SFB 823

Bayesian outlier detection in INGARCH time series

Discussion Paper

Roland Fried, Inoncent Agueusop,
Björn Bornkamp, Konstantinos Fokianos,
Jana Fruth, Katja Ickstadt

Nr. 26/2012

SFB 823

Bayesian outlier detection in INGARCH time series

Roland Fried[a], Inoncent Agueusop[a], Björn Bornkamp[b], Konstantinos Fokianos[c], Jana Fruth[a] & Katja Ickstadt[a]

[a] *TU Dortmund University, Germany,* [b] *Novartis, Basel, Switzerland,* [c] *University of Cyprus, Cyprus*

Abstract: INGARCH models for time series of counts arising, e.g., in epidemiology assume the observations to be Poisson distributed conditionally on the past, with the conditional mean being an affine-linear function of the previous observations and the previous conditional means. We model outliers within such processes, assuming that we observe a contaminated process with additive Poisson distributed contamination, affecting each observation with a small probability. Our particular concern are additive outliers, which do not enter the dynamics of the process and can represent measurement artifacts and other singular events influencing a single observation. Such outliers are difficult to handle within a non-Bayesian framework since the uncontaminated values entering the dynamics of the process at contaminated time points are unobserved. We propose a Bayesian approach to outlier modeling in INGARCH processes, approximating the posterior distribution of the model parameters by application of a componentwise Metropolis-Hastings algorithm. Analyzing real and simulated data sets, we find Bayesian outlier detection with non-informative priors to work well if there are some outliers in the data.

Keywords: Generalized Linear Models; Time Series of Counts; Additive Outliers; Level Shift.

1 Introduction

Models for outliers and intervention effects in Gaussian time series and methods for their detection are well established nowadays. Dependencies within Gaussian time series are commonly modeled by assuming a simple parametric form of the conditional mean of the current observation Y_t given its past $Y_{t-1}, Y_{t-2}, \dots$ and adding a random error e_t representing an innovation at time t which cannot be predicted from the past; compare the popular autoregressive moving average (ARMA) models. Outlier effects are usually added directly to the observations within this framework, see Charles and Darne (2005), for instance. Fox (1972) distinguishes between innovative outliers and additive outliers in such time series models: the former correspond to outlying innovations e_t and describe, e.g., major technological advances influencing the whole process according to its dynamics. The latter represent single outlying observations which may be caused, e.g., by measurement artifacts.

Little work has been done so far on the analysis of outliers in time series of counts arising in epidemiology, insurance industry, economics and communications, among others. Following similar lines as Chen and Liu (1993) in the Gaussian framework, Fokianos and Fried (2010) propose an iterative procedure for outlier detection and correction in so called integer-valued GARCH (INGARCH) models, which have been developed and investigated by Ferland, Latour and Oraichi (2006) and Fokianos, Rahbek and Tjostheim (2009), among others. These models are developed within the framework of generalized linear models. Denoting the whole information up to time $t-1$ by $\mathcal{F}_{t-1}^{Y}$, dependencies between subsequent observations are incorporated by expressing the conditional mean $\lambda_t = \mathrm{E}(Y_t|\mathcal{F}_{t-1}^{Y})$ of the current observation Y_t in terms of past observations and past conditional means, using the identity link. More precisely, an INGARCH process $\{Y_t\}$ of order (p, q), abbreviated by INGARCH(p, q), is defined by the relationships

$$
\begin{aligned}
Y_t | \mathcal{F}_{t-1}^{Y} &\sim \text{Poisson}(\lambda_t), \quad t \geq 1, \\
\lambda_t &= \beta_0 + \sum_{i=1}^{q} \beta_i Y_{t-i} + \sum_{j=1}^{p} \alpha_j \lambda_{t-j} \ .
\end{aligned} \tag{1}
$$

Here, $\mathcal{F}_{t-1}^{Y}$ is the σ–field generated by $\{Y_{1-q}, \ldots, Y_{t-1}, \lambda_{1-p}, \ldots, \lambda_0\}$, $\beta_0 > 0$ is an intercept and $\beta_i > 0$, $i = 1, \ldots, q$, and $\alpha_j > 0$, $j = 1, \ldots, p$, are regression coefficients. A stationary solution of (1) with mean $\beta_0/(1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j)$ exists if $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. Similar models, in which $\{\lambda_t\}$ is regressed on past values of the observed process and past values of $\{\lambda_t\}$ itself, have been studied before by Rydberg and Shephard (2000), Streett (2000) and Heinen (2003). These models for count time series include a feedback mechanism to achieve parsimony–an idea similar to the GARCH model (Bollerslev, 1986). In addition, stationarity and geometric ergodicity are guaranteed by simple conditions on the parameters (Doukhan, Fokianos and Tjostheim, 2012).

Fokianos and Fried (2010) model different types of outliers and intervention effects through the conditional mean. Parameter estimation is accomplished by maximization of the conditional likelihood. However, purely additive outliers describing measurement artifacts or other events not entering the dynamics cannot be treated easily within this frequentist framework. The reason is that the conditional means given the past of the process, which are needed for maximum likelihood estimation and other purposes, depend on the unobserved uncontaminated observations. Even the spiky outliers considered by Fokianos and Fried (2010) affect the future of the process via the evolution of the conditional mean, similar to the innovation outliers in Gaussian ARMA models. We will show that additive outliers are straightforward to deal with within a Bayesian framework. A Bayesian approach to the detection of multiple outliers in ARMA models has been suggested by Justel, Peña and Tsay (2001), while Abanto-Valle *et al.* (2010) and

Wang, Chan and Choy (2011) develop Bayesian methods for treating outliers in stochastic volatility models. Silva, Frias and Pereira (2012) treat additive outliers in integer autoregressive models for time series of counts, which are constructed by means of binomial thinning and are very different from the generalized linear modeling framework considered here.

Like Fokianos and Fried (2010) we restrict ourselves to INGARCH models, though our approach can be generalized to other link functions. Our focus is on the INGARCH(1,1) model, since it is simple, but yet sufficiently flexible for approximating many realistic dependence structures, which are observed in real count time series. Section 2 presents a Bayesian extension of the INGARCH(1,1) model and proposes a definition of additive outliers within this context. Intervention effects are added directly to the observations $\{Y_t\}$ following the linear time series methodology, so that the next observations are not contaminated. We develop Bayesian methods for the estimation of regression parameters and intervention effects using Markov Chain Monte Carlo (MCMC) techniques. Section 3 describes some examples and a simulation study for investigating the reliability of this procedure. Section 4 provides an application to real data and Section 5 draws some conclusions.

2 Bayesian outlier modeling in INGARCH series

In the following we develop Bayesian approaches for dealing with outliers in INGARCH processes, starting from the uncontaminated process without outliers.

2.1 Bayesian modeling of INGARCH processes

To implement the Bayesian version of the INGARCH(1,1) model (1) we use the software OpenBUGS (Bayesian inference Using Gibbs Sampling, Lunn et al., 2009), and the R2WinBUGS interface (Sturtz, Ligges and Gelman, 2005). In the absence of prior information we use weakly-informative prior distributions for the parameters $\tilde{\theta} = (\beta_0, \beta_1, \alpha_1, \lambda_0, Y_0)$, with β_0, (β_1, α_1) and (λ_0, Y_0) assumed to be independent random variables. For the positive parameters β_0 and λ_0, which are the regression intercept and the value initializing the dynamics of the INGARCH process, independent gamma priors are used with parameters (0.1,0.1), which are traditional weakly informative prior distributions in the Poisson model. The unobserved observation Y_0 also needed for initialization is drawn from a Poisson distribution with mean λ_0. To guarantee stationarity of the resulting process the dependence parameters α_1 and β_1 need to be positive with $\alpha_1 + \beta_1 < 1$. A prior distribution suited for encoding this constraint is the two-dimensional Dirichlet distribution for (α_1, β_1). The three components of a vector simulated from this distribution are positive and sum to one, so that the two first components sum up to a random value between 0 and 1, while the last one is determined by the first two components. The parameters of the Dirichlet are chosen as (1,1,1), corresponding to the uniform distribution on the 2-dimensional probability simplex, and implying Beta(1,2) marginal distributions for α_1 and β_1. To obtain a sample from the posterior distribution of $\tilde{\theta}$, the Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953, Hastings, 1970) is used.

2.2 Modeling of additive outliers

Additive outliers correspond to single contaminated observations, so that the outlier effect does not carry over to the subsequent observations. This does not apply if an outlier changes the conditional mean λ_t of the process, as it enters the dynamics (Fokianos and Fried, 2010). Additive outliers can be modeled assuming that we observe a contaminated process $\{Z_t\}$ instead of the clean INGARCH process $\{Y_t\}$. For this, we complement model (1) by an observation equation, contaminating each clean value Y_t with probability π_t by an additive outlier of random size X_t following a Poisson distribution,

$$\begin{aligned} Z_t &= Y_t + \delta_t X_t \qquad (2) \\ \text{with } X_t &\sim Pois(\omega) \quad \text{and} \quad \delta_t \sim Bern(\pi_t). \end{aligned}$$

We assume $X_1, \delta_1, \ldots, X_n, \delta_n$ to be independent and independent from the latent process $\{Y_t\}$ of clean data. If we have prior information about the distribution of outliers, all probabilities π_t can be chosen equal to a deterministic constant, $\pi_t \equiv \pi$, such that the product $n\pi$ is the expected numbers of outliers in the data. The parameter ω must be strictly positive.

Model (2) is different from the model studied by Fokianos and Fried (2010) since additive outliers do not contaminate the observations thereafter, *i.e.* the conditional means λ_t are unaffected. It is difficult to analyze this model within a non-Bayesian framework since the conditional mean λ_{t+1} of an observation right after a contaminated time point depends on the uncontaminated value y_t, which is not observed. An EM algorithm (e.g., O'Hagan, Murphy and Gormley, 2012) could be used if the time of the outlier is known, but the arising computations become cumbersome if there are multiple interventions at unknown time points. We implement a Bayesian version of model (2) in R (R Development Core Team, 2009).

All model parameters can be collected in an extended vector

$$\theta = (\beta_0, \beta_1, \alpha_1, \omega, \delta_1, \cdots, \delta_n, \pi_1, \cdots, \pi_n, \lambda_0, Y_0)$$

with $(2n+6)$ components. Here, λ_0 and Y_0 are starting values for the unobserved conditional mean process $\{\lambda_t\}$ and the INGARCH process $\{Y_t\}$. In addition to the prior specifications provided in Subsection 2.1 for $(\beta_0, \beta_1, \alpha_1, \lambda_0, Y_0)$, we use independent beta priors with parameter $(1, 10)$ for the π_t and a gamma prior with parameter (0.1,0.1) for ω.

Denoting random variables by capital letters and their realizations by small letters, i.e. w is the realization of W, and using $p(w)$ as a generic symbol for the density of a random variable W, the a posteriori distribution of the parameters is obtained from Bayes theorem as $p(\theta|z) = \frac{p(z|\theta)\cdot p(\theta)}{p(z)} \propto p(z|\theta)\cdot p(\theta)$ with $p(z|\theta) = L(\theta)$ being the conditional likelihood function,

$$L(\theta) = \prod_{t=1}^{n} \frac{(\lambda_t + \delta_t\omega)^{z_t}}{z_t!} e^{-(\lambda_t+\delta_t\omega)} \ .$$

We simulate samples $\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(N)}$ from the posterior distribution using the componentwise Metropolis-Hastings algorithm as described in Ntzoufras (2009), for instance. Dividing the vector θ into subvectors of correlated elements (parameter blocking) and updating it sequentially can improve the convergence. In our case, we update the parameters α_1 and β_1 jointly, while all other blocks correspond to single components. Let θ_j denote the j-th block, θ_{-j} the vector θ excluding the block θ_j, and $\theta^{(k)}$ be the value of θ after updating all blocks of θ the k-th time.

1. Set initial values $\theta^{(0)}$ by giving values to all blocks $\theta_j^{(0)}$.
2. For $k = 1, \ldots, N$ repeat the following steps
   a) Set $\theta = \theta^{(k-1)}$
   b) For $j = 1, \ldots, (2n+5)$, generate a new candidate value $\theta_j^{'}$ for the block θ_j from a proposal distribution $q(\theta_j^{'}|\theta)$. Calculate

   $$\begin{aligned} \alpha(\theta_j^{'}, \theta_{-j}) &= \min\left(1, \frac{p(\theta_j^{'}|\theta_{-j}, z)q(\theta_j|\theta_j^{'}, \theta_{-j})}{p(\theta_j|\theta_{-j}, z)q(\theta_j^{'}|\theta_j, \theta_{-j})}\right) \\ &= \min\left(1, \frac{p(z|\theta_j^{'}, \theta_{-j})p(\theta_j^{'}, \theta_{-j})q(\theta_j|\theta_j^{'}, \theta_{-j})}{p(z|\theta_j, \theta_{-j})p(\theta_j, \theta_{-j})q(\theta_j^{'}|\theta_j, \theta_{-j})}\right) \end{aligned}$$

   Set $\theta_j = \theta_j^{'}$ with probability $\alpha(\theta_j^{'}, \theta_{-j})$.
   c) Set $\theta^{(k)} = \theta$.

For the simulations, we need the full conditional distributions $p(\theta_j|\theta_{-j}, z)$, that is the posteriori distribution of each block conditional on all the other blocks of θ and the data $z = (z_1, \ldots, z_n)$. For the parameters

π_t and δ_t we obtain analytically known full conditional distributions so that the quantity $\alpha(\theta_j^{'}, \theta_{-j})$ is equal to 1 in the Metropolis-Hastings algorithm for these particular steps. We make use of the fact that for δ_t known the distribution of Z_t is again a Poisson distribution with mean equal to the sum of the mean parameters of Y_t and $\delta_t X_t$.

2.3 Full conditional distribution of β_0

The a-posteriori distribution of β_0 given all other parameters and the data is proportional to the likelihood function times the prior distribution. Assuming β_0 to be a-priori $\Gamma(\alpha_{\beta_0}, \beta_{\beta_0})$ distributed and excluding all terms without β_0 we obtain

$$p(\beta_0|\theta_{-1}, z) \propto \prod_{t=1}^{n} [(\beta_0 + r_t)^{z_t} \cdot e^{-(\beta_0 + r_t)}] \cdot \beta_0^{\alpha_{\beta_0}-1} e^{-\beta_{\beta_0} \cdot \beta_0},$$

with $r_t = \beta_1 y_{t-1} + \alpha_1 \lambda_{t-1} + \omega \delta_t$. Since this full conditional distribution does not correspond to a known distribution, the MH-Algorithm is used for β_0 with a Gamma distribution as proposal distribution, using the slightly disturbed values from the previous iteration for the parameters. The full condition distributions for ω, λ_0 and Z_0 are simulated similarly.

2.4 Full conditional distribution of (β_1, α_1)

To ensure that $\beta_1 + \alpha_1 < 1$ we use the Dirichlet distribution with parameter $a_{\beta_1}, a_{\alpha_1}, a_{\alpha_2}$ as prior distribution for (β_1, α_1). This leads to a full conditional distribution with a density

$$p(\beta_1, \alpha_1, \alpha_2|\theta_{-2}, z) \propto \prod_{t=1}^{n} (\beta_0 + \beta_1 y_{t-1} + \alpha_1 \lambda_{t-1} + \delta_t \omega)^{z_t} \cdot e^{-(\beta_0 + \beta_1 y_{t-1} + \alpha_1 \lambda_{t-1} + \delta_t \omega)}$$
$$\cdot \beta_1^{a_{\beta_1}-1} \cdot \alpha_1^{a_{\alpha_1}-1} \cdot (1 - \alpha_1 - \beta_1)^{a_{\alpha_2}-1}$$

Again, this is a non-standard distribution, so that we use an MH-Algorithm step with a Dirichlet distribution with the parameters updated according to the values accepted in the previous iteration as proposal distribution.

2.5 Full conditional distribution of δ_t

We assume δ_t to be a-priori Bernoulli distributed with parameter π_t. Analogously to the previous subsection, the full conditional distribution can be computed excluding the terms without δ_t as follows:

$$p(\delta_t|\theta_{-(t+4)}, z) \propto (\lambda_t + \delta_t\omega)^{z_t} e^{-(\lambda_t+\delta_t\omega)} \pi_t^{\delta_t} (1 - \pi_t)^{1-\delta_t}.$$

Because δ_t only takes the values 0 and 1, we can simplify the above expression,

$$\begin{aligned} p(\delta_t = 1|\theta_{-(t+4)}) &\propto (\lambda_t + \omega)^{z_t} e^{-(\lambda_t+\omega)} \pi_t = A_t \\ p(\delta_t = 0|\theta_{-(t+4)}) &\propto \lambda_t^{z_t} e^{-\lambda_t} (1 - \pi_t) = B_t \end{aligned}$$

The probabilities need to sum to 1, implying that we can norm these expressions to become $A_t/(A_t+B_t)$ and $B_t/(A_t+B_t)$, respectively. The values δ_t are sampled directly using Gibbs Sampling.

2.6 Full conditional distribution of π_t

The hierarchical structure of model (2) ensures that each π_t is independent of all other model parameters except δ_t. Choosing the prior distribution of π_t as a Beta distribution with parameters a and b, the full conditional for π_t is given by

$$\begin{aligned} p(\pi_t|\theta_{-(4+n+t)}, z) &= p(\pi_t|\delta_t) \\ &\propto \pi_t^{\delta_t} (1 - \pi_t)^{1-\delta_t} \cdot \pi_t^{a-1} (1 - \pi_t)^{b-1} \\ \Rightarrow \pi_t|\theta_{-(4+n+t)}, z &\sim Beta(\delta_t + a, b + 1 - \delta_t) \ . \end{aligned} \quad (3)$$

Alternatively, the π_t can be set to the same constant, since the model described above is not affected strongly by this.

3 Simulation experiments

Deviating slightly from the fitted additive outlier model (2), we simulate INGARCH(1,1) time series with additive outliers at known time points.

Using fixed time points for the outliers, corresponding to fixed values of $\pi_t \in \{0, 1\}$, allows to control the number of outliers in the time series and to calculate the relative frequencies of time series generated from the same model, for which each of the outliers was detected.

We apply the procedure given above to estimate the model parameters by the posterior means. The posterior means of the outlier probabilities for each observation are approximated by calculating the mean of the vectors $\delta^{(k)}$ selected from $\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(N)}$. The average value of δ_t obtained in the iterations can be used to approximate the a-posteriori probability of Z_t being outlying. Alternatively, the mean value of π_t could also be used for this purpose, but the difference will usually be small in practice. Each observation can be classified as outlying or not choosing a cutoff, e.g., 0.5, for one of these two quantities.

In the following we first illustrate our procedure using a single artificial data example, before we study it more systematically varying the number of outliers and the parameter values.

3.1 A first example

The first example refers to an INGARCH(1,1) series of length $n = 150$ with parameters $(\beta_0, \beta_1, \alpha_1) = (2, 0.3, 0.4)$ and eight additive outliers at times 15, 48, 83, 101, 126, 136, 137 and 138. The outlier sizes are generated from the Poisson distribution with mean $\omega = 15$, see Figure 1. The posterior means estimate the true parameters rather well, see Table 1.

Table 2 illustrates that the Bayesian fitting procedure does not lead to any false detection in this data set and identifies seven of the eight outliers correctly since the a-posteriori probabilities of outlyingness are close to one. The only exception is the rather small outlier at time 138. This can happen since the Poisson distribution used for the contamination can take on small values occasionally. Additionally, there can be masking effects because of the two outliers occurring immediately

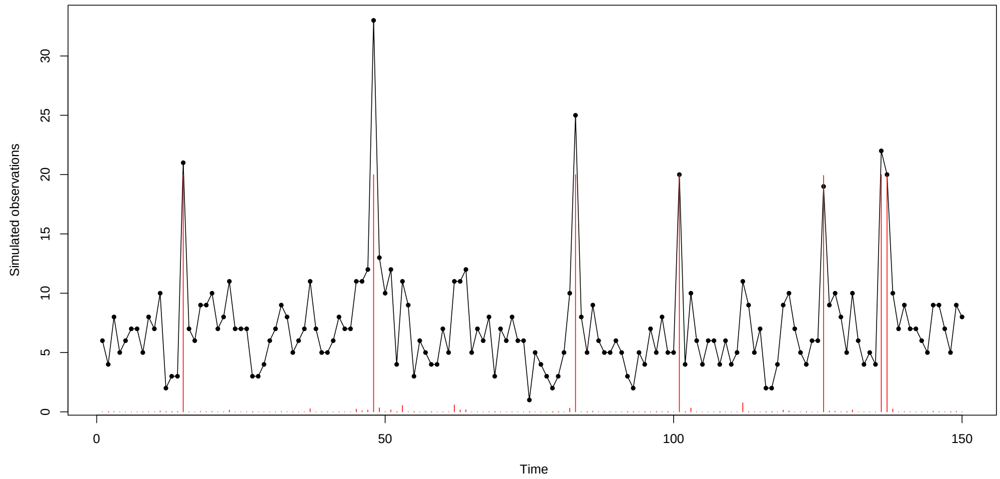

Figure 1: First example: Simulated INGARCH(1,1) time series with eight additive outliers and a-posteriori probabilities of outlyingness, multiplied by 20 for the reason of illustration.

Table 1: First example: True parameters and parameter estimates obtained as the posterior means with their standard deviations.

Parameter	β_0	β_1	α_1	ω
True value	2	0.3	0.4	15
Estimate	2.810	0.357	0.213	15.830
Standard deviation	0.801	0.083	0.127	1.965

before it, if these are not detected or if their size is underestimated.

In the next subsection we study the performance of the Bayesian procedure more carefully in a simulation study and compare it to ordinary conditional likelihood estimation.

3.2 Simulation study

We start our simulation study by comparing the classical conditional maximum likelihood (CML) and the Bayesian estimators described above in the absence of outliers. The biases, standard deviations and root of the mean square errors of the estimators are approximated by

Table 2: First example: Times of the true and the detected outliers along with the a-posteriori probabilities of outlyingness.

True outliers at	15	48	83	101	126	136	137	138
Observed value	21	33	25	20	19	22	20	10
Detected outliers	15	48	83	101	126	136	137	—
posterior probabilities	0.999	1	0.994	0.999	0.996	0.996	0.994	0.010

analyzing 10000 time series by conditional maximum likelihood and 100 time series by the Bayesian procedure for each of several models, with the length of the time series being $n = 150$. The Bayesian procedure is applied to fewer time series because of its larger computation times.

In case of clean data, both estimators are somewhat positively biased for β_0 and slightly negatively biased for α_1, and stronger so if α_1 is large, see Table 3. While the biases of the estimators are of similar size, the Bayesian procedure leads to smaller MSEs because of smaller variabilities. Imposing a prior distribution apparently stabilizes the estimates.

Table 4 shows the results for INGARCH(1,1) time series with three or six positive additive outliers. The conditional maximum likelihood estimator becomes positively biased for β_0 and negatively biased for the dependence parameters β_1 and α_1. The Bayesian procedure performs considerably better in this respect and shows much smaller biases, except if $\alpha_1 + \beta_1$ is large, i.e., as the model approaches non-stationarity. Outlier detection is very difficult in almost non-stationary time series, as becomes manifest in the strong underestimation of the mean ω of the outliers in this situation. Incorporating the possibility of additive outliers in the analysis is helpful in the other cases. Since additionally the standard deviation of the Bayesian estimators is again smaller than that of the CML estimators, the Bayesian approach leads to much smaller MSEs. The estimators of $\beta_0, \beta_1, \alpha_1$ perform better if there are only a few outliers, since fewer outliers affect the estimators less. How-

Table 3: Results for clean data: Bias, standard deviation and root of the mean square error of the estimators based on the Bayesian procedure with additive outliers and of the conditional maximum likelihood estimators, in case of time series generated from different INGARCH(1,1) models without outliers.

	Bayesian estimator					CML-estimator		
	β_0	β_1	α_1	ω	no. of outliers	β_0	β_1	α_1
true	*1*	*0.2*	*0.3*	*0*	-	*1*	*0.2*	*0.3*
bias	0.068	-0.013	-0.030	0.217		0.096	-0.008	-0.041
std.dev.	0.191	0.069	0.091	0.174		0.463	0.081	0.246
$\sqrt{\text{MSE}}$	0.203	0.070	0.096	0.278		0.473	0.081	0.250
true	*1*	*0.5*	*0.3*	*0*	-	*1*	*0.5*	*0.3*
bias	0.221	0.011	-0.060	0.400		0.179	-0.010	-0.028
std.dev.	0.369	0.068	0.091	0.356		0.439	0.083	0.126
$\sqrt{\text{MSE}}$	0.430	0.069	0.109	0.536		0.474	0.084	0.129
true	*2.5*	*0.2*	*0.3*	*0*	-	*2.5*	*0.2*	*0.3*
bias	0.035	-0.005	-0.007	0.362		0.231	-0.008	-0.039
std.dev.	0.425	0.072	0.098	0.238		1.152	0.081	0.246
$\sqrt{\text{MSE}}$	0.427	0.072	0.099	0.434		1.175	0.081	0.249

ever, the Bayesian estimator of ω gets less biased as the number of outliers increases, because more information on the outlier mechanisms becomes available. Hence, non-informative Bayesian modelling of additive outliers becomes more effective when several outliers exist in the data, and this is when protection from outliers is more important.

The results obtained by the Bayesian estimators can be improved by choosing an improper $\Gamma(2,0)$ prior for the expected outlier size ω instead of the $\Gamma(0.1,0.1)$ prior, which is a common standard for positive parameters. The underlying idea of this improper prior is that one knows a-priori that outliers are large, and that an observation is more probable to be an outlier the larger it is. So it makes sense to choose a monotone increasing function as "prior" density. The improper $\Gamma(2,0)$ prior corresponds to a linearly increasing function, $p(\theta) \propto \theta$. The results also provided in Table 4 indicate that the bias and to some

Table 4: Bias, standard deviation and root of the mean square error of the conditional maximum likelihood estimators and the estimators based on the Bayesian modelling of additive outliers with different priors for ω, in case of time series generated from different INGARCH(1,1) models with $n_o \in \{3, 6\}$ additive outliers.

	CML-estimator				$\Gamma(0.1, 0.1)$ prior for ω				$\Gamma(2, 0)$ prior for ω			
	β_0	β_1	α_1	n_o	β_0	β_1	α_1	ω	β_0	β_1	α_1	ω
true	*1*	*0.2*	*0.3*	*6*	*1*	*0.2*	*0.3*	*5*	*1*	*0.2*	*0.3*	*5*
bias	.334	-.068	-.039		.108	-.058	-.002	-2.53	.060	-.056	-.001	-1.48
std.dev.	.581	.080	.278		.225	.055	.094	1.08	.186	.058	.072	.62
$\sqrt{\text{MSE}}$	.670	.105	.281		.249	.080	.094	2.75	.196	.081	.072	1.60
true				*3*								
bias	.199	-.038	-.034		.082	-.033	-.006	-3.87	.052	-.033	-.020	-2.14
std.dev.	.523	.082	.261		.197	.072	.085	.94	.184	.064	.078	.63
$\sqrt{\text{MSE}}$	.560	.091	.263		.213	.079	.086	3.98	.191	.073	.080	2.23
true	*1*	*0.5*	*0.3*	*6*	*1*	*0.5*	*0.3*	*5*	*1*	*0.5*	*0.3*	*5*
bias	.306	-.066	.013		.286	-.057	.003	-3.53	.197	-.060	.011	-1.38
std.dev.	.515	.086	.138		.461	.074	.103	1.39	.435	.090	.104	.88
$\sqrt{\text{MSE}}$	.599	.109	.139		.542	.094	.103	3.80	.478	.108	.104	1.63
true				*3*								
bias	.239	-.039	-.007		.242	-.035	-.018	-4.36	.214	-.039	-.016	-1.66
std.dev.	.480	.085	.133		.410	.082	.099	.61	.400	.070	.095	.96
$\sqrt{\text{MSE}}$	.536	.093	.134		.476	.089	.100	4.40	.454	.080	.096	1.92
true	*2.5*	*0.2*	*0.3*	*6*	*2.5*	*0.2*	*0.3*	*5*	*2.5*	*0.2*	*0.3*	*5*
bias	.496	-.039	-.038		.062	-.031	.028	-3.98	.120	-.039	.007	-1.37
std.dev.	1.275	.079	.258		.479	.061	.106	1.01	.448	.059	.093	.94
$\sqrt{\text{MSE}}$	1.368	.088	.261		.483	.068	.110	4.11	.464	.071	.093	1.66
true				*3*								
Bias	.342	-.024	-.034		.140	-.029	.004	-4.20	-.071	-.015	.014	-1.91
std.dev.	1.214	.080	.252		.478	.063	.096	.89	.467	.067	.091	.70
$\sqrt{\text{MSE}}$	1.261	.083	.254		.498	.070	.096	4.29	.472	.069	.092	2.03

extent also the variability of the estimators of ω and β_0 are reduced with this prior choice, resulting in substantially smaller MSEs. The results for β_1 and α_1 are little affected.

Since we expect outlier sizes in Poisson data to be much larger than 1, which is the expected value of the $\Gamma(0.1, 0.1)$ prior used above, another reasonable alternative is to replace it by a more 'realistic' distribution with a larger expectation. Further simulation results based on a $\Gamma(0.25, 0.05)$ prior for ω not reported here indicate further possible substantial improvements for the bias and the MSE for the estimation of ω, for the expense of a somewhat increased bias (but not necessarily MSE) for the estimation of β_0.

4 Application to real data

Finally, we apply the iterative procedure with the $\Gamma(0.1, 0.1)$ prior for ω to the campylobacterosis data analyzed by Fokianos and Fried (2010), among others, see Figure 2. We detect additive outliers at times 100, 101, 113 and 125, see Table 5. These findings match those of Fokianos and Fried (2010) in a sense, as these authors detect a level shift at time 84 and a spiky outlier at time 100. This is an alternative explanation of the increased variability at the end of the time series. A spiky outlier with carry-over effect as considered by Fokianos and Fried (2010) can well lead to two subsequent deviating observations, which are then detected as additive outliers by the approach taken here. Moreover, an increase of the Poisson parameter (corresponding to a simultaneous shift of level and scale) will lead to several observations which look like additive outliers. Note, however, the substantial differences between the parameter estimates reported in Table 6: Similar to the case of linear Gaussian time series, additive outliers not entering the dynamics strongly affect classical estimates of dependence parameters, as could already be seen in Table 4. The correction estimators based on data

Table 5: Additive outliers detected in the campylobacterosis data and corresponding a-posteriori probabilities.

Outliers(Time)	100	101	113	125
Observed values	55	47	33	25
A-posteriori probabilities	1	0.998	0.964	0.928

cleaning suggested by Fokianos and Fried (2010) assume outliers to enter the dynamics and thus lead to substantially different estimates of the dependence parameters α_1 and β_1 than the Bayesian approach for additive outliers suggested here. Note that the Bayesian estimate of the marginal mean $\beta_0/(1-\beta_1-\alpha_1)$ can be interpreted as a weighted average of the cleaning estimates of the means before and after the shift of height ω, $\beta_0/(1-\beta_1-\alpha_1)$ and $(\beta_0+\omega)/(1-\beta_1-\alpha_1)$.

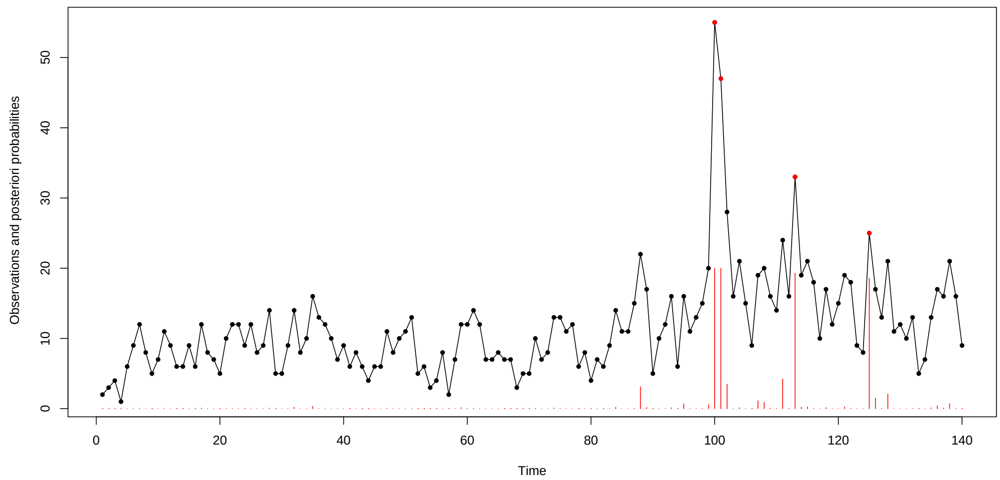

Figure 2: Campylobacterosis data and estimated outlier probabilities.

Table 6: Model parameters estimated for the campylobacterosis data by conditional likelihood, applying the classical outlier detection procedure, or the Bayesian procedure with additive outliers.

Parameter	β_0	β_1	α_1	ω
Conditional likelihood	2.439	0.196	0.591	—
Cleaning estimate	3.584	0.352	0.230	2.93 / 41.65
Bayesian estimate	1.692	0.431	0.417	21.892
Standard deviation	0.690	0.094	0.135	4.667

5 Conclusions

We have developed a Bayesian framework for the analysis of INGARCH models for time series of counts, including the treatment of outlier effects. As compared to a classical likelihood approach, the Bayesian paradigm offers increased flexibility and provides feasible strategies for including different outlier effects in this context. While Fokianos and Fried (2010) analyzed outliers which enter the dynamics of the process, we have focused on additive outliers not entering the dynamics since they can be handled more easily by our Bayesian analysis than by a likelihood approach. Having the possibility to distinguish and classify the different outlier patterns would be interesting. The flexibility of the Bayesian paradigm is promising for such a unified framework, but much more work is necessary for this.

We have seen that several outliers are needed to provide enough information for reliable estimation of the model parameters, including the expected size ω of the outliers, when using non-informative priors. Suitable informative priors for ω can provide better results if only a few outliers are present. Estimators which are able to deal with multiple outliers at unknown time points like the Bayesian ones presented here are interesting and useful since such scenarios are difficult to deal with when using classical estimators and simple iterative procedures for outlier detection and cleaning.

Acknowledgement

The financial support of the Deutsche Forschungsgemeinschaft (SFB 823, "Statistical modelling of nonlinear dynamic processes") is gratefully acknowledged.

References

Abanto-Valle, C. A., Bandyopadhyay, D., Lachos, V. H. and 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.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Charles, A. and Darne, O. (2005). Outliers and GARCH models in financial data. Economics Letters 86, 347–352.

Chen, C. and Liu, L.M. (1993). Joint estimation of model parameters and outlier effects in time series. Journal of the American Statistical Association 88, 284–297.

Doukhan, P ., Fokianos, K., and Tjostheim, D. (2012). On weak dependence conditions for Poisson autoregressions. Statistics & Probability Letters 82, 942–948.

Ferland, R., Latour, A. and Oraichi, D. (2006). Integer-valued GARCH process. Journal of Time Series Analysis 27, 923–942.

Fokianos, K. and Fried, R. (2010), Interventions in INGARCH processes. Journal of Time Series Analysis 31, 210–225.

Fokianos, K., Rahbek, A. and Tjostheim, D. (2009). Poisson autoregression. Journal of the American Statistical Association 104, 1430–1439.

Fox, A.J. (1972). Outliers in time series. Journal of the Royal Statistical Society Series B 34, 350–363.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.

Heinen, A. (2003). Modelling time series count data: an autoregressive conditional Poisson model. MPRA Paper 8113, University Library of Munich, Germany. Available at http://mpra.ub.uni-muenchen.de/8113/.

Justel., A. Peña, D. and Tsay, R.S. (2001). Detection of outlier patches in autoregressive time series. Statistica Sinica 11, 651-673.

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009). The BUGS project: Evolution, critique, and future directions, Statistics in Medicine 28, 3049–3067.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equation of state calculations by fast computing machines. The Journal of Chemical Physics 21, 1087–10092.

Ntzoufras, I. (2009). Bayesian modeling using WinBUGS. John Wiley & Sons Inc, New York.

O'Hagan, A., Murphy, T.B. and Gormley, I. C. (2012). Computational aspects of fitting mixture models via the expectation-maximization algorithm. Computational Statistics & Data Analysis 56, 3843–3864.

R Development Core Team (2009). R: A Language and Environment for Statistical Computing.. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rydberg, T.H. and Shephard, N. (2000). A modelling framework for the prices and times of trades made on the New York stock exchange. In: Fitzgerald, W.J., Smith, R.L., Walden, A.T., Young, P.C. (Eds.), Nonlinear and Nonstationary Signal Processing. Cambridge: Isaac Newton Institute and Cambridge University Press, pp. 217–46.

Silva, M.E., Frias, R. and Pereira, I. (2012). Detection of additive outliers in Poisson INteger-valued AutoRegressive time series. Available at http://arxiv.org/pdf/1204.6516.pdf

Streett, S. (2000). Some observation driven models for time series of counts. PhD-thesis, Colorado State University, Department of Statistics.

Sturtz, S., Ligges, U. and Gelman, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R. Journal of Statistical Software 12, 1–16.

Wang, J. J., Chan J. S. K. and Choy S. T. B. (2011). Stochastic volatility models with leverage and heavy-tailed distributions: A Bayesian approach using scale mixtures. Computational Statistics & Data Analysis 55, 852–862.