



First-Order Integer Valued AR Processes with Zero-Inflated Innovations

Aldo M. Garay¹(✉), Franciyelle L. Medina¹, Isaac Jales C. S.²,
and Patrice Bertail³

¹ Department of Statistics, Federal University of Pernambuco, Recife, Brazil
`{agaray,francy}@de.ufpe.br`

² Department of Mathematics and Statistics,
State University of Rio Grande do Norte, Mossoró, Brazil
`isaacjales@uern.br`

³ Université Paris Nanterre, MODAL'X, Paris, France
`patrice.bertail@parisnanterre.fr`

Abstract. To deal with time series process with excess of zeros, we extend the INAR(1) process by considering that the innovations follow different zero-inflated models, called the ZI-INAR(1) model. We present some of its theoretical properties, develop an efficient EM algorithm for parameter estimation and propose several bootstrap techniques to construct confidence intervals for the parameters. Finally, we present the relevance and applicability the of proposed ZI-INAR(1) model through simulation studies and an application to a real dataset.

Keywords: Count time series · ZINAR(1) process · Bootstrap · EM algorithm

1 Introduction

Recently the modeling and analysis of count time series have been carried out by considering the integer-valued autoregressive (INAR) process, introduced by [3, 22, 29]. In this process, the innovations are assumed to follow a Poisson distribution, but in practice, the innovations may be overdispersed. A frequent manifestation of this incidence of zero counts greater than expected.

In the context of non-negative integer value time series with excess zeros, [15] proposed the ZINAR(1) process, which assumes that the innovations of the process follow a zero-inflated Poisson distribution. The authors developed some structural properties of the process and estimated the unknown parameters by conditional or approximate full maximum likelihood.

To analyze discrete time series with excess of zeros, in this manuscript we extend the ZINAR(1) model, considering that the innovations follow a class of zero-inflated (ZI) models, called the ZI-INAR(1) model. Some properties, such as mean, variance and joint distribution are developed. We propose an EM algorithm to estimate the unknown parameters by maximizing the conditional likelihood function. For this purpose, we start by introducing some definitions,

notations and properties about the first-order integer valued autoregressive processes INAR(1) and the zero-inflated (ZI) models.

Definition 1. Let X be a non-negative integer-valued random variable (r.v.) and $\alpha \in [0, 1]$. Then, the *thinning* operator ‘ \circ ’ [29] is defined as:

$$\alpha \circ X = \sum_{i=1}^X Z_i, \quad (1)$$

where $\{Z_i\}_{i \geq 1}$ is a sequence of independent identically distributed (iid) Bernoulli random variables independent of X , with $\mathbb{P}(Z_i = 1) = \alpha$.

Thus, considering Definition 1, the INAR(1) processes have the following stochastic structure:

$$Y_t = \alpha \circ Y_{t-1} + V_t, \quad t \in \mathbb{Z}, \quad (2)$$

where $\{V_t\}_{t \in \mathbb{Z}}$ is a sequence of non-negative integer-valued random iid variables, called *innovations*, with $E(V_t) = \mu < \infty$, $Var(V_t) = \sigma^2 < \infty$ and independent of Y_{t-1} , for all t .

It is important to note that given $Y_{t-1} = y_{t-1} > 0$, the r.v. $\alpha \circ Y_{t-1}$ follows a binomial distribution, with parameters y_{t-1} and α , and given $Y_{t-1} = 0$, we have that $\alpha \circ Y_{t-1}$ is a r.v. degenerate at zero. As discussed by [3, 9], if $\alpha \in (0, 1)$, then $\{Y_t\}_{t \in \mathbb{Z}}$ is a stationary process, whereas $\alpha = 0$ and $\alpha = 1$ imply, respectively, independence and non-stationarity. Also, if $0 \leq \alpha < 1$, the INAR(1) process is second-order stationary.

Several zero-inflated models have been proposed in the literature to accommodate simultaneously both overdispersion and excess zeros: (i) the zero-inflated Poisson (ZIP) model, [19], in which the zero counts can come from two sources: from the Poisson distribution (*sampling zeros*) or from Bernoulli distribution (*structural zeros*); (ii) the zero-inflated negative binomial distribution (ZINB), [10, 12, 28], which is obtained by mixing a Bernoulli distribution with a baseline negative binomial (NB) distribution over the ZIP model, among others.

In the following, we define the zero-inflated distributions, through their hierarchical formulation and introduce some further properties and three particular cases.

Definition 2. The discrete r.v. V follows a zero-inflated (ZI) distribution, with parameters ρ and $\boldsymbol{\lambda}$, if it has the following stochastic representation:

$$V = BU, \quad B \perp U,$$

where B is a Bernoulli r.v., with $\mathbb{P}(B = 1) = 1 - \rho$ and $0 \leq \rho < 1$. U is a non-negative discrete r.v., with probability mass function (pmf) $h_U(u|\boldsymbol{\lambda})$. $\boldsymbol{\lambda}$ can be a scalar or vector parameter, indexing the distribution of U . $B \perp U$ indicates that the r.v. B and U are independent.

As consequence of Definition 2, we can obtain the pmf of V , given by:

$$\mathbb{P}(V = v) = \begin{cases} \rho + (1 - \rho) h_U(0|\boldsymbol{\lambda}) & v = 0 \\ (1 - \rho) h_U(v|\boldsymbol{\lambda}) & v \geq 1, \end{cases} \quad (3)$$

where $h_U(v|\boldsymbol{\lambda}) = \mathbb{P}(U = v)$. We denote $V \sim \text{ZI}(\rho, \boldsymbol{\lambda}; h_U(\cdot))$.

$$E[V] = (1 - \rho) E[U] \quad \text{and} \quad \text{Var}[V] = (1 - \rho) (\text{Var}[U] + \rho E^2[U]). \quad (4)$$

The distribution of the r.v. U determines the form of the ZI distribution. Thus, we describe briefly three particular cases of the flexible ZI models:

- *The zero-inflated Poisson (ZIP) model:*

In this case, we assume that the r.v. U follows a Poisson distribution, with mean λ . Thus, a pmf of the r.v. V , defined in Eq. (3), takes the form:

$$\mathbb{P}(V = v) = \begin{cases} \rho + (1 - \rho) e^{-\lambda}, & v = 0 \\ (1 - \rho) \frac{e^{-\lambda} \lambda^v}{v!} & v \geq 1. \end{cases} \quad (5)$$

We use the notation $V \sim \text{ZIP}(\rho, \lambda)$ so that:

$$E[V] = (1 - \rho) \lambda \quad \text{and} \quad \text{Var}[V] = (1 - \rho) \lambda (1 + \rho \lambda).$$

- *The zero-inflated negative binomial (ZINB) model:*

The ZINB distribution is the result of considering that the r.v. U follows a negative binomial distribution, as in Eq. (3). We denote by $V \sim \text{ZINB}(\rho, \mu, \phi)$ and its pmf is given by:

$$\mathbb{P}(V = v) = \begin{cases} \rho + (1 - \rho) \left(\frac{\phi}{\mu + \phi} \right)^\phi, & v = 0 \\ (1 - \rho) \frac{\Gamma(\phi + v)}{\Gamma(v + 1) \Gamma(\phi)} \left(\frac{\mu}{\mu + \phi} \right)^v \left(\frac{\phi}{\mu + \phi} \right)^\phi, & v \geq 1 \end{cases}$$

where $\mu \geq 0$ and $\phi > 0$ is the dispersion parameter. $\Gamma(\cdot)$ represents the gamma function. We have that:

$$E[V] = (1 - \rho) \mu \quad \text{and} \quad \text{Var}[V] = (1 - \rho) \mu \left(1 + \frac{\mu}{\phi} + \rho \mu \right).$$

When $\rho = 0$, the r.v. V follows a negative binomial distribution, with mean μ and dispersion parameter ϕ , denoted by $V \sim \text{NB}(\mu, \phi)$. For more details, see [10–12, 21, 28].

- *The zero-inflated Poisson inverse Gaussian (ZIPIG) model:*

This case arises when we consider that U , given in Eq. (3), follows a Poisson inverse Gaussian (PIG) distribution. The pmf of the r.v. V , denoted by $V \sim \text{ZIPIG}(\rho, \lambda, \phi)$, is:

$$\mathbb{P}(V = v) = \begin{cases} \rho + (1 - \rho) e^{\phi - \sqrt{\phi(\phi + 2\lambda)}}, & v = 0; \\ (1 - \rho) \sqrt{\frac{2}{\rho}} [\phi(\phi + 2\lambda)]^{-\frac{(v-1/2)}{2}} \frac{e^{\phi(\lambda\phi)^v}}{v!} K_{v-1/2} \left(\sqrt{\phi(\phi + 2\lambda)} \right), & v \geq 1 \dots \end{cases}$$

where, $\mu > 0$ is the mean, ϕ is a dispersion parameter and $K_\lambda(t) = \frac{1}{2} \int_0^\infty u^{\lambda-1} e^{-\frac{t}{2}(u+\frac{1}{u})} du$ is the modified Bessel function of the third kind, [1]. When $\rho = 0$, the r.v. V follows a PIG distribution, with mean λ and dispersion parameter ϕ , denoted by $V \sim \text{PIG}(\mu, \phi)$, and we have that:

$$E[V] = (1 - \rho)\mu \quad \text{and} \quad \text{Var}[V] = (1 - \rho)\mu \left(1 + \frac{\mu}{\phi} + \rho\mu\right).$$

The ZIPIG distribution is a particular case of the mixed Poisson (MP) distribution, with hierarchical representation: $V|Z = z \sim \text{Poisson}(\mu z)$, where Z follows an Inverse Gaussian (IG) distribution, with mean 1 and dispersion parameter ϕ , denoted by $Z \sim \text{IG}(1, \phi)$. More details are presented and discussed by [4, 16].

This manuscript is organized as follows. Section 2 outlines the proposed ZI-INAR(1) model and discusses some mathematical properties, including the likelihood function. Section 3 presents the implementation of an EM-type algorithm to estimate the parameters of the model and proposes several bootstrap methods to construct confidence intervals for the parameters. The suitability and applicability of the process are illustrated in Sects. 4 and 5, through simulation studies and analysis of real dataset, respectively. Finally, Sect. 6 concludes with short remarks and some possible avenues for future research.

2 The ZI-INAR(1) Process

The ZI-INAR(1) process is an integer-valued first-order autoregressive process, with ZI innovations, as presented in Eq. (2), given by:

$$Y_t = \alpha \circ Y_{t-1} + V_t, \quad t \in \mathbb{Z}, \quad (6)$$

where $V_t \sim \text{ZI}(\rho, \lambda; h_U(\cdot))$.

2.1 Mathematical Properties

In this subsection, we present some mathematical and structural properties of the corresponding marginal distributions of the process:

Proposition 1. *Let $\{Y_t\}_{t \in \mathbb{Z}}$ be a stationary ZI-INAR(1) process; then:*

$$E[Y_t] = \frac{(1 - \rho)E[U_t]}{1 - \alpha} \quad \text{and} \quad \text{Var}[Y_t] = \frac{(1 - \rho)(\alpha E[U_t] + \rho E[U_t]^2 + \text{Var}[U_t])}{1 - \alpha^2},$$

where U_t denotes a r.v. with density function (or pmf) $h_U(u|\lambda)$, for all $t \in \mathbb{Z}$.

Proof. By using the expectation obtained in Eq. (4), we have that:

$$\begin{aligned} E[Y_t|Y_{t-1}] &= E[\alpha \circ Y_{t-1} + V_t|Y_{t-1}] \\ &= E[\alpha \circ Y_{t-1}|Y_{t-1}] + E[V_t|Y_{t-1}] \\ &= \alpha Y_{t-1} + (1 - \rho) E[U_t]. \end{aligned} \quad (7)$$

Thus,

$$\begin{aligned}
 E[Y_t] &= E[E(Y_t|Y_{t-1})] \\
 &= E[\alpha Y_{t-1} + (1 - \rho) E(U_t)] \\
 &= \alpha E[Y_{t-1}] + (1 - \rho) E[U_t].
 \end{aligned} \tag{8}$$

Equation (8) is a consequence of the fact that the process is stationary, that is, $E[Y_t] = \mu_Y$, for all $t \in \mathbb{Z}$.

To obtain the variance of the process, we use:

$$\begin{aligned}
 \text{Var}[Y_t|Y_{t-1} = y_{t-1}] &= \text{Var}[\alpha \circ Y_{t-1} + V_t|Y_{t-1} = y_{t-1}] \\
 &= \text{Var}[\alpha \circ Y_{t-1}|Y_{t-1} = y_{t-1}] + \text{Var}[V_t|Y_{t-1} = y_{t-1}] \\
 &= \alpha(1 - \alpha)y_{t-1} + (1 - \rho) (\text{Var}[U_t] + \rho E^2[U_t]).
 \end{aligned} \tag{9}$$

Thus, by (7) and (9) we obtain:

$$\begin{aligned}
 \text{Var}[Y_t] &= \text{Var}[E(Y_t|Y_{t-1})] + E[\text{Var}(Y_t|Y_{t-1})] \\
 &= \alpha^2 \text{Var}[Y_{t-1}] + \alpha(1 - \alpha)E[Y_{t-1}] + (1 - \rho)(\text{Var}[U_t] + \rho E^2[U_t]) \\
 &= \alpha^2 \text{Var}[Y_{t-1}] + \alpha(1 - \rho)E[U_t] + (1 - \rho)(\text{Var}[U_t] + \rho E^2[U_t]),
 \end{aligned} \tag{10}$$

where (10) is obtained by the expectation given in Eq. (8).

2.2 The Likelihood Function

Let $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ be realizations of the ZI-INAR(1) process. Then the likelihood function of the unknown parameters $\boldsymbol{\theta} = (\alpha, \rho, \boldsymbol{\lambda})^\top$, given \mathbf{y} , can be written as:

$$\begin{aligned}
 \mathbb{L}(\boldsymbol{\theta}|\mathbf{y}) &= \mathbb{P}(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) \\
 &= \mathbb{P}(Y_1 = y_1) \mathbb{P}(Y_2 = y_2|Y_1 = y_1) \dots \mathbb{P}(Y_n = y_n|Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) \\
 &= \mathbb{P}(Y_1 = y_1) \prod_{t=2}^n \mathbb{P}(Y_t = y_t|Y_{t-1} = y_{t-1}),
 \end{aligned} \tag{11}$$

where

$$\begin{aligned}
 \mathbb{P}(Y_t = y_t|Y_{t-1} = y_{t-1}) &= \sum_{k=0}^{\min\{y_{t-1}, y_t\}} \binom{y_{t-1}}{k} \alpha^k (1 - \alpha)^{y_{t-1}-k} \\
 &\quad \times [\rho \mathbb{I}_{\{0\}}(y_t - k) + (1 - \rho)h(y_t - k|\boldsymbol{\lambda})].
 \end{aligned} \tag{12}$$

$\mathbb{I}_{\mathcal{A}}(\cdot)$ denotes the indicator function, i.e., $\mathbb{I}_{\mathcal{A}}(y) = 1$, if $y \in \mathcal{A}$ and $\mathbb{I}_{\mathcal{A}}(y) = 0$ otherwise.

Note that Eq. (12) represents the probability transition of a stationary Markov chain, of state y_{t-1} to y_t . Thus, the marginal probability function can be defined by:

$$\mathbb{P}(Y_t = y_t) = \sum_{y_{t-1}=0}^{\infty} \mathbb{P}(Y_{t-1} = y_{t-1}) \mathbb{P}(Y_t = y_t|Y_{t-1} = y_{t-1}). \tag{13}$$

From Eqs. (11)–(13), we have that:

$$\begin{aligned} \mathbb{L}(\boldsymbol{\theta}|\mathbf{y}) = \mathbb{P}(Y_1 = y_1) \prod_{t=2}^n \sum_{k=0}^{\min\{y_{t-1}, y_t\}} \binom{y_{t-1}}{k} \alpha^k (1-\alpha)^{y_{t-1}-k} \\ \times [\rho \mathbb{I}_{\{0\}}(y_t - k) + (1-\rho)h(y_t - k|\boldsymbol{\lambda})]. \end{aligned} \quad (14)$$

The marginal distribution is intractable, so a simple approach to deal with this is to condition it on the observed Y_1 , and estimate the parameters by conditional maximum likelihood (CML); see [3, 15]. To estimate the parameters of the ZI-INAR(1) process by maximizing this conditional likelihood function directly prevents the possibility of analytical solutions. One alternative is to maximize the complete conditional likelihood using the expectation-maximization (EM) algorithm [6], which is stable and straightforward to implement, since the iterations converge monotonically and no second derivatives are required. In the next section, we discuss a technique to find the ML estimates of the parameters vector $\boldsymbol{\theta}$, based on the EM algorithm.

3 Maximum Likelihood Estimation and Bootstrap Resampling Methods

3.1 Parameter Estimation via the EM Algorithm

In this section we develop an EM-type algorithm for maximum likelihood estimation of the parameters of the ZI-INAR(1) process. The key to the development of our EM-type algorithm is to consider the presence of latent variables and treat the problem as if these variables were in fact observed.

As suggested by [12, 25], we define the latent variables $\mathbf{W} = (W_1, \dots, W_n)^\top$ and $\mathbf{S} = (S_1, \dots, S_n)^\top$ for all $t \geq 1$, where:

- S_t is defined by $S_t = \alpha \circ Y_{t-1}$. Thus $S_t|Y_{t-1} = y_{t-1}, \alpha \sim \text{Bin}(y_{t-1}; \alpha)$, if $y_{t-1} > 0$ and it is a degenerate distribution at zero if $y_{t-1} = 0$, where $\text{Bin}(y; \alpha)$ represents the binomial distribution with parameters y and $\alpha \in [0, 1]$.
- W_t is a latent dichotomous variable, that is:

$$W_t = \begin{cases} 1 & \text{if } V_t \text{ is from the zero state} \\ 0 & \text{if } V_t \sim h(\cdot|\boldsymbol{\lambda}), \end{cases}$$

with $\mathbb{P}(W_t = 1) = \rho$ and $h(\cdot|\boldsymbol{\lambda})$ as presented in Definition 2.

Let $\mathbf{Y}_c = (\mathbf{Y}, \mathbf{W}, \mathbf{S})$ be the complete data vector, where \mathbf{Y} and $\{\mathbf{W}, \mathbf{S}\}$ represent the observed data and the missing data, respectively, with $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$. Then the joint probability function for $\mathbf{Y}_{c_t} = (Y_t, S_t, W_t)$ is given by:

$$\begin{aligned}
\mathbb{P}(\mathbf{Y}_{c_t} = \mathbf{y}_{c_t}) &= \mathbb{P}(Y_t = y_t, S_t = s_t, W_t = w_t | Y_{t-1} = y_{t-1}) \\
&= \mathbb{P}(Y_t = y_t, W_t = w_t | S_t = s_t, Y_{t-1} = y_{t-1}) \mathbb{P}(S_t = s_t | Y_{t-1} = y_{t-1}) \\
&= \left[\rho^{w_t} ((1 - \rho) h(y_t - s_t | \lambda))^{1-w_t} \right] \\
&\quad \times \binom{y_{t-1}}{s_t} \alpha^{s_t} (1 - \alpha)^{y_{t-1} - s_t}, \quad \text{for all } t \geq 1.
\end{aligned}$$

Thus, the complete likelihood function is defined by:

$$\begin{aligned}
\mathbb{L}_c(\boldsymbol{\theta} | \mathbf{y}_c) &= \mathbb{P}(Y_1 = y_1) \prod_{t=2}^n \mathbb{P}(Y_t = Y_t, S_t = s_t, W_t = w_t | Y_{t-1} = y_{t-1}) \\
&\propto \prod_{t=2}^n \left\{ \rho^{w_t} ((1 - \rho) h(y_t - s_t | \lambda))^{1-w_t} \binom{y_{t-1}}{s_t} \alpha^{s_t} (1 - \alpha)^{y_{t-1} - s_t} \right\}.
\end{aligned}$$

Hence, the complete log-likelihood function is given by:

$$\begin{aligned}
\ell_c(\boldsymbol{\theta} | \mathbf{y}_c) &= \log(\mathbb{L}_c(\boldsymbol{\theta} | \mathbf{y}_c)) \\
&\propto \sum_{t=2}^n w_t \log(\rho) + \sum_{t=2}^n (1 - w_t) \log(1 - \rho) + \sum_{t=2}^n (1 - w_t) \log(h(y_t - s_t | \lambda)) \\
&\quad + \sum_{t=2}^n s_t \log(\alpha) + \sum_{t=2}^n (y_{t-1} - s_t) \log(1 - \alpha).
\end{aligned}$$

The EM algorithm has several appealing properties relative to other iterative algorithms such as Newton-Raphson and Fisher's scoring method for finding MLEs, [23]. One of them is that the sequence of estimates from the EM algorithm increases the likelihood function $\ell(\boldsymbol{\theta} | \mathbf{y})$ at each iteration, and under standard regularity conditions the sequence converges to a stationary point of the likelihood.

Thus, the EM-algorithm proceeds in two steps:

- **E-step:** Let $\hat{\boldsymbol{\theta}}^{(k)}$ be the current k -th step estimate of $\boldsymbol{\theta}$. By using the property of conditional expectation, we compute the $Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)})$ function given by:

$$Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)}) = E \left[\ell_c(\boldsymbol{\theta} | \mathbf{y}_c) | \mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)} \right]. \quad (15)$$

- **M-step:** Maximize $Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$, obtaining $\hat{\boldsymbol{\theta}}^{(k+1)}$.

Observe that the expression of the Q-function, in Eq. (15), is determined by the knowledge of the following expectations:

$$\begin{aligned}
\hat{s}_t^{(k)} &= E \left[S_t | \mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)} \right] \\
\hat{w}_t^{(k)} &= E \left[W_t | \mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)} \right] \quad \text{and} \\
Q_t^*(\boldsymbol{\lambda} | \boldsymbol{\theta}^{(k)}) &= E \left[(1 - W_t) \log \left(h(y_t - S_t | \boldsymbol{\lambda}) | \mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)} \right) \right].
\end{aligned}$$

Thus, the Q-function can be written in more simple form as:

$$\begin{aligned} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) &\propto \sum_{t=2}^n \hat{w}_t^{(k)} \log(\rho) + \sum_{t=2}^n \left(1 - \hat{w}_t^{(k)}\right) \log(1 - \rho) + \sum_{t=2}^n Q_t^*(\boldsymbol{\lambda}|\boldsymbol{\theta}^{(k)}) \\ &\quad + \sum_{t=2}^n \hat{s}_t^{(k)} \log(\alpha) + \sum_{t=2}^n \left(y_{t-1} - \hat{s}_t^{(k)}\right) \log(1 - \alpha). \end{aligned}$$

At each step, $\hat{s}_t^{(k)}$ and $\hat{w}_t^{(k)}$ can be obtained by using the following results:

$$\begin{aligned} \mathbb{P}\left(S_t = s_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) &= \frac{\mathbb{P}\left(S_t = s_t, Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\ &= \frac{\mathbb{P}\left(Y_t = y_t | S_t = s_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) \times \mathbb{P}\left(S_t = s_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} \mathbb{P}\left(Y_t = y_t, S_t = s_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\ &= \frac{\binom{y_{t-1}}{s_t} (\alpha^{(k)})^{s_t} (1 - \alpha^{(k)})^{y_{t-1} - s_t} \left[\rho^{(k)} \mathbb{I}_{\{0\}}(y_t - s_t) + (1 - \rho^{(k)}) h(y_t - s_t | \lambda^{(k)})\right]}{\sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} \mathbb{P}\left(Y_t = y_t, S_t = s_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \quad (16) \end{aligned}$$

and

$$\begin{aligned} \mathbb{P}\left(W_t = w_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) &= \frac{\mathbb{P}\left(W_t = w_t, Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{P\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\ &= \frac{\mathbb{P}\left(W_t = w_t, Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} \mathbb{P}\left(Y_t = y_t, S_t = s_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}. \end{aligned}$$

Then, for all $t \geq 1$, we have that:

$$\hat{s}_t^{(k)} = \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} s_t \mathbb{P}\left(S_t = s_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) \quad (17)$$

$$\begin{aligned} \hat{w}_t^{(k)} &= 1 \times \mathbb{P}\left(W_t = 1 | \mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)}\right) \\ &= \frac{\mathbb{P}\left(W_t = 1, Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} \mathbb{P}\left(Y_t = y_t, S_t = s_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}. \quad (18) \end{aligned}$$

Therefore, our EM algorithm for the ZI-INAR(1) process can be summarized in the following way:

- **E-step:** Given $\hat{\boldsymbol{\theta}}^{(k)}$, for $t \geq 1$ we compute $\hat{s}_t^{(k)}$, $\hat{w}_t^{(k)}$, as given in Eq. (17)–(18), respectively, and $Q_t^*(\boldsymbol{\lambda}|\boldsymbol{\theta}^{(k)})$.

- **M-step:** Update $\hat{\boldsymbol{\theta}}^{(k)}$ by maximizing $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$ over $\boldsymbol{\theta}$, which leads to the following expressions:

$$\begin{aligned}\hat{\alpha}^{(k+1)} &= \frac{\sum_{t=2}^n \hat{s}_t}{\sum_{t=2}^n y_{t-1}}, \\ \hat{\rho}^{(k+1)} &= \frac{\sum_{t=2}^n \hat{w}_t}{n-1}, \\ \hat{\lambda}^{(k+1)} &= \arg \max_{\lambda} \left\{ \sum_{t=2}^n Q_t^*(\lambda|\boldsymbol{\theta}^{(k)}) \right\}.\end{aligned}\quad (19)$$

In the following, we develop the procedure to obtain the expressions $Q_t^*(\lambda|\boldsymbol{\theta}^{(k)})$ and $\hat{\lambda}^{(k+1)}$, considering the three particular cases of the ZI models seen before, i.e., when $V \sim \text{ZI}(\rho, \lambda)$. For this, we define the expectation $\widehat{b_t s_t}$ with $B_t = 1 - W_t$, given by:

$$\begin{aligned}\widehat{b_t s_t} &= E[B_t S_t | \mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)}] \\ &= \frac{(1 - \hat{\rho}^{(k)}) \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} s_t \binom{y_{t-1}}{s_t} (\hat{\alpha}^{(k)})^{s_t} (1 - \hat{\alpha}^{(k)})^{y_{t-1}-s_t} h(y_t - s_t | \hat{\lambda}^{(k)})}{\mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)})}.\end{aligned}$$

Thus,

- If $V_t \sim \text{ZIP}(\rho, \lambda)$, then from Eq. (5) $h(\cdot|\lambda)$ represents a pmf of the Poisson distribution, with parameter λ . Consequently,

$$Q_t^*(\lambda|\boldsymbol{\theta}^{(k)}) \propto -\lambda(1 - \hat{w}_t^{(k)}) + \log(\lambda)(1 - \hat{w}_t^{(k)})y_t + \log(\lambda)\widehat{b_t s_t}^{(k)} \quad (20)$$

and from Eq. (19) and (20):

$$\hat{\lambda}^{(k+1)} = \frac{\sum_{t=2}^n (1 - \hat{w}_t^{(k)})y_t - \sum_{t=2}^n \widehat{b_t s_t}^{(k)}}{\sum_{t=2}^n (1 - \hat{w}_t^{(k)})}.$$

- If $V_t \sim \text{ZINB}(\rho, \mu, \phi)$, then $h(\cdot|\lambda)$ represents the pmf of the negative binomial distribution with parameters $\lambda = (\mu, \phi)$. Thus,

$$\begin{aligned}Q_t^*(\lambda|\boldsymbol{\theta}^{(k)}) &\propto \widehat{g_t(\phi)}^{(k)} + \left(\log(\mu) - \log(\mu + \phi) \right) \left((1 - \hat{w}_t^{(k)})y_t - \widehat{b_t s_t}^{(k)} \right) \\ &\quad + \left(-\log(\Gamma(\phi)) + \phi(\log(\phi) - \log(\mu + \phi)) \right) \left(1 - \hat{w}_t^{(k)} \right)\end{aligned}\quad (21)$$

with

$$\begin{aligned}\widehat{g_t(\phi)}^{(k)} &= E[B_t \log \Gamma(y_t - S_t + \phi) | \mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)}] \\ &= \frac{(1 - \hat{\rho}^{(k)}) \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} \log \Gamma(y_t - s_t + \phi) \mathbb{P}(S_t = s_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}) h(y_t - s_t | \hat{\boldsymbol{\theta}}^{(k)})}{\mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)})}.\end{aligned}\quad (22)$$

When the M-step turns out to be analytically intractable, it can be replaced by a sequence of conditional maximization (CM) steps. The procedure is known as the ECM algorithm [24]. Thus, from Eq. (19), (21) and (22), we have that for the ZINB-INAR(1) process, $\hat{\boldsymbol{\lambda}}^{(k+1)} = (\hat{\mu}^{(k+1)}, \hat{\phi}^{(k+1)})$ are given by:

$$\hat{\mu}^{(k+1)} = \frac{\sum_{t=2}^n (1 - \hat{w}_t^{(k)}) y_t - \sum_{t=2}^n \widehat{b_t s_t}^{(k)}}{\sum_{t=2}^n (1 - \hat{w}_t^{(k)})}.$$

$$\hat{\phi}^{(k+1)} = \arg \max_{\phi} \left\{ \sum_{t=2}^n Q_t^*(\mu^{(k+1)}, \phi | \hat{\theta}^{(k)}) \right\}$$

$\hat{\phi}^{(k+1)}$ is obtained using the “*optim*” routine in the \mathbb{R} software [27].

- If $V_t \sim \text{ZIPIG}(\rho, \mu, \phi)$, then $\boldsymbol{\lambda} = (\mu, \phi)$ and Its hierarchical representation is:

$$V_t | Z_t = z_t, W_t = 0 \sim \text{Poisson}(\mu z_t)$$

$$V_t | Z_t = z_t, W_t = 1 \text{ follows a degenerate distribution at zero;} \quad (23)$$

$$Z_t | W_t = 0 \sim IG(1, \phi), \quad (24)$$

$$W_t \sim \text{Bin}(1, \rho), \quad t = 2, 3, \dots, \quad (25)$$

the density function of $Z_t | W_t = w_t$ is:

$$f(z_t | w_t) = \left\{ \phi^{1/2} z_t^{-3/2} \exp \left(-\frac{\phi}{2} \frac{(z_t - 1)^2}{z_t} \right) \right\}^{1-w_t}. \quad (26)$$

In order to develop the EM algorithm for the ZIPIG-INAR model, we add the latent variable Z_t . Thus, the complete-data is defined by $\mathbf{Y}_c = (\mathbf{Y}, \mathbf{W}, \mathbf{S}, \mathbf{Z})$. Thus,

$$Q_t^*(\mu, \phi | \theta^{(k)}) \propto -\mu \widehat{b_t z_t}^{(k)} - \log(\mu) \widehat{b_t s_t}^{(k)} + \log(\mu) (1 - \widehat{w}_t^{(k)}) y_t$$

$$+ \frac{\log(\phi)}{2} (1 - \widehat{w}_t^{(k)}) + \phi (1 - \widehat{w}_t^{(k)}) - \frac{\phi}{2} \left(\widehat{b_t z_t}^{(k)} + \widehat{b_t / z_t}^{(k)} \right), \quad (27)$$

where the expectations $\widehat{b_t z_t}^{(k)}$ and $\widehat{b_t / z_t}^{(k)}$ are defined by:

$$\widehat{b_t z_t}^{(k)} = E \left[E \left(B_t Z_t | S_t, \mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)} \right) \right]$$

$$= \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} \left\{ \left(1 - \hat{\rho}^{(k)} \right) (y_t - s_t + 1) h \left(y_t - s_t + 1 | \hat{\boldsymbol{\lambda}}^{(k)} \right) \right.$$

$$\times \mathbb{P} \left(S_t = s_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)} \right)$$

$$\times \left. \frac{1}{\hat{\mu}^{(k)} \left[\hat{\rho}^{(k)} \mathbb{I}_{\{y_t=s_t\}} + (1 - \hat{\rho}^{(k)}) h \left(y_t - s_t | \hat{\boldsymbol{\lambda}}^{(k)} \right) \right]} \right\} \quad (28)$$

and

$$\begin{aligned}
\widehat{b_t/z_t}^{(k)} &= E \left[E \left(B_t Z_t^{-1} | S_t, \mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)} \right) \right] \\
&= \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} \left\{ \frac{\widehat{\mu}^{(k)} h(y_t - s_t - 1 | \widehat{\boldsymbol{\lambda}}^{(k)}) \mathbb{P}(S_t = s_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)})}{(y_t - s_t) h(y_t - s_t | \widehat{\boldsymbol{\lambda}}^{(k)})} \mathbb{I}_{\{y_t > s_t\}} \right. \\
&\quad + \left(\sqrt{\widehat{\rho}^{(k)} (\widehat{\rho}^{(k)} + 2\widehat{\mu}^{(k)})} + 1 \right) (1 - \widehat{\rho}^{(k)}) h(0 | \widehat{\boldsymbol{\lambda}}^{(k)}) \\
&\quad \times \mathbb{P}(S_t = y_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)}) \\
&\quad \left. \times \frac{1}{\widehat{\rho}^{(k)} [\widehat{\rho}^{(k)} + (1 - \widehat{\rho}^{(k)}) h(0 | \widehat{\boldsymbol{\lambda}}^{(k)})]} \mathbb{I}_{\{y_t = s_t\}} \right\}, \tag{29}
\end{aligned}$$

where $\mathbb{P}(S_t = s_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)})$ is given in Eq. (16). Thus, from Eq. (19) and (27), we have that:

$$\begin{aligned}
\widehat{\mu}^{(k+1)} &= \frac{\sum_{t=2}^n (1 - \widehat{w}_t^{(k)}) y_t - \sum_{t=2}^n \widehat{b_t s_t}^{(k)}}{\sum_{t=2}^n \widehat{b_t z_t}^{(k)}}. \\
\widehat{\phi}^{(k+1)} &= \frac{\sum_{t=2}^n (1 - \widehat{w}_t^{(k)})}{\sum_{t=2}^n \widehat{b_t z_t}^{(k)} + \sum_{t=2}^n \widehat{b_t/z_t}^{(k)} - 2 \sum_{t=2}^n (1 - \widehat{w}_t^{(k)})}.
\end{aligned}$$

This process is iterated until some convergence rule is satisfied. Here, we use the Aitken acceleration-based stopping criterion [23] as a convergence rule. This criterion is based on the fact that the limit of the sequence $\ell^{(k+1)} = \ell(\widehat{\boldsymbol{\theta}}^{(k+1)} | \mathbf{y})$, denoted by ℓ_∞ , can be approximated by $\ell_\infty^{(k+1)} = \ell^{(k)} + (\ell^{(k+1)} - \ell^{(k)}) / (1 - c^{(k)})$, where $c^{(k)} = (\ell^{(k+1)} - \ell^{(k)}) / (\ell^{(k)} - \ell^{(k-1)})$. As suggested by [31], we decided to stop the algorithm when $\left\| \ell_\infty^{(k+1)} - \ell^{(k+1)} \right\| < \varepsilon = 10^{-5}$.

3.2 Bootstrap Resampling Methods

In this section, we propose several bootstrap alternatives to construct confidence intervals for the parameter. Actually, we can indifferently use the posterior mean of the Bayesian approach or the standard maximum likelihood estimators (mle), or any other efficient estimators. In any case, the standard error can be difficult to compute, so several bootstrap approaches can be used to solve this issue. Here we propose two types of approaches, based respectively on some strong parametric assumptions, on the mixing properties of the INAR(p) process and on the Markov property of this process.

(a) **Parametric bootstrap approach:**

Let $\theta = (\alpha, \rho, \lambda)$ be the parameters of the ZI-INAR(1) process, where we denote $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\rho}_n, \hat{\lambda}_n)$ as the mle estimator obtained using the EM algorithm developed in Sect. 3.1. The properties of mle for general INAR(p) models are studied, for instance, in [9], who showed strong consistency of the mle, and [20] who studied the efficiency and asymptotic normality in Proposition 6.1. Other estimators, based on saddle point approximation of the likelihood, which may be easier to implement, were proposed in [26]. In the following, we denote by $\hat{\theta}_n$ any kind of asymptotically efficient estimator of the parameter θ .

The parametric bootstrap method simply consists of generating new data in the model with estimated parameters. That is:

Step 1: Generate the random variable V_t^* from the ZI distribution, with parameters $\hat{\theta}_n$.

Step 2: Generate a ZI-INAR(1) model recursively as follows:

$$\begin{aligned} Y_1^* &= y_1 \\ Y_t^* &= \hat{\alpha}_n \circ Y_{t-1}^* + V_t^*, \quad t = 2, \dots, n, \end{aligned}$$

meaning that if $Y_{t-1}^* > 0$, then $\hat{\alpha}_n \circ Y_{t-1}^* | Y_{t-1}^* \stackrel{d}{=} \sum_{i=1}^{Y_{t-1}^*} Z_i^*$ with $Z_i^* \sim \text{Bin}(1, \hat{\alpha}_n)$ or $\hat{\alpha}_n \circ Y_{t-1}^* | Y_{t-1}^*$ is degenerate at zero if $Y_{t-1}^* = 0$.

Notice that the first observation of the bootstrap ZI-INAR(1) process is set to the first observation of the observed data: asymptotically, even if the generated process is not stationary, this will not perturb the asymptotic properties of the bootstrap process. Now the bootstrap counterparts of $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\rho}_n, \hat{\lambda}_n)$, say $\hat{\theta}_n^* = (\hat{\alpha}_n^*, \hat{\rho}_n^*, \hat{\lambda}_n^*)$, are the mle obtained using our proposed EM algorithm, considering the bootstrap process $\mathbf{Y}^* = (Y_1^*, \dots, Y_n^*)$.

Step 3: Since the exact bootstrap distribution of $\hat{\theta}_n^*$ may be difficult (and time consuming) to compute, it is replaced by a Monte Carlo approximation that is repeats the procedure in Step 2, B times, by generating times series $\mathbf{y}^{*(b)}$, $b = 1, \dots, B$, for realization of the bootstrap process \mathbf{Y}^* .

Then we can compute the corresponding $\hat{\theta}_n^{*(b)}$, $b = 1, \dots, B$ estimators. The centered and normalized bootstrap distribution is then given by, for $\mathbf{u} = (u_1, \dots, u_4) \in \mathbb{R}^4$:

$$K_n^{*(B)}(\mathbf{u}) = \frac{1}{B} \sum_{b=1}^B \mathbb{I}_{\{\sqrt{n}(\hat{\theta}_n^{*(b)} - \hat{\theta}_n) \leq \mathbf{u}\}}.$$

If the bootstrap distribution is asymptotically valid, that is, it is a convergent distribution (at least in probability) of $\mathbb{P}(\sqrt{n}(\hat{\theta}_n - \theta) \leq \mathbf{u})$ then it is easy to use the quantile of the Monte Carlo bootstrap distribution to construct (simultaneous) confidence intervals for the parameter θ or any regular (at least differentiable) functional of the parameter.

Concerning the choice of B , it is known from the work by [13] that one needs to choose B so that it at least has the same order as n and such that $(B+1)(1-\gamma)/2$ is an integer (when constructing asymptotic two-sided confidence intervals of level γ . So typically for $\gamma = 95\%$ (for size of n smaller than 1000) we choose $B = 999$.

Proposition 2. *Assume that the parameter space of $\theta = (\alpha, \rho, \lambda)$ with $\lambda = (\mu, \phi)$ is given by $\Theta =]0, 1[\times]0, 1[\times]0, \infty[\times]0, \infty[$ and consider $\hat{\theta}_n$ to be the mle of the ZI-INAR(1) process. Then the parametric bootstrap is asymptotically correct, meaning that almost surely along the sample,*

$$\sup_{\mathbf{u} \in \mathbb{R}^4} \left| \mathbb{P}(\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \leq \mathbf{u}) - \mathbb{P}(\sqrt{n}(\hat{\theta}_n - \theta) \leq \mathbf{u}) \right| \rightarrow 0 \text{ when } n \rightarrow \infty$$

Proof: In [5], Theorem 2.1 gives necessary and sufficient conditions for the validity of the bootstrap of mle estimators or efficient estimators in regular models (in the LeCam sense). These conditions essentially reduce to two conditions: 1) LAN (locally asymptotically normal) property of the log-likelihood at the true parameter (see Definition 2.1 in [5]); and 2) LAE (locally asymptotically equivariant) condition for the estimator $\hat{\theta}_n$ (see Definition 2.2 in [5]). In the open domain Θ , we have that $l(\theta|\mathbf{y})$ is twice differentiable with a non degenerate information matrix, which implies differentiability in quadratic mean in the LeCam sense (see [30], Lemma 7.6, p. 95). As a consequence, it has the LAN property (see 7.14 and 7.15 in [30] p. 104). As far as the LAE condition is concerned, this follows from the fact that the ZI-INAR(1) model may be written as a regular model on the domain Θ (see 7.16, p. 104 as well as the references in [26]): notice that any efficient estimator has the LAE property, so that one may use indifferently the mle in the construction or the estimators proposed in [26].

(b) **Moving block bootstrap approach**

When no specific assumption is made about the distribution of the residuals, it is still possible to implement semiparametric estimators of the ZI-INAR process, as described in [26]. In that case, a parametric bootstrap process cannot be used. Moreover, to assess the robustness of the parametric assumptions, it may also be interesting to implement a more robust version of the bootstrap process. The more general method is based on splitting the original time series into overlapping blocks which are then resampled to reconstruct the original time series. The procedure in our case is as follows:

Step 1: Choose a length b (which will typically be of size $b = o(\sqrt{n})$). Define the overlapping blocks

$$B_1 = (Y_1, \dots, Y_b), B_2 = (Y_2, \dots, Y_{b+1}), \dots, B_{n-b+1} = (Y_{n-b+1}, \dots, Y_n)$$

Use circular block bootstrap, or even better, apply stationary block bootstrap (see [7]), which allows simulating a stationary version of the moving block bootstrap, in place of the moving block bootstrap method.

- Step 2: Draw without replacement $\lfloor \frac{n}{b} \rfloor + 1$ blocks, which are bound together (and possible truncated at the end) to form a new time series of size n . After this, compute the statistics of interest, either the mle or estimators based on approximations (see [26]) of the corresponding time series.
- Step 3: Just like in the Step 3 of the parametric bootstrap, use a Monte Carlo method to obtain an approximation of the bootstrap distribution.

It has been shown in [17] that under some strong mixing conditions and provided that $b_n \rightarrow \infty$ and $\frac{b_n}{\sqrt{n}} \rightarrow 0$, the moving block bootstrap is asymptotically valid. A lot of variation and modifications have been proposed in the literature to obtain valid second-order approximations (see for instance [18] for a complete overview and references). However, it should be noticed that according to the distribution of V_t , the INAR(p) process may not strongly mix. The process can be shown to strongly mix when the residuals V_t have Poisson distribution. However, it can be shown that the process is only weakly mixing, or ψ -weak mixing for more general distributions; see [8]. This includes the ZI cases studied here, when $0 < \alpha < 1$, since the moments of all residuals are finite. The validity of the stationary moving block bootstrap method (as well as variations including the circular block bootstrap) in the weak mixing case has been studied in [14], see Theorem 3.2. As noticed by the authors, their theorem can be easily extended to functionals of means, Frechet differentiable functionals (which can be simply linearized) or even Hadamard differentiable functionals (including M-estimators). As a consequence, this method will be asymptotically valid for the mle, which is LAE (it is less obvious for the estimators proposed in [26]).

3.3 Confidence Intervals

In the second simulation study, we use and compare two different types of asymptotically valid confidence intervals, mainly bootstrap variance based asymptotic intervals and the standard percentile method. We briefly describe these methods in our context.

Consider $\hat{\theta}_k$ the components of the mle of $\theta = (\theta_1, \dots, \theta_4) = (\alpha, \rho, \mu, \phi)$ and denote by $\hat{\theta}_{k,l}^{(b)}$, $b = 1, \dots, B$, the corresponding Bootstrap replications obtained at the Monte-Carlo step, either by the parametric ($l = 1$) or moving block ($l = 2$) techniques. In the simulation we will choose B such that $(B + 1)(1 - \gamma)/2$ is an integer, as suggested by [13].

3.3.1 Variance Based Asymptotic Interval

The bootstrap variance of θ_k is simply given by:

$$\hat{V}_{k,l}^* = \frac{1}{B} \sum_{b=1}^B \left(\hat{\theta}_{k,l}^{*(b)} - \frac{1}{B} \sum_{i=1}^B \hat{\theta}_{k,l}^{*(i)} \right)^2.$$

Thus, the asymptotic normal approximation leads to an asymptotic γ confidence interval of the type

$$\left[\hat{\theta}_k - u_{\frac{1+\gamma}{2}} \hat{V}_{k,l}^{*1/2}, \hat{\theta}_k + u_{\frac{1+\gamma}{2}} \hat{V}_{k,l}^{*1/2} \right],$$

where $u_{\frac{1+\gamma}{2}}$ is a quantile of order $\frac{1+\gamma}{2}$ of a standard normal distribution.

3.3.2 Standard Percentile Method

Define respectively $\hat{\theta}_{k,l(\frac{1-\gamma}{2})}^*$ and $\hat{\theta}_{k,l(\frac{1+\gamma}{2})}^*$ as the $\frac{1-\gamma}{2}$ and $\frac{1+\gamma}{2}$ quantiles of the empirical distribution of $\hat{\theta}_{k,l}^{*(b)}$, $b = 1, \dots, B$. Then the standard percentile confidence interval (which has the same asymptotic property as the preceding one) is given by:

$$\left[\hat{\theta}_{k,l(\frac{1-\gamma}{2})}^*, \hat{\theta}_{k,l(\frac{1+\gamma}{2})}^* \right].$$

4 Simulation Study

In this section we show the performance of the estimation procedure for the ZI-INAR(1) process, using the EM algorithm proposed and bootstrap resampling methods described before. Thus, we present examples considering artificial datasets, in different scenarios. All the computational codes were implemented using the \mathbb{R} software [27] and the program codes are available from us on request.

We present two simulation studies. The first one investigate the asymptotic properties of the maximum likelihood estimation. In the second one, we use and compare different type of asymptotically valid confidence intervals.

4.1 Simulation 1: Asymptotic Properties

We consider the set of sample sizes $n \in \{100, 300, 500, 1000\}$ to analyze the consistency of the mle of parameters. We fix $\alpha = 0.3$, $\rho \in \{0.3, 0.6\}$. Thus, by considering the eight different combinations of n and ρ , we simulate $N = 300$ Monte Carlo replicates for the following ZI-INAR(1) processes: (i) ZIP-INAR(1) with $\lambda = 2$; (ii) ZINB-INAR(1) and ZIPIG-INAR(1) with $\mu = 2$ and $\phi \in \{0.75, 1.5, 2.5\}$. We compute the relative bias (RB) and root relative square error (RRSE), defined by:

$$\text{RB}(\hat{\theta}_i) = \frac{1}{N} \sum_{j=1}^N \frac{\hat{\theta}_{ij} - \theta_i}{\theta_i} \quad \text{and} \quad \text{RRSE}(\hat{\theta}_i) = \sqrt{\frac{1}{N} \sum_{j=1}^N \left(\frac{\hat{\theta}_{ij} - \theta_i}{\theta_i} \right)^2},$$

where $\hat{\theta}_{ij}$ is the mle of parameter θ_i , computed in the j -th sample.

Tables 1, 2 and 3 show that both RB and RRSE decrease when the sample size increases, for all the parameter estimates.

Table 1. RB and RRSE of the parameter estimates of ZIP-INAR(1) processes, with $\alpha = 0.3$ and $\lambda = 2$.

ρ	n	$\hat{\alpha}$		$\hat{\rho}$		$\hat{\lambda}$	
		RB	RRSE	RB	RRSE	RB	RRSE
0.3	100	-0.0668	0.3141	-0.0764	0.3436	-0.0035	0.1163
	300	-0.0166	0.1635	-0.0090	0.1943	0.0058	0.0777
	500	-0.0229	0.1188	-0.0128	0.1372	0.0033	0.0579
	1000	-0.0020	0.0870	-0.0150	0.0995	-0.0024	0.0395
0.6	100	-0.0433	0.2362	-0.0212	0.1337	0.0007	0.1582
	300	-0.0118	0.1289	-0.0068	0.0707	-0.0058	0.0895
	500	-0.0117	0.0985	-0.0050	0.0488	0.0021	0.0650
	1000	-0.0095	0.0750	-0.0031	0.0388	-0.0040	0.0472

Table 2. RB and RRSE of the parameter estimates of ZINB-INAR(1) processes, with $\alpha = 0.3$ and $\mu = 2$.

ϕ	ρ	n	$\hat{\alpha}$		$\hat{\rho}$		$\hat{\mu}$		$\hat{\phi}$	
			RB	RRSE	RB	RRSE	RB	RRSE	RB	RRSE
0.75	0.3	100	0.0051	0.2080	-0.0001	0.7010	0.0971	0.3614	0.3417	1.8875
		300	0.0042	0.1117	-0.0365	0.5423	0.0344	0.2355	0.1165	0.6861
		500	0.0022	0.0842	-0.0238	0.4571	0.0143	0.1873	0.0696	0.4845
		1000	0.0019	0.0595	-0.0424	0.3500	-0.0048	0.1376	0.0111	0.2943
	0.6	100	0.0052	0.1893	-0.1443	0.4576	0.0306	0.4736	0.2758	3.2307
		300	-0.0148	0.1091	-0.0688	0.2908	0.0106	0.2980	0.0990	0.9289
		500	0.0083	0.0830	-0.0684	0.2552	-0.0158	0.2584	0.0125	0.6428
		1000	-0.0021	0.0612	-0.0242	0.1472	-0.0076	0.1692	0.0602	0.3940
	1.5	100	-0.0157	0.2241	-0.0666	0.6120	0.0183	0.2399	0.2144	1.4996
		300	-0.0015	0.1219	-0.0834	0.4179	-0.0147	0.1604	0.0298	0.6910
		500	-0.0019	0.0899	-0.0185	0.3245	0.0084	0.1294	0.0585	0.4348
		1000	0.0020	0.0687	-0.0333	0.2436	-0.0043	0.0950	-0.0039	0.2898
1.5	0.3	100	-0.0200	0.2105	-0.1236	0.3800	-0.0118	0.3557	0.0959	2.1109
		300	0.0005	0.1204	-0.0407	0.1930	-0.0130	0.2056	0.0401	1.1545
		500	-0.0109	0.0949	-0.0256	0.1351	-0.0167	0.1607	0.0213	0.6549
		1000	-0.0004	0.0603	-0.0218	0.0883	-0.0184	0.1161	-0.0314	0.3832
	0.6	100	-0.0388	0.2513	-0.1162	0.5348	0.0030	0.2104	0.0319	0.9771
		300	-0.0050	0.1370	-0.0642	0.3803	-0.0077	0.1390	0.1027	0.9715
		500	-0.0212	0.1089	-0.0502	0.2699	-0.0061	0.1037	0.0178	0.5092
		1000	-0.0066	0.0764	-0.0281	0.1829	-0.0082	0.0755	-0.0169	0.3569
	2.5	100	-0.0169	0.2278	-0.1465	0.3334	-0.0959	0.2975	-0.1273	1.3087
		300	-0.0125	0.1194	-0.0165	0.1262	-0.0176	0.1730	0.0525	1.0700
		500	-0.0030	0.0889	-0.0088	0.0954	-0.0022	0.1286	0.0730	0.8590
		1000	-0.0006	0.0697	-0.0065	0.0690	-0.0025	0.0977	<0.001	0.4792

4.2 Simulation 2: Bootstrap Confidence Intervals

In this simulation study we analyze the efficiency of the bootstrap methods presented in Sect. 3. We simulate $N = 100$ samples of size $n = 196$ from the following processes with parameters $\alpha = 0.25$ and $\rho = 0.4$: (i) ZIP-INAR(1), with $\lambda = 4$; and (ii) ZINB-INAR(1) and ZIPIG-INAR(1), with $\mu = 4$ and $\phi = 3$. For each sample, we generate $B = 999$ bootstrap replicates and the length of the circular block bootstrap adopted is $b = 14$.

Tables 4 and 5 show the mean value of the bootstrap estimator and the proportion of times that the confidence intervals contain the true value of the parameter (Cov.), as well as the percentage of times that the parameter value is less than the lower bound ($\theta_k < LL$) and greater than the upper bound ($UL < \theta_k$). In general, the proportion of the intervals containing the true value of the parameters is close to the nominal level (95%). However, these proportions are smaller than 90% for the intervals based on percentiles of the ZINB-INAR(1) process.

Table 3. RB and RRSE of the parameter estimates of ZIPIG-INAR(1) processes, with $\alpha = 0.3$ and $\mu = 2$.

ϕ	ρ	n	$\hat{\alpha}$		$\hat{\rho}$		$\hat{\mu}$		$\hat{\phi}$		
			RB	RRSE	RB	RRSE	RB	RRSE	RB	RRSE	
0.75	0.3	100	0.0010	0.2044	-0.0046	0.5836	0.0391	0.2632	0.1688	1.8383	
		300	-0.0056	0.1175	-0.0665	0.3940	-0.0078	0.1739	0.0431	0.5878	
		500	-0.0002	0.0887	-0.0243	0.2975	0.0004	0.1293	0.0237	0.3576	
		1000	-0.0016	0.0614	-0.0117	0.2135	-0.0004	0.0918	0.0301	0.2578	
	0.6	100	-0.0205	0.2034	-0.0825	0.3159	0.0369	0.3589	0.2246	2.6273	
		300	-0.0007	0.1125	-0.0345	0.1567	-0.0168	0.2096	-0.0025	0.7321	
		500	-0.0073	0.0871	-0.0121	0.1040	-0.0096	0.1539	0.0717	0.5155	
		1000	-0.0027	0.0677	-0.0095	0.0790	-0.0110	0.1204	0.0187	0.3109	
	1.5	0.3	100	-0.0340	0.2193	-0.0740	0.5251	0.0183	0.2156	0.1622	1.3835
			300	-0.0068	0.1367	-0.0243	0.3257	0.0016	0.1346	0.0398	0.7542
			500	-0.0050	0.1032	-0.0162	0.2326	0.0007	0.0959	0.0268	0.4768
			1000	-0.0065	0.0702	-0.0273	0.1705	-0.0001	0.0762	-0.0163	0.2293
0.6		100	-0.0320	0.2046	-0.0452	0.2319	0.0160	0.2772	0.1455	1.6193	
		300	-0.0052	0.1066	-0.0212	0.1103	-0.0034	0.1659	0.0623	0.9166	
		500	-0.0051	0.0913	-0.0096	0.0885	-0.0015	0.1253	0.0543	0.9582	
		1000	0.0022	0.0640	0.0025	0.0573	0.0051	0.0942	0.0481	0.3459	
2.5	0.3	100	-0.0437	0.2472	-0.0750	0.4556	0.0187	0.1978	0.1961	1.3687	
		300	-0.0095	0.1382	-0.0194	0.2708	-0.0006	0.1141	0.0972	0.9488	
		500	-0.0040	0.1188	0.0002	0.2095	0.0061	0.0938	0.0628	0.4610	
		1000	-0.0042	0.0720	-0.0017	0.1384	0.0070	0.0621	0.0338	0.3133	
	0.6	100	-0.0486	0.2160	-0.0850	0.2530	-0.0336	0.2658	-0.0715	1.1974	
		300	-0.0107	0.1299	-0.0161	0.0976	-0.0046	0.1420	0.0965	0.9962	
		500	-0.0088	0.0918	-0.0143	0.0777	-0.0097	0.1051	0.0423	0.7257	
		1000	0.0025	0.0653	-0.0031	0.0503	-0.0027	0.0776	0.0077	0.3785	

Table 4. Parametric bootstrap approach results using confidence intervals.

Model		Value	Boots. mean	Asymptotic interval			Standard percentile method		
				$\theta_k < LL$	$UL < \theta_k$	Cov.	$\theta_k < LL$	$UL < \theta_k$	Cov.
ZIP-INAR(1)	α	0.25	0.2504	3%	4%	93%	3%	6%	91%
	ρ	0.40	0.4038	3%	1%	96%	3%	1%	96%
	μ	4.00	3.9864	1%	4%	95%	1%	4%	95%
ZINB-INAR(1)	α	0.25	0.2430	1%	2%	97%	0%	3%	97%
	ρ	0.40	0.3772	3%	1%	96%	3%	3%	94%
	μ	4.00	3.9360	2%	6%	92%	2%	6%	92%
	ϕ	3.00	4.4043	0%	4%	96%	5%	2%	93%
ZIPIG-INAR(1)	α	0.25	0.2441	4%	3%	93%	4%	3%	93%
	ρ	0.40	0.3863	0%	1%	99%	0%	2%	98%
	μ	4.00	4.0022	7%	1%	92%	7%	1%	92%
	ϕ	3.00	5.0738	0%	2%	98%	4%	0%	96%

Table 5. Circular block bootstrap results using confidence intervals.

Model		Value	Boots. mean	Asymptotic interval			Standard percentile method		
				$\theta_k < LL$	$UL < \theta_k$	Cov.	$\theta_k < LL$	$UL < \theta_k$	Cov.
ZIP-INAR(1)	α	0.25	0.2375	2%	5%	93%	2%	5%	93%
	ρ	0.40	0.3945	3%	3%	94%	3%	3%	94%
	μ	4.00	3.9912	2%	4%	94%	2%	5%	93%
ZINB-INAR(1)	α	0.25	0.2249	0%	9%	91%	0%	11%	89%
	ρ	0.40	0.3575	2%	4%	94%	1%	9%	90%
	μ	4.00	3.9057	3%	9%	88%	3%	9%	88%
	ϕ	3.00	4.1189	0%	7%	93%	4%	7%	89%
ZIPIG-INAR(1)	α	0.25	0.2258	0%	6%	94%	0%	6%	94%
	ρ	0.40	0.3693	0%	8%	92%	0%	9%	91%
	μ	4.00	3.9847	3%	1%	96%	5%	1%	94%
	ϕ	3.00	4.8409	0%	5%	95%	4%	3%	93%

5 Real Dataset: Drug Offenses

To further demonstrate the usefulness of our approach, we analyze a real dataset of the monthly number of drug offenses recorded from January 1990 to December 2001, with 144 observations, in Pittsburgh's census tract 2206. The data were obtained from Forecasting Principles site: <http://www.forecastingprinciples.com>.

Figure 1 contains the panels (a) and (b), which represent the time evolution and the bar plot of the monthly count of drug offenses recorded, respectively. The two plots indicate a significant presence of zero value observations (43%). In panel (c), the partial autocorrelation function (PACF) suggests a first-order dependence structure.

For comparative purposes, we consider the INAR(1) process with Poisson (Po), NB and PIG innovations and ZI-INAR(1) processes.

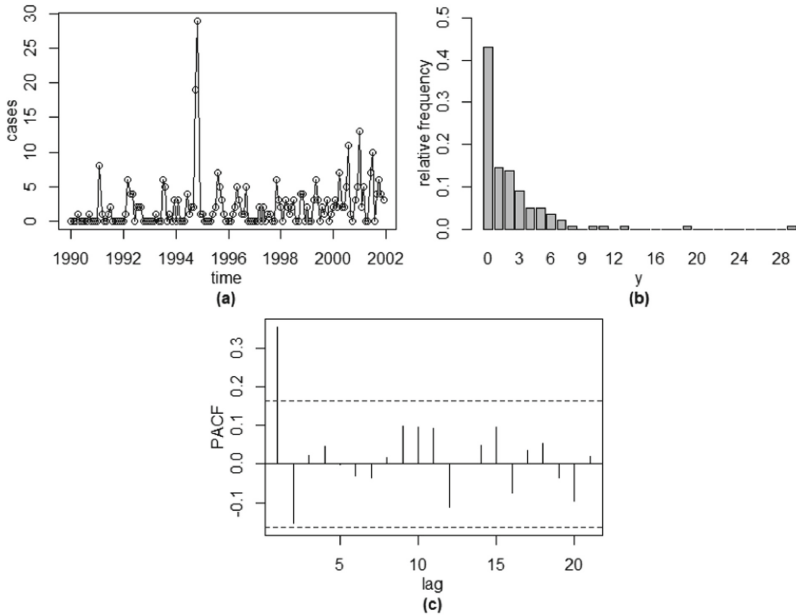


Fig. 1. Drugs offence dataset. (a) Time series plot. (b) Bar plot. (c) PACF.

Table 6 presents the parameter estimates and the Akaike information criterion (AIC) values, [2], of all the considered processes. This criterion indicates that the ZIPIG-INAR(1) process is more appropriate for this dataset. Table 7 shows the parameter estimates of the ZIPIG-INAR(1) process and the confidence interval, using parametric bootstrap (B-PAR) and circular block bootstrap (CBB).

Table 6. Parameter estimates

Processo	$\hat{\alpha}$	$\hat{\rho}$	$\hat{\mu}$	$\hat{\phi}$	AIC
Po-INAR(1)	0.212	-	1.679	-	590.74
ZIP-INAR(1)	0.181	0.512	3.577	-	565.77
NB-INAR(1)	0.071	-	1.977	0.471	550.43
ZINB-INAR(1)	0.070	0.138	2.296	0.630	552.20
PIG-INAR(1)	0.072	-	1.973	0.336	554.53
ZIPIG-INAR(1)	0.065	0.325	2.946	0.903	549.41

Table 7. Parameter estimates of ZIPIG-INAR(1) processes using B-PAR (parametric bootstrap) and CBB (circular block bootstrap).

Parameter	EM	B-PAR	CBB	CI (B-PAR)	CI (CBB)
α	0.065	0.066	0.060	(0; 0.141)	(0; 0.131)
ρ	0.325	0.320	0.325	(0.138; 0.503)	(0.169; 0.482)
μ	2.946	2.957	2.960	(2.029; 3.883)	(2.000; 3.921)
ϕ	0.903	1.095	1.385	(0.005; 2.185)	(0; 4.245)

6 Conclusion

This article presents a study of the ZI-INAR(1) processes, which are an extension of the ZINAR(1) process, introduced by [15]. The aim is to propose models with more flexibility and that can deal with excess zeros. To fit the models, we develop an efficient EM algorithm to estimate the parameters and apply several alternative bootstrap techniques to construct confidence intervals for the parameters. The performance of our proposals was evaluated through simulation studies and application to a real dataset of monthly number of drug offenses recorded. In this illustrative example, there is strong evidence of a first-order dependence structure and zero-inflated counts. We conjecture that our methods can be extended by adding a moving average structure to the process. An in-depth investigation of such extension is beyond the scope of the present manuscript, but is an interesting topic for further research.

Acknowledgements. Aldo M. Garay would like to acknowledge the support of the National Council for Scientific and Technological Development, CNPq-Brazil, Grant 420082/2016-6. This research has also been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01).

References

1. Abramowitz, M., Stegun, I.: Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables Applied Mathematics Series, pp. 295–300. National Bureau of Standards, Washington, DC (1964)
2. Akaike, H.: A new look at the statistical model identification. IEEE Trans. Autom. Control **19**, 716–723 (1974)
3. Al-Osh, M., Alzaid, A.A.: First-order integer-valued autoregressive (INAR (1)) process. J. Time Ser. Anal. **8**(3), 261–275 (1987)
4. Barreto-Souza, W., Simas, A.B.: General mixed Poisson regression models with varying dispersion. Stat. Comput. **26**(6), 1263–1280 (2016)
5. Beran, R.: Diagnosing bootstrap success. Ann. Inst. Statist. Math. **49**(1), 1–24 (1997)
6. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. Ser. B **39**(1), 1–38 (1977)
7. Politis, D.N., Romano, J.P.: The stationary bootstrap. J. Am. Stat. Assoc. **89**(428), 1303–1313 (1994)

8. Doukhan, P., Fokianos, K., Li, X.: On weak dependence conditions: the case of discrete valued processes. *Stat. Probab. Lett.* **82**(11), 1941–1948 (2012)
9. Du, J.G., Li, Y.: The integer valued autoregressive (INAR(p)) model. *J. Time Ser. Anal.* **12**, 129–142 (1991)
10. Garay, A.M., Hashimoto, E.M., Ortega, E.M., Lachos, V.H.: On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Comput. Stat. Data Anal.* **55**, 1304–1318 (2011)
11. Garay, A.M., Lachos, V.H., Bolfarine, H.: Bayesian estimation and case influence diagnostics for the zero-inflated Negative Binomial regression model. *J. Appl. Stat.* **42**(6), 1148–1165 (2015)
12. Hall, D.: Zero-inflated Poisson and Binomial regression with random effects: a case study. *Biometrics* **56**(4), 1030–1039 (2000)
13. Hall, P.: On the number of bootstrap simulations required to construct a confidence interval. *Ann. Statist.* **14**(4), 1453–1462 (1986)
14. Hwang, E., Shin, D.W.: Strong consistency of the stationary bootstrap under ψ -weak dependence. *Stat. Probab. Lett.* **82**(3), 488–495 (2012)
15. Jazi, M.A., Jones, G., Lai, C.-D.: First-order integer valued AR processes with zero inflated Poisson innovations. *J. Time Ser. Anal.* **33**(6), 954–963 (2012)
16. Karlis, D.: A general EM approach for maximum likelihood estimation in mixed Poisson regression models. *Stat. Model.* **1**(4), 305–318 (2001)
17. Kunsch, H.R.: The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17**(3), 1217–1241 (1989)
18. Lahiri, S., *Resampling Methods for Dependent Data*. Springer Series in Statistics. Springer, New York (2003). <https://doi.org/10.1007/978-1-4757-3803-2>
19. Lambert, D.: Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**(1), 1–14 (1992)
20. Latour, A.: Existence and stochastic structure of a non negative integer-valued autoregressive process. *J. Time Ser. Anal.* **19**, 439–455 (1998)
21. Martinez-Flores, G., Bolfarine, H., Gomez, H.: Asymmetric regression models with limited responses with an application to antibody response to vaccine. *Biom. J.* **55**, 156–172 (2013)
22. McKenzie, E.: Some ARMA models for dependent sequences of Poisson counts. *Adv. Appl. Probab.* **20**(4), 822–835 (1988)
23. McLachlan, G.J., Krishnan, T.: *The EM Algorithm and Extensions*, 2nd edn. Wiley, New York (2008)
24. Meng, X.-L., Rubin, D.B.: Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* **80**(2), 267–278 (1993)
25. Neal, P., Subba Rao, T.: MCMC for integer-valued ARMA processes. *J. Time Ser. Anal.* **28**(1), 92–110 (2007)
26. Pedeli, X., Davison, A.C., Fokianos, K.: Likelihood estimation for the INAR(p) model by saddlepoint approximation. *J. Am. Stat. Assoc.* **110**(511), 1229–1238 (1982)
27. R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2019). ISBN 3-900051-07-0
28. Ridout, M., Hinde, J., Demétrio, C.: A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics* **57**(1), 219–223 (2001)
29. Steutel, F.W., Harn, K.V.: Discrete analogues of self-decomposability and stability. *Ann. Probab.* **7**(5), 893–899 (1979)

30. van der Vaart, A.W.: Asymptotic Statistics, 7th edn. Cambridge University Press (2006)
31. Zeller, C.B., Cabral, C.R.B., Lachos, V.H., Benites, L.: Finite mixture of regression models for censored data based on scale mixtures of normal distributions. *Adv. Data Anal. Classif.* **13**(1), 89–116 (2018). <https://doi.org/10.1007/s11634-018-0337-y>