Journal of Statistical Modelling: Theory and Applications
Vol. 4, No. 1, 2023, pp. 45-58
Yazd University Press 2023

JSMTA

Research Paper

Modeling zero-inflated and zero-deflated count data time series using the INMA(1) process

Eisa Mahmoudi*, Ameneh Rostami
[1]Department of Statistics, Yazd University, Yazd, Iran

Received: May 28, 2023/ Revised: September 09, 2023/ Accepted: December 23, 2023

Abstract: In the real world, we may come across with zero-inflated or zero-deflated count data that have a very short-run autocorrelation. Integer-valued moving average processes are suitable for modeling these data. In this paper, a non-negative integer-valued moving average process of the first order with zero-modified geometric innovations is introduced. This model is called zero-modified geometric INMA(1) process which contains geometric INMA(1) process as a particular case. Some statistical properties of the process are obtained. The parameters of the model are estimated by the Yule-Walker method. Then, using the simulation study, we evaluate the performance of this estimators. Finally, the model is applied to two examples of real time series of the monthly number of rubella cases and the annually number of earthquakes magnitude 8.0 to 9.9. Then, we exhibit the ability of the model for fitting and predicting count data with excess and deficit of zeros.

Keywords: INMA(1) process; Zero-deflated; Zero-inflated; Zero-modified geometric distribution.
Mathematics Subject Classification (2010): 62M10.

1 Introduction

In recent decades, the integer-valued time series is used in different fields, such as economic, insurance, medicine, ecology and biology. Sometimes, in real life, there are situations where the zeros appear in the count data with a greater or a lesser tendency, such as the daily number of hospitalized patients in a hospital and the annually number of earthquakes in the world. Count data with this feature is also known as zero-inflation or zero-deflation.

*Corresponding author: `emahmoudi@yazd.ac.ir`

One of the methods for analyzing and modeling count data with inflation or deflation of zeros are zero modified count time series models. Some researchers introduced zero modified INAR models, such as Jazi et al. (2012), Barreto-Souza (2015) and Li et al. (2015) studied INAR(1) process with zero-inflated Poisson innovations, INAR(1) model with zero-modified geometric marginal distribution based on negative binomial thinning operator and INAR(1) process with zero-inflated generalized power series innovations, respectively. Recently, Bakouch et al. (2018) proposed a zero-inflated geometric INAR(1) process with random coefficient and Bourguinon (2018) introduced an INAR(1) process with zero-modified geometric innovations.

For modeling this type of count data, the zero-modified INAR models often are not the best choice when data have a very short-run autocorrelation. In this case, we need to introduce a zero-modified model based on the INMA process that was introduced by McKenzie (1988) and Al-Osh and Alzaid (1988).

The aim of this paper is to introduce a new INMA(1) process with zero-modified geometric innovations based on the binomial thinning operator. Advantages of the new process are these: (i) the model is suitable for modeling count data with excess or deficit zeros that has a very short-run autocorrelation. (ii) in the model, innovations come from the zero-modified geometric distribution that is very flexible. (iii) the model is used to fit the count data which has overdispersion, underdispersion and equidispersion.

The paper is organized as follows: The INMA(1) process with zero-modified geometric innovations is defined in Section 2 and some statistical properties of this model are derived. In Section 3, the Yule-Walker estimators of the parameters is obtained. Some simulation results for the estimators are given in Section 4. Section 5 presents two real applications of the proposed model. Finally, we conclude the paper in Section 6.

2 Construction of the ZMGINMA(1) process

In this section, we introduce an INMA(1) model with zero-modified geometric innovations generated by the binomial thinning operator. The binomial thinning operator $\alpha o X$ (Steutel and Van Harn, 1979) for a random variable X is generated by a sequence of independent and identically distributed (i.i.d.) Bernoulli random variables $\{Z_i\}_{i\geq 1}$ with mean α. For a given random variable X, the random variable $\alpha o X$ has the binomial distribution with mean αX.

As we mention above, we consider a model which innovations have the zero-modified geometric distribution. Firstly, we review the zero-modified geometric distribution. A non-negative integer-valued random variable X is said to follow a zero-modified geometric distribution (ZMG) with parameters $\mu > 0$ and $\pi \in (-1/\mu, 1)$, if its probability mass function is given by

$$P(X = x) = \begin{cases} \pi + \frac{1-\pi}{1+\mu}, & \text{if } x = 0, \\ \frac{(1-\pi)\mu^x}{(1+\mu)^{x+1}}, & \text{if } x = 1, 2, \ldots. \end{cases} \quad (1)$$

This distribution when $\pi \in (-1/\mu, 0)$ and $\pi \in (0, 1)$ have a proportion of less and more zeros than the geometric distribution, respectively. Thus, this distribution becomes a usual geometric distribution if $\pi = 0$; a zero-deflated geometric distribution when $\pi \in (-1/\mu, 0)$; a zero-inflated distribution when $\pi \in (0,1)$.

The probability generating function (pgf) of X is

$$\varphi(s) = \frac{1+\pi\mu(1-s)}{1+\mu(1-s)}, \quad |s| < 1.$$

The expectation, variance, skewness and kurtosis of X are given, respectively, by

$$\begin{aligned}
E(X) &= \mu(1-\pi), \\
Var(X) &= \mu(1-\pi)[1+\mu(1+\pi)], \\
Sk &= \frac{(6\mu^2+4\mu+1)-3\mu(1-\pi)(1+2\mu\pi)}{\mu^{\frac{1}{2}}(1-\pi)^{\frac{1}{2}}[1+\mu(1+\pi)]^{\frac{3}{2}}} + \frac{2\mu^2(1-\pi)^2}{\mu^{\frac{1}{2}}(1-\pi)^{\frac{1}{2}}[1+\mu(1+\pi)]^{\frac{3}{2}}}, \\
Kur &= \frac{(24\mu^3+36\mu^2+12\mu+5)}{\mu(1-\pi)[1+\mu(1+\pi)]^2} - \frac{4\mu(1-\pi)(6\mu^2+4\mu+1)}{\mu(1-\pi)[1+\mu(1+\pi)]^2} \\
&\quad + \frac{6\mu^2(1-\pi)^2(1+2\mu\pi)-3\mu^3(1-\pi)^3}{\mu(1-\pi)[1+\mu(1+\pi)]^2},
\end{aligned}$$

The index of dispersion is given by $ID(x) = 1 + \mu(1-\pi)$. Thus,
(i) If $\pi = -1$, the ZMG distribution is equidispersed.
(ii) For $\mu \in (0,1)$ and $\pi \in (-1/\mu, -1)$, the ZMG distribution presents underdispersed.
(iii) For $\pi \in (-1,1)$, the ZMG distribution is overdispersed.

Definition 2.1. *A time series model $\{Y_t\}$, which is given by*

$$Y_t = \alpha o \varepsilon_{t-1} + \varepsilon_t, \ t \in \{0, \pm 1, \pm 2, \dots\}, \tag{2}$$

is called the integer-valued moving average model with zero-modified geometric innovations (ZMGINMA(1)) if the following conditions are satisfied: (i) $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with a ZMG distribution given by (1), *(ii) the counting series Z_i, incorporated in $\alpha o \varepsilon_t$, is a sequence of i.i.d. Bernoulli random variables with parameter $\alpha \in (0,1)$ for all i, (iii) all the counting series incorporated in $\alpha \ominus \varepsilon_s$ and $\alpha \ominus \varepsilon_t$ are independent for all $s \neq t$, (iv) the counting series Z_i is independent of ε_t for all t and i.*

Under the above assumptions, the expectation and variance of Y_t are given, respectively, by

$$\begin{aligned}
E(Y_t) &= (1+\alpha)\mu(1-\pi), \\
Var(Y_t) &= \mu(1-\pi)[(1+\alpha^2)\mu(1+\pi)+1+\alpha].
\end{aligned}$$

Also, the index of dispersion is given by

$$ID(x) = \frac{(1+\alpha^2)\mu(1+\pi)+1+\alpha}{1+\alpha}.$$

This model is equidispersed or overdispersed if $\pi = -1$ or $\pi \in (-1,1)$, respectively, and is underdispersed for $\mu \in (0,1)$ and $\pi \in (-1/\mu, -1)$.

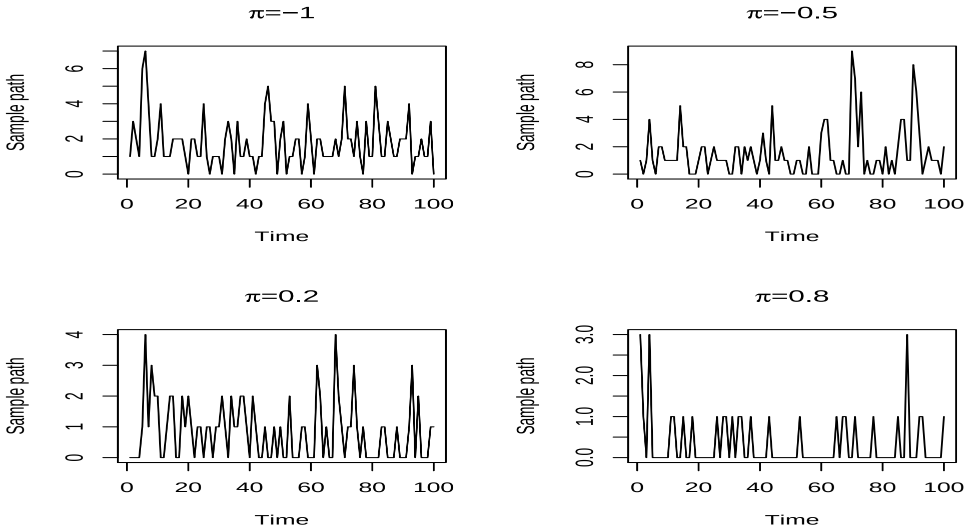

Figure 1: Sample paths of ZMGINMA(1) process for $\mu = 0.5$, $\alpha = 0.3$, and different values of π.

Figure 1 shows the sample paths of simulated INMA(1) process with ZMG(π,μ) innovations for $\mu = 0.7$, $\alpha = 0.2$, and different values of π. In Figure 1, we observe that the number of zeros increase as π increases. The autocovariance function of the proposed model is given by

$$\gamma(k) = Cov(Y_t, Y_{t-k}) = \begin{cases} \alpha[\mu(1-\pi)[1+\mu(1+\pi)]], & k = 1, \\ 0, & k > 1. \end{cases}$$

Thus, the autocorrelation function is

$$\rho(k) = \begin{cases} \frac{\alpha[\mu(1-\pi)[1+\mu(1+\pi)]]}{\mu(1-\pi)[(1+\alpha^2)\mu(1+\pi)+1+\alpha]}, & k = 1, \\ 0, & k > 1. \end{cases}$$

We can see that $\rho(1)$ is non-negative and bounded above by 1/2.

Theorem 2.2. *The ZMGINMA(1) process has the following properties:*
(i) A covariance stationary process.
(ii) Is ergodic in the mean and autocovariance function.

Proof. (i) Since the expectation and variance of the model are constant and autocovariance function does not depend on time, the model (2) is covariance stationary.
(ii) the proof is similar to the proof of Theorem 7 from Yu and Zou (2015) hence the details are avoided. □

Using the pgf of the ZMG distribution and the independency of the counting series and the random variables ε_{t-1} and ε_t, we obtain that the pgf of the random variable

Y_t is given by

$$\varphi_{Y_t}(s) = \varphi_{\varepsilon_{t-1}}(1-\alpha+\alpha s)\varphi_{\varepsilon_t}(s) = \frac{1+\pi\mu\alpha(1-s)}{1+\mu\alpha(1-s)} \times \frac{1+\pi\mu(1-s)}{1+\mu(1-s)},$$

which implies that the random variable Y_t is distributed as the random variable $X+W$, where X and W are independent random variables and have the ZMG distributions with parameters (π,μ) and $(\pi,\alpha\mu)$, respectively. Thus the probability mass function of the random variable Y_t can be calculated by

$$P(Y_t = k) = \begin{cases} \frac{(\pi\mu+1)(\pi\mu\alpha+1)}{(1+\mu)(1+\mu\alpha)}, & \text{if } k=0, \\ \frac{(1-\pi)\mu^k[\alpha^k(1+\mu\pi)(1+\mu)^k+(1+\pi\mu\alpha)(1+\alpha\mu)^k]}{(1+\mu)^{k+1}(1+\alpha\mu)^{k+1}} & \\ +\frac{(1-\pi)^2\mu^k}{(1+\mu)^k(1+\mu\alpha)^k}\sum_{i=1}^{k-1}\alpha^i(1+\mu)^{i-1}(1+\mu\alpha)^{k-1-i}, & \text{if } k=1,2,\dots. \end{cases}$$

In this part, we will consider the regression of the proposed model and will show this model like the GINMA(1) and PINMAPS(1) (Mahmoudi and Rostami, 2020) models has a non-linear regression. So, we first derive the joint pgf of the random variables Y_{t-1} and Y_t. It is given by

$$\varphi_{Y_{t-1},Y_t}(s_1,s_2) = \frac{1+\pi\mu(1-s_2)}{1+\mu(1-s_2)} \times \frac{1+\pi\mu\alpha(1-s_1)}{1+\mu\alpha(1-s_1)} \times \frac{1+\pi\mu(1-s_1+\alpha s_1-\alpha s_1 s_2)}{1+\mu(1-s_1+\alpha s_1-\alpha s_1 s_2)}.$$

The joint pgf can be used for calculating of the conditional pgf of $Y_t|Y_{t-1}=x$; which is given in the following theorem.

Theorem 2.3. *The conditional pgf of $Y_t|Y_{t-1}=x$, $x\in\{0,1,2,\dots\}$, is given by*

$$\varphi_{Y_t|Y_{t-1}=x}(s) = \frac{\frac{1+\pi\mu(1-s)}{1+\mu(1-s)}\sum_{j=0}^{x}\binom{x}{j}\frac{j!(x-j)!\mu^x\alpha^j(1-\alpha+\alpha s)^{x-j}}{(1+\mu\alpha)^{j+1}(1+\mu)^{x-j+1}}}{\sum_{j=0}^{x}\binom{x}{j}\frac{j!(x-j)!\mu^x\alpha^j}{(1+\mu\alpha)^{j+1}(1+\mu)^{x-j+1}}}.$$

Proof. According to Theorem 1.3.1 from Kocherlakota and Kocherlakota (1992) the conditional pgf of the random variable Y_t for given $Y_{t-1}=x$ can be derived from the joint pgf $\varphi_{Y_{t-1},Y_t}(s_1,s_2)$ as

$$\varphi_{Y_t|Y_{t-1}=x}(s) = \frac{\frac{\partial^x\varphi_{Y_{t-1},Y_t}(0,s)}{\partial s_1^x}}{\frac{\partial^x\varphi_{Y_{t-1},Y_t}(0,1)}{\partial s_1^x}}. \tag{3}$$

By the Leibniz's rule, we have

$$\begin{aligned}\frac{\partial^x\varphi_{Y_{t-1},Y_t}(s_1,s_2)}{\partial s_1^x} &= \frac{1+\pi\mu(1-s_2)}{1+\mu(1-s_2)}\sum_{j=0}^{x}\binom{x}{j}\frac{\partial^j(\frac{1+\pi\mu\alpha(1-s_1)}{1+\mu\alpha(1-s_1)})}{\partial s_1^j} \\ &\quad\times\frac{\partial^{x-j}(\frac{1+\pi\mu(1-s_1+\alpha s_1-\alpha s_1 s_2)}{1+\mu(1-s_1+\alpha s_1-\alpha s_1 s_2)})}{\partial s_1^{x-j}}.\end{aligned} \tag{4}$$

It is easy to derive the kth partial derivatives of $\frac{1+\pi\mu(1-s_1+\alpha s_1-\alpha s_1 s_2)}{1+\mu(1-s_1+\alpha s_1-\alpha s_1 s_2)}$ as

$$\frac{\partial^k(\frac{1+\pi\mu(1-s_1+\alpha s_1-\alpha s_1 s_2)}{1+\mu(1-s_1+\alpha s_1-\alpha s_1 s_2)})}{\partial s_1^k} = \frac{k!\mu^k(1-\pi)(1-\alpha+\alpha s_2)^k}{(1+\mu(1-s_1+\alpha s_1-\alpha s_1 s_2))^{k+1}}, \quad k=0,1,\dots,x. \tag{5}$$

On the other hand, after some calculations, we obtain the kth partial derivatives of $\frac{1+\pi\mu\alpha(1-s_1)}{1+\mu\alpha(1-s_1)}$ as

$$\frac{\partial^k\left(\frac{1+\pi\mu\alpha(1-s_1)}{1+\mu\alpha(1-s_1)}\right)}{\partial s_1^k} = \frac{k!\mu^k\alpha^k(1-\pi)}{(1+\mu\alpha(1-s_1))^{k+1}}, \quad k = 0, 1, \ldots, x. \tag{6}$$

Finally, replacing (5) and (6) in (4) with $s_1 = 0$ and $s_2 = s$, we obtain the numerator of (3). In a similar way, one can obtain the denominator of (3) which proves the theorem. □

Now, we derive the regression of the proposed model. It is given in the following corollary.

Corollary 2.4. *The regression of Y_t given $Y_{t-1} = x$ is a non-linear function given by*

$$E(Y_t|Y_{t-1} = x) = \frac{\sum_{j=0}^{x}\binom{x}{j}\frac{j!(x-j)!\mu^x\alpha^j}{(1+\mu\alpha)^{j+1}(1+\mu)^{x-j+1}}(\mu(1-\pi)+(x-j)\alpha)}{\sum_{j=0}^{x}\binom{x}{j}\frac{j!(x-j)!\mu^x\alpha^j}{(1+\mu\alpha)^{j+1}(1+\mu)^{x-j+1}}}. \tag{7}$$

Proof. The proof follows from Corollary 1.3.1 given by Kocherlakota and Kocherlakota (1992) and Theorem 2.3. □

The conditional mean and variance of Y_t given $\mathcal{F}_{t-1}$ are obtained as follows

$$\begin{aligned} E(Y_t|\mathcal{F}_{t-1}) &= \alpha\varepsilon_{t-1} + \mu_\varepsilon, \\ Var(Y_t|\mathcal{F}_{t-1}) &= \alpha(1-\alpha)\varepsilon_{t-1} + \sigma_\varepsilon^2, \end{aligned}$$

where $\mathcal{F}_{t-1}$ is the information set at time $t-1$.

Note that in this paper, the ZMGINMA(1) model is applied for analyzing count time series with excess or deficit of zeros, therefore, we derive the distribution of zero values in the series.

Lemma 2.5. *The transition probability from zero to zero and zero to non-zero of the ZMGINMA(1) model are given, respectively, by*

$$\begin{aligned} P(Y_t = 0 \mid Y_{t-1} = 0) &= \frac{\pi\mu+1}{1+\mu}, \\ P(Y_t \neq 0 \mid Y_{t-1} = 0) &= \frac{(1-\pi)\mu}{1+\mu}. \end{aligned}$$

Proof. According to Theorem 2.3, the conditional pgf of $Y_t|Y_{t-1} = 0$ is given by

$$\varphi_{Y_t|Y_{t-1}=0}(s) = \frac{1+\pi\mu(1-s)}{1+\mu(1-s)},$$

thus, the random variable Y_t given $Y_{t-1} = 0$ has the ZMG distribution with parameters μ and π. Therefore,

$$\begin{aligned} P(Y_t = 0 \mid Y_{t-1} = 0) &= \frac{\pi\mu+1}{1+\mu}, \\ P(Y_t \neq 0 \mid Y_{t-1} = 0) &= 1 - P(Y_t = 0 \mid Y_{t-1} = 0) = \frac{(1-\pi)\mu}{1+\mu}. \end{aligned}$$

□

In the ZMGINMA(1) model like the INAR(1) model with ZMG innovations (Bourguinon (2018)), the run length of zeros, S, has a geometric distribution with termination probability $\frac{(1-\pi)\mu}{1+\mu}$. Thus, the average of S in the proposed model is given by $E(S) = \frac{1+\mu}{(1-\pi)\mu}$. On the other hand, the average run length of zeros in the GINMA(1) model is $E(S_0) = \frac{1+\mu}{\mu}$, therefore,

$$\begin{aligned} E(S) &\geq E(S_0), \qquad \text{for} \quad \pi \in [0, 1), \\ E(S) &\leq E(S_0), \qquad \text{for} \quad \pi \in (-1/\mu, 0). \end{aligned}$$

Remark 2.6. *The proportion of zeros in the ZMGINMA(1) model is given by*

$$P(Y_t = 0) = \frac{(\pi\mu + 1)(\pi\mu\alpha + 1)}{(1 + \mu)(1 + \mu\alpha)}.$$

3 Estimation of the unknown parameters

In this section, the Yule-Walker (YW) estimatiors of the model parameters are obtained. For this purpose, we consider a random sample of size T from ZMGINMA(1) model. The YW estimators of α, π and μ are obtained by solving the following equations:

$$\begin{aligned} &\bar{Y} - (1 + \alpha)\mu(1 - \pi) = 0, \\ &\hat{\gamma}(0) - \mu(1 - \pi)[((1 + \alpha^2)(1 + \mu(1 + \pi)) + \alpha - \alpha^2] = 0, \\ &\hat{\gamma}(1) - \alpha\mu(1 - \pi)[1 + \mu(1 + \pi)] = 0, \end{aligned}$$

where $\bar{Y}$, $\hat{\gamma}(0)$ and $\hat{\gamma}(1)$ are the sample mean, variance and autocovariance function at lag 1, respectively. Thus, estimator $\hat{\alpha}_{YW}$ is obtained by solving the cubic equation

$$(\hat{\gamma}(1) - \bar{Y})\alpha^3 + (\hat{\gamma}(1) + \bar{Y} - \hat{\gamma}(0))\alpha^2 + (\hat{\gamma}(1) - \hat{\gamma}(0))\alpha + \hat{\gamma}(1) = 0.$$

After some mathematical computations, the YW estimators of μ and π are given by

$$\begin{aligned} \hat{\mu}_{YW} &= \frac{\bar{Y}}{2(1 + \hat{\alpha}_{YW})} + \frac{(1 + \hat{\alpha}_{YW})\hat{\gamma}(1) - \hat{\alpha}_{YW}\bar{Y}}{2\hat{\alpha}_{YW}\bar{Y}}, \\ \hat{\pi}_{YW} &= 1 - \frac{\bar{Y}}{\hat{\mu}_{YW}(1 + \hat{\alpha}_{YW})}. \end{aligned}$$

Theorem 3.1. *The YW estimators of the parameters α, μ and π are consistent.*

Proof. According to Theorem 2.2, we have $\bar{Y} \xrightarrow{p} \mu_Y$, $\hat{\gamma}(0) \xrightarrow{p} \gamma(0)$ and $\hat{\gamma}(1) \xrightarrow{p} \gamma(1)$ where $\mu_Y := E(Y_t)$ and $\gamma(0) := Var(Y_t)$. Thus, using the properties of convergence in probability, the consistency of the YW estimators is achieved. □

4 Simulation study

In this section, we simulated 10000 samples of size $T = 500, 800, 1500$ from the ZMGIN MA(1) model for $\alpha = 0.3$ and different values of μ and π. We derived the YW estimates,

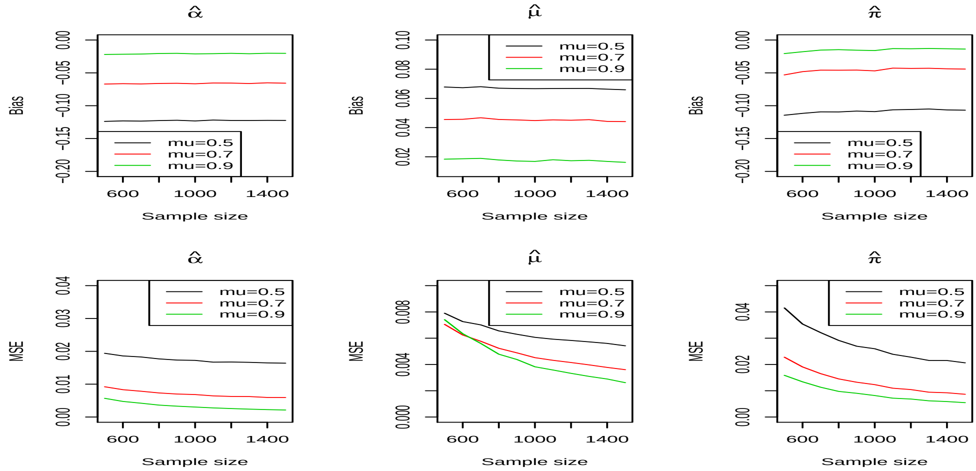

Figure 2: Bias and MSE of the simulated YW estimates of α, μ and π for $\alpha = 0.3$, $\pi = -1$, different values of μ.

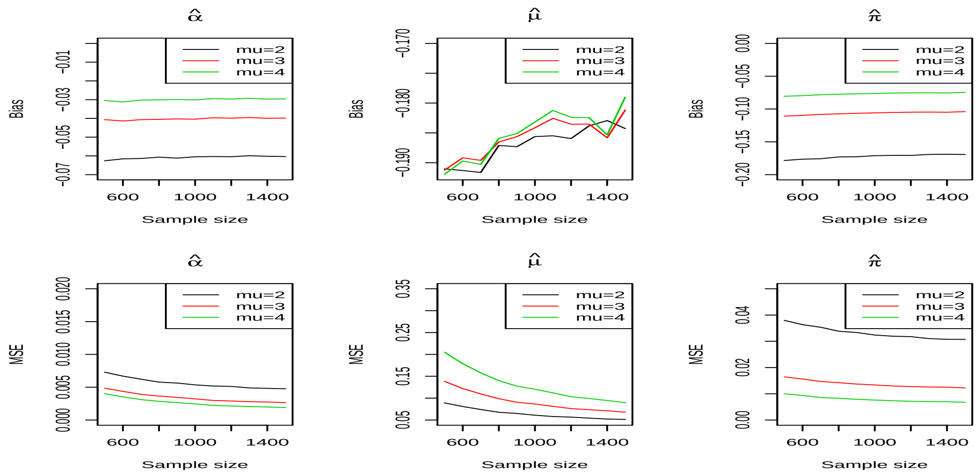

Figure 3: Bias and MSE of the simulated YW estimates of α, μ and π for $\alpha = 0.3$, $\pi = 0.5$, different values of μ.

bias and mean square error (MSE) of the estimators of α, μ and π. The results are presented in Tables 1 and 2. Based on Tables 1 and 2, we observe that as the sample size T increases, the YW estimates converge to the true values.

Figures 2 and 3 show the biases and MSEs of the simulated YW estimators of α, μ and π. Due to the Figures 2 and 3 and Tables 1 and 2, we can see that the bias and MSE of the estimators of α and π decrease by increasing μ and the MSE of the estimates of μ, by increasing μ, decrease when $\pi < 0$ and increase when $\pi > 0$.

Table 1: YW Estimates of the parameters, bias and MSE (in parentheses) of the estimates of α, μ and π for $\alpha = 0.3$ and different values of μ and π.

Sample size	π	$\hat{\alpha}_{YW}$	$\hat{\mu}_{YW}$	$\hat{\pi}_{YW}$
		$(\alpha, \mu) = (0.3, 0.5)$		
$T = 500$	-1	0.1767(0.0194)	0.5678(0.0078)	$-1.1141(0.0408)$
Bias		-0.1233	0.0677	-0.1141
	-0.8	0.1659(0.0219)	0.5632(0.0073)	$-0.9841(0.0620)$
Bias		-0.1340	0.0632	-0.1842
$T = 800$	-1	0.1772(0.0177)	0.5663(0.0065)	$-1.1113(0.0295)$
Bias		-0.1227	0.0664	-0.1112
	-0.8	0.1659(0.0205)	0.5616(0.0059)	$-0.9810(0.0501)$
Bias		-0.1340	0.0616	-0.1810
$T = 1500$	-1	0.1778(0.0162)	0.5656(0.0054)	$-1.1074(0.0210)$
Bias		-0.1222	0.0656	-0.1074
	-0.8	0.1676(0.0188)	0.5613(0.0049)	$-0.9742(0.0397)$
Bias		-0.1324	0.0613	-0.1742
		$(\alpha, \mu) = (0.3, 0.7)$		
$T = 500$	-1	0.2344(0.0092)	0.7450(0.0070)	$-1.0517(0.0225)$
Bias		-0.0655	0.0451	-0.0517
	-0.8	0.2157(0.0115)	0.7494(0.0074)	$-0.8982(0.0294)$
Bias		-0.0843	0.0494	-0.0982
$T = 800$	-1	0.2338(0.0074)	0.7455(0.0052)	$-1.0469(0.0147)$
Bias		-0.0662	0.0455	-0.0469
	-0.8	0.2171(0.0096)	0.7485(0.0055)	$-0.8915(0.0207)$
Bias		-0.0829	0.0485	-0.0915
$T = 1500$	-1	0.2342(0.0059)	0.7442(0.0036)	$-1.0438(0.0087)$
Bias		-0.0657	0.0442	-0.0438
	-0.8	0.2166(0.0084)	0.7486(0.0040)	$-0.8887(0.0142)$
Bias		-0.0834	0.0486	-0.0887
		$(\alpha, \mu) = (0.3, 0.9)$		
$T = 500$	-1	0.2790(0.0057)	0.9182(0.0076)	$-1.0212(0.0163)$
Bias		-0.0209	0.0182	-0.0212
	-0.8	0.2528(0.0067)	0.9338(0.0081)	$-0.8513(0.0175)$
Bias		-0.0472	0.0338	-0.0513
$T = 800$	-1	0.2787(0.0037)	0.9169(0.0046)	$-1.0174(0.0100)$
Bias		-0.0212	0.0169	-0.0174
	-0.8	0.2537(0.0050)	0.9335(0.0054)	$-0.8449(0.0116)$
Bias		-0.0462	0.0335	-0.0449
$T = 1500$	-1	0.2794(0.0021)	0.9163(0.0027)	$-1.0134(0.0054)$
Bias		-0.0206	0.0163	-0.0134
	-0.8	0.2542(0.0036)	0.9315(0.0033)	$-0.8444(0.0070)$
Bias		-0.0457	0.0315	-0.0444

5 Real time series data studies

In this section, the ZMGINMA(1) model is fitted to two real data sets with excess and deficit of zeros. Then, the results for the proposed model is compared with the GINMA(1) model (Alzaid and Al-Osh, 1988).

Table 2: YW Estimates of the parameters, Bias and MSE (in parentheses) of the estimates of α, μ and π for $\alpha = 0.3$ and different values of μ and π.

Sample size	π	$\hat{\alpha}_{YW}$	$\hat{\mu}_{YW}$	$\hat{\pi}_{YW}$
		$(\alpha, \mu) = (0.3, 2)$		
T=500	0.2	0.2503 (0.0058)	1.9463(0.0387)	0.0812(0.0209)
Bias		-0.0497	-0.0536	-0.1188
	0.5	0.2368(0.0073)	1.8071(0.0885)	0.3214(0.0380)
Bias		−0.0631	−0.1929	−0.1786
$T = 800$	0.2	0.2511(0.0044)	1.9496(0.0244)	0.0856(0.0173)
Bias		−0.0489	−0.0503	−0.1144
	0.5	0.2389(0.0059)	1.8144(0.0675)	0.3275(0.0337)
Bias		−0.0610	−0.1856	−0.1725
$T = 1500$	0.2	0.2526(0.0034)	1.9534(0.0141)	0.0910(0.0142)
Bias		−0.0474	−0.0460	−0.1090
	0.5	0.2398(0.0047)	1.8178(0.0505)	0.3311(0.0306)
Bias		−0.0602	−0.1822	−0.1689
		$(\alpha, \mu) = (0.3, 3)$		
$T = 500$	0.2	0.2682(0.0042)	2.9493(0.0702)	0.1293(0.0103)
Bias		−0.0318	−0.0506	−0.0707
	0.5	0.2584(0.00481)	2.8055(0.1423)	0.3885(0.0167)
Bias		−0.0415	−0.1944	−0.1114
$T = 800$	0.2	0.27(0.0029)	2.9557(0.0450)	0.1332(0.0078)
Bias		−0.0300	−0.0443	−0.0667
	0.5	0.2595(0.0036)	2.8110(0.0993)	0.3928(0.0142)
Bias		−0.0405	−0.1890	−0.1072
$T = 1500$	0.2	0.2707(0.0019)	2.9577(0.0243)	0.1362(0.0059)
Bias		−0.0293	−0.0423	−0.0637
	0.5	0.2609(0.0026)	2.8169(0.0681)	0.3959(0.0122)
Bias		−0.0391	−0.1830	−0.1040
		$(\alpha, \mu) = (0.3, 4)$		
$T = 500$	0.2	0.2759(0.0037)	3.9533(0.1089)	0.1487(0.0072)
Bias		−0.0241	−0.0466	−0.0513
	0.5	0.2692(0.0041)	3.8099(0.2111)	0.4199(0.01)
Bias		−0.0307	−0.1901	−0.0801
$T = 800$	0.2	0.2786(0.0024)	3.9629(0.0715)	0.1543(0.0050)
Bias		−0.0214	−0.0371	−0.0456
	0.5	0.2699(0.0028)	3.8108(0.1452)	0.4222(0.0083)
Bias		−0.0301	−0.1891	−0.0777
$T = 1500$	0.2	0.2791(0.0015)	3.9618(0.0382)	0.1564(0.0034)
Bias		−0.0208	−0.0382	−0.0436
	0.5	0.2706(0.0019)	3.8178(0.0908)	0.4249(0.0068)
Bias		−0.0294	−0.1821	−0.0751

5.1 Number of rubella cases

The first example assumes the number of rubella cases, monthly from Aug 2012 - Dec 2018 in Spain. The sample path, autocorrelation and partial autocorrelation functions of the series are shown in Figure 4. Due to this Figure 4, we deduce that an INMA(1) model can be appropriate for modeling this data set, because there exists a cut-off after lag 1 in the sample autocorrelation. The sample mean, variance and empirical index of dispersion are, respectively, 0.43, 1.54 and 3.59. Since the index of dispersion exceeds 1, the rubella series is overdispersed. The proportion of zeros in the rubella series is

%74 which it shows exist inflation of zeros in the series. Thus, a zero-inflated count time series model must be assumed for modeling this series.

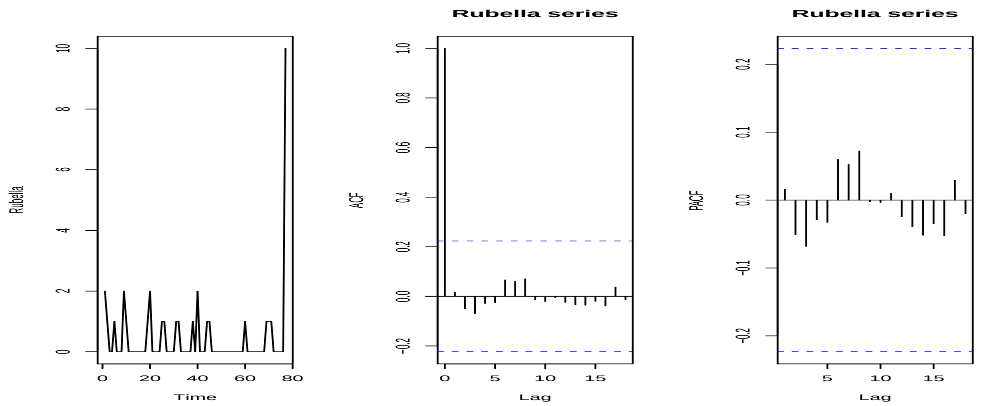

Figure 4: The sample path, ACF and PACF plots of the number of rubella cases, monthly from Aug 2012 - Dec 2018 in Spain.

We fit the ZMGINMA(1) and GINMA(1) models to this data set. For both models, we calculate the YW estimates, lower and upper bounds of the %95 confidence intervals (by performing bootstrap resampling) of the parameters and the root mean square of difference between observations and predicted values (RMS). The results are reported in Table 3. According to this Table, we observed that the ZMGINMA(1) model presents a better fit and forecast than the GINMA(1) model to this series, because its RMS is smaller. Figure 5 shows the plot of the rubella data series and their predicted values which are obtained by (7).

The positive value of the estimate of π in Table 3 shows the inflation of zeros in the ZMGINMA(1) model. The proportions of zeros based on the estimated ZMGINMA(1) and GINMA(1) models are %82 and %70, respectively. Now, we test the hypothesis $H_0 : \pi = 0$(GINMA(1)) versus the hypothesis $H_1 : \pi \neq 0$(ZMGINMA(1)). Due to the Table 3, since the confidence interval of π does not contain the zero value, thus, we reject H_0. Therefore, the ZMGINMA(1) model is suitable for fitting to this data set.

Table 3: Estimated parameters, lower and upper bounds of the %95 confidence intervals of the parameters and RMS for the rubella series.

Model	Parameter	YW Estimates	Lower	Upper	RMS
	α	0.02	0.0034	0.3734	
ZMGINMA(1)	μ	1.52	0.4527	2.9989	1.2262
	π	0.72	0.3909	0.8621	
GINMA(1)	α	0.02	0.0038	0.3475	1.2265
	μ	0.42	0.2472	0.5930	

5.2 Number of earthquakes

The second example assumes the number of earthquakes magnitude 8.0 to 9.9, annually from 1970-2018 in the world. The sample path, autocorrelation and partial autocor-

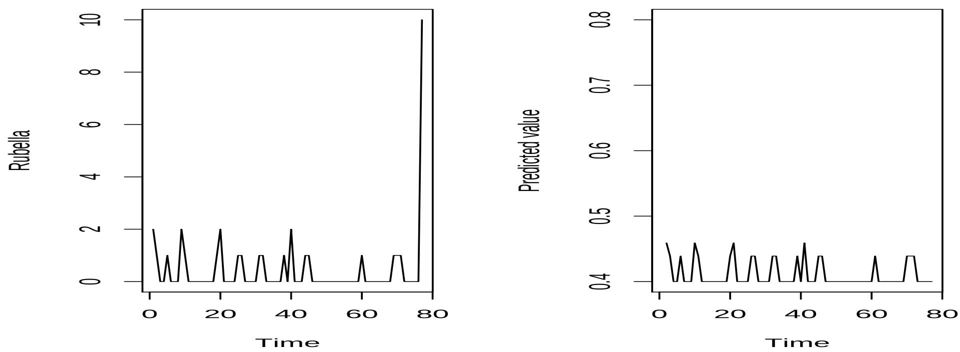

Figure 5: rubella data and their predicted values.

relation functions of the series are shown in Figure 6. Due to the Figure 6, one can realize that an INMA(1) process can be suitable for modeling this data set, because exists a cut-off after lag 1 in the sample autocorrelation. The sample mean, variance and empirical index of dispersion are, respectively, 0.8, 0.75 and 0.94. Since the index of dispersion lower than 1, the earthquakes series is underdispersed. The proportion of zeros in the earthquakes series is %42.9 which it shows exist deflation of zeros in the series. Thus, this series must be modeled by a zero-deflated count time series model.

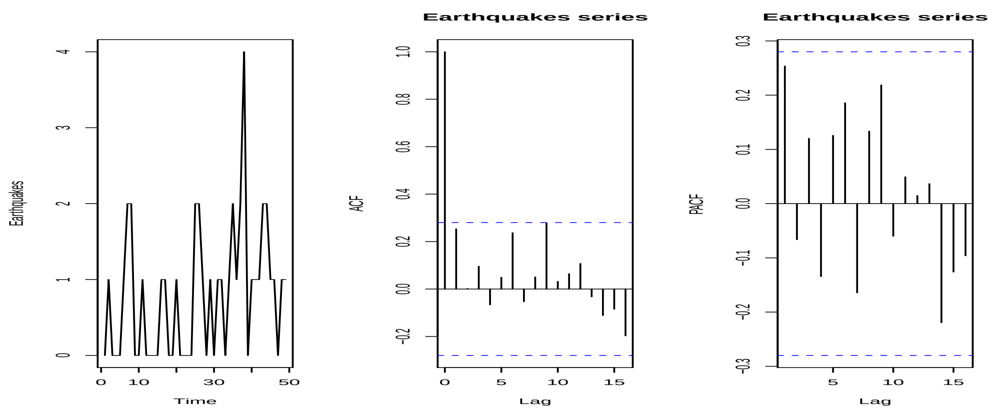

Figure 6: The sample path, ACF and PACF plots of the number of earthquakes magnitude 8.0 to 9.9, annually from 1970-2018 in the world.

We fit the ZMGINMA(1) and GINMA(1) models to this data set. For both models, we obtain the Yule-Walker estimates, lower and upper bounds of the %95 confidences interval (by performing bootstrap resampling) of the parameters and RMS. The results are reported in Table 4. According to the values of Table 4, we observed that the ZMGINMA(1) model gives a better fit and forecast than the GINMA(1) model to this series, because its RMS is smaller. Figure 7 shows the plots of the earthquakes data series and their predicted values which are obtained by (7). The negative value of the estimate of π in Table 4 shows the deflation of zeros in the ZMGINMA(1) model. The proportions of zeros based on the estimated ZMGINMA(1) and GINMA(1) models are %43.1 and %52, respectively. Now, we test the hypothesis $H_0 : \pi = 0$ (GINMA(1)) versus the hypothesis $H_1 : \pi \neq 0$(ZMGINMA(1)). Due to Table 4, since the confidence interval of π does not contain the zero value, thus, we reject H_0. Therefore, the ZMGINMA(1) model is suitable for fitting to this data.

Table 4: Estimated parameters, lower and upper bounds of the %95 confidence intervals of the parameters and RMS for the earthquakes series.

Model	Parameter	YW Estimates	Lower	Upper	RMS
	α	0.35	0.0121	0.5832	
ZMGINMA(1)	μ	0.26	0.1081	0.6229	0.8329
	π	-1.27	-5.1330	-0.3998	
GINMA(1)	α	0.35	0.0121	0.6244	0.8334
	μ	0.591	0.5422	1.2367	

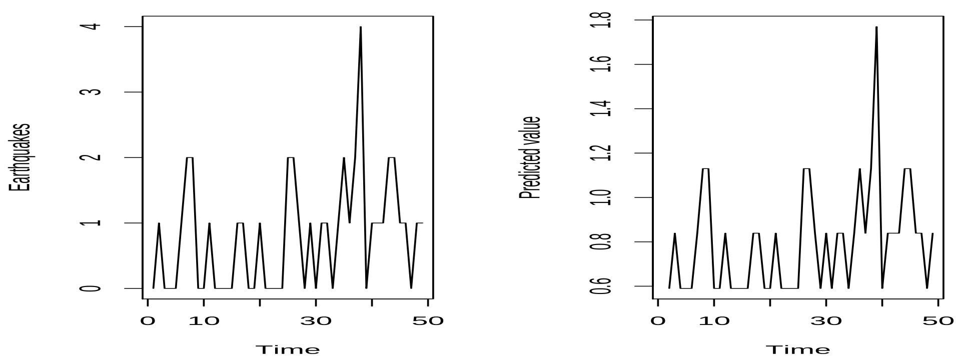

Figure 7: Earthquakes data and their predicted values.

6 Discussion and conclusions

This paper provides a new INMA(1) process generated by the binomial thinning operator. This process is appropriate for modeling zero-inflated or zero-deflated count time series that have a very short-run autocorrelation. Some statistical properties of the process have been obtained. The estimators of the model parameters are obtained using the YW method and also, are shown these estimators are consistent. Then the performance of the YW estimators are evaluated via simulation. At the end, we apply the proposed model to two real data sets and conclude that the model gives the better performance the GINMA(1) model for fitting and predicting future values of zero-inflated and zero-deflated count data.

Acknowledgement

The authors thank the anonymous referees for their valuable comments and careful reading, which led to an improvement of the presentation and results of the article.

References

Al-Osh, M. and Alzaid, A.A. (1988). Integer-valued moving average (INMA) process. *Statistical Papers*, **2**(1):281–300.

Alzaid, A.A. and Al-Osh, M. (1988). First-order integer-valued autoregressive (INAR(1)) process: Distributional and regression properties. *Statistica Neerlandica*, **42**(1):53–61.

Bakouch, H.S., Mohammadpour, M. and Shirozhan, M. (2018). A zero-inflated geometric INAR(1) process with random coefficient. *Applications of Mathematics*, **63**(1):79–105.

Barreto-Souza, W. (2015). Zero-modified geometric INAR(1) process for modeling count time series with deflation or inflation of zeros. *Journal of Time Series Analysis*, **36**(6):839–852.

Bourguinon, M. (2018). Modelling time series of counts with deflation or inflation of zeros. *Statistics and Its Interface*, **11**(4):631–639.

Jazi, M.A., Jones, G. and Lai, C.D. (2012). First order-integer valued AR processes with zero inflated poisson innovations. *Journal of Time Series Analysis*, **33**(6):954–963.

Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions, Statistics: Textbooks and Monographs*. New York: Markel Dekker.

Li, C., Wang, D. and Zhang, H. (2015). First-order mixed integer-valued autoregressive processes with zero-inflated generalized power series innovations. *Journal of the Korean Statistical Society*, **44**(2):232–246.

Mahmoudi, E. and Rostami, A. (2020). First-order integer-valued moving average process with power series innovations. *Journal of Statistical Theory and Applications*, **19**(3):415–431.

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, **20**(4):822–835.

Steutel, F.W. and van Harn, K. (1979). Discrete analogues of self decomposability and stability. *The Annals of Probability*, **7**(5):893–899.

Yu, K. and Zou, H. (2015). The combined Poisson INMA(q) models for time series of counts. *Journal of Applied Mathematics*, **2015**:457842.