Multivariate mixed Poisson Generalized Inverse Gaussian INAR(1) regression

Zezhun Chen[1] · **Angelos Dassios[1]** · **George Tzougas[2]**

Abstract

In this paper, we present a novel family of multivariate mixed Poisson-Generalized Inverse Gaussian INAR(1), MMPGIG-INAR(1), regression models for modelling time series of overdispersed count response variables in a versatile manner. The statistical properties associated with the proposed family of models are discussed and we derive the joint distribution of innovations across all the sequences. Finally, for illustrative purposes different members of the MMPGIG-INAR(1) class are fitted to Local Government Property Insurance Fund data from the state of Wisconsin via maximum likelihood estimation.

1 Introduction

In recent years, there has been a growing interest in modelling integer-valued time series of univariate and multivariate count data in a plethora of different scientific fields such as sociology, econometrics, manufacturing, engineering, agriculture, biology, biometrics, genetics, medicine, sports, marketing, and insurance. In particular, regarding the univariate case (Al-Osh and Alzaid 1987) and McKenzie (1985) were the first to consider an INAR(1) model based on the so-called binomial thinning operator. Subsequently, many articles focused on extending this setup by applying different thinning operators or by varying the distribution of innovations. For more details, the interested reader can refer to Weiß (2018), Davis et al. (2016),

✉ Zezhun Chen
  z.chen58@lse.ac.uk

[1] Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, UK

[2] Department of Actuarial Mathematics and Statistics and Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK

Scotto et al. (2015), Weiß (2008) among many more. The INAR(1) model with Poisson marginal distribution (Poisson INAR(1)) has been the most popular choice due to the simplicity of its log-likelihood function that implies that the formality of parameter estimation via maximum likelihood (ML) estimation is straightforward. Also, Freeland and McCabe (2004) considered an extension of the model by allowing for regression specifications on the mean of the Poisson innovation as well as parameter of binomial thinning operator. On the other hand, the literature which focuses on the multivariate case is less developed. In particular, Latour (1997) introduced a multivariate GINAR(p) model with a generalized thinning operator. Karlis and Pedeli (2013) and Pedeli and Karlis (2011, 2013a, b) focused on the diagonal case under which the thinning operators do not introduce cross correlation among different counts. In this case, the dependence structure introduced by innovations. Additionally, Ristić et al. (2012), Popović (2016), Popović et al. (2016) and Nastić et al. (2016) constructed multivariate INAR distributions with cross correlations among counts and random coefficients thinning. Finally, Karlis and Pedeli (2013) extended the setup of the previous articles by allowing for negative cross correlation via a copula-based approach for modelling the innovations.

In this paper, we extend the model proposed by Pedeli and Karlis (2011) by introducing the multivariate mixed Poisson-Generalized Inverse Gaussian INAR(1), MMP-GIG-INAR(1), regression model for multivariate count time series data. The MMP-GIG-INAR(1) is a general three parameter distribution family of INAR(1) models driven by mixed Poisson regression innovations where the mixing densities are chosen from the Generalized Inverse Gaussian class of distributions. Thus, the proposed modelling framework can provide the appropriate level of flexibility for modelling positive correlations of different magnitudes among time series of different types of overdispersed count response variables. In particular, depending on the values taken by the shape parameter, the MMPGIG-INAR(1) family includes many members, such as the mixed Poisson-Inverse Gaussian (PIG), as special cases and several others as limiting cases, such as the Negative Binomial, or Poisson-Gamma, the Poisson-Inverse Gamma (PIGA), the Poisson-Inverse Exponential, the Poisson-Inverse Chi Squared and the Poisson-Scaled Inverse Chi Squared distributions. Therefore, it can accommodate different levels of overdispersion depending on the chosen parametric form of the mixing density. Furthermore, the MMPGIG-INAR(1) family of models is constructed by assuming that the probability mass function (pmf) of the MMPGIG innovations is parameterized in terms of the mean parameter which results in a more orthogonal parameterization that facilitates maximum likelihood (ML) estimation when regression specifications are allowed for the mean parameters of the MMPGIG-INAR(1) regression model. For expository purposes, we derive the joint probability mass functions and the derivatives of several special cases of the MMPGIG-INAR(1) family which are used as innovations. These models are fitted to time series of claim count data from the Local Government Property Insurance Fund (LGPIF) data in the state of Wisconsin. At this point it is worth noting that modelling the correlation between different types of claims from the same and/or different types of coverage it is very important from a practical business standpoint. Many articles have been devoted to this topic, see for example, Bermúdez and Karlis (2011), Bermúdez and Karlis (2012), Shi and Valdez (2014a, b), Abdallah et al. (2016), Bermúdez and Karlis (2017), Pechon et al. (2018),

Pechon et al. (2019), Bolancé and Vernic (2019), Denuit et al. (2019), Fung et al. (2019), Bolancé et al. (2020), Pechon et al. (2021), Jeong and Dey (2021), Gómez-Déniz and Calderín-Ojeda (2021), Tzougas and di Cerchiara (2021a, b). However, with the exception of very few articles, such as Bermúdez et al. (2018) and Bermúdez and Karlis (2021), the construction of bivariate INAR(1) models which can capture the serial correlation between the observations of the same policyholder over time and the correlation between different claim types remains a largely uncharted territory. This is an additional contribution of this study.

The rest of the paper proceeds as follows. Section 2 presents the derivation of the MMPGIG-INAR(1) model. Statistical properties of the MMPGIG innovations are discussed in Sect. 3. In Sect. 4, we present a description of the alternative special cases of the MMPGIG-INAR(1) family. Section 5 discusses the parameter estimation for these models based on the maximum likelihood method and integer-valued prediction. Section 6 contains our empirical analysis for the LGPIF data set. Finally, concluding remarks are given in Sect. 7.

2 Generalized setting

Let $\mathbf{X}$ and $\mathbf{R}$ be non-negative integer-valued random vectors in $\mathbb{R}^m$. Let $\mathbf{P}$ be a diagonal matrix in $\mathbb{R}^{m \times m}$ with elements $p_i \in (0, 1)$. The multivariate Poisson-Generalized Inverse Gaussian INAR(1) is defined as

$$\mathbf{X}_t = \mathbf{P} \circ \mathbf{X}_{t-1} + \mathbf{R}_t = \begin{bmatrix} p_1 & 0 & \dots & 0 & 0 \\ 0 & p_2 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & & p_m \end{bmatrix} \circ \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \\ \vdots \\ X_{m,t-1} \end{bmatrix} + \begin{bmatrix} R_{1,t} \\ R_{2,t} \\ \vdots \\ R_{m,t} \end{bmatrix} \qquad (2.1)$$

where the thinning operator $\circ$ is the widely used binomial thinning operator such that $p_i \circ X_{i,t} = \sum_{k=1}^{X_{i,t}} U_k$ where 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

$$f_{p_i}(x, X_{i,t}) = \binom{X_{i,t}}{x} p_i^x (1 - p_i)^{X_{i,t} - x} \qquad (2.2)$$

Note that given $X_{i,t}, X_{j,t}$ $i \neq j$, $p_i \circ X_{i,t}$ and $p_j \circ X_{j,t}$, are independent of each other. To adapt the heteroscedasticity arising from the data, $\{R_{i,t}\}_{i=1,\dots,m}$ are mixed Poisson random variables $Po(\theta_t \lambda_{i,t})$ with the random effect θ_t. The rate $\lambda_{i,t}$ is characterized by its observed covariate $z_{i,t} \in \mathbb{R}^{a_i \times 1}$ for some positive integer a_i and they are connected through a log link function such that $\log(\lambda_{i,t}) = z_{i,t}^T \beta_i$ where $\beta_i \in \mathbb{R}^{a_i \times 1}$. Furthermore, $\{R_{i,t}\}_{i=1,\dots,m}$ share the same random effect θ_t with mixing distribution $G(\theta)$, which means the dependent structure among $X_{i,t}$ can be controlled by the choice of distribution and its corresponding size of parameters. The joint distribution of $\mathbf{R}_t$ is

$$f_\phi(\mathbf{k}, t) = \mathcal{P}(R_{1,t} = k_1, \ldots, R_{m,t} = k_m)$$
$$= \mathbb{E}\left[\mathcal{P}(R_{1,t} = k_1, \ldots, R_{m,t} = k_m | \theta_t)\right]$$
$$= \prod_{j=1}^{m} \frac{\lambda_{j,t}^{k_j}}{k_j!} \int_0^\infty e^{-\theta \sum_{i=1}^m \lambda_{i,t}} \theta^{\sum_{i=1}^m k_i} dG(\theta) \tag{2.3}$$

We let θ_t be a continuous random variable from the Generalized Inverse Gaussian distribution with density function $g(\theta)$

$$g(\theta) = \frac{(\psi/\chi)^{\frac{\nu}{2}}}{2K_\nu(\sqrt{\psi\chi})} \theta^{\nu-1} \exp\left\{-\frac{1}{2}\left(\psi\theta + \frac{\chi}{\theta}\right)\right\}, \tag{2.4}$$

where $-\infty \leq \nu \leq \infty, \psi > 0, \chi > 0$ and $K_\nu(\omega)$ is the modified Bessel function of the third kind of order ν and argument ω such that

$$K_\nu(\omega) = \int_0^\infty z^{\nu-1} \exp\left\{-\frac{1}{2}\omega\left(z + \frac{1}{z}\right)\right\} dz$$

The Generalized Inverse Gaussian distribution is a widely used family. For example, it includes the Inverse Gaussian as special case and the Gamma and Inverse Gamma as limiting cases. To avoid identification problems for mixed Poisson regression random variable $\mathbf{R}_t$, the mean of θ_t is restricted to one, i.e. $\mathbb{E}[\theta_t] = 1$, and all the parameters ν, ψ, χ will be either fixed or a function of another parameter ϕ. With these two constraints, there is only one parameter that is free to vary. (e.g. for Inverse Gaussian distribution, $\nu = -\frac{1}{2}$ and $\psi = \chi = \phi$). The joint distribution of $\mathbf{R}_t$ becomes an MPGIG distribution

$$f_\phi(\mathbf{k}, t) = \frac{(\psi/\chi)^{\frac{\nu}{2}}}{2K_\nu(\sqrt{\psi\chi})} \prod_{j=1}^m \frac{\lambda_{j,t}^{k_j}}{k_j!} \int_0^\infty e^{-\theta \sum_{i=1}^m \lambda_{i,t}} \theta^{\sum_{i=1}^m k_i} \theta^{\nu-1} \exp\left\{-\frac{1}{2}\left(\psi\theta + \frac{\chi}{\theta}\right)\right\} d\theta$$
$$= \frac{(\psi/\chi)^{\frac{\nu}{2}}}{(\Delta/\chi)^{\frac{\nu+\sum_i k_i}{2}}} \frac{K_{\nu+\sum_i k_i}(\sqrt{\Delta\chi})}{K_\nu(\sqrt{\psi\chi})} \prod_{j=1}^m \frac{\lambda_{j,t}^{k_j}}{k_j!}, \tag{2.5}$$

where $\Delta = \psi + 2\sum_{i=1}^m \lambda_{i,t}$. In Sect. 5, we will discuss in detail the distribution function $f_\phi(\mathbf{k}, t)$ for some special cases. Finally, it should be noted that several articles discuss multivariate versions of MPGIG distribution and/or the MPIG distribution which is a special case for $\nu = -0.5$, see, for instance, Barndorff-Nielsen et al. (1992), Ghitany et al. (2012) Amalia et al. (2017), Mardalena et al. (2020), Tzougas and di Cerchiara (2021b) and Mardalena et al. (2021). However, this is the first time that the MMPGIG-INAR(1) distribution family of INAR(1) models driven by mixed Poisson regression innovations are considered for modelling time series of count response variables.

3 Properties of innovations $\mathbf{R}_t$

Proposition 3.1 (The moments of $\mathbf{R}_t$) *The mean, variance of $R_{i,t}$ and covariance between $R_{i,t}, R_{j,t}, i \neq j$ are given by*

$$\mathbb{E}[R_{i,t}] = \mathbb{E}[\mathbb{E}[R_{i,t}|\theta_t]] = \lambda_{i,t}$$
$$Var(R_{i,t}) = Var(\mathbb{E}[R_{i,t}|\theta_t]) + \mathbb{E}[Var(R_{i,t}|\theta_t)]$$
$$= \sigma_\theta^2 \lambda_{i,t}^2 + \lambda_{i,t} \tag{3.1}$$
$$Cov(R_{i,t}, R_{j,t}) = Cov(\mathbb{E}[R_{i,t}|\theta_t], \mathbb{E}[R_{j,t}|\theta_t]) + \mathbb{E}[Cov(R_{i,t}, R_{j,t}|\theta_t)]$$
$$= \lambda_{i,t} \lambda_{j,t} \sigma_\theta^2$$

where σ_θ^2 is the variance for the random effect θ_t and $i, j = 1, \ldots, m$.

Proposition 3.2 (*Marginal property*) *The joint distribution function $f_\phi(\mathbf{k}, t)$ is closed to marginalization, i.e. the marginal distribution for $R_{i,t}$ is given by $f_\phi(k_i, t)$ such that*

$$f_\phi(k_i, t) = \int_0^\infty \frac{\lambda_{i,t}^{k_i}}{k_i!} \theta^{k_i} e^{-\lambda_{i,t}\theta} dG(\theta)$$
$$= \frac{(\psi/\chi)^{\frac{\nu}{2}}}{((\psi + \lambda_{i,t})/\chi)^{\frac{\nu+k_i}{2}}} \frac{K_{\nu+k_i}(\sqrt{(\psi + \lambda_{i,t})\chi})}{K_\nu(\sqrt{\psi\chi})} \frac{\lambda_{i,t}^{k_i}}{k_i!} \tag{3.2}$$

which is a univariate mixed Poisson regression random variable. In general, this result is valid for any m'-variate mixed Poisson regression random variable with $m' < m$

Proof We will show the result for univariate case. The m'-variate case can be derived similarly by reducing the number of following sum to $m - m'$

$$f_\phi(k_i, t) = \sum_{k_1=0} \cdots \sum_{k_{i-1}=0} \sum_{k_{i+1}=0} \cdots \sum_{k_m=0} f_\phi(\mathbf{k}, t)$$
$$= \int_0^\infty \prod_{j\neq i} \left(\sum_{k_j=0}^\infty \frac{e^{-\theta\lambda_{j,t}}(\theta\lambda_{j,t})^{k_j}}{k_j!} \right) \frac{e^{-\theta\lambda_{i,t}}(\theta\lambda_{i,t})^{k_i}}{k_i!} dG(\theta)$$
$$= \int_0^\infty \frac{\lambda_{i,t}^{k_i}}{k_i!} \theta^{k_i} e^{-\lambda_{i,t}\theta} dG(\theta)$$

$\square$

The marginalization property can enable, for example insurers, to easily price those policyholders who only engage in some but not all lines of business. The last property is about the identifiability of $\mathbf{R}_t$, which will ensure the uniqueness of the model.

Proposition 3.3 (Identifiability of joint distribution $\mathbf{R}_t$) *Assume that the covariate space $\mathbf{z}_t = (z_{1,t}, \ldots, z_{m,t})$ is of full rank. Denote the parameter set $\Theta_R = \{\beta_i, \phi | i = 1, \ldots, m\}$ and $\tilde{\Theta}_R = \{\tilde{\beta}_i, \tilde{\phi} | i = 1, \ldots, m\}$, the joint distribution $f_\phi(\mathbf{k}, t)$ is identifiable such that*

$$f_\phi(\mathbf{k}, t) = f_{\tilde{\phi}}(\mathbf{k}, t)$$

if and only if $\Theta_R = \tilde{\Theta}_R$.

Proof With the assumption that the covariate $\mathbf{z}$ is full rank and the log-link function is monotonic such that $\log(\lambda_{i,t}) = z_{i,t}^T \beta_i$, it is obvious that the identification problem for the mixed Poisson regression random variable $\mathbf{R}_t$ reduces to identification for mixed Poisson random variable (without regression), which means the set of parameter can be re-parametrized as $\Theta_R^* = \{\lambda_{i,t}, \phi | i = 1, \ldots, m\}$ and $\tilde{\Theta}_R^* = \{\tilde{\lambda}_{i,t}, \tilde{\phi} | i = 1, \ldots, m\}$.

Then the 'if' statement is obvious since the same set of parameters will definitely lead to the same joint distribution function. For the 'only if' statement, to match two distribution functions, all the moments (mean, variance, covariance) must reconcile. From the moment properties above, matching the $\mathbb{E}[R_{i,t}]$ will lead to $\lambda_{i,t} = \tilde{\lambda}_{i,t}$. Likewise, given that the first moment is matched, only $\phi = \tilde{\phi}$ will lead to the same $Var(R_{i,t})$. Matching these moments already leads to $\Theta_R^* = \tilde{\Theta}_R^*$, then the covariance $Cov(R_{i,t}, R_{j,t})$ must match with each other. $\qquad\square$

4 Model specification

The distributional properties of $\mathbf{X}_t$, in particular the correlation structure and 'tailedness' of the distribution, are mainly determined by the innovation $\mathbf{R}_t$, more specifically, the mixing density $g(\theta)$. On the other hand, the explicit form of the derivatives of $f_\phi(\mathbf{k}, t)$ can significantly accelerate the computational speed when performing estimation. Hence, the distribution function $f_\phi(\mathbf{k}, t)$ as well as its derivatives are derived for two limiting cases (Gamma, Inverse Gamma) and some other special cases (GIG with unit mean and different values of v). Throughout this session, we define $S_t^\lambda = \sum_{i=1}^m \lambda_{i,t}$ and $S^k = \sum_{i=1}^m k_i$.

4.1 Mixing by Gamma distribution

If $\mathbf{R}_t$ is univariate, the resulting distribution is known as the negative binomial distribution and this result can be easily extended to the multivariate case which is called the multivariate negative binomial distribution (see e.g. Marshall and Olkin 1990; Boucher et al. 2008; Cheon et al. 2009). The gamma density is obtained by letting

$\nu = \phi, \psi = 2\phi$ and $\chi = 0$ in generalized Inverse Gaussian density in Sect. 2.4. The resulting mixing density has the following form:

$$g(\theta) = \frac{\phi^\phi}{\Gamma(\phi)} \theta^{\phi-1} e^{-\phi\theta} \tag{4.1}$$

with unit mean and variance $\frac{1}{\phi}$. Then the expectation 2.3 can be evaluated explicitly

$$
\begin{aligned}
f_\phi(\mathbf{k}, t) &= \prod_{i=1}^{m} \frac{\lambda_{i,t}^{k_i}}{k_i!} \mathbb{E}[e^{-(S_t^\lambda)\theta} \theta^{S^k}] \\
&= \frac{\Gamma(\phi + S^k)}{\Gamma(\phi) \prod_{i=1}^{m} \Gamma(k_i + 1)} \frac{\phi^\phi \prod_{i=1}^{m} \lambda_{i,t}^{k_i}}{(\phi + S_t^\lambda)^{\phi+S^k}}
\end{aligned}
\tag{4.2}
$$

Proposition 4.1 *The derivatives of the distribution function $f_\phi(\mathbf{k}, t)$ with respect to $\Theta_R = \{\phi, \beta_i \mid i = 1, \ldots, m\}$ when $\theta_t \sim Gamma(\phi, \phi)$ are given by*

$$
\begin{aligned}
\frac{\partial f_\phi(\mathbf{k}, t)}{\partial \phi} &= f_\phi(\mathbf{k}, t) \left(\sum_{n=1}^{S^k} \frac{1}{n + \phi - 1} + \log\left(\frac{\phi}{\phi + S_t^\lambda}\right) + \frac{\sum_{i=1}^{m}(\lambda_{i,t} - k_i)}{\phi + S_t^\lambda} \right) \\
\frac{\partial f_\phi(\mathbf{k}, t)}{\partial \beta_i} &= f_\phi(\mathbf{k}, t) \left(\frac{k_i}{\lambda_{i,t}} - \frac{\phi + S^k}{\phi + S_t^\lambda} \right) \lambda_{i,t} z_{i,t} ,
\end{aligned}
\tag{4.3}
$$

where the sum $\sum_{n=1}^{S^k} \frac{1}{n+\phi-1} = 0$ when $S^k = 0$.

Proof The derivatives $\frac{\partial f_\phi(\mathbf{k},t)}{\partial \beta_i}$ can be figured out easily except $\frac{\partial f_\phi(\mathbf{k},t)}{\partial \phi}$ which involves the gamma function. The derivative of the gamma function can be derived by utilizing the alternative Weierstrass's definition such that

$$\Gamma(z + 1) = e^{-\gamma z} \prod_{n \geq 1} \left(1 + \frac{z}{n}\right)^{-1} e^{\frac{z}{n}},$$

which is valid for all complex number z except non-positive integers and γ is Euler–Mascheroni constant. Then the derivative can be derived by differentiating its log transform $\log \Gamma(z + 1)$, which leads to the series expansion of digamma function

$$\Psi(z + 1) = \frac{\Gamma'(z + 1)}{\Gamma(z + 1)} = -\gamma + \sum_{n \geq 1} \left(\frac{1}{n} - \frac{1}{n + z}\right)$$

Then the derivative $\frac{\partial f_\phi(\mathbf{k},t)}{\partial \phi}$ can be derived steps by steps. First let us simplify the expression of $f_\phi(\mathbf{k}, t)$ such that

$$f_\phi(\mathbf{k}, t) = c_1 \frac{N(\phi)}{D(\phi)}$$

$$c_1 = \prod_{i=1}^{m} \frac{\lambda_{i,t}^{k_i}}{\Gamma(k_i + 1)}, \ N(\phi) = \Gamma(\phi + S^k)\phi^\phi, \ D(\phi) = \Gamma(\phi)\big(\phi + S_t^\lambda\big)^{\phi + S^k}$$

The derivative is then

$$\frac{\partial f_\phi(\mathbf{k}, t)}{\partial \phi} = c_1 \frac{N'(\phi)D(\phi) - N(\phi)D'(\phi)}{D^2(\phi)}$$

$$= c_1 \frac{N(\phi)}{D(\phi)} \left(\sum_{n \geq 1} \left(\frac{1}{n + \phi - 1} - \frac{1}{n + \phi + S^k - 1} \right) \right.$$

$$\left. + 1 + \log\phi - \log(\phi + S_t^\lambda) - \frac{\phi + S^k}{\phi + S_t^\lambda} \right)$$

$$\square$$

4.2 Mixing by Inverse Gamma

The Inverse gamma distribution, which is another limiting case of generalized Inverse Gaussian distribution, is discussed in Sect. 9.3 (Johnson et al. 1995). Inverse gamma random variable has a relatively thicker right tail and a low probability in taking the values closed to 0. In this case, the density function $g(\theta)$ is obtained by letting $\psi = 0$, $\chi = 2\phi$ and $\nu = -\phi - 1$ such that

$$g(\theta) = \frac{\phi^{\phi+1}}{\Gamma(\phi + 1)} \theta^{-\phi-2} e^{-\frac{\phi}{\theta}}, \tag{4.4}$$

with mean 1 and variance $\frac{1}{\phi - 1}$ for $\phi > 1$. It is also called the reciprocal gamma distribution such that $\theta = 1/x$ where $x \sim \text{Gamma}(\phi + 1, \phi)$. The distribution function $f_\phi(\mathbf{k}, t)$ becomes

$$f_\phi(\mathbf{k}, t) = \prod_{i=1}^{m} \frac{\lambda_{i,t}^{k_i}}{k_i!} \mathbb{E}\left[e^{-S_t^\lambda \theta_t} \theta_t^{S^k} \right]$$

$$= \frac{2K_\nu(\omega)}{\Gamma(\phi + 1) \prod_{i=1}^{m} \Gamma(k_i + 1)} \frac{\phi^{\frac{\nu}{2}+\phi+1} \prod_{i=1}^{m} \lambda_{i,t}^{k_i}}{(S_t^\lambda)^{\frac{\nu}{2}}}, \tag{4.5}$$

where $\nu = S^k - \phi - 1$ and $\omega = 2\sqrt{\phi S_t^\lambda}$. The derivatives of $f_\phi(\mathbf{k}, t)$ with respect to the parameter set $\Theta_R = \{\phi, \beta_i \mid i = 1, \ldots, m\}$ are given by

$$\frac{\partial f_\phi(\mathbf{k}, t)}{\partial \phi} = \left(\log \frac{\omega}{2} + \frac{S^k + \phi + 1}{2\phi} + \frac{\partial \log K_\nu(\omega)}{\partial \phi} - \Psi(\phi + 1) \right) f_\phi(\mathbf{k}, t)$$

$$\frac{\partial f_\phi(\mathbf{k}, t)}{\partial \beta_i} = \left(\frac{k_i}{\lambda_{i,t}} f_\phi(\mathbf{k}, t) - \frac{k_i + 1}{\lambda_{i,t}} f_\phi(\mathbf{k} + \mathbf{1}_i, \phi) \right) \lambda_{i,t} z_{i,t} \tag{4.6}$$

In this case, numerical differentiation is applied to calculate $\frac{\partial \log K_\nu(\omega)}{\partial \phi}$ since the parameter ϕ appears both in the order ν and argument ω of the modified Bessel function $K_\nu(\omega)$.

4.3 Mixing by Generalized Inverse Gaussian

Likewise, if $\mathbf{R}_t$ is univariate, the distribution of $\mathbf{R}_t$ is known as the Poisson Generalized Inverse Gaussian distribution. To comply with constraints we made in Sect. 2, the mixing density function has following form

$$g(\theta) = \frac{c^\nu}{2K_\nu(\phi)} \theta^{\nu-1} \exp\left\{ -\frac{\phi}{2} \left(c\theta + \frac{1}{c\theta} \right) \right\} \tag{4.7}$$

with unit mean and variance $var(\theta_t) = \frac{1}{c^2} + \frac{2(\nu+1)}{c\phi} - 1$, where $c = \frac{K_{\nu+1}(\phi)}{K_\nu(\phi)}$, $\phi > 0$ and $\nu \in \mathbb{R}$. Then the distribution function $f_\phi(\mathbf{k}, t)$ becomes

$$\begin{aligned} f_\phi(\mathbf{k}, t, \nu) &= \prod_{i=1}^m \frac{\lambda_{i,t}^{k_i}}{k_i!} \mathbb{E}\left[e^{-\theta_t S_t^\lambda} \theta_t^{S^k} \right] \\ &= \frac{K_p(\sqrt{ab})}{K_\nu(\phi)} c^\nu \left(\frac{b}{a} \right)^{\frac{p}{2}} \prod_{i=1}^m \frac{\lambda_{i,t}^{k_i}}{k_i!} \end{aligned} \tag{4.8}$$

where $a = \phi c + 2S_t^\lambda$, $b = \frac{\phi}{c}$ and $p = S^k + \nu$. Furthermore, we let ν be constant and fixed in order to avoid potential identification problems which may appear when performing estimation. In general, however, the derivative with respect to ϕ is really hard to find since the constant c involves the Bessel function. On the other hand, it is worth noting that $var(\theta_t)$ is roughly unbounded when $\nu \in [-2, 0]$ and the skewness and kurtosis are decreasing with respect to ν, which can be easily verified by some statistical software on computer. So, we will discuss cases where $\nu = -\frac{1}{2}, -\frac{3}{2}, -\frac{3}{4}$, two of which have 'explicit' distributions in the sense that the constant c can be evaluated in closed form.

4.3.1 Generalized Inverse Gaussian with $\nu = -\frac{1}{2}$

In this case, the resulting distribution known as the Poisson Inverse Gaussian distribution is investigated by many authors (see,e.g Sichel 1974, 1982; Atkinson and Yeh 1982; Stein and Juritz 1988 among others). When $\nu = -\frac{1}{2}$, $c = 1$ and the distribution function f becomes

$$f_\phi(\mathbf{k}, t) = \prod_{i=1}^{m} \frac{\lambda_{i,t}^{k_i}}{k_i!} \sqrt{\frac{2}{\pi}} \phi^{\frac{1}{2}} e^{\phi} K_p(\sqrt{\phi(\phi + 2S_t^\lambda)}) \left(\frac{\phi}{\phi + 2S_t^\lambda} \right)^{\frac{p}{2}} \tag{4.9}$$

For convenience, we reparametrize the above density by squaring the parameter ϕ such that

$$f_\phi(\mathbf{k}, t) = \prod_{i=1}^{m} \frac{\lambda_{i,t}^{k_i}}{k_i!} \sqrt{\frac{2}{\pi}} \phi e^{\phi^2} K_p(\phi \Delta) \left(\frac{\phi}{\Delta} \right)^{\nu} \tag{4.10}$$

where $p = S^k - \frac{1}{2}$ and $\Delta = \sqrt{\phi^2 + 2S^\lambda}$. The derivatives of $f_\phi(\mathbf{k}, t)$ with respect to different parameters can be derived by making use of the derivative of $K_\nu(\omega)$ with respect to its argument such that

$$\frac{\partial K_\nu(\omega)}{\omega} = \frac{\nu}{\omega} K_\nu(\omega) - K_{\nu+1}(\omega), \tag{4.11}$$

then it leads to the following derivatives

$$\begin{aligned}
\frac{\partial f_\phi(\mathbf{k}, t)}{\partial \phi} &= \left(2\phi + \frac{1 + 2\nu}{\phi} \right) f_\phi(\mathbf{k}, t) - \left(\phi + \frac{\Delta^2}{\phi} \right) \frac{k_1 + 1}{\lambda_{1,t}} f_\phi(\mathbf{k} + \mathbf{1}_1, \phi), \\
\frac{\partial f_\phi(\mathbf{k}, t)}{\partial \beta_i} &= \left(\frac{k_i}{\lambda_{i,t}} f_\phi(\mathbf{k}, t) - \frac{k_i + 1}{\lambda_{i,t}} f_\phi(\mathbf{k} + \mathbf{1}_i, \phi) \right) \lambda_{i,t} z_{i,t}
\end{aligned} \tag{4.12}$$

where $\mathbf{1}_i = (0, \dots, 0, 1, 0, \dots, 0)^T \in \mathbb{R}^{m \times 1}$ is vector with i-th element being one and 0 elsewhere.

4.3.2 Generalized Inverse Gaussian with $\nu = -\frac{3}{2}$

In this case, the constant $c = \frac{\phi}{1+\phi}$ and the variance $var(\theta_t) = \frac{1}{\phi}$ which is exactly the same as the variance of Inverse Gaussian case but the random effect θ_t will in general have larger skewness and kurtosis. The resulting distribution function is

$$f_\phi(\mathbf{k}, t) = \prod_{i=1}^{m} \frac{\lambda_{i,t}^{k_i}}{k_i!} \sqrt{\frac{2}{\pi}} (\phi + 1)^{S^k - 1} e^{\phi} \omega^{-p} K_p(\omega). \tag{4.13}$$

where $p = S^k - \frac{3}{2}$ and $\omega = \sqrt{\phi^2 + 2(\phi + 1)S_t^\lambda}$. The derivatives with respect to different parameters can be derived similar to that of Inverse Gaussian case

$$\begin{aligned}
\frac{\partial f_\phi(\mathbf{k}, t)}{\partial \phi} &= \left(\frac{\phi + S^k}{\phi + 1} \right) f_\phi(\mathbf{k}, t) - \left(\frac{k_1 + 1}{\lambda_{1,t}} \frac{\phi + S_t^\lambda}{\phi + 1} \right) f_\phi(\mathbf{k} + \mathbf{1}_1, t) \\
\frac{\partial f_\phi(\mathbf{k}, t)}{\partial \beta_i} &= \left(\frac{k_i}{\lambda_{i,t}} f_\phi(\mathbf{k}, t) - \frac{k_i + 1}{\lambda_{i,t}} f_\phi(\mathbf{k} + \mathbf{1}_i, \phi) \right) \lambda_{i,t} z_{i,t}
\end{aligned} \tag{4.14}$$

Table 1 Parametrization of mixing density based on GIG density 2.4

Mixing density	ψ	χ	ν	Range of parameter
Gamma	2ϕ	0	ϕ	$\phi > 0$
GIG with unit mean	$c\phi$	$\frac{\phi}{c}$	fixed constant in $\mathbb{R}$	$c = \frac{K_\nu(\phi+1)}{K_\nu(\phi)}, \phi > 0$
Inverse Gamma	0	2ϕ	$-\phi - 1$	$\phi > 1$

Table 2 Moments for the random effect θ_t. Ex.Kurtosis = Kurtosis $- 3$

Mixing density $g(\theta)$	Variance	Skewness	Ex.Kurtosis
Gamma	$\frac{1}{\phi}$	$\frac{2}{\sqrt{\phi}}$	$\frac{6}{\phi}$
Inverse Gaussian	$\frac{1}{\phi^2}$	$\frac{3}{\phi}$	$\frac{15}{\phi^2}$
GIG $\nu = -\frac{3}{2}$	$\frac{1}{\phi}$	$\frac{\phi^{\frac{3}{2}} + \phi^{\frac{1}{2}}}{\phi^2}$	$\frac{3 + 12\phi + 15\phi^2}{\phi^3}$
Inverse Gamma	$\frac{1}{\phi-1}$	$\frac{4\sqrt{\phi-1}}{\phi-2}$	$\frac{6(5\phi-1)}{(\phi-2)(\phi-3)}$

The remaining case where $\nu = -\frac{3}{4}$ cannot be simplified since the $c = \frac{K_{1/4}(\phi)}{K_{3/4}(\phi)}$ cannot be written down in terms of basic functions. Hence numerical differentiation has to be applied when evaluating $\frac{\partial f_\phi(\mathbf{k}, t)}{\partial \phi}$ and $\frac{\partial f_\phi(\mathbf{k}, t)}{\partial \beta_i}$. Finally, Table 1 summarise the parametrization of all mixing densities and Table 2 shows the moments formula for each mixing density.

Although the formula for variances is slightly different due to its parametrization, they can be easily reparameterized and compared with each other. It turns out that the Inverse Gamma has the largest skewness and kurtosis while the Gamma density has the smallest, which means the 'tailedness' of those density increases in a 'top-down' order according to the Table. Hence, one can choose different density to accommodate different tail structure encountered in real data.

5 Model fitting and prediction

5.1 Maximum likelihood estimation for the MMPGIG-INAR(1) model

In this section, we derive the log likelihood function and score function of the MMPGIG-INAR(1) model defined above for the general case. Let the whole parameter set be $\Theta = \{p_i, \beta_i, \phi | i = 1, \ldots, m\}$ and then the log likelihood function $\ell(\Theta)$ for this discrete Markov chain is just the product of their conditional probability function such that $\ell(\Theta) = \prod_t \mathcal{P}_\Theta(\mathbf{X}_t | \mathbf{X}_{t-1})$, where the conditional probability is the convolution of m+1 distribution functions such that

$$\mathcal{P}(\mathbf{X}_t|\mathbf{X}_{t-1}) = \mathbb{E}\left[\prod_{i=1}^{2} \mathcal{P}(p_i \circ X_{i,t-1} + R_{i,t} = X_{i,t-1}|X_{i,t-1}, \theta_t)\right]$$

$$= \mathbb{E}\left[\prod_{i=1}^{m} \sum_{k=0}^{s_i} f_{p_i}(k, X_{i,t-1}) f_{R_i}(X_{i,t} - k, t)\right], \quad s_i = \min\{X_{i,t-1}, X_{i,t}\}$$

$$= \sum_{k_1=0}^{s_1} \cdots \sum_{k_m=0}^{s_m} f_\phi(\mathbf{k}, t) \prod_{i=1}^{m} f_{p_i}(X_{i,t} - k_i, X_{i,t-1}),$$

(5.1)

where the expectation is taken with respect to the random variable θ_t. The following proposition gives $\ell(\Theta)$ and its score functions.

Proposition 5.1 *Suppose there is a multivariate random sequence* $(\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n)$ *generated from the MMPGIG-INAR*(1) *model, the log likelihood function* $\ell(\Theta)$ *and score functions are given by*

$$\ell(\Theta) = \sum_{t=1}^{n} \log \mathcal{P}(\mathbf{X}_t|\mathbf{X}_{t-1})$$

$$= \sum_{t=1}^{n} \log \sum_{k_1=0}^{s_1} \cdots \sum_{k_m=0}^{s_m} f_\phi(\mathbf{k}, t) \prod_{i=1}^{m} f_{p_i}(X_{i,t} - k_i, X_{i,t-1}) \qquad (5.2)$$

$$\frac{\partial \ell(\Theta)}{\partial \vartheta} = \sum_{t=1}^{n} \frac{1}{\mathcal{P}(\mathbf{X}_t|\mathbf{X}_{t-1})} \frac{\partial \mathcal{P}(\mathbf{X}_t|\mathbf{X}_{t-1})}{\partial \vartheta}, \quad \vartheta \in \Theta$$

The derivatives inside the sum are given by

$$\frac{\partial \mathcal{P}(\mathbf{X}_t|\mathbf{X}_{t-1})}{\partial p_j} = \sum_{k_1=0}^{s_1} \cdots \sum_{k_m=0}^{s_m} f_\phi(\mathbf{k}, t) \frac{\partial f_{p_j}(X_{ij,t} - k, X_{j,t-1})}{\partial p_j} \prod_{i \neq j} f_{p_j}(X_{i,t} - k_i, X_{i,t-1})$$

$$\frac{\partial \mathcal{P}(\mathbf{X}_t|\mathbf{X}_{t-1})}{\partial \vartheta_1} = \sum_{k_1=0}^{s_1} \cdots \sum_{k_m=0}^{s_m} \frac{\partial f_\phi(\mathbf{k}, t)}{\vartheta_1} \prod_{i=1}^{m} f_{p_i}(X_{i,t} - k_i, X_{i,t-1})$$

$$\vartheta_1 \in \{\beta_1, \beta_2, \phi\}$$

(5.3)

where the derivative $\frac{\partial f_{p_j}(\omega, X_{j,t-1})}{\partial p_j}$ *has the same form for all* $j = 1, \ldots, m.$

$$\frac{\partial f_{p_j}(\omega, X_{j,t})}{\partial p_j} = f_{p_j}(\omega, X_{j,t}) \frac{\omega - p_j X_{j,t}}{p_j(1 - p_j)}$$

The derivatives $\frac{\partial f_\phi(\mathbf{k}, t)}{\vartheta_1}$ are already discussed in Sect. 4 for different cases. Hence, the maximum likelihood estimators can be obtained through numerical algorithms, for example Newton-raphson, Quasi-Newton and so on. However, optimization will be computational intensive as m increases. One can solve this issue by adopting the

composite likelihood method introduced in Pedeli and Karlis (2013a), where the high dimensional likelihood function was reduced to a sum of bivariate cases.

5.2 Integer-valued prediction

Based on the estimates obtained by maximum likelihood and the random sequence $(\mathbf{X}_1, \dots, \mathbf{X}_n)$, the h-steps ahead distribution of $\mathbf{X}_{n+h}$ conditional on $\mathbf{X}_n$ is given by

$$\mathbf{X}_{n+h} \stackrel{D}{=} \hat{\mathbf{P}}^h \circ \mathbf{X}_n + \sum_{k=1}^{h} \hat{\mathbf{P}}^{h-k} \circ \mathbf{R}_{n+k}, \tag{5.4}$$

where $\hat{\mathbf{P}}$ is obtained from above estimation procedure. In the classical time series model, one would minimise $\mathrm{MSE}(h) = \mathbb{E}[(\hat{\mathbf{X}}_{n+h} - \mathbf{X}_{n+h})^2 | \mathbf{X}_n]$ to obtain the optimal linear predictor such that $\hat{\mathbf{X}}_{n+h} = \mathbb{E}[\mathbf{X}_{n+h} | \mathbf{X}_n]$. However, this would inevitably introduce real value for $\hat{\mathbf{X}}_{n+h}$, which is not coherent to the integer-valued nature of MMPGIG-INAR(1) model. To solve this, one can instead use the median $\tilde{\mathbf{X}}_{n+h}$ of $\mathbf{X}_{n+h}$, the 50% quantile, as prediction value for the model, which is also discussed by Pavlopoulos and Karlis (2008) and Homburg et al. (2019) . In the univariate case, the median is obtained by minimising the mean absolute error $\mathrm{MAE}(h) = \mathbb{E}[|\tilde{X}_{n+h} - X_{n+h}||X_n]$. The idea here can be extended to the multivariate case so that the median $\tilde{\mathbf{X}}_{n+h}$ is called geometric median, which is calculated by minimising the expected Euclidean distance

$$MAE(h) = \mathbb{E}[||\tilde{\mathbf{X}}_{n+h} - \mathbf{X}_{n+h}||_2 | X_n] \tag{5.5}$$

On the other hand, the expectation can be evaluated numerically by simulating the random samples of $\mathbf{X}_{n+h}$.

6 Empirical analysis

The data used in this section come from the Local Government Property Insurance Fund (LGPIF) from the state of Wisconsin. This fund provides property insurance to different types of government units, which includes villages, cities, counties, towns

Table 3 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 $X_{1,t}$	85.75009	86.64794	41.84915	43.01832	126.68793
Kurtosis of $X_{2,t}$	53.91835	61.77408	111.28108	184.13950	133.92283
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 years
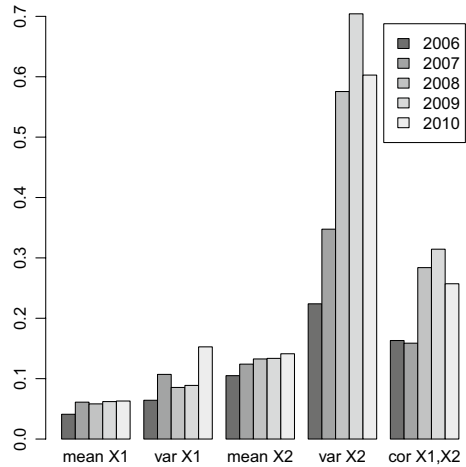
and schools. The LGPIF contains three major groups of property insurance coverage, namely building and contents (BC), inland marine (IM) and motor vehicles (PN, PO, CN, CO). For exploratory purposes, we focus on modelling jointly the claim frequency of IM, denoted as $X_{1,t}$, and comprehensive new vehicles collision (CN), denoted as $X_{2,t}$. The insurance data cover the period over 2006–2011 with 1234 policyholder records in total. Only $n_1 = 1048$ of them have complete data over the period 2006–2010 which will be used as the training data set. The last year 2011 with $n_2 = 1025$ policyholders out of 1048 in the data set will be the test data set. Denote the IM type and CN type claim frequency for a particular policyholder as $X_{1,t}^{(j)}, X_{2,t}^{(j)}$ respectively, where j is the identifier for each policyholder. Then the relationship between $X_{i,t}$ and $X_{i,t}^{(j)}$ is simply $X_{i,t} = \sum_{j=1}^{n_1} X_{i,t}^{(j)}$ with $i = 1, 2$ while t would take the values from 1 to 5 corresponding to the year 2006 to 2010.

In what follows, basic statistical analysis is shown in Table 3 and Figs. 1 and 2. The proportion of zeros for the two types of claims is higher than 90% during the period 2006–2010. Also, both types of claims exhibit overdispersion, since their variances exceeds their means during this period. Furthermore, the overdispersion for $X_{2,t}$ is even stronger than that of $X_{1,t}$, which indicates the need to employ an overdispersed distribution for this data. Additionally, the correlation tests for $X_{1,t}$ and $X_{2,t}$ show a positive correlation between the two claim types. At this point it is worth noting that modelling positively correlated claims has been explored by many articles. See for example, Bermúdez and Karlis (2011), Bermúdez and Karlis (2012), Shi and Valdez (2014a, b), Abdallah et al. (2016), Bermúdez and Karlis (2017), Bermúdez et al. (2018), Bermúdez et al. (2018), Pechon et al. (2018), Pechon et al. (2019), Bolancé and Vernic (2019), Denuit et al. (2019), Fung et al. (2019), Bolancé et al. (2020), Pechon et al. (2021), Jeong and Dey (2021), Gómez-Déniz and Calderín-Ojeda (2021), Tzougas and di Cerchiara (2021a, b) and Bermúdez and Karlis (2021). Finally, the proportion of zeros and kurtosis show that the marginal distributions of $X_{1,t}, X_{2,t}$ are positively skewed and exhibit a fat-tailed structure which

Table 4 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

indicates the appropriateness of adopting a positive skewed and fat-tailed distribution (GIG distribution).

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 4. 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 explained part of the heterogeneity between $X_{1,t}$ and $X_{2,t}$.

The MMPGIG-INAR(1) with $m = 2$, is applied to model the joint behaviour of $X_{1,t}^{(j)}, X_{2,t}^{(j)}$ across all the policyholders. Note that when Gamma mixing density is used in MPGIG INAR(1), the resulting model will be the "BINAR(1) Process with BVNB Innovations" in Pedeli and Karlis (2011), which we will used as comparison benchmark for other choices of mixing density. The the likelihood function would simply become

$$\ell(\Theta) = \sum_{j=1}^{n_1} \ell_j(\Theta) = \sum_{j=1}^{n_1} \sum_{t=1}^{4} \log \Pr(X_{1,t+1}^{(j)}, X_{2,t+1}^{(j)} | X_{1,t}^{(j)}, X_{2,t}^{(j)}), \qquad (6.1)$$

where $\ell_j(\Theta)$ is the likelihood function for policyholder j. Note that all the policyholders with the same type of claim $X_{i,\cdot}$ will share the same set of parameters p_i, β_i

and ϕ will be same for both claim types. In addition, it is necessary to show the appropriateness of introducing correlation and time-series component (binomial thinning) in MPGIG INAR(1). Then we also fit the data to following models.

1. The joint distribution of $X_{1,t}^{(j)}$ and $X_{2,t}^{(j)}$ are assumed to be bivariate mixed Poisson distribution (BMP) with probability mass function $f_\phi(\mathbf{k}, t)$ which we already discussed in Sect. 4.
2. The joint distribution of $X_{1,t}^{(j)}$ and $X_{2,t}^{(j)}$ are characterized by two independent INAR(1) models (TINAR)

$$X_{1,t}^{(j)} = p_1 \circ X_{1,t-1}^{(j)} + R_{1,t}$$
$$X_{2,t}^{(j)} = p_2 \circ X_{2,t-1}^{(j)} + R_{2,t},$$

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. Similarly, the likelihood functions for these models will have the same form as Eq. (6.1) but different joint distribution $\Pr(X_{1,t+1}^{(j)}, X_{2,t+1}^{(j)} | X_{1,t}^{(j)}, X_{2,t}^{(j)})$. For comparison purposes, we fit the bivariate Poisson mixture regression model with the training data starting from 2007 because BMP model does not need to consider lag responses.

All the estimations is implemented in R software by the 'optim' function with method 'BFGS' (quasi-Newton method). The gradient functions with respect to all the parameters are derived in Sects. 4 and 5 and they can be input as gradient argument in 'optim' function, which will significantly decrease the amount of computational time compared to numerical gradient function in default setting.

Model fitting results are shown in Tables 5 and 6. All the results show a great improvement by adopting a time series model compared to BMP results in Table 6. When focusing on the results of BINAR in Table 6, except the case where the mixing density is GIG $v = -\frac{3}{2}$, there is an significant improvement by introducing the fat-tailed distribution as mixing density in $\mathbf{R}_t$ compared to Gamma case. On the other hand, the improvement from the optimal TINAR to the optimal BINAR (cells are in bold face) is obvious, which is indicated by lower AIC and BIC of BINAR with GIG $v = -\frac{3}{4}$ compared to TINAR with GIG $v = -\frac{3}{4}$ and Inverse Gaussian. It implies that there is significant correlation between two claim sequences. Maximum likelihood estimates for three cases are given in Table 7 as well as their standard deviations. The standard derivations are estimated by inverting the numerical Hessian matrix. From Table 7 we see that the estimates for p_i, β_i are very close to each other while the estimated ϕ is significantly different among three mixing densities, which is expected because ϕ influences the tail and correlation structure of the bivariate sequence $X_{1,t}, X_{2,t}$. Furthermore, we see that the explanatory variables have a similar effect (positive and/or negative) and are almost identical for both response variables in the case of of all three models. Finally, the variables which are statistically significant at a 5 % threshold for $X_{1,t}$ are TypeCounty, TypeMisc, TypeVillage, NoClaimCreditIM, and those which are statistically significant at a 5 % threshold for $X_{2,t}$ are TypeCity, TypeCounty, TypeVillage, CoverageIM, and CoverageCN.

The Fig. 2 below presents prediction for both types of claims at $t = 2011$ with $n_2 = 1025$ policyholders based on geometric median Eq. (5.5). It seems that the

Table 5 The AIC and BIC when fitting as two independent INAR model with different combination

Mixing density	Gamma		Inverse Gaussian		GIG $\nu = -\frac{3}{4}$		GIG $\nu = -\frac{3}{2}$		Inverse Gamma	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Gamma	2999.957	3133.117	2999.433	3132.592	2999.590	3132.749	3000.48	3133.64	3002.352	3135.512
Inverse Gaussian	2998.876	3132.036	2998.351	3131.511	2998.508	3131.668	2999.399	3132.558	3001.271	3134.430
GIG $\nu = -\frac{3}{4}$	2998.661	3131.820	**2998.136**	**3131.296**	2998.293	3131.453	2999.184	3132.343	3001.056	3134.215
GIG $\nu = -\frac{3}{2}$	3004.415	3137.574	3003.890	3137.050	3004.047	3137.207	3004.938	3138.097	3006.810	3139.969
Inverse Gamma	2998.778	3131.938	2998.254	3131.413	2998.410	3131.570	2999.301	3132.461	3001.173	3134.333

row = Gamma, column = Inverse Gaussian means that gamma mixing density for $X_{1,t}$ and Inverse Gaussian mixing density for $X_{2,t}$. For each entry, the number on the left is AIC and the other one is BIC

Table 6 The AIC and BIC when fitting bivariate sequence as bivariate mixed Poisson regression model and BINAR model

Mixing density	Gamma		Inverse Gaussian		GIG $v = -\frac{3}{4}$		GIG $v = -\frac{3}{2}$		Inverse Gamma	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
BMP	3073.149	3187.285	3061.892	3176.028	3060.961	3175.098	3079.179	3193.315	3059.374	3173.510
BINAR(1)	2996.291	3123.109	2992.953	3119.771	**2992.854**	**3119.672**	3008.458	3135.277	2995.348	3122.167

For each entry, the number on the left is AIC and the other one is BIC

Table 7 Maximum likelihood estimation for MMPGIG-INAR(1) of insurance's claim frequency data when $m = 2$

Estimate	Mixing density					
	Gamma		Inverse Gaussian		GIG $v = -\frac{3}{4}$	
	$X_{1,t}$	$X_{2,t}$	$X_{1,t}$	$X_{2,t}$	$X_{1,t}$	$X_{2,t}$
p_i	0.1238	0.2904	0.1200	0.2768	0.1194	0.2750
	(0.0373)	(0.0378)	(0.0376)	(0.0388)	(0.0376)	(0.0388)
ϕ	0.9495		0.8885		0.7944	
	(0.1662)		(0.0931)		(0.0136)	
Intercept	− 3.7980	− 5.9744	− 3.8228	− 5.9967	− 3.8287	− 6.0029
	(0.5158)	(0.4428)	(0.5206)	(0.4503)	(0.5213)	(0.4489)
TypeCity	− 0.2242	0.6673	− 0.2316	0.6510	− 0.2337	0.6480
	(0.2555)	(0.2823)	(0.2577)	(0..2861)	(0.2586)	(0.2866)
TypeCounty	0.5682	1.3290	0.5784	1.3019	0.5814	1.2953
	(0.2811)	(0.2643)	(0.2836)	(0.2674)	(0.2794)	(0.2673)
TypeMisc	− 2.0110	− 0.1141	− 2.0213	− 0.1313	− 2.0226	− 0.1342
	(1.0210)	(0.6567)	(1.0223)	(0.6592)	(1.0244)	(0.6577)
TypeSchool	− 0.0387	0.1559	− 0.0638	0.1323	− 0.0692	0.1279
	(0.3587)	(0.2811)	(0.3570)	(0.2837)	(0.3534)	(0.2841)
TypeTown	− 0.3565	− 0.8941	− 0.3661	− 0.9155	− 0.3680	− 0.9195
	(0.3037)	(0.4794)	(0.3048)	(0.4816)	(0.3054)	(0.4820)
CoverageIM	1.4543		1.4309		1.4259	
	(0.2126)		(0.2115)		(0.2080)	
InDeductIM	0.0170		0.0243		0.0259	
	(0.0788)		(0.0795)		(0.0792)	
NoClaimCreaditIM	− 0.4569		− 0.4501		− 0.4482	
	(0.1570)		(0.1579)		(0.1579)	
CoverageCN		2.4227		2.4596		2.4675
		(0.2210)		(0.2260)		(0.2249)
NoClaimCreaditCN		− 0.3047		− 0.3231		− 0.3261
		(0.1811)		(0.1814)		(0.1801)

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

prediction for number of policyholders who make no claims are reasonably good while the prediction for $X_{1,t}$ are generally underestimated at tail and the prediction for $X_{2,t}$ are overestimated at the tail. On the other hand, Table 8 shows the prediction sum of squared error (PSSE) and frequency of some basic combination of observations, namely (0, 0), (1, 0), (0, 1), (1, 1) for the best fitted models within three classes, bivariate mixed Poisson regression, Two independent INAR(1) and bivariate INAR(1). It is again clear that the introduction of autoregressive part makes sense as it greatly reduce the prediction error. Although the best TINAR model has the closet frequency of (0, 0), the best BINAR model has the lowest overall prediction error.
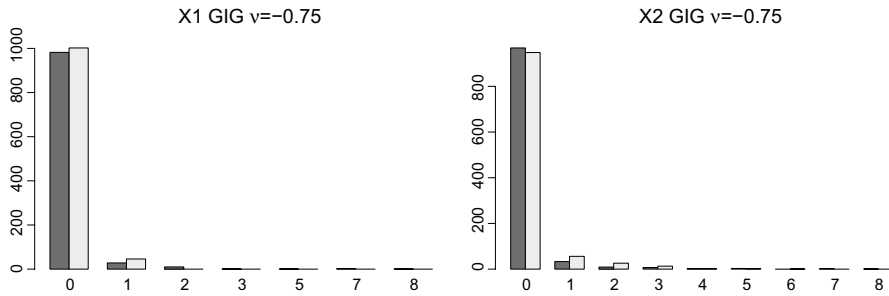
Fig. 2 Observed (dark) and Predicted (grey) frequency of the test data set based on estimated BINAR with GIG $v = -\frac{3}{4}$ as mixing density

Table 8 Summary of prediction on test data

Model	PSSE			Frequency			
	X_1	X_2	$X_1 + X_2$	(0,0)	(1,0)	(0,1)	(1,1)
BMP Inverse gamma	264	221	485	970	0	24	0
TINAR GIG $v = -0.75$ & Inverse gamma	212	164	376	964	4	33	0
BINAR GIG $v = -0.75$	212	156	368	966	3	32	0
Observed				940	20	26	6

7 Concluding remarks

In this paper we proposed the MMPGIG-INAR(1) regression model for modelling multiple time series of different types of count response variables. The proposed model, which is an extension of BINAR(1) regression model that was introduced by Pedeli and Karlis (2011), can accommodate positive correlation and multivariate overdispersion in a flexible manner. In particular, the Generalized Inverse Gaussian class includes many distributions as its special and limiting cases that can be used for modelling the innovations $\mathbf{R}_t$. Thus, the proposed modelling framework can efficiently capture the stylized characteristics of alternative complex data sets. Furthermore, due to the simple form of its density function, statistical inference for the MMPGIG-INAR(1) model is straightforward via the ML method, whereas other models that have been proposed in the literature, such as copula-based models, may result in numerical instability during the ML estimation procedure. For demonstration purposes different members of the proposed family of models were fitted to LGPIF data from the state of Wisconsin. Finally, it is worth mentioning that a possible line of further research could be to also consider cross correlation, meaning that the non-diagonal elements of $\mathbf{P}$ can take positive values.

Declarations

References

Abdallah A, Boucher J-P, Cossette H (2016) Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. Insurance 68:120–133

Al-Osh M, Alzaid AA (1987) First-order integer-valued autoregressive (INAR (1)) process. J Time Ser Anal 8(3):261–275

Amalia J, Purhadi, Otok BW (2017) Parameter estimation and statistical test of geographically weighted bivariate Poisson Inverse Gaussian regression models. In: AIP Conference Proceedings, vol 1905. AIP Publishing LLC, p 050005

Atkinson A, Yeh L (1982) Inference for Sichel's compound Poisson distribution. J Am Stat Assoc 77(377):153–158

Barndorff-Nielsen O, Blaesild P, Seshadri V (1992) Multivariate distributions with generalized inverse Gaussian marginals, and associated Poisson mixtures. Can J Stat 109–120

Bermúdez L, Karlis D (2011) Bayesian multivariate Poisson models for insurance ratemaking. Insurance 48(2):226–236

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

Bermúdez L, Karlis D (2017) A posteriori ratemaking using bivariate Poisson models. Scand Actuar J 2017(2):148–158

Bermúdez L, Karlis D (2021) Multivariate Inar (1) regression models based on the Sarmanov distribution. Mathematics 9(5):505

Bermúdez L, Guillén M, Karlis D (2018) Allowing for time and cross dependence assumptions between claim counts in ratemaking models. Insurance 83:161–169

Bolancé C, Vernic R (2019) Multivariate count data generalized linear models: three approaches based on the Sarmanov distribution. Insurance 85:89–103

Bolancé C, Guillen M, Pitarque A (2020) A Sarmanov distribution with beta marginals: an application to motor insurance pricing. Mathematics 8(11):2020

Boucher J-P, Denuit M, Guillen M (2008) Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions. Variance 2(1):135–162

Cheon S, Song SH, Jung BC (2009) Tests for independence in a bivariate negative binomial model. J Korean Stat Soc 38(2):185–190

Davis RA, Holan SH, Lund R, Ravishanker N (2016) Handbook of discrete-valued time series. CRC Press, Boca Raton

Denuit M, Guillen M, Trufin J et al (2019) Multivariate credibility modelling for usage-based motor insurance pricing with behavioural data. Ann Actuar Sci 13(2):378–399

Freeland RK, McCabe BP (2004) Analysis of low count time series data by Poisson autoregression. J Time Ser Anal 25(5):701–722

Fung TC, Badescu AL, Lin XS (2019) A class of mixture of experts models for general insurance: application to correlated claim frequencies. ASTIN Bull 49(3):647–688

Ghitany M, Karlis D, Al-Mutairi D, Al-Awadhi F (2012) An EM algorithm for multivariate mixed Poisson regression models and its application. Appl Math Sci 6(137):6843–6856

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

Homburg A, Weiß CH, Alwan LC, Frahm G, Göb R (2019) Evaluating approximate point forecasting of count processes. Econometrics 7(3):30

Jeong H, Dey DK (2021) Multi-peril frequency credibility premium via shared random effects. Available at SSRN 3825435

Johnson NL, Kotz S, Balakrishnan N (1995) Continuous univariate distributions, volume 1, vol 289. Wiley, New York

Karlis D, Pedeli X (2013) Flexible bivariate INAR (1) processes using Copulas. Commun Stat-Theory Methods 42(4):723–740

Latour A (1997) The multivariate GINAR (p) process. Adv Appl Probab 29(1):228–248

Mardalena S, Purhadi P, Purnomo JDT, Prastyo DD (2020) Parameter estimation and hypothesis testing of multivariate Poisson Inverse Gaussian regression. Symmetry 12(10):1738

Mardalena S, Purnomo J, Prastyo D, et al (2021) Bivariate poisson inverse gaussian regression model with exposure variable: infant and maternal death case study. In: Journal of Physics: Conference Series, vol 1752. IOP Publishing, p 012016

Marshall AW, Olkin I (1990) Multivariate distributions generated from mixtures of convolution and product families. Lecture Notes-Monograph Series, pp 371–393

McKenzie E (1985) Some simple models for discrete variate time series. J Am Water Resourc Assoc 21(4):645–650

Nastić AS, Ristić MM, Popović PM (2016) Estimation in a bivariate integer-valued autoregressive process. Commun Stat-Theory Methods 45(19):5660–5678

Pavlopoulos H, Karlis D (2008) INAR(1) modeling of overdispersed count series with an environmental application. Environmetrics 19(4):369–393

Pechon F, Trufin J, Denuit M et al (2018) Multivariate modelling of household claim frequencies in motor third-party liability insurance. Astin Bull 48(3):969–993

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

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

Pedeli X, Karlis D (2011) A bivariate INAR (1) process with application. Stat Model 11(4):325–349

Pedeli X, Karlis D (2013a) On composite likelihood estimation of a multivariate INAR (1) model. J Time Ser Anal 34(2):206–220

Pedeli X, Karlis D (2013b) On estimation of the bivariate Poisson INAR process. Commun Stat-Simul Comput 42(3):514–533

Popović PM (2016) A bivariate INAR (1) model with different thinning parameters. Stat Pap 57(2):517–538

Popović PM, Ristić MM, Nastić AS (2016) A geometric bivariate time series with different marginal parameters. Stat Pap 57(3):731–753

Ristić MM, Nastić AS, Jayakumar K, Bakouch HS (2012) A bivariate INAR (1) time series model with geometric marginals. Appl Math Lett 25(3):481–485

Scotto MG, Weiß CH, Gouveia S (2015) Thinning-based models in the analysis of integer-valued time series: a review. Stat Model 15(6):590–618

Shi P, Valdez EA (2014a) Longitudinal modeling of insurance claim counts using jitters. Scand Actuar J 2014(2):159–179

Shi P, Valdez EA (2014b) Multivariate negative binomial models for insurance claim counts. Insurance 55:18–29

Sichel H (1974) On a distribution representing sentence-length in written prose. J R Stat Soc 137(1):25–34

Sichel H (1982) Asymptotic efficiencies of three methods of estimation for the inverse Gaussian-Poisson distribution. Biometrika 69(2):467–472

Stein GZ, Juritz JM (1988) Linear models with an Inverse Gaussian Poisson error distribution. Commun Stat-Theory Methods 17(2):557–571

Tzougas G, di Cerchiara AP (2021a) The multivariate mixed negative binomial regression model with an application to insurance a posteriori ratemaking. Insurance 101:602–625

Tzougas G, di Cerchiara AP (2021b) Bivariate mixed Poisson regression models with varying dispersion. N Am Actuar J 1–31

Weiß CH (2008) Thinning operations for modeling time series of counts—a survey. Adv Stat Anal 92(3):319–341

Weiß CH (2018) An introduction to discrete-valued time series. Wiley, New York

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