

Original Research Paper

Robust Modelling of Coronavirus Cases in Egypt: Poisson INARCH and Negative Binomial INARCH

Hanan Elsaied

Department of Statistic, Faculty of Commerce, Suez Canal University, Egypt

Article history

Received: 03-05-2021

Revised: 04-06-2021

Accepted: 08-06-2021

E-mail: hanan167@yahoo.com

Abstract: This study compares robust Poisson INARCH(P) models (more briefly: RP-INARCH) and robust negative binomial INARCH(p) models (more briefly: RNB-INARCH) to fit the new daily confirmed cases for the first wave of COVID 19 in Egypt. The robust estimation of these models is based on some modifications of the Conditional Maximum Likelihood Estimates (CMLE). The simulation results show that RNB-INARCH is more robust than RP-INARCH, but less efficient if the data contain isolated or patched additive outliers in terms of the bias calculation, whereby the low-bias model is more robust. These results are confirmed by the application study on COVID-19 data. The Akaike Information Criterion (AIC) is also compared for these models.

Keywords: COVID 19 Data, Robust Poisson INARCH(p) Models, Robust Negative Binomial INARCH(p) Models, Isolated Additive Outliers, Patch Additive Outliers

Introduction

In 2020, Corona virus or “SARS-COV-2” (abbreviation of Corona-2 virus, severe active respiratory syndrome) also known as “COVID 19” spread globally. In 11 March 2020, the World Health Organization declared this virus as a pandemic. By the end of 2020, the virus has spread in at least one case in 218 countries. In Egypt, the primary case was reported on 14/02/2020. At the end of 2020, it confirmed its higher new cases of COVID-19 that occurred every day and it's become essential from the statistical point of view to analyze the behavior of this disease. Biostatistics as an important branch of statistic provides us with models which help to solve most problems of human health and diseases. Elsaied (2021) proposed some examples in the medical field, which used different statistical models: Heinen (2003) used a conditional autoregressive Poisson model (CAP) to fit the time series of monthly polio cases in the United States. Promprou *et al.* (2006) used ARIMA model to predict dengue haemorrhagic fever cases in southern Thailand. Martinez *et al.* (2011) used SARIMA model to predict dengue fever cases in Campinas, São Paulo State, Brazil. Ahdika and Lusiyana (2017) compared the INAR (1) Poisson model and Markov prediction model in West Java,

Indonesia to predict the number of Dengue Hemorrhagic fever (DH) patients. On the other hand, Elsaied (2021) expanded other models that were applied to COVID-19 data: Giuliani *et al.* (2020) used GLM model class to model and predict the spatio-temporal spread of Corona virus disease 2019 (COVID-19) in Italy. Fahmy *et al.* (2020) used a generalized and modified version of classical SIR/SEIR” Susceptible-Exposed-Infectious-Recovered” models to analyze the dynamics outbreak of COVID-19 in Egypt, Qatar and Saudi Arabia. Asamoah *et al.* (2020) extended the generalized SEIR epidemic model to capture the dynamics of COVID-19 in Ghana and Egypt. Lauer *et al.* (2020) assumed totally different distribution: Log normal, gamma, Weibull and Erlang to estimate the period of time of COVID-19 from publicly reportable confirmed cases. Babaei *et al.* (2021a) investigated a stochastic SEIAQHR model for transmission of Corona-virus disease. Their model was established due to several safety protocols, for instance social-distancing, mask and quarantine. Also, Babaei *et al.* (2021b) introduced a stochastic model to describe the spread of coronavirus with considering several disease compartments related to different age groups. Khan *et al.* (2021), proposed a stochastic model to discuss the dynamics of novel corona virus disease. They formulated the model to study the long run behavior in varying

population environment. Additionally, Elsaied (2021) recommended to use the RP-INARCH(p) models to fit the number of new daily confirmed cases of “COVID 19” in Egypt, wherein those models proved its reasonability than the non robust ones to model such data. These models are considered to be the simplest subclasses of the INGARCH (p, q) models. The INGARCH (p, q) were developed early by Ferland *et al.* (2006) and Fokianos *et al.* (2009) among others. This study suggests to compare between RP-INARCH(p) and RNB-INARCH(p) models to fit “COVID 19” count data. In (2014), Elsaied and Fried analyzed the robust estimation of the Poisson-INARCH model for counting time series data. Also, negative binomial INARCH(P) (NB-INARCH) models and their comparison with the Poisson-INGARCH(P) (P-INARCH) models were proposed for time series data. For non-robust data: Davis and Wu (2009) studied the negative binomial model of time series counting and provided a performance estimator, which is an efficiency estimator, very close to the ML estimator. Zhu (2010) compared between Yule-Walker, the conditional least squares estimator and the maximum likelihood estimator (MLE) in terms of the mean absolute deviation error (MADE) under the shape parameter of NB-INGARCH model is known. Jamaludin *et al.* (2020) used P-INGARCH and NB-INGARCH models to investigate the behavior of asthma disease in Johor Bahru. In keeping with their results NB-INGARCH with identity and log link function is adequate in representing the asthma data than P-INGARCH model. As far as we know that there are few researches applied for robust time series data: Xiong and Zhu (2019) conferred the robust quasi-likelihood estimation for the NB-INGARCH (1,1) model within the presence of transient shifts and additive outliers with an application to transaction counts data. Recently, Elsaied and Fried (2021) discussed robust M-estimation for NB-INARCH(p) models. In their study, they mixed the M-estimation approaches developed by Elsaied and Fried (2014) for the restricting Poisson INARCH model with Aeberhard *et al.* (2014) for the negative binomial regression model. The remainder of this study, through its sections, will compare between RP-INARCH (1) and RNB-INARCH (1) to fit the new daily confirmed cases for the first wave of COVID 19 in Egypt. Section 2 provides the theoretical basis for modelling and estimating these models. Section 3 uses some built-in functions in the R program to calculate the parameter estimates for our model. The simulation comparison between our functions is given in section 4. The use of COVID-19 count data is described in section 5. Section 6 contains some conclusions based on the analysis results in sections 4 and 5.

Modeling and Estimation

Modeling

Let $\{Y_t : t \in \mathbb{N}\}$ denotes a count time series and $Y_t | F_{t-1}$ with F_{t-1} stands for the history of the process generated by Y_{t-1} up to time $t-1$. When the process allows equal-dispersion, we assume the observations at each time point conditionally on the past to follow a Poisson ($Y_t | F_{t-1} \sim Pois(\mu_t)$). But when the process allows over-dispersion, we assume the observations at each time point conditionally on the past to follow the negative binomial $Y_t | F_{t-1} \sim NB(r, p_t)$. Here, $r \in \mathbb{N}$ is a constant number of successes and p_t is a probability of success that varies with time. Aeberhard *et al.* (2014) represented the density function for the linear NB-model in terms of the inverse probability $\mu_t = (1 - p_t)/p_t$ with $\kappa = r^{-1} \geq 0$ is used as a measure of the over-dispersion parameter, as shown below:

$$P(Y_t = y_t | \mathcal{F}_{t-1}) = \frac{\Gamma(y_t + \kappa^{-1})}{\Gamma(\kappa^{-1})\Gamma(y_t + 1)} \left(\frac{\kappa^{-1}}{\kappa^{-1} + \mu_t} \right)^{1/\kappa} \left(\frac{\mu_t}{\kappa^{-1} + \mu_t} \right)^{y_t} \quad (1)$$

The conditional Poisson model can be obtained from (1) when κ converges to zero.

The INARCH(p) models of orders p can be defined in terms of the condition mean for the both models as follows:

$$\mu_t = \alpha_0 + \sum_{i=1}^p \alpha_i Y_{t-i} \quad (2)$$

Here, $t \geq 1$, $\alpha_0 > 0$ is an intercept and α_i , $i = 1, \dots, p$, are positive regression parameters. A stationary process that satisfies (2) with mean $\alpha_0 / (1 - \sum_{i=1}^p \alpha_i)$ exists if $\sum_{i=1}^p \alpha_i < 1$. The conditional mean and the conditional variance for the NB-INARCH(p) are given by: $E(Y_t | F_{t-1}) = \mu_t$ and $Var(Y_t | F_{t-1}) = \sigma_t^2 = \mu_t + \kappa \mu_t^2$. For the P-INARCH(p), the conditional mean is equal to the conditional variance ($\sigma_t^2 = \mu_t$) if $\kappa \rightarrow 0$.

Estimation

According to Elsaied and Fried (2021), P-INARCH(p) and NB-INARCH(p) models can be estimated non-robustly by the conditional maximum likelihood estimation (CMLE) as an optimization problem as follows: let \hat{z} be the log-conditional likelihood function, $z = (\theta, \kappa)$ is the parameter vector, with $\theta = (\alpha_0, \dots, \alpha_p)$, the Conditional Maximum Likelihood Estimation (CMLE) estimation equation of NB-INARCH(p) can be calculated by finding the roots of the score equation:

$$\begin{aligned} \frac{\partial \ell(z)}{\partial z} &= \left(\sum_{i=1}^n S'_{i,\theta}(z), \sum_{i=1}^n S_{i,\kappa}(z) \right)' = 0 \\ S_{i,\theta}(z) &= \frac{y_i - \mu_i}{\mu_i + \kappa \mu_i^2} \frac{\partial \mu_i}{\partial \theta} \\ S_{i,\kappa}(z) &= -\kappa^{-2} \left(\frac{\partial \ln \Gamma(y_i + 1/\kappa)}{\partial (y_i + 1/\kappa)} - \frac{\partial \ln \Gamma(1/\kappa)}{\partial \kappa} - \ln(\kappa \mu_i + 1) - \frac{\kappa(y_i - \mu_i)}{\kappa \mu_i + 1} \right) \\ &= -\frac{1}{\kappa^2} \left(M(y_i + 1/\kappa) - M(1/\kappa) - \ln(\kappa \mu_i + 1) - \frac{\kappa(y_i - \mu_i)}{\kappa \mu_i + 1} \right) \end{aligned} \tag{3}$$

Here, $\frac{\partial \mu_i}{\partial \theta} = (1, y_{i-1}, \dots, y_{i-p})^0$, $M(u)$ represents the digamma function, $M(u) = \partial \ln \Gamma(u) / \partial u$.

By solving the score equation $S_{i,\theta}(z)$ with $\kappa = 0$, we obtain the CMLE of θ for the P-INARCH(p) models.

For κ , we use the moment estimation defined in (Breslow, 1984; Lawless, 1987; Christou and Fokianos, 2014) in the following:

$$\frac{1}{n-2p-1} \sum_{i=p+1}^n \left(\frac{y_i - \mu_i}{\sqrt{\mu_i + \kappa \mu_i^2}} \right)^2 = 1 \tag{4}$$

here $p+1$ is that the dimension of θ .

Following to Elsaied and Fried (2021), an M-estimation approach estimation to estimate θ given in (3) has the following form:

$$\begin{aligned} \sum_{i=p+1}^n S_{i,0}(y_i, z) &= \sum_{i=p+1}^n \psi(r_i) \frac{1}{\sigma_i} - (n-p)d_{i,0} = 0, \\ \sum_{i=p+1}^n S_{i,i}(y_i, z) &= \sum_{i=p+1}^n \psi(r_i) \frac{1}{\sigma_i} - \left[\sigma \psi \left(\frac{y_{i-i} - \mu}{\sigma} \right) + \mu \right] \\ -(n-p)d_{i,i} &= 0, \quad i = 1, \dots, p. \end{aligned} \tag{5}$$

with related conditional Pearson-residuals $r_i = \left(\frac{y_i - \mu_i}{\sigma_i} \right)$

and σ_i^2 as given above. $d_{i,i}$: $i = 0, \dots, p$ denotes the bias correction terms. The bias corrections can be calculated in the same way as in Elsaied and Fried (2014) by:

$$\begin{aligned} d_{i,0} &= E_z \left[\psi \left(\frac{Y_i - \mu_i}{\sigma_i} \right) / \sigma_i \mid \mathcal{F}_{i-1} \right] \text{ and} \\ d_{i,i} &= E_z \left[\psi \left(\frac{Y_i - \mu_i}{\sigma_i} \right) \left[\sigma \psi \left(\frac{Y_{i-i} - \mu}{\sigma} \right) + \mu \right] / \sigma_i \mid \mathcal{F}_{i-1} \right], \quad i = 1, \dots, p \end{aligned}$$

To estimate κ , we use the M-estimator of Aeberhard *et al.* (2014) as follows:

$$\begin{aligned} \sum_{i=p+1}^n \left[\psi(r_i) \frac{1}{\kappa^2 r_i} \left(M(y_i + 1/\kappa) - M(1/\kappa) - \ln(\kappa \mu_i + 1) - \frac{\kappa(y_i - \mu_i)}{\kappa \mu_i + 1} \right) \right] \\ \Psi(y_{i-i}, \mu, \sigma) = 0, \end{aligned}$$

with:

$$\Psi(y_{i-i}, \mu, \sigma) = \frac{\left[\sigma \psi \left(\frac{y_{i-i} - \mu}{\sigma} \right) + \mu \right]}{y_{i-i}}$$

For computation, we can fix κ and drive a new estimate of κ iteratively as follows:

$$\Rightarrow \kappa = \frac{\sum_{i=p+1}^n \left[\psi(r_i) \frac{1}{r_i} \left(M(y_i + 1/\kappa) - M(1/\kappa) - \ln(\kappa \mu_i + 1) \right) \right] \Psi(y_{i-i}, \mu, \sigma)}{\sum_{i=p+1}^n \frac{y_i - \mu_i}{\kappa \mu_i + 1} \psi(y_{i-i} - i, \mu, \sigma)}$$

Initialize σ from Poisson and then retrieve κ in an iterative process until the numerical convergence is obtained. Here, we use the ψ function for Tukey, which can be abbreviated as:

$$\psi_k(y, \theta) = (y - \theta) \left(C^2 - (y - \theta)^2 \right)^2 I_{[-k,k]}(y - \theta)$$

with I_A is the indicator function of a real set A .

The tuning constant C determines the robustness and the efficiency of the estimators. Choosing a higher value of C can improve efficiency, but will reduce the robustness to outliers.

The asymptotic properties of the consistency and normal convergence of the M-estimators for the RP-INARCH and RNB-INARCH models are discussed in Elsaied and Fried (2014) and Elsaied and Fried (2021), respectively.

Computation

Elsaied (2021) proposed two functions implemented in the *R* program to compute the estimates of the P-

INARCH (1) parameters in a robust and non-robust way. We recall the following notation for using those functions: “CML-Pois” to compute the conditional maximum likelihood estimator and “Tukeycor-Pois” to compute M-estimation estimator with bias correction term. Similarly, we provide two new functions to calculate the parameter estimates of the NB-INARCH model (1) given in (3) and (5), by using “CML-NB” to denote the first function and “Tukeycor-NB” to call the second one.

We follow the steps below to calculate the “CML-NB” algorithm in the R-program:

1. Define a function in data(y), the parameter vector $\theta = (\alpha_0, \alpha_1)$ to calculate the score equation $S_{t,\theta}(z)$ with $\kappa = 0$ defined in (3) to get the Poisson likelihood estimator of θ
2. Define the function “CML-NB” as a function on θ and data (y) in order to solve the estimation score equation $S_{t,\theta}(z)$ defined in (3) to obtain the NB quasi-likelihood estimator of θ . These estimates are obtained using the “constrOptim” R function with initial values obtained under Poisson case resulted from Step 1
3. In order to simplify the calculation of κ , assume that $\kappa = 0.3$ is a known value, as given by Xiong and Zhu (2019)

While the steps to calculate “Tukeycor-NB” algorithm in R program are as follows:

1. Define a function in data (y), the parameter vector $\theta = (\alpha_0, \alpha_1)$ and in the tuning constant (C). Then solve the score equation $S_{t,\theta}(z)$ with $\kappa = 0$ given in (5) to obtain M-estimation for Poisson quasi-likelihood estimator of θ
2. Define the function “Tukeycor-NB” as a function of y and C. This function solves the estimation equation $S_{t,\theta}(z)$ defined in (5) to calculate the NB M-estimation of θ with bias correction term. As above again, we use the “constrOptim” function with initialization resulted from step 1
3. Here, also κ is assumed to be known as described above

In order to determine which of the two models is the best, we suggest to use Akaike Information Criterion (AIC). The general AIC formula for a given modal using the CMLE of the parameter vector θ is as follows:

$$AIC = (-2)LML(\hat{\theta}) + 2N \quad (6)$$

with LML is the log likelihood of the fitted model and N is the number of parameters in the model. Different modified robust versions of (6) are given in Ronchetti (1997) and Tharmaratnam and Claeskens (2013) among others. In order to simplify the calculation of AIC, we call the “robmixglm” function in the package robmixglm in R-program. The “robmixglm” is a function in: [Formula(y, $\sim y_{t-1}$ with $t \geq 2$), family = c(“Poisson”, “nbinom”, ...)]. It returns AIC as a “robmixglm” object by using the following

command: $AIC(\text{robmixglm}(y_t, \sim y_{t-1}))$. We note that this package offers the possibility to calculate AIC criteria which are not available under “glmrob” function in the R package robustbase, for more details see Beath (2017).

Simulations

In this section, we run some simulation experiments to compare the performance of the functions defined above: “CML-Pois”, “CML-NB”, “Tukeycor-Pois” and “Tukeycor-NB”. We consider scenarios with no outliers and those with isolated or patched additive outliers and real parameter vector $z = (\alpha_0, \alpha_1, \kappa) = (0.55, 0.4, 0.3)$, which are similar to those selected by Xiong and Zhu (2019).

For Clean Data

We start by comparing the efficiency for data without outliers and generate 500 time series with the length $n = 1000$ from P-INARCH (1) and NB-INARCH (1) models as a function in the tuning constant C . The efficiencies for “Tukeycor-Pois” and “Tukeycor-NB” are measured relatively for “CML-Pois” and “CML-NB”, respectively. As shown in Fig. 1 (both top), the “Tukeycor-Pois” estimators achieve about 85% level of efficiency for α_0 and about 90% level of efficiency for α_1 . While the “Tukeycor-NB estimator improves this level to reach about 95% with tuning constant $C > 9$. Figure 1 (both below) shows the biases also as function of the tuning constant. There are slight differences in biases for all estimators with tuning constant $C > 7$. These differences become larger when $C < 7$ especially for α_0 .

According to Fig. 1 and since the suitable selection of the tuning constants depends in principle on the parameters of the model, we suggest to use $C \in \{7, 9, 11\}$ where for:

$$C = \begin{cases} 7, & \text{“Tukeycor – pois” efficiency is} > \text{“Tukeycor – NB” efficiency} \\ 9, & \text{“Tukeycor – pois” efficiency is} = \text{“Tukeycor – NB” efficiency} \\ 11, & \text{“Tukeycor – pois” efficiency is} < \text{“Tukeycor – NB” efficiency} \end{cases} \quad (7)$$

In the next two subsections, we will compare the estimated values in cases of presence isolated or patch additive outliers and increase or fix their sizes using the previously suggested values of the tuning constant C to control the efficiency level.

Isolated Additive Outliers

In this section, we will compare the biases between our estimators to increase 4, 8, ..., 40 isolated outliers of size 5 or 10 at random selected positions. We use 100 time series, each with a sample size of 500. For size 5 in Fig. 2, the loss of robustness of the “CML-Pois” or “CML-NB” is expressed as an increase in positive (negative) bias for α_0 (α_1), especially for the larger number of outliers. For α_0 , “Tukeycor-NB”, with $C = 7$ has the smallest bias. Whereas for α_1 , “Tukeycor-Pois” with $C = 7$ has the smallest bias. Both followed by the biases with $C = 9$. But CMLEs give the largest biases for both

α_0 and α_1 . For size 10 in Fig. 3, all estimators show similar biases behavior however the CMLEs give the largest biases. Here, the differences in biases between “CML-Pois”, “CMLNB” in one side and “Tukeycor-Pois”, “Tukeycor-NB” on the other hand, it becomes larger than in case of size 5 for both for both α_0 and α_1 .

Patch Additive Outliers

Here, we consider a scenario that has a consecutive patch outlier of the same size. This type may occur due to temporary interruptions or temporary level shifts. Figure 4 compares our estimated biases of 20 patch outliers of fixed size 5 at positions 101: 120. We use 100 time series for each with sample size = 500. All estimators show

small biases differences. “Tukeycor-NB” with $C = 7$ and $C = 9$ give smaller biases than with $C = 11$ for both α_0 and α_1 . Figure 5 compares the biases in case of 20 patch of additive outliers. The size increases from 1 to 20. For α_0 , all estimators show small biases differences, except for “CML-Pois” with a large number of outliers >10 . But for α_1 , “Tukeycor-NB” with $C = 7$ and $C = 9$ give smaller biases than those for “Tukeycor-Pois”. Followed by the biases with $C=11$. Whereas “CML-Pois” and “CML-NB” have the largest biases, especially for number of outliers >6 .

Altogether, we conclude that “Tukeycor-Pois” and “Tukeycor-NB” with $C = 7$ and $C = 9$ give better results than those with $C=11$. Also, they are better than the CMLEs under the both scenarios of outliers: Isolated or patch additive.

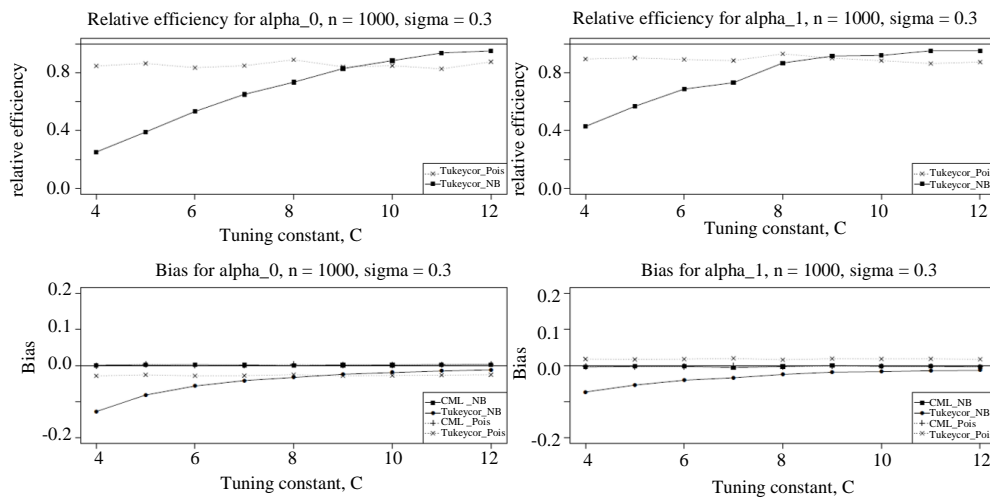


Fig. 1: Simulated efficiencies (both above) for α_0 on the left and α_1 on the right as a function of the tuning constant C for M-estimators “Tukeycor-Pois” and “Tukeycor-NB” relatively to CMLEs. Similar experimental biases (both bottom) for α_0 (left) and α_1 (right) with $z = (0:55, 0:4, 0:3)$ and $n = 1000$

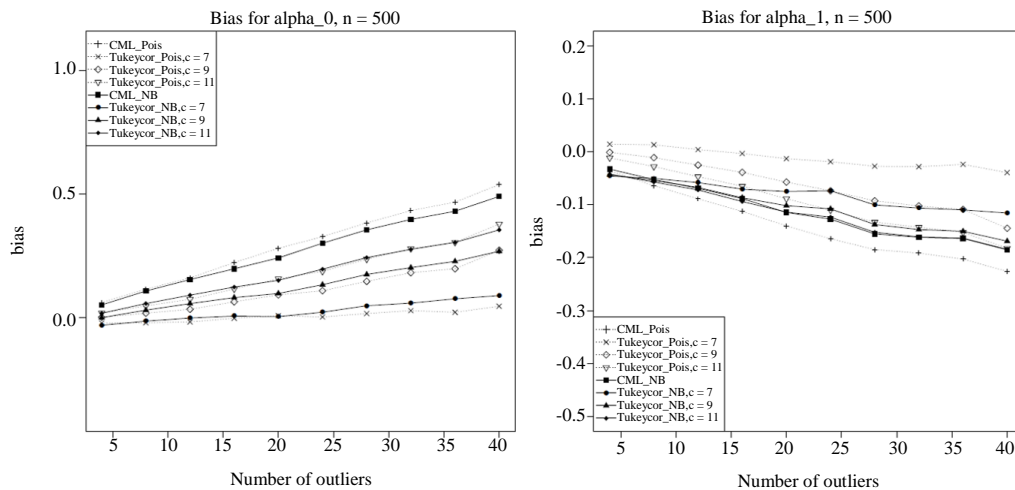


Fig. 2: Simulated biases for M-estimators and CMLEs with various C for α_0 (in left) and α_1 (in right) for 40 isolated outliers of size 5 with $z = (0:55, 0:4, 0:3)$ and $n = 500$

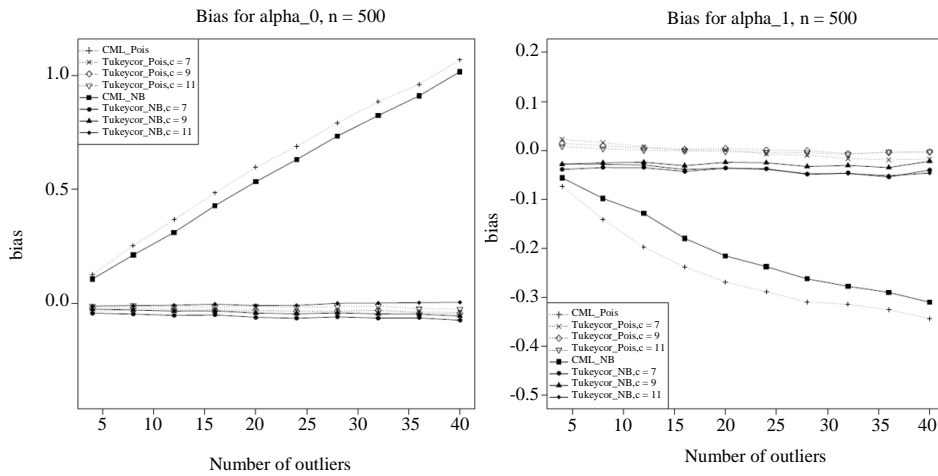


Fig. 3: Simulated biases for M-estimators and CMLEs with various C for α_0 (in left) and α_1 (in right) for 40 isolated outliers of size 10 with $z = (0:55, 0:4, 0:3)$ and $n = 500$

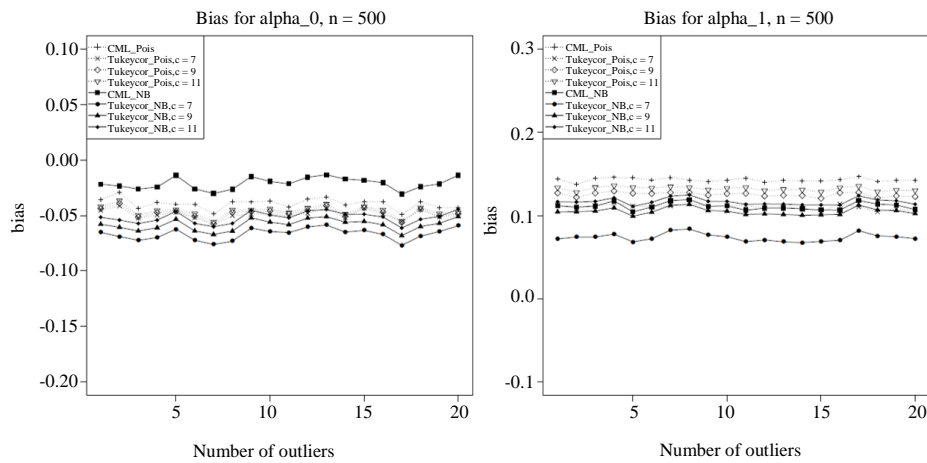


Fig. 4: The biases for M-estimators and CMLEs with various C for α_0 (in left) and α_1 (in right) for 20 patched-outliers of size 5 with $z = (0:55, 0:4, 0:3)$ and $n = 500$

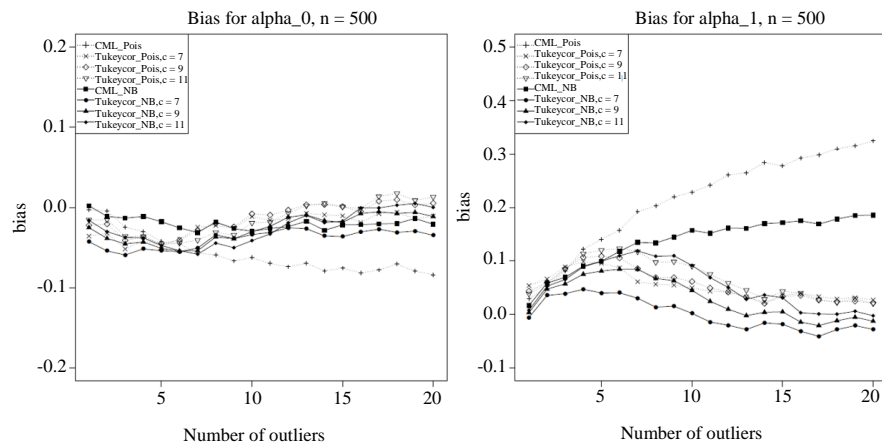


Fig. 5: The biases for M-estimators and CMLEs with various C for α_0 (in left) and α_1 (in right) for 20 patched-outliers, their size increases with $z = (0:55, 0:4, 0:3)$ and $n = 500$

Using COVID-19 Data

This section compares the performance of our best function resulted above, which are: Tukeycor-Pois and Tukeycor-NB with tuning constants 7 and 9 through fitting RP-INARCH(1) and RNB-INARCH(1) for the variable of the number of new “COVID-19” confirmed cases in Egypt. Figure 6 shows 265 observations of each day new confirmed cases of “COVID 19” from March 7, 2020 to November 26, 2020. This period corresponds to the time until the end of the first wave of COVID 19 in Egypt as it is declared. The source of this data and its description are given in Elsaied (2021). Here, we follow the steps below to reach which one of the two models is the best choice:

1. Fit RP-INARCH (1) and RNB-INARCH (1) to the real confirmed COVID-19 cases using the functions “Tukeycor-Pois” and “Tukeycor-NB” with tuning constants 7 and 9
2. Use the modified version of the “interv-multiple” function in the R-package “tscount” to detect and remove the outlier. This function returns the data after each step of the correction process using the same setting parameter as given in Elsaied (2021)
3. Repeat Step 1 again, but after removing the outliers according to Step 2
4. Compare the values of AIC criteria for our functions (the less AIC is the better) depending on the function “robmixglm” as defined before.

Table 1 (right) shows the results of the final estimates by RP-INARCH(1) and RNB-INARCH(1), which depend on the data obtained from the last step of the correction

procedure. Here is step 24, Fig. 6. By using of the “interv. multiple” function, we detect 10 essential outliers without repeating and without including the small level shifts at the beginning of the series with 7 of 10 (70% of the determined outliers) were level shift outliers. These outliers at point times: 45, 69, 117, 119, 133, 237 and 255. As expected for clean data, “Tukeycor-Pois” and “Tukeycor-NB” with tuning constants 7 and 9 give similar results for $\hat{\alpha}_0$ and $\hat{\alpha}_1$. Also, the differences in “RMSEs” are negligible. In Table 1 (left), however, all estimators give different values for both for $\hat{\alpha}_0$ and $\hat{\alpha}_1$. “Tukeycor-NB” with C=7 and C=9 give smaller RMSEs” than for “Tukeycor-Pois”. For both types of data, the “RMSEs” are computed by simulation using 100 data sets of sizes 265 from P-INARCH (1) and RP-INARCH (1) models with true parameter values computed by the CMLE.

For both data types, “Tukeycor-NB” has smaller AICs than that of “Tukeycor-Pois” and we find that the values of AIC do not affect the values of the tuning constant C. This is because that the function “robmixglm” does not depend on the tuning parameter values to be determined in advance, Beath (2017).

We note that for the dependence parameter α_1 , there’s nearly no serial relation and α_0 absorbs all outliers’ effects. An explanation of this problem is that the presence of seven level shifts can be explained as the long sequence of outliers, that is, the high degree contamination.

According to our situation here, we deduce that: At tuning constants 9: Both estimators have the same efficiency and RNB-INARCH(1) is more robust. At tuning constants 7: RP-INARCH(1) is more efficiency but less robust than RNB-INARCH(1) and the choice of any one depends on the aim of the study or the desire properties of the estimator.

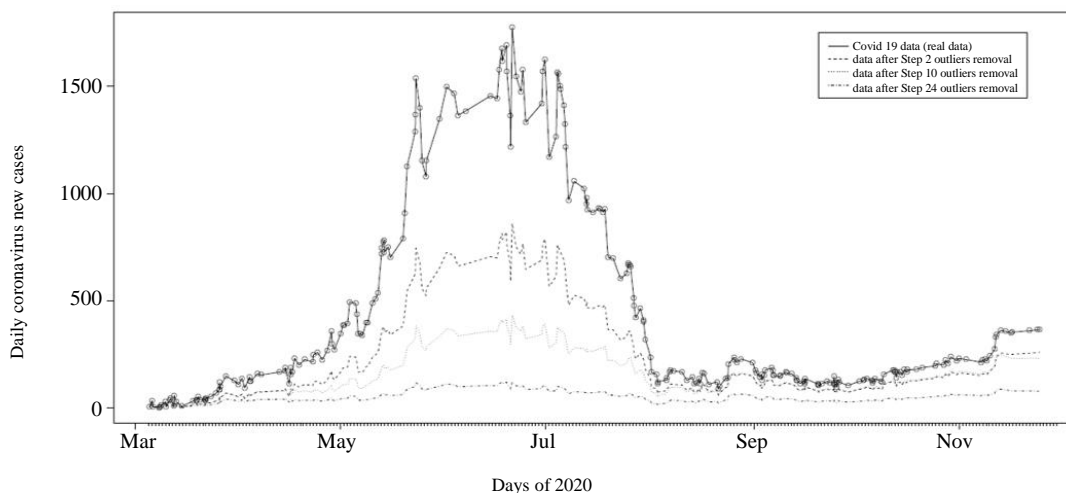


Fig. 6: COVID 19 data (in solid line) and data after the outliers are removed (in dashed line)

Table 1: Parameter vector $\theta = (\alpha_0, \alpha_1)$ estimates, their related roots of the Mean Square Errors (RMSEs) between the parentheses and AIC for Covid 19 data (left) and for the cleaned Covid 19 (right)

Estimates model estimators	Real data			Clean data		
	$\hat{\alpha}_0$	$\hat{\alpha}_1$	AIC	$\hat{\alpha}_0$	$\hat{\alpha}_1$	AIC
“Tukeycor-Pois, C= 7”	14.783 (2.2346)	0.013 (0.1378)	3823	2.718 (1.5068)	0.956 (0.0424)	1975
“Tukeycor-Pois, C= 9”	35.373 (2.2187)	0.088 (0.1397)	3823	2.893 (1.5163)	0.952 (0.0422)	1975
“Tukeycor-NB, C= 7”	1.74 (1.766)	1.27*10 ⁻⁹ (0.1147)	3242	0.005 (1.4565)	0.850 (0.0605)	1955
“Tukeycor-NB, C= 9”	37.369 (1.5143)	0.006 (0.1124)	3242	0.006 (1.4426)	0.918 (0.0581)	1955

Conclusion

We have compared the robust parameter estimates of INARCH (1) model with conditional Poisson or negative binomial distributions. Our proposal of negative binomial is more robust with the tuning constant = 7, but has a lower efficiency compared to Poisson. For the tuning constant = 9, both have similar efficiencies, but the negative binomial is more robust. Obviously, it is more difficult to choose a tuning constant in order to achieve high efficiency and high robustness for different outlier scenarios (isolated or patched). Therefore, when robustness is important, we prefer to use negative binomial. When efficiency is important, we prefer Poisson. Alternatively for future work, we suggest to use the concept of numerical quadrature, ordinary differential equation with block method, non-linear partial differential equation with B-Spline collocation and differential transformation technique instead of Poisson process.

Ethics

The author confirms that this article is original and contains unpublished material and the author has read and approved the manuscript and no ethical issues involved.

References

Aeberhard, W. H., Cantoni, E., & Heritier, S. (2014). Robust inference in the negative binomial regression model with an application to falls data. *Biometrics*, 70(4), 920-931. <https://doi.org/10.1111/biom.12212>

Ahdika, A., & Lusiyana, N. (2017, February). Comparison of INAR (1)-Poisson model and Markov prediction model in forecasting the number of DHF patients in west java Indonesia. In *Journal of Physics: Conference Series* (Vol. 814, No. 1, p. 012002). IOP Publishing. <https://doi.org/10.1088/1742-6596/814/1/012002>

Asamoah, J., Jin, Z., Seidu, B., Odoro, F. T., Sun, G. Q., & Alzahrani, F. (2020). Mathematical modelling and sensitivity assessment of COVID-19 outbreak for Ghana and Egypt. Available at SSRN 3612877. <https://doi.org/10.2139/ssrn.3612877>

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 33(1), 38-44. <https://doi.org/10.2307/2347661>

Babaei, A., Jafari, H., Banihashemi, S., & Ahmadi, M. (2021a). Mathematical analysis of a stochastic model for spread of Coronavirus. *Chaos, Solitons & Fractals*, 145, 110788. <https://doi.org/10.1016/j.chaos.2021.110788>

Babaei, A., Jafari, H., Banihashemi, S., & Ahmadi, M. (2021b). A stochastic mathematical model for COVID-19 according to different age groups. *Applied and Computational Mathematics*, 20(1), 140-159.

Beath, K. J. (2017). Robmixglm: An R Package for Robust Analysis Based on Mixtures.

Christou, V., & Fokianos, K. (2014). Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis*, 35(1), 55-78. <https://doi.org/10.1111/jtsa.12050>

Davis, R. A., & Wu, R. (2009). A negative binomial model for time series of counts. *Biometrika*, 96(3), 735-749. <https://doi.org/10.1093/biomet/asp029>

Elsaied, H. (2021). Using Observation Driven Models to fit count daily new confirmed cases of “COVID 19” in Egypt. *Journal Advances and Applications in Statistics* Vol. 68, No. 1. <https://doi.org/10.17654/AS068010071>

Elsaied, H., & Fried, R. (2014). Robust fitting of INARCH models. *Journal of Time Series Analysis*, 35(6), 517-535. <https://doi.org/10.1111/jtsa.12079>

Elsaied, H., & Fried, R. (2021). On robust estimation of negative binomial INARCH models. *METRON*, 1-22. <https://doi.org/10.1007/s40300-021-00207-8>

Fahmy, A. E., Eldesouky, M. M., & Mohamed, A. S. (2020). Epidemic Analysis of COVID-19 in Egypt, Qatar and Saudi Arabia using the Generalized SEIR Model. medRxiv. <https://doi.org/10.1101/2020.08.19.20178129>

Ferland, R., Latour, A., & Oraichi, D. (2006). Integer-valued GARCH process. *Journal of time series analysis*, 27(6), 923-942. <https://doi.org/10.1111/j.1467-9892.2006.00496.x>

- Fokianos, K., Rahbek, A., & Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association* 104, 1430-1439. <https://doi.org/10.1198/jasa.2009.tm08270>
- Giuliani, D., Dickson, M. M., Espa, G., & Santi, F. (2020). Modelling and predicting the spread of coronavirus (covid-19) infection in nuts-3 italian regions. *arXiv preprint arXiv:2003.06664*. <https://doi.org/10.21203/rs.3.rs-20501/v2>
- Heinen, A. (2003). Modelling time series count data: an autoregressive conditional Poisson model. Available at SSRN 1117187. <https://doi.org/10.2139/ssrn.1117187>
- Jamaludin, A. R., Yusof, F., & Suhartono, S. (2020). Modelling Asthma Cases by Count Analysis Approach: Poisson INGARCH and Negative Binomial INGARCH. *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*, 36(1), 15-30. <https://doi.org/10.11113/matematika.v36.n1.1158>
- Khan, T., Zaman, G., & El-Khatib, Y. (2021). Modeling the dynamics of novel coronavirus (COVID-19) via stochastic epidemic model. *Results in Physics*, 24, 104004. <https://doi.org/10.1016/j.rinp.2021.104004>
- Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 209-225. <https://doi.org/10.2307/3314912>
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., ... & Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, 172(9), 577-582. <https://doi.org/10.7326/M20-0504>
- Martinez, E. Z., Silva, E. A. S. D., & Fabbro, A. L. D. (2011). A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. *Revista da Sociedade Brasileira de Medicina Tropical*, 44(4), 436-440. <https://doi.org/10.1590/S0037-86822011000400007>
- Promprou, S., Jaroensutasinee, M., & Jaroensutasinee, K. (2006). Forecasting Dengue Haemorrhagic Fever Cases in Southern Thailand using ARIMA Models.
- Ronchetti, E. (1997). Robustness aspects of model choice. *Statistica Sinica*, 327-338.
- Tharmaratnam, K., & Claeskens, G. (2013). A comparison of robust versions of the AIC based on M-, S- and MM-estimators. *Statistics*, 47(1), 216-235. <https://doi.org/10.1080/02331888.2011.568120>
- Xiong, L., & Zhu, F. (2019). Robust quasi-likelihood estimation for the negative binomial integer-valued GARCH (1, 1) model with an application to transaction counts. *Journal of Statistical Planning and Inference*, 203, 178-198. <https://doi.org/10.1016/j.jspi.2019.03.010>
- Zhu, F. (2011). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis*, 32(1), 54-67. <https://doi.org/10.1111/j.1467-9892.2010.00684.x>