One-misrecorded Poisson INAR(1) model via two random operators with application to crime and economics data

M. M. Gabr[1] · Hassan S. Bakouch[2,3] · Shimaa M. El-Hadidy[4]

Abstract

In this paper, we propose a first-order non-negative integer-valued autoregressive (INAR(1)) model with one-misrecorded Poisson (OMP) marginal via a combination of the generalized binomial thinning and mixture Pegram operators for zero-inflated and one deflated count time series. The suggested model is suitable for multimodal, equi- and over-dispersed data modeling. It contains two particular cases: the mixture of Pegram and thinning of a first-order integer-valued autoregressive (MPT(1)) with Poisson and one-misrecorded Poisson. The distribution of the innovation term is derived as a mixture of degenerate distributions at 0 and 1, and two Poisson distributions with certain parameters. Regression and several statistical properties of the proposed model are discussed. We investigate the distribution of the runs (the lengths of zeros and ones). The parameters of the model are estimated using the maximum-likelihood, modified Yule–Walker, and modified conditional least squares methods. The estimation of the parameters, their behavior, and their performance are presented through a simulation study. Two practical data sets on the monthly cases of criminal records and weekly sales are applied to check the proposed process's performance against other relevant INAR(1) models, showing its capabilities in the challenging case of over-dispersed count data. Furthermore, the proposed model discusses data forecasting.

Communicated by Clémentine Prieur.

✉ Shimaa M. El-Hadidy
shymaaelhadidy2015@yahoo.com

1 Department of Mathematics, Faculty of Science, Alexandria University, Alexandria, Egypt

2 Department of Mathematics, College of Science, Qassim University, Buraydah, Saudi Arabia

3 Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt

4 Department of Mathematics, Faculty of Science, Mansoura University, Mansoura, Egypt

1 Introduction and basics

In the recent years, there has been a lot of interest in modeling and forecasting the temporal dependence and evolution of time series of low counts. Integer-valued time series arise in many applications, such as queueing systems, finance, reliability theory, medicine, and epidemiology, among others. The non-negative integer-valued autoregressive (INAR) model was developed by Al-Osh and Alzaid (1987) based on the binomial thinning operator introduced by Steutel and van Harn (1979) and McKenzie (1985). The INAR process has undergone numerous generalizations and modifications thanks to the efforts of many authors, including Alzaid and Al-Osh (1990), Al-Osh and Aly (1992), and Alzaid and Al-Osh (1993). Models based on the negative binomial thinning operator were also constructed and taken into consideration (Aly and Bouzar 1994; Ristić et al. 2009; Ristić and Nastić 2012). All of the models previously described have one common feature, which is considering on independent counting random variables to the thinning operation. The assumption of independence simplifies the computations and enables obtaining simple, not complex, models with nice characteristics. However, the fundamental drawback of such models is that they cannot be used in practice with data that do not support this assumption. For example, it is reasonable to expect some dependence between the survival or bankruptcy of some big companies in the economy, or it is logical to expect that the occurrence of some natural disasters affects the appearance or non-appearance of some other disasters in meteorology, etc. Therefore, a model with dependent counting sequences is needed to be established. Brännäs and Hellström (2001), Ristić et al. (2013), Ilić (2016), and Nastić et al. (2017) introduced thinning operators with dependent counting series. In this paper, the proposed model is based on the generalized thinning with dependent Bernoulli counting series and Pegram's mixture operators that were introduced by Ilić (2016) and Biswas and Song (2009), respectively.

The generalized binomial thinning operator is defined as follows:

Definition 1.1 (*Generalized binomial thinning operator with a dependent Bernoulli counting series, GBT*) As in Ilić (2016), the GBT operator $\alpha \star_\vartheta$ was defined as

$$\alpha \star_\vartheta X \overset{d}{=} \begin{cases} X, & w.p. \ 1 - (1-\alpha)/\vartheta, \\ Bin(X, 1-\vartheta), & w.p. \ (1-\alpha)/\vartheta, \end{cases} \tag{1.1}$$

where X is a non-negative integer-valued random variable (RV), $0 \leq 1 - \alpha \leq \vartheta$, and $\vartheta \in (0, 1]$.

Ilić (2016) defined the first-order dependent counting INAR model [NDCINAR(1)] based on the GBT operator ($\star_\vartheta$) as

$$X_t = \alpha \star_\vartheta X_{t-1} + \xi_t, \ t \in \mathbb{Z}, \tag{1.2}$$

where $\{\xi_t\}$ is a sequence of $i.i.d.$ RVs such that $Cov(\xi_t, X_s) = 0$ for $s < t$. For more details about this operator and its properties, see Ilić (2016).

Pegram (1980) introduced the discrete $AR(p)$ that looks like $ARMA$ models and permits some of the serial correlations to be negative. Biswas and Song (2009) used the Pegram operator to define the discrete $ARMA$ model.

Khoo et al. (2017) introduced an INAR process with Poisson marginal via mixing the binomial thinning that consists of independent Bernoulli counting series, with the mixture Pegram operators (abbreviated MPT), and Khoo (2016) handled the MPT(1) process with some discrete marginal distributions, such as binomial, geometric, and negative binomial. Shirozhan and Mohammadpour (2018) and Shirozhan et al. (2019) presented the INAR(1)

process with Poisson and geometric marginal distributions by combining the mixture Pegram with the generalized binomial thinning, which was based on the Bernoulli dependent counting series (Ristić et al. 2013).

The models based on combining the mixture Pegram and thinning operators are easy to interpret, have a simple form, and are more flexible in terms of the range of correlation. They are appropriate for any discrete marginal distributions, including infinite and finite divisible distributions, unlike the models based on the thinning operator, which are applied only for infinite range distributions, and also applied to the multimodal data.

The conditional variance of the INAR(1) based on the independent counting series is linear in X_t, but for a Pegram model, it is a quadratic function like the INAR(1) with dependent counting series. Therefore, the quadratic property is not satisfied for INAR(1) with independent counting series. This property implies large values of variance, which may be appropriate to model over-dispersed data. The one-misrecorded Poisson distribution—also known as a simple modified Poisson distribution or a Cohen Poisson distribution—forms the basis of the proposed model and is introduced in the following definition.

Definition 1.2 (*The one-misrecorded Poisson*) The one-misrecorded Poisson distribution, abbreviated as (OMP), of a RV X with parameters $\lambda > 0$, $0 \leq \phi \leq 1$, has the probability mass function (p.m.f.)

$$P(X = x) = \begin{cases} e^{-\lambda}(1 + \lambda\phi) & x = 0, \\ e^{-\lambda}\lambda(1 - \phi) & x = 1, \\ e^{-\lambda}\frac{\lambda^x}{x!} & x = 2, 3, \ldots. \end{cases} \tag{1.3}$$

The OMP was invented by Cohen (1960) to model the data (or misclassified data) that contain errors in reporting one as a zero, or it sometimes happens that a few of the observations taking value one are misclassified and reported as zero. ϕ is the probability of misclassification, or the proportion of ones reported as zeros. The OMP distribution has its application in situations where the true values of ones are erroneously recorded as zeros with probability ϕ, while values of two and more are recorded correctly from a Poisson. For example, when determining the number of defects per unit or item examined, an inspector may make an error by reporting units that actually contain a single defect as being perfect or defect-free. The OMP is reduced to a Poisson distribution when $\phi = 0$, the equi-dispersion case. Also, all the observations falling into class one will be reported as zero when $\phi = 1$.

For the OMP distribution, the corresponding probability generating function ($p.g.f.$), mean, variance, second, third, and fourth moments are, respectively, given by

$$\mathbf{G}_X(s) = \lambda\phi e^{-\lambda}(1 - s) + e^{\lambda(s-1)}, \tag{1.4}$$

$$E(X) = \mu_X = \lambda(1 - \phi e^{-\lambda}), \tag{1.5}$$

$$Var(X) = C_2(0) = \sigma_X^2 = \lambda^2 + \lambda(1 - \phi e^{-\lambda})(1 - \lambda(1 - \phi e^{-\lambda})), \tag{1.6}$$

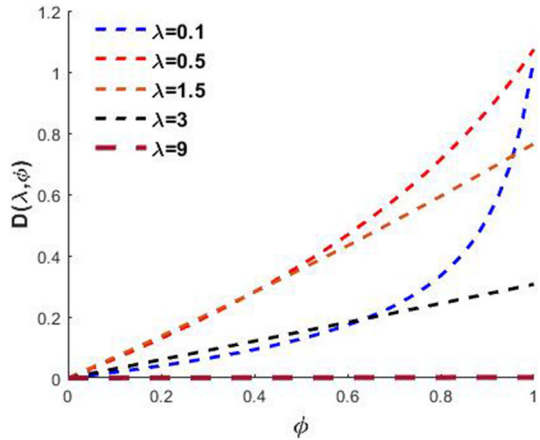$$E(X^2) = \mu_{(0)} = \lambda^2 + \lambda(1 - \phi e^{-\lambda}), \tag{1.7}$$

$$E(X^3) = \mu_{(0,0)} = \lambda^3 + 3\lambda^2 + \lambda(1 - \phi e^{-\lambda}), \tag{1.8}$$

$$E(X^4) = \mu_{(0,0,0)} = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda(1 - \phi e^{-\lambda}). \tag{1.9}$$

Therefore, the Fisher index of dispersion (variance to mean ratio) is obtained as follows:

$$I_X = \frac{Var(X)}{E(X)} = 1 + \frac{\lambda\phi e^{-\lambda}(2 - \phi e^{-\lambda})}{(1 - \phi e^{-\lambda})} = 1 + D(\phi, \lambda),$$

Fig. 1 Behavior of the function $D(\phi, \lambda)$ on the index dispersion

where $D(\phi, \lambda) = \frac{\lambda \phi e^{-\lambda}(2 - \phi e^{-\lambda})}{(1 - \phi e^{-\lambda})}$. $D(\phi, \lambda)$ increases as ϕ increases from 0 to 1 and the degree of over-dispersion increases. The value of function $D(\phi, \lambda) = 0$ for large values of λ or for $\phi = 0$. The behavior of this function can be seen graphically in Fig. 1.

We can say that the OMP distribution is zero-inflated (there is an excess of zeros) and one-deflated (there are fewer ones than expected), according to the two definitions given by Ahmad (1968):

The first definition is that a class in a distribution will be said to be inflated if some of the observations falling in other classes are erroneously misclassified in this class.

The second definition is that a class in a distribution will be said to be deflated if some of the observations falling in the class are erroneously misclassified in another class.

Also, according to Johnson et al. (2005), pages (211 and 353), in describing the modified distribution's zero probability, we can conclude that if $\phi_0 = \lambda \phi e^{-\lambda} > 0$, then the Poisson's zero probability is increased (zero-inflation), while the Poisson's one probability is decreased for $\phi_1 = -\lambda \phi e^{-\lambda} < 0$ (one-deflation).

Moreover, according to Puig and Valero (2006), a random variable X is zero-inflated when its proportion of zeros is greater than the proportion of zeros of a Poisson variate with the same expectation; zero inflation can be studied in a standard way using the zero-inflation index, I_{Zero}, as

$$I_{\text{Zero}} = 1 + \frac{\ln(P_0)}{\mu_X} = 1 + \frac{-\lambda + \ln(1 + \lambda\phi)}{\lambda(1 - \phi e^{-\lambda})} = 1 + Z(\lambda, \phi). \tag{1.10}$$

The value of function $Z(\lambda, \phi) = -1$ for $\phi = 0$ and any value of λ, and this implies that $I_{\text{Zero}} = 0$, consequently, X is a Poisson distribution. The value of $Z(\lambda, \phi)$ is greater than -1 since $Z(\lambda, \phi)$ increases as ϕ increases from 0 to 1. In particular, the behavior of the function $Z(\lambda, \phi)$ is shown graphically in Fig. 2. As a result, the values of $I_{\text{Zero}} > 0$ and from the plot in Fig. 2, we conclude that $I_{\text{Zero}} \in (0, 1)$, which indicates zeroinflation (an excess of zero with respect to a Poisson distribution).

Practical over-dispersed INAR(1) processes have been presented in the literature, such as the negative binomial INAR(1) process (McKenzie 1986) and the compound Poisson INAR(1) process (Schweer and Weiß 2014). These over-dispersed classes of INAR(1) models fail when the data set contains a large number of zeros. To analyze count data with excess zeros, Jazi et al. (2012) and Barreto-Souza (2015) introduced INAR(1) with zero-inflated Poisson (ZIP) innovations and zero-modified geometric INAR(1) models, respectively. The

Fig. 2 Behavior of the function
$Z(\phi, \lambda)$ on the zero index
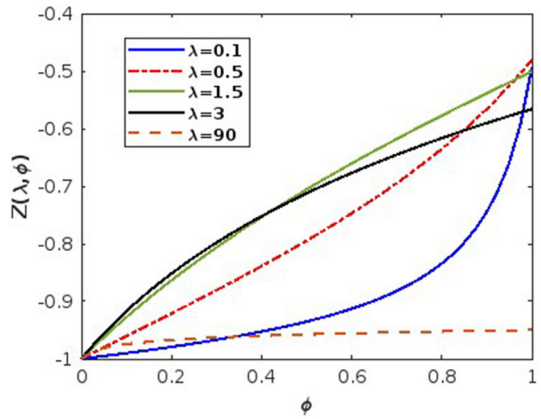
over-dispersed and zero-inflated INAR(1) models, however, fail when a data set contains both a large number of zeros and ones, which occurs frequently in practice, for example, in the number of patients with a rare disease in various areas, the monthly counts of Computer-Aided Dispatch (CAD) drug calls, etc. Furthermore, the illustrative examples in Sect. 4 show real-life data examples with excesses of zero-and-one. So that a change in the p.m.fs. for modeling zero-and-one observation is necessary. Qi et al. (2019), Mohammadi et al. (2022a, b) introduced INAR(1) models with zero-and-one inflated Poisson (ZOIP), Poisson–Lindely (ZOIPL), and geometric (ZOIG) that are useful tools for capturing the characteristics of such count data. In these processes, the authors assume that the innovation term has a zero-and-one inflated distribution.

The main aim of this paper is to construct an INAR(1) process for modeling over-, equi-dispersed, and multimodality data by assuming that the main time series $\{X_t\}$ follows the OMP distribution based on combining the GBT and mixture Pegram operators. Also, the proposed process can be used to model zero-inflated and one-deflated data, and the misclassification of ones as zeros (errors in recording at one or one-misrecorded data).

This paper is constructed as follows: In Sect. 2, an extension or generalization of the MPT(1) model based on a dependent Bernoulli counting series with OMP distribution (OMP-EMPT(1)) is defined, and some properties of the innovation process and its p.m.f. are derived. Also, several statistical and conditional properties of the model, including conditional and joint distributions and jumps, are shown, along with a study of the distribution of zeros and ones in the proposed process. In Sect. 3, the parameters of the process are estimated via the maximum-likelihood (ML), modified conditional least-squares (MCLS), and modified Yule–Walker (MYW) methods. Also, the performance of the estimates is checked via a simulation study. In Sect. 4, we provide two real-life data sets to model over-dispersed data, illustrating the performance of the process throughout the goodness-of-fit statistics. Also, the model's adequacy is checked by the mean logarithmic score, benchmark chart, and non-randomized probability integral transform (PIT). The predictive capacity of the proposed model is checked through forecasting.

2 Construction of the model and its statistical properties

This section is implemented as follows: An extension of the MPT(1) with OMP marginal distribution is constructed using a generalized binomial thinning (abbreviated as OMP-

EMPT(1)). The p.m.f. for the innovation term of the model is obtained. The bivariate $p.g.f.$, conditional $p.g.f.$, conditional mean and variance, $k-$step-ahead conditional mean and variance, transition probability, autocorrelation function, and the number of zeros and ones in the model are discussed.

The proposed model is defined as follows:

Definition 2.1 (*Extended MPT(1) with OMP, OMP-EMPT(1)*) The OMP-EMPT(1) $\{X_t\}$ is defined as

$$X_t = \begin{cases} \alpha \star_\vartheta X_{t-1} & \text{with probability } p, \\ \xi_t & \text{with probability } 1-p, \end{cases}$$

also, this model can be expressed as

$$X_t = \begin{cases} X_{t-1} & \text{w.p. } p\left(1 - \frac{1-\alpha}{\vartheta}\right), \\ (1-\vartheta) \circ X_{t-1} & \text{w.p. } p\frac{(1-\alpha)}{\vartheta}, \\ \xi_t & \text{w.p. } 1-p, \end{cases}$$

where "$\alpha \star_\vartheta$" refers to the GBT operator, and $\{\xi_t\}$ is a sequence of $i.i.d.$ RVs such that $\vartheta \in (0, 1]$, $p \in (0, 1)$, $0 \le 1 - \alpha \le \vartheta$, $\lambda > 0$, and $\phi \in [0, 1]$.

Remark 2.1 Note the following sub-models:

- When $\vartheta = 1 - \alpha$, the GBT is reduced to binomial thinning, and hence, the MPT(1) with OMP marginal, OMP-MPT(1), process is obtained, which has not been discussed before.
- When $\phi = 0$ and $\vartheta = 1 - \alpha$, the constructed process is reduced to the MPT(1) with Poisson distribution that is given by Khoo et al. (2017).

2.1 Probability distribution of the innovation term

The following proposition provides the distribution of the innovation process:

Proposition 2.1 *The $p.m.f.$ for ξ_t has the following representations:*

$$P(\xi_t = 0) = \frac{1}{1-p}$$
$$\times \left[\lambda\phi e^{-\lambda}(1 - p\alpha) + e^{-\lambda}\left(1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)\right) - \left(\frac{p(1-\alpha)e^{-\lambda(1-\vartheta)}}{\vartheta}\right) \right],$$
$$P(\xi_t = 1) = \frac{1}{1-p}\left[-\lambda\phi e^{-\lambda}(1 - p\alpha) + e^{-\lambda}\lambda\left(1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)\right)\right.$$
$$\left. - \left(\frac{p(1-\alpha)(\lambda(1-\vartheta))e^{-\lambda(1-\vartheta)}}{\vartheta}\right) \right],$$
$$P(\xi_t = j) = \frac{1}{1-p}\left[e^{-\lambda}\frac{\lambda^j}{j!}\left(1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)\right)\right.$$
$$\left. - \left(\frac{p(1-\alpha)(\lambda(1-\vartheta))^j e^{-\lambda(1-\vartheta)}}{j!\vartheta}\right) \right], \quad j = 2, 3, \ldots, \quad (2.1)$$

for $\lambda > 0$, $\vartheta \in (0, 1]$, $\alpha \in [0, 1]$, and $0 \le 1 - \alpha \le \vartheta$. The condition

$$0 \le p \le \min(C_1, C_2), \quad (2.2)$$

where C_1 and C_2 are defined as

$$C_1 = \frac{1 + \lambda\phi}{\lambda\phi\alpha + \left(1 - \frac{1-\alpha}{\vartheta}\right) + \left(\frac{1-\alpha}{\vartheta}\right)e^{\lambda\vartheta}}, \tag{2.3}$$

$$C_2 = \frac{1 - \phi}{\frac{1-\alpha}{\vartheta}(1 - \vartheta)e^{\lambda\vartheta} + \left(1 - \frac{1-\alpha}{\vartheta}\right) - \alpha\phi}, \tag{2.4}$$

is necessary for the p.m.f. of the innovation process ξ_t in Eq. (2.1) *to be well defined and non-negative for any j.*

Proof By using the definition of $(1 - \vartheta) \circ X = \sum_{i=1}^{X} Z_i$, where $Z_i \sim Ber((1 - \vartheta))$, as a result the $p.g.f.$ of $(1 - \vartheta) \circ X$ is given by $G_{(1-\vartheta)\circ X} = G_X(\vartheta + (1 - \vartheta)s)$. First, we get the $p.g.f.$ of the innovation process $\{\xi_t\}$ as a mixture representation of ones to the $\{X_t\}$ as follows:

For this purpose, the $p.g.f.$ of $\{X_t\}$ given by Definition (2.1) has the expression

$$\mathbf{G}_{X_t}(s) = p\mathbf{G}_{\alpha\star\vartheta X_{t-1}}(s) + (1 - p)\mathbf{G}_{\xi_t}(s)$$

$$= p\left(1 - \frac{1-\alpha}{\vartheta}\right)\mathbf{G}_{X_{t-1}}(s) + p\left(\frac{1-\alpha}{\vartheta}\right)\mathbf{G}_{(1-\vartheta)\circ X_{t-1}}(s) + (1 - p)\mathbf{G}_{\xi_t}(s)$$

$$= p\left(1 - \frac{1-\alpha}{\vartheta}\right)\mathbf{G}_{X_{t-1}}(s) + p\left(\frac{1-\alpha}{\vartheta}\right)\mathbf{G}_{X_{t-1}}(\vartheta + (1 - \vartheta)s) + (1 - p)\mathbf{G}_{\xi_t}(s).$$

Hence, the $p.g.f.$ of the innovation process $\{\xi_t\}$ is

$$\mathbf{G}_{\xi_t}(s) = \frac{1}{1 - p}\left[\mathbf{G}_{X_t}(s) - p\left(1 - \frac{1-\alpha}{\vartheta}\right)\mathbf{G}_{X_{t-1}}(s) - p\left(\frac{1-\alpha}{\vartheta}\right)\mathbf{G}_{X_{t-1}}(\vartheta + (1 - \vartheta)s)\right]$$

$$= \frac{1}{1 - p}\left[\lambda\phi e^{-\lambda}(1 - s) + e^{\lambda(s-1)} - p\left(1 - \frac{1-\alpha}{\vartheta}\right)(\lambda\phi e^{-\lambda}(1 - s) + e^{\lambda(s-1)})\right.$$

$$\left. - p\left(\frac{1-\alpha}{\vartheta}\right)(\lambda\phi e^{-\lambda}(1 - (\vartheta + (1 - \vartheta)s)) + e^{\lambda((\vartheta+(1-\vartheta)s)-1)})\right]$$

$$= \frac{1}{1 - p}\left[\lambda\phi e^{-\lambda}(1 - p\alpha)(1 - s) + \left(1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)\right)e^{\lambda(s-1)}\right.$$

$$\left. - p\left(\frac{1-\alpha}{\vartheta}\right)e^{\lambda(1-\vartheta)(s-1)}\right]. \tag{2.5}$$

From Eq. (2.5), it can be noted that $\mathbf{G}_{\xi_t}(1) = 1$ and $\mathbf{G}_{\xi_t}(0) = p(\xi_t = 0)$ with (2.2).

Using Eq. (2.5), the innovation process $\{\xi_t\}$ has a mixture distribution of degenerate distributions at 0 and 1, $Poisson(\lambda)$ and $Poisson((1 - \vartheta)\lambda)$ with mixing weights $\frac{\lambda\phi e^{-\lambda}(1-p\alpha)}{1-p}$, $-\frac{\lambda\phi e^{-\lambda}(1-p\alpha)}{1-p}$, $\frac{1-p(1-\frac{1-\alpha}{\vartheta})}{1-p}$, and $\frac{-p(\frac{1-\alpha}{\vartheta})}{1-p}$, respectively. Hence, the p.m.f. of ξ_t is represented by equation's system given in Eq. (2.1). Finally, we demonstrate as follows that this function is a p.m.f.

By the definition of generalized mixture of the distribution functions (Mohammadpour et al. 2018), the sum of preceding mixing weights equals 1, and hence, it follows that $\sum_{j=0}^{\infty} P(\xi_t = j) = 1$.

The following step will demonstrate and confirm that the system of equations outlined in Eq. (2.1) yields non-negative probabilities under the condition specified in by Eq. (2.1). For each $j = 0, 1, 2, \ldots$, we aim to prove that $P(\xi_t = j) \geq 0$ as outlined below.

When $j = 0$, ensuring $P(\xi_t = 0) \geq 0$ is achieved through solving the subsequent inequality

$$\frac{1}{1 - p} \left[\lambda\phi e^{-\lambda}(1 - p\alpha) + \left(1 - p\left(1 - \frac{1 - \alpha}{\vartheta}\right)\right) e^{-\lambda} - p\left(\frac{1 - \alpha}{\vartheta}\right) e^{-\lambda(1-\vartheta)} \right] \geq 0;$$

consequently,

$$\lambda\phi e^{-\lambda} + e^{-\lambda} \geq p \left[\alpha\lambda\phi e^{-\lambda} + \left(1 - \frac{1 - \alpha}{\vartheta}\right) e^{-\lambda} + \left(\frac{1 - \alpha}{\vartheta}\right) e^{-\lambda(1-\vartheta)} \right].$$

Upon solving the aforementioned inequality for p, the following result is obtained:

$$p \leq \frac{\lambda\phi + 1}{\alpha\phi\lambda + (1 - \frac{1-\alpha}{\vartheta}) + (\frac{1-\alpha}{\vartheta})e^{\lambda\vartheta}}. \tag{2.6}$$

In the case of $j = 1$, the condition $P(\xi_t = 1) \geq 0$ is satisfied if

$$\frac{1}{1 - p} \left[-\lambda\phi e^{-\lambda}(1 - p\alpha) + e^{-\lambda}\lambda(1 - p(1 - \frac{1 - \alpha}{\vartheta})) - \left(\frac{p(1 - \alpha)(\lambda(1 - \vartheta))e^{-\lambda(1-\vartheta)}}{\vartheta}\right) \right]$$
$$\geq 0;$$

therefore, we have

$$-\lambda\phi e^{-\lambda} + \lambda e^{-\lambda} \geq p \left[\frac{1 - \alpha}{\vartheta}(\lambda(1 - \vartheta))e^{-\lambda(1-\vartheta)} + \left(1 - \frac{1 - \alpha}{\vartheta}\right)\lambda e^{-\lambda} - \alpha\phi\lambda e^{-\lambda} \right].$$

Following the resolution for the preceding inequality, the condition for p is determined to be

$$p \leq \frac{1 - \phi}{-\alpha\phi + \left(1 - \frac{1-\alpha}{\vartheta}\right) + \frac{1-\alpha}{\vartheta}(1 - \vartheta)e^{\lambda\vartheta}}. \tag{2.7}$$

Finally, when $j = 2, 3, 4, \ldots$, the condition $P(\xi_t = j) \geq 0$ is satisfied if

$$\frac{1}{1 - p} \left[e^{-\lambda}\frac{\lambda^j}{j!} \left(1 - p\left(1 - \frac{1 - \alpha}{\vartheta}\right)\right) - \left(\frac{p(1 - \alpha)(\lambda(1 - \vartheta))^j e^{-\lambda(1-\vartheta)}}{j!\vartheta}\right) \right] \geq 0,$$

or equivalently

$$1 - p\left(1 - \frac{1 - \alpha}{\vartheta}\right) \geq p\frac{(1 - \alpha)(1 - \vartheta)^j e^{\lambda\vartheta}}{\vartheta}.$$

Consequently, the condition for p will be

$$p \leq \frac{1}{\left(1 - \frac{1-\alpha}{\vartheta}\right) + \frac{1-\alpha}{\vartheta}(1 - \vartheta)^j e^{\lambda\vartheta}}. \tag{2.8}$$

Let us assume that

$$C_1 = \frac{1 + \lambda\phi}{\lambda\phi\alpha + \left(1 - \frac{1-\alpha}{\vartheta}\right) + \left(\frac{1-\alpha}{\vartheta}\right)e^{\lambda\vartheta}},$$

$$C_2 = \frac{1 - \phi}{\frac{1-\alpha}{\vartheta}(1 - \vartheta)e^{\lambda\vartheta} + \left(1 - \frac{1-\alpha}{\vartheta}\right) - \alpha\phi},$$

$$C_3 = \frac{1}{1 - \frac{1-\alpha}{\vartheta} + \frac{1-\alpha}{\vartheta}(1 - \vartheta)^j e^{\lambda\vartheta}}.$$

Hence, it is evident that C_3 increases with an increase in j, reaching its minimum at $j = 2$. This is depicted in Fig. 3 and supported by the numeric findings (conditions, confirmation numeric values, and folder of plots) available at https://drive.google.com/drive/folders/1oDmUr-vuSw98V9eYDmIsLbKA6Ha35J1t?usp=sharing. Hence, C_3 is consistently greater than or equal to C_2 or C_1, leading to the conclusion that $\min(C_1, C_2, C_3) = \min(C_1, C_2)$. □

Remark 2.2 Taking the minimum of C_1 and C_2 guarantees that the p.m.f. of ξ_t is non-negative for $j = 0, 1, 2, \ldots$.

From the numeric results and the graphs of the two conditions (C_1 and C_2), the minimum is changed from C_1 to C_2, and this means that there is an inflection point or intersection point of the two conditions. By setting $C_1 = C_2$, the value of the intersection point, ϕ, can be obtained as follows:

$$\phi = \phi_0 = \frac{\vartheta e^{\lambda \vartheta}}{\vartheta + (e^{\lambda \vartheta} - 1)(1 + \lambda - \lambda \vartheta)}.$$

ϕ_0 is considered as an inflection point between the two conditions, and at ϕ_0, we have $C_1 = C_2$. Figure 4 illustrates the two conditions, C_1 and C_2, plotted under the specified intervals for ϕ, with $0 \leq \phi \leq \phi_0$ and $\phi_0 \leq \phi \leq 1$, and for $\lambda = 0.5, 1, 5, 10, \alpha = 0.83$, and $\vartheta = 0.21$. The illustrations in Fig. 4, coupled with the confirmation numeric values file (https://drive.google.com/file/d/1-hHVIoHLUMZUn2Z9o-AxghExHADjzHPY/view?usp=drive_link), demonstrate that:

- When $\phi < \phi_0$, C_1 becomes the minimum. Therefore, p should be within the interval $0 \leq p \leq C_1$.
- When $\phi > \phi_0$, C_2 becomes the minimum. Thus, p should be within the interval $0 \leq p \leq C_2$.
- If $\phi = \phi_0$, then $C_2 = C_1 = C$. Hence, p should be within the interval $0 \leq p \leq C$.

Remark 2.3   • If $\phi = 0$, then $\min(C_1, C_2) = C_1$, which implies that p falls within the interval $0 \leq p \leq C_1$. The condition in Eq. (2.2) is reduced to that proposed by Khoo et al. (2017) for the case where $\phi = 0$ and $\vartheta = 1 - \alpha$, in the context of the MPT(1) with Poisson distribution, which is a sub-model of the proposed model.
- If $\phi = 1$, then $\min(C_1, C_2) = C_2$ and consequently $0 \leq p \leq C_2$.

As a visual illustration of the p.m.f. under the mentioned conditions:
For $\lambda = 10; \alpha = 0.83; \vartheta = 0.21$, the p.m.f. of the innovation process, ξ_t, **is shown in**:

- Figure 5a with $\phi = \phi_0 = 0.02679988802$ and $C_1 = C_2 = 0.1805334165$, so p must be in the range $0 \leq p \leq 0.1805334165$.
- Figure 6a with $\phi = 0.5 > \phi_0$ and taking $p \leq C2$, i.e., $p \leq 0.1000412914$.
  For $\lambda = 1; \alpha = 0.6; \vartheta = 0.5$, the p.m.f of ξ_t is represented in:

- Figure 5a with $\phi = \phi_0 = 0.5596162930$ and $C_1 = C_2 = 0.8408782799$, so $0 \leq p \leq 0.8408782799$.
- Figure 6b with $\phi = 0.7 > \phi_0$ and taking $p \leq C2$, i.e., $p \leq 0.6826117049$.

Consequently, based on the aforementioned proposition, the mean, variance, and the second and third moments of $\{\xi_t\}$ are provided as follows, respectively:

$$E(\xi_t) = \frac{\lambda(1 - p\alpha)(1 - \phi e^{-\lambda})}{1 - p} = \mu_{\xi_t}, \tag{2.9}$$

(a) $\phi \leq 0.01$.   (b) $\phi \leq 1$
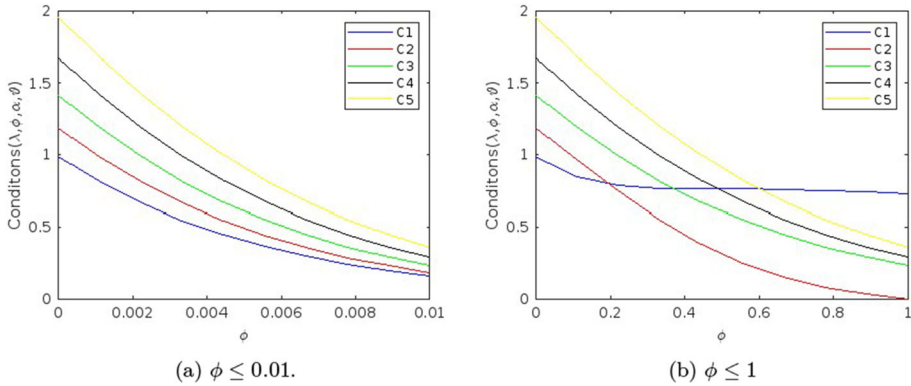
Fig. 3 Conditions $C_1, C_2,$ and C_3 at $j = 2, 3, 4$; $\lambda = 0.1\ to\ 10$; $\alpha = 0.83$; $\vartheta = 0.21$

$$Var(\xi_t) = \frac{1}{1-p}\Big[(\lambda^2(1 + p + p\alpha(-2 + \vartheta) - p\vartheta))$$
$$+(\lambda(1 - \phi e^{-\lambda}))(1 - p\alpha)\left(1 - \frac{(1 - p\alpha)(1 - \phi e^{-\lambda})\lambda}{1 - p}\right)\Big], \quad (2.10)$$

$$E(\xi_t^2) = \frac{(1 - p\alpha)\lambda(1 - \phi e^{-\lambda}) + \lambda^2(1 + p + p\alpha(-2 + \vartheta) - p\vartheta)}{1 - p}, \quad (2.11)$$

and

$$E(\xi_t^3) = \frac{(1 - p\alpha)\lambda(1 - \phi e^{-\lambda}) + 3(\lambda^2(1 + p + p\alpha(-2 + \vartheta) - p\vartheta))}{1 - p}$$
$$- \frac{\lambda^3(-1 + p(-2 + 3\vartheta - \vartheta^2 + \alpha(3 - 3\vartheta + \vartheta^2)))}{1 - p}. \quad (2.12)$$

2.2 Conditional properties of the process, joint distributions, and jumps

Regression properties are useful for estimating unknown parameters as well as forecasting future values. In particular, the conditional expectation is used for the conditional least-squares estimation method, k-step-ahead expectation for forecasting the future values of time series data, and a conditional probability function for the maximum-likelihood estimation method. The joint distributions are used to check the model's time reversibility.

Let $\{X_t\}$ be a stationary OMP-EMPT(1) process, then the conditional probability generating function ($c.p.g.f.$) and its corresponding conditional probability (one-step transition probability) function are as follows:

$$\mathbf{G}_{X_t|X_{t-1}}(s) = E(s^{X_t}|X_{t-1})$$
$$= pE(s^{\alpha \star_\vartheta X_{t-1}}|X_{t-1}) + (1 - p)E(s^{\xi_t}|X_{t-1})$$
$$= p\left[\left(1 - \frac{1-\alpha}{\vartheta}\right)s^{X_{t-1}} + \frac{1-\alpha}{\vartheta}(\vartheta + (1 - \vartheta)s)^{X_{t-1}}\right] + (1 - p)\mathbf{G}_{\xi_t}(s)$$
$$= p\left(1 - \left(\frac{1-\alpha}{\vartheta}\right)\right)s^{X_{t-1}}$$
$$+ p\left(\frac{1-\alpha}{\vartheta}\right)(\vartheta + (1 - \vartheta)s)^{X_{t-1}} + \lambda\phi e^{-\lambda}(1 - p\alpha)(1 - s)$$

$$+ \left(1 - p \left(1 - \frac{1-\alpha}{\vartheta} \right) \right) e^{\lambda(s-1)} - p \left(\frac{1-\alpha}{\vartheta} \right) e^{\lambda(1-\vartheta)(s-1)}, \qquad (2.13)$$

$$P(X_t = i | X_{t-1} = j) = P(i|j) = p P(\alpha \star_\vartheta X_{t-1} = i | X_{t-1} = j) + (1-p) P(\xi_t = i)$$

$$= p \left(1 - \frac{1-\alpha}{\vartheta} \right) \mathbf{I}_{\{j\}}$$

$$+ p \mathbf{I}_{\{i \le j\}} \frac{1-\alpha}{\vartheta} \binom{j}{i} (1-\vartheta)^i \vartheta^{j-i} + (1-p) P(\xi_t = i), \qquad (2.14)$$

(a) $\alpha = 0.83$, $\vartheta = 0.21$, $\lambda = 0.5$ for $\phi \ge \phi_0$

(b) $\alpha = 0.83$, $\vartheta = 0.21$, $\lambda = 0.5$ for $\phi \le \phi_0$

(c) $\alpha = 0.83$, $\vartheta = 0.21$, $\lambda = 1$ for $\phi \le \phi_0$

(d) $\alpha = 0.83$, $\vartheta = 0.21$, $\lambda = 1$ for $\phi \ge \phi_0$

(e) $\alpha = 0.83$, $\vartheta = 0.21$, $\lambda = 5$ for $\phi \ge \phi_0$

(f) $\alpha = 0.83$, $\vartheta = 0.21$, $\lambda = 5$ for $\phi \le \phi_0$

(g) $\alpha = 0.83$, $\vartheta = 0.21$, $\lambda = 10$ for $\phi \ge \phi_0$

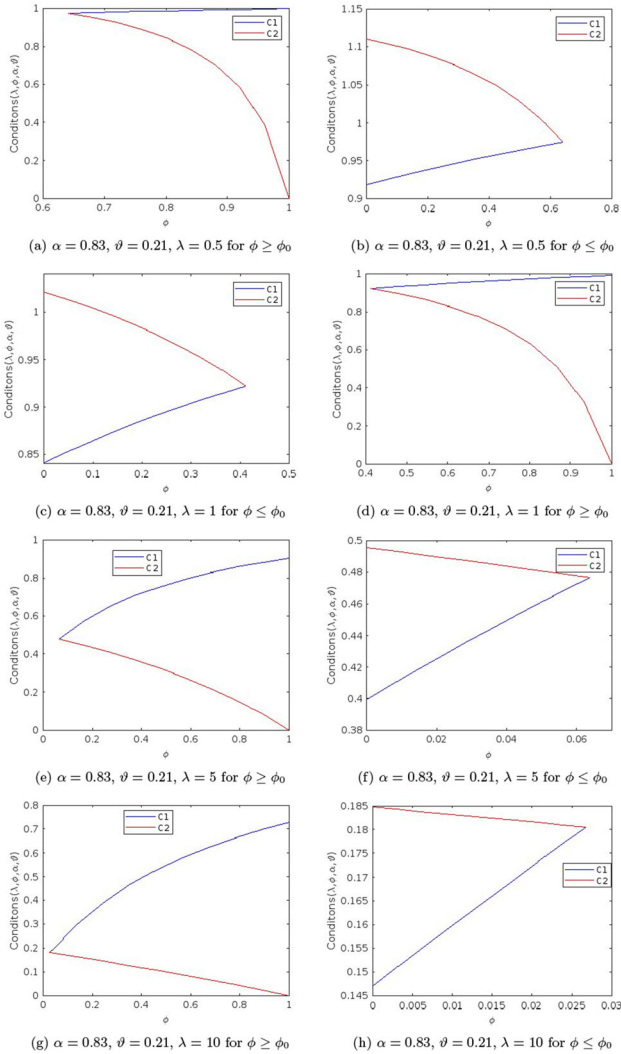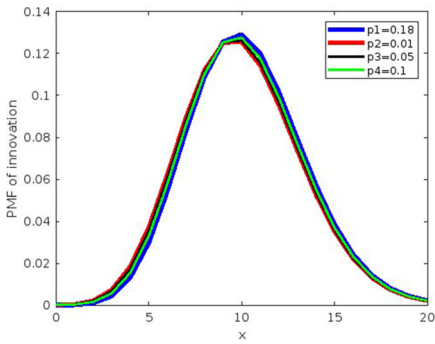(h) $\alpha = 0.83$, $\vartheta = 0.21$, $\lambda = 10$ for $\phi \le \phi_0$
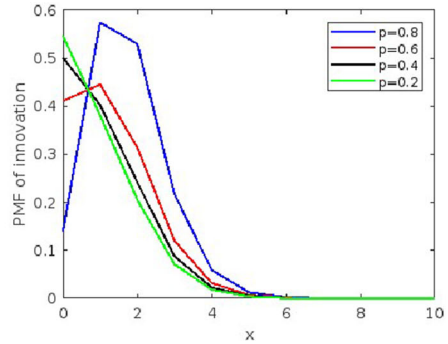
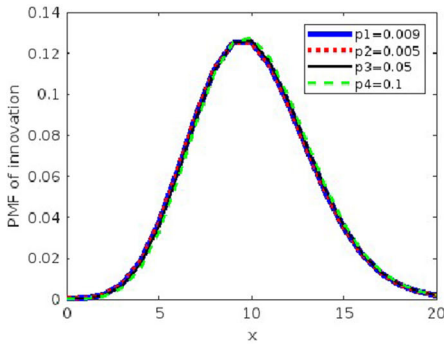**Fig. 4** The conditions C_1 and C_2

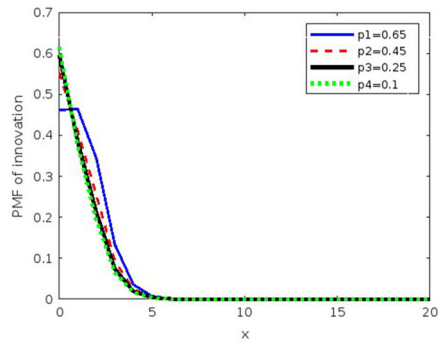(a) $\lambda = 10; \alpha = 0.83; \vartheta = 0.21, \phi = 0.02679988802$ (b) $\lambda = 1; \alpha = 0.6; \vartheta = 0.5, \phi = 0.5596162930$

Fig. 5 PMF of the innovation process at $\phi = \phi_0$, $C_1 = C_2 = C$, and $p \leq C$

(a) $\lambda = 10; \alpha = 0.83; \vartheta = 0.21, \phi = 0.5$ (b) $\lambda = 1; \alpha = 0.6; \vartheta = 0.5; \phi = 0.7$

Fig. 6 PMF of the innovation process at $\phi > \phi_0$ and $p \leq C_2$

where $P(\xi_t = i)$, the $p.m.f.$ of $\{\xi_t\}$ is given by Eq. (2.1), t=0, 1, 2, ..., and $|s| \leq 1$, and $I_{\{j\}}$ is the indicator function with a degenerate random variable defined as

$$I_{\{j\}} = \begin{cases} 1 & X_{t-1} = j \\ 0 & X_{t-1} \neq j. \end{cases} \tag{2.15}$$

Proposition 2.2 *Let $\{X_t\}$ be a stationary OMP-EMPT(1) process, then*

1. *The one-step-ahead conditional expectation is*

$$E(X_{t+1}|X_t) = p\alpha X_t + (1 - p\alpha)\lambda(1 - \phi e^{-\lambda}), \tag{2.16}$$

2. *The k-step-ahead conditional expectation is formulated as*

$$E(X_{t+k}|X_t) = (p\alpha)^k X_t + (1 - (p\alpha)^k)\lambda(1 - \phi e^{-\lambda}). \tag{2.17}$$

3. *The one-step-ahead conditional variance is*

$$\begin{aligned} Var(X_{t+1}|X_t) = &\, p[(2\alpha - \alpha\vartheta + \vartheta - 1) - p\alpha^2]X_t^2 \\ &+ p[(1 - \alpha)(1 - \vartheta) - 2\alpha\lambda(1 - p\alpha)(1 - \phi e^{-\lambda})]X_t \\ &+ ((1 - p\alpha)\lambda(1 - \phi e^{-\lambda})(1 - \lambda(1 - p\alpha)(1 - \phi e^{-\lambda})) \end{aligned}$$

$$+\lambda^2(1 + p + p\alpha(-2 + \vartheta) - p\vartheta). \tag{2.18}$$

4. *The k-step-ahead conditional variance is given as*

$$Var(X_{t+k}|X_t) = (p\mathbf{A_0})^k X_t^2 + p^k \mathbf{B_0} \left(\frac{\mathbf{A_0}^k - \alpha^k}{\mathbf{A_0} - \alpha} \right) X_t$$

$$+ ((1 - p\alpha)\lambda(1 - \phi e^{-\lambda}) + \lambda^2(1 - p\mathbf{A_0})) \left(\frac{1 - (p\mathbf{A_0})^k}{1 - p\mathbf{A_0}} \right)$$

$$+ \left(\frac{(1 - p\alpha)\lambda(1 - \phi e^{-\lambda})}{\mathbf{A_0} - \alpha} \right) \mathbf{B_0} \left(\frac{(p\mathbf{A_0})^k - 1}{p\mathbf{A_0} - 1} - \frac{(p\alpha)^k - 1}{p\alpha - 1} \right)$$

$$- ((p\alpha)^k X_t + (1 - (p\alpha)^k)\lambda(1 - \phi e^{-\lambda}))^2, \tag{2.19}$$

where $\mathbf{A_0} = (2\alpha - \alpha\vartheta + \vartheta - 1)$, $\mathbf{B_0} = (1 - \alpha)(1 - \vartheta)$.

Proof The proof of Proposition 2.2 is given in Appendix A.1. □

The autocovariance, autocorrelation, and partial autocorrelation functions at lag k of $\{X_t\}$ are, respectively, given as follows:

$$\gamma_X(k) = (p\alpha)^k [\lambda^2 + \lambda(1 - \phi e^{-\lambda})(1 - \lambda(1 - \phi e^{-\lambda}))], \tag{2.20}$$

$$\rho_X(k) = (p\alpha)^k, \ k \in \mathbb{Z}, \tag{2.21}$$

$$\beta_X(k) = Corr(X_{k+1} - E(X_{k+1}|X_2, \dots, X_k), X_1)$$
$$= Corr(X_{k+1} - (p\alpha)X_k - (1 - p)\mu_\xi, X_1) = (p\alpha)^k - p\alpha(p\alpha)^{k-1} = 0, \ k > 1, \tag{2.22}$$

$$\beta_X(1) = Corr(X_{k+1}, X_k) = p\alpha. \tag{2.23}$$

Remark 2.4 • The unconditional mean and variance of the OMP-EMPT(1) process are obtained by taking the limit of Eqs. (2.17) and (2.19) at $k \longrightarrow \infty$, that is

$$\lim_{k \to \infty} E(X_{t+k}|X_t) = \lambda(1 - \phi e^{-\lambda}) = E(X_t), \tag{2.24}$$

$$\lim_{k \to \infty} Var(X_{t+k}|X_t) = [(1 - p\alpha)\lambda(1 - \phi e^{-\lambda}) + \lambda^2(1 - p\mathbf{A_0})] \left(\frac{1}{1 - p\mathbf{A_0}} \right)$$

$$+ \left(\frac{(1 - p\alpha)\lambda(1 - \phi e^{-\lambda})}{\mathbf{A_0} - \alpha} \right) \mathbf{B_0} \left(\frac{-1}{p\mathbf{A_0} - 1} - \frac{-1}{p\alpha - 1} \right) - (\lambda(1 - \phi e^{-\lambda}))^2$$

$$= \frac{\lambda(1 - \phi e^{-\lambda})(1 - p\alpha)}{1 - p\mathbf{A_0}} + \lambda^2 - \frac{\lambda(1 - \phi e^{-\lambda})(1 - p\alpha)\mathbf{B_0}}{-\mathbf{B_0}(p\mathbf{A_0} - 1)}$$

$$+ \frac{\lambda(1 - \phi e^{-\lambda})(1 - p\alpha)\mathbf{B_0}}{-\mathbf{B_0}(p\alpha - 1)} - (\lambda(1 - \phi e^{-\lambda}))^2$$

$$= \lambda^2 - (\lambda(1 - \phi e^{-\lambda}))^2 + \frac{(1 - p\alpha)\lambda(1 - \phi e^{-\lambda})}{1 - p\mathbf{A_0}}[1 - 1] + \lambda(1 - \phi e^{-\lambda})$$

$$= \lambda^2 + \lambda(1 - \phi e^{-\lambda})(1 - \lambda(1 - \phi e^{-\lambda})) = Var(X_t). \tag{2.25}$$

• From Eq. (2.19), it is concluded that the $k-$step-ahead conditional variance, $Var(X_{t+k}|X_t)$, depends on X_t^2, that is, quadratically depends on the past of the process, and the coefficient of this dependence is $(p\mathbf{A_0})^k - (p\alpha)^{2k}$. Thus, if $\vartheta \neq 0$ and

$\alpha \in (0, 1)$, which is a general case, then large values of variance could be expected when the process reaches some significant realizations. This goes with the fact that $Var(X_t)$ is approximated by $Var(X_{t+k}|X_t)$ for large values of k. Therefore, OMP-EMPT(1) model provides a very convenient property in the application to over-dispersed data.

- The process is stationary, since the mean, variance, and autocovariance are finite and independent of time.
- The autocorrelation and partial autocorrelation functions can be thought as model specification tools. When $k \longrightarrow \infty$ in Eq. (2.21), the autocorrelation function decays exponentially to zero. Also, we have from Eq. (2.22), and noting that $\beta(1) = p\alpha$, the PACF of OMP-EMPT(1) cuts off after lag 1. As a result, it could be remarked that the OMP-EMPT(1) model is an autoregressive process of order one.

The joint $p.g.f.$ of two consecutive observations of this process, X_t and X_{t-1}, is

$$
\begin{aligned}
\mathbf{G_{X_{t-1}, X_t}}(s_1, s_2) = {}& p\left(1 - \frac{1-\alpha}{\vartheta}\right)[\lambda\phi e^{-\lambda}(1 - s_1 s_2) + e^{\lambda(s_1 s_2 - 1)}] \\
& + p\frac{1-\alpha}{\vartheta}(\lambda\phi e^{-\lambda}(1 - s_1(\vartheta + (1 - \vartheta)s_2)) + e^{\lambda(s_1(\vartheta + (1-\vartheta)s_2) - 1)}) \\
& + (\lambda\phi e^{-\lambda}(1 - s_1) + e^{\lambda(s_1 - 1)})\left[\lambda\phi e^{-\lambda}(1 - p\alpha)(1 - s_2) \right. \\
& \left. + \left(1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)\right)e^{\lambda(s_2 - 1)} - p\left(\frac{1-\alpha}{\vartheta}\right)e^{\lambda(1-\vartheta)(s_2 - 1)}\right].
\end{aligned}
$$
(2.26)

It is clear from Eq. (2.26) that $\mathbf{G_{X_{t-1}, X_t}}(s_1, s_2)$ is not symmetric in s_1 and s_2, implying that the process is not time-reversible, and thus, the joint distribution of (X_{t-1}, X_t) is not equal to the joint distribution of (X_t, X_{t-1}).

The joint probability function of X_{t-1}, X_t is

$$
P_{X_{t-1}, X_t}(i, j) = P(X_t = j|X_{t-1} = i)P(X_{t-1} = i),
$$
(2.27)

where $P(X_t = j|X_{t-1} = i)$ is the one-step transition probability function given by Eq. (2.13) and $P(X_{t-1} = i)$ is the marginal probability distribution defined by Eq.(1.2).

The adequacy or sufficiency of the fitted models can be checked by jumps (see Weiß (2009)). The jump process is also used to construct control charts that show changes in the serial dependence structure. The $p.g.f.$ of the jump process, $J_t \equiv X_t - X_{t-1}$, for $t \geq 2$ is obtained by replacing s_1 and s_2 by u^{-1} and u, respectively, in Eq. (2.26), as

$$
\begin{aligned}
\mathbf{G}_{X_{t-1}, X_t}(u^{-1}, u) = {}& p\left(1 - \frac{1-\alpha}{\vartheta}\right) \\
& + p\frac{1-\alpha}{\vartheta}\left(\lambda\phi e^{-\lambda}\left(1 - \left(\frac{\vartheta}{u} + (1 - \vartheta)\right)\right) + e^{\lambda\left(\left(\frac{\vartheta}{u} + (1-\vartheta)\right) - 1\right)}\right) \\
& + (\lambda\phi e^{-\lambda}(1 - u^{-1}) + e^{\lambda(u^{-1} - 1)})\left[\lambda\phi e^{-\lambda}(1 - p\alpha)(1 - u) \right. \\
& \left. + \left(1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)\right)e^{\lambda(u-1)} - p\left(\frac{1-\alpha}{\vartheta}\right)e^{\lambda(1-\vartheta)(u-1)}\right].
\end{aligned}
$$
(2.28)

Utilizing Eq. (2.28) and the definition of the jump process, it is determined that the mean, variance, autocovariance, and autocorrelation of J_t are, respectively, specified by

$$E(J_t) = E(X_t - X_{t-1}) = 0, \tag{2.29}$$

$$Var(J_t) = Var(X_t - X_{t-1}) = 2Var(X_t) - 2Cov(X_t, X_{t-1})$$
$$= 2(1 - p\alpha)[\lambda^2 + \lambda(1 - \phi e^{-\lambda})(1 - \lambda(1 - \phi e^{-\lambda}))], \tag{2.30}$$

$$\gamma_J(k) = Cov(J_t, J_{t+k})$$
$$= (p\alpha)^{k-1}(2p\alpha - 1 - (p\alpha)^2)[\lambda^2 + \lambda(1 - \phi e^{-\lambda})(1 - \lambda(1 - \phi e^{-\lambda}))], \tag{2.31}$$

$$\rho_J(k) = Corr(J_t, J_{t+k}) = (p\alpha)^{k-1}(2p\alpha - 1 - (p\alpha)^2). \tag{2.32}$$

2.3 Distribution of zeros and ones

The run is defined by Mood (1940), as a succession of similar events that are preceded and succeeded by different events, and the length of the run is determined by its number. The distribution of runs has been extensively studied in the context of $i.i.d.$ set-ups, and a well-known application of runs is the runs test in non-parametric hypothesis testing. Jacobs and Lewis (1978) introduced the distribution of the run length for a stationary discrete autoregressive process of first-order. The expected time of its stay in state i at a stress after entering the process at state i from another state j can be evaluated using the run length. The expected length of zeros and ones will be discussed in two different ways. Let $\{X_t\}$ be the OMP-EMPT(1) process, for a fixed state $i \in \{0, 1, 2, \ldots\}$, suppose $\mathcal{N}_i = \inf\{n \geq 1 : X_n \neq i\} - 1$, $\mathcal{N}_i$ is the run length of state i starting at time 1, where the length can be 0, 1, 2, $\cdots$. Then, the survival function, probability of $\mathcal{N}_i$ greater than or equal to n, of $\mathcal{N}_i$ is

$$P(\mathcal{N}_i \geq n) = P(X_1 = X_2 = \ldots = X_n = i)$$
$$= P(X_1 = i)(P(X_n = i | X_{n-1} = i))^{n-1}$$
$$= P(X_1 = i)(P(i|i))^{n-1}, \ n \geq 1 \tag{2.33}$$

and $P(\mathcal{N}_i = 0) = P(X_1 \neq i) = 1 - P(X_1 = i)$, where $P(X_1 = i)$ and $P(i|i) = P(X_n = i | X_{n-1} = i) = p(1 - \frac{1-\alpha}{\vartheta})I_{\{i\}} + p\frac{1-\alpha}{\vartheta}(1 - \vartheta)^i + (1 - p)P(\xi_n = i)$ are, respectively, the stationary distribution and conditional distribution of the process given by Eqs. (1.3) and (2.13).

Using the stationary and Markov properties of the proposed process, and based on the expression of the survival function of $\mathcal{N}_i$ given by Eq.(2.33), the $p.m.f.$ of $\mathcal{N}_i$ is obtained as

$$P(\mathcal{N}_i = n) = P(\mathcal{N}_i \geq n) - P(\mathcal{N}_i \geq n + 1)$$
$$= P(X_{n+1} \neq i, X_n = i, X_{n-1} = i, \ldots, X_1 = i)$$
$$= P(X_{n+1} \neq i | X_n = i)P(X_n = i | X_{n-1} = i) \ldots P(X_2 = i | X_1 = i)P(X_1 = i)$$
$$= P(X_1 = i)[1 - P(X_{n+1} = i | X_n = i)](P(X_n = i | X_{n-1} = i))^{n-1}$$
$$= -P(X_1 = i)(P(X_{n+1} = i | X_n = i))^n$$
$$= (1 - P(X_{n+1} = i | X_n = i))$$
$$= P(X_1 = i)[1 - P(i|i)](P(i|i))^{n-1}, n \geq 1. \tag{2.34}$$

Therefore, based on the previous expression for the p.m.f. of $\mathcal{N}_i$ given by Eq. (2.34), the expected run length of state i is given by

$$
\begin{aligned}
E(\mathcal{N}_i = n) &= \sum_{n=0}^{\infty} n P(\mathcal{N}_i = n) \\
&= \sum_{n=1}^{\infty} n P(X_1 = i)[1 - P(i|i)](P(i|i))^{n-1} \\
&= P(X_1 = i)[1 - P(i|i)] \sum_{n=1}^{\infty} n(P(i|i))^{n-1} \\
&= P(X_1 = i)[1 - P(i|i)] \left(\frac{1}{[1 - P(i|i)]^2} \right) \\
&= \frac{P(X_1 = i)}{1 - P(i|i)}.
\end{aligned}
\tag{2.35}
$$

The $p.g.f.$ of $\mathcal{N}_i$ is

$$
\begin{aligned}
\mathbf{G}_{\mathcal{N}_i}(s) &= \sum_{n=0}^{\infty} s^n P(\mathcal{N}_i = n) \\
&= P(\mathcal{N}_i = 0) + \sum_{n=1}^{\infty} s^n p(\mathcal{N}_i = n) \\
&= (1 - P(X_1 = i)) + P(X_1 = i) \sum_{n=1}^{\infty} s^n \\
&\quad \times [(1 - P(X_n = i|X_{n-1} = i))(P(X_n = i|X_{n-1} = i))^{n-1}] \\
&= (1 - P(X_1 = i)) + \frac{s(1 - P(i|i))P(X_1 = i)}{1 - sP(i|i)}.
\end{aligned}
\tag{2.36}
$$

For the OMP-EMPT(1) process, the transition probability functions from zero to zero and non-zero are, respectively, given by

$$
\begin{aligned}
P(X_n = 0|X_{n-1} = 0) &= p\left(1 - \frac{1-\alpha}{\vartheta}\right) I_{\{0\}} + p\frac{1-\alpha}{\vartheta} + \lambda\phi e^{-\lambda}(1 - p\alpha) \\
&\quad - p\frac{1-\alpha}{\vartheta} e^{-\lambda(1-\vartheta)} + e^{-\lambda}\left(1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)\right),
\end{aligned}
\tag{2.37}
$$

and

$$
\delta_0 = P(X_n \neq 0|X_{n-1} = 0) = 1 - P(X_n = 0|X_{n-1} = 0).
\tag{2.38}
$$

Also, the transition probability functions from one to one and non-one are, respectively, given by

$$
\begin{aligned}
P(X_n = 1|X_{n-1} = 1) &= p\left(1 - \frac{1-\alpha}{\vartheta}\right) I_{\{1\}} + p\frac{1-\alpha}{\vartheta}(1 - \vartheta) - \lambda\phi e^{-\lambda}(1 - p\alpha) \\
&\quad - p\frac{1-\alpha}{\vartheta} e^{-\lambda(1-\vartheta)}\lambda(1-\vartheta) + \lambda e^{-\lambda}\left(1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)\right),
\end{aligned}
\tag{2.39}
$$

and

$$\eta_1 = P(X_n \neq 1 | X_{n-1} = 1) = 1 - P(X_n = 1 | X_{n-1} = 1). \tag{2.40}$$

Using Eq. (2.35), the expected length of the runs of zeros and ones, the number of zeros between two non-zero values, and the number of ones between two non-one values of the process can be, respectively, derived as

$$E(\mathcal{N}_0) = \frac{P(X_1 = 0)}{1 - P(X_n = 0 | X_{n-1} = 0)} = \frac{e^{-\lambda}(1 + \lambda\phi)}{P(X_n \neq 0 | X_{n-1} = 0)},$$

$$E(\mathcal{N}_0) = \frac{e^{-\lambda}(1 + \lambda\phi)}{1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right) I_{\{0\}} - p\frac{1-\alpha}{\vartheta} - \lambda\phi e^{-\lambda}(1 - p\alpha) + p\frac{1-\alpha}{\vartheta}e^{-\lambda(1-\vartheta)}}{[-e^{-\lambda}\left(1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)\right)]}. \tag{2.41}$$

$$E(\mathcal{N}_1) = \frac{P(X_n = 1)}{1 - P(X_n = 1 | X_{n-1} = 1)} = \frac{\lambda e^{-\lambda}(1 - \phi)}{P(X_n \neq 1 | X_{n-1} = 1)},$$

$$E(\mathcal{N}_1) = \frac{\lambda e^{-\lambda}(1 - \phi)}{1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right) I_{\{1\}} - p\frac{1-\alpha}{\vartheta}(1 - \vartheta) + \lambda\phi e^{-\lambda}(1 - p\alpha)}{[+ p\frac{1-\alpha}{\vartheta}e^{-\lambda(1-\vartheta)}\lambda(1 - \vartheta) - \lambda e^{-\lambda}\left(1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)\right)]}. \tag{2.42}$$

Using Eq. (2.33), the survival functions of $\mathcal{N}_0$ and $\mathcal{N}_1$ are, respectively, given by

$$P(\mathcal{N}_0 \geq n) = P(X_1 = 0)(P(X_n = 0 | X_{n-1} = 0))^{n-1}, \; n \geq 1 \tag{2.43}$$

and

$$P(\mathcal{N}_1 \geq n) = P(X_1 = 1)(P(X_n = 1 | X_{n-1} = 1))^{n-1}, \; n \geq 1. \tag{2.44}$$

Using the stationarity of the process, as in Barreto- Souza W (2015), and Eq. (2.34), the *p.m.f.* for the number of zeros and ones is given as follows, respectively:

$$\begin{aligned} P(\mathcal{N}_0 = n) &= P(X_n \neq 0, X_{n-1} = 0, \ldots, X_1 = 0 | X_0 = 0) \\ &= P(X_n \neq 0 | X_{n-1} = 0)P(X_{n-1} = 0 | X_{n-2} = 0) \ldots P(X_1 = 0 | X_0 = 0) \\ &= \delta_0(1 - \delta_0)^{n-1}, \end{aligned}$$

for $n \geq 2$ and $P(\mathcal{N}_0 = 1) = P(X_1 \neq 0 | X_0 = 0) = \delta_0$.

$$\begin{aligned} P(\mathcal{N}_1 = n) &= P(X_n \neq 1, X_{n-1} = 1, \ldots, X_1 = 1 | X_0 = 1) \\ &= P(X_n \neq 1 | X_{n-1} = 1)P(X_{n-1} = 1 | X_{n-2} = 1) \ldots P(X_1 = 1 | X_0 = 1) \\ &= \eta_1(1 - \eta_1)^{n-1}, \end{aligned}$$

for $n \geq 2$ and $P(\mathcal{N}_1 = 1) = P(X_1 \neq 1 | X_0 = 1) = \eta_1$. Based on Eq. (2.36), the *p.g.f.* for $\mathcal{N}_0$ and $\mathcal{N}_1$ are, respectively, given by

$$\mathbf{G}_{\mathcal{N}_0}(s) = (1 - P(X_1 = 0)) + \frac{s(1 - P(0|0))P(X_1 = 0)}{1 - sP(0|0)}, \tag{2.45}$$

and

$$\mathbf{G}_{\mathcal{N}_1}(s) = (1 - P(X_1 = 1)) + \frac{s(1 - P(1|1))P(X_1 = 1)}{1 - sP(1|1)}. \tag{2.46}$$

Remark 2.5 1. The expected run length of zeros and ones in the process, as in Jazi et al. (2012) and Barreto-Souza (2015), can be determined in a different way, because the run length of zeros and ones has a geometric distribution with probabilities δ_0 and η_1, that is, $P(\mathcal{N}_0 = n) = \delta_0(1 - \delta_0)^{n-1}, n \geq 1$, where δ_0 is given by Eq. (2.38) and $P(\mathcal{N}_1 = n) = \eta_1(1 - \eta_1)^{n-1}, n \geq 1$, where η_1 is given by Eq. (2.40), respectively. The expected run length zeros and ones are calculated as

$$E(\mathcal{N}_0) = \frac{1}{\begin{array}{c} 1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)I_{\{0\}} - p\frac{1-\alpha}{\vartheta} - \lambda\phi e^{-\lambda}(1 - p\alpha) \\ [+ p\frac{1-\alpha}{\vartheta}e^{-\lambda(1-\vartheta)} - e^{-\lambda}\left(1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)\right)] \end{array}}, \tag{2.47}$$

and

$$E(\mathcal{N}_1) = \frac{1}{\begin{array}{c} 1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)I_{\{1\}} - p\frac{1-\alpha}{\vartheta}(1 - \vartheta) + \lambda\phi e^{-\lambda}(1 - p\alpha) \\ [+ p\frac{1-\alpha}{\vartheta}e^{-\lambda(1-\vartheta)}\lambda(1 - \vartheta) - \lambda e^{-\lambda}\left(1 - p\left(1 - \frac{1-\alpha}{\vartheta}\right)\right)] \end{array}}. \tag{2.48}$$

2. The expected run length for the MPT(1) with Poisson marginal is obtained by setting $\phi = 0$ and $1 - \alpha = \vartheta$ in Eqs. (2.41 and 2.42) or Eqs. (2.47 and 2.48).

3 Estimation of parameters and a simulation study

In this section, the unknown parameters of the OMP-EMPT(1) model will be estimated based on the maximum-likelihood (ML), modified Yule–Walker (MYW), and modified conditional least-squares (MCLS) methods. Suppose that $\Theta = (\alpha, \vartheta, \phi, \lambda, p)^T$ is the vector of parameters that will be estimated using a realization $X_1, X_2, \ldots, X_n$ from the process, where n is the sample size.

3.1 Maximum-likelihood estimation

The considered maximum-likelihood (ML) estimators take into account the temporal dependence between time series ordered records. The ML function of the considered realization is given by

$$L(\Theta; x) = P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \prod_{j=2}^{n} P(X_j = x_j | X_{j-1} = x_{j-1}).$$

The corresponding log-likelihood function can be written as

$$l(\Theta; x) = \log L(\Theta; x) = \log p_{x_1} + \sum_{j=2}^{n} \log P(X_j = x_j | X_{j-1} = x_{j-1}), \tag{3.1}$$

where $P(X_j = x_j | X_{j-1} = x_{j-1})$ is the one-step transition probability given by Eq. (2.13) and p_{x_1} is given by Eq. (1.3). The maximum-likelihood (ML) estimator of Θ is the value $\hat{\Theta}_{ML} = (\hat{\alpha}_{ML}, \hat{\vartheta}_{ML}, \hat{\phi}_{ML}, \hat{\lambda}_{ML}, \hat{p}_{ML})^T$, which maximizes the log-likelihood function in Eq. (3.1). The ML estimators are computed using the numerical methods found in most statistical software to solve the nonlinear system obtained from $\frac{\partial l(\Theta;x)}{\partial \Theta} = 0$. To maximize the log-likelihood function of the proposed process, the optim function in R software is used.

3.2 Modified Yule–Walker estimation

The Modified Yule–Walker (MYW) estimators are computed via a combination of Yule–Walker estimators and ML estimators that have been introduced by Miletić Ilić et al. (2018). The MYW estimators of the parameters $\alpha, p, \lambda, \phi, \theta$ are based on the sample values of $\rho_X(1), E(X_t^2) - E(X_t), E(X_t), Cov(X_t^2, X_{t-1})$. Recall that $\rho_X(1) = Corr(X_t, X_{t-1}) = p\alpha$, $E(X_t) = \lambda(1 - \phi e^{-\lambda})$ and $E(X_t^2) = \lambda^2 + \lambda(1 - \phi e^{-\lambda})$. Using the earlier sample values, the estimators for α, p, λ, and ϕ are obtained as

$$\widehat{\lambda}_{YW} = \sqrt{\frac{\sum_{t=1}^n X_t^2 - \sum_{t=1}^n X_t}{n}},$$

$$\widehat{\phi}_{YW} = e^{\widehat{\lambda}_{YW}} \left(1 - \frac{\overline{X}}{\widehat{\lambda}_{YW}}\right),$$

$$\widehat{\alpha}_{YW} = \widehat{\rho}_X(1) = \frac{\sum_{t=2}^n (X_t - \overline{X})(X_{t-1} - \overline{X})}{\widehat{p}_{ML} \sum_{t=1}^n (X_t - \overline{X})^2},$$

where $\widehat{p}_{ML}$ is the ML estimator of p and it is used as the initial estimator of the parameter. Then, the YW estimator of p is given by

$$\widehat{p}_{YW} = \frac{\sum_{t=2}^n (X_t - \overline{X})(X_{t-1} - \overline{X})}{\widehat{\alpha}_{YW} \sum_{t=1}^n (X_t - \overline{X})^2}.$$

To estimate the parameter ϑ, first, the $Cov(X_t^2, X_{t-1})$ is computed

$$\begin{aligned} Cov(X_t^2, X_{t-1}) &= pCov((\alpha \star_\vartheta X_{t-1})^2, X_{t-1}) + (1 - p)Cov(\xi_t^2, X_{t-1}) \\ &= p(2\alpha - \alpha\vartheta + \vartheta - 1)[E(X_{t-1}^3) - \mu_X E(X_{t-1}^2)] \\ &\quad + p(1 - \alpha)(1 - \vartheta)Var(X_{t-1}), \end{aligned}$$

using Eqs. (1.8), (1.7), and (1.6) for $E(x_{t-1}^3), E(X_{t-1}^2)$ and $Var(X_t)$, then

$$\begin{aligned} Cov(X_t^2, X_{t-1}) &= p(2\alpha - 1)(\lambda^3 + 3\lambda^2 + \lambda(1 - \phi e^{-\lambda})(1 - \lambda^2 - \lambda(1 - \phi e^{-\lambda}))) \\ &\quad + p(1 - \alpha)[\lambda^2 + \lambda(1 - \phi e^{-\lambda})(1 - \lambda(1 - \phi e^{-\lambda})) \\ &\quad + \lambda^2(2 + \lambda\phi e^{-\lambda})p\vartheta(1 - \alpha). \end{aligned}$$

By solving the previous equation to ϑ, then the YW estimator of the parameter ϑ is obtained as

$$\widehat{\vartheta}_{YW} = \frac{\widehat{C} - \widehat{p}_{YW}[(2\widehat{\alpha}_{YW} - 1)\widehat{M}_1 + (1 - \widehat{\alpha}_{YW})\widehat{V}_1]}{\widehat{p}_{YW}(1 - \widehat{\alpha}_{YW})\widehat{\lambda}_{YW}^2(2 + \widehat{\lambda}_{YW}\widehat{\phi}_{YW}e^{-\widehat{\lambda}_{YW}})},$$

where $\widehat{M}_1$ and $\widehat{V}_1$ are defined as

$$\widehat{M}_1 = \widehat{\lambda}_{YW}^3 + 3\widehat{\lambda}_{YW}^2 + \widehat{\lambda}_{YW}(1 - \widehat{\phi}_{YW}e^{-\widehat{\lambda}_{YW}})(1 - \widehat{\lambda}_{YW}^2 - \widehat{\lambda}_{YW}(1 - \widehat{\phi}_{YW}e^{-\widehat{\lambda}_{YW}})),$$

$$\widehat{V}_1 = \widehat{\lambda}_{YW}^2 + \widehat{\lambda}_{YW}(1 - \widehat{\phi}_{YW}e^{-\widehat{\lambda}_{YW}})(1 - \widehat{\lambda}_{YW}(1 - \widehat{\phi}_{YW}e^{-\widehat{\lambda}_{YW}})),$$

and $\widehat{C}$ is the estimator of $Cov(X_t^2, X_{t-1})$ given by

$$\widehat{C} = \frac{1}{n-1} \sum_{t=2}^n X_t^2 X_{t-1} - \left(\frac{1}{n-1} \sum_{t=2}^n X_t^2\right)\left(\frac{1}{n-1} \sum_{t=2}^n X_{t-1}\right).$$

3.3 Modified conditional least-squares estimation

In this section, the unknown parameters of the proposed model are estimated by modified conditional least-squares (MCLS). The MCLS is presented by Miletić Ilić et al. (2018) and Karlsen and TjøStheim (1988) to estimate the parameters, and it is based on a combination of CLS and ML estimators. The parameters α, ϑ, p, ϕ, and λ are estimated by the two-step CLS method. The first step is estimating the parameters α, p, ϕ, and, λ by minimizing the function $Q_n(\alpha, \ p, \ \phi, \ \lambda)$

$$
\begin{aligned}
Q_n(\Theta) &= \sum_{t=2}^{n}(X_t - E(X_t|X_{t-1}))^2 \\
&= \sum_{t=2}^{n}(X_t - p\alpha X_{t-1} - (1 - p\alpha)\lambda(1 - \phi e^{-\lambda}))^2;
\end{aligned}
$$

the estimators are then obtained by solving the system of equations: $\frac{\partial Q_n(\Theta)}{\partial (\alpha, \ p, \ \phi, \ \lambda)^T} = 0$ and combining the CLS method with the ML approach. Because the parameters α and p cannot be separable and because when the Q_n function is differentiated with respect to ϕ and λ, they lead to the same estimated equation, the method of moments is used to estimate one of these parameters. With this, we use the CLS to derive the estimators of α, ϕ, and p and suppose that λ will be estimated using the method of moments; also, we have that $\frac{\partial Q_n(\Theta)}{\partial (\alpha, \ p, \ \phi)^T} = 0$, then the estimators of α, λ, p and ϕ are, respectively, given by

$$
\widehat{\alpha}_{\text{cls}} = \frac{\sum_{t=2}^{n} X_t X_{t-1} - \frac{1}{n-1} \sum_{t=2}^{n} X_t \sum_{t=2}^{n} X_{t-1}}{\widehat{p}_{\text{ML}} \left(\sum_{t=2}^{n} X_{t-1}^2 - \frac{1}{n-1} \left(\sum_{t=2}^{n} X_{t-1} \right)^2 \right)}, \tag{3.2}
$$

$$
\widehat{\lambda}_{\text{cls}} = \widehat{\lambda}_{\text{mm}} = \sqrt{\frac{\sum_{t=1}^{n} X_t^2 - \sum_{t=1}^{n} X_t}{n}}, \tag{3.3}
$$

where $\widehat{p}_{\text{ML}}$ is the ML estimator of the parameter p used as the initial estimate of the parameter. Thus, the $\widehat{p}_{\text{cls}}$ is given by

$$
\widehat{p}_{\text{cls}} = \frac{\sum_{t=2}^{n} X_t X_{t-1} - \frac{1}{n-1} \sum_{t=2}^{n} X_t \sum_{t=2}^{n} X_{t-1}}{\widehat{\alpha}_{\text{cls}} \left(\sum_{t=2}^{n} X_{t-1}^2 - \frac{1}{n-1} \left(\sum_{t=2}^{n} X_{t-1} \right)^2 \right)}.
$$

Consequently, we get

$$
\widehat{\phi}_{\text{cls}} = e^{\widehat{\lambda}_{\text{cls}}} \left(1 - \left(\frac{\sum_{t=2}^{n} X_t - \widehat{\alpha}_{\text{cls}} \widehat{p}_{\text{cls}} \sum_{t=2}^{n} X_{t-1}}{(n-1)\widehat{\lambda}_{\text{cls}}(1 - \widehat{p}_{\text{cls}}\widehat{\alpha}_{\text{cls}})} \right) \right). \tag{3.4}
$$

The calculations for the expressions in Eqs. (3.2), (3.3), and (3.4) are provided in Appendix A.2.

In the second step, the CLS estimator of the parameter ϑ is obtained using the method introduced by Karlsen and TjøStheim (1988) by minimizing the function

$$
S_n(\vartheta) = \sum_{t=2}^{n}((X_t - E(X_t|X_{t-1}))^2 - Var(X_t|X_{t-1}))^2,
$$

where $E(X_t|X_{t-1})$ and $Var(X_t|X_{t-1})$ are given by (2.16) and (2.18), respectively. Therefore, by solving the equation $\frac{\partial S_n(\vartheta)}{\partial \vartheta} = 0$, the estimator of ϑ is given by

$$\widehat{\vartheta}_{\text{cls}} = \frac{\sum_{t=2}^{n}(X_{t-1}^2 - X_{t-1} - \lambda_{\text{cls}}^2)(\widehat{\mathcal{V}}_t - \widehat{\mathcal{H}}_{t-1})}{(1 - \widehat{\alpha}_{\text{cls}})\widehat{p}_{\text{cls}}\sum_{t=2}^{n}(X_{t-1}^2 - X_{t-1} - \lambda_{\text{cls}}^2)^2},$$

where

$$\widehat{\mathcal{V}}_t = (X_t - \widehat{p}_{\text{cls}}\widehat{\alpha}_{\text{cls}}X_{t-1} - (1 - \widehat{p}_{\text{cls}}\widehat{\alpha}_{\text{cls}})\widehat{\lambda}_{\text{cls}}(1 - \widehat{\phi}_{\text{cls}}e^{-\widehat{\lambda}_{\text{cls}}}))^2,$$

and

$$
\begin{aligned}
\widehat{\mathcal{H}}_{t-1} = &[(2\widehat{\alpha}_{\text{cls}} - 1)\widehat{p}_{\text{cls}} - \widehat{p}_{\text{cls}}^2\widehat{\alpha}_{\text{cls}}^2]X_{t-1}^2 \\
&+ (\widehat{p}_{\text{cls}}(1 - \widehat{\alpha}_{\text{cls}}) - 2\widehat{\alpha}_{\text{cls}}\widehat{p}_{\text{cls}}(1 - \widehat{p}_{\text{cls}}\widehat{\alpha}_{\text{cls}})\widehat{\lambda}_{\text{cls}}(1 - \widehat{\phi}_{\text{cls}}e^{-\widehat{\lambda}_{\text{cls}}}))X_{t-1} \\
&+ (1 - \widehat{p}_{\text{cls}}\widehat{\alpha}_{\text{cls}})\widehat{\lambda}_{\text{cls}}(1 - \widehat{\phi}_{\text{cls}}e^{-\widehat{\lambda}_{\text{cls}}})(1 - (1 - \widehat{p}_{\text{cls}}\widehat{\alpha}_{\text{cls}})\widehat{\lambda}_{\text{cls}}(1 - \widehat{\phi}_{\text{cls}}e^{-\widehat{\lambda}_{\text{cls}}})) \\
&+ \widehat{\lambda}_{\text{cls}}^2(1 + \widehat{p}_{\text{cls}} - 2\widehat{p}_{\text{cls}}\widehat{\alpha}_{\text{cls}}).
\end{aligned}
$$

Remark 3.1 The OMP-EMPT(1) process is an irreducible Feller chain, stationary, strictly stationary, and ergodic. The proof of these properties is similar to the ones in Khoo et al. (2017), Shirozhan and Mohammadpour (2018), Nastić et al. (2012), and Schweer and Weiß (2014).

3.4 Simulation

Some simulated trajectories of the OMP-EMPT(1) process for sample size $n = 100$ and $\lambda = 0.5, 1.5, 5$ with $\alpha = 0.5$; $\vartheta = 0.6$; $p = 0.2$; $p = 0.02$; $\phi = 0.5$, and $\alpha = 0.7$; $\vartheta = 0.3$; $p = 0.1$; $\phi = 0.7$ are presented in Fig. 7. From these simulated series, one can conclude that the proposed process is stationary, and for lower values of λ, the number of ones and zeros increases. When λ is large, the number of zeros and ones decreases, resulting in larger values for the sample path. The details of the simulations and how the innovation process ξ_t has been simulated are shown by the R code for simulated series at $n = 100$; $\alpha = 0.5$; $\vartheta = 0.6$; $p = 0.2$; $\lambda = 1.5$; $\phi = 0.5$; see Appendix A.3.

A Monte Carlo simulation study is performed for different parameter values and different sample sizes to evaluate the performance of the ML, MYW, and MCLS estimators. The mean square errors (MSE) and empirical biases were calculated over 100 replications with the sample sizes $n = 100, 500, 1000, 5000$ to evaluate the performance of the estimates and check their properties. The true values of the parameters in the simulations are: $(\alpha, \vartheta, \phi, \lambda, p) = (0.7, 0.3, 0.01, 1, 0.1)$, $(0.4, 0.6, 0.1, 1, 0.1)$, $(0.5, 0.6, 0.2, 0.5, 0.5)$, and $(0.5, 0.6, 0.2, 1, 0.5)$. Using this simulation, Tables 1 and 2 show the bias and MSE (in brackets) of the estimates of different parameter values ($\alpha, \vartheta, \phi, \lambda, p$) and through different sample sizes. From these tables, we conclude that the MSEs and biases are reduced by increasing the sample sizes for the three methods. The estimation using using the ML approach is better than MYW and MCLS. The ML-estimated parameters converged to their true values faster. Furthermore, we can conclude that the MLEs have the smallest MSE for all parameters and thus give the best performance, as expected.

Table 1 Simulated values of biases (MSEs within parentheses) of estimates of $\hat{\alpha}$, $\hat{\vartheta}$, $\hat{\phi}$, $\hat{\lambda}$, and p

n	$\hat{\alpha}_{ML}$	$\hat{\vartheta}_{ML}$	$\hat{\phi}_{ML}$	$\hat{\lambda}_{ML}$	p_{ML}	$\hat{\alpha}_{cls}$	$\hat{\vartheta}_{cls}$	$\hat{\phi}_{cls}$	$\hat{\lambda}_{cls}$	p_{cls}
$(\alpha, \vartheta, \phi, \lambda, p) = (0.7, 0.3, 0.01, 1, 0.1)$										
100	−0.053407	0.003968	0.027017	−0.015913	0.01821	−0.367142	−0.035956	0.085526	−0.008885	0.017038
	(0.074297)	(0.038052)	(0.004228)	(0.010594)	(0.005666)	(0.156224)	(0.022528)	(0.010502)	(0.01017)	(0.000293)
500	0.032668	0.016307	0.006962	−0.014425	0.00176	−0.273695	−0.017516	0.057925	−0.007375	0.000752
	(0.010141)	(0.022565)	(0.000567)	(0.001751)	(0.001305)	(0.083925)	(0.019427)	(0.005899)	(0.00227)	(0.000001)
1000	0.033815	−0.01825	0.007524	−0.012575	0.001685	−0.234406	−0.027259	0.042393	−0.021483	0.000678
	(0.006011)	(0.016816)	(0.000507)	(0.000915)	(0.000604)	(0.062644)	(0.017988)	(0.003376)	(0.001805)	(0.0000005)
5000	0.015763	−0.010174	0.001841	−0.01939	0.001619	−0.204076	0.035708	0.022496	−0.026088	0.000613
	(0.001346)	(0.00578)	(0.000087)	(0.000824)	(0.000132)	(0.044648)	(0.008152)	(0.000881)	(0.000967)	(0.0000004)
$(\alpha, \vartheta, \phi, \lambda, p) = (0.4, 0.6, 0.1, 1, 0.1)$										
100	0.000894	−0.036423	−0.020198	−0.046639	0.039845	−0.131435	−0.120193	0.033239	−0.037957	0.038457
	(0.029456)	(0.066274)	(0.003799)	(0.012531)	(0.008435)	(0.038191)	(0.062566)	(0.007136)	(0.016314)	(0.001494)
500	0.044806	0.005047	−0.018504	−0.032659	−0.001105	−0.042855	−0.117664	0.017472	−0.048667	−0.002084
	(0.013979)	(0.025583)	(0.001507)	(0.003611)	(0.002375)	(0.017631)	(0.058139)	(0.004755)	(0.005535)	(0.000004)
1000	0.013079	0.003895	−0.015886	−0.030699	0.011169	−0.040213	−0.110037	0.002621	−0.058378	0.010067
	(0.008784)	(0.018462)	(0.001192)	(0.003039)	(0.002014)	(0.010811)	(0.056063)	(0.003209)	(0.004916)	(0.000102)
5000	0.00394	0.007916	−0.012316	−0.029339	0.004899	0.001275	−0.095387	−0.005101	−0.06654	0.00386
	(0.005525)	(0.00788)	(0.000511)	(0.002258)	(0.000822)	(0.006072)	(0.052663)	(0.000996)	(0.004828)	(0.000015)

Table 1 continued

n	$\hat{\alpha}_{ML}$	$\hat{\vartheta}_{ML}$	$\hat{\phi}_{ML}$	$\hat{\lambda}_{ML}$	p_{ML}	$\hat{\alpha}_{cls}$	$\hat{\vartheta}_{cls}$	$\hat{\phi}_{cls}$	$\hat{\lambda}_{cls}$	p_{cls}
$(\alpha, \vartheta, \phi, \lambda, p) = (0.5, 0.6, 0.2, 0.5, 0.5)$										
100	0.032848	−0.043902	−0.024204	−0.154267	−0.035758	0.010441	−0.036302	0.00972	−0.114516	−0.04035
	(0.02574)	(0.077284)	(0.022594)	(0.033907)	(0.026414)	(0.0101173)	(0.038181)	(0.010563)	(0.020216)	(0.001645)
500	0.009337	−0.059336	−0.004952	−0.125749	0.005523	0.05223	−0.033697	0.005468	−0.133832	0.000518
	(0.011681)	(0.055142)	(0.007804)	(0.020988)	(0.011991)	(0.00772)	(0.030982)	(0.009661)	(0.019732)	(2.71×10^{-7})
1000	0.01048	−0.041435	−0.00761	−0.123168	−0.00128	0.043336	0.011288	0.02813	−0.131905	−0.006217
	(0.009163)	(0.045158)	(0.004666)	(0.018687)	(0.010773)	(0.005255)	(0.018767)	(0.0081)	(0.018605)	(0.000039)
5000	0.005361	−0.009257	0.008447	−0.11011	0.001405	0.027238	0.119611	0.038336	−0.133644	−0.003559
	(0.005499)	(0.015634)	(0.001125)	(0.016097)	(0.005826)	(0.001894)	(0.017971)	(0.002909)	(0.018344)	(0.000013)
$(\alpha, \vartheta, \phi, \lambda, p) = (0.5, 0.6, 0.2, 1, 0.5)$										
100	0.027581	−0.037458	−0.005995	−0.282658	0.036159	−0.084988	0.068232	0.034416	−0.321495	0.030847
	(0.015383)	(0.054098)	(0.013674)	(0.104375)	(0.019093)	(0.013128)	(0.013048)	(0.013197)	(0.1217897)	(0.000961)
500	0.01382	−0.030213	0.04267	−0.278687	0.013368	−0.0195597	0.020238	0.084163	−0.3034194	0.008284
	(0.011068)	(0.025956)	(0.005511)	(0.089124)	(0.012011)	(0.003551)	(0.010142)	(0.011181)	(0.096279)	(0.0000693)
1000	−0.004084	−0.0039	0.050537	−0.271822	0.010956	−0.023360	0.042026	0.071138	−0.302077	0.005896
	(0.007861)	(0.01677)	(0.00548)	(0.086306)	(0.008446)	(0.002603)	(0.009519)	(0.006577)	(0.093634)	(0.0000351)
5000	0.011665	−0.018246	0.056114	−0.265964	−0.008747	0.0033357	0.0012461	0.0871569	−0.2997216	−0.0136095
	(0.002222)	(0.004406)	(0.004152)	(0.082517)	(0.002037)	(0.0009155)	(0.0055597)	(0.0077997)	(0.0910393)	(0.0001871)

Table 2 Simulated values of biases (MSEs within parentheses) of estimates of α, ϑ, ϕ, λ, and p based on MYW

n	$\hat{\alpha}_{yw}$	$\hat{\vartheta}_{yw}$	$\hat{\phi}_{yw}$	$\hat{\lambda}_{yw}$	$\hat{p}_{yw}$
$(\alpha, \vartheta, \phi, \lambda, p) = (0.7, 0.3, 0.01, 1, 0.1)$					
100	−0.37002	−0.046511	0.084134	−0.008885	0.017038
	(0.158232)	(0.022348)	(0.010251)	(0.01017)	(0.000293)
500	−0.240772	−0.027209	0.064304	−0.010018	0.000752
	(0.072056)	(0.019636)	(0.007122)	(0.003196)	(0.000001)
1000	−0.234799	−0.014381	0.042265	−0.021483	0.000678
	(0.062821)	(0.020029)	(0.003366)	(0.001805)	(0.000000465)
5000	−0.181715	0.068443	0.021291	−0.025767	0.000613
	(0.035806)	(0.013168)	(0.0009997)	(0.0009548)	(0.000000379)
$(\alpha, \vartheta, \phi, \lambda, p) = (0.4, 0.6, 0.1, 1, 0.1)$					
100	−0.138074	−0.243424	0.032166	−0.039507	0.038457
	(0.050377)	(0.105396)	(0.00858)	(0.012214)	(0.001494)
500	−0.064423	−0.18819	0.015093	−0.053533	−0.002084
	(0.01896)	(0.085177)	(0.005456)	(0.006025)	(0.000004)
1000	−0.0238	−0.114106	0.003074	−0.061914	0.010067
	(0.012097)	(0.047538)	(0.002638)	(0.005368)	(0.000102)
5000	0.00118	−0.011227	−0.005158	−0.06654	0.00386
	(0.00607)	(0.014182)	(0.000997)	(0.004828)	(0.000015)
$(\alpha, \vartheta, \phi, \lambda, p) = (0.5, 0.6, 0.2, 0.5, 0.5)$					
100	0.005104	−0.25892	0.025346	−0.114516	−0.04035
	(0.010113)	(0.113075)	(0.011025)	(0.020216)	(0.001645)
500	0.051259	−0.175498	0.008877	−0.133832	0.000518
	(0.007659)	(0.067743)	(0.00966)	(0.019732)	(2.704×10^{-7})
1000	0.042842	−0.144442	0.02978	−0.131905	−0.006217
	(0.005217)	(0.041903)	(0.008159)	(0.018605)	(0.000039)
5000	0.025393	−0.056939	0.035407	−0.132658	−0.003559
	(0.001301)	(0.011009)	(0.00278)	(0.017994)	(0.000013)
$(\alpha, \vartheta, \phi, \lambda, p) = (0.5, 0.6, 0.2, 1, 0.5)$					
100	−0.085317	−0.112745	0.002839	−0.323747	0.030847
	(0.014727)	(0.047759)	(0.011621)	(0.114894)	(0.000961)
500	−0.020441	−0.0261532	0.0810343	−0.3034194	0.0082843
	(0.0035852)	(0.0142071)	(0.0106941)	(0.0962788)	(0.0000693)
1000	−0.0237843	−0.0110103	0.0696525	−0.3020769	0.0058964
	(0.0026129)	(0.0104978)	(0.0063613)	(0.0936337)	(0.0000351)
5000	0.0032367	−0.0038256	0.0868606	−0.2997216	−0.0136095
	(0.0009141)	(0.0042571)	(0.0077478)	(0.0910393)	(0.0001871)

(a) n=100; $\alpha = 0.5; \vartheta = 0.6; p = 0.2; \lambda = 0.5; \phi = 0.5$

(b) n=100; $\alpha = 0.5; \vartheta = 0.6; p = 0.2; \lambda = 1.5; \phi = 0.5$

(c) n=100; $\alpha = 0.5; \vartheta = 0.6; p = 0.02; \lambda = 5; \phi = 0.5$

(d) n=100; $\alpha = 0.7; \vartheta = 0.3; p = 0.1; \lambda = 0.5; \phi = 0.7$

(e) n=100; $\alpha = 0.7; \vartheta = 0.3; p = 0.1; \lambda = 1.5; \phi = 0.7$

(f) n=100; $\alpha = 0.7; \vartheta = 0.3; p = 0.1; \lambda = 5; \phi = 0.7$
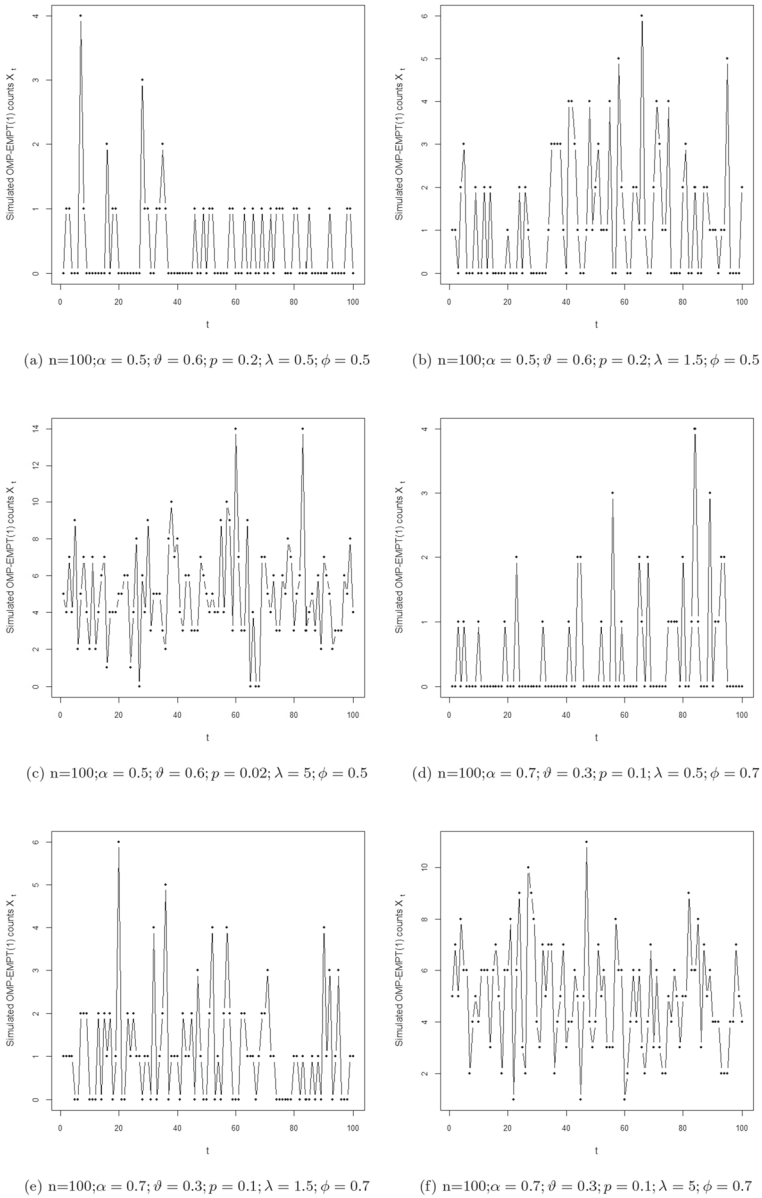
Fig. 7 Sample paths of OMP-EMPT(1) process

4 Real-life data analysis, model selection, diagnostic tools, and forecasting

In this section, the usefulness and applicability of the OMP-EMPT(1) process are shown using two real-life data sets. The assessment of the adequacy of the models proposed and fitted to the data under analysis is a critical step in any statistical investigation. In the literature,

several methods for model validation and diagnostics in discrete-valued time series have been proposed. These methods are broadly classified as follows: model comparisons using information criteria and scores; residual-based methods; and predictive distribution-based methods.

4.1 Data and model selection based on information criteria

The application of the proposed process is investigated using economic and crime data sets:

- The first data set represents $n = 242$ weekly sales of a particular soap product in a supermarket, obtained from a database provided by the Kilts Center for Marketing, Graduate School of Business of the University of Chicago (available at http://research. chicagobooth.edu/marketing/databases/dominicks/categories/soa.aspx; the product is 'Zest White Water 15 oz', with code 3700031165), or data can also be accessed from https://drive.google.com/drive/folders/165XirSF7gKiACX-Mg0Ho16QZ0J0KlqIe?usp =sharing.
- The second data set represents $n = 142$ monthly counts of liquor law violations, reported in the 12th police car beat in Pittsburgh Plus, reported in a period from January 1990 to October 2001, and obtained from the Crime data section of the Forecasting Principles site http://www.forecastingprinciples.com or it can be obtained by https://drive.google. com/drive/folders/1b4a1Z2y9MPSFl1e3uMMmWc8zC-kjYVxU?usp=sharing. Table 3 displays some descriptive statistics for the two data sets. The time series, sample autocorrelation, partial autocorrelation functions (SACF and SPACF, respectively), and empirical p.m.f. for the weekly sales counts and monthly counts of liquor law violations are displayed in Figs. 8 and 9, respectively. From these plots, we conclude that a first-order autoregressive model may be appropriate for the two data sets, as the first autocorrelation is more significant than the other and the SPACF cuts off after lag one. Moreover, the behavior of the two data sets indicates that they may be stationary, and we confirm that result using the test of stationarity. The stationarity of the data is justified throughout the Augmented Dickey–Fuller (ADF) test based on the p-value of the ADF test; the stationarity of the data is supported, and the null hypothesis (non-stationary) is rejected.

The empirical p.m.f. displayed in black in Figs. 8d and 9d, while gray bars refer to the fitted OMP distribution below. From these figures, the empirical p.m.f. and the estimated OMP p.m.f. are close to each other.

As can be seen in Table 3, the sample variance is greater than the sample mean for the data sets; therefore, they are over-dispersed. Schweer and Weiß (2014) proposed an over-dispersion test that is used to check this assumption, with a test statistic based on the empirical dispersion index $\hat{I}_x = \frac{S^2}{\overline{X}}$, where S^2 and $\overline{X}$ are, respectively, the sample variance and mean. The null hypothesis $H_0 : x_1, \ldots, x_n$ is derived from an equi-dispersed PINAR(1) process (Al-Osh and Alzaid 1987), while the alternative $H_1 : x_1, \ldots, x_n$ stems from an over-dispersed INAR(1) process, with a level of significance of $\beta_0 = 0.05$. If the observed value of $\hat{I}_x$ exceeds the critical value $1 + z_{1-\beta_0}\sqrt{\frac{2(1+\alpha^2)}{n(1-\alpha^2)}}$, or if the p-value $1 - \Phi\left(\sqrt{\frac{n(1-\alpha^2)}{2(1+\alpha^2)}} \times (\hat{I}_x - 1)\right)$ falls below β_0, the null hypothesis is rejected, where $z_{1-\beta_0}$ and $\Phi(.)$ represents the $(1 - \beta_0)-$ quantile and the distribution function of the standard normal distribution, respectively, and $\alpha = \hat{\rho}(1)$. The considered data sets are empirically over-dispersed, with dispersion indices, $\hat{I}_x$ given in Table 3. By applying the over-dispersion test, the critical values and the p-values for the weekly sales and monthly liquor law violations counts are (1.165164, 1.232448) and

Table 3 Descriptive statistics of weekly sales and monthly liquor law violations counts

data sets	n	Min	Q_1	Q_2	Mean	$\hat{\rho}(1)$	Q_3	Max	Var	p-value (ADF)	$\hat{I}_X$
Weekly sales	242	0.0000	3.0000	4.000	5.4421	0.3924	7.000	22.000	15.4012	0.01	2.818292
liquor law violations	142	0.000	0.000	0.000	0.6549	0.50726	1.000	11.000	2.185	0.01	3.312812

(0.000,0.000), respectively. Therefore, the data sets stem from an over-dispersed INAR(1) process.

For comparison purposes, the two data sets are used to compare the proposed process, OMP-EMPT(1), with the following relevant INAR(1) models: PINAR(1) (Al-Osh and Alzaid 1987), Pegram AR(1) model with Poisson marginal (Biswas and Song 2009), Poisson MPT(1) (Khoo et al. 2017), MPDBPINAR(1) (Shirozhan and Mohammadpour 2018). We use the following goodness-of-fit statistics: the Akaike information criterion (AIC), the Hannan–Quinn information criterion (HQIC), and the corrected Akaike information criterion (AICc), where the best model has the ones with the lowest values. The results of these goodness-of-fit statistics and ML estimates with their standard error (S.E.) are displayed in Tables 4 and 5. These tables show that the values of the AIC, HQIC, and AICc are the smallest for the OMP-EMPT(1) process. Hence, we conclude that the proposed model works well for the two data sets. Also, we can conclude that the OMP-EMPT(1) model is significantly better than the Poisson-MPT(1) model for explaining both data sets, as we have $\hat{\phi} = 0.3185468$, 0.6040971, i.e., the proportion of ones that were misclassified as zeros in the process does not equal zero for the considered data sets. When $\hat{\phi} = 0.6040971$, this means that most ones are misclassified and reported as zeros, and this implies a zero inflation as seen in Fig. 9d. This means that when determining the number of weekly sales or liquor violations per unit, an inspector may make an error by reporting units that actually contain one item or one person as having no item or no one, respectively. The R script that was used to fit the data to the OMP-EMPT(1) model is included in the Appendix A.4.

4.2 Diagnostic tools and model adequacy

In this subsection, we check the model's adequacy throughout the logarithmic scoring rule, non-randomized probability integral transform (PIT), and benchmark chart.

Scoring rules can be used to assess a model's relative performance within a given group of competing models. In decision analysis, scoring rules are frequently used to evaluate the accuracy of probabilistic predictions by assigning a numerical score based on the predictive distribution and observed data. To compare the observation x_t realized at time t with the conditional distribution $p_{.|x_{t-1}}$ based on the previous observation, typical scoring rules are of the form $s(p_{.|x_{t-1}}, x_t)$, with smaller score values indicating better agreement. The mean score $\frac{1}{T-1} \sum_{t=2}^{T} s(p_{.|x_{t-1}}, x_t)$ is used to evaluate the model's overall predictive performance with respect to the time series $x_1, \cdots, x_T$. There are many scoring rules, such as quadratic, logarithmic, ranked probability, and spherical scores, to select the model with the smallest value of its mean. For the two data sets, we compute the logarithmic score defined as

$$s_{ls}(p_{(.|x_{t-1})}, x_t) = -\ln p_{x_t|x_{t-1}}, \tag{4.1}$$

Table 4 ML estimates with their corresponding standard error (S.E.) in brackets and some goodness-of-fit statistics for weekly sales

Model	ML estimates (S.E.)	AIC	HQIC	AICc
i.i.d Poisson	$\hat{\lambda} = 5.460581(0.1505257)$	1418.0350	1419.4405	1418.0852
Poisson INAR(1)	$\hat{\lambda} = 4.1669088(0.21054858)$	1371.320	1374.131	1371.421
	$\hat{\alpha} = 0.2332869(0.03230576)$			
Poisson Pegram's AR(1)	$\hat{\lambda} = 5.43232677(0.15655427)$	1407.4598	1410.2708	1407.5607
	$\hat{\phi} = 0.05120388(0.02061672)$			
Poisson MPT(1)	$\hat{\alpha} = 0.8541167(0.04089583)$	1373.2362	1377.4526	1373.4049
	$\hat{\lambda} = 5.2793810(0.16848405)$			
	$\hat{\phi} = 0.1287824(0.03365876)$			
MPDBPINAR(1)	$\hat{\lambda} = 5.2734328(0.16828282)$	1373.8812	1379.5031	1374.1354
	$\hat{\alpha} = 0.8503549(0.03970675)$			
	$\hat{\theta} = 0.2068880(0.12359867)$			
	$\hat{\phi} = 0.1307891(0.03295645)$			
OMP-EMPT(1)	$\hat{\alpha} = 0.8269426(0.0395708)$	1366.1007	1373.1281	1366.4582
	$\hat{\vartheta} = 0.2100749(0.0447467)$			
	$\hat{p} = 0.1490073(0.0343505)$			
	$\hat{\lambda} = 5.3034063(0.1705476)$			
	$\hat{\phi} = 0.3185468(0.1263267)$			

Table 5 ML estimates with their corresponding standard error (S.E.) and some goodness-of-fit statistics for monthly liquor law violations counts

Model	ML estimate(S.E.)	AIC	HQIC	AICc
i.i.d Poisson	$\hat{\lambda} = 0.6312058(0.06690768)$	377.2913	378.4925	377.3777
Poisson INAR(1)	$\hat{\lambda} = 0.4539531(0.06079951)$	346.2811	348.6834	346.4551
	$\hat{\alpha} = 0.3159473(0.06051378)$			
Poisson Pegram's AR(1)	$\hat{\lambda} = 0.8951231(0.14007193)$	364.1458	366.5480	364.3197
	$\hat{\phi} = 0.3438248(0.08675592)$			
Poisson MPT(1)	$\hat{\alpha} = 0.6149809(0.10692628)$	350.8832	354.4866	351.1752
	$\hat{\lambda} = 0.7885890(0.11757260)$			
	$\hat{\phi} = 0.5023093(0.09847041)$			
MPDBPINAR(1)	$\hat{\lambda} = 0.7901840(0.1190384)$	352.6239	357.4284	353.0650
	$\hat{\alpha} = 0.6006368(0.1328089)$			
	$\hat{\theta} = 0.2555253(0.2953373)$			
	$\hat{\phi} = 0.5185557(0.1142482)$			
OMP-EMPT(1)	$\hat{\alpha} = 0.5626468(0.17990509)$	327.6184	333.6240	328.2406
	$\hat{\vartheta} = 0.5636788(0.14690364)$			
	$\hat{p} = 0.3142644(0.14164205)$			
	$\hat{\lambda} = 0.8869830(0.10475618)$			
	$\hat{\phi} = 0.6040971(0.09361425)$			

(a) Sample path of weekly sales

(b) SPACF

(c) SACF
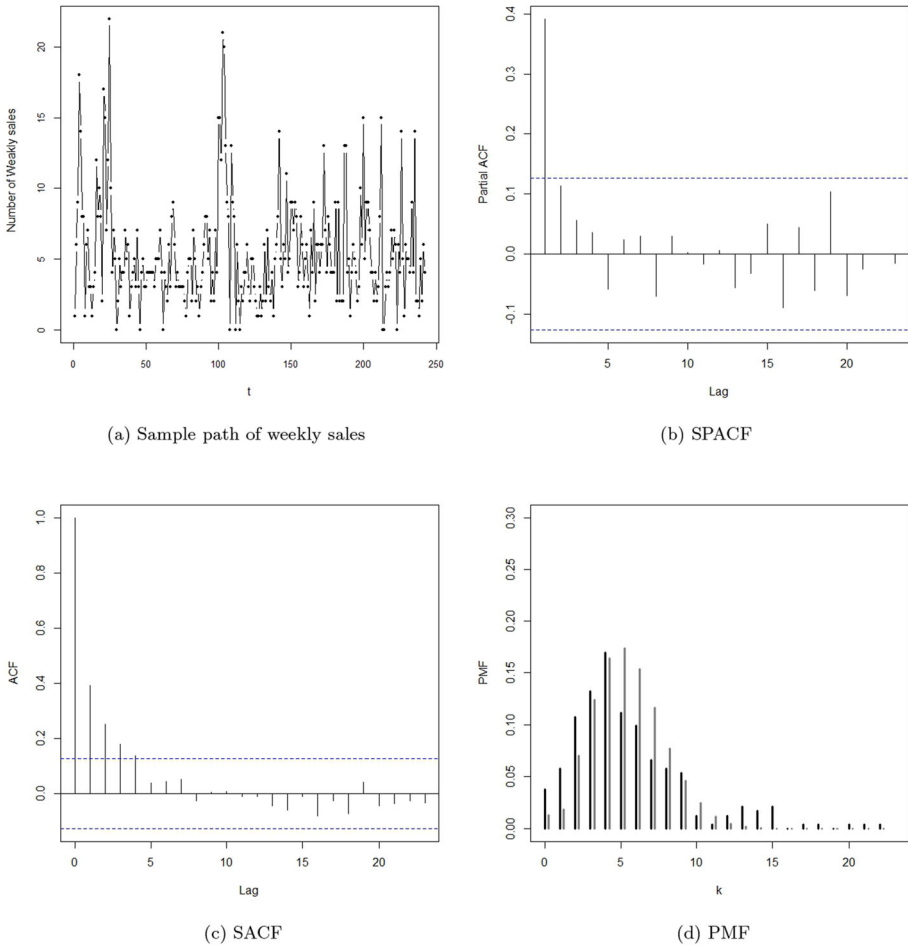
(d) PMF

Fig. 8 Sample path, SPACF, SACF, and PMF of the number of weekly sales

which corresponds to the conditional log-likelihood calculation defined by Eq. (2.13). Table 6 shows the mean logarithmic score $(\overline{S_{ls}})$ for the proposed fitted process and the competing models of the two data series. The mean score recommends the OMP-EMPT(1) model among the competitive INAR models.

The adequacy of the proposed fitted model is checked using the jumps $J_t = X_t - X_{t-1}$, for $t = 2, 3, \cdots, n$, $E(J_t)$, and $Var(J_t)$, which are given by Eq.(2.29) and Eq.(2.30), respectively. Figures 10 and 11a represent the jumps against time (Shewhart control chart) of the two data sets with $\pm 3\sigma_J$ limits chosen as the benchmark chart as proposed by Weiß (2009), where $\sigma_J = \sqrt{Var(J_t)} = 3.072668$, and 1.288924, respectively. The Shewhart control chart for the two considered data sets indicates that there is no particular point causing a huge impact in the model and implies that the proposed model fits the two data sets well. If a model is suitable for the data series, the non-randomized probability integral transform (PIT) histogram, Czado et al. (2009), should be uniform. Figures 10 and 11b show the PIT for the two data sets, indicating that the PIT histogram is close to uniform and that the proposed model fits the two data sets well, i.e., the model is adequate for these data series.

(a) Sample path of monthly liquor law violations counts

(b) SPACF

(c) SACF
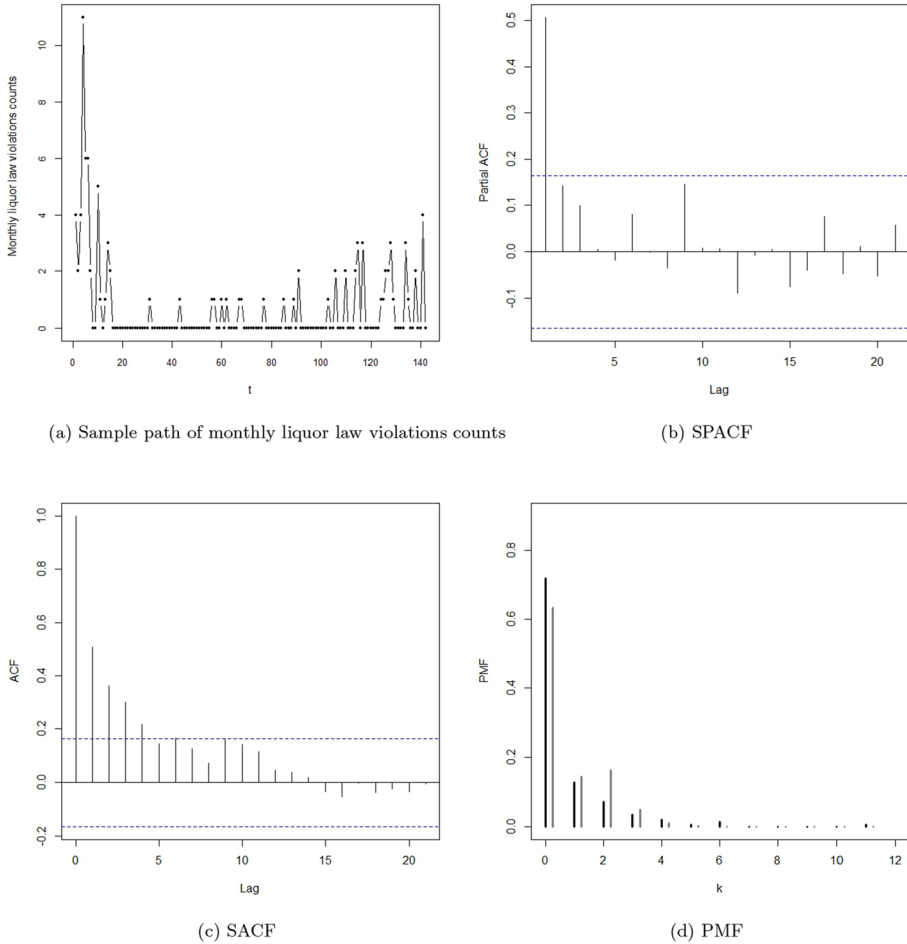
(d) PMF

Fig. 9 Sample path, SACF, SPACF, and PMF of the number of monthly liquor law violations counts

Table 6 The mean logarithmic score for the considered data sets

Model	$\overline{S_{ls}}$ (weekly sales)	$\overline{S_{ls}}$ (liquor law violations)
Poisson INAR(1)	2.821237	1.174885
Pegram's AR(1) Poisson	2.896224	1.245082
Poisson MPT(1)	2.821587	1.18812
MPDBPINAR(1)	2.818796	1.187246
OMP-EMPT(1)	2.796812	1.094074

4.3 Forecasting

The minimum mean square error is one of the most common procedures for optimally fore-casting future values of a model. The conditional mean finds a forecast, $\hat{X}_{t+k}$ of X_{t+k} that minimizes the expected squared error given a sample. Using Eqs. (2.16) and (2.17), the one-

(a) The Shewhart control chart for the jumps   (b) PIT histogram
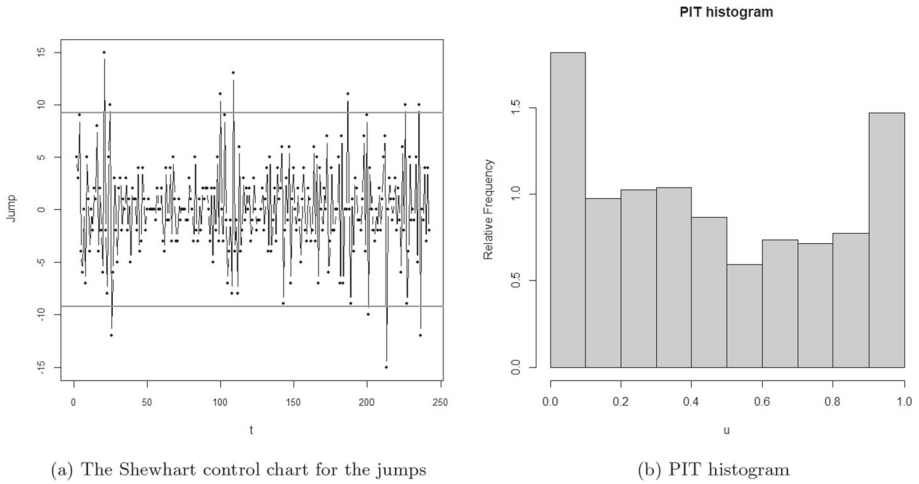
Fig. 10 Jumps against time and PIT of fitted OMP-EMPT(1) for weekly sales

(a) The Shewhart control chart for the jumps   (b) PIT histogram
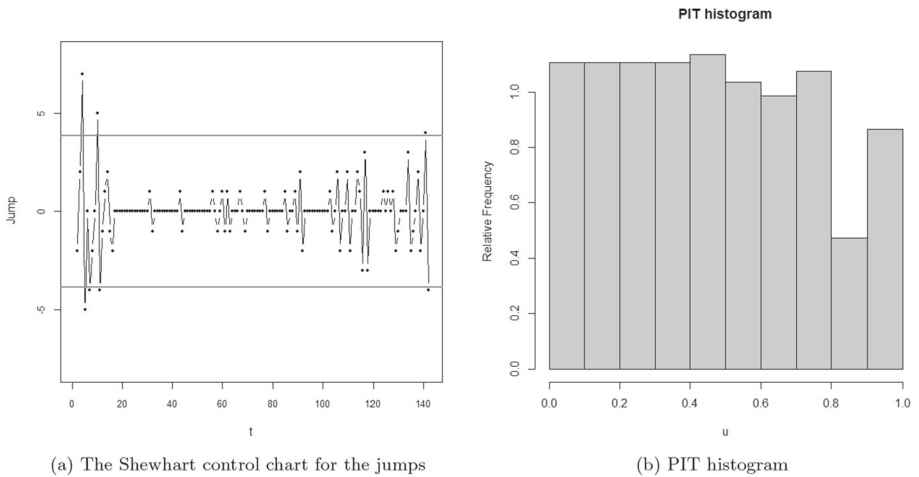
Fig. 11 Jumps against time and PIT of fitted OMP-EMPT(1) for monthly liquor law violations counts

and k-step-ahead forecasts of X_{t+1} and X_{t+k} are, respectively, given by

$$\hat{X}_{t+1} = E(X_{t+1}|X_t) = (p\alpha)X_t + (1 - (p\alpha))\lambda(1 - \phi e^{-\lambda}),$$

where $\hat{X}_1 = (1 - (p\alpha))\lambda(1 - \phi e^{-\lambda})$ and

$$\hat{X}_{t+k} = E(X_{t+k}|X_t) = (p\alpha)^k X_t + (1 - (p\alpha)^k)\lambda(1 - \phi e^{-\lambda}), k \geq 1.$$

In practice, the parameters p, α, λ, and ϕ are replaced by their ML estimates. Therefore

$$\hat{X}_{t+1} = E(X_{t+1}|X_t) = (\hat{p}\hat{\alpha})X_t + (1 - (\hat{p}\hat{\alpha}))\hat{\lambda}(1 - \hat{\phi} e^{-\hat{\lambda}}), \qquad (4.2)$$

and

$$\hat{X}_{t+k} = E(X_{t+k}|X_t) = (\hat{p}\hat{\alpha})^k X_t + (1 - (\hat{p}\hat{\alpha})^k)\hat{\lambda}(1 - \hat{\phi} e^{-\hat{\lambda}}), k \geq 1. \qquad (4.3)$$

Series **Predicted series**
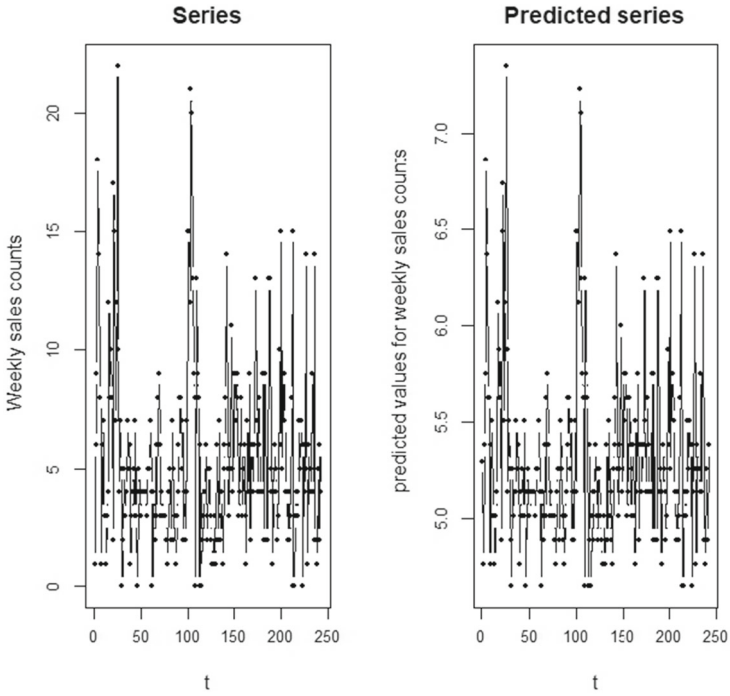
Fig. 12 Actual and predicted values of weekly sales

Series **Predicted series**
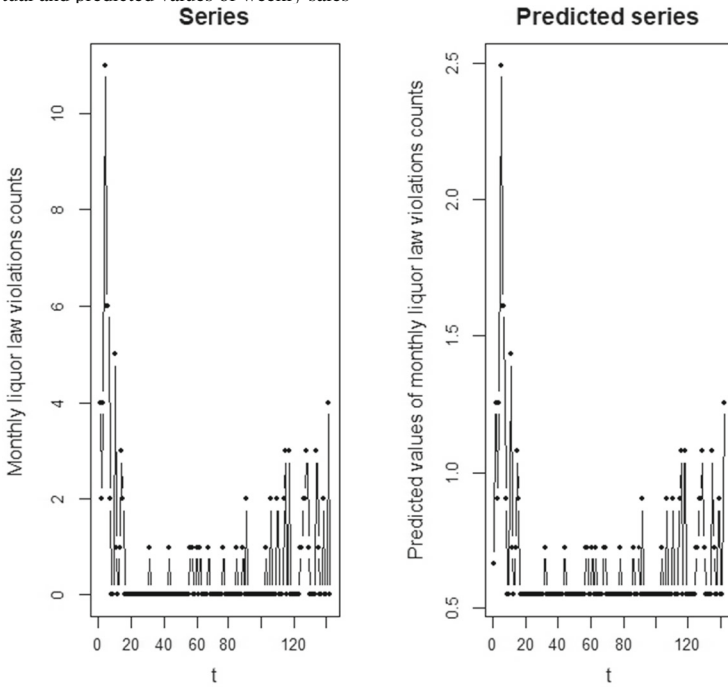
Fig. 13 Actual and predicted values of monthly liquor law violations counts

The two actual data sets and their predicted values in the classical way using Eq. (4.2) are displayed and plotted in Figs. 12 and 13 to show the predictive ability of the OMP-EMPT(1) process, where $\hat{\alpha}$, $\hat{\phi}$, $\hat{p}$, and $\hat{\lambda}$ are presented in Tables 4 and 5. From these figures, we observe that the predicted values are close to the original ones in the data sets, indicating that the OMP-EMPT(1) process can provide a good forecast for these data sets.

Data Availability The application section of this research cites the URLs to the used data sets, and the authors can provide the data upon request.

Declarations

Conflict of interest There is no conflict of interest.

A Appendix

A.1 Proof of Proposition 2.2

Proof 1. The one-step-ahead conditional expectation is calculated as

$$E(X_{t+1}|X_t) = p E(\alpha \star_\vartheta X_t | X_t) + (1-p) E(\xi_{t+1}|X_t)$$
$$= p\alpha X_t + (1-p)\mu_{\xi_t}.$$

2. By using mathematical induction, the $k-$step-ahead conditional expectation is given as follows: For $k=1$, as determined by Eq. (2.16); similarly, $k=2$

$$E(X_{t+2}|X_t) = E_{X_{t+1}|X_t}(E(X_{t+2}|X_{t+1}))$$
$$= E_{X_{t+1}|X_t}((p\alpha X_{t+1} + (1-p)\mu_{\xi_{t+2}})|X_{t+1})$$
$$= p\alpha E(X_{t+1}|X_t) + (1-p)\mu_{\xi_t}$$
$$= (p\alpha)^2 X_t + (1-p)\mu_{\xi_t}(1+p\alpha),$$

where μ_{ξ_t} is given by Eq. (2.9).
Now, suppose that the $(k-1)-$step-ahead conditional expectation is

$$E(X_{t+k-1}|X_t) = (p\alpha)^{k-1} X_t + (1-p)\mu_{\xi_t} \frac{1-(p\alpha)^{k-1}}{1-p\alpha}.$$

Hence, the $k-$step-ahead conditional expectation is obtained as follows:

$$E(X_{t+k}|X_t) = E_{X_{t+1}|X_t}(E(X_{t+k}|X_{t+1}))$$

$$= E\left[\left((p\alpha)^{k-1}X_{t+1} + (1-p)\mu_{\xi_{t+1}}\frac{1-(p\alpha)^{k-1}}{1-p\alpha}\right)|X_t\right]$$

$$= (p\alpha)^{k-1}[p\alpha X_t + (1-p)\mu_{\xi_t}] + (1-p)\frac{1-(p\alpha)^{k-1}}{1-p\alpha}\mu_{\xi_{t+1}}$$

$$= (p\alpha)^k X_t + (1-p)\mu_{\xi_t}\left[(p\alpha)^{k-1} + \frac{1-(p\alpha)^{k-1}}{1-p\alpha}\right].$$

3. The one-step-ahead conditional variance is computed as follows:

$$Var(X_{t+1}|X_t) = E(X_{t+1}^2|X_t) - (E(X_{t+1}|X_t))^2.$$

Thus, the expression for $E(X_{t+1}^2|X_t)$ is calculated to get $Var(X_{t+1}|X_t)$

$$E(X_{t+1}^2|X_t) = pE((\alpha\star_{\vartheta}X_t)^2|X_t) + (1-p)E(\xi_{t+1}^2|X_t)$$
$$= p[(2\alpha - \alpha\vartheta + \vartheta - 1)X_t^2 + (1-\alpha)(1-\vartheta)X_t] + (1-p)E(\xi_t^2). \tag{A.1}$$

Hence

$$Var(X_{t+1}|X_t) = p[(2\alpha - \alpha\vartheta + \vartheta - 1)X_t^2 + (1-\alpha)(1-\vartheta)X_t]$$
$$+(1-p)E(\xi_t^2) - (p\alpha X_t + (1-p)\mu_{\xi_t})^2,$$

using Eqs. (2.16), (2.9), and (2.11) for $E(X_{t+1}|X_t)$, μ_{ξ}, and $E(\xi^2)$, respectively, then

$$Var(X_{t+1}|X_t) = p[(2\alpha - \alpha\vartheta + \vartheta - 1) - p\alpha^2]X_t^2$$
$$+p\left[(1-\alpha)(1-\vartheta) - 2\alpha(1-p)\frac{\lambda(1-p\alpha)(1-\phi e^{-\lambda})}{1-p}\right]X_t$$
$$+(1-p)\left(\frac{(1-p\alpha)\lambda(1-\phi e^{-\lambda}) + \lambda^2(1+p+p\alpha(-2+\vartheta)-p\vartheta)}{1-p}\right)$$
$$-(1-p)^2\left(\frac{\lambda(1-p\alpha)(1-\phi e^{-\lambda})}{1-p}\right)^2.$$

After some simplification, the one-step conditional variance in Eq. (2.18) is derived.

4. To compute the k-step conditional variance, $Var(X_{t+k}|X_t)$, first compute the expression $E(X_{t+k}^2|X_t)$. The first-step of the conditional second moment that is given by Eq. (A.1) can be written as follows: by setting $\mathbf{A_0} = (2\alpha - \alpha\vartheta + \vartheta - 1)$ and $\mathbf{B_0} = (1-\alpha)(1-\vartheta)$:

$$E(X_{t+1}^2|X_t) = p\mathbf{A_0}X_t^2 + p\mathbf{B_0}X_t + (1-p)E(\xi_t^2).$$

For $k = 2$:

$$E(X_{t+2}^2|X_t) = E(E(X_{t+2}^2|X_{t+1})|X_t)$$
$$= pE((\alpha\star_{\vartheta}X_{t+1})^2|X_t) + (1-p)E(\xi_t^2)$$
$$= p[E((\mathbf{A_0}X_{t+1}^2 + \mathbf{B_0}X_{t+1})|X_t)] + (1-p)E(\xi_t^2)$$
$$= (p\mathbf{A_0})^2 X_t^2 + p^2\mathbf{B_0}(\mathbf{A_0}+\alpha)X_t + (1-p)(1+p\mathbf{A_0})E(\xi_t^2) + (1-p)p\mathbf{B_0}\mu_{\xi_t}.$$

For $k = 3$:

$$E(X_{t+3}^2|X_t) = E(E(X_{t+3}^2|X_{t+1})|X_t)$$
$$= E([(p\mathbf{A_0})^2 X_{t+1}^2 + p^2\mathbf{B_0}(\mathbf{A_0}+\alpha)X_{t+1}]|X_t)$$
$$+(1-p)(1+p\mathbf{A_0})E(\xi_t^2) + (1-p)p\mathbf{B_0}\mu_{\xi_t}$$

$$= (p\mathbf{A_0})^3 X_t^2 + p^3 \mathbf{B_0} X_t \left(\frac{\mathbf{A_0^3} - \alpha^3}{\mathbf{A_0} - \alpha} \right) + (1-p) E(\xi_t^2) \left(\frac{(p\mathbf{A_0})^3 - 1}{(p\mathbf{A_0}) - 1} \right)$$

$$+ \frac{(1-p)\mathbf{B_0}\mu_{\xi_t}}{(\mathbf{A_0} - \alpha)} \left(\frac{(p\mathbf{A_0})^3 - 1}{p\mathbf{A_0} - 1} - \frac{(p\alpha)^3 - 1}{p\alpha - 1} \right).$$

Hence, by continuing the procedure and observing the obtained formulas, it can be concluded that the conditional second moment for $k-$step, $E(X_{t+k}^2|X_t)$, is

$$E(X_{t+k}^2|X_t) = (p\mathbf{A_0})^k X_t^2 + p^k \mathbf{B_0} X_t \left(\frac{\mathbf{A_0^k} - \alpha^k}{\mathbf{A_0} - \alpha} \right) + (1-p) E(\xi_t^2) \left(\frac{(p\mathbf{A_0})^k - 1}{(p\mathbf{A_0}) - 1} \right)$$

$$+ \frac{(1-p)\mathbf{B_0}\mu_{\xi_t}}{(\mathbf{A_0} - \alpha)} \left(\frac{(p\mathbf{A_0})^k - 1}{p\mathbf{A_0} - 1} - \frac{(p\alpha)^k - 1}{p\alpha - 1} \right). \tag{A.2}$$

Therefore, using Eqs. (A.2) and (2.17) and the expression $Var(X_{t+k}|X_t) = E(X_{t+k}^2|X_t) - (E(X_{t+k}|X_t))^2$, the conditional variance for $k-$step is obtained as

$$Var(X_{t+k}|X_t)$$

$$= (p\mathbf{A_0})^k X_t^2 + p^k \mathbf{B_0} X_t \left(\frac{\mathbf{A_0^k} - \alpha^k}{\mathbf{A_0} - \alpha} \right) + (1-p) E(\xi_t^2) \left(\frac{(p\mathbf{A_0})^k - 1}{(p\mathbf{A_0}) - 1} \right)$$

$$+ \frac{(1-p)\mathbf{B_0}\mu_{\xi_t}}{(\mathbf{A_0} - \alpha)} \left(\frac{(p\mathbf{A_0})^k - 1}{p\mathbf{A_0} - 1} - \frac{(p\alpha)^k - 1}{p\alpha - 1} \right)$$

$$- ((p\alpha)^k X_t + (1-p)\mu_{\xi_t} \frac{(p\alpha)^k - 1}{p\alpha - 1})^2$$

$$= ((p\mathbf{A_0})^k - (p\alpha)^{2k}) X_t^2 + (p^k \mathbf{B_0} \left(\frac{\mathbf{A_0^k} - \alpha^k}{\mathbf{A_0} - \alpha} \right) - 2(p\alpha)^k \mu_X (1 - (p\alpha)^k)) X_t$$

$$+ \left[(1-p) E(\xi_t^2) \left(\frac{(p\mathbf{A_0})^k - 1}{(p\mathbf{A_0}) - 1} \right) + \frac{(1-p\alpha)\mathbf{B_0}\mu_X}{(\mathbf{A_0} - \alpha)} \left(\frac{(p\mathbf{A_0})^k - 1}{p\mathbf{A_0} - 1} - \frac{(p\alpha)^k - 1}{p\alpha - 1} \right) \right.$$

$$\left. - (\mu_X (1 - (p\alpha)^k))^2 \right],$$

using $E(\xi_t^2)$ that is given by Eq. (2.11) and $\mu_X = \lambda(1 - \phi e^{-\lambda})$, the $k-$step conditional variance for the process is obtained.

□

A.2 Derivations of the CLS estimators for α, ϕ, and λ

Proof The CLS estimators given by the expressions in Eqs. (3.2), (3.4), and (3.3) are obtained as follows.

Recall that

$$Q_n(\Theta) = \sum_{t=2}^{n} (X_t - E(X_t|X_{t-1}))^2$$

$$= \sum_{t=2}^{n} (X_t - p\alpha X_{t-1} - (1 - p\alpha)\mu_X)^2,$$

with $\mu_X = \lambda(1 - \phi e^{-\lambda})$, then

$$\frac{\partial Q_n(\Theta)}{\partial \alpha} = 2p \sum_{t=2}^{n} (X_t - p\alpha X_{t-1} - (1 - p\alpha)\mu_X)(-X_{t-1} + \mu_X),$$

and

$$\frac{\partial Q_n(\Theta)}{\partial \mu_X} = -2(1 - p\alpha) \sum_{t=2}^{n} (X_t - p\alpha X_{t-1} - (1 - p\alpha)\mu_X).$$

Consequently, $\frac{\partial Q_n(\Theta)}{\partial \Theta} = 0$, which implies

$$\sum_{t=2}^{n} X_t - p\alpha \sum_{t=2}^{n} X_{t-1} - (n-1)(1 - p\alpha)\mu_X = 0,$$

$$\sum_{t=2}^{n} (X_t - p\alpha X_{t-1} - (1 - p\alpha)\mu_X)\mu_X = \sum_{t=2}^{n} (X_t - p\alpha X_{t-1} - (1 - p\alpha)\mu_X)X_{t-1};$$

as a result, we have

$$\widehat{\mu}_X = \frac{\sum_{t=2}^{n} X_t - p\hat{\alpha} \sum_{t=2}^{n} X_{t-1}}{(n-1)(1 - p\hat{\alpha})},$$

$$\underbrace{\sum_{t=2}^{n} (X_t - p\alpha X_{t-1} - (1 - p\alpha)\mu_X)\mu_X}_{(I)} = \underbrace{\sum_{t=2}^{n} (X_t - p\alpha X_{t-1} - (1 - p\alpha)\mu_X)X_{t-1}}_{(II)}.$$

We obtain that $(I) = 0$ by rearranging it and substituting $\widehat{\mu}_X$ inside the brackets as follows:

$$(I) = \widehat{\mu}_X \left\{ \sum_{t=2}^{n} X_t - p\hat{\alpha} \sum_{t=2}^{n} X_{t-1} - (n-1)(1 - p\hat{\alpha})\hat{\mu}_X \right\}$$

$$= \widehat{\mu}_X \left\{ \sum_{t=2}^{n} X_t - p\hat{\alpha} \sum_{t=2}^{n} X_{t-1} - \left(\sum_{t=2}^{n} X_t - p\hat{\alpha} \sum_{t=2}^{n} X_{t-1} \right) \right\}$$

$$= 0.$$

Using (II) and replacing $\widehat{\mu}_X$, we obtain the estimator for α as follows:

$$(II) = \sum_{t=2}^{n} X_t X_{t-1} - p\hat{\alpha} \sum_{t=2}^{n} X_{t-1}^2 - \sum_{t=2}^{n} X_{t-1}$$

$$\times \left(\frac{\sum_{t=2}^{n} X_t - p\hat{\alpha} \sum_{t=2}^{n} X_{t-1}}{(n-1)} \right)$$

Therefore, as (II)=(I)=0;

$$\hat{\alpha} p \left(\sum_{t=2}^{n} X_{t-1}^2 - \frac{\left(\sum_{t=2}^{n} X_{t-1} \right)^2}{(n-1)} \right) = \sum_{t=2}^{n} X_t X_{t-1} - \frac{\sum_{t=2}^{n} X_{t-1} \sum_{t=2}^{n} X_t}{(n-1)}.$$

$$\hat{\alpha}_{\text{cls}} = \frac{\sum_{t=2}^{n} X_t X_{t-1} - \frac{1}{(n-1)} \sum_{t=2}^{n} X_{t-1} \sum_{t=2}^{n} X_t}{p \left(\sum_{t=2}^{n} X_{t-1}^2 - \frac{1}{(n-1)} \left(\sum_{t=2}^{n} X_{t-1} \right)^2 \right)},$$

by replacing p by some consistent estimators, $\hat{p}_{\text{ML}}$, as an initial value, we obtain Eq. (3.2). Hence, the CLS estimator for p is given by

$$\hat{p}_{\text{cls}} = \frac{\sum_{t=2}^{n} X_t X_{t-1} - \frac{1}{(n-1)} \sum_{t=2}^{n} X_{t-1} \sum_{t=2}^{n} X_t}{\hat{\alpha}_{\text{cls}} \left(\sum_{t=2}^{n} X_{t-1}^2 - \frac{1}{(n-1)} \left(\sum_{t=2}^{n} X_{t-1} \right)^2 \right)}.$$

The parameter λ is estimated using the method of moments based on the following statistics:

$$E(X_t) = \lambda(1 - \phi e^{-\lambda})$$
$$E(X_t^2) = \lambda^2 + \lambda(1 - \phi e^{-\lambda}).$$

Therefore, using the first and second sample means, we obtain the estimator of λ from the expression

$$\overline{X^2} - \bar{X} = \widehat{\lambda}^2.$$

Therefore, the expression in Eq. (3.3) is obtained as follows:

$$\hat{\lambda} = \sqrt{\overline{X^2} - \bar{X}} = \sqrt{\frac{\sum_{t=1}^{n} X_t^2 - \sum_{t=1}^{n} X_t}{n}}.$$

The estimator for ϕ is obtained as

$$\widehat{\lambda}(1 - \widehat{\phi} e^{-\widehat{\lambda}}) = \widehat{\mu}_X = \frac{\sum_{t=2}^{n} X_t - \hat{p}\widehat{\alpha} \sum_{t=2}^{n} X_{t-1}}{(n-1)(1 - \widehat{p\alpha})}.$$

As a result, $\widehat{\phi}$ is provided as

$$\hat{\phi}_{\text{cls}} = \hat{\lambda}_{\text{cls}} \left(1 - \frac{\sum_{t=2}^{n} X_t - \hat{p}_{\text{cls}} \sum_{t=2}^{n} X_{t-1}}{\hat{\lambda}_{\text{cls}}(n-1)(1 - \hat{p}_{\text{cls}}\hat{\alpha}_{\text{cls}})} \right).$$

$\square$

A.3 R code for simulated series at $n = 100$; $\alpha = 0.5$; $\vartheta = 0.6$; $p = 0.2$; $\lambda = 1.5$; $\phi = 0.5$

```
rm(list=ls(all=TRUE))
n=100;alpha=0.5;theta=0.6;p=0.2;lambda=1.5;phi=0.5
error.MRPinar < − function(n,alpha,theta,p,lambda,phi){
epsilonmr < − rep(NA, times = n)
for (i in 1:n){
u=runif(1,0,1)
if(u < (1/(1-p))*(lambda*phi*exp(-lambda)*(1-p*alpha))){
epsilonmr[i]=0
}else if (u>= (1/(1-p))*(lambda*phi*exp(-lambda)*(1-p*alpha)) &
u < ((1/(1-p))*(lambda*phi*exp(-lambda)*(1-p*alpha))
+((1/(1-p))*(1-(p*(alpha+theta-1)/theta))))){
epsilonmr[i]=rpois(1,lambda)}else if(u >=((1/(1-p))*(lambda*phi*exp(-lambda)
*(1-p*alpha)))
+((1/(1-p))*(1-(p*(alpha+theta-1)/theta)))
& u <((1/(1-p))*(lambda*phi*exp(-lambda)*(1-p*alpha))
+((1/(1-p))*(1-(p*(alpha+theta-1)/theta)))-(p/(1-p))*((1-alpha)/theta))){
```

```
epsilonmr [i]=rpois(1,(1-theta)*lambda)}else{
epsilonmr[i]=1}}
return(epsilonmr)}
xt < − rep(NA, times = n)
epsilonmr < − error.MRPinar(n,alpha,theta,p,lambda,phi)
xt[1] = round(lambda*(1-phi*exp(-lambda)))
for (t in 2:n){
u< −runif(1,0,1)
if(u< p*((1-alpha)/theta)){
xt[t]< −rbinom(1,xt[t-1],(1-theta))}
else if(u>=p*((1-alpha)/theta)&u<p){xt[t]=xt[t-1]}
else {xt[t]=epsilonmr[t]}
}
windows()
plot(xt, type="b", pch=19, cex=0.5, xlab = "t", ylab = expression("Simulated OMP-EMPT(1)
counts X\t"[t]), cex.axis=0.85)
```

A.4 R code that was used to fit the data to the OMP-EMPT(1) model

```
rm(list=ls(all=TRUE))
library(tseries)
fx.p< − function(k,lambda,phi){
if(k==0){
fx.p< − exp(-lambda)*(1+lambda*phi)
}else if(k==1){fx.p< − lambda*exp(-lambda)*(1-phi)}
else {fx.p< − (exp(-lambda)*lambda^k)/factorial(k)}
}
xt < − scan(file.choose())
adf.test(xt)
windows()
acf(xt,main=" ")
windows()
pacf(xt,main=" ")
rho1 < − acf(xt, plot=FALSE)[[1]][2]
rho1
Xlen <- length(xt) barX < - mean(xt)
barX
sX < − var(xt)
sX
ID < − (Xlen-1)/Xlen*sX/barX
ID
meanID < − 1-1/Xlen*((1+rho1)/(1-rho1))
meanID
sdID < − sqrt(2/Xlen * ((1 + rho1^2)/(1 − rho1^2)))
sdID
meanID+qnorm(0.95)*sdID
round(1-pnorm(ID, mean=meanID, sd=sdID),6)
```

```
probEpsm < − function(i,alpha,theta,p,lambda,phi) {
if(i==0){
probEpsm < − (((1-p*alpha)*lambda*phi*exp(-lambda)-(p*((1-alpha)/theta))*exp(-lambda*(1-theta))
+(1-(p*(1-((1-alpha)/theta))))*exp(-lambda)))/(1-p)
}
else if (i==1) {
probEpsm < − (((p*alpha-1)*lambda*phi*exp(-lambda)-(p*((1-alpha)/theta))*lambda*(1-theta)
*exp(-lambda*(1-theta))
+(1-(p*(1-((1-alpha)/theta))))*lambda*exp(-lambda)))/(1-p)
}
else {
probEpsm < − ((p*((alpha-1)/theta))*dpois(i,lambda*(1-theta))+(1-(p*(1-((1-alpha)/theta))))
*dpois(i,lambda))/(1-p)
}
}
conditionalPomis < − function(i,j,alpha,theta,p,lambda,phi){
I < − ifelse(j==i,1,0)
(p*((1-alpha)/theta))*dbinom(i,j,1-theta)+(p*(1-((1-alpha)/theta)))*I+(1-p)
*probEpsm(i,alpha,theta,p,lambda,phi)
}
loglimsrp< − function(par,xt) {
alpha< − par[1];theta< − par[2];p< − par[3];lambda< − par[4];phi< − par[5]
T< −length(xt);
value< − -log(fx.p(xt[1],lambda,phi))
for (t in c(2:T)) {
value< − value-log(conditionalPomis(xt[t],xt[t-1],alpha,theta,p,lambda,phi))}
value}
estmsrpo< − suppressWarnings(optim(par=c(0.4,0.7,0.2,0.2,0.2),loglimsrp,xt=xt,method = "BFGS",hessian
= TRUE))
alphaestml < − estmsrpo$par[[1]]
thetaestml < − estmsrpo$par[[2]]
pestml < − estmsrpo$par[[3]]
lambdaestml < − estmsrpo$par[[4]]
phiestml < − estmsrpo$ par[[5]]
ofiest < − estmsrpo$hessian
neglmax < − estmsrpo$value
estcov < − solve(ofiest)
c(alphaestml,thetaestml,pestml,lambdaestml,phiestml)
c(sqrt(diag(estcov)))
k=5
AIC < − 2*neglmax+2*k
HQIC< − 2*neglmax+2*k*log(log(Xlen))
AICc< − AIC+2*(((k+1)*(k+2))/(Xlen-k-2))
c(neglmax, AIC,HQIC,AICc)
```

References

Ahmad M (1968) Truncated multivariate Poisson distributions. Iowa State University, Iowa

Al-Osh MA, Aly E-EA (1992) First order autoregressive time series with negative binomial and geometric marginals. Commun Stat Theory Methods 21(9):2483–2492

Al-Osh M, Alzaid AA (1987) First-order integer-valued autoregressive (INAR (1)) process. J Time Ser Anal 8(3):261–275

Aly E-EA, Bouzar N (1994) On some integer-valued autoregressive moving average models. J Multivar Anal 50(1):132–151

Alzaid A, Al-Osh M (1990) An integer-valued pth-order autoregressive structure (INAR(p)) process. J Appl Probab 27(2):314–324

Alzaid AA, Al-Osh M (1993) Some autoregressive moving average processes with generalized Poisson marginal distributions. Ann Inst Stat Math 45(2):223–232

Barreto-Souza W (2015) Zero-modified geometric INAR (1) process for modelling count time series with deflation or inflation of zeros. J Time Ser Anal 36(6):839–852

Biswas A, Song PX-K (2009) Discrete-valued ARMA processes. Stat Probab Lett 79(17):1884–1889

Brännäs K, Hellström J (2001) Generalized integer-valued autoregression. Econom Rev 20(4):425–443

Cohen AC Jr (1960) Estimating the parameters of a modified Poisson distribution. J Am Stat Assoc 55(289):139–143

Czado C, Gneiting T, Held L (2009) Predictive model assessment for count data. Biometrics 65(4):1254–1261

Ilić AVM (2016) A geometric time series model with a new dependent Bernoulli counting series. Commun Stat Theory Methods 45(21):6400–6415

Jacobs PA, Lewis PA (1978) Discrete time series generated by mixtures. I: Correlational and runs properties. J R Stat Soc Ser B (Methodol) 40(1):94–105

Jazi MA, Jones G, Lai C-D (2012) First-order integer valued AR processes with zero inflated Poisson innovations. J Time Series Anal 33(6):954–963

Johnson NL, Kotz S, Kemp AW (2005) Univariate discrete distributions. Wiley, New York

Karlsen H, Tjøstheim D (1988) Consistent estimates for the NEAR (2) and NLAR (2) time series models. J R Stat Soc Ser B (Methodol) 50(2):313–320

Khoo WC (2016) Statistical modelling of time series of counts for a new class of mixture distributions/Khoo Wooi Chen. PhD thesis, University of Malaya

Khoo WC, Ong SH, Biswas A (2017) Modeling time series of counts with a new class of INAR (1) model. Stat Pap 58(2):393–416

McKenzie E (1985) Some simple models for discrete variate time series 1. JAWRA J Am Water Resour Assoc 21(4):645–650

McKenzie E (1986) Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. Adv Appl Probab 18(3):679–705

Miletić Ilić AV, Ristić MM, Nastić AS, Bakouch HS (2018) An INAR (1) model based on a mixed dependent and independent counting series. J Stat Comput Simul 88(2):290–304

Mohammadi Z, Sajjadnia Z, Bakouch H, Sharafi M (2022a) Zero-and-one inflated Poisson–Lindley INAR (1) process for modelling count time series with extra zeros and ones. J Stat Comput Simul 92(10):2018–2040

Mohammadi Z, Sajjadnia Z, Sharafi M, Mamode Khan N (2022b) Modeling medical data by flexible integer-valued AR (1) process with zero-and-one-inflated geometric innovations. Iran J Sci Technol Trans A Sci 46(3):891–906

Mohammadpour M, Bakouch HS, Shirozhan M (2018) Poisson–Lindley INAR (1) model with applications. Braz J Probab Stat 32(2):262–280

Mood AM (1940) The distribution theory of runs. Ann Math Stat 11(4):367–392

Nastić AS, Ristić MM, Bakouch HS (2012) A combined geometric INAR (p) model based on negative binomial thinning. Math Comput Model 55(5–6):1665–1672

Nastić AS, Ristić MM, Ilić AVM (2017) A geometric time-series model with an alternative dependent Bernoulli counting series. Commun Stat Theory Methods 46(2):770–785

Pegram G (1980) An autoregressive model for multilag Markov chains. J Appl Probab 17(2):350–362

Puig P, Valero J (2006) Count data distributions: some characterizations with applications. J Am Stat Assoc 101(473):332–340

Qi X, Li Q, Zhu F (2019) Modeling time series of count with excess zeros and ones based on INAR (1) model with zero-and-one inflated Poisson innovations. J Comput Appl Math 346:572–590

Ristić MM, Nastić AS (2012) A mixed INAR (p) model. J Time Ser Anal 33(6):903–915

Ristić MM, Bakouch HS, Nastić AS (2009) A new geometric first-order integer-valued autoregressive (NGINAR (1)) process. J Stat Plann Inference 139(7):2218–2226

Ristić MM, Nastić AS, Miletić Ilić AV (2013) A geometric time series model with dependent Bernoulli counting series. J Time Ser Anal 34(4):466–476

Schweer S, Weiß CH (2014) Compound Poisson INAR (1) processes: stochastic properties and testing for overdispersion. Comput Stat Data Anal 77:267–284

Shirozhan M, Mohammadpour M (2018) A new class of INAR (1) model for count time series. J Stat Comput Simul 88(7):1348–1368

Shirozhan M, Mohammadpour M, Bakouch HS (2019) A new geometric INAR (1) model with mixing Pegram and generalized binomial thinning operators. Iran J Sci Technol Trans A Sci 43(3):1011–1020

Steutel FW, van Harn K (1979) Discrete analogues of self-decomposability and stability. Ann Probab 7(5):893–899

Weiß CH (2009) Jumps in binomial AR (1) processes. Stat Probab Lett 79(19):2012–2019

Springer SBMAC