



On Outliers and Interventions in Count Time Series following GLMs

Roland Fried Tobias Liboschik Hanan Elsaied S. Kitromilidou K. Fokianos
TU Dortmund TU Dortmund Suez Canal Univ. Univ. of Cyprus Univ. of Cyprus

Abstract

We discuss the analysis of count time series following generalised 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 extensions to the problem of robust parameter estimation, identification of the model orders by robust estimation of autocorrelations and partial autocorrelations, and online surveillance by sequential testing for outlyingness is provided.

Keywords: discrete data, model identification, robustness, (partial) autocorrelations, surveillance.

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 generalised 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). Focusing on first order models, we consider time series $(Y_t : t \in \mathbb{N}_0)$ following a Poisson model

$$\begin{aligned} Y_t | \mathcal{F}_{t-1}^Y &\sim \text{Pois}(\lambda_t), \\ \eta(\lambda_t) &= \beta_0 + \beta_1 \eta(Y_{t-1} + c) + \gamma_1 \eta(\lambda_{t-1}), \quad t \geq 1, \end{aligned} \tag{1}$$

where \mathcal{F}_{t-1}^Y stands for the σ -algebra created by $\{Y_{t-1}, \dots, Y_0, \lambda_0\}$, while $\beta_0, \beta_1, \gamma_1$ 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 γ_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.

The remainder of the paper is organised as follows. Section 2 generalises 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. Section 3 reviews first attempts of robust fitting of models with known link function and model orders. Section 4 reports a first study of model identification for the linear model applying the identity link, using robust estimators of the autocorrelations and partial autocorrelations. Section 5 provides an outlook to surveillance, that is online monitoring by sequential outlier detection.

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,

$$\begin{aligned} Z_t | \mathcal{F}_{t-1}^Z &\sim \text{Pois}(\lambda_t^c), \\ \eta(\lambda_t^c) &= \beta_0 + \beta_1\eta(Z_{t-1} + c) + \gamma_1\eta(\lambda_{t-1}^c) + \omega\delta^{t-\tau}I(t \geq \tau), \quad t \geq 1. \end{aligned} \quad (2)$$

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, e.g. due to particular weather conditions or other expositions, 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

$$\begin{aligned} Z_t | \mathcal{F}_{t-1}^Z &\sim \text{Pois}(\lambda_t^c), \\ \eta(\lambda_t^c) &= \eta(\lambda_t) + \omega \delta^{t-\tau} I(t \geq \tau), \\ \eta(\lambda_t) &= \beta_0 + \beta_1 \eta(Z_{t-1} + c) + \gamma_1 \eta(\lambda_{t-1}), \quad t \geq 1. \end{aligned} \quad (3)$$

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

$$\eta(\lambda_t^c) = \beta_0 + \beta_1 \eta(Z_{t-1} + c) + \gamma_1 (\eta(\lambda_{t-1}^c) - \omega \delta^{t-1-\tau} I(t-1 \geq \tau)) + \omega \delta^{t-\tau} I(t \geq \tau).$$

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 maximising 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 maximises 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 analysed 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. (2013) 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 analyse an artificial time series of length $n = 200$ generated from model (2) with $\eta = id$, $\beta_0 = 3$, $\beta_1 = 0.4$, $\gamma_1 = 0.3$, 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 $\tau_2 = 150$.

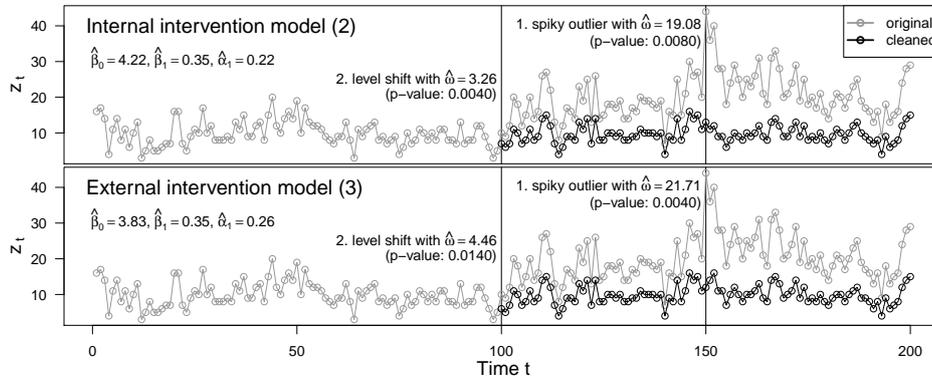


Figure 1: Results obtained from fitting both intervention models to a time series with an internal level shift at time 100 and an internal spike at time 150.

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. Robust estimation

First attempts are available concerning the robust estimation of the model parameters in the presence of outliers and intervention effects. This is even more important because of the difficulties in specifying intervention effects correctly and because of the remaining difficulties in dealing with a single or a few additive outliers outlined above.

M-estimators are a popular generalisation 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

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

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

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

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 $k \rightarrow \infty$ 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 re-descending 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 generalised 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 those based on the Huber function 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 down-weight 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 $\gamma_1 = 0$, 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 on 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 ($\gamma_1 \neq 0$) indicates that the Cantoni and Ronchetti (2001) estimation approach performs better with weights (Kitromilidou and Fokianos, 2014).

Maronna, Martin and Yohai (2006) recommend Tukey's ψ -function since its re-descending 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 (4) 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 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

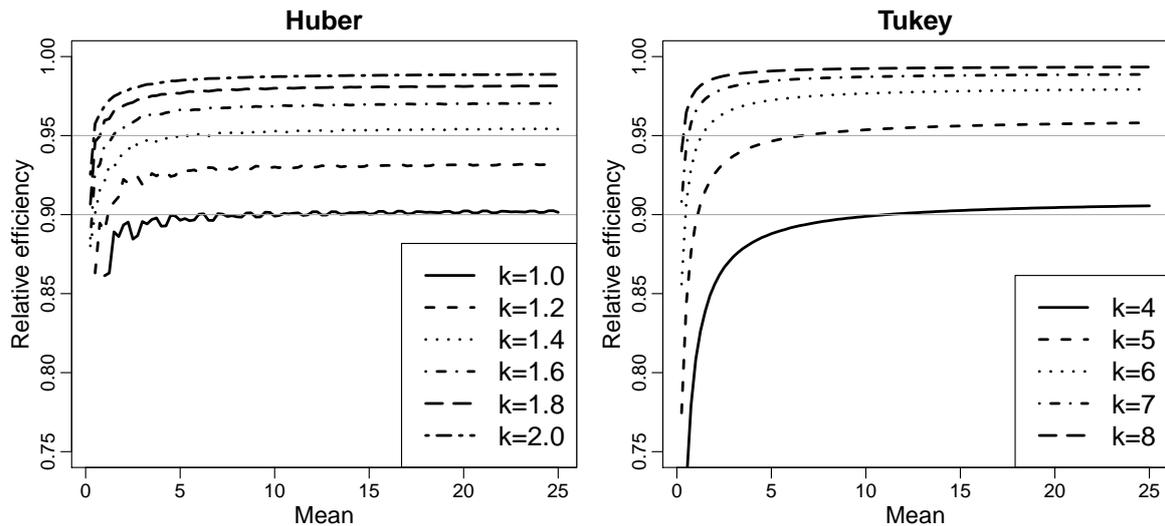


Figure 2: Asymptotical efficiencies of the Huber and the Tukey M-estimator with different tuning constants k for different values of the mean θ .

for all possible values of θ . For further investigations in this respect and a first approach to robust M-estimation for model (1) with the identity link see Elsaied (2012).

4. Robust model identification

Besides the robust estimation of the parameters of a specific model, the proper identification of the link function and the model orders gets more complicated in the presence of outliers. In the following we provide a first robustness study for the identification of the model orders in case of a linear model with the identity link.

Two common tools for the choice of the model orders of linear time series models are the sample autocorrelation function (SACF) and the sample partial autocorrelation function (SPACF). However, these are strongly affected by outlying observations so that there is a need for robust and efficient alternatives. Let $\mathbf{y} = (y_1, \dots, y_n)'$ be an observed time series. We consider estimation of the autocorrelation at lag h by a robust bivariate correlation estimator applied to the vector $\mathbf{y}_t^h = (y_{1+h}, \dots, y_n)'$ and the vector of lagged observations $\mathbf{y}_{t-h}^h = (y_1, \dots, y_{n-h})'$. We consider the rank-based correlation estimators Spearman's ρ , Kendall's τ and Gaussian rank (for a comparison in the bivariate context see Boudt et al., 2012). Another class of autocorrelation estimators, which is based on an idea of Gnanadesikan and Kettenring (1972), employs any robust univariate scale estimator $\widehat{\text{var}}(\cdot)$. We use a variant bounded between -1 to 1 inclusive, which at lag h is given by

$$\widehat{\text{acf}}_{GK}(\mathbf{y}; h) = \frac{\widehat{\text{var}}(\mathbf{y}_t^h + \mathbf{y}_{t-h}^h) - \widehat{\text{var}}(\mathbf{y}_t^h - \mathbf{y}_{t-h}^h)}{\widehat{\text{var}}(\mathbf{y}_t^h + \mathbf{y}_{t-h}^h) + \widehat{\text{var}}(\mathbf{y}_t^h - \mathbf{y}_{t-h}^h)}.$$

Ma and Genton (2000) study this Gnanadesikan-Kettenring (GK) approach in the Gaussian framework, using the highly robust Q_n estimator of scale proposed by Croux and Rousseeuw (1992). We additionally consider the median absolute deviation from the median (MAD), the 10% and 20% winsorised variance, the interquartile range (IQR), as well as the highly robust S_n (Croux and Rousseeuw, 1992) and τ (Maronna and Zamar, 2002) estimators of scale. Apart from the winsorised variance, these estimators are on the scale of the original data and need to be squared.

We compare estimators which are corrected such that they achieve consistency at the normal distribution. Note that the normal distribution is a limiting case of a Poisson distribution

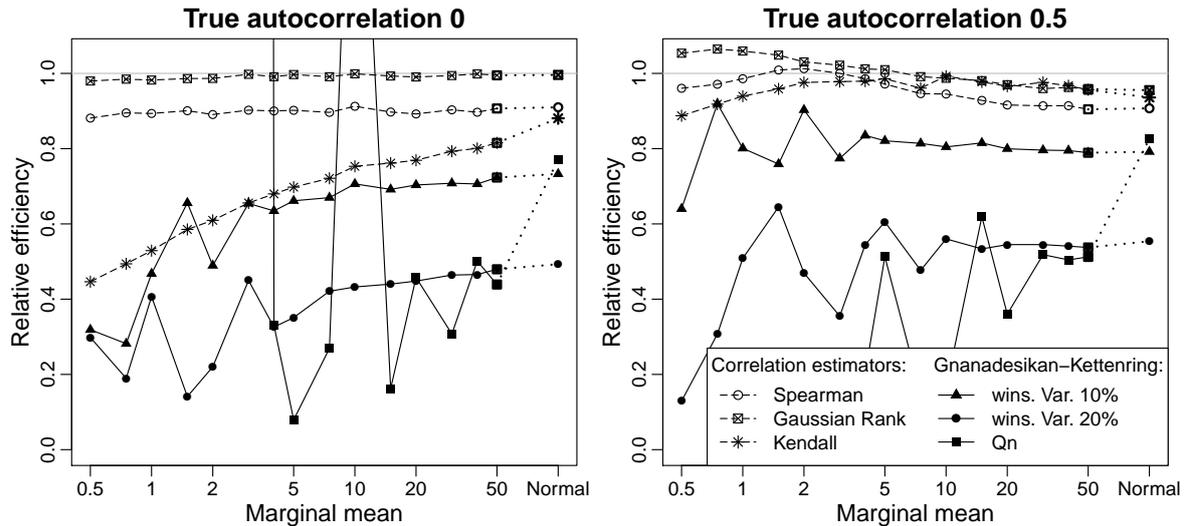


Figure 3: Efficiency of autocorrelation estimators at lag $h = 1$ relatively to the SACF. Time series of length 100 are simulated from model (1) with the marginal mean given on the horizontal axis and from a $N(\lambda_t, \lambda_t)$ model with a marginal mean of 50 (points on the very right of each plot).

with mean tending to infinity. However, we cannot expect this Fisher-consistency correction to hold true, especially in the case of a clearly skewed Poisson distribution with a small mean. Moreover, the marginal distribution of a time series from model (1) is strictly speaking only Poisson under the null hypothesis of independence.

In our simulation study we generate time series with 100 observations from the first order linear Poisson model (1) with $\eta = id$, $c = 0$ and $\gamma_1 = 0$. We consider scenarios with a true autocorrelation at lag $h = 1$ of zero ($\beta_1 = 0$) and of 0.5 ($\beta_1 = 0.5$). The results are averaged over 10 000 repetitions for each scenario and reported as a function of the marginal mean $\mu = \beta_0 / (1 - \beta_1)$. The shown relative efficiencies are the ratio of the mean square errors of the SACF and the respective estimator.

The GK autocorrelation estimators based on Q_n (see Figure 3), S_n , MAD and IQR are unsuitable for small counts, as these estimators are unstable due to the high proportion of ties in such data. It frequently happens that the scale estimations $\widehat{\text{var}}(\mathbf{y}_t^h + \mathbf{y}_{t-h}^h)$ and $\widehat{\text{var}}(\mathbf{y}_t^h - \mathbf{y}_{t-h}^h)$ coincide, resulting in an autocorrelation estimate of zero, or that one or both of them collapse to zero, resulting in an estimate of ± 1 or a non-computable autocorrelation estimation, respectively. Particularly for small marginal means, we get zero estimates with high probability, causing a super-efficient performance if the true autocorrelation is zero. Implosion, that is breakdown to zero, is a known problem of many robust scale estimators. But not even the Q_n estimator, which showed the best performance with respect to implosion among many other alternatives in a study of Gather and Fried (2003), does perform acceptably in the case of small counts. We also tried variants of the Q_n using the 50%- and 75%-quantile of the pairwise distances, instead of the 25%-quantile as it is usually employed. Yet, for counts with low means none of these alternatives perform well. The τ estimator of scale as implemented by Maronna and Zamar (2002) is based on the variance estimation of the MAD and hence also performs poorly. We conclude that none of these popular highly robust scale estimators seems to be appropriate for small counts. Particularly for a low winsoring proportion, the winsorised variance estimator results in smaller problems with stability than the estimators mentioned before and will be considered further.

Figure 3 reconfirms the result that the efficiency of the estimators relatively to the SACF tends to its value achieved under a normal distribution. The Gaussian rank estimator has a very high relative efficiency both for uncorrelated and autocorrelated data, which does not

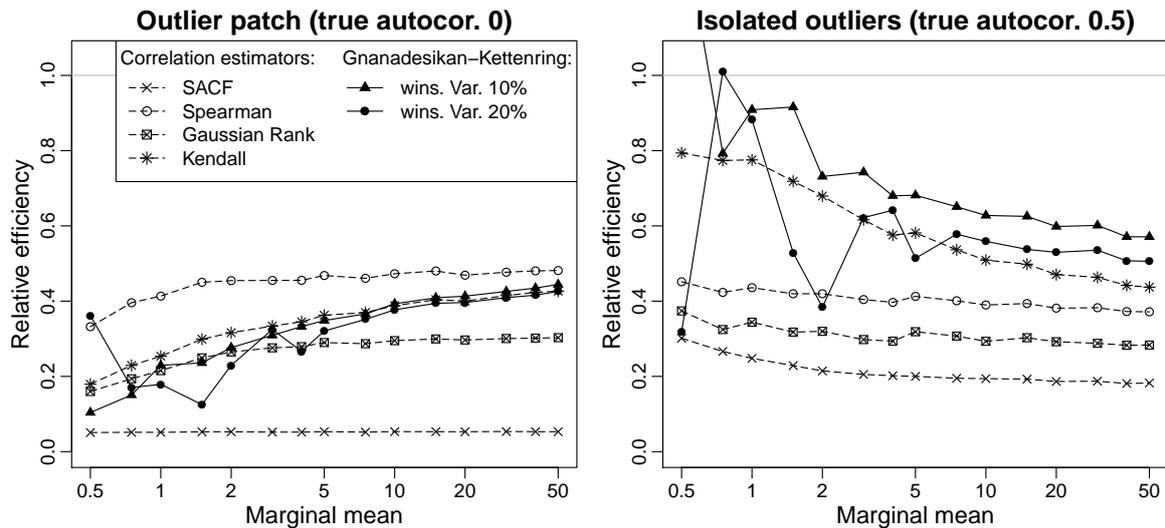


Figure 4: Efficiency of autocorrelation estimators at lag $h = 1$ for contaminated Poisson data relatively to the SACF for uncontaminated Poisson data. We contaminated 5% of the 100 observations with additive outliers of size five times the marginal standard deviation. Left: Patchy outliers in the centre. Right: Isolated outliers at arbitrarily chosen positions 17, 40, 55, 72 and 92.

depend a lot on the marginal mean. Spearman's ρ correlation estimator behaves in a similar fashion, but has a lower relative efficiency of about 90% on uncorrelated data. In contrast, the relative efficiency of Kendall's τ depends very much on the marginal mean. In case of uncorrelated data its relative efficiency is below 50% for small means and even for large means slightly below Spearman's ρ .

To study the robustness properties of the estimators, we contaminate the time series of independent data, that is $\beta_1 = 0$, with a patch of 5% additive outliers in the centre and the autocorrelated ones with 5% of isolated additive outliers. The first outlier scenario is known to bias the estimation towards one and the latter one biases towards zero, which is away from the true values of zero and 0.5, respectively. For autocorrelation estimation when $\beta_1 = 0$, outlier patches are the worst case, whereas for time series with $\beta_1 > 0$ they can even compensate for an existing downward bias in finite samples. The simulation results in Figure 4 can be interpreted as the loss of efficiency compared to the SACF for uncontaminated data from the same model.

The outlier patch has a strong effect on the efficiency of the autocorrelation estimators for uncorrelated data (see Figure 4 left). The ordinary SACF is not robust and drops down to a relative efficiency of around 5%. The rank-based autocorrelation estimators show qualitatively the same pattern of increasing relative efficiency for increasing marginal mean. The Gaussian rank correlation, which has been the most efficient rank-based estimator for clean uncorrelated data, is the least robust one, because it gives more influence to the largest and the smallest observations. The 10%-winsorised variance has an efficiency of around 10% relative to the SACF for clean data, which also increases with the marginal mean to about 40%. The 20%-winsorised variance is in principle slightly less efficient and shows a similar behaviour but is, as for uncontaminated data, quite unstable for low means.

The same number of isolated outliers for moderately correlated data has a weaker effect on the efficiency of the autocorrelation estimators than the outlier patch for uncorrelated data (see Figure 4 right). Unlike in the latter situation, we observe a decreasing relative efficiency for an increasing marginal mean for all estimators, except for the instability of the GK estimation based on the 20%-winsorised variance, which has been discussed before. Again, the Gaussian rank based estimator is the least efficient among the rank-based estimators, but this time

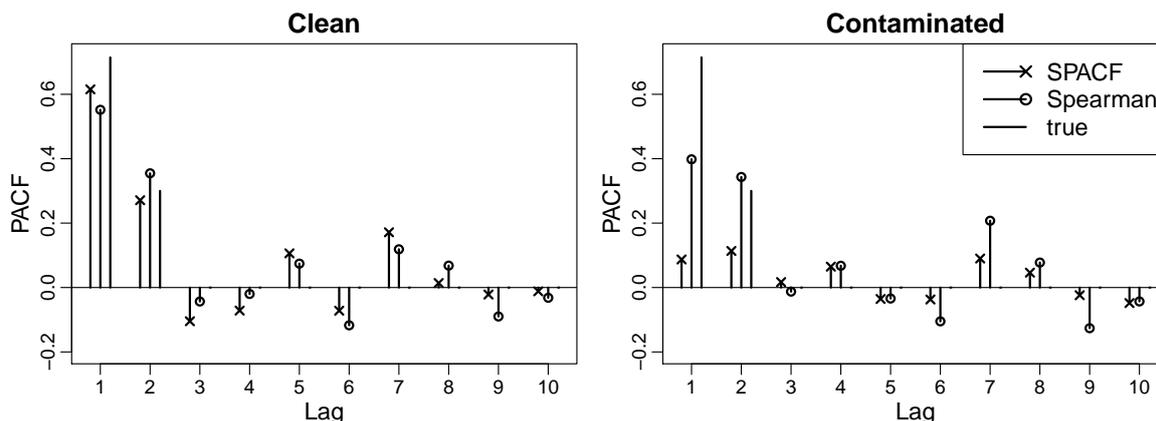


Figure 5: Estimated PACF of a simulated INARCH(2) time series of length 100 with parameters $\beta_0 = 0.4$, $\beta_1 = 0.5$ and $\beta_2 = 0.3$. Left: Clean data. Right: Contaminated with five additive outliers of size five times the marginal standard deviation at arbitrarily chosen positions 17, 40, 55, 72 and 92.

Kendall's τ is much more efficient than Spearman's ρ , particularly for low marginal means.

Because of the instability of most of the other estimators we recommend to use one of the rank-based autocorrelation estimators for count time series with small counts. When choosing an autocorrelation estimator one should take into account both, the desired efficiency at clean data and the desired robustness properties.

We illustrate the usefulness of robust autocorrelation estimation for identification of the model order with a simulated example. Consider a time series $(Y_t : t \in \mathbb{N}_0)$ from an integer-valued ARCH model of unknown order $p \in \mathbb{N}_0$, called INARCH(p), with $Y_t | \mathcal{F}_{t-1}^Y \sim \text{Pois}(\lambda_t)$ and conditional mean equation $\lambda_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_p Y_{t-p}$ for $t \geq 1$. We want to determine the model order p . The time series $(Y_t : t \in \mathbb{N}_0)$ has the same second-order properties as an AR(p) model (cf. Ferland et al., 2006). Hence, it is known that the partial autocorrelation function (PACF) is non-zero for lags up to p and zero for larger lags. We obtain the estimated partial autocorrelation function from the estimated autocorrelation function by applying the Durbin-Levinson algorithm (see for example Morettin, 1984).

Looking at Figure 5, we see that one can correctly identify the model order of an INARCH(2) model by looking at the SPACF or at the estimated PACF derived from the ACF estimation based on Spearman's ρ : both estimations are clearly larger than zero for the first two lags and close to zero for all other lags. In case of a contamination with isolated outliers the non-robust estimation with the SPACF is pushed towards zero, such that one might falsely identify a model of order $p = 0$. As opposed to this, the robust estimation of the PACF with Spearman's ρ is not so strongly affected by the outliers and would still allow a correct model specification.

Since the Spearman correlation coefficient measures monotone, but not necessarily linear dependence, one might speculate about its possible value for the identification of the model orders in case of models applying (monotone) link functions different from the identity. However, a thorough examination of this is beyond the scope of this work.

5. Surveillance

The methods for detection of intervention effects in count time series described above can be applied retrospectively, i.e. when we observe the whole time series before it is analysed. 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 for example in epidemiology, where we want to detect

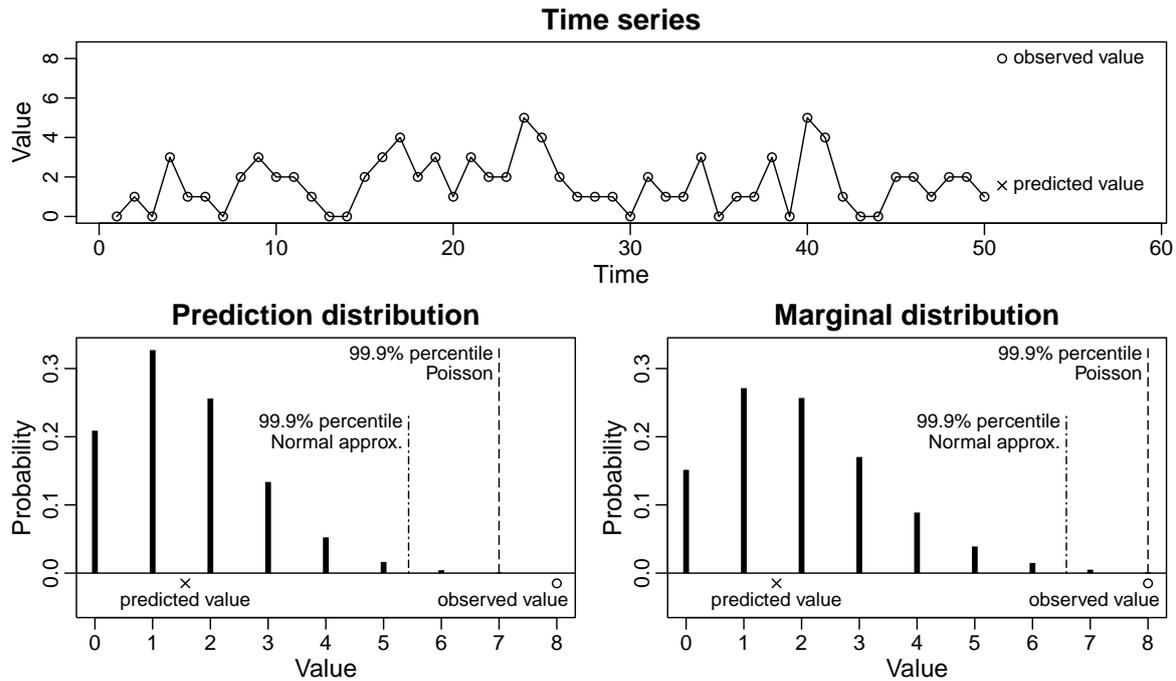


Figure 6: Simulated example for the proposed monitoring procedure. The value observed at time 51 is beyond the 99.9% percentile of the prediction, but not the marginal distribution, and would thus be identified as an outlier. A normal approximation would provide somewhat different critical values but the same conclusion in this case.

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 fitting model (1) from the data observed until time point n , plugging in the estimated parameters into the formula for $\eta(\lambda_{n+1})$ and applying the inverse transform η^{-1} . Given such a model, there is evidence of an extraordinary effect at time $n+1$ if y_{n+1} is larger than the upper $1 - \alpha_N$ percentile of a Poisson distribution with mean $\hat{\lambda}_{n+1}$. Assuming the model and its parameters to be known exactly, choosing $\alpha_N = 1 - (1 - \alpha)^{1/N}$ ensures that we do not falsely detect any outlier with probability $1 - \alpha$ when applying this rule to N subsequent predictions. This follows along the same lines as in Davies and Gather (1993), who treat the independent case, since we control the probability of detecting an outlier conditionally on the past \mathcal{F}_n^Y . As an example, for $N = 50$ predictions an individual level of $\alpha_N = 0.1\%$ yields a global level of $\alpha = 4.9\%$. Other error probabilities α_N can be chosen for tuning the sensitivity and the specificity of the sequential detection procedure. For large means $\hat{\lambda}_{n+1}$ of the prediction distribution one would also need to consider downward outliers. In this case one defines, in the terminology of Davies and Gather (1993), an outlier identifier by a lower and an upper bound both depending on α_N .

We illustrate the approach outlined above with a simulated example. We generate a time series from the first order linear Poisson model (1) with $\eta = id$, $c = 0$, $\beta_0 = 1$, $\beta_1 = 0.3$ and $\gamma_1 = 0.2$ (see Figure 6 top). In order to assess whether observation y_{51} is notably large, we fit the model on the previous observations y_1, \dots, y_{50} and, based on this, compute its 1-step ahead prediction $\hat{\lambda}_{51}$. Compared with the 99.9% percentile of the 1-step prediction distribution for y_{51} , a Poisson with mean $\hat{\lambda}_{51}$, the observed value y_{51} is large and therefore identified as a potential outlier (see Figure 6 bottom left). In this case, one would have come to the same decision if we compare y_{51} with the 99.9% percentile of a $N(\hat{\lambda}_{51}, \hat{\lambda}_{51})$, a normal approximation of the 1-step prediction distribution. Note that we would not have identified this observation as a potential outlier if we compare it with the 99.9% percentile of the marginal distribution of the process (see Figure 6 bottom right). Since no analytical

formula for this percentile is available, we approximated it by simulation of a time series with 100 000 observations.

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 $y_{n+2}, y_{n+3}, \dots, y_{n+m}$ 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 $y_{n+1}, \dots, y_{n+h-1}$. 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 $h = 2, 3, \dots, m$. 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.

Acknowledgements

The work of K. Fokianos and S. Kitromilidou is supported by Cyprus Research Promotion Foundation TEXNOLOGIA/THEPIS/0609(BE)/02. The work of R. Fried and T. Liboschik is supported by the German Research Foundation (DFG, SFB 823 “Statistical modelling of nonlinear dynamic processes”).

References

- Boudt K., Cornelissen J., Croux C. (2012). The Gaussian Rank Correlation Estimator: Robustness Properties. *Statistics and Computing*. Vol. **22**, pp. 471–483.
- 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.
- 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.
- Cantoni E., Ronchetti E. (2001). Robust Inference for Generalized Linear Models. *Journal of the American Statistical Association*. Vol. **96**, pp. 1022–1030.
- Chang I., Tiao G.C., Chen C. (1988). Estimation of Time Series Parameters in the Presence of Outliers. *Technometrics*. Vol. **30**, pp. 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*. Vol. **88**, pp. 284–297.
- Croux C., Rousseeuw P.J. (1993). Time-Efficient Algorithms for two Highly Robust Estimators of Scale. In: Dodge Y., Whittaker J. *Computational Statistics Volume 1*, pp. 284–297. Physika, Heidelberg.
- Davies L., Gather U. (1993). The Identification of Multiple Outliers. *Journal of the American Statistical Association*. Vol. **88**, pp. 782–792.
- Elsaied H. (2012). Robust Modelling of Count Data. *Unpublished PhD thesis. Department of Statistics, TU Dortmund University, Germany*. <http://hdl.handle.net/2003/29404>.
- Ferland R., Latour A., Oraichi D. (2006). Integer-valued GARCH Processes. *Journal of Time Series Analysis*. Vol. **27**, pp. 923–942.

- Fokianos K., Fried R. (2010). Interventions in INGARCH Processes. *Journal of Time Series Analysis*. Vol. **31**, pp. 210–225.
- Fokianos K., Fried R. (2012). Interventions in Log-linear Poisson Autoregression. *Statistical Modelling*. Vol. **12**, pp. 299–322.
- Fokianos K., Tjøstheim D. (2011). Log-linear Poisson Autoregression. *Journal of Multivariate Analysis*. Vol. **102**, pp. 563–578.
- Fokianos K., Rahbek A., Tjøstheim D. (2009). Poisson Autoregression. *Journal of the American Statistical Association*. Vol. **104**, pp. 1430–1439.
- Fox A.J. (1972). Outliers in Time Series. *Journal of the Royal Statistical Society. Series B*. Vol. **34**, pp. 350–363.
- Fried R., Agueusop I., Bornkamp B., Fokianos K., Fruth J., Ickstadt K. (2013). Bayesian Outlier Detection in INGARCH Time Series. *Statistics and Computing*. To appear.
- Gather U., Fried R. (2003). Robust Estimation of Scale for Local Linear Temporal Models. *Tatra Mathematical Publications*. Vol. **26**, pp. 87–101.
- Gnanadesikan R., Kettenring J.R. (1972). Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data. *Biometrics*. Vol. **28**, pp. 81–124.
- Kedem B., Fokianos K. (2002). *Regression Models for Time Series Analysis*. Wiley, Hoboken, NJ.
- Kitromilidou S., Fokianos K. (2014). *Robust Estimation Methods for a Class of Count Time Series Log-Linear Models*. Submitted for publication.
- Liboschik T., Kerschke P., Fokianos K., Fried R. (2013). Modelling Interventions in INGARCH Processes. *SFB 823 Discussion Paper 03/13, TU Dortmund University, Germany*. <http://hdl.handle.net/2003/29878>.
- Ma Y., Genton M.G. (2000). Highly Robust Estimation of the Autocovariance Function. *Journal of Time Series Analysis*. Vol. **21** pp. 663–684.
- Maronna R.A., Martin R.D., Yohai V.J. (2006). *Robust Statistics*. Wiley, New York.
- Maronna R.A., Zamar R.H. (2002). Robust Estimates of Location and Dispersion of High-dimensional Datasets. *Technometrics*. Vol. **44**, pp. 307–317.
- Morettin P.A. (1984). The Levinson Algorithm and its Applications in Time Series Analysis. *International Statistical Review*. Vol. **52**, pp. 83–92.
- Tsay R.S. (1986). Time Series Model Specification in the Presence of Outliers. *Journal of the American Statistical Association*. Vol. **81**, pp. 132–141.

Affiliation:

Roland Fried

Statistics in Biosciences

Department of Statistics

TU Dortmund University

44221 Dortmund, Germany

E-mail: fried@statistik.tu-dortmund.deURL: <http://www.statistik.tu-dortmund.de/fried.html>