A ZERO-INFLATED GEOMETRIC INAR(1) PROCESS WITH RANDOM COEFFICIENT

Hassan S. Bakouch, Tanta,
Mehrnaz Mohammadpour, Masumeh Shirozhan, Babolsar

Received March 26, 2017. First published January 25, 2018.

Abstract. Many real-life count data are frequently characterized by overdispersion, excess zeros and autocorrelation. Zero-inflated count time series models can provide a powerful procedure to model this type of data. In this paper, we introduce a new stationary first-order integer-valued autoregressive process with random coefficient and zero-inflated geometric marginal distribution, named $\mathrm{ZIGINAR}_{\mathrm{RC}}(1)$ process, which contains some sub-models as special cases. Several properties of the process are established. Estimators of the model parameters are obtained and their performance is checked by a small Monte Carlo simulation. Also, the behavior of the inflation parameter of the model is justified. We investigate an application of the process using a real count climate data set with excessive zeros for the number of tornados deaths and illustrate the best performance of the proposed process as compared with a set of competitive INAR(1) models via some goodness-of-fit statistics. Consequently, forecasting for the data is discussed with estimation of the transition probability and expected run length at state zero. Moreover, for the considered data, a test of the random coefficient for the proposed process is investigated.

Keywords: randomized binomial thinning; geometric minima; estimation; likelihood ratio test; mixture distribution; realization with random size

MSC 2010: 62M10

1. Introduction

Integer-valued time series are encountered in several life situations, e.g., the number of days with storm, the number of road accidents, the number of foggy days and so on. The most common integer-valued time series models are constructed via the binomial thinning operator $\circ$, which was first introduced by [22] in the form

$$\alpha \circ X = S_X = \sum_{i=1}^{X} Y_i, \quad \alpha \in (0,1),$$

DOI: 10.21136/AM.2018.0082-17

where $S_0 = Y_0 = 0$, X is a non-negative integer-valued random variable and $\{Y_i\}$ is a sequence of independent identically distributed (i.i.d.) random variables with Bernoulli(α) distribution and is independent of X. The first-order non-negative integer value autoregressive (INAR(1)) process was first introduced in [15] and [2] based on the operator $\alpha\circ$. Let X_t be a non-negative integer-valued random variable observed at time t, then the INAR(1) process is defined as

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t, \tag{1.1}$$

where X_t can follow a certain marginal distribution, the innovation (or white noise) $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with some discrete distributions and also $\{\varepsilon_t\}$ is independent of the Bernoulli counting series $\{Y_i^{(t)}\}$ and X_{t-l}, for $l \geqslant 1$. The coefficient α may be explained as the proportion of observations counted at time $t-1$ that still remains at time t. Moreover, operator $\alpha\circ$ mimics the scalar multiplication used in Gaussian time series and reduces the random sum, S_X, to a single integer value (i.e. thinning), which leads to integer-valued time series models.

Modeling of INAR(1) time series based on (1.1) was first introduced using the Poisson marginal distribution and then using the geometric marginal distribution [4], negative binomial marginal distribution [4], generalized Poisson marginal distribution [5] and zero truncated Poisson distribution [6]. Other models of INAR time series are introduced via the negative binomial thinning operator (see [3], [19], [20]). An overdispersed INAR(1) model with innovations following a finite mixture of Poisson distribution of order k, $k \geqslant 1$, is introduced in [17]. Jazi et al. [12] worked on an overdispersed INAR(1) model with geometric innovations. Random coefficient INAR(1) processes, in which the autoregressive parameter itself is a random variable, are considered by a few authors such as [9], [27], [25]. Further, [11], [14] introduced INAR(1) models with zero-inflated Poisson and zero-inflated generalized power series innovations, respectively, and [7] introduced a zero-modified geometric INAR(1) model based on negative binomial thinning operator for modeling count series with inflation or deflation of zeros. Recently, [21] introduced a general family of INAR(1) models with compound Poisson innovations. [8] constructed an INAR(1) model with power series innovations and [13] presented a new model based on the Pegram and thinning operators.

In many real count data, we may observe that the number of zeros has a large proportion, which is called zero inflation. An example of zero inflation is the hydrologic data in arid and semiarid regions, like monthly precipitation in dry seasons, annual peak flow discharges, etc. Such data can be analyzed by zero-inflated statistical models, e.g. Poisson and negative binomial distributions. Zero-inflation is common in many fields, e.g. manufacturing defects, medical consultations, hydrology, road safety, ecology and econometrics. Ignoring zero inflation in the analysis can have two consequences: firstly, the estimated parameters and standard errors may be biased, and secondly, the excessive number of zeros can cause overdispersion ([28]).

Zero-inflated models are extensively investigated in regression contexts (see, e.g., [18]) and have received limited attention in time series contexts, although count data with many zeros may exhibit strong autocorrelation and can often be modeled by the INAR(1) model. Moreover, most of the introduced integer-valued time series models can analyze only the non-negative integer values without excess zeros. Therefore, there is a need to introduce other integer valued time series models besides the current ones. In this paper, the zero-inflated geometric marginal distribution is used directly to model the observed INAR(1) time series $\{X_t\}$ with the operator $\circ$ defined by (1.1), and the distribution of the innovations of the INAR(1) is determined. Therefore, this approach is different from the one used in [11], [14], as the latter specify the INAR(1) model by assuming a certain distribution for the innovations. Further, our approach defines an INAR(1) model with a random coefficient, which is different from the model of [7] because of the nature of the autoregressive parameter of the INAR(1) model, α. To show this let X_t be the number of patients in inpatient wards in the t-th month, hence X_t obeys an INAR process and represents the sum of the number of surviving patients from the previous month (denoted by $\alpha \circ X_{t-1}$) and the newly admitted patients in the current month (ε_t). In this example, the survival rate α may be affected by several environmental factors such as the state of health of patients, the quality of health care, nature of disease, etc., hence it could vary randomly over time and be denoted by α_t. Hence, X_t could potentially be modeled as the sum of $\alpha_t \circ X_{t-1}$ and ε_t. This situation and the issue of excess zeros were the motivations to our model.

Definition 1.1. A random variable X is said to be a zero-inflated geometric (ZIG) distribution with parameters μ and p, denoted as $\mathrm{ZIG}(p, \mu/(1+\mu))$, if its probability mass function (pmf) is defined as

$$P(X = x) = \begin{cases} p, & x = 0, \\ (1-p)\dfrac{\mu^x}{(1+\mu)^{x+1}}, & x = 0, 1, \ldots, \end{cases}$$

where $0 \leqslant p < 1$, $\mu > 0$.

Remark 1.1. (i) The probability generating function (pgf) of X is $\varphi_X(s) = (1 + \mu p(1-s))/(1 + \mu(1-s))$.

(ii) The mean and the variance of X are $\mu_X = (1-p)\mu$, $\sigma_X^2 = (1-p)\mu((1+p)\mu+1)$, respectively, and hence the $\mathrm{ZIG}(p, \mu/(1+\mu))$ is overdispersed.

(iii) The coefficient of variation of X is

$$C.V_{\text{ZIG}} = \frac{\sqrt{(1-p)\mu(1+(1+p)\mu)}}{(1-p)\mu}.$$

Since the coefficient of variation of the negative binomial $(r, \mu/(1+\mu))$ is $C.V_{\text{NB}} = \sqrt{r\mu(1+\mu)}/r\mu$, we conclude that $C.V_{\text{ZIG}} > C.V_{\text{NB}}$.

By virtue of (ii), overdispersion represents another motivation of the proposed model besides its use in modeling and analysis of the non-negative integer-valued time series with excess zeros.

The rest of the paper is organized as follows. In Section 2, we construct the zero-inflated geometric INAR(1) process and obtain the pmf of innovations of the process. Various properties of the process are established in Section 3, including the autocorrelation function, joint probability generating function, k-step ahead conditional expectation and variance, $k = 1, 2, \ldots$, the one-step transition probabilities, extreme order statistics and survival function of the run length with its expectation. Also, estimators of the model parameters are obtained by the conditional least squares and maximum likelihood estimation methods. The performance of the estimators of both methods is checked by a small Monte Carlo simulation, as well as the behavior of the inflation parameter of the model. In Section 4, we give an application of the process to real count climate data with excessive zeros for the number of tornado deaths and illustrate the best performance of the proposed process as compared with some competitive INAR(1) models via the Akaike information criterion, Bayesian information criterion, Hannan-Quinn information criterion and consistent Akaike information criterion. Based on the best model selection, forecasting for the tornado data is discussed with estimation of the transition probability and expected run length at state zero. Moreover, for the tornado deaths data, a test of random coefficient for the proposed process is investigated. The concluding remarks are given in Section 5.

2. The ZIGINAR$_{\text{RC}}$(1) model

In this section, we propose a randomized binomial thinning operator and then introduce a zero-inflated geometric INAR(1) process with random coefficient that admits non-negative integer values with excess zeros based on such an operator. The pmf of the innovation term of the process is obtained.

2.1. Properties of the randomized binomial thinning operator.

In this section, we define the randomized binomial thinning operator and justify some of its properties.

Definition 2.1. Given a non-negative integer-valued random variable X and a binary random variable α_t independent of X, such that $P(\alpha_t = 0) = \beta = 1 - P(\alpha_t = \alpha)$ for any given real numbers $\alpha, \beta \in (0, 1)$, the (standard) binomial thinning operation $\alpha \circ X$ is hereby extended to a randomized binomial thinning operation, defined and denoted by the random variable

$$\alpha_t \circ X = \begin{cases} \alpha \circ X, & \text{w.p. } 1 - \beta, \\ 0, & \text{w.p. } \beta. \end{cases}$$

Based on this definition, we note that the binomial thinning operator with random coefficient is specified by binary random variable α_t, that is, '$\alpha_t \circ$' is either equal to the 'standard' binomial thinning operator '$\alpha \circ$' (w.p. $1-\beta$) or equal to 0 (w.p. β). Moreover, the probability generating function (pgf) of $\alpha_t \circ X$ is expressed as follows in terms of the parameters α, β and the pgf of X:

$$\begin{aligned} \varphi_{\alpha_t \circ X}(s) = E(s^{\alpha_t \circ X}) &= \beta + (1 - \beta) E(s^{\alpha \circ X}) \\ &= \beta + (1 - \beta) E(E(s^{\alpha \circ X} \mid X)) \\ &= \beta + (1 - \beta) \varphi_X(1 - \alpha + \alpha s). \end{aligned}$$

Now, we give some properties of the operator $\alpha_t \circ$ as follows.

Property 2.1. *If* $X \sim \mathrm{ZIG}(p, \mu/(1 + \mu))$, *then*

$$\alpha_t \circ X \sim \mathrm{ZIG}(\beta + p(1 - \beta), \alpha\mu/(1 + \alpha\mu)).$$

See Appendix A for details.

Property 2.2. *If* $X \sim \mathrm{ZIG}(p, \mu/(1 + \mu))$, *then*

$$p(\alpha_t \circ X = 0) = \frac{1 + \alpha\mu(p + \beta(1 - p))}{1 + \alpha\mu} > \frac{1 + \alpha\mu p}{1 + \alpha\mu} = p(\alpha \circ X = 0).$$

See Appendix B for details. Property 2.2 justifies the role of β as inflation parameter for the probability of 0.

Property 2.3. *Given non-negative integer-valued random variables* X *and* Y, *having an arbitrary discrete distribution, and a binary random variable* α_t *defined by Definition* 2.1, *such that* X, Y, α_t *are mutually independent, then the following equality of distributions holds:*

$$\alpha_t \circ (X + Y) \stackrel{\mathrm{d}}{=} \alpha_t \circ X + \alpha_t \circ Y.$$

See Appendix C for details.

Property 2.4. *Given a non-negative integer-valued random variable X, having an arbitrary discrete distribution, and binary random variables α_{t_1} and α_{t_2} with $P(\alpha_{t_i} = 0) = \beta_i = 1 - P(\alpha_{t_i} = \alpha_i)$, $\alpha_i, \beta_i \in (0,1)$, where $i = 1, 2$, such that X, α_{t_1}, α_{t_2} are mutually independent, then the following equality of distributions holds:*

$$\alpha_{t_1} \circ (\alpha_{t_2} \circ X) \stackrel{\mathrm{d}}{=} (\alpha_{t_1}\alpha_{t_2}) \circ X.$$

See Appendix D for details.

2.2. Formulation of the model. In this subsection, we introduce a new strictly stationary integer-valued autoregressive process $\{X_t\}$ of the first order with $\mathrm{ZIG}(p, \mu/(1+\mu))$ marginals based on the randomized binomial thinning.

The model is constructed as follows. Consider an i.i.d. sequence of binary thinning coefficients $\{\alpha_t\}$ and another i.i.d. sequence of non-negative integer-valued innovations $\{\varepsilon_t\}$. Moreover, the sequences $\{\alpha_t\}$ and $\{\varepsilon_t\}$ are assumed to be mutually independent, with each thinning coefficient, α_t, following the binary distribution $P(\alpha_t = 0) = \beta = 1 - P(\alpha_t = \alpha)$, with $\alpha, \beta \in (0,1)$ constrained by the condition $p/(\beta + p(1-\beta)) < \alpha < 1$. Note that the condition on α is required to get a proper probability mass function of $\{\varepsilon_t\}$ as it will be shown later by Proposition 2.1. Hence, the process $\{X_t\}$ is defined as

$$X_t = \alpha_t \circ X_{t-1} + \varepsilon_t; \tag{2.1}$$

also note the independence of each ε_t of the past of the solution $\{X_s;\ s < t\}$ and the independence of ε_t of the corresponding pair (α_t, X_{t-1}). We will refer to this model as a zero-inflated geometric INAR(1) with random coefficient ($\mathrm{ZIGINAR_{RC}}(1)$). Note that the $\mathrm{ZIGINAR_{RC}}(1)$ process and ZTPINAR(1) process defined by [6] are similar in their general form but the ZTPINAR(1) takes positive integer values while the $\mathrm{ZIGINAR_{RC}}(1)$ takes non-negative integer values with excess zeros.

Note that the definition of $\alpha_t \circ X_{t-1}$ implies inflation of probabilities at 0 for the model (2.1), since exploiting the fact that $(\alpha \circ X_{t-1} \mid X_{t-1}) \sim \mathrm{Bin}(X_{t-1}, \alpha)$, we have $P(\alpha_t \circ X_{t-1} = 0) = (1 + \alpha\mu(p + \beta(1-p)))/(1+\alpha\mu)$, as pointed out in Property 2.2.

Based on the properties of the randomized binomial thinning operator and iterating (2.1), we obtain

$$X_t = \left(\prod_{i=0}^{k-1} \alpha_{t-i}\right) \circ X_{t-k} + \sum_{i=1}^{k-1}\left(\prod_{l=0}^{i-1} \alpha_{t-l}\right) \circ \varepsilon_{t-i} + \varepsilon_t.$$

Therefore,

$$E\Bigg[X_t - \sum_{i=1}^{k-1}\Bigg(\prod_{l=0}^{i-1}\alpha_{t-l}\Bigg)\circ\varepsilon_{t-i} - \varepsilon_t\Bigg]^2 = E\Bigg[\Bigg(\prod_{i=0}^{k-1}\alpha_{t-i}\Bigg)\circ X_{t-k}\Bigg]^2$$
$$= (1-\beta)^k E[(\alpha^k \circ X_{t-k})^2];$$

conditioning on X_{t-k} and using the fact that $(\alpha^k \circ X_{t-k} \mid X_{t-k}) \sim \mathrm{Bin}(X_{t-k}, \alpha^k)$, we get

$$E\Bigg[X_t - \sum_{i=1}^{k-1}\Bigg(\prod_{l=0}^{i-1}\alpha_{t-l}\Bigg)\circ\varepsilon_{t-i} - \varepsilon_t\Bigg]^2 = (1-\beta)^k E[\alpha^k(1-\alpha^k)X_{t-k} + \alpha^{2k}X_{t-k}^2]$$
$$= (1-\beta)^k\alpha^k(1-\alpha^k)\mu_X + (1-\beta)^k\alpha^{2k}(\mu_X^2+\sigma_X^2) \to 0 \quad \text{as } k\to\infty,$$

where μ_X and σ_X^2 are, respectively, the mean and variance of the stationary solution $\{X_t\}$. Hence,

$$X_t \stackrel{\mathrm{d}}{=} \sum_{i=1}^{\infty}\Bigg(\prod_{l=0}^{i-1}\alpha_{t-l}\Bigg)\circ\varepsilon_{t-i} + \varepsilon_t. \tag{2.2}$$

As the expectation of the infinite sum in (2.2) exists, it converges with probability one.

Using Theorem 2.7 of [16] and the expression (2.2) for ZIGINAR$_{\mathrm{RC}}$(1) process, we have the following theorem.

Theorem 2.1. *The* ZIGINAR$_{\mathrm{RC}}$(1) *process* (2.1) *has a unique, strictly stationary solution given by* (2.2).

Based on the solution of the model (2.1) and the specification of the processes $\{\alpha_t\}$ and $\{\varepsilon_t\}$ mentioned above, another representation of the model (2.1) in terms of the standard (instead of randomized) thinning operation can be given as

$$X_t = \begin{cases} \varepsilon_t, & \text{w.p. } \beta, \\ \alpha\circ X_{t-1}+\varepsilon_t, & \text{w.p. } 1-\beta, \end{cases} \qquad t\geqslant 1,\ 0<\beta<1. \tag{2.3}$$

From equation (2.1) and the fact that ε_t is stochastically independent of the pair (α_t, X_{t-1}) for each t, the pgf of the innovation $\{\varepsilon_t\}$ is obtained as

$$\begin{aligned} \varphi_{\varepsilon_t}(s) &= \varphi_{X_t}(s)/\varphi_{\alpha_t\circ X_{t-1}}(s) \\ &= \frac{(1+\mu p-\mu ps)(1+\alpha\mu-\alpha\mu s)}{(1+\mu-\mu s)(1+\alpha\mu(\beta+p(1-\beta))-\alpha\mu(\beta+p(1-\beta))s)}, \end{aligned} \tag{2.4}$$

where $\varphi_X(\cdot)$ is the pgf of $\{X_t\}$. The degree of the numerator and the denominator of (2.4) with respect to s is two. Hence, we first divide the fraction as

$$\varphi_\varepsilon(s) = \frac{p}{\beta+p(1-\beta)} + \frac{s[\frac{p}{\beta+p(1-\beta)} - p(1-\alpha) - \alpha]\mu + (1+\mu p + \alpha\mu - \alpha\mu p - \frac{p(1+\mu)}{\beta+p(1-\beta)})}{(1+\mu-\mu s)(1+\alpha\mu(\beta+p(1-\beta)) - \alpha\mu(\beta+p(1-\beta))s)}.$$

Now, the degree of the numerator of the second part is one and we can decompose it. So, $\varphi_\varepsilon(s)$ can be put in the form

$$\varphi_\varepsilon(s) = A + B\frac{1}{1+\mu-\mu s} + C\frac{1}{1+\alpha\mu(\beta+p(1-\beta)) - \alpha\mu(\beta+p(1-\beta))s},$$

where

$$A = \frac{p}{\beta+p(1-\beta)}, \quad B = \frac{(1-p)(1-\alpha)}{1-\alpha[\beta+p(1-\beta)]}$$

and

$$C = \frac{(1-p)(1-\beta)[\alpha(\beta+p(1-\beta)) - p]}{(1-\alpha[\beta+p(1-\beta)])(\beta+p(1-\beta))}.$$

It can be shown that $A+B+C=1$, $\lim_{s\to\infty}\varphi_\varepsilon(s) = A$, $\varphi_\varepsilon(1)=1$, and $\varphi_\varepsilon(0)<1$. Further, the next proposition proves that the coefficients of the series expansion of $\varphi_\varepsilon(s)$, say $g(x)$, are non-negative and having the sum one, which implies that $\varphi_\varepsilon(s)$ is a proper pgf.

Proposition 2.1. *The mixture*

$$g(x) = \frac{p}{\beta+p(1-\beta)}I_{(x)} + \frac{(1-p)(1-\alpha)}{1-\alpha[\beta+p(1-\beta)]}\frac{1}{1+\mu}\Big(\frac{\mu}{1+\mu}\Big)^x + \frac{(1-p)(1-\beta)[\alpha(\beta+p(1-\beta)) - p]}{(1-\alpha[\beta+p(1-\beta)])(\beta+p(1-\beta))} \times \frac{1}{1+\alpha\mu(\beta+p(1-\beta))}\Big[\frac{\alpha\mu(\beta+p(1-\beta))}{1+\alpha\mu(\beta+p(1-\beta))}\Big]^x$$

is a pmf *for* $p/(\beta+p(1-\beta)) < \alpha < 1$, *where*

$$I_{(x)} = \begin{cases} 1, & x=0, \\ 0, & x\neq 0. \end{cases}$$

Proof. From the definition of A, B and C, we find that $A > 0$, $B > 0$ and $C > 0$ for $\alpha > p/(\beta + p(1-\beta))$, then $g(x) \geqslant 0$. Since $A + B + C = 1$, we have

$$\sum_{x=1}^{\infty} g(x) = 1.$$

Therefore, $g(x)$ is a pmf. $\square$

In Proposition 2.1, the constraint is imposed by the desired marginal law $\mathrm{ZIG}(p, \mu/(1+\mu))$ of the stationary solution $\{X_t\}$ to guarantee the existence of a proper probability mass function of the marginal law of the i.i.d. innovations $\{\varepsilon_t\}$.

From Proposition 2.1, the innovation process $\{\varepsilon_t\}$ has a mixture distribution of a degenerate distribution at 0, $\mathrm{Geometric}(\mu/(1+\mu))$ and $\mathrm{Geometric}((\alpha\mu(\beta + p(1-\beta)))/(1+\alpha\mu(\beta+p(1-\beta))))$, with mixing portions A, B, and C, respectively. The mean and variance of $\{\varepsilon_t\}$ are

$$\mu_\varepsilon = (1-(1-\beta)\alpha)(1-p)\mu,$$

and

$$\sigma_\varepsilon^2 = (1-p)\mu[(1-(1-\beta)\alpha^2)((1+p)\mu+1) - (1-\beta)\alpha(1-\alpha)],$$

respectively.

Remark 2.1. If $p = 0$, then the $\mathrm{ZIGINAR_{RC}}(1)$ process defined by (2.3) is reduced to an INAR(1) model with $\mathrm{Geometric}(\mu/(1+\mu))$ marginal, which is a submodel of the process and will be denoted by $\mathrm{GINAR_{RC}}(1)$.

3. Probabilistic properties of the process and estimation

In this section we investigate some statistical and conditional properties of the $\mathrm{ZIGINAR_{RC}}(1)$ process, including the autocorrelations, spectral density, one-step transition probabilities, multi-step conditional mean and variance, extreme order statistics, expected run length T_i, survival function of T_i and the pgf of T_i. Also, estimation of the process parameters is investigated.

3.1. Statistical properties. The autocovariance function of the $\mathrm{ZIGINAR_{RC}}(1)$ process, $\{X_t\}$, is obtained as

$$\mathrm{Cov}(X_t, X_{t-k}) = (1-p)\mu((1+p)\mu+1)(1-\beta)^k\alpha^k,$$

see Appendix E for details.

The autocorrelation function is $\varrho_k = \mathrm{Corr}(X_t, X_{t-k}) = (1-\beta)^k\alpha^k$. Consequently, the spectral density function

$$f_{xx}(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \mathrm{Cov}(X_t, X_{t-k})\mathrm{e}^{-\mathrm{i}\omega k}$$

of the ZIGINAR$_{\mathrm{RC}}$(1) process is

$$f_{xx}(\omega) = \frac{(1-p)\mu((1+p)\mu+1)(1-(\alpha(1-\beta))^2)}{2\pi(1+(\alpha(1-\beta))^2 - 2\alpha(1-\beta)\cos\omega)}, \quad \omega \in (-\pi, \pi].$$

Moreover, the joint pgf of the process is given by

$$\varphi_{X_t,X_{t-1}}(s_1,s_2) = \varphi_{\varepsilon_t}(s_1)\Big[\beta\frac{1+\mu p(1-s_2)}{1+\mu(1-s_2)} + (1-\beta)\frac{1+\mu p(1-s_2(1-\alpha+\alpha s_1))}{1+\mu(1-s_2(1-\alpha+\alpha s_1))}\Big],$$

where $\varphi_{\varepsilon_t}(s)$ is given by (2.4), and the proof of the above expression is outlined in Appendix F. Because of this, the process is not time reversible, and hence, the joint distribution of (X_t, X_{t-1}) is not equal to joint distribution of (X_{t-1}, X_t).

3.2. Conditional properties. The ZIGINAR$_{\mathrm{RC}}$(1) process is a stationary discrete time Markov chain with the one-step transition probabilities from state i to state j:

$$\begin{aligned}(3.1) \qquad P_{ij} = p(X_t = j \mid X_{t-1} = i) &= \beta p(\varepsilon_t = j) \\ &+ (1-\beta)\sum_{k=0}^{\min(i,j-1)}\binom{i}{k}\alpha^k(1-\alpha)^{i-k}p(\varepsilon_t = j-k),\end{aligned}$$

where $p(\varepsilon_t = j)$ is the pmf of $\{\varepsilon_t\}$ defined by Proposition 2.1 and i, $j = 0, 1, \ldots$ Due to this, the transition probabilities satisfy $P_{ij} > 0$, and hence the process $\{X_t\}$ is an irreducible and aperiodic Markov chain. Thus, it is either positive recurrent or $\lim_{n\to\infty} P_{ij}^n = 0$.

Also, for the ZIGINAR$_{\mathrm{RC}}$(1) process we get

$$\begin{aligned}P_{00} = &\frac{p}{p+\beta-p\beta} + \frac{(1-p)(1-\alpha)}{1-\alpha(p+\beta-p\beta)}\frac{1}{1+\mu} \\ &+ \frac{(1-p)(1-\beta)(\alpha\beta(1-p)-p(1-\alpha))}{(1-\alpha(p+\beta-p\beta))(p+\beta-p\beta)}\frac{1}{1+\alpha\mu(\beta+p(1-\beta))},\end{aligned}$$

see Appendix G for the details.

Conditional expectation is one of the most common techniques for forecasting time series data, so it is obtained in the next proposition together with the conditional variance.

Proposition 3.1. *For the* $\mathrm{ZIGINAR}_{\mathrm{RC}}(1)$ *process, the* k *step-ahead conditional expectation and variance are*

$$E(X_{t+k} \mid X_t = x) = (1-\beta)^k \alpha^k x + \frac{1-(1-\beta)^k\alpha^k}{1-(1-\beta)\alpha}\mu_\varepsilon, \tag{3.2}$$

and

$$\operatorname{Var}(X_{t+k} \mid X_t) = \sum_{j=0}^{k-1}(1-\beta)^j[\alpha^j(1-\alpha^j)\mu_\varepsilon + \alpha^{2j}\sigma_\varepsilon^2] + (1-\beta)^k\alpha^k(1-\alpha^k)\sigma_X^2,$$

respectively. Here μ_ε *and* σ_ε^2 *are the unconditional mean and the unconditional variance of the innovation* $\{\varepsilon_t\}$*, respectively.*

P r o o f. Using the properties of the binomial thinning operator, the proof of the first part follows easily by induction. For the second part, we find that

$$\begin{aligned}\operatorname{Var}(X_{t+k} \mid X_t) &= \operatorname{Var}(\varepsilon_{t+k}) + (1-\beta)\operatorname{Var}(\alpha \circ X_{t+k-1} \mid X_t) = \dots \\ &= \sum_{j=0}^{k-1}(1-\beta)^j \operatorname{Var}(\alpha^j \circ \varepsilon_{t+k-j}) + (1-\beta)^k \operatorname{Var}(\alpha^k \circ X_t \mid X_t) \\ &= \sum_{j=0}^{k-1}(1-\beta)^j[\alpha^j(1-\alpha^j)\mu_\varepsilon + \alpha^{2j}\sigma_\varepsilon^2] + (1-\beta)^k\alpha^k(1-\alpha^k)\sigma_X^2,\end{aligned}$$

hence the required result is obtained. □

Corollary 3.1. *From equation* (3.2) *we get*

$$\lim_{k\to\infty} E(X_{t+k} \mid X_t = x) = \mu_X,$$

which is the unconditional mean of the process as expected, since the process is ergodic.

Corollary 3.2. *Using equation* (3.2)*, we find that*

$$\lim_{k\to\infty} \operatorname{Var}(X_{t+k} \mid X_t) = (1-p)\mu(1+\mu(1+p)) = \sigma_X^2,$$

which is the unconditional variance of the process.

3.3. Distributional properties of length of run of zeros. A run is defined as a succession of similar events preceded and succeeded by different events and the number of elements in a run is referred to as its length. [10] obtained the distributions of the run length of state i for a stationary discrete process. Using the run length, we can evaluate the expected time it would spend in state i at a stress after entering the process at state i from another state j. For a fixed state $i \in \{0, 1, \ldots\}$, let $T_i = \inf\{t \geqslant 1, X_t \neq i\} - 1$ be the run length of state i, starting at time epoch 1.

Proposition 3.2. *For the* $\mathrm{ZIGINAR_{RC}}(1)$ *process, we have:*

(i) *The* pmf *of* T_i *at* $t = 0$ *is*

$$P(T_i = 0) = 1 - \Big(pI_{(i)} + \frac{(1-p)\mu^i}{(1+\mu)^{i+1}}\Big). \tag{3.3}$$

(ii) *At* $t \in \{1, 2, \ldots\}$, *the* pmf *of* T_i *is*

$$P(T_i = t) = \Big(pI_{(i)} + \frac{(1-p)\mu^i}{(1+\mu)^{i+1}}\Big)(1 - P_{ii})P_{ii}^{t-1}. \tag{3.4}$$

(iii) *The survival function of* T_i *is*

$$P(T_i \geqslant t) = \begin{cases} 1, & t = 0, \\ \Big(pI_{(i)} + \dfrac{(1-p)\mu^i}{(1+\mu)^{i+1}}\Big)P_{ii}^{t-1}, & t = 1, 2, \ldots \end{cases} \tag{3.5}$$

(iv) *The expected run length of the state* i *is*

$$E(T_i) = \frac{pI_{(i)} + (1-p)\mu^i/(1+\mu)^{i+1}}{1 - P_{ii}}. \tag{3.6}$$

(v) *The* pgf *of* T_i *is*

$$\varphi_{T_i}(s) = 1 - \Big(pI_{(i)} + \frac{(1-p)\mu^i}{(1+\mu)^{i+1}}\Big) + \frac{s(1-P_{ii})(pI_{(i)} + (1-p)\mu^i/(1+\mu)^{i+1})}{1 - sP_{ii}}, \tag{3.7}$$

where P_{ii} *is obtained by* (3.1).

Proof. The proofs of (i) and (ii) are obtained directly by noting that $P(T_i = 0) = P\{X_1 \neq i\}$, $P(T_i = t) = P\{X_1 = i,\ X_2 = i, \ldots, X_t = i,\ X_{t+1} \neq i\}$ and using Markovian property of the process. (iii) is obtained by the identity

$$P(T_i \geqslant t) = P\Big\{\bigcup_{k=0}^{\infty}(X_1 = i,\ X_2 = i, \ldots, X_{t+k} = i,\ X_{t+k+1} \neq i)\Big\}.$$

Statements (iv) and (v) are obtained by definition of the expectation and pgf. □

R e m a r k 3.1. Equations (3.4), (3.5) yield $\sum_{t=0}^{\infty} P(T_i = t) = 1$.

3.4. Estimation and simulation comparison. Various features of the model depend on its parameters, so estimation of the model parameters is essential. In this section, we describe the estimation of the unknown parameters of the ZIGINAR$_{\mathrm{RC}}(1)$ process, based on a realization $X_1, \ldots, X_n$ of this process, and conduct a small Monte Carlo simulations in terms of their mean and standard deviations to gain an idea on the estimation methods and also the behavior of the inflation parameter of the model.

3.4.1. Estimation. Let $\gamma = \alpha(1-\beta)$ and $\tau = (1-p)\mu$. Then, using by virtue of (3.2), the conditional least squares (CLS) estimators of the parameters γ and τ are obtained by minimizing the function

$$S_n(\gamma,\tau) = \sum_{t=2}^{n} (X_t - \gamma X_{t-1} - (1-\gamma)\tau)^2, \tag{3.8}$$

and are given by

$$\hat{\gamma}_{\mathrm{cls}} = \frac{(n-1)\sum\limits_{t=2}^{n}(X_{t-1}X_t) - \sum\limits_{t=2}^{n} X_t \sum\limits_{t=2}^{n} X_{t-1}}{(n-1)\sum\limits_{t=2}^{n} X_{t-1}^2 - \Big(\sum\limits_{t=2}^{n} X_{t-1}\Big)^2},$$

$$\hat{\tau}_{\mathrm{cls}} = \frac{1}{(n-1)(1-\hat{\gamma}_{\mathrm{cls}})}\Big(\sum_{t=2}^{n} X_t - \hat{\gamma}_{\mathrm{cls}} \sum_{t=2}^{n} X_{t-1}\Big).$$

The estimators $\hat{p}$, $\hat{\alpha}$, $\hat{\beta}$, $\hat{\mu}$ are obtained by numerical solution of (3.8).

Now, we get the asymptotic properties of the estimators $\hat{\gamma}_{\mathrm{cls}}$ and $\hat{\tau}_{\mathrm{cls}}$. All the conditions of Theorem 3.1 in [23] are satisfied. Thus, it follows that the conditional least squares estimators $\hat{\gamma}_{\mathrm{cls}}$ and $\hat{\tau}_{\mathrm{cls}}$ are strongly consistent estimators. Also, making use of Theorem 3.2 in [23] implies

$$\sqrt{n}\begin{pmatrix} \hat{\gamma}_{\mathrm{cls}} - \gamma \\ \hat{\tau}_{\mathrm{cls}} - \tau \end{pmatrix} \xrightarrow{d} N(0, \mathbf{J}),$$

where $N(0,\mathbf{J})$ denotes the bivariate normal distribution with mean zero vector and covariance matrix

$$\mathbf{J} = \begin{bmatrix} \frac{\mathrm{var}(\varepsilon)\,\mathrm{var}(X_t) + \gamma(1-\alpha)[E(X_t^3) + E^3(X_t) - 2E(X_t^2)E(X_t)]}{(\mathrm{var}(X_t))^2} & \frac{\gamma(1-\alpha)}{1-\gamma} \\ \frac{\gamma(1-\alpha)}{1-\gamma} & \frac{\tau((1+p)\mu+1)(1-(1-\beta)\alpha^2)}{(1-\gamma)^2} \end{bmatrix},$$

with

$$E(X_t^3) = (1-p)\mu[1 + 6\mu^2 + 6\mu].$$

The maximum likelihood estimators (MLEs) of the unknown parameters p, α, β and μ are obtained by maximization of the log-likelihood function

$$
\begin{aligned}
(3.9)\qquad l(\boldsymbol{\theta}) &= \log L(X_1,\ldots,X_n;p,\alpha,\beta,\mu) \\
&= \log\Big[pI_{(X_1)} + \frac{(1-p)\mu^{X_1}}{(1+\mu)^{X_1+1}}\Big] + \sum_{l=2}^{n} \log p(X_l \mid X_{l-1};p,\alpha,\beta,\mu),
\end{aligned}
$$

where $p(X_l \mid X_{l-1};p,\alpha,\beta,\mu)$ is given by (3.1) and $\boldsymbol{\theta} \equiv (p,\alpha,\beta,\mu)$. The MLEs can be easily computed by using the function nlm from statistical package R, taking the CLS estimators as initial values of the function nlm.

Based on MLEs, $\hat{\boldsymbol{\theta}}$, we provide some well-known measures of goodness-of-fit statistics to check the adequacy of a time series model (model selection) as compared with a finite set of models. These statistics are the Akaike information criterion (AIC), Bayesian information criterion (BIC), Hannan-Quinn information criterion (HQIC), consistent Akaike information criterion (CAIC). The smaller values of the statistics are, the better the fit. In Section 4 these statistics are used.

3.4.2. Simulation comparison. Here, we assess the performance of the ML estimators with respect to sample size n. Obviously, the normal equations do not have an explicit solution and the MLEs must be numerically calculated. We provide an algorithm to estimate the model parameters.

Algorithm:

Step 1. Generate $t = 1000$ samples from the ZIGINAR$_{\mathrm{RC}}(1)$ process.

Step 2. Compute the ML estimators using the function "nlm" from statistical software R.

Step 3. For $n = 50, 100, 500, 1000, 5000, 10000$ repeat step 1 to 2.

Step 4. Compute the mean value of the estimates, their standard errors and the asymptotic standard deviations.

The performance of the CLS and ML estimates is checked through a small Monte Carlo simulation using different sample sizes ($n = 50, 100, 500, 1000, 5000, 10000$), where 1000 samples are simulated from the ZIGINAR$_{\mathrm{RC}}(1)$ process. Based on this simulation, Figure 1 displays some sample paths of ZIGINAR$_{\mathrm{RC}}(1)$ process with different choices of the inflation parameter p, where $p = 0.1, 0.3, 0.5, 0.8$ with fixed values $\alpha = 0.9$, $\beta = 0.7$, $\mu = 1$. It can be noted that for larger values of p, the sample paths tend to be smaller values and frequently return zeros. Also, the number of zeros increases by increasing the values of p.

Table 1 gives the mean and standard deviation (in brackets) for the ML and CLS estimators for different values of the parameters p, α, β and μ with different

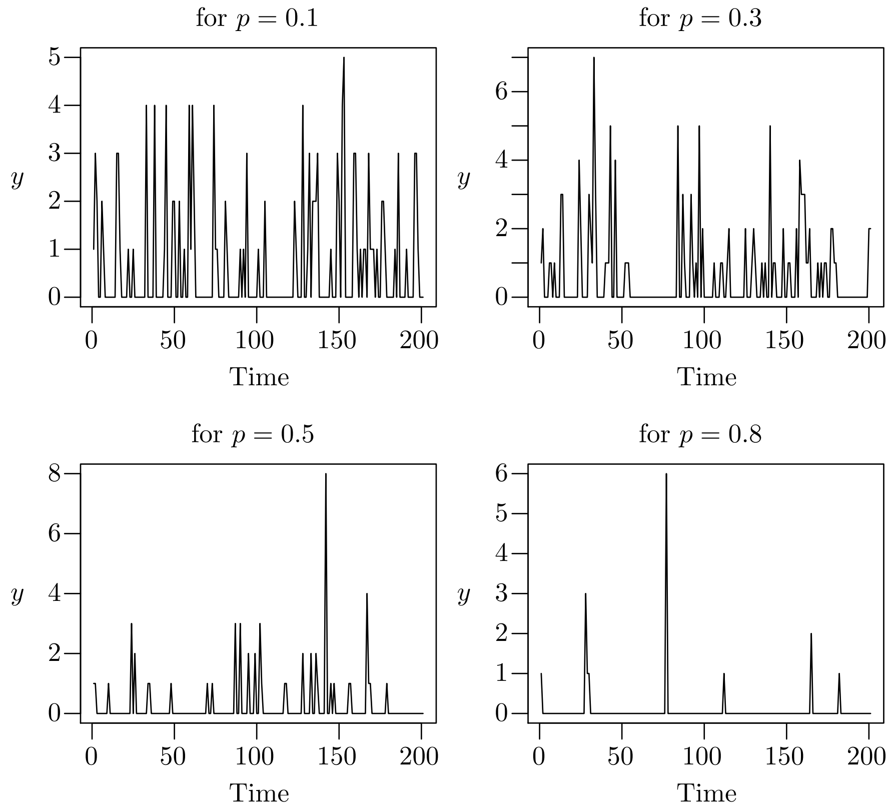

Figure 1. Sample paths of ZIGINAR(1) process for $p = 0.1, 0.3, 0.5, 0.8$ with fixed values to $\alpha = 0.9$, $\beta = 0.7$, $\mu = 1$.

sample sizes. Based on this table, we found that the estimates obtained from the two estimation methods are convergent in their values. Also, increasing the sample size implies smaller standard deviation and the MLEs converge faster to the true values of the parameters. Moreover, we can conclude that the MLEs have the smallest standard deviations and hence the MLEs provide the best performance, which was expected, as the CLS estimates are not solved directly.

On the other hand, the asymptotic behavior of the MLEs is studied via simulation too. We simulated samples of size 1000 and estimated the parameters α, p, β, μ. This procedure was repeated 500 times. Figure 2 shows the Q-Q plots for all parameters ($p = 0.3$, $\alpha = 0.7$, $\beta = 0.5$, $\mu = 2.5$). Also, the K-S test was performed, and we found that the p-values of the estimated parameters α, p, β and μ are 0.7319, 0.5855, 0.6566, 0.9794, respectively. As the Q-Q plots represented by straight line and the p-values are significant, we conclude that the MLEs are asymptotically normal.

	$\hat{p}_{ML}$	$\hat{p}_{\text{CLS}}$	$\hat{\alpha}_{ML}$	$\hat{\alpha}_{\text{CLS}}$	$\hat{\beta}_{ML}$	$\hat{\beta}_{CLS}$	$\hat{\mu}_{ML}$	$\hat{\mu}_{\text{CLS}}$
$(p,\alpha,\beta,\mu) = (0.1, 0.5, 0.7, 0.5)$								
50	0.0979752	0.1048887	0.4939831	0.4894182	0.7128314	0.716774	0.49598213	0.4938558
	(0.0164383)	(0.02044416)	(0.0624852)	(0.09883159)	(0.1323461)	(0.1556039)	(0.0294339)	(0.03481835)
100	0.0987902	0.09779116	0.4972097	0.4926184	0.710575	0.7131973	0.4972911	0.5054949
	(0.0043922)	(0.02012676)	(0.048421)	(0.06861304)	(0.1129023)	(0.1165385)	(0.0207813)	(0.0303127)
500	0.0999054	0.1000696	0.4992737	0.4961113	0.7010677	0.7044796	0.4998507	0.5021612
	(0.0017532)	(0.00730943)	(0.0075823)	(0.02620362)	(0.0064889)	(0.0771452)	(0.0064081)	(0.0271989)
1000	0.09998526	0.1000208	0.5000372	0.4992143	0.6999735	0.70020523	0.5000314	0.49982314
	(0.0002863)	(0.00169942)	(0.0013775)	(0.00782502)	(0.0009248)	(0.00521815)	(0.0032271)	(0.0081962)
5000	0.0999931	0.10000548	0.5000063	0.499868787	0.70000324	0.70003159	0.499999832	0.4999919
	(0.00006724)	(0.0007286)	(0.0004552)	(0.000793)	(0.00005014)	(0.0008299)	(0.00018427)	(0.00071629)
10000	0.1000037	0.0999892	0.50000011	0.4999691	0.69999982	0.70001641	0.50000094	0.49999892
	(0.00001247)	(0.0001195)	(0.0000729)	(0.0001851)	(0.0000023)	(0.00010633)	(0.00004368)	(0.00008525)
$(p,\alpha,\beta,\mu) = (0.3, 0.7, 0.5, 2.5)$								
	$\hat{p}_{ML}$	$\hat{p}_{\text{CLS}}$	$\hat{\alpha}_{ML}$	$\hat{\alpha}_{\text{CLS}}$	$\hat{\beta}_{ML}$	$\hat{\beta}_{CLS}$	$\hat{\mu}_{ML}$	$\hat{\mu}_{\text{CLS}}$
50	0.2972049	0.3050916	0.6902085	0.6837624	0.4992157	0.5119355	2.502221	2.492566
	(0.0582048)	(0.06682771)	(0.0732794)	(0.0951907)	(0.01387967)	(0.06300845)	(0.0548825)	(0.07650927)
100	0.2992074	0.3028493	0.6960248	0.6918238	0.5005822	0.5083244	2.498963	2.510992
	(0.0194723)	(0.0420851)	(0.0313975)	(0.0867492)	(0.01320152)	(0.0605254)	(0.0547093)	(0.06146961)
500	0.2998213	0.2991872	0.7041359	0.6940831	0.5001227	0.504889	2.5007152	2.4909765
	(0.0024892)	(0.0069118)	(0.0079307)	(0.0413913)	(0.01112545)	(0.0263371)	(0.04128757)	(0.05959829)
1000	0.2999895	0.3000649	0.6991842	0.70289301	0.4999602	0.5008471	2.5002069	2.499668
	(0.0006401)	(0.0010641)	(0.0034319)	(0.0082461)	(0.0007804)	(0.01523114)	(0.005417)	(0.01961732)
5000	0.2999995	0.2999923	0.6999607	0.7009923	0.5000049	0.5000502	2.5000638	2.50008214
	(0.0001184)	(0.0004292)	(0.0002287)	(0.0009746)	(0.0000455)	(0.001943)	(0.0007259)	(0.0025017)
10000	0.2999999	0.3000003	0.70000206	0.6999581	0.5000011	0.5000862	2.5000103	2.5000563
	(0.0000263)	(0.0001074)	(0.0000627)	(0.0001493)	(0.0000073)	(0.000382)	(0.0002816)	(0.00053872)
$(p,\alpha,\beta,\mu) = (0.4, 0.8, 0.6, 1.33)$								
	$\hat{p}_{ML}$	$\hat{p}_{\text{CLS}}$	$\hat{\alpha}_{ML}$	$\hat{\alpha}_{\text{CLS}}$	$\hat{\beta}_{ML}$	$\hat{\beta}_{CLS}$	$\hat{\mu}_{ML}$	$\hat{\mu}_{\text{CLS}}$
50	0.3979273	0.3976135	0.8084732	0.7949524	0.6058391	0.6060093	1.3308501	1.331173
	(0.0404739)	(0.04945894)	(0.0330429)	(0.03964038)	(0.0572902)	(0.0773563)	(0.0170805)	(0.02247761)
100	0.3998743	0.3996208	0.79994003	0.7968134	0.5998024	0.60420485	1.3303911	1.3307295
	(0.0284379)	(0.03738236)	(0.0086328)	(0.0102717)	(0.0118632)	(0.0206231)	(0.0160328)	(0.02015414)
500	0.3999218	0.40037512	0.7999891	0.79979542	0.6003787	0.60148419	1.3300382	1.33008255
	(0.0052873)	(0.00825819)	(0.0013047)	(0.00710528)	(0.0056028)	(0.0062197)	(0.00672903)	(0.00811545)
1000	0.3999858	0.40008104	0.8000221	0.79994388	0.6000972	0.6005732	1.3300108	1.330028271
	(0.0032966)	(0.0056607)	(0.0006421)	(0.0010524)	(0.0008233)	(0.0018121)	(0.0008623)	(0.00251135)
5000	0.4000012	0.4000215	0.79999932	0.79999898	0.59999713	0.60001276	1.33000471	1.33000629
	(0.0001192)	(0.0007411)	(0.000362)	(0.0008074)	(0.0004304)	(0.0007329)	(0.0004249)	(0.0006217)
10000	0.4000005	0.4000072	0.80000015	0.7999989	0.5999999	0.6000072	1.33000029	1.33000108
	(0.0000382)	(0.0002299)	(0.0000518)	(0.0002385)	(0.0000725)	(0.0005825)	(0.0001071)	(0.0002513)

Table 1. Some simulation results for the estimates of some true values of the parameters p, α, β, μ with their standard errors in brackets.

4. A climate application

In this section, we investigate an application for the $\text{ZIGINAR}_{\text{RC}}(1)$ process using a real count climate data series with excessive zeros. The series represents the number of tornado deaths recorded monthly and obtained from the National Oceanic and Atmospheric Administration (NOAA), United States, Storm Prediction Center (`http://www.spc.noaa.gov/climo/online/monthly/newm.html#latestmts`), from 2008 to 2014.

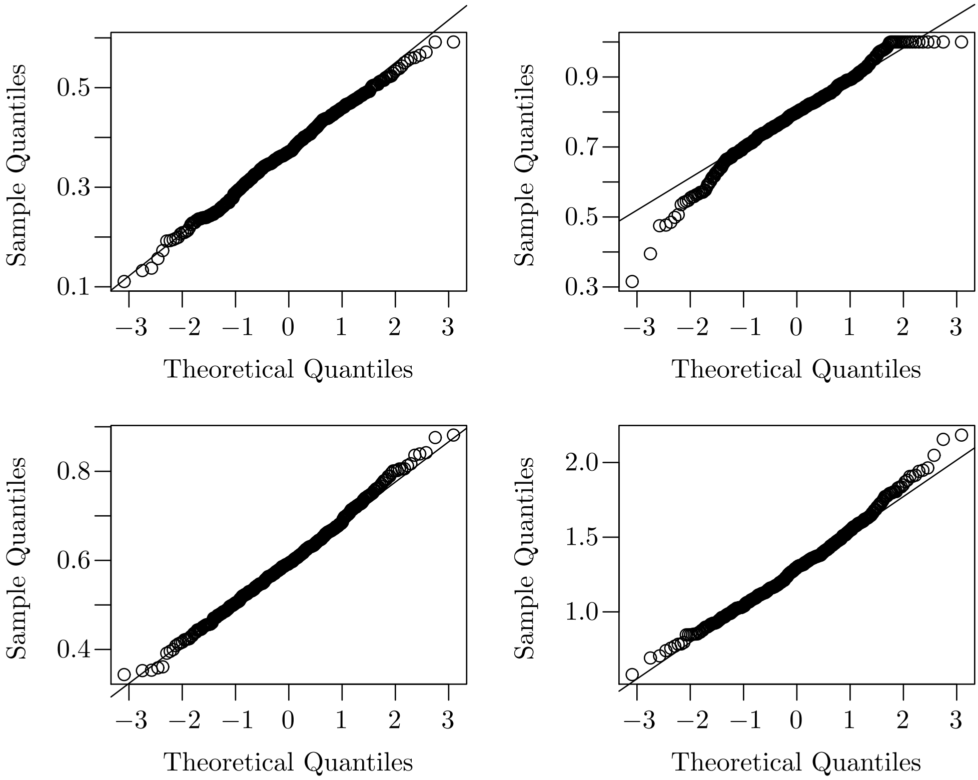

Figure 2. Normal Q-Q plots.

Now, we illustrate how the number of tornado deaths can be modeled by an INAR model with random coefficient. Let X_t be the number of tornado deaths in the t-th month, hence X_t acts as an INAR process and represents the sum of the number of tornado deaths from the previous month denoted by $\alpha \circ X_{t-1}$ and the newly admitted deaths in the current month ε_t, noting that the rate α may be affected by several environmental factors, like the strength and direction of tornado, the population density in the place of tornado, etc., that is α varies randomly over time and then it can be denoted by α_t, therefore X_t could modeled as sum of $\alpha_t \circ X_{t-1}$ and ε_t.

4.1. Data analysis and model selection. The sample paths, autocorrelation functions (ACFs), partial autocorrelation functions (PACFs) and Pareto charts of the series are displayed in Figures 3 and 4. Figure 3 suggests that the first-order autoregressive model is appropriate for the data series. Pareto chart shows that the zeros have the greatest frequency among the other values for the data series and hence the zero-inflated INAR model is more appropriate for the data. Also, the Augmented Dickey-Fuller Test for stationarity has p-value $= 0.019$ that demonstrates

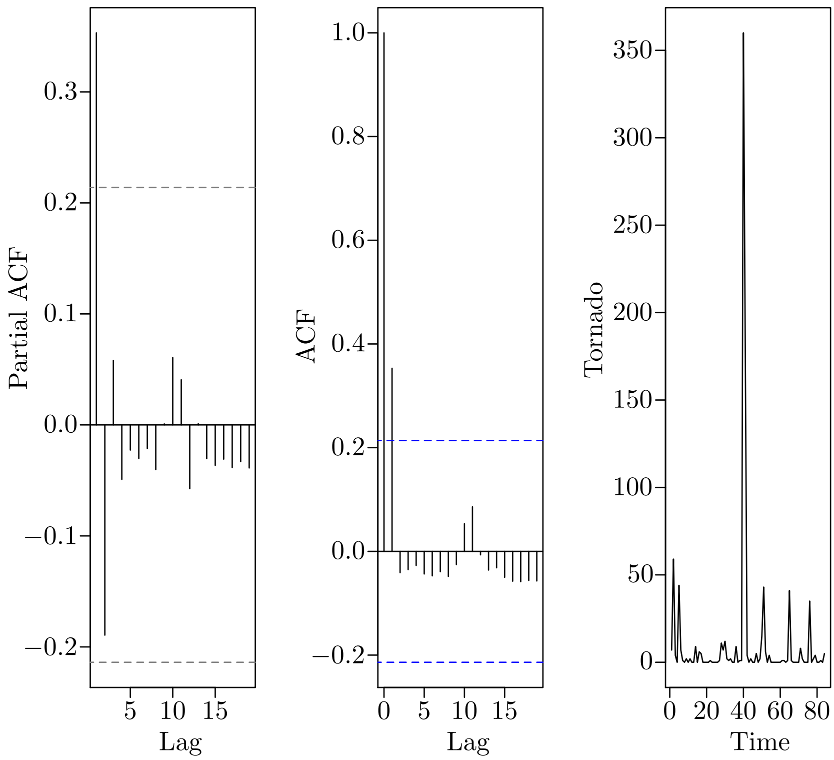

Figure 3. The sample paths, ACFs and PACFs of the tornados data.

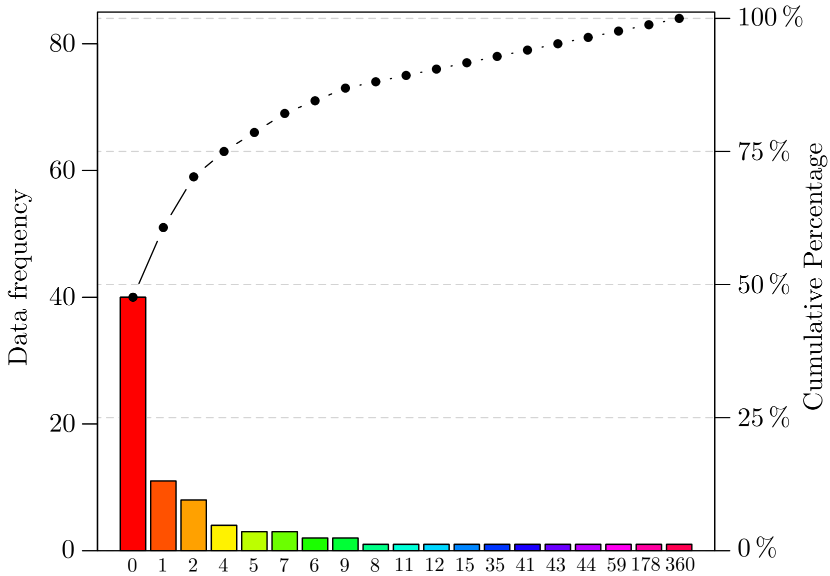

Figure 4. The Pareto chart of the tornados data.

the stationarity of the tornado data. Also, we expect some positive correlation in the data series because climate parameters are changing gradually over time and this expectation is confirmed by noting the value of the sample autocorrelation of the data series at Table 2.

$\overline{X} \equiv$ Sample mean, $S_x^2 \equiv$ Sample variance, $\hat{\varrho} \equiv$ Sample autocorrelation							
$\overline{X}$	S_x^2	Sample $C.V$	$\hat{\varrho}$	$C.V_{\text{ZIG}}$	$C.V_{\text{NB}}$	LRT	p-value
10.89286	1959.109	4.064448	0.3489932	4.852401	2.550726	59.06	1.532108e−14

Table 2. Some measures of the tornado data.

To identify a significant empirical overdispersion towards the data, we test H_0: $\{X_t\}$ is equidispersed versus H_1: $\{X_t\}$ is overdispersed, with significance level $\beta = 0.05$. We will reject H_0 on significance level β if the observed value of the index dispersion, $\hat{I}_x$, exceeds the critical value $1 + z_{1-\beta}\sqrt{2(1+\alpha^2)/n(1-\alpha^2)}$, where $\hat{I}_x = S_x^2/\bar{x}$ and $z_{1-\beta}$ denotes the $(1-\beta)$-quantile of the $N(0,1)$ distribution. Alternatively, we can check if the p-value $1-\Phi\big(\sqrt{n(1-\alpha^2)/2(1+\alpha^2)}(\hat{I}_x - 1)\big)$ falls bellow β, noting that we can replace α by $\hat{\varrho}_x(1)$, see [21].

For tornado data we have mean of 10.89286 and variance of 1959.109. So $\hat{I}_x = 179.85$ and the critical value is 1.286679 and the p-value is 0. The observed value of the index of dispersion, $\hat{I}_x$, exceeds the critical value completely, hence the data series does not stem from an equidispersed INAR(1) process. This conclusion is supported again by comparing the p-value of the data to the considered significance level.

On the other hand, testing zero-inflation for the data set is equivalent to testing $H_0\colon\ p = 0$ against $H_1\colon\ p \neq 0$, and for this purpose, we use the likelihood ratio test (LRT) statistic $-2\{\log L_0(\alpha,\beta,\mu) - \log L_1(p,\alpha,\beta,\mu)\}$, where $L_0(\cdot)$ and $L_1(\cdot)$ are the likelihood values of the $\text{GINAR}_{\text{RC}}(1)$ and the $\text{ZIGINAR}_{\text{RC}}(1)$ models, respectively. Note that the LRT statistic follows the chi-square distribution (asymptotically) with 1 degree of freedom which equals to the number of additional parameters in the model with extra parameters. The LRT statistic and its corresponding p-value of the data set are shown in Table 2. We note that the LRT statistic is greater than the critical value of the test $\chi^2_{(1,0.05)} = 3.841459$ for the data set, hence the statistic indicates zero-inflation in the data and therefore the $\text{ZIGINAR}_{\text{RC}}(1)$ model would fit the data better.

As we can see from Table 2, the coefficient of variation $(C.V)$ of the ZIG distribution is greater than the $C.V$ of the negative binomial (NB) distribution for the data, which adapts (iii) of Remark 1.1. Hence, the ZIG distribution offers more flexibility for modeling the data than the NB distribution.

Based on the results above, the $\mathrm{ZIGINAR_{RC}}(1)$ appears to be more appropriate than the geometric and negative binomial INAR(1) models for the tornado data set.

For selecting the best model for the data series among a finite set of INAR(1) models, we compared the $\mathrm{ZIGINAR_{RC}}(1)$ model to some competitive INAR(1) models with geometric and negative binomial marginals which are GINAR(1) ([4]), NBI-INAR(1) ([1]), NBRCINAR(1) ([24]), NGINAR(1) ([19]), ZMGINAR(1) ([7]) and $\mathrm{GINAR_{RC}}(1)$ (Remark 2.1). For each INAR model, we obtained the maximum likelihood estimates, the four goodness-of-fit statistics (AIC, BIC, HQIC and CAIC) and the root mean squares of the differences of observations and predicted values (RMS). The obtained results, for the data series, are shown in Table 3. As it can be seen from this table, the values of the four goodness-of-fit statistics are the smallest for the $\mathrm{ZIGINAR_{RC}}(1)$ model and also the RMS values are the smallest for the $\mathrm{ZIGINAR_{RC}}(1)$ model among other models. Therefore, we conclude that the $\mathrm{ZIGINAR_{RC}}(1)$ model presents the best model among the other INAR models for fitting the tornado data.

Finally the hypothesis $H_0\colon \sigma_\alpha^2 := \mathrm{Var}(\alpha_t) = 0$ against $H_1\colon \sigma_\alpha^2 := \mathrm{Var}(\alpha_t) > 0$, which is equivalent to $H_0\colon \alpha_t = \alpha$, is checked for the $\mathrm{ZIGINAR_{RC}}(1)$ model as follows. Let $R_t = X_t - E(X_t \mid X_{t-1})$ and $Z_t = (X_{t-1}^2, X_{t-1}, 1)'$. Then by [26] we find that

$$\sqrt{n}(\widehat{\sigma}_\alpha^2 - \sigma_\alpha^2) \xrightarrow{d} N(0, \widetilde{T}'\Gamma^{-1}\omega\Gamma^{-1}\widetilde{T}),$$

where $\widetilde{T} = (1,0,0)'$, $\Gamma = E(Z_t Z_t')$, $\omega = E(Z_t Z_t'(R_t^2 - Z_t' l)^2)$, $l = (\sigma_\alpha^2, \alpha(1-\alpha)-\sigma_\alpha^2, \sigma_\varepsilon^2)$ and $\sigma_\alpha^2 = \beta(1-\beta)$. As the p-value $= 1 - \Phi\big(\sqrt{n}\beta(1-\beta)/\widetilde{T}'\Gamma^{-1}\omega\Gamma^{-1}\widetilde{T}\big) \simeq 0$, the hypothesis of constant coefficient α_t for the $\mathrm{ZIGINAR_{RC}}(1)$ model is rejected.

4.2. Forecasting. Forecasting is an integral part of time series analysis, as it is a planning tool that helps decision makers to foresee the future uncertainty based on the behavior of the past and current observations. As in Subsection 3.2, conditional expectation is one of the most common procedures for forecasting the mean value of time series data. Based on (3.2), the forecast for the value of X at time $t+1$ that is made at time t equals the one-step ahead conditional mean (predictor of the mean) and is given by

$$\hat{X}_{t+1} = E(X_{t+1} \mid X_t) = (1-\beta)\alpha X_t + (1-p)(1-(1-\beta)\alpha)\mu.$$

In practice, the parameters p, α, β, and μ are replaced by their corresponding MLEs $\hat{p}$, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\mu}$. Therefore

$$\hat{X}_{t+1} = (1-\hat{\beta})\hat{\alpha} X_t + (1-\hat{p})(1-(1-\hat{\beta})\hat{\alpha})\hat{\mu}, \tag{4.1}$$

where $t = 1, 2, \ldots, n$, noting that $\hat{X}_1 = E(X_t) = (1 - \hat{p})\hat{\mu}$ and n is the size of the realization of a data set.

To check the adequacy and predictive ability of the selected model $\text{ZIGINAR}_{\text{RC}}(1)$, the actual data series and their predicted values based on (4.1) are plotted and displayed in Figure 5, where $\hat{p}$, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\mu}$ of the data series are given in Tables 3. The predicted values are close to the original data series, which indicates that the $\text{ZIGINAR}_{\text{RC}}(1)$ model can provide a better forecasting for the data series.

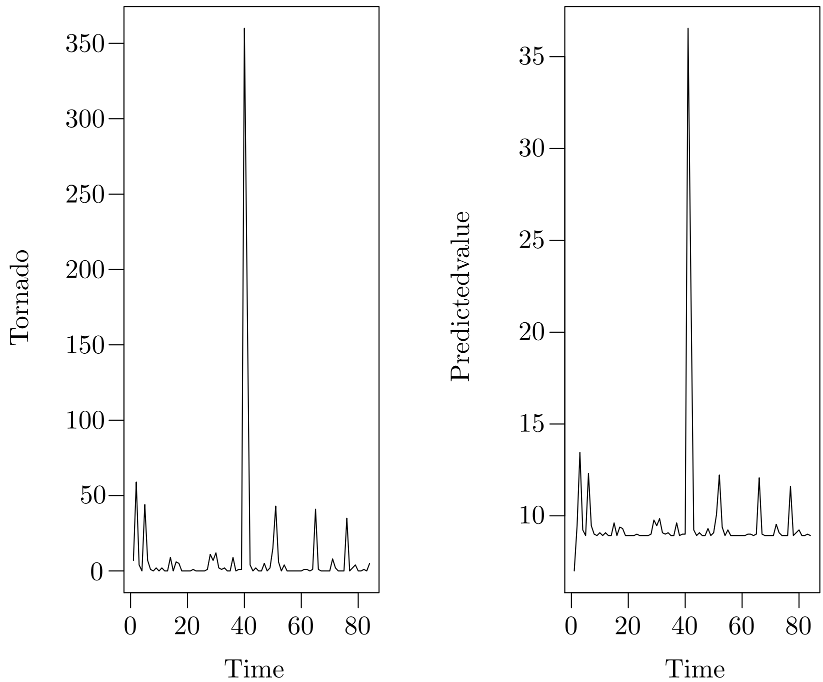

Figure 5. Actual tornados data and its predicted value using ZIGINAR(1).

4.3. Estimation of P_{00} and $E(T_0)$. We showed that the $\text{ZIGINAR}_{\text{RC}}(1)$ process is the best model for the tornado data, so it is of interest to estimate transition probability P_{00} and the expected run length at state 0, $E(T_0)$, of this process. The estimations of P_{00} and $E(T_0)$ are summarized in Table 4 and done by replacing the parameters p, α, β, and μ of their formulas with the corresponding MLEs of the tornado data in Table 3.

The probability $\hat{P}_{00}$ indicates that 50.27 % of tornado data are zeros, and the corresponding average run length is 0.93078.

Model	MLE	AIC	BIC	HQIC	CAIC	RMS
GINAR(1)	$\hat{p} = 0.9370294$ (0.0944031)	554.80	559.66	556.75	554.94	43.26
	$\hat{\alpha} = 0.0818242$ (0.01893175)					
NGINAR(1)	$\hat{p} = 6.1340857$ (0.08569405)	610.83	615.69	612.78	610.97	44.87
	$\hat{\alpha} = -0.0217589$ (0.01987791)					
NBRCINAR(1)	$\hat{n} = 0.16901623$ (0.03357513)	479.86	487.15	482.79	480.16	43.24
	$\hat{p} = 0.01640751$ (0.00610436)					
	$\hat{\varrho} = 0.05513365$ (0.07460729)					
NBIINAR(1)	$\hat{n} = 0.16226302$ (0.02907992)	475.65	482.94	478.58	475.95	41.79
	$\hat{p} = 0.01928195$ (0.00650967)					
	$\hat{\varrho} = 0.22505260$ (0.1818972)					
ZMGINAR(1)	$\hat{\pi} = 0.44146$ (0.006259023)	556.17	563.46	559.1	556.47	44.65
	$\hat{\alpha} = 0.01583$ (0.009439554)					
	$\hat{\mu} = 6.14212$ (2.635587e-05)					
$\text{GINAR}_{\text{RC}}(1)$	$\hat{\alpha} = 0.4835057$ (0.03027246)	518.27	525.56	521.21	518.57	41.97
	$\hat{\beta} = 0.1432110$ (0.05264126)					
	$\hat{\mu} = 15.3530698$ (1.786727)					
$\text{ZIGINAR}_{\text{RC}}(1)$	$\hat{p} = 0.4312814$ (0.06568031)	461.20	470.93	465.11	461.71	41.19
	$\hat{\alpha} = 0.4918816$ (0.02940932)					
	$\hat{\beta} = 0.8440253$ (0.09410729)					
	$\hat{\mu} = 17.0005565$ (1.817263)					

Table 3. Estimated parameters with standard errors in brackets and some goodness-of-fit statistics for tornado data series.

$\hat{P}_{00}$	$\hat{E}(T_0)$
0.5027057	0.9307887

Table 4. Estimation of P_{00} and $E(T_0)$ for the tornado data.

5. Concluding remarks

A new stationary first-order integer-valued autoregressive model with random coefficient and zero-inflated geometric marginal distribution, named $\text{ZIGINAR}_{\text{RC}}(1)$, is introduced, with some sub-models as special cases. Various statistical properties of the model are obtained, for example, the autocorrelation function, spectral density, joint probability generating function, multi-step ahead conditional expectation and variance, the one-step transition probabilities and the survival function of the run length with its expectation. Estimation of the model parameters is assessed by two methods and performance of the estimates of both methods is checked, as well as the behavior of the inflation parameter of the model. The $\text{ZIGINAR}_{\text{RC}}(1)$ model is compared to some competitive INAR(1) models with geometric and negative binomial marginal distribution, and its merit among those models validated using a real climate data set. We conclude that the introduced model can deal with integer-valued time series models with excess zeros and the autoregressive parameter varying with time.

The results for this study can be extended to INAR models with orders higher than one.

Appendices

A. Proof of Property 1

The pgf of $\alpha_t \circ X$ is obtained as

$$\begin{aligned}\varphi_{\alpha_t \circ X}(s) &= E(s^{\alpha_t \circ X}) \\ &= \beta + (1-\beta)E(s^{\alpha \circ X}) \\ &= \beta + (1-\beta)E(E(s^{\alpha \circ X} \mid X)) \\ &= \beta + (1-\beta)\varphi_X(1-\alpha+\alpha s) \\ &= (1+\alpha\mu(\beta + p(1-\beta))(1-s))/(1+\alpha\mu(1-s)).\end{aligned}$$

B. Proof of Property 2

$$\begin{aligned}p(\alpha_t \circ X = 0) &= \beta + (1-\beta)p(\alpha \circ X = 0) = \beta + (1-\beta)\sum_{i=0}^{\infty}(1-\alpha)^i p(X=i) \\ &= \beta + (1-\beta)E((1-\alpha)^X) = \beta + (1-\beta)\varphi_X(1-\alpha) \\ &= \frac{1+\alpha\mu(p+\beta(1-p))}{1+\alpha\mu}, \\ p(\alpha \circ X = 0) &= \frac{1+\alpha\mu p}{1+\alpha\mu} < \frac{1+\alpha\mu(p+\beta(1-p))}{1+\alpha\mu} = p(\alpha_t \circ X = 0).\end{aligned}$$

C. Proof of Property 3

We aim at proving that

$$\alpha_t \circ (X+Y) \overset{\text{d}}{=} \alpha_t \circ X + \alpha_t \circ Y.$$

The pgf of the right-hand side is

$$\begin{aligned}\varphi_{\alpha_t \circ X + \alpha_t \circ Y}(s) &= E(s^{\alpha_t \circ X + \alpha_t \circ Y}) = E(E(s^{\alpha_t \circ X + \alpha_t \circ Y} \mid \alpha_t)) \\ &= \beta + (1-\beta)E(s^{\alpha \circ X + \alpha \circ Y}) \\ &= \beta + (1-\beta)E(s^{\alpha \circ X})E(s^{\alpha \circ Y}) = \beta + (1-\beta)\varphi_{\alpha \circ X}(s)\varphi_{\alpha \circ Y}(s) \\ &= \beta + (1-\beta)E(1-\alpha+\alpha s)^{X+Y}.\end{aligned}$$

Due to $X+Y=Z$, the pgf of the left hand side is

$$\begin{aligned}\varphi_{\alpha_t\circ(X+Y)}(s) = \varphi_{\alpha_t\circ Z}(s) = E(s^{\alpha_t\circ Z}) &= \beta+(1-\beta)E(s^{\alpha\circ Z})\\ &= \beta+(1-\beta)E(1-\alpha+\alpha s)^Z.\end{aligned}$$

By the equality of two pgfs we complete the proof.

D. Proof of Property 4
We aim at proving that

$$\alpha_{t_1}\circ(\alpha_{t_2}\circ X) \overset{\mathrm{d}}{=} (\alpha_{t_1}\alpha_{t_2})\circ X.$$

We know $\alpha_{t_i} = \begin{cases} 0 & \beta_i, \\ \alpha_i & 1-\beta_i, \end{cases}$ $i=1,2$, now consider $\gamma=\alpha_{t_1}\alpha_{t_2}$, hence

$$\gamma = \begin{cases} 0 & 1-(1-\beta_1)(1-\beta_2), \\ \alpha_1\alpha_2 & (1-\beta_1)(1-\beta_2). \end{cases}$$

The pgf of the right-hand side is given by

$$\begin{aligned}\varphi_{(\alpha_{t_1}\alpha_{t_2})\circ X}(s) &= \varphi_{\gamma\circ X}(s) = E(s^{\gamma\circ X})\\ &= \beta_1+\beta_2-\beta_1\beta_2+(1-\beta_1)(1-\beta_2)E(s^{(\alpha_1\alpha_2)\circ X})\\ &= \beta_1+\beta_2-\beta_1\beta_2+(1-\beta_1)(1-\beta_2)\varphi_X(1-\alpha_1\alpha_2+\alpha_1\alpha_2 s),\end{aligned}$$

while the pgf for the left-hand side is

$$\begin{aligned}\varphi_{\alpha_{t_1}\circ(\alpha_{t_2}\circ X)}(s) &= E(s^{\alpha_{t_1}\circ(\alpha_{t_2}\circ X)}) = E[E(s^{\alpha_{t_1}\circ(\alpha_{t_2}\circ X)}) \mid \alpha_{t_1}]\\ &= \beta_1+(1-\beta_1)E(s^{\alpha_1\circ(\alpha_{t_2}\circ X)}) = \beta_1+(1-\beta_1)E[E(s^{\alpha_1\circ(\alpha_{t_2}\circ X)} \mid \alpha_{t_2})]\\ &= \beta_1+(1-\beta_1)[\beta_2+(1-\beta_2)E(s^{\alpha_1\circ(\alpha_2\circ X)})]\\ &= \beta_1+\beta_2-\beta_1\beta_2+(1-\beta_1)(1-\beta_2)E(s^{\alpha_1\circ(\alpha_2\circ X)})\end{aligned}$$

and we know that for the standard binomial thinning operator we have

$$\alpha_1\circ(\alpha_2\circ X) \overset{\mathrm{d}}{=} (\alpha_1\alpha_2)\circ X,$$

hence the proof is completed.

E. Autocovariance function

The autocovariance function of the $\mathrm{ZIGINAR_{RC}}(1)$ process, $\{X_t\}$, is obtained as

$$\begin{aligned}\operatorname{Cov}(X_t, X_{t-k}) &= \beta \operatorname{Cov}(\varepsilon_t, X_{t-k}) + (1-\beta) \operatorname{Cov}(\alpha \circ X_{t-1} + \varepsilon_t, X_{t-k}) \\ &= (1-\beta) \operatorname{Cov}(\alpha \circ X_{t-1}, X_{t-k}) = (1-\beta)\alpha \operatorname{Cov}(X_{t-1}, X_{t-k}) \\ &\;\;\vdots \\ &= (1-\beta)^k \alpha^k \operatorname{Cov}(X_{t-k}, X_{t-k}) = (1-\beta)^k \alpha^k (1-p)\mu((1+p)\mu+1).\end{aligned}$$

F. Joint probability generating function

The joint pgf of the process is obtained as follows:

$$\begin{aligned}\varphi_{X_t,X_{t-1}}(s_1,s_2) &= \beta\varphi_{\varepsilon_t,X_{t-1}}(s_1,s_2) + (1-\beta)\varphi_{\alpha\circ X_{t-1}+\varepsilon_t,X_{t-1}}(s_1,s_2) \\ &= \varphi_{\varepsilon_t}(s_1)[\beta\varphi_{X_{t-1}}(s_2) + (1-\beta)\varphi_{\alpha\circ X_{t-1},X_{t-1}}(s_1,s_2)] \\ &= \varphi_{\varepsilon_t}(s_1)[\beta\varphi_{X_{t-1}}(s_2) + (1-\beta)\varphi_{X_{t-1}}(s_2(1-\alpha+\alpha s_1))] \\ &= \varphi_{\varepsilon_t}(s_1)\Big[\beta\frac{1+\mu p(1-s_2)}{1+\mu(1-s_2)} + (1-\beta)\frac{1+\mu p(1-s_2(1-\alpha+\alpha s_1))}{1+\mu(1-s_2(1-\alpha+\alpha s_1))}\Big].\end{aligned}$$

G. Transition probability

The transition probability from state 0 to state 0 is obtained as

$$\begin{aligned}P_{00} &= p(X_t = 0 \mid X_{t-1} = 0) = p(\varepsilon_t = 0) \\ &= \frac{p}{p+\beta-p\beta} + \frac{(1-p)(1-\alpha)}{1-\alpha(p+\beta-p\beta)}\frac{1}{1+\mu} \\ &\quad + \frac{(1-p)(1-\beta)(\alpha\beta(1-p)-p(1-\alpha))}{(1-\alpha(p+\beta-p\beta))(p+\beta-p\beta)}\frac{1}{1+\alpha\mu(\beta+p(1-\beta))}.\end{aligned}$$

A c k n o w l e d g e m e n t. The authors express their sincere thanks to the referees for careful reading of the article and providing valuable comments. Their effort enabled us to improve the paper in a constructive way.

References

[1] *M. A. Al-Osh, E.-E. A. A. Aly*: First order autoregressive time series with negative binomial and geometric marginals. Commun. Stat. Theory. Methods. *21* (1992), 2483–2492. zbl MR doi
[2] *M. A. Al-Osh, A. A. Alzaid*: First-order integer-valued autoregressive (INAR(1)) process. J. Time. Ser. Anal. *8* (1987), 261–275. zbl MR doi
[3] *E.-E. A. A. Aly, N. Bouzar*: On some integer-valued autoregressive moving average models. J. Multivariate Anal. *50* (1994), 132–151. zbl MR doi
[4] *A. A. Alzaid, M. A. Al-Osh*: First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties. Stat. Neerl. *42* (1988), 53–61. zbl MR doi

[5] *A. A. Alzaid, M. A. Al-Osh*: Some autoregressive moving average processes with generalized Poisson marginal distributions. Ann. Inst. Stat. Math. *45* (1993), 223–232. zbl MR doi
[6] *H. S. Bakouch, M. M. Ristić*: Zero truncated Poisson integer-valued AR(1) model. Metrika *72* (2010), 265–280. zbl MR doi
[7] *W. Barreto-Souza*: Zero-modified geometric INAR(1) process for modelling count time series with deflation or inflation of zeros. J. Time. Ser. Anal. *36* (2015), 839–852. zbl MR doi
[8] *M. Bourguignon, K. L. P. Vasconcellos*: First order non-negative integer valued autoregressive processes with power series innovations. Braz. J. Probab. Stat. *29* (2015), 71–93. zbl MR doi
[9] *D. Gomes, L. C. e Castro*: Generalized integer-valued random coefficient for a first order structure autoregressive (RCINAR) process. J. Stat. Plann. Inference *139* (2009), 4088–4097. zbl MR doi
[10] *P. A. Jacobs, P. A. W. Lewis*: Discrete time series generated by mixtures. I: Correlational and runs properties. J. R. Stat. Soc., Ser. B. *40* (1978), 94–105. zbl MR
[11] *M. A. Jazi, G. Jones, C.-D. Lai*: First-order integer valued AR processes with zero inflated Poisson innovations. J. Time. Ser. Anal. *33* (2012), 954–963. zbl MR doi
[12] *M. A. Jazi, G. Jones, C.-D. Lai*: Integer valued AR(1) with geometric innovations. J. Iran. Stat. Soc. JIRSS *11* (2012), 173–190. zbl MR
[13] *W. C. Khoo, S. H. Ong, A. Biswas*: Modeling time series of counts with a new class of INAR(1) model. Stat. Pap. *58* (2017), 393–416. zbl MR doi
[14] *C. Li, D. Wang, H. Zhang*: First-order mixed integer-valued autoregressive processes with zero-inflated generalized power series innovations. J. Korean Stat. Soc. *44* (2015), 232–246. zbl MR doi
[15] *E. McKenzie*: Some simple models for discrete variate time series. J. Water Res. Bull. *21* (1985), 645–650. doi
[16] *D. F. Nicholls, B. G. Quinn*: Random Coefficient Autoregressive Models: An Introduction. Lecture Notes in Statistics 11, Springer, New York, 1982. zbl MR doi
[17] *H. Pavlopoulos, D. Karlis*: INAR(1) modeling of overdispersed count series with an environmental application. Environmetrics *19* (2008), 369–393. MR doi
[18] *M. Ridout, C. G. B. Demétrio, J. Hinde*: Models for count data with many zeros. Proceedings of the 19th International Biometrics Conference, Cape Town, South Africa. 1998, pp. 179–190.
[19] *M. M. Ristić, H. S. Bakouch, A. S. Nastić*: A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. J. Stat. Plan. Inference. *139* (2009), 2218–2226. zbl MR doi
[20] *M. M. Ristić, A. S. Nastić, H. S. Bakouch*: Estimation in an integer-valued autoregressive process with negative binomial marginals (NBINAR(1)). Commun. Stat., Theory Methods *41* (2012), 606–618. zbl MR doi
[21] *S. Schweer, C. H. Weiß*: Compound Poisson INAR(1) processes: Stochastic properties and testing for overdispersion. J. Comput. Stat. Data Anal. *77* (2014), 267–284. MR doi
[22] *F. W. Steutel, K. van Harn*: Discrete analogues of self-decomposability and stability. Ann. Probab. *7* (1979), 893–899. zbl MR doi
[23] *D. Tjøstheim*: Estimation in nonlinear time series models. Stochastic Processes Appl. *21* (1986), 251–273. zbl MR doi
[24] *C. H. Weiß*: Thinning operations for modeling time series of count—a survey. AStA, Adv. Stat. Anal. *92* (2008), 319–341. MR doi
[25] *H. Zhang, D. Wang*: Inference for random coefficient INAR(1) process based on frequency domain analysis. Commun. Stat., Simulation Comput. *44* (2015), 1078–1100. zbl MR doi
[26] *Z. Zhao, Y. Hu*: Statistical inference for first-order random coefficient integer-valued autoregressive processes. J. Inequal. Appl. *2015* (2015), Paper No. 359, 12 pages. zbl MR doi
[27] *H. Zheng, I. V. Basawa, S. Datta*: First-order random coefficient integer-valued autoregressive processes. J. Stat. Plan. Inference *137* (2007), 212–229. zbl MR doi

[28] *A. F. Zuur, E. N. Ieno, N. J. Walker, A. A. Saveliev, G. M. Smith*: Mixed Effects Models and Extensions in Ecology with R. Statistics for Biology and Health, Springer, New York, 2009. zbl MR doi

Authors' addresses: *Hassan S. Bakouch*, Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt, e-mail: hassan.bakouch@science.tanta.edu.eg, hnbakouch@yahoo.com; *Mehrnaz Mohammadpour*, *Masumeh Shirozhan*, Department of Statistics, Faculty of Mathematical Sciences, University of Mazandaran, Babolsar, Iran, e-mail: m.mohammadpour@umz.ac.ir, ms.shirozhan@gmail.com.