Article

A New Extension of Thinning-Based Integer-Valued Autoregressive Models for Count Data

Zhengwei Liu and Fukang Zhu * [ID]

School of Mathematics, Jilin University, 2699 Qianjin Street, Changchun 130012, China; zhengweil18@mails.jlu.edu.cn
* Correspondence: fzhu@jlu.edu.cn

Abstract: The thinning operators play an important role in the analysis of integer-valued autoregressive models, and the most widely used is the binomial thinning. Inspired by the theory about extended Pascal triangles, a new thinning operator named extended binomial is introduced, which is a general case of the binomial thinning. Compared to the binomial thinning operator, the extended binomial thinning operator has two parameters and is more flexible in modeling. Based on the proposed operator, a new integer-valued autoregressive model is introduced, which can accurately and flexibly capture the dispersed features of counting time series. Two-step conditional least squares (CLS) estimation is investigated for the innovation-free case and the conditional maximum likelihood estimation is also discussed. We have also obtained the asymptotic property of the two-step CLS estimator. Finally, three overdispersed or underdispersed real data sets are considered to illustrate a superior performance of the proposed model.

Keywords: extended binomial distribution; INAR; thinning operator; time series of counts

1. Introduction

Counting time series naturally occur in many contexts, including actuarial science, epidemiology, finance, economics, etc. The last few years have witnessed the rapid development of modeling time series of counts. One of the most common approaches for modeling integer-valued autoregressive (INAR) time series is based on thinning operators. In order to fit different kinds of situations, many corresponding operators have been developed; see [1] for a detailed discussion on thinning-based INAR models.

The most popular thinning operator is the binomial thinning operator introduced by [2]. Let X be a non-negative integer-valued random variable and $\alpha \in (0,1)$, the binomial thinning operator is defined as

$$\alpha \circ X = \sum_{i=1}^{X} B_i, \quad X > 0, \tag{1}$$

and 0 otherwise, where $\{B_i\}$ is a sequence of independent identically distributed (i.i.d.) Bernoulli random variables with fixed success probability α, and B_i is independent of X. Based on the binomial thinning operator, [3,4] independently proposed an INAR(1) model as follows

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \quad t \in \mathbb{Z}, \tag{2}$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. integer-valued random variables with finite mean and variance. Since this seminal work, the INAR-type models have received considerable attention. For recent literature on this topic, see [5,6], among others.

Note that B_i in (1) follows a Bernoulli distribution, so $\alpha \circ X$ is always less than or equal to X; in other words, the first part of the right side in (2) cannot be greater than X_{t-1},

which limits the flexibility of the model. Although it has such a shortcoming, the simple form makes it easy to estimate the parameter, and it also has many similar properties to the multiplication operator in the continuous case. For this reason, there have still been many extensions of the binomial thinning operator since its emergence. Zhu and Joe [7] proposed the expectation thinning operator, which is the generalization of binomial thinning from the perspective of a probability generating function (pgf). Although this extension is very successful, the estimation procedure is a little complicated. Compared with this extension, the thinning operator we proposed is simpler and more intuitive. For recent developments, Yang et al. [8] proposed the generalized Poisson (GP) thinning operator, which is defined by replacing B_i with a GP counting series. Although the GP thinning operator is flexible and adaptable, we argue that it has a potential drawback: the GP distribution is not a strict probability distribution in the conventional sense. Recently, Aly and Bouzar [9] introduced a two-parameter expectation thinning operator based on a linear fractional probability generating function, which can be regarded as a general case of at least nine thinning operators. Kang et al [10] proposed a new flexible thinning operator, which is named GSC because of three initiators of the counting series: Gómez-Déniza, Sarabia and Calderín-Ojeda.

Although the binomial thinning operator is very popular, it may not perform very well in large numerical value counting time series. This is because under such circumstances, the predicted data are often volatile, and the data are more likely to be non-stationary when the numerical value is large. We intend to establish a new thinning operator which meets the following requirements: (i) it is an extension of the binomial thinning operator; (ii) it contains two parameters to achieve flexibility, (iii) it has a simple structure and is easy to implement.

Based on the above considerations, we propose a new thinning operator based on the extended binomial (EB) distribution. The operator has two parameters: real-valued α and integer-valued m ($0 \leq \alpha \leq 1$, $m \geq 2$), which is more flexible compared to some single parameter thinning, and the binomial thinning operator (1) can be regarded as a special case of $m = 2$ in the EB thinning. The case of $m > 2$ in the EB thinning usually performs better than $m = 2$ in some large value data sets. In other words, the EB thinning alleviates the main defect of the binomial thinning to some extent. Since the EB thinning is not a special case of the expectation thinning in [9], we have further extended the framework of thinning-based INAR models to provide a new way in practical application. Therefore, an INAR(1) model is proposed based on the EB thinning operator, which is an extension of the model (2) and can more accurately and flexibly capture the dispersed features in real data.

This paper is organized as follows. In Section 2, we review the properties of the EB distribution and then introduce the EB thinning operator. Based on the new thinning operator, we propose a new INAR(1) model. In Section 3, two-step conditional least squares estimation is investigated for the innovation-free case of the model and the asymptotic property of the estimator is obtained. The conditional maximum likelihood estimation is discussed and the numerical simulations. In Section 4, we focus on forecasting and introduce two criteria to compare the prediction results for three overdispersed or underdispersed real data sets, which are considered to illustrate a better performance of the proposed model. In Section 5, we give some conclusions and related discussions.

2. A New INAR(1) Model

The EB distribution comes from the theory about Pascal's triangles, which can be regarded as a multivariate case of the binomial distribution; see [11] for more details. Based on this distribution, we introduce the EB thinning operator and propose a corresponding INAR(1) model.

2.1. EB Distribution

The EB random variable $X_n(m, \alpha)$, denoted as $EB(m, n, \alpha)$, which is defined as follows:

$$P(X_n(m, \alpha) = r) = C_m(n, r)\alpha^r \beta^{(m-1)n-r}, \quad 0 \le r \le (m-1)n, \tag{3}$$

where m and n are both integers satisfying $m \ge 2$ and $n \ge 1$; $C_m(n, r)$ can be calculated as

$$C_m(n, r) = \sum_{s=0}^{s_1} (-1)^s \binom{n}{s} \binom{r + n - sm - 1}{n - 1},$$

where $s_1 = \min\{n, \text{integer part in } r/m\}$; and α and β in (3) satisfy the following restriction:

$$\beta^{m-1} + \alpha\beta^{m-2} + \alpha^2\beta^{m-3} + \ldots + \alpha^{m-1} = 1, \quad 0 \le \alpha \le 1, \ 0 \le \beta \le 1. \tag{4}$$

The above restriction is equivalent to $\beta^m - \alpha^m = \beta - \alpha$. The mean and variance of EB random variables $X_n(m, \alpha)$ are

$$E(X_n(m, \alpha)) = n\alpha \frac{1 - m\alpha^{m-1}}{\beta - \alpha}, \quad \text{Var}(X_n(m, \alpha)) = n\alpha\beta \frac{1 - m^2(\alpha\beta)^{m-1}}{(\beta - \alpha)^2},$$

respectively. The pgf of $X_n(m, \alpha)$ can be written as $G(t) = E(t^{X_n(m, \alpha)}) = \left(\frac{\beta^m - \alpha^m t^m}{\beta - \alpha t}\right)^n$.

As $X_n(m, \alpha)$ can be expressed as a convolution, the EB distribution has the reproductive property. Specifically, if $Y_1, Y_2, \ldots, Y_k$ are independent random variables with $Y_i \sim EB(m, n_i, \alpha)$ in (3), then $\sum_{i=1}^k Y_i \sim EB(m, \sum_{i=1}^k n_i, \alpha)$. Notice that random variable $Y_i \sim EB(2, 1, \alpha)$ is equivalent to a Bernoulli random variable satisfying $P(Y_i = 1) = 1 - P(Y_i = 0) = 1 - \alpha$.

2.2. EB Thinning Operator

According to discussions in 2.1, we construct the EB thinning operator based on the configuration of $n = 1$. Let $\{U_i(m, \alpha)\}$ be a sequence of i.i.d. random variables with common distribution $EB(m, 1, \alpha)$, i.e., $P(U_i(m, \alpha) = z) = \alpha^z \beta^{(m-1)-z}, z = 0, \ldots, m - 1$, where α and β satisfy (4). Note that the mean and variance of U_i are

$$\mu = E(U_i) := \alpha \frac{1 - m\alpha^{m-1}}{\beta - \alpha}, \quad \sigma^2 = \text{Var}(U_i) := \alpha\beta \frac{1 - m^2(\alpha\beta)^{m-1}}{(\beta - \alpha)^2}. \tag{5}$$

One can easily see that $\mu < \sigma^2$ if and only if

$$m\alpha^{m-2}(m\beta^m - \beta + \alpha) < 1. \tag{6}$$

For any $3 \le m < \infty$, the left-hand side of (6) approaches 0 as $\alpha \to 0$ and m as $\alpha \to 1$, respectively. Hence, $\mu < \sigma^2$ or $\mu \ge \sigma^2$ is possible. When $m = 2$, which corresponds to the binomial distribution, and (5) gives $\mu = \alpha$ and $\sigma^2 = \alpha\beta$ with $\alpha + \beta = 1$, then we have $\mu > \sigma^2$ for all $0 < \alpha < 1$. When $\alpha \ge \beta$ and $m \to \infty$, one can easily show that $\mu = \theta/(1 - \theta)$, $\sigma^2 = \theta/(1 - \theta)^2$, where $\theta = \alpha/\beta$. Therefore, $\mu \le \sigma^2$ for all $0 < \theta < 1$ in this case.

Based on the reproductive property of the EB distribution, we define the EB thinning operator "⊙" as follows: for a non-negative integer-valued random variable X,

$$(m, \alpha) \odot X = \sum_{i=1}^X U_i(m, \alpha), \quad X > 0,$$

where $(m, \alpha) \odot X = 0$ if $X = 0$. Note that the EB thinning operator reduces to the binomial operator (1) when $m = 2$. It is easy to know that $(m, \alpha) \odot X \le X$ or $> X$, so the EB thinning operator is quite flexible when dealing with the overdispersed or underdispersed data sets.

Remark 1. *The computation of* (α, m) *for given* (μ, σ^2) *is based on* (4) *and* (5)*. The solution can be obtained by solving these nonlinear equations. When* $m = 3$*,* $\beta = (-\alpha + \sqrt{4 - 3\alpha^2})/2$ *and when* $m = 4$*,*

$$
\beta = \sqrt[3]{-\left(\frac{10\alpha^3}{27} - \frac{1}{2}\right) + \sqrt{\left(\frac{10\alpha^3}{27} - \frac{1}{2}\right)^2 + \left(\frac{2\alpha^2}{9}\right)^3}}
$$
$$
+ \sqrt[3]{-\left(\frac{10\alpha^3}{27} - \frac{1}{2}\right) - \sqrt{\left(\frac{10\alpha^3}{27} - \frac{1}{2}\right)^2 + \left(\frac{2\alpha^2}{9}\right)^3}} - \frac{\alpha}{3}.
$$

For more complex cases $(m \geq 5)$*, we can derive the solution* (α, β, m) *by solving these large-scale nonlinear systems, and a more detailed calculation procedure is given in Section 3.3.*

2.3. EB-INAR(1) Model

Based on the EB thinning operator, we define the EB-INAR(1) model as follows:

$$
X_t = (m, \alpha) \odot X_{t-1} + \epsilon_t, \quad t = 1, 2, \ldots \tag{7}
$$

where $\alpha \in (0, 1)$, $\{X_t\}$ is a sequence of non-negative integer-valued random variables; the innovation process $\{\epsilon_t\}$ is a sequence of i.i.d. integer-valued random variables with finite mean and variance; and ϵ_t is independent of $\{X_s, s < t\}$.

In order to obtain the estimation equations, we give some conditional or unconditional moments of the EB-INAR(1) model in the following proposition.

Proposition 1. *Suppose* $\{X_t\}$ *is a stationary process defined by* (7) *and let* $\mu < 1$*; then for* $t \geq 1$*,*

1. $E(X_t | X_{t-1}) = \mu X_{t-1} + \mu_\epsilon$;
2. $E(X_t) = \frac{\mu_\epsilon}{1 - \mu}$;
3. $\mathrm{Var}(X_t | X_{t-1}) = \sigma^2 X_{t-1} + \sigma_\epsilon^2$;
4. $\mathrm{Var}(X_t) = \frac{\sigma^2 \mu_\epsilon + \sigma_\epsilon^2 (1 - \mu)}{(1 - \mu)^2 (1 + \sigma^2)}$;
5. $\mathrm{Cov}(X_t, X_{t-h}) = \mu^h \mathrm{Var}(X_{t-h})$ *and* $\mathrm{Corr}(X_t, X_{t-h}) = \mu^h$*, for* $h = 0, 1, 2, \ldots$

where μ_ϵ *and* σ_ϵ^2 *are the expectation and variance of the innovation* ϵ_t*, respectively.*

The proof of some of these properties mentioned above is given in Appendix A.

Remark 2. *Inspired by the INAR(p) model in* [12]*, we can further extend this model to INAR(p); the EB-INAR(p) model is defined as follows:*

$$
X_t = (m, \alpha_1) \odot X_{t-1} + \ldots + (m, \alpha_p) \odot X_{t-p} + \epsilon_t, \quad t = 2, 3, \ldots
$$

where $\alpha_1, \ldots, \alpha_p \in (0, 1)$*,* m *is an integer satisfying* $m \geq 2$*,* $\{X_t\}$ *is a sequence of non-negative integer-valued random variables, the innovation process* $\{\epsilon_t\}$ *is a sequence of i.i.d. integer-valued random variables with finite mean and variance, and* ϵ_t *is independent with* $\{X_s, s < t\}$*.*

We will show that the new model can accurately and flexibly capture the dispersion features of real data in Section 4.

3. Estimation

We use the two-step conditional least squares estimation proposed by [13] to investigate the innovation-free case and the asymptotic properties of the estimators are obtained. Conditional maximum likelihood estimation for the parametric cases are also discussed. Finally, we demonstrate the finite sample performance via simulation studies.

3.1. Two-Step Conditional Least Squares Estimation

Denote $\theta_1 = (\mu, \mu_\epsilon)^\top$, $\theta_2 = (\sigma^2, \sigma_\epsilon^2)^\top$ and $\theta = (\theta_1^\top, \theta_2^\top)^\top$. The two-step CLS estimation will be conducted by the following two steps.

Step 1.1. The estimator for θ_1.

Let $g_1(\theta_1, X_{t-1}) = E(X_t|X_{t-1}) = \mu X_{t-1} + \mu_\epsilon$, $q_{1t}(\theta_1) = (X_t - g_1(\theta_1, X_{t-1}))^2$. Let

$$Q_1(\theta_1) = \sum_{t=1}^{n} q_{1t}(\theta_1)$$

be the CLS criterion function. Then the CLS estimator $\hat{\theta}_{1,CLS} := (\hat{\mu}_{CLS}, \hat{\mu}_{\epsilon,CLS})^\top$ of θ_1 can be obtained by solving the score equation $\frac{\partial Q_1(\theta_1)}{\partial \theta_1} = 0$, which implies a closed-form solution:

$$\hat{\theta}_{1,CLS} = \begin{pmatrix} \sum_{t=1}^{n} X_{t-1}^2 & \sum_{t=1}^{n} X_{t-1} \\ \sum_{t=1}^{n} X_{t-1} & n \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{n} X_t X_{t-1} \\ \sum_{t=1}^{n} X_t \end{pmatrix}.$$

Step 1.2. The estimator for θ_2.

Let $Y_t = X_t - E(X_t|X_{t-1})$, $g_2(\theta_2, X_{t-1}) = \mathrm{Var}(X_t|X_{t-1}) = \sigma^2 X_{t-1} + \sigma_\epsilon^2$. Then

$$\begin{aligned} E(Y_t^2|X_{t-1}) &= E((X_t - E(X_t|X_{t-1})^2)|X_{t-1}) \\ &= \mathrm{Var}(X_t|X_{t-1}) = g_2(\theta_2, X_{t-1}). \end{aligned}$$

Let $q_{2t}(\theta_2) = (Y_t^2 - g_2(\theta_2, X_{t-1}))^2$; then the CLS criterion function for θ_2 can be written as

$$Q_2(\theta_2) = \sum_{t=1}^{n} q_{2t}(\theta_2).$$

By solving the score equation $\frac{\partial Q_2(\theta_2)}{\partial \theta_2} = 0$, we can obtain the CLS estimator $\hat{\theta}_{2,CLS} := (\hat{\sigma}_{CLS}^2, \hat{\sigma}_{\epsilon,CLS}^2)^\top$ of θ_2, which also is a closed-form solution:

$$\hat{\theta}_{2,CLS} = \begin{pmatrix} \sum_{t=1}^{n} X_{t-1}^2 & \sum_{t=1}^{n} X_{t-1} \\ \sum_{t=1}^{n} X_{t-1} & n \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{n} Y_t^2 X_{t-1} \\ \sum_{t=1}^{n} Y_t^2 \end{pmatrix}.$$

Step 2. Estimating parameters (m, α) via the method of moments.

The estimator $(\hat{m}, \hat{\alpha})$ of (m, α), which is called a two-step CLS estimator, can be obtained by solving the following estimation equations:

$$\begin{cases} \hat{\mu}_{CLS} = \alpha \frac{1 - m\alpha^{m-1}}{\beta - \alpha}, \\ \hat{\sigma}_{CLS}^2 = \alpha\beta \frac{1 - m^2(\alpha\beta)^{m-1}}{(\beta - \alpha)^2}, \end{cases} \tag{8}$$

where α and β satisfy (4).

Therefore, the resulting CLS estimator is $\hat{\Theta}_{CLS} = (\hat{m}_{CLS}, \hat{\alpha}_{CLS}, \hat{\mu}_{\epsilon,CLS}, \hat{\sigma}_{\epsilon,CLS}^2)^\top$. To study the asymptotic behaviour of the estimator, we make the following assumptions:

Assumption 1. $\{X_t\}$ *is a stationary and ergodic process;*

Assumption 2. $EX_t^4 < \infty$.

Proposition 2. *Under assumptions 1 and 2, the CLS estimator $\hat{\theta}_{1,CLS}$ is strongly consistent and asymptotically normal:*

$$\sqrt{n}(\hat{\theta}_{1,CLS} - \theta_{1,0}) \xrightarrow{L} N(0, V_1^{-1} W_1 V_1^{-1}),$$

where $V_1 := E\left(\frac{\partial}{\partial\theta_1}g_1(\theta_{1,0},X_0)\frac{\partial}{\partial\theta_1^\top}g_1(\theta_{1,0},X_0)\right)$, $W_1 := E\left(q_{11}(\theta_{1,0})\frac{\partial}{\partial\theta_1}g_1(\theta_{1,0},X_0)\frac{\partial}{\partial\theta_1^\top}$
$g_1(\theta_{1,0},X_0)\right)$, and $\theta_{1,0} = (\mu_0,\mu_{\epsilon_0})$ denotes the true value of θ_1.

To obtain the asymptotic normality of $\hat{\theta}_{2,CLS}$, we make a further assumption:

Assumption 3. $EX_t^6 < \infty$.

Then we have the following proposition.

Proposition 3. *Under assumptions 1 and 3, the CLS estimator $\hat{\theta}_{2,CLS}$ is strongly consistent and asymptotically normal:*

$$\sqrt{n}(\hat{\theta}_{2,CLS} - \theta_{2,0}) \xrightarrow{L} N(0, V_2^{-1}W_2V_2^{-1}),$$

where $V_2 := E\left(\frac{\partial}{\partial\theta_2}g_2(\theta_{2,0},X_0)\frac{\partial}{\partial\theta_2^\top}g_2(\theta_{2,0},X_0)\right)$, $W_2 := E\left(q_{21}(\theta_{2,0})\frac{\partial}{\partial\theta_2}g_2(\theta_{2,0},X_0)\frac{\partial}{\partial\theta_2^\top}$
$g_2(\theta_{2,0},X_0)\right)$, and $\theta_{2,0} = (\sigma_0^2,\sigma_{\epsilon_0}^2)$ denotes the true value of θ_2.

Based on Propositions 2 and 3 and Theorem 3.2 in [14], we have the following proposition.

Proposition 4. *Under assumptions 1 and 3, the CLS estimator $\hat{\theta}_{CLS} = (\hat{\theta}_{1,CLS}, \hat{\theta}_{2,CLS})^\top$ is strongly consistent and asymptotically normal:*

$$\sqrt{n}(\hat{\theta}_{CLS} - \theta_0) \xrightarrow{L} N(0, \Omega),$$

where

$$\Omega = \begin{pmatrix} V_1^{-1}W_1V_1^{-1} & V_1^{-1}MV_2^{-1} \\ V_2^{-1}M^\top V_1^{-1} & V_2^{-1}W_2V_2^{-1} \end{pmatrix},$$

$M = E\left(\sqrt{q_{11}(\theta_{1,0})q_{21}(\theta_{2,0})}\frac{\partial}{\partial\theta_1}g_1(\theta_{1,0},X_0)\frac{\partial}{\partial\theta_2^\top}g_2(\theta_{2,0},X_0)\right)$, and $\theta_0 = (\theta_{1,0},\theta_{2,0})^\top$ denotes the true value of θ.

We do the following preparation to establish Proposition 5. Based on (5), solve the equation about (m,α), and denote the solution as $(h_1(\mu,\sigma^2), h_2(\mu,\sigma^2))$. Let

$$D = D(\mu,\sigma^2) = \begin{pmatrix} \partial h_1/\partial\mu & \partial h_1/\partial\sigma^2 \\ \partial h_2/\partial\mu & \partial h_2/\partial\sigma^2 \end{pmatrix}. \tag{9}$$

Based on Proposition 4, we state the strong consistency and asymptotic normality of $(\hat{m},\hat{\alpha})^\top$ in the following proposition.

Proposition 5. *Under assumptions 1 and 3, the CLS estimator $(\hat{m}_{CLS}, \hat{\alpha}_{CLS})^\top$ is strongly consistent and asymptotically normal:*

$$\sqrt{n}\begin{pmatrix} \hat{m}_{CLS} - m_0 \\ \hat{\alpha}_{CLS} - \alpha_0 \end{pmatrix} \xrightarrow{L} N(0, D\Sigma D^\top),$$

where D is given in (9); $\Sigma = diag(IV_1^{-1}W_1V_1^{-1}I^\top, IV_2^{-1}W_2V_2^{-1}I^\top)$ with $I = (1,0)$; m_0 and α_0 denote the true values of m and α, respectively.

The brief proofs of Propositions 2–5 are given in Appendix A.

3.2. Conditional Maximum Likelihood Estimation

We maximize the likelihood function with respect to the model parameters $\theta = (m, \alpha, \delta)$ to get the conditional maximum likelihood (CML) estimate of the parametric case

$$L(X_1 = x_1, \ldots, X_N = x_N | \theta) = P_\theta(X_1 = x_1) \prod_{i=1}^{N} P_\theta(X_i = x_i | X_{i-1} = x_{i-1}, \ldots, X_1 = x_1)$$

$$= P_\theta(X_1 = x_1) \prod_{i=1}^{N} P_\theta(X_i = x_i | X_{i-1} = x_{i-1}),$$

where δ is the parameter of ϵ_i, P_{X_1} is the pmf for X_1 and $P_\theta(X_{i+1}|X_i)$ is the conditional pmf. Since the marginal distribution is difficult to obtain in general, a simple approach is conditional on the observed X_1. By essentially ignoring the dependency on the initial value and considering the CML estimate given X_1 as an estimate for θ by maximizing the conditional log-likelihood

$$l(X_1 = x_1, \ldots, X_N = x_N | \theta) = \sum_{i=2}^{N} \log P_\theta(X_i | X_{i-1})$$

over Θ, we denote the CML estimate by $\hat{\theta} = (\hat{m}, \hat{\alpha}, \hat{\delta})$. The log-likelihood function is as follows:

$$l(X_1 = x_1, \ldots, X_N = x_N | \theta) = \sum_{i=2}^{N} \log \left\{ \sum_{w=0}^{\min\{(m-1)x_{i-1}, x_i\}} C_m(x_{i-1}, w) \alpha^w \beta^{(m-1)x_{i-1}-w} \cdot P(\epsilon_i = x_i - w) \right\},$$

where α and β satisfy (4); ϵ_i follows a non-negative discrete distribution with a parameter δ. In what follows, we consider two cases: $m = 3, 4$.

Case 1: For $m=3$ with Poisson innovation, i.e., $\epsilon_t \sim P(\delta)$.

$$l(X_1 = x_1, \ldots, X_N = x_N | \theta) = \sum_{i=2}^{N} \log \left\{ \sum_{w=0}^{\min\{2x_{i-1}, x_i\}} \sum_{t_1=\max\{0, x_{i-1}-w\}}^{x_{i-1}-\frac{w}{2}} \binom{x_{i-1}}{t_1} \binom{x_{i-1}-t_1}{2x_{i-1}-2t_1-w} \right.$$

$$\left. \cdot (\beta^2)^{t_1} (\alpha\beta)^{2x_{i-1}-2t_1-w} (\alpha^2)^{t_1-x_{i-1}+w} \frac{\delta^{(x_i-w)}}{(x_i-w)!} e^{-\delta} \right\},$$

where β is given in Remark 1.

Case 2: For $m=4$ with geometric innovation, i.e. $\epsilon_t \sim Ge(\delta) = (1-\delta)^k \delta$ for $k = 0, 1, 2, \ldots$

$$l(X_1 = x_1, \ldots, X_N = x_N | \theta) = \sum_{i=2}^{N} \log \left\{ \sum_{w=0}^{\min\{3x_{i-1}, x_i\}} \sum_{t_1=x_{i-1}-w}^{x_{i-1}-\frac{w}{3}} \sum_{t_2=\max\{0, 2x_{i-1}-2t_1-w\}}^{\frac{3x_{i-1}-3t_1-w}{2}} \binom{x_{i-1}}{t_1} \binom{x_{i-1}-t_1}{t_2} \right.$$

$$\cdot \binom{x_{i-1}-t_1-t_2}{3x_{i-1}-3t_1-2t_2-w} (\beta^3)^{t_1} (\alpha\beta^2)^{t_2} (\alpha^2\beta)^{3x_{i-1}-3t_1-2t_2-w}$$

$$\left. \cdot (\alpha^3)^{2t_1+t_2-2x_{i-1}+w} (1-\delta)^{(x_i-w)} \delta \right\},$$

where β is given in Remark 1. For higher order m, the formula is a little tedious, which is omitted here. For the estimate of EB-INAR(p), the CML estimation is too complicated, but the two-step CLS estimation is quite feasible, the procedure is similar to the case of $p = 1$. For this reason, we only consider the case of EB-INAR(1) in simulation studies.

3.3. Simulation

A Monte Carlo simulation study was conducted to evaluate the finite sample performance of the estimator. For CLS estimation, we used the package BB in R for solving and optimizing large-scale nonlinear systems to solve Equations (4) and (8). For CML estimation, we used the package maxLik in R to maximize the log-likelihood function.

We considered the following configurations of the parameters:

- Poisson INAR(1) models with $\theta = (m, \alpha, \delta)^\top$:
  (A1) $= (3, 0.2, 1)^\top$;　(A2) $= (3, 0.1, 0.5)^\top$;　(A3) $= (4, 0.2, 1)^\top$;　(A4) $= (4, 0.1, 0.5)^\top$;
- Geometric INAR(1) models with $\theta = (m, \alpha, \delta)^\top$:
  (B1) $= (3, 0.3, 0.5)^\top$;　(B2) $= (3, 0.4, 2/3)^\top$;　(B3) $= (4, 0.3, 0.5)^\top$;　(B4) $= (4, 0.4, 2/3)^\top$.

In simulations, we chose sample sizes $n = 100, 200$ and 400 with $M = 500$ replications for each choice of parameters. The root mean squared error (RMSE) was calculated to evaluate the performance of the estimator according to the following formula: RMSE $= \sqrt{\dfrac{1}{M-1} \sum_{j=1}^{M} (\hat{\xi}_j - \xi_0)^2}$, where $\hat{\xi}_j$ is the estimator of ξ_0 in the jth replication.

For the CLS estimate, the solutions of (4) and (8) are sensitive to $\hat{\mu}$ and $\hat{\sigma}^2$, so we adopted the following estimation procedure. First, calculate 500 groups of $\hat{\mu}$ and $\hat{\sigma}^2$ estimates, then use the mean values of $\hat{\mu}$ and $\hat{\sigma}^2$ to solve the Equations (4) and (8). The simulation results of CLS are summarized in Table 1. We found that the estimation values are closer to the true value and the values of RMSE gradually decrease as the sample size increases.

Table 1. Means of estimates, RMSEs (within parentheses) by CLS.

Case	n	$\hat{m}$	$\hat{\alpha}$	$\hat{\delta}$
A1	100	2.9100	0.1758	1.0220(0.1595)
	200	2.9219	0.1931	1.0142(0.1148)
	400	2.9995	0.1984	1.0090(0.0771)
A2	100	2.7628	0.0886	0.5014(0.0893)
	200	2.8169	0.0873	0.5049(0.0668)
	400	2.9132	0.0942	0.5004(0.0449)
A3	100	3.7885	0.1713	1.0265(0.1594)
	200	3.8450	0.1902	1.0165(0.1127)
	400	3.8957	0.1952	1.0113(0.0789)
A4	100	3.8421	0.0912	0.5059(0.0912)
	200	3.8483	0.0912	0.5031(0.0603)
	400	3.9590	0.0981	0.5012(0.0439)
B1	100	2.9074	0.3122	0.4981(0.0511)
	200	2.9588	0.3115	0.5008(0.0365)
	400	2.9858	0.3087	0.4984(0.0270)
B2	100	3.0719	0.3841	0.6578(0.0553)
	200	3.0523	0.3877	0.6569(0.0400)
	400	3.0986	0.3924	0.6600(0.0307)
B3	100	3.6127	0.2703	0.4962(0.0536)
	200	3.7937	0.2961	0.4980(0.0404)
	400	3.9217	0.2984	0.4970(0.0269)
B4	100	3.9575	0.3935	0.6417(0.0653)
	200	4.0388	0.3929	0.6556(0.0478)
	400	4.0027	0.3953	0.6606(0.0353)

Note: RMSE, root mean squared error.

As it is a little difficult to estimate the parameter m in CML estimation, we considered m as known. The simulation results of CML estimators are given in Table 2. For all cases, all estimates generally show small values of RMSE, and the values of RMSE gradually decrease as the sample size increases.

Table 2. Means of estimates, RMSEs (within parentheses) by CML.

Case	n	100	200	400
A1	$\hat{\alpha}$	0.1858(0.0656)	0.1927(0.0476)	0.1970(0.0347)
	$\hat{\delta}$	1.0070(0.1420)	1.0043(0.1057)	1.0057(0.0806)
A2	$\hat{\alpha}$	0.1041(0.0575)	0.1027(0.0453)	0.0973(0.0357)
	$\hat{\delta}$	0.4939(0.0804)	0.4964(0.0593)	0.4973(0.0422)
A3	$\hat{\alpha}$	0.1827(0.0592)	0.1923(0.0406)	0.1966(0.0306)
	$\hat{\delta}$	1.0186(0.1358)	1.0030(0.1059)	1.0002(0.0789)
A4	$\hat{\alpha}$	0.0938(0.0500)	0.0940(0.0414)	0.0987(0.0342)
	$\hat{\delta}$	0.4993(0.0790)	0.5052(0.0583)	0.4967(0.0420)
B1	$\hat{\alpha}$	0.2982(0.0456)	0.2980(0.0329)	0.2992(0.0237)
	$\hat{\delta}$	0.5086(0.0469)	0.5013(0.0323)	0.5016(0.0224)
B2	$\hat{\alpha}$	0.3888(0.0455)	0.3949(0.0331)	0.3980(0.0222)
	$\hat{\delta}$	0.6685(0.0501)	0.6681(0.0374)	0.6664(0.0256)
B3	$\hat{\alpha}$	0.2904(0.0441)	0.2958(0.0276)	0.2990(0.0219)
	$\hat{\delta}$	0.5039(0.0480)	0.5003(0.0338)	0.5006(0.0264)
B4	$\hat{\alpha}$	0.3867(0.0348)	0.3940(0.0240)	0.3965(0.0163)
	$\hat{\delta}$	0.6650(0.0539)	0.6674(0.0382)	0.6662(0.0271)

Note: RMSE, root mean squared error.

4. Real Data Examples

In this section, three real data sets, including overdispersed and underdispersed settings, are considered to illustrate the better performance of the proposed model. The first example is overdispersed crime data in Pittsburgh; the second is overdispersed stock data in New York Stock Exchange (NYSE); and the third is underdispersed crime data in Pittsburgh, which was also analyzed by [15]. As is well known, in time series analysis, forecasting is very important in model evaluation. We first introduce two criteria on forecasting, and other preparations.

4.1. Forecasting

Before introducing the evaluation criterion, we briefly introduce the basic procedure as follows: First, we divide the $n_1 + n_2$ data into two parts, the training set with the first n_1 data and the prediction set with the last n_2 data. The training set is used to estimate the parameters and evaluate the fitness of the model. Then we can evaluate the efficiency of each model by comparing the following criteria between prediction data and the real data in the prediction set.

Similar to the procedure in [16], which performs an out-of-sample experiment to compare forecasting performances of two model-based bootstrap approaches, we introduce the forecasting procedure as follows: For each $t = (n_1 + 1), \ldots, (n_1 + n_2 - 5)$ we estimate an INAR(1) model for the data $x_1, \ldots, x_t$, then we use the fitted result based on $x_1, \ldots, x_t$ to generate the next five forecasts, which is called the 5-step ahead forecast $x_{t+1}^F, \ldots, x_{t+5}^F$ for each t in $\{(n_1 + 1), \ldots, (n_1 + n_2 - 5)\}$, where x_t^F is the forecast at time t. In this way we obtain many sequences of $1, 2, \ldots, 5$ step-ahead forecasts, finally we replicate the whole

procedure P times. Then we can evaluate the point forecast accuracy by the forecast mean square error (FMSE) defined as

$$\text{FMSE} = \frac{1}{P} \sum_{i=(n_1+1)}^{n_2} (x_i - \bar{x}_i^F)^2,$$

and forecast mean absolute error (FMAE) defined as

$$\text{FMAE} = \frac{1}{P} \sum_{i=(n_1+1)}^{n_2} |x_i - \bar{x}_i^F|,$$

where x_i is the true value of the data, $\bar{x}_i^F$ is the mean of all the forecasts at i and P is the number of replicates.

4.2. Overdispersed Cases

We consider two overdispersed data sets, the first one contains 144 observations and represents monthly tallies of crime data from the Forecasting Principles website http://www.forecastingprinciples.com, and these crimes are reported in the police car beats in Pittsburgh from January 1990 to December 2001; the second one is Empire District Electric Company (EDE) data set from the Trades and Quotes (TAQ) set in NYSE, which contains 300 observations, and it was also analyzed by [17].

4.2.1. P1V Data

The 45th P1V (Part 1 Violent Crimes) data set contains crimes of murder, rape, robbery and other kinds; see more details in the data dictionary on the Forecasting Principles website. Figure 1 plots the time series plot, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of 45th data of P1V series, respectively. The maximum value of the data is 15 and the minimum is 0; the mean is 4.3333; the variance is 7.4685. From the ACF plot, we found that the data are dependent. From the PACF plots, we can see that only the first sample is significant, which strongly suggests an INAR(1) model.

First, we divided the data set into two parts–the training set with the first $n_1 = 134$ counting data and the prediction set with the last $n_2 = 10$ data. We fit the training set by the following models: expectation thinning INAR(1) (ETINAR(1)) model in [9], GSC thinning INAR(1) (GSCINAR(1)) model in [10], the binomial thinning INAR(1) model and EB thinning EB-INAR(1) models with $m = 3, 4$. According to the mean and variance of P1V data, we used one of the most common settings–geometric distribution–as the distribution of the innovation in above models.

In order to compare the effectiveness of the models, we consider the following evaluation criteria: (1) AIC. (2) The mean and standard error of Pearson residual r_t and its related Ljung–Box statistics, where the Pearson residuals are defined as

$$r_t = \frac{X_t - \hat{\mu} X_{t-1} - \hat{\mu}_\epsilon}{[\hat{\sigma}^2 X_{t-1} + \hat{\sigma}_\epsilon^2]^{\frac{1}{2}}}, \quad t = 1, 2, \ldots,$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are the estimated expectation and variance for related thinning operators, respectively. (3) Three goodness-of-fit statistics: RMS (root mean square error), MAE (mean absolute error) and MdAE (median absolute error), where the error is defined by $X_t - E(X_t|X_{t-1})$, $t = 1, \ldots, n_1$. (4) The mean of the data $\hat{\bar{x}}$ on the training set calculated by the estimated results.

Next, focusing on forecasting, we generated $P = 100$ replicates based on the training set for each model. Then we calculated the FMSE and FMAE for each model.

All results of the fitted models are given in Table 3. There is no evidence of any correlation within the residuals of all five models, which is also supported by the Ljung–Box statistic based on 15 lags (because $\chi_{0.05}^2(14) = 23.6847$). There were no significant

differences for the RMS, MAE, MdAE and $\hat{\bar{x}}$ values (the true mean of the 134 training set was 4.3880) of the models. In other words, no model performed the best in terms of these four criteria, so we also considered AIC. Since the CML estimator cannot be adopted in GSCINAR(1), one can only compare other criteria.

Considering the fitness on the training set, the EB-INAR(1) with $m = 3$ has the smallest AIC, EB-INAR(1) with $m = 4$ has almost the same AIC as $m = 3$. For the results on forecasting, EB-INAR(1) with $m = 4$ has the smallest FMSE and the second smallest FMAE among all models. EB-INAR(1) with $m = 3$ has the second smallest FMSE and the smallest FMAE. Based on these results, we conclude that EB-INAR(1) with $m = 3, 4$ performs better than INAR(1), ETINAR(1) and GSCINAR(1).
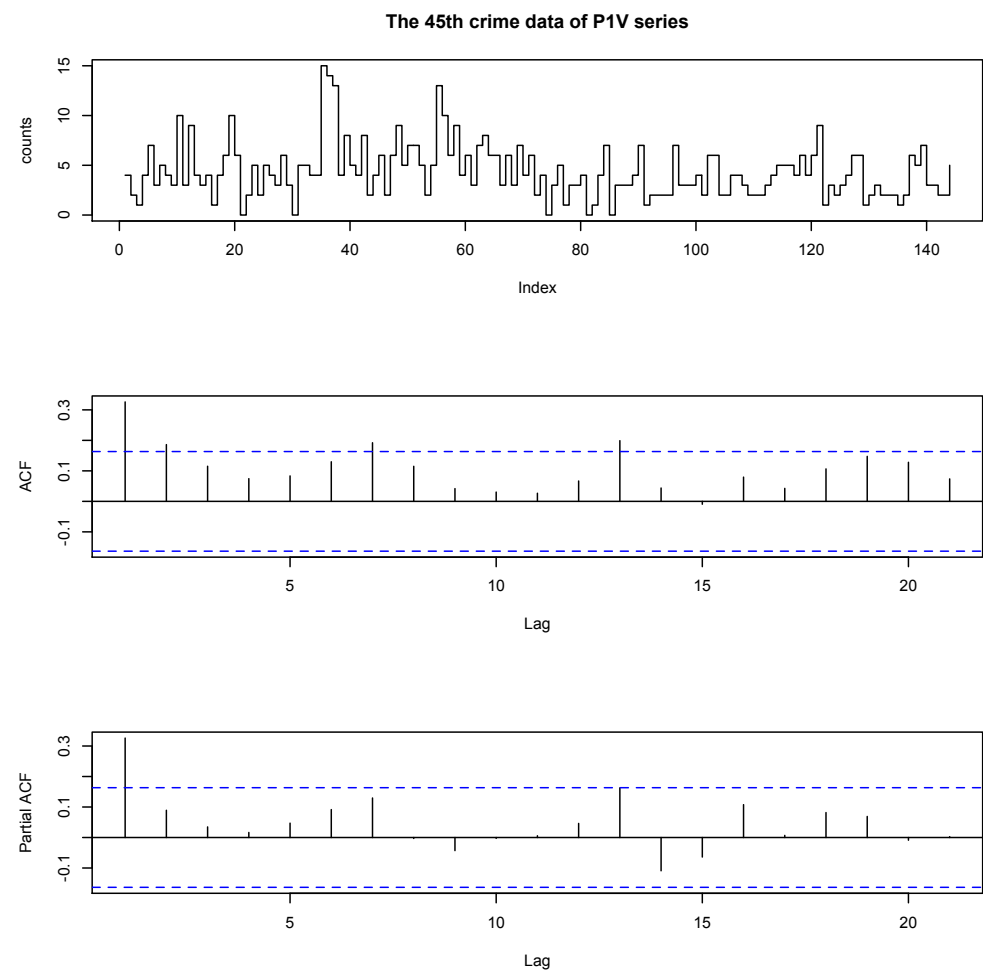
Figure 1. The data, autocorrelation function (ACF) and partial autocorrelation function (PACF) of 45th P1V series.

4.2.2. Stock Data

We analyzed another overdispersed data set of Empire District Electric Company (EDE) from the Trades and Quotes (TAQ) data set in NYSE. The data are about the number of trades in 5 min intervals between 9:45 a.m. and 4:00 p.m. in the first quarter of 2005 (3 January–31 March 2005, 61 trading days). Here we analyze a portion of the data between first to fourth trading days. As there are 75 5 min intervals per day, the sample size was $T = 300$.

Figure 2 plots the time series plot, the ACF and the PACF of the EDE series. The maximum value of the data is 25 and the minimum is 0; the mean is 4.6933; and the variance is 14.1665. It seems that the series is not completely stationary with several outliers or

influential observations based on the time series plot. Zhu et al. [18] analyzed the Poisson autoregression for the stock transaction data with extreme values, which can be considered in the current setting. From the ACF plot, we found that the data are dependent. From the PACF plots, we can see that only the first sample is significant, which strongly suggests an INAR(1) model. We used the same procedures and criteria as before. We used the geometric distribution as the distribution of the innovation in above models.

First divide the data set into two parts–the training set with the first $n_1 = 270$ data and the prediction set with the last $n_2 = 30$ data. All results of the fitted models are given in Table 4. Among all models, EB-INAR(1) with $m = 4$ has the smallest AIC, and there is no evidence of any correlation within the residuals of all five models, which is also supported by the Ljung–Box statistic based on 15 lags. There are no significant differences for the RMS, MAE, MdAE and $\hat{\bar{x}}$ values (the true mean of the 270 training set was 4.3407) of all considered models. For the results of prediction, EB-INAR(1) with $m = 4$ has the smallest FMSE and FMAE among all models. Based on the above results, we conclude that EB-INAR(1) with $m = 4$ performs best for this data set.
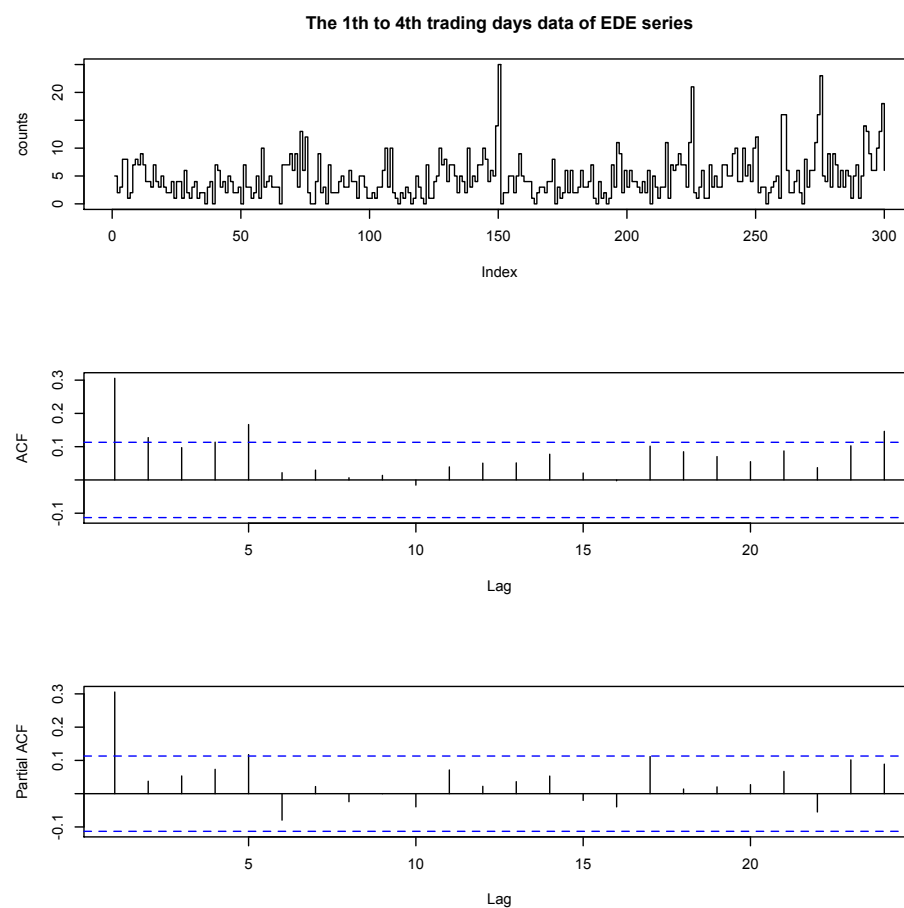

Figure 2. The data, ACF and PACF of first to fourth trading days of EDE series.

Table 3. Fitting results, AIC and some characteristics of P1V data.

Model	Estimates	AIC	$\overline{r_t}$	std(r_t)	Ljung-Box	RMS	MAE	MdAE	$\hat{\overline{x}}$	FMSE	FMAE
INAR(1)	$\hat{\alpha} = 0.3955$ $\hat{\delta} = 0.2741$	641.9136	0.0020	0.8008	12.5545	2.6384	2.0236	1.5828	4.3811	0.3988	0.1809
ETINAR(1)	$\hat{q} = 0.4942$ $\hat{r} = 0.3739$ $\hat{\delta} = 0.3112$	639.3243	0.0134	0.8392	16.6092	2.6736	2.0628	1.6958	4.3758	0.3847	0.1821
GSCINAR(1)	$\hat{\gamma} = 0.8944$ $\hat{\delta} = 0.2515$	——	0.0005	0.6805	12.7469	2.6304	2.0027	1.7395	4.3838	0.4115	0.1935
EB-INAR(1) with $m = 3$	$\hat{\alpha} = 0.3299$ $\hat{\delta} = 0.3050$	636.2386	0.0097	0.8525	15.7722	2.6666	2.0563	1.7167	4.3771	0.3770	0.1748
EB-INAR(1) with $m = 4$	$\hat{\alpha} = 0.3164$ $\hat{\delta} = 0.3156$	636.2887	0.0145	0.8526	17.0785	2.6789	2.0673	1.6815	4.3756	0.3589	0.1791

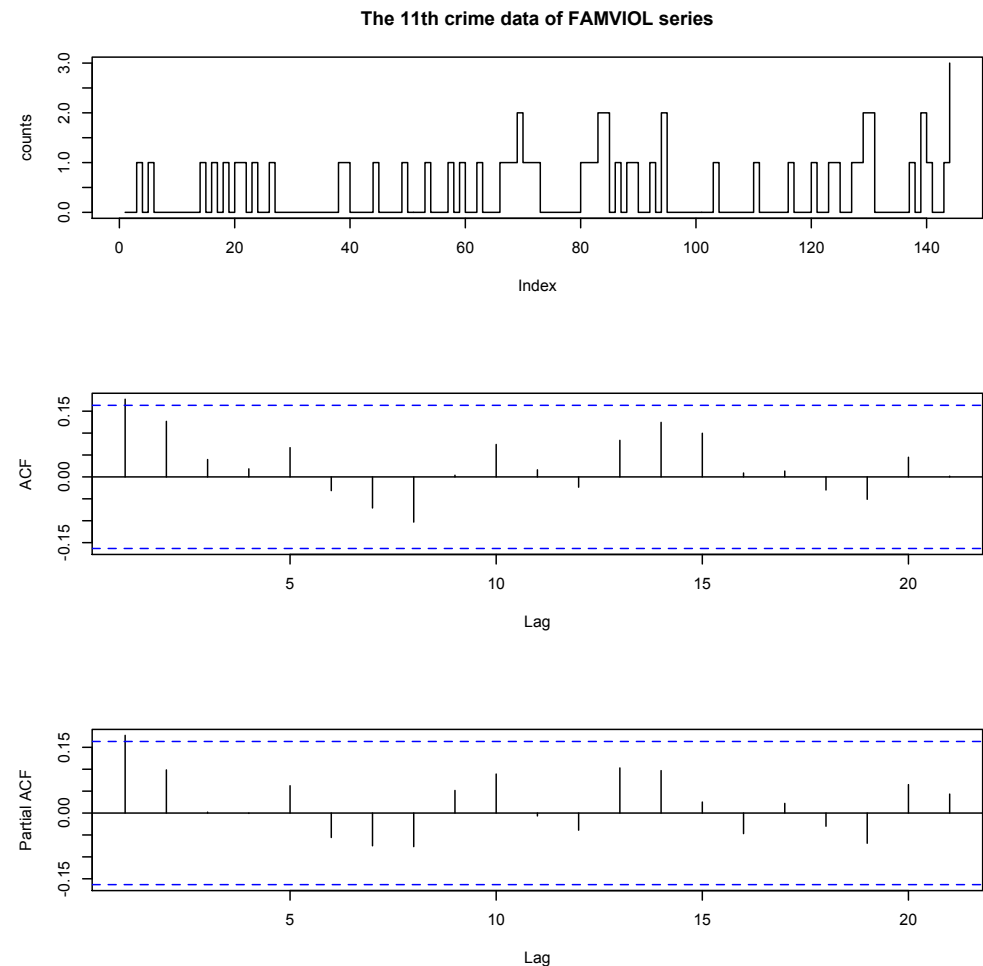**Table 4.** Fitting results, AIC and some characteristics of EDE data.

Model	Estimates	AIC	$\overline{r_t}$	std(r_t)	Ljung-Box	RMS	MAE	MdAE	$\hat{\overline{x}}$	FMSE	FMAE
INAR(1)	$\hat{\alpha} = 0.2457$ $\hat{\delta} = 0.2341$	1341.648	0.0003	0.8590	13.0696	3.3407	2.4817	2.0075	4.3358	10.2602	1.2368
ETINAR(1)	$\hat{q} = 0.3204$ $\hat{r} = 0.4125$ $\hat{\delta} = 0.2533$	1339.397	0.0041	0.8811	16.4070	3.3570	2.4910	2.0534	4.3357	10.6103	1.2688
GSCINAR(1)	$\hat{\gamma} = 0.9578$ $\hat{\delta} = 0.2284$	——	-0.0003	0.8027	13.1596	3.3397	2.4812	2.0409	4.3361	10.4577	1.2362
EB-INAR(1) with $m = 3$	$\hat{\alpha} = 0.2200$ $\hat{\delta} = 0.2448$	1338.071	0.0018	0.8801	14.5878	3.3477	2.4863	2.0299	4.3352	10.1038	1.2317
EB-INAR(1) with $m = 4$	$\hat{\alpha} = 0.2194$ $\hat{\delta} = 0.2486$	1337.554	0.0027	0.8838	15.3238	3.3514	2.4880	2.0224	4.3350	10.0463	1.2279

4.3. Underdispersed Case

The 11th FAMVIOL data set contains the crimes of family violence, which can also be obtained from the Forecasting Principles website. Figure 3 plots the time series plot, the ACF and the PACF of the 11th data set of FAMVIOL series. The maximum value of the data is 3 and the minimum is 0; the mean is 0.4027; and the variance is 0.3820. We use the procedures and criteria in Section 4.2.1 to compare different models. According to the mean and the variance of FAMVIOL data, we use one of the most common settings-Poisson distribution as the distribution of the innovation in above models.

All results of the fitted models are given in Table 5. There is no evidence of any correlation within the residuals of all five models, which is also supported by the Ljung–Box statistic based on 15 lags. There are no significant differences about the criteria on the fitness and forecasting of all models. ETINAR(1) with the biggest AIC, performed the worst in these models.

Now let us have a brief summary. For the P1V data and stock data, which are overdispersed with slightly high-count data, the EB-INAR(1) of $m > 2$ is obviously better than $m = 2$. For the FAMVIOL data, which is underdispersed with small-count data, the EB-INAR(1) with $m > 2$ is also competitive.

Figure 3. The data, ACF and PACF of the 11th data set of the FAMVIOL series.

Table 5. Fitting results, AIC and some characteristics of FAMVIOL data.

Model	Estimates	AIC	$\overline{r_t}$	std(r_t)	Ljung-Box	RMS	MAE	MdAE	$\hat{\overline{x}}$	FMSE	FMAE
INAR(1)	$\hat{\alpha} = 0.1750$ $\hat{\delta} = 0.3101$	206.5983	-0.0005	0.9157	8.6672	0.5596	0.4940	0.4851	0.3759	0.1111	0.0790
ETINAR(1)	$\hat{q} = 0.1774$ $\hat{r} = 0.0267$ $\hat{\delta} = 0.3092$	208.5826	-0.0003	0.9122	8.6344	0.5596	0.4939	0.4866	0.3759	0.1205	0.0837
GSCINAR(1)	$\hat{\gamma} = 0.9741$ $\hat{\delta} = 0.3045$	——	0.0023	0.8609	8.3735	0.5595	0.4932	0.4944	0.3759	0.1020	0.0814
EB-INAR(1) with $m = 3$	$\hat{\alpha} = 0.1505$ $\hat{\delta} = 0.3068$	206.5335	0.0007	0.9005	8.5485	0.5595	0.4935	0.4900	0.3757	0.1135	0.0792
EB-INAR(1) with $m = 4$	$\hat{\alpha} = 0.1371$ $\hat{\delta} = 0.3131$	206.8414	-0.0014	0.8943	8.4928	0.5597	0.4944	0.4802	0.3759	0.1149	0.0808

5. Conclusions

This paper proposes an EB-INAR(1) model based on the newly constructed EB thinning operator, which is an extension of the thinning-based INAR models. We gave the estimation method for parameters and established the asymptotic properties of the estimators for the innovation-free case. Based on the simulations and real data analysis, the EB-INAR(1) model can accurately and flexibly capture the dispersion features of the data, which shows its effectiveness and practicality. Compared with other models, such as ETINAR(1) and GSCINAR(1), our model is competitive.

We point out that many existing integer-valued models can be generalized by replacing the binomial thinning operator with the EB thinning operator, such as those models in [19–23]. In addition, we can extend the considered first-order INAR model to the higher-order one. More research will be studied in the future.

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

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof of Proposition 1. Since 1–4 are easy to verify, we only prove 5. By the law of total covariance, we have

$$
\begin{aligned}
Cov(X_t, X_{t-h}) &= Cov\big(E(X_t|X_{t-1}, \dots), E(X_{t-h}|X_{t-1}, \dots)\big) + E\big(Cov(X_t, X_{t-h}|X_{t-1}, \dots)\big) \\
&= Cov(\mu X_{t-1} + \mu_\epsilon, X_{t-h}) \\
&\quad + E\big(E(X_t - E(X_t|X_{t-1}, \dots))(X_{t-h} - E(X_{t-h}|X_{t-1}, \dots))|X_{t-1}, \dots\big) \\
&= \mu \cdot Cov(X_{t-1}, X_{t-h}) + 0 \\
&= \cdots \\
&= \mu^h \cdot \mathrm{Var}(X_{t-h}).
\end{aligned}
$$

Thus, the autocorrelation function $Corr(X_t, X_{t-h}) = \mu^h$. □

Proof of Propositions 2 and 3. Propositions 2 and 3 are similar to Theorems 1 and 2 in [8], which can be proved by verifying the regularity conditions of Theorems 3.1 and 3.2 in [24]. For instance, in the proof of Proposition 2, the partial derivatives $\partial g(\alpha, F_{m-1})/\partial \alpha_i$ have finite fourth moments in [24], which correspond to Assumption 2 in Section 3.1, $u_m^2(\alpha)$ in [24] is corresponds to $q_{1t}(\theta_1)$ in Step 1.1. Hence, Proposition 2 can be regarded as a direct conclusion of Theorem 3.2.

Besides, the proof of Proposition 3 is similar to Proposition 2; the procedure is almost the same as Theorem 3.2 in [24]. □

Proof of Proposition 4. Similarly to the Theorem in [25], based on Theorem 3.2 in [14], we have

$$
\sqrt{n}(\hat{\theta}_{CLS} - \theta_0) \xrightarrow{L} N(0, \Omega),
$$

where

$$
\Omega = \begin{pmatrix} V_1^{-1} W_1 V_1^{-1} & V_1^{-1} M V_2^{-1} \\ V_2^{-1} M^\top V_1^{-1} & V_2^{-1} W_2 V_2^{-1} \end{pmatrix}.
$$

Based on the proof of the Theorem in [25], $\sqrt{q_{1t}(\theta_1)}$ corresponds to u_t and $\sqrt{q_{2t}(\theta_2)}$ corresponds to U_t in [25]. Based on the result of V_1, W_1 in Proposition 2 and V_2, W_2 in Proposition 3, we can obtain $M = E\left(\sqrt{q_{11}(\theta_{1,0})q_{21}(\theta_{2,0})}\frac{\partial}{\partial\theta_1}g_1(\theta_{1,0}, X_0)\frac{\partial}{\partial\theta_2^\top}g_2(\theta_{2,0}, X_0)\right)$. □

Proof of Proposition 5. Since the solutions $(h_1(\mu, \sigma^2), h_2(\mu, \sigma^2))$ about (m, α) in (3.1) are real-valued and have a nonzero differential, Proposition 5 is an application of the δ-method, for example, which can be found in Theorem A on p.122 of [26]. □

References

1. Weiß, C.H. *An Introduction to Discrete-Valued Time Series*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2018.
2. Steutel, F.W.; van Harn, K. Discrete analogues of self-decomposability and stability. *Ann. Probab.* **1979**, *7*, 893–899. [CrossRef]
3. McKenzie, E. Some simple models for discrete variate time series. *Water Resour. Bull.* **1985**, *21*, 645–650. [CrossRef]
4. Al-Osh, M.; Alzaid, A. First-order integer-valued autoregressive (INAR(1)) processes. *J. Time Ser. Anal.* **1987**, *8*, 261–275. [CrossRef]
5. Schweer, S. A goodness-of-fit test for integer-valued autoregressive processes. *J. Time Series Anal.* **2016**, *37*, 77–98. [CrossRef]
6. Chen, C.W.S.; Lee, S. Bayesian causality test for integer-valued time series models with applications to climate and crime data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **2017**, *66*, 797–814. [CrossRef]
7. Zhu, R.; Joe, H. Negative binomial time series models based on expectation thinning operators. *J. Statist. Plann. Inference* **2010**, *140*, 1874–1888. [CrossRef]
8. Yang, K.; Kang, Y.; Wang, D.; Li, H.; Diao, Y. Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued autoregressive processes. *Metrika* **2019**, *82*, 863–889. [CrossRef]
9. Aly, E.; Bouzar, N. Expectation thinning operators based on linear fractional probability generating functions. *J. Indian Soc. Probab. Stat.* **2019**, *20*, 89–107. [CrossRef]
10. Kang, Y.; Wang, D.; Yang, K.; Zhang, Y. A new thinning-based INAR(1) process for underdispersed or overdispersed counts. *J. Korean Statist. Soc.* **2020**, *49*, 324–349. [CrossRef]
11. Balasubramanian, K.; Viveros, R.; Balakrishnan, N. Some discrete distributions related to extended Pascal triangles. *Fibonacci Quart.* **1995**, *33*, 415–425.
12. Du, J.; Li, Y. The integer-valued autoregressive (INAR(p)) model. *J. Time Ser. Anal.* **1991**, *12*, 129–142.
13. Li, Q.; Zhu, F. Mean targeting estimator for the integer-valued GARCH(1,1) model. *Statist. Pap.* **2020**, *61*, 659–679. [CrossRef]
14. Nicholls, D.F.; Quinn, B.G. *Random Coefficient Autoregressive Models: An Introduction*; Springer: New York, NY, USA, 1982.
15. Bakouch, H.S.; Ristić, M.M. A mixed thinning based geometric INAR(1) model. *Metrika* **2010**, *72*, 265–280. [CrossRef]
16. Bisaglia, L.; Gerolimetto, M. Model-based INAR bootstrap for forecasting INAR(p) models. *Comput. Statist.* **2019**, *34*, 1815–1848. [CrossRef]
17. Jung, R.C.; Liesenfeld, R.; Richard, J.-F. Dynamic factor models for multivariate count data: An application to stock-market trading activity. *J. Bus. Econom. Statist.* **2011**, *29*, 73–85. [CrossRef]
18. Zhu, F.; Liu, S.; Shi, L. Local influence analysis for Poisson autoregression with an application to stock transaction data. *Stat. Neerl.* **2016**, *70*, 4–25. [CrossRef]
19. Zhang, H.; Wang, D.; Zhu, F. Inference for INAR(p) processes with signed generalized power series thinning operator. *J. Statist. Plann. Inference* **2010**, *140*, 667–683. [CrossRef]
20. Zhang, H.; Wang, D.; Zhu, F. Generalized RCINAR(1) process with signed thinning operator. *Comm. Statist. Theory Methods* **2012**, *41*, 1750–1770. [CrossRef]
21. Qi, X.; Li, Q.; Zhu, F. Modeling time series of count with excess zeros and ones based on INAR(1) model with zero-and-one inflated Poisson innovations. *J. Comput. Appl. Math.* **2019**, *346*, 572–590. [CrossRef]
22. Liu, Z.; Li, Q.; Zhu, F. Random environment binomial thinning integer-valued autoregressive process with Poisson or geometric marginal. *Braz. J. Probab. Stat.* **2020**, *34*, 251–272. [CrossRef]
23. Qian, L.; Li, Q.; Zhu, F. Modelling heavy-tailedness in count time series. *Appl. Math. Model.* **2020**, *82*, 766–784. [CrossRef]
24. Klimko, LA.; Nelson, P.I. On conditional least squares estimation for stochastic processes. *Ann. Statist.* **1978**, *6*, 629–642. [CrossRef]
25. Zhu, F.; Wang, D. Estimation of parameters in the NLAR(p) model. *J. Time Ser. Anal.* **2008**, *29*, 619–628. [CrossRef]
26. Serfling, R.J. *Approximation Theorems of Mathematical Statistics*; John Wiley & Sons, Inc.: New York, NY, USA, 1980.