

# Quasi-likelihood inference for self-exciting threshold integer-valued autoregressive processes

Han Li<sup>1</sup> · Kai Yang<sup>1</sup> · Dehui Wang<sup>1</sup>

Received: 21 November 2016 / Accepted: 27 June 2017 / Published online: 6 July 2017  
© Springer-Verlag GmbH Germany 2017

**Abstract** This article redefines the self-exciting threshold integer-valued autoregressive (SETINAR(2,1)) processes under a weaker condition that the second moment is finite, and studies the quasi-likelihood inference for the new model. The ergodicity of the new processes is discussed. Quasi-likelihood estimators for the model parameters and the asymptotic properties are obtained. Confidence regions of the parameters based on the quasi-likelihood method are given. A simulation study is conducted for the evaluation of the proposed approach and an application to a real data example is provided.

**Keywords** SETINAR process · Integer-valued threshold models · Confidence region

## 1 Introduction

There has been an increasing interest in developing models for time series of (small) counts because of their wide range of applications, including epidemiology, finance, disease modeling, etc. The majority of these models are based on the thinning operators, see [Al-Osh and Alzaid \(1987\)](#), [Du and Li \(1991\)](#), [Ristić et al. \(2009\)](#), [Zhang et al. \(2010\)](#) and [Li et al. \(2015\)](#), among others. [Weiß \(2008\)](#) and [Scotto et al. \(2015\)](#) gave detailed reviews of the development of integer-valued time series models. Among the above models, the class of Poisson integer-valued autoregressive (INAR) moving average models ([Al-Osh and Alzaid 1987, 1991, 1992; Du and Li 1991](#)) play a central role. However, when dealing with the nonlinear time series of counts such as volatility

---

✉ Dehui Wang  
wangdh@jlu.edu.cn

<sup>1</sup> School of Mathematics, Jilin University, Changchun 130012, Jilin, People's Republic of China

changes in time, high threshold exceedances appearing in clusters, and the so-called piecewise phenomenon, such models will not work well.

On modelling the non-linear phenomena, especially the piecewise phenomenon, the threshold models (Tong 1978; Tong and Lim 1980) have attracted much attention and have been widely used in diverse areas. For dealing with time series of counts exhibiting piecewise-type patterns, Monteiro et al. (2012) introduced a class of SETINAR(2,1) models driven by a Poisson distribution. Wang et al. (2014) proposed a self-excited threshold Poisson autoregressive (SETPAR) model and applied it to the world major earthquakes data. Yang et al. (2017) proposed an integer-valued threshold autoregressive process based on negative binomial thinning operator (NBTINAR(1)), and compared the performances of the above threshold models. Möller and Weiß (2015) presented a brief survey of threshold models of integer-valued time series.

A drawback of SETINAR(2,1) model proposed by Monteiro et al. (2012) is the mean and variance of Poisson distribution are equal and this property is not always found in the real data. The goal of this paper is to weaken the conditions of original SETINAR(2,1) model by removing the assumption of Poisson distribution, and to present a quasi-likelihood (QL) inference for the new SETINAR(2,1) process. Quasi-likelihood method, a nonparametric inference method was initially introduced by Wedderburn (1974). QL methods are useful because (i) they can be used in cases where exact distributional information is not available, (ii) only second moment assumptions are required and, (iii) they enjoy a certain robustness of validity. Hence the proposed method allows precise estimation of the relationship between the response and the covariate variables without requiring exact distributional assumptions. QL method has been widely applied in many fields, including generalized linear models (McCullagh and Nelder 1989; Sutardhar and Rao 2001; Lu et al. 2006), stochastic volatility models (Ruiz 1994), semiparametric models (Severini and Staniswalis 1994), median regression models (Jung 1996), autoregressive models (Azrak and Mélard 1998, 2006; Ling 2007), nonstationary time series models (Kim and Park 2008; Aue and Horváth 2011), integer-valued time series models (Zheng et al. 2006a, b; Niaparast and Schwabe 2013; Christou and Fokianos 2014), among others.

The rest of this paper is organized as follows. In Sect. 2, we redefine the SETINAR(2,1) processes, and consider the quasi-likelihood inference for the unknown parameters of interest. In Sect. 3, some numerical results of the estimates are presented. In Sect. 4, we give an application of the proposed QL method to a real data set. Some concluding remarks are given in Sect. 5. All proofs are postponed to “Appendix”.

## 2 Main results

The SETINAR(2,1) process (originally proposed by Monteiro et al. 2012), is defined by the following recursive equation:

$$X_t = I_{1,t}(\alpha_1 \circ X_{t-1}) + I_{2,t}(\alpha_2 \circ X_{t-1}) + Z_t, \quad (2.1)$$

where

- (i)  $I_{1,t} := I\{X_{t-1} \leq r\}$ ,  $I_{2,t} := 1 - I_{1,t} = I\{X_{t-1} > r\}$ ,  $r$  is the known threshold variable;
- (ii) The thinning operator “ $\circ$ ”, introduced in [Steutel and Harn \(1979\)](#), is defined as

$$\alpha_i \circ X = \sum_{k=1}^X B_k^{(i)}, \quad \alpha_i \in (0, 1),$$

- where  $\{B_k^{(i)}\}$  is a sequence of independent and identically distributed (i.i.d.) Bernoulli random variables independent of  $X$ , satisfying  $P(\alpha_i = 1) = 1 - P(\alpha_i = 0) = \alpha_i$ ,  $i = 1, 2$ ;
- (iii)  $\{Z_t\}$  is a sequence of i.i.d. random variables with  $E(Z_t) = \lambda$  and  $\text{Var}(Z_t) = \sigma_z^2 < \infty$ ;
  - (iv) For fixed  $t$  and  $i$  ( $i = 1, 2$ ),  $Z_t$  is assumed to be independent of counting series in  $\{\alpha_i \circ X_{t-l}, l \geq 1\}$  and  $\{X_{t-l}, l \geq 1\}$ .

*Remark 2.1* In [Monteiro et al. \(2012\)](#), the SETINAR(2,1) process is defined with  $Z_t$  follows a Poisson distribution with mean  $\lambda$ . In this paper, we remove the assumption of Poisson distribution (in (iii)) and use  $E(Z_t) = \lambda$  and  $\text{Var}(Z_t) = \sigma_z^2 < \infty$  instead, so that the model is more flexible.

*Remark 2.2* The conditional expectation and conditional variance of the SETINAR(2,1) process are given by

- (i)  $E(X_t|X_{t-1}) = \alpha_1 X_{t-1} I_{1,t} + \alpha_2 X_{t-1} I_{2,t} + \lambda, \quad t = 1, 2, \dots$
- (ii)  $\text{Var}(X_t|X_{t-1}) = \alpha_1(1 - \alpha_1) X_{t-1} I_{1,t} + \alpha_2(1 - \alpha_2) X_{t-1} I_{2,t} + \sigma_z^2, \quad t = 1, 2, \dots$

*Remark 2.3* Following [Monteiro et al. \(2012\)](#), we assume that  $r$  is known. In practice, we can use NeSS algorithm (see, e.g., [Li and Tong 2016](#); [Yang et al. 2017](#)) to estimate  $r$  first.

The following theorem states the ergodicity of the SETINAR(2,1) process (2.1). This property will be useful in deriving the asymptotic properties of the estimators.

**Proposition 2.1** *The SETINAR(2,1) process  $\{X_t\}_{t \in \mathbb{Z}}$  defined in (2.1) is an ergodic Markov chain.*

Suppose we have a series of observations  $\{X_t\}_{t=1}^n$  generated from the SETINAR(2,1) process and we want to estimate the parameter  $\beta = (\alpha_1, \alpha_2, \lambda)^\top$ . [Monteiro et al. \(2012\)](#) considered the conditional least squares (CLS) estimation and the conditional maximum likelihood (CML) estimation of  $\beta$ . In what follows, we will consider the maximum quasi-likelihood (MQL) estimation of  $\beta$  first, and then give the QL confidence regions of  $\beta$ .

Denote  $\theta = (\theta_1, \theta_2, \sigma_z^2)^\top$  with  $\theta_i = \alpha_i(1 - \alpha_i), i = 1, 2$ , then the variance of  $X_t$  conditional on  $X_{t-1}$  fixed is given by

$$V_\theta(X_t|X_{t-1}) = \text{Var}(X_t|X_{t-1}) = \theta_1 X_{t-1} I_{1,t} + \theta_2 X_{t-1} I_{2,t} + \sigma_z^2. \quad (2.2)$$

As discussed in Wedderburn (1974), we have the standard QL estimating equations:

$$\sum_{t=1}^n \mathbf{m}_t(\boldsymbol{\beta}) = \mathbf{0}, \tag{2.3}$$

where

$$\begin{aligned} \mathbf{m}_t(\boldsymbol{\beta}) &= (m_{1t}(\boldsymbol{\beta}), m_{2t}(\boldsymbol{\beta}), m_{3t}(\boldsymbol{\beta}))^T, \quad m_{it}(\boldsymbol{\beta}) \\ &= \frac{X_t - E(X_t|X_{t-1})}{V_{\theta}(X_t|X_{t-1})} \frac{\partial E(X_t|X_{t-1})}{\partial \beta_i}, \quad i = 1, 2, 3, \end{aligned}$$

with  $\beta_1 = \alpha_1, \beta_2 = \alpha_2, \beta_3 = \lambda$ . Note that the presence of  $\theta$  in the expression for the conditional variance makes the estimating equations (2.3) complicated and intractable in the general case. Consequently, we propose substituting a suitable consistent estimator  $\hat{\theta}$  of  $\theta$  obtained by other means and then solve the resulting MQL estimating equations for the primary parameters of interest. This approach leads to the following closed form estimator of  $\boldsymbol{\beta}$ :

$$\hat{\boldsymbol{\beta}}_{MQL} = \mathbf{Q}_n^{-1} \mathbf{q}_n, \tag{2.4}$$

where

$$\mathbf{Q}_n = \begin{pmatrix} \sum V_{\hat{\theta}}^{-1}(X_t|X_{t-1})I_{1,t}X_{t-1}^2 & 0 & \sum V_{\hat{\theta}}^{-1}(X_t|X_{t-1})I_{1,t}X_{t-1} \\ 0 & \sum V_{\hat{\theta}}^{-1}(X_t|X_{t-1})I_{2,t}X_{t-1}^2 & \sum V_{\hat{\theta}}^{-1}(X_t|X_{t-1})I_{2,t}X_{t-1} \\ \sum V_{\hat{\theta}}^{-1}(X_t|X_{t-1})I_{1,t}X_{t-1} & \sum V_{\hat{\theta}}^{-1}(X_t|X_{t-1})I_{2,t}X_{t-1} & \sum V_{\hat{\theta}}^{-1}(X_t|X_{t-1}) \end{pmatrix},$$

and

$$\mathbf{q}_n = \begin{pmatrix} \sum_{t=1}^n V_{\hat{\theta}}^{-1}(X_t|X_{t-1})I_{1,t}X_tX_{t-1} \\ \sum_{t=1}^n V_{\hat{\theta}}^{-1}(X_t|X_{t-1})I_{2,t}X_tX_{t-1} \\ \sum_{t=1}^n V_{\hat{\theta}}^{-1}(X_t|X_{t-1})X_t \end{pmatrix}.$$

To study the asymptotic behaviour of the estimator, we make the following assumptions:

- (C1)  $\{X_t\}$  is a stationary process.
- (C2)  $E|X_t|^4 < \infty$ .

We now state the asymptotic properties of MQL-estimators  $\hat{\boldsymbol{\beta}}_{MQL}$ . Theorem 2.1 shows the MQL-estimators  $\hat{\boldsymbol{\beta}}_{MQL}$  is asymptotic normally distributed if  $\theta$  can be estimated consistently; Theorem 2.2 constructs a consistent estimator of  $\theta$ .

**Theorem 2.1** *Under the assumptions (C1)–(C2), the MQL-estimators  $\hat{\boldsymbol{\beta}}_{MQL}$  is asymptotically normal,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{MQL} - \boldsymbol{\beta}) \xrightarrow{L} N(\mathbf{0}, \mathbf{T}^{-1}(\boldsymbol{\theta})),$$

where

$$\begin{aligned}
 \mathbf{T}(\boldsymbol{\theta}) &= (T_{ij}(\boldsymbol{\theta}))_{3 \times 3}, \quad T_{ij}(\boldsymbol{\theta}) = T_{ji}(\boldsymbol{\theta}), \quad 1 \leq i \leq j \leq 3; \\
 T_{ii}(\boldsymbol{\theta}) &= E[V_{\boldsymbol{\theta}}^{-1}(X_1|X_0)X_0^2I_{i,1}] \quad (i = 1, 2), \quad T_{33}(\boldsymbol{\theta}) = E[V_{\boldsymbol{\theta}}^{-1}(X_1|X_0)]; \\
 T_{12}(\boldsymbol{\theta}) &= 0, \quad T_{13}(\boldsymbol{\theta}) = E[V_{\boldsymbol{\theta}}^{-1}(X_1|X_0)X_0I_{1,1}], \quad T_{23}(\boldsymbol{\theta}) = E[V_{\boldsymbol{\theta}}^{-1}(X_1|X_0)X_0I_{2,1}],
 \end{aligned}$$

Furthermore, the matrix  $\mathbf{T}(\boldsymbol{\theta})$  can be estimated consistently by

$$\widehat{\mathbf{T}}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n \mathbf{m}_t(\hat{\boldsymbol{\beta}}_{MQL}) \mathbf{m}_t(\hat{\boldsymbol{\beta}}_{MQL})^T. \tag{2.5}$$

*Remark 2.4* Since the MQL-estimators  $\hat{\boldsymbol{\beta}}_{MQL}$  given in (2.4) is the result from a two-stage estimation procedure mentioned above, the true asymptotic variance is supposed to be larger than  $\mathbf{T}(\boldsymbol{\theta})$  in Theorem 2.1. Fortunately, the estimator  $\widehat{\mathbf{T}}(\boldsymbol{\theta})$  given in (2.5) can give a estimate of the true asymptotic variance consistently.

Note that the consistency of  $\hat{\boldsymbol{\beta}}_{MQL}$  follows readily from the above result.

**Theorem 2.2** *Under the assumptions (C1)–(C2), the following estimators are consistent:*

$$\hat{\sigma}_z^2 = \frac{1}{n} \sum_{t=1}^n (X_t - \hat{\alpha}_1 X_{t-1} I_{1,t} - \hat{\alpha}_2 X_{t-1} I_{2,t} - \hat{\lambda})^2 - \frac{1}{n} \sum_{i=1}^2 \sum_{t=1}^n \hat{\alpha}_i (1 - \hat{\alpha}_i) X_{t-1} I_{i,t}, \tag{2.6}$$

and

$$\hat{\boldsymbol{\theta}} = (\hat{\alpha}_1(1 - \hat{\alpha}_1), \hat{\alpha}_2(1 - \hat{\alpha}_2), \hat{\sigma}_z^2)^T, \tag{2.7}$$

where  $\hat{\alpha}_i$  and  $\hat{\lambda}$  are consistent estimators of  $\alpha_i$  ( $i = 1, 2$ ) and  $\lambda$ . In practice, we can use the CLS-estimators of  $\alpha_i$  ( $i = 1, 2$ ) and  $\lambda$  given in Theorem 3.1 of [Monteiro et al. \(2012\)](#).

Theorem 2.2 gives a consistent estimator  $\hat{\boldsymbol{\theta}}$  of  $\boldsymbol{\theta}$  which depends on some consistent estimators  $\hat{\alpha}_i$  ( $i = 1, 2$ ) and  $\hat{\lambda}$ . Sometimes we may be interested in the variance of  $\hat{\boldsymbol{\theta}}$ . Since the consistent estimators  $\hat{\alpha}_i$  ( $i = 1, 2$ ) and  $\hat{\lambda}$  may have many forms, we refer to the moving block bootstrap (MBB) methods ([Kreiss and Lahiri 2012](#)) to estimate the variance of  $\hat{\boldsymbol{\theta}}$ . Suppose the sample  $(X_1, X_2, \dots, X_n)$  is observed. Let  $\ell$  be an integer satisfying  $1 < \ell < n$ . We can use the following MBB method to estimate the variance of  $\hat{\boldsymbol{\theta}}$ :

*Step 1* Define the overlapping blocks  $\mathbb{B}_1, \dots, \mathbb{B}_N$  of length  $\ell$  as,  $\mathbb{B}_1 = (X_1, X_2, \dots, X_{\ell})$ ,  $\mathbb{B}_2 = (X_2, \dots, X_{\ell}, X_{\ell+1})$ ,  $\dots$ ,  $\mathbb{B}_N = (X_{n-\ell+1}, \dots, X_n)$ , where  $N = n - \ell + 1$ ;

*Step 2* To generate the MBB samples, select  $b$  blocks at random with replacement from the collection  $\{\mathbb{B}_1, \dots, \mathbb{B}_N\}$ . Denote the selected MBB samples by  $(X_1^*, X_2^*, \dots, X_b^*)$ ;

*Step 3* For  $j = 1, \dots, b$ , calculate the  $\hat{\theta}_j$  by (2.6) and (2.7) with the MBB sample  $X_j^*$ ;

*Step 4* Calculate the variance of  $\hat{\theta}$  by

$$\text{Var}(\hat{\theta}) = \frac{1}{b} \sum_{j=1}^b (\hat{\theta}_j - \bar{\theta})(\hat{\theta}_j - \bar{\theta})^T, \quad \bar{\theta} = \sum_{j=1}^b \hat{\theta}_j. \quad (2.8)$$

Next, we give the QL confidence region of  $\beta$  based on the above theorems.

**Theorem 2.3** *Under the assumptions (C1)–(C2), for  $0 < \delta < 1$ , the 100(1 -  $\delta$ )% confidence region of  $\beta$  is given by:*

$$C_{\beta}\{\delta\} = \{\beta \in \mathbb{R}^3 : n(\hat{\beta}_{MQL} - \beta)^T \widehat{T}(\hat{\theta})(\hat{\beta}_{MQL} - \beta) \leq \chi_{(3)}^2(\delta)\}, \quad (2.9)$$

where  $\chi_3^2(\delta)$  denotes the  $\delta$ -upper quantile of  $\chi^2$  distribution with degrees of freedom 3.

Equation (2.9) is called the normal approximation (NA) confidence region based on MQL method. By Theorems 3.1 and 3.3 in Monteiro et al. (2012), we can easily construct the NA confidence regions based on CLS and CML methods. In the next section, we will compare the performances of these confidence regions in terms of coverage rates.

### 3 Simulation studies

To report the performances of the proposed method described in the previous section, we conduct simulation studies under the following two models:

Model I. Assume that  $Z_t$  follows a Poisson distribution with mean  $\lambda$ .

Model II. Assume that  $Z_t$  follows a Geometric distribution with probability mass function (p.m.f.) given by

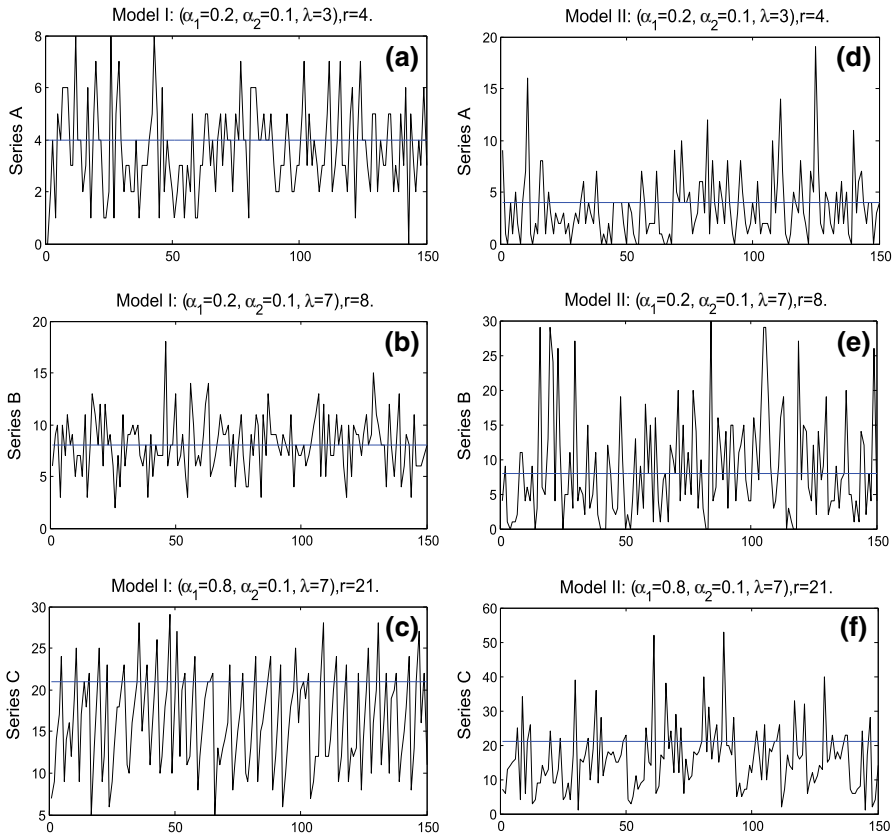
$$p(Z_t = z) = \frac{\lambda^z}{(1 + \lambda)^{z+1}}, \quad z = 0, 1, 2, \dots \quad (3.1)$$

It is easy to see that  $E(Z_t) = \lambda$ ,  $\text{Var}(Z_t) = \lambda(1 + \lambda)$ . These properties are different from Model I.

For all the following simulations, we generating the data with  $X_0 = 0$ . All simulation results are calculated under MATLAB software based on 1000 replications.

#### 3.1 Performances of the MQL-estimators

In this subsection, we first show the performances of the MQL-estimators  $\hat{\beta}_{MQL}$ , then compare the performances with the CLS-estimator  $\hat{\beta}_{CLS}$  and CML-estimators  $\hat{\beta}_{CML}$  obtained by Monteiro et al. (2012) under Models I and II. For generating the data of Models I and II, we each consider the following three series:



**Fig. 1** Sample paths for series A–C of Model I (a–c) and Model II (d–f)

Series A.  $(\alpha_1, \alpha_2, \lambda) = (0.2, 0.1, 3), r = 4$ .  
 Series B.  $(\alpha_1, \alpha_2, \lambda) = (0.2, 0.1, 7), r = 8$ .  
 Series C.  $(\alpha_1, \alpha_2, \lambda) = (0.8, 0.1, 7), r = 21$ .

*Remark 3.1* The parameter setting of the above three series are the same with models (A1), (B1) and (B3) in [Monteiro et al. \(2012\)](#). So that we can compare our simulation results with the corresponding results in [Monteiro et al. \(2012\)](#).

For each of the above series, the values of  $r$  was chosen such that the observations in each regime is at least 10% of the total sample size. As mentioned in [Li and Tong \(2016\)](#), when the proportion of observations in one regime to the whole is less than 5%, the estimate result may not be reliable. The sample paths are plotted in [Fig. 1](#). The simulation results are summarized in [Tables 1, 2, 3 and 4](#).

It is shown in [Fig. 1](#) that the threshold  $r$  of the series A and B are relatively moderate, and the threshold  $r$  of series C (especially in Model I) is slightly larger. Compare subfigures (a)–(c) and (d)–(f) we can find that the sample paths of Model II fluctuate greatly. This is because the variance of  $Z_t$  of Model II is much larger than its mean.

**Table 1** Simulation results of MQL for Models I and II: mean bias and MSE

Models	N	Para.	Series A			Series B			Series C		
			True	Mean	MSE	True	Mean	MSE	Ture	Mean	MSE
I	100	$\alpha_1$	0.2	0.194	0.030	0.2	0.195	0.036	0.8	0.804	0.005
		$\alpha_2$	0.1	0.086	0.009	0.1	0.094	0.013	0.1	0.099	0.003
		$\lambda$	3	3.028	0.229	7	7.040	1.368	7	6.957	1.046
	200	$\alpha_1$	0.2	0.196	0.016	0.2	0.202	0.018	0.8	0.801	0.003
		$\alpha_2$	0.1	0.095	0.005	0.1	0.098	0.006	0.1	0.100	0.001
		$\lambda$	3	3.021	0.115	7	7.005	0.698	7	7.001	0.548
	300	$\alpha_1$	0.2	0.196	0.011	0.2	0.199	0.014	0.8	0.801	0.002
		$\alpha_2$	0.1	0.095	0.004	0.1	0.100	0.005	0.1	0.100	0.001
		$\lambda$	3	3.016	0.085	7	7.007	0.537	7	6.990	0.374
II	100	$\alpha_1$	0.2	0.164	0.093	0.2	0.189	0.130	0.8	0.803	0.022
		$\alpha_2$	0.1	0.090	0.011	0.1	0.086	0.011	0.1	0.094	0.007
		$\lambda$	3	3.061	0.408	7	7.097	2.160	7	7.060	3.471
	200	$\alpha_1$	0.2	0.202	0.045	0.2	0.196	0.063	0.8	0.797	0.012
		$\alpha_2$	0.1	0.096	0.005	0.1	0.095	0.006	0.1	0.096	0.004
		$\lambda$	3	3.016	0.188	7	7.051	1.141	7	7.052	1.872
	300	$\alpha_1$	0.2	0.199	0.031	0.2	0.203	0.044	0.8	0.798	0.008
		$\alpha_2$	0.1	0.096	0.003	0.1	0.095	0.004	0.1	0.098	0.002
		$\lambda$	3	3.006	0.127	7	7.047	0.833	7	7.038	1.267

**Table 2** Comparison of simulation results for Series A of Model I: CLS, MQL and CML

N	Para.	CLS <sup>a</sup>		MQL		CML <sup>b</sup>	
		Mean	MSE	Mean	MSE	Mean	MSE
50	$\alpha_1 = 0.2$	0.413	0.095	0.184	0.063	0.388	0.067
	$\alpha_2 = 0.1$	0.176	0.020	0.079	0.020	0.177	0.021
	$\lambda = 3$	2.556	0.464	3.073	0.458	2.565	0.422
100	$\alpha_1 = 0.2$	0.336	0.051	0.194	0.030	0.322	0.042
	$\alpha_2 = 0.1$	0.141	0.010	0.086	0.009	0.141	0.011
	$\lambda = 3$	2.762	0.215	3.028	0.229	2.765	0.201
200	$\alpha_1 = 0.2$	0.280	0.031	0.196	0.016	0.276	0.025
	$\alpha_2 = 0.1$	0.120	0.005	0.095	0.005	0.121	0.005
	$\lambda = 3$	2.880	0.113	3.021	0.115	2.873	0.107
500	$\alpha_1 = 0.2$	0.222	0.011	0.200	0.007	0.221	0.010
	$\alpha_2 = 0.1$	0.101	0.002	0.099	0.002	0.101	0.002
	$\lambda = 3$	2.991	0.049	3.005	0.048	2.986	0.045

<sup>a</sup> Remark: This column are the CLS results for model (A1) in [Monteiro et al. \(2012\)](#).

<sup>b</sup> Remark: This column are the CML results for model (A1) in [Monteiro et al. \(2012\)](#)



**Table 3** Comparison of simulation results for Series B of Model I: CLS, MQL and CML

N	Para.	CLS <sup>a</sup>		MQL		CML <sup>b</sup>	
		Mean	MSE	Mean	MSE	Mean	MSE
50	$\alpha_1 = 0.2$	0.470	0.120	0.198	0.200	0.428	0.085
	$\alpha_2 = 0.1$	0.212	0.027	0.091	0.082	0.197	0.020
	$\lambda = 7$	5.630	3.431	7.039	8.075	5.831	2.462
100	$\alpha_1 = 0.2$	0.352	0.052	0.195	0.036	0.341	0.043
	$\alpha_2 = 0.1$	0.162	0.013	0.094	0.013	0.160	0.011
	$\lambda = 7$	6.252	1.442	7.040	1.368	6.293	1.197
200	$\alpha_1 = 0.2$	0.263	0.019	0.202	0.018	0.266	0.018
	$\alpha_2 = 0.1$	0.124	0.005	0.098	0.006	0.126	0.005
	$\lambda = 7$	6.700	0.579	7.005	0.698	6.677	0.563
500	$\alpha_1 = 0.2$	0.212	0.008	0.199	0.008	0.215	0.007
	$\alpha_2 = 0.1$	0.102	0.003	0.098	0.003	0.104	0.002
	$\lambda = 7$	6.965	0.286	7.007	0.334	6.943	0.254

<sup>a</sup> Remark: This column are the CLS results for model (B1) in [Monteiro et al. \(2012\)](#).

<sup>b</sup> Remark: This column are the CML results for model (B1) in [Monteiro et al. \(2012\)](#)

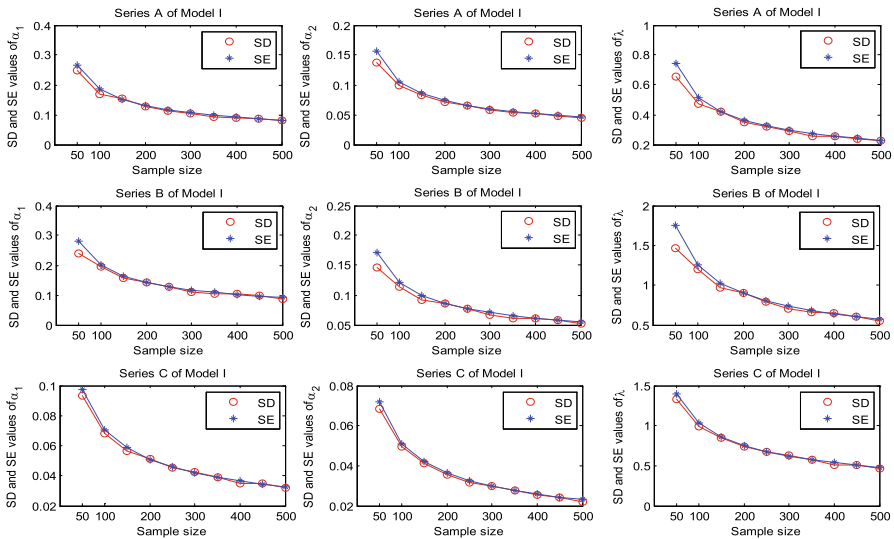
**Table 4** Comparison of simulation results for Series C of Model I: CLS, MQL and CML

N	Para.	CLS <sup>a</sup>		MQL		CML <sup>a</sup>	
		Mean	MSE	Mean	MSE	Mean	MSE
50	$\alpha_1 = 0.8$	0.821	0.008	0.805	0.008	0.814	0.004
	$\alpha_2 = 0.1$	0.112	0.004	0.099	0.004	0.109	0.002
	$\lambda = 7$	6.733	1.684	6.948	1.597	6.808	0.935
100	$\alpha_1 = 0.8$	0.804	0.006	0.804	0.005	0.807	0.002
	$\alpha_2 = 0.1$	0.103	0.003	0.099	0.003	0.105	0.001
	$\lambda = 7$	6.947	1.213	6.957	1.046	6.897	0.549
200	$\alpha_1 = 0.8$	0.799	0.003	0.801	0.003	0.801	0.001
	$\alpha_2 = 0.1$	0.098	0.001	0.100	0.001	0.100	0.001
	$\lambda = 7$	7.025	0.549	7.001	0.548	6.986	0.269
500	$\alpha_1 = 0.8$	0.800	0.001	0.800	0.001	0.801	4.9e-04
	$\alpha_2 = 0.1$	0.100	0.001	0.101	0.001	0.100	3.4e-04
	$\lambda = 7$	6.998	0.233	6.997	0.233	6.988	0.114

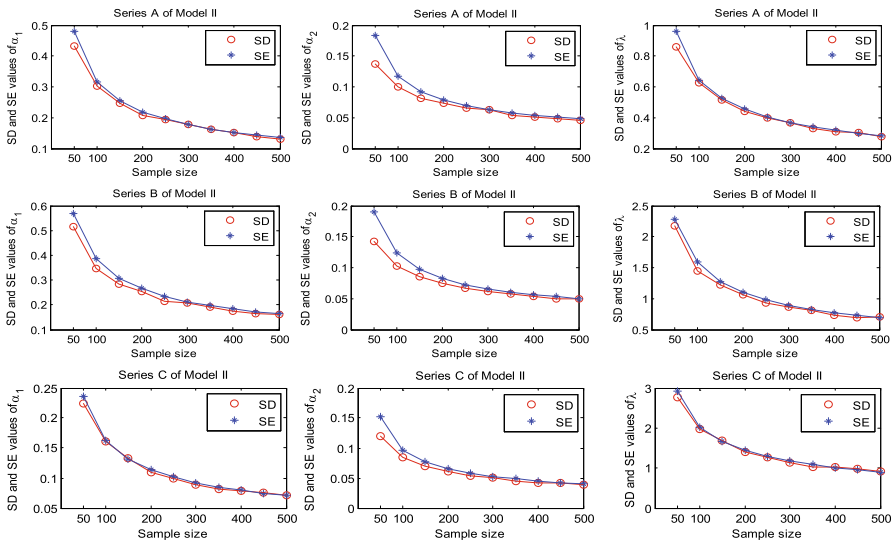
<sup>a</sup> Remark: This column are the CLS results for model (B3) in [Monteiro et al. \(2012\)](#).

<sup>b</sup> Remark: This column are the CML results for model (B3) in [Monteiro et al. \(2012\)](#)

Table 1 reports the means and mean square errors (MSE) of the MQL-estimators  $\hat{\beta}_{MQL}$ . From Table 1 we can find that all the simulation results perform better as  $N$  increases, which imply that our estimators are consistent for all the parameters. Most of the biases (the means minus the corresponding true values) and MSE of Model II



**Fig. 2** Plot for SD and SE values of Model I: Series A–C



**Fig. 3** Plot for SD and SE values of Model II: Series A to C

are bigger than those in Model I. This is because the variances of  $Z_t$  in Model II are bigger than in Model I.

Figures 2 and 3 show the standard deviations (SD) of the MQL-estimators  $\hat{\beta}_{MQL}$  across 1000 replications, and the standard errors (SE) of the MQL-estimators  $\hat{\beta}_{MQL}$ . The SE are calculated by the mean of the square roots of the asymptotic variances divided by  $N$ . From Figs. 2 and 3 we can find that the SD and SE of the  $\hat{\beta}_{MQL}$  are very

**Table 5** Comparison of simulation results for Series D of Model I: CLS, MQL and CML

N	Para.	CLS		MQL		CML	
		Mean	MSE	Mean	MSE	Mean	MSE
50	$\alpha_1 = 0.2$	0.190	0.023	0.207	0.021	0.208	0.028
	$\alpha_2 = 0.8$	0.778	0.003	0.783	0.003	0.786	0.003
	$\lambda = 7$	7.637	2.344	7.483	2.014	7.444	1.889
100	$\alpha_1 = 0.2$	0.194	0.012	0.204	0.011	0.200	0.013
	$\alpha_2 = 0.8$	0.789	0.001	0.793	0.001	0.793	0.001
	$\lambda = 7$	7.319	1.096	7.222	0.985	7.228	0.905
200	$\alpha_1 = 0.2$	0.190	0.009	0.195	0.008	0.195	0.007
	$\alpha_2 = 0.8$	0.793	0.001	0.795	0.001	0.796	0.001
	$\lambda = 7$	7.191	0.647	7.143	0.591	7.124	0.474
500	$\alpha_1 = 0.2$	0.190	0.005	0.192	0.005	0.198	0.003
	$\alpha_2 = 0.8$	0.796	0.000	0.797	0.000	0.800	0.000
	$\lambda = 7$	7.103	0.390	7.085	0.364	7.004	0.183

close, indicating that the MQL method converges fast. The values of SD and SE in Model II (shown in Fig. 3) are bigger than the corresponding series in Model I (shown in Fig. 2), indicating that bigger variances of  $Z_t$  will result slower convergence rate of the estimate.

Tables 2, 3 and 4 show the simulation results based on different methods. As can be seen in Table 2 that most biases of  $\hat{\beta}_{MQL}$  are smaller than those of CLS ones, almost all of the MSE of  $\hat{\beta}_{MQL}$  are smaller than the  $\hat{\beta}_{CLS}$ . The CML method considered by Monteiro et al. (2012) can be used as a benchmark here. From Table 2 we can see that the MSE of the MQL and CML methods is basically the same, when the sample size is small, most of biases of MQL estimates are less than the CML estimates, which indicates that the MQL results are credible. We can obtain similar conclusions in Tables 3 and 4. In addition, when  $n = 50$  in Table 3, the MSE of the MQL-estimator  $\hat{\beta}_{MQL}$  is relatively large. This may be because the QL method uses a two-step estimation. Before estimating the parameters of interest, the variance  $\sigma_z^2$  of  $Z_t$  needs to be estimated. When the sample size is small, the estimator  $\hat{\sigma}_z^2$  is occasionally less effective. Based on the above simulation results, we conclude that the MQL-estimators  $\hat{\beta}_{CML}$  is better than the CLS-estimators, and the CML method is not unanimously better than the MQL method.

Considering that different samples may have some impact on the estimates, we consider another simulations based on another series of Model I, i.e., Series D:  $(\alpha_1, \alpha_2, \lambda) = (0.2, 0.8, 7)$ ,  $r = 16$ . In Series D, we calculate the CLS, MQL and CML estimates under the same samples. The simulation results are summarized in Table 5. As is shown in Table 5, the MQL estimates perform better than the CLS ones, and the CML estimates are not unanimously better than the MQL ones, i.e., we can get the same conclusions from Table 5 as shown in Tables 2, 3 and 4.

Next we continue to show the performances of the variance of  $\hat{\theta}$  calculated by (2.8). For generating the bootstrap samples, we choose  $\ell = 80\%$  of the total sample size and  $b = 500$ . The simulation results are summarized in Table 6. We can see from Table 6 that all variances  $\text{Var}(\hat{\theta})$  convergence to zero as  $N$  increases, implying the  $\hat{\theta}$  is consistent. Further, we can find that the variances  $\text{Var}(\hat{\theta})$  of each series of Model II are bigger than the corresponding series in Model I. This is because the true variance of Model II is much bigger than Model I. To study the sample size that can make the variance of  $\hat{\sigma}_z^2$  in Model II close to zero, we do some additional simulations. The simulation results (not given here) show that more than 5000 samples are needed, meaning that the convergence rate of  $\text{Var}(\hat{\sigma}_z^2)$  in Model II is much slower than that of Model I.

### 3.2 Coverage rates of the confidence regions

In this subsection, we first consider the performances of the coverage rates of the QL confidence regions, then compare the simulation results of QL confidence regions with NA confidence regions based on CLS and CML methods in terms of coverage rates.

To show the performances of QL confidence regions given in (2.9), we consider the same series of Models I and II mentioned above. Table 7 reports the coverage rates of the QL confidence regions under Models I and II based on confidence levels 0.90 and 0.95, respectively. It can be seen from Table 7 that as the sample size  $N$  increases, the coverage rates of three series (in each model) all increase. These results indicate that the QL confidence regions perform well in practice.

Figure 4 plots the QL confidence regions for Series A of Model I based on confidence level 95% with sample size  $N = 300$ , while Fig. 5 plots the 95% QL confidence regions for Model II under the same sample size and the same series. From Figs. 4 and 5 we can see that the points which fall within the confidence regions (blue dots) are dense near the center of the confidence regions and are sparse at the edges. There are about 4% of the points (red plus) scattered outside the confidence regions, which is consistent with the confidence level 95%. Similar results are obtained for other series of Models I and II, but are not given here.

To demonstrate the robustness of the QL method, we consider the following mixed model:

Model III: Assume that  $Z_t$  has the following mixed structure:

$$Z_t = \delta_t Z_{1t} + (1 - \delta_t) Z_{2t}, \quad (3.2)$$

where  $Z_{1t}$  follows a Poisson distribution with mean  $\lambda$ ,  $Z_{2t}$  follows a Geometric distribution with p.m.f. given in (3.1),  $\delta_t$  follows a Bernoulli distribution with p.m.f. given by  $P(\delta_t = 1) = 1 - P(\delta_t = 0) = \gamma$ , meaning that  $Z_{1t}$  is contaminated by  $Z_{2t}$  with probability  $1 - \gamma$ . Calculate to see that  $E(Z_t) = \lambda$ ,  $\text{Var}(Z_t) = \lambda^2(1 - \gamma) + \lambda$ .

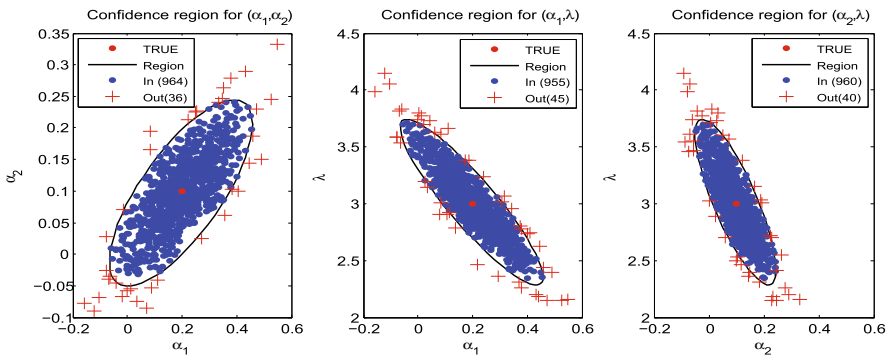
For generating the data, we set the parameters setting the same as Series A, B and C mentioned above. The coefficient  $\gamma$  is chosen as  $\gamma = 1, 0.95, 0.90$ , respectively, so that we can compare the performances of NA confidence regions based on CLS, MQM and CML methods. Table 8 shows the simulation results of the coverage rates of

**Table 6** Estimation the variances of  $\hat{\theta}$  of Models I and II via the MBB method

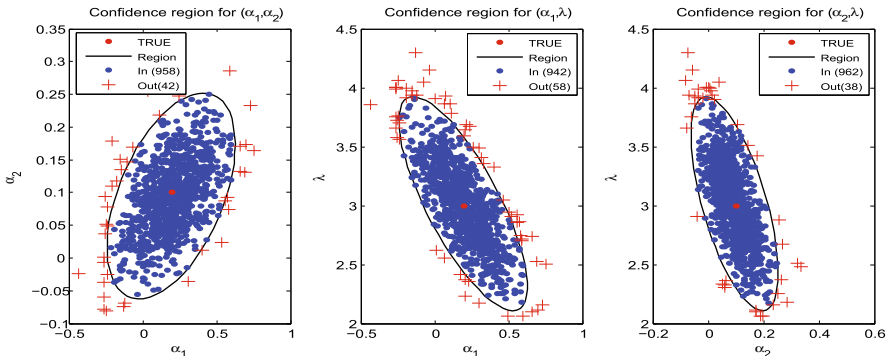
N	Var( $\hat{\theta}$ )	Model I				Model II			
		Series A		Series B		Series C		Series D	
		Series A	Series B	Series C	Series D	Series A	Series B	Series C	Series D
100	Var( $\hat{\theta}_1$ )	0.002	0.003	0.000	0.002	0.008	0.013	0.001	0.045
	Var( $\hat{\theta}_2$ )	0.001	0.001	0.000	0.000	0.001	0.001	0.001	0.000
	Var( $\hat{\sigma}_2^2$ )	0.041	0.262	0.182	0.246	1.219	23.332	22.621	20.217
	Var( $\hat{\theta}_1$ )	0.001	0.001	0.000	0.001	0.002	0.003	0.000	0.003
300	Var( $\hat{\theta}_2$ )	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Var( $\hat{\sigma}_2^2$ )	0.014	0.082	0.064	0.091	0.382	9.810	8.713	9.322
	Var( $\hat{\theta}_1$ )	0.000	0.000	0.000	0.001	0.001	0.001	0.000	0.001
	Var( $\hat{\theta}_2$ )	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
500	Var( $\hat{\theta}_2$ )	0.008	0.045	0.040	0.056	0.246	5.221	4.926	5.265

**Table 7** Coverage rates based on QL method for Models I and II

N	level	Model I			Model II		
		Series A	Series B	Series C	Series A	Series B	Series C
100	0.95	0.936	0.930	0.960	0.928	0.933	0.928
	0.90	0.893	0.870	0.916	0.876	0.878	0.871
200	0.95	0.940	0.944	0.954	0.943	0.942	0.931
	0.90	0.897	0.898	0.911	0.892	0.882	0.882
300	0.95	0.952	0.951	0.947	0.948	0.950	0.949
	0.90	0.899	0.899	0.892	0.901	0.901	0.897



**Fig. 4** Confidence regions for Series A of Models I with confidence level 95%



**Fig. 5** Confidence regions for Series A of Models II with confidence level 95%

the NA confidence regions based on different methods under contaminated samples. From Table 8 we can see that when the sample is not been contaminated ( $\gamma = 1$ ), the coverage rates of NA confidence regions based on CML method performs better than the other two methods. However, once the samples are contaminated ( $\gamma = 0.95, 0.90$ ), the results of the confidence regions based on CML method are no longer accurate, the corresponding coverage rates decrease rapidly. This shows that the CML method is

**Table 8** Comparison of the NA confidence regions based on different methods under Model III

Series	$r$	$\gamma = 1$			$\gamma = 0.95$			$\gamma = 0.90$		
		CLS	MQL	CML	CLS	MQL	CML	CLS	MQL	CML
(0.2,0.1,3,4)	0.95	0.956	0.959	0.949	0.949	0.952	0.935	0.959	0.954	0.897
	0.90	0.889	0.896	0.898	0.872	0.890	0.863	0.886	0.891	0.789
(0.2,0.1,7,8)	0.95	0.956	0.961	0.953	0.949	0.951	0.890	0.942	0.950	0.791
	0.90	0.882	0.896	0.906	0.878	0.896	0.765	0.878	0.888	0.580
(0.8,0.1,7,21)	0.95	0.946	0.949	0.948	0.940	0.944	0.779	0.956	0.948	0.464
	0.90	0.892	0.913	0.895	0.893	0.898	0.657	0.879	0.896	0.382

not robust. The coverage accuracy of confidence regions based on CLS method is also affected (slightly) with the decrease of  $\gamma$ . When  $\gamma$  down to 0.9, the coverage accuracy of the proposed QL method has exceeded that of the other two methods, indicating that the quasi-likelihood method is more robust than the least square method.

### 3.3 Performances of $\hat{r}$

In this subsection, we investigate the performances of  $\hat{r}$  using the NeSS algorithm proposed by [Li and Tong \(2016\)](#). [Yang et al. \(2017\)](#) used this algorithm to estimate the threshold  $r$  in the NBTINAR(1) model. Following [Li and Tong \(2016\)](#), let

$$J_n(r) = \sum_{t=1}^n \left( X_t - \frac{\sum_{t=1}^n X_t X_{t-1} - \lambda \sum_{t=1}^n X_{t-1}}{\sum_{t=1}^n X_{t-1}^2} \cdot X_{t-1} - \lambda \right)^2 - S_n(r),$$

where

$$S_n(r) = \sum_{t=1}^n \left( X_t - \sum_{k=1}^2 \frac{\sum_{t=1}^n X_t X_{t-1} I_{k,t}(r) - \lambda \sum_{t=1}^n X_{t-1} I_{k,t}}{\sum_{t=1}^n X_{t-1}^2 I_{k,t}(r)} \cdot X_{t-1} I_{k,t}(r) - \lambda \right)^2,$$

with  $I_{1,t}(r) := I\{X_{t-1} \leq r\}$ ,  $I_{2,t}(r) := 1 - I_{1,t}(r) = I\{X_{t-1} > r\}$ . Then, the threshold  $r$  can be estimated by maximizing  $J_n(r)$ , i.e.,

$$\hat{r} = \arg \max_{r \in [\underline{r}, \bar{r}]} J_n(r). \tag{3.3}$$

For more details of the NeSS algorithm, please see [Li and Tong \(2016\)](#) and [Yang et al. \(2017\)](#). [Table 9](#) reports the means, medians, MSE and the percentage which correctly estimate the threshold value of  $\hat{r}$  across 1000 replications for Series A - C of Models I and II, respectively, under sample size  $N = 100, 300, 500$ . From [Table 9](#) we can find that all the simulation results perform better as  $N$  increases implying that the algorithm is consistent. We also find that the results of  $\hat{r}$  in Model I have smaller biases, MSE

**Table 9** The performances of  $\hat{r}$  for Models I and II

Series	$N$	$r$	Model I				Model II			
			Mean	Median	MSE	Percent	Mean	Median	MSE	Percent
A.	100	4	3.540	4	2.844	0.263	3.829	4	4.331	0.199
	300	4	3.788	4	1.714	0.473	3.971	4	3.257	0.252
	500	4	3.906	4	1.110	0.583	3.904	4	2.886	0.332
B.	100	8	7.746	8	1.880	0.413	7.542	7	15.954	0.109
	300	8	7.789	8	0.543	0.675	7.794	8	9.410	0.171
	500	8	7.896	8	0.250	0.818	7.864	8	6.562	0.262
C.	100	21	20.998	21	0.002	0.998	20.998	21	0.424	0.785
	300	21	21	21	0	1	21.010	21	0.014	0.989
	500	21	21	21	0	1	21.002	21	0.002	0.998



and higher correct percentages than in Model II. This may be because the variance of Model II is larger than Model I for each series. Furthermore, we find that for the integer-valued  $r$  the median estimated to be better than the mean.

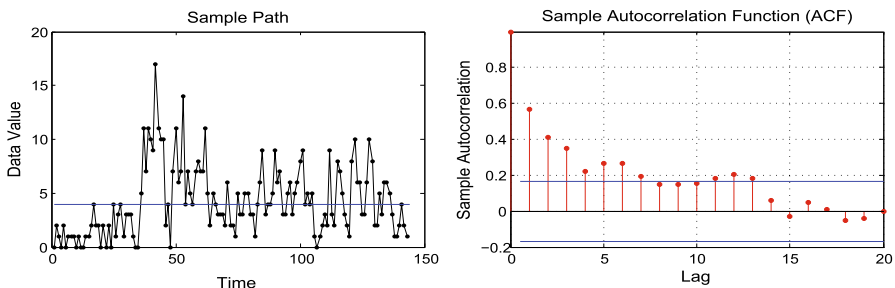
### 4 Real data example

In this section, we will use the new SETINAR(2,1) model to fit a series of criminal data, which can be downloaded from the Forecasting Principles website (<http://www.forecastingprinciples.com>). The analyzed data set is about the counts of drug in Pittsburgh, which consist of 144 monthly observations, starting from January 1990 to December 2001. The sample mean and sample variance are 4.181 and 10.597, respectively, showing considerable over-dispersion. Figure 6 shows the sample path and the ACF plot of the observations. From the sample path and the ACF plot shown in Fig. 6 we can see that the analyzed data set is a time series showing piecewise phenomenon with the threshold value  $r = 4$ . The threshold value of  $r$  is calculated by the NeSS algorithm discussed in Li and Tong (2016) and Yang et al. (2017). It also displays positive serial dependence, as can be seen in the ACF plot in Fig. 6.

We use the following competitive counts models to fit the data set, and compare different models via the AIC criterion and BIC criterion.

- i.i.d. Poisson distribution with mean  $\lambda$ , denoted by i.i.d. Poisson.
- i.i.d. Geometric distribution with p.m.f. given in (3.1), denoted by i.i.d. Geometric.
- The Poisson INAR(1) model proposed by Al-Osh and Alzaid (1987), denoted by Po-INAR(1).
- The Geometric INAR(1) model proposed by Alzaid and Al-Osh (1988), denoted by Ge-INAR(1).
- SETINAR(2,1) model with innovations  $Z_t$  follows a Poisson distribution with mean  $\lambda$ , denoted by SETINAR(2,1)-I.
- SETINAR(2,1) model with innovations  $Z_t$  follows a Geometric distribution with p.m.f. given in (3.1), denoted by SETINAR(2,1)-II.

For each model, we use the CLS (if available), MQL and CML methods to estimate the parameters, and summarized the fitting results in Table 10. As can be seen from Table 10, the fitting results of the CML, CLS and MQL of SETINAR(2,1)-II are



**Fig. 6** Sample path (left) and the ACF plot (right) of the criminal data. The blue line in the sample path estimated threshold dividing the range into two regimes

**Table 10** The fitting results of criminal data

	CLS	MQL	CML	AIC	BIC
i.i.d. Poisson	– <sup>a</sup>	$\hat{\lambda} = 4.181$	$\hat{\lambda} = 4.181$	792.283	795.253
i.i.d. Geometric	–	$\hat{\lambda} = 4.181$	$\hat{\lambda} = 4.181$	733.954	736.924
Po-INAR(1)	$\hat{\alpha} = 0.577$	$\hat{\alpha} = 0.598$	$\hat{\alpha} = 0.424$	708.607	714.547
	$\hat{\lambda} = 1.776$	$\hat{\lambda} = 1.688$	$\hat{\lambda} = 2.389$		
Ge-INAR(1)	$\hat{\alpha} = 0.578$	$\hat{\alpha} = 0.591$	$\hat{\alpha} = 0.461$	694.299	700.238
	$\hat{p} = 0.192$	$\hat{p} = 0.192$	$\hat{p} = 0.244$		
SETINAR(2,1)-I	$\hat{\alpha}_1 = 0.335$	$\hat{\alpha}_1 = 0.364$	$\hat{\alpha}_1 = 0.123$	690.919	699.808
	$\hat{\alpha}_2 = 0.545$	$\hat{\alpha}_2 = 0.563$	$\hat{\alpha}_2 = 0.446$		
	$\hat{\lambda} = 2.173$	$\hat{\lambda} = 2.083$	$\hat{\lambda} = 2.726$		
SETINAR(2,1)-II	$\hat{\alpha}_1 = 0.335$	$\hat{\alpha}_1 = 0.364$	$\hat{\alpha}_1 = 0.395$	654.135	663.023
	$\hat{\alpha}_2 = 0.545$	$\hat{\alpha}_2 = 0.563$	$\hat{\alpha}_2 = 0.503$		
	$\hat{\lambda} = 2.173$	$\hat{\lambda} = 2.083$	$\hat{\lambda} = 2.226$		

<sup>a</sup> Remark: “–” stand for not available

approximately the same, and the SETINAR(2,1)-II model takes the smallest AIC and BIC values. Thus, we can conclude that SETINAR(2,1)-II is appropriate for this data set. Moreover, this example also shows that it is necessary to extend the original SETINAR(2,1) model proposed by [Monteiro et al. \(2012\)](#).

## 5 Conclusions

This article extended the original SETINAR(2,1) model proposed by [Monteiro et al. \(2012\)](#) by removing the assumption of Poisson distribution of  $Z_t$ , and redefined the new SETINAR(2,1) process under the conditions of finite second moment of  $Z_t$ . The ergodicity of the new process is established. MQL-estimators of the model parameters and confidence regions based on QL method are derived and the asymptotic properties of the estimators are obtained. A real data example reveals that the new SETINAR(2,1) model with Geometric innovations is appropriate for the criminal data. Potential issues of future research include to test the linearity against the nonlinear model, the marginal distributions, extend the results to multivariate cases.

**Acknowledgements** This work is supported by National Natural Science Foundation of China (Nos. 11271155, 11371168, J1310022, 11571138, 11501241, 11571051, 11301137), National Social Science Foundation of China (16BTJ020), Science and Technology Research Program of Education Department in Jilin Province for the 12th Five-Year Plan (440020031139) and Jilin Province Natural Science Foundation (20150520053JH).

## Appendix

*Proof of Proposition 2.1* According to Theorem 3.1 of [Tweedie \(1975\)](#)(or Proposition 2.2 of [Zheng and Basawa 2008](#)), the sufficient condition of  $\{X_t\}$  to be ergodic is that

there exists a set  $K$  and a non-negative measurable function  $g$  on state space  $\mathbb{N}_0$  such that

$$\int_{\mathbb{N}_0} P(x, dy)g(y) \leq g(x) - 1, \quad x \notin K, \tag{5.1}$$

and for some fixed  $B$ ,

$$\int_{\mathbb{N}_0} P(x, dy)g(y) = \lambda(x) \leq B < \infty, \quad x \in K, \tag{5.2}$$

where  $P(x, A) = P(X_1 \in A|X_0 = x)$ . Let  $g(x) = x$ , we have

$$\begin{aligned} \int_{\mathbb{N}_0} g(y)dP(X_1 = y|X_0 = x_0) &= E(X_1|X_0 = x_0) = \alpha_1 x_0 I_{0,1} + \alpha_2 x_0 I_{0,2} + \lambda \\ &\leq \alpha_{\max} x_0 + \lambda, \end{aligned}$$

where  $\alpha_{\max} = \max\{\alpha_1, \alpha_2\} < 1$ . let  $N = \lceil \frac{1+\lambda}{1-\alpha_{\max}} \rceil + 1$ , where  $\lceil x \rceil$  denotes the integer part of  $x$ . Then for  $x_0 \geq N$ , we have

$$\alpha_{\max} x_0 + \lambda \leq x_0 - 1 = g(x_0) - 1,$$

and for  $0 \leq x_0 \leq N - 1$ ,

$$\int_{\mathbb{N}_0} g(y)dP(X_1 = y|X_0 = x_0) = E(X_1|X_0 = x_0) \leq \alpha_{\max} x_0 + \lambda \leq N + \lambda < \infty.$$

Let  $K = \{0, 1, \dots, N - 1\}$ , then (5.1) and (5.2) both hold which completes the proof. □

*Proof of Theorem 2.1* First, we suppose  $\theta$  is known. Let  $\mathcal{F}_t = \sigma(X_0, X_1, \dots, X_t)$  be the  $\sigma$ -field generated by  $\{X_0, X_1, \dots, X_t\}$ . For the following estimation equations:

$$S_n^{(1)}(\theta, \beta) = \sum_{t=1}^n V_{\theta}^{-1}(X_t|X_{t-1})(X_t - \alpha_1 X_{t-1} I_{1,t} - \alpha_2 X_{t-1} I_{2,t} - \lambda) I_{1,t} X_{t-1},$$

we have

$$\begin{aligned} &E[V_{\theta}^{-1}(X_t|X_{t-1})(X_t - \alpha_1 X_{t-1} I_{1,t} - \alpha_2 X_{t-1} I_{2,t} - \lambda) I_{1,t} X_{t-1} | \mathcal{F}_{t-1}] \\ &= V_{\theta}^{-1}(X_t|X_{t-1}) I_{1,t} X_{t-1} E[(X_t - \alpha_1 X_{t-1} I_{1,t} - \alpha_2 X_{t-1} I_{2,t} - \lambda) | \mathcal{F}_{t-1}] \\ &= 0, \end{aligned}$$

and

$$\begin{aligned} E[S_t^{(1)}(\boldsymbol{\theta}, \boldsymbol{\beta}) | \mathcal{F}_{t-1}] &= E[(V_{\boldsymbol{\theta}}^{-1}(X_t | X_{t-1})(X_t - \alpha_1 X_{t-1} I_{1,t} - \alpha_2 X_{t-1} I_{2,t} - \lambda) I_{1,t} X_{t-1} \\ &\quad + S_{t-1}^{(1)}(\boldsymbol{\theta}, \boldsymbol{\beta})) | \mathcal{F}_{t-1}] \\ &= S_{t-1}^{(1)}(\boldsymbol{\theta}, \boldsymbol{\beta}). \end{aligned}$$

Thus,  $\{S_t^{(1)}(\boldsymbol{\theta}, \boldsymbol{\beta}), \mathcal{F}_t, t \geq 0\}$  is a martingale. By (C1) and Theorem 1.1 in Billingsley (1961),

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^n V_{\boldsymbol{\theta}}^{-2}(X_t | X_{t-1})(X_t - \alpha_1 X_{t-1} I_{1,t} - \alpha_2 X_{t-1} I_{2,t} - \lambda)^2 X_{t-1}^2 I_{1,t} \\ &\xrightarrow{a.s.} E\left(V_{\boldsymbol{\theta}}^{-2}(X_1 | X_0)(X_1 - \alpha_1 X_0 I_{1,1} - \alpha_2 X_0 I_{2,1} - \lambda)^2 X_0^2 I_{1,1}\right) \\ &= E\left(E[V_{\boldsymbol{\theta}}^{-2}(X_1 | X_0)(X_1 - \alpha_1 X_0 I_{1,1} - \alpha_2 X_0 I_{2,1} - \lambda)^2 X_0^2 I_{1,1} | X_0]\right) \\ &= E[V_{\boldsymbol{\theta}}^{-1}(X_1 | X_0) X_0^2 I_{1,1}] \\ &= T_{11}(\boldsymbol{\theta}). \end{aligned}$$

Hence, by Corollary 3.2 in Hall and Heyde (1980) and the central limit theorem of martingale, we have,

$$\frac{1}{\sqrt{n}} S_n^{(1)}(\boldsymbol{\theta}, \boldsymbol{\beta}) \xrightarrow{L} N(0, T_{11}(\boldsymbol{\theta})).$$

Similarly,

$$S_n^{(2)}(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{t=1}^n V_{\boldsymbol{\theta}}^{-1}(X_t | X_{t-1})(X_t - \alpha_1 X_{t-1} I_{1,t} - \alpha_2 X_{t-1} I_{2,t} - \lambda) I_{2,t} X_{t-1},$$

and

$$S_n^{(3)}(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{t=1}^n V_{\boldsymbol{\theta}}^{-1}(X_t | X_{t-1})(X_t - \alpha_1 X_{t-1} I_{1,t} - \alpha_2 X_{t-1} I_{2,t} - \lambda),$$

We can verify that  $\{S_t^{(i)}(\boldsymbol{\theta}, \boldsymbol{\beta}), \mathcal{F}_t, t \geq 0\}$  ( $i = 2, 3$ ) are also martingales. Similar to the previous discussion, we have

$$\frac{1}{\sqrt{n}} S_n^{(i)}(\boldsymbol{\theta}, \boldsymbol{\beta}) \xrightarrow{L} N(0, T_{ii}(\boldsymbol{\theta})), \quad i = 2, 3.$$

By Cramer-Wold device, for any  $\mathbf{c}^T = (c_1, c_2, c_3) \in \mathbb{R}^3$  and  $(c_1, c_2, c_3) \neq (0, 0, 0)$ , we have

$$\begin{aligned} \frac{\mathbf{c}^\top}{\sqrt{n}} \begin{pmatrix} S_n^{(1)}(\boldsymbol{\theta}, \boldsymbol{\beta}) \\ S_n^{(2)}(\boldsymbol{\theta}, \boldsymbol{\beta}) \\ S_n^{(3)}(\boldsymbol{\theta}, \boldsymbol{\beta}) \end{pmatrix} &= \frac{1}{\sqrt{n}} \sum_{i=1}^3 c_i S_n^{(i)}(\boldsymbol{\theta}, \boldsymbol{\beta}) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n V_{\boldsymbol{\theta}}^{-1}(X_t|X_{t-1})(X_t - \alpha_1 X_{t-1} I_{1,t} - \alpha_2 X_{t-1} I_{2,t} - \lambda) \\ &\quad \cdot (c_1 I_{1,t} X_{t-1} + c_2 I_{2,t} X_{t-1} + c_3) \\ &\xrightarrow{L} N(0, E[V_{\boldsymbol{\theta}}^{-1}(X_1|X_0)(c_1 X_0 I_{1,1} + c_2 X_0 I_{2,1} + c_3)^2]), \end{aligned}$$

implying

$$\frac{1}{\sqrt{n}} \begin{pmatrix} S_n^{(1)}(\boldsymbol{\theta}, \boldsymbol{\beta}) \\ S_n^{(2)}(\boldsymbol{\theta}, \boldsymbol{\beta}) \\ S_n^{(3)}(\boldsymbol{\theta}, \boldsymbol{\beta}) \end{pmatrix} \xrightarrow{L} N(\mathbf{0}, \mathbf{T}(\boldsymbol{\theta})).$$

Now, we replace  $V_{\boldsymbol{\theta}}(X_t|X_{t-1})$  with  $V_{\hat{\boldsymbol{\theta}}}(X_t|X_{t-1})$ , where  $\hat{\boldsymbol{\theta}}$  is a consistent estimator of  $\boldsymbol{\theta}$ . Then we want

$$\frac{1}{\sqrt{n}} \begin{pmatrix} S_n^{(1)}(\hat{\boldsymbol{\theta}}, \boldsymbol{\beta}) \\ S_n^{(2)}(\hat{\boldsymbol{\theta}}, \boldsymbol{\beta}) \\ S_n^{(3)}(\hat{\boldsymbol{\theta}}, \boldsymbol{\beta}) \end{pmatrix} \xrightarrow{L} N(\mathbf{0}, \mathbf{T}(\boldsymbol{\theta})). \tag{5.3}$$

To prove (5.3), we need to check that

$$\frac{1}{\sqrt{n}} S_n^{(i)}(\hat{\boldsymbol{\theta}}, \boldsymbol{\beta}) - \frac{1}{\sqrt{n}} S_n^{(i)}(\boldsymbol{\theta}, \boldsymbol{\beta}) \xrightarrow{P} 0, \quad i = 1, 2, 3. \tag{5.4}$$

Let  $R_n(\boldsymbol{\theta}) = (1/\sqrt{n})S_n^{(1)}(\boldsymbol{\theta}, \boldsymbol{\beta})$ . Then for any  $\varepsilon > 0$  and  $\delta > 0$ , we have

$$\begin{aligned} P(|R_n(\hat{\boldsymbol{\theta}}) - R_n(\boldsymbol{\theta})| > \varepsilon) &\leq \sum_{i=1}^2 P(|\hat{\theta}_i - \theta_i| > \delta) + P(|\hat{\sigma}_z^2 - \sigma_z^2| > \delta) \\ &\quad + P(\sup_D |R_n(\boldsymbol{\theta}_1) - R_n(\boldsymbol{\theta})| > \varepsilon), \end{aligned}$$

where  $\boldsymbol{\theta}_1 = (\theta_1^1, \theta_2^1, \sigma_1^2)^\top$ ,  $D := \{|\theta_1^1 - \theta_1| < \delta, |\theta_2^1 - \theta_2| < \delta, |\sigma_1^2 - \sigma_z^2| < \delta\}$ . If  $\hat{\boldsymbol{\theta}}$  is a consistent estimator of  $\boldsymbol{\theta}$ , then we just need to prove that

$$P\left(\sup_D |R_n(\boldsymbol{\theta}_1) - R_n(\boldsymbol{\theta})| > \varepsilon\right) \xrightarrow{P} 0.$$

By Markov inequality,

$$\begin{aligned}
 &P\left(\sup_D |R_n(\theta_1) - R_n(\theta)| > \varepsilon\right) \\
 &\leq \frac{1}{\varepsilon^2} E\left(\sup_D (R_n(\theta_1) - R_n(\theta))^2\right) \\
 &= \frac{1}{\varepsilon^2} E\left(\sup_D \frac{1}{n} \sum_{t=1}^n (V_{\theta_1}^{-1}(X_t|X_{t-1}) - V_{\theta}^{-1}(X_t|X_{t-1}))^2 \left(X_t - \sum_{i=1}^2 \alpha_i I_{i,t} X_{t-1} - \lambda\right)^2 X_{t-1}^2 I_{1,t}\right) \\
 &= \frac{1}{\varepsilon^2} E\left(\sup_D (V_{\theta_1}^{-1}(X_1|X_0) - V_{\theta}^{-1}(X_1|X_0))^2 \left(X_1 - \sum_{i=1}^2 \alpha_i I_{i,1} X_0 - \lambda\right)^2 X_0^2 I_{1,1}\right) \\
 &= \frac{1}{\varepsilon^2} E\left(\sup_D \frac{\left(\sum_{i=1}^2 (\theta_i - \theta_i^1) X_0 I_{i,1} + (\sigma_z^2 - \sigma_1^2)\right)^2}{V_{\theta_1}^2(X_1|X_0) V_{\theta}^2(X_1|X_0)} \left(X_1 - \sum_{i=1}^2 \alpha_i I_{i,1} X_0 - \lambda\right)^2 X_0^2 I_{1,1}\right) \\
 &= \frac{1}{\varepsilon^2} E\left(\sup_D \frac{\left(\sum_{i=1}^2 (\theta_i - \theta_i^1) X_0 I_{i,1} + (\sigma_z^2 - \sigma_1^2)\right)^2}{V_{\theta_1}^2(X_1|X_0) V_{\theta}^2(X_1|X_0)} X_0^2 I_{1,1}\right) \\
 &\leq \frac{1}{\varepsilon^2} \sup_D \left\{ (\theta_1 - \theta_1^1)^2 m_1 + (\theta_2 - \theta_2^1)^2 m_2 + (\sigma_z^2 - \sigma_1^2)^2 m_3 + 2m_4 |\theta_1 - \theta_1^1| |\theta_2 - \theta_2^1| \right. \\
 &\quad \left. + 2m_5 |\theta_1 - \theta_1^1| |\sigma_z^2 - \sigma_1^2| + 2m_6 |\theta_2 - \theta_2^1| |\sigma_z^2 - \sigma_1^2| \right\} X_0^2 I_{1,1} \\
 &\leq \frac{C\delta^2}{\varepsilon^2}, \tag{5.5}
 \end{aligned}$$

where  $m_i$  ( $i = 1, 2, \dots, 6$ ) denote some finite moments of process  $\{X_t\}$ ,  $C$  is a positive constant. Similar argument can be used for  $1/\sqrt{n} S_n^{(i)}(\theta, \beta)$  ( $i = 2, 3$ ). For any fixed  $\varepsilon > 0$ , letting  $\delta \rightarrow 0$ , we get our assertion which in turn establishes (5.3).

Finally, by the ergodic theorem, we have

$$\frac{1}{n} \mathbf{Q}_n \xrightarrow{P} \mathbf{T}(\theta).$$

After some algebra, we have,

$$\hat{\beta}_{MQL} - \beta = \mathbf{Q}_n^{-1} \begin{pmatrix} S_n^{(1)}(\hat{\theta}, \beta) \\ S_n^{(2)}(\hat{\theta}, \beta) \\ S_n^{(3)}(\hat{\theta}, \beta) \end{pmatrix},$$

Therefore,

$$\sqrt{n}(\hat{\beta}_{MQL} - \beta) = \left(\frac{1}{n} \mathbf{Q}_n\right)^{-1} \frac{1}{\sqrt{n}} \begin{pmatrix} S_n^{(1)}(\hat{\theta}, \beta) \\ S_n^{(2)}(\hat{\theta}, \beta) \\ S_n^{(3)}(\hat{\theta}, \beta) \end{pmatrix} \xrightarrow{L} N(\mathbf{0}, \mathbf{T}^{-1}(\theta)). \tag{5.6}$$

The proof is complete. □

*Proof of Theorem 2.2* The proof follows by the ergodic theorem. □

*Proof of Theorem 2.3* The proof follows from Theorem 2.1. □

## References

- Al-Osh MA, Alzaid AA (1987) First-order integer-valued autoregressive (INAR(1)) process. *J Time Ser Anal* 8:261–275
- Al-Osh MA, Alzaid AA (1991) Binomial autoregressive moving average models. *Commun Stat Stoch Models* 7:261–282
- Al-Osh MA, Alzaid AA (1992) First order autoregressive time series with negative binomial and geometric marginals. *Commun Stat Theory Methods* 21:2483–2492
- Alzaid AA, Al-Osh MA (1988) First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties. *Stat Neerlandica* 42(1):53–61
- Aue A, Horváth L (2011) Quasi-likelihood estimation in stationary and nonstationary autoregressive models with random coefficients. *Stat Sin* 21:973–999
- Azrak R, Mélard G (1998) The exact quasi-likelihood of time-dependent ARMA models. *J Stat Plan Inference* 68:31–45
- Azrak R, Mélard G (2006) Asymptotic properties of quasi-maximum likelihood estimators for ARMA models with time-dependent coefficients. *Stat Inference Stoch Process* 9:279–330
- Billingsley P (1961) *Statistical inference for Markov processes*. The University of Chicago Press, Chicago
- Christou V, Fokianos K (2014) Quasi-likelihood inference for negative binomial time series models. *J Time Ser Anal* 35:55–78
- Du JG, Li Y (1991) The integer-valued autoregressive (INAR( $p$ )) model. *J Time Ser Anal* 12:129–142
- Hall P, Heyde CC (1980) *Martingale limit theory and its application*. Academic Press, New York
- Jung SH (1996) Quasi-likelihood for median regression models. *J Am Stat Assoc* 91:251–257
- Kim HY, Park Y (2008) A non-stationary integer-valued autoregressive model. *Stat Pap* 49:485–502
- Kreiss J, Lahiri SN (2012) Bootstrap methods for time series. In: Rao TS, Rao SS, Rao CR (eds) *Handbook of statistics*, vol 30. Elsevier, Amsterdam, pp 1–24
- Li D, Tong H (2016) Nested sub-sample search algorithm for estimation of threshold models. *Stat Sin* 26:1543–1554
- Li C, Wang D, Zhang H (2015) First-order mixed integer-valued autoregressive processes with zero-inflated generalized power series innovations. *J Korean Stat Soc* 44:232–246
- Ling S (2007) Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *J Econom* 140:849–873
- Lu JC, Chen D, Zhou W (2006) Quasi-likelihood estimation for GLM with random scales. *J Stat Plan Inference* 136:401–429
- McCullagh P, Nelder JA (1989) *Generalized linear models*, 2nd edn. Springer, New York
- Monteiro M, Scotto MG, Pereira I (2012) Integer-valued self-exciting threshold autoregressive processes. *Commun Stat Theory Method* 41:2717–2737
- Möller TA, Weiß CH (2015) Stochastic models, statistics and their applications. In: Steland (ed) *Threshold models for integer-valued time series with infinite and finite range* Springer Proceedings in Mathematics and Statistics. Springer, Berlin, pp 327–334
- Niaparast M, Schwabe R (2013) Optimal design for quasi-likelihood estimation in Poisson regression with random coefficients. *J Stat Plan Inference* 143:296–306
- Ristić MM, Bakouch HS, Nastić AS (2009) A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *J Stat Plan Inference* 139:2218–2226
- Ruiz E (1994) Quasi-maximum likelihood estimation of stochastic volatility models. *J Econom* 194(63):289–306
- Severini TA, Staniswalis JG (1994) Quasilielihood estimation in semiparametric models. *J Am Stat Assoc* 89:501–511
- Scotto MG, Weiß CH, Gouveia S (2015) Thinning-based models in the analysis of integer-valued time series: a review. *Stat Model* 15:590–618
- Steutel F, van Harn K (1979) Discrete analogues of self-decomposability and stability. *Ann Probab* 7:893–899

- Sutardhar BC, Rao RP (2001) On marginal quasi-likelihood inference in generalized linear mixed models. *J Multivar Anal* 76:1–34
- Tong H (1978) On a threshold model. In: Chen CH (ed) *Pattern recognition and signal processing*. Sijthoff and Noordhoff, Amsterdam, pp 575–586
- Tong H, Lim KS (1980) Threshold autoregressive, limit cycles and cyclical data. *J R Stat Soc B* 42:245–292
- Tweedie RL (1975) Sufficient conditions for regularity, recurrence and ergodicity of Markov process. *Stoch Process Appl* 3:385–403
- Wang C, Liu H, Yao J, Davis RA, Li WK (2014) Self-excited threshold Poisson autoregression. *J Am Stat Assoc* 2014(109):777–787
- Wedderburn RWM (1974) Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 1974(61):439–447
- Weiß CH (2008) Thinning operations for modeling time series of counts—a survey. *Adv Stat Anal* 92:319–343
- Yang K, Wang D, Jia B, Li H (2017) An integer-valued threshold autoregressive process based on negative binomial thinning. *Stat Pap*. doi:[10.1007/s00362-016-0808-1](https://doi.org/10.1007/s00362-016-0808-1)
- Zhang H, Wang D, Zhu F (2010) Inference for INAR( $p$ ) processes with signed generalized powerseries thinning operator. *J Stat Plan Inference* 140:667–683
- Zheng H, Basawa IV (2008) First-order observation-driven integer-valued autoregressive processes. *Stat Prob Lett* 78:1–9
- Zheng H, Basawa IV, Datta S (2006a) Inference for  $p$ th order random coefficient integer-valued autoregressive processes. *J Time Ser Anal* 27:411–440
- Zheng H, Basawa IV, Datta S (2006b) First-order random coefficient integer-valued autoregressive processes. *J Stat Plan Inference* 137:212–229