SORT 43 (2) July-December 2019, 355-384 DOI: 10.2436/20.8080.02.92

Forecasting with two generalized integer-valued autoregressive processes of order one in the mutual random environment

Predrag M. Popović[1], Petra N. Laketa[2] and Aleksandar S. Nastić[3]

Abstract

In this article, we consider two univariate random environment integer-valued autoregressive processes driven by the same hidden process. A model of this kind is capable of describing two correlated non-stationary counting time series using its marginal variable parameter values. The properties of the model are presented. Some parameter estimators are described and implemented on the simulated time series. The introduction of this bivariate integer-valued autoregressive model with a random environment is justified at the end of the paper, where its real-life data-fitting performance was checked and compared to some other appropriate models. The forecasting properties of the model are tested on a few data sets, and forecasting errors are discussed through the residual analysis of the components that comprise the model.

MSC: 62M10.

Keywords: INAR, negative binomial thinning, random states, time series of count, non-stationary process.

1 Introduction

After many scientific proposals of possible models of counting processes in the last decades of the 20th century, so far the best results have been obtained by the thinning-based integer-valued autoregressive models of order one (INAR(1)) introduced almost simultaneously by McKenzie (1985) and Al-Osh and Alzaid (1987). For the first time, they used an idea of defining the deterministic part of the counting process in a certain moment, designated by X_n, for the given $X_{n-1} = x_{n-1}$, using the random sum of x_{n-1} independent and identically distributed (i.i.d.) Bernoulli variables. Precisely,

$$X_n = \sum_{i=1}^{x_{n-1}} v_i + \varepsilon_n,$$

[1] popovicpredrag@yahoo.com, University of Niš, Faculty of Civil Engineering and Architecture, Serbia.

[2] petra.laketa@pmf.edu.rs, University of Niš, Faculty of Sciences and Mathematics, Serbia.

[3] anastic78@gmail.com, University of Niš, Faculty of Sciences and Mathematics, Serbia.

Received: April 2019

Accepted: November 2019

where $\{v_i\}$ is a counting sequence of i.i.d. Bernoulli random variables and ε_n are the innovation process. It is assumed that ε_n and X_{n-k} are independent for all $k > 0$. The INAR model of order one (INAR(1)) can be expressed such:

$$X_n = \alpha \circ x_{n-1} + \varepsilon_n.$$

where the operator $\alpha \circ X_{n-1} | X_{n-1} = x_{n-1}$ is equal to $\alpha \circ x_{t-1} = \sum_{i=1}^{x_{n-1}} v_i$, and it is called the binomial thinning or binomial subsampling operator. The first addend of the model above can be interpreted as a survival process. Therefore, these kinds of processes were ideal for modeling counts generated by limited surviving entities. During the adaptation of this INAR model to many different counting time series, many modifications and generalizations were done. Some researchers were focused on the thinning operator, and their innovative results can be found in Aly and Bouzar (1994), Latour (1998), Zheng, Basawa and Datta (2006, 2007), Zhu and Joe (2006), Ristić, Bakouch and Nastić (2009) and Zhu and Joe (2010). Even though Al-Osh and Aly (1992) as well as Alzaid and Al-Osh (1993) focused on the marginal distribution of the process, other authors preferred to concentrate on the distribution of the innovations, like Jazi, Jones and Lai (2012a, 2012b), Fernández-Fontelo, Fontdecaba and Puig (2017). Also, a certain modification of the innovation process was studied recently in Qi, Li and Zhu (2019).

Later, more attention was paid to the correlation characteristics of the observed processes, i.e. the additional assumptions about the dependence in the counting sequence were introduced. Initial results on the INAR models based on the thinning operator defined using dependent counting sequences were given by Ristić, Nastić and Miletić Ilić (2013). Also, the possibility of serially dependent innovations of the INAR model was studied and well-presented in Weiß (2015). In addition, Weiß, Homburg and Puig (2019) considered testing for zero inflation and overdispersion in INAR(1) models.

Parameter-driven models provided another approach to modeling counting processes. A good insight into these models can be found in Fokianos (2011) and some recent progress is presented in Chakraborty and Bhati (2016) (see also Chakraborty and Bhati, 2017) and Rydén (2017).

In addition to all the preceding models and given aspects of counting processes construction, there were many other approaches which resulted, especially in the last decade, with significant number of papers covering this area of time series research. Although, a great majority of them referred to the problems of modeling stationary processes, in the past years, some authors have been working to accommodate potential patterns of trend and seasonality in INAR models. Significant results in this area can be found in Moriña et al. (2011), Fernández-Fontelo et al. (2017).

Since non-stationarity may be noted in many real life situations, inspired by the work of Tang and Wang (2014), and in order to provide more efficient INAR modeling, a new random environment INAR process of order one (INAR(1) with variable marginal distribution was introduced in Nastić, Laketa and Ristić (2016). This model was non-stationary, which made it more applicable to counting processes in practice. The same

authors also presented a higher-order r-states random environment non-stationary INAR model, which can be found in Nastić, Laketa and Ristić (2019) and Laketa, Nastić and Ristić (2018). Fernández-Fontelo et al. (2016) gave an under-reported data analysis with INAR-hidden Markov chains. However, in the matter of modelling two correlated simultaneous integer-valued series, significant results were achieved by introducing bivariate INAR models which can be found in Pedeli and Karlis (2011), Ristić et al. (2012) and Popović, Ristić and Nastić (2016). The first model is based on the binomial thinning operator, and the dependence between time series was introduced through the innovation processes. The second model is based on the negative binomial thinning operator, considering geometric marginal distribution with the same mean parameters. The last model also has a geometric marginal distribution but assuming different mean parameters. Besides, while in the first model, the dependence between the series is considered in the innovation process, in the last two models, this dependence is considered in the survival components, i.e. the components defined through the thinning operator.

In this article, we focus on the bivariate random environment INAR model which is composed of the two univariate models discussed in Laketa et al. (2018). The two univariate series follow the same hidden process which determines the states of the observed processes. Thus, simultaneously with the observed process we have a Markov process $\{Z_n\}$, with a finite state space $E_r = \{1, 2, \ldots, r\}$, called the random state process. Its realized values z_n define marginal distribution parameter values. So, since each value from E_r corresponds to one state of the process environment, then the marginal distribution is directly dependent on the possible random states of the observed process environment. This can be found in nature every time we consider two random variables in the same circumstances. These variables do not have to be correlated directly, but only through their distributions which depend on the same conditions, i.e. random states. Also, considering such a bivariate model, we present its forecasting properties by conducting the residual analysis of its univariate components.

Like all random environment INAR models, the model proposed here is good for the data which are non-stationary (to be precise, they are part by part stationary), where we can suppose that the conditions in which they are measured can change and affect the measured values. So, this model is better than the other bivariate models for such data.

In the second section of this article, we give a short review of random environment INAR models. Then, in the following section, we introduce the corresponding bivariate model based on the realizations of the random environment process. Section 4 is mainly devoted to moment-based estimators. Also, a brief construction of the likelihood-based estimator is given. Section 5 deals with the residual analysis of the model. The quality of defined estimators is confirmed using simulated series of different sizes, presented in Section 6. The next section contains some real-life examples of the application of the introduced model to certain counting processes, where the model performance is compared to some other competitive bivariate INAR models. Also, the errors produced by one-step-ahead forecasting are analysed. Finally, all the proofs of the theorems are given in the Appendix.

2 A short review of random environment INAR models

The first random environment integer-valued autoregressive model was introduced in Nastić et al. (2016), and that is the random environment INAR(1) model. It is based on the random environment process, which represents the conditions of the environment in which the counting process is observed. Also, the corresponding process $\{Z_n\}$ is said to be an r-state random environment process if it is a Markov chain of order one and takes values from the set $E_r = \{1,2,\ldots,r\}$. The main assumption of the observed process is that the environment conditions have an effect on its marginal distribution. Thus, the r-state random environment INAR(1) process with the determined geometric marginal distribution, based on the negative binomial thinning operator (RrNGINAR(1)), is given by the equation

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

where $\{z_n\}$ is a realization of the process $\{Z_n\}$. The notation $X_n(z_n)$ is used to emphasize the fact that the distribution of X_n depends on z_n. The value z_n determines the value μ_{z_n} from the supposed set of marginal parameter values $\{\mu_1, \mu_1, \ldots, \mu_r\}$ where, $X_n(z_n)$ has the geometric distribution with the expectation μ_{z_n}, since we supposed that its probability mass function (pmf) is defined as

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

Here we gave an explanation on how the observable component X_n of the process depends on its latent component z_n. In addition, the denotation $\alpha*$ stands for the negative binomial thinning operator, which is defined by

$$\alpha * X = \sum_{i=1}^{X} U_i,$$

for an integer-valued random variable X, where $\alpha \in (0,1)$ and $\{U_i\}$, $i \in \mathbb{N}$, is a sequence of i.i.d. random variables with pmf given by

$$P(U_i = u) = \frac{\alpha^u}{(1+\alpha)^{u+1}},\ u \in \mathbb{N}_0.$$

In Laketa et al. (2018), this (RrNGINAR(1)) model is generalized, assuming that the realized random environment sequence $\{z_n\}$ determines not only the marginal distribution of the model, but also the order of the process and the thinning parameter value. In order to accurately present the models from Laketa et al. (2018), the following sets should be previously introduced: the set $\mathscr{M} = \{\mu_1, \mu_2, \ldots, \mu_r\}$ which consists of the possible mean values of the process in the corresponding states, the set $\mathscr{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ containing possible values of the thinning parameters corresponding to different states, and the set $\mathscr{P} = \{p_1, p_2, \ldots, p_r\}$ considering the order of the process. For example, when $z_n = i$,

the RrNGINAR(1) model is in its i-th state, and this means that the counting process is observed in the i-th environment state. Additionally, the model parameters in the i-th state are μ_i, α_i and p_i. In fact, in Laketa et al. (2018) two different RrNGINAR are introduced: RrNGINAR$_{max}(\mathscr{M},\mathscr{A},\mathscr{P})$ and RrNGINAR$_1(\mathscr{M},\mathscr{A},\mathscr{P})$. The set $\mathscr{P}$ contains actually the maximal orders for all states. The difference between these models (the one indexed by max, and other by 1) relies on the way of reaching these maximal orders. Let explain now this in more details, starting from the general form of these two models

$$X_n(z_n)=\begin{cases}\alpha_{z_n}*X_{n-1}(z_{n-1})+\varepsilon_n(z_{n-1},z_n), & \text{w.p. } \phi_{1,P_n}^{(z_n)},\\ \alpha_{z_n}*X_{n-2}(z_{n-2})+\varepsilon_n(z_{n-2},z_n), & \text{w.p. } \phi_{2,P_n}^{(z_n)},\\ \vdots & \vdots\\ \alpha_{z_n}*X_{n-P_n}(z_{n-P_n})+\varepsilon_n(z_{n-P_n},z_n), & \text{w.p. } \phi_{P_n,P_n}^{(z_n)},\end{cases}, \tag{2}$$

where $X_n(z_n)$ has geometric distribution with expectation μ_{z_n}. Since the distribution of the residuals would be complicated to obtain when $P_n=p_{z_n}$, P_n should be defined differently. Thus, for the RrNGINAR$_{max}(\mathscr{M},\mathscr{A},\mathscr{P})$ model, named INAR process with r-states, distribution parameters set $\mathscr{M}$, thinning parameters set $\mathscr{A}$ and maximal order set $\mathscr{P}$, it holds that

$$P_n=\min\{p_{z_n},p_n^*\},$$

$$p_n^*=\max\{i\in\{1,2,\ldots,n\}: z_{n-1}=z_{n-2}=\cdots=z_{n-i}\}.$$

From here, when the state change occurs, $z_n\neq z_{n-1}$, the process order becomes one, i.e. $P_n=1$, and afterwards it starts rising by 1 in every moment of the process, until it reaches its maximum value for that state, which equals p_{z_n}. Then it remains at maximum until the process state changes again. The alternative way, for the other type of the considered model (RrNGINAR$_1(\mathscr{M},\mathscr{A},\mathscr{P})$), is letting the value P_n equal 1 (instead of making it growing gradually), but still considering the value at the same moment at when the previously explained model RrNGINAR$_{max}(\mathscr{M},\mathscr{A},\mathscr{P})$ reaches the maximal order. Accordingly, for the RrNGINAR$_1(\mathscr{M},\mathscr{A},\mathscr{P})$ model, the only possible order values corresponding to the process state i are 1 and p_{z_n}

$$P_n=\begin{cases}p_{z_n}, & p_n^*\geq p_{z_n}\\ 1, & p_n^*<p_{z_n}\end{cases}$$

This model is named the random environment INAR process with r-states, distribution parameters set $\mathscr{M}$, thinning parameters set $\mathscr{A}$ and the order set $\mathscr{P}$.

If, as a special case, it holds that $p_1=p_2=\cdots=p_r=1$, then both models are the same and of order one. Also, the RrNGINAR(1) model is a special case of these two models, when $p_1=p_2=\cdots=p_r=1$ and $\alpha_1=\alpha_2=\cdots=\alpha_r$.

Explaining these two models from Laketa et al. (2018) further, let us now recall the Theorem 1 from that paper, which makes a point about residual distribution (see Appendix for details). Considering the models in Laketa et al. (2018), we should com-

prehend these random environment INAR models as an attempt of fitting counting processes in time-varying conditions, which directly affect to certain parameters of the observed process. As long as the conditions of the process environment do not change, the process itself has the same (and unchanged) value of its latent component z_n. However, when the environment eventually changes (e.g., social circumstances or economic factors), then the random environment INAR models introduced in this paper adapt to these changes. In fact, these models accommodate these environment changes by adequately modifying the values of specific parameters, including even the order of the process. Notice that these models are stationary while they are in the same state z_n, and their non-stationarity starts when changing this state. The latter is a consequence of changing the marginal distribution of the models, the thinning operator value, and the order of the process. When we observe the order, we notice that after the process state is changed from z_{n-1} to z_n, the process order is reduced to 1, which is necessary because of the definition of the model. The way it reaches its value of p_{z_n} depends on the model type (i.e., whether its type is "1" or "max", which was explained earlier). Finally, it should also be emphasized that we consider non-stationary processes, which are "part-by-part stationary", where each "part" corresponds to the period of a random process $\{Z_n\}$ remaining in the same state.

3 Considered models

Now, we focus on the model introduced in this article. Let $\{X_n(z_n)\}$ and $\{Y_n(z_n)\}$ be the RrNGINAR$_1(\mathscr{M}_1,\mathscr{A}_1,\mathscr{P}_1)$ and RrNGINAR$_1(\mathscr{M}_2,\mathscr{A}_2,\mathscr{P}_2)$ processes, respectively, where $\mathscr{M}_1=\{\mu_1,\mu_2,\ldots,\mu_r\}$, $\mathscr{M}_2=\{\nu_1,\nu_2,\ldots,\nu_r\}$, $\mathscr{A}_1=\{\alpha_1,\alpha_2,\ldots,\alpha_r\}$, $\mathscr{A}_2=\{\beta_1,\beta_2,\ldots,\beta_r\}$ and $\mathscr{P}_1=\mathscr{P}_2=\{1\}$. Then, they are defined with the following relations

$$X_n(z_n)=\alpha_{z_n}*X_{n-1}(z_{n-1})+\varepsilon_n(z_{n-1},z_n),\ n\in\mathbb{N}, \tag{3}$$

$$Y_n(z_n)=\beta_{z_n}*Y_{n-1}(z_{n-1})+\eta_n(z_{n-1},z_n),\ n\in\mathbb{N}. \tag{4}$$

In order to give a precise definition of the processes introduced in (3) and (4), we add some additional assumptions:

(C1) $\{\varepsilon_n(1,1)\},\{\varepsilon_n(1,2)\},\ldots,\{\varepsilon_n(r,r)\},\{\eta_n(1,1)\},\{\eta_n(1,2)\},\ldots,\{\eta_n(r,r)\}$ are mutually independent for all $n\in\mathbb{N}_0$,

(C2) $\varepsilon_m(i,j)$ and $\eta_m(i,j)$ are independent of $Y_n(k)$ and $X_n(k)$, respectively, for $n<m$ and for all $i,j,k\in E_r$,

(C3) the covariance between $X_n(z_n)$ and $Y_n(z_n)$ is the same as the covariance between $X_m(z_m)$ and $Y_m(z_m)$, when $z_n=z_m$.

Based on the Theorem 1 from Laketa et al. (2018), the distributions of the innovation series $\{\varepsilon_n\}$ and $\{\eta_n\}$ are given by the following relations:

$$\varepsilon_n(i,j) \stackrel{d}{=} \begin{cases} Geom\left(\frac{\mu_j}{1+\mu_j}\right), & \text{w.p. } 1-\frac{\alpha_j\mu_i}{\mu_j-\alpha_j}, \\ Geom\left(\frac{\alpha_j}{1+\alpha_j}\right), & \text{w.p. } \frac{\alpha_j\mu_i}{\mu_j-\alpha_j}. \end{cases} \tag{5}$$

$$\eta_n(i,j) \stackrel{d}{=} \begin{cases} Geom\left(\frac{\nu_j}{1+\nu_j}\right), & \text{w.p. } 1-\frac{\beta_j\nu_i}{\nu_j-\beta_j}, \\ Geom\left(\frac{\beta_j}{1+\beta_j}\right), & \text{w.p. } \frac{\beta_j\nu_i}{\nu_j-\beta_j}. \end{cases} \tag{6}$$

Now, we will present some new results of considered models. For the simplicity of notation, in the following text, we shall use X_n and Y_n instead of $X_n(z_n)$ and $Y_n(z_n)$, respectively. We will consider $\mathbf{X}_n = (X_n, Y_n)$ as a bivariate process named BRrNGINAR(1). Also, let us define vector $\boldsymbol{\mu}_n = \begin{bmatrix} \mu_{z_n} \\ \nu_{z_n} \end{bmatrix}$ and matrix $\mathbf{A}_n = \begin{bmatrix} \alpha_{z_n} & 0 \\ 0 & \beta_{z_n} \end{bmatrix}$.

The following theorem explains the process correlation structure.

Theorem 1 *(a) The covariance matrix of random variables $\mathbf{X}_n$ and $\mathbf{X}_{n-k}$, $k \in \{0,1,\ldots,n\}$, is given as*

$$Cov(\mathbf{X}_k, \mathbf{X}_0) = \mathbf{A}_1\mathbf{A}_2\ldots\mathbf{A}_k Cov(\mathbf{X}_0, \mathbf{X}_0), \tag{7}$$

(b) The correlation matrix of random variables $\mathbf{X}_n$ and $\mathbf{X}_{n-k}$, $k \in \{0,1,\ldots,n\}$, is given as

$$Corr(\mathbf{X}_k, \mathbf{X}_0) = \begin{bmatrix} \sqrt{\frac{Var(X_0)}{Var(X_k)}} & 0 \\ 0 & \sqrt{\frac{Var(Y_0)}{Var(Y_k)}} \end{bmatrix} \mathbf{A}_1\mathbf{A}_2\ldots\mathbf{A}_k Corr(\mathbf{X}_0, \mathbf{X}_0), \tag{8}$$

where $Var(X_i) = \frac{\mu_{z_i}}{1+\mu_{z_i}}$ and $Var(Y_i) = \frac{\nu_{z_i}}{1+\nu_{z_i}}$.

The proof is given in the Appendix.

Following theorem contains the results of the conditional expectations and variances.

Theorem 2 *a) The conditional expectation of the random variable $\mathbf{X}_{n+k}$ on $\mathbf{X}_n$ is given by*

$$E(\mathbf{X}_{n+k}|\mathbf{X}_n) = \mathbf{A}_{n+1}\mathbf{A}_{n+2}\ldots\mathbf{A}_{n+k}[\mathbf{X}_n - \boldsymbol{\mu}_n] + \boldsymbol{\mu}_{n+k},\ k \in \mathbb{N}_0, \tag{9}$$

b) The conditional variance of X_{n+k} on $\mathbf{X}_n$ is given by

$$\begin{aligned} Var(X_{n+k}|X_n,Y_n) = \Bigg\{ & \alpha_{z_{n+1}}(1+\alpha_{z_{n+1}}) \left(\prod_{s=2}^{k} \alpha_{z_{n+s}}^2 I\{k>1\} + I\{k=1\} \right) \\ & + \sum_{i=2}^{k-1} \left(\prod_{s=1}^{i-1} \alpha_{z_{n+s}} \right) \alpha_{z_{n+i}}(1+\alpha_{z_{n+i}}) \left(\prod_{s=i+1}^{k} \alpha_{z_{n+s}}^2 \right) I\{k>2\} \\ & + \left(\prod_{s=1}^{k} \alpha_{z_{n+s}} \right) \alpha_{z_{n+k}}(1+\alpha_{z_{n+k}}) I\{k>1\} \Bigg\} (X_n - \mu_{z_n}) \\ & + \mu_{z_{n+k}}(1+\mu_{z_{n+k}}) - \left(\prod_{s=1}^{k} \alpha_{z_{n+s}}^2 \right) \mu_{z_n}(1+\mu_{z_n}), \end{aligned}$$

and the conditional variance of Y_{n+k} on $\mathbf{X}_n$ is analogous.

c) The conditional probability mass function is given by

$$\begin{aligned} & P(X_n = x_n, Y_n = y_n | X_{n-1} = x_{n-1}, Y_{n-1} = y_{n-1}, Z_n = z_n, Z_{n-1} = z_{n-1}) \\ & = P(X_n = x_n | X_{n-1} = x_{n-1}, Z_n = z_n, Z_{n-1} = z_{n-1}) \\ & \cdot P(Y_n = y_n | Y_{n-1} = y_{n-1}, Z_n = z_n, Z_{n-1} = z_{n-1}), \end{aligned}$$

where

$$\begin{aligned} & P(X_n = x_n | X_{n-1} = x_{n-1}, Z_n = z_n, Z_{n-1} = z_{n-1}) \\ & = \sum_{k=0}^{x_n} \binom{x_{n-1}+k-1}{x_{n-1}-1} \frac{\alpha_{z_{n-1}}^k}{(1+\alpha_{z_{n-1}})^{k+x_{n-1}}} \\ & \cdot \left[\left(1 - \frac{\alpha_{z_n}\mu_{z_{n-1}}}{\mu_{z_n} - \alpha_{z_n}} \right) \frac{\mu_{z_n}^{x_n-k}}{(1+\mu_{z_n})^{x_n-k+1}} + \frac{\alpha_{z_n}\mu_{z_{n-1}}}{\mu_{z_n} - \alpha_{z_n}} \cdot \frac{\alpha_{z_n}^{x_n-k}}{(1+\alpha_{z_n})^{x_n-k}} \right] I_{\{x_{n-1}\neq 0\}} \\ & + \left[\left(1 - \frac{\alpha_{z_n}\mu_{z_{n-1}}}{\mu_{z_n} - \alpha_{z_n}} \right) \frac{\mu_{z_n}^{x_n}}{(1+\mu_{z_n})^{x_n+1}} + \frac{\alpha_{z_n}\mu_{z_{n-1}}}{\mu_{z_n} - \alpha_{z_n}} \cdot \frac{\alpha_{z_n}^{x_n}}{(1+\alpha_{z_n})^{x_n}} \right] I_{\{x_{n-1}=0\}}, \end{aligned}$$

and the analogous formula holds for Y_n.

The proofs are given in the Appendix.

Remark 1 Regarding the correlation between $\{X_n(z_n)\}$ and $\{Y_n(z_n)\}$, the following can be said. Values of the processes $\{X_n(z_n)\}$ and $\{Y_n(z_n)\}$ are determined by the random process realization. Namely, certain parameter values of one component may only occur with the corresponding parameter values of another component. This explains the correlation between $\{X_n\}$ and $\{Y_n\}$, which cannot be calculated, since it is not a correlation in the standard sense and definition. However, as $\{z_n\}$ is determined by the clustering of the observed counting processes, it is actually this sequence, $\{z_n\}$, that contains in-

formation about this kind of correlation. Beside this, the standard covariance function $Cov(X_n, Y_m)$, for some $m,n \in \mathbb{N}$, can be different from zero, which is in fact used in the section about the Yule-Walker method of estimation of the unknown process parameters.

4 Parameter estimation

Let $X_1, X_2, \ldots, X_N$ and $Y_1, Y_2, \ldots, Y_N$ be samples from the RrNGINAR$_1(\mathscr{M}_1, \mathscr{A}_1, \mathscr{P}_1)$ and RrNGINAR$_1(\mathscr{M}_2, \mathscr{A}_2, \mathscr{P}_2)$ processes, respectively, where $\mathscr{M}_1 = \{\mu_1, \mu_2, \ldots, \mu_r\}$, $\mathscr{M}_2 = \{\nu_1, \nu_2, \ldots, \nu_r\}$, $\mathscr{A}_1 = \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$, $\mathscr{A}_2 = \{\beta_1, \beta_2, \ldots, \beta_r\}$ and $\mathscr{P}_1 = \mathscr{P}_2 = \{1\}$ are the corresponding sets of unknown parameters. In the following subsections, two methods for parameter estimation are given: the Yule-Walker method and the conditional maximum likelihood method.

4.1 Yule-Walker estimation

The Yule-Walker parameter estimators are defined matching theoretical and empirical values of the correlation structure of the model. Recall that, usually the Yule-Walker estimation method (YW) assumes that the process is stationary. Since this assumption does not hold for the models with a random environment, because they have different states, it is necessary to define the Yule-Walker estimators using some parts of the sample, which can be considered stationary.

Let us define the set $I_k = \{i \in \{1,2,\ldots,N\} | z_i = z_{i+1} = k\}$ of indices i of the process elements $X_i(z_i)$ and $Y_i(z_i)$ corresponding to the state k, whose followers $X_{i+1}(z_{i+1})$ and $Y_{i+1}(z_{i+1})$ are also in the same state k and denote its number of elements by $n_k = |I_k|$.

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

$$\widehat{\nu}_k = \frac{1}{n_k}\sum_{i\in I_k} Y_i(k), \quad \widehat{\gamma}_0^{Y,k} = \frac{1}{n_k}\sum_{i\in I_k}(Y_i(k)-\widehat{\nu}_k)^2,$$

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

$$\widehat{\gamma}_1^{Y,k} = \frac{1}{n_k}\sum_{i\in I_k}(Y_{i+1}(k)-\widehat{\nu}_k)(Y_i(k)-\widehat{\nu}_k),$$

$$\widehat{\gamma}_{10}^{X,Y,k} = \frac{1}{n_k}\sum_{i\in I_k}(X_{i+1}(k)-\widehat{\mu}_k)(Y_i(k)-\widehat{\nu}_k),$$

$$\widehat{\gamma}_{01}^{X,Y,k} = \frac{1}{n_k}\sum_{i\in I_k}(Y_{i+1}(k)-\widehat{\nu}_k)(X_i(k)-\widehat{\mu}_k),$$

$$\widehat{\gamma}_{00}^{X,Y,k} = \frac{1}{n_k}\sum_{i\in I_k}(X_i(k)-\widehat{\mu}_k)(Y_i(k)-\widehat{\nu}_k),$$

where

$$\begin{aligned}
&\gamma_1^{X,k} = Cov(X_n, X_{n+1}),\ \ \gamma_1^{Y,k} = Cov(Y_n, Y_{n+1}) \ \text{ if } \ z_n = z_{n+1} = k,\\
&\gamma_0^{X,k} = Var(X_n),\ \ \gamma_0^{Y,k} = Var(Y_n) \ \text{ if } \ z_n = k,\\
&\gamma_{00}^{X,Y,k} = Cov(X_n, Y_n) \ \text{ if } \ z_n = k,\\
&\gamma_{01}^{X,Y,k} = Cov(X_n, Y_{n+1}),\ \ \gamma_{10}^{X,Y,k} = Cov(X_{n+1}, Y_n) \ \text{ if } \ z_n = z_{n+1} = k.
\end{aligned}$$

These estimators are all strongly consistent, which can be shown by a similar proof as in Nastić et al. (2016). From the covariance properties, analysed in the previous section, it follows that

$$\begin{aligned}
Cov(X_{n+1}(z_{n+1}), Y_n(z_n)) &= \alpha_{z_{n+1}} Cov(X_n(z_n), Y_n(z_n)),\\
Cov(X_{n+1}(z_{n+1}), X_n(z_n)) &= \alpha_{z_{n+1}} Cov(X_n(z_n), X_n(z_n)),
\end{aligned}$$

so we can write

$$\alpha_{z_{n+1}} = \frac{1}{2}\left(\frac{Cov(X_{n+1}(z_{n+1}), Y_n(z_n))}{Cov(X_n(z_n), Y_n(z_n))} + \frac{Cov(X_{n+1}(z_{n+1}), X_n(z_n))}{Cov(X_n(z_n), X_n(z_n))}\right).$$

Let us now consider $X_n(z_n)$, such that $n \in I_k$. Then,

$$\alpha_k = \frac{1}{2}\left(\frac{Cov(X_{n+1}(k), Y_n(k))}{Cov(X_n(k), Y_n(k))} + \frac{Cov(X_{n+1}(k), X_n(k))}{Cov(X_n(k), X_n(k))}\right).$$

Therefore, we can estimate α_k in the following way

$$\widehat{\alpha}_k = \frac{1}{2}\left(\frac{\widehat{\gamma}_{10}^{X,Y,k}}{\widehat{\gamma}_{00}^{X,Y,k}} + \frac{\widehat{\gamma}_1^{X,k}}{\widehat{\gamma}_0^{X,k}}\right).$$

Similarly, we get

$$\widehat{\beta}_k = \frac{1}{2}\left(\frac{\widehat{\gamma}_{01}^{X,Y,k}}{\widehat{\gamma}_{00}^{X,Y,k}} + \frac{\widehat{\gamma}_1^{Y,k}}{\widehat{\gamma}_0^{Y,k}}\right).$$

From the consistency of the modified sample covariances follows the consistency of $\widehat{\alpha}_k$ and $\widehat{\beta}_k$.

4.2 Conditional maximum likelihood estimation

We also consider likelihood-based estimation method (CML), where we conduct the maximization of the log-likelihood function for the given sample $\{(X_1(z_1),Y_1(z_1)),\ldots,(X_N(z_N),Y_N(z_N))\}$. The function that needs to be maximized is of the form

$$\log L = \sum_{i=2}^{N} \log P((X_i,Y_i)=(x_i,y_i)|(X_{i-1},Y_{i-1})=(x_{i-1},y_{i-1})).$$

The conditional probability mass function is given by Theorem 2, where values X_0 and Y_0 are treated as known. The maximization procedure is conducted numerically using the `optim` function in the programming language `R`.

5 Analysis of prediction errors

In this section, we give the equations for the analysis of one-step-ahead prediction errors. Since model's prediction is conducted with two processes, survival and innovation, we analyse the prediction errors of these two processes separately. Since these two processes are unobservable, we will discuss their prediction errors in terms of conditional expectations. Namely, knowing the realization of the processes $\{X_n\}$ and $\{Y_n\}$ at the moment n, we calculate the conditional expectations of survival and innovation processes for that moment. This approach was discussed in detail in Freeland and McCabe (2004) for the univariate case and in Popović, Nastić and Ristić (2018) for the bivariate case. Here we use the similar methodology as in Popović et al. (2018). Notice that the survival and the innovation processes are mutually independent for known realization of the process $\{Z_n\}$.

Knowing all states up to moment n, we want to determine $P(\alpha_{z_n} * X_{n-1}(z_{n-1}) = m|X_n = x_n, Y_n = y_n, Z_n = z_n, X_{n-1} = x_{n-1}, Y_{n-1} = y_{n-1}, Z_{n-1} = z_{n-1})$ and $P(\varepsilon_n(z_n,z_{n-1}) = x_n - m|X_n = x_n, Y_n = y_n, Z_n = z_n, X_{n-1} = x_{n-1}, Y_{n-1} = y_{n-1}, Z_{n-1} = z_{n-1})$, and similarly for $\beta_{z_n} * Y_{n-1}(z_{n-1})$ and $\eta_n(z_n,z_{n-1})$. As stated above, we consider a model which is based on a realization of the process $\{Z_n\}$ i.e. $\{z_n\}$. Thus,

$$\begin{aligned} &P(\alpha_{z_n} * X_{n-1}(z_{n-1}) = m|X_n = x_n, Y_n = y_n, Z_n = z_n, X_{n-1} = x_{n-1}, Y_{n-1} = y_{n-1}, Z_{n-1} = z_{n-1}) \\ &= \frac{P(\alpha_{z_n} * X_{n-1}(z_{n-1}) = m, X_n = x_n|Z_n = z_n, X_{n-1} = x_{n-1}, Z_{n-1} = z_{n-1})}{P(X_n = x_n|Z_n = z_n, X_{n-1} = x_{n-1}, Z_{n-1} = z_{n-1})} = f(m), \quad (10) \end{aligned}$$

where we have in mind that the survival component of the process X_n is independent of Y_n for known X_n, X_{n-1}, Z_n and Z_{n-1}. Function $f(m)$ is introduced here for practical reasons. The denominator is given in Theorem 2. Further, we calculate the nominator having in mind the definition of the process X_n, i.e. $X_n = \alpha_{z_n} * X_{n-1}(z_{n-1}) + \varepsilon_n(z_n,z_{n-1})$. Thus, for known X_n and X_{n-1}, the probability $P(\alpha_{z_n} * X_{n-1}(z_{n-1}) = m, X_n = x_n)$ is the

same as $P(\alpha_{z_n} * X_{n-1}(z_{n-1}) = m, \varepsilon_n(z_{n-1}, z_n) = x_n - m)$. According to the definition of the process, X_{n-1} is independent from ε_n (statement (C2)), so the nominator of the above equation is obtained as

$$\begin{aligned} &P(\alpha_{z_n} * X_{n-1}(z_{n-1}) = m, \varepsilon_n(z_{n-1}, z_n) = x_n - m | Z_n = z_n, X_{n-1} = x_{n-1}, Z_{n-1} = z_{n-1}) \\ &= P(NB(x_{n-1}, \alpha_{z_n}) = m) \cdot P(\varepsilon_n(z_{n-1}, z_n) = x_n - m). \end{aligned}$$

$NB(x_{n-1}, \alpha_{z_n})$ stands for a random variable with a negative binomial distribution with stated parameters. The probability of the random variable $\varepsilon_n(z_{n-1}, z_n)$ is given by equation (5) and it is equal to

$$P(\varepsilon_n(z_{n-1}, z_n) = x_n - m) = \left(1 - \frac{\alpha_{z_n}\mu_{z_{n-1}}}{\mu_{z_n} - \alpha_{z_n}}\right) \frac{\mu_{z_n}^{x_n - m}}{(1+\mu_{z_n})^{x_n - m + 1}} + \frac{\alpha_{z_n}\mu_{z_{n-1}}}{\mu_{z_n} - \alpha_{z_n}} \frac{\alpha_{z_n}^{x_n - m}}{(1+\alpha_{z_n})^{x_n - m + 1}}.$$

Further, the conditional distribution of the innovation process can be obtained following computations similar to those presented above for equation (10). Thus we have

$$\begin{aligned} &P(\varepsilon_n(z_{n-1}, z_n) = m | X_n = x_n, Y_n = y_n, Z_n = z_n, X_{n-1} = x_{n-1}, Y_{n-1} = y_{n-1}, Z_{n-1} = z_{n-1}) \\ &= f(x_n - m). \end{aligned} \tag{11}$$

By using equations (10) and (11), we can derive the conditional expectations for the survival and innovation components, respectively. With $\mathscr{F}_n$, we denote the σ-algebra generated with (X_n, Y_n, Z_n), $(X_{n-1}, Y_{n-1}, Z_{n-1})$, $\ldots$, (X_0, Y_0, Z_0). Then, we have that

$$\begin{aligned} E(\alpha_{z_n} * X_{n-1}(z_{n-1}) | \mathscr{F}_n) = &\frac{x_{n-1}\alpha_{z_{n-1}}}{P(X_n = x_n | Z_n = z_n, X_{n-1} = x_{n-1}, Z_{n-1} = z_{n-1})} \\ &\cdot P(X_n = x_n - 1 | Z_n = z_n, X_{n-1} = x_{n-1} + 1, Z_{n-1} = z_{n-1}), \end{aligned} \tag{12}$$

and

$$\begin{aligned} E(\varepsilon_n(z_{n-1}, z_n) | \mathscr{F}_n) = &\frac{1}{P(X_n = x_n | Z_n = z_n, X_{n-1} = x_{n-1}, Z_{n-1} = z_{n-1})} \\ &\cdot \left[x_n P(X_n = x_n | Z_n = z_n, X_{n-1} = x_{n-1}, Z_{n-1} = z_{n-1}) \right. \\ &\left. - x_{n-1}\alpha_{z_{n-1}} P(X_n = x_n - 1 | Z_n = z_n, X_{n-1} = x_{n-1} + 1, Z_{n-1} = z_{n-1})\right]. \end{aligned} \tag{13}$$

The detailed derivations of equations (12) and (13) are given in Appendix. The analogue equations stand for $E(\beta_{z_n} * Y_{n-1}(z_{n-1}) | \mathscr{F}_n)$ and $E(\eta_n(z_n, z_{n-1}) | \mathscr{F}_n)$.

According to equation (9), the one-step-ahead prediction error at moment n, denoted as r_n, is equal to

$$\begin{aligned} r_n &= x_n(z_n) - (\alpha_{z_n} x_{n-1}(z_{n-1}) + \mu_{z_n} - \alpha_{z_n}\mu_{z_{n-1}}) \\ &= E(x_n(z_n)|\mathscr{F}_n) - \alpha_{z_n} x_{n-1}(z_{n-1}) - \mu_{z_n} + \alpha_{z_n}\mu_{z_{n-1}} \\ &= E(\alpha_{z_n} * x_{n-1}(z_{n-1}) + \varepsilon_{z_n}(z_{n-1}, z_n)|\mathscr{F}_n) - \alpha_{z_n} x_{n-1}(z_{n-1}) - \mu_{z_n} + \alpha_{z_n}\mu_{z_{n-1}} \\ &= E(\alpha_{z_n} * x_{n-1}(z_{n-1})|\mathscr{F}_n) - \alpha_{z_n} x_{n-1}(z_{n-1}) + E(\varepsilon_{z_n}(z_{n-1}, z_n)|\mathscr{F}_n) - \mu_{z_n} + \alpha_{z_n}\mu_{z_{n-1}}. \end{aligned}$$

We can conclude that the prediction error can be decomposed into two components. The first one is the prediction error of the survival process $r_{sur} = E(\alpha_{z_n} * X_{n-1}(z_{n-1})|\mathscr{F}_n) - \alpha_{z_n} X_{n-1}(z_{n-1})$ and the second one is the prediction error of the innovation process $r_{inn} = E(\varepsilon_{z_n}(z_{n-1}, z_n)|\mathscr{F}_n) - \mu_{z_n} + \alpha_{z_n}\mu_{z_{n-1}}$.

6 Model simulations

In this section, we test two methods for estimating the parameters of the BRrNGINAR(1) model on simulated data sets. The first method is the conditional maximum likelihood method where the conditional likelihood function can be obtained from Theorem 2, statement c). The second one is the Yule-Walker method presented in Section 4.1.

We simulate 100 samples of lengths 100, 500, 1000 and 5000. Using the Monte Carlo method, we generate a time series that evolves according to equations (3) and (4). The values for ε_n and η_n are picked randomly from the distribution determined by equations (5) and (6), respectively. Further, the values for components $\alpha_{z_n} * X_{n-1}(z_{n-1})$ and $\beta_{z_n} * Y_{n-1}(z_{n-1})$ are random numbers generated from the appropriate negative binomial distribution (where we take $X_0 = \varepsilon_0$ and $Y_0 = \eta_0$ as initial values).

The following parameters were used for the simulation procedure: a) $\alpha_1 = 0.1$, $\alpha_2 = 0.2$, $\beta_1 = 0.15$, $\beta_2 = 0.25$, $\mu_1 = 1$, $\mu_2 = 2$, $\nu_1 = 1$, $\nu_2 = 3$; b) $\alpha_1 = 0.45$, $\alpha_2 = 0.5$, $\beta_1 = 0.55$, $\beta_2 = 0.65$, $\mu_1 = 2$, $\mu_2 = 3$, $\nu_1 = 4$, $\nu_2 = 5$; c) $\alpha_1 = 0.1$, $\alpha_2 = 0.2$, $\alpha_3 = 0.25$, $\beta_1 = 0.15$, $\beta_2 = 0.25$, $\beta_3 = 0.25$, $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 3$, $\nu_1 = 1$, $\nu_2 = 2$, $\nu_3 = 3$; d) $\alpha_1 = 0.35$, $\alpha_2 = 0.4$, $\alpha_3 = 0.4$, $\beta_1 = 0.4$, $\beta_2 = 0.25$, $\beta_3 = 0.35$, $\mu_1 = 2$, $\mu_2 = 3$, $\mu_3 = 4$, $\nu_1 = 3$, $\nu_2 = 4$ $\nu_2 = 5$. These values were chosen according to our experience in testing BRrNGINAR(1) as well as other bivariate models. We tried to determine the sets of parameters that are most likely to be found with real data sets. In all cases, we take into account the appropriate boundaries for the thinning parameters. The random environment processes with 2 and 3 random states are considered. For the cases a) and b), the probability vector of states is $(0.5, 0.5)$, while this vector has values $(0.3, 0.4, 0.3)$ for cases c) and d), so all the states are nearly equally probable. We set the transition probability matrix from state i to state j as $\begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$ for cases a) and b), and $\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$ for cases c) and d). They are chosen in such way that diagonal elements are the biggest in the matrices, so that the corresponding processes stay in the same state long enough. The estimated values obtained with the YW method are presented in Table 2 and Table 4, and with the

CML method in Table 1 and Table 3. Besides the estimated values of the parameters, there are also standard deviations of the estimates.

Table 1: *Estimated values of unknown parameters for the BRrNGINAR(1) model with two states using the conditional maximum likelihood method. The standard errors of estimates are given in the brackets.*

a)	$\alpha_1=0.1$	$\alpha_2=0.2$	$\beta_1=0.15$	$\beta_2=0.25$	$\mu_1=1$	$\mu_2=2$	$\nu_1=1$	$\nu_2=3$
100	0.1149	0.1879	0.1438	0.2118	0.9979	1.9679	0.9589	2.4212
	(0.095)	(0.0959)	(0.078)	(0.080)	(0.145)	(0.260)	(0.129)	(0.524)
500	0.1082	0.1982	0.1439	0.2345	1.0044	2.022	0.9773	2.715
	(0.045)	(0.058)	(0.034)	(0.046)	(0.083)	(0.162)	(0.095)	(0.367)
1000	0.099	0.202	0.1459	0.2423	1.0067	2.0106	0.9905	2.7554
	(0.0337)	(0.0350	(0.024)	(0.036)	(0.067)	(0.123)	(0.071)	(0.327)
5000	0.0988	0.1909	0.1527	0.2441	1.0022	1.9972	0.9982	2.9955
	(0.0138)	(0.014)	(0.012)	(0.015)	(0.022)	(0.042)	(0.028)	(0.051)
b)	$\alpha_1=0.45$	$\alpha_2=0.5$	$\beta_1=0.55$	$\beta_2=0.65$	$\mu_1=2$	$\mu_2=3$	$\nu_1=4$	$\nu_2=5$
100	0.4316	0.4454	0.5283	0.5894	1.989	2.9757	3.9493	4.9333
	(0.057)	(0.064)	(0.064)	(0.064)	(0.159)	(0.212)	(0.248)	(0.444)
500	0.4489	0.4745	0.5416	0.6257	2.0079	2.992	4.0002	4.9824
	(0.041)	(0.030)	(0.040)	(0.032)	(0.103)	(0.127)	(0.133)	(0.141)
1000	0.4516	0.4803	0.5444	0.6344	2.007	2.9871	4.0101	4.9915
	(0.029)	(0.025)	(0.027)	(0.027)	(0.096)	(0.109)	(0.130)	(0.155)
5000	0.4425	0.4877	0.5431	0.6401	1.985	2.988	3.9874	4.9854
	(0.021)	(0.012)	(0.020)	(0.014)	(0.064)	(0.065)	(0.074)	(0.092)

Table 2: *Estimated values of unknown parameters for BRrNGINAR(1) model with two states using Yule-Walker method. The standard errors of estimates are given in the brackets.*

a)	$\alpha_1=0.1$	$\alpha_2=0.2$	$\beta_1=0.15$	$\beta_2=0.25$	$\mu_1=1$	$\mu_2=2$	$\nu_1=1$	$\nu_2=3$
100	0.1681	0.2175	0.1671	0.2017	0.9671	1.9477	1.0168	2.8605
	(0.123)	(0.127)	(0.127)	(0.123)	(0.247)	(0.359)	(0.214)	(0.560)
500	0.1163	0.2004	0.1459	0.2382	1.0035	1.9938	1.0155	2.9531
	(0.075)	(0.079)	(0.073)	(0.074)	(0.097)	(0.186)	(0.104)	(0.262)
1000	0.1061	0.195	0.1516	0.2402	1.0007	1.9983	1.016	3.0034
	(0.057)	(0.058)	(0.048)	(0.055)	(0.065)	(0.119)	(0.077)	(0.221)
5000	0.1009	0.2029	0.1484	0.2529	0.994	1.9956	1.0003	3.0076
	(0.025)	(0.024)	(0.025)	(0.024)	(0.027)	(0.057)	(0.031)	(0.091)
b)	$\alpha_1=0.45$	$\alpha_2=0.5$	$\beta_1=0.55$	$\beta_2=0.65$	$\mu_1=2$	$\mu_2=3$	$\nu_1=4$	$\nu_2=5$
100	0.4149	0.449	0.5316	0.5641	1.9574	2.972	4.0829	5.1293
	(0.169)	(0.169)	(0.178)	(0.159)	(0.513)	(0.739)	(0.930)	(1.279)
500	0.4409	0.4803	0.5359	0.6358	1.9907	2.9747	3.9991	5.0536
	(0.085)	(0.077)	(0.076)	(0.075)	(0.235)	(0.358)	(0.447)	(0.588)
1000	0.4457	0.4877	0.5396	0.6361	1.9744	2.9839	4.0285	5.0532
	(0.064)	(0.057)	(0.054)	(0.053)	(0.163)	(0.245)	(0.288)	(0.398)
5000	0.4472	0.4988	0.5477	0.6486	1.9963	2.9922	3.9944	5.0088
	(0.027)	(0.029)	(0.025)	(0.024)	(0.064)	(0.100)	(0.130)	(0.197)

Table 3: *Estimated values of unknown parameters for the BRrNGINAR(1) model with three states using the conditional maximum likelihood method. The standard errors of estimates are given in the brackets.*

c)	$\alpha_1=0.1$	$\alpha_2=0.2$	$\alpha_3=0.25$	$\beta_1=0.15$	$\beta_2=0.25$	$\beta_3=0.25$	$\mu_1=1$	$\mu_2=2$	$\mu_3=3$	$\nu_1=1$	$\nu_2=2$	$\nu_3=3$
100	0.1424	0.1901	0.2078	0.1648	0.2044	0.2018	1.0127	1.985	2.9899	1.0063	1.9992	3.0107
	(0.067)	(0.052)	(0.052)	(0.060)	(0.049)	(0.048)	(0.097)	(0.129)	(0.081)	(0.057)	(0.046)	(0.115)
500	0.1162	0.1965	0.2231	0.1526	0.2255	0.2179	1.0156	1.9939	2.9887	1.0197	1.9997	2.9795
	(0.039)	(0.033)	(0.031)	(0.036)	(0.030)	(0.032)	(0.060)	(0.068)	(0.101)	(0.060)	(0.070)	(0.097)
1000	0.1048	0.2003	0.2272	0.1441	0.2337	0.2288	1.018	1.9874	2.9805	1.0161	2.0023	2.9915
	(0.031)	(0.030)	(0.024)	(0.035)	(0.022)	(0.023)	(0.040)	(0.059)	(0.083)	(0.042)	(0.050)	(0.075)
5000	0.099	0.1989	0.2403	0.1534	0.2372	0.2365	1.0029	2.0027	2.9992	1.0039	1.9941	3.0019
	(0.012)	(0.018)	(0.015)	(0.013)	(0.011)	(0.014)	(0.017)	(0.029)	(0.033)	(0.021)	(0.031)	(0.037)
d)	$\alpha_1=0.35$	$\alpha_2=0.4$	$\alpha_3=0.4$	$\beta_1=0.4$	$\beta_2=0.25$	$\beta_3=0.35$	$\mu_1=2$	$\mu_2=3$	$\mu_3=4$	$\nu_1=3$	$\nu_2=4$	$\nu_2=5$
100	0.3211	0.343	0.3357	0.3571	0.2844	0.3417	2.0001	3.0044	3.9411	3.0038	4.0015	4.98
	(0.070)	(0.059)	(0.071)	(0.078)	(0.072)	(0.072)	(0.098)	(0.237)	(0.264)	(0.135)	(0.192)	(0.191)
500	0.3422	0.3659	0.3666	0.3827	0.2644	0.3446	2.009	3.0076	3.9788	2.9994	3.9926	4.9844
	(0.034)	(0.036)	(0.036)	(0.043)	(0.039)	(0.043)	(0.100)	(0.100)	(0.174)	(0.139)	(0.221)	(0.227)
1000	0.3471	0.3797	0.3746	0.3965	0.2586	0.3475	1.9973	3.004	3.9739	3.0007	3.9998	4.9766
	(0.029)	(0.025)	(0.028)	(0.037)	(0.033)	(0.035)	(0.051)	(0.097)	(0.131)	(0.080)	(0.09)	(0.123)
5000	0.3489	0.3926	0.3835	0.3961	0.2518	0.3418	2.0004	2.9981	3.9959	3.0011	4.0003	5.0014
	(0.011)	(0.012)	(0.015)	(0.014)	(0.016)	(0.016)	(0.035)	(0.042)	(0.042)	(0.040)	(0.050)	(0.052)

Table 4: *Estimated values of unknown parameters for the BRrNGINAR(1) model with three states using the Yule-Walker method. The standard errors of estimates are given in the brackets.*

c)	$\alpha_1=0.1$	$\alpha_2=0.2$	$\alpha_3=0.25$	$\beta_1=0.15$	$\beta_2=0.25$	$\beta_3=0.25$	$\mu_1=1$	$\mu_2=2$	$\mu_3=3$	$\nu_1=1$	$\nu_2=2$	$\nu_3=3$
100	0.2347	0.3121	0.3808	0.2396	0.3024	0.2697	1.0124	1.973	2.9647	1.0669	1.9626	2.9263
	(0.241)	(0.203)	(0.249)	(0.224)	(0.259)	(0.164)	(0.250)	(0.445)	(0.637)	(0.283)	(0.429)	(0.639)
500	0.1532	0.1906	0.2612	0.1478	0.2529	0.2612	0.9999	1.9868	2.9664	0.998	2.0213	2.9839
	(0.107)	(0.102)	(0.122)	(0.097)	(0.147)	(0.117)	(0.111)	(0.205)	(0.311)	(0.117)	(0.203)	(0.276)
1000	0.1221	0.1816	0.2544	0.1384	0.2597	0.2377	1.0099	1.9862	2.9945	0.9963	2.0264	2.987
	(0.073)	(0.090)	(0.086)	(0.083)	(0.102)	(0.096)	(0.085)	(0.144)	(0.232)	(0.090)	(0.151)	(0.200)
5000	0.0978	0.1949	0.2461	0.1492	0.2534	0.2492	0.9983	1.9849	2.9967	1.0083	2.0047	3.0044
	(0.045)	(0.040)	(0.039)	(0.041)	(0.044)	(0.041)	(0.036)	(0.068)	(0.090)	(0.043)	(0.073)	(0.090)
d)	$\alpha_1=0.35$	$\alpha_2=0.4$	$\alpha_3=0.4$	$\beta_1=0.4$	$\beta_2=0.25$	$\beta_3=0.35$	$\mu_1=2$	$\mu_2=3$	$\mu_3=4$	$\nu_1=3$	$\nu_2=4$	$\nu_2=5$
100	0.35	0.4054	0.3785	0.4183	0.3304	0.3102	1.9593	2.9636	3.8851	3.0181	3.9338	5.0608
	(0.172)	(0.203)	(0.236)	(0.235)	(0.207)	(0.217)	(0.501)	(0.708)	(0.895)	(0.719)	(0.809)	(1.313)
500	0.3356	0.4027	0.3935	0.3819	0.2583	0.3533	2.0347	3.0377	4.0598	2.9842	3.9981	5.0265
	(0.134)	(0.108)	(0.120)	(0.118)	(0.117)	(0.121)	(0.230)	(0.335)	(0.407)	(0.322)	(0.447)	(0.525)
1000	0.3311	0.4026	0.3979	0.386	0.2496	0.355	2.0031	3.0253	4.0108	3.0082	4.0012	5.033
	(0.094)	(0.094)	(0.081)	(0.089)	(0.081)	(0.082)	(0.162)	(0.231)	(0.292)	(0.216)	(0.284)	(0.351)
5000	0.3458	0.4028	0.4004	0.3987	0.2452	0.3486	1.9961	2.9871	4.0118	3.0009	3.9956	4.9814
	(0.043)	(0.041)	(0.039)	(0.040)	(0.038)	(0.037)	(0.068)	(0.110)	(0.135)	(0.102)	(0.126)	(0.154)

From the presented results, we can conclude that the estimates converge to the true values with the growth of the sample length, which is followed by the decrease of the standard deviation of the estimates. We can notice that both methods perform better when the true values of the parameters are larger (this will be important when we discuss the results from the application section). A probable reason for that is that when the parameters take small values, the generated series have a lot of zeros. So, these methods need bigger samples to estimate parameters of such "flattened" series.

The CML method provides good results even for samples of length 100. Also, there is no influence on the estimates with respect to the number of random states. On the other hand, the number of random states has a large impact on YW estimates when the sample length is 100. When there are only two states, YW performs similarly as the CML method. With three states, YW provides quite unprecise estimates for samples of length 100. The reason for that lies in the fact that the correlation functions are calculated on small sub-samples, thus their values are not very reliable. So, we can notice large deviations from the true values in the test c). The estimates are much better when the length of the sample is 500 or larger. The estimates of the parameters μ_i and ν_i, $i \in \{1,2,3\}$ converge very quickly with both methods, regardless of the number of states.

The probability vector of states and the transition probability matrix are estimated regardless of YW and CML methods. The probability vector is estimated by dividing the number of occurrences of a state by the length of the sample, while for the transition probability matrix, the number of transitions from state i to state j is divided by the total number of occurrences of states i. This way, we obtain very precise results for all studied samples, thus we omit a detailed discussion here.

We can conclude that CML is much more reliable for small samples (when the length of series is 100). On the other hand, a disadvantage of the CML method is that CML estimates are obtained numerically, thus the CML method is much more time consuming. The YM method provides estimates quite close to the real values when the sample size is 500 or greater and, since it has the analytical solution, it proves to be a better choice than CML for large samples.

The estimation procedure was conducted by using the Monte Carlo simulation. Thus, for each of 100 sample paths we estimate the model parameters. So, for each of these parameters, we get series of 100 values. The mean values and the sample standard deviations of these series are presented respectively as the estimations and their standard errors in Tables 1-4.

7 Application

This section is devoted to the practical aspect of the model. We test the model on a real data set and compare the results to some other known bivariate models. The comparison is based on the ability of the model to predict a value one step ahead for the observed

time series. The goodness of fit is measured in terms of the root mean square error (RMS). Also, we provide values for the Akaike information criterion (AIC), but since we are focused on the forecasting ability of the model, the main attention is paid to the values of RMS.

Parameters of the model are estimated using the conditional maximum likelihood method. As we have concluded in the previous section, the YW method is not very reliable for samples of length 100. Since the series that we deal with in this section have between 105 and 144 observations, we will only use the CML method for parameter estimation.

We compare the BRrNGINAR(1) model with three other bivariate models. Two of these models are with constant coefficients and dependent innovation processes, where one evolves under the Poisson bivariate distribution (BVPOIBINAR(1) model) and the other evolves under the negative binomial distribution (BVNBIBINAR(1) model). Both models were presented in Pedeli and Karlis (2011). The third model that we use for comparison was presented in Popović et al. (2016), it has random coefficients and independent innovation processes (BVGGINAR(1) model).

We test our model on three data sets. First, we consider the data set that contains two series of different events observed in the same region. Then, we focus on bivariate time series composed of data of the same event, observed in different regions. The third test considers two series of data of the same type of event that evolve in the same environment. In all three cases, we assume that the same factors influence both observed series.

7.1 Different events observed in the same region

First, we will test our model on the same data series as in Popović et al. (2016). These series are monthly counts of robberies (ROBB) and aggravated assaults (AGGASS) from January 1990 to December 2001 (for more details about these time series, see Popović et al. (2016)). The observed series together with their ACF and PACF are given in Figure 1. The bar plots in Figure 1 imply a higher level of activities in the first half compared to the end of the series. The series fluctuate around different means during two periods, so the BRrNGINAR(1) model might be appropriate since it has the ability to capture these changes of the frequency.

The results can be found in Table 5. It can be noticed that RMS for both series is the lowest for the BRrNGINAR(1) model. For the observed series, we have detected two states. According to this conclusion, we define the BRrNGINAR(1) model. For the BRrNGINAR(1) model, the main drawback is the number of parameters, but as we can see the model produces the lowest prediction errors, especially for the AGGASS series, and the lowest AIC value.

Table 5: *Parameter estimates of INAR models, RMS and AIC for ROBB and AGGASS data series.*

Model	CML estimates	RMS ROBB	RMS AGGASS	AIC
BRrNGINAR(1)	$\hat{\alpha}_1 = 0.515(0.008)$, $\hat{\alpha}_2 = 0.568(0.021)$ $\hat{\beta}_1 = 0.259(0.105)$, $\hat{\beta}_2 = 0.37(0.059)$ $\hat{\mu}_1 = 2.388(0.001)$, $\hat{\mu}_2 = 3.205(0.001)$ $\hat{\nu}_1 = 1.117(0.001)$, $\hat{\nu}_2 = 2.018(0.001)$	2.376	1.648	1044.45
BVGGINAR(1)	$\hat{\alpha} = 0.499(0.052)$, $\hat{p} = 0.887(0.12)$, $\hat{a} = 2.877(0.328)$ $\hat{\beta} = 0.281(0.058)$, $\hat{q} = 0.805(0.192)$, $\hat{b} = 1.765(0.187)$	2.496	1.827	1065.83
BVPOIBINAR(1)	$\hat{\alpha}_1 = 0.413(0.042)$, $\hat{\lambda}_1 = 1.664(0.148)$ $\hat{\alpha}_2 = 0.21(0.053)$, $\hat{\lambda}_2 = 1.389(0.128)$, $\hat{\phi} = 0.443(0.107)$	2.541	1.857	1183.76
BVNBIBINAR(1)	$\hat{\alpha}_1 = 0.413(0.046)$, $\hat{\lambda}_1 = 1.665(0.205)$ $\hat{\alpha}_2 = 0.169(0.061)$, $\hat{\lambda}_2 = 1.461(0.182)$, $\hat{\beta} = 0.883(0.176)$	2.541	1.88	1077.39

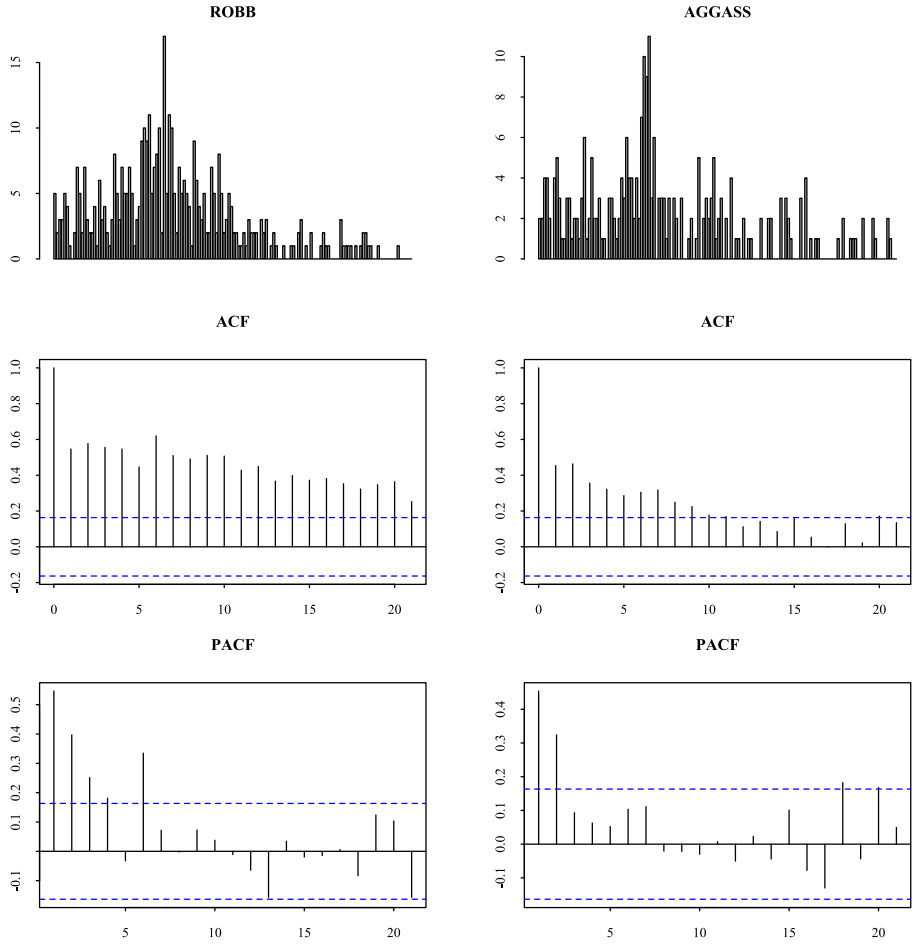

Figure 1: *Bar plots, autocorrelation and partial autocorrelation functions of robberies and aggravated assaults recorded in one police station.*

Table 5 contains the estimated values of model parameters as well as the standard errors of these estimates. Since the estimates are obtained with the CML method, these standard errors are computed as the square root of the diagonal elements of the inverse of the Hessian (the `optim` function from the programming language `R` can return the Hessian). The same holds for Table 6 and Table 7 that are going to be discussed in the next two subsections.

Table 6: *Parameter estimates of INAR models, RMS and AIC for SIMPASS-A and SIMPASS-B data series.*

Model	CML estimates	RMS SIMPASS-A	RMS SIMPASS-B	AIC
BRrNGINAR(1)	$\hat{\alpha}_1 = 0.502(0.131)$, $\hat{\alpha}_2 = 0.507(0.001)$ $\hat{\beta}_1 = 0.52(0.088)$, $\hat{\beta}_2 = 0.63(0.191)$ $\hat{\mu}_1 = 1.768(0.001)$, $\hat{\mu}_2 = 2.485(0.001)$ $\hat{\nu}_1 = 3.877(0.552)$, $\hat{\nu}_2 = 4.52(0.454)$	1.448	2.164	1066.63
BVGGINAR(1)	$\hat{\alpha} = 0.492(0.001)$, $\hat{p} = 0.558(0.075)$, $\hat{a} = 2.076(0.001)$ $\hat{\beta} = 0.65(0.041)$, $\hat{q} = 0.291(0.053)$, $\hat{b} = 2.075(0.001)$	1.612	2.48	1118.61
BVPOIBINAR(1)	$\hat{\alpha}_1 = 0.315(0.065)$, $\hat{\lambda}_1 = 1.544(0.171)$ $\hat{\alpha}_2 = 0.294(0.064)$, $\hat{\lambda}_2 = 2.96(0.303)$, $\hat{\phi} = 0.42(0.201)$	1.588	2.243	1053.74
BVNBIBINAR(1)	$\hat{\alpha}_1 = 0.33(0.067)$, $\hat{\lambda}_1 = 1.512(0.183)$ $\hat{\alpha}_2 = 0.345(0.066)$, $\hat{\lambda}_2 = 2.744(0.319)$, $\hat{\beta} = 0.168(0.065)$	1.584	2.238	1043.91

Table 7: *Parameter estimates of INAR models, RMS and AIC for Bitfinex and Kraken data series.*

Model	CML estimates	RMS Bitfinex	RMS Kraken	AIC
BRrNGINAR(1)	$\hat{\alpha}_1 = 0.819(0.001)$, $\hat{\alpha}_2 = 0.767(0.084)$ $\hat{\beta}_1 = 0.83(0.001)$, $\hat{\beta}_2 = 0.829(0.011)$ $\hat{\mu}_1 = 20.782(1.647)$, $\hat{\mu}_2 = 24.362(1.619)$ $\hat{\nu}_1 = 10.056(1.252)$, $\hat{\nu}_2 = 11.117(0.985)$	9.611	4.142	1333.53
BVGGINAR(1)	$\hat{\alpha} = 0.497(0.012)$, $\hat{p} = 0.783(0.069)$, $\hat{a} = 23.361(0.001)$ $\hat{\beta} = 0.433(0.001)$, $\hat{q} = 0.233(0.054)$, $\hat{b} = 10.138(0.001)$	11.415	4.727	1522.22
BVPOIBINAR(1)	$\hat{\alpha}_1 = 0.515(0.022)$, $\hat{\lambda}_1 = 11.409(0.583)$ $\hat{\alpha}_2 = 0.558(0.037)$, $\hat{\lambda}_2 = 4.529(0.405)$, $\hat{\phi} = 4.465(0.416)$	10.802	4.287	1603.54
BVNBIBINAR(1)	$\hat{\alpha}_1 = 0.611(0.023)$, $\hat{\lambda}_1 = 9.142(1.141)$ $\hat{\alpha}_2 = 0.666(0.026)$, $\hat{\lambda}_2 = 3.432(0.466)$, $\hat{\beta} = 1.198(0.252)$	10.509	4.263	1273.17

7.2 The same event observed in the different regions

The BRrNGINAR(1) model evolves under hidden time series that represents certain states of the observed series. Thus, the observed series are affected by some common factor. To find the most realistic scenario, we will focus on the time series of the same event that took place in different regions. From the database that can be found on website `http://www.forecastingprinciples.com`, we examine the number of

simple assaults recorded in two police stations located in Rochester. These data were recorded from January 1990 to December 2000 in police stations number 36055009401 and 36055009602, so we denote the data series SIMPASS-A and SIMPASS-B, respectively. The mean values of these series are 2.24 and 4.23, while the variances are 3.11 and 5.79, respectively. The correlation coefficient between the two series is 0.29. The autocorrelation coefficients at lag one are 0.45 and 0.36 for SIMPASS-A and SIMPASS-B, respectively.

The bar plots of the observed series are given in Figure 2. We can notice some similar patterns in the evolution of these two series. The bar plots in Figure 2 imply a higher rate of activities at the beginning compared to the end of the observed data. This suggests the existence of two or more random states for the BRrNGINAR(1) model. The BRrNGINAR(1) with two random states shows better performance than the model with three random states in terms of RMS. Models with more than three random states are not adequate for these series, since the observed data set is not long enough to properly estimate all parameters of such models.

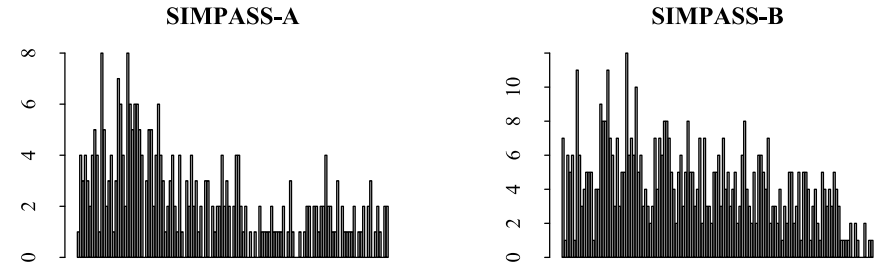

Figure 2: *Bar plots for simple assaults recorded in the two police stations.*

Since the random states have to be the same for both series, they are defined in the following way. The step one is to determine states for each series separately. This procedure is performed by using the quantiles of the observed series. Since we have only two states, we use the median as the boundary for determining states. Then, the states for the BRrNGINAR(1) model are determined as rounded average values of the states from step one for each observed moment. The states are given in Figure 3. In some cases, two observed values of one series have different states although they are equal. This is the consequence of determining random states for both series. But, in spite of this, it can be noticed that observed values are grouped into clusters.

Once again, we will compare the BRrNGINAR(1) model to three bivariate models mentioned above. The results are summarized in Table 6. We can notice that the BRrNGINAR(1) model achieved a much lower RMS for both observed series. Since we examine two time series of the same criminal activity, we can expect that the same factors affect the generation of these series. For example, unemployment or lack of police officers will encourage someone to commit a criminal act such as a simple as-

sault. Our model is based on this assumption, and as such it provides the best results from the forecasting point of view. We can notice that the AIC value is quite close to the BVPOIBINAR(1) and BVNBIBINAR(1) models. Since the BRrNGINAR(1) model depends on a larger number of parameters, it was expected to have a little bit larger value of the AIC.

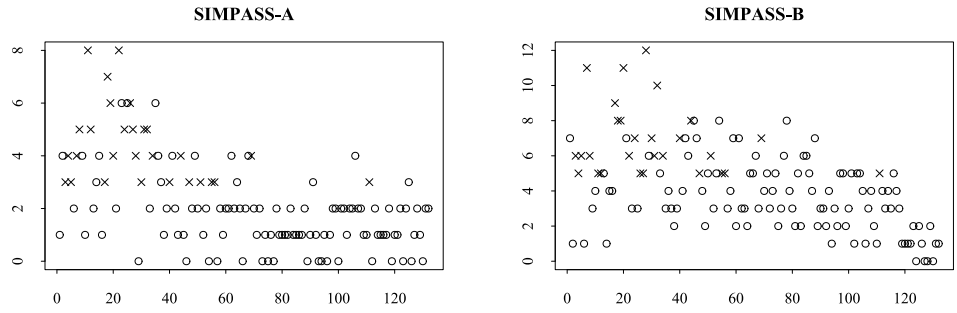

Figure 3: *States for SIMPASS-A and SIMPASS-B data series. The state one is denoted with* $\times$ *and the state two with* $\circ$*.*

For the analysis of the prediction error made by the BRrNGINAR(1) model, we use the approach discussed in Section 5. On the data sets SIMPASS-A and SIMPASS-B, the model makes the root mean square errors of 1.448 and 2.164, respectively. It can be said that these errors are produced by two sources, the prediction of the survival process and the prediction of the innovation process. We measure the prediction error of the survival component as the difference between the value calculated from equation (12) and the first addend of equation (14) when $k = 1$. Similarly, the prediction error of the innovation component is the difference between the value calculated from equation (13) and the second addend of equation (14) when $k = 1$. The residuals are presented in Figure 4.

The black line shows the series of the prediction errors created by the survival component, while the gray one represents the error of the innovation component. The dots are the actual prediction errors that we get when we apply the BRrNGINAR(1) model to the two observed series. It can be noticed that the two components produce errors with the opposite sign. Actually, the correlation coefficient between the two components for SIMPASS-A series is -0.55, and for SIMPASS-B it is -0.75. As a result of these negatively correlated errors, the actual prediction error is reduced. The most obvious consequence of this kind of behaviour can be noticed on the tenth observed value of the SIMPASS-A data set and on the seventh observed values of the SIMPASS-B data set.

It cannot be said that one or the other component produces larger errors. The behaviours of the survival and the innovation processes are quite similar. One of the most probable reasons for this is that they are both driven by the same hidden process which determines the states of the observed series.

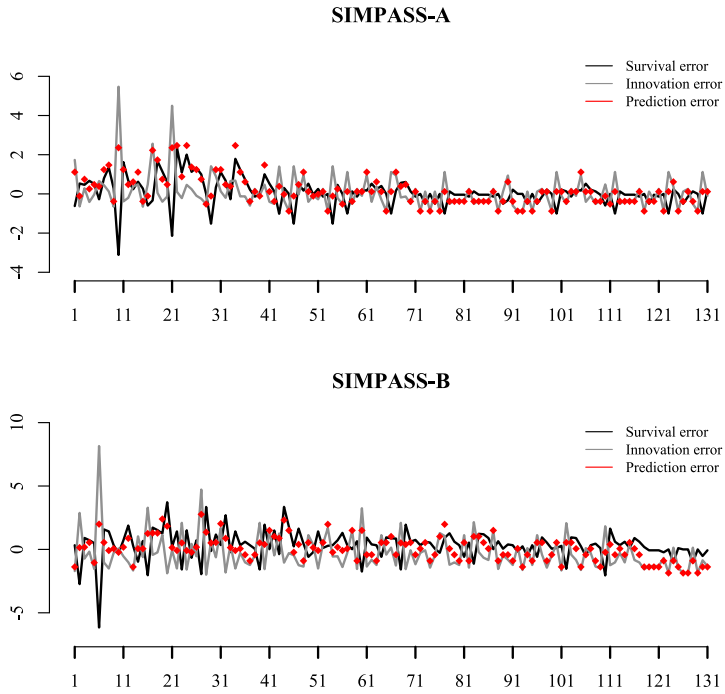

Figure 4: *Prediction errors produced by the survival and the innovation processes.*

7.3 The same type of event in the same environment

In order to more clearly motivate the introduction of the BRrNGINAR(1) model, we give another example where we test the model on two time series that nowadays spark a lot of interest among many people. Namely, we observe the volumes of two cryptocurrencies traded on weekly bases. The data set contains the traded volumes from the beginning of April 2017, until the end of March 2019, for cryptocurrencies Bitfinex and Kraken. We denote the smallest fraction of a coin that can be traded as a unit. Since these cryptocurrencies are traded in vary small fractions, the data that we present here are in 10^{12} units. So, the average values of these series are 23.36 and 10.14, respectively (which are actually 23.36×10^{12} and 10.14×10^{12} units). The standard deviations for the two series are 14.93 and 5.64, respectively, while the correlation between the series is 0.53. The autocorrelation coefficients on lag one are 0.71 and 0.64 for Bitfinex and Kraken, respectively.

Both series are presented in Figure 5. From the bar plots, we can conclude that similar factors influence weekly volumes for these two cryptocurrencies. We can clearly distinguish periods of high and low trading intensity. Thus, a stationary model for these two series would not be the best choice. Also, for the second series we can notice that the three periods of high volumes are followed by low market activities, which is the

usual trading behavior. These three peeks occurred on the last week of May 2017, the second week of February 2018 and the last week of November 2018. Thus, we cannot conclude that high trading volumes are connected to a specific time of year, nor that they occur after a certain period.

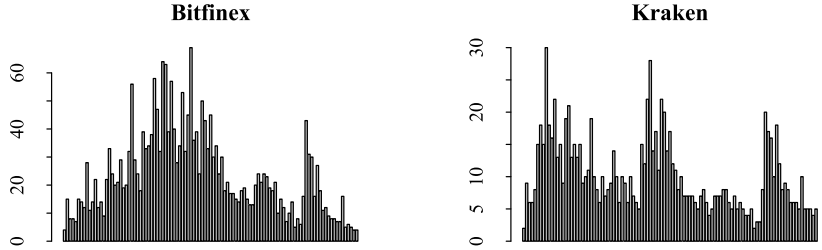

Figure 5: *Bar plots for weekly traded volumes of Bitfinex and Kraken cryptocurrencies.*

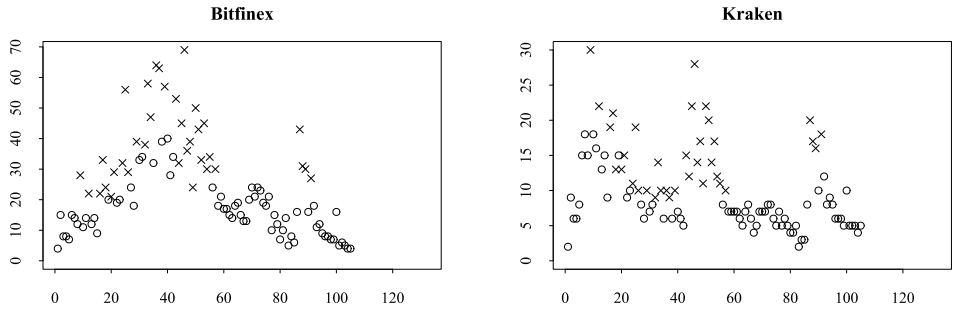

Figure 6: *States for weekly traded volumes of Bitfinex and Kraken cryptocurrencies. The state one is denoted with* $\times$ *and the state two with* $\circ$.

We notice two states of trading intensities for our model, which implies that we define the BRrNGINAR(1) with two random states, i.e. $r = 2$. These states are presented in Figure 6. Similarly as in two previous examples, following these states, we estimate the coefficient with the CML method and compare the results with the other three mentioned models. The obtained results are given in Table 7.

The values in Table 7 suggest that the BRrNGINAR(1) model has the smallest RMS. Thus, from the forecasting perspective, this model shows the best results. The advantage of the BRrNGINAR(1) model can be noticed especially with Bitfinex series, and some improvements are present with Kraken series as well. The reason for that probably lies in the fact that the observed series is non-stationary. As we can see with Bitfinex series, the BRrNGINAR(1) model estimated the mean value as $\hat{\mu}_1 = 20.782$ and $\hat{\mu}_2 = 24.362$, depending on the state. Other tested models have only one parameter for modelling the mean value. Even with the Kraken series where the difference between parameters $\hat{\nu}_1$ and $\hat{\nu}_2$ is not that big, we can see the improvements with RMS. The BRrNGINAR(1)

model achieves the second best value of AIC, which is a consequence of a larger number of parameters. Beside the fact that the number of parameters increases the AIC value, the estimated values of all these parameters have some deviation from the real values when the series are of length 105, as they are is this case. This fact also increases AIC value to a certain extent, having in mind the definition of AIC.

The purpose of all this testing is not to point out one model as the best model, but to demonstrate the type of series for which the BRrNGINAR(1) model is an adequate one. All observed series have in common that they fluctuate around different means during their evolution, which is expected to see when observing non-stationary series. This kind of behaviour looks like the series have trend, but not trend that can be easily captured with some linear or quadratic function, for example. These series take values from different intervals in different time frames which can be captured (in some degree) with the presented model.

8 Conclusion

The paper discusses a bivariate integer-valued autoregressive model of order one. The model is composed of two univariate models driven by the same hidden process. This hidden process is determined by the states that are assigned to the observed data. So, the hidden process allows the model to adjust itself to environment changes. As such, the model is non-stationary. Besides the main properties of the model, the focus is placed on its forecasting ability. Through tests on real data sets, it was shown that the model produces the smallest one-step-ahead prediction errors in terms of the root mean square error. Also, prediction errors are analysed in more detail by investigating prediction errors of each model component, the survival and the innovation component. These two components produce negatively correlated one-step-ahead prediction errors. This fact contributes to the reduction of the prediction errors which the model makes. The model contains a large number of parameters, so it requires a little bit larger data set for parameter estimation.

9 Appendix

Theorem 1

Proof.

a) Using the properties of the negative binomial thinning operator we have

$$Cov(X_k, Y_l) = \alpha_{z_k} Cov(X_{k-1}, Y_l), \quad Cov(X_k, X_l) = \alpha_{z_k} Cov(X_{k-1}, X_l),$$

$$Cov(X_k,Y_l)=\beta_{z_l}Cov(X_k,Y_{l-1}),\quad Cov(X_k,X_l)=\beta_{z_l}Cov(X_k,X_{l-1}).$$

Now, from these equalities, we simply get what is required.

b) This is obvious, based on the results given in a) and the fact that correlation is defined using the covariance. ∎

Theorem 2

Proof. It holds that

$$E(X_{n+k}|X_n)=\left(\prod_{s=1}^{k}\alpha_{z_{n+s}}\right)X_n+\sum_{l=1}^{k-1}\left(\prod_{s=1}^{k}\alpha_{z_{n+s}}\right)E(\varepsilon_{n+l})+E(\varepsilon_{n+k}),\qquad(14)$$

from the properties of the negative binomial thinning operator. If we take into account the distribution of the residuals, we get

$$E(X_{n+k}|X_n)=\left(\prod_{s=1}^{k}\alpha_{z_{n+s}}\right)(X_n-\mu_{z_n})+\mu_{z_{n+k}}.$$

The analogous relation holds for the Y component, so the required relation in a) is valid.

For the proof of b), the recurrent relation

$$\begin{aligned}Var(X_{n+k}|X_n,Y_n) &= \alpha^2_{z_{n+k}}Var(X_{n+k-1}|X_n,Y_n)\\ &\quad+\alpha_{z_{n+k}}(1+\alpha_{z_{n+k}})E(X_{n+k-1}|X_n)+Var(\varepsilon_{n+k})\end{aligned}$$

is used.

The statement c) follows from the fact that X_n and Y_n are independent for known Z_n, X_{n-1}, Y_{n-1} and Z_{n-1}. Also, Z_n is independent from X_{n-1} and Y_{n-1} for known Z_{n-1}. From the definition of the process $\{(X_n,Y_n)\}$, we have that

$$P(X_n=x_n|X_{n-1}=x_{n-1},Z_n=z_n,Z_{n-1}=z_{n-1})=P\left(\sum_{i=1}^{x_{n-1}}U_i^{z_{n-1}}+\varepsilon_n(z_{n-1},z_n)=x_n\right),$$

$$P(Y_n=y_n|Y_{n-1}=y_{n-1},Z_n=z_n,Z_{n-1}=z_{n-1})=P\left(\sum_{i=1}^{y_{n-1}}V_i^{z_{n-1}}+\eta_n(z_{n-1},z_n)=y_n\right).$$

Therefore, the statement c) is obtained using the above equations and properties of the residuals. ∎

Equation (12)

Proof. For simplicity, we will denote the probability mass function in the denominator as $P(X_n=x_n|Z_n=z_n,X_{n-1}=x_{n-1},Z_{n-1}=z_{n-1})=P(X_n=x_n|z_n,x_{n-1},z_{n-1})$. Now we

have

$$
\begin{aligned}
&E(\alpha_{z_n} * X_{n-1}(z_{n-1})|\mathscr{F}_n) = \sum_{j=0}^{x_n} j \cdot f(j) \\
&= \sum_{j=0}^{x_n} j \cdot \frac{P(NB(x_{n-1},\alpha_{z_n}) = j) \cdot P(\varepsilon_n(z_{n-1},z_n) = x_n - j)}{P(X_n = x_n|z_n,x_{n-1},z_{n-1})} \\
&= \frac{\sum_{j=0}^{x_n} j \binom{x_{n-1}+j-1}{j} \frac{\alpha_{z_{n-1}}^j}{(1+\alpha_{z_{n-1}})^{x_{n-1}+j}} P(\varepsilon_n(z_{n-1},z_n) = x_n - j)}{P(X_n = x_n|z_n,x_{n-1},z_{n-1})} \\
&= \frac{x_{n-1}\frac{\alpha_{z_{n-1}}}{1+\alpha_{z_{n-1}}} \sum_{j=1}^{x_n} \binom{x_{n-1}+j-1}{j-1} \frac{\alpha_{z_{n-1}}^{j-1}}{(1+\alpha_{z_{n-1}})^{x_{n-1}+j-1}} P(\varepsilon_n(z_{n-1},z_n) = x_n - j)}{P(X_n = x_n|z_n,x_{n-1},z_{n-1})} \\
&= \frac{x_{n-1}\frac{\alpha_{z_{n-1}}}{1+\alpha_{z_{n-1}}} \sum_{j=0}^{x_n-1} \binom{x_{n-1}+1+j-1}{j} \frac{\alpha_{z_{n-1}}^j}{(1+\alpha_{z_{n-1}})^{x_{n-1}+j}} P(\varepsilon_n(z_{n-1},z_n) = x_n - 1 - j)}{P(X_n = x_n|z_n,x_{n-1},z_{n-1})} \\
&= x_{n-1}\alpha_{z_{n-1}} \sum_{j=0}^{x_n-1} \frac{\binom{x_{n-1}+1+j-1}{j} \frac{\alpha_{z_{n-1}}^j}{(1+\alpha_{z_{n-1}})^{x_{n-1}+1+j}} P(\varepsilon_n(z_{n-1},z_n) = x_n - 1 - j)}{P(X_n = x_n|z_n,x_{n-1},z_{n-1})} \\
&= \frac{x_{n-1}\alpha_{z_{n-1}}}{P(X_n = x_n|Z_n = z_n,X_{n-1} = x_{n-1},Z_{n-1} = z_{n-1})} \\
&\cdot P(X_n = x_n - 1|Z_n = z_n,X_{n-1} = x_{n-1} + 1,Z_{n-1} = z_{n-1}),
\end{aligned}
$$

■

Equation (13)

Proof. For simplicity, we will introduce the following notation $P(A = a|Z_n = z_n,X_{n-1} = x_{n-1},Z_{n-1} = z_{n-1}) = P(A = a|z_n,x_{n-1},z_{n-1})$. Now we have

$$
\begin{aligned}
&E(\varepsilon_n(z_{n-1},z_n)|\mathscr{F}_n) = \sum_{i=0}^{x_n} i \cdot f(x_n - i) = \sum_{i=0}^{x_n} (x_n - i) \cdot f(i) \\
&= \frac{1}{P(X_n = x_n|z_n,x_{n-z},z_{n-1})} \\
&\cdot \sum_{i=0}^{x_n} (x_n - i) P(\varepsilon_n(z_{n-1},z_n) = x_n - i|z_n,x_{n-1},z_{n-1}) \cdot P(\alpha * X_{n-1}(Z_{n-1}) = i|z_n,x_{n-1},z_{n-1}) \\
&= \frac{1}{P(X_n = x_n|z_n,x_{n-z},z_{n-1})}
\end{aligned}
$$

$$\cdot \left[x_n \sum_{i=0}^{x_n} P(\varepsilon_n(z_{n-1}, z_n) = x_n - i | z_n, x_{n-1}, z_{n-1}) \cdot P(\alpha * X_{n-1}(Z_{n-1}) = i | z_n, x_{n-1}, z_{n-1}) \right.$$
$$\left. - \sum_{i=0}^{x_n} i \cdot P(\varepsilon_n(z_{n-1}, z_n) = x_n - i | z_n, x_{n-1}, z_{n-1}) \cdot P(\alpha * X_{n-1}(Z_{n-1}) = i | z_n, x_{n-1}, z_{n-1}) \right]$$
$$= \frac{1}{P(X_n = x_n | Z_n = z_n, X_{n-1} = x_{n-1}, Z_{n-1} = z_{n-1})}$$
$$\cdot [x_n P(X_n = x_n | Z_n = z_n, X_{n-1} = x_{n-1}, Z_{n-1} = z_{n-1})$$
$$- x_{n-1} \alpha_{z_{n-1}} P(X_n = x_n - 1 | Z_n = z_n, X_{n-1} = x_{n-1} + 1, Z_{n-1} = z_{n-1})],$$

where the second term inside the brackets is derived in the same way as equation (12). ■

Theorem 1 from Laketa et al. (2018)

Let $\{X_n(z_n)\}$ be the RrNGINARmax$(\mathscr{M}, \mathscr{A}, \mathscr{P})$ or the RrNGINAR$_1(\mathscr{M}, \mathscr{A}, \mathscr{P})$ process. Let us suppose that $z_n = j$ and $z_{n-1} = i$ for some i and $j \in E_r$. If $0 \leq \alpha_j \leq \frac{\mu_j}{1+\max_{k \in E_r} \mu_k}$, then the distribution of the random variable $\varepsilon_n(i, j)$ can be written as a mixture of two geometric distributed random variables with means μ_j and α_j as follows

$$\varepsilon_n(i, j) \stackrel{d}{=} \begin{cases} Geom\left(\frac{\mu_j}{1+\mu_j}\right), & w.p.\ 1 - \frac{\alpha_j \mu_i}{\mu_j - \alpha_j}, \\ Geom\left(\frac{\alpha_j}{1+\alpha_j}\right), & w.p.\ \frac{\alpha_j \mu_i}{\mu_j - \alpha_j}. \end{cases} \tag{15}$$

Acknowledgements

The authors are grateful to Professor Miroslav Ristić for valuable suggestions. Also, the authors highly appreciate all the comments and instructions from the reviewers of the SORT journal, which significantly improved the quality of the manuscript. The first author acknowledges the grant of MNTR 174026 and the second and the third author acknowledge the grant of MNTR 174013 for carrying out this research.

References

Al-Osh, M.A. and Aly, E.E.A.A. (1992). First order autoregressive time series with negative binomial and geometric marginals. *Communications in Statistics - Theory and Methods*, 21, 2483–2492.

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

Aly, E.E.A.A. and Bouzar, N. (1994). On some integer-valued autoregressive moving average models. *Journal of Multivariate Analysis*, 50, 132–151.

Alzaid, A.A. and Al-Osh, M.A. (1993). Some autoregressive moving average processes with generalized Poisson marginal distributions. *Annals of the Institute of Statistical Mathematics*, 45, 223–232.

Chakraborty, S. and Bhati, D. (2016). Transmuted geometric distribution with applications in modeling and regression analysis of count data. *SORT*, 40, 153–176.

Chakraborty, S. and Bhati, D. (2017). Corrigendum to "Transmuted geometric distribution with applications in modeling and regression analysis of count data". *SORT*, 41, 117-118.

Fernández-Fontelo, A., Cabana, A., Puig, P. and Morina, D. (2016). Under-reported data analysis with INAR-hidden Markov chains. *Statistics in Medicine*, 35, 4875–4890.

Fernández-Fontelo, A., Fontdecaba, S. and Puig, P. (2017). Integer-valued AR processes with Hermite innovations and time-varying parameters: An application to bovine fallen stock surveillance at a local scale. *Statistical Modelling*, 17, 1–24.

Fokianos, K. (2011). Some recent progress in count time series. *Statistics*, 45, 49–58.

Freeland, R.K. and McCabe B.P.M. (2004). Analysis of low count time series data by poisson autoregression. *Journal of Time Series Analysis*, 25, 701–722.

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

Jazi, M.A., Jones, G. and Lai, C.D. (2012b). Integer-valued AR with geometric innovations. *Journal of Iranian Statistical Society*, 2, 173–190.

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, 19, 439–455.

Laketa, P.N., Nastić, A.S. and Ristić, M.M. (2018). Generalized random environment INAR models of higher order. *Mediterranean Journal of Mathematics*, 15, 1–9.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, 21, 645–650.

Moriña, D., Puig, P., Ríos, J., Vilella, A. and Trilla, A. (2011). A statistical model for hospital admissions caused by seasonal diseases. *Statistics in Medicine*, 30, 3125–3136.

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

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

Pedeli, X. and Karlis, D. (2011). A bivariate INAR(1) process with application. *Statistical Modelling*, 11, 325–349.

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

Popović, P.M., Nastić, A.S. and Ristić, M.M. (2018). Residual analysis with bivariate INAR(1) models. *Revstat*, 16, 349–363.

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

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

Ristić, M.M., Nastić, A.S., Jayakumar, K. and Bakouch, H.S. (2012). A bivariate INAR(1) time series model with geometric marginals. *Applied Mathematics Letters*, 25, 481–485.

Ristić, M.M., Nastić, S.A. and Miletić Ilić, V.A. (2013). A geometric time series model with dependent Bernoulli counting series. *Journal of Time Series Analysis*, 34, 466–476.

Rydén, J. (2017). Statistical modeling of warm-spell duration series using hurdle models. *SORT*, 41, 177–188.

Tang, M. and Wang, Y. (2014). Asymptotic behavior of random coefficient INAR model under random environment defined by difference equation. *Advances in Difference Equations*, 2014, 1–9.

Weiß, C.H. (2015). A Poisson INAR(1) model with serially dependent innovations. *Metrika*, 78, 829–851.

Weiß, C.H., Homburg, A. and Puig, P. (2019). Testing for zero inflation and overdispersion in INAR(1) models. *Statistical Papers*, 60, 823–848.

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

Zheng, H., Basawa, I.V. and Datta, S. (2007). First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference*, 137, 212–229.

Zhu, R. and Joe, H. (2010). Negative binomial time series model based on expectation thinning operators. *Journal of Statistical Planning and Inference*, 140, 1974–1888.

Zhu, R. and Joe, H. (2006). Modelling count data time series with Markov processes based on binomial thinning. *Journal of Time Series Analysis*, 27, 725–738.