FACTA UNIVERSITATIS (NIŠ)
Ser. Math. Inform. Vol. 36, No 1 (2021), 143-156
https://doi.org/10.22190/FUMI200420012P

Original Scientific Paper

A MIXED (NONLINEAR) INAR(1) MODEL

Predrag M. Popović

Faculty of Civil Engineering and Architecture, Department of Mathematics, Informatics and Physics, 18000 Niš, Serbia

Abstract. The paper introduces a new autoregressive model of order one for time series of counts. The model is comprised of a linear as well as nonlinear autoregressive component. These two components are governed by random coefficients. The autoregression is achieved by using the negative binomial thinning operator. The method of moments and the conditional maximum likelihood method are discussed for the parameter estimation. The practicality of the model is presented on a real data set.
Keywords: Time series of counts, Negative binomial thinning operator, Linear model, Nonlinear model.

1. Introduction

In the past few decades, time series modeling has been drawing a lot of attention to researchers as well as practitioners. Understanding the dependence and the evolution of an observed series is an important task. A significant contribution in this field is modeling time series of counts. Time series of counts arises in many real-life situations. For example, number of infected persons, number of stock transactions, number of spaces, number of committed crimes, etc. Studding of these types of time series started after the introduction of the thinning operator in [15]. Some of the first integer-valued autoregressive (INAR) models based on the thinning parameter are presented in [11], [1], [2]. These models experienced various modifications regarding their structure, the definition of thinning operator and the dimensionality. A comprehensive review of INAR models can be found in [16] and [14]. The extension to bivariate INAR models can be found in [7], [9], [8].

Received April 20, 2020; accepted May 28, 2020.
Corresponding Author: Predrag M. Popović, Faculty of Civil Engineering and Architecture, Department of Mathematics, Informatics and Physics, 18000 Niš, Serbia | E-mail: popovicpredrag@yahoo.com
2010 *Mathematics Subject Classification.* Primary 62M10

In general, the INAR models are composed of the survival and the innovation process. The survival process is an autoregressive component, which is defined through the thinning operator. Some of the most exploited thinning operators are the binomial thinning operator introduced in [15], and the negative binomial thinning operator introduced in [13]. The autoregressive component, usually named the survival process, is of the form

$$\alpha \circ X = \sum_{i=1}^{X} W_i,$$

where W_i is a counting sequence. And the major drawback of the autoregressive models is that they put too much or too little weight on thier previous value when predicting the next one. Some of the solutions to this problem were given in [5] where the thinning parameter α is governed by an external process. The generalization of this model was discussed in [6] and [4]. Some other modifications of the autoregressive dependence are based on introducing a bilinear autoregressive component, [3]. Also, there are autoregressive INAR modes that are dealing with the excess number of zeros and ones [12].

The aim of this paper is to introduce a model whose autoregressive part is comprised of a linear as well as a nonlinear component. The nonlinear component is defined through the current state of the innovation process. The idea for that lies in the fact that the survival process might depend on the innovation process. For example, if we have a lot of new specimens of some population, probably the environment conditions are adequate for that species so the survival rate will be higher. Random coefficients determine whether the autoregressive component is linear or not. The linear, as well as the nonlinear component, are defined through the negative binomial thinning operator. Even though the model has this complex definition of the autoregressive component, the conditional expectation can be determined. This fact increases the practical aspect of the model, since the one-step-ahead prediction is possible. Also, the model is proved to be stationary.

The next section gives us the definition of the model. In Section 3. the main properties of the model are derived. Section 4. proposes two methods for the parameter estimation, whose efficiency are tested in Section 5. Section 6. discusses the practical aspect of the model. The concluding remarks are given in Section 7.

2. Model definition

In this section, we introduce the Mixed nonlinear INAR(1) model (MNLINAR(1)) in a general form, without specifying a distribution of the innovation process. For such a model, we prove the existence and the strict stationarity. Also, the main properties of the model are derived.

Let $\{X_t\}$ be a non-negative integer-valued time series. Then, the MNLINAR(1) model is defined as follows:

$$X_t = \begin{cases} \alpha * X_{t-1} + \varepsilon_t, & \text{w.p. } p \\ \alpha * (X_{t-1}\varepsilon_t) + \varepsilon_t, & \text{w.p. } 1-p \end{cases} \tag{2.1}$$

where the negative binomial thinning operator is defined as $\alpha * X = \sum_{i=1}^{X} W_i$, where $\{W_t\}$ is independent identically distributed random variables with geometric marginal distribution $Geom(\alpha/(1+\alpha))$, whose probability mass function is $P(W_i = w) = \frac{\alpha^w}{(1+\alpha)^{w+1}}$. The counting sequence $\{W_t\}$ are independent of $\{X_t\}$ and $\{\varepsilon_t\}$. Further, the random variable ε_t is independent of X_s for $s < t$.

As we can see, the MNLINAR(1) model evolves as a linear model with probability p and as a nonlinear model with probability $1-p$. So, the model can be expressed with random variables U_t and V_t where $P(U_t = \alpha, V_t = 0) = 1 - P(U_t = 0, V_t = \alpha) = p$. Than the MNLINAR(1) model is defined as

$$X_t = U_t * X_{t-1} + V_t * (X_{t-1}\varepsilon_t) + \varepsilon_t. \tag{2.2}$$

Theorem 2.1. *There exist a unique strictly stationary bivariate time series $\{X_t\}$ that satisfies equation (2.2), when $\alpha(p + (1-p)\lambda) < 1$, $\alpha^2(p + (1-p)E(\varepsilon_t^2)) < 1$ and $E(\varepsilon_t^2) < \infty$, where λ stands for $E(\varepsilon_t)$.*

Proof. Let us introduce a series $\{X_t^{(n)}\}$ in the following way:

$$X_t^{(n)} = \begin{cases} 0, & n < 0 \\ \varepsilon_t, & n = 0 \\ U_{(t)} * X_{t-1}^{(n-1)} + V_{(t)} * (X_{t-1}^{(n-1)}\varepsilon_t) + \varepsilon_t, & n > 0 \end{cases}.$$

Here, notations $U_{(t)}$ and $V_{(t)}$ implies that the counting series that figure in $U_{(t)} * X^{(n)}$ are fixed at time t for all n. Now, we define the Hilbert space $L^2(\Omega, \mathcal{F}, P) = \{X : E(X^2) < \infty\}$, where the measure between two random variables is defined as $E(XY)$. The idea is to prove that $\{X_t^{(n)}\}$ is strictly stationary, and then to show that $\{X_t^{(n)}\}$ is a Cauchy sequence that belongs to just defined L^2 space.

Using the same approach as in [3], it can be proved that the series $\{X_t^{(n)}\}$ is strictly stationary, so we omit that proof here.

To show that $X_t^{(n)}$ belong to the above defined Hilbert space, we need to prove that $E(X_t^{(n)})^2 < \infty$. For $n \leq 0$ it obviously holds, thus let us focus on $n > 0$. We obtain the following equation:

$$
\begin{aligned}
E\left(X_t^{(n)}\right)^2 &= pE(\alpha * X_{t-1}^{(n-1)} + \varepsilon_t)^2 + (1-p)E(\alpha * (X_{t-1}^{(n-1)}\varepsilon_t) + \varepsilon_t) \\
&= pE((\alpha * X_{t-1}^{(n-1)})^2 + 2\alpha * X_{t-1}^{(n-1)}\varepsilon_t + \varepsilon_t^2) \\
&\quad +(1-p)E((\alpha * (X_{t-1}^{(n-1)}\varepsilon_t))^2 + 2\alpha * (X_{t-1}^{(n-1)}\varepsilon_t)\varepsilon_t + \varepsilon_t^2) \\
&= p\Big[\alpha^2 E(X_{t-1}^{(n-1)})^2 + \alpha(1+\alpha)E(X_{t-1}^{(n-1)}) \\
&\quad +2\alpha E(X_{t-1}^{(n-1)})E(\varepsilon_t) + E(\varepsilon_t^2)\Big] \\
&+(1-p)\Big[\alpha E(X_{t-1}^{(n-1)}\varepsilon_t)^2 + \alpha(1+\alpha)E(X_{t-1}^{(n-1)}\varepsilon_t) \\
&\quad +2\alpha E(X_{t-1}^{(n-1)}\varepsilon_t^2) + E(\varepsilon_t^2)\Big] = \dots \\
&= E\left(X_{t-1}^{(n-1)}\right)^2 \left(p\alpha^2 + (1-p)\alpha^2 E(\varepsilon_t^2) + E(X_{t-1}^{(n-1)})\right) \\
&\quad \cdot \left[\alpha(1+\alpha)\left(p + (1-p)E(\varepsilon_t)\right) + 2\alpha\left(pE(\varepsilon_t) + (1-p)E(\varepsilon_t^2)\right)\right] + E(\varepsilon_t^2).
\end{aligned}
$$

Since the series $\{X_t^{(n)}\}$ is strictly stationary, it follows that $E\left(X_t^{(n)}\right)^2 < \infty$ if $1 - \alpha^2(p + (1-p)E(\varepsilon_t^2)) > 0$, which is satisfied by the condition of the theorem. In the above derivation, we used some known properties of the negative binomial thinning operator which can be found in [13].

Now, let us prove that $\{X_t^{(n)}\}$ is a Cauchy sequence. Notice that equation (2.3) holds if and only if the sequence $\{X_t^{(n)}\}$ is non-decreasing.

$$
(2.3) \quad X_t^{(n)} - X_t^{(n-1)} = U_{(t)} * (X_{t-1}^{(n-1)} - X_{t-1}^{(n-2)}) + V_{(t)} * ((X_{t-1}^{(n-1)} - X_{t-1}^{(n-2)})\varepsilon_t)
$$

To show that the sequence is non-decreasing we use mathematical induction. Notice that

$$
X_t^{(1)} = U_{(t)} * X_{t-1}^{(0)} + V_{(t)} * (X_{t-1}^{(0)}\varepsilon_t) + \varepsilon_t \geq \varepsilon_t = X_t^{(0)}.
$$

Suppose that $X_t^{(k)} > X_t^{(k-1)}$ for some k and let's prove it for $k+1$.

$$
\begin{aligned}
X_t^{(k)} &= U_{(t)} * X_{t-1}^{(k-1)} + V_{(t)} * (X_{t-1}^{(k-1)}\varepsilon_t) + \varepsilon_t \leq \\
&\leq U_{(t)} * X_{t-1}^{(k)} + V_{(t)} * (X_{t-1}^{(k)}\varepsilon_t) + \varepsilon_t = X_t^{(k+1)}.
\end{aligned}
$$

So, $\{X_t^{(n)}\}$ is non-decreasing and equation (2.3) holds. Taking expectation of the both sides of equation (2.3) we obtain that

$$\begin{aligned}
&E(X_t^{(n)} - X_t^{(n-1)}) \\
&= p\alpha E(X_{t-1}^{(n-1)} - X_{t-1}^{(n-2)}) + (1-p)\alpha E[(X_{t-1}^{(n-1)} - X_{t-1}^{(n-2)})\varepsilon_t] \\
&= p\alpha E(X_{t-1}^{(n-1)} - X_{t-1}^{(n-2)}) + (1-p)\alpha\lambda E(X_{t-1}^{(n-1)} - X_{t-1}^{(n-2)}) \\
&= (p\alpha + (1-p)\alpha\lambda) E(X_{t-1}^{(n-1)} - X_{t-1}^{(n-2)}) = \dots \\
&= (p\alpha + (1-p)\alpha\lambda)^{n-1} E(X_{t-1}^{(1)} - X_{t-1}^{(0)}) \\
&\qquad + (p\alpha + (1-p)\alpha\lambda)^n E(\varepsilon_t^2).
\end{aligned}$$

We can conclude that

$$E(X_t^{(n)} - X_t^{(n-1)}) \xrightarrow[n\to\infty]{} 0 \Longleftrightarrow p\alpha + (1-p)\alpha\lambda < 1.$$

Thus, $\{X_t^{(n)}\}$ is a Cauchy sequence in the above-defined Hilbert space which implies that the Cauchy sequence converges, i.e. $\lim_{n\to\infty} X_t^{(n)} = X_t$. Since the series $\{X_t^{(n)}\}$ is strictly stationary it follows that its limit is strictly stationary as well.

The uniqueness of the solution of equation (2.2) can be proved using the same approach as [3], so we omit it here. □

3. Properties of the model

In this section, we derive the most important properties of the MNLINAR(1) model, including the first and the second moments as well as the conditional expectation and the conditional probability mass function.

From the model definition given by equation (2.1), and the properties of the negative binomial thinning operator we obtain

$$E(X_t) = \alpha p E(X_{t-1}) + \alpha(1-p)E(X_{t-1})E(\varepsilon_t) + E(\varepsilon_t).$$

Having in mind that $\{X_t\}$ is a strictly stationary process, and relying on the conditions of Theorem 2.1, it follows that

$$E(X_t) = \frac{E(\varepsilon_t)}{1 - \alpha(p + (1-p)E(\varepsilon_t))}. \tag{3.1}$$

For the derivation of the second moment we use the same extensive technique as in Theorem (2.1), so we only notice that

$$\begin{aligned}
&E\left(X_t^2\right) = E\left(X_{t-1}\right)^2 (p\alpha^2 + (1-p)\alpha^2 E(\varepsilon_t^2) \\
&+E(X_{t-1})\left[\alpha(1+\alpha)(p + (1-p)E(\varepsilon_t)) + 2\alpha(pE(\varepsilon_t) + (1-p)E(\varepsilon_t^2))\right] + E(\varepsilon_t^2).
\end{aligned}$$

Under the conditions of Theorem (2.1), it follows that

$$\begin{aligned}
&E\left(X_t^2\right) = \\
&\frac{E(X_{t-1})\left[\alpha(1+\alpha)(p+(1-p)E(\varepsilon_t))+2\alpha(pE(\varepsilon_t) + (1-p)E(\varepsilon_t^2))\right]+E(\varepsilon_t^2)}{1-\alpha^2(p+(1-p)E(\varepsilon_t^2))}.
\end{aligned} \tag{3.2}$$

Further, let us pay the attention on the expected value of the product X_tX_{t-k}. It is equal to

$$\begin{aligned}
&E(X_tX_{t-k}) = pE((\alpha * X_{t-1} + \varepsilon_t)X_{t-k}) + (1-p)E((\alpha * (X_{t-1}\varepsilon_t) + \varepsilon_t)X_{t-k})\\
&= \alpha p E(X_{t-1}X_{t-k}) + \alpha(1-p)E(X_{t-1}X_{t-k})E(\varepsilon_t) + E(\varepsilon_t)E(X_{t-k})\\
&= (\alpha p + \alpha(1-p)E(\varepsilon_t))E(X_{t-1}X_{t-k}) + E(\varepsilon_t)E(X_{t-k}) = \dots\\
&= (\alpha p + \alpha(1-p)E(\varepsilon_t))^k E(X_{t-k}^2) + E(X_{t-k})E(\varepsilon_t)\sum_{j=0}^{k-1}(\alpha p + \alpha(1-p)E(\varepsilon_t))^j.
\end{aligned}$$

It can be notice that, under the conditions of Theorem 2.1,

$$E(X_tX_{t-k}) \xrightarrow[k\to\infty]{} \frac{E(X_{t-k})E(\varepsilon_t)}{1-\alpha(p+(1-p)E(\varepsilon_t))}.$$

Substituting $E(\varepsilon_t)$ by using equation (3.1), we obtain

$$E(X_tX_{t-k}) \xrightarrow[k\to\infty]{} E(X_t)E(X_{t-k}).$$

For further discussion, it will be particularly important the case when $k = 1$, so let us notice that

$$\begin{aligned}
E(X_tX_{t-1}) &= (\alpha p + \alpha(1-p)E(\varepsilon_t))E(X_{t-1}^2) + E(X_{t-1})E(\varepsilon_t)\\
&= (\alpha p + \alpha(1-p)E(\varepsilon_t))E(X_t^2) + E(X_t)E(\varepsilon_t).
\end{aligned} \tag{3.3}$$

But the autocorrelation structure of the series $\{X_t\}$ would be much easier to observe through the autocovariance function directly. Namely, we obtain the following:

$$\begin{aligned}
&Cov(X_t, X_{t-k}) = E(X_tX_{t-k}) - E(X_t)E(X_{t-k})\\
&= (\alpha p + \alpha(1-p)E(\varepsilon_t))E(X_{t-1}X_{t-k}) + E(\varepsilon_t)E(X_{t-k})\\
&= (\alpha p + \alpha(1-p)E(\varepsilon_t))Cov(X_{t-1}, X_{t-k})\\
&+E(\varepsilon_t)E(X_{t-k}) + (\alpha p + \alpha(1-p)E(\varepsilon_t))E(X_{t-1})E(X_{t-k}) - E(X_t)E(X_{t-k})\\
&= (\alpha p + \alpha(1-p)E(\varepsilon_t))Cov(X_{t-1}, X_{t-k}) = \dots\\
&= (\alpha p + \alpha(1-p)E(\varepsilon_t))^k Cov(X_{t-k}, X_{t-k}) = (\alpha p + \alpha(1-p)E(\varepsilon_t))^k Var(X_t).
\end{aligned}$$

The above equation follows from the property of the negative binomial thinning operator, which can be found in Lemma 3 of [13]. Now, it can be easily concluded that, under assumption of Theorem 2.1, the autocorrelation tends to zero when k tends to infinity.

Regarding the practicality of the MNLINAR(1) model, the most important aspect of the model is the ability to predict forthcoming values of a modeled series. Unlike some other nonlinear models ([3], [10]), for the MNLINAR(1) model the conditional expectation can be derived as

$$\begin{aligned}
E(X_t|X_{t-1}) &= pE(\alpha * X_{t-1} + \varepsilon_t|X_{t-1}) + (1-p)E(\alpha * (X_{t-1}\varepsilon_t) + \varepsilon_t|X_{t-1})\\
&= p(\alpha X_{t-1} + E(\varepsilon_t)) + (1-p)(\alpha E(X_{t-1}\varepsilon_t|X_{t-1}) + E(\varepsilon_t))\\
&= \alpha(p + (1-p)E(\varepsilon_t))X_{t-1} + E(\varepsilon_t).
\end{aligned}$$

Finally, we focus on the conditional probability mass function, where we focus on the one-step-ahead conditional probability.

$$
\begin{aligned}
P(X_t = x|X_{t-1} = u) &= pP(\alpha * X_{t-1} + \varepsilon_t = x|X_{t-1} = u) \\
&\quad + (1-p)P(\alpha * (X_{t-1}\varepsilon_t) + \varepsilon_t = x|X_{t-1} = u) \\
&= p\sum_{i=0}^{x} P(\alpha * X_{t-1} = i|X_{t-1} = u)P(\varepsilon_t = x-i) \\
&\quad + (1-p)\sum_{i=0}^{x} P(\alpha * (X_{t-1}\varepsilon_t) = i|X_{t-1} = u, \varepsilon_t = x-i)P(\varepsilon_t = x-i) \\
&= p\sum_{i=0}^{x} P(\alpha * u = i)P(\varepsilon_t = x-i) \\
&\quad + (1-p)\sum_{i=0}^{x} P(\alpha * (u(x-i)) = i)P(\varepsilon_t = x-i) \\
&= p\sum_{i=0}^{x} P(N = i)P(\varepsilon_t = x-i) \\
&\quad + (1-p)\sum_{i=0}^{x} P(M = i)P(\varepsilon_t = x-i), \qquad (3.4)
\end{aligned}
$$

where N and M are random variables with negative binomial distribution with parameters (α, u) and $(\alpha, u(x-i))$, respectively.

3.1. Specification of the innovation process

So far, we have not specified the marginal distribution of the innovation process $\{\varepsilon_t\}$. And as we could notice, that didn't affect the derivation of the MNLINAR(1) model properties. In order to complete the definition of the MNLINAR(1) model, we introduce the assumption about the distribution of ε_t. In the succeeding sections, we assume that ε_t follows the geometric distribution with parameter $\lambda/(1+\lambda)$. The corresponding probability mass function is equal to $P(\varepsilon_t = k) = \frac{\lambda^k}{(1+\lambda)^{k+1}}$. Notice that this model can be easily adjusted for a different type of series by introducing different distributions of the innovation process.

4. Parameter estimation

In this section, we propose two methods for the estimation of unknown parameters of the MNLINAR(1) model. First, we discuss in detail the method of moments, and then the conditional maximum likelihood method. At the end, we test the efficiency of the presented methods on simulated data sets.

4.1. Method of moments

Assume that we have a realization of the series given by the equation (2.1) of length N. Then, for the given series $\{X_1, X_2, \ldots, X_N\}$, first sample moment is denoted as $\overline{X}_N$, the second sample moment as $\overline{X^2}_N$ and $E(X_k X_{k-1})$ as γ.

Since we have that $E(\varepsilon_t) = \lambda$ from equation (3.1) we obtain the estimate parameter λ as

$$\lambda = \frac{\overline{X}_N(1-\alpha p)}{1+\alpha(1-p)\overline{X}_N}. \tag{4.1}$$

Then, we can easily solve equation (3.3) for λ, since it is a linear equation with respect to λ.

$$\lambda = \frac{\gamma - \alpha p \overline{X^2}_N}{\alpha(1-p)\overline{X^2}_N + \overline{X}_N}. \tag{4.2}$$

The left sides of equations (4.1) and (4.2) are equal, so it follows that

$$\frac{\overline{X}_N(1-\alpha p)}{1+\alpha(1-p)\overline{X}_N} = \frac{\gamma - \alpha p \overline{X^2}_N}{\alpha(1-p)\overline{X^2}_N + \overline{X}_N}.$$

After some algebraic transformations, we can solve the above equation for α, where we obtain

$$\begin{aligned} \alpha &= \frac{\gamma - \left(\overline{X}_N\right)^2}{(1-p)(\overline{X^2}_N - \gamma)\overline{X}_N + p(\overline{X^2}_N - \left(\overline{X}_N\right)^2)} \\ &= \frac{C_x}{(1-p)(\overline{X^2}_N - \gamma)\overline{X}_N + pD_x}. \end{aligned} \tag{4.3}$$

where C_x is the sample lag-one covariance, and D_x is the sample variance. Further, since the equation (3.2) is liner with respect to p, the estimate of parameter p we obtain from equation (3.2) as

$$p = \frac{\overline{X^2}_N - \alpha^2 \overline{X^2}_N E(\varepsilon_t^2) - \alpha(1+\alpha)\lambda\overline{X}_N - 2\alpha\overline{X}_N E(\varepsilon_t^2) - E(\varepsilon_t^2)}{\alpha^2\overline{X^2}_N(1-E(\varepsilon_t^2)) + \alpha\overline{X}_N[(1+\alpha)(1-E(\varepsilon_t) + 2(E(\varepsilon_t) - E(\varepsilon_t^2)]} \tag{4.4}$$

Note that under the assumption introduced in Subsection 3.1. we have $E(\varepsilon_t^2) = \lambda(2\lambda+1)$.

The system of equations (4.1), (4.3) and (4.4) cannot be solved analytically. Thus, we apply the following numerical procedure. For a given p_0, we can calculate α_0 from equation (4.3), and then with these two values we get λ_0 from equation (4.1). From equation (4.4) we obtain p_1. We repeat the procedure until $|p_{k+1} - p_k| + |\lambda_{k+1} - \lambda_k| + |\alpha_{k+1} - \alpha_k| < \delta$, where δ is set to be a sufficiently small value.

4.2. Conditional maximum likelihood

For the given series $\{X_1, X_2, \dots, X_N\}$, we estimate the parameters of the MNLINAR(1) model using the conditional maximum likelihood method (CML). The likelihood function that we maximize here, is actually the log-likelihood function determined through equation (3.4). Let us denote the set of parameters of the MNLINAR(1) model as vector $\boldsymbol{\theta} = (\alpha, p, \lambda)$. Then, the estimate of the vector $\boldsymbol{\theta}$ is obtained as

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}), \tag{4.5}$$

where

$$\begin{aligned} L(\boldsymbol{\theta}) &= \sum_{i=2}^{N} \ln P(X_i = x_i | X_{i-1} = x_{i-1}) \\ &= \sum_{i=2}^{N} \ln \Bigg[\frac{p\lambda^{x_i}}{(1+\alpha)^{x_{i-1}}(1+\lambda)^{x_i+1}} \sum_{j=0}^{x_i} \binom{j + x_{i-1} - 1}{x_{i-1} - 1} \left(\frac{\alpha(1+\lambda)}{\lambda(1+\alpha)} \right)^j \\ &+ \frac{(1-p)\lambda^{x_i}}{(1+\alpha)^{x_{i-1}x_i}(1+\lambda)^{x_i+1}} \sum_{j=0}^{x_i} \binom{j + x_{i-1}(x_i - j) - 1}{x_{i-1}(x_i - j) - 1} \left(\frac{\alpha(1+\lambda)}{\lambda(1+\alpha)^{1-x_{i-1}}} \right)^j \Bigg]. \end{aligned}$$

Since this maximization procedure cannot be done analytically, some numerical approach must be applied. For that purpose, we use built-in functions of the program language `R`.

5. Simulation

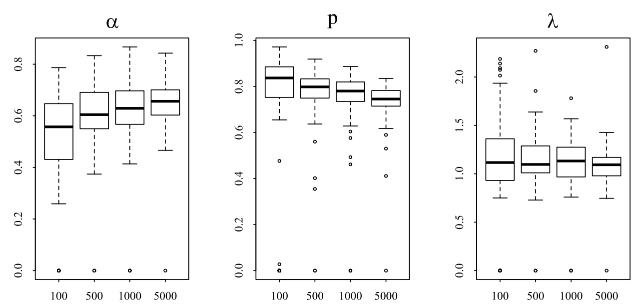
In this section, by using the Monte Carlo method, we generate time series according to equation (2.1). We conduct this procedure using different sets of parameters that figure in the MNLINAR(1) model. On these simulated series we test the efficiency of the MM and CML methods described in the previous section. The efficiency of the proposed methods is measured with respect to the bias and the standard deviation of the obtained estimates.

We have chosen four sets of parameters, considering conditions of Theorem 2.1. The following parameter were used for the simulation purpose: a) $\alpha = 0.7$, $p = 0.7$, $\lambda = 1$; b) $\alpha = 0.3$, $p = 0.3$, $\lambda = 2$; c) $\alpha = 0.5$, $p = 0.9$, $\lambda = 3$; d) $\alpha = 0.1$, $p = 0.9$, $\lambda = 7$. The estimates obtained by the MM and CML methods are given in Table 8.1 (the table can be found in Appendix).

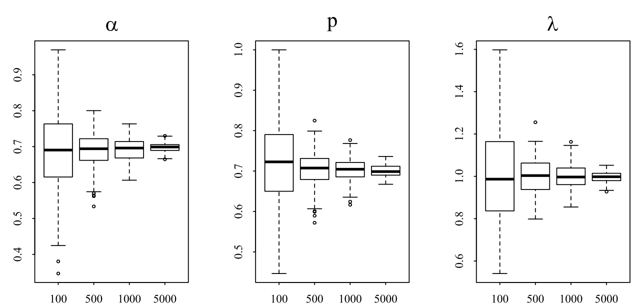
According to the results presented in Table 8.1, we can conclude that both methods converge to the true value of parameters. Also, the standard error of estimates is reducing with the increase of the sample size. It should be noticed that the MM method is not very accurate estimates when the length of a sample is 100 and even 500. But for samples whose length is 1000 or 5000, the estimates are quite adequate. On the other side, CML demonstrates remarkable precision even for samples of length 100.

The MM method is conducted through the iterative procedure described in Subsection 4.1. The maximal number of iterations was set to be 100, and the estimation procedure is very fast. The method usually converges in less than 100 iterations.

The CML method is based on the numerical maximization of the function given by equation (4.5). The numerical procedure is obtained using `nlm` function of the programming language `R`. It doesn't take too much of computation time except for the samples of length 5000.



(a) The method of moments estimates.



(b) The conditional maximum likelihood method estimates.

FIG. 5.1: The box plots of estimates for the set of parameters $\alpha = 0.7$, $p = 0.7$, $\lambda = 1$, obtained by the method of moments (upper) and the conditional maximum likelihood method (lower).

For the MM method, approximately, one of ten estimates is outside the feasible range. On the other side, the CML method had only a few estimates outside the feasible range, and only for the case when the length of the series was 100. The distribution of the estimates for the parameter set a) is given in Figure 5.1. Also, from Figure 5.1 we can notice the convergence of the estimates toward the true values.

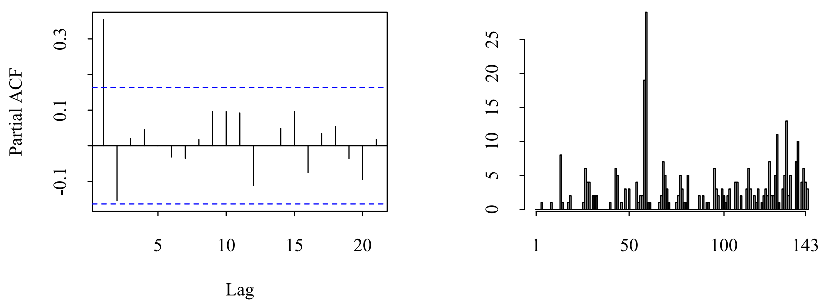


FIG. 6.1: The partial autocorrelation function and the bar plot for DRUG series.

6. Real data example

In this section, we will demonstrate a practical aspect of the MNLINAR(1) model. We test the ability of this model to capture and predict values of an observed time series. Our goal is not to compare the MNLINAR(1) with all known models, but to see what is the effect of having a model with the linear as well as the nonlinear component, in comparison with the models that have only linear (named LINAR(1) model which is actually the model presented in [13]) or only nonlinear component (named NLINAR(1) model). The criteria for the goodness-of-fit are going to be the root mean square error (RMS), the Akaike information criterion and the Bayesian information criterion (BIC).

For this test, we use time series of criminal records, collected by the Pittsburgh police station number 2206. The data can be found on the link `http://www.forecastingprinciples.com/`. We focus on series of monthly drug offenses (DRUG) that took place between January 1990 and December 2001. There are 144 observed values, whose mean value is 2.1 and the standard deviation 12.9. The bar plot of the series is presented in Figure 6.1. In Figure 6.1 there is also the partial autocorrelation diagram. Although the MNLINAR(1) model is not a standard autocorrelated model, it has some properties of the autocorrelated model of order one. Figure 6.1 shows that the observed series is autocorrelated on lag one.

The results obtained from the three tested models are presented in Table 6.1. As we can see that by introducing the linear and the nonlinear component, we have reduced the one-step ahead prediction error, while also reduced the values of AIC and BIC measures. Having in mind that the MNLINAR(1) model has one more parameter than the other two models, and considering values of AIC and BIC, we can conclude that the best fit of the observed series is provided by the MNLINAR(1) model.

Table 6.1: Estimated parameters, standard errors of the estimates, the root mean square error for one step ahead prediction, AIC and BIC values for MNLINAR(1) , LINAR(1) and NLINAR(1) models.

Model	Estimates	RMS	AIC	BIC
MNLINAR(1)	$\hat{\alpha} = 0.174$ (0.068) $\hat{p} = 0.395$ (0.204) $\hat{\lambda} = 1.49$ (0.211)	3.37	548.17	557.08
LINAR(1)	$\hat{\alpha} = 0.044$ (0.037) $\hat{\lambda} = 2.033$ (0.343)	3.52	565.06	573.97
NLINAR(1)	$\hat{\alpha} = 0.108$ (0.075) $\hat{\lambda} = 1.615$ (0.192)	3.42	551.15	560.06

7. Conclusion

The model discussed in this paper is the INAR model of order one. Although it is not a pure autoregressive model, it still preserves some of the autoregressive properties. The survival component is composed of linear and nonlinear processes, both defined through the negative binomial thinning operator, while the innovation component is driven by the geometrical marginal distribution. The method of moments and the conditional maximum likelihood method are presented for the estimation of the model parameters. While the method of moments showed to be unreliable for small samples, the conditional maximum likelihood provides very accurate estimates for all testes samples. The practicality of the model was discussed on a real data set, where the surplus of having both linear and nonlinear components was demonstrated.

Some further modifications of the model can be based on choosing different thinning operators or different marginal distribution of the innovation process. Both components of the model can be adjusted in order to better model an observed series.

REFERENCES

1. Al-Osh, M.A., Alzaid, A.A. (1987) First-order integer-valued autoregressive (INAR (1)) process, Journal of Time Series Analysis 8, 261–275.
2. Alzaid, A., Al-Osh, M. (1988) First-order integer-valued autoregressive (INAR (1)) process: distributional and regression properties, Statistica Neerlandica 42, 53–61.
3. Doukhan, P., Latour, A., Oraichi, D. (2006) A simple integer-valued bilinear time series model, Advances in Applied Probability 38, 559–578.
4. Laketa, P.N., Nastić, A.S. (2019) Conditional least squares estimation of the parameters of higher order Random environment INAR models, Facta Universitatis, Series: Mathematics and Informatics 34, 525–535.
5. Nastić, A.S., Laketa, P.N., Ristić, M.M. (2016) Random environment integer-valued autoregressive process, Journal of Time Series Analysis 37, 267–287.

6. Nastić, A.S., Laketa, P.N., Ristić, M.M. (2019) Random environment INAR models of higher order, REVSTAT-Statistical Journal 17, 35–65.

7. Pedeli, X., Karlis, D. (2013) Some properties of multivariate INAR (1) processes, Computational Statistics & Data Analysis 67, 213–225.

8. Popović, P. (2015) Random coefficient bivariate INAR (1) process, Facta Universitatis, Series: Mathematics and Informatics 30, 263–280.

9. Popović, P.M., Ristić, M.M., Nastić, A.S. (2016) A geometric bivariate time series with different marginal parameters, Statistical Papers 57, 731–753.

10. Popović, P.M., Bakouch, H.S. (2020) A bivariate integer-valued bilinear autoregressive model with random coefficients, Statistical Papers 61, 1819–1840.

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

12. Qi, X., Li, Q., 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.

13. Ristić, M.M., Bakouch, H.S., Nastić, A.S. (2009) A new geometric first-order integer-valued autoregressive (NGINAR (1)) process, Journal of Statistical Planning and Inference 139, 2218–2226.

14. Scotto, M.G., Weiß, C.H., Gouveia, S. (2015) Thinning-based models in the analysis of integer-valued time series: a review, Statistical Modelling 15, 590–618.

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

16. Weiß, C.H. (2008) Thinning operations for modeling time series of counts—a survey, AStA Advances in Statistical Analysis 92, 319–341.

8. Appendix

Table 8.1: The bias and the standard errors of the estimates obtained by the method of moments and the conditional maximum likelihood method.

	MM			CML		
N	α	p	λ	α	p	λ
a) $\alpha = 0.7$, $p = 0.7$, $\lambda = 1$						
100	-0.039	-0.195	-0.168	-0.007	-0.004	0.004
	0.141	0.209	0.355	0.094	0.151	0.307
500	-0.069	-0.127	-0.156	-0.002	0.018	0.017
	0.127	0.157	0.312	0.035	0.084	0.125
1000	-0.057	-0.154	-0.10	-0.003	0.007	0.012
	0.121	0.183	0.187	0.026	0.062	0.094
5000	-0.034	-0.053	-0.07	-0.001	-0.002	0.008
	0.104	0.164	0.048	0.01	0.025	0.039
b) $\alpha = 0.3$, $p = 0.3$, $\lambda = 2$						
100	0.159	-0.134	-0.189	0.035	-0.021	-0.032
	0.14	0.118	0.548	0.114	0.127	0.24
500	0.087	-0.088	-0.163	0.011	0.005	-0.007
	0.112	0.075	0.218	0.05	0.05	0.091
1000	0.058	-0.07	-0.111	0.006	0.004	-0.005
	0.103	0.087	0.202	0.03	0.032	0.061
5000	0.034	-0.045	-0.07	0.003	0.002	0.007
	0.08	0.051	0.153	0.01	0.022	0.011
c) $\alpha = 0.5$, $p = 0.9$, $\lambda = 3$						
100	0.027	-0.063	-0.634	-0.002	0.023	-0.113
	0.098	0.033	0.973	0.048	0.238	1.279
500	0.032	-0.047	-0.533	0.001	0.002	-0.094
	0.066	0.03	0.545	0.019	0.087	0.498
1000	0.017	-0.044	-0.366	0.001	-0.002	-0.081
	0.05	0.023	0.289	0.013	0.055	0.324
5000	0.014	-0.031	-0.259	-0.002	-0.007	0.005
	0.04	0.02	0.265	0.02	0.021	0.019
d) $\alpha = 0.1$, $p = 0.9$, $\lambda = 7$						
100	-0.033	-0.063	0.185	-0.003	0.001	-0.065
	0.062	0.034	0.87	0.07	0.069	0.58
500	-0.023	-0.034	0.103	0.001	0.003	-0.026
	0.048	0.033	0.477	0.029	0.025	0.241
1000	-0.013	-0.028	0.066	-0.001	-0.001	-0.011
	0.04	0.036	0.423	0.022	0.02	0.181
5000	-0.001	-0.008	-0.018	-0.001	-0.001	0.007
	0.02	0.025	0.231	0.019	0.011	0.059