On robust estimation of negative binomial INARCH models

Hanan Elsaied[1] · Roland Fried[2]

Abstract

We discuss robust estimation of INARCH models for count time series, where each observation conditionally on its past follows a negative binomial distribution with a constant scale parameter, and the conditional mean depends linearly on previous observations. We develop several robust estimators, some of them being computationally fast modifications of methods of moments, and some rather efficient modifications of conditional maximum likelihood. These estimators are compared to related recent proposals using simulations. The usefulness of the proposed methods is illustrated by a real data example.

1 Introduction

Let $Y_1, \ldots, Y_n$ be a time series of counts like the weekly number of people falling ill in epidemiology, the number of transactions per minute in finance, or jobs sent to a server during an hour. Zhu [28] extends the Poisson integer valued GARCH (more briefly called INGARCH) model, which has been put forward by [12,14], among others, to scenarios where the conditional distribution of Y_t given the past of the process exhibits overdispersion. In the arising NBINGARCH model this conditional distribution is assumed to belong to the negative binomial family,

✉ Roland Fried
fried@statistik.tu-dortmund.de

Hanan Elsaied
hanan_abdelkareem@commerce.suez.edu.eg

[1] Department of Statistics, Faculty of Commerce, Suez Canal University, Ismailia 41522, Egypt

[2] Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany

$$Y_t | \mathcal{F}_{t-1} \sim NB(r, p_t).$$

Here, $\mathcal{F}_{t-1}$ is the σ-algebra describing the information provided by the past of the process up to time $t-1$, $r \in \mathbb{N}$ is the constant number of successes, and p_t is the time-varying probability of success. He assumes a linear model for the inverse odds $\lambda_t = (1 - p_t)/p_t$, which is regressed on past values $Y_{t-1}, \ldots, Y_{t-p}, \lambda_{t-1}, \ldots, \lambda_{t-q}$. The negative binomial is a natural extension of the Poisson and covers a broad variety of overdispersed unimodal distributions. Zhu [28] studies Yule–Walker estimators for the parameter $\theta = (\alpha_0, \ldots, \alpha_p)'$ of NBINGARCH($p, 0$) models, also called NBINARCH(p) models, and conditional least squares and conditional maximum likelihood estimators for the parameter of NBINGARCH(p, q) models for given values of r.

A first approach for robust fitting of NBINGARCH(p, q) models has been published by [27]. Like these authors we parameterize the negative binomial distribution in terms of the conditional mean $\mu_t \geq 0$ and the parameter $\kappa = 1/r \geq 0$. The conditional mean and variance of Y_t are

$$\mu_t = E(Y_t | \mathcal{F}_{t-1}) = r\lambda_t = r \frac{1 - p_t}{p_t}$$
$$\text{and } \sigma_t^2 = Var(Y_t | \mathcal{F}_{t-1}) = r\lambda_t + r\lambda_t^2 = \mu_t + \kappa\mu_t^2,$$

respectively, i.e., κ measures the amount of overdispersion. The conditional Poisson model corresponds to the limiting special case $\kappa = 0$. An analogous parametrization has also been used by [1] in the context of generalized linear models for independent observations because of its greater numerical stability. While κ can take any positive value, r is usually restricted to be an integer. The linear model of [28] mentioned above can be expressed equivalently in terms of μ_t as

$$\mu_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{-i} + \sum_{j=1}^{q} \beta_j \mu_{t-j},$$

with $\alpha_0, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ being positive parameters.

The interest of [27] is in transaction counts, where very large numbers of observations are available for model fitting. Accordingly, they focus on scenarios with at least $n = 1000$ observations. Moreover, they usually fix the dispersion parameter κ at the true value in their simulations. Our interest is also in scenarios with only a few hundred observations, where robust joint estimation of the shape parameter κ and the regression parameters $\beta_1, \ldots, \beta_q$ measuring the effects of unobserved past conditional means becomes very difficult. We avoid the latter and focus on the simpler class of NBINARCH(p) models, where we regress on past observations, only. NBINARCH models offer the advantage that the observed process $(Y_t : t \in \mathbb{N}_0)$ forms a p-th order Markov chain, which implies some simplifications in practice. Furthermore, Xiong and Zhu recommend the MCD [24] for dealing with outlying past observations, although the MCD has been designed for multivariate elliptically symmetric continuous measurements [17]. This may eliminate observations from the estimating equations although just a single preceding value is spurious. We instead shrink outlying past values towards an estimate of the marginal mean to avoid unnecessary eliminations.

The rest of the paper is organized as follows. In Sect. 2, we investigate robust versions of method of moment estimators, which are useful to get a first idea about a suitable model and its order p. In Sect. 3, we discuss robustifications of conditional maximum likelihood estimators for NBINARCH models. Section 4 illustrates the methods by a data example and Sect. 5 concludes.

2 Estimation derived from method of moments

Moment type estimators are computationally tractable even in case of high model orders if the moments are estimated individually without the need of multidimensional optimization. A possible drawback arises if the parameters of interest are nonlinear transforms of the moments used, since estimation errors can increase a lot in this transformation then. Nevertheless, moment type estimators are convenient for getting a first idea about suitable model orders and for initialization of more complex estimators like those discussed in the next section.

Here and in the following, we make use of popular ψ-functions as proposed by Huber and by Tukey. The Huber ψ-function compromises least squares and least absolute deviations,

$$\psi_c(x) = x \min\left(1, \frac{c}{|x|}\right). \tag{1}$$

Tukey's biweight ψ-function is

$$\psi_c(x) = x \left(c^2 - x^2\right)^2 I_{[-c,c]}(x), \tag{2}$$

where I_A is the indicator function of a set A. The ψ-functions are applied to standardized values $(y - \mu)/\sigma$. The tuning constant c regulates the robustness and the efficiency of the estimators. For both these ψ-functions, larger values of c increase the efficiency but reduce the robustness to outliers. The Poisson and the negative binomial distributions do not form location-scale families and the tail behaviour depends on the parameters, so that suitable choices of the tuning constant in principle depend on the parameters.

For our calculations we need the formula for the conditional probabilities corresponding to the negative binomial distribution in terms of μ_t and κ, which is

$$P(Y_t = y_t | \mathcal{F}_{t-1}) = \frac{\Gamma(y_t + 1/\kappa)}{y_t! \, \Gamma(1/\kappa)} (\kappa \mu_t + 1)^{-1/\kappa} \left(\frac{\kappa \mu_t}{\kappa \mu_t + 1}\right)^{y_t}, \quad y_t \in \mathbb{N}_0.$$

2.1 Conditions for mean and second-order stationarity

Many of [28] formulae simplify if the parametrization in terms of κ and the regression parameters $\theta = (\alpha_0, \ldots, \alpha_p)'$ for the conditional mean is used. The necessary and sufficient condition for the NBINARCH(p) model to be stationary in the mean stated in his Theorem 1 then becomes that all roots of the equation

$$1 - \sum_{i=1}^{p} \alpha_i z^{-1} = 0$$

lie inside the unit circle. Under this assumption, the marginal mean is

$$\mu = E(Y_t) = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i}, \tag{3}$$

as in the limiting conditional Poisson case. The additional condition for second-order stationarity is that all roots of $1 - C_1 z^{-1} - \cdots - C_p z^{-p} = 0$ lie inside the unit circle, where

$$C_u = (1 + \kappa) \left(\alpha_u^2 - \alpha_u \sum_{v=1}^{p-1} \sum_{|i-j|=v} \alpha_i \alpha_j b_{vu} \right), \quad u = 1, \ldots, p-1,$$

$$C_p = (1 + \kappa) \alpha_p^2,$$

with b_{vu} being the elements of the inverse of the matrix $(\beta_{vu})_{v,u=1}^{p-1}$, $\beta_{vv} = \sum_{|i-v|=v} \alpha_i - 1$ and $\beta_{vu} = \sum_{|i-v|=u} \alpha_i$, $v \neq u$. Like in the standard AR(p) model, the autocorrelation function $(\rho_Y(h), h \in \mathbb{N})$ then satisfies

$$\rho_Y(h) = \sum_{i=1}^{p} \alpha_i \rho_Y(|h-i|), \quad h \geq 1. \tag{4}$$

The above conditions simplify if p is small. An NBINARCH(2) model is mean stationary iff $\alpha_1 + \alpha_2 < 1$, and it is second-order stationary iff additionally

$$(1 + \kappa) \left[\alpha_1^2 + \alpha_2^2 + \frac{2\alpha_1^2 \alpha_2}{1 - \alpha_2} \right] < 1.$$

The restrictions for α_1 and α_2 to achieve second-order stationarity are thus more stringent for larger values of κ.

In the simplest interesting case, the second-order stationary NBINARCH(1)-model, the marginal mean, variance and the autocorrelations are

$$\mu = E(Y_t) = \frac{\alpha_0}{1 - \alpha_1},$$

$$\sigma^2 = Var(Y_t) = \frac{\mu + \kappa \mu^2}{1 - \alpha_1^2(1 + \kappa)},$$

$$\rho_Y(h) = \alpha_1^{h-1}, \quad h \geq 1.$$

Note that the model is overdispersed in case of $\alpha_1 > 0$, $E(Y_t) < Var(Y_t)$, even if $\kappa = 0$, imposing a tendency to overestimate κ if autocorrelation is neglected.

2.2 Definition of the estimators

Zhu [28] studies Yule–Walker (YW) estimation of $\alpha_1, \ldots, \alpha_p$ based on the ordinary sample autocovariances in NBINARCH(p) models. He suggests combining them with a method of moments (MoM) estimator of α_0 based on the formula (3) for the marginal mean μ. A moment type estimator of κ [4,8,19] can be obtained solving

$$\frac{1}{n - 2p - 1} \sum_{t=p+1}^{n} \left(\frac{y_t - \mu_t}{\sigma_t} \right)^2 = 1, \tag{5}$$

where $p + 1$ is the dimension of θ. Remember that $\sigma_t^2 = \mu_t + \kappa \mu_t^2$.

We can use the autocorrelation $(\rho_Y(h), h \in \mathbb{N})$ instead of the autocovariance function to write down the Yule–Walker estimators. We plug the sample autocorrelations $\hat{\rho}_Y(1), \ldots, \hat{\rho}_Y(p)$ into the equations (4) for $\rho_Y(1), \ldots, \rho_Y(p)$ and solve for the unknown parameters $\alpha_1, \ldots, \alpha_p$, imposing the basic positivity restriction. A suitable model order p can be identified from the partial autocorrelation function, as it is zero for all lags $h > p$. The partial autocorrelations can be estimated from estimates of the ordinary autocorrelations

using the Durbin-Levinson algorithm, setting negative estimates to 0 because of the positivity restrictions.

Fried et al. [15] find Spearman's rank autocorrelation to work well in case of Poisson INARCH models, being rather efficient but more robust than the sample autocorrelations. We study YW estimation of NBINARCH models, applying Spearman's rank correlation to all pairs (Y_t, Y_{t-h}), $t = h + 1, \ldots, n$, for $h = 1, \ldots, p$, and plugging the resulting autocorrelation estimates $\hat{\rho}_R(h)$, $h = 1, \ldots, p$, into (4).

Suitable score transforms can improve the asymptotical efficiencies of rank based autocorrelation estimators in case of autoregressive models with continuous innovation densities; see e.g. [16]. Let R_t be the rank of Y_t and J a score function, which is often chosen as the quantile function F^{-1} of a suitable distribution function F. Using this notation, the concept of empirical autocorrelation can be applied to the transformed ranks $J(R_t/(n + 1))$, $t = 1, \ldots, n$. The negative binomial distribution functions $F_{\mu,\kappa}$ do not form a location-scale family and there is no linear relationship between the quantiles of different negative binomials. A suitable score function will thus depend on μ and κ, which will be unknown in practice. We investigate rank estimators with negative binomial (NB) scores, applying a stepwise approach: for initialization we calculate estimates $\hat{\mu}, \hat{\kappa}$ of the negative binomial parameters under the simplifying assumption of observing independent data. Then estimate $\alpha_1, \ldots, \alpha_p$ from ranks transformed by the quantile function $J = F^{-1}_{\hat{\mu},\hat{\kappa}}$. Thereafter use these NB scores estimators to estimate α_0 and κ as described below. The process can be iterated using the new estimates of μ and κ to improve on the estimates of $\alpha_1, \ldots, \alpha_p$.

A robust method of moments type estimator of α_0 can be derived by estimating the marginal mean μ for a given value of κ modifying the Huber or the Tukey M-estimator for the Poisson distribution investigated by [5,11], respectively. Setting $\sigma^2 = \mu + \kappa\mu^2$, we can solve

$$\frac{1}{n} \sum_{t=1}^{n} \psi_c \left(\frac{y_t - \mu}{\sigma} \right) = a, \qquad (6)$$

where $a = a(\mu, \kappa) = E\psi_c((Y_1 - \mu)/\sigma)$ is a bias correction to achieve Fisher consistency. In our calculations, we fix the value of κ at an initial estimate and ignore the effects of the temporal dependence in this part of the algorithm for simplicity. Elsaied and Fried [11] suggest the tuning constants $c = 1.8$ for the Huber and $c = 5.5$ for the Tukey M-estimator in case of independent Poisson data. Since negative binomial distributions are more heavy-tailed than the Poisson, we choose larger tuning constants $c \geq 2$ and $c \geq 6$, respectively.

Moreover, we suggest a robustification of the moment type estimator of κ based on equation (5), applying a robust ψ-function and solving

$$\frac{1}{n - 2p - 1} \sum_{t=p+1}^{n} \psi_c^2 \left(\frac{y_t - \mu_t}{\sigma_t} \right) = d, \qquad (7)$$

where d is the expectation of the term on the left hand side. Since our interest in this equation is in a robust estimate of the amount of overdispersion κ, a rather large tuning constant like $c = 10$ can be chosen for the Tukey ψ-function here to restrict the resulting negative bias, setting $d(n, p) = 1$, which is its limiting value as $c \to \infty$.

We combine the ideas presented above as follows, initializing with an a-priori guess of κ (e.g. $\kappa = 0$) and estimating $\alpha_1, \ldots, \alpha_p$ by plugging the first p Spearman rank autocorrelations into the YW equations. In steps 1 and 3 we use the R (R Core Team [21]) function optimise to minimize the squared difference between the terms on the left and the right hand side of the corresponding estimation equation.

0. Estimate $\alpha_1, \ldots, \alpha_p$ using the YW equations and the first p Spearman rank autocorre-
   lations.
1. Estimate μ from (6) using the current estimate of κ.
2. Estimate α_0 from (3) using the current estimates of μ and $\alpha_1, \ldots, \alpha_p$.
3. Estimate κ from (7) using the current estimate of $\theta = (\alpha_0, \ldots, \alpha_p)'$.
4. (Optional) Estimate $\alpha_1, \ldots, \alpha_p$ using the YW equations by plugging in the NB score
   transforms of the first p Spearman rank autocorrelations based on the current estimates
   of μ and κ.

Steps 1.–4. can be iterated until some convergence criterion is met. We found the changes
of the parameter estimates to be very small after the second iteration. We call the estimator
obtained after the first iteration the 1-step NB scores estimator.

   The estimators of μ and κ corresponding to (6) and (7), respectively, can be shown to be
consistent at least for known values of the other parameters, using similar arguments as for the
joint estimator of all model parameters presented in the next section. In the next subsection
we will inspect the performance of the estimators presented here via simulations.

2.3 Simulations

We evaluate the performance of the estimators derived from methods of moments, applying
them to time series of length $n = 200$ and $n = 500$ from several NBINARCH(1) and
NBINARCH(2) models. We consider first order models with parameters

$$(\alpha_0, \alpha_1, \kappa) \in \{(0.55, 0.3, 0.3), (0.55, 0.7, 0.0), (0.55, 0.7, 0.3), (0.55, 0.7, 0.7),$$
$$(1.5, 0.2, 0.0), (1.5, 0.8, 0.2), (1.5, 0.2, 0.8), (1.5, 0.6, 0.8)\}.$$

The two models with $\kappa = 0$ correspond to Poisson INARCH models with low or high lag-one
autocorrelation, $\alpha_1 = 0.2$ or $\alpha_1 = 0.7$. Two other models with $\kappa \in \{0.2, 0.3\}$ correspond to
moderate deviations from the Poisson case, again one of them with low and one with high
lag-one autocorrelation. Four other models with $\kappa \in \{0.7, 0.8\}$ correspond to substantial
deviations from the Poisson case, two of them with low and two with high lag-one autocorre-
lation, combined with a small or a moderately large intercept $\alpha_0 \in \{0.55, 1.5\}$. Varying these
models, we also consider eight second order models with parameters $(\alpha_0, \alpha_1, \alpha_2, \kappa)$ in

$$\{(0.55, 0.7, 0.2, 0.0), (0.55, 0.7, 0.1, 0.3), (0.55, 0.6, 0.1, 0.7), (0.55, 0.3, 0.2, 0.3),$$
$$(1.5, 0.0, 0.2, 0.0), (1.5, 0.5, 0.3, 0.8), (1.5, 0.7, 0.1, 0.2), (1.5, 0.1, 0.2, 0.8)\}.$$

We consider time series of length $n \in \{200, 500\}$ and four different data scenarios for each of
these models, namely clean time series without outliers, time series with 5% additive outliers
of size either 4 marginal standard deviations or 8 marginal standard deviations, rounded to
the nearest integer, at time points chosen independently from a uniform distribution, and time
series with 5% additive outliers occurring as a block at the end of the time series. The latter
scenario resembles the onset of a level shift.

   For each of these 128 combinations of an NBINARCH model, an outlier scenario and a
sample size, we generate 1000 time series and fit NBINARCH(2) models. From the derived
estimates we calculate the empirical mean square errors (MSE) for the estimation of each
of the parameters $\alpha_0, \alpha_1, \alpha_2$ and κ, and take the efficiency (as measured by the ratio of the
MSEs) relatively to the estimator with the smallest MSE for the respective scenario among
those considered here. Then we calculate summary measures like the average mean square
error or the median relative efficiency of an estimator over sets of data scenarios.

Table 1 summarizes the simulation results for the estimation of α_1 and α_2 by showing the average root of the mean square error and the median relative efficiencies separately for the scenarios without outliers, with isolated additive outliers of size 4σ or 8σ, or with a patch of outliers. We set the tuning constant of the Tukey score function to $c = 6$ if not stated otherwise. Results for $c = 7$ are quite similar and not shown here.

The YW estimators based on the ordinary sample autocorrelations offer little advantage as compared to the rank based estimators considered here in terms of efficiency in clean data scenarios, and they are usually inferior in the outlier scenarios. Transforming the ranks by the NB scores improves the efficiency of the rank based YW estimators in case of clean data, but reduces their robustness against outliers to some extent. Similarly, the 1-step NB scores estimator shows somewhat better efficiencies in case of clean data but somewhat larger MSE in case of outliers as compared to Spearman's rank autocorrelation. This difference appears to be smaller for the larger sample size $n = 500$.

The significance of the coefficients of the sample partial autocorrelation function is commonly checked by comparing their absolute values to the 97.5% percentile of the standard normal distribution divided by the square root of the sample size. We apply this rule to the different estimators of α_2 and calculate the percentage of time series for which the estimate of α_2 is found to be significant separately for 1st and 2nd order models. This gives us an idea of the size and the power of such tests based on the different estimators, see Table 1. The rank based tests work similarly well as the one based on the ordinary sample autocorrelations in case of clean data and better in the presence of outliers. Spearman's rank autocorrelation can be recommended as it works well in case of all outlier scenarios considered here if $n = 200$, and it shows the least size distortions in case of $n = 500$ with outliers.

Table 2 shows the results for the estimation of α_0 and κ. Combining the Tukey M-estimator of the marginal mean with Spearman's rank autocorrelation leads to the most stable results for α_0 among the methods considered here, followed by the estimators based on the NB scores transformed ranks. Increasing the tuning constant to $c = 7$ (not shown here) improved the efficiency for α_0 in case of clean data and patchy outliers, but increased the MSE in case of isolated outliers. The iterated NB scores estimator gives the best results for the estimation of κ. Starting the iterated estimator from Spearman's rank autocorrelation (not shown here) leads to rather similar results.

These findings indicate that the partial autocorrelation estimates arising from Spearman's rank autocorrelation can be used to get a first idea about a suitable NBINARCH model order p even in the presence of some outlier contamination. Simple yet quite robust estimates of the autoregressive parameters $\alpha_1, \ldots, \alpha_p$ can then be obtained by applying the Yule–Walker equations to the first p Spearman rank autocorrelations, and of α_0 by combining these estimates $\hat{\alpha}_1, \ldots, \hat{\alpha}_p$ with Tukey's M-estimator of the marginal mean using a tuning constant $c \geq 6$. Finally, estimate κ solving Eq. (7) using the estimates $\hat{\mu}_t = \hat{\alpha}_0 + \hat{\alpha}_1 y_{t-1} + \cdots + \hat{\alpha}_p y_{t-p}$ and $\hat{\sigma}_t = \hat{\mu}_t + \kappa \hat{\mu}_t^2$.

3 Joint M-estimation

In Sect. 2 we have seen that rank based autocorrelation estimators perform similarly well as ordinary sample autocorrelations in terms of efficiency in NBINARCH models, while being more robust. Now we investigate whether the efficiency and robustness of these simple estimators can be improved further if we robustify the conditional maximum likelihood estimators for a given model order p. We expect substantial improvements in particular concerning the

Table 1 Results for Yule–Walker type estimators of α_1 and α_2 if $n = 200$ (left) or $n = 500$ (right): ordinary sample acf (SC), rank-based Spearman acf (RC), NB scores based on Tukey M-estimation of (μ, κ) with tuning constant c (NB0$_c$) and after 1 iteration (NB1$_c$)

	SC	RC	NB0$_6$	NB1$_6$	SC	RC	NB0$_6$	NB1$_6$
	Estimation of α_1: ARMSE							
Clean	11.8	13.6	12.0	11.1	8.8	11.8	9.5	8.1
AO4	32.7	20.9	24.7	25.9	29.6	19.4	23.0	23.9
AO8	44.0	21.1	26.5	28.0	42.3	19.9	25.4	26.8
PO8	31.3	11.0	11.5	13.3	33.5	9.3	11.1	13.7
	Estimation of α_1: MedEff							
Clean	84.5	68.4	88.8	95.4	81.5	49.9	78.4	91.7
AO4	51.6	100.0	79.8	72.9	50.0	100.0	77.1	70.1
AO8	29.1	100.0	67.4	63.1	25.1	100.0	62.9	59.9
PO8	8.6	49.2	39.3	66.7	3.7	60.0	37.6	59.2
	Estimation of α_2: ARMSE							
Clean	8.0	7.1	6.9	7.0	7.0	6.0	5.8	5.9
AO4	9.8	7.8	8.6	9.0	8.9	6.6	7.5	8.0
AO8	11.0	7.9	8.9	9.3	10.0	6.7	7.9	8.4
PO8	9.9	7.8	8.8	9.1	9.4	7.1	9.0	9.4
	Estimation of α_2: MedEff							
Clean	78.6	86.8	92.5	90.8	77.2	83.6	89.9	90.1
AO4	68.1	100.0	88.7	81.4	55.5	100.0	83.0	75.8
AO8	58.7	100.0	79.7	75.9	55.0	100.0	71.4	66.0
PO8	100.0	76.3	54.1	48.5	82.3	66.0	42.7	37.5
	Test of $\alpha_2 = 0$: size							
Clean	1.3	1.0	0.8	0.8	1.9	1.3	1.4	1.1
AO4	13.6	5.1	9.7	11.3	36.3	15.6	30.4	34.2
AO8	10.0	5.1	11.1	12.7	23.1	16.6	31.3	35.1
PO8	1.3	4.3	9.6	10.3	25.1	16.7	38.8	37.6
	Test of $\alpha_2 = 0$: power							
Clean	33.6	34.2	35.9	34.8	52.3	58.1	58.6	56.5
AO4	39.2	40.9	45.2	44.2	79.0	83.4	87.0	85.9
AO8	24.3	41.5	45.8	44.6	49.4	83.8	84.7	83.5
PO8	22.1	43.2	45.1	42.0	41.2	62.8	65.1	61.3

Results for clean data (Clean), 5% additive outliers or 5% patchy outliers of size s marginal standard deviations (AOs or POs). Average root of the mean square error (ARMSE, multiplied by 100), median relative efficiency MedEff, empirical size and empirical power for the respective scenarios

estimation of α_0, as the simple method of moment type estimator $\hat{\mu}/(1 - \hat{\alpha}_1 - \cdots - \hat{\alpha}_p)$ suffers from the non-linear effects of estimation errors in $\hat{\alpha}_1, \ldots, \hat{\alpha}_p$, especially if the sum of these parameters is close to 1. This can possibly be improved by joint estimation of $\alpha_0, \alpha_1, \ldots, \alpha_p$ as considered in the following but for the price of larger computational costs. We focus on first order models for the reason of simplicity. After reviewing conditional maximum likelihood estimation we summarize the estimation approach of [27], before explaining ours. Then we compare the estimators in a simulation study, including also some of the estimators which showed the best performance in Sect. 2.

Table 2 Results for method of moments type estimators of α_0 and κ if $n = 200$ (left) or $n = 500$ (right): ordinary moments and sample autocorrelations (MoM), Tukey M-estimator with Spearman acf (TR), Tukey M-estimator and NB scores estimation of (μ, κ) with tuning constant c (NB0$_c$) and after 1 iteration (NB1$_c$)

	MoM	TR	NB0$_6$	NB1$_6$	MoM	TR	NB0$_6$	NB1$_6$
	Estimation of α_0: ARMSE							
Clean	49.9	25.8	29.2	26.2	0.93	0.52	0.50	0.60
AO4	178.1	36.2	49.9	64.3	4.61	0.76	0.91	1.36
AO8	334.2	36.5	54.9	69.2	13.56	0.74	0.99	1.42
PO8	60.9	32.5	41.4	40.1	0.49	0.36	0.28	0.28
	Estimation of α_0: MedEff							
Clean	58.7	96.3	75.1	80.3	59.5	94.1	99.4	91.6
AO4	7.5	100.0	66.2	28.2	22.5	100.0	94.4	56.7
AO8	2.0	100.0	62.2	35.6	6.4	100.0	85.0	58.3
PO8	33.7	76.5	43.8	43.7	97.3	75.7	89.0	86.6
	Estimation of κ: ARMSE							
Clean	41.0	49.9	48.8	16.2	0.34	0.46	0.44	0.28
AO4	53.3	49.9	49.6	20.2	0.38	0.46	0.45	0.27
AO8	66.3	49.9	49.4	17.4	0.45	0.46	0.45	0.27
PO8	66.2	49.9	70.2	14.0	0.45	0.46	0.45	0.27
	Estimation of κ: MedEff							
Clean	20.3	14.3	13.0	88.4	84.9	48.1	54.2	94.6
AO4	3.5	14.0	13.7	68.1	65.1	40.8	40.9	84.5
AO8	6.2	11.5	11.7	75.7	52.9	49.1	50.4	95.5
PO8	4.1	12.2	9.5	87.8	54.1	44.8	44.9	92.8

Results for clean data (Clean), 5% additive outliers or 5% patchy outliers of size s marginal standard deviations (AOs or POs). Average root of the mean square error (ARMSE, multiplied by 100) and median relative efficiency MedEff

3.1 Conditional maximum likelihood estimation

The score equations of the conditional maximum likelihood estimator (CMLE) for the parameter vector $\omega = (\theta', \kappa)'$, with $\theta = (\alpha_0, \ldots, \alpha_p)'$, are

$$\frac{\partial \ell(\omega)}{\partial \omega} = \left(\sum_{t=1}^{n} S'_{t,\theta}(\omega), \sum_{t=1}^{n} S_{t,\kappa}(\omega) \right)' = 0, \text{ where}$$

$$S_{t,\theta}(\omega) = \frac{y_t - \mu_t}{\mu_t + \kappa \mu_t^2} \frac{\partial \mu_t}{\partial \theta},$$

$$S_{t,\kappa}(\omega) = -\kappa^{-2} \left(\frac{\partial \ln \Gamma(y_t + 1/\kappa)}{\partial (y_t + 1/\kappa)} - \frac{\partial \ln \Gamma(1/\kappa)}{\partial \kappa} - \ln(\kappa \mu_t + 1) - \frac{\kappa(y_t - \mu_t)}{\kappa \mu_t + 1} \right)$$

$$= -\frac{1}{\kappa^2} \left(G(y_t + 1/\kappa) - G(1/\kappa) - \ln(\kappa \mu_t + 1) - \frac{\kappa(y_t - \mu_t)}{\kappa \mu_t + 1} \right),$$

$$\frac{\partial \mu_t}{\partial \theta} = \left(1, y_{t-1}, \ldots, y_{t-p} \right)' = \tilde{X}_t.$$

Hereby, $G(u)$ denotes the digamma function, $G(u) = \partial \ln \Gamma(u)/\partial u$. Solving the score equations can be initialized using the Poisson quasi-likelihood estimator of θ, which corresponds to the assumption $\kappa = 0$, and the moment estimator of κ defined in (5).

3.2 M-estimation using multivariate outlyingness

Xiong and Zhu [27] robustify the first of the above score equations of the CML estimator using the Pearson residuals $r_t = (y_t - \mu_t)/\sigma_t$ and Mallow's quasi-likelihood estimator proposed by Cantoni and Ronchetti (2001),

$$\sum_{t=p+1}^{n} \left[\frac{\psi(r_t) w_t}{\sigma_t} \frac{\partial \mu_t}{\partial \theta} - a_t(\theta) \right] = 0,$$

where $a_t(\theta) = E\left[\psi(r_t) w_t (\sigma_t)^{-1} \partial \mu_t/\partial \theta \,\middle|\, \mathcal{F}_{t-1}\right]$ is a bias correction to achieve Fisher consistency. They consider weights $w_t = \sqrt{1 - h_{t,t}}$ based on the diagonal elements $h_{t,t}$ of the hat matrix $H = X(X'X)^{-1}X'$, with $X = (\tilde{X}_1, \ldots, \tilde{X}_{n-p})'$, as well as hard rejection weights $w_t = I(D_t^2 \leq \chi_p^2(0.95))$ based on the squared Mahalanobis distances D_t^2 measuring multivariate outlyingness.

For estimation of κ, they suggest the weighted maximum likelihood estimator proposed by [1], using the estimation equation

$$0 = \sum_{t=p+1}^{n} \left[\frac{\psi(r_t)}{r_t} S_{t,\kappa}(\omega) w_t - b_t(\kappa) \right]$$

with another Fisher consistency correction $b_t(\kappa) = E(S_{t,\kappa} w_t \psi(r_t)/r_t | \mathcal{F}_{t-1})$.

Xiong and Zhu [27] perform a simulation study for time series consisting of $n = 1000$ observations without, or with up to 40 isolated, or a patch of 52 subsequent outliers. Based on this they recommend hard rejection weights using the minimum covariance determinant (MCD) estimator [25] in combination with Tukey's ψ-function for both, the estimation of θ and κ, using constants $c_1 = 7$ and $c_2 = 6$, respectively.

Using multivariate outlyingness allows taking positive correlations among the regressors into account, but it may downweight the contribution of many observations substantially although just one of its regressor components looks suspicious; see the R-package cellWise [22], and the references cited therein for related discussions. Moreover, the MCD has been designed for (continuous) unimodal elliptically symmetric but not for discrete data, see e.g. [17]. Indeed, methods based on subsets of the observations like the MCD may have problems with many identical values, as they can occur in count data; see Duerre et al. (2015) for a discussion in the context of robust autocorrelation estimation. These are possible drawbacks of the method by Xiong and Zhu.

3.3 M-estimation using componentwise shrinking

We propose another variant of robust M-estimation for NBINARCH(p) models, combining the M-estimation approaches developed by [10] for the limiting Poisson INARCH model and by [1] for the negative binomial regression model.

We robustify the CML estimator along the same lines as in [10] in the Poisson case, applying M-estimation to the Pearson residuals $r_t = (y_t - \mu_t)/\sigma_t$,

$$\sum_{t=p+1}^{n} s_{t,0}(y_t, \omega) = \sum_{t=p+1}^{n} \left(\frac{\psi(r_t)}{\sigma_t} - c_{t,0} \right) = 0,$$

$$\sum_{t=p+1}^{n} s_{t,i}(y_t, \omega) = \sum_{t=p+1}^{n} \left(\frac{\psi(r_t)}{\sigma_t} \left[\sigma \psi \left(\frac{y_{t-i} - \mu}{\sigma} \right) + \mu \right] - c_{t,i} \right) = 0, \quad i = 1, \dots, p.$$

Hereby, we use bias corrections $c_{t,0} = c_{t,0}(\omega) = E_\omega(\psi((Y_t - \mu_t)/\sigma_t)/\sigma_t | \mathcal{F}_{t-1})$ and $c_{t,i} = c_{t,i}(\omega) = E_\omega \left(\psi((Y_t - \mu_t)/\sigma_t) \left[\sigma \psi \left(\frac{Y_{t-i} - \mu}{\sigma} \right) + \mu \right] / \sigma_t | \mathcal{F}_{t-1} \right)$, $i = 1, \dots, p$. While [27] downweight the whole vector $S_{t,\theta}(\omega)$ in the score equation according to a measure of the multivariate outlyingness of the corresponding regressor values, including the component for the intercept, we shrink each regressor towards the center of its distribution. Note that replacing $c_{t,i}(\omega)$ by its expectation $c_i(\omega) = E(c_{t,i}(\omega))$ is simpler and asymptotically equivalent because of the general ergodic theorem [18] and Slutsky's Lemma.

For estimation of κ we also make use of the M-estimator of [1]. To obtain starting values for an iterative scheme to solve these equations, [1] recommend plugging in the estimates resulting from Poisson M-estimation into the equation for κ and then iterate. Alternatively, we can use the estimator of κ proposed in the previous section.

In the Appendix we prove the consistency of the joint estimator $\hat{\omega}_n = (\hat{\theta}_n', \hat{\kappa}_n)'$ and conjecture its asymptotic normality,

$$\sqrt{n}(\hat{\omega}_n - \omega^{(0)}) \to N\left(0, A(\omega^{(0)})^{-1} B(\omega^{(0)}) A(\omega^{(0)})^{-1} \right),$$

with $A(\omega^{(0)}) = E(\partial s_t(Y_t, \omega)/\partial \omega)_{\omega^{(0)}}$, $B(\omega^{(0)}) = E(s_t(Y_t, \omega^{(0)}) \times s_t(Y_t, \omega^{(0)})')$, $s_t(y_t, \omega) = (s_{t,0}(y_t, \omega), \dots, s_{t,p}(y_t, \omega))'$, and $\omega^{(0)} = (\theta^{(0)'}, \kappa^{(0)})'$ being the true parameter vector. The matrices $A(\omega^{(0)})$ and $B(\omega^{(0)})$ can be estimated using their empirical counterparts as in [10], replacing the unknown parameters by their robust estimations and the expectations by the corresponding averages across the realizations; the arising estimate inherits some robustness when using a bounded ψ-function with a bounded derivative.

3.4 Simulations

We perform some simulations to compare the performance of the CML estimator, the robust method of moments estimator based on Spearman's rank autocorrelation combined with the Tukey M-estimator of the marginal mean, the 1-step NB scores transformed rank estimator, as well as our Tukey M-estimator and [27] M-estimators using the hat matrix or the MCD. We consider the performance in scenarios without outliers as well as with an increasing number of isolated or patchy additive outliers with the same size. We focus on scenarios similar to those inspected by [27] with time series of length $n = 1000$. This is in part due to problems when running their algorithms with the recommended MCD for shorter series lengths n, presumably caused by the discreteness of the data. We applied several estimators for initialization of our joint M-estimator, namely the estimators presented in Sect. 2, in particular that using the ordinary Spearman rank autocorrelations, and the robust Poisson INARCH estimator of [10] assuming $\kappa = 0$. We can combine these possibilities to check for the possibility of multiple roots. The solutions obtained by the different initializations in the following simulations can usually be regarded as identical.
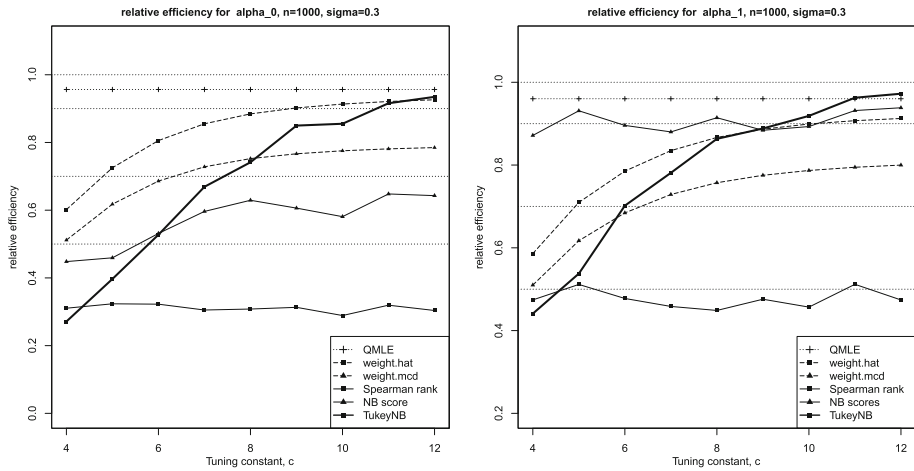
Fig. 1 Simulated efficiencies for α_0 (left) and α_1 (right) as a function of the tuning constant in case of $n = 1000$ data points from an NBINARCH(1) process with parameters $\alpha_0 = 0.55$, $\alpha_1 = 0.4$ and $\kappa = 0.3$

3.4.1 Scenarios without outliers

First we inspect the efficiency in case of data without outliers, generating 1000 time series of length $n = 1000$ from an NBINARCH(1) model with parameter $\omega = (0.55, 0.4, 0.3)$, see Fig. 1. The initial method of moments estimators achieve about 50% efficiency for α_1 and about 30% efficiency for α_0. The 1-step NB scores rank estimator improves on this as its efficiencies are about 90% for α_1 and 60% for α_0. The estimator of [27] with a recommended tuning constant c about 6 or 7 achieves an efficiency of more than 70% using hard rejection based on the MCD and even about 80% efficiency when using soft rejection based on the hat matrix. Our estimator shows a stronger dependence on c. It achieves high efficiencies of about 90% when using $c = 10$, like Xiong and Zhu's method, while for $c = 7$ it is somewhat less efficient than their estimator.

Figure 2 compares normal qq-plots of our joint M-estimator with $c = 10$ and of the conditional maximum likelihood estimator in case of time series of length $n = 200, n = 500$ or $n = 1000$. The finite-sample distributions of both estimators are right-skewed for α_0 if $n = 200$ but reasonably close to normality for the larger sample sizes, with little difference between the estimators.

3.4.2 Scenarios with an increasing number of isolated outliers

Now we analyze the behavior of the estimators in the presence of outliers. In a first experiment we generate a time series of length $n = 300$ from an NBINARCH(1) model with parameters $\alpha_0 = 0.55$, $\alpha_1 = 0.4$ and $\kappa = 0.3$ and include a single additive outlier of increasing size. Figure 3 shows the arising sensitivity curves of several estimators for estimation of α_0 and α_1, corresponding to the difference between the estimated value with and without this outlier, multiplied by n. Except for the CMLE, the other estimators show bounded sensitivities, with the effects being largest for the 1-step NB score estimator, followed by our joint M-estimator.

Next we include an increasing number of 4,8,…,40 isolated additive outliers of size 5 or 10 at positions chosen at random, analyzing 500 time series for each scenario. Figure 4 depicts the resulting biases as a function of the number of outliers. The lack of robustness of
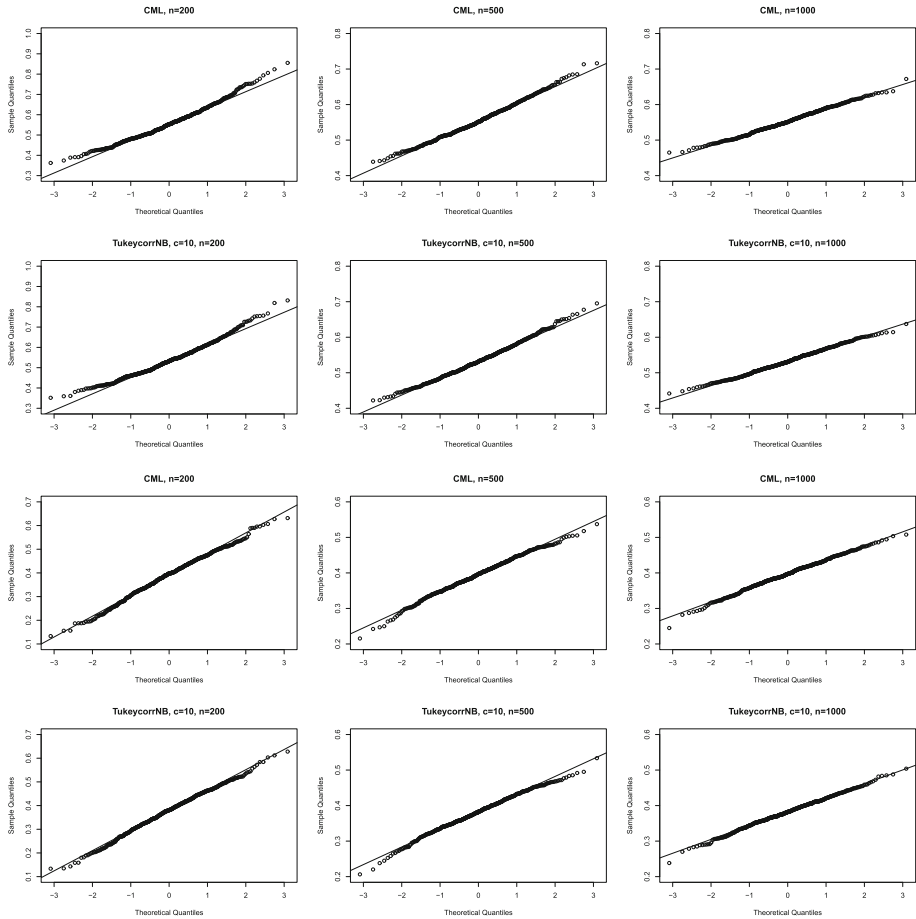
Fig. 2 Normal qq-plots for the CML (upper panel) and our joint Tukey M-estimator (lower panel) of α_0 (top) and α_1 (bottom) in case of $n = 200$ (left), $n = 500$ (middle) or $n = 1000$ (right) data points from an NBINARCH(1) process with parameters $\alpha_0 = 0.55$, $\alpha_1 = 0.4$ and $\kappa = 0.3$

the CMLE and the Poisson quasi-likelihood estimator manifests itself in an increasingly large positive (negative) bias for α_0 (α_1), particularly for the larger outlier size. Spearman's rank autocorrelation is negatively biased for α_1, and the resulting estimator for α_0 thus positively biased, which explains its moderately large efficiency for large sample sizes observed before. The 1-step NBscore transform estimator reduces this bias in case of clean data but is more affected by the outliers. Our Tukey M-estimator shows some bias for the estimation of α_1 in case of the smaller outlier size considered here, but performs better than the other estimators otherwise, except for M-estimator with hard-thresholding based on the MCD [27]. These two estimators perform similarly well although we have chosen a larger tuning constant $c = 10$ for our estimator to obtain a high efficiency in case of clean data, instead of the $c = 6$ recommended by Xiong and Zhu for their estimator. Qualitatively similar results have been obtained for time series with parameters $\alpha_0 = 1$, $\alpha_1 = 0.6$ and $\kappa = 0.3$ (not shown here).
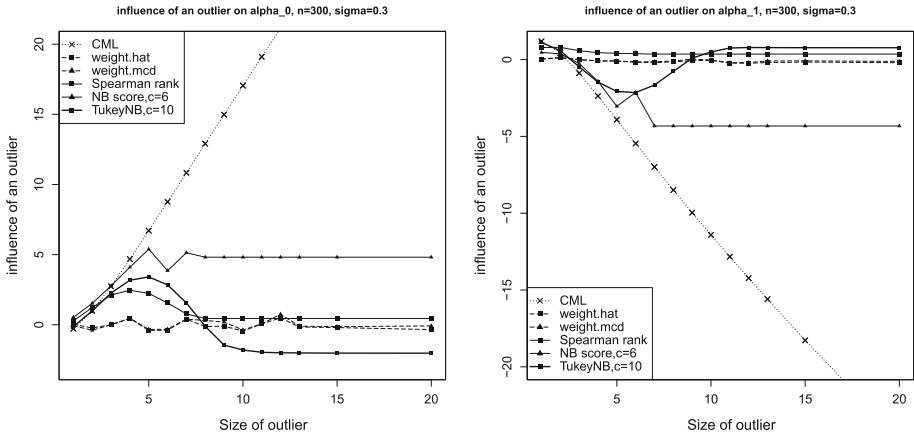
Fig. 3 Sensitivities of the estimators of α_0 (left) and α_1 (right) against a single additive outlier of increasing size in case of data generated from an NBINARCH(1) process with parameters $\alpha_0 = 0.55$, $\alpha_1 = 0.4$ and $\kappa = 0.3$

3.5 Scenarios with a patch of additive outliers

We also investigate scenarios with a patch of consecutive additive outliers of the same size, which can arise for instance because of a temporary disturbance or a temporary level shift. Figure 5 depicts the results obtained for an increasing number of 4, 8, …, 52 additive outliers of identical size 5 or 10 starting at time point 250 in NBINARCH(1) time series with parameters $\alpha_0 = 0.55$, $\alpha_1 = 0.4$, $\kappa = 0.3$. A patch of consecutive outliers imposes a positive bias on non-robust estimators of α_1, which can then lead to a negative bias for the estimation of α_0. This can be seen for the 1-step NB scores rank estimator, which shows a similarly strong bias for α_1 as the QMLE and the CML here. Estimation using Spearman's rank autocorrelation performs better than this in terms of bias and similar to our Tukey M-estimator with a tuning constant $c = 10$, which is large relatively to the height of the shift. Application of the latter with a smaller tuning constant $c = 6$ resists these outliers better and similarly well as the M-estimator with hard rejection proposed by [27] in the scenarios considered here. The version of their estimator applying soft-thresholding based on the hat matrix shows considerable problems for the estimation of the intercept here, and this also applies to the version using the MCD in case of the smaller outlier size.

4 Data example

For illustration we analyze the number of campylobacterosis infections from January 1990 to the end of October 2000 in the north of the Province of Québek, Canada. There is one observation every 4 weeks, that is 13 observations per year. [12] used this data set shown in Fig. 6 to exemplify the INGARCH model with a conditional Poisson distribution. Under the same basic model assumption, [13] found a possible level shift at time $t = 84$ and a so-called spiky outlier at time $t = 100$ by applying an iterative procedure for the detection and elimination of different types of intervention effects. Later on, [20] found evidence that the conditional distribution might be better described by a negative binomial than by a Poisson distribution, albeit their estimate of the overdispersion parameter was small ($\hat{\kappa} \approx 0.0297$). In
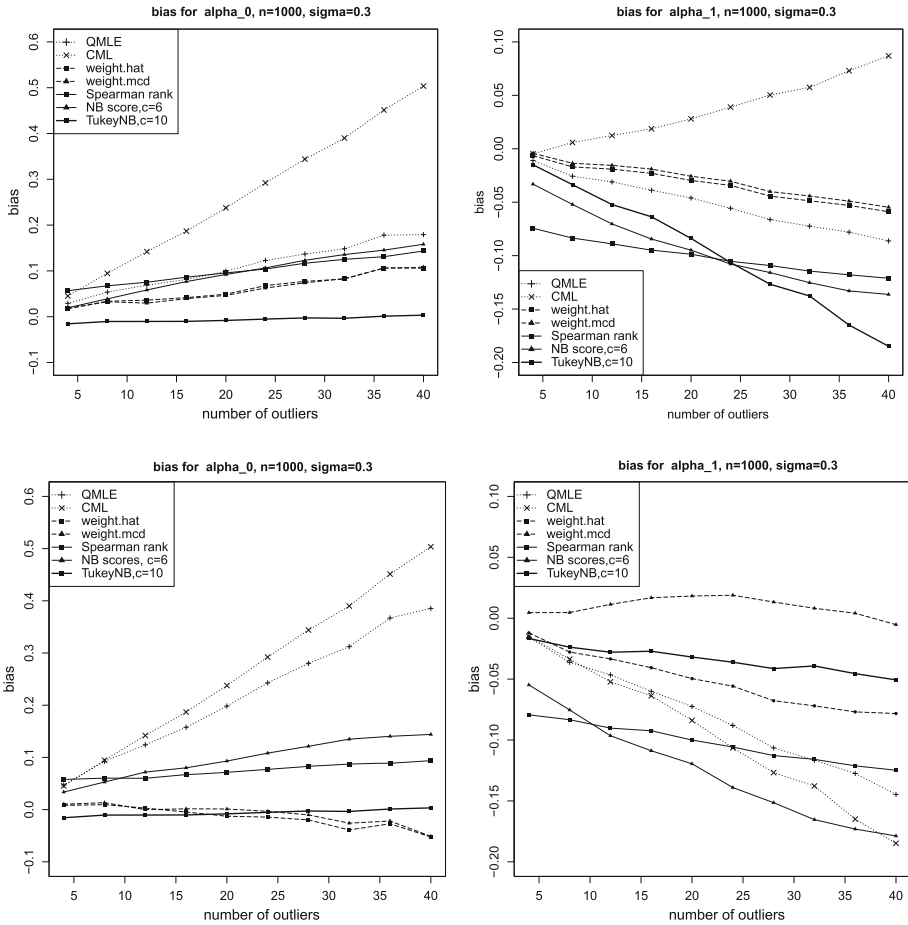
Fig. 4 Simulated biases for α_0 (left) and α_1 (right) in case of an increasing number of 4,8,…,20,24,…,36,40 isolated additive outliers of size 5 (top) or 10 (bottom) at randomly chosen positions in time series of length $n = 1000$ generated from an NBINARCH(1) process with parameters $\alpha_0 = 0.55$, $\alpha_1 = 0.4$ and $\kappa = 0.3$

consequence, the findings of [13] might have been influenced by using the wrong conditional distribution, and in turn the findings of [20] might be in part due to extraordinary effects in these data, which might have been estimated incorrectly or even been missed. We re-analyze these data using robust methods to get additional insights. Further work is needed to include structural changes or covariate information into the methods presented in Sect. 3. We instead concentrate on the methods presented in Sect. 2.2 as they can be modified easily to account for a level shift:

1. First we want to get an idea about suitable model orders, taking a possible level shift after observation number $t = 84$ into account, i.e., in the middle of the seventh year of the observation period. For this we calculate Spearman's rank autocorrelation and also the ordinary sample autocorrelation from the first six years of data. Both types of autocorrelation point at the suitability of an INARCH(1) model without a seasonal effect, with the estimates of α_1 being about 0.44 and thus substantially larger than their asymptotical standard error under the assumption of white noise, which is about 0.11. Calculating the ordinary sample autocorrelation at lag 1 or the conditional maximum
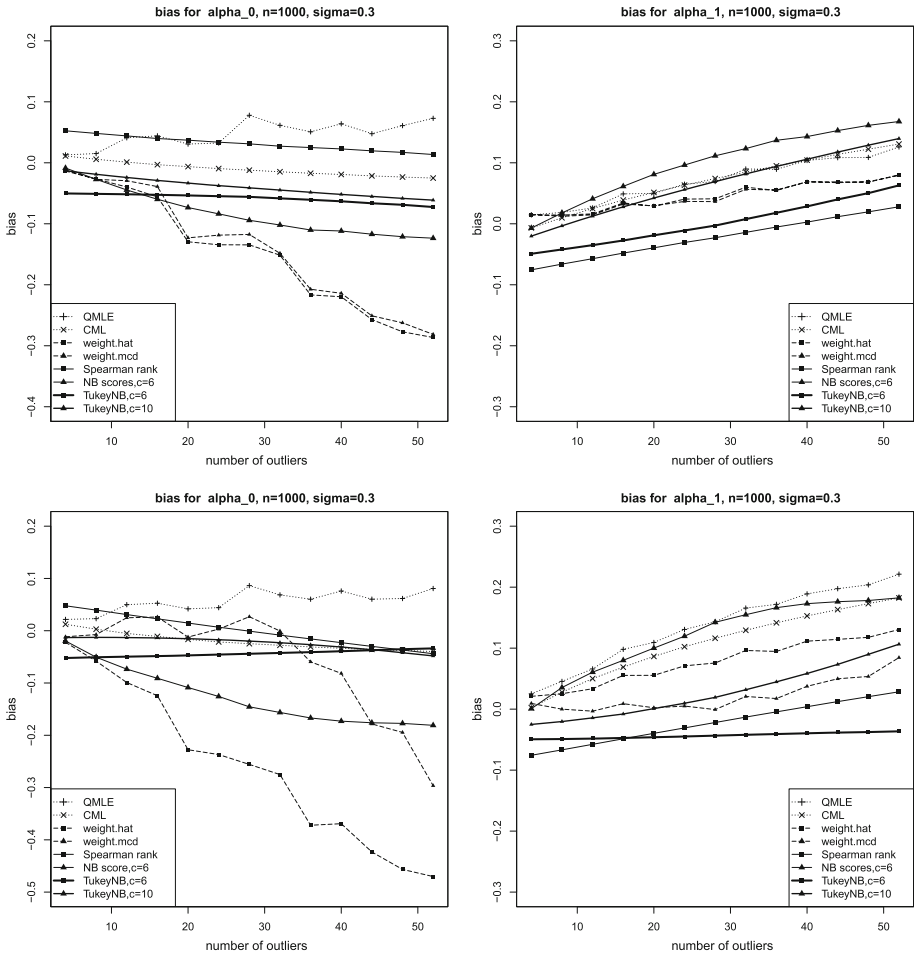
Fig. 5 Simulated biases for α_0 (left) and α_1 (right) in case of an outlier patch consisting of an increasing number of 4,8,…,52 additive outliers of size 5 (top) or 10 (bottom) in time series of length $n = 1000$ generated from an NBINARCH(1) process with parameters $\alpha_0 = 0.55$, $\alpha_1 = 0.4$ and $\kappa = 0.3$ (top) or $\alpha_0 = 1$, $\alpha_1 = 0.6$ and $\kappa = 0.3$ (bottom)

likelihood estimate of α_1 from the full data set leads to a much larger value than this, namely about 0.64, which is just within the range of a 95% confidence interval based on the first six years.

2. Second we fit an INARCH(1) model with a level shift after $t = 84$ to the data, i.e., we allow for a change of the intercept α_0. In order to make use of the full data set in spite of a possible level shift, we split the time series into several non-overlapping subsequences and estimate the parameters of interest from each of them separately. This approach has been analysed by [23] for estimation of the Hurst parameter and by [3] for estimation of the variance. Here, we split the time series into five subsequences of length 28 each, which fits well to the total sample size of 140 and also to a possible level shift right at the end of the third subsequence. Due to the robustness properties of ranks we expect the effect of a single spiky outlier on the estimate to be small and take the average of the estimates obtained from the five subsequences as our final estimate. The average lag
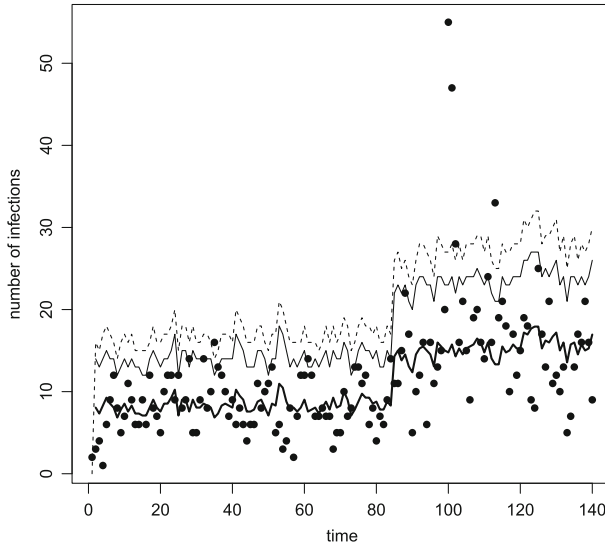
Fig. 6 Number of campylobacterosis infections (dots), conditional mean fitted by robustified method of moments (bold solid line), and 95% as well as 99% percentiles of the conditional negative binomial distributions (solid line and dashed line)

one Spearman rank autocorrelation is $\hat{\alpha}_1 = 0.368$, which is again large as compared to the asymptotic standard error 0.085 under white noise and very close to the value 0.369 obtained by [20] using an iterative procedure for detecting and modelling different types of intervention effects. The robust Tukey type method of moments estimate of α_0 derived from the first three subsequences is 5.27, and the estimate of the change in intercept because of a level shift, derived from the difference with the estimate calculated from the fourth and fifth subsequence, is 4.20. A parametric bootstrap confidence interval for the height of the shift is [3.04, 6.69].

3. It remains to estimate κ, taking the conditional mean function $\mu_t = 5.27 + 4.20I(t > 84) + 0.368y_{t-1}$ derived in step 2 into account. The robustified method of moments type estimate with the tuning constant $c = 10$ is $\tilde{\kappa} = 0.0179$. Application of a parametric bootstrap, generating 200 time series from a Poisson INARCH model with the same conditional mean function, leads to a bootstrap p-value of 0.01, so that this estimate is small but significantly different from 0. In another parametric bootstrap we generate 1000 time series for each of several INARCH(1) models, all with the conditional mean function estimated for the real data but different values of $\kappa = 0.001, \ldots, 0.1$. In a bootstrap confidence interval we include all values of κ for which the estimate $\tilde{\kappa} = 0.0179$ lies between the 2.5% and the 97.5% percentile of the estimates obtained from the corresponding simulated data sets, resulting in the interval [0.006, 0.093]. For a larger tuning constant $c = 12$ we get a larger estimate $\tilde{\kappa} = 0.0303$, which is very close to the estimate 0.0297 obtained by [20] applying a sequential outlier detection and modelling procedure.

Our robust method of moment type estimator $\tilde{\kappa}$ of κ is obviously biased, with a bias which is larger for smaller tuning constants c. The smaller estimate obtained for $c = 10$ might thus be due to a negative bias, but the difference to the less robust estimates could also be due to outlier effects. To correct the bias of $\tilde{\kappa}$ we can multiply it with a correction

factor $f_c(\theta, \kappa)$, which will likely depend on the true parameters $\theta = (\alpha_0, \alpha_1)$ and κ as well. To derive suitable finite sample correction factors we generate 1000 NBINARCH(1) time series with the same length $n = 140$, the same parameters and the same level shift as found for the real data, and estimate the value of κ for each of them. Then we set $f_c(\tilde{\theta}, \tilde{\kappa})$ to the ratio between the value of κ, from which these artificial data sets have been generated, and the average of these estimates. Multiplication of $\tilde{\kappa}$ with this factor gives a new bias-reduced estimate. Since $f_c(\tilde{\theta}, \tilde{\kappa})$ only approximates the factor $f_c(\theta, \kappa)$ needed, we can iterate this process until it stabilizes after a few steps. This leads us to an estimate of about 0.04, a little larger than those reported before but well within the bootstrap CI.

Overall our analysis confirms that a NBINGARCH model describes this data set well after taking a possible level shift at observation 84 and an outlier at time 100 into account. Figure 6 illustrates that all observations except for a few ones at or right after time point 100 are below or at least close to the respective 95% percentile of the fitted 1-step ahead predictive distribution. The observations at $t = 100$ and $t = 101$ are far outlying as compared even to the 99% percentile. Note that the differences between these percentiles and those obtained by a Poisson INARCH(1) model with the same conditional mean sequence is at most 2 in case of the 95% and at most 3 in case of the 99% percentile, so that a Poisson model suits most purposes here.

5 Conclusions

We have proposed robustifications of method of moments and ML-estimation for fitting INARCH models with conditional negative binomial distributions. The former avoid multi-dimensional optimizations and are useful for getting information about a suitable model and as initial estimates for calculation of the more efficient and robust joint M-estimators of the autoregressive parameters.

Our proposal of negative binomial scores increases the efficiency of rank based autocorrelation and partial autocorrelation estimators in the NBINARCH model substantially but reduces its robustness. Similar findings are known for van der Waerden scores under Gaussian model assumptions. If robustness against outliers is important, we prefer the ordinary Spearman's rank (partial) autocorrelation. Our joint M-estimation approach is an alternative to the one suggested by [27]. Our method seems to be computationally simpler, so that it can be applied to moderately large data sets, where we encountered problems for the Xiong and Zhu's algorithm. A drawback of our approach is that the choice of tuning constants to achieve both high efficiency and high robustness against different outlier scenarios (like isolated and patchy outliers) seems to be more difficult. Nevertheless, in our experiments it achieves better efficiency and robustness than the Yule–Walker estimates based on Spearman's rank autocorrelation if the model order is known.

For illustration we have analyzed a famous data set from the literature, the campylobacterosis data studied e.g. by [12]. In doing so we have modified the robust method of moment estimators such that patterns of change like a level shift found in these data by several authors (e.g. [13]) can be incorporated in the estimation. The results obtained agree well with those obtained by [20] applying sophisticated stepwise detection and correction procedures.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Data availability The real data analyzed here are included in the R-package tscount which is available at https://cran.r-project.org/.

Code availability The authors would like to thank Dr. Zhu for sharing the programme code for his estimator. The programme code for the other methods developed here is available from the authors upon request.

Appendix

In the following we establish consistency and asymptotic normality of our joint M-estimator. For consistency we check the assumptions of Theorem 4.1.2 in [2], for asymptotic normality we additionally use the second part of Theorem 3.2.23 in [26]. We use the following notation, setting $Y = (Y_1, \ldots, Y_n)'$, $y = (y_1, \ldots, y_n)'$ and $r_t = (y_t - \mu_t)/\sigma_t$:

$$Q_n(y, \omega) = -S_n(y, \omega)' S_n(y, \omega)/n,$$

$$S_n(y, \omega) = \sum_{t=p+1}^{n} s_t(y, \omega),$$

$$s_t(y, \omega) = (s_{t,0}(y, \omega), \ldots, s_{t,p}(y, \omega), u_t(y, \omega))',$$

$$s_{t,0}(y, \omega) = \frac{\psi(r_t)}{\sigma_t} - c_{t,0}(\omega),$$

$$s_{t,i}(y, \omega) = \frac{\psi(r_t)}{\sigma_t} \left[\sigma \psi \left(\frac{y_{t-i} - \mu}{\sigma} \right) + \mu \right] - c_{t,i}(\omega), \ i = 1, \ldots, p,$$

$$u_t(y, \omega) = \frac{\psi(r_t)}{r_t} S_{t,\kappa}(\omega) w_t - b_t(\omega)$$

with $c_{t,i}(\omega)$ being the conditional expectation of the term before with respect to the σ-field $\mathcal{F}_{t-1}$ of the respective past, calculated assuming that ω is the true parameter value. Note that a value $\hat{\omega}_n$ with $S_n(y, \hat{\omega}_n) = 0$ is a global maximizer of $Q_n(y, \omega)$.

For weak consistency of a sequence of solutions $(\hat{\omega}_n : n \in \mathbb{N})$, we assume for the true parameter value $\omega^{(0)} = (\alpha_0^{(0)}, \ldots, \alpha_p^{(0)}, \kappa^{(0)})' \in \Theta$, which is an open bounded subset of $\mathbb{R}_+^{p+1}$ with a strictly positive lower bound for the first coordinate to keep the mean and the variance away from zero. $Q_n(y, \omega)$ is a measurable function of y, and its gradient vector $\partial Q_n(y, \omega)/\partial \omega$ exists and is continuous in case of differentiable ψ-functions like the one of Tukey. In case of the Huber function, we can use Theorem 4.1.1 of [2] because of the uniqueness of the solution, avoiding the need of a continuous derivative.

Part a) of the following Lemma 1 and the continuous mapping theorem imply that $Q_n(Y, \omega)/n$ converges to a nonstochastic limit $Q(\omega) = -E(s_t(Y, \omega)'s_t(Y, \omega))$ in probability. Uniformity of the convergence of $Q_n(Y, \omega)/n - Q(\omega)$ to 0 can be seen applying Theorem 4.2.1 of [2] to

$$g(Y; \omega) = s_t(Y, \omega)'s_t(Y, \omega) - E(s_t(Y, \omega)'s_t(Y, \omega)).$$

The condition $E \sup_{\omega \in \Theta} |g(Y; \omega)| < \infty$ is obviously fulfilled,

$$E \sup_{\omega \in \Theta} |g(Y; \omega)| \leq 2E \sup_{\omega \in \Theta} |s_t(Y, \omega)'s_t(Y, \omega)| < \infty,$$

because $|s_t(Y, \omega)'s_t(Y, \omega)|$ is bounded with probability 1 in view of the boundedness of the ψ-function and the parameter space.

Since $Q(\omega) \leq 0 = Q(\omega^{(0)})$ by construction, there is a global maximum at $\omega^{(0)}$, so that $\hat{\omega}_n$ is consistent if it is initialized by a consistent estimator.

For asymptotic normality, note that $\partial S_n(y, \omega)/\partial \omega$ exists and is continuous in an open convex neighborhood of $\omega^{(0)}$ if we use a differentiable ψ-function, since the other terms μ_t, σ_t, μ and σ involved in the construction of $S_n(y, \omega)$ are also differentiable. The condition that $n^{-1}\partial S_n(y, \omega)/\partial \omega$ converges to a finite nonsingular matrix $A(\omega^{(0)}) = E(\partial s_t(Y_t, \omega)/\partial \omega)_{\omega^{(0)}}$ in probability can be deduced by applying the general ergodic theorem [18] to the Markov chain $(Y_t, \ldots, Y_{t-p})', t \in \mathbb{N}$.

The condition $\lim_{n\to\infty} \sup_{\delta \searrow 0}(n\delta)^{-1}|T_n(y, \omega^\star)| < \infty$ a.s., where $T_n(y, \omega^\star) = (\partial S_n(y, \omega)/\partial \omega)_{\omega^\star} - (\partial S_n(y, \omega)/\partial \omega)_{\omega^{(0)}}$, can be verified along the same lines as in the proof of Lemma 3 of [27], albeit the handling of the derivatives is much more cumbersome in our case.

For the final condition $S_n(Y, \omega^{(0)})/\sqrt{n} \to N(0, B(\omega^{(0)}))$, where $B(\omega^{(0)}) = E(s_t(Y, \omega^{(0)}) \times s_t(Y, \omega^{(0)})')$, we use Lemma 1b). Altogether this results in

$$\sqrt{n}(\hat{\omega}_n - \omega^{(0)}) \to N\left(0, A(\omega^{(0)})^{-1}B(\omega^{(0)})A(\omega^{(0)})^{-1}\right).$$

Lemma 1 *We have as $n \to \infty$*

a) $S_n(Y, \omega)/n \xrightarrow{a.s.} E(s_t(Y, \omega))$

b) $S_n(Y, \omega^{(0)})/\sqrt{n} \xrightarrow{d} N(0, B(\omega^{(0)}))$, with $B(\omega^{(0)}) = E(Var(s_t(Y, \theta^{(0)})|\mathcal{F}_{-1}))$.

Proof By construction $(s_t(Y, \omega) - E(s_t(Y, \omega)|\mathcal{F}_{t-1}))$ is a martingale difference sequence and $(S_n(Y, \omega) - \sum_{t=p+1}^{n} E(s_t(Y, \omega)|\mathcal{F}_{t-1}))$ is a martingale. We thus make use of the strong law of large numbers [7] and the central limit theorem for martingales.

For a) it suffices that $s_t(Y, \omega) - E(s_t(Y, \omega)|\mathcal{F}_{t-1}))$ is square integrable, $E||s_t(Y, \omega) - E(s_t(Y, \omega)|\mathcal{F}_{t-1}))||^2 < \infty$, for the true set of parameter values. Because of $E||E(Z|\mathcal{F}_{t-1})||^2 \leq E||Z||^2$ for any random variable Z we have

$$E||s_t(Y, \omega) - E(s_t(Y, \omega)|\mathcal{F}_{t-1})||^2 \leq 2E||s_{t,0}(Y, \omega)||^2 + \cdots$$
$$+ 2E||s_{t,p}(Y, \omega)||^2 + 2E||u_t(Y, \omega)||^2.$$

The boundedness of $E||u_t(Y, \omega)||^2$ has already been verified by [27]. For the other terms $E||s_{t,i}||^2, i = 0, \ldots, p$, this is obvious due to the boundedness of the ψ-function, as long as σ and σ_t are bounded away from 0. This is guaranteed since $\alpha_0 > 0$.

For b) we show the conditional Lindeberg condition using

$$\frac{1}{n} \sum_{t=p+1}^{n} E\left[||s_t||^2 I(||s_t|| > \sqrt{n}\epsilon|\mathcal{F}_{t-1}) \leq \frac{1}{n^2\epsilon^2} \sum_{t=p+1}^{n} E[||s_t||^4|\mathcal{F}_{t-1}]\right].$$

It is hence sufficient to show $E||s_t||^4 < \infty$. This follows using

$$||s_t||^4 = [||s_{t,0}||^2 + \cdots + ||s_{t,p}||^2 + ||u_t||^2]^2 \leq [(p+1)\max\{||s_{t,0}||^2, \ldots, ||s_{t,p}||^2, ||u_t||^2\}]^2,$$

i.e., it is sufficient that $E||s_{t,i}||^4$, $i = 0, \ldots, p$, and $E||u_t||^4$ are all finite. For $E||u_t||^4$ see [27], and for $E||s_{t,i}||^4$, $i = 0, \ldots, p$, this is again obvious because of the boundedness of ψ_c and $\alpha_0 > 0$. Moreover,

$$\frac{1}{n}\sum_{t=p+1}^{n} Var(s_t|\mathcal{F}_{t-1}) \xrightarrow{p} E[Var(s_t s_t'|\mathcal{F}_{t-1})] = B(\omega^{(0)}).$$

Application of the central limit theorem for martingales [26, Theorem 1.1.13] proves

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} s_t \xrightarrow{d} N(0, B(\omega^{(0)})).$$

$\square$

References

1. Aeberhard, W.H., Cantoni, E., Heritier, S.: Robust inference in the negative binomial regression model with an application to falls data, Biometrics, 70, 920–931 (2014)
2. Amemiya, T.: Advanced Econometrics. Harvard University Press, Cambridge (1985)
3. Axt, I., Fried, R.: On variance estimation under shifts in the mean. AStA Advances in Statistical Analysis **104**, 417–457 (2020)
4. Breslow, N.E.: Extra-Poisson variation in log-linear models. Journal of the Royal Statistical Society, Series C **33**, 38–44 (1984)
5. Cadigan, N.G., Chen, J.: Properties of robust M-estimators for Poisson and negative binomial data. Journal of Statistical Computation and Simulation 70, 273–288 (2001)
6. Cantoni, E., Ronchetti, E.: Robust inference for generalized linear models. Journal of the American Statistical Association **96**, 1022–1030 (2001)
7. Chow, Y.S.: On a strong law of large numbers for martingales. The Annals of Mathematical Statistics, 38(2), 610 (1967)
8. Christou, V., Fokianos, K.: Quasi-Likelihood Inference for Negative Binomial Time Series Models. J. Time Ser. Anal. **35**(1), 55–78 (2014)
9. Durre, A., Fried, R., Liboschik, T.: Robust estimation of (partial) autocorrelation. WIREs Computational Statistics **7**(3), 205–222 (2015)
10. Elsaied, H., Fried R.: Robust fitting of INARCH models. Journal of Time Series Analysis 35 (6), 517–535 (2014)
11. Elsaied, H., Fried, R.: Tukey's M-estimator of the Poisson parameter with a special focus on small means. Statistical Methods and Applications 25(2), 191–209 (2016)
12. Ferland, R., Latour, A., Oraichi, D.: Integer valued GARCH processes. J. Time Series Anal. 27, 923–942 (2006)
13. Fokianos, K., Fried, R.: Interventions in INGARCH processes. Journal of Time Series Analysis 31, 210–225 (2011)
14. Fokianos, K., Rahbek, A., Tjostheim, D.: Poisson autoregression. J. Amer. Statist. Assoc. **104**, 1430–1439 (2009)
15. Fried, R., Liboschik, T., Elsaied, H., Kitromilidou, S., Fokianos, K.: On Outliers and Interventions in Count Time Series following GLMs. Austrian Journal of Statistics **43**(3), 181–193 (2014). https://doi.org/10.17713/ajs.v43i3.30
16. Garel, B., Hallin, M.: Rank-Based Autoregressive Order Identification. J. Am. Stat. Assoc. 94(448), 1357-1371 (1999)
17. Hubert, M., Debruyne, M.: Minimum Covariance Determinant. WIREs Comp Stat **2**, 36–43 (2010)
18. Jensen, S.T., Rahbek, A.: On the law of large numbers for (geometrically) ergodic Markov chains. Econometric Theory **23**, 761–766 (2007)
19. Lawless, J.F.: Negative binomial and mixed Poisson regression. The Canadian Journal of Statistics **15**, 209–225 (1987)

20. Liboschik, T., Fokianos, K., Fried, R.: tscount: an R package for analysis of count time series following generalized linear models. J. Stat. Softw. **82**, 5 (2017). https://doi.org/10.18637/jss.v082.i05

21. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2019). https://www.R-project.org/

22. Raymaekers, J., Rousseeuw, P.J., Van den Bossche, W., Hubert, M.: CellWise. https://cran.r-project.org/web/packages/cellWise/cellWise.pdf (2019)

23. Rooch, A., Zelo, I., Fried, R.: Estimation methods for the LRD parameter under a change in the mean. Statistical Papers, 60 (1), 313–347 (2019)

24. Rousseeuw, P.J.: Multivariate estimation with highbreakdown point. In: Grossmann, W., Pflug, G., Vincze, I., Wertz, W. (eds.) Mathematical Statistics and Applications, vol. B, pp. 283–297. Reidel Publishing Company, Dordrecht (1985)

25. Rousseeuw, P.J., van Zomeren, B.C.: Unmasking multivariate outliers and leverage points. J. Amer. Statist. Assoc. **85**, 633–639 (1990)

26. Taniguchi, M., Kakizawa, Y.: Asymptotic Theory of Statistical Inference for Time Series. Springer, New York (2000)

27. Xiong, L., Zhu, F.: Robust quasi-likelihood estimation for the negative binomial integer-valued GARCH(1,1) model with an application to transaction counts. J. Statist. Plan. Inference **203**, 178–198 (2019)

28. Zhu, F.: A negative binomial integer-valued GARCH model. J. Time Series Anal. **32**, 54–67 (2011)