Hacettepe Journal of
Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 51 (5) (2022), 1458 – 1471
DOI : 10.15672/hujms.989627

Research Article

A special integer-valued bilinear time series model with applications

Sakineh Ramezani[1], Mehrnaz Mohammadpour*[2]

[1] *Mazandaran University of Medical Sciences, Sari, Iran*
[2] *Faculty of Mathematical Sciences, University of Mazandaran, Babolsar, Iran*

Abstract

The present work proposes a special integer-valued bilinear time series model based on the thinning operators. Basic probabilistic and statistical properties of this class of models are discussed. Moreover, parameter estimation methods in the time and frequency domains and forecasting are addressed. Finally, the performances of the estimation methods are illustrated through a simulation study and an empirical application to two data sets.

Mathematics Subject Classification (2020). 62M10, 62M20

Keywords. integer-valued bilinear process, sieve bootstrap prediction, spectral density function, thinning operator, Whittle criterion

1. Introduction

Bilinear models have attracted much attention and been widely used to model nonlinear phenomena in medical, economic and financial fields. The evidence of nonlinearity which is usually found in the dynamic behavior of such data imply that classical linear models are not appropriate for modelling these series. Bilinear model was first introduced by [7] with the applications to seismological and financial data. The bilinear process $\{X_t\}$ satisfies the stochastic difference equation

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=1}^{q} c_j e_{t-j} + \sum_{k=1}^{P}\sum_{l=1}^{Q} b_{kl} X_{t-k} e_{t-l} + e_t.$$

Bilinear time series models have received great attention and discussed by many authors, among them we refer to [1, 4, 12, 13, 16].

In recent years, many developments have been done in the modelling and analysis of count time series. On modelling the non-linear phenomena of counts, Doukhan et al. [5] extended the bilinear models to integer-valued bilinear model, INBL(p, q, P, Q), as

$$X_t = \sum_{i=1}^{p} a_i \circ X_{t-i} + \sum_{j=1}^{q} c_j \circ e_{t-j} + \sum_{k=1}^{P}\sum_{l=1}^{Q} b_{kl} \circ X_{t-k} e_{t-l} + e_t,$$

*Corresponding Author.
Email addresses: sakineh.ramzani@gmail.com (S. Ramezani), m.mohammadpour@umz.ac.ir (M. Mohammadpour)
Received: 02.09.2021; Accepted: 07.06.2022

where "$\circ$" is the thinning operator that is defined as

$$a \circ X = \sum_{i=1}^{X} Y_i,$$

where X is a non-negative integer valued random variable, $a > 0$ and $\{Y_i\}$ is a sequence of i.i.d non-negative integer valued random variables called counting series [11]. In the field of integer-valued time series modeling, limited research has been carried out so far to develop models to non-linear time series. Doukhan et al. [5] obtained the sufficient condition for stationarity of the INBL(1,0,1,1) process, estimated the parameters by Yule-Walker method and obtained the asymptotic distribution of estimators. Drost et al. [6] determined the existence of stationary solution for super diagonal INBL model. Recently, Bentarzi and Bentarzi [2] introduced the periodic INBL(1,0,1,1) model and Mohammadpour et al. [8] worked on the new class of INBL(1,0,1,1) model by mixing the thinning and Pegram operators.

On modelling the non-linear phenomena of counts, the worked models are only applicable to integer-valued time series with significant sample PACF only at lag 1. In this paper, we propose a new class of INBL process where the sample PACF is virtually only significant at lag 2, see Figure 1. We propose a class of INBL(2,0,2,1) model as

$$X_t = a \circ X_{t-2} + b \circ X_{t-2}e_{t-1} + e_t$$

and discuss some properties of the model such as strictly stationarity, the existence of the moments, spectral density, parameters estimation methods. We estimate the parameters via Yule-Walker, conditional least square and Whittle criterion methods. The performance of the methods is assessed by a simulation study. We apply the proposed model to two real count data sets. Consequently, forecasting of the data sets is considered.

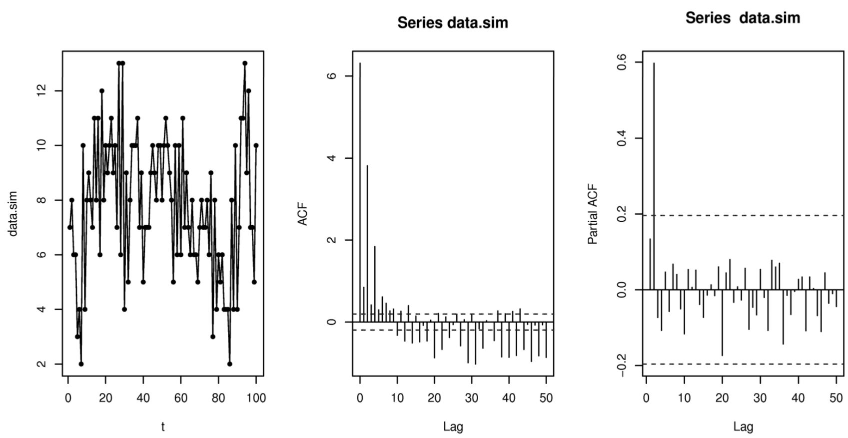

Figure 1. Sample path, sample ACF and PACF for the simulated data of the model.

2. The model and its properties

Consider the INBL(2,0,2,1) process $\{X_t\}_{t\in Z}$ as follows:

$$X_t = a \circ X_{t-2} + b \circ X_{t-2}e_{t-1} + e_t, \tag{2.1}$$

where "$\circ$" stands for the thinning operator, $\{e_t\}$ is a sequence of i.i.d non-negative integer-valued random variables with finite mean μ and finite variance σ^2 and it is independent of $\{X_s\}_{s<t}$ and $a, b \in (0,1)$. Also, the counting sequences $\{Y_i\}$ and $\{\widetilde{Y}_i\}$ associated with the operators $a\circ$ and $b\circ$ have the means a and b and variances α and β.

First, we give a sufficient condition under which the process $\{X_t\}_{t\in Z}$ is strictly stationary.

Theorem 2.1. *A sufficient condition for the INBL(2,0,2,1) process $\{X_t\}_{t\in Z}$ defined in (2.1) to be stationary is that $(a+b\mu)^2+b^2\sigma^2<1$.*

Proof. Let $\{X_t^n\}_{t\in Z}$ be a sequence of random variables as

$$X_t^n=\left\{\begin{array}{cc} 0 & n<1\\ e_t & n=1\\ a\circ X_{t-2}^{n-2}+b\circ X_{t-2}^{n-2}e_{t-1}+e_t & n>1. \end{array}\right\}$$

where sequence $\{X_s^n\}_{s<t}$ is independent of e_t.

C1. *The process $\{X_t^n\}_{t\in Z}$ is strictly stationary for any $n\in N$.*

It is enough to show that the two vectors $(X_2^n,...,X_k^n)^T$ and $(X_{2+h}^n,...,X_{k+h}^n)^T$ are identically distributed. It is clear that the process $\{X_t^n\}_{t\in Z}$ is strictly stationary for $n=1$. Now, we suppose process $\{X_t^m\}_{t\in Z}$ is strictly stationary for all $1\le m\le n-1$. Hence we have for $m=n$;

$$\begin{pmatrix}X_2^n\\ \vdots\\ X_k^n\end{pmatrix}=\begin{pmatrix}a\circ & \cdots & 0\circ\\ \vdots & \ddots & \vdots\\ 0\circ & \cdots & a\circ\end{pmatrix}\begin{pmatrix}X_0^{n-2}\\ \vdots\\ X_{k-2}^{n-2}\end{pmatrix}+\begin{pmatrix}b\circ & \cdots & 0\circ\\ \vdots & \ddots & \vdots\\ 0\circ & \cdots & b\circ\end{pmatrix}\begin{pmatrix}X_0^{n-2}e_1\\ \vdots\\ X_{k-2}^{n-2}e_{k-1}\end{pmatrix}+\begin{pmatrix}e_2\\ \vdots\\ e_k\end{pmatrix}$$

and

$$\begin{pmatrix}X_{2+h}^n\\ \vdots\\ X_{k+h}^n\end{pmatrix}=\begin{pmatrix}a\circ & \cdots & 0\circ\\ \vdots & \ddots & \vdots\\ 0\circ & \cdots & a\circ\end{pmatrix}\begin{pmatrix}X_h^{n-2}\\ \vdots\\ X_{k+h-2}^{n-2}\end{pmatrix}+\begin{pmatrix}b\circ & \cdots & 0\circ\\ \vdots & \ddots & \vdots\\ 0\circ & \cdots & b\circ\end{pmatrix}\begin{pmatrix}X_h^{n-2}e_{h+1}\\ \vdots\\ X_{k+h-2}^{n-2}e_{k+h-1}\end{pmatrix}+\begin{pmatrix}e_{h+2}\\ \vdots\\ e_{k+h}\end{pmatrix}.$$

According to the induction hypothesis and the random variables involved in the right expressions, it is deduced that the vectors $(X_2^n,...,X_k^n)^T$ and $(X_{2+h}^n,...,X_{k+h}^n)^T$ are identically distributed.

C2. *The sequence $\{X_t^n\}_{t\in Z}$ belongs to the space $\pounds^2=\{X|EX^2<\infty\}$.*

Let $\mu_n=E(X_t^n)$. We have

$$\mu_n=(a+b\mu)\mu_{n-2}+\mu=\cdots=(a+b\mu)^{[n/2]}\mu+\mu\sum_{i=1}^{[\frac{n}{2}]}(a+b\mu)^{i-1},$$

and under the strictly stationary condition, $\mu_n<\infty$ for all n. Now we show that $E(X_t^n)^2<\infty$ for all n,

$$\begin{aligned}E(X_t^n)^2 &= [(a+b\mu)^2+b^2\sigma^2]E(X_{t-2}^{n-2})^2+((\alpha+\beta\mu)+2(a+b\mu)\mu)E(X_{t-2}^{n-2})+E(e_t^2)\\ &= \Big((a+b\mu)^2+b^2\sigma^2\Big)^{[n/2]}E(e_{t-2[n/2]}^2)\\ &\quad+((\alpha+\beta\mu)+2(a+b\mu)\mu)\left[\sum_{i=1}^{[n/2]}((a+b\mu)^2+b^2\sigma^2)^{i-1}E(X_{t-2i}^{n-2i})\right.\\ &\quad\left.+E(e_t^2)\sum_{i=1}^{[n/2]}((a+b\mu)^2+b^2\sigma^2)^{i-1}\right].\end{aligned}$$

So under the strictly stationary condition, $E(X_t^n)^2<\infty$ for all n.

C3. *The sequence $\{X_t^n\}$ is Cauchy.*

Let $\Psi(t,n,m) = |X_t^n - X_t^{n-m}|$, $m = 1, 2, \ldots$. Using the definition of $\{X_t^n\}$, we have

$$
\begin{aligned}
E\Psi(t,n,m) &\le E|a \circ (X_{t-2}^{n-2} - X_{t-2}^{n-m-2})| + E|b \circ (X_{t-2}^{n-2} - X_{t-2}^{n-m-2})e_{t-1}| \\
&\le E|a \circ (X_{t-2}^{n-2} - X_{t-2}^{n-m-2})| + E|b \circ (X_{t-2}^{n-2} - X_{t-2}^{n-m-2})|E(e_{t-1}) \\
&\le |a + b\mu| E\Psi(t-2, n-2, m) = |a + b\mu|^2 E\Psi(t-4, n-4, m) \\
&\le |a + b\mu|^3 E\Psi(t-6, n-6, m) \\
&\vdots \\
&\le |a + b\mu|^{[\frac{n}{2}]} E\Psi(t - 2[\frac{n}{2}], 0, m) = |a + b\mu|^{[\frac{n}{2}]} E|X^0_{t-2[\frac{n}{2}]}| = |a + b\mu|^{[\frac{n}{2}]}\mu_{|e|}.
\end{aligned}
$$

As $n \to \infty$, under the strictly stationary condition, $E\Psi(t,n,m)$ converges to 0. Similarly, we can obtain that

$$
\begin{aligned}
E\Psi^2(t,n,m) &\le ((a + b\mu)^2 + b^2\sigma^2)E\Psi^2(t-2, n-2, m) + (\alpha + \beta\mu)E\Psi(t-2, n-2, m) \\
&\vdots \\
&\le ((a + b\mu)^2 + b^2\sigma^2)^{[\frac{n}{2}]} E\Psi^2(t - 2[\frac{n}{2}], 0, m) \\
&\quad + (\alpha + \beta\mu) \sum_{i=1}^{[\frac{n}{2}]} ((a + b\mu)^2 + b^2\sigma^2)^{i-1} E\Psi(t - 2i, n - 2i, m).
\end{aligned}
$$

As $n \to \infty$, it easily can be seen that $E\Psi^2(t,n,m)$ converges to 0. So $\{X_t^n\}$ is a Cauchy sequence, [10]. Let $X_t = \lim_{n\to\infty} X_t^n$, then $X_t \in L^2$.

C4. *The process $\{X_t\}$ satisfies Eqn. (2.1).*

Since $X_t^n \to X_t$, so by applying the properties of thinning operators we have

$$E|a \circ X_{t-2}^{n-2} - a \circ X_{t-2}|^2 = a^2 E|X_{t-2}^{n-2} - X_{t-2}|^2 + \alpha E|X_{t-2}^{n-2} - X_{t-2}|$$

and

$$E|b \circ (X_{t-2}^{n-2} - b \circ X_{t-2})e_{t-1}|^2 = b^2(\mu^2 + \sigma^2)E|X_{t-2}^{n-2} - X_{t-2}|^2 + \beta\mu E|(X_{t-2}^{n-2} - X_{t-2})|$$

converge to zero. Hence the process $\{X_t\}$ satisfies Eqn. (2.1).

C5. *Uniqueness.*

To show the uniqueness of the process $\{X_t\}$, assume that there exists another process $\{X_t^*\}$ such that $X_t^{(n)} \to X_t^*$. By Minkowski inequality, we have

$$E^{1/2}(|X_t - X_t^*|^2) \le E^{1/2}(|X_t^n - X_t^*|^2) + E^{1/2}(|X_t^n - X_t|^2).$$

Hence $X_t = X_t^*$ *a.s.*

C6. *Strictly stationarity.*

Since the process $\{X_t^n\}$ is strictly stationary and $X_t^n \xrightarrow{L^2} X_t$, using Cramer-Wold device [3], it can be concluded that

$$(X_0^n, \ldots, X_k^n) \Rightarrow (X_0, \ldots, X_k)$$

and

$$(X_h^n, \ldots, X_{k+h}^n) \Rightarrow (X_h, \ldots, X_{k+h}).$$

Since $(X_0^n, ..., X_k^n)$ and $(X_h^n, ..., X_{k+h}^n)$ have the same distribution, hence $(X_0, ..., X_k)$ and $(X_h, ..., X_{k+h})$ have the same distribution. This completes the proof. □

In the following, we investigate some properties of the process $\{X_t\}_{t\in Z}$.

Theorem 2.2. *Suppose the INBL(2,0,2,1) process $\{X_t\}_{t\in Z}$ is stationary. Then we obtain the following properties*
(i) $E(X_t) = \frac{\mu}{1-(a+b\mu)}$.
(ii) $E(X_t^2) = \frac{(\mu^2+\sigma^2)+[2\mu(a+b\mu)+(\alpha+\beta\mu)]E(X_t)}{1-[(a+b\mu)^2+b^2\sigma^2]}$.
(iii) $E(X_t^p) < \infty$.

Proof. (i) The first moment obtains easily by Eqn. (2.1).
(ii) To obtain $E(X_t^2)$, we have

$$E(X_t^2) = E(S_{t-2}^2) + E(e_t^2) + 2E(S_{t-2})E(e_t),$$

where $S_{t-2} = a \circ X_{t-2} + b \circ X_{t-2}e_{t-1}$. Using the properties of the thinning operator, we have

$$E(S_{t-2}^2) = (\alpha + \beta\mu)E(X_t) + ((a + b\mu)^2 + b^2\sigma^2)E(X_t^2).$$

Therefore

$$E(X_t^2) = \frac{(\mu^2 + \sigma^2) + [2\mu(a + b\mu) + (\alpha + \beta\mu)]E(X_t)}{1 - [(a + b\mu)^2 + b^2\sigma^2]}.$$

(iii) Let $I_p = ||X_t||_p = E^{1/p}(X_t^p)$. Obviously, we have

$$I_p = ||X_t||_p \leq ||a \circ X_{t-2}||_p + ||b \circ X_{t-2}e_{t-1}||_p + ||e_t||_p.$$

Using the counting sequences $\{Y_i\}$ and $\{\tilde{Y}_j\}$ in operators $a\circ$ and $b\circ$ and the convexity of the function $f(l) = l^p$, for $p > 4$ we have [5]

$$||a \circ X_{t-2}||_p^p = E(a \circ X_{t-2})^p = E[\sum_{i=1}^{X_{t-2}} Y_i]^p \leq E(Y^p)E(X_{t-2}^p)$$

and

$$||b \circ X_{t-2}e_{t-1}||_p^p = E(b \circ X_{t-2}e_{t-1})^p = E[\sum_{i=1}^{X_{t-2}e_{t-1}} \tilde{Y}_i]^p \leq E(\tilde{Y}^p)E(X_{t-2}^p)E(e_{t-1}^p).$$

Hence

$$I_p \leq E^{1/p}(Y^p)I_p + E^{1/p}(\tilde{Y}^p)E^{1/p}(e_t^p)I_p + E^{1/p}(e_t^p).$$

Therefore

$$I_p \leq \frac{E^{1/p}(e_t^p)}{1 - (E^{1/p}(Y^p) + E^{1/p}(\tilde{Y}^p)E^{1/p}(e_t^p))},$$

which completes the proof. □

Theorem 2.3. *The autocovariance functions of INBL(2,0,2,1) process are obtained as follows:*

$$\gamma_X(1) = \frac{b\sigma^2\mu}{(1 - (a + b\mu))^2}, \tag{2.2}$$

$$\gamma_X(2h) = (a + b\mu)^h\gamma_X(0), \tag{2.3}$$

and

$$\gamma_X(2h + 1) = (a + b\mu)^h\gamma_X(1).$$

Proof. Multiplying (2.1) by X_t and $X_{t-1}e_t$, taking expectation on both sides and using the stationarity of the process, we obtain

$$E(X_{t+1}X_t) = aE(X_tX_{t-1}) + bE(X_tX_{t-1}e_t) + E(X_t)E(e_{t+1}) \tag{2.4}$$

and

$$(1-b\mu)E(X_tX_{t-1}e_t) = a\mu E(X_{t-1}X_{t-2}) + E(X_t)E(e_t^2). \tag{2.5}$$

Substituting (2.5) to (2.4), we get

$$(1-b\mu)E(X_{t+1}X_t) = a(1-b\mu)E(X_tX_{t-1}) + ab\mu E(X_{t-1}X_{t-2}) + bE(e_t^2)E(X_t) + \mu(1-b\mu)E(X_t).$$

So

$$(1-(a+b\mu))E(X_{t+1}X_t) = bE(e_t^2)E(X_t) + \mu(1-b\mu)E(X_t).$$

Consequently, by a simple substitution and using $E(X_t)$ in Theorem 2.1, we have the above expression for $\gamma_X(1)$. Also $\gamma_X(2h)$ and $\gamma_X(2h+1)$ will be obtained in a similar manner. □

By applying the autocovariance function, the spectral density function of the process $\{X_t\}_{t\in Z}$ which is applied in frequency domain approaches obtained as follows:

$$\begin{aligned} f_X(\omega) &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_X(k)e^{-i\omega k} = \frac{1}{2\pi}(\gamma_X(0) + \sum_{k\neq 0}\gamma_X(k)e^{-i\omega k}) && (2.6)\\ &= \frac{1}{2\pi}\frac{(1-(a+b\mu))[\gamma_X(0)(1+(a+b\mu)) + 2\gamma_X(1)\cos\omega]}{1+(a+b\mu)^2 - 2(a+b\mu)\cos 2\omega}. && (2.7)\end{aligned}$$

The conditional expectations of the process are obtained as follows:

$$\begin{aligned} E(X_{t+1}|t) &= (a+be_t)X_{t-1} + \mu\\ E(X_{t+2}|t) &= (a+b\mu)X_t + \mu,\end{aligned}$$

where $E(.|t)$ is the conditional expectation with respect to the σ-field $\mathcal{F}_t^e$ which is generated by random variables $e_s, s \le t$. By induction, we can conclude that

$$\begin{aligned} E(X_{t+k}|t) &= (a+b\mu)^{k/2}X_t + \mu\sum_{i=1}^{k/2}(a+b\mu)^{i-1}, && k=2h\\ E(X_{t+k}|t) &= (a+b\mu)^{[k/2]}E(X_{t+1}|t) + \mu\sum_{i=1}^{[k/2]}(a+b\mu)^{i-1}, && k=2h+1. \quad (2.8)\end{aligned}$$

Also

$$\begin{aligned} E(X_{t+1}^2|t) &= (a+be_t)^2X_{t-1}^2 + [(\alpha+\beta e_t) + 2\mu(a+be_t)]X_{t-1} + (\mu^2+\sigma^2),\\ E(X_{t+2}^2|t) &= [(a+b\mu)^2 + b^2\sigma^2]X_t^2 + [(\alpha+\beta\mu) + 2\mu(a+b\mu)]X_t + (\mu^2+\sigma^2),\end{aligned}$$

and for even and odd k, respectively

$$E(X_{t+k}^2|t) = H^{k/2}X_t^2 + [(\alpha+\beta\mu) + 2\mu(a+b\mu)]\textstyle\sum_{i=1}^{k/2} H^{i-1}E(X_{t+k-2i}|t) + (\mu^2+\sigma^2)\sum_{i=1}^{k/2}(H^{i-1}),$$

and

$$E(X_{t+k}^2|t) = H^{[k/2]}E(X_{t+1}^2|t) + [(\alpha+\beta\mu) + 2\mu(a+b\mu)]\textstyle\sum_{i=1}^{[k/2]} H^{i-1}E(X_{t+k-2i}|t) + (\mu^2+\sigma^2)\sum_{i=1}^{[k/2]}(H^{i-1}), \tag{2.9}$$

where $H = [(a+b\mu)^2 + b^2\sigma^2]$.

Note 1. From Eqns. (2.8) and (2.9), $\lim_{k\to\infty} E(X_{t+k}|t) \to E(X_t)$ and $\lim_{k\to\infty} E(X_{t+k}^2|t) \to E(X_t^2)$.

3. Estimation and simulation

In this section, we will investigate several methods for parameter estimation of the stationary INBL(2,0,2,1) model based on a realization $X_1, ..., X_n$ of this process. In the estimation procedure, we assume that the distribution of $\{e_t\}$ is Poisson with parameter μ.

3.1. Yule-Walker method

We investigate the Yule-Walker (YW) estimators of the unknown parameters a, b and μ. Using Theorem 2.2, Eqns. (2.2) and (2.3) and also the sample first moment ($\bar{X}$), second moment ($\bar{X}^2$) and autocovariance function ($\hat{\gamma}_X(\cdot)$), the YW estimators are obtained as

$$\begin{aligned}
\hat{b}_{YW} &= \frac{\hat{\gamma}_X(1)}{\bar{X}^2} \\
\hat{\mu}_{YW} &= \bar{X}(1-\hat{M}) \\
\hat{a}_{YW} &= \hat{M} - \hat{b}\hat{\mu},
\end{aligned}$$

where $\hat{M} = \frac{\hat{\gamma}_X(2)}{\hat{\gamma}_X(0)}$.

The asymptotic distribution of the YW estimators $\hat{a}_{YW}, \hat{b}_{YW}$ and $\hat{\mu}_{YW}$ is

$$\left(\hat{a}_{YW}, \hat{b}_{YW}, \hat{\mu}_{YW}\right) \sim N((a,b,\mu), n^{-1}\mathbf{V}\Sigma\mathbf{V}^T),$$

where

$$\mathbf{V}^T = \begin{bmatrix}
\frac{\gamma_X(1)}{\mu^2}(1-\frac{\gamma_X(2)}{\gamma_X(0)}) & \frac{-2\gamma_X(1)}{\mu^3} & 1-\frac{\gamma_X(2)}{\gamma_X(0)} \\
-\frac{\gamma_X(2)}{\gamma_X^2(0)}[\frac{\gamma_X(1)}{\mu}+1] & 0 & \frac{\gamma_X(2)\mu}{\gamma_X^2(0)} \\
-\frac{1}{\mu}(1-\frac{\gamma_X(2)}{\gamma_X(0)}) & \frac{1}{\mu^2} & 0 \\
\frac{1}{\gamma_X(0)}(1+\frac{\gamma_X(1)}{\mu}) & 0 & -\frac{\mu}{\gamma_X(0)}
\end{bmatrix},$$

and Σ is the variance-covariance matrix of the random vector $\left(\bar{X}, \hat{\gamma}_X(0), \hat{\gamma}_X(1), \hat{\gamma}_X(2)\right)$, for more details of the matrix Σ see [5].

3.2. Conditional least squares method

The conditional least squares (CLS) estimators of the parameters a, b and μ are obtained by minimizing the squares of the conditional errors

$$Q(\Theta) = \sum_{t=3}^{n}(X_t - (a+be_{t-1})X_{t-2} - \mu)^2,$$

where $\Theta = (a, b, \mu)$ and $e_t = X_t - aX_{t-2} - bX_{t-2}e_{t-1}$. The minimization is achieved through Newton Raphson (NR) algorithm as

$$\theta^{(j+1)} = \theta^{(j)} - S^{-1}(\theta^{(j)})G(\theta^{(j)}),$$

where G and S are Gradient vector and Hessian matrix, respectively. Also, the partial derivations of Q with respect to the parameters are as

$$\begin{aligned}
\frac{\partial Q}{\partial a} &= 2\sum_{t=3}^{n}(e_t-\mu)(-X_{t-2} - bX_{t-2}\frac{\partial e_{t-1}}{\partial a}) \\
\frac{\partial Q}{\partial b} &= 2\sum_{t=3}^{n}(e_t-\mu)(-X_{t-2}e_{t-1} - bX_{t-2}\frac{\partial e_{t-1}}{\partial b}) \\
\frac{\partial Q}{\partial \mu} &= 2\sum_{t=3}^{n}(e_t-\mu)(-1 - bX_{t-2}\frac{\partial e_{t-1}}{\partial \mu})
\end{aligned}$$

and

$$\frac{\partial^2 Q}{\partial\theta_i\partial\theta_j} = 2\sum_{t=3}^{n}\frac{\partial e_t}{\partial\theta_i}\frac{\partial e_t}{\partial\theta_j} + 2\sum_{t=3}^{n}\frac{\partial^2 e_t}{\partial\theta_i\partial\theta_j}.$$

Since that the terms e_1 and e_2 can not be computed, without loss of generality we can choose the conditions $e_1 = e_2 = 1$ and $\frac{\partial e_t}{\partial\theta_i} = \frac{\partial^2 e_t}{\partial\theta_i\partial\theta_j} = 0$ for $t = 1, 2$ and $i, j = 1, 2, 3$. The NR algorithm requires initial values to estimate the parameters. The CLS or YW estimates of INAR(2) model is used as an initial estimates, because the structure of INBL(2,0,2,1) model in terms of its ACF and PACF is similar to INAR(2) model. Given the database of X_t, the iteration procedure can be continued until $|\theta_i^{(n+1)} - \theta_i^{(n)}| < 10^{-3}$.

Note 2. The CLS estimators of the parameters $M = (a + b\mu)$ and μ have the closed form by minimizing the squares of the conditional errors as

$$Q(a, b, \mu) = \sum_{i=3}^{n}(X_t - E(X_t|t-2)) = \sum_{i=3}^{n}(X_t - (a+b\mu)X_{t-2} - \mu)^2$$

and they will be derived as the following form:

$$\hat{M}_{CLS} = \frac{(n-2)\sum_{t=3}^{n} X_t X_{t-2} - \sum_{t=3}^{n} X_t \sum_{t=3}^{n} X_{t-2}}{(n-2)\sum_{t=3}^{n} X_{t-2}^2 - (\sum_{t=3}^{n} X_{t-2})^2}$$

and

$$\hat{\mu}_{CLS} = \frac{\sum_{t=3}^{n} X_t - \hat{M}_{CLS}\sum_{t=3}^{n} X_{t-2}}{n-2}.$$

Theorem 3.1. *The CLS and YW estimators of the parameters M and μ are asymptotically equivalent.*

Proof. The proof follows by showing the two conditions

$$\hat{\mu}_{CLS} - \hat{\mu}_{YW} = o(n^{-\frac{1}{2}})$$

$$\hat{M}_{CLS} - \hat{M}_{YW} = o(n^{-\frac{1}{2}}).$$

Using the definition of $\hat{\mu}_{CLS}$ and $\hat{\mu}_{YW}$, we have

$$\lim_{n\to\infty}[\sqrt{n-2}(\hat{\mu}_{CLS} - \hat{\mu}_{YW})] = \lim_{n\to\infty}[\frac{M(X_{n-1}+X_n) - (X_1+X_2)}{\sqrt{n-2}}] = 0,$$

and

$$\begin{aligned}\lim_{n\to\infty}[\sqrt{n}(\hat{M}_{CLS} - \hat{M}_{YW})] &= \lim_{n\to\infty}\frac{H(X_n, X_{n-1}, \bar{X})}{\sqrt{n}}\frac{\sum_{t=3}^{n} X_t X_{t-2} - \frac{\sum_{t=3}^{n} X_t}{n-2}\frac{\sum_{t=3}^{n} X_{t-2}}{n-2}}{[\frac{\sum_{t=3}^{n} X_{t-2}^2}{n-2} - (\frac{\sum_{t=3}^{n} X_{t-2}}{n-2})^2][\frac{\sum_{t=1}^{n} X_t^2}{n} - \bar{X}^2]} \\ &= 0,\end{aligned}$$

where $H(X_n, X_{n-1}, \bar{X}) = X_n^2 + X_{n-1}^2 - 2(X_n + X_{n-1})\bar{X} + \frac{(X_n+X_{n-1})^2}{n}$, which give the results. □

3.3. Whittle method

We describe a frequency domain estimation procedure based on Whittle criterion. This approach was originally proposed by [14,15]. The main motivation for the Whittle criterion is the fact that the spectral density function of a process may not be easy obtain whereas an exact likelihood. Now for INBL(2,0,2,1) process, we obtain the Whittle (W) estimator of parameters $\Theta = (a, b, \mu)$ by minimizing the function

$$\hat{l}_n(\Theta) = \frac{1}{n}\sum_{j=1}^{[n/2]}(log f_X(\omega_j) + \frac{I_n(\omega_j)}{f_X(\omega_j)}),$$

where $f_X(\omega_j)$ is the spectral density function of INBL(2,0,2,1) process in (2.6), $\omega_j = \frac{2\pi j}{n}$ is the Fourier frequency and $I_n(\omega_j)$ is the periodogram. The numerical minimization is achieved using the function nlm in statistical package. The initial values of nlm function is obtained by the YW estimates.

3.4. Simulation study

The performance of the estimates is checked by a small Monte Carlo simulation using different sample sizes over 1000 replications. In the simulation procedure, we assume that "$\circ$" is the binomial thinning operator. Based on these simulations, Table 1 gives the bias and mean square error (MSE) for the estimates for different values of the parameters a, b and μ with different sample sizes. Based on Table 1, we find that increasing the sample size implies smaller MSE and Whittle estimates converge faster to the true values of the parameters.

4. Applications

Here, we investigate an application for the INBL(2,0,2,1) process by using two real count data that obtained from the Forecasting Principles site†. The first data set represents the monthly counts of crimes, reported in the 21th aggregation of police car beat and the second data set represents the monthly counts of drug calls, reported in the 2508th of police carbeat in Pittsburgh. The sample paths, autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) of the two data sets are displayed in Figures 2 and 3, respectively. Moreover, the stationarity of the data series is justified by using the Phillips-Perron test. The test rejected the null hypothesis of non-stationarity, the p-value for the test for both data sets being 0.01 (with a significance level at 0.05). Also, we apply Keenan test for the non-linearity test. The p-value of the Keenan test for the data sets are 0.00015 and 0.015. Consequently, we reject linearity hypothesis with a significance level at 0.05. Therefore, the data can not be described by linear time series models. The Figures 2 and 3 suggest that INBL$(2, 0, 2, 1)$ model is appropriate for modeling the data sets. The parameters of INBL model are estimated by Whittle method. The results are presented in Table 2.

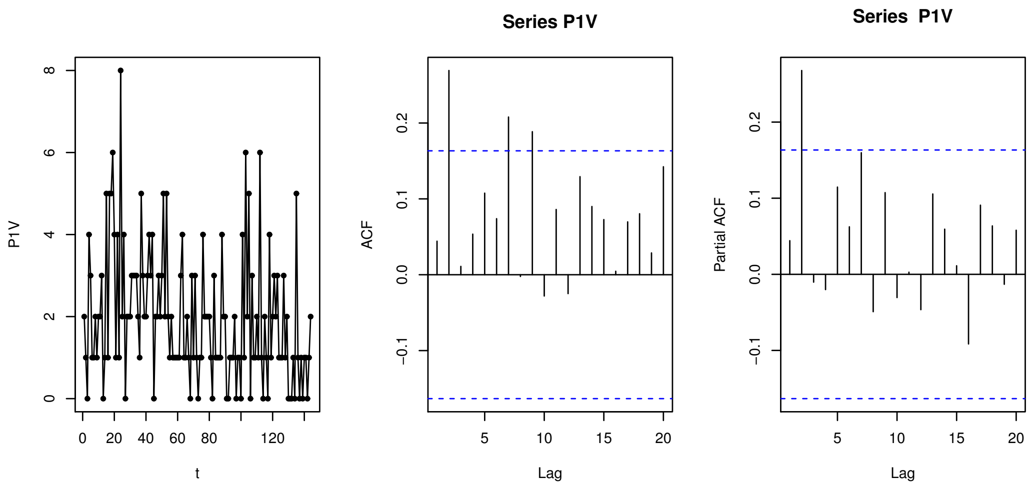

Figure 2. Sample path, sample ACF and PACF for crime data.

†http://www.forecastingprinciples.com

Table 1. The bias and MSE (in parentheses) of the YW, CLS and W estimates for true values of parameter of the model.

	$\hat{a}_{YW}$	$\hat{b}_{YW}$	$\hat{\mu}_{YW}$	$\hat{a}_{CLS}$	$\hat{b}_{CLS}$	$\hat{\mu}_{CLS}$	$\hat{a}_{W}$	$\hat{b}_{W}$	$\hat{\mu}_{W}$
n	$(a,b,\mu)=(0.2,0.1,1)$								
50	−0.028(0.012)	0.005(0.008)	0.055(0.053)	0.006(0.017)	−0.011(0.006)	0.044(0.061)	−0.021(0.002)	−0.021(0.002)	−0.002(0.000)
100	0.001(0.009)	0.002(0.006)	0.001(0.031)	0.008(0.013)	−0.021(0.003)	−0.003(0.027)	−0.022(0.002)	−0.022(0.002)	−0.002(0.000)
1000	−0.006(0.001)	0.007(0.001)	−0.001(0.003)	0.016(0.001)	−0.038(0.002)	0.013(0.003)	−0.025(0.001)	−0.025(0.001)	−0.001(0.000)
5000	0.000(0.000)	−0.001(0.000)	0.000(0.000)	0.021(0.000)	−0.032(0.001)	0.003(0.001)	−0.024(0.001)	−0.024(0.001)	−0.001(0.000)
	$(a,b,\mu)=(0.3,0.05,2)$								
50	−0.057(0.021)	−0.001(0.002)	0.101(0.209)	−0.007(0.016)	−0.011(0.002)	0.096(0.229)	−0.004(0.000)	−0.009(0.001)	0.000(0.000)
100	−0.027(0.012)	0.011(0.002)	0.051(0.112)	0.016(0.009)	−0.016(0.001)	0.047(0.11)	−0.009(0.000)	−0.017(0.001)	0.000(0.000)
1000	0.002(0.001)	−0.002(0.000)	0.013(0.012)	0.032(0.002)	−0.018(0.000)	−0.007(0.012)	−0.001(0.001)	−0.017(0.001)	0.000(0.000)
5000	0.001(0.000)	−0.001(0.000)	0.004(0.002)	0.032(0.001)	−0.018(0.000)	−0.006(0.002)	−0.000(0.000)	−0.015(0.000)	0.000(0.000)
	$(a,b,\mu)=(0.4,0.1,3)$								
50	−0.123(0.037)	−0.005(0.003)	0.653(1.223)	0.021(0.021)	−0.017(0.003)	0.117(0.611)	−0.016(0.000)	−0.047(0.003)	−0.001(0.000)
100	−0.057(0.021)	−0.004(0.004)	0.441(0.991)	0.024(0.007)	−0.013(0.001)	0.039(0.332)	−0.011(0.000)	−0.032(0.002)	−0.001(0.000)
1000	−0.012(0.002)	−0.002(0.000)	0.091(0.012)	0.000(0.000)	0.000(0.000)	0.000(0.000)	−0.006(0.000)	−0.019(0.000)	−0.001(0.000)
5000	−0.000(0.000)	−0.000(0.000)	0.013(0.015)	0.000(0.000)	0.000(0.000)	0.000(0.000)	−0.006(0.000)	−0.017(0.000)	−0.000(0.000)
	$(a,b,\mu)=(0.6,0.005,3)$								
50	−0.187(0.053)	0.026(0.001)	0.481(0.881)	−0.073(0.021)	0.013(0.001)	0.312(0.889)	0.002(0.000)	0.009(0.000)	0.000(0.000)
100	−0.112(0.018)	0.017(0.000)	0.291(0.472)	−0.039(0.009)	0.005(0.000)	0.171(0.432)	0.002(0.000)	0.006(0.000)	0.000(0.000)
1000	−0.015(0.001)	0.004(0.000)	0.021(0.051)	−0.001(0.001)	0.000(0.000)	0.015(0.039)	−0.002(0.000)	−0.003(0.0001)	0.000(0.000)
5000	−0.004(0.000)	−0.001(0.000)	0.005(0.007)	0.003(0.000)	−0.001(0.000)	0.014(0.008)	0.000(0.000)	0.001(0.000)	0.000(0.000)
	$(a,b,\mu)=(0.6,0.05,3)$								
50	−0.184(0.051)	0.236(0.003)	0.162(1.511)	0.039(0.013)	−0.027(0.002)	0.484(1.565)	−0.008(0.000)	−0.027(0.001)	−0.006(0.000)
100	−0.091(0.023)	0.012(0.001)	0.362(0.821)	0.048(0.012)	−0.029(0.001)	0.221(1.011)	−0.008(0.000)	−0.025(0.001)	0.000(0.000)
1000	−0.012(0.001)	0.000(0.000)	0.082(0.079)	0.088(0.009)	−0.032(0.001)	0.031(0.142)	−0.005(0.000)	−0.016(0.000)	0.000(0.000)
5000	−0.003(0.000)	0.000(0.000)	0.014(0.021)	0.081(0.008)	−0.028(0.001)	−0.053(0.091)	0.000(0.000)	−0.151(0.000)	0.000(0.000)
	$(a,b,\mu)=(0.8,0.005,5)$								
50	−0.362(0.145)	0.617(0.004)	0.269(1.606)	−0.052(0.014)	0.006(0.007)	0.741(3.101)	0.000(0.000)	0.000(0.000)	0.000(0.000)
100	−0.172(0.039)	0.029(0.001)	0.281(1.612)	−0.005(0.001)	0.000(0.000)	0.221(1.242)	0.000(0.000)	0.000(0.000)	0.000(0.000)
1000	−0.023(0.001)	0.003(0.000)	0.169(0.271)	0.004(0.000)	−0.003(0.000)	0.193(0.412)	0.000(0.000)	0.000(0.000)	0.000(0.000)
5000	−0.003(0.000)	0.000(0.000)	0.034(0.062)	0.014(0.000)	−0.003(0.000)	0.061(0.0513)	0.000(0.000)	0.000(0.000)	0.000(0.000)

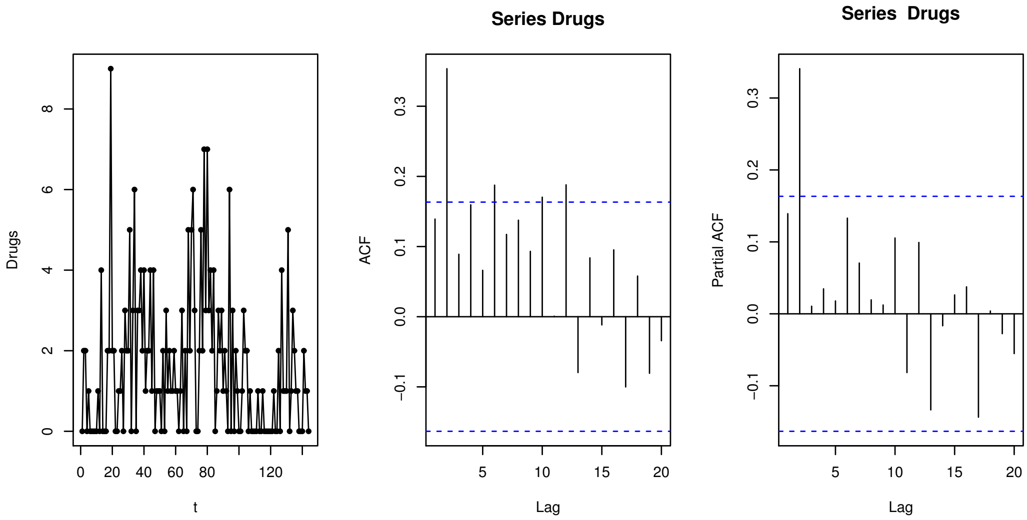

Figure 3. Sample path, sample ACF and PACF for drug call data.

Table 2. Estimated parameters for the two data sets.

Model	$\hat{a}_W$	$\hat{b}_W$	$\hat{\mu}_W$
Crime data	0.213	0.008	1.419
Drug call data	0.152	0.148	1.019

4.1. The residual analysis and model adequacy

In the following, we assess the adequacy of the model proposed and fitted to data. Tools in the diagnostic checking of dynamic structure can be based on the standardized Pearson residuals given by

$$r_t = \frac{X_t - E(X_t|t-1)}{\sqrt{Var(X_t|t-1)}},$$

where the population quantities are replaced by their estimated counterparts in INBL(2,0,2,1) model. The residual analysis are shown in Figures 4 and 5. We depict the sample ACF and cumulative periodogram plots of the Pearson residuals for both data series. Figures 4 and 5, plots the ACF of the Pearson residuals. There is no evidence of any correlation within the residuals and findings are confirmed by p-values 0.22 and 0.65 of Ljung-Box test for both data sets. Based on cumulative periodogram plots in Figures 4 and 5, it is clearly shown that the residuals are randomly distributed and do not have specified trend. The right panel in the figures shows the results of the parametric resampling method. We generate 1000 artificial data sets of length 144 using the fitted INBL(2,0,2,1) model with Poisson innovations. Based on these bootstrap data sets, 1000 autocorrelation functions are computed. For each fixed lag of the ACF, $100(1-\alpha/2)\%$ and $100(\alpha/2)\%$ ($\alpha = 0.05$) quantiles of the ACFs are obtained to constitute the bounds of an acceptance region. These bounds are shown as "$+$" symbols in the figure with the sample ACF presented by "o". Based on Figures 4 and 5, the adequacy of the fitted model is concluded.

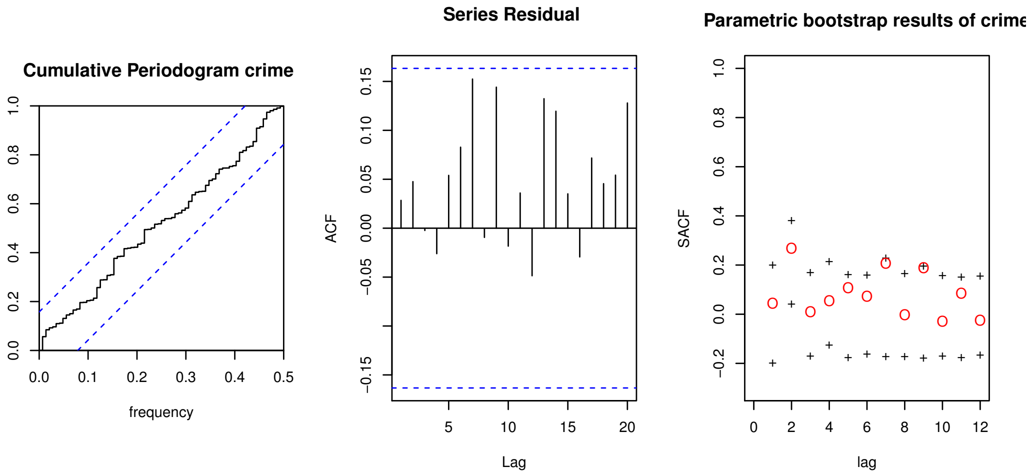

Figure 4. Cumulative periodogram, ACF plots of the Pearson residuals and parametric bootstrap result for crime data.

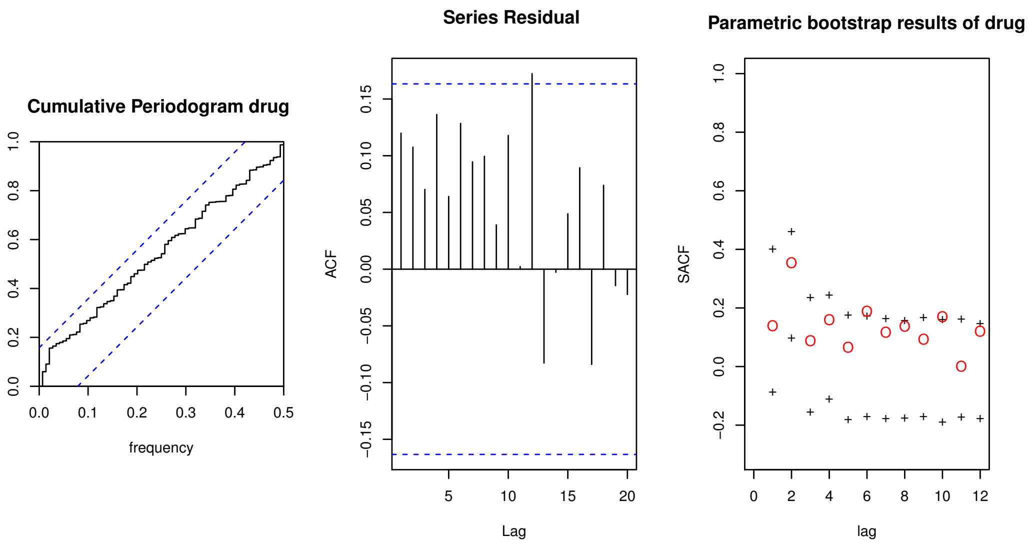

Figure 5. Cumulative periodogram, ACF plots of the Pearson residuals and parametric bootstrap result for drug data.

4.2. Forecasting

Here, we discuss the classical and bootstrap methods for forecasting the model.

4.2.1. Classical prediction. One of the most common procedures for classical prediction in time series models is to use conditional expectation. Based on (2.8), the k-step ahead predictors are given by

$$\hat{X}_{t+1} = (a + be_t)X_{t-1} + \mu$$

and

$$\hat{X}_{t+k} = (a + b\mu)\hat{X}_{t+k-2} + \mu.$$

In practice, the parameters a, b and μ are replaced by their corresponding Whittle estimates.

4.2.2. Sieve bootstrap prediction. The classical predictions do not preserve the integer-valued nature of the data in generating forecasts when the time series is integer-valued. To preserve the integer-valued nature of data, we use a bootstrap approach as a distribution free alternative. Some bootstrap approaches have been proposed. Among others, we employ the bootstrap method proposed by [9] after some modifications to INBL(2,0,2,1) model as the following steps.

(1) Estimate the parameters (a, b, μ) using Whittle method.
(2) Compute residuals $\hat{e}_t = x_t - \hat{a}x_{t-2} - \hat{b}x_{t-2}\hat{e}_{t-1}$ for $t = 3, ..., n$.
(3) Construct the empirical distribution for modified residuals $\tilde{e}_t$ defined by $\tilde{e}_t = [\hat{e}_t]$ as $\hat{e}_t > 0$ else $\tilde{e}_t$ equal zero. Also $[\cdot]$ represents the value rounded to the nearest integer.
(4) For $b = 1, ..., B$ define the bootstrap series X_t^b by

$$X_t^b = \hat{a}\circ X_{t-2}^b + \hat{b}\circ X_{t-2}^b e_{t-1}^b + e_t^b,$$

where e_t^b for $t = 1, 2, ..., n$ is an i.i.d. sample from the residuals computed previously.
(5) Based on X_t^b, compute the Whittle estimates of the parameters $(\hat{a}^b, \hat{b}^b, \widehat{\mu}^b)$ as in step 1.
(6) Estimates of (a, b, μ_e) can be obtained using sample mean: $\hat{a}^* = \frac{\sum_{b=1}^{B} \hat{a}^b}{B}$, $\hat{b}^* = \frac{\sum_{b=1}^{B} \hat{b}^b}{B}$ and $\mu^* = \frac{\sum_{b=1}^{B} \widehat{\mu}^b}{B}$.
(7) Compute future observations from k-step ahead prediction by recursion;

$$X_{t+k}^b = \hat{a}^b\circ X_{t+k-2}^b + \hat{b}^b\circ X_{t+k-2}^b e_{t+k-1}^b + e_{t+k}^b. \qquad k \geqslant 1$$

The classic and sieve bootstrap forecasts of the real data series are presented in Table 3, where the Whittle parameter estimates of the data sets are given in Table 2.

Table 3. k-step ahead predictions of the two data sets.

	Crime data			Drug call data		
k	Data	Classic	Bootstrap	Data	Classic	Bootstrap
1	5	1.630	0	2	1.599	0
2	1	1.419	1	1	1.951	0
3	0	1.773	1	1	1.535	4
4	1	1.727	2	0	1.640	2
5	0	1.804	2	0	1.516	0
6	1	1.794	2	0	1.548	1
7	1	1.810	1	2	1.511	1
8	0	1.808	0	1	1.520	1
9	1	1.812	1	1	1.509	1
10	2	1.811	3	0	1.512	0
	SMAPE	0.971	0.773		1.038	0.986

In order to evaluate and compare the different prediction methodologies, we compute k-step ahead predictions ($k = 1, 2, ..., 10$) and use the symmetric mean absolute percentage error (SMAPE) as follows:

$$\text{SMAPE} = \frac{1}{H}\sum_{k=1}^{H} \frac{2 \mid X_{n+k} - \hat{X}_{n+k} \mid}{X_{n+k} + \hat{X}_{n+k}},$$

where H represents the number of predictions realized. The classic and sieve bootstrap predictions are presented in Table 3. As it expects, it can be noted that SMAPE values of bootstrap predictions are smaller than classic predictions, also bootstrap predictors are integer, same as the nature of the real data.

5. Conclusion

In this paper, we have considered INBL(2,0,2,1) model for modelling nonlinear counts. The practical prominence of the model was also confirmed by two real data sets. As the classical predictions do not coherently preserve the integer nature of the data, the bootstrap approach is also used as a distribution free alternative.

References

[1] B. Basrak, R.A. Davis and T. Mikosch, *The sample ACF of a simple bilinear process*, Stoch. Process. Their Appl. **83** (9), 1-14, 1999.

[2] M. Bentarzi and W. Bentarzi, *Periodic integer-valued bilinear time series model*, Comm. Statist. Theory Methods **46** (3), 1184-1201, 2017.

[3] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*, 2nd ed., Springer, 1991.

[4] R.A. Davis and S.I. Resnick, *Limit theory for bilinear processes with heavy-tailed noise*, Ann. Appl. Probab. **6** (4), 1191-1210, 1996.

[5] P. Doukhan, A. Latour and D. Oraichi, *Simple integer-valued bilinear time series model*, Adv. in Appl. Probab. **38** (2), 559-578, 2006.

[6] F.C. Drost, R. van den Akker and B.J.M. Werker, *Note on integer-valued bilinear time series*, Statist. Probab. Lett. **38** (8), 559-578, 2008.

[7] C.W.J. Granger and A.P. Andersen, *An Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Gottingen, 1978.

[8] M. Mohammadpour, H.S. Bakouch and S. Ramzani, *An integer-valued bilinear time series model via two random operators*, Math. Comput. Model. Dyn. Syst. **25** (4), 429-446, 2019.

[9] L. Pascual, J. Romo and E. Ruiz, *Bootstrap predictive inference for ARIMA processes*, J. Time Series Anal. **25** (4), 449-465, 2004.

[10] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, New York, 1976.

[11] F.W. Steutel and K. van Harn, *Discrete analogues of self-decomposability and stability*, Ann. Probab. **7** (5), 893-899, 1979.

[12] T. Subba Rao, *On the theory of bilinear time series models*, J. R. Stat. Soc. Ser. B. Stat. Methodol. **43** (2), 244-255, 1981.

[13] K.F. Turkman and M.A.A. Turkman, *Extremes of bilinear time series models*, J. Time Series Anal. **18** (3), 305-319, 1997.

[14] P. Whittle, *Estimation and information in stationary time series*, Ark. Mat. **2** (5), 423-434, 1953.

[15] P. Whittle, *A Study in the Analysis of Stationary Time Series*, Almquist and Wiksell, Stockholm, 1954.

[16] Z. Zhang and H. Tong, *Some distributional properties of a first-order nonnegative bilinear time series model*, J. Appl. Probab. **38** (3), 659-671, 2001.