stats

Article

A New Bivariate INAR(1) Model with Time-Dependent Innovation Vectors

Huaping Chen [1,*], Fukang Zhu [2] and Xiufang Liu [3]

[1] School of Mathematics and Statistics, Henan University, Kaifeng 475004, China
[2] School of Mathematics, Jilin University, Changchun 130012, China
[3] College of Mathematics, Taiyuan University of Technology, Taiyuan 030024, China
* Correspondence: chenhp0107@henu.edu.cn

Abstract: Recently, there has been a growing interest in integer-valued time series models, especially in multivariate models. Motivated by the diversity of the infinite-patch metapopulation models, we propose an extension to the popular bivariate INAR(1) model, whose innovation vector is assumed to be time-dependent in the sense that the mean of the innovation vector is linearly increased by the previous population size. We discuss the stationarity and ergodicity of the observed process and its subprocesses. We consider the conditional maximum likelihood estimate of the parameters of interest, and establish their large-sample properties. The finite sample performance of the estimator is assessed via simulations. Applications on crime data illustrate the model.

Keywords: bivariate INAR model; bivariate poisson distribution; time-dependent innovation; time series of counts; stability; parameters estimation

1. Introduction

Bivariate count data occur in many contexts, often as the counts of two events, objects or individuals during a certain period of time. For example, such counts occur in epidemiology when two kinds of related diseases are examined, in criminology when two kinds of crimes are committed, in business when the volume of sales of two correlated products are observed or in manufacturing when two similar products are produced.

In real application, the observed time series data are often discrete, over-dispersed (the empirical variance is greater than the empirical mean) and have other features such as time dependence. Many univariate models have been proposed to deal with integer-valued time series data based on the univariate binomial thinning operator " $\circ$ ", which is proposed by Steutel and van Harn [1]:

$$\alpha \circ X = \sum_{i=1}^{X} W_i, \tag{1}$$

where X is a non-negative integer-valued random variable and $P(W_i = 1) = 1 - P(W_i = 0) = \alpha$. The INAR(1) model [2], The BAR(1) model [3], the INAR(p) model [4], The PDINAR(1) model [5] and the BARIO model [6] are very popular in analyzing non-negative integer-valued time series; see Weiß [7], Scotto et al. [8] and Davis et al. [9] for recent reviews on this topic. Motivated by infinite-patch metapopulation models discussed in Buckley and Pollett [10], Weiß [11] proposed an extension to the popular Poisson INAR(1) model, which is characterized by time-dependent innovations, i.e., the mean of the innovation is linearly increased by the previous population size. An important advantage of this model is that it gives a reasonable interpretation for immigration, which becomes more attractive if the current population is large; see Weiß [11] for an application to iceberg order data.

Univariate models are extensively investigated in the literature, but relatively few multivariate models, especially for bivariate versions, have been studied in detail. Franke and Rao [12] proposed a multivariate INAR(1) model, which is generalized to the p-order case

check for updates

Citation: Chen, H.; Zhu, F.; Liu, X. A New Bivariate INAR(1) Model with Time-Dependent Innovation Vectors. *Stats* **2022**, *5*, 819–840. https://doi.org/10.3390/stats5030048

Academic Editor: Wei Zhu

Received: 11 July 2022
Accepted: 15 August 2022
Published: 19 August 2022

by Latour [13]. Pedeli and Karlis [14] discussed a tractable bivariate INAR(1) model, which can be used to deal with bivariate count data with equi-dispersion and over-dispersion, but with small flexibility. See Pedeli and Karlis [15] for the estimation of the BINAR model and Pedeli and Karlis [16] for a further discussion of the properties of the multivariate INAR(1) model. Based on hierarchical dynamic models, Ravishanker et al. [17] described a Bayesian framework for estimation and prediction for multivariate times series of counts. Popović [18] proposed a bivariate INAR(1) model with random coefficients based on different binomial thinning operators. The above models assumed the innovations of their marginal models are independent and identically distributed. Based on a finite range of counts, Scotto et al. [19] considered the density-dependent bivariate binomial autoregressive models by using their state-dependent thinning concept. Li et al. [20] proposed a bivariate random coefficient INAR(1) model with asymmetric Hermite innovations.

Inspired by Weiß [11], we aim at providing a bivariate INAR model to analyze bivariate time series with time dependence and cross-correlation. The first contribution is that this paper gives an available method to capture the time-dependence trend by imposing the past information in the distribution of the innovational vector, which in turn makes the cross-correlation between the two entries into an innovation vector. The second contribution is that the new model not only allows autocorrelation but also allows cross-correlation. The third contribution is that this paper illustrates the stationarity and ergodicity of the extended bivariate INAR process and its two subprocesses, which is important to derive the consistence and asymptotic normality of the CML estimation.

The remainder of the paper is organized as follows. In Section 2, we first give brief reviews of the bivariate Poisson distribution and the bivariate binomial thinning operator, based on which we give the definition of the new bivariate INAR(1) model. Conditional maximum likelihood (CML) estimates and the asymptotic properties of unknown parameters are discussed in Section 3. A simulation and two real data examples that show the effectiveness of the new model are given in Sections 4 and 5, respectively. Conclusions are made in Section 6.

2. A New Bivariate INAR(1) Model

For readability, we first give a brief review of the bivariate Poisson distribution.

Definition 1. *If the joint probability mass function (pmf) of (X, Y) satisfies*

$$P(X = x, Y = y) = \mathrm{e}^{-(\lambda_1+\lambda_2-\phi)} \frac{(\lambda_1 - \phi)^x}{x!} \frac{(\lambda_2 - \phi)^y}{y!} \sum_{i=0}^{\min(x,y)} \binom{x}{i} \binom{y}{i} i! \left[\frac{\phi}{(\lambda_1 - \phi)(\lambda_2 - \phi)} \right]^i, \quad (2)$$

where $\lambda_1, \lambda_2 > 0$ and $\phi \in (0, \min(\lambda_1, \lambda_2))$, then (X, Y) is called a bivariate Poisson random variable with parameters $(\lambda_1, \lambda_2, \lambda_3)$, i.e., $\mathrm{BP}(\lambda_1, \lambda_2, \phi)$.

From Kocherlakota and Kocherlakota [21], we obtain the fact that if (X, Y) follows $\mathrm{BP}(\lambda_1, \lambda_2, \phi)$, there must exist three mutually independent random variables Z_1, Z_2, Z_3 such that $X = Z_1 + Z_3$ and $Y = Z_2 + Z_3$, where Z_1, Z_2 and Z_3 follow Poisson$(\lambda_1 - \phi)$, Poisson$(\lambda_2 - \phi)$ and Poisson(ϕ), respectively. Then, we have the conclusion that $\mathrm{Cov}(X, Y) = \phi$. In addition, $P(X = x, Y = y)$, given in (2), is continuous and differentiable. For convenience, we denote $f(x, y, \lambda_1, \lambda_2, \phi) = P(X = x, Y = y)$. By using Lemma A3 in Li et al. [20], we obtain that

$$\frac{\partial f(x,y,\lambda_1,\lambda_2,\phi)}{\partial \lambda_1} = f(x-1,y,\lambda_1,\lambda_2,\phi) - f(x,y,\lambda_1,\lambda_2,\phi), \tag{3}$$

$$\frac{\partial f(x,y,\lambda_1,\lambda_2,\phi)}{\partial \lambda_2} = f(x,y-1,\lambda_1,\lambda_2,\phi) - f(x,y,\lambda_1,\lambda_2,\phi), \tag{4}$$

$$\begin{aligned}\frac{\partial f(x,y,\lambda_1,\lambda_2,\phi)}{\partial \phi} &= f(x,y,\lambda_1,\lambda_2,\phi) - f(x-1,y,\lambda_1,\lambda_2,\phi) - f(x,y-1,\lambda_1,\lambda_2,\phi) \\ &\quad + f(x-1,y-1,\lambda_1,\lambda_2,\phi).\end{aligned} \tag{5}$$

Applying the univariate binomial thinning operator " $\circ$ " given in (1) to the bivariate case with $\boldsymbol{X} = (X_1, X_2)^\top$ leads to the bivariate binomial thinning operator:

$$\boldsymbol{A} \circ \boldsymbol{X} = \begin{pmatrix} \alpha_{11} \circ X_1 + \alpha_{12} \circ X_2 \\ \alpha_{21} \circ X_1 + \alpha_{22} \circ X_2 \end{pmatrix} \text{with } \boldsymbol{A} = (\alpha_{ij})_{2\times 2},$$

where $\alpha_{ij} \in (0,1), i,j = 1,2$, X_1 and X_2 are non-negative integer-valued random variables, and all the thinnings are performed independent of each other.

By calculation, $E(\boldsymbol{A} \circ \boldsymbol{X}) = \boldsymbol{A}E(\boldsymbol{X})$. Denoting $\boldsymbol{V}$ as the 2×2 variance matrix of the Bernoulli random variables $\alpha_{ij} \circ X_j$ with $(\boldsymbol{V})_{ij} = \alpha_{ij}(1-\alpha_{ij})$, $i,j = 1,2$, we obtain that $E((\boldsymbol{A} \circ \boldsymbol{X})(\boldsymbol{A} \circ \boldsymbol{X})^\top) = \boldsymbol{A}E(\boldsymbol{X}\boldsymbol{X}^\top)\boldsymbol{A}^\top + \mathrm{diag}(\boldsymbol{V}E(\boldsymbol{X}))$. Furthermore, if all the counting series of $\boldsymbol{A} \circ \boldsymbol{X}$ and $\boldsymbol{B} \circ \boldsymbol{Y}$ are independent, $E\big((\boldsymbol{A} \circ \boldsymbol{X})(\boldsymbol{B} \circ \boldsymbol{Y})^\top\big) = \boldsymbol{A}E(\boldsymbol{X}\boldsymbol{Y}^\top)\boldsymbol{B}^\top$.

In the following, we give the definition of the new bivariate INAR(1) model, which not only includes the property of the models defined by Pedeli and Karlis [14,16] but also allows the innovation vectors $\{\boldsymbol{\epsilon}_t\}$ to be time-dependent.

Definition 2. *Let $\boldsymbol{X}_t = (X_{1t}, X_{2t})^\top$ be non-negative integer-valued bivariate random vector. If the process $\{\boldsymbol{X}_t\}$ satisfies*

$$\boldsymbol{X}_t = \boldsymbol{A} \circ \boldsymbol{X}_{t-1} + \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z}, \tag{6}$$

then $\{\boldsymbol{X}_t\}$ is said to follow the extended bivariate INAR(1) process, where $\boldsymbol{A} = (\alpha_{ij})_{2\times 2}$, $0 < \alpha_{ij} < 1$, for any $i,j = 1,2, \boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t})^\top \sim \mathrm{BP}(\lambda_{1t}, \lambda_{2t}, \phi)$ with $(\lambda_{1t}, \lambda_{2t})^\top = \boldsymbol{B}\boldsymbol{X}_{t-1} + \boldsymbol{C}, \boldsymbol{B} = (b_{ij})_{2\times 2}$, $\boldsymbol{C} = (c_1, c_2)^\top, 0 < b_{ij} < 1, c_i > 0, i,j = 1,2$.

For simplicity, we denote the new model as the EBINAR(1) model. It is easy to see that the ith equation of model (6) is presented by:

$$X_{it} = \alpha_{i1} \circ X_{1,t-1} + \alpha_{i2} \circ X_{2,t-1} + \epsilon_{it}, \quad i = 1,2. \tag{7}$$

Notice that the model given by (7) is similar to the one discussed in Weiß [11], the main difference is that X_{it} involves two paralleled survivors $X_{1,t-1}$ and $X_{2,t-1}$. It is known that the EBINAR(1) process $\{\boldsymbol{X}_t, t \in \mathbb{Z}\}$ has two parts: the first part consists of the survivors of the elements of the system at the preceding time $t-1$, denoted by $\boldsymbol{X}_{t-1}$; the other part is comprised by the time-dependent innovation vector $\boldsymbol{\epsilon}_t$, which implies that the mean of the innovation vector is linearly increased by the previous population size.

Remark 1. *(1). If both $\boldsymbol{A}$ and $\boldsymbol{B}$ are diagonal matrices, the component equation given in (7) becomes to the one discussed by Weiß [11].*

(2). If $\boldsymbol{A}$ is diagonal and $\boldsymbol{B} = \boldsymbol{0}$, model (6) becomes the one discussed in Pedeli and Karlis [14], but it is worth mentioning that the autoregression matrix in Pedeli and Karlis [14] is diagonal, which means that it causes no cross-correlation in the counts.

(3). If $\boldsymbol{A}$ is non-diagonal and $\boldsymbol{B} = \boldsymbol{0}$, model (6) becomes the one discussed in Pedeli and Karlis [16], which accounts for cross-correlation in the counts, but they still keep the innovations of their marginal models independent and identically distributed such that the time dependence can not to be captured.

To derive the pmf the EBINAR(1) process, we first denote $h(k, m_1, m_2, \alpha_1, \alpha_2) := P(X + Y = k)$ is the convolution of $X + Y, \forall k \geq 0$ with $X \sim \text{Bin}(m_1, \alpha_1)$ and $Y \sim \text{Bin}(m_2, \alpha_2)$. By calculation, we obtain that

$$h(k, m_1, m_2, \alpha_1, \alpha_2) = \sum_{j=0}^{s} P(X = j|m_1, \alpha_1)P(Y = k - j|m_2, \alpha_2). \tag{8}$$

Furthermore, by using Lemma A3 in Li et al. [20],

$$\frac{\partial h(k, m_1, m_2, \alpha_1, \alpha_2)}{\partial \alpha_1} = m_1\big(h(k-1, m_1 - 1, m_2, \alpha_1, \alpha_2) - h(k, m_1 - 1, m_2, \alpha_1, \alpha_2)\big), \tag{9}$$
$$\frac{\partial h(k, m_1, m_2, \alpha_1, \alpha_2)}{\partial \alpha_2} = m_2\big(h(k-1, m_1, m_2 - 1, \alpha_1, \alpha_2) - h(k, m_1, m_2 - 1, \alpha_1, \alpha_2)\big). \tag{10}$$

Second, we denote $\boldsymbol{\varsigma} = (\varsigma_1, \varsigma_2)^\top$, $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2)^\top$, $\boldsymbol{k} = (k_1, k_2)^\top$ and let $x = \varsigma_1 - k_1$ and $y = \varsigma_2 - k_2$. Then, the conditional probability distribution of the EBINAR(1) process takes the following form:

$$\begin{aligned}
P(\boldsymbol{\varsigma}|\boldsymbol{\vartheta}) &:= P(\boldsymbol{X}_t = \boldsymbol{\varsigma}|\boldsymbol{X}_{t-1} = \boldsymbol{\vartheta}) = \sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} P(\boldsymbol{A} \circ \boldsymbol{X}_{t-1} = \boldsymbol{k}|\boldsymbol{\epsilon}_t = \boldsymbol{\varsigma} - \boldsymbol{k})P(\boldsymbol{\epsilon}_t = \boldsymbol{\varsigma} - \boldsymbol{k}) \\
&= \sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} \big(P(\alpha_{11} \circ X_{1,t-1} + \alpha_{12} \circ X_{2,t-1} = k_1) \times P(\alpha_{21} \circ X_{1,t-1} + \alpha_{22} \circ X_{2,t-1} = k_2) \\
&\quad \times P(\epsilon_{1t} = \varsigma_1 - k_1, \epsilon_{2t} = \varsigma_2 - k_2)\big) \\
&= \sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} h(k_1, \vartheta_1, \vartheta_2, \alpha_{11}, \alpha_{12})h(k_2, \vartheta_1, \vartheta_2, \alpha_{21}, \alpha_{22})f(x, y, \lambda_{1t}, \lambda_{2t}, \phi),
\end{aligned} \tag{11}$$

where $g_1 = \min(\varsigma_1, \vartheta_1)$, $g_2 = \min(\varsigma_2, \vartheta_2)$,

$$\begin{aligned}
&f(x, y, \lambda_{1t}, \lambda_{2t}, \phi) = P(\epsilon_{1t} = x, \epsilon_{2t} = y) \\
&\overset{(2)}{=} \exp(\lambda_{1t} + \lambda_{2t} - \phi)\frac{(\lambda_{1t} - \phi)^x}{x!}\frac{(\lambda_{2t} - \phi)^y}{y!}\sum_{i=0}^{\min(x,y)}\binom{x}{i}\binom{y}{i}i!\left[\frac{\phi}{(\lambda_{1t} - \phi)(\lambda_{2t} - \phi)}\right]^i
\end{aligned}$$

with $\lambda_{1t} = b_{11}X_{1,t-1} + b_{12}X_{2,t-1} + c_1$ and $\lambda_{2t} = b_{21}X_{1,t-1} + b_{22}X_{2,t-1} + c_2$.

If the largest eigenvalue of non-negative matrix $\boldsymbol{A}$ is less than 1, then the bivariate marginal distribution of model (6) can be expressed in terms of the bivariate innovation vectors:

$$\boldsymbol{X}_t \overset{d}{=} \boldsymbol{A}^k \circ \boldsymbol{X}_{t-k} + \sum_{j=0}^{k-1} \boldsymbol{A}^j \circ \boldsymbol{\epsilon}_{t-j} = \boldsymbol{A}^t \circ \boldsymbol{X}_0 + \sum_{j=0}^{t-1} \boldsymbol{A}^j \circ \boldsymbol{\epsilon}_{t-j}, k = 1, 2, \cdots, t, \tag{12}$$

where $\boldsymbol{A}^0$ is an identity matrix, and $\boldsymbol{X}_0$ is the initial value of the process.

In what follows, we first discuss the stationarity and ergodicity of processes (6) and (7), respectively. Second, we obtain the first two-moment of $\{\boldsymbol{X}_t\}$ and $\{\boldsymbol{\epsilon}_t\}$, respectively. Third, we give a necessary and sufficient condition for the existence of $E(X_{1t})^k$ and $E(X_{2t})^k$ for any fixed positive integer k. These properties are necessary to derive the asymptotic properties of the estimators.

Theorem 1. *Let $\{\boldsymbol{X}_t = (X_{1t}, X_{2t})^\top\}$ follow (6), $\boldsymbol{\Gamma} = \boldsymbol{A} + \boldsymbol{B} = (\gamma_{ij})_{i,j=1,2}$ with $0 < \gamma_{ij} < 1$. If the largest eigenvalue of $\boldsymbol{\Gamma}$ is less than 1, there exists a strictly stationary and ergodic process satisfying (7).*

Proof. Let $W_{t,k}$, $V_{t,l}$ and δ_{1t} be independent of each other and each of them be independent and identically distributed, i.e., $W_{t,k} \sim \text{Bin}(1, \alpha_{11}) + \text{Poi}(b_{11})$, $V_{t,l} \sim \text{Bin}(1, \alpha_{12}) + \text{Poi}(b_{12})$ and $\delta_{1t} \sim \text{Poi}(c_1)$, where $\text{Bin}(1, \alpha_{1j}) + \text{Poi}(b_{1j})$ means the convolution of the distributions $\text{Bin}(1, \alpha_{1j})$ and $\text{Poi}(b_{1j})$, $k = 1, 2, \cdots, X_{1,t-1}$ and $l = 1, 2, \cdots, X_{2,t-1}$; see Weiß [11] for

details. According to the concepts of bivariate binomial thinning and the additivity of binomial distribution and Poisson distribution, (7) can be rewritten as

$$X_{1t} \overset{d}{=} W_{t,1} + \cdots + W_{t,X_{1,t-1}} + V_{t,1} + \cdots + V_{t,X_{2,t-1}} + \delta_{1t}. \tag{13}$$

Since $\gamma = \max(\alpha_{ij} + b_{ij}) < 1$, we have $E(W_{t,k}) = \alpha_{11} + b_{11} < 1$ and $E(V_{t,l}) = \alpha_{12} + b_{12} < 1$. Denote $H(n) = \sum_{k=1}^{n} \frac{1}{k}$ and $H(0) = 0$, then $E(H(\delta_{1t})) = \sum_{k=1}^{+\infty} \frac{1}{k} P(\delta_{1t} \geq k)$. In addition, that $H(\delta_{1t}) \leq \delta_{1t}$ and $E(\delta_{1t}) = c_1 < \infty$, thus, $EH(\delta_{1t}) \leq E(\delta_{1t}) < \infty$. Therefore, the Theorem of Heathcote [22] holds. Hence, there exists a stationary marginal distribution of (7), i.e., there exists a strictly stationary process satisfying (7). Similarly, we also have a similar conclusion for X_{2t}. □

To prove the stationaity of the EBINAR(1) process, we first introduce a sequence of random variables $\{\boldsymbol{X}_t^{(n)}\}$ that could be considered as approximations to $\{\boldsymbol{X}_t\}$ with

$$\boldsymbol{X}_t^{(n)} = \begin{cases} \mathbf{0}, & n < 0, \\ \boldsymbol{R}_t, & n = 0, \\ \boldsymbol{A} \circ \boldsymbol{X}_{t-1}^{(n-1)} + \boldsymbol{B}\boldsymbol{X}_{t-1}^{(n-1)} + \boldsymbol{R}_t, & n > 0, \end{cases}$$

where the largest eigenvalues of the non-negative matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{\Gamma} := \boldsymbol{A} + \boldsymbol{B}$ are less than 1, all of the non-negative matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $I - \boldsymbol{\Gamma}$ are invertible, $\boldsymbol{R}_t = (R_{1t}, R_{2t})^\top$, R_{1t} is independent with R_{2t} and R_{it} follows a Poisson distribution with the parameter c_i, $i = 1, 2$.

Theorem 2. *If the conditions of Theorem 1 hold, there exists a strictly stationary process satisfying (6).*

Proof. Because

$$\begin{pmatrix} \boldsymbol{X}_1^{(0)} \\ \vdots \\ \boldsymbol{X}_k^{(0)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{R}_1 \\ \vdots \\ \boldsymbol{R}_k \end{pmatrix} \text{and} \begin{pmatrix} \boldsymbol{X}_{h+1}^{(0)} \\ \vdots \\ \boldsymbol{X}_{h+k}^{(0)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{R}_{h+1} \\ \vdots \\ \boldsymbol{R}_{h+k} \end{pmatrix}$$

are identically distributed for $(\boldsymbol{R}_1, \cdots, \boldsymbol{R}_k)^\top$ and $(\boldsymbol{R}_{h+1}, \cdots, \boldsymbol{R}_{h+k})^\top$ are identically distributed. Thus, $\{\boldsymbol{X}_t^{(0)}\}$ is strictly stationary. Now, we suppose $\{\boldsymbol{X}_t^{(n)}\}$ is strictly stationary. Then,

$$\begin{pmatrix} \boldsymbol{X}_1^{(n+1)} \\ \vdots \\ \boldsymbol{X}_k^{(n+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{R}_1 \\ \vdots \\ \boldsymbol{R}_k \end{pmatrix} + \begin{pmatrix} \boldsymbol{A} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{A} \end{pmatrix} \circ \begin{pmatrix} \boldsymbol{X}_0^{(n)} \\ \vdots \\ \boldsymbol{X}_{k-1}^{(n)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{B} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{B} \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_0^{(n)} \\ \vdots \\ \boldsymbol{X}_{k-1}^{(n)} \end{pmatrix} \tag{14}$$

and

$$\begin{pmatrix} \boldsymbol{X}_{h+1}^{(n+1)} \\ \vdots \\ \boldsymbol{X}_{h+k}^{(n+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{R}_{h+1} \\ \vdots \\ \boldsymbol{R}_{h+k} \end{pmatrix} + \begin{pmatrix} \boldsymbol{A} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{A} \end{pmatrix} \circ \begin{pmatrix} \boldsymbol{X}_h^{(n)} \\ \vdots \\ \boldsymbol{X}_{h+k-1}^{(n)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{B} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{B} \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_h^{(n)} \\ \vdots \\ \boldsymbol{X}_{h+k-1}^{(n)} \end{pmatrix}. \tag{15}$$

Because the joint distributions of the variables involved in the right-hand sides of (14) and (15) are identical, thus, $\left(\boldsymbol{X}_1^{(n+1)}, \cdots, \boldsymbol{X}_k^{(n+1)}\right)^\top$ and $\left(\boldsymbol{X}_{h+1}^{(n+1)}, \cdots, \boldsymbol{X}_{h+k}^{(n+1)}\right)^\top$ in the left-hand side of the two equations are identically distributed. Hence, the process $\{\boldsymbol{X}_t^{(n)}\}$ is strictly stationary. Therefore, $\{\boldsymbol{X}_t\}$ is a strictly stationary process. □

Theorem 3. *If the conditions of Theorem 1 hold, $\{\boldsymbol{X}_t\}$ is a null recurrent and ergodic Markov chain.*

Proof. First, we prove $\{\boldsymbol{X}_t\}$ is null recurrent. Because $\boldsymbol{X}_t \overset{d}{=} \sum_{j=0}^{\infty} \boldsymbol{A}^j \circ \boldsymbol{\epsilon}_{t-j}$, then $P^t_{\mathbf{0},\mathbf{0}} = P(\boldsymbol{X}_t = \mathbf{0} | \boldsymbol{X}_0 = \mathbf{0}) = \prod_{j=0}^{k-1} P(\boldsymbol{A}^j \circ \boldsymbol{\epsilon}_{t-j} = \mathbf{0})$ with probability one. Let $\boldsymbol{A}^j = (p_{ij})$ and $\gamma = \max(p_{ij}), \forall i = 1,2,\ \forall j = 1,2$. Then, we obtain that

$$
\begin{aligned}
&P(\boldsymbol{A}^j \circ \boldsymbol{\epsilon}_{t-j} \neq \mathbf{0}) = P(\{p_{11} \circ \epsilon_{1,t-j} + p_{12} \circ \epsilon_{2,t-j} \geq 1\} \cup \{p_{21} \circ \epsilon_{1,t-j} + p_{22} \circ \epsilon_{2,t-j} \geq 1\}) \\
&\leq P(p_{11} \circ \epsilon_{1,t-j} + p_{12} \circ \epsilon_{2,t-j} \geq 1) + P(p_{21} \circ \epsilon_{1,t-j} + p_{22} \circ \epsilon_{2,t-j} \geq 1) \\
&\leq P(p_{11} \circ \epsilon_{1,t-j} \geq 1) + P(p_{12} \circ \epsilon_{2,t-j} \geq 1) + P(p_{21} \circ \epsilon_{1,t-j} \geq 1) + P(p_{22} \circ \epsilon_{2,t-j} \geq 1) \\
&\leq 2\Big[P(\gamma^j \circ \epsilon_{1,t-j} \geq 1) + P(\gamma^j \circ \epsilon_{2,t-j} \geq 1)\Big] \\
&\leq 2\Big[E(\gamma^j \circ \epsilon_{1,t-j}) + E(\gamma^j \circ \epsilon_{2,t-j})\Big] = 2\gamma^j(\mu_{\epsilon_1} + \mu_{\epsilon_2}).
\end{aligned}
$$

According to Theorem 2, there exists an $M > 0$ such that $\mu_{\epsilon_i} \leq M/4,\ i = 1,2$. Then, we have $P(\boldsymbol{A}^j \circ \boldsymbol{\epsilon}_{t-j} = \mathbf{0}) \geq 1 - M\gamma^j$. Hence,

$$
\begin{aligned}
P^t_{\mathbf{0},\mathbf{0}} \geq \prod_{j=0}^{t-1} P(\boldsymbol{A}^j \circ \boldsymbol{\epsilon}_{t-j} = \mathbf{0}) \geq \prod_{j=0}^{t-1}(1 - M\gamma^j) = \exp\{\textstyle\sum_{j=0}^{t-1} \log(1 - M\gamma^j)\} \\
\geq \exp\{\log(1 - M\gamma^\tau)/(1 - r)\} > 0,\ \forall t > \tau.
\end{aligned}
$$

Therefore, $\lim\limits_{t\to\infty} P^t_{\mathbf{0},\,\mathbf{0}} > 0$. Thus, $\sum_{k=0}^{\infty} P^t_{\mathbf{0},\mathbf{0}} = \infty$, i.e., $\mathbf{0}$ is a recurrent state.

Second, we illustrate the ergodicity. For all states $\boldsymbol{\varsigma}, \boldsymbol{\vartheta}, \boldsymbol{\kappa}_{t-2}, \boldsymbol{\kappa}_{t-3}, \cdots$, we have

$$
P(\boldsymbol{X}_t = \boldsymbol{\varsigma} | \boldsymbol{X}_{t-1} = \boldsymbol{\vartheta}, \boldsymbol{X}_{t-2} = \boldsymbol{\kappa}_{t-2}, \cdots) = P(\boldsymbol{X}_t = \boldsymbol{\varsigma} | \boldsymbol{X}_{t-1} = \boldsymbol{\vartheta}) = P(\boldsymbol{\vartheta}, \boldsymbol{\varsigma}),
$$

where $P(\boldsymbol{\vartheta}, \boldsymbol{\varsigma})$ denotes the transition probability from state $\boldsymbol{\vartheta}$ to state $\boldsymbol{\varsigma}$. Thus, $\{\boldsymbol{X}_t\}$ is a homogeneous Markov chain. Since $\alpha_{ij}, b_{ij} \in (0,1)$, thus $P(\epsilon_{1t} = \varsigma_1 - k_1, \epsilon_{2t} = \varsigma_2 - k_2) > 0$. Denote n-state transition probability from state $\boldsymbol{\varsigma}$ to state $\boldsymbol{\vartheta}$ with $P^n_{\boldsymbol{\varsigma}\boldsymbol{\vartheta}}$. For a given $\boldsymbol{X}_{t-1}$, the conditional probability function of the random vector $\boldsymbol{X}_t$ is derived by:

$$
\begin{aligned}
&P(X_{1t} = \varsigma_1, X_{2t} = \varsigma_2 | X_{1,t-1} = \vartheta_1, X_{2,t-1} = \vartheta_2) \\
&\quad = P(X_{1t} = \varsigma_1 | X_{1,t-1} = \vartheta_1, X_{2,t-1} = \vartheta_2) P(X_{2t} = \varsigma_2 | X_{1,t-1} = \vartheta_1, X_{2,t-1} = \vartheta_2, X_{1t} = \varsigma_1),
\end{aligned}
$$

then $P^1_{\boldsymbol{\tau}\boldsymbol{v}} > 0$ for all $\boldsymbol{\tau}, \boldsymbol{v} \in \mathbb{N}_0^2$. According to (12), every state can be reached from any other state with positive probability in a finite number of steps, analogously. Hence, $\{\boldsymbol{X}_t\}$ is irreducible. By (12), k steps of conditional probability distribution $P^k_{\mathbf{0},\mathbf{0}}$ are obtained with:

$$
\begin{aligned}
P^k_{\mathbf{0},\mathbf{0}} = P(\boldsymbol{X}_t = \mathbf{0} | \boldsymbol{X}_{t-k} = \mathbf{0}) = P(\boldsymbol{A}^k \circ \boldsymbol{X}_{t-k} + \textstyle\sum_{j=0}^{k-1} \boldsymbol{A}^j \circ \boldsymbol{\epsilon}_{t-j} = \mathbf{0} | \boldsymbol{X}_{t-k} = \mathbf{0}) \\
= \underbrace{P(\boldsymbol{A}^k \circ \boldsymbol{X}_{t-k} = \mathbf{0} | \boldsymbol{X}_{t-k} = \mathbf{0})}_{\text{(V)}} \underbrace{\prod_{j=0}^{k-1} P(\boldsymbol{A}^j \circ \boldsymbol{\epsilon}_{t-j} = \mathbf{0})}_{\text{(VI)}}.
\end{aligned}
$$

Note that the first multiplier (V) is positive, which can be obtained by a similar method to (11). Denoting $\boldsymbol{A}^j = (p_{ij})_{i,j=1,2}$, then we have:

$$
\begin{aligned}
&P(\boldsymbol{A}^j \circ \boldsymbol{\epsilon}_{t-j} = \mathbf{0}) = P(p_{11} \circ \epsilon_{1,t-j} + p_{12} \circ \epsilon_{2,t-j} = 0, p_{21} \circ \epsilon_{1,t-j} + p_{22} \circ \epsilon_{2,t-j} = 0) \\
&= \textstyle\sum_{k=0}^{\infty} \sum_{s=0}^{\infty} (1 - p_{11})^k (1 - p_{12})^s (1 - p_{21})^k (1 - p_{22})^s P(\epsilon_{1,t-j} = k, \epsilon_{2,t-j} = s) > 0,
\end{aligned}
$$

thus, the second part, (VI), is also positive. Therefore, $P^k_{\mathbf{0},\mathbf{0}} > 0$, with probability one, i.e., $\{\boldsymbol{X}_t\}$, is aperiodic. Hence, $\{\boldsymbol{X}_t\}$ is an ergodic Markov chain. □

Note that $E(\boldsymbol{X}_t^{(n)}) = (\boldsymbol{I} - \boldsymbol{A} - \boldsymbol{B})^{-1}\boldsymbol{C} < \infty$ and

$$\begin{aligned} E\left(\boldsymbol{X}_t^{(n)}\boldsymbol{X}_t^{(n)\top}\right) &= \boldsymbol{\Gamma} E\left(\boldsymbol{X}_t^{(n-1)}\boldsymbol{X}_t^{(n-1)\top}\right)\boldsymbol{\Gamma}^\top + \boldsymbol{\Psi} \\ &= \cdots = \boldsymbol{\Gamma}^n E\left(\boldsymbol{X}_t^{(0)}\boldsymbol{X}_t^{(0)\top}\right)(\boldsymbol{\Gamma}^n)^\top + \boldsymbol{\Gamma}^{n-1}\boldsymbol{\Psi}(\boldsymbol{\Gamma}^{n-1})^\top + \cdots + \boldsymbol{\Gamma}\boldsymbol{\Psi}\boldsymbol{\Gamma}^\top + \boldsymbol{\Psi}, \end{aligned}$$

where $\boldsymbol{\Psi}$ involves the first moments of $\boldsymbol{X}_t^{(n)}$ and $\boldsymbol{R}_t$. Hence, the first two moments of $\boldsymbol{X}_t^{(n)}$ are finite. Thus, $\{\boldsymbol{X}_t^{(n)}\}$ is stationary and ergodic by Theorem 2, Theorem 3 and Shumway and Stoffer [23].

Theorem 4. *If the conditions of Theorem 1 hold, the first two moments and covariance matrix of $\{\boldsymbol{X}_t\}$ exist and*

(1). $E(\boldsymbol{X}_t|\boldsymbol{X}_{t-1}) = (\boldsymbol{A} + \boldsymbol{B})\boldsymbol{X}_{t-1} + \boldsymbol{C}$;

(2). $E(\boldsymbol{X}_t) = (\boldsymbol{I} - \boldsymbol{A} - \boldsymbol{B})^{-1}\boldsymbol{C}$, if $(\boldsymbol{I} - \boldsymbol{A} - \boldsymbol{B})^{-1}$ exists, where $\boldsymbol{I}$ denotes the identity matrix;

(3). $R(k) = \mathrm{Cov}(\boldsymbol{X}_{t+k}, \boldsymbol{X}_t) = (\boldsymbol{A} + \boldsymbol{B})^k R(0)$, $k = 1, 2, \cdots$.

In addition, if $k = 0$, $R(0) = \boldsymbol{A}R(0)\boldsymbol{A}^\top + \boldsymbol{H}^* + \boldsymbol{A}R(0)\boldsymbol{B}^\top + \boldsymbol{B}R(0)\boldsymbol{A}^\top + \Sigma$, where $\boldsymbol{H}^* = \mathrm{diag}(\sum_{j=1}^{M} \boldsymbol{V}_{ij} E(X_{j,t-1}))$, $\boldsymbol{V}_{ij} = \alpha_{ij}(1 - \alpha_{ij})$ and $\Sigma = \mathrm{Cov}(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t)$. Specifically, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are diagonal matrices, $R(0) = (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^\top - 2\boldsymbol{A}\boldsymbol{B}^\top)^{-1}\Sigma + \boldsymbol{H}^*$.

Proof. (1) and (2) are easy to prove by the moment property of $\boldsymbol{A}\circ$, and we omit them. Here, we only give the proof of (3):

$$\begin{aligned} R(k) &= \mathrm{Cov}(\boldsymbol{A} \circ \boldsymbol{X}_{t+k-1} + \boldsymbol{\epsilon}_{t+k}, \boldsymbol{X}_t) = \mathrm{Cov}(\boldsymbol{A} \circ \boldsymbol{X}_{t+k-1}, \boldsymbol{X}_t) + \mathrm{Cov}(\boldsymbol{\epsilon}_{t+k}, \boldsymbol{X}_t) \\ &= \boldsymbol{A}\mathrm{Cov}(\boldsymbol{X}_{t+k-1}, \boldsymbol{X}_t) + \mathrm{Cov}(\boldsymbol{B}\boldsymbol{X}_{t+k-1} + \boldsymbol{C}, \boldsymbol{X}_t) = (\boldsymbol{A} + \boldsymbol{B})\mathrm{Cov}(\boldsymbol{X}_{t+k-1}, \boldsymbol{X}_t) \\ &= (\boldsymbol{A} + \boldsymbol{B})\boldsymbol{R}(k-1) = \cdots = (\boldsymbol{A} + \boldsymbol{B})^k R(0). \end{aligned}$$

In fact, $\mathrm{Cov}(\boldsymbol{X}_{t-1}, \boldsymbol{\epsilon}_t) = \mathrm{Cov}(\boldsymbol{X}_{t-1}, \boldsymbol{B}\boldsymbol{X}_{t-1} + \boldsymbol{C}) = \mathrm{Cov}(\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-1})\boldsymbol{B}^\top$ and $\mathrm{Cov}(\boldsymbol{\epsilon}_t, \boldsymbol{X}_{t-1}) = \mathrm{Cov}(\boldsymbol{B}\boldsymbol{X}_{t-1} + \boldsymbol{C}, \boldsymbol{X}_{t-1}) = \boldsymbol{B}\mathrm{Cov}(\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-1})$. Hence,

$$\begin{aligned} R(0) &= \mathrm{Cov}(\boldsymbol{X}_t, \boldsymbol{X}_t) = \mathrm{Cov}(\boldsymbol{A} \circ \boldsymbol{X}_{t-1} + \boldsymbol{\epsilon}_t, \boldsymbol{A} \circ \boldsymbol{X}_{t-1} + \boldsymbol{\epsilon}_t) \\ &= \boldsymbol{A}\mathrm{Cov}(\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-1})\boldsymbol{A}^\top + \boldsymbol{H}^* + \boldsymbol{A}\mathrm{Cov}(\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-1})\boldsymbol{B}^\top + \boldsymbol{B}\mathrm{Cov}(\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-1})\boldsymbol{A}^\top + \Sigma \\ &= \boldsymbol{A}R(0)\boldsymbol{A}^\top + \boldsymbol{H}^* + \boldsymbol{A}R(0)\boldsymbol{B}^\top + \boldsymbol{B}R(0)\boldsymbol{A}^\top + \Sigma, \end{aligned}$$

where $\boldsymbol{H}^* = \mathrm{diag}(\sum_{j=1}^{2} \boldsymbol{V}_{ij} E(X_{j,t-1}))$. Let λ, λ_1 and λ_2 be the largest eigenvalues of $\boldsymbol{A}\boldsymbol{A}^\top + 2\boldsymbol{A}\boldsymbol{B}^\top$, $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are diagonal matrices,

$$|\lambda| \le |\lambda_1^2 + 2\lambda_1\lambda_2| \le |\lambda_1(\lambda_1 + \lambda_2) + \lambda_1\lambda_2| \le \lambda_1|(\lambda_1 + \lambda_2)| + \lambda_2|\lambda_1| \le \lambda_1 + \lambda_2 < 1,$$

then $\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^\top - 2\boldsymbol{A}\boldsymbol{B}^\top$ is a nonsingular matrix. Hence, $R(0)$ is obtained. □

Theorem 5. *If the conditions of Theorem 1 hold, the first two moments and covariance matrices of $\{\boldsymbol{\epsilon}_t\}$ exist and:*

(1). $E(\boldsymbol{\epsilon}_t|\boldsymbol{X}_{t-1}) = \boldsymbol{B}\boldsymbol{X}_{t-1} + \boldsymbol{C}$;

(2). $E(\boldsymbol{\epsilon}_t) = (\boldsymbol{I} - \boldsymbol{A})(\boldsymbol{I} - \boldsymbol{A} - \boldsymbol{B})^{-1}\boldsymbol{C}$, if $(\boldsymbol{I} - \boldsymbol{A} - \boldsymbol{B})^{-1}$ exists;

(3). $R_{\boldsymbol{\epsilon}}(k) = \mathrm{Cov}(\boldsymbol{\epsilon}_{t+k}, \boldsymbol{\epsilon}_t) = \boldsymbol{B}(\boldsymbol{A} + \boldsymbol{B})^k R(0)\boldsymbol{B}^\top$, $k = 0, 1, 2, \cdots$.

Proof. (1) is easy to prove by the distribution of $\boldsymbol{\epsilon}_t$. We only need to prove (2) and (3).

(2). $E(\boldsymbol{\epsilon}_t) = \boldsymbol{B}\boldsymbol{\mu}_t + \boldsymbol{C} = \boldsymbol{B}(\boldsymbol{I} - \boldsymbol{A} - \boldsymbol{B})^{-1}\boldsymbol{C} + (\boldsymbol{I} - \boldsymbol{A} - \boldsymbol{B})(\boldsymbol{I} - \boldsymbol{A} - \boldsymbol{B})^{-1}\boldsymbol{C} = (\boldsymbol{I} - \boldsymbol{A})E(\boldsymbol{X}_t)$ by the definition of $\boldsymbol{\epsilon}_t$. $E(\boldsymbol{\epsilon}_t)$ can be obtained directly by (6).

(3). According to the construction of the EBINAR(1) model, we have:

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{X}_t,\boldsymbol{\epsilon}_t) &= \mathrm{Cov}(\boldsymbol{A}\circ\boldsymbol{X}_{t-1},\boldsymbol{B}\boldsymbol{X}_{t-1}+\boldsymbol{C})+\mathrm{Cov}(\boldsymbol{B}\boldsymbol{X}_{t-1}+\boldsymbol{C},\boldsymbol{B}\boldsymbol{X}_{t-1}+\boldsymbol{C})\\ &= \boldsymbol{A}\mathrm{Cov}(\boldsymbol{X}_{t-1},\boldsymbol{X}_{t-1})\boldsymbol{B}^\top+\boldsymbol{B}\mathrm{Cov}(\boldsymbol{X}_{t-1},\boldsymbol{X}_{t-1})\boldsymbol{B}^\top=(\boldsymbol{A}+\boldsymbol{B})R(0)\boldsymbol{B}^\top, \qquad (16)\end{aligned}$$

$$\begin{aligned}R_\epsilon(k) &= \mathrm{Cov}(\boldsymbol{B}\boldsymbol{X}_{t+k-1}+\boldsymbol{C},\boldsymbol{\epsilon}_t)=\boldsymbol{B}\mathrm{Cov}(\boldsymbol{X}_{t+k-1},\boldsymbol{\epsilon}_t)\\ &= \boldsymbol{B}\mathrm{Cov}(\boldsymbol{A}\circ\boldsymbol{X}_{t+k-2}+\boldsymbol{\epsilon}_{t+k-1},\boldsymbol{\epsilon}_t)=\boldsymbol{B}\boldsymbol{A}\mathrm{Cov}(\boldsymbol{X}_{t+k-2},\boldsymbol{\epsilon}_t)+\boldsymbol{B}\mathrm{Cov}(\boldsymbol{\epsilon}_{t+k-1},\boldsymbol{\epsilon}_t)\\ &= \boldsymbol{B}\boldsymbol{A}\mathrm{Cov}(\boldsymbol{X}_{t+k-2},\boldsymbol{\epsilon}_t)+\boldsymbol{B}\mathrm{Cov}(\boldsymbol{B}\boldsymbol{X}_{t+k-2}+\boldsymbol{C},\boldsymbol{\epsilon}_t)\\ &= \boldsymbol{B}\boldsymbol{A}\mathrm{Cov}(\boldsymbol{X}_{t+k-2},\boldsymbol{\epsilon}_t)+\boldsymbol{B}^2\mathrm{Cov}(\boldsymbol{X}_{t+k-2},\boldsymbol{\epsilon}_t)=\boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})\mathrm{Cov}(\boldsymbol{X}_{t+k-2},\boldsymbol{\epsilon}_t)\\ &= \cdots=\boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{k-1}\mathrm{Cov}(\boldsymbol{X}_t,\boldsymbol{\epsilon}_t), \qquad (17)\end{aligned}$$

then $R_\epsilon(k)$ is achieved by substituting (16) into (17), i.e., $R_\epsilon(k)=\boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^kR(0)\boldsymbol{B}^\top$. Note that $\mathrm{Cov}(\boldsymbol{\epsilon}_t,\boldsymbol{\epsilon}_t)=\mathrm{Cov}(\boldsymbol{B}\boldsymbol{X}_{t-1}+\boldsymbol{C},\boldsymbol{B}\boldsymbol{X}_{t-1}+\boldsymbol{C})=\boldsymbol{B}R(0)\boldsymbol{B}^\top$, i.e., the formula holds for $k=0$. □

Theorem 6. *For any fixed positive integer k, it is a necessary and sufficient condition that $E(X_{it})^k<\infty$ is $\gamma<1$, $i=1,2$.*

Proof. For convenience, let $\boldsymbol{A}$ and $\boldsymbol{B}$ be diagonal matrices.

Necessity. According to Lemma 2.1 of Silva and Oliveira [24], $E[(\alpha_{11}\circ X_{1,t-1})^i(\epsilon_{1t})^{k-i}]=\alpha_{11}^ib_{11}^{k-i}E(X_{1,t-1})^k+\psi_1$, where $\psi_1=\psi_1(X_{1,t-1})$ involves the moments of $X_{1,t-1}$ of order $\le(k-1)$ and $i=0,1,2,\cdots,k$. Then,

$$\begin{aligned}E(X_{1t})^k&=E(\alpha_{11}\circ X_{1,t-1}+\epsilon_{1t})^k=\textstyle\sum_{i=0}^k\binom{k}{i}E[(\alpha_{11}\circ X_{1,t-1})^i(\epsilon_{1t})^{k-i}]\\ &=\textstyle\sum_{i=0}^k\binom{k}{i}\alpha_{11}^ib_{11}^{k-i}E(X_{1,t-1})^k+\psi_1=(\alpha_{11}+b_{11})^kE(X_{1,t-1})^k+\psi_1. \qquad (18)\end{aligned}$$

Thus, $E(X_{1t})^k=\dfrac{\psi_1}{1-(\alpha_{11}+b_{11})^k}$ by (18). Hence, $1-(\alpha_{11}+b_{11})^k>0$ if $E(X_{1t})^k<\infty$, i.e., $\alpha_{11}+b_{11}<1$. Similarly, $\alpha_{22}+b_{22}<1$ if $E(X_{2t})^k<\infty$. Hence, $\gamma<1$ if $E(X_{it})^k<\infty$, $i=1,2$.

Sufficiency. We know that $E(X_{it})^k<\infty$ holds for $k=1,2$ by Theorems 4 and 5. The sufficient condition can be proved by induction with respect to k. Now suppose that $E(X_{it})^{k-1}<\infty$, $k\ge3$. According to (13), we define

$$X_{1t}^{(n)}=\begin{cases}0, & n<0;\\ \delta_{1t}, & n=0;\\ \delta_{1t}+\sum_{j=1}^{X_{1,t-1}^{(n-1)}}W_{t,j}, & n>0\end{cases}\quad\text{and}\quad X_{2t}^{(n)}=\begin{cases}0, & n<0;\\ \delta_{2t}, & n=0;\\ \delta_{2t}+\sum_{s=1}^{X_{2,t-1}^{(n-1)}}V_{t,s}, & n>0,\end{cases}$$

where $W_{t,j}$, δ_{1t}, $V_{t,s}$ and δ_{2t} are independent of each other and each of them is independent and identically distributed, i.e., $W_{t,j}\sim\mathrm{Bin}(1,\alpha_{11})+\mathrm{Poi}(b_{11})$, $\delta_{1t}\sim\mathrm{Poi}(c_1)$, $V_{t,s}\sim\mathrm{Bin}(1,\alpha_{22})+\mathrm{Pois}(b_{22})$ and $\delta_{2t}\sim\mathrm{Pois}(c_2)$. Using the univariate binomial thinning operator, $X_{1t}^{(n)}$ and $X_{2t}^{(n)}$ admit the representations:

$$X_{1t}^{(n)}=\delta_{1t}+(\alpha_{11}\circ X_{1,t-1}^{(n-1)}+Z_{1t}), \qquad (19)$$

$$X_{2t}^{(n)}=\delta_{2t}+(\alpha_{22}\circ X_{2,t-1}^{(n-1)}+Z_{2t}), \qquad (20)$$

where $Z_{1t} \sim \text{Poi}(b_{11}X_{1,t-1}^{(n-1)})$ and $Z_{2t} \sim \text{Poisson}(b_{22}X_{2,t-1}^{(n-1)})$. It is easy to see both $\{X_{1t}^{(n)}\}_{n\in\mathbb{N}}$ and $\{X_{2t}^{(n)}\}_{n\in\mathbb{N}}$ are non-decreasing. According to Lemma 2.1 of [24], we have:

$$E(\alpha_{11} \circ X_{1,t-1}^{(n-1)} + Z_{1t})^k = (\alpha_{11} + b_{11})^k E(X_{1,t-1}^{(n-1)})^k + \psi_2 \le (\alpha_{11} + b_{11})^k E(X_{1,t-1}^{(n)})^k + \psi_4,$$
$$E(\alpha_{22} \circ X_{2,t-1}^{(n-1)} + Z_{2t})^k = (\alpha_{22} + b_{22})^k E(X_{2,t-1}^{(n-1)})^k + \psi_3 \le (\alpha_{22} + b_{22})^k E(X_{2,t-1}^{(n)})^k + \psi_5,$$

where $\psi_2 = \psi_2(X_{1,t-1}^{(n-1)})$ and $\psi_3 = \psi_3(X_{2,t-1}^{(n-1)})$ involve the moments of $X_{1,t-1}^{(n-1)}$ and $X_{2,t-1}^{(n-1)}$ of order $\le (k-1)$, and $\psi_4 = \psi_4(X_{1,t-1}^{(n)})$ and $\psi_5 = \psi_5(X_{2,t-1}^{(n)})$ involve the moments of $X_{1,t-1}^{(n)}$ and $X_{2,t-1}^{(n)}$ of order $\le (k-1)$, respectively. According to (19) and (20), we obtain:

$$\begin{aligned}
E(X_{1t}^{(n)})^k &= E(\delta_{1t})^k + (\alpha_{11} \circ X_{1,t-1}^{(n-1)} + Z_{1t})^k + \sum_{j=1}^{k-1} \binom{k}{j} E(\delta_{1t})^{k-j} E(\alpha_{11} \circ X_{1,t-1}^{(n-1)} + Z_{1t})^j \\
&\le E(\delta_{1t})^k + (\alpha_{11} + b_{11})^k E(X_{1,t-1}^{(n)})^k + \psi_4 + \sum_{j=1}^{k-1} \binom{k}{j} E(\delta_{1t})^{k-j} E(\alpha_{11} \circ X_{1,t-1}^{(n)} + Z_{1t})^j \\
&\le c_1^k + \gamma^k E(X_{1,t-1}^{(n)})^k + \psi_4 + \sum_{j=1}^{k-1} \binom{k}{j} E(\delta_{1t})^{k-j} E(\alpha_{11} \circ X_{1,t-1}^{(n)} + Z_{1t})^j, \qquad (21) \\
E(X_{2t}^{(n)})^k &= E(\delta_{2t})^k + (\alpha_{22} \circ X_{2,t-1}^{(n-1)} + Z_{2t})^k + \sum_{j=1}^{k-1} \binom{k}{j} E(\delta_{2t})^{k-j} E(\alpha_{22} \circ X_{2,t-1}^{(n-1)} + Z_{2t})^j \\
&\le E(\delta_{2t})^k + (\alpha_{22} + b_{22})^k E(X_{2,t-1}^{(n)})^k + \psi_5 + \sum_{j=1}^{k-1} \binom{k}{j} E(\delta_{2t})^{k-j} E(\alpha_{22} \circ X_{2,t-1}^{(n)} + Z_{2t})^j \\
&\le c_2^k + \gamma^k E(X_{2,t-1}^{(n)})^k + \psi_5 + \sum_{j=1}^{k-1} \binom{k}{j} E(\delta_{2t})^{k-j} E(\alpha_{22} \circ X_{2,t-1}^{(n)} + Z_{2t})^j. \qquad (22)
\end{aligned}$$

Using (21) and (22),

$$E(X_{1t}^{(n)})^k + E(X_{2t}^{(n)})^k \le \frac{\sum_{i=1}^{2} \left[c_i^k + \sum_{j=1}^{k-1} \binom{k}{j} E(\delta_{it})^{k-j} E(\alpha_{ii} \circ X_{i,t-1}^{(n)} + Z_{it})^j \right] + \psi_6}{1-\gamma^k}, \qquad (23)$$

where $\psi_6 = \psi_4 + \psi_5$. Note that the numerator in (23) involves the moments of $X_{1,t-1}^{(n)}$ and $X_{2,t-1}^{(n)}$ of order $\le k-1$ and is finite; thus, $E(X_{1t}^{(n)})^k + E(X_{2t}^{(n)})^k$ is finite if $\gamma < 1$. In addition that both $E(X_{1t}^{(n)})$ and $E(X_{2t}^{(n)})$ are non-negative; thus, $E(X_{1t}^{(n)})$ and $E(X_{2t}^{(n)})$ are finite. $\square$

3. Parameter Estimation

In this section, we consider the conditional maximum likelihood estimation for model (6). Let $\boldsymbol{\theta} = \left(\alpha_{ij}, b_{ij}, c_i, \phi\right)^\top, i, j = 1, 2$. Suppose that $\boldsymbol{X}_0, \boldsymbol{X}_1, \cdots, \boldsymbol{X}_T$ are generated by the EBINAR(1) model with the true parameter value $\boldsymbol{\theta}_0$.

By (11), the conditional log-likelihood function can be written as:

$$\ell(\boldsymbol{\theta}) = \textstyle\sum_{t=1}^{T} \ln P_{\boldsymbol{\theta}}(\boldsymbol{X}_t | \boldsymbol{X}_{t-1}), \qquad (24)$$

where

$$P_{\boldsymbol{\theta}}(\mathbf{X}_t|\mathbf{X}_{t-1}) = P(X_{1t} = X_{1t}, X_{2t} = X_{2t} | X_{1,t-1} = X_{1,t-1}, X_{2,t-1} = X_{2,t-1}) = \sum_{k_1=0}^{g_1} \sum_{k_2=0}^{g_2}$$
$$\Big(h(k_1, X_{1,t-1}, X_{2,t-1}, \alpha_{11}, \alpha_{12}) h(k_2, X_{1,t-1}, X_{2,t-1}, \alpha_{21}, \alpha_{22}) f(X_{1t} - k_1, X_{2t} - k_2, \lambda_{1t}, \lambda_{2t}, \phi) \Big)$$

with $\lambda_{1t} = b_{11}X_{1,t-1} + b_{12}X_{2,t-1} + c_1$, $\lambda_{2t} = b_{21}X_{1,t-1} + b_{22}X_{2,t-1} + c_2$, $g_1 = \min(X_{1t}, X_{1,t-1})$, $g_2 = \min(X_{2t}, X_{2,t-1})$, $f()$ and $h()$ are given in (2) and (8), respectively.

By using (3)–(5), and (9) and (10), we can derive the score equation:

$$\frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{T} \frac{1}{P_{\boldsymbol{\theta}_0}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})} \frac{\partial P_{\boldsymbol{\theta}_0}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \tag{25}$$

where

$$\begin{aligned}
\frac{\partial P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{\partial \alpha_{11}} &= \sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} X_{1,t-1} h(k_2, X_{1,t-1}, X_{2,t-1}, \alpha_{21}, \alpha_{22}) f(X_{1t}-k_1, X_{2t}-k_2, \lambda_{1t}, \lambda_{2t}, \phi) \\
&\quad \times \big(h(k_1-1, X_{1,t-1}-1, X_{2,t-1}, \alpha_{11}, \alpha_{12}) - h(k_1, X_{1,t-1}-1, X_{2,t-1}, \alpha_{11}, \alpha_{12})\big), \\
\frac{\partial P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{\partial \alpha_{12}} &= \sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} X_{2,t-1} h(k_2, X_{1,t-1}, X_{2,t-1}, \alpha_{21}, \alpha_{22}) f(X_{1t}-k_1, X_{2t}-k_2, \lambda_{1t}, \lambda_{2t}, \phi) \\
&\quad \times \big(h(k_1-1, X_{1,t-1}, X_{2,t-1}-1, \alpha_{11}, \alpha_{12}) - h(k_1, X_{1,t-1}, X_{2,t-1}-1, \alpha_{11}, \alpha_{12})\big), \\
\frac{\partial P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{\partial \alpha_{21}} &= \sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} X_{1,t-1} h(k_1, X_{1,t-1}, X_{2,t-1}, \alpha_{11}, \alpha_{12}) f(X_{1t}-k_1, X_{2t}-k_2, \lambda_{1t}, \lambda_{2t}, \phi) \\
&\quad \times \big(h(k_2-1, X_{1,t-1}-1, X_{2,t-1}, \alpha_{21}, \alpha_{22}) - h(k_2, X_{1,t-1}-1, X_{2,t-1}, \alpha_{21}, \alpha_{22})\big), \\
\frac{\partial P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{\partial \alpha_{22}} &= \sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} X_{2,t-1} h(k_1, X_{1,t-1}, X_{2,t-1}, \alpha_{11}, \alpha_{12}) f(X_{1t}-k_1, X_{2t}-k_2, \lambda_{1t}, \lambda_{2t}, \phi) \\
&\quad \times \big(h(k_2-1, X_{1,t-1}, X_{2,t-1}-1, \alpha_{21}, \alpha_{22}) - h(k_2, X_{1,t-1}, X_{2,t-1}-1, \alpha_{21}, \alpha_{22})\big), \\
\frac{\partial P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{\partial b_{11}} &= \sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} h(k_1, X_{1,t-1}, X_{2,t-1}, \alpha_{11}, \alpha_{12}) h(k_2, X_{1,t-1}, X_{2,t-1}, \alpha_{21}, \alpha_{22}) \\
&\quad \times X_{1,t-1} f(X_{1t}-k_1-1, X_{2t}-k_2, \lambda_{1t}, \lambda_{2t}, \phi), \\
\frac{\partial P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{\partial b_{12}} &= \sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} h(k_1, X_{1,t-1}, X_{2,t-1}, \alpha_{11}, \alpha_{12}) h(k_2, X_{1,t-1}, X_{2,t-1}, \alpha_{21}, \alpha_{22}) \\
&\quad \times X_{2,t-1} f(X_{1t}-k_1-1, X_{2t}-k_2, \lambda_{1t}, \lambda_{2t}, \phi), \\
\frac{\partial P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{\partial c_{1}} &= \sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} h(k_1, X_{1,t-1}, X_{2,t-1}, \alpha_{11}, \alpha_{12}) h(k_2, X_{1,t-1}, X_{2,t-1}, \alpha_{21}, \alpha_{22}) \\
&\quad \times f(X_{1t}-k_1-1, X_{2t}-k_2, \lambda_{1t}, \lambda_{2t}, \phi), \\
\frac{\partial P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{\partial b_{21}} &= \sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} h(k_1, X_{1,t-1}, X_{2,t-1}, \alpha_{11}, \alpha_{12}) h(k_2, X_{1,t-1}, X_{2,t-1}, \alpha_{21}, \alpha_{22}) \\
&\quad \times X_{1,t-1} f(X_{1t}-k_1, X_{2t}-k_2-1, \lambda_{1t}, \lambda_{2t}, \phi), \\
\frac{\partial P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{\partial b_{22}} &= \sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} h(k_1, X_{1,t-1}, X_{2,t-1}, \alpha_{11}, \alpha_{12}) h(k_2, X_{1,t-1}, X_{2,t-1}, \alpha_{21}, \alpha_{22}) \\
&\quad \times X_{2,t-1} f(X_{1t}-k_1-1, X_{2t}-k_2-1, \lambda_{1t}, \lambda_{2t}, \phi), \\
\frac{\partial P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{\partial c_{2}} &= \sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} h(k_1, X_{1,t-1}, X_{2,t-1}, \alpha_{11}, \alpha_{12}) h(k_2, X_{1,t-1}, X_{2,t-1}, \alpha_{21}, \alpha_{22}) \\
&\quad \times f(X_{1t}-k_1, X_{2t}-k_2-1, \lambda_{1t}, \lambda_{2t}, \phi),
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{\partial \phi} = &\sum_{k_1=0}^{g_1}\sum_{k_2=0}^{g_2} h(k_1, X_{1,t-1}, X_{2,t-1}, \alpha_{11}, \alpha_{12}) h(k_2, X_{1,t-1}, X_{2,t-1}, \alpha_{21}, \alpha_{22}) \\
&\times \Big(f(X_{1t}-k_1, X_{2t}-k_2, \lambda_{1t}, \lambda_{2t}, \phi) - f(X_{1t}-k_1-1, X_{2t}-k_2, \lambda_{1t}, \lambda_{2t}, \phi) \\
&- f(X_{1t}-k_1, X_{2t}-k_2-1, \lambda_{1t}, \lambda_{2t}, \phi) + f(X_{1t}-k_1-1, X_{2t}-k_2-1, \lambda_{1t}, \lambda_{2t}, \phi) \Big).
\end{aligned}
$$

The maximizer $\hat{\boldsymbol{\theta}}_T$ of (24) is the CML estimate of $\boldsymbol{\theta}_0$, which is obtained by numerically maximizing the log-likelihood (24) or by solving the score Equation (25). To study the asymptotic behaviour of the estimator, we make the following two Assumptions about the parameter space and the underlying process.

Assumption 1. *The parametric space Θ is compact with $\Theta = \{\boldsymbol{\theta}, \boldsymbol{\theta} = \{\alpha_{ij}, b_{ij}, c_1, c_2, \phi\}^\top,\ i,j = 1,2\}$, where $\underline{\delta} \le \alpha_{ij}, b_{ij} \le \bar{\delta}$, $\underline{c} \le c_i \le \bar{c}$, $\underline{\phi} \le \phi \le \bar{\phi}$, $\gamma = \max(\alpha_{ij} + b_{ij}) < 1$, $\underline{\delta}$, $\bar{\delta}$, $\underline{c}$, $\bar{c}$, $\underline{\phi}$ and $\bar{\phi}$ are finite positive constants, and $\boldsymbol{\theta}_0$ is an interior point in Θ.*

Assumption 2. *If there exists a $t \ge 1$, such that $\boldsymbol{X}_t(\boldsymbol{\theta}_0) = \boldsymbol{X}_t(\boldsymbol{\theta})$, $P_{\boldsymbol{\theta}_0}$ a.s., then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, where $P_{\boldsymbol{\theta}_0}$ is the probability measure under the true parameter $\boldsymbol{\theta}_0$.*

To derive the identification of the EBINAR(1) model, we give the following two Lemmas.

Lemma 1. *Let $g_1(x,y,b_{11},b_{12},c_1) = b_{11}x + b_{12}y + c_1$, $b_{11}, b_{12}, c_1 > 0$ for $(x,y) \in \mathbb{R}^+ \times \mathbb{R}^+$. Then, if $g_1(x,y,b_{11},b_{12},c_1) = g_1(x,y,b_{11}^0,b_{12}^0,c_1^0)$, then $b_{11} = b_{11}^0, b_{12} = b_{12}^0, c_1 = c_1^0$.*

Proof. By the assumption:

$$
\begin{aligned}
\frac{\partial g_1(x,y,b_{11},b_{12},c_1)}{\partial x} &= \frac{\partial g_1(x,y,b_{11}^0,b_{12}^0,c_1^0)}{\partial x}, \\
\frac{\partial g_1(x,y,b_{11},b_{12},c_1)}{\partial y} &= \frac{\partial g_1(x,y,b_{11}^0,b_{12}^0,c_1^0)}{\partial y}, \\
g_1(0,0,b_{11},b_{12},c_1) &= g_1(0,0,b_{11}^0,b_{12}^0,c_1^0).
\end{aligned}
$$

we obtain: $b_{11} = b_{11}^0, b_{12} = b_{12}^0, c_1 = c_1^0$. □

Similarly, we denote $g_2(x,y,b_{21},b_{22},c_2) = b_{21}x + b_{22}y + c_2$. If $g_2(x,y,b_{21},b_{22},c_2) = g_2(x,y,b_{21}^0,b_{22}^0,c_2^0)$, then we have $b_{21} = b_{21}^0, b_{22} = b_{22}^0, c_2 = c_2^0$ by the same method.

Lemma 2. *If $\{\boldsymbol{X}_t\}$ is the strictly stationary and ergodic solution of model (6), Assumptions 1 and 2 hold, then model (6) is identifiable.*

Proof. According to Lemma 1, we conclude that if $\lambda_{1t}(b_{11},b_{12},c_1) = \lambda_{1t}(b_{11}^0,b_{12}^0,c_1^0)$, then $b_{11} = b_{11}^0, b_{12} = b_{12}^0, c_1 = c_1^0$. Similarly, if $\lambda_{2t}(b_{21},b_{22},c_2) = \lambda_{2t}(b_{21}^0,b_{22}^0,c_2^0)$, then $b_{21} = b_{21}^0, b_{22} = b_{22}^0, c_2 = c_2^0$. Thus, if $\boldsymbol{\epsilon}_t(b_{ij},c_i,\phi) = \boldsymbol{\epsilon}_t(b_{ij}^0,c_i^0,\phi^0)$, $t \ge 1, i,j = 1,2$, then $b_{ij} = b_{ij}^0$, $c_i = c_i^0$, $\phi = \phi^0$. According to (7), we have $\epsilon_{it} = X_{it} - \alpha_{i1} \circ X_{1,t-1} - \alpha_{i2} \circ X_{2,t-1}$, $i = 1,2$. If $\epsilon_{it}(\boldsymbol{\theta}) = \epsilon_{it}(\boldsymbol{\theta}_0)$, then we have $\alpha_{i1} = \alpha_{i1}^0, \alpha_{i2} = \alpha_{i2}^0$, otherwise

$$
0 = E(\epsilon_{it}(\boldsymbol{\theta})) - E(\epsilon_{it}(\boldsymbol{\theta}_0)) = (\alpha_{i1} - \alpha_{i1}^0)E(X_{1t}) + (\alpha_{i2} - \alpha_{i2}^0)E(X_{2t}),
$$

then $E(X_{1t}) = 0$ and $E(X_{2t}) = 0$, which contradicts the fact that $E(X_{it}) > 0$, $i = 1,2$.

By Assumption 2, for given $X_{1,t-1}$ and $X_{2,t-1}$, we have

$$
\phi = \mathrm{Cov}(X_{1t}(\boldsymbol{\theta}), X_{2t}(\boldsymbol{\theta})) = \mathrm{Cov}(X_{1t}(\boldsymbol{\theta}_0), X_{2t}(\boldsymbol{\theta}_0)) = \phi^0.
$$

Thus, $\phi = \phi^0$. Hence, model (6) is identifiable. $\square$

Theorem 7. *Suppose that $\{\boldsymbol{X}_t\}$ is the strictly stationary and ergodic solution of model (6) and Assumptions 1 and 2 hold. As $T \to \infty$, there exists an estimator $\hat{\boldsymbol{\theta}}_T$ such that $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0$.*

Proof. To prove the strong consistence of $\hat{\boldsymbol{\theta}}_T$, we need to check all the assumptions given in Theorems 4.1.2 and 4.1.3 in Amemiya [25]. Let $W_t(\boldsymbol{\theta}) = \ln P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})$, then $\ell(\boldsymbol{\theta}) = \sum_{t=1}^{T} W_t(\boldsymbol{\theta})$. We observe that $W_t(\boldsymbol{\theta})$ is a measurable function of $\boldsymbol{X}_t$ for all $\boldsymbol{\theta} \in \Theta$, and is continuous in an open and convex neighborhood $N(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$, then there at least exists a point $\overline{\boldsymbol{\theta}} \in N(\boldsymbol{\theta}_0)$ such that $W_t(\boldsymbol{\theta})$ attains the maximum value at $\overline{\boldsymbol{\theta}}$.

Thus,

$$E\left(\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} W_t(\boldsymbol{\theta})\right) = E\big(\ln P_{\overline{\boldsymbol{\theta}}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})\big) \le \ln E\big(P_{\overline{\boldsymbol{\theta}}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})\big) < \infty.$$

Note that $\{\boldsymbol{X}_t\}$ is a stationary and ergodic time series, and in terms of Theorem 4.1.2 in Amemiya [25], $\frac{1}{T}\sum_{t=1}^{T} W_t(\boldsymbol{\theta}) \to EW_t(\boldsymbol{\theta})$ in probability as $T \to \infty$. By Jensen's inequality, we have:

$$E(W_t(\boldsymbol{\theta})) - E(W_t(\boldsymbol{\theta}_0)) = E \ln \frac{P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{P_{\boldsymbol{\theta}_0}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})} \le \ln E \frac{P_{\boldsymbol{\theta}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{P_{\boldsymbol{\theta}_0}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})} = 0. \tag{26}$$

Thus, $EW_t(\boldsymbol{\theta})$ attains a strict local maximum at $\boldsymbol{\theta}_0$ by (26) and Lemma 2. Hence, the conditions of Theorem 4.1.2 in Amemiya [25] are fulfilled; thus, there exists an estimator $\hat{\boldsymbol{\theta}}_T$ such that $\hat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta}_0$, $T \to \infty$. $\square$

Theorem 8. *If the conditions of Theorem 7 hold, as $T \to \infty$,*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\Big(\boldsymbol{0}, (\boldsymbol{J}(\boldsymbol{\theta}_0))^{-1}\boldsymbol{I}(\boldsymbol{\theta}_0)(\boldsymbol{J}(\boldsymbol{\theta}_0))^{-1}\Big),$$

where $\boldsymbol{I}(\boldsymbol{\theta}_0) = \lim_{T\to\infty} T^{-1} E\left(\frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\left(\frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)^{\top}\right)$, $\boldsymbol{J}(\boldsymbol{\theta}_0) = \lim_{T\to\infty} T^{-1} E\left(\frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\top}}\right)$.

Proof. To prove the asymptotic normality of $\hat{\boldsymbol{\theta}}_T$, we need to verify the assumptions of Theorem 4.1.3 in Amemiya [25].

First, by Proposition 1 in Freeland and McCabe [26], it is easy to obtain all the partial derivatives in a similar way, i.e., $\frac{\partial W_t(\boldsymbol{\theta})}{\partial \theta_i}$ exist and are three times continuous differentiable in Θ; thus, $\frac{\partial^2 W_t(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$ exists and is continuous in $N(\boldsymbol{\theta}_0)$, for any $i, j, k = 1, 2, \cdots, 11$. Thus, there at least exists a point $\tilde{\boldsymbol{\theta}} \in N(\boldsymbol{\theta}_0)$ such that $\frac{\partial^2 W_t(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$ attains the maximum value at $\tilde{\boldsymbol{\theta}}$. Hence,

$$E\left(\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \frac{\partial^2 W_t(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right) = E\left(\frac{\partial^2 W_t(\tilde{\boldsymbol{\theta}})}{\partial \theta_i \partial \theta_j}\right) < \infty.$$

For convenience, we denote: $\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\top}} = \boldsymbol{G}(\boldsymbol{X}_t, \boldsymbol{\theta}) = \big(g_{ij}(\boldsymbol{X}_t, \boldsymbol{\theta})\big)$ and $E\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\top}}\right) = \boldsymbol{G}(\boldsymbol{\theta}) = \big(g_{ij}(\boldsymbol{X}_t, \boldsymbol{\theta})\big)$. We only need to prove $g_{ij}(\boldsymbol{X}_t, \boldsymbol{\theta})$ converges to a finite and non-stochastic function $g_{ij}(\boldsymbol{\theta}) = E\big(g_{ij}(\boldsymbol{X}_t, \boldsymbol{\theta})\big)$. Let $h(\boldsymbol{X}_t, \boldsymbol{\theta}) = g_{ij}(\boldsymbol{X}_t, \boldsymbol{\theta}) - E\big[g_{ij}(\boldsymbol{X}_t, \boldsymbol{\theta})\big]$, then $Eh(\boldsymbol{X}_t, \boldsymbol{\theta}) = 0$. Hence, the conditions of Theorem 4.1.3 in [25] are fulfilled. Thus, $T^{-1}\sum_{t=1}^{T} h(\boldsymbol{X}_t, \boldsymbol{\theta})$ converges to 0 in probability uniformly in $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0)$. Furthermore, $T^{-1}\sum_{t=1}^{T} h(\boldsymbol{X}_t, \boldsymbol{\theta}_T^{\star})$ converges to 0 in probability, when $\boldsymbol{\theta}_T^{\star} \to \boldsymbol{\theta}_0$, $T \to \infty$.

Second, it is easy to see $\mathrm{Cov}(\partial W_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}) = E\left[\partial W_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial W_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}^\top\right]$ because $E(\partial W_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}) = \mathbf{0}$.

Using the ergodic theorem,

$$\frac{1}{T}\frac{\partial\ell(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \xrightarrow{p} E\frac{1}{P_{\boldsymbol{\theta}_0}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}\frac{\partial P_{\boldsymbol{\theta}_0}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})}{\partial\boldsymbol{\theta}}.$$

Using the martingale central limit theorem and the Cramér–Wold device, it is direct to show that

$$\frac{1}{\sqrt{T}}\partial\ell(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{I}(\boldsymbol{\theta}_0)) \text{ with } \boldsymbol{I}(\boldsymbol{\theta}_0) = \lim_{T\to\infty} T^{-1}E\left[\partial\ell(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}(\partial\ell(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta})^\top\right].$$

Third, there exists an $H(X_{1t}, X_{2t})$ such that $\left|\frac{\partial^3 \ln \ell(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| \le H(X_{1t}, X_{2t})$ and $E[H(X_{1t}, X_{2t})] < \infty$ by Theorem 4. By the Taylor expansion, we have

$$\frac{\partial\ell(\hat{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}} = \frac{\partial\ell(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \frac{\partial^2\ell(\boldsymbol{\theta}_T^\star)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0), \tag{27}$$

where $\boldsymbol{\theta}_T^\star$ lies in between $\hat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}_0$. We observe that the $\frac{\partial\ell(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = \mathbf{0}$ in (27) by (25), then

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) = \left[\frac{1}{T}\frac{\partial^2\ell(\boldsymbol{\theta}_T^\star)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right]^{-1}\frac{1}{\sqrt{T}}\frac{\partial\ell(\hat{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}}. \tag{28}$$

Hence, the asymptotic normality of $\hat{\boldsymbol{\theta}}_T$ follows from (28). □

4. Simulation

In this section, we conduct a simulation study to illustrate the finite sample property of the CML estimate. The simulation is carried out in `R` by using the `optim` function for the optimization of the conditional log-likelihood function.

In the simulation, we generate data from the non-diagonal EBINAR(1) model and the diagonal EBINAR(1) model. The sizes of samples are chosen to be 50, 100, 200, 500 and 1000 to reflect relatively small, small, moderate, large and relatively large sample sizes, and we use 500 replications. For the simulated sample, performances of the estimators are evaluated by mean squared error (MSE) and mean absolute deviation error (MADE), where $\mathrm{MSE} = \frac{1}{m}\sum_{i=1}^m(\hat{\varphi}_i - \varphi)^2$, $\mathrm{MADE} = \frac{1}{m}\sum_{i=1}^m|\hat{\varphi}_i - \varphi|$, where $\hat{\varphi}_i$ is the estimator of φ in the ith replication and m denotes replication times. The used parameter combinations of $\boldsymbol{\theta} = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, b_{11}, b_{12}, b_{21}, b_{22}, c_1, c_2, \phi)^\top$ are listed as follows:

(1). For a non-diagonal model: $\boldsymbol{\theta} = (0.3, 0.1, 0.1, 0.1, 0.2, 0.1, 0.1, 0.3, 0.6, 0.6, 0.5)^\top$;

(2). For a diagonal model: $\boldsymbol{\theta} = \mathrm{I} : (0.2, 0, 0, 0.3, 0.3, 0, 0, 0.2, 0.6, 0.6, 0.5)^\top$, $\mathrm{II} : (0.2, 0, 0, 0.3, 0.3, 0, 0, 0.2, 2, 2, 1)^\top$, $\mathrm{III} : (0.1, 0, 0, 0.4, 0.4, 0, 0, 0.1, 0.6, 0.6, 0.5)^\top$, $\mathrm{IV} : (0.1, 0, 0, 0.4, 0.4, 0, 0, 0.1, 2, 2, 1)^\top$.

Tables 1–5 show that the MSE and MADE decrease with the increase in T for diagonal and non-diagonal models, which implies that the estimators are consistent.

To illustrate the location and dispersion of the estimates, we present the boxplots of the estimates for the non-diagonal and I of diagonal parameter combinations in Figures 1 and 2; the others are similar.

Table 1. Results for non-diagonal EBINAR(1) model.

Size		α_{11}	α_{12}	α_{21}	α_{22}	b_{11}	b_{12}	b_{21}	b_{22}	c_1	c_2	ϕ
50	MSE	0.0095	0.0122	0.0082	0.0177	0.0048	0.0187	0.0100	0.0380	0.0172	0.0246	0.0160
	MADE	0.0561	0.0537	0.0390	0.0460	0.0371	0.0487	0.0409	0.0750	0.0857	0.0991	0.0811
100	MSE	0.0069	0.0043	0.0033	0.0127	0.0063	0.0070	0.0099	0.0090	0.0128	0.0210	0.0150
	MADE	0.0473	0.0269	0.0287	0.0417	0.0399	0.0355	0.0378	0.0578	0.0763	0.0913	0.0736
200	MSE	0.0044	0.0019	0.0034	0.0108	0.0058	0.0063	0.0054	0.0063	0.0178	0.0185	0.0170
	MADE	0.0421	0.0281	0.0326	0.0291	0.0426	0.0363	0.0372	0.0532	0.0901	0.0876	0.0824
500	MSE	0.0033	0.0008	0.0011	0.0105	0.0044	0.0027	0.0008	0.0061	0.0029	0.0044	0.0063
	MADE	0.0317	0.0161	0.0213	0.0469	0.0379	0.0293	0.0225	0.0556	0.0446	0.0529	0.0465
1000	MSE	0.0002	0.0001	0.0005	0.0041	0.0014	0.0007	0.0002	0.0006	0.0006	0.0015	0.0048
	MADE	0.0116	0.0079	0.0167	0.0413	0.0280	0.0225	0.0114	0.0190	0.0199	0.0345	0.0379

Table 2. Results for diagonal EBINAR(1) model with parameter I.

Size		α_{11}	α_{22}	b_{11}	b_{22}	c_1	c_2	ϕ
50	MSE	0.0031	0.0146	0.0205	0.0030	1.2188	0.6134	0.3256
	MADE	0.0406	0.1021	0.1250	0.0398	0.7710	0.5880	0.5059
100	MSE	0.0020	0.0062	0.0134	0.0023	0.7887	0.4527	0.2778
	MADE	0.0323	0.0665	0.0976	0.0330	0.6125	0.4978	0.4843
200	MSE	0.0015	0.0045	0.0088	0.0010	0.4832	0.3250	0.2572
	MADE	0.0300	0.0524	0.0775	0.0217	0.5016	0.3995	0.4742
500	MSE	0.0007	0.0031	0.0043	0.0010	0.2240	0.1528	0.2198
	MADE	0.0202	0.0396	0.0495	0.0227	0.3655	0.2670	0.4312
1000	MSE	0.0005	0.0020	0.0022	0.0004	0.1965	0.1147	0.1789
	MADE	0.0156	0.0352	0.0377	0.0142	0.3100	0.2084	0.3954

Table 3. Results for diagonal EBINAR(1) model with parameter II.

Size		α_{11}	α_{22}	b_{11}	b_{22}	c_1	c_2	ϕ
50	MSE	0.0154	0.0183	0.0122	0.0191	0.7510	1.1007	0.3183
	MADE	0.0871	0.0988	0.0903	0.0949	0.6894	0.8269	0.5008
100	MSE	0.0059	0.0089	0.0059	0.0072	0.4333	0.6728	0.2336
	MADE	0.0470	0.0742	0.0599	0.0582	0.4957	0.5889	0.4442
200	MSE	0.0042	0.0044	0.0041	0.0053	0.2876	0.4939	0.1983
	MADE	0.0411	0.0499	0.0475	0.0470	0.3796	0.4866	0.4193
500	MSE	0.0027	0.0035	0.0036	0.0025	0.1038	0.3240	0.1899
	MADE	0.0344	0.0400	0.0414	0.0326	0.2636	0.4216	0.4107
1000	MSE	0.0013	0.0022	0.0017	0.0011	0.0730	0.0855	0.1352
	MADE	0.0238	0.0303	0.0307	0.0204	0.1978	0.2221	0.3512

Table 4. Results for diagonal EBINAR(1) model with parameter III.

Size		α_{11}	α_{22}	b_{11}	b_{22}	c_1	c_2	ϕ
50	MSE	0.0027	0.0048	0.0473	0.0083	0.0078	0.0083	0.0013
	MADE	0.0258	0.0428	0.1533	0.0546	0.0620	0.0586	0.0316
100	MSE	0.0036	0.0064	0.0429	0.0102	0.0089	0.0087	0.0017
	MADE	0.0359	0.0485	0.1486	0.0640	0.0680	0.0638	0.0330
200	MSE	0.0060	0.0059	0.0380	0.0059	0.0054	0.0047	0.0017
	MADE	0.0341	0.0469	0.1239	0.0469	0.0541	0.0507	0.0321
500	MSE	0.0018	0.0042	0.0082	0.0031	0.0046	0.0039	0.0016
	MADE	0.0312	0.0429	0.0638	0.0380	0.0477	0.0426	0.0287
1000	MSE	0.0011	0.0030	0.0037	0.0020	0.0020	0.0016	0.0005
	MADE	0.0253	0.0395	0.0380	0.0331	0.0384	0.0341	0.0182

Table 5. Results for diagonal EBINAR(1) model with parameter IV

Size		α_{11}	α_{22}	b_{11}	b_{22}	c_1	c_2	ϕ
50	MSE	0.0031	0.0146	0.0205	0.0030	1.2188	0.6134	0.3256
	MADE	0.0406	0.1021	0.1250	0.0398	0.7710	0.5880	0.5059
100	MSE	0.0020	0.0062	0.0134	0.0023	0.7887	0.4527	0.2778
	MADE	0.0323	0.0665	0.0976	0.0330	0.6125	0.4978	0.4843
200	MSE	0.0015	0.0045	0.0088	0.0010	0.4832	0.3250	0.2572
	MADE	0.0300	0.0524	0.0775	0.0217	0.5016	0.3995	0.4742
500	MSE	0.0007	0.0031	0.0043	0.0010	0.2240	0.1528	0.2198
	MADE	0.0202	0.0396	0.0495	0.0227	0.3655	0.2670	0.4312
1000	MSE	0.0005	0.0020	0.0022	0.0004	0.1965	0.1147	0.1789
	MADE	0.0156	0.0352	0.0377	0.0142	0.3100	0.2084	0.3954

Figures 1 and 2 illustrate the large sample properties of the estimators on a limited sample size. In general, the estimated medians are apparently closer to the real parameter values with the sample size increases. Regarding dispersion issues, both the interquartile ranges and the overall ranges of the produced values are narrower with the sample size increases.

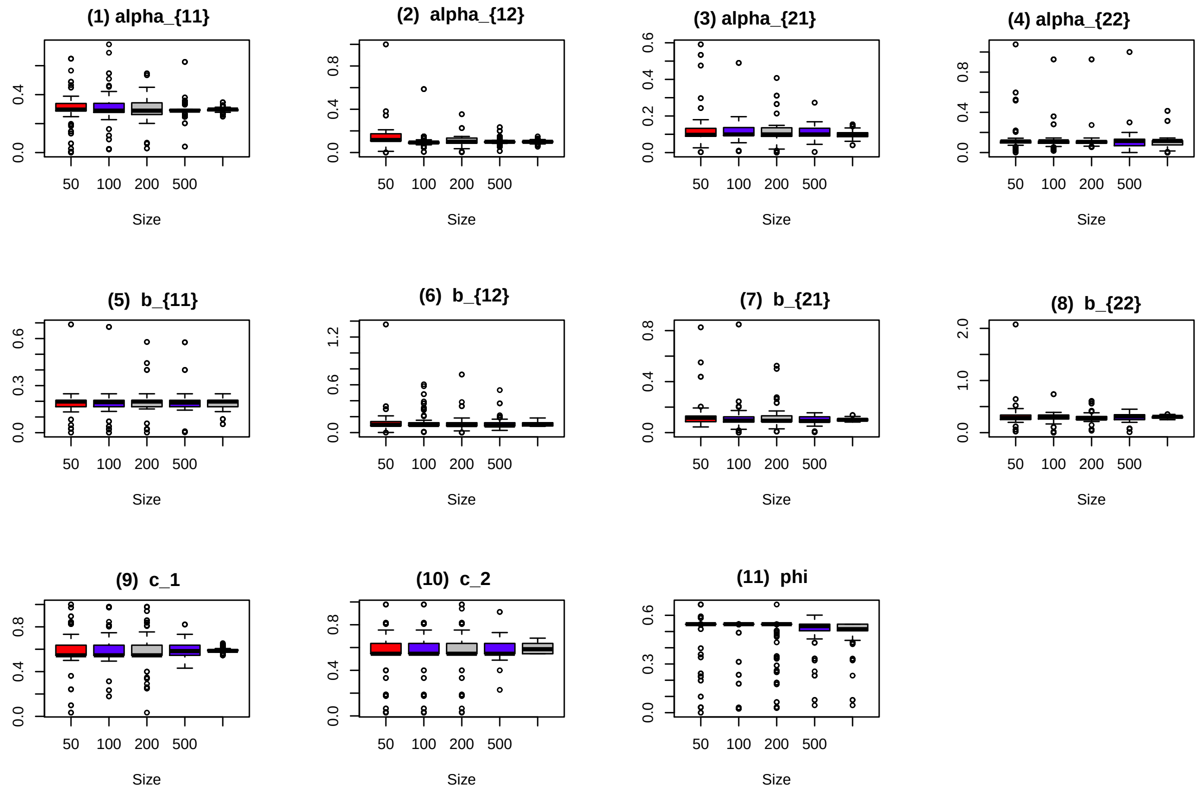

Figure 1. Boxplots of the CML estimates for non-diagonal EBINAR(1) model.

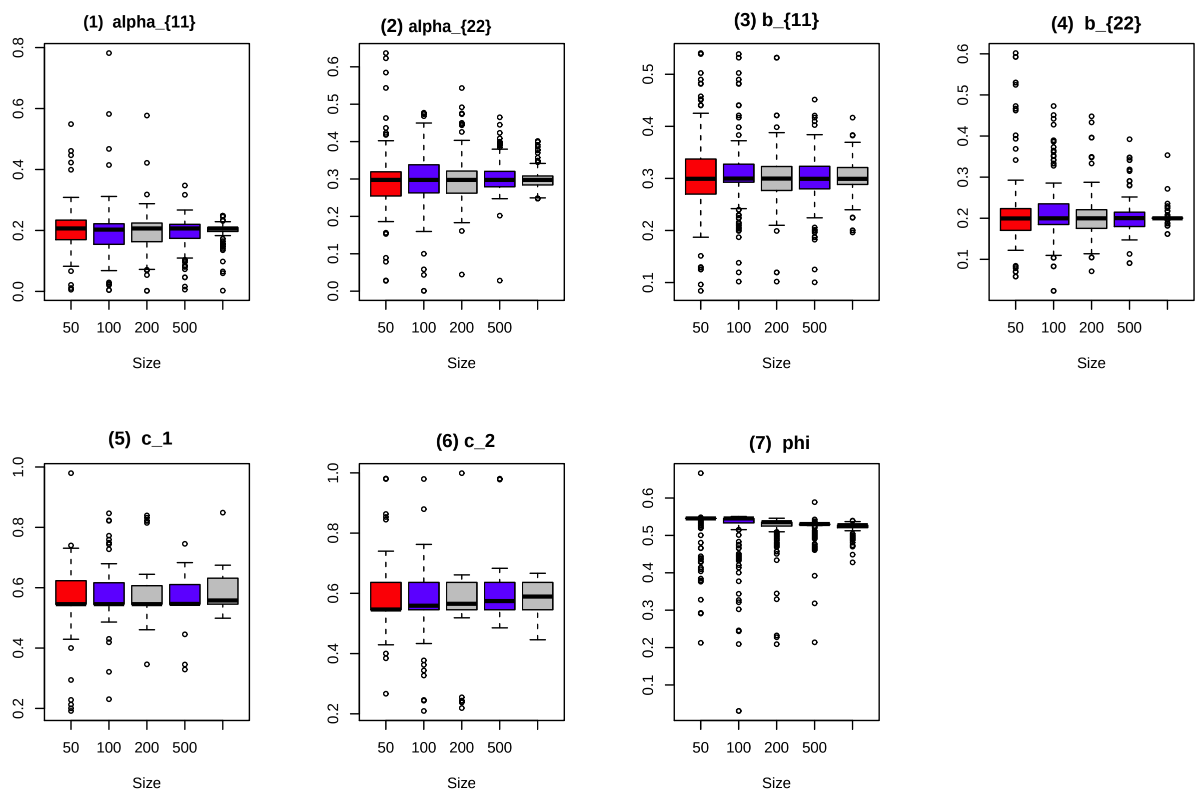

Figure 2. Boxplots of the CML estimates for diagonal EBINAR(1) model with parameter I.

5. Illustrative Examples

In this section, we apply the proposed model to two crime datasets coming from different number of car beats, which is the unique ID for the observation unit's geography in Pittsburgh Police Department. The crime data is available online at "The Forecasting Principles" site (http://www.forecastingprinciples.com/index.php/crimedata) in the section about Crime data and download on 23 September 2016.

According to Cohen and Gorr [27], the occurrence of criminal mischief may be accompanied by burglary behavior, so does for the robbery. Hence, the monthly counts of burglary and CMIS (or those of burglary and robbery) may exhibit dependence. In this section, we take the monthly counts of burglary and CMIS in beat 11 and the monthly counts of burglary and robbery in beat 26 as examples.

5.1. Monthly Counts of Burglary and CMIS in Beat 11

In this part, we consider the monthly number of burglary and criminal mischief (CMIS) from January 1990 to December 2001 in the geographic ID = 11. Table 6 gives the statistics of the counts of burglaries and CMIS.

Table 6. Summary statistics for the monthly number of burglaries and CMIS in beat 11.

Data	Mean	Variance	Minimum	Median	Maximum
Burglary	2.8819	4.1188	0	3	10
CMIS	6.3819	10.0839	1	6	22

Table 6 shows that both the counts of burglaries and CMIS are over-dispersed because their variances are greater than their means. In contrast, this relationship can also be illustrated by the cross-correlation graph of the samples, which are given in Figure 3.

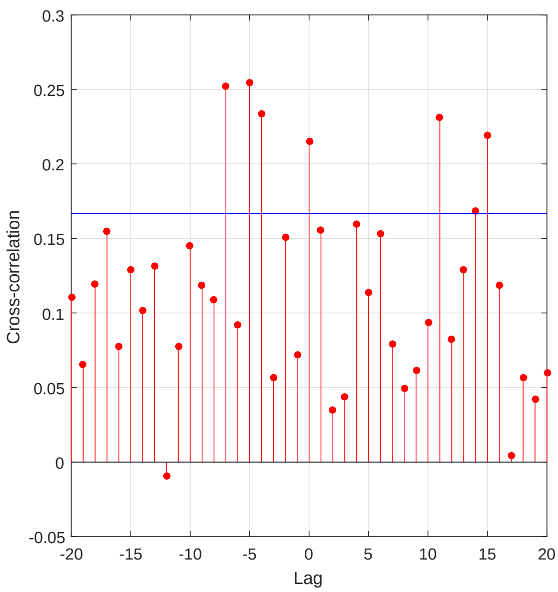

Figure 3. Cross-correlation between the monthly number of burglaries and CMIS in beat 11.

From Figure 3, the counts of burglaries are weakly dependent with those of CMIS. Their plots of sample path, autocorrelation function (ACF) and partial autocorrelation function (PACF) are given in Figure 4, which show that the analyzed data sets are bivariate integer-valued time series with some characteristics of mutual influence.

To give quantitative results about cross-correlation, we compare our model with the following models:

- Full BINAR-BP with ϵ_t following BP$(\lambda_1, \lambda_2, \phi)$ [16];

- Full BINAR-NB with ϵ_t following bivariate negative binomial distribution with parameters $(\lambda_1, \lambda_2, \beta)$; see [14,16] for detail.

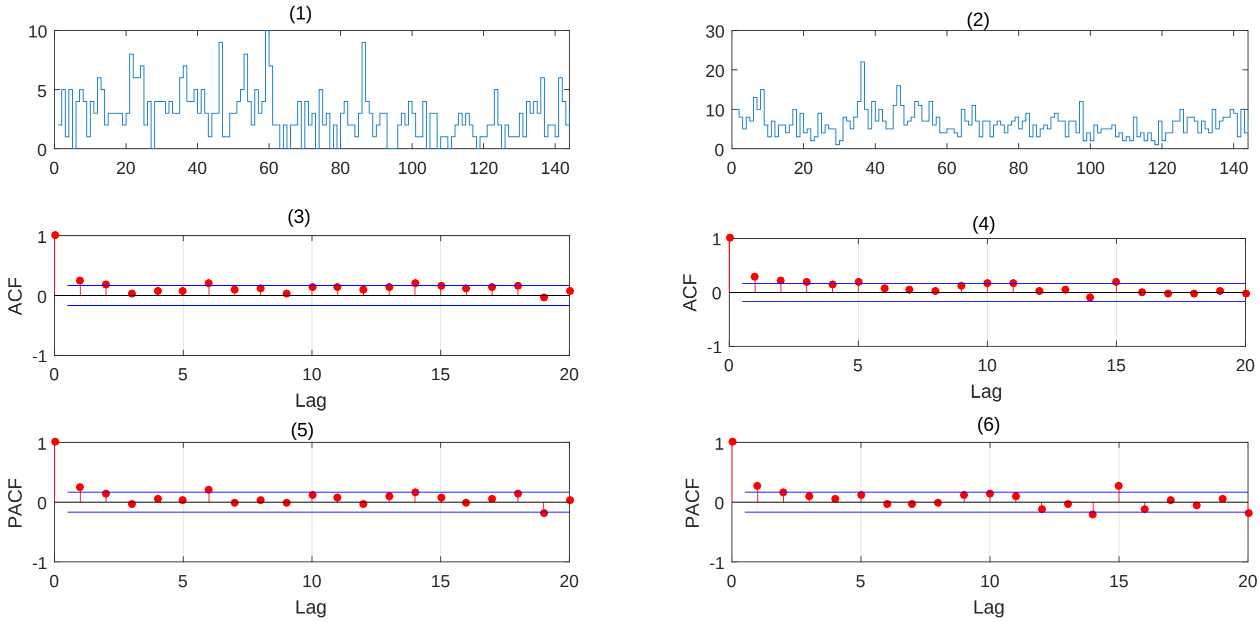

Figure 4. Beat 11: (**1**) monthly number of burglary, (**2**) monthly number of CMIS, (**3**) ACF of burglary, (**4**) ACF of CMIS, (**5**) PACF of burglary, (**6**) PACF of CMIS.

As the goodness-of-fit criteria, we use the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the mean square error of the Pearson residuals (PRMS), which is equal to $\sum_{t=1}^{n} Z_t^2/(n-p^*)$, where p^* denotes the number of estimated parameters and Z_t denotes standardized Pearson residuals.

The CML estimate and approximated standard error (SE) of the parameter, including the fitted values of PRMS, AIC, BIC and log-likelihood function (Log Lik), are summarized in Table 7, where the approximated standard error is computed by using the estimated version of the robust sandwich matrix $(\boldsymbol{J}(\boldsymbol{\theta}_0))^{-1}\boldsymbol{I}(\boldsymbol{\theta}_0)(\boldsymbol{J}(\boldsymbol{\theta}_0))^{-1}$; see Theorem 8 for details.

Table 7. Estimates for the monthly numbers of burglaries and those of CMIS in beat 11.

	EBINAR(1)			Full BINAR(1)-NB			Full BINAR(1)-BP	
Para.	**Estimate**	**SE**	**Para.**	**Estimate**	**SE**	**Para.**	**Estimate**	**SE**
$\hat{\alpha}_{11}$	0.1689	0.1559	$\hat{\alpha}_{11}$	0.2784	0.0665	$\hat{\alpha}_{11}$	0.2993	0.0838
$\hat{\alpha}_{12}$	0.0179	0.0411	$\hat{\alpha}_{12}$	0.0217	0.0092	$\hat{\alpha}_{12}$	0.0217	0.0215
$\hat{\alpha}_{21}$	0.0390	0.1447	$\hat{\alpha}_{21}$	0.1060	0.0550	$\hat{\alpha}_{21}$	0.1060	0.0719
$\hat{\alpha}_{22}$	0.1131	0.1236	$\hat{\alpha}_{22}$	0.5010	0.0295	$\hat{\alpha}_{22}$	0.1934	0.0551
$\hat{b}_{11}$	0.0690	0.1460						
$\hat{b}_{12}$	0.0093	0.0559						
$\hat{b}_{21}$	0.1014	0.1809						
$\hat{b}_{22}$	0.1354	0.1414						
$\hat{c}_1$	1.0007	0.4372	$\hat{\lambda}_1$	1.9814	0.0186	$\hat{\lambda}_1$	1.5164	0.2561
$\hat{c}_2$	3.3190	0.5478	$\hat{\lambda}_2$	2.3137	0.0166	$\hat{\lambda}_2$	4.5258	0.4493
$\hat{\phi}$	0.5273	0.2628	$\hat{\beta}$	0.1374	0.9759	$\hat{\phi}$	0.4044	0.2274
PRMS	0.0064			0.0245			0.0103	
AIC	1315.4620			1387.8913			1350.9488	
BIC	1348.1300			1408.6800			1371.7375	
Log Lik	−646.7310			−686.9457			−668.4744	

Table 7 shows that the EBINAR(1) model takes the highest Log Lik value and the lowest AIC, BIC and PRMS for the monthly number of burglaries and CMIS. Hence, the EBINAR(1) model is more suitable for the data sets.

5.2. Monthly Counts of Burglaries and Robberies in Beat 26

In this part, we consider the monthly number of burglaries and robberies from January 1990 to December 2001 in the geographic ID = 26; see Table 8 for some of their statistics.

Table 8. Summary statistics for the monthly number of burglaries and robberies in beat 26.

Data	Mean	Variance	Minimum	Median	Maximum
Burglary	3.9306	9.7434	0	3	15
Robbery	3.0625	9.6394	0	2	17

Table 8 shows the monthly number of burglary and robbery are over-dispersed. In contrast, this relationship can also be illustrated by their cross-correlation graph given in Figure 5, which shows that the counts of burglary are significantly dependent on those of robbery.

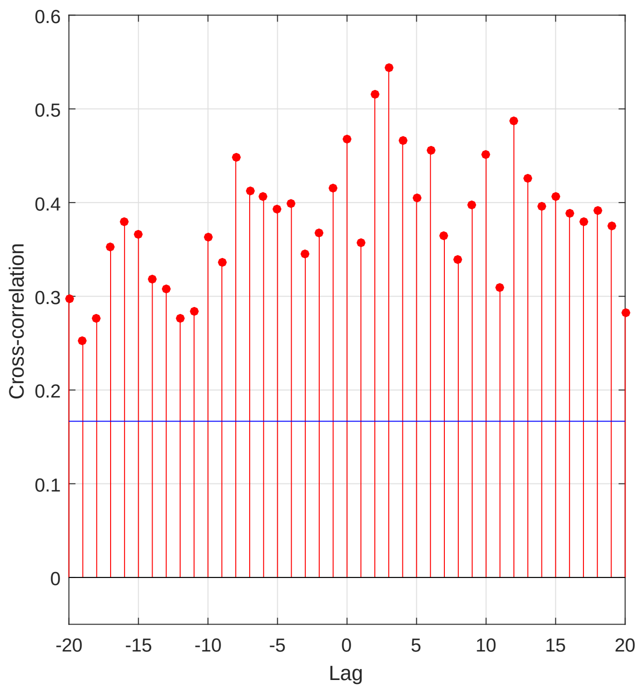

Figure 5. Cross-correlation between the monthly number of burglaries and robberies in beat 26.

To further illustrate the the monthly number of burglaries and robberies in beat 26, we present their sample path, ACF and PACF plots in Figure 6, from which we can conclude that the analyzed data set exhibits some characteristics of mutual influence.

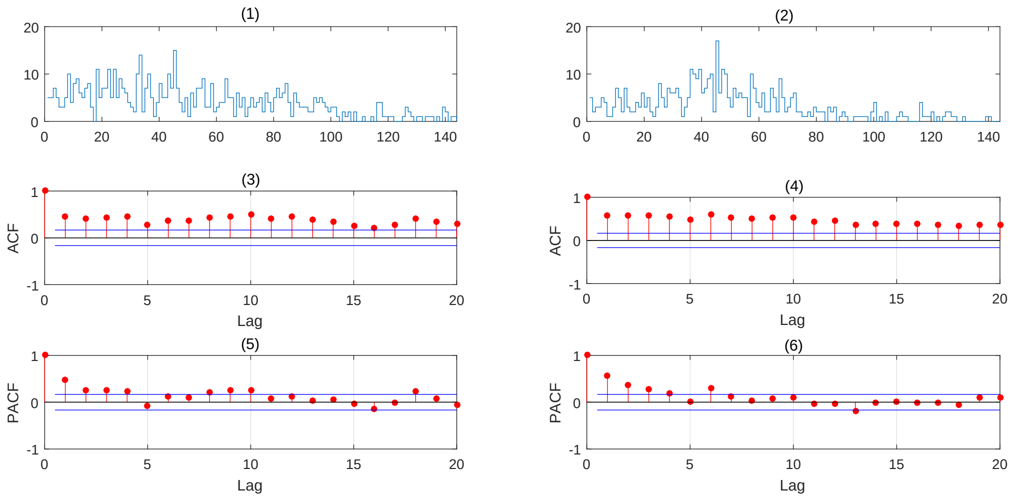

Figure 6. Beat 26: (**1**) monthly number of burglaries, (**2**) monthly number of robberies, (**3**) ACF of burglaries, (**4**) ACF of robberies, (**5**) PACF of burglaries, (**6**) PACF of robberies.

To give quantitative result about cross-correlation, we compare our model with the Full BINAR-BP and Full BINAR-NB models. The CML estimate and SE, including the fitted PRMS, AIC, BIC and Log Lik, are summarized in Table 9.

Table 9. Estimates for the monthly number of burglaries and robberies in beat 26.

	EBINAR(1)			Full BINAR(1)-NB			Full BINAR(1)-BP	
Para.	**Estimate**	**SE**	**Para.**	**Estimate**	**SE**	**Para.**	**Estimate**	**SE**
$\hat{\alpha}_{11}$	0.3117	0.0654	$\hat{\alpha}_{11}$	0.2314	0.3042	$\hat{\alpha}_{11}$	0.2765	0.0537
$\hat{\alpha}_{12}$	0.2086	0.0611	$\hat{\alpha}_{12}$	0.3172	0.2442	$\hat{\alpha}_{12}$	0.0927	0.0471
$\hat{\alpha}_{21}$	0.0900	0.0511	$\hat{\alpha}_{21}$	0.1099	0.2834	$\hat{\alpha}_{21}$	0.0001	0.0000
$\hat{\alpha}_{22}$	0.1906	0.1163	$\hat{\alpha}_{22}$	0.4361	0.2244	$\hat{\alpha}_{22}$	0.4249	0.0415
$\hat{b}_{11}$	0.0671	0.0706						
$\hat{b}_{12}$	0.2280	0.0653						
$\hat{b}_{21}$	0.1233	0.0511						
$\hat{b}_{22}$	0.3358	0.1161						
$\hat{c}_1$	0.2043	0.2048	$\hat{\lambda}_1$	2.2310	0.0026	$\hat{\lambda}_1$	1.7652	0.2048
$\hat{c}_2$	0.4139	0.1139	$\hat{\lambda}_2$	1.1708	0.0076	$\hat{\lambda}_2$	0.9604	0.1601
$\hat{\phi}$	0.5599	0.1187	$\hat{\beta}$	0.4073	0.7189	$\hat{\phi}$	0.7778	0.1494
PRMS	0.0087			0.0748			0.0992	
AIC	1320.8092			1344.6968			1357.7718	
BIC	1353.4771			1365.4855			1378.5604	
Log Lik	−649.4046			−665.3484			−671.8859	

Table 9 shows that the EBINAR(1) model takes the highest Log Lik value and the lowest AIC, BIC and PRMS for burglaries and robberies in beat 26. Hence, EBINAR(1) model is more suitable.

To sum up, our findings reveal that there are some connections for the burglary and CMIS in beat 11 and those for the burglary and robbery in beat 26, which agrees with the conclusion of Cohen and Gorr [27]. Of course, the counts of burglary may be affected by other crime activities, such as simple assault, vagrancy and trespassing, which will be studied in a further study.

Remark 2. *For the two real datasets, our EBINAR(1) model performs best, but it is not clear enough regarding predicting unseen data. To further illustrate the better performance of the new model in prediction, one available solution is to conduct a further experiment when dividing the*

considered data into a training set and test set. In addition such experiment will be considered in future study of the crime data.

6. Concluding Remarks

This paper proposes a more flexible model for bivariate integer-valued time series data, i.e., the EBINAR(1) model, whose innovation vector is time-dependent. It is a generalization of the EINAR(1) model [11] to the two-dimensional case as well as a generalization of the BINAR(1) model [14,16], but with more flexibility. We discuss some necessary properties of the model, the CML estimators of parameters and their large-sample properties. Simulation was conducted to examine the finite sample performance of estimators. Real data examples are provided to illustrate our model to be effective relative to existing models.

To make the bivariate INAR-type models more flexible with respect to real-data applications in some cases, it may be interesting to include explanatory covariates or periodicity in the model to account for dependence through thinning operations on several factors, which will be considered in another project: see Aknouche et al. [28] and Chen and Khamthong [29].

Author Contributions: Conceptualization, H.C. and F.Z.; methodology, H.C. and F.Z.; software, H.C., F.Z. and X.L.; validation, H.C., F.Z. and X.L.; formal analysis, H.C.; investigation, H.C.; resources, F.Z.; data curation, X.L.; writing—original draft preparation, H.C. and F.Z.; writing—review and editing, H.C. and F.Z.; visualization, H.C. and F.Z.; supervision, F.Z. All authors have read and agreed to the published version of the manuscript.

Funding: Chen's work is supported by the Natural Science Foundation of Henan province (No.222300420127). Zhu's work is supported by the National Natural Science Foundation of China (Nos.11871027, 11731015) and the Natural Science Foundation of Jilin Province (No.20210101143JC). Liu's work is supported by the Basic Research Programs of Shanxi Province (No.202103021223084).

Data Availability Statement: Crime data are available online at The Forecasting Principles site and were downloaded on 23 September 2016 from http://www.forecastingprinciples.com/index.php/crimedata.

Acknowledgments: We are very grateful to anonymous referees for providing several exceptionally helpful comments which led to a significant improvement in the manuscript.

Conflicts of Interest: The authors declare no conflict of interests in the publication of this paper.

Abbreviations

The following abbreviations are used in this manuscript:

$\boldsymbol{A}^{-1}$	inverse of matrix $\boldsymbol{A}$;
$\boldsymbol{A}^{\top}$	transpose of the matrix or vector $\boldsymbol{A}$;
$\|\cdot\|$	Euclidean norm of a matrix or vector;
$\lvert\cdot\rvert$	absolute value of a univariate variable;
$\xrightarrow{d}$	convergence in distribution;
$\xrightarrow{p}$	convergence in probability one;
pmf	probability mass function;
CML	conditional maximun likelihood;
AIC	Akaike information criterion;
BIC	Bayesian information criterion;
SE	standard error;
PRMS	mean square error of the Pearson residual;
Para.	parameter.

References

1. Steutel, F.W.; van Harn, K. Discrete analogues of self-decomposability and stability. *Ann. Probab.* **1979**, *7*, 893–899. [CrossRef]
2. Al-Osh, M.A.; Alzaid, A.A. First-order integer-valued autoregressive process. *J. Time Ser. Anal.* **1987**, *8*, 261–275. [CrossRef]
3. McKenzie, E. Some simple models for discrete variate time series. *Water Resoure Bull.* **1985**, **21**, 645–650. [CrossRef]

4. Du, J.; Li, Y. The integer valued autoregressive INAR(p) model. *J. Time Ser. Anal.* **1991**, *12*, 129–142.
5. Alzaid, A.A.; Omair, M.A. Poisson difference integer valued autoressive model of order one. *Bull. Malays. Math. Sci. Soc.* **2014**, *37*, 465–485.
6. Chen, H.; Li, Q.; Zhu, F. Binomial AR(1) processes with innovational outliers. *Commun. Stat. Theory Methods* **2021**, *50*, 446–472. [CrossRef]
7. Weiß, C.H. Thinning operations for modeling time series of counts—A survey. *Adv. Stat. Anal.* **2008**, *92*, 319–341. [CrossRef]
8. Scotto, M.G.; Wei, C.H.; Gouveia, S. Thinning-based models in the analysis of integer-valued time series: A review. *Stat. Model.* **2015**, *15*, 590–618. [CrossRef]
9. Davis, R.A.; Fokianos, K.; Holan, S.H.; Joe, H.; Livsey, J.; Lund, R.; Pipiras, V.; Ravishanker, N. Count time series: A methodological review. *J. Am. Stat. Assoc.* **2021**, *116*, 1533–1547. [CrossRef]
10. Buckley, F.M.; Pollett, P.K. Limit theorems for discrete-time metapopulation models. *Probab. Surv.* **2010**, *7*, 53–83. [CrossRef]
11. Weiß, C.H. A Poisson INAR(1) model with serially dependent innovations. *Metrika* **2015**, *78*, 829–851. [CrossRef]
12. Franke, J.; Rao, T.S. *Multivariate First-Order Integer Valued Autoregressions*; Technical Report; Department of Mathematics, UMIST: Manchester, UK, 1995.
13. Latour, A. The multivariate GINAR(p) process. *Adv. Appl. Probab.* **1997**, *29*, 228–248. [CrossRef]
14. Pedeli, X.; Karlis, D. A bivariate INAR(1) processes with application. *Stat. Model.* **2011**, *11*, 325–349. [CrossRef]
15. Pedeli, X.; Karlis, D. On estimation of the bivariate Poisson INAR process. *Commun. Stat. Simul. Comput.* **2013**, *42*, 514–533. [CrossRef]
16. Pedeli, X.; Karlis, D. Some properties of multivariate INAR(1) processes. *Comput. Stat. Data Anal.* **2013**, *67*, 213–225. [CrossRef]
17. Ravishanker, N.; Serhiyenko, V.; Willig, M.R. Hierarchical dynamic models for multivariate times series of counts. *Stat. Its Interface* **2014**, *7*, 559–570. [CrossRef]
18. Popović, P.M. A bivariate INAR(1) model with different thinning parameters. *Stat. Pap.* **2016**, *57*, 517–538. [CrossRef]
19. Scotto, M.G.; Wei, C.H.; Silva, M.E.; Pereira, I. Bivariate binomial autoregressive models. *J. Multivar. Anal.* **2014**, *125*, 233–251. [CrossRef]
20. Li, Q.; Chen, H.; Liu, X. A new bivariate random coefficient INAR(1) model with applications. *Symmetry* **2022**, *14*, 39. [CrossRef]
21. Kocherlakota, S.; Kocherlakota, K. *Bivariate Discrete Distributions*; Marcel Dekker: New York, NY, USA, 1992; pp. 87–97.
22. Heathcote, C.R. Corrections and comments on the paper "A branching process allowing immigration". *J. R. Stat. Soc. Ser. B* **1966**, *28*, 213–217. [CrossRef]
23. Shumway, R.H.; Stoffer, D.S. *Time Series Analysis and Its Applications with R examples*, 3rd ed.; Springer: New York, NY, USA, 2011.
24. Silva, M.E.; Oliveira, V.L. Difference equations for the higher-order moments and cumulants of the INAR(1) model. *J. Time Ser. Anal.* **2004**, *25*, 317–333. [CrossRef]
25. Amemiya, T. *Advanced Econometrics*; Harvard University Press: Cambridge, MA, USA, 1985; pp. 110–112.
26. Freeland, R.K.; McCabe, B.P.M. Analysis of low count time series data by poisson autoregression. *J. Time Ser. Anal.* **2004**, *25*, 701–722. [CrossRef]
27. Cohen, J.; Gorr, W.L. *Development of Crime Forecasting and Mapping Systems for Use by Police*; Inter-University Consortium for Political and Social Research: New York, NY, USA, 2005. [CrossRef]
28. Aknouche, A.; Bentarzi, W.; Demouche, N. On periodic ergodicity of a general periodic mixed Poisson autoregression. *Stat. Probab. Lett.* **2018**, *134*, 15–21. [CrossRef]
29. Chen, C.W.S.; Khamthong, K. Bayesian modelling of nonlinear negative binomial integer-valued GARCHX models. *Stat. Model.* **2020**, *20*, 537–561. [CrossRef]