Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 251–272
https://doi.org/10.1214/18-BJPS421

Random environment binomial thinning integer-valued autoregressive process with Poisson or geometric marginal

Zhengwei Liu[a], **Qi Li**[b] **and Fukang Zhu**[a]

[a]*Jilin University*
[b]*Changchun Normal University*

Abstract. To predict time series of counts with small values and remarkable fluctuations, an available model is the r states random environment process based on the negative binomial thinning operator and the geometric marginal. However, we argue that the aforementioned model may suffer from the following two drawbacks. First, under the condition of no prior information, the overdispersed property of the geometric distribution may cause the predictions fluctuate greatly. Second, because of the constraints on the model parameters, some estimated parameters are close to zero in real-data examples, which may not objectively reveal the correlation relationship. For the first drawback, an r states random environment process based on the binomial thinning operator and the Poisson marginal is introduced. For the second drawback, we propose a generalized r states random environment integer-valued autoregressive model based on the binomial thinning operator to model fluctuations of data. Yule–Walker and conditional maximum likelihood estimates are considered and their performances are assessed via simulation studies. Two real-data sets are conducted to illustrate the better performances of the proposed models compared with some existing models.

1 Introduction

Integer-valued autoregressive (INAR) time series models play an important role in the theoretical research and real-life applications over the last few years. These models are constructed usually based on the binomial thinning operator introduced by Steutel and van Harn (1979), which is defined as follows

$$\alpha \circ X := \sum_{i=1}^{X} \xi_i, \quad X > 0, \tag{1.1}$$

and 0 otherwise, where $\xi := \{\xi_i\}$ is a collection of independent and identically distributed (i.i.d.) Bernoulli counting random variables with fixed success probability $\alpha \in [0, 1]$, and X is a non-negative integer-valued random variable independent of ξ. McKenzie (1986) and Al-Osh and Alzaid (1987) introduced the INAR(1) model as follows

$$X_t = \alpha \circ X_{t-1} + Z_t, \quad t \in \mathbb{Z}, \tag{1.2}$$

where $\{Z_t\}$ is a sequence of i.i.d. integer-valued random variables with finite mean and variance. Furthermore, it is also assumed that all thinning operations are performed mutually independent and each of them is independent of $\{Z_t\}$. The autocorrelation function (ACF) of model (1.2) takes the form $\mathrm{Corr}(X_t, X_{t-k}) = \alpha^k$, and the probability generating function (pgf) satisfies $P_X(s) = P_X(1 - \alpha + \alpha s)P_z(s)$.

The binomial thinning operator (1.1) and related time series models were generalized in many ways. Joe (1996) proposed a general way to obtain stationary INAR(1) models of the

Key words and phrases. Binomial thinning, INAR, random environment, time series of counts.
Received July 2018; accepted October 2018.

form $X_t = M_t(X_{t-1}, \alpha) + Z_t$, where M_t is a random operator. Zheng, Basawa and Datta (2006) generalized α in (1.1) to be random and proposed the random coefficient thinning operator, which was further considered by Li et al. (2018) among others. Jazi, Jones and Lai (2012) introduced a new stationary INAR(1) process with zero-inflated Poisson innovations and it was further generalized to zero-and-one inflated case by Qi, Li and Zhu (2019). For the pth-order INAR model, Drost, van den Akker and Werker (2008) provided an efficient estimator of the parameters, showed that it has the local asymptotic normality property. Barczy, Ispány and Pap (2011) obtained the asymptotic behavior of an unstable INAR(p) process. McCabe, Martin and Harris (2011) derived the efficient probabilistic forecasts of the integer-valued random variables by estimating the forecast distribution non-parametrically and proved the asymptotic (non-parametric) efficiency. Parameter estimation for INAR-type models has been discussed extensively. Latour (1998) considered conditional least-squares estimation. Bu, McCabe and Hadri (2008) developed a general maximum likelihood analysis of INAR(p) processes. However it becomes difficult to implement as the order p increases, thus Pedeli, Davison and Fokianos (2015) proposed a simple saddlepoint approximation to the log-likelihood. Drost, van den Akker and Werker (2009) considered a more realistic semiparametric INAR(p) model where there is essentially no restrictions on the innovation distribution, and proposed an (semiparametrically) efficient estimator of the autoregression parameters. Weiß (2018) gave excellent reviews about INAR models.

Nastić, Laketa and Ristić (2016) and Nastić, Laketa and Ristić (2018) have introduced non-stationary integer-valued time series models based on a random environment process. In the first paper, a random environment INAR model has been introduced, while the second paper generalizes this model to high-order models, and this model is further generalized by Laketa, Nastić and Ristić (2018) by relaxing the assumption about the equality of thinning operators and the equality of the maximal orders for the random states. We make the following contributions compared with Nastić, Laketa and Ristić (2016): (1) A simple but practicable INAR model with Poisson marginal is introduced to model the non-stationary nature of time series of counts. (2) We propose an r states random environment binomial thinning INAR model with Poisson marginal, which performs better than existing models. (3) A new generalized r states random environment binomial thinning INAR model and its new related empirical criterion for classifying random environment are introduced.

This paper is organized as follows. The definition of two classes r states random environment INAR(1) processes with Poisson or geometric marginal distribution are introduced in Section 2. In Section 3, we derive some basic stochastic properties, like the conditional moment and the distributional properties. The parameter estimation is given in Section 4 and simulations are given in Section 5. Section 6 is devoted to the application of the proposed models to two real datasets. All proofs are given in the Appendix.

2 Definition

First, we introduce an INAR model. A sequence of the INAR process in time n, denoted by $\{X_n\}$, which is obtained under a certain combination of the environment conditions. We assume that there is a finite number of combinations, denoted by $r \in \{1, 2, \ldots\}$, which also represents the number of different parameter values of the marginal distribution, and possible sets of environment factors are represented by $E_r = \{1, 2, \ldots, r\}$. A sequence of random variables $\{Z_n\}_{n\in\mathbb{N}_0}$, where $\mathbb{N}_0 \equiv \mathbb{N} \cup 0$ and $\mathbb{N} \equiv \{1, 2, \ldots\}$, is called the r states random environment process if it is a Markov chain taking values in E_r.

A sequence of non-negative integer-valued random variables $\{X_n(Z_n)\}_{n\in\mathbb{N}_0}$, where $X_n(Z_n)$ is defined as $\sum_{z=1}^{r} X_n(z_n) I_{\{Z_n=z_n\}}$, is called the r states random environment

INAR(1) process (see Definition 2 in Nastić, Laketa and Ristić (2016)), if it has the following expression

$$X_n(Z_n) = \sum_{i=1}^{X_{n-1}(Z_{n-1})} U_i + \varepsilon_n(Z_{n-1}, Z_n), \quad n \in \mathbb{N},$$

where

$$\varepsilon(Z_{n-1}, Z_n) = \sum_{z_{n-1}=1}^{r} \sum_{z_n=1}^{r} \varepsilon_n(z_{n-1}, z_n) I_{\{Z_{n-1}=z_{n-1}, Z_n=z_n\}},$$

$\{U_i\}$ is a counting sequence of i.i.d. random variables, $\{Z_n\}$ is an r states random environment process, z_n is the realization of the random environment state in time n, I is the indicator function and $\{\varepsilon_n(i, j)\}$, $n \in \mathbb{N}_0$, $i, j \in E_r$, are sequences of i.i.d. random variables, which meet the following assumptions:

(A1) $\{Z_n\}$, $\{\varepsilon_n(1, 1)\}$ and $\{\varepsilon_n(1, 2)\}, \ldots, \{\varepsilon_n(r, r)\}$ are mutually independent for all $n \in \mathbb{N}_0$;
(A2) $\{Z_m\}$ and $\varepsilon_n(i, j)$ are independent of $X_n(l)$ for $n < m$ and any $i, j, l \in E_r$.

The inspiration of our new INAR processes comes from Tang and Wang (2014) and Nastić, Laketa and Ristić (2016), the former discussed a first-order random coefficient INAR model under random environment by introducing a Markov chain with finite state space, while the latter introduced a first-order random coefficient INAR model under random environment with geometric distribution. An r states random environment INAR(1) processes based on the binomial thinning operator with Poisson or geometric marginal distribution is given by the following definition.

Definition 1. A sequence $\{X_n(z_n)\}_{n\in\mathbb{N}_0}$, where z_n is the realization of the r states random environment process $\{Z_n\}$ in time n, is the r states random environment INAR(1) process based on the binomial thinning operator (RrINAR(1)), if $X_n(z_n)$ is defined as

$$X_n(z_n) = \alpha \circ X_{n-1}(z_{n-1}) + \varepsilon_n(z_{n-1}, z_n), \quad n \in \mathbb{N}, \tag{2.1}$$

where $\alpha \in (0, 1)$, $\alpha \circ X_{n-1}(z_{n-1}) = \sum_{i=1}^{X_{n-1}(z_{n-1})} U_i$, the counting sequence $\{U_i\}$, $i \in \mathbb{N}$, incorporated in $\alpha\circ$ makes a sequence of i.i.d. random variables with probability mass function (pmf) given as $P(U_i = 1) = \alpha$, $P(U_i = 0) = 1 - \alpha$, $X_n(z_n)$ has the following two different marginal distributions:

Case 1.1

$$P\big(X_n(z_n) = x\big) = \frac{\mu_{z_n}^x}{x!} e^{-\mu_{z_n}}, \quad x \in \mathbb{N}_0.$$

Case 1.2

$$P\big(X_n(z_n) = x\big) = \frac{\mu_{z_n}^x}{(1 + \mu_{z_n})^{x+1}}, \quad x \in \mathbb{N}_0.$$

Proposition 1. *The bivariate time series $\{X_n(z_n), z_n\}$ given by (2.1) is a first-order Markov time series.*

Proposition 1 can be proved with arguments similar to those in Nastić, Laketa and Ristić (2016), and the handling skills can also be found in Tang and Wang (2014).

Definition 2. A sequence $\{X_n\}_{n\in\mathbb{N}_0}$, is the generalized r states random environment INAR(1) process based on the binomial thinning operator (GRrINAR(1)), if X_n is defined as

$$X_n = \alpha_{z_{n-1}} \circ X_{n-1} + \varepsilon_n(z_{n-1}), \quad n \in \mathbb{N}, \tag{2.2}$$

where $\alpha_{z_{n-1}} \in (0,1)$, $\alpha_{z_{n-1}} \circ X_{n-1} = \sum_{i=1}^{X_{n-1}} U_{i,z_{n-1}}$, the counting sequence $\{U_{i,z_{n-1}}\}_{i\in\mathbb{N}}$, incorporated in $\alpha\circ$ makes a sequence of i.i.d. random variables with pmf given as $P(U_{i,z_{n-1}} = 1) = \alpha_{z_{n-1}}$, $P(U_{i,z_{n-1}} = 0) = 1 - \alpha_{z_{n-1}}$, X_n has the following two different marginal distributions:

Case 2.1

$$P(X_n = x) = \frac{\mu^x}{x!}e^{-\mu}, \quad x \in \mathbb{N}_0.$$

Case 2.2

$$P(X_n = x) = \frac{\mu^x}{(1+\mu)^{x+1}}, \quad x \in \mathbb{N}_0.$$

Note that Definition 1 considers the variation of the marginal distribution, while Definition 2 considers the variation of the thinning operator. Both models are dynamic, but they have different emphasize: one focuses on marginal distribution, the other concentrates on the fluctuation, which behave differently in prediction, see Figure 7 in Section 6 for an intuitive explanation.

3 Properties

In this section, we derive some properties of models (2.1) and (2.2), such as the distribution of the innovation, the correlation structure and the conditional variance of the processes. Because of the structural similarity between two kinds of models, the aforementioned properties are similar. The proofs of all theorems and corollaries are given in the Appendix.

3.1 Properties of RrINAR(1) model

We begin with the distributions of the random variables $\varepsilon_n(1,1)$, $\varepsilon_n(1,2)$, $\ldots$, $\varepsilon_n(r,r)$ in process (2.1).

Theorem 1. *Let $\{X_n(z_n)\}_{n\in\mathbb{N}_0}$ be the process given by (2.1), $\mu_1 > 0$, $\mu_2 > 0$, $\ldots$, $\mu_r > 0$. If $z_{n-1} = i$, $z_n = j$, $i, j \in E_r$, $0 \le \alpha \le \min\{\frac{\mu_s}{\mu_t}; s,t \in E_r\}$, then random variable $\varepsilon_n(i,j)$ has the following distribution.*

Case 1.1

$$\varepsilon_n(i,j) \overset{d}{=} P(\mu_j - \alpha\mu_i).$$

Case 1.2

$$\varepsilon_n(i,j) \overset{d}{=} \begin{cases} \mathrm{Geom}\left(\dfrac{\mu_j}{1+\mu_j}\right), & w.p. \quad 1 - \dfrac{\alpha\mu_i}{\mu_j}, \\ 0, & w.p. \quad \dfrac{\alpha\mu_i}{\mu_j}. \end{cases}$$

Corollary 1. *Let $z_{n-1} = i$ and $z_n = j$, $i, j \in E_r$. The expectation and variance of a random variable $\varepsilon_n(i,j)$ in Definition 1 are given as the following.*

Case 1.1 $\mathrm{E}(\varepsilon_n(i,j)) = \mu_j - \alpha\mu_i$ *and* $\mathrm{Var}(\varepsilon_n(i,j)) = \mu_j - \alpha\mu_i$, *respectively.*

Case 1.2 $\mathrm{E}(\varepsilon_n(i,j)) = \mu_j - \alpha\mu_i$ *and* $\mathrm{Var}(\varepsilon_n(i,j)) = \mu_j(1+\mu_j) - \alpha\mu_i(1+\alpha\mu_i)$, *respectively.*

Next, we consider the covariance and correlation structure of process (2.1).

Theorem 2. *Let* $\{X_n(z_n)\}_{n\in\mathbb{N}_0}$ *be the process given by* (2.1), $\mu_1 > 0$, $\mu_2 > 0$, …, $\mu_r > 0$. *Then* (1) *The covariance function of the random variables* $X_n(z_n)$ *and* $X_{n-k}(z_{n-k})$, $k \in \{0, 1, \ldots, n\}$, *is positive and is given as*:

Case 1.1

$$\gamma_n(k) \equiv \operatorname{Cov}\big(X_n(z_n), X_{n-k}(z_{n-k})\big) = \alpha^k \mu_{z_{n-k}}.$$

Case 1.2

$$\gamma_n(k) \equiv \operatorname{Cov}\big(X_n(z_n), X_{n-k}(z_{n-k})\big) = \alpha^k \mu_{z_{n-k}}(1 + \mu_{z_{n-k}}).$$

(2) *The correlation function of the random variables* $X_n(z_n)$ *and* $X_{n-k}(z_{n-k})$, $k \in \{0, 1, \ldots, n\}$, *is positive*, *less than* 1 *and is given as*:

Case 1.1

$$\rho_n(k) \equiv \operatorname{Corr}\big(X_n(z_n), X_{n-k}(z_{n-k})\big) = \alpha^k \sqrt{\frac{\mu_{z_{n-k}}}{\mu_{z_n}}}.$$

Case 1.2

$$\rho_n(k) \equiv \operatorname{Corr}\big(X_n(z_n), X_{n-k}(z_{n-k})\big) = \alpha^k \sqrt{\frac{\mu_{z_{n-k}}(1 + \mu_{z_{n-k}})}{\mu_{z_n}(1 + \mu_{z_n})}}.$$

The proof of Theorem 2 is similar to Theorem 3 in Nastić, Laketa and Ristić (2016), so the details are omitted here. The following theorem are some regression properties of the RrINAR(1) process.

Theorem 3. *Let* $\{X_n(z_n)\}_{n\in\mathbb{N}_0}$ *be the process given by* (2.1), $\mu_1 > 0$, $\mu_2 > 0$, …, $\mu_r > 0$. *Then*, *the k-step conditional mean and variance of* $X_{n+k}(z_{n+k})$ *on* $X_n(z_n)$ *are the following cases*.

Case 1.1

$$\begin{aligned}
\mathrm{E}\big(X_{n+k}(z_{n+k})|X_n(z_n)\big) &= \alpha^k X_n(z_n) + \mu_{z_{n+k}} - \alpha^k \mu_{z_n}, \quad k \in \mathbb{N}_0,\\
\operatorname{Var}\big(X_{n+k}(z_{n+k})|X_n(z_n)\big) &= \alpha^k(1 - \alpha^k)\big(X_n(z_n) - \mu_{z_n}\big)\\
&\quad + \mu_{z_{n+k}} - \alpha^{2k}\mu_{z_n}, \quad k \in \mathbb{N}_0.
\end{aligned}$$

Case 1.2

$$\begin{aligned}
\mathrm{E}\big(X_{n+k}(z_{n+k})|X_n(z_n)\big) &= \alpha^k X_n(z_n) + \mu_{z_{n+k}} - \alpha^k \mu_{z_n}, \quad k \in \mathbb{N}_0,\\
\operatorname{Var}\big(X_{n+k}(z_{n+k})|X_n(z_n)\big) &= \alpha^k(1 - \alpha^k)\big(X_n(z_n) - \mu_{z_n}\big) + \mu_{z_{n+k}}(1 + \mu_{z_{n+k}})\\
&\quad - \alpha^{2k}\mu_{z_n}(1 + \mu_{z_n}), \quad k \in \mathbb{N}_0.
\end{aligned}$$

3.2 Properties of GRrINAR(1) model

We derive some properties of the GRrINAR(1) process (2.2), the following theorem is about the distributions of the random variables $\varepsilon_n(1), \varepsilon_n(2), \ldots, \varepsilon_n(r)$.

Theorem 4. *Let* $\{X_n\}_{n\in\mathbb{N}_0}$ *be the process given by* (2.2), $\alpha_i \in (0, 1)$, $i \in 1, \ldots, r$. *If* $z_{n-1} = i$, $i \in E_r$, *then random variable* $\varepsilon_n(i)$ *has the following distribution*:

Case 2.1

$$\varepsilon_n(i) \stackrel{d}{=} P(\mu - \alpha_i \mu).$$

Case 2.2

$$\varepsilon_n(i) \stackrel{d}{=} \begin{cases} \text{Geom}\left(\dfrac{\mu}{1+\mu}\right), & w.p. \quad 1-\alpha_i, \\ 0, & w.p. \quad \alpha_i. \end{cases}$$

Corollary 2. *Let $z_{n-1} = i$, $i \in E_r$. The expectation and variance of a random variable $\varepsilon_n(i)$ in Definition 2 are given as*:

Case 2.1 $\mathrm{E}(\varepsilon_n(i)) = \mu - \alpha_i\mu$ *and* $\mathrm{Var}(\varepsilon_n(i)) = \mu - \alpha_i\mu$, *respectively.*

Case 2.2 $\mathrm{E}(\varepsilon_n(i)) = \mu - \alpha_i\mu$ *and* $\mathrm{Var}(\varepsilon_n(i)) = \mu(1+\mu) - \alpha_i\mu(1+\alpha_i\mu)$, *respectively.*

Next, we consider the correlation structure of the GRrINAR(1) model.

Theorem 5. *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be the process given by Definition 2, $\alpha_i \in (0,1)$, $i \in 1,\ldots,r$. The one-step conditional variance of X_{n+1} on X_n is*:

Case 2.1

$$\mathrm{Var}(X_{n+1}|X_n) = \alpha_{z_n}(1-\alpha_{z_n})X_n + (1-\alpha_{z_n})\mu.$$

Case 2.2

$$\mathrm{Var}(X_{n+1}|X_n) = \alpha_{z_n}(1-\alpha_{z_n})X_n + \mu(1+\mu) - \alpha_{z_n}\mu(1+\alpha_{z_n}\mu).$$

According to Theorem 5, in Case 2.1, for fixed μ, if $X_n = 0$, then $\mathrm{Var}(X_{n+1}|X_n) = (1-\alpha_{z_n})\mu$; if $X_n \neq 0$, the conditional variance is a quadratic function of X_n, the symmetry axis is $\frac{X_n-\mu}{2X_n}$, we can calculate the limit of the expectation of the symmetry axis as $\lim_{n\to\infty}\mathrm{E}(\frac{X_n-\mu}{2X_n}) \to 0^-$. Similarly, in Case 2.2, for fixed μ, if $X_n = 0$, then $\mathrm{Var}(X_{n+1}|X_n) = \mu(1+\mu) - \alpha_{z_n}\mu(1+\alpha_{z_n}\mu)$, the symmetry axis is $-\frac{1}{2\mu}$; if $X_n \neq 0$, the symmetry axis is $\frac{X_n-\mu}{2(X_n+\mu^2)}$, the limit of the expectation of the symmetry axis is $\lim_{n\to\infty}\mathrm{E}(\frac{X_n-\mu}{2(X_n+\mu^2)}) \to 0^-$. In this sense, with α_{z_n} increases from 0 to 1, the one-step conditional variance decreases monotonically in both two cases.

4 Estimation

We can see that the RrINAR(1) and GRrINAR(1) models are two dynamic ones, which can adjust their marginal distributions to varying circumstances or their innovation distributions through time. Suppose we have data, a set of realizations: $\{x_1, x_2, \ldots, x_N\}$. It is required to estimate the parameters to fit the data. For RrINAR(1), we can use the K-means clustering to partition N observation into r clusters in which each observation belongs to the cluster with the nearest mean. In this simulation, we use the statistical software R to obtain the sequence $z_1, z_2, \ldots, z_N$. For GRrINAR(1), we can use the relationship between α_{z_i} and its one-step conditional variance. Take $r = 2$ as an example, consider the absolute value between x_i and x_{i+1}, if $|x_{i+1} - x_i| \leq \sigma$, where σ is the standard deviation of its marginal distribution, then $z_i = 1$; else if, $z_i = 2$. Similarly, for $r = 3$, if $|x_{i+1} - x_i| \leq \sigma$, then $z_i = 1$; else if, $\sigma < |x_{i+1} - x_i| \leq 2\sigma$, then $z_i = 2$; else, $z_i = 3$. Using this criterion, we can derive the sequence $z_1, z_2, \ldots, z_{N-1}$.

We consider the Yule–Walker (YW) estimation for the RrINAR(1) model, conditional maximum likelihood (CML) estimation for both RrINAR(1) and GRrINAR(1) models. For GRrINAR(1) model, which contains several different thinning operators, the YW estimation is not considered since the moment estimation usually shows little efficiency in small samples.

4.1 Yule–Walker estimation

The ergodicity of the INAR(1) process ensures the consistency of the sample mean and sample covariance. The following is a brief description of proving consistency. Consider a sample $X_1(z_1)$, $X_2(z_2)$, …, $X_N(z_N)$ of the random environment INAR(1) process $\{X_n(z_n)\}$. Let $z_k \neq i$, $z_{k+1} = z_{k+2} = \cdots = z_n = i$, $z_{n+1} \neq i$, where k and n are natural numbers. The subsample $X_{k+1}(i)$, $X_{k+2}(i)$, …, $X_n(i)$ corresponds to the circumstance i is maximal in the sense that it cannot be expanded neither to the left nor to the right. In other words, all of its elements are in the same circumstance. On account of the ergodicity of the INAR(1) process, we know that these estimators are strongly consistent.

We use the method similar to Nastić, Laketa and Ristić (2016). First, partition the samples into the equivalence classes in the following way:

$$I_k = \{i \in \{1, 2, \ldots, N\} \mid z_i = k\}, \quad k \in \{1, 2, \ldots, r\},$$

$$\bigcup_{k=1}^{r} I_k = \{1, 2, \ldots, N\}, \qquad |I_k| = n_k, \quad n_1 + n_2 + \cdots + n_r = N,$$

$$U_k = \big(X_{k_1}(k), X_{k_2}(k), \ldots, X_{k_{n_k}}(k)\big), \quad k_i \in I_k,$$

$$k_i < k_{i+1}, \quad \forall i \in \{1, 2, \ldots, n_k - 1\},$$

where U_k is the subsample of the initial sample, and it contains all the elements corresponding to the circumstance k but not contain elements belong to other circumstances. This subsample can be partitioned into the maximal subsamples, where "maximal" has the same meaning mentioned above. Let i_k be the number of all the maximal subsamples of the sample U_k. Denote these subsamples as $U_{k,1}, U_{k,2}, \ldots, U_{k,i_k}$. Let us introduce the following symbols $R_{k,l} = \{i \in \{1, 2, \ldots, N\} \mid X_i(z_i) \in U_{k,l}\}$, define the number of elements that $R_{k,l}$ contained is $|R_{k,l}| = n_{k,l}$ and $n_{k,1} + n_{k,2} + \cdots n_{k,i_k} = n_k$. Using the previous conclusions, estimators obtained from the samples $U_{k,l}$ are strongly consistent. We denote $c_{k,l} = (n_{k,l} - 1)I(n_{k,l} > 1)$, where I is the indicator function, these estimators are

$$\widehat{\mu}_{k,l} = \frac{1}{n_{k,l}} \sum_{i \in R_{k,l}} X_i(k), \qquad \widehat{\gamma}_{0,l}^{(k)} = \frac{1}{n_{k,l}} \sum_{i \in R_{k,l}} \big(X_i(k) - \widehat{\mu}_{k,l}\big)^2,$$

$$\widehat{\gamma}_{1,l}^{(k)} = \frac{1}{c_{k,l}} \sum_{\{i,i+1\} \subseteq R_{k,l}} \big(X_{i+1}(k) - \widehat{\mu}_{k,l}\big)\big(X_i(k) - \widehat{\mu}_{k,l}\big).$$

Definition 3. We notice that if under the circumstance k, the estimators obtained from the subsample U_k are defined as

$$\widehat{\mu}_k = \frac{1}{n_k} \sum_{i \in I_k} X_i(k), \qquad \widehat{\gamma}_0^{(k)} = \frac{1}{n_k} \sum_{i \in I_k} \big(X_i(k) - \widehat{\mu}_k\big)^2,$$

$$\widehat{\gamma}_1^{(k)} = \frac{1}{s_k} \sum_{\{i,i+1\} \subseteq I_k} \big(X_{i+1}(k) - \widehat{\mu}_k\big)\big(X_i(k) - \widehat{\mu}_k\big),$$

where $s_k = \sum_{l \in I(n_{k,l} > 1)} c_{k,l}$.

The parameter α can be calculated easily,

$$\widehat{\alpha} = \sum_{k=1}^{r} \frac{n_k}{N} \widehat{\alpha}_k, \quad \text{where } \widehat{\alpha}_k = \frac{\widehat{\gamma}_1^{(k)}}{\widehat{\gamma}_0^{(k)}}.$$

The proposed process is not stationary, but it can be proved similarly as Nastić, Laketa and Ristić (2016) that $\widehat{\alpha}$ is strongly consistent.

4.2 Conditional maximum likelihood estimation

We can obtain the log-likelihood function of the process given by Definition 1, which can be used for parameter estimation of unknown parameters $\mu_1, \mu_2, \ldots, \mu_r, \alpha$. Assume that we have already obtained values x_0 and z_0, which make the joint mass function more readable. Let $Y_i = (X_i, Z_i)$, $y_i = (x_i, z_i)$ and $A = \{Y_s = y_s, 0 \le s < i-1\}$, where x_n is the realization of the process in time n, from Proposition 1 we have

$$P(Y_i = y_i \mid Y_{i-1} = y_{i-1}, A) = P(Z_i = z_i \mid Z_{i-1} = z_{i-1}) \cdot P\left(\sum_{k=1}^{x_{i-1}} U_k + \varepsilon_i(Z_{i-1}, Z_i) = x_i\right),$$

omit the factor $p_{i-1,i} = P(Z_i = z_i \mid Z_{i-1} = z_{i-1})$ and derive the joint log-likelihood function as follows

$$\begin{aligned}\log L &= \log L(x_1, z_1, \ldots x_N, z_N \mid \mu_1, \mu_2, \ldots, \mu_r, \alpha) \\ &= \sum_{i=2}^{N} \log P\big(X_i(z_i) = x_i \mid X_{i-1}(z_{i-1}) = x_{i-1}\big).\end{aligned}$$

Then we have the following cases.

Case 1.1

$$\log L = \sum_{i=2}^{N} \log \left\{ \sum_{k=o}^{\min\{x_{i-1}, x_i\}} \binom{x_{i-1}}{k} \alpha^k (1-\alpha)^{x_{i-1}-k} \cdot \frac{(\mu_{z_i} - \alpha\mu_{z_{i-1}})^{(x_i-k)}}{(x_i-k)!} \cdot e^{-(\mu_{z_i} - \alpha\mu_{z_{i-1}})} \cdot I_{\{x_{i-1}\neq 0\}} + \frac{(\mu_{z_i} - \alpha\mu_{z_{i-1}})^{x_i}}{x_i!} e^{-(\mu_{z_i} - \alpha\mu_{z_{i-1}})} \cdot I_{\{x_{i-1}=0\}} \right\}.$$

Case 1.2

$$\begin{aligned}\log L = \sum_{i=2}^{N} \log \Bigg\{ \Bigg\{ \Bigg\{ \sum_{k=o}^{x_i-1} \binom{x_{i-1}}{k} \alpha^k (1-\alpha)^{x_{i-1}-k} \left(1 - \frac{\alpha\mu_{z_{i-1}}}{\mu_{z_i}}\right) \cdot \frac{\mu_{z_i}^{x_i-k}}{(1+\mu_{z_i})^{x_i-k+1}} \Bigg\} I_{\{x_{i-1}<x_i\}} + \Bigg\{ (1-\alpha)^{x_{i-1}} \left(\left(1 - \frac{\alpha\mu_{z_{i-1}}}{\mu_{z_i}}\right) \frac{1}{1+\mu_{z_i}} \right. \\ \left. + \frac{\alpha\mu_{z_{i-1}}}{\mu_{z_i}} \right) \Bigg\} I_{\{x_i=0\}} + \Bigg\{ \sum_{k=o}^{x_i-1} \binom{x_{i-1}}{k} \alpha^k (1-\alpha)^{x_{i-1}-k} \left(1 - \frac{\alpha\mu_{z_{i-1}}}{\mu_{z_i}}\right) \cdot \frac{\mu_{z_i}^{x_i-k}}{(1+\mu_{z_i})^{x_i-k+1}} + \binom{x_{i-1}}{x_i} \alpha^{x_i} \cdot (1-\alpha)^{x_{i-1}-x_i} \left(\left(1 - \frac{\alpha\mu_{z_{i-1}}}{\mu_{z_i}}\right) \frac{1}{1+\mu_{z_i}} \right. \\ \left. + \frac{\alpha\mu_{z_{i-1}}}{\mu_{z_i}} \right) \Bigg\} I_{\{x_{i-1}\geq x_i\}} \cdot I_{\{x_i\neq 0\}} \Bigg\} I_{\{x_{i-1}\neq 0\}} + \Bigg\{ \Bigg\{ \left(1 - \frac{\alpha\mu_{z_{i-1}}}{\mu_{z_i}}\right) \cdot \frac{\mu_{z_i}^{x_i}}{(1+\mu_{z_i})^{x_i+1}} \Bigg\} I_{\{x_i\neq 0\}} + \Bigg\{ \left(\left(1 - \frac{\alpha\mu_{z_{i-1}}}{\mu_{z_i}}\right) \frac{1}{1+\mu_{z_i}} \right. \\ \left. + \frac{\alpha\mu_{z_{i-1}}}{\mu_{z_i}} \right) \Bigg\} I_{\{x_i=0\}} \Bigg\} I_{\{x_{i-1}=0\}} \Bigg\}.\end{aligned}$$

Similarly, from Definition 2, we have:

Case 2.1

$$\log L = \sum_{i=2}^{N} \log \left\{ \sum_{k=o}^{\min\{x_{i-1},x_i\}} \binom{x_{i-1}}{k} \alpha_{z_{i-1}}^{k} (1-\alpha_{z_{i-1}})^{x_{i-1}-k} \cdot \frac{((1-\alpha_{z_{i-1}})\mu)^{(x_i-k)}}{(x_i-k)!} e^{-(1-\alpha_{z_{i-1}})\mu} \cdot I_{\{x_{i-1}\neq 0\}} + \frac{((1-\alpha_{z_{i-1}})\mu)^{x_i}}{x_i!} e^{-(1-\alpha_{z_{i-1}})\mu} \cdot I_{\{x_{i-1}=0\}} \right\}.$$

Case 2.2

$$\log L = \sum_{i=2}^{N} \log \left\{ \left\{ \left\{ \sum_{k=o}^{x_{i-1}} \binom{x_{i-1}}{k} \alpha_{z_{i-1}}^{k} (1-\alpha_{z_{i-1}})^{x_{i-1}-k} \frac{\mu^{x_i-k}}{(1+\mu)^{x_i-k+1}} \right\} \cdot I_{\{x_{i-1}<x_i\}} + \left\{ (1-\alpha_{z_{i-1}})^{x_{i-1}} \left(\frac{1-\alpha_{z_{i-1}}}{1+\mu} + \alpha_{z_{i-1}} \right) \right\} I_{\{x_i=0\}} + \left\{ \sum_{k=o}^{x_i-1} \binom{x_{i-1}}{k} \alpha_{z_{i-1}}^{k} (1-\alpha_{z_{i-1}})^{x_{i-1}-k} \frac{\mu^{x_i-k}}{(1+\mu)^{x_i-k+1}} + \binom{x_{i-1}}{x_i} \alpha_{z_{i-1}}^{x_i} \cdot (1-\alpha_{z_{i-1}})^{x_{i-1}-x_i} \left(\frac{1-\alpha_{z_{i-1}}}{1+\mu} + \alpha_{z_{i-1}} \right) \right\} \cdot I_{\{x_{i-1}\geq x_i\}} \cdot I_{\{x_i\neq 0\}} \right\} I_{\{x_{i-1}\neq 0\}} + \left\{ \left\{ (1-\alpha_{z_{i-1}}) \frac{\mu_{z_i}^{x_i}}{(1+\mu_{z_i})^{x_i+1}} \right\} \cdot I_{\{x_i\neq 0\}} + \left\{ \left(\frac{1-\alpha_{z_{i-1}}}{1+\mu} + \alpha_{z_{i-1}} \right) \right\} I_{\{x_i=0\}} \right\} I_{\{x_{i-1}=0\}} \right\}.$$

Finally, the CML estimates can be obtained by maximizing these functions in terms of some numerical algorithms. We use the package MaxLik in R to optimize these functions.

5 Numerical simulations

A simulation study was conducted to evaluate the finite sample performances of the YW and CML estimates. We have simulated 500 replications with sample size $n = 200, 400$ for each model, and discussed two practicable cases: the case of three or two states. For the estimation of the parameters, the most important is to determine the Markov chain, we set vector p as the discrete distribution of z_1, and matrix P as the process transition probability matrix. The matrix sets the frequencies of the realized circumstances and it shapes the dynamical structure of the RrINAR(1) and GRrINAR(1) processes. For this, these set-ups were considered as follows:

For Poisson RrINAR(1) model (short as P-RrINAR(1)) and Geometric RrINAR(1) model (short as G-RrINAR(1)):

(1) The P-R3INAR(1) model (three states random environment process based on binomial thinning operator with Poisson marginal distribution (2.1), Case (1.1)) and the G-R3INAR(1) model (three states random environment process based on binomial thinning operator with geometric marginal distribution (2.1), Case (1.2)) with the following three scenarios

Scenario (a): $(\mu_1, \mu_2, \mu_3, \alpha) = (1, 2, 3, 0.2)$, $p = (0.33, 0.34, 0.33)$, $P = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}$;

Scenario (b): $(\mu_1, \mu_2, \mu_3, \alpha) = (4, 5, 6, 0.5)$, $p = (0.1, 0.8, 0.1)$, $P = \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}$;

Scenario (c): $(\mu_1, \mu_2, \mu_3, \alpha) = (1, 2, 6, 0.1)$, $p = (0.45, 0.3, 0.25)$, $P = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.35 & 0.4 & 0.25 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$.

(2) P-R2INAR(1) and G-R2INAR(1) with the following three scenarios

Scenario (d): $(\mu_1, \mu_2, \alpha) = (1, 2, 0.3)$, $p = (0.5, 0.5)$, $P = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}$;

Scenario (e): $(\mu_1, \mu_2, \alpha) = (4, 5, 0.6)$, $p = (0.5, 0.5)$, $P = \begin{pmatrix} 0.3 & 0.7 \\ 0.7 & 0.3 \end{pmatrix}$;

Scenario (f): $(\mu_1, \mu_2, \alpha) = (1, 3, 0.2)$, $p = (0.8, 0.2)$, $P = \begin{pmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \end{pmatrix}$.

For Poisson GRrINAR(1) model (short as P-GRrINAR(1)) and Geometric GRrINAR(1) model (short as G-GRrINAR(1)):

(3) The P-GR3INAR(1) model (the general three states random environment process based on binomial thinning operator with Poisson marginal distribution (2.2), Case (2.1)); and the G-GR3INAR(1) model (the general three states random environment process based on binomial thinning operator with geometric marginal distribution (2.2), Case (2.2)) with the following three scenarios

Scenario (g): $(\alpha_1, \alpha_2, \alpha_3, \mu) = (0.5, 0.3, 0.1, 5)$, $p = (0.33, 0.34, 0.33)$, $P = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}$;

Scenario (h): $(\alpha_1, \alpha_2, \alpha_3, \mu) = (0.7, 0.5, 0.3, 2)$, $p = (0.1, 0.8, 0.1)$, $P = \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}$;

Scenario (i): $(\alpha_1, \alpha_2, \alpha_3, \mu) = (0.6, 0.4, 0.2, 3)$, $p = (0.45, 0.3, 0.25)$, $P = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.35 & 0.4 & 0.25 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$,

where α_1, α_2, α_3 represent the different thinning operators belonging to three different random environment states (Definition 2).

(4) P-GR2INAR(1) and G-GR2INAR(1) with the following three scenarios

Scenario (j): $(\alpha_1, \alpha_2, \mu) = (0.7, 0.3, 3)$, $p = (0.5, 0.5)$, $P = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}$;

Scenario (k): $(\alpha_1, \alpha_2, \mu) = (0.6, 0.4, 2)$, $p = (0.5, 0.5)$, $P = \begin{pmatrix} 0.3 & 0.7 \\ 0.7 & 0.3 \end{pmatrix}$;

Scenario (l): $(\alpha_1, \alpha_2, \mu) = (0.5, 0.2, 5)$, $p = (0.8, 0.2)$, $P = \begin{pmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \end{pmatrix}$.

In this simulation, the transition probability matrices were given beforehand to obtain the desired process dynamics, which are responsible for setting the dynamic structure of the process but not the parameters of the process. The simulation results are given in Tables 1–4. It can be seen that as the sample size increases, the estimates seem converge to the true parameter values with the root mean square errors (RMSEs) decreasing towards 0. The two estimation methods both perform well, and CML gave smaller RMSEs than YW in most cases.

6 Illustrative examples

In applications of INAR models, researchers may face that the time series are not stationary. The stationary INAR models may not be the best choice but usually unavoidable. We have obtained two time series representing a monthly counting of drug reselling (DRUGS) from the Forecasting Principles website (http://www.forecastingprinciples.com). These crimes are reported in the 27th (Figure 1) and the 24th (Figure 5) police car beats in Pittsburgh from January 1990 to December 2001, each constituting a sequence of 144 observations. The data in the 27th police car beat was also discussed in Nastić, Laketa and Ristić (2016).

The first step in standard INAR modeling is to obtain the plots of the time series: the ACF and partial ACF (PACF). From the ACF and PACF plots (Figures 1 and 5), we find that modeling the counts using INAR(1) is reasonable. In the time series plots, it is not difficult to see that besides small jumps, in the last 2 years, there is a steady and permanent significant

Table 1 *Mean of estimates*, *RMSE* (*within parentheses*) *for* P-R3INAR(1) *and* G-R3INAR(1)

Scenario	N	Method	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\mu_3}$	$\hat{\alpha}$
P (a)	200	YW	1.0269 (0.1381)	2.0311 (0.2012)	3.0253 (0.2385)	0.1949 (0.1245)
		CML	0.9988 (0.1299)	2.0008 (0.1852)	3.0224 (0.2199)	0.2041 (0.0657)
	400	YW	1.0181 (0.0952)	2.0122 (0.1363)	3.0020 (0.1616)	0.1981 (0.0868)
		CML	0.9976 (0.0925)	2.0030 (0.1292)	3.0079 (0.1688)	0.1975 (0.0435)
(b)	200	YW	4.0198 (0.3497)	5.0080 (0.3536)	5.9958 (0.4200)	0.4787 (0.1200)
		CML	4.0001 (0.3011)	5.0039 (0.3217)	5.9976 (0.3795)	0.4998 (0.0489)
	400	YW	3.9892 (0.2537)	4.9965 (0.2327)	5.9856 (0.3091)	0.4893 (0.0901)
		CML	4.0010 (0.2146)	4.9951 (0.2163)	6.0051 (0.2641)	0.5011 (0.0326)
(c)	200	YW	1.0336 (0.1411)	2.0078 (0.1851)	6.0022 (0.3019)	0.0986 (0.1288)
		CML	0.9922 (0.1326)	1.9928 (0.1836)	6.0192 (0.3309)	0.0991 (0.0639)
	400	YW	1.0128 (0.0871)	2.0056 (0.1251)	5.9933 (0.2191)	0.1012 (0.0893)
		CML	1.0019 (0.0954)	2.0019 (0.1277)	6.0021 (0.2187)	0.0990 (0.0431)
Geo (a)	200	YW	1.0195 (0.1839)	2.0017 (0.3314)	2.9859 (0.4569)	0.1944 (0.1307)
		CML	1.0014 (0.1804)	1.9817 (0.3188)	3.0203 (0.4303)	0.1986 (0.0402)
	400	YW	1.0197 (0.1319)	2.0039 (0.2290)	3.0233 (0.3226)	0.1959 (0.0935)
		CML	1.0010 (0.1237)	1.9948 (0.2135)	2.9906 (0.3020)	0.2001 (0.0272)
(b)	200	YW	4.0079 (0.7932)	5.0184 (0.7712)	6.0141 (1.1440)	0.4712 (0.1458)
		CML	3.9647 (0.6584)	4.9762 (0.7384)	5.9574 (0.9193)	0.5008 (0.0233)
	400	YW	3.9964 (0.5490)	5.0242 (0.5713)	5.9948 (0.8045)	0.4813 (0.1027)
		CML	4.0032 (0.4394)	4.9809 (0.4994)	5.9941 (0.6342)	0.4995 (0.0170)
(c)	200	YW	1.0342 (0.1857)	2.0200 (0.3231)	5.9937 (0.7944)	0.1017 (0.1300)
		CML	0.9813 (0.1810)	1.9675 (0.3006)	6.0350 (0.8026)	0.1011 (0.0314)
	400	YW	1.0158 (0.1289)	2.0218 (0.2126)	5.9913 (0.5910)	0.0995 (0.0896)
		CML	1.0022 (0.1213)	2.0003 (0.2173)	6.0031 (0.5616)	0.1015 (0.0211)

increase of the observed period. So we assume that these two counting sequences may arise from different environments, or may have different parameter turbulence through time, we can apply our models to them.

We model these two datasets using RrINAR(1) in (2.1), GRrINAR(1) in (2.2) and Geometric RrINAR(1) based on the negative binomial thinning operator (short as RrNGINAR(1)) from Nastić, Laketa and Ristić (2016). For Example 1 (27th police car beat), we fit the models mentioned above based on the whole sequence of 144 observations, obtain the result of which random environment each data belongs to, then estimate the parameters, and generate the fitting data according to the known random environment. Finally, we calculate the value of Akaike information criterion (AIC), Bayesian information criterion (BIC) and the RMSE between 144 fitting data and its corresponding observations. For Example 2 (24th police car beat), we divide the sequence of 144 observations into two parts, the first is the former 132 observations, the second is the last 12 observations. The transition probability matrix is estimated by the method in Anderson and Goodman (1957). According to this matrix, we predict which random environment the last 12 data belonging to, and then generate the last 12 predictions. Finally, we calculate the value of AIC and BIC. We also calculate the RMSE between the last 12 predictions and the last 12 observations.

Example 1

Establishing the above models according to the the sequence of 144 observations. Then we calculate AIC and BIC. We also calculate RMSE between the observations and the pre-

Table 2 *Mean of estimates*, *RMSE* (*within parentheses*) *for* P-R2INAR(1) *and* G-R2INAR(1)

Scenario	N	Method	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\alpha}$
P (d)	200	YW	1.0331 (0.1236)	2.0173 (0.1678)	0.2921 (0.0946)
		CML	0.9991 (0.1163)	1.9975 (0.1731)	0.2940 (0.0646)
	400	YW	1.0040 (0.0863)	2.0091 (0.1239)	0.2939 (0.0695)
		CML	0.9945 (0.0818)	1.9928 (0.1139)	0.3001 (0.0418)
(e)	200	YW	4.0068 (0.2920)	4.9902 (0.3189)	0.5846 (0.1427)
		CML	4.0113 (0.2896)	5.0154 (0.3321)	0.5983 (0.0402)
	400	YW	4.0097 (0.2210)	5.0042 (0.2436)	0.5924 (0.1023)
		CML	3.9776 (0.2167)	4.9859 (0.2479)	0.5987 (0.0294)
(f)	200	YW	1.0349 (0.1247)	2.9969 (0.1788)	0.1927 (0.0955)
		CML	1.0048 (0.1067)	3.0098 (0.1842)	0.1965 (0.0663)
	400	YW	1.0146 (0.0803)	3.0026 (0.1325)	0.1970 (0.0696)
		CML	1.0029 (0.0822)	3.0009 (0.1305)	0.2027 (0.0397)
Geo (d)	200	YW	1.0350 (0.1794)	2.0197 (0.2995)	0.2877 (0.0971)
		CML	0.9868 (0.1651)	1.9990 (0.2995)	0.2966 (0.0424)
	400	YW	1.0061 (0.1254)	1.9981 (0.2083)	0.2954 (0.0750)
		CML	0.9953 (0.1196)	1.9905 (0.1955)	0.2964 (0.0312)
(e)	200	YW	4.0169 (0.6968)	5.0554 (0.8436)	0.5811 (0.1764)
		CML	4.0171 (0.6076)	4.9637 (0.7147)	0.6000 (0.0223)
	400	YW	4.0252 (0.4566)	5.0410 (0.5786)	0.5946 (0.1380)
		CML	3.9782 (0.4064)	4.9619 (0.5038)	0.5992 (0.0158)
(f)	200	YW	1.0414 (0.1644)	3.0233 (0.3739)	0.1905 (0.0999)
		CML	0.9910 (0.1511)	2.9958 (0.3549)	0.1984 (0.0379)
	400	YW	1.0177 (0.1200)	2.9907 (0.2687)	0.1961 (0.0750)
		CML	0.9974 (0.1113)	2.9953 (0.2495)	0.1993 (0.0268)

Table 3 *Mean of estimates*, *RMSE* (*within parentheses*) *for* P-GR3INAR(1) *and* G-GR3INAR(1)

Scenario	N	Method	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\alpha_3}$	$\hat{\mu}$
P (g)	200	CML	0.4989 (0.0896)	0.3003 (0.1147)	0.0972 (0.1318)	5.0119 (0.2088)
	400		0.5022 (0.0584)	0.2940 (0.0796)	0.0971 (0.0897)	4.9974 (0.1579)
(h)	200	CML	0.6965 (0.0612)	0.4922 (0.0800)	0.2913 (0.1224)	1.9983 (0.1619)
	400		0.6966 (0.0427)	0.4970 (0.0577)	0.2925 (0.0817)	1.9941 (0.1242)
(i)	200	CML	0.5946 (0.0731)	0.3926 (0.1006)	0.2047 (0.1126)	2.9990 (0.1723)
	400		0.5962 (0.0511)	0.3957 (0.0682)	0.1954 (0.0814)	3.0052 (0.1307)
Geo (g)	200	CML	0.4976 (0.0443)	0.3024 (0.0526)	0.1003 (0.0659)	5.0250 (0.4941)
	400		0.4997 (0.0317)	0.3012 (0.0363)	0.0996 (0.0420)	4.9920 (0.3455)
(h)	200	CML	0.7009 (0.0487)	0.5009 (0.0532)	0.3030 (0.0792)	2.0071 (0.2880)
	400		0.6997 (0.0340)	0.4982 (0.0407)	0.3036 (0.0524)	2.0018 (0.1970)
(i)	200	CML	0.5992 (0.0491)	0.3987 (0.0605)	0.1995 (0.0735)	2.9910 (0.3526)
	400		0.5979 (0.0326)	0.3977 (0.0417)	0.2008 (0.0465)	3.0057 (0.2497)

dicted values. For P-RrINAR(1), G-RrINAR(1) and RrNGINAR(1), we separate the sequence for two or three possible random states by K-means clustering algorithm (Figure 2). For P-GRrINAR(1) and G-GRrINAR(1), we use the criterion in Section 4, and then separate the random environment for two or three random states. After that, using CML in Section 4.2

Table 4 *Mean of estimates*, *RMSE* (*within parentheses*) *for* P-GR2INAR(1) *and* G-GR2INAR(1)

Scenario	N	Method	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\mu}$
P (j)	200	CML	0.7005 (0.0471)	0.2952 (0.0838)	3.0014 (0.2044)
	400		0.6963 (0.0326)	0.2964 (0.0654)	2.9963 (0.1424)
(k)	200	CML	0.5920 (0.0588)	0.3956 (0.0849)	2.0044 (0.1690)
	400		0.6015 (0.0386)	0.3953 (0.0549)	2.0026 (0.1194)
(l)	200	CML	0.4947 (0.0792)	0.1988 (0.0931)	4.9885 (0.2251)
	400		0.4977 (0.0540)	0.1962 (0.0680)	5.0052 (0.1604)
Geo (j)	200	CML	0.6993 (0.0302)	0.2985 (0.0541)	2.9877 (0.3946)
	400		0.6985 (0.0223)	0.2985 (0.0387)	2.9898 (0.2779)
(k)	200	CML	0.5961 (0.0457)	0.3934 (0.0559)	2.0035 (0.2788)
	400		0.6011 (0.0299)	0.4003 (0.0377)	2.0060 (0.2008)
(l)	200	CML	0.4996 (0.0376)	0.2025 (0.0434)	4.9998 (0.5060)
	400		0.4989 (0.0265)	0.2007 (0.0303)	4.9899 (0.3524)

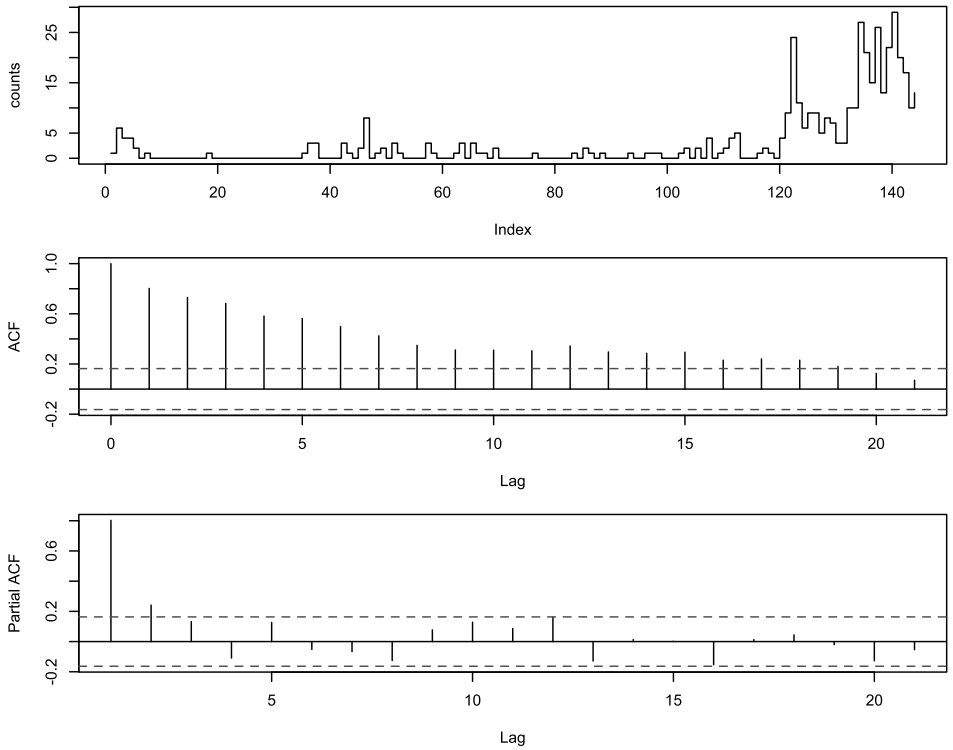

Figure 1 *27th DRUGS series*, *autocorrelations and partial autocorrelations*.

to estimate the parameters. The results in Table 5 about RrNGINAR(1) are not exactly the same as those of Nastić, Laketa and Ristić (2016). Because the result of K-means cluster analysis using R is often not unique and the adoption of different algorithms, the little difference is reasonable. According to the results of Table 5, P-R3INAR(1) has the smallest AIC, BIC and RMSE, to this set of data, which means that P-R3INAR(1) perform best among those models.

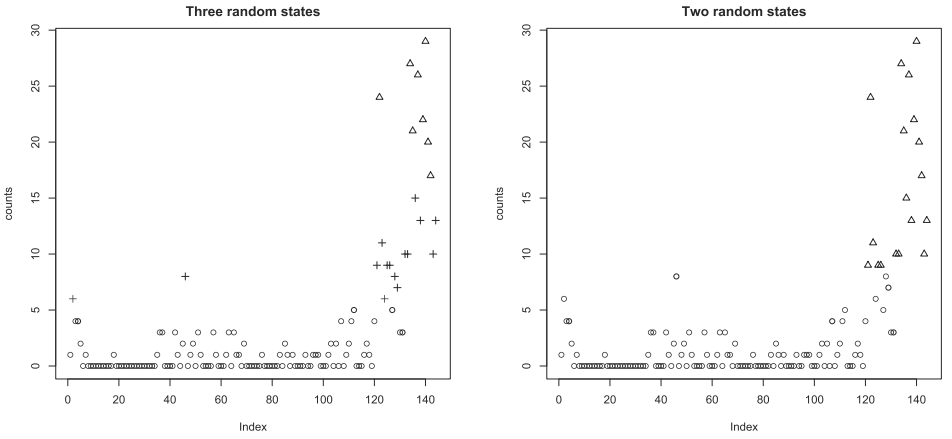

Figure 2 *27th DRUGS data in three or two states.*

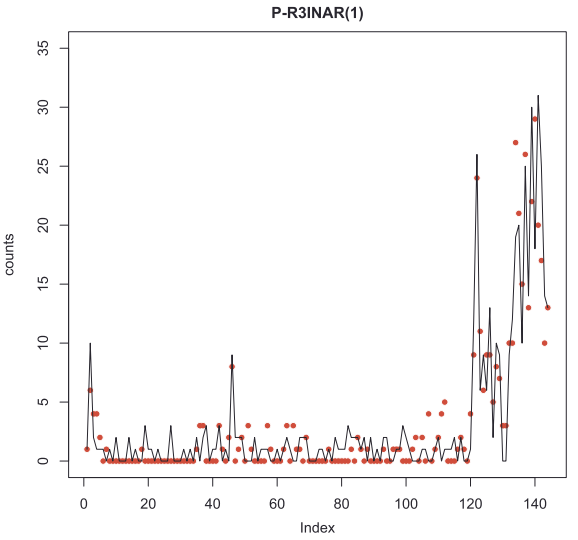

Figure 3 *27th DRUGS data* (•) *and fitted values from* P-R3INAR(1).

The fitted P-R3INAR(1) model for 27th DRUGS data is given in Figure 3. To further examine the adequacy of the fitted model, let us consider the Pearson residuals, defined by

$$r_{1t} = \frac{X_t - \hat{\alpha} X_{t-1}(z_{t-1}) - \hat{\mu}_{z_t} + \hat{\alpha}\hat{\mu}_{z_{t-1}}}{[\hat{\alpha}(1-\hat{\alpha})(X_{t-1}(z_{t-1}) - \hat{\mu}_{z_{t-1}}) + \hat{\mu}_{z_t} - \hat{\alpha}^2\hat{\mu}_{z_{t-1}}]^{\frac{1}{2}}}.$$

We can also calculate the mean square error of Pearson residuals, which is equal to $\sum_{t=1}^{n} r_{1t}^2/(n-p)$, where p denotes the number of estimated parameters. Table 6 gives some characteristics of the residuals for P-R3INAR(1) model. Figure 4 plots the ACF and PCF of the Pearson residuals. There is no evidence of any correlation within the residuals, a finding is supported by the Ljung–Box statistic of 20.4391 based on 15 lags (because $\chi^2_{0.05}(14) = 23.6847$).

Example 2

Dividing the sequence of 144 observations into two parts, the first is the former 132 observations, remaining the last 12 observations as the second part. We establish the above models according to the first part, then estimate the transition probability matrix between the

Table 5 *Parameter estimates, RMSE, AIC and BIC for modeling of the 27th DRUGS data*

model	CML	RMSE	AIC	BIC
P-R3INAR(1)	$\mu_1 = 0.8166$ $\mu_2 = 23.0785$ $\mu_3 = 9.3196$ $\alpha = 0.0348$	2.4352	453.3236	465.2029
G-R3INAR(1)	$\mu_1 = 0.8241$ $\mu_2 = 21.6645$ $\mu_3 = 8.5302$ $\alpha = 0.0348$	5.1565	471.4125	483.2917
R3NGINAR(1)	$\mu_1 = 0.8215$ $\mu_2 = 23.6046$ $\mu_3 = 8.5531$ $\alpha = 0.0333$	4.3389	471.5221	483.4014
P-GR3INAR(1)	$\alpha_1 = 0.8802$ $\alpha_2 = 0.4463$ $\alpha_3 = 0.1207$ $\mu = 8.9221$	7.1209	752.1834	764.0626
G-GR3INAR(1)	$\alpha_1 = 0.4411$ $\alpha_2 = 0.6265$ $\alpha_3 = 0.4740$ $\mu = 2.9025$	6.8526	535.7508	547.6301
P-R2INAR(1)	$\mu_1 = 1.0712$ $\mu_2 = 15.7599$ $\alpha = 0.0431$	3.7776	564.7871	573.6965
G-R2INAR(1)	$\mu_1 = 1.1144$ $\mu_2 = 12.7787$ $\alpha = 0.0557$	4.8769	497.4740	506.3834
R2NGINAR(1)	$\mu_1 = 1.1085$ $\mu_2 = 12.9136$ $\alpha = 0.0519$	4.4024	497.6248	506.5342
P-GR2INAR(1)	$\alpha_1 = 0.8755$ $\alpha_2 = 0.3391$ $\mu = 8.5725$	7.0440	756.8653	765.7748
G-GR2INAR(1)	$\alpha_1 = 0.4427$ $\alpha_2 = 0.5519$ $\mu = 2.9762$	6.7889	536.0080	544.9174

Table 6 *Some characteristics of the 27th DRUGS data residuals*

model	mean	standard deviation	MSE	Ljung–Box statistic
P-R3INAR(1)	0.0055	1.3203	1.7810	20.4391

environment states in the former 132 observations. Later, we forecast the last 12 probable environment states by using the transition probability matrix and then forecast its related observations. Finally, we calculate the AIC, BIC of the former 132 observations, and RMSE of difference between the last 12 observations and the related predictions.

For P-RrINAR(1), G-RrINAR(1) and RrNGINAR(1), we separate the sequence for two or three random states by K-means clustering algorithm (Figure 6). For P-GRrINAR(1) and G-GRrINAR(1), using the criterion in Section 4, we separate the random environment for

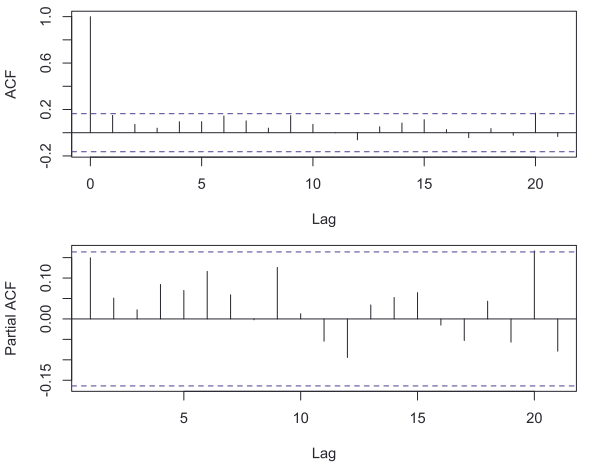

Figure 4 *Pearson residual analysis for* 27*th DRUGS data*: *the autocorrelation function and the partial autocorrelation function of residuals*.

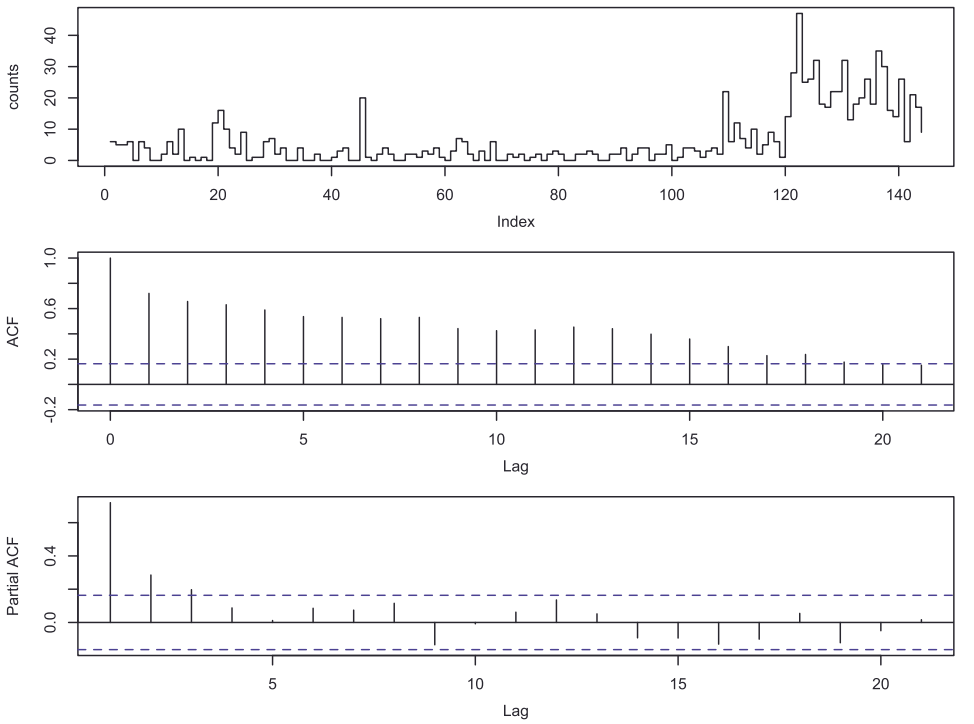

Figure 5 24*th DRUGS series*, *autocorrelations and partial autocorrelations*.

two or three random states. We can estimate the transition probability matrix by the method in Anderson and Goodman (1957). After that, using CML in Section 4.2 to estimate the parameters of the INAR models. According to the results of Table 7, P-R3INAR(1) has the smallest AIC, BIC and second-smallest RMSE, which means that P-R3INAR(1) perform best among those models. Figure 7 gives the prediction of the last 12 observations based on ten models of the 24th DRUGS data.

From these two examples, we show that P-R3INAR(1) model has the best performance among these models when fitting the above two datasets. The reason may be the sample sizes of the two samples are not large, the number of subsamples of per states is small, the variance

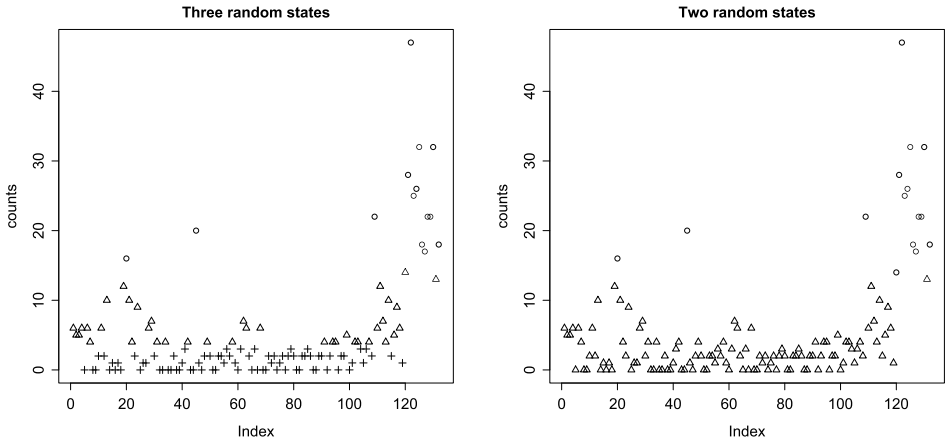

Figure 6 *24th DRUGS data in three or two states.*

of Poisson marginal distribution is smaller than geometric distribution when they have the same expectation. Larger fluctuation of geometric distribution makes it difficult to fit the marginal distribution. We also found that the number of data states has a great impact on the estimation of these models. For example, in Example 1, P-R3INAR(1) performs better than G-R3INAR(1), but G-R2INAR(1) performs better than P-R2INAR(1). How to determine the number of random states before estimation is still open. If there is prior information, we can choose the model according to the actual situation. If not, estimating by Poisson distribution is more accurate than geometric distribution in some real datasets, especially for small samples with sudden fluctuations.

7 Conclusion

In this paper, we propose two classes of INAR(1) models (RrINAR(1) and GRrINAR(1)) and illustrate that they perform better than existing RrNGINAR(1) in some cases. But there still is space for further study. First, the marginal distributions of $X_n(z_n)$ in RrINAR(1) may be generalized to some other discrete distributions, like the class of zero-modified distributions. Second, the prediction of GRrINAR(1) models is not accurate enough because it just knows that the fluctuation whether bigger or smaller, not the value of the data, using signed binomial thinning operator maybe more accurate.

Appendix

Proof of Theorem 1.

Case 1.1 When $z_{n-1} = i$ and $z_n = j$, where $i, j \in E_r$, we have $E(s^{X_n(j)}) = E(s^{\alpha\circ X_{n-1}(i)+\varepsilon_n(i,j)})$. The left side becomes $E(s^{X_n(j)}) = e^{(s-1)\mu_j}$. Consider the right side. Random variable $X_{n-1}(i)$ is independent of the random variable $\varepsilon_n(i, j)$, we have

$$\begin{aligned} E\big(s^{\alpha\circ X_{n-1}(i)+\varepsilon_n(i,j)}\big) &= E\big(s^{\alpha\circ X_{n-1}(i)}\big)E\big(s^{\varepsilon_n(i,j)}\big) = E\big(\big(Es^{U_1}\big)^{X_{n-1}(i)}\big)E\big(s^{\varepsilon_n(i,j)}\big) \\ &= E\big((s\alpha+1-\alpha)^{X_{n-1}(i)}\big)E\big(s^{\varepsilon_n(i,j)}\big) = e^{(s-1)\alpha\mu_i}E\big(s^{\varepsilon_n(i,j)}\big). \end{aligned}$$

The pgf of the random variable $\varepsilon_n(i, j)$ is given as

$$E\big(s^{\varepsilon_n(i,j)}\big) = e^{(s-1)(\mu_j-\alpha\mu_i)},$$

Table 7 *Transition matrix, parameter estimates, RMSE, AIC and BIC of the 24th DRUGS data*

model	matrix	CML	RMSE	AIC	BIC
P-R3INAR(1)	$\begin{pmatrix} 0.6923 & 0.2307 & 0.0769 \\ 0.0769 & 0.3846 & 0.5384 \\ 0.0253 & 0.2531 & 0.7215 \end{pmatrix}$	$\mu_1 = 24.5130$ $\mu_2 = 6.4364$ $\mu_3 = 1.1339$ $\alpha = 0.0197$	10.1571	513.3920	524.9232
G-R3INAR(1)		$\mu_1 = 23.5916$ $\mu_2 = 6.3504$ $\mu_3 = 1.1246$ $\alpha = 0.0288$	17.0269	581.4121	592.9433
R3NGINAR(1)		$\mu_1 = 23.4396$ $\mu_2 = 6.3293$ $\mu_3 = 1.1234$ $\alpha = 0.0343$	14.9080	581.3765	592.9077
P-GR3INAR(1)	$\begin{pmatrix} 0.9210 & 0.0614 & 0.0175 \\ 0.5555 & 0.2222 & 0.2222 \\ 0.5714 & 0 & 0.4285 \end{pmatrix}$	$\alpha_1 = 0.7015$ $\alpha_2 = 0.2127$ $\alpha_3 = 0.2031$ $\mu = 7.9750$	11.6583	1040.9420	1052.4730
G-GR3INAR(1)		$\alpha_1 = 0.2314$ $\alpha_2 = 0$ $\alpha_3 = 0.1990$ $\mu = 5.7085$	16.4519	708.5603	720.0915
P-R2INAR(1)	$\begin{pmatrix} 0.7142 & 0.2857 \\ 0.0427 & 0.9572 \end{pmatrix}$	$\mu_1 = 22.8077$ $\mu_2 = 2.7794$ $\alpha = 0.0684$	9.7724	721.6241	730.2725
G-R2INAR(1)		$\mu_1 = 22.3401$ $\mu_2 = 2.7588$ $\alpha = 0.0263$	18.5965	636.1490	644.7974
R2NGINAR(1)		$\mu_1 = 22.2584$ $\mu_2 = 2.7582$ $\alpha = 0.0277$	16.7854	636.1300	644.7784
P-GR2INAR(1)	$\begin{pmatrix} 0.9210 & 0.0789 \\ 0.5625 & 0.4375 \end{pmatrix}$	$\alpha_1 = 0.7013$ $\alpha_2 = 0.2066$ $\mu = 7.9692$	10.3963	1038.9560	1047.6040
G-GR2INAR(1)		$\alpha_1 = 0.2105$ $\alpha_2 = 0.1061$ $\mu = 5.3808$	13.7325	707.3969	716.0453

if $0 \le \alpha \le \frac{\mu_j}{\mu_i}$, the random variable $\varepsilon_n(i, j)$ has the distribution given in Case 1.1. Since i and j are arbitrary numbers from the set E_r, it follows that the random variables $\varepsilon_n(1, 1)$, $\varepsilon_n(1, 2)$, $\ldots$, $\varepsilon_n(r, r)$ will have well-defined distributions for $\alpha \in \bigcap_{k,l \in E_r} [0, \frac{\mu_l}{\mu_k}]$, that is, for $0 \le \alpha \le \min\{\frac{\mu_l}{\mu_k}; k, l \in E_r\}$.

Case 1.2 When $z_{n-1} = i$ and $z_n = j$, where $i, j \in E_r$, we have $E(s^{X_n(j)}) = E(s^{\alpha \circ X_{n-1}(i) + \varepsilon_n(i,j)})$. The left-hand side becomes $E(s^{X_n(j)}) = \frac{1}{1+\mu_j - \mu_j s}$. Consider the right-

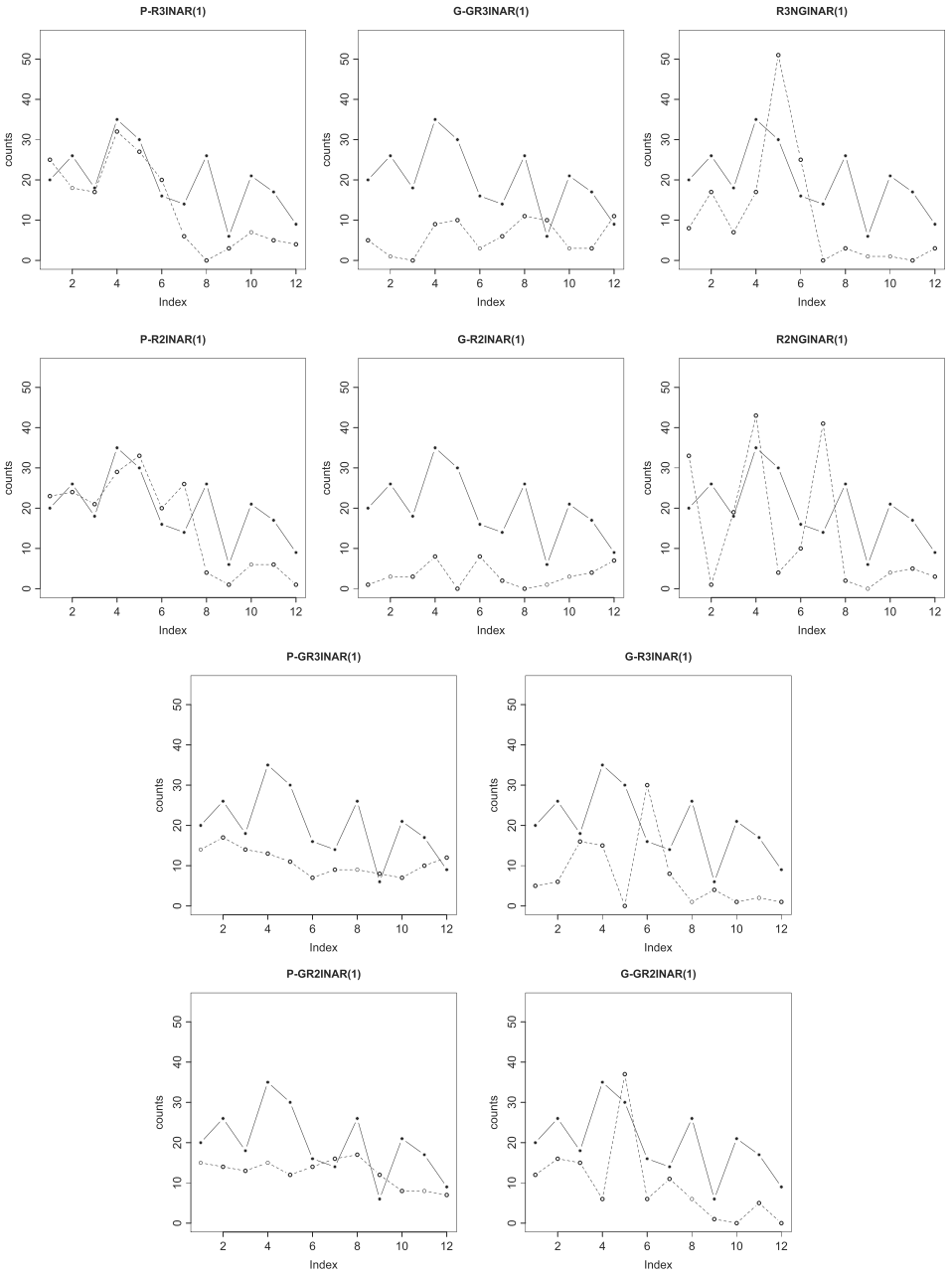

Figure 7 *The prediction of the last* 12 *observations based on ten models of the* 24*th DRUGS data.*

hand side. Random variable $X_{n-1}(i)$ is independent of random variable $\varepsilon_n(i, j)$, we have

$$\begin{aligned} &E\left(s^{\alpha\circ X_{n-1}(i)+\varepsilon_n(i,j)}\right) \\ &\quad = E\left(s^{\alpha\circ X_{n-1}(i)}\right)E\left(s^{\varepsilon_n(i,j)}\right) = E\left(\left(Es^{U_1}\right)^{X_{n-1}(i)}\right)E\left(s^{\varepsilon_n(i,j)}\right) \\ &\quad = E\left((s\alpha+1-\alpha)^{X_{n-1}(i)}\right)E\left(s^{\varepsilon_n(i,j)}\right) = \frac{1}{1+\mu_i-\mu_i(s\alpha+1-\alpha)}E\left(s^{\varepsilon_n(i,j)}\right). \end{aligned}$$

The pgf of the random variable $\varepsilon_n(i, j)$ is given as

$$E\left(s^{\varepsilon_n(i,j)}\right) = \frac{1+\alpha\mu_i-\alpha\mu_i s}{1+\mu_j-\mu_j s} = \left(1-\frac{\alpha\mu_i}{\mu_j}\right)\left(\frac{1}{1+\mu_j-\mu_j s}\right)+\frac{\alpha\mu_i}{\mu_j},$$

if $0 \le \alpha \le \frac{\mu_j}{\mu_i}$, the random variable $\varepsilon_n(i, j)$ has the distribution in Case 1.2. Since i and j are arbitrary numbers from the set E_r, it follows that the random variables $\varepsilon_n(1, 1)$, $\varepsilon_n(1, 2)$, $\ldots$, $\varepsilon_n(r, r)$ have well-defined distributions for $\alpha \in \bigcap_{k,l \in E_r}[0, \frac{\mu_l}{\mu_k}]$, that is, for $0 \le \alpha \le \min\{\frac{\mu_l}{\mu_k}; k, l \in E_r\}$. □

Proof of Corollaries 1 and 2. The results follow from the fact that $E(\varepsilon_n(i, j)) = \Phi'_\varepsilon(1)$ and $\mathrm{Var}(\varepsilon_n(i, j)) = \Phi''_\varepsilon(1) + \Phi'_\varepsilon(1)(1 - \Phi'_\varepsilon(1))$, where $\Phi_\varepsilon(s)$ is the pgf of the random variable $\varepsilon_n(i, j)$, one can derive the result through a simple calculation. □

Proof of Theorem 3. Let $\mu_{n+k|n} = E(X_{n+k}|X_n)$ and $\mu_{\varepsilon_n} = E(\varepsilon_n)$. According to Definition 1 and the independence of random variables X_{n+k-1} and ε_{n+k}, we obtain the conditional expectation of X_{n+k} on X_n satisfies the equation $\mu_{n+k|n} = \alpha\mu_{n+k-1|n} + \varepsilon_{n+k}$. Using the equation $k - 1$ times and the fact that $\mu_{n|n} = X_n$, we obtain

$$\mu_{n+k|n} = \alpha^k X_n + \sum_{l=0}^{k-1} \alpha^l \varepsilon_{n+k-l}.$$

Using the result of Corollary 1 for the expectations of the random variables ε_{n+k-l}, $l \in \{0, 1, \ldots, k - 1\}$, we obtain the expression for the conditional expectation.

Next consider the conditional variance. Let $\sigma^2_{n+k|n} = \mathrm{Var}(X_{n+k}(z_{n+k}) \mid X_n(z_n))$ and $\sigma^2_{\varepsilon_{n+k}} = \mathrm{Var}(\varepsilon_{n+k})$. Using the similar argument and properties of the binomial thinning operator, the conditional variance satisfies the following equation

$$\sigma^2_{n+k|n} = \alpha^2\sigma^2_{n+k-1|n} + \alpha(1 - \alpha)\mu_{n+k-1|n} + \sigma^2_{\varepsilon_{n+k}},$$

using the equation $k - 1$ times, we obtain

$$\sigma^2_{n+k|n} = \alpha^{2k}\sigma^2_{n|n} + \alpha(1 - \alpha)\sum_{l=0}^{k-1} \alpha^{2l}\mu_{n+k-1-l|n} + \sum_{l=0}^{k-1} \alpha^{2l}\sigma^2_{\varepsilon_{n+k-l}}.$$

Finally, using the fact $\sigma^2_{n|n} = 0$ and Corollary 1 for the variances of the random variables ε_{n+k-l}, $l \in \{0, 1, \ldots, k - 1\}$, which completes the proof. □

Proof of Theorem 4.

Case 2.1 When $z_{n-1} = i$, $i \in E_r$, we have $E(s^{X_n}) = E(s^{\alpha_i \circ X_{n-1} + \varepsilon_n(i)})$. The left side becomes $E(s^{X_n}) = e^{(s-1)\mu}$. Let us consider the right side, X_{n-1} is independent of $\varepsilon_n(i)$, we have

$$\begin{aligned} E\big(s^{\alpha_i \circ X_{n-1} + \varepsilon_n(i)}\big) &= E\big(s^{\alpha_i \circ X_{n-1}}\big)E\big(s^{\varepsilon_n(i)}\big) = E\big(\big(Es^{(U_i)_1}\big)^{X_{n-1}}\big)E\big(s^{\varepsilon_n(i)}\big) \\ &= E\big((s\alpha_i + 1 - \alpha_i)^{X_{n-1}}\big)E\big(s^{\varepsilon_n(i)}\big) = e^{(s-1)\alpha_i\mu}E\big(s^{\varepsilon_n(i)}\big). \end{aligned}$$

The pgf of the random variable $\varepsilon_n(i, j)$ is given as

$$E\big(s^{\varepsilon_n(i)}\big) = e^{(s-1)(\mu - \alpha_i\mu)}.$$

Since i is arbitrary number from set E_r, it follows that $\varepsilon_n(1)$, $\varepsilon_n(2)$, $\ldots$, $\varepsilon_n(r)$ are well defined, if $0 \le \alpha_i \le 1$, then $\varepsilon_n(i)$ has the distribution in Case 2.1.

Case 2.2 When $z_{n-1} = i$, where $i \in E_r$, we have $E(s^{X_n}) = E(s^{\alpha_i \circ X_{n-1} + \varepsilon_n(i)})$. The left-hand side becomes $E(s^{X_n}) = \frac{1}{1+\mu-\mu s}$. Let us consider the right-hand side, X_{n-1} is independent of $\varepsilon_n(i)$, we have

$$\begin{aligned} E\big(s^{\alpha_i \circ X_{n-1} + \varepsilon_n(i)}\big) &= E\big(s^{\alpha_i \circ X_{n-1}}\big)E\big(s^{\varepsilon_n(i)}\big) = E\big(\big(Es^{(U_i)_1}\big)^{X_{n-1}}\big)E\big(s^{\varepsilon_n(i)}\big) \\ &= E\big((s\alpha_i + 1 - \alpha_i)^{X_{n-1}}\big)E\big(s^{\varepsilon_n(i)}\big) \\ &= \frac{1}{1 + \mu - \mu(s\alpha_i + 1 - \alpha_i)}E\big(s^{\varepsilon_n(i)}\big). \end{aligned}$$

The pgf of the random variable $\varepsilon_n(i)$ is given as

$$E\left(s^{\varepsilon_n(i)}\right) = \frac{1+\alpha_i\mu-\alpha_i\mu s}{1+\mu-\mu s} = (1-\alpha_i)\left(\frac{1}{1+\mu-\mu s}\right)+\alpha_i.$$

Since i is arbitrary number from the set E_r, it follows that $\varepsilon_n(1)$, $\varepsilon_n(2)$, …, $\varepsilon_n(r)$ are well defined, if $0 \le \alpha_i \le 1$, then $\varepsilon_n(i)$ has the distribution in Case 2.2. □

Proof of Theorem 5. The proof is similar to the situation $k = 1$ in the second part of Theorem 3. □

Acknowledgments

We thank the Editor and the anonymous referee for their constructive comments and suggestions that have greatly improved the paper. This work is supported by National Natural Science Foundation of China (Nos. 11871027, 11731015), Science and Technology Developing Plan of Jilin Province (No. 20170101057JC), Science and Technology Program of Jilin Educational Department during the "13th Five-Year" Plan Period (No. 2016-399), and Cultivation Plan for Excellent Young Scholar Candidates of Jilin University. Zhu is the corresponding author.

References

Al-Osh, M. A. and Alzaid, A. A. (1987). First order integer-valued autoregressive (INAR(1)) processes. *Journal of Time Series Analysis* **8**, 261–275. MR0903755

Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics* **28**, 89–110. MR0084903

Barczy, M., Ispány, M. and Pap, G. (2011). Asymptotic behavior of unstable INAR(p) processes. *Stochastic Processes and Their Applications* **121**, 583–608. MR2763097

Bu, R., McCabe, B. P. M. and Hadri, K. (2008). Maximum likelihood estimation of higher-order integer-valued autoregressive processes. *Journal of Time Series Analysis* **29**, 973–994. MR2464949

Drost, F. C., van den Akker, R. and Werker, B. J. M. (2008). Local asymptotic normality and efficient estimation for INAR(p) models. *Journal of Time Series Analysis* **29**, 783–801. MR2450896

Drost, F. C., van den Akker, R. and Werker, B. J. M. (2009). Efficient estimation of auto-regression parameters and innovation distributions for semiparametric integer-valued AR(p) models. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **71**, 467–485. MR2649605

Jazi, M. A., Jones, G. and Lai, C. D. (2012). First-order integer valued AR processes with zero inflated Poisson innovations. *Journal of Time Series Analysis* **33**, 954–963. MR2991911

Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability* **33**, 664–677. MR1401464

Laketa, P. N., Nastić, A. S. and Ristić, M. M. (2018). Generalized random environment INAR models of higher order. *Mediterranean Journal of Mathematics* **15**. MR3740339 https://doi.org/10.1007/s00009-017-1054-z

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive processes. *Journal of Time Series Analysis* **4**, 439–455. MR1652193

Li, H., Yang, K., Zhao, S. and Wang, D. (2018). First-order random coefficients integer-valued threshold autoregressive processes. *AStA Advances in Statistical Analysis* **102**, 305–331. MR3829551

McCabe, B. P. M., Martin, G. M. and Harris, D. (2011). Efficient probabilistic forecasts for counts. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **73**, 253–272. MR2814495

McKenzie, E. (1986). Autoregressive moving-average processes with negative binomial and geometric distributions. *Advances in Applied Probability* **18**, 679–705. MR0857325

Nastić, A. S., Laketa, P. N. and Ristić, M. M. (2016). Random environment integer-valued autoregressive process. *Journal of Time Series Analysis* **37**, 267–287. MR3511585

Nastić, A. S., Laketa, P. N. and Ristić, M. M. (2018). Random environment INAR models of higher order. *REVSTAT Statistical Journal*. To appear.

Pedeli, X., Davison, A. C. and Fokianos, K. (2015). Likelihood estimation for the INAR(p) model by saddlepoint approximation. *Journal of the American Statistical Association* **110**, 1229–1238. MR3420697

Qi, X., Li, Q. and Zhu, F. (2019). Modeling time series of count with excess zeros and ones based on INAR(1) model with zero-and-one inflated Poisson innovations. *Journal of Computational and Applied Mathematics* **346**, 572–590. MR3864182 https://doi.org/10.1016/j.cam.2018.07.043

Steutel, F. W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability* **7**, 893–899. MR0542141

Tang, M. and Wang, Y. (2014). Asymptotic behavior of random coefficient INAR model under random environment defined by difference equation. *Advances in Difference Equations* **2014**, 99.

Weiß, C. H. (2018). *An Introduction to Discrete-Valued Time Series*. Chichester: John Wiley & Sons.

Zheng, H., Basawa, I. V. and Datta, S. (2006). Inference for pth-order random coefficient integer-valued autoregressive processes. *Journal of Time Series Analysis* **27**, 411–440. MR2328539

Z. Liu
F. Zhu
School of Mathematics
Jilin University
2699 Qianjin
Changchun 130012
China
E-mail: 914590404@qq.com
zfk8010@163.com

Q. Li
College of Mathematics
Changchun Normal University
Changchun 130032
China
E-mail: 46968158@qq.com