left[\sum_{j=1}^n L_{1j} - (u_0 + u_1 + u_2)L_2 - \sum_{j \in \mathbb{J}_1} w_j L_{3j} - \sum_{j \in \mathbb{J}_2} z_j L_{4j} \right] \frac{\partial c_2}{\partial \alpha_2},$$

where $\ell_2^* \hat{=} \ell_2^*(\boldsymbol{\theta}|Y_{\text{com}})$ which is given in (32).

Appendix C: The joint pmf of the Type II bivariate ZIGP$_\lambda$ distribution

Let $(Y_1, Y_2)^\top \sim \text{ZIGP}_\lambda^{(\text{II})}(\phi_1, \phi_2; \theta_1, \theta_2, \alpha_1, \alpha_2)$. The joint pmf of $(Y_1, Y_2)^\top$ is given by

$$\Pr(Y_1 = y_1, \ Y_2 = y_2) = \Pr(Z_1 X_1 = y_1, \ Z_2 X_2 = y_2).$$

We consider the following four cases. (i) If $y_1 = 0$ and $y_2 = 0$, we have

$$\Pr(Y_1 = 0, Y_2 = 0)$$

$$= \ \Pr(Z_1 = 0, Z_2 = 0) + \Pr(Z_1 = 0, Z_2 = 1, X_2 = 0)$$

$$+ \ \Pr(Z_1 = 1, X_1 = 1, Z_2 = 0) + \Pr(Z_1 = 1, X_1 = 0, Z_2 = 1, X_2 = 0)$$

$$= \ \phi_1 \phi_2 + \phi_1 (1 - \phi_2) e^{-\theta_2} + (1 - \phi_1) \phi_2 e^{-\theta_1} + (1 - \phi_1)(1 - \phi_2) e^{-(\theta_1 + \theta_2)} [1 + \lambda(1 - c_1)(1 - c_2)],$$

where c_1 and c_2 defined in (1). (ii) If $y_1 = 0$ and $y_2 > 0$, we have

$$\Pr(Y_1 = 0, Y_2 = y_2)$$

$$= \ \Pr(Z_1 = 0, Z_2 = 1, X_2 = y_2) + \Pr(Z_1 = 1, X_1 = 0, Z_2 = 1, X_2 = y_2)$$

$$= \ (1 - \phi_2) \frac{\theta_2^{y_2} (1 + \alpha_2 y_2)^{y_2 - 1}}{y_2!} \exp[-\theta_2(1 + \alpha_2 y_2)] \left\{ \phi_1 + (1 - \phi_1) e^{-\theta_1} [1 + \lambda(1 - c_1)(e^{-y_2} - c_2)] \right\}.$$

(iii) If $y_1 > 0$ and $y_2 = 0$, we have

$$\Pr(Y_1 = y_1, Y_2 = 0)$$

$$= \ \Pr(Z_1 = 1, X_1 = y_1, Z_2 = 0) + \Pr(Z_1 = 1, X_1 = y_1, Z_2 = 1, X_2 = 0)$$

$$= \ (1 - \phi_1) \frac{\theta_1^{y_1} (1 + \alpha_1 y_1)^{y_1 - 1}}{y_1!} \exp[-\theta_1(1 + \alpha_1 y_1)] \left\{ \phi_2 + (1 - \phi_2) e^{-\theta_2} [1 + \lambda(e^{-y_1} - c_1)(1 - c_2)] \right\}.$$

(iv) If $y_1 > 0$ and $y_2 > 0$, we have

$$\Pr(Y_1 = y_1, Y_2 = y_2)$$

$$= \ \Pr(Z_1 = 1, X_1 = y_1, Z_2 = 1, X_2 = y_2)$$

$$= \ (1 - \phi_1)(1 - \phi_2) \prod_{i=1}^{2} \left[\frac{\theta_i (1 + \alpha_i y_i)^{y_i - 1}}{y_i!} \exp[-\theta_i(1 + \alpha_i y_i)] \right] [1 + \lambda(e^{-y_1} - c_1)(e^{-y_2} - c_2)].$$

By combining above four cases, we obtain the joint pmf of $(Y_1, Y_2)^\top$ given by (29).

REFERENCES

1. A.J. Ademola, and M.U. Ahamefula, *Multivariate generalized Poisson distribution for inference on selected non-communicable diseases in Lagos State, Nigeria*, Journal of Modern Applied Statistical Methods, vol. 11, no. 2, pp. 524–529, 2012.
2. H. Akaike, *A new look at the statistical model identification*, IEEE Transactions on Automatic Control, vol. 19, no. 6, pp. 716–723, 1974.
3. T. Baghfalaki, and M. Ganjali, *A fast algorithm for using semi-parametric random effects model for analyzing longitudinal data*, Statistics, Optimization & Information, vol. 2, no. 4, pp. 339-351, 2014.
4. D.A. Berry, *Logarithmic transformations in ANOVA*, Biometrics, vol. 43, pp. 439–456, 1987.
5. A. Cameron, and P. Trivedi, *Regression Analysis of Count Data (2nd Edition)*. Cambridge University Press, New York, 2013.
6. H. Chernoff, *On the distribution of the likelihood ratio*, The Annals of Mathematical Statistics, vol. 25, pp. 573–578, 1954.
7. P.C. Consul, *Generalized Poisson Distributions: Properties and Applications*, Marcel Dekker, New York, 1989.
8. T.J. DiCiccio, and B. Efron, *Bootstrap confidence intervals*, Statistical Science, vol. 11, no. 3, pp. 189–212, 1996.
9. F. Famoye, and P.C. Consul, *Bivariate generalized Poisson distribution with some applications*, Metrika, vol. 42, no. 1, pp. 127–138, 1995.
10. F. Famoye, J.T. Wulu, and K.P. Singh, *On the generalized Poisson regression model with an application to accident data*, Journal of Data Science, vol. 2, pp. 287–295, 2004.
11. F. Famoye, and K.P. Singh, *Zero-inflated generalized Poisson regression model with an application to domestic violence data*, Journal of Data Science, vol. 4, no. 1, pp. 117–130, 2006.
12. F. Famoye, *On the bivariate negative binomial regression model*, Journal of Applied Statistics, vol. 37, no. 6, pp. 969–981, 2010.
13. F. Famoye, *A new bivariate generalized Poisson distribution*, Statistica Neerlandica, vol. 64, no. 1, pp. 112–124, 2010.
14. Z.D. Feng, and C.E. McCulloch, *Statistical inference using maximum likelihood estimation and the generalized likelihood ratio when the true parameter is on the boundary of the parameter space*, Statistics & Probability Letters, vol. 13, no. 4, pp. 325–332, 1992.
15. P. Holgate, *Estimation for the bivariate Poisson distribution*, Biometrika, vol. 51, no. 1-2, pp. 241–287, 1964.
16. N. Jansakul, and J.P. Hinde, *Score tests for zero-inflated Poisson models*, Computational Statistics & Data Analysis, vol. 40, pp. 75–96, 2002.
17. H. Joe, and R. Zhu, *Generalized Poisson distribution: The property of mixture of Poisson and comparison with negative binomial distribution*, Biometrical Journal, vol. 47, no. 2, pp. 219–229, 2005.
18. D. Karlis, and I. Ntzoufras, *Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R*, Journal of Statistical Software, vol. 14, no. 10, 2005.
19. P.M. Krishna, and S.D. Tukaram, *Bivariate zero-inflated power series distribution*, Applied Mathematics, vol. 2, no. 7, pp. 824–829, 2011.
20. J. Lakshminarayana, S.N.N. Pandit, and S.K. Rao, *On a bivariate Poisson distribution*, Communications in Statistics—Theory and Methods, vol. 28, no. 2, pp. 267–276, 1999.
21. C.S. Li, J.C. Lu, J. Park, K. Kim, P.A. Brinkley, and J.P. Peterson, *Multivariate zero-inflated Poisson models and their applications*, Technometrics, vol. 41, no. 1, pp. 29–38, 1999.
22. Y. Liu, and G.L. Tian, *Type I multivariate zero-inflated Poisson distribution with applications*, Computational Statistics & Data Analysis, vol. 83, pp. 200–222, 2015.
23. T.A. Louis, *Finding the observed information matrix when using the EM algorithm*, Journal of the Royal Statistical Society, Series B, vol. 44, no. 2, pp. 226–233, 1982.
24. G. Schwarz, *Estimating the dimension of a model*, The Annals of Statistics, vol. 6, no. 2, pp. 461–464, 1978.
25. S.G. Self, and K.Y. Liang, *Asymptotic properties of the maximum likelihood estimators and likelihood ratio tests under nonstandard conditions*, Journal of the American Statistical Association, vol. 82, pp. 605–610, 1987.