

Article

# A New First-Order Integer-Valued Autoregressive Model with Bell Innovations

Jie Huang and Fukang Zhu \* 

School of Mathematics, Jilin University, 2699 Qianjin Street, Changchun 130012, China; jiehuang19@mails.jlu.edu.cn

\* Correspondence: fzhu@jlu.edu.cn

**Abstract:** A Poisson distribution is commonly used as the innovation distribution for integer-valued autoregressive models, but its mean is equal to its variance, which limits flexibility, so a flexible, one-parameter, infinitely divisible Bell distribution may be a good alternative. In addition, for a parameter with a small value, the Bell distribution approaches the Poisson distribution. In this paper, we introduce a new first-order, non-negative, integer-valued autoregressive model with Bell innovations based on the binomial thinning operator. Compared with other models, the new model is not only simple but also particularly suitable for time series of counts exhibiting overdispersion. Some properties of the model are established here, such as the mean, variance, joint distribution functions, and multi-step-ahead conditional measures. Conditional least squares, Yule–Walker, and conditional maximum likelihood are used for estimating the parameters. Some simulation results are presented to access these estimates’ performances. Real data examples are provided.

**Keywords:** Bell distribution; count time series; estimation; INAR; overdispersion



**Citation:** Huang, J.; Zhu, F. A New First-Order Integer-Valued Autoregressive Model with Bell Innovations. *Entropy* **2021**, *23*, 713. <https://doi.org/10.3390/e23060713>

Received: 5 April 2021  
Accepted: 1 June 2021  
Published: 4 June 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent years, studying count time series has attracted a lot of attention in different fields, such as finance, medical science, and insurance. There are many models for count data that have been proposed by scholars. The most famous model was first introduced by McKenzie (1985) [1] and Al-Osh and Alzaid (1987) [2] based on the binomial thinning  $\circ$  operator (Steutel and van Harn 1979 [3]) called the first-order integer-valued autoregressive (INAR(1)) process. Given a non-negative integer-valued random variable (r.v.)  $X$  and a constant  $\alpha \in (0, 1)$ , the binomial thinning operator  $\circ$  is defined as  $\alpha \circ X = \sum_{i=1}^X \zeta_i$ , where the counting series  $\zeta_i$  is a sequence of independent identically distributed (i.i.d.) Bernoulli r.v.s with  $P(\zeta_i = 1) = 1 - P(\zeta_i = 0) = \alpha$ . Then, the form of the INAR(1) model is

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \quad t = 0, 1, 2, \dots, \quad (1)$$

where  $\epsilon_t$  is a sequence of i.i.d. discrete r.v.s, with the mean  $\mu_\epsilon$  and finite variance  $\sigma_\epsilon^2$ .  $\epsilon_t$  is independent of  $\zeta_i$  and  $X_{t-s}$  for  $s \geq 1$ . According to Alzaid and Al-Osh (1988) [4], we know that the mean and variance of the INAR(1) model are

$$\mu := \mu_X = \frac{\mu_\epsilon}{1 - \alpha} \quad \text{and} \quad \sigma^2 := \sigma_X^2 = \frac{\sigma_\epsilon^2 + \alpha\mu_\epsilon}{1 - \alpha^2}, \quad \text{respectively.}$$

For innovation  $\epsilon_t$ , the Poisson distribution is often assumed as the distribution of  $\epsilon_t$  in the INAR(1) model. A natural characteristic of the Poisson distribution is equidispersion; i.e., its mean and variance are equal to each other. In practice, however, many data examples are overdispersed (variance is greater than mean) relative to the Poisson distribution. For this reason, the INAR(1) model with Poisson innovations is not always suitable for modeling integer-valued time series. Therefore, several models which describe the overdispersion phenomena have been discussed in the statistical literature.

One common approach is to change the thinning operation in the INAR(1) model. Weiß (2018) [5] summarized several alternative thinning operators, such as random coefficient thinning, iterated thinning and quasi-binomial thinning. Ristić et al. (2009) [6] proposed the negative binomial thinning operator and defined the corresponding INAR(1) process with geometric marginal distributions. Liu and Zhu (2021) [7] generalized the binomial thinning operator to the extended binomial one.

Changing the distribution of innovations is also used to modify the INAR(1) model. Jung et al. (2005) [8] indicated that the INAR(1) model with negative binomial innovations (NB-INAR(1)) is appropriate for generating overdispersion. Jazi et al. (2012) [9] defined a zero-inflated Poisson ZIP( $\rho, \lambda$ ) for innovations (ZIP-INAR(1)), because a frequent occurrence in overdispersion is that the incidence of zero counts is greater than expected from the Poisson distribution. Jazi et al. (2012) [10] proposed a modification of the INAR(1) model with geometric innovations (G-INAR(1)) for modeling overdispersed count data. Schweer and Weiß (2014) [11] investigated the compound Poisson INAR(1) (CP-INAR(1)) model, which is suitable for fitting datasets with overdispersion. According to Schweer and Weiß (2014) [11], we can also know that the negative binomial distribution and the geometric distribution both belong to the compound Poisson distribution. Livio et al. (2018) [12] presented the INAR(1) model with the Poisson–Lindley innovations, i.e., PL-INAR(1). Bourguignon et al. (2019) [13] introduced the INAR(1) model with the double Poisson (DP-INAR(1)) and generalized Poisson innovations (GP-INAR(1)). Qi et al. (2019) [14] considered zero-and-one inflated INAR(1)-type models, and Cunha et al. (2021) [15] introduced an INAR(1) model with Borel innovations to model zero truncated count time series.

This paper applies the second approach to dealing with overdispersion. Although several models have been proposed in recent years, most of the considered distributions are based on some generalizations of the Poisson distribution and have more than one parameter, such as the zero-inflated Poisson, compound Poisson, double Poisson, and generalized Poisson distributions. Here we use a relatively simple distribution introduced by Castellares et al. (2018) [16] for the innovations, i.e., the Bell distribution. It has the advantages of having only one parameter, belonging to the exponential family, having a simple probability mass function, and having infinite divisibility. Infinite divisibility is significant for constructing the binomial thinning INAR(1) model. Further, the Bell distribution is suitable for modeling some overdispersed count data. Therefore, we introduce a new INAR(1) model with Bell innovations (BL-INAR(1)), which can account for overdispersion in an INAR(1) framework.

In order to observe whether the BL-INAR(1) model has advantages, we compare it with the INAR(1) model with Poisson innovations (P-INAR(1)), G-INAR(1), PL-INAR(1), NB-INAR(1), ZIP-INAR(1), DP-INAR(1), and GP-INAR(1) models. Different information criteria, such as Akaike's information criterion (AIC) [17], the Bayesian information criterion (BIC) [18], the consistent Akaike information criterion (CAIC) [19], and the Hannan–Quinn information criterion (HQIC) [20], are used to compare the above eight models. By comparing the results of different information criteria, it can be seen that the BL-INAR(1) model is competitive when modeling the overdispersed integer-valued time series data, which shows that the proposed BL-INAR(1) model is meaningful; see Section 5 for more details.

We organize the remaining parts of this paper as follows. In Section 2, we briefly review the Bell distribution, including its definition and some properties. Then we propose the BL-INAR(1) model, and its basic properties are constructed; conditional mean and variance are obtained. Section 3 discusses estimates of the model parameters by using the conditional least squares (CLS), Yule–Walker (YW), and conditional maximum likelihood (CML) methods. In Section 4, a numerical simulation of the estimates is presented with some discussions. In Section 5, we compare the proposed model with the other seven INAR(1)-type models when fitting two real data examples, which show the superior performances of the proposed model. The paper concludes in Section 6.

## 2. The BL-INAR(1) Model

In this section, we present a brief review of the Bell distribution (Castellares et al., 2018 [16]). Its definition and some properties are presented. Later we introduce the BL-INAR(1) model and derive some basic properties of it.

### 2.1. The Bell Distribution

At first, we introduce the Bell numbers. Bell (1934) [21] has provided the following expansion:

$$\exp(e^x - 1) = \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n, \quad x \in \mathbb{R},$$

where  $B_n$  is the Bell number defined by

$$B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}. \quad (2)$$

The Bell number  $B_n$  is the  $n$ -th moment of the Poisson distribution with parameter equal to 1. Some Bell numbers are listed as follows.  $B_0 = B_1 = 1, B_2 = 2, B_3 = 5, B_4 = 15, B_5 = 52, B_6 = 203, B_7 = 877, B_8 = 4140, B_9 = 21,147, B_{10} = 115,975, B_{11} = 678,570, B_{12} = 4,213,597$  and  $B_{13} = 27,644,437$ .

For the convenience of the reader, we introduce the following definition and properties of the Bell distribution described in Castellares et al. (2018) [16]:

**Definition 1.** A discrete r.v.  $Z$  taking values in  $\mathbb{N}_0 = \{0, 1, 2, \dots\}$  has a Bell distribution with parameter  $\theta > 0$ , denoted as  $Z \sim \text{Bell}(\theta)$ , if its probability mass function is given by

$$\Pr(Z = z) = \frac{\theta^z e^{-e^\theta + 1} B_z}{z!}, \quad z \in \mathbb{N}_0, \quad (3)$$

where  $B_z$  is the Bell number in (2).

We can see that the Bell distribution has only one parameter, and it belongs to the one-parameter exponential family of distributions. If  $Z \sim \text{Bell}(\theta)$ , the probability generating function is

$$G_Z(s) = E(s^Z) = \exp(e^{s\theta} - e^\theta), \quad |s| < 1.$$

The mean and variance of  $Z$  are

$$E(Z) = \theta e^\theta \text{ and } \text{Var}(Z) = \theta(1 + \theta)e^\theta, \text{ respectively.} \quad (4)$$

Note that  $\text{Var}(Z)/E(Z) = 1 + \theta > 1$ ; hence, the Bell distribution is overdispersed, which means the Bell distribution may be suitable for count data with overdispersion in certain situations.

There are some other interesting properties of the Bell distribution, including the following: (i) the Poisson distribution is not nested in the Bell family, but for small values of the parameter, the Bell distribution approaches the Poisson distribution; (ii) it is identifiable, strongly unimodal and infinitely divisible; (iii) a r.v.  $Z \sim \text{Bell}(\theta)$  has the same distribution as  $Y_1 + Y_2 + \dots + Y_N$ , where  $Y_n$  has zero-truncated Poisson distribution with parameter  $\theta$ , and  $N \sim \text{Poisson}(e^\theta - 1)$ . See Castellares et al. (2018) [16] for more properties.

Additionally, there are some papers based on the Bell distribution, and the following are a few related references: Batsidis et al. (2020) [22] proposed and studied a goodness-of-fit test for the Bell distribution, which is consistent against fixed alternatives; Castellares et al. (2020) [23] presented a new two-parameter Bell-Touchard discrete distribution; Lemonte et al. (2020) [24] introduced a zero-inflated Bell regression model for count data; Muhammad et al. (2021) [25] proposed a Bell ridge regression as a solution to the multicollinearity problems.

### 2.2. Definition and Properties of the BL-INAR(1) Process

In this section, we give the definition of the BL-INAR(1) process, and its basic statistical properties are derived.

**Definition 2.** Let  $\{X_t\}_{t \in \mathbb{N}_0}$  be an INAR(1) process according to (1). It refers to a BL-INAR(1) model if the innovations  $\{\epsilon_t\}_{t \in \mathbb{N}_0}$  are a sequence of i.i.d. Bell( $\theta$ ) r.v.s given by (3); i.e.,

$$\begin{cases} X_t = \alpha \circ X_{t-1} + \epsilon_t, & t \geq 1, \\ \epsilon_t \sim \text{Bell}(\theta), \end{cases} \tag{5}$$

where  $0 < \alpha < 1$  and  $\theta > 0$ , and  $\epsilon_t$  is independent of  $\xi_i$  and  $X_{t-1}$  for  $t \geq 1$ .

According to Equation (4), we know the mean and variance of  $\epsilon_t$  are finite; therefore, the process of  $\{X_t\}_{t \in \mathbb{N}_0}$  in (5) is an ergodic stationary Markov chain (Du and Li, (1991) [26]) with transition probabilities

$$\begin{aligned} P_{ij} &= P(X_t = i | X_{t-1} = j) = P(\alpha \circ X_{t-1} + \epsilon_t = i | X_{t-1} = j) \\ &= \sum_{m=0}^{\min(i,j)} P(\alpha \circ X_{t-1} = m | X_{t-1} = j) P(\epsilon_t = i - m) \\ &= \sum_{m=0}^{\min(i,j)} \binom{j}{m} \alpha^m (1 - \alpha)^{j-m} \frac{\theta^{i-m} e^{-e^\theta + 1} B_{i-m}}{(i-m)!}, \quad i, j = 0, 1, \dots \end{aligned}$$

Further, we can obtain the joint probability function as follows:

$$\begin{aligned} f(i_1, i_2, \dots, i_T) &= P(X_1 = i_1, X_2 = i_2, \dots, X_T = i_T) \\ &= P(X_1 = i_1) P(X_2 = i_2 | X_1 = i_1) \dots P(X_T = i_T | X_{T-1} = i_{T-1}) \\ &= P(X_1 = i_1) \prod_{k=1}^{T-1} \left[ \sum_{m=0}^{\min(i_k, i_{k+1})} \binom{i_k}{m} \alpha^m (1 - \alpha)^{i_k - m} P(\epsilon_{k+1} = i_{k+1} - m) \right]. \end{aligned} \tag{6}$$

The conditional mean, conditional variance, mean, variance, covariance and autocorrelation function of the BL-INAR(1) process are given in the following lemma.

**Lemma 1.** Let  $X_t$  be the process in Definition 2. Then it has the following properties:

- (i)  $E[X_t | X_{t-1}] = \alpha X_{t-1} + \mu_\epsilon = \alpha X_{t-1} + \theta e^\theta$ ;
- (ii)  $\text{Var}[X_t | X_{t-1}] = \alpha(1 - \alpha) X_{t-1} + \sigma_\epsilon^2 = \alpha(1 - \alpha) X_{t-1} + \theta(1 + \theta) e^\theta$ ;
- (iii)  $\mu := E[X_t] = \frac{\theta e^\theta}{1 - \alpha}$ ;
- (iv)  $\sigma^2 := \text{Var}[X_t] = \frac{\theta e^\theta (1 + \alpha + \theta)}{1 - \alpha^2}$ ;
- (v)  $\gamma_k := \text{Cov}(X_k, X_{k+1}) = \alpha^k \sigma^2$ ;
- (vi)  $\rho_k := \text{Corr}(X_k, X_{k+1}) = \alpha^k$ .

The proof of Lemma 1 is similar to that of Theorem 1 of Qi et al. (2019) [14], so it is omitted.

According to Lemma 1, the dispersion index (Fisher, 1950 [27]) of  $X_t$  is derived as follows:

$$I_x := \frac{\sigma^2}{\mu} = 1 + \frac{\theta}{1 + \alpha} > 1;$$

thus, the BL-INAR(1) process is suited for overdispersed integer-valued time series.

Additionally, we can obtain the  $k$ -step ahead conditional mean and  $k$ -step ahead conditional variance of the BL-INAR(1) process in the following theorem.

**Theorem 1.** *The  $k$ -step ahead conditional mean and  $k$ -step ahead conditional variance of the BL-INAR(1) process are given, respectively, by:*

$$E(X_{t+k}|X_t) = \alpha^k X_t + \mu_\epsilon \frac{1 - \alpha^k}{1 - \alpha},$$

and

$$\text{Var}(X_{t+k}|X_t) = \alpha^k (1 - \alpha^k) X_t + \mu_\epsilon \frac{(\alpha - \alpha^k)(1 - \alpha^k)}{1 - \alpha^2} + \sigma_\epsilon^2 \frac{1 - \alpha^{2k}}{1 - \alpha^2}.$$

For more details about the proof of this theorem, see Qi et al. (2019) [14] and Ristić, Bakouch, and Nastić (2009) [6]. It is easy to see that if  $k \rightarrow \infty$ ,  $E(X_{t+k}|X_t) \rightarrow \frac{\mu_\epsilon}{1 - \alpha} = \frac{\theta e^\theta}{1 - \alpha}$  and  $\text{Var}(X_{t+k}|X_t) \rightarrow \frac{\alpha \mu_\epsilon + \sigma_\epsilon^2}{1 - \alpha^2} = \frac{\theta e^\theta (1 + \alpha + \theta)}{1 - \alpha^2}$ , which are the unconditional mean and unconditional variance of  $X_t$ , respectively.

### 3. Estimation of Parameters

The true values of parameters  $\alpha$  and  $\theta$  are unknown in practice; therefore, we need to estimate the value of  $(\alpha, \theta)$ . Sometimes we have to give an estimate of  $(\alpha, \mu)$  first to get the estimate of  $(\alpha, \theta)$ . In this section, we consider three methods for estimating parameters, namely, CLS, YW and CML.

#### 3.1. Conditional Least Squares Estimation

The CLS estimates of the parameters  $\alpha$  and  $\theta$  are obtained by

$$(\hat{\alpha}, \hat{\theta}) = \arg \min \sum_{t=2}^T [X_t - E(X_t|X_{t-1})]^2,$$

and the CLS estimates of  $(\alpha, \mu)$  are given by

$$\hat{\alpha}_{CLS} = \frac{(T - 1) \sum_{t=2}^T X_t X_{t-1} - \sum_{t=2}^T X_t \sum_{t=2}^T X_{t-1}}{(T - 1) \sum_{t=2}^T X_{t-1}^2 - \left( \sum_{t=2}^T X_{t-1} \right)^2},$$

and

$$\hat{\mu}_{CLS} = \frac{\sum_{t=2}^T X_t - \hat{\alpha}_{CLS} \sum_{t=2}^T X_{t-1}}{(T - 1)(1 - \hat{\alpha}_{CLS})}.$$

Then, the CLS estimate of  $\theta$  can be obtained by solving the equation  $\hat{\theta}_{CLS} e^{\hat{\theta}_{CLS}} = \hat{\mu}_{CLS} (1 - \hat{\alpha}_{CLS})$ .

According to Theorems 3.1 and 3.2 in Tjøstheim (1986) [28], we can establish the consistency and asymptotic normality of the CLS estimates  $\hat{\alpha}_{CLS}$  and  $\hat{\mu}_{CLS}$  in the following theorem. The proofs of Theorem 2 and the following theorem are given in Appendix A.

**Theorem 2.** *Let  $\hat{\alpha}_{CLS}$  and  $\hat{\mu}_{CLS}$  be the CLS estimates of the BL-INAR(1) process; then  $(\hat{\alpha}_{CLS}, \hat{\mu}_{CLS})'$  is strongly consistent for  $(\alpha, \mu)$ ; and the asymptotic distribution follows as:*

$$\sqrt{T}(\hat{\alpha}_{CLS} - \alpha, \hat{\mu}_{CLS} - \mu)' \xrightarrow{d} \mathbf{N}(0, \Sigma),$$

where

$$\Sigma = \begin{bmatrix} \frac{\alpha(1+\alpha)\mu + \sigma_\epsilon^2}{(1-\alpha)^2} & \frac{\alpha\sigma^2}{\mu(1+\mu)} \\ \frac{\alpha\sigma^2}{\mu(1+\mu)} & \frac{\alpha(1-\alpha)(\mu_3 - 2\mu\sigma^2 - \mu^3) + \sigma_\epsilon^2\sigma^2}{\mu^2(1+\mu)^2} \end{bmatrix},$$

and  $\mu_3 = E(X_t^3) = \frac{(1-\alpha^3)\mu^3 + (1+2\alpha^2-3\alpha^3)\mu\sigma^2 + \alpha^2(1-\alpha)\sigma^2}{1-\alpha^3}$ .

Using the delta method, we can obtain the limit distribution of  $(\hat{\alpha}, \hat{\theta})$ , and we can also know that  $\hat{\theta}$  is consistent.

### 3.2. Yule–Walker Estimation

Let  $X_1, \dots, X_T$  come from the process  $\{X_t\}$  in Definition 2. The sample mean is  $\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t$ , and the sample autocorrelation function is

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2}.$$

From Lemma 1, we know  $\rho_k = \alpha^k$ , thus the Yule–Walker (YW) estimate of  $\alpha$  is given by

$$\hat{\alpha}_{YW} = \hat{\rho}(1) = \frac{\sum_{t=1}^{T-1} (X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2},$$

and

$$\hat{\mu}_{YW} = \bar{X},$$

with  $\mu = \frac{\theta e^\theta}{1 - \alpha}$ ; then the estimate of  $\theta$  can be obtained.

For asymptotic properties of the YW estimates, Freeland and McCabe (2005) [29] showed that the YW and CLS estimates are asymptotically equivalent for a Poisson INAR(1) process. The next theorem shows that the conclusion holds for our BL-INAR(1) process.

**Theorem 3.** *In the BL-INAR(1) process, CLS and YW estimates are asymptotically equivalent, i.e.,*

$$\hat{\alpha}_{CLS} - \hat{\alpha}_{YW} = o_p(T^{-\frac{1}{2}}) \quad \text{and} \quad \hat{\theta}_{CLS} - \hat{\theta}_{YW} = o_p(T^{-\frac{1}{2}}).$$

### 3.3. Conditional Maximum Likelihood Estimation

According to the joint probability function (6), the likelihood function can be obtained as:

$$\begin{aligned} f(x_1, x_2, \dots, x_T) &= P(X_1 = x_1) \prod_{t=1}^{T-1} P(X_{t+1} = x_{t+1} | X_t = x_t) \\ &= f(x_1) \prod_{t=1}^{T-1} \left[ \sum_{m=0}^{\min(x_t, x_{t+1})} \binom{x_t}{m} \alpha^m (1 - \alpha)^{x_t - m} P(\epsilon_{t+1} = x_{t+1} - m) \right]. \end{aligned}$$

To condition on variable  $X_1$ , we can obtain the conditional log likelihood function as:

$$L(\alpha, \theta) = \sum_{t=1}^{T-1} \log P(X_{t+1} = x_{t+1} | X_t = x_t),$$

the CML estimates of  $(\alpha, \theta)$  are the values of  $(\hat{\alpha}_{CML}, \hat{\theta}_{CML})$  obtained by maximizing the conditional log likelihood function  $L(\alpha, \theta)$ . It is easy to check that the BL-INAR(1) process satisfies conditions (C1)–(C6) of Franke and Seligmann (1993) [30]; thus, the CML estimates  $(\hat{\alpha}_{CML}, \hat{\theta}_{CML})$  are consistent and asymptotically normal. The proof is similar to those of Theorems 22.4 and 22.5 of Franke and Seligmann (1993) [30], so it is omitted.

## 4. Simulation

A Monte Carlo simulation was conducted to study the performances of the CLS, YW, and CML estimates of the BL-INAR(1) model. The CML estimates were obtained by using the BFGS quasi-Newton nonlinear optimization algorithm with numerical derivatives. We considered YW estimates as initial values for the algorithm. The simulation was conducted using R programming language, and the size of the sample was 100, 250, 500, or 1000.

The number of replicates was 1000. For the true values of parameters, we considered  $\alpha = 0.25, 0.5$ , and  $0.75$  and  $\theta = 0.5$ , and  $1.5$ .

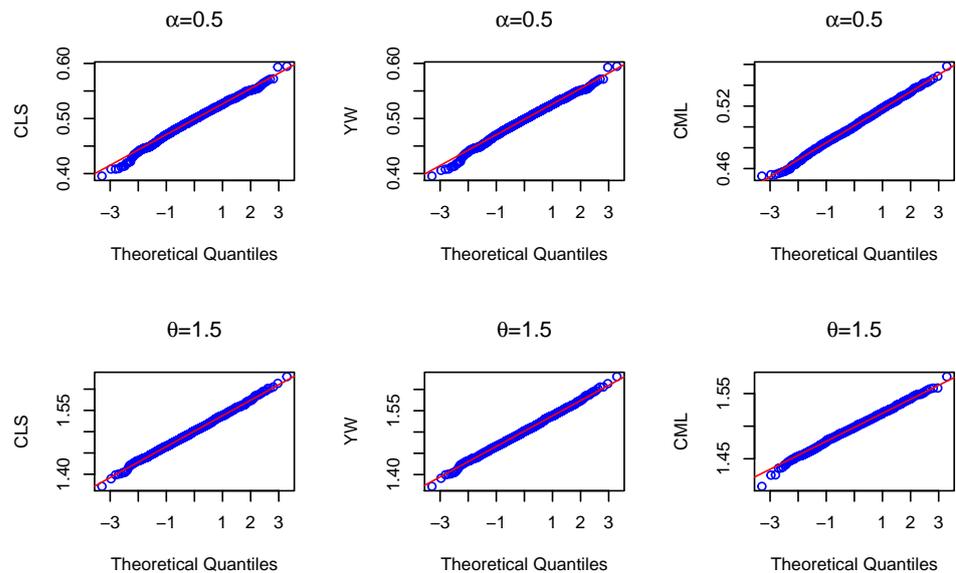
First, we give the Q–Q plots of the CLS, YW, and CML estimates for the BL-INAR(1) model with sample size  $T = 1000$ ,  $\alpha = 0.5$ , and  $\theta = 1.5$  in Figure 1. From the six Q–Q plots, we can see that they contain roughly straight lines; i.e., the estimates of the parameters are normally distributed. Then, the numerical simulation results are presented in Tables 1 and 2. By comparing the two tables, we can find that with the same  $\theta$  and  $T$ , the mean squared error (MSE) for the estimate of  $\theta$  increased with the increase of  $\alpha$ , but the MSE for the estimate of  $\alpha$  decreased. Additionally, the MSE for the estimate of  $\theta$  increased with the increase of  $\theta$  with the same  $\alpha$  and  $T$ , but the MSE for the estimate of  $\alpha$  decreased. Furthermore, we can observe that the estimates of CLS and YW are similar, and the bias tended toward zero for all estimates as the sample size increased. The estimates of CML converged faster to the true parameter values. We conclude that the CML estimates produced the smallest mean square errors, and CML performed better than CLS and YW.

**Table 1.** Empirical means and mean squared errors (in parentheses) of the estimates of the parameters for some values of  $\alpha$  and  $\theta$  of the BL-INAR(1) model.

$T$	$\hat{\alpha}_{CLS}$	$\hat{\theta}_{CLS}$	$\hat{\alpha}_{YW}$	$\hat{\theta}_{YW}$	$\hat{\alpha}_{CML}$	$\hat{\theta}_{CML}$
$(\alpha, \theta) = (0.25, 0.5)$						
100	0.220445 (0.011615)	0.507901 (0.003871)	0.218464 (0.011564)	0.508838 (0.003814)	0.238497 (0.007120)	0.500617 (0.003049)
250	0.240011 (0.004665)	0.502610 (0.001520)	0.239107 (0.004653)	0.503039 (0.001501)	0.246433 (0.002658)	0.500099 (0.001146)
500	0.245254 (0.002286)	0.500707 (0.000758)	0.244710 (0.002284)	0.500965 (0.000759)	0.247766 (0.001300)	0.499777 (0.000603)
1000	0.246458 (0.001195)	0.500868 (0.000379)	0.246197 (0.001197)	0.500974 (0.000380)	0.249120 (0.000714)	0.499768 (0.000290)
$(\alpha, \theta) = (0.5, 0.5)$						
100	0.475430 (0.010198)	0.508229 (0.005396)	0.469977 (0.010256)	0.512782 (0.005296)	0.495566 (0.004046)	0.497046 (0.003083)
250	0.488517 (0.003895)	0.504259 (0.002128)	0.486491 (0.003911)	0.505890 (0.002123)	0.497636 (0.001723)	0.499045 (0.001332)
500	0.493426 (0.001857)	0.502160 (0.001026)	0.492388 (0.001868)	0.503029 (0.001025)	0.498222 (0.000866)	0.499355 (0.000643)
1000	0.496426 (0.000914)	0.501635 (0.000530)	0.495922 (0.000916)	0.502043 (0.000529)	0.499262 (0.000412)	0.499936 (0.000322)
$(\alpha, \theta) = (0.75, 0.5)$						
100	0.714977 (0.006966)	0.535092 (0.011276)	0.707308 (0.007639)	0.543643 (0.011838)	0.745993 (0.001321)	0.500460 (0.003355)
250	0.736256 (0.002354)	0.513360 (0.004357)	0.733057 (0.002456)	0.517222 (0.004432)	0.748974 (0.000494)	0.498915 (0.001352)
500	0.743674 (0.001052)	0.505799 (0.001967)	0.742084 (0.001079)	0.507695 (0.001983)	0.749245 (0.000243)	0.499568 (0.000681)
1000	0.746006 (0.000546)	0.504828 (0.001001)	0.745283 (0.000554)	0.505726 (0.001011)	0.749925 (0.000132)	0.500221 (0.000309)

**Table 2.** Empirical means and mean squared errors (in parentheses) of the estimates of the parameters for some values of  $\alpha$  and  $\theta$  of the BL-INAR(1) model.

$T$	$\hat{\alpha}_{CLS}$	$\hat{\theta}_{CLS}$	$\hat{\alpha}_{YW}$	$\hat{\theta}_{YW}$	$\hat{\alpha}_{CML}$	$\hat{\theta}_{CML}$
$(\alpha, \theta) = (0.25, 1.5)$						
100	0.230059 (0.010409)	1.508538 (0.007500)	0.227601 (0.010294)	1.510489 (0.007323)	0.252786 (0.004877)	1.492707 (0.004375)
250	0.243290 (0.003994)	1.503077 (0.002928)	0.242313 (0.003976)	1.503896 (0.002898)	0.250278 (0.001810)	1.498450 (0.001660)
500	0.244804 (0.001917)	1.503143 (0.001459)	0.244310 (0.001914)	1.503531 (0.001451)	0.249992 (0.000913)	1.499429 (0.000829)
1000	0.248715 (0.000984)	1.500420 (0.000745)	0.248470 (0.000983)	1.500628 (0.000745)	0.251744 (0.000477)	1.498222 (0.000422)
$(\alpha, \theta) = (0.5, 1.5)$						
100	0.472192 (0.008714)	1.522593 (0.011950)	0.467254 (0.008913)	1.528069 (0.011884)	0.497401 (0.002653)	1.497773 (0.004999)
250	0.489125 (0.003127)	1.509054 (0.004609)	0.487244 (0.003148)	1.511225 (0.004598)	0.499745 (0.000991)	1.498361 (0.001856)
500	0.496116 (0.001650)	1.502407 (0.002493)	0.495032 (0.001660)	1.503670 (0.002487)	0.501865 (0.000584)	1.496562 (0.001078)
1000	0.497904 (0.000826)	1.501246 (0.001314)	0.497432 (0.000827)	1.501800 (0.001314)	0.500976 (0.000274)	1.498100 (0.000502)
$(\alpha, \theta) = (0.75, 1.5)$						
100	0.721350 (0.005627)	1.547389 (0.025790)	0.713291 (0.006188)	1.565523 (0.026764)	0.749555 (0.000827)	1.495159 (0.005581)
250	0.736975 (0.002181)	1.522062 (0.011299)	0.733930 (0.002278)	1.529286 (0.011488)	0.749880 (0.000343)	1.497782 (0.002363)
500	0.742692 (0.000919)	1.512717 (0.005007)	0.741144 (0.000947)	1.516338 (0.005076)	0.749888 (0.000158)	1.498329 (0.001101)
1000	0.747785 (0.000476)	1.503485 (0.002670)	0.747046 (0.000479)	1.505296 (0.002670)	0.750224 (0.000083)	1.499187 (0.000541)



**Figure 1.** The Q–Q plots of the CLS, YW, and CML estimates for the BL-INAR(1) model with sample size  $T = 1000$ .

**5. Real Data Examples**

In this section, we present two applications of the BL-INAR(1) model to real datasets, and compare it with the P-INAR(1), G-INAR(1), PL-INAR(1), NB-INAR(1), ZIP-INAR(1),

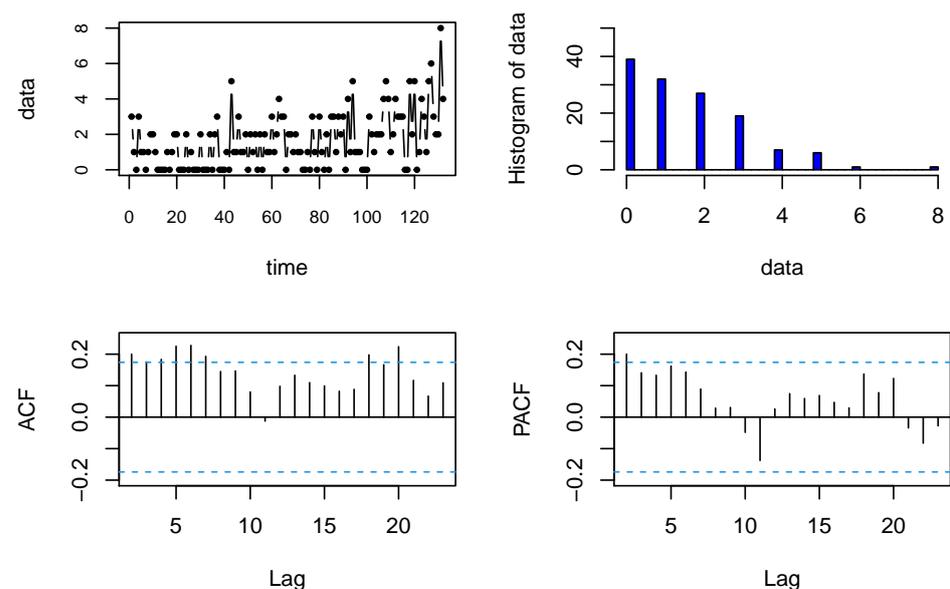
DP-INAR(1), and GP-INAR(1) models. Results of the comparison are discussed here as well.

### 5.1. Disconduct Data

The first dataset is a monthly count of misconduct in the first census tract in Rochester, which can be obtained from Available online: <http://www.forecastingprinciples.com> (accessed on 8 May 2012). The data comprise 132 observations ( $T = 132$ ) starting from January 1991 and ending in December 2001.

The time plot, histogram, autocorrelation function (ACF), and partial autocorrelation function (PACF) are provided in Figure 2. We applied the Ljung–Box test (Ljung and Box (1978) [31]) to check whether this time series dataset has any autocorrelation. The  $p$ -value of the Ljung–Box test is  $1.317 \times 10^{-5}$ , which is less than 0.05. This means that the time series data have some autocorrelation, and according to the PACF diagram, the data are first-order autocorrelated, which shows that the AR(1)-type process is appropriate for modeling this dataset.

The sample mean and variance of the data are  $\bar{X} = 1.6288$  and  $S_X^2 = 2.4455$ , respectively. Thus, we got the dispersion index  $\hat{I}_x = S_X^2/\bar{X} = 1.5014$ . According to the overdispersion test of Schweer and Weiß (2014) [11], the critical value of the data is 1.1994. The dispersion index  $\hat{I}_x$  exceeds the critical value, which means that the equidispersed P-INAR(1) model is not a good choice for the data.



**Figure 2.** The time plot, histogram, ACF, and PACF of misconduct data.

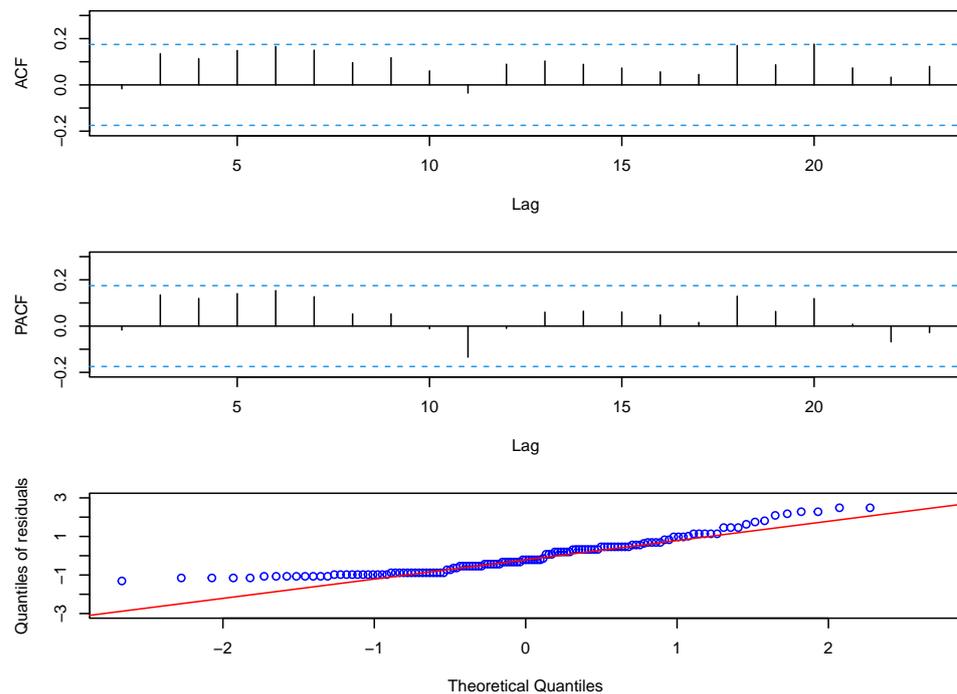
For comparison, we calculated the CML estimates of parameters, and the AIC, BIC, CAIC, HQIC, fitted mean, and fitted variance of the BL-INAR(1) model, the P-INAR(1) model, the G-INAR(1) model, the PL-INAR(1) model, the ZIP-INAR(1) model, the NB-INAR(1) model, the DP-INAR(1) model, and the GP-INAR(1) model. Among the eight models, the first four are two-parameter models and the last four are three-parameter models. The results are presented in Table 3. We found that the AIC, BIC, CAIC, and HQIC of the BL-INAR(1) model were smaller than those of others. We also found that the fitted means of all eight models were near to the sample mean, and the fitted mean of the PL-INAR(1) model was the closest to the sample mean. In terms of fitted variance, Table 3 shows that the fitted variance of the BL-INAR(1) model performed better than those of the other seven models.

**Table 3.** CML estimates, AIC, BIC, CAIC, HQIC, fitted mean, fitted variance, and RMSE for eight INAR(1) models of disconduct data.

Model	Parameters	AIC	BIC	CAIC	HQIC	Mean	Variance	RMSE
BL-INAR	$\hat{\alpha} = 0.1882$ $\hat{\theta} = 0.6718$	<b>441.7380</b> <sup>1</sup>	<b>447.5036</b> <sup>1</sup>	<b>449.5036</b> <sup>1</sup>	<b>444.0809</b> <sup>1</sup>	1.6201	<b>2.5361</b> <sup>2</sup>	3.2205
P-INAR	$\hat{\alpha} = 0.1496$ $\hat{\lambda} = 1.3773$	456.4653	462.2309	464.2309	458.8082	1.6197	2.1512	3.2497
G-INAR	$\hat{\alpha} = 0.2405$ $\hat{\tau} = 0.4482$	446.1416	451.9072	453.9072	448.4845	1.6207	3.2290	<b>3.1803</b> <sup>1</sup>
PL-INAR	$\hat{\alpha} = 0.2197$ $\hat{\theta} = 1.1545$	444.3542	450.1198	452.1198	446.6971	<b>1.6254</b> <sup>2</sup>	0.2928	3.1929
NB-INAR	$\hat{\alpha} = 0.1845$ $\hat{n} = 1.9345$ $\hat{\tau} = 0.5942$	445.4351	454.0835	457.0835	448.9494	1.6201	2.5542	3.2233
ZIP-INAR	$\hat{\alpha} = 0.1992$ $\hat{\lambda} = 1.8674$ $\hat{\rho} = 0.3052$	442.7224	451.3708	454.3708	446.2367	1.6202	2.3903	3.2121
DP-INAR	$\hat{\alpha} = 0.1865$ $\hat{\mu} = 1.2121$ $\hat{\phi} = 0.5122$	443.7622	452.4106	455.4106	447.2765	1.4900	2.6859	3.3155
GP-INAR	$\hat{\alpha} = 0.1820$ $\hat{\mu} = 1.0254$ $\hat{\phi} = 0.2262$	445.8156	454.4640	457.4640	449.3300	1.6200	2.5386	3.2252

<sup>1</sup> Bold text means the smallest value in the column. <sup>2</sup> Bold text means that this value is the closest in the column to the sample value described in the text.

For the prediction, we used the first 126 observations to estimate the parameters, and then predicted the last six observations. The predicted values of the disconduct data could be given by  $E(X_{t+k} | X_t) = \alpha^k X_t + \mu_\epsilon \frac{1 - \alpha^k}{1 - \alpha}$ . For a further comparison of models, we calculated the root mean square values of the prediction errors (RMSEs) for the last 6 months of the data, and the RMSE is defined as  $RMSE = \sqrt{\frac{1}{6} \sum_{k=1}^6 (X_{t+k} - \hat{X}_{t+k})^2}$ . We present the RMSE results of eight models in the last column of Table 3. From the table, we can see that the RMSE of the G-INAR(1) model was best. The RMSE of the BL-INAR(1) model is smaller than those of the P-INAR(1) model, the NB-INAR(1) model, the DP-INAR(1) model, and the GP-INAR(1) model; and a little larger than those of the G-INAR(1) model, the PL-INAR(1) model, and the ZIP-INAR(1) model. Although the fitted mean and RMSE of the BL-INAR(1) model are not the best, it is the best choice under the other five criteria. Further, we analyze the Pearson residuals, and Figure 3 plots the ACF, PACF, and Q-Q plots of residuals. The ACF and PACF graphs show no correlation between residuals, which is supported by the result of the Ljung-Box test with a  $p$ -value of  $0.05251 > 0.05$ . The Q-Q plots appear to be roughly normally distributed, as we expected. Hence, we can conclude that the BL-INAR(1) model is the most suitable among those available for this dataset.



**Figure 3.** The ACF, PACF, and Q–Q plots of the Pearson residual for disconduct data using the BL-INAR(1) model.

### 5.2. Strikes Data

The second dataset, which was analyzed by Weiß (2010) [32], is the monthly number of work stoppages (strikes and lock-outs) of 1000 or more workers for the period 1994–2002. It was published by the US Bureau of Labor Statistics and can be obtained by online at the address Available online: <http://www.bls.gov/wsp/> (accessed on 8 May 2012). The data contain 108 observations, and the time plot, histogram, ACF, and PACF are provided in Figure 4. As with the previous example, the Ljung–Box test was used to check whether the strike data have any autocorrelation. The  $p$ -value of the Ljung–Box test was  $2.372 \times 10^{-8}$ , which shows that the time series data have some autocorrelation, and according to the PACF diagram, it is also first-order autocorrelated, so an AR(1)-type process is appropriate for modeling this dataset.

The sample mean, variance, and dispersion index were calculated to be 4.9444, 7.8488, and 1.5874, respectively. According to the overdispersion test, the critical value of the data is 1.2808, and we observe that it was inappropriate to use the P-INAR(1) model to fit the data. The CML estimates, AIC, BIC, CAIC, HQIC, fitted mean, and fitted variance for the BL-INAR(1), P-INAR(1), G-INAR(1), PL-INAR(1), NB-INAR(1), ZIP-INAR(1), DP-INAR(1), and GP-INAR(1) models were obtained and are shown in Table 4. We see that the AIC, BIC, CAIC, and HQIC of the BL-INAR(1) model are smaller than those of others, and the fitted mean of the BL-INAR(1) model is not much different from those of the other seven models. Further, we can see that the BL-INAR(1) model performed better than others when calculating the fitted variance. Similarly to the previous example, the first 102 observations were used to estimate the parameters and predict the last six observations. The RMSE of the predictions is also presented in Table 4. We can observe that the RMSE of the G-INAR(1) model is the smallest; however, it is only 0.05 less than the RMSE of the BL-INAR(1) model. As in the previous example, although the BL-INAR(1) model was not the best under the fitted mean and RMSE criteria, it performed best under the other five criteria. Additionally, we show the Pearson residuals analysis. Figure 5 gives the ACF, PACF, and Q–Q plots of the residuals. We found that there is no evidence of any significant correlation within the residuals, a finding also supported by the Ljung–Box test with a  $p$ -value of 0.9522, which is greater than 0.05. The Q–Q plot also appears to be roughly normally distributed. Thus,

according to above discussions and its simplicity, we can conclude that the BL-INAR(1) model was the most appropriate.

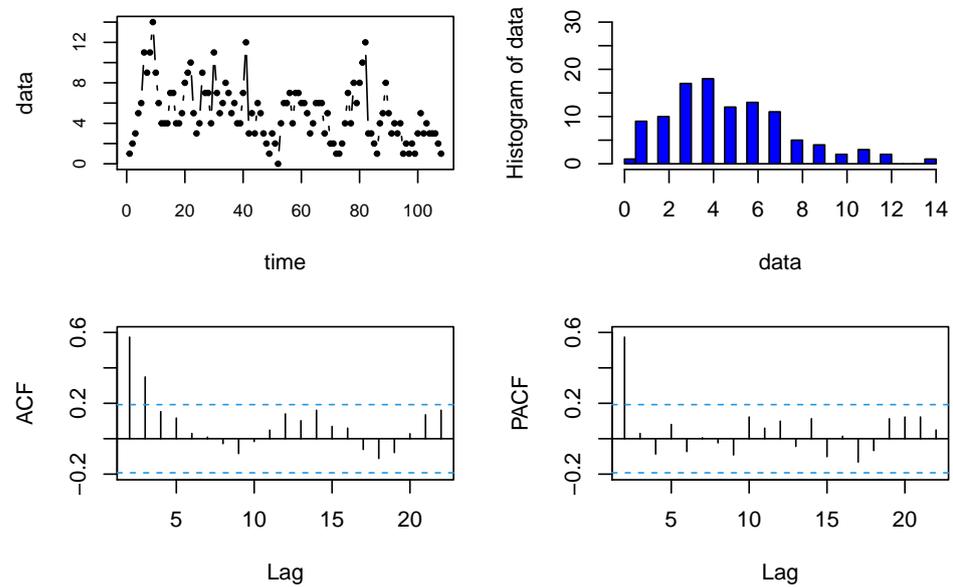


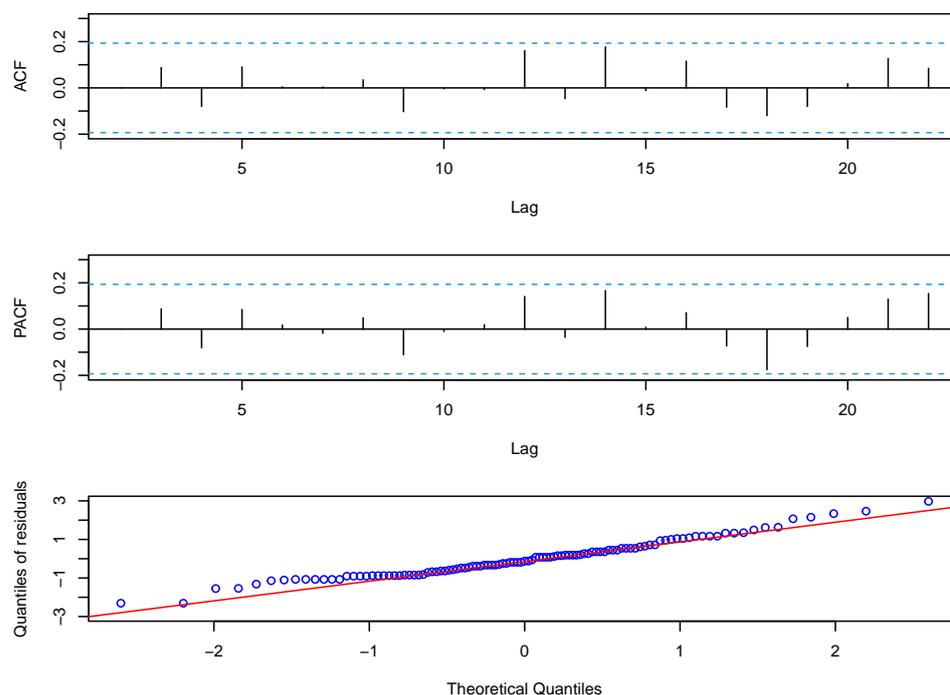
Figure 4. The time plot, histogram, ACF, and PACF of data on strikes.

Table 4. CML estimates, AIC, BIC, CAIC, HQIC, fitted mean, fitted variance, and RMSE from eight INAR(1) models of strike data.

Model	Parameters	AIC	BIC	CAIC	HQIC	Mean	Variance	RMSE
BL-INAR	$\hat{\alpha} = 0.5789$ $\hat{\theta} = 0.8747$	<b>468.1557</b> <sup>1</sup>	<b>473.5199</b> <sup>1</sup>	<b>475.5199</b> <sup>1</sup>	<b>470.3307</b> <sup>1</sup>	4.9813	<b>7.7408</b> <sup>2</sup>	2.2659
P-INAR	$\hat{\alpha} = 0.5061$ $\hat{\lambda} = 2.4603$	473.0936	478.4578	480.4578	475.2686	4.9813	9.8110	2.3331
G-INAR	$\hat{\alpha} = 0.6235$ $\hat{\tau} = 0.3478$	475.3209	480.6852	482.6852	477.4960	4.9813	10.7361	<b>2.2121</b> <sup>1</sup>
PL-INAR	$\hat{\alpha} = 0.6062$ $\hat{\theta} = 0.7911$	471.9345	477.2987	479.2987	474.1095	5.0016	1.8876	2.2489
NB-INAR	$\hat{\alpha} = 0.5483$ $\hat{n} = 3.8582$ $\hat{\tau} = 0.6317$	469.6850	477.7314	480.7314	472.9476	4.9813	6.8573	2.2969
ZIP-INAR	$\hat{\alpha} = 0.5785$ $\hat{\lambda} = 2.6343$ $\hat{\rho} = 0.2030$	470.9985	479.0449	482.0449	474.2610	4.9813	6.6692	2.2663
DP-INAR	$\hat{\alpha} = 0.5617$ $\hat{\mu} = 2.1727$ $\hat{\phi} = 0.5924$	469.5585	477.6048	480.6048	472.8210	<b>4.9576</b> <sup>2</sup>	7.1420	2.2659
GP-INAR	$\hat{\alpha} = 0.5464$ $\hat{\mu} = 1.8003$ $\hat{\phi} = 0.2032$	469.7467	477.7930	480.7930	473.0092	4.9813	6.8335	2.2986

<sup>1</sup> Bold means the smallest value in the column. <sup>2</sup> Bold means that this value is the closest in the column to the sample value described in the text.

Combined with the above two examples and the advantages of the Bell distribution with one parameter and a simple form, the BL-INAR(1) model is competitive with the other seven models.



**Figure 5.** The ACF, PACF, and Q–Q plots of the Pearson residual for strike data using the BL-INAR(1) model.

## 6. Conclusions

A new INAR(1) model with Bell innovations based on the binomial thinning operator was introduced in this paper. Based on the overdispersed property of the Bell distribution, we found that the BL-INAR(1) model is suitable for overdispersed data. Some basic properties of the model were obtained, such as transition probabilities, conditional mean, conditional variance, mean, variance, covariance, autocorrelation function, and  $k$ -step ahead conditional mean and variance. For unknown parameters, CLS, YW, and CML methods are used to estimate them. The Q–Q plots showed that the estimates of the parameters are normally distributed. The simulated results revealed that the CML estimates of parameters of the BL-INAR(1) model were better than the CLS and YW estimates. Finally, by comparing the AIC values, BIC values, CAIC values, HQIC values, fitted means, fitted variances, and RMSE values of the predictions among eight INAR(1) models, two real datasets both showed that the BL-INAR(1) model fits better than other INAR(1) models. The analysis of residuals also shows that the BL-INAR(1) model provided adequate fits to those datasets.

Although there are many overdispersed INAR(1) models, some interesting properties of the Bell distribution, such as having one parameter, infinitely divisibility, having a simple probability mass function, belonging to the one-parameter exponential family of distributions, and for a parameter with a small value, having the Bell distribution approach the Poisson distribution, make the BL-INAR(1) model competitive. Some extended distributions of the Bell distribution, such as the zero-inflated Bell distribution and the Bell–Touchard distribution, provide ideas for us to study related INAR models in the future.

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

**Funding:** Zhu’s work is supported by National Natural Science Foundation of China, grant numbers 11871027 and 11731015.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The disconduct data and the strike data are available at <http://www.forecastingprinciples.com> (accessed on 1 June 2021) and <http://www.bls.gov/wsp/> (accessed on 1 June 2021), respectively.

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

## Appendix A

### Appendix A.1. Proof of Theorem 2

To prove this theorem, we need to show that the conditions given in Theorems 3.1 and 3.2 of Tjøstheim (1986) [28] are satisfied.

Define  $\phi = (\alpha, \mu)$ , and the true value of the unknown parameter  $\phi_0 = (\alpha_0, \mu_0)$ . According to Lemma 1, we know that  $E[X_t^2] < \infty$  and that  $E[X_t | X_{t-1}]$  is almost surely three times differentiable in an open set  $\Phi$  containing  $\phi_0$ .

Condition 1:

$$E \left\{ \left| \frac{\partial E(X_t | X_{t-1})}{\partial \phi_i}(\phi_0) \right|^2 \right\} < \infty \text{ and } E \left\{ \left| \frac{\partial^2 E(X_t | X_{t-1})}{\partial \phi_i \partial \phi_j}(\phi_0) \right|^2 \right\} < \infty, \text{ for } i, j = 1, 2.$$

According to  $E(X_t | X_{t-1}) = \alpha X_{t-1} + (1 - \alpha)\mu$ , we have

$$E \left\{ \left| \frac{\partial E(X_t | X_{t-1})}{\partial \alpha}(\phi_0) \right|^2 \right\} = E \{ |X_{t-1} - \mu_0|^2 \} = \text{Var}(X_{t-1}) < \infty \text{ and}$$

$$E \left\{ \left| \frac{\partial E(X_t | X_{t-1})}{\partial \mu}(\phi_0) \right|^2 \right\} = E \{ |1 - \alpha_0|^2 \} = (1 - \alpha_0)^2 < \infty.$$

For the second derivative of  $E(X_t | X_{t-1})$ , we have

$$E \left\{ \left| \frac{\partial^2 E(X_t | X_{t-1})}{\partial \alpha^2}(\phi_0) \right|^2 \right\} = E \left\{ \left| \frac{\partial^2 E(X_t | X_{t-1})}{\partial \mu^2}(\phi_0) \right|^2 \right\} = 0 < \infty \text{ and}$$

$$E \left\{ \left| \frac{\partial^2 E(X_t | X_{t-1})}{\partial \alpha \partial \mu}(\phi_0) \right|^2 \right\} = 1 < \infty.$$

Condition 2:

The vectors  $\partial E(X_t | X_{t-1})(\theta_0) / \partial \phi_i, i, j = 1, 2$  are linearly independent in the sense that if  $a_1$  and  $a_2$  are arbitrary real numbers such that

$$E \left\{ \left| \sum_{i=1}^2 a_i \frac{\partial E(X_t | X_{t-1})}{\partial \phi_i}(\phi_0) \right|^2 \right\} = 0,$$

then  $a_1 = a_2 = 0$ . Note that

$$E \left\{ \left| a_1 \frac{\partial E(X_t | X_{t-1})}{\partial \alpha}(\phi_0) + a_2 \frac{\partial E(X_t | X_{t-1})}{\partial \mu}(\phi_0) \right|^2 \right\} = 0 \Rightarrow$$

$$E \{ |a_1(X_{t-1} - \mu_0) + a_2(1 - \alpha_0)|^2 \} = 0 \Rightarrow$$

$$\underbrace{a_1^2 \text{Var}(X_{t-1})}_{\geq 0} + \underbrace{a_2^2(1 - \alpha_0)^2}_{\geq 0} = 0 \Rightarrow$$

$$a_1^2 \underbrace{\text{Var}(X_{t-1})}_{> 0} = 0 \text{ and } a_2^2 \underbrace{(1 - \alpha_0)^2}_{> 0} = 0 \Rightarrow$$

$$a_1^2 = 0 \text{ and } a_2^2 = 0.$$

Then  $a_1 = a_2 = 0$ .

Condition 3:

For  $\phi \in \Phi$ , there exist functions  $G_{t-1}^{ijk}(X_1, \dots, X_{t-1})$  and  $H_t^{ijk}(X_1, \dots, X_t)$  for  $i, j = 1, 2$  such that

$$M_{t-1}^{ijk}(\phi) = \left| \frac{\partial E(X_t | X_{t-1})}{\partial \phi_i}(\phi) \frac{\partial^2 E(X_t | X_{t-1})}{\partial \phi_j \partial \phi_k}(\phi) \right| \leq G_{t-1}^{ijk}, E(G_{t-1}^{ijk}) < \infty,$$

$$N_t^{ijk}(\phi) = \left| \{X_t - E(X_t | X_{t-1})\} \frac{\partial^3 E(X_t | X_{t-1})}{\partial \phi_i \partial \phi_j \partial \phi_k}(\phi) \right| \leq H_t^{ijk}, E(H_t^{ijk}) < \infty.$$

Note that  $M_{t-1}^{111}(\phi) = M_{t-1}^{122}(\phi) = M_{t-1}^{211}(\phi) = M_{t-1}^{222}(\phi) = 0$  and

$$M_{t-1}^{112}(\phi) = M_{t-1}^{121}(\phi) = |X_{t-1} - \mu|,$$

$$M_{t-1}^{212}(\phi) = M_{t-1}^{221}(\phi) = |\alpha - 1| < 1;$$

then we can choose  $G_{t-1}^{ijk}(\phi) = (X_{t-1} - \mu)^2 + 1, \forall i, j, k = 1, 2$ , which guarantees that  $M_{t-1}^{ijk}(\phi) < G_{t-1}^{ijk}(\phi)$  and  $E(G_{t-1}^{ijk}) = \text{Var}(X_{t-1} + 1) < \infty$ .

For  $N_t^{ijk}(\phi)$ , it is easy to know that  $N_t^{ijk}(\phi) = 0, \forall i, j, k = 1, 2$ . So we choose  $H_t^{ijk}(\phi) = 0, \forall i, j, k = 1, 2$  to satisfy  $N_t^{ijk}(\phi) < H_t^{ijk}(\phi)$  and  $E(H_t^{ijk}) = 0 < \infty$ .

The above three conditions ensure that  $(\hat{\alpha}_{CLS}, \hat{\mu}_{CLS})$  is a strongly consistent estimator for  $(\alpha, \mu)$ . According to Theorem 3.2 in Tjøstheim (1986) [28], the asymptotic distribution of  $(\hat{\alpha}_{CLS}, \hat{\mu}_{CLS})$  is

$$\sqrt{T}(\hat{\alpha}_{CLS} - \alpha, \hat{\mu}_{CLS} - \mu)' \xrightarrow{d} \mathbf{N}(0, \Sigma),$$

where  $\Sigma = U^{-1}RU^{-1}$ ,

$$U = E \left\{ \frac{\partial E(X_t | X_{t-1})^T}{\partial \phi}(\phi) \cdot \frac{\partial E(X_t | X_{t-1})}{\partial \phi}(\phi) \right\},$$

$$R = E \left\{ \frac{\partial E(X_t | X_{t-1})^T}{\partial \phi}(\phi) f_{t|t-1}(\phi) \frac{\partial E(X_t | X_{t-1})}{\partial \phi}(\phi) \right\},$$

and

$$f_{t|t-1}(\phi) = E \left\{ (X_t - E(X_t | X_{t-1}))(X_t - E(X_t | X_{t-1}))^T | X_{t-1} \right\}.$$

We can then find that

$$\Sigma = \begin{bmatrix} \frac{\alpha(1+\alpha)\mu + \sigma_\epsilon^2}{(1-\alpha)^2} & \frac{\alpha\sigma^2}{\mu(1+\mu)} \\ \frac{\alpha\sigma^2}{\mu(1+\mu)} & \frac{\alpha(1-\alpha)(\mu_3 - 2\mu\sigma^2 - \mu^3) + \sigma_\epsilon^2\sigma^2}{\mu^2(1+\mu)^2} \end{bmatrix},$$

where  $\mu_3 = E(X_t^3) = \frac{(1-\alpha^3)\mu^3 + (1+2\alpha^2-3\alpha^3)\mu\sigma^2 + \alpha^2(1-\alpha)\sigma^2}{1-\alpha^3}$ , which follows from the following derivation:

$$\begin{aligned} \mu_3 &= E[X_t^3] = E[X_t^2(\alpha \circ X_{t-1} + \epsilon_t)] = E[E[X_t^2(\alpha \circ X_{t-1} + \epsilon_t) | X_{t-1}]] \\ &= E[\alpha X_{t-1} E[X_t^2 | X_{t-1}] + \mu_\epsilon E[X_t^2 | X_{t-1}]] \\ &= E[\alpha X_{t-1} E[X_t^2 | X_{t-1}]] + \mu_\epsilon E[X_t^2], \end{aligned}$$

according to Lemma 1,

$$\begin{aligned} E[X_t^2|X_{t-1}] &= \text{Var}[X_t|X_{t-1}] + (E[X_t|X_{t-1}])^2 \\ &= \alpha^2 X_{t-1}^2 + (\alpha(1 - \alpha) + 2\alpha\mu_\epsilon)X_t + \mu_\epsilon^2 + \sigma_\epsilon^2; \end{aligned}$$

then, we have

$$\begin{aligned} \mu_3 &= E[\alpha X_{t-1} E[X_t^2|X_{t-1}]] + \mu_\epsilon E[X_t^2] \\ &= \alpha^3 E[X_{t-1}^3] + \alpha^2(1 - \alpha)E[X_{t-1}^2] + 2\alpha^2\mu_\epsilon E[X_{t-1}] + \alpha\mu(\mu_\epsilon^2 + \sigma_\epsilon^2) + \mu_\epsilon E[X_t^2] \\ &\quad (\text{where } \mu_\epsilon = (1 - \alpha)\mu \text{ and } \sigma_\epsilon^2 = (1 - \alpha^2)\sigma^2 - \alpha(1 - \alpha)\mu) \\ &= \alpha^3\mu_3 + (1 - \alpha^3)\mu^3 + (1 + 2\alpha^2 - 3\alpha^3)\mu\sigma^2 + \alpha^2(1 - \alpha)\sigma^2. \end{aligned}$$

Thus, we obtain that

$$\mu_3 = \frac{(1 - \alpha^3)\mu^3 + (1 + 2\alpha^2 - 3\alpha^3)\mu\sigma^2 + \alpha^2(1 - \alpha)\sigma^2}{1 - \alpha^3}.$$

### Appendix A.2. Proof of Theorem 3

The proof is similar to that of Theorem 4.2 in Cunha et al. (2021) [15]. For estimator  $\hat{\alpha}$ , we have

$$\begin{aligned} \sqrt{T}(\hat{\alpha}_{YW} - \hat{\alpha}_{CLS}) &= \frac{(D_{CLS} - D_{YW})}{D_{CLS}D_{YW}} \cdot \frac{\sum_{t=2}^T X_t X_{t-1}}{\sqrt{T}} \\ &\quad - \frac{(D_{CLS} \sum_{t=2}^T \frac{X_t}{T} - D_{YW} \sum_{t=2}^T \frac{X_t}{T-1})}{D_{CLS}D_{YW}} \cdot \frac{\sum_{t=1}^T X_t}{\sqrt{T}} \\ &\quad - \frac{D_{YW}}{D_{CLS}D_{YW}} \cdot \sum_{t=2}^T \frac{X_t}{T-1} \frac{X_T}{\sqrt{T}} + \frac{D_{CLS}}{D_{CLS}D_{YW}} \cdot \bar{X} \frac{(X_T - \bar{X})}{\sqrt{T}} \\ &= o_p(1)O_p(1) - o_p(1)O_p(1) - O_p(1)O_p(1)o_p(1) + O_p(1)O_p(1)o_p(1) \\ &= o_p(1), \end{aligned}$$

where  $D_{CLS} = \frac{1}{T} \left[ \sum_{t=2}^T X_{t-1}^2 - \frac{1}{T-1} \left( \sum_{t=2}^T X_{t-1} \right)^2 \right]$  and  $D_{YW} = \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^2 = \frac{1}{T} \sum_{t=1}^T X_t^2 - \bar{X}^2$ . For estimator  $\hat{\theta}$ , we only need to prove  $\sqrt{T-1}(\hat{\mu}_{\epsilon,CLS} - \hat{\mu}_{\epsilon,YW})$  is  $o_p(1)$ .

$$\begin{aligned} &\sqrt{T-1}(\hat{\mu}_{\epsilon,CLS} - \hat{\mu}_{\epsilon,YW}) \\ &= \sqrt{T-1} \left( \frac{\sum_{t=2}^T X_t - \hat{\alpha}_{CLS} \sum_{t=2}^T X_{t-1}}{T-1} - \bar{X}(1 - \hat{\alpha}_{YW}) \right) \\ &= \frac{1}{\sqrt{T-1}} \left( \sum_{t=2}^T X_t - \hat{\alpha}_{CLS} \sum_{t=2}^T X_{t-1} - T\bar{X}(1 - \hat{\alpha}_{YW}) + \bar{X}(1 - \hat{\alpha}_{YW}) \right) \\ &= \frac{\hat{\alpha}_{CLS}X_T - X_1}{\sqrt{T-1}} - \sqrt{\frac{T}{T-1}} \cdot \sqrt{T}(\hat{\alpha}_{CLS} - \hat{\alpha}_{YW}) \bar{X} + \frac{\bar{X}}{\sqrt{T-1}}(1 - \hat{\alpha}_{YW}) \\ &= o_p(1) - o_p(1) + o_p(1) = o_p(1). \end{aligned}$$

### References

1. McKenzie, E. Some simple models for discrete variate time series. *Water Resour. Bull.* **1985**, *21*, 645–650. [CrossRef]
2. Al-Osh, M.A.; Alzaid, A.A. First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* **1987**, *8*, 261–275. [CrossRef]
3. Steutel, F.W.; van Harn, K. Discrete analogues of self-decomposability and stability. *Ann. Probab.* **1979**, *7*, 893–899. [CrossRef]
4. Alzaid, A.A.; Al-Osh, M.A. First-order integer-valued autoregressive process: Distributional and regression properties. *Stat. Neerl.* **1988**, *42*, 53–61. [CrossRef]

5. Weiß, C.H. *An Introduction to Discrete-Valued Time Series*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
6. Ristić, M.M.; Bakouch, H.S.; Nastić, A.S. A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *J. Stat. Plan. Inference* **2009**, *139*, 2218–2226. [[CrossRef](#)]
7. Liu, Z.; Zhu, F. A new extension of thinning-based integer-valued autoregressive models for count data. *Entropy* **2021**, *23*, 62. [[CrossRef](#)] [[PubMed](#)]
8. Jung, R.C.; Ronning, G.; Tremayne, A.R. Estimation in conditional first order autoregression with discrete support. *Stat. Pap.* **2005**, *46*, 195–224. [[CrossRef](#)]
9. Jazi, M.A.; Jones, G.; Lai, C.-D. First-order integer valued AR processes with zero inflated Poisson innovations. *J. Time Ser. Anal.* **2012**, *33*, 954–963. [[CrossRef](#)]
10. Jazi, M.A.; Jones, G.; Lai, C.-D. Integer valued AR(1) with geometric innovations. *J. Iran. Stat. Soc.* **2012**, *11*, 173–190.
11. Schweer, S.; Weiß, C.H. Compound Poisson INAR(1) processes: Stochastic properties and testing for overdispersion. *Comput. Stat. Data Anal.* **2014**, *77*, 267–284. [[CrossRef](#)]
12. Livio, T.; Mamode Khan, N.; Bourguignon, M.; Bakouch, H.S. An INAR(1) model with Poisson–Lindley innovations. *Econ. Bull.* **2018**, *38*, 1505–1513.
13. Bourguignon, M.; Rodrigues, J.; Santos-Neto, M. Extended Poisson INAR(1) processes with equidispersion, underdispersion and overdispersion. *J. Appl. Stat.* **2019**, *46*, 101–118. [[CrossRef](#)]
14. 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](#)]
15. Cunha, E.T.D.; Bourguignon, M.; Vasconcellos, K.L.P. On shifted integer-valued autoregressive model for count time series showing equidispersion, underdispersion or overdispersion. *Commun. Stat.-Theory Methods* **2021**. [[CrossRef](#)]
16. Castellares, F.; Ferrari, S.L.P.; Lemonte, A.J. On the Bell distribution and its associated regression model for count data. *Appl. Math. Model.* **2018**, *56*, 172–185. [[CrossRef](#)]
17. Akaike, H. Information theory as an extension of the maximum likelihood principle. In *Proceedings of the Second International Symposium on Information Theory*; Petrov, B.N., Csaki, F., Eds.; Akadémiai Kiado: Budapest, Hungary, 1973; pp. 267–281.
18. Schwarz G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
19. Bozdogan, H. Model selection and Akaike’s Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika* **1978**, *52*, 345–370. [[CrossRef](#)]
20. Hannan, E.J.; Quinn, B.G. The Determination of the Order of an Autoregression. *J. R. Stat. Soc. Ser. B* **1979**, *41*, 190–195. [[CrossRef](#)]
21. Bell, E.T. Exponential polynomials. *Ann. Math.* **1934**, *35*, 258–277. [[CrossRef](#)]
22. Batsidis, A.; Jiménez-Gamero, M.D.; Lemonte, A.J. On goodness-of-fit tests for the Bell distribution. *Metrika* **2020**, *83*, 297–319. [[CrossRef](#)]
23. Castellares, F.; Lemonte, A.J.; Moreno-Arenas, G. On the two-parameter Bell–Touchard discrete distribution. *Commun. Stat.-Theory Methods* **2020**, *49*, 4834–4852. [[CrossRef](#)]
24. Lemonte, A.J.; Moreno-Arenas, G.; Castellares, F. Zero-inflated Bell regression models for count data. *J. Appl. Stat.* **2020**, *47*, 265–286. [[CrossRef](#)]
25. Muhammad, A.; Muhammad, N.A.; Abdul, M. On the estimation of Bell regression model using ridge estimator. *Commun. Stat.-Simul. Comput.* **2021**. [[CrossRef](#)]
26. Du, J.G.; Li, Y. The integer valued autoregressive (INAR( $p$ )) model. *J. Times Ser. Anal.* **1991**, *12*, 129–142.
27. Fisher, R.A. The significance of deviations from expectation in a Poisson series. *Biometrics* **1950**, *6*, 17–24. [[CrossRef](#)]
28. Tjøstheim, D. Estimation in nonlinear time series models. *Stoch. Process. Their Appl.* **1986**, *21*, 251–273. [[CrossRef](#)]
29. Freeland, R.K.; McCabe, B. Asymptotic properties of CLS estimates in the Poisson AR(1) model. *Stat. Probab. Lett.* **2005**, *73*, 147–153. [[CrossRef](#)]
30. Franke, J.; Seligmann, T. Conditional maximum likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. In *Developments in Time Series Analysis*; Rao, T.S., Ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 1993; pp. 310–330.
31. Ljung, G.M.; Box, G.E.P. On a measure of lack of fit in time series models. *Biometrika* **1978**, *65*, 297–303. [[CrossRef](#)]
32. Weiß, C.H. The INARCH(1) model for overdispersed time series of Counts. *Commun. Stat.-Simul. Comput.* **2010**, *39*, 1269–1291. [[CrossRef](#)]