Residual analysis with bivariate INAR(1) models

Authors: Predrag M. Popović
– Faculty of Civil Engineering and Architecture, Serbia, University of Niš, (popovicpredrag@yahoo.com)

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

Miroslav M. Ristić
– Faculty of Sciences and Mathematics, Serbia, University of Niš, (miristic72@gmail.com)

Abstract:

- In this paper we analyze forecasting errors made by random coefficients bivariate integer-valued autoregressive models of order one. These models are based on the thinning operator to support discreteness of data. In order to achieve a comprehensive analysis, we introduce a model that implements a binomial as well as a negative binomial thinning operator. There are two components of the model: survival and innovation. Forecasting errors made by each of these two sources of uncertainty are unobservable in the classic way. Thus, we derive predictive distributions from which we obtain the expected value of each component of the model. We provide an example of residual analysis on real data.

Key-Words:

- *Bivariate INAR(1) model, residual analysis, predictive distribution, binomial thinning, negative binomial thinning, geometric marginal distribution*

AMS Subject Classification:

- 62M10

1. Introduction

An integer-valued time series is a sequence of integer data points measured at uniform time intervals. Such series are present in many fields of sciences. For example, in medicine the number of infected persons represents such a series, in finance the number of defaults, in criminology the number of committed crimes, in biology the size of the population of a species, etc. Thus, modelling integer-valued time series and better understanding their nature is a point of interest for many researchers. Some of the first models for non-negative integer-valued time series were introduced by [7], [1] and [2]. These models have autoregressive structure, where the autoregression is achieved through the thinning operator. Models with the full autoregressive-moving average structure were investigated in [8]. Following these ideas, many models have been developed. A survey on integer-valued autoregressive processes can be found in [15].

While these are models with constant coefficients, [17] defined an integer-valued random coefficient model. Using the approach proposed by [3], [12] introduced a bivariate integer-valued random coefficient model. The dependence between processes that this model consists of is achieved through their autoregressive components, which are based on the negative binomial thinning operator. Some modifications of this model regarding the thinning operator and the marginal distribution are discussed in [9]. In this paper we focus on analyzing prediction errors made by these types of models. Since these models are composed of two components, survival and innovation, there are two sources of uncertainty. We try to estimate the portion of prediction error made by the survival and by the innovation component separately. Since these residuals are unobservable, we derive predictive distribution and calculate expected values of these components. Some aspects of predictive distributions for univariate models were presented in [13] and [14]. Residual analysis for univariate models was discussed in [5] and [16]. We extend the research on the bivariate models with random coefficients.

In addition, to cover two types of thinning operators, we introduce a bivariate model whose survival components are generated by different thinning operators, namely, binomial and negative binomial. This mix of thinning operators makes it possible to model two dependent processes whose survival parts have different properties. While the survival component generated by the negative binomial thinning operator does produce new members of the series, the other one generated by the binomial thinning does not and new members depends only on the innovation component. To motivate the model we consider two data series: monthly count of motor vehicle thefts and monthly count of drug dealing activities. The first series is characterized by the fact that offended persons are not provoked to commit the same criminal act, but the second series is to a large extent generated by itself since some amount of drugs has been resold many times.

The paper is organized as follows. In Section 2 we discuss the general form of bivariate integer-valued autoregressive models of order one with random coefficients. Section 3 introduces a bivariate model with both, binomial and negative binomial, thinning operators. We discuss residual analysis in Section 4. Real data modelling is considered in Section 5.

2. Bivariate INAR models

In this section we state a general form of the random coefficient bivariate integer-valued autoregressive models of order one (BINAR(1)). In order to define a model suitable for the time series of count, we use the thinning operators. The binomial thinning operator is defined as $\alpha_1 \circ X = \sum_{i=1}^{X} B_i$, where X is a non-negative integer-valued random variable and $\{B_i\}$ is a sequence of i.i.d. Bernoulli random variables with mean parameter α_1. The negative binomial thinning operator is defined as $\alpha_1 * X = \sum_{i=1}^{X} G_i$, where $\{G_i\}$ is a sequence of i.i.d. geometric random variables with mean parameter α_1. For the time being we will not specify the thinning operator in the definition of BINAR(1) model. Let us denote a nonnegative bivariate time series of counts by $\boldsymbol{Z}_n$ and introduce a random matrix $\boldsymbol{A}_n = \begin{bmatrix} U_{1n} & U_{2n} \\ V_{1n} & V_{2n} \end{bmatrix}$, whose elements have the joint probability mass function defined as $P(U_{1n} = \alpha_1, U_{2n} = 0) = p = 1 - P(U_{1n} = 0, U_{2n} = \alpha_1)$ and $P(V_{1n} = \alpha_2, V_{2n} = 0) = q = 1 - P(V_{1n} = 0, V_{2n} = \alpha_2)$, where $\alpha_1, \alpha_2 \in (0,1)$

and $p, q \in [0,1]$. Then, the structure of BINAR(1) model is given by

$$\boldsymbol{Z}_n = \boldsymbol{A}_n \star \boldsymbol{Z}_{n-1} + \boldsymbol{e}_n, \quad n \geq 1, \tag{2.1}$$

where $\{\boldsymbol{e}_n\}$ represents the innovation process, which is composed of two independent series. The thinning operator is denoted with $\star$ and it acts as the matrix multiplication. The two processes that figure in $\boldsymbol{Z}_n$ are mutually dependent and their dependence is achieved trough autoregressive components, which are named survival processes. Coefficients that figure in (2.1) are random variables, which make this model significantly different from the similar multivariate INAR models (such as the one presented in [4] and [6]). Notice that

$$E(\boldsymbol{A}_n) = \boldsymbol{A} = \begin{bmatrix} \alpha_1 p & \alpha_1(1-p) \\ \alpha_2 q & \alpha_2(1-q) \end{bmatrix}.$$

It is easy to show that $E(\boldsymbol{A}_n \star \boldsymbol{Z}_n) = \boldsymbol{A}E(\boldsymbol{Z}_n)$. Following the discussion from [6], $\boldsymbol{I} - \boldsymbol{A}$ is a non singular matrix if all eigenvalues of $\boldsymbol{A}$ are inside the unit circle, which is proved for matrix $\boldsymbol{A}$ in [12]. All this implies that $E(\boldsymbol{Z}_n) = (\boldsymbol{I} - \boldsymbol{A})^{-1}E(\boldsymbol{e}_n)$. Since $(\boldsymbol{I} - \boldsymbol{A})^{-1}$ is a matrix of finite values, $E(\boldsymbol{Z}_0) < \infty$ iff $E(\boldsymbol{e}_1) < \infty$. The conditional expectation for process (2.1) is $E(\boldsymbol{Z}_{n+k}|\boldsymbol{Z}_n) = \boldsymbol{A}^k\boldsymbol{Z}_n + (\boldsymbol{I}-\boldsymbol{A})^{-1}(\boldsymbol{I}-\boldsymbol{A}^k)(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{\mu}$, where $\boldsymbol{\mu} = E(\boldsymbol{Z}_n)$. The correlation structure of BINAR(1) model is given as $Cov(\boldsymbol{Z}_{n+k}, \boldsymbol{Z}_n) = \boldsymbol{A}^k Var(\boldsymbol{Z}_n)$, $k \geq 0$. Since the eigenvalues of matrix $\boldsymbol{A}$ are inside the unit circle, covariance tends to zero and conditional expectation tends to the unconditional one, as k tends to infinity. More details on the correlation structure can be found in [9].

3. Model with mixed thinning operators

In this section we introduce a new bivariate time series model $\{(X_{1,n}, X_{2,n})\}$, $n \in \mathbb{N}_0$, where the two time series are dependent but evolve under different thinning operators. Let $\{X_{1,n}\}$ and $\{X_{2,n}\}$ be the two nonnegative integer-valued time series with probability mass function $P(X_{i,n} = k) = \mu^k/(1+\mu)^{k+1}$, $k \geq 0$, $\mu > 0$ and $i \in \{1,2\}$. A mixed geometric bivariate autoregressive process of order one (BVMIXGINAR(1)) is given by the following equations

$$X_{1,n} = \begin{cases} \alpha_1 \circ X_{1,n-1} + \varepsilon_{1,n}, & \text{w.p. } p, \\ \alpha_1 \circ X_{2,n-1} + \varepsilon_{1,n}, & \text{w.p. } 1-p, \end{cases} \tag{3.1}$$

$$X_{2,n} = \begin{cases} \alpha_2 * X_{1,n-1} + \varepsilon_{2,n}, & \text{w.p. } q, \\ \alpha_2 * X_{2,n-1} + \varepsilon_{2,n}, & \text{w.p. } 1-q, \end{cases} \tag{3.2}$$

where $\{\varepsilon_{1,n}\}$ and $\{\varepsilon_{2,n}\}$ are i.i.d. sequences. The random vectors $(\varepsilon_{1,n}, \varepsilon_{2,n})$ and $(X_{1,m}, X_{2,m})$ are independent for all $m < n$. The thinning operators are defined in previous section and the counting series in $\alpha_1 \circ X_{1,n}$, $\alpha_1 \circ X_{2,n}$, $\alpha_2 * X_{1,n}$ and $\alpha_2 * X_{2,n}$ are mutually independent for all $n \in \mathbb{N}_0$ and are also independent of innovation processes $\{\varepsilon_{1,n}\}$ and $\{\varepsilon_{2,n}\}$. The distributions of the innovation processes are given by the following theorem.

Theorem 3.1. *Let $X_{1,0}$ and $X_{2,0}$ have the $Geom(\frac{\mu}{1+\mu})$ distribution, where $\mu > 0$. The stationary bivariate time series $\{(X_{1,n}, X_{2,n})\}_{n\in\mathbb{N}_0}$ given by equations (3.1) and (3.2) has $Geom(\frac{\mu}{1+\mu})$ marginal distributions if and only if the processes $\{\varepsilon_{1,n}\}$ and $\{\varepsilon_{2,n}\}$ are distributed as*

$$\varepsilon_{1,n} \stackrel{d}{=} \begin{cases} Geom(\frac{\mu}{1+\mu}), & w.p.\ 1-\alpha_1, \\ 0, & w.p.\ \alpha_1, \end{cases} \tag{3.3}$$

$$\varepsilon_{2,n} \stackrel{d}{=} \begin{cases} Geom(\frac{\mu}{1+\mu}), & w.p.\ \frac{\mu(1-\alpha_2)-\alpha_2}{\mu-\alpha_2}, \\ Geom(\frac{\alpha_2}{1+\alpha_2}), & w.p.\ \frac{\alpha_2\mu}{\mu-\alpha_2}, \end{cases} \tag{3.4}$$

where $\alpha_1 \in (0,1)$, $\alpha_2 \in (0, \frac{\mu}{1+\mu}]$ and $p, q \in [0,1]$.

Proof: Let us assume that the stationary time series $\{(X_{1,n}, X_{2,n})\}$ has the geometric marginal distribution $Geom(\frac{\mu}{1+\mu})$, $\mu > 0$. Since the random variables $X_{1,n}$ and $X_{2,n}$ are equal in distribution, considering probability generating functions we obtain $\Phi_{X_{1,n}}(s) = \Phi_{\varepsilon_{1,n}}(s)\Phi_{X_{1,n-1}}(\Phi_{B_i}(s))$, which follows from (3.1). From their geometric distribution we obtain

$$\Phi_{\varepsilon_{1,n}}(s) = \frac{1+\mu\alpha_1(1-s)}{1+\mu(1-s)} = \alpha_1 + (1-\alpha_1)\frac{1}{1+\mu-\mu s},$$

which proves equation (3.3). In a similar manner we derive the probability generating function for $\varepsilon_{2,n}$ and obtain

$$\begin{aligned} \Phi_{\varepsilon_{2,n}}(s) &= \frac{(1+\mu)(1+\alpha_2-\alpha_2 s)-\mu}{(1+\mu-\mu s)(1+\alpha_2-\alpha_2 s)} \\ &= \frac{\mu(1-\alpha_2)-\alpha_2}{\mu-\alpha_2}\frac{1}{a+\mu-\mu s} + \frac{\alpha_2\mu}{\mu-\alpha_2}\frac{1}{1+\alpha_2-\alpha_2 s}. \end{aligned}$$

Now, equation (3.4) follows under the constraints given in [11] for NGINAR(1) model.

Conversely, let us assume that the distributions of the random variables $\varepsilon_{1,n}$ and $\varepsilon_{2,n}$ are given by equations (3.3) and (3.4), respectively. Since $X_{1,0} \stackrel{d}{=} X_{2,0} \stackrel{d}{=} Geom(\mu/(1+\mu))$, we obtain

$$\begin{aligned} \Phi_{X_{1,1}}(s) &= \Phi_{X_{1,0}}(1-\alpha_1+\alpha_1 s)\Phi_{\varepsilon_{1,1}}(s) \\ &= \frac{1}{1+\mu-\mu(1-\alpha_1+\alpha_1 s)}\frac{1+\mu\alpha_1(1-s)}{1+\mu(1-s)} = \frac{1}{1+\mu-\mu s} \end{aligned}$$

and

$$\begin{aligned} \Phi_{X_{2,1}}(s) &= \Phi_{X_{2,0}}\left(\frac{1}{1+\alpha_2-\alpha_2 s}\right)\Phi_{\varepsilon_{2,1}}(s) \\ &= \frac{1}{1+\mu(1-\frac{1}{1+\alpha_2-\alpha_2 s})} \cdot \frac{(1+\mu)(1+\alpha_2-\alpha_2 s)-\mu}{(1+\mu-\mu s)(1+\alpha_2-\alpha_2 s)} = \frac{1}{1+\mu-\mu s}. \end{aligned}$$

Thus, $X_{1,1}$ and $X_{2,1}$ have geometric distribution with parameter $\mu/(1+\mu)$. Using mathematical induction we can prove that $X_{1,n} \stackrel{d}{=} X_{2,n} \stackrel{d}{=} Geom(\mu/(1+\mu))$ for any $n \in \mathbb{N}_0$. □

Even if we assume that $X_{1,0}$ and $X_{2,0}$ have the same arbitrary distribution, $X_{1,n}$ as well as $X_{2,n}$ converges to geometric distribution $Geom(\mu/(1+\mu))$, as $n \to \infty$, if random variables $\varepsilon_{1,n}$ and $\varepsilon_{2,n}$ have the distribution given by Theorem 3.1. This can be proved with the following two equations. The first equation is

$$\begin{aligned}
\Phi_{X_{1,n}}(s) &= \Phi_{X_{1,n-1}}(1-\alpha_1+\alpha_1 s)\Phi_{\varepsilon_{1,n}}(s) \\
&= \Phi_{X_{1,0}}(1-\alpha_1^n+\alpha_1^n s)\prod_{k=0}^{n-1}\frac{1+\mu\alpha_1^{k+1}(1-s)}{1+\mu\alpha_1^k(1-s)} \\
&= \Phi_{X_{1,0}}(1-\alpha_1^n+\alpha_1^n s)\frac{1+\mu\alpha_1^n(1-s)}{1+\mu(1-s)} \xrightarrow{n\to\infty} \frac{1}{1+\mu-\mu s}.
\end{aligned}$$

Also,

$$\begin{aligned}
&\Phi_{X_{2,n}}(s) = \Phi_{X_{2,n-1}}\left(\frac{1}{1+\alpha_2-\alpha_2 s}\right)\Phi_{\varepsilon_{2,n}}(s) \\
&= \Phi_{X_{2,0}}\left(\frac{1-\alpha_2+\alpha_2(1-s)(1-\alpha_2^{n-1})}{1-\alpha_2+\alpha_2(1-s)(1-\alpha_2^{n})}\right)\prod_{k=0}^{n-1}\Phi_{\varepsilon_{2,k}}\left(\frac{1-\alpha_2+\alpha_2(1-s)(1-\alpha_2^{k-1})}{1-\alpha_2+\alpha_2(1-s)(1-\alpha_2^{k})}\right) \\
&= \Phi_{X_{2,0}}\left(\frac{1-\alpha_2+\alpha_2(1-s)(1-\alpha_2^{n-1})}{1-\alpha_2+\alpha_2(1-s)(1-\alpha_2^{n})}\right) \\
&\quad\times\frac{(1-\alpha_2^2(1-s)-\alpha_2 s)(1+\alpha_2^n\mu(1-s)-\alpha_2^{n+1}(1+\mu)(1-s)-\alpha_2 s)}{(1-\alpha_2^{n+1}(1-s)-\alpha_2 s)(1+\alpha_2\mu(1-s)-\alpha_2^2(1+\mu)(1-s)-\alpha_2 s)} \\
&\xrightarrow{n\to\infty} \frac{1}{1+\mu-\mu s}.
\end{aligned}$$

Random variables $X_{1,n}$ and $X_{2,n}$ are independent for known $X_{1,n-1}$ and $X_{2,n-1}$. Thus, the conditional distribution of $(X_{1,n}, X_{2,n})$, given $(X_{1,n-1}, X_{2,n-1})$, is defined as

$$\begin{aligned}
&P(X_{1,n}=x, X_{2,n}=y|X_{1,n-1}=u, X_{2,n-1}=v) \\
&= P(X_{1,n}=x|X_{1,n-1}=u, X_{2,n-1}=v)P(X_{2,n}=y|X_{1,n-1}=u, X_{2,n-1}=v).
\end{aligned}$$

The conditional probability mass function of the random variable $X_{1,n}$ for given $X_{1,n-1}$ and $X_{2,n-1}$ has the form

$$\begin{aligned}
&P(X_{1,n}=x|X_{1,n-1}=u, X_{2,n-1}=v) \\
&= p\sum_{k=0}^{\min(x,u)} P(\varepsilon_{1,n}=x-k)P(\alpha_1\circ X_{1,n-1}=k|X_{1,n-1}=u) \\
&+ (1-p)\sum_{k=0}^{\min(x,v)} P(\varepsilon_{1,n}=x-k)P(\alpha_1\circ X_{2,n-1}=k|X_{2,n-1}=v). \qquad (3.5)
\end{aligned}$$

Similarly, for $X_{2,n}$ the form is

$$\begin{aligned} &P(X_{2,n}=y|X_{1,n-1}=u,X_{2,n-1}=v)\\ &=q\sum_{k=0}^{y}P(\varepsilon_{2,n}=y-k)P(\alpha_2*X_{1,n-1}=k|X_{1,n-1}=u)\\ &+(1-q)\sum_{k=0}^{y}P(\varepsilon_{2,n}=y-k)P(\alpha_2*X_{2,n-1}=k|X_{2,n-1}=v). \end{aligned} \tag{3.6}$$

The random variables $\alpha_1 \circ X$ and $\alpha_2 * X$ under the condition $X = u$ have binomial and negative binomial distribution with parameters (u, α_1) and $(u, \frac{\alpha_2}{1+\alpha_2})$, respectively (where the probability mass function for negative binomial distribution is taken as $P(\alpha_2 * X = k|X = u) = \binom{u+k-1}{k}\frac{\alpha_2^k}{(1+\alpha_2)^{k+u}}$). Notice that the probability mass functions for the innovation processes are, respectively,

$$P(\varepsilon_{1,n}=x-k)=1_{\{x=k\}}\alpha_1+(1-\alpha_1)\frac{\mu^{x-k}}{(1+\mu)^{x-k+1}},$$

$$P(\varepsilon_{2,n}=y-k)=\frac{\mu(1-\alpha_2)-\alpha_2}{\mu-\alpha_2}\frac{\mu^{y-k}}{(1+\mu)^{y-k+1}}+\frac{\alpha_2\mu}{\mu-\alpha_2}\frac{\alpha_2^{y-k}}{(1+\alpha_2)^{y-k+1}},$$

where 1_A is the indicator function of a random event A.

The estimation of unknown parameters of the bivariate INAR(1) models with random coefficients is discussed in details in [9]. We consider the conditional maximum likelihood method for parameters estimation of the model presented in this paper.

For the given values $\{(X_{1,k}, X_{2,k})\}_{k=\overline{0,n}}$, we set the conditional likelihood function as

$$L_1(\boldsymbol{\theta})=\sum_{i=1}^{n}\ln P(X_{1,i}=x_{1,i},X_{2,i}=x_{2,i}|X_{1,i-1}=x_{1,i-1},X_{2,i-1}=x_{2,i-1},\boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = (\alpha_1, \alpha_2, p, q, \mu)$ is a vector of unknown parameters. The probability mass function is defined as a product of functions (3.5) and (3.6). The maximization of the log-likelihood function is obtained by numerical procedure, which, in our case, is conducted through the programming language `R` and the function `nlm`.

4. Residuals

The standard statistic used for determining a goodness of fit is obtained by summing squared residuals. The residuals are obtained as a difference between a value at time n and an expected value of the process in time n for the given value at $n-1$, i.e.,

$$r_{X_{1,n}} = X_{1,n} - \alpha_1 p X_{1,n-1} - \alpha_1(1-p)X_{2,n-1} - \mu_{\varepsilon_1},$$
$$r_{X_{2,n}} = X_{2,n} - \alpha_2 q X_{1,n-1} - \alpha_2(1-q)X_{2,n-1} - \mu_{\varepsilon_2},$$

where μ_{ε_i} are the expected values of the random variables ε_i, $i \in \{1,2\}$. Since our process is composed of two sources of uncertainty (survival process and innovation process) it would be useful to track residuals of each source separately. This idea for one-dimensional INAR process is presented in [5] and [16], while we extend it for the bivariate case where the coefficients of the model are random variables. The residual analysis for a bivariate model is also investigated in [10], but the model has constant coefficients, independent survival processes and dependent innovation processes, which makes it significantly different from BVMIXGINAR(1) model.

If we introduce the two pairs of random variables (U_{1n}, U_{2n}) and (V_{1n}, V_{2n}) defined in Section 2, we can present BVMIXGINAR(1) model as

$$X_{1,n} = U_{1n} \circ X_{1,n-1} + U_{2n} \circ X_{2,n-1} + \varepsilon_{1,n}, \tag{4.1}$$
$$X_{2,n} = V_{1n} * X_{1,n-1} + V_{2n} * X_{2,n-1} + \varepsilon_{2,n}. \tag{4.2}$$

The process in this form is more tractable in terms of survival and innovation components. Therefore, we get two sets of residuals: $r^{sur}_{X_{1,n}} = U_{1n} \circ X_{1,n-1} + U_{2n} \circ X_{2,n-1} - \alpha_1 p X_{1,n-1} - \alpha_1(1-p)X_{2,n-1}$ and $r^{in}_{X_{1,n}} = \varepsilon_{1,n} - \mu_{\varepsilon_1}$ (analogous for the process $\{X_{2,n}\}$). The problem that arises here is that the binomial thinning component and the innovation component are not observable. Thus, we have to consider their conditional expectation with respect to the σ-algebra generated by vectors $(X_{1,n}, X_{2,n}), (X_{1,n-1}, X_{2,n-1}), ..., (X_{1,0}, X_{2,0})$. Since the process $\{(X_{1,n}, X_{2,n})\}$ is lag-one dependent, we investigate conditional expectations with respect to the σ-algebra generated only with random vectors at moments n and $n-1$, $\mathcal{F}_n(X_{1,n}, X_{2,n}, X_{1,n-1}, X_{2,n-1})$.

Let us introduce the notations $P_{u,v}(A) = P(A|X_{1,n-1} = u, X_{2,n-1} = v)$ and $P_{x_1,x_2,u,v}(A) = P(A|X_{1,n} = x_1, X_{2,n} = x_2, X_{1,n-1} = u, X_{2,n-1} = v)$. The conditional probability mass function of the first addend in equation (4.1) with respect to the σ-algebra $\mathcal{F}_n$, for $m, x, y, u, v \in \mathbb{N}_0$, is

$$\begin{aligned}
&P_{x_1,x_2,u,v}(U_{1n} \circ X_{1,n-1} = m)\\
&= \frac{P_{u,v}(U_{1n} \circ X_{1,n-1} = m, U_{1n} \circ X_{1,n-1} + U_{2n} \circ X_{2,n-1} + \varepsilon_{1,n} = x_1)}{P_{u,v}(X_{1,n} = x_1)}\\
&= \frac{1}{P_{u,v}(X_{1,n} = x_1)}[pP_{u,v}(\alpha_1 \circ X_{1,n-1} = m, 0 \circ X_{2,n-1} + \varepsilon_{1,n} = x_1 - m)\\
&\qquad + (1-p)P_{u,v}(0 \circ X_{1,n-1} = m, \alpha_1 \circ X_{2,n-1} + \varepsilon_{1,n} = x_1 - m)]\\
&= \frac{1}{P_{u,v}(X_{1,n} = x_1)}[pP(Bin(u,\alpha_1) = m)P(\varepsilon_{1,n} = x_1 - m)\\
&\qquad + (1-p)1_{\{m=0\}}P(Bin(v,\alpha_1) + \varepsilon_{1,n} = x_1 - m)],
\end{aligned}$$

where $Bin(u,\alpha)$ denotes a random variable with binomial distribution and parameters u i α. In a similar manner, for $r,k,s\in\mathbb{N}_0$, we obtain the following three equations,

$$P_{x_1,x_2,u,v}(U_{2n}\circ X_{2,n-1}=r)=\frac{1}{P_{u,v}(X_{1,n}=x_1)}[pI_{\{r=0\}}P(Bin(u,\alpha_1)+\varepsilon_{1,n}=x_1)$$
$$+(1-p)P(Bin(v,\alpha_2)=r)P(\varepsilon_{1,n}=x_1-r)],$$

$$P_{x_1,x_2,u,v}(V_{1n}*X_{1,n-1}=k)=\frac{1}{P_{u,v}(X_{2,n}=x_2)}[qP(NB(u,\alpha_2)=k)P(\varepsilon_{2,n}=x_2-k)$$
$$+(1-q)I_{\{k=0\}}P(NB(v,\alpha_2)+\varepsilon_{2,n}=x_2)],$$

$$P_{x_1,x_2,u,v}(V_{2n}*X_{2,n-1}=s)=\frac{1}{P_{u,v}(X_{2,n}=x_2)}[qI_{\{s=0\}}P(NB(u,\alpha_2)+\varepsilon_{2,n}=x_2)$$
$$+(1-q)P(NB(v,\alpha_2)=s)P(\varepsilon_{2,n}=y-s)].$$

In the last two equations the notation $NB(u,\alpha)$ stands for a random variable with negative binomial distribution with parameters u and $\frac{\alpha}{1+\alpha}$.

With these results in mind and applying some algebra, we obtain the following equations

$$\begin{aligned}
&E(U_{in}\circ X_{i,n-1}|\mathcal{F}_n)=\sum_{j=1}^{u_i} jP_{x_1,x_2,u,v}(U_{in}\circ X_{i,n-1}=j)\\
&=\frac{p_i}{P_{u,v}(X_{1,n}=x_1)}\sum_{j=1}^{\min(u_i,x_1)} j\binom{u_i}{j}\alpha_1^j(1-\alpha_1)^{u_i-j}P(\varepsilon_{1,n}=x_1-j)\\
&=\frac{\alpha_1p_iu_i}{P_{u,v}(X_{1,n}=x_1)}\sum_{j=0}^{\min(u_i-1,x_1-1)}\binom{u_i-1}{j}\alpha_1^j(1-\alpha_1)^{u_i-1-j}P(\varepsilon_{1,n}=x_1-1-j)\\
&=\frac{\alpha_1p_iu_i}{P_{u,v}(X_{1,n}=x_1)}P'_{i,u_i-1}(\alpha_1\circ X_{i,n-1}+\varepsilon_{1,n}=x_1-1)
\end{aligned}$$

and

$$\begin{aligned}
&E(V_{in}*X_{i,n-1}|\mathcal{F}_n)=\sum_{j=1}^{x_2} jP_{x_1,x_2,u,v}(V_{in}*X_{i,n-1}=j)\\
&=\frac{q_i}{P_{u,v}(X_{2,n}=x_2)}\sum_{j=1}^{x_2} j\binom{u_i+j-1}{j}\frac{\alpha_2^j}{(1+\alpha_2)^{u_i+j}}P(\varepsilon_{2,n}=x_2-j)\\
&=\frac{\alpha_2q_iu_i}{P_{u,v}(X_{2,n}=x_2)}\sum_{j=0}^{x_2-1}\binom{u_i+1+j-1}{j}\frac{\alpha_2^j}{(1+\alpha_2)^{u_i+1+j}}P(\varepsilon_{2,n}=x_2-1-j)\\
&=\frac{\alpha_2q_iu_i}{P_{u,v}(X_{2,n}=x_2)}P'_{i,u_i+1}(\alpha_2*X_{i,n-1}+\varepsilon_{2,n}=x_2-1),
\end{aligned}$$

where we introduced the notations $P'_{i,x}(A)=P(A|X_{i,n-1}=x)$, $i=1,2$, $p_1=p$, $p_2=1-p$, $q_1=q$, $q_2=1-q$, $u_1=u$ and $u_2=v$. Thus, we can conclude that the conditional expectation of the survival part of the process (4.1) is calculated as

$$
\begin{aligned}
&E(U_{1n}\circ X_{1,n-1}+U_{2n}\circ X_{2,n-1}|\mathcal{F}_n)=pE(\alpha_1\circ X_{1,n-1}|\mathcal{F}_n)+(1-p)E(\alpha_1\circ X_{2,n-1}|\mathcal{F}_n)\\
&=\frac{1}{P_{u,v}(X_{1,n}=x_1)}\cdot[\alpha_1 puP_{1,u-1}(\alpha_1\circ X_{1,n-1}+\varepsilon_{1,n}=x_1-1)\\
&\quad+\alpha_1(1-p)vP_{2,v-1}(\alpha_1\circ X_{2,n-1}+\varepsilon_{1,n}=x_1-1)]
\end{aligned}
$$

and for the process (4.2) as

$$
\begin{aligned}
&E(V_{1n}*X_{1,n-1}+V_{2n}*X_{2,n-1}|\mathcal{F}_n)=qE(\alpha_2*X_{1,n-1}|\mathcal{F}_n)+(1-q)E(\alpha_2*X_{2,n-1}|\mathcal{F}_n)\\
&=\frac{1}{P_{u,v}(X_{2,n}=x_2)}\cdot[qu\alpha_2P_{1,u+1}(\alpha_2*X_{1,n-1}+\varepsilon_{2,n}=x_2-1)\\
&\quad+(1-q)v\alpha_2P_{2,v+1}(\alpha_2*X_{2,n-1}+\varepsilon_{2,n}=x_2-1)]\,.
\end{aligned}
$$

We have defined the innovation processes such that $(\varepsilon_{1,n},\varepsilon_{2,n})$ is independent of $(X_{1,m},X_{2,m})$ for $m<n$. Since we observed the conditional expectation at time n with respect to the σ-algebra $\mathcal{F}_n$, we need to pay special attention here. The conditional probability mass functions are

$$
\begin{aligned}
P_{x_1,x_2,u,v}(\varepsilon_{1,n}=x_1-k)=&\frac{P(\varepsilon_{1,n}=x_1-k)}{P_{u,v}(X_{1,n}=x_1)}\\
&\times[pP_{u,v}(\alpha_1\circ X_{1,n-1}=k)+(1-p)P_{u,v}(\alpha_1\circ X_{2,n-1}=k)]
\end{aligned}
$$

and

$$
\begin{aligned}
P_{x_1,x_2,u,v}(\varepsilon_{2,n}=x_2-k)=&\frac{P(\varepsilon_{2,n}=x_2-k)}{P_{u,v}(X_{2,n}=x_2)}\\
&\times[qP_{u,v}(\alpha_2*X_{1,n-1}=k)+(1-q)P_{u,v}(\alpha_2*X_{2,n-1}=k)]\,.
\end{aligned}
$$

Hance, the corresponding conditional expectations are

$$
\begin{aligned}
&E(\varepsilon_{1,n}|\mathcal{F}_n)=\frac{1}{P_{u,v}(X_{1,n}=x_1)}\\
&\times[p\,(x_1P_{1,u}(\alpha_1\circ X_{1,n-1}+\varepsilon_{1,n}=x_1)-\alpha_1uP_{1,u-1}(\alpha_1\circ X_{1,n-1}+\varepsilon_{1,n}=x_1-1))\\
&+(1-p)\,(x_1P_{2,v}(\alpha_1\circ X_{2,n-1}+\varepsilon_{1,n}=x_1)-\alpha_1vP_{2,v-1}(\alpha_1\circ X_{2,n-1}+\varepsilon_{1,n}=x_1-1))]
\end{aligned}
$$

and

$$
\begin{aligned}
&E(\varepsilon_{2,n}|\mathcal{F}_n)=\frac{1}{P_{u,v}(X_{2,n}=x_2)}\\
&\times[q\,(x_2P_{1,u}(\alpha_2*X_{1,n-1}+\varepsilon_{2,n}=x_2)-\alpha_2uP_{1,u+1}(\alpha_2*X_{1,n-1}+\varepsilon_{2,n}=x_2-1))\\
&+(1-q)\,(x_2P_{2,v}(\alpha_2*X_{2,n-1}+\varepsilon_{2,n}=x_2)-\alpha_2vP_{2,v+1}(\alpha_2*X_{2,n-1}+\varepsilon_{2,n}=x_2-1))]\,.
\end{aligned}
$$

Now we can distinguish between the error from the survival and the error from the innovation process. If we sum these two values we obtain the following results

$$\begin{aligned} r^{surr}_{X_{1,n}} + r^{in}_{X_{1,n}} =& E(p\alpha_1 \circ X_{1,n-1} + (1-p)\alpha_1 \circ X_{2,n-1} | X_{1,n}, X_{1,n-1}) \\ & - \alpha_1 p X_{1,n-1} - \alpha_1(1-p)X_{2,n-1} + E(\varepsilon_{1,n}|X_{1,n}, X_{1,n-1}) - \mu_\varepsilon \\ =& E(p\alpha_1 \circ X_{1,n-1} + (1-p)\alpha_1 \circ X_{2,n-1} + \varepsilon_{1,n} | X_{1,n}, X_{1,n-1}) \\ & - \alpha_1 p X_{1,n-1} - \alpha_1(1-p)X_{2,n-1} - \mu_{\varepsilon_{1,n}} \\ =& X_{1,n} - \alpha_1 p X_{1,n-1} - \alpha_1(1-p)X_{2,n-1} - \mu_{\varepsilon_1} = r_{X_{1,n}}. \end{aligned} \tag{4.3}$$

We can conclude that the sum of these two error terms is equal to the error term obtained by using conditional expectation for the process $\{X_{1,n}\}$ with respect to σ-algebra $\mathcal{F}_{n-1}$. The conclusion is analogous for the process $\{X_{2,n}\}$.

5. Application

In this section, we discuss the characteristics of data for which BVMIXGINAR(1) model is the most adequate. We compare results of BVMIXGINAR(1) model with results of some other bivariate models. At the end of the section we analyze prediction errors of BVMIXGINAR(1) model and suggest how the model can be improved.

We analyze data from the Pittsburgh police department number 407, which can be found on the website www.forecastingprinciples, where we focus on the number of stolen vehicles (MVTHEFT) and the number of reported drug activities (C_DRUG) per month from January 1990 to December 2001. The average values of these two series are 1.74 and 1.5, and the variances are 2.98 and 5.01, respectively. The correlation between the series is 0.22. The bar plots and corelograms are given in Figure 1. Both corelograms show the presence of lag 1 autocorrelation. Although there are some autocorrelations on higher lags for series C_DRUG, the value on the first lag is dominant. High positive correlation between the series, overdispersion and the first lag autocorrelation imply that BVMIXGINAR(1) might be adequate.

We compare BVMIXGINAR(1) model with models BVNGINAR(1) introduced in [12], and BVPOINAR(1) introduced in [9], since both of these models have random coefficients and a similar structure. For the three models, we compare their values of the log-likelihood functions and the root mean square errors (RMS) made by one step ahead prediction. The results are presented in Table 1.

According to the test results, BVMIXGINAR(1) is the most adequate for these data. Notice that models with geometric distribution obtain higher values of the likelihood function. BVMIXGINAR(1) achieves slightly higher log-likelihood values than BVNGINAR(1) but much lower RMS for MVTHEFT series. Modelling C_DRUG series with geometric distribution where survival processes evolve under negative binomial thinning provides the best results. RMS for C_DRUG are the same for BVMIXGINAR(1) and BVNGINAR(1). The improvement with

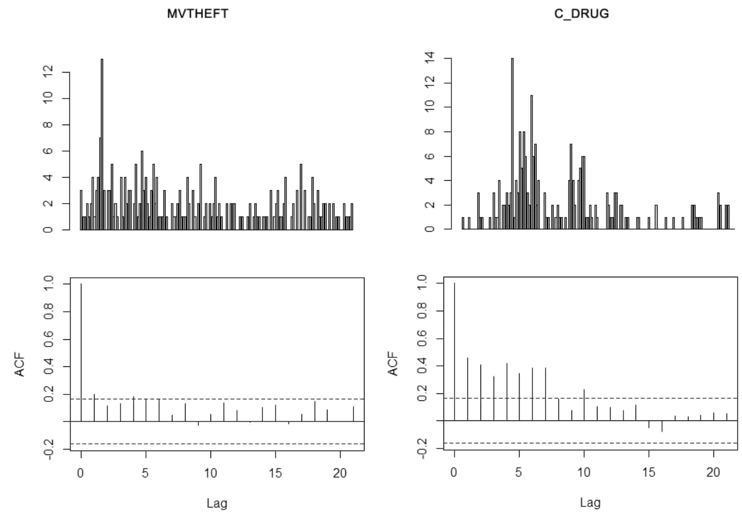

Figure 1: Data series and autocorrelation function for MVTHEFT and C_DRUG series

Model	CML estimates	log-likelihood	RMS MVTHEFT	RMS C_DRUG
BVMIXGINAR(1)	$\widehat{\alpha_1} = 0.148(0.1)$, $\widehat{\alpha_2} = 0.472(0.095)$, $\hat{p} = 0.469(0.325)$, $\hat{q} = 0.02(0.08)$, $\hat{\mu} = 1.598(0.164)$	-477.51	1.695	1.990
BVNGINAR(1)	$\widehat{\alpha_1} = 0.242(0.166)$, $\widehat{\alpha_2} = 0.473(0.095)$, $\hat{p} = 0.39(0.244)$, $\hat{q} = 0.02(0.078)$, $\hat{\mu} = 1.604(0.171)$	-478.5	1.703	1.990
BVPOINAR(1)	$\widehat{\alpha_1} = 0.217(0.066)$, $\widehat{\alpha_2} = 0.41(0.055)$, $\hat{p} = 0.328(0.172)$, $\hat{q} = 0.068(0.066)$, $\hat{\lambda} = 1.687(0.11)$	-516.62	1.701	1.998

Table 1: Parameter estimates of INAR models, root mean square errors and the log-likelihood function for MVTHEFT and C_DRUG data series

BVMIXGINAR(1) is with RMS for MVTHEFT. The assumption that one survival process evolves under binomial and the other survival process under negative binomial thinning improves prediction performance. We need this mix of thinning operators when we model two series with different behavior, as the case here. Since once sold drugs are often resold, but once stolen vehicle cannot be stolen again, we have here one process that is self-generated and one that is not.

The estimated parameters of BVMIXGINAR(1) indicate that drug activities influence the number of stolen vehicles in this area, while vice versa does not hold since the value of parameter q is statistically equal to zero.

We continue with the prediction performance analysis by focusing on the prediction errors made by the survival and innovation components separately. We calculate these residuals and plot them to assess the adequacy of each component. As given by equation (4.3), the sum of these two residuals is equal to the residuals obtained by the usual definition. The residuals are presented in Figure 2. It can be noticed that the residuals of the innovation processes are much higher than the residuals of the survival processes, apart from the few cases of C_DRUG series. Further, the correlation between the two type of residuals is 0.425 for MVTHEFT

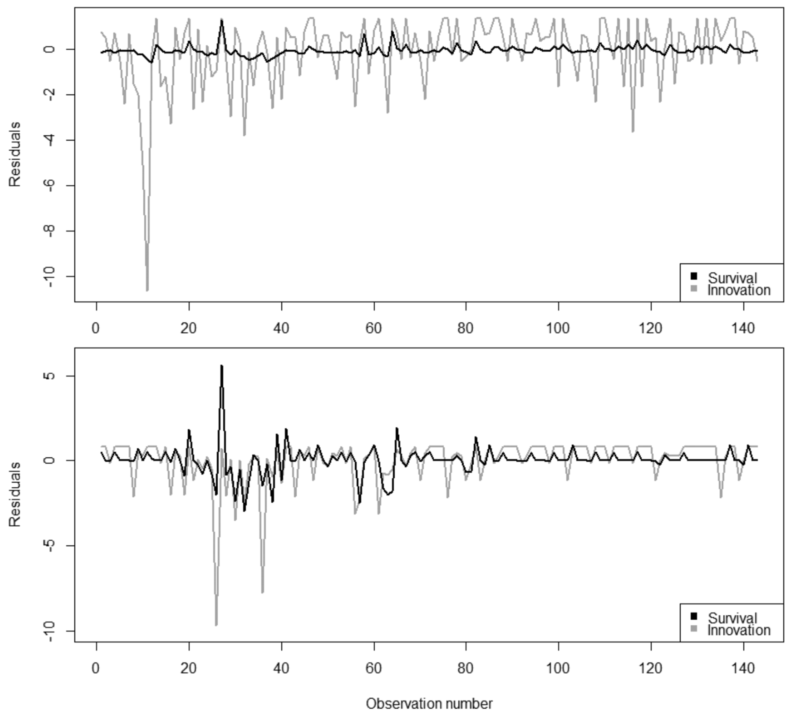

Figure 2: Upper figure shows the residuals for MVTHEFT and lower for C_DRUG series.

and 0.506 for C_DRUG series. The correlation is positive but not as high as one might expect. These results also add value to the model since an imprecise prediction of one component can be absorbed by the prediction of the other component. Another interesting point is a low correlation of only 0.11 between the innovation processes of the two series, which supports the structural assumption that innovation processes are independent. Higher residuals generated by the innovation processes indicate that future work should focus on improving the innovation processes.

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

[3] Dewald, S.L., Lewis, P.A.W., McKenzie, E. (1989) A bivariate First-Order Autoregressive Time Series Modiel in Exponential Variables (BEAR(1)), Management Science 35, 1236–1246.

[4] Franke, J., Rao, T.S. (1993) Multivariate first-order integer-valued autoregressions, Technical Report, Mathematics Department, UMIST.

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

[6] Latour, A. (1997) The Multivatiate GINAR(p) Process, Advances in Applied Probability 29, 228–248.

[7] McKenzie, E. (1985) Some simple models for discrete variate time series, Journal of the American Water Resources Association 21, 645–650.

[8] McKenzie, E. (1988) Some ARMA models for dependent sequences of Poisson counts, Advances in Applied Probability 20, 822–835.

[9] Nastić, A.S., Ristić, M.M., Popović P.M. (2016) Estimation in a Bivariate Integer-Valued Autoregressive Process, Communications in Statistics - Theory and Methods 45, 5660–5678.

[10] Pedeli, X., Karlis, D. (2011) A Bivariate INAR(1) Process with Application, Statistical Modelling 11, 325–349.

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

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

[13] Silva, N., Pereira, I., Silva, E.M. (2008) Estimation and forecasting in SUINAR (1) model, REVSTAT 6, 253–277.

[14] Silva, N., Pereira, I., Silva, E.M. (2009) Forecasting in INAR (1) model, REVSTAT 7, 119–134.

[15] Weiss, H.C. (2008) Thinning operations for modeling time series of counts–a survey, Advances in Statistical Analysis 92, 319–341.

[16] Park, Y., Kim, H.Y. (2012) Diagnostic checks for integer-valued autoregressive models using expected residuals, Statistical Papers 53, 951–971.

[17] Zheng, H., Basawa, I.V., Datta, S. (2005) Inference for p^{th}-order random coefficient integer-valued autoregressive processes, Journal Of Time Series Analysis 27, 411–440.