



# Modeling Medical Data by Flexible Integer-Valued AR(1) Process with Zero-and-One-Inflated Geometric Innovations

Zohreh Mohammadi<sup>1</sup> · Zahra Sajjadnia<sup>2</sup> · Maryam Sharafi<sup>2</sup> · Naushad Mamode Khan<sup>3</sup>

Received: 24 January 2022 / Accepted: 8 April 2022 / Published online: 23 May 2022  
© The Author(s), under exclusive licence to Shiraz University 2022

## Abstract

In this paper, we introduce a new stationary first-order integer-valued autoregressive process (INAR) with zero-and-one-inflated geometric innovations that is useful for modeling medical practical data. Basic probabilistic and statistical properties of the model are discussed. Conditional least squares and maximum likelihood estimators are proposed to estimate the model parameters. The performance of the estimation methods is assessed by some Monte Carlo simulation experiments. The zero-and-one-inflated INAR process is subsequently applied to analyze two medical series that include the number of new COVID-19-infected series from Barbados and Poliomyelitis data. The proposed model is compared with other popular competing zero-inflated and zero-and-one-inflated INAR models on the basis of some goodness-of-fit statistics and selection criteria, where it shows to provide better fitting and hence can be considered as another important commendable model in the class of INAR models.

**Keywords** INAR process · Geometric distribution · Zero-and-one-inflated geometric distribution · Binomial thinning operator · Estimation · Runs

## 1 Introduction

Time series of counts are emerging in almost every domain of applications now, be in economics, medicine, or life sciences. Some examples include the monthly cases of crimes and offenses as studied in Bakouch and Ristić (2010), Ristić et al. (2009, 2012), Bourguignon and Vasconcellos (2015), Mamode Khan et al. (2020a), the daily

number of newly infected and deaths due to SARs-Cov 2 patients (Mamode Khan et al. 2020b), the weekly number of syphilis cases, (Bourguignon et al. 2018), the number of daily fatal road traffic accidents (Pedeli and Karlis 2011), the tick by tick intra-day transactions of stocks (Pedeli and Karlis 2013; Sunecher et al. 2018) and amongst others. In such applications, the counting series are usually characterized by frequent low figures that include mainly zeros and ones and this happens mostly when the unit of the collection is at a very micro level. Likewise, the daily SARs-Cov 2 death and newly infected series in small island developing states like Barbados, Guinea-Bissau, Sao Tome consist of mainly 0's and 1's. The same remark can be made to the sex offenses, arsenic, domestic violence data that are available at <http://www.forecastingprinciples.com>. The interested reader may consult more examples in Li et al. (2015) and the references therein. Such excess of zeros or ones leads to overdispersion in the series. This paper, therefore, proposes an integer-valued time series model of auto-regressive nature of order 1 to model such data series but with a zero-and-one-inflated type innovation structure.

✉ Maryam Sharafi  
msharafi@shirazu.ac.ir

Zohreh Mohammadi  
z.mohammadi@jahromu.ac.ir

Zahra Sajjadnia  
sajjadnia@shirazu.ac.ir

Naushad Mamode Khan  
n.mamodekhan@uom.ac.mu

<sup>1</sup> Department of Statistics, Jahrom University, Persian Gulf Boulevard, Jahrom, Fars 7413188941, Iran

<sup>2</sup> Department of Statistics, Shiraz University, Adabiat Four Square, Shiraz, Fars 7145685464, Iran

<sup>3</sup> Department of Economics and Statistics, University of Mauritius, Reduit, Moka 80837, Mauritius

McKenzie (1985) and Al-Osh and Alzadi (1987), independently, introduced the integer-valued autoregressive (INAR) process  $\{X_t\}_{t \in \mathbb{N}_0}$  with one lag using a binomial thinning operator as follows

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \quad t \in \mathbb{N} \quad (1)$$

where  $0 \leq \alpha < 1$ ,  $\{\epsilon_t\}_{t \in \mathbb{N}}$  is a sequence of independent and identically distributed integer-valued random variables, called innovations, with  $\epsilon_t$  independent of  $X_{t-k}$  for all  $k \geq 1$ ,  $E(\epsilon_t) = \mu_\epsilon$  and  $\text{Var}(\epsilon_t) = \sigma_\epsilon^2$ . The binomial thinning operator “ $\circ$ ” is defined by Steutel and van Harn (1979) as  $\alpha \circ X_{t-1} = \sum_{j=1}^{X_{t-1}} Y_j$ , where the counting series  $\{Y_j\}_{j \geq 1}$  is a sequence of independent and identically distributed Bernoulli random variables with  $P(Y_j = 1) = 1 - P(Y_j = 0) = \alpha$ . From the results of Al-Osh and Alzadi (1987), we have that  $\alpha \in [0, 1)$  and  $\alpha = 1$  are the conditions of stationarity and non-stationarity of the process  $\{X_t\}_{t \in \mathbb{N}_0}$ , respectively. Also,  $\alpha = 0$  ( $\alpha > 0$ ) implies the independence (dependence) of the observations of  $\{X_t\}_{t \in \mathbb{N}_0}$ . The following representation for the marginal distribution of the INAR(1) model, provided by Al-Osh and Alzadi (1987), is expressed in terms of the innovation sequence  $\epsilon_t$

$$X_t = \sum_{i=0}^{\infty} \alpha^i \circ \epsilon_{t-i}. \quad (2)$$

Modeling of INAR(1) time series based on (1) was first introduced using the Poisson marginal distribution by Al-Osh and Alzadi (1987) and McKenzie (1988), denoted by PINAR(1). It is a simple model and is appropriate for modeling equidispersed time series data. In many practical scenarios, as discussed above, data are overdispersed. To cater for this phenomenon in counting series, Alzadi and Al-Osh (1988) considered INAR(1) processes with geometric marginal distribution for time series of overdispersed counts. Other useful overdispersed INAR(1) models have been proposed, such as the negative binomial INAR(1) process (McKenzie 1986), generalized Poisson INAR(1) process (Alzadi and Al-Osh 1993). Jazi et al. (2012a) and Schweer and Weiß (2014) introduced an overdispersed INAR(1) model with geometric and compound Poisson innovations, respectively. Basically, there is an ongoing vast literature on the handling of the overdispersion in the simple INAR process (Weiß 2008; Awale et al. 2021; Huang and Zhu 2021; Weiß 2020). However, we note that the construction of the INAR process, in addition to the self-decomposability properties, becomes simpler with assuming the distribution of the innovation series, and without compromising on the marginal distribution of the counting series (See Bourguignon et al. 2019; Livio et al. 2018). In fact, Livio et al. (2018) confirms that

such a later INAR process with the pre-specified innovation yields lower AICs than other competing INAR(1)s in Mohammadpour et al. (2018).

On the other hand, where the data set contains a large number of zeros, Jazi et al. (2012b) introduced an INAR(1) process with zero-inflated Poisson innovations and showed that the marginal distribution of the process is also zero-inflated. However, in the construction of the INAR(1) process, it is not always direct to derive the distribution of the counting series similar to the distribution of the zero-inflated innovation series. In this sense, Barreto-Souza (2015), Bakouch and Ristić (2010) and Bourguignon et al. (2018) studied novel INAR(1) models with zero-modified geometric and zero-truncated Poisson marginal distribution, respectively, similar to the construction process in Livio et al. (2018). Furthermore, Li et al. (2015) developed the mixed INAR(1) process with zero-inflated generalized power series innovations, while Bakouch et al. (2018) investigated the zero-inflated geometric INAR(1) process with random coefficient until recently, Sharafi et al. (2020) proposed the INAR(1) model with zero-modified Poisson–Lindley innovations.

However, when a data set is subject to zero inflation along with one-inflation, the previous models are not very useful. In this research, we restrict our attention to modeling such data. Qi et al. (2019) introduced a stationary INAR(1) process with zero-and-one-inflated Poisson innovations. Also, Mohammadi et al. (2021) introduced the ZOIPLINAR(1) model which is the stationary INAR(1) model with zero-and-one-inflated Poisson–Lindley distributed innovations. The geometric distribution is one of the most important distributions used to analyze count data. Many authors such as McKenzie (1986), Ristić et al. (2009, 2012), Jazi et al. (2012a, b) used geometric distribution to analyze count time series data. This fact motivated us to introduce the flexible INAR(1) model with zero-and-one-inflated geometric innovations to model count data, especially in the analyzing of the COVID-19 real data.

It should be mentioned that in the COVID-19 data time series analysis based on the PACF plot, in most applications, it seems that order 1 is not suitable and the higher-order time series are needed. Recently, Foroughi et al. (2021) introduced a new portmanteau test to examine the null hypothesis  $H_0 : X_t \sim \text{GINAR}(1)$  versus the alternative  $H_1 : X_t \sim \text{GINAR}(p)$  for  $p > 1$  and a wide group of INAR processes, called generalized INAR. They developed some portmanteau test statistics to check the adequacy of the fitted model. In this paper, we use the above test statistics to check the adequacy of our introduced model which is applied to the practical data example.

The paper is organized as follows. In Sect. 2, we introduce and construct a flexible INAR(1) model and

obtain some of its statistical and conditional properties. Section 3 is devoted to parameter estimation of the model which is included two estimation methods, maximum likelihood, and conditional least square estimators. In Sect. 4, we present some simulation experiments and real-life data applications to assess the performance of the proposed zero-and-one-inflated INAR model.

## 2 Model Construction and Properties

In this section, we introduce a flexible INAR(1) process with zero-and-one-inflated geometric-distributed innovations denoted by INARZOIG(1) and present some of its properties.

Based on the Eq. (1), we define the INARZOIG(1) as follow:

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \quad t \in \mathbb{N} \tag{3}$$

where  $0 < \alpha < 1$  and the innovation process  $\{\epsilon_t\}$  is said to have zero-and-one-inflated geometric distribution, denoted by ZOIG( $\phi_0, \phi_1, \theta$ ), with the following probability mass function (pmf),

$$P(\epsilon_t = k) = \begin{cases} \phi_0 + (1 - \phi_0 - \phi_1) \frac{1}{1 + \theta}, & \text{if } k = 0 \\ \phi_1 + (1 - \phi_0 - \phi_1) \frac{\theta}{(1 + \theta)^2}, & \text{if } k = 1 \\ (1 - \phi_0 - \phi_1) \frac{1}{1 + \theta} \left(\frac{\theta}{1 + \theta}\right)^k, & \text{if } k = 2, 3, \dots, \end{cases} \tag{4}$$

where  $0 \leq \phi_0, \phi_1 < 1, \theta > 0$ . The parameter  $\theta$  is the mean of the traditional geometric distribution and the parameters  $\phi_0$  and  $\phi_1$  denote the unknown proportions for incorporating extra zeros and ones than those allowed by the considered a traditional geometric distribution, respectively. Also,  $\epsilon_t$  is independent of  $X_s$  for all  $t > s$  and it is independent of the counting series contained in the binomial thinning operator “ $\circ$ .”

Based on Du and Li (1991) and Dion et al. (1995), it can be easily shown that this process is stationary if and only if  $0 < \alpha < 1$ . This process is reduced into INARZIG(1) when  $\phi_1 = 0$  and INAROIG(1) when  $\phi_0 = 0$ , respectively.

In the following proposition, some moments and conditional moments of the INARZOIG(1) process are summarized for the coming use.

**Proposition 1** *Let  $\{X_t\}$  be the process defined by (3). Then*

- a)  $E(X_t|X_{t-1}) = \alpha X_{t-1} + \mu_\epsilon$
- b)  $\mu = E(X_t) = \frac{\mu_\epsilon}{1 - \alpha}$ ,
- c)  $\text{Var}(X_t|X_{t-1}) = \alpha(1 - \alpha)X_{t-1} + \sigma_\epsilon^2$ ,
- d)  $\sigma^2 = \text{Var}(X_t) = \frac{\alpha\mu_\epsilon + \sigma_\epsilon^2}{1 - \alpha^2}$ ,

- e)  $\gamma(k) = \text{Cov}(X_t, X_{t-k}) = \alpha^k \gamma(0)$ ,
- $\rho(X_t, X_{t-k}) = \text{Corr}(X_t, X_{t-k}) = \alpha^k$ ,
- f)  $\psi_{X_t}(s) = \prod_{i=0}^{\infty} \left[ \phi_0 + \phi_1(1 - \alpha^i + \alpha^i s) + \frac{\phi_2}{1 + \theta \alpha^i - \theta \alpha^i s} \right]$ ,

where  $\phi_2 = 1 - \phi_0 - \phi_1, \mu_\epsilon = E(\epsilon_t) = \phi_1 + \phi_2 \theta$  and  $\sigma_\epsilon^2 = \text{Var}(\epsilon_t) = \mu_\epsilon - \mu_\epsilon^2 + 2\phi_2 \theta^2$ .

The proof of Proposition 1 is similar to Theorem 1 in Qi et al. (2019), we omit the details here and refer the reader to Qi et al. (2019).

Using conditional mean is one of the most common techniques for forecasting time series processes. In the next proposition, conditional mean and variance of INARZOIG(1) process is obtained.

**Proposition 2** *For INARZOIG(1) process, the  $(h + 1)$ -step ahead forecast which is conditional mean, and the conditional variance are*

$$\hat{X}_{t+h} = E(X_{t+h}|X_{t-1}) = \alpha^{h+1} X_{t-1} + \frac{(1 - \alpha^{h+1})\mu_\epsilon}{(1 - \alpha)}, \tag{5}$$

and

$$\text{Var}(X_{t+h}|X_{t-1}) = \alpha^{h+1}(1 - \alpha^{h+1})X_{t-1} + \frac{\sigma_\epsilon^2(1 - \alpha^{2(h+1)})}{1 - \alpha^2} + \frac{\mu_\epsilon(1 - \alpha^h)(\alpha - \alpha^{h+2})}{1 - \alpha^2}, \tag{6}$$

respectively.

**Proof** The proof of Proposition 2 is given in Appendix.□

It is clear that  $E(X_{t+h}|X_{t-1}) \rightarrow \frac{\mu_\epsilon}{(1 - \alpha)}$  as  $h \rightarrow \infty$ , which is the unconditional mean of the process. Also, the  $(h + 1)$ -step ahead conditional variance converges to  $\frac{\alpha\mu_\epsilon + \sigma_\epsilon^2}{1 - \alpha^2} = \sigma^2$  as  $h \rightarrow \infty$ .

According to Proposition 1, the Fisher index of dispersion for the model can be calculated as

$$FI_X = \frac{\sigma^2}{\mu} = \frac{\alpha + FI_\epsilon}{1 + \alpha}, \tag{7}$$

where  $FI_\epsilon = 1 - \mu_\epsilon + 2 \frac{\theta^2 \phi_2}{\mu_\epsilon}$ , then the dispersion of the INARZOIG(1) process is similar to the dispersion of its innovation process, i.e., it is overdispersed (underdispersed) if the innovations  $\{\epsilon_t\}$  is overdispersed (underdispersed).

**Remark 1** After some calculation, it is easy to show that  $\epsilon_t$  is overdispersed if  $\theta > \frac{\phi_1}{\sqrt{2\phi_2 - \phi_2}}$ , it is underdispersed if  $0 < \theta < \frac{\phi_1}{\sqrt{2\phi_2 - \phi_2}}$  and it is equidispersed if  $\theta = \frac{\phi_1}{\sqrt{2\phi_2 - \phi_2}}$ .

Since the model (3) forms a stationary discrete-time Markov chain, the transition probabilities obtained as (see, e.g., Weiß 2008):

$$P_{ij} = P(X_t = j | X_{t-1} = i) = \sum_{k=0}^{\min(i,j)} \binom{i}{k} \alpha^k (1-\alpha)^{i-k} P(\epsilon_t = j-k) \tag{8}$$

where  $P(\epsilon_t = j)$  is the pmf of  $\{\epsilon_t\}$  defined by (4) and  $i, j = 0, 1, \dots$

Hence, the marginal probability function of  $X_t$  of INARZOIG(1) is obtained as:

$$P_j = P(X_t = j) = \sum_{i=0}^{\infty} P(X_t = j | X_{t-1} = i) P(X_{t-1} = i) \\ = \sum_{i=0}^{\infty} P_{ij} P_i = \sum_{i=0}^{\infty} \sum_{k=0}^{\min(i,j)} \binom{i}{k} \alpha^k (1-\alpha)^{i-k} P(\epsilon_t = j-k) P_i, \quad j = 0, 1, \dots \tag{9}$$

Also, the joint probability of the processes using the first-order dependence can be calculated as:

$$f(i_1, i_2, \dots, i_n) = P(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) \\ = P_{i_1} \prod_{s=1}^{n-1} \left\{ \sum_{k=0}^{\min(i_s, i_{s+1})} \binom{i_s}{k} \alpha^k (1-\alpha)^{i_s-k} P(\epsilon_t = i_{s+1} - k) \right\}. \tag{10}$$

### 2.1 Distributions of Lengths of Zeros and Ones

Mood (1940) presented a definition of the number of the “succession” of similar events preceded and succeeded by different events, and called it “the Run.” In this section, we find the expected length of the runs of zeros and the lengths of the runs of ones for the INARZOIG(1) process.

**Theorem 1** *The expected length of the runs of zeros for the INARZOIG(1) process is*

$$\eta_0 = \frac{1 + \theta}{(1 + \theta)\phi_1 + \theta\phi_2} \tag{11}$$

and the expected length of the runs of ones for the INARZOIG(1) process is

$$\eta_1 = \frac{(1 + \theta)^2}{(1 + \theta)^2 [1 - \alpha\phi_0 - (1 - \alpha)\phi_1] - \phi_2(\theta + \alpha)}. \tag{12}$$

**Proof** The zero-to-zero transition probability for the INARZOIG(1) process is obtained as:

$$P_{00} = P(X_t = 0 | X_{t-1} = 0) = P(\epsilon_t = 0) = \phi_0 + \frac{\phi_2}{1 + \theta}. \tag{13}$$

Therefore, the transition probabilities from zero to nonzero for the INARZOIG(1) process can be obtained as

$$p^* = P(X_t \neq 0 | X_{t-1} = 0) = 1 - \left\{ \phi_0 + \frac{\phi_2}{1 + \theta} \right\} \\ = \phi_1 + \frac{\theta\phi_2}{1 + \theta}. \tag{14}$$

Since the run length of zeros is defined as the number of zeros between two nonzero values, it can be shown that it follows from a geometric distribution with the parameter  $p^*$ , and hence, the expected run length of zeros in the process is  $\frac{1}{p^*}$ . The expected run length of ones can be obtained similarly.  $\square$

The expected length of the runs of zeros is independent of  $\alpha$ . If  $\phi_0 = 0$  or  $\phi_1 = 0$ , we obtain the expected length of the runs for the INARZOIG(1) or INARZIG(1) process, respectively.

**Theorem 2** *The proportion of zeros in the INARZOIG(1) process is given by*

$$P(X_t = 0) = \prod_{i=0}^{\infty} \left[ \phi_0 + \phi_1(1 - \alpha^i) + \frac{\phi_2}{1 + \theta\alpha^i} \right], \tag{15}$$

and the proportion of ones in the INARZOIG(1) process is

$$P(X_t = 1) = \sum_{j=0}^{\infty} \left[ \left( \phi_1 \alpha^j + \frac{\phi_2 \theta \alpha^j}{(1 + \theta\alpha^j)^2} \right) \prod_{i \neq j} \left[ \phi_0 + \phi_1(1 - \alpha^i) + \frac{\phi_2}{1 + \theta\alpha^i} \right] \right]. \tag{16}$$

**Proof** Using part (f) of Proposition 1 and based on the following relationship between the probability generating function (pgf) and pmf,

$$P(X_t = k) = \frac{1}{k!} \psi_{X_t}^{(k)}(0)$$

where  $\psi_{X_t}^{(k)}(\cdot)$  denotes the  $k$ th derivative of the pgf  $\psi_{X_t}(\cdot)$ , the proof is completed by calculating the following statements.

$$P(X_t = 0) = \psi_{X_t}^{(0)}(0) = \psi_{X_t}(0),$$

and

$$P(X_t = 1) = \psi_{X_t}^{(1)}(0).$$

$\square$

### 3 Parameter Estimation

Let  $X = (X_1, \dots, X_n)$  be observations from the model (3) and  $\lambda \in \Lambda = \{(\alpha, \theta, \phi_0, \phi_1)'; 0 < \alpha < 1, 0 \leq \phi_0 < 1, 0 \leq \phi_1 < 1, \theta > 0\}$  denote the parameter vector. In the study of integer-valued time series, different estimation methods are applied. In this section, we are going to estimate the parameters of the INARZOIG(1) model using conditional maximum likelihood (CML) and conditional least squares (CLS) estimation methods.

#### 3.1 Conditional Maximum Likelihood Estimation

For simplicity of notations, we can write the likelihood function through the joint probability function (10) as

$$P(X_1 = x_1) \prod_{i=1}^{n-1} P_\lambda(X_{i+1} = x_{i+1} | X_i = x_i) = P_{X_1} \prod_{i=1}^{n-1} \left\{ \sum_{k=0}^{\min(X_i, X_{i+1})} \binom{X_i}{k} \alpha^k (1-\alpha)^{X_i-k} P(\epsilon_i = X_{i+1} - k) \right\}, \tag{17}$$

where  $P_{X_1}$  is the pmf of  $X_1$  and  $P_\lambda(X_{i+1}|X_i)$  is the conditional pmf. To overcome the complexity of the marginal distribution, a simple approach is to find the conditional pmf conditioned on the first observation  $X_1$ , essentially ignoring the dependency on the initial value and obtain the conditional maximum likelihood (CML) estimate given  $X_1$  as an estimate of  $\lambda$  by maximizing the conditional log-likelihood.

$$l_n(\lambda|X_1) = \sum_{i=1}^n \log P_\lambda(X_{i+1}|X_i), \tag{18}$$

over  $\lambda$ . Since there is no closed form for the CML estimates, these estimates are achieved using numerical methods. The asymptotic properties of the CML estimators follow from Freeland and McCabe (2004).

#### 3.2 Conditional Least Squares Estimation

In this subsection, we describe the estimation of the unknown parameters of the INARZOIG(1) process using the two-step CLS estimation method proposed by Karlsen and Tjøstheim (1988) which is conducted by the following two steps.

##### Step 1

Let  $\beta_1 = (\alpha, \mu_\epsilon)'$ , then the conditional least square (CLS) estimators of the parameters  $\alpha$  and  $\mu_\epsilon$  are obtained by minimizing the function

$$Q_n(\beta_1) = \sum_{t=2}^n [X_t - g_1(\beta_1, X_{t-1})]^2, \tag{19}$$

where  $g_1(\beta_1, X_{t-1}) = E(X_t|X_{t-1}) = \alpha X_{t-1} + \mu_\epsilon$ , and are given by

$$\hat{\alpha}_{cls} = \frac{(n-1) \sum_{t=2}^n X_t X_{t-1} - \sum_{t=2}^n X_t \sum_{t=2}^n X_{t-1}}{(n-1) \sum_{t=2}^n X_{t-1}^2 - (\sum_{t=2}^n X_{t-1})^2}, \tag{20}$$

and

$$\hat{\mu}_{\epsilon,cls} = \frac{\sum_{t=2}^n (X_t - \hat{\alpha}_{cls} X_{t-1})}{n-1}. \tag{21}$$

##### Step 2

Let  $\beta_2 = (\phi_0, \phi_1)'$  and  $Y_t = X_t - E(X_t|X_{t-1})$ ,  $g_2(\beta_2, X_{t-1}) = \text{Var}(X_t|X_{t-1})$ . Then

$$E(Y_t^2|X_{t-1}) = E([X_t - E(X_t|X_{t-1})]^2|X_{t-1}) = \text{Var}(X_t|X_{t-1}) = g_2(\beta_2, X_{t-1})$$

where  $\text{Var}(X_t|X_{t-1}) = \hat{\alpha}_{cls}(1 - \hat{\alpha}_{cls})X_{t-1} + \hat{\mu}_{\epsilon,cls} - \hat{\mu}_{\epsilon,cls}^2 + \frac{2(\hat{\mu}_{\epsilon,cls} - \phi_1)^2}{(1 - \phi_0 - \phi_1)}$ . Therefore, the CLS criterion function for  $\beta_2$  can be written as

$$Q_n^*(\beta_2) = \sum_{t=2}^n [Y_t^2 - g_2(\beta_2, X_{t-1})]^2 \tag{22}$$

The CLS estimator  $\hat{\beta}_{2,cls} = (\hat{\phi}_{0,cls}, \hat{\phi}_{1,cls})'$  of  $\beta_2$  are obtained by numerical solution of (22).

##### Step 3

Based on the results from Steps 1 and 2, the estimator  $\hat{\theta}_{cls}$  of  $\theta$  can be obtained by considering the following equation:

$$\hat{\mu}_{\epsilon,cls} = \hat{\phi}_{1,cls} + (1 - \hat{\phi}_{0,cls} - \hat{\phi}_{1,cls})\theta,$$

Therefore, the resulting CLS estimators is  $(\hat{\alpha}_{cls}, \hat{\theta}_{cls}, \hat{\phi}_{0,cls}, \hat{\phi}_{1,cls})'$ . To study the asymptotic behavior of the estimators, we make the following assumptions,

- (C1)  $X_t$  is a stationary and ergodic process.
- (C2)  $E(X_t^4) < \infty$ .

**Proposition 3** Under the assumptions (C1) and (C2), the CLS estimator  $\hat{\beta}_{1,cls} = (\hat{\alpha}_{cls}, \hat{\mu}_{\epsilon,cls})'$  is strongly consistent and asymptotically normal,

$$\sqrt{n}(\hat{\beta}_{1,cls} - \beta_{1,0}) \xrightarrow{L} N(\mathbf{0}, V_1^{-1} W_1 V_1^{-1})$$

where  $\beta_{1,0} = (\alpha_0, \mu_{\epsilon,0})'$  denote the true value of  $\beta_1$ ,  $V_1 = E[\frac{\partial}{\partial \beta_1} g_1(\beta_{1,0}, X_1) \frac{\partial}{\partial \beta_1} g_1(\beta_{1,0}, X_1)]$ ,

$$q_{11}(\beta_{1,0}) = [X_2 - g_1(\beta_{1,0}, X_1)]^2, \quad \text{and} \quad W_1 = E \left[ q_{11}(\beta_{1,0}) \frac{\partial}{\partial \beta_1} g_1(\beta_{1,0}, X_1) \frac{\partial}{\partial \beta_1} g_1(\beta_{1,0}, X_1) \right].$$

**Proposition 4** Under the assumptions (C1) and (C2), the CLS estimator  $\hat{\beta}_{2,\text{cls}} = (\hat{\phi}_{0,\text{cls}}, \hat{\phi}_{1,\text{cls}})'$  is strongly consistent and asymptotically normal,

$$\sqrt{n}(\hat{\beta}_{2,\text{cls}} - \beta_{2,0}) \xrightarrow{L} N(\mathbf{0}, V_2^{-1}W_2V_2^{-1})$$

where  $\beta_{2,0} = (\phi_{0,0}, \phi_{1,0})'$  denotes the true value of  $\beta_2$ ,  $V_2 = E[\frac{\partial}{\partial \beta_2} g_2(\beta_{2,0}, X_1) \frac{\partial}{\partial \beta_2} g_2(\beta_{2,0}, X_1)']$ ,  $q_{21}(\beta_{2,0}) = [Y_2^2 - g_2(\beta_{2,0}, X_1)]^2$ , and  $W_2 = E[q_{21}(\beta_{2,0}) \frac{\partial}{\partial \beta_2} g_1(\beta_{2,0}, X_1) \frac{\partial}{\partial \beta_2} g_2(\beta_{2,0}, X_1)']$ .

Based on Propositions 3 and 4 and Theorem 3.2 in Nicholls and Quinn (1982), we have the following proposition.

**Proposition 5** Under the assumptions (C1) and (C2), the CLS estimator  $\hat{\beta}_{\text{cls}} = (\hat{\beta}_{1,\text{cls}}, \hat{\beta}_{2,\text{cls}})'$  is strongly consistent and asymptotically normal,

$$\sqrt{n}(\hat{\beta}_{\text{cls}} - \beta_0) \xrightarrow{L} N(\mathbf{0}, \Omega),$$

where

$$\Omega = \begin{bmatrix} V_1^{-1}W_1V_1^{-1} & V_1^{-1}MV_2^{-1} \\ V_2^{-1}MV_1^{-1} & V_2^{-1}W_2V_2^{-1} \end{bmatrix},$$

$M = E[\sqrt{q_{11}(\beta_{1,0})q_{21}(\beta_{2,0})} \frac{\partial}{\partial \beta_1} g_1(\beta_{1,0}, X_1) \frac{\partial}{\partial \beta_2} g_2(\beta_{2,0}, X_1)]$ , and  $\beta_0 = (\beta_{1,0}, \beta_{2,0})'$  denotes the true value of  $\beta$ .

Based on the above proposition, we state the strong consistency and asymptotic normality of  $\hat{\lambda}_{\text{cls}}$  in the following proposition.

**Proposition 6** Under the assumptions (C1) and (C2), the CLS estimator  $\hat{\lambda}_{\text{cls}}$  is strongly consistent and asymptotically normal,

$$\sqrt{n}(\hat{\lambda}_{\text{cls}} - \lambda_0) \xrightarrow{L} N(\mathbf{0}, D\Omega D')$$

where  $\lambda_0 = (\alpha_0, \theta_0, \phi_{0,0}, \phi_{1,0})'$  denotes the true values of  $\lambda = (\alpha, \theta, \phi_0, \phi_1)'$  and

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\phi_2} & \frac{1}{\phi_2^2} & \frac{\phi_0 + \phi_1 + \theta\phi_2 - 1}{\phi_2^2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and  $\phi_2 = 1 - \phi_0 - \phi_1$ .

The brief proofs of Propositions 3–6 are given in Appendix.

## 4 Numerical Illustration

This part of the paper includes two subsections. In the first part, the performance of the estimation methods, which are presented in the previous section, is evaluated through a simulation study. Moreover, the empirical distribution of the simulated sample path in points zero and one are compared with the results of the Eqs. (15) and (16). To ensure the practical performance of the proposed process, the second part is focused on two real-life application series: the number of daily infected cases due to COVID-19 in Barbados, available in <https://ourworldindata.org/covid-cases> and the Poliomyelitis data from Zeger (1988) and Maiti et al. (2018).

### 4.1 Simulation

To conduct the simulation study, we need to generate a random sample from the INARZOIG(1) process. Based on the second stochastic representation in Zhang et al. (2016), we first generate a random sample  $\epsilon_1, \dots, \epsilon_n$  from ZOIG( $\phi_0, \phi_1, \frac{\theta}{1+\theta}$ ) and then simulate  $\{X_t\}_{t=1}^n$  from INARZOIG(1) model. The simulation comprised the following steps:

**Step 1** Generate  $Z_1, \dots, Z_n$  form Bernoulli( $1 - \phi$ ),

**Step 2.** From Bernoulli( $p$ ) generate  $\eta_1, \dots, \eta_n$ ,

**Step 3.** From Ge( $\frac{\theta}{1+\theta}$ ) generate  $T_1, \dots, T_n$ ,

**Step 4.** Use  $\epsilon_i = (1 - Z_i)\eta_i + Z_iT_i$  for  $i = 1, \dots, n$ , generate  $\epsilon_1, \dots, \epsilon_n$ ,

where  $\phi_0 = \phi(1 - p)$  and  $\phi_1 = \phi p$ .

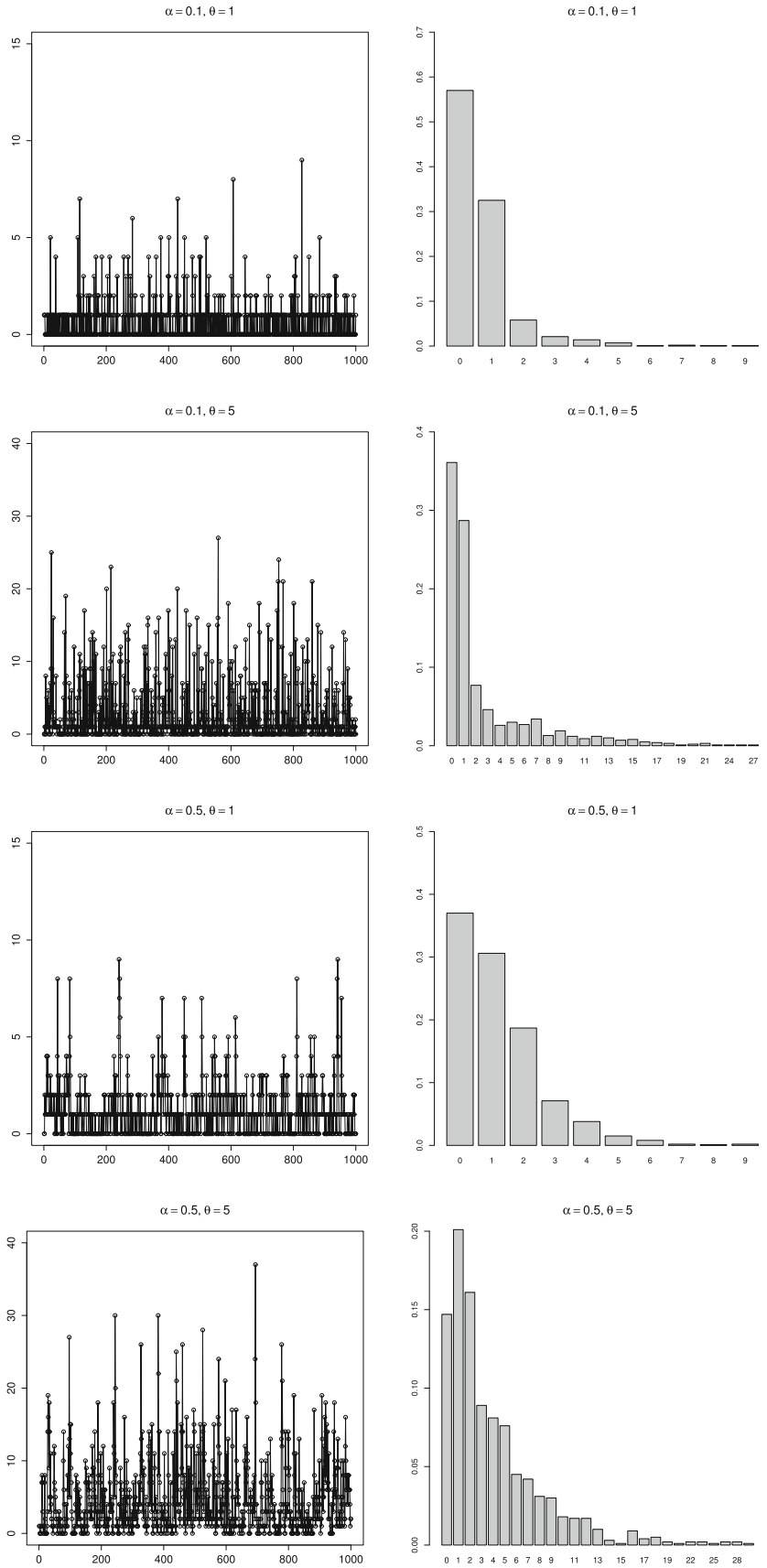
According to the above algorithm, we generate a random sample (with  $n = 1000$ ) from the INARZOIG(1) process with  $\phi_0 = 0.4$ ,  $\phi_1 = 0.2$ ,  $\theta = 1, 5$  and  $\alpha = 0.1, 0.5$ . The sample path and barplot of the marginal distribution of this simulated count time series is presented in Fig. 1.

As can be seen from Fig. 1, for all values of  $\alpha$  and larger values of  $\theta$ , the sample path tends to have larger values. But for all values of  $\alpha$  and smaller values of  $\theta$ , the process has a strong tendency to return to zero or one values with less mean and variance which is clear from parts (b) and (d) of Proposition 1.

In addition, Fig. 1 shows that the number of zeros and ones increases by decreasing the values of  $\theta$ .

To compare the performance of the CML and the CLS estimators, we simulate the data for  $n = 50, 100, 200, 500, 1000$ ,  $\alpha = 0.2$ ,  $\phi_0 = 0.1, 0.4$ ,  $\phi_1 = 0.1$  and  $\theta = 1, 3$  with 10,000 replications. Mean and mean squared error (MSE) of the estimates are computed to evaluate the estimates. The function “nminb” in “R” is used to obtaining these estimates. The results of the simulation are given in Tables 1 and 2. These tables show that the CML estimate is

**Fig. 1** Barplots of limiting marginal distribution and sample paths of the simulated INARZOIG(1) process for  $\phi_0 = 0.4, \phi_1 = 0.2, \theta = 1, 5$  and  $\alpha = 0.1, 0.5$



performed better than CLS estimate because of smaller MSE (except for a few cases).

In Table 3, we compare the empirical distribution of the simulated sample path with Eqs. (15) and (16) and it can be seen that for different values of  $n$  and other parameters of the model the estimated values of the proportion of zeros and ones are near to the theoretical values of them.

### 4.2 Real Data

In this subsection, using two real-life data sets, we show the applicability of the INARZOIG(1). In the first example, we use the data of new infected cases in Barbados from March 17, 2020, until January 02, 2021, and in the second, we considered the Poliomyelitis data which are the monthly cases in the USA from 1970 to 1983. To compare INARZOIG(1) model with various INAR(1) models such as OMINAR(1) (one modified geometric INAR(1) model), PINAR(1) (INAR(1) process with Poisson-distributed innovations), ZIPINAR(1) (INAR(1) process with zero-inflated Poisson-distributed innovations),

OIPINAR(1) (INAR(1) process with one-inflated Poisson-distributed innovations), ZOIPINAR(1) (INAR(1) process with zero-one-inflated Poisson-distributed innovations), ZOIPLINAR(1) (INAR(1) process with zero-one-inflated Poisson–Lindley distributed innovations), INARG(1) (INAR(1) process with geometric-distributed innovations), INARZIG(1) (INAR(1) process with zero-inflated geometric-distributed innovations) and INAROIG(1) (INAR(1) process with one-inflated geometric-distributed innovations) for these data sets. We use the AIC (Akaike information criterion), loglik (log-likelihood function), AICc (corrected version of the AIC), BIC (Bayesian information criterion), PMAE( $h$ ) (predicted mean absolute error), and the PTP( $h$ ) (percentage of true prediction ) criteria where the last two criteria are the  $h$ -step ahead forecasting accuracy measures.

To calculate the last two measures, we divide the data into two parts. The first part is used to fit the considered models, and the second part which is the last 20 observations is used to compute the  $\hat{X}_{t+h}$  and then the PMAE( $h$ ) and PTP( $h$ ) are computed for  $h = 1$ .

**Table 1** Mean and MSE for CML and CLS estimators for  $\theta = 1$

$n$	Method	$\alpha$		$\phi_0$		$\phi_1$		$\theta$	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
$\alpha = 0.2, \phi_0 = 0.1, \phi_1 = 0.1$									
50	CML	0.1914	0.0121	0.1287	0.0146	0.1244	0.0111	1.0248	0.0384
	CLS	0.1994	0.0147	0.0994	0.0033	0.1010	0.0033	1.0062	0.0928
100	CML	0.1965	0.0069	0.1134	0.0083	0.1096	0.0059	1.0104	0.0155
	CLS	0.1916	0.0094	0.0992	0.0032	0.0996	0.0033	1.1016	0.0518
200	CML	0.1977	0.0035	0.1051	0.0049	0.1025	0.0033	1.0039	0.0076
	CLS	0.1915	0.0054	0.0998	0.0032	0.1007	0.0033	1.0144	0.0286
500	CML	0.1992	0.0014	0.1007	0.0022	0.1008	0.0013	1.0004	0.0026
	CLS	0.1955	0.0022	0.0997	0.0033	0.1001	0.0033	1.0101	0.0151
1000	CML	0.1993	0.0006	0.0995	0.0011	0.1002	0.0006	1.0001	0.0012
	CLS	0.1978	0.0011	0.0999	0.0033	0.1000	0.0032	1.0088	0.0104
$\alpha = 0.2, \phi_0 = 0.4, \phi_1 = 0.1$									
50	CML	0.1853	0.0139	0.4024	0.0198	0.1115	0.0080	1.0062	0.0201
	CLS	0.1981	0.0152	0.2507	0.0428	0.0997	0.0033	0.8215	0.1700
100	CML	0.1918	0.0071	0.3998	0.0109	0.1061	0.0046	0.9999	0.0094
	CLS	0.1902	0.0100	0.2499	0.0437	0.1003	0.0033	0.8245	0.1242
200	CML	0.1967	0.0036	0.3985	0.0060	0.1027	0.0025	0.9984	0.0056
	CLS	0.1924	0.0059	0.2488	0.0437	0.1006	0.0033	0.8169	0.0977
500	CML	0.1978	0.0014	0.3983	0.0023	0.1005	0.0009	0.9989	0.0024
	CLS	0.1960	0.0025	0.2507	0.0431	0.0985	0.0033	0.8145	0.0826
1000	CML	0.1997	0.0006	0.3998	0.0011	0.1003	0.0004	0.9995	0.0011
	CLS	0.1986	0.0012	0.2506	0.0431	0.0992	0.0032	0.8123	0.0782



**Table 2** Mean and MSE for CML and CLS estimators for  $\theta = 3$

$n$	Method	$\alpha$		$\phi_0$		$\phi_1$		$\theta$	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
$\alpha = 0.2, \phi_0 = 0.1, \phi_1 = 0.1$									
50	CML	0.1912	0.0124	0.1292	0.0144	0.1217	0.0111	1.0259	0.0393
	CLS	0.1943	0.0136	0.1005	0.0033	0.0992	0.0033	3.0492	0.6482
100	CML	0.1944	0.0067	0.1125	0.0081	0.1102	0.0059	1.0100	0.0161
	CLS	0.1918	0.0088	0.0993	0.0033	0.1005	0.0033	3.0630	0.3971
200	CML	0.1993	0.0016	0.1014	0.0037	0.1010	0.0028	3.0014	0.0020
	CLS	0.1925	0.0049	0.0996	0.0033	0.1004	0.0033	3.0539	0.2405
500	CML	0.1995	0.0006	0.0999	0.0015	0.1006	0.0012	2.9992	0.0004
	CLS	0.1961	0.0021	0.1004	0.0033	0.0996	0.0033	3.0403	0.1432
1000	CML	0.1996	0.0003	0.1002	0.0007	0.1002	0.0005	2.9996	0.0001
	CLS	0.1982	0.0010	0.1003	0.0033	0.0993	0.0033	3.0319	0.1055
$\alpha = 0.2, \phi_0 = 0.4, \phi_1 = 0.1$									
50	CML	0.1973	0.0064	0.3951	0.0139	0.1089	0.0066	2.9966	0.0135
	CLS	0.1912	0.0146	0.2505	0.0431	0.0991	0.0033	2.4848	1.3653
100	CML	0.1987	0.0031	0.3978	0.0067	0.1043	0.0036	2.9991	0.0022
	CLS	0.1874	0.0091	0.2521	0.0427	0.1002	0.0033	2.4954	0.9963
200	CML	0.1991	0.0014	0.3992	0.0033	0.1014	0.0019	2.9999	0.0004
	CLS	0.1919	0.0051	0.2513	0.0042	0.0993	0.0033	2.4793	0.8224
500	CML	0.2001	0.0005	0.4002	0.0013	0.1009	0.0007	2.9998	0.0001
	CLS	0.1965	0.0021	0.2473	0.0443	0.0994	0.0033	2.4504	0.7604
1000	CML	0.2002	0.0002	0.3997	0.0006	0.1004	0.0003	3.0001	0.00001
	CLS	0.1985	0.0011	0.2479	0.0440	0.1003	0.0033	2.4443	0.7251

**Table 3** Estimated values of the proportion of zeros and ones in the simulated data from INARZOIG(1) processes for different values of  $n$

$n$	$\hat{P}_0$	$\hat{P}_1$	$n$	$\hat{P}_0$	$\hat{P}_1$
$\alpha = 0.2, \phi_0 = 0.1, \phi_1 = 0.1, \theta = 1$ $P(X_t = 0) = 0.4, P(X_t = 1) = 0.31$			$\alpha = 0.2, \phi_0 = 0.1, \phi_1 = 0.1, \theta = 3$ $P(X_t = 0) = 0.18, P(X_t = 1) = 0.21$		
50	0.4	0.32	50	0.18	0.16
100	0.38	0.30	100	0.21	0.19
200	0.41	0.31	200	0.17	0.21
500	0.39	0.33	500	0.18	0.22
1000	0.41	0.31	1000	0.18	0.21
$\alpha = 0.2, \phi_0 = 0.4, \phi_1 = 0.1, \theta = 1$ $P(X_t = 0) = 0.57, P(X_t = 1) = 0.26$			$\alpha = 0.2, \phi_0 = 0.4, \phi_1 = 0.1, \theta = 3$ $P(X_t = 0) = 0.39, P(X_t = 1) = 0.22$		
50	0.60	0.24	50	0.40	0.20
100	0.55	0.27	100	0.40	0.30
200	0.57	0.26	200	0.36	0.26
500	0.58	0.27	500	0.41	0.22
1000	0.58	0.27	1000	0.40	0.23

**4.2.1 COVID-19 Data in Barbados**

In this subsection, using a real data set, we show the applicability of the INARZOIG(1). We use the data of new infected cases in Barbados from the 17th of March 2020 until the 2nd of January 2021. This data set has 292 observations for which 148 (51%) of observations are zero

and 64 (22%) of observations are one, and the other 80 (27%) of observations had infected cases more than one. The mean and variance of observations are 1.35 and 5.60, respectively, and hence, the Fisher index of them is given as 4.15 and it shows that the data are overdispersed. The barplot, series plot, ACF and PACF are plotted in Figs. 2 and 3, respectively. It is noted that the PACF yields

significant lags with values greater than one, and it seems that the INAR with an order greater than 1 is suitable for the data. To determine the order of the process, we use the portmanteau tests introduced by Foroughi et al. (2021) with  $m = 2, 3, 4, 8, 12$ . We want to test the null hypothesis that the data set follows the INAR(1) versus the alternative hypothesis that the data follow the INAR( $p$ ) with  $p > 1$ . The  $p$  values are reported in the Table 4 and show that order 1 is appropriate for this data set.

The relative frequencies of zeros and ones and Fig. 2 show the inflation in zeros and ones. The extra zeros and ones motivated us to use INARZOIG for this data set. To show the adequacy of the INARZOIG(1), we compare the proposed model with OMGINAR(1), PINAR(1), ZIPI-NAR(1), OIPINAR(1), ZOIPINAR(1), ZOIPLINAR(1), INARG(1), INARZIG(1) and INAROIG(1) based on the mentioned criteria. The results in the Table 5, show that the INARZOIG(1) has the best fit since it has the largest Loglik and smallest values of other criteria except for BIC which are indicated by bold numbers. In the sense of BIC, INARZIG(1) is the best model and to show that INARZOIG(1) is more suitable for this data set than the INARZIG(1), we use the likelihood ratio test (LRT) with the following hypothesis.

$$\begin{cases} H_0 : \phi_1 = 0 \\ H_1 : \phi_1 \neq 0 \end{cases} \quad (23)$$

The LRT statistics is equal to 3.937 and the critical value at level 0.05 is equal to 3.841. Hence, we can conclude that the null hypothesis rejects and the zero-and-one-inflated distribution is more suitable than zero-inflated model for

this data set. Also, we calculated two forecasting accuracy measures; however, they are the same for all models and PMAE is equal to 4.45 and PTP is equal to 20.

The last figure shows the daily new infected cases of COVID-19 in Barbados and their predicted values using INARZOIG(1). As can be seen, the predicted values are closed to the original data, which indicates the good performance of the proposed fitted model in the sense of forecasting (Fig. 4).

### 4.2.2 Poliomyelitis Data

In this subsection, we considered the Poliomyelitis data which are the monthly cases in the USA from 1970 to 1983. These data were analyzed by Zeger (1988) for the first time. This data set has 168 observations for which 64 (38%) of observations are zero and 55 (32%) of observations are one, and the other 49 (30%) of observations had monthly cases more than one. The mean and variance of observations are 1.33 and 3.50, respectively, and hence, the Fisher index of them is given as 2.63. The value of the Fisher index indicates that the data are overdispersed. Recently, Maiti et al. (2018) considered these data and fitted most of the existing INAR(1) models including Poisson INAR(1), overdispersed models such as geometric INAR(1) and compound Poisson INAR(1), zero-inflated models like zero-inflated and zero-modified INAR(1) and their proposed sub-model, the one-modified geometric INAR(1)(OMGINAR(1)). Using some goodness of fit criteria and 1-step ahead forecasting accuracy measures, they showed that OMGINAR(1) had the best fit among all considered models.

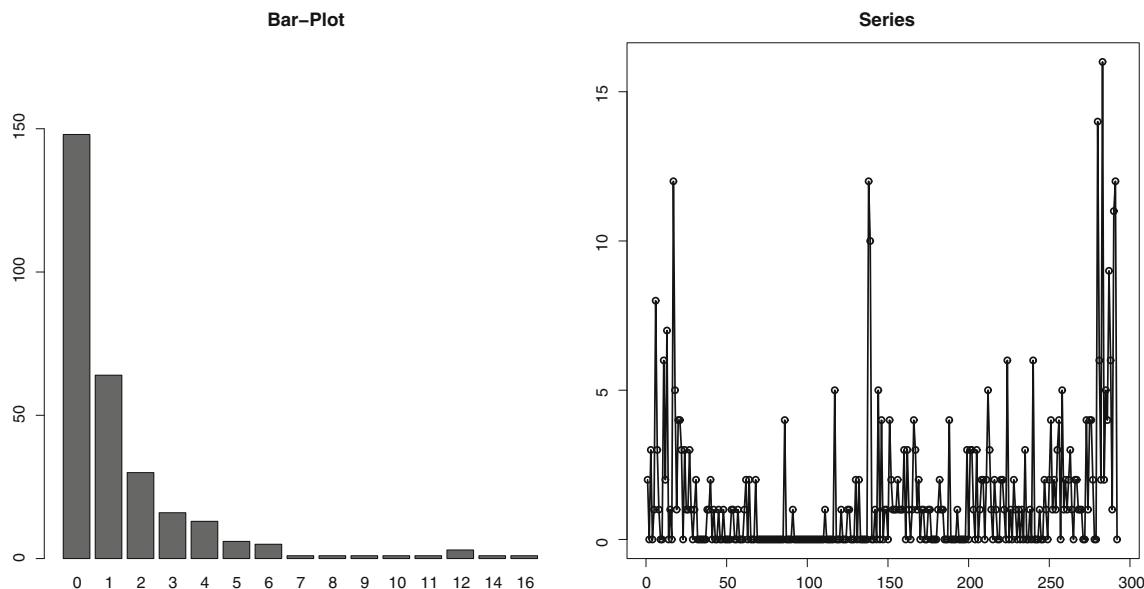
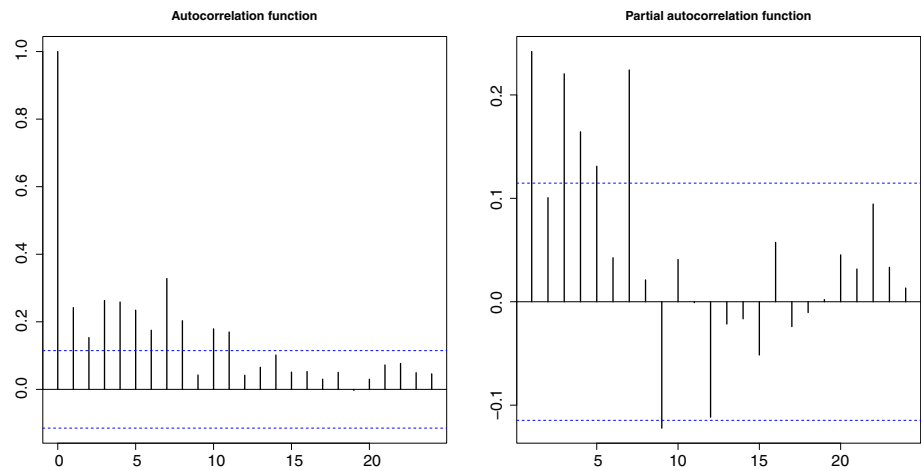


Fig. 2 Barplot and series plot of the new infected cases in Barbados

**Fig. 3** ACF and PACF of the new infected cases in Barbados



**Table 4** The  $p$  values of the portmanteau tests for different values of  $m$

$m$	Test statistics		
	$Q_{LB}$	$Q_{BP}$	$Q_{LM}$
2	0.3669	0.3703	0.3661
3	0.4645	0.4386	0.4381
4	0.3779	0.3829	0.3824
8	0.3458	0.3528	0.3522
12	0.4927	0.4995	0.4986

Now, we analyze the data further. First, we plot the barplot and the series plot in Fig. 5. These figures and the frequencies of the observed zeros and ones show the extra number of zeros and ones. This fact and the overdispersion of the data, motivated us to fit the INARZOIG(1) model into this data set. The ACF and PACF of the data are plotted in Fig. 6.

Based on the conclusions of Maiti et al. (2018) about the considered data set, we compare our model with OMGINAR(1) and used the reported criteria in that paper for this model. Also, we considered the ZOIPLINAR(1), introduced by Mohammadi et al. (2021), as another alternative to compare with. We use the Loglik, AIC, AICc, and BIC criteria, and the results are reported in Table 6. As can be seen, the INARZOIG(1) model has the largest Loglik and smallest AIC, AICc, but the value of the BIC of the OMGINAR(1) is the smallest BIC. Nevertheless, based on Raftery (1995), since the difference between these values is less than 2, it is not significant and the other criteria show that the INARZOIG(1) is more suitable for this data set. We can conclude that our introduced model has the best fit on this data set; however, the forecasting accuracy measures are the same when PMAE is equal to 0.95 and PTP is equal to 45 for all considered models. Moreover, from Fig. 7 that shows the plot of the Poliomyelitis data and

their predicted values, it can be seen that the predicted values are found to be almost close to the real data. This figure indicates the good performance of the INARZOIG(1) in the sense of forecasting, too.

### 5 Conclusion

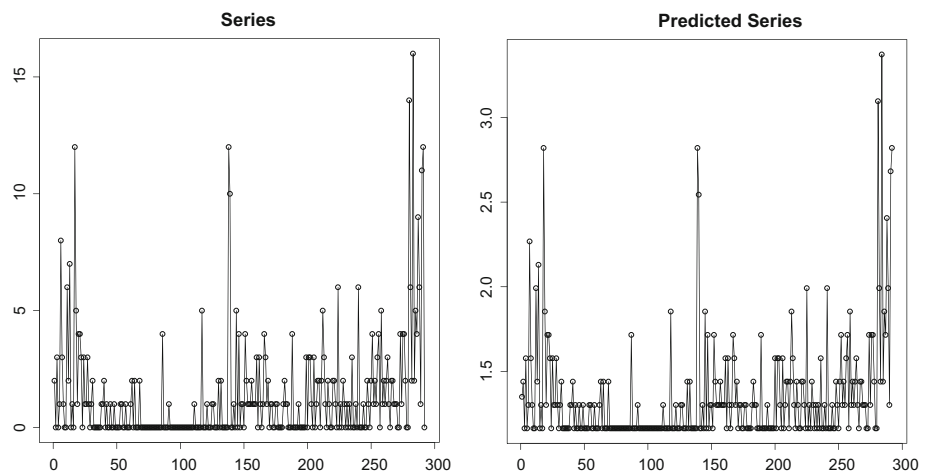
This paper analyzes the zero-and-one-inflated time series using a flexible INAR(1) model based on the zero-and-one-inflated geometric-distributed innovation. The main properties of this novel INAR(1) process are established, and its model parameters are estimated via the CML and CLS approaches. The performance of the two estimation techniques is assessed through some Monte Carlo experiments wherein both approaches are shown to provide consistent estimates, but with the CML approach providing lesser biased estimates.

Furthermore, the INARZOIG(1) model is applied to analyze the COVID-19 series from Barbados, which is found to consist of a more frequent number of zeros and ones. Also, using the portmanteau test we indicate that order 1 is suitable for this data set. As a next example, this model is applied to another real data set which is the monthly cases in the USA from 1970 to 1983 that was analyzed by Zeger (1988) for the first time. Under the data applications, the INARZOIG(1) model is shown to provide better fitting criteria than the existing competing models. Evidently, the statistical performance of the INARZOIG(1) depends on the nature of the data as well, but overall, the INARZOIG(1) model has a worthy contribution to the class of INAR models.

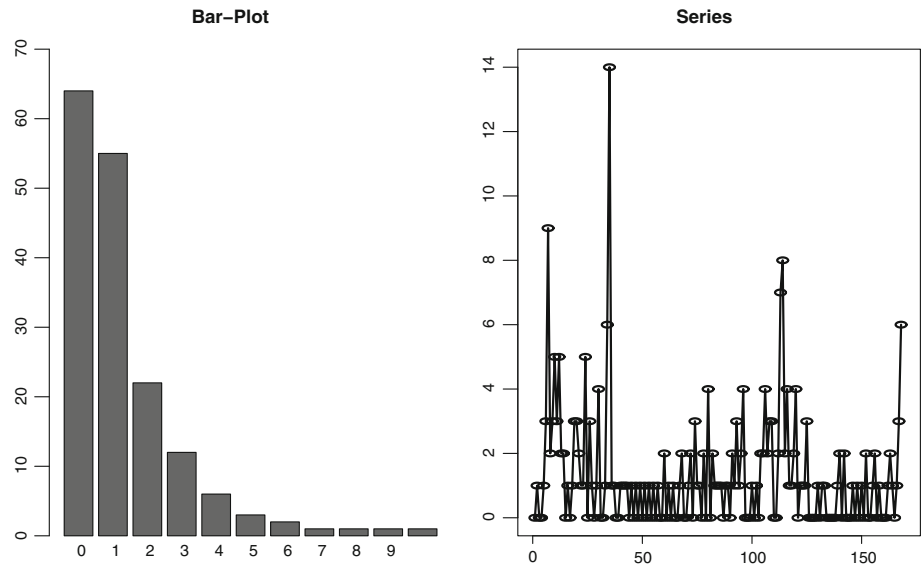
**Table 5** Parameter estimations and their standard errors and Loglik, AIC, AICc and BIC criteria for compared models that are fitted to daily new infected cases of COVID-19 in Barbados

Model	Estimated values (SE)	AIC	Loglik	AICc	BIC
PINAR(1)	$\hat{\alpha} = 0.1482(0.0305)$	1184.856	- 590.428	1184.897	1192.210
ZIPINAR(1)	$\hat{\lambda} = 1.1493(0.0712)$	992.236	- 493.118	992.319	1003.266
	$\hat{\alpha} = 0.1903(0.0314)$				
OIPINAR(1)	$\hat{\phi}_0 = 0.6033(0.0350)$	1147.017	- 570.509	1147.100	1158.047
	$\hat{\alpha} = 0.1858(0.0311)$				
	$\hat{\lambda} = 1.0747(0.0590)$				
ZOIPINAR(1)	$\hat{\phi}_1 = 0.0000(0.0433)$	949.333	- 470.666	949.471	964.039
	$\hat{\alpha} = 0.1669(0.0370)$				
	$\hat{\lambda} = 3.9909(0.3100)$				
ZOIPLINAR(1)	$\hat{\phi}_0 = 0.5890(0.0350)$	908.542	- 450.271	908.682	923.249
	$\hat{\phi}_1 = 0.1723(0.0303)$				
	$\hat{\alpha} = 0.1391(0.0393)$				
	$\hat{\theta} = 0.6411(0.0816)$				
INARG(1)	$\hat{\phi}_0 = 0.4793(0.0507)$	933.106	- 464.553	933.148	940.460
	$\hat{\phi}_1 = 0.0970(0.0367)$				
INARZIG(1)	$\hat{\alpha} = 0.0763(0.0398)$	930.972	- 462.486	931.553	942.002
	$\hat{\theta} = 1.2472(0.1105)$				
INAROIG(1)	$\hat{\alpha} = 0.1445(0.0371)$	908.344	- 451.172	908.428	<b>919.375</b>
	$\hat{\theta} = 1.8385(0.2208)$				
	$\hat{\phi}_0 = 0.3720(0.0624)$				
INARZOIG(1)	$\hat{\alpha} = 0.1078(0.0399)$	<b>906.407</b>	<b>- 449.204</b>	<b>906.547</b>	<b>921.114</b>
	$\hat{\theta} = 1.1887(0.1035)$				
	$\hat{\phi}_1 = 0.000(0.0389)$				
INARZOIG(1)	$\hat{\alpha} = 0.1381(0.0393)$	<b>906.407</b>	<b>- 449.204</b>	<b>906.547</b>	<b>921.114</b>
	$\hat{\theta} = 2.1965(0.0587)$				
	$\hat{\phi}_0 = 0.4284(0.0383)$				
	$\hat{\phi}_1 = 0.0772(0.0334)$				

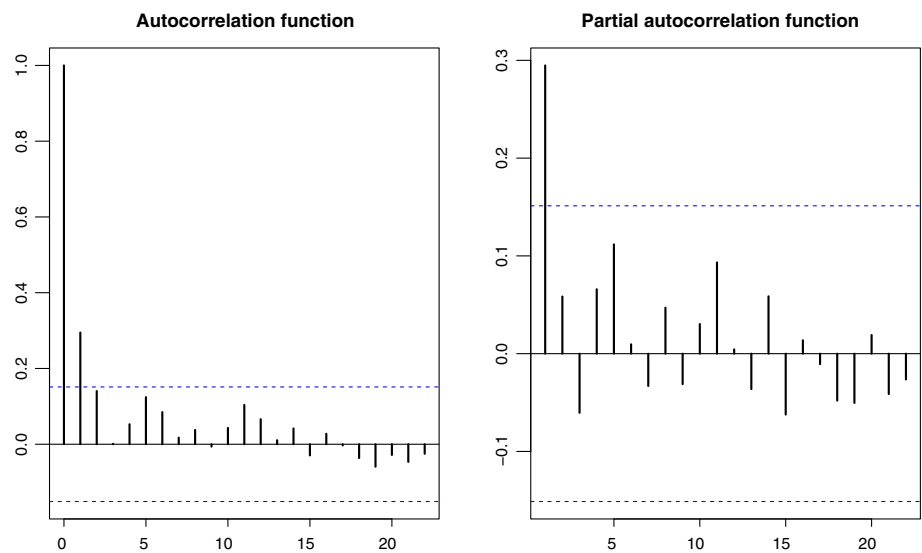
**Fig. 4** Daily new infected cases of COVID-19 in Barbados and their predicted values using INARZOIG(1)



**Fig. 5** Barplot and series plot of monthly cases of poliomyelitis data in the USA from 1970 to 1983



**Fig. 6** ACF and PACF of the monthly cases of poliomyelitis data in the USA from 1970 to 1983



**Table 6** Parameter estimation and Loglik, AIC, BIC and AICc criteria for compared models for Poliomyelitis data

Model	Estimated values	Loglik	AIC	AICc	BIC
OMGINAR(1)	$\hat{\alpha} = 0.068$				
	$\hat{p} = 0.882$	- 264.005	534.01	534.1563	543.3819
	$\hat{\theta} = 0.577$				
ZOIPLINAR(1)	$\hat{\alpha} = 0.0845(0.0493)$				
	$\hat{\theta} = 0.9116(0.1613)$	- 262.411	532.823	533.0685	545.318
	$\hat{\phi}_0 = 0.1887(0.0970)$ $\hat{\phi}_1 = 0.1881(0.0660)$				
INARZOIG(1)	$\hat{\alpha} = 0.0817(0.0496)$				
	$\hat{\theta} = 1.4812(0.3066)$	- 262.0769	532.1538	532.3992	544.6497
	$\hat{\phi}_0 = 0.1124(0.1151)$ $\hat{\phi}_1 = 0.1656(0.0691)$				

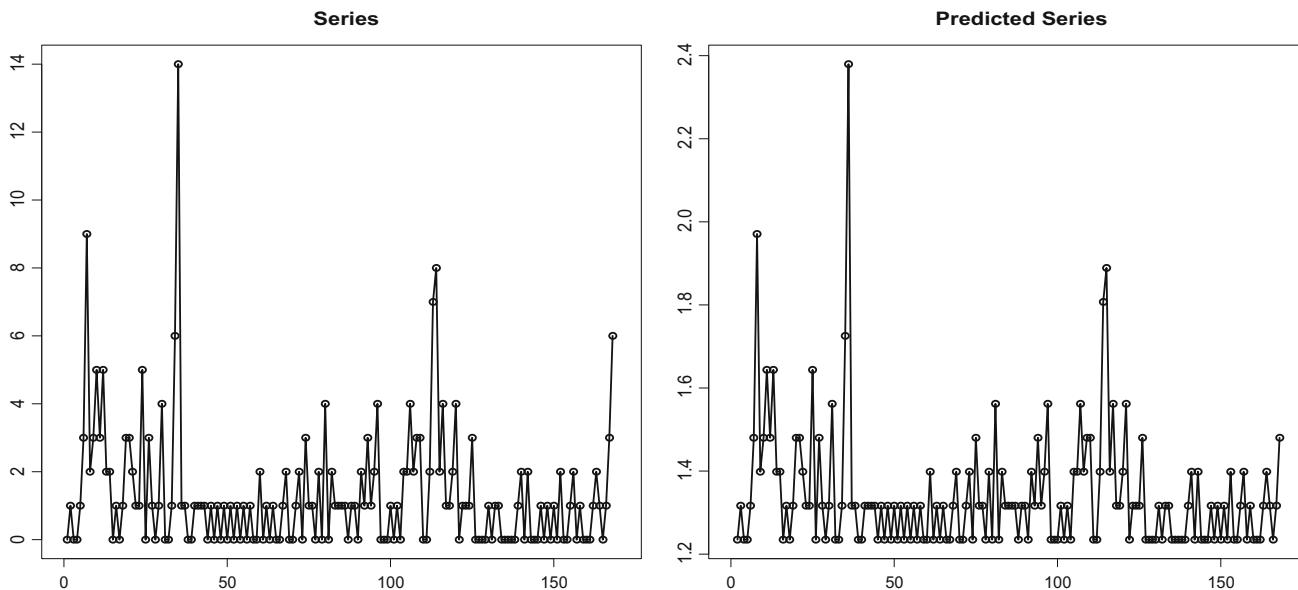


Fig. 7 Poliomyelitis data and their predicted values using INARZOIG(1)

### Appendix

#### Proof of Proposition 2

$$\begin{aligned}
 E(X_{t+h} | X_{t-1}) &= E(\alpha^{h+1} \circ X_{t-1} \\
 &+ \sum_{i=0}^h \alpha^i \circ \epsilon_{t+h-i} | X_{t-1}) \\
 &= \alpha^{h+1} X_{t-1} + E(\epsilon_t) \sum_{i=0}^h \alpha^i \\
 &= \alpha^{h+1} X_{t-1} + \frac{(1 - \alpha^{h+1}) \mu_\epsilon}{(1 - \alpha)} \\
 \text{Var}(X_{t+h} | X_{t-1}) &= \text{Var}(\alpha^{h+1} \circ X_{t-1} \\
 &+ \sum_{i=0}^h \alpha^i \circ \epsilon_{t+h-i} | X_{t-1}) \\
 &= \alpha^{h+1} (1 - \alpha^{h+1}) X_{t-1} + \sum_{i=0}^h \alpha^{2i} \text{Var}(\epsilon_t) \\
 &+ \sum_{i=0}^h \alpha^i (1 - \alpha^i) E(\epsilon_t) \\
 &= \alpha^{h+1} (1 - \alpha^{h+1}) X_{t-1} \\
 &+ \frac{(1 - \alpha^{2(h+1)}) \sigma_\epsilon^2}{1 - \alpha^2} \\
 &+ \frac{(1 - \alpha^h)(\alpha - \alpha^{h+2}) \mu_\epsilon}{1 - \alpha^2}
 \end{aligned}$$

**Proof of Propositions 3 and 4** These two Propositions are similar to Theorem 1 and 2 in Yang et al. (2019), which can be proved by verifying the regularity conditions of

Theorems 3.1 and 3.2 in Klimko and Nelson (1978). For instance, in the proof of Proposition 3, the partial derivatives  $\frac{\partial g(\alpha, F_{m-1})}{\partial \alpha_i}$  have finite fourth moments in Klimko and Nelson (1978),  $u_m(\alpha^0)$  in Klimko and Nelson (1978) is corresponded to  $q_{11}(\beta_{1,0})$  in Step 1. Hence, Proposition 3 can be regarded as a direct conclusion of Theorem 3.2.

**Proof of Proposition 5** The proof of Proposition 5 is similar to Proposition 4 in Liu and Zhu (2021), we omit the details here and refer the reader to Liu and Zhu (2021).

**Proof of Proposition 6** This is an application of the  $\delta$ -method. For completeness, we refer the reader to Theorem A on p. 122 of Serfling (1980) for a proof.

**Author Contributions** All authors made substantial contributions to the conception and design and finally approved the manuscript. Material preparation, data collection and analysis were performed by ZM, MS, ZS and NMK. All authors contributed to the drafting and preparation of the manuscript and the study protocol and provided the approval of the final version of the manuscript.

**Funding** There is no funding for this manuscript.

**Availability of data and material (data transparency)** The websites that contain the data are presented in the main manuscript.

**Code availability** We use R software and the codes can be sent to anyone who is interested in.

### Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Ethics approval (include appropriate approvals)** Not applicable

**Consent to participate (include appropriate statements)** Not applicable.

**Consent for publication (include appropriate statements)** Not applicable.

## References

- Al-Osh MA, Alzaid AA (1987) First-order integer-valued autoregressive (INAR(1)) process. *J Time Ser Anal* 8:261–275
- Alzaid AA, Al-Osh MA (1988) First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties. *Stat Neerl* 42:53–61
- Alzaid AA, Al-Osh MA (1993) Some autoregressive moving average processes with generalized Poisson marginal distributions. *Ann Inst Stat Math* 45:223–232
- Awale M, Kashikar AS, Ramanathan TV (2021) Forecasting overdispersed INAR(1) count time series with negative binomial marginal. *Commun Stat Simul Comput*. <https://doi.org/10.1080/03610918.2021.1908559>
- Bakouch HS, Ristić MM (2010) Zero-truncated Poisson integer-valued AR(1) model. *Metrika* 72:265–280
- Bakouch HS, Mohammadpour M, Shirozhan MA (2018) A zero inflated geometric INAR(1) process with random coefficient. *Appl Math* 63:79–105
- Barreto-Souza W (2015) Zero-modified geometric INAR(1) process for modelling count time series with deflation or inflation of zeros. *J Time Ser Anal* 36:839–852
- Bourguignon M, Vasconcellos KLP (2015) First order non-negative integer valued autoregressive processes with power series innovations. *Braz J Probab Stat* 29(1):71–93
- Bourguignon M, Borges P, Molinares FF (2018) A new geometric INAR(1) process based on counting series with deflation or inflation of zeros. *J Stat Comput Simul* 88(17):3338–3348
- Bourguignon M, Rodrigues J, Santos-Neto M (2019) Extended Poisson INAR(1) processes with equidispersion, underdispersion and overdispersion. *J Appl Stat* 46(1):101–118
- Dion JP, Gauthier G, Latour A (1995) Branching processes with immigration and integer-valued time series. *Serdica Math J* 21:123–136
- Du JG, Li Y (1991) The integer-valued autoregressive (INAR(p)) model. *J Time Ser Anal* 12:129–142
- Freeland RK, McCabe BPM (2004) Analysis of low count time series data by Poisson autoregression. *J Time Ser Anal* 25(5):701–722
- Forughi M, Shishebor Z, Zamani A (2021) Portmanteau tests for generalized integer-valued autoregressive time series models. *Stat Pap*. <https://doi.org/10.1007/s00362-021-01274-9>
- Huang J, Zhu F (2021) A new first-order integer-valued autoregressive model with Bell innovations. *Entropy* 23(6):713
- Jazi MA, Jones G, Lai CD (2012a) Integer valued AR(1) with geometric innovations. *J Iran Stat Soc* 11:173–190
- Jazi MA, Jones G, Lai CD (2012b) First-order integer valued AR processes with zero inflated Poisson innovations. *J Time Ser Anal* 33:954–963
- Karlsen H, Tjøstheim D (1988) Consistent estimates for the NEAR(2) and NLAR(2) time series models. *J R Stat Soc B Ser* 50:313–320
- Klimko LA, Nelson PI (1978) On conditional least squares estimation for stochastic processes. *Ann Stat* 6(3):629–642
- Livio T, Mamode Khan N, Bourguignon M, Bakouch HS (2018) An INAR(1) model with Poisson–Lindley innovations. *Econ Bull* 38:1505–1513
- Li C, Wang D, Zhang H (2015) First-order mixed integer-valued autoregressive processes with zero-inflated generalized power series innovations. *J Korean Stat Soc* 44:232–246
- Liu Z, Zhu F (2021) A new extension of thinning-based integer-valued autoregressive models for count data. *Entropy* 23(1):62
- Maiti R, Biswas A, Das S (2015) Time series of zero inflated counts and their coherent forecasting. *J Forecast* 34:694–707
- Maiti R, Biswas A, Chakraborty B (2018) Modelling of low count heavy tailed time series data consisting large number of zeros and ones. *Stat Methods Appl* 27(3):407–435
- Mamode Khan N, Cekim HO, Ozel G (2020a) The family of the bivariate integer-valued autoregressive process (BINAR(1)) with Poisson–Lindley (PL) innovations. *J Stat Comput Simul* 90(4):624–637
- Mamode Khan N, Soobhug AD, Mamode Khan MH (2020b) Studying the trend of the novel coronavirus series in Mauritius and its implications. *PLoS ONE* 15(7):e0235730
- McKenzie E (1985) Some simple models for discrete variate time series. *Water Resour Bull* 21:645–650
- McKenzie E (1986) Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Adv Appl Probab* 18:679–705
- McKenzie E (1988) Some ARMA models for dependent sequences of Poisson counts. *Adv Appl Probab* 20(4):822–835
- Mohammadi Z, Sajjadnia Z, Bakouch HS, Sharafi M (2021) Zero-and-one inflated Poisson–Lindley INAR(1) process for modelling count time series with extra zeros and ones. *J Stat Comput Simul*. <https://doi.org/10.1080/00949655.2021.2019255>
- Mohammadpour M, Bakouch HS, Shirozhan M (2018) Poisson–Lindley INAR(1) model with applications. *Braz J Probab Stat* 32(2):262–280
- Mood AM (1940) The distribution theory of runs. *Ann Math Stat* 11:367–392
- Nicholls DF, Quinn BG (1982) Random coefficient autoregressive models: an introduction. Springer, New York
- Pedeli X, Karlis D (2011) A bivariate INAR(1) process with application. *Stat Model* 11:325–349
- Pedeli X, Karlis D (2013) Some properties of multivariate INAR(1) processes. *Comput Stat Data Anal* 67:213–225
- Qi X, Li Q, Zhu F (2019) Modeling time series of count with excess zeros and ones based on INAR(1) model with zero-one inflated Poisson innovations. *J Comput Appl Math* 346:572–590
- Raftery AE (1995) Bayesian model selection in social research. *Sociol Methodol* 25:111–161
- Ristić MM, Bakouch HS, Nastić AS (2009) A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *J Stat Plan Inference* 139:2218–2226
- Ristić MM, Nastić AS, Bakouch HS (2012) Estimation in an integer-valued autoregressive process with negative binomial marginals (NBINAR(1)). *Commun Stat Theory Methods* 41:606–618
- Schweer S, Weiß CH (2014) Compound Poisson INAR(1) processes: stochastic properties and testing for overdispersion. *Comput Stat Data Anal* 77:267–284
- Serfling RJ (1980) Approximation theorems of mathematical statistics. Wiley, New York
- Sharafi M, Sajjadnia Z, Zamani A (2020) A first-order integer-valued autoregressive process with zero-modified Poisson–Lindley distributed innovations. *Commun Stat Simul Comput* 1–18. <https://doi.org/10.1080/03610918.2020.1864644>
- Steutel FW, van Harn K (1979) Discrete analogues of self-decomposability and stability. *Ann Probab* 7:893–899
- Sunecher Y, Mamode Khan N, Jowaheer V (2018) A case study of MCB and SBMH stock transaction using a novel BINMA(1) with non-stationary NB correlated innovations. *J Appl Stat* 46(2):304–323

- Tjøstheim D (1986) Estimation in nonlinear time series models. *Stoch Process Appl* 21:251–273
- Weiß CH (2008) Thinning operations for modeling time series of counts—a survey. *Adv Stat Anal* 92:319–343
- Weiß CH (2020) Stationary count time series models. *Wiley Interdiscip Rev Comput Stat* 13(1):e1502
- Yang K, Kang Y, Wang D, Li H, Diao Y (2019) Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued autoregressive processes. *Metrika* 82(7):863–889
- Zeger SL (1988) A regression model for time series of counts. *Biometrika* 75(4):621–629
- Zhang C, Tian G, Ng KW (2016) Properties of the zero-and-one inflated Poisson distribution and likelihood-based inference methods. *Stat Interface* 9:11–32