Article

Statistical Inference for Periodic Self-Exciting Threshold Integer-Valued Autoregressive Processes

Congmin Liu [1], Jianhua Cheng [1] and Dehui Wang [2,*]

[1] School of Mathematics, Jilin University, 2699 Qianjin Street, Changchun 130012, China; lcm18@mails.jlu.edu.cn (C.L.); chengjh@jlu.edu.cn (J.C.)

[2] School of Economics, Liaoning University, Shenyang 110036, China

* Correspondence: wangdh@jlu.edu.cn

Abstract: This paper considers the periodic self-exciting threshold integer-valued autoregressive processes under a weaker condition in which the second moment is finite instead of the innovation distribution being given. The basic statistical properties of the model are discussed, the quasi-likelihood inference of the parameters is investigated, and the asymptotic behaviors of the estimators are obtained. Threshold estimates based on quasi-likelihood and least squares methods are given. Simulation studies evidence that the quasi-likelihood methods perform well with realistic sample sizes and may be superior to least squares and maximum likelihood methods. The practical application of the processes is illustrated by a time series dataset concerning the monthly counts of claimants collecting short-term disability benefits from the Workers' Compensation Board (WCB). In addition, the forecasting problem of this dataset is addressed.

Keywords: periodic autoregression; integer-valued threshold models; parameter estimation

1. Introduction

There has been considerable interest in integer-valued time series because of their wide range of applications, including epidemiology, finance, and disease modeling. Examples of such data are as follows: the number of major global earthquakes per year, monthly crimes in a particular country or region, and patient numbers in a hospital per month over a period of time, etc. Following the first-order integer-valued autoregressive (INAR(1)) models introduced by Al-Osh and Alzaid [1], INAR models have been widely used, see Du and Li [2], Jung et al. [3], Weiß [4], Ristić et al. [5], Zhang et al. [6], Li et al. [7], Kang et al. [8] and Yu et al. [9], among others. However, for so-called piecewise phenomenon such as high thresholds, sudden bursts of large values, and time volatility, the INAR model will not work well. The threshold models (Tong [10]; Tong and Lim [11]) have attracted much attention and have been widely used to model nonlinear phenomena. To capture the piecewise phenomenon of integer-valued time series, Monteiro et al. [12] introduced a class of self-exciting threshold integer-valued autoregressive (SETINAR) models driven by independent Poisson-distributed random variables. Wang et al. [13] proposed a self-excited threshold Poisson autoregressive (SETPAR) model. Yang et al. [14] considered a class of SETINAR processes that properly capture flexible asymmetric and nonlinear responses without assuming the distributions for the errors. Yang et al. [15] introduced an integer-valued threshold autoregressive process based on a negative binomial thinning operator (NBTINAR(1)).

In addition, there are many sources of business, economic and meteorology time series data showing a periodically varying phenomenon that repeats itself after a regular period of time. It may be affected by seasonal factors and human activities. For dealing with the processes exhibiting periodic patterns, Bennett [16] and Gladyshev [17] proposed periodically correlated random processes. Then, Bentarzi and Hallin [18], Lund and Basawa [19], Basawa and Lund [20], and Shao [21], among other authors, studied

the periodic autoregressive moving-average (PARMA) models in some detail. To capture the periodic phenomenon of integer-valued time series, Monteiro et al. [22] proposed the periodic integer-valued autoregressive models of order one (PINAR(1)) with period T, driven by a periodic sequence of independent Poisson-distributed random variables. Hall et al. [23] considered the extremal behavior of periodic integer-valued moving-average sequences. Santos et al. [24] introduced a multivariate PINAR model with time-varying parameters. The analysis of periodic self-exciting threshold integer-valued autoregressive (PSETINAR$(2; 1, 1)_T$) processes was introduced by Pereira et al. [25]. Manaa and Bentarzi [26] established the existence of high moment and the strict periodic stationarity for the PSETINAR$(2; 1, 1)_T$ processes. The CLS and CML methods are applied to estimate the parameters while using the nested sub-sample search (NeSS) algorithm proposed by Li and Tong [27] to estimate the periodic threshold parameters. A drawback of this PSETINAR$(2; 1, 1)_T$ model is that the mean and variance of Poisson distribution are equal, which is not always true in the real data. Therefore, in this paper, we remove the assumption of Poisson distribution, only specify the relationship between mean and variance of observations, develop quasi-likelihood inference for the PSETINAR$(2; 1, 1)_T$ processes, and consider the estimation of thresholds.

Quasi-likelihood is a non-parametric inference method proposed by Wedderburn [28]. It is very useful in cases where the exact distributional information is not available, while only the relation between mean and variance of the observation is given, and it enjoys a certain robustness of validity. Quasi-likelihood has been widely applied. For example, Azrak and Mélard [29] proposed a simple and efficient algorithm to evaluate the exact quasi-likelihood of ARMA models with time-dependent coefficients; Christou and Fokianos [30] studied probabilistic properties and quasi-likelihood estimation for negative binomial time series models; Li et al. [31] studied the quasi-likelihood inference for the self-exciting threshold integer-valued autoregressive (SETINAR(2,1)) processes under a weaker condition; Yang et al. [32] modeled overdispersed or underdispersed count data with generalized Poisson integer-valued autoregressive (GPINAR(1)) processes and investigated the maximum quasi- likelihood estimators.

The remainder of this paper is organized as follows. In Section 2, we redefine the PSETINAR(2; 1, 1)$_T$ processes under weak conditions and discuss their basic properties. In Section 3, we consider the quasi-likelihood inference for the unknown parameters. Thresholds estimation is also discussed. Section 4 presents some simulation results for the estimates. In Section 5, we give an application of the proposed processes to a real dataset. The forecasting problem of this dataset is addressed. Concluding remarks are given in Section 6. All proofs are postponed to the Appendix A.

2. The Model and Its Properties

The periodic self-exciting threshold integer-valued autoregressive model of order one with two regimes (PSETINAR$(2; 1, 1)_T$) (originally proposed by Pereira et al. [25], and further studied by Manaa and Bentarzi [26]) is defined by the recursive equation:

$$X_t = \begin{cases} \alpha_t^{(1)} \circ X_{t-1} + Z_t, & X_{t-1} \leq r_t, \\ \alpha_t^{(2)} \circ X_{t-1} + Z_t, & X_{t-1} > r_t, \end{cases} \quad t \in \mathbb{Z} \tag{1}$$

with threshold parameters $r_t = r_j$, autoregressive coefficients $\alpha_t^{(k)} = \alpha_j^{(k)} \in (0, 1)$, for $k = 1, 2$, $t = j + sT, j = 1, 2, \ldots, T, s \in \mathbb{Z}$, and $T \in \mathbb{N}_0$. Note that Equation (1) admits the representation

$$X_{j+sT} = \left(\alpha_j^{(1)} \circ X_{j+sT-1} \right) I_{j+sT-1}^{(1)} + \left(\alpha_j^{(2)} \circ X_{j+sT-1} \right) I_{j+sT-1}^{(2)} + Z_{j+sT}, \tag{2}$$

where

(i) $I_{j+sT-1}^{(1)} := I\{X_{j+sT-1} \leq r_j\}, I_{j+sT-1}^{(2)} := 1 - I_{j+sT-1}^{(1)} = I\{X_{j+sT-1} > r_j\}$, in which $\{r_j, j = 1, 2, \ldots, T\}$ is a set of thresholds value;

(ii) The thinning operator "$\circ$" is defined as

$$\alpha_j^{(k)} \circ X_{j+sT-1} = \sum_{i=1}^{X_{j+sT-1}} U_{i,j+sT}\left(\alpha_j^{(k)}\right), \tag{3}$$

in which $\{U_{i,j+sT}\left(\alpha_j^{(k)}\right), j = 1, 2, \ldots, T, s \in \mathbb{Z}\}$ is a sequence of independent periodic Bernoulli random variables with $P\left(U_{i,j+sT}\left(\alpha_j^{(k)}\right) = 1\right) = 1 - P\left(U_{i,j+sT}\left(\alpha_j^{(k)}\right) = 0\right) = \alpha_j^{(k)}, k = 1, 2$;

(iii) $\{Z_{j+sT}, j = 1, 2, \ldots, T, s \in \mathbb{Z}\}$ constitutes a sequence of independent periodic random variables with $E(Z_{j+sT}) = \lambda_j, Var(Z_{j+sT}) = \sigma_{z,j}^2$, which is assumed to be independent of $\{X_{j+sT-1}\}$ and $\{\alpha_j^{(k)} \circ X_{j+sT-1}\}$.

Remark 1. *The innovation of PSETINAR$(2; 1, 1)_T$ process defined by Pereira et al. [25] and Manaa and Bentarzi [26] is a sequence of independent periodic Poisson-distributed random variables with mean λ_j, that is $\{Z_t\} \sim P(\lambda_j)$, where $t = j + sT, j = 1, 2, \ldots, T, s \in \mathbb{Z}$. In this paper, we use $E(Z_{j+sT}) = \lambda_j, Var(Z_{j+sT}) = \sigma_{z,j}^2$ instead of the assumption of periodic Poisson distribution for $\{Z_{j+sT}\}$, so that the model is more flexible.*

The following proposition establishes the conditional mean and the conditional variance of the PSETINAR$(2; 1, 1)_T$ process, which plays an important role in the study of the process properties and parameter estimations.

Proposition 1. *For any fixed $j = 1, 2, \ldots, T$, with $T \in \mathbb{N}_0$, the conditional mean and the conditional variance of the process $\{X_t\}$ for $t = j + sT$ and $s \in \mathbb{Z}$ defined in (2) are given by*

(i) $E(X_{j+sT}|X_{j+sT-1}) = \alpha_j^{(1)} X_{j+sT-1} I_{j+sT-1}^{(1)} + \alpha_j^{(2)} X_{j+sT-1} I_{j+sT-1}^{(2)} + \lambda_j$,

(ii) $Var(X_{j+sT}|X_{j+sT-1}) = \sum_{k=1}^{2} \alpha_j^{(k)}\left(1 - \alpha_j^{(k)}\right) X_{j+sT-1} I_{j+sT-1}^{(k)} + \sigma_{z,j}^2$.

The following theorem states the ergodicity of the PSETINAR$(2; 1, 1)_T$ process (2). This property is useful in deriving the asymptotic properties of the parameter estimators.

Theorem 1. *For any fixed $j = 1, 2, \ldots, T$, with $T \in \mathbb{N}_0$, the process $\{X_t\}$ for $t = j + sT$ and $s \in \mathbb{Z}$ defined in (2) is an ergodic Markov chain.*

3. Parameters Estimation

Suppose we have a series of observations $\{X_{j+sT}, j = 1, 2, \ldots, T, s \in \mathbb{N}_0\}$ generated from the PSETINAR$(2; 1, 1)_T$ process. The goal of this section is to estimate the unknown parameters vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{3T})' \triangleq (\alpha_1^{(1)}, \alpha_1^{(2)}, \lambda_1, \alpha_2^{(1)}, \alpha_2^{(2)}, \lambda_2, \ldots, \alpha_T^{(1)}, \alpha_T^{(2)}, \lambda_T)'$ and threshold parameters vector $\boldsymbol{r} = (r_1, r_2, \ldots, r_T)'$. This section is divided into two subsections. In Section 3.1, we estimate the parameters vector $\boldsymbol{\beta}$ by using the maximum quasi-likelihood (MQL) method when the thresholds value is known. We consider the maximum quasi-likelihood (MQL) and conditional least square (CLS) estimators of thresholds $\boldsymbol{r}$ in Section 3.2.

3.1. Estimation of Parameters $\boldsymbol{\beta}$

As described in Proposition 1 (ii), we have the variance of X_t conditional on X_{t-1}, let $\boldsymbol{\theta}_j \triangleq \left(\theta_j^{(1)}, \theta_j^{(2)}, \sigma_{z,j}^2\right)'$ with $\theta_j^{(k)} = \alpha_j^{(k)}\left(1 - \alpha_j^{(k)}\right), k = 1, 2, j = 1, 2, \ldots, T$, then the $Var(X_{j+sT}|X_{j+sT-1})$ admits the representation

$$V_{\boldsymbol{\theta}_j}\left(X_{j+sT}|X_{j+sT-1}\right) \overset{\triangle}{=} Var\left(X_{j+sT}|X_{j+sT-1}\right) = \theta_j^{(1)} X_{j+sT-1} I_{j+sT-1}^{(1)} + \theta_j^{(2)} X_{j+sT-1} I_{j+sT-1}^{(2)} + \sigma_{z,j}^2,$$

for $\forall j = 1, 2, \ldots, T, s \in \mathbb{N}_0$.

As discussed in Wedderburn [28], we have the set of standard quasi-likelihood estimating equations:

$$L(\boldsymbol{\beta}) = \sum_{s=0}^{N-1} \sum_{j=1}^{T} \frac{X_{j+sT} - E\left(X_{j+sT}|X_{j+sT-1}\right)}{V_{\boldsymbol{\theta}_j}\left(X_{j+sT}|X_{j+sT-1}\right)} \frac{\partial E\left(X_{j+sT}|X_{j+sT-1}\right)}{\partial \beta_i} = 0, \tag{4}$$

for $i = 1, \ldots, 3T$, where N is the total number of cycles. By solving (4), the quasi-likelihood estimator can be obtained.

This method is essentially a two-step estimation, if $\boldsymbol{\theta}_j$ is unknown, we propose substituting a suitable consistent estimator of $\boldsymbol{\theta}_j$ obtained by other means, getting modified quasi-likelihood estimating equations and then solving them for the primary parameters of interest. In the modified quasi- likelihood estimating equations, we replace $\boldsymbol{\theta}_j$ with a suitable consistent estimator $\hat{\boldsymbol{\theta}}_j$. For simplicity in notation, we define $V_{\hat{\boldsymbol{\theta}}_j}^{-1} \overset{\triangle}{=} V_{\hat{\boldsymbol{\theta}}_j}^{-1}\left(X_{j+sT}|X_{j+sT-1}\right)$.

This approach leads to the modified quasi-likelihood estimator $\hat{\boldsymbol{\beta}}_{MQL}$ of $\boldsymbol{\beta}$ (see Zheng, Basawa and Datta [33]):

$$\hat{\boldsymbol{\beta}}_{MQL} = Q_N^{-1} \boldsymbol{q}_N, \tag{5}$$

where

$$Q_N = \begin{bmatrix} \boldsymbol{Q}_{1,N} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_{2,N} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{Q}_{T,N} \end{bmatrix},$$

and

$$\boldsymbol{q}_N = \left(\boldsymbol{q}_{1,N}, \boldsymbol{q}_{2,N}, \ldots, \boldsymbol{q}_{T,N}\right)',$$

moreover, the $\boldsymbol{0}$'s are (3×3)-null matrices, $\boldsymbol{Q}_{j,N}$ and $\boldsymbol{q}_{j,N}$ $(j = 1, 2, \ldots, T)$ given by

$$\boldsymbol{Q}_{j,N} = \begin{bmatrix} \sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} X_{j+sT-1}^2 I_{j+sT-1}^{(1)} & 0 & \sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} X_{j+sT-1} I_{j+sT-1}^{(1)} \\ 0 & \sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} X_{j+sT-1}^2 I_{j+sT-1}^{(2)} & \sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} X_{j+sT-1} I_{j+sT-1}^{(2)} \\ \sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} X_{j+sT-1} I_{j+sT-1}^{(1)} & \sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} X_{j+sT-1} I_{j+sT-1}^{(2)} & \sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} \end{bmatrix},$$

$$\boldsymbol{q}_{j,N} = \left(\sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} X_{j+sT} X_{j+sT-1} I_{j+sT-1}^{(1)} \quad \sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} X_{j+sT} X_{j+sT-1} I_{j+sT-1}^{(2)} \quad \sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} X_{j+sT} \right)'.$$

Note that we use consistent estimator $\hat{\boldsymbol{\theta}}_j = \left(\hat{\alpha}_j^{(1)}\left(1 - \hat{\alpha}_j^{(1)}\right), \hat{\alpha}_j^{(2)}\left(1 - \hat{\alpha}_j^{(2)}\right), \hat{\sigma}_{z,j}^2 \right)'$ instead of $\boldsymbol{\theta}_j$.

Next, the proposition gives consistent estimators $\hat{\sigma}_{z,j}^2$ of $\sigma_{z,j}^2$, which depends on some consistent estimators $\hat{\alpha}_j^{(k)}$ and $\hat{\lambda}_j$ with $k = 1, 2$, $j = 1, 2, \ldots, T$.

Proposition 2. *The following variance estimators for $\{Z_{j+sT}\}$ with $j = 1, 2, \ldots, T, s \in \mathbb{N}_0$ are consistent:*

$$(i) \quad \hat{\sigma}^2_{1,z,j} = \frac{1}{N} \sum_{s=0}^{N-1} \left(X_{j+sT} - \sum_{k=1}^{2} \hat{\alpha}^{(k)}_j X_{j+sT-1} I^{(k)}_{j+sT-1} - \hat{\lambda}_j \right)^2$$

$$- \frac{1}{N} \sum_{k=1}^{2} \sum_{s=0}^{N-1} \hat{\alpha}^{(k)}_j \left(1 - \hat{\alpha}^{(k)}_j \right) X_{j+sT-1} I^{(k)}_{j+sT-1}, \tag{6}$$

$$(ii) \quad \hat{\sigma}^2_{2,z,j} = \hat{\sigma}^2_{x,j} - \hat{p}_j \left[\hat{\alpha}^{(1)^2}_j \hat{\sigma}^{2(1)}_j + \hat{\alpha}^{(1)}_j \left(1 - \hat{\alpha}^{(1)}_j \right) \hat{\mu}^{(1)}_j \right]$$

$$- (1 - \hat{p}_j) \left[\hat{\alpha}^{(2)^2}_j \hat{\sigma}^{2(2)}_j + \hat{\alpha}^{(2)}_j \left(1 - \hat{\alpha}^{(2)}_j \right) \hat{\mu}^{(2)}_j \right]$$

$$- \hat{p}_j (1 - \hat{p}_j) \left(\hat{\alpha}^{(1)}_j \hat{\mu}^{(1)}_j - \hat{\alpha}^{(2)}_j \hat{\mu}^{(2)}_j \right)^2, \tag{7}$$

for $k = 1, 2, j = 1, 2, \ldots, T, s \in \mathbb{N}_0$, in which $\hat{\alpha}^{(k)}_j$ and $\hat{\lambda}_j$ are consistent estimators of $\alpha^{(k)}_j$ and λ_j (for example, we can use the CLS estimators given in Theorem 3.1 of Pereira et al. [25]), furthermore

$$\bar{X}_j = \frac{1}{N} \sum_{s=0}^{N-1} X_{j+sT}, \quad \hat{\sigma}^2_{x,j} = \frac{1}{N} \sum_{s=0}^{N-1} \left(X_{j+sT} - \bar{X}_j \right)^2,$$

$$N^{(k)}_j = \sum_{s=0}^{N-1} I^{(k)}_{j+sT-1}, \quad \hat{\mu}^{(k)}_j = \frac{1}{N^{(k)}_j} \sum_{s \in \{I^{(k)}_{j+sT-1}=1\}} X_{j+sT},$$

$$\hat{p}_j = \frac{1}{N} \sum_{s=0}^{N-1} I^{(1)}_{j+sT-1}, \quad \hat{\sigma}^{2(k)}_j = \frac{1}{N^{(k)}_j} \sum_{s \in \{I^{(k)}_{j+sT-1}=1\}} \left(X_{j+sT} - \hat{\mu}^{(k)}_j \right)^2.$$

The two estimations are based on conditional variance $Var\left(X_{j+sT} | X_{j+sT-1} \right)$ and variance $Var\left(X_{j+sT} \right)$, respectively. The details can be found in the Appendix A.

To study the asymptotic behavior of the estimator $\hat{\beta}_{MQL}$, we make the following assumptions about the process of $\{X_t\}$:

(C1) *By Proposition 1 in Pereira et al. [25], we assume the $\{X_t\}$ is a strictly ciclostationary process;*
(C2) $E|X_t|^4 < \infty$.

Now for the asymptotic properties of the quasi-likelihood estimator $\hat{\beta}_{MQL}$ given by (5), we have the following asymptotic distribution.

Theorem 2. *Let $\{X_t\}$ be a PSETINAR$(2; 1, 1)_T$ process defined in (2), then under the assumptions (C1)-(C2), the estimator $\hat{\beta}_{MQL}$ given by (5) is asymptotically normal,*

$$\sqrt{N} \left(\hat{\beta}_{MQL} - \beta \right) \to \mathcal{N} \left(0, H^{-1}(\theta) \right),$$

where

$$H(\theta) = \begin{bmatrix} H_1(\theta) & 0 & \cdots & 0 \\ 0 & H_2(\theta) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_T(\theta) \end{bmatrix},$$

with matrices H_j $(j = 1, 2, \ldots, T)$ given by

$$\boldsymbol{H}_j(\boldsymbol{\theta}) = \begin{bmatrix} E\left(V_{\boldsymbol{\theta}_j}^{-1}(X_j|X_{j-1})X_{j-1}^2 I_{j-1}^{(1)}\right) & 0 & E\left(V_{\boldsymbol{\theta}_j}^{-1}(X_j|X_{j-1})X_{j-1} I_{j-1}^{(1)}\right) \\ 0 & E\left(V_{\boldsymbol{\theta}_j}^{-1}(X_j|X_{j-1})X_{j-1}^2 I_{j-1}^{(2)}\right) & E\left(V_{\boldsymbol{\theta}_j}^{-1}(X_j|X_{j-1})X_{j-1} I_{j-1}^{(2)}\right) \\ E\left(V_{\boldsymbol{\theta}_j}^{-1}(X_j|X_{j-1})X_{j-1} I_{j-1}^{(1)}\right) & E\left(V_{\boldsymbol{\theta}_j}^{-1}(X_j|X_{j-1})X_{j-1} I_{j-1}^{(2)}\right) & E\left(V_{\boldsymbol{\theta}_j}^{-1}(X_j|X_{j-1})\right) \end{bmatrix}.$$

It is worth mentioning that this theorem reflects the consistency of the estimator $\hat{\boldsymbol{\beta}}_{MQL}$.

3.2. Estimation of Thresholds Vector $\boldsymbol{r}$

Note that in the real data application, the threshold values are also unknown. In this subsection, we estimate the thresholds vector $\boldsymbol{r} = (r_1, r_2, \ldots, r_T)'$. Here, we further promote the nested sub-sample search (NeSS) algorithm (see, e.g., Yang et al. [15], Li and Tong [27], and Li et al. [31]) and use conditional least squares (CLS) and modified quasi-likelihood (MQL) principles to estimate $\boldsymbol{r}$.

For some fixed $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_T)'$, the application of the conditional least squares principle yields the sum of squared errors:

$$S_N(\boldsymbol{r}, \boldsymbol{\lambda})$$

$$= \sum_{s=0}^{N-1}\sum_{j=1}^{T}\left(X_{j+sT} - \sum_{k=1}^{2} \frac{\sum_{s=0}^{N-1}\left(X_{j+sT}X_{j+sT-1}I_{j+sT-1}^{(k)} - \lambda_j X_{j+sT-1}I_{j+sT-1}^{(k)}\right)}{\sum_{s=0}^{N-1}X_{j+sT-1}^2 I_{j+sT-1}^{(k)}} X_{j+sT-1}I_{j+sT-1}^{(k)} - \lambda_j\right)^2,$$

and then the thresholds vector $\boldsymbol{r}$ can be estimated by minimizing $S_N(\boldsymbol{r}, \boldsymbol{\lambda})$,

$$\hat{\boldsymbol{r}} = \arg\min_{\boldsymbol{r}\in[\underline{\boldsymbol{r}},\bar{\boldsymbol{r}}]} S_N(\boldsymbol{r}, \boldsymbol{\lambda}), \tag{8}$$

where $\underline{\boldsymbol{r}}$ and $\bar{\boldsymbol{r}}$ are some known lower and upper bounds of $\boldsymbol{r}$. In practice, they can be selected as the minimum and maximum values in each cycle of the sample. For convenience, we consider an alternative objective function

$$J_N(\boldsymbol{r}, \boldsymbol{\lambda}) = S_N - S_N(\boldsymbol{r}, \boldsymbol{\lambda}),$$

where

$$S_N = \sum_{s=0}^{N-1}\sum_{j=1}^{T}\left(X_{j+sT} - \frac{\sum_{s=0}^{N-1}\left(X_{j+sT}X_{j+sT-1} - \lambda_j X_{j+sT-1}\right)}{\sum_{s=0}^{N-1}X_{j+sT-1}^2} X_{j+sT-1} - \lambda_j\right)^2.$$

Now, the optimization in (8) is equivalent to

$$\hat{\boldsymbol{r}}_{CLS} = \arg\max_{\boldsymbol{r}\in[\underline{\boldsymbol{r}},\bar{\boldsymbol{r}}]} J_N(\boldsymbol{r}, \boldsymbol{\lambda}), \tag{9}$$

where $\hat{\boldsymbol{r}}_{CLS}$ is the conditional least squares estimator of the thresholds vector $\boldsymbol{r}$.

Inspired by the method of conditional least squares, we investigate the performances of $\boldsymbol{r}$ by using the quasi-likelihood principle. The modified quasi-likelihood estimator $\hat{\boldsymbol{r}}_{MQL}$ of $\boldsymbol{r}$ is obtained by maximizing the expression

$$\tilde{J}_N(\boldsymbol{r}, \boldsymbol{\lambda}) = \tilde{S}_N - \tilde{S}_N(\boldsymbol{r}, \boldsymbol{\lambda}),$$

which yields

$$\hat{\boldsymbol{r}}_{MQL} = \arg\max_{\boldsymbol{r}\in[\underline{\boldsymbol{r}},\bar{\boldsymbol{r}}]} \tilde{J}_N(\boldsymbol{r}, \boldsymbol{\lambda}), \tag{10}$$

where

$$\tilde{S}_N(\boldsymbol{r}, \boldsymbol{\lambda})$$

$$= \sum_{s=0}^{N-1} \sum_{j=1}^{T} V_{\hat{\boldsymbol{\theta}}_j}^{-1} \left(X_{j+sT} - \sum_{k=1}^{2} \frac{\sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} \cdot \left(X_{j+sT} X_{j+sT-1} I_{j+sT-1}^{(k)} - \lambda_j X_{j+sT-1} I_{j+sT-1}^{(k)} \right)}{\sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} \cdot X_{j+sT-1}^2 I_{j+sT-1}^{(k)}} X_{j+sT-1} I_{j+sT-1}^{(k)} - \lambda_j \right)^2,$$

and

$$\tilde{S}_N = \sum_{s=0}^{N-1} \sum_{j=1}^{T} V_{\hat{\boldsymbol{\theta}}_j}^{-1} \left(X_{j+sT} - \frac{\sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} \cdot \left(X_{j+sT} X_{j+sT-1} - \lambda_j X_{j+sT-1} \right)}{\sum_{s=0}^{N-1} V_{\hat{\boldsymbol{\theta}}_j}^{-1} \cdot X_{j+sT-1}^2} X_{j+sT-1} - \lambda_j \right)^2.$$

It is worth mentioning that there are unknown parameters λ_j with $j = 1, \ldots, T$ when we use (9) and (10) to estimate thresholds vector $\boldsymbol{r}$. As argued in Li and Tong [27], Yang et al. [14], and Yang et al. [15], when λ and r are one-dimensional parameters, we can choose any positive number as the value of λ without worrying about getting a wrong result of $\hat{r}$. Fortunately, we also find out by simulations that the estimations of $\boldsymbol{r}$ by maximizing $\tilde{J}_N(\boldsymbol{r}, \boldsymbol{\lambda})$ and $J_N(\boldsymbol{r}, \boldsymbol{\lambda})$ do not depend on the value of $\boldsymbol{\lambda}$. In order to give an intuitive impression of $\tilde{J}_N(\boldsymbol{r}, \boldsymbol{\lambda})/N$, we generate a set of data with Model I (given in Section 4, i.e., $T = 2, N = 50, \boldsymbol{\beta} = (0.2, 0.1, 3, 0.8, 0.1, 7)', \boldsymbol{r} = (8, 4)'$), and plot the shapes of $\tilde{J}_N(\boldsymbol{r}, \boldsymbol{\lambda})/N$. From Figure 1, we can see that for different values of $\boldsymbol{\lambda}$, the shape of $\tilde{J}_N(\boldsymbol{r}, \boldsymbol{\lambda})/N$ changes, but the maximum value in each subfigure is obtained at the true thresholds vector $\boldsymbol{r} = (8, 4)'$. In practice, we can choose the mean in each cycle of the samples for $\lambda_j, j = 1, 2, \ldots, T$.

Actually, using the quasi-likelihood method to estimate the thresholds is a three-step estimation procedure, and we now present the algorithm to implement our estimation procedure as follows:

Step 1: *Choose the upper bound $\bar{\boldsymbol{r}}$ and lower bound $\underline{\boldsymbol{r}}$ of $\boldsymbol{r}$, solve (9) to get the $\hat{\boldsymbol{r}}_{CLS}$ with $\lambda_j = \bar{X}_j = \frac{1}{N} \sum_{s=0}^{N-1} X_{j+sT}, j = 1, 2, \ldots, T$;*

Step 2: *Fix $\hat{\boldsymbol{r}}_{CLS}$ at the current value, solve (6) or (7) to get the $\hat{\sigma}_{z,j}^2, j = 1, 2, \ldots, T$, where $\alpha_j^{(k)}$ and λ_j with $k = 1, 2$ can be estimated by other methods, then solve (5) to get $\hat{\boldsymbol{\beta}}_{MQL}$.*

Step 3: *Fix $\hat{\boldsymbol{\theta}}_j = \left(\hat{\alpha}_{j,MQL}^{(1)} \left(1 - \hat{\alpha}_{j,MQL}^{(1)} \right), \hat{\alpha}_{j,MQL}^{(2)} \left(1 - \hat{\alpha}_{j,MQL}^{(2)} \right), \hat{\sigma}_{z,j}^2 \right)', j = 1, 2, \ldots, T$ at its estimated value from Step 2, choose the same upper bound $\bar{\boldsymbol{r}}$ and lower bound $\underline{\boldsymbol{r}}$ as in Step 1, solve (10) to get $\hat{\boldsymbol{r}}_{MQL}$.*

(**a**) Shape of $\tilde{J}_N(r,\lambda)/N$ with $\lambda = (3,7)$

(**b**) Shape of $\tilde{J}_N(r,\lambda)/N$ with $\lambda = (3,8)$

(**c**) Shape of $\tilde{J}_N(r,\lambda)/N$ with $\lambda = (2,7)$

(**d**) Shape of $\tilde{J}_N(r,\lambda)/N$ with $\lambda = (2,8)$
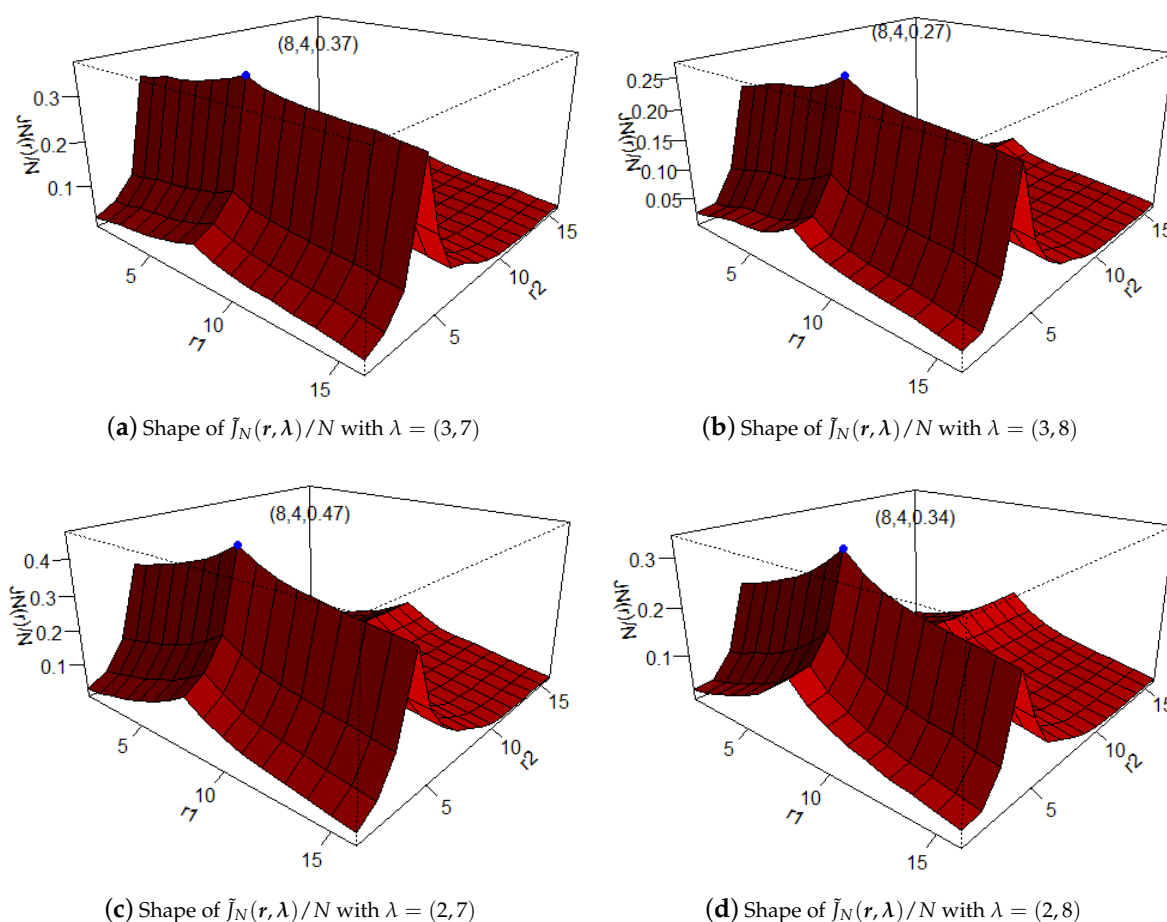
Figure 1. The shapes of $\tilde{J}_N(r,\lambda)/N$.

4. Simulation Study

In this section, we conduct simulation studies to illustrate the finite sample performances of the estimates. The initial value X_0 is fixed at 0. In order to capture the characteristics of the data from the PSETINAR$(2;1,1)_T$ process, we first generate a set of data with the distribution of innovations $\{Z_t\}$ given by Model I (mentioned below in this section) and parameters $\beta = (0.2, 0.45, 1, 0.2, 0.45, 2, 0.65, 0.45, 1, 0.65, 0.45, 2, 0.2, 0.45, 3, 0.2, 0.45,$ $7, 0.8, 0.45, 7, 0.2, 0.1, 3, 0.8, 0.1, 7, 0.2, 0.1, 7, 0.8, 0.45, 2)'$, $r = (3,3,3,1,3,3,5,9,3,6,7)'$, $T = 11$, $N = 50$. The parameter vectors we choose here are randomly selected, and there are slight differences between the parameters of each cycle, the thresholds vector of r was chosen such that there are enough data in each regime. We give the sample path in the first six cycles in Figure 2, of which $N = 6$. We can see that even if there are slight differences between the parameters of each cycle, the dataset still exhibits periodic characteristics.

To report the performances of the estimates, we conduct simulation studies under the following three models:

Model I. Assume that $\{Z_t\}$ is a sequence of i.i.d periodic Poisson distributed random variables with mean $E(Z_t) = Var(Z_t) = \lambda_j$ for $t = j + sT, j = 1, 2, \ldots, T, s \in \mathbb{N}_0$.

Model II. Assume that $\{Z_t\}$ is a sequence of i.i.d. periodic Geometric distributed random variables with p.m.f. given by

$$p\left(Z_{j+sT} = z\right) = \frac{\lambda_j^z}{\left(1 + \lambda_j\right)^{1+z}}, z = 0, 1, 2, \ldots$$

with $E(Z_t) = \lambda_j, Var(Z_t) = \lambda_j(1 + \lambda_j)$ for $t = j + sT, j = 1, 2, \ldots, T, s \in \mathbb{N}_0$.

Model III. Assume that $\{Z_t\}$ is a sequence of i.i.d mixed distributed random variables,

$$Z_t = \Delta_t Z_{1t} + (1 - \Delta_t)Z_{2t},$$

where $\{\Delta_t\}$ is a sequence of i.i.d periodic Bernoulli distributed random variables with $P(\Delta_t = 1) = 1 - P(\Delta_t = 0) = \rho_j, \boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_T)$ for $t = j + sT, j = 1, 2, \ldots, T, s \in \mathbb{N}_0$, which is independent of $\{Z_{it}\}, i = 1, 2$.

For $\{Z_{1t}\}$ given in Model I and $\{Z_{2t}\}$ given in Model II, we can easily see that $E(Z_t) = \lambda_j, Var(Z_t) = \lambda_j^2(1 - \rho_j) + \lambda_j$.

For each model, we generate the data with $X_0 = 0$, set $T = 3$ and the sample sizes $n = NT = 150, 300, 900$. All the calculations are performed under the $\mathcal{R}3.6.2$ software with 1000 replications. We use the command *constrOptim* to optimize the objective function of the maximum likelihood estimation. The threshold vector is calculated by the algorithms discussed in Section 3.2. Other algorithms are based on the explicit expressions.
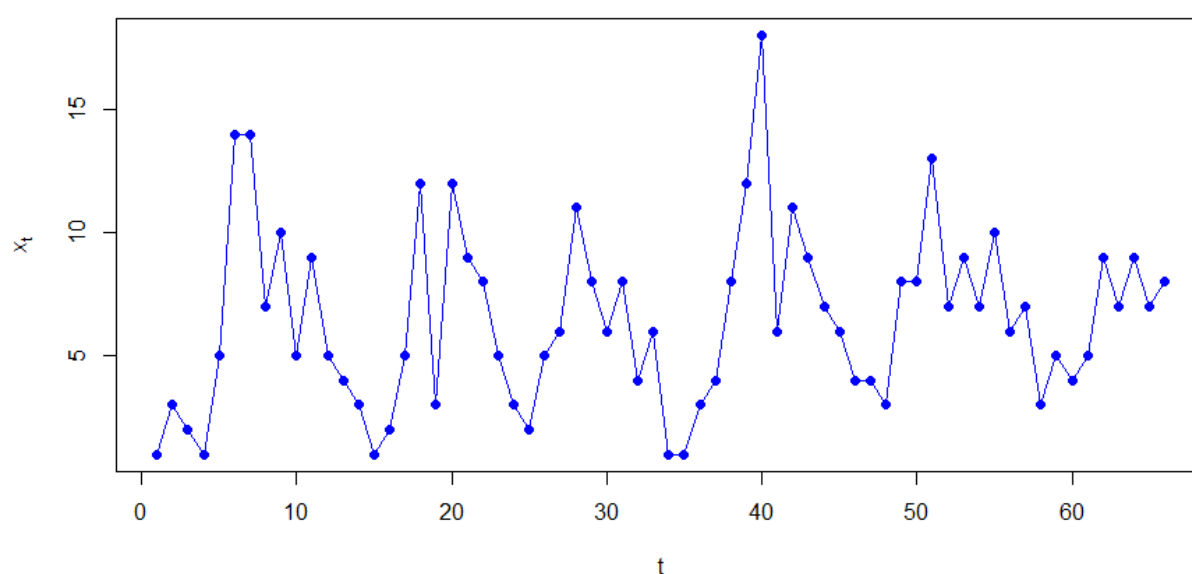
Figure 2. Sample path of the first six cycles.

4.1. Performances of the $\hat{\boldsymbol{\beta}}_{CLS}$, $\hat{\boldsymbol{\beta}}_{MQL}$ and $\hat{\boldsymbol{\beta}}_{CML}$

Pereira et al. [25] provided a theoretical basis for the conditional least squares (CLS) and conditional maximum likelihood (CML) estimators of the parameters vector $\boldsymbol{\beta}$ in the PSETINAR$(2; 1, 1)_T$ process but did not conduct simulation research. Manaa and Bentarzi [26] provided the asymptotic properties of the estimators and compared their performance through a simulation study. To compare the performance of the three estimators $\hat{\boldsymbol{\beta}}_{CLS}$, $\hat{\boldsymbol{\beta}}_{CML}$ and $\hat{\boldsymbol{\beta}}_{MQL}$ (given in Section 3), we conduct simulation studies for these three estimators under Models I to III. The parameters are selected as follows:

Series A. $\boldsymbol{\beta} = (0.2, 0.45, 1, 0.2, 0.45, 2, 0.8, 0.45, 2)', \boldsymbol{r} = (3, 2, 2)'$.

Series B. $\boldsymbol{\beta} = (0.65, 0.45, 1, 0.65, 0.45, 2, 0.35, 0.45, 2)', \boldsymbol{r} = (2, 2, 3)'$.

Series C. $\boldsymbol{\beta} = (0.2, 0.45, 3, 0.2, 0.45, 7, 0.8, 0.45, 7)', \boldsymbol{r} = (12, 7, 9)'$.

To eliminate the influence of the change of parameters on estimates, we choose the series randomly and change the parameters with fixed $\boldsymbol{\alpha}^{(k)}, k = 1, 2$ or $\boldsymbol{\lambda}$ separately. The selection of these thresholds ensures there are enough data in each regime.

Spectral analysis starts from finding hidden periodicity, and it is an important subject of time series frequency domain analysis. The approach for studying hidden periods based on frequency domain analysis is the periodogram method, proposed by Schuster [34]; the rigorous examination is shown in Fisher [35]. For a series of observations $\{X_t\}$, $t = 1, 2, \ldots, n$, the periodogram is defined as

$$I_n(f_k) = \frac{1}{n}|\sum_{t=1}^{n} X_t e^{-i2\pi f_k t}|^2 = a_k^2 + b_k^2,\tag{11}$$

where

$$a_k = \begin{cases} \frac{1}{\sqrt{n}}(\sum_{t=1}^{n} X_t \cos(2\pi f_k t))^2, & k = 1, 2, \ldots, \left[\frac{n-1}{2}\right], \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(-1)^t X_t, & k = \frac{n}{2}, \end{cases}$$

$$b_k = \begin{cases} \frac{1}{\sqrt{n}}(\sum_{t=1}^{n} X_t \sin(2\pi f_k t))^2, & k = 1, 2, \ldots, \left[\frac{n-1}{2}\right], \\ 0, & k = \frac{n}{2}, \end{cases}$$

and the period $T = [1/\mathrm{argmax}_f I_n(f_k)]$, where $[\cdot]$ denotes the integer part of a number.

The sample path and periodogram of the Series A, B and C under Model I are plotted in Figure 3 to show the periodic characteristics. Because the period is three and short, it is difficult to see the period from the sample path, but the periodogram can show the period very well. In addition, the simulation results are summarized in Tables 1–9.

As expected, biases and MSE of the estimators decrease as the sample size N increases, which is in agreement with the asymptotic properties of the estimators: asymptotic unbiasedness and consistency. Most of the biases and MSE in Model II are larger than those in Model I. Maybe this is because the variance of $\{Z_t\}$ in Model II is larger than that in Model I, which leads to the fluctuation of data.

Tables 1–6 summarize the simulation results for different series under Model I and Model II. From these tables, we can see that most of the biases and MSE of $\hat{\beta}_{MQL}$ are smaller than $\hat{\beta}_{CLS}$. Perhaps it is because that the MQL method uses more information about the data than the CLS method. Therefore, the MQL method can obtain the optimal value more accurately. In addition, most of the biases of $\hat{\beta}_{MQL}$ are smaller than $\hat{\beta}_{CML}$, while the MSE is larger, which is because the CML uses the distribution. If the distribution is correct, it is indeed better than the MQL. It is worth mentioning that the CML method is more complicated and time-consuming than the MQL method in the simulation procedure. We can conclude that the MQL estimators are better than CLS estimators, and the CML estimators are not unanimously better than MQL estimators.

Table 1. Bias and MSE for Series A of Model I (MSE in parentheses): CLS, MQL and CML.

N	Method	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	λ_1	$\alpha_2^{(1)}$	$\alpha_2^{(2)}$	λ_2	$\alpha_3^{(1)}$	$\alpha_3^{(2)}$	λ_3
50	CLS	0.001	−0.001	0.001	−0.018	−0.004	0.006	0.008	0.005	−0.025
		(0.052)	(0.014)	(0.253)	(0.131)	(0.024)	(0.230)	(0.160)	(0.024)	(0.326)
	MQL	0.000	−0.002	0.006	−0.015	−0.004	0.002	0.011	0.006	−0.030
		(0.054)	(0.014)	(0.266)	(0.126)	(0.023)	(0.220)	(0.156)	(0.024)	(0.316)
	CML	0.024	0.010	−0.047	0.054	0.019	−0.079	0.003	0.007	−0.027
		(0.024)	(0.008)	(0.117)	(0.062)	(0.016)	(0.126)	(0.047)	(0.013)	(0.134)
100	CLS	0.004	0.000	−0.006	0.013	−0.001	−0.005	0.002	−0.003	0.008
		(0.026)	(0.007)	(0.132)	(0.058)	(0.011)	(0.108)	(0.085)	(0.012)	(0.168)
	MQL	0.004	0.000	−0.006	0.013	−0.001	−0.006	−0.001	−0.004	0.012
		(0.024)	(0.007)	(0.120)	(0.057)	(0.011)	(0.105)	(0.082)	(0.011)	(0.162)
	CML	0.012	0.004	−0.023	0.036	0.007	−0.034	0.003	0.000	−0.001
		(0.014)	(0.004)	(0.067)	(0.036)	(0.008)	(0.073)	(0.024)	(0.006)	(0.066)
300	CLS	−0.003	−0.002	0.009	0.002	0.000	−0.005	−0.002	0.000	−0.001
		(0.010)	(0.003)	(0.051)	(0.020)	(0.004)	(0.034)	(0.028)	(0.004)	(0.055)
	MQL	−0.002	−0.001	0.007	0.001	0.000	−0.004	−0.003	0.000	0.000
		(0.009)	(0.002)	(0.045)	(0.019)	(0.004)	(0.033)	(0.027)	(0.003)	(0.053)
	CML	0.000	0.000	0.000	0.003	0.001	−0.007	0.001	0.002	−0.006
		(0.005)	(0.001)	(0.025)	(0.014)	(0.003)	(0.024)	(0.007)	(0.002)	(0.020)

Table 2. Bias and MSE for Series B of Model I (MSE in parentheses): CLS, MQL and CML.

N	Method	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	λ_1	$\alpha_2^{(1)}$	$\alpha_2^{(2)}$	λ_2	$\alpha_3^{(1)}$	$\alpha_3^{(2)}$	λ_3
50	CLS	0.009	0.001	0.003	0.014	0.003	−0.015	−0.013	−0.009	0.032
		(0.119)	(0.015)	(0.238)	(0.166)	(0.031)	(0.365)	(0.105)	(0.026)	(0.525)
	MQL	0.010	0.001	0.003	0.013	0.003	−0.014	−0.012	−0.009	0.031
		(0.129)	(0.015)	(0.241)	(0.161)	(0.030)	(0.354)	(0.104)	(0.026)	(0.516)
	CML	0.006	0.003	0.001	0.014	0.006	−0.020	0.008	0.003	−0.019
		(0.043)	(0.007)	(0.090)	(0.062)	(0.016)	(0.150)	(0.045)	(0.014)	(0.229)
100	CLS	0.007	0.000	−0.001	−0.022	−0.009	0.042	−0.004	−0.002	0.003
		(0.061)	(0.008)	(0.133)	(0.076)	(0.014)	(0.173)	(0.046)	(0.012)	(0.222)
	MQL	0.008	0.000	−0.003	−0.023	−0.010	0.044	−0.004	−0.002	0.003
		(0.055)	(0.007)	(0.116)	(0.076)	(0.014)	(0.172)	(0.045)	(0.012)	(0.216)
	CML	0.002	0.000	−0.001	−0.004	−0.001	0.013	0.000	0.001	−0.008
		(0.018)	(0.003)	(0.040)	(0.031)	(0.008)	(0.078)	(0.027)	(0.007)	(0.127)
300	CLS	0.003	0.000	−0.003	0.002	0.000	0.002	−0.003	−0.001	−0.001
		(0.020)	(0.003)	(0.043)	(0.026)	(0.005)	(0.060)	(0.017)	(0.004)	(0.081)
	MQL	0.003	−0.001	−0.002	0.001	0.000	0.004	−0.002	0.000	−0.004
		(0.019)	(0.002)	(0.039)	(0.025)	(0.005)	(0.058)	(0.016)	(0.004)	(0.077)
	CML	−0.002	−0.002	0.003	0.003	0.001	−0.001	−0.003	0.000	−0.002
		(0.006)	(0.001)	(0.014)	(0.009)	(0.002)	(0.025)	(0.009)	(0.003)	(0.043)

Table 3. Bias and MSE for Series C of Model I (MSE in parentheses): CLS, MQL and CML.

N	Method	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	λ_1	$\alpha_2^{(1)}$	$\alpha_2^{(2)}$	λ_2	$\alpha_3^{(1)}$	$\alpha_3^{(2)}$	λ_3
50	CLS	−0.013	−0.010	0.146	−0.010	−0.003	0.053	−0.010	−0.007	0.054
		(0.022)	(0.011)	(2.088)	(0.082)	(0.022)	(1.915)	(0.078)	(0.026)	(3.823)
	MQL	−0.010	−0.008	0.117	−0.010	−0.003	0.052	−0.014	−0.009	0.079
		(0.022)	(0.010)	(2.000)	(0.082)	(0.021)	(1.913)	(0.075)	(0.025)	(3.709)
	CML	0.003	0.001	−0.015	0.044	0.021	−0.201	0.003	0.000	−0.033
		(0.012)	(0.006)	(1.119)	(0.044)	(0.013)	(1.054)	(0.025)	(0.010)	(1.286)
100	CLS	0.001	−0.002	0.015	−0.003	0.001	0.013	0.002	−0.003	0.022
		(0.014)	(0.006)	(1.323)	(0.043)	(0.011)	(1.046)	(0.038)	(0.012)	(1.772)
	MQL	0.000	−0.003	0.034	−0.002	0.001	0.008	0.001	−0.003	0.027
		(0.012)	(0.006)	(1.203)	(0.042)	(0.011)	(1.027)	(0.037)	(0.012)	(1.726)
	CML	0.006	0.001	−0.029	0.018	0.010	−0.085	0.011	0.003	−0.043
		(0.007)	(0.003)	(0.672)	(0.026)	(0.007)	(0.657)	(0.012)	(0.005)	(0.620)
300	CLS	0.000	0.000	0.006	0.002	0.002	−0.014	0.006	0.003	−0.040
		(0.006)	(0.003)	(0.586)	(0.014)	(0.004)	(0.350)	(0.013)	(0.004)	(0.606)
	MQL	0.001	0.000	0.002	0.001	0.001	−0.010	0.005	0.002	−0.032
		(0.005)	(0.002)	(0.527)	(0.014)	(0.004)	(0.341)	(0.012)	(0.004)	(0.589)
	CML	0.002	0.001	−0.013	0.005	0.003	−0.030	0.003	0.002	−0.026
		(0.003)	(0.001)	(0.262)	(0.011)	(0.003)	(0.267)	(0.004)	(0.002)	(0.201)

Table 4. Bias and MSE for Series A of Model II (MSE in parentheses): CLS, MQL and CML.

N	Method	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	λ_1	$\alpha_2^{(1)}$	$\alpha_2^{(2)}$	λ_2	$\alpha_3^{(1)}$	$\alpha_3^{(2)}$	λ_3
50	CLS	−0.013	−0.005	0.021	−0.016	−0.014	0.024	−0.025	−0.011	0.044
		(0.073)	(0.011)	(0.247)	(0.291)	(0.032)	(0.408)	(0.330)	(0.024)	(0.449)
	MQL	−0.011	−0.005	0.019	−0.012	−0.013	0.019	−0.020	−0.010	0.040
		(0.067)	(0.011)	(0.228)	(0.287)	(0.032)	(0.402)	(0.330)	(0.024)	(0.439)
	CML	0.014	0.005	−0.026	0.041	0.011	−0.050	0.004	0.008	−0.016
		(0.016)	(0.005)	(0.076)	(0.040)	(0.010)	(0.158)	(0.020)	(0.007)	(0.153)
100	CLS	0.003	0.003	−0.011	−0.013	−0.011	0.027	−0.002	0.001	−0.005
		(0.032)	(0.005)	(0.116)	(0.145)	(0.016)	(0.195)	(0.170)	(0.012)	(0.219)
	MQL	0.001	0.002	−0.006	−0.011	−0.010	0.024	−0.001	0.001	−0.006
		(0.030)	(0.005)	(0.104)	(0.143)	(0.016)	(0.194)	(0.169)	(0.011)	(0.215)
	CML	0.006	0.004	−0.014	0.020	0.006	−0.021	0.005	0.004	−0.019
		(0.007)	(0.002)	(0.039)	(0.022)	(0.005)	(0.080)	(0.011)	(0.003)	(0.072)

Table 4. *Cont.*

N	Method	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	λ_1	$\alpha_2^{(1)}$	$\alpha_2^{(2)}$	λ_2	$\alpha_3^{(1)}$	$\alpha_3^{(2)}$	λ_3
300	CLS	0.001	0.000	−0.005	−0.003	0.000	0.009	−0.001	0.000	−0.001
		(0.011)	(0.002)	(0.039)	(0.050)	(0.006)	(0.067)	(0.052)	(0.004)	(0.077)
	MQL	0.000	0.000	−0.005	−0.003	−0.001	0.010	0.000	0.000	−0.002
		(0.010)	(0.001)	(0.034)	(0.049)	(0.006)	(0.067)	(0.052)	(0.004)	(0.076)
	CML	0.005	0.001	−0.011	0.000	0.004	0.000	0.002	0.002	−0.007
		(0.003)	(0.001)	(0.013)	(0.008)	(0.002)	(0.026)	(0.004)	(0.001)	(0.026)

Table 5. Bias and MSE for Series B of Model II (MSE in parentheses): CLS, MQL and CML.

N	Method	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	λ_1	$\alpha_2^{(1)}$	$\alpha_2^{(2)}$	λ_2	$\alpha_3^{(1)}$	$\alpha_3^{(2)}$	λ_3
50	CLS	0.009	0.003	−0.019	0.005	−0.012	0.016	0.006	−0.003	−0.002
		(0.038)	(0.007)	(2.068)	(0.382)	(0.043)	(5.495)	(0.217)	(0.026)	(5.702)
	MQL	0.008	0.003	−0.017	−0.055	−0.022	0.070	0.009	−0.002	−0.008
		(0.037)	(0.006)	(1.995)	(0.378)	(0.043)	(5.461)	(0.220)	(0.026)	(5.718)
	CML	0.007	0.004	−0.017	0.015	0.004	−0.025	0.014	0.007	−0.031
		(0.005)	(0.002)	(0.590)	(0.025)	(0.006)	(1.380)	(0.008)	(0.004)	(1.326)
100	CLS	−0.001	−0.002	0.007	−0.006	−0.004	0.007	−0.006	−0.002	0.011
		(0.019)	(0.003)	(1.143)	(0.190)	(0.023)	(3.017)	(0.114)	(0.011)	(2.871)
	MQL	0.000	−0.002	0.004	−0.005	−0.004	0.007	−0.006	−0.003	0.012
		(0.018)	(0.003)	(1.091)	(0.189)	(0.023)	(3.001)	(0.115)	(0.012)	(2.882)
	CML	0.006	0.002	−0.012	0.008	0.004	−0.017	0.001	0.007	−0.017
		(0.002)	(0.001)	(0.238)	(0.012)	(0.003)	(0.691)	(0.004)	(0.002)	(0.660)
300	CLS	−0.003	−0.001	0.004	−0.004	−0.001	−0.006	0.003	−0.002	−0.006
		(0.006)	(0.001)	(0.361)	(0.062)	(0.007)	(0.889)	(0.033)	(0.004)	(0.848)
	MQL	−0.003	0.000	0.003	−0.002	0.000	−0.008	0.004	−0.002	−0.007
		(0.006)	(0.001)	(0.345)	(0.062)	(0.007)	(0.887)	(0.033)	(0.004)	(0.849)
	CML	0.000	0.001	−0.001	0.001	0.001	−0.011	0.004	0.002	−0.015
		(0.001)	(0.000)	(0.069)	(0.004)	(0.001)	(0.205)	(0.001)	(0.001)	(0.222)

Table 6. Bias and MSE for Series C of Model II (MSE in parentheses): CLS, MQL and CML.

N	Method	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	λ_1	$\alpha_2^{(1)}$	$\alpha_2^{(2)}$	λ_2	$\alpha_3^{(1)}$	$\alpha_3^{(2)}$	λ_3
50	CLS	−0.004	−0.002	0.069	−0.019	−0.008	0.061	−0.011	−0.008	0.131
		(0.038)	(0.007)	(2.068)	(0.382)	(0.043)	(5.495)	(0.217)	(0.026)	(5.702)
	MQL	−0.004	−0.002	0.067	−0.016	−0.007	0.051	−0.009	−0.007	0.122
		(0.037)	(0.006)	(1.995)	(0.378)	(0.043)	(5.461)	(0.220)	(0.026)	(5.718)
	CML	0.010	0.005	−0.019	0.037	0.014	−0.152	0.013	0.009	−0.038
		(0.005)	(0.002)	(0.590)	(0.025)	(0.006)	(1.380)	(0.008)	(0.004)	(1.326)
100	CLS	0.000	0.000	−0.005	−0.020	−0.004	0.054	0.001	−0.008	0.046
		(0.019)	(0.003)	(1.143)	(0.190)	(0.023)	(3.017)	(0.114)	(0.011)	(2.871)
	MQL	−0.002	−0.001	0.006	−0.020	−0.004	0.054	0.002	−0.008	0.045
		(0.018)	(0.003)	(1.091)	(0.189)	(0.023)	(3.001)	(0.115)	(0.012)	(2.882)
	CML	0.008	0.003	−0.059	0.016	0.005	−0.068	0.009	0.003	−0.047
		(0.002)	(0.001)	(0.238)	(0.012)	(0.003)	(0.691)	(0.004)	(0.002)	(0.660)
300	CLS	0.000	−0.001	−0.007	−0.005	−0.001	0.010	−0.014	−0.004	0.071
		(0.006)	(0.001)	(0.361)	(0.062)	(0.007)	(0.889)	(0.033)	(0.004)	(0.848)
	MQL	0.000	−0.001	−0.008	−0.005	−0.001	0.011	−0.014	−0.004	0.072
		(0.006)	(0.001)	(0.345)	(0.062)	(0.007)	(0.887)	(0.033)	(0.004)	(0.849)
	CML	0.000	0.000	−0.012	0.005	0.001	−0.020	0.004	0.002	−0.021
		(0.001)	(0.000)	(0.069)	(0.004)	(0.001)	(0.205)	(0.001)	(0.001)	(0.222)

To demonstrate the robustness of the MQL method, we consider the simulations about Model III with different series by using CLS, MQL and CML methods, and set $N = 300$, $\rho = (0.9, 0.9, 0.9), (0.8, 0.8, 0.8)$, respectively. From Tables 7–9, we can see that when ρ varies from (0.9, 0.9, 0.9) down to (0.8, 0.8, 0.8), the effect on CLS and MQL estimators is slight. Most of the biases and MSE of MQL estimators are smaller than CLS. But due to

incorrect distribution used, the biases and MSE of CML estimators increase. This indicates that the MQL method is more robust than CLS and CML methods.

Table 7. Bias and MSE for Series A of Model III with $N = 300$ (MSE in parentheses).

ρ	Method	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	λ_1	$\alpha_2^{(1)}$	$\alpha_2^{(2)}$	λ_2	$\alpha_3^{(1)}$	$\alpha_3^{(2)}$	λ_3
(0.9, 0.9, 0.9)	CLS	0.002	0.002	−0.004	0.009	0.004	−0.014	−0.007	0.000	0.002
		(0.010)	(0.002)	(0.049)	(0.022)	(0.004)	(0.041)	(0.026)	(0.004)	(0.055)
	MQL	0.002	0.002	−0.004	0.009	0.004	−0.014	−0.007	−0.001	0.003
		(0.009)	(0.002)	(0.042)	(0.021)	(0.004)	(0.040)	(0.026)	(0.004)	(0.053)
	CML	−0.021	−0.009	0.046	−0.043	−0.018	0.057	−0.055	−0.022	0.081
		(0.006)	(0.001)	(0.027)	(0.013)	(0.003)	(0.030)	(0.012)	(0.003)	(0.034)
(0.8, 0.8, 0.8)	CLS	−0.001	−0.001	0.000	0.005	−0.004	0.005	−0.005	−0.004	0.012
		(0.010)	(0.002)	(0.048)	(0.026)	(0.004)	(0.044)	(0.030)	(0.004)	(0.056)
	MQL	−0.001	−0.001	0.000	0.005	−0.004	0.006	−0.008	−0.005	0.016
		(0.009)	(0.002)	(0.042)	(0.026)	(0.004)	(0.043)	(0.030)	(0.004)	(0.054)
	CML	−0.042	−0.018	0.088	−0.080	−0.040	0.122	−0.121	−0.049	0.183
		(0.007)	(0.002)	(0.033)	(0.015)	(0.004)	(0.041)	(0.028)	(0.004)	(0.067)

Table 8. Bias and MSE for Series B of Model III with $N = 300$ (MSE in parentheses).

ρ	Method	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	λ_1	$\alpha_2^{(1)}$	$\alpha_2^{(2)}$	λ_2	$\alpha_3^{(1)}$	$\alpha_3^{(2)}$	λ_3
(0.9, 0.9, 0.9)	CLS	0.003	0.001	−0.001	0.001	0.000	0.000	0.004	0.000	−0.006
		(0.020)	(0.003)	(0.041)	(0.031)	(0.005)	(0.068)	(0.018)	(0.004)	(0.083)
	MQL	0.002	0.001	0.001	0.001	−0.001	0.001	0.003	−0.001	−0.003
		(0.018)	(0.002)	(0.036)	(0.030)	(0.005)	(0.065)	(0.017)	(0.004)	(0.080)
	CML	−0.023	−0.009	0.041	−0.080	−0.033	0.122	−0.065	−0.030	0.140
		(0.007)	(0.001)	(0.018)	(0.019)	(0.004)	(0.050)	(0.014)	(0.003)	(0.069)
(0.8, 0.8, 0.8)	CLS	0.001	0.001	−0.006	−0.005	−0.005	0.017	−0.002	−0.002	0.009
		(0.023)	(0.003)	(0.045)	(0.033)	(0.006)	(0.070)	(0.018)	(0.004)	(0.083)
	MQL	0.002	0.002	−0.008	−0.004	−0.005	0.016	−0.002	−0.002	0.010
		(0.021)	(0.002)	(0.039)	(0.032)	(0.005)	(0.067)	(0.017)	(0.004)	(0.078)
	CML	−0.043	−0.015	0.064	−0.156	−0.065	0.240	−0.122	−0.054	0.263
		(0.009)	(0.001)	(0.021)	(0.040)	(0.008)	(0.104)	(0.023)	(0.005)	(0.119)

Table 9. Bias and MSE for Series C of Model III with $N = 300$ (MSE in parentheses).

| ρ	Method	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	λ_1	$\alpha_2^{(1)}$	$\alpha_2^{(2)}$	λ_2	$\alpha_3^{(1)}$	$\alpha_3^{(2)}$	λ_3
(0.9, 0.9, 0.9)	CLS	0.003	0.001	−0.024	−0.014	−0.007	0.078	−0.002	−0.002	0.021
		(0.006)	(0.002)	(0.534)	(0.020)	(0.005)	(0.485)	(0.014)	(0.004)	(0.643)
	MQL	0.003	0.001	−0.020	−0.013	−0.007	0.077	−0.001	−0.002	0.018
		(0.005)	(0.002)	(0.472)	(0.020)	(0.005)	(0.484)	(0.014)	(0.004)	(0.630)
	CML	−0.044	−0.028	0.432	−0.122	−0.064	0.631	−0.186	−0.097	1.250
		(0.005)	(0.002)	(0.470)	(0.020)	(0.006)	(0.595)	(0.044)	(0.013)	(2.143)
(0.8, 0.8, 0.8)	CLS	0.001	−0.001	−0.003	−0.008	−0.003	0.036	0.000	−0.001	0.005
		(0.005)	(0.002)	(0.448)	(0.023)	(0.005)	(0.494)	(0.016)	(0.004)	(0.668)
	MQL	0.000	−0.001	0.002	−0.008	−0.002	0.034	−0.001	−0.002	0.007
		(0.005)	(0.002)	(0.407)	(0.023)	(0.005)	(0.490)	(0.015)	(0.004)	(0.661)
	CML	−0.074	−0.045	0.706	−0.158	−0.085	0.811	−0.296	−0.144	1.907
		(0.008)	(0.003)	(0.754)	(0.027)	(0.009)	(0.800)	(0.098)	(0.024)	(4.230)
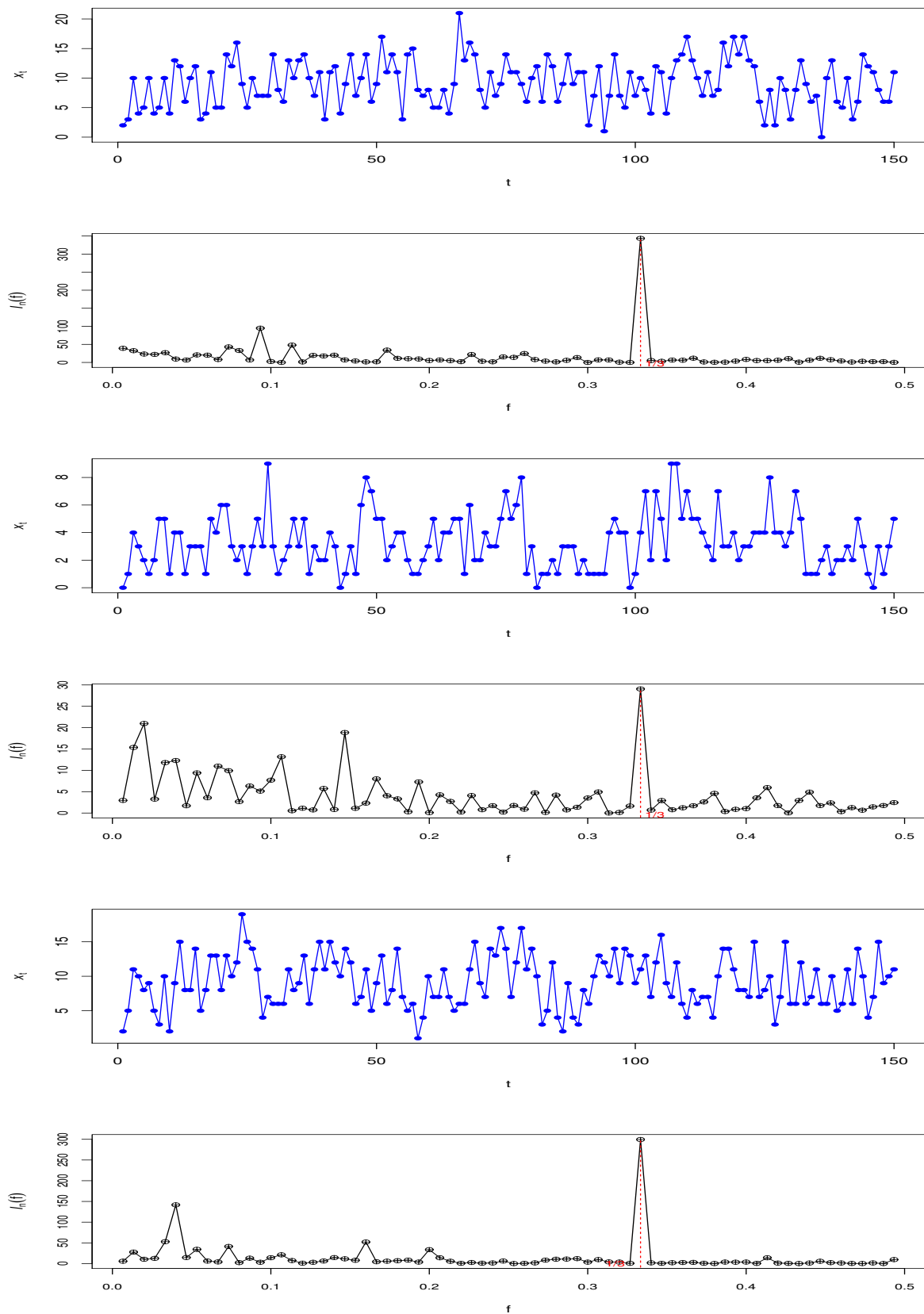
Figure 3. The sample path and periodogram of Series A(top), B(middle) and C(bottom) in Model I.

4.2. Performances of $\hat{r}_{MQL}$ and $\hat{r}_{CLS}$

As discussed in Section 3.2, we estimate the thresholds vector by using conditional least squares and modified quasi-likelihood methods. The performances of $\hat{r}_{MQL}$ and $\hat{r}_{CLS}$ are compared in this subsection through simulation studies. From the simulation results in Section 4.1, we find that the contaminated data generated from Model III has little influence on least squares and quasi-likelihood estimators, so we only simulate thresholds estimation for different series under Model I and Model II. We assess the performance of r by the bias, MSE and bias median, where the bias median is defined by:

$$\text{Bias median} = \underset{i \in \{1,2,\dots,K\}}{\text{median}} (\hat{r}_{ij} - r_{0j}), j = 1, 2, \dots, T,$$

where $\hat{r}_{ij}$ is the estimator of r_{0j}, r_{0j} is the true value with $j = 1, 2, \dots, T$, and K is the number of replications. The simulation results are summarized in Tables 10–15.

Table 10. Bias, bias median and MSE for Series A of Model I.

N	Para.	MQL			CLS		
		Bias	Median	MSE	Bias	Median	MSE
50	r_1	−0.167	0	0.447	0.042	0	0.550
	r_2	0.422	0	1.986	0.723	0	2.841
	r_3	0.457	0	1.975	0.947	0	3.779
100	r_1	−0.107	0	0.151	−0.003	0	0.137
	r_2	0.224	0	1.378	0.570	0	2.428
	r_3	0.245	0	0.861	0.505	0	1.903
300	r_1	−0.007	0	0.007	0.000	0	0.002
	r_2	0.027	0	0.283	0.117	0	0.477
	r_3	0.021	0	0.035	0.066	0	0.200

Table 11. Bias, bias median and MSE for Series B of Model I.

N	Para.	MQL			CLS		
		Bias	Median	MSE	Bias	Median	MSE
50	r_1	0.499	0	2.129	1.294	1	4.176
	r_2	0.538	0	2.320	0.868	0	3.142
	r_3	0.139	0	2.687	0.634	0	3.610
100	r_1	0.555	1	1.933	1.301	1	3.597
	r_2	0.283	0	1.437	0.643	0	2.473
	r_3	0.107	0	2.537	0.599	0	3.431
300	r_1	0.480	1	1.518	1.215	1	2.485
	r_2	0.021	0	0.213	0.141	0	0.489
	r_3	−0.095	0	1.191	0.261	0	1.825

Table 12. Bias, bias median and MSE for Series C of Model I.

N	Para.	MQL			CLS		
		Bias	**Median**	**MSE**	**Bias**	**Median**	**MSE**
50	r_1	−0.012	0	0.588	0.023	0	0.661
	r_2	0.268	0	5.378	0.541	0	5.909
	r_3	0.155	0	1.433	0.216	0	1.750
100	r_1	0.015	0	0.079	0.023	0	0.081
	r_2	0.072	0	2.332	0.254	0	2.972
	r_3	0.041	0	0.325	0.050	0	0.330
300	r_1	0.000	0	0.000	0.000	0	0.000
	r_2	−0.015	0	0.317	0.027	0	0.457
	r_3	0.002	0	0.004	0.002	0	0.004

Table 13. Bias, bias median and MSE for Series A of Model II.

N	Para.	MQL			CLS		
		Bias	**Median**	**MSE**	**Bias**	**Median**	**MSE**
50	r_1	0.027	0	1.231	0.407	0	2.227
	r_2	1.025	0	4.897	1.293	1	6.051
	r_3	1.582	1	7.600	2.003	1	9.905
100	r_1	−0.013	0	0.489	0.185	0	0.723
	r_2	0.944	0	4.808	1.271	0	6.215
	r_3	1.539	0	8.391	2.005	1	11.269
300	r_1	−0.042	0	0.066	0.022	0	0.070
	r_2	0.652	0	3.560	0.940	0	5.088
	r_3	0.605	0	3.243	1.062	0	6.540

Table 14. Bias, bias median and MSE for Series B of Model II.

N	Para.	MQL			CLS		
		Bias	**Median**	**MSE**	**Bias**	**Median**	**MSE**
50	r_1	1.231	1	5.527	2.134	2	9.638
	r_2	1.307	1	6.063	1.633	1	7.439
	r_3	0.840	0	6.658	1.237	1	8.211
100	r_1	1.070	1	3.954	1.972	2	8.050
	r_2	1.208	0	5.772	1.561	1	7.375
	r_3	0.998	0	7.652	1.488	1	9.644
300	r_1	1.059	1	3.143	1.829	2	5.611
	r_2	0.717	0	3.465	1.031	0	4.961
	r_3	0.617	0	5.925	1.153	0	8.549

Table 15. Bias, bias median and MSE for Series C of Model II.

N	Para.	MQL			CLS		
		Bias	Median	MSE	Bias	Median	MSE
50	r_1	−1.066	0	11.494	−0.859	0	12.671
	r_2	0.006	0	18.430	0.149	0	19.137
	r_3	−0.337	−1	27.211	−0.206	−1	27.764
100	r_1	−0.130	0	4.220	0.078	0	5.250
	r_2	0.538	0	22.610	0.696	0	23.536
	r_3	0.241	0	26.911	0.386	0	28.340
300	r_1	−0.040	0	0.236	−0.016	0	0.262
	r_2	1.213	0	26.909	1.389	0	28.515
	r_3	0.794	0	19.586	0.961	0	21.521

From Tables 10–15, we can see that all the simulation results perform better as sample size N increases, which implies that the estimators are consistent. The results in Tables 10–12 have smaller biases, bias medians and MSE than in Tables 13–15. This might be because the variance of Model II is larger than Model I for each series. Moreover, almost all the biases, bias medians and MSE of MQL estimators are smaller than CLS estimators, and the MSE of some MQL estimators are even half of the CLS. Because the thresholds are integer values, when we assess the accuracy of the estimators, the bias medians estimated can be more reasonable. It is concluded that it is much better to estimate the thresholds with the MQL method than CLS.

In the process of simulation, we generate the data with $X_0 = 0$; however, 0 is not the mean of the process, so we generate a set of data, discard some data generated first, and use the remaining data for inference, namely, "burn in" samples. Here, we generate a set of data with a length of 1800. We do the simulations for Series A of Model I, Model II and Model III ($\rho = (0.8, 0.8, 0.8)$). Other simulation settings are the same as before. The simulation results are listed in Tables 16–20. From these tables, we can see that under the "burn in" samples, the estimated results are similar to that when the initial value is 0, which indicates that the initial value will not affect our estimated results.

Table 16. Bias and MSE for Series A of Model I with "burn in" samples (MSE in parentheses): CLS, MQL and CML.

N	Method	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	λ_1	$\alpha_2^{(1)}$	$\alpha_2^{(2)}$	λ_2	$\alpha_3^{(1)}$	$\alpha_3^{(2)}$	λ_3
50	CLS	−0.002	−0.008	0.012	−0.018	−0.012	0.029	0.001	0.004	−0.008
		(0.067)	(0.017)	(0.338)	(0.132)	(0.024)	(0.241)	(0.168)	(0.025)	(0.351)
	MQL	0.001	−0.006	0.006	−0.016	−0.012	0.027	0.002	0.004	−0.007
		(0.066)	(0.017)	(0.331)	(0.134)	(0.024)	(0.240)	(0.174)	(0.024)	(0.347)
	CML	0.032	0.008	−0.061	0.043	0.007	−0.044	0.000	0.004	−0.005
		(0.027)	(0.008)	(0.124)	(0.056)	(0.015)	(0.125)	(0.046)	(0.012)	(0.125)
100	CLS	−0.005	−0.006	0.014	−0.011	−0.005	0.012	−0.001	−0.002	0.011
		(0.030)	(0.009)	(0.153)	(0.063)	(0.011)	(0.106)	(0.081)	(0.012)	(0.166)
	MQL	−0.006	−0.006	0.017	−0.012	−0.006	0.013	0.000	−0.002	0.008
		(0.028)	(0.008)	(0.138)	(0.061)	(0.010)	(0.103)	(0.078)	(0.012)	(0.158)
	CML	0.006	−0.001	−0.010	0.025	0.006	−0.031	0.000	0.001	0.003
		(0.015)	(0.004)	(0.069)	(0.035)	(0.007)	(0.069)	(0.021)	(0.006)	(0.061)
300	CLS	−0.001	−0.002	0.002	−0.003	−0.001	0.009	0.002	0.001	0.000
		(0.010)	(0.003)	(0.052)	(0.019)	(0.004)	(0.034)	(0.024)	(0.004)	(0.050)
	MQL	0.000	−0.001	0.000	−0.003	−0.001	0.008	0.001	0.001	0.002
		(0.009)	(0.002)	(0.047)	(0.019)	(0.003)	(0.033)	(0.024)	(0.003)	(0.049)
	CML	0.001	−0.001	−0.003	0.003	0.001	0.001	0.001	0.001	0.000
		(0.006)	(0.002)	(0.027)	(0.015)	(0.003)	(0.025)	(0.007)	(0.002)	(0.021)

Table 17. Bias and MSE for Series A of Model II with "burn in" samples (MSE in parentheses): CLS, MQL and CML.

N	Method	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	λ_1	$\alpha_2^{(1)}$	$\alpha_2^{(2)}$	λ_2	$\alpha_3^{(1)}$	$\alpha_3^{(2)}$	λ_3
50	CLS	0.009	0.000	−0.017	0.023	−0.011	0.005	−0.011	−0.004	0.005
		(0.067)	(0.011)	(0.242)	(0.306)	(0.035)	(0.424)	(0.303)	(0.026)	(0.479)
	MQL	0.010	0.000	−0.017	0.028	−0.010	0.000	−0.018	−0.005	0.012
		(0.065)	(0.010)	(0.227)	(0.302)	(0.035)	(0.420)	(0.355)	(0.026)	(0.505)
	CML	0.022	0.006	−0.042	0.053	0.013	−0.048	0.007	0.007	−0.032
		(0.015)	(0.004)	(0.075)	(0.045)	(0.010)	(0.151)	(0.022)	(0.007)	(0.148)
100	CLS	−0.011	−0.001	0.026	−0.013	0.000	0.015	−0.024	−0.007	0.022
		(0.034)	(0.005)	(0.123)	(0.151)	(0.016)	(0.210)	(0.157)	(0.012)	(0.223)
	MQL	−0.011	−0.001	0.026	−0.013	−0.001	0.016	−0.022	−0.006	0.019
		(0.033)	(0.005)	(0.117)	(0.148)	(0.016)	(0.208)	(0.156)	(0.012)	(0.219)
	CML	0.006	0.005	−0.004	0.018	0.007	−0.013	0.009	0.003	−0.015
		(0.008)	(0.002)	(0.039)	(0.025)	(0.005)	(0.075)	(0.010)	(0.003)	(0.076)
300	CLS	−0.004	−0.001	0.005	−0.001	−0.002	0.008	0.000	−0.005	0.006
		(0.010)	(0.002)	(0.037)	(0.050)	(0.005)	(0.069)	(0.052)	(0.003)	(0.074)
	MQL	−0.003	−0.001	0.004	−0.001	−0.002	0.009	0.001	−0.005	0.004
		(0.010)	(0.001)	(0.034)	(0.049)	(0.005)	(0.068)	(0.051)	(0.003)	(0.073)
	CML	0.001	0.001	−0.005	0.007	0.001	−0.001	0.000	−0.001	−0.002
		(0.003)	(0.001)	(0.013)	(0.008)	(0.002)	(0.028)	(0.003)	(0.001)	(0.028)

Table 18. Bias and MSE for Series A of Model III ($\rho = (0.8, 0.8, 0.8)$) with "burn in" samples (MSE in parentheses): CLS, MQL and CML.

N	Method	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	λ_1	$\alpha_2^{(1)}$	$\alpha_2^{(2)}$	λ_2	$\alpha_3^{(1)}$	$\alpha_3^{(2)}$	λ_3
50	CLS	−0.087	−0.040	0.214	0.018	−0.004	−0.007	0.019	0.002	−0.030
		(0.068)	(0.016)	(0.339)	(0.153)	(0.025)	(0.248)	(0.203)	(0.026)	(0.381)
	MQL	−0.011	−0.007	0.026	0.019	−0.003	−0.008	0.019	0.002	−0.031
		(0.065)	(0.014)	(0.292)	(0.155)	(0.024)	(0.244)	(0.203)	(0.026)	(0.376)
	CML	−0.016	−0.009	0.039	−0.012	−0.024	0.047	−0.109	−0.045	0.153
		(0.022)	(0.007)	(0.118)	(0.042)	(0.015)	(0.140)	(0.091)	(0.017)	(0.245)
100	CLS	−0.044	−0.017	0.103	−0.015	−0.006	0.020	−0.005	−0.003	0.008
		(0.033)	(0.008)	(0.162)	(0.075)	(0.012)	(0.132)	(0.100)	(0.013)	(0.199)
	MQL	−0.008	−0.002	0.013	−0.015	−0.006	0.020	−0.004	−0.003	0.008
		(0.030)	(0.007)	(0.137)	(0.074)	(0.012)	(0.129)	(0.099)	(0.012)	(0.197)
	CML	−0.043	−0.017	0.088	−0.057	−0.033	0.093	−0.129	−0.048	0.186
		(0.014)	(0.004)	(0.073)	(0.027)	(0.008)	(0.083)	(0.062)	(0.010)	(0.156)
300	CLS	−0.016	−0.006	0.036	0.000	−0.002	0.000	0.004	−0.001	−0.003
		(0.010)	(0.002)	(0.048)	(0.026)	(0.004)	(0.043)	(0.030)	(0.004)	(0.057)
	MQL	−0.003	−0.001	0.003	−0.001	−0.002	0.003	0.004	−0.001	−0.003
		(0.009)	(0.002)	(0.043)	(0.025)	(0.004)	(0.042)	(0.029)	(0.004)	(0.054)
	CML	−0.047	−0.020	0.097	−0.081	−0.037	0.113	−0.112	−0.046	0.169
		(0.007)	(0.002)	(0.035)	(0.016)	(0.004)	(0.038)	(0.025)	(0.004)	(0.061)

Table 19. Bias, bias median and MSE for Series A of Model I with "burn in" samples.

N	Para.	MQL				CLS		
		Bias	Median	MSE		Bias	Median	MSE
50	r_1	−0.180	0	0.400		0.053	0	0.393
	r_2	0.390	0	1.960		0.720	0	2.894
	r_3	0.580	0	2.322		0.963	0	3.583
100	r_1	−0.099	0	0.143		−0.007	0	0.081
	r_2	0.198	0	1.142		0.491	0	1.975
	r_3	0.218	0	0.800		0.455	0	1.585
300	r_1	−0.015	0	0.015		−0.004	0	0.004
	r_2	0.018	0	0.268		0.098	0	0.416
	r_3	0.018	0	0.036		0.058	0	0.170

Table 20. Bias, bias median and MSE for Series A of Model II with "burn in" samples.

N	Para.	MQL				CLS		
		Bias	Median	MSE		Bias	Median	MSE
50	r_1	−0.071	0	0.835		0.252	0	1.394
	r_2	1.156	0	5.640		1.436	1	6.878
	r_3	1.616	1	7.974		2.046	1	10.284
100	r_1	−0.110	0	0.320		0.099	0	0.473
	r_2	1.172	0	5.662		1.477	1	7.063
	r_3	1.518	1	7.508		1.947	1	10.059
300	r_1	−0.041	0	0.055		0.027	0	0.055
	r_2	0.648	0	3.364		0.940	0	4.884
	r_3	0.574	0	3.532		0.854	0	5.338

5. Real Data Example

In this section, we use the $PSETINAR(2;1,1)_T$ process to fit the series of monthly counts of claimants collecting short-term disability benefits. In the dataset, all the claimants are male, have cuts, lacerations or punctures, and are between the ages of 35 and 54. In addition, they all work in the logging industry and collect benefits from the Workers' Compensation Board (WCB) of British Columbia. The dataset consists of 120 observations, from 1985 to 1994 (Freeland [36]). The simulations were performed on the $\mathcal{R}3.6.2$ software. The threshold vector was calculated by the algorithms (the three-step algorithm of NeSS combined with quasi-likelihood principle and the algorithm of NeSS combined with least squares principle) described in Section 3.2. We uses the command *constrOptim* to optimize the objective function of the maximum likelihood estimation. Figure 4 shows the sample path, ACF and PACF plots of the observations. It can be seen from Figure 4 that this dataset is a dependent counting time series with periodic characteristic.

Sample Path

(**a**)

Sample Autocorrelation Function

(**b**)

Sample Partial Autocorrelation Function

(**c**)
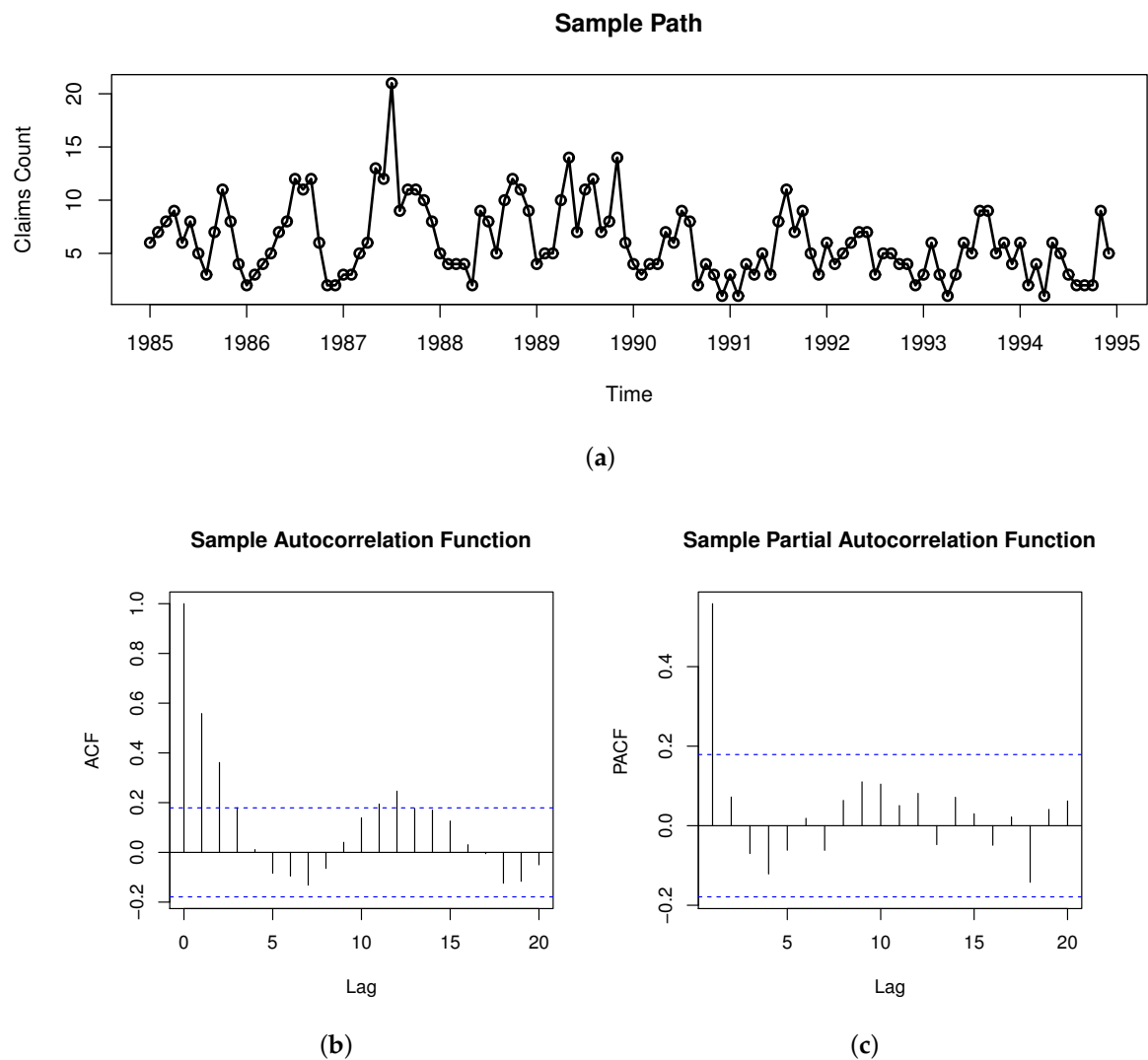
Figure 4. The sample path plot (**a**), ACF and PACF plots (**b**,**c**) for the counts of claimants.

We use the periodogram method to determine the period about this dataset and draw Figure 5, from which it can be seen that $I_n(f_k)$ reach maximum at $f_k = 1/12$, and concluded that $T = 12$. This displays the periodic characteristic of the data and exhibits a form of periodic change per year.
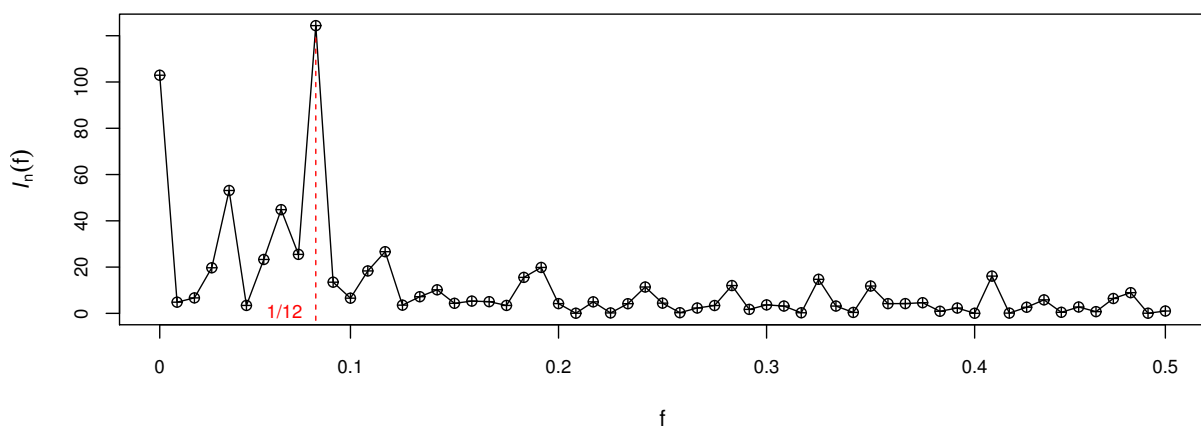
Figure 5. The periodogram plot for the monthly counts of claimants.

Table 21 displays the descriptive statistics for the monthly counts of claimants collecting short-term disability benefits from WCB. From Table 21, we can see that the mean and variance are approximately equal in some months. We can assume that the distribution of the innovations is a periodic Poisson. However, some months and the total data indicate overdispersion. We find that the dataset has no zero and the minimum value is one. This leads us to consider the periodic Poisson, periodic Geometric, zero-truncated periodic Poisson and zero-truncated periodic Geometric distributions for the innovations to fit the model, respectively. Before the model fitting, we first estimate the threshold vector. The $\hat{r}_{CLS}$ is calculated by (9) and the $\hat{r}_{MQL}$ is calculated through (10) by using the three-step algorithm. Table 22 summarizes the fitting results of $\hat{r}_{CLS}$ and $\hat{r}_{MQL}$. Due to the lesser data, to fit the model better, when the number of data in each regime is relatively smaller than two or the threshold is the maximum or minimum value of the boundary, we think that these monthly data do not have a piecewise phenomenon, that is, March, July, and August do not have piecewise phenomena.

Table 21. Summary statistics for the monthly counts of claimants.

	Whole Dataset	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
Mean	6.1	4.2	3.8	4.6	4.9	7.0	7.1	8.5	7.5	7.2	7.2	7.2	4.4
Variance	11.8	2.2	3.3	1.8	9.0	14.7	5.9	28.9	12.5	12.0	12.2	14.8	6.9
Maximum	21	6	7	8	10	14	12	21	12	12	12	14	19
Minimum	1	2	1	3	1	2	3	3	2	2	2	2	1

Table 22. Threshold estimators for the monthly counts of claimants.

	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
$\hat{r}_{CLS}$	3	4	7	5	5	6	10	4	9	6	7	6
$\hat{r}_{MQL}$	3	4	7	5	5	6	10	4	9	6	7	5

To capture the piecewise phenomenon of this time series dataset, we use $\text{PINAR}(1)_T$ and $\text{PSETINAR}(2;1,1)_T$ models with period $T = 12$ to fit the dataset, respectively. The PINAR(1) process proposed by Monteiro et al. [22] with the following recursive equation

$$X_t = \alpha_t \circ X_{t-1} + Z_t, \tag{12}$$

with $\alpha_t = \alpha_j \in (0,1)$ for $t = j + sT (j = 1, \ldots, T, s \in \mathbb{N}_0)$, the definition of thinning operator "$\circ$" and innovation process $\{Z_t\}$ is the same as the $\text{PSETINAR}(2;1,1)_T$ process.

It is worth mentioning that for this dataset, the conditional least squares and quasi-likelihood methods produce non-admissible estimators for some months, so we use the conditional maximum likelihood approach to estimate the parameters. Next, we use $\text{PSETINAR}(2;1,1)_{12}$ and $\text{PINAR}(1)_{12}$ models to fit the dataset in combination with the four innovation distributions mentioned before. Here, the threshold vectors are based on $\hat{r}_{MQL}$. The AIC and BIC are listed in Table 23. When we fit the dataset, we hope to get smaller AIC and BIC values. From the results in Table 23, we can conclude that the $\text{PSETINAR}(2;1,1)_{12}$ model with zero-truncated periodic Poisson distribution is more suitable. Then, we do the conditional maximum likelihood estimation, and the results are listed in Table 24. Some estimators of the parameters in Table 24, for example, the $\alpha^{(2)}$ of January, May, June, September, October and November, are not statistically significant, suggesting that on those months, the number of claims is mainly modeled through the innovation process.

Table 23. The AIC and BIC of the claims data.

PSETINAR$(2;1,1)_{12}$	AIC	BIC	PINAR$(1)_{12}$	AIC	BIC
Pois.	586.63	596.61	Pois.	592.12	599.38
Zero-truncated Pois.	581.65	591.64	Zero-truncated Pois.	594.44	601.71
Geom.	610.45	620.43	Geom.	605.56	612.82
Zero-truncated Geom.	586.36	596.34	Zero-truncated Geom.	595.15	602.42

Table 24. CML estimators in the dataset.

Month	$\alpha^{(1)}$	$\alpha^{(2)}$	λ
Jan.	0.112	8.907×10^{-08}	3.819
Feb.	0.227	0.032	3.060
Mar.	0.692	-	1.969
Apr.	0.999	0.240	2.048
May	0.586	8.521×10^{-09}	4.889
Jun.	0.265	4.316×10^{-08}	5.507
Jul.	0.360	-	5.942
Aug.	0.390	-	4.186
Sep.	0.380	3.366×10^{-07}	5.218
Oct.	0.502	1.027×10^{-07}	4.044
Nov.	0.433	2.776×10^{-08}	4.990
Dec.	0.508	0.222	1.000

Remark: "-" stand for not available.

To check the predictability of the PSETINAR$(2;1,1)_T$ model, we carry out the *h*-step-ahead forecasting for varying *h* of the PSETINAR$(2;1,1)_T$ model. The *h*-step-ahead conditional expectation point predictor of the PSETINAR$(2;1,1)_T$ model is given by

$$\hat{X}_{j+sT+h} = E\left[X_{j+sT+h}|X_{j+sT}\right], \, h = 1, 2, \ldots.$$

Specifically, the one-step-ahead conditional expectation point predictor is given by

$$\hat{X}_{j+sT+1} = E\left[X_{j+sT+1}|X_{j+sT}\right] = \alpha_{j+1}^{(1)}X_{j+sT}I_{j+sT}^{(1)} + \alpha_{j+1}^{(2)}X_{j+sT}I_{j+sT}^{(2)} + \lambda_{j+1}.$$

However, the conditional expectation will seldom produce integer-valued forecasts. Recently, coherent forecasting techniques have been recommended, which only produce forecasts in $\mathbb{N}_0$. This is achieved by computing the *h*-step-ahead forecasting conditional distribution. As pointed out by Möller et al. [37], this approach leads to forecasts themselves being easily obtained from the median or the mode of the forecasting distribution. In addition, Li et al. [38] and Kang et al. [8] have applied this method to forecast the integer-valued processes. Homburg et al. [39] discussed the prediction methods based on conditional distributions and Gaussian approximations and applied them to some integer-valued processes and compared them. For the PSETINAR$(2;1,1)_T$ process, the one-step-ahead conditional distribution of X_{j+sT+1} given X_{j+sT} is given by

$$P\left(X_{j+sT+1} = x_{j+sT+1}|X_{j+sT} = x_{j+sT}\right)$$

$$= \sum_{i=1}^{\min\{x_{j+sT}, x_{j+sT+1}\}} \sum_{k=1}^{2} \binom{x_{j+sT}}{i} \alpha_{j+1}^{(k)^i}\left(1 - \alpha_{j+1}^{(k)}\right)^{x_{j+sT} - i} I_{j+sT}^{(k)} P\left(Z_{j+sT+1} = x_{j+sT+1} - i\right).$$

Due to the existence of the threshold, while we use the conditional expectation method to predict $X_{j+sT+h}, h > 1$, we have to predict the previous moment of $X_{j+sT+h-1}$ first and compare it with the corresponding threshold before we do the next prediction. We do the same for the conditional distribution method. (To prevent confusion, we call this method a

point-wise conditional distribution forecast. The predictors completely based on *h*-step-ahead conditional distribution without intermediate step prediction will be discussed later.) The mode of *h*-step-ahead point-wise conditional distribution can be viewed as the point prediction. Here we compare the two forecasting methods, a standard descriptive measure of forecasting accuracy, namely, *h*-step-ahead predicted root mean squared error (PRMSE) is adopted. This measure can be given by

$$PRMSE = \sqrt{\frac{1}{K - K_0} \sum_{t=K_0+1}^{K} \left(X_{t+h} - \hat{X}_{t+h}\right)^2}, \ h = 1, 2, \ldots,$$

where *K* is the full sample size, we split the data into two parts, and the last $K - K_0$ observations as a forecasting evaluation sample. We forecast the value of the last year when $h = 1, 2, 3, 12$.

The PRMSEs of the *h*-step-ahead point predictors are list in Table 25. For conditional expectation point predictors, the PRMSEs of $PSETINAR(2; 1, 1)_{12}$ with zero-truncated periodic Poisson distribution are smaller than the $PINAR(1)_{12}$ with periodic Poisson and zero-truncated periodic Poisson distributions. This further shows the superiority of our model. The PRMSEs of the one-step-ahead point predictors are smaller than others. This is very natural because we use the value of the previous moment as the explanatory variable. For $PSETINAR(2; 1, 1)_{12}$ with zero-truncated periodic Poisson distribution, the PRMSEs of twelve-step-ahead predictors are smaller than other *h*-step-ahead predictors except for one-step-ahead. This may be because our period is 12. The PRMSE of one-step-ahead conditional expectation point predictors is smaller than point-wise conditional distribution point predictors. Thus, the former method is better for this dataset.

The PRMSEs of the one-step-ahead fitted series calculated by conditional expectation and conditional distribution are 2.434 and 3.565, respectively. This further illustrates that for our dataset, one-step-ahead forecasting conditional expectation is better than conditional distribution. The original data and the fitted series (calculated by the one-step-ahead conditional expectation based on the observations of the previous moments) by the $PSETINAR(2; 1, 1)_{12}$ model with zero-truncated periodic Poisson distribution are plotted in Figure 6. It is observed that the trend is similar to the original data. Except for the points with large value (the unexpected prediction may be due to the wrong judgement of regime), this model fits the data well.
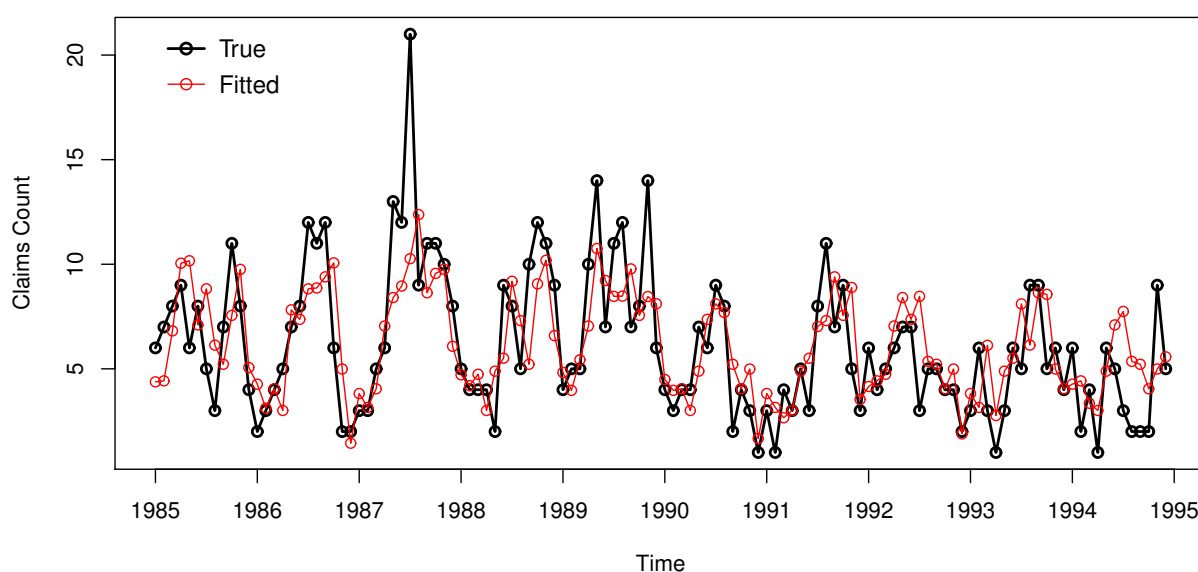
Figure 6. Plot of fitted curves of the claims data.

Table 25. PRMSE of the *h*-step-ahead point predictors.

	h	1	2	3	12
Conditional expectation	PSETINAR$(2;1,1)_{12}$ (Zero-truncated Pois.)	2.641	3.019	3.433	2.929
	PINAR$(1)_{12}$ (Zero-truncated Pois.)	2.753	3.377	3.567	3.788
	PINAR$(1)_{12}$ (Pois.)	2.724	3.407	3.704	4.008
Conditional distribution	PSETINAR$(2;1,1)_{12}$ (Zero-truncated Pois.)	2.814	3.000	3.109	2.930

Actually, we can get the *h*-step-ahead conditional distribution; here, we list the two-step-ahead and three-step-ahead conditional distributions as an example,

$$P\big(X_{j+sT+2} = x_{j+sT+2}|X_{j+sT} = x_{j+sT}\big)$$
$$= \sum_{m=0}^{n} P\big(X_{j+sT+1} = m|X_{j+sT} = x_{j+sT}\big)P\big(X_{j+sT+2} = x_{j+sT+2}|X_{j+sT+1} = m\big),$$

and

$$P\big(X_{j+sT+3} = x_{j+sT+3}|X_{j+sT} = x_{j+sT}\big)$$
$$= \sum_{m=0}^{n} P\big(X_{j+sT+2} = m|X_{j+sT} = x_{j+sT}\big)P\big(X_{j+sT+3} = x_{j+sT+3}|X_{j+sT+2} = m\big),$$

where $m \in \{0, 1, \ldots, n\}$ is the possible domain of X_{j+sT}, $j = 1, \ldots, T$, and $s \in \mathbb{N}_0$. When $h = 1, 2, 3$, we show the plots of the *h*-step-ahead conditional distribution in Figure 7, where x_{j+sT} represents the count of claimants in December 1993 and February 1994, respectively. The mode of h-step-ahead conditional distribution can be viewed as the point prediction. The PRMSEs of the two-step-ahead and three-step-ahead point predictors for the last year are 3.227 and 3.215, respectively, which is larger than the point-wise conditional distribution method described before. Maybe for other datasets or models, the *h*-step-ahead forecasting conditional distribution will show some advantages. We will not go into details here.
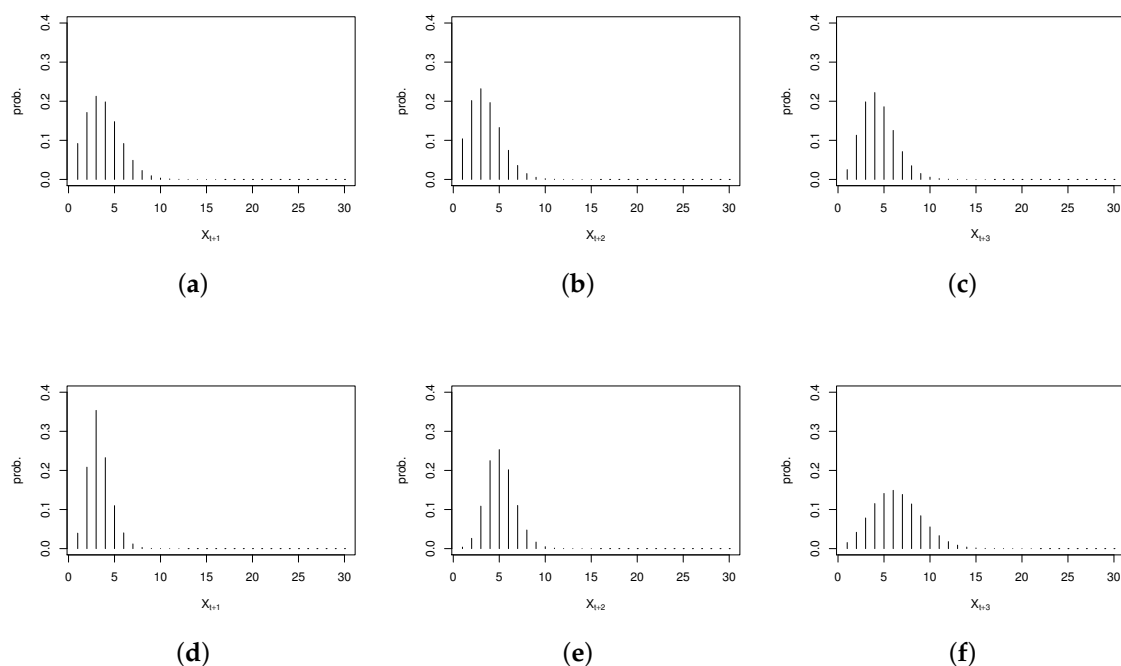
Figure 7. The *h*-step-ahead forecasting conditional distribution for the counts of claimants: (**a**–**c**) conditional on the count of claimants in December 1993; (**d**–**f**) conditional on the count of claimants in February 1994.

6. Conclusions

This paper extended the PSETINAR$(2; 1, 1)_T$ process proposed by Pereira et al. [25], by removing the assumption of Poisson distribution of $\{Z_t\}$ and considered the PSETINAR $(2; 1, 1)_T$ process under weak conditions that the second moment of $\{Z_t\}$ is finite. The ergodicity of the process is established. MQL-estimators of the model parameters vector β, MQL-estimators and CLS-estimators of the thresholds vector r are obtained. Moreover, through simulation, we can see the advantages of the quasi-likelihood method by comparing with the conditional maximum likelihood and conditional least square methods. An application to a real dataset is presented. In addition, the forecasting problem of this dataset is addressed.

In this paper, we only discuss the PSETINAR$(2; 1, 1)_T$ process for univariate time series. Hence, an extension for the multivariate PSETINAR$(2; 1, 1)_T$ process with a diagonal or cross-correlation autoregressive matrix is a topic for future investigation. Furthermore, it is also important to stress that beyond this extension, there are a number of interesting problems for future research in this area. For example, even a simple periodic model can have an inordinately large number of parameters. This is also true for PSETINAR$(2; 1, 1)_T$ models and even multi-period models. Therefore, the development of procedures of dimensionality reduction to overcome the computational difficulties is an impending problem. This remains a topic of future research.

Appendix A

Proof of Theorem 1. According to Theorem 2 of Tweedie [40] (see also, Zheng and Basawa [41]), for the process defined by (2), and $\forall j = 1, 2, \ldots, T, s \in \mathbb{Z}$, we have

$$E\big(X_{j+sT} | X_{j+sT-1} = x\big) = \alpha_j^{(1)} x I_{j+sT-1}^{(1)} + \alpha_j^{(2)} x I_{j+sT-1}^{(2)} + \lambda_j \leq \alpha_{j,\max} x + \lambda_j,$$

where $\alpha_{j,\max} = \max\{\alpha_j^{(1)}, \alpha_j^{(2)}\} < 1$.

Let $K = \left[\frac{1+\lambda_j}{1-\alpha_{j,\max}}\right] + 1$, where $[\cdot]$ denotes the integer part of a number. Then for $x \geq K$, we have

$$E\big(X_{j+sT} | X_{j+sT-1} = x\big) \leq x - 1,$$

and for $x < K$,

$$E\big(X_{j+sT} | X_{j+sT-1} = x\big) \leq \alpha_{j,\max} x + \lambda_j \leq K + \lambda_j < \infty.$$

Therefore, the process $\{X_t\}$ for $t = j + sT$ defined in (2) is an ergodic Markov chain. □

Proof of Proposition 2. (i) From Proposition 1, we have

$$Var(X_{j+sT}|X_{j+sT-1}) = E\{X_{j+sT} - E(X_{j+sT}|X_{j+sT-1})\}^2$$
$$= E\left(X_{j+sT} - \alpha_j^{(1)} X_{j+sT-1} I_{j+sT-1}^{(1)} - \alpha_j^{(2)} X_{j+sT-1} I_{j+sT-1}^{(2)} - \lambda_j\right)^2,$$

and

$$Var(X_{j+sT}|X_{j+sT-1}) = \sum_{k=1}^{2} \alpha_j^{(k)}\left(1 - \alpha_j^{(k)}\right) X_{j+sT-1} I_{j+sT-1}^{(k)} + \sigma_{z,j}^2,$$

with $k = 1, 2, j = 1, \ldots, T, s \in \mathbb{N}_0$, so by substituting suitable consistent estimators of $\alpha_j^{(k)}$ and λ_j, we can get the consistent estimation of $\sigma_{z,j}^2$,

$$\hat{\sigma}_{1,z,j}^2 = \frac{1}{N} \sum_{s=0}^{N-1}\left(X_{j+sT} - \sum_{k=1}^{2} \hat{\alpha}_j^{(k)} X_{j+sT-1} I_{j+sT-1}^{(k)} - \hat{\lambda}_j\right)^2$$
$$- \frac{1}{N} \sum_{k=1}^{2} \sum_{s=0}^{N-1} \hat{\alpha}_j^{(k)}\left(1 - \hat{\alpha}_j^{(k)}\right) X_{j+sT-1} I_{j+sT-1}^{(k)}.$$

(ii) Moreover, from model (2), we have

$$Var(X_{j+sT}) = \sum_{k=1}^{2} Var\left\{\left(\alpha_j^{(k)} \circ X_{j+sT-1}\right) I_{j+sT-1}^{(k)}\right\} + Var(Z_{j+sT})$$
$$+ 2Cov\left\{\left(\alpha_j^{(1)} \circ X_{j+sT-1}\right) I_{j+sT-1}^{(1)}, \left(\alpha_j^{(2)} \circ X_{j+sT-1}\right) I_{j+sT-1}^{(2)}\right\},$$

where

$$Var\left\{\left(\alpha_j^{(k)} \circ X_{j+sT-1}\right) I_{j+sT-1}^{(k)}\right\}$$
$$= Var\left\{E\left[\left(\alpha_j^{(k)} \circ X_{j+sT-1}\right) I_{j+sT-1}^{(k)}|X_{j+sT-1}\right]\right\} + E\left\{Var\left[\left(\alpha_j^{(k)} \circ X_{j+sT-1}\right) I_{j+sT-1}^{(k)}|X_{j+sT-1}\right]\right\}$$
$$= \alpha_j^{(k)2} Var\left(X_{j+sT-1} I_{j+sT-1}^{(k)}\right) + \alpha_j^{(k)}\left(1 - \alpha_j^{(k)}\right) E\left(X_{j+sT-1} I_{j+sT-1}^{(k)}\right),$$

and

$$2Cov\left\{\left(\alpha_j^{(1)} \circ X_{j+sT-1}\right) I_{j+sT-1}^{(1)}, \left(\alpha_j^{(2)} \circ X_{j+sT-1}\right) I_{j+sT-1}^{(2)}\right\}$$
$$= -2E\left\{\left(\alpha_j^{(1)} \circ X_{j+sT-1}\right) I_{j+sT-1}^{(1)}\right\} E\left\{\left(\alpha_j^{(2)} \circ X_{j+sT-1}\right) I_{j+sT-1}^{(2)}\right\}$$
$$= -2\alpha_j^{(1)} \alpha_j^{(2)} E\left\{X_{j+sT-1} I_{j+sT-1}^{(1)}\right\} E\left\{X_{j+sT-1} I_{j+sT-1}^{(2)}\right\}.$$

Note that

$$E\left\{X_{j+sT-1} I_{j+sT-1}^{(1)}\right\}$$
$$= E\left\{E\left[X_{j+sT-1} I_{j+sT-1}^{(1)}|I_{j+sT-1}^{(1)} = 1\right]\right\}$$
$$= E\left\{X_{j+sT-1} I_{j+sT-1}^{(1)}|I_{j+sT-1}^{(1)} = 1\right\} P\left(I_{j+sT-1}^{(1)} = 1\right)$$
$$= E\left\{X_{j+sT-1}|X_{j+sT-1} \le r_j\right\} P\left(I_{j+sT-1}^{(1)} = 1\right).$$

Let $p_j = P\left(I_{j+sT-1}^{(1)} = 1\right)$ with $j = 1, \ldots, T, s \in \mathbb{N}_0$, we can estimate it with $\hat{p}_j = \frac{1}{N}\sum_{s=0}^{N-1} I_{j+sT-1}^{(1)}$.

Therefore, by substituting a suitable consistent estimator of $\alpha_j^{(k)}$, based on moment estimation, we can get the estimator $\hat{\sigma}_{2,z,j}^2$ in Proposition 2. $\quad\square$

Proof of Theorem 2. Let $\mathcal{F}_{j+sT} = \sigma\left(X_0, X_1, \ldots, X_{j+sT}\right)$ with $j = 1, \ldots, T, s \in \mathbb{N}_0$. First, we suppose $\boldsymbol{\theta}$ is known, for the following estimation equations:

$$S_{N,j}^{(1)}(\boldsymbol{\theta}_j, \beta_j) = \sum_{s=0}^{N-1} V_{\boldsymbol{\theta}_j}^{-1}\left(X_{j+sT} - \alpha_j^{(1)} X_{j+sT-1} I_{j+sT-1}^{(1)} - \alpha_j^{(2)} X_{j+sT-1} I_{j+sT-1}^{(2)} - \lambda_j\right) X_{j+sT-1} I_{j+sT-1}^{(1)},$$

we have

$$E\left[V_{\boldsymbol{\theta}_j}^{-1}\left(X_{j+sT} - \alpha_j^{(1)} X_{j+sT-1} I_{j+sT-1}^{(1)} - \alpha_j^{(2)} X_{j+sT-1} I_{j+sT-1}^{(2)} - \lambda_j\right) X_{j+sT-1} I_{j+sT-1}^{(1)} | \mathcal{F}_{j+sT-1}\right]$$

$$= V_{\boldsymbol{\theta}_j}^{-1} X_{j+sT-1} I_{j+sT-1}^{(1)} E\left[\left(X_{j+sT} - E\left(X_{j+sT} | X_{j+sT-1}\right)\right) | \mathcal{F}_{j+sT-1}\right]$$

$$= V_{\boldsymbol{\theta}_j}^{-1} X_{j+sT-1} I_{j+sT-1}^{(1)} E\left[\left(X_{j+sT} - E\left(X_{j+sT} | X_{j+sT-1}\right)\right) | X_{j+sT-1}\right]$$

$$= 0,$$

and

$$E\left[S_{s,j}^{(1)}(\boldsymbol{\theta}_j, \beta_j) | \mathcal{F}_{j+sT-1}\right]$$

$$= E\left[V_{\boldsymbol{\theta}_j}^{-1}\left(X_{j+sT} - \alpha_j^{(1)} X_{j+sT-1} I_{j+sT-1}^{(1)} - \alpha_j^{(2)} X_{j+sT-1} I_{j+sT-1}^{(2)} - \lambda_j\right) X_{j+sT-1} I_{j+sT-1}^{(1)} + S_{s-1,j}^{(1)}(\boldsymbol{\theta}_j, \beta_j) | \mathcal{F}_{j+sT-1}\right]$$

$$= E\left[S_{s-1,j}^{(1)}(\boldsymbol{\theta}_j, \beta_j) | \mathcal{F}_{j+sT-1}\right]$$

$$= S_{s-1,j}^{(1)}(\boldsymbol{\theta}_j, \beta_j),$$

so $\{S_{s,j}^{(1)}(\boldsymbol{\theta}_j, \beta_j), \mathcal{F}_{j+sT}, j = 1, 2, \ldots, T, s \in \mathbb{N}_0\}$ is a martingale. By Theorem 1.1 of Billingsley [42], we have

$$\frac{1}{N}\sum_{s=0}^{N-1} V_{\boldsymbol{\theta}_j}^{-2}\left(X_{j+sT} | X_{j+sT-1}\right)\left(X_{j+sT} - \alpha_j^{(1)} X_{j+sT-1} I_{j+sT-1}^{(1)} - \alpha_j^{(2)} X_{j+sT-1} I_{j+sT-1}^{(2)} - \lambda_j\right)^2 X_{j+sT-1}^2 I_{j+sT-1}^{(1)}$$

$$\xrightarrow{a.s.} E\left[V_{\boldsymbol{\theta}_j}^{-2}\left(X_j | X_{j-1}\right)\left(X_j - \alpha_j^{(1)} X_{j-1} I_{j-1}^{(1)} - \alpha_j^{(2)} X_{j-1} I_{j-1}^{(2)} - \lambda_j\right)^2 X_{j-1}^2 I_{j-1}^{(1)}\right]$$

$$= E\left\{E\left[V_{\boldsymbol{\theta}_j}^{-2}\left(X_j | X_{j-1}\right)\left(X_j - \alpha_j^{(1)} X_{j-1} I_{j-1}^{(1)} - \alpha_j^{(2)} X_{j-1} I_{j-1}^{(2)} - \lambda_j\right)^2 X_{j-1}^2 I_{j-1}^{(1)} | X_{j-1}\right]\right\}$$

$$= E\left\{V_{\boldsymbol{\theta}_j}^{-2}\left(X_j | X_{j-1}\right) X_{j-1}^2 I_{j-1}^{(1)} E\left[\left(X_j - \alpha_j^{(1)} X_{j-1} I_{j-1}^{(1)} - \alpha_j^{(2)} X_{j-1} I_{j-1}^{(2)} - \lambda_j\right)^2 | X_{j-1}\right]\right\}$$

$$= E\left\{V_{\boldsymbol{\theta}_j}^{-1}\left(X_j | X_{j-1}\right) X_{j-1}^2 I_{j-1}^{(1)}\right\}$$

$$\triangleq H_{j,11}(\boldsymbol{\theta}_j).$$

Thus, by the central limit theorem of martingale, we get

$$\frac{1}{\sqrt{N}} S_{N,j}^{(1)}(\boldsymbol{\theta}_j, \beta_j) \xrightarrow{L} \mathcal{N}\left(0, H_{j,11}(\boldsymbol{\theta}_j)\right).$$

Similarly,

$$\frac{1}{\sqrt{N}} S_{N,j}^{(2)}(\boldsymbol{\theta}_j, \beta_j) \xrightarrow{L} \mathcal{N}\left(0, H_{j,22}(\boldsymbol{\theta}_j)\right),$$

$$\frac{1}{\sqrt{N}} S_{N,j}^{(3)}(\boldsymbol{\theta}_j, \beta_j) \xrightarrow{L} \mathcal{N}\big(0, H_{j,33}(\boldsymbol{\theta}_j)\big),$$

where

$$S_{N,j}^{(2)}(\boldsymbol{\theta}_j, \beta_j) = \sum_{s=0}^{N-1} V_{\boldsymbol{\theta}_j}^{-1} \Big(X_{j+sT} - \alpha_j^{(1)} X_{j+sT-1} I_{j+sT-1}^{(1)} - \alpha_j^{(2)} X_{j+sT-1} I_{j+sT-1}^{(2)} - \lambda_j \Big) X_{j+sT-1} I_{j+sT-1}^{(2)},$$

and

$$S_{N,j}^{(3)}(\boldsymbol{\theta}_j, \beta_j) = \sum_{s=0}^{N-1} V_{\boldsymbol{\theta}_j}^{-1} \Big(X_{j+sT} - \alpha_j^{(1)} X_{j+sT-1} I_{j+sT-1}^{(1)} - \alpha_j^{(2)} X_{j+sT-1} I_{j+sT-1}^{(2)} - \lambda_j \Big).$$

For any $c = (c_1, c_2, \ldots, c_T)' \in \mathbb{R}^{3T} \setminus (0, 0, \ldots, 0)$, $c_j = \left(c_j^{(1)}, c_j^{(2)}, c_j^{(3)} \right)'$ with $j = 1, 2, \ldots, T$, to simplify, let

$$j + sT \neq i + kT, i, j = 1, 2, \ldots, T, s, k \in \mathbb{N}_0,$$

$$u_{j,s} = c_j^{(1)} X_{j+sT-1} I_{j+sT-1}^{(1)} + c_j^{(2)} X_{j+sT-1} I_{j+sT-1}^{(2)} + c_j^{(3)},$$

$$w_{j,s} = X_{j+sT} - E\big(X_{j+sT} | X_{j+sT-1}\big) = X_{j+sT} - \alpha_j^{(1)} X_{j+sT-1} I_{j+sT-1}^{(1)} - \alpha_j^{(2)} X_{j+sT-1} I_{j+sT-1}^{(2)} - \lambda_j,$$

and $n(T, N)$ is a constant associated with N and T, then

$$E\left[\sum_{j=1}^{T} \sum_{s=0}^{N-1} V_{\boldsymbol{\theta}_j}^{-1} \big(X_{j+sT} - E\big(X_{j+sT}|X_{j+sT-1}\big) \big) \left(c_j^{(1)} X_{j+sT-1} I_{j+sT-1}^{(1)} + c_j^{(2)} X_{j+sT-1} I_{j+sT-1}^{(2)} + c_j^{(3)} \right) \right]^2$$

$$= \sum_{j=1}^{T} \sum_{s=0}^{N-1} E\left[V_{\boldsymbol{\theta}_j}^{-2} w_{j,s}^2 u_{j,s}^2 \right] + n(T, N) E\left[V_{\boldsymbol{\theta}_j}^{-1} V_{\boldsymbol{\theta}_i}^{-1} w_{j,s} w_{i,k} u_{j,s} u_{i,k} \right], \tag{A1}$$

for the first item in the right side of Equation (A1), we have

$$E\left[V_{\boldsymbol{\theta}_j}^{-2} w_{j,s}^2 u_{j,s}^2 \right]$$

$$= E\left\{ E\left[V_{\boldsymbol{\theta}_j}^{-2} w_{j,s}^2 u_{j,s}^2 | X_{j+sT-1} \right] \right\}$$

$$= E\left\{ V_{\boldsymbol{\theta}_j}^{-2} u_{j,s}^2 E\left[w_{j,s}^2 | X_{j+sT-1} \right] \right\}$$

$$= E\left[V_{\boldsymbol{\theta}_j}^{-1} \left(c_j^{(1)} X_{j+sT-1} I_{j+sT-1}^{(1)} + c_j^{(2)} X_{j+sT-1} I_{j+sT-1}^{(2)} + c_j^{(3)} \right)^2 \right],$$

for the second item in the right side of Equation (A1), we have

$$E\left[V_{\boldsymbol{\theta}_j}^{-1} V_{\boldsymbol{\theta}_i}^{-1} w_{j,s} w_{i,k} u_{j,s} u_{i,k} \right]$$

$$= E\left\{ V_{\boldsymbol{\theta}_j}^{-1} V_{\boldsymbol{\theta}_i}^{-1} u_{j,s} u_{i,k} E\left[w_{j,s} w_{i,k} | X_{j+sT-1}, X_{i+kT-1} \right] \right\}$$

$$= 0,$$

which imply that $Cov(S_{N,j}, S_{N,i}) = 0$, where $S_{N,j} = \left(S_{N,j}^{(1)}(\boldsymbol{\theta}_j, \beta_j), S_{N,j}^{(2)}(\boldsymbol{\theta}_j, \beta_j), S_{N,j}^{(3)}(\boldsymbol{\theta}_j, \beta_j) \right)'$, $\forall i, j = 1, 2, \ldots, T, i \neq j$, then we have

$$\frac{c_j^T}{\sqrt{N}}\left(S_{N,j}^{(1)}(\boldsymbol{\theta}_j,\beta_j),S_{N,j}^{(2)}(\boldsymbol{\theta}_j,\beta_j),S_{N,j}^{(3)}(\boldsymbol{\theta}_j,\beta_j)\right)'$$

$$=\frac{1}{\sqrt{N}}\sum_{i=1}^{3}c_j^{(i)}S_{N,j}^{(i)}(\boldsymbol{\theta}_j,\beta_j)$$

$$=\frac{1}{\sqrt{N}}\sum_{s=0}^{N-1}V_{\boldsymbol{\theta}_j}^{-1}\left[X_{j+sT}-E(X_{j+sT}|X_{j+sT-1})\right]\left(c_j^{(1)}X_{j+sT-1}I_{j+sT-1}^{(1)}+c_j^{(2)}X_{j+sT-1}I_{j+sT-1}^{(2)}+c_j^{(3)}\right)$$

$$\xrightarrow{L}\mathcal{N}\left(0,E\left[V_{\boldsymbol{\theta}_j}^{-1}(X_j|X_{j-1})\left(c_j^{(1)}X_{j-1}I_{j-1}^{(1)}+c_j^{(2)}X_{j-1}I_{j-1}^{(2)}+c_j^{(3)}\right)^2\right]\right),\,as\,N\to\infty,$$

therefore, the $\frac{c_j^T}{\sqrt{N}}\left(S_{N,j}^{(1)}(\boldsymbol{\theta}_j,\beta_j),S_{N,j}^{(2)}(\boldsymbol{\theta}_j,\beta_j),S_{N,j}^{(3)}(\boldsymbol{\theta}_j,\beta_j)\right)'$ converges to a normal distribution with mean zero and variance $E\left[V_{\boldsymbol{\theta}_j}^{-1}(X_j|X_{j-1})\left(c_j^{(1)}X_{j-1}I_{j-1}^{(1)}+c_j^{(2)}X_{j-1}I_{j-1}^{(2)}+c_j^{(3)}\right)^2\right]$. Thus, by Cramer-wold device, it follows that

$$\frac{1}{\sqrt{N}}\begin{pmatrix}S_{N,1}^{(1)}(\boldsymbol{\theta}_1,\beta_1)\\S_{N,1}^{(2)}(\boldsymbol{\theta}_1,\beta_1)\\S_{N,1}^{(3)}(\boldsymbol{\theta}_1,\beta_1)\\\vdots\\S_{N,T}^{(1)}(\boldsymbol{\theta}_T,\beta_T)\\S_{N,T}^{(2)}(\boldsymbol{\theta}_T,\beta_T)\\S_{N,T}^{(3)}(\boldsymbol{\theta}_T,\beta_T)\end{pmatrix}\xrightarrow{L}\mathcal{N}\left(0,\begin{bmatrix}\boldsymbol{H}_1(\boldsymbol{\theta})&0&\cdots&0\\0&\boldsymbol{H}_2(\boldsymbol{\theta})&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&\boldsymbol{H}_T(\boldsymbol{\theta})\end{bmatrix}\right),$$

the 0's are (3×3)-null matrices. Now, we replace $V_{\boldsymbol{\theta}_j}(X_{j+sT}|X_{j+sT-1})$ by $V_{\hat{\boldsymbol{\theta}}_j}(X_{j+sT}|X_{j+sT-1})$, where $\hat{\boldsymbol{\theta}}_j$ is a consistent estimator of θ_j. We aim to get the result

$$\frac{1}{\sqrt{N}}\begin{pmatrix}S_{N,1}^{(1)}(\hat{\boldsymbol{\theta}}_1,\beta_1)\\S_{N,1}^{(2)}(\hat{\boldsymbol{\theta}}_1,\beta_1)\\S_{N,1}^{(3)}(\hat{\boldsymbol{\theta}}_1,\beta_1)\\\vdots\\S_{N,T}^{(1)}(\hat{\boldsymbol{\theta}}_T,\beta_T)\\S_{N,T}^{(2)}(\hat{\boldsymbol{\theta}}_T,\beta_T)\\S_{N,T}^{(3)}(\hat{\boldsymbol{\theta}}_T,\beta_T)\end{pmatrix}\xrightarrow{L}\mathcal{N}\left(0,\begin{bmatrix}\boldsymbol{H}_1(\boldsymbol{\theta})&0&\cdots&0\\0&\boldsymbol{H}_2(\boldsymbol{\theta})&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&\boldsymbol{H}_T(\boldsymbol{\theta})\end{bmatrix}\right). \tag{A2}$$

To prove (A2), we need to check the following conclusion

$$\frac{1}{\sqrt{N}}S_{N,j}^{(i)}(\hat{\boldsymbol{\theta}}_j,\beta_j)-\frac{1}{\sqrt{N}}S_{N,j}^{(i)}(\boldsymbol{\theta}_j,\beta_j)\xrightarrow{P}0,j=1,2,\ldots,T,i=1,2,3,N\to\infty. \tag{A3}$$

For $\forall\epsilon>0,\exists\delta>0$, we have

$$P\left(|\frac{1}{\sqrt{N}}S_N^{(1)}(\hat{\boldsymbol{\theta}}_j,\beta_j)-\frac{1}{\sqrt{N}}S_N^{(1)}(\boldsymbol{\theta}_j,\beta_j)|>\epsilon\right)\le\sum_{k=1}^{2}P\left(|\theta_{j_1}^{(k)}-\theta_j^{(k)}|>\delta\right)+P\left(|\sigma_{z,j_1}^2-\sigma_{z,j}^2|>\delta\right)$$
$$+P\left(\sup_{D}|\frac{1}{\sqrt{N}}S_N^{(1)}(\boldsymbol{\theta}_{j_1},\beta_j)-\frac{1}{\sqrt{N}}S_N^{(1)}(\boldsymbol{\theta}_j,\beta_j)|>\epsilon\right),$$

where $\boldsymbol{\theta}_{j_1} = \left(\theta_{j_1}^{(1)}, \theta_{j_1}^{(2)}, \sigma_{z,j_1}^2\right)'$, $D = \{\boldsymbol{\theta}_{j_1} : |\theta_{j_1}^{(1)} - \theta_j^{(1)}| < \delta, |\theta_{j_1}^{(2)} - \theta_j^{(2)}| < \delta, \sigma_{z,j_1}^2 - \sigma_{z,j}^2| < \delta\}$.
If $\hat{\boldsymbol{\theta}}_j$ is a consistent estimator of $\boldsymbol{\theta}_j$, then we just need to prove that

$$P\left(\sup_D |\frac{1}{\sqrt{N}}S_N^{(1)}(\boldsymbol{\theta}_{j_1}, \beta_j) - \frac{1}{\sqrt{N}}S_N^{(1)}(\boldsymbol{\theta}_j, \beta_j)| > \epsilon\right) \xrightarrow{P} 0, N \to \infty.$$

By the Markov inequality,

$$P\left(\sup_D |\frac{1}{\sqrt{N}}S_N^{(1)}(\boldsymbol{\theta}_{j_1}, \beta_j) - \frac{1}{\sqrt{N}}S_N^{(1)}(\boldsymbol{\theta}_j, \beta_j)| > \epsilon\right)$$

$$\leq \frac{1}{\epsilon^2}E\left(\sup_D\left(\frac{1}{\sqrt{N}}S_N^{(1)}(\boldsymbol{\theta}_{j_1}, \beta_j) - \frac{1}{\sqrt{N}}S_N^{(1)}(\boldsymbol{\theta}_j, \beta)\right)^2\right)$$

$$\leq \frac{1}{\epsilon^2}E\{\sup_D \frac{1}{N}\left[\sum_{s=0}^{N-1}\left(V_{\boldsymbol{\theta}_{j_1}}^{-1} - V_{\boldsymbol{\theta}_j}^{-1}\right)^2\left(X_{j+sT} - E(X_{j+sT}|X_{j+sT-1})\right)^2 X_{j+sT-1}^2 I_{j+sT-1}^{(1)}\right]\}$$

$$= \frac{1}{\epsilon^2}E\left[\sup_D\left(V_{\boldsymbol{\theta}_{j_1}}^{-1}(X_j|X_{j-1}) - V_{\boldsymbol{\theta}_j}^{-1}(X_j|X_{j-1})\right)^2\left(X_j - \alpha_j^{(1)}X_{j-1}I_{j-1}^{(1)} - \alpha_j^{(2)}X_{j-1}I_{j-1}^{(2)} - \lambda_j\right)^2 X_{j-1}^2 I_{j-1}^{(1)}\right]$$

$$= \frac{1}{\epsilon^2}E\{\sup_D \frac{\left[\left(\theta_j^{(1)} - \theta_{j_1}^{(1)}\right)X_{j-1}I_{j-1}^{(1)} + \left(\theta_j^{(1)} - \theta_{j_1}^{(2)}\right)X_{j-1}I_{j-1}^{(2)} + \left(\sigma_{z,j}^2 - \sigma_{z,j_1}^2\right)\right]^2}{V_{\boldsymbol{\theta}_{j_1}}^2(X_j|X_{j-1})V_{\boldsymbol{\theta}_j}(X_j|X_{j-1})}X_{j-1}^2 I_{j-1}^{(1)}\}$$

$$\leq \frac{1}{\epsilon^2}\sup_D\{[\left(\theta_j^{(1)} - \theta_{j_1}^{(1)}\right)^2 m_1 + \left(\theta_j^{(2)} - \theta_{j_1}^{(2)}\right)^2 m_2 + \left(\sigma_{z,j}^2 - \sigma_{z,j_1}^2\right)^2 m_3 + 2m_4|\theta_j^{(1)} - \theta_{j_1}^{(1)}||\theta_j^{(2)} - \theta_{j_1}^{(2)}|$$

$$+ 2m_5|\theta_j^{(1)} - \theta_{j_1}^{(1)}||\sigma_{z,j}^2 - \sigma_{z,j_1}^2| + 2m_6|\theta_j^{(2)} - \theta_{j_1}^{(2)}||\sigma_{z,j}^2 - \sigma_{z,j_1}^2|]X_{j-1}^2 I_{j-1}^{(1)}\}$$

$$\leq \frac{c\delta^2}{\epsilon^2},$$

where $m_1, m_2, \ldots, m_6$ are some finite moments of process $\{X_t\}$ under assumption (C2), and c is a positive constant. A similar argument can be used for $\frac{1}{\sqrt{N}}S_{N,j}^{(2)}(\boldsymbol{\theta}_j, \beta_j)$ and $\frac{1}{\sqrt{N}}S_{N,j}^{(3)}(\boldsymbol{\theta}_j, \beta_j)$, $j = 1, \ldots, T$. Let $\delta \to 0$, we can get (A3).
By the ergodic theorem, we have

$$\frac{1}{N}Q_N \xrightarrow{P} H(\boldsymbol{\theta}).$$

After some calculation, we have

$$Q_N\left(\hat{\boldsymbol{\beta}}_{MQL} - \boldsymbol{\beta}\right)$$
$$= \left(S_{N,1}^{(1)}(\hat{\boldsymbol{\theta}}_1, \beta_1), S_{N,1}^{(2)}(\hat{\boldsymbol{\theta}}_1, \beta_1), S_{N,1}^{(3)}(\hat{\boldsymbol{\theta}}_1, \beta_1), \ldots, S_{N,T}^{(1)}(\hat{\boldsymbol{\theta}}_T, \beta_T), S_{N,T}^{(2)}(\hat{\boldsymbol{\theta}}_T, \beta_T), S_{N,T}^{(3)}(\hat{\boldsymbol{\theta}}_T, \beta_T)\right)',$$

Therefore,

$$\sqrt{N}\left(\hat{\boldsymbol{\beta}}_{MQL} - \boldsymbol{\beta}\right) \xrightarrow{L} \mathcal{N}\left(\mathbf{0}, H^{-1}(\boldsymbol{\theta})\right).$$

This completes the proof. □

References

1. Al-Osh, M.A.; Alzaid, A.A. First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* **1987**, *8*, 261–275. [CrossRef]
2. Du, J.; Li, Y. The integer-valued autoregressive (INAR(p)) model. *J. Time Ser. Anal.* **1991**, *12*, 129–142.
3. Jung, R.C.; Ronning, G.; Tremayne, A.R. Estimation in conditional first order autoregression with discrete support. *Stat. Pap.* **2005**, *46*, 195–224. [CrossRef]
4. Weiß, C.H. Thinning operations for modeling time series of counts-a survey. *Asta-Adv. Stat. Anal.* **2008**, *92*, 319–341. [CrossRef]

5. Ristić, M.M.; Bakouch, H.S.; Nastić, A.S. A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *J. Stat. Plan. Infer.* **2009**, *139*, 2218–2226. [CrossRef]

6. Zhang, H.; Wang, D.; Zhu, F. Inference for INAR(p) processes with signed generalized power series thinning operator. *J. Stat. Plan. Infer.* **2010**, *140*, 667–683. [CrossRef]

7. Li, C.; Wang, D.; Zhang, H. First-order mixed integer-valued autoregressive processes with zero-inflated generalized power series innovations. *J. Korean Stat. Soc.* **2015**, *44*, 232–246. [CrossRef]

8. Kang, Y.; Wang, D.; Yang, K. A new INAR(1) process with bounded support for counts showing equidispersion, underdispersion and overdispersion. *Stat. Pap.* **2021**, *62*, 745–767. [CrossRef]

9. Yu, M.; Wang, D.; Yang, K.; Liu, Y. Bivariate first-order random coefficient integer-valued autoregressive processes. *J. Stat. Plan. Inference* **2020**, *204*, 153–176. [CrossRef]

10. Tong, H. On a threshold model. In *Pattern Recognition and Signal Processing*; Chen, C.H., Ed.; Sijthoff and Noordhoff: Amsterdam, The Netherlands, 1978; pp. 575–586.

11. Tong, H.; Lim, K.S. Threshold autoregression, limit cycles and cyclical data. *J. R. Stat. Soc. B* **1980**, *42*, 245–292. [CrossRef]

12. Monteiro, M.; Scotto, M.G.; Pereira, I. Integer-valued self-exciting threshold autoregressive processes. *Commun. Stat-Theory Methods* **2012**, *41*, 2717–2737. [CrossRef]

13. Wang, C.; Liu, H.; Yao, J.; Davis, R.A.; Li, W.K. Self-excited threshold poisson autoregression. *J. Am. Stat. Assoc.* **2014**, *109*, 777–787. [CrossRef]

14. Yang, K.; Li, H.; Wang, D. Estimation of parameters in the self-exciting threshold autoregressive processes for nonlinear time series of counts. *Appl. Math. Model.* **2018**, *57*, 226–247. [CrossRef]

15. Yang, K.; Wang, D.; Jia, B.; Li, H. An integer-valued threshold autoregressive process based on negative binomial thinning. *Stat. Pap.* **2018**, *59*, 1131–1160. [CrossRef]

16. Bennett, W.R. Statistics of regenerative digital transmission. *Bell Syst. Tech. J.* **1958**, *37*, 1501–1542. [CrossRef]

17. Gladyshev, E.G. Periodically and almost-periodically correlated random processes with a continuous time parameter. *Theory Probab. Appl.* **1963**, *8*, 173–177. [CrossRef]

18. Bentarzi, M.; Hallin, M. On the invertibility of periodic moving-average models. *J. Time Ser. Anal.* **1994**, *15*, 263–268. [CrossRef]

19. Lund, R.; Basawa, I.V. Recursive Prediction and Likelihood Evaluation for Periodic ARMA Models. *J. Time Ser. Anal.* **2000**, *21*, 75–93. [CrossRef]

20. Basawa, I.V.; Lund, R. Large sample properties of parameter estimates for periodic ARMA models. *J. Time Ser. Anal.* **2001**, *22*, 651–663. [CrossRef]

21. Shao, Q. Mixture periodic autoregressive time series models. *Stat. Probabil. Lett.* **2006**, *76*, 609–618. [CrossRef]

22. Monteiro, M.; Scotto, M.G.; Pereira, I. Integer-valued autoregressive processes with periodic structure. *J. Stat. Plan. Inference* **2010**, *140*, 1529–1541. [CrossRef]

23. Hall, A.; Scotto, M.; Cruz, J. Extremes of integer-valued moving average sequences. *Test* **2010**, *19*, 359–374. [CrossRef]

24. Santos, C.; Pereira, I.; Scotto, M.G. On the theory of periodic multivariate INAR processes. *Stat. Pap.* **2021**, *62*, 1291–1348. [CrossRef]

25. Pereira, I.; Scotto, M.G.; Nicolette, R. Integer-valued self-exciting periodic threshold autoregressive processes. In *Contributions in Statistics and Inference. Celebrating Nazaré Mendes Lopes' Birthday*; Gonçalves, E., Oliveira, P.E., Tenreiro, C., Eds.; Departamento de Matemática, Universidade de Coimbra/Mathematics Department of the University of Coimbra: Coimbra, Portugal, 2015; Volume 47, pp. 81–92.

26. Manaa, A.; Bentarzi, M. On a periodic SETINAR model. *Commun. Stat.-Simul. Comput.* **2021**. [CrossRef]

27. Li, D.; Tong, H. Nested sub-sample search algorithm for estimation of threshold models. *Stat. Sin.* **2016**, *26*, 1543–1554. [CrossRef]

28. Wedderburn, R.W.M. Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* **1974**, *61*, 439–447.

29. Azrak, R.; Mélard, G. The exact quasi-likelihood of time-dependent ARMA models. *J. Stat. Plan. Inference* **1998**, *68*, 31–45. [CrossRef]

30. Christou, V.; Fokianos, K. Quasi-likelihood inference for negative binomial time series models. *J. Time Ser. Anal.* **2014**, *35*, 55–78. [CrossRef]

31. Li, H.; Yang, K.; Wang, D. Quasi-likelihood inference for self-exciting threshold integer-valued autoregressive processes. *Comput. Stat.* **2017**, *32*, 1597–1620. [CrossRef]

32. Yang, K.; Kang, Y.; Wang, D.; Li, H.; Diao, Y. Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued autoregressive processes. *Metrika* **2019**, *82*, 863–889. [CrossRef]

33. Zheng, H.; Basawa, I.V.; Datta, S. Inference for *p*th-order random coefficient integer-valued autoregressive processes. *J. Time Ser. Anal.* **2006**, *27*, 411–440. [CrossRef]

34. Schuster, A. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terr. Magn.* **1898**, *3*, 13–41. [CrossRef]

35. Fisher, R.A. Tests of significance in harmonic analysis. *Proc. Roy. Soc. A Math. Phys.* **1929**, *125*, 54–59.

36. Freeland, R.K. Statistical Analysis of Discrete Time Series with Application to the Analysis of Workers' Compensation Claims Data. Ph.D. Thesis, Management Science Division, Faculty of Commerce and Business Administration, University of British Columbia, Vancouver, BC, Canada, 1998.

37. Möller T.A.; Silva, M.E.; Weiß, C.H.; Scotto, M.G.; Pereira, I. Self-exciting threshold binomial autoregressive processes. *Asta-Adv. Stat. Anal.* **2016**, *100*, 369–400. [CrossRef]
38. Li, H.; Yang, K.; Zhao, S.; Wang, D. First-order random coefficients integer-valued threshold autoregressive processes. *Asta-Adv. Stat. Anal.* **2018**, *102*, 305–331. [CrossRef]
39. Homburg, A.; Weiß, C.H.; Alwan, L.C.; Frahm, G.; Göb, R. Evaluating Approximate Point Forecasting of Count Processes. *Econometrics* **2019**, *7*, 30. [CrossRef]
40. Tweedie, R.L. Sufficient conditions for regularity, recurrence and ergodicity of Markov processes. *Proc. Camb. Philos. Soc.* **1975**, *78*, 125–136. [CrossRef]
41. Zheng, H.; Basawa, I.V. First-order observation-driven integer-valued autoregressive processes. *Stat. Probabil. Lett.* **2008**, *78*, 1–9. [CrossRef]
42. Billingsley, P. *Statistical Inference for Markov Processes*; The University of Chicago Press: Chicago, IL, USA, 1961.