Annals of the Institute of Statistical Mathematics (2020) 72:1449–1477
https://doi.org/10.1007/s10463-019-00732-4

Flexible bivariate Poisson integer-valued GARCH model

Yan Cui[1] · Qi Li[2] · Fukang Zhu[1]

Received: 17 April 2019 / Revised: 3 August 2019 / Published online: 16 September 2019

Abstract
Integer-valued time series models have been widely used, especially integer-valued autoregressive models and integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) models. Recently, there has been a growing interest in multivariate count time series. However, existing models restrict the dependence structures imposed by the way they constructed. In this paper, we consider a class of flexible bivariate Poisson INGARCH(1,1) model whose dependence is established by a special multiplicative factor. Stationarity and ergodicity of the process are discussed. The maximization by parts algorithm and its modified version together with the alternative method by using R package Template Model Builder are employed to estimate the parameters of interest. The consistency and asymptotic normality for estimates are obtained, and the finite sample performance of estimators is given via simulations. A real data example is also provided to illustrate the model.

Keywords Bivariate · INGARCH model · Multiplicative factor · Poisson distribution · Time series of counts

1 Introduction

Integer-valued time series are commonly encountered in many practical situations, such as the number of goods sold in a shopping mall, the monthly number of insurance claim, the daily number of transaction in stock market and so on. Recently, there have

Zhu's work is supported by National Natural Science Foundation of China (Nos. 11871027, 11731015), Science and Technology Developing Plan of Jilin Province (No. 20170101057JC), and Cultivation Plan for Excellent Young Scholar Candidates of Jilin University. Li's work is supported by Natural Science Foundation of Changchun Normal University (No. 2018-004).

✉ Fukang Zhu
zfk8010@163.com

[1] School of Mathematics, Jilin University, 2699 Qianjin Street, Changchun 130012, China

[2] College of Mathematics, Changchun Normal University, Changchun 130032, China

been plenty of attempts to deal with them, see Weiß (2008) and Scotto et al. (2015) for some excellent reviews on INAR models.

As an alternative, the INGARCH model proposed by Ferland et al. (2006) and Fokianos et al. (2009) is also very popular, which is defined as follows:

$$\begin{cases} X_t|\mathcal{F}_{t-1} \sim \mathbb{P}(\lambda_t), & \forall t \in \mathbb{Z}, \\ \lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \lambda_{t-i} + \sum_{j=1}^{q} \beta_j X_{t-j}, \end{cases} \tag{1}$$

where $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, i = 1, \ldots, p, j = 1, \ldots, q, p \geq 0, q \geq 1$, and $\mathcal{F}_{t-1}$ is the σ-field generated by $\{X_{t-1}, X_{t-2}, \ldots\}$. Zhu (2011, 2012a, b) and Davis and Liu (2016) generalized the Poisson assumption to negative binomial, generalized Poisson, zero-inflated Poisson, negative binomial and exponential family distributions, respectively. Inferential aspects of model (1) and its generalized forms have been well established, including ergodicity, estimating methods and goodness-of-fit tests, see Neumann (2011), Doukhan et al. (2012), Fokianos and Neumann (2013), Ahmad and Francq (2016) and Douc et al. (2017), among others. Fokianos (2012, 2016) and Tjøstheim (2012, 2016) reviewed some recent progress in this field.

In many cases, the data are observed across time leading to multivariate time series data. Research on univariate time series of counts is quite mature; however, the literature on multivariate time series models for count data is more limited, especially for bivariate versions. Franke and Rao (1995) proposed a multivariate INAR(1) model and it was generalized to the p-order case by Latour (1997). Pedeli and Karlis (2011) introduced bivariate INAR(1) model with Poisson and negative binomial innovations, Popović (2016) proposed a bivariate INAR(1) model with random coefficients based on different binomial thinning operators, and Liu (2012) conducted a bivariate Poisson INGARCH model. But the above models only allow for positive correlations between the two series.

This leads researchers to select feasible distributions that can deal with negative correlation. Aitchison and Ho (1989) proposed the bivariate Poisson-lognormal model to produce negative correlation. Karlis and Meligkotsidou (2007) developed the finite mixtures of multivariate Poisson distributions allowing for both overdispersion in the marginal distributions and negative correlation. Later, Cui and Zhu (2018) introduced a new bivariate INGARCH model which can capture both positive and negative correlations. However, the dependence structure is limited and do not offer enough flexibility for dependence modelling. So we consider some other alternative forms to construct bivariate distribution with more complex dependence structures.

As we all know, copulas are very popular in statistical applications since they provide a general way of introducing dependence among several series with given continuous distributions. Because of its flexibility, copula function has also become a widely accepted tool in the integer-valued (discrete) field during recent years. Heinen and Rengifo (2007) used the continued extension argument to apply copulas with discrete marginals and introduced contemporaneous correlation of multivariate counts. Karlis and Pedeli (2013) considered a bivariate INAR(1) process by specifying the innovation distribution via finite differences of the copula cumulative distribution function. Fokianos et al. (2019) proposed a joint distribution of counts with copula structure on the waiting time of the Poisson process instead of imposing it on discrete random variables. But in this paper, we aim to deal with the discrete case from a different prospective. Motivated by the idea of Lakshminarayana et al. (1999), we investigate a new class of bivariate Poisson distribution whose probability mass function is a product of Poisson marginals with multiplicative factors, and then introduce the flexible bivariate Poisson INGARCH(1,1) model that accommodates both positive and negative correlations.

Fokianos et al. (2019) provided a plausible approach to construct bivariate Poisson distribution, whose aim is to introduce a copula form via a vector of continuous random variables. However, their model can only consider the quasi-likelihood inference due to the implicit copula structure. Therefore, they need to determine the copula structure and estimate the corresponding copula parameter by the bootstrap procedure based on a parametric copula form. On the one hand, this method is time-consuming on estimation of copula parameter together with its standard error. On the other hand, picking out the optimal copula structure is challenging for real data analysis, as Fokianos et al. (2019) mentioned. In contrast, our model has several merits as follows. First, the positive or negative correlation between two random variates directly depends on the dependency parameter in the expression of the multiplicative factor. Second, since the log-likelihood function of our proposed model can be viewed as an explicit additive decomposition, we can estimate the total parameters of interest at the same time using the maximization by parts algorithm and its modified version. Finally, due to the fact that above two methods might be a little bit time-consuming, we also consider another approach using an R package Template Model Builder to obtain the accurate estimates with faster computational speed, even when the dependency between two random variables is quite high.

The rest of this paper is organized as follows. In Sect. 2, we first briefly review the bivariate Poisson distribution, then present the new models and discuss some important properties. The estimation procedure for the parameters of interest is given in Sect. 3. In Sect. 4, we give some simulation results. Section 5 provides a real data example and we conclude in Sect. 6.

Throughout this paper, we use the following notations. The term $|\cdot|$ denotes the absolute value of a univariate variable; $\|\boldsymbol{x}\|$ and $\|\boldsymbol{x}\|_p$ are the Euclidean norm and p-norm of vector $\boldsymbol{x}$. $\|\boldsymbol{J}\|_p = \max_{x\neq 0}\{\|\boldsymbol{J}\boldsymbol{x}\|_p / \|\boldsymbol{x}\|_p\}$ stands for the p-induced norm of matrix $\boldsymbol{J}$ for $1 \leq p \leq \infty$. $\rho(\boldsymbol{A})$ means the spectral radius, i.e., the largest absolute eigenvalue of $\boldsymbol{A}$. The terms $\dot{l}(\boldsymbol{\theta})$ and $\ddot{l}(\boldsymbol{\theta})$ are the first- and second-order derivatives concerning the parameter $\boldsymbol{\theta}$, respectively.

2 Flexible bivariate Poisson INGARCH model

2.1 The bivariate Poisson distribution

First, we recall the bivariate Poisson distribution $\mathrm{BP}(\lambda_1, \lambda_2, \delta)$ defined by Lakshminarayana et al. (1999). It is a product of Poisson marginals with a multiplicative factor, whose probability mass function (pmf) is given by

$$\mathbb{P}(Y_1 = y_1, Y_2 = y_2) = \frac{\lambda_1^{y_1}\lambda_2^{y_2}}{y_1!y_2!}\exp\{-(\lambda_1+\lambda_2)\}[1+\delta(\mathrm{e}^{-y_1}-\mathrm{e}^{-c\lambda_1})(\mathrm{e}^{-y_2}-\mathrm{e}^{-c\lambda_2})], \tag{2}$$

where $c = 1-\mathrm{e}^{-1}$. One can see that the marginal pmf of Y_1 and Y_2 are Poisson distribution with parameters λ_1 and λ_2, respectively. The mean vector of the above distribution is $(\lambda_1, \lambda_2)^\top$ and the correlation coefficient turns out to be $\rho = \delta c^2\sqrt{\lambda_1\lambda_2}\mathrm{e}^{-c(\lambda_1+\lambda_2)}$. Thus, the dependence between the variates can be positive, zero or negative depending on the value of the correlation parameter δ. It is important to mention that the function $1 + \delta(\mathrm{e}^{-y_1} - \mathrm{e}^{-c\lambda_1})(\mathrm{e}^{-y_2} - \mathrm{e}^{-c\lambda_2})$ in (2) in a sense is used to link the marginal distributions. However, the simplex structure is limited and cannot provide enough flexibility for dependence modelling.

In view of this construction for BP$(\lambda_1, \lambda_2, \delta)$ distribution, we introduce three types of bivariate Poisson (BP) distributions with different multiplicative factors, that is, BP with Gaussian factor (BPG), BP with Frank factor (BPF) and BP with Farlie–Gumbel–Morgenstern factor (BPFGM). These functions are given as follows:

$$c_\rho(u_1, u_2) = \frac{1}{\sqrt{1-\rho^2}}\exp\left(-\frac{\rho^2(q_1^2+q_2^2)-2\rho q_1 q_2}{2(1-\rho^2)}\right), \tag{3}$$

$$c_\gamma(u_1, u_2) = \frac{-\gamma(\mathrm{e}^{-\gamma}-1)\mathrm{e}^{-(u_1+u_2)\gamma}}{[(\mathrm{e}^{-\gamma}-1)+(\mathrm{e}^{-u_1\gamma}-1)(\mathrm{e}^{-u_2\gamma}-1)]^2}, \tag{4}$$

$$c_\sigma(u_1, u_2) = 1+\sigma(1-2u_1)(1-2u_2), \tag{5}$$

where $q_i = \Phi^{-1}(u_i)$, Φ^{-1} is the inverse of the standard univariate normal distribution and $u_i \in [0, 1], i = 1, 2$. $\rho \in (-1, 1)$, $\gamma \in (-\infty, +\infty)\backslash\{0\}$, $\sigma \in [-1, 1]$ are regarded as dependency parameters for BP distributions.

Consider random variables Y_1 and Y_2, which follow independent Poisson distributions marginally with parameters λ_1 and λ_2, respectively. We define the joint pmf of newly proposed BP distribution given by the product of marginal densities and multiplicative factors as

$$\mathbb{P}(Y_1 = y_1, Y_2 = y_2) = \frac{1}{Z(\lambda_1, \lambda_2, \theta)}\frac{\lambda_1^{y_1}\lambda_2^{y_2}}{y_1!y_2!}\exp\{-(\lambda_1+\lambda_2)\}c(F_1(y_1), F_2(y_2)), \tag{6}$$

where $c(F_1(y_1), F_2(y_2))$ stands for anyone of the above multiplicative factors (see (3)–(5)), $F_i(y_i)$ is the Poisson marginal cumulative distribution function (cdf) of Y_i, $i = 1, 2$ and $Z(\lambda_1, \lambda_2, \theta) := \sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty}\frac{\lambda_1^{y_1}\lambda_2^{y_2}}{y_1!y_2!}\exp\{-(\lambda_1+\lambda_2)\}c(F_1(y_1), F_2(y_2))$ with θ being the dependency parameter. Note that the functions (3) and (4) are both bounded away from 0, and (5) is nonnegative on $[0, 1]^2$. With the above definition, (6) satisfies the conditions for a probability of bivariate Poisson distribution. Similar to the discussion of Appendix B in Shmueli et al. (2005), we know that the series Z converges and both Z and Z^{-1} are bounded. As a result, we can treat the factor Z^{-1} as a constant weight when conducting estimation, see more details in Appendix A.

Since the weight Z^{-1} does not affect the dependence between the variates, we can interpret (6) as a bivariate Poisson distribution constructed by a scaled copula function with $F_1(y_1)$ and $F_2(y_2)$ as marginals. In particular, $\mathbb{E}(Y_1) = \lambda_1$, $\mathbb{E}(Y_2) = \lambda_2$, and the dependency parameters ρ, γ and σ can capture both positive and negative correlations between the variates according to the values chosen in their ranges.

2.2 Model formulation and stability theory

Denote $\boldsymbol{Y}_t = (Y_{t,1}, Y_{t,2})^\top$ as the bivariate observations at time t, that is, $\{Y_{t,1}, t \geq 1\}$ and $\{Y_{t,2}, t \geq 1\}$ are two time series under consideration. As a generalization of traditional BP model in Liu (2012), Cui and Zhu (2018) defined an INGARCH model of order (1,1) based on BP$(\lambda_1, \lambda_2, \delta)$ as follows:

$$\boldsymbol{Y}_t|\mathcal{F}_{t-1} \sim \text{BP}(\lambda_{t,1}, \lambda_{t,2}, \delta), \ \boldsymbol{\lambda}_t = (\lambda_{t,1}, \lambda_{t,2})^\top = \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1} + \boldsymbol{B}\boldsymbol{Y}_{t-1},$$

where $\mathcal{F}_t = \sigma\{\boldsymbol{\lambda}_1, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t\}$, $\boldsymbol{\omega} = (\omega_1, \omega_2)^\top \in \mathbb{R}_+^2$, $\boldsymbol{A}, \boldsymbol{B}$ are both 2×2 matrices with nonnegative entries and the correlation $\text{Corr}(Y_{t,1}, Y_{t,2}|\mathcal{F}_{t-1}) = \delta c^2\sqrt{\lambda_{t,1}\lambda_{t,2}}\text{e}^{-c(\lambda_{t,1}+\lambda_{t,2})}$ for capturing dependence between $Y_{t,1}$ and $Y_{t,2}$. Similar to this, here we propose a new class of BP-INGARCH(1,1) model with flexible multiplicative factor as follows:

$$\boldsymbol{Y}_t|\mathcal{F}_{t-1} \sim \mathcal{BP}(\boldsymbol{\lambda}_t), \ \boldsymbol{\lambda}_t = (\lambda_{t,1}, \lambda_{t,2})^\top = \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1} + \boldsymbol{B}\boldsymbol{Y}_{t-1}, \tag{7}$$

where $\mathcal{BP}$ stands for one of three BP distributions defined in Sect. 2.1, denoted as BPG$(\lambda_{t,1}, \lambda_{t,2}, \rho)$, BPF$(\lambda_{t,1}, \lambda_{t,2}, \gamma)$ and BPFGM$(\lambda_{t,1}, \lambda_{t,2}, \sigma)$. It is easy to see that the conditional mean and variance of $\boldsymbol{Y}_t$ are both equal to $\boldsymbol{\lambda}_t$. Furthermore,

$$\begin{aligned}\text{Var}(\boldsymbol{Y}_t) &= \mathbb{E}(\text{Var}(\boldsymbol{Y}_t|\mathcal{F}_{t-1})) + \text{Var}(\mathbb{E}(\boldsymbol{Y}_t|\mathcal{F}_{t-1})) \\ &= \mathbb{E}(\boldsymbol{\lambda}_t) + \text{Var}(\boldsymbol{\lambda}_t) > \mathbb{E}(\boldsymbol{\lambda}_t) = \mathbb{E}(\boldsymbol{Y}_t),\end{aligned}$$

which indicates the model (7) can account for overdispersion. The study focuses on the bivariate Markov chain $\{\boldsymbol{\lambda}_t, t \geq 1\}$. Note that by iteration, for any $s \geq 1$, we have

$$\boldsymbol{\lambda}_t = (\boldsymbol{I} + \boldsymbol{A} + \cdots + \boldsymbol{A}^{s-1})\boldsymbol{\omega} + \boldsymbol{A}^s\boldsymbol{\lambda}_{t-s} + \sum_{k=0}^{s-1}\boldsymbol{A}^k\boldsymbol{B}\boldsymbol{Y}_{t-k-1}, \tag{8}$$

where $\boldsymbol{I}$ is the identity matrix. Now further assume that $\rho(\boldsymbol{A}) < 1$ for some $p \in [1, \infty]$, then we have

$$\boldsymbol{\lambda}_t = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{\omega} + \sum_{k=0}^{\infty}\boldsymbol{A}^k\boldsymbol{B}\boldsymbol{Y}_{t-k-1}. \tag{9}$$

Hence, under the condition $\rho(\boldsymbol{A}) < 1$, (9) implies $\boldsymbol{\lambda}_t \geq (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{\omega}$ for all t. In addition, $\{\boldsymbol{\lambda}_t, t \geq 1\}$ can be represented as an iterated random function following the notation used by Wu and Shao (2004). To facilitate the investigation, the random function $f_{\boldsymbol{d}}(\boldsymbol{\lambda})$ according to the pmf (6) is defined as

$$f_{\boldsymbol{d}}(\boldsymbol{\lambda}) = \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{\lambda} + \boldsymbol{B}F_{\boldsymbol{\lambda}}^{-1}(\boldsymbol{d}),$$

where $\boldsymbol{d} = (d_1, d_2)^\top \in [0, 1]^2$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$, $F_{\boldsymbol{\lambda}}^{-1}(\boldsymbol{d}) = (F_{\lambda_1}^{-1}(d_1), F_{\lambda_2}^{-1}(d_2))^\top \in \mathbb{N}_0^2$, and $F^{-1}(d) = \inf\{t \geq 0 : F(t) \geq d\}$. Thus, it can be seen that for all t, $\boldsymbol{\lambda}_t = f_{\boldsymbol{D}_t}(\boldsymbol{\lambda}_{t-1})$, where $\{\boldsymbol{D}_t, t \geq 1\}$ follows independent uniform distribution on $[0, 1]^2$.

Next, the stability properties of the proposed model are given in the following theorem.

Theorem 1 *Suppose $\{\boldsymbol{Y}_t, t \geq 1\}$ follow the model defined by (7), $\boldsymbol{\omega}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ have nonnegative entries.*

(a) *If $\rho(\boldsymbol{A} + \boldsymbol{B}) < 1$, then there exists at least one stationary distribution to $\{\boldsymbol{\lambda}_t\}$. Moreover, if $\|\boldsymbol{A}\|_p < 1$ for some $1 \leq p \leq \infty$, then the stationary distribution is unique.*

(b) *If $\|\boldsymbol{A}\|_p + 2^{(1-1/p)} \|\boldsymbol{B}\|_p < 1$ for some $1 \leq p \leq \infty$, then $\{\boldsymbol{\lambda}_t\}$ is a geometric moment contraction Markov chain with a unique stationary and ergodic distribution, denoted by π.*

Proofs for the above and next theorems are deferred to the Appendix B.

To make a further comparison among the existing bivariate process, we recall the bivariate Poisson INGARCH models defined by Liu (2012) and Fokianos et al. (2019). Liu (2012) defined an INGARCH(1,1) model based on the bivariate Poisson distribution $BP^*(\lambda_1, \lambda_2, \phi)$ with the best known method *Trivariate Reduction*. In the same setting of Sect. 2.2, Liu's model has the following form:

$$\boldsymbol{Y}_t|\mathcal{F}_{t-1} \sim \mathrm{BP}^*(\lambda_{t,1}, \lambda_{t,2}, \phi),\ \boldsymbol{\lambda}_t = (\lambda_{t,1}, \lambda_{t,2})^\top = \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1} + \boldsymbol{B}\boldsymbol{Y}_{t-1},$$

where the covariance $\mathrm{Cov}(Y_{t,1}, Y_{t,2}) = \phi \geq 0$. While Fokianos et al. (2019) let the count time series follow by Poisson distributions marginally and then impose a copula on the waiting times to accommodate dependence. Fokianos's linear model is defined as

$$Y_{t,i}|\mathcal{F}_{t-1} \text{ is marginally Poisson}(\lambda_{t,i}),\ i = 1, 2,$$
$$\boldsymbol{\lambda}_t = (\lambda_{t,1}, \lambda_{t,2})^\top = \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1} + \boldsymbol{B}\boldsymbol{Y}_{t-1},$$

the copula structure with copula parameter ψ is contained in the conditional innovation $\boldsymbol{Y}_t|\boldsymbol{\lambda}_t$.

3 Estimation

For ease of presentation, let $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$ be observations from model (7) with the assumption that $\boldsymbol{A}$ and $\boldsymbol{B}$ are nonnegative diagonal matrices. Extension to the case of nonnegative non-diagonal matrices for $\boldsymbol{A}$ and $\boldsymbol{B}$ is direct, but we omit it for simplicity. Then, write the true value of the parameter as $\boldsymbol{\theta}^0 = (\boldsymbol{\theta}_1^{0\top}, \theta_2^0)^\top$, the parameter vector turns to be $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \theta_2)^\top$, where $\boldsymbol{\theta}_1 = (\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2)^\top$, $\theta_2 = \rho,\ \gamma$ or σ depending on which multiplicative factor $c(\cdot, \cdot)$ to be used. As discussed in Appendix A, we can regard Z^{-1} as a constant weight, and then, the log-likelihood function is given by, up to a constant free of $\boldsymbol{\theta}$,

$$l(\boldsymbol{\theta}) = \sum_{t=2}^{n} l_t(\boldsymbol{\theta}) = \sum_{t=2}^{n} [Y_{t,1} \ln \lambda_{t,1}(\boldsymbol{\theta}_1) + Y_{t,2} \ln \lambda_{t,2}(\boldsymbol{\theta}_1) - \lambda_{t,1}(\boldsymbol{\theta}_1) - \lambda_{t,2}(\boldsymbol{\theta}_1)] + \sum_{t=2}^{n} [\ln c(F_1(Y_{t,1}; \lambda_{t,1}(\boldsymbol{\theta}_1)), F_2(Y_{t,2}; \lambda_{t,2}(\boldsymbol{\theta}_1)))]. \quad (10)$$

where $F_i(Y_{t,i}; \lambda_{t,i})$ is the marginal cdf with the parameter $\lambda_{t,i}$ of $Y_{t,i}$, $i = 1, 2$.

One can see that the objective function is very complicated and the second-order derivatives are hard to derive analytically. So we resort to the alternative approaches: the maximization by parts (MBP) algorithm proposed by Song et al. (2005) and the modified maximization by parts (MMBP) algorithm studied by Liu and Luger (2009). These two algorithms can yield closed-form expressions for the iterative estimates of the marginal parameters and make the calculation simpler. To the end, we also utilize a faster and more flexible method based on the R package Template Model Builder (TMB) to estimate the parameters.

3.1 MBP algorithm

Rewrite the log-likelihood function $l(\boldsymbol{\theta}) = l_m(\boldsymbol{\theta}_1) + l_c(\boldsymbol{\theta}_1, \theta_2) = \sum_{t=2}^{n} l_{m,t}(\boldsymbol{\theta}_1) + \sum_{t=2}^{n} l_{c,t}(\boldsymbol{\theta}_1, \theta_2)$, where the marginal density part $l_m(\boldsymbol{\theta}_1)$ is the first term of the last equation in (10) and the multiplicative factor part $l_c(\boldsymbol{\theta}_1, \theta_2)$ is the left term. The score equation of the full log-likelihood function in (10) is given by

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial l_m(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}} + \frac{\partial l_c(\boldsymbol{\theta}_1, \theta_2)}{\partial \boldsymbol{\theta}}.$$

Song et al. (2005) provided several examples where the likelihood can be decomposed in this way and the maximum likelihood estimator (MLE) is difficult to compute directly. The MBP algorithm proceeds as follows:

$$\text{Step 1 : Solve } \frac{\partial l_m(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} = \mathbf{0} \text{ for } \hat{\boldsymbol{\theta}}_{1,n}^{1} \text{ and } \frac{\partial l_c(\hat{\boldsymbol{\theta}}_{1,n}^{1}, \theta_2)}{\partial \theta_2} = 0 \text{ for } \hat{\theta}_{2,n}^{1}.$$

$$\text{Step } k \text{ : Solve } \frac{\partial l_m(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} = -\frac{\partial l_c(\hat{\boldsymbol{\theta}}_{1,n}^{k-1}, \hat{\theta}_{2,n}^{k-1})}{\partial \boldsymbol{\theta}_1} \text{ for } \hat{\boldsymbol{\theta}}_{1,n}^{k}$$

$$\text{and } \frac{\partial l_c(\hat{\boldsymbol{\theta}}_{1,n}^{k-1}, \theta_2)}{\partial \theta_2} = 0 \text{ for } \hat{\theta}_{2,n}^{k},\ k = 2, 3, \ldots.$$

As a promising alternative to direct maximization of the full likelihood function, this algorithm improves efficiency of the estimates through an iterative approach. Let $\hat{\boldsymbol{\theta}}_{n}^{k} = (\hat{\boldsymbol{\theta}}_{1,n}^{k\top}, \hat{\theta}_{2,n}^{k})^{\top}$, to establish asymptotic properties of the estimators $\hat{\boldsymbol{\theta}}_{n}^{k}$, we introduce a neighborhood of the true parameter $\boldsymbol{\theta}^0$, i.e., $\mathcal{U}_0 = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| < \xi\}$. Now we study the large sample properties in terms of the MBP algorithm. To formulate the results, first we need the following assumptions.

Assumption 1 $l(\boldsymbol{\theta}), l_m(\boldsymbol{\theta}_1)$ and $l_c(\boldsymbol{\theta}_1, \theta_2)$ are twice continuously differentiable for $\boldsymbol{\theta} \in \mathcal{U}_0$.

Assumption 2 (Information dominance) $\left\|\mathcal{I}_m^{-1}\mathcal{I}_c\right\| < 1$, where $\mathcal{I}_m = -n^{-1}\mathbb{E}\ddot{l}_m(\boldsymbol{\theta}^0)$ and $\mathcal{I}_c = -n^{-1}\mathbb{E}\ddot{l}_c(\boldsymbol{\theta}^0)$.

Then, the following theorems regarding the consistency and asymptotic normality of the estimators $\hat{\boldsymbol{\theta}}_n^k$ for both a fixed k and as $k \to \infty$ hold true.

Theorem 2 *If $\hat{\boldsymbol{\theta}}_n^1$ is consistent under Assumption 1, then $\hat{\boldsymbol{\theta}}_n^k$ is consistent for each $k = 2, 3, \ldots$.*

Suppose the solution $\hat{\boldsymbol{\theta}}_n^k$ satisfies $\hat{\boldsymbol{\theta}}_n^k = M_n(\hat{\boldsymbol{\theta}}_n^{k-1})$, where $M_n(\cdot)$ is an asymptotic contraction mapping, then according to Dominitz and Sherman (2005), there exists a fixed point such that $\hat{\boldsymbol{\theta}}_n = \lim_{k\to\infty} \hat{\boldsymbol{\theta}}_n^k$, which is always the MLE.

Remark 1 If $\hat{\boldsymbol{\theta}}_n^1$ is not a consistent estimator of $\boldsymbol{\theta}^0$, Dominitz and Sherman (2005) has verified that as long as $M_n(\cdot)$ is an asymptotic contraction mapping conditional on the data and the sample size n, the sequence $\{\hat{\boldsymbol{\theta}}_n^k\}_{k=1}^{\infty}$ will still converge to the MLE $\hat{\boldsymbol{\theta}}_n$.

For ease of readability, we introduce the following notations. Let

$$\dot{l}_{i(j)} = \dot{l}_{i(j)}(\boldsymbol{\theta}^0) = \partial l_i(\boldsymbol{\theta}^0)/\partial \boldsymbol{\theta}_j, \quad \ddot{l}_{i(jk)} = \ddot{l}_{i(jk)}(\boldsymbol{\theta}^0) = \partial^2 l_i(\boldsymbol{\theta}^0)/\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k^\top \tag{11}$$

and $\mathcal{I}_{i(jk)} = -n^{-1}\mathbb{E}\ddot{l}_{i(jk)}(\boldsymbol{\theta}^0)$, $i = m, c$ and $j, k = 1, 2$. Then at $\boldsymbol{\theta}^0$, we have the following results:

$$\begin{aligned}
\boldsymbol{D}_n &= \begin{pmatrix} -n^{-1}\ddot{l}_{m(11)} & 0 \\ 0 & -n^{-1}\ddot{l}_{c(22)} \end{pmatrix} = \begin{pmatrix} \mathcal{I}_{m(11)} & 0 \\ 0 & \mathcal{I}_{c(22)} \end{pmatrix} + o(1) \equiv \boldsymbol{D} + o(1), \\
\boldsymbol{T}_n &= \begin{pmatrix} n^{-1}\ddot{l}_{c(11)} & n^{-1}\ddot{l}_{c(12)} \\ n^{-1}\ddot{l}_{c(21)} & 0 \end{pmatrix} = -\begin{pmatrix} \mathcal{I}_{c(11)} & \mathcal{I}_{c(12)} \\ \mathcal{I}_{c(21)} & 0 \end{pmatrix} + o(1) \equiv \boldsymbol{T} + o(1), \\
\boldsymbol{V}_n &= \begin{pmatrix} 0 & 0 \\ n[\ddot{l}_{c(22)}]^{-1}[\ddot{l}_{c(21)}][\ddot{l}_{m(11)}]^{-1} & 0 \end{pmatrix} = -\begin{pmatrix} 0 & 0 \\ \mathcal{I}_{c(22)}^{-1}\mathcal{I}_{c(21)}\mathcal{I}_{m(11)}^{-1} & 0 \end{pmatrix} + o(1) \\
&\equiv \boldsymbol{V} + o(1),
\end{aligned}$$

and $\boldsymbol{\Gamma} = \lim_{n\to\infty} \boldsymbol{D}_n^{-1}\boldsymbol{T}_n = \boldsymbol{D}^{-1}\boldsymbol{T}$.

Theorem 3 *Consider model (7) and suppose that at the true value $\boldsymbol{\theta}^0$, Assumptions 1 holds. Then, $\hat{\boldsymbol{\theta}}_n^k$ is asymptotically normal with mean vector $\boldsymbol{\theta}^0$ and covariance matrix $n^{-1}\boldsymbol{\Sigma}_k$, where $\boldsymbol{\Sigma}_k = \boldsymbol{\Psi}_k \boldsymbol{\Omega}_V \boldsymbol{\Psi}_k^\top$, $\boldsymbol{\Psi}_k = [\boldsymbol{\Psi}_{k1}, \boldsymbol{\Psi}_{k2}]^\top$,*

$$\begin{aligned}
\boldsymbol{\Omega}_V &= \begin{pmatrix} n^{-1}\mathbb{E}(\dot{l}_{m(1)} + \dot{l}_{c(1)})(\dot{l}_{m(1)} + \dot{l}_{c(1)})^\top & n^{-1}\mathbb{E}\dot{l}_{c(2)}(\dot{l}_{m(1)} + \dot{l}_{c(1)})^\top \\ n^{-1}\mathbb{E}(\dot{l}_{m(1)} + \dot{l}_{c(1)})\dot{l}_{c(2)}^\top & n^{-1}\mathbb{E}\dot{l}_{c(2)}\dot{l}_{c(2)}^\top \end{pmatrix}, \\
\boldsymbol{\Psi}_{k1} &= (\boldsymbol{I} - \boldsymbol{\Gamma}^k)\mathcal{I}^{-1} + \boldsymbol{\Gamma}^{k-1}\boldsymbol{V}, \text{ and } \boldsymbol{\Psi}_{k2} = (\boldsymbol{I} - \boldsymbol{\Gamma}^{k-1})\mathcal{I}^{-1},
\end{aligned}$$

where $\mathcal{I}^{-1} = -n^{-1}\mathbb{E}\ddot{l}(\boldsymbol{\theta}^0)$ *is the inverse of the Fisher information. Moreover, when Assumption 2 holds,* $\boldsymbol{\Gamma}^k \to 0$ *as* $k \to \infty$*, then* $\boldsymbol{\Sigma}_k \to \mathcal{I}^{-1}$.

It is necessary to note that the marginal function needs to satisfy the information dominance condition (Assumption 2) to ensure convergence of the MBP algorithm. In other words, it requires the marginal density function to be more informative about the true parameter values relative to the multiplicative factor part. A necessary and sufficient condition for Assumption 2 to hold is that the spectral radius of $\boldsymbol{\Gamma}(\boldsymbol{\theta})$ is less than one. In effect, numerical check on the information dominance condition can be computed by a consistent estimate of $\boldsymbol{\Gamma}(\boldsymbol{\theta})$ at the first step of MBP estimate $\hat{\boldsymbol{\theta}}_n^1$. Let $\boldsymbol{\Gamma}(\hat{\boldsymbol{\theta}}_n^1) = \{\mathcal{I}_m(\hat{\boldsymbol{\theta}}_n^1)\}^{-1}\mathcal{I}_c(\hat{\boldsymbol{\theta}}_n^1)$, where $\mathcal{I}_m(\boldsymbol{\theta}) = -n^{-1}\sum_{t=1}^n \ddot{l}_{m,t}(\boldsymbol{\theta}), \mathcal{I}_c(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta}) - \mathcal{I}_m(\boldsymbol{\theta})$ and $\mathcal{I}(\boldsymbol{\theta}) = n^{-1}\sum_{t=1}^n \dot{l}_t(\boldsymbol{\theta})\dot{l}_t(\boldsymbol{\theta})^\top$. Hence, if $\rho(\boldsymbol{\Gamma}(\hat{\boldsymbol{\theta}}_n^1)) < 1$, then the information dominance condition is confirmed.

The importance of this issue is illustrated by Song et al. (2005) in a simulation study of a bivariate Gaussian copula model with exponential marginal distributions. Their results show that the model may fail to satisfy the information dominance condition when the correlation between the two random variables becomes large. In other words, MBP needs more iterations to recover the full information as the correlation increases, and it fails to work when the correlation is too high.

3.2 MMBP algorithm

To overcome the convergence difficulties of the original MBP algorithm and improve the efficiency of the estimates, we will introduce the MMBP algorithm proposed by Liu and Luger (2009). The log-likelihood function can be decomposed as

$$l(\boldsymbol{\theta}) = l_m^*(\boldsymbol{\theta}_1) + l_c^*(\boldsymbol{\theta}_1, \theta_2) = [l_m(\boldsymbol{\theta}_1) + l_c(\boldsymbol{\theta}_1, \hat{\theta}_{2,n}^1)] + [l_c(\boldsymbol{\theta}_1, \theta_2) - l_c(\boldsymbol{\theta}_1, \hat{\theta}_{2,n}^1)]. \quad (12)$$

The steps of MMBP are:

$$\text{Step 1 : Solve } \frac{\partial l_m(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} = \mathbf{0} \text{ for } \hat{\boldsymbol{\theta}}_{1,n}^1 \text{ and } \frac{\partial l_c(\hat{\boldsymbol{\theta}}_{1,n}^1, \theta_2)}{\partial \theta_2} = 0 \text{ for } \hat{\theta}_{2,n}^1.$$

$$\text{Step 2 : Solve } \frac{\partial l_m(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} + \frac{\partial l_c(\boldsymbol{\theta}_1, \hat{\theta}_{2,n}^1)}{\partial \boldsymbol{\theta}_1} = \mathbf{0} \text{ for } \hat{\boldsymbol{\theta}}_{1,n}^2$$

$$\text{and } \frac{\partial l_c(\hat{\boldsymbol{\theta}}_{1,n}^2, \theta_2)}{\partial \theta_2} = 0 \text{ for } \hat{\theta}_{2,n}^2.$$

$$\text{Step } k \text{ : Solve } \frac{\partial l_m(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} + \frac{\partial l_c(\boldsymbol{\theta}_1, \hat{\theta}_{2,n}^1)}{\partial \boldsymbol{\theta}_1} = -\left(\frac{\partial l_c(\hat{\boldsymbol{\theta}}_{1,n}^{k-1}, \hat{\theta}_{2,n}^{k-1})}{\partial \boldsymbol{\theta}_1} - \frac{\partial l_c(\hat{\boldsymbol{\theta}}_{1,n}^{k-1}, \hat{\theta}_{2,n}^1)}{\partial \boldsymbol{\theta}_1}\right)$$

$$\text{for } \hat{\boldsymbol{\theta}}_{1,n}^k \text{ and } \frac{\partial l_c(\hat{\boldsymbol{\theta}}_{1,n}^{k-1}, \theta_2)}{\partial \theta_2} = 0 \text{ for } \hat{\theta}_{2,n}^k, \quad k = 3, 4, \ldots.$$

This new algorithm is a natural choice for the models whose log-likelihood function can be additively decomposed and provides a better approximation to the true likelihood since it uses the correctly specified joint function. The estimates from each step naturally serve as starting points for the successive updating step. The main advantage of the MMBP algorithm is that it resolves the convergence issues of MBP when the correlation between the variables is high. In our simulation studies in Sect. 4, when the correlation is high, MBP fails to ensure convergence but MMBP performs better in convergence speed. In addition, MMBP algorithm also restores the consistency and asymptotic normality of the iterative estimator.

3.3 Alternative method using TMB

The above two algorithms both use the idea of iteration, which may cause the computational speed a little bit slower when the sample size becomes larger. Here, we also consider an alternative approach using R package TMB (see Kristensen et al. 2016), which is designed for estimating complex nonlinear models both with and without random effects. For example, Berentsen et al. (2018) utilized this method to estimate the parameters of a Markov-switching Poisson log-linear autoregressive model and address the inference about the underlying regimes.

Now, we first specify the log-likelihood function (10) as a C++ user template function, and then use the R function to compile, link and optimize it with the TMB package. In fact, the TMB package supplies an object with functions to evaluate the likelihood function and gradient exactly, which will improve the speed and accuracy. On the other hand, to our knowledge, there are no results about asymptotic theory based on TMB method. Thereby, we preserve MBP and MMBP algorithms to show the consistency and asymptotic normality of the MLE.

In the current paper, our model cannot consider the covariates due to its linearity. While the new proposed models can be easily extend to their log-linear version, which can include or exclude the exogenous covariates. A similar study of this problem has been discussed by Berentsen et al. (2018) and we will leave it to future research.

4 Simulation

A simulation study is conducted to evaluate the performances of the estimators with MBP, MMBP and TMB methods. For simplicity, we choose four set-ups of our proposed BP-INGARCH(1,1) models with both $\boldsymbol{A}$ and $\boldsymbol{B}$ are diagonal as follows:

$$\begin{aligned}
&\text{(A1) BPG} : \boldsymbol{\theta}_1 = (0.5, 0.4, 0.5, 0.5, 0.4, 0.5)^\top \\
&\qquad \text{with dependency parameter } \rho = 0, 0.3, 0.9; \\
&\text{(A2) BPG} : \boldsymbol{\theta}_1 = (1, 0.5, 0.4, 0.7, 0.4, 0.5)^\top \\
&\qquad \text{with dependency parameter } \rho = 0, 0.3, 0.9; \\
&\text{(B) BPF} : \boldsymbol{\theta}_1 = (0.5, 0.45, 0.5, 0.5, 0.45, 0.5)^\top \\
&\qquad \text{with dependency parameter } \gamma = -1, -0.5, 0.5, 1;
\end{aligned}$$

Table 1 Simulation results for model A1 based on MBP&MMBP

Para		ω_1	α_1	β_1	ω_2	α_2	β_2	ρ
$\rho = 0$								
$n = 200$	MADE	0.2064	0.0733	0.0574	0.1743	0.0733	0.0604	0.0785
	MSE	0.0817	0.0084	0.0052	0.0575	0.0093	0.0058	0.0091
$n = 500$	MADE	0.0928	0.0436	0.0353	0.1257	0.0490	0.0358	0.0654
	MSE	0.0155	0.0031	0.0020	0.0270	0.0036	0.0019	0.0056
$\rho = 0.3$								
$n = 200$	MADE	0.2582	0.0839	0.0692	0.2317	0.0967	0.0751	0.0541
	MSE	0.1360	0.0115	0.0074	0.1061	0.0137	0.0086	0.0043
$n = 500$	MADE	0.1183	0.0528	0.0421	0.1514	0.0546	0.0426	0.0388
	MSE	0.0259	0.0045	0.0028	0.0416	0.0049	0.0027	0.0023
$\rho = 0.9$								
$n = 200$	MADE	0.3344	0.1016	0.0914	0.3398	0.1067	0.0919	0.0153
	MSE	0.2445	0.0183	0.0132	0.2596	0.0200	0.0131	0.0004
$n = 500$	MADE	0.1391	0.0673	0.0586	0.1391	0.0706	0.0608	0.0141
	MSE	0.0353	0.0076	0.0059	0.0342	0.0082	0.0060	0.0003

MADE, mean absolute deviation error; MSE, mean squared error

$$\text{(C) BPFGM} : \boldsymbol{\theta}_1 = (0.5, 0.45, 0.5, 0.5, 0.45, 0.5)^\top$$
$$\text{with dependency parameter } \sigma = -0.5, -0.2, 0.2, 0.5.$$

For the estimation of the parameters, we use the method of randomly choosing from a uniform distribution to find out the initial values and set the sample size $n = 200$ and 500 with $m = 200$ replications for each choice of parameters. The estimators are compared in terms of their mean absolute deviation error (MADE) and mean squared error (MSE) according to the following formulas:

$$\text{MADE} = \frac{1}{m}\sum_{j=1}^{m} |\hat{\vartheta}_j - \vartheta^0|, \quad \text{MSE} = \frac{1}{m}\sum_{j=1}^{m} (\hat{\vartheta}_j - \vartheta^0)^2, \tag{13}$$

where $\hat{\vartheta}_j$ is the estimator of ϑ^0 in the jth replication.

4.1 Simulation based on MBP & MMBP algorithms

First, we conduct simulations about the above four set-ups of models with MBP and MMBP algorithms by Matlab. More specifically, in both two iteration algorithms, the first step is followed by the constrained nonlinear optimization function fmincon to obtain the minimum of constrained nonlinear multivariable function. Then for the next iteration steps, we use lsqnonlin function to solve the function by minimizing the sum of components' squares with bound constraints.

Table 2 Simulation results for model A2 based on MBP&MMBP

Para		ω_1	α_1	β_1	ω_2	α_2	β_2	ρ
$\rho = 0$								
$n = 200$	MADE	0.3253	0.0677	0.0591	0.2306	0.0735	0.0608	0.0612
	MSE	0.1705	0.0071	0.0056	0.0996	0.0095	0.0058	0.0056
$n = 500$	MADE	0.2036	0.0423	0.0333	0.1654	0.0454	0.0364	0.0425
	MSE	0.0671	0.0031	0.0018	0.0472	0.0033	0.0020	0.0027
$\rho = 0.3$								
$n = 200$	MADE	0.4003	0.0776	0.0684	0.2844	0.0916	0.0716	0.0503
	MSE	0.2766	0.0096	0.0075	0.1687	0.0132	0.0081	0.0038
$n = 500$	MADE	0.2500	0.0517	0.0396	0.1778	0.0508	0.0406	0.0283
	MSE	0.1091	0.0043	0.0024	0.0562	0.0041	0.0024	0.0013
$\rho = 0.9$								
$n = 200$	MADE	0.6322	0.1126	0.0967	0.4608	0.1151	0.0964	0.0205
	MSE	0.2027	0.0221	0.0141	0.4745	0.0244	0.0151	0.0054
$n = 500$	MADE	0.2832	0.0647	0.0522	0.2158	0.0702	0.0572	0.0087
	MSE	0.1373	0.0074	0.0048	0.0797	0.0081	0.0053	0.0001

MADE, mean absolute deviation error; MSE, mean squared error

Table 3 Simulation results for model B based on MBP&MMBP

Para		ω_1	α_1	β_1	ω_2	α_2	β_2	γ
$\gamma = -1$								
$n = 200$	MADE	0.2192	0.0620	0.0552	0.2048	0.0634	0.0568	0.4218
	MSE	0.0862	0.0061	0.0046	0.0915	0.0068	0.0052	0.2597
$n = 500$	MADE	0.1266	0.0401	0.0373	0.1218	0.0389	0.0360	0.3834
	MSE	0.0263	0.0025	0.0021	0.0249	0.0023	0.0019	0.1881
$\gamma = -0.5$								
$n = 200$	MADE	0.2220	0.0607	0.0538	0.2044	0.0645	0.0551	0.2989
	MSE	0.0879	0.0058	0.0044	0.0868	0.0063	0.0047	0.1214
$n = 500$	MADE	0.1316	0.0395	0.0366	0.1229	0.0384	0.0358	0.2737
	MSE	0.0281	0.0024	0.0020	0.0261	0.0023	0.0020	0.0995
$\gamma = 0.5$								
$n = 200$	MADE	0.2377	0.0616	0.0537	0.2130	0.0581	0.0518	0.3100
	MSE	0.1010	0.0061	0.0045	0.1060	0.0055	0.0042	0.1478
$n = 500$	MADE	0.1432	0.0411	0.0375	0.1311	0.0399	0.0376	0.2274
	MSE	0.0331	0.0025	0.0021	0.0298	0.0025	0.0022	0.0872
$\gamma = 1$								
$n = 200$	MADE	0.2486	0.0638	0.0552	0.2294	0.0604	0.0545	0.3193
	MSE	0.1112	0.0065	0.0048	0.1254	0.0060	0.0047	0.1579
$n = 500$	MADE	0.1510	0.0429	0.0391	0.1417	0.0433	0.0405	0.2154
	MSE	0.0365	0.0028	0.0023	0.0356	0.0029	0.0025	0.0775

MADE, mean absolute deviation error; MSE, mean squared error

Table 4 Simulation results for model C based on MBP&MMBP

Para		ω_1	α_1	β_1	ω_2	α_2	β_2	σ
$\sigma = -0.5$								
$n = 200$	MADE	0.1714	0.0699	0.0642	0.1837	0.0586	0.0503	0.2327
	MSE	0.0558	0.0076	0.0065	0.0588	0.0054	0.0042	0.0694
$n = 500$	MADE	0.1108	0.0384	0.0383	0.0977	0.0374	0.0352	0.1720
	MSE	0.0227	0.0025	0.0024	0.0186	0.0021	0.0019	0.0393
$\sigma = -0.2$								
$n = 200$	MADE	0.1806	0.0624	0.0592	0.1665	0.0569	0.0540	0.2062
	MSE	0.0676	0.0060	0.0054	0.0521	0.0051	0.0045	0.0709
$n = 500$	MADE	0.1171	0.0399	0.0331	0.1222	0.0367	0.0322	0.1422
	MSE	0.0234	0.0023	0.0018	0.0292	0.0021	0.0017	0.0295
$\sigma = 0.2$								
$n = 200$	MADE	0.2297	0.0631	0.0549	0.2513	0.0697	0.0597	0.1756
	MSE	0.1100	0.0055	0.0042	0.1185	0.0080	0.0059	0.0458
$n = 500$	MADE	0.1215	0.0382	0.0358	0.1269	0.0356	0.0327	0.1135
	MSE	0.0235	0.0023	0.0019	0.0317	0.0020	0.0017	0.0191
$\sigma = 0.5$								
$n = 200$	MADE	0.2133	0.0638	0.0557	0.2266	0.0641	0.0621	0.1632
	MSE	0.0930	0.0068	0.0049	0.0848	0.0070	0.0062	0.0378
$n = 500$	MADE	0.1606	0.0390	0.0342	0.1261	0.0372	0.0332	0.1028
	MSE	0.0473	0.0023	0.0018	0.0287	0.0023	0.0020	0.0171

MADE, mean absolute deviation error; MSE, mean squared error

The summary of the simulation results are given in Tables 1, 2, 3 and 4. As the sample size increases, the values of MADE and MSE gradually decrease. Note that the special case $\rho = 0$ reduces the bivariate case to two independent Poisson distributions. As discussed by Liu and Luger (2009), MBP and MMBP algorithms show almost the same efficiencies under lower correlation values, but MBP fails to work at all under high correlation values. For example, $\rho = 0$ or 0.3 in Tables 1 and 2, MBP and MMBP yield the same values of MADE and MSE for each estimate of interest, whereas when ρ turns to be 0.9, MBP algorithm fails to converge so we only display the results for MMBP algorithm.

Tables 3 and 4 refer to BPF and BPFGM INGARCH(1,1) models, respectively. To evaluate the performance of estimates under two algorithms, we choose the relative small values of the correlations. It is obvious to see that the autoregressive conditional heteroscedastic and GARCH effect parameters in each set-up show smaller values of MADE and MSE compared to other parameters. In addition, the average iteration steps of MBP and MMBP algorithms are reported in Table 5. From it, we can find that the iteration steps also decrease as the sample size increases. On the whole, MMBP algorithm uses less iteration steps to converge than MBP in each configuration of parameters.

Table 5 Average iteration steps of MBP and MMBP algorithms

Model A1	n	MBP	MMBP	Model A2	n	MBP	MMBP
$\rho = 0$	200	5.48	4.51	$\rho = 0$	200	4.85	3.94
	500	4.97	3.76		500	4.16	3.34
$\rho = 0.3$	200	7.65	5.84	$\rho = 0.3$	200	7.07	6.83
	500	6.94	4.36		500	6.48	4.34
$\rho = 0.9$	200	–	4.41	$\rho = 0.9$	200	–	5.07
	500	–	3.59		500	–	3.82
Model B	n	MBP	MMBP	Model C	n	MBP	MMBP
$\gamma = -1$	200	7.82	7.50	$\sigma = -0.5$	200	6.92	6.39
	500	7.17	5.86		500	6.57	5.67
$\gamma = -0.5$	200	5.60	4.81	$\sigma = -0.2$	200	5.43	4.50
	500	5.06	4.02		500	4.58	3.72
$\gamma = 0.5$	200	5.83	4.78	$\sigma = 0.2$	200	5.20	4.24
	500	5.53	4.11		500	4.72	3.58
$\gamma = 1$	200	6.08	5.07	$\sigma = 0.5$	200	5.67	4.91
	500	5.33	4.28		500	5.05	3.99

4.2 Simulation based on TMB

In this subsection, we also conduct some simulations based on TMB method. For saving space of the paper, we only consider the first two set-ups (A1 & A2) to see the performance of this algorithm, especially when the dependency parameter becomes larger. In our case, we opt for the R-routine constrOptim to obtain the MLE with specified constraints.

Similar to the previous results in Tables 1 and 2, all seven estimates show quite small values of MADE and MSE, which indicates the feasibility and rationality of this algorithm. Furthermore, the values of MADE and MSE both decrease as the sample size increases. From Tables 6 and 7, we find that under the low correlations ($\rho = 0$ or 0.3), the results based on TMB do not show much difference with the results of MBP and MMBP. What's most interest us is the performance when the dependency parameter ρ becomes larger. For the case $\rho = 0.9$, one can see that in both models, most of the estimates for ω_1 and ω_2 show smaller MADE and MSE compared to the corresponding case in Tables 1 and 2. Besides, this TMB method indeed increases the computational speed. Although the simulations based on the above three methods are conducted by different software (say TMB by R versus to MBP & MMBP by Matlab), it is feasible to compare the time by intuition. In the practical implementation, we find that with TMB method, the estimation for model A1 takes only 5 s with 500 observations at one replication. However in the same settings, MBP and MMBP need a few minutes at one replication, which further verifies that TMB can speed up the computation.

Table 6 Simulation results for model A1 based on TMB

Para		ω_1	α_1	β_1	ω_2	α_2	β_2	ρ
$\rho = 0$								
$n = 200$	MADE	0.1723	0.0708	0.0536	0.1760	0.0751	0.0571	0.0947
	MSE	0.0569	0.0081	0.0046	0.0554	0.0088	0.0053	0.0126
$n = 500$	MADE	0.0953	0.0423	0.0331	0.1000	0.0413	0.0311	0.0802
	MSE	0.0148	0.0028	0.0017	0.0193	0.0029	0.0016	0.0083
$\rho = 0.3$								
$n = 200$	MADE	0.1376	0.0692	0.0528	0.1342	0.0647	0.0541	0.0774
	MSE	0.0324	0.0075	0.0045	0.0331	0.0065	0.0047	0.0086
$n = 500$	MADE	0.0949	0.0403	0.0330	0.0907	0.0410	0.0352	0.0613
	MSE	0.0137	0.0026	0.0017	0.0135	0.0027	0.0019	0.0050
$\rho = 0.9$								
$n = 200$	MADE	0.1462	0.0512	0.0539	0.1308	0.0521	0.0539	0.0159
	MSE	0.0296	0.0047	0.0047	0.0249	0.0045	0.0045	0.0004
$n = 500$	MADE	0.1320	0.0331	0.0420	0.1308	0.0340	0.0420	0.0148
	MSE	0.0226	0.0017	0.0025	0.0218	0.0017	0.0025	0.0003

MADE, mean absolute deviation error; MSE, mean squared error.

Table 7 Simulation results for model A2 based on TMB

Para		ω_1	α_1	β_1	ω_2	α_2	β_2	ρ
$\rho = 0$								
$n = 200$	MADE	0.3498	0.0728	0.0569	0.2952	0.0737	0.0525	0.0704
	MSE	0.2281	0.0084	0.0052	0.1623	0.0082	0.0045	0.0077
$n = 500$	MADE	0.2106	0.0408	0.0293	0.1281	0.0379	0.0303	0.0542
	MSE	0.0823	0.0026	0.0014	0.0313	0.0024	0.0014	0.0044
$\rho = 0.3$								
$n = 200$	MADE	0.2749	0.0633	0.0509	0.1833	0.0662	0.0561	0.0600
	MSE	0.1542	0.0065	0.0040	0.0618	0.0067	0.0049	0.0052
$n = 500$	MADE	0.1858	0.0431	0.0330	0.1197	0.0382	0.0313	0.0405
	MSE	0.0547	0.0029	0.0018	0.00237	0.0023	0.0016	0.0025
$\rho = 0.9$								
$n = 200$	MADE	0.2216	0.0491	0.0401	0.1607	0.0507	0.0479	0.0103
	MSE	0.0721	0.0037	0.0026	0.0378	0.0042	0.0036	0.0002
$n = 500$	MADE	0.1775	0.0331	0.0271	0.1084	0.0329	0.0341	0.0088
	MSE	0.0450	0.0018	0.0011	0.0170	0.0017	0.0017	0.0001

MADE, mean absolute deviation error; MSE, mean squared error

Table 8 Descriptive statistics for the number of syphilis cases

States	No.	Mean	Variance	Minimum	Median	Maximum	Cross-Corr.
Pennsylvania	209	3.5167	6.8759	0	3	12	−0.1355
Maryland	209	3.4737	9.2794	0	3	15	

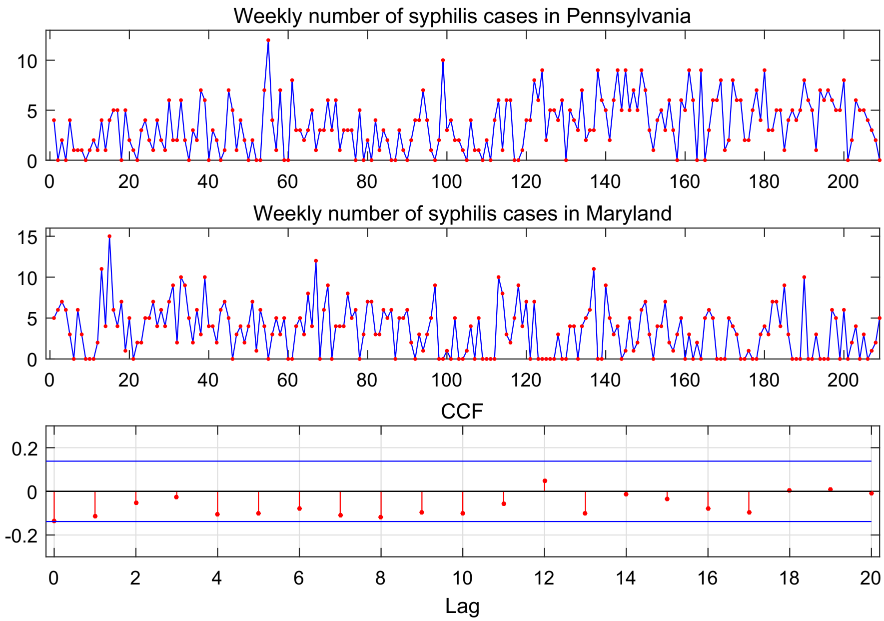

Fig. 1 Top: Number of syphilis cases in Pennsylvania. Middle: Number of syphilis cases in Maryland. Bottom: CCF of two data series

5 Application

We now illustrate the ability and flexibility of our proposed BP-INGARCH(1,1) models with a real time series data set. Here, we still consider the case that parameter matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are both diagonal for ease of presentation. And the extension to non-diagonal case is straightforward. Now, we consider the weekly numbers of syphilis cases from 2007 to 2010 in two states of the USA, Pennsylvania and Maryland. This is a part of data set given in the ZIM package available for download at the website https://cran.r-project.org/web/packages/ZIM/ZIM.pdf, which was also studied by Borges et al. (2017).

Table 8 displays some descriptive statistics for these data. From it, one can find that the sample variance in each series is much larger than the sample mean, which both show overdispersion. The cross-correlation coefficient between two series is −0.1355, and we use the MMBP algorithm to obtain the estimates of our proposed models. Two data series paths together with their cross-correlation function (CCF) are plotted in Figure 1. The CCF reveals the negative correlation between these two series. Besides, the

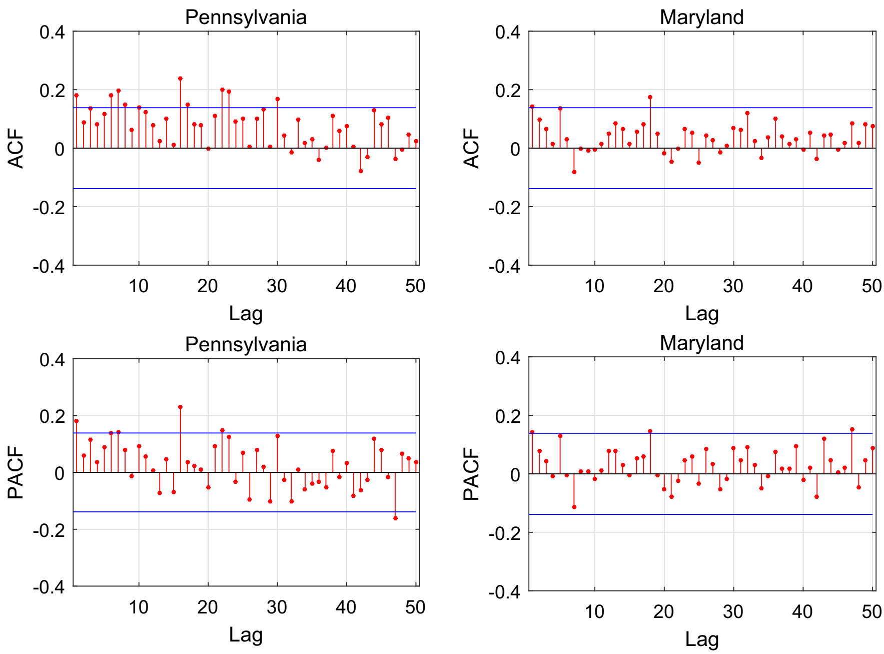

Fig. 2 Top left: ACF of Pennsylvania. Top right: ACF of Maryland. Bottom left: PACF of Pennsylvania. Bottom right: PACF of Maryland

sample autoregression function (ACF) and the sample partial autoregression function (PACF) are presented in Figure 2.

For comparison, we fit BPG, BPF and BPFGM models along with the aforementioned BP model (Cui and Zhu 2018), Liu's model (Liu 2012) and Fokianos's model (Fokianos et al. 2019). To see the accuracy of the estimates of parameters, we calculate the standard errors (SE) for parameters in each model. Table 9 summarizes the estimate of parameter with SE in parentheses, Akaike information criterion (AIC) and Bayesian information criterion (BIC) values. From it, one can see that the first six parameters show similar estimates in the above six different models. It is important to mention that in Fokianos's model, we use the quasi maximum likelihood estimation (QMLE) to obtain the first six parameters and then conduct a parametric bootstrap to estimate copula parameter.

Now, we focus on the parameter for describing the dependence structure in each model. As can be seen, the dependency parameters of our proposed models together with Fokianos's model in Table 9 all yield negative estimates which are in accordance with the negative correlation (-0.1355) between two data series. However, the parameter δ in BP model still shows positive value with a much larger SE. The estimate of $\phi(= 0.0001)$ in Liu's model in fact is the predetermined lower bound of the algorithm, which further illustrates that their model cannot capture negative cross-correlation. Moreover, AIC and BIC values reveal the improvement in fit when we use the models defined through multiplicative factors. Notice that although there are six

Table 9 Comparison of different models for syphilis cases

Models	Estimates		AIC	BIC
BPG	$\hat{\omega}_1 = 0.1590(0.8565)$	$\hat{\omega}_2 = 1.5425(0.7376)$	2160.7015	2183.8593
	$\hat{\alpha}_1 = 0.9018(0.2883)$	$\hat{\alpha}_2 = 0.3886(0.2360)$		
	$\hat{\beta}_1 = 0.0570(0.0566)$	$\hat{\beta}_2 = 0.1617(0.0713)$		
	$\hat{\rho} = -0.0597(0.0577)$			
BPF	$\hat{\omega}_1 = 0.1621(0.9726)$	$\hat{\omega}_2 = 1.5041(0.7797)$	2158.3271	2181.4849
	$\hat{\alpha}_1 = 0.9010(0.3225)$	$\hat{\alpha}_2 = 0.4061(0.2517)$		
	$\hat{\beta}_1 = 0.0569(0.0576)$	$\hat{\beta}_2 = 0.1561(0.0711)$		
	$\hat{\gamma} = -0.6331(0.2803)$			
BPFGM	$\hat{\omega}_1 = 0.1644(0.8165)$	$\hat{\omega}_2 = 1.5185(0.8646)$	2159.3170	2182.4749
	$\hat{\alpha}_1 = 0.9006(0.3390)$	$\hat{\alpha}_2 = 0.4016(0.2573)$		
	$\hat{\beta}_1 = 0.0567(0.0597)$	$\hat{\beta}_2 = 0.1571(0.0786)$		
	$\hat{\sigma} = -0.2499(0.1249)$			
BP	$\hat{\omega}_1 = 0.1810(0.0565)$	$\hat{\omega}_2 = 1.7186(0.7538)$	2162.4678	2185.6257
	$\hat{\alpha}_1 = 0.8965(0.0402)$	$\hat{\alpha}_2 = 0.3460(0.2374)$		
	$\hat{\beta}_1 = 0.0575(0.0336)$	$\hat{\beta}_2 = 0.1629(0.0713)$		
	$\hat{\delta} = 0.7468(0.8760)$			
Liu's	$\hat{\omega}_1 = 0.1755(0.0650)$	$\hat{\omega}_2 = 1.7113(0.7406)$	2163.7861	2186.9440
	$\hat{\alpha}_1 = 0.8971(0.0442)$	$\hat{\alpha}_2 = 0.3432(0.2394)$		
	$\hat{\beta}_1 = 0.0571(0.0349)$	$\hat{\beta}_2 = 0.1637(0.0713)$		
	$\hat{\phi} = 0.0001(0.1519)$			
Fokianos's	$\hat{\omega}_1 = 0.1756(0.0653)$	$\hat{\omega}_2 = 1.7113(0.7342)$	2161.7842	2181.6635
	$\hat{\alpha}_1 = 0.8971(0.0453)$	$\hat{\alpha}_2 = 0.3432(0.2405)$		
	$\hat{\beta}_1 = 0.0574(0.0361)$	$\hat{\beta}_2 = 0.1637(0.0691)$		
	$\hat{\psi} = -0.2176(0.1028)$			

AIC, Akaike information criterion; BIC, Bayesian information criterion

parameters by QMLE procedure in Fokianos's model, their AIC is larger than our proposed models. It is worth noting that for estimating the copula parameter in Fokianos's model, we let the parametric bootstrap algorithm choose between the Gaussian and Clayton copula with 500 realizations. Indeed, there are 262 cases out of 500 (52.4%) selecting Clayton copula, and we estimate ψ by the average of these realizations. Nevertheless, this way for selecting the correct copula can be suspicious because we cannot tell which copula structure is better for this data.

We also consider the Pearson residuals defined by $e_t = \frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}$ to further examine the adequacy fitting of our models. Under the correct model, the sequence e_t should be a white noise sequence with constant variance (see Kedem and Fokianos 2002, Sect. 1.6.3). And we can substitute the λ_t by $\lambda_t(\hat{\theta})$ to obtain $\hat{e}_t$. Furthermore, we show cumulative periodogram plot for BPF model (see Brockwell and Davis 1991, Sect. 10.2). From Figure 3, we can see that the standardized cumulative periodogram lies in the Kolmogorov–Smirnov bounds with level $\alpha = 0.05$, which indicates the

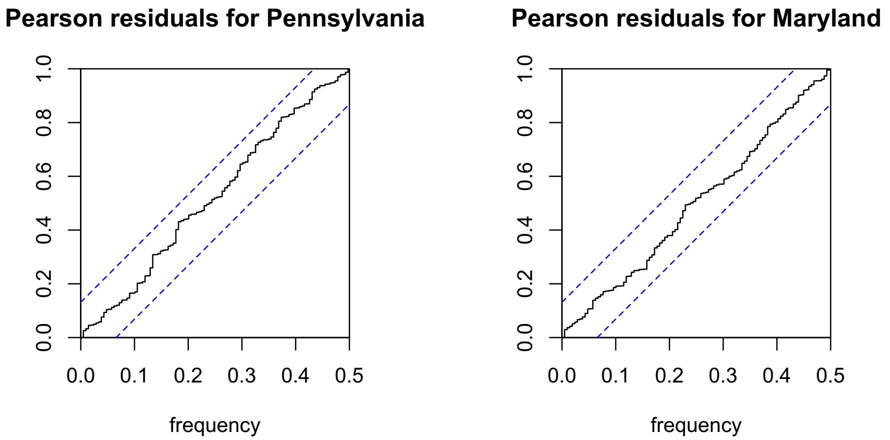

Fig. 3 Cumulative periodogram plot of the Pearson residuals of the weekly numbers of syphilis cases fitted by BPF model

Table 10 MADE and MSE for predicted values for each model

Model	BPG	BPF	BPFGM	BP	Liu's	Fokianos's
MADE	3.1768	3.1812	3.1812	3.2032	3.1955	3.2004
MSE	3.5584	3.5629	3.5634	3.5845	3.5924	3.5925

whiteness of the Pearson residuals. Similar results hold for the other two models (BPG & BPFGM), thus we omit them.

In addition, we perform an out-of-sample forecasting exercise on the aforementioned models for comparison. First of all, we split the data in two parts. The first part has size $T_0 = 157$ (year 2007–2009), with the observations $\{\boldsymbol{Y}_t, t = 1, \ldots, T_0\}$ being used for initial estimation of the model, while the remaining observations $\{\boldsymbol{Y}_t, t = T_0 + 1, \ldots, T(= 209)\}$ will be used for a forecasting. Then, we predict the number of syphilis cases during the year 2010 by computing the one-step ahead forecast of $\boldsymbol{Y}_t$ using the conditional expectation $\mathbb{E}(\boldsymbol{Y}_t|\mathcal{F}_{t-1})$ and repeat the above exercise for $t = T_0 + 1, \ldots, T$. Given the forecast path $\widehat{\boldsymbol{Y}}_t$, we use the MADE and MSE to evaluate the performance of these models, where the error terms are measured by supremum norm and Euclidean norm, respectively. Table 10 shows the results and we can see that the our proposed models based on multiplicative factors perform better in forecasting. In general, we choose BPF model as the preferred model among these models for fitting this data set based on its smallest AIC and BIC.

6 Conclusion

The current paper considers a class of flexible bivariate Poisson INGARCH models by introducing multiplicative factors for dependence modelling. Some important properties such as stationarity and ergodicity of the process are discussed. For estimating the parameters of the models, MBP, MMBP algorithms together with the method based on TMB have been employed and the consistency and asymptotic normality of estimates are also established. The numerical simulation shows that the estimation results are reliable as long as the sample size is large enough. Finally, we apply the newly proposed BP-INGARCH(1,1) models to an empirical example to demonstrate their better performances.

Acknowledgements The authors are very grateful to the Editor, Associate Editor and anonymous referee for providing several constructive comments which led to a significant improvement of the paper.

Appendix A

Let X and Y be two independent random variables from Poisson distribution with cdf F_1 and F_2. Then, the joint cdf of newly proposed BP distribution $C(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ should satisfy the definition of bivariate probability distribution and the following properties:

$$
\begin{aligned}
&(a)\ \text{Right continuity}: \ C(x, y+0) = C(x, y);\ C(x+0, y) = C(x, y);\\
&(b)\ \text{Boundedness}: \ \lim_{x\to 0^-} C(x, y) = \lim_{y\to 0^-} C(x, y) = 0,\\
&\qquad \lim_{x,y\to +\infty} C(x, y) = 1;\\
&(c)\ \text{Monotonicity}: \ C(x, y) \text{ is monotonically non-decreased for } x \text{ and } y,\\
&\qquad \text{respectively};\\
&(d)\ \text{Nonnegativity}: \ \text{for any } x_1 \leq x_2,\ y_1 \leq y_2 \text{ such that}\\
&\qquad C(x_2, y_2) - C(x_1, y_2) - C(x_2, y_1) + C(x_1, y_1) \geq 0.
\end{aligned}
$$

It is worth mentioning that from a different prospective, our model (6) can be viewed as a bivariate Poisson distribution constructed by linking function, copula. By imposing a positive multiplicative factor Z^{-1}, the properties (a) and (b) can be easily verified by relation of events and nature of probability. Now, we turn to consider properties (c) and (d). By the form of equation (6) and the definition of Eqs. (3)–(5), we find that the multiplicative factors $c(\cdot, \cdot)$ are all nonnegative bounded functions on $[0, 1]^2$ for suitably chosen range of values of the dependency parameters. Hence, when $x_2 \geq x_1$, we have

$$
C(x_2, y) - C(x_1, y) = Z^{-1} \sum_{k=x_1+1}^{x_2} \sum_{s=0}^{y} \frac{\lambda_1^k \lambda_2^s}{k!s!} \exp\{-(\lambda_1 + \lambda_2)\} c(F_1(k), F_2(s)) \geq 0.
$$

The case for $y_2 \geq y_1$ is similar, so (c) is verified. As for nonnegativity, we have the following result by definition of the cdf, for any $x_1 \leq x_2$, $y_1 \leq y_2$,

$$
\begin{aligned}
&C(x_2, y_2) - C(x_1, y_2) - C(x_2, y_1) + C(x_1, y_1)\\
&= \mathbb{P}(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) \geq 0.
\end{aligned}
$$

Hence, property (d) is also verified. □

In fact, $Z(\lambda_1, \lambda_2, \theta)$ is a bounded factor because the series converges by its definition. It also can be viewed as a weight, which makes the sum of our model's pdf equals to one. One can find similar methods to define some distribution families, for example, Efron (1986) considered the double exponential families by adding a nonlinear constant $c(\mu, \theta, n)$; Shmueli et al. (2005) tackled the important task of characterizing the Conway–Maxwell–Poisson distribution by dividing a series as weight. In our bivariate case, it is difficult to give the explicit approximation to $Z^{-1}(\lambda_1, \lambda_2, \theta)$ due to the copula structure. Instead, we could use the truncated double summation at some k_1, k_2 to compute Z^{-1}, which will give the reasonable approximation with a small error. More specifically, we write $Z(\lambda_1, \lambda_2, \theta) = \widehat{Z}_{k_1,k_2} + R_{k_1,k_2}$, where

$$\widehat{Z}_{k_1,k_2} = \sum_{y_1=0}^{k_1}\sum_{y_2=0}^{k_2} \frac{\lambda_1^{y_1}\lambda_2^{y_2}}{y_1!y_2!} \exp\{-(\lambda_1+\lambda_2)\} c(F_1(y_1), F_2(y_2)),$$

$$R_{k_1,k_2} = \left(\sum_{y_1=k_1+1}^{\infty}\sum_{y_2=0}^{k_2} + \sum_{y_1=0}^{k_1}\sum_{y_2=k_2+1}^{\infty} + \sum_{y_1=k_1+1}^{\infty}\sum_{y_2=k_2+1}^{\infty}\right) \frac{\lambda_1^{y_1}\lambda_2^{y_2}}{y_1!y_2!} \times \exp\{-(\lambda_1+\lambda_2)\} c(F_1(y_1), F_2(y_2)).$$

Similar to Shmueli et al. (2005), we define the relative truncation error (TE) as

$$\mathrm{TE} = \frac{\widehat{Z}_{k_1,k_2}^{-1} - Z^{-1}}{\widehat{Z}_{k_1,k_2}^{-1}} = \frac{R_{k_1,k_2}}{Z}. \tag{14}$$

Here, we show some numerical simulations of different combinations with our proposed model. We choose the Poisson intensity pairs as $(\lambda_1, \lambda_2) = (1, 1),\ (1, 2)$ and (3,2). For sake of illustration, we let $k = k_1 = k_2$ be the same truncated number, which will specified in different cases.

$$(a)\ \mathrm{BPG}: \rho = -0.5, -0.2, 0.2, 0.5;$$
$$(b)\ \mathrm{BPF}: \gamma = -1, -0.5, 0.5, 1;$$
$$(c)\ \mathrm{BGFGM}: \sigma = -0.5, -0.2, 0.2, 0.5.$$

Based on the fact that the kurtosis of univariate Poisson distribution is λ^{-1} (λ as the Poisson intensity), leading to the flatter curve plot of pmf as λ increases. Therefore, we know that the truncated number k needs to be increased to approximate Z as λ_1, λ_2 increase. In the above three scenarios, we choose the largest value of k to be 6 as λ_1 and λ_2 are relative small. If we increase λ_1 or λ_2, then k may be 10 or much larger to approximate the infinite sum. From Table 11, we conclude that the truncated sum can be an approximate of Z in practice.

Furthermore, following by the discussion of Appendix B in Shmueli et al. (2005) and elementary analysis, we know that Z converges and both Z and Z^{-1} are bounded by some positive constant. Therefore, going back to our model (6), we can view Z^{-1} as

Table 11 TE for the above three cases

BPG		ρ			
(λ_1, λ_2)	k	−0.5	−0.2	0.2	0.5
(1,1)	4	0.0013	0.0049	0.0093	0.0113
(1,2)	4	0.0223	0.0434	0.0667	0.0795
(3,2)	6	0.0205	0.0314	0.0431	0.0495
BPF		γ			
(λ_1, λ_2)	k	−1	−0.5	0.5	1
(1,1)	4	0.0065	0.0069	0.0077	0.0080
(1,2)	4	0.0492	0.0526	0.0590	0.0619
(3,2)	6	0.0345	0.0361	0.0391	0.0405
BPFGM		σ			
(λ_1, λ_2)	k	−0.5	−0.2	0.2	0.5
(1,1)	4	0.0075	0.0074	0.0072	0.0070
(1,2)	4	0.0534	0.0549	0.0570	0.0587
(3,2)	6	0.0427	0.0397	0.0355	0.0322

a regularity constant weight. The rest terms by removing Z^{-1} at the right-hand side of (6) will dominate the pmf and can be convenient to use, thus we only consider the rest dominated terms for estimation and inference parts. As mentioned by Efron (1986), he suggested to leave out the highly nonlinear multiplicative factor when estimating the parameters.

Appendix B

Proof of Theorem 1 When considering BPFGM$(\lambda_{t,1}, \lambda_{t,2}, \sigma)$ model, this theorem can be proved using arguments similar to Cui and Zhu (2018, Theorem 1). So we only consider BPG$(\lambda_{t,1}, \lambda_{t,2}, \rho)$ and BPF$(\lambda_{t,1}, \lambda_{t,2}, \gamma)$ models. The proof employs the theory of Markov chain again, but a little difference from those in Cui and Zhu (2018) is that we introduce a uniform constant bound for the corresponding multiplicative function. First note that $\{\boldsymbol{\lambda}_t\}$ has at least one stationary distribution, refer to Liu (2012) for more details. From (8), it is easy to see that $(\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{\omega}$ is a reachable state if $\boldsymbol{Y}_{t-1} = \boldsymbol{Y}_{t-2} = \cdots = \boldsymbol{0}$ for some $t \in \mathbb{N}$ large enough. Then we only have to show that $\{\boldsymbol{\lambda}_t\}$ is an e-chain, which can naturally guarantee the existence of a unique invariant probability measure by the main virtue of Meyn and Tweedie (2009, Theorem 18.8.4). To see this, we recall the definition of e-chain, i.e., for any continuous function f with compact support defined on $[0, \infty) \times [0, \infty)$ and $\varepsilon > 0$, there exists an $\eta > 0$ such that $|\mathbb{P}^k_{\boldsymbol{x}_1} f - \mathbb{P}^k_{\boldsymbol{z}_1} f| < \varepsilon$, for $\|\boldsymbol{x}_1 - \boldsymbol{z}_1\|_p < \eta$ and all $k \geq 1$, where $\boldsymbol{x}_1 = (x_{1,1}, x_{1,2})^\top$, $\boldsymbol{z}_1 = (z_{1,1}, z_{1,2})^\top$, $\mathbb{P}^k_{\boldsymbol{x}_1} f = \mathbb{E}\{f(\boldsymbol{\lambda}_k)|\boldsymbol{\lambda}_0 = \boldsymbol{x}\}$. Without loss of generality, assume $|f| \leq 1$, M is a finite constant and take ε' and η sufficiently small such that $\varepsilon' + 8M\eta/(1 - \|\boldsymbol{A}\|_p) < \varepsilon$ and

$|f(\boldsymbol{x}_1) - f(\boldsymbol{z}_1)| < \varepsilon'$ whenever $\|\boldsymbol{x}_1 - \boldsymbol{z}_1\|_p < \eta$, for some $p \in [1, \infty]$. When we choose Gaussian factor, denote $c_1 = Z^{-1}(x_{1,1}, x_{1,2}, \rho)c_\rho(F_1(x_{1,1}), F_2(x_{1,2}))$, $c_2 = Z^{-1}(z_{1,1}, z_{1,2}, \rho)c_\rho(F_1(z_{1,1}), F_2(z_{1,2}))$. When it comes to Frank factor, they turn to be $c_1 = Z^{-1}(x_{1,1}, x_{1,2}, \gamma)c_\gamma(F_1(x_{1,1}), F_2(x_{1,2}))$, $c_2 = Z^{-1}(z_{1,1}, z_{1,2}, \gamma)c_\gamma(F_1(z_{1,1}), F_2(z_{1,2}))$. Hence, according to (3), (4) and the fact Z^{-1} is bounded, one can easily see that there exists a finite constant M, such that $|c_1|, |c_2| \le M$.

For the case $k = 1$,

$$
\begin{aligned}
&|\mathbb{P}_{\boldsymbol{x}_1} f - \mathbb{P}_{\boldsymbol{z}_1} f|\\
&= \Bigg|\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}[f(\boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{B}(m,n)^\top)p(m,n|\boldsymbol{x}_1)\\
&\quad - |f(\boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{z}_1 + \boldsymbol{B}(m,n)^\top)p(m,n|\boldsymbol{z}_1)]\Bigg|\\
&\le \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} p(m,n|\boldsymbol{x}_1)|f(\boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{B}(m,n)^\top) - f(\boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{z}_1 + \boldsymbol{B}(m,n)^\top)|\\
&\quad + \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} |p(m,n|\boldsymbol{x}_1) - p(m,n|\boldsymbol{z}_1)||f(\boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{z}_1 + \boldsymbol{B}(m,n)^\top)|\\
&= I_1 + I_2,
\end{aligned}
$$

where $p(m, n|\boldsymbol{x}_1)$ and $p(m, n|\boldsymbol{z}_1)$ are the pmfs of BPG$(x_{1,1}, x_{1,2}, \rho)$ (or BPF $(x_{1,1}, x_{1,2}, \gamma)$) and BPG $(z_{1,1}, z_{1,2}, \rho)$ (or BPF$(z_{1,1}, z_{1,2}, \gamma)$) given by (6). We start to formulate the main part of I_2, first suppose $c_1 \le c_2$,

$$
\begin{aligned}
&\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} |p(m,n|\boldsymbol{x}_1) - p(m,n|\boldsymbol{z}_1)|\\
&\le \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \left|\frac{x_{1,1}^m x_{1,2}^n}{m!n!}\mathrm{e}^{-(x_{1,1}+x_{1,2})}c_1 - \frac{z_{1,1}^m z_{1,2}^n}{m!n!}\mathrm{e}^{-(z_{1,1}+z_{1,2})}c_2\right|\\
&\le \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} M\left|\frac{x_{1,1}^m}{m!}\mathrm{e}^{-x_{1,1}} - \frac{z_{1,1}^m}{m!}\mathrm{e}^{-z_{1,1}}\right|\frac{x_{1,2}^n}{n!}\mathrm{e}^{-x_{1,2}}\\
&\quad + \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} M\left|\frac{x_{1,2}^n}{n!}\mathrm{e}^{-x_{1,2}} - \frac{z_{1,2}^n}{n!}\mathrm{e}^{-z_{1,2}}\right|\frac{z_{1,1}^m}{m!}\mathrm{e}^{-z_{1,1}}\\
&\le M\sum_{i=0}^{\infty}|p(i|x_{1,1}) - p(i|z_{1,1})| + M\sum_{i=0}^{\infty}|p(i|x_{1,2}) - p(i|z_{1,2})|.
\end{aligned}
$$

By the proof of Wang et al. (2014, Lemma 6.4), we know that $\sum_{i=0}^{\infty}|p(i|x_1) - p(i|z_1)| \le 2(1-\mathrm{e}^{-|x_1-z_1|})$, where $p(i|x)$ is the pmf of a univariate Poisson distribution with intensity x evaluated at i. And since $|x_{1,i} - z_{1,i}| \le \|\boldsymbol{x}_1 - \boldsymbol{z}_1\|_1 \le c_p \|\boldsymbol{x}_1 - \boldsymbol{z}_1\|_p$, for $i = 1, 2$ and any $1 \le p \le \infty$, where $c_p = 2^{1-1/p} \le 2$, so for any $\boldsymbol{x}_1, \boldsymbol{z}_1$ and $p \in [1, \infty]$, we have

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}|p(m,n|\boldsymbol{x}_1)-p(m,n|\boldsymbol{z}_1)|\leq 4M(1-\mathrm{e}^{-2\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p}). \tag{A.1}$$

When $c_1 > c_2$, we can obtain the same results. So it follows from $|f| \leq 1$ that $I_2 \leq 4M(1-\mathrm{e}^{-2\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p})$. As for I_1, since

$$\begin{aligned}&\left\|\boldsymbol{\omega}+\boldsymbol{A}\boldsymbol{x}_1+\boldsymbol{B}(m,n)^\top-(\boldsymbol{\omega}+\boldsymbol{A}\boldsymbol{z}_1+\boldsymbol{B}(m,n)^\top)\right\|_p\\&=\|\boldsymbol{A}(\boldsymbol{x}_1-\boldsymbol{z}_1)\|_p\leq\|\boldsymbol{A}\|_p\,\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p\leq\eta,\end{aligned}$$

so $I_1 \leq \varepsilon'$. Therefore, we have

$$|\mathbb{P}_{\boldsymbol{x}_1}f-\mathbb{P}_{\boldsymbol{z}_1}f|\leq\varepsilon'+4M(1-\mathrm{e}^{-2\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p}). \tag{A.2}$$

For the case that $k = 2$, it follows from

$$\mathbb{E}\{f(\boldsymbol{\lambda}_2)|\boldsymbol{\lambda}_0=\boldsymbol{x}\}=\mathbb{E}\{\mathbb{E}[f(\boldsymbol{\lambda}_2)|\boldsymbol{\lambda}_1]|\boldsymbol{\lambda}_0=\boldsymbol{x}\},$$

then

$$\begin{aligned}|\mathbb{P}^2_{\boldsymbol{x}_1}f-\mathbb{P}^2_{\boldsymbol{z}_1}f|&=\left|\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}[p(m,n|\boldsymbol{x}_1)\mathbb{P}_{\boldsymbol{x}_2}f-p(m,n|\boldsymbol{z}_1)\mathbb{P}_{\boldsymbol{z}_2}f]\right|\\&\leq\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}p(m,n|\boldsymbol{x}_1)|\mathbb{P}_{\boldsymbol{x}_2}f-\mathbb{P}_{\boldsymbol{z}_2}f|+\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}|p(m,n|\boldsymbol{x}_1)-p(m,n|\boldsymbol{z}_1)||\mathbb{P}_{\boldsymbol{z}_2}f|,\end{aligned}$$

where $\boldsymbol{x}_2=\boldsymbol{\omega}+\boldsymbol{A}\boldsymbol{x}_1+\boldsymbol{B}(m,n)^\top$ and $\boldsymbol{z}_2=\boldsymbol{\omega}+\boldsymbol{A}\boldsymbol{z}_1+\boldsymbol{B}(m,n)^\top$. Since $\|\boldsymbol{x}_2-\boldsymbol{z}_2\|_p=\|\boldsymbol{A}(\boldsymbol{x}_1-\boldsymbol{z}_1)\|_p\leq\|\boldsymbol{A}\|_p\,\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p\leq\eta$, so it follows from (A.1) and (A.2) that

$$\begin{aligned}|\mathbb{P}^2_{\boldsymbol{x}_1}f-\mathbb{P}^2_{\boldsymbol{z}_1}f|&\leq\varepsilon'+4M(1-\mathrm{e}^{-2\|\boldsymbol{x}_2-\boldsymbol{z}_2\|_p})+4M(1-\mathrm{e}^{-2\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p})\\&\leq\varepsilon'+4M(1-\mathrm{e}^{-2\|\boldsymbol{A}\|_p\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p})+4M(1-\mathrm{e}^{-2\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p}).\end{aligned}$$

Hence by induction, we have for any $k \geq 1$ that

$$\begin{aligned}|\mathbb{P}^k_{\boldsymbol{x}_1}f-\mathbb{P}^k_{\boldsymbol{z}_1}f|&\leq\varepsilon'+4M\sum_{s=0}^{k-1}(1-\mathrm{e}^{-2\|\boldsymbol{A}\|_p^s\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p})\\&\leq\varepsilon'+8M\sum_{s=0}^{\infty}\|\boldsymbol{A}\|_p^s\,\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p\leq\varepsilon'+\frac{8M\eta}{1-\|\boldsymbol{A}\|_p}\leq\varepsilon,\end{aligned}$$

which proves that $\{\boldsymbol{\lambda}_t\}$ is an e-chain. Therefore there exists a unique stationary distribution to $\{\boldsymbol{\lambda}_t\}$.

As for (b), it holds similar arguments to the proof of in Liu (2012, Proposition 4.2.1). □

Proof of Theorem 2 The proof follows the technique from Song et al. (2005, Theorem 1) and we just rewrite them in component form. Note that we only prove the consistency of $\hat{\boldsymbol{\theta}}_n^2$, because the estimators $\hat{\boldsymbol{\theta}}_n^k$ for $k > 2$ can be derived from it in the same manner. Suppose that $\hat{\boldsymbol{\theta}}_n^1 \xrightarrow{\mathbb{P}} \boldsymbol{\theta}^0$, then $\hat{\boldsymbol{\theta}}_n^1 = \boldsymbol{\theta}^0 + o_{\mathbb{P}}(1)$. Because $\hat{\boldsymbol{\theta}}_n^2$ satisfies equations

$$\dot{l}_{m(1)}(\hat{\boldsymbol{\theta}}_{1,n}^2) + \dot{l}_{c(1)}(\hat{\boldsymbol{\theta}}_{1,n}^1, \hat{\theta}_{2,n}^1) = \mathbf{0}, \text{ and } \dot{l}_{c(2)}(\hat{\boldsymbol{\theta}}_{1,n}^1, \hat{\theta}_{2,n}^2) = 0.$$

By Taylor's expansion, we have

$$\begin{aligned}
\mathbf{0} &= \dot{l}_{m(1)}(\hat{\boldsymbol{\theta}}_{1,n}^2) + \dot{l}_{c(1)}(\hat{\boldsymbol{\theta}}_{1,n}^1, \hat{\theta}_{2,n}^1) \\
&= \dot{l}_{m(1)}(\boldsymbol{\theta}_1^0) + \ddot{l}_{m(11)}(\boldsymbol{\theta}_1^0)(\hat{\boldsymbol{\theta}}_{1,n}^2 - \boldsymbol{\theta}_1^0) + o(\|\hat{\boldsymbol{\theta}}_{1,n}^2 - \boldsymbol{\theta}_1^0\|^2) + \dot{l}_{c(1)}(\hat{\boldsymbol{\theta}}_{1,n}^1, \hat{\theta}_{2,n}^1), \\
0 &= \dot{l}_{c(2)}(\hat{\boldsymbol{\theta}}_{1,n}^1, \hat{\theta}_{2,n}^2) = \dot{l}_{c(2)}(\hat{\boldsymbol{\theta}}_{1,n}^1, \theta_2^0) + \ddot{l}_{c(22)}(\hat{\boldsymbol{\theta}}_{1,n}^1, \theta_2^0)(\hat{\theta}_{2,n}^2 - \theta_2^0) + o(\|\hat{\theta}_{2,n}^2 - \theta_2^0\|^2).
\end{aligned}$$

Then we can obtain

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{1,n}^2 - \boldsymbol{\theta}_1^0 &= [-n^{-1}\ddot{l}_{m(11)}(\boldsymbol{\theta}_{1,n}^*)]^{-1} n^{-1} \\
&\quad \times [\dot{l}_{m(1)}(\boldsymbol{\theta}_1^0) + \dot{l}_{c(1)}(\hat{\boldsymbol{\theta}}_{1,n}^1, \hat{\theta}_{2,n}^1) + o(\|\hat{\boldsymbol{\theta}}_{1,n}^2 - \boldsymbol{\theta}_1^0\|^2)], \qquad \text{(A.3)} \\
\hat{\theta}_{2,n}^2 - \theta_2^0 &= [-n^{-1}\ddot{l}_{c(22)}(\hat{\boldsymbol{\theta}}_{1,n}^1, \theta_{2,n}^*)]^{-1} n^{-1} [\dot{l}_{c(2)}(\hat{\boldsymbol{\theta}}_{1,n}^1, \theta_2^0) + o(\|\hat{\theta}_{2,n}^2 - \theta_2^0\|^2)]. \qquad \text{(A.4)}
\end{aligned}$$

Under the regularity conditions, $[-n^{-1}\ddot{l}_{m(11)}(\boldsymbol{\theta}_{1,n}^*)]$ and $[-n^{-1}\ddot{l}_{c(22)}(\hat{\boldsymbol{\theta}}_{1,n}^1, \theta_{2,n}^*)]$ are bounded and due to the consistency of $\hat{\boldsymbol{\theta}}_n^1$, it follows that

$$\begin{aligned}
&\lim_{n\to\infty} [n^{-1}\dot{l}_{m(1)}(\boldsymbol{\theta}_1^0) + n^{-1}\dot{l}_{c(1)}(\hat{\boldsymbol{\theta}}_{1,n}^1, \hat{\theta}_{2,n}^1) + n^{-1} o(\|\hat{\boldsymbol{\theta}}_{1,n}^2 - \boldsymbol{\theta}_1^0\|^2)] \\
&= \lim_{n\to\infty} n^{-1}\dot{l}(\boldsymbol{\theta}^0) = \mathbf{0}, \\
&\lim_{n\to\infty} [n^{-1}\dot{l}_{c(2)}(\hat{\boldsymbol{\theta}}_{1,n}^1, \theta_2^0) + n^{-1} o(\|\hat{\theta}_{2,n}^2 - \theta_2^0\|^2)] \\
&= \lim_{n\to\infty} n^{-1}\dot{l}_{c(2)}(\boldsymbol{\theta}_1^0, \theta_2^0) = 0.
\end{aligned}$$

It is easy to find that $\hat{\boldsymbol{\theta}}_n^2 \xrightarrow{\mathbb{P}} \boldsymbol{\theta}^0$ according to (A.3) and (A.4). □

Proof of Theorem 3 We employ the similar arguments of Song et al. (2005, Theorem 3). First, it is important to mention that under the regularity conditions, the following result is obvious:

$$n^{-1/2} \begin{pmatrix} \dot{l}_m(\boldsymbol{\theta}^0) \\ \dot{l}_c(\boldsymbol{\theta}^0) \end{pmatrix} \to N(0, \boldsymbol{\Omega}),$$

where

$$\boldsymbol{\Omega} = \lim_{n\to\infty} n^{-1} \begin{pmatrix} \mathbb{E}\dot{l}_m(\boldsymbol{\theta}^0)\dot{l}_m^\top(\boldsymbol{\theta}^0) & \mathbb{E}\dot{l}_m(\boldsymbol{\theta}^0)\dot{l}_c^\top(\boldsymbol{\theta}^0) \\ \mathbb{E}\dot{l}_c(\boldsymbol{\theta}^0)\dot{l}_m^\top(\boldsymbol{\theta}^0) & \mathbb{E}\dot{l}_c(\boldsymbol{\theta}^0)\dot{l}_c^\top(\boldsymbol{\theta}^0) \end{pmatrix}.$$

Then according to Theorem 2, the consistency of $\boldsymbol{\theta}_n^k$ holds and satisfies equations

$$\dot{l}_{m(1)}(\hat{\boldsymbol{\theta}}_{1,n}^k) + \dot{l}_{c(1)}(\hat{\boldsymbol{\theta}}_{1,n}^{k-1}, \hat{\theta}_{2,n}^{k-1}) = \mathbf{0}, \text{ and } \dot{l}_{c(2)}(\hat{\boldsymbol{\theta}}_{1,n}^{k-1}, \hat{\theta}_{2,n}^k) = 0.$$

By Taylor's expansion without the remainder terms, we have at Step k

$$\begin{aligned} &\dot{l}_{m(1)}(\boldsymbol{\theta}_1^0) + \ddot{l}_{m(11)}(\boldsymbol{\theta}_1^0)(\hat{\boldsymbol{\theta}}_{1,n}^k - \boldsymbol{\theta}_1^0) \\ &\quad + \dot{l}_{c(1)}(\boldsymbol{\theta}^0) + \ddot{l}_{c(11)}(\boldsymbol{\theta}^0)(\hat{\boldsymbol{\theta}}_{1,n}^{k-1} - \boldsymbol{\theta}_1^0) + \ddot{l}_{c(12)}(\boldsymbol{\theta}^0)(\hat{\theta}_{2,n}^{k-1} - \boldsymbol{\theta}_2^0) = \mathbf{0}, \\ &\dot{l}_{c(2)}(\boldsymbol{\theta}^0) + \ddot{l}_{c(21)}(\boldsymbol{\theta}^0)(\hat{\boldsymbol{\theta}}_{1,n}^{k-1} - \boldsymbol{\theta}_1^0) + \ddot{l}_{c(22)}(\boldsymbol{\theta}^0)(\hat{\theta}_{2,n}^k - \boldsymbol{\theta}_2^0) = 0. \end{aligned}$$

Rewriting these in a matrix form, we can obtain

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^k - \boldsymbol{\theta}^0) = \boldsymbol{D}_n^{-1}\boldsymbol{T}_n\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{k-1} - \boldsymbol{\theta}^0) + \boldsymbol{D}_n^{-1}[n^{-1/2}\dot{l}(\boldsymbol{\theta}^0)]. \tag{A.5}$$

Hence by recursion, (A.5) turns to be

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^k - \boldsymbol{\theta}^0) = (\boldsymbol{D}_n^{-1}\boldsymbol{T}_n)^{k-1}\sqrt{n}(\hat{\boldsymbol{\theta}}_n^1 - \boldsymbol{\theta}^0) + \sum_{s=0}^{k-2}(\boldsymbol{D}_n^{-1}\boldsymbol{T}_n)^s\boldsymbol{D}_n^{-1}[n^{-1/2}\dot{l}(\boldsymbol{\theta}^0)].$$

Because $\hat{\boldsymbol{\theta}}_{1,n}^1$ is used to define $\hat{\theta}_{2,n}^1$, a Taylor's expansion at Step 1 leads to

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\theta}}_n^1 - \boldsymbol{\theta}^0) &= \begin{pmatrix} -n^{-1}\ddot{l}_{m(11)} & 0 \\ -n^{-1}\ddot{l}_{c(21)} & -n^{-1}\ddot{l}_{c(22)} \end{pmatrix}^{-1} \begin{pmatrix} -n^{-1/2}\dot{l}_{m(1)} \\ -n^{-1/2}\dot{l}_{c(2)} \end{pmatrix} \\ &= \boldsymbol{D}_n^{-1}\begin{pmatrix} -n^{-1/2}\dot{l}_{m(1)} \\ -n^{-1/2}\dot{l}_{c(2)} \end{pmatrix} + \boldsymbol{V}_n\begin{pmatrix} -n^{-1/2}\dot{l}_{m(1)} \\ -n^{-1/2}\dot{l}_{c(2)} \end{pmatrix}. \end{aligned}$$

Thus,

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\theta}}_n^k - \boldsymbol{\theta}^0) &= \sum_{s=0}^{k-1}(\boldsymbol{D}_n^{-1}\boldsymbol{T}_n)^s\boldsymbol{D}_n^{-1}\begin{pmatrix} -n^{-1/2}\dot{l}_{m(1)} \\ -n^{-1/2}\dot{l}_{c(2)} \end{pmatrix} \\ &\quad + \sum_{s=0}^{k-2}(\boldsymbol{D}_n^{-1}\boldsymbol{T}_n)^s\boldsymbol{D}_n^{-1}\begin{pmatrix} -n^{-1/2}\dot{l}_{c(1)} \\ 0 \end{pmatrix} \\ &\quad + (\boldsymbol{D}_n^{-1}\boldsymbol{T}_n)^{k-1}\boldsymbol{V}_n\begin{pmatrix} -n^{-1/2}\dot{l}_{m(1)} \\ -n^{-1/2}\dot{l}_{c(2)} \end{pmatrix} \\ &= [\boldsymbol{I} - (\boldsymbol{D}_n^{-1}\boldsymbol{T}_n)^k][\boldsymbol{I} - \boldsymbol{D}_n^{-1}\boldsymbol{T}_n]^{-1}\boldsymbol{D}_n^{-1}\begin{pmatrix} -n^{-1/2}\dot{l}_{m(1)} \\ -n^{-1/2}\dot{l}_{c(2)} \end{pmatrix} \end{aligned}$$

$$+ [\boldsymbol{I} - (\boldsymbol{D}_n^{-1}\boldsymbol{T}_n)^{k-1}][\boldsymbol{I} - \boldsymbol{D}_n^{-1}\boldsymbol{T}_n]^{-1}\boldsymbol{D}_n^{-1}\begin{pmatrix} -n^{-1/2}\dot{l}_{c(1)} \\ 0 \end{pmatrix}$$
$$+ (\boldsymbol{D}_n^{-1}\boldsymbol{T}_n)^{k-1}\boldsymbol{V}_n\begin{pmatrix} -n^{-1/2}\dot{l}_{m(1)} \\ -n^{-1/2}\dot{l}_{c(2)} \end{pmatrix}.$$

Note that $[\boldsymbol{I} - \boldsymbol{D}_n^{-1}\boldsymbol{T}_n]^{-1}\boldsymbol{D}_n^{-1} = (\boldsymbol{D}_n^{-1} - \boldsymbol{T}_n)^{-1} = [-n^{-1}\ddot{l}(\boldsymbol{\theta}^0)]^{-1}$, then it follows that

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\theta}}_n^k - \boldsymbol{\theta}^0) = &\{[\boldsymbol{I} - (\boldsymbol{D}_n^{-1}\boldsymbol{T}_n)^k][-n^{-1}\ddot{l}(\boldsymbol{\theta}^0)]^{-1} \\ &+ (\boldsymbol{D}_n^{-1}\boldsymbol{T}_n)^{k-1}\boldsymbol{V}_n\}\begin{pmatrix} -n^{-1/2}\dot{l}_{m(1)} \\ -n^{-1/2}\dot{l}_{c(2)} \end{pmatrix} \\ &+ [\boldsymbol{I} - (\boldsymbol{D}_n^{-1}\boldsymbol{T}_n)^{k-1}][-n^{-1}\ddot{l}(\boldsymbol{\theta}^0)]^{-1}\begin{pmatrix} -n^{-1/2}\dot{l}_{c(1)} \\ 0 \end{pmatrix} \\ &\to N(0, \boldsymbol{\Sigma}_k), \quad \text{as } n \to \infty,\end{aligned}$$

where $\boldsymbol{\Gamma}_k$ is defined in the statement of Theorem 3.

Furthermore, when Assumption 2 holds, i.e., the marginal function l_m satisfies the information dominance condition. Then we have $\boldsymbol{\Gamma}^k \to 0$ as $k \to \infty$, thus the asymptotic variance–covariance matrix becomes

$$\begin{aligned}\boldsymbol{\Sigma}_\infty &= \mathcal{I}^{-1}\begin{pmatrix} n^{-1}\mathbb{E}(\dot{l}_{m(1)} + \dot{l}_{c(1)})(\dot{l}_{m(1)} + \dot{l}_{c(1)})^\top & n^{-1}\mathbb{E}\dot{l}_{c(2)}(\dot{l}_{m(1)} + \dot{l}_{c(1)})^\top \\ n^{-1}\mathbb{E}(\dot{l}_{m(1)} + \dot{l}_{c(1)})\dot{l}_{c(2)}^\top & n^{-1}\dot{l}_{c(2)}\dot{l}_{c(2)}^\top \end{pmatrix}\mathcal{I}^{-1} \\ &= \mathcal{I}^{-1}\{\lim_{n\to\infty} n^{-1}\mathbb{E}\dot{l}(\boldsymbol{\theta}^0)\dot{l}(\boldsymbol{\theta}^0)^\top\}\mathcal{I}^{-1} = \mathcal{I}^{-1}.\end{aligned}$$

□

References

Ahmad, A., Francq, C. (2016). Poisson QMLE of count time series models. *Journal of Time Series Analysis*, *37*, 291–314.

Aitchison, J., Ho, C. (1989). The multivariate Poisson-log normal distribution. *Biometrika*, *75*, 621–629.

Berentsen, G. D., Bulla, J., Maruotti, A., Støve, B. (2018). Modelling corporate defaults: A Markov-switching Poisson log-linear autoregressive model. arXiv:1804.09252.

Borges, P., Bourguignon, M., Molinares, F. (2017). A generalised NGINAR(1) process with inflated-parameter geometric counting series. *Australian and New Zealand Journal of Statistics*, *59*, 137–150.

Brockwell, P. J., Davis, R. A. (1991). *Time series*: *Data analysis and theory* 2nd ed. New York: Springer.

Cui, Y., Zhu, F. (2018). A new bivariate integer-valued GARCH model allowing for negative cross-correlation. *Test*, *27*, 428–452.

Davis, R. A., Liu, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica*, *26*, 1673–1707.

Dominitz, J., Sherman, R. P. (2005). Some convergence theory for iterative estimation procedures with an application to semiparametric estimation. *Econometric Theory*, *21*, 838–863.

Douc, R., Fokianos, K., Moulines, E. (2017). Asymptotic properties of quasi-maximum likelihood estimators in observation-driven time series models. *Electronic Journal of Statistics*, *11*, 2707–2740.

Doukhan, P., Fokianos, K., Tjøstheim, D. (2012). On weak dependence conditions for Poisson autoregressions. *Statistics and Probability Letters*, *82*, 942–948.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, *81*, 709–721.
Ferland, R., Latour, A., Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, *27*, 923–942.
Fokianos, K. (2012). Count time series models. In T. S. Rao, S. S. Rao & C. R. Rao (Eds.), *Handbook of statistics*: *Time series analysis-methods and applications*, pp. 315–347. Amsterdam: Elsevier.
Fokianos, K. (2016). Statistical analysis of count time series models: A GLM perspective. In R. A. Davis, S. H. Holan, R. Lund & N. Ravishanker (Eds.), *Handbook of discrete-valued time series*, pp. 3–27. Boca Raton: Chapman and Hall/CRC.
Fokianos, K., Neumann, M. H. (2013). A goodness-of-fit test for Poisson count processes. *Electronic Journal of Statistics*, *7*, 793–819.
Fokianos, K., Rahbek, A., Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, *104*, 1430–1439.
Fokianos, K., Støve, B., Tjøstheim, D., Doukhan, P. (2019). Multivariate count autoregression. *Bernoulli*. arXiv:1704.02097. **(forthcoming)**.
Franke, J., Rao, T. S. (1995). *Multivariate first-order integer-valued autoregressions*. Department of Mathematics, UMIST: Technical Report.
Heinen, A., Rengifo, E. (2007). Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance*, *14*, 564–583.
Karlis, D., Meligkotsidou, L. (2007). Finite mixtures of multivariate Poisson distributions with application. *Journal of Statistical Planning and Inference*, *137*, 1942–1960.
Karlis, D., Pedeli, X. (2013). Flexible bivariate INAR(1) processes using copulas. *Communications in Statistics-Theory and Methods*, *42*, 723–740.
Kedem, B., Fokianos, K. (2002). *Regression models for time series analysis*. Hoboken, NJ: Wiley.
Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*,. https://doi.org/10.18637/jss.v070.i05.
Lakshminarayana, J., Pandit, S. N. N., Rao, K. S. (1999). On a bivariate Poisson distribution. *Communication in Statistics-Theory and Methods*, *28*, 267–276.
Latour, A. (1997). The multivariate GINAR(p) process. *Advances in Applied Probability*, *29*, 228–248.
Liu, H. (2012). *Some models for time series of counts*. Dissertations, Columbia University. https://doi.org/10.7916/D8Z325SW.
Liu, Y., Luger, R. (2009). Efficient estimation of copula-GARCH models. *Computational Statistics and Data Analysis*, *53*, 2284–2297.
Meyn, S. P., Tweedie, R. L. (2009). *Markov chains and stochastic stability*, 2nd ed., New York: Cambridge University Press.
Neumann, M. H. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli*, *17*, 1268–1284.
Pedeli, X., Karlis, D. (2011). A bivariate INAR(1) process with application. *Statistical Modelling*, *11*, 325–349.
Popović, P. M. (2016). A bivariate INAR(1) model with different thinning parameters. *Statistical Papers*, *57*, 517–538.
Scotto, M. G., Weiß, C. H., Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: A review. *Statistical Modelling*, *15*, 590–618.
Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *Journal of the Royal Statistical Society*: *Series C*, *54*, 127–142.
Song, P. X. K., Fan, Y., Kalbfleisch, J. D. (2005). Maximization by parts in likelihood inference. *Journal of the American Statistical Association*, *100*, 1145–1158.
Tjøstheim, D. (2012). Some recent theory for autoregressive count time series (with discussions). *Test*, *21*, 413–476.
Tjøstheim, D. (2016). Count time series with observation-driven autoregressive parameter dynamics. In R. A. Davis, S. H. Holan, R. Lund & N. Ravishanker (Eds.), *Handbook of discrete-valued time series*, pp. 77–100. Boca Raton: Chapman and Hall/CRC.
Wang, C., Liu, H., Yao, J. F., Davis, R. A., Li, W. K. (2014). Self-excited threshold Poisson autoregression. *Journal of American Statistical Association*, *109*, 776–787.
Weiß, C. H. (2008). Thinning operations for modeling time series of counts—A survey. *Advances in Statistical Analysis*, *92*, 319–341.

Wu, W., Shao, X. (2004). Limit theorems for iterated random functions. *Journal of Applied Probability*, *41*, 425–436.

Zhu, F. (2011). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis*, *32*, 54–67.

Zhu, F. (2012a). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *Journal of Mathematical Analysis and Applications*, *389*, 58–71.

Zhu, F. (2012b). Zero-inflated Poisson and negative binomial integer-valued GARCH models. *Journal of Statistical Planning and Inference*, *142*, 826–839.