TEST (2018) 27:428–452
https://doi.org/10.1007/s11749-017-0552-4

ORIGINAL PAPER

A new bivariate integer-valued GARCH model allowing for negative cross-correlation

Yan Cui[1] · **Fukang Zhu**[1]

Received: 19 January 2017 / Accepted: 5 September 2017 / Published online: 19 September 2017

Abstract Univariate integer-valued time series models, including integer-valued autoregressive (INAR) models and integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) models, have been well studied in the literature, but there is little progress in multivariate models. Although some multivariate INAR models were proposed, they do not provide enough flexibility in modeling count data, such as volatility of numbers of stock transactions. Then, a bivariate Poisson INGARCH model was suggested by Liu (Some models for time series of counts, Dissertations, Columbia University, 2012), but it can only deal with positive cross-correlation between two components. To remedy this defect, we propose a new bivariate Poisson INGARCH model, which is more flexible and allows for positive or negative cross-correlation. Stationarity and ergodicity of the new process are established. The maximum likelihood method is used to estimate the unknown parameters, and consistency and asymptotic normality for estimators are given. A simulation study is given to evaluate the estimators for parameters of interest. Real and artificial data examples are illustrated to demonstrate good performances of the proposed model relative to the existing model.

Keywords Bivariate · GARCH model · MLE · Negative cross-correlation · Poisson · Time series of counts

Mathematics Subject Classification 62M10 · 62F10 · 62F12

✉ Fukang Zhu
zfk8010@163.com

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

1 Introduction

With a surge in real-world applications including epidemiology, marketing, insurance and environmental science, there has been increasing interest in developing models for time series of counts. In many practical situations, one often encounters integer-valued time series. There has been a great number of attempts to deal with them, see Scotto et al. (2015) for a resent review on INAR models based on binomial thinning operators.

As an alternative, the INGARCH model proposed by Ferland et al. (2006) and Fokianos et al. (2009), which can deal with time series of counts, is also very popular. For a Poisson INGARCH, which is commonly used, the observations $\{X_t\}$ given the intensity process $\{\lambda_t\}$ follow a Poisson distribution and λ_t is a linear combination of its lagged values and lagged X_ts, i.e.,

$$\begin{cases} X_t|\mathcal{F}_{t-1} \sim \mathrm{P}(\lambda_t), \\ \lambda_t = \omega_0 + \alpha_0\lambda_{t-1} + \beta_0 X_{t-1}, \end{cases}$$

where $\omega_0 > 0$, $\alpha_0 \geq 0$, $\beta_0 \geq 0$, and $\mathcal{F}_{t-1}$ is the σ-field generated by $\{X_{t-1}, X_{t-2}, \ldots\}$. Zhu (2011), Gonçalves et al. (2016) and Davis and Liu (2016) generalized the Poisson distribution assumption to the negative binomial distribution, infinitely divisible distributions and exponential family distributions, respectively. Fokianos and Tjøstheim (2011, 2012) generalized the linear assumption to log-linear and nonlinear cases. Later Gonçalves et al. (2015) considered zero-inflated INGARCH models with general compound Poisson deviates. Neumann (2011) and Doukhan et al. (2012) established ergodicity of the model. Christou and Fokianos (2014) and Ahmad and Francq (2016) estimated parameters using the Poisson quasi maximum likelihood method. Fokianos (2012, 2016) and Tjøstheim (2012, 2016) gave excellent summaries about recent progress in this field.

In some situations in finance, criminology and accidents analysis, the data are observed across time leading to multivariate time series data. This great variety of application areas is illustrated with data examples by many authors. For example, Heinen and Rengifo (2007) worked with the five most important US department stores stocks traded on the New York Stock Exchange (NYSE) from January 2nd 1999 to December 30th 1999, while Pedeli and Karlis (2011) considered the daily number of daytime and nighttime road accidents in Schiphol area (Netherlands) in the year 2001.

Referring to the existing literature, we found many studies about multivariate integer-valued time series models in recent years. Latour (1997) proposed a stationary and causal multivariate INAR process and discussed estimation of parameters as well as time series forecasting. Pedeli and Karlis (2011) introduced bivariate INAR(1) with Poisson and negative binomial innovations, while Pedeli and Karlis (2013a) generalized it to a full bivariate INAR(1) process and the method of conditional maximum likelihood was suggested for estimating its unknown parameters. To overcome the computational difficulties of the maximum likelihood approach, Pedeli and Karlis (2013b) suggested the method of composite likelihood. But the above models do not provide enough flexibility in modeling count data, such as volatility of numbers of stock transactions. Liu (2012) conducted bivariate Poisson INGARCH models and proved the stationarity and ergodicity under certain conditions. At the time near completion of this paper, we noticed that Lee et al. (2017) established large-sample properties for estimators of parameters and considered parameter change test for the model proposed by Liu (2012).

In this paper, we mostly focus on a new bivariate INGARCH model based on the bivariate Poisson distribution proposed by Lakshminarayana et al. (1999), which is capable of modeling the dependence between two time series of counts. The conventional model is developed by trivariate reduction with the drawback that it does not support negative cross-correlation values, refer to Liu (2012) for more details. However, our new model can break this limitation by modeling either positive or negative cross-correlation between the two components.

The rest organization of this paper is as follows. In Sect. 2, we first briefly introduce the used bivariate Poisson distribution and then present the stability properties of our new bivariate Poisson integer-valued GARCH model. The maximum likelihood estimators for unknown parameters and related theoretical properties are proposed in Sect. 3. Section 4 presents a simulation study. In Sect. 5, we apply the proposed model to real and artificial data examples to demonstrate excellent performances relative to the competing model. Section 6 concludes.

2 A new bivariate Poisson INGARCH model

2.1 The bivariate Poisson distribution

One of the best known methods for defining the bivariate Poisson distribution is the *Trivariate Reduction*. Consider random variables $X_k, k = 1, 2, 3$, which follow independent Poisson distributions with parameters $\lambda_1 - \phi, \lambda_2 - \phi, \phi$, respectively, and then the random variables $Y_1 = X_1 + X_3$ and $Y_2 = X_2 + X_3$ follow jointly a bivariate Poisson distribution $\mathrm{BP}^*(\lambda_1, \lambda_2, \phi)$ with probability mass function (pmf)

$$P(Y_1 = y_1, Y_2 = y_2) = \exp\{-(\lambda_1 + \lambda_2 - \phi)\}\frac{(\lambda_1 - \phi)^{y_1}(\lambda_2 - \phi)^{y_2}}{y_1!y_2!} \times \sum_{k=0}^{\min(y_1,y_2)} \binom{y_1}{k}\binom{y_2}{k}k!\left(\frac{\phi}{(\lambda_1 - \phi)(\lambda_2 - \phi)}\right)^k,$$

where $\phi \in [0, \min(\lambda_1, \lambda_2))$. Marginally each random variable follows a Poisson distribution with means $E(Y_1) = \lambda_1$ and $E(Y_2) = \lambda_2$, and covariance $\mathrm{Cov}(Y_1, Y_2) = \phi \geq 0$. The main drawback of this bivariate Poisson distribution is that it does not support negative correlation values.

Lakshminarayana et al. (1999) defined a bivariate Poisson distribution as a product of Poisson marginals with a multiplicative factor, whose pmf is given by

$$P(Y_1 = y_1, Y_2 = y_2) = \frac{\lambda_1^{y_1}\lambda_2^{y_2}}{y_1!y_2!}\exp\{-(\lambda_1+\lambda_2)\}[1+\delta(\mathrm{e}^{-y_1}-\mathrm{e}^{-c\lambda_1})(\mathrm{e}^{-y_2}-\mathrm{e}^{-c\lambda_2})], \tag{2.1}$$

where $c = 1 - \mathrm{e}^{-1}$. It is denoted by BP $(\lambda_1, \lambda_2, \delta)$. It is obvious to see that the marginal pmf of Y_1 and Y_2 are Poisson with parameters λ_1 and λ_2, respectively. The mean vector of the above distribution is $(\lambda_1, \lambda_2)^\top$ and covariance matrix is

$$\begin{pmatrix} \lambda_1 & \delta c^2 \lambda_1 \lambda_2 \mathrm{e}^{-c(\lambda_1+\lambda_2)} \\ \delta c^2 \lambda_1 \lambda_2 \mathrm{e}^{-c(\lambda_1+\lambda_2)} & \lambda_2 \end{pmatrix}.$$

Hence, the correlation coefficient turns out to be $\rho = \delta c^2 \sqrt{\lambda_1 \lambda_2} \mathrm{e}^{-c(\lambda_1+\lambda_2)}$, where δ should lie in the range

$$|\delta| \leq \frac{1}{(1-\mathrm{e}^{-c\lambda_1})(1-\mathrm{e}^{-c\lambda_2})}. \tag{2.2}$$

It is easy to verify that ρ should satisfy

$$|\rho| \leq \min\left\{ \frac{c^2\sqrt{\lambda_1\lambda_2}\mathrm{e}^{-c(\lambda_1+\lambda_2)}}{(1-\mathrm{e}^{-c\lambda_1})(1-\mathrm{e}^{-c\lambda_2})}, 1 \right\}.$$

Thus, the correlation coefficient can be positive, zero or negative depending on the value of the multiplicative factor parameter δ.

For ease of acquiring some visual information regarding the values of δ and ρ, we provide a few combinations of Poisson rates λ_1 and λ_2 in Table 1.

2.1.1 Algorithm for generating samples of the bivariate Poisson distribution

Since we know the joint pmf and marginal pmf of BP$(\lambda_1, \lambda_2, \delta)$, the key step is how to generate i.i.d. samples of $(Y_1, Y_2)^\top$. Note that $Y_1 \sim \mathrm{P}(\lambda_1)$ and the conditional pmf of $(Y_2|Y_1)$ is

$$\begin{aligned} P(y_2|y_1) &:= P(Y_2 = y_2 | Y_1 = y_1) \\ &= \frac{\lambda_2^{y_2}}{y_2!}\mathrm{e}^{-\lambda_2}[1 + \delta(\mathrm{e}^{-y_1} - \mathrm{e}^{-c\lambda_1})(\mathrm{e}^{-y_2} - \mathrm{e}^{-c\lambda_2})],\ \ y_2 = 0, 1, 2, ... \end{aligned}$$

Thus, we propose the conditional sampling and inverse transformation method to generate sample $(y_1, y_2)^\top$ of $(Y_1, Y_2)^\top$ as follows,

Step 1 Begin with the values of parameters $\lambda_1, \lambda_2, \delta$;
Step 2 Generate the observation $y_1 \sim \mathrm{P}(\lambda_1)$;
Step 3 Generate the random number $u \sim \mathrm{U}(0, 1)$;
Step 4 If $u < P(0|y_1)$, then let $y_2 = 0$;
Step 5 If there exists some k such that $P(k-1|y_1) \leq u < P(k|y_1)$, then let $y_2 = k$.

Table 1 Three cases of realizations regarding the values of δ and ρ

λ_1	λ_2	δ	ρ
1	2	±2.5	±0.2121
5	3	±1	±0.0098
0.5	0.5	±8	±0.8494

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. Liu (2012) defined an INGARCH(1,1) model based on $\mathrm{BP}^*(\lambda_1, \lambda_2, \phi)$ as follows

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

where $\mathcal{F}_t = \sigma\{\boldsymbol{\lambda}_1, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t\}$, $\boldsymbol{\omega} = (\omega_1, \omega_2)^\top \in \mathbb{R}_+^2$, $\boldsymbol{A}, \boldsymbol{B}$ are both 2×2 matrices with nonnegative entries and $\mathrm{Cov}(Y_{t,1}, Y_{t,2}|\mathcal{F}_{t-1}) = \phi \geq 0$ for modeling dependence between $Y_{t,1}$ and $Y_{t,2}$. Similar to Liu's INGARCH model, we propose a new bivariate INGARCH model of order (1,1) as follows

$$\begin{aligned}\boldsymbol{Y}_t|\mathcal{F}_{t-1} \sim \mathrm{BP}(\lambda_{t,1}, \lambda_{t,2}, \delta),\ \boldsymbol{\lambda}_t &= (\lambda_{t,1}, \lambda_{t,2})^\top \\ &= \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1} + \boldsymbol{B}\boldsymbol{Y}_{t-1}.\end{aligned} \tag{2.3}$$

It is easy to see that the conditional covariance $\mathrm{Cov}(Y_{t,1}, Y_{t,2}|\mathcal{F}_{t-1}) = \delta c^2 \lambda_{t,1}\lambda_{t,2}$ $\mathrm{e}^{-c(\lambda_{t,1}+\lambda_{t,2})}$. Before stating the main result, we first introduce some relevant notations for a general matrix $\boldsymbol{J} \in \mathbb{C}^{m \times n}$. Define $\|\boldsymbol{J}\|_p$ as the p-induced norm of matrix $\boldsymbol{J}$ for $1 \leq p \leq \infty$, i.e., $\|\boldsymbol{J}\|_p = \max_{x \neq 0}\{\|\boldsymbol{J}\boldsymbol{x}\|_p / \|\boldsymbol{x}\|_p : \boldsymbol{x} \in \mathbb{C}^n\}$, where $\|\boldsymbol{x}\|_p$ is the p-norm of vector $\boldsymbol{x}$.

The study focus on the bivariate Markov chain $\{\boldsymbol{\lambda}_t, t \geq 1\}$. Note that by recursion, for any $s \geq 1$, we have

$$\boldsymbol{\lambda}_t = (\boldsymbol{I} + \boldsymbol{A} + \cdots + \boldsymbol{A}^{s-1})\,\boldsymbol{\omega} + \boldsymbol{A}^s\boldsymbol{\lambda}_{t-s} + \sum_{k=0}^{s-1} \boldsymbol{A}^k \boldsymbol{B}\boldsymbol{Y}_{t-k-1}, \tag{2.4}$$

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

$$\boldsymbol{\lambda_t} = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{\omega} + \sum_{k=0}^{\infty} \boldsymbol{A}^k \boldsymbol{B}\boldsymbol{Y}_{t-k-1}. \tag{2.5}$$

Hence, under the condition $\rho(\boldsymbol{A}) < 1$, we have $\boldsymbol{\lambda_t} \geq (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{\omega}$ 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 $f_{\boldsymbol{u}}(\boldsymbol{\lambda})$ according to the pmf (2.1) is defined as $f_{\boldsymbol{u}}(\boldsymbol{\lambda}) = \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{\lambda} + \boldsymbol{B}F_{\boldsymbol{\lambda}}^{-1}(\boldsymbol{u})$, where $\boldsymbol{u} = (u_1, u_2)^\top \in [0, 1]^2$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$, $F_{\boldsymbol{\lambda}}^{-1}(\boldsymbol{u}) = (F_{\lambda_1}^{-1}(u_1), F_{\lambda_2}^{-1}(u_2))^\top \in \mathbb{N}_0^2$, and $F^{-1}(u) = \inf\{t \geq 0 : F(t) \geq u\}$. Hence, it can be seen that for all t, $\boldsymbol{\lambda}_t = f_{\boldsymbol{U}_t}(\boldsymbol{\lambda}_{t-1})$, where $\{\boldsymbol{U}_t, t \geq 1\}$ follows independent uniform distribution on $[0, 1]^2$.

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

Theorem 1 *Suppose* $\{\boldsymbol{Y}_t, t \geq 1\}$ *follow* (2.3), $\boldsymbol{\omega}$, $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *have nonnegative entries.*

(a) *If* $\rho(\boldsymbol{A}+\boldsymbol{B}) < 1$, *then there exists at least one stationary distribution to* $\{\boldsymbol{\lambda}_t\}$. *In addition, 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.

Proof for the above and next theorems or lemmas are relegated to Appendix.

One particular consequence of the above theorem is that the unconditional moment of the process $\{\boldsymbol{Y}_t\}$ can be calculated by

$$\begin{aligned} E(\boldsymbol{Y}_t) &= E(E(\boldsymbol{Y}_t|\mathcal{F}_{t-1})) = E(\boldsymbol{\lambda}_t) = \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1} + \boldsymbol{B}\boldsymbol{Y}_{t-1} \\ &= \boldsymbol{\omega} + (\boldsymbol{A}+\boldsymbol{B})E(\boldsymbol{Y}_{t-1}), \end{aligned}$$

thus we have $E(\boldsymbol{Y}_t) = E(\boldsymbol{\lambda}_t) = (\boldsymbol{I}-\boldsymbol{A}-\boldsymbol{B})^{-1}\boldsymbol{\omega}$.

3 Estimation

Let $\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_n$ be observations from model (2.3) with the assumption of $\boldsymbol{A}$ and $\boldsymbol{B}$ have nonnegative entries. Then, the parameter vector turns to be $\boldsymbol{\theta} = (\omega_1, \alpha_{11}, \alpha_{12}, \beta_{11}, \beta_{12}, \omega_2, \alpha_{21}, \alpha_{22}, \beta_{21}, \beta_{22}, \delta)^\top = (\theta_1, \theta_2, ..., \theta_{11})^\top$, and write the true value of the parameter as $\boldsymbol{\theta}^0 = (\omega_1^0, \alpha_{11}^0, \alpha_{12}^0, \beta_{11}^0, \beta_{12}^0, \omega_2^0, \alpha_{21}^0, \alpha_{22}^0, \beta_{21}^0, \beta_{22}^0, \delta^0)^\top$. The likelihood function conditional on $\boldsymbol{\lambda}_1 = (\lambda_{1,1}, \lambda_{1,2})^\top$ is therefore given by

$$L(\boldsymbol{\theta}|\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_n, \boldsymbol{\lambda}_1) = \prod_{t=2}^{n} \frac{\lambda_{t,1}^{Y_{t,1}}\lambda_{t,2}^{Y_{t,2}}}{Y_{t,1}!Y_{t,2}!}\exp\{-(\lambda_{t,1}+\lambda_{t,2})\}\varphi_t,$$

where $\varphi_t = \varphi_t(\boldsymbol{\theta}) = 1 + \delta(\mathrm{e}^{-Y_{t,1}} - \mathrm{e}^{-c\lambda_{t,1}})(\mathrm{e}^{-Y_{t,2}} - \mathrm{e}^{-c\lambda_{t,2}})$. Here, we use the bivariate Poisson assumption, $\boldsymbol{\lambda}_t(\boldsymbol{\theta}) = \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1}(\boldsymbol{\theta}) + \boldsymbol{B}\boldsymbol{Y}_{t-1}$ by model (2.3). Thus, the log-likelihood function is given, up to a constant free of $\boldsymbol{\theta}$, by

$$l(\boldsymbol{\theta}) = \sum_{t=2}^{n}[Y_{t,1}\ln\lambda_{t,1}(\boldsymbol{\theta}) + Y_{t,2}\ln\lambda_{t,2}(\boldsymbol{\theta}) - \lambda_{t,1}(\boldsymbol{\theta}) - \lambda_{t,2}(\boldsymbol{\theta}) + \ln\varphi_t(\boldsymbol{\theta})]. \quad (3.1)$$

The score function is defined by

$$S_n(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=2}^{n}\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \quad (3.2)$$

and the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ is a solution to the equation $S_n(\boldsymbol{\theta}) = 0$. The Hessian matrix is given by

$$H_n(\boldsymbol{\theta}) = -\sum_{t=2}^{n} \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}. \tag{3.3}$$

Detailed expressions for partial derivatives in (3.2) and (3.3) are given in Appendix. To study the asymptotic properties of the maximum likelihood estimator (MLE) $\widehat{\boldsymbol{\theta}}$, we introduce lower and upper values of each component of $\boldsymbol{\theta}^0$, $0 < \omega_L \leq \omega_i^0 \leq \omega_U$, $0 < \alpha_L \leq \alpha_{ij}^0 \leq \alpha_U$, and $0 < \beta_L \leq \beta_{ij}^0 \leq \beta_U$, $i, j = 1, 2$. To formulate the results, we make the following assumptions.

Assumption 1 A fixed compact neighborhood of $\boldsymbol{\theta}^0$ is defined as

$$\begin{aligned} O(\boldsymbol{\theta}^0) = \{\boldsymbol{\theta} | 0 < \omega_L \leq \omega_i \leq \omega_U,\ 0 < \alpha_L \leq \alpha_{ij} \leq \alpha_U,\ 0 < \beta_L \leq \beta_{ij} \\ \leq \beta_U,\ and\ |\delta| \leq \delta_U,\ i, j = 1, 2\}. \end{aligned} \tag{3.4}$$

Assumption 2 $0 < \varphi_L \leq |\varphi_t| \leq \varphi_U$.

Assumption 3 There exists a $p \in [1, \infty]$ such that $\|\boldsymbol{A}\|_p + 2^{(1-1/p)} \|\boldsymbol{B}\|_p < 1$.

We first introduce some lemmas to establish the large-sample properties of the estimator $\widehat{\boldsymbol{\theta}}$.

Lemma 1 *The score function defined by (3.2) and evaluated at the true value $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ satisfies*

$$\frac{1}{\sqrt{n}} S_n(\boldsymbol{\theta}) \xrightarrow{d} \mathrm{N}(0, G),$$

where the matrix G is defined as $G(\boldsymbol{\theta}) = E\left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right)$. A consistent estimator of G is given by $G_t(\widehat{\boldsymbol{\theta}})$, where $G_t(\boldsymbol{\theta}) = \mathrm{Var}\left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{t-1}\right)$.

Lemma 2 *The Hessian matrix defined by (3.3) and evaluated at the true value $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ satisfies*

$$\frac{1}{n} H_n(\boldsymbol{\theta}) \xrightarrow{P} G.$$

Lemma 3 *With the neighborhood $O(\boldsymbol{\theta}^0)$ defined in (3.4), it holds under Assumptions 1–3 that*

$$\max_{i,j,k=1,\ldots,11} \sup_{\boldsymbol{\theta} \in O(\boldsymbol{\theta}^0)} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| \leq M_n := \frac{1}{n} \sum_{t=1}^{n} m_t.$$

In addition,

$$\begin{aligned} &m_t = C g(Y_{t,i}) h(\mu_{ts}),\ i = 1, 2, \\ &\mu_{ts} = \beta_U \sum_{j=1}^{t-s} k_{j,s} \alpha_U^{j-1} Y_{t-s-j,i},\ s = 0, 1, 2, 3, \\ &k_{j,0} = 0,\ k_{j,1} = j,\ k_{j,2} = j(j+1)\ and\ k_{j,3} = j(j+1)(j+2), \end{aligned}$$

where $g(\cdot)$ and $h(\cdot)$ are functions of $Y_{t,i}$ and μ_{ts}, respectively. $0 \leq M_n \rightarrow M$, *M and C are both constants.*

Then, the following theorem regarding the properties of the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ holds true.

Theorem 2 *Consider model (2.3) and suppose that at the true value $\boldsymbol{\theta}^0$, Assumptions 1–3 hold. Then, there exists a fixed neighborhood $O(\boldsymbol{\theta}^0)$ of $\boldsymbol{\theta}^0$ such that the log-likelihood function (3.1) has a unique maximum point $\widehat{\boldsymbol{\theta}}$ with probability tending to one. Furthermore, $\widehat{\boldsymbol{\theta}}$ is consistent and asymptotically normal, i.e., $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{d}$* $\mathrm{N}(0, G^{-1})$, *where G is defined in Lemma 1.*

As for forecasting, we employ the method of conditional expectation to predict Poisson rates. By using recursion formula (2.4), a natural one-step ahead forecast of $\boldsymbol{\lambda}_t$ is

$$\hat{\boldsymbol{\lambda}}_{t+1} = E(\boldsymbol{\lambda}_{t+1}|\mathcal{F}_t) = \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{\lambda}_t + \boldsymbol{B}\boldsymbol{Y}_t = \sum_{k=0}^{t-1}(\boldsymbol{A}^k\boldsymbol{\omega} + \boldsymbol{A}^k\boldsymbol{B}\boldsymbol{Y}_{t-k}) + \boldsymbol{A}^t\boldsymbol{\lambda}_1$$

with parameter matrices being replaced by their corresponding estimates. More generally, a h-step ahead forecast of $\boldsymbol{\lambda}_t$, for some $h > 1$, can be obtained through the same recursive scheme, and we do not repeat it here.

4 Simulation

A simulation study is conducted to evaluate the performances of the estimators. For the estimation of the parameters, we use the method of randomly choosing from a uniform distribution to find out the initial values and the constrained nonlinear optimization function fmincon in Matlab to maximize the log-likelihood function.

In simulations, we choose sample of size $n = 200$ and 500 with $m = 200$ replications for each choice of parameters. The mean absolute deviation error (MADE) and the mean squared error (MSE) are calculated to evaluate the performances of the estimators according to the following formulas:

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

where $\hat{\vartheta}_j$ is the estimator of ϑ^0 in the jth replication. We consider the following configurations of the parameters:

- $\boldsymbol{A}$ and $\boldsymbol{B}$ are both diagonal, $\boldsymbol{\theta} = (\omega_1, \alpha_{11}, \beta_{11}, \omega_2, \alpha_{22}, \beta_{22}, \delta)^\top$:

(A1) $(1, 0.4, 0.2, 0.5, 0.3, 0.4, 0.5)^\top$; (A2) $(0.3, 0.2, 0.5, 0.5, 0.4, 0.3, 0.7)^\top$;
(A3) $(0.5, 0.1, 0.8, 0.5, 0.2, 0.7, -1)^\top$; (A4) $(0.5, 0.6, 0.1, 0.3, 0.8, 0.1, -0.5)^\top$.

- $\boldsymbol{A}$ is diagonal and $\boldsymbol{B}$ is non-diagonal, $\boldsymbol{\theta} = (\omega_1, \alpha_{11}, \beta_{11}, \beta_{12}, \omega_2, \alpha_{22}, \beta_{21}, \beta_{22}, \delta)^\top$:

$$\text{(B) } (0.3, 0.2, 0.5, 0.1, 0.5, 0.4, 0.3, 0.2, 0.7)^\top.$$

- $\boldsymbol{A}$ is non-diagonal and $\boldsymbol{B}$ is diagonal, $\boldsymbol{\theta} = (\omega_1, \alpha_{11}, \alpha_{12}, \beta_{11}, \omega_2, \alpha_{21}, \alpha_{22}, \beta_{22}, \delta)^\top$:

$$\text{(C) } (0.5, 0.2, 0.3, 0.2, 0.3, 0.4, 0.2, 0.3, -0.5)^\top.$$

- $\boldsymbol{A}$ and $\boldsymbol{B}$ are both non-diagonal, $\boldsymbol{\theta} = (\omega_1, \alpha_{11}, \alpha_{12}, \beta_{11}, \beta_{12}, \omega_2, \alpha_{21}, \alpha_{22}, \beta_{21}, \beta_{22}, \delta)^\top$:

$$\text{(D) } (0.5, 0.3, 0.2, 0.1, 0.2, 0.3, 0.1, 0.3, 0.3, 0.2, 0.4)^\top.$$

Mean and standard deviation (SD) of MLE estimators, MADE and MSE are summarized in Tables 2 and 3. We find that the estimation approach has reasonable estimators which generally show small values of SD, MADE and MSE. As the sample size increases, the estimators seem to converge to the true parameter values.

5 Illustrative examples

To show that the proposed model has flexibility in modeling cross-correlation, we use a real example to model positive cross-correlation, two artificial examples and another real example to model negative cross-correlation. In this section, two circumstances are considered in each example, i.e., $\boldsymbol{A}$ and $\boldsymbol{B}$ are both diagonal or non-diagonal.

5.1 Modeling positive cross-correlation

First, we apply the proposed model to the numbers of transactions in 5-min intervals during the first week of 2005 (January 3rd−7th) for two stocks traded at NYSE: P.H. Glatfelter Company (GLT) and Empire District Electric Company (EDE). This is a part of Trades and Quotes (TAQ) dataset, provided by the NYSE. There are 75 available observations between 9:45 a.m. and 4:00 p.m. per day, and the sample size is 375.

Note that the series of GLT and EDE have mean values (variances) equal to 6.147 (21.853) and 4.477 (13.849), respectively, which both show overdispersion. To give an idea about the data structure, Fig. 1 presents the original data series and the cross-correlation function (CCF) between two data series. The sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) of both GLT and EDE are provided in Fig. 2.

Table 4 summarizes the estimates of parameters, estimated log-likelihood, Akaike information criterion (AIC) and Bayesian information criterion (BIC) values. In addition, the cross-correlation between two series is 0.155, and the estimates of ϕ and δ are both positive. It turns out that the models fitted above are capturing the high volatility of the data. Using the consequence in Sect. 2.2, we can calculate that the estimated sample means of diagonal and non-diagonal cases are $(6.0407, 4.6964)^\top$

Table 2 Simulation results for the proposed model with diagonal coefficient matrices

Model	n	ω_1	α_{11}	β_{11}	ω_2	α_{22}	β_{22}	δ
A1	200							
	Mean	1.1331	0.3507	0.1934	0.5486	0.2815	0.3874	0.5773
	SD	0.4916	0.2282	0.0748	0.1695	0.1302	0.0765	0.8453
	MADE	0.4091	0.1944	0.0589	0.1306	0.1029	0.0606	0.7054
	MSE	0.2583	0.0543	0.0056	0.0310	0.0172	0.0060	0.8461
	500							
	Mean	1.0719	0.3755	0.1949	0.5238	0.2863	0.3954	0.4555
	SD	0.3598	0.1638	0.0487	0.1185	0.0968	0.0504	0.5393
	MADE	0.2892	0.1325	0.0379	0.0960	0.0783	0.0404	0.4404
	MSE	0.1340	0.0273	0.0024	0.0145	0.0095	0.0025	0.2973
A2	200							
	Mean	0.3284	0.1785	0.4840	0.5747	0.3557	0.2988	0.6535
	SD	0.0969	0.1313	0.0896	0.2301	0.1738	0.0724	0.6063
	MADE	0.0798	0.1079	0.0706	0.1837	0.1426	0.0579	0.4851
	MSE	0.0101	0.0176	0.0082	0.0583	0.0320	0.0052	0.3748
	500							
	Mean	0.3114	0.1934	0.4917	0.5334	0.3842	0.2976	0.7138
	SD	0.0621	0.0817	0.0535	0.1484	0.1115	0.0489	0.4253
	MADE	0.0476	0.0648	0.0426	0.1158	0.0891	0.0391	0.3364
	MSE	0.0040	0.0067	0.0029	0.0230	0.0126	0.0024	0.1810
A3	200							
	Mean	0.5510	0.1028	0.7808	0.5916	0.1850	0.6862	−0.9375
	SD	0.1649	0.0772	0.0745	0.2005	0.0907	0.0764	1.2599
	MADE	0.1273	0.0626	0.0614	0.1554	0.0745	0.0628	1.0340
	MSE	0.0297	0.0059	0.0059	0.0484	0.0084	0.0060	1.5833
	500							
	Mean	0.5194	0.1005	0.7943	0.5367	0.1940	0.6976	−1.0136
	SD	0.0966	0.0490	0.0457	0.1075	0.0526	0.0447	0.9408
	MADE	0.0732	0.0392	0.0370	0.0864	0.0414	0.0350	0.7935
	MSE	0.0097	0.0024	0.0021	0.0129	0.0028	0.0020	0.8809
A4	200							
	Mean	0.6671	0.5105	0.0906	0.3910	0.7807	0.0863	−0.6143
	SD	0.4655	0.2935	0.0613	0.3365	0.1561	0.0576	0.9019
	MADE	0.3607	0.2341	0.0514	0.1628	0.0875	0.0456	0.7336
	MSE	0.2435	0.0937	0.0038	0.1209	0.0246	0.0035	0.8224
	500							
	Mean	0.5935	0.5483	0.0953	0.3621	0.7887	0.0903	−0.5357
	SD	0.3412	0.2243	0.0401	0.2023	0.0786	0.0304	0.5205
	MADE	0.2453	0.1660	0.0327	0.1161	0.0515	0.0254	0.4085
	MSE	0.1246	0.0528	0.0016	0.0446	0.0063	0.0010	0.2709

Mean sample mean of estimates, *SD* standard deviation, *MADE* mean absolute deviation error, *MSE* mean squared error

Table 3 Simulation results for the proposed model with non-diagonal coefficient matrices

Model	n	ω_1	α_{11}	α_{12}	β_{11}	β_{12}	ω_2	α_{21}	α_{22}	β_{21}	β_{22}	δ
B	200											
	Mean	0.3415	0.1874		0.4825	0.1004	0.5347		0.3894	0.3102	0.1879	0.7700
	SD	0.1368	0.1185		0.0775	0.0527	0.1965		0.1433	0.0770	0.0763	0.7701
	MADE	0.1091	0.0970		0.0615	0.0426	0.1529		0.1152	0.0623	0.0634	0.6319
	MSE	0.0204	0.0141		0.0063	0.0028	0.0396		0.0205	0.0060	0.0059	0.5950
	500											
	Mean	0.3222	0.1969		0.4886	0.0984	0.5226		0.3915	0.3055	0.1972	0.7042
	SD	0.0899	0.0789		0.0449	0.0346	0.1309		0.0918	0.0496	0.0474	0.5187
	MADE	0.0716	0.0608		0.0356	0.0273	0.1025		0.0716	0.0388	0.0372	0.4276
	MSE	0.0085	0.0062		0.0021	0.0012	0.0176		0.0085	0.0025	0.0022	0.2677
C	200											
	Mean	0.4953	0.2234	0.2960	0.1849		0.3925	0.3471	0.2057		0.2943	−0.4458
	SD	0.3258	0.2184	0.1841	0.0676		0.3267	0.2236	0.1696		0.0683	0.6753
	MADE	0.2606	0.1776	0.1470	0.0543		0.2829	0.1901	0.1386		0.0551	0.5280
	MSE	0.1057	0.0480	0.0337	0.0048		0.1147	0.0525	0.0287		0.0047	0.4566
	500											
	Mean	0.5081	0.2044	0.2985	0.1895		0.3283	0.3994	0.1944		0.2964	−0.5262
	SD	0.2461	0.1620	0.1185	0.0466		0.2796	0.1654	0.1221		0.0438	0.4165
	MADE	0.2017	0.1354	0.0942	0.0370		0.2233	0.1332	0.0991		0.0342	0.3330
	MSE	0.0603	0.0261	0.0140	0.0023		0.0786	0.0272	0.0149		0.0019	0.1733
D	200											
	Mean	0.5601	0.2775	0.2274	0.0766	0.1953	0.3495	0.1527	0.2389	0.3023	0.1854	0.2377
	SD	0.2996	0.2431	0.2059	0.0583	0.0764	0.2754	0.1947	0.1828	0.0686	0.0749	0.7166
	MADE	0.2283	0.2131	0.1780	0.0511	0.0617	0.2098	0.1539	0.1583	0.0554	0.0608	0.6225
	MSE	0.0924	0.0595	0.0428	0.0039	0.0058	0.0765	0.0407	0.0368	0.0047	0.0057	0.5357

Table 3 continued

Model	n	ω_1	α_{11}	α_{12}	β_{11}	β_{12}	ω_2	α_{21}	α_{22}	β_{21}	β_{22}	δ
	500											
	Mean	0.5147	0.3107	0.1972	0.0923	0.1946	0.3121	0.1656	0.2292	0.3038	0.1974	0.3729
	SD	0.1894	0.2005	0.1755	0.0440	0.0473	0.1707	0.1820	0.1665	0.0442	0.0447	0.4899
	MADE	0.1466	0.1704	0.1512	0.0372	0.0371	0.1372	0.1526	0.1482	0.0354	0.0363	0.4083
	MSE	0.0355	0.0395	0.0309	0.0020	0.0023	0.0290	0.0374	0.0332	0.0020	0.0020	0.2396

Mean sample mean of estimates, *SD* standard deviation, *MADE* mean absolute deviation error, *MSE* mean squared error

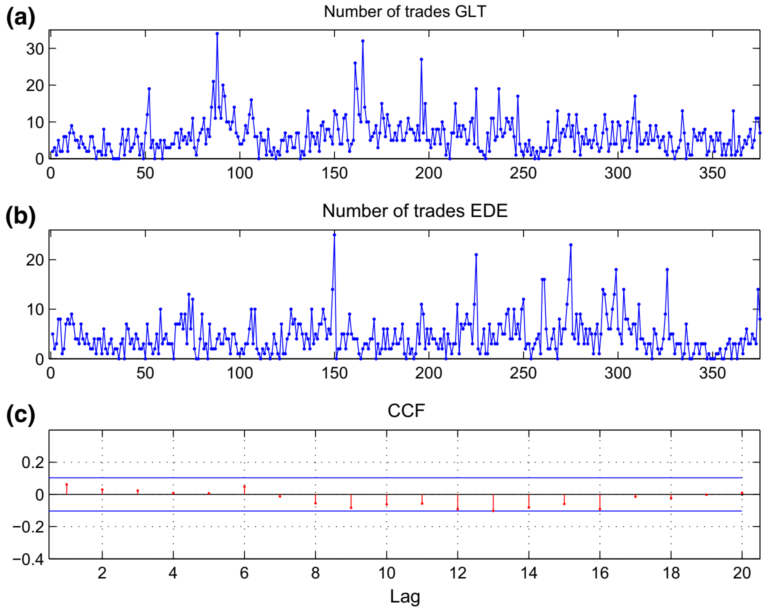

Fig. 1 **a** Number of trades GLT **b** Number of trades EDE **c** CCF of two data series

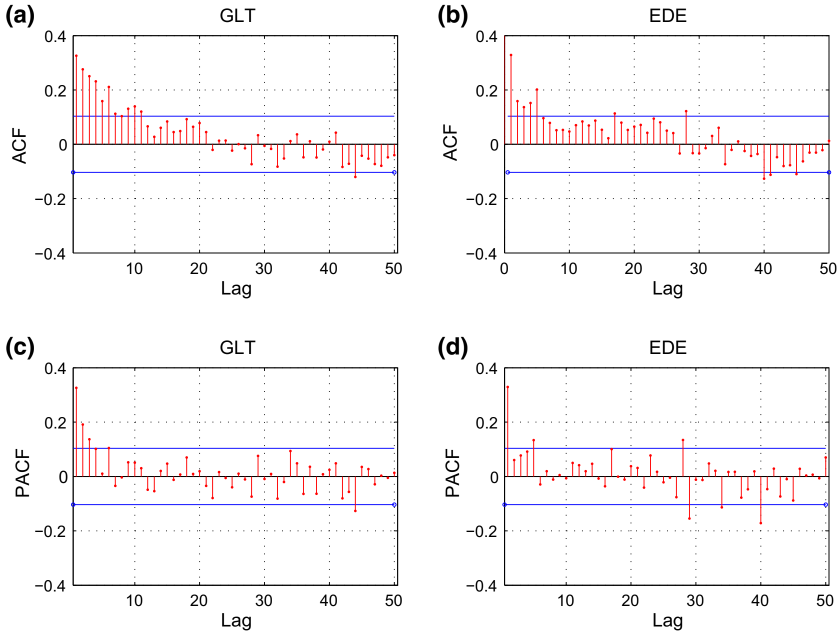

Fig. 2 **a** ACF of GLT series **b** ACF of EDE series **c** PACF of GLT series **d** PACF of EDE series

Table 4 Estimates for Liu's and proposed models for the transaction data

Parameter	Diagonal		Non-diagonal	
	Liu's	Proposed	Liu's	Proposed
ω_1	0.7989	0.7575	0.7354	0.7855
ω_2	0.8493	1.0224	0.8345	1.0201
α_{11}	0.6767	0.6702	0.6759	0.6695
α_{12}			0.0001	0.0001
α_{21}			0.0001	0.0001
α_{22}	0.5645	0.5413	0.5593	0.5306
β_{11}	0.2027	0.2044	0.2027	0.2042
β_{12}			0.0050	0.0033
β_{21}			0.0043	0.0057
β_{22}	0.2481	0.2410	0.2493	0.2432
δ		5.7010		5.7617
ϕ	0.5408		0.5384	
Log-Lik	−2137.631	−2135.097	−2137.600	−2135.086
AIC	4289.262	4284.194	4297.199	4292.172
BIC	4316.618	4311.551	4340.068	4335.041

and $(6.3451, 4.6724)^\top$, respectively. As seen from Table 4, the proposed model has larger log-likelihood function, and smaller AIC and BIC values, which seems to overperform Liu's model in each circumstance. In summary, our proposed model with diagonal matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is more suitable and exhibits better performances.

5.2 Modeling negative cross-correlation

Here, we will show how our proposed model can be used to model negative cross-correlation.

5.2.1 Two artificial data examples

To begin with, we choose two setups to generate Poisson INGARCH(1, 1) processes, i.e.,

$$X_t|\mathcal{F}_{t-1} \sim \mathrm{P}(\lambda_t), \quad \lambda_t = \omega + \alpha\lambda_{t-1} + \beta X_{t-1}$$

with parameter values:

$$\text{(E1)}\ (\omega, \alpha, \beta)^\top = (0.5, 0.05, 0.7)^\top; \quad \text{(E2)}\ (\omega, \alpha, \beta)^\top = (1, 0.35, 0.45)^\top.$$

Each setup has 1000 samples, let $Y_{t,1} = X_t$ and $Y_{t,2} = X_{t+500}$ for $t = 1, \ldots, 500$. By calculating, we know that the two data series have mean (variance) values equal to

Table 5 Estimates for Liu's and proposed models for artificial data E1

Parameter	Diagonal		Non-diagonal	
	Liu's	Proposed	Liu's	Proposed
ω_1	0.5588	0.5663	0.5531	0.5606
ω_2	0.5122	0.5167	0.5069	0.5117
α_{11}	0.0001	0.0001	0.0001	0.0001
α_{12}			0.0001	0.0001
α_{21}			0.0001	0.0001
α_{22}	0.0724	0.0724	0.0726	0.0725
β_{11}	0.7265	0.7241	0.7270	0.7246
β_{12}			0.0001	0.0001
β_{21}			0.0001	0.0001
β_{22}	0.7210	0.7190	0.7214	0.7193
δ		−0.7310		−0.7285
ϕ	0.0001		0.0001	
Log-Lik	−1597.856	−1596.881	−1598.018	−1597.037
AIC	3209.711	3207.762	3218.036	3216.073
BIC	3239.115	3237.165	3264.152	3262.189

Table 6 Estimates for Liu's and proposed models for artificial data E2

Parameter	Diagonal		Non-diagonal	
	Liu's	Proposed	Liu's	Proposed
ω_1	1.0059	1.0304	0.9664	1.0204
ω_2	1.0204	1.0225	1.0064	1.0119
α_{11}	0.3912	0.3918	0.3918	0.3918
α_{12}			0.0001	0.0001
α_{21}			0.0001	0.0001
α_{22}	0.3737	0.3720	0.3747	0.3721
β_{11}	0.4180	0.4147	0.4180	0.4147
β_{12}			0.0001	0.0001
β_{21}			0.0001	0.0001
β_{22}	0.4261	0.4276	0.4258	0.4275
δ		−11.7795		−11.7205
ϕ	0.0001		0.0001	
Log-Lik	−2188.743	−2185.975	−2188.768	−2186.043
AIC	4391.487	4385.950	4399.537	4394.086
BIC	4420.890	4415.354	4445.653	4440.201

(E1) 2.056 (4.799) and 2.424 (4.846); (E2) 5.212 (7.967) and 5.086 (7.337). It is easy to see both artificial data exhibit overdispersion. Furthermore, the cross-correlations between two time series are (E1) −0.219, (E2) −0.111, respectively.

In order to model the two data examples, we compare our proposed model with Liu's model. Parameters estimates, log-likelihood function, AIC and BIC values of each data set can be seen in Tables 5 and 6. Both models can capture overdispersion of each individual data series. Besides, we find the estimate of δ presents negative in our proposed model, while the estimate of ϕ is always 0.0001 which in fact is the predetermined lower bound in our algorithm. Furthermore, as Liu's model cannot capture negative cross-correlation, the estimate of ϕ naturally appears close to the lower bound. In summary, our proposed model cannot only model the negative cross-correlation with negative values of δ, but also yield smaller AIC and BIC values in two circumstances, which shows good performances compared with Liu's model.

5.2.2 Crime data of Pittsburgh

Here, we provide a real data example which shows negative cross-correlation. We mainly consider the data series of Robbery (ROB) and Criminal mischief (CMIS) in census tract 101 of Pittsburgh over the period 1990 through 1999. The data are available from the website http://www.forecastingprinciples.com/index.php/crimedata. And the time period covered is $T = 120$ observations.

Empirical mean values (variances) of ROB and CMIS are 2.083 (2.531) and 6.050 (14.435), respectively, which show overdispersion. Furthermore, by calculating the cross-correlation between the two series is -0.131, which is negative indeed. The original data series and CCF between two data series are plotted in Fig. 3.

Analogously to former example, we continue to compare our proposed model with Liu's model. Parameters estimates, log-likelihood function, AIC and BIC values

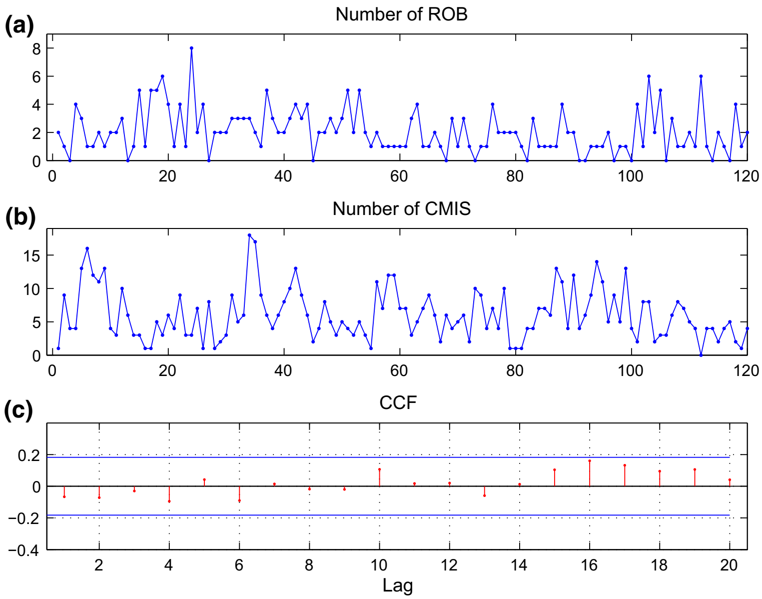

Fig. 3 **a** Number of ROB **b** Number of CMIS **c** CCF of two data series

Table 7 Estimates for Liu's and proposed models for crime data

Parameter	Diagonal		Non-diagonal	
	Liu's	Proposed	Liu's	Proposed
ω_1	0.4561	0.4930	0.4443	0.4820
ω_2	3.5322	3.5274	3.5324	3.5313
α_{11}	0.7051	0.6859	0.7044	0.6848
α_{12}			0.0001	0.0001
α_{21}			0.0001	0.0001
α_{22}	0.0929	0.0923	0.0920	0.0909
β_{11}	0.0804	0.0741	0.0811	0.0747
β_{12}			0.0001	0.0001
β_{21}			0.0001	0.0001
β_{22}	0.3288	0.3299	0.3290	0.3300
δ		−6.4350		−6.4363
ϕ	0.0001		0.0001	
Log-Lik	−537.391	−535.900	−537.448	−535.950
AIC	1088.782	1085.801	1096.897	1093.899
BIC	1107.874	1104.892	1126.502	1123.504

Table 8 MADE and MSE of predicted Poisson rates for crime data

	Diagonal				Non-diagonal			
	ROB		CMIS		ROB		CMIS	
	Liu's	Proposed	Liu's	Proposed	Liu's	Proposed	Liu's	Proposed
MADE	1.3482	1.3329	2.1436	2.1387	1.4155	1.3859	1.8836	1.8699
MSE	3.0598	3.0092	7.2945	7.2732	3.0515	2.9831	6.4678	6.4092

of each data set are given in Table 7. By using the proposed model, the estimated sample means of diagonal and non-diagonal cases are $(2.0542, 6.1049)^\top$ and $(2.0092, 6.1093)^\top$, which show quite little difference. As in Sect. 5.2.1, we also find the estimate of ϕ is the predetermined lower bound 0.0001. Similar to the previous analysis, our proposed model can model the negative cross-correlation with negative values of δ and generate smaller AIC and BIC values as well. So, the proposed model fits better and sounds more reasonable.

We also perform an out-of-sample forecasting exercise on two models above for comparison. First of all, we split the data in two parts. The first part has size $T_0(= 108)$, with the observations $\{\boldsymbol{Y}_t, t = 1, \ldots, T_0\}$ being used for initial estimation of the model, while the remaining observations $\{\boldsymbol{Y}_t, t = T_0 + 1, \ldots, T\}$ will be used for a forecasting. Then, we predict the crime data of the year 1999 by computing the one-step ahead forecast of $\boldsymbol{\lambda}_t$ using only information up to time t as discussed in Sect. 3 and repeat the above exercise for $t = T_0 + 1, \ldots, T$. Given the forecast path $\hat{\boldsymbol{\lambda}}_{t+1}$, we use MADE and MSE to evaluate the performance of two models with diagonal and non-diagonal cases. Table 8 shows MADE and MSE of predicted Poisson rates.

In general, we see that the our proposed model performs better for both diagonal and non-diagonal cases. In terms of MSE, the non-diagonal case with smaller MSE values provides a better forecasting than diagonal case.

6 Conclusion

The main focus in this paper is on bivariate time series models for count data. We propose a new bivariate INGARCH(1,1) model which can deal with either positive or negative cross-correlation. Besides, we give the conditions for the existence and ergodicity of the new model. The maximum likelihood estimators for the parameters are considered and asymptotic properties of the estimators are established. A simulation study shows that the estimation results are reliable as long as the sample size is reasonably large. We finally apply the proposed model to real examples and artificial data sets, which shows better performances compared with the existing model.

Acknowledgements We thank the Editor, Associate Editor and two referees for their valuable suggestions and comments, which greatly improved the presentation of the paper. This work is supported by National Natural Science Foundation of China (Nos. 11731015, 11371168), Science and Technology Developing Plan of Jilin Province (No. 20170101057JC), Science and Technology Program of Jilin Educational Department during the "13th Five-Year" Plan Period (No. 2016-399), and Cultivation Plan for Excellent Young Scholar Candidates of Jilin University.

Appendix

Proof of Theorem 1 First note that $\{\boldsymbol{\lambda}_t\}$ has at least one stationary distribution, refer to Liu (2012) for the details. From (2.5), it is easy to see that $(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{\omega}$ is a reachable state if $\boldsymbol{Y}_{t-1}=\boldsymbol{Y}_{t-2}=\cdots=\mathbf{0}$ for some $t\in\mathbb{N}$ large enough. Then, to verify a unique invariant probability measure exists, which is the main idea of Theorem 18.8.4 of Meyn and Tweedie (2009). We need to show that $\{\boldsymbol{\lambda}_t\}$ is an 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 $|P^k_{\boldsymbol{x}_1}f-P^k_{\boldsymbol{z}_1}f|<\varepsilon$, for $\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p<\eta$ and all $k\geq 1$, where $\boldsymbol{x}_1=(x_{1,1},x_{1,2})^\top$, $\boldsymbol{z}_1=(z_{1,1},z_{1,2})^\top$, $P^k_{\boldsymbol{x}_1}f=E\{f(\boldsymbol{\lambda}_k)|\boldsymbol{\lambda}_0=\boldsymbol{x}\}$. Without loss of generality, assume $|f|\leq 1$. Take ε' and η sufficiently small such that $\varepsilon'+16\eta/(1-\|\boldsymbol{A}\|_p)<\varepsilon$ and $|f(\boldsymbol{x}_1)-f(\boldsymbol{z}_1)|<\varepsilon'$ whenever $\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p<\eta$, for some $p\in[1,\infty]$. Here, denote $\varphi_1=1+\delta(\mathrm{e}^{-m}-\mathrm{e}^{-cx_{1,1}})(\mathrm{e}^{-n}-\mathrm{e}^{-cx_{1,2}})$, $\varphi_2=1+\delta(\mathrm{e}^{-m}-\mathrm{e}^{-cz_{1,1}})(\mathrm{e}^{-n}-\mathrm{e}^{-cz_{1,2}})$, and according to (2.2), it is clear to see that $|\varphi_{1\vee 2}|=\max(\varphi_1,\varphi_2)\leq 2$.

Then, for the case $k=1$,

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

$$\leq \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} p(m,n|\boldsymbol{x}_1)|f(\boldsymbol{\omega}+\boldsymbol{A}\boldsymbol{x}_1+\boldsymbol{B}(m,n)^\top)-f(\boldsymbol{\omega}+\boldsymbol{A}\boldsymbol{z}_1+\boldsymbol{B}(m,n)^\top)|$$
$$+\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}|p(m,n|\boldsymbol{x}_1)-p(m,n|\boldsymbol{z}_1)||f(\boldsymbol{\omega}+\boldsymbol{A}\boldsymbol{z}_1+\boldsymbol{B}(m,n)^\top)| = I_1+I_2,$$

where $p(m,n|\boldsymbol{x}_1)$ and $p(m,n|\boldsymbol{z}_1)$ are the pmfs of $\mathrm{BP}(x_{1,1},x_{1,2},\delta)$ and $\mathrm{BP}(z_{1,1},z_{1,2},\delta)$ given by (2.1).

We start to formulate the core part of I_2,

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

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

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}|p(m,n|\boldsymbol{x}_1)-p(m,n|\boldsymbol{z}_1)|\leq 8(1-\mathrm{e}^{-2\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p}). \tag{A.1}$$

So, it follows from $|f|\leq 1$ that $I_2\leq 8(1-\mathrm{e}^{-2\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p})$. As for I_1, since $\left\|\boldsymbol{\omega}+\boldsymbol{A}\boldsymbol{x}_1+\boldsymbol{B}(m,n)^\top-(\boldsymbol{\omega}+\boldsymbol{A}\boldsymbol{z}_1+\boldsymbol{B}(m,n)^\top)\right\|_p=\|\boldsymbol{A}(\boldsymbol{x}_1-\boldsymbol{z}_1)\|_p\leq\|\boldsymbol{A}\|_p\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p\leq\eta$, so $I_1\leq\varepsilon'$. Hence

$$|P_{\boldsymbol{x}_1}f-P_{\boldsymbol{z}_1}f|\leq\varepsilon'+8(1-\mathrm{e}^{-2\|\boldsymbol{x}_1-\boldsymbol{z}_1\|_p}). \tag{A.2}$$

For the case that $k = 2$, it follows from $E\{f(\boldsymbol{\lambda}_2)|\boldsymbol{\lambda}_0 = \boldsymbol{x}\} = E\{E[f(\boldsymbol{\lambda}_2)|\boldsymbol{\lambda}_1]|\boldsymbol{\lambda}_0 = \boldsymbol{x}\}$, then

$$
\begin{aligned}
|P_{\boldsymbol{x}_1}^2 f - P_{z_1}^2 f| &= \left|\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}[p(m,n|\boldsymbol{x}_1)P_{\boldsymbol{x}_2} f - p(m,n|z_1)P_{z_2} f]\right| \\
&\leq \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} p(m,n|\boldsymbol{x}_1)|P_{\boldsymbol{x}_2} f - P_{z_2} f| \\
&\quad + \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} |p(m,n|\boldsymbol{x}_1) - p(m,n|z_1)||P_{z_2} f|,
\end{aligned}
$$

where $\boldsymbol{x}_2 = \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{B}(m,n)^\top$ and $z_2 = \boldsymbol{\omega} + \boldsymbol{A}z_1 + \boldsymbol{B}(m,n)^\top$. Since $\|\boldsymbol{x}_2 - z_2\|_p = \|\boldsymbol{A}(\boldsymbol{x}_1 - z_1)\|_p \leq \|\boldsymbol{A}\|_p \|\boldsymbol{x}_1 - z_1\|_p \leq \eta$, so it follows from (A.1) and (A.2) that

$$
\begin{aligned}
|P_{\boldsymbol{x}_1}^2 f - P_{z_1}^2 f| &\leq \varepsilon' + 8(1 - \mathrm{e}^{-2\|\boldsymbol{x}_2 - z_2\|_p}) + 8(1 - \mathrm{e}^{-2\|\boldsymbol{x}_1 - z_1\|_p}) \\
&\leq \varepsilon' + 8(1 - \mathrm{e}^{-2\|\boldsymbol{A}\|_p\|\boldsymbol{x}_1 - z_1\|_p}) + 8(1 - \mathrm{e}^{-2\|\boldsymbol{x}_1 - z_1\|_p}).
\end{aligned}
$$

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

$$
\begin{aligned}
|P_{\boldsymbol{x}_1}^k f - P_{z_1}^k f| &\leq \varepsilon' + 8\sum_{s=0}^{k-1}(1 - \mathrm{e}^{-2\|\boldsymbol{A}\|_p^s\|\boldsymbol{x}_1 - z_1\|_p}) \\
&\leq \varepsilon' + 16\sum_{s=0}^{\infty}\|\boldsymbol{A}\|_p^s\|\boldsymbol{x}_1 - z_1\|_p \leq \varepsilon' + \frac{16\eta}{1 - \|\boldsymbol{A}\|_p} \leq \varepsilon,
\end{aligned}
$$

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 Proposition 4.2.1 in Liu (2012).

Partial derivatives in Sect. 3 Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, \delta)^\top$, where $\boldsymbol{\theta}_i = (\omega_i, \alpha_{ij}, \beta_{ij})^\top$, $i, j = 1, 2$. The score function is given by $S_n(\boldsymbol{\theta}) = \sum_{t=1}^n \partial l_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ with

$$
\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_1} = \left[\frac{Y_{t,1}}{\lambda_{t,1}} - 1 + \frac{\delta c\mathrm{e}^{-c\lambda_{t,1}}(\mathrm{e}^{-Y_{t,2}} - \mathrm{e}^{-c\lambda_{t,2}})}{\varphi_t}\right]\frac{\partial\lambda_{t,1}}{\partial\boldsymbol{\theta}_1}, \tag{A.3}
$$

$$
\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_2} = \left[\frac{Y_{t,2}}{\lambda_{t,2}} - 1 + \frac{\delta c\mathrm{e}^{-c\lambda_{t,2}}(\mathrm{e}^{-Y_{t,1}} - \mathrm{e}^{-c\lambda_{t,1}})}{\varphi_t}\right]\frac{\partial\lambda_{t,2}}{\partial\boldsymbol{\theta}_2}, \tag{A.4}
$$

$$
\frac{\partial l_t(\boldsymbol{\theta})}{\partial\delta} = \frac{(\mathrm{e}^{-Y_{t,1}} - \mathrm{e}^{-c\lambda_{t,1}})(\mathrm{e}^{-Y_{t,2}} - \mathrm{e}^{-c\lambda_{t,2}})}{\varphi_t}, \tag{A.5}
$$

$$
\begin{aligned}
&\frac{\partial\lambda_{t,i}}{\partial\omega_i} = 1 + \alpha_{ii}\frac{\partial\lambda_{t-1,i}}{\partial\omega_i}, \quad \frac{\partial\lambda_{t,i}}{\partial\alpha_{ij}} = \lambda_{t-1,j} + \alpha_{ii}\frac{\partial\lambda_{t-1,i}}{\partial\alpha_{ij}}, \\
&\frac{\partial\lambda_{t,i}}{\partial\beta_{ij}} = Y_{t-1,j} + \alpha_{ii}\frac{\partial\lambda_{t-1,i}}{\partial\beta_{ij}}, \quad i, j = 1, 2.
\end{aligned} \tag{A.6}
$$

The second derivatives of (3.3) are expressed as

$$
\begin{aligned}
\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^\top} =& -\left[\frac{Y_{t,1}}{\lambda_{t,1}^2} + \frac{\delta c^2 \mathrm{e}^{-c\lambda_{t,1}}(\mathrm{e}^{-Y_{t,2}} - \mathrm{e}^{-c\lambda_{t,2}})}{\varphi_t}\right.\\
&\left. + \frac{\delta^2 c^2 \mathrm{e}^{-2c\lambda_{t,1}}(\mathrm{e}^{-Y_{t,2}} - \mathrm{e}^{-c\lambda_{t,2}})^2}{\varphi_t^2}\right] \frac{\partial \lambda_{t,1}}{\partial \boldsymbol{\theta}_1} \frac{\partial \lambda_{t,1}}{\partial \boldsymbol{\theta}_1^\top}\\
&+ \left[\frac{Y_{t,1}}{\lambda_{t,1}} - 1 + \frac{\delta c \mathrm{e}^{-c\lambda_{t,1}}(\mathrm{e}^{-Y_{t,2}} - \mathrm{e}^{-c\lambda_{t,2}})}{\varphi_t}\right] \frac{\partial^2 \lambda_{t,1}}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^\top},\\
\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top} =& -\left[\frac{Y_{t,2}}{\lambda_{t,2}^2} + \frac{\delta c^2 \mathrm{e}^{-c\lambda_{t,2}}(\mathrm{e}^{-Y_{t,1}} - \mathrm{e}^{-c\lambda_{t,1}})}{\varphi_t}\right.\\
&\left. + \frac{\delta^2 c^2 \mathrm{e}^{-2c\lambda_{t,2}}(\mathrm{e}^{-Y_{t,1}} - \mathrm{e}^{-c\lambda_{t,1}})^2}{\varphi_t^2}\right] \frac{\partial \lambda_{t,2}}{\partial \boldsymbol{\theta}_2} \frac{\partial \lambda_{t,2}}{\partial \boldsymbol{\theta}_2^\top}\\
&+ \left[\frac{Y_{t,2}}{\lambda_{t,2}} - 1 + \frac{\delta c \mathrm{e}^{-c\lambda_{t,2}}(\mathrm{e}^{-Y_{t,1}} - \mathrm{e}^{-c\lambda_{t,1}})}{\varphi_t}\right] \frac{\partial^2 \lambda_{t,2}}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top},\\
\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \delta^2} =& -\frac{(\mathrm{e}^{-Y_{t,1}} - \mathrm{e}^{-c\lambda_{t,1}})^2(\mathrm{e}^{-Y_{t,2}} - \mathrm{e}^{-c\lambda_{t,2}})^2}{\varphi_t^2},\\
\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_2^\top} =& \left[\frac{\delta c^2 \mathrm{e}^{-c(\lambda_{t,1}+\lambda_{t,2})}}{\varphi_t}\right.\\
&\left. - \frac{\delta^2 c^2 \mathrm{e}^{-c(\lambda_{t,1}+\lambda_{t,2})}(\mathrm{e}^{-Y_{t,1}} - \mathrm{e}^{-c\lambda_{t,1}})(\mathrm{e}^{-Y_{t,2}} - \mathrm{e}^{-c\lambda_{t,2}})}{\varphi_t^2}\right] \frac{\partial \lambda_{t,1}}{\partial \boldsymbol{\theta}_1} \frac{\partial \lambda_{t,2}}{\partial \boldsymbol{\theta}_2^\top},\\
\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1 \partial \delta} =& \left[\frac{c \mathrm{e}^{-c\lambda_{t,1}}(\mathrm{e}^{-Y_{t,2}} - \mathrm{e}^{-c\lambda_{t,2}})}{\varphi_t}\right.\\
&\left. - \frac{\delta c \mathrm{e}^{-c\lambda_{t,1}}(\mathrm{e}^{-Y_{t,1}} - \mathrm{e}^{-c\lambda_{t,1}})(\mathrm{e}^{-Y_{t,2}} - \mathrm{e}^{-c\lambda_{t,2}})^2}{\varphi_t^2}\right] \frac{\partial \lambda_{t,1}}{\partial \boldsymbol{\theta}_1},\\
\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2 \partial \delta} =& \left[\frac{c \mathrm{e}^{-c\lambda_{t,2}}(\mathrm{e}^{-Y_{t,1}} - \mathrm{e}^{-c\lambda_{t,1}})}{\varphi_t}\right.\\
&\left. - \frac{\delta c \mathrm{e}^{-c\lambda_{t,2}}(\mathrm{e}^{-Y_{t,1}} - \mathrm{e}^{-c\lambda_{t,1}})^2(\mathrm{e}^{-Y_{t,2}} - \mathrm{e}^{-c\lambda_{t,2}})}{\varphi_t^2}\right] \frac{\partial \lambda_{t,2}}{\partial \boldsymbol{\theta}_2},\\
\frac{\partial^2 \lambda_{t,i}}{\partial \omega_i^2} =& 0, \quad \frac{\partial^2 \lambda_{t,i}}{\partial \beta_{ij}^2} = 0, \quad \frac{\partial^2 \lambda_{t,i}}{\partial \omega_i \partial \beta_{ij}} = 0,\\
\frac{\partial^2 \lambda_{t,i}}{\partial \alpha_{ij}^2} =& \begin{cases} 2\frac{\partial \lambda_{t-1,i}}{\partial \alpha_{ii}} + \alpha_{ii}\frac{\partial^2 \lambda_{t,i}}{\partial \alpha_{ii}^2} & i = j,\\ 0 & i \neq j. \end{cases}
\end{aligned}
$$

$$\frac{\partial^2\lambda_{t,i}}{\partial\omega_i\partial\alpha_{ij}} = \begin{cases} \frac{\partial\lambda_{t-1,i}}{\partial\omega_i} + \alpha_{ii}\frac{\partial^2\lambda_{t-1,i}}{\partial\omega_i\partial\alpha_{ii}} & i = j, \\ 0 & i \neq j. \end{cases}$$

$$\frac{\partial^2\lambda_{t,i}}{\partial\alpha_{ij}\partial\beta_{ij}} = \begin{cases} \frac{\partial\lambda_{t-1,i}}{\partial\beta_{ii}} + \alpha_{ii}\frac{\partial^2\lambda_{t-1,i}}{\partial\alpha_{ii}\partial\beta_{ii}} & i = j, \\ 0 & i \neq j. \end{cases} \quad i, j = 1, 2. \tag{A.7}$$

Proof of Lemma 1 We know that $\partial l_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ given in (A.3)–(A.6) is a martingale difference sequence with respect to $\mathcal{F}_{t-1}$. It follows that at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, $E(\partial l_t/\partial\boldsymbol{\theta}|\mathcal{F}_{t-1}) = 0$. Furthermore, from (2.3) and (A.6), we obtain the following results by iteration.

$$\frac{\partial\lambda_{t,i}}{\partial\omega_i} = \frac{1-(\alpha_{ii}^0)^t}{1-\alpha_{ii}^0}, \quad \frac{\partial\lambda_{t,i}}{\partial\alpha_{ij}} = \sum_{k=0}^{t-1}(\alpha_{ii}^0)^k\lambda_{t-1-k,j},$$
$$\frac{\partial\lambda_{t,i}}{\partial\beta_{ij}} = \sum_{k=0}^{t-1}(\alpha_{ii}^0)^k Y_{t-1-k,j}, \quad i, j = 1, 2.$$

Observe that, as $\alpha_L \le \alpha_{ij}^0 \le \alpha_U, \lambda_{t,i} \ge \omega_L, EY_{t,j}^2 < \infty, E\lambda_{t,j}^2 < \infty$, then $E(\partial\lambda_{t,i}/\partial\omega_i)^2, E(\partial\lambda_{t,i}/\partial\alpha_{ij})^2, E(\partial\lambda_{t,i}/\partial\beta_{ij})^2, i, j = 1, 2$ are all finite. Under Assumption 2, we have

$$E\left(\left[\frac{Y_{t,1}}{\lambda_{t,1}} - 1 + \frac{\delta c \mathrm{e}^{-c\lambda_{t,1}}(\mathrm{e}^{-Y_{t,2}} - \mathrm{e}^{-c\lambda_{t,2}})}{\varphi_t}\right]^2 \middle| \mathcal{F}_{t-1}\right) \le \frac{1}{\omega_L} + \frac{4\delta_U^2 c^2}{\varphi_L^2},$$
$$E\left(\left[\frac{Y_{t,2}}{\lambda_{t,2}} - 1 + \frac{\delta c \mathrm{e}^{-c\lambda_{t,2}}(\mathrm{e}^{-Y_{t,1}} - \mathrm{e}^{-c\lambda_{t,1}})}{\varphi_t}\right]^2 \middle| \mathcal{F}_{t-1}\right) \le \frac{1}{\omega_L} + \frac{4\delta_U^2 c^2}{\varphi_L^2},$$
$$E\left(\frac{(\mathrm{e}^{-Y_{t,1}} - \mathrm{e}^{-c\lambda_{t,1}})^2(\mathrm{e}^{-Y_{t,2}} - \mathrm{e}^{-c\lambda_{t,2}})^2}{\varphi_t^2} \middle| \mathcal{F}_{t-1}\right) \le \frac{16}{\varphi_L^2}.$$

Hence, by Hölder's inequality, we conclude that $E\|\partial l_t/\partial\boldsymbol{\theta}\| < \infty$. Using the martingale central limit theorem and the Cramér–Wold device, we know that $\frac{1}{\sqrt{n}}S_n(\boldsymbol{\theta}) \xrightarrow{d} \mathrm{N}(0, G)$. A consistent estimator of G is given by $G_t(\widehat{\boldsymbol{\theta}})$. □

Proof of Lemma 2 Through calculating the conditional expectation, we know that

$$E\left(-\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j^\top}\right) = E\left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i}\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j^\top}\right),$$
$$E\left(-\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\delta\partial\boldsymbol{\theta}_i^\top}\right) = E\left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial\delta}\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i^\top}\right), \quad i, j = 1, 2.$$

Thus,

$$E\left(-\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right)=E\left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\right).$$

Following the arguments similar to the proof for Lemma 3.3 of Fokianos et al. (2009), the result can be established. □

Proof of Lemma 3 It is simple to see that all terms that do not contain the partial derivatives of $\lambda_{t,i}, i=1,2$ can be controlled, without loss of generality, we choose $\theta_i=\theta_j=\theta_k=\alpha_{11}$ to verify.

$$\begin{aligned}
\frac{\partial^3 l_t(\boldsymbol{\theta})}{\partial\alpha_{11}^3}=&\left[\frac{2Y_{t,1}}{\lambda_{t,1}^3}+\frac{\delta c^3 \mathrm{e}^{-c\lambda_{t,1}-Y_{t,1}}}{\varphi_t}\right.\\
&\left.+\frac{\delta^2 c^3\mathrm{e}^{-2c\lambda_{t,1}}(\mathrm{e}^{-Y_{t,2}}-\mathrm{e}^{-c\lambda_{t,2}})^2(3\varphi_t+2\delta\mathrm{e}^{-c\lambda_{t,1}})}{\varphi_t^3}\right]\left(\frac{\partial\lambda_{t,1}}{\partial\alpha_{11}}\right)^3-\left[\frac{3Y_{t,1}}{\lambda_{t,1}^2}\right.\\
&\left.+\frac{2\delta c^2\mathrm{e}^{-c\lambda_{t,1}}(\mathrm{e}^{-Y_{t,2}}-\mathrm{e}^{-c\lambda_{t,2}})(\varphi_t+\delta\mathrm{e}^{-Y_{t,2}}-\delta\mathrm{e}^{-c\lambda_{t,2}})}{\varphi_t^2}\right]\frac{\partial^2\lambda_{t,1}}{\partial\alpha_{11}^2}\frac{\partial\lambda_{t,1}}{\partial\alpha_{11}}\\
&+\left[\frac{Y_{t,1}}{\lambda_{t,1}}-1+\frac{\delta c\mathrm{e}^{-c\lambda_{t,1}}(\mathrm{e}^{-Y_{t,2}}-\mathrm{e}^{-c\lambda_{t,2}})}{\varphi_t}\right]\frac{\partial^3\lambda_{t,1}}{\partial\alpha_{11}^3}\\
\le&\left[\frac{2Y_{t,1}}{\omega_L^3}+\frac{\delta_U c^3(\varphi_L^2+12\delta_U\varphi_L+16\delta_U^2)}{\varphi_L^3}\right]\left(\frac{\partial\lambda_{t,1}}{\partial\alpha_{11}}\right)^3\\
&+\left[\frac{3Y_{t,1}}{\omega_L}+\frac{4\delta_U c^2(\varphi_L+2\delta_U)}{\varphi_L^2}\right]\frac{\partial^2\lambda_{t,1}}{\partial\alpha_{11}^2}\frac{\partial\lambda_{t,1}}{\partial\alpha_{11}}\\
&+\left(\frac{Y_{t,1}}{\omega_L}+\frac{2\delta_U c}{\varphi_L}-1\right)\frac{\partial^3\lambda_{t,1}}{\partial\alpha_{11}^3}.
\end{aligned}$$

Next, turn to the derivatives where the first- and second-order derivatives are given by (A.6) and (A.7). Similar to the proof for Lemma 3.4 of Fokianos et al. (2009), we know that $\partial\lambda_{t,1}/\partial\alpha_{11}\le\mu_{t1}$, $\partial^2\lambda_{t,1}/\partial\alpha_{11}^2\le\mu_{t2}$, $\partial^3\lambda_{t,1}/\partial\alpha_{11}^3\le\mu_{t3}$. Hence there exists a constant C such that

$$\left|\frac{\partial^3 l_t(\boldsymbol{\theta})}{\partial\alpha_{11}^3}\right|\le m_t:=Cg(Y_{t,1})h(\mu_{ts}).$$

Considering the expectation of each term in m_t separately, we can see that $M:=Em_t<\infty$, likewise for other terms. Finally, it can be shown that the conclusion holds. □

Proof of Theorem 2 By using Lemmas 1–3, all conditions of Lemma 1 in Jensen and Rahbek (2004) have been verified for model (2.3), the probability that $\hat{\boldsymbol{\theta}}$ is a unique solution to $S_n(\boldsymbol{\theta})=0$ in $O(\boldsymbol{\theta}^0)$ tends to one. Furthermore, by Taylor's expansion,

$$0 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} + \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_t(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0),$$

where $\boldsymbol{\theta}^*$ is an intermediate point between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^0$. We will find that $\hat{\boldsymbol{\theta}}$ is asymptotically normal, and thus Theorem 2 holds. □

References

Ahmad A, Francq C (2016) Poisson QMLE of count time series models. J Time Ser Anal 37:291–314

Christou V, Fokianos K (2014) Quasi-likelihood inference for negative binomial time series models. J Time Ser Anal 35:55–78

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

Doukhan P, Fokianos K, Tjøtheim D (2012) On weak dependence conditions for Poisson autoregressions. Stat Probab Lett 82:942–948

Ferland R, Latour A, Oraichi D (2006) Integer-valued GARCH process. J Time Ser Anal 27:923–942

Fokianos K (2012) Count time series models. In: Rao TS, Rao SS, Rao CR (eds) Time series analysis: methods and applications, vol 30. Handbook of statistics. Elsevier, Amsterdam, pp 315–347

Fokianos K (2016) Statistical analysis of count time series models: a GLM perspective. In: Davis RA, Holan SH, Lund R, Ravishanker N (eds) Handbook of discrete-valued time series. Chapman and Hall/CRC, Boca Raton, pp 3–27

Fokianos K, Tjøtheim D (2011) Log-linear Poisson autoregression. J Multivar Anal 102:563–578

Fokianos K, Tjøtheim D (2012) Nonlinear Poisson autoregression. Ann Inst Stat Math 64:1205–1225

Fokianos K, Rahbek A, Tjøstheim D (2009) Poisson autoregression. J Am Stat Assoc 104:1430–1439

Gonçalves E, Mendes-Lopes N, Silva F (2015) Infinitely divisible distributions in integer-valued GARCH models. J Time Ser Anal 36:503–527

Gonçalves E, Mendes-Lopes N, Silva F (2016) Zero-inflated compound Poisson distributions in integer-valued GARCH models. Statistics 50:558–578

Heinen A, Rengifo E (2007) Multivariate autoregressive modeling of time series count data using copulas. J Empir Financ 14:564–583

Jensen ST, Rahbek A (2004) Asymptotic inference for nonstationary GARCH. Econom Theor 20:1203–1226

Lakshminarayana J, Pandit SNN, Rao KS (1999) On a bivariate Poisson distribution. Commun Stat Theor Method 28:267–276

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

Lee S, Lee Y, Chen CWS (2016) Parameter change test for zero-inflated generalized Poisson autoregressive models. Statistics 50:540–557

Lee S, Lee Y, Tjøtheim D (2017) Asymptotic normality and parameter change test for bivariate Poisson INGARCH models. Test (forthcoming)

Liu H (2012) Some models for time series of counts. Dissertations, Columbia University. https://core.ac.uk/download/pdf/27293903.pdf

Meyn SP, Tweedie RL (2009) Markov chains and stochastic stability, 2nd edn. Cambridge University Press, New York

Neumann MH (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. Stat Model 11:325–349

Pedeli X, Karlis D (2013a) Some properties of multivariate INAR(1) processes. Comput Stat Data Anal 67:213–225

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

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

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: Davis RA, Holan SH, Lund R, Ravishanker N (eds) Handbook of discrete-valued time series. Chapman and Hall/CRC, Boca Raton, pp 77–100
Wang C, Liu H, Yao JF, Davis RA, Li WK (2014) Self-excited threshold Poisson autoregression. J Am Stat Assoc 109:776–787
Wu W, Shao X (2004) Limit theorems for iterated random functions. J Appl Probab 41:425–436
Zhu F (2011) A negative binomial integer-valued GARCH model. J Time Ser Anal 32:54–67