



**EM estimation for bivariate mixed poisson INAR(1) claim count regression models with correlated random effects**

**LSE Research Online URL for this paper:** <http://eprints.lse.ac.uk/118826/>

Version: Published Version

---

**Article:**

Chen, Zezhun Chen, Dassios, Angelos ORCID: 0000-0002-3968-2366 and Tzougas, George (2023) EM estimation for bivariate mixed poisson INAR(1) claim count regression models with correlated random effects. *European Actuarial Journal*. ISSN 2190-9733

<https://doi.org/10.1007/s13385-023-00351-7>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>



# EM estimation for bivariate mixed poisson INAR(1) claim count regression models with correlated random effects

Zezhun Chen<sup>1</sup> · Angelos Dassios<sup>1</sup> · George Tzougas<sup>2</sup>

Received: 13 September 2022 / Revised: 6 February 2023 / Accepted: 17 April 2023  
© The Author(s) 2023

## Abstract

This article considers bivariate mixed Poisson INAR(1) regression models with correlated random effects for modelling correlations of different signs and magnitude among time series of different types of claim counts. This is the first time that the proposed family of INAR(1) models is used in a statistical or actuarial context. For expository purposes, the bivariate mixed Poisson INAR(1) claim count regression models with correlated Lognormal and Gamma random effects paired via a Gaussian copula are presented as competitive alternatives to the classical bivariate Negative Binomial INAR(1) claim count regression model which only allows for positive dependence between the time series of claim count responses. Our main achievement is that we develop novel alternative Expectation-Maximization type algorithms for maximum likelihood estimation of the parameters of the models which are demonstrated to perform satisfactory when the models are fitted to Local Government Property Insurance Fund data from the state of Wisconsin.

**Keywords** Count data time series · Binomial-mixed Poisson INAR(1) regression models with correlated random effects · Overdispersion · Gaussian copula · Correlations of different signs and magnitude

---

Angelos Dassios and George Tzougas have contributed equally to this work.

---

✉ Zezhun Chen  
z.chen58@lse.ac.uk

Angelos Dassios  
a.dassios@lse.ac.uk

George Tzougas  
george.tzougas@hw.ac.uk

<sup>1</sup> Department of Statistics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, England, UK

<sup>2</sup> Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Campus The Avenue, Edinburgh EH14 4AS, Scotland, UK

## 1 Introduction

Over the past decade, there has been a growing literature on bivariate (and/or multivariate) claim count regression models which can efficiently capture the dependence between claims from the same policy and/or different coverages bundled into a single policy. The interested reader is referred to Refs. [1–7, 9, 11, 13, 15, 23–25, 28, 29], among many others.

Pechon et al. [23] proposed the use of bivariate mixed Poisson count regression models, with correlated random effects for capturing the interactions between the different coverages purchased by members of the same household. In particular, Pechon et al. [23] considered the bivariate Poisson-Gamma (BPGGA) regression model with Gaussian copula and the bivariate Poisson-Lognormal (BPLN) regression model. In the former model the random effects are distributed according to two Gamma distributions with unit means and the dependence between the random effects is introduced by means of a Gaussian bivariate copula whereas in the latter model these random effects are distributed according to the bivariate Lognormal mixing distribution. Bermúdez et al. [4], following the setup of Pedeli and Karlis ([26] and [27]), were the first to derive a bivariate Poisson integer-valued autoregressive process of order 1 (BINAR(1)) claim count regression model which can account both for cross-sectional and temporal dependence between multiple claim types. The model they developed was employed for addressing the ratemaking problem of pricing an insurance contract in the case of positively correlated claims from different types of coverage in non-life insurance. Finally, Bermúdez and Karlis [5] built on the previous paper by using a multivariate INAR(1) (MINAR(1)) regression model based on the Sarmanov family of distributions. The MINAR(1) regression models based on the Sarmanov family of distributions are also restricted to a positive correlation structure between the claim count response variables. However, it enjoys some advantages compared to a different approach which can allow for both positive and negative correlations by using copulas for the specification of the joint distribution of the innovations. See, for instance, Refs. [8, 17, 18, 20–22] among others. Firstly, it avoids identifiability issues which may arise when a continuous copula distribution is paired with discrete marginals, see Ref. [12]. As is well known, the lack of identifiability means that it cannot be guaranteed that model fitting is unique and this may lead to problems in statistical inference, for example, one might receive no meaningful values for the standard errors of the parameters. Secondly, the computational intensity for discrete copula-based models increases as the dimension of the model increases and hence, as is also mentioned by the authors, their approach, which relies on the use of the Sarmanov family, provides models that are less computationally intensive to estimate and can still have a reasonable range for positive dependence structure between the claim count responses.

In this study, we introduce a family of bivariate mixed Poisson INAR(1) claim count regression models with correlated random effects for modelling the dependence structure between times series of different types of claim counts from the same and/or different types of coverage. The bivariate mixed Poisson INAR(1) regression models with correlated random effects are a broad class of models which can accommodate overdispersion, which is a direct consequence of unobserved heterogeneity due to systematic effects in the data, and correlations of different signs and magnitude.

For demonstration purposes, we consider the bivariate mixed Poisson INAR(1) claim count regression models which are derived by using the bivariate Lognormal and Gaussian copula paired with gamma marginals as mixing densities, which we refer to as BINAR(1)-LN and BINAR(1)-GGA claim count regression models respectively. Both models can be regarded as extensions of the classical bivariate Negative Binomial INAR(1) claim count regression model with a shared gamma random effect, which we refer to as BINAR(1)-GA claim count regression model, in the sense that they provide more flexibility for modelling overdispersed bivariate time series of count data compared to the BINAR(1)-GA model which is derived by pre-imposing the restrictive positive correlation assumption between time series of different claim types of claim counts, since in some cases negative correlations may be of interest as well. Furthermore, unlike previous copula-based count regression models for which identifiability issues can arise when a continuous copula distribution is paired with discrete marginals, in the proposed family of models identifiability of the bivariate distribution of the innovations is guaranteed by imposing a unit mean constraint for the Gamma continuous mixing densities which are paired with a Gaussian copula.

The main contributions we make are as follows:

- Firstly, before we introduce the time series components, we present a unified framework for statistical inference via the Expectation-Maximization (EM) algorithm for the BPGA, BPLN and BPGGA regression models.<sup>1</sup>
- Secondly, we develop novel EM type algorithms for maximum likelihood (ML) estimation of the BINAR(1)-GA, BINAR(1)-LN and BINAR(1)-GGA regression models, which has not been explored in the literature so far. The main reason for this is because the joint distribution of the innovations cannot be written in closed form in either model and hence its maximization is not possible via standard numerical optimization as is done in Refs. [4, 5, 17, 26, 27].

The rest of the paper is organized as follows. Section 2 presents the model specifications for the bivariate mixed Poisson regression models we consider and describes their ML estimation via the EM algorithm. Section 3 presents the derivation of their INAR(1) extensions that we first proposed herein and outlines the EM type algorithms we developed for statistical inference. Section 4 presents our empirical analysis which is based on the LGPIF dataset. Concluding remarks are given in Sect. 5. The interpretation of abbreviations used in the paper and some other technical details are provided in Appendix A.1.

## 2 The bivariate mixed Poisson regression model

### 2.1 Model specifications

The bivariate Poisson mixture is constructed by two independent Poisson random variables conditional on a random effect vector (or scalar)  $\theta = (\theta_1, \theta_2)$  such that  $N^{(i)} \sim$

<sup>1</sup> Note that EM estimation for the parameters of the BPGA regression model with a shared random effect and the BPLN regression model has been discussed in Refs. [14] and [30] respectively. However, this is the first time that the EM algorithm is used for estimating the parameters of the BPGA regression model with Gaussian copula.

$Pois(\lambda_{i,t}\theta_i)$ ,  $i = 1, 2$ . The bivariate mixed Poisson regression is then constructed by further allowing the rate  $\lambda_i$  to be modelled as functions of explanatory variables  $\mathbf{z}_{i,t}$  such that  $\lambda_{i,t} = \exp\{\mathbf{z}_{i,t}^T \boldsymbol{\beta}_i\}$ . Denote the mixing density function of the random effect as  $f_\phi(\boldsymbol{\theta})$  parametrized by  $\boldsymbol{\phi}$ . To avoid the identifiability issue, we have to restrict the expectation  $\mathbb{E}[\theta_i]$  to be a fixed constant. One usually lets  $\mathbb{E}[\theta_i] = 1$  so that  $\boldsymbol{\lambda}_t := (\lambda_{1,t}, \lambda_{2,t})$  will fully explain the frequency of a event and  $\boldsymbol{\phi}$  will explain the variation and correlation of the whole bivariate sequence. In the following, we will discuss three different mixing densities, univariate gamma (shared random effect), bivariate Lognormal and Gaussian copula paired with Gamma marginals.

(a) **Univariate Gamma density**

In this case, the bivariate mixed Poisson regression model shares the same random effect  $N_t^{(i)} \sim Pois(\lambda_{i,t}\theta)$   $i = 1, 2$ . Denote the mixing density function as  $f_\phi(\theta) = f_\phi(\boldsymbol{\theta})$  and it has following expression

$$f_\phi(\theta) = \frac{\phi^\phi}{\Gamma(\phi)} \theta^{\phi-1} e^{-\phi\theta}, \tag{1}$$

which has unit mean and variance  $\frac{1}{\phi}$ . Then the unconditional probability mass function  $f_{PG}(\mathbf{k}, t)$  of  $\mathbf{N}_t := (N_t^{(1)}, N_t^{(2)})$  can be written down in a closed form

$$\begin{aligned} f_{PG}(\mathbf{k}, t) &= \frac{\lambda_{1,t}}{k_1!} \frac{\lambda_{2,t}}{k_2!} \int_0^\infty e^{-(\lambda_{1,t} + \lambda_{2,t})\theta} \theta^{k_1+k_2} f_\phi(\theta) d\theta \\ &= \frac{\Gamma(\phi + k_1 + k_2)}{\Gamma(\phi)\Gamma(k_1 + 1)\Gamma(k_2 + 1)} \frac{\phi^\phi \lambda_{1,t}^{k_1} \lambda_{2,t}^{k_2}}{(\phi + \lambda_{1,t} + \lambda_{2,t})^{\phi+k_1+k_2}}. \end{aligned} \tag{2}$$

(b) **Bivariate Lognormal density**

Suppose now the random vector  $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2)$  follows bivariate normal distribution, with mean vector  $(-\frac{\phi_1^2}{2}, -\frac{\phi_2^2}{2})$  and covariance matrix  $\Sigma$

$$\Sigma = \begin{bmatrix} \phi_1^2 & \rho\phi_1\phi_2 \\ \rho\phi_1\phi_2 & \phi_2^2 \end{bmatrix} \tag{3}$$

Then the random effect vector  $\boldsymbol{\theta} = e^\boldsymbol{\epsilon} = (e^{\epsilon_1}, e^{\epsilon_2})$  has Lognormal distribution with unit mean. Denote the density function of  $\boldsymbol{\epsilon}$  as  $f_\Sigma^N$  and  $f_\Sigma^{LN}$  for Lognormal density. Then they have the following expressions

$$\begin{aligned} f_\Sigma^N(\boldsymbol{\epsilon}) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{\epsilon_1 + 0.5\sigma_1^2}{\sigma_1} \right)^2 \right. \right. \\ &\quad \left. \left. - 2\rho \left( \frac{\epsilon_1 + 0.5\sigma_1^2}{\sigma_1} \right) \left( \frac{\epsilon_2 + 0.5\sigma_2^2}{\sigma_2} \right) \right] \right\} \end{aligned}$$

$$f_{\phi}(\boldsymbol{\theta}) = \frac{1}{\theta_1 \theta_2} f_{\Sigma}^N(\log \boldsymbol{\theta}) = f_{\Sigma}^{LN}(\boldsymbol{\theta}).$$

The unconditional distribution  $f_{PLN}(\mathbf{k}, t)$  of  $\mathbf{N}_t$  is expressed as a double integral

$$\begin{aligned} f_{PLN}(\mathbf{k}, t) &= \int_0^{\infty} \int_0^{\infty} \frac{\lambda_{1,t}^{k_1}}{k_1!} \frac{\lambda_{2,t}^{k_2}}{k_2!} e^{-\lambda_{1,t}\theta_1} e^{-\lambda_{2,t}\theta_2} \theta_1^{k_1} \theta_2^{k_2} f_{\Sigma}^{LN}(\boldsymbol{\theta}) d\theta_1 d\theta_2 \\ &= \int_R \int_R \frac{\lambda_{1,t}^{k_1}}{k_1!} \frac{\lambda_{2,t}^{k_2}}{k_2!} \exp\{-\lambda_{1,t}e^{\epsilon_1} - \lambda_{2,t}e^{\epsilon_2} + k_1\epsilon_1 + k_2\epsilon_2\} \\ &\quad f_{\Sigma}^N(\boldsymbol{\epsilon}) d\epsilon_1 d\epsilon_2. \end{aligned} \quad (4)$$

All the double integrals with respect to Lognormal density  $f_{\Sigma}^{LN}$  can be transformed into double integrals with respect to normal density  $f_{\Sigma}^N$  so that they can be evaluated by Gauss-Hermite quadrature. See details in the Appendix A.2.

### (c) Gaussian copula paired with Gamma marginals

Suppose now the random vector  $\boldsymbol{\theta}$  is distributed as a meta Gaussian copula such that its marginals are two independent Gamma random variables with parameter  $(\phi_1, \phi_2)$  respectively. Define uniform random vector  $\mathbf{u} = (F_{\phi_1}(\theta_1), F_{\phi_2}(\theta_2))$ . The distribution function  $F_{GC}(\boldsymbol{\theta})$  and density function  $f_{GC}(\boldsymbol{\theta})$  can be written as

$$\begin{aligned} F_{GC}(\boldsymbol{\theta}) &= C_{\rho}(\mathbf{u}) = F_{\rho}(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \\ f_{\phi}(\boldsymbol{\theta}) &= f_{GC}(\boldsymbol{\theta}) = \frac{f_{\rho}(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}{f_{sn}(\Phi^{-1}(u_1))f_{sn}(\Phi^{-1}(u_2))} f_{\phi_1}(\theta_1) f_{\phi_2}(\theta_2) \\ &:= c_{\rho}(\mathbf{u}) f_{\phi_1}(\theta_1) f_{\phi_2}(\theta_2), \end{aligned} \quad (5)$$

where  $f_{\rho}(\cdot, \cdot)$ ,  $F_{\rho}(\cdot, \cdot)$  are the density function and cumulative distribution of bivariate normal random variable with the following expression

$$f_{\rho}(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2} \frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{1-\rho^2}\right\}. \quad (6)$$

The  $\Phi(x)$  is the cdf of standard normal random variable with  $\Phi^{-1}(x)$  as its quantile function and  $f_{sn}(x)$  is the density function of the standard normal random variable. Finally,  $f_{\phi_i}(x)$  and  $F_{\phi_i}(x)$  are the pdf and cdf of Gamma density function defined in 1 for  $i = 1, 2$ . Then a bivariate Poisson Gamma random vector is constructed as  $N_t^{(i)} \sim Pois(\lambda_{i,t}\theta_i)$ ,  $i = 1, 2$  with probability mass function  $f_{PGC}(\mathbf{k}, t)$  such that

$$\begin{aligned}
 f_{PGC}(\mathbf{k}, t) &= \frac{\lambda_{1,t}^{k_1} \lambda_{2,t}^{k_2}}{k_1! k_2!} \int_0^\infty \int_0^\infty \exp\{-\lambda_{1,t}\theta_1 - \lambda_{2,t}\theta_2\} \theta_1^{k_1} \theta_2^{k_2} f_{GC}(\theta_1, \theta_2) d\theta_1 d\theta_2 \\
 &= \frac{\lambda_{1,t}^{k_1} \lambda_{2,t}^{k_2}}{k_1! k_2!} \int_0^1 \int_0^1 e^{-\lambda_{1,t}F_{\phi_1}^{-1}(u_1) - \lambda_{2,t}F_{\phi_2}^{-1}(u_2)} \\
 &\quad F_{\phi_1}^{-1}(u_1)^{k_1} F_{\phi_2}^{-1}(u_2)^{k_2} c_\rho(u_1, u_2) du_1 du_2.
 \end{aligned}$$

Then the double integral can be evaluated by Gauss-Legendre quadrature. See details in Appendix A.3.

### 2.2 The EM algorithm

For statistical inference of above model, the classical maximum likelihood estimation is not straightforward to apply because the log likelihood function

$$\ell(\Theta) = \sum_{t=1}^n \log \left( \frac{\lambda_{1,t}^{k_{1,t}} \lambda_{2,t}^{k_{2,t}}}{k_{1,t}! k_{2,t}!} \int_0^\infty \int_0^\infty \theta_1^{k_{1,t}} \theta_2^{k_{2,t}} e^{-\lambda_{1,t}\theta_1 - \lambda_{2,t}\theta_2} f_\phi(\theta) d\theta_1 d\theta_2 \right) \tag{7}$$

is not computational tractable and its maximum likelihood estimators are not straightforward to achieve. Alternatively, we can apply the EM algorithm to estimate the parameters  $\Theta = \{\beta_1, \beta_2, \phi\}$ . For given random samples  $(\mathbf{k}_1, \dots, \mathbf{k}_n)$ , suppose now we observe the random effect  $(\theta_1, \dots, \theta_n)$ , then the complete likelihood function  $\ell_c(\Theta)$  is given by

$$\ell_c(\Theta) = \sum_{t=1}^n \left[ \left( \sum_{i=1}^2 k_{i,t} \log(\lambda_{i,t}\theta_{i,t}) - \lambda_{i,t}\theta_{i,t} - \log(k_{i,t}!) \right) + \log f_\phi(\theta_t) \right]. \tag{8}$$

Compared to  $\ell(\Theta)$ , the complete log likelihood function  $\ell_c(\Theta)$  are simplified in the sense that there is no integration and mixture likelihood are decomposed into Poisson likelihood and the likelihood for mixing density.

However, to evaluate  $\ell_c(\Theta)$  we need to find out the conditional (posterior) distribution of  $\theta$  given the random samples. Then we define  $\eta(\theta|\lambda_t, \mathbf{k}_t) = e^{-\lambda_{1,t}\theta_1 - \lambda_{2,t}\theta_2} \theta_1^{k_{1,t}} \theta_2^{k_{2,t}}$  and posterior density

$$\begin{aligned}
 \pi(\theta|\Theta^{(j)}, \mathbf{k}_t) &= \frac{f_\phi(\theta) \prod_{i=1}^2 f_{Po}(k_{i,t}|\lambda_{i,t}\theta_i)}{\int_0^\infty \int_0^\infty f_\phi(\theta) \prod_{i=1}^2 f_{Po}^{(j)}(k_{i,t}|\lambda_{i,t}\theta_i) d\theta_1 d\theta_2} \\
 &= \frac{\eta(\theta|\lambda_t, \mathbf{k}_t) f_\phi(\theta)}{\int_0^\infty \int_0^\infty \eta(\theta|\lambda_t, \mathbf{k}_t) f_\phi(\theta) d\theta_1 d\theta_2}.
 \end{aligned} \tag{9}$$

Then posterior expectation for any real value function  $h(\theta)$  is given by

$$\begin{aligned} \mathbb{E}[h(\theta)|\Theta^{(j)}, \mathbf{k}_t] &= \int_0^\infty \int_0^\infty h(\theta)\pi(\theta|\Theta^{(j)}, \mathbf{k}_t)d\theta_1d\theta_2 \\ &=: \mathbb{E}_{\theta,t}^{(j)}[h(\theta)], \end{aligned} \tag{10}$$

where  $f_{Po}(k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$  is the probability mass function of a Poisson random variable with rate  $\lambda$  and the condition  $\Theta^{(j)}$  means that the posterior density function is evaluated with the parameters estimated at j-th iteration. The subscript  $\theta$  of  $\mathbb{E}_{\theta,t}^{(j)}$  means that the expectation is taken with respect to the  $\theta$  for t-th observation.

- **E-step:** Evaluating the Q function  $Q(\Theta; \Theta^{(j)})$  given the the parameters estimated at j-th iteration

$$\begin{aligned} Q(\Theta; \Theta^{(j)}) &\propto \sum_{t=1}^n \sum_{i=1}^2 k_{i,t} \log(\lambda_{i,t}) - \lambda_{i,t} \mathbb{E}[\theta_i|\Theta^{(j)}, \mathbf{k}_t] + \sum_{t=1}^n \mathbb{E}[\log f_\phi(\theta)|\Theta^{(j)}, \mathbf{k}_t] \\ &= \sum_{t=1}^n \sum_{i=1}^2 k_{i,t} \log(\lambda_{i,t}) - \lambda_{i,t} \mathbb{E}_{\theta,t}^{(j)}[\theta_i] + \sum_{t=1}^n \mathbb{E}_{\theta,t}^{(j)}[\log f_\phi(\theta)]. \end{aligned} \tag{11}$$

- **M-step:** After finding out the Q function, we update the parameters for the next iteration,  $\Theta^{(j+1)}$ , which can be achieved by finding the gradient functions  $g(\cdot)$  and the Hessian matrix  $H(\cdot)$  of Q functions and then apply the Newton-Raphson algorithm to maximize the Q function for the next iteration. The parameters can be updated separately as Poisson part  $\beta_1, \beta_2$  and random effect part  $\phi$ .

– For the Poisson part

$$\begin{aligned} \beta_i^{(j+1)} &= \beta_i^{(j)} - H^{-1}(\beta_i^{(j)})g(\beta_i^{(j)}), \quad i = 1, 2 \\ g(\beta_i^{(j)}) &= \mathbf{Z}_i^T \mathbf{V}_i^{(g)} \quad H(\beta_i^{(j)}) = \mathbf{Z}_i^T \mathbf{D}_i^{(H)} \mathbf{Z}_i \\ \mathbf{V}_i^{(g)} &= \left( \left\{ k_{i,t} - \lambda_{i,t} \mathbb{E}_{\theta,t}^{(j)}[\theta_i] \right\}_{t=1, \dots, n} \right) \\ \mathbf{D}_i^{(H)} &= \text{diag} \left( \left\{ -\lambda_{i,t} \mathbb{E}_{\theta,t}^{(j)}[\theta_i] \right\}_{t=1, \dots, n} \right) \end{aligned} \tag{12}$$

- For the random effect part, we need to derive the first and second order derivatives of  $\log f_\phi(\theta)$  and then the take posterior expectation to construct its gradient functions and Hessian matrix. In the following, we derive the derivatives for those three mixing densities defined in the last session. Different mixing densities will affect the way we calculate the posterior expectation, and in many cases, we have to rely on numerical evaluation. However, some posterior expectations can be simplified to reduce computational cost when implementing the EM algorithm in practice.

(a) **Univariate Gamma density**

This can be regarded as a special case because the posterior density is known in



closed form as another univariate Gamma density with different parameters.

$$\theta | \Theta^{(j)}, \mathbf{k}_t \sim \text{Gamma} \left( \phi^{(j)} + k_{1,t} + k_{2,t}, \phi^{(j)} + \lambda_{1,t}^{(j)} + \lambda_{2,t}^{(j)} \right) \tag{13}$$

Then, the posterior expectation when updating  $\beta_i$  can be simplified as

$$\mathbb{E}_{\theta,t}^{(j)}[\theta_1] = \mathbb{E}_{\theta,t}^{(j)}[\theta_2] = \frac{\phi^{(j)} + k_{1,t} + k_{2,t}}{\phi^{(j)} + \lambda_{1,t}^{(j)} + \lambda_{2,t}^{(j)}}. \tag{14}$$

Finally, to update  $\phi$

$$\begin{aligned} \phi^{(j+1)} &= \phi^{(j)} - \frac{g(\phi^{(j)})}{h(\phi^{(j)})}, \\ g(\phi^{(j)}) &= n(\log \phi^{(j)} - \Psi(\phi^{(j)} + 1) + \sum_{t=1}^n \left( \mathbb{E}_{\theta,t}^{(j)}[\log \theta] - \mathbb{E}_{\theta,t}^{(j)}[\theta] \right)) \\ h(\phi^{(j)}) &= n((\phi^{(j)})^{-1} - \Psi'(\phi^{(j)})), \end{aligned} \tag{15}$$

where  $\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$  and  $\Psi'(x)$  are digamma and trigamma functions respectively. The posterior expectation  $\mathbb{E}_{\theta,t}^{(j)}[\log \theta]$  is given by

$$\mathbb{E}_{\theta,t}^{(j)}[\log \theta] = \Psi \left( \phi^{(j)} + k_{1,t} + k_{2,t} \right) - \log \left( \phi^{(j)} + \lambda_{1,t}^{(j)} + \lambda_{2,t}^{(j)} \right) \tag{16}$$

**(b) Bivariate Lognormal density**

In this case, there is no analytic expression for the posterior density. However, it can be transformed in the following way

$$\begin{aligned} \pi(\boldsymbol{\theta} | \Theta^{(j)}, \mathbf{k}_t) &= \frac{\eta(\boldsymbol{\theta} | \boldsymbol{\lambda}_t, \mathbf{k}_t) f_{\Sigma}^{LN}(\boldsymbol{\theta})}{\int_0^\infty \int_0^\infty \eta(\boldsymbol{\theta} | \boldsymbol{\lambda}_t, \mathbf{k}_t) f_{\Sigma}^{LN}(\boldsymbol{\theta}) d\theta_1 d\theta_2} \\ &= \frac{\eta(e^\epsilon | \boldsymbol{\lambda}_t, \mathbf{k}_t) f_{\Sigma}^N(\boldsymbol{\epsilon})}{\int_0^\infty \int_0^\infty \eta(e^\epsilon | \boldsymbol{\lambda}_t, \mathbf{k}_t) f_{\Sigma}^N(\boldsymbol{\epsilon}) d\epsilon_1 d\epsilon_2} \\ &=: \pi(\boldsymbol{\epsilon} | \Theta^{(j)}, \mathbf{k}_t). \end{aligned} \tag{17}$$

Then, all posterior expectations with respect to  $\boldsymbol{\theta}$  can be transformed into expectations with respect to  $\boldsymbol{\epsilon}$  such that  $\mathbb{E}_{\theta,t}^{(j)}[h(\boldsymbol{\theta})] = \mathbb{E}_{\epsilon,t}^{(j)}[h(e^\epsilon)]$ . Under this transformation, all the posterior expectations can be evaluated by Gauss-Hermite quadrature. Furthermore,

$$\mathbb{E}_{\theta,t}^{(j)}[\log f_{\Sigma}^{LN}(\boldsymbol{\theta})] = \mathbb{E}_{\epsilon,t}^{(j)}[\log f_{\Sigma}^N(\boldsymbol{\epsilon}) - \epsilon_1 - \epsilon_2] \propto \mathbb{E}_{\epsilon,t}^{(j)}[\log f_{\Sigma}^N(\boldsymbol{\epsilon})].$$

To update  $\phi = \{\phi_1, \phi_2, \rho\}$ ,

$$\begin{aligned}\phi^{(j+1)} &= \phi^{(j)} - H^{-1}(\phi^{(j)})g(\phi^{(j)}) \\ g(\phi^{(j)})_r &= \sum_{t=1}^n \mathbb{E}_{\epsilon,t}^{(j)} \left[ \frac{\partial \log f_{\Sigma}^N(\epsilon)}{\partial \phi_r} \right] \\ H(\phi^{(j)})_{r,s} &= \sum_{t=1}^n \mathbb{E}_{\epsilon,t}^{(j)} \left[ \frac{\partial^2 \log f_{\Sigma}^N(\epsilon)}{\partial \phi_r \partial \phi_s} \right],\end{aligned}\quad (18)$$

where the subscript  $r$  denotes the  $r$ -th element of a vector and  $r, s$  denotes  $r$ -th row  $s$ -th column entry of a matrix. The first and second order derivatives are given by

$$\begin{aligned}\frac{\partial \log f_{\Sigma}^N(\epsilon)}{\partial \phi_i} &= -\frac{1}{1-\rho^2} \left( -\frac{1}{\phi_i^3} \epsilon_i^2 + \frac{\rho \phi_{3-i}}{2\phi_i^2} \epsilon_i - \frac{\rho}{2\phi_{3-i}} \epsilon_{3-i} + \frac{\rho}{\phi_i^2 \phi_{3-i}} \epsilon_1 \epsilon_2 \right. \\ &\quad \left. + \frac{\phi_i - \rho \phi_{3-i}}{4} \right) - \frac{1}{\phi_i} \\ \frac{\partial \log f_{\Sigma}^N(\epsilon)}{\partial \rho} &= -\frac{\rho}{(1-\rho^2)^2} \left( \frac{1}{\phi_1^2} \epsilon_1^2 + \frac{1}{\phi_2^2} \epsilon_2^2 + (1 - \frac{\rho \phi_2}{\phi_1}) \epsilon_1 + (1 - \frac{\rho \phi_1}{\phi_2}) \epsilon_2 \right. \\ &\quad \left. + \frac{\phi_1^2 + \phi_2^2}{4} - \frac{\rho \phi_1 \phi_2}{2} \right) \\ &\quad + \frac{1}{1-\rho^2} \left( \frac{\phi_2}{2\phi_1} \epsilon_1 + \frac{\phi_1}{2\phi_2} \epsilon_2 + \frac{1}{\phi_1 \phi_2} \epsilon_1 \epsilon_2 + \frac{\phi_1 \phi_2}{4} \right) + \frac{\rho}{1-\rho^2} \\ \frac{\partial^2 \log f_{\Sigma}^N(\epsilon)}{\partial \phi_i^2} &= -\frac{1}{1-\rho^2} \left( \frac{3}{\phi_i^4} \epsilon_i^2 - \frac{\rho \phi_{3-i}}{\phi_i^3} \epsilon_i - \frac{2\rho}{\phi_i^3 \phi_{3-i}} \epsilon_1 \epsilon_2 + \frac{1}{4} \right) + \frac{1}{\phi_i^2} \\ \frac{\partial^2 \log f_{\Sigma}^N(\epsilon)}{\partial \rho^2} &= \frac{1+3\rho^2}{(1-\rho^2)^3} \left( \frac{1}{\phi_1^2} \epsilon_1^2 + \frac{1}{\phi_2^2} \epsilon_2^2 + (1 - \frac{\rho \phi_2}{\phi_1}) \epsilon_1 + (1 - \frac{\rho \phi_1}{\phi_2}) \epsilon_2 \right) \\ &\quad + \frac{1+3\rho^2}{(1-\rho^2)^3} \left( -\frac{2\rho}{\phi_1 \phi_2} \epsilon_1 \epsilon_2 + \frac{\phi_1^2 + \phi_2^2}{4} - \frac{\rho \phi_1 \phi_2}{4} \right) \\ &\quad + \frac{4\rho}{(1-\rho^2)^2} \left( \frac{\phi_2}{2\phi_1} \epsilon_1 + \frac{\phi_1}{\phi_2} \epsilon_2 + \frac{1}{\phi_1 \phi_2} \epsilon_1 \epsilon_2 + \frac{\phi_1 \phi_2}{4} \right) \\ &\quad + \frac{1+\rho^2}{(1-\rho^2)^2} \\ \frac{\partial^2 \log f_{\Sigma}^N(\epsilon)}{\partial \phi_1 \partial \phi_2} &= \frac{\rho}{1-\rho^2} \left( -\frac{1}{2\phi_1^2} \epsilon_1 - \frac{1}{2\phi_2^2} \epsilon_2 + \frac{1}{\phi_1^2 \phi_2^2} \epsilon_1 \epsilon_2 + \frac{1}{4} \right) \\ \frac{\partial^2 \log f_{\Sigma}^N(\epsilon)}{\partial \rho \partial \phi_i} &= -\frac{2\rho}{(1-\rho^2)^2} \left( -\frac{1}{\phi_i^3} \epsilon_i^2 + \frac{\rho \phi_{3-i}}{2\phi_i^2} \epsilon_i - \frac{\rho}{2\phi_{3-i}} \epsilon_{3-i} \right.\end{aligned}$$

$$\begin{aligned}
 & + \frac{\rho}{\phi_i^2 \phi_{3-i}} \epsilon_1 \epsilon_2 + \frac{1}{4} (\phi_i - \rho \phi_{3-i}) \\
 & + \frac{1}{1 - \rho^2} \left( -\frac{\phi_{3-i}}{2\phi_i^2} \epsilon_i + \frac{1}{2\phi_{3-i}} \epsilon_{3-i} - \frac{1}{\phi_i^2 \phi_{3-i}} \epsilon_1 \epsilon_2 + \frac{\phi_{3-i}}{4} \right).
 \end{aligned}$$

Notice that all the derivatives are in the linear form of  $\epsilon_1^2, \epsilon_2^2, \epsilon_1, \epsilon_2, \epsilon_1 \epsilon_2$ . Hence, we can evaluate these posterior expectations in each iteration once to avoid repeating calculations.

(c) **Gaussian copula paired with Gamma marginals**

In this case, there is no simplification either for the posterior density or for the posterior expectation. To update  $\boldsymbol{\phi} = \{\phi_1, \phi_2, \rho\}$ , we have almost the same procedure as for the bivariate Lognormal case.

$$\begin{aligned}
 \boldsymbol{\phi}^{(j+1)} &= \boldsymbol{\phi}^{(j)} - H^{-1}(\boldsymbol{\phi}^{(j)})g(\boldsymbol{\phi}^{(j)}) \\
 g(\boldsymbol{\phi}^{(j)})_r &= \sum_{t=1}^n \mathbb{E}_{\boldsymbol{\theta}, t}^{(j)} \left[ \frac{\partial \log f_{GC}(\boldsymbol{\theta})}{\partial \phi_r} \right] \\
 H(\boldsymbol{\phi}^{(j)})_{r,s} &= \sum_{t=1}^n \mathbb{E}_{\boldsymbol{\theta}, t}^{(j)} \left[ \frac{\partial^2 \log f_{GC}(\boldsymbol{\theta})}{\partial \phi_r \partial \phi_s} \right], \tag{19}
 \end{aligned}$$

where the first and second order partial derivatives are given by

$$\begin{aligned}
 u_i &= \int_0^{\phi_i} \frac{\phi_i^{\phi_i}}{\Gamma(\phi_i)} y^{\phi_i-1} e^{-\phi_i y} dy \\
 \frac{\partial \log f_{GC}(\boldsymbol{\theta})}{\partial \phi_i} &= \left( -\frac{\rho^2}{1 - \rho^2} \frac{\Phi^{-1}(u_i)}{f_{sn}(\Phi^{-1}(u_i))} + \frac{\rho}{1 - \rho^2} \frac{\Phi^{-1}(u_{3-i})}{f_{sn}(\Phi^{-1}(u_i))} \right) \frac{\partial u_i}{\partial \phi_i} \\
 &\quad + 1 + \log(\phi) - \Psi(\phi) + \log(\theta) - \theta \\
 \frac{\partial \log f_{GC}(\boldsymbol{\theta})}{\partial \rho} &= \frac{\rho}{1 - \rho^2} - \frac{\rho}{(1 - \rho^2)^2} \left( \Phi^{-1}(u_1) + \Phi^{-1}(u_2) \right) \\
 &\quad + \frac{1 + \rho^2}{(1 - \rho^2)^2} \Phi^{-1}(u_1) \Phi^{-1}(u_2) \\
 \frac{\partial^2 \log f_{GC}(\boldsymbol{\theta})}{\partial \phi_i^2} &= \left( -\frac{\rho^2}{1 - \rho^2} \frac{1 + \Phi^{-1}(u_i)^2}{f_{sn}(\Phi^{-1}(u_i))^2} + \frac{\rho}{1 - \rho^2} \frac{\Phi^{-1}(u_1) \Phi^{-1}(u_2)}{f_{sn}(\Phi^{-1}(u_i))} \right) \\
 &\quad \times \left( \frac{\partial u_i}{\partial \phi_i} \right)^2 + \frac{\partial \log f_{GC}(\boldsymbol{\theta})}{\partial \phi_i} \frac{\partial^2 u_i}{\partial \phi_i^2} + \frac{1}{\phi} - \Psi'(\phi_i) \\
 \frac{\partial^2 \log f_{GC}(\boldsymbol{\theta})}{\partial \rho^2} &= \frac{1 + \rho^2}{(1 - \rho^2)^2} - \frac{1 + 3\rho^2}{(1 - \rho^2)^3} \left( \Phi^{-1}(u_1) + \Phi^{-1}(u_2) \right) \\
 &\quad + \frac{2\rho^3 + 6\rho}{(1 - \rho^2)^3} \Phi^{-1}(u_1) \Phi^{-1}(u_2)
 \end{aligned}$$

$$\frac{\partial^2 \log f_{GC}(\boldsymbol{\theta})}{\partial \phi_1 \partial \phi_2} = \frac{\rho}{1 - \rho^2} \frac{\frac{\partial u_1}{\partial \phi_1} \frac{\partial u_2}{\partial \phi_2}}{f_{sn}(\Phi^{-1}(u_1)) f_{sn}(\Phi^{-1}(u_2))}$$

$$\frac{\partial^2 \log f_{GC}(\boldsymbol{\theta})}{\partial \rho \partial \phi_i} = \left( -\frac{2\rho}{(1 - \rho^2)^2} \frac{\Phi^{-1}(u_i)}{f_{sn}(\Phi^{-1}(u_i))} + \frac{1 + \rho^2}{(1 - \rho^2)^2} \frac{\Phi^{-1}(u_{3-i})}{f_{sn}(\Phi^{-1}(u_i))} \right) \frac{\partial u_i}{\partial \phi_i}.$$

### 3 The bivariate mixed Poisson INAR(1) regression model

#### 3.1 Model specifications

Let  $\mathbf{X}$  and  $\mathbf{R}$  be non-negative integer-valued random vectors in  $\mathbb{R}^2$ . Let  $\mathbf{P}$  be a diagonal matrix in  $\mathbb{R}^{2 \times 2}$  with elements  $p_i \in (0, 1)$ . The bivariate first-order integer-valued autoregressive model (Bivariate INAR(1)) is defined as

$$\mathbf{X}_t = \mathbf{P} \circ \mathbf{X}_{t-1} + \mathbf{R}_t = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix} \circ \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} R_{1,t} \\ R_{2,t} \end{bmatrix}, \tag{20}$$

where the thinning operator  $\circ$  is the widely used binomial thinning operator such that  $p_i \circ X_{i,t} = \sum_{k=1}^{X_{i,t}} U_k$  and  $U_k$  are independent identically distributed Bernoulli random variables with success probability  $p_i$ , i.e.  $\mathcal{P}(U_k = 1) = p_i$ . Hence  $p_i \circ X_{i,t}$  is binomially distributed with size  $X_{i,t}$  and success probability  $p_i$ . Then the distribution function  $f_{p_i}(x, X_{i,t})$  can be easily written down as

$$f_{p_i}(k, X_{i,t}) = \binom{X_{i,t}}{k} p_i^k (1 - p_i)^{X_{i,t}-k}. \tag{21}$$

Note that  $p_i \circ X_{i,t}$  and  $p_j \circ X_{j,t}, i \neq j$  are independent of each other. To adapt the heteroscedasticity arising from the data,  $\mathbf{R}_t$  is bivariate mixed Poisson regression model such that  $R_{i,t} \sim Po(\lambda_{i,t}, \theta_i)$  defined in the last session. The joint distribution of the bivariate sequence  $\mathbf{X}_{t+1}$  conditional on the last state  $\mathbf{X}_t$  is given by

$$\mathcal{P}(\mathbf{X}_{t+1} | \mathbf{X}_t) = \sum_{k_1=0}^{s_{1,t}} \sum_{k_2=0}^{s_{2,t}} f_{p_1}(k_1, X_{1,t}) f_{p_2}(k_2, X_{2,t}) f_{\mathbf{R}}(X_{1,t+1} - k_1, X_{2,t+1} - k_2)$$

$$f_{\mathbf{R}}(\mathbf{k}, t) = \frac{\lambda_{1,t}^{k_1}}{k_1!} \frac{\lambda_{2,t}^{k_2}}{k_2!} \int_0^\infty \int_0^\infty \eta(\boldsymbol{\theta} | \lambda_t, \mathbf{k}_t) f_{\boldsymbol{\phi}}(\boldsymbol{\theta}) d\theta_1 d\theta_2$$

$$s_{i,t} = \min\{X_{i,t+1}, X_{i,t}\}, \tag{22}$$

where  $f_{\mathbf{R}}(\mathbf{k}, t)$  is a probability mass function of a bivariate mixed Poisson regression model with mixing density  $f_{\boldsymbol{\phi}}(\boldsymbol{\theta})$ . Under this construction, the bivariate sequence  $\mathbf{X}_t$  is correlated with each other and its correlation structure mainly depends on the correlation structure of innovation  $\mathbf{R}_t$ .

### 3.2 The EM algorithm

Similarly, the maximum likelihood estimation is not straightforward to apply as the log likelihood function

$$\ell(\Theta) = \sum_{t=1}^n \log \left( \sum_{k_1=0}^{s_{1,t}} \sum_{k_2=0}^{s_{2,t}} f_{p_1}(k_1, X_{1,t}) f_{p_2}(k_2, X_{2,t}) f_{\mathbf{R}}(X_{1,t+1}-k_1, X_{2,t+1}-k_2) \right)$$

has discrete convolution and double integrals. Then we can use similar techniques to decompose the log likelihood function as we did in Sect. 2.

Given the observed bivariate sequence  $\{\mathbf{X}_t\}_{t=1,\dots,n}$ . Let  $Y_{i,t} = p_i \circ X_{i,t-1}$  and  $\Theta = \{p_1, p_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\phi}\}$  be the parameter space for this model. Suppose now we observe the latent variable  $\{Y_t\}_{t=1,\dots,n}$ , then the log likelihood function becomes

$$\begin{aligned} \ell(\Theta|\mathbf{Y}) &\propto \sum_{t=1}^n \sum_{i=1}^2 (Y_{i,t} \log p_i + (X_{i,t} - Y_{i,t}) \log(1 - p_i)) + \sum_{t=1}^n \log f_{\mathbf{R}}(\mathbf{R}_t, t) \\ R_{i,t} &= X_{i,t} - Y_{i,t}. \end{aligned} \tag{23}$$

Notice that there are still unobserved random variables  $\boldsymbol{\theta}$  in  $R_t$ . In some of the examples we discuss in the last section,  $f_{\mathbf{R}}(\mathbf{k}, t)$  may not have analytic expression and hence we would like to further break down the likelihood function. Suppose further that we observe the random effect  $\{\boldsymbol{\theta}_t\}_{t=1,\dots,n}$ , then the complete log likelihood becomes

$$\begin{aligned} \ell(\Theta|\mathbf{Y}, \boldsymbol{\theta}) &\propto \sum_{t=1}^n \sum_{i=1}^2 (Y_{i,t} \log p_i + (X_{i,t} - Y_{i,t}) \log(1 - p_i)) \\ &+ \sum_{t=1}^n \sum_{i=1}^2 (R_{i,t} \log(\lambda_{i,t}) - \lambda_{i,t} \theta_{i,t}) + \sum_{t=1}^n \log f_{\boldsymbol{\phi}}(\boldsymbol{\theta}). \end{aligned} \tag{24}$$

Define the following posterior density functions

$$\begin{aligned} \pi_1(\mathbf{y}|\Theta^{(j)}, \mathbf{X}_t, \mathbf{X}_{t-1}) &= \frac{f_{\mathbf{R}}(\mathbf{X}_{t-1} - \mathbf{y}) \prod_{i=1}^2 f_{p_i}(y_i, X_{i,t-1})}{\mathcal{P}(\mathbf{X}_t|\mathbf{X}_{t-1})} \\ \pi_2(\boldsymbol{\theta}|\Theta^{(j)}, \mathbf{R}_t) &= \frac{\eta(\boldsymbol{\theta}|\boldsymbol{\lambda}_t, \mathbf{R}_t) f_{\boldsymbol{\phi}}(\boldsymbol{\phi})}{\int_0^\infty \int_0^\infty \eta(\boldsymbol{\theta}|\boldsymbol{\lambda}_t, \mathbf{R}_t) f_{\boldsymbol{\phi}}(\boldsymbol{\theta}) d\theta_1 d\theta_2}, \end{aligned} \tag{25}$$

Define the posterior expectations with respect to real-value functions  $h(\cdot, \cdot)$

$$\begin{aligned} \mathbb{E}_{\mathbf{y},t}^{(j)}[h(\mathbf{y})] &= \sum_{y_1=0}^{s_{1,t-1}} \sum_{y_2=0}^{s_{2,t-1}} h(\mathbf{y}) \pi_1(\mathbf{y}|\Theta^{(j)}, \mathbf{X}_t, \mathbf{X}_{t-1}) \\ \mathbb{E}_{\boldsymbol{\theta},t}^{(j)}[h(\boldsymbol{\theta})|\mathbf{R}_t] &= \int_0^\infty \int_0^\infty h(\boldsymbol{\theta}) \pi_2(\boldsymbol{\theta}|\Theta^{(j)}, \mathbf{R}_t) d\theta_1 d\theta_2. \end{aligned} \tag{26}$$

- **E-step:** Evaluating the Q function  $Q(\Theta; \Theta^{(j)})$  given the the parameters estimated in the  $j$ -th iteration,

$$\begin{aligned}
 Q(\Theta; \Theta^{(j)}) &= \sum_{t=1}^n \sum_{i=1}^2 (y_{i,t}^{(j)} \log p_i + (X_{i,t-1} - y_{i,t}^{(j)}) \log(1 - p_i)) \\
 &\quad + \sum_{t=1}^n \sum_{i=1}^2 (r_{i,t}^{(j)} \log(\lambda_{i,t}) - \lambda_{i,t} \hat{\theta}_{i,t}^{(j)}) + \sum_{t=1}^n \mathbb{E}_{\mathbf{y},t}^{(j)} [\mathbb{E}_{\theta,t}^{(j)} [\log f_{\phi}(\theta) | \mathbf{R}_t]] \\
 y_{i,t}^{(j)} &= \mathbb{E}_{\mathbf{y},t}^{(j)} [Y_i], \quad r_{i,t}^{(j)} = X_{i,t} - y_{i,t}^{(j)}, \quad \hat{\theta}_{i,t}^{(j)} = \mathbb{E}_{\mathbf{y},t}^{(j)} [\mathbb{E}_{\theta,t}^{(j)} [\theta_i | \mathbf{R}_t]].
 \end{aligned}
 \tag{27}$$

After breaking down the log likelihood function, it is obvious that except for the log likelihood contributed by binomial distribution, the rest of the terms are almost the same as that of the Q-function of bivariate mixed Poisson regression model discussed in the last session, which means the updating procedure for  $\beta_i, \phi$  will be exactly the same, but we need to evaluate different posterior expectations in this case.

- **M-step:** Similarly, we apply the Newton–Raphson algorithm to update the parameters. Based on the structure of  $Q(\Theta; \Theta^{(j)})$ , the parameters can be updated separately for binomial part  $p$ , Poisson part  $\beta_i$  and random effect part  $\phi$ 
  - The binomial part can be updated simply as the following gradient function has a unique solution

$$\begin{aligned}
 g(p_i) &= \frac{\sum_{t=1}^n y_{i,t}^{(j)}}{p_i} - \frac{\sum_{t=1}^n (X_{i,t-1} - y_{i,t}^{(j)})}{1 - p_i} = 0 \\
 p_i^{(j+1)} &= \frac{\sum_{t=1}^n y_{i,t}^{(j)}}{\sum_{t=1}^n X_{i,t-1}}, \quad i = 1, 2 \\
 y_{i,t}^{(j)} &= \begin{cases} \frac{p_i^{(j)} X_{i,t-1} \mathcal{P}(\mathbf{X}_t - \mathbf{1}_i | \mathbf{X}_{t-1} - \mathbf{1}_i)}{\mathcal{P}(\mathbf{X}_t | \mathbf{X}_{t-1})}, & X_{i,t} \neq 0 \text{ and } X_{i,t-1} \neq 0 \\ 0, & \text{otherwise} \end{cases} \\
 \mathbf{1}_1 &= (1, 0)^T \quad \mathbf{1}_2 = (0, 1)^T.
 \end{aligned}
 \tag{28}$$

See Appendix A.4 for the derivation of  $y_{i,t}^{(j)}$ .

- For the Poisson part, the updating equations are the same with different posterior expectation

$$\begin{aligned}
 \beta_i^{(j+1)} &= \beta_i^{(j)} - H^{-1}(\beta_i^{(j)}) g(\beta_i^{(j)}), \quad i = 1, 2 \\
 g(\beta_i^{(j)}) &= \mathbf{Z}_i^T \mathbf{V}_i^{(g)} \quad H(\beta_i^{(j)}) = \mathbf{Z}_i^T \mathbf{D}_i^{(H)} \mathbf{Z}_i \\
 \mathbf{V}_i^{(g)} &= \left( \left\{ k_{i,t} - \lambda_{i,t}^{(j)} \hat{\theta}_{i,t}^{(j)} \right\}_{t=1, \dots, n} \right) \\
 \mathbf{D}_i^{(H)} &= \text{diag} \left( \left\{ -\lambda_{i,t}^{(j)} \hat{\theta}_{i,t}^{(j)} \right\}_{t=1, \dots, n} \right).
 \end{aligned}
 \tag{29}$$

Note that when the mixing density  $f_{\phi}(\boldsymbol{\theta})$  is univariate Gamma, the posterior expectation for  $\boldsymbol{\theta}$  has a simple expression

$$\hat{\theta}_t^{(j)} = \hat{\theta}_{1,t}^{(j)} = \hat{\theta}_{2,t}^{(j)} = \frac{\phi^{(j)} + r_{1,t}^{(j)} + r_{2,t}^{(j)}}{\phi^{(j)} + \lambda_{1,t}^{(j)} + \lambda_{2,t}^{(j)}}.$$

– Similarly, for the random effect part  $\boldsymbol{\phi}$ ,

(a) **Univariate Gamma density**

$$\begin{aligned} \phi^{(j+1)} &= \phi^{(j)} - \frac{g(\phi^{(j)})}{h(\phi^{(j)})}, \\ g(\phi^{(j)}) &= n(\log \phi^{(j)} - \Psi(\phi^{(j)} + 1) + 1) + \sum_{t=1}^n \left( \mathbb{E}_{\mathbf{y},t}^{(j)} [\mathbb{E}_{\theta,t}^{(j)} [\log \theta | \mathbf{R}_t]] - \hat{\theta}_t^{(j)} \right) \\ h(\phi^{(j)}) &= n((\phi^{(j)})^{-1} - \Psi'(\phi^{(j)})), \end{aligned} \tag{30}$$

(b) **Bivariate Lognormal**

$$\begin{aligned} \boldsymbol{\phi}^{(j+1)} &= \boldsymbol{\phi}^{(j)} - H^{-1}(\boldsymbol{\phi}^{(j)})g(\boldsymbol{\phi}^{(j)}) \\ g(\boldsymbol{\phi}^{(j)})_r &= \sum_{t=1}^n \mathbb{E}_{\mathbf{y},t}^{(j)} \left[ \mathbb{E}_{\boldsymbol{\epsilon},t}^{(j)} \left[ \frac{\partial \log f_{\Sigma}^N(\boldsymbol{\epsilon})}{\partial \boldsymbol{\phi}_r} \mid \mathbf{R}_t \right] \right] \\ H(\boldsymbol{\phi}^{(j)})_{r,s} &= \sum_{t=1}^n \mathbb{E}_{\mathbf{y},t}^{(j)} \left[ \mathbb{E}_{\boldsymbol{\epsilon},t}^{(j)} \left[ \frac{\partial^2 \log f_{\Sigma}^N(\boldsymbol{\epsilon})}{\partial \boldsymbol{\phi}_r \partial \boldsymbol{\phi}_s} \mid \mathbf{R}_t \right] \right], \end{aligned} \tag{31}$$

(c) **Gaussian copula paired with Gamma marginals**

$$\begin{aligned} \boldsymbol{\phi}^{(j+1)} &= \boldsymbol{\phi}^{(j)} - H^{-1}(\boldsymbol{\phi}^{(j)})g(\boldsymbol{\phi}^{(j)}) \\ g(\boldsymbol{\phi}^{(j)})_r &= \sum_{t=1}^n \mathbb{E}_{\mathbf{y},t}^{(j)} \left[ \mathbb{E}_{\boldsymbol{\theta},t}^{(j)} \left[ \frac{\partial \log f_{GC}(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}_r} \mid \mathbf{R}_t \right] \right] \\ H(\boldsymbol{\phi}^{(j)})_{r,s} &= \sum_{t=1}^n \mathbb{E}_{\mathbf{y},t}^{(j)} \left[ \mathbb{E}_{\boldsymbol{\theta},t}^{(j)} \left[ \frac{\partial^2 \log f_{GC}(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}_r \partial \boldsymbol{\phi}_s} \mid \mathbf{R}_t \right] \right]. \end{aligned} \tag{32}$$

The partial derivatives inside expectations are derived in the last section.

**Remark** This model as well as the EM algorithm can be extended to multivariate case straightforwardly. All the steps and the general form of the formula of the EM algorithm in the multivariate case are exactly the same. The only problem is that it would become cumbersome to evaluate the transition probability  $\mathcal{P}(\mathbf{X}_t | \mathbf{X}_{t-1})$  as dimension of  $\mathbf{X}_t$  increases.

## 4 Empirical analysis

### 4.1 Data description and model fitting

The data used in this section come from the Local Government Property Insurance Fund (LGPIF) from the state of Wisconsin. On previous application on this dataset, interested reader can refer to Refs. [28–30]. This fund provides property insurance to different types of government units, which includes villages, cities, counties, towns and schools. The LGPIF contains three major groups of property insurance coverage, namely building and contents (BC), contractors' equipment (IM) and motor vehicles (PN, PO, CN, CO). For exploratory purposes, we focus on modelling jointly the claim frequency of IM, denoted as  $X_1$ , and comprehensive new vehicles collision (CN), denoted as  $X_2$  as they are both related to land transport. The insurance data cover the period over 2006–2010 with 1234 policyholders in total. Only  $n_1 = 1048$  of them have complete data over the period 2006–2010, which will be the training dataset. The last year 2011 with  $n_2 = 1025$  policyholders, which is the same set of policyholders as in the training dataset, out of 1098 policyholders will be the test dataset. Denote the IM type and CN type claim frequency for a particular policyholder as  $X_{1,t}^{(h)}$ ,  $X_{2,t}^{(h)}$  respectively, where  $h$  is the identifier for each policyholder and  $t$  is the year. Then the relationship between  $X_{i,t}$  and  $X_{i,t}^{(h)}$  is simply  $X_{i,t} = \sum_h X_{i,t}^{(h)}$  with  $i = 1, 2$ .

Some basic statistical analysis is shown in the following Table 1 and Fig. 1. The proportion of zeros for two types of claims are all over 90% during 2006–2010. Both types of claim shows overdispersion as the variance are all higher than their mean over years and the overdispersion for  $X_{2,t}$  is even stronger than that of  $X_{1,t}$ , which indicate the need to apply overdispersed distribution model for the data. The correlation tests over years imply that it is reasonable to introduce correlation structure between  $X_{1,t}$  and  $X_{2,t}$ . The proportion of zeros and kurtosis show that the marginal distributions of  $X_{1,t}$ ,  $X_{2,t}$  are all positively skewed and exhibit a fat-tailed structure which indicates the appropriateness of adopting a positive skewed and fat-tailed distribution (Log Normal distribution). Last but not least, the correlation tests illustrated in Table 2 do support the appropriateness of introduction of time series term in modelling the claim sequence.

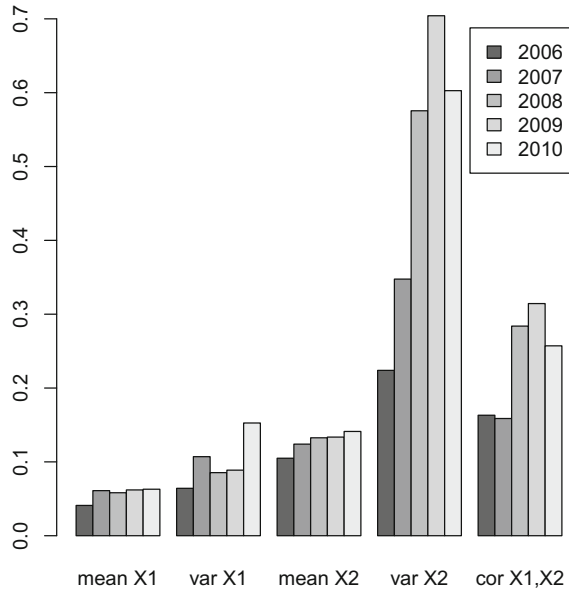
**Table 1** Summary statistics of two types of claims over years

	2006	2007	2008	2009	2010
Proportion of zeros $X_{1,t}$	0.9685	0.9542	0.9552	0.9504	0.9590
Proportion of zeros $X_{2,t}$	0.9342	0.9332	0.9399	0.9370	0.9323
Kurtosis of $X_{1,t}$	85.7500	86.6479	41.8491	43.0183	126.6879
Kurtosis of $X_{2,t}$	53.9183	61.7740	111.2810	184.1395	133.9228
P-value of correlation test	0.0000	0.0000	0.0000	0.0000	0.0000

The correlations test is a one-sided test where the alternative hypothesis is “The sample correlation is greater than 0”



**Fig. 1** Summary statistics (mean, variance and correlation) for each type of claims across all the policyholders over the years



The description and some summary statistics for all the explanatory variables (covariates  $z_{1,t}, z_{2,t}$ ) that are relevant to  $X_{1,t}, X_{2,t}$  are shown in Table 3. Variables 1–5 including ‘TypeVillage’ are categorical variables to indicate the entity types of a policyholder. Due to the strongly heavy-tailed structure appearing in variables 6 and 9, which can drastically distort the model fitting, those variables are transformed by means of the ‘rank’ function in R software and then standardized, which can mitigate the effect of outliers. Variables 6–8 are relevant to IM claim  $X_{1,t}$ , while variables 9,10 provide information for CN claims  $X_{2,t}$ . The covariate  $z_{1,t}$  includes variables 1-8, and  $z_{2,t}$  contains variables 1-5 and variables 9,10. These covariates act as the regression part for  $\lambda_{i,t}$  mentioned in Sect. 2, which may help to explain part of the heterogeneity within  $X_{1,t}$  and  $X_{2,t}$ .

Due to the large computational cost for evaluating the partial derivatives of copula case (large sample size), all the models except the copula case discussed in Sects. 2 and 3 are applied to model the joint behaviour of  $X_{1,t}^{(h)}, X_{2,t}^{(h)}$  across all the policyholders. Instead, a simulation study in the Appendix A.5 shows that the EM algorithm does work for copula case.

**Table 2** Correlation test for  $X_{i,t}$  and  $X_{i,t-1}$

	$X_1$	$X_2$
Correlation	0.4062	0.7478
p-value	0.0000	0.0000

The test is a one-sided test where the alternative hypothesis is “The sample correlation is greater than 0”

**Table 3** Summary statistics for the explanatory variables

Variable index	Variable name	Type	Description	Proportion/Mean
1	TypeCity	Categorical	Indicator for city entity	0.1400
2	TypeCounty	Categorical	Indicator for county entity	0.0578
3	TypeMisc	Categorical	Indicator for miscellaneous entity	0.1104
4	TypeSchool	Categorical	Indicator for school entity	0.2817
5	TypeTown	Categorical	Indicator for town entity	0.1728
-	TypeVillage	Categorical	Indicator for village entity (reference category)	0.2373
6	CoverageIM	Continuous	Coverage amount of IM(transformed)	0
7	InDeductIM	Continuous	Log deductible amount for inland marine	5.3400
8	NoClaimCreditIM	Binary	Indicator for no IM claims in prior year	0.4210
9	CoverageCN	Countinuous	Coverage amount of CN (transformed)	0
10	NoClaimCreditCN	Binary	Indicator for no CN claims in prior year	0.0897

Since we would like to model the whole behaviour rather than the individual one, the the likelihood function would simply become

$$\ell(\Theta) = \sum_{h=1}^{n_1} \ell_h(\Theta) = \sum_{h=1}^{n_1} \sum_{t=1}^4 \log \Pr(X_{1,t+1}^{(h)}, X_{2,t+1}^{(h)} | X_{1,t}^{(h)}, X_{2,t}^{(h)}), \tag{33}$$

where  $\ell_h(\Theta)$  is the log likelihood function for policyholder  $h$ . Note that all the policyholders with the same type of claim  $X_i$ , will share the same set of parameters  $\{p_1, p_2, \beta_1, \beta_2, \phi\}$ . In addition, it is necessary to show the appropriateness of introducing crosscorrelation and autocorrelation in BINAR(1) model. Then we also fit the data to following models.

1. The joint distribution of  $X_{1,t}^{(h)}$  and  $X_{2,t}^{(h)}$  are characterized by two independent mixed Poisson (TMP)

$$X_1^{(h)} \sim Pois(\lambda_1\theta_1), \quad X_2^{(h)} \sim Pois(\lambda_2\theta_2), \tag{34}$$

where  $\theta_1$  and  $\theta_2$  are independent random variables, either Gamma or Log Normal.

2. The joint distribution of  $X_{1,t}^{(h)}$  and  $X_{2,t}^{(h)}$  are assumed to be bivariate mixed Poisson distribution (BMP) with different probability mass function according to the choice of mixing densities, see equations (2) and (4).
3. The joint distribution of  $X_{1,t}^{(h)}$  and  $X_{2,t}^{(h)}$  are characterized by two independent INAR(1) models (TINAR)

$$\begin{aligned} X_{1,t}^{(h)} &= p_1 \circ X_{1,t-1}^{(h)} + R_{1,t} \\ X_{2,t}^{(h)} &= p_2 \circ X_{2,t-1}^{(h)} + R_{2,t}, \end{aligned}$$

where  $R_{i,t} \sim Pois(\lambda_{i,t}\theta_{i,t}), i = 1, 2$  and random effect  $\theta_{i,t}$  is independent of  $i$  and  $t$ .

For comparison purpose, we fit these univariate and bivariate Poisson mixture models with training dataset starting from 2007 because they do not need to consider the lag responses. When it comes to the initial values, we use the following. Lag one correlation of each sequence serves as the initial value of  $p_i$ . We fit a Poisson generalized linear model for each sequence to obtain the initial values of  $\beta_i$ . Finally, we used the moment estimates of the bivariate Poisson mixture model (without regression) for initial values of  $\phi = \{\phi_1, \phi_2, \rho\}$ . All the estimation is performed in R software where we implement the EM algorithms derived in previous sections. The standard deviations of the estimators are calculated by inverting the observed information of matrix from the incomplete log-likelihood function (the log likelihood function without unobserved latent variables).

Model fitting results are shown in Table 4. Within the same class of models, compared to univariate Gamma as mixing density, the Log Normal case allows more flexible structures to capture different distributional behaviour within two types of claims. Hence we can observe the improvement of AIC from univariate Gamma case

**Table 4** Goodness of fit for different models with different choices of mixing densities

	Mixing density	Gamma AIC  BIC	Log Normal AIC BIC
TMP	Gamma	3073.332   3193.810	3067.119   3187.596
	Log Normal	3072.287   3192.765	<b>3066.074   3186.551</b>
TINAR	Gamma	2999.957   3133.117	2999.843   3133.003
	Log Normal	2998.326   3131.485	<b>2998.212   3131.371</b>
BMP	–	3073.176   3191.329	<b>3055.066   3181.885</b>
BINAR	–	2996.291   3123.109	<b>2990.744   3130.245</b>

For the class TMP and TINAR, row and column stand for mixing density  $\theta_1, \theta_2$ , respectively. The bold cells indicate the best one within the same class of models

to Log Normal case and hence it is no surprise that the Log Normal is always the best choice within the same class of model. Among different classes of models, it is clear that the adoption of autocorrelation component significantly improves the model fitting. Finally, the significant improvement in terms of AIC from TINAR to BINAR, as well as from TMP to BMP, indicates that it is appropriate to introduce cross correlation between two sequences  $X_1, X_2$ . The estimated parameters via EM algorithm are shown in Table 5.

## 4.2 Predictive performance

In insurance claims modelling, it is more useful to check the overall distribution for all policyholders rather than prediction of the claim frequency for each policyholder, which can be used for premium calculation, risk management, and so forth. To evaluate the predictive performance, we then calculate the predicted claim frequencies  $\text{Freq}(\mathbf{X}_t | \mathbf{X}_{t-1}, \hat{\Theta})$ , which are the sum of individual probabilities  $\mathcal{P}(\mathbf{X}_t^{(h)} | \mathbf{X}_{t-1}^{(h)}, \hat{\Theta})$  of joint events  $(X_1^{(h)}, X_2^{(h)}) \in \{(i, j), 0 \leq i, j \leq 10\}$  based on the estimated parameters, and compare these to the observed frequencies from the test sample  $(X_{1,2011}, X_{2,2011})$  (year 2011). In addition to our proposed BINAR model with Log Normal mixing density, we also compute predictive performance of the best TMP, TINAR, BMP models from Table 4 as the benchmark for comparison purposes. Based on the predictive claim frequencies, one can also compute expected number of claims marginally ( $\mathbb{E}[X_1], \mathbb{E}[X_2]$ ),

$$\begin{aligned} \mathbb{E}[X_1] &= \sum_{k_1=0}^{10} \sum_{k_2=0}^{10} k_1 \text{Freq}(k_1, k_2 | \mathbf{X}_{2010}, \hat{\Theta}) \\ \mathbb{E}[X_2] &= \sum_{k_1=0}^{10} \sum_{k_2=0}^{10} k_2 \text{Freq}(k_1, k_2 | \mathbf{X}_{2010}, \hat{\Theta}) \end{aligned} \quad (35)$$

and measure the Predictive Sum of Square error:

$$\text{PSSE} = (\mathbb{E}[X_1] - X_{1,2011})^2 + (\mathbb{E}[X_2] - X_{2,2011})^2. \quad (36)$$

**Table 5** Parameter estimation for the Bivariate mixed Poisson regression model and the bivariate INAR model of insurance claim frequency data with EM algorithm

Estimate	Model							
	BMP				BINAR			
	BPGA		BPLN		BINAR(1)-GA		BINAR(1)-LN	
	$X_{1,t}$	$X_{2,t}$	$X_{1,t}$	$X_{2,t}$	$X_{1,t}$	$X_{2,t}$	$X_{1,t}$	$X_{2,t}$
$p_i$					0.1238 (0.0373)	0.2904 (0.0378)	0.1118 (0.0384)	0.2761 (0.03950)
$\phi_i$	1.1742 (0.1656)		1.0668 (0.1167)	0.8408 (0.0696)	0.9495 (0.1662)		1.0559 (0.1274)	0.9403 (0.0931)
$\rho$			0.5895 (0.1485)				0.6063 (0.1782)	
Intercept	-3.7300 (0.4835)	-5.4277 (0.3612)	-3.6811 (0.5175)	-5.4836 (0.3743)	-3.7980 (0.5158)	-5.9744 (0.4428)	-3.7966 (0.5412)	-6.0363 (0.4607)
TypeCity	-0.2183 (0.2405)	0.5354 (0.2259)	-0.2830 (0.2565)	0.5234 (0.2306)	-0.2242 (0.2823)	0.6673 (0.2577)	-0.2735 (0.2657)	0.6625 (0.2888)
TypeCounty	0.5269 (0.2605)	1.2034 (0.2074)	0.6190 (0.2809)	1.1312 (0.2119)	0.5682 (0.2811)	1.3290 (0.2643)	0.6200 (0.2930)	1.2744 (0.2698)
TypeMisc	-2.0565 (1.0192)	-0.4822 (0.6349)	-2.0968 (1.0251)	-0.4916 (0.6355)	-2.0110 (1.0210)	-0.1141 (0.6567)	-2.0360 (1.0253)	-0.1279 (0.6604)
TypeSchool	0.0678 (0.3242)	-0.0068 (0.2291)	-0.0768 (0.3401)	-0.0190 (0.2318)	-0.0387 (0.3587)	0.1559 (0.2811)	-0.0945 (0.36074)	0.1367 (0.2852)
TypeTown	-0.4060 (0.2992)	-1.1058 (0.4289)	-0.4427 (0.3064)	-1.1294 (0.4309)	-0.3565 (0.3037)	-0.8941 (0.4794)	-0.3830 (0.3091)	-0.9257 (0.4832)

Table 5 continued

Estimate	Model							
	BMP				BINAR			
	BPGA		BPLN		BINAR(1)-GA		BINAR(1)-LN	
	$X_{1,t}$	$X_{2,t}$	$X_{1,t}$	$X_{2,t}$	$X_{1,t}$	$X_{2,t}$	$X_{1,t}$	$X_{2,t}$
CoverageIM	1.5299 (0.1971)		1.4519 (0.2062)		1.4543 (0.2126)		1.4219 (0.2163)	
InDeductIM	0.0241 (0.0730)		0.0274 (0.0799)		0.0170 (0.0788)		0.0233 (0.0835)	
NoClaimCreditIM	-0.6557 (0.1421)		-0.6053 (0.1552)		-0.4569 (0.1570)		-0.4382 (0.1665)	
CoverageCN		2.3947 (0.1819)		2.4422 (0.1903)		2.4227 (0.2210)		2.4818 (0.2210)
NoClaimCreditCN		-0.6752 (0.1652)		-0.6786 (0.1663)		-0.3047 (0.1811)		-0.3139 (0.1811)

For each entry, the upper one is the estimate and the estimated standard deviations are indicated in round brackets

**Table 6** Predictive and observed joint frequencies for each models

	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)	(2,0)	(2,1)	(2,2)
TMP	901.40	36.53	11.95	38.30	4.03	2.20	5.37	0.97	0.62
TINAR	903.99	37.47	11.70	37.42	4.16	2.30	4.87	0.93	0.61
BMP	904.09	36.87	11.75	36.91	4.09	2.41	4.81	0.81	0.63
BINAR	906.03	37.67	11.55	36.31	4.16	2.45	4.45	0.79	0.61
Observed	940	26	6	20	6	2	4	1	1

**Table 7** Predictive marginal claim frequency

	TMP	TINAR	BMP	BINAR	Observed
$\mathbb{E}[X_1]$	79.18	74.17	77.06	72.55	78
$\mathbb{E}[X_2]$	158.89	153.28	157.34	152.68	117
PSSE	1756.164	1330.907	1628.199	1302.765	–
TLL	–348.3537	–340.4619	–345.9555	–338.4115	–

On the other hand, the log likelihood on test samples (TLL) can also be a measure of predictive performance for each model.

$$TLL = \sum_h^{n_2} \log \mathcal{P}(\mathbf{X}_t^{(h)} | \mathbf{X}_{t-1}^{(h)}, \hat{\Theta}). \quad (37)$$

All the results are summarised in Tables 6 and 7 and it is clear that our proposed model, bivariate INAR(1), has the best predictive performance with the smallest PSSE among all other models. Furthermore, TLL result shows that the bivariate INAR outperforms all other models, which is consistent with the model fitting result in Table 4.

### 4.3 Application to ratemaking

In this subsection, the analysis of best fitted models from Table 4 for ratemaking is conducted. We select three representative risk profiles under different models, named Good, Average and Bad, illustrated in Table 8. These three risk profiles are selected according to the sign and size of the coefficients in Table 5 and those variables are not mentioned in the following table are taken to be 0. Note that CoverageIM and CoverageCN are selected according to their empirical distribution on test data.

We then evaluate the mean and variance of  $X_{1,t}^{(h)} + X_{2,t}^{(h)}$  under each best TMP, TINAR, BMP and BINAR according to Table 4. The mean and variance for one policyholder is given by two quantity, i.e.  $\mathbb{E}[X_1^{(h)} + X_2^{(h)} | \hat{\Theta}, \mathbf{X}_{t-1}, \mathbf{Z}_1, \mathbf{Z}_2]$  and

**Table 8** The risk profiles for two claim sequences

		TypeCounty	TypeMisc	CoverageIM
$\mathbf{Z}_1$	Good	0	1	-1.43
	Average	1	0	0
	Bad	1	0	1
		TypeCounty	TypeMisc	CoverageCN
$\mathbf{Z}_2$	Good	0	1	-0.56
	Average	1	0	1.6
	Bad	1	0	2.15

$\text{Var}(X_1^{(h)} + X_2^{(h)} | \hat{\Theta}, \mathbf{X}_{t-1}, \mathbf{Z}_1, \mathbf{Z}_2)$ . They have following explicit formulae

$$\begin{aligned}
 \mathbb{E}[R_{1,t} + R_{2,t} | \hat{\Theta}, \mathbf{Z}_1, \mathbf{Z}_2] &= \hat{\lambda}_1 + \hat{\lambda}_2 = e^{\mathbf{Z}_1 \hat{\beta}_1} + e^{\mathbf{Z}_2 \hat{\beta}_2} \\
 \text{Var}(R_{1,t} + R_{2,t} | \hat{\Theta}, \mathbf{Z}_1, \mathbf{Z}_2) &= \hat{\lambda}_1^2 (e^{\hat{\sigma}_1} - 1) + \hat{\lambda}_2^2 (e^{\hat{\sigma}_2} - 1) + \hat{\lambda}_1 + \hat{\lambda}_2 \\
 &\quad + 2\hat{\lambda}_1 \hat{\lambda}_2 (e^{\hat{\rho} \hat{\sigma}_1 \hat{\sigma}_2} - 1) \\
 \mathbb{E}[X_1^{(h)} + X_2^{(h)} | \hat{\Theta}, \mathbf{X}_{t-1}, \mathbf{Z}_1, \mathbf{Z}_2] &= \hat{p}_1 X_{1,t-1}^{(h)} + \hat{p}_2 X_{2,t-1}^{(h)} \tag{38} \\
 &\quad + \mathbb{E}[R_{1,t} + R_{2,t} | \hat{\Theta}, \mathbf{Z}_1, \mathbf{Z}_2] \\
 \text{Var}(X_1^{(h)} + X_2^{(h)} | \hat{\Theta}, \mathbf{X}_{t-1}, \mathbf{Z}_1, \mathbf{Z}_2) &= \hat{p}_1 (1 - \hat{p}_1) X_{1,t-1}^{(h)} + \hat{p}_2 (1 - \hat{p}_2) X_{2,t-1}^{(h)} \\
 &\quad + \text{Var}(R_{1,t} + R_{2,t} | \hat{\Theta}, \mathbf{Z}_1, \mathbf{Z}_2)
 \end{aligned}$$

Tables 9 and 10 summarise the mean and variance under different risk profiles and different claim history structure  $(X_{1,t-1}^{(h)}, X_{2,t-1}^{(h)})$ . As TMP and BMP do not depend on claim history, their mean and variance are all the same within the same risk profiles. It is interesting to see that the variance of INAR models are smaller than that of mixed Poisson models in many cases.

### 5 Concluding remarks

In this paper, we consider a new family of bivariate mixed Poisson INAR(1) regression models for modelling multiple time series of different types of claim counts. The proposed family of models accounts for bivariate overdispersion and, similarly to copula-based models, allows for interactions of different signs and magnitude among the two count response variables without using the finite differences of the copula representation which may result in numerical instability in the ML estimation procedure. For illustrative purposes, we derived the BINAR(1)-LN and BINAR(1)-GGA regression models which can be regarded as competitive alternatives to the BINAR(1)-GA regression model for modelling time series of count data. Furthermore, from a computational statistics standpoint, the EM type algorithms we developed for ML estimation of the parameters of all the models were easily implementable and were shown to perform well when we exemplified our approach on LGPIF data from the



**Table 9** Premium calculations from different models: Means

Profile	$(X_{1,t-1}^{(h)}, X_{2,t-1}^{(h)})$	TMP	TINAR	BMP	BINAR
Good	(0,0)	0.0010	0.0009	0.0010	0.0009
	(0,1)	0.0010	0.2856	0.0010	0.2771
	(1,0)	0.0010	0.1194	0.0010	0.1128
	(1,1)	0.0010	0.4040	0.0010	0.3889
Average	(0,0)	0.6361	0.4690	0.6877	0.4951
	(0,1)	0.6361	0.7537	0.6877	0.7713
	(1,0)	0.6361	0.5875	0.6877	0.6070
	(1,1)	0.6361	0.8722	0.6877	0.8832
Bad	(0,0)	2.4873	1.8631	2.6553	1.9484
	(0,1)	2.4873	2.1478	2.6553	2.2246
	(1,0)	2.4873	1.9816	2.6553	2.0603
	(1,1)	2.4873	2.2663	2.6553	2.3364

**Table 10** Premium calculations from different models: Variances

Profile	$(X_{1,t-1}^{(h)}, X_{2,t-1}^{(h)})$	TMP	TINAR	BMP	BINAR
Good	(0,0)	0.0010	0.0009	0.0010	0.0009
	(0,1)	0.0010	0.2045	0.0010	0.2008
	(1,0)	0.0010	0.1053	0.0010	0.1003
	(1,1)	0.0010	0.3090	0.0010	0.3001
Average	(0,0)	0.8746	0.6869	1.1563	0.8221
	(0,1)	0.8746	0.8905	1.1563	1.0220
	(1,0)	0.8746	0.7914	1.1563	0.9215
	(1,1)	0.8746	0.9950	1.1563	1.1214
Bad	(0,0)	6.0635	5.2554	9.6217	6.9969
	(0,1)	6.0635	5.4591	9.6217	7.1968
	(1,0)	6.0635	5.3599	9.6217	7.0962
	(1,1)	6.0635	5.5635	9.6217	7.2961

state of Wisconsin. At this point, it should be noted that we considered the bivariate case and the Gamma and Lognormal correlated random effects for expository purposes. Moreover, the EM estimation framework we proposed is sufficiently flexible and can be used for other continuous mixing densities with a unit mean and, unlike copula-based models, which also allow for both positive and negative correlations, generalizations to any vector size response variables are straightforward. However, in the latter case, EM estimation may be chronologically demanding due to algebraic intractability. Nevertheless, in such cases, due to the structure of the EM algorithm for multivariate INAR(1) models with correlated random effects, the E- and M-steps can be executed in parallel across multiple threads to exploit the processing power available in multicore machines.

Finally, an interesting topic for further research would be to also take into account cross autocorrelation, proceeding along similar lines as in Ref. [4].

**Data Availability** The data that support the findings of this study are available from the corresponding author, Zezhun Chen, upon reasonable request.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A

### A.1 Abbreviations

Here is a table for all the abbreviations used in this paper (Table 11).

### A.2 The Gauss-Hermite quadrature in the high dimensional setting

In this session, we introduce how to transform an integral with respect to multivariate normal density function into a multi-dimension Gauss-Hermite quadrature rule. Starting from one dimensional case, the way we calculate the following integral

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x) \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx,$$

**Table 11** The explanation of the Abbreviations used in

Abbreviation	Interpretation
BP..	Bivariate Mixed Poisson regression model...
BPGA	~with univariate Gamma as mixing density
BPLN	~with bivariate log normal as mixing density
BPGGA	~with bivariate Gaussian Copula paired with univariate Gamma as mixing density
BINAR(1) - ...	Bivariate Integer-valued autoregressive model ...
BINAR(1) - GA	~with BPGA as innovations
BINAR(1) - LN	~with BPLN as innovations
BINAR(1) - GGA	~with BPGGA as innovations

where  $X \sim N(\mu, \sigma^2)$ , is first to make a linear transformation of integrand and then apply the quadrature rule directly:

$$\begin{aligned} \mathbb{E}[h(x)] &= \int_{-\infty}^{\infty} h(\sqrt{2}\sigma y + \mu) \frac{1}{\sqrt{\pi}} \exp\{-y^2\} dy \\ &\approx \frac{1}{\sqrt{\pi}} \sum_{i=1}^n h(\sqrt{\sigma}\xi_i + \mu) w_i \end{aligned}$$

where  $\xi_i$  are the roots of Hermite polynomial of degree  $n$ , with a certain weight  $w_i$ . The quadrature rule approximation of integral will be accurate only when the function  $h$  can be well-approximated by a polynomial of degree  $2n - 1$  or less. Those values can be found from the R function **gauss.quad** in the package **statmod**. The idea to extend the result to high dimensional setting is straightforward. Specifically, we need to first transform the density function into the form  $\exp\{\mathbf{y}^T \mathbf{y}\}$ , where  $\mathbf{y} \in \mathbb{R}^{k \times 1}$ , then the  $k$ -dimensional integral reduces to a  $k$ -fold Gauss-Hermit integral. Suppose the  $k$ -dimensional random vector  $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , where  $\boldsymbol{\mu} \in \mathbb{R}^{k \times 1}$  and  $\boldsymbol{\Sigma} \in \mathbb{R}^{k \times k}$ . Then a linear transform for this random vector is through eigen decomposition of  $\boldsymbol{\Sigma}$  such that

$$\mathbf{x} = \sqrt{2}\mathbf{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{y} + \boldsymbol{\mu},$$

where  $\mathbf{Q} = (\mathbf{v}_1, \dots, \mathbf{v}_k)$  is the matrix formed by eigen vectors and  $\boldsymbol{\Lambda}$  is the diagonal matrix with eigen values  $(\lambda_1, \dots, \lambda_k)$  such that  $\boldsymbol{\Sigma}\mathbf{v}_i = \lambda_i\mathbf{v}_i, \quad i = 1, \dots, k$ . Then the exponent of multivariate normal density becomes

$$\begin{aligned} &\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \\ &= \frac{1}{2}(\sqrt{2}\mathbf{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{y} + \boldsymbol{\mu} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\sqrt{2}\mathbf{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{y} + \boldsymbol{\mu} - \boldsymbol{\mu}) \\ &= (\mathbf{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{y})^T (\mathbf{Q}\boldsymbol{\Lambda}^{-1}\mathbf{Q}^{-1})(\mathbf{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{y}) \\ &= \mathbf{y}^T \boldsymbol{\Lambda}^{\frac{1}{2}}(\mathbf{Q}^T\mathbf{Q})\boldsymbol{\Lambda}^{-1}(\mathbf{Q}^{-1}\mathbf{Q})\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{y} \\ &= \mathbf{y}^T \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{y} \\ &= \mathbf{y}^T \mathbf{y}. \end{aligned}$$

Since  $\boldsymbol{\Sigma}$  is symmetric, then  $\mathbf{Q}^T = \mathbf{Q}^{-1}$ . Finally, the  $k$ -dimensional integral becomes

$$\begin{aligned} \mathbb{E}[h(\mathbf{X})] &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h(\mathbf{x}) \frac{1}{(2\pi)^{\frac{k}{2}}|\boldsymbol{\Sigma}|} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\} dx_1 \dots dx_k \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h(\sqrt{2}\mathbf{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{y} + \boldsymbol{\mu}) \pi^{-\frac{k}{2}} \exp\{-\mathbf{y}^T \mathbf{y}\} dy_1 \dots dy_k \\ &\approx \pi^{-\frac{k}{2}} \sum_{i_1=1}^n \dots \sum_{i_k=1}^n h(\sqrt{2}\mathbf{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi} + \boldsymbol{\mu}) w_{i_1} \dots w_{i_k}. \end{aligned}$$

### A.3 The Gauss-Legendre Quadrature in the high dimensional setting

The extension of the Gauss-Legendre quadrature rule into high dimensional situation is much more straightforward. The following  $m$ -dimensional integral can be approximated by

$$\begin{aligned} & \int_{a_1}^{b_1} \dots \int_{a_m}^{b_m} h(y_1, \dots, y_m) dy_1 \dots dy_m \\ & \approx \left( \prod_{i=1}^m \frac{b_i - a_i}{2} \right) \sum_{i_1=1}^n \dots \\ & \quad \sum_{i_m=1}^n h \left( \frac{b_1 - a_1}{2} \xi_{i_1} + \frac{b_1 + a_1}{2}, \dots, \frac{b_m - a_m}{2} \xi_{i_m} + \frac{b_m + a_m}{2} \right) w_{i_1} \dots w_{i_m} \end{aligned}$$

where  $\xi_i$  are roots of Legendre polynomials of degree  $n$  and  $w_i$  are the corresponding weights. These can also be found easily in R by the ‘gauss.quad’ function in the package ‘statmod’. Similarly, the  $h$  function should be well-approximated by a polynomial of degree  $2n - 1$  or less to ensure accuracy of the approximation.

### A.4 Derivation of conditional expectation

The conditional expectation  $\mathbb{E}_{y,t}^{(j)}[Y_i]$  can be derived explicitly as follows. For simplicity, we just write  $p_1, p_2$  instead of  $p_1^{(j)}, p_2^{(j)}$

$$\begin{aligned} \mathbb{E}_{y,t}^{(j)}[Y_1] &= \frac{1}{\mathcal{P}(\mathbf{X}_t | \mathbf{X}_{t-1})} \sum_{k_1=0}^{s_{1,t}} \sum_{k_2=0}^{s_{2,t}} k_1 \binom{X_{1,t-1}}{k_1} p_1^{k_1} (1-p_1)^{X_{1,t-1}-k_1} \\ & \quad \times f_{p_2}(y_2, X_{2,t-1}) f_{\mathbf{R}}(X_{1,t} - k_1, X_{2,t} - k_2) \\ &= \frac{p_1 X_{1,t-1}}{\mathcal{P}(\mathbf{X}_t | \mathbf{X}_{t-1})} \sum_{k_1=1}^{s_{1,t}} \sum_{k_2=0}^{s_{2,t}} \binom{X_{1,t-1} - 1}{k_1 - 1} p_1^{k_1-1} (1-p_1)^{X_{1,t-1}-k_1} \\ & \quad \times f_{p_2}(y_2, X_{2,t-1}) f_{\mathbf{R}}(X_{1,t} - k_1, X_{2,t} - k_2) \\ &= \frac{p_1 X_{1,t-1}}{\mathcal{P}(\mathbf{X}_t | \mathbf{X}_{t-1})} \sum_{k'_1=0}^{s'_{1,t}} \sum_{k_2=0}^{s_{2,t}} \binom{X_{1,t-1} - 1}{k'_1} p_1^{k'_1} (1-p_1)^{X_{1,t-1}-1-k'_1} \\ & \quad \times f_{p_2}(y_2, X_{2,t-1}) f_{\mathbf{R}}(X_{1,t} - 1 - k'_1, X_{2,t} - k_2) \\ &= p_1 X_{1,t-1} \frac{\mathcal{P}(\mathbf{X}_t - \mathbf{1}_1 | \mathbf{X}_{t-1} - \mathbf{1}_1)}{\mathcal{P}(\mathbf{X}_t | \mathbf{X}_{t-1})}. \end{aligned}$$

**Table 12** Parameter setting for simulation

Parameters	$\boldsymbol{p}$	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\phi_1, \phi_2$	$\rho$	$\mathbf{Z}_1$	$\mathbf{Z}_2$
Values	0.4,0.5	-2,0.8,0.5	-1.5,0.5,0.3	2,3	0.5	MVN( $\boldsymbol{\mu}_1, \mathbf{D}$ )	MVN( $\boldsymbol{\mu}_2, \mathbf{D}$ )

### A.5 Simulation study for Gaussian copula paired with Gamma marginals

This is to verify that the EM algorithms work for both the bivariate Poisson mixture model and the bivariate INAR model when the random effect is characterized by copula. Two random samples of size 500 are generated from these two models separately with pre-specified parameters, which are listed in the Table 12

where  $\boldsymbol{\mu}_1 = (1, 0.3, 0.5)^T$ ,  $\boldsymbol{\mu}_2 = (1, 0.2, 0.4)^T$ ,  $\mathbf{D} = \text{diag}\{0, 1, 1\}$  and MVN stands for multivariate normal distribution. Then each model is fitted by two methods: maximising the incomplete likelihood and EM algorithms. These two methods should give almost the same results for  $\boldsymbol{p}$ ,  $\boldsymbol{\beta}_1$ ,  $\boldsymbol{\beta}_2$  which determine the mean of the model and hence have a relatively large contribution to likelihood. On the other hand, this may not be the case for other parameters  $\phi_1, \phi_2, \rho$  which determine the variation and correlation of the model, and only contribute relative small part of the likelihood. The final log likelihood values would normally be larger than the log likelihood value evaluated at pre-specific parameters. The estimated results are given in Table 13

The difference between estimated parameters  $\phi_1, \phi_2, \rho$  and their true values seems larger than others. This is reasonable because these parameters control the variation of distribution and the log likelihood would be less sensitive to them.

**Table 13** Simulation study for two estimation methods, where OPTIM is the optimization function implemented in R with conjugate gradient (CG) as method

	Parameters	$p$	$\beta_1$	$\beta_2$	$\phi_1, \phi_2$	$\rho$	log likelihood
BPGGA	TRUE		-2 0.8 0.5	-1.5 0.5 0.3	2 3	0.5	-674.2437
	CG		-1.9428 0.8328 0.4920	-1.5659 0.4712 0.2778	3.2321 3.083	0.5666	-672.4953
	EM		-1.9554 0.8337 0.5019	-1.5662 0.4727 0.2782	3.4162 3.1606	0.6444	-672.44
BINAR(1)-GGA	TRUE	0.4 0.5	-2 0.8 0.5	-1.5 0.5 0.3	2 3	0.5	-825.1105
	CG	0.4022 0.5253	-2.0019 0.7560 0.4778	-1.7251 0.4788 0.3315	1.8934 1.2583	0.3665	-821.6101
	EM	0.4014 0.5232	-2.0028 0.7587 0.4793	-1.7198 0.4767 0.3322	2.1990 1.5932	0.4004	-821.7555

## References

1. Abdallah A, Boucher J-P, Cossette H (2016) Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. *Insur Math Econ* 68:120–133
2. Bermúdez L, Karlis D (2011) Bayesian multivariate Poisson models for insurance ratemaking. *Insur Math Econ* 48(2):226–236
3. Bermúdez L, Karlis D (2012) A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking. *Comput Stat Data Anal* 56(12):3988–3999
4. Bermúdez L, Guillén M, Karlis D (2018) Allowing for time and cross dependence assumptions between claim counts in ratemaking models. *Insur Math Econ* 83:161–169
5. Bermúdez L, Karlis D (2021) Multivariate INAR (1) regression models based on the Sarmanov distribution. *Mathematics* 9(5):505
6. Bolancé C, Vernic R (2019) Multivariate count data generalized linear models: three approaches based on the Sarmanov distribution. *Insur Math Econ* 85:89–103
7. Bolancé C, Guillen M, Pitarque A (2020) A Sarmanov distribution with beta marginals: an application to motor insurance pricing. *Mathematics* 8(11)
8. Cameron AC, Li T, Trivedi PK, Zimmer DM (2004) Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. *Econom J* 7(2):566–584
9. Denuit M, Guillen M, Trufin J (2019) Multivariate credibility modelling for usage-based motor insurance pricing with behavioural data. *Ann Actuar Sci* 13(2):378–399
10. Frees EW, Lee G, Yang L (2016) Multivariate frequency-severity regression models in insurance. *Risks* 4(1):4
11. Fung TC, Badescu AL, Lin XS (2019) A class of mixture of experts models for general insurance: application to correlated claim frequencies. *ASTIN Bull J IAA* 49(3):647–688
12. Genest C, Nešlehová J (2007) A primer on copulas for count data. *ASTIN Bull J IAA* 37(2):475–515
13. Gómez-Déniz E, Calderín-Ojeda E (2021) A priori ratemaking selection using multivariate regression models allowing different coverages in auto insurance. *Risks* 9(7):137
14. Gurmu S, Elder J (2000) Generalized bivariate count data regression models. *Econ Lett* 68(1):31–36
15. Jeong H, Dey DK (2021) Multi-peril frequency credibility premium via shared random effects. Available at SSRN 3825435
16. Jeong H, Tzougas G, Fung TC (2023) Multivariate claim count regression model with varying dispersion and dependence parameters. *J R Stat Soc Ser A Stat Soc* 186(1):61–83
17. Karlis D, Pedeli X (2013) Flexible bivariate INAR(1) processes using copulas. *Commun Stat Theory Methods* 42(4):723–740
18. Lee A (1999) Applications: modelling rugby league data via bivariate negative binomial regression. *Aust N Z J Stat* 41(2):141–152
19. Lee GY, Shi P (2019) A dependent frequency-severity approach to modeling longitudinal insurance claims. *Insur Math Econ* 87:115–129
20. Nikoloulopoulos AK, Karlis D (2010) Regression in a copula model for bivariate count data. *J Appl Stat* 37(9):1555–1568
21. Nikoloulopoulos AK (2013) Copula-based models for multivariate discrete response data. In: *Copulae in mathematical and quantitative finance*. Springer, pp 231–249
22. Nikoloulopoulos AK (2016) Efficient estimation of high-dimensional multivariate normal copula models with discrete spatial responses. *Stoch Environ Res Risk Assess* 30(2):493–505
23. Pechon F, Trufin J, Denuit M (2018) Multivariate modelling of household claim frequencies in motor third-party liability insurance. *ASTIN Bull J IAA* 48(3):969–993
24. Pechon F, Denuit M, Trufin J (2019) Multivariate modelling of multiple guarantees in motor insurance of a household. *Eur Actuar J* 9(2):575–602
25. Pechon F, Denuit M, Trufin J (2021) Home and motor insurance joined at a household level using multivariate credibility. *Ann Actuar Sci* 15(1):82–114
26. Pedeli X, Karlis D (2011) A bivariate INAR(1) process with application. *Stat Model* 11(4):325–349
27. Pedeli X, Karlis D (2013) On composite likelihood estimation of a multivariate INAR(1) model. *J Time Ser Anal* 34(2):206–220
28. Shi P (2014) Valdez EA (2014) Longitudinal modeling of insurance claim counts using jitters. *Scand Actuar J* 2:159–179
29. Shi P, Valdez EA (2014) Multivariate negative binomial models for insurance claim counts. *Insur Math Econ* 55:18–29

30. Silva A, Rothstein SJ, McNicholas PD, Subedi S (2019) A multivariate Poisson-log normal mixture model for clustering transcriptome sequencing data. *BMC Bioinform* 20(1):1–11

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.