ON OUTLIERS AND INTERVENTIONS IN COUNT TIME SERIES FOLLOWING GLMs

R. Fried, H. Elsaied, T. Liboschik
Department of Statistics, TU Dortmund University
44221 Dortmund, Germany
e-mail: {fried,saied,liboschik}@statistik.tu-dortmund.de
K. Fokianos, S. Kitromilidou
Department of Mathematics and Statistics, University of Cyprus
e-mail: {fokianos,kitromelidou.stella}@ucy.ac.cy

Abstract

We discuss the analysis of count time series following generalized linear models in the presence of outliers and intervention effects. Different modifications of such models are formulated which allow to incorporate, detect and to a certain degree distinguish extraordinary events (interventions) of different types in count time series retrospectively. An outlook on future extensions to the problem of online surveillance and robust parameter estimation is provided.

1 Introduction

Time series of counts are measured in various disciplines whenever a number of events is counted during certain time periods. Examples are the monthly number of car accidents in a region, the weekly number of new cases in epidemiology, the number of transactions at a stock market per minute in finance, or the number of photon arrivals per microsecond in a biological experiment. A natural modification of the popular autoregressive moving average (ARMA) models for continuous variables is based on the assumption that the observation Y_t at time t is generated by a generalized linear model (GLM) conditionally on the past, choosing an adequate distribution for count data like the Poisson and a link function $\eta(\cdot)$. This approach of time series following a GLM is pursued e.g. by Kedem and Fokianos (2002). Restricting ourselves to first order models, we consider time series $(Y_t : t \in \mathbb{N}_0)$ following a Poisson model

$$
\begin{aligned}
Y_t|\mathcal{F}_{t-1}^{Y} &\sim Pois(\lambda_t), \\
\eta(\lambda_t) &= \beta_0 + \beta_1\eta(Y_{t-1}+c) + \alpha_1\eta(\lambda_{t-1}),\ t \geq 1,
\end{aligned}
\tag{1}
$$

where $\mathcal{F}_{t-1}^{Y}$ stands for the σ-algebra created by $\{Y_{t-1}, \ldots, Y_0, \lambda_0\}$, while $\beta_0, \beta_1, \alpha_1$ are unknown parameters, and c is a known constant. Models employing other distributions like the negative binomial could be treated similarly.

The natural choice for η is the logarithm, and this is the reason for adding the constant c to Y_{t-1} in the term $\eta(Y_{t-1} + c)$, since we need to avoid difficulties arising from observations which are equal to 0. Following Fokianos and Tjøstheim (2011), who develop ergodicity conditions for a subclass of the arising log–linear models, we set $c = 1$. Another choice for η which has received some attention is the identity,

$\eta = id$, see e.g. Ferland, Latour and Oraichi (2006). In this case we can set c to 0. For ergodicity conditions for this model class see Fokianos, Rahbek and Tjøstheim (2009).

We briefly discuss possible interpretations of models like those given in (1) in the context of epidemiology, with Y_t denoting the number of new cases observed at time t. For a fixed population size, the conditional mean λ_t measures the risk of a person to fall ill at time t then. Our model assumes that all effects on λ_t are linear after transformation to a suitable scale by η. The term $\eta(Y_{t-1} + c)$ in the second equation models the dependence of the transformed conditional mean $\eta(\lambda_t)$ and thus of the observation Y_t on the previous value Y_{t-1}, with β_1 measuring the strength of this dependence. A large number of cases Y_{t-1} at time $t-1$ can cause a large number of cases Y_t at time t because the risk of infection increases. The term $\eta(\lambda_{t-1})$ additionally describes that there can be periods of increased risk also because of certain weather conditions or expositions, for instance, and α_1 measures the size of such dependencies.

Given a model as formulated in (1), a basic question is whether it properly describes all the observations of a given time series, or whether some observations have been influenced by extraordinary effects, which are called interventions in what follows. Outlier and intervention analysis for ARMA processes of continuous variables has been developed by Fox (1972), Box and Tiao (1975), Tsay (1986), Chang, Tiao and Chen (1988) and Chen and Liu (1993), among others. However, counts are positive and typically right-skewed, causing a need for especially designed models and procedures.

In the following sections, we review the intervention models proposed by Fokianos and Fried (2010, 2012) for time series which are Poisson conditionally on the past, with η being the identity and the log-link, respectively, and describe some extensions.

2 Models for Intervention Analysis

A possibility to introduce an extraordinary effect on a time series (Y_t) generated by (1) is the assumption that from a time point τ on the underlying conditional mean process is changed by adding terms $\omega\delta^{t-\tau}I(t \geq \tau)$ to $\eta(\lambda_t)$, so that instead of (Y_t) we observe a contaminated process (Z_t) generated from a model with contamination,

$$\begin{aligned} Z_t|\mathcal{F}_{t-1}^Z &\sim Pois(\lambda_t^c), \\ \eta(\lambda_t^c) &= \beta_0 + \beta_1\eta(Z_{t-1} + c) + \alpha_1\eta(\lambda_{t-1}^c) + \omega\delta^{t-\tau}I(t \geq \tau),\ t \geq 1. \end{aligned} \tag{2}$$

In obvious notation, (λ_t^c) is the contaminated process of conditional means, which coincides with (λ_t) until time $\tau - 1$ and then becomes affected, while $\mathcal{F}_{t-1}^Z$ denotes the σ-algebra representing the information on the past of the contaminated process and the initial values, analogous to $\mathcal{F}_{t-1}^Y$. The new parameter ω determines the size of the effect, $I(t \geq \tau)$ indicates whether $t \geq \tau$ or not, and $\delta \in [0, 1]$ determines whether the effect is concentrated on time τ (in case of $\delta = 0$), causing a spiky outlier, whether the whole level is shifted from time τ on ($\delta = 1$), or whether a geometrically decaying transient shift with rate $\delta \in (0, 1)$ occurs. Note that even in case of $\delta = 0$ the whole future of the process is affected by an intervention, since its effect enters the dynamics both via Z_t and $\eta(\lambda_t^c)$, $t \geq \tau$. Continuing the explanations given above in the context of epidemiology, an intervention according to (2) can be interpreted as an internal change of the data generating process. For some reason, the conditional mean of the process (the risk) changes in an unpredictable manner at time τ, and this changes the observation for that time point, and also the observations thereafter.

Liboschik et al. (2013) explore another intervention model in case of the identity link. In their approach, an intervention affects the observation at time τ, but not the underlying conditional mean. This can be understood as an external change, as the contaminated observation Z_τ equals the sum of the uncontaminated value Y_τ plus a random number C_τ, which arises because of extraordinary reasons and enters the dynamics of the process in the same way as Y_τ, while the underlying risk λ_τ initially is not affected. An example might be people being infected due to external reasons, e.g. on a journey. The modified intervention model with a general link function η reads

$$\begin{aligned} Z_t|\mathcal{F}_{t-1}^Z &\sim Pois(\lambda_t^c), \\ \eta(\lambda_t^c) &= \eta(\lambda_t) + \omega\delta^{t-\tau} I(t \geq \tau), \\ \eta(\lambda_t) &= \beta_0 + \beta_1 \eta(Z_{t-1} + c) + \alpha_1 \eta(\lambda_{t-1}),\ t \geq 1. \end{aligned} \tag{3}$$

The last two equations describing the conditional mean process can be summarized as

$$\eta(\lambda_t^c) = \beta_0 + \beta_1 \eta(Z_{t-1} + c) + \alpha_1 \left(\eta(\lambda_{t-1}^c) - \omega\delta^{t-1-\tau} I(t-1 \geq \tau)\right) + \omega\delta^{t-\tau} I(t \geq \tau)\ .$$

This shows the difference to model (2) more clearly.

If the time point τ and the type of an intervention, i.e. the value of δ, both are known, an intervention model as formulated in (2) or (3) can be fitted by maximizing the conditional likelihood iteratively, starting from suitable initial values. The existence of such a known intervention can be confirmed by comparing the test statistics of the corresponding score test to the upper percentiles of its asymptotical χ_1^2-distribution, as described in the papers mentioned above. If only the time point τ is unknown, but the type is known, simulation experiments indicate that parametric bootstrap procedures work rather well: fit the model without intervention effects and calculate the score test statistics for all time points. Use the maximum of all score test statistics for all time points as the final test statistic. Then generate artificial time series without interventions from the fitted model and calculate the corresponding maximum score test statistic as well. Opt for an intervention at that time point which maximizes the score test statistic for the real data, if it is among the largest 100α-percent of all maximum score test statistics. If the type of the intervention is unknown as well, the maximum score test statistics can be calculated for each type given either model (2) or (3). The simulations suggest that preference should be given to level shifts ($\delta = 1$) if they turn out to be significant, since a level shift usually causes the test statistics for the other types of intervention effects also to become large, while the reverse effect is much less pronounced. Multiple interventions can be dealt with by estimating the effect of a detected intervention and subtracting it from the time series, before the cleaned data are analyzed with respect to further interventions.

Note that the above intervention models are not able to describe so called additive outliers representing e.g. pure measurement or reporting errors, i.e. the case where a single observation is changed without any effects on the future of the process. Actually, such additive outliers are difficult to deal with by a frequentist approach, since we would need to condition on the unobserved value Y_τ instead of the contaminated Z_τ. Fried et al. (2012) develop a Bayesian approach for additive outliers, applying Markov Chain Monte Carlo techniques. Their simulation results provide evidence that in this way it is possible to deal with additive outliers if there are several of them. A single or very few additive outliers pose difficulties to a Bayesian approach based on little informative prior distributions, since they do not provide enough information on that component of the underlying mixture distribution which causes the outliers.

Furthermore it should be noted that we implicitly assume intervention effects to be additive when using the identity link, and multiplicative on the original scale when using the log-link, since for simplicity we introduce the intervention effects in the same way as the dependencies on the past. Another assumption underlying the intervention models formulated above, and also the common outlier and intervention models which have been proposed for ARMA processes in the literature, is that the dynamics of the process does not change and follows the same model after an intervention as before it.

For an illustration we analyze an artificial time series of length $n = 200$ generated from model (2) with $\eta = id$, $\beta_0 = 3$, $\beta_1 = 0.4$, $\alpha_1 = 0.3$, an internal level shift of size $\omega_1 = 4$ at time $\tau_1 = 100$ and an internal spike of size $\omega_2 = 30$ at time $\tau_2 = 150$.

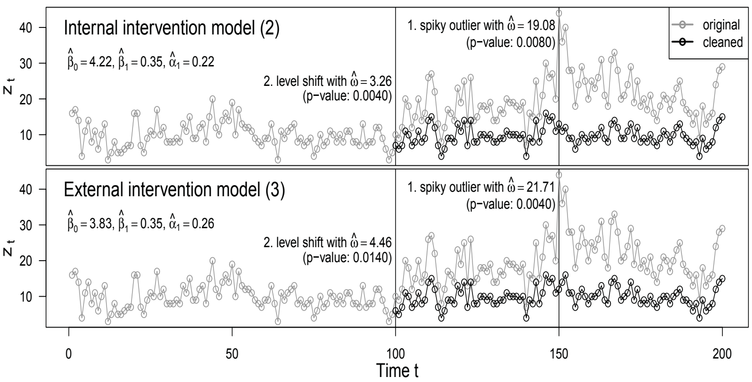

Figure 1: Results obtained from fitting both intervention models to a time series with an internal level shift at time 100 and an internal spike at time 150.

The results obtained from fitting both intervention models to these data are illustrated in Figure 1. The spike and the level shift are detected when using either of these two models, albeit with some differences between the estimated parameter values and outlier sizes, according to the different influences of such patterns on the time series. These findings confirm those of Liboschik et al. (2013): interventions can be detected successfully even if the wrong model is used. This is good news and also bad news: it is good news since it implies a certain robustness against model misspecification, but it makes a statement about the cause of an intervention effect and about its mechanism (internal / external) difficult. More work on model selection is needed for this.

3 Extensions and Outlook

3.1 Surveillance

The methods for detection of intervention effects in count time series mentioned above work retrospectively, i.e., we observe the whole time series before it is analyzed. An open problem so far is how these models can be used for surveillance, i.e. online detection of changes. This is an interesting problem not only in epidemiology, where we want to detect the outbreak of an epidemic with only short time delays. An intuitive approach is to compare an incoming observation y_{n+1} to its 1-step prediction $\hat{\lambda}_{n+1}$, obtained by estimating the parameters of model (1) from the data observed until time point n, plugging in these estimates into the formula for $\eta(\lambda_{n+1})$ and applying the inverse transform η^{-1}. There is evidence of an extraordinary effect if y_{n+1} is larger than the upper, say, 99% percentile of a Poisson distribution with parameter $\hat{\lambda}_{n+1}$.

An analysis of a single observation cannot tell us which type of intervention occurs, e.g. whether there is a spiky outlier or a level shift. For this we need to wait some more time points until further values $y_{n+2}, y_{n+3}, \ldots, y_{n+m}$ are observed, with a suitably chosen delay $m \in \mathbb{N}$. Instead of its 1-step ahead prediction, a comparison of y_{n+h} to its h-step ahead prediction might be advantageous then, since the 1-step ahead prediction will strongly be affected by a level shift at time $n+1$ due to its use of $y_{n+1}, \ldots, y_{n+h-1}$. To the best of our knowledge, so far there are no simple formulae available for the conditional expectation of Y_{n+h} given $\mathcal{F}_n^Y$ if $h \geq 2$, which is the natural candidate for h-step ahead prediction, so that we would need to rely on simulating the future given the fitted model, or use simple linear predictions instead, sticking the previous predictions $\hat{y}_{t+h-1} = \hat{\lambda}_{t+h-1}$ into the formula for $\eta(\hat{\lambda}_{t+h})$ for $h = 2, 3, \ldots, m$. However, note that the conditional distribution of Y_{n+h} given $\mathcal{F}_n^Y$ is not Poisson for $h \geq 2$, so that there is need for more research on these models.

3.2 Robust estimation

Further open questions remain concerning the robust estimation of the model parameters in the presence of outliers and intervention effects. This is even more important because it is difficult to specify intervention effects correctly and because of the difficulties in dealing with a single or a few additive outliers outlined above.

M-estimators are a popular generalization of (conditional) maximum likelihood estimators which provide some robustness against outliers by replacing the log-likelihood or the score function by more robust alternatives. An M-estimator of a parameter θ can be defined as the solution of a score equation

$$\sum_{t=1}^{n} \psi(y_t, \hat{\theta}) = 0 \; . \tag{4}$$

Maximum likelihood estimation is derived by choosing $\psi(y, \theta)$ as the derivative of the log-density $\ln f_\theta(y)$ with respect to θ, i.e. as the usual score function, while $\psi(y, \theta) = y - \theta$ corresponds to least squares and $\psi(y, \theta) = \text{sign}(y - \theta)$ to least absolute deviation estimation of location. The popular Huber M-estimator of the location parameter θ in a location-scale model with known (or preliminarily estimated) scale σ is derived from

$$\psi(y,\theta) = \frac{y-\theta}{\sigma} I(-k\sigma \leq y-\theta \leq k\sigma) + k \ \mathrm{sign}(y-\theta) I(|y-\theta| > k\sigma),$$

where k is a tuning constant which determines the efficiency and the robustness of the resulting estimator. For $k = 0$ we get least absolute deviations and for $k \to \infty$ we get least squares. The score function of the Huber M-estimator is monotone. This guarantees a unique solution which can easily be determined iteratively starting from any initial value. The score function of the Tukey M-estimator,

$$\psi(y,\theta) = \frac{y-\theta}{\sigma}\left(k^2 - \frac{(y-\theta)^2}{\sigma^2}\right)^2 I(-k\sigma \leq y-\theta \leq k\sigma),$$

however, is redescending to 0 as $y-\theta$ approaches $\pm k\sigma$. This leads to the possibility of multiple solutions of the defining score equations (4).

M-estimation of generalized linear models using the Huber ψ-function has been treated by Cantoni and Ronchetti (2001). However, in our basic model (1) we regress on previous observations and previous conditional means, and it is well known that monotone M-estimators like the Hubers need further modifications to become robust against outlying regressors. Cantoni and Ronchetti (2001) consider covariates following an elliptical distribution and use weights based on robustly estimated Mahalanobis distances to downweight observations with outlying regressors. This approach is not natural in our context, since we regress on previous observations, which are conditionally Poisson, or some transformation of them. Empirical work on model (2) with the log-link and $\alpha_1 = 0$, that is a model without feedback, indicates that in the cases of level shift and transient shift there are no significant differences between the classical maximum likelihood estimation and the approach based by Cantoni and Ronchetti (2001). This agrees with findings for Gaussian ARMA models, that maximum likelihood and least squares work rather well in case of outliers which conform to the dynamics of the process. In the case of additive outliers, the weighted approach through robust Mahalanobis distances was found to perform much better than the classical maximum likelihood estimation, especially as the number of outliers increases. In fact, some further empirical work on the feedback case ($\alpha_1 \neq 0$) indicates that the Cantoni and Ronchetti (2001) estimation approach performs better with weights.

Maronna, Martin and Yohai (2006) recommend Tukey's ψ-function since its redescending behavior completely eliminates the influence of huge outliers and provides some robustness even in the case of outlying regressors. However, we need to use highly robust initial parameter estimates then, in order not to get trapped in a wrong solution when trying to solve (1) iteratively. This and the discreteness and strong asymmetries of Poisson models pose further problems which are not encountered in ordinary symmetric location-scale models. This will briefly be illustrated in the context of independent Poisson data in the following.

Cadigan and Chen (2001) investigate a modification of the Huber score function for the Poisson distribution. Under Poisson assumptions, the variance σ^2 equals the

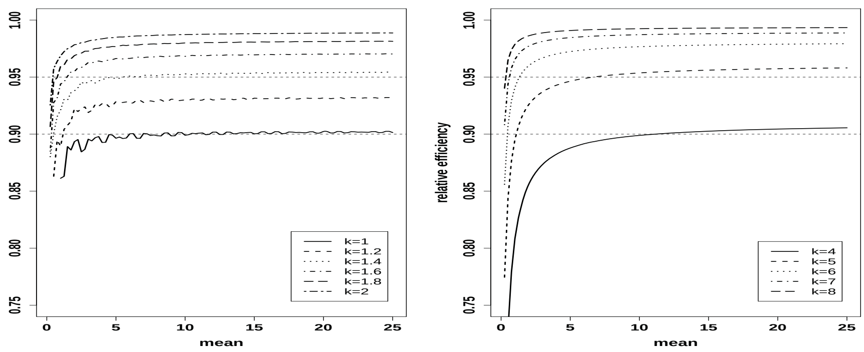

Figure 2: Asymptotical efficiencies of the Huber and the Tukey M-estimator with different tuning constants k for different values of the mean θ.

mean θ, so that we can replace σ by $\sqrt{\theta}$ in the above score functions, see also Elsaied (2012). Furthermore, the expectation of $\psi(Y,\theta)$ has to be zero for getting asymptotically unbiased estimates. This can be accomplished by introducing a bias correction a and replacing $(y-\theta)/\sigma$ by $(y-\theta)/\sqrt{\theta}-a$ in the above formulae. Given the need for a highly robust initial estimate when using the Tukey ψ-function, we might want to apply the median of the data, but this only works if it is not zero because of our scaling by $\sqrt{\hat{\theta}}$, and it provides only a very rough estimate if the sample median is small. Elsaied (2012) proposes an adaptive estimate instead, combining the sample median with an estimate derived from the frequency of zero observations.

The asymptotical distribution of an M-estimator under suitable regularity conditions is $N(\theta, V_\psi(\theta))$, with the asymptotical variance $V_\psi(\theta) = E(\psi(Y,\theta)/B_\theta)^2$, where $B_\theta = \partial E\psi(Y,\theta)/\partial\theta$, see e.g. Maronna, Martin and Yohai (2006). The relative efficiency of an M-estimator as compared to the maximum likelihood estimator, which is the sample mean, under these conditions thus becomes $\theta/V_\psi(\theta)$, and is illustrated in Figure 2. Note that an estimator with a fixed tuning constant k does not achieve a desirable high level of efficiency for all possible values of θ. For further investigations in this respect and a first attempt to formulate robust M-estimators for model (1) with $p=1$, $q \in \{0,1\}$ and the identity link see Elsaied (2012).

Acknowledgement: The work of K. Fokianos and S. Kitromilidou is supported by Cyprus Research Promotion Foundation TEXNOLOGIA/THEPIS/0609(BE)/02.

References

[1] Box G.E.P., Tiao G.C. (1975). Intervention analysis with applications to economics and environmental problems. *Journal of the American Statistical Association*. Vol. **70**, pp. 70–79.

[2] Cadigan N.G., Chen J. (2001). Properties of robust M-estimators for Poisson and negative binomial data. *Journal of Statistical Computation and Simulation*. Vol. **70**, pp. 273–288.

[3] Cantoni E., Ronchetti E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*. Vol. **96**, pp. 1022–1030.

[4] Chang I., Tiao G.C., Chen C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*. Vol. **30**, pp. 193–204.

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

[6] Elsaied H. (2012). Robust Modelling of Count Data. *Unpublished PhD-thesis. Department of Statistics, TU Dortmund University, Germany.*

[7] Ferland R., Latour A., Oraichi D. (2006). Integer–valued GARCH processes. *Journal of Time Series Analysis*. Vol. **27**, pp. 923–942.

[8] Fokianos K., Fried R. (2010). Interventions in INGARCH processess. *Journal of Time Series Analysis*. Vol. **31**, pp. 210–225.

[9] Fokianos K., Fried R. (2012). Interventions in log-linear Poisson autoregression. *Statistical Modelling*. Vol. **12**, pp. 299–322.

[10] Fokianos K., Tjøstheim D. (2011). Log–linear Poisson autoregression. *Journal of Multivariate Analysis*. Vol. **102**, pp. 563–578.

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

[12] Fokianos K., Rahbek A., Tjøstheim D. (2009). Poisson autoregression. *Journal of the American Statistical Association*. Vol. **104**, pp. 1430–1439.

[13] Fried R., Agueusop I., Bornkamp B., Fokianos K., Fruth J., Ickstadt K. (2012). Bayesian outlier detection in INGARCH time series. *Discussion Paper 26/12, SFB 823, TU Dortmund, Germany.*

[14] Kedem B., Fokianos K. (2002). *Regression Models for Time Series Analysis*. Wiley, Hoboken, NJ.

[15] Liboschik T., Kerschke P., Fokianos K., Fried R. (2013). Modelling interventions in INGARCH processes. *Discussion Paper 03/13, SFB 823, TU Dortmund, Germany.*

[16] Maronna R.A., Martin R.D., Yohai V.J. (2006). *Robust Statistics*. Wiley, New York.

[17] Tsay R.S. (1986). Time series model specification in the presence of outliers. *Journal of the American Statistical Association*. Vol. **81**, pp. 132–141.