Statistical Papers (2024) 65:5873–5901
https://doi.org/10.1007/s00362-024-01605-6

REGULAR ARTICLE

A new integer-valued threshold autoregressive process based on modified negative binomial operator driven by explanatory variables

Yixuan Fan[1] · Jianhua Cheng[1] · Dehui Wang[2]

Received: 29 November 2023 / Revised: 2 May 2024 / Published online: 1 November 2024

Abstract
In this article, a new random coefficient integer-valued self-exciting threshold autoregressive process based on modified negative binomial operator is introduced. The autoregressive coefficients are driven by the explanatory variables via a logistic regression structure. Basic probabilistic and statistical properties of this process are discussed. Estimators of the unknown parameters are obtained via the conditional least squares and conditional maximum likelihood methods, as well as the asymptotic properties. The nonlinearity test of the proposed model and the existence test of explanatory variables are performed using the Wald-type test. Monte Carlo simulations are provided to illustrate the finite-sample performance of the estimators and the hypothesis tests. A real example is applied to illustrate the superiority of the proposed model.

Keywords Integer-valued time series · Threshold autoregressive model · Random coefficient · Logistic regression · Explanatory variables

1 Introduction

Integer-valued time series have attracted great attention in the recent years. The literature covers a broad range of studies from the number of crime incidents (Yang et al. (2018)) to the occurrence of certain diseases in epidemiology (Chen et al. 2019). Following the well-known first-order integer-valued autoregressive (INAR(1)) process introduced by (Al-Osh and Alzaid (1987)), the INAR model based on a binomial thinning operator has been applied in various research fields, such as meteorology (Chen and Khamthong (2020)), seismology (Wang et al. (2014)) and economics (Liu et al. (2019)). Further summary about the INAR(1) process based on binomial thinning operator can be found in Davis et al. (2021). However, the classical INAR(1)

✉ Jianhua Cheng
chengjh@jlu.edu.cn

1 School of Mathematics, Jilin University, Changchun 130012, China

2 School of Mathematics and Statistics, Liaoning University, Shenyang 110036, China

process based on the binomial thinning operator is not always suitable for the analysis of integer-valued time series, as the Bernoulli counting series only take values of zero or one. To cope with this problem, Ristić et al. (2009) first introduced the negative binomial operator for the INAR(1) process, where the counting series follow geometric distribution. Thereafter, integer-valued autoregressive models based on negative binomial operator have been widely used in fitting the time series of counts. Zhang et al. (2020) proposed an extended negative binomial operator " $\diamond$ " to introduce new members for the INAR-type model, where $\alpha \diamond 0 \neq 0$ for all values of α. This operator can be defined as

$$\alpha \diamond X = \sum_{j=1}^{X+1} G_j, \tag{1}$$

where X is an integer-valued random variable, $\{G_j\}$ is a sequence of independent and identically distributed geometric random variables with parameter $\frac{\alpha}{1+\alpha}$, independently of X. Based on this operator, Aleksić and Ristić (2021) constructed a new minification INAR model to cope with the problem of zero constantly behaviour over time in the existing minification model, and renamed it as the modified negative binomial operator. To model count data with a finite range, Kang et al. (2020) introduced an extended binomial AR(1) model based on the generalized binomial thinning operator. After that, Chen et al. (2023) proposed the Conway-Maxwell-Poisson-Binomial thinning operator and defined a new binomial AR(1) model based on it for more complex bounded time series of counts.

A great number of studies ascertain the important role that the threshold variable has in integer-valued time series model based on the binomial thinning operator. As Liu et al. (2020) pointed out, the piecewise structure in the threshold autoregressive model makes the nonlinearity-related properties realizable. Related works include Monteiro et al. (2012); Möller et al. (2016) and Li et al. (2018). To remedy the drawback of Bernoulli counting series in the integer-valued threshold autoregressive models, Yang et al. (2018) considered the self-exciting threshold INAR(1) model based on negative binomial operator (NBTINAR(1)) with a unknown threshold variable. To better describe the characteristic such as overdispersion or structural change for the time series of counts, Wang et al. (2021) redefined this NBTINAR(1) process under a weaker condition that the expectation of the innovations is finite. Zhang et al. (2023) introduced a two-threshold-variable INAR model for fitting the stock dataset. Furthermore, Yang et al. (2023) proposed a bivariate threshold Poisson INAR(1) process to model the bivariate time series of counts which exhibit piecewise phenomena. After that, Yang et al. (2023) considered a class of multivariate threshold INAR(1) processes driven by explanatory variables to capture the correlations in the multivariate count time series.

To handle the effect of exogenous variables on the observations, Zheng and Basawa (2008) proposed an observation-driven INAR(1) model, where the autoregressive parameter was a random sequence depending on the past observation. Ding and Wang (2016) incorporated explanatory variables into the autoregressive parameter for INAR(1) model. Yu et al. (2019) proposed a class of observation-driven random coefficient integer-valued autoregressive processes based on negative binomial operator, where the autoregressive parameter is a measurable function of X_{t-1}. Yang et al. (2021) developed a new integer-valued self-exciting threshold autoregressive process with the random coefficients driven by explanatory variables (RCTINAR(1)-X). Qian and Zhu (2022) proposed a new minification INAR(1) process driven by explanatory variables to make it more practical and flexible. To capture higher-order autocorrelation structure and introduce the driving effect of covariates for time series of counts, Li et al. (2023) proposed a pth-order random coefficient mixed binomial autoregressive process with explanatory variables. Chen et al. (2023) proposed a covariate-driven beta-binomial INGARCH model to illustrate extra-binomial variation and high volatility for bounded counts. Zhang and Wang (2023) considered a new binomial autoregressive process with explanatory variables for modelling the dataset with changing dynamic behaviours.

To capture the impact of covariates for the observed data and nonlinear responses, we combine the integer-valued self-exciting autoregressive process based on the modified negative binomial operator with the threshold nonlinear form. The motivation of this study arises from the need to make the integer-valued time series model more flexible and practical by introducing the randomness of the autoregressive coefficients. Furthermore, the modified negative binomial operator overcomes the problem that the Bernoulli counting series only take values on 0 and 1. Therefore, to better describe the characteristic for time series of counts, we consider the integer-valued threshold autoregressive process based on the modified negative binomial operator. Actually, our proposed model is characterized not only by the switching mechanism of the random coefficients between two regimes, but also the specification of random coefficients through a logistic regression driven by explanatory variables.

This paper is organized as follows. In Sect. 2, the random coefficient self-exciting threshold INAR(1) process based on modified negative binomial operator driven by explanatory variables (RCTMNBINAR(1)-X) is proposed and the basic properties are derived. In Sect. 3, the parameter estimation problem is considered. In Sect. 4, two hypothesis testing problems about the proposed model are discussed. In Sect. 5, forecasting problem for the RCTMNBINAR(1)-X process is addressed. In Sect. 6, Monte Carlo simulations about the estimation procedure and the test problems are presented. In Sect. 7, a real example is provided to illustrate the superiority of the proposed model. In Sect. 8, some possible extensions of the RCTMNBINAR(1)-X model are discussed. Section 9 concludes.

2 The RCTMNBINAR(1)-X process

In this section, we propose a new random coefficient integer-valued threshold time series model based on modified negative binomial operator driven by explanatory variables, called RCTMNBINAR(1)-X process. This process is defined by the following recursive equation:

$$X_t = (\alpha_{1,t} \diamond X_{t-1}) I_{1,t}(r) + (\alpha_{2,t} \diamond X_{t-1}) I_{2,t}(r) + Z_t, \ t \in \mathbb{N}_0, \tag{2}$$

where

(i) $I_{1,t}(r) = I\{0 \leqslant X_{t-1} \leqslant r\}$, $I_{2,t}(r) = I\{r < X_{t-1} < \infty\}$, where $r \in [\underline{r}, \bar{r}]$ is the threshold variable, $\underline{r}$ and $\bar{r}$ are known lower and upper bounds of r.

(ii) For $i = 1, 2$, $\alpha_{i,t}$ is the autoregressive parameter with the following covariates-driven structure

$$\log\left(\frac{\alpha_{i,t}}{1-\alpha_{i,t}}\right) = \omega_i + \boldsymbol{W}_t^T \boldsymbol{\beta}_i, \tag{3}$$

where $\boldsymbol{\beta}_i = (\beta_{i,1}, \cdots, \beta_{i,q})^T$ are regression coefficients, $\{\boldsymbol{W}_t := (W_{1,t}, \cdots, W_{q,t})^T\}$ is a q-dimensional explanatory variable which can be observed, ω_i is a constant.

(iii) For fixed $\alpha_{i,t}$, the modified negative binomial operator " $\diamond$ " is defined in (1).

(iv) $\{Z_t\}$ is a sequence of independent and identically distributed random variables.

(v) For fixed t and i ($i = 1, 2$), Z_t is assumed to be independent of $\boldsymbol{W}_t$, counting series in $\{\alpha_{i,t} \diamond X_{t-1}\}$ and X_{t-s} for $\forall s \geqslant 1$.

It can be seen that we extend the integer-valued self-exciting threshold process based on modified negative binomial operator (SETMNBINAR(2,1)), i.e. $X_t = (\alpha_1 \diamond X_{t-1})I_{1,t}(r) + (\alpha_2 \diamond X_{t-1})I_{2,t}(r) + Z_t$, via making the autoregressive coefficients vary with time. By allowing the autoregressive coefficients to depend on explanatory variables $\boldsymbol{W}_t$, we make the process more applicable. Here the upper and lower bounds of r ensure the rigor of the model and do not affect the parameter estimation or the hypothesis testing of model. Moreover, it is reasonable to assume that $(\omega_1, \boldsymbol{\beta}_1^T) \neq (\omega_2, \boldsymbol{\beta}_2^T)$ throughout this paper. Otherwise, the existence of the threshold variable r will make no sense. When $(\omega_1, \boldsymbol{\beta}_1^T) = (\omega_2, \boldsymbol{\beta}_2^T)$, the regime switching mechanism will no longer exist. The reduced model is a random coefficient INAR(1) process based on modified negative binomial operator with explanatory variables. In addition, if $\beta_{i,j} = 0$ for $i = 1, 2$ and $j = 1, 2, \cdots, q$, the model (2) will degenerate to the integer-valued self-exciting threshold process based on modified negative binomial operator. In the following, we will omit (r) in $I_{i,t}(r)$ $(i = 1, 2)$ to make the notations easy without ambiguity.

Notice that (3) implies that

$$\alpha_{i,t} = \frac{\exp(\omega_i + \boldsymbol{W}_t^T \boldsymbol{\beta}_i)}{1 + \exp(\omega_i + \boldsymbol{W}_t^T \boldsymbol{\beta}_i)}, \quad i = 1, 2. \tag{4}$$

Zhang et al. (2020) indicated that the bivariate INAR (1) process based on modified negative binomial operator is a Markov chain. Similar to Zhang et al. (2020), we state that model (2) is an ergodic Markov chain on $\mathbb{N}_0$ with the transition probabilities $P(X_t = x_t | X_{t-1} = x_{t-1}, \boldsymbol{W}_t)$ as follows. Throughout this paper, we consider a particular case, i.e. the innovation sequence $\{Z_t\}$ follows the geometric distribution Geom(μ).

Proposition 2.1 *Let $\{X_t\}_{t\in\mathbb{N}_0}$ be the process defined in (2) with geometric distributed innovations, then $\{X_t\}_{t\in\mathbb{N}_0}$ is a stationary and ergodic Markov chain with the following transition probabilities*

$$P(X_t = x_t | X_{t-1} = x_{t-1}, \boldsymbol{W}_t) = p(x_{t-1}, x_t, \alpha_{1,t} I_{1,t} + \alpha_{2,t} I_{2,t}, \mu), \tag{5}$$

where

$$p(x_{t-1}, x_t, \alpha_{i,t}, \mu) = \sum_{k=0}^{x_t} \binom{x_{t-1}+k}{k} \frac{\alpha_{i,t}^k \mu^{x_t-k}}{(1+\alpha_{i,t})^{x_{t-1}+k+1}(1+\mu)^{x_t-k+1}}, \quad i = 1, 2. \tag{6}$$

According to the previous discussion, if $(\omega_1, \boldsymbol{\beta}_1^T) = (\omega_2, \boldsymbol{\beta}_2^T)$, the second regime of model (2) will be removed and (5) becomes the transition probabilities of the SETMNBINAR(2,1) process. Furthermore, we can derive the conditional expectation and the conditional variance in the following proposition.

Proposition 2.2 *Let* $\{X_t\}_{t\in\mathbb{N}_0}$ *be the process defined in (2) with geometric distributed innovations, then for* $t \geqslant 1$,

(i) $E(X_t|X_{t-1}, \boldsymbol{W}_t) = \sum_{i=1}^{2} \left[\frac{\exp(\omega_i + \boldsymbol{W}_t^T \boldsymbol{\beta}_i)}{1+\exp(\omega_i + \boldsymbol{W}_t^T \boldsymbol{\beta}_i)} (X_{t-1}+1) I_{i,t} \right] + \mu,$

(ii) $Var(X_t|X_{t-1}, \boldsymbol{W}_t) = \sum_{i=1}^{2} \frac{\exp(\omega_i + \boldsymbol{W}_t^T \boldsymbol{\beta}_i)}{1+\exp(\omega_i + \boldsymbol{W}_t^T \boldsymbol{\beta}_i)} \big(1 + \frac{\exp(\omega_i + \boldsymbol{W}_t^T \boldsymbol{\beta}_i)}{1+\exp(\omega_i + \boldsymbol{W}_t^T \boldsymbol{\beta}_i)}\big)(X_{t-1} + 1) I_{i,t} + \mu(1+\mu).$

3 Parameter estimation

In this section, suppose that $X_0, X_1, \cdots, X_n$ is a series of observations generated from the RCTMNBINAR(1)-X process with the true parameter vector $\boldsymbol{\theta}_0$. We discuss the estimation problem for the unknown parameter $\boldsymbol{\theta} = (\omega_1, \omega_2, \boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \mu)^T$ using the conditional least squares (CLS) and conditional maximum likelihood (CML) methods, as well as their asymptotic properties. We first consider the threshold variable r is known, and then consider the estimation of r.

3.1 Conditional least squares estimation

Let $g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t) = E(X_t|X_{t-1}, \boldsymbol{W}_t)$, then the conditional sum of squares is given by

$$S(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left\{ X_t - \sum_{i=1}^{2} \left[\frac{\exp(\omega_i + \boldsymbol{W}_t^T \boldsymbol{\beta}_i)}{1+\exp(\omega_i + \boldsymbol{W}_t^T \boldsymbol{\beta}_i)} (X_{t-1}+1) \right] I_{i,t} - \mu \right\}^2. \tag{7}$$

The CLS-estimator $\hat{\boldsymbol{\theta}}_{CLS} = (\hat{\omega}_{1,CLS}, \hat{\omega}_{2,CLS}, \hat{\boldsymbol{\beta}}_{1,CLS}^T, \hat{\boldsymbol{\beta}}_{2,CLS}^T, \hat{\mu}_{CLS})^T$ is obtained by minimizing the CLS criterion function as follows,

$$\hat{\boldsymbol{\theta}}_{CLS} = \arg\min_{\boldsymbol{\theta}} S(\boldsymbol{\theta}). \tag{8}$$

Let $S_t(\boldsymbol{\theta}) = (X_t - g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t))^2$ and

$$-\frac{1}{2}\partial S_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \boldsymbol{M}_t(\boldsymbol{\theta}).$$

Then the CLS-estimator $\hat{\boldsymbol{\theta}}_{CLS}$ is a solution of

$$\frac{\partial S(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \boldsymbol{0} \text{ or } \sum_{t=1}^{n}\boldsymbol{M}_t(\boldsymbol{\theta}) = \boldsymbol{0}.$$

The following theorem states the strong consistency and the asymptotic normality of the CLS-estimator $\hat{\boldsymbol{\theta}}_{CLS}$.

Theorem 3.1 *Let $\{X_t\}_{t\in\mathbb{N}_0}$ be the RCTMNBINAR(1)-X process defined in (2). Assume that $\sup_t E(X_t^4) < \infty$, the CLS-estimator $\hat{\boldsymbol{\theta}}_{CLS}$ is strongly consistent and asymptotically normal, i.e.*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}_0) \xrightarrow{L} N(\boldsymbol{0}, \boldsymbol{V}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{W}(\boldsymbol{\theta}_0)\boldsymbol{V}^{-1}(\boldsymbol{\theta}_0)),$$

where $\xrightarrow{L}$ denotes convergence in distribution, $\boldsymbol{W}(\boldsymbol{\theta}_0)=E[\boldsymbol{M}_t(\boldsymbol{\theta}_0)\boldsymbol{M}_t(\boldsymbol{\theta}_0)^T]$, $u_t(\boldsymbol{\theta}_0) = X_t - g_t(\boldsymbol{\theta}_0, X_{t-1}, \boldsymbol{W}_t)$, and

$$\boldsymbol{V}(\boldsymbol{\theta}_0)=E\left(u_t^2(\boldsymbol{\theta}_0)\frac{\partial g_t(\boldsymbol{\theta}_0, X_{t-1}, \boldsymbol{W}_t)}{\partial\boldsymbol{\theta}}\frac{\partial g_t(\boldsymbol{\theta}_0, X_{t-1}, \boldsymbol{W}_t)}{\partial\boldsymbol{\theta}^T}\right).$$

The proof of this theorem is postponed to Appendix.

3.2 Conditional maximum likelihood estimation

In this subsection, we focus on the CML estimation for the parameter vector $\boldsymbol{\theta}$. The CML-estimator $\hat{\boldsymbol{\theta}}_{CML}$ can be obtained by directly maximizing the conditional log-likelihood function

$$\log L(\boldsymbol{\theta}) = \sum_{t=1}^{n}\log P(X_t = x_t|X_{t-1} = x_{t-1}, \boldsymbol{W}_t), \tag{9}$$

where $P(X_t = x_t|X_{t-1} = x_{t-1}, \boldsymbol{W}_t)$ denotes the transition probability given in (5). Then the CML-estimator $\hat{\boldsymbol{\theta}}_{CML} = (\hat{\omega}_{1,CML}, \hat{\omega}_{2,CML}, \hat{\boldsymbol{\beta}}^T_{1,CML}, \hat{\boldsymbol{\beta}}^T_{2,CML}, \hat{\mu}_{CML})^T$ is obtained by maximizing the conditional log-likelihood function

$$\hat{\boldsymbol{\theta}}_{CML} = \arg\max_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}). \tag{10}$$

Note that no closed-form expressions can be found for the CLS- and CML-estimator. Thus, we need to employ numerical procedure. In this paper, we optimize the functions in (8) and (10) by using optim() in $\mathcal{R}$ software.

The following theorem establish the strong consistency and the asymptotic normality of the CML-estimator $\hat{\boldsymbol{\theta}}_{CML}$.

Theorem 3.2 Let $\{X_t\}_{t\in\mathbb{N}_0}$ be the RCTMNBINAR(1)-X process defined in (2), then the CML-estimator $\hat{\boldsymbol{\theta}}_{CML}$ is strongly consistent and asymptotically normal, i.e.

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{CML} - \boldsymbol{\theta}_0) \xrightarrow{L} N(\boldsymbol{0},\ \boldsymbol{I}^{-1}(\boldsymbol{\theta}_0)),$$

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

The proof of this theorem is provided in the Appendix.

3.3 Unknown *r* cases

For the RCTMNBINAR(1)-X model, the unknown threshold parameter r can be easily estimated by using the two-step maximum likelihood method proposed in Wang et al. (2014). The idea of this algorithm is to estimate the threshold by searching the integer in the interval $[\underline{r}, \bar{r}]$, which minimizes the conditional log-likelihood function in (9). In practice, the algorithm can be done in the following two steps:

Step 1. For each $r \in [\underline{r}, \bar{r}] \cap \mathbb{N}$, calculate $\hat{\boldsymbol{\theta}}^{(r)}$ such that $\hat{\boldsymbol{\theta}}^{(r)} = \arg\max_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta})$ according to (9).

Step 2. The threshold r is estimated by searching over all candidates,

$$\hat{r} = \arg \max_{r\in[\underline{r},\bar{r}]\cap\mathbb{N}} \log L(\hat{\boldsymbol{\theta}}^{(r)}),$$

where we can use the lower and upper δth-quantile of the observations as the minimum and maximum values of the threshold variable.

4 Testing problems of the RCTMNBINAR(1)-X model

Given the observations $X_0, X_1, \cdots, X_n$, we are concerned about the nonlinearity for the threshold model and the existence of explanatory variables for our proposed model.

4.1 Testing nonlinearity structure for the threshold model

We first focus on the issue of testing the threshold effects for the RCTMNBINAR(1)-X model. Denote $\boldsymbol{\theta}_i = (\omega_i, \boldsymbol{\beta}_i^T)^T,\ i = 1, 2$. By definition of the RCTMNBINAR(1)-X process, the null hypothesis and the alternative hypothesis are:

$$\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \quad vs \quad \mathcal{H}_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2. \tag{11}$$

With the idea of the Wald-type test proposed by Yang et al. (2023), we assume that the following regular conditions are satisfied:

(A1) Let $\boldsymbol{\xi} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ be the parameter vector of interest, $\hat{\boldsymbol{\xi}} = (\hat{\boldsymbol{\theta}}_1^T, \hat{\boldsymbol{\theta}}_2^T)^T$ is a consistent estimator of the true value $\boldsymbol{\xi}_0$.

(A2) $\hat{\boldsymbol{\xi}}$ is asymptotically normal distributed around $\boldsymbol{\xi}_0$, i.e.

$$\sqrt{n}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) \xrightarrow{L} N(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

for some covariance matrix $\boldsymbol{\Sigma}$.

Theorem 4.1 *Under conditions (A1)-(A2), the test statistic for testing problem (11) is*

$$u_n = n(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2)^T (\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^T)^{-1}(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2), \tag{12}$$

where $\boldsymbol{C} = (\boldsymbol{I}_{(q+1)\times(q+1)}, -\boldsymbol{I}_{(q+1)\times(q+1)})$, $\boldsymbol{I}_{(q+1)\times(q+1)}$ *is an identity matrix of order* $(q+1)$, $\hat{\boldsymbol{\Sigma}}$ *is the consistent estimator of* $\boldsymbol{\Sigma}$. *Then under* $\mathcal{H}_0$,

$$u_n \xrightarrow{L} \chi^2_{q+1}, \ as\ n \to \infty.$$

The proof of Theorem 4.1 is omitted since it follows easily by the properties of the normal distribution and the Slutsky's Theorem. We can use Theorem 4.1 to test the nonlinearity for our proposed model. In practice, the estimator $\hat{\boldsymbol{\theta}}$ can be any consistent estimator of the unknown parameter $\boldsymbol{\theta}$. We use the CML-estimator obtained in Sect. 3.2 for this study.

4.2 Testing the existence of explanatory variables

Another interesting problem is to test whether the explanatory variables exist. If the explanatory variables do not exist, our proposed model (2) will degenerate into the integer-valued self-exciting threshold process based on modified negative binomial operator. Inspired by the success of testing the constancy for time series models, see (Han and McCabe (2013); Zhao et al. (2013); Awale et al. (2019)) among others, we consider the following hypothesis test

$$\mathcal{H}_0 : \beta_{i,j}=0,\ i=1,2,\ j=1,\cdots,q\ vs\ \mathcal{H}_1 : \text{at least one } \beta_{i,l} \neq 0,\ i=1,2,\ 1\leqslant l \leqslant q. \tag{13}$$

Testing problem (13) is equivalent to the following hypothesis:

$$\mathcal{H}_0 : \boldsymbol{D\xi} = \boldsymbol{0}\ vs\ \mathcal{H}_1 : \boldsymbol{D\xi} \neq \boldsymbol{0}, \tag{14}$$

where $\boldsymbol{D} = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B} \end{pmatrix}$ is a block matrix with $\boldsymbol{B} = (\boldsymbol{0}_{q\times 1}, \boldsymbol{I}_{q\times q})$.

To solve testing problem (13), a Wald-type test will also work. We state the result in the following theorem.

Theorem 4.2 *Under conditions (A1)-(A2), the test statistic for testing problem (13) is*

$$\nu_n = n\hat{\boldsymbol{\xi}}^T \boldsymbol{D}^T (\boldsymbol{D}\hat{\boldsymbol{\Sigma}}\boldsymbol{D}^T)^{-1}\boldsymbol{D}\hat{\boldsymbol{\xi}}. \tag{15}$$

where $\hat{\boldsymbol{\Sigma}}$ is the consistent estimator of $\boldsymbol{\Sigma}$. Then under $\mathcal{H}_0$,

$$\nu_n \xrightarrow{L} \chi^2_{2q}, \text{ as } n \to \infty.$$

Similar to Theorem 4.1, the proof of Theorem 4.2 is omitted. Theorem 4.2 can be used to test whether the autoregressive parameters are constants. This test is specifically designed to check whether the RCTMNBINAR(1)-X model is suitable for a given dataset. Although we need to know the value of q before applying the hypothesis test, it can still be used to check whether the specific explanatory variables are included in the model.

5 Forecasting for RCTMNBINAR(1)-X process

In this section, we focus on the forecasting problem for the RCTMNBINAR(1)-X process. Methods of prediction generally consist of two methods. One approach is to use the conditional mean as the point forecast, which is optimal in the sense of mean squared error. But it usually leads to a non-integer value, which is unsatisfactory for integer-valued time series of counts. Another approach is to employ the coherent forecasting method suggested by Freeland and McCabe (2004), which will only produce forecasts on $\mathbb{N}_0$. Specifically, we can use the median of the h-step-ahead forecast distribution as the point prediction to overcome the disadvantage of point forecast.

In this paper, we mainly focus on the one-step-ahead forecast, which is frequently applied in practice. To forecast the RCTMNBINAR(1)-X process, we use the method mentioned by Bu et al. (2008) and Weiß (2010) to deal with the problem that the number of states where X_t can take values is infinite. We can choose $M \in \mathbb{N}$ sufficiently large and then the approximated transition matrix is constructed as $\boldsymbol{P}_M = (p_{ij})_{i,j=0,1,\cdots,M}$, where p_{ij} is the transition probability given in (5). The h-step-ahead forecasting conditional distribution $p_h(x|X_n, \boldsymbol{W}_n, \boldsymbol{\theta})$ can be obtained approximatively as follows

$$p_h(x_{n+h}|X_n, \boldsymbol{W}_{n+h}, \boldsymbol{\theta}) = P(X_{n+h} = x_{n+h}|X_n = x_n, \boldsymbol{W}_{n+h}, \boldsymbol{\theta}) = [\boldsymbol{P}_M^h]_{x_{n+h},x_n}.$$

In practice, we can use the CML-estimator $\hat{\boldsymbol{\theta}}_{CML}$ to compute $p_h(x|X_n, \boldsymbol{W}_{n+h}, \boldsymbol{\theta})$. Since the maximum likelihood estimators $\hat{\boldsymbol{\theta}}_{CML}$ are asymptotically normal distributed around the true value $\boldsymbol{\theta}_0$ (see Theorem 3.2), we can work out the asymptotic distribution of $p_h(x|X_n, \boldsymbol{W}_{n+h}, \hat{\boldsymbol{\theta}}_{CML})$ as follows for prediction.

Theorem 5.1 *For a fixed $x \in \mathbb{N}_0$, the quantity $p_h(x|X_n, \boldsymbol{W}_{n+h}, \hat{\boldsymbol{\theta}})$ has an asymptotically normal distribution,*

$$\sqrt{n}\big(p_h(x|X_n, \boldsymbol{W}_{n+h}, \hat{\boldsymbol{\theta}}) - p_h(x|X_n, \boldsymbol{W}_{n+h}, \boldsymbol{\theta}_0)\big) \xrightarrow{L} N\big(0, \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^T\big).$$

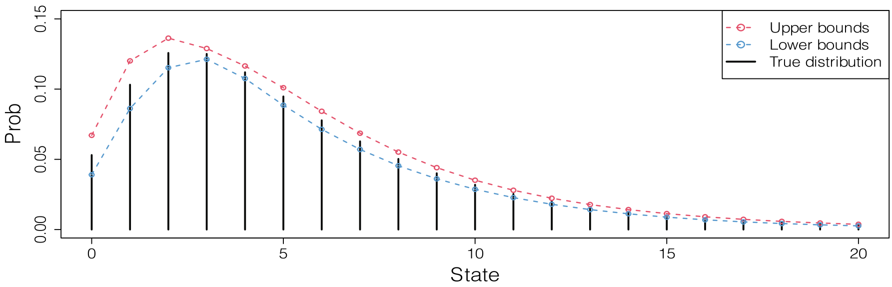

Fig. 1 One-step ahead forecasting distribution and the 95% forecasting confidence intervals

where $\boldsymbol{\Gamma} = \partial p_h(x|X_n, \boldsymbol{W}_{n+h}, \boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}_0^T$, $\hat{\boldsymbol{\theta}}$ *is a consistent estimator of* $\boldsymbol{\theta}_0$.

In fact, $\hat{\boldsymbol{\theta}}$ can be chosen as the CML-estimator $\hat{\boldsymbol{\theta}}_{CML}$ discussed in Sect. 3.2, then $\boldsymbol{\Sigma}$ would be $\boldsymbol{I}(\boldsymbol{\theta})^{-1}$ accordingly. Furthermore, the $100(1-\alpha)\%$ confidence interval for $p_h(x|X_n, \boldsymbol{W}_{n+h}, \hat{\boldsymbol{\theta}})$ is given by

$$\left(p_h(x|X_n, \boldsymbol{W}_{n+h}, \hat{\boldsymbol{\theta}}) - \frac{u_{1-\frac{\alpha}{2}}\sigma_h}{\sqrt{n}},\ p_h(x|X_n, \boldsymbol{W}_{n+h}, \hat{\boldsymbol{\theta}}) + \frac{u_{1-\frac{\alpha}{2}}\sigma_h}{\sqrt{n}}\right),$$

where $\sigma_h = \sqrt{\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^T}$, $u_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$-upper quantile of the standard normal distribution $N(0, 1)$.

As an illustration, we draw the one-step forecasting distribution and 95% forecasting confidence intervals for the RCTMNBINAR(1)-X process. The parameters are chosen the same as Scenario A in Sect. 6, i.e. $(\omega_1, \omega_2, \beta_{11}, \beta_{21}, \mu)^T = (0.2, 0.8, -0.4, -1.2, 4)^T$ and $r = 4$. Figure 1 shows us that the forecasting distribution is an unimodal distribution. We note that the probabilities are almost zero in larger states, which indicates that the approximation method mentioned before is feasible. The forecasting distribution covers the true distribution and the corresponding interval length in the smaller states is greater than others. Compared with the point forecast method, Fig. 1 provides more comprehensive statistical information about the prediction distribution.

6 Simulation

6.1 Finite-sample behavior of the estimators

In this section, we conduct simulation studies to evaluate the finite-sample performance of the CLS-estimator and CML-estimator. All the following simulations are performed under the $\mathcal{R}$ software based on 1000 replications. We conduct simulation studies using the following three different scenarios with sample size $n = 400, 800, 1500$ respectively.

Scenario A: $\boldsymbol{\theta}_0 = (\omega_1, \omega_2, \beta_{11}, \beta_{21}, \mu)^T = (0.2, 0.8, -0.4, -1.2, 4)^T$ and $r = 4$. In this scenario, we assume that the threshold variable r is known. The

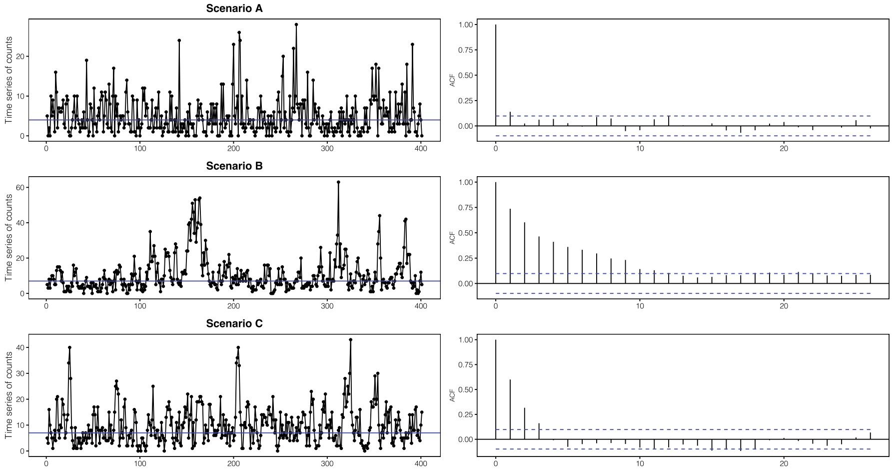

Fig. 2 Time series and ACF plots of X_t under Scenarios A-C with $n = 400$. The horizontal line shows the threshold values of each scenario

explanatory variables $\{W_{1,t}\}_{t \geqslant 1}$ are generated from Poisson distribution Pois(λ) with $\lambda = 3$.

Scenario B: $\boldsymbol{\theta}_0 = (\omega_1, \omega_2, \beta_{11}, \beta_{21}, \mu)^T = (0.3, 0.9, -0.5, -1.4, 3)^T$ and $r = 7$. In this scenario, we assume that the threshold variable r is known. The explanatory variables $\{W_{1,t}\}_{t \geqslant 1}$ are generated from normal distribution $N(0, 1)$.

Scenario C: $\boldsymbol{\theta}_0 = (\omega_1, \omega_2, \beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}, \mu)^T = (0.8, 0.2, -0.4, -0.6, 0.3, 0.9, 3)^T$ and $r = 7$. In this scenario, we assume that the threshold variable r is unknown. The explanatory variables $\{W_{1,t}\}_{t \geqslant 1}$ are generated from normal distribution $N(0, 1)$ and $\{W_{2,t}\}_{t \geqslant 1}$ are generated from an AR(1) process satisfying $W_{2,t} = 0.5W_{2,t-1} + \eta_t,\ \eta_t \sim N(0, 1)$.

Figure 2 shows the time series and the autocorrelation function (ACF) plots for each scenario under sample size $n = 400$. As seen in Fig. 2, the observations demonstrate different levels of autocorrelation. The autocorrelation function decays exponentially, which is similar to the AR(1)-type models. Furthermore, the piecewise characteristic of the threshold variable is not as clear as the traditional threshold models. This phenomenon is greatly different from the fixed coefficient integer-valued threshold time series models, which is probably caused by the explanatory variables introduced in the autoregressive coefficient. We also calculate the sample means and the variances for the three simulated series. The sample mean for the Scenarios A-C are 4.935, 10.466, 9.8 and the corresponding variance are 22.216, 97.179, 52.04. All the variances of the three series are larger than their means, showing different overdispersion features.

Tables 1 and 2 summarize the empirical bias, standard deviation (SD) and mean squared error (MSE) of the estimates across 1000 replications. The CLS-estimator $\hat{\boldsymbol{\theta}}_{CLS}$ and the CML-estimator $\hat{\boldsymbol{\theta}}_{CML}$ for Scenarios A and B are reported when the threshold variable r is known, while Table 3 reports the simulation results of Scenario C with r to be unknown. As seen in Tables 1 and 2, all the bias, SD and MSE decrease

Table 1 Simulation results for Scenario A with $r = 4$: bias, SD and SE

n	Para	CLS			CML		
		Bias	SD	MSE	Bias	SD	MSE
400	ω_1	0.2969	1.3371	1.8761	0.1897	0.8696	0.7921
	ω_2	0.1631	0.6437	0.4409	0.1086	0.4959	0.2577
	β_{11}	−0.2615	0.8545	0.7985	−0.0764	0.4067	0.1713
	β_{21}	−0.1615	0.5384	0.316	−0.0592	0.2958	0.0910
	μ	−0.0513	0.4409	0.1971	−0.0576	0.2723	0.0775
800	ω_1	0.1452	0.8698	0.7776	0.0594	0.5631	0.3206
	ω_2	0.0639	0.4444	0.2016	0.0267	0.3527	0.1251
	β_{11}	−0.1525	0.5337	0.3081	−0.0232	0.2134	0.0461
	β_{21}	−0.0736	0.3551	0.1315	−0.0152	0.2089	0.0439
	μ	−0.0083	0.3061	0.0938	−0.0206	0.1930	0.0377
1500	ω_1	0.0767	0.5613	0.3209	0.0294	0.3685	0.1367
	ω_2	0.0194	0.3030	0.0922	0.0126	0.2335	0.0547
	β_{11}	−0.0546	0.3051	0.0961	−0.0087	0.1314	0.0173
	β_{21}	−0.0232	0.2258	0.0515	−0.0072	0.1387	0.0193
	μ	−0.0122	0.2167	0.0471	−0.0111	0.1391	0.0195

as the sample size increases. This implies our estimates are consistent for all the parameters. In addition, most of the empirical bias, SD and MSE of CML-estimator $\hat{\boldsymbol{\theta}}_{CML}$ is smaller than the corresponding CLS-estimator $\hat{\boldsymbol{\theta}}_{CLS}$. This means the CML-estimator has better robustness and perform better than the CLS-estimator. We can obtain a similar conclusion from Table 3 with the threshold parameter r to be estimated together.

6.2 Powers of the tests

In this subsection, we conduct simulations to illustrate the performances of the tests discussed in Sect. 4. In the following, Scenarios D-F are used to study the powers of the test problem (11) and Scenarios G-I are used to study the powers of the test problem (13).

Scenario D: $\boldsymbol{\theta}_0 = (\omega_1, \omega_2, \beta_{11}, \beta_{21}, \mu)^T = (1.2, 1.2, -1.5, -1.5, 3)^T$ and $r = 5$. In this scenario, we assume that the threshold variable r is known. The explanatory variables $\{W_{1,t}\}_{t\geqslant 1}$ are generated from Poisson distribution Pois(λ) with $\lambda = 1$.

Scenario E: $\boldsymbol{\theta}_0 = (\omega_1, \omega_2, \beta_{11}, \beta_{21}, \mu)^T = (0.8, 0.8, -1.2, -1.2, 2)^T$ and $r = 6$. In this scenario, we assume that the threshold variable r is known. The explanatory variables $\{W_{1,t}\}_{t\geqslant 1}$ are generated from normal distribution $N(0, 1)$.

Scenario F: $\boldsymbol{\theta}_0 = (\omega_1, \omega_2, \beta_{11}, \beta_{21}, \mu)^T = (0.6, 0.6, -1.8, -1.8, 3)^T$ and $r = 8$. In this scenario, we assume that the threshold variable r is known. The

Table 2 Simulation results for Scenario B with $r = 7$: bias, SD and SE

n	Para	CLS			CML		
		Bias	SD	MSE	Bias	SD	MSE
400	ω_1	−0.0036	0.5803	0.3367	0.0737	0.3365	0.1186
	ω_2	−0.0089	0.2552	0.0652	0.0268	0.1844	0.0347
	β_{11}	−0.0529	0.3425	0.1201	−0.0252	0.2725	0.0749
	β_{21}	−0.0301	0.2171	0.0480	−0.0219	0.1828	0.0339
	μ	0.0478	0.5743	0.3321	−0.0646	0.3219	0.1078
800	ω_1	−0.0396	0.3688	0.1376	0.0156	0.2091	0.0440
	ω_2	−0.0231	0.1723	0.0302	0.0007	0.1234	0.0152
	β_{11}	−0.0281	0.2154	0.0472	−0.0181	0.1850	0.0345
	β_{21}	−0.0034	0.1481	0.0219	−0.0058	0.1210	0.0147
	μ	0.0655	0.4033	0.1669	−0.0081	0.2303	0.0531
1500	ω_1	0.0022	0.2630	0.0692	0.0172	0.1548	0.0242
	ω_2	−0.0052	0.1256	0.0158	0.0020	0.0889	0.0079
	β_{11}	−0.0187	0.1488	0.0225	−0.0129	0.1279	0.0165
	β_{21}	−0.0027	0.1055	0.0111	−0.0020	0.0888	0.0079
	μ	0.0112	0.2944	0.0868	−0.0137	0.1689	0.0287

explanatory variables $\{W_{1,t}\}_{t\geqslant 1}$ are generated from an AR(1) process satisfying $W_{1,t} = 0.5W_{1,t-1} + \eta_t,\ \eta_t \sim N(0, 1)$.

Scenario G: $\boldsymbol{\theta}_0 = (\omega_1, \omega_2, \beta_{11}, \beta_{21}, \mu)^T = (0.4, 0.6, 0, 0, 2)^T$ and $r = 6$. In this scenario, we assume that the threshold variable r is known. The explanatory variables $\{W_{1,t}\}_{t\geqslant 1}$ are generated from Poisson distribution Pois(λ) with $\lambda = 1$.

Scenario H: $\boldsymbol{\theta}_0 = (\omega_1, \omega_2, \beta_{11}, \beta_{21}, \mu)^T = (0.8, 1.2, 0, 0, 2)^T$ and $r = 9$. In this scenario, we assume that the threshold variable r is known. The explanatory variables $\{W_{1,t}\}_{t\geqslant 1}$ are generated from normal distribution $N(0, 1)$.

Scenario I: $\boldsymbol{\theta}_0 = (\omega_1, \omega_2, \beta_{11}, \beta_{21}, \mu)^T = (0.4, 1.5, 0, 0, 1)^T$ and $r = 6$. In this scenario, we assume that the threshold variable r is known. The explanatory variables $\{W_{1,t}\}_{t\geqslant 1}$ are generated from an AR(1) process satisfying $W_{1,t} = 0.5W_{1,t-1} + \eta_t,\ \eta_t \sim N(0, 1)$.

Firstly, we consider the empirical size and the empirical power of the test statistic (12) for testing problem (11) with sample size $n = 400, 800, 1500$. It is clear that Scenarios D-F are cases where $\mathcal{H}_0$ is true. Then we consider the empirical size and the empirical power of the test statistic (15) for testing problem (14) with sample size $n = 400, 800, 1500$, where Scenarios G-I are cases that $\mathcal{H}_0$ is true. Each experiment is based on 1000 replications using the CML-estimator. The selected nominal significance level $\alpha = 1\%, 5\%, 10\%$. Tables 4 and 5 reports the simulated results of the empirical size and the empirical power for selected nominal significance level $\alpha = 1\%, 5\%, 10\%$. From Table 4, we can see that the empirical size convergences to the significance

Table 3 Simulation results for Scenario C with $r = 7$: bias, SD and SE

n	Para	CLS			CML		
		Bias	SD	MSE	Bias	SD	MSE
400	ω_1	0.0673	0.773	0.6021	0.2008	0.5659	0.3605
	ω_2	−0.0517	0.2291	0.2418	0.0114	0.1550	0.1951
	$\beta_{1,1}$	−0.0444	0.4890	0.1795	−0.0339	0.4415	0.1533
	$\beta_{2,1}$	0.0321	0.4213	0.0535	0.0140	0.3901	0.0242
	$\beta_{1,2}$	−0.0106	0.1461	0.0215	−0.0068	0.1268	0.0161
	$\beta_{2,2}$	0.0113	0.1668	0.0279	0.0085	0.1392	0.0194
	μ	0.1142	0.5864	0.3569	−0.0620	0.3426	0.1212
	r	$(\hat{r}_{Bias}, \hat{r}_{SD}, \hat{r}_{MSE})$=(0.131, 1.4526, 2.1271)					
800	ω_1	−0.0045	0.4563	0.2083	−0.0091	0.3046	0.0929
	ω_2	−0.027	0.1588	0.0559	−0.0147	0.1126	0.0264
	$\beta_{1,1}$	−0.0225	0.2349	0.0359	−0.0099	0.1618	0.0184
	$\beta_{2,1}$	0.0214	0.1882	0.0257	0.0100	0.1355	0.0128
	$\beta_{1,2}$	−0.0086	0.1037	0.0108	−0.0018	0.0745	0.0056
	$\beta_{2,2}$	0.0107	0.1133	0.0130	0.0007	0.0750	0.0056
	μ	0.0694	0.4162	0.178	0.0462	0.2937	0.0884
	r	$(\hat{r}_{Bias}, \hat{r}_{SD}, \hat{r}_{MSE})$=(0.093, 0.6187, 0.3914)					
1500	ω_1	−0.0091	0.3046	0.0929	0.0318	0.1927	0.0382
	ω_2	−0.0147	0.1126	0.0264	0.0056	0.0731	0.0200
	$\beta_{1,1}$	−0.0099	0.1618	0.0184	−0.0074	0.1414	0.0149
	$\beta_{2,1}$	0.0100	0.1355	0.0128	0.0094	0.1219	0.0054
	$\beta_{1,2}$	−0.0018	0.0745	0.0056	0.0000	0.0635	0.0040
	$\beta_{2,2}$	0.0007	0.0750	0.0056	0.0007	0.0642	0.0041
	μ	0.0462	0.2937	0.0884	−0.0122	0.1681	0.0284
	r	$(\hat{r}_{Bias}, \hat{r}_{SD}, \hat{r}_{MSE})$=(0.009, 0.1922, 0.037)					

level as the sample size n increases. This implies that the asymptotic distribution in Theorem 4.1 is correct. Meanwhile, for Scenarios A-C where $\mathcal{H}_0$ of (11) is not true, all the results show a high empirical power of the test, since the empirical power grows to one very fast as n increases. As seen in Table 5, for Scenarios A-C where $\mathcal{H}_0$ of (13) is not true, the empirical power of the test statistic (15) approaches unity. This implies the proposed test statistics performs well in practice. Empirical sizes for Scenarios G-I are close to the nominal size in all cases, which implies that the conclusion of Theorem 4.2 is correct.

7 Application

In this section, we consider the monthly counts of assault in North Sydney from January 1996 to December 2022 to illustrate the superiority of our proposed model.

Table 4 Empirical size of test (11) based on CML method with different significance level

n	Nominal significance level	$\mathcal{H}_0$ is not true				$\mathcal{H}_0$ is true	
		Scenario A	Scenario B	Scenario C	Scenario D	Scenario E	Scenario F
400	$\alpha = 1\%$	0.576	0.627	0.776	0.01	0.008	0.004
	$\alpha = 5\%$	0.769	0.788	0.917	0.045	0.034	0.031
	$\alpha = 10\%$	0.845	0.845	0.957	0.079	0.078	0.067
800	$\alpha = 1\%$	0.911	0.894	0.993	0.012	0.004	0.012
	$\alpha = 5\%$	0.961	0.947	0.99	0.056	0.039	0.047
	$\alpha = 10\%$	0.978	0.971	0.99	0.1	0.078	0.111
1500	$\alpha = 1\%$	0.994	0.998	0.859	0.017	0.01	0.01
	$\alpha = 5\%$	0.999	0.998	0.974	0.046	0.048	0.042
	$\alpha = 10\%$	0.999	0.993	0.988	0.087	0.087	0.093

Table 5 Empirical power of test (13) based on CML estimator with different significance level

		$\mathcal{H}_0$ is not true			$\mathcal{H}_0$ is true		
n	Nominal significance level	Scenario A	Scenario B	Scenario C	Scenario G	Scenario H	Scenario I
400	$\alpha = 1\%$	0.983	1	1	0.002	0.003	0.003
	$\alpha = 5\%$	0.996	1	1	0.032	0.04	0.047
	$\alpha = 10\%$	1	1	1	0.08	0.071	0.093
800	$\alpha = 1\%$	1	1	1	0.005	0.009	0.008
	$\alpha = 5\%$	1	1	1	0.033	0.043	0.045
	$\alpha = 10\%$	1	1	1	0.098	0.081	0.008
1500	$\alpha = 1\%$	1	1	1	0.007	0.009	0.009
	$\alpha = 5\%$	1	1	1	0.051	0.058	0.051
	$\alpha = 10\%$	1	1	1	0.099	0.101	0.093

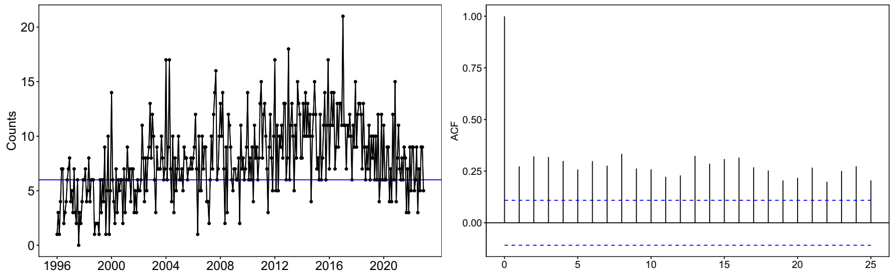

Fig. 3 Time series and ACF plots of the monthly counts of assault in North Sydney

It includes the subcategory of 'domestic violence related assault' and 'non-domestic violence related assault' during this time period. We handle the former subcategory as an example and label it as 'assault'. Let X_t denotes the monthly number of assault at time t. The dataset consists of 324 observations, and the sample mean and variance of $\{X_t\}$ are 7.83 and 12.773, which means that the analyzed series show considerable over-dispersion feature. Figure 3 shows the sample path and the sample autocorrelation function (ACF) plot for the time series $\{X_t\}$. From Fig. 3, we can see that the dataset may come from an integer-valued AR(1) process.

We choose the covariates according to previous studies by many scholars. Chen and Lee (2017) indicated that criminal activities can be influenced by environmental factors. Specifically, the local temperature would have an effect on criminal activities. Therefore, the first covariate is selected as the monthly mean maximum temperature in North Sydney. It is the average of all available daily maximum for the month, recorded by station 066037, and can be downloaded from the URL

$$http://www.bom.gov.au/climate/data/index.shtml.$$

Considering that the original temperature sequence $\{W_{1,t}^*\}$ exhibits typical seasonal characteristic, we conduct a 12-step difference on it, and denote the resulting series by $\{W_{1,t}\}$. The second covariate is selected as the monthly number of 'possession and/or use of amphetamines' in North Sydney. To remove the linear trends in the original series $\{W_{2,t}^*\}$, we conduct a first-order difference on it and denote the resulting series by $\{W_{2,t}\}$. As seen in Figs. 4-5, the differenced series are stationary series with no trend. Furthemore, we conduct the augmented Dickey-Fuller (ADF) test for the dataset and the covariate series, the p-values of $\{X_t\}$, $\{W_{1,t}\}$ and $\{W_{2,t}\}$ are both smaller than 0.01. These results indicate that the observed series and the covariates series are all stationary.

We use the proposed RCTMNBINAR(1)-X model and the following models to fit the dataset, and compare them via the AIC criterion.

- NBTINAR(1) process proposed by Yang et al. (2018):

$$X_t = (\alpha_1 * X_{t-1} + Z_{1,t})I_{1,t} + (\alpha_2 * X_{t-1} + Z_{2,t})I_{2,t},$$

where the innovations $Z_{i,t} \sim \mathrm{NB}(\nu, \frac{\alpha_i}{1+\alpha_i}),\ i = 1, 2.$

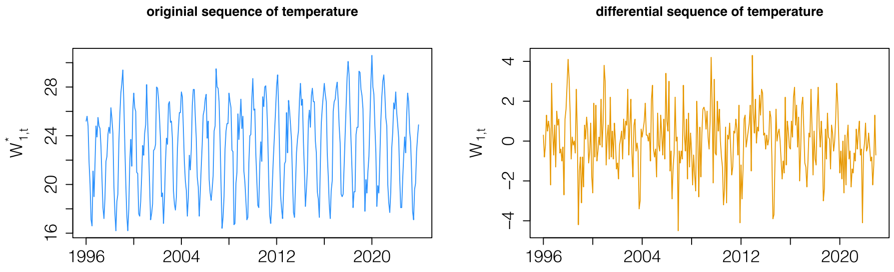

Fig. 4 Time series plots of $\{W^*_{1,t}\}$ and $\{W_{1,t}\}$

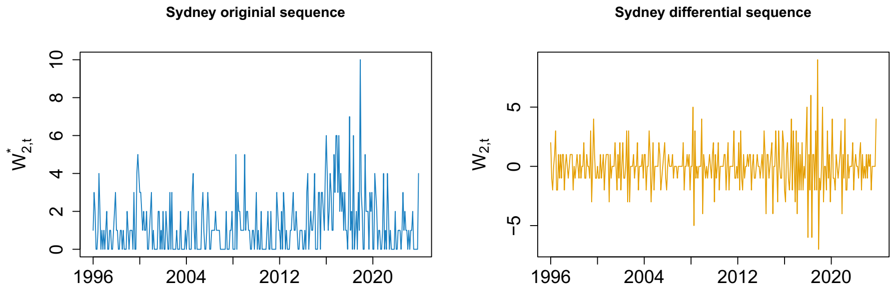

Fig. 5 Time series plots of $\{W^*_{2,t}\}$ and $\{W_{2,t}\}$

- RCMNBTINAR(1) process, proposed by Fan et al. (2023):

$$X_t = (\alpha_{1,t} \diamond X_{t-1} + Z_{1,t})I_{1,t}(r) + (\alpha_{2,t} \diamond X_{t-1} + Z_{2,t})I_{2,t}(r), \;\; t \in \mathbb{N}_0,$$

where $\alpha_{i,t}$ follows the Beta prime distribution with parameter m_i and known constant l and the innovations $Z_{i,t} \sim \text{Geom}(\mu_i)$, $i = 1, 2$.
- RCTINAR(1)-X model proposed by Yang et al. (2021):

$$X_t = (\alpha_{1,t} \circ X_{t-1})I_{1,t} + (\alpha_{2,t} \circ X_{t-1})I_{2,t} + Z_t,$$

where $\alpha_{i,t} = \frac{\exp(\omega_i+\beta_{i,1}W_{1,t})}{1+\exp(\omega_i+\beta_{i,1}W_{1,t})}$, $i = 1, 2$ and the innovations $Z_t \sim \text{Poi}(\lambda)$.

Table 6 reports the CML estimates, SD of the unknown parameters, AIC and BIC values for all the competing models. Note that when $\beta_{1,2} = \beta_{2,2} = 0$, the RCTMNBINAR(1)-X model is denoted by the RCTMNBINAR(1)-X1 model and when $\beta_{1,1} = \beta_{2,1} = 0$ the RCTMNBINAR(1)-X model is denoted by the RCTMNBINAR(1)-X2 model. The latter model is not mentioned in Table 6 because its nonlinearity test is not significant and thereby be removed from the fitted models. For all the fitted threshold models, the threshold value are calculated using the algorithm discussed in Sect. 3.3. We conduct the nonlinearity test discussed in Sect. 4.1, the p-values of the nonlinearity test for the NBTINAR(1) process, the RCTINAR(1)-X process, the RCTMNBINAR(1)-X process and the RCTMNBINAR(1)-X1 process are smaller than 0.05, which indicates strong threshold effect for the analyzed dataset. Furthermore, we also conduct the existence test for the explanatory variables discussed in

Table 6 Fitting results of the monthly counts of assault under different models

	Para	Estimator	SD	AIC	BIC
NBTINAR(1)	α_1	0.7364	0.0955	1752.719	1764.061
	α_2	0.5612	0.0464		
	ν	5	0.4723		
	r	6			
RCMNBTINAR(1)	m_1	2.1899	0.3674	1894.325	1909.448
	m_2	1.4231	0.1961		
	μ_1	2.8490	0.5852		
	μ_2	3.3831	0.6019		
	r	6			
RCTINAR(1)-X	ω_1	−2.6289	0.3565	1768.082	1786.986
	ω_2	0.3618	0.1208		
	$\beta_{1,1}$	−1.5917	0.3239		
	$\beta_{2,1}$	0.2504	0.1232		
	λ	4	0.3124		
	r	7			
RCMNBTINAR(1)-X	ω_1	0.6832	0.2417	1735.915	1762.38
	ω_2	0.3830	0.1605		
	$\beta_{1,1}$	−0.4559	0.7324		
	$\beta_{1,2}$	−0.2877	0.6105		
	$\beta_{2,1}$	0.1083	0.0703		
	$\beta_{2,2}$	−0.0673	0.0459		
	μ	1.8534	0.3382		
	r	6			
RCMNBTINAR(1)-X1	ω_1	2.8103	1.2630	1729.302	1748.205
	ω_2	0.3569	0.1628		
	$\beta_{1,1}$	−0.3292	0.6938		
	$\beta_{2,1}$	0.1019	0.0701		
	μ	1.8870	0.3555		
	r	6			

Sect. 4.2. The resulting p-values for the RCTINAR(1)-X model, RCTMNBINAR(1)-X model and RCTMNBINAR(1)-X1 model are smaller than 0.05, showing a strong evidence for the existence of explanatory variables. It also means that these explanatory variables play an important role in data modeling. All the fitting results are summarized in Table 6. To select the best model among all the competing models, we adopt the AIC and BIC criteria.

As can be seen from Table 6, the AIC and BIC value of the proposed RCTMNBINAR(1)-X1 model are smaller than the other competing models. Moreover, if we take a serious look at Table 6, one could find that the AIC value of the RCTINAR(1)-X model is bigger than the other models based on negative binomial

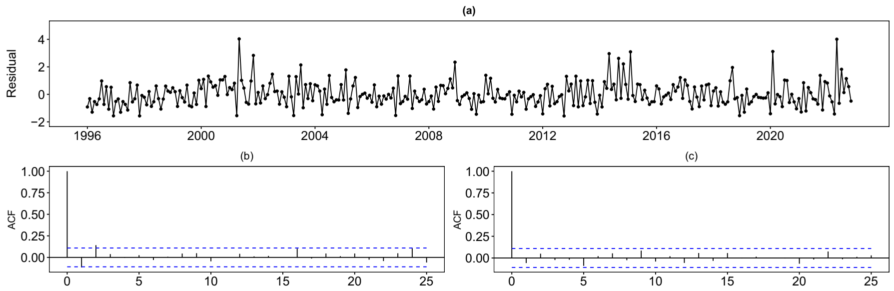

Fig. 6 Diagnostic checking plots of the fitted RCTMNBINAR(1)-X1 model. (a) The traces of standard residuals. (b) ACF plots of standard residuals. (c) ACF plots of squared residuals

operator with random coefficient, while the AIC value of other models are relatively close. It suggests that the integer-valued autoregressive threshold models based on negative binomial operator as an extension of the models based on binomial thinning operator can better capture the effect of exogenous variables for this dataset. In conclusion, Table 6 indicates that the proposed RCTMNBINAR(1)-X1 model is the most appropriate model for fitting this dataset.

To show the fitting details of the RCTMNBINAR(1)-X1 model, we draw the diagnostic checking plots in Fig. 6 based on the standardized Pearson residuals. The standardized Pearson residuals are used to check the adequacy of the fitting result of the RCTMNBINAR(1)-X1 model, which is defined by

$$e_t = \frac{X_t - E(X_t|X_{t-1}, \boldsymbol{W}_t)}{\sqrt{Var(X_t|X_{t-1}, \boldsymbol{W}_t)}}.$$

In practice, we can substitute the CML estimates into the conditional expectation and conditional variance equations to calculate $\hat{e}_t$. For an adequate model, these residuals will be uncorrelated, with mean about 0 and variance about 1. The time series plot, ACF of the fitted standardized and squared Person residuals are displayed in Fig. 6. We also conduct the ADF test for the standardized residuals. The p-value is smaller than 0.01, meaning that the residuals is a stationary sequence. For our fitted model, the mean and variance of the standardized Pearson residuals are 0.0138 and 0.9182, respectively. The residual of maximum observed value is 4.0324. As stated in Yang et al. (2023), we also consider introducing covariates in the innovation term to further account for heterogeneity of the observations in the future research.

The original dataset and the fitted series calculated by the median of the one-step ahead forecasting distribution of the RCTMNBINAR(1)-X1 model is presented in Fig. 7. It can be observed that the threshold model fits the data better when the observation X_t is larger. In other words, its improvement is mainly in the upper regime. In addition, as is pointed by Li et al. (2015) and Yang et al. (2023), a self-exciting hysteretic RCTMNBINAR(1)-X model would capture the piecewise features for time series of counts more precisely. Thus, it may be more flexible and practical in data fitting. Since this problem is beyond the scope of the present paper, we will complete these issues in a future study.

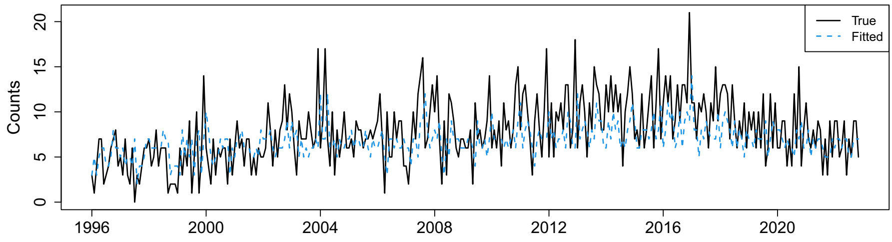

Fig. 7 Plot of fitted curves of the monthly counts of assult in North Sydney

8 Possible extension of the RCTMNBINAR(1)-X model

The RCTMNBINAR(1)-X model defined in (2) is a basic first-order model with two regimes and with delay parameter one. It can be extended to higher-order autoregressions by adapting the INAR(p) model of Du and Li (1991) in each regime with the autoregressive coefficients driven by logistic regression. This method leads to a pth-order Markov process called RCTMNBINAR(p)-X process, i.e.

$$X_t = (\sum_{j=1}^{p} \alpha_{j,t}^{(1)} \diamond X_{t-j}) I_{1,t}(r) + (\sum_{j=1}^{p} \alpha_{j,t}^{(2)} \diamond X_{t-j}) I_{2,t}(r) + Z_t,\ \ t \in \mathbb{N}_0, \tag{16}$$

where $I_{1,t}(r) = I\{0 \leqslant X_{t-1} \leqslant r\}$, $I_{2,t}(r) = I\{r < X_{t-1} < \infty\}$, r is the threshold variable. The autoregressive coefficient $\alpha_{j,t}^{(i)}$ in each regime is driven by logistic regression, i.e.

$$\log\left(\frac{\alpha_{j,t}^{(i)}}{1-\alpha_{j,t}^{(i)}}\right) = \omega_j^{(i)} + \boldsymbol{W}_t^T \boldsymbol{\beta}_j^{(i)},\ \ i \in \{1,2\},\ \ j \in \{1,2,\cdots,p\}, \tag{17}$$

where $\boldsymbol{\beta}_j^{(i)} = (\beta_{j,1}^{(i)}, \beta_{j,2}^{(i)}, \cdots, \beta_{j,q}^{(i)})$ are the regression coefficients. $\{\boldsymbol{W}_t := (W_{1,t}, \cdots, W_{q,t})^T\}$ is a q-dimensional explanatory variable which can be observed. $\{Z_t\}$ is a sequence of independent and identically distributed geometric random variables. Conditional on X_{t-1}, the counting series $\alpha_{j,t}^{(i)} \diamond X_{t-j}$ and Z_t are mutually independent, $i \in \{1,2\}$, $j \in \{1,2,\cdots,p\}$. It is easy to verify that the process $\{X_t\}_{t\in\mathbb{N}_0}$ defined in (16) with the following transition probabilities

$$\begin{aligned}
&P(X_t = x_t | X_{t-1} = x_{t-1}, \cdots, X_{t-p} = x_{t-p}, \boldsymbol{W}_t) \\
&= \sum_{i_1=0}^{x_t} \sum_{i=1}^{2} \binom{x_{t-1}+i_1}{i_1} (\frac{\alpha_{1,t}^{(i)}}{1+\alpha_{1,t}^{(i)}})^{i_1} (\frac{1}{1+\alpha_{1,t}^{(i)}})^{x_{t-1}+1} I_{i,t} \\
&\times \sum_{i_2=0}^{x_t} \sum_{i=1}^{2} \binom{x_{t-2}+i_2}{i_2} (\frac{\alpha_{2,t}^{(i)}}{1+\alpha_{2,t}^{(i)}})^{i_2} (\frac{1}{1+\alpha_{2,t}^{(i)}})^{x_{t-2}+1} I_{i,t} \\
&\times \cdots
\end{aligned}$$

$$\times \sum_{i_p=0}^{x_t} \sum_{i=1}^{2} \binom{x_{t-p}+i_p}{i_p} (\frac{\alpha_{p,t}^{(i)}}{1+\alpha_{p,t}^{(i)}})^{i_p} (\frac{1}{1+\alpha_{p,t}^{(i)}})^{x_{t-p}+1} I_{i,t} \times P(Z_t = x_t - \sum_{s=1}^{p} i_s), \tag{18}$$

where $\alpha_{j,t}^{(i)}$ is defined in (17). Similar to Chen et al. (2022), we can introduce a time-dependent innovational vector. Motivated by Tzougas and di Cerchiara (2021), we can assume that Z_t follows the mixed negative binomial distribution, that is

$$Z_t \sim \mathrm{NB}(\vartheta, \frac{\vartheta}{\vartheta + \psi_{\mathrm{t}}}),$$

where $\psi_t = \exp(\zeta X_{t-1})$, ζ is a constant. By imposing the past information in the distribution of Z_t, the model can better capture the time-dependence trend. The transition probabilities given in (18) can be used to construct the loglikelihood function to estimate the unknown parameters of the RCTMNBINAR(p)-X model. The RCTMNBINAR(p)-X model defined in (16) is a generalization of the RCTMNBINAR(1)-X model. It would be interesting to consider the statistical inference problem for the model parameters, including the autoregressive parameters $\alpha_{j,t}^{(i)}$, the threshold parameter r and the delay parameter. In addition, we can also consider introducing the hysteresis zone proposed by Li et al. (2015) into the regime-switching structure to further extend the autoregressive model. However, as these problems are somewhat beyond the scope of the present paper and need deeper attention, we will discuss these issues in the future project.

9 Conclusion

In this paper, a new random coefficient integer-valued self-exciting threshold autoregressive process based on modified negative binomial operator is proposed, where the autoregressive parameters depend on explanatory variables. The CLS-estimator and the CML-estimator of the unknown parameters are derived when the threshold value is known, as well as their consistency and asymptotic properties. The testing problems including nonlinearity and the existence of the explanatory variables of the RCTMNBINAR(1)-X model are considered. The coherent forecasts for the RCTMNBINAR(1)-X model are also addressed. Monte Carlo simulation suggests that the parameter estimation and the hypothesis test results are reliable. A real example is applied to illustrate the performance of the proposed model. Finally, some possible extensions of the RCTMNBINAR(1)-X model are provided. We leave this issue as our future work since it needs more attention.

Appendix

Proof of Proposition 2.1: Suppose that i and j are arbitrary non-negative integers. The transition probabilities can be defined by the following equation

$$
\begin{aligned}
& P(X_t = x_t | X_{t-1} = x_{t-1}, \boldsymbol{W}_t) \\
&= P((\alpha_{1,t} \diamond X_{t-1}) I_{1,t} + (\alpha_{2,t} \diamond X_{t-1}) I_{2,t} + Z_t = x_t | X_{t-1} = x_{t-1}, \boldsymbol{W}_t) \\
&= p(x_{t-1}, x_t, \alpha_{1,t}, \mu) I_{1,t} + p(x_{t-1}, x_t, \alpha_{2,t}, \mu) I_{2,t},
\end{aligned}
$$

where

$$
p(x_{t-1}, x_t, \alpha_{i,t}, \mu) = \sum_{k=0}^{x_t} \binom{x_{t-1}+k}{k} \frac{\alpha_{i,t}^k \mu^{x_t - k}}{(1+\alpha_{i,t})^{x_{t-1}+k+1}(1+\mu)^{x_t - k + 1}}, \; i = 1, 2.
$$

According to Theorem 3.1 of Tweedie (1975), the sufficient condition of $\{X_t\}_{t \in \mathbb{N}_0}$ to be ergodic is that there exists a set K and non-negative measurable function g on state space $\mathbb{N}_0$ satisfying the following equation

$$
\int_{\mathbb{N}_0} P(x, dy) g(y) \leqslant g(x) - 1, \; x \in K^C, \tag{19}
$$

where K^C is a complement of the set K, and for some fixed B,

$$
\int_{\mathbb{N}_0} P(x, dy) g(y) = \lambda(x) \leqslant B < \infty, \; x \in K, \tag{20}
$$

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

$$
\begin{aligned}
& \int_{\mathbb{N}_0} g(y) dP(X_1 = y | X_0 = x_0) = E(X_1 | X_0 = x_0) \\
&\quad = [E(\alpha_{1,1} | X_0 = x_0)(x_0 + 1)] I_{1,1} + [E(\alpha_{2,1} | X_0 = x_0)(x_0 + 1)] I_{2,1} + \mu \\
&\quad \leqslant c_1 x_0 + c_2,
\end{aligned}
$$

where $c_1 = \max\{E(\alpha_{1,1} | X_0 = x_0), E(\alpha_{2,1} | X_0 = x_0)\} < 1$ and $c_2 = \max\{E(\alpha_{1,1} | X_0 = x_0), E(\alpha_{2,1} | X_0 = x_0)\} + \mu$. □

Let $N = [\frac{c_2+1}{1-c_1}] + 1$, where $[x]$ denotes the integer part of x. Then for $x_0 > N$, we have

$$
c_1 x_0 + c_2 \leqslant x_0 - 1 = g(x_0) - 1,
$$

Then for $0 \leqslant x_0 \leqslant N - 1$, we have

$$
\int_{\mathbb{N}_0} g(y) dP(X_1 = y | X_0 = x_0) \leqslant x_0 + c_2 \leqslant N + c_2 < \infty.
$$

Let $K = \{0, 1, \cdots, N-1\}$ and $K^C = \{N, N+1, \cdots\}$, then (19) and (20) both hold. Furthermore, as $P(X_t = x_t | X_{t-1} = x_{t-1}, \boldsymbol{W}_t) > 0$, we can see that $\{X_t\}$ is an irreducible, aperiodic Markov chain. As $\lim_{t\to\infty} P(X_t = 0 | X_0 = 0) > 0$, it follows that state 0 is positive recurrent. Therefore, $\{X_t\}_{t\in\mathbb{N}_0}$ is a positive recurrent and ergodic Markov chain. By Proposition 2.1 in Yang et al. (2018), the existence of a strictly stationary distribution of process (2) could be ensured, which completes the proof.

Proof of Theorem 3.1: According to Theorem 3.1 in Zheng and Basawa (2008) (or Klimko and Nelson (1978)), since $g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t) = E(X_t | X_{t-1}, \boldsymbol{W}_t)$, Theorem 3.1 holds if the following conditions are satisfied.

(B1) $\partial g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t)/\partial\theta_i, \partial^2 g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t)/\partial\theta_i\partial\theta_j, \partial^3 g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t)/\partial\theta_i\partial\theta_j\theta_k$, $1 \leqslant i, j, k \leqslant 2q+3$ exist and continuous;

(B2) $E|u_t(\boldsymbol{\theta})\partial g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t)/\partial\theta_i| < \infty$, $E|u_t(\boldsymbol{\theta})\partial^2 g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t)/\partial\theta_i\partial\theta_j| < \infty$, $E|\partial g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t)/\partial\theta_i \cdot \partial g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t)/\partial\theta_j| < \infty$ for $1 \leqslant i, j, k \leqslant 2q+3$;

(B3) for $1 \leqslant i, j, k \leqslant q$, there exist functions

$$H^{(0)}(X_{t-1}, \cdots, X_0), H_i^{(1)}(X_{t-1}, \cdots, X_0), H_{ij}^{(2)}(X_{t-1}, \cdots, X_0),$$
$$H_{ijk}^{(3)}(X_{t-1}, \cdots, X_0),$$

such that

$$|g| \leqslant H^{(0)}, |\partial g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t)/\partial\theta_i| \leqslant H_i^{(1)}, |\partial^2 g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t)/\partial\theta_i\partial\theta_j| \leqslant H_{ij}^{(2)},$$
$$|\partial^3 g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t)/\partial\theta_i\partial\theta_j\partial\theta_k| \leqslant H_{ijk}^{(3)},$$

and

$$E\left|X_t H_{ijk}^{(3)}(X_{t-1}, \cdots, X_0)\right| < \infty,$$
$$E\left|H^{(0)}(X_{t-1}, \cdots, X_0) H_{ijk}^{(3)}(X_{t-1}, \cdots, X_0)\right| < \infty,$$
$$E\left|H_i^{(1)}(X_{t-1}, \cdots, X_0) H_{ij}^{(2)}(X_{t-1}, \cdots, X_0)\right| < \infty;$$

(B4) for $t \geqslant 1$, $E(X_t | X_{t-1}, \cdots, X_0) = E(X_t | X_{t-1})$ $a.s.$, and

$$E(u_t^2(\boldsymbol{\theta})|\partial g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t)/\partial\theta_i \cdot \partial g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t)/\partial\theta_j|) < \infty.$$

To verify the above conditions, notice that $g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t) = (X_{t-1} + 1)[\frac{\exp(\omega_1 + \boldsymbol{W}_t^T\boldsymbol{\beta}_1)}{1+\exp(\omega_1 + \boldsymbol{W}_t^T\boldsymbol{\beta}_1)} I_{1,t} + \frac{\exp(\omega_2 + \boldsymbol{W}_t^T\boldsymbol{\beta}_2)}{1+\exp(\omega_2 + \boldsymbol{W}_t^T\boldsymbol{\beta}_2)} I_{2,t}] + \mu$. Then we have

$$g_t(\boldsymbol{\theta}, X_{t-1}, \boldsymbol{W}_t) \leqslant (X_{t-1} + 1) + \mu.$$

$\square$

It is easy to verify that

$$\left|\frac{\partial E(X_1|X_0,\boldsymbol{W}_1)}{\partial \omega_i}\right| \leqslant X_0+1,\ i=1,2,$$
$$\left|\frac{\partial E(X_1|X_0,\boldsymbol{W}_1)}{\partial \beta_{i,j}}\right| \leqslant (X_0+1)W_{j,t},\ i=1,2,\ j=1,2,\cdots,q,$$
$$\left|\frac{\partial E(X_1|X_0,\boldsymbol{W}_1)}{\partial \mu}\right| \leqslant 1.$$

Thus the second and third derivatives are finite. All the conditions are satisfied and hence Theorem 3.1 holds.

Proof of Theorem 3.2: To prove Theorem 3.2, we want to use Theorem 2.1 and 2.2 in Billingsley (1961) about estimates for the parameters of general stochastic process with discrete time. It is sufficient to check following conditions (C1)-(C6) hold. For simplicity, we denote $f(k,\mu) := P(Z_t=k) = \frac{\mu^k}{(1+\mu)^{k+1}}$ as the probability mass function of Z_t.

(C1) The set $\{k : f(k,\mu)>0\}$ does not depend on μ.
(C2) $E(Z_t^3)<\infty$.
(C3) For $\forall k$, $f(k,\mu)$ is three times continuously differentiable with respect to μ.
(C4) For $\forall \mu' \in B$, where B is an open subset of $\mathbb{R}$. There exists a neighborhood U of μ' such that

$$\sum_{k=0}^{\infty}\sup_{\mu\in B} f(k,\mu)<\infty,\quad \sum_{k=0}^{\infty}\sup_{\mu\in B}\left|\frac{\partial f(k,\mu)}{\partial\mu}\right|<\infty,\quad \sum_{k=0}^{\infty}\sup_{\mu\in B}\left|\frac{\partial^2 f(k,\mu)}{\partial\mu^2}\right|<\infty.$$

(C5) For $\forall\mu'\in B$, there exists a neighborhood U of μ' and increasing sequences $\psi_1(n)=const1\cdot n$, $\psi_{11}(n)=const2\cdot n^2$ and $\psi_{111}(n)=const3\cdot n^3$, where $const1,\ const2,\ const3$ are suitable constants, and $n\geqslant 0$ such as $\forall\mu\in B$ and $k\leqslant n$ with nonvanishing $f(k,\mu)$,

$$\left|\frac{\partial f(k,\mu)}{\partial\mu}\right|\leqslant\psi_1(n)f(k,\mu),\ \left|\frac{\partial^2 f(k,\mu)}{\partial\mu^2}\right|\leqslant\psi_{11}(n)f(k,\mu),\ \left|\frac{\partial^3 f(k,\mu)}{\partial\mu^3}\right|\leqslant\psi_{111}(n)f(k,\mu),$$

and with respect to the stationary distribution of the process $\{X_t\}_{t\in\mathbb{N}_0}$,

$$E[\psi_1^3(X_1)]<\infty,\ E[X_1\psi_{11}(X_2)]<\infty,\ E[\psi_1(X_1)\psi_{11}(X_2)]<\infty,$$
$$E[\psi_{111}(X_1)]<\infty.$$

(C6) Let $\boldsymbol{I}(\boldsymbol{\theta})=(\sigma_{ij})_{(2q+3)\times(2q+3)}$ denotes the Fisher information matrix with elements $\sigma_{ij}=E(\frac{\partial}{\partial\theta_i}\log P(X_1,X_2)\frac{\partial}{\partial\theta_j}\log P(X_1,X_2))$. $P(X_1,X_2)$ denotes the transition probability given in (5). $\boldsymbol{I}(\boldsymbol{\theta})$ is non-singular matrix.

As Franke and Seligmann (1993) and Monteiro et al. (2012) stated, (C1)-(C5) are easy to verfied. Thus for the RCTMNBINAR(1)-X model, we first need to check all elements in $\boldsymbol{I}(\boldsymbol{\theta})$ are finite to prove condition (C6). In other words, we need to check the following statements are both true.

(D1) $E\left|\frac{\partial}{\partial\theta_i}\log P(X_1, X_2)\right|^2 < \infty,\ i = 1, 2, \cdots, 2q+3;$

(D2) $E\left|\frac{\partial}{\partial\theta_i}\log P(X_1, X_2)\frac{\partial}{\partial\theta_j}\log P(X_1, X_2)\right|^2 < \infty,\ i \neq j,\ i, j \in \{1, 2, \cdots, 2q+3\}.$

□

First, we have

$$-\frac{x_{t-1}+1}{1+\alpha_{i,t}} \leqslant \frac{\partial \log p(x_{t-1}, x_t, \alpha_{i,t}, \mu)}{\partial\omega_i} \leqslant \frac{x_t}{1+\alpha_{i,t}},\ i = 1, 2.$$

Thus we have

$$E\left|\frac{\partial}{\partial\beta_{i,j}}\log P(X_1, X_2)\right|^2 \leqslant CE(X^2),$$

since $\left|\frac{\partial \log p(x_{t-1}, x_t, \alpha_{i,t}, \mu)}{\partial\omega_i}\right| \leqslant \max\{x_t, x_{t-1}+1\}$, C is some suitable constant.

Second,

$$-\frac{(x_{t-1}+1)W_{j,t}}{1+\alpha_{i,t}} \leqslant \frac{\partial \log p(x_{t-1}, x_t, \alpha_{i,t}, \mu)}{\partial\beta_{i,j}} \leqslant \frac{x_t W_{j,t}}{1+\alpha_{i,t}},\ \text{if } W_{j,t} \geqslant 0,$$
$$\frac{x_t W_{j,t}}{1+\alpha_{i,t}} \leqslant \frac{\partial \log p(x_{t-1}, x_t, \alpha_{i,t}, \mu)}{\partial\beta_{i,j}} \leqslant -\frac{(x_{t-1}+1)W_{j,t}}{1+\alpha_{i,t}},\ \text{if } W_{j,t} < 0.$$

Therefore, we have

$$E\left|\frac{\partial}{\partial\beta_{i,j}}\log P(X_1, X_2)\right|^2 < CE(X^2)E(W^2),\ i = 1, 2,\ j = 1, \cdots, q.$$

Third,

$$\begin{aligned}\left|\frac{\partial \log p(x_{t-1}, x_t, \alpha_{i,t}, \mu)}{\partial\mu}\right| &= \frac{1}{p(x_{t-1}, x_t, \alpha_{i,t}, \mu)}\sum_{k=0}^{j}\binom{i+k}{k}\frac{\alpha_{i,t}^k}{(1+\alpha_{i,t})^{x_{t-1}+k+1}} \\ &\quad \times \frac{(x_t-k-\mu)\mu^{x_t-k-1}}{(1+\mu)^{x_t-k+2}} \\ &\leqslant \frac{x_t}{\mu(1+\mu)}.\end{aligned}$$

Thus we have $E\left|\frac{\partial}{\partial\mu}\log P(X_1, X_2)\right|^2 < E(X^2) < \infty.$

Acknowledgements This work is supported by National Natural Science Foundation of China (Grant Nos. 12271231, 12001229, 11901053) and China Scholarship Council (Grant No. CSC202206170056).

References

Aleksić MS, Ristić MM (2021) A geometric minification integer-valued autoregressive model. Appl Math Model 90:265–280. https://doi.org/10.1016/j.apm.2020.08.047

Al-Osh MA, Alzaid AA (1987) First-order integer-valued autoregressive (INAR(1)) process. J Time Ser Anal 8(3):261–275. https://doi.org/10.1111/j.1467-9892.1987.tb00438.x

Awale M, Balakrishna N, Ramanathan TV (2019) Testing the constancy of the thinning parameter in a random coefficient integer autoregressive model. Stat Pap 60:1515–1539. https://doi.org/10.1007/s00362-017-0884-x

Billingsley P (1961) Statistical inference for Markov processes. The University of Chicago Press, Chicago

Bu R, McCabe B, Hadri K (2008) Maximum likelihood estimation of higher-order integer-valued autoregressive processes. J Time Ser Anal 29:973–994. https://doi.org/10.1111/j.1467-9892.2008.00590.x

Chen CWS, Lee S (2016) Generalized Poisson autoregressive models for time series of counts. Comput Stat Data Anal 99:51–67. https://doi.org/10.1016/j.csda.2016.01.009

Chen CWS, Khamthong K, Lee S (2019) Markov switching integer-valued GARCH models for dengue counts. J R Stat Soc Ser C 68:963–983. https://doi.org/10.1111/rssc.12344

Chen CWS, Khamthong K (2020) Bayesian modelling of nonlinear negative binomial integer-valued GARCHX models. Stat Model 20(6):537–561. https://doi.org/10.1177/1471082X19845541

Chen H, Zhu F, Liu X (2022) A new bivariate INAR(1) model with time-dependent innovation vectors. Stats 5(3):819–840. https://doi.org/10.3390/stats5030048

Chen H, Li Q, Zhu F (2023) A covariate-driven beta-binomial integer-valued GARCH model for bounded counts with an application. Metrika. https://doi.org/10.1007/s00184-023-00894-5

Davis RA, Fokianos K, Holan SH et al (2021) Count time series: a methodological review. J Am Stat Assoc 116(535):1533–1547. https://doi.org/10.1080/01621459.2021.1904957

Ding X, Wang D (2016) Empirical likelihood inference for INAR(1) model with explanatory variables. J Korean Stat Soc 45(4):623–632. https://doi.org/10.1016/j.jkss.2016.05.004

Du J, Li Y (1991) The integer-valued autoregressive (INAR(p)) model. J Time Ser Anal 12(2):129–142. https://doi.org/10.1111/j.1467-9892.1991.tb00073.x

Fan Y, Wang D, Cheng J (2023) A new threshold INAR(1) model based on modified negative binomial operator with random coefficient. J Stat Comput Simul. https://doi.org/10.1080/00949655.2023.2282742

Franke J, Seligmann T (1993) Conditional maximum likelihood estimates for INAR(1) processes and their application to modeling eplieptic seizure counts. In: Subba Rao T (ed) Developments in Time Series Analysis. Chapman & Hall, London, pp 310–330

Freeland RK, McCabe BPM (2004) Analysis of low count time series data by Poisson autoregression. J Time Ser Anal 25:701–722. https://doi.org/10.1111/j.1467-9892.2004.01885.x

Han L, McCabe B (2013) Testing for parameter constancy in non-Gaussian time series. J Time Ser Anal 34:17–29. https://doi.org/10.1111/j.1467-9892.2012.00810.x

Kang Y, Wang D, Yang K (2020) Extended binomial AR(1) processes with generalized binomial thinning operator. Commun Stat Theor Methods 49:3498–3520

Klimko LA, Nelson PI (1978) On conditional least squares estimation for stochastic processes. Ann Stat 6(3):629–642. https://doi.org/10.1214/aos/1176344207

Li H, Yang K, Zhao S et al (2018) First-order random coefficients integer-valued threshold autoregressive processes. AStA Adv Stat Anal 102:305–331. https://doi.org/10.1007/s10182-017-0306-3

Li H, Liu Z, Yang K et al (2023) A pth-order random coefficients mixed binomial autoregressive process with explanatory variables. Comput Stat. https://doi.org/10.1007/s00180-023-01396-8

Li G, Guan B, Li WK et al (2015) Hysteretic autoregressive time series models. Biometrika 102:717–723. https://doi.org/10.1093/biomet/asv017

Liu M, Li Q, Zhu F (2019) Threshold negative binomial autoregressive model. Statistics 53(1):1–25. https://doi.org/10.1080/02331888.2018.1546307

Liu M, Li Q, Zhu F (2020) Self-excited hysteretic negative binomial autoregression. AStA Adv Stat Anal 104:325–361. https://doi.org/10.1007/s10182-019-00360-6

Monteiro M, Scotto MG, Pereira I (2012) Integer-valued self-exciting threshold autoregressive processes. Commun Stat Theor Methods 41:2717–2737. https://doi.org/10.1080/03610926.2011.556292

Möller TA, Silva ME, Weiß CH et al (2016) Self-exciting threshold binomial autoregressive processes. AStA Adv Stat Anal 100:369–400. https://doi.org/10.1007/s10182-015-0264-6

Qian L, Zhu F (2022) A new minification integer-valued autoregressive process driven by explanatory variables. Aust N Z J Stat 64(4):478–494. https://doi.org/10.1111/anzs.12379

Ristić MM, Bakouch HS, Nastić AS (2009) A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. J Stat Plan Inference 139:2218–2226. https://doi.org/10.1016/j.jspi.2008.10.007

Tweedie RL (1975) Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. Stochastic Processes and their Applications 3:385–403. https://doi.org/10.1016/0304-4149(75)90033-2

Tzougas G, di Cerchiara AP (2021) The multivariate mixed negative binomial regression model with an application to insurance a posteriori ratemaking. Insur Math Econ 101:602–625. https://doi.org/10.1016/j.insmatheco.2021.10.001

Wang C, Liu H, Yao J et al (2014) Self-excited threshold poisson autoregression. J Am Stat Assoc 109:777–787. https://doi.org/10.1080/01621459.2013.872994

Wang X, Wang D, Yang K et al (2021) Estimation and testing for the integer-valued threshold autoregressive models based on negative binomial thinning. Commun Stat Simul Comput 50:1622–1644. https://doi.org/10.1080/03610918.2019.1586929

Weiß CH (2010) The INARCH(1) model for overdispersed time series of counts. Commun Stat Simul Comput 39:1269–1291. https://doi.org/10.1080/03610918.2010.490317

Yang K, Li H, Wang D et al (2021) Random coefficients integer-valued threshold autoregressive processes driven by logistic regression. AStA Adv Stat Anal 105:533–557. https://doi.org/10.1007/s10182-020-00379-0

Yang K, Li H, Wang D (2018) Estimation of parameters in the self-exciting threshold autoregressive processes for nonlinear time series of counts. Appl Math Model 57:226–247. https://doi.org/10.1016/j.apm.2018.01.003

Yang K, Wang D, Jia B et al (2018) An integer-valued threshold autoregressive process based on negative binomial thinning. Stat Papers 59(3):1131–1160. https://doi.org/10.1007/s00362-016-0808-1

Yang K, Zhao X, Dong X et al (2023) Self-exciting hysteretic binomial autoregressive processes. Stat Pap. https://doi.org/10.1007/s00362-023-01444-x

Yang K, Zhao Y, Li H et al (2023) On bivariate threshold Poisson integer-valued autoregressive processes. Metrika 86:931–963. https://doi.org/10.1007/s00184-023-00899-0

Yang K, Xu N, Li H et al (2023) Multivariate threshold integer-valued autoregressive processes with explanatory variables. Appl Math Model 124:142–166. https://doi.org/10.1016/j.apm.2023.07.030

Yu M, Wang D, Yang K (2019) A class of observation-driven random coefficient INAR(1) processes based on negative binomial thinning. J Korean Stat Soc 48:248–264. https://doi.org/10.1016/j.jkss.2018.11.004

Zhao Z, Wang D, Peng C (2013) Coefficient constancy test in generalized random coefficient autoregressive model. Appl Math Comput 219(20):10283–10292. https://doi.org/10.1016/j.amc.2013.03.135

Zhang Q, Wang D, Fan X (2020) A negative binomial thinning-based bivariate INAR(1) process. Stat Neerl 74:517–537. https://doi.org/10.1111/stan.12210

Zhang R, Wang D (2023) A new binomial autoregressive process with explanatory variables. J Comput Appl Math 420:114–814. https://doi.org/10.1016/j.cam.2022.114814

Zheng H, Basawa IV, Datta S (2007) First-order random coefficient integer-valued autoregressive processes. J Stat Plan Inference 137:212–229. https://doi.org/10.1016/j.jspi.2005.12.003

Zheng H, Basawa IV (2008) First-order observation-driven integer-valued autoregressive processes. Stat Probab Lett 78:1–9. https://doi.org/10.1016/j.spl.2007.04.017

Zhang J, Zhu F, Chen H (2023) Two-threshold-variable integer-valued autoregressive model. Mathematics 11(16):3586. https://doi.org/10.3390/math11163586