Detection of Additive Outliers in Poisson INAR(1) Time Series

Maria Eduarda Silva and Isabel Pereira

Abstract Outlying observations are commonly encountered in the analysis of time series. In this paper a Bayesian approach is employed to detect additive outliers in order one Poisson integer-valued autoregressive time series. The methodology is informative and allows the identification of the observations which require further inspection. The procedure is illustrated with simulated and observed data sets.

1 Introduction

It is well known that unusual observations and intervention effects often occur in data sets and can have adverse effects on model identification and parameter estimation. Time series of counts are no exception. In the last decades time series of counts have become available in a wide variety of fields including: actuarial science, computer science, economics, epidemiology, finance, hydrology, meteorology and environmental studies. These data are naturally non-normal and present non linear characteristics. The need to analyse such data adequately led to a multiplicity of approaches and a diversification of models that explicitly account for the discreteness of the data, see [10] for a recent review. In this paper we focus on the class of Poisson integer valued autoregressive models of order 1, INAR(1). This model, first proposed by [1], has been extensively studied in the literature and applied to many real-world problems because of its simplicity and easiness of interpretation. In fact, any data series that may be thought of as the number of members (e.g. people, firms or orders) of a queue, the number of units in a stock or inventory, or the outcome of a birth-and-death process, or a branching process with immigration may be modelled by the INAR class. The point is that the INAR class has found applications across many disciplines. Hence, it is timely to study the problem of outlier detection given its relevance for inference and diagnostics.

M.E. Silva (✉)
CIDMA & Faculdade de Economia, Universidade do Porto, Porto, Portugal
e-mail: mesilva@fep.up.pt

I. Pereira
CIDMA & Departamento de Matemática, Universidade de Aveiro, Aveiro, Portugal
e-mail: isabel.pereira@ua.pt

J.-P. Bourguignon et al. (eds.), *Mathematics of Energy and Climate Change*,
CIM Series in Mathematical Sciences 2, DOI 10.1007/978-3-319-16121-1_19

In the framework of Gaussian linear time series the problem of detecting and estimating outliers and other intervention effects has been investigated by several authors including [4, 6, 11, 15]. However, the problem of modelling outliers and other intervention effects in the context of time series of counts has, as yet, received little attention in the literature albeit its relevance for inference and diagnostics. Moreover, in this context additional motivation stems from the fact that the usual techniques for outlier removal are not adequate since often lead to non integer values. In the context of time series of counts, [7] investigate the problem of modelling intervention effects in INGARCH models and [2, 3] consider Conditional Least Squares (CLS) estimation of the parameters of an INAR(1) model contaminated, at known time periods, with innovational and additive outliers, respectively. Here the problem of detecting outliers is considered under a Bayesian perspective. Bayesian approaches have been used to detect outliers in ARMA models by [11] and in bilinear models by [5]. This approach has the advantage of not requiring beforehand knowledge on the number and location of outliers in the series and of treating equally all the observations (with and without outliers). In fact, all the observations have the same prior probability of being an outlier. Then, at each time point we estimate the posterior probability of occurrence of an outlier via Gibbs sampling. The Gibbs sampler [8] is a Markovian updating scheme enabling the obtention of samples from a joint distribution via iterated sampling from full conditional distributions. The method may be briefly described considering the case of three parameters $(\theta_1, \theta_2, \theta_3)$ with a complex (posterior) joint and marginal distributions. Let $\mathbf{y}$ be the observed time series and $f_i(\theta_i|\theta_k, \theta_l, \mathbf{y})$ be the conditional distribution of θ_i given the remainder parameters θ_k, θ_l and data, $\mathbf{y}$. The Gibbs sampler employed in this paper then works as follows: given initial values $(\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)})$, draw $\theta_1^{(1)}$ from $f_1(\theta_1|\theta_2^{(0)}, \theta_3^{(0)}, \mathbf{y})$, then draw $\theta_2^{(1)}$ from $f_2(\theta_2|\theta_3^{(0)}, \theta_1^{(1)}, \mathbf{y})$ and finally, complete the first iteration by drawing $\theta_3^{(1)}$ from $f_3(\theta_3|\theta_1^{(1)}, \theta_2^{(1)}, \mathbf{y})$. After a large number of iterations, say $M+N$ we obtain a sample $(\theta_1^{(j)}, \theta_2^{(j)}, \theta_3^{(j)})$, $j = M+1, M+2, \ldots, M+N$ whose empirical distribution can approximate the desired posterior marginals. This methodology provides estimates for the probability of outlier occurrence at each time point leading to an effective outlier detection.

To motivate our approach, we represent in Fig. 1a a data set studied by [16] concerning the number of different IP addresses (approximately equivalent to the number of different users) accessing the server of the pages of the Department of Statistics of the University of Würzburg in 2-min periods from 10 am to 6 pm on the 29th November 2005, in a total of 241 observations. Henceforth, this data set will be denoted as IP data and is analysed in detail in Sect. 4.1. Figure 1b represents the posterior probability of outlier occurrence at time t and clearly indicates $t = 224$ as an outlying observation. This result agrees with that of [16] who uses statistical process control techniques.

The paper is organized as follows. Section 2 introduces the first order Poisson integer-valued autoregressive model contaminated with outliers. Section 3 explains the procedure for outlier detection and discusses several computational issues.

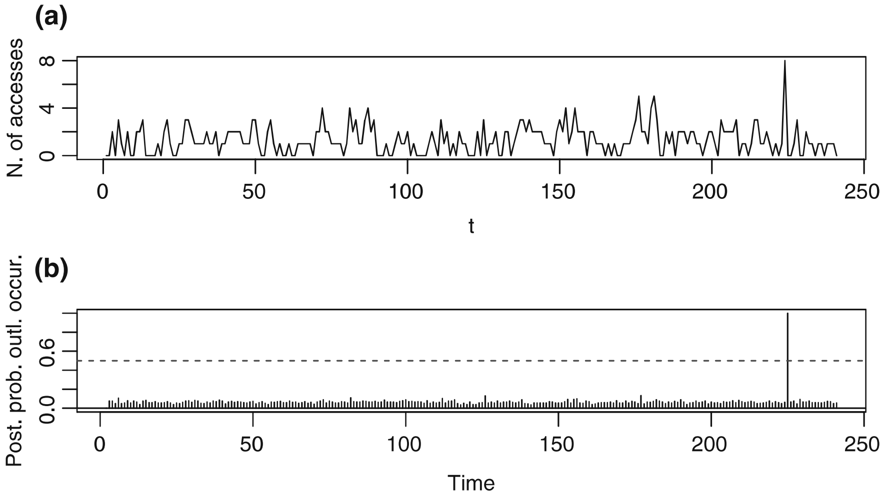

Fig. 1 Number of different IP addresses accessing the server of the pages of the Department of Statistics of the University of Würzburg between 10 am and 6 pm on 29 November 2005 (**a**); posterior probability of outlier occurrence (**b**)

Section 4 illustrates the methodology on several sets of simulated data as well as on the IP data set. Section 5 concludes the paper.

2 INAR(1) Models with Additive Outliers

The Poisson INAR(1), PoINAR(1), model, first introduced by [1] and [12] is defined by the recursive equation

$$X_t = \alpha \circ X_{t-1} + e_t, \quad t \in \mathbb{N}_0, \tag{1}$$

where $\circ$ denotes the binomial thinning operator and $\{e_t\}$, the arrival process, is a sequence of independent and identically distributed Poisson variables, $e_t \sim \text{Po}(\lambda)$, independent of the thinning operations. The binomial thinning is defined as $\alpha \circ X_{t-1} \overset{\mathcal{D}}{=} \sum_{j=1}^{X_{t-1}} \xi_{t,j}$, with $\xi_{t,j}, j = 1, \ldots, X_{t-1}$ a sequence of independent Bernoulli random variables (r.v.'s) with probability of success $\text{P}(\xi_{tj} = 1) = \alpha$. Thus $\alpha \circ X_{t-1}|X_{t-1} \sim Bi(X_{t-1}, \alpha)$, ensuring the discreteness of the process. In fact, the thinning operator $\circ$ acts as the analogue of the usual multiplication used in the continuous-valued autoregressive, AR(1), processes. This concept of thinning is well known in classical probability theory and has been used in the Bienaymé-Galton-Watson branching processes literature as well as in the theory of stopped-sum distributions. Under the above conditions if $X_0 \sim \text{Po}(\lambda/(1-\alpha))$, then the process is strictly stationary and $X_t \sim \text{Po}(\lambda/(1-\alpha))$, yielding a Poisson marginal. X_t behaves like a queue, with arrivals at time t represented by e_t and survivors remaining in the queue, from $t-1$ to t, by $\alpha \circ X_{t-1}$. Alternatively the model may be thought of as a birth-and-death, or stock, process, with additions (births) being generated by e_t and losses (deaths) by $X_{t-1} - \alpha \circ X_{t-1}$.

Note that the Poisson INAR(1) process is a Markov process and that the distribution of X_t given X_{t-1}, $\mathrm{p}(X_t|X_{t-1})$ is the convolution of the two components, binomial and Poisson, as follows:

$$\mathrm{p}(x_t|x_{t-1}) = \sum_{i=0}^{M_t} \binom{x_{t-1}}{i} \alpha^i \, (1-\alpha)^{x_{t-1}-i} \, \frac{\mathrm{e}^{-\lambda} \, \lambda^{x_t - i}}{(x_t - i)!} \tag{2}$$

where $M_t = \min(x_{t-1}, x_t)$ and $\binom{\cdot}{\cdot}$ is the standard combinatorial symbol.

Assume now that the observed time series of counts $y_1, \ldots, y_n$ may be contaminated with one or more additive outliers at unknown time points. Roughly speaking an additive outlier can be interpreted as a measurement error or as an impulse due to some unspecified exogenous source at time τ_i, $i = 1, \ldots, k$. When outliers are present, X_t is unobservable. Then the proposed model for Y_t is the following

$$Y_t = X_t + \eta_t \delta_t, \quad 1 \le t \le n \tag{3}$$

where X_t is a PoINAR(1) process satisfying (1), $\delta_1, \ldots, \delta_n$ are Bernoulli variables with $\mathrm{P}(\delta_t = 1) = \epsilon$, independent of X_t and $\eta_1, \ldots, \eta_n$ are integer valued independent random variables, also independent of X_t and of δ_t. This means that if $\delta_t = 1$ the observed value Y_t is contaminated with an additive outlier (AO) of magnitude η_t. Henceforth, model (3) will be called a Poisson INAR(1) contaminated with outliers.

To obtain the likelihood of the data let $\mathbf{y} = (y_1, \ldots, y_n)$, $\boldsymbol{\Theta} = (\alpha, \lambda)$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)$ and assume that there is no outlier in the first observation, that is $y_1 = x_1$. Moreover, under (3) $X_t = Y_t - \eta_t \delta_t$ is a PoINAR(1). Then conditioning on the first observation the likelihood of $\mathbf{y}$ is given by

$$\begin{aligned} L(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon, \mathbf{y}) &= \prod_{t=2}^{n} \epsilon^{\delta_t} (1-\epsilon)^{1-\delta_t} \\ &\times \sum_{i=0}^{M_t} \frac{\lambda^{y_t - \delta_t \eta_t - i}}{(y_t - \delta_t \eta_t - i)!} \binom{y_{t-1} - \delta_{t-1} \eta_{t-1}}{i} \mathrm{e}^{-\lambda} \, \alpha^i \, (1-\alpha)^{y_{t-1} - \delta_{t-1} \eta_{t-1} - i} \end{aligned}$$

where now $M_t = \min(y_{t-1} - \eta_{t-1} \delta_{t-1}, y_t - \eta_t \delta_t)$.

3 Bayesian Outlier Detection in PoINAR(1) Models

In this section we describe the Bayesian approach via Gibbs sampling to estimate model (3).

In addition to the data and the likelihood, the Bayesian model specification also requires a prior distribution on the parameters. The prior distribution for the contamination parameter ϵ is $\epsilon \sim \text{Be}(h, g)$, with expectation $E(\epsilon) = h/(h+g)$.The prior distribution for η_t is $Po(\beta)$. Regarding the PoINAR(1) parameters α and λ we choose for prior distributions the conjugate of Binomial and Poisson, respectively and thus $\alpha \sim \text{Be}(a, b)$, $\lambda \sim \text{Ga}(c, d)$ [14]. The choice of the set of hyperparameters a, b, c, d, β, h, g is discussed in Sect. 3.2.

Under the above assumptions the prior distribution for $(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon)$, denoted $\pi(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon)$ is given by

$$\pi(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon) \propto \mathrm{e}^{-d\lambda}\, \lambda^{c-1}\, \alpha^{a-1}\, (1-\alpha)^{b-1}\, \epsilon^{h-1}\, (1-\epsilon)^{g-1} \prod_{t=2}^{n} \mathrm{e}^{-\beta} \frac{\beta^{\eta_t}}{\eta_t!}. \tag{4}$$

The posterior distribution for $(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon)$ is then given by

$$\begin{aligned}\pi(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon|\mathbf{y}) &\propto \pi(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon)\, L(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon, \mathbf{y}) \\ &\propto \mathrm{e}^{-[d\lambda+n\beta]}\, \lambda^{c-1}\, \alpha^{a-1}\, (1-\alpha)^{b-1}\, \epsilon^{h-1}\, (1-\epsilon)^{g-1} \\ &\quad \frac{\beta^{\sum_{t=2}^{n} \eta_t}}{\prod_{t=2}^{n} \eta_t!}\, L(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon, \mathbf{y})\end{aligned} \tag{5}$$

with $0 < \alpha < 1$, $\lambda > 0$, $0 < \epsilon < 1$, $\delta_t = 0, 1$ and $\eta_t = 0, 1, \ldots$, $t = 2, 3, \ldots, n$.

This approach is attractive since it enables to measure the likelihood that at each time point the observed value Y_t is affected by an outlier as well as describe the distribution of the outlier size. However, the complexity of the posterior distribution (5) (and consequently of the marginals) makes them analytically intractable. Hence MCMC (Marlov Chain Monte Carlo) methods are required. The model parameters are then estimated by sampling from the complete conditional distribution of each parameter, conditional on the previous sampled values of the other parameters.

3.1 Full Posterior Distributions

The full conditional posterior distributions for α and λ are given by [14]

$$\begin{aligned}\pi(\alpha|\lambda, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon, \mathbf{y}) \propto\;& \alpha^{a-1}\, (1-\alpha)^{b-1} \\ \prod_{t=2}^{n} \epsilon^{\delta_t} (1-\epsilon)^{1-\delta_t} & \sum_{i=0}^{M_t} \frac{\lambda^{y_t - \eta_t\delta_t - i}}{(y_t - \eta_t\delta_t - i)!} \binom{y_{t-1} - \eta_{t-1}\delta_{t-1}}{i} \\ & \alpha^{i}\, (1-\alpha)^{y_{t-1} - \eta_{t-1}\delta_{t-1} - i}\end{aligned} \tag{6}$$

and

$$\pi(\lambda|\alpha, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon, \mathbf{y}) \propto \quad \lambda^{c-1} \mathrm{e}^{-(d+(n-1))\lambda}$$
$$\prod_{t=2}^{n} \epsilon^{\delta_t}(1-\epsilon)^{1-\delta_t} \sum_{i=0}^{M_t} \frac{\lambda^{y_t-\eta_t\delta_t-i}}{(y_t-\eta_t\delta_t-i)!} \binom{y_{t-1}-\eta_{t-1}\delta_{t-1}}{i}$$
$$\alpha^i \, (1-\alpha)^{y_{t-1}-\eta_{t-1}\delta_{t-1}-i}. \tag{7}$$

Now, with respect to the full conditional distribution of $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$ we reason as follows. Let $j = 2, \ldots, n$, and for each j define $\boldsymbol{\Upsilon_\delta} = (\alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)})$ and $\boldsymbol{\Upsilon_\eta} = (\alpha, \lambda, \boldsymbol{\delta}_{(-j)}, \epsilon, \boldsymbol{\eta}_{(-j)})$ where $\boldsymbol{\delta}_{(-j)}$ and $\boldsymbol{\eta}_{(-j)}$ denote the vectors $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$, respectively, each with the jth component deleted. To derive the full conditional distribution of $\boldsymbol{\delta}$ first note that $\delta_j|(\boldsymbol{y}, \boldsymbol{\Upsilon_\delta}) \sim Ber(p_j)$. Accordingly, we can write

$$p_j = \mathrm{P}(\delta_j = 1|\boldsymbol{y}, \boldsymbol{\Upsilon_\delta}) = \frac{\mathrm{P}(\delta_j = 1, \boldsymbol{y}|\boldsymbol{\Upsilon_\delta})}{f(\mathbf{y}|\boldsymbol{\Upsilon_\delta})}. \tag{8}$$

But

$$f(\mathbf{y}|\boldsymbol{\Upsilon_\delta}) = f(\mathbf{y}|\delta_j = 1, \boldsymbol{\Upsilon_\delta})\mathrm{P}(\delta_j = 1|\boldsymbol{\Upsilon_\delta}) + f(\mathbf{y}|\delta_j = 0, \boldsymbol{\Upsilon_\delta})\mathrm{P}(\delta_j = 0|\boldsymbol{\Upsilon_\delta})$$

with $\mathrm{P}(\delta_j = 1|\boldsymbol{\Upsilon_\delta}) = \epsilon$.

Therefore

$$p_j = \frac{\epsilon f(\mathbf{y}|\delta_j = 1, \boldsymbol{\Upsilon_\delta})}{\epsilon f(\mathbf{y}|\delta_j = 1, \boldsymbol{\Upsilon_\delta}) + (1-\epsilon) f(\mathbf{y}|\delta_j = 0, \boldsymbol{\Upsilon_\delta})}. \tag{9}$$

To compute $f(\mathbf{y}|\delta_j = 1, \boldsymbol{\Upsilon_\delta})$ first note that Y_t inherits the Markovian property of X_t and consequently the outlier at time j affects the model for $t = j$ and $t = j + 1$. Therefore

$$\begin{aligned} f(\mathbf{y}|\delta_j = 1, \boldsymbol{\Upsilon_\delta}) &= f(y_j, y_{j+1}|y_{j-1}, \delta_j = 1, \boldsymbol{\Upsilon_\delta}) \\ &= f(y_j, y_{j+1}|y_{j-1}, \delta_j = 1, \alpha, \lambda, \eta_j) \\ &= f(y_j|y_{j-1}, \delta_j = 1, \alpha, \lambda, \eta_j) \\ &\quad \times f(y_{j+1}|y_j, \delta_j = 1, \alpha, \lambda, \eta_j). \end{aligned} \tag{10}$$

Moreover, assuming that $Y_{j-1} = X_{j-1}$ and $Y_{j+1} = X_{j+1}$ meaning that there are no patches of outliers we have

$$f(y_j|y_{j-1}, \delta_j = 1, \alpha, \lambda, \eta_j) = \mathrm{e}^{-\lambda} \sum_{i=0}^{M_j^{**}} \binom{y_{j-1}}{i} \alpha^i \, (1-\alpha)^{y_{j-1}-i} \, \frac{\lambda^{y_j-\eta_j-i}}{(y_j-\eta_j-i)!} \tag{11}$$

and

$$f(y_{j+1}|y_j, \delta_j = 1, \alpha, \lambda, \eta_j)$$
$$= e^{-\lambda} \sum_{i=0}^{M_j^*} \binom{y_j - \eta_j}{i} \alpha^i (1-\alpha)^{y_j-\eta_j-i} \frac{\lambda^{y_{j+1}-i}}{(y_{j+1}-i)!} \tag{12}$$

with $M_t^{**} = \min(y_{t-1}, y_t - \eta_t)$ and $M_t^* = \min(y_t - \eta_t,\ y_{t+1})$.

Similarly, if $\delta_j = 0$ then $X_j = Y_j$ and therefore

$$f(\mathbf{y}|\delta_j = 0, \boldsymbol{\Upsilon_\delta}) = f(\mathbf{y}|\delta_j = 0, \alpha, \lambda, \eta_j)$$
$$= \prod_{t=j}^{j+1} \sum_{i=0}^{M_t} \binom{y_{t-1}}{i} \alpha^i (1-\alpha)^{y_{t-1}-i} e^{-\lambda} \frac{\lambda^{y_t-i}}{y_t - i!} \tag{13}$$

Now, to derive the conditional posterior distribution of $\boldsymbol{\eta}$ note that if $\delta_j = 0$, no outlier at $t = j$, there is no information about η_j except the prior. Then $\eta_j|(\mathbf{y}, \delta_j = 0, \boldsymbol{\Upsilon_\eta}) \sim Po(\beta)$. However, if $\delta_j = 1$, $\mathbf{y}$ contains information about η_j. Therefore we have

$$\pi(\eta_j\,|\mathbf{y}, \delta_j = 1, \boldsymbol{\Upsilon_\eta}) = \frac{\pi(\eta_j|\delta_j = 1, \boldsymbol{\Upsilon_\eta}) f(\mathbf{y}|\delta_j = 1, \eta_j, \boldsymbol{\Upsilon_\eta})}{\sum_{\eta_j=0}^{\infty} \pi(\eta_j|\delta_j = 1, \boldsymbol{\Upsilon_\eta}) f(\mathbf{y}|\delta_j = 1, \eta_j, \boldsymbol{\Upsilon_\eta})}$$

$$\propto\ e^{-\beta}\beta^{\eta_j}/(\eta_j!)\, f(y_j, y_{j+1}\,|\eta_j, \delta_j = 1, \alpha, \lambda, \epsilon),$$
$$\eta_j = 0, 1, 2, \ldots \tag{14}$$

with $f(y_j, y_{j+1}\,|\eta_j, \delta_j = 1, \alpha, \lambda, \epsilon)$ as given in (10).

Finally, the conditional posterior distribution for ϵ depends only on $\boldsymbol{\delta}$. Since the prior distribution of ϵ is $Be(h, g)$ the conditional posterior is given by

$$\epsilon|\mathbf{y}, \lambda, \boldsymbol{\eta}, \boldsymbol{\delta} \equiv \epsilon|\boldsymbol{\delta} \sim Be(h + k, g + n - 1 - k) \tag{15}$$

where k is the number of outliers (number of δ_j's equal to 1).

3.2 Computational Issues

The full conditional distributions of α, λ, $\boldsymbol{\delta} = (\delta_2, \ldots, \delta_n)$, $\boldsymbol{\eta} = (\eta_2, \ldots, \eta_n)$ and ϵ do not have standard forms, therefore we have to use the Metropolis-Hastings algorithm to draw a sample of a Markov chain which converges to the joint posterior distribution of the parameters. Since they are not log-concave densities we use the Gibbs methodology within the Metropolis step. In particular Adaptive Rejection Metropolis sampling—ARMS [9]—is used inside the Gibbs sampler. When the number of iterations is sufficiently large, the Gibbs draw can be regarded as a sample from the joint posterior distribution. Thus, complete distributions for the estimated parameters are obtained.

Two key issues in the successful implementation of this methodology are: deciding the length of the chain and the burn-in period and establishing the convergence of the chain. We use a burn-in period of M iterations and then iterate the Gibbs sampler for a further N iterations, but retain only each Lth element in the sample. This thinning strategy reduces the autocorrelation within the chain.

We now discuss the other relevant issue in the proposed Bayesian approach: the choice of the hyperparameters for prior distributions. Recall from Sect. 2 that the prior distributions for α and λ are $\text{Be}(a, b)$ and $\text{Ga}(c, d)$, respectively. In the absence of further or inside information we set $a = b = c = d = 0.001$ to use non informative prior distributions (Beta and Gamma distributions with large variability). For the prior distribution for the size of the outlier at time t, $\eta_t \sim Po(\beta)$ two approaches are pursued: an informative setup in which β_{info} is set equal to three times the standard deviation of the 1-step-ahead prediction error and also a non-informative setup with $\beta_{ninfo} = 30$ to reflect large variability. Finally, regarding the prior distribution for the probability of outliers occurrence, $\epsilon \sim Be(h, g)$, we choose $h = 5$, $g = 95$ to express the view that outliers occur occasionally. The posterior probability of outlier occurrence is then estimated and inspected to identify potential outliers. In an automated procedure a cut-off value, typically $c = 0.2$, may be used.

4 Illustration

In this section we document the performance of the above procedure with simulated data sets of 100 observations. In all the examples the Gibbs sampler is iterated $M + N = 5005$ times and the $L = 5$th value of the last $N = 2505$ iterations is kept, providing sample sizes of 501 values from which the probability of outlier occurrence at each time point as well as all the other parameter estimates are computed. The parameters α and λ are computed as the posterior mean. To ensure an integer value, the size of the outlier η is computed as the posterior median. The results are reported for $\beta_{ninfo} = 30$ since they do not differ from those obtained with β_{info}. We simulate time series from several PoINAR(1) processes without and with outliers of different sizes introduced at different times. The range of values considered for α and λ allow to illustrate the performance of the methodology for time series with small and large variability.

Table 1 reports the results from the application of the methodology to time series simulated from INAR(1) models with $\alpha = 0.15, 0.5, 0.85$ and $\lambda = 1, 3, 5$, all contaminated with three outliers of different sizes. The data generating model is identified in the column headed *Model* by the parameters of the contaminated PoINAR(1) model: α, λ and η_S, which indicates contamination with an outlier of size η at time S. Finally, all the outliers detected by the algorithm, that is all the time points for which the posterior probability of outlier occurrence is over the threshold $\hat{p} = 0.2$, are indicated by the time of occurrence, estimated size, and associated posterior probability.

The results indicate that the procedure detects all the additive outliers in PoINAR(1) time series with high (near 1) estimated probabilities of occurrence. Moreover, the occurrence of false detections was null in the contaminated time series as well as in outlier free time series whose results are not reported here.

Note that the convergence of the MCMC algorithm was duly analysed with the usual diagnostic tests available in [13].

4.1 IP Data Example

Let us consider once again the motivating example of Sect. 1, regarding the number of different IP addresses accessing the server of the Department of Statistics of the University of Würzburg on November 29th, 2005, between 10 am and 6 pm, represented in Fig. 1 [16]. The sample mean and variance of the series are $\bar{x} = 1.32, \hat{\sigma}^2 = 1.39$. The autocorrelation and partial autocorrelation functions indicate that a model of order one is appropriate. CLS estimates for α and λ are $\hat{\alpha} = 0.22$ and $\hat{\lambda} = 1.03$, respectively. The result of applying the proposed methodology is represented in Fig. 1b indicating the possible occurrence of an outlier at time $t = 224$. The estimated size of the outlier is $\hat{\eta} = 7$. It is interesting to note that setting the time of the outlier to $t = 224$ and using the results from [2] the CLS estimate for η is $\hat{\eta}_{CLS} = 6.73$. Removing the effect of the outlier at $t = 224$ the mean and variance of the resulting series are 1.29 and 1.2, respectively. The autocorrelation and partial autocorrelation functions still indicate that a model of order one is appropriate. CLS estimates for the parameters are now $\hat{\alpha}_{CLS} = 0.29$ and $\hat{\lambda}_{CLS} = 0.91$ in accordance with the estimates obtained from the Gibbs sampling, $\hat{\alpha}_{Bayes} = 0.27$ and $\hat{\lambda}_{Bayes} = 0.89$, whose posterior distributions are represented in Fig. 2.

Table 1 Results from Gibbs sampling in simulated INAR(1) time series with parameters α, λ, contaminated at time S with an outlier of size η_S

Model		Estimates	Outliers detected			Model		Estimates	Outliers detected		
			Time	Size	Probability				Time	Size	Probability
α	0.15	0.07				α	0.85	0.90			
λ	1	1.27				λ	1	0.83			
η_{34}	7		34	8	0.89	η_9	7		9	7	0.87
η_{50}	5		50	8	0.89	η_{29}	13		29	12	0.99
η_{63}	9		63	9	1.00	η_{75}	18		75	19	0.99
α	0.15	0.01				α	0.85	0.86			
λ	3	3.60				λ	3	2.62			
η_{24}	9		24	11	0.99	η_9	31		9	29	0.92
η_{28}	13		28	13	0.99	η_{29}	13		29	10	0.99
η_{65}	6		65	7	0.99	η_{75}	22		75	22	0.99
α	0.15	0.01				α	0.85	0.85			
λ	5	5.3				λ	5	4.60			
η_{33}	7		33	11	0.99	η_{38}	40		38	37	0.92
η_{70}	12		70	13	1.00	η_{41}	28		41	27	0.99
η_{10}	16		10	18	0.98	η_{78}	17		78	20	0.99
α	0.5	0.41				α	0.85	0.90			
λ	1	0.94				λ	1	0.83			
η_9	10		9	11	0.90	η_9	7		9	7	0.87
η_{27}	4		27	7	0.85	η_{29}	13		29	12	0.99
η_{97}	7		97	8	0.81	η_{75}	18		75	19	0.99
α	0.5	0.59				α	0.85	0.86			
λ	3	2.28				λ	3	2.62			
η_{99}	17		99	17	0.99	η_9	31		9	29	0.92
η_{17}	12		17	16	0.99	η_{29}	13		29	10	0.99
η_7	7		7	7	0.97	η_{75}	22		75	22	0.99
α	0.5	0.51				α	0.85	0.85			
λ	5	4.30				λ	5	4.60			
η_{29}	10		29	14	0.91	η_{38}	40		38	37	0.92
η_{22}	21		22	22	0.99	η_{41}	28		41	27	0.99
η_{19}	15		19	17	0.99	η_{78}	17		78	20	0.99

5 Concluding Remarks

In this paper, a retrospective analysis of the Poisson INAR(1) model for time series of counts under a Bayesian approach is carried out. The Bayesian framework is more flexible than a classical likelihood approach leading to the identification of observations that may require further scrutinizing. In fact, by estimating the probability that each observation is affected by an outlier under a certain model, the procedure is useful for detecting suspicious observations but also possible model

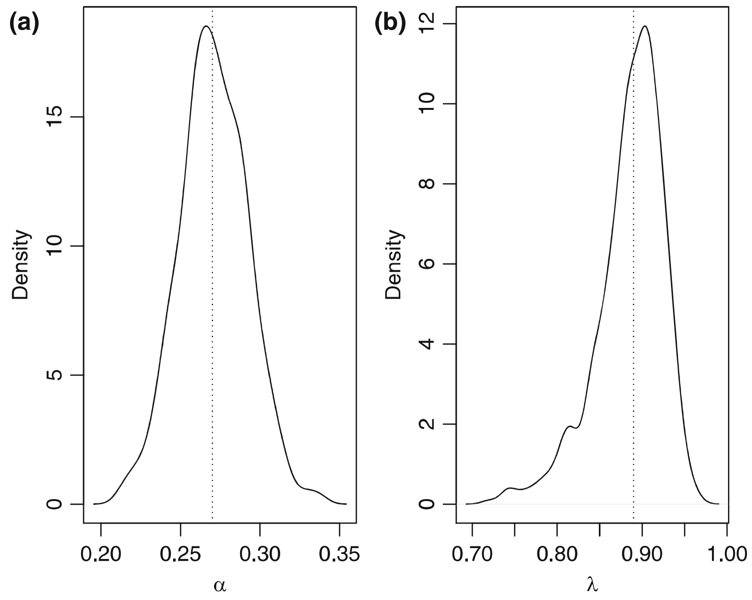

Fig. 2 Posterior distribution of α and λ. The *dotted lines* represent the estimates $\hat{\alpha}_{Bayes} = 0.27$ and $\hat{\lambda}_{Bayes} = 0.89$

inadequacies since the presence of many outliers may indicate the wrong choice of model. There are, thus, several extensions to this work that are being investigated, namely: the detection of patches of outliers that may cause masking and swamping effects; development of strategies for including different outliers effects and other interventions; other distributional assumptions; higher-order models.

Acknowledgements This work was supported by Portuguese funds through the CIDMA—Center for Research and Development in Mathematics and Applications, and the Portuguese Foundation for Science and Technology (FCT-Fundação para a Ciência e a Tecnologia), within project PEst-OE/MAT/UI4106/2014.

References

1. Al-Osh, M.A., Alzaid, A.A.: First-order integer-valued autoregressive (INAR(1)) process. J. Time Ser. Anal. **8**, 261–275 (1987)
2. Barczy, M., Ispány, M., Pap, G., Scotto, M., Silva, M.E.: Additive outliers in INAR(1) models. Stat. Pap. **53**, 935–949 (2011)
3. Barczy, M., Ispány, M., Pap, G., Scotto, M., Silva, M.E.: Innovational outliers in INAR(1) models. Commun. Stat. Theory Methods **39**, 3343–3362 (2010)
4. Chang, I., Tiao, G.C., Chen, C.: Estimation of time series parameters in the presence of outliers. Technometrics **30**, 193–204 (1988)

5. Chen, C.W.: Detection of additive outliers in bilinear time series. Comput. Stat. Data Anal. **24**, 283–294 (1997)
6. Chen, C., Liu, L.M.: Joint estimation of model parameters and outlier effects in time series. J. Am. Stat. Assoc. **88**, 284–297 (1993)
7. Fokianos, K., Fried, R.: Interventions in ingarch processes. J. Time Ser. Anal. **31**, 210–225 (2010)
8. Gelfand, A.E., Smith, A.F.M.: Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. **85**, 398–409 (1990)
9. Gilks, W.R., Best, N.G., Tan, K.K.C.: Adaptive rejection metropolis sampling within Gibbs sampling. J. R. Stat. Soc. Ser. C (Appl. Stat.) **44**, 455–472 (1995)
10. Jung, R., Tremayne, A.: Useful models for time series of counts or simply wrong ones? Adv. Stat. Anal. **55**, 59–91 (2011)
11. Justel, A., Peña, D., Tsay, R.: Detection of outlier patches in autoregressive time series. Stat. Sin. **11**, 651–673 (2001)
12. McKenzie, E.: Some simple models for discrete variate time series. J. Am. Water Resour. Assoc. **21**, 645–650 (1985)
13. Plummer, M., Best, N., Cowles, K., Vines, K.: CODA: convergence diagnosis and output analysis for MCMC. R News **6**, 7–11 (2006)
14. Silva, I., Silva, M., Pereira, I., Silva, N.: Replicated INAR(1) processes. Methodol. Comput. Appl. Probab. **7**, 517–542 (2005)
15. Tsay, R.S.: Time series model specification in the presence of outliers. J. Am. Stat. Assoc. **81**, 132–141 (1986)
16. Weiß, C.H.: Controlling correlated processes of poisson counts. Qual. Reliab. Eng. Int. **23**, 741–754 (2007)