

# Bayesian outlier detection in non-Gaussian AutoRegressive time series\*

Maria Eduarda Silva<sup>†</sup> Isabel Pereira<sup>‡</sup> Brendan McCabe<sup>§</sup>

## Abstract

This work investigates outlier detection and modelling in non Gaussian autoregressive time series models with margins in the class of a convolution closed parametric family. This framework allows for a wide variety of models for count and positive data types. The paper investigates additive outliers which do not enter the dynamics of the process but whose presence may adversely influence statistical inference based on the data. The Bayesian approach proposed here allows one to estimate, at each time point, the probability of an outlier occurrence and its corresponding size thus identifying the observations that require further investigation. The methodology is illustrated using simulated and observed data sets.

## 1 Introduction

Observations that look discordant in a data set are often encountered in time series. These observations, outliers according to Barnett and Lewis (1994), may result from gross errors in the collection or processing of the data, or from non repetitive exogenous interventions such as strikes, wars and other sudden or unexpected shocks. Neglecting the presence of outliers in a time series hinders meaningful statistical inference, leading to model misspecification, biased parameter estimation, and poor forecasts. Since the fundamental article of Fox (1972) there has been an increasing amount of literature on outliers in time series. Two major approaches for outlier handling may be distinguished. One approach advocates the use of robust estimators to reduce the effect of the outlying observations. Denby and Martin (1979) and Martin and Yohai (1986), among others use this approach for ARIMA models, Beran (1994) for Gaussian long-memory models, Chan and Cheung (1994) for threshold autoregressive

---

\* Accepted author's manuscript (AAM) to be published in [Journal Time Series Analysis]. [DOI:10.1111/jtsa.12439]

<sup>†</sup> Corresponding author: mesilva@fep.up.pt, Faculdade de Economia, Universidade do Porto, Rua Dr Roberto Frias, s/n, 4200-464 Porto, Portugal & CIDMA

<sup>‡</sup> isabel.pereira@ua.pt, Departamento de Matemática, Universidade de Aveiro, Portugal & CIDMA

<sup>§</sup> Brendan.Mccabe@liverpool.ac.uk, Management School, Chatham Building Chatham Street, University of Liverpool, L69 7ZH

models, Kleiner and Martin (1979) consider robust frequency domain analysis. However, this approach often leads to ignoring observations which may be the most interesting. The fact that outlying observations have led to the discovery of important underlying phenomena (e.g. holes in the atmospheric ozone layer), suggests the importance of the alternative approach which detects possible outliers and estimates the corresponding sizes. Several methodologies for this, and other intervention effects, have been established in the framework of Gaussian linear time series. The emphasis is on iterative procedures and likelihood based statistics, Tsay (1986); Chang et al. (1988); Chen and Liu (1993). In particular, the residual based iterative procedure of Chen and Liu (1993) is standard now in retrospective analysis of outliers with linear models. More recently, several authors considered the detection of outliers in nonlinear time series models: Chen (1997) considered additive outliers in bilinear models and Battaglia and Orfei (2005) considered detecting the presence of outliers when the series is generated by a general nonlinear model that includes as particular cases the bilinear, the self-exciting threshold autoregressive (SETAR) model and the exponential autoregressive model, among others. In financial time series modeling, Doornik and Ooms (2005); Carnero et al. (2007); Grané and Veiga (2010) and Hotta and Tsay (2012) presented procedures for detecting outliers in GARCH models, while Abanto-Valle et al. (2010) and Wang et al. (2011) develop Bayesian methods for treating outliers in stochastic volatility models.

Apparently there is not much work, so far, on the analysis of outliers in positive-valued and count time series. These time series arise in a wide variety of fields including: telecommunications, actuarial science, computer science, economics, epidemiology, finance, hydrology, meteorology and environmental studies. These data are naturally non Gaussian and typically right-skewed, creating a need for especially designed models and procedures. In fact, the usual procedures to detect outliers in time series would suggest real-valued sizes for outliers and these are not appropriate for integer-valued time series. In the context of count data, Silva and Pereira (2015) proposed an approach for detecting single additive outliers in Poisson integer-valued autoregressive processes while Fokianos and Fried (2010) and Fried et al. (2015) analyse outliers and other interventions for count time series within the framework of generalised linear models. In this paper we consider a general class of convolution closed infinitely divisible (CCID) models as proposed by Joe (1996).

The CCID class, while being simple and flexible, is useful for modelling positive valued and integer valued time series possessing an autoregressive structure with non negative serial correlation. The CCID class includes the Poisson, negative binomial (with fixed probability parameter), gamma (with fixed scale parameter), generalised Poisson (with one fixed parameter), inverse Gaussian (with one fixed parameter) and Gaussian distributions. Here we consider the problem of modelling outliers in the CCID class, focussing on the integer-valued time series models although the same approach may be taken for any members of the class. We conduct a retrospective analysis of purely additive outliers, occurring at unknown time locations, and which do not enter the dynamics of the model. To this end, we specify a state space model where the observations

are a contaminated version of a state (latent) CCID process. The unknown contamination mechanism is a sequence of independent variables, each one containing information on the occurrence (or not) of a contaminating outlier and its size. A natural approach to estimate such a model is in a Bayesian framework, as this does not require advance knowledge of the number and location of outliers in the span of the series. Moreover, this approach allows estimation of the contamination process itself, providing, at each time point, the probability that outlier occurred and an estimate of the corresponding size.

The structure of the paper is as follows. Section 2 introduces the CCID class for non-Gaussian time series, section 3 describes the contaminated state space model, the Bayesian priors and MCMC schemes to compute the posteriors. The methodology is illustrated with synthetic Poisson integer-valued autoregressive time series with multiple outliers and two observed time series in section 4. Some final remarks are made in section 5.

## 2 Convolution closed models for non Gaussian time series

Let  $R(\cdot)$  denote a random operator such that the conditional distribution of  $R(X)$  given  $X = x$  has a distribution  $G_{\alpha\mu, (1-\alpha)\mu, x}$ , with  $\mu > 0$  and  $0 < \alpha < 1$ , and  $R(X)$  has a marginal distribution  $F_{\alpha\mu}$  when  $X$  has a distribution  $F_\mu$  belonging to the CCID parametric family. In particular,  $F_\mu$  is assumed to satisfy  $F_{\mu_1} * F_{\mu_2} = F_{\mu_1 + \mu_2}$ , where  $*$  is the convolution operator. A time series process  $\{X_t; t = 0, \pm 1, \pm 2, \dots\}$  satisfying the CCID(1) equation

$$X_t = R_t(X_{t-1}) + e_t \quad (1)$$

where the innovations  $e_t$  are independently and identically distributed (i.i.d.) with distribution  $F_{(1-\alpha)\mu}$  and  $\{R_t(\cdot) : t = 0, \pm 1, \pm 2, \dots\}$  are independent replications of the operator  $R(\cdot)$ , is a stationary time series with margin  $F_\mu$  i.e.  $X_t \sim F_\mu$  and the lag 1 autocorrelation  $\rho(1) = \alpha$  (Joe, 1996). Model (1) encompasses many non Gaussian AR(1) time series models proposed in the literature for integer valued and positive valued time series.

If one chooses  $F_{(1-\alpha)\mu}$  as  $\text{Poisson}((1-\alpha)\mu)$ , and the random operation as the usual binomial thinning operation (based on underlying Bernoulli random variables)  $R_t(X_{t-1}) = \sum_{i=1}^{X_{t-1}} \xi_{ti}$ ,  $\xi_{ti} \sim^{iid} \text{Ber}(\alpha)$  then  $F_\mu$  is  $\text{Poisson}(\mu)$  and the Poisson integer-valued autoregressive model, PoINAR(1), model proposed by McKenzie (1988) and Al-Osh and Alzaid (1987) is recovered. In this case,  $R_t(X_{t-1})$  is denoted by  $\alpha \circ X_{t-1}$  and the probability mass function, pmf, of  $R_t$  given  $X_{t-1}$ , is  $g(s|X_{t-1}; \alpha) = \binom{X_{t-1}}{s} \alpha^s (1-\alpha)^{X_{t-1}-s}$ , where  $\binom{\cdot}{\cdot}$  is the usual combinatorial factor. Similarly, if one needs to account for overdispersion present in the count time series, then a negative binomial marginal may be entertained. In this case,  $F_\mu$  is the negative binomial distribution with parameters  $\mu$  and  $\xi$

(fixed),  $NB(\mu, \xi)$  and pmf

$$f(s; \mu, \xi) = \frac{\Gamma(\mu + s)}{\Gamma(s + 1)\Gamma(\mu)} \xi^\mu (1 - \xi)^s, \quad s = 0, 1, \dots,$$

where  $\Gamma(\cdot)$  is the gamma function<sup>1</sup>. Then, with arrivals distributed as  $F_{(1-\alpha)\mu}$  and  $G_{\alpha\mu, (1-\alpha)\mu, x}$  set to be beta-binomial with parameters  $x, \mu\alpha$  and  $\mu(1 - \alpha)$ ,  $BB(x, \mu\alpha, \mu(1 - \alpha))$  with pmf

$$g(s|x; \mu, \alpha) = \binom{x}{s} \frac{B(\alpha\mu + s, (1 - \alpha)\mu + x - s)}{B(\alpha\mu, (1 - \alpha)\mu)}, \quad s = 0, 1, \dots,$$

where  $B(\cdot, \cdot)$  is the beta function, the observations are also marginally negative binomial i.e.  $X_t \sim F_\mu$ . This model corresponds to the beta-binomial thinning of McKenzie (1986). Table 1 summarizes some particular non-Gaussian models, indicating the marginal distribution, the random operation and its pmf or pdf and the set of parameters  $\theta$  for each model. The Poisson INAR(1), PoINAR(1), the negative binomial INAR(1), NBINAR(1), and the generalised Poisson INAR(1), GPINAR(1) (Alzaid and Al-Osh, 1993), have been widely used in the literature to model time series of counts, see *inter alia* Jung and Tremayne (2011) and Han and McCabe (2013), among others. The gamma AR(1) which is the same model proposed by Lewis et al. (1989) is suitable for positive valued time series.

Since model (1) is Markovian (Joe, 1996), given a time series  $\mathbf{x} = (X_1, \dots, X_n)$ , the conditional likelihood is as follows

$$L(\theta) = \prod_{t=2}^n f_{X_t|X_{t-1}}(x_t|x_{t-1}) \quad (2)$$

with

$$f_{X_t|X_{t-1}}(k|l) = P(X_t = k|X_{t-1} = l) = \sum_{j=0}^{\min\{k,l\}} g(j|l)P(e_t = k - j) \quad (3)$$

for discrete valued models such as Poisson, generalised Poisson and negative binomial and

$$f_{X_t|X_{t-1}}(y|x) = \int_0^\infty g(w|x)f_{e_t}(y - w)dw \quad (4)$$

for the continuous valued models such as gamma.

### 3 Bayesian outlier modelling in CCID(1) time series

Given a model as in (1), a basic question is whether it properly describes all the observations of a given time series, or whether some observations have been

<sup>1</sup>In applications we do not consider  $\xi$  to be fixed but we continue to use the CCID nomenclature for ease of reference when extra parameters are required

Marginal distribution	Random operator	$g(s X_{t-1}; \alpha)$	Innovations	$\theta$
Poisson $Po(\mu)$	binomial thinning	$\binom{X_{t-1}}{s} \alpha^s (1-\alpha)^{X_{t-1}-s}$	$Po(\lambda)$	$(\mu, \alpha)$
Negative binomial $NB(\mu, \xi)$	beta binomial thinning	$\binom{X_{t-1}}{s} \frac{B(\alpha\mu+s, \lambda+X_{t-1}-s)}{B(\alpha\mu, \lambda)}$	$NB(\lambda, \xi)$	$(\mu, \alpha, \xi)$
Generalised Poisson $GP(\mu, \xi)$	quasi binomial thinning	$\binom{X_{t-1}}{s} \alpha (\alpha + s(\xi/\mu))^{s-1} (1-\alpha-s(\xi/\mu))^{X_{t-1}-s}$	$GP(\lambda, \xi)$	$(\mu, \alpha, \xi)$
Gamma( $\mu, \nu$ )	$A_t X_{t-1}$ $A_t \sim \text{Beta}(\alpha\mu, \lambda)$		Gamma( $\lambda, \nu$ )	$(\mu, \alpha, \nu)$

Table 1: Methods for constructing non Gaussian AR(1) models with specified marginals  $F_\mu$  and innovations  $e_t \stackrel{iid}{=} F_\lambda$ ,  $\lambda = \mu(1-\alpha)$ .  $B(\cdot, \cdot)$  denotes the beta function.

affected by extraordinary effects which do not enter the dynamics of the process. Here we describe a model for dealing with additive outliers in CCID(1) processes and develop a Bayesian approach to achieve its estimation.

### 3.1 Modelling additive outliers

Additive outliers correspond to the contamination of single observations so that the effect is not carried over to subsequent observations by the dynamics. Therefore, additive outliers can be modelled assuming that a contaminated process,  $Y_t$  is observed instead of the clean CCID(1) process,  $X_t$ . Thus, we consider (1) as a latent process and add an observation equation, contaminating, with probability  $p_t$ , each clean value  $X_t$  with an additive outlier of random size  $\eta_t$ . Hence

$$\begin{aligned}
Y_t &= X_t + \eta_t \delta_t, \\
X_t &= R_t(X_{t-1}) + e_t \\
\delta_t &\sim \text{Ber}(p_t), \quad \eta_t \sim F_{\beta_t} \\
\eta_t \text{ and } \delta_t &\text{ independent.}
\end{aligned} \tag{5}$$

The binary sequence  $\delta_t$  indicates, for each  $t$ , the occurrence (or not) with probability  $p_t$ , of an outlier. The size of the outlier at time  $t$ ,  $\eta_t$ , is a random variable with the same support as  $X_t$  and mean  $\beta_t$  :  $\eta_t \sim Po(\beta_t)$  for Poisson,

negative binomial and generalised Poisson models and  $\eta_t \sim \text{Gamma}(\beta, 1)$  for the gamma model (c.f. table 1). We assume that the occurrence of an outlier is independent of its size and of the latent clean process, that is  $\delta_t$  and  $\eta_t$  are independent and independent of  $X_t$ . This model is quite flexible since we do not fix either the probability of occurrence nor the outlier size over the span of the time series. Once priors have been specified, the Bayesian Additive Integer Outlier Detection (BAIOD) method inspects the plots of the posterior values of  $p_t$  to determine the presence of outliers or not. Automatic outlier detection can be accomplished by setting a threshold probability value which may depend on the application.

### 3.2 Prior Distributions

A Bayesian analysis of the contaminated model (5) requires two sets of prior specifications; one for the parameters of the CCID model,  $\theta$ , and a second for the parameters of the outlier generating process.<sup>2</sup>

In the absence of prior information we use weakly informative prior distributions for  $\theta$ . Thus, for the parameter  $0 < \alpha < 1$  we choose a non-informative beta prior with parameters  $(a, b)$ , while for the positive parameter  $\mu$  we choose a non-informative Gamma(shape, rate) prior with parameters  $(c, d)$ . Additionally, the prior for the parameter  $\xi \in ]0, 1[$  in NBINAR and GPINAR models is a Beta  $(e, f)$  and for  $\nu > 0$  in the gamma model is a gamma distribution.

The specification of priors for the contamination parameters  $p_t$  of the indicator variables  $\delta_t$  and those for the mean size of the outliers,  $\beta_t$ , is a more sensitive issue as these affect directly the frequency and magnitude of the outliers.

The priors for  $p_t$  are Beta  $(g, h)$  and for the mean size of the outliers  $\beta_t$  we use non informative gamma distributions with parameters  $(l, m)$ . The hyperparameters  $a$  to  $m$  are specified on a case by case basis.

### 3.3 Posterior Distributions

For the remainder of the paper, we denote by  $\mathbf{y} = (Y_1, \dots, Y_n)$  the observed time series and assume that there is no outlier in the first observation, that is  $Y_1 = X_1$ . Let  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$ ,  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)$ . The posterior distribution for  $(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta})$  is then

$$\pi(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta} | \mathbf{y}) \propto \pi(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}) L(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}) \quad (6)$$

where  $\pi(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta})$  denotes the prior distribution for  $(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta})$  and  $L(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta})$  the (conditional) likelihood.

To illustrate the complexity of the posterior distribution consider the model PoINAR(1). In this case,  $\boldsymbol{\theta} = (\mu, \alpha)$  and using the prior distributions and the independence assumptions, we have

<sup>2</sup>Note, within a non-Bayesian framework (conditional) likelihood based inference is difficult, if not impossible, since the uncontaminated values enter the dynamics of the process at contaminated but unknown time points.

$$\pi(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}) \propto e^{-d\mu} \mu^{c-1} \alpha^{a-1} (1-\alpha)^{b-1} \prod_{t=1}^n p_t^{\delta_t+g-1} (1-p_t)^{h-\delta_t} e^{-(m+1)\beta_t} \beta_t^{l-1+\eta_t} / \eta_t! \quad (7)$$

and

$$\pi(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta} | \mathbf{y}) \propto \pi(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}) L(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}) \quad (8)$$

with

$$L(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}) = \prod_{t=2}^n \sum_{i=0}^{M_t} \frac{(\mu(1-\alpha))^{Y_t-\delta_t\eta_t-i}}{(Y_t-\delta_t\eta_t-i)!} \binom{Y_{t-1}-\delta_{t-1}\eta_{t-1}}{i} e^{-\mu(1-\alpha)} \alpha^i (1-\alpha)^{Y_{t-1}-\delta_{t-1}\eta_{t-1}-i} \quad (9)$$

where  $M_t = \min(Y_{t-1} - \eta_{t-1}\delta_{t-1}, Y_t - \eta_t\delta_t)$ .

Thus, given the complexity of the posterior distribution, Markov Chain Monte Carlo techniques are required for sampling from the full conditional distributions.

### 3.4 Full conditional distributions

To obtain the full conditional distributions for each parameter we condition the posterior on all the other parameters (and the data). Thus, the full conditional distributions for the parameters  $\theta_i$  in  $\boldsymbol{\theta}$  and  $\mathbf{p}$  are easily obtained. To obtain the full conditional distributions for  $\boldsymbol{\delta}$  and  $\boldsymbol{\eta}$  we must also take into account the presence of outliers. In what follows, define for each  $j = 2, \dots, n$ ,  $\boldsymbol{\Upsilon}_\delta = (\boldsymbol{\theta}, \boldsymbol{\delta}_{(-j)}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta})$  and  $\boldsymbol{\Upsilon}_\eta = (\boldsymbol{\theta}, \boldsymbol{\delta}_{(-j)}, \boldsymbol{\eta}_{(-j)}, \mathbf{p}, \boldsymbol{\beta})$  where  $\boldsymbol{\nu}_{(-j)}$  denote the vectors  $\boldsymbol{\nu}$  with the  $j$ th component deleted. The full conditional distributions are illustrated for the PoINAR(1) model. The Appendix establishes likelihood and full conditional distributions for the NBINAR(1).

#### 3.4.1 Full Conditional Distributions for $\theta$

The full conditional distributions for the parameters  $\theta_i$  in  $\boldsymbol{\theta}$  are obtained from

$$\pi(\theta_i | \boldsymbol{\theta}_{(-i)}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}, \mathbf{y}) \propto \pi(\theta_i) L(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}).$$

In the case of the PoINAR(1),  $\boldsymbol{\theta} = (\mu, \alpha)$  and thus

$$\pi(\mu | \alpha, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}, \mathbf{y}) \propto \mu^{c-1} e^{-(d+(n-1)(1-\alpha))\mu} \prod_{t=2}^n \sum_{i=0}^{M_t} \mu^{Y_t-\eta_t\delta_t-i} K(t, i) \quad (10)$$

with  $K(t, i) = \frac{(1-\alpha)^{Y_t - \eta_t \delta_t - i}}{(Y_t - \eta_t \delta_t - i)!} \binom{Y_{t-1} - \eta_{t-1} \delta_{t-1}}{i} \alpha^i (1-\alpha)^{Y_{t-1} - \eta_{t-1} \delta_{t-1} - i}$  and

$$\begin{aligned} \pi(\alpha | \mu, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}, \mathbf{y}) &\propto \alpha^{a-1} (1-\alpha)^{b-1} e^{\alpha \mu (n-1)} \\ &\prod_{t=2}^n \sum_{i=0}^{M_t} (1-\alpha)^{Y_t - \eta_t \delta_t - i} \alpha^i (1-\alpha)^{Y_{t-1} - \eta_{t-1} \delta_{t-1} - i} T(t, i) \end{aligned} \quad (11)$$

with  $T(t, i) = \frac{\mu^{Y_t - \eta_t \delta_t - i}}{(Y_t - \eta_t \delta_t - i)!} \binom{Y_{t-1} - \eta_{t-1} \delta_{t-1}}{i}$

### 3.4.2 Full Conditional Distribution for $\delta_j$

Recalling that  $\delta_j$  takes only the values 0 and 1, the specification of  $\pi(\delta_j | \boldsymbol{\Upsilon}_\delta, \mathbf{y})$  reduces to the computation of

$$\begin{aligned} \pi_j &= \text{P}(\delta_j = 1 | \boldsymbol{\Upsilon}_\delta, \mathbf{y}) = \frac{\text{P}(\delta_j = 1, \mathbf{y} | \boldsymbol{\Upsilon}_\delta)}{f(\mathbf{y} | \boldsymbol{\Upsilon}_\delta)} \\ &= \frac{\text{P}(\delta_t = 1 | \boldsymbol{\Upsilon}_\delta) f(\mathbf{y} | \delta_j = 1, \boldsymbol{\Upsilon}_\delta)}{\text{P}(\delta_t = 1 | \boldsymbol{\Upsilon}_\delta) f(\mathbf{y} | \delta_j = 1, \boldsymbol{\Upsilon}_\delta) + (1 - \text{P}(\delta_t = 1 | \boldsymbol{\Upsilon}_\delta)) f(\mathbf{y} | \delta_j = 0, \boldsymbol{\Upsilon}_\delta)} \end{aligned} \quad (12)$$

with  $\text{P}(\delta_t = 1 | \boldsymbol{\Upsilon}_\delta) = p_j$ .

To obtain  $f(\mathbf{y} | \delta_j = 1, \boldsymbol{\Upsilon}_\delta)$  (and  $f(\mathbf{y} | \delta_j = 0, \boldsymbol{\Upsilon}_\delta)$ ) we first note that  $Y_t$  inherits the Markovian property of  $X_t$  and also that at time  $t = j - 1$  an outlier may have occurred and at time  $j + 1$  an outlier may occur. Thus taking into account these possible outlier configurations we write

$$\begin{aligned} f(\mathbf{y} | \delta_j = 1, \boldsymbol{\Upsilon}_\delta) &= \\ &\left[ f(Y_j | Y_{j-1}, \delta_j = \delta_{j-1} = 1) f(Y_{j+1} | Y_j, \delta_j = \delta_{j+1} = 1) \text{P}(\delta_{j+1} = 1) \right. \\ &\quad \left. + f(Y_j | Y_{j-1}, \delta_j = \delta_{j-1} = 1) f(Y_{j+1} | Y_j, \delta_j = 1, \delta_{j+1} = 0) (1 - \text{P}(\delta_{j+1} = 1)) \right] \\ &\mathbf{I}_{\{\delta_{j-1}=1\}} \\ &+ \left[ f(Y_j | Y_{j-1}, \delta_j = 1, \delta_{j-1} = 0) f(Y_{j+1} | Y_j, \delta_j = \delta_{j+1} = 1) \text{P}(\delta_{j+1} = 1) \right. \\ &\quad \left. + f(Y_j | Y_{j-1}, \delta_j = 1, \delta_{j-1} = 0) f(Y_{j+1} | Y_j, \delta_j = 1, \delta_{j+1} = 0) (1 - \text{P}(\delta_{j+1} = 1)) \right] \\ &\mathbf{I}_{\{\delta_{j-1}=0\}} \end{aligned} \quad (13)$$

and

$$\begin{aligned}
& f(\mathbf{y}|\delta_j = 0, \boldsymbol{\Upsilon}_\delta) = \\
& \left[ f(Y_j|Y_{j-1}, \delta_j = 0, \delta_{j-1} = 1)f(Y_{j+1}|Y_j, \delta_j = 0, \delta_{j+1} = 1)P(\delta_{j+1} = 1) \right. \\
& \quad \left. + f(Y_j|Y_{j-1}, \delta_j = 0, \delta_{j-1} = 1)f(Y_{j+1}|Y_j, \delta_j = \delta_{j+1} = 0)(1 - P(\delta_{j+1} = 1)) \right] \\
& \mathbf{I}_{\{\delta_{j-1}=1\}} \\
& + \left[ f(Y_j|Y_{j-1}, \delta_j = \delta_{j-1} = 0)f(Y_{j+1}|Y_j, \delta_j = 0, \delta_{j+1} = 1)P(\delta_{j+1} = 1) \right. \\
& \quad \left. + f(Y_j|Y_{j-1}, \delta_j = \delta_{j-1} = 0)f(Y_{j+1}|Y_j, \delta_j = \delta_{j+1} = 0)(1 - P(\delta_{j+1} = 1)) \right] \\
& \mathbf{I}_{\{\delta_{j-1}=0\}}
\end{aligned} \tag{14}$$

where  $\mathbf{I}_A$  is the indicator function of the set  $A$  and the conditional distributions  $f(Y_t|Y_{t-1}, \delta_t = u, \delta_{t-1} = v)$ ,  $u, v = 0, 1$ ,  $t = j, j + 1$ , are defined as

$$f(Y_t|Y_{t-1}, \delta_t = u, \delta_{t-1} = v) = f_{Y_t^*|Y_{t-1}^*}(Y_t^*|Y_{t-1}^*)$$

with

$$Y_t^* = \begin{cases} Y_t, & \delta_t = 0 \\ Y_t - \eta_t, & \delta_t = 1 \end{cases}$$

and  $f_{Y_t^*|Y_{t-1}^*}(Y_t^*|Y_{t-1}^*)$  obtained from (3) or (4). In particular for the PoINAR(1) we have

$$\begin{aligned}
f(Y_t^*|Y_{t-1}^*, \delta_t = u, \delta_{t-1} = v) &= e^{-(\mu(1-\alpha))} \\
& \sum_{i=0}^{M_t^*} \binom{Y_{t-1}^*}{i} \alpha^i (1-\alpha)^{Y_{t-1}^*-i} \frac{(\mu(1-\alpha))^{Y_t^*-i}}{(Y_t^*-i)!}
\end{aligned} \tag{15}$$

with  $u, v = 0, 1$ ,  $t = j, j + 1$ ,  $M_t^* = \min(Y_{t-1}^*, Y_t^*)$ .

### 3.4.3 Full Conditional Distributions for $p_j$

The conditional posterior distribution for  $p_j$  depends only on  $\delta_j$ . Since the prior distribution of  $p_j$  is  $\text{Beta}(g, h)$  the full conditional is given by

$$\pi(p_j|\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}_{(-j)}\boldsymbol{\beta}, \mathbf{y}) \propto p_j^{\delta_j+g-1} (1-p_j)^{h-\delta_j} \tag{16}$$

that is a  $\text{Beta}(\delta_j + g, h - \delta_j + 1)$  distribution.

### 3.4.4 Full Conditional Distributions for $\eta_j$

The hierarchical structure of model (5) implies that  $\eta_j$  is independent of all other parameters except of  $\delta_j$ . In fact, if  $\delta_j = 0$ , no outlier at  $t = j$ , there is no information about  $\eta_j$  except that contained in the prior  $\pi(\eta_j)$ . However, if  $\delta_j = 1$ ,  $\mathbf{y}$  contains information about  $\eta_j$ . Therefore we have

$$\begin{aligned} \pi(\eta_j | \delta_j = 1, \mathbf{Y}_\eta, \mathbf{y}) \\ \propto \pi(\eta_j) f(\mathbf{y} | \eta_j, \delta_j = 1, \mathbf{Y}_\eta) \end{aligned} \quad (17)$$

for  $\eta_j = 0, 1, 2, \dots$  with  $f(\mathbf{y} | \eta_j, \delta_j = 1, \mathbf{Y}_\eta)$  as given in (13).

In case of the PoINAR(1),  $\pi(\eta_j) = \frac{e^{-\beta_j} \beta_j^{\eta_j}}{\eta_j!}$ .

### 3.4.5 Full Conditional Distributions for $\beta_j$

Finally, the conditional posterior distribution for  $\beta_j$  depends only on  $\eta_j$ . Then

$$\pi(\beta_j | \boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \mathbf{y}) \propto e^{-(m+1)\beta_j} \beta_j^{\eta_j + l - 1} \quad (18)$$

## 4 Illustrations

In this section we document the performance of the above procedure with count time series simulated from equi and over dispersed CCID(1) processes with additive outliers introduced at fixed time points. The simulation allows control of the number, size and configuration of the outliers. In addition, we apply the procedure to two time series of monthly counts: the first is of poliomyelitis cases in the US between 1970 and 1983 and the second is of bruise injuries incurred during logging operations in British Columbia, Canada between 1985 and 1994. Taken together, these examples suggest that the new procedure is effective at detecting potential outliers and is not prone to excessive false alarms. We compare the results from our BAIOD approach with two other generic outlier detection procedures available in the literature. The comparators are from the R packages `forecast` (Hyndman and Khandakar, 2008; Hyndman, 2017) and `tsoutliers` (de Lacalle, 2017). The first uses function `tsoutliers` (1) which implements a non parametric outlier detection procedure based on exponential smoothing while the second, `tso` (2), implements a residual based outlier detection strategy designed for ARMA models. We begin by addressing some computational issues that arise in the implementation of the procedure.

### 4.1 Computational Issues and Prior Hyper-Parameters

The full conditional distributions are not standard distributions and therefore MCMC methods (Metropolis-Hastings and Gibbs sampling) are needed. In particular, the Adaptive Rejection Metropolis Sampling (ARMS) is used for the parameters  $\boldsymbol{\theta}$ , and, within Gibbs sampling, the values  $\delta_t$  are sampled from

a Bernoulli distribution with parameter given by (12). In the absence of any information on the underlying process or on the occurrence (time and size) of outliers, the prior distributions must be weakly informative and the hyper-parameters chosen to meet this requirement. In the examples considered here, the hyper-parameters for the gamma and beta priors for parameters of the Poisson and negative binomial marginal distributions are  $c = d = 0.1$  and  $a = b = 0.01$ , respectively. The prior for  $\alpha$  is also a Beta(0.01, 0.01). These are the weakly informative priors for the PoINAR(1) used in Silva et al. (2005). The probabilities  $p_t$  are assigned priors Beta(5, 95) (mean 0.05) to reflect that outliers occur infrequently. Finally, the priors for the outlier sizes,  $\eta_t$ , are Gamma(10, 1) (mean 10).

Our BAIOD procedure is implemented in R (R Core Team, 2014) and, in particular, the package HI (Petris and original C code for ARMS by Wally Gilks, 2013) is used for the ARMS algorithm. The algorithm is iterated 22000 times, the 2000 initial burn-in iterations were discarded and only every 40th value of the last iterations is kept to reduce the autocorrelation within the chain. Nevertheless, the convergence of the MCMC algorithm was duly analysed with the usual diagnostic tests available in Plummer et al. (2006). The initial estimates are obtained by conditional least squares and the model parameters are estimated by posterior means or medians.

## 4.2 Simulation

The first example is a time series of length  $n = 120$  simulated from the PoINAR(1) model with parameters  $\lambda = \mu(1 - \alpha) = 1$ ,  $\alpha = 0.85$ . There are five outliers all of the same size  $\eta$  at times  $\tau = 7, 26, 60, 90, 91$ . As is well known, it is expected that detection of the outlier patch at  $t = 90, 91$  will prove problematic. We consider two different values for  $\eta = 6, 9$ . The results are presented in Figures 1 and 2 respectively. Panels (a) plot the data while Panels (b) give the posterior probabilities  $p_t$ .

Figure 1 gives the results for  $\eta = 9$ . The point estimates (standard error) for latent PoINAR(1) model are  $\hat{\lambda} = \widehat{\mu(1 - \alpha)} = 0.94$  (0.0093) and  $\hat{\alpha} = 0.87$  (0.0653) and these exhibit some bias. This is in accordance with previous results see Silva et al. (2005). The data plot in Panel (a) shows that the outliers are visually identifiable but nevertheless it is reassuring that the BAIOD plot in Panel (b) clearly identifies the outliers including the patch of two outliers that occurs at times  $t = 90, 91$ . The estimated *sizes of the outliers* themselves are  $\hat{\eta}_\tau = 7, 8, 8, 4, 7$  which are reasonably close to the true value of 9 except for  $t = 90$ .

Applying the function *tsoutliers* (1) to this data, it identifies correctly the outliers, including the patch, and it produces quite good (real valued) estimates for the sizes: 7.0, 4.0, 9.5, 8.3, 8.7. However function *tso* (2), which uses ARMA residuals, finds only 3 outliers and does not find the patch.

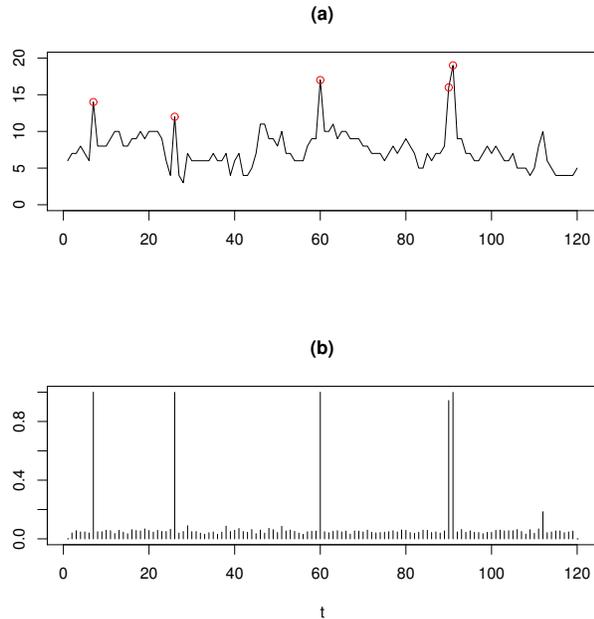


Figure 1: (a) Simulated PoINAR(1) time series with  $\alpha = 0.85$ ,  $\lambda = \mu(1 - \alpha) = 1$ , outliers at times  $t = 7, 26, 60, 90, 91$ , all with same size  $\eta_t = 9$ ; (b) posterior probability of outlier occurrence at each time point.

Figure 2 covers the case when  $\eta = 6$ . The data plot shows that the outliers are less easily visually identified and this is reflected in the somewhat lower  $p_t$  probabilities in Panel (b). In particular, the outlier at  $t = 90$ , part of the patch, is assigned very low probability. Nevertheless the remaining outliers are clearly identified by the  $p_t$  plot. The estimated *sizes of the outliers* themselves are  $\hat{\eta}_\tau = 4, 5, 5, 1, 5$ .

When  $\eta = 6$ , *tsoutliers* (1) finds only 1 outlier at time  $t = 91$  while *tso* (2) finds 4 outliers but does not identify the patch. Over the two scenarios, the new BAIOD method would appear to be the procedure of choice.

We also conducted some experiments to assess the effect of the prior specification of the contaminated component of the model. Figure 3 gives the results for the previous PoINAR(1) model when  $\eta = 9$  but the prior for the size of the outlier (only) was changed from Gamma(10, 1) (mean 10) to Gamma(0.01, 0.01) (mean 1). The only significant difference is that the first observation in the outlier patch at  $t = 90, 91$  is no longer identified as an outlier.

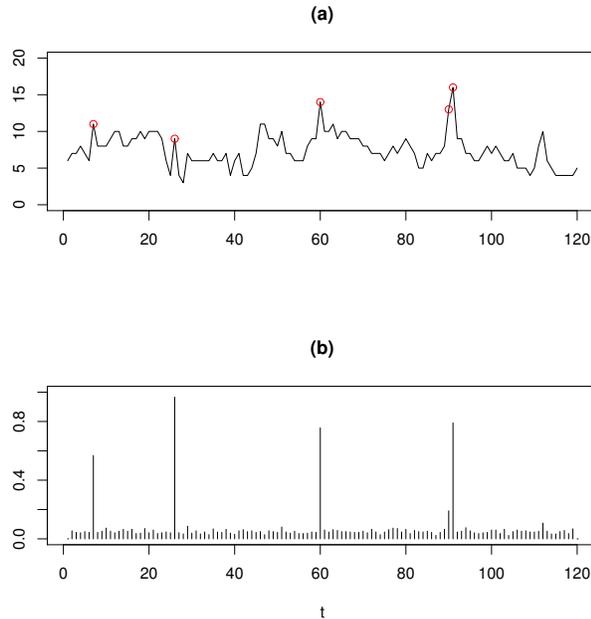


Figure 2: (a) Simulated PoINAR(1) time series with  $\alpha = 0.85$ ,  $\lambda = \mu(1 - \alpha) = 1$ , outliers at times  $t = 7, 26, 60, 90, 91$ , all with same size  $\eta = 6$ ; (b) posterior probability of outlier occurrence at each time point.

Figure 4 illustrates the effect of changing the prior on  $p_t$  (only) from an informative Beta(5, 95) (mean 0.05) to the Bayes-Laplace non informative prior Beta(1, 1) (mean 0.5) thus significantly increasing the prior probability of observations being classed as outliers. In this case, Panel (b) assigns a large proportion of the observations a probability of being an outlier in excess of 0.5, indicating the responsiveness of the analysis to the presence of valuable prior information in this dimension.

### 4.3 Empirical Examples

This section looks at two empirical examples and checks for the presence of outliers. The first example is the U.S. polio incidence counts, henceforth *polio data*, which is one of the most famous time series of counts, introduced by Zeger (1988) and which serves as a benchmark data set in the literature. The second is a time series of monthly counts of claimants collecting wage benefits from the Workers Compensation Board of British Columbia, Canada.

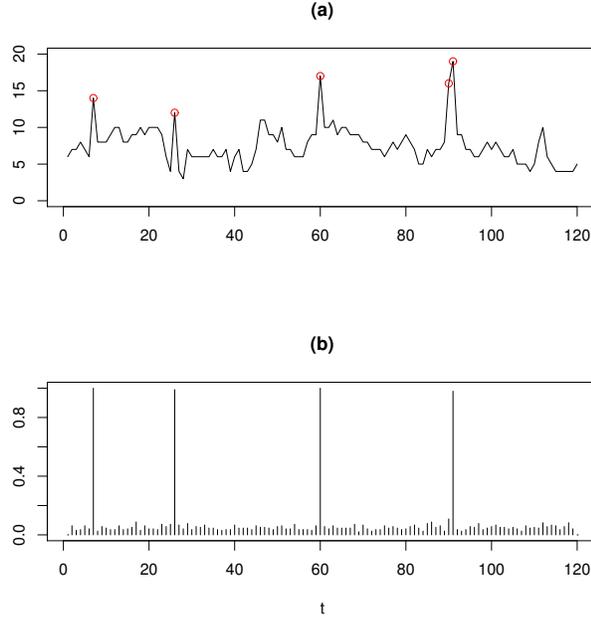


Figure 3: (a) Simulated PoINAR(1) time series with  $\alpha = 0.85$ ,  $\lambda = \mu(1 - \alpha) = 1$ , outliers at times  $t = 7, 26, 60, 90, 91$ , all with same size  $\eta = 9$ ; (b) posterior probability of outlier occurrence at each time point when non informative prior for the mean size of outliers is used.

#### 4.3.1 Polio Data

The polio data set consists of monthly observations from January 1970 to December 1983, a total of 168 data points. Figure 5 (a) plots the data. The counts range from 0 to 14 with a sample mean of  $\bar{x} = 1.33$  and a sample variance of  $s^2 = 3.5$ . The variance mean ratio is 2.63, indicating overdispersion relative to Poisson. The data presents an unusually large count of 14 in October 1972 which has been treated as an outlier by several authors. Here we use a NBINAR(1) with additive outliers to model the data. The priors for the negative binomial (NB) parameters of the marginal distributions are beta with  $a = b = 0.01$  for  $\xi$  and  $\alpha$  and gamma with  $c = d = 0.1$  for  $\mu$  leading to weakly informative priors.

The estimated parameters of the model are  $\hat{\mu} = 1.33$ ,  $\hat{\xi} = 0.503$  and  $\hat{\alpha} = 0.106$ . The priors for the probability of outlier occurrence  $p_t$  and the mean size of the outliers  $\beta_t$  are beta with  $g = 5$  and  $h = 95$  (mean 0.05) and gamma with  $l = 10$  and  $m = 1$  (mean 10). The posterior probabilities of outlier occurrence, Figure 5 (b), indicate that the observation in October 1972 is an outlier with probability  $\hat{p}_{34} = 0.56$ . The estimated size of the mean is 9.0 and the estimated size of the outlier is 8.

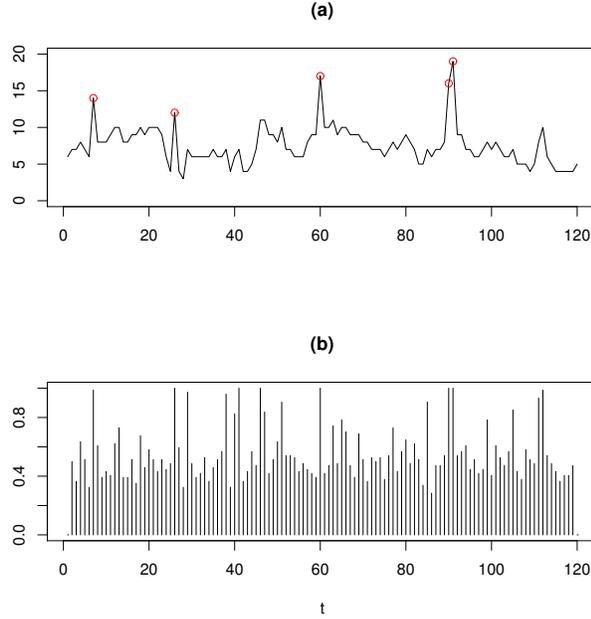


Figure 4: (a) Simulated PoINAR(1) time series with  $\alpha = 0.85$ ,  $\lambda = \mu(1-\alpha) = 1$ , outliers at times  $t = 7, 26, 60, 90, 91$ , all with same size  $\eta = 9$ ; (b) posterior probability of outlier occurrence at each time point when a Beta (1,1) prior for the probability of outliers is used.

By contrast, function *tsoutliers* (1) indicates that this data set presents outliers at times June 1970, September 1972, October 1972, April 1979 and May 1979. Also function *tso* (2) suggests additive outliers at times June 1970, October 1972, April 1979 and November 1983.

To assess the adequacy of the underlying negative binomial model we analyse the Pearson (studentised) residuals after having removed the outlier. These residuals and the acf are presented in Figure 6 which indicates that there is still some degree of correlation remaining but it is not substantive. A probability integral transform (PIT) analysis is given in Figure 7 and it shows some slight deviation from the NB distribution. Overall though, the negative binomial model may be deemed acceptable.

#### 4.3.2 Wage Benefit Data

In this section, we investigate the existence of outliers in a time series of monthly counts of claimants collecting wage loss benefits from the Workers Compensation Board of British Columbia, Canada, for injuries due to bruises in the logging

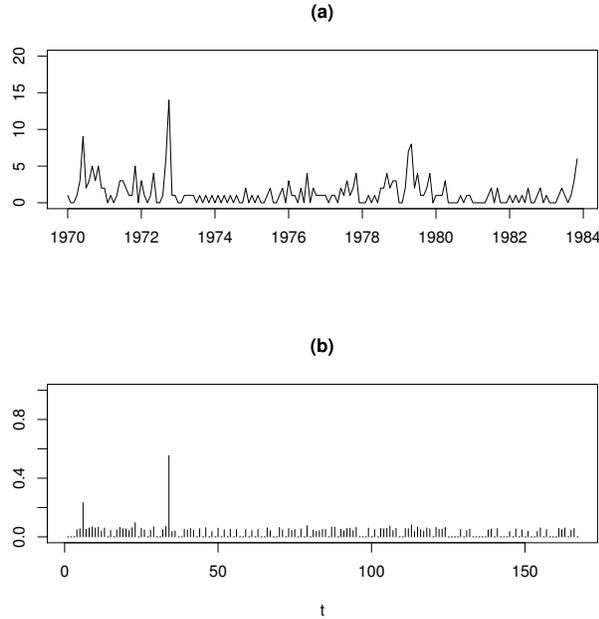


Figure 5: Incidence of U.S. Polio, Jan 1970 - Dec 1983 and posterior probabilities of outlier occurrence at each time point.

industry from January 1985 to December 1994. The bruise counts, represented in figure 8(a), range from 4 to 17, with a mean of 9.83 and a sample variance of 9.56. A PoINAR(1) model with additive outliers is fitted to the data. The estimated model with parameters  $\hat{\mu} = 9.61$ , and  $\hat{\alpha} = 0.49$  agrees well with the empirical mean and variance and first order autocorrelation. Since the PoINAR(1) model can be interpreted as an infinite server queue with a geometric service time and Poisson arrival process these results indicate that a newly injured worker is expected to collect benefit for 1.95 months.

The posterior probabilities of outlier occurrence, Figure 8 (b), indicate absence of outlier occurrences. In addition, neither of the two competitors to BAIOD suggest that outliers are present. As before, model fit assessed by residual analysis and both the acf (Figure 9) and the PIT diagram (Figure 10) indicate the Poisson model is adequate for the data.

## 5 Concluding remarks

This paper describes a Bayesian framework for the retrospective analysis of CCID(1) models for positive valued time series with multiple additive outlier effects. The Bayesian paradigm allows for the computation of the probability

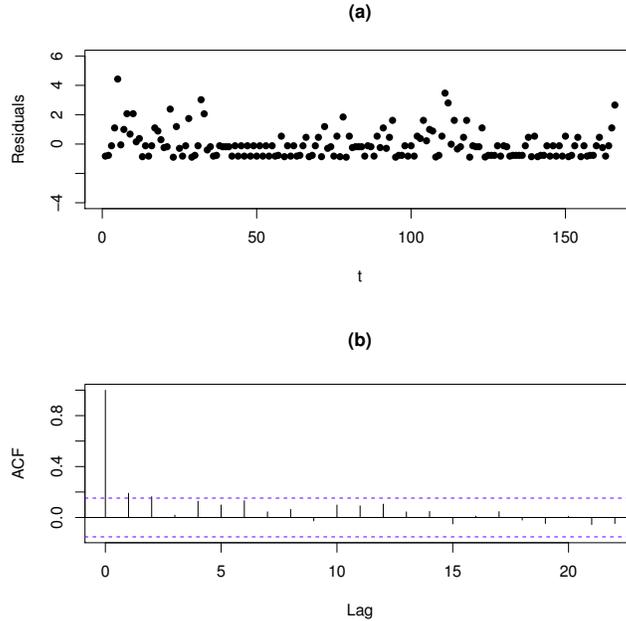


Figure 6: Pearson residuals (a) acf of residuals (b) for Incidence of U.S. Polio, Jan 1970 - Dec 1983.

that an outlier occurred at each time point and an estimate of its size. The procedure has been presented as useful for detecting suspicious observations given a certain model but it can also be used for detecting possible model deficits, for example, suggesting that distributions with fatter tails or dummy variables be employed.

Further extensions of this work are possible. The first and foremost, is the extension to higher order models i.e. models with longer lags. The necessary mathematical expressions are easily derived from the likelihood function. However, the likelihood and consequently the full conditional posterior distributions are highly complex, leading to additional computing effort. Another extension is the incorporation of other effects such as transient or permanent level shifts and covariates. Finally, possible future work might attempt to improve the component-wise Metropolis-Hastings sampling algorithm, using approaches such as Hamiltonian Monte Carlo Neal (2010) or the Wang-Landau algorithm Atchadeé and Liu (2010).

## Acknowledgements

We are grateful to the anonymous referees for helpful comments on an earlier version

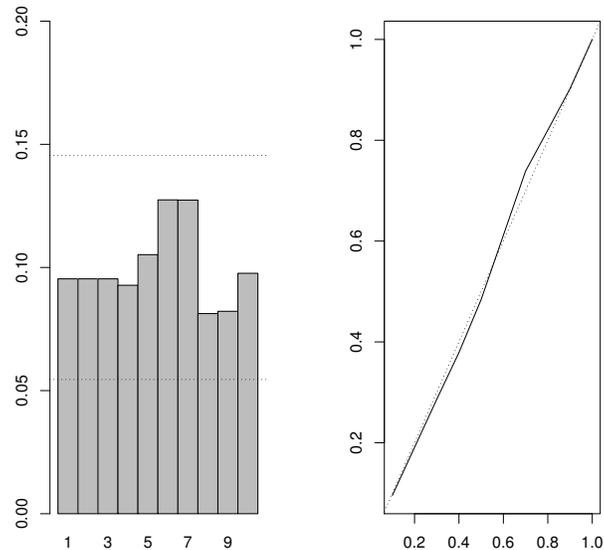


Figure 7: PIT analysis for Incidence of U.S. Polio, Jan 1970 - Dec 1983.

of the article, which have led to numerous improvements. This work is partially supported by Portuguese funds through CIDMA and the Portuguese Foundation for Science and Technology (FCT -Fundação para a Ciência e a Tecnologia), within project UID/MAT/04106/2013.

## References

- Abanto-Valle, C., Bandyopadhyay, D., Lachos, V., Enriquez, I., 2010. Robust bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics & Data Analysis* 54, 2883 – 2898.
- Al-Osh, M.A., Alzaid, A.A., 1987. First-order integer-valued autoregressive (inar(1)) process. *Journal of Time Series Analysis* 8, 261–275.
- Alzaid, A., Al-Osh, M., 1993. Some autoregressive moving average processes with generalized Poisson marginal distributions. *Annals of the Institute of Mathematical Statistics* 45, 223–232.
- Atchadeé, Y.F., Liu, J., 2010. The wang-landau algorithm in general state spaces: applications and convergence analysis. *Statistica Sinica* 20, 209–233.

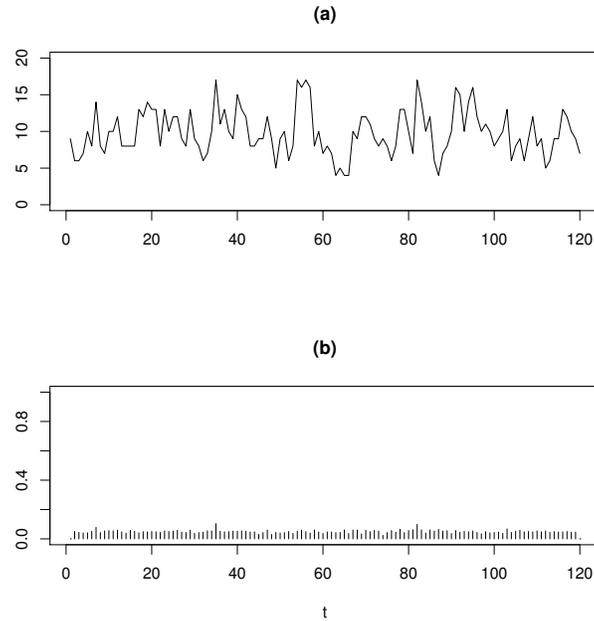


Figure 8: Monthly counts of claimants collecting wage benefits for injuries due to bruises, Jan 1985 to Dec 1994 and posterior probabilities of outlier occurrence at each time point.

- Barnett, V., Lewis, T., 1994. *Outliers in statistical data*. John Wiley & Sons, Chichester. 3 edition.
- Battaglia, F., Orfei, L., 2005. Outlier detection and estimation in nonlinear time series. *Journal of Time Series Analysis* 26, 107–121.
- Beran, J., 1994. *Statistics for Long-Memory Processes*. Chapman and Hall-CRC, Boca Raton.
- Carnero, M.A., Peña, D., Ruiz, E., 2007. Effects of outliers on the identification and estimation of garch models. *Journal of Time Series Analysis* 28, 471–497.
- Chan, W., Cheung, S., 1994. On robust estimation of threshold autoregressions. *Journal of Forecasting* 13, 37 – 49.
- Chang, I., Tiao, G.C., Chen, C., 1988. Estimation of time series parameters in the presence of outliers. *Technometrics* 30, 193–204.
- Chen, C., Liu, L.M., 1993. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association* 88, 284–297.

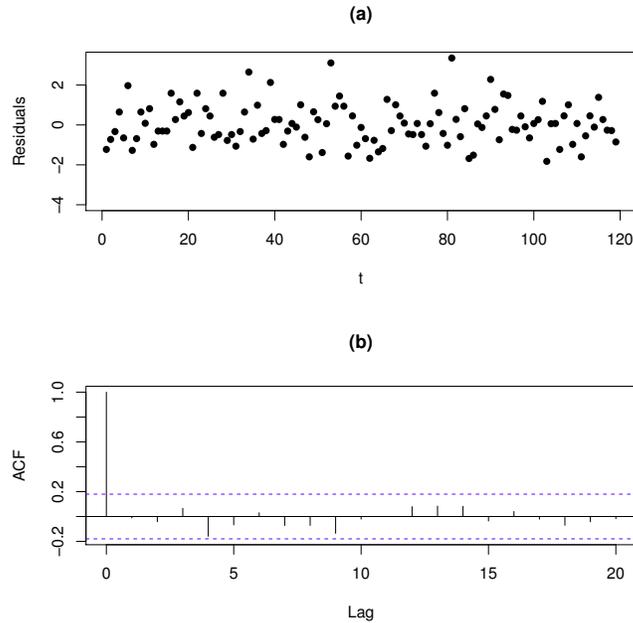


Figure 9: Residuals (a) acf of residuals (b) for bruises data.

- Chen, C.W., 1997. Detection of additive outliers in bilinear time series. *Computational Statistics and Data Analysis* 24, 283 – 294.
- Denby, L., Martin, R., 1979. Robust estimation of the first-order autoregressive parameter. *Journal of American Statistical Association* 74, 140 – 146.
- Doornik, J.A., Ooms, M., 2005. Outlier detection in garch models. Tinbergen Institute Discussion Paper, No. 05-092/4 .
- Fokianos, K., Fried, R., 2010. Interventions in ingarch processes. *Journal of Time Series Analysis* 31, 210–225.
- Fox, A.J., 1972. Outliers in time series. *Journal of the Royal Statistical Society. Series B (Methodological)* 34, 350–363.
- Fried, R., Agueusop, I., Bornkamp, B., Fokianos, K., Fruth, J., Ickstadt, K., 2015. Retrospective bayesian outlier detection in INGARCH series. *Statistics and Computing* 25, 365–374.
- Grané, A., Veiga, H., 2010. Wavelet-based detection of outliers in financial time series. *Computational Statistics & Data Analysis* 54, 2580 – 2593.

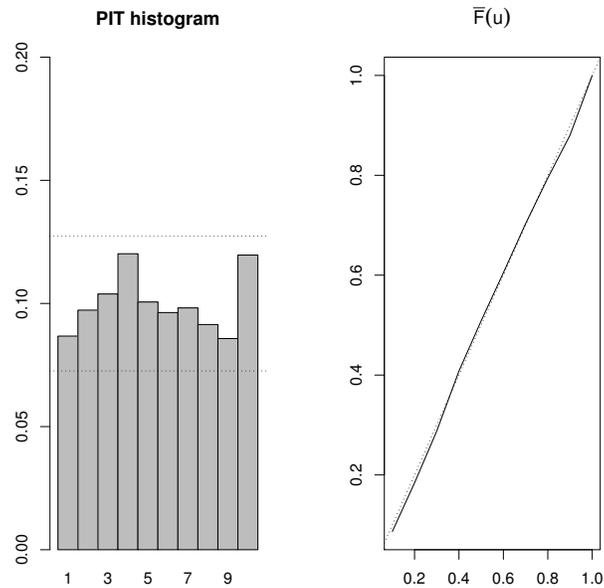


Figure 10: PIT for bruises data.

- Han, L., McCabe, B., 2013. Testing for parameter constancy in non-Gaussian time series. *Journal of Time Series Analysis* 34, 17–29.
- Hotta, L.K., Tsay, R., 2012. Outliers in garch processes, in: *Economic Time Series: Modeling and Seasonality*. Chapman & Hall CRC Press, pp. 337–358.
- Hyndman, R.J., 2017. forecast: Forecasting functions for time series and linear models. R package version 8.0.
- Hyndman, R.J., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software* 26, 1–22.
- Joe, H., 1996. Time series model with univariate margins in the convolution-closed infinitely divisible class. *Applied Probability* 33, 664–677.
- Jung, R.C., Tremayne, A.R., 2011. Useful models for time series of counts or simply wrong ones? *AStA Advances in Statistical Analysis* 95, 59–91.
- Kleiner, B., Martin, D., 1979. Robust estimation of power spectra. *Journal Royal Statistical Society B* 41, 313–351.
- de Lacalle, J.L., 2017. tsoutliers: Detection of Outliers in Time Series. R package version 0.6-6.

- Lewis, P., McKenzie, E., Hugus, D., 1989. Gamma processes. *Communications Statistics. Stochastic Models* 5, 1–30.
- Martin, R., Yohai, V., 1986. Robustness in time series and estimating arima models, in: Hannan, E., Krishnaiah, P. (Eds.), *Handbook of Statistics*. Elsevier. volume 5, pp. 119 – 155.
- McKenzie, E., 1986. Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability* 18, 679–705.
- McKenzie, E., 1988. Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability* 20, 822–835.
- Neal, R.M., 2010. MCMC using Hamiltonian dynamics, in: Brooks, S., Gelman, A., Jones, G.L., Meng, X.L. (Eds.), *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Press. volume 30, pp. 113–162.
- Petris, G., original C code for ARMS by Wally Gilks, L.T., 2013. HI: Simulation from distributions supported by nested hyperplanes. R package version 0.4.
- Plummer, M., Best, N., Cowles, K., Vines, K., 2006. Coda: Convergence diagnosis and output analysis for MCMC. *R News* 6, 7–11.
- R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.
- Silva, I., Silva, M.E., Pereira, I., Silva, N., 2005. Replicated INAR(1) processes. *Methodology and Computing in Applied Probability* 7, 517–542.
- Silva, M.E., Pereira, I., 2015. Detection of Additive Outliers in Poison INAR(1) Time Series, in: Bourguignon, J.P., Pinto, A., Viana, M. (Eds.), *Mathematics of Energy and Climate Change, International Conference and Advanced School Planet Earth, Portugal, March 21-28, 2013*. Springer. volume 2 of *CIM Series in Mathematical Sciences*, pp. 377–388.
- Tsay, R.S., 1986. Time series model specification in the presence of outliers. *Journal of the American Statistical Association* 81, 132–141.
- Wang, J.J., Chan, J.S., Choy, S.B., 2011. Stochastic volatility models with leverage and heavy-tailed distributions: A bayesian approach using scale mixtures. *Computational Statistics & Data Analysis* 55, 852 – 862.
- Zeger, S.L., 1988. A regression model for time series of counts. *Biometrika* 75, 621–629.

## Appendix

This section contains the calculations needed to perform Bayesian outlier detection when we consider the negative binomial INAR(1) model with the beta-binomial thinning operator (see Table 1 ).

-Joint prior:

$$\pi(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}) \propto e^{-d\mu} \mu^{c-1} \alpha^{a-1} (1-\alpha)^{b-1} \xi^{e-1} (1-\xi)^{f-1} e^{-(m+1)\sum_{t=1}^n \beta_t} \prod_{t=1}^n p_t^{\delta_t+g-1} (1-p_t)^{h-\delta_t} \beta_t^{\eta_t+l-1} / \eta_t! \quad (19)$$

where  $\boldsymbol{\theta} = (\mu, \alpha, \xi)$ .

-Conditional likelihood:

$$L(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}) = \prod_{t=2}^n \sum_{i=0}^{M_t} \frac{\Gamma(\mu(1-\alpha) + Y_t - \delta_t \eta_t - i)}{(Y_t - \delta_t \eta_t - i)! \Gamma(\mu(1-\alpha))} (1-\xi)^{Y_t - \delta_t \eta_t - i} \xi^{\mu(1-\alpha)} \binom{Y_{t-1} - \delta_{t-1} \eta_{t-1}}{i} \frac{B(\alpha\mu + i, (1-\alpha)\mu + Y_{t-1} - \delta_{t-1} \eta_{t-1} - i)}{B(\alpha\mu, (1-\alpha)\mu)} \quad (20)$$

where  $M_t = \min(Y_{t-1} - \delta_{t-1} \eta_{t-1}, Y_t - \delta_t \eta_t)$ .

-Full conditional distributions:

$$\pi(\mu | \alpha, \xi, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}, \mathbf{y}) \propto \frac{\mu^{c-1} e^{-d\mu} \xi^{\mu(1-\alpha)(n-1)}}{[\Gamma(\mu(1-\alpha)) B(\mu\alpha, (1-\alpha)\mu)]^{n-1}} \prod_{t=2}^n \sum_{i=0}^{M_t} \Gamma(\mu(1-\alpha) + Y_t - \delta_t \eta_t - i) B(\alpha\mu + i, (1-\alpha)\mu + Y_{t-1} - \delta_{t-1} \eta_{t-1} - i) K(t, i) \quad (21)$$

$$\text{with } K(t, i) = \frac{(1-\xi)^{Y_t - \delta_t \eta_t - i}}{(Y_t - \delta_t \eta_t - i)!} \binom{Y_{t-1} - \delta_{t-1} \eta_{t-1}}{i}$$

$$\pi(\alpha | \mu, \xi, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{p}, \boldsymbol{\beta}, \mathbf{y}) \propto \frac{\alpha^{a-1} (1-\alpha)^{b-1} \xi^{\mu(1-\alpha)(n-1)}}{[\Gamma(\mu(1-\alpha)) B(\mu\alpha, (1-\alpha)\mu)]^{n-1}} \prod_{t=2}^n \sum_{i=0}^{M_t} \Gamma(\mu(1-\alpha) + Y_t - \delta_t \eta_t - i) B(\alpha\mu + i, (1-\alpha)\mu + Y_{t-1} - \delta_{t-1} \eta_{t-1} - i) K(t, i) \quad (22)$$

$$\pi(\xi|\mu, \alpha, \boldsymbol{\delta}, \boldsymbol{\eta}, \boldsymbol{p}, \boldsymbol{\beta}, \mathbf{y}) \propto (1 - \xi)^{f-1} \xi^{\mu(1-\alpha)(n-1)+e-1} \prod_{t=2}^n \sum_{i=0}^{M_t} (1 - \xi)^{Y_t - \eta_t \delta_t - i} T(t, i) \quad (23)$$

$$\text{with } T(t, i) = \frac{\Gamma(\mu(1-\alpha) + Y_t - \delta_t \eta_t - i) B(\alpha\mu + i, (1-\alpha)\mu + Y_{t-1} - \delta_{t-1} \eta_{t-1} - i)}{(Y_t - \eta_t \delta_t - i)!} \binom{Y_{t-1} - \eta_{t-1} \delta_{t-1}}{i}$$

The full conditional distribution for  $\delta_t$ , depends on the expression of  $f(Y_t|Y_{t-1}, \delta_t = u, \delta_{t-1} = v)$  which changes to:

$$\begin{aligned} f(Y_t|Y_{t-1}, \delta_t = u, \delta_{t-1} = v) &= \frac{\xi^{\mu(1-\alpha)}}{B(\mu\alpha, (1-\alpha)\mu)\Gamma(\mu(1-\alpha))} \\ &\sum_{i=0}^{M_t^*} \frac{\Gamma(\mu(1-\alpha) + Y_t^* - i)}{(Y_t^* - i)!} \binom{Y_{t-1}^*}{i} (1 - \xi)^{Y_{t-1}^* - i} B(\alpha\mu + i, (1-\alpha) + Y_{t-1}^* - i) \end{aligned} \quad (24)$$

with  $u, v = 0, 1$ ,  $t = j, j + 1$ ,  $M_t^* = \min(Y_{t-1}^*, Y_t^*)$ .

Full conditional distributions for  $p_t$ ,  $\eta_t$  and  $\beta_t$  remain the same as in PoINAR(1) context.