

PAPER • OPEN ACCESS

Outliers detection in INAR (1) time series

To cite this article: Hua Shang and Beibei Zhang 2018 *J. Phys.: Conf. Ser.* **1053** 012094

View the [article online](#) for updates and enhancements.

Related content

- [The multiple outliers detection using agglomerative hierarchical methods in circular regression model](#)
Siti Zanariah Satari, Nur Faraidah
Muhammad Di and Roslinazairimah
Zakaria
- [An Illustration of New Methods in Machine Condition Monitoring, Part II: Adaptive outlier detection](#)
I. Antoniadou, K. Worden, S. Marchesiello
et al.
- [Outlier detection and classification in sensor data streams for proactive decision support systems](#)
M V Shcherbakov, A Brebels, N.L.
Shcherbakova et al.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices
to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of
every title for free.

Outliers detection in INAR (1) time series

Hua Shang^{1,3} and Beibei Zhang²

¹College of Science, North China Institute of Science and Technology, Hebei, 065201

²School of Statistics, Capital University of Economics and Business, Beijing, 100070

³Email: 514741255@qq.com

Abstract. This article considers the problem of detecting outliers of discrete time series. The underlying process is modeled as Poisson INAR (1) models, a special class of integer-valued autoregressive process. Based on this model, we develop a Gibbs sampling algorithm to estimate the model parameters and the size of outliers. This current method can detect outliers and specify the type of outliers simultaneously, without any additional computation to distinguish between additive outliers (AO) and innovational outliers (IO). The approach is illustrated with simulation studies as well as a real data analysis.

1. Introduction

The detection of outliers in time series is an important issue because their presence may have serious effects on the analysis in many different ways. For instance, outliers can lead to biased parameter estimation and poor forecasts, even though the time series model is well specified. Fox [1] (1972) first introduced two type of outliers into time series models, namely, additive outliers (AO) and innovational outliers (IO). Other references on this topic include Tsay [2] (1986), Muirhead [3] (1986), Chang, Tiao and Chen [4] (1988), Chen and Liu [5] (1993), Justel [6] (2001), Galeano, Peña, and Tsay [7] (2006), and among others.

Most of these methods are developed for continuous variables. However, many interesting empirical questions can be recorded by count time series, which are measured in various disciplines whenever a number of events are counted during certain time periods. Examples can be found in a great variety of contexts: the weekly number of the incidence of a disease, the monthly number of car accidents in a region, the number of transactions at a stock market per minute in finance and so on.

Outlier detecting of count time series is not trivial. Count time series are positive and typically right-skewed, causing a need for especially designed models and procedures. The general methods of detecting and removing outliers are not suitable, since they often lead to non-integer values. Hence, special outlier detecting procedures for discrete time series are motivated to account for the discrete nature of certain data sets.

However, to the best of our knowledge, such studies for integer valued-dependent data are missing, although their development is important for inference and diagnostics. Only in recent years, the attention of researchers has shifted to the detection of outliers in count time series, which is also our research interest. Fokianos and Fried [8] (2010) suggested detection and estimation method of various types of outliers in the framework of integer-valued GARCH model. Fried, Liboschik, Elsaied, Kitromilidou and Fokianos [9] (2014) proposed a robust fitting of count time series following generalized linear models in the presence of outliers and intervention effects. Kitromilid and Fokianos



[10] (2016) developed a robust estimation of a log-linear Poisson model for count time series under three different type of outliers: additive outliers (AOs), transient shifts (TSs) and level shifts (LSs).

The Integer-valued AutoRegressive model of order 1 (INAR (1)), is one of the most successful integer-valued time series models proposed in the existing literatures. In fact, this model introduced independently by Al-Osh and Alzaid [11] (1987) and McKenzie [12] (1985) to model count time series, has been extensively studied and applied to many real-world problems. Here, we will focus on outliers detection in INAR (1) time series model. Several procedures are available in the literature to handle outliers in INAR (1) model. Barczy [13][14] (2010, 2012) considered CLS estimation of the parameters of an INAR (1) model contaminated, at known time periods, with innovational and additive outliers, respectively. Silva and Pereira [15] (2012) detected outlying observations in INAR (1) process contaminated with additive outliers. In their research, the models contaminated with only one type of outliers were assumed. This paper considers the problem of modeling a Poisson integer valued autoregressive time series contaminated with additive and innovational outliers. We show how Gibbs sampling can be used to detect outlying observations in INAR (1) processes. Our methodology can detect outliers and specify the type of outliers simultaneously, without an additional procedure distinguishing between AO and IO outliers.

The paper is organized as follows. Section 2 sets up the model and priors. Section 3 discusses the algorithms of outlier detection. Section 4 reports the simulation results as well as the real data. Finally, some comments are given in Section 5.

2. The setup

2.1. The model

Let $\{z_t\}$ be a Poisson INAR (1) process

$$z_t = \alpha \circ z_{t-1} + \varepsilon_t = \sum_{j=1}^{z_{t-1}} \xi_{t,j} + \varepsilon_t \quad (1)$$

i.e.

$$\phi(B)z_t = \varepsilon_t \quad (2)$$

where $\phi(B) = I - \alpha \circ B$ such that $\phi(B)z_t = z_t - \alpha \circ z_{t-1} = z_t - \sum_{j=1}^{z_{t-1}} \xi_{t,j}$, $(\xi_{t,j})$ is a sequence of Bernoulli r.v. with mean $\alpha \in [0,1]$, and $\{\varepsilon_t\}$ is the arrival process, a sequence of i.i.d. Poisson variables $\varepsilon_t \sim P \circ (\lambda)$.

According to the definition of AO and IO (Abraham and Box [16] 1979), observed data $\{x_t\}$ with AO and IO can be expressed as

$$x_t = z_t + w_t^{\text{AO}} \delta_t^{\text{AO}} + \phi^{-1}(B)w_t^{\text{IO}} \delta_t^{\text{IO}} \quad (3)$$

where $\{z_t\}$ is an INAR(1) model, $\phi^{-1}(B) = I + \alpha \circ B + \alpha^2 \circ B^2 + \dots$, with $(\alpha^j \circ B^j)z_t = \alpha^j \circ z_{t-j} = \sum_{j=1}^{z_{t-j}} \xi_{t,j}$, and $(\xi_{t,j})$ is a sequence of Bernoulli r.v. with mean α^j .

Let

$$y_t = z_t + \phi^{-1}(B)w_t^{\text{IO}} \delta_t^{\text{IO}} \quad (4)$$

which means y_t is only influenced by IO. From (2), (3) and (4), we get

$$\begin{cases} \phi(B)(y_t - \phi^{-1}(B)w_t^{\text{IO}} \delta_t^{\text{IO}}) = \varepsilon_t \\ x_t = y_t + w_t^{\text{AO}} \delta_t^{\text{AO}} \end{cases} \quad (5)$$

i.e.

$$\begin{cases} y_t = \alpha \circ y_{t-1} + \varepsilon_t + w_t^{\text{IO}} \delta_t^{\text{IO}} \\ x_t = y_t + w_t^{\text{AO}} \delta_t^{\text{AO}} \end{cases} \quad (6)$$

where δ_t^{AO} is an additive indicator function taking the value 1 if x_t is affected by an AO with size w_t^{AO} and 0 otherwise; δ_t^{IO} is an innovational indicator function taking the value 1 if x_t is affected by an IO with size w_t^{IO} and 0 otherwise.

2.2. Priors

We will consider Bayesian inference on detecting outliers. The following assumptions and prior distributions are given.

- Each observation has a priori probability of γ to be either AO or IO, which means $p(\delta_t^{\text{AO}} = 1) = p(\delta_t^{\text{IO}} = 1) = \gamma$. The prior distribution of γ is $\gamma \sim \text{Be}(h, g)$. Let the hyperparameters $h = 5$, $g = 95$, so that $E(\gamma) = 0.05$. This means we believe AO or IO outliers occur at any time with probability 0.05.
- $\{w_t^{\text{AO}}\}$ is a sequence of i.i.d. variables, the prior of which is $w_t^{\text{AO}} \sim P \circ (\beta)$. For simplicity, $\{w_t^{\text{IO}}\}$ is also a sequence of i.i.d. variables with $w_t^{\text{IO}} \sim P \circ (\beta)$. Also, w_t^{AO} and δ_t^{AO} , δ_t^{IO} and w_t^{IO} are independent for all t . There are two approaches to determine the hyperparameter parameter of β : one is a non-informative setup with $\beta_{\text{inf}} = 30$ that means large variability; the other is an informative setup that set β_{inf} equal to three times the standard deviation of the 1-step-ahead prediction error.
- The prior distribution of α and λ are $\alpha \sim \text{Be}(a, b)$, $\lambda \sim \text{Ga}(c, d)$ (Silva, 2005). Let the hyperparameters $a = b = c = d = 0.01$, i.e. non-information prior distributions (Beta and Gamma distributions with large variability).

Since $p(\delta_j^{\text{AO}} = 1 | \mathbf{X}) + p(\delta_j^{\text{AO}} = 0 | \mathbf{X}) = 1$, a cut-off point of 0.5 is used for detecting outliers, i.e. when $p(\delta_j^{\text{AO}} = 1 | \mathbf{X}) > 0.5$, there is a possible additive outlier. Similarly, when $p(\delta_j^{\text{IO}} = 1 | \mathbf{X}) > 0.5$, there is a possible innovation outlier.

3. Bayesian computation

Theoretically, the joint posterior distribution can be obtained by high dimensional integration. This is not easy in practice. Thus we rely on Markov Chain Monte Carlo (MCMC) simulation from the posterior distribution. In particular we resort to an iterative Gibbs sampler.

In this section we describe the Bayesian approach via Gibbs sampling to estimate model (3). Let $x_1 = y_1 = z_1$, that means there is no AO and IO outliers in the first observation, and $\Theta = (\alpha, \lambda)$, $\mathbf{w}^{\text{AO}} = (w_1^{\text{AO}}, w_2^{\text{AO}}, \dots, w_n^{\text{AO}})^T$, $\mathbf{w}^{\text{IO}} = (w_1^{\text{IO}}, w_2^{\text{IO}}, \dots, w_n^{\text{IO}})^T$, $\delta^{\text{AO}} = (\delta_1^{\text{AO}}, \delta_2^{\text{AO}}, \dots, \delta_n^{\text{AO}})^T$, $\delta^{\text{IO}} = (\delta_1^{\text{IO}}, \delta_2^{\text{IO}}, \dots, \delta_n^{\text{IO}})^T$, $\mathbf{w} = ((\mathbf{w}^{\text{AO}})^T, (\mathbf{w}^{\text{IO}})^T)^T$, $\delta = ((\delta^{\text{AO}})^T, (\delta^{\text{IO}})^T)^T$. Now we need to derive the conditional posterior distribution of $\Theta, \delta, \mathbf{w}$ and γ .

Conditioning on y_1 , the likelihood of \mathbf{X} is given by

$$L(\Theta, \delta, \mathbf{w}, \gamma) = \pi(\mathbf{X} | \Theta, \delta, \mathbf{w}, \gamma) \propto e^{-n\lambda} \prod_{t=2}^n \sum_{i=0}^{M_t} \frac{\lambda^{x_t^* - i}}{(x_t^* - i)!} C_i^{y_{t-1}} \alpha^i (1 - \alpha)^{y_{t-1} - i} \quad (7)$$

where

$$x_t^* = x_t - w_t^{\text{AO}} \delta_t^{\text{AO}} - w_t^{\text{IO}} \delta_t^{\text{IO}} \text{ and } M_t = \min(y_{t-1}, x_t^*), t = 2, \dots, n.$$

Let $\pi(\Theta, \delta, \mathbf{w}, \gamma)$ denote the prior distribution of $(\Theta, \delta, \mathbf{w}, \gamma)$, Then

$$\pi(\Theta, \delta, \mathbf{w}, \gamma) \propto e^{-d\lambda} \lambda^{c-1} \alpha^{a-1} (1-\alpha)^{b-1} \gamma^{h-1} (1-\gamma)^{g-1} \left(\prod_{t=2}^n e^{-\beta} \frac{\beta^{w_t^{\text{AO}}}}{w_t^{\text{AO}}!} \right) \left(\prod_{t=2}^n e^{-\beta} \frac{\beta^{w_t^{\text{IO}}}}{w_t^{\text{IO}}!} \right) \quad (8)$$

The posterior distribution of $(\Theta, \delta, \mathbf{w}, \gamma)$ is then given by

$$\begin{aligned} \pi(\Theta, \delta, \mathbf{w}, \gamma | \mathbf{X}) &\propto L(\Theta, \delta, \mathbf{w}, \gamma) \pi(\Theta, \delta, \mathbf{w}, \gamma) \\ &\propto e^{-[d\lambda+2(n-1)\beta]} \lambda^{c-1} \alpha^{a-1} (1-\alpha)^{b-1} \gamma^{h-1} (1-\gamma)^{g-1} \frac{\beta^{\sum_{t=2}^n (w_t^{\text{AO}} + w_t^{\text{IO}})}}{(\prod_{t=2}^n w_t^{\text{AO}}!)(\prod_{t=2}^n w_t^{\text{IO}}!)} L(\Theta, \delta, \mathbf{w}, \gamma) \end{aligned} \quad (9)$$

with $0 < \alpha < 1, \lambda > 0, 0 < \gamma < 1$ and $w_t^{\text{AO}} = 0, 1, \dots, w_t^{\text{IO}} = 0, 1, \dots, t = 2, 3, \dots, n$.

The full conditional posterior distribution for α, λ can be derived using approaches similar to that of Silva and Pereira [17] (2005).

$$\pi(\alpha | \mathbf{X}, \lambda, \mathbf{w}, \delta, \gamma) \propto \alpha^{a-1} (1-\alpha)^{b-1} \prod_{t=2}^n \sum_{i=0}^{M_t} T(t, i) \alpha^i (1-\alpha)^{y_{t-1}-i} \quad (10)$$

$$\text{with } T(t, i) = \frac{\lambda^{x_t^* - i}}{(x_t^* - i)!} C_i^{y_{t-1}}.$$

$$\pi(\lambda | \mathbf{X}, \alpha, \mathbf{w}, \delta, \gamma) \propto \lambda^{c-1} e^{-(d+n)\lambda} \prod_{t=2}^n \sum_{i=0}^{M_t} U(t, i) \lambda^{x_t^* - i} \quad (11)$$

$$\text{with } U(t, i) = \frac{1}{(x_t^* - i)!} C_i^{y_{t-1}} \alpha^i (1-\alpha)^{y_{t-1}-i}.$$

The full conditional distribution δ^{AO} of is given by

$$\delta_j^{\text{AO}} | \mathbf{X}, \Theta, \mathbf{w}, \delta_{(\cdot j)}^{\text{AO}}, \delta^{\text{IO}}, \gamma \sim \text{Ber}(1, p_j^{\text{AO}}) \quad (12)$$

where $\delta_{(\cdot j)}^{\text{AO}} = (\delta_1^{\text{AO}}, \dots, \delta_{j-1}^{\text{AO}}, \delta_{j+1}^{\text{AO}}, \dots, \delta_n^{\text{AO}})^T$, for each $j = 2, \dots, n$.

We know

$$p(\delta_j^{\text{AO}} = 1 | \mathbf{X}, \Theta, \mathbf{w}, \delta_{(\cdot j)}^{\text{AO}}, \delta^{\text{IO}}, \gamma) + p(\delta_j^{\text{AO}} = 0 | \mathbf{X}, \Theta, \mathbf{w}, \delta_{(\cdot j)}^{\text{AO}}, \delta^{\text{IO}}, \gamma) = 1 \quad (13)$$

$$\begin{aligned} p(\delta_j^{\text{AO}} | \mathbf{X}, \Theta, \mathbf{w}, \delta_{(\cdot j)}^{\text{AO}}, \delta^{\text{IO}}, \gamma) &\propto \pi(\mathbf{X} | \Theta, \mathbf{w}, \delta, \gamma) p(\delta_j^{\text{AO}}) \\ &\propto e^{-n\lambda} \left(\prod_{t=2}^n \sum_{i=0}^{M_t} \frac{\lambda^{x_t^* - i}}{(x_t^* - i)!} C_i^{y_{t-1}} \alpha^i (1-\alpha)^{y_{t-1}-i} \right) \gamma^{\delta_j^{\text{AO}}} (1-\gamma)^{1-\delta_j^{\text{AO}}} \end{aligned} \quad (14)$$

$$\begin{aligned} p(\delta_j^{\text{AO}} = 1 | \mathbf{X}, \Theta, \mathbf{w}, \delta_{(\cdot j)}^{\text{AO}}, \delta^{\text{IO}}, \gamma) \\ \propto \gamma \left(\sum_{i=0}^{M_j} \frac{\lambda^{x_j - w_j^{\text{AO}} - w_j^{\text{IO}} \delta_j^{\text{IO}} - i}}{(x_j - w_j^{\text{AO}} - w_j^{\text{IO}} \delta_j^{\text{IO}} - i)!} C_i^{y_{j-1}} \alpha^i (1-\alpha)^{y_{j-1}-i} \right) \left(\sum_{i=0}^{M_{j+1}^*} \frac{\lambda^{x_{j+1}^* - i}}{(x_{j+1}^* - i)!} C_i^{x_j - w_j^{\text{AO}}} \alpha^i (1-\alpha)^{x_j - w_j^{\text{AO}} - i} \right) \end{aligned} \quad (15)$$

with $M_{j+1}^* = \min(x_j - w_j^{\text{AO}}, x_{j+1}^*)$.

and

$$\begin{aligned} p(\delta_j^{\text{AO}} = 0 | \mathbf{X}, \Theta, \mathbf{w}, \delta_{(\cdot j)}^{\text{AO}}, \delta^{\text{IO}}, \gamma) \\ \propto (1-\gamma) \left(\sum_{i=0}^{M_j} \frac{\lambda^{x_j - w_j^{\text{IO}} \delta_j^{\text{IO}} - i}}{(x_j - w_j^{\text{IO}} \delta_j^{\text{IO}} - i)!} C_i^{y_{j-1}} \alpha^i (1-\alpha)^{y_{j-1}-i} \right) \left(\sum_{i=0}^{M_{j+1}^{**}} \frac{\lambda^{x_{j+1}^* - i}}{(x_{j+1}^* - i)!} C_i^{x_j} \alpha^i (1-\alpha)^{x_j - i} \right) \end{aligned} \quad (16)$$

with $M_{j+1}^{**} = \min(x_j, x_{j+1}^*)$.

Therefore

$$p(\delta_j^{\text{AO}} = 1 | \mathbf{X}, \mathbf{\Theta}, \mathbf{w}, \delta_{(j)}^{\text{AO}}, \delta_j^{\text{IO}}, \gamma) = \frac{q_{1j}^{\text{AO}}}{q_{1j}^{\text{AO}} + q_{2j}^{\text{AO}}} \triangleq p_j^{\text{AO}} \quad (17)$$

with

$$q_{1j}^{\text{AO}} = \gamma \left(\sum_{i=0}^{M_j} \frac{\lambda^{x_j - w_j^{\text{AO}} - w_j^{\text{IO}} \delta_j^{\text{IO}} - i}}{(x_j - w_j^{\text{AO}} - w_j^{\text{IO}} \delta_j^{\text{IO}} - i)!} C_i^{y_{j-1}} \alpha^i (1-\alpha)^{y_{j-1}-i} \right) \left(\sum_{i=0}^{M_{j+1}^*} \frac{\lambda^{x_{j+1}^* - i}}{(x_{j+1}^* - i)!} C_i^{x_j - w_j^{\text{AO}}} \alpha^i (1-\alpha)^{x_j - w_j^{\text{AO}} - i} \right) \quad (18)$$

$$q_{2j}^{\text{AO}} = (1-\gamma) \left(\sum_{i=0}^{M_j} \frac{\lambda^{x_j - w_j^{\text{IO}} \delta_j^{\text{IO}} - i}}{(x_j - w_j^{\text{IO}} \delta_j^{\text{IO}} - i)!} C_i^{y_{j-1}} \alpha^i (1-\alpha)^{y_{j-1}-i} \right) \left(\sum_{i=0}^{M_{j+1}^{**}} \frac{\lambda^{x_{j+1}^* - i}}{(x_{j+1}^* - i)!} C_i^{x_j} \alpha^i (1-\alpha)^{x_j - i} \right) \quad (19)$$

Similarly, the full conditional distribution of δ_j^{IO} is given by

$$\delta_j^{\text{IO}} | \mathbf{X}, \mathbf{\Theta}, \mathbf{w}, \delta_{(j)}^{\text{AO}}, \delta_{(j)}^{\text{IO}}, \gamma \sim \text{Ber}(1, p_j^{\text{IO}}) \quad (20)$$

with $\delta_{(j)}^{\text{IO}} = (\delta_1^{\text{IO}}, \dots, \delta_{j-1}^{\text{IO}}, \delta_{j+1}^{\text{IO}}, \dots, \delta_n^{\text{IO}})^T$, for each $j = 2, \dots, n$.

$$p(\delta_j^{\text{IO}} = 1 | \mathbf{X}, \mathbf{\Theta}, \mathbf{w}, \delta_{(j)}^{\text{AO}}, \delta_{(j)}^{\text{IO}}, \gamma) + p(\delta_j^{\text{IO}} = 0 | \mathbf{X}, \mathbf{\Theta}, \mathbf{w}, \delta_{(j)}^{\text{AO}}, \delta_{(j)}^{\text{IO}}, \gamma) = 1 \quad (21)$$

$$p(\delta_j^{\text{IO}} | \mathbf{X}, \mathbf{\Theta}, \mathbf{w}, \delta_{(j)}^{\text{AO}}, \delta_{(j)}^{\text{IO}}, \gamma) \propto \pi(\mathbf{X} | \mathbf{\Theta}, \mathbf{w}, \delta, \gamma) p(\delta_j^{\text{IO}}) \propto e^{-n\lambda} \left(\prod_{t=2}^n \sum_{i=0}^{M_t} \frac{\lambda^{y_t - w_t^{\text{IO}} \delta_t^{\text{IO}} - i}}{(y_t - w_t^{\text{IO}} \delta_t^{\text{IO}} - i)!} C_i^{y_{t-1}} \alpha^i (1-\alpha)^{y_{t-1}-i} \right) \gamma^{\delta_j^{\text{IO}}} (1-\gamma)^{1-\delta_j^{\text{IO}}} \quad (22)$$

with $M_t = \min(y_{t-1}, x_t^*) = \min(y_{t-1}, y_t - w_t^{\text{IO}} \delta_t^{\text{IO}})$.

$$p(\delta_j^{\text{IO}} = 1 | \mathbf{X}, \mathbf{\Theta}, \mathbf{w}, \delta_{(j)}^{\text{AO}}, \delta_{(j)}^{\text{IO}}, \gamma) \propto \gamma \sum_{i=0}^{M_j^*} \frac{\lambda^{y_j - w_j^{\text{IO}} - i}}{(y_j - w_j^{\text{IO}} - i)!} C_i^{y_{j-1}} \alpha^i (1-\alpha)^{y_{j-1}-i} \quad (23)$$

with $M_j^* = \min(y_{j-1}, y_j - w_j^{\text{IO}})$.

And

$$p(\delta_j^{\text{IO}} = 0 | \mathbf{X}, \mathbf{\Theta}, \mathbf{w}, \delta_{(j)}^{\text{AO}}, \delta_{(j)}^{\text{IO}}, \gamma) \propto (1-\gamma) \sum_{i=0}^{M_j^{**}} \frac{\lambda^{y_j - i}}{(y_j - i)!} C_i^{y_{j-1}} \alpha^i (1-\alpha)^{y_{j-1}-i} \quad (24)$$

with $M_j^{**} = \min(y_{j-1}, y_j - w_j^{\text{IO}})$.

Therefore

$$p(\delta_j^{\text{IO}} = 1 | \mathbf{X}, \mathbf{\Theta}, \mathbf{w}, \delta_{(j)}^{\text{AO}}, \delta_{(j)}^{\text{IO}}, \gamma) = \frac{q_{1j}^{\text{IO}}}{q_{1j}^{\text{IO}} + q_{2j}^{\text{IO}}} \triangleq p_j^{\text{IO}} \quad (25)$$

with

$$q_{1j}^{\text{IO}} = \gamma \sum_{i=0}^{M_j^*} \frac{\lambda^{y_j - w_j^{\text{IO}} - i}}{(y_j - w_j^{\text{IO}} - i)!} C_i^{y_{j-1}} \alpha^i (1-\alpha)^{y_{j-1}-i} \quad (26)$$

$$q_{2j}^{\text{IO}} = (1-\gamma) \sum_{i=0}^{M_j^{**}} \frac{\lambda^{y_j - i}}{(y_j - i)!} C_i^{y_{j-1}} \alpha^i (1-\alpha)^{y_{j-1}-i} \quad (27)$$

We note that, when $\delta_j^{\text{AO}} = 0$, there will be no additive outlier at $t = j$, which further implies w_j^{AO} contain no new information expect the prior. Then conditional posterior distribution of w_j^{AO} is:

$$w_j^{\text{AO}} \mid \mathbf{X}, \boldsymbol{\Theta}, \delta_j^{\text{AO}} = 0, \boldsymbol{\delta}_{(j)}^{\text{AO}}, \mathbf{w}_{(j)}^{\text{AO}}, \gamma \sim P \circ (\beta) \quad (28)$$

However, if $\delta_j^{\text{AO}} = 1$

$$\begin{aligned} & p(w_j^{\text{AO}} \mid \mathbf{X}, \boldsymbol{\Theta}, \delta_j^{\text{AO}} = 1, \boldsymbol{\delta}_{(j)}^{\text{AO}}, \mathbf{w}_{(j)}^{\text{AO}}, \gamma) \\ & \propto p(w_j^{\text{AO}} \mid \boldsymbol{\Theta}, \delta_j^{\text{AO}} = 1, \boldsymbol{\delta}_{(j)}^{\text{AO}}, \mathbf{w}_{(j)}^{\text{AO}}, \gamma) p(\mathbf{X} \mid \boldsymbol{\Theta}, \delta_j^{\text{AO}} = 1, \boldsymbol{\delta}_{(j)}^{\text{AO}}, \mathbf{w}_{(j)}^{\text{AO}}, \gamma) \\ & \propto p(w_j^{\text{AO}} \mid \delta_j^{\text{AO}} = 1) p(\delta_j^{\text{AO}} = 1 \mid \mathbf{X}, \boldsymbol{\Theta}, \boldsymbol{\delta}_{(j)}^{\text{AO}}, \mathbf{w}_{(j)}^{\text{AO}}, \gamma) \\ & \propto \frac{e^{-\beta} \beta^{w_j^{\text{AO}}}}{w_j^{\text{AO}}!} q_{1j}^{\text{AO}} \quad (w_j^{\text{AO}} = 0, 1, 2, \dots) \end{aligned} \quad (29)$$

Similarly, the full conditional distribution of w_j^{IO} when $\delta_j^{\text{IO}} = 0$ is

$$w_j^{\text{IO}} \mid \mathbf{X}, \boldsymbol{\Theta}, \delta_j^{\text{IO}} = 0, \boldsymbol{\delta}_{(j)}^{\text{IO}}, \mathbf{w}_{(j)}^{\text{IO}}, \gamma \sim P \circ (\beta) \quad (30)$$

However, if $\delta_j^{\text{IO}} = 1$

$$p(w_j^{\text{IO}} \mid \mathbf{X}, \boldsymbol{\Theta}, \delta_j^{\text{IO}} = 1, \boldsymbol{\delta}_{(j)}^{\text{IO}}, \mathbf{w}_{(j)}^{\text{IO}}, \gamma) \propto \frac{e^{-\beta} \beta^{w_j^{\text{IO}}}}{w_j^{\text{IO}}!} q_{1j}^{\text{IO}} \quad (w_j^{\text{IO}} = 0, 1, 2, \dots) \quad (31)$$

Finally, the conditional posterior distribution of γ depends only on $\boldsymbol{\delta}$, and the prior distribution of γ is $\gamma \sim Be(h, g)$, so the conditional posterior is given by

$$\gamma \mid \mathbf{X}, \boldsymbol{\Theta}, \mathbf{w}, \boldsymbol{\delta} \equiv \gamma \mid \boldsymbol{\delta} \sim Be(h + k, g + 2n - 2 - k) \quad (32)$$

Where k is the estimated number of outliers (including additive outliers and innovation outliers).

If let $\boldsymbol{\delta}^{\text{IO}} = (\delta_1^{\text{IO}}, \delta_2^{\text{IO}}, \dots, \delta_n^{\text{IO}})^T = \mathbf{0}$, i.e. there is no innovation outliers, then the full posterior distributions of parameters are the same as that of Silva and Pereira [17] (2012).

4. Illustration

4.1. Simulated data sets

In this section we report the results of simulation studies detecting AO and IO outliers using different parameters and outlier sizes.

We generated $n = 100$ observations from different INAR (1) models. We chose $\alpha = 0.15, 0.5, 0.85$ and $\lambda = 3, 5$ with three outliers, one innovation outlier and two additive outliers of sizes equal to three, five and/or seven times the standard deviation. The locations of the outliers are generated randomly.

The Gibbs sampler is used to obtain the Bayesian estimates. The sampling procedure iterates 5000 times the value of the last 2500 iterations are kept, and the estimate is given by the mean of the last 500 iterations.

For each scenario we run our Gibbs sampling algorithm with hyperparameters determined as described in Section 2.2.

The results for all simulated models are summarized in table 1 for time series contaminated with three outliers including one innovational outlier and two additive outliers. Table 1 contains the values of the parameters α and λ that are used to generate the series with additive outliers of size δ_t^{AO} and innovational outliers of size δ_t^{IO} at times t . It also includes the estimator for the parameters α and λ using conditional least squares assuming no outlier, *CLS*. We also have obtained the estimator of parameters α and λ by Gibbs sampling, *Bayes*, the estimated probability of outlier occurrence,

Probability, and the estimated outlier size for all the time points for which that probability is over the threshold 0.5.

The results presented in table 1 indicate that the procedure is able to detect innovational and additive outliers in INAR (1) models. In comparison, the obtained Bayes estimator of parameters α and λ (*Bayes*) are very close to the given initial values. It illustrates the negative impact of the outliers on the estimates of α and λ (*Initial CLS*).

Table 1. Gibbs sampling results for INAR (1) time series simulations with two different parameters α and λ , three outliers including one innovation outlier and two additive outliers.

Parameter	True	Estimates		<i>Probability</i>
		<i>CLS</i>	<i>Bayes</i>	
α	0.15	0.05	0.14	-
λ	3	3.44	2.71	-
δ_{26}^{AO}	9	-	10	0.67
δ_{49}^{IO}	6	-	10	0.69
δ_{79}^{AO}	13	-	11	0.63
α	0.15	0.09	0.16	-
λ	5	5.88	5.01	-
δ_{17}^{AO}	7	-	11	0.58
δ_{26}^{AO}	16	-	12	0.73
δ_{91}^{IO}	12	-	11	0.60
α	0.5	0.27	0.48	-
λ	3	4.71	2.98	-
δ_{26}^{IO}	12	-	13	0.91
δ_{49}^{AO}	17	-	14	0.85
δ_{76}^{AO}	8	-	11	0.75
α	0.5	0.23	0.51	-
λ	5	8.17	5.12	-
δ_{11}^{AO}	10	-	7	0.99
δ_{28}^{IO}	15	-	16	0.99
δ_{91}^{AO}	21	-	17	0.99
α	0.85	0.72	0.85	-
λ	3	5.78	2.73	-
δ_{37}^{IO}	31	-	28	0.63
δ_{49}^{AO}	22	-	24	0.99
δ_{76}^{AO}	13	-	23	0.85
α	0.85	0.40	0.84	-
λ	5	20.3	4.89	-
δ_6^{AO}	28	-	28	0.99
δ_{15}^{AO}	40	-	29	0.99
δ_{51}^{IO}	17	-	23	0.93

We measured the computational costs of the algorithm with different iterations, number of outliers, number of observations, and then, generate a table with runtime for each situation in table 2. The test was carried on an ACER laptop with an Intel Core i3-6100U CPU running at 2.3GHz, a 10GB RAM,

and running on Windows 7 SP1 64-Bit operating system. The algorithm was written in R and not optimized.

From the above table, we could see that the runtime is not quite affected by the number of outliers, and is linear to both the number of iterations and observations. The time complexity of the algorithm is very likely to be of $O(n)$.

Table 2. Runtime at different iterations, number of outliers, and number of observations.

Iterations	Number of outliers	Observations	Runtime (min)
5000	1	100	2.6651
5000	3	100	2.7653
5000	5	100	3.8575
5000	1	500	18.1593
5000	3	500	18.4848
5000	5	500	18.2354
10000	1	100	7.9704
10000	3	100	8.2880
10000	5	100	6.8059
10000	1	500	40.5298
10000	3	500	39.5765
10000	5	500	40.3201

4.2. Application to IP data

We apply the proposed method to a data set studied by Silva and Pereira (2005, 2012) concerning the number of different IP addresses accessing the server of the Department of Statistics of the University of Würzburg on November 29th, 2005, from 10:00 to 18:00. The result is represented in figure 1.

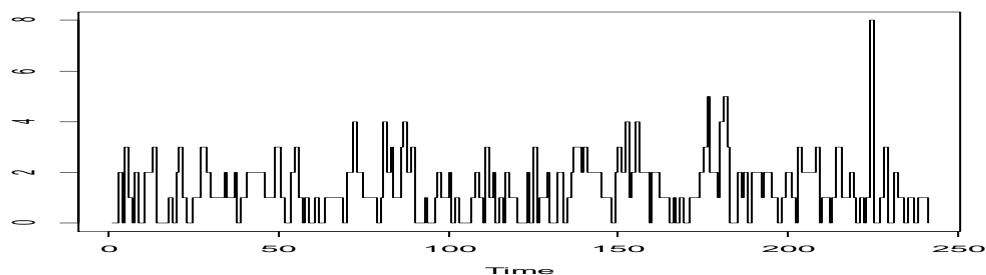


Figure 1. The number of different IP addresses accessing the server of the Department of Statistics of the University of Würzburg from 10:00 to 18:00, Nov. 29, 2005.

We will use a model of order one for this data set here whose appropriateness is justified in Silva (2015). The CLS estimates of α and λ are $\hat{\alpha}_{CLS} = 0.22$ and $\hat{\lambda}_{CLS} = 1.03$ respectively. The posterior probability of each observation applying the proposed methodology is represented in figure 2. Figure 2(a) indicates the possible occurrence of an additive outlier at time $t=224$. Figure 2(b) indicates

no innovational outlier occurrence. The estimated size of the additive outlier is $\delta_{224}^{AO} = 7$. Simultaneously, the estimates for α and λ are $\hat{\alpha}_{Bayes} = 0.24$ and $\hat{\lambda}_{Bayes} = 0.98$, respectively. The result is consistent with the conclusion in Silva and Pereira[17] (2012).

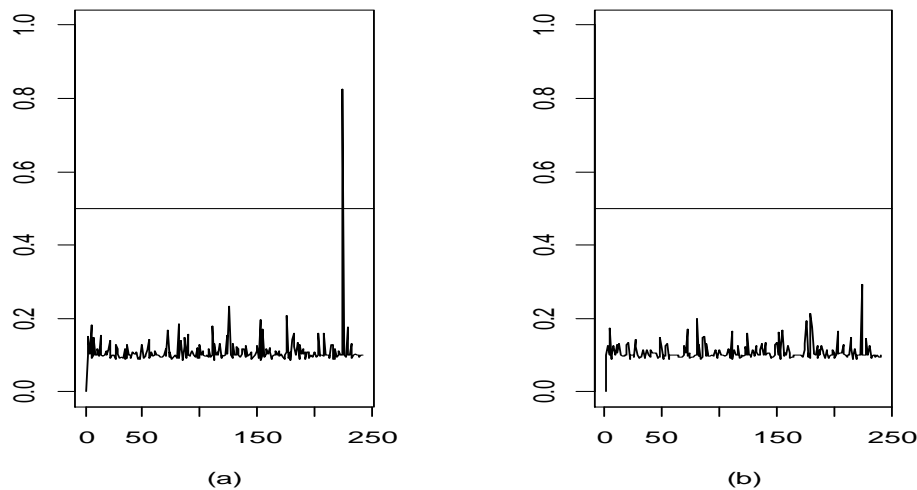


Figure 2. (a) Posterior probability of additive outlier occurrence for the IP data; (b) Posterior probability of innovation outlier occurrence for the IP data.

5. Summary

We propose an algorithm to detect outliers and identify the type as either additive or innovation in Poisson INAR (1) time series. This method is designed for computing the posterior probabilities, estimating parameters and the size of outliers based on Gibbs sampling. The outliers can be detected by comparing the posterior probabilities with a cut-off point. The method can be used without knowing the type of outlier, the number of outliers. Extensive simulation studies and an experiment on real-world datasets, show that the proposed algorithm performs well.

A promising research direction is to extend our algorithm for higher order models, INAR (p) ($p > 1$). Since the full conditional posterior distributions of INAR (p) are similar but more complex, the implementation of the methodology for higher order models requires additional computing effort. Masking and swamping effects caused by patches of outliers may occur within the patch. The solution of these problems is being investigated.

Acknowledgements

The work of Shang H. and Zhang B. B. is supported by the Fundamental Research Funds for the Central Universities (3142017005, 3142015042, 3142017109), Social science research project of Beijing Municipal Commission of Education (SM201310038007), and the Hebei province science and technology project (162176438), Key Project of Probability Theory and Mathematical Statistics Discipline, North China Institute of Science and Technology (06DV09).

References

- [1] Fox A J 1972 Outliers in time series *J. Roy. Sta. B.* **34** 350-363
- [2] Tsay R S 1986 Time series model specification in the presence of outliers *J. Am. Stat. A.* **81** 132-141
- [3] Muirhead C R 1986 Distinguishing outlier types in time series *J. Roy. Sta. B* **48** 39-47
- [4] Chang I, Tiao G C and Chen C 1988 Estimation of time series parameters in the presence of outliers. *Technomet.* **30** 193–204

- [5] Chen C and Liu L M 1993 Joint estimation of model parameters and outlier effects in time series *J. Am. Stat. A.* **88(421)** 284-297
- [6] Justel A, Peña D and Tsay R 2001 Detection of outlier patches in autoregressive time series *Stat. Sinica.* **11** 651-673
- [7] Galeano P, Peña D and Tsay R S 2006 Outlier detection in multivariate time series by projection pursuit *J. Am. Stat. A.* **101(474)** 654-669
- [8] Fokianos K and Fried R 2010 Interventions in ingarch processes *J. Time Ser. Anal.* **31** 210-225
- [9] Fried R, Liboschik T, Elsaied H, Kitromilidou S, and Fokianos K 2014 On outliers and interventions in count time series following glms *Aus. J. St.* **43** 181-193
- [10] Kitromilid S and Fokianos K 2016 Robust estimation methods for a class of log-linear count time series models *J. Stat. Comp.* **86(4)** 740-755
- [11] Al-Osh M A and Alzaid A A 1987 First-order integer-valued autoregressive (inar(1)) process *J. Time Ser. Anal.* **8** 261-275
- [12] McKenzie E 1985 Some simple models for discrete variate time series *Jaw. J. Am. Wat. Re.* **21** 645-650
- [13] Barczy M, Ispány M, Pap G, Scotto M and Silva M E 2010 Innovational outliers in inar(1) models *Comm. St-the.* **39** 3343-3362
- [14] Barczy M 2012 Additive outliers in inar(1) models *Stat. Pap.* **53(4)** 935-949
- [15] Silva M E and Pereira I 2012 Detection of additive outliers in poisson integer-valued autoregressive time series *Sta.*
- [16] Abraham B and Box G E P 1979 Bayesian analysis of some outlier problems in time series *Biometrika* **66(2)** 229-236
- [17] Silva I, Silva M, Pereira I and Silva N 2005 Replicated inar(1) processes *Methodol Comp. Appl Probab.* **7** 517-542