ISSN 2581-3463

AJMS

Asian Journal of Mathematical Sciences

RESEARCH ARTICLE

On Performance of Integer-valued Autoregressive and Poisson Autoregressive Models in Fitting and Forecasting Time Series Count Data with Excess Zeros

Saleh Ibrahim Musa[1], N. O. Nwezeand[2], M. O. Adenomon[2]

[1]*Department of Statistics, Federal University of Lafia, PMB 146 Lafia, Nigeria,* [2]*Department of Statistics, Nasarawa State University Keffi, PMB 122 Keffi, Nigeria*

Received: 30-04-2021; Revised: 29-05-2021; Accepted: 26-06-2021

ABSTRACT

Time series data often entail counts. Time series count data, which refer to the number of times an item or an event occurs within a fixed period of time, are essential in many fields most of the works on time series count data do not exhaustively consider the effect excess zeros in modeling. This study, therefore, seeks to examine the performance integer-valued autoregressive (INAR) and Poisson autoregressive models on count data under the influence of excess zeros. The effect of sample sizes, n =30, 60,…, 300, on the performance of the models were also studied. At every sample size, the best status of the orders p and q where p, q = 1, 2 are, respectively, determined for the levels of the excess zeros through simulations. The predictive ability of the models was observed at h-steps ahead, h = 5, 10, 15,…, 50 for the models with excess zeros data structures. It was concluded that the best model to fit and forecast data with excess zeros is INAR at different sample sizes. The predictive abilities of the four fitted models increased as sample size and number of steps ahead were increased

Key words: Count data, excess zeros, integer-valued autoregressive, Poisson autoregressive

INTRODUCTION

Time series is the values of some statistical variables measured over a uniform set of time points. Examples of time series data are monthly sales in a store, monthly HIV/AIDS cases recorded in a hospital, yearly production by a company, daily number of eggs laid by fowls in a farm, consumption of electricity in kilowatts, and data on population motor registration per day. Time series data often entail counts, such as the number of road accidents, the number of patients in a certain hospital, and the number of customers waiting for service at a certain time. Count data can be found in many practical lifetime studies, such as the number of days before death in certain diseases or the number of cycles (runs) until a machine stops working and so on. Hence, a number of statistical distributions have been applied to model the case of a count random variable (RV) with a non-negative integer value. A good overview of these distributions can be found in Johnson *et al*. (2005).[9]

Research in economics, ecology, environmental sciences, medical, and public health-related fields, it is often practical that the pattern of outcomes is relatively infrequent behaviors. Data of this type consist of excess zeros. It has been reported that the traditional Poisson model provided the popular frame work for fitting count data but not suitable for time series data with excess zeros and further considered using integer-valued autoregressive (INAR) (1) model with restricted application to zeros and ones data (Qi *et al*., 2019).[21] Data with many zeros are usually frequent in research studies when counting the occurrence of certain behavioral events, such as number of purchases made, number of school absences, number of cigarettes smoked, or number of hospitalizations. These types of data are called count data and their values are usually non-negative with a lower bound of zero. Common issues when dealing with count data are typically zero inflation or excessive zeros (Akeyede *et al*., 2021).[1]

Address for correspondence:
Saleh Ibrahim Musa
E-mail: saleh.musa@science.fulafia.edu.ng

Indeed, there already exists a number of literatures, for instance, the work of Yusuf *et al*., 2018, compares the performance of hurdle Poisson, hurdle negative binomial, and zero negative inflated binomial (ZINB) on 2013 health survey data set, where he recommended that ZINB model outperformed others on excess zeros count data.[24] Qian *et al*., 2020, considered modeling of heavy tailed count time series data on number of traded stock in 5 min for interval Empire District Electric Company using heavy tailed probabilities, he further recommend the use of INAR of order p to analyze heavy tailed count time series data.[17] In the work of Hinde *et al*., 2001, the author uses Poisson regression and negative binomial regression models as standard methods for modeling count outcomes but the common problems associated with overdispersion and excess zeros were not accounted for in the work.[7] Siti and Jamil, 2017, used Poisson regression on data with excess zeros and discovered a presence of over-dispersion in analyzing zero values and finally recommended zero-inflated Poisson to be a better model for such type of data.[20]

To account for the excessive proportion of zeros, either the hurdle or zero-inflated model is used. Mullahy (1986) studied the hurdle model for univariate count data.[16] An extension for longitudinal or clustered count data with excessive zeros was considered by Min and Agresti (2005). A separate strand of literature is devoted to zero-inflated model. Lambert (1992) and Greene (1994) studied zero inflation for cross-sectional count data, and the multilevel extension was the focus of Lee *et al*. (2006). Min and Agresti (2005) and Lee *et al*. (2006) introduced two separate and possibly correlated subject-specific random effects, one in the count and the other in the zero inflation part.[2,5,10,15]

In this study, the analysis of data with excess zeros of count data was carried out using INAR model and Poisson autoregressive (PAR) models. Further, model selection criteria were also considered to give a simple and effective way of selecting a model for the excess zeros of the count time series process.

METHODOLOGY[6,8,12,23]

Data set was simulated in R statistical software with sample sizes of 30, 60, 90,… and 300, from Poisson and negative binomial distribution to produce generated series count data with Poisson excess zero and negative binomial excess zero, respectively. The two models under study, namely, INAR and PAR were fitted to the simulated data so as to examine the effect of the proportion of excess zero on their performance. Levels excess zero were imposed at different percent on the simulated data for observation yi in different data set generated, which were randomized and replicated 1000 times each for the respective selected sample sizes.

In simulation, we set our parameters to be $\emptyset_1=1\emptyset_2=1$ to ensure discrete nature of count data generated. The response Y_{ti} in (1) was generated from Poisson and negative binomial distributions. The two models under study were considered to analyze how well each of the model fits the selected data sets having some degree of excess zeros. To compare the forecasting accuracy of the model, a multicriterion performance evaluation procedure earlier mentioned will be used in this study. The model with the minimum criteria shall be considered as the best for the fitting and forecasting. Note that a number of steps ahead will be forecasted from each model.

Data were generated from linear second orders of autoregressive functions given as follows:

Model 1. AR (2): $Y_{ti}=0.2Y_{ti-1}+0.4Y_{ti-2}+e_t$ (1)
t=30, 60, 90, 120, 150, 180, 210, 240, 270, 300. i = 1, 2,…,1000

Where, Y_{ti} will be simulated from Poisson and negative binomial families for equidispersed excess zeros, respectively, as follows:

The basic count model is the Poisson regression model which is based on the Poisson distribution with probability density function

$$\frac{\lambda^{y_i} e^{-\lambda_i}}{y_i!}, for\ y_i = 0,1,2,\ldots \quad (2)$$

Thus, for the Poisson models $E(y_i)=V(y_i)=\mu_i$. The restrictive condition that the mean must equal the variance is often violated by excess zeros data (where variance may exceeds the mean).Since excess zero count data often violates the equality of mean and variance feature, Poisson model is generally considered inappropriate for count data, because count data contains excess zeros and are usually highly skewed and overdispersed (Cameron and Trivedi, 2008).[4]

The over-dispersion is achieved from the negative binomial distribution function given as follows;

$$p(y_i;\lambda_i,\alpha_i)=\frac{\Gamma\left(y_i+\frac{1}{\alpha_i}\right)}{\Gamma(y_i+1)\Gamma\left(\frac{1}{\alpha_i}\right)}\left(\frac{1}{1+\alpha_i\lambda_i}\right)^{\frac{1}{\alpha_i}}\left(\frac{\alpha_i\lambda_i}{1+\alpha_i\lambda_i}\right)^{y_i},i=1,\ldots,1000 \quad (3)$$

Here, the dispersion parameter $\alpha_i>0$, $\lambda_i=E(Y_i)$; and $V(Y_i)=\lambda_i+\alpha_i\lambda_i^2$

The negative binomial model can be used to impose the over-dispersion problem on by creating larger values of variances than means. Lawal (2011) argued that the negative binomial (NB) model might be a suitable alternative to the Poisson model, especially for excess zeros which lead to overdispersion in count data.[11] This is because the NB model in this case would account for the heterogeneity in the data by introducing the dispersion parameter α. To compare the modeling and forecasting accuracy of the models, AIC and BIC criteria for performance evaluation procedure were used in this study. The model with the minimum criteria values was considered as the best for the fitting and forecasting. Note that a number of steps ahead were forecasted from each model.

Zero inflated Poisson (ZIP) model

The ZIP regression model was introduced by Lambert (1992) for analyzing manufacturing data and investigating the number of defects in equipment.[10] Poisson models are mixed with zeros to allow for the excessive zeros in the data, which is commonly encountered in real life. The ZIP can be derived as follows:

$$f(y_i|x_i,z_i)=\begin{cases}\varnothing_i+(1-\varnothing_i)e^{-\mu_i}, & y_i=0\\ \dfrac{(1-\varnothing_i)\mu_i^{y_i}e^{-\mu_i}}{y_i!}, & y_i=1,2,\ldots\end{cases} \quad (4)$$

The parameter μ (>0) is the mean (and also the variance) of Poisson distribution. Then, for a sample y_i, $i = 1, 2,\ldots, n$, and within the framework of generalized linear model, this distribution is generalized by allowing to be related to a set of covariates $x_i = (x_{i1}, x_{i2}, \ldots x_{ip})$ with corresponding parameters α,through the log link function, such that;

$\mu_i=e\text{–}x'^{i\alpha}$ 2.1

This was chosen to ensure that μ_i remains positive in order that its predicted values will always be positive. Thus, the response variable y represents the frequencies of an event of interest. The x and z are the set of covariates with α coefficients, respectively.

Although the common practice in analyzing count data in many disciplines as widely considered is Poisson model being the most basic model, regrettably, the model reliance on a single parameter often limits its use on real data. This is due to the fact that, most data violated feature of Poisson distribution, which is sameness of mean and variance, known as equidispersed.[7] Van den Broek (1995) was first to proposed a score test so as to test whether the ZIP distribution should be used as an alternative to the ordinary Poisson distribution, the work was later extended to the setting where count data are correlated.[22] Despite the popularity of the ZIP models, the literature for time series count data with excess zeros is sparse.

Zero-inflated negative binomial (ZINB) model

ZINB models have been described as an extended version of the negative binomial models for excess zero count data[5] The ZINB model can be derived. Thus,

$$f(Y|y_t)=\begin{cases}\rho_t+(1-\rho_t)(1+k\mu_t)^{-1/k}, & y_t=0\\ (1-\rho_t)\dfrac{\tilde{A}\left(y_t+\frac{1}{k}\right)}{\tilde{A}(y_t+1)\tilde{A}\left(\frac{1}{k}\right)}\dfrac{(k\mu_t)^{y_t}}{(1+k\mu_t)^{y_t+1/k}}, & y_t>0\end{cases} \quad (5)$$

with probability ρ_t and κ as overall dispersion parameter.

The ZINB autoregression can be used to account for simultaneous excess zero and over-dispersion in many count time series data.

INAR model

In this work, we are interested in a special class models, the so-called INAR process introduced by

(McKenzie, 1985; Al-Osh and Alzaid, 1987).[2,14] The theoretical properties and practical applications of INAR and related processes have been discussed extensively in the literature. Silva *et al*. (2005) considered independent replications of count time series modeled by INAR (1) and proposed several estimation methods using the classical and Bayesian approaches in time and frequency domains.[18]

Point prediction for INAR (1) process

Suppose a non-negative integer-valued RV X and λ ∈ [0, 1], the generalized thinning operation which is denoted by "∘"is given by;

$$\lambda \circ X = \sum_{j=1}^{X} Y_j \quad (6)$$

Where, $\{Y_j\}$, j =1,…, X, is a sequence of independent and identically distributed non-negative integer-valued random variables, independent of X, with finite mean λ and variance σ^2. The sequence is known as the counting series of $\lambda \circ$ X. When $\{Y_j\}$ is a sequence of Bernoulli random variables, the thinning operation is called binomial thinning operation and was defined by Steutel and van Harn, 1979.[19] The well-known INAR(1) process $\{X_t; t = 0, \pm1, \pm2,\ldots\}$ is defined on the discrete support N_o by the equation

$$X_t = \lambda \circ X_{t-1} + \varepsilon_t \quad (7)$$

Where, $0 < \lambda < 1$, $\{\varepsilon_t\}$ is a sequence of independent and identically distributed integer-valued random variables, with $E[\varepsilon_t] = \mu_\varepsilon$ and $Var[\varepsilon_t] = \sigma_\varepsilon^2$.

Poisson autoregressive (PAR) model

The PAR (p) model can be defined as

$$P\left(\frac{Q_t}{s_t}\right) = \frac{s_t^{q_t} e^{-s_t}}{q_t!} \quad (8)$$

Where, s_t is the conditional mean of the linear autoregressive AR process with $E(q_t/Q_{t-1})$ in (15). This represents the measurement equation for the observed data

The one step ahead for the conditional PAR (p) model forecast is given by

$$E(q_{t+1}/Q_t) = s_{t+1/t} = \sum_{i=1}^{k} \rho_i s_{\frac{t}{t}-1} + (1 - \sum_{i=1}^{k} \rho_i)\mu \quad (9)$$

$$Var(q_{t+1}/Q_t) = \frac{1+\sigma_{t+1/t}}{\sigma_{t+1/t}} s_{t+1/t} \quad (10)$$

Where, ρ, δ, s_t, and σ_t are the optimized values of a PAR series, the induced covariance X_t has the $\mu = e(X_t^\delta)$ (Brandt and Williams, 2001).[3]

PAR (p) forecast density for the one step ahead distribution

The PAR (p) forecast density is given by

$$P(q_t/Q_{t-1}) = \int_\emptyset Pr(q_t/\emptyset_t) Pr(\emptyset_t/Q_{t-1}) d\emptyset) \quad (11)$$

$$= \int_\emptyset \frac{\emptyset_t^{q_t} e^{-\emptyset_t}}{q_t} \cdot \frac{e^{-\sigma_{t/t-1}\emptyset} \emptyset_t^{\sigma_{t/t-1}s_{t/t-1}} \sigma_t^{\sigma_{t/t-1}s_{t/t-1}}}{\tilde{A}(\sigma_{t/t-1}s_{t/t-1})}$$

$$= \frac{\tilde{A}(\sigma_{t/t-1}s_{t/t-1} + q_t)}{\tilde{A}(q_t+1)\tilde{A}(\sigma_{t/t-1}s_{t/t-1})} (\sigma_{t/t-1})^{\sigma_{t/t-1}s_{t/t-1}} \times (1+\sigma_{t/t-1})^{\sigma_{t/t-1}s_{t/t-1}+q_t} \quad (12)$$

This is a negative binomial distribution function with a gamma function Γ(·).

The forecast function for the conditional mean and variance of a PAR (p) series realizations is based on the optimized values of ρ, δ, s_t, and σ_t. The log-likelihood function for the PAR (p) model is given as

$$L(s_{t-1,}\sigma_{t-1}/q_t,\ldots,q_T;Q_{t-1}) = ln\prod_{t=1}^{T} P(q_t/Q_{t-1}) \quad (13)$$

$$= \sum_{t=1}^{T} ln\tilde{A}(\sigma_{t/t-1}s_{t/t-1} + q_t) - ln\tilde{A}(q_t+1) - ln\tilde{A}(\sigma_{t-1}s_{t-1}) + \sigma_{t-1}s_{t-1}\ln(\sigma_{t-1}) - (\sigma_{t-1}s_{t-1} + q_t) + \ln(1+\sigma_{t-1}) \quad (14)$$

Using the linear autoregressive equation

$$E(q_t/Q_{t-1}) = \sum_{i=1}^{p} \rho_i Q_{t-i} + \lambda \quad (15)$$

Where, ρ_i and λ are any real number values.

We can obtain AR (1) for q_t which yield PAR (1) model with a negative binomial predictive distribution, for order p can also be generated as well. There is no restriction for the linear AR process with respect to the density $P(q_t/Q_{t-1})$. The

q_t density choice resulted constraints to ρ_i and λ to require admissible values.

Theil's U statistics

Theil's U statistics is the relative accuracy measure that compares forecasted results with the results of forecasting with minimal historical data it also requires the deviations to give more weight to large errors and to exaggerate errors, which can help eliminate methods with large errors.[13] U>1 indicates that the forecasting technique is better than guessing, U = 1 indicates that the forecasting technique is as good as guessing, U<1 indicates that the forecasting technique is worse than guessing. Theil's U-statistic: The U-statistic developed by Theil (1966) is an accuracy measure that emphasizes the importance of large errors (as in MSE) as well as providing a relative basis for comparison with naïve forecasting methods.[21] Makridakis *et al*. (1998) have simplified Theil's equation to the form

$$U = \sqrt{\frac{\sum_{i=1}^{n-1}\left(\frac{\hat{Y}_{i+1} - Y_{i+1}}{Y_t}\right)^2}{\sum_{i=1}^{n-1}\left(\frac{\hat{Y}_{i+1} - Y_t}{Y_t}\right)^2}} \qquad (16)$$

[13]

DATA ANALYSIS AND RESULTS

The results of the analyses of simulated data from different levels of excess zeros and sample size through the Poisson and negative binomial distributions are presented in Tables 1-5. The tables revealed the relative performance of INAR and PAR models of different orders.

The fitted models' AIC values as reported in Table 1 show that INAR (4) has exhibited a linear pattern of performance across sample sizes and seem the best fitted followed by INAR (3). PAR (3) fitted best at low sample sizes below 90, overtaken by PAR (1) with sharp fall at sample sizes above 120, then PAR (4) above 120 with minimum values of AIC, as shown in [Figures 1a,1b,2a,3a and 4a].

The fitted models' BIC values as reported in Table 2 show that INAR (4) has exhibited a linear pattern of performance across sample sizes and seem to be the best fitted followed by INAR (3). PAR (4) fitted best at sample sizes below 270, overtaken by PAR (1) as sample sizes approach 300 with minimum values of BIC, as plotted in Figure 2b.

The average values of AIC of each model at various sample sizes when there are excess zeros are presented in Table 3. The best fitted among INAR models is INAR (4). Relatively, PAR (3) displays good trend pattern below sample sizes of

Table 1: AIC of models' performance for negative binomial excess zeros

Sample sizes	30	60	90	120	150	180	210	240	270	300
Models										
INAR (1)	110.963	172.337	218.014	253.497	281.400	303.262	320.100	332.638	341.417	346.855
INAR (2)	110.912	172.236	217.862	253.294	281.147	302.958	319.746	332.234	340.961	346.349
INAR (3)	110.801	172.013	217.528	252.849	280.590	302.290	318.967	331.343	339.960	345.236
INAR (4)	110.79	171.991	217.495	252.805	280.535	302.224	318.890	331.255	339.860	345.125
PAR (1)	53.383	147.743	186.573	16.483	325.911	422.383	510.818	563.182	645.428	684.645
PAR (2)	48.914	144.209	188.815	253.065	320.519	420.304	511.761	565.215	647.099	719.556
PAR (3)	48.325	142.280	188.623	253.362	320.535	420.289	508.944	563.643	644.759	715.474
PAR (4)	48.643	142.978	186.573	252.368	320.206	418.23	500.949	555.913	644.181	614.034

Table 2: BIC of models' performance for negative binomial excess zeros

Sample sizes	30	60	90	120	150	180	210	240	270	300
Models										
INAR (1)	116.568	180.715	228.013	264.647	293.442	316.033	333.489	346.561	355.811	361.670
INAR (2)	116.517	180.613	227.861	264.444	293.189	315.730	333.134	346.156	355.355	361.164
INAR (3)	116.406	180.391	227.528	263.999	292.633	315.062	332.355	345.266	354.353	360.051
INAR (4)	116.395	180.369	227.494	263.955	292.577	314.996	332.278	345.177	354.254	359.941
PAR (1)	58.987	156.121	196.572	257.633	337.954	435.155	524.207	577.104	659.822	699.46
PAR (2)	54.519	152.586	198.815	264.215	332.561	433.076	525.149	579.137	661.493	734.371
PAR (3)	53.93	151.157	195.622	262.512	332.578	433.061	522.332	577.566	659.153	730.289
PAR (4)	54.248	151.355	196.572	263.518	332.249	431.002	514.337	569.836	658.575	729.649

Table 3: AIC of models' performance for Poisson excess zeros

Sample sizes	30	60	90	120	150	180	210	240	270	300
Models										
INAR (1)	−10.283	−6.495	−8.7	18.675	28.079	25.507	30.514	25.277	23.143	15.096
INAR (2)	−10.848	−7.799	−17.248	−1.218	6.69	3.743	2.069	−2.826	−7.524	−18.891
INAR (3)	−11.189	−8.107	−17.557	−1.259	6.322	3.375	1.850	−3.638	−9.887	−21.671
INAR (4)	−11.779	−8.652	−17.569	−2.581	5.308	1.233	0.109	−5.260	−10.075	−21.867
PAR (1)	11.529	48.819	88.699	98.742	121.149	150.003	174.987	192.859	205.416	205.648
PAR (2)	11.534	63.08	59.638	82.147	109.641	132.779	146.591	165.773	181.205	193.866
PAR (3)	11.525	35.02	51.984	72.446	106.109	130.152	169.095	184.534	199.974	201.032
PAR (4)	11.530	36.323	56.551	74.811	131.067	157.96	164.504	177.54	214.33	199.189

Table 4: BIC of models' performance for Poisson excess zeros

Sample sizes	30	60	90	120	150	180	210	240	270	300
Models										
INAR (1)	−4.678	1.883	1.299	29.825	40.121	38.279	43.902	39.200	37.536	29.912
INAR (2)	−5.244	0.579	−7.249	9.932	18.733	16.515	15.457	11.097	6.870	−4.076
INAR (3)	−5.585	0.271	−7.558	9.891	18.364	16.147	15.238	10.285	4.507	−6.856
INAR (4)	−6.175	−0.274	−7.569	8.569	17.351	14.005	13.497	8.663	4.319	−7.052
PAR(1)	17.134	57.196	98.699	109.892	133.192	162.775	188.376	206.781	219.809	220.463
PAR(2)	17.139	71.457	69.637	93.297	121.683	145.551	159.98	179.695	195.598	208.681
PAR(3)	17.130	53.397	61.983	82.596	118.151	142.924	182.484	198.457	214.367	215.848
PAR(4)	17.135	44.7	66.55	85.961	143.109	170.732	177.892	191.463	228.724	214.004

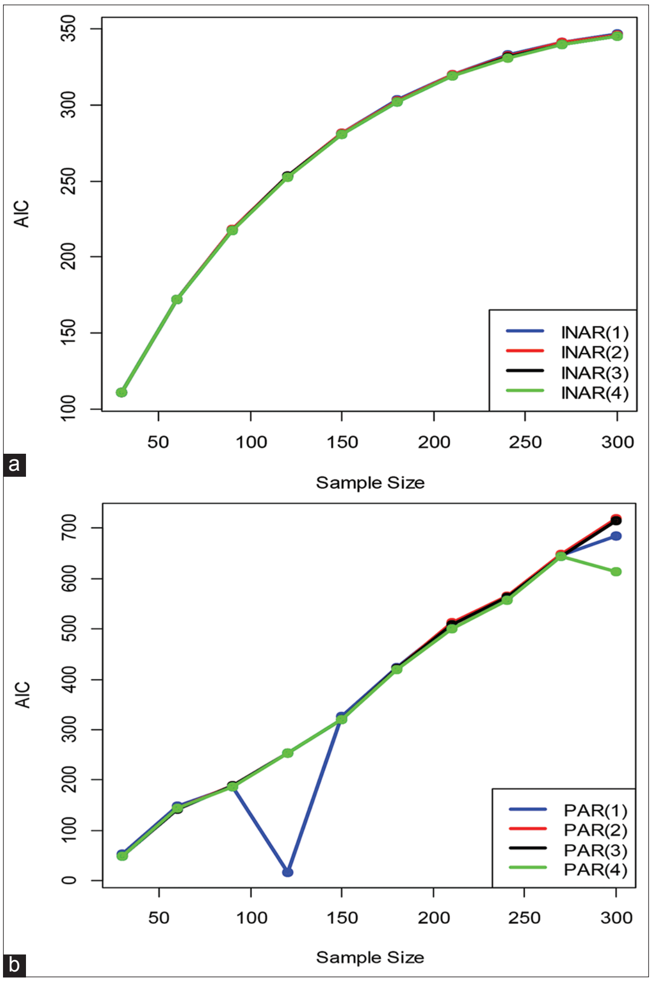

Figure 1: (a) AIC of the fitted INAR (p) models with negative binomial excess zeros. (b) AIC of the fitted PAR (p) models with negative binomial excess zeros

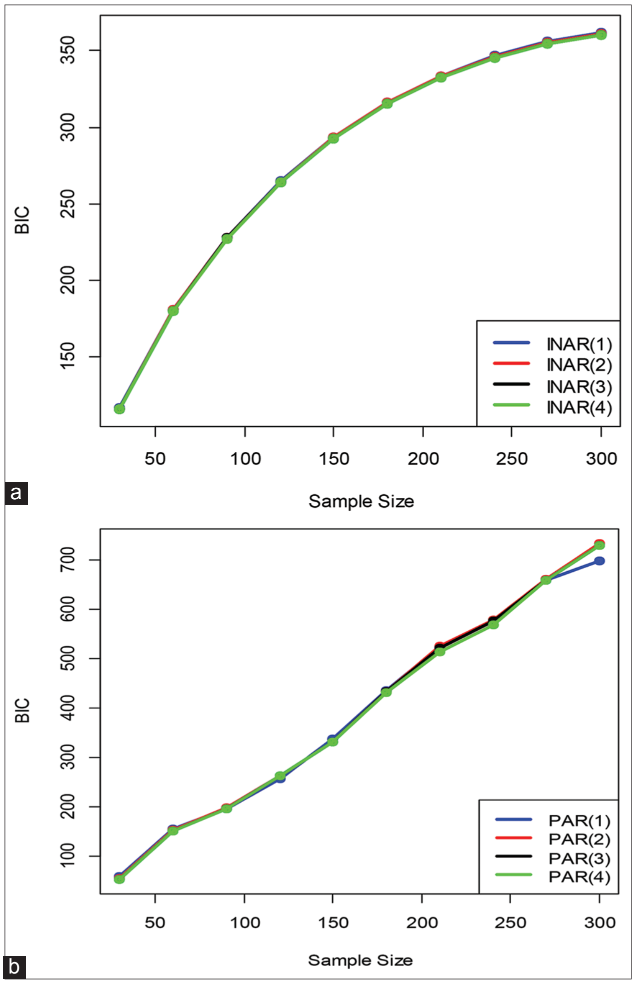

Figure 2: (a) BIC of the fitted INAR (p) models with negative binomial excess zeros. (b) BIC of the fitted PAR (p) models with negative binomial excess zeros

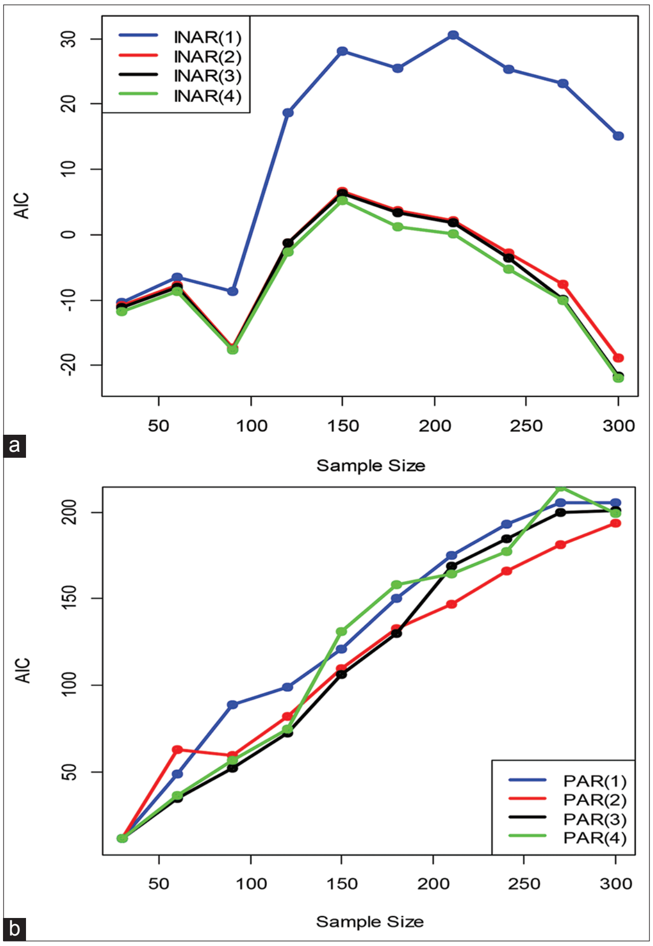

Figure 3: (a) AIC of the fitted INAR (p) models with Poisson excess zeros. (b) AIC of the fitted PAR (p) models with Poisson excess zeros

180, followed by PAR (2) in closed linear trend in Figure 3b based on the minimum reported AIC values.

The average values of BIC of each model at various sample sizes when there are excess zeros are presented in Table 4. The best fitted among INAR models is INAR (4) while in PAR models, PAR (4) performed well at low sample sizes below 60, PAR (3) displays good trend pattern below sample sizes of 210, followed by PAR (2) in closed linear trend at sample sizes above 210 as shown in Figure 4b based on the minimum reported BIC values.

Based on the Theil's analysis in Table 5, the ACP has the highest forecasting power due to their values greater than 1 and also greater than other values of the models across the steps ahead; this is followed by INAR and PAR. However, the Theil's values of PAR, at higher steps ahead are close to zero, hence it is not as good as other models in forecasting. Indeed, the forecasting ability of all the models decreases as steps ahead increases.

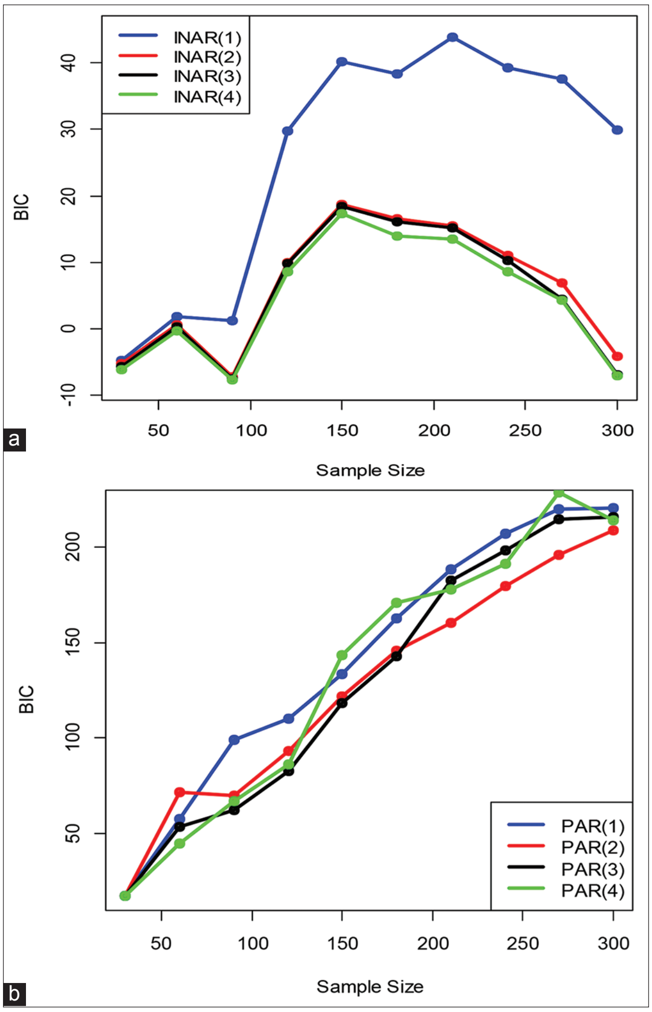

Figure 4: (a) BIC of the fitted INAR (p) models with Poisson excess zeros. (b) BIC of the fitted PAR (p) models with Poisson excess zeros

Table 5: Forecast performance of the models with zero-inflated negative binomial and zero-inflated Poisson using Theil's U statistic

Steps ahead	Zero-inflated Poisson		Zero-inflated negative binomial	
	INAR	PAR	INAR	PAR
5	2.0421	2.0031	2.4982	1.7847
10	2.0098	1.9542	2.3997	1.7262
15	1.9775	1.9184	2.3112	1.6677
20	1.9652	1.7915	2.2527	1.6092
25	1.9329	1.5646	2.1942	1.5507
30	1.9106	1.4337	2.1357	1.4922
35	1.8483	1.3908	2.0772	1.1136
40	1.816	1.3339	2.0187	0.9867
45	1.7837	1.3110	1.9602	0.7900
50	1.7514	1.1311	1.9017	0.6779

CONCLUSION

This study discovered that the highest performing model in fitting and forecasting different count time series data with different levels of excess zeros was the INAR model based on all criteria of assessment. The model has the speedy fitting capabilities at both high and low sample sizes compare to. PAR models have the slowest fitting speed across sample sizes. Specifically, the INAR (4) has the highest performance followed by INAR (3) among all the models in fitting any time series count data with the underlying features reported in this study.

RECOMMENDATION

Based on the findings of this study, the following recommendations were made.

i. INAR (4) can also be used for time series count data when sample size is small
ii. PAR (4) can be used as alternative for count data with excess zeros at all sample sizes

REFERENCES

1. Akeyede I, Bakari HR, Muhammad BR. Robustness of ARIMA and ACP Models too verdispersion in Analysis of Count data. J Niger Stat Assoc 2021;33:35-41.
2. Al-Osh MA, Alzaid AA. First-order integer-valued autoregressive (INAR(1)) process. J Time Ser Anal 1987;8:261-75.
3. Brandt P, Williams J. A linear Poisson autoregressive model: The Poisson AR(p) model. Pol Anal 2001;9:164-84.
4. Cameron CA, Trivedi PK. Regression Analysis of Count Data Econometric Society Monographs No. 30. New York: Cambridge University Press; 2008. p. 96-97, 221.
5. Greene W. Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models, Working Paper ec-94-10. New York: Department of Economics, New York University; 1994. p. 9-10.
6. Hasmida H. Water quality Trend at the Upper Part of Johor River in Relation to Rainfall and Runoff Pattern. Malaysia: University of Technology Malaysia; 2009.
7. Hinde J, Ridout M, Demétrio CG. A score test for testing a zero-inflated Poisson regression model againstzero-inflated negative binomial alternatives. Biometrics 2001;57:219-23.
8. Hilbe JM. *Modeling Count Data*. Cambridge: Cambridge University Press; 2014.
9. Johnson NL, Kemp AW, Kotz S. Univariate Discrete Distribution. New York: Wiley; 2005.
10. Lambert D. Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics 1992;34:1-14.
11. Lawal HB. On zero-truncated generalized poisson count regression model. Pioneer J Theoret Appl Stat 2011;1:16.
12. Lee AH, Wang K, Scott J, Yau KK, McLachlan GJ. Multilevel zero-inflated Poisson regression modelling of correlated count data with excess zeros. Stat Methods Med Res 2006;15:47-61.
13. Makridakis S, Wheelwright SC. Forecasting Methods and Applications. New York: John Wiley & Sons; 1998.
14. McKenzie E. Some simple models for discrete variate time series. Water Resour Bull 1985;21:645-50.
15. Min Y, Agresti A. Random effect models for repeated measures of zero-inflated count data. Stat Model 2005;5:1-19.
16. Mullahy J. Specification and testing of some modified count data models. J Econ 1986;33:341-65.
17. Qian L, Li Q, Zhu F. Modelling heavy-tailedness in count time series. Appl Mathe Model 2020;82:766-84.
18. Silva I, Silva ME, Pereira I, Silva N. Replicated INAR(1) process, Methodol Comput Appl Probabil 2005;7:517-42.
19. Jamil SA, Abdullah MA, Long KS, Nor ME, Mohamed M, Ismail N. Detecting overdispersion in count data: A zeroinflated Poisson regression analysis. IOP Conf Ser J Phys 2017;890:012170.
20. Steutel FW, Harn KV. Discrete analogues of self-decomposability and stability. Ann Probab 1979;7:893-9.
21. Theil H. Applied Economic Forecasting. Chicago: Rand McNally; 1966.
22. Van den Broek J. A score test for zero inflation in a Poisson distribution. Biometrics 1995;51:738-43.
23. Qi X, Li Q, Zhu F. Modelling time series of count with excess zeros and ones based on INAR (1) model with zero-and-one inflated Poisson innovations. J Comput Appl Mathe 1995;356:572-90.
24. Yusuf O, Afolabi RF, Agbaje AS. Modelling excess zeros in count data with application to antenatal care utilisation. Int J Stat Probabil 2018;7:22-35.