FACTA UNIVERSITATIS (NIŠ)
Ser. Math. Inform. Vol. 29, No 1 (2014), 25–42

ON SUITABILITY OF NEGATIVE BINOMIAL MARGINALS AND GEOMETRIC COUNTING SEQUENCE IN SOME APPLICATIONS OF COMBINED INAR(p) MODEL *

Aleksandar S. Nastić

Abstract. A combined negative binomial integer-valued autoregressive process of order p is defined. Correlation structure and regression properties are presented. Model parameters are estimated using conditional least squares and Yule-Walker methods and the asymptotic distributions of the obtained estimators are derived. Model interpretation is provided, especially focusing on usage of geometric counting sequence and negative binomial marginals and further it is justified by application of the introduced model to certain counting data, where it is compared with some other possible known model solutions.

1. Introduction

In recent years there has been an exponential growth of interest and research as well, in the area of the discrete valued time series modeling. It has all begun by Cox and Miller [8] using Markov chains. Later, some significant results were obtained by Jacobs and Lewis [9, 10, 11], designing the discrete ARMA models. Finally, real foundation of a contemporary discrete valued time series analysis was made by defining the integer-valued autoregressive (INAR) models, which were introduced, independently of each other by McKenzie [13] and Al-Osh and Alzaid [2]. They used a binomial thinning operator based on the Bernoulli counting series in order to define the dependance among the random variables of non-negative integer-valued time series. Mainly defined by Poisson or geometric marginal distribution, these models were highly adequate for modeling counting data concerning the number of random events or some population elements which could enter into or either survive or disappear from the observed system during counting time intervals. However, responding over time to more demanding modeling requirements, there have been many modifications and generalizations of the introduced INAR models. Some of them were in respect of marginal distribution and they can be found in

Received March 06, 2014.; Accepted March 31, 2014.
2010 *Mathematics Subject Classification.* Primary 62M10; Secondary 62M15
*The author acknowledges the grant of MNTR 174013 for carrying out this research.

[14], [1], [5] and [6]. Other authors have focused their attention on the thinning operator, such as [3], [12] and [22, 23]. Also, the main contribution in [17] was an introduction of the new integer-valued autoregressive process based on the negative binomial thinning operator.

Here we construct the INAR process in order to model a certain kind of crime data in the best possible way. Our interest is focused on the counting of committing light criminal activities periodically in time. In this case the observed population elements may interact among themselves, producing in this way newly generated counting objects, which probability of occurrence decay over time. Therefore, geometric distribution seems to be a promising choice for the random variables of counting sequence. Further, we noted that these kind of time series contain only few zeroes and smaller non negative integers as well, on the other hand they are mostly made of slightly larger two-digit integers. Based on the intuitive and empirical distribution interpretation, this has made us to consider negative binomial marginals. Besides a few new process characterizations, this is the main step forward in relation to the combined geometric INAR(p) model introduced in [15], where geometric marginal distribution was taken into account.

The outline of the paper is as follows. In the next section we introduce the combined negative binomial integer-valued autoregressive process of the order p and we present its main features, including conditional stochastic properties. In Section 3, Yule-Walker and conditional least squares methods are used for model parameter estimation. Also, we discuss asymptotic behavior and distributional properties of the corresponding statistics. In the last section we elaborate the main contribution of the paper. Namely, we give a detailed model interpretation, reflecting its key features, especially the choice of marginal distribution and counting sequence, on the data characteristics. Also, the compatibility of the introduced model with the observed counting series is confirmed by its comparison with some other possible INAR model solutions.

2. Construction of the model

In this section we introduce a combined integer-valued autoregressive process with negative binomial marginal distribution based on the negative binomial thinning operator "$*_n$", defined in [15] as,

$$\alpha *_n X = \sum_{i=1}^{X} W_i^{(n)},$$

where $\left\{W_i^{(n)}\right\}$ is a sequence of independent random variables, independent of a non-negative integer-valued random variable X, geometrically distributed with parameter $\alpha/(1+\alpha)$, $\alpha \in [0,1]$, i.e. with probability mass function (pmf) given as $P\left(W_i^{(n)} = x\right) = \alpha^x/(1+\alpha)^{x+1}$, $x = 0, 1, \ldots$. Index n is used as a time notation, meaning that the thinning "$*_n$" is realized at time point n.

Definition 2.1. A time series $\{X_n\}$, which is given by

$$X_n = \begin{cases} \alpha *_n X_{n-1} + \varepsilon_n, & \text{with probability } \phi_1, \\ \alpha *_n X_{n-2} + \varepsilon_n, & \text{with probability } \phi_2, \\ \vdots & \\ \alpha *_n X_{n-p} + \varepsilon_n, & \text{with probability } \phi_p, \end{cases} \tag{2.1}$$

where $\alpha \in (0,1)$, $0 \leqslant \phi_1, \phi_2, \dots, \phi_p \leqslant 1$, $\phi_1 + \phi_2 + \dots + \phi_p = 1$, $p \geqslant 1$, and $X_n : \mathcal{NB}\left(\theta, \frac{q}{1+q}\right)$ has a negative binomial probability mass function $P(X_n = x) = \binom{\theta-1+x}{\theta-1}\frac{q^x}{(1+q)^{x+\theta}}$, $\theta > 0$, $q > 0$ is called a combined negative binomial integer-valued autoregressive process of order p (CNBINAR(p)), if the following conditions are satisfied:

(i) $\{\varepsilon_n\}$ is an i.i.d. sequence of random variables, where ε_n is independent of X_m, for any $m < n$,

(ii) counting sequences $\left\{W_i^{(m)}\right\}$ are mutually independent, and not correlated to ε_n, for any m and n,

(iii) conditioned on X_n, random variables $\alpha *_{n+1} X_n, \dots, \alpha *_{n+p} X_n$ are independent,

(iv) random variables $\alpha *_{n+1} X_n$, $\alpha *_{n+2} X_n, \dots$, $\alpha *_{n+p} X_n$ do not depend on H_{n-1}, which represent the process history generated by all random variables X_m, $\alpha *_{m+j} X_m$, $m < n$, $j \in \{1, 2, \dots, p\}$.

Remark 2.1. If $p = 1$, then the process $\{X_n\}$ introduced in Definition 2.1 is reduced to the NBINAR(1) defined in [18].

In order to discuss the distributional properties of the innovation sequence $\{\varepsilon_n\}$, we focus on its probability generating function (pgf). Since,

$$\Phi_{X_n}(s) = \sum_{i=1}^{p} \phi_i \Phi_{\alpha * X_{n-i}+\varepsilon_n}(s) = \Phi_X(\Phi_W(s))\Phi_\varepsilon(s),$$

then this problem is reduced to the case of $p = 1$. Thus, as in [18], it is resolved obtaining that the pgf of ε_n is

$$\Phi_\varepsilon(s) = \left(\frac{1}{1+\alpha-\alpha s}\right)^\theta \left(\frac{1+\alpha(1+q)-\alpha(1+q)s}{1+q-qs}\right)^\theta.$$

Using this, it follows that $\varepsilon_n = Y_n + Z_n$, where Y_n is $\mathcal{NB}\left(\theta, \frac{\alpha}{1+\alpha}\right)$ distributed and $Z_n = \sum_{i=1}^{N}\left(\frac{\alpha(1+q)}{q}\right)^{R_i} \circ V_i$, where $N \stackrel{d}{=} \mathcal{P}\left(-\theta \log \frac{\alpha(1+q)}{q}\right)$, $R_i \stackrel{d}{=} \mathcal{U}(0,1)$, $V_i \stackrel{d}{=} Geom\left(\frac{q}{1+q}\right)$, $\circ$

is binomial thinning and N, R_i, V_i are independent random variables. Therefore, the pmf of ε_n is obtained as

$$\begin{aligned} P(\varepsilon_n = 0) &= \left(\frac{1}{1+\alpha}\right)^{\theta} \left(\frac{1+\alpha(1+q)}{1+q}\right)^{\theta}, \\ P(\varepsilon_n = l) &= \frac{\theta}{l} \sum_{j=0}^{l-1} P(\varepsilon_n = j) \left\{ \left(\frac{\alpha}{1+\alpha}\right)^{l-j} \right. \\ &\quad \left. - \left(\frac{\alpha(1+q)}{1+\alpha(1+q)}\right)^{l-j} + \left(\frac{q}{1+q}\right)^{l-j} \right\}, \end{aligned} \tag{2.2}$$

where $l \in \{1, 2, \ldots\}$. Also, the mean and the variance of ε_n are $\mu_\varepsilon = \theta q(1-\alpha)$ and $\sigma_\varepsilon^2 = \theta q(1+\alpha)((1+q)(1-\alpha)-\alpha)$, respectively.

2.1. Model properties

Here, we present some characteristic features of the introduced model. The properties which are the same as those given in case of CGINAR(p) in [15] will just be stated without any derivation, while the others, which are newly introduced or dependent of marginal distribution, will be presented in much more detail.

The autocorrelation function satisfies the following equation

$$\rho(k) = \alpha \sum_{j=1}^{p} \phi_j \rho(|k-j|), \tag{2.3}$$

where $\rho(k)$ is decreasing exponentially to zero as k tends to infinity. Also, the CNBINAR(p) is a pth order Markov process, which is strictly stationary and ergodic. Therefore, in order to have the joint probability function it is enough to derive the transition probabilities. Let $I = \{j | x_{n-j} = 0, j = 1, \ldots, p\}$, then

$$\begin{aligned} P(X_n = x_n | H_{n-1}) &= \sum_{j \notin I} \phi_j P\left(\sum_{i=1}^{X_{n-j}} W_i + \varepsilon_n = x_n | H_{n-1}\right) + \sum_{j \in I} \phi_j P\left(\varepsilon_n = x_n\right) \\ &= \sum_{j \notin I} \phi_j P\left(\sum_{i=1}^{x_{n-j}} W_i + \varepsilon_n = x_n\right) + \sum_{j \in I} \phi_j P\left(\varepsilon_n = x_n\right), \end{aligned}$$

where

$$P\left(\sum_{i=1}^{x_{n-j}} W_i + \varepsilon_n = x_n\right) = \sum_{k=0}^{x_n} \binom{x_{n-j} - 1 + x_n - k}{x_{n-j} - 1} \frac{\alpha^{x_n - k}}{(1+\alpha)^{x_n - k + x_{n-j}}} P(\varepsilon_n = k)$$

and probabilities of the innovations ε_n are given by (2.2).

Now, let investigate the time reversibility of the process. Since, it is a Markov process of order p, we ought to check whether its joint probability generating function $\Phi_{X_n,X_{n-1},\dots,X_{n-p}}(s_1,s_2,\dots,s_{p+1})$ is symmetric in $s_1,s_2,\dots,s_{p+1}$. This pgf is as follows

$$E\left(s_1^{X_n}s_2^{X_{n-1}}\dots s_{p+1}^{X_{n-p}}\right)=\sum_{i=1}^{p}\phi_i E\left(s_1^{\alpha*X_{n-i}+\varepsilon_n}s_2^{X_{n-1}}\dots s_{p+1}^{X_{n-p}}\right)$$

Hence,

$$\Phi_{X_n,\dots,X_{n-p}}(s_1,\dots,s_{p+1})=\Phi_{\varepsilon_n}(s_1)\sum_{i=1}^{p}\phi_i E\left(s_1^{\alpha*X_{n-i}}s_2^{X_{n-1}}\dots s_{p+1}^{X_{n-p}}\right), \tag{2.4}$$

where, using notation $p(x_{n-1},\dots,x_{n-p})$ for $P(X_{n-1}=x_{n-1},\dots,X_{n-p}=x_{n-p})$, we have that

$$\begin{aligned}
& E\left(s_1^{\alpha*X_{n-1}}s_2^{X_{n-1}}\dots s_{p+1}^{X_{n-p}}\right)= \\
= & \sum_{x_{n-1}=0}^{\infty}\dots\sum_{x_{n-p}=0}^{\infty}E\left(s_1^{\sum_{i=1}^{x_{n-1}}W_i}s_2^{x_{n-1}}\dots s_{p+1}^{x_{n-p}}\right)p(x_{n-1},\dots,x_{n-p}) \\
= & \sum_{x_{n-1}=0}^{\infty}\dots\sum_{x_{n-p}=0}^{\infty}(\Phi_W(s_1)s_2)^{x_{n-1}}\,s_3^{x_{n-2}}\dots s_{p+1}^{x_{n-p}}p(x_{n-1},\dots,x_{n-p}) \\
= & \Phi_{X_{n-1},X_{n-2},\dots,X_{n-p}}\left(\Phi_W(s_1)s_2,s_3,\dots,s_{p+1}\right),
\end{aligned}$$

which is a part of the first term in (2.4). Also, all other terms in (2.4) are similarly calculated, providing the joint pgf in the following form.

$$\begin{aligned}
& \Phi_{X_n,X_{n-1},\dots,X_{n-p}}(s_1,s_2,\dots,s_{p+1})= \\
= & \Phi_{\varepsilon_n}(s_1)\Big[\phi_1\Phi_{X_{n-1},X_{n-2},\dots,X_{n-p}}(\Phi_W(s_1)s_2,s_3,\dots,s_{p+1})+ \\
+ & \phi_2\Phi_{X_{n-1},X_{n-2},\dots,X_{n-p}}(s_2,\Phi_W(s_1)s_3,\dots,s_{p+1}) \\
+ & \dots+\phi_p\Phi_{X_{n-1},X_{n-2},\dots,X_{n-p}}(s_2,s_3,\dots\Phi_W(s_1)s_{p+1})\Big]
\end{aligned}$$

Now, when it can be confirmed that

$$\Phi_{X_n,X_{n-1},\dots,X_{n-p}}(s_1,s_2,\dots,s_{p+1})\neq\Phi_{X_n,X_{n-1},\dots,X_{n-p}}(s_2,s_1,\dots,s_{p+1}),$$

we conclude that the process is not time reversible.

Finally, we focus our attention on the model conditional properties. Using process definition (2.1) and some of pgf properties we obtain the process probability generating function.

$$\begin{aligned}
\Phi_{X_{n+1}|H_n}(s) &= \Phi_\varepsilon(s)\sum_{i=1}^{p}\phi_i\Phi_W(s)^{X_{n-i+1}} \\
\Phi_{X_{n+k}|H_n}(s) &= \Phi_\varepsilon(s)\left(\sum_{i=1}^{k-1}\phi_i\Phi_{X_{n+k-i}|H_n}(\Phi_W(s))\right.
\end{aligned}$$

$$+\sum_{i=k}^{p}\phi_i\Phi_W(s)^{X_{n+k-i}}\Bigg),\quad 2\leqslant k\leqslant p, \tag{2.5}$$

$$\Phi_{X_{n+k}|H_n}(s) \;=\; \Phi_\varepsilon(s)\sum_{i=1}^{p}\phi_i\Phi_{X_{n+k-i}|H_n}(\Phi_W(s)),\quad k\geqslant p+1.$$

Then, using the equality $\Phi'_{X_{n+k}|H_n}(1)=E(X_{n+k}|H_n)$, it directly follows that the process regression properties can be obtained from the following equations.

$$\begin{aligned}
E(X_{n+1}|H_n) &= \alpha\sum_{i=1}^{p}\phi_iX_{n+1-i}+\mu_\varepsilon\\
E(X_{n+k|H_n}) &= \alpha\left(\sum_{i=1}^{k-1}\phi_iE(X_{n+k-i}|H_n)\right.\\
&\quad\left.+\sum_{i=k}^{p}\phi_iX_{n+k-i}\right)+\mu_\varepsilon,\quad 2\leqslant k\leqslant p, \qquad (2.6)\\
E(X_{n+k|H_n}) &= \alpha\sum_{i=1}^{p}\phi_iE(X_{n+k-i}|H_n)+\mu_\varepsilon,\quad k\geqslant p+1.
\end{aligned}$$

If we denoted the corresponding sum in the right part of (2.6) with $S^{(m)}$, for $mp+1\leqslant k\leqslant(m+1)p$, then by simple recursive derivation we could obtain the following

$$\begin{aligned}
E(X_{n+k}|H_n) &= \alpha^{m+1}\sum_{i=1}^{p}\phi_iS^{(m)}+\mu_\varepsilon\left(1+\alpha+\alpha^2+\ldots+\alpha^m\right)\\
&= \alpha^{(m+1)}S^{(m+1)}+\mu_\varepsilon\frac{1-\alpha^{m+1}}{1-\alpha}.
\end{aligned}$$

Therefore,

$$\lim_{k\to\infty}E(X_{n+k}|H_n)=\frac{\mu_\varepsilon}{1-\alpha}=\theta q=E(X_n).$$

On the other hand, by the same approach and using the fact that the second order moment equals $\Phi''_X+\Phi'_X$, k-step ahead conditional variance is derived as

$$\begin{aligned}
Var(X_{n+k}|H_n) &= \alpha^2\sum_{i=1}^{p}\phi_iVar(X_{n+k-i})-\left(\alpha\sum_{i=1}^{p}\phi_iE(X_{n+k-i}|H_n)\right)^2\\
&+ (\alpha+2\alpha^2)\sum_{i=1}^{p}\phi_iE(X_{n+k-i}|H_n)+\sigma_\varepsilon^2.
\end{aligned}$$

Easily, we show that

$$\begin{aligned}
\lim_{k\to\infty}Var(X_{n+k}|H_n) &= \theta q(q-\alpha^2q-2\alpha^2-\alpha+1)\\
&+ (\alpha+\alpha^2)E(X_n)+\alpha^2E(X_n^2)-\alpha^2(E(X_n))^2\\
&= \theta q(q+1)=Var(X_n). \qquad (2.7)
\end{aligned}$$

Since in the case of standard INAR processes the conditional variance is a linear function of X_n, it is interesting to notice that our process conditional variance quadratically depends on its history values, where this impact is realized through α^2 value. Based on (2.7), it means that especially strongly correlated overdispersed data series might be a good candidate for modeling by here introduced model.

3. Parameter estimation

Here, in order to estimate the unknown model parameters $\alpha, \phi_1, \phi_2, \ldots, \phi_p$, q and θ, we present some non-parametric procedures. The obtained statistics are based on the finite process random sample $X_1, X_2, \ldots, X_n$ and are derived using approach of Yule-Walker and the conditional least squares method.

3.1. Method of moments

Using the results of [15], we already have the following strongly consistent estimates with asymptotical normal distribution.

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

where $\mu_X = E(X_n) = q\theta$ and $D_1, D_2, \ldots, D_p$ and D are the Crammer's Rule determinants, used in solving the linear system, defined by (2.3).

Further, using that $Var(X_n) = \sigma_X^2 = q\theta(1+q) = \mu_X(1+q)$, we have that

$$\hat{q}^{yw} = \frac{\hat{\sigma^2}_X^{yw}}{\hat{\mu}_X^{yw}} - 1 = \frac{\sum_{i=1}^{N}(X_i - \overline{X}_N)^2 - \sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} X_i}, \tag{3.1}$$

which is, based on the process ergodic property, a strongly consistent estimator. Now, using the Theorem 1 of [20], it follows that $[\hat{\mu}_X^{yw}, \hat{\sigma^2}_X^{yw}]^T$ is an asymptotically normally distributed, strongly consistent estimator of $[q\theta, q\theta(1+q)]^T$, where after applying Proposition 6.4.3, from [7], it is obtained that $\hat{q}^{yw}$ has an asymptotic normal distribution.

Finally, based on preceding derivation, pameter θ is estimated via strongly consistent statistics

$$\hat{\theta}^{yw} = \frac{\hat{\mu}_X^{yw}}{\hat{q}^{yw}} = \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{\sum_{i=1}^{N}(X_i - \overline{X}_N)^2 - \sum_{i=1}^{N} X_i}. \tag{3.2}$$

Since $\hat{\mu}_X^{yw}$ has an asymptotic normal distribution and $\hat{q}^{yw} \xrightarrow{w.p.1} q$, then using Slutsky theorem, asymptotic normality of $\hat{\theta}^{yw}$ directly follows, furthermore with distribution parameters equal to asymptotical mean and variance of $\hat{\mu}^{yw}$, given in [15] and divided by q and q^2, respectively.

3.2. Modified conditional least squares method

As the previous one, this method will partially be based on the corresponding conditional least squares estimating procedures used for the model presented in [15]. Namely, we minimize the sum of squares

$$\sum_{n=p+1}^{N}\left(X_n-\alpha\phi_1X_{n-1}-\ldots-\alpha\phi_pX_{n-p}-\left(1-\sum_{i=1}^{p}\alpha_i\right)\mu_X\right)^2,$$

by equating to zero the corresponding partial derivatives in respect to unknown parameters and solving the obtained problem, we have

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

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

where D^* and $D_i^*, i=1,2,\ldots,p$, are the determinants of Cramer's rule applied to the corresponding linear system. All these estimators are strongly consistent and asymptotically normally distributed.

We still need to estimate parameters q and θ. Using $\sigma_X^2=q\theta(1+q)$ and the dispersion Yule-Walker estimator we obtain the modified conditional least squares estimator via

$$\hat{q}_{mcls}=\frac{\hat{\sigma^2}_X^{yw}}{\hat{\mu}_X^{cls}}-1, \tag{3.3}$$

which is obviously a strongly consistent estimator. Now, since from [15] it follows that $N^{\frac{1}{2}}\left(\hat{\mu}_X^{cls}-\hat{\mu}_X^{yw}\right)=o(1)$, for $N\to\infty$, we have that as $N\to\infty$

$$N^{\frac{1}{2}}\left(\hat{q}^{yw}-\hat{q}^{mcls}\right)=N^{\frac{1}{2}}\left(\frac{\hat{\sigma^2}_X^{yw}}{\hat{\mu}_X^{yw}}-\frac{\hat{\sigma^2}_X^{yw}}{\hat{\mu}_X^{cls}}\right)=\frac{\hat{\sigma^2}_x^{yw}N^{\frac{1}{2}}\left(\hat{\mu}_X^{cls}-\hat{\mu}_X^{yw}\right)}{\hat{\mu}_X^{cls}\hat{\mu}_X^{yw}}=o(1).$$

This is a sufficient condition for applying Proposition 6.3.3 [7], from which follows that $\hat{q}^{mcls}$ has the same asymptotic normal distribution as $\hat{q}^{yw}$.

The estimator of parameter θ is

$$\hat{\theta}^{mcls}=\frac{\hat{\mu}_X^{cls}}{\hat{q}^{mcls}}. \tag{3.4}$$

Since, as above, $N^{\frac{1}{2}}\left(\hat{\theta}^{mcls}-\hat{\theta}^{yw}\right)=o(1)$, $N\to\infty$, then modified conditional least squares estimator $\hat{\theta}^{mcls}$ is strongly consistent and asymptotically normally distributed with the same "mean" and "variance" as $\hat{\theta}^{yw}$.

4. Empirical results

The main results, referring to the subject of this paper, are contained in this section. At first, we discuss the reasons of very convenient application of INAR models based on geometric counting sequence to dynamical, self-generating counting processes. Further, we describe the situations in which the negative binomial marginal distribution is more appropriate to choose than the geometric one. In the second part of this section, we corroborate this discussion with real data example.

4.1. Interpretation

The INAR models which were first developed and probably the most commonly used in practice were those based on Bernoulli counting sequence, i.e. binomial thinning operator. Such models are ideal for counting processes where the observed population members or random events can contribute to the overall sum by 1 or 0, or in other words may survive or vanish through time. However, when the observed unit is capable of generating more counting objects or produce more new random events, then Bernoulli random variable is no more the best choice for constructing the counting sequence. In order to cope with this problem [17] introduced a negative binomial thinning, which was based on the geometrically distributed counting sequence. Considering the nature of the distribution, it was more appropriate for modeling counting processes, which referred to population elements or random events capable of replication or production of other elements or events. Briefly speaking, these counting objects might contribute to the overall sum by 0, 1, 2 or more. Based on this fact, [15] obtained better performance in modeling crime counting data using their Combined INAR(p) model based on negative binomial thinning than by Combined INAR(p) based on binomial thinning, introduced in [21].

Although, due to the negative binomial thinning, CGINAR(p) has proved to be a quite good choice for these dynamical data, there are situations in which this model could be further significantly improved. Thus, there are certain data which are not enough compatible with geometric marginal distribution. This can be noticed especially in counting processes which contain only few zeros and also are comprised mainly of two-digit non-negative integers, i.e. which sample mode is grater than zero. It turned out that this is a characteristic for many of the light criminal activity counting series. So, we came up with the idea of using a negative binomial marginal distribution, which happened to be a much better fit to the data described. We can explain this in the following. Suppose that we want to model a monthly counting of light criminal activities, such as purse snatching, simple assaults or motor vehicle thefts, through a period of several years. Now, let A represent a random event of a "registered theft of a motor vehicle by a police station", where $P(A) = \frac{q}{1+q}$, which correspond to our notation in preceding sections. Process realization $X_i = x$ means that during the ith month, there have been registered a number of x motor vehicle thefts. So, if the marginal probability mass function is geometric defined by $P(X_i = x) = \frac{q^x}{(1+q)^{x+1}}$, than $\{X_i = x\}$ represents the realization of the compound random event $\underbrace{AA\ldots A}_{x} A^c$, i.e. after x registered vehicle thefts, one theft was not registered or it was just an attempt of a crime. Unfortunately, in real life after one realization of a A^c a sequence of crimes $AA\ldots A$ might continue in the same month. Therefore, during one month it is much more realistic to expect to happen something like this $\underbrace{AA\ldots A}_{y_1} A^c \underbrace{AA\ldots A}_{y_2} A^c \ldots \underbrace{AA\ldots A}_{y_\theta} A^c$, where $y_1 + y_2 + \ldots + y_\theta = x$. Here, during the ith month in the counting series, x represents the number of A realized events, and θ stands for the number of realizations of A^c, which is equivalent to $\{X_i = x\}$, where $X_i : \mathcal{NB}\left(\theta, \frac{q}{1+q}\right)$ has a negative binomial probability mass function defined by $P(X_i = x) = \binom{\theta-1+x}{\theta-1}\frac{q^x}{(1+q)^{x+\theta}}$, $\theta > 0$, $q > 0$. After all, it is obvious that negative binomial marginals are much more realistic choice than the geometric. This is also supported by the fact that the mode of the negative binomial distribution equals $q(\theta-1)$, which is grater than zero far all $\theta > 1$, which is much more compatible with the characteristics of the considered counting data than the geometric mode, which is always zero. All the reasons discussed above, motivate us to introduce a combined negative binomial integer-valued autoregressive model of order p which qualities will be empirically tested on the real data series in the following.

4.2. Real data example

According to previous model interpretation, here we try to find the most appropriate INAR modeling of a data series representing a counting of a certain light criminal activity. Namely, from a web sight Forecasting Principles we have obtained a time series of monthly count of the motor vehicle theft (MVTheft). These crimes are reported in the 11th police car beat in Pittsburgh in a period from January 1990 to December 2001, constituting a sequence of 144 observations. Its sample mean, variance and autocorrelation are 4.917, 10.678 and 0.354, so the overdispersion is evident. The plots of the time series, the autocorrelation and partial autocorrelation functions are given in Figure 4.1, from which we can conclude that it is justify to use INAR(2) modeling. However, in order to perform a more complete survey, we have decided to compare our CNBINAR model to some competitive models of order 1 and order 2. With the same objective, though the data series is overdispersed and its sample mode equals 3, i.e. more than zero, we shall still consider the models with Poisson and geometric marginals, too. Thus, in the case of the first order model application we have compared CNBINAR(1) to the following INAR(1) models: INAR(1) model with Poisson marginals introduced in [2], Quasi-binomial INAR(1) model with generalized Poisson marginals given in [5], Geometric INAR(1) model defined in [4], New Geometric INAR(1) of [17], Negative binomial INAR(1) given in [24, 25], Iterated INAR(1) model with negative binomial marginals constructed in [1], Random Coefficient INAR(1) model with negative binomial marginals introduced in [23] and Mixed INAR(1) model with geometric marginals defined in [16].

Also, we have tested the performance of the CNBINAR(2) against some competitive known models of order 2. These are the Combined INAR(2) model with Poisson marginals given in [21], Combined Geometric INAR(2) model introduced in [15] and Mixed Geometric INAR(2) model defined in [19]. We carried out the models quality comparison by calculating the Akaike and Bayesian information criteria (AIC and BIC), as well as the root mean squares of differences between the observations and predicted values (RMS). These values together with the maximum likelihood estimates of the model parameters are presented in Table 4.1 and Table 4.2.

Table 4.1: ML parameter estimates, AIC, BIC and RMS for INAR(1) modeling of the MVTheft counts data.

Model	MLE	AIC	BIC	RMS
PoINAR(1)	$\hat{\lambda} = 3.6782$			
	$\hat{\alpha} = 0.2512$	747.6862	753.6258	3.0723
GPQINAR(1)	$\hat{\lambda} = 2.3023$			
	$\hat{\theta} = 0.2938$			
	$\hat{\rho} = 0.337$	705.9824	714.8918	3.2402
GINAR(1)	$\hat{q} = 0.7847$			
	$\hat{\alpha} = 0.4555$	750.9045	756.8442	3.1436
NGINAR(1)	$\hat{\mu} = 4.3321$			
	$\hat{\alpha} = 0.6445$	727.2790	733.2187	3.1996
NBINAR(1)	$\hat{q} = 0.4866$			
	$\hat{\theta} = 4.6561$			
	$\hat{\alpha} = 0.3333$	702.9605	711.8699	3.0540
NBIINAR(1)	$\hat{n} = 4.2965$			
	$\hat{p} = 1.3959$			
	$\hat{\rho} = 0.3730$	704.5925	713.5020	3.0536
NBRCINAR(1)	$\hat{n} = 4.1971$			
	$\hat{p} = 0.4598$			
	$\hat{\rho} = 0.3837$	703.4900	712.3994	3.0545
MGINAR(1)	$\hat{\mu} = 4.3254$			
	$\hat{\alpha} = 0.6470$			
	$\hat{p} = 0.0393$	729.2264	738.1359	3.2021
CNBINAR(1)	$\hat{q} = 1.1870$			
	$\hat{\theta} = 4.1399$			
	$\hat{\alpha} = 0.3707$	702.5546	711.4641	3.0535

Providing the smallest values of the observed criteria, we notice that our model is the most appropriate to use for the considered data series. This is particularly

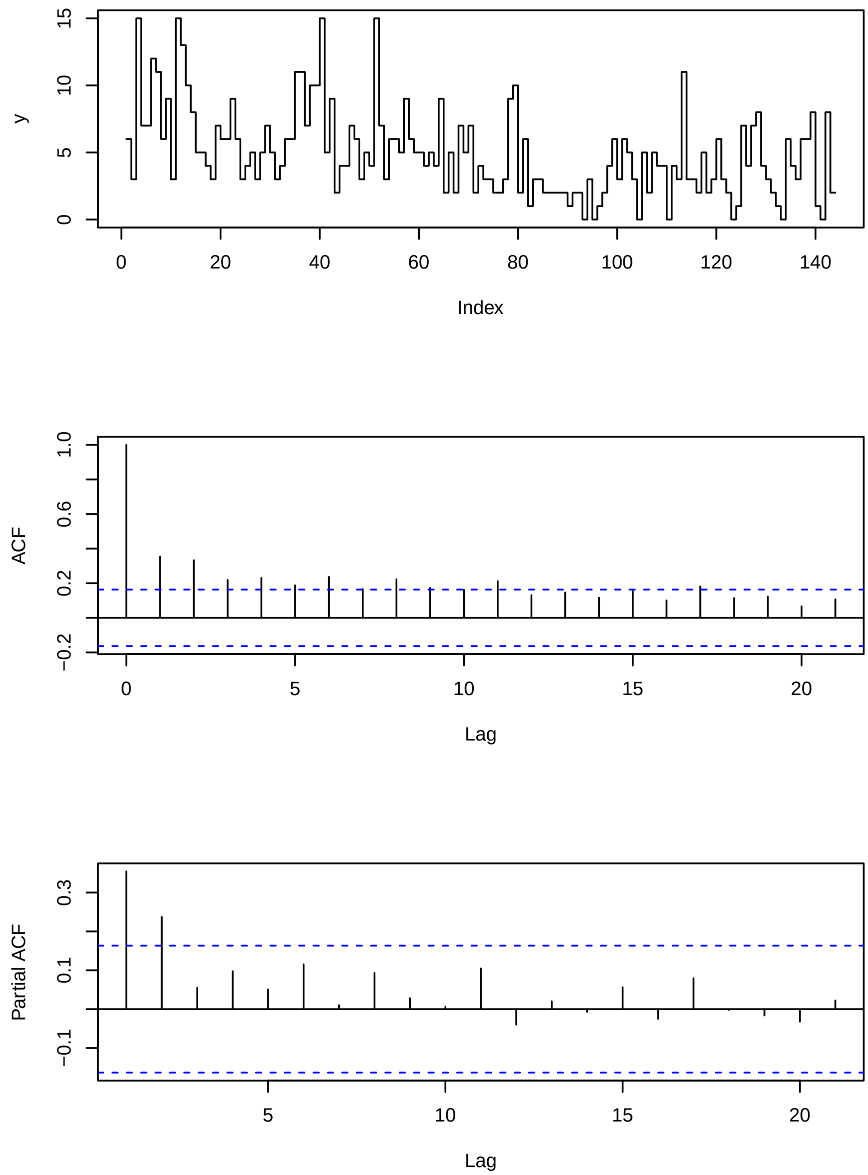

FIG. 4.1: MVTheft series, autocorrelations and partial autocorrelations

Table 4.2: ML parameter estimates, AIC, BIC and RMS for INAR(2) modeling of the MVTheft counts data.

Model	MLE	AIC	BIC	RMS
CPoINAR(2)	$\hat{\lambda} = 3.1594$			
	$\hat{\alpha} = 0.3576$			
	$\hat{\phi}_1 = 0.5040$	728.3877	737.2971	2.9986
CGINAR(2)	$\hat{\mu} = 4.2736$			
	$\hat{\alpha} = 0.7218$			
	$\hat{\phi}_1 = 0.5452$	707.6046	716.5140	3.0256
MGINAR(2)	$\hat{\mu} = 3.9349$			
	$\hat{\alpha} = 0.7185$			
	$\hat{\phi}_1 = 0.5452$	706.8205	715.7299	3.0595
CNBINAR(2)	$\hat{q} = 1.2008$			
	$\hat{\theta} = 4.0780$			
	$\hat{\alpha} = 0.4633$			
	$\hat{\phi}_1 = 0.5329$	690.0126	701.8919	2.9716

evident among the second order models where it shows the best performance, despite the fact of having the largest number of unknown parameters. In order to understand better this behavior, we might investigate the adequacy of the used marginal distributions in case of the observed counting. For this purpose we have derived the expected probabilities for all the considered distributions, which are given in respect of the observed probabilities in Table 4.3. In support of this, the observed frequencies and the expected frequencies for each of the applied processes marginal distributions are presented in Figure 4.2 and Figure 4.3, respectively. It is easy to see that the negative binomial and the generalized Poisson distribution provide the best fits of the considered data. Since, based on their graphs it is not clear which of these two distributions is the most appropriate, we have applied a χ^2 fit test. The results are given in Table 4.4. Based on the p-value it seems that the generalized Poisson distribution gives the best fit. However, our model still shows the best performance. This could only be justified by implementation of the thinning operator based on the geometric counting sequence, which happened to be in better accord with the self-generating nature of the crime data than the binomial thinning operator, used in the corresponding GPQINAR model.

Finally, we can conclude that the CNBINAR is the only one, among all the considered models, which have both, the negative binomial thinning operator and the negative binomial marginals. So, these together with the occurring nature of MVTheft events, according to the model interpretation given above, provide an explanation of the most appropriate performance of here introduced model.

Table 4.3: The observed and expected probabilities for the MVTheft counts data.

	Observed probabilities	Geometric	Poisson	NB	GP
0	0.0486	0.1690	0.0073	0.0386	0.0356
1	0.0417	0.1404	0.0360	0.0874	0.0860
2	0.1528	0.1167	0.0885	0.1225	0.1241
3	0.1597	0.0970	0.1451	0.1365	0.1395
4	0.1111	0.0806	0.1783	0.1325	0.1350
5	0.1181	0.0670	0.1753	0.1172	0.1184
6	0.1250	0.0557	0.1437	0.0969	0.0969
7	0.0764	0.0462	0.1009	0.0762	0.0754
8	0.0278	0.0384	0.0620	0.0575	0.0565
9	0.0417	0.0319	0.0339	0.0421	0.0411
10	0.0278	0.0265	0.0167	0.0299	0.0292
11	0.0278	0.0221	0.0074	0.0208	0.0203
12	0.0069	0.0183	0.0031	0.0142	0.0140
13	0.0069	0.0152	0.0012	0.0096	0.0095
14	0.0000	0.0127	0.0004	0.0063	0.0064
15	0.0278	0.0623	0.0002	0.0118	0.012

Table 4.4: χ^2 fit tests.

Distribution	Geometric	Poisson	NB	GP
χ^2 test	59.6451	613.4258	14.3829	14.1152
p-value	0.0000	0.0000	0.3474	0.3658

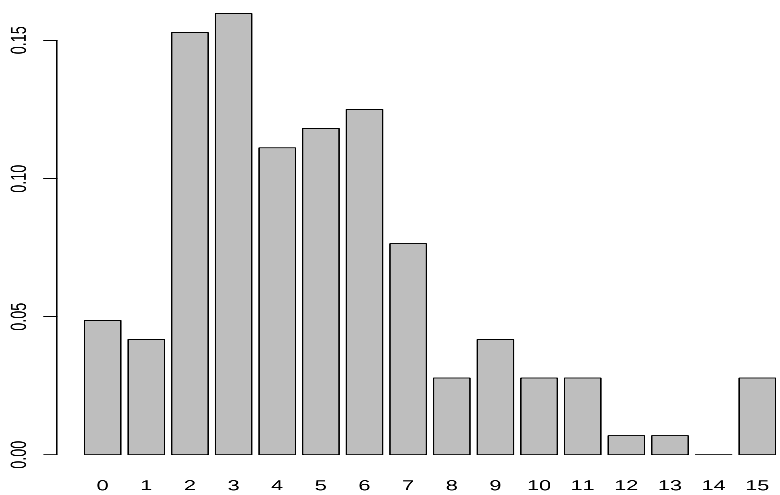

FIG. 4.2: MVTheft observed frequencies

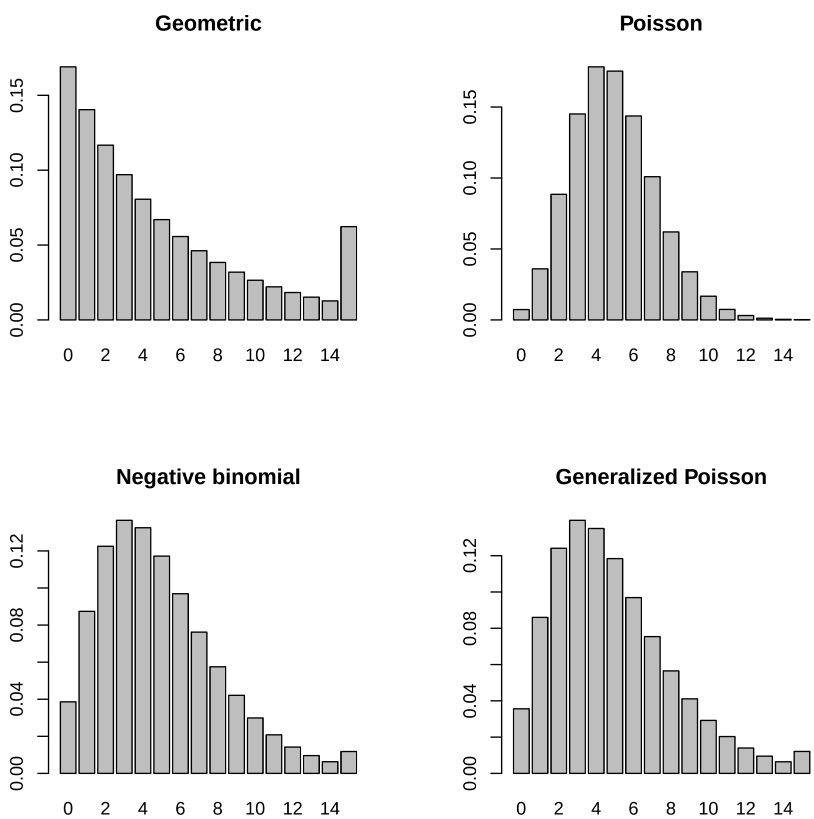

FIG. 4.3: MVTheft expected frequencies

REFERENCES

1. M.A. Al-Osh and E.E.A.A. Aly: *First order autoregressive time series with negative binomial and geometric marginals*. Commun. Statist. Theory Meth. **21** (1992), 2483–2492.
2. M.A. Al-Osh and A.A. Alzaid: *First-order integer-valued autoregressive* ($INAR(1)$) *process*. J. Time Ser. Anal. **8** (1987), 261–275.
3. E.E.A.A. Aly and N. Bouzar: *On Some Integer-Valued Autoregressive Moving Average Models*. J. Mult. Anal. **50** (1994), 132–151.
4. A.A. Alzaid and M.A. Al-Osh: *First-order integer-valued autoregressive* ($INAR(1)$) *process: distributional and regression properties*. Stat. Neerlandica **42** (1988), 53–61.
5. A.A. Alzaid and M.A. Al-Osh: *Some autoregressive moving average processes with generalized Poisson marginal distributions*. Ann. Inst. Statist. Math. **45** (1993), 223–232.
6. H.S. Bakouch and M.M. Ristić: *Zero Truncated Poisson Integer Valued AR(1) Model*. Metrika **72(2)** (2010), 265–280.
7. P.J. Brockwell and R.A. Davis: *Time Series: Theory and Methods*. New York: Springer-Verlag, 1987.
8. D.R. Cox and H.D. Miller: *The Theory of Stochastic Processes*. London: Methuen, 1965.
9. P.A. Jacobs and P.A.W. Lewis: *Discrete Time Series generated by Mixtures I: Correlation and Runs Properties*. J.R. Statist. Soc. (B) **40** (1987a), 94–105.
10. P.A. Jacobs and P.A.W. Lewis: *Discrete Time Series generated by Mixtures II: Asymptotic Properties*. J.R. Statist. Soc. (B) **40** (1987b), 222–228.
11. P.A. Jacobs and P.A.W. Lewis: *Discrete Time Series generated by Mixtures III: Autoregressive Processes (DAR(p))*. Naval Postgraduate School Technical Report, NPS55Lw 73061A, 1987c.
12. A. Latour: *Existence and stochastic structure of a non-negative integer-valued autoregressive process*. J. Time Ser. Anal. **19** (1998), 439–455.
13. E. McKenzie: *Some simple models for discrete variate time series*. Water Resour. Bull. **21** (1985), 645–650.
14. E. McKenzie: *Autoregressive moving-average processes with negative binomial and geometric distributions*. Adv. Appl. Prob. **18** (1986), 679–705.
15. A.S. Nastić, M.M. Ristić and H.S. Bakouch: *A combined geometric* $INAR(p)$ *model based on negative binomial thinning*. Math. Comp. Modell. **55** (2012), 1665–1672.
16. A.S. Nastić and M.M. Ristić: *Some geometric mixed integer-valued autoregressive (INAR) models*. Stat. Prob. Lett. **82** (2012), 805–811.
17. M.M. Ristić, H.S. Bakouch and A.S. Nastić: *A new geometric first-order integer-valued autoregressive (NGINAR(1)) process*. J. Stat. Plan. Inf. **139** (2009), 2218–2226.
18. M.M. Ristić, A.S. Nastić and H.S. Bakouch: *Estimation in an Integer-Valued Autoregressive Process with Negative Binomial Marginals (NBINAR(1))*. Commun. Statist. Theory Meth. **41** (2012), 606–618.
19. M.M. Ristić and A.S. Nastić: *A mixed INAR(p) model*. J. Time Ser. Anal. **33** (2012), 903–915.
20. I. Silva and M.E. Silva: *Asymptotic distribution of the Yule-Walker estimator for* $INAR(p)$ *processes*. Stat. Prob. Lett. **76** (2006), 1655–1663.

21. C.H. Weiss: *The combined INAR(p) models for time series of counts*. Stat. Prob. Lett. **72** (2008), 1817–1822.

22. H. Zheng, I.V. Basawa and S. Datta: *Inference for pth-order random coefficient integer-valued autoregressive processes*. J. Time Ser. Anal. **27** (2006), 411–440.

23. H. Zheng, I.V. Basawa and S. Datta: *First-order random coefficient integer-valued autoregressive processes*. J. Stat. Plann. Inf. **137** (2007), 212–229.

24. R. Zhu and H. Joe: *Modelling count data time series with Markov processes based on binomial thinning*. J. Time Ser. Anal. **27** (2006), 727–738.

25. R. Zhu and H. Joe: *Negative binomial time series models based on expectation thinning operators*. J. Stat. Plann. Inf. **140** (2010), 1874–1888.

Aleksandar S. Nastić
University of Niš
Faculty of Sciences and Mathematics
Department of Mathematics
Niš, Serbia
anastic78@gmail.com