Acta Mathematicae Applicatae Sinica, English Series
Vol. 37, No. 4 (2021) 758–772
https://doi.org/10.1007/s10255-021-1043-7
http://www.ApplMath.com.cn & www.SpringerLink.com

Acta Mathematicae Applicatae Sinica, English Series

Statistical Inference for the Covariates-driven Binomial AR(1) Process

De-hui WANG[1,2,†], **Shuai CUI**[1], **Jian-hua CHENG**[1], **Shu-hui WANG**[1]

[1]School of Mathematics, Jilin University, Changchun, 130012, China (E-mail: wangdh@jlu.edu.cn)
[2] School of Economics, Liaoning University, Shenyang 110036, China

Abstract The binomial autoregressive (BAR(1)) process is very useful to model the integer-valued time series data defined on a finite range. It is commonly observed that the autoregressive coefficient is assumed to be a constant. To make the BAR(1) model more practical, this paper introduces a new random coefficient binomial autoregressive model, which is driven by covariates. Basic probabilistic and statistical properties of this model are discussed. Conditional least squares and conditional maximum likelihood estimators of the model parameters are derived, and the asymptotic properties are obtained. The performance of these estimators is compared via a simulation study. An application to a real data example is also provided. The results show that the proposed model and methods perform well for the simulations and application.

Keywords Covariates-driven binomial autoregressive (BAR(1)) model; conditional least squares; conditional maximum likelihood

2000 MR Subject Classification 62M10; 62P20

1 Introduction

In recent years, the analysis and application of integer-valued time series data have become a popular research field. In particular, non-negative integer-valued time series has received growing attention due to the strong application value. Generally, time series of counts can be broadly classified into two categories: count data with an infinite range $\{0, 1, \cdots\}$ and count data on a finite range $\{0, 1, \cdots, n\}$. Modelling count data without upper limit has been studied intensively by researchers and many useful models have been proposed. Due to the application background, count data on a finite range has been paid more attention recently. For example, Weiß and Kim[15] considered the numbers of $n = 22$ companies in the securities business being traded (per 5 min.) at the Korea stock market; Weiß and Kim[16] considered $n = 15$ workstations and monitored (per min.) the number of occupied workstations among them.

One of the main methods to fit count data is to establish the integer-valued autoregressive model through the so-called thinning operator. The earliest thinning operator is binomial thinning operator "$\circ$" proposed by Steutal and van Harn[12], which is defined as

$$\alpha \circ X = \sum_{i=1}^{X} W_i,$$

where $\alpha \in [0, 1)$, X is a non-negative integer-valued random variable, $\{W_i\}$ is an independent

Manuscript received on December 28, 2020. Accepted on May 18, 2021.
This paper is supported by the National Natural Science Foundation of China (Nos. 11871028, 11731015, 11901053) and the Natural Science Foundation of Jilin Province (No. 20180101216JC).
†Corresponding author.

and identically distributed (i.i.d.) Bernoulli random variable sequence and satisfies $P(W_i = 1) = 1 - P(W_i = 0) = \alpha$. Besides, W_i and X are independent of each other.

The most common way for fitting count data with upper limit is the binomial autoregressive (BAR(1)) model proposed by McKenzie[10] based on binomial thinning operator. After the model was put forward, it has become a hot research topic of scholars. Weiß[13, 14] extended BAR(1) model to higher order, proposed BAR(p) model, and studied the control chart of BAR(1) model. Cui and Lund[3] studied the statistical inference of BAR(1) model. On the basis of previous studies, Weiß and Kim[15] studied the parameter estimation of BAR(1) model, and applied it to financial and industrial fields. Meanwhile, Weiß and Kim[16] considered higher-order moments and proposed a new estimation method. Kim and Weiß[7] conducted goodness of fit tests on BAR(1) model. Kim et al.[8] constructed a test for the existence of zero inflation in BAR(1) model. Yang et al.[17] used empirical likelihood method to estimate the parameters of threshold BAR(1) model. Chen et al.[2] considered the BAR(1) model with outliers. Kang et al.[5] proposed a hybrid BAR(1) model which can characterize zero inflation. Kang et al.[6] proposed a generalized BAR(1) model based on generalized binomial thinning operators.

However, the BAR(1) model still has some limitations in explaining some practical phenomena, one of the most important problems lies in that the thinning probability is non-random, which is only described by a constant parameter. In order to solve this shortcoming, some scholars took the influence of observations on the thinning probability into account, and proposed a dependent BAR(1) model. This kind of model introduces the influence of the internal factors of the system on the thinning probability, which greatly enhances the ability of BAR(1) model to explain many practical problems. However considering the influence of external factors on thinning probability may be more common, and most of the external factors are observable, for example, the number of different securities companies being traded for the stock at time t could depend on the composite stock price index and other market characteristics varying over time, in order to better fit the real data, the goal of this paper is to extend the BAR(1) model to a covariates-driven BAR(1) model.

The rest of the paper is organized as follows. In Section 2, we introduce our model and discuss its basic probabilistic and statistical properties. In Section 3, we propose two estimation methods for the model parameters. Section 4 presents some simulation results for the estimation methods. A real data example is given in Section 5.

2 The Covariates-driven BAR(1) Model

Definition 2.1. *Let $\alpha_t, \beta_t \in (0,1)$, for fixed $n \in \mathbb{N}$, the covariates-driven BAR(1) (CDBAR(1)) process $\{X_t\}$ is defined by the following recursive equation:*

$$\begin{cases} X_t = \alpha_t \circ X_{t-1} + \beta_t \circ (n - X_{t-1}), \\ \operatorname{logit}(\alpha_t) = \boldsymbol{\theta}' \boldsymbol{Z_t}, \\ \operatorname{logit}(\beta_t) = \boldsymbol{\lambda}' \boldsymbol{Z_t}, \end{cases} \tag{2.1}$$

where $\boldsymbol{Z_t} = (Z_{t1}, \cdots, Z_{tp})'$ is an observable p-dimensional covariate, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)'$ and $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_p)'$ are two p-dimensional unknown vectors. $\boldsymbol{Z_t}$ and X_{t-1} are independent.

Since the moments and conditional moments will be useful in obtaining the appropriate estimating equations for parameter estimation, we first discuss some moment properties of CDBAR(1) process.

Proposition 2.1. *Let $\{X_t\}$ be the process defined by (2.1), then we have*

$$(i)\quad E(X_t|X_{t-1}, \boldsymbol{Z_t}) = \Big(\frac{e^{\boldsymbol{\theta}' \boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\theta}' \boldsymbol{Z_t}}} - \frac{e^{\boldsymbol{\lambda}' \boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\lambda}' \boldsymbol{Z_t}}}\Big) X_{t-1} + n \frac{e^{\boldsymbol{\lambda}' \boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\lambda}' \boldsymbol{Z_t}}},$$

$$(ii)\quad \mathrm{Var}(X_t|X_{t-1},\boldsymbol{Z_t}) = \Big(\frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}} - \frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)\Big(1 - \frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}} - \frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)X_{t-1} + n\frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big(1 - \frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big),$$

$$(iii)\quad \mathrm{Corr}(X_t, X_{t-1}) = \Big(\frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}} - \frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big).$$

Proof. (i) By the definition, it is easy to obtain that

$$\begin{aligned} E(X_t|X_{t-1},\boldsymbol{Z_t}) &= E[\alpha_t \circ X_{t-1} + \beta_t \circ (n - X_{t-1})|X_{t-1},\boldsymbol{Z_t}] \\ &= E\Big[\sum_{i=1}^{X_{t-1}} W_i^{(\alpha_t)} + \sum_{j=1}^{n-X_{t-1}} W_j^{(\beta_t)}|X_{t-1},\boldsymbol{Z_t}\Big] \\ &= \frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}X_{t-1} + \frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}(n - X_{t-1}) \\ &= \Big(\frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}} - \frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)X_{t-1} + n\frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}. \end{aligned}$$

(ii) By direct calculation, we have

$$\begin{aligned} & E(X_t^2|X_{t-1},\boldsymbol{Z_t}) \\ =& E\{[\alpha_t \circ X_{t-1} + \beta_t \circ (n - X_{t-1})]^2|X_{t-1},\boldsymbol{Z_t}\} \\ =& E\Big[\Big(\sum_{i=1}^{X_{t-1}} W_i^{(\alpha_t)}\Big)^2 + \Big(\sum_{j=1}^{n-X_{t-1}} W_j^{(\beta_t)}\Big)^2 + 2\sum_{i=1}^{X_{t-1}} W_i^{(\alpha_t)} \sum_{j=1}^{n-X_{t-1}} W_j^{(\beta_t)}|X_{t-1},\boldsymbol{Z_t}\Big] \\ =& \sum_{i=1}^{X_{t-1}} E(W_i^{(\alpha_t)})^2 + X_{t-1}(X_{t-1}-1)E[(W_1^{(\alpha_t)})(W_2^{(\alpha_t)})] \\ &+ \sum_{j=1}^{n-X_{t-1}} E(W_j^{(\beta_t)})^2 + (n - X_{t-1})(n - X_{t-1} - 1)E[(W_1^{(\beta_t)})(W_2^{(\beta_t)})] \\ &+ 2X_{t-1}(n - X_{t-1})E[W_1^{(\alpha_t)})(W_1^{(\beta_t)})] \\ =& \frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}X_{t-1} + \frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}(n - X_{t-1}) \\ &+ 2\frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}\frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'Z_t}}X_{t-1}(n - X_{t-1}) + \Big(\frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}\Big)^2 X_{t-1}(X_{t-1}-1) \\ &+ \Big(\frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)^2(n - X_{t-1})(n - X_{t-1} - 1), \end{aligned}$$

from which it holds that

$$\begin{aligned} \mathrm{Var}\,(X_t|X_{t-1},\boldsymbol{Z_t}) &= E(X_t^2|X_{t-1},\boldsymbol{Z_t}) - [E(X_t|X_{t-1},\boldsymbol{Z_t})]^2 \\ &= \alpha_t X_{t-1} + \beta_t(n - X_{t-1}) + 2\alpha_t\beta_t X_{t-1}(n - X_{t-1}) + \alpha_t^2 X_{t-1}(X_{t-1}-1) \\ &\quad + \beta_t^2(n - X_{t-1})(n - X_{t-1} - 1) - [\alpha_t X_{t-1} + \beta_t(n - X_{t-1})]^2 \\ &= \alpha_t X_{t-1} + \beta_t n - \beta_t X_{t-1} + 2\alpha_t\beta_t X_{t-1}(n - X_{t-1}) + \alpha_t^2 X_{t-1}^2 - \alpha_t^2 X_{t-1} \end{aligned}$$

$$
\begin{aligned}
&+ \beta_t^2(n - X_{t-1})^2 - \beta_t^2(n - X_{t-1}) - \alpha_t^2(X_{t-1})^2 - \beta_t^2(n - X_{t-1})^2 \\
&- 2\alpha_t\beta_t X_{t-1}(n - X_{t-1}) \\
=\ & \alpha_t X_{t-1} + n\beta_t - \beta_t X_{t-1} - \alpha_t^2 X_{t-1} - \beta_t^2 n + \beta_t^2 X_{t-1} \\
=\ & (\alpha_t - \beta_t - \alpha_t^2 + \beta_t^2)X_{t-1} + n\beta_t(1 - \beta_t) \\
=\ & (\alpha_t - \beta_t)(1 - \alpha_t - \beta_t)X_{t-1} + n\beta_t(1 - \beta_t) \\
=\ & \Big(\frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}} - \frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)\Big(1 - \frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}} - \frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)X_{t-1} \\
&+ n\frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big(1 - \frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big).
\end{aligned}
$$

(iii) The covariance can be obtained as

$$
\begin{aligned}
\operatorname{Cov}(X_t, X_{t-1}) &= \operatorname{Cov}[\alpha_t \circ X_{t-1} + \beta_t \circ (n - X_{t-1}), X_{t-1}] \\
&= \operatorname{Cov}(\alpha_t \circ X_{t-1}, X_{t-1}) + \operatorname{Cov}[\beta_t \circ (n - X_{t-1}), X_{t-1}] \\
&= E[(\alpha_t \circ X_{t-1})X_{t-1}] - E(\alpha_t \circ X_{t-1})E(X_{t-1}) \\
&\quad + E\{[\beta_t \circ (n - X_{t-1})]X_{t-1}\} - E[\beta_t \circ (n - X_{t-1})]E(X_{t-1}) \\
&= E(\alpha_t X_{t-1}^2) - \alpha_t[E(X_{t-1})]^2 \\
&\quad + E[\beta_t X_{t-1}(n - X_{t-1})] - E[\beta_t(n - X_{t-1})]E(X_{t-1}) \\
&= \alpha_t E(X_{t-1}^2) - \alpha_t[E(X_{t-1})]^2 \\
&\quad + \beta_t E[X_{t-1}(n - X_{t-1})] - \beta_t E(n - X_{t-1})E(X_{t-1}) \\
&= \alpha_t E(X_{t-1}^2) - \alpha_t[E(X_{t-1})]^2 + \beta_t n E(X_{t-1}) - \beta_t E[X_{t-1}^2] \\
&\quad - \beta_t n E(X_{t-1}) + \beta_t[E(X_{t-1})]^2 \\
&= (\alpha_t - \beta_t)E(X_{t-1}^2) - (\alpha_t - \beta_t)[E(X_{t-1})]^2 \\
&= (\alpha_t - \beta_t)\operatorname{Var}(X_{t-1}) \\
&= \Big(\frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}} - \frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)\operatorname{Var}(X_{t-1}),
\end{aligned}
$$

then it follows that

$$
\operatorname{Corr}(X_t, X_{t-1}) = \frac{\operatorname{Cov}(X_t, X_{t-1})}{\operatorname{Var}(X_t)} = \Big(\frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}} - \frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big).
$$

This completes the proof. □

According to the Definition 2.1, it is easy to see that $\{X_t\}$ is a Markov chain on the state space $S = \{1, 2, \cdots, n\}$, and the transition probabilities are given by

$$
\begin{aligned}
p_{k|l} &= p(X_t = k | X_{t-1} = l, \boldsymbol{Z_t}) \\
&= \sum_{m=\max\{0,k+l-n\}}^{\min\{k,l\}} \binom{l}{m}\binom{n-l}{k-m}\Big(\frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}\Big)^m \Big(\frac{1}{1 + e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}\Big)^{l-m} \\
&\quad \times \Big(\frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)^{k-m} \Big(\frac{1}{1 + e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)^{n-l+m-k},
\end{aligned}
\tag{2.2}
$$

in which α_t and β_t can be rewritten through a linear transformation as:

$$
\alpha_t = \frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}, \qquad \beta_t = \frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1 + e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}.
$$

For the strict stationarity and ergodicity of CDBAR(1) process, we state the following result.

Proposition 2.2. *Let $\{X_t\}$ be the process defined by (2.1), then it is an irreducible, aperiodic and positive recurrent (and hence ergodic) Markov chain. Therefore, there exists a strictly stationary process satisfying (2.1).*

Proof. Note that $0 < \alpha_t < 1$ and $0 < \beta_t < 1$, it holds that

$$p_{k|l} = p(X_t = k|X_{t-1} = l, \boldsymbol{Z_t}) > 0, \qquad \text{for all} \quad i, j \in S,$$

then it is known that the Markov chain $\{X_t\}$ is irreducible. The transition probabilities (2.2) imply that $\{X_t\}$ is aperiodic. Furthermore, $\{X_t\}$ is positive recurrent because of the finite state space. Thus the conclusion follows from Theorem 4.3.3 in Ross[11]. □

3 Parameters Estimation

For the CDBAR(1) model, our primary interest lies in estimating the parameters $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)^{'}$ and $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_p)^{'}$. We mainly consider two methods, i.e., the conditional least squares (CLS) and the conditional maximum likelihood (CML) estimation.

3.1 CLS Estimators of the CDBAR(1) Model

CLS estimation is one of the most commonly used methods in integer-valued time series analysis. The CLS estimators of the CDBAR(1) model can be obtained by minimizing the conditional sum of squares function

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\lambda}) &= \sum_{t=2}^{T} [X_t - E(X_t|X_{t-1}, \boldsymbol{Z_t})]^2 \\ &= \sum_{t=2}^{T} \Big[X_t - \Big(\frac{e^{\boldsymbol{\theta}^{'}\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}^{'}\boldsymbol{Z_t}}} - \frac{e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}\Big)X_{t-1} - n\frac{e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}\Big]^2. \end{aligned}$$

Let the partial derivatives of $Q(\boldsymbol{\theta}, \boldsymbol{\lambda})$ equal to zero, we obtain for $i = 1, \cdots, p$, that

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \theta_i} = {} & -2\sum_{t=2}^{T} \Big[X_t - \Big(\frac{e^{\boldsymbol{\theta}^{'}\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}^{'}\boldsymbol{Z_t}}} - \frac{e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}\Big)X_{t-1} - n\frac{e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}\Big] \\ & \times \frac{Z_{ti}e^{\boldsymbol{\theta}^{'}\boldsymbol{Z_t}}}{(1+e^{\boldsymbol{\theta}^{'}\boldsymbol{Z_t}})^2}X_{t-1} = 0, \end{aligned}$$

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \lambda_i} = {} & -2\sum_{t=2}^{T} \Big[X_t - \Big(\frac{e^{\boldsymbol{\theta}^{'}\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}^{'}\boldsymbol{Z_t}}} - \frac{e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}\Big)X_{t-1} - n\frac{e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}\Big] \\ & \times \Big(\frac{nZ_{ti}e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}{(1+e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}})^2} - \frac{Z_{ti}e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}}}{(1+e^{\boldsymbol{\lambda}^{'}\boldsymbol{Z_t}})^2}X_{t-1}\Big) = 0, \end{aligned}$$

By solving the above equations, we can get the CLS estimators $\widehat{\boldsymbol{\theta}}_{\text{CLS}}$ and $\widehat{\boldsymbol{\lambda}}_{\text{CLS}}$.

Rewrite the parameters of the model as $\boldsymbol{\eta} = (\boldsymbol{\theta}', \boldsymbol{\lambda}')'$, and denote their true values by $\boldsymbol{\eta}_0 = (\boldsymbol{\theta}_0', \boldsymbol{\lambda}_0')'$. The following theorem states the asymptotic properties of the CLS estimators.

Theorem 3.1. *Let $\{X_t\}$ be a strictly stationary and ergodic CDBAR(1) process, then the CLS estimators $\widehat{\boldsymbol{\eta}}_{\mathrm{CLS}}$ are strongly consistent and asymptotically normal, i.e.,*

$$\sqrt{T}(\widehat{\boldsymbol{\eta}}_{\mathrm{CLS}}-\boldsymbol{\eta}_0)\xrightarrow{L} N(\mathbf{0},\boldsymbol{V}^{-1}(\boldsymbol{\eta}_0)\boldsymbol{W}(\boldsymbol{\eta}_0)\boldsymbol{V}^{-1}(\boldsymbol{\eta}_0)),\qquad T\to+\infty, \tag{3.1}$$

where

$$\boldsymbol{W}(\boldsymbol{\eta}_0)=E\left(u_t^2(\boldsymbol{\eta}_0)\frac{\partial g_t(\boldsymbol{\eta}_0)}{\partial\boldsymbol{\eta}}\frac{\partial g_t(\boldsymbol{\eta}_0)}{\partial\boldsymbol{\eta}'}\right),\quad \boldsymbol{V}(\boldsymbol{\eta}_0)=E\left(\frac{\partial g_t(\boldsymbol{\eta}_0)}{\partial\boldsymbol{\eta}}\frac{\partial g_t(\boldsymbol{\eta}_0)}{\partial\boldsymbol{\eta}'}-u_t(\boldsymbol{\eta}_0)\frac{\partial^2 g_t(\boldsymbol{\eta}_0)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\right),$$

with

$$g_t(\boldsymbol{\eta})=E(X_t|X_{t-1},\boldsymbol{Z_t}),\qquad u_t(\boldsymbol{\eta})=X_t-g_t(\boldsymbol{\eta}).$$

Proof. Note that

$$g_t(\boldsymbol{\eta})=\frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}X_{t-1}+\frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}(n-X_{t-1}),$$

we can prove (3.1) by three steps.

(i) For $1\le i,j,k\le 2p$, it is easy to check that $\frac{\partial g_t(\boldsymbol{\eta})}{\partial\eta_i}$, $\frac{\partial^2 g_t(\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j}$ and $\frac{\partial^3 g_t(\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j\partial\eta_k}$ exist and are continuous.

(ii) Denote $\varphi_1(\boldsymbol{\eta})=\boldsymbol{\theta}'\boldsymbol{Z_t}$ and $\varphi_2(\boldsymbol{\eta})=\boldsymbol{\lambda}'\boldsymbol{Z_t}$, and define

$$\begin{aligned}&H^{(0)}(X_0,\cdots,X_{t-1})=n, && H_i^{(1)}(X_0,\cdots,X_{t-1})=|Z_{ti}|X_{t-1},\\ &H_{ij}^{(2)}(X_0,\cdots,X_{t-1})=|Z_{ti}Z_{tj}|X_{t-1}, && H_{ijk}^{(3)}(X_0,\cdots,X_{t-1})=7|Z_{ti}Z_{tj}Z_{tk}|X_{t-1},\end{aligned}$$

then for any $1\le i,j,k\le p$, we can verify that

$$\begin{aligned}|g_t(\boldsymbol{\eta})|&<H^{(0)}(X_0,\cdots,X_{t-1}),\\ \left|\frac{\partial g_t(\boldsymbol{\eta})}{\partial\eta_i}\right|&=\left|X_{t-1}h'(\varphi_1)\frac{\partial\varphi_1(\boldsymbol{\eta})}{\partial\eta_i}\right|\le\left|\frac{\partial\varphi_1(\boldsymbol{\eta})}{\partial\eta_i}\right|X_{t-1}=H_i^{(1)}(X_0,\cdots,X_{t-1}),\\ \left|\frac{\partial^2 g_t(\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j}\right|&=X_{t-1}\left|h''(\varphi_1)\frac{\partial\varphi_1(\boldsymbol{\eta})}{\partial\eta_i}\frac{\partial\varphi_1(\boldsymbol{\eta})}{\partial\eta_j}+h'(\varphi_1)\frac{\partial^2\varphi_1(\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j}\right|\\ &<X_{t-1}\left(\left|\frac{\partial\varphi_1(\boldsymbol{\eta})}{\partial\eta_i}\frac{\partial\varphi_1(\boldsymbol{\eta})}{\partial\eta_j}\right|+\left|\frac{\partial^2\varphi_1(\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j}\right|\right)=H_{ij}^{(2)}(X_0,\cdots,X_{t-1}),\\ \left|\frac{\partial^3 g_t(\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j\partial\eta_k}\right|&<X_{t-1}\left(7\left|\frac{\partial\varphi_1(\boldsymbol{\eta})}{\partial\eta_i}\frac{\partial\varphi_1(\boldsymbol{\eta})}{\partial\eta_j}\frac{\partial\varphi_1(\boldsymbol{\eta})}{\partial\eta_k}\right|+\left|\frac{\partial^2\varphi_1(\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j}\frac{\partial\varphi_1(\boldsymbol{\eta})}{\partial\eta_k}\right|+\left|\frac{\partial^2\varphi_1(\boldsymbol{\eta})}{\partial\eta_i\partial\eta_k}\frac{\partial\varphi_1(\boldsymbol{\eta})}{\partial\eta_j}\right|\right)\\ &\quad+X_{t-1}\left(\left|\frac{\partial^2\varphi_1(\boldsymbol{\eta})}{\partial\eta_j\partial\eta_k}\frac{\partial\varphi_1(\boldsymbol{\eta})}{\partial\eta_i}\right|+\left|\frac{\partial^3\varphi_1(\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j\partial\eta_k}\right|\right)=H_{ijk}^{(3)}(X_0,\cdots,X_{t-1}).\end{aligned}$$

Furthermore, from the fact that

$$E(X_t^m)<+\infty,\quad\text{for any}\quad m\ge 1,\quad t\ge 1,$$

and the Hölder inequality, it follows that

$$\begin{aligned}&E\left|X_tH_{ijk}^{(3)}(X_0,\cdots,X_{t-1})\right|<+\infty,\\ &E\left|H^{(0)}(X_0,\cdots,X_{t-1})H_{ijk}^{(3)}(X_0,\cdots,X_{t-1})\right|<+\infty,\\ &E\left|H_i^{(1)}(X_0,\cdots,X_{t-1})H_{ij}^{(2)}(X_0,\cdots,X_{t-1})\right|<+\infty.\end{aligned}$$

Similarly, these results can be extended to the case of $1 \le i, j, k \le 2p$.

(iii) Applying Hölder inequality again, we have for any $1 \le i, j, k \le 2p$ that

$$E\left|u_t(\boldsymbol{\eta})\frac{\partial g_t(\boldsymbol{\eta})}{\partial \eta_i}\right| < +\infty, \qquad E\left|u_t(\boldsymbol{\eta})\frac{\partial^2 g_t(\boldsymbol{\eta})}{\partial \eta_i \partial \eta_j}\right| < +\infty,$$
$$E\left|\frac{\partial g_t(\boldsymbol{\eta})}{\partial \eta_i}\frac{\partial g_t(\boldsymbol{\eta})}{\partial \eta_j}\right| < +\infty, \quad E\left|u_t^2(\boldsymbol{\eta})\frac{\partial g_t(\boldsymbol{\eta})}{\partial \eta_i}\frac{\partial g_t(\boldsymbol{\eta})}{\partial \eta_j}\right| < +\infty.$$

By Theorem 3.1 and 3.2 in Klimko and Nelson[9], we conclude that (3.1) holds. □

3.2 CML Estimators of the CDBAR(1) Model

We now consider the conditional maximum likelihood estimation for CDBAR(1) model. For any fixed $n \ge 1$, it is easy to get the conditional log-likelihood function from model (2.1) and the conditional probability mass function of X_t as

$$l_T(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \log\Big[\prod_{t=2}^{T} p(X_t|X_{t-1}, \boldsymbol{Z_t})\Big] = \sum_{t=2}^{T}\log[p(X_t|X_{t-1}, \boldsymbol{Z_t})].$$

The CML estimators $\widehat{\boldsymbol{\theta}}_{\mathrm{CML}}$ and $\widehat{\boldsymbol{\lambda}}_{\mathrm{CML}}$ are obtained by maximizing $l_T(\boldsymbol{\theta}, \boldsymbol{\lambda})$. To this end, let the partial derivatives of $l_T(\boldsymbol{\theta}, \boldsymbol{\lambda})$ equal to zero, then the CML estimators are the solutions to the score equations that for $i = 1, \cdots, p$ are given by

$$\begin{aligned}
\frac{\partial l_T(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \theta_i} &= \sum_{t=2}^{T}\frac{1}{p(X_t = k|X_{t-1} = l, \boldsymbol{Z_t})}\Big[\sum_{m=\max\{0,k+l-n\}}^{\min\{k,l\}}\binom{l}{m}\binom{n-l}{k-m}\Big(\frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)^{k-m} \\
&\quad \Big(\frac{1}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)^{n-l+m-k}\frac{Z_{ti}e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{(1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}})^2}\Big(\frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}\Big)^{m-1}\Big(\frac{1}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}\Big)^{l-m-1} \\
&\quad \times\Big(\frac{m}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}} - \frac{(l-m)e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}\Big)\Big] = 0, \\
\frac{\partial l_T(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \lambda_i} &= \sum_{t=2}^{T}\frac{1}{p(X_t = k|X_{t-1} = l, \boldsymbol{Z_t})}\Big[\sum_{m=\max\{0,k+l-n\}}^{\min\{k,l\}}\binom{l}{m}\binom{n-l}{k-m}\Big(\frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}\Big)^{m} \\
&\quad \Big(\frac{1}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}\Big)^{l-m}\frac{Z_{ti}e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{(1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}})^2}\Big(\frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)^{k-m-1}\Big(\frac{1}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)^{n-l+m-k-1} \\
&\quad \times\Big(\frac{k-m}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}} - \frac{(n-l+m-k)e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)\Big] = 0.
\end{aligned}$$

These estimators can be easily found by using numerical method in most statistical and data analysis packages.

Let $\widehat{\boldsymbol{\eta}}_{\mathrm{CML}} = (\widehat{\boldsymbol{\theta}}'_{\mathrm{CML}}, \widehat{\boldsymbol{\lambda}}'_{\mathrm{CML}})'$ denote the CML estimators of $\widehat{\boldsymbol{\eta}}$, then the strong consistency and the asymptotic normality of $\widehat{\boldsymbol{\eta}}_{\mathrm{CML}}$ are established by the following theorem.

Theorem 3.2. *Let $\{X_t\}$ be a strictly stationary and ergodic CDBAR(1) process, then the CML estimators $\widehat{\boldsymbol{\eta}}_{\mathrm{CML}}$ are strongly consistent and asymptotically normal, i.e.,*

$$\sqrt{T}(\widehat{\boldsymbol{\eta}}_{\mathrm{CML}} - \boldsymbol{\eta}_0) \xrightarrow{L} N(\mathbf{0}, \boldsymbol{I}^{-1}(\boldsymbol{\eta}_0)), \qquad T \to +\infty, \tag{3.2}$$

where $\boldsymbol{I}(\boldsymbol{\eta}_0)$ is the Fisher information matrix.

Proof. We prove (3.2) by three steps.

(i) By (2.2), it holds that the set D of (k, l) such that $p_{k|l}(\boldsymbol{\eta}) > 0$ is independent of $\boldsymbol{\eta}$, and each $p_{k|l}(\boldsymbol{\eta})$ has continuous partial derivatives of third order with respect to $\boldsymbol{\eta}$.

(ii) From Proposition 2.2, we know that there is only one ergodic set and there are no transient states for each $\boldsymbol{\eta}$.

(iii) Letting $k = 0$ in (2.2) leads to

$$p_{k|l} = \Big(\frac{1}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}\Big)^{l}\Big(\frac{1}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\Big)^{n-l}, \qquad l = 1, 2, \cdots, n.$$

Without loss of generality, suppose that $n > 2p$, then we can get

$$\frac{\partial p_{k|l}}{\partial \eta_i} = lZ_{ti}\left(\frac{1}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}\right)^{l}\left(\frac{1}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\right)^{n-l}\left(\frac{e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}\right), \qquad 1 \le i \le p,$$

$$\frac{\partial p_{k|l}}{\partial \eta_i} = (n-l)Z_{ti}\left(\frac{1}{1+e^{\boldsymbol{\theta}'\boldsymbol{Z_t}}}\right)^{l}\left(\frac{1}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\right)^{n-l}\left(\frac{e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}{1+e^{\boldsymbol{\lambda}'\boldsymbol{Z_t}}}\right), \qquad p+1 \le i \le 2p,$$

from which it can be verified that $\big(\frac{\partial p_{0|l}}{\partial \eta_i}\big)_{l,i=1,2,\cdots,2p}$, and furthermore $\big(\frac{\partial p_{k|l}}{\partial \eta_i}\big)_{(k,l)\in D,\ i=1,2,\cdots,2p}$ has rank $2p$.

(i), (ii) and (iii) show that Condition 5.1 of Billingsley[1] holds, then Theorems 2.1 and 2.2 of Billingsley[1] guarantee that there exists a consistent CML estimator being asymptotically normally distributed. □

4 Simulation

In order to compare the performance of the proposed estimators described in the previous section, for the CDBAR(1) process

$$\begin{cases} X_t = \alpha_t \circ X_{t-1} + \beta_t \circ (n - X_{t-1}), \\ \text{logit}(\alpha_t) = \boldsymbol{\theta}'\boldsymbol{Z_t}, \\ \text{logit}(\beta_t) = \boldsymbol{\lambda}'\boldsymbol{Z_t}, \end{cases}$$

we conduct simulation studies under the following two cases.

First, we consider the one-dimensional covariate $\{Z_t\}$, and focus on the following models:

Model I: $\{Z_t\}$ is an i.i.d. sequence of random variables that follow uniform distribution $U(0, 1)$;

Model II: $\{Z_t\}$ is an i.i.d. sequence of random variables that follow normal distribution $N(0, 1)$;

Model III: $\{Z_t\}$ is an AR process as follow:

$$Z_t = \alpha \cdot Z_{t-1} + \varepsilon_t, \qquad t \ge 1,$$

where $\{\varepsilon_t\}$ is an i.i.d. sequence of normal random variables with mean 0 and variance 1. α takes 0.3, 0.5 and 0.7, denoted as AR(1)-I, AR(1)-II and AR(1)-III respectively.

For the above three models, the parameters are selected as

(a) $(\theta, \lambda) = (-1, 0.5)$; (b) $(\theta, \lambda) = (-0.5, 0.1)$;

(c) $(\theta, \lambda) = (-0.1, -0.5)$; (d) $(\theta, \lambda) = (1, -1)$.

Second, we consider the two-dimensional covariate $\{Z_t\}$, and focus on the following models:

Model IV: $\{Z_t\}$ is an i.i.d. sequence of random vectors that follow bivariate normal distribution $N_2(0, 1, 0, 1, 0.5)$. The parameters are selected as

(a) $(\theta_1, \lambda_1, \theta_2, \lambda_2)$=(0.6,0.2,0.4,0.5); (b) $(\theta_1, \lambda_1, \theta_2, \lambda_2)$=(0.2,0.6,0.5,0.4);
(c) $(\theta_1, \lambda_1, \theta_2, \lambda_2)$=(0.2,−0.1,0.4,−0.5); (d) $(\theta_1, \lambda_1, \theta_2, \lambda_2)$=(0.2,−0.1,−0.4,0.9).

To report the performance of the CLS estimators and the CML estimators described in the previous section, we conduct simulation studies using sample size $n = 100, 200, 300$ and 500.

Figure 1 shows the typical sample path, ACF plot and PACF plot of Model I with different parameter combinations for a sample size 300 (the sample paths, ACF plots and PACF plots of the other 3 models are not listed here because they are similar with Figure 1). We use mean absolute deviation error (MADE) and mean square error (MSE) to compare the effects of the two methods. Taking the parameter α for example, these two criteria are defined as

$$\text{MADE} = \frac{1}{m}\sum_{j=1}^{m}|\hat{\alpha}_j - \alpha_0|, \qquad \text{MSE} = \frac{1}{m}\sum_{j=1}^{m}(\hat{\alpha}_j - \alpha_0)^2,$$

where m is the replication times, and $\hat{\alpha}_j$ is the estimator of α at the jth replication. Here we choose $m = 1000$. The results are showed in Tables 1-7.

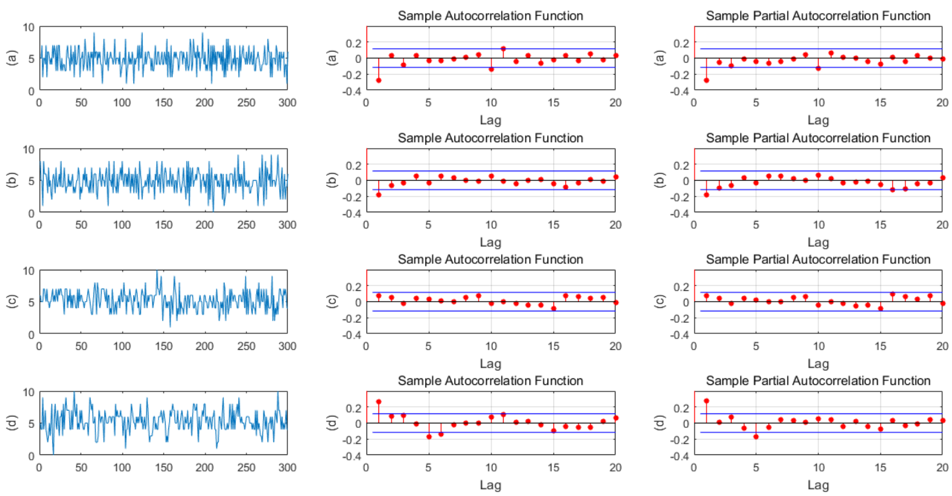

Figure 1. Sample paths, ACF and PACF plots of Models I with different parameters

As we can see that, the values of MSE and MADE gradually decrease as n increases, which implies that the estimators are consistent for all the parameters. Moreover, the CML estimators have smaller MSE and MADE, which implies that the CML estimators perform better than the CLS-estimators. Therefore, we recommend CML estimation as its main estimation method.

5 Real Data Analysis

In this section, we use the proposed model to fit a set of stock trading data, which has been analyzed by Weiß and Kim[15], to illustrate the application of CDBAR(1) process. This data set consists of the numbers of 22 securities companies being traded every 5 minutes in the Korea stock market from 9:00 to 14:50 on February 8, 2011. There are totally 70 observations, and the range of data is from 0 to 22. Figure 2 presents the original series, the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the data.

We use the proposed CDBAR(1) model and BAR(1) model to fit the data set, and compare different models via the AIC criterion and BIC criterion. For the covariates of the CDBAR (1) model, we consider the following five cases:

Table 1. Simulation results for Models I

Sample size	Parameter	CLS		CML	
		MADE	MSE	MADE	MSE
$T = 100$	$\theta = -1$	0.3458	0.1884	0.3364	0.1760
	$\lambda = 0.5$	0.2848	0.1255	0.2685	0.1158
$T = 200$	$\theta = -1$	0.2486	0.0978	0.2399	0.0944
	$\lambda = 0.5$	0.2064	0.0661	0.2034	0.0637
$T = 300$	$\theta = -1$	0.2139	0.0707	0.1999	0.0604
	$\lambda = 0.5$	0.1679	0.0470	0.1594	0.0424
$T = 500$	$\theta = -1$	0.1609	0.0440	0.1447	0.0371
	$\lambda = 0.5$	0.1259	0.0249	0.1152	0.0213
$T = 100$	$\theta = -0.5$	0.3169	0.1498	0.3195	0.1548
	$\lambda = 0.25$	0.2721	0.1173	0.2850	0.1301
$T = 200$	$\theta = -0.5$	0.2310	0.0816	0.2280	0.0790
	$\lambda = 0.25$	0.2046	0.0655	0.2001	0.0618
$T = 300$	$\theta = -0.5$	0.1799	0.0515	0.1765	0.0489
	$\lambda = 0.25$	0.1712	0.0482	0.1675	0.0456
$T = 500$	$\theta = -0.5$	0.1375	0.0303	0.1337	0.0288
	$\lambda = 0.25$	0.1263	0.0261	0.1218	0.0244
$T = 100$	$\theta = 0.5$	0.3031	0.1405	0.3123	0.1606
	$\lambda = -0.25$	0.3343	0.1734	0.3449	0.1960
$T = 200$	$\theta = 0.5$	0.1952	0.0635	0.1904	0.0614
	$\lambda = -0.25$	0.2085	0.0742	0.2072	0.0742
$T = 300$	$\theta = 0.5$	0.1646	0.0425	0.1639	0.0421
	$\lambda = -0.25$	0.1689	0.0459	0.1699	0.0466
$T = 500$	$\theta = 0.5$	0.1298	0.0261	0.1283	0.0261
	$\lambda = -0.25$	0.1350	0.0287	0.1347	0.0287
$T = 100$	$\theta = 1$	0.3313	0.1732	0.3234	0.1674
	$\lambda = -0.5$	0.3422	0.1884	0.3347	0.1842
$T = 200$	$\theta = 1$	0.2119	0.0698	0.2001	0.0623
	$\lambda = -0.5$	0.2020	0.0665	0.1906	0.0607
$T = 300$	$\theta = 1$	0.1762	0.0508	0.1682	0.0459
	$\lambda = -0.5$	0.1878	0.0553	0.1797	0.0494
$T = 500$	$\theta = 1$	0.1315	0.0274	0.1243	0.0251
	$\lambda = -0.5$	0.1341	0.0291	0.1313	0.0274

Table 2. Simulation results for Models II

Sample size	Parameter	CLS		CML	
		MADE	MSE	MADE	MSE
$T = 100$	$\theta = -1$	0.2502	0.1118	0.1910	0.0606
	$\lambda = 0.5$	0.1678	0.0494	0.1331	0.0292
$T = 200$	$\theta = -1$	0.1693	0.0482	0.1331	0.0283
	$\lambda = 0.5$	0.1190	0.0223	0.0992	0.0149
$T = 300$	$\theta = -1$	0.1319	0.0305	0.1098	0.0196
	$\lambda = 0.5$	0.0973	0.0156	0.0775	0.0099
$T = 500$	$\theta = -1$	0.1011	0.0164	0.0843	0.0108
	$\lambda = 0.5$	0.0726	0.0082	0.0576	0.0054
$T = 100$	$\theta = -0.5$	0.1941	0.0589	0.1880	0.0567
	$\lambda = 0.25$	0.1652	0.0433	0.1622	0.0443
$T = 200$	$\theta = -0.5$	0.1242	0.0246	0.1157	0.0205
	$\lambda = 0.25$	0.1066	0.0203	0.0975	0.0167
$T = 300$	$\theta = -0.5$	0.1081	0.0177	0.1000	0.0157
	$\lambda = 0.25$	0.0962	0.0149	0.0913	0.0135
$T = 500$	$\theta = -0.5$	0.0828	0.0108	0.0763	0.0095
	$\lambda = 0.25$	0.0717	0.0081	0.0690	0.0075
$T = 100$	$\theta = 0.5$	0.2010	0.0639	0.1881	0.0572
	$\lambda = -0.25$	0.1795	0.0499	0.1721	0.0490
$T = 200$	$\theta = 0.5$	0.1417	0.0303	0.1291	0.0256
	$\lambda = -0.25$	0.1151	0.0219	0.1081	0.0197
$T = 300$	$\theta = 0.5$	0.1013	0.0164	0.0969	0.0149
	$\lambda = -0.25$	0.0924	0.0124	0.0862	0.0116
$T = 500$	$\theta = 0.5$	0.0774	0.0095	0.0725	0.0080
	$\lambda = -0.25$	0.0713	0.0077	0.0624	0.0062
$T = 100$	$\theta = 1$	0.2278	0.1067	0.1921	0.0603
	$\lambda = -0.5$	0.1664	0.0467	0.1384	0.0298
$T = 200$	$\theta = 1$	0.1622	0.0476	0.1360	0.0303
	$\lambda = -0.5$	0.1184	0.0221	0.0983	0.0154
$T = 300$	$\theta = 1$	0.1238	0.0249	0.0955	0.0138
	$\lambda = -0.5$	0.0949	0.0144	0.0750	0.0089
$T = 500$	$\theta = 1$	0.0930	0.0134	0.0797	0.0093
	$\lambda = -0.5$	0.0675	0.0072	0.0570	0.0051

Table 3. Simulation results for Models III (AR(1)-I $\{Z_t\}$)

Sample size	Parameter	CLS		CML	
		MADE	MSE	MADE	MSE
$T = 100$	$\theta = -1$ $\lambda = 0.5$	0.2288 0.1606	0.0844 0.0461	0.1864 0.1298	0.0586 0.0282
$T = 200$	$\theta = -1$ $\lambda = 0.5$	0.1474 0.1042	0.0349 0.0168	0.1260 0.0884	0.0250 0.0124
$T = 300$	$\theta = -1$ $\lambda = 0.5$	0.1383 0.0981	0.0302 0.0152	0.1000 0.0725	0.0150 0.0080
$T = 500$	$\theta = -1$ $\lambda = 0.5$	0.1007 0.0659	0.0156 0.0067	0.0785 0.0525	0.0096 0.0045
$T = 100$	$\theta = -0.5$ $\lambda = 0.25$	0.1957 0.1738	0.0583 0.0452	0.1837 0.1660	0.0515 0.0421
$T = 200$	$\theta = -0.5$ $\lambda = 0.25$	0.1278 0.1130	0.0247 0.0209	0.1230 0.1086	0.0233 0.0191
$T = 300$	$\theta = -0.5$ $\lambda = 0.25$	0.1078 0.0948	0.0188 0.0145	0.0892 0.0824	0.0135 0.0104
$T = 500$	$\theta = -0.5$ $\lambda = 0.25$	0.0858 0.0700	0.0112 0.0080	0.0742 0.0624	0.0091 0.0065
$T = 100$	$\theta = 0.5$ $\lambda = -0.25$	0.1738 0.1550	0.0482 0.0394	0.1644 0.1470	0.0459 0.0369
$T = 200$	$\theta = 0.5$ $\lambda = -0.25$	0.1346 0.1159	0.0278 0.0218	0.1272 0.1158	0.0256 0.0213
$T = 300$	$\theta = 0.5$ $\lambda = -0.25$	0.0911 0.0858	0.0131 0.0117	0.0827 0.0791	0.0108 0.0100
$T = 500$	$\theta = 0.5$ $\lambda = -0.25$	0.0808 0.0747	0.0106 0.0094	0.0727 0.0702	0.0087 0.0078
$T = 100$	$\theta = 1$ $\lambda = -0.5$	0.2287 0.1622	0.0823 0.0422	0.1743 0.1252	0.0465 0.0245
$T = 200$	$\theta = 1$ $\lambda = -0.5$	0.1594 0.1136	0.0414 0.0209	0.1215 0.0920	0.0232 0.0137
$T = 300$	$\theta = 1$ $\lambda = -0.5$	0.1245 0.0799	0.0241 0.0111	0.1052 0.0693	0.0164 0.0077
$T = 500$	$\theta = 1$ $\lambda = -0.5$	0.0919 0.0683	0.0134 0.0071	0.0777 0.0570	0.0095 0.0050

Table 4. Simulation results for Models III (AR(1)-II $\{Z_t\}$)

Sample size	Parameter	CLS		CML	
		MADE	MSE	MADE	MSE
$T = 100$	$\theta = -1$ $\lambda = 0.5$	0.2126 0.0786	0.0715 0.0095	0.1160 0.0514	0.0204 0.0044
$T = 200$	$\theta = -1$ $\lambda = 0.5$	0.1307 0.0515	0.0302 0.0047	0.0854 0.0356	0.0117 0.0022
$T = 300$	$\theta = -1$ $\lambda = 0.5$	0.0956 0.0401	0.0150 0.0025	0.0724 0.0299	0.0084 0.0014
$T = 500$	$\theta = -1$ $\lambda = 0.5$	0.0767 0.0322	0.0089 0.0015	0.0526 0.0224	0.0042 0.0007
$T = 100$	$\theta = -0.5$ $\lambda = 0.25$	0.1006 0.0630	0.0178 0.0066	0.0867 0.0536	0.0131 0.0048
$T = 200$	$\theta = -0.5$ $\lambda = 0.25$	0.0818 0.0442	0.0101 0.0033	0.0674 0.0378	0.0074 0.0024
$T = 300$	$\theta = -0.5$ $\lambda = 0.25$	0.0671 0.0415	0.0070 0.0027	0.0485 0.0299	0.0037 0.0014
$T = 500$	$\theta = -0.5$ $\lambda = 0.25$	0.0485 0.0299	0.0036 0.0014	0.0392 0.0240	0.0024 0.0009
$T = 100$	$\theta = 0.5$ $\lambda = -0.25$	0.0828 0.0852	0.0110 0.0112	0.0692 0.0657	0.0075 0.0069
$T = 200$	$\theta = 0.5$ $\lambda = -0.25$	0.0582 0.0607	0.0057 0.0059	0.0468 0.0479	0.0035 0.0038
$T = 300$	$\theta = 0.5$ $\lambda = -0.25$	0.0501 0.0481	0.0038 0.0037	0.0401 0.0391	0.0025 0.0024
$T = 500$	$\theta = 0.5$ $\lambda = -0.25$	0.0360 0.0380	0.0020 0.0022	0.0268 0.0323	0.0011 0.0015
$T = 100$	$\theta = 1$ $\lambda = -0.5$	0.1241 0.1004	0.0242 0.0147	0.0808 0.0639	0.0103 0.0066
$T = 200$	$\theta = 1$ $\lambda = -0.5$	0.0789 0.0708	0.0104 0.0077	0.0585 0.0506	0.0056 0.0040
$T = 300$	$\theta = 1$ $\lambda = -0.5$	0.0633 0.0556	0.0062 0.0051	0.0454 0.0375	0.0031 0.0023
$T = 500$	$\theta = 1$ $\lambda = -0.5$	0.0510 0.0451	0.0039 0.0030	0.0351 0.0311	0.0020 0.0014

Table 5. Simulation results for Models III (AR(1)-III $\{Z_t\}$)

Sample size	Parameter	CLS		CML	
		MADE	MSE	MADE	MSE
$T = 100$	$\theta = -1$	0.1948	0.0620	0.1377	0.0342
	$\lambda = 0.5$	0.1292	0.0260	0.1040	0.0170
$T = 200$	$\theta = -1$	0.1390	0.0306	0.0952	0.0150
	$\lambda = 0.5$	0.0885	0.0129	0.0661	0.0067
$T = 300$	$\theta = -1$	0.1008	0.0165	0.0741	0.0091
	$\lambda = 0.5$	0.0690	0.0072	0.0522	0.0040
$T = 500$	$\theta = -1$	0.0785	0.0095	0.0595	0.0056
	$\lambda = 0.5$	0.0495	0.0037	0.0370	0.0021
$T = 100$	$\theta = -0.5$	0.1520	0.0403	0.1441	0.0344
	$\lambda = 0.25$	0.1297	0.0318	0.1189	0.0254
$T = 200$	$\theta = -0.5$	0.1056	0.0184	0.0934	0.0145
	$\lambda = 0.25$	0.0896	0.0129	0.0796	0.0104
$T = 300$	$\theta = -0.5$	0.0776	0.0098	0.0724	0.0083
	$\lambda = 0.25$	0.0722	0.0082	0.0665	0.0069
$T = 500$	$\theta = -0.5$	0.0629	0.0063	0.0542	0.0047
	$\lambda = 0.25$	0.0493	0.0040	0.0424	0.0031
$T = 100$	$\theta = 0.5$	0.1495	0.0377	0.1373	0.0328
	$\lambda = -0.25$	0.1236	0.0250	0.1127	0.0205
$T = 200$	$\theta = 0.5$	0.0956	0.0147	0.0910	0.0126
	$\lambda = -0.25$	0.0807	0.0108	0.0766	0.0094
$T = 300$	$\theta = 0.5$	0.0823	0.0099	0.0689	0.0073
	$\lambda = -0.25$	0.0718	0.0082	0.0613	0.0061
$T = 500$	$\theta = 0.5$	0.0675	0.0069	0.0597	0.0056
	$\lambda = -0.25$	0.0600	0.0053	0.0530	0.0043
$T = 100$	$\theta = 1$	0.1906	0.0616	0.1365	0.0299
	$\lambda = -0.5$	0.1073	0.0216	0.0923	0.0134
$T = 200$	$\theta = 1$	0.1318	0.0280	0.0938	0.0150
	$\lambda = -0.5$	0.0819	0.0118	0.0623	0.0067
$T = 300$	$\theta = 1$	0.1085	0.0188	0.0826	0.0106
	$\lambda = -0.5$	0.0648	0.0078	0.0515	0.0043
$T = 500$	$\theta = 1$	0.0848	0.0111	0.0638	0.0066
	$\lambda = -0.5$	0.0557	0.0048	0.0418	0.0027

Table 6. Simulation results for Models IV

Sample size	Parameter	CLS		CML	
		MADE	MSE	MADE	MSE
$T = 100$	$\theta_1 = 0.6$	0.1805	0.0523	0.1824	0.0532
	$\lambda_1 = 0.2$	0.2322	0.0787	0.2403	0.0853
	$\theta_2 = 0.4$	0.2398	0.0913	0.2384	0.0907
	$\lambda_2 = 0.5$	0.1708	0.0460	0.1754	0.0498
$T = 200$	$\theta_1 = 0.6$	0.1165	0.0234	0.1157	0.0233
	$\lambda_1 = 0.2$	0.2075	0.0586	0.2082	0.0589
	$\theta_2 = 0.4$	0.2246	0.0686	0.2248	0.0685
	$\lambda_2 = 0.5$	0.1220	0.0253	0.1232	0.0248
$T = 300$	$\theta_1 = 0.6$	0.0918	0.0136	0.0914	0.0137
	$\lambda_1 = 0.2$	0.2112	0.0541	0.2116	0.0542
	$\theta_2 = 0.4$	0.2028	0.0542	0.1989	0.0525
	$\lambda_2 = 0.5$	0.0980	0.0171	0.0982	0.0169
$T = 500$	$\theta_1 = 0.6$	0.0729	0.0086	0.0723	0.0082
	$\lambda_1 = 0.2$	0.2036	0.0491	0.2040	0.0491
	$\theta_2 = 0.4$	0.2006	0.0492	0.2004	0.0489
	$\lambda_2 = 0.5$	0.0775	0.0095	0.0771	0.0097
$T = 100$	$\theta_1 = 0.2$	0.1645	0.0426	0.1638	0.0411
	$\lambda_1 = 0.6$	0.1894	0.0566	0.1958	0.0628
	$\theta_2 = 0.5$	0.1874	0.0588	0.1861	0.0566
	$\lambda_2 = 0.4$	0.1618	0.0448	0.1626	0.0469
$T = 200$	$\theta_1 = 0.2$	0.1102	0.0200	0.1102	0.0207
	$\lambda_1 = 0.6$	0.1535	0.0369	0.1527	0.0369
	$\theta_2 = 0.5$	0.1383	0.0299	0.1412	0.0312
	$\lambda_2 = 0.4$	0.1143	0.0216	0.1117	0.0207
$T = 300$	$\theta_1 = 0.2$	0.0846	0.0117	0.0842	0.0115
	$\lambda_1 = 0.6$	0.1259	0.0251	0.1274	0.0254
	$\theta_2 = 0.5$	0.1256	0.0229	0.1242	0.0226
	$\lambda_2 = 0.4$	0.0904	0.0135	0.0890	0.0129
$T = 500$	$\theta_1 = 0.2$	0.0626	0.0063	0.0624	0.0063
	$\lambda_1 = 0.6$	0.1165	0.0200	0.1161	0.0196
	$\theta_2 = 0.5$	0.1105	0.0177	0.1096	0.0174
	$\lambda_2 = 0.4$	0.0683	0.0071	0.0651	0.0068

Table 7. Simulation results for Models IV (continued)

Sample size	Parameter	CLS		CML	
		MADE	MSE	MADE	MSE
$T = 100$	$\theta_1 = 0.2$	0.1989	0.0653	0.2123	0.0739
	$\lambda_1 = -0.1$	0.5165	0.3254	0.5236	0.3356
	$\theta_2 = 0.4$	0.5004	0.3217	0.5090	0.3319
	$\lambda_2 = -0.5$	0.2207	0.0831	0.2217	0.0836
$T = 200$	$\theta_1 = 0.2$	0.1468	0.0348	0.1506	0.0357
	$\lambda_1 = -0.1$	0.5205	0.3026	0.5186	0.3011
	$\theta_2 = 0.4$	0.5112	0.2986	0.5114	0.2996
	$\lambda_2 = -0.5$	0.1586	0.0414	0.1556	0.0399
$T = 300$	$\theta_1 = 0.2$	0.1085	0.0181	0.1108	0.0190
	$\lambda_1 = -0.1$	0.5049	0.2726	0.5031	0.2718
	$\theta_2 = 0.4$	0.5027	0.2748	0.5032	0.2759
	$\lambda_2 = -0.5$	0.1203	0.0221	0.1218	0.0226
$T = 500$	$\theta_1 = 0.2$	0.0809	0.0098	0.0813	0.0100
	$\lambda_1 = -0.1$	0.5047	0.2661	0.5043	0.2658
	$\theta_2 = 0.4$	0.4956	0.2583	0.4964	0.2591
	$\lambda_2 = -0.5$	0.0915	0.0130	0.0917	0.0132
$T = 100$	$\theta_1 = 0.2$	0.1856	0.0548	0.1769	0.0467
	$\lambda_1 = -0.1$	0.3012	0.1274	0.2965	0.1221
	$\theta_2 = -0.4$	0.3433	0.1782	0.3421	0.1626
	$\lambda_2 = 0.9$	0.2336	0.0920	0.2166	0.0732
$T = 200$	$\theta_1 = 0.2$	0.1206	0.0229	0.1138	0.0219
	$\lambda_1 = -0.1$	0.3091	0.1134	0.3043	0.1091
	$\theta_2 = -0.4$	0.3321	0.1368	0.3360	0.1365
	$\lambda_2 = 0.9$	0.1544	0.0361	0.1490	0.0344
$T = 300$	$\theta_1 = 0.2$	0.0967	0.0150	0.0906	0.0134
	$\lambda_1 = -0.1$	0.3011	0.1068	0.2995	0.1040
	$\theta_2 = -0.4$	0.2992	0.1055	0.3016	0.1043
	$\lambda_2 = 0.9$	0.1413	0.0296	0.1321	0.0261
$T = 500$	$\theta_1 = 0.2$	0.0762	0.0094	0.0747	0.0088
	$\lambda_1 = -0.1$	0.3014	0.0985	0.3005	0.0967
	$\theta_2 = -0.4$	0.3112	0.1103	0.3104	0.1085
	$\lambda_2 = 0.9$	0.1026	0.0163	0.0922	0.0128

Table 8. Analysis results of the stock trading data

Model	CML	AIC	BIC
BAR(1)	$\widehat{p} = 0.428$	283.8	288.3
	$\widehat{\rho} = 0.510$		
CDBAR(1)-I	$\widehat{\theta}_1 = 0.454$	283.7	288.2
	$\widehat{\lambda}_1 = -0.639$		
CDBAR(1)-II	$\widehat{\theta}_1 = 1.969$	277.7	282.2
	$\widehat{\lambda}_1 = -2.831$		
CDBAR(1)-III	$\widehat{\theta}_1 = 0.556$	324.0	328.5
	$\widehat{\lambda}_1 = -0.307$		
CDBAR(1)-IV	$\widehat{\theta}_1 = 0.517$	279.0	283.5
	$\widehat{\lambda}_1 = -0.762$		
CDBAR(1)-V	$\widehat{\theta}_1 = -0.525$	327.1	336.1
	$\widehat{\lambda}_1 = -0.173$		
	$\widehat{\theta}_2 = 0.545$		
	$\widehat{\lambda}_2 = 0.050$		

(a) CDBAR(1)-I model: In KRX (the Korea Exchange), KOSPI (Korean Composite Stock Price Index) is the flagship market. As a result, we consider KOSPI on February 8, 2011 as the covariate. However, we only know the open value (2090.47), the close value (2069.70), the low value (2066.71) and the high value (2092.43), so we fix KOSPI at 9:00 to be the open value, let KOSPI at 14.50 be the close value, then select randomly two different points from the time per 5 minutes between 9:00 and 14:50, and assign them the low value and the high value, respectively. The other KOSPI are generated from the uniform distribution $U(2066.71, 2092.43)$.

(b) CDBAR(1)-II model: $\{Z_t\}$ is an i.i.d. sequence of random variables that follow uniform distribution $U(0, 1)$;

(c) CDBAR(1)-III model: $\{Z_t\}$ is an i.i.d. sequence of random variables that follow normal distribution $N(0, 1)$;

(d) CDBAR(1)-IV model: $\{Z_t\}$ is an AR(1) process:

$$Z_t = 0.5 \cdot Z_{t-1} + \varepsilon_t, \qquad t \geq 1;$$

(e) CDBAR(1)-V model: $\boldsymbol{Z}_t \sim N_2(0, 1, 0, 1, 0.5)$.

For each model, we use the CML method to estimate the parameters, and the AIC and BIC values are summarized in Table 8. As can be seen that, CDBAR(1)-I only improves the classical BAR(1) model slightly, one of the reasons could be that the KOSPI values of the whole day are very close to each other. On the other hand, CDBAR(1)-II and CDBAR(1)-IV have the smaller AIC value and BIC value, from which we can conclude these two models are competitive for this data set, showing that it is necessary to extend the classical BAR(1) model, because some appropriate covariates may be very useful.

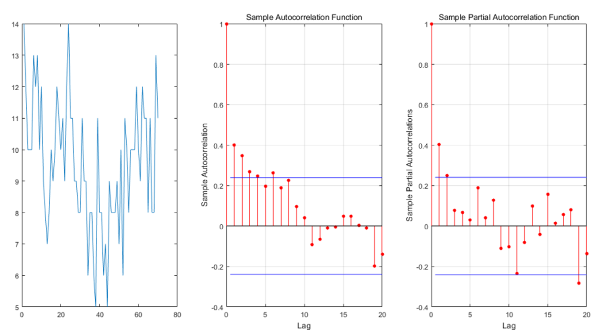

Figure 2. Sample path, the ACF plot and PACF plot of the real data

6 Summary and Conclusion

In this paper, we first extend the classical BAR(1) model, and introduce a covariates-driven BAR(1) (CDBAR(1)) process for count data of finite range by using logistic function. Next, the probabilistic and statistical properties of the model are discussed, such as the conditional expectation, conditional variance and correlation coefficient. CLS estimators and CML estimators of the model parameters are derived, as well as their asymptotic results are obtained. Then, we compare the performance of the CLS estimators and CML estimators by a simulation study. It is found that these two estimators are satisfactory and the CML estimators perform better than the CLS estimators. Finally, a real data example reveals that the CDBAR(1) model is appropriate for the stock trading data.

References

[1] Billingsley, P. Statistical Inference for Markov Processes. University of Chicago Press, Chicago, 1961

[2] Chen, H.P., Li, Q., Zhu, F.K. Binomial AR(1) processes with innovational outliers. *Communications in Statistics-Theory and Methods*, 50: 446–472 (2021)

[3] Cui, Y., Lund, R. Inference in binomial AR(1) models. *Statistics and Probability Letters*, 80: 1985–1990 (2010)

[4] Daley, D., Gani, J. Epidemic Modelling: an Introduction, Cambridge Studies in Mathematical Biology. Cambridge University Press, UK, 1999

[5] Kang, Y., Wang, D.H., Yang, K. A new INAR(1) process with bounded support for counts showing equidispersion, underdispersion and overdispersion. *Statistical Papers*, 62: 745–767 (2021)

[6] Kang, Y, Wang, D.H., Yang, K. Extended binomial AR(1) processes with generalized binomial thinning operator. *Communications in Statistics - Theory and Methods*, 49, 3498–3520 (2020)

[7] Kim, H.Y., Weiß C.H. Goodness-of-fit tests for binomial AR(1) processes. *Statistics*, 2015, 49: 291–315

[8] Kim, H.Y., Weiß, C.H., Möller, T.A. Testing for an excessive number of zeros in time series of bounded counts. *Statistical Methods and Applications*, 27: 689–714 (2018)

[9] Klimko, L.A., Nelson, P.I. On conditional least squares estimation for stochastic processes. *The Annals of Statistics*, 6: 629–642 (1978)

[10] McKenzie, E. Some simple models for discrete variate time series. *Water Resources Bulletin*, 21: 645–650 (1985)

[11] Ross, S.M. Stochastic processes. John Wiley & Sons, New York, 1996

[12] Steutel, F.W., Van Harn, K. Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 7: 893–899 (1979)

[13] Weiß, C.H. A New Class of Autoregressive Models for Time Series of Binomial Counts. *Communications in Statistics-Theory and Methods*, 38: 447–460 (2009)

[14] Weiß, C.H. Jumps in binomial AR(1) processes. *Statistics and Probability Letters*, 79: 2012–2019 (2009)

[15] Weiß, C.H., Kim, H.Y. Parameter estimation for binomial AR(1) models with applications in finance and industry. *Statistical Papers*, 54: 563–590 (2013)

[16] Weiß, C.H., Kim, H.Y. Binomial AR(1) processes: moments, cumulants, and estimation. *Statistics*, 47: 494–510 (2013)

[17] Yang, K., Wang, D.H., Li, H. Threshold autoregression analysis for finite-range time series of counts with an application on measles data. *Journal of Statistical Computation and Simulation*, 88: 597–614 (2018)