J Syst Sci Complex (2023) 36: 843–865

Empirical Likelihood for a First-Order Generalized Random Coefficient Integer-Valued Autoregressive Process*

CHENG Jianhua · WANG Xu · WANG Dehui

DOI: 10.1007/s11424-023-1051-1
Received: 28 February 2021 / Revised: 12 September 2021

Abstract In this paper, the authors consider the empirical likelihood method for a first-order generalized random coefficient integer-valued autoregressive process. The authors establish the log empirical likelihood ratio statistic and obtain its limiting distribution. Furthermore, the authors investigate the point estimation, confidence regions and hypothesis testing for the parameters of interest. The performance of empirical likelihood method is illustrated by a simulation study and a real data example.

Keywords Empirical likelihood, generalized random coefficient, integer-valued time series, thinning operator.

1 Introduction

Integer-valued time series models have been well developed since the 1970s, because various kinds of correlated count data are commonly encountered in practice. The areas that could often produce these types of data, such as the numbers of patients, insurance claims, accidents, wet days, and so on, include epidemiology, economics, communications and meteorology, etc. One of the most popular approaches to analyze integer-valued time series is the class of autoregressive moving average processes based on different thinning operators. See for example [1–4] for detailed and recent surveys on these models.

Pioneering works on the thinning-operator-based model are due to [5] and [6], the authors propose the first-order integer-valued autoregressive process (INAR(1)) that is written as

$$X_t = \phi \circ X_{t-1} + \varepsilon_t, \quad t \geq 1, \tag{1}$$

CHENG Jianhua · WANG Xu
School of Mathematics, Jilin University, Changchun 130012, China.
WANG Dehui (Corresponding author)
School of Mathematics and Statistics, Liaoning University, Shenyang 110036, China.
Email: wangdh@jlu.edu.cn.
*This work was supported by the National Natural Science Foundation of China under Grant Nos. 11871028 and 11731015.
◇ *This paper was recommended for publication by Editor DONG Yuexiao.*

where $\phi \in [0,1)$, and the so-called binomial thinning operator "$\circ$" is introduced by [7] with the following definition

$$\phi \circ X_{t-1} = \sum_{k=1}^{X_{t-1}} B_k^{(t)}, \tag{2}$$

in which $\{B_k^{(t)}, k \geq 1\}$ is a sequence of i.i.d. Bernoulli random variables with mean ϕ and independent of X_{t-1}.

During the last decades, the binomial thinning operator and its generalizations have been widely implemented to construct more flexible count data models for practical purposes. For instance, negative binomial thinning operator[8], expectation thinning operator[9], Poisson thinning operator[10], signed binomial thinning operator[11], signed generalized power series thinning operator[12], Pegram's thinning operator[13], GSC thinning operator[14], extended negative binomial thinning operator[15], and so on. In a similar way as the RCAR (Random Coefficient Autoregressive) model, [16–18] consider random coefficient thinning to describe the situation that the parameter ϕ may vary randomly over time because of various environmental factors. [19] and [20] study models in which the thinning parameters are driven by observation and explanatory variables, respectively.

In this paper, we focus on a generalized random coefficient thinning operator that is proposed by [21], in analogy to and extend the concept of generalized thinning operator in [22] which allows $\{B_k^{(t)}, k \geq 1\}$ in (2) to be i.i.d. integer-valued random variables but not necessarily Bernoulli-distributed, such that the binomial thinning operator, the negative binomial thinning operator, the expectation thinning operator and Poisson thinning operator are included as its special cases. Our main purpose is to develop the empirical likelihood (EL) method for a first-order integer-valued autoregressive process presented by [21] based on their generalized random coefficient thinning operator. As is known to all, the EL method has been widely used in many different situations since it was introduced by [23] and [24], because this nonparametric inference idea could produce confidence regions with shape and orientation determined entirely by the data and Bartlett correction, as well as conduct the parameter-free asymptotic distribution of an EL test statistic. Contributions to the application of EL method to integer-valued processes include, among others, [25–29].

The layout of this paper is as follows. In Sections 2, we introduce the model and some distributional properties. In Section 3, we investigate the EL method for the parameters of interest. In Section 4, we report the simulation results. In Section 5, we give a real data analysis based on the daily numbers of confirmed COVID-19 cases imported from abroad in China. All proofs are postponed to the Appendix.

2 Modelling and Distributional Properties

The first-order generalized random coefficient integer-valued autoregressive (GRCINAR(1)) process proposed by [21], could be expressed by the following recursive equation:

$$X_t = \phi_t \circ^G X_{t-1} + \varepsilon_t, \quad t \geq 1, \tag{3}$$

where

- i) $\{\phi_t, t \geq 1\}$ is an i.i.d. sequence with finite mean $\phi = E(\phi_t)$, variance $\sigma_1^2 = \text{Var}(\phi_t)$ and cumulative distribution function P_{ϕ_1}.
- ii) $\{\varepsilon_t, t \geq 1\}$ is an i.i.d. nonnegative integer-valued sequence with finite mean $\lambda = E(\varepsilon_t)$, variance $\sigma_2^2 = \text{Var}(\varepsilon_t)$ and probability mass function f_ε.
- iii) The generalized thinning operator is defined by

$$\phi_t \circ^G X_{t-1}|\phi_t, X_{t-1} \sim G(\phi_t X_{t-1}, \delta_t X_{t-1}),$$

  in which $G(\phi_t X_{t-1}, \delta_t X_{t-1})$ represents a given discrete type distribution with mean $\phi_t X_{t-1}$ and variance $\delta_t X_{t-1}$, respectively.
- iv) X_0, $\{\phi_t, t \geq 1\}$ and $\{\varepsilon_t, t \geq 1\}$ are independent.

Remark 2.1 From the definition, we know that

$$\text{Var}(\phi_t \circ^G X_{t-1}|\phi_t, X_{t-1}) = \delta_t X_{t-1},$$

in which δ_t is possibly dependent on ϕ_t and X_{t-1}. However, following [21], we also assume that δ_t only depends on ϕ_t for simplicity.

Remark 2.2 EL method for random coefficient INAR(1) process based on binomial thinning operator has been considered by [25]. This kind of model is a special case of GRCINAR(1) process with binomial distribution G and $\delta_t = \phi_t(1 - \phi_t)$. However, it is known that GRCINAR(1) process also includes many other models, such as INAR(1) process based on negative binomial thinning operator and Poisson thinning operator, which respectively correspond to the situations that G is negative binomial distribution with $\delta_t = \phi_t(1 + \phi_t)$, and G is Poisson distribution with $\delta_t = \phi_t$. In this paper, we extend the results of [25] to general GRCINAR(1) process. Besides, we also discuss the forecasting problem for GRCINAR(1) process.

The following propositions present several important distributional properties of GRCINAR(1) process established by [21].

Proposition 2.3 *The GRCINAR(1) process $\{X_t, t \geq 1\}$ is a Markov chain on $\{0, 1, \cdots\}$ with the transition probabilities*

$$\begin{aligned} P_{ij} &= P(X_t = j|X_{t-1} = i) \\ &= P(\phi_t \circ^G X_{t-1} + \varepsilon_t = j|X_{t-1} = i) \\ &= \sum_{k=\kappa_1(t)}^{\kappa_2(t)} P(V_t = j - k)P(\varepsilon_t = k), \end{aligned}$$

in which $V_t|\phi_t \sim G(\phi_t i, \delta_t i)$, while $\kappa_1(t)$ and $\kappa_2(t)$ are determined by the distributions of V_t and ε_t, respectively.

Proposition 2.4 *For any $t \geq 1$, we have*

1) $E(X_t|X_{t-1}) = \phi X_{t-1} + \lambda$;

2) $\mathrm{Var}(X_t|X_{t-1}) = \sigma_1^2 X_{t-1}^2 + E(\delta_t)X_{t-1} + \sigma_2^2$;

3) $\mathrm{Cov}(X_{t+h}, X_t) = \phi^h \mathrm{Var}(X_t),\ h \geq 0$.

Proposition 2.5 *If $0 < \sigma_1^2 + \phi^2 < 1$, then there is a unique non-negative integer-valued weakly stationary process $\{X_t, t \geq 1\}$ that satisfies* (3).

Remark 2.6 By the same methods used in [12] and [19], we can also prove that the process stated in Proposition 2.5 is strictly stationary and ergodic. We omit the details here for brevity.

3 EL Inference for the GRCINAR(1) Process

In the GRCINAR(1) process, let $\boldsymbol{\theta} = (\phi, \lambda)^{\mathrm{T}}$ be the unknown parameters we are interested in, and their true values are denoted by $\boldsymbol{\theta}_0 = (\phi_0, \lambda_0)^{\mathrm{T}}$. Based on the observations $X_0, X_1, \cdots, X_n$, we consider three aspects of EL method: Estimate, confidence regions and test. To this end, we assume that the following two conditions holds throughout the rest of the paper:

(C1) $\{X_t, t \geq 0\}$ is a strictly stationary and ergodic process;

(C2) $E(X_t)^4 < +\infty,\ t \geq 0$.

For the process (3), define the conditional least squares (CLS) criterion function as

$$S(\boldsymbol{\theta}) = \sum_{t=1}^{n}[X_t - E(X_t|X_{t-1})]^2 = \sum_{t=1}^{n}(X_t - \phi X_{t-1} - \lambda)^2, \tag{4}$$

which leads to an estimating equation by taking derivative

$$-\frac{1}{2} \times \frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \boldsymbol{m}_t(\boldsymbol{\theta}) = 0,$$

where

$$\boldsymbol{m}_t(\boldsymbol{\theta}) = (m_{1t}(\boldsymbol{\theta}), m_{2t}(\boldsymbol{\theta}))^{\mathrm{T}}$$

with

$$m_{1t}(\boldsymbol{\theta}) = X_{t-1}(X_t - \phi X_{t-1} - \lambda), \quad m_{2t}(\boldsymbol{\theta}) = X_t - \phi X_{t-1} - \lambda.$$

It is obvious that the solution to (4) yields the CLS estimators of $\boldsymbol{\theta}$:

$$\widehat{\phi}_{CLS} = \frac{n\sum_{t=1}^{n} X_t X_{t-1} - \sum_{t=1}^{n} X_t \sum_{t=1}^{n} X_{t-1}}{n\sum_{t=1}^{n} X_{t-1}^2 - (\sum_{t=1}^{n} X_{t-1})^2}, \quad \widehat{\lambda}_{CLS} = \frac{1}{n}\left(\sum_{t=1}^{n} X_t - \widehat{\phi}\sum_{t=1}^{n} X_{t-1}\right).$$

Following the Remark in Section 4 of [21], or according to [30], we can have

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}_0) \xrightarrow{L} N(\mathbf{0}, \boldsymbol{V}^{-1}\boldsymbol{W}(\boldsymbol{\theta}_0)\boldsymbol{V}^{-1}), \quad n \longrightarrow +\infty,$$

in which

$$\boldsymbol{W}(\boldsymbol{\theta}_0) = \begin{pmatrix} E[X_0^2(X_1 - \phi_0 X_0 - \lambda_0)^2] & E[X_0(X_1 - \phi_0 X_0 - \lambda_0)^2] \\ E[X_0(X_1 - \phi_0 X_0 - \lambda_0)^2] & E[(X_1 - \phi_0 X_0 - \lambda_0)^2] \end{pmatrix}, \tag{5}$$

$$\boldsymbol{V} = \begin{pmatrix} E(X_0^2) & E(X_0) \\ E(X_0) & 1 \end{pmatrix}. \tag{6}$$

Therefore, a $100(1-\alpha)\%$ normal approximation (NA) confidence region of $\boldsymbol{\theta}$ based on CLS method would be constructed as follows:

$$C_{\alpha,n}^{CLS} = \{\boldsymbol{\theta} | n(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta})^{\mathrm{T}} \widehat{\boldsymbol{V}}^{\mathrm{T}} \widehat{\boldsymbol{W}}^{-1}(\widehat{\boldsymbol{\theta}}_{CLS}) \widehat{\boldsymbol{V}} (\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}) \leq c_\alpha\},$$

where $0 < \alpha < 1$, c_α satisfies $P(\chi^2(2) \geq c_\alpha) = \alpha$, and

$$\widehat{\boldsymbol{W}}(\widehat{\boldsymbol{\theta}}_{CLS}) = \frac{1}{n}\sum_{t=1}^{n} \begin{pmatrix} X_{t-1}^2(X_t - \widehat{\phi}X_{t-1} - \widehat{\lambda})^2 & X_{t-1}(X_t - \widehat{\phi}X_{t-1} - \widehat{\lambda})^2 \\ X_{t-1}(X_t - \widehat{\phi}X_{t-1} - \widehat{\lambda})^2 & (X_t - \widehat{\phi}X_{t-1} - \widehat{\lambda})^2 \end{pmatrix},$$

$$\widehat{\boldsymbol{V}} = \frac{1}{n}\sum_{t=1}^{n} \begin{pmatrix} X_{t-1}^2 & X_{t-1} \\ X_{t-1} & 1 \end{pmatrix},$$

which are the consistent estimators of $\boldsymbol{W}(\boldsymbol{\theta})$ and $\boldsymbol{V}$, respectively.

Now, we discuss the EL method. Based on the score function $\boldsymbol{m}_t(\boldsymbol{\theta})$, the profile empirical likelihood ratio (ELR) function can be given by

$$R(\boldsymbol{\theta}) = \max\left\{\prod_{t=1}^{n} np_t : p_t \geq 0,\ \sum_{t=1}^{n} p_t = 1,\ \sum_{t=1}^{n} p_t \boldsymbol{m}_t(\boldsymbol{\theta}) = 0\right\}.$$

By the standard Lagrange multiplier approach, we may verify that the optimal value of p_t is found to be

$$p_t = \frac{1}{n}\frac{1}{1 + \boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{m}_t(\boldsymbol{\theta})}, \quad t = 1, 2, \cdots, n,$$

where the Lagrange multiplier $\boldsymbol{b}(\boldsymbol{\theta})$ satisfies

$$g(\boldsymbol{b}(\boldsymbol{\theta})) = \frac{1}{n}\sum_{t=1}^{n} \frac{\boldsymbol{m}_t(\boldsymbol{\theta})}{1 + \boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{m}_t(\boldsymbol{\theta})} = \mathbf{0}. \tag{7}$$

Therefore, we obtain the log ELR function as follows:

$$l(\boldsymbol{\theta}) = -2\log R(\boldsymbol{\theta}) = 2\sum_{t=1}^{n} \log(1 + \boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{m}_t(\boldsymbol{\theta})). \tag{8}$$

The limiting distribution of the ELR statistic $l(\boldsymbol{\theta})$ can be obtained consequently.

Theorem 3.1 *Under the assumptions* (C1) *and* (C2), *we have*

$$l(\boldsymbol{\theta}_0) \xrightarrow{L} \chi^2(2), \quad n \longrightarrow +\infty.$$

The confidence region of the parameters can be constructed based on the above theorem.

Theorem 3.2 *Under the assumptions* (C1) *and* (C2), *the* $100(1-\alpha)\%$ *EL confidence region of* $\boldsymbol{\theta}$ *is*

$$C_{\alpha,n}^{EL} = \{\boldsymbol{\theta} | l(\boldsymbol{\theta}) \leq c_\alpha\},$$

in which $0 < \alpha < 1$, *and* c_α *satisfies* $P(\chi^2(2) \geq c_\alpha) = \alpha$.

On the other hand, by minimizing $l(\boldsymbol{\theta})$, we can obtain $\widehat{\boldsymbol{\theta}}_{MEL}$, which is the maximum empirical likelihood estimator of $\boldsymbol{\theta}$. The following theorem gives its asymptotic property.

Theorem 3.3 *Under the assumptions* (C1) *and* (C2)*, the MEL estimator* $\widehat{\boldsymbol{\theta}}_{MEL}$ *is consistent, moreover we have*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{MEL} - \boldsymbol{\theta}_0) \stackrel{L}{\longrightarrow} N(\mathbf{0}, \boldsymbol{V}^{-1}\boldsymbol{W}(\boldsymbol{\theta}_0)\boldsymbol{V}^{-1}), \quad n \longrightarrow +\infty,$$

where $\boldsymbol{\theta}_0$ *is the true value of* $\boldsymbol{\theta}$*,* $\boldsymbol{W}(\boldsymbol{\theta}_0)$ *and* $\boldsymbol{V}$ *are defined by* (5) *and* (6)*, respectively.*

In practice, we could also consider the following hypothesis problem about $\boldsymbol{\theta}$:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{v.s.} \quad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0.$$

Based on EL method, we can get the ELR statistic as

$$Q(\boldsymbol{\theta}) = l(\boldsymbol{\theta}) - l(\widehat{\boldsymbol{\theta}}_{MEL}),$$

which has limit distribution given in the subsequent theorem.

Theorem 3.4 *Under the assumptions* (C1) *and* (C2)*, when the null hypothesis* H_0 *is true, we have*

$$Q(\boldsymbol{\theta}_0) \stackrel{L}{\longrightarrow} \chi^2(2), \quad n \longrightarrow +\infty.$$

Therefore, the rejection region for significance level α is given by

$$W_{n,\alpha} = \{Q(\boldsymbol{\theta}) \geq c_\alpha\},$$

where c_α satisfies $P(\chi^2(2) \geq c_\alpha) = \alpha$.

4 Simulation Studies

The purpose of this section is to compare the performances of CLS and MEL estimators. To this end, we take an important special case of GRCINAR(1) process as an example:

$$X_t = \phi_t \circ^G X_{t-1} + \varepsilon_t, \quad t \geq 1, \tag{9}$$

in which

$$\phi_t \circ^G X_{t-1} = \sum_{i=1}^{X_{t-1}} W_i^{(t)},$$
$$\phi_t = \phi + U_t, \quad U_t \sim U(-0.1, 0.1), \quad \varepsilon_t \sim P(\lambda),$$

and both $\{\varepsilon_t, t \geq 1\}$ and $\{U_t, t \geq 1\}$, independent of each other, are sequences of i.i.d. random variables. Furthermore, we focus on the following two models:

1) RCNBINAR(1) process: Random coefficient INAR(1) process based on negative binomial thinning operator, in which $\{W_i^{(t)}, i \geq 1\}$ is assumed to be a sequence of conditional i.i.d. geometric random variables with probability mass function

$$P(W_i^{(t)} = x|\phi_t) = \frac{\phi_t^x}{(1 + \phi_t)^{x+1}}, \quad x = 0, 1, \cdots.$$

2) RCPINAR(1) process: Random coefficient INAR(1) process based on Poisson thinning operator, in which $\{W_i^{(t)}, i \geq 1\}$ is assumed to be a sequence of conditional i.i.d. Poisson random variables with probability mass function

$$P(W_i^{(t)} = x|\phi_t) = \frac{\phi_t^x}{x!}\mathrm{e}^{-\phi_t}, \quad x = 0, 1, \cdots.$$

In each simulation, we generate the data with $X_0 = 1$, and perform all the calculations based on 1000 replications under the following 4 scenarios:

(i) $(\phi, \lambda) = (0.3, 1)$; (ii) $(\phi, \lambda) = (0.3, 2)$; (iii) $(\phi, \lambda) = (0.7, 1)$; (iv) $(\phi, \lambda) = (0.7, 2)$.

4.1 Estimate

For the generated samples with size n chosen to be 50, 100, 300 and 1000, we calculate the average estimates across m repetitions, and evaluate the effectiveness of CLS and MEL estimators by applying the empirical bias (BIAS) and the mean squared errors (MSE), which are defined for parameter $\boldsymbol{\theta}$ as

$$\mathrm{BIAS} = \frac{1}{m}\sum_{k=1}^{m}\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0, \quad \mathrm{MSE} = \frac{1}{m}\sum_{k=1}^{m}(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)^2,$$

where $m = 1000$ is the replication times, $\widehat{\boldsymbol{\theta}}_k$ is the estimator of $\boldsymbol{\theta}$ at the kth replication, and $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$.

In addition, we also use the conditional maximum likelihood (CML) method as a benchmark to estimate the parameters. The CML estimators of $\boldsymbol{\theta}$ can be obtained by maximizing the conditional log-likelihood function

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

where for RCNBINAR(1) process, we have

$$P(X_t = x_t|X_{t-1} = x_{t-1}) = \sum_{k=0}^{x_t}\frac{\lambda^{x_t-k}}{(x_t-k)!}\mathrm{e}^{-\lambda}\binom{x_t+k-1}{k}\int_{-0.1}^{+0.1}\frac{5(\phi+u)^k}{(1+\phi+u)^{x_{t-1}+k-1}}du,$$

and for RCPINAR(1) process, we have

$$P(X_t = x_t|X_{t-1} = x_{t-1}) = \sum_{k=0}^{x_t}\frac{\lambda^{x_t-k}}{(x_t-k)!}\mathrm{e}^{-\lambda}\int_{-0.1}^{+0.1}\frac{5(\phi+u)^k x_{t-1}^k}{k!}\mathrm{e}^{-(\phi+u)x_{t-1}}du.$$

We give the summary of the simulation results in Tables 1–8. It can be seen that the CML method has the best performance in terms of BIAS and MSE under the correct model. However, as nonparametric methods, CLS and MEL are also efficient methods because they produce good estimators which can be overall comparable to the CML results. Besides, MEL estimators have less BIAS and smaller MSE than CLS estimators in most cases, showing that MEL is a competitive choice compared with CLS method. On the other, both BIAS and MSE of the three methods gradually decrease as expected, when the sample size increases, implying that the estimators are consistent for all the parameters.

Table 1 Simulation results for $(\phi, \lambda) = (0.3, 1)$ in RCNBINAR(1) process

n	Parameter	CLS			MEL			CLM		
		Estimate	BIAS	MES	Estimate	BIAS	MSE	Estimate	BIAS	MSE
50	$\phi = 0.3$	0.2447	−0.0553	0.0259	0.2576	−0.0424	0.0210	0.2628	−0.0372	0.0171
	$\lambda = 1$	1.0649	0.0649	0.0685	1.0528	0.0528	1.0459	0.0459	−0.0424	0.0501
100	$\phi = 0.3$	0.2687	−0.0313	0.0128	0.2752	−0.0248	0.0126	0.2768	−0.0232	0.0103
	$\lambda = 1$	1.0407	0.0407	0.0334	1.0319	0.0319	0.0341	1.0306	0.0306	0.0262
300	$\phi = 0.3$	0.2862	−0.0138	0.0043	0.2884	−0.0116	0.0041	0.2888	−0.0112	0.0036
	$\lambda = 1$	1.0181	0.0181	0.0095	1.0125	0.0125	0.0104	1.0123	0.0123	0.0092
1000	$\phi = 0.3$	0.2967	−0.0033	0.0012	0.2969	−0.0031	0.0012	0.2972	−0.0028	0.0010
	$\lambda = 1$	1.0059	0.0059	0.0030	1.0042	0.0042	0.0029	1.0038	0.0038	0.0025

Table 2 Simulation results for $(\phi, \lambda) = (0.3, 2)$ in RCNBINAR(1) process

n	Parameter	CLS			MEL			CLM		
		Estimate	BIAS	MES	Estimate	BIAS	MSE	Estimate	BIAS	MSE
50	$\phi = 0.3$	0.2514	−0.0486	0.0257	0.2592	−0.0408	0.0201	0.2624	−0.0376	0.0159
	$\lambda = 2$	2.1327	0.1327	0.2488	2.1063	0.1063	0.1760	2.1003	0.1003	0.1478
100	$\phi = 0.3$	0.2750	−0.0250	0.0127	0.2783	−0.0217	0.0114	0.2791	−0.0209	0.0093
	$\lambda = 2$	2.0686	0.0686	0.1191	2.0505	0.0505	0.0968	2.0502	0.0502	0.0796
300	$\phi = 0.3$	0.2904	−0.0096	0.0035	0.2929	−0.0071	0.0036	0.2935	−0.0065	0.0032
	$\lambda = 2$	2.0257	0.0257	0.0311	2.0192	0.0192	0.0319	2.0289	0.0289	0.0269
1000	$\phi = 0.3$	0.2982	−0.0018	0.0011	0.2975	−0.0025	0.0010	0.2976	−0.0024	0.0009
	$\lambda = 2$	2.0083	0.0083	0.0106	2.0031	0.0031	0.0091	2.0030	0.0030	0.0079

Table 3 Simulation results for $(\phi, \lambda) = (0.7, 1)$ in RCNBINAR(1) process

n	Parameter	CLS			MEL			CLM		
		Estimate	BIAS	MES	Estimate	BIAS	MSE	Estimate	BIAS	MSE
50	$\phi = 0.7$	0.6000	−0.1000	0.0302	0.5987	−0.1013	0.0291	0.6298	−0.0702	0.0217
	$\lambda = 1$	1.2704	0.2704	0.3074	1.2754	0.2754	0.2606	1.1633	0.1633	0.1576
100	$\phi = 0.7$	0.6401	−0.0599	0.0137	0.6428	−0.0572	0.0126	0.6656	−0.0344	0.0085
	$\lambda = 1$	1.1595	0.1595	0.1163	1.1482	0.1482	0.1211	1.0777	0.0777	0.0546
300	$\phi = 0.7$	0.6774	−0.0226	0.0042	0.6774	−0.0226	0.0039	0.6876	−0.0124	0.0022
	$\lambda = 1$	1.0690	0.0690	0.0352	1.0578	0.0578	0.0317	1.0249	0.0249	0.0144
1000	$\phi = 0.7$	0.6937	−0.0063	0.0012	0.6949	−0.0051	0.0012	0.6974	−0.0026	0.0007
	$\lambda = 1$	1.0172	0.0172	0.0108	1.0131	0.0131	0.0101	1.0056	0.0056	0.0041

Table 4 Simulation results for $(\phi, \lambda) = (0.7, 2)$ in RCNBINAR(1) process

n	Parameter	CLS			MEL			CLM		
		Estimate	BIAS	MES	Estimate	BIAS	MSE	Estimate	BIAS	MSE
50	$\phi = 0.7$	0.6141	−0.0859	0.0254	0.6205	−0.0795	0.0214	0.6382	−0.0618	0.0167
	$\lambda = 2$	2.5056	0.5056	1.1108	2.4653	0.4653	0.8145	2.3388	0.3388	0.5843
100	$\phi = 0.7$	0.6581	−0.0419	0.0100	0.6561	−0.0439	0.0094	0.6663	−0.0337	0.0069
	$\lambda = 2$	2.2430	0.2430	0.3779	2.2518	0.2518	0.3566	2.1822	0.1822	0.2388
300	$\phi = 0.7$	0.6824	−0.0176	0.0030	0.6854	−0.0146	0.0031	0.6897	−0.0103	0.0021
	$\lambda = 2$	2.1056	0.1056	0.1122	2.0837	0.0837	0.1080	2.0546	0.0546	0.0665
1000	$\phi = 0.7$	0.6962	−0.0038	0.0009	0.6962	−0.0038	0.0009	0.6978	−0.0022	0.0005
	$\lambda = 2$	2.0197	0.0197	0.0317	2.0197	0.0197	0.0334	2.0106	0.0106	0.0165

Table 5 Simulation results for $(\phi, \lambda) = (0.3, 1)$ in RCPINAR(1) process

n	Parameter	CLS			MEL			CLM		
		Estimate	BIAS	MES	Estimate	BIAS	MSE	Estimate	BIAS	MSE
50	$\phi = 0.3$	0.2462	−0.0538	0.0250	0.2606	−0.0394	0.0216	0.2749	−0.0251	0.0169
	$\lambda = 1$	1.0626	0.0626	0.0653	1.0512	0.0512	0.0641	1.0332	0.0332	0.0564
100	$\phi = 0.3$	0.2711	−0.0289	0.0116	0.2770	−0.0230	0.0122	0.2816	−0.0184	0.0104
	$\lambda = 1$	1.0344	0.0344	0.0310	1.0268	0.0268	0.0301	1.0209	0.0209	0.0270
300	$\phi = 0.3$	0.2908	−0.0092	0.0039	0.2942	−0.0058	0.0040	0.2948	−0.0052	0.0036
	$\lambda = 1$	1.0092	0.0092	0.0099	1.0050	0.0050	0.0102	1.0042	0.0042	0.0094
1000	$\phi = 0.3$	0.2976	−0.0024	0.0012	0.2954	−0.0046	0.0012	0.2958	−0.0042	0.0011
	$\lambda = 1$	1.0019	0.0019	0.0031	1.0042	0.0042	0.0030	1.0036	0.0036	0.0028

Table 6 Simulation results for $(\phi, \lambda) = (0.3, 2)$ in RCPINAR(1) process

n	Parameter	CLS			MEL			CLM		
		Estimate	BIAS	MES	Estimate	BIAS	MSE	Estimate	BIAS	MSE
50	$\phi = 0.3$	0.2456	−0.0544	0.0248	0.2610	−0.0390	0.0205	0.2715	−0.0285	0.0158
	$\lambda = 2$	2.1458	0.1458	0.2499	2.1068	0.1068	0.1911	2.0799	0.0799	0.1609
100	$\phi = 0.3$	0.2813	−0.0187	0.0109	0.2771	−0.0229	0.0106	0.2818	−0.0182	0.0097
	$\lambda = 2$	2.0530	0.0530	0.1024	2.0512	0.0512	0.0976	2.0455	0.0455	0.0921
300	$\phi = 0.3$	0.2927	−0.0073	0.0035	0.2965	−0.0035	0.0045	0.2973	−0.0027	0.0033
	$\lambda = 2$	2.0211	0.0211	0.0349	2.0257	0.0257	0.0302	2.0229	0.0229	0.0292
1000	$\phi = 0.3$	0.3002	0.0002	0.0010	0.2966	−0.0034	0.0010	0.2981	−0.0019	0.0010
	$\lambda = 2$	1.9992	−0.0008	0.0096	2.0070	0.0070	0.0090	2.0082	0.0082	0.0087

Table 7 Simulation results for $(\phi, \lambda) = (0.7, 1)$ in RCPINAR(1) process

n	Parameter	CLS			MEL			CLM		
		Estimate	BIAS	MES	Estimate	BIAS	MSE	Estimate	BIAS	MSE
50	$\phi = 0.7$	0.6196	−0.0804	0.0246	0.6203	−0.0797	0.0239	0.6372	−0.0628	0.0197
	$\lambda = 1$	1.2272	0.2272	0.2661	1.2132	0.2132	0.2363	1.1639	0.1639	0.1822
100	$\phi = 0.7$	0.6580	−0.0420	0.0100	0.6573	−0.0427	0.0099	0.6717	−0.0283	0.0076
	$\lambda = 1$	1.1243	0.1243	0.1056	1.1232	0.1232	0.0994	1.0761	0.0761	0.0743
300	$\phi = 0.7$	0.6842	−0.0158	0.0030	0.6850	−0.0150	0.0030	0.6893	−0.0107	0.0022
	$\lambda = 1$	1.0445	0.0445	0.0285	1.0433	0.0433	0.0272	1.0305	0.0305	0.0193
1000	$\phi = 0.7$	0.6948	−0.0052	0.0008	0.6945	−0.0055	0.0009	0.6964	−0.0036	0.0006
	$\lambda = 1$	1.0164	0.0164	0.0077	1.0150	0.0150	0.0080	1.0085	0.0085	0.0051

Table 8 Simulation results for $(\phi, \lambda) = (0.7, 2)$ in RCPINAR(1) process

n	Parameter	CLS			MEL			CLM		
		Estimate	BIAS	MES	Estimate	BIAS	MSE	Estimate	BIAS	MSE
50	$\phi = 0.7$	0.6248	−0.0752	0.0248	0.6312	−0.0689	0.0179	0.6408	−0.0592	0.0156
	$\lambda = 2$	2.4707	0.4707	1.1547	2.4124	0.4124	0.7133	2.3504	0.3504	0.6157
100	$\phi = 0.7$	0.6577	−0.0423	0.0103	0.6634	−0.0366	0.0078	0.6701	−0.0299	0.0066
	$\lambda = 2$	2.2634	0.2634	0.4431	2.2405	0.2405	0.3309	2.1959	0.1959	0.2745
300	$\phi = 0.7$	0.6871	−0.0129	0.0026	0.6842	−0.0158	0.0024	0.6872	−0.0128	0.0020
	$\lambda = 2$	2.0756	0.0756	0.1069	2.0963	0.0963	0.0990	2.0660	0.0660	0.0772
1000	$\phi = 0.7$	0.6952	−0.0048	0.0007	0.6961	−0.0039	0.0007	0.6974	−0.0026	0.0006
	$\lambda = 2$	2.0281	0.0281	0.0275	2.0242	0.0242	0.0272	2.0173	0.0173	0.0217

4.2 Confidence Region

In what follows, we conduct a simulation to compare the CLS and EL methods by overage probabilities of the confidence regions based on $C_{\alpha,n}^{CLS}$ and $C_{\alpha,n}^{EL}$, respectively. For both of the above two models under Scenarios 1–4, the sample size n is chosen to be 50, 100, 300 and 1000, and the coverage probabilities are calculated based on 1000 replications at confidence levels 0.90 and 0.95, respectively.

From Tables 9 and 10, we can see that coverage probabilities of the two methods increase and coverage to the confidence levels as the sample size n increases. By comparison, the EL confidence regions have larger and more accurate coverage probabilities than those of the CLS ones, especially for small sample sizes. As a result, we conclude that the EL method performs better in terms of coverage accuracy.

Table 9 The coverage probabilities of confidence regions for RCNBINAR(1) process

Nominal	Parameter	$n = 50$		$n = 100$		$n = 300$		$n = 1000$	
Levels	(ϕ, λ)	El	CLS	EL	CLS	EL	CLS	EL	CLS
0.90	$(0.3, 1)$	0.854	0.818	0.875	0.862	0.892	0.884	0.900	0.900
	$(0.3, 2)$	0.836	0.810	0.876	0.847	0.904	0.904	0.889	0.889
	$(0.7, 1)$	0.801	0.770	0.847	0.822	0.861	0.855	0.896	0.897
	$(0.7, 2)$	0.820	0.782	0.856	0.845	0.891	0.881	0.889	0.887
0.95	$(0.3, 1)$	0.902	0.874	0.933	0.919	0.929	0.929	0.943	0.942
	$(0.3, 2)$	0.898	0.873	0.935	0.917	0.946	0.943	0.947	0.945
	$(0.7, 1)$	0.861	0.846	0.898	0.889	0.924	0.918	0.946	0.945
	$(0.7, 2)$	0.897	0.873	0.927	0.913	0.939	0.939	0.953	0.947

Table 10 The coverage probabilities of confidence regions for RCPINAR(1) process

Nominal	Parameter	$n = 50$		$n = 100$		$n = 300$		$n = 1000$	
Levels	(ϕ, λ)	El	CLS	EL	CLS	EL	CLS	EL	CLS
0.90	$(0.3, 1)$	0.857	0.837	0.879	0.854	0.903	0.893	0.890	0.888
	$(0.3, 2)$	0.849	0.813	0.874	0.866	0.901	0.897	0.897	0.895
	$(0.7, 1)$	0.824	0.792	0.861	0.847	0.885	0.881	0.911	0.910
	$(0.7, 2)$	0.828	0.786	0.868	0.835	0.878	0.875	0.917	0.910
0.95	$(0.3, 1)$	0.922	0.891	0.940	0.923	0.950	0.942	0.938	0.935
	$(0.3, 2)$	0.904	0.883	0.930	0.920	0.945	0.940	0.953	0.953
	$(0.7, 1)$	0.888	0.859	0.923	0.908	0.935	0.927	0.957	0.957
	$(0.7, 2)$	0.890	0.859	0.911	0.891	0.938	0.932	0.956	0.954

4.3 Performances of the EL Test

In this subsection, we investigate to illustrate the empirical size and power of the EL test. To this end, we consider the following hypothesis testing problem that could occur in practical applications:

$$H_0 : \phi = 0.5 \quad \text{v.s.} \quad H_1 : \phi \neq 0.5.$$

We select the sample size n as 50, 100, 200 and 300, and calculate the results based on 1000 replications at significance levels 0.90 and 0.95, respectively. Specifically, we compare the empirical likelihood ratio statistic with the critical value, to get the observed percentage of rejecting the null hypothesis. In particular, as to study the power, we generate data from RCNBINAR(1) process and RCPINAR(1) process under Scenarios 1–4, but estimate them by restricting $\phi = 0.5$, i.e, under the null hypothesis.

From the results summed up in Tables 9–12, we find that the sizes of the EL test decrease to the corresponding significance level and the powers of the EL test increase to 1 as the sample size n increase, indicating that the EL method is reasonable and effective.

Table 11 Empirical size and power of the EL test for RCNBINAR(1) process

Significance Levels	Parameter	Size				Power			
		$n=50$	$n=100$	$n=200$	$n=300$	$n=50$	$n=100$	$n=200$	$n=300$
0.10	$(0.3,1)$	0.156	0.131	0.119	0.106	0.537	0.787	0.956	0.994
	$(0.3,2)$	0.138	0.127	0.129	0.100	0.741	0.945	0.999	1.000
	$(0.7,1)$	0.217	0.154	0.121	0.103	0.601	0.815	0.970	0.993
	$(0.7,2)$	0.156	0.134	0.125	0.100	0.819	0.961	0.999	1.000
0.05	$(0.3,1)$	0.102	0.084	0.066	0.060	0.434	0.682	0.914	0.991
	$(0.3,2)$	0.078	0.079	0.066	0.045	0.623	0.892	0.997	1.000
	$(0.7,1)$	0.124	0.090	0.060	0.060	0.522	0.744	0.956	0.990
	$(0.7,2)$	0.102	0.094	0.055	0.050	0.741	0.934	0.999	1.000

Table 12 Empirical size and power of the EL test for RCPINAR(1) process

Significance Levels	Parameter	Size				Power			
		$n=50$	$n=100$	$n=200$	$n=300$	$n=50$	$n=100$	$n=200$	$n=300$
0.10	$(0.3,1)$	0.145	0.142	0.120	0.108	0.583	0.811	0.981	0.998
	$(0.3,2)$	0.128	0.122	0.108	0.103	0.779	0.965	1.000	1.000
	$(0.7,1)$	0.181	0.155	0.101	0.107	0.728	0.919	0.990	0.999
	$(0.7,2)$	0.146	0.122	0.116	0.106	0.897	0.993	1.000	1.000
0.05	$(0.3,1)$	0.099	0.073	0.066	0.065	0.451	0.701	0.956	0.996
	$(0.3,2)$	0.076	0.064	0.062	0.053	0.676	0.935	1.000	1.000
	$(0.7,1)$	0.115	0.079	0.067	0.061	0.658	0.975	0.986	0.999
	$(0.7,2)$	0.079	0.073	0.057	0.058	0.885	0.984	1.000	1.000

5 Real Data Analysis

In this section, we apply GRCINAR(1) process and EL method to fit the real life situations. For this purpose, we consider a data set that shows the daily numbers of new confirmed COVID-19 cases imported from abroad in China. The data set is reported by National Health Commission, and consists of a total of 546 observations starting from Mar. 4th 2020 to Aug. 31st, 2021. The sample path, ACF and PACF of the series are given in Figure 1, from which we can reasonably assume that these data come from INAR(1) process.

Except for RCNBINAR(1) process and RCPINAR(1) process, we also use the following models to fit the data:

3) RCBINAR(1) process: Random coefficient INAR(1) process based on binomial thinning operator, in which $\{W_i^{(t)}, i \geq 1\}$ in (9) is assumed to be a sequence of conditional i.i.d. Bernoulli random variables with probability mass function

$$P(W_i^{(t)}=1|\phi_t)=1-P(W_i^{(t)}=0|\phi_t)=\phi_t.$$

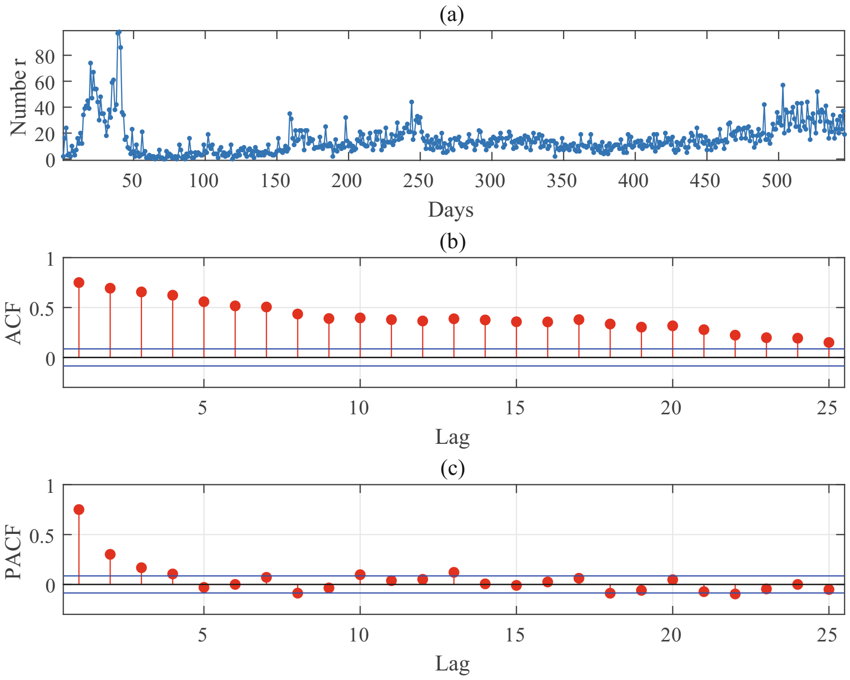

Figure 1 The sample path, ACF and PACF of the imported cases of COVID-19 in China

4) NBINAR(1) process: Reduced RCNBINAR(1) process with $\phi_t = \phi$.

5) PINAR(1) process: Reduced RCPINAR(1) process with $\phi_t = \phi$.

6) BINAR(1) process: Reduced RCBINAR(1) process with $\phi_t = \phi$.

In addition, it should be noted that we need not specify the distribution of the error term ε_t, when CLS method and EL method are used during the discussion. Moreover, we consider mean absolute deviation error (MADE) for in-sample prediction (see [31]) to compare different estimation methods. This kind of criteria is defined as

$$\text{MADE} = \frac{1}{m}\sum_{i=1}^{m}|\widehat{X}_{n-m+i} - X_{n-m+i}|,$$

in which

$$\widehat{X}_{n-m+i} = E(X_{n-m+i}|X_{n-m}) = \widehat{\phi}^i X_{n-m} + \frac{1-\widehat{\phi}^i}{1-\widehat{\phi}}\widehat{\lambda}. \tag{10}$$

Table 13 reports the point estimates of the model parameters and the MADE for prediction of last m observations of the daily numbers of new confirmed COVID-19 cases imported from abroad in China. It is observed that EL method has better performance, since it leads to smaller MADE than CLS method.

Table 13 The results of estimates and MADE

Method	Estimate		MADE		
	ϕ	λ	$m=5$	$m=10$	$m=30$
CLS	0.7472	3.8824	6.9203	8.3937	12.2080
EL	0.7503	3.8359	6.7760	8.2801	12.1056

In what follows, we perform an out-of-sample experiment to further compare the aforementioned 6 models. Specifically, we use all the 546 observations of the COVID-19 data set to estimate the parameters, and then obtain the coherent predictions for the observations from Sep. 1st to Sep. 5th, 2021. It is widely known that conditional expectation (CE) given by (10) is one of the most common approach to construct forecasts. However, CE will produce the same predictors for all the 6 models, so that we can not distinguish them from each other. Besides, CE falls short of preserving the integer-valued nature in generating forecasts for count data time series. Therefore, we apply the model-based INAR bootstrap (MBB) technique proposed by [32] to achieve our goal. The algorithm steps of this method (taking (9) as an example) are the following:

Step 1 Estimate the model parameters ϕ by EL method.

Step 2 Compute the residuals

$$\widehat{\varepsilon}_t = X_t - \phi_t \circ^G X_{t-1}, \quad t = 2, 3, \cdots, n,$$

where $\phi_t = \widehat{\phi} + U_t$ and $\{U_t\}$ is a sequence of i.i.d. random variables drawn from $U(-0.1, 0.1)$.

Step 3 Construct the empirical distribution $\widehat{F}_\varepsilon$ of the modified residuals $\widetilde{\varepsilon}_t$ defined by

$$\widetilde{\varepsilon}_t = \begin{cases} \widehat{\varepsilon}_t, & \text{if } \widehat{\varepsilon}_t \geq 0, \\ 0, & \text{if } \widehat{\varepsilon}_t < 0, \quad t = 2, 3, \cdots, n. \end{cases}$$

Step 4 For $b = 1, 2, \cdots, B$, define the bootstrapped series X_t^b by

$$X_t^b = \phi_t^b \circ^G X_{t-1} + \varepsilon_t^b, \quad t = 2, 3, \cdots, n,$$

in which $\phi_t^b = \widehat{\phi} + U_t^b$, while $\{U_t^b\}$ and $\{\varepsilon_t^b\}$ are i.i.d. random samples from $U(-0.1, 0.1)$ and $\widehat{F}_\varepsilon$, respectively.

Step 5 Based on $\{X_1^b, X_2^b, \cdots, X_n^b\}$, compute the estimators $\widehat{\phi}^b$ of ϕ as in Step 1.

Step 6 Denote H the largest prediction horizon, and compute the forecasts as

$$X_{n+h}^b = \phi_{n+h}^b \circ^G X_{n+h-1}^b + \varepsilon_{n+h}^b, \quad h = 1, 2, \cdots, H,$$

in which $X_n^b = X_n$, $\phi_{n+h}^b = \widehat{\phi} + U_{n+h}^b$, while $\{U_{n+h}^b, h = 1, 2, \cdots, H\}$ and $\{\varepsilon_{n+h}^b, h = 1, 2, \cdots, H\}$ are i.i.d. random samples from $U(-0.1, 0.1)$ and $\widehat{F}_\varepsilon$, respectively.

Step 7 From the replicates $\{X_{n+h}^1, X_{n+h}^2, \cdots, X_{n+h}^B\}$, obtain the point forecast $\widehat{X}_{n+h}$ of X_{n+h} $(h = 1, 2, \cdots, H)$ by considering the median of $\widehat{F}_{X_{n+h}}$.

For our analysis, we set $B = 1001$ to select the final model. Furthermore, the forecast mean absolute error (FMAE) statistic of h-step-ahead forecasts, that is used to assess the forecasting performance, is defined by

$$\text{FMAE} = \frac{1}{H}\sum_{h=1}^{H} |\widehat{X}_{n+h} - X_{n+h}|,$$

where $n = 546$ and $H = 5$. Figure 2 illustrates the forecasts of the 6 models based on MBB method, showing the prediction differences among them. Furthermore, from the results reported in Table 14, it is evident that MBB works better than CE, and we could state the superiority of RCNBINAR(1) process to the other 5 models in view of its lowest FMAE.

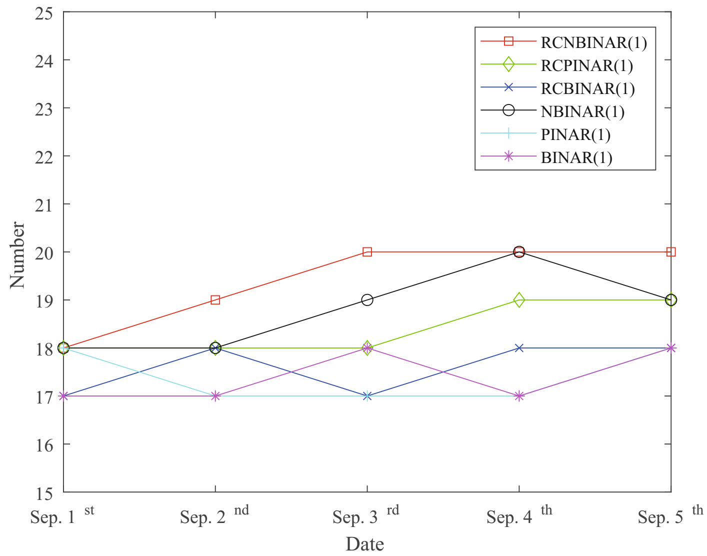

Figure 2 Forecasts of the 6 models based on MBB method

Table 14 Forecasting performance results

Date	Sep. 1^{st}	Sep. 2^{nd}	Sep. 3^{rd}	Sep. 4^{th}	Sep. 5^{th}	FMAE
Observation	27	28	27	28	18	-
Forecast-CE	18.1093	17.4406	16.9386	16.5618	16.2789	8.5342
Forecast-MBB for RCNBINAR(1)	18	19	20	20	20	7.0000
Forecast-MBB for RCPINAR(1)	18	18	18	19	19	7.6000
Forecast-MBB for RCBINAR(1)	17	18	17	18	18	8.0000
Forecast-MBB for NBINAR(1)	18	18	19	20	19	7.2000
Forecast-MBB for PINAR(1)	18	17	17	17	18	8.2000
Forecast-MBB for BINAR(1)	17	17	18	17	18	8.2000

6 Concluding Remarks

In this paper, we studied the nonparametric statistical inference for a GRCINAR(1) process by the EL method. Based on the ELR statistic and its limiting distribution, we give the EL confidence regions and the MEL estimators of the model parameters, as well as their asymptotic properties. The simulation results and a real data example illustrate the good performance of the EL method for the model, and support the recommendation to use GRCINAR(1)) process effectively in some practical applications. As for the future work, we could consider the estimation of σ_1^2, the variance of ϕ_t, and discuss the test for randomness of the coefficient by EL method. Other potential issues include investigating the observation-driven or covariate-driven GRCINAR(1) process and threshold GRCINAR(1) process, as well as higher-order GRCINAR process to improve the accuracy of prediction for real data.

References

[1] Weiß C H, Thinning operations for modeling time series of counts — A survey, *AStA Advances in Statistical Analysis*, 2008, **92**(3): 319–343.

[2] Scotto M G, Weiß C H, and Gouveia S, Thinning-based models in the analysis of integer-valued time series: A review, *Statistical Modelling*, 2015, **15**(6): 590–618.

[3] Weiß C H, *An Introduction to Discrete-Valued Time Series*, John Wiley and Sons Ltd, New Jersey, 2018.

[4] Weiß C H, Stationary count time series models, *WIREs Computational Statistics*, 2021, **13**(1): e1502.

[5] McKenzie E, Some simple models for discrete variate time series, *Journal of the American Water Resources Association*, 1985, **21**(4): 645–650.

[6] Al-Osh M A and Alzaid A A, First-order integer-valued autoregressive (INAR(1)) process, *Journal of Time Series Analysis*, 1987, **8**(3): 261–275.

[7] Steutel F W and van Harn K, Discrete analogues of self-decomposability and stability, *The Annals of Probability*, 1979, **7**(5): 893–899.

[8] Ristić M M, Bakouch H S, and Nastić A S, A new geometric first-order integer-valued autoregressive (NGINAR(1)) process, *Journal of Statistical Planning and Inference*, 2009, **139**(7): 2218–2226.

[9] Zhu R and Joe H, Negative binomial time series models based on expectation thinning operators, *Journal of Statistical Planning and Inference*, 2010, **140**(7): 1874–1888.

[10] Weiß C H, A Poisson INAR(1) model with serially dependent innovations, *Metrika*, 2015, **78**(7): 829–851.

[11] Kim H Y and Park Y, A non-stationary integer-valued autoregressive model, *Statistical Papers*, 2008, **49**(3): 485–502.

[12] Zhang H X, Wang D H, and Zhu F K, Inference for INAR(p) processes with signed generalized power series thinning operator, *Journal of Statistical Planning and Inference*, 2010, **140**(3): 667–883.

[13] Biswas A and Song P X K, Discrete-valued ARMA processes, *Statistics and Probability Letters*, 2009, **79**(17): 1884–1889.

[14] Kang Y, Wang D H, Yang K, et al., A new thinning-based INAR(1) process for underdispersed or overdispersed counts, *Journal of the Korean Statistical Society*, 2020, **49**(2): 324–349.

[15] Zhang Q C, Wang D H, and Fan X D, A negative binomial thinning-based bivariate INAR(1) process, *Statistica Neerlandica*, 2020, **74**(4): 517–537.

[16] Joe H, Time series models with univariate margins in the convolution-closed infinitely divisible class, *Journal of Applied Probability*, 1996, **33**(3): 664–677.

[17] Zheng H, Basawa I V, and Datt S, Inference for pth-order random coefficient integer-valued autoregressive processes, *Journal of Time Series Analysis*, 2006, **27**(3): 411–440.

[18] Zheng H, Basawa I V, and Datta S, First-order random coefficient integer-valued autoregressive process, *Journal of Statistical Planning and Inference*, 2007, **137**(1): 212–229.

[19] Zheng H and Basawa I V, First-order observation-driven integer-valued autoregressive processes, *Statistics and Probability Letters*, 2008, **78**(1): 1–9.

[20] Yu M J, Wang D H, and Yang K, A class of observation-driven random coefficient INAR(1) processes based on negative binomial thinning, *Journal of the Korean Statistical Society*, 2019, **48**(2): 248–264.

[21] Gomes D and Canto e Castro L, Generalized integer-valued random coefficient for a first order structure autoregressive (RCINAR) process, *Journal of Statistical Planning and Inference*, 2009, **139**(12): 4088–4097.

[22] Latour A, Existence and stochastic structure of a non-negative integer-valued autoregressive processes, *Journal of Time Series Analysis*, **19**(4): 439–455.

[23] Owen A B, Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 1988, **75**(2): 237–249.

[24] Owen A B, Empirical likelihood ratio confidence regions, *Annals of Statistics*, 1990, **18**(1): 90–120.

[25] Zhang H X, Wang D H, and Zhu F K, Empirical likelihood for first-order random coefficient integer-valued autoregressive processes, *Communications in Statistics — Theory and Methods*, 2011, **40**(3): 492–509.

[26] Zhang H X, Wang D H, and Zhu F K, Empirical likelihood inference for random coefficient INAR(p) process, *Journal of Time Series Analysis*, 2011, **32**(2): 195–223.

[27] Zhu F K and Wang D H, Empirical likelihood for linear and log-linear INGARCH models, *Journal of the Korean Statistical Society*, 2015, **44**(1): 150–160.

[28] Ding X and Wang D H, Empirical likelihood inference for INAR(1) model with explanatory variables, *Journal of the Korean Statistical Society*, 2016, **45**(4): 623–632.

[29] Yang K, Li H, and Wang D H, Estimation of parameters in the self-exciting threshold autoregressive processes for nonlinear time series of counts, *Applied Mathematical Modelling*, 2018, **57**(1): 226–247.

[30] Klimko L A and Nelson P I, On conditional least squares estimation for stochastic processes, *The Annals of Statistics*, 1978, **6**(3): 629–642.

[31] Li Q, Lian H, and Zhu F K, Robust closed-form estimators for the integer-valued GARCH (1,1) model, *Computational Statistics and Data Analysis*, 2016, **101**: 209–225.

[32] Bisaglia L and Gerolimetto M, Model-based INAR bootstrap for forecasting INAR(p) models, *Computational Statistics*, 2019, **34**(4): 1815–1848.

[33] Billingsley P, *Statistical Inference for Markov Processes*, The University of Chicago Press, Chicago, 1961.
[34] Owen A, *Empirical Likelihood*, CRC Press, New York, 2001.
[35] Qin J and Lawless J, Empirical likelihood and general estimating equations, *The Annals of Statistics*, 1994, **22**(1): 300–325.

Appendix

Before proving Theorem 3.1, we first show some crucial lemmas.

Lemma 1 Under the assumptions (C1) and (C2), we have

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0) \xrightarrow{\text{a.s.}} \boldsymbol{W}(\boldsymbol{\theta}_0), \quad n \longrightarrow +\infty,$$

in which $\boldsymbol{W}(\boldsymbol{\theta}_0)$ is defined by (5).

Proof By the strict stationarity and ergodicity, it is easy to know that

$$\frac{1}{n}\sum_{t=1}^{n} m_{1t}^2(\boldsymbol{\theta}_0) \xrightarrow{\text{a.s.}} E[X_0^2(X_1-\phi_0X_0-\lambda_0)^2], \quad n \longrightarrow +\infty,$$
$$\frac{1}{n}\sum_{t=1}^{n} m_{1t}(\boldsymbol{\theta}_0)m_{2t}(\boldsymbol{\theta}_0) \xrightarrow{\text{a.s.}} E[X_0(X_1-\phi_0X_0-\lambda_0)^2], \quad n \longrightarrow +\infty,$$
$$\frac{1}{n}\sum_{t=1}^{n} m_{2t}^2(\boldsymbol{\theta}_0) \xrightarrow{\text{a.s.}} E[(X_1-\phi_0X_0-\lambda_0)^2], \quad n \longrightarrow +\infty.$$

This completes the proof. ▮

Lemma 2 Under the assumptions (C1) and (C2), we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0) \xrightarrow{L} N(\boldsymbol{0}, \boldsymbol{W}(\boldsymbol{\theta}_0)), \quad n \longrightarrow +\infty.$$

Proof For any fixed $\boldsymbol{c} = (c_1, c_2)^{\mathrm{T}} \in \mathbb{R}^2$, by Theorems 1.1 and 1.2 in [33], we conclude that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{c}^{\mathrm{T}}\boldsymbol{m}_t(\boldsymbol{\theta}_0) \xrightarrow{L} N(\boldsymbol{0}, \boldsymbol{c}^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{\theta}_0)\boldsymbol{c}), \quad n \longrightarrow +\infty.$$

Thus, the lemma holds in view of Cramér-Wold device. ▮

Lemma 3 Under the assumptions (C1) and (C2), we have

$$\max_{1\le t\le n}\|\boldsymbol{m}_t(\boldsymbol{\theta})\| = o_p(n^{1/2}).$$

Proof By (C2), it holds that

$$E(\|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\|^2) = E(\boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{m}_t(\boldsymbol{\theta}_0)) < +\infty, \quad t \ge 1,$$

therefore, for any $\eta > 0$, we have

$$\lim_{n\longrightarrow\infty} nP(\|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\|^2 > n\eta^2) = 0, \quad t \geq 1,$$

which leads to

$$\lim_{n\longrightarrow\infty} nP(\|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\| > n^{1/2}\eta) = 0, \quad t \geq 1.$$

Furthermore, (C2) implies that

$$\begin{aligned} P\Big(\max_{1\leq t\leq n} \|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\| > n^{1/2}\eta\Big) &\leq \sum_{t=1}^{n} P(\|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\| > n^{1/2}\eta) \\ &= nP(\|\boldsymbol{m}_1(\boldsymbol{\theta}_0)\| > n^{1/2}\eta) \longrightarrow 0, \quad n \longrightarrow +\infty, \end{aligned}$$

which means

$$\max_{1\leq t\leq n} \|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\| = o_p(n^{1/2}).$$

This completes the proof. ∎

Lemma 4 Under the assumptions (C1) and (C2), we have

$$\frac{1}{n}\sum_{t=1}^{n} \|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\|^3 = o_p(n^{1/2}).$$

Proof By Lemmas 1 and 3, we obtain

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n} \|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\|^3 &\leq \max_{1\leq t\leq n} \|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\| \cdot \frac{1}{n}\sum_{t=1}^{n} \|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\|^2 \\ &= \max_{1\leq t\leq n} \|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\| \cdot \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{m}_t(\boldsymbol{\theta}_0) \\ &= o_p(n^{1/2})O_p(1) = o_p(n^{1/2}). \end{aligned}$$

The conclusion then follows. ∎

Lemma 5 Under the assumptions (C1) and (C2), we have

$$\|\boldsymbol{b}(\boldsymbol{\theta}_0)\| = O_p(n^{-1/2}),$$

in which $\boldsymbol{b}_t(\boldsymbol{\theta}_0)$ is defined in (7).

Proof As that in [34], let $\boldsymbol{b}(\boldsymbol{\theta}_0) = \|\boldsymbol{b}(\boldsymbol{\theta}_0)\|\boldsymbol{\nu}(\boldsymbol{\theta}_0)$ in which

$$\boldsymbol{\nu}(\boldsymbol{\theta}_0) = \frac{\boldsymbol{b}(\boldsymbol{\theta}_0)}{\|\boldsymbol{b}(\boldsymbol{\theta}_0)\|}$$

and denote

$$Y_t(\boldsymbol{\theta}_0) = \boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{m}_t(\boldsymbol{\theta}_0), \quad Z_n^*(\boldsymbol{\theta}_0) = \max_{1\leq t\leq n} \|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\|.$$

Noting that $1/(1+Y_t(\boldsymbol{\theta}_0)) = 1 - Y_t(\boldsymbol{\theta}_0)/(1+Y_t(\boldsymbol{\theta}_0))$ and $\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)g(\boldsymbol{b}(\boldsymbol{\theta}_0)) = 0$, we get

$$0 = \boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)g(\boldsymbol{b}(\boldsymbol{\theta}_0))$$

$$
\begin{aligned}
&= \boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\left(1-\frac{Y_t(\boldsymbol{\theta}_0)}{1+Y_t(\boldsymbol{\theta}_0)}\right)\\
&= \boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0) - \boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{m}_t(\boldsymbol{\theta}_0)\boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{\nu}(\boldsymbol{\theta}_0)\|\boldsymbol{b}(\boldsymbol{\theta}_0)\|}{1+Y_t(\boldsymbol{\theta}_0)}\\
&= \boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0) - \|\boldsymbol{b}(\boldsymbol{\theta}_0)\|\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{m}_t(\boldsymbol{\theta}_0)\boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0)}{1+Y_t(\boldsymbol{\theta}_0)}\boldsymbol{\nu}(\boldsymbol{\theta}_0),
\end{aligned}
$$

then it follows that

$$
\|\boldsymbol{b}(\boldsymbol{\theta}_0)\|\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{m}_t(\boldsymbol{\theta}_0)\boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0)}{1+Y_t(\boldsymbol{\theta}_0)}\boldsymbol{\nu}(\boldsymbol{\theta}_0) = \boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0).
$$

Since

$$
p_t(\boldsymbol{\theta}_0) = \frac{1}{n}\frac{1}{1+Y_t(\boldsymbol{\theta}_0)} > 0,
$$

we obtain

$$
1+Y_t(\boldsymbol{\theta}_0) > 0
$$

and hence

$$
\begin{aligned}
&\|\boldsymbol{b}(\boldsymbol{\theta}_0)\|\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{\nu}(\boldsymbol{\theta}_0)\\
\le&\|\boldsymbol{b}(\boldsymbol{\theta}_0)\|\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{m}_t(\boldsymbol{\theta}_0)\boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0)}{1+Y_t(\boldsymbol{\theta}_0)}\boldsymbol{\nu}(\boldsymbol{\theta}_0)\Big(1+\max_{1\le t\le n}Y_t(\boldsymbol{\theta}_0)\Big)\\
\le&\|\boldsymbol{b}(\boldsymbol{\theta}_0)\|\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{m}_t(\boldsymbol{\theta}_0)\boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0)}{1+Y_t(\boldsymbol{\theta}_0)}\boldsymbol{\nu}(\boldsymbol{\theta}_0)\left(1+\|\boldsymbol{b}(\boldsymbol{\theta}_0)\|Z_n^*(\boldsymbol{\theta}_0)\right)\\
=&\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\left(1+\|\boldsymbol{b}(\boldsymbol{\theta}_0)\|Z_n^*(\boldsymbol{\theta}_0)\right),
\end{aligned}
$$

therefore,

$$
\begin{aligned}
&\|\boldsymbol{b}(\boldsymbol{\theta}_0)\|\left(\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{\nu}(\boldsymbol{\theta}_0) - Z_n^*(\boldsymbol{\theta}_0)\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\right)\\
\le&\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0).
\end{aligned}
\tag{11}
$$

From Lemma 2, it is obvious that

$$
\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0) = O_p(1),
$$

together with Lemma 3, we get

$$
Z_n^*(\boldsymbol{\theta}_0)\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}}Z_n^*(\boldsymbol{\theta}_0)\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0) = o_p(1)O_p(1) = o_p(1) \tag{12}
$$

and

$$\sqrt{n}\left(\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\right)=\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)=O_p(1),$$

which leads to

$$\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)=O_p(n^{-1/2}). \tag{13}$$

On the other hand, we know that from Lemma 1

$$\sigma_{\min}(\boldsymbol{\theta}_0)+o_p(1)\le\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{\nu}(\boldsymbol{\theta}_0)\le\sigma_{\max}(\boldsymbol{\theta}_0)+o_p(1),$$

where $\sigma_{\min}(\boldsymbol{\theta}_0)$ and $\sigma_{\max}(\boldsymbol{\theta}_0)$ represent the largest and smallest eigenvalues of $\boldsymbol{W}(\boldsymbol{\theta}_0)$. Then by (11)–(13), it holds that

$$\|\boldsymbol{b}(\boldsymbol{\theta}_0)\|\left(\boldsymbol{\nu}^{\mathrm{T}}(\boldsymbol{\theta}_0)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{\nu}(\boldsymbol{\theta}_0)+o_p(1)\right)=O_p(n^{-1/2}),$$

from which we conclude

$$\|\boldsymbol{b}(\boldsymbol{\theta}_0)\|=O_p(n^{-1/2}).$$

The proof thus is completed. ∎

Now, we are ready to prove Theorem 3.1.

Proof of Theorem 3.1

Note that

$$Y_t(\boldsymbol{\theta}_0)=\boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{m}_t(\boldsymbol{\theta}_0),$$

it follows from Lemmas 3 and 5 that

$$\max_{1\le t\le n}|Y_t(\boldsymbol{\theta}_0)|\le\|Y_t(\boldsymbol{\theta}_0)\|\max_{1\le t\le n}\|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\|=O_p(n^{-1/2})o_p(n^{-1/2})=o_p(1).$$

By Taylor's expansion, we get

$$\begin{aligned}l(\boldsymbol{\theta}_0)&=2\sum_{t=1}^{n}\log(1+\boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{m}_t(\boldsymbol{\theta}_0))\\&=2\sum_{t=1}^{n}\log(1+Y_t(\boldsymbol{\theta}_0))\\&=2\sum_{t=1}^{n}Y_t(\boldsymbol{\theta}_0)-\sum_{t=1}^{n}Y_t^2(\boldsymbol{\theta}_0)+2\sum_{t=1}^{n}\eta_t,\end{aligned}$$

in which as $n\longrightarrow+\infty$,

$$P(|\eta_t|\le A|Y_t(\boldsymbol{\theta}_0)|^3,1\le t\le n)\longrightarrow 1$$

for some $A>0$.

The fact $\frac{1}{1+Y_t(\boldsymbol{\theta}_0)} = 1 - Y_t(\boldsymbol{\theta}_0) + \frac{Y_t^2(\boldsymbol{\theta}_0)}{1+Y_t(\boldsymbol{\theta}_0)}$ results in

$$\begin{aligned}
\mathbf{0} &= \frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{m}_t(\boldsymbol{\theta}_0)}{1+\boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{m}_t(\boldsymbol{\theta}_0)}\\
&= \frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{m}_t(\boldsymbol{\theta}_0)}{1+Y_t(\boldsymbol{\theta}_0)}\\
&= \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\left(1 - Y_t(\boldsymbol{\theta}_0) + \frac{Y_t^2(\boldsymbol{\theta}_0)}{1+Y_t(\boldsymbol{\theta}_0)}\right)\\
&= \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0) - \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{b}(\boldsymbol{\theta}_0) + \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\frac{Y_t^2(\boldsymbol{\theta}_0)}{1+Y_t(\boldsymbol{\theta}_0)},
\end{aligned}$$

thus, we can obtain

$$\boldsymbol{b}(\boldsymbol{\theta}_0) = \boldsymbol{C}_n^{-1}\boldsymbol{B}_n + \boldsymbol{C}_n^{-1}\boldsymbol{D}_n,$$

where

$$\boldsymbol{B}_n = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0),\quad \boldsymbol{C}_n = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\boldsymbol{m}_t^{\mathrm{T}}(\boldsymbol{\theta}_0),\quad \boldsymbol{D}_n = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\frac{Y_t^2(\boldsymbol{\theta}_0)}{1+Y_t(\boldsymbol{\theta}_0)}.$$

Therefore, we have

$$\begin{aligned}
l(\boldsymbol{\theta}_0) &= 2\sum_{t=1}^{n}Y_t(\boldsymbol{\theta}_0) - \sum_{t=1}^{n}Y_t^2(\boldsymbol{\theta}_0) + 2\sum_{t=1}^{n}\eta_t\\
&= 2\sum_{t=1}^{n}\boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{m}_t(\boldsymbol{\theta}_0) - \sum_{t=1}^{n}\boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{m}_t(\boldsymbol{\theta}_0)\boldsymbol{m}^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{b}_t(\boldsymbol{\theta}_0) + 2\sum_{t=1}^{n}\eta_t\\
&= 2n\boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{B}_n - n\boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta}_0)\boldsymbol{C}_n\boldsymbol{b}_t(\boldsymbol{\theta}_0) + 2\sum_{t=1}^{n}\eta_t\\
&= n\boldsymbol{B}_n^{\mathrm{T}}\boldsymbol{C}_n^{-1}\boldsymbol{B}_n - n\boldsymbol{D}_n^{\mathrm{T}}\boldsymbol{C}_n^{-1}\boldsymbol{D}_n + 2\sum_{t=1}^{n}\eta_t. \qquad (14)
\end{aligned}$$

First, by Lemmas 1 and 2, it is easy to know

$$\boldsymbol{C}_n \xrightarrow{\text{a.s.}} \boldsymbol{W}(\boldsymbol{\theta}_0),\quad \sqrt{n}\boldsymbol{N}_n \xrightarrow{L} N(\mathbf{0}, \boldsymbol{W}(\boldsymbol{\theta}_0)),\quad n \longrightarrow +\infty,$$

from which it holds that

$$n\boldsymbol{B}_n^{\mathrm{T}}\boldsymbol{C}_n^{-1}\boldsymbol{B}_n = (\sqrt{n}\boldsymbol{B}_n^{\mathrm{T}})\boldsymbol{C}_n^{-1}(\sqrt{n}\boldsymbol{B}_n) \xrightarrow{L} \chi^2(2),\quad n \longrightarrow +\infty. \qquad (15)$$

Second, by Lemmas 4 and 5, we have

$$\begin{aligned}
\|\boldsymbol{D}_n\| &= \left\|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{m}_t(\boldsymbol{\theta}_0)\frac{Y_t^2(\boldsymbol{\theta}_0)}{1+Y_t(\boldsymbol{\theta}_0)}\right\|\\
&\le \|\boldsymbol{b}(\boldsymbol{\theta}_0)\|^2 \cdot \frac{1}{n}\sum_{t=1}^{n}\|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\|^3 \cdot \frac{1}{|1 - \max_{1\le t\le n}|Y_t(\boldsymbol{\theta}_0)||}
\end{aligned}$$

$$= (O_p(n^{-1/2}))^2 o_p(n^{1/2}) O_p(1)$$
$$= o_p(n^{-1/2}),$$

therefore, it follows that

$$|n\boldsymbol{D}_n^{\mathrm{T}} \boldsymbol{C}_n^{-1} \boldsymbol{D}_n| \le n\|\boldsymbol{D}_n\| \cdot \|\boldsymbol{C}_n^{-1}\|_F \cdot \|\boldsymbol{D}_n\| = n o_p(n^{-1/2}) O_p(1) o_p(n^{-1/2}) = o_p(1), \tag{16}$$

in which $\|\cdot\|_F$ denotes the Frobenius norm.

Last of all, we can verify that

$$\left|2\sum_{t=1}^{n} \eta_t\right| \le 2A\|\boldsymbol{b}(\boldsymbol{\theta}_0)\|^3 \sum_{t=1}^{n} \|\boldsymbol{m}_t(\boldsymbol{\theta}_0)\|^3 = n(O_p(n^{-1/2}))^3 o_p(n^{1/2}) = o_p(1). \tag{17}$$

We can conclude from (14)–(17) that

$$l(\boldsymbol{\theta}_0) \xrightarrow{L} \chi^2(2).$$

This completes the proof. ▮

Proof of Theorem 3.3

The conclusion can be verified using similar argument in Section 6 in [25]. ▮

Proof of Theorem 3.4

This theorem can be proved with standard argument in Theorem 2 in [35]. ▮