Statistics and Its Interface Volume 10 (2017) 291–311

Type I multivariate zero-inflated generalized Poisson distribution with applications

Xi-Fen Huang, Guo-Liang Tian*, Chi Zhang, and Xuejun Jiang

Excessive zeros in multivariate count data are often encountered in practice. Since the Poisson distribution only possesses the property of equi-dispersion, the existing Type I multivariate zero-inflated Poisson distribution (Liu and Tian, 2015, CSDA) [15] cannot be used to model multivariate zero-inflated count data with over-dispersion or under-dispersion. In this paper, we extend the univariate *zero-inflated generalized Poisson* (ZIGP) distribution to Type I multivariate ZIGP distribution via stochastic representation aiming to model positively correlated multivariate zero-inflated count data with over-dispersion or under-dispersion. Its distributional theories and associated properties are derived. Due to the complexity of the ZIGP model, we provide four useful algorithms (a very fast Fisher-scoring algorithm, an expectation/conditional-maximization algorithm, a simple EM algorithm and an explicit majorization–minimization algorithm) for finding maximum likelihood estimates of parameters of interest and develop efficient statistical inference methods for the proposed model. Simulation studies for investigating the accuracy of point estimates and confidence interval estimates and comparing the likelihood ratio test with the score test are conducted. Under both AIC and BIC, our analyses of the two data sets show that Type I multivariate ZIGP model is superior over Type I multivariate zero-inflated Poisson model.

Keywords and phrases: AIC, BIC, EM algorithm, Fisher scoring algorithm, MM algorithm, Multivariate zero-inflated generalized Poisson distribution, Zero-inflated count data.

1. INTRODUCTION

Count data with excessive zeros are frequently encountered in a number of research fields such as medicine, public health, agriculture, ecology, econometrics, manufacturing and so on. Several distributions of mixture including the *zero-inflated Poisson* (ZIP), *zero-inflated binomial* (ZIB), *zero-inflated negative binomial* (ZINB) have been proposed to handle such count data. For example, Lambert (1992) [12] introduced a ZIP regression model with an application to defects in manufacturing; Hall (2000) [7] described a ZIB regression model and incorporated random effects into ZIP and ZIB models; Lee *et al.* (2001) [13] generalized the ZIP model by incorporating the extent of individual exposure; and Minami *et al.* (2007) [16] proposed the ZINB model and applied it to model the shark by catch data. Other existing models in the literature include the hurdle model (Mullahy, 1986) [17], the two-part model (Heibron, 1994) [8], and the semi-parametric model (Li, 2012) [14].

The equality of mean and variance characterizes the Poisson distribution. It has also been observed that in a population the probability of the occurrence of an event does not remain constant and changes with time and/or previous occurrences, resulting in unequal mean and variance in the data. As a useful generalization of the standard Poisson distribution, the *generalized Poisson* (GP) distribution was introduced firstly by Consul and Jain (1973) [1] as a limiting form of the generalized negative binomial distribution, implying that there is some changing tendency in the parameter with successive occurrences by adding an additional parameter. It is an important competitor to the negative binomial model when the count data are over-dispersed since the variance of the GP distribution cpuld be greater than, equal to or smaller than its mean depending on if the additional parameter is positive, zero or negative. A non-negative integer valued random variable X is said to have a GP distribution with parameter $\lambda \in \mathbb{R}_+$ and dispersion parameter θ, if its *probability mass function* (pmf) is given by (Consul and Jain, 1973 [1]; Consul and Shoukri, 1985 [2])

$$f(x;\lambda,\theta) = \begin{cases} \dfrac{\lambda(\lambda+\theta x)^{x-1}\mathrm{e}^{-\lambda-\theta x}}{x!}, & x=0,1,\ldots, \\ 0, & \text{for } x>q \text{ when } \theta<0, \end{cases} \tag{1.1}$$

where $\max(-1,-\lambda/q) < \theta \leqslant 1$ and $q \geqslant 4$ is the largest positive integer for which $\lambda+\theta q>0$ when $\theta<0$. We denote it by $X \sim \mathrm{GP}(\lambda,\theta)$. When $\theta=0$, the $\mathrm{GP}(\lambda,\theta)$ distribution reduces to the Poisson(λ) distribution with the property of equi-dispersion. When $\theta>0$ (or $\theta<0$), the $\mathrm{GP}(\lambda,\theta)$ distribution can be used to model count data with over-dispersion (or under-dispersion). When $\lambda=0$, the $\mathrm{GP}(\lambda,\theta)$ reduces to the degenerate distribution Degenerate(0).

Based on (1.1), some researchers developed so-called *zero-inflated generalized Poisson* (ZIGP) and *zero-adjusted gen-*

*Corresponding author.

eralized Poisson (ZAGP) models as alternatives to ZIP, ZIB and ZINB models for the analysis of count data with extra zeros. For example, Gupta *et al.* (1996) [5] studied general zero-adjusted count data models, proposed a ZAGP distribution and investigated the relative error incurred by ignoring the adjustment. They also provided real data sets where the ZAGP distribution fits very well. Gupta *et al.* (2004) [6] studied the ZIGP regression model and developed a score test to determine whether an adjustment for zero inflation is necessary. Famoye and Singh (2006) [4] developed a ZIGP regression model to model domestic violence data. Xie and Wei (2010) [21] extended the ZIP mixed regression model to the ZIGP mixed regression model and Xie *et al.* (2014) [22] provided a Markov chain Monte Carlo method for dealing with the complexity of the ZIGP model.

Excessive zeros in multivariate count data are often encountered in practice, e.g., when events involve different types of defects in a manufacturing process near its perfect state, the univariate zero-inflated count distributions are no longer appropriate. To model dependent structure in multivariate count data, some authors have extended the univariate ZIP distribution to multivariate ZIP distribution, for example, Liu and Tian (2015) [15] introduced the Type I multivariate ZIP distribution in order to model correlated multivariate count data with extra zeros. Since the Poisson distribution only possesses the property of equi-dispersion, the existing Type I multivariate ZIP distribution cannot be used to model multivariate zero-inflated count data with over-dispersion or under-dispersion. In this paper, we extend the univariate ZIGP distribution to Type I multivariate ZIGP distribution via stochastic representation aiming to model positively correlated multivariate zero-inflated count data with over-dispersion or under-dispersion. Its distributional theories and associated properties are derived. Due to the complexity of the ZIGP model, we will provide four useful algorithms (a very fast Fisher-scoring algorithm, an ECM algorithm, a simple EM algorithm and an explicit MM algorithm) for finding *maximum likelihood estimates* (MLEs) of parameters of interest and will develop efficient statistical inference methods for the proposed model.

The rest of the paper is organized as follows. In Section 2, we introduce the Type I multivariate ZIGP distribution, and study the distributional theories and corresponding properties. In Section 3, the likelihood-based statistical inferences about parameters of interest are provided. Simulation studies for investigating the accuracy of point estimates and confidence interval estimates and comparing the likelihood ratio test with the score test are conducted in Section 4. In Section 5, two real examples are used to illustrate the proposed methods and to compare with existing methods. A discussion is given in Section 6. Some detailed technical proofs are put in the Appendices.

2. TYPE I MULTIVARIATE ZERO-INFLATED GENERALIZED POISSON DISTRIBUTION

Let $Z \sim \text{Bernoulli}(1-\phi)$, $X \sim \text{GP}(\lambda, \theta)$ and $Z \perp\!\!\!\perp X$. The random variable $Y \sim \text{ZIGP}(\phi, \lambda, \theta)$ has the following *stochastic representation* (SR):

$$Y \overset{\text{d}}{=} ZX = \begin{cases} 0, & \text{with probability } \phi, \\ X, & \text{with probability } 1-\phi, \end{cases} \tag{2.1}$$

where the symbol "$\overset{\text{d}}{=}$" means that the random variables on both sides of the equality have the same distribution. When $\theta = 0$, the $\text{ZIGP}(\phi, \lambda, \theta)$ reduces to the zero-inflated Poisson distribution $\text{ZIP}(\phi, \lambda)$. Alternative to (2.1), we obtain the following mixture representation:

$$Z \sim \text{Bernoulli}(1-\phi) \quad \text{and} \quad Y|(Z=z) \sim \text{GP}(\lambda z, \theta).$$

From (2.1), we immediately obtain

$$\begin{cases} E(Y) & = & \dfrac{(1-\phi)\lambda}{(1-\theta)}, \\ E(Y^2) & = & \dfrac{(1-\phi)\lambda}{(1-\theta)^3} + \dfrac{(1-\phi)\lambda^2}{(1-\theta)^2}, \\ \text{Var}(Y) & = & \dfrac{(1-\phi)\lambda}{(1-\theta)^3} + \dfrac{\phi(1-\phi)\lambda^2}{(1-\theta)^2}, \end{cases}$$

where $\theta < 1$.

Motivated by the SR (2.1) of the univariate ZIGP distribution, we can extend it to the multivariate version by means of an SR in a vector form with a common Z to characterize the correlation structure among the components, as shown in the following definition.

Definition 1. Let $Z \sim \text{Bernoulli}\,(1-\phi)$, $\mathbf{x} = (X_1, \ldots, X_m)^\top$, $X_i \sim \text{GP}(\lambda_i, \theta_i)$ for $i = 1, \ldots, m$, and $(Z, X_1, \ldots, X_m)$ are mutually independent. An m-dimensional discrete random vector $\mathbf{y} = (Y_1, \ldots, Y_m)^\top$ is said to have a Type I multivariate ZIGP distribution if

$$\mathbf{y} \overset{\text{d}}{=} Z\,\mathbf{x} = \begin{cases} \mathbf{0}, & \text{with probability } \phi, \\ \mathbf{x}, & \text{with probability } 1-\phi, \end{cases} \tag{2.2}$$

where $\phi \in [0,1)$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^\top \in \mathbb{R}_+^m$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^\top$, $\max(-1, -\lambda_i/q_i) < \theta_i \leqslant 1$ and $q_i \geqslant 4$ is the largest positive integer for each $\lambda_i + \theta_i q_i > 0$ when $\theta_i < 0$. We write $\mathbf{y} \sim \text{ZIGP}_m^{(\text{I})}(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})$ or $\mathbf{y} \sim \text{ZIGP}^{(\text{I})}(\phi; \lambda_1, \ldots, \lambda_m, \theta_1, \ldots, \lambda_m)$ and call $\mathbf{x}$ the base vector of the $\mathbf{y}$. ¶

2.1 Joint pmf and joint cumulative distribution function

The joint pmf of $\mathbf{y} \sim \text{ZIGP}_m^{(\text{I})}(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})$ is denoted by $f(\boldsymbol{y}|\phi, \boldsymbol{\lambda}, \boldsymbol{\theta}) = \Pr(\mathbf{y} = \boldsymbol{y}) = \Pr(ZX_i = y_i,\ 1 \leqslant i \leqslant m)$. If $\boldsymbol{y} = \mathbf{0}_m$, we have

$$\begin{aligned} & f(\boldsymbol{y}|\phi,\boldsymbol{\lambda},\boldsymbol{\theta}) \\ = \; & \Pr(ZX_i = 0,\; 1 \leqslant i \leqslant m) \\ = \; & \Pr(Z=0) + \Pr(Z=1,\; X_i = 0,\; 1 \leqslant i \leqslant m) \\ = \; & \phi + (1-\phi)\mathrm{e}^{-\lambda_+}, \end{aligned} \tag{2.3}$$

where $\lambda_+ = \sum_{i=1}^m \lambda_i$. If $\boldsymbol{y} \neq \mathbf{0}_m$, we have

$$\begin{aligned} & f(\boldsymbol{y}|\phi,\boldsymbol{\lambda},\boldsymbol{\theta}) \\ = \; & \Pr(Z=1,\; X_i = y_i,\; 1 \leqslant i \leqslant m) \\ = \; & (1-\phi)\mathrm{e}^{-\lambda_+ - \sum_{i=1}^m \theta_i y_i} \prod_{i=1}^m \frac{\lambda_i(\lambda_i + \theta_i y_i)^{y_i - 1}}{y_i!} \\ \hat{=} \; & (1-\phi)a, \end{aligned} \tag{2.4}$$

By combining (2.3) with (2.4), we obtain

$$\begin{aligned} & f(\boldsymbol{y}|\phi,\boldsymbol{\lambda},\boldsymbol{\theta}) \\ = \; & [\phi + (1-\phi)\mathrm{e}^{-\lambda_+}]I(\boldsymbol{y} = \mathbf{0}) + (1-\phi)aI(\boldsymbol{y} \neq \mathbf{0}) \\ = \; & \phi \Pr(\boldsymbol{\xi} = \boldsymbol{y}) + (1-\phi)\Pr(\mathbf{x} = \boldsymbol{y}), \end{aligned} \tag{2.5}$$

where $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)^\top$ and $\{\xi_i\}_{i=1}^m \overset{\text{iid}}{\sim}$ Degenerate(0).

Let $\mathbf{y} \sim \mathrm{ZIGP}_m^{(\mathrm{I})}(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})$. For any non-negative real vector $\boldsymbol{y} = (y_1, \ldots, y_m)^\top$, the joint cumulative distribution function of $\mathbf{y}$ is given by

$$\begin{aligned} & \Pr(\mathbf{y} \leqslant \boldsymbol{y}) \\ = \; & \phi \Pr(\boldsymbol{\xi} = \mathbf{0}) + (1-\phi)\Pr(\mathbf{x} \leqslant \boldsymbol{y}) \\ = \; & \phi + (1-\phi)\prod_{i=1}^m \Pr(X_i \leqslant y_i) \\ = \; & \phi + (1-\phi)\prod_{i=1}^m \left[\sum_{k_i=0}^{y_i} \frac{\lambda_i(\lambda_i + \theta_i k_i)^{k_i-1}\mathrm{e}^{-\lambda_i - \theta_i k_i}}{k_i!} \right] \end{aligned}$$

for $y_1, \ldots, y_m \geqslant 0$.

2.2 Mixed moments and moment generating function

From (2.2), it is not difficult to obtain

$$\begin{cases} E(\mathbf{y}) = (1-\phi)\,\boldsymbol{\alpha}, \\ E(\mathbf{y}\mathbf{y}^\top) = (1-\phi)[\mathrm{diag}(\boldsymbol{\beta}) + \boldsymbol{\alpha}\boldsymbol{\alpha}^\top], \\ \mathrm{Var}(\mathbf{y}) = (1-\phi)[\mathrm{diag}(\boldsymbol{\beta}) + \phi\,\boldsymbol{\alpha}\boldsymbol{\alpha}^\top], \end{cases}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)^\top$, $\alpha_i = \lambda_i/(1-\theta_i)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^\top$, $\beta_i = \lambda_i/(1-\theta_i)^3$, $\theta_i < 1$ for $i = 1, \ldots, m$. Thus we have

$$\begin{aligned} & \mathrm{Corr}(Y_i, Y_j) \\ = \; & \sqrt{\frac{\lambda_i\lambda_j(1-\theta_i)(1-\theta_j)}{[(1-\theta_i)\lambda_i + 1/\phi][(1-\theta_j)\lambda_j + 1/\phi]}} \end{aligned}$$

for $i \neq j$. In particular, when $\lambda_i = \lambda_j = \lambda$ and $\theta_i = \theta_j = \theta$, we obtain

$$\mathrm{Corr}(Y_i, Y_j) = \frac{\phi\lambda(1-\theta)}{1 + \phi\lambda(1-\theta)}, \qquad i \neq j.$$

By using the formula of $E(\xi) = E[E(\xi|\eta)]$, we can obtain the moment generating function of $\mathbf{y} \sim \mathrm{ZIGP}^{(\mathrm{I})}(\phi; \lambda_1, \ldots, \lambda_m, \theta_1, \ldots, \theta_m)$, given by

$$\begin{aligned} M_{\mathbf{y}}(\boldsymbol{t}) &= E[\exp(\boldsymbol{t}^\top\mathbf{y})] = E[\exp(Z \cdot \boldsymbol{t}^\top\mathbf{x})] \\ &= E\Big\{E[\exp(Z\boldsymbol{t}^\top\mathbf{x})|Z]\Big\} = E\left[\prod_{i=1}^m M_{X_i}(t_i Z)\right] \\ &= \phi\prod_{i=0}^m M_{X_i}(0) + (1-\phi)\prod_{i=1}^m M_{X_i}(t_i), \end{aligned}$$

where

$$\begin{aligned} M_{X_i}(0) &= \exp\left\{-\frac{\lambda_i}{\theta_i}\Big[W(-\theta_i\mathrm{e}^{-\theta_i}) + \theta_i\Big]\right\} \quad \text{and} \\ M_{X_i}(t_i) &= \exp\left\{-\frac{\lambda_i}{\theta_i}\Big[W(-\theta_i\mathrm{e}^{-\theta_i + t_i}) + \theta_i\Big]\right\} \end{aligned}$$

for $i = 1, \ldots, m$; the Lambert $W(\cdot)$ function is defined by $W(x)\exp[W(x)] = x$, for more details about this function see Corless *et al.* (1996) [3].

2.3 Marginal distributions

Let $\mathbf{y} \sim \mathrm{ZIGP}^{(\mathrm{I})}(\phi; \lambda_1, \ldots, \lambda_m, \theta_1, \ldots, \theta_m)$. Partition $\mathbf{y}$ into two parts

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{pmatrix}, \text{ where } \mathbf{y}^{(1)} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_r \end{pmatrix}, \quad \mathbf{y}^{(2)} = \begin{pmatrix} Y_{r+1} \\ \vdots \\ Y_m \end{pmatrix}.$$

We can partition $\mathbf{x}$ in the same fashion. According to Definition 1, we obtain

$$\begin{cases} \mathbf{y}^{(1)} \overset{\mathrm{d}}{=} Z\mathbf{x}^{(1)} \sim \mathrm{ZIGP}^{(\mathrm{I})}(\phi; \lambda_1, \ldots, \lambda_r, \theta_1, \ldots, \theta_r) \quad \text{and} \\ \mathbf{y}^{(2)} \overset{\mathrm{d}}{=} Z\mathbf{x}^{(2)} \sim \mathrm{ZIGP}^{(\mathrm{I})}(\phi; \lambda_{r+1}, \ldots, \lambda_m, \theta_{r+1}, \ldots, \theta_m). \end{cases} \tag{2.6}$$

In fact, for any positive integers $i_1, \ldots, i_r$ satisfying $1 \leqslant i_1 < \cdots < i_r \leqslant m$, we have

$$\begin{aligned} \begin{pmatrix} Y_{i_1} \\ \vdots \\ Y_{i_r} \end{pmatrix} &\overset{\mathrm{d}}{=} Z\begin{pmatrix} X_{i_1} \\ \vdots \\ X_{i_r} \end{pmatrix} \\ &\sim \mathrm{ZIGP}^{(\mathrm{I})}(\phi; \lambda_{i_1}, \ldots, \lambda_{i_r}, \theta_{i_1}, \ldots, \theta_{i_r}). \end{aligned} \tag{2.7}$$

2.4 Conditional distributions

2.4.1 Conditional distribution of $\mathbf{y}^{(1)}|\mathbf{y}^{(2)}$

From (2.5) and (2.6), the conditional distribution of $\mathbf{y}^{(1)}|\mathbf{y}^{(2)}$ is given by

$$\Pr(\mathbf{y}^{(1)} = \boldsymbol{y}^{(1)}|\mathbf{y}^{(2)} = \boldsymbol{y}^{(2)}) = \frac{f(\boldsymbol{y}|\phi,\boldsymbol{\lambda},\boldsymbol{\theta})}{\Pr(\mathbf{y}^{(2)} = \boldsymbol{y}^{(2)})} \tag{2.8}$$

$$= \frac{[\phi + (1-\phi)\mathrm{e}^{-\lambda_+}]I(\boldsymbol{y}=\mathbf{0}) + (1-\phi)aI(\boldsymbol{y}\neq\mathbf{0})}{\left[\phi+(1-\phi)\mathrm{e}^{-\lambda_+^{(2)}}\right]I(\boldsymbol{y}^{(2)}=\mathbf{0})+(1-\phi)a_2 I(\boldsymbol{y}^{(2)}\neq\mathbf{0})},$$

where a is defined by (2.4), $\lambda_+^{(2)} = \sum_{i=r+1}^m \lambda_i = \lambda_+ - \lambda_+^{(1)}$, and

$$a_2 = \mathrm{e}^{-\lambda_+^{(2)} - \sum_{i=r+1}^m \theta_i y_i} \prod_{i=r+1}^m \frac{\lambda_i(\lambda_i+\theta_i y_i)^{y_i-1}}{y_i!}. \tag{2.9}$$

We consider two cases. Case I: $\boldsymbol{y}^{(2)} \neq \mathbf{0}$. Under Case I, it is obvious that $\boldsymbol{y} \neq \mathbf{0}$. From (2.8) and (2.9), it is easy to obtain

$$\begin{aligned} &\Pr(\mathbf{y}^{(1)} = \boldsymbol{y}^{(1)} | \mathbf{y}^{(2)} = \boldsymbol{y}^{(2)}) \\ = &\prod_{i=1}^r \frac{\lambda_i(\lambda_i+\theta_i y_i)^{y_i-1}\mathrm{e}^{-\lambda_i-\theta_i y_i}}{y_i!} = \frac{a}{a_2} \mathrel{\hat=} a_1. \end{aligned} \tag{2.10}$$

This implies $\mathbf{y}^{(1)}|(\mathbf{y}^{(2)} = \boldsymbol{y}^{(2)} \neq \mathbf{0}) \stackrel{\mathrm{d}}{=} \mathbf{x}^{(1)}$, not depending on Z. In other words, given $\boldsymbol{y}^{(2)} \neq \mathbf{0}$, $(Y_1, \ldots, Y_r)$ are mutually independent and $Y_i|(\mathbf{y}^{(2)} = \boldsymbol{y}^{(2)} \neq \mathbf{0}) \stackrel{\mathrm{d}}{=} X_i \sim \mathrm{GP}(\lambda_i, \theta_i)$, being free from ϕ.

Case II: $\boldsymbol{y}^{(2)} = \mathbf{0}$. Under Case II, it is possible that $\boldsymbol{y}^{(1)} = \mathbf{0}$ or $\boldsymbol{y}^{(1)} \neq \mathbf{0}$. When $\boldsymbol{y}^{(1)} = \mathbf{0}$, from (2.8), we obtain

$$\begin{aligned} \Pr(\mathbf{y}^{(1)} = \mathbf{0}|\mathbf{y}^{(2)} = \mathbf{0}) &= \frac{\phi + (1-\phi)\mathrm{e}^{-\lambda_+}}{\phi + (1-\phi)\mathrm{e}^{-\lambda_+^{(2)}}} \\ &= \phi^* + (1-\phi^*)\mathrm{e}^{-\lambda_+^{(1)}}, \end{aligned} \tag{2.11}$$

where $\phi^* \mathrel{\hat=} \phi\,\mathrm{e}^{\lambda_+^{(2)}}/(\phi\,\mathrm{e}^{\lambda_+^{(2)}} + 1 - \phi)$. When $\boldsymbol{y}^{(1)} \neq \mathbf{0}$, from (2.8), we have

$$\begin{aligned} &\Pr(\mathbf{y}^{(1)} = \boldsymbol{y}^{(1)}|\mathbf{y}^{(2)} = \mathbf{0}) \\ = &\frac{(1-\phi)\mathrm{e}^{-\lambda_+ - \sum_{i=1}^r \theta_i y_i} \prod_{i=1}^r \lambda_i(\lambda_i+\theta_i y_i)^{y_i-1}/y_i!}{\phi + (1-\phi)\mathrm{e}^{-\lambda_+^{(2)}}} \\ = &\frac{(1-\phi)\mathrm{e}^{-\lambda_+^{(1)} - \sum_{i=1}^r \theta_i y_i} \prod_{i=1}^r \lambda_i(\lambda_i+\theta_i y_i)^{y_i-1}/y_i!}{\phi\,\mathrm{e}^{\lambda_+^{(2)}} + (1-\phi)} \\ = &(1-\phi^*)\mathrm{e}^{-\lambda_+^{(1)} - \sum_{i=1}^r \theta_i y_i} \prod_{i=1}^r \frac{\lambda_i(\lambda_i+\theta_i y_i)^{y_i-1}}{y_i!} \\ \stackrel{(2.10)}{=} &(1-\phi^*)a_1. \end{aligned} \tag{2.12}$$

By combining (2.11) with (2.12), we obtain

$$\begin{aligned} &\Pr(\mathbf{y}^{(1)} = \boldsymbol{y}^{(1)}|\mathbf{y}^{(2)} = \mathbf{0}) \\ = &\left[\phi^* + (1-\phi^*)\mathrm{e}^{-\lambda_+^{(1)}}\right] I(\mathbf{y}^{(1)} = \mathbf{0}) \\ &+ (1-\phi^*)a_1 I(\mathbf{y}^{(1)} \neq \mathbf{0}), \end{aligned}$$

i.e., $\mathbf{y}^{(1)}|(\mathbf{y}^{(2)} = \mathbf{0}) \sim \mathrm{ZIGP}^{(\mathrm{I})}(\phi^*; \lambda_1, \ldots, \lambda_r, \theta_1, \ldots, \theta_r)$.

2.4.2 Conditional distribution of $Z|\mathbf{y}$

Since $Z \sim \mathrm{Bernoulli}(1-\phi)$, Z only takes the value 0 or 1. Note that

$$\begin{aligned} \Pr(Z = 1|\mathbf{y} = \boldsymbol{y}) &= \frac{\Pr(Z=1, \mathbf{x} = \boldsymbol{y})}{f(\boldsymbol{y}|\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})} \\ &\stackrel{(2.4)}{=} \frac{(1-\phi)\,a}{f(\boldsymbol{y}|\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})} \\ &\stackrel{(2.5)}{=} \begin{cases} \dfrac{(1-\phi)\mathrm{e}^{-\lambda_+}}{\phi + (1-\phi)\mathrm{e}^{-\lambda_+}}, & \text{if } \boldsymbol{y} = \mathbf{0}, \\ 1, & \text{if } \boldsymbol{y} \neq \mathbf{0}. \end{cases} \end{aligned}$$

Therefore,

$$Z|(\mathbf{y} = \boldsymbol{y}) \sim \begin{cases} \mathrm{Bernoulli}(\psi), & \text{if } \boldsymbol{y} = \mathbf{0}, \\ \mathrm{Degenerate}(1), & \text{if } \boldsymbol{y} \neq \mathbf{0}, \end{cases} \tag{2.13}$$

where

$$\psi \mathrel{\hat=} \frac{(1-\phi)\mathrm{e}^{-\lambda_+}}{\phi + (1-\phi)\mathrm{e}^{-\lambda_+}}. \tag{2.14}$$

2.4.3 Conditional distribution of $\mathbf{x}|\mathbf{y}$

If $\boldsymbol{y} = \mathbf{0}$, we have

$$\begin{aligned} &\Pr(\mathbf{x} = \boldsymbol{x}|\mathbf{y} = \mathbf{0}) = \frac{\Pr(\mathbf{x} = \boldsymbol{x}, \mathbf{y} = \mathbf{0})}{\Pr(\mathbf{y} = \mathbf{0})} \\ = &\frac{\Pr(\mathbf{x} = \mathbf{0}, \mathbf{y} = \mathbf{0})}{f(\mathbf{0}|\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})} I(\boldsymbol{x} = \mathbf{0}) \\ &+ \frac{\Pr(\mathbf{x} = \boldsymbol{x}, Z = 0)}{f(\mathbf{0}|\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})} I(\boldsymbol{x} \neq \mathbf{0}) \\ = &\frac{\Pr(\mathbf{x} = \mathbf{0})}{f(\mathbf{0}|\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})} I(\boldsymbol{x} = \mathbf{0}) + \frac{\phi\Pr(\mathbf{x} = \boldsymbol{x})}{f(\mathbf{0}|\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})} I(\boldsymbol{x} \neq \mathbf{0}) \\ \stackrel{(2.5)}{=} &\frac{\mathrm{e}^{-\lambda_+}}{\phi + (1-\phi)\mathrm{e}^{-\lambda_+}} I(\boldsymbol{x} = \mathbf{0}) \\ &+ \frac{\phi\prod_{i=1}^m \lambda_i(\lambda_i+\theta_i x_i)^{x_i-1}\mathrm{e}^{-\lambda_i-\theta_i x_i}/x_i! \cdot I(\boldsymbol{x} \neq \mathbf{0})}{\phi + (1-\phi)\mathrm{e}^{-\lambda_+}} \\ \stackrel{(2.14)}{=} &[\psi + (1-\psi)\mathrm{e}^{-\lambda_+}]I(\boldsymbol{x} = \mathbf{0}) \\ &+ \left[(1-\psi)\prod_{i=1}^m \frac{\lambda_i(\lambda_i+\theta_i x_i)^{x_i-1}\mathrm{e}^{-\lambda_i-\theta_i x_i}}{x_i!}\right] I(\boldsymbol{x} \neq \mathbf{0}), \end{aligned}$$

i.e.,

$$\mathbf{x}|(\mathbf{y} = \mathbf{0}) \sim \mathrm{ZIGP}^{(\mathrm{I})}(\psi; \lambda_1, \ldots, \lambda_m, \theta_1, \ldots, \theta_m). \tag{2.15}$$

If $\boldsymbol{y} \neq \mathbf{0}$, we have

$$\begin{aligned} \Pr(\mathbf{x} = \boldsymbol{x}|\mathbf{y} = \boldsymbol{y}) &= \frac{\Pr(\mathbf{x} = \boldsymbol{x}, \mathbf{y} = \boldsymbol{y})}{\Pr(\mathbf{y} = \boldsymbol{y})} \\ &= \frac{\Pr(\mathbf{x} = \boldsymbol{y}, Z = 1)}{f(\boldsymbol{y}|\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})} = 1. \end{aligned}$$

Thus, given $\mathbf{y} = \boldsymbol{y} \neq \mathbf{0}$, $(X_1, \ldots, X_m)$ are independent and

(2.16) $\quad X_i|(\mathbf{y} = \boldsymbol{y} \neq \mathbf{0}) \sim \text{Degenerate}(y_i), \quad i = 1, \ldots, m.$

2.4.4 Conditional distribution of $X_i|(Y_i = y_i = 0), i = 1, \ldots, m$

From (2.7), we have $Y_i \sim \text{ZIGP}(\phi, \lambda_i, \theta_i)$. Thus,

$$
\begin{aligned}
\Pr(X_i = x_i|Y_i = 0) &= \frac{\Pr(X_i = x_i, Y_i = 0)}{\Pr(Y_i = 0)} \\
&= \frac{\Pr(X_i = 0, Y_i = 0)}{f(0|\phi, \lambda_i, \theta_i)} I(x_i = 0) \\
&\quad + \frac{\Pr(X_i = x_i, Z = 0)}{f(0|\phi, \lambda_i, \theta_i)} I(x_i > 0) \\
&= \frac{\Pr(X_i = 0)}{f(0|\phi, \lambda_i, \theta_i)} I(x_i = 0) + \frac{\phi \Pr(X_i = x_i)}{f(0|\phi, \lambda_i, \theta_i)} I(x_i > 0) \\
&= \frac{\mathrm{e}^{-\lambda_i}}{\phi + (1-\phi)\mathrm{e}^{-\lambda_i}} I(x_i = 0) \\
&\quad + \frac{\phi}{\phi + (1-\phi)\mathrm{e}^{-\lambda_i}} \frac{\lambda_i(\lambda_i + \theta_i x_i)^{x_i - 1}\mathrm{e}^{-\lambda_i - \theta_i x_i}}{x_i!} I(x_i > 0) \\
&= [\phi_i^* + (1-\phi_i^*)\mathrm{e}^{-\lambda_i}] I(x_i = 0) \\
&\quad + (1-\phi_i^*) \frac{\lambda_i(\lambda_i + \theta_i x_i)^{x_i - 1}\mathrm{e}^{-\lambda_i - \theta_i x_i}}{x_i!} I(x_i > 0),
\end{aligned}
$$

i.e., $X_i|(Y_i = 0) \sim \text{ZIGP}^{(\text{I})}(\phi_i^*, \lambda_i, \theta_i)$, where

$$\phi_i^* = \frac{(1-\phi)\mathrm{e}^{-\lambda_i}}{\phi + (1-\phi)\mathrm{e}^{-\lambda_i}}.$$

2.4.5 Conditional distribution of $X_i|(Y_i = y_i > 0),\ i = 1, \ldots, m$

Since

$$
\begin{aligned}
\Pr(X_i = x_i|Y_i = y_i) &= \frac{\Pr(X_i = x_i, Y_i = y_i)}{\Pr(Y_i = y_i)} \\
&= \frac{\Pr\{X_i = y_i, Z = 1\}}{f(y_i|\phi, \lambda_i, \theta)} = 1,
\end{aligned}
$$

we obtain $X_i|(Y_i = y_i > 0) \sim \text{Degenerate}(y_i)$.

3. LIKELIHOOD-BASED STATISTICAL INFERENCES

Suppose that $\mathbf{y}_1, \ldots, \mathbf{y}_n$ is a random sample of size n from the Type I m-dimensional ZIGP distribution $\text{ZIGP}^{(\text{I})}(\phi; \lambda_1, \ldots, \lambda_m, \theta_1, \ldots, \theta_m)$, where $\mathbf{y}_j = (Y_{1j}, \ldots, Y_{mj})^\top$ for $j = 1, \ldots, n$. Let $\boldsymbol{y}_j = (y_{1j}, \ldots, y_{mj})^\top$ denote the realization of the random vector $\mathbf{y}_j$, and $Y_{\text{obs}} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$ be the observed data. Furthermore, Let $\mathbb{J}_0 = \{j|\boldsymbol{y}_j = \mathbf{0}, j = 1, \ldots, n\}$ and $n_0 = \sum_{j=1}^n I(\boldsymbol{y}_j = \mathbf{0})$ denote the number of elements in $\mathbb{J}_0$. Then, the observed-data likelihood function

$$
\begin{aligned}
& L(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta}|Y_{\text{obs}}) \\
& \propto [\phi + (1-\phi)\mathrm{e}^{-\lambda_+}]^{n_0} (1-\phi)^{n-n_0} \mathrm{e}^{-(n-n_0)\lambda_+} \\
& \quad \times \prod_{j=1}^n \prod_{i=1}^m \lambda_i(\lambda_i + \theta_i y_{ij})^{y_{ij}-1} \mathrm{e}^{-\theta_i y_{ij}},
\end{aligned} \tag{3.1}
$$

so that the log-likelihood function is

$$
\begin{aligned}
\ell &= \ell(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta}|Y_{\text{obs}}) \\
&= n_0 \log[\phi + (1-\phi)\mathrm{e}^{-\lambda_+}] \\
&\quad + (n - n_0)\log(1-\phi) - (n-n_0)\lambda_+ \\
&\quad + \sum_{j=1}^n \sum_{i=1}^m [\log \lambda_i + (y_{ij} - 1)\log(\lambda_i + \theta_i y_{ij}) - \theta_i y_{ij}].
\end{aligned}
$$

3.1 MLEs via the Fisher scoring algorithm

In this subsection, the Fisher scoring algorithm is employed to calculate the MLEs of ϕ, $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$. The score vector $\nabla\ell$ and the Hessian matrix $\nabla^2\ell$ are given by

$$
\nabla\ell = \left(\frac{\partial\ell}{\partial\phi}, \frac{\partial\ell}{\partial\boldsymbol{\lambda}^\top}, \frac{\partial\ell}{\partial\boldsymbol{\theta}^\top}\right)^\top \quad \text{and}
$$

$$
\nabla^2\ell = \begin{pmatrix}
\frac{\partial^2\ell}{\partial\phi^2} & \frac{\partial^2\ell}{\partial\phi\partial\boldsymbol{\lambda}^\top} & \frac{\partial^2\ell}{\partial\phi\partial\boldsymbol{\theta}^\top} \\
\frac{\partial^2\ell}{\partial\boldsymbol{\lambda}\partial\phi} & \frac{\partial^2\ell}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^\top} & \frac{\partial^2\ell}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\theta}^\top} \\
\frac{\partial^2\ell}{\partial\boldsymbol{\theta}\partial\phi} & \frac{\partial^2\ell}{\partial\boldsymbol{\theta}\partial\boldsymbol{\lambda}^\top} & \frac{\partial^2\ell}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}
\end{pmatrix},
$$

respectively, where

$$
\begin{aligned}
\frac{\partial\ell}{\partial\phi} &= \frac{n_0(1-\mathrm{e}^{-\lambda_+})}{\phi + (1-\phi)\mathrm{e}^{-\lambda_+}} - \frac{n-n_0}{1-\phi}, \\
\frac{\partial\ell}{\partial\lambda_i} &= -\frac{n_0(1-\phi)\mathrm{e}^{-\lambda_+}}{\phi + (1-\phi)\mathrm{e}^{-\lambda_+}} - (n-n_0) \\
&\quad + \sum_{j=1}^n \left(\frac{1}{\lambda_i} + \frac{y_{ij}-1}{\lambda_i + \theta_i y_{ij}}\right), \\
\frac{\partial\ell}{\partial\theta_i} &= \sum_{j=1}^n \left[\frac{y_{ij}(y_{ij}-1)}{\lambda_i + \theta_i y_{ij}} - y_{ij}\right], \\
\frac{\partial^2\ell}{\partial\phi^2} &= -\frac{n_0(1-\mathrm{e}^{-\lambda_+})^2}{[\phi + (1-\phi)\mathrm{e}^{-\lambda_+}]^2} - \frac{n-n_0}{(1-\phi)^2}, \\
\frac{\partial^2\ell}{\partial\lambda_i^2} &= \frac{n_0\phi(1-\phi)\mathrm{e}^{-\lambda_+}}{[\phi + (1-\phi)\mathrm{e}^{-\lambda_+}]^2} \\
&\quad - \sum_{j=1}^n \left[\frac{1}{\lambda_i^2} + \frac{y_{ij}-1}{(\lambda_i + \theta_i y_{ij})^2}\right], \\
\frac{\partial\ell}{\partial\theta_i^2} &= -\sum_{j=1}^n \frac{y_{ij}^2(y_{ij}-1)}{(\lambda_i + \theta_i y_{ij})^2}, \\
\frac{\partial^2\ell}{\partial\lambda_i\partial\phi} &= \frac{n_0\mathrm{e}^{-\lambda_+}}{[\phi + (1-\phi)\mathrm{e}^{-\lambda_+}]^2},
\end{aligned}
$$

$$\frac{\partial^2 \ell}{\partial \lambda_i \partial \lambda_k} = \frac{n_0 \phi(1-\phi)\mathrm{e}^{-\lambda_+}}{[\phi + (1-\phi)\mathrm{e}^{-\lambda_+}]^2},$$
$$\frac{\partial^2 \ell}{\partial \lambda_i \partial \theta_i} = -\sum_{j=1}^{n} \frac{y_{ij}(y_{ij}-1)}{(\lambda_i + \theta_i y_{ij})^2},$$
$$\frac{\partial^2 \ell}{\partial \lambda_i \partial \theta_k} = \frac{\partial^2 \ell}{\partial \theta_i \partial \theta_k} = \frac{\partial^2 \ell}{\partial \theta_i \partial \phi} = 0,$$

for $i, k = 1, \ldots, m$ and $i \neq k$. By replacing n_0,

$$\left\{ \frac{y_{ij} - 1}{(\lambda_i + \theta_i y_{ij})^2} \right\}_{i=1}^{m}, \left\{ \frac{y_{ij}^3 - y_{ij}^2}{(\lambda_i + \theta_i y_{ij})^2} \right\}_{i=1}^{m}, \left\{ \frac{y_{ij}^2 - y_{ij}}{(\lambda_i + \theta_i y_{ij})^2} \right\}_{i=1}^{m}$$

in the above second partial derivatives with their expectations (see Appendix A)

(3.2)
$$\begin{cases} E\left[\sum_{j=1}^{n} I(\mathbf{y}_j = \mathbf{0})\right] = n[\phi + (1-\phi)\mathrm{e}^{-\lambda_+}], \\ E\left[\frac{Y_{ij} - 1}{(\lambda_i + \theta_i Y_{ij})^2}\right] = \frac{1-\phi}{\lambda_i} - \frac{1}{\lambda_i^2} - \frac{\theta_i(1-\phi)}{\lambda_i + 2\theta_i}, \\ E\left[\frac{Y_{ij}^3 - Y_{ij}^2}{(\lambda_i + \theta_i Y_{ij})^2}\right] = \frac{\lambda_i(1-\phi)}{1-\theta_i} + \frac{2\lambda_i(1-\phi)}{\lambda_i + 2\theta_i}, \\ E\left[\frac{Y_{ij}^2 - Y_{ij}}{(\lambda_i + \theta_i Y_{ij})^2}\right] = \frac{\lambda_i(1-\phi)}{\lambda_i + 2\theta_i}, \end{cases}$$

we can calculate the Fisher information matrix

$$\mathbf{J}(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta}) = E[-\nabla^2 \ell(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta} | Y_{\text{obs}})].$$

Let $(\phi^{(0)}, \boldsymbol{\lambda}^{(0)}, \boldsymbol{\theta}^{(0)})$ be the initial values of the MLEs $(\hat{\phi}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}})$. If $(\phi^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)})$ denote the t-th approximations of $(\hat{\phi}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}})$, then their $(t+1)$-th approximations can be obtained by the following Fisher scoring algorithm:

$$\begin{pmatrix} \phi^{(t+1)} \\ \boldsymbol{\lambda}^{(t+1)} \\ \boldsymbol{\theta}^{(t+1)} \end{pmatrix} = \begin{pmatrix} \phi^{(t)} \\ \boldsymbol{\lambda}^{(t)} \\ \boldsymbol{\theta}^{(t)} \end{pmatrix}$$
$$+ \mathbf{J}^{-1}(\phi^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)}) \nabla \ell(\phi^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)} | Y_{\text{obs}}). \tag{3.3}$$

The standard errors of the MLEs $(\hat{\phi}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}})$ are the square roots of the diagonal elements J^{kk} of the inverse Fisher information matrix $\mathbf{J}^{-1}(\hat{\phi}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}})$. Thus the $(1-\alpha)100\%$ asymptotic Wald *confidence intervals* (CIs) of ϕ, $\{\lambda_i\}_{i=1}^m$ and $\{\theta_i\}_{i=1}^m$ are given by

(3.4)
$$\left[\hat{\phi} - z_{\alpha/2}\sqrt{J^{11}},\ \ \hat{\phi} + z_{\alpha/2}\sqrt{J^{11}}\right],$$
$$\left[\hat{\lambda}_i - z_{\alpha/2}\sqrt{J^{i+1,i+1}},\ \ \hat{\lambda}_i + z_{\alpha/2}\sqrt{J^{i+1,i+1}}\right],$$
$$\left[\hat{\theta}_i - z_{\alpha/2}\sqrt{J^{i+m+1,i+m+1}},\ \ \hat{\theta}_i + z_{\alpha/2}\sqrt{J^{i+m+1,i+m+1}}\right],$$

for $i = 1, \ldots, m$, respectively, where z_α denotes the α-th upper quantile of the standard normal distribution.

3.2 MLEs via two EM-type algorithms

Although we derive the Fisher scoring algorithm to estimates the parameters in the Type I multivariate ZIGP model, it is sensitive to the choice of initial values. In other words, the Fisher scoring algorithm may be divergent if a poor initial value is chosen. Thus in this subsection we develop two EM-type algorithms: the first one is an ECM algorithm and the second one is an EM algorithm.

3.2.1 An ECM algorithm based on SR

For each $\boldsymbol{y}_j = (y_{1j}, \ldots, y_{mj})^\top$ with $j \in \{1, \ldots, n\}$, based on the SR (2.2) we introduce independent latent variables $Z_j \sim \text{Bernoulli}(1-\phi)$, $X_{ij} \sim \text{GP}(\lambda_i, \theta_i)$ for $i = 1, \ldots, m$. We denote the latent/missing data by $Y_{\text{mis}} = \{z_j, \boldsymbol{x}_j\}_{j=1}^n$ such that $\boldsymbol{y}_j = z_j \boldsymbol{x}_j$ and the complete data are $Y_{\text{com}} = \{Y_{\text{obs}}, Y_{\text{mis}}\} = Y_{\text{mis}}$, where $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{mj})^\top$, z_j and x_{ij} denote the realizations of Z_j and X_{ij}, respectively. The complete-data likelihood function is

$$L_1(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta} | Y_{\text{com}}) = \prod_{j=1}^{n} \left[(1-\phi)^{z_j} \phi^{(1-z_j)} \prod_{i=1}^{m} \frac{\lambda_i(\lambda_i + \theta_i x_{ij})^{x_{ij}-1} \mathrm{e}^{-\lambda_i - \theta_i x_{ij}}}{x_{ij}!} \right],$$

so that the complete-data log-likelihood function is

$$\begin{aligned} &\ell_1(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta} | Y_{\text{com}}) \\ = \ & c_1 + \sum_{j=1}^{n} \Big[z_j \log(1-\phi) + (1-z_j)\log(\phi) \Big] \\ &+ \sum_{j=1}^{n} \sum_{i=1}^{m} \Big[\log \lambda_i + (x_{ij}-1)\log(\lambda_i + \theta_i x_{ij}) - \lambda_i - \theta_i x_{ij} \Big]. \end{aligned}$$

Then, the complete-data MLEs of ϕ and $\{\lambda_i\}_{i=1}^m$ are given by

$$\phi = \frac{n - \sum_{j=1}^n z_j}{n} \quad \text{and} \quad \lambda_i = \frac{\sum_{j=1}^n x_{ij}}{n}(1-\theta_i), \tag{3.5}$$

for $i = 1, \ldots, m$, while the complete-data MLE of θ_i is the root of the equation

$$H_i(\theta_i | \lambda_i) = \sum_{j=1}^{n} \frac{x_{ij}^2 - x_{ij}}{\lambda_i + \theta_i x_{ij}} - \sum_{j=1}^{n} x_{ij} = 0, \tag{3.6}$$

for $i = 1, \ldots, m$. The E-step is to replace $\{z_j\}_{j=1}^n$, $\{x_{ij}\}_{j=1}^n$ and $\left\{ \frac{x_{ij}^2 - x_{ij}}{\lambda_i + \theta_i x_{ij}} \right\}_{j=1}^n$ in (3.5)–(3.6) by their conditional expectations:

$$\begin{aligned} &E(Z_j | Y_{\text{obs}}, \phi, \boldsymbol{\lambda}, \boldsymbol{\theta}) \\ \overset{(2.13)}{=} \ &\psi I(\boldsymbol{y}_j = \mathbf{0}) + I(\boldsymbol{y}_j \neq \mathbf{0}), \end{aligned} \tag{3.7}$$

$$(3.8)\qquad E(X_{ij}|Y_{\rm obs},\phi,\boldsymbol{\lambda},\boldsymbol{\theta}) \overset{(2.15)\ \&\ (2.16)}{=} \frac{(1-\psi)\lambda_i}{1-\theta_i} I(\boldsymbol{y}_j=\mathbf{0}) + y_{ij} I(\boldsymbol{y}_j \neq \mathbf{0}),$$

$$(3.9)\qquad E\left(\frac{X_{ij}^2-X_{ij}}{\lambda_i+\theta_i X_{ij}}\middle| Y_{\rm obs},\phi,\boldsymbol{\lambda},\boldsymbol{\theta}\right) = \frac{(1-\psi)\lambda_i}{1-\theta_i} I(\boldsymbol{y}_j=\mathbf{0}) + \frac{y_{ij}^2-y_{ij}}{\lambda_i+\theta_i y_{ij}} I(\boldsymbol{y}_j\neq\mathbf{0}),$$

where ψ is defined by (2.14) and the proof of (3.9) is given in Appendix B.

By combining (3.5)–(3.9), we have the following ECM iterations: Let $t=0$ and given $\phi^{(t)}$ and $\{\lambda_i^{(t)}\}_{i=1}^m$,

M-Step 1: From (3.7) and the first formula of (3.5), we calculate

$$(3.10)\qquad \phi^{(t+1)} = \frac{n_0(1-\psi^{(t)})}{n},$$

where

$$\psi^{(t)} = \frac{(1-\phi^{(t)})\,\mathrm{e}^{-\lambda_+^{(t)}}}{\phi^{(t)} + (1-\phi^{(t)})\,\mathrm{e}^{-\lambda_+^{(t)}}},$$

$\lambda_+^{(t)} = \sum_{i=1}^m \lambda_i^{(t)}$ and $n_0 = \sum_{j=1}^n I(\boldsymbol{y}_j=\mathbf{0})$.

CM-Step 2: From (3.9), (3.8) and (3.6), we calculate $\theta_i^{(t+1)}$, which is the root of the equation

$$(3.11)\qquad H_i(\theta_i) = \sum_{j=1}^n \frac{y_{ij}^2-y_{ij}}{\lambda_i^{(t)}+\theta_i y_{ij}} - \sum_{j=1}^n y_{ij} = 0,$$

for $i=1,\ldots,m$.

CM-Step 3: From (3.8) and the second formula of (3.5), we calculate

$$(3.12)\qquad \lambda_i^{(t+1)} = \lambda_i^{(t)}\phi^{(t+1)} + (1-\theta_i^{(t+1)})\bar{y}_{i\cdot},$$

where

$$(3.13)\qquad \bar{y}_{i\cdot} = \frac{1}{n}\sum_{j=1}^n y_{ij}, \quad i=1,\ldots,m.$$

3.2.2 A simple EM algorithm by introducing only one latent variable

In the previous subsection, we proposed an ECM algorithm by introducing $n(1+m)$ latent variables. It is well known that an ECM algorithm generally converges much slower than the corresponding EM algorithm. Thus in this subsection, we will provide a simple EM algorithm by introducing only one latent variable.

Note that the observed zero vectors from a Type I multivariate ZIGP distribution can be classified into two categories: One is called the *extra zero vectors* resulted from degenerate distribution at point zero because of population variability; while the other is called the *structural zero vectors* came from the independent ordinary GP distributions. Thus, we can partition

$$\mathbb{J}_0 = \{j|\boldsymbol{y}_j=\mathbf{0},\ j=1,\ldots,n\}$$

as the union of $\mathbb{J}_{\rm extra}$ and $\mathbb{J}_{\rm structural}$. The major obstacle for obtaining explicit solutions of MLEs of parameters from (3.1) is the first term of (3.1). To overcome this difficulty, we augment $Y_{\rm obs}$ with a latent variable W that denotes the number of $\mathbb{J}_{\rm extra}$ to split n_0 into W and n_0-W. The resultant conditional predictive distribution of W given $Y_{\rm obs}$ and $(\phi,\boldsymbol{\lambda},\boldsymbol{\theta})$ is

$$W|(Y_{\rm obs},\phi,\boldsymbol{\lambda},\boldsymbol{\theta}) \sim \text{Binomial}\left(n_0,\ \frac{\phi}{\phi+(1-\phi)\mathrm{e}^{-\lambda_+}}\right) = \text{Binomial}(n_0, 1-\psi),$$

where ψ is defined by (2.14). The complete-data likelihood

$$\begin{aligned} & L_2(\phi,\boldsymbol{\lambda},\boldsymbol{\theta}|Y_{\rm com}) \\ \propto\ & \phi^w[(1-\phi)\mathrm{e}^{-\lambda_+}]^{n_0-w}(1-\phi)^{n-n_0}\mathrm{e}^{-(n-n_0)\lambda_+} \\ & \times \prod_{j=1}^n\prod_{i=1}^m \lambda_i(\lambda_i+\theta_i y_{ij})^{y_{ij}-1}\mathrm{e}^{-\theta_i y_{ij}} \\ =\ & \phi^w(1-\phi)^{n-w}\mathrm{e}^{-(n-w)\lambda_+} \\ & \times \prod_{j=1}^n\prod_{i=1}^m \lambda_i(\lambda_i+\theta_i y_{ij})^{y_{ij}-1}\mathrm{e}^{-\theta_i y_{ij}}, \end{aligned}$$

so that the complete-data log-likelihood function is

$$\begin{aligned} & \ell_2(\phi,\boldsymbol{\lambda},\boldsymbol{\theta}|Y_{\rm com}) \\ =\ & w\log\phi + (n-w)\log(1-\phi) - (n-w)\lambda_+ \\ & + \sum_{j=1}^n\sum_{i=1}^m \Big[\log\lambda_i + (y_{ij}-1)\log(\lambda_i+\theta_i y_{ij}) - \theta_i y_{ij}\Big]. \end{aligned}$$

Hence, the complete-data MLEs of ϕ and $\{\lambda_i\}_{i=1}^m$ are given by

$$(3.14)\qquad \phi = \frac{w}{n}, \qquad \lambda_i = \frac{n\bar{y}_{i\cdot}(1-\theta_i)}{n-w}, \qquad i=1,\ldots,m,$$

where $\bar{y}_{i\cdot}$ is defined by (3.13), and complete-data MLE of θ_i is the root of the equation

$$(3.15)\qquad H_i(u) = \sum_{j=1}^n \frac{y_{ij}(y_{ij}-1)(n-w)}{n\bar{y}_{i\cdot}(1-u)+uy_{ij}(n-w)} - n\bar{y}_{i\cdot} = 0,$$

for $i=1,\ldots,m$. Thus, the E-step is to replace w in the above expressions by its conditional expectation

$$(3.16)\qquad E(W|Y_{\rm obs},\phi,\boldsymbol{\lambda},\theta) = \frac{n_0\phi}{\phi+(1-\phi)\mathrm{e}^{-\lambda_+}} = n_0(1-\psi).$$

3.3 MLEs via the MM algorithm

Although the proposed two EM-type algorithms provided relatively simple iterations to find the MLEs of parameters in the Type I multivariate ZIGP distribution, we have to solve the root of m one-dimensional nonlinear equations specified by (3.11) or (3.15) at each step by employing the Newton's method, whose convergence depends on the choice of initial values. The situation becomes much complicated when such EM-type algorithms are utilized to calculate the confidence intervals of parameters via bootstrap methods as shown in the next subsection. In other words, we do not know how to specify so many initial values in these Newton's methods such that they can converge. In this subsection, we will develop a novel MM algorithm with explicit expressions at each iteration through constructing a Q function to separate the parameters ϕ, $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$.

For convenience, we first define

$$\begin{aligned}
\mathbb{J}_0 &= \{j|\boldsymbol{y}_j = \mathbf{0},\ j = 1, \ldots, n\}, \\
n_0 &= \sum_{j=1}^{n} I(\boldsymbol{y}_j = \mathbf{0}) = \#\{\mathbb{J}_0\}, \\
\mathbb{J} &= \{j|\boldsymbol{y}_j \neq \mathbf{0},\ j = 1, \ldots, n\}, \\
\mathbb{J}_i &= \{j|y_{ij} \neq 0,\ j = 1, \ldots, n\}, \qquad\qquad i = 1, \ldots, m, \\
\mathbb{J}_{i0} &= \{j|y_{ij} = 0,\ \boldsymbol{y}_j \neq \mathbf{0},\ j = 1, \ldots, n\}, \quad n_{i0} = \#\{\mathbb{J}_{i0}\}.
\end{aligned}$$

Then, we have $\#\{\mathbb{J}\} = n - n_0$ and

$$\#\{\mathbb{J}_i\} = \#\{\mathbb{J}\} - \#\{\mathbb{J}_{i0}\} = n - n_0 - n_{i0}.$$

The observed-data likelihood function can be rewritten as

$$\begin{aligned}
& L(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta}|Y_{\text{obs}}) \\
= \ & [\phi + (1-\phi)\mathrm{e}^{-\lambda_+}]^{n_0}(1-\phi)^{n-n_0} \\
& \times \prod_{j\in\mathbb{J}}\prod_{i=1}^{m} \frac{\lambda_i(\lambda_i + \theta_i y_{ij})^{y_{ij}-1}\mathrm{e}^{-\lambda_i-\theta_i y_{ij}}}{y_{ij}!} \\
\propto \ & [\phi + (1-\phi)\mathrm{e}^{-\lambda_+}]^{n_0}(1-\phi)^{n-n_0} \\
& \times \prod_{i=1}^{m}\prod_{j\in\mathbb{J}} \lambda_i(\lambda_i + \theta_i y_{ij})^{y_{ij}-1}\mathrm{e}^{-\lambda_i-\theta_i y_{ij}} \\
= \ & [\phi + (1-\phi)\mathrm{e}^{-\lambda_+}]^{n_0}(1-\phi)^{n-n_0} \\
& \times \prod_{i=1}^{m} \mathrm{e}^{-n_{i0}\lambda_i} \prod_{j\in\mathbb{J}_i} \lambda_i(\lambda_i + \theta_i y_{ij})^{y_{ij}-1}\mathrm{e}^{-\lambda_i-\theta_i y_{ij}},
\end{aligned}$$

so that the log-likelihood function is

$$\begin{aligned}
& \ell(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta}|Y_{\text{obs}}) \\
& = n_0 \log[\phi + (1-\phi)\mathrm{e}^{-\lambda_+}] + (n-n_0)\log(1-\phi) - \sum_{i=1}^{m} n_{i0}\lambda_i \\
& \quad + \sum_{i=1}^{m}\sum_{j\in\mathbb{J}_i}\Big[\log(\lambda_i) + (y_{ij}-1)\log(\lambda_i + \theta_i y_{ij}) - \lambda_i - \theta_i y_{ij}\Big] \\
& \hat{=}\ \ell_0(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta}) + \ell_1(\phi, \boldsymbol{\lambda}) + \sum_{i=1}^{m}\sum_{j\in\mathbb{J}_i} \ell_{ij}(\boldsymbol{\lambda}, \boldsymbol{\theta}),
\end{aligned}$$

where

$$\text{(3.17)} \quad \begin{cases}
\ell_0(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta}) = (n-n_0)\log(1-\phi) - \sum_{i=1}^{m}(\sum_{j\in\mathbb{J}_i} y_{ij})\theta_i \\
\qquad\qquad + \sum_{i=1}^{m}\Big[(n-n_0-n_{i0})\log(\lambda_i) - (n-n_0)\lambda_i\Big], \\
\ell_1(\phi, \boldsymbol{\lambda}) = n_0 \log[\phi + (1-\phi)\mathrm{e}^{-\lambda_+}], \\
\ell_{ij}(\boldsymbol{\lambda}, \boldsymbol{\theta}) = (y_{ij}-1)\log(\lambda_i + \theta_i y_{ij}),\ j \in \mathbb{J}_i,\ i = 1, \ldots, m.
\end{cases}$$

For any concave function $f(\cdot)$, Jensen's inequality implies that

$$\text{(3.18)} \qquad f\left(\sum_{j=1}^{n} \alpha_j h_j(\boldsymbol{x})\right) \geqslant \sum_{j=1}^{n} \alpha_j f(h_j(\boldsymbol{x})),$$

where $\alpha_j \geqslant 0$ and $\sum_{j=1}^{n} \alpha_j = 1$. For $\ell_1(\phi, \boldsymbol{\lambda})$ in (3.17), we apply (3.18) to $n_0 \log(\cdot)$ and can construct a Q_1 function as follows:

$$\begin{aligned}
& Q_1(\phi, \boldsymbol{\lambda}|\phi^{(t)}, \boldsymbol{\lambda}^{(t)}) \\
= \ & C_1 + \frac{n_0\phi^{(t)}}{\beta^{(t)}}\log(\phi) + \frac{n_0(\beta^{(t)} - \phi^{(t)})}{\beta^{(t)}}\log(1-\phi) \\
& - \frac{n_0(\beta^{(t)} - \phi^{(t)})}{\beta^{(t)}}\lambda_+ \\
\leqslant \ & \ell_1(\phi, \boldsymbol{\lambda}),
\end{aligned}$$

where C_1 is a constant not involving $(\phi, \boldsymbol{\lambda})$, and

$$\beta^{(t)} = \phi^{(t)} + (1-\phi^{(t)})\mathrm{e}^{-\lambda_+^{(t)}}.$$

Similarly, for $\ell_{ij}(\boldsymbol{\lambda}, \boldsymbol{\theta})$ in (3.17), we can construct a Q_{ij} function as follows:

$$\begin{aligned}
Q_{ij}(\boldsymbol{\lambda}, \boldsymbol{\theta}|\boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)}) &= C_2 + \frac{\lambda_i^{(t)}(y_{ij}-1)}{\lambda_i^{(t)} + \theta_i^{(t)} y_{ij}}\log(\lambda_i) \\
& \quad + \frac{\theta_i^{(t)} y_{ij}(y_{ij}-1)}{\lambda_i^{(t)} + \theta_i^{(t)} y_{ij}}\log(\theta_i) \\
& \leqslant \ell_{ij}(\boldsymbol{\lambda}, \boldsymbol{\theta}),
\end{aligned}$$

Hence, we can construct the Q function for $\ell(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta}|Y_{\text{obs}})$ as follows:

$$\begin{aligned}
& Q(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta}|\phi^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)}) \\
= \ & \ell_0(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta}) + Q_1(\phi, \boldsymbol{\lambda}|\phi^{(t)}, \boldsymbol{\lambda}^{(t)}) \\
& + \sum_{i=1}^{m}\sum_{j\in\mathbb{J}_i} Q_{ij}(\boldsymbol{\lambda}, \boldsymbol{\theta}|\boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)}) \\
= \ & C + Q_{(\mathrm{I})}(\phi|\phi^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)}) + Q_{(\mathrm{II})}(\boldsymbol{\lambda}|\phi^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)}) \\
& + Q_{(\mathrm{III})}(\boldsymbol{\theta}|\phi^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)}),
\end{aligned}$$

where the parameters ϕ, $\boldsymbol{\lambda}$, $\boldsymbol{\theta}$ are separated, C is a constant not involving $(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})$, and

$$\left\{\begin{array}{rl} & Q_{(\mathrm{I})}(\phi|\phi^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)}) \\ = & \dfrac{n_0\phi^{(t)}}{\beta^{(t)}}\log(\phi) + \left(n - \dfrac{n_0\phi^{(t)}}{\beta^{(t)}}\right)\log(1-\phi), \\ & Q_{(\mathrm{II})}(\boldsymbol{\lambda}|\phi^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)}) \\ = & -\dfrac{n_0(\beta^{(t)}-\phi^{(t)})}{\beta^{(t)}}\lambda_+ + \displaystyle\sum_{i=1}^m \sum_{j\in\mathbb{J}_i} \frac{(y_{ij}-1)\lambda_i^{(t)}}{\lambda_i^{(t)}+\theta_i^{(t)}y_{ij}}\log(\lambda_i) \\ & + \displaystyle\sum_{i=1}^m \Big[(n-n_0-n_{i0})\log(\lambda_i) - (n-n_0)\lambda_i\Big], \\ & Q_{(\mathrm{III})}(\boldsymbol{\theta}|\phi^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)}) \\ = & \displaystyle\sum_{i=1}^m \sum_{j\in\mathbb{J}_i}\left[\frac{\theta_i^{(t)}y_{ij}(y_{ij}-1)}{\lambda_i^{(t)}+\theta_i^{(t)}y_{ij}}\log(\theta_i) - \theta_i y_{ij}\right]. \end{array}\right.$$

Therefore, the explicit MM iterations are given by

$$\left\{\begin{array}{rcl} \phi^{(t+1)} & = & \dfrac{n_0\phi^{(t)}}{n\beta^{(t)}}, \\ \lambda_i^{(t+1)} & = & \dfrac{n - n_0 - n_{i0} + \displaystyle\sum_{j\in\mathbb{J}_i}\frac{(y_{ij}-1)\lambda_i^{(t)}}{\lambda_i^{(t)}+\theta_i^{(t)}y_{ij}}}{n - n\phi^{(t+1)}}, \\ \theta_i^{(t+1)} & = & \dfrac{\sum_{j\in\mathbb{J}_i}\dfrac{\theta_i^{(t)}y_{ij}(y_{ij}-1)}{\lambda_i^{(t)}+\theta_i^{(t)}y_{ij}}}{\sum_{j\in\mathbb{J}_i} y_{ij}}, \end{array}\right. \tag{3.19}$$

for $i = 1, \ldots, m$.

3.4 Bootstrap confidence intervals for small sample sizes

The Wald *confidence interval* (CI) of ϕ specified by (3.4) may fall outside the unit interval $[0,1]$. The Wald CIs of $\{\lambda_i\}_{i=1}^m$ and $\{\theta_i\}_{i=1}^m$ given by (3.4) are reliable only for large sample sizes. For small sample sizes, the bootstrap method is a useful tool to find CI for an arbitrary function of ϕ, $\{\lambda_i\}_{i=1}^m$ and $\{\theta_i\}_{i=1}^m$, say, $\vartheta = h(\phi, \lambda_1, \ldots, \lambda_m, \theta_1, \ldots, \theta_m)$. Let $\hat\vartheta = h(\hat\phi, \hat\lambda_1, \ldots, \hat\lambda_m, \hat\theta_1, \ldots, \hat\theta_m)$ denote the MLE of ϑ, where $\hat\phi$, $\{\hat\lambda_i\}_{i=1}^m$ and $\{\hat\theta_i\}_{i=1}^m$ represent the respective MLEs of ϕ, $\{\lambda_i\}_{i=1}^m$ and $\{\theta_i\}_{i=1}^m$ calculated by means of the second EM algorithm (3.14)–(3.16) or the MM algorithm (3.19). Based on the obtained MLEs $\hat\phi$, $\{\hat\lambda_i\}_{i=1}^m$ and $\{\hat\theta_i\}_{i=1}^m$, we can generate

$$\mathbf{y}_1^*, \ldots, \mathbf{y}_n^* \stackrel{\text{iid}}{\sim} \mathrm{ZIGP}^{(\mathrm{I})}(\hat\phi; \hat\lambda_1, \ldots, \hat\lambda_m, \hat\theta_1, \ldots, \hat\theta_m).$$

Having obtained $Y_{\text{obs}}^* = \{\boldsymbol{y}_1^*, \ldots, \boldsymbol{y}_n^*\}$, we can calculate the bootstrap replication $\hat\phi^*$, $\{\hat\lambda_i^*\}_{i=1}^m$ and $\{\hat\theta_i^*\}_{i=1}^m$ and get $\hat\vartheta^* = h(\hat\phi^*, \hat\lambda_1^*, \ldots, \hat\lambda_m^*, \hat\theta_1^*, \ldots, \hat\theta_m^*)$. Independently repeating this process G times, we obtain G bootstrap replications $\{\hat\vartheta_g^*\}_{g=1}^G$. Consequently, the standard error, $\mathrm{se}(\hat\vartheta)$, of $\hat\vartheta$ can be estimated by the sample standard deviation of the G replications, i.e.,

$$\widehat{\mathrm{se}}(\hat\vartheta) = \left\{\frac{1}{G-1}\sum_{g=1}^G [\hat\vartheta_g^* - (\hat\vartheta_1^* + \cdots + \hat\vartheta_G^*)/G]^2\right\}^{1/2}. \tag{3.20}$$

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

$$\left[\hat\vartheta - z_{\alpha/2}\cdot\widehat{\mathrm{se}}(\hat\vartheta),\ \hat\vartheta + z_{\alpha/2}\cdot\widehat{\mathrm{se}}(\hat\vartheta)\right]. \tag{3.21}$$

Alternatively, if $\{\hat\vartheta_g^*\}_{g=1}^G$ is non-normally distributed, the second $(1-\alpha)100\%$ bootstrap CI of ϑ can be obtained as

$$[\hat\vartheta_{\mathrm{L}},\ \hat\vartheta_{\mathrm{U}}], \tag{3.22}$$

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

3.5 Testing hypotheses for large sample sizes

3.5.1 Likelihood ratio test for zero inflation

Suppose we want to test the null hypothesis

$$H_0\colon \phi = 0 \quad \text{against} \quad H_1\colon \phi > 0. \tag{3.23}$$

Under H_0, the *likelihood ratio test* (LRT) statistic (Jansakul and Hinde, 2002, p. 78 [9]; Joe and Zhu, 2005, p. 225 [10])

$$\begin{aligned} T_1 &= -2\{\ell(0, \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\theta}}_0|Y_{\text{obs}}) - \ell(\hat\phi, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}}|Y_{\text{obs}})\} \\ &\dot\sim\ 0.5\chi^2(0) + 0.5\chi^2(1), \end{aligned} \tag{3.24}$$

where $\hat{\boldsymbol{\lambda}}_0$ and $\hat{\boldsymbol{\theta}}_0$ are the MLEs of $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ under H_0, $(\hat\phi, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}})$ are the unconstrained MLEs of $(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})$, and $\chi^2(0)$ denotes the degenerate distribution with all mass at zero. The corresponding p-value is

$$p_{\mathrm{v}1} = \Pr(T_1 > t_1|H_0) = \frac{1}{2}\Pr\{\chi^2(1) > t_1\}, \tag{3.25}$$

where t_1 is the realization of the LRT statistic T_1.

3.5.2 Score test for zero inflation

In this subsection, we will develop a score test for testing zero inflation in the Type I multivariate ZIGP model by reparametrization. Let

$$\gamma = \frac{\phi}{1-\phi}, \tag{3.26}$$

then, testing H_0 specified by (3.23) is equivalent to testing $H_0^*\colon \gamma = 0$. The observed-data log-likelihood function now becomes

$$\ell^* = \ell(\gamma, \boldsymbol{\lambda}, \boldsymbol{\theta}|Y_{\text{obs}})$$

$$= n_0 \log(\gamma + \mathrm{e}^{-\lambda_+}) - n\log(1+\gamma) - (n-n_0)\lambda_+ + \sum_{j=1}^{n}\sum_{i=1}^{m} [\log\lambda_i + (y_{ij}-1)\log(\lambda_i + \theta_i y_{ij}) - \theta_i y_{ij}] .$$

The score vector is

$$U(\gamma, \boldsymbol{\lambda}, \boldsymbol{\theta}) = \left(\frac{\partial \ell^*}{\partial \gamma}, \frac{\partial \ell^*}{\partial \boldsymbol{\lambda}}, \frac{\partial \ell^*}{\partial \boldsymbol{\theta}}\right)^\top$$

and the Fisher information matrix is

$$\mathbf{J}(\gamma, \boldsymbol{\lambda}, \boldsymbol{\theta}) = (J_{ik}) = E[\mathbf{I}(\gamma, \boldsymbol{\lambda}, \boldsymbol{\theta}|Y_{\mathrm{obs}})],$$

see Appendix C.

Under H_0^*, the score test statistic

$$T_2 = U^\top(\hat{\gamma}_0, \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\theta}}_0)\mathbf{J}^{-1}(\hat{\gamma}_0, \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\theta}}_0)U(\hat{\gamma}_0, \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\theta}}_0) \dot{\sim} \chi^2(1), \tag{3.27}$$

where $\hat{\gamma}_0 = 0$, $\hat{\boldsymbol{\lambda}}_0$ and $\hat{\boldsymbol{\theta}}_0$ denote the MLEs of $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ under H_0^*, The corresponding p-value is given by

$$p_{\mathrm{v2}} = \Pr(T_2 > t_2|H_0) = \Pr\{\chi^2(1) > t_2\}, \tag{3.28}$$

where t_2 is the realization of the score test statistic T_2.

3.5.3 Likelihood ratio test for testing equality of all λ_i's

Suppose we want to test the null hypothesis

$$H_0\colon \lambda_1 = \cdots = \lambda_m = \lambda \quad \text{vs} \quad H_1\colon H_0 \text{ is not true.} \tag{3.29}$$

Under H_0, the LRT statistic

$$T_3 = -2\{\ell(\hat{\phi}_0, \hat{\lambda}_0, \hat{\boldsymbol{\theta}}_0|Y_{\mathrm{obs}}) - \ell(\hat{\phi}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}}|Y_{\mathrm{obs}})\} \dot{\sim} \chi^2(m-1), \tag{3.30}$$

where $(\hat{\phi}_0, \hat{\lambda}_0, \hat{\boldsymbol{\theta}}_0)$ are the MLEs of $(\phi, \lambda, \boldsymbol{\theta})$ under H_0, and $(\hat{\phi}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}})$ are the unconstrained MLEs of $(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})$. The corresponding p-value is given by

$$p_{\mathrm{v3}} = \begin{cases} 2\min\{\Pr(T_3 > t_3|H_0),\ \Pr(T_3 \leqslant t_3|H_0)\}, & \text{if } m > 3, \\ \Pr(T_3 > t_3|H_0), & \text{if } m = 2, 3, \end{cases} \tag{3.31}$$

where t_3 is the realization of the LRT statistic T_3. When $p_{\mathrm{v3}} > \alpha$, we cannot reject the null hypothesis H_0 at the α level of significance. However, if H_0 specified by (3.29) is rejected, we could consider to test H_0': $\lambda_i = \lambda_j$ for a fixed pair (i, j), where $i, j = 1, \ldots, m$; $i \neq j$, and the corresponding test statistic follows $\chi^2(1)$.

3.5.4 Score test for testing equality of all λ_i's

Let γ be defined by (3.26), then, we apply score test to test H_0 specified by (3.29). Under H_0, the score test statistic

$$T_4 = U^\top(\hat{\gamma}_0, \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\theta}}_0)\mathbf{J}^{-1}(\hat{\gamma}_0, \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\theta}}_0)U(\hat{\gamma}_0, \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\theta}}_0) \dot{\sim} \chi^2(m-1), \tag{3.32}$$

where $\hat{\boldsymbol{\lambda}}_0 = \hat{\lambda}\mathbf{1}_m$, and $(\hat{\gamma}_0, \hat{\lambda}, \hat{\boldsymbol{\theta}}_0)$ are the MLEs of $(\gamma, \lambda, \boldsymbol{\theta})$ under H_0. Hence, the p-value is

$$p_{\mathrm{v4}} = \begin{cases} 2\min\{\Pr(T_4 > t_4|H_0),\ \Pr(T_4 \leqslant t_4|H_0)\}, & \text{if } m > 3, \\ \Pr(T_4 > t_4|H_0), & \text{if } m = 2, 3, \end{cases} \tag{3.33}$$

where t_4 is the realization of the score test statistic T_4.

3.5.5 Likelihood ratio test for testing equality of all θ_i's

Suppose we want to test the null hypothesis

$$H_0\colon \theta_1 = \cdots = \theta_m = \theta \quad \text{vs} \quad H_1\colon H_0 \text{ is not true.} \tag{3.34}$$

Under H_0, the LRT statistic

$$T_5 = -2\{\ell(\hat{\phi}_0, \hat{\boldsymbol{\lambda}}_0, \hat{\theta}_0|Y_{\mathrm{obs}}) - \ell(\hat{\phi}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}}|Y_{\mathrm{obs}})\} \dot{\sim} \chi^2(m-1), \tag{3.35}$$

where $(\hat{\phi}_0, \hat{\boldsymbol{\lambda}}_0, \hat{\theta}_0)$ are the MLEs of $(\phi, \boldsymbol{\lambda}, \theta)$ under H_0, and $(\hat{\phi}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}})$ are the unconstrained MLEs of $(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})$. The corresponding p-value is given by

$$p_{\mathrm{v5}} = \begin{cases} 2\min\{\Pr(T_5 > t_5|H_0),\ \Pr(T_5 \leqslant t_5|H_0)\}, & \text{if } m > 3, \\ \Pr(T_5 > t_5|H_0), & \text{if } m = 2, 3, \end{cases} \tag{3.36}$$

where t_5 is the realization of the LRT statistic T_5. When $p_{\mathrm{v5}} > \alpha$, we cannot reject the null hypothesis H_0 at the α level of significance. If H_0 specified by (3.34) cannot be rejected and $\theta = 0$, we can say this $\mathrm{ZIGP}^{(\mathrm{I})}$ model reduced to $\mathrm{ZIP}^{(\mathrm{I})}$ model. However, if H_0 specified by (3.34) is rejected, we could consider to test H_0': $\theta_i = \theta_j$ for a fixed pair (i, j), where $i, j = 1, \ldots, m$; $i \neq j$, and the corresponding test statistic follows $\chi^2(1)$.

3.5.6 Score test for testing equality of all θ_i's

Let γ be defined by (3.26), then, we apply score test to test H_0 specified by (3.34). Under H_0, the score test statistic

$$T_6 = U^\top(\hat{\gamma}_0, \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\theta}}_0)\mathbf{J}^{-1}(\hat{\gamma}_0, \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\theta}}_0)U(\hat{\gamma}_0, \hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\theta}}_0) \dot{\sim} \chi^2(m-1), \tag{3.37}$$

where $\hat{\boldsymbol{\theta}}_0 = \hat{\theta}\mathbf{1}_m$, and $(\hat{\gamma}_0, \hat{\boldsymbol{\lambda}}_0, \hat{\theta})$ are the MLEs of $(\gamma, \boldsymbol{\lambda}, \theta)$ under H_0. Hence, the p-value is

$$p_{\mathrm{v6}} = \begin{cases} 2\min\{\Pr(T_6 > t_6|H_0),\ \Pr(T_6 \leqslant t_6|H_0)\}, & \text{if } m > 3, \\ \Pr(T_6 > t_6|H_0), & \text{if } m = 2 \text{ or } 3, \end{cases} \tag{3.38}$$

where t_6 is the realization of the score test statistic T_6.

Table 1. MLEs and bootstrap CIs of parameters for $m = 2$

Parameter	True Value	MLE	Width	CP	True Value	MLE	Width	CP
ϕ	0.1	0.1001	0.1280	0.937	0.2	0.1998	0.1571	0.953
λ_1	2	2.0285	0.8970	0.960	3	3.0551	1.2652	0.942
λ_2	2	2.0437	0.9021	0.941	3	3.0396	1.2692	0.940
θ_1	0.3	0.2881	0.2604	0.933	0.4	0.3875	0.2359	0.929
θ_2	0.3	0.2834	0.2602	0.925	0.4	0.3891	0.2353	0.932

Note: MLE is the mean of the 1000 point estimates via the EM algorithm (3.14)–(3.16); width and CP are the average width and coverage proportion of 1000 bootstrap CIs.

Table 2. MLEs and bootstrap CIs of parameters for $m = 3$

Parameter	True Value	MLE	Width	CP	True Value	MLE	Width	CP
ϕ	0.1	0.1004	0.1150	0.937	0.2	0.1994	0.1559	0.925
λ_1	2	2.0218	0.8579	0.941	3	3.0389	1.2772	0.936
λ_2	4	4.0471	1.4900	0.941	5	5.0745	1.9290	0.929
λ_3	6	6.1214	2.1171	0.930	7	7.1299	2.5927	0.940
θ_1	0.3	0.2898	0.2555	0.920	0.2	0.1867	0.2871	0.916
θ_2	0.5	0.4926	0.1865	0.934	0.4	0.3919	0.2213	0.920
θ_3	0.7	0.6911	0.1197	0.928	0.6	0.5912	0.1557	0.923

Note: MLE is the mean of the 1000 point estimates via the EM algorithm (3.14)–(3.16); width and CP are the average width and coverage proportion of 1000 bootstrap CIs.

4. SIMULATION STUDIES

To evaluate the performance of the proposed statistical methods in Section 3 for the Type I multivariate ZIGP distribution, we first investigate the accuracy of point estimates and confidence interval estimates for different parameter settings via simulation studies. Second, we assess the performance of the LRT with the score test by comparing their type I error rates and powers.

4.1 Accuracy of point estimates and interval estimates

In this subsection, we compare the accuracy of point estimates and confidence intervals by considering both cases of two-dimensional (i.e., $m = 2$) and three-dimensional (i.e., $m = 3$). When $m = 2$, the parameters $(\phi, \lambda_1, \lambda_2, \theta_1, \theta_2)$ are set to be $(0.1, 2, 2, 0.3, 0.3)$ and $(0.2, 3, 3, 0.4, 0.4)$. When $m = 3$, the parameters $(\phi, \lambda_1, \lambda_2, \lambda_3, \theta_1, \theta_2, \theta_3)$ are set to be $(0.1, 2, 4, 6, 0.3, 0.5, 0.7)$ and $(0.2, 3, 5, 7, 0.2, 0.4, 0.6)$. For each parameter configuration, we generate

$$\{\mathbf{y}_j\}_{j=1}^n \overset{\text{iid}}{\sim} \text{ZIGP}_m^{(\text{I})}(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})$$

with $n = 100$, and calculate the MLEs via the second EM algorithm (3.14)–(3.16) and the 95% bootstrap CIs with $G = 1{,}000$. Here, we independently repeat this process 1,000 times and report the corresponding mean of the MLEs, the average width and the *coverage probability* (CP) of the bootstrap CIs in Tables 1 and 2, respectively.

4.2 Comparison of the LRT with the score test

4.2.1 Tests for zero inflation

In this subsection, we compare the corresponding type I error rates (with H_0: $\phi = 0$) and powers (with H_1: $\phi > 0$) between the LRT and the score test for various sample sizes via simulations, where the values of ϕ in H_1 are chosen to be 0.01, 0.03, 0.05, 0.07, 0.10, 0.15. For a given pair of (n, ϕ), we first draw

$$Z_1^{(l)}, \ldots, Z_n^{(l)} \overset{\text{iid}}{\sim} \text{Bernoulli}(1 - \phi)$$

for $l = 1, \ldots, L\,(L = 1{,}000)$, and then independently generate

$$X_{11}^{(l)}, \ldots, X_{1n}^{(l)} \overset{\text{iid}}{\sim} \text{GP}(\lambda_1, \theta_1)$$

and

$$X_{21}^{(l)}, \ldots, X_{2n}^{(l)} \overset{\text{iid}}{\sim} \text{GP}(\lambda_2, \theta_2),$$

where only $\lambda_1 = 5, \theta_1 = 0.4$ and $\lambda_2 = 3, \theta_2 = 0.6$ are considered. Finally, we set

$$\mathbf{y}_j^{(l)} = \begin{pmatrix} Y_{1j}^{(l)} \\ Y_{2j}^{(l)} \end{pmatrix} = Z_j^{(l)} \begin{pmatrix} X_{1j}^{(l)} \\ X_{2j}^{(l)} \end{pmatrix}, \quad j = 1, \ldots, n.$$

All hypothesis testings are conducted at the significant level $\alpha = 0.05$. Let r_k denote the number of rejecting the null hypothesis H_0: $\phi = 0$ by the test statistics T_k $(k = 1, 2)$ given by (3.24) and (3.27), respectively. Hence, the actual significance level can be estimated by r_k/L with $\phi = 0$ and

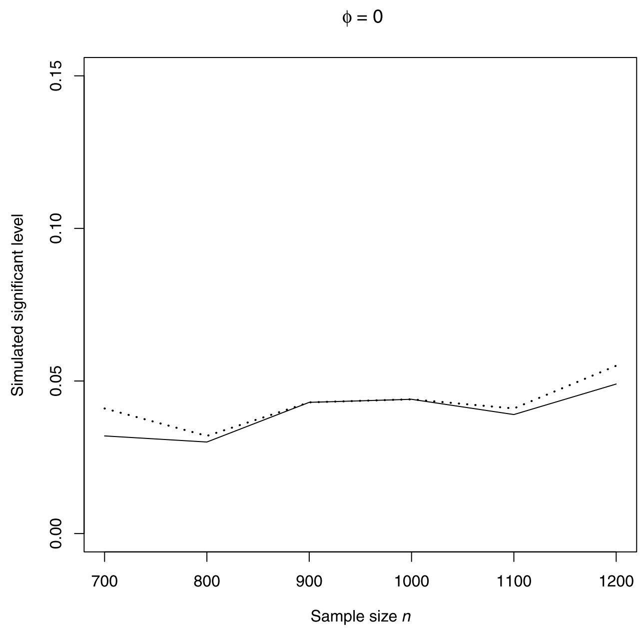

Figure 1. Comparison of type I error rates between the LRT (solid line) and the score test (dotted line).

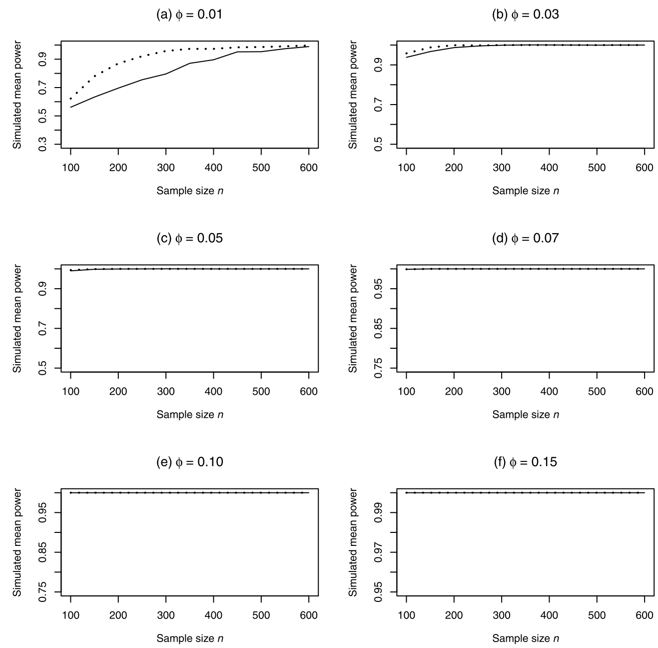

Figure 2. Comparison of powers between the LRT (solid line) and the score test (dotted line).

the power of the test statistic T_k can be estimated by r_k/L with $\phi > 0$.

Figure 1 shows that the comparison of type I error rates between the LRT and the score test. In general, we can see the LRT test have a explicitly better performance in controlling its type I error rates around the pre-chosen nominal level.

Figure 2 gives the comparison of powers between the LRT and the score test for different values of $\phi > 0$. It is not difficult to find that there is no significant difference between the powers of the two tests when ϕ is larger than 0.03. But when $\phi = 0.01$, the score test is slightly more powerful than the LRT.

4.2.2 Tests for equality of λ_1 and λ_2

In this subsection, we compare the respective type I error rates (with H_0: $\lambda_1 = \lambda_2$) and powers (with H_1: $\lambda_1 \neq \lambda_2$) between the LRT and the score test for various sample sizes and different combinations of (λ_1, λ_2) via simulations, where the values of (λ_1, λ_2) are set to be $(4, 4)$ and $(5, 8)$. For a given combination of $(n, \lambda_1, \lambda_2)$, we first generate

$$Z_1^{(l)}, \ldots, Z_n^{(l)} \overset{\text{iid}}{\sim} \text{Bernoulli}(1-\phi)$$

for $l = 1, \ldots, L\,(L = 1{,}000)$, and then independently generate

$$X_{11}^{(l)}, \ldots, X_{1n}^{(l)} \overset{\text{iid}}{\sim} \text{GP}(\lambda_1, \theta_1)$$
and
$$X_{21}^{(l)}, \ldots, X_{2n}^{(l)} \overset{\text{iid}}{\sim} \text{GP}(\lambda_2, \theta_2),$$

where only $\phi = 0.5, \theta_1 = 0.4, \theta_2 = 0.6$ are considered. Then, we have

$$\mathbf{y}_j^{(l)} = \begin{pmatrix} Y_{1j}^{(l)} \\ Y_{2j}^{(l)} \end{pmatrix} = Z_j^{(l)} \begin{pmatrix} X_{1j}^{(l)} \\ X_{2j}^{(l)} \end{pmatrix}, \quad j = 1, \ldots, n.$$

All hypothesis testings are conducted at the significant level $\alpha = 0.05$. Let r_k denote the number of rejecting the null hypothesis H_0: $\lambda_1 = \lambda_2$ by the statistics T_k $(k = 3, 4)$ given by (3.30) and (3.32), respectively. Hence, the actual significance level can be estimated by r_k/L with $\lambda_1 = \lambda_2$ and the power of the test statistic T_k can be estimated by r_k/L with $\lambda_1 \neq \lambda_2$.

Figure 3 shows that some comparison of type I error rates between the LRT and the score test. In general, there is no significance difference between the two tests' performances in controlling their type I error rates around the pre-chosen nominal level.

Figure 4 gives the comparison of powers between the LRT and the score test for one case with $\lambda_1 \neq \lambda_2$. It is not difficult to find that the LRT almost has the same power as the score test is, no matter the sample size is small or large.

4.2.3 Tests for equality of θ_1 and θ_2

In this subsection, we compare the respective type I error rates (with H_0: $\theta_1 = \theta_2$) and powers (with H_1: $\theta_1 \neq \theta_2$) between the LRT and the score test for various sample sizes and different combinations of (θ_1, θ_2) via simulations, where the values of (θ_1, θ_2) are set to be $(0.6, 0.6)$ and $(0.3, 0.7)$.

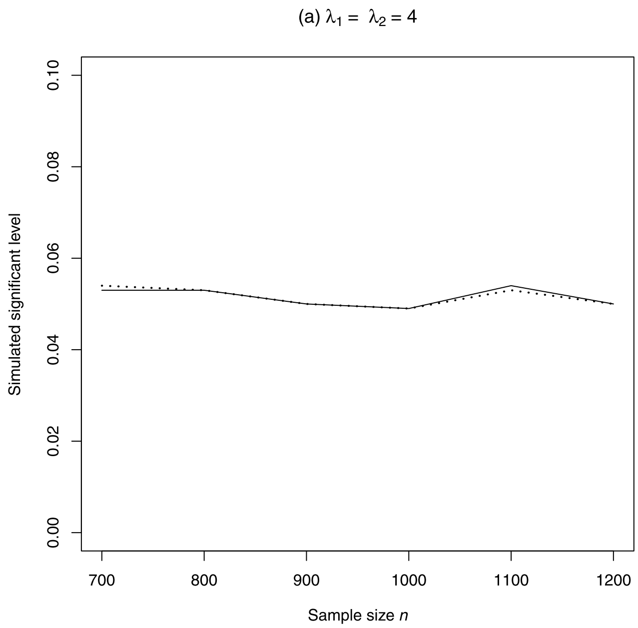

Figure 3. Comparison of type I error rates between the LRT (solid line) and the score test (dotted line).

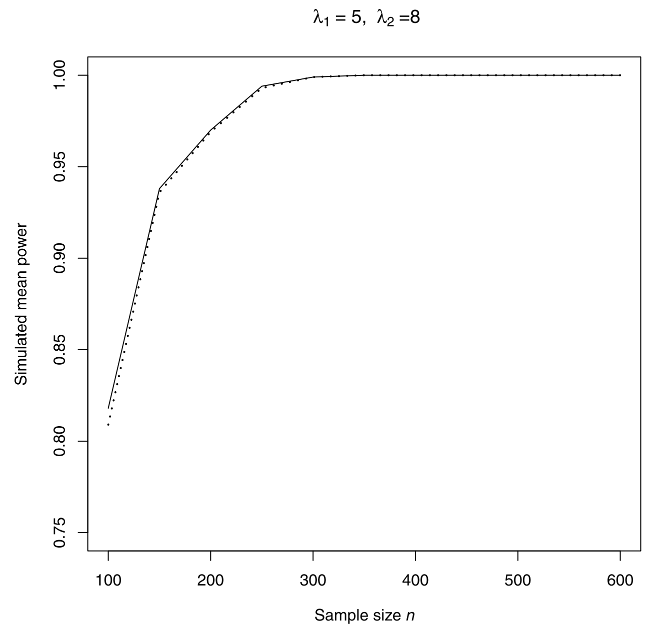

Figure 4. Comparison of powers between the LRT (solid line) and the score test (dotted line).

For a given combination of (n, θ_1, θ_2), we first generate

$$Z_1^{(l)}, \ldots, Z_n^{(l)} \overset{\text{iid}}{\sim} \text{Bernoulli}(1-\phi)$$

for $l = 1, \ldots, L\,(L = 1{,}000)$, and then independently generate

$$X_{11}^{(l)}, \ldots, X_{1n}^{(l)} \overset{\text{iid}}{\sim} \text{GP}(\lambda_1, \theta_1)$$
and
$$X_{21}^{(l)}, \ldots, X_{2n}^{(l)} \overset{\text{iid}}{\sim} \text{GP}(\lambda_2, \theta_2),$$

where only $\phi = 0.5, \lambda_1 = 5, \lambda_2 = 8$ are considered. Then, we have

$$\mathbf{y}_j^{(l)} = \begin{pmatrix} Y_{1j}^{(l)} \\ Y_{2j}^{(l)} \end{pmatrix} = Z_j^{(l)} \begin{pmatrix} X_{1j}^{(l)} \\ X_{2j}^{(l)} \end{pmatrix}, \quad j = 1, \ldots, n.$$

All hypothesis testings are conducted at the significant level $\alpha = 0.05$. Let r_k denote the number of rejecting the null hypothesis H_0: $\theta_1 = \theta_2$ by the statistics T_k $(k = 3, 4)$ given by (3.35) and (3.37), respectively. Hence, the actual significance level can be estimated by r_k/L with $\theta_1 = \theta_2$ and the power of the test statistic T_k can be estimated by r_k/L with $\theta_1 \neq \theta_2$.

Figure 5 shows that some comparison of type I error rates between the LRT and the score test. In general, there is no significance difference between the two tests' performances in controlling their type I error rates around the pre-chosen nominal level.

Figure 6 gives the comparison of powers between the LRT and the score test for one case with $\theta_1 \neq \theta_2$. It is not difficult to find that the LRT almost has the same power as the score test is, no matter the sample size is small or large.

5. TWO REAL EXAMPLES

In this section, two real data sets are used to illustrate the proposed methods, where the Newton–Raphson algorithm for finding the MLEs of parameters does not work for the two examples because the corresponding observed information matrices are nearly singular, while the Fisher-scoring algorithm is always sensitive to the initial values. Unfortunately, the first EM algorithm does not work in the second example. As expected, the second EM algorithm and the MM algorithm work well in the two examples.

5.1 The children's absenteeism data in Indonesia

In a survey of Indonesian family life conducted by Strauss *et al.* (2004) [18], the participants included 7,000 households sampled from 321 communities randomly selected from 13 of the nation's 26 Provinces, in which 83% of the Indonesian population lived. Among those households with one child per household, 437 household heads were asked questions about the health of their children. Let Y_1 denote the number of days the children missed their primary activities due to illness in the last four weeks and Y_2 denote the number of days the children spent in bed due to illness in the last four weeks. Table 3 shows the children's absenteeism data from this survey.

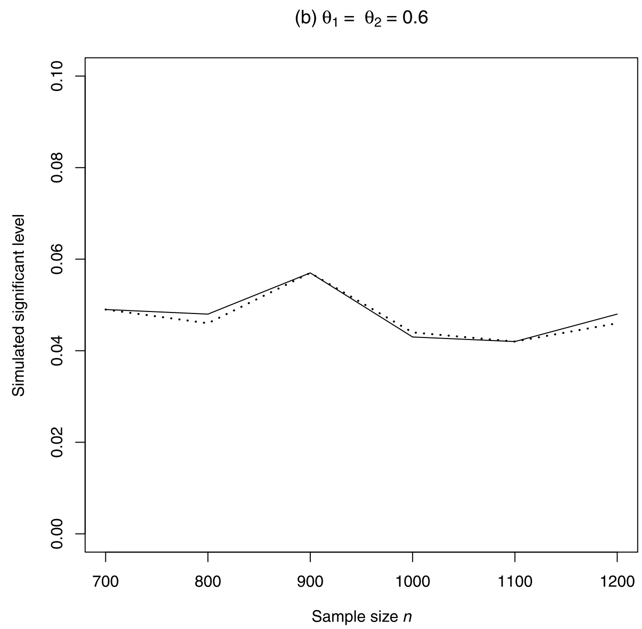

Figure 5. Comparison of type I error rates between the LRT (solid line) and the score test (dotted line).

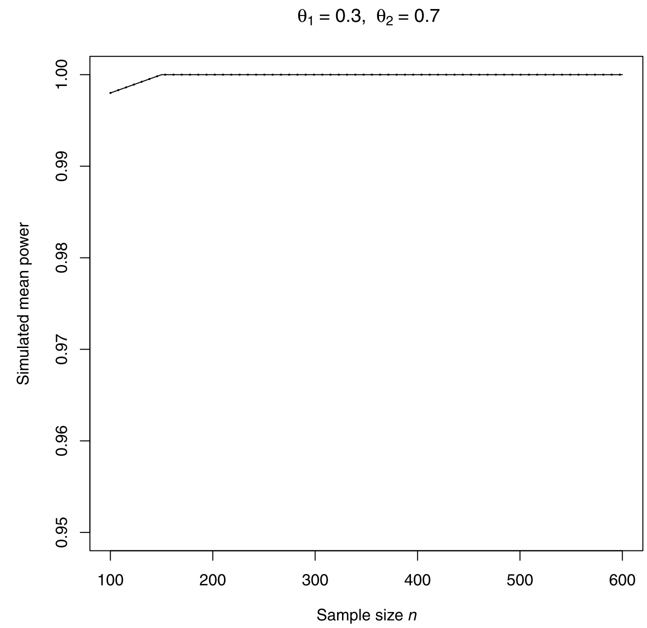

Figure 6. Comparison of powers between the LRT (solid line) and the score test (dotted line).

Table 3. The children's absenteeism data in the Indonesian family life survey (Cheung and Lam, 2006)

$Y_1 \backslash Y_2$	0	1	2	3	4	5	6	7	Total
0	323	0	0	0	0	0	0	0	323
1	10	8	0	0	0	0	0	0	18
2	22	0	9	0	0	0	0	0	31
3	19	1	3	6	0	0	0	0	29
4	3	3	3	0	0	0	0	0	9
5	0	0	0	0	0	1	0	0	1
6	0	0	0	1	0	0	1	0	2
7	12	0	1	2	0	0	0	5	20
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0
10	2	0	0	0	0	0	0	0	2
11	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	1	1
15	0	0	0	1	0	0	0	0	1
Total	391	12	16	10	0	1	1	6	437

5.1.1 Likelihood-based inferences

Let $\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{\text{iid}}{\sim} \text{ZIGP}^{(\text{I})}(\phi; \lambda_1, \lambda_2, \theta_1, \theta_2)$, where $\mathbf{y}_j = (Y_{1j}, Y_{2j})^\top$ for $j = 1, \ldots, n\,(n = 437)$. To find the MLEs of $(\phi, \lambda_1, \lambda_2, \theta_1, \theta_2)$, we randomly choose $(\phi^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)}, \theta_1^{(0)}, \theta_2^{(0)}) = (0.5, 5, 5, 0.5, 0.5)$ as their initial values of the two EM algorithms and MM algorithm, and carefully choose initial values $(\phi^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)}, \theta_1^{(0)}, \theta_2^{(0)}) = (0.5, 1, 1, 0.2, 0.2)$ for the Fisher-scoring algorithm. The MLEs of $(\phi, \lambda_1, \lambda_2, \theta_1, \theta_2)$ converged to $(\hat\phi, \hat\lambda_1, \hat\lambda_2, \hat\theta_1, \hat\theta_2)$ as shown in the second column of Table 4 in 13 iterations for the Fisher-scoring algorithm (3.3), in 105 iterations for the first EM algorithm (3.10)–(3.12), in 17 iterations for the second EM algorithm (3.14)–(3.16) and in 89 iterations for the MM algorithm (3.19). The standard errors of the MLEs $(\hat\phi, \hat\lambda_1, \hat\lambda_2, \hat\theta_1, \hat\theta_2)$ are given in the third column and 95% asymptotic Wald CIs (i.e., (3.4)) of the five parameters are listed in the fourth column of Table 4. With $G = 10{,}000$ bootstrap replications, the two 95% bootstrap CIs of $(\phi, \lambda_1, \lambda_2, \theta_1, \theta_2)$ are shown in the sixth and seventh columns of Table 4.

Suppose that we want to test the null hypothesis H_0: $\phi = 0$ against the alternative hypothesis H_1: $\phi > 0$. According to (3.24) and (3.27), we calculate the values of the LRT statistic and score test statistic, which are given by $t_1 = 182.6755$ and $t_2 = 178.9367$, respectively. Then from (3.25) and (3.28), we have $p_{\text{v}1} = p_{\text{v}2} \approx 0 \ll \alpha = 0.05$. Thus, we should reject H_0.

If we want to test the null hypothesis H_0: $\lambda_1 = \lambda_2$ against the alternative hypothesis H_1: $\lambda_1 \neq \lambda_2$. According to (3.30) and (3.32), we calculate the values of the LRT statistic and score test statistic, which are given by $t_3 = 96.44369$ and $t_4 = 85.29277$, respectively. Then from (3.31) and (3.33), we have $p_{\text{v}3} = p_{\text{v}4} \approx 0 \ll 0.05$. As a result, the H_0 should be rejected.

Suppose that we want to test the null hypothesis H_0: $\theta_1 = \theta_2$ against the alternative hypothesis H_1: $\theta_1 \neq \theta_2$. According to (3.35) and (3.37), we calculate the values of the LRT statistic and score test statistic, which are given

Table 4. MLEs and CIs of parameters for the children's absenteeism data in Indonesia

Parameter	MLE	$\mathrm{std}^{\mathrm{F}}$	95% Wald CI	$\mathrm{std}^{\mathrm{B}}$	95% CI†	95% CI‡
ϕ	0.7252	0.0225	[0.6811, 0.7693]	0.0147	[0.6964, 0.7538]	[0.6961, 0.7536]
λ_1	2.4618	0.2563	[1.9595, 2.9642]	0.1708	[2.1420, 2.8117]	[2.1553, 2.8269]
λ_2	0.5208	0.0758	[0.3721, 0.6694]	0.0501	[0.4247, 0.6210]	[0.4285, 0.6259]
θ_1	0.2772	0.0581	[0.1633, 0.3911]	0.0389	[0.1965, 0.3489]	[0.1932, 0.3448]
θ_2	0.5076	0.0708	[0.3688, 0.6464]	0.0477	[0.4087, 0.5957]	[0.4047, 0.5896]

$\mathrm{std}^{\mathrm{F}}$: The square roots of the diagonal elements of the inverse Fisher information matrix $\mathbf{J}^{-1}(\hat{\phi}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}})$. $\mathrm{std}^{\mathrm{B}}$: The sample standard deviation of the boostrap samples, cf. (3.20). CI†: Normal-based bootstrap CI, cf. (3.21). CI‡: Non-normal-based bootstrap CI, cf. (3.22).

Table 5. Comparisons for Type I multivariate ZIGP distribution and the Type I multivariate ZIP distribution

Model	Criterion	
	AIC	BIC
Type I multivariate ZIGP distribution	1324.855	1345.255
Type I multivariate ZIP distribution	1435.731	1447.971

by $t_5 = 7.086542925$ and $t_6 = 7.914848236$, respectively. Then from (3.36) and (3.38), we have $p_{\mathrm{v1}} = 0.007766492 < 0.05$, $p_{\mathrm{v2}} = 0.004903069 < 0.05$. Thus, we should reject H_0.

5.1.2 Model comparison

Now we focus on the comparison between the Type I multivariate ZIGP model with the Type I multivariate ZIP model under AIC and BIC based on the full likelihood function. In Table 5, we can find that both AIC and BIC of the Type I multivariate ZIGP model are less than those of the Type I multivariate ZIP model, indicating that the proposed Type I multivariate ZIGP model is more appropriate to fit the data set.

5.2 Voluntary and involuntary job changes data

Jung and Winkelmann (1993) [11] provided data on both the numbers of voluntary and involuntary job changes of males during ten period 1974–1984. The samples contain 2124 males who started their working career before or in 1974 and did not retire before 1984. The cross tabulation is given in Table 6.

Table 6. Cross tabulation of voluntary and involuntary job changes (Jung and Winkelmann, 1993)

$Y_1 \backslash Y_2$	0	1	2	3	4	5	6	7	8	9	10	12	Total
0	1227	319	109	27	20	5	1	2	1	0	2	0	1713
1	150	83	23	10	1	3	2	1	0	1	0	0	274
2	34	16	6	6	2	2	1	1	0	0	0	1	69
3	20	5	1	0	2	0	0	0	0	0	0	0	28
4	8	2	0	2	0	0	0	0	0	0	1	0	13
5	6	2	0	0	0	0	0	0	0	0	0	0	8
6	2	0	0	0	0	0	0	0	0	1	0	0	3
7	2	0	0	0	0	0	0	0	0	0	0	0	2
8	3	0	0	0	0	0	0	0	0	0	0	0	3
9	3	0	0	0	0	0	0	0	0	0	0	0	3
10	7	0	0	0	0	0	0	0	0	0	0	0	7
15	1	0	0	0	0	0	0	0	0	0	0	0	1
Total	1463	427	139	45	25	10	4	4	1	2	3	1	2124

5.2.1 Likelihood-based inferences

Let $\mathbf{y}_1, \dots, \mathbf{y}_n \overset{\text{iid}}{\sim} \mathrm{ZIGP}^{(\mathrm{I})}(\phi; \lambda_1, \lambda_2, \theta_1, \theta_2)$, where $\mathbf{y}_j = (Y_{1j}, Y_{2j})^{\top}$ for $j = 1, \dots, n\,(n = 2124)$. To find the MLEs of $(\phi, \lambda_1, \lambda_2, \theta_1, \theta_2)$, we randomly choose $(\phi^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)}, \theta_1^{(0)}, \theta_2^{(0)}) = (0.5, 5, 5, 0.5, 0.5)$ as their initial values of the second EM algorithm and MM algorithm, and carefully choose initial values $(\phi^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)}, \theta_1^{(0)}, \theta_2^{(0)}) = (0.2, 1, 1, 0.1, 0.1)$ for the Fisher-scoring algorithm. The MLEs of $(\phi, \lambda_1, \lambda_2, \theta_1, \theta_2)$ converged to $(\hat{\phi}, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\theta}_1, \hat{\theta}_2)$ as shown in the second column of Table 7 in 12 iterations for the Fisher-scoring algorithm (3.3), in 180 iterations for the second EM algorithm (3.14)–(3.16) and in 219 iterations for the MM algorithm (3.19). The standard errors of the MLEs $(\hat{\phi}, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\theta}_1, \hat{\theta}_2)$ are given in the third column and 95% asymptotic Wald CIs (i.e., (3.4)) of the five parameters are listed in the fourth column of Table 7. With $G = 10{,}000$ bootstrap replications, the two 95% bootstrap CIs of $(\phi, \lambda_1, \lambda_2, \theta_1, \theta_2)$ are shown in the sixth and seventh columns of Table 7.

Suppose that we want to test the null hypothesis H_0: $\phi = 0$ against the alternative hypothesis H_1: $\phi > 0$. According to (3.24) and (3.27), we calculate the values of the LRT statistic and score test statistic, which are given by $t_1 = 15.02807$ and $t_2 = 16.98990$, respectively. Then from (3.25) and (3.28), we have $p_{\mathrm{v1}} = 0.000106 \ll 0.05$, $p_{\mathrm{v2}} = 0.000019 \ll \alpha = 0.05$. Thus, we should reject H_0.

If we want to test the null hypothesis H_0: $\lambda_1 = \lambda_2$ against the alternative hypothesis H_1: $\lambda_1 \neq \lambda_2$. According to (3.30) and (3.32), we calculate the values of the LRT statistic and score test statistic, which are given by $t_3 = 85.2319$ and $t_4 = 85.99736$, respectively. Then from (3.31) and (3.33), we have $p_{\mathrm{v3}} = p_{\mathrm{v4}} \approx 0 \ll 0.05$. As a result, the H_0 should be rejected at the level of $\alpha = 0.05$.

Suppose that we want to test the null hypothesis H_0: $\theta_1 = \theta_2$ against the alternative hypothesis H_1: $\theta_1 \neq \theta_2$. According to (3.35) and (3.37), we calculate the values of

Table 7. MLEs and CIs of parameters for the voluntary and involuntary job changes data

Parameter	MLE	std$^{\text{F}}$	95% Wald CI	std$^{\text{B}}$	95% CI†	95% CI‡
ϕ	0.1937	0.0434	[0.1087, 0.2788]	0.0943	[0.0388, 0.4083]	[0.0918, 0.4796]
λ_1	0.2680	0.0208	[0.2272, 0.3089]	0.0539	[0.1814, 0.3929]	[0.2211, 0.4467]
λ_2	0.4738	0.0360	[0.4032, 0.5444]	0.2031	[0.1455, 0.9415]	[0.3956, 1.1850]
θ_1	0.3928	0.0254	[0.3430, 0.4426]	0.0396	[0.3074, 0.4628]	[0.3024, 0.4594]
θ_2	0.2690	0.0219	[0.2260, 0.3120]	0.1493	[-0.0688, 0.5165]	[-0.2428, 0.3255]

std$^{\text{F}}$: The square roots of the diagonal elements of the inverse Fisher information matrix $\mathbf{J}^{-1}(\hat{\phi}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}})$. std$^{\text{B}}$: The sample standard deviation of the boostrap samples, cf. (3.20). CI†: Normal-based bootstrap CI, cf. (3.21). CI‡: Non-normal-based bootstrap CI, cf. (3.22).

Table 8. Comparisons for Type I multivariate ZIGP distribution and the Type I multivariate ZIP distribution

Model	Criterion	
	AIC	BIC
Type I multivariate ZIGP distribution	7182.795	7211.100
Type I multivariate ZIP distribution	7894.818	7911.801

the LRT statistic and score test statistic, which are given by $t_5 = 15.7782$ and $t_6 = 16.94288$, respectively. Then from (3.36) and (3.38), we have $p_{\text{v}1} = 0.000071 \ll 0.05$, $p_{\text{v}2} = 0.000038 \ll 0.05$. Thus, we should reject H_0 at 0.05 level of significance.

5.2.2 Model comparison

Now we focus on the comparison between the Type I multivariate ZIGP model with the Type I multivariate ZIP model under AIC and BIC based on the full likelihood function. In Table 8, we can see that both AIC and BIC of the Type I multivariate ZIGP model are less than that of the Type I multivariate ZIP model, indicating that the data set is fitted more appropriately by using the proposed Type I multivariate ZIGP model compared with the Type I multivariate ZIP model.

6. DISCUSSION

In this paper we have introduced a multivariate ZIGP distribution, called the Type I multivariate ZIGP model, and developed the distribution theory and its important properties. We have also investigated the efficient likelihood inference approaches via four different algorithms concerning the model parameters. Two data sets in literature have been used to illustrate the applications. For the likelihood inference, actually, the MLE procedure for the ZIGP model is difficult especially when the dimension is large. For the propose model, however, the MLEs, the confidence intervals and the bootstrap method can be easily calculated via the four proposed algorithms: from Fisher scoring to two EM algorithms, to MM algorithm, thus offering substantial computational advantages.

ACKNOWLEDGMENTS

Xuejun Jiang's research was partially supported by a grant from the Natural Science Foundation of China (No. 11101432). The authors contributed equally to this work.

Appendix A: The derivation of (3.2)

Since $\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{\text{iid}}{\sim} \text{ZIGP}_m^{(\text{I})}(\phi, \boldsymbol{\lambda}, \boldsymbol{\theta})$, we have

$$\begin{aligned} E\left[\textstyle\sum_{j=1}^n I(\mathbf{y}_j = \mathbf{0})\right] &= nE[I(\mathbf{y}_1 = \mathbf{0})] = n\Pr(\mathbf{y}_1 = \mathbf{0}) \\ &\overset{(2.3)}{=} n[\phi + (1-\phi)\mathrm{e}^{-\lambda_+}], \end{aligned}$$

which implies the first formula of (3.2). In the follows, we assume that $Y \sim \text{ZIGP}(\phi, \lambda, \theta)$ and only need to prove

$$\text{(A.1)} \quad \begin{cases} E\left[\dfrac{Y-1}{(\lambda+\theta Y)^2}\right] = \dfrac{1-\phi}{\lambda} - \dfrac{1}{\lambda^2} - \dfrac{\theta(1-\phi)}{\lambda+2\theta}, \\ E\left[\dfrac{Y^3-Y^2}{(\lambda+\theta Y)^2}\right] = \dfrac{\lambda(1-\phi)}{1-\theta} + \dfrac{2\lambda(1-\phi)}{\lambda+2\theta}, \\ E\left[\dfrac{Y^2-Y}{(\lambda+\theta Y)^2}\right] = \dfrac{\lambda(1-\phi)}{\lambda+2\theta}. \end{cases}$$

Let

$$s = \sum_{y=0}^{\infty} \frac{(\lambda+\theta y)^{y-1}\mathrm{e}^{-(\lambda+\theta y)}}{y!}.$$

Because

$$\lambda \times s = \sum_{y=0}^{\infty} \frac{\lambda(\lambda+\theta y)^{y-1}\mathrm{e}^{-(\lambda+\theta y)}}{y!} = 1,$$

then we obtain $s = 1/\lambda$. On the one hand, we have

$$\begin{aligned} s &= \sum_{y=0}^{\infty} \frac{(\lambda+\theta y)^{y-1}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\ &= \sum_{y=0}^{\infty} \frac{\lambda(\lambda+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\ &\quad + \sum_{y=0}^{\infty} \frac{\theta y(\lambda+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \end{aligned}$$

$$
\begin{aligned}
&= \lambda\sum_{y=0}^{\infty}\frac{(\lambda+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\
&\quad + \theta\sum_{y=1}^{\infty}\frac{[\lambda+\theta+\theta(y-1)]^{y-1-1}\mathrm{e}^{-[\lambda+\theta+\theta(y-1)]}}{(y-1)!} \\
&= \lambda\sum_{y=0}^{\infty}\frac{(\lambda+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\
&\quad + \theta\sum_{y=0}^{\infty}\frac{[\lambda+\theta+\theta y]^{y-1}\mathrm{e}^{-[\lambda+\theta+\theta y]}}{y!} \\
&= \lambda\sum_{y=0}^{\infty}\frac{(\lambda+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta y)}}{y!} + \frac{\theta}{\lambda+\theta} \\
&\hat{=} \lambda\times s_1 + \frac{\theta}{\lambda+\theta},
\end{aligned}
$$

so that

$$
s_1 = \frac{1}{\lambda^2} - \frac{\theta}{\lambda(\lambda+\theta)}. \tag{A.2}
$$

On the other hand,

$$
\begin{aligned}
s_1 &= \sum_{y=0}^{\infty}\frac{(\lambda+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\
&= \sum_{y=0}^{\infty}\frac{\lambda(\lambda+\theta y)^{y-3}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\
&\quad + \sum_{y=0}^{\infty}\frac{\theta y(\lambda+\theta y)^{y-3}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\
&= \lambda\sum_{y=0}^{\infty}\frac{(\lambda+\theta y)^{y-3}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\
&\quad + \theta\sum_{y=1}^{\infty}\frac{[\lambda+\theta+\theta(y-1)]^{y-1-2}\mathrm{e}^{-[\lambda+\theta+\theta(y-1)]}}{(y-1)!} \\
&= \lambda\sum_{y=0}^{\infty}\frac{(\lambda+\theta y)^{y-3}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\
&\quad + \theta\sum_{y=0}^{\infty}\frac{[\lambda+\theta+\theta y]^{y-2}\mathrm{e}^{-[\lambda+\theta+\theta y]}}{y!} \\
&= \lambda\sum_{y=0}^{\infty}\frac{(\lambda+\theta y)^{y-3}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\
&\quad + \theta\left[\frac{1}{(\lambda+\theta)^2} - \frac{\theta}{(\lambda+\theta)(\lambda+2\theta)}\right] \\
&\hat{=} \lambda\times s_2 + \theta\left[\frac{1}{(\lambda+\theta)^2} - \frac{\theta}{(\lambda+\theta)(\lambda+2\theta)}\right],
\end{aligned}
$$

that is,

$$
\begin{aligned}
&\frac{1}{\lambda^2} - \frac{\theta}{\lambda(\lambda+\theta)} \\
&\overset{(A.2)}{=} s_1 \\
&= \lambda\times s_2 + \theta\left[\frac{1}{(\lambda+\theta)^2} - \frac{\theta}{(\lambda+\theta)(\lambda+2\theta)}\right],
\end{aligned}
$$

so that

$$
s_2 = \frac{1}{\lambda^3} - \frac{\theta}{\lambda^2(\lambda+\theta)} - \frac{\theta}{\lambda(\lambda+\theta)^2} + \frac{\theta^2}{\lambda(\lambda+\theta)(\lambda+2\theta)}. \tag{A.3}
$$

Based on (A.3) and (A.2), we can obtain

$$
\begin{aligned}
&E\left[\frac{1}{(\lambda+\theta Y)^2}\right] \\
&= \frac{\phi}{\lambda^2} + (1-\phi)\sum_{y=0}^{\infty}\frac{1}{(\lambda+\theta y)^2}\cdot\frac{\lambda(\lambda+\theta y)^{y-1}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\
&= \frac{\phi}{\lambda^2} + \lambda(1-\phi)\sum_{y=0}^{\infty}\frac{(\lambda+\theta y)^{y-3}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\
&\overset{(A.3)}{=} \frac{\phi}{\lambda^2} + \lambda(1-\phi)\left[\frac{1}{\lambda^3} - \frac{\theta}{\lambda^2(\lambda+\theta)}\right. \\
&\qquad \left. - \frac{\theta}{\lambda(\lambda+\theta)^2} + \frac{\theta^2}{\lambda(\lambda+\theta)(\lambda+2\theta)}\right] \\
&= \frac{1}{\lambda^2} - \frac{\theta(1-\phi)}{\lambda(\lambda+\theta)} - \frac{\theta(1-\phi)}{(\lambda+\theta)^2} + \frac{\theta^2(1-\phi)}{(\lambda+\theta)(\lambda+2\theta)}, \qquad \text{(A.4)}
\end{aligned}
$$

and

$$
\begin{aligned}
&E\left[\frac{Y}{(\lambda+\theta Y)^2}\right] \\
&= (1-\phi)\sum_{y=0}^{\infty}\frac{y}{(\lambda+\theta y)^2}\cdot\frac{\lambda(\lambda+\theta y)^{y-1}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\
&= (1-\phi)\sum_{y=0}^{\infty}\frac{y\lambda(\lambda+\theta y)^{y-3}\mathrm{e}^{-(\lambda+\theta y)}}{y!} \\
&= (1-\phi)\lambda\sum_{y=1}^{\infty}\frac{[\lambda+\theta+\theta(y-1)]^{y-1-2}\mathrm{e}^{-[\lambda+\theta+\theta(y-1)]}}{(y-1)!} \\
&= (1-\phi)\lambda\sum_{y=0}^{\infty}\frac{(\lambda+\theta+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta+\theta y)}}{y!} \\
&\qquad [\text{Let } \lambda+\theta=\lambda^*] \\
&\overset{(A.2)}{=} (1-\phi)\lambda\cdot\left[\frac{1}{(\lambda+\theta)^2} - \frac{\theta}{(\lambda+\theta)(\lambda+2\theta)}\right] \\
&\text{(A.5)} \\
&= \frac{\lambda(1-\phi)}{(\lambda+\theta)^2} - \frac{\lambda\theta(1-\phi)}{(\lambda+\theta)(\lambda+2\theta)}.
\end{aligned}
$$

By combining (A.4) with (A.5), we immediately obtain the first formula of (A.1). Next,

$$
E\left[\frac{Y^2}{(\lambda+\theta Y)^2}\right]
$$

$$
\begin{aligned}
&= (1-\phi)\sum_{y=0}^{\infty}\frac{y^2}{(\lambda+\theta y)^2}\cdot\frac{\lambda(\lambda+\theta y)^{y-1}\mathrm{e}^{-(\lambda+\theta y)}}{y!}\\
&= (1-\phi)\lambda\sum_{y=1}^{\infty}\frac{y[\lambda+\theta+\theta(y-1)]^{y-1-2}\mathrm{e}^{-[\lambda+\theta+\theta(y-1)]}}{(y-1)!}\\
&= (1-\phi)\lambda\sum_{y=0}^{\infty}\frac{(y+1)(\lambda+\theta+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta+\theta y)}}{y!}\\
&= (1-\phi)\lambda\sum_{y=0}^{\infty}\frac{y(\lambda+\theta+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta+\theta y)}}{y!}\\
&\quad+\ (1-\phi)\lambda\sum_{y=0}^{\infty}\frac{(\lambda+\theta+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta+\theta y)}}{y!}\\
&= (1-\phi)\lambda\sum_{y=1}^{\infty}\frac{[\lambda+2\theta+\theta(y-1)]^{y-1-1}\mathrm{e}^{-[\lambda+2\theta+\theta(y-1)]}}{(y-1)!}\\
&\quad+\ (1-\phi)\lambda\left[\frac{1}{(\lambda+\theta)^2}-\frac{\theta}{(\lambda+\theta)(\lambda+2\theta)}\right]\\
&= \frac{\lambda(1-\phi)}{\lambda+2\theta}+\frac{\lambda(1-\phi)}{(\lambda+\theta)^2}-\frac{\lambda\theta(1-\phi)}{(\lambda+\theta)(\lambda+2\theta)},
\end{aligned}
$$

(A.6)

By combining (A.5) with (A.6), we immediately obtain the third formula of (A.1). To obtain $E[Y^3/(\lambda+\theta Y)^2]$, we need to calculate

$$
c_1=\sum_{y=0}^{\infty}\frac{y^2(\lambda+\theta+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta+\theta y)}}{y!}
$$

and

$$
c_2=\sum_{y=0}^{\infty}\frac{y(\lambda+\theta+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta+\theta y)}}{y!}.
$$

In fact,

$$
\begin{aligned}
c_1 &= \sum_{y=0}^{\infty}\frac{y^2(\lambda+\theta+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta+\theta y)}}{y!}\\
&= \sum_{y=1}^{\infty}\frac{y[\lambda+2\theta+\theta(y-1)]^{y-1-1}\mathrm{e}^{-[\lambda+2\theta+\theta(y-1)]}}{(y-1)!}\\
&= \sum_{y=0}^{\infty}\frac{(y+1)(\lambda+2\theta+\theta y)^{y-1}\mathrm{e}^{-(\lambda+2\theta+\theta y)}}{y!}\\
&= \sum_{y=0}^{\infty}\frac{y(\lambda+2\theta+\theta y)^{y-1}\mathrm{e}^{-(\lambda+2\theta+\theta y)}}{y!}\\
&\quad+\sum_{y=0}^{\infty}\frac{(\lambda+2\theta+\theta y)^{y-1}\mathrm{e}^{-(\lambda+2\theta+\theta y)}}{y!}\\
&= \sum_{y=1}^{\infty}\frac{[\lambda+3\theta+\theta(y-1)]^{y-1}\mathrm{e}^{-[\lambda+3\theta+\theta(y-1)]}}{(y-1)!}+\frac{1}{\lambda+2\theta}\\
&= \frac{1}{1-\theta}+\frac{1}{\lambda+2\theta},
\end{aligned}
$$

and

$$
\begin{aligned}
c_2 &= \sum_{y=0}^{\infty}\frac{y(\lambda+\theta+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta+\theta y)}}{y!}\\
&= \sum_{y=1}^{\infty}\frac{[\lambda+2\theta+\theta(y-1)]^{y-1-1}\mathrm{e}^{-[\lambda+2\theta+\theta(y-1)]}}{(y-1)!}\\
&= \sum_{y=0}^{\infty}\frac{(\lambda+2\theta+\theta y)^{y-1}\mathrm{e}^{-(\lambda+2\theta+\theta y)}}{y!}\\
&= \frac{1}{\lambda+2\theta}.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&E\left[\frac{Y^3}{(\lambda+\theta Y)^2}\right]\\
&= (1-\phi)\sum_{y=0}^{\infty}\frac{y^3}{(\lambda+\theta y)^2}\cdot\frac{\lambda(\lambda+\theta y)^{y-1}\mathrm{e}^{-(\lambda+\theta y)}}{y!}\\
&= (1-\phi)\lambda\sum_{y=0}^{\infty}\frac{y^3(\lambda+\theta y)^{y-3}\mathrm{e}^{-(\lambda+\theta y)}}{y!}\\
&= (1-\phi)\lambda\sum_{y=1}^{\infty}\frac{y^2[\lambda+\theta+\theta(y-1)]^{y-1-2}\mathrm{e}^{-[\lambda+\theta+\theta(y-1)]}}{(y-1)!}\\
&= (1-\phi)\lambda\sum_{y=0}^{\infty}\frac{(y+1)^2(\lambda+\theta+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta+\theta y)}}{y!}\\
&= (1-\phi)\lambda\sum_{y=0}^{\infty}\frac{y^2(\lambda+\theta+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta+\theta y)}}{y!}\\
&\quad+2(1-\phi)\lambda\sum_{y=0}^{\infty}\frac{y(\lambda+\theta+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta+\theta y)}}{y!}\\
&\quad+(1-\phi)\lambda\sum_{y=0}^{\infty}\frac{(\lambda+\theta+\theta y)^{y-2}\mathrm{e}^{-(\lambda+\theta+\theta y)}}{y!}\\
&= (1-\phi)\lambda\cdot c_1+2(1-\phi)\lambda\cdot c_2\\
&\quad+(1-\phi)\lambda\cdot\left[\frac{1}{(\lambda+\theta)^2}-\frac{\theta}{(\lambda+\theta)(\lambda+2\theta)}\right],\\
&= \lambda(1-\phi)\cdot\left[\frac{1}{1-\theta}+\frac{1}{\lambda+2\theta}\right]+\frac{2\lambda(1-\phi)}{\lambda+2\theta}\\
&\quad+\lambda(1-\phi)\cdot\left[\frac{1}{(\lambda+\theta)^2}-\frac{\theta}{(\lambda+\theta)(\lambda+2\theta)}\right]\\
&= \frac{\lambda(1-\phi)}{1-\theta}+\frac{3\lambda(1-\phi)}{\lambda+2\theta}+\frac{\lambda(1-\phi)}{(\lambda+\theta)^2}-\frac{\lambda\theta(1-\phi)}{(\lambda+\theta)(\lambda+2\theta)}.
\end{aligned}
$$

By combining (A.6) with the above formula, we immediately obtain the second formula of (A.1).

Appendix B: The derivation of (3.9)

From (2.16), since $X_i|(\mathbf{y}=\boldsymbol{y}\neq\mathbf{0})\sim\text{Degenerate}(y_i)$ for $i=1,\ldots,m$, we obtain that when $\mathbf{y}=\boldsymbol{y}\neq\mathbf{0}$,

$$E\left(\frac{X_{ij}^2 - X_{ij}}{\lambda_i + \theta_i X_{ij}}\middle| Y_{\text{obs}}, \phi, \boldsymbol{\lambda}, \boldsymbol{\theta}\right) = \frac{y_{ij}^2 - y_{ij}}{\lambda_i + \theta_i y_{ij}}.$$

From (2.15), we know that when $\mathbf{y} = \boldsymbol{y} = \mathbf{0}$, $X_i|(\mathbf{y} = \mathbf{0}) \sim \text{ZIGP}(\psi; \lambda_i, \theta_i)$. To obtain (3.9), we assume that $X \sim \text{ZIGP}(\psi, \lambda, \theta)$ and only need to prove

$$E\left(\frac{X^2 - X}{\lambda + \theta X}\right) = \frac{(1-\psi)\lambda}{1-\theta}. \tag{B.1}$$

Note that

$$\begin{aligned}
&E\left(\frac{X}{\lambda + \theta X}\right)\\
= \;& (1-\psi)\sum_{x=0}^{\infty} \frac{x}{\lambda + \theta x} \cdot \frac{\lambda(\lambda + \theta x)^{x-1} \mathrm{e}^{-(\lambda + \theta x)}}{x!}\\
= \;& (1-\psi)\lambda \sum_{x=1}^{\infty} \frac{[\lambda + \theta + \theta(x-1)]^{x-1-1} \mathrm{e}^{-[\lambda+\theta+\theta(x-1)]}}{(x-1)!}\\
= \;& (1-\psi)\lambda \sum_{x=0}^{\infty} \frac{(\lambda + \theta + \theta x)^{y-1} \mathrm{e}^{-(\lambda+\theta+\theta x)}}{x!}\\
= \;& \frac{(1-\psi)\lambda}{\lambda + \theta},
\end{aligned}$$

and

$$\begin{aligned}
&E\left(\frac{X^2}{\lambda + \theta X}\right)\\
= \;& (1-\psi)\sum_{x=0}^{\infty} \frac{x^2}{\lambda + \theta x} \cdot \frac{\lambda(\lambda + \theta x)^{x-1} \mathrm{e}^{-(\lambda + \theta x)}}{x!}\\
= \;& (1-\psi)\lambda \sum_{x=1}^{\infty} \frac{x[\lambda + \theta + \theta(x-1)]^{x-1-1} \mathrm{e}^{-[\lambda+\theta+\theta(x-1)]}}{(x-1)!}\\
= \;& (1-\psi)\lambda \sum_{x=0}^{\infty} \frac{(x+1)(\lambda + \theta + \theta x)^{x-1} \mathrm{e}^{-(\lambda+\theta+\theta x)}}{x!}\\
= \;& (1-\psi)\lambda \sum_{x=0}^{\infty} \frac{x(\lambda + \theta + \theta x)^{x-1} \mathrm{e}^{-(\lambda+\theta+\theta x)}}{x!}\\
& + \frac{(1-\psi)\lambda}{\lambda+\theta}\\
= \;& (1-\psi)\lambda \sum_{x=1}^{\infty} \frac{[\lambda + 2\theta + \theta(x-1)]^{x-1} \mathrm{e}^{-[\lambda+2\theta+\theta(x-1)]}}{(x-1)!}\\
& + \frac{(1-\psi)\lambda}{\lambda+\theta}\\
= \;& (1-\psi)\lambda \sum_{x=0}^{\infty} \frac{(\lambda + 2\theta + \theta x)^{x} \mathrm{e}^{-(\lambda+\theta+\theta x)}}{x!} + \frac{(1-\psi)\lambda}{\lambda+\theta}\\
= \;& \lambda(1-\psi)(\lambda+2\theta) \sum_{x=0}^{\infty} \frac{(\lambda + 2\theta + \theta x)^{x-1} \mathrm{e}^{-(\lambda+2\theta+\theta x)}}{x!}\\
& + \frac{\lambda\theta(1-\psi)}{1-\theta} + \frac{(1-\psi)\lambda}{\lambda+\theta}\\
= \;& \lambda(1-\psi) + \frac{(1-\psi)\lambda}{\lambda+\theta} + \frac{\lambda\theta(1-\psi)}{1-\theta}\\
= \;& \frac{(1-\psi)\lambda}{\lambda+\theta} + \frac{\lambda(1-\psi)}{1-\theta}.
\end{aligned}$$

By combining the two formulae, we obtain (B.1).

Appendix C: The score vector and Fisher information matrix in Section 3.5.2

The elements in the score vector $U(\gamma, \boldsymbol{\lambda}, \boldsymbol{\theta})$ and observed information matrix $\mathbf{I}(\gamma, \boldsymbol{\lambda}, \boldsymbol{\theta}|Y_{\text{obs}})$ are

$$\begin{aligned}
\frac{\partial \ell^*}{\partial \gamma} &= -\frac{n}{1+\gamma} + \frac{n_0}{\gamma + \mathrm{e}^{-\lambda_+}},\\
\frac{\partial \ell^*}{\partial \lambda_i} &= -\frac{n\mathrm{e}^{-\lambda_+} + (n-n_0)\gamma}{\gamma + \mathrm{e}^{-\lambda_+}}\\
&\quad + \sum_{j=1}^{n}\left(\frac{1}{\lambda_i} + \frac{y_{ij}-1}{\lambda_i + \theta_i y_{ij}}\right),\\
\frac{\partial \ell^*}{\partial \theta_i} &= \sum_{j=1}^{n}\left[\frac{y_{ij}(y_{ij}-1)}{\lambda_i + \theta_i y_{ij}} - y_{ij}\right],\\
\frac{\partial^2 \ell^*}{\partial \gamma^2} &= \frac{n}{(1+\gamma)^2} - \frac{n_0}{(\gamma + \mathrm{e}^{-\lambda_+})^2},\\
\frac{\partial^2 \ell^*}{\partial \lambda_i^2} &= \frac{n_0\gamma\mathrm{e}^{-\lambda_+}}{(\gamma + \mathrm{e}^{-\lambda_+})^2}\\
&\quad - \sum_{j=1}^{n}\left[\frac{1}{\lambda_i^2} + \frac{y_{ij}-1}{(\lambda_i + \theta_i y_{ij})^2}\right],\\
\frac{\partial^2 \ell^*}{\partial \theta_i^2} &= -\sum_{j=1}^{n} \frac{y_{ij}^2(y_{ij}-1)}{(\lambda_i + \theta_i y_{ij})^2},\\
\frac{\partial^2 \ell^*}{\partial \lambda_i \partial \gamma} &= \frac{n_0\mathrm{e}^{-\lambda_+}}{(\gamma + \mathrm{e}^{-\lambda_+})^2},\\
\frac{\partial^2 \ell^*}{\partial \lambda_i \partial \lambda_k} &= \frac{n_0\gamma\mathrm{e}^{-\lambda_+}}{(\gamma + \mathrm{e}^{-\lambda_+})^2},\\
\frac{\partial^2 \ell^*}{\partial \lambda_i \partial \theta_i} &= \sum_{j=1}^{n} -\frac{y_{ij}(y_{ij}-1)}{(\lambda_i + \theta_i y_{ij})^2},\\
\frac{\partial^2 \ell^*}{\partial \lambda_i \partial \theta_k} &= \frac{\partial^2 \ell^*}{\partial \theta_i \partial \theta_k} = \frac{\partial^2 \ell^*}{\partial \theta_i \partial \gamma} = 0,
\end{aligned}$$

for $i, k = 1, \ldots, m$ and $i \neq k$. By using (3.2), we can calculate the Fisher information matrix $\mathbf{J}(\gamma, \boldsymbol{\lambda}, \boldsymbol{\theta})$, whose elements are given by

$$\begin{aligned}
J_{11} &= -E\left(\frac{\partial^2 \ell^*}{\partial \gamma^2}\right) \\
&= -\frac{n}{(1+\gamma)^2} + \frac{n}{(1+\gamma)(\gamma+\mathrm{e}^{-\lambda_+})}, \\
J_{i+1,\ i+1} &= -E\left(\frac{\partial^2 \ell^*}{\partial \lambda_i^2}\right) \\
&= -\frac{n\gamma \mathrm{e}^{-\lambda_+}}{(1+\gamma)(\gamma+\mathrm{e}^{-\lambda_+})} + \frac{n}{(1+\gamma)(\lambda_i+\theta_i)} \\
&\quad - \frac{n\theta_i}{(1+\gamma)(\lambda_i+2\theta_i)} + \frac{n\theta_i}{\lambda_i(\lambda_i+\theta_i)(1+\gamma)},
\end{aligned}$$

$$\begin{aligned}
J_{i+m+1,\ i+m+1} &= -E\left(\frac{\partial^2 \ell^*}{\partial \theta_i^2}\right) \\
&= \frac{n\lambda_i}{(1-\theta_i)(1+\gamma)} + \frac{2n\lambda_i}{(\lambda_i+2\theta_i)(1+\gamma)}, \\
J_{1,\ i+1} &= -E\left(\frac{\partial^2 \ell^*}{\partial \gamma \partial \lambda_i}\right) = -\frac{n\mathrm{e}^{-\lambda_+}}{(1+\gamma)(\gamma+\mathrm{e}^{-\lambda_+})}, \\
J_{1,\ i+m+1} &= -E\left(\frac{\partial^2 \ell^*}{\partial \gamma \partial \theta_i}\right) = 0, \\
J_{i+1,\ k+1} &= -E\left(\frac{\partial^2 \ell^*}{\partial \lambda_i \partial \lambda_k}\right) = -\frac{n\gamma \mathrm{e}^{-\lambda_+}}{(1+\gamma)(\gamma+\mathrm{e}^{-\lambda_+})}, \\
J_{i+1,\ k+m+1} &= -E\left(\frac{\partial^2 \ell^*}{\partial \lambda_i \partial \theta_k}\right) = 0, \\
J_{i+m+1,\ k+m+1} &= -E\left(\frac{\partial^2 \ell^*}{\partial \theta_i \partial \theta_k}\right) = 0, \\
J_{i+1,\ i+m+1} &= -E\left(\frac{\partial^2 \ell^*}{\partial \lambda_i \partial \theta_i}\right) = \frac{n\lambda_i}{(\lambda_i+2\theta_i)(1+\gamma)},
\end{aligned}$$

for $i, k = 1, \ldots, m$ and $i \neq k$.

Received 21 May 2015

REFERENCES

[1] Consul, P. C. and Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics* **15**(4), 791–799. MR0391332

[2] Consul, P. C. and Shoukri, M. M. (1985). The generalized Poisson distribution when the sample mean is larger than the sample variance. *Communications in Statistics – Simulation and Computation* **14**(3), 667–681. MR0821106

[3] Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J. and Knuth, D. E. (1996). On the Lambert W function. *Advances in Computational Mathematics* **5**(1), 329–359. MR1414285

[4] Famoye, F. and Singh, K. P. (2006). Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science* **4**, 117–130.

[5] Gupta, P. L., Gupta, R. C. and Tripathi, R. C. (1996). Analysis of zero-adjusted count data. *Computational Statistics and Data Analysis* **23**, 207–218.

[6] Gupta, P. L., Gupta, R. C. and Tripathi, R. C. (2004). Score test for zero inflated generalized Poisson regression model. *Communications in Statistics – Theory and Methods* **33**(1), 47–64. MR2040552

[7] Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics* **56**, 1030–1039. MR1815581

[8] Heibron, D. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal* **36**, 531–547.

[9] Jansakul, N. and Hinde, J. P. (2002). Score tests for zero-inflated Poisson models. *Computational Statistics & Data Analysis* **40**, 75–96. MR1921123

[10] Joe, H. and Zhu, R. (2005). Generalized Poisson distribution: The property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal* **47**(2), 219–229. MR2137236

[11] Jung, R. C. and Winkelmann, R. (1993). Two aspects of labor mobility: A bivariate Poisson regression approach. *Empirical Economics* **18**(3), 543–556.

[12] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**(1), 1–14.

[13] Lee, A. H., Wang, K. and Yau, K. K. W. (2001). Analysis of zero-inflated Poisson data incorporating extent of exposure. *Biometrical Journal* **43**, 963–975. MR1878270

[14] Li, C. S. (2012). Score tests for semiparametric zero-inflated Poisson models. *International Journal of Statistics and Probability* **1**(2), 1–7.

[15] Liu, Y. and Tian, G. L. (2015). Type I multivariate zero-inflated Poisson distribution with applications. *Computational Statistics and Data Analysis* **83**, 200–222. MR3281806

[16] Minami, M., Lennert-Cody, C. E., Gao, W. and Román-Verdesoto, M. (2007). Modeling shark bycatch: The zero-inflated negative binomial regression model with smoothing. *Fisheries Research* **84**(2), 210–221.

[17] Mullahy, J. (1986). Specification and testing of some modified count models. *Journal of Econometrics* **33**, 341–365. MR0867980

[18] Strauss, J., Beegle, K., Sikoki, B., Dawiyanto, A., Herawati, Y. and Witoelar, Y. (2004). *The Third Wave of the Indonesia Family Life Survey (IFLS): Overview and Field Report, WR-144/1-NIA/NICHD*, RAND Corporation, Santa Monica, CA.

[19] Vernic, R. (1997). On the bivariate generalized Poisson distribution. *Astin Bulletin* **27**(1), 23–31.

[20] Vernic, R. (2000). A multivariate generalization of the generalized Poisson distribution. *Astin Bulletin* **30**(1), 57–67. MR1945547

[21] Xie, F. C. and Wei, B. C. (2010). Influence analysis for count data based on generalized Poisson regression models. *Statistics* **44**(4), 341–360. MR2739528

[22] Xie, F. C., Lin, J. G. and Wei, B. C. (2014). Bayesian zero-inflated generalized Poisson regression model: Estimation and case influence diagnostics. *Journal of Applied Statistics* **41**(6), 1383–1392. MR3268901

Xi-Fen Huang
Department of Statistics and Actuarial Science
The University of Hong Kong, Pokfulam Road
Hong Kong
P. R. China
E-mail address: u3003396@hku.hk

Guo-Liang Tian
Department of Mathematics
South University of Science and Technology of China
Shenzhen, Guangdong
P. R. China
E-mail address: tiangl@sustc.edu.cn

Chi Zhang
Department of Statistics and Actuarial Science
The University of Hong Kong, Pokfulam Road
Hong Kong
P. R. China
E-mail address: zhangchi@hku.hk

Xuejun Jiang
Department of Mathematics
South University of Science and Technology of China
Shenzhen, Guangdong
P. R. China
E-mail address: jiangxj@sustc.edu.cn