Eur. J. Stat. 6 (2026) 11
doi: 10.28924/ada/stat.6.11

Zero Inflated Poisson Xgamma INAR(1) Model and Its Applications

Nuzhat Ul Mohi U Din, Peer Bilal Ahmad*

Department of Mathematical Sciences, Islamic University of Science and Technology, Kashmir, India
rashijiya85@gmail.com, bilalahmadpz@gmail.com
[*] Correspondence: bilalahmadpz@gmail.com

Abstract. The first order non-negative integer valued autoregressive process is widely used for counts of events over time. It is obvious that if the innovations are Poisson distributed, then the marginal model is Poisson. However, this marginal process may not hold for the overdispersed count data. An example of one common type of overdispersion is when zero counts are more likely than claimed by the Poisson marginal model. In this regard, we introduce a zero inflated version of a discrete compound model and developed its first integer autoregressive process using zero inflated Poisson Xgamma (ZIPXG) innovations. We derive different structural properties of the INAR(1) model and estimated the parameters by conditional maximum likelihood, Yule-Walker and Bayesian methods. Further a simulation study has been performed to investigate the behavior of these estimated parameters. To illustrate the utility of the newly proposed INAR(1) model, a time series data on submissions to animal health laboratories has been analyzed.

1. Introduction

Time series models hold a prominent place for analyzing time dependent data. In particular, when the data are from a random count process, we are in the context of integer valued time series which preserves the discrete nature of the related data. Such a time series model can be motivated in many contexts of research, including epidemiology, economics, social sciences, insurance, and environmental studies. Traditional Gaussian autoregressive moving average (GARMA) models have limitations when modeling integer valued time series, especially if we deal with overdispersion. Integer valued autoregressive (INAR) have been put forward as a possible direction. INAR models preserve the non-negative and discrete nature of the count variable but can imply temporal dependence. The integer valued autoregressive INAR process based on the binomial thinning operator of [20] which was presented first as an innovative process by [14] & [2] has emerged as a topic of interest in the literature over the last decade. Innovations are considered to follow a Poisson because the Poisson is the classical and natural distribution

Received: 7 Apr 2026.

Key words and phrases. Anorexia; Bayesian estimation; Forecasting; Integer valued autoregressive process; Poisson Xgamma.

for discrete data when the data is equidispersed (i.e., variance equals mean). The INAR(1) process with compound Poisson innovations was first presented by [18]. The count time series with compound Poisson innovations offers the possibility of accounting for some features of equidispersion, underdispersion, and overdispersion. [1] proposed the INAR(1) process with Geometric innovation. A novel mixture of Geometric INAR(1) processes was proposed by [10] to model overdispersed count time series characterized by very high frequencies of zeros and ones. Their model effectively overcomes the limitations of existing variants such as the Geometric and Poisson INAR(1), which tend to mis-estimate zero and one counts and poorly capture heavy tails. [8] explored the discrete pseudo-Lindley INAR(1) process, providing new theoretical developments, on the other hand, [13] presented an INAR(1) model using Poisson–extended exponential innovations, which offered a wider range of specification flexibility. One of the previous studies that worked on INAR(1) processes with varied innovations by [16] have given a full comparison of several estimation and forecasting methods, all of which were given to the zero inflated Geometric INAR(1) model with random coefficients. In general, Nasirzadeh and Bakouch's study is a superior work that brings new or expanded frameworks to understand and estimate INAR(1) processes to account for overdispersion, excess zeros, and autocorrelation. [7] analyzed INAR(1) models with discrete mixture exponential innovations using survival discretization. [3] managed to advance the Geometric distribution framework, becoming one of the first authors contributing to its extension by featuring INAR(1) processes with alternative innovation distributions, most notably Poisson, Poisson Lindley, and modified Poisson. In the same year, [11] provided an extended INAR(1) based on 3-parameter Lindley innovation, which helped expand the diversity of INAR innovations in order to include aspects of time, or at least the severity of an unusual event.

Count data often includes a greater proportion of zeros, known as zero inflation. An example of zero inflation can be observed in hydrologic data collected from arid and semiarid areas, such as the amounts of precipitation during dry months or the annual peak flow discharges. Data such as these can be analyzed using zero inflated statistical models, which include Poisson and Negative Binomial distributed models, etc. Zero inflation is observed across many scientific disciplines, which include manufacturing defects, medical visits, hydrology, road safety, spatial ecology and econometrics. Not accounting for zero inflation during analysis will lead to two main problems: first, we estimate parameters and standard errors to be biased, and secondly, the grossly excessive number of zeros leads to overdispersion [22]. Models for zero inflation have been widely studied within regression contexts, see [17] and however, they have not gained traction in a space and time series setting, even if count data with excessive numbers of zeros is usually highly autocorrelated and typically modeled as an INAR(1) process. Furthermore, most of the work done on integer valued time series models only focuses on non-negative integer values, excluding any zeros. Therefore, further integer valued time series models are needed besides those that have been developed. In this paper, a novel zero inflated Poisson Xgamma distribution is developed to manage overdispersion and excess zero counts. This distribution provides a maximum likelihood inference and model selection (AIC/BIC) criteria, as well as an extension of the time series. We directly use the zero inflated Poisson Xgamma marginal distribution to model the observed INAR(1) time series X_t with operator $\circ$, and also characterize the distribution of the INAR(1) innovations. Therefore, the approach we put forward is different from the work of [8] and [9], in which the authors specify an INAR(1) model and assume a particular distribution of the innovations. In this paper, we introduce a new INAR(1) with ZIPXG innovations called the ZIPXGINAR(1) model. The primary contributions of this work are as follows:

(1) We introduce a new zero inflated Poissson Xgamma distribution and study its statistical properties.
(2) We introduce the ZIPXGINAR(1) model by adopting the ZIPXG distribution as the innovation component within the INAR(1) framework.
(3) We investigate and compare three parameter estimation methods, Conditional Maximum Likelihood (CML), Yule-Walker (YW) and Bayesian method for the newly proposed model.
(4) Through simulation studies, we demonstrate that the model can replicate the characteristics of overdispersed count time series data.
(5) We apply the ZIPXGINAR(1) process to a real count time series exhibiting an excess of zeros and overdispersion, and we compare the fit and predictive accuracy with the existing INAR(1) models.

2. Zero inflated Poisson Xgamma Distribution

A random variable X is said to follow a zero inflated Poisson Xgamma (ZIPXG) distribution with parameters α and θ, if its probability mass function (pmf) is expressed as

$$P[X=x]=\begin{cases}\alpha+\dfrac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Big[2(\theta+1)^2+2\theta\Big], & x=0,\\[2ex] \dfrac{(1-\alpha)\theta^2}{2(\theta+1)^{4+x}}\Big[2(\theta+1)^2+\theta(1+x)(2+x)\Big], & x=1,2,3,...\end{cases} \tag{1}$$

or equivalently,

$$P[X=x]=\alpha I_{(0)}(x)+(1-\alpha)\frac{\theta^2}{2(\theta+1)^{4+x}}\Big[2(\theta+1)^2+\theta(1+x)(2+x)\Big], x=0,1,2,...$$

where $0\le\alpha<1$, $\theta>0$, and the indicator function $I_A(x)$ is equal to 1 if $x\in A$, else equals to 0. Also, the pmf plots of zero inflated Poisson Xgamma distribution are shown in Figure 1a.

2.1. Moment Generating Function.

The moment generating function (mgf) of X is

$$\alpha+\frac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Big[2(\theta+1)^2+2\theta+2(\theta+1)^2\Big(\frac{e^t}{1+\theta-e^t}\Big)+2\theta e^t\Big(\frac{e^{2t}-3e^t(1+\theta)+3(1+\theta)^2}{(1+\theta-e^t)^3}\Big)\Big]. \tag{2}$$

Proof: We have,

$$M_x(t) = E(e^{tx}) \quad = \quad \sum_{x=0}^{\infty} e^{tx} P[X = x] = P[x = 0] + \sum_{x=1}^{\infty} e^{tx} P[X = x]$$

$$= \quad \alpha + \frac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Big[2(\theta+1)^2 + 2\theta\Big] + \sum_{x=1}^{\infty} e^{tx} \frac{(1-\alpha)\theta^2}{2(\theta+1)^{4+x}}\Big[2(\theta+1)^2 + \theta(1+x)(2+x)\Big]$$

$$= \quad \alpha + \frac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Big[2(\theta+1)^2 + 2\theta + \sum_{x=1}^{\infty}\Big(\frac{e^t}{\theta+1}\Big)^x 2(\theta+1)^2 + \sum_{x=1}^{\infty}\Big(\frac{e^t}{\theta+1}\Big)^x \theta(1+x)(2+x)$$

$$\Rightarrow M_x(t) = \alpha + \frac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Big[2(\theta+1)^2 + 2\theta + 2(\theta+1)^2\Big(\frac{e^t}{1+\theta-e^t}\Big)$$
$$+ 2\theta e^t\Big(\frac{e^{2t} - 3e^t(\theta+1) + 3(\theta+1)^2}{(1+\theta-e^t)^3}\Big)\Big],$$

on differentiating equation (2), we get

$$M_x'(t) = \frac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Bigg[2(\theta+1)^2 + 2\theta + 2(\theta+1)^3\Big(\frac{e^t}{(1+\theta-e^t)^3}\Big) + \frac{6e^{3t}(\theta+1)}{(1+\theta-e^t)^4}$$
$$-\frac{6e^{2t}\theta(\theta+1)(2\theta+e^t+2)}{(1+\theta-e^t)^4} + \frac{6e^t\theta(\theta+1)^2(\theta+2e^t+1)}{(\theta-e^t+1)^4}\Bigg],$$

at t=0,

$$M_x'(t) = \frac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Bigg[2(\theta+1)^2 + 2\theta + 2(\theta+1)^3\Big(\frac{1}{\theta^2}\Big) + \frac{6\theta(\theta+1)}{\theta^4} -$$
$$\frac{6\theta(\theta+1)(2\theta+3)}{\theta^4} + \frac{6\theta(\theta+1)^2(\theta+3)}{\theta^4}\Bigg]$$

$$\Rightarrow M_x'(t=0) = (1-\alpha)\frac{3+\theta^2}{\theta(\theta+1)}.$$

So, the mean and variance of X are

$$E[X] = (1-\alpha)\frac{3+\theta^2}{\theta(\theta+1)},$$

and

$$Var[X] = (1-\alpha)\frac{3(1+5\theta+\theta^3)-\theta^2}{\theta^2(\theta+1)^2} + \alpha(1-\alpha)\Big[\frac{3+\theta^2}{\theta(\theta+1)}\Big]^2,$$

respectively. Clearly, if $\alpha > 0$, the ZIPXG (α, θ) is overdispersed $(\sigma_x^2 > \mu_x)$. In this case the ZIPXG distribution is actually a mixture of a degenerate distribution at zero when $\alpha = 1$, and when $\alpha = 0$, X reduces to the base Poisson Xgamma distribution. Also, as long as $P(x = 0) \geq 0$, the lower bound of α should be limited so that probabilities remain non-nagative.

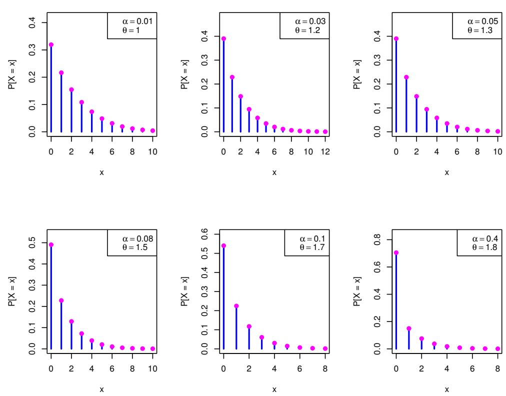
Index of Dispersion (IOD) is given as

$$IOD \quad = \quad \frac{(1-\alpha)\frac{3(1+5\theta+\theta^3)-\theta^2}{\theta^2(\theta+1)^2} + \alpha(1-\alpha)\Big[\frac{3+\theta^2}{\theta(\theta+1)}\Big]^2}{(1-\alpha)\frac{3+\theta^2}{\theta(\theta+1)}}$$

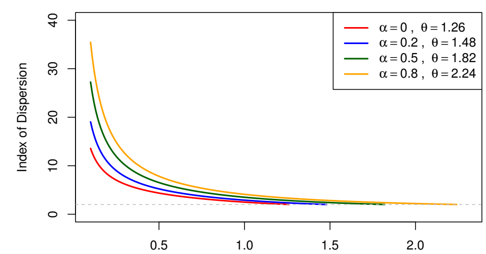
$$\Rightarrow IOD \quad = \quad \frac{3(1+5\theta+\theta^3)-\theta^2}{(3+\theta^2)\theta(\theta+1)} + \frac{\alpha(3+\theta^2)}{\theta(\theta+1)}.$$

The IOD plots of zero inflated Poisson Xgamma distribution are given in Figure 1b.

It is clear from the Figure 1b that the ZIPXG distribution shows strong overdispersion, especially for lower values of θ and higher values of α. Overdispersion occurs when the variance of the distribution is greater than the mean, as indicated by an index of dispersion that is much greater than 1. For the choice of parameter values indicated, the index of dispersion that is greater than 1 suggests that the ZIPXG distribution is very versatile for count data with excess zeros and high variability. Thus, the ZIPXG model is best for modeling count data with excess zeros and overdispersion.



(A) The pmf plots of ZIPXG model



(B) IOD plots of ZIPXG

FIGURE 1. pmf and iod plots of ZIPXG model.

2.2. **Probability Generating Function.** The Probability generating function (PGF) is an important technique when dealing with discrete random variables, as it explains the distribution of $X+Y$ when they are independent.

The PGF of ZIPXG distribution is

$$P_x(t) = \alpha + \frac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Big[2(\theta+1)^2 + 2\theta + 2(\theta+1)^2\Big(\frac{t}{\theta+1-t}\Big) + \theta\Big(\frac{6t(\theta+1)^2 - 6t^2(\theta+1) + 2t^3}{(\theta+1-t)^3}\Big)\Big].$$

Similarly, the characteristic function ($\phi_x(t)$) of ZIPXG is obtained by replacing e^t with $e^{\iota t}$ in above equation, and is given as

$$\phi_x(t) = \alpha + \frac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Big[2(\theta+1)^2 + 2\theta + 2(\theta+1)^2\Big(\frac{e^{\iota t}}{\theta+1-e^{\iota t}}\Big) + 2\theta e^{\iota t}\Big(\frac{e^{2\iota t} - 3e^{\iota t}(\theta+1) + 3(\theta+1)^2}{(\theta+1-e^{\iota t})^3}\Big)\Big].$$

3. Parametric Estimation

To estimate the parameters of the ZIPXG distribution, we have used the Maximum Likelihood Estimation (MLE) method as follows

3.1. **Maximum Likelihood Estimation (MLE).** Let's have a random sample of variables $x_1, x_2, ...x_n$ from ZIPXG distribution as defined in equation (1) and $i = 1, 2, 3, .., n$

$$a_i = \begin{cases} 1; & if\ x = 0, \\ \\ 0; & otherwise, \end{cases}$$

then for $i = 1, 2, 3, ..n$, $p[X = x]$ can be written as

$$P[X = x_i] = \left(\alpha + \frac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Big[2(\theta+1)^2 + 2\theta\Big]\right)^{a_i}$$
$$\left(\frac{(1-\alpha)\theta^2}{2(\theta+1)^{4+x_i}}\Big[2(\theta+1)^2 + \theta(1+x_i)(2+x_i)\Big]\right)^{1-a_i}.$$

Thus, the likelihood function $L(\alpha, \theta, x_1, x_2, ...x_n)$ will be

$$L = \prod_{i=1}^{n}\left[\left(\alpha + \frac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Big[2(\theta+1)^2 + 2\theta\Big]\right)^{a_i}\right.$$
$$\left.\left(\frac{(1-\alpha)\theta^2}{2(\theta+1)^{4+x_i}}\Big[2(\theta+1)^2 + \theta(1+x_i)(2+x_i)\Big]\right)^{1-a_i}\right]$$

$$\Rightarrow L = \left(\alpha + \frac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Big[2(\theta+1)^2 + 2\theta\Big]\right)^{\sum_{i=1}^{n} a_i}$$
$$\prod_{i=1}^{n}\left(\frac{(1-\alpha)\theta^2}{2(\theta+1)^{4+x_i}}\Big[2(\theta+1)^2 + \theta(1+x_i)(2+x_i)\Big]\right)^{1-a_i}$$

$$\Rightarrow L = \left(\alpha + \frac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Big[2(\theta+1)^2 + 2\theta\Big]\right)^{\sum_{i=1}^{n} a_i}\Big(1-\alpha\Big)^{n-\sum_{i=1}^{n} a_i}$$
$$\prod_{i=1}^{n}\left(\frac{\theta^2}{2(\theta+1)^{4+x_i}}\Big[2(\theta+1)^2 + \theta(1+x_i)(2+x_i)\Big]\right)^{1-a_i},$$

put $\sum_{i=1}^{n} a_i = m_0$ & $1 - a_i = l_i$ in above equation. Therefore,

$$log\ L = m_0\ log\left(\alpha + \frac{(1-\alpha)\theta^2}{2(\theta+1)^4}\Big[2(\theta+1)^2 + 2\theta\Big]\right) + (n-m_0)log(1-\alpha)$$
$$+ log\sum_{i\geq 1} l_i\left(\frac{\theta^2[2(\theta+1)^2 + \theta(1+x_i)(2+x_i)]}{2(\theta+1)^{4+x_i}}\right).$$

In order to find estimates of $\alpha\,\&\,\theta$, we will differentiate above equation with respect to $\alpha\,\&\,\theta$ and equate to '0', respectively, & let $\theta^2\Big[2(\theta+1)^4\Big(2(\theta+1)^2 + 2\theta\Big)\Big] = j$, in the above equation, we have

$$\begin{aligned} log\ L =& m_0\ log\Big(\alpha + (1-\alpha)j\Big) + (n-m_0)log(1-\alpha) \\ &+ log\sum_{i\geq 1} l_i\left(\frac{\theta^2[2(\theta+1)^2+\theta(1+x_i)(2+x_i)]}{2(\theta+1)^{4+x_i}}\right) \\ &\Rightarrow \frac{\delta\ logL}{\delta\,\alpha} = \frac{m_0(1-j)}{\alpha(1-j)+j} + \frac{m_0-n}{\alpha-1} = 0 \\ &\Rightarrow \alpha = \frac{j(n-m_0)}{(1-j)(2m_0-n)}. \end{aligned}$$

Now, on replacing actual value of j in the above equation, we will get the estimate of α as follows

$$\hat{\alpha} = \frac{\Big(2(\theta+1)^4 m_0\Big) - \Big(\theta^2[2(\theta+1)^2+2\theta][2m_0-n]\Big)}{\Big(2(\theta+1)^4 m_0 - \theta^2[2(\theta+1)^2+2\theta]\Big)\Big(2m_0-n\Big)}.$$

Similarly,

$$\begin{aligned} \frac{\delta\ logL}{\delta\,\theta} =& \frac{(1-\alpha)(\theta^3+7\theta^2+2\theta)}{\alpha(\theta+1)(\theta^4+4\theta^3+6\theta^2+4\theta+\alpha)} + \frac{1}{\theta+1} \\ &+ \frac{(2\theta^4+4\theta^3+2\theta^2)(1+x_i)(2+x_i)}{2\theta^2(\theta+1)^2+\theta(1+x_i)(2+x_i)}. \end{aligned}$$

Since the equation is not in closed form, thus R software will be used to estimate θ.

4. Testing

In this section, the statistical significance of the inflation parameter α was tested using the likelihood ratio test.

4.1. **Likelihood Ratio Test (LRT).**

To test the significance of the α, the likelihood ratio test is carried to differentiate between ZIPXG (α,θ) and PXG (θ). The decision on the model fit is based upon the following null and alternative hypothesis

$$H_0 : \alpha = 0\ vs\ H_1 : \alpha \neq 0$$

and the likelihood ratio test statistic is given as

$$-2ln\zeta = 2(l_1 - l_2),$$

where, $l_1 = lnL(\hat{\alpha},\hat{\theta} : x)$, here $\hat{\alpha},\hat{\theta}$ are the maximum likelihood estimates and $l_2 = lnL(\hat{\theta^*} : x)$, θ^* is the MLE for θ under null hypothesis.

Since the critical value at 5% (d.f.=1) level of significance is 3.84. It is clearly seen from the Table (1), that the calculated values are greater than the critical values which results in the rejection of null hypothesis in both the data sets. Thus the inflation parameter of our model ZIPXG is significant.

TABLE 1. Calculated value of Test Statistic

Data set	$lnL(\hat{\theta}^* : x)$	$lnL(\hat{\alpha}, \hat{\theta} : x)$	Test Statistic
1	496.026	492.01	8.032
2	1593.374	1442.05	308.648

4.2. **Wald Test.** In this test, the significance of inflation parameter α is checked by using the following test statistics

$$W_\alpha = \Big[\frac{\hat{\alpha}}{\sqrt{Var(\hat{\alpha})}}\Big]^2,$$

and the null hypothesis is

$$H_0 : \alpha = 0 \ vs \ H_1 : \alpha \neq 0$$

Here the variances have been calculated using R software, which are $Var(\hat{\alpha})$= 0.004481 and $Var(\hat{\alpha})$=0.000445 for data set 1 and 2 respectively. Since, the critical value of $\chi^2(1, 0.05)$=3.84.

TABLE 2. Test Statistic of Wald Test

Data Set	Test Statistic
1	10.765
2	836.598

From Table (2), it is clear that the calculated values for both data sets are greater than critical value. Thus, we conclude that the null hypothesis is rejected in both data sets and the inflation parameter is significant.

5. SIMULATION

This section represents the simulation study of the ZIPXG model. This simulation study is conducted to evaluate the finite sample behaviour of parameters of the ZIPXG model for different sample sizes (n = 25, 75, 100, 300, 600). Four parametric combinations $[(\alpha = 0.2, \theta = 2.0), (\alpha = 0.3, \theta = 1.0), (\alpha = 0.4, \theta = 0.8), (\alpha = 0.6,\ \theta = 1.5)]$ were used to conduct this study. Table (3) provides the outcomes of this study. The table demonstrates that Bias, Variance and MSE all tend to '0' as sample size increases.

Table 3. simulation results for ZIPXG model

Sample Size	Parameter	($\alpha = 0.2,\ \theta = 2.0$)			($\alpha = 0.3,\ \theta = 1.0$)		
		Bias	Var	MSE	Bias	Var	MSE
25	α	0.0090	0.0417	0.0417	0.0306	0.0276	0.0285
	θ	0.2136	0.9114	0.9553	0.0791	0.1148	0.1209
75	α	0.0031	0.0222	0.0222	0.0170	0.0098	0.0098
	θ	0.0493	0.1981	0.2002	0.0359	0.0296	0.0309
100	α	-0.0099	0.0179	0.0180	0.0152	0.0072	0.0074
	θ	0.0483	0.1555	0.1576	0.0265	0.0236	0.0242
300	α	-0.0060	0.0071	0.0072	0.0031	0.0023	0.0023
	θ	0.0226	0.0530	0.0534	0.0012	0.0058	0.0058
600	α	-0.0009	0.0040	0.0040	0.0020	0.0012	0.0012
	θ	0.0106	0.0285	0.0286	0.0010	0.0028	0.0028
Sample Size	Parameter	($\alpha = 0.4,\ \theta = 0.8$)			($\alpha = 0.6,\ \theta = 1.5$)		
		Bias	Var	MSE	Bias	Var	MSE
25	α	-0.0267	0.0223	0.0229	0.0702	0.0529	0.0577
	θ	0.0805	0.0784	0.0847	0.6183	0.3530	0.7286
75	α	-0.0141	0.0076	0.0076	0.0236	0.0122	0.0128
	θ	0.0248	0.0191	0.0197	0.1130	0.1835	0.1959
100	α	-0.0056	0.0058	0.0058	-0.186	0.0103	0.0107
	θ	0.0099	0.0127	0.0128	0.0839	0.1761	0.1828
300	α	-0.0022	0.0018	0.0018	-0.0048	0.0030	0.0031
	θ	0.0051	0.0038	0.0038	0.0289	0.0448	0.0455
600	α	-0.0000	0.0009	0.0009	-0.0017	0.0015	0.0015
	θ	0.0014	0.0018	0.0018	0.0127	0.0208	0.0209

6. INAR(1) process with Zero Inflated Poisson Xgamma innovations

The INAR(1) process is defined by the following recursive relationship,

$$X_t = p \circ X_{t-1} + \epsilon_t,$$

where, X_t is the count at time t, $p \in [0, 1]$ is the autoregressive parameter that predicts that the percentage of the count in the past will carry over to the present. The operator $\circ$ is the binomial thinning operator used to keep the integer nature in the process. ϵ_t is the innovation term at time t. It is usually simulated with a Poisson or Negative Binomial distribution. In this study, we propose an INAR(1) model with zero inflated Poisson Xgamma innovations called ZIPXGINAR(1). We presented some structural and mathematical features of the marginal distribution, including mean, variance, autocovariance, and (conditional) likelihood.

6.1. **Some mathematical properties.** The mean and variance of X_t are

$$E(X_t) = \mu = \frac{\mu_\epsilon}{1-p},$$

$$\mu = \frac{(1-\alpha)(\theta^2+3)}{(1-p)\theta(\theta+1)},$$

and

$$Var(X_t) = \sigma^2 = \frac{p(1-p)\mu + \sigma_\epsilon^2}{1-p^2},$$

$$\sigma^2 = \frac{1-\alpha}{(1-p^2)\theta(\theta+1)}\Big[p(\theta^2+3) + \frac{3(1+5\theta+\theta^3) - \theta^2 + \alpha(3+\theta^2)^2}{\theta(\theta+1)}\Big].$$

Also, the dispersion index of X_t is

$$\begin{aligned} DI &= \frac{Var(X_t)}{E(X_t)} \\ &= \frac{1}{(1+p)}\Big[p + \frac{3(1+5\theta+\theta^3) - \theta^2 + \alpha(3+\theta^2)^2}{\theta(\theta+1)(3+\theta^2)}\Big]. \end{aligned}$$

Using the results of [2], the conditional expectation and variance are given as

$$\begin{aligned} E[X_t|X_{t-1}] &= E[p \circ X_{t-1} + \epsilon_t | X_{t-1}] \\ &= E[p \circ X_{t-1}|X_{t-1}] + E[\epsilon_t|X_{t-1}] \\ \Rightarrow E[X_t|X_{t-1}] &= pX_{t-1} + (1-\alpha)(3+\theta^2)[\theta(1+\theta)]^{-1}, \end{aligned}$$

and

$$\begin{aligned} Var[X_t|X_{t-1}] &= Var[p \circ X_{t-1} + \epsilon_t|X_{t-1}] \\ &= V[p \circ X_{t-1}|X_{t-1}] + V[\epsilon_t|X_{t-1}] \\ &= p(1-p)X_{t-1} + (1-\alpha)(3(1+5\theta+\theta^3) - \theta^2)[\theta(\theta+1)]^{-2} \\ &\qquad + \alpha(1-\alpha)(3+\theta^2)^2[\theta(\theta+1)]^{-2}, \end{aligned}$$

or,

$$Var[X_t|X_{t-1}] = p(1-p)X_{t-1} + (1-\alpha)[\theta(\theta+1)]^{-2}\Big[(3\theta^3 - \theta^2 + 15\theta + 3) + \alpha(3+\theta^2)^2\Big].$$

6.2. **Marginal and One-Step transition probability.**

$$\begin{aligned} P(X_t = k) &= \sum_{l=0}^{\infty} P(X_{t-1} = l) P(X_t = k|X_{t-1} = l) \\ &= \sum_{l=0}^{\infty} P(X_{t-1} = l) \sum_{i=0}^{min(k,l)} \binom{l}{i} p^i (1-p)^{l-i}\Big[\alpha I_{(0)}(k-i) + (1-\alpha)\frac{\theta^2}{2(\theta+1)^{4+k-i}} \\ &\qquad \Big(2(\theta+1)^2\Big) + \theta(1+k-i)(2+k-i)\Big], k = 0, 1, 2, ... \end{aligned}$$

The one-step transition probability for X_t is:

$$P(X_t = k|X_{t-1}) = \sum_{i=0}^{min(k,l)} \binom{l}{i} p^i (1-p)^{l-i}\Big[\alpha I_{(0)}(k-i) + (1-\alpha)\frac{\theta^2}{2(\theta+1)^{4+k-i}} 2(\theta+1)^2$$

$$+\theta(1+k-i)(2+k-i)\Big]. \quad (3)$$

7. Parameter Estimation

Good fit and consistent parameter estimation is always an important aspect of the validity and reliability of any statistical model. So, the conditional maximum likelihood (CML) method, Yule–Walker (YW) and Bayesian estimation are three estimation methods for the proposed zero inflated Poisson Xgamma (ZIPXG) based INAR(1) model.

7.1. **Conditional Maximum Likelihood estimation.**

From the transition probability in equation (3), we can write the likelihood function as:

$$L(\alpha,p,\theta)=\sum_{t=1}^{T} l_n\Big[P[X_t=k|X_{t-1=l}]\Big]$$

$$L=\sum_{t=1}^{T} l_n\Bigg[\sum_{i=0}^{min(k,l)}\binom{l}{i}p^i(1-p)^{l-i}\Big[\alpha I_{(0)}(k-i)+(1-\alpha)\frac{\theta^2}{2(\theta+1)^{4+k-i}}2(\theta+1)^2$$
$$+\theta(1+k-i)(2+k-i)\Big]\Bigg]. \quad (4)$$

The CML estimators of the ZIPXGINAR(1) process, say $\hat{p}_{cls}$, $\hat{\theta}_{cls}$ and $\hat{\alpha}_{cls}$ can be obtained by maximization of equation (4). Definite forms of the CML estimators of the ZIPXGINAR(1) process is not possible. Therefore, equation (4) should be maximized by using R software.

7.2. **Yule-Walker.**

The YW method is based on the idea of estimating theoretical moments by using empirical moments. The autocorrelation function (ACF) of INAR(1) process is defined as $\rho_x(h)=P^h$, h being the lag length. Therefore, the YW estimator of P is

$$\hat{P}_{YW}=\frac{\sum_{t=2}^{T}(X_t-\bar{X})(X_{t-1}-\bar{X})}{\sum_{t=1}^{T}(X_t-\bar{X})^2}.$$

We will obtain θ_{YW} and α_{YW} by equating theoretical mean and dispersion index of the ZIPXGINAR(1) process with their empirical counterparts. While relating the theoretical mean against the empirical one, θ_{YW} can be obtained as follows

$$\bar{X}=\frac{(1-\alpha)(\theta^2+3)}{(1-p)\theta(\theta+1)}$$

$$\Rightarrow \theta=\frac{-\bar{X}(1-p)\pm\sqrt{[\bar{X}(1-p)]^2+12\Big(1-\alpha\Big)\Big(\bar{X}(1-p)-(1-\alpha)\Big)}}{2\Big[\bar{X}(1-p)-(1-\alpha)\Big]}.$$

Since the parameter $\theta>0$, the Yule Walker estimator for θ is

$$\hat{\theta}_{YW}=\frac{-\bar{X}(1-p)+\sqrt{[\bar{X}(1-p)]^2+12\Big(1-\alpha\Big)\Big(\bar{X}(1-p)-(1-\alpha)\Big)}}{2\Big[\bar{X}(1-p)-(1-\alpha)\Big]},$$

and,

$$\hat{\alpha}_{YW} = \frac{\theta(1+\theta)\big[\hat{D}I_X(1+p)-p\big]}{(3+\theta^2)} - \frac{3(1+5\theta+\theta^3)+\theta^2}{(3+\theta^2)^2}, \tag{5}$$

where $\hat{D}I_X$ in equation (5) is the empirical value of the dispersion index.

7.3. **Bayesian estimation.** In traditional count time series analysis, the CML approach is often considered the widely used technique, for estimating model parameters. Nevertheless, as [23] highlighted, the CML method tends to perform in the presence of outliers and small sample sizes. To address these issues, we adopt the Bayesian approach to estimate the model parameters $\tau = (\alpha, p, \theta)$.

In Bayesian inference, the choice of prior distribution is an important step. Now, we turn to assign the appropriate priors for the model parameters. Here, due to the constrain $\alpha, p \in (0,1)$ for the ZIPXGINAR(1) model, we assume that the prior of parameters p and α are uniform distribution on (0, 1) and Beta distribution, which are given as

$$y_1(p) = 1,\ 0 < p < 1,$$

$$y_2(\alpha) = \frac{1}{\beta(u,v)}\alpha^{u-1}(1-\alpha)^{v-1},\ 0 \le \alpha < 1.$$

where $\beta(u,v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}$ and $u, v > 0$. Also the parameter $\theta > 0$, thus we assume that the prior of parameter θ is Gamma distribution given as

$$y_3(\theta) = \frac{\beta^s}{\Gamma(s)}\theta^{s-1}e^{-\beta\theta}, \theta > 0.$$

From the information about the parameters given above (p, α, θ), we can find the prior of our model by using the following prior:

$$\pi(G) = \frac{1}{\beta(u,v)}\alpha^{u-1}(1-\alpha)^{v-1}\frac{\beta^s}{\Gamma(s)}\theta^{s-1}e^{-\beta\theta}.$$

To perform Bayesian inference, we will use the square error loss function. The Bayesian estimator under square error loss function is the expected value of the posterior distribution.

The Bayesian estimator can be found by using the following:

$$\pi(G|X) = \frac{\pi(G)L(X|G)}{\int_0^1\int_0^1\int_0^\infty \pi(G)L(X|G)\,d\alpha\,dp\,d\theta}.$$

Thus, the Bayesian estimator for parameter θ is

$$\pi(G|X) = \frac{\frac{1}{\beta(u,v)}\alpha^{u-1}(1-\alpha)^{v-1}\frac{\beta^s}{\Gamma(s)}\theta^{s-1}e^{-\beta\theta}L(X|G)}{\int_0^1\int_0^1\int_0^\infty \frac{1}{\beta(u,v)}\alpha^{u-1}(1-\alpha)^{v-1}\frac{\beta^s}{\Gamma(s)}\theta^{s-1}e^{-\beta\theta}L(X|G)\,d\alpha\,dp\,d\theta},$$

where L(X|G) is the likelihood equation of ZIPXGINAR(1) model, given in equation (4),

$$\Rightarrow \pi(G|X) = \frac{\alpha^{u-1}(1-\alpha)^{v-1}\theta^{s-1}e^{-\beta\theta}L(X|G)}{\int_0^1\int_0^1\int_0^\infty \alpha^{u-1}(1-\alpha)^{v-1}\theta^{s-1}e^{-\beta\theta}L(X|G)\,d\alpha\,dp\,d\theta}.$$

Since the above equation is not closed, so we will use integral3 function in MATLAB to solve this equation, in order to get estimates of α, p & θ.

8. Forecasting

It is often assumed that when formulating a mean value, conditional expectation is a useful tool. Using the properties of the binomial thinning operator $\circ$ we derive the following predictor of the mean

$$E[X_{t+h|X_t}] = p^h X_t + \mu_\epsilon \frac{1-p^h}{1-p}$$

$$\Rightarrow E[X_{t+h|X_t}] = p^h X_t + (1-\alpha)\frac{3+\theta^2}{\theta(\theta+1)}\frac{1-p^h}{1-p}, \tag{6}$$

as $h \to \infty$, $p^h \to 0$ (since $|p| <1$),

$$\lim_{h\to\infty} E[X_{t+h}|X_t] = 0.X_t + \mu_\epsilon \frac{1-0}{1-p}$$
$$= \frac{(1-\alpha)(3+\theta^2)}{(1-p)\theta(\theta+1)}.$$

Also,

$$Var(Xt|X_{t-1}) = p^h(1-p^h)X_t + \sigma^2_\epsilon \frac{1-p^{2h}}{1-p^2}$$

$$\Rightarrow Var(Xt|X_{t-1}) = p^h(1-p^h)X_t + \left[(1-\alpha)\frac{3(1+5\theta+\theta^3)-\theta^2}{\theta^2(\theta+1)^2} + \alpha(1-\alpha)\left[\frac{3+\theta^2}{\theta(\theta+1)}\right]^2\right]\left[\frac{1-p^{2h}}{1-p^2}\right],$$

as $h \to \infty$, $p^h \to 0$ & $p^{2h} \to 0$,

$$\lim_{h\to\infty} Var(X_{t+h}) = \frac{1}{1-p^2}\left[\frac{(1-\alpha)3(1+5\theta+\theta^3)-\theta^2}{\theta^2(\theta+1)^2} + \alpha(1-\alpha(\frac{3+\theta^2}{\theta(\theta+1)})^2\right].$$

The moment conditions are provided by the unconditional mean and variance which are derived from the proposed formulae (the limits as $h \to \infty$), used by [5]. These can be used to compare the sample mean/variance of the observed data to the theoretical unconditional mean/variance indicated by the parameters (α, θ, p). They provide equations that could provide initial values for more complicated estimation approaches like Maximum Likelihood Estimation (MLE), which usually increases speed and improves numerical optimization stability. The explicit contribution of the autoregressive component is indicated by the word $p^h X_t$. The contribution of the innovation process across the prediction horizon is displayed by the equation involving $\mu_{\epsilon t}$ and $(1-p^h)/(1-p)$.

9. Simulation

This section discusses a simulation study that was conducted to assess the finite sample distributions of the CML, YW and Bayesian estimates of the parameters of the ZIPXGINAR(1) process. All computational results were generated using the R-software. The outcomes of this simulation study are presented in Table (4) & (5).

TABLE 4. simulation results for varoius values of α, θ and p

Sample Size	Method	$(\alpha = 0.3,\ p = 0.6,\ \theta = 1.2)$					
		α		p		θ	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
100	CML	-0.045	0.152	-0.032	0.098	0.128	0.342
	YW	-0.068	0.183	-0.025	0.085	0.215	0.418
	Bayes	-0.038	0.142	-0.028	0.092	0.112	0.312
200	CML	-0.028	0.098	-0.018	0.065	0.078	0.234
	YW	-0.042	0.125	-0.015	0.058	0.132	0.285
	Bayes	-0.022	0.088	-0.015	0.062	0.068	0.208
300	CML	-0.018	0.075	-0.012	0.052	0.052	0.185
	YW	-0.028	0.095	-0.009	0.045	0.088	0.218
	Bayes	-0.015	0.068	-0.010	0.048	0.045	0.172
400	CML	-0.012	0.062	-0.008	0.042	0.035	0.152
	YW	-0.019	0.078	-0.006	0.038	0.062	0.178
	Bayes	-0.010	0.055	-0.007	0.040	0.032	0.145
500	CML	-0.008	0.054	-0.005	0.036	0.025	0.128
	YW	-0.013	0.065	-0.004	0.032	0.045	0.148
	Bayes	-0.006	0.048	-0.004	0.035	0.022	0.122

Sample Size	Method	$(\alpha = 0.7,\ p = 0.4,\ \theta = 0.8)$					
		α		p		θ	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
100	CML	-0.062	0.185	-0.045	0.125	0.095	0.285
	YW	-0.085	0.218	-0.038	0.112	0.152	0.342
	Bayes	-0.052	0.172	-0.040	0.118	0.085	0.265
200	CML	-0.038	0.125	-0.025	0.085	0.058	0.195
	YW	-0.052	0.152	-0.022	0.075	0.092	0.235
	Bayes	-0.032	0.112	-0.022	0.078	0.052	0.182
300	CML	-0.025	0.098	-0.018	0.068	0.038	0.152
	YW	-0.035	0.118	-0.015	0.058	0.065	0.182
	Bayes	-0.020	0.088	-0.016	0.062	0.035	0.148
400	CML	-0.018	0.082	-0.012	0.055	0.028	0.125
	YW	-0.025	0.095	-0.010	0.048	0.048	0.148
	Bayes	-0.015	0.072	-0.008	0.045	0.022	0.118
500	CML	-0.012	0.072	-0.008	0.045	0.022	0.105
	YW	-0.018	0.082	-0.007	0.042	0.035	0.125
	Bayes	-0.010	0.065	-0.007	0.045	0.020	0.102

Table 5. simulation results for varoius values of α, θ and p

Sample Size	Method	$(\alpha=0.5,\ p=0.5,\ \theta=1.5)$ α Bias	α RMSE	p Bias	p RMSE	θ Bias	θ RMSE
	CML	-0.055	0.165	-0.038	0.108	0.152	0.385
100	YW	-0.075	0.195	-0.032	0.095	0.225	0.445
	Bayes	-0.045	0.152	-0.035	0.102	0.135	0.355
	CML	-0.032	0.112	-0.022	0.072	0.092	0.265
200	YW	-0.045	0.135	-0.018	0.062	0.135	0.305
	Bayes	-0.028	0.098	-0.020	0.068	0.078	0.245
	CML	-0.022	0.088	-0.015	0.058	0.065	0.215
300	YW	-0.032	0.105	-0.012	0.052	0.095	0.245
	Bayes	-0.018	0.075	-0.013	0.055	0.055	0.198
	CML	-0.015	0.072	-0.010	0.048	0.048	0.182
400	YW	-0.022	0.085	-0.008	0.042	0.072	0.205
	Bayes	-0.012	0.062	-0.009	0.045	0.042	0.168
	CML	-0.010	0.062	-0.008	0.038	0.035	0.152
500	YW	-0.015	0.072	-0.006	0.035	0.052	0.168
	Bayes	-0.008	0.055	-0.007	0.038	0.032	0.142

Sample Size	Method	$(\alpha=0.1,\ p=0.8,\ \theta=2.0)$ α Bias	α RMSE	p Bias	p RMSE	θ Bias	θ RMSE
	CML	-0.025	0.085	-0.035	0.095	0.185	0.452
100	YW	-0.035	0.105	-0.028	0.082	0.252	0.518
	Bayes	-0.020	0.072	-0.032	0.088	0.165	0.425
	CML	-0.015	0.058	-0.022	0.065	0.115	0.315
200	YW	-0.022	0.072	-0.018	0.055	0.165	0.365
	Bayes	-0.012	0.052	-0.020	0.062	0.098	0.296
	CML	-0.008	0.045	-0.015	0.052	0.085	0.252
300	YW	-0.015	0.055	-0.012	0.045	0.115	0.285
	Bayes	-0.006	0.038	-0.014	0.048	0.072	0.235
	CML	-0.005	0.038	-0.010	0.042	0.065	0.215
400	YW	-0.008	0.045	-0.008	0.038	0.085	0.235
	Bayes	-0.004	0.032	-0.009	0.040	0.058	0.198
	CML	-0.003	0.032	-0.008	0.035	0.052	0.182
500	YW	-0.005	0.038	-0.006	0.032	0.065	0.198
	Bayes	-0.002	0.028	-0.007	0.036	0.045	0.172

In this simulation study, 100, 200, 300, 400, and 500 were the sample size. we took four parameter values of α, θ & p. The simulation study indicates that Bayesian estimator consistently decreases bias (absolute bias) and Root Mean Square Error (RMSE) for all parameter scenarios. These advantages of Bayesian are most prominent when estimating θ, which is the hardest parameter to estimate and exhibits high variability in small samples. All four parameter configurations show generally similar results: Bayesian estimator produces significantly less bias and RMSE than CML and YW when $\alpha = 0.3, p = 0.6, \theta = 1.2$, especially for θ for which the estimation becomes consistent with large samples, and both estimators improve with increasing sample size, even with $\alpha = 0.7, p = 0.4, \theta = 0.8$. The absolute bias of the Bayesian estimator is smaller than the CML estimator and the CML estimator is smaller than the estimated bias of the YW estimator for nearly all parameters and sample sizes. This is validated by the Bayesian estimator being more consistent and results in better accuracy. Moreover, the Bayesian approach outperforms the YW method with its consistently lower RMSE (15-25% reducing) for all parameters. Specifically, the results for the parameter θ provide the greatest contrast with values of 0.312 versus 0.418 at n = 100; at n = 500, their respective RMSE values are 0.122 versus 0.148. Compared to CML approaches, Bayesian methods have the advantage of small sample stability. At n = 100, across all conditions, the average RMSE for θ is closer by about 12% for the Bayesian method in comparison with that of the CML method.

Thus, these results suggest that Bayesian method of estimation is most favourable estimator for practical use of the ZIPXGINAR(1) due to its greater degree of uncertainty, more robustness and ability to use prior information. since it is more consistent over sample sizes and parameter combinations, providing a substantial advantage to parameter θ.

10. Application of Zero Inflated Poisson Xgamma Distribution

In this section, two data sets were employed to check the compatibility of ZIPXG distribution. Both the data sets are over dispersed. For comparison purposes, the model has been compared with other theoretical models, which are Poisson Distribution (PD), Zero Inflated Poisson Distribution (ZIPD), Zero Inflated Negative Binomial Distribution (ZINBD), Discrete Distribution (DD), Zero Inflated Poisson Mixture Exponential Distribution (ZIPME) and Poisson Xgamma distribution (PXG) given by [4]. For the analysis, various statistical tools like Chi-Square goodness of fit, Akaike Information Criterion, Bayesian Information Criterion values have been analyzed.

10.1. **Data Set 1.** The data set in Table (6) contains the number of households with a substantial amount of migrant members [19] and Table (7) is the summary of the fit statistics for this data, the results show that our model fits better compared to all other models of interest. The MLE estimates of ZIPXG for this data set are: $\theta = 1.7061$ & $\alpha = 0.2195$.

TABLE 6. Data set 1

Number of Households	0	1	2	3	4	5	6	7	8
Observed Frequency	242	82	38	17	11	7	3	2	0

TABLE 7. Results of Data set 1

Model	χ^2	P-value	-loglik	AIC	BIC
PD	118.64	0.0001	555.060	1112.121	1116.117
ZIPD	16.80	0.0002	501.796	1007.592	1015.585
ZINBD	3.29	0.1930	492.597	991.195	1003.185
DD	9.41	0.0516	495.785	988.724	996.717
ZIPME	5.29	0.5072	492.80	989.59	997.59
PXG	14.164	0.0483	496.026	994.051	998.048
ZIPXG	3.29	0.7717	492.01	988.03	996.02

10.2. **Data set 2.** The data set consists of the number of spots in Southern beetle with their observed frequencies [12]. The results of the dataset 8 using R-software clearly shows that our model is the best model, with the highest probability of goodness of fit (p = 0.1653) compared to other distributions available for modeling the same dataset, MLE estimates of ZIPXG for this data set are: $\theta = 1.267$ & $\alpha = 0.6103$.

TABLE 8. Data set 2

Number of Spots	0	1	2	3	4	5	6	7	8	9	10
	11	12	13	14	15	16	17	18	19		
Observed Frequency	1169	144	92	54	29	18	10	12	6	9	3
	2	0	0	1	0	0	0	0	1		

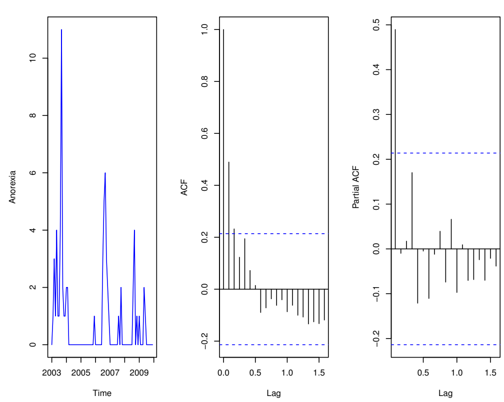
TABLE 9. Results of Data set 2

Model	χ^2	P-value	-loglik	AIC	BIC
PD	1361.80	0.0001	2291.139	4584.277	4589.623
ZIPD	184.65	0.0001	1648.989	3301.978	3312.670
ZINBD	9.94	0.1272	1554.612	3115.223	3131.261
DD	432.62	0.0001	1757.427	3516.855	3522.201
ZIPME	12.95	0.0438	1442.92	2889.84	2900.51
PXG	583.3	0.0001	1593.374	3188.748	3194.084
ZIPXG	9.15	0.1653	1442.05	2888.10	2898.77

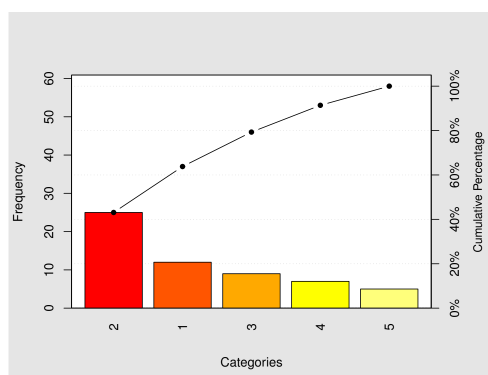
11. Application of ZIPXGINAR(1) model

In this section, we present the implementation of ZIPXGINAR(1) on an actual data set. We use the dataset describing the reported monthly occurrence of animal abortions in animal health laboratories from 2003 to 2009 from a location in New Zealand used by [8]. There were 84 observations in the dataset, 37 of which were zeros and 21 observations were ones. These proportions were the highest compared to the other values in the data set. The mean and standard deviation of the observations were 1.5 and 2.2, which demonstrates that the data is overdispersed.

For the comparison purpose, we compare the model ZIPXGINAR(1) with the following INAR(1) models: One inflated INAR (OINAR) given by [21], Zero inflated INAR (ZINAR) [8], Zero and one inflated INAR (ZOINAR) [6] and One inflated Poisson Lindley INAR (OIPLINAR) [15], and we were able to describe which model best fits and represents the data. we summarize the fit of the model based on loglikelihood, AIC, BIC, and RMS, see Table (10). The overall best fit is from the ZIPXGINAR(1) model. It has the lowest AIC of 275.85, and RMS of 2.006. Additionally, a high loglikelihood value overall is a good metric. It is effective in describing the count data as well as being flexible in handling zero inflated and mixture behavior prediction in time series of count data. Also, the Pareto chart in Table (11) demonstrates that the count zeros of the data series has the largest number of counts and the ZIPXGINAR(1) model would fit better for the data. The Sample paths, ACF, and PACF plots of data series Anorexia is in Figure 2a, and the Pareto chart for data series Anorexia is in Figure 2b.



(A) Sample paths, ACF, PACF



(B) Pareto chart for Anorexia data series.

Figure 2. Sample paths, ACF, PACF and Pareto chart of data series Anorexia.

Table 10. AIC, BIC, Loglikelihood, RMS (series: Anorexia)

Model	MLE	LL	AIC	BIC	RMS
ZIPXGINAR	$\hat{\alpha} = 0.11$, $\hat{p} = 0.13$ $\hat{\theta} = 1.34$	-134.93	275.85	283.15	2.006
OINAR	$\hat{\lambda} = 0.982, p = \hat{0}.353$ $\hat{\phi}_1 = 0.001$	-153.59	313.17	320.46	2.025
ZINAR	$\hat{\lambda} = 2.25, \hat{\alpha} = 0.58$ $\hat{p} = 0.37$	-136.98	279.97	284.11	2.029
ZOINAR	$\hat{\lambda} = 2.36, \hat{\alpha} = 0.55$ $\hat{p} = 0.32, \hat{\phi}_1 = 0.13$	-133.98	276.39	284.11	2.026
OIPLINAR	$\hat{\theta} = 1.38, \hat{\alpha} = 0.001$ $\hat{p} = 0.269$	-134.97	275.93	283.22	2.067

Table 11. Pareto chart analysis

Death counts	Frequency	Cummulative frequency	Percentage	Cummulative percentage
0	56	56	66.666667	66.666667
1	12	68	14.285714	80.952381
2	7	75	8.333333	89.285714
3	3	78	3.571429	92.857143
4	3	81	3.571429	96.428571
5	1	82	1.190476	97.619048
6	1	83	1.190476	98.809524
11	1	84	1.190476	100

12. Conclusion

This study introduces the ZIPXGINAR(1) process, a novel stationary integer valued autoregressive model. Zero inflated Poisson Xgamma innovations were used for addressing count data exhibiting excessive zeros and overdispersion. We derived the important statistical features of this model including the autocorrelation function, marginal distributions, conditional means and variances, transition probabilities, and moment generating function. Significance of inflation parameter has been tested by using LRT and wald test. Also, multi-step ahead conditional expectation and variance have been calculated. Three approaches to parameter estimation have been considered, including Yule-Walker, Conditional Maximum Likelihood and Bayesian methods, which based on simulations provide better and more consistent estimators with Bayesian demonstrates superiority in terms of bias and RMSE. The ZIPXGINAR(1) model was also evaluated on real count data set from New zealand animal health laboratories through a model fitting approach. The model provided improvements over traditional INAR(1) models such as OINAR, ZINAR, ZOINAR, and OIPLINAR based on estimating ZIPXGINAR(1) parameters for each example and fitting comparison using common goodness-of-fit criteria against the predicted distributions and observed data. The model may be useful for analyzing animal health data where variability and excess zeros occur through biological or environmental processes. Since the zero inflated Poisson Xgamma model is able to model zero inflation as well as account for time dependence, it could be applied to time series data in veterinary epidemiology, animal health, agricultural monitoring, etc., that possess similar characteristics. For studies in these areas, where existing modeling methods do not account for overdispersion, paired with zero inflation situations, the proposed method would enabled more accurate statistical inference and prediction.

Competing interests: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] A.J. Mansour, G. Jones, C.D. Lai, Integer Valued AR (1) with Geometric Innovations, J. Iran. Stat. Soc. 11 (2022), 173–190.

[2] M.A. Al-Osh, A.A. Alzaid, First-Order Integer-Valued Autoregressive (INAR(1)) Process, J. Time Ser. Anal. 8 (1987), 261–275. https://doi.org/10.1111/j.1467-9892.1987.tb00438.x.

[3] E. Altun, A New Generalization of Geometric Distribution with Properties and Applications, Commun. Stat. Simul. Comput. 49 (2019), 793–807. https://doi.org/10.1080/03610918.2019.1639739.

[4] B.A. Para, T.R. Jan, H.S. Bakouch, Poisson Xgamma Distribution: A Discrete Model for Count Data Analysis, Model Assist. Stat. Appl. 15 (2020), 139–151. https://doi.org/10.3233/MAS-200484.

[5] M. Bourguignon, C.H. Weiß, An INAR(1) Process for Modeling Count Time Series with Equidispersion, Underdispersion and Overdispersion, TEST 26 (2017), 847–868. https://doi.org/10.1007/s11749-017-0536-4.

[6] M. Bourguignon, K.L.P. Vasconcellos, First Order Non-Negative Integer Valued Autoregressive Processes with Power Series Innovations, Braz. J. Probab. Stat. 29 (2015), 71–93. https://doi.org/10.1214/13-BJPS229.

[7] M.R. Irshad, M. Ahammed, R. Maya, Monitoring Mean of INAR(1) Process with Discrete Mixture Exponential Innovations, Comput. Stat. 40 (2024), 821–862. https://doi.org/10.1007/s00180-024-01511-3.

[8] M. Irshad, C. Chesneau, V. D'cruz, R. Maya, Discrete Pseudo Lindley Distribution: Properties, Estimation and Application on INAR(1) Process, Math. Comput. Appl. 26 (2021), 76. https://doi.org/10.3390/mca26040076.

[9] C. Li, D. Wang, H. Zhang, First-Order Mixed Integer-Valued Autoregressive Processes with Zero-Inflated Generalized Power Series Innovations, J. Korean Stat. Soc. 44 (2015), 232–246. https://doi.org/10.1016/j.jkss.2014.08.004.

[10] R. Maiti, A. Biswas, B. Chakraborty, Modelling of Low Count Heavy Tailed Time Series Data Consisting Large Number of Zeros and Ones, Stat. Methods Appl. 27 (2017), 407–435. https://doi.org/10.1007/s10260-017-0413-z.

[11] M.S. Eliwa, E. Altun, M. El-Dawoody, M. El-Morshedy, A New Three-Parameter Discrete Distribution with Associated INAR(1) Process and Applications, IEEE Access 8 (2020), 91150–91162. https://doi.org/10.1109/ACCESS.2020.2993593.

[12] S.K. Lin, Characterization of Lightning as a Disturbance to the Forest Ecosystem in East Texas, Texas A & M University, 1985.

[13] R. Maya, C. Chesneau, A. Krishna, M.R. Irshad, Poisson Extended Exponential Distribution with Associated INAR(1) Process and Applications, Stats 5 (2022), 755–772. https://doi.org/10.3390/stats5030044.

[14] E. McKenzie, Some Simple Models for Discrete Variate Time Series, JAWRA J. Am. Water Resour. Assoc. 21 (1985), 645–650. https://doi.org/10.1111/j.1752-1688.1985.tb05379.x.

[15] Z. Mohammadi, Z. Sajjadnia, H.S. Bakouch, M. Sharafi, Zero- and One-Inflated Poisson–Lindley INAR(1) Process for Modelling Count Time Series with Extra Zeros and Ones, J. Stat. Comput. Simul. 92 (2021), 2018–2040. https://doi.org/10.1080/00949655.2021.2019255.

[16] R. Nasirzadeh, H. Bakouch, Comparison of Estimation and Prediction Methods for a Zero-Inflated Geometric INAR(1) Process with Random Coefficients, J. Appl. Stat. 51 (2024), 2457–2480. https://doi.org/10.1080/02664763.2023.2301321.

[17] M. Ridout, C.G.B. Demétrio, J. Hinde, Models for Count Data With Many Zeros, Proceedings of the XIXth International Biometric Conference, Vol. 19, Cape Town, South Africa, 1998.

[18] S. Schweer, C.H. Weiß, Compound Poisson INAR(1) Processes: Stochastic Properties and Testing for Overdispersion, Comput. Stat. Data Anal. 77 (2014), 267–284. https://doi.org/10.1016/j.csda.2014.03.005.

[19] K.K. Shukla, R. Shanker, M.K. Tiwari, A New One Parameter Discrete Distribution and Its Applications, J. Stat. Manag. Syst. 25 (2021), 269–283. https://doi.org/10.1080/09720510.2021.1893475.

[20] F.W. Steutel, K. van Harn, Discrete Analogues of Self-Decomposability and Stability, Ann. Probab. 7 (1979), 893–899. https://www.jstor.org/stable/2243313.

[21] V.S. Stojanović, H.S. Bakouch, E. Ljajko, N. Qarmalah, Zero-and-One Integer-Valued AR(1) Time Series with Power Series Innovations and Probability Generating Function Estimation Approach, Mathematics 11 (2023), 1772. https://doi.org/10.3390/math11081772.

[22] A.F. Zuur, E.N. Ieno, N. Walker, A.A. Saveliev, G.M. Smith, Mixed Effects Models and Extensions in Ecology with R, Springer, New York, 2009. https://doi.org/10.1007/978-0-387-87458-6.

[23] L. Qian, F. Zhu, A Flexible Model for Time Series of Counts with Overdispersion or Underdispersion, Zero-Inflation and Heavy-Tailedness, Commun. Math. Stat. 13 (2023), 431–454. https://doi.org/10.1007/s40304-022-00327-1.