A combined geometric INAR(p) model based on negative binomial thinning

Aleksandar S. Nastić, Miroslav M. Ristić, Hassan S. Bakouch

University of Niš, Faculty of Sciences and Mathematics, Serbia
Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt

ARTICLE INFO

Article history:
Received 11 June 2011
Received in revised form 28 October 2011
Accepted 31 October 2011

Keywords:
INAR(p) models
Negative binomial thinning
Geometric marginal distribution
Estimation
Strictly stationary
Ergodicity

ABSTRACT

In this paper, we introduce a new (combined) integer-valued autoregressive model of order p with geometric marginal distributions, denoted by CGINAR(p), using the negative binomial thinning introduced by Ristić et al. Several properties of the model are constructed and discussed, including one-step ahead conditional statistical measures. Some methods for estimating the model parameters are considered and the asymptotic properties of the obtained estimators are derived. A real-life data example is investigated to assess the performance of the model.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

In many real-life situations there is a need to model and mathematically describe time series of correlated count observations. These situations could be found in reliability theory, meteorology, insurance theory, communications, medicine, law and social sciences. Such time series may be counts of accidents, detected errors, transmitted messages, patients, crime victimization and so on. When realizations of a process contain very large integers, then the process could be well approximated by continuous standard autoregressive processes, but unfortunately, this rarely happens in practice; so integer-valued models have to be used. In achieving this goal, the first significant results and efforts were made by introducing Markov chains, whose properties and structure can be referred to [1]. However, these models were generally overparametrized and have limitations in correlation structure. Then, Jacobs and Lewis [2–4] introduced the discrete autoregressive moving average (DARMA) models, which were based on the well-known ARMA models. After that, there were many other attempts, with various approaches, to fit and forecast these kinds of series. Some of the best results were obtained by using integer-valued autoregressive (INAR) models. Most of INAR models are based on binomial thinning operator and its generalizations, which was first defined by Steutel and Van Harn [5] and after that was widely implemented in the construction of integer-valued models.

The binomial thinning operator is generated by counting series of independent Bernoulli distributed random variables; therefore an INAR process based on such an operator is ideal for modeling the counting in the case where the observed population members can contribute the value 0 or 1 to the overall sum. Depending on the nature of the observed population, many of the discrete self-decomposable distributions can be viewed as marginal distributions of the integer valued time series models. At first, the Poisson distribution was generally considered, whereas its corresponding Poisson INAR processes
have been studied by Al-Osh and Alzaid [6], Du and Li [7], Freeland and McCabe [8] and others. However, other distributions were also taken into account, since the equality of mean and variance could not be confirmed by many real-life data. Therefore, Al-Osh and Aly [9] and McKenzie [10–12] have introduced and discussed in some detail the integer-valued time series with negative-binomial marginals, which has successfully covered some of the overdispersed cases. In addition, some more marginal distributions have been introduced, such as geometric [13], generalized Poisson [14] and zero truncated Poisson [15].

Further development and generalization have been made by introducing integer-valued higher-order models (INAR(p)), which were considered independently from each other by Alzaid and Al-Osh [16], on the one hand, their approach led to models that are analogous to the standard ARMA(p, p − 1) and Du and Li [7] on the other hand, their INAR(p) models had AR(p)-like autocorrelation structure. Recently, Zhang et al. [17] have introduced a pth-order integer-valued autoregressive processes with signed generalized power series thinning operator. Another generalization has been made in respect of thinning operator and can be found in [12,18,19]. Also, Zheng et al. [20,21] introduced the random coefficient integer-valued autoregressive processes by replacing the scalar coefficient of the thinning operator with a random variable, thereby providing non-parametric and parametric parameter estimators. Recently, Fokianos [22] has given a review on new progress on the analysis of a count time series.

Concerning the real-life over-dispersed time series of counts, where each population member can contribute any value of a non-negative integer to the overall sum, a new negative binomial thinning process with geometric marginals was introduced in [19]. However, in this paper we will generalize this process to order p. Also, we will rely on the known facts about the INAR(p) processes, but in order to overcome the difficulties of the complex structure of such processes introduced by Alzaid and Al-Osh [16] and Du and Li [7], we will use an alternative approach of Zhu and Joe [23] instead, that combined the EAR(p) model of Lawrance and Lewis [24] and the INAR(1) model. This new approach was later generalized and investigated in much more detail by Weiss [25,26], as a combined INAR(p) model for series of counts. The importance of the combined INAR(p) model comes from its ability to model the integer-valued autoregressive processes by discrete self-decomposable distributions (e.g. Poisson and negative binomial) for any order p than the well-known INAR(p) models. Weiss [25] modeled real-data by the Poisson CINAR(p) model based on binomial thinning. However, the Poisson distribution is not always suitable for modeling and analyzing the integer-valued time series because of its equi-dispersion. So, we introduce a new (combined) integer-valued autoregressive model of order p with geometric marginal distributions, denoted by CGINAR(p), using the negative binomial thinning introduced by Ristić et al. [19]. The CGINAR(p) model can model the over-dispersed time series data of counts and has memoryless property among the other integer-valued autoregressive models. Also, it can be used very well for modeling a series of counts in several real-life situations found in law, social studies and crime investigation.

The rest of the paper is arranged as follows. In Section 2, we introduce a new (combined) integer-valued autoregressive model of order p with geometric marginal distributions, denoted by CGINAR(p), using the negative binomial thinning introduced by Ristić et al. [19]. Also, several properties of the CGINAR(p) model are constructed and discussed, such as the autocorrelations and one-step ahead conditional expectation and variance. Section 3 deals with the estimation problem of the unknown parameters of the model by using conditional least squares estimation and Yule–Walker estimation. Moreover, the asymptotic properties and asymptotic distributions of the conditional least squares estimators and Yule–Walker estimators are investigated. In Section 4, we discuss a real-life data example to assess the performance of the CGINAR(p) model.

2. Construction and properties of the model

In this section we construct a new CINAR(p) model with geometric marginals denoted by CGINAR(p) via the negative binomial thinning operator $\ast_n$ defined as

\[ \alpha \ast_n X = \sum_{j=1}^{X} \theta_j^{(n)}, \]

where \( \{\theta_j^{(n)}\} \) is a sequence of i.i.d. geometric(\( \alpha/(1+\alpha) \)) random variables, i.e. a sequence of i.i.d. random variables with probability mass function \( P(\theta_j^{(n)} = x) = \alpha^x/(1+\alpha)^{x+1}, x = 0, 1, \ldots \), that is independent of a stochastic process \( \{X_n\} \), and is known as a counting series. Also, note that the random variables of the counting series \( \{\theta_j^{(n)}\} \) are counted at time \( n \), so the time index \( n \) below the negative binomial thinning operator $\ast_n$ means that the thinning is performed at time \( n \). Moreover, properties of the operator $\ast_n$ are the same as properties of the negative binomial thinning operator introduced by Ristić et al. [19]. We first start with the definition of a CINAR(p) model generated by the negative binomial thinning operator.

**Definition 1.** A time series \( \{X_n\} \) given by

\[ X_n = \begin{cases} \alpha \ast_n X_{n-1} + \varepsilon_n, & \text{w.p. } \phi_1 \\ \alpha \ast_n X_{n-2} + \varepsilon_n, & \text{w.p. } \phi_2 \\ \vdots \\ \vdots \\ \alpha \ast_n X_{n-p} + \varepsilon_n, & \text{w.p. } \phi_p \end{cases} \]
Theorem 1. The random variable \( \varepsilon_n \) is a mixture of two geometrically distributed random variables

\[
\varepsilon_n \sim \begin{cases} 
\text{Geom}(\mu/(1+\mu)), & \text{w.p. } 1-\alpha \mu/(\mu-\alpha), \\
\text{Geom}(\alpha/(1+\alpha)), & \text{w.p. } \alpha \mu/(\mu-\alpha).
\end{cases}
\]

Proof. The distribution of the random variable \( \varepsilon_n \) can be determined by following Ristić et al. [19], Weiß [25] and Definition 1. \( \square \)

Remark 1. If \( \alpha = 0 \), then \( X_n = \varepsilon_n \), for \( n \geq 1 \), and the CGINAR\((p)\) process \( \{X_n\} \) is a sequence of i.i.d. geometric\((\mu/(1+\mu))\) random variables.

Remark 2. If \( p = 1 \), then the process \( \{X_n\} \) given by Definition 1 is reduced to the NGINAR\((1)\) process introduced by Ristić et al. [19].

Now, let us consider some properties of the introduced CGINAR\((p)\) model, such as one-step ahead conditional expectation, variance and probability generating function that have a major role in forecasting. We start first with the autocorrelation function. The properties of the autocorrelation function are given in the following theorem.

Theorem 2. The autocorrelation function satisfies the equation

\[
\rho(k) = \alpha \sum_{j=1}^{p} \phi_j \rho(|k-j|)
\]

and \( \rho(n) \) is decreasing exponentially to zero as \( n \) tends to infinity.

Proof. Following the proof of Theorem 2 [25] and Definition 1, we can show that the autocorrelation function satisfies Eq. (1) since we have assumed that the condition (vi) is satisfied. Starting from the previous formula for \( k \in \{1, 2, \ldots, p\} \) and the fact that the inequality \( \rho(l) \leq 1 \) is valid for each \( l \geq 0 \), by direct calculation we get that \( \rho(k) \leq \alpha \) for \( k \in \{1, 2, \ldots, p\} \). Then, in the same way, we get \( \rho(k) \leq \alpha^2 \) for \( k \in \{p+1, p+2, \ldots, 2p\} \). Finally, repeating this procedure \( n \) times, \( \rho(k) \leq \alpha^n \) for \( k \in \{(n-1)p+1, (n-1)p+2, \ldots, np\} \) is obtained, and from this \( \rho(np) \leq \alpha^n \) is valid, and it follows that \( \rho(n) \) is decreasing exponentially to zero as \( n \) tends to infinity. \( \square \)

From Definition 1, it is obvious that the CGINAR\((p)\) model is a \( p \)th order Markov process. Therefore, in order to obtain the joint probability distribution, it is sufficient to determine transition probabilities, which can be directly obtained by Definition 1:

\[
P(X_n = x_n|H_{n-1}) = \sum_{j \in l} \phi_j P \left( \sum_{i=1}^{X_{n-j}} W_i + \varepsilon_n = x_n|H_{n-1} \right) + \sum_{j \in l} \phi_j P (\varepsilon_n = x_n) ,
\]

where \( l = \{j|X_{n-j} = 0, j = 1, \ldots, p\} \) and \( P (\varepsilon_n = x_n) \) are probabilities of the process \( \{\varepsilon_n\} \). Furthermore, it is not difficult to calculate

\[
P \left( \sum_{i=1}^{X_{n-j}} W_i + \varepsilon_n = x_n|H_{n-1} \right) = \sum_{k=0}^{x_n} \binom{x_n-j+k-1}{x_{n-j}-1} \left( \frac{\alpha}{1+\alpha} \right)^k \left( \frac{1}{1+\alpha} \right)^{x_{n-j}} \times \left( 1 - \frac{\alpha \mu}{\mu-\alpha} \right) \mu^{x_{n-k}} + \frac{\alpha \mu}{\mu-\alpha} \frac{\alpha^{x_{n-k}}}{(1+\alpha)^{x_{n-k+1}}} .
\]

Next, we consider the strict stationarity and ergodicity of the CGINAR\((p)\) model. We have the following theorem.
Theorem 3. The CGINAR(p) model is strictly stationary and ergodic.

Proof. It is easy to show that the process is strictly stationary. Now, let us discuss the ergodic property of the process. Namely, following Proposition 2 of Shiryaev [27, p. 441], a strictly stationary process is ergodic if the probability of an event which is invariant in respect of this process is 0 or 1. Let A be an event, arbitrary and invariant in respect of \( \{X_n\} \). Then, there is a set \( B \), such that for each \( n \in \mathbb{Z}, A = \{o| (X_n, X_{n-1}, \ldots) \in B\} \). It follows that \( A \isin \mathcal{F} (X_n, X_{n-1}, \ldots) \), for each \( n \in \mathbb{Z} \), where \( \mathcal{F} (X_n, X_{n-1}, \ldots) \) is \( \sigma \)-algebra generated by \( \{X_n, X_{n-1}, \ldots\} \). From Definition 1 we have that \( \mathcal{F} \ (X_n, X_{n-1}, \ldots) \subseteq \mathcal{F} (\varepsilon_n, \varphi^{(n)}, \varepsilon_{n-1}, \varphi^{(n-1)}, \ldots) \), where \( \varphi^{(n)} \) is the corresponding sequence of i.i.d. geometric random variables \( \{W_j^{(n)}\} \), known as the counting series of the thinning operator. Therefore, \( A \isin \mathcal{F} (\varepsilon_n, \varphi^{(n)}, \varepsilon_{n-1}, \varphi^{(n-1)}, \ldots) \), for each \( n \in \mathbb{Z} \), and consequently

\[
A \isin \bigcap_{n=0}^{\infty} \mathcal{F} (\varepsilon_n, \varphi^{(n)}, \varepsilon_{n-1}, \varphi^{(n-1)}, \ldots).
\]

Note that the right side of (2) is the intersection of an infinite sequence of \( \sigma \)-algebras, generated by independent random series \( \{\varepsilon_n, \varphi^{(n)}\} \), representing in this way, a tail \( \sigma \)-algebra. Therefore, following Theorem 1 (Kolmogorov’s 0-1 law) in [27, p. 441], \( A \) is a tail event, and \( P(A) = 0 \) or \( 1 \). This finalizes the proof of ergodicity of the CGINAR(p) process. \( \square \)

By direct calculation, based on properties of conditional expectation, it is not difficult to obtain the one-step ahead conditional probability generating function (cpgf) of the CGINAR(p) process. Considering the definition of a negative binomial thinning operator and Definition 1, it is obtained that

\[
E \left( s^{X_{n+1}} \mid H_n \right) = \left( \frac{1 + \alpha (1 + \mu) - \alpha (1 + \mu)s}{1 + \alpha - \alpha s} \right) X_{n+1-i}.
\]

Applying the well known pgf properties, the immediate consequences of the previous results are the one-step ahead conditional expectation and variance of the CGINAR(p) process, which can be expressed, respectively, as

\[
E(X_{n+1} \mid H_n) = \alpha \sum_{i=1}^{p} \phi_i X_{n-i+1} + (1 - \alpha) \mu
\]

and

\[
\text{Var}(X_{n+1} \mid H_n) = \text{Var}(\varepsilon_n) + \alpha \sum_{i=1}^{p} \phi_i X_{n-i+1} \left( 1 + \alpha + \alpha X_{n-i+1} - \alpha \sum_{j=1}^{p} \phi_j X_{n-j+1} \right).
\]

3. Estimation of the unknown parameters

In this section we will deal with the problem of estimating the unknown parameters \( \alpha, \phi_1, \phi_2, \ldots, \phi_p \) and \( \mu \) of the CGINAR(p) model with negative binomial thinning. Some estimators will be derived, such as Yule–Walker and conditional least squares estimators. All these estimators are based on a realization \( (X_1, X_2, \ldots, X_N) \) of the CGINAR(p) process.

3.1. Conditional least squares estimators

As is well known, these kinds of estimators are obtained as the statistics which minimize the following sum of squares:

\[
Q_n(\theta) = \sum_{n=p+1}^{N} \left( X_n - \theta_1 X_{n-1} - \cdots - \theta_p X_{n-p} - \left( 1 - \sum_{i=1}^{p} \theta_i \right) \mu \right)^2,
\]

where \( \theta_i = \alpha \phi_i \), for \( i \in \{1, \ldots, p\} \) and \( \theta = (\theta_1, \theta_2, \ldots, \theta_p, \mu) \). Solving the corresponding linear system, obtained by equating partial derivatives of \( Q_n(\theta) \) to zero, it is easy to get that \( \mu = \frac{1}{\left( 1 - \sum_{i=1}^{p} \theta_i \right)} \left( X^{(0)} - \sum_{i=1}^{p} \theta_i X^{(i)} \right) \), where \( X^{(i)} = \frac{1}{n-p} \sum_{n=p+1}^{N} X_{n-j} \) for \( j = 0, 1, \ldots, p \). Then, substituting \( \mu \) using the previous result, the corresponding system of linear equations is reduced to

\[
\sum_{j=1}^{p} \theta_j \hat{\rho}^*(l-j) = \hat{\rho}^*(l), \quad l = 1, 2, \ldots, p,
\]

where \( \hat{\rho}^*(l-j) = \frac{1}{n-p} \sum_{n=p+1}^{N} X_{n-j} X_n - X^{(0)} X^{(i)} \), or equivalently, introducing \( \hat{\rho}^*(k) = \frac{\hat{\rho}^*(k)}{\hat{\rho}^*(0)} \), it can be rewritten as

\[
\sum_{j=1}^{p} \theta_j \hat{\rho}^*(l-j) = \hat{\rho}^*(l), \quad l = 1, 2, \ldots, p.
\]
Solving the preceding system of equations, estimators of $\hat{\theta}_j$, $j = 1, 2, \ldots, p$ and $\mu$ can be obtained in the following form

$$\hat{\mu}^\text{cls} = \frac{D_j^*}{D^*}, \quad j = 1, 2, \ldots, p$$

and

$$\hat{\mu}^\text{cls} = \left(\frac{1}{1 - \sum_{i=1}^p \frac{D_i^*}{D^*}} - \frac{1}{N - p} \sum_{n=p+1}^N X_n - \sum_{j=1}^p \frac{D_j^*}{D^*} \sum_{n=p+1}^N X_n \right),$$

where $D^*$ and $D_i^*$, $i = 1, 2, \ldots, p$ are well known determinants of Cramer’s rule for the matrix of system (3). Since $\sum_{i=1}^p \theta_i = \alpha$ and $\phi_i = \frac{\alpha}{\alpha}$, using the above results, the estimators of parameter $\mu$, $\alpha$ and $\phi_j$, $j = 1, 2, \ldots, p$ are calculated, respectively, as

$$\hat{\mu}^\text{cls} = \left(\frac{D_j^*}{D^*} \sum_{i=1}^p \frac{D_i^*}{D^*} (N - p) \sum_{n=p+1}^N X_n - \sum_{j=1}^p \frac{D_j^*}{D^*} \sum_{n=p+1}^N X_n \right),$$

$$\hat{\alpha}^\text{cls} = \frac{p}{\sum_{i=1}^p D_i^*} \quad \text{and} \quad \hat{\phi}_j^\text{cls} = \frac{D_j^*}{\sum_{i=1}^p D_i^*}, \quad j = 1, 2, \ldots, p.$$

In order to describe the properties of the conditional least squares estimators $\left(\hat{\theta}_1^\text{cls}, \hat{\theta}_2^\text{cls}, \ldots, \hat{\theta}_p^\text{cls}, \hat{\mu}^\text{cls}\right)$ we will take the estimators $\left(\hat{\theta}_1^\text{cls}, \hat{\theta}_2^\text{cls}, \ldots, \hat{\theta}_p^\text{cls}, \hat{\mu}^\text{cls}\right)$ as a base for getting further analysis, because they are more suitable for the application of relevant results from [28].

**Lemma 4.** The estimator $\left(\hat{\theta}_1^\text{cls}, \hat{\theta}_2^\text{cls}, \ldots, \hat{\theta}_p^\text{cls}, \hat{\mu}^\text{cls}\right)$ is strongly consistent and has an asymptotical normal distribution with “mean” $\theta$ and “variance” $N^{-1}U^{-1}RU^{-1}$, where $N$ is the sample size and $U$ is a matrix defined as in Theorem 3.2 [28].

**Proof.** First, it is not difficult to prove that the statistics $\left(\hat{\theta}_1^\text{cls}, \hat{\theta}_2^\text{cls}, \ldots, \hat{\theta}_p^\text{cls}, \hat{\mu}^\text{cls}\right)$ meet all the conditions of Theorem 3.1 in [28]. Namely, conditions C1 and C3 are trivially verified, but the condition C2 still requires some more effort. Basically, it boils down to proving the implication $\text{Var}(\sum_{i=1}^p a_i X_{n-i}) = 0 \Rightarrow a_1 = a_2 = \cdots = a_p = 0$, for arbitrary real numbers $a_i$, $i = 1, 2, \ldots, p$, which is a direct consequence of the regularity of autocovariance matrix $\Gamma$ of random vector $(X_{n-1}, X_{n-2}, \ldots, X_{n-p})$. The regularity property of this matrix can be easily confirmed by applying Proposition 5.1.1 [29], given that all its conditions are satisfied. Therefore, a strong consistency is established for these estimators.

On the asymptotic normality of the conditional least squares estimators, note that the estimators $\hat{\theta}_j^\text{cls}$, $i = 1, 2, \ldots, p$ and $\hat{\mu}^\text{cls}$ satisfy the requirements of Theorem 3.2 [28], because the CGINAR(p) process is a pth order Markovian. The conditions of Theorem 3.1 [28] are already met, and it is not hard to prove that all the elements of the corresponding matrix $R$, defined in Theorem 3.2 [28], are finite. Hence, $\hat{\theta}_j^\text{cls} = \left(\hat{\theta}_1^\text{cls}, \ldots, \hat{\theta}_p^\text{cls}, \hat{\mu}^\text{cls}\right)$ has an asymptotical normal distribution with “mean” $\theta$ and “variance” $N^{-1}U^{-1}RU^{-1}$, where $N$ is sample size and $U$ is a matrix defined as in Theorem 3.2 [28], for which $U^{-1}$ is well defined, since

$$U = \begin{bmatrix} \Gamma & 0 \\ 0 & \left(1 - \sum_{i=1}^p \theta_i\right)^{-1} \end{bmatrix}.$$
3.2. Yule–Walker estimators

Now, let us consider the estimators of the unknown parameters obtained by the method of moments. Since $E(X) = \mu$ and $\rho(k) = \alpha \sum_{j=0}^{p} \phi_j \rho(|k-j|)$, $k = 1, 2, \ldots, p$, introducing $\theta_j = \alpha \phi_j$, $j = 1, 2, \ldots, p$ and using the fact that $\sum_{j=1}^{p} \theta_j = \alpha$, the Yule–Walker estimators of $\mu$, $\alpha$ and $\phi_1, \phi_2, \ldots, \phi_p$ are obtained from the preceding equations and they are

$$\hat{\mu}^{yw} = \frac{1}{N} \sum_{i=1}^{N} X_n, \quad \hat{\alpha}^{yw} = \frac{\sum_{i=1}^{p} D_i}{D}, \quad \hat{\phi}_j^{yw} = \frac{D_j}{\sum_{i=1}^{p} D_i}, \quad j = 1, 2, \ldots, p,$$

respectively, where $D_1, D_2, \ldots, D_p$ and $D$ are corresponding determinants of Cramer’s rule, implemented in solving the linear system, defined by the relations of autocorrelation functions.

Now, let us examine the character of consistency and asymptotic distribution of these estimators.

**Lemma 6.** The conditional least squares estimators and the Yule–Walker estimators are asymptotically equivalent.

**Proof.** For this purpose it is enough to prove the validity of the following conditions:

$$N^{\frac{1}{2}} (\hat{\alpha}^{cls} - \hat{\alpha}^{yw}) = o(1), \quad N \to \infty,$$

$$N^{\frac{1}{2}} (\hat{\mu}^{cls} - \hat{\mu}^{yw}) = o(1), \quad N \to \infty,$$

$$N^{\frac{1}{2}} (\hat{\phi}_j^{cls} - \hat{\phi}_j^{yw}) = o(1), \quad N \to \infty, \quad j = 1, 2, \ldots, p.$$

So, we briefly prove one of these conditions, namely the first one

$$N^{\frac{1}{2}} (\hat{\alpha}^{cls} - \hat{\alpha}^{yw}) = \frac{N^{\frac{1}{2}} \sum_{i=1}^{p} D_i^*}{D^*} - \frac{N^{\frac{1}{2}} \sum_{i=1}^{p} D_i}{D},$$

where $D, D^*, D_i, D_i^*$, for $i = 1, 2, \ldots, p$, are the determinants introduced earlier. Since the right side of the previous equation may be written as the difference

$$f \left( \hat{\gamma}^*(0), \ldots, \hat{\gamma}^*(p), N^{\frac{1}{2}} \hat{\gamma}^*(0), \ldots, N^{\frac{1}{2}} \hat{\gamma}^*(p) \right) - f \left( \hat{\gamma}(0), \ldots, \hat{\gamma}(p), N^{\frac{1}{2}} \hat{\gamma}(0), \ldots, N^{\frac{1}{2}} \hat{\gamma}(p) \right),$$

where $f$ is obviously a continuous function, it is enough to prove that

$$N^{\frac{1}{2}} \hat{\gamma}^*(k) - N^{\frac{1}{2}} \hat{\gamma}(k) = o(1),$$

as $N$ tends to $\infty$, for $k = 0, 1, 2, \ldots, p$. Using the representations

$$\hat{\gamma}^*(k) = \frac{1}{N-p} \sum_{n=p+1}^{N} \left( X_n X_{n-k} - \bar{X}^{(0)} \bar{X}^{(k)} \right),$$

$$\hat{\gamma}(k) = \frac{1}{N} \sum_{n=k+1}^{N} \left( X_n - \bar{X} \right) \left( X_{n-k} - \bar{X} \right),$$

where $\bar{X} = \frac{1}{N} \sum_{n=1}^{N} X_n$, the left side of (5) equals $o(1)$, w.p. 1, as $N \to \infty$. The other two conditions of (4) can also be similarly proved. $\square$

Finally, we obtain the asymptotic properties of the Yule–Walker estimators.

**Theorem 7.** The Yule–Walker estimator $\left( \hat{\alpha}^{yw}, \hat{\phi}_1^{yw}, \hat{\phi}_2^{yw}, \ldots, \hat{\phi}_p^{yw}, \hat{\mu}^{yw} \right)$ is strongly consistent and has an asymptotical normal distribution with “mean” and “variance” as the corresponding conditional least squares estimator $\left( \hat{\alpha}^{cls}, \hat{\phi}_1^{cls}, \ldots, \hat{\phi}_p^{cls}, \hat{\mu}^{cls} \right)$.

**Proof.** Since the conditions (4) are checked and considering that the strong consistency of conditional least squares estimators is already proved, as a direct consequence, the strong consistency of Yule–Walker estimators is confirmed. On the other hand, those conditions are sufficient for applying the Proposition 6.3.3 [29]. Hence, the asymptotic normality of the Yule–Walker estimators is confirmed, even more, the asymptotic normal distribution of the estimator $\left( \hat{\alpha}^{yw}, \hat{\phi}_1^{yw}, \hat{\phi}_2^{yw}, \ldots, \hat{\phi}_p^{yw}, \hat{\mu}^{yw} \right)$ has the same “mean” and “variance” as the corresponding conditional least squares estimator $\left( \hat{\alpha}^{cls}, \hat{\phi}_1^{cls}, \ldots, \hat{\phi}_p^{cls}, \hat{\mu}^{cls} \right)$ which is analyzed in the previous subsection. $\square$
4. Real data example

In order to compare the introduced CGINAR\((p)\) model, a time series of real count data is observed. The data are obtained from the Crime data section of the Forecasting Principles site (http://www.forecastingprinciples.com). This data series represents the counting of receiving stolen property, reported in the 32th police car beat in Pittsburgh. There are 144 observations, starting in January 1990 and ending in December 2001. The sample mean, variance and autocorrelation are 0.8958, 1.4626 and 0.278, respectively. The over-dispersion is evident. The plots of the time series, its autocorrelation and partial autocorrelation functions are given in Fig. 1. Analyzing the diagrams we conclude that the 2nd order autoregression models are appropriate for the given data series. Therefore, we compared our CGINAR\((2)\) to the Combined PoINAR\((2)\) model of Weiβ [25]. Namely, this PoINAR\((2)\) is also, like our model, a probabilistic mixture of lag 1 and lag 2 thinning operators, being different, in using the binomial thinning operator. Maximum likelihood parameter estimates and their standard errors are calculated. The values of information criteria AIC and BIC and the root mean squares of differences of observed and predicted values RMS are obtained, too. The results are given in Table 1. As can be seen, AIC, BIC and RMS are smaller for the Combined INAR\((2)\) model with geometric marginals generated by the negative binomial thinning operator. Besides the over-dispersion, a very important reason for the better performance of the CGINAR\((p)\) model in the case of these data series may be found in the nature of the observed population. Namely, as was mentioned above, the numbers of the series refers to the counting of receiving the stolen property. But when some property is received, after a while it could be resold and delivered to one or more different buyers, which generates one or more new received properties. In this way a buyer who becomes
Table 1
Parameter estimates, AIC, BIC and RMS for stolen property data. The standard errors are given in brackets.

<table>
<thead>
<tr>
<th>Model</th>
<th>ML estimates</th>
<th>AIC</th>
<th>BIC</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoINAR(2)</td>
<td>$\lambda = 0.5523 (0.0582)$</td>
<td>363.1</td>
<td>372.0</td>
<td>1.0657</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.3285 (0.0742)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NGINAR(2)</td>
<td>$\mu = 0.8007 (0.1571)$</td>
<td>352.3</td>
<td>361.2</td>
<td>1.0597</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.4171 (0.1157)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\phi_1 = 0.4427 (0.1875)$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

the seller, which represents one observed element of the population, might produce more than one random event which are the object of our counting. Also, often it is not easy for the seller to sell goods to multiple buyers, i.e. the probability of achieving this, decreases geometrically with the number of buyers. Therefore, since one population member may contribute to the overall counting result by zero, one or more with the decreasing probability, the modeling of such a counting process is performed much more naturally by applying a geometric counting series than by using the Bernoulli distribution. So in this case the negative binomial thinning operator is more logical and a natural choice than the binomial thinning.

Acknowledgments

The authors are very grateful to the referees for their valuable suggestions and comments which have greatly improved this paper. The first two authors acknowledge the grant MNTR 174013 for carrying out this research.

References