



OPEN

Some developments on seasonal INAR processes with application to influenza data

Fatimah E. Almuhayfith^{1✉}, Emmanuel W. Okereke², Manik Awale³, Hassan S. Bakouch^{4,5} & Hana N. Alqifari⁶

Influenza epidemic data are seasonal in nature. Zero-inflation, zero-deflation, overdispersion, and underdispersion are frequently seen in such number of cases of disease (count) data. To explain these counts' features, this paper introduces a flexible model for nonnegative integer-valued time series with a seasonal autoregressive structure. Some probabilistic properties of the model are discussed for general seasonal INAR(p) model and three estimation methods are used to estimate the model parameters for its special case seasonal INAR(1) model. The performance of the estimation procedures has been studied using simulation. The proposed model is applied to analyze weekly influenza data from the Breisgau- Hochschwarzwald county of Baden–Württemberg state, Germany. The empirical findings show that the suggested model performs better than existing models.

Many epidemic diseases follow a seasonal pattern and so is the Influenza. Influenza shows a seasonal pattern¹ in temperate regions. The weekly/monthly/yearly epidemic data form a time series of counts. Analyzing and forecasting time series of counts remain a useful technique of getting information needed for successful policy making and management of epidemics. Before modeling a count time series, an analyst should be familiar with specific characteristics of the series in order to select a good model for the data. Dispersion characteristic, stationarity, autocorrelation structure and seasonality remain factors that help a researcher to determine the model that should be fitted to a count time series. Seasonality refers to the tendency of a time series (including low count time series) to exhibit patterns or movements that are completed within a year and repeat themselves every year. Seasonality deals with regular and predictable patterns in a time series. For example, a weekly time series of low counts is said to be seasonal with seasonal period $s = 52$ if its associated autocorrelation function has spikes at multiples of lag 52.

New INAR processes have been constructed and used to model stationary nonseasonal count time series from a variety of fields since the independent works of Al-Osh and Alzaid², and McKenzie³, which paved the way for research on the applications and construction of nonnegative integer-valued autoregressive (INAR) processes. Several models were constructed for stationary, overdispersed nonseasonal series based on the binomial thinning operator and first-order autoregressive correlation structure^{4,5}. Authors who used other thinning operators to build models for the count time series include Liu and Zhu⁶, and Ristić et al.⁷. There is also evidence of stationary, first-order integer-valued autoregressive processes, which were constructed using innovation distributions, that can exhibit any of the underdispersion, equidispersion and overdispersion phenomena^{8–10}.

An INAR(1) process with an innovation distribution that is basically a standard discrete or compound Poisson distribution has been described as inappropriate for modeling any of underdispersed and overdispersed series with either inflation or a deflation of zeros¹¹. Zero-modified distributions are useful in modeling count data with an inflation or deflation of zeros. The zero-modified versions of some discrete distributions have been introduced and used as innovation distributions to construct INAR(1) models by some authors. For example, Barreto-Souza¹¹ constructed a zero-modified geometric INAR(1) process using the negative binomial thinning operator and illustrated empirically the applicability of the model using two practical data sets. In particular, the model may be used to describe an underdispersed or overdispersed series with deflationary or inflationary zeros. Sharafi et al.¹² suggested an INAR(1) process using the zero-modified Poisson–Lindley (ZMPL) distributed

¹Department of Mathematics and Statistics, College of Science, King Faisal University, Alahsa 31982, Saudi Arabia. ²Department of Statistics, Michael Okpara University of Agriculture, Umudike, Nigeria. ³Department of Statistics, Savitribai Phule Pune University, Pune 411007, India. ⁴Department of Mathematics, College of Science, Qassim University, Buraydah 51452, Saudi Arabia. ⁵Department of Mathematics, Faculty of Science, Tanta University, Tanta 31111, Egypt. ⁶Department of Statistics and Operation Research, College of Science, Qassim University, Buraydah 51482, Saudi Arabia. ✉email: falmuhaifeez@kfu.edu.sa

innovations. They fitted the model to both zero-deflated and zero-inflated series and compared the model's fit to each data set to that of competing models. Empirically speaking, their model outperformed the other models fitted to each of the data sets.

Sometimes, it is necessary to analyze time series that exhibit seasonal variations. Such time series are called seasonal time series. Seasonal time series of counts are seen in several domains, including epidemiology¹³. Though authors have analyzed several seasonal epidemiological data using the Gaussian multiplicative seasonal autoregressive moving average (SARIMA) model^{14–16} and Holt–Winter's Method¹⁶, these models may not be suitable for certain nonnegative count time series, especially those involving small counts^{17–19}. In the case of seasonal low count time series, a seasonal nonnegative integer-valued autoregressive (INAR) model is needed. Unlike the SARIMA model, the seasonal INAR model takes into consideration the discrete nature of the data through an appropriate discrete innovation distribution. It can also be used to generate integer-valued forecasts. It can also be constructed to account for properties of count time series, such as overdispersion, underdispersion, zero-inflation and zero-deflation. Despite the extensive literature on INAR models, INAR models for stationary seasonal count time series have received little attention^{20–22}. Remarkably, none of the seasonal models was developed for the purpose of analyzing an underdispersed or overdispersed seasonal count time series that is either zero-inflated or zero-deflated. This study intends to achieve two principal objectives. First, to propose a seasonal INAR process of order one that is suitable for underdispersed, overdispersed, zero-inflated and zero-deflated count time series data. Second, application of the model to influenza data. In view of the absence of the general seasonal INAR model in the literature, the theoretical framework is developed for the general seasonal INAR model in this paper.

The remainder of this article is organized as follows: “**Introduction**” discusses the general seasonal INAR model. In “**The general seasonal INAR model**”, we deal with the construction and properties of the general seasonal INAR(p) model. Methods of estimating the parameters of the model are investigated, namely the Yule–Walker and conditional least squares. The theoretical aspect of forecasting based on the proposed model, simulation and real-world data application for a special case of the general seasonal model, namely the INAR(1)_s ZMPL process are considered in “**Construction of the proposed seasonal INAR(1) process and its properties**”. The conclusions of this research are summarized in “**Conclusion**”.

The general seasonal INAR model

In response to the comment made by one of the anonymous reviewers, and considering that the literature lacks a general seasonal INAR model, we introduce the general seasonal INAR model in this section. Let “ $*$ ” be the negative binomial thinning (NBT) operator and $\lambda * W = \sum_{i=1}^W Z_i$, where Z_i refers to the sequence of independent and identically distributed (i.i.d.) geometric variables having the probability mass function (PMF)

$$P[Z_i = w] = \frac{\lambda^w}{(1 + \lambda)^{w+1}}, w \in \mathbb{N}_0, \lambda \in [0, 1).$$

An elaborate discussion of the properties of the NBT operator is available in Ristić et al.⁷. The general seasonal INAR model of order P and seasonal period s (INAR(P)_s model) is defined by

$$W_t = \lambda_1 * W_{t-s} + \lambda_2 * W_{t-2s} + \dots + \lambda_P * W_{t-Ps} + v_t, \quad t \in \mathbb{N}_0. \quad (1)$$

In (1), the counting series are mutually independent, $\lambda_l \in [0, 1]$, $l = 1, 2, \dots, P$ and $\{v_t\}$ are independent of all the counting series.

Theorem 2.1 contains the condition for the existence of a unique stationary INAR(P)_s model.

Theorem 2.1 *Let $\{v_t\}$ be i.i.d., nonnegative integer-valued random variables whose mean and variance are $E(v_t) = \mu_v$ and $\text{Var}(v_t) = \sigma_v^2$ respectively. Suppose that $\lambda_l \in [0, 1]$, $l = 1, 2, \dots, P$. If the roots of the equation*

$$z^P - \lambda_1 z^{P-1} - \lambda_2 z^{P-2} - \dots - \lambda_P = 0$$

lie inside the unit circle, then there is a unique stationary nonnegative integer-valued series $\{W_t\}$ that satisfies the equation

$$W_t = \lambda_1 * W_{t-s} + \lambda_2 * W_{t-2s} + \dots + \lambda_P * W_{t-Ps} + v_t, \quad t \in \mathbb{N}_0.$$

Proof The proof of Theorem 2.1 is based on some properties of the NBT operator. These properties, which were established by Ristić et al.⁷ are as follows

- (i) $E[\prod_{m=1}^r (\lambda_m * W_m)] = \prod_{m=1}^r \lambda_m E(W_m)$, $r \geq 1$
- (ii) $E(\lambda * W)^2 = \lambda^2 E(W)^2 + \lambda(1 + \lambda)E(W)$
- (iii) $E(\lambda * W - \lambda * Y)^2 = \lambda(1 + \lambda)E(|W - Y|) + \lambda^2 E(|W - Y|)^2$, if each of the counting series of $\lambda * W$ and $\lambda * Y$ has the geometric $\left(\frac{\lambda}{1+\lambda}\right)$ distribution. The remaining part of the proof can be deduced from the proof of Theorem 2.1 in Jin-Guan and Yuan²³.

□

Taking expectation of both sides of (1) and assuming stationarity, the mean of the INAR(P)_s model becomes

$$\mu_W = E(W_t) = \frac{\mu_v}{1 - \sum_{l=1}^P \lambda_l} \tag{2}$$

Next, we study the autocorrelation structure of the INAR(P)_{*s*} model. Let $\mathbf{W}_t = (W_t, W_{t-s}, W_{t-2s}, \dots, W_{t-(P+1)s})'$.

Then (1) can be written as

$$\mathbf{W}_t = \mathbf{C} * \mathbf{W}_{t-s} + \mathbf{v}_t, \tag{3}$$

where

$$\mathbf{C} = \begin{pmatrix} \lambda_1 & \lambda_2 & \dots & \lambda_{P-1} & \lambda_P \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \text{ and } \mathbf{v}_t = \begin{pmatrix} v_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The vector of the autocovariances is defined by

$$\begin{aligned} \boldsymbol{\gamma}_k &= E[(\mathbf{W}_t - E(\mathbf{W}_t))(\mathbf{W}_{t-k} - E(\mathbf{W}_{t-k}))^\top] \\ &= E(\mathbf{W}_t \mathbf{W}_{t-k}^\top) - E(\mathbf{W}_t)E(\mathbf{W}_{t-k}^\top) \\ &= \mathbf{C}E(\mathbf{W}_{t-s} \mathbf{W}_{t-k}^\top) + E(\mathbf{v}_t)E(\mathbf{W}_{t-k}^\top) - E(\mathbf{W}_t)E(\mathbf{W}_{t-k}^\top) \\ &= \mathbf{C}\boldsymbol{\gamma}_{(k-s)} + (\mathbf{C} - \mathbf{I})E(\mathbf{W}_{t-s})E(\mathbf{W}_{t-k}^\top) + E(\mathbf{v}_t)E(\mathbf{W}_{t-k}^\top) \end{aligned}$$

Since $E(\mathbf{W}_t) = \mathbf{C}E(\mathbf{W}_{t-s}) + E(\mathbf{v}_t)$, it is easy to verify that $(\mathbf{C} - \mathbf{I})E(\mathbf{W}_{t-s})E(\mathbf{W}_{t-k}^\top) + E(\mathbf{v}_t)E(\mathbf{W}_{t-k}^\top)$ is a null matrix. Thus

$$\boldsymbol{\gamma}_k = \mathbf{C}\boldsymbol{\gamma}_{(k-s)}, \quad k > 0. \tag{4}$$

It follows that the autocovariance function at lag k is

$$\gamma_k = \lambda_1 \gamma_{k-s} + \lambda_2 \gamma_{k-2s} + \dots + \lambda_P \gamma_{k-Ps}. \tag{5}$$

The associated autocorrelation coefficient has the form:

$$\rho_k = \lambda_1 \rho_{k-s} + \lambda_2 \rho_{k-2s} + \dots + \lambda_P \rho_{k-Ps}. \tag{6}$$

In view of (6), it is obvious that the model under consideration and the Gaussian AR model of order P and seasonal period ' s ' have similar autocorrelation structures. Hence, the identification of the model can be done using the autocorrelation function (ACF) and partial autocorrelation function (PACF). Theoretically speaking, the ACF of the INAR(P)_{*s*} model has nonzero values at the seasonal lags. The nonzero values at the seasonal lags tail off while the related partial ACF cuts off after lag P . When we have a stationary seasonal time series with sample ACF and sample partial ACF that have patterns that are akin to the theoretical ACF and sample partial ACF, we are expected to fit the INAR(P)_{*s*} model to the data. In situations where, the sample ACF and sample partial ACF do not look exactly like their theoretical counterparts, a variety of models can be fitted to the given data and the best model is determined using model selection criteria²⁴.

The parameters $\lambda_l, l = 1, 2, \dots, P$ of the model can be estimated using any of the Yule-Walker and conditional least squares approaches. For $k = s, 2s, \dots, Ps$, we derive P equations from (6). These equations are written in matrix form as

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\rho},$$

where

$$\Delta = \begin{pmatrix} 1 & \rho_s & \dots & \rho_{(P-1)s} \\ \rho_s & 1 & \dots & \rho_{(P-2)s} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{(P-1)s} & \rho_{(P-2)s} & \dots & 1 \end{pmatrix}, \quad \boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_P \end{pmatrix} \text{ and } \boldsymbol{\rho} = \begin{pmatrix} \rho_s \\ \rho_{2s} \\ \vdots \\ \rho_{Ps} \end{pmatrix}.$$

The Yule-Walker estimator $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ is given as

$$\hat{\boldsymbol{\lambda}} = \hat{\Delta}^{-1} \hat{\boldsymbol{\rho}}. \tag{7}$$

For example, for a second order monthly count time series, $s = 12$ and the estimates $\hat{\lambda}_1$ and $\hat{\lambda}_2$ based on sample autocorrelation coefficients at lags 12 and 24 are given as.

$$\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_1 \end{pmatrix} = \begin{pmatrix} 1 & \hat{\rho}_{12} \\ \hat{\rho}_{12} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_{12} \\ \hat{\rho}_{24} \end{pmatrix}. \tag{8}$$

In (7),

$$\hat{\Delta} = \begin{pmatrix} 1 & \hat{\rho}_s & \cdots & \hat{\rho}_{(P-1)s} \\ \hat{\rho}_s & 1 & \cdots & \hat{\rho}_{(P-2)s} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{(P-1)s} & \hat{\rho}_{(P-2)s} & \cdots & 1 \end{pmatrix}, \hat{\lambda} = \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \\ \vdots \\ \hat{\lambda}_P \end{pmatrix} \text{ and } \hat{\rho} = \begin{pmatrix} \hat{\rho}_s \\ \hat{\rho}_{2s} \\ \vdots \\ \hat{\rho}_{Ps} \end{pmatrix}.$$

Furthermore, the Yule–Walker estimate of μ_v is

$$\hat{\mu}_v = (1 - \hat{\lambda}_1 - \hat{\lambda}_2 - \cdots - \hat{\lambda}_P) \bar{W}. \tag{9}$$

The corresponding estimate of σ_v^2 is

$$\hat{\sigma}_v^2 = \frac{\sum_{t=Ps+1}^n (\hat{v}_t - \bar{v}_t)^2}{n - Ps}. \tag{10}$$

Notably,

$$\hat{v}_t = W_t - \hat{\lambda}_1 W_{t-s} - \hat{\lambda}_2 W_{t-2s} - \cdots - \hat{\lambda}_P W_{t-Ps}$$

and

$$\bar{v}_t = \frac{\sum_{t=Ps+1}^n \hat{v}_t}{n - Ps}.$$

Let $\Theta = (\mu_v, \lambda_1, \lambda_2, \dots, \lambda_P)^\top$, $g(\Theta, F_t) = \lambda_1 W_{t-s} + \lambda_2 W_{t-2s} + \cdots + \lambda_P W_{t-Ps} + \mu_v$ and $Q(\Theta) = \sum_{t=Ps+1}^n (W_t - g(\Theta, F_t))^2$.

The CLS estimator $\hat{\Theta}$ of Θ is obtained by minimizing $Q(\Theta)$. In this case, we solve the following equations simultaneously:

$$\begin{aligned} \frac{\partial Q(\Theta)}{\partial \mu_v} &= 0; \\ \frac{\partial Q(\Theta)}{\partial \lambda_l} &= 0, \quad l = 1, 2, \dots, P. \end{aligned}$$

Certain properties of CLS estimators are well-known (see, Klimko and Nelson²⁵). The minimum mean squared error (MMSE) predictor of W_{n+1} is

$$\begin{aligned} \hat{W}_n(1) &= E(W_{n+1} | \mathcal{F}_n) \\ &= \lambda_1 W_{n+1-s} + \lambda_2 W_{n+1-2s} + \cdots + \lambda_P W_{n+1-Ps} + \mu_v \end{aligned}$$

In general, the MMSE of W_{n+m} becomes

$$\begin{aligned} \hat{W}_n(m) &= \sum_{l=1}^P E(W_{n+m} | \mathcal{F}_n) \\ &= \sum_{l=1}^P E(\lambda_l * W_{n+m-ls} | \mathcal{F}_n) + \mu_v \\ &= \sum_{l=1}^P E[E(\lambda_l * W_{n+m-ls} | W_{n+m-ls}, \mathcal{F}_n) | \mathcal{F}_n] \\ &= \sum_{l=1}^P \lambda_l E(W_{n+m-ls} | \mathcal{F}_n) \\ &= \sum_{l=1}^P \lambda_l \hat{W}_n(m - ls) + \mu_v. \end{aligned} \tag{11}$$

We have carried out a simulation study to assess the performance of the Yule–Walker (YW) and conditional least squares (CLS) estimates. For the sake of brevity we consider INAR(2)_s model and its parameter estimation. We simulated 1000 series from the proposed model for various parameter combinations and for various sample sizes as given in Table 1. We have computed the mean estimates and their related MSEs based on 1000 simulations. In this study, we estimated two autoregressive parameters (λ_1 and λ_2) and the mean of

| n | $\hat{\lambda}_{1YW}$ | $\hat{\lambda}_{2YW}$ | $\hat{\mu}_{vYW}$ | $\hat{\lambda}_{1CLS}$ | $\hat{\lambda}_{2CLS}$ | $\hat{\mu}_{vCLS}$ |
|--|-----------------------|-----------------------|-------------------|------------------------|------------------------|--------------------|
| $\lambda_1 = 0.2, \lambda_2 = 0.3, \mu_v = 0.9$ | | | | | | |
| 300 | 0.1608 | 0.1193 | 1.1578 | 0.1986 | 0.2923 | 0.9120 |
| | 0.0055 | 0.0351 | 0.0788 | 0.0081 | 0.0101 | 0.0225 |
| 500 | 0.1924 | 0.1997 | 1.0531 | 0.1950 | 0.2935 | 0.9154 |
| | 0.0025 | 0.0119 | 0.0314 | 0.0031 | 0.0040 | 0.0117 |
| 1000 | 0.2035 | 0.2547 | 0.9654 | 0.1989 | 0.2963 | 0.9069 |
| | 0.0015 | 0.0033 | 0.0094 | 0.0014 | 0.0015 | 0.0054 |
| 1500 | 0.2026 | 0.2711 | 0.9432 | 0.1985 | 0.2974 | 0.9047 |
| | 0.0009 | 0.0017 | 0.0054 | 0.0009 | 0.0009 | 0.0035 |
| $\lambda_1 = 0.3, \lambda_2 = 0.4, \mu_v = 0.75$ | | | | | | |
| 300 | 0.2152 | 0.1824 | 1.0753 | 0.2900 | 0.3937 | 0.7690 |
| | 0.0116 | 0.0510 | 0.1295 | 0.0073 | 0.0103 | 0.0259 |
| 500 | 0.2718 | 0.2556 | 0.9927 | 0.2959 | 0.3885 | 0.7670 |
| | 0.0038 | 0.0233 | 0.0738 | 0.0032 | 0.0043 | 0.0145 |
| 1000 | 0.2978 | 0.3346 | 0.8809 | 0.2969 | 0.3938 | 0.7674 |
| | 0.0014 | 0.0055 | 0.0252 | 0.0014 | 0.0016 | 0.0085 |
| 1500 | 0.3022 | 0.3563 | 0.8388 | 0.2994 | 0.3962 | 0.7567 |
| | 0.0009 | 0.0029 | 0.0129 | 0.0009 | 0.0010 | 0.0050 |
| $\lambda_1 = 0.4, \lambda_2 = 0.3, \mu_v = 1.3333$ | | | | | | |
| 300 | 0.3944 | 0.1469 | 1.7433 | 0.3933 | 0.2954 | 1.3582 |
| | 0.0110 | 0.0631 | 0.4273 | 0.0090 | 0.0114 | 0.0574 |
| 500 | 0.3941 | 0.1455 | 1.7445 | 0.3980 | 0.2938 | 1.3547 |
| | 0.0027 | 0.0251 | 0.1878 | 0.0036 | 0.0042 | 0.0301 |
| 1000 | 0.4165 | 0.2295 | 1.5181 | 0.3990 | 0.2947 | 1.3560 |
| | 0.0017 | 0.0062 | 0.0473 | 0.0014 | 0.0015 | 0.0168 |
| 1500 | 0.4163 | 0.2525 | 1.4479 | 0.3990 | 0.2973 | 1.3502 |
| | 0.0012 | 0.0031 | 0.0232 | 0.0010 | 0.0010 | 0.0124 |
| $\lambda_1 = 0.6, \lambda_2 = 0.2, \mu_v = 1$ | | | | | | |
| 300 | 0.4066 | 0.0557 | 1.5797 | 0.5912 | 0.1984 | 1.0170 |
| | 0.0439 | 0.0244 | 0.4013 | 0.0084 | 0.0097 | 0.0483 |
| 500 | 0.5138 | 0.1025 | 1.4430 | 0.5915 | 0.1957 | 1.0293 |
| | 0.0113 | 0.0119 | 0.2321 | 0.0044 | 0.0043 | 0.0304 |
| 1000 | 0.5759 | 0.1510 | 1.2525 | 0.5982 | 0.1965 | 1.0136 |
| | 0.0024 | 0.0036 | 0.0806 | 0.0017 | 0.0018 | 0.0159 |
| 1500 | 0.5893 | 0.1679 | 1.1633 | 0.5965 | 0.1985 | 1.0156 |
| | 0.0012 | 0.0019 | 0.0395 | 0.0012 | 0.0011 | 0.0125 |

Table 1. Parameter estimates and their mean squared errors for INAR(2)_S model.

the innovation distribution (μ_v). For simulation from an innovation distribution, we use the parameter combinations: $\alpha = (4, 1, 2, 0.5)$ and $\delta = (-2, 0.5, -1, 0.7)$, which give the mean of the innovation distribution as $\mu_v = (0.9, 0.75, 1.3333, 1.00)$. The estimation of AR coefficient and all the parameters of innovation distribution have been studied in detail for INAR(1)_S model in “Construction of the proposed seasonal INAR(1) process and its properties”. From Table 1, it can be seen that the estimates perform well, and their mean squared errors (MSEs) decrease as the sample size increases. It also can be seen that the CLS estimation performs much better than the YW estimation in terms of the MSE. As the joint distribution of $\lambda_1 * W_{t-s}$ and $\lambda_2 * W_{t-2s}$ is not tractable, we can not find the conditional distribution and hence maximum likelihood estimation cannot be attempted, as it has been done for INAR(1)_S model in “Construction of the proposed seasonal INAR(1) process and its properties”.

Construction of the proposed seasonal INAR(1) process and its properties

In this section, the definition of the first-order seasonal INAR process based on zero-modified Poisson–Lindley (ZMPL) innovations¹² and NBT operation is given, with an extensive study. Using the NBT operator, we define the proposed seasonal INAR(1) process as follows.

Definition 3.1 Let $\{W_t\}$ be a discrete-time nonnegative integer-valued process. Then, the process is a seasonal INAR(1) process with zero-modified Poisson–Lindley innovations if

$$W_t = \lambda * W_{t-s} + v_t, \quad t \in \mathbb{N}_0, \quad (12)$$

where ‘s’ represents the seasonal period such that $s \in \mathbb{N}^+$, \mathbb{N}^+ is the set of positive integers, $\{W_t\}$ stands for a sequence of identically distributed nonnegative random variables, $\{v_t\}$ is an innovation sequence of i.i.d. ZMPL

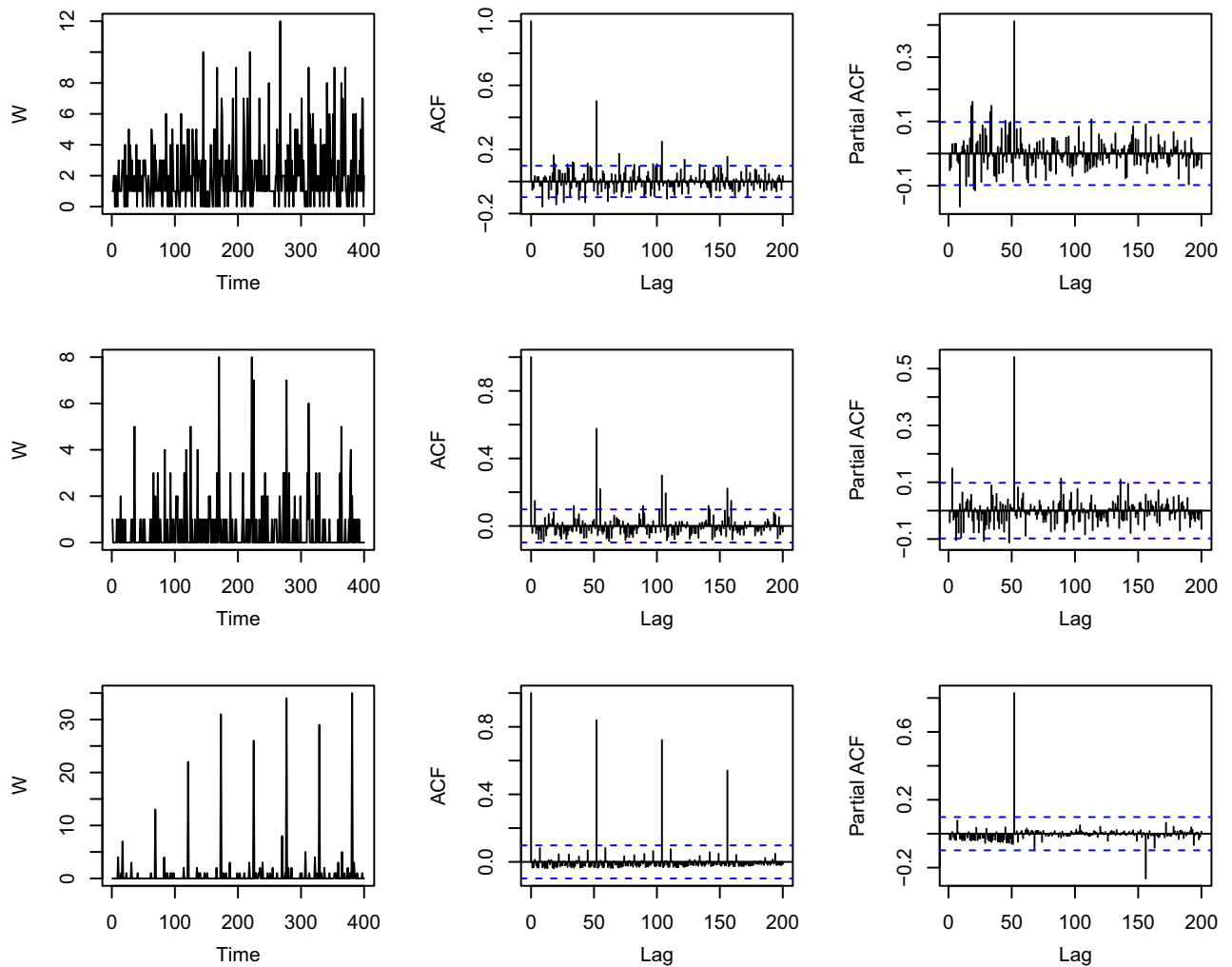


Figure 1. Sample path, sample ACF and sample PACF for simulated data: first row- $\lambda = 0.6, \alpha = 4, \delta = -2$; second row- $\lambda = 0.7, \alpha = 5, \delta = 0$; Third row- $\lambda = 0.8, \alpha = 8, \delta = 0.5$, seasonal period $S = 52$.

random variables not depending on the past values of $\{W_t\}$ and the thinnings of $\{W_t\}$ and the random variables $\{W_{t-s}\}, \{W_{t-2s}\}, \dots$, are independent.

In the sequel, the notation $\text{INAR}(1)_S\text{ZMPL}$ process is used to represent the model defined in Eq. (12). A random variable X follows the ZMPL distribution with parameters α and δ if its PMF is

$$P(X = x) = \begin{cases} \delta + (1 - \delta) \frac{\alpha^2(\alpha+2)}{(\alpha+1)^3}, & x = 0 \\ (1 - \delta) \frac{\alpha^2(x+\alpha+2)}{(\alpha+1)^{x+3}}, & x \in \mathbb{N}^+, \end{cases} \tag{13}$$

where $\alpha > 0$ and $-\frac{\alpha^2(\alpha+2)}{\alpha^2+3\alpha+1} \leq \delta \leq 1$. Thus, the unconditional mean and variance of the random variable $\{v_t\}$ are, respectively, given by

$$E(v_t) = \mu_v = (1 - \delta) \frac{\alpha + 2}{\alpha(\alpha + 1)} \tag{14}$$

and

$$\text{Var}(v_t) = \sigma_v^2 = (1 - \delta) \frac{\alpha^3 + 4\alpha^2 + 6\alpha + 2 + \delta(\alpha + 2)^2}{\alpha^2(\alpha + 1)^2}. \tag{15}$$

Following²⁶, the distribution of the random variable (RV) v_t is overdispersed if $\delta \in [0, 1)$, while it is underdispersed when $\alpha > \sqrt{2}$ and $\delta \in \left(\frac{-\alpha^2(\alpha+2)}{\alpha^2+3\alpha+1}, 0\right)$. The $\text{INAR}(1)_S\text{ZMPL}$ process comprises ‘s’ mutually independent $\text{INAR}(1)$ processes with ZMPL innovations, which are constructed using the NBT operator such that they have the autoregressive parameter λ and an innovation distribution. The simulated sample paths, sample ACFs and sample PACFs of the process for various parameter combinations are given in the Fig. 1. The seasonality in the

simulated data can be seen from the time series plots as well as ACF plots in the figure. Let $W_t^{(r)} = W_{ts+r}$, $t \in \mathbb{N}_0$, where \mathbb{N}_0 is a set of nonnegative integers. For $r = 0, 1, 2, \dots, s - 1$, the process $\{W_t^{(r)}\}$ satisfies the model:

$$W_t^{(r)} = \lambda * W_{t-1}^{(r)} + v_t^{(r)}, \quad t \in \mathbb{N}_0.$$

Here, the innovation sequences $\{v_t^{(r)}\}$ satisfy the equation $v_t^{(r)} = v_{ts+r}$. The independence of $\{W_t^{(r)}\}$ is the direct implication of the independence of $\{v_t^{(r)}\}$ and the counting processes that define the requisite thinning operators. Additionally, the process defined in Eq. (12) is an s -step Markov chain, implying that

$$P(W_t = w_t | W_{t-1} = w_{t-1}, \dots, W_0 = w_0) = P(W_t = w_t | W_{t-s} = w_{t-s}), \quad w_0, w_1, \dots, w_{s-1} \in \mathbb{N}_0.$$

The conditional mean of the process is

$$\begin{aligned} E(W_t | W_{t-s}) &= E(\lambda * W_{t-s} + v_t | W_{t-s}) \\ &= \lambda W_{t-s} + \mu_v. \end{aligned} \tag{16}$$

Hence, the unconditional expectation of the process is

$$\begin{aligned} E(W_t) &= (E(W_t | W_{t-s})) \\ &= \lambda E(W_{t-s}) + \mu_v. \end{aligned} \tag{17}$$

Assuming stationarity and using Eq. (17), the unconditional mean is found to be

$$E(W_t) = \frac{\mu_v}{1 - \lambda} = \mu_W.$$

The conditional variance satisfies the equation:

$$\begin{aligned} \text{Var}(W_t | W_{t-s}) &= \text{Var}(\lambda * W_{t-s} + v_t | W_{t-s}) \\ &= \text{Var}(\lambda * W_{t-s} | W_{t-s}) + \sigma_v^2 \\ &= \lambda^2 E(W_{t-s}^2 | W_{t-s}) + \lambda(\lambda + 1)E(W_{t-s} | W_{t-s}) - \lambda^2 (E(W_{t-s} | W_{t-s}))^2 \\ &\quad + \sigma_v^2 \\ &= \lambda^2 \text{Var}(W_{t-s} | W_{t-s}) + \lambda(\lambda + 1)E(W_{t-s} | W_{t-s}) + \sigma_v^2 \\ &= \lambda(\lambda + 1)W_{t-s} + \sigma_v^2. \end{aligned} \tag{18}$$

The unconditional variance is obtained as

$$\begin{aligned} \text{Var}(W_t) &= E(\text{Var}(W_t | W_{t-s})) + \text{Var}(E(W_t | W_{t-s})) \\ &= \lambda(\lambda + 1)E(W_{t-s}) + \sigma_v^2 + \lambda^2 \text{Var}(W_{t-s}). \end{aligned} \tag{19}$$

Hence, it follows that

$$\text{Var}(W_t) = \frac{\lambda(\lambda + 1)\mu_v + (1 - \lambda)\sigma_v^2}{(1 - \lambda)(1 - \lambda^2)} = \sigma_W^2.$$

Suppose that $W_{(k+h)s+j}$ and W_{ks+i} satisfy the process in Eq. (12), $h \in \mathbb{N}^+$ and $k \in \mathbb{N}_0$, $i, j = 1, 2, \dots, s$. In order to gain insight into conditional moments based on the seasonal period as well as the autocorrelation function of the process, we proceed to give the following propositions:

Proposition 3.1 *The conditional expectation of $W_{(k+h)s+j}$ given W_{ks+i} is*

$$E(W_{(k+h)s+j} | W_{ks+i}) = \begin{cases} \frac{\mu_v}{1 - \lambda}, & \text{if } i \neq j \\ \lambda^h W_{ks+i} + (1 - \lambda^h)\mu_v, & \text{if } i = j \end{cases}$$

and $\lim_{h \rightarrow \infty} E(W_{(k+h)s+j} | W_{ks+i}) = \mu_W$.

Proof Let $W_{(k+h)s+j} = W_{k+h}^{(j)}$ and $W_{ks+i} = W_k^{(i)}$. If $i \neq j$, the conditional expectation of $W_{k+h}^{(j)}$ given $W_k^{(i)}$ is equal to $\frac{\mu_v}{1 - \lambda}$, which is the unconditional expectation of W_t . If $i = j$, we have

$$\begin{aligned}
 E(W_{(k+h)s+j} | W_{ks+i}) &= E(W_{k+h}^{(j)} | W_k^{(j)}) \\
 &= \lambda E(W_{k+h-1}^{(j)} | W_k^{(j)}) + E(v_{k+h}^{(j)} | W_k^{(j)}) \\
 &= \lambda E(W_{k+h-1}^{(j)} | W_k^{(j)}) + \mu_v \\
 &= \lambda^2 E(W_{k+h-2}^{(j)} | W_k^{(j)}) + (\lambda + 1)\mu_v \\
 &\dots \\
 &= \lambda^h E(W_k^{(j)} | W_k^{(j)}) + (\lambda^{h-1} + \lambda^{h-2} + \dots + \lambda + 1)\mu_v \\
 &= \lambda^h W_k^{(j)} + \frac{1 - \lambda^h}{1 - \lambda} \mu_v \\
 &= \lambda^h W_{ks+i} + (1 - \lambda^h)\mu_W.
 \end{aligned}$$

Consequently,

$$\begin{aligned}
 \lim_{h \rightarrow \infty} E(W_{(k+h)s+j} | W_{ks+i}) &= \lim_{h \rightarrow \infty} E\left(\lambda^h W_k^{(j)} + \frac{1 - \lambda^h}{1 - \lambda} \mu_v\right) \\
 &= \frac{\mu_v}{1 - \lambda} \\
 &= \mu_W.
 \end{aligned}$$

The proof comes to an end here. □

Proposition 3.2 *The conditional variance of $W_{(k+h)s+j}$ given W_{ks+i} is*

$$\text{Var}(W_{(k+h)s+j} | W_{ks+i}) = \begin{cases} \frac{\lambda(\lambda+1)\mu_v + (1-\lambda)\sigma_v^2}{(1-\lambda)(1-\lambda^2)}, & \text{if } i \neq j \\ \left(\lambda + 1\right) \left(\frac{\lambda^h(1-\lambda^h)}{1-\lambda}\right) W_k^{(j)} + \frac{\lambda(\lambda+1)}{1-\lambda} \left(\frac{1-\lambda^{2(h-1)}}{1-\lambda^2} - \lambda^{h-1} \left(\frac{1-\lambda^{(h-1)}}{1-\lambda}\right)\right) \mu_v + \left(\frac{1-\lambda^{2h}}{1-\lambda^2}\right) \sigma_v^2, & \text{if } i = j \end{cases}$$

and $\lim_{h \rightarrow \infty} \text{Var}(W_{(k+h)s+j} | W_{ks+i}) = \sigma_W^2$.

Proof If $i \neq j$, $W_{k+h}^{(j)}$ and $W_k^{(i)}$ are independent. Thus, the conditional variance of $W_{(k+h)s+j}$ given W_{ks+i} is $\frac{\lambda(\lambda+1)\mu_v + (1-\lambda)\sigma_v^2}{(1-\lambda)(1-\lambda^2)}$, the unconditional variance of W_t .

If $i = j$, then

$$\begin{aligned}
 \text{Var}(W_{(k+h)s+j} | W_{ks+i}) &= \text{Var}(W_{k+h}^{(j)} | W_k^{(j)}) \\
 &= \text{Var}(\lambda * W_{k+h-1}^{(j)} + v_{k+h}^{(j)} | W_k^{(j)}) \\
 &= \text{Var}(\lambda * W_{k+h-1}^{(j)} | W_k^{(j)}) + \text{Var}(v_{k+h}^{(j)} | W_k^{(j)}) \\
 &= \lambda^2 \text{Var}(W_{k+h-1}^{(j)} | W_k^{(j)}) + \lambda(\lambda + 1)E(v_{k+h-1}^{(j)} | W_k^{(j)}) + \sigma_v^2 \\
 &= \lambda^2 \text{Var}(\lambda * W_{k+h-2}^{(j)} + v_{k+h-1}^{(j)} | W_k^{(j)}) \\
 &\quad + \lambda(\lambda + 1)E(W_{k+h-1}^{(j)} | W_k^{(j)}) + \sigma_v^2 \\
 &= \lambda^4 E((W_{k+h-2}^{(j)})^2 | W_k^{(j)}) - \lambda^4 (E(W_{k+h-2}^{(j)} | W_k^{(j)}))^2 \\
 &\quad + \lambda^3(\lambda + 1)E(W_{k+h-2}^{(j)} | W_k^{(j)}) \\
 &\quad + \lambda(\lambda + 1)E(W_{k+h-1}^{(j)} | W_k^{(j)}) + \lambda^2 \sigma_v^2 + \sigma_v^2 \\
 &= \lambda^4 \text{Var}(W_{k+h-2}^{(j)} | W_k^{(j)}) + \lambda^3(\lambda + 1)E(W_{k+h-2}^{(j)} | W_k^{(j)}) \\
 &\quad + \lambda(\lambda + 1)E(W_{k+h-1}^{(j)} | W_k^{(j)}) + \lambda^2 \sigma_v^2 + \sigma_v^2 \\
 &= \dots \\
 &= \lambda^{2h} \text{Var}(W_k^{(j)} | W_k^{(j)}) + \lambda(\lambda + 1)M \\
 &\quad + (\lambda^{2(h-1)} + \lambda^{2(h-2)} + \dots + \lambda^2 + 1)\sigma_v^2,
 \end{aligned}$$

where

$$M = \lambda^{2(h-1)} E\left(W_k^{(j)} | W_k^{(j)}\right) + \lambda^{2(h-2)} E\left(W_{k+1}^{(j)} | W_k^{(j)}\right) + \dots + \lambda^2 E\left(W_{k+h-2}^{(j)} | W_k^{(j)}\right) + E\left(W_{k+h-1}^{(j)} | W_k^{(j)}\right).$$

Notably, $\text{Var}\left(W_k^{(j)} | W_k^{(j)}\right) = 0$. Also, after some algebra with the earlier discussions, we can find that,

$$M = \frac{\lambda^{(h-1)}(1 - \lambda^h)}{1 - \lambda} W_k^{(j)} + \frac{1}{1 - \lambda} \left(\frac{1 - \lambda^{2(h-1)}}{1 - \lambda^2} - \frac{\lambda^{(h-1)}(1 - \lambda^{(h-1)})}{1 - \lambda} \right) \mu_v.$$

In the light of the above, we have

$$\begin{aligned} \text{Var}\left(W_{(k+h)s+j} | W_{ks+i}\right) &= (1 + \lambda) \frac{\lambda^h(1 - \lambda^h)}{1 - \lambda} W_k^{(j)} \\ &+ \frac{\lambda(1 + \lambda)}{1 - \lambda} \left(\frac{1 - \lambda^{2(h-1)}}{1 - \lambda^2} - \frac{\lambda^{(h-1)}(1 - \lambda^{(h-1)})}{1 - \lambda} \right) \mu_v \\ &+ \left(\frac{1 - \lambda^{2h}}{1 - \lambda^2} \right) \sigma_v^2. \end{aligned}$$

Hence,

$$\lim_{h \rightarrow \infty} \text{Var}\left(W_{(k+h)s+j} | W_{ks+i}\right) = \frac{\lambda(\lambda + 1)\mu_v + (1 - \lambda)\sigma_v^2}{(1 - \lambda)(1 - \lambda^2)} = \text{Var}(W_t).$$

□

Proposition 3.3 *The covariance of $W_{(k+h)s+j}$ and W_{ks+i} has the form*

$$\text{Covar}\left(W_{(k+h)s+j}, W_{ks+i}\right) = \begin{cases} 0, & \text{if } i \neq j \\ \lambda^h \frac{\lambda(\lambda+1)\mu_v + (1-\lambda)\sigma_v^2}{(1-\lambda)(1-\lambda^2)}, & \text{if } i = j. \end{cases}$$

Proof Once $i \neq j$, the variables $W_{(k+h)s+j}$ and W_{ks+i} are uncorrelated. In this case, $\text{Covar}\left(W_{(k+h)s+j}, W_{ks+i}\right) = 0$.

$$\begin{aligned} \text{Covar}\left(W_{(k+h)s+j}, W_{ks+i}\right) &= \text{Covar}\left(W_{k+h}^{(j)}, W_k^{(j)}\right) \\ &= E\left(\lambda * W_{k+h-1}^{(j)} + v_{k+h}^{(j)} W_k^{(j)}\right) \\ &\quad - E\left(\lambda * W_{k+h-1}^{(j)} + v_{k+h}^{(j)}\right) E\left(W_k^{(j)}\right) \\ &= E\left(\left(\sum_{q=1}^{W_{k+h-1}^{(j)}} Z_q\right) W_k^{(j)}\right) + E\left(v_{k+h}^{(j)}\right) E\left(W_k^{(j)}\right) \\ &\quad - \lambda E\left(W_{k+h-1}^{(j)}\right) E\left(W_k^{(j)}\right) - E\left(v_{k+h}^{(j)}\right) E\left(W_k^{(j)}\right) \\ \text{On the other side,} \quad &= E\left[E\left(\sum_{q=1}^{W_{k+h-1}^{(j)}} Z_q | W_{k+h-1}^{(j)}\right) E\left(W_k^{(j)} | W_{k+h-1}^{(j)}\right)\right] \\ &\quad - \lambda E\left(W_{k+h-1}^{(j)}\right) E\left(W_k^{(j)}\right) \\ &= E\left[W_{k+h-1}^{(j)} W_k^{(j)} E(Z_q)\right] - \lambda E\left(W_{k+h-1}^{(j)}\right) E\left(W_k^{(j)}\right) \\ &= \lambda \text{Cov}\left(W_{k+h-1}^{(j)}, W_k^{(j)}\right) \\ &\dots \\ &= \lambda^h \text{Cov}\left(W_k^{(j)}, W_k^{(j)}\right) \\ &= \lambda^h \text{Var}\left(W_k^{(j)}\right) = \lambda^h \frac{\lambda(\lambda + 1)\mu_v + (1 - \lambda)\sigma_v^2}{(1 - \lambda)(1 - \lambda^2)}. \end{aligned}$$

Dividing the autocovariance function by $\text{Var}(W_t)$, the autocorrelation function (ACF) is obtained as

$$\rho(h) = \begin{cases} \lambda^{\frac{h}{s}}, & \text{if } h \text{ is a multiple of } s \\ 0, & \text{elsewhere.} \end{cases}$$

Here, we have an exponentially decaying autocorrelation function. □

Techniques for estimating model parameters

Three popular and widely used methods of point estimation for the parameters of the INAR(1)_sZMPL are adopted in this section. These are, the Yule–Walker, conditional least squares and conditional maximum likelihood methods. In each of the methods, we assume a realization of the seasonal count time series.

Yule–Walker approach

To estimate the parameters λ, δ and α of the INAR(1)_sZMPL process, we form three equations by equating $\hat{\rho}(s)$, $E(W_t)$ and $\text{Var}(W_t)$ to $\frac{\hat{\gamma}(s)}{\hat{\gamma}(0)}$, \bar{W} and $\hat{\gamma}(0)$, respectively.

Notably, $\bar{W} = \frac{\sum_{t=1}^n W_t}{n}$, $\hat{\gamma}(s) = \frac{1}{n} \sum_{t=s+1}^n (W_t - \bar{W})(W_{t-s} - \bar{W})$, and n are, respectively, the sample mean, sample autocorrelation function and length of the time series. If $\hat{\lambda}_{YW}, \hat{\delta}_{YW}$ and $\hat{\alpha}_{YW}$ are the Yule–Walker (YW) estimators of λ, δ and α in that order, then the following equations are needed:

$$\hat{\lambda}_{YW} = \frac{\sum_{t=s+1}^n (W_t - \bar{W})(W_{t-s} - \bar{W})}{\sum_{t=1}^n (W_t - \bar{W})^2}, \tag{20}$$

$$(1 - \hat{\delta}_{YW}) \frac{\hat{\alpha}_{YW} + 2}{\hat{\alpha}_{YW}(\hat{\alpha}_{YW} + 1)(1 - \hat{\lambda}_{YW})} = \bar{W}, \tag{21}$$

and
$$\frac{A}{(1 - \hat{\lambda}_{YW}^2)\hat{\alpha}_{YW}(\hat{\alpha}_{YW} + 1)(\hat{\alpha}_{YW} + 2)} = \hat{\gamma}(0), \tag{22}$$

where

$$\begin{aligned} A = & \hat{\lambda}_{YW}(\hat{\lambda}_{YW} + 1)\bar{W}\hat{\alpha}_{YW}(\hat{\alpha}_{YW} + 1)(\hat{\alpha}_{YW} + 2) \\ & + (1 - \hat{\lambda}_{YW})\bar{W}\left((1 - (1 - \hat{\lambda}_{YW})\bar{W})\hat{\alpha}_{YW}^3 + (5 - 3(1 - \hat{\lambda}_{YW})\bar{W})\hat{\alpha}_{YW}^2\right) \\ & + (1 - \hat{\lambda}_{YW})\bar{W}\left((10 - 2(1 - \hat{\lambda}_{YW})\bar{W})\hat{\alpha}_{YW} + 6\right). \end{aligned}$$

From Eq. (20), we obtain

$$(1 - \hat{\delta}_{YW}) = \frac{\hat{\alpha}_{YW}(\hat{\alpha}_{YW} + 1)(1 - \hat{\lambda}_{YW})}{\hat{\alpha}_{YW} + 2} \bar{W}. \tag{23}$$

Using Eqs. (21) and (22), the following cubic equation is obtained:

$$c_1\hat{\alpha}_{YW}^3 + c_2\hat{\alpha}_{YW}^2 + c_3\hat{\alpha}_{YW} + c_4 = 0, \tag{24}$$

where

$$\begin{aligned} c_1 = & \left[\hat{\lambda}_{YW}(\hat{\lambda}_{YW} + 1) + (1 - \hat{\lambda}_{YW})(1 - (1 - \hat{\lambda}_{YW})\bar{W}) \right] \bar{W} - (1 - \hat{\lambda}_{YW}^2)\hat{\gamma}(0), \\ c_2 = & \left[3\hat{\lambda}_{YW}(\hat{\lambda}_{YW} + 1) + (1 - \hat{\lambda}_{YW})(5 - 3(1 - \hat{\lambda}_{YW})\bar{W}) \right] \bar{W} - 3(1 - \hat{\lambda}_{YW}^2)\hat{\gamma}(0), \\ c_3 = & \left[2\hat{\lambda}_{YW}(\hat{\lambda}_{YW} + 1) + (1 - \hat{\lambda}_{YW})(10 - 2(1 - \hat{\lambda}_{YW})\bar{W}) \right] \bar{W} - 2(1 - \hat{\lambda}_{YW}^2)\hat{\gamma}(0), \end{aligned}$$

and
$$c_4 = 6(1 - \hat{\lambda}_{YW})\bar{W}.$$

R packages for solving polynomial equations, particularly the cubic equation, abound. In “Yule–Walker approach” of this study, polyroot R function is used to obtain the roots of Eq. (23) after the coefficients have been calculated from the given time series. Though the equation has three roots, only the root that gives an acceptable value of $\hat{\delta}_{YW}$ will be used for further computations.

Approach to conditional least squares estimation

Consider the process in Eq. (12). Let $\hat{\xi}_{CLS} = (\hat{\lambda}_{CLS}, \hat{\delta}_{CLS}, \hat{\alpha}_{CLS})^T$ be the vector of the conditional least squares (CLS) estimators of $\xi_{CLS} = (\lambda_{CLS}, \delta_{CLS}, \alpha_{CLS})^T$. Then

$$\hat{\xi}_{CLS} = \underset{\xi}{\operatorname{argmin}}(C_n(\xi)),$$

where $C_n(\xi) = \sum_{t=s+1}^n \left(W_t - \lambda W_{t-s} - (1 - \delta) \frac{\alpha + 2}{\alpha(\alpha + 1)} \right)^2$.

Here, it is necessary to obtain and equate each of $\frac{\partial C_n(\xi)}{\partial \lambda} \Big|_{\lambda=\hat{\lambda}_{CLS}, \delta=\hat{\delta}_{CLS}, \alpha=\hat{\alpha}_{CLS}}$, $\frac{\partial C_n(\xi)}{\partial \delta} \Big|_{\lambda=\hat{\lambda}_{CLS}, \delta=\hat{\delta}_{CLS}, \alpha=\hat{\alpha}_{CLS}}$ and $\frac{\partial C_n(\xi)}{\partial \alpha} \Big|_{\lambda=\hat{\lambda}_{CLS}, \delta=\hat{\delta}_{CLS}, \alpha=\hat{\alpha}_{CLS}}$ to zero, leading to the following equations, respectively:

$$\sum_{t=s+1}^n W_t W_{t-s} - \sum_{t=s+1}^n W_{t-s}^2 - (1 - \hat{\delta}_{CLS}) \frac{\hat{\alpha}_{CLS} + 2}{\hat{\alpha}_{CLS}(\hat{\alpha}_{CLS} + 1)} \sum_{t=s+1}^n W_{t-s} = 0; \tag{25}$$

$$\sum_{t=s+1}^n W_t - \hat{\lambda}_{CLS} \sum_{t=s+1}^n W_{t-s} - (n - s)(1 - \hat{\delta}_{CLS}) \frac{\hat{\alpha}_{CLS} + 2}{\hat{\alpha}_{CLS}(\hat{\alpha}_{CLS} + 1)} = 0; \tag{26}$$

$$\sum_{t=s+1}^n W_t - \hat{\lambda}_{CLS} \sum_{t=s+1}^n W_{t-s} - (n - s)(1 - \hat{\delta}_{CLS}) \frac{\hat{\alpha}_{CLS} + 2}{\hat{\alpha}_{CLS}(\hat{\alpha}_{CLS} + 1)} = 0. \tag{27}$$

Equations (26) and (27) are identical. Hence, we solve for $\hat{\lambda}_{CLS}$ using Eqs. (25) and (26). The concerned estimator has the form:

$$\hat{\lambda}_{CLS} = \frac{(n - s) \sum_{t=s+1}^n W_t W_{t-s} - \sum_{t=s+1}^n W_t \sum_{t=s+1}^n W_{t-s}}{(n - s) \sum_{t=s+1}^n W_{t-s}^2 - \left(\sum_{t=s+1}^n W_{t-s} \right)^2}. \tag{28}$$

Suppose $C_n(\delta, \alpha)$ is a function obtained when we substitute $\hat{\lambda}_{CLS}$ for λ in $C_n(\xi)$. We minimize $C_n(\delta, \alpha)$ in order to find $\hat{\delta}_{CLS}$ and $\hat{\alpha}_{CLS}$. The minimization process is done using a suitable numerical method. To summarize, no closed form expression can be found for any of $\hat{\delta}_{CLS}$ and $\hat{\alpha}_{CLS}$.

Conditional maximum likelihood estimation

In view of the s-step Markovian property of the seasonal INAR(1)_sZMPL process, we define the transition probabilities as

$$\begin{aligned} P_{xy} &= P(X_t = y | X_{t-s} = x) \\ &= P(\lambda * X_{t-s} + v_t = y | X_{t-s} = x) \\ &= \sum_{r=0}^y P(\lambda * x = r) P(v_t = y - r). \end{aligned}$$

Using this we can write the detailed transition probabilities as

$$P_{xy} = \begin{cases} \delta + (1 - \delta) \frac{\alpha^2(\alpha + 2)}{(\alpha + 1)^3}, & \text{if } x = 0, y = 0, \\ (1 - \delta) \frac{\alpha^2(y + \alpha + 2)}{(\alpha + 1)^{y+3}}, & \text{if } x = 0, y \geq 1, \\ \frac{1}{(1 + \lambda)^x} \left(\delta + (1 - \delta) \frac{\alpha^2(\alpha + 2)}{(\alpha + 1)^3} \right), & \text{if } x \geq 1, y = 0, \\ \binom{x + y - 1}{y} \frac{\lambda^y}{(1 + \lambda)^{x+y}} \left(\delta + (1 - \delta) \frac{\alpha^2(\alpha + 2)}{(\alpha + 1)^3} \right) \\ + \sum_{r=0}^{y-1} \binom{x + r - 1}{r} \frac{\lambda^r}{(1 + \lambda)^{x+r}} \\ \times \left(\delta + (1 - \delta) \frac{\alpha^2(y - r + \alpha + 2)}{(\alpha + 1)^{(y-r+3)}} \right) & \text{if } x \geq 1, y \geq 1. \end{cases} \tag{29}$$

Let $\hat{\xi}_{CML}$ be the conditional maximum likelihood(CML) estimator of ξ . To find $\hat{\xi}_{CML}$, maximize the conditional log-likelihood function below:

$$l(\xi) = \sum_{t=s+1}^n \log P(X_t = x_t | X_{t-s} = x_{t-s}).$$

Obviously, $\hat{\xi}_{CML}$ has no closed form. Furthermore, the required estimators can only be found through a numerical technique.

Simulation study for the INAR(1)_sZMPL

In this section, we have carried out a simulation study to assess the performance of the proposed estimation methods for INAR(1)_sZMPL process, paying attention to the cases of zero-inflation, overdispersion, zero-deflation and underdispersion based on the INAR(1)_sZMPL process. We have simulated 1000 series of the sample sizes 100, 300, 500 and 1000 for various parameter combinations and for seasonal period $s = 52$. We have computed the mean and the mean squared errors (MSEs) of the estimates over all the 1000 simulations. All the quantities are given in the Table 2, the first row for each parameter and sample size (n) combination represents the mean estimate and second row represents the MSE. From the table, it can be seen that the MSE is decreasing with the increase in sample size, indicating the consistency of the estimates. The numerical optimization of the likelihood function to obtain CML estimates is done using the function ‘*constrOptim*’ in R software. In our setup, constraints are imposed on all the parameters. In particular, the upper limit of δ is 1 while its lower limit is a negative quantity that is a function of α . These constraints are considered while numerical optimization using ‘*constrOptim*’. In all, CML estimates correspond to minimum MSEs compared to the YW and CLS estimates.

Model-based forecasting

Time series models are usually constructed with a view to forecasting future values of time series data. In the case of INAR(1)_sZMPL process, we handle the problem of forecasting a future observation W_{n+h} , $h \in \mathbb{N}$ using the information \mathcal{F}_n available upto time n . From Eq. (12), we deduce using induction and properties of the NBT that

$$W_{n+h} \stackrel{d}{=} \lambda^q * W_{n+h-qs} + \sum_{j=0}^{q-1} \lambda^j * v_{n+h-js}, \quad h \in \mathbb{N}^+, \quad (30)$$

where $\stackrel{d}{=}$ implies equality in distribution, $q = \lceil \frac{h}{s} \rceil$ is the integer part of $\frac{h}{s}$. That is $\lceil y \rceil = \min \{n \in \mathbb{N}^+ | y \leq n\}$. We round off this section with the following proposition.

Proposition 3.4 Consider the INAR(1)_sZMPL process in Eq. (12). Then

(a) The h -step conditional expectation is

$$E(W_{n+h} | \mathcal{F}_n) = \lambda^q (W_{n+h-qs} - E(W_t)) + \mu_v.$$

(b) The h -step conditional variance is

$$\begin{aligned} \text{Var}(W_{n+h} | \mathcal{F}_n) &= (1 + \lambda) \frac{\lambda^q (1 - \lambda^q)}{1 - \lambda} W_{n+h-qs} \\ &+ \frac{\lambda(1 + \lambda)}{1 - \lambda} \left(\frac{1 - \lambda^{2(q-1)}}{1 - \lambda^2} - \frac{\lambda^{(q-1)}(1 - \lambda^{(q-1)})}{1 - \lambda} \right) \mu_v \\ &+ \left(\frac{1 - \lambda^{2q}}{1 - \lambda^2} \right) \sigma_v^2. \end{aligned}$$

(c) $\lim_{q \rightarrow \infty} E(W_{n+h} | \mathcal{F}_n) = E(W_t)$.

(d) $\lim_{q \rightarrow \infty} \text{Var}(W_{n+h} | \mathcal{F}_n) = \text{Var}(W_t)$.

It is not difficult to prove Proposition 3.4 using knowledge of the proofs of Propositions 3.1 to 3.2 and Eq. (30). Therefore, we omit the proof of Proposition 3.4. The h -step ahead forecast can be written as

$$\hat{W}_{n+h} = \hat{\lambda}^q \left(\hat{W}_{n+h-qs} - \hat{\mu}_W \right) + \hat{\mu}_v.$$

Model comparison with seasonal ARIMA

In this section, we have compared the forecast performance of the proposed model and its counterpart SARIMA(0, 0, 0)(1, 0, 0)_s model. Here, we have simulated 1000 series each of size 500 from the INAR(1)_sZMPL model with various parameter combinations given in Table 3 and the seasonal period $s = 52$. For each series, the first 495 observations are used for the estimation of the parameters and last five observations are used to check the forecast accuracy. We have used three forecast accuracy measures to assess the forecast performance of the model. These measures include prediction root mean squared error (PRMSE), prediction mean absolute error (PMAE) and percentage of true prediction (PTP). Notably,

$$\text{PRMSE}(k) = \sqrt{\frac{1}{m} \sum_{i=1}^m \left(X_{(t+i)} - \hat{X}_{(t+i)}^{(k)} \right)^2},$$

where $\hat{X}_{(t+i)}^{(k)} = \widehat{\text{mean}}(X_{t+i} | X_{t-k+i})$ is the k -step ahead conditional mean of the fitted process,

| n | $\hat{\lambda}_{YW}$ | $\hat{\alpha}_{YW}$ | $\hat{\delta}_{YW}$ | $\hat{\lambda}_{CLS}$ | $\hat{\alpha}_{CLS}$ | $\hat{\delta}_{CLS}$ | $\hat{\lambda}_{CML}$ | $\hat{\alpha}_{CML}$ | $\hat{\delta}_{CML}$ |
|--|----------------------|---------------------|---------------------|-----------------------|----------------------|----------------------|-----------------------|----------------------|----------------------|
| $\lambda = 0.3, \alpha = 0.5, \delta = 0.7$ | | | | | | | | | |
| 100 | 0.2739 | 0.6347 | 0.6307 | 0.2747 | 0.5749 | 0.6324 | 0.2952 | 0.5234 | 0.6919 |
| | 0.0157 | 0.0799 | 0.0205 | 0.0133 | 0.0159 | 0.0259 | 0.0081 | 0.0164 | 0.0054 |
| 300 | 0.2954 | 0.5363 | 0.6827 | 0.2947 | 0.5504 | 0.6641 | 0.2963 | 0.5073 | 0.6956 |
| | 0.0046 | 0.0083 | 0.0042 | 0.0046 | 0.0072 | 0.0077 | 0.0022 | 0.0036 | 0.0014 |
| 500 | 0.2953 | 0.5229 | 0.6868 | 0.2954 | 0.5396 | 0.6721 | 0.2993 | 0.5088 | 0.6963 |
| | 0.0029 | 0.0043 | 0.0026 | 0.0026 | 0.0046 | 0.0038 | 0.0013 | 0.0022 | 0.0009 |
| 1000 | 0.2966 | 0.5142 | 0.6919 | 0.2970 | 0.5303 | 0.6785 | 0.3005 | 0.5025 | 0.6998 |
| | 0.0013 | 0.0021 | 0.0012 | 0.0012 | 0.0029 | 0.0019 | 0.0006 | 0.0011 | 0.0004 |
| $\lambda = 0.5, \alpha = 1, \delta = 0.5$ | | | | | | | | | |
| 100 | 0.4319 | 1.1772 | 0.3866 | 0.4656 | 1.0993 | 0.4106 | 0.4838 | 1.0845 | 0.4713 |
| | 0.0198 | 0.4565 | 0.0848 | 0.0177 | 0.0358 | 0.0545 | 0.0138 | 0.0986 | 0.0190 |
| 300 | 0.4756 | 1.0852 | 0.4525 | 0.4864 | 1.0718 | 0.4472 | 0.4912 | 1.0121 | 0.4920 |
| | 0.0054 | 0.0560 | 0.0168 | 0.0050 | 0.0160 | 0.0147 | 0.0040 | 0.0209 | 0.0054 |
| 500 | 0.4865 | 1.0564 | 0.4689 | 0.4905 | 1.0668 | 0.4505 | 0.4954 | 1.0179 | 0.4905 |
| | 0.0032 | 0.0282 | 0.0097 | 0.0030 | 0.0114 | 0.0099 | 0.0022 | 0.0115 | 0.0028 |
| 1000 | 0.4928 | 1.0226 | 0.4867 | 0.4930 | 1.0622 | 0.4576 | 0.4988 | 1.0087 | 0.4973 |
| | 0.0015 | 0.0116 | 0.0043 | 0.0014 | 0.0089 | 0.0058 | 0.0011 | 0.0062 | 0.0014 |
| $\lambda = 0.6, \alpha = 1.5, \delta = -0.5$ | | | | | | | | | |
| 100 | 0.5188 | 1.4169 | -0.4893 | 0.5621 | 1.4184 | -0.4468 | 0.5688 | 1.5475 | -0.5601 |
| | 0.0172 | 0.9482 | 0.5433 | 0.0150 | 0.0833 | 0.0372 | 0.0124 | 0.1645 | 0.0796 |
| 300 | 0.5784 | 1.5685 | -0.5892 | 0.5838 | 1.4341 | -0.4443 | 0.5919 | 1.5326 | -0.5312 |
| | 0.0043 | 0.4783 | 0.2972 | 0.0039 | 0.0382 | 0.0268 | 0.0028 | 0.0712 | 0.0354 |
| 500 | 0.5851 | 1.5689 | -0.5748 | 0.5917 | 1.4460 | -0.4464 | 0.5943 | 1.5242 | -0.5266 |
| | 0.0026 | 0.2143 | 0.1535 | 0.0023 | 0.0261 | 0.0188 | 0.0017 | 0.0478 | 0.0239 |
| 1000 | 0.5936 | 1.5606 | -0.5558 | 0.5961 | 1.4491 | -0.4494 | 0.5971 | 1.5090 | -0.5107 |
| | 0.0011 | 0.0649 | 0.0558 | 0.0011 | 0.0130 | 0.0130 | 0.0010 | 0.0225 | 0.0109 |
| $\lambda = 0.7, \alpha = 2, \delta = -1$ | | | | | | | | | |
| 100 | 0.5804 | 1.4330 | -0.5686 | 0.6590 | 1.8734 | -0.9271 | 0.6653 | 1.9439 | -0.9821 |
| | 0.0240 | 1.8765 | 1.2004 | 0.0153 | 0.1675 | 0.0744 | 0.0122 | 0.3010 | 0.1417 |
| 300 | 0.6713 | 1.7131 | -0.8222 | 0.6858 | 1.8965 | -0.9120 | 0.6891 | 2.0113 | -1.0204 |
| | 0.0045 | 1.7318 | 1.0939 | 0.0038 | 0.0983 | 0.0559 | 0.0025 | 0.1510 | 0.0841 |
| 500 | 0.6848 | 1.8997 | -0.9649 | 0.6906 | 1.9002 | -0.9101 | 0.6936 | 2.0205 | -1.0274 |
| | 0.0022 | 1.2074 | 0.8342 | 0.0022 | 0.0732 | 0.0491 | 0.0015 | 0.1214 | 0.0667 |
| 1000 | 0.6930 | 2.0364 | -1.0658 | 0.6932 | 1.9088 | -0.9172 | 0.6982 | 2.0339 | -1.0286 |
| | 0.0011 | 0.7614 | 0.5599 | 0.0010 | 0.0393 | 0.0289 | 0.0007 | 0.0678 | 0.0378 |

Table 2. Parameter estimates and their mean squared errors.

$$PMAE(k) = \frac{1}{m} \sum_{i=1}^m |X_{(t+i)} - \hat{X}_{(t+i)}^{(k)}|,$$

where $\hat{X}_{(t+i)}^{(k)} = \widehat{median}(X_{t+i}|X_{t-k+i})$ is the k -step ahead conditional median of the fitted process, and

$$PTP(k) = \frac{1}{m} \sum_{i=1}^m I(X_{(t+i)} = \hat{X}_{(t+i)}^{(k)}) \times 100\%,$$

where $I(\cdot)$ denotes the indicator function. Here, $X_{(t+i)}$ is the actual i th observation at time point $(t + i)$ and $\hat{X}_{(t+i)}^{(k)}$ is the k -step ahead forecast value at the same time point.

We have obtained the conditional mean of the process as traditional forecast while the median and mode from one-step ahead forecast distribution serve as coherent forecasts. For the computation of PMAE in the SARIMA model, we have used mean forecast while in the INAR(1), ZMPL model, we have used the median of the one-step ahead forecast distribution. In the SARIMA model, we can have only mean as the forecast, for computation of percentage of true prediction (PTP) based on mean we have considered rounded mean for both models. We have computed median and mode from the forecast distribution for the computation of PTP-Med and PTP-Mode. In INAR literature, the median and mode of the forecast distribution are called the coherent forecasts.

| k | SARIMA | | | INAR(1) _s ZMPL | | | | |
|--|--------|-------|-------|---------------------------|-------|----------|---------|----------|
| | PRMSE | PMAE | PTP | PRMSE | PMAE | PTP-Mean | PTP-Med | PTP-Mode |
| $\lambda = 0.3, \alpha = 0.5, \delta = 0.7$ | | | | | | | | |
| 1 | 2.376 | 1.555 | 10.00 | 2.374 | 1.186 | 9.80 | 58.10 | 61.30 |
| 2 | 2.622 | 1.727 | 8.90 | 2.618 | 1.411 | 9.50 | 55.00 | 57.30 |
| 3 | 2.602 | 1.679 | 8.10 | 2.592 | 1.300 | 8.20 | 59.20 | 60.20 |
| 4 | 2.545 | 1.643 | 9.10 | 2.546 | 1.244 | 8.50 | 58.60 | 62.80 |
| 5 | 2.443 | 1.595 | 10.20 | 2.438 | 1.203 | 9.80 | 60.30 | 62.30 |
| $\lambda = 0.5, \alpha = 1, \delta = 0.5$ | | | | | | | | |
| 1 | 1.787 | 1.310 | 17.50 | 1.784 | 1.140 | 17.50 | 43.10 | 47.50 |
| 2 | 1.671 | 1.254 | 16.60 | 1.673 | 1.035 | 16.90 | 45.20 | 51.10 |
| 3 | 1.791 | 1.322 | 16.10 | 1.788 | 1.122 | 16.00 | 43.30 | 50.60 |
| 4 | 1.809 | 1.335 | 15.30 | 1.810 | 1.134 | 15.00 | 42.30 | 49.50 |
| 5 | 1.834 | 1.335 | 17.00 | 1.832 | 1.128 | 17.00 | 46.00 | 48.70 |
| $\lambda = 0.6, \alpha = 1.5, \delta = -0.5$ | | | | | | | | |
| 1 | 2.391 | 1.730 | 19.60 | 2.391 | 1.701 | 19.60 | 22.20 | 23.70 |
| 2 | 2.209 | 1.677 | 19.00 | 2.213 | 1.645 | 18.70 | 20.70 | 22.20 |
| 3 | 2.483 | 1.868 | 15.60 | 2.470 | 1.839 | 15.10 | 18.00 | 21.60 |
| 4 | 2.307 | 1.705 | 19.50 | 2.302 | 1.674 | 18.90 | 20.80 | 22.00 |
| 5 | 2.300 | 1.719 | 19.50 | 2.294 | 1.683 | 19.20 | 20.90 | 22.70 |
| $\lambda = 0.7, \alpha = 2, \delta = -1$ | | | | | | | | |
| 1 | 2.584 | 1.875 | 19.20 | 2.587 | 1.858 | 18.90 | 19.90 | 20.30 |
| 2 | 2.448 | 1.856 | 17.30 | 2.442 | 1.841 | 17.30 | 18.10 | 20.30 |
| 3 | 2.586 | 1.873 | 18.90 | 2.575 | 1.849 | 18.50 | 20.50 | 22.00 |
| 4 | 2.565 | 1.862 | 20.90 | 2.562 | 1.846 | 20.30 | 21.10 | 19.80 |
| 5 | 2.752 | 1.959 | 21.30 | 2.736 | 1.953 | 21.00 | 21.60 | 20.20 |

Table 3. Forecast comparison with SARIMA.

From Table 3, it can be seen that in terms of PRMSE both models perform equally well, as their forecasts are the same. However, the PMAE for the SARIMA model is higher than that of the PMAE for the proposed model, showing the better performance of the proposed model than the SARIMA model. This may be because median is used in the proposed model and mean is used in the SARIMA model for the computation of the PMAE. But when it comes to the coherent forecast performance, the proposed model outperforms the SARIMA model, which can be seen from the PTP for median and mode. When data are overdispersed, as it is the case in the top panel in the table, the coherent forecast performs five times better than the traditional forecast. Both models perform equally well when the data are underdispersed.

Application of INAR(1)_sZMPL model

We have analyzed weekly Influenza data from the Breisgau- Hochschwarzwald county of Baden–Württemberg state, Germany for the years 2001 to 2008 (<https://survstat.rki.de>). The data have mean 0.4735 and variance 2.5969. Clearly the data are overdispersed. From the relative frequency plot in Fig. 2, we can see that the number of zeros in the data is excessively high. Hence, this series can be modeled using the proposed zero-modified model. The seasonality of the data can be seen from the ACF and PACF plot in Fig. 3, as it is characterized by the sinusoidal autocorrelation pattern. Also, the data are seasonal because the corresponding ACF plot has significant peaks at multiples of 54. From Fig. 4, it can be seen that the AIC and BIC values are small for the period $s = 54$. Hence the seasonal period of the series ($s = 54$) is confirmed.

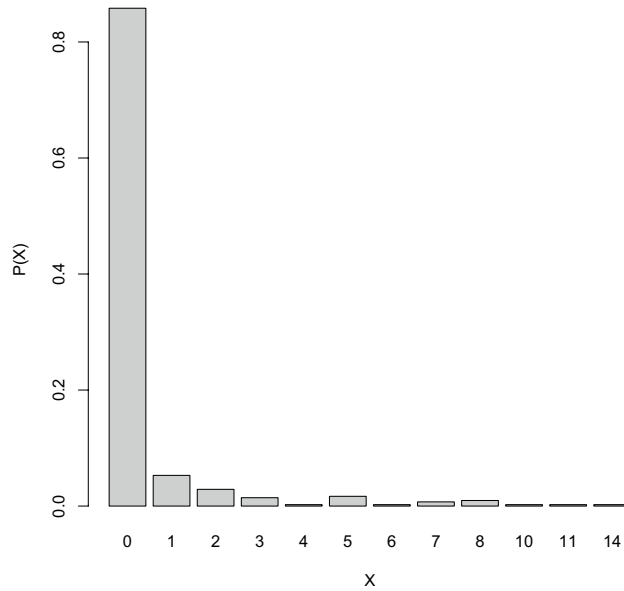


Figure 2. Relative frequency plot for the influenza data.

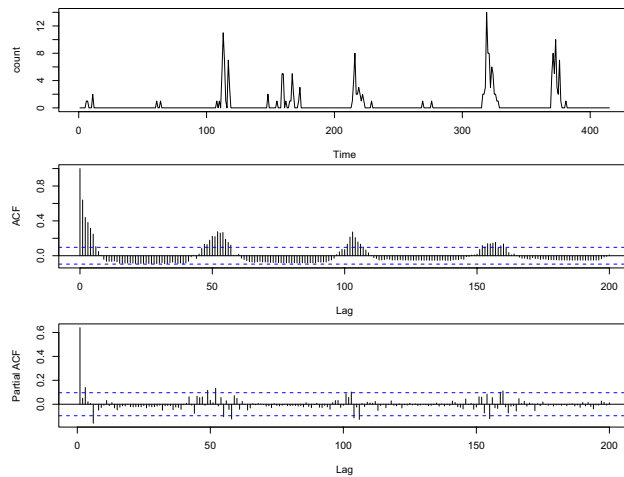


Figure 3. Sample path and sample ACF influenza data.

From Table 4, it can be seen that the proposed model is the best model for the data. We have obtained point forecast using the $INAR(1)_sZMPL$ model for the given data. The data has 416 observations. We have used 411 observations (training data) for estimating the parameters and last five observations are predicted using the fitted model. We have used the conditional ' h ' step ahead mean as the forecast function. The parameter estimates (CML) based on training data are $\hat{\lambda} = 0.2497$, $\hat{\alpha} = 0.5289$, $\hat{\delta} = 0.8619$. The mean forecast \hat{X}_{t+h} and the rounded mean forecast $[\hat{X}_{t+h}]$ are given in the Table 5. From this table it can be seen that the model predicts the future observations with good accuracy.

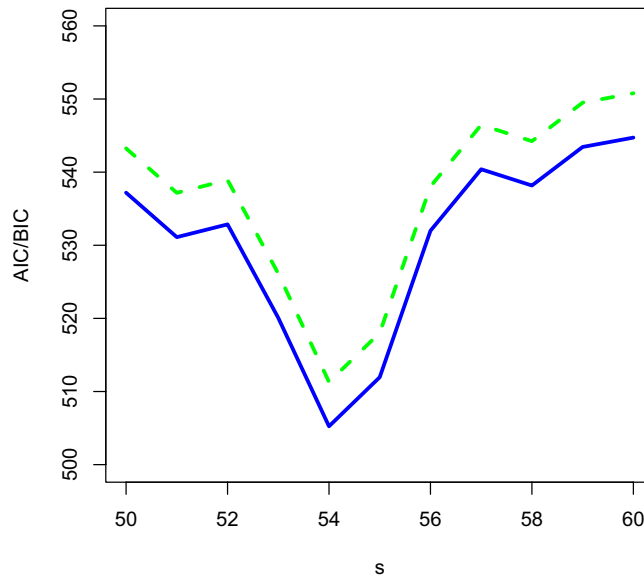


Figure 4. Akaike information criterion (AIC) and Bayesian information criterion (BIC) for influenza data for various seasonal periods ‘s’. Green dashed line is for the BIC and Blue solid line is for AIC.

| Model | Parameter estimates | AIC | BIC |
|-------------------------------------|--|---------|---------|
| PINAR(1) _s ¹⁴ | $\hat{\phi}=0.2435, \hat{\lambda}=0.4288$ | 939.950 | 948.011 |
| GINAR(1) _s ¹³ | $\hat{\phi}=0.4891, \hat{\theta}=0.4993$ | 627.946 | 636.946 |
| SNGINAR(s) ¹⁶ | $\hat{\alpha}=0.2900, \hat{\mu}=0.6499$ | 660.493 | 668.554 |
| INAR(1) _s ZMPL | $\hat{\lambda} = 0.2509, \hat{\alpha} = 0.5286, \hat{\delta}=0.8640$ | 505.243 | 517.535 |

Table 4. Model selection among seasonal INAR models using AIC and BIC.

| h | X_{411+h} | \hat{X}_{411+h} | $[\hat{X}_{411+h}]$ |
|---|-------------|-------------------|---------------------|
| 1 | 0 | 0.2850 | 0 |
| 2 | 0 | 0.2850 | 0 |
| 3 | 0 | 0.2850 | 0 |
| 4 | 0 | 0.2850 | 0 |
| 5 | 0 | 0.2850 | 0 |

Table 5. Point forecasts for the influenza data.

Conclusion

In this paper, we have introduced a seasonal, nonnegative, integer-valued autoregressive model, with multiple features, using the negative binomial thinning operator and zero-modified Poisson-Lindley distributed innovations. General seasonal INAR model of order P has been discussed in the paper and it has been illustrated in detail for order one. This model, denoted by INAR(1)_sZMPL model, is the first of its kind for modeling zero-deflated or zero-inflated seasonal time series of counts. Some theoretical results are determined for the model. Specifically, means, variances and autocorrelation function are obtained. In estimating the parameters of the model, the Yule-Walker, conditional least squares and conditional maximum likelihood approaches are given due consideration. The simulation results obtained in this study demonstrate the superiority of the conditional maximum likelihood method over the other two point estimation methods of estimating the model parameters. A real-life application of the model was investigated by analyzing a zero-inflated overdispersed seasonal count time series. The model fit was compared to the fits of three competing models. In the final analysis, the proposed model outperforms the other models.

Data availability

The datasets used and/or analysed during the current study can be provided by the corresponding author upon reasonable request.

Received: 12 June 2023; Accepted: 30 November 2023

Published online: 12 December 2023

References

1. Tamerius, J. *et al.* Global influenza seasonality: Reconciling patterns across temperate and tropical regions. *Environ. Health Perspect.* **119**, 439–445 (2011).
2. Al-Osh, M. A. & Alzaid, A. A. First-order integer-valued autoregressive (INAR (1)) process. *J. Time Ser. Anal.* **8**, 261–275 (1987).
3. McKenzie, E. Some simple models for discrete variate time series 1. *J. Am. Water Resour. Assoc.* **21**, 645–650 (1985).
4. Altun, E. A new one-parameter discrete distribution with associated regression and integer-valued autoregressive models. *Math. Slov.* **70**, 979–994 (2020).
5. Mohammadpour, M., Bakouch, H. S. & Shirozhan, M. Poisson–Lindley INAR(1) model with applications. *Braz. J. Probab. Stat.* **32**, 262–280 (2018).
6. Liu, Z. & Zhu, F. A new extension of thinning-based integer-valued autoregressive models for count data. *Entropy* **23**, 62 (2020).
7. Ristić, M. M., Bakouch, H. S. & Nastić, A. S. A new geometric first-order integer-valued autoregressive (NGINAR (1)) process. *J. Stat. Plan. Inference* **139**, 2218–2226 (2009).
8. Bourguignon, M., Rodrigues, J. & Santos-Neto, M. Extended Poisson INAR (1) processes with equidispersion, underdispersion and overdispersion. *J. Appl. Stat.* **46**, 101–118 (2019).
9. Bourguignon, M. & Weiß, C. H. An INAR (1) process for modeling count time series with equidispersion, underdispersion and overdispersion. *TEST* **26**, 847–868 (2017).
10. Eliwa, M. S., Altun, E., El-Dawoody, M. & El-Morshedy, M. A new three-parameter discrete distribution with associated INAR (1) process and applications. *IEEE Access* **8**, 91150–91162 (2020).
11. Barreto-Souza, W. Zero-modified geometric INAR (1) process for modelling count time series with deflation or inflation of zeros. *J. Time Ser. Anal.* **36**, 839–852 (2015).
12. Sharafi, M., Sajjadnia, Z. & Zamani, A. A first-order integer-valued autoregressive process with zero-modified Poisson–Lindley distributed innovations. *Commun. Stat. Simul. Comput.* **20**, 1–18 (2020).
13. Awale, M., Kashikar, A. & Ramanathan, T. Modeling seasonal epidemic data using integer autoregressive model based on binomial thinning. *Model. Assist. Stat. Appl.* **15**, 1–17 (2020).
14. Arwaekaji, M., Sillabutra, J., Viwatwongkasem, C. & Soontornpipit, P. Forecasting influenza incidence in public health region 8 Udonthani, Thailand by SARIMA model. *Curr. Appl. Sci. Technol.* **20**, 10–55003 (2022).
15. Zhao, Z. *et al.* Study on the prediction effect of a combined model of SARIMA and LSTM based on SSA for influenza in Shanxi province, China. *BMC Infect. Dis.* **23**, 71 (2023).
16. Riaz, M. *et al.* Epidemiological forecasting models using ARIMA, SARIMA, and Holt-Winter multiplicative approach for Pakistan. *J. Environ. Public Health* **20**, 23 (2023).
17. Qaddus, M. A. Time series count data models: An empirical application to traffic accidents. *Accid. Anal. Prev.* **40**, 1732–1741 (2008).
18. Homburg, A., Weiß, C. H., Alwan, L. C., Frahm, G. & Göb, R. A performance analysis of prediction intervals for count time series. *J. Forecast.* **40**, 603–625 (2021).
19. Kong, J. & Lund, R. Seasonal count time series. *J. Time Ser. Anal.* **44**, 93–124 (2023).
20. Bourguignon, M., LP Vasconcellos, K., Reisen, V. A. & Ispány, M. A Poisson INAR (1) process with a seasonal structure. *J. Stat. Comput. Simul.* **86**, 373–387 (2016).
21. Okereke, E. W., Gideon, S. N. & Omekara, C. O. A seasonal INAR (1) process with geometric innovation for over dispersed count time series. *Int. J. Stat. Reliab. Eng.* **6**, 82–100 (2020).
22. Tian, S., Wang, D. & Cui, S. A seasonal geometric INAR process based on negative binomial thinning operator. *Stat. Pap.* **61**, 2561–2581 (2020).
23. Jin-Guan, D. & Yuan, L. The integer-valued autoregressive (INAR (p)) model. *J. Time Ser. Anal.* **12**, 129–142 (1991).
24. Merabet, F. & Zeghdoud, H. On modelling seasonal ARIMA series: Comparison, application and forecast (number of injured in road accidents in northeast algeria). *Wseas Trans. Syst. Control* **15**, 235–246 (2020).
25. Klimko, L. A. & Nelson, P. I. On conditional least squares estimation for stochastic processes. *Ann. Stat.* **20**, 629–642 (1978).
26. Xavier, D., Santos-Neto, M., Bourguignon, M. & Tomazella, V. Zero-modified Poisson–Lindley distribution with applications in zero-inflated and zero-deflated count data. [arXiv:1712.04088](https://arxiv.org/abs/1712.04088) (arXiv preprint) (2017).

Acknowledgements

The work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. 3290].

Keywords: Seasonal INAR time series, Forecasting, Statistical model, Estimation, Simulation.

Author contributions

Conceptualization, H.S.B. and E.W.O.; methodology, E.W.O., F.E.A. and H.S.B.; software, M.A.; formal analysis, E.W.O., M.A., F.E.A. and H.S.B.; investigation, E.W.O., M.A., F.E.A. and H.A.; writing—original draft preparation, E.W.O. and M.A.; writing—review and editing, F.E.A. and H.A.; funding acquisition, F.E.A. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to F.E.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023