Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 4, 685–711
https://doi.org/10.1214/19-BJPS452

Two classes of dynamic binomial integer-valued ARCH models

Huaping Chen[a], **Qi Li**[b] **and Fukang Zhu**[a]

[a]*Jilin University*
[b]*Changchun Normal University*

Abstract. This paper introduces two classes of binomial integer-valued ARCH models with dynamic survival probabilities, each of which is controlled by a stochastic recurrence equation. Stationarity and ergodicity of the process are established, and stochastic properties are given. Conditional least squares and conditional maximum likelihood estimators for the parameters of interest are considered, and their large-sample properties are established. The performances of these estimators are compared via simulation studies. Finally, we demonstrate the usefulness of the proposed models by analyzing real datasets.

1 Introduction

The analysis and modeling of integer-valued time series with a finite range have become a popular research area during the last decades. One of the earliest models for time series of bounded counts, the binomial AR(1) model, was proposed by McKenzie (1985). Its definition is based on the probabilistic operation of binomial thinning introduced by Steutel and van Harn (1979): $\alpha \circ X = \sum_{i=1}^{X} Y_i$, where X is a non-negative integer-valued random variable and Y_i is independent and identically distributed (i.i.d.) Bernoulli random variables with success probability $\alpha \in (0, 1)$, that is, $P(Y_i = 1) = 1 - P(Y_i = 0) = \alpha$. Using the thinning operator "$\circ$" with $\alpha, \beta \in (0, 1)$, McKenzie (1985) defined the binomial AR(1) process $\{X_t\}$ by the difference equation

$$X_t = \alpha \circ X_{t-1} + \beta \circ (n - X_{t-1}), \quad t = 1, 2, \dots,$$

where X_0 follows the binomial distribution with $P(X_0 = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, $\alpha = \beta + \gamma$ and $\beta = (1 - \gamma)p$ with $\gamma \in (\max(-p/(1 - p), -(1 - p)/p), 1)$. All the counting series in "$\alpha\circ$" and "$\beta\circ$" are mutually independent sequences of independent Bernoulli distributed random variables with parameters α and β, respectively, and the counting series at time t are independent of the random variables $X_s, \forall s < t$. This model was further investigated by Weiß (2009a, 2009b), Cui and Lund (2010), Weiß and Pollett (2012) and it was generalized to the higher-order case by Weiß (2009c), Kim and Park (2010a, 2010b), see Möller et al. (2016) and Yang, Wang and Li (2018) for threshold binomial autoregressive processes.

One important limitation of the above models is that they assume the thinning probabilities are not to be affected by various environmental factors. Based on this point, Zhang, Wang and Zhu (2011a, 2011b, 2012) proposed the random coefficient integer-valued autoregressive processes. Weiß and Kim (2014) proposed the beta-binomial AR(1) model to describe time-dependent counts with extra-binomial variation. Weiß and Pollett (2014) proposed a binomial AR(1) model with density dependent thinning, that is, $X_t|\mathcal{F}_{t-1} : \mathrm{Bin}(n, \alpha_t), t \in \{\dots, -1, 0, 1, \dots\}$, where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{X_{t-k}\}_{k\geq 1}$ and α_t is given by $\alpha_t = a_0 + a_1 \frac{X_{t-1}}{n}$ with $a_0 > 0$ and $a_1 \geq 0$. This model is referred to as the BARCH(1) model and it was generalized to the pth-order case by a linear link function in Ristić, Weiß and

Key words and phrases. Binomial ARCH, parameter estimation, stationarity, time series of counts.
Received January 2019; accepted July 2019.

Janjić (2016). Lee and Lee (2019) discussed a binomial integer-valued GARCH(1, 1) model by a similar linear link function, see Scotto et al. (2014) and Ristić and Popović (2019) for bivariate binomial time series models.

However, smooth change in thinning probability can not be described by above models. Zheng and Basawa (2008) provided a dynamic structure INAR(1) model whose thinning probability is a sequence of dependent (on past observations) random variables and updated by using past information. Creal, Koopman and Lucas (2013) proposed generalized autoregressive score models to describe the smooth change by using the score of the log-likelihood function, see Harvey (2013), Blasques, Koopman and Lucas (2014), Blasques et al. (2018a), Blasques, Lucas and Silde (2018b) and Bazzi et al. (2017) for recent developments. Based on this point, Gorgi (2018) proposed a dynamic structure INAR model whose thinning probability is driven by a stochastic recurrence equation to describe the smooth change in thinning probability. A common characteristic of above models is that they are all dedicated to the infinite time series, but there exist few literatures on the time series with bounded support. To fill this gap, we provide two classes of random coefficient binomial integer-valued ARCH models with dynamic structures. The first one is a random coefficient binomial integer-valued ARCH(p) model with hysteric property, which makes the survival probability updated by the past information, and the model will be referred to as the binomial logit-ARCH(p) model. The second one is a binomial ARCH model with a time-varying survival probability, which is controlled through a stochastic recurrence equation driven by the score of the predictive log-likelihood. The new model not only updates the survival probability at each time period by using past information, but also describes the smooth change, and it will be referred to as the binomial score-ARCH(1) model.

The paper is organized as follows. The two classes of random coefficient binomial integer-valued ARCH models with dynamic structures are discussed in Section 2. Conditional least squares (CLS) and conditional maximum likelihood (CML) estimates and their asymptotic properties are established in Section 3. A simulation study and two real datasets which show the effectiveness of the new models are given in Sections 4 and 5, respectively. Conclusions are made in Section 6. All proofs of theorems are given in Appendix A and some auxiliary results are given in Appendix B.

Throughout this paper, we use the following notations. $|\cdot|$ denotes the absolute value of a random variable; $\|\cdot\|$ denotes the Euclidean norm of a matrix or vector; $\|\cdot\|_\Theta$ denotes the uniform norm, that is, $\|u\|_\Theta = \sup_{\boldsymbol{\eta}\in\Theta} |u(\boldsymbol{\eta})|$ for all function $u(\boldsymbol{\eta})$ mapping from Θ into $\mathbb{R}$; $\xrightarrow{e.a.s.}$ denotes exponential almost sure uniform convergence, that is, a sequence of non-negative random variables $\{\alpha_t\}$ converges e.a.s. to zero if there exists a constant $r > 1$ such that $r^t \alpha_t \xrightarrow{a.s.} 0$ as t diverges.

2 Dynamic binomial integer-valued ARCH models

In this section, we will introduce two classes of binomial integer-valued ARCH models with dynamic structures.

2.1 Binomial logit-ARCH model

Let $\mathcal{F}_{t-1}$ be the σ-field generated by the random variables $\{X_{t-1}, \dots, X_{t-p}\}$, $p = 1, 2, \dots$, $n \in \mathbb{N}$ be the predetermined upper limit of the range and $\operatorname{logit}(x) = \log(x/(1-x))$, $\forall x \in$

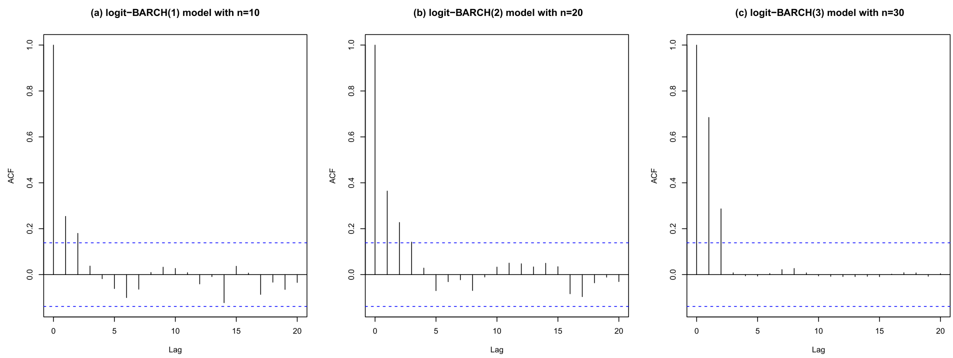

Figure 1 *Examples of autocorrelation function of* logit-BARCH(p) *model for different values of upper limit* n: (a) $n = 10$, (b) $n = 20$, (c) $n = 30$.

$(0, 1)$. Then the binomial logit-ARCH(p) model is defined as:

$$\begin{cases} X_t|\mathcal{F}_{t-1} : \mathrm{Bin}(n, \alpha_t), \\ \mathrm{logit}(\alpha_t) = r_0 + \displaystyle\sum_{k=1}^{p} r_k X_{t-k}, \quad t = 1, 2, \ldots, \end{cases} \tag{2.1}$$

where $r_j \in \mathbb{R}$, $j = 0, 1, \ldots, p$.

The time series $\{X_t\}$ given by (2.1) will be denoted as logit-BARCH(p) model. Specially, for $k = 1, 2, \ldots, p$, if $r_k = 0$ and $r_0 \neq 0$, $\{X_t\}$ follows a binomial distribution with the constant probability $\exp(r_0)/(1 + \exp(r_0))$; if $r_k \neq 0$, $\{X_t\}$ follows the binomial distribution with a time-varying survival probability which allows for flexible dynamics of the number of counts in terms of past counts, which are captured by $\sum_{k=1}^{p} r_k X_{t-k}$.

For model (2.1), the conditional probability mass function of X_t has the form

$$P_n(X_t = x_t|\mathcal{F}_{t-1}) = \binom{n}{x_t} \alpha_t^{x_t} (1 - \alpha_t)^{n-x_t}. \tag{2.2}$$

with $\alpha_t = \frac{\exp(r_0 + \sum_{k=1}^{p} r_k X_{t-k})}{1+\exp(r_0 + \sum_{k=1}^{p} r_k X_{t-k})}$.

In what follows, we first discuss the stationarity and ergodicity of processes $\{X_t\}$, then we will illustrate the stochastic properties of $\{X_t\}$ by examples.

Theorem 1. *The binomial* logit-ARCH(p) *process* $\{X_t\}$ *is an ergodic and strictly stationary process.*

Note that $\mathrm{Cov}(X_{t-k}, X_t) = n\,\mathrm{Cov}(\alpha_t, X_{t-k})$ by assumption (2.1). Thus, the second-order correlation structure of the logit-BARCH(p) process is not similar to that of the BARCH(p) process in Ristić, Weiß and Janjić (2016). In Figure 1, we present some examples of the autocorrelation function of the logit-BARCH(p) model with different upper limit of the range n and $p \in \{1, 2, 3\}$, when fixing $r_0 = -1$, $r_1 = 0.1$, $r_2 = 0.1$, $r_3 = 0.1$ and sample size $T = 200$.

2.2 Binomial score-ARCH(1) model

Let $\{X_t, t \in \mathbb{Z}\}$ and $\{\alpha_t\}$ be time series of counts with fixed α_0 and X_0, $n \in \mathbb{Z}^+$ be the predetermined upper limit of the range, $\mathcal{D}_{t-1}$ denote the information available on X_t up to time $t - 1$, that is, $\mathcal{D}_{t-1} = \sigma(X_s, \forall s < t)$. Then the binomial score-ARCH(1) process $\{X_t, t \in \mathbb{Z}\}$

satisfies

$$\begin{cases} X_t|\mathcal{D}_{t-1} : \mathrm{Bin}(n, \alpha_t), \\ \mathrm{logit}(\alpha_t) = w + \beta\,\mathrm{logit}(\alpha_{t-1}) + \tau s_t \end{cases} \tag{2.3}$$

with

$$\begin{cases} s_t = \dfrac{\partial \log P_n(X_{t-1} = x|\alpha_{t-1}, X_{t-2})}{\partial\,\mathrm{logit}(\alpha_{t-1})}, \\ P_n(X_{t-1} = x|\alpha_{t-1}, X_{t-2}) = \dbinom{n}{x} \alpha_{t-1}^x (1 - \alpha_{t-1})^{n-x}, \end{cases} \tag{2.4}$$

where $w, \tau \in \mathbb{R}$, $|\beta| \in (0, 1)$ and $\log(\max(|\beta - \frac{1}{4}\tau n|, |\beta + \frac{1}{4}\tau n|, |\beta|)) < 0$.

The time series $\{X_t\}$ will be denoted as score-BARCH(1) model. According to (2.4), we have $s_t = X_{t-1} - n\alpha_{t-1}$. Hence, the success probability α_t satisfies the following stochastic recurrence equation:

$$\mathrm{logit}(\alpha_t) = w + \beta\,\mathrm{logit}(\alpha_{t-1}) + \tau(X_{t-1} - n\alpha_{t-1}). \tag{2.5}$$

Note that $E(s_t|\alpha_{t-1}, X_{t-2}) = 0$ and $E(s_t^2|\alpha_{t-1}, X_{t-2}) = \mathrm{Var}(X_t) < \infty$. Hence, the score s_t can be regarded as the innovation of the dynamic system and (2.3) looks like a GARCH(1, 1) model. Furthermore, if we let $u_t = \mathrm{logit}(\alpha_t)$ and $\beta \in (0, 1)$, then

$$u_t = w + \beta u_{t-1} + \tau s_t.$$

Thus, $u_t = \sum_{j=0}^{t-1} \beta^j w + \beta^t u_0 + \tau \sum_{j=0}^{t-1} \beta^j s_{t-j}$. Hence, $E(u_t) = \sum_{j=0}^{t-1} \beta^j w + \beta^t u_0 \to \frac{w}{1-\beta}$, as $t \to \infty$.

The score used to update $\mathrm{logit}(\alpha_t)$ defines a steepest ascent direction for improving the local fit of the score-BARCH(1) model in terms of the likelihood at time $t - 1$ given the previous position of the parameter α_t. If $|\beta| \in (0, 1)$ and $\tau \neq 0$, $\mathrm{logit}(\alpha_t)$ allows for flexible dynamics of the number of counts in terms of past counts, which are captured by s_t and u_{t-1}.

A fundamental problem in the analysis of the score-BARCH(1) model is to prove the stability and the ergodicity of the time-varying parameter $\{\alpha_t\}$. Similar to the method in Gorgi (2018), we assume that the observed data is generated by an unknown stationary and ergodic data generating process with a finite range and Assumption 1 holds.

Assumption 1. The parametric space Ω is compact with $\Omega = \{\boldsymbol{\eta} : \boldsymbol{\eta} = (w, \beta, \tau)^\top\}$, $|\beta| \in (0, 1)$, and $\log(\max(|\beta - \frac{1}{4}\tau n|, |\beta + \frac{1}{4}\tau n|, |\beta|)) < 0$.

Theorem 2. *Let $\{X_t, t \in \mathbb{Z}\}$ be a stationary and ergodic sequence of random variable sequence with a finite range. If Assumption 1 holds, then the time-varying parameter $\{\alpha_t(\boldsymbol{\eta}), t \in \mathbb{Z}\}$ defined by (2.3) converges e.a.s. and uniformly to a unique stationary and ergodic sequence $\{\widetilde{\alpha}_t\}$, that is,*

$$\left\|\mathrm{logit}(\alpha_t) - \mathrm{logit}(\widetilde{\alpha}_t)\right\|_\Omega \xrightarrow{e.a.s.} 0, \quad \text{as } t \to \infty,$$

for any initialization α_0.

Note that we discuss the stability of $\{\alpha_t\}$ under the unknown data generating process in Theorem 2. Blasques, Koopman and Lucas (2015), Blasques, Lucas and Silde (2018b) and Bazzi et al. (2017) also considered similar assumptions about the data generating process. For illustrative purposes, we let $P_n^o(x|\mathcal{D}_{t-1}) = P_n^o(X_t = x|\mathcal{D}_{t-1})$ be the true conditional probability of the unknown data generating process $\{X_t\}$, where the true conditional probability of

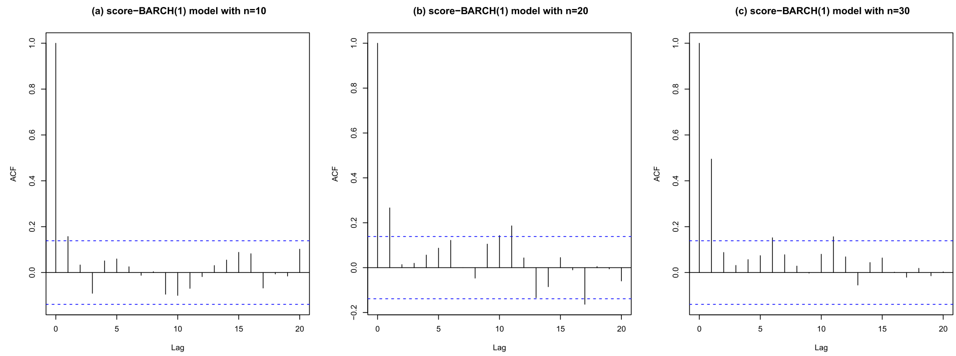

Figure 2 *Examples of autocorrelation function of score-BARCH model for different values of upper limit* n: (a) $n = 10$, (b) $n = 20$, (c) $n = 30$.

$\{X_t\}$ is conditional on the past observations $\mathcal{D}_{t-1} = \sigma(X_{t-1}, X_{t-2}, \dots)$ and n is the predetermined upper limit of $\{X_t\}$. Denote $P_n(x|\widetilde{\alpha}_t(\boldsymbol{\eta}), \mathcal{D}_{t-1}) = P_n(X_t = x|\widetilde{\alpha}_t(\boldsymbol{\eta}), \mathcal{D}_{t-1})$ is the postulated conditional probability under the time-varying parameter $\widetilde{\alpha}_t(\boldsymbol{\eta})$. Then the conditional Kullback–Leibler (KL) divergence between the true conditional probability $P_n^o(x|\mathcal{D}_{t-1})$ and the postulated conditional probability $P_n(X_t = x|\widetilde{\alpha}_t(\boldsymbol{\eta}), \mathcal{D}_{t-1})$ is

$$\mathrm{KL}_t(\boldsymbol{\eta}) = \sum_{x=0}^{n} \log\left(\frac{P_n^o(x|\mathcal{D}_{t-1})}{P_n(x|\widetilde{\alpha}_t(\boldsymbol{\eta}), \mathcal{D}_{t-1})}\right) P_n^o(x|\mathcal{D}_{t-1}),$$

and the average KL divergence is $\mathrm{KL}(\boldsymbol{\eta}) = E(\mathrm{KL}_t(\boldsymbol{\eta}))$ under the condition $E(|\log P_n^o(x|\mathcal{D}_{t-1})|) < \infty$.

Corollary 1. *Let $\{X_t, t \in \mathbb{Z}\}$ be a stationary and ergodic sequence of random variable sequence with a finite range. If Assumption 1 holds, then we have the following results*:

(1) $\|\alpha_t - \widetilde{\alpha}_t\|_\Omega \xrightarrow{e.a.s.} 0$, *as $t \to \infty$, for any initialization α_0.*
(2) *Let $\boldsymbol{\eta}_0$ be the minimizer of the average KL divergence* $\mathrm{KL}(\boldsymbol{\eta})$, *then we immediately obtain that* $\|\operatorname{logit}(\alpha_t(\boldsymbol{\eta}_0)) - \operatorname{logit}(\widetilde{\alpha}_t(\boldsymbol{\eta}_0))\|_\Omega \xrightarrow{e.a.s.} 0$, $t \to \infty$.
(3) *In addition, if there exists a sequence $\hat{\boldsymbol{\eta}}_l$ such that $\hat{\boldsymbol{\eta}}_l \xrightarrow{a.s.} \boldsymbol{\eta}_0$, $l \to \infty$, then we have*

$$\begin{cases} \left|\operatorname{logit}\big(\alpha_t(\hat{\boldsymbol{\eta}}_l)\big) - \operatorname{logit}\big(\widetilde{\alpha}_t(\boldsymbol{\eta}_0)\big)\right| \xrightarrow{a.s.} 0, \\ \left|P_n\big(x|\alpha_t(\hat{\boldsymbol{\eta}}_l), \mathcal{D}_{t-1}\big) - P_n\big(x|\widetilde{\alpha}_t(\boldsymbol{\eta}_0), \mathcal{D}_{t-1}\big)\right| \xrightarrow{a.s.} 0, \end{cases}$$

as $l \to \infty$, $t \to \infty$, where $P_n(x|\alpha_t(\hat{\boldsymbol{\eta}}_l), \mathcal{D}_{t-1}) = P_n(X_t = x|\alpha_t(\hat{\boldsymbol{\eta}}_l), \mathcal{D}_{t-1})$ is the conditional probability under the time-varying parameter $\alpha_t(\hat{\boldsymbol{\eta}}_l)$ and the parameter $\hat{\boldsymbol{\eta}}_l$.

Note that $\operatorname{Cov}(X_{t-k}, X_t) = n \operatorname{Cov}(\alpha_t, X_{t-k})$ by assumption (2.5). Thus, the second-order correlation structure of the score-BARCH process is not similar to that of BARCH(1) process in Ristić, Weiß and Janjić (2016). In Figure 2, we present some examples of the autocorrelation function of the model for different upper limit of the range n, when fixing $w = -1$, $\beta = 0.2$, $\tau = 0.1$ and sample size $T = 200$.

3 Parameter estimation

In this section, we use the conditional least squares (CLS) and conditional maximum likelihood (CML) methods to estimate the parameters in the binomial logit-ARCH(p) model and the binomial score-ARCH(1) model.

3.1 Binomial logit-ARCH(p) model

Let $\boldsymbol{\theta} = (r_0, r_1, \dots, r_p)^\top$ and $X_0, X_1, \dots, X_T$ be generated by the logit-BARCH(p) model with the true parameter value $\boldsymbol{\theta}_0$, where $T \in \mathbb{N}$ represents the size of the sample. Here, the parameter n (upper limit of the range) is considered as a known quantity. To estimate $\boldsymbol{\theta}_0$, we will first briefly discuss the CLS estimation and then develop the CML estimation. To study the asymptotic behaviour of the estimator, we make the following assumptions about the underlying process and the parameter space.

Assumption 2. The parametric space Θ is compact with $\Theta = \{\boldsymbol{\theta} : \boldsymbol{\theta} = (r_0, r_1, r_2, \dots, r_p)^\top\}$ and $\boldsymbol{\theta}_0$ is an interior point in Θ.

Assumption 3. If there exists a $t \geq 1$, such that $X_t(\boldsymbol{\theta}_0) = X_t(\boldsymbol{\theta})$, $P_n(x|\mathcal{F}_{t-1})_{\boldsymbol{\theta}_0}$ a.s., then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, where $P_n(x|\mathcal{F}_{t-1})_{\boldsymbol{\theta}_0} = P_n(X_t = x|\mathcal{F}_{t-1})_{\boldsymbol{\theta}_0}$ is the probability measure under the true parameter $\boldsymbol{\theta}_0$ and $\mathcal{F}_{t-1}$, n is the upper limit of X_t.

3.1.1 *Conditional least squares estimation.* Let $\boldsymbol{Y}_t = (1, X_{t-1}, \dots, X_{t-p})^\top$ and $g_t(\boldsymbol{\theta}) = E(X_t|\mathcal{F}_{t-1}) = \frac{n \exp(\boldsymbol{Y}_t^\top \boldsymbol{\theta})}{1+\exp(\boldsymbol{Y}_t^\top \boldsymbol{\theta})}$, then the CLS estimate $\hat{\boldsymbol{\theta}}_T^{\text{cls}}$ is obtained by minimizing the function

$$Q(\boldsymbol{\theta}) = \sum_{t=p+1}^{T} \big(X_t - E(X_t|\mathcal{F}_{t-1})\big)^2 = \sum_{t=p+1}^{T} \big(X_t - g_t(\boldsymbol{\theta})\big)^2, \tag{3.1}$$

that is, $\hat{\boldsymbol{\theta}}^{\text{cls}}$ is a solution of $\frac{\partial Q(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$, where

$$\begin{cases} \dfrac{\partial Q(\boldsymbol{\theta})}{\partial r_0} = \displaystyle\sum_{t=p+1}^{T} \big(X_t - n g_t(\boldsymbol{\theta})\big) g_t(\boldsymbol{\theta})\big(1 - g_t(\boldsymbol{\theta})\big) = 0, \\ \dfrac{\partial Q(\boldsymbol{\theta})}{\partial r_j} = \displaystyle\sum_{t=p+1}^{T} \big(X_t - n g_t(\boldsymbol{\theta})\big) g_t(\boldsymbol{\theta})\big(1 - g_t(\boldsymbol{\theta})\big) X_{t-j} = 0, \quad j = 1, 2, \dots, p. \end{cases}$$

The following theorem gives the asymptotic properties of the CLS estimator.

Theorem 3. *Let* $\{X_t\}$ *be the* logit-BARCH(p) *process and assumptions* 2 *and* 3 *hold. Then the CLS estimator* $\hat{\boldsymbol{\theta}}^{\text{cls}}$ *is consistent and has the following asymptotic distribution*:

$$\sqrt{T}\big(\hat{\boldsymbol{\theta}}_T^{\text{cls}} - \boldsymbol{\theta}_0\big) \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \boldsymbol{V}^{-1} \boldsymbol{W} \boldsymbol{V}^{-1}\big), \quad T \to \infty,$$

where $\boldsymbol{W} = E\Big(u_t^2 \frac{\partial g_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial g_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\Big)_{\boldsymbol{\theta}_0}$, $\boldsymbol{V} = E\Big(\frac{\partial g_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial g_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} - u_t \frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\Big)_{\boldsymbol{\theta}_0}$ *with* $u_t = X_t - g_t(\boldsymbol{\theta})$.

3.1.2 *Conditional maximum likelihood estimation.* Using (2.2), the conditional log-likelihood function can be written as:

$$\ell(\boldsymbol{\theta}) = \sum_{t=p+1}^{T} \log P_n(X_t|\mathcal{F}_{t-1}) \quad \text{with } \alpha_t = \frac{\exp(\boldsymbol{Y}_t^\top \boldsymbol{\theta})}{1 + \exp(\boldsymbol{Y}_t^\top \boldsymbol{\theta})} = f\big(\boldsymbol{Y}_t^\top \boldsymbol{\theta}\big). \tag{3.2}$$

Then the CML estimate $\hat{\boldsymbol{\theta}}_T^{\text{cml}}$ is obtained by minimizing (3.2), that is, $\hat{\boldsymbol{\theta}}_T^{\text{cml}}$ is a solution of the score equation $\frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = 0$, where

$$\begin{cases} \dfrac{\partial \ell(\boldsymbol{\theta})}{\partial r_0} = \displaystyle\sum_{t=p+1}^{T} \frac{f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})(1 - f(\boldsymbol{Y}_t^\top \boldsymbol{\theta}))}{P_n(X_t|\mathcal{F}_{t-1})} \frac{d P_n(X_t|\mathcal{F}_{t-1})}{d\alpha_t} = 0, \\ \dfrac{\partial \ell(\boldsymbol{\theta})}{\partial r_j} = \displaystyle\sum_{t=p+1}^{T} \frac{f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})(1 - f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})) X_{t-j}}{P_n(X_t|\mathcal{F}_{t-1})} \frac{d P_n(X_t|\mathcal{F}_{t-1})}{d\alpha_t} = 0 \end{cases}$$

with $\frac{dP_n(X_t|\mathcal{F}_{t-1})}{d\alpha_t} = \frac{n}{1-\alpha_t}[P_{n-1}(X_t - 1|\mathcal{F}_{t-1}) - P_n(X_t|\mathcal{F}_{t-1})]$ by Lemma 1, for $j = 1, 2, \dots, p$.

Let $P_{\boldsymbol{\theta}_0}$ be the probability measure under the true parameter $\boldsymbol{\theta}_0$ and, unless otherwise indicated, $E(\cdot)^k$ is taken under $\boldsymbol{\theta}_0$, $\forall k \geq 1$. Lemma 2 in the Appendix B establishes the identification of model (2.1). The following theorem gives the asymptotic properties of the CML estimator.

Theorem 4. *Let $\{X_t\}$ be the* logit-BARCH(p) *process and assumptions* 2 *and* 3 *hold. Then, as $T \to \infty$,*

(1) *there exists an estimator $\hat{\boldsymbol{\theta}}_T^{\text{cml}}$ such that $\hat{\boldsymbol{\theta}}_T^{\text{cml}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$;*
(2) $\sqrt{T}(\hat{\boldsymbol{\theta}}_T^{\text{cml}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{I}^{-1}(\boldsymbol{\theta}_0))$,

where $\boldsymbol{I}(\boldsymbol{\theta}_0) = E\Big[\frac{\partial \log P_n(X_t|\mathcal{F}_{t-1})}{\partial \boldsymbol{\theta}} \frac{\partial \log P_n(X_t|\mathcal{F}_{t-1})}{\partial \boldsymbol{\theta}^\top}\Big]_{\boldsymbol{\theta}_0}$.

3.2 Binomial score-ARCH(1) model

Let $\boldsymbol{\eta}_0$ be the minimizer of the average KL divergence KL$(\boldsymbol{\eta})$, the observed sequence $X_1, \dots, X_T$ be a realized path of an unknown data generating process $\{X_t\}$ and the true conditional probability of $\{X_t\}$ be denoted as $P_n^o(x|\mathcal{D}_{t-1})$, where $\mathcal{D}_{t-1} = \sigma(X_s)_{s<t}$, n is considered as a known quantity and $T \in \mathbb{N}$ represents the size of the sample.

Note that Theorem 2 implies that the initialization logit(α_0) is irrelevant asymptotically. Without loss of generality, we choose logit$(\alpha_0) = w/(1-\beta)$. For given X_0, we have

$$\text{logit}(\alpha_t) = w + \beta\, \text{logit}(\alpha_{t-1}) + \tau(X_{t-1} - n\alpha_{t-1}), \quad t = 1, 2, 3, \dots$$

by (2.5). Before discussing the CLS and the CML estimation, we make the following assumptions.

Assumption 4. The moment condition $E|\log P_n^o(X_t|\mathcal{D}_{t-1})| < \infty$ holds.

Assumption 5. If there exists a $t \geq 1$ such that $X_t(\boldsymbol{\eta}_0) = X_t(\boldsymbol{\eta})$, $P_n(x|\widetilde{\alpha}_t(\boldsymbol{\eta}_0), \mathcal{D}_{t-1})$ a.s., then $\boldsymbol{\eta} = \boldsymbol{\eta}_0$, where $P_n(x|\widetilde{\alpha}_t(\boldsymbol{\eta}), \mathcal{D}_{t-1}) = P_n(X_t = x|\widetilde{\alpha}_t(\boldsymbol{\eta}), \mathcal{D}_{t-1})$ is the conditional probability under the time-varying parameter $\widetilde{\alpha}_t(\boldsymbol{\eta})$ and n is the predetermined upper limit of X_t.

The item $E|\log P_n^o(X_t|\mathcal{D}_{t-1})| < \infty$ in Assumption 4 ensures the existence of the average KL divergence. Lemma 3 in Appendix B establishes the identification of the model $X_t \sim \text{Bin}(n, \widetilde{\alpha}_t)$ with $\widetilde{\alpha}_t$ satisfying the stochastic recurrence equation (2.5) under Assumption 5.

3.2.1 *Conditional least squares estimation.* Let $h(u_t) = \alpha_t = \frac{\exp(u_t)}{1+\exp(u_t)}$ with $u_t = w + \beta u_{t-1} + \tau s_t$, then $g_{\boldsymbol{\eta}}(u_t) = E(X_t|\alpha_t, \mathcal{D}_{t-1}) = nh(u_t)$. Then the CLS estimate $\hat{\boldsymbol{\eta}}_T^{\text{cls}}$ is obtained by minimizing the function

$$Q(\boldsymbol{\eta}) = \sum_{t=2}^{T} (X_t - g_{\boldsymbol{\eta}}(u_t))^2 = \sum_{t=2}^{T} (X_t - nh(u_t))^2, \tag{3.3}$$

that is, $\hat{\boldsymbol{\eta}}_T^{\mathrm{cls}}$ is a solution of $\frac{\partial Q(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = 0$, where

$$\begin{cases} \dfrac{\partial Q(\boldsymbol{\eta})}{\partial w} = n \sum_{t=2}^{T} (X_t - nh(u_t))h(u_t)(1-h(u_t)) = 0, \\ \dfrac{\partial Q(\boldsymbol{\eta})}{\partial \beta} = n \sum_{t=2}^{T} (X_t - nh(u_t))h(u_t)(1-h(u_t))u_{t-1} = 0, \\ \dfrac{\partial Q(\boldsymbol{\eta})}{\partial \tau} = n \sum_{t=2}^{T} (X_t - nh(u_t))h(u_t)(1-h(u_t))s_t = 0. \end{cases}$$

Note that if $\{\alpha_t\}$ is stationary and ergodic under the assumption of the data generating process $\{X_t\}$, then the asymptotic theory of the CLS estimator $\hat{\boldsymbol{\eta}}_T^{\mathrm{cls}}$ can be developed directly. A fundamental problem is that we do not know precisely what conditions guarantee the ergodicity of $\{\alpha_t\}$. However, the assumptions of Theorem 2 guarantee that $\{\alpha_t\}$ converges e.a.s. and uniformly to a unique stationary and ergodic sequence $\{\widetilde{\alpha}_t\}$. Hence, we use the ergodic properties of $\{\widetilde{\alpha}_t\}$ to study the asymptotic properties of the CLS estimators.

Denote

$$\widetilde{Q}(\boldsymbol{\eta}) = \sum_{t=2}^{T} (X_t - g_{\boldsymbol{\eta}}(\widetilde{u}_t))^2 \quad \text{and} \quad \overline{\widetilde{Q}}(\boldsymbol{\eta}) = E(X_t - g_{\boldsymbol{\eta}}(\widetilde{u}_t))^2$$

with $\widetilde{u}_t = \mathrm{logit}(\widetilde{\alpha}_t)$ and $g_{\boldsymbol{\eta}}(\widetilde{u}_t) = E(X_t|\widetilde{\alpha}_t, \mathcal{D}_{t-1})$. Let $\widetilde{\boldsymbol{\eta}}_T^{\mathrm{cls}}$ be a solution of the score equations $\frac{\partial \widetilde{Q}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = 0$. Similar to Theorem 3, the CLS estimator $\widetilde{\boldsymbol{\eta}}_T^{\mathrm{cls}}$ is consistent and asymptotic normal. Then the consistency and asymptotic normality of $\hat{\boldsymbol{\eta}}_T^{\mathrm{cls}}$ is constructed via the relation between $\hat{\boldsymbol{\eta}}_T^{\mathrm{cls}}$ and $\widetilde{\boldsymbol{\eta}}_T^{\mathrm{cls}}$. The following theorem gives the strong consistency and the asymptotic properties of $\hat{\boldsymbol{\eta}}_T^{\mathrm{cls}}$.

Theorem 5. *Let $\{X_t, t \in \mathbb{Z}\}$ be a stationary and ergodic sequence and Assumptions 1, 4 and 5 hold. Then the CLS estimator $\hat{\boldsymbol{\eta}}_T^{\mathrm{cls}}$ is consistent and has the following asymptotic distribution:*

$$\sqrt{T}(\hat{\boldsymbol{\eta}}_T^{\mathrm{cls}} - \boldsymbol{\eta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{V}^{-1}\boldsymbol{W}\boldsymbol{V}^{-1}), \quad T \to \infty,$$

where $\boldsymbol{V} = E\Big(\frac{\partial g_{\boldsymbol{\eta}}(u_t)}{\partial \boldsymbol{\eta}} \frac{\partial g_{\boldsymbol{\eta}}(u_t)}{\partial \boldsymbol{\eta}^{\top}} - U_t(\boldsymbol{\eta}) \frac{\partial^2 g_{\boldsymbol{\eta}}(u_t)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^{\top}}\Big)_{\boldsymbol{\eta}_0}$, $U_t(\boldsymbol{\eta}) = X_t - g_{\boldsymbol{\eta}}(u_t)$ *and* $\boldsymbol{W} = E\Big(U_t(\boldsymbol{\eta})^2 \frac{\partial g_{\boldsymbol{\eta}}(u_t)}{\partial \boldsymbol{\eta}} \frac{\partial g_{\boldsymbol{\eta}}(u_t)}{\partial \boldsymbol{\eta}^{\top}}\Big)_{\boldsymbol{\eta}_0}$.

3.2.2 *Conditional maximum likelihood estimation.* Using (2.4), the conditional log-likelihood function can be written as:

$$\ell(\boldsymbol{\eta}) := \sum_{t=2}^{T} l_t(\boldsymbol{\eta}) = \sum_{t=2}^{T} \log P_n(X_t|\alpha_t, \mathcal{D}_{t-1}) \quad \text{with } \alpha_t = \frac{\exp(u_t)}{1+\exp(u_t)}. \tag{3.4}$$

Then the CML estimate $\hat{\boldsymbol{\eta}}_T^{\mathrm{cml}}$ is obtained by minimizing (3.4), that is, $\hat{\boldsymbol{\eta}}_T^{\mathrm{cml}}$ is a solution of the score equation $\frac{\partial \ell(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = 0$, where

$$\begin{cases} \dfrac{\partial \ell(\boldsymbol{\eta})}{\partial w} = \sum_{t=2}^{T} \dfrac{h(u_t)(1-h(u_t))}{P_n(X_t|\alpha_t, \mathcal{D}_{t-1})} \dfrac{dP_n(X_t|\mathcal{F}_{t-1})}{d\alpha_t} = 0, \\ \dfrac{\partial \ell(\boldsymbol{\eta})}{\partial \beta} = \sum_{t=2}^{T} \dfrac{h(u_t)(1-h(u_t))u_{t-1}}{P_n(X_t|\alpha_t, \mathcal{D}_{t-1})} \dfrac{dP_n(X_t|\mathcal{F}_{t-1})}{d\alpha_t} = 0, \\ \dfrac{\partial \ell(\boldsymbol{\eta})}{\partial \tau} = \sum_{t=2}^{T} \dfrac{h(u_t)(1-h(u_t))s_t}{P_n(X_t|\alpha_t, \mathcal{D}_{t-1})} \dfrac{dP_n(X_t|\mathcal{F}_{t-1})}{d\alpha_t} = 0 \end{cases}$$

with $\frac{dP_n(X_t|\mathcal{F}_{t-1})}{d\alpha_t} = \frac{n}{1-\alpha_t}[P_{n-1}(X_t-1|\mathcal{F}_{t-1}) - P_n(X_t|\mathcal{F}_{t-1})]$ by Lemmas 1 and 4.

To study the asymptotic properties of the CML estimator, we denote

$$\overline{\widetilde{\ell}}(\boldsymbol{\eta}) = E\widetilde{l}_t(\boldsymbol{\eta}) \quad \text{and} \quad \widetilde{\ell}(\boldsymbol{\eta}) = \sum_{t=2}^{T} \widetilde{l}_t(\boldsymbol{\eta}) \quad \text{with } \widetilde{l}_t(\boldsymbol{\eta}) = \log P\big(X_t|\widetilde{\alpha}_t(\boldsymbol{\eta}), \mathcal{D}_{t-1}\big),$$

where $\{\widetilde{\alpha}_t\}$ is the time-varying parameter. Let $\widetilde{\boldsymbol{\eta}}_T^{\text{cml}}$ be a solution of the corresponding score equations $\frac{\partial \widetilde{\ell}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = 0$. Similar to Theorem 4, $\widetilde{\boldsymbol{\eta}}_T^{\text{cml}}$ is consistent and asymptotic normal. Then the consistency and asymptotic normality of $\hat{\boldsymbol{\eta}}_T^{\text{cml}}$ is constructed via the relation between $\hat{\boldsymbol{\eta}}_T^{\text{cml}}$ and $\widetilde{\boldsymbol{\eta}}_T^{\text{cml}}$. The following theorem gives the strong consistency and the asymptotic properties of the CML estimator.

Theorem 6. *Let $\{X_t, t \in \mathbb{Z}\}$ be a stationary and ergodic sequence and assumptions 1, 4, 5 hold. Then, as $T \to \infty$,*

(1) *there exists an estimator $\hat{\boldsymbol{\eta}}_T^{\text{cml}}$ such that $\hat{\boldsymbol{\eta}}_T^{\text{cml}} \xrightarrow{a.s.} \boldsymbol{\eta}_0$;*
(2) $\sqrt{T}(\hat{\boldsymbol{\eta}}_T^{\text{cml}} - \boldsymbol{\eta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{I}^{-1}(\boldsymbol{\eta}_0))$, *where* $\boldsymbol{I}(\boldsymbol{\eta}_0) = E\Big[\frac{\partial l_t(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \frac{\partial l_t(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^\top}\Big]_{\boldsymbol{\eta}_0}$.

4 Simulation studies

A simulation study is conducted to check the finite-sample performance of the two estimation methods for the logit-BARCH(p) model and the score-BARCH(1) model considered in the previous section. To estimate the parameter vector $(r_0, r_1, \ldots, r_p)$ for logit-BARCH(p) model and (w, β, τ) for score-BARCH(1) model, we first choose some values of $r_0 \in [-2, 2]$ for logit-BARCH(p) model and $w \in [-2, 2]$ for score-BARCH(1) model to guarantee thinning probability α_t changes over $(0.1, 0.9)$. Considering that α_t is updated by past observations, then, we choose some values of $r_1, \ldots, r_p$ for logit-BARCH(p) model and β for score-BARCH(1) model in $[-0.3, 0.3]$ to guarantee the slow change of the gradient of α_t. Last, we choose some values of τ for score-BARCH(1) model under Assumption 1. Among these, some representative parameter combinations are listed as follows:

- logit-BARCH(1) model with $(r_0, r_1)^\top =$ (A1): $(-1, 0.1)$, (A2): $(-1, 0.2)$, (A3): $(-1, -0.1)$, (A4): $(-1, -0.2)$, (A5): $(1, 0.1)$, (A6): $(1, -0.1)$, (A7): $(1, -0.2)$
- logit-BARCH(2) model with $(r_0, r_1, r_2)^\top =$ (B1): $(-1, -0.1, -0.1)$, (B2): $(-1, -0.1, 0.1)$, (B3): $(-1, 0.1, -0.1)$, (B4): $(-2, 0.1, 0.1)$, (B5): $(1, -0.1, -0.1)$, (B6): $(1, -0.1, 0.1)$, (B7): $(1, 0.1, 0.1)$
- score-BARCH(1) model with $(w, \beta, \tau)^\top =$ (C1): $(-1, -0.1, -0.1)$, (C2): $(-1, -0.1, 0.1)$, (C3): $(-1, 0.1, -0.1)$, (C4): $(-1, 0.1, 0.1)$, (C5): $(0.5, 0.3, 0.1)$, (C6): $(1, 0.3, -0.1)$, (C7): $(1, -0.3, 0.1)$

Here, we fix $n = 20$ and use the `optim` function in `R` for the optimization of the functions (3.1) and (3.2), and (3.3) and (3.4). The size of sample is 100, 200, 500 and 1000 and we use 10,000 replications. For the simulated sample, performances of mean and standard deviation (sd) are given. For a scale parameter φ, sd $= \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(\hat{\varphi}_i - \varphi)^2}$, where $\hat{\varphi}_i$ is the estimator of φ in the ith replication and m denotes repetition times.

A summary of the simulation results are given in Tables 1 to 3, which represent the logit-BARCH(1), logit-BARCH(2) and score-BARCH(1) models, respectively.

These studies indicate that both of the estimation methods seem to perform reasonably well, but the CML gives smaller standard deviations than those of the CLS, and the means

Table 1 *Mean and sd in parentheses of estimates for* logit-BARCH(1) *model*

Method:		CML				CLS			
	T:	100	200	500	1000	100	200	500	1000
(A1)	r_0	−0.9529	−0.9764	−0.9896	−0.9961	−0.9535	−0.9768	−0.9898	−0.9962
		(0.1863)	(0.1328)	(0.0835)	(0.0582)	(0.1866)	(0.1330)	(0.0836)	(0.0583)
	r_1	0.0952	0.0976	0.0990	0.0996	0.0953	0.0976	0.0990	0.0996
		(0.0181)	(0.0129)	(0.0081)	(0.0056)	(0.0181)	(0.0129)	(0.0081)	(0.0056)
(A2)	r_0	−0.9242	−0.9427	−0.9688	−0.9782	−0.9139	−0.9311	−0.9590	−0.9727
		(0.4454)	(0.4270)	(0.3891)	(0.3446)	(0.4438)	(0.4279)	(0.3928)	(0.3498)
	r_1	0.1962	0.1971	0.1984	0.1989	0.1957	0.1965	0.1979	0.1986
		(0.0245)	(0.0232)	(0.0210)	(0.0186)	(0.0239)	(0.0230)	(0.0210)	(0.0188)
(A3)	r_0	−0.9973	−0.9992	−0.9987	−1.0008	−0.9966	−0.9988	−0.9985	−0.1008
		(0.1276)	(0.0911)	(0.0580)	(0.0408)	(0.1288)	(0.0920)	(0.0585)	(0.0412)
	r_1	−0.1012	−0.1005	−0.1003	−0.0998	−0.1014	−0.1006	−0.1004	−0.0998
		(0.0304)	(0.0216)	(0.0138)	(0.0097)	(0.0308)	(0.0218)	(0.0140)	(0.0098)
(A4)	r_0	−1.0044	−1.0018	−1.0002	−1.0009	−1.0030	−1.0012	−1.0000	−1.0007
		(0.1176)	(0.0818)	(0.0519)	(0.0371)	(0.1214)	(0.0849)	(0.0536)	(0.0381)
	r_1	−0.1994	−0.1998	−0.2002	−0.1997	−0.2001	−0.2000	−0.2003	−0.1998
		(0.0352)	(0.0245)	(0.0154)	(0.0109)	(0.0373)	(0.0259)	(0.0162)	(0.0115)
(A5)	r_0	1.0475	1.0405	1.0296	1.0142	1.0502	1.0434	1.0314	1.0150
		(0.4573)	(0.4443)	(0.4106)	(0.3767)	(0.4573)	(0.4448)	(0.4120)	(0.3789)
	r_1	0.0977	0.0980	0.0985	0.0993	0.0975	0.0978	0.0984	0.0992
		(0.0248)	(0.0238)	(0.0218)	(0.0200)	(0.0245)	(0.0236)	(0.0218)	(0.0201)
(A6)	r_0	0.9949	0.9977	0.9976	0.9990	0.9957	0.9980	0.9977	0.9990
		(0.1817)	(0.1311)	(0.0819)	(0.0585)	(0.1820)	(0.1312)	(0.0819)	(0.0586)
	r_1	−0.0996	−0.0998	−0.0998	−0.0999	−0.0996	−0.0998	−0.0998	−0.0999
		(0.0176)	(0.0128)	(0.0079)	(0.0057)	(0.0177)	(0.0128)	(0.0079)	(0.0057)
(A7)	r_0	0.9885	0.9921	0.9987	0.9983	0.9909	0.9934	0.9991	0.9985
		(0.1132)	(0.0782)	(0.0481)	(0.0348)	(0.1151)	(0.0793)	(0.0486)	(0.0353)
	r_1	−0.1985	−0.1990	−0.1998	−0.1998	−0.1989	−0.1992	−0.1999	−0.1998
		(0.0145)	(0.0101)	(0.0062)	(0.0044)	(0.0149)	(0.0104)	(0.0063)	(0.0045)

of CML are closer to the true parameter values than those of the CLS in most cases. As the sample size increases, the estimates seem to converge to the true parameter values.

To illustrate the location and dispersion of the estimates, we present the boxplots of the estimates for the parameter combinations (A1), (B1) and (C1) in Figures 3 to 5, others are similar. Figures 3–5 illustrate the large-sample properties of the estimators on a limited sample size.

In general, the estimated medians are apparently closer to the real parameter values with the sample size increase. Regarding dispersion issues, both the interquartile ranges and the overall range of the produced values become narrower with the sample size increase, which indicates the consistency of the CML and CLS estimators.

5 Real data examples

In this section, we consider the possible applications of the dynamic binomial integer-valued ARCH models in the field of biostatistics and meteorology. We use the binomial logit-ARCH(p) and score-ARCH(1) models, BARCH(p) model (Ristić, Weiß and Janjić, 2016) and binomial GARCH(1, 1) model (BGARCH(1, 1)) (Lee and Lee, 2019) to fit the

Table 2 *Mean and sd in parentheses of estimates for* logit-BARCH(2) *model*

Method:		CML				CLS			
	T:	100	200	500	1000	100	200	500	1000
(B1)	r_0	−0.9827	−0.9923	−0.9969	−0.9986	−0.9820	−0.9913	−0.9965	−0.9986
		(0.1740)	(0.1275)	(0.0819)	(0.0581)	(0.1786)	(0.1308)	(0.0836)	(0.0593)
	r_1	−0.1020	−0.1010	−0.1006	−0.1002	−0.1021	−0.1011	−0.1007	−0.1002
		(0.0366)	(0.0259)	(0.0163)	(0.0116)	(0.0374)	(0.0263)	(0.0166)	(0.0117)
	r_2	−0.1045	−0.1021	−0.1008	−0.1004	−0.1048	−0.1023	−0.1009	−0.1004
		(0.0317)	(0.0240)	(0.0157)	(0.0113)	(0.0335)	(0.0251)	(0.0162)	(0.0116)
(B2)	r_0	−0.9404	−0.9706	−0.9887	−0.9929	−0.9403	−0.9697	−0.9884	−0.9926
		(0.2350)	(0.1717)	(0.1067)	(0.0755)	(0.2385)	(0.1746)	(0.1083)	(0.0766)
	r_1	−0.1034	−0.1016	−0.1007	−0.1004	−0.1038	−0.1018	−0.1008	−0.1004
		(0.0248)	(0.0176)	(0.0110)	(0.0078)	(0.0255)	(0.0181)	(0.0112)	(0.0080)
	r_2	0.0925	0.0963	0.0985	0.0991	0.0927	0.0963	0.0985	0.0991
		(0.0214)	(0.0160)	(0.0101)	(0.0072)	(0.0217)	(0.0162)	(0.0103)	(0.0073)
(B3)	r_0	−0.9823	−0.9879	−0.9972	−0.9980	−0.9796	−0.9866	−0.9966	−0.9976
		(0.1536)	(0.1125)	(0.0711)	(0.0505)	(0.1544)	(0.1128)	(0.0716)	(0.0508)
	r_1	0.0971	0.0982	0.0993	0.0996	0.0973	0.0983	0.0994	0.0996
		(0.0236)	(0.0168)	(0.0106)	(0.0074)	(0.0238)	(0.0169)	(0.0107)	(0.0075)
	r_2	−0.1005	−0.1004	−0.1000	−0.1000	−0.1014	−0.1008	−0.1002	−0.1001
		(0.0227)	(0.0164)	(0.0107)	(0.0076)	(0.0232)	(0.0167)	(0.0108)	(0.0077)
(B4)	r_0	−1.8816	−1.9433	−1.9807	−1.9912	−1.8834	−1.9448	−1.9814	−1.9915
		(0.1722)	(0.1147)	(0.0662)	(0.0454)	(0.1730)	(0.1155)	(0.0666)	(0.0458)
	r_1	0.0957	0.0982	0.0995	0.0998	0.0960	0.0984	0.0995	0.0998
		(0.0219)	(0.0148)	(0.0092)	(0.0065)	(0.0222)	(0.0144)	(0.0092)	(0.0065)
	r_2	0.0912	0.0959	0.0985	0.0993	0.0912	0.0959	0.0985	0.0993
		(0.0211)	(0.0143)	(0.0092)	(0.0065)	(0.0213)	(0.0144)	(0.0092)	(0.0065)
(B5)	r_0	0.9984	1.0015	0.9985	1.0000	1.0010	1.0032	0.9992	1.0005
		(0.2286)	(0.1695)	(0.1079)	(0.0772)	(0.2301)	(0.1705)	(0.1084)	(0.0775)
	r_1	−0.0997	−0.1001	−0.0999	−0.1000	−0.0999	−0.1002	−0.0999	−0.1001
		(0.0192)	(0.0140)	(0.0088)	(0.0063)	(0.0193)	(0.0141)	(0.0088)	(0.0063)
	r_2	−0.1002	−0.1001	−0.1000	−0.1000	−0.1004	−0.1002	−0.1000	−0.1001
		(0.0181)	(0.0134)	(0.0086)	(0.0062)	(0.0182)	(0.0135)	(0.0087)	(0.0062)
(B6)	r_0	1.0641	1.0500	1.0272	1.0153	1.0646	1.0529	1.0289	1.0163
		(0.3695)	(0.3264)	(0.2533)	(0.1911)	(0.3736)	(0.3306)	(0.2579)	(0.1946)
	r_1	−0.0994	−0.1002	−0.1003	−0.1002	−0.0994	−0.1003	−0.1004	−0.1003
		(0.0181)	(0.0145)	(0.0104)	(0.0077)	(0.0183)	(0.0147)	(0.0106)	(0.0079)
	r_2	0.0949	0.0967	0.0985	0.0992	0.0949	0.0967	0.0984	0.0992
		(0.0168)	(0.0133)	(0.0098)	(0.0072)	(0.0173)	(0.0136)	(0.0099)	(0.0073)
(B7)	r_0	1.1302	1.1089	1.0792	1.0675	1.1422	1.1269	1.0897	1.0770
		(0.4549)	(0.4540)	(0.4460)	(0.4423)	(0.4497)	(0.4470)	(0.4434)	(0.4370)
	r_1	0.1115	0.1114	0.1108	0.1096	0.1249	0.1132	0.1120	0.1117
		(0.0679)	(0.0678)	(0.0666)	(0.0652)	(0.1008)	(0.0704)	(0.0630)	(0.0620)
	r_2	0.0845	0.0843	0.0858	0.0874	0.1044	0.0860	0.0841	0.0848
		(0.0701)	(0.0679)	(0.0654)	(0.0631)	(0.1072)	(0.0748)	(0.0670)	(0.0659)

data by the CML method. We compare their approximated standard error (SE), Bayesian information criterion (BIC), where the approximated standard error is computed by using the matrix $\boldsymbol{I}^{-1}(\boldsymbol{\theta}_0)$ (or $\boldsymbol{I}^{-1}(\boldsymbol{\eta}_0)$), see Theorem 4 (or Theorem 6) for detail.

Note that the number of summation terms in the log-likelihood decreases with the increasing of order p, which affects the values of the information criteria BIC. To correct for this issue, we follow the suggestion in Weiß (2018) to compute the values of BIC (denoted as

Table 3 *Mean and sd in parentheses of estimates for* score-BARCH(1) *model*

Method:		CML				CLS			
	T:	100	200	500	1000	100	200	500	1000
(C1)	ω	−0.9611	−0.9801	−1.0038	−0.9968	−0.9614	−0.9804	−1.0040	−0.9969
		(0.2231)	(0.1614)	(0.0855)	(0.0715)	(0.2242)	(0.1620)	(0.0857)	(0.0717)
	β	−0.0547	−0.0772	−0.1038	−0.0964	−0.0550	−0.0775	−0.1040	−0.0965
		(0.2380)	(0.1725)	(0.0906)	(0.0769)	(0.2392)	(0.1731)	(0.0909)	(0.0771)
	τ	−0.1029	−0.1014	−0.1006	−0.1003	−0.1030	−0.1014	−0.1006	−0.1003
		(0.0254)	(0.0176)	(0.0112)	(0.0078)	(0.0257)	(0.0178)	(0.0113)	(0.0079)
(C2)	ω	−1.0405	−1.0192	−1.0182	−1.0044	−1.0408	−1.0194	−1.0183	−1.0045
		(0.2224)	(0.1618)	(0.0893)	(0.0718)	(0.2230)	(0.1624)	(0.0896)	(0.0719)
	β	−0.1433	−0.1204	−0.1196	−0.1045	−0.1437	−0.1205	−0.1198	−0.1046
		(0.2388)	(0.1727)	(0.0950)	(0.0768)	(0.2395)	(0.1732)	(0.0952)	(0.0769)
	τ	0.0984	0.0990	0.0996	0.0998	0.0984	0.0990	0.0996	0.0998
		(0.0248)	(0.0175)	(0.0110)	(0.0079)	(0.0250)	(0.0177)	(0.0111)	(0.0079)
(C3)	ω	−0.9389	−0.9649	−0.9746	−0.9951	−0.9388	−0.9650	−0.9746	−0.9951
		(0.2893)	(0.2150)	(0.1161)	(0.0953)	(0.2903)	(0.2161)	(0.1164)	(0.0956)
	β	0.1570	0.1326	0.1231	0.1049	0.1572	0.1326	0.1232	0.1049
		(0.2566)	(0.1911)	(0.1024)	(0.0845)	(0.2575)	(0.1922)	(0.1027)	(0.0848)
	τ	−0.1035	−0.1015	−0.1007	−0.1003	−0.1037	−0.1016	−0.1007	−0.1003
		(0.0279)	(0.0197)	(0.0125)	(0.0087)	(0.0282)	(0.0198)	(0.0126)	(0.0087)
(C4)	ω	−1.0549	−1.0338	−0.9947	−1.0077	−1.0555	−1.0344	−0.9947	−1.0079
		(0.2788)	(0.2133)	(0.1096)	(0.0946)	(0.2795)	(0.2140)	(0.1098)	(0.0949)
	β	0.0521	0.0706	0.1051	0.0933	0.0515	0.0701	0.1050	0.0931
		(0.2476)	(0.1892)	(0.0968)	(0.0837)	(0.2484)	(0.1898)	(0.0970)	(0.0840)
	τ	0.0976	0.0991	0.0995	0.0997	0.0977	0.0991	0.0995	0.0997
		(0.0273)	(0.0188)	(0.0119)	(0.0084)	(0.0275)	(0.0190)	(0.0120)	(0.0085)
(C5)	ω	0.5423	0.5203	0.5074	0.5041	0.5424	0.5204	0.5074	0.5042
		(0.1607)	(0.1142)	(0.0725)	(0.0515)	(0.1610)	(0.1144)	(0.0726)	(0.0516)
	β	0.2421	0.2723	0.2894	0.2944	0.2420	0.2723	0.2894	0.2944
		(0.2137)	(0.1532)	(0.0974)	(0.0693)	(0.2141)	(0.1535)	(0.0976)	(0.0694)
	τ	0.0978	0.0987	0.0995	0.0997	0.0979	0.0987	0.0995	0.0997
		(0.0233)	(0.0166)	(0.0104)	(0.0072)	(0.0234)	(0.0166)	(0.0104)	(0.0072)
(C6)	ω	0.9089	0.9446	0.9763	0.9882	0.9065	0.9430	0.9757	0.9878
		(0.3544)	(0.2903)	(0.1929)	(0.1398)	(0.3563)	(0.2918)	(0.1945)	(0.1407)
	β	0.3659	0.3395	0.3169	0.3085	0.3675	0.3406	0.3174	0.3088
		(0.2466)	(0.2025)	(0.1347)	(0.0973)	(0.2477)	(0.2035)	(0.1359)	(0.0979)
	τ	−0.1046	−0.1023	−0.1010	−0.1004	−0.1050	−0.1025	−0.1011	−0.1004
		(0.0335)	(0.0232)	(0.0148)	(0.0102)	(0.0340)	(0.0235)	(0.0149)	(0.0104)
(C7)	ω	1.0315	1.0187	1.0065	1.0037	1.0312	1.0186	1.0066	1.0037
		(0.1730)	(0.1260)	(0.0800)	(0.0572)	(0.1732)	(0.1264)	(0.0803)	(0.0574)
	β	−0.3393	−0.3234	−0.3084	−0.3051	−0.3390	−0.3233	−0.3084	−0.3051
		(0.2118)	(0.1563)	(0.0990)	(0.0711)	(0.2122)	(0.1568)	(0.0994)	(0.0713)
	τ	0.0985	0.0994	0.0997	0.0997	0.0986	0.0994	0.0997	0.0997
		(0.0234)	(0.0167)	(0.0104)	(0.0074)	(0.0235)	(0.0168)	(0.0105)	(0.0074)

BIC*) by multiplying the maximized log-likelihood by the factor $T/(T-p)$, see Weiß and Feld (2019) for more detail. In addition, we also evaluate the performance by comparing the root mean square error (RMSE) of above models, where

$$\text{RMSE} = \sqrt{\frac{1}{(T-p)} \sum_{t=p+1}^{T} (X_t - n\hat{\alpha}_t)^2},$$

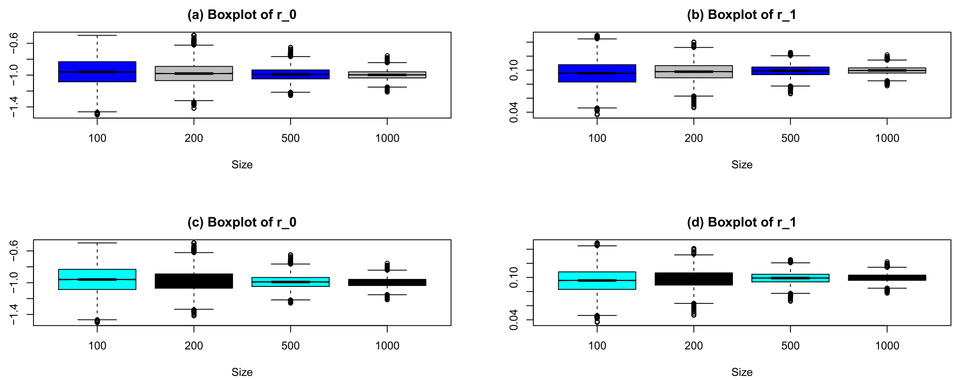

Figure 3 (a)–(b): *Boxplots of CML estimates for* (A1), (d)–(e): *boxplots of CLS estimates for* (A1).

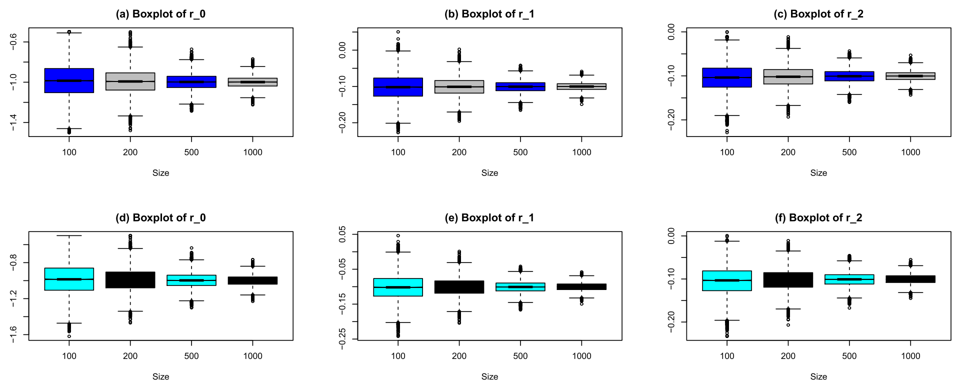

Figure 4 (a)–(c): *Boxplots of CML estimates for* (B1), (d)–(f): *boxplots of CLS estimates for* (B1).

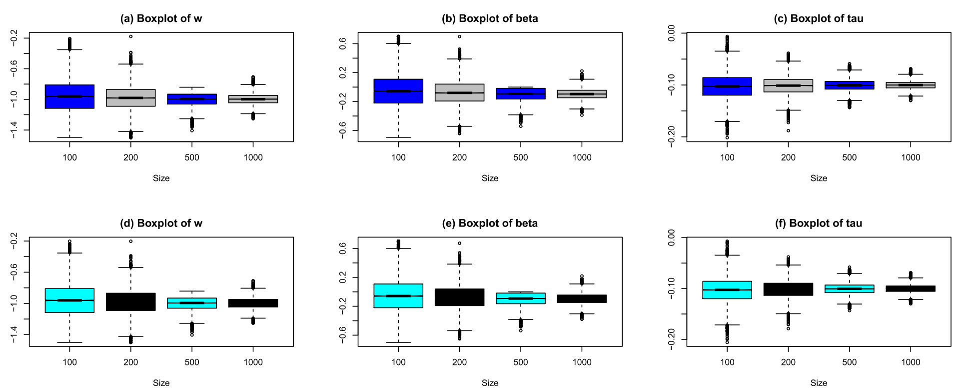

Figure 5 (a)–(c): *Boxplots of CML estimates for* (C1), (d)–(f): *boxplots of CLS estimates for* (C1).

with

- $\hat{\alpha}_t = \hat{a}_0 + \sum_{t=p+1}^{T} \hat{a}_k X_{t-k}/n$ for BARCH(p) model (Ristić, Weiß and Janjić, 2016),
- $\hat{\alpha}_t = \hat{w} + \hat{w}_1\hat{\alpha}_{t-1} + \hat{w}_2 X_{t-1}/n$ for BGARCH(1, 1) (Lee and Lee, 2019),
- $\hat{\alpha}_t = \frac{\exp(\hat{r}_0+\sum_{t=p+1}^{T}\hat{r}_k X_{t-k})}{1+\exp(\hat{r}_0+\sum_{t=p+1}^{T}\hat{r}_k X_{t-k})}$ for logit-BARCH(p) model,
- $\hat{\alpha}_t = \frac{\exp(\sum_{t=2}^{T}\hat{\omega}+\hat{\beta}\,\mathrm{logit}(\hat{\alpha}_{t-1})+\hat{\tau}s_t)}{1+\exp(\sum_{t=2}^{T}\hat{\omega}+\hat{\beta}\,\mathrm{logit}(\hat{\alpha}_{t-1})+\hat{\tau}s_t)}$ for score-BARCH(1) model.

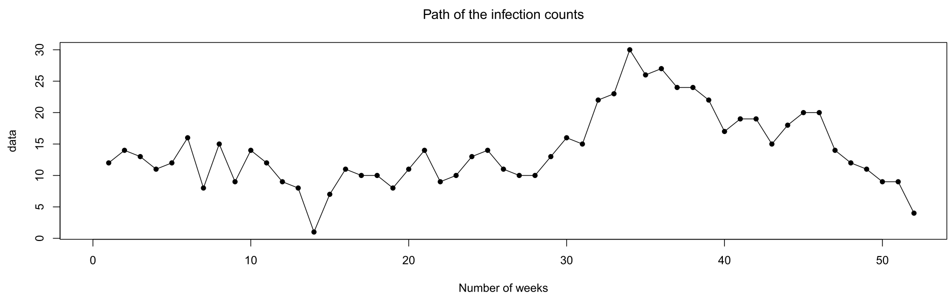

Figure 6 *Path of the infection counts.*

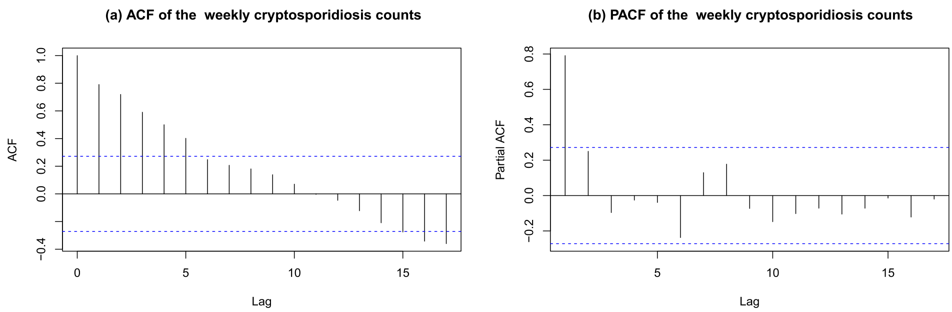

Figure 7 *Infection counts*: (a) *the ACF* (b) *the PACF.*

5.1 Infection counts

In this section, we consider the number of districts with new cases of cryptosporidiosis infections per week of the year 2013 reported in $n = 38$ Germany's districts. The dataset is taken from the "SurvStat" data (https://survstat.rki.de/Content/Query/Main.aspx), which has been reported to the Robert-Koch-Institut via local and state health departments. The length of the data is 52 and the sample mean and variance are 14.0577 and 35.8986, respectively. The sample path and the ACF, PACF plots of the observations are given in Figures 6 and 7, respectively.

Figure 7 shows that an autoregressive model of order $p \leq 2$ appears to be reasonable. The CML estimates and approximated standard errors of parameters, including the fitted values of BIC* and RMSE, are summarized in Table 4.

From Table 4, we have the following observations. For $p = 1$, the logit-BARCH(1) model takes the smallest BIC* and RMSE. For $p = 2$, the logit-BARCH(p) model takes the smaller BIC* and RMSE than those of the corresponding BARCH(p) model. For all the considered models, the logit-BARCH(2) model takes the smallest BIC* and RMSE. Hence, the logit-BARCH(2) model is more suitable for the analysed dataset.

To further check the adequacy of this model, we first use the parametric bootstrap based on the fitted model (Tsay, 1992), which was also considered in Weiß (2015), Ristić, Weiß and Janjić (2016) and Ristić and Popović (2019). For parameter values $r_0 = -2.0215$, $r_1 = 0.0747$ and $r_2 = 0.0275$, we simulate 10,000 samples of size $T = 52$ from the logit-BARCH(2) model. For each simulated sample, we compute the sample ACF, and for each fixed lag, we derive the 2.5% and 97.5% quantiles and draw the bootstrap confidence intervals in Figure 8.

From this figure, we can conclude that the logit-BARCH(2) model adequately describes the autocorrelation structure of the infection counts. Second, we analyze the Pearson residuals

Table 4 *Estimates for the cryptosporidiosis infection counts, SE are shown in parentheses*

Model	estimates			BIC*	RMSE
	$\hat{w}$	$\hat{w}_1$	$\hat{w}_2$		
BGARCH(1, 1)	0.0718	0.0299	0.7743	275.0295	3.5336
	(0.0293)	(0.1281)	(0.1008)		
	$\hat{w}$	$\hat{\beta}$	$\hat{\tau}$		
score-BARCH(1)	−0.0826	0.8610	0.0865	273.8296	3.4462
	(0.0656)	(0.0823)	(0.0149)		
	$\hat{a}_0$	$\hat{a}_1$			
BARCH(1)	0.0759	0.7929		277.1845	3.5137
	(0.0232)	(0.0608)			
	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$		
BARCH(2)	0.0497	0.6092	0.2485	266.9672	3.4258
	(0.0263)	(0.1065)	(0.1101)		
	$\hat{r}_0$	$\hat{r}_1$			
logit-BARCH(1)	−1.9417	0.0970		273.0364	3.4322
	(0.1361)	(0.0086)			
	$\hat{r}_0$	$\hat{r}_1$	$\hat{r}_2$		
logit-BARCH(2)	−2.0215	0.0747	0.0275	264.1300	3.3701
	(0.1421)	(0.0148)	(0.0147)		

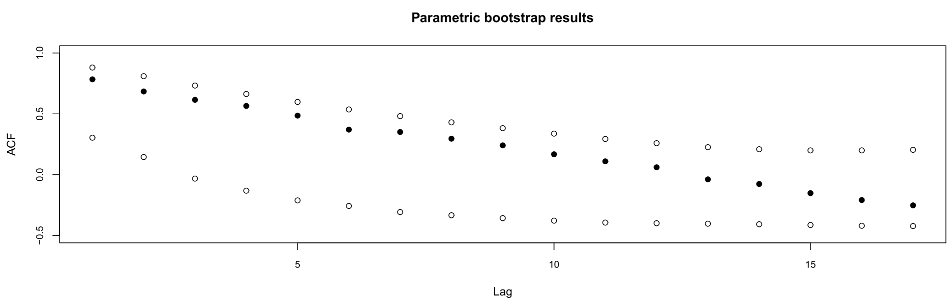

Figure 8 *ACF for cryptosporidiosis infection counts with* 95% *bootstrap confidence intervals.*

of the logit-BARCH(2) model, which is defined by

$$e_t = \frac{X_t - n\hat{\alpha}_t}{\sqrt{n\hat{\alpha}_t(1-\hat{\alpha}_t)}} \quad \text{with } \hat{\alpha}_t = \frac{\exp(\hat{r}_0 + \sum_{t=p+1}^{T} \hat{r}_k X_{t-k})}{1+\exp(\hat{r}_0 + \sum_{t=p+1}^{T} \hat{r}_k X_{t-k})}. \tag{5.1}$$

The mean and variance of Pearson residuals of the fitted logit-BARCH(2) model are 0.0029 and 1.3886, respectively. The residual analysis in Figure 9 shows that this model does rather well.

Thus, we conclude that our logit-BARCH(2) model is an adequate model for cryptosporidiosis counts.

Remark 1. Note that the bootstrap confidence intervals are rather wide in Figure 8, but this was not surprising because the sample size is small in the analyzed data. An intuitive idea is to analyze the number of districts with new cases of cryptosporidiosis infection per week in two or more years, but cryptosporidiosis is known to be a seasonal infection, see Current and Garcia (1991). To model the seasonal data, one can consider the seasonal log-linear model

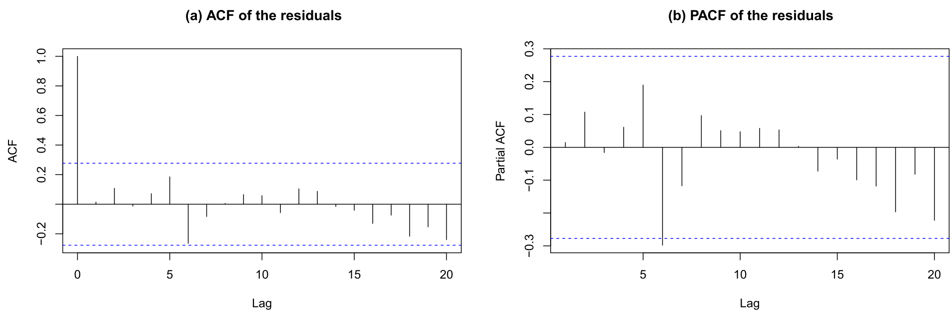

Figure 9 *Pearson residual analysis of cryptosporidiosis infection counts* (a) *ACF of the residuals* (b) *PACF of the residuals*.

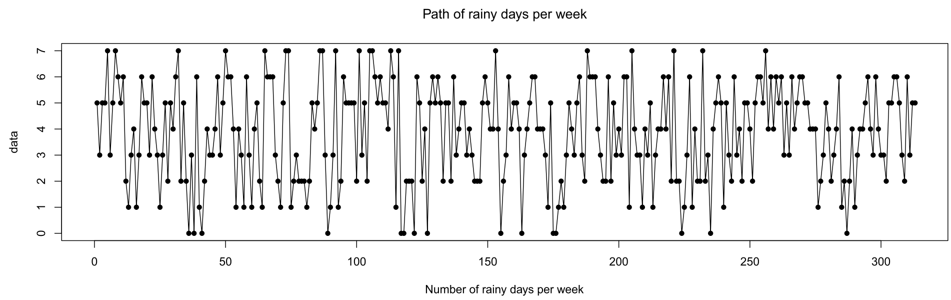

Figure 10 *Path of the rainy days in Hamburg-Neuwiedenthal*.

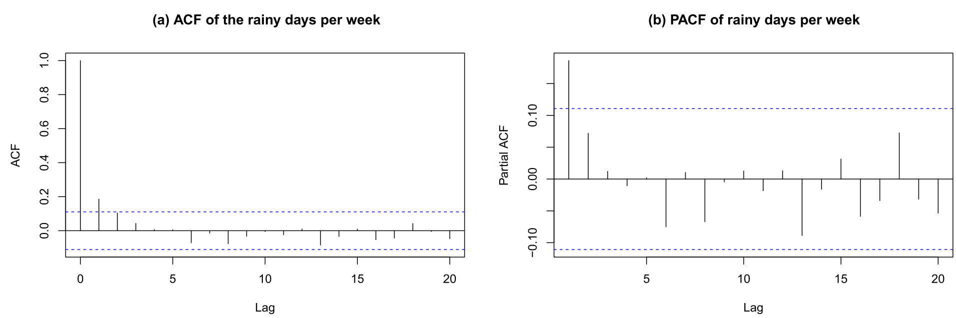

Figure 11 *Rainy days in Hamburg-Neuwiedenthal*: (a) *the ACF* (b) *the PACF*.

in Höhle and Paul (2008), see Weiß (2018) and Weiß and Feld (2019) for more discussion. A relevant future study will be considered to model data with seasonal and non-stationary characteristics.

5.2 Rainy days time series

In this section, we consider the number of rainy days per week at Hamburg-Neuwiedenthal in Germany. The data was collected from January 1st 2005 till December 31th 2010 by the German Weather Service (DWD = "Deutscher WetterDienst", http://www.dwd.de/), where weeks are defined from Saturday to Friday and $n = 7$. The length of the data is 313 and the sample mean and variance are 3.8371 and 3.6753, respectively. The sample path and the ACF, PACF plots of the observations are given in Figures 10 and 11, respectively.

Figure 11 shows that an autoregressive model of order $p \leq 2$ appears to be reasonable. The CML estimates and approximated standard errors of parameters, including the fitted values of BIC* and RMSE, are summarized in Table 5.

Table 5 *Estimates for the rainy days in Hamburg-Neuwiedenthal, SE are shown in parentheses*

Model	estimates			BIC*	RMSE
	$\hat{w}$	$\hat{w}_1$	$\hat{w}_2$		
BGARCH(1, 1)	0.2308	0.4064	0.1724	1376.8676	1.8820
	(0.1136)	(0.2099)	(0.0391)		
	$\hat{w}$	$\hat{\beta}$	$\hat{\tau}$		
score-BARCH(1)	0.0837	0.5721	0.1002	1376.9969	1.8814
	(0.0585)	(0.2090)	(0.0230)		
	$\hat{a}_0$	$\hat{a}_1$			
BARCH(1)	0.4460	0.1853		1379.2257	1.8824
	(0.0238)	(0.0384)			
	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$		
BARCH(2)	0.4132	0.1736	0.0726	1374.1191	1.8801
	(0.0296)	(0.0392)	(0.0392)		
	$\hat{r}_0$	$\hat{r}_1$			
logit-BARCH(1)	−0.2210	0.1080		1379.1103	1.8823
	(0.0962)	(0.0226)			
	$\hat{r}_0$	$\hat{r}_1$	$\hat{r}_2$		
logit-BARCH(2)	−0.3524	0.1010	0.0419	1374.0993	1.8801
	(0.1203)	(0.0230)	(0.0230)		

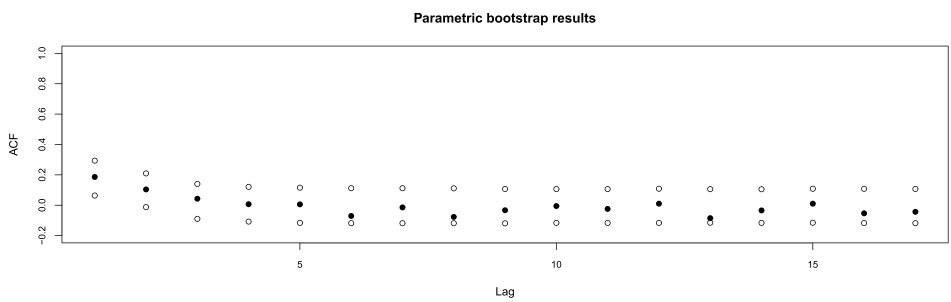

Figure 12 *ACF for rainy days counts with* 95% *bootstrap confidence intervals.*

From Table 5, we have the following observations. For $p = 1$, the score-BARCH(1) model takes the smallest RMSE, while its BIC* is slightly greater than that of BGARCH(1, 1) model. For $p = 2$, the logit-BARCH(2) and BARCH(2) models take the same value of the RMSE, but the logit-BARCH(2) model takes the smaller BIC* than that of the corresponding BARCH(2) model. For all the considered models, the logit-BARCH(2) and BARCH(2) models take the same minimum value of the RMSE, but the logit-BARCH(2) model takes the smallest BIC*. Hence, the logit-BARCH(2) model is more suitable for the analyzed dataset.

To further check the adequacy of this model, we first use the parametric bootstrap based on the fitted logit-BARCH(2) model. For parameter values $r_0 = -0.3524$, $r_1 = 0.1010$ and $r_2 = 0.0419$, we simulate 10,000 samples of size $T = 313$ from the logit-BARCH(2) model. For each simulated sample, we compute the sample ACF, and for each fixed lag, we derive the 2.5% and 97.5% quantiles and draw the bootstrap confidence intervals in Figure 12.

From this graph, we can conclude that the logit-BARCH(2) model adequately describes the autocorrelation structure of the rainy days counts. Second, we analyze the Pearson residual (defined by (5.1)) of the logit-BARCH(2) model. Its mean and variance are −0.0020 and

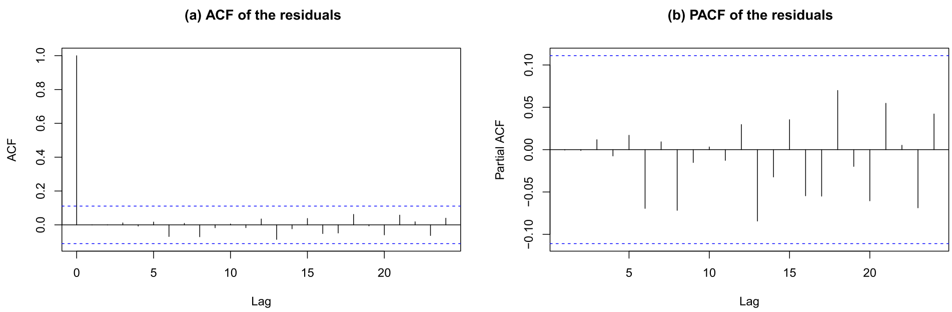

Figure 13 *Pearson residual analysis of rainy days counts in Hamburg-Neuwiedenthal* (a) *ACF of the residuals* (b) *PACF of the residuals.*

2.0609, respectively. The residual analysis in Figure 13 shows that this model does rather well.

Thus, we conclude that our logit-BARCH(2) model is an adequate model for the analyzed rainy days counts.

Remark 2. Note that the bootstrap confidence intervals are narrower in Figure 12 than that in Figure 8, which implies that increasing the sample size is a method to reduce the width of the bootstrap confidence interval.

Remark 3. For the two real datasets, our logit-BARCH(2) model performs best. But there exists a common drawback that this model can not fully capture the actual dispersion since the variances of the model's Pearson residuals are a little greater than one (1.3886 for the infection counts and 2.0609 for the rainy days counts). One possible solution to this problem is that we replace the binomial distribution to the beta-binomial distribution, see Weiß and Kim (2014) for the beta-binomial AR(1) model, which deserves a further study.

6 Concluding remarks

This paper considers two alternative approaches to construct dynamic binomial ARCH model, where time-varying coefficient α_t is updated at each time period by using past information. We discuss some properties of the models, the estimate of the parameters and its large-sample properties. Simulations are conducted to examine the finite sample performance of estimators. Real data examples are provided to illustrate the applicability of the new models.

After having defining the dynamic binomial ARCH models, there are a number of important issues for future research. As already discussed in Section 2, the autocorrelation function of the models should be treated in more detail. In Section 2.1, the higher order score-BARCH model would deserve a detailed analysis in a future project. Finally, one may also try to find alternative approaches for modeling bounded count time series to capture the additional dispersion besides the suggestion in Remark 3.

Appendix A: Proof of theorems

Proof of Theorem 1. The proof of Theorem 1 is similar to that of Theorem 1 in Ristić, Weiß and Janjić (2016) and we omit it. □

Proof of Theorem 2. Because the range of X_t is finite, any moments exist. According to (2.4), we have

$$\frac{\partial s_t(\alpha_{t-1})}{\partial \mathrm{logit}(\alpha_{t-1})} = -n\alpha_{t-1}(1-\alpha_{t-1}).$$

Hence,

$$E\log\left(\sup_{\eta\in\Theta}\left|\beta+\tau\frac{\partial s_t(\alpha_{t-1})}{\partial \mathrm{logit}(\alpha_{t-1})}\right|\right) \leq \log\left(\max\left(\left|\beta-\frac{1}{4}\tau n\right|,\left|\beta+\frac{1}{4}\tau n\right|,|\beta|\right)\right) < 0$$

holds by Assumption 1 and Lemma 5. Similar to the proof of Proposition 3.1 in Gorgi (2018), we obtain that the conditions in Theorem 2 of Wintenberger (2013) hold, hence, there exists a unique stationary and ergodic sequence $\{\widetilde{\alpha}_t\}$ such that the time-varying parameter $\{\alpha_t\}$ converges e.a.s. and uniformly to $\{\widetilde{\alpha}_t\}$, $t\to\infty$. Thus, the result of Theorem 2 hold. □

Proof of Corollary 1.

$$\begin{aligned}(1)\quad \|\alpha_t-\widetilde{\alpha}_t\|_\Omega &= \left\|\frac{\exp(\mathrm{logit}(\alpha_t))}{1+\exp(\mathrm{logit}(\alpha_t))}-\frac{\exp(\mathrm{logit}(\widetilde{\alpha}_t))}{1+\exp(\mathrm{logit}(\widetilde{\alpha}_t))}\right\|_\Omega\\ &\leq \frac{1}{4}\|\mathrm{logit}(\alpha_t)-\mathrm{logit}(\widetilde{\alpha}_t)\|_\Omega\end{aligned}$$

by the second assertion of Lemma 4. Hence, result (1) holds by Theorem 2.

The second result is directly obtained by Theorem 2. Then we prove the third assertion.

(3) In the following, we first prove (I) $:= |\mathrm{logit}(\alpha_t(\hat{\boldsymbol{\eta}}_l))-\mathrm{logit}(\widetilde{\alpha}_t(\boldsymbol{\eta}_0))| \xrightarrow{a.s.} 0$, if $\hat{\boldsymbol{\eta}}_l \xrightarrow{a.s.} \boldsymbol{\eta}_0$, $l\to\infty$. Note that

$$\text{(I)} \leq \underbrace{\left|\mathrm{logit}\big(\alpha_t(\hat{\boldsymbol{\eta}}_l)\big)-\mathrm{logit}\big(\widetilde{\alpha}_t(\hat{\boldsymbol{\eta}}_l)\big)\right|}_{\text{(II)}}+\underbrace{\left|\mathrm{logit}\big(\widetilde{\alpha}_t(\hat{\boldsymbol{\eta}}_l)\big)-\mathrm{logit}\big(\widetilde{\alpha}_t(\boldsymbol{\eta}_0)\big)\right|}_{\text{(III)}}. \tag{A.1}$$

According to Theorem 2, part (II) $\xrightarrow{a.s} 0$. In the following, we prove part (III) $\xrightarrow{a.s} 0$. Note that

$$\begin{cases}\mathrm{logit}\big(\widetilde{\alpha}_t(\hat{\boldsymbol{\eta}}_l)\big)=\hat{w}_l+\hat{\beta}_l\,\mathrm{logit}\big(\widetilde{\alpha}_{t-1}(\hat{\boldsymbol{\eta}}_l)\big)+\hat{\tau}_l\big(X_{t-1}-n\widetilde{\alpha}_{t-1}(\hat{\boldsymbol{\eta}}_l)\big),\\ \mathrm{logit}\big(\widetilde{\alpha}_t(\boldsymbol{\eta}_0)\big)=w_0+\beta_0\,\mathrm{logit}\big(\widetilde{\alpha}_{t-1}(\boldsymbol{\eta}_0)\big)+\tau_0\big(X_{t-1}-n\widetilde{\alpha}_{t-1}(\boldsymbol{\eta}_0)\big),\end{cases}$$

then, we have

$$\begin{aligned}\text{(III)} &\leq |\hat{w}_l-w_0|+|X_{t-1}||\hat{\tau}_l-\tau_0|\\ &\quad+\left|\hat{\beta}_l\,\mathrm{logit}\big(\widetilde{\alpha}_{t-1}(\hat{\boldsymbol{\eta}}_l)\big)-\beta_0\,\mathrm{logit}\big(\widetilde{\alpha}_{t-1}(\boldsymbol{\eta}_0)\big)\right|+n\left|\hat{\tau}_l\widetilde{\alpha}_{t-1}(\hat{\boldsymbol{\eta}}_l)-\tau_0\widetilde{\alpha}_{t-1}(\boldsymbol{\eta}_0)\right|\\ &\leq |\hat{w}_l-w_0|+|X_{t-1}||\hat{\tau}_l-\tau_0|+\left|\mathrm{logit}\big(\widetilde{\alpha}_{t-1}(\boldsymbol{\eta}_0)\big)\right||\hat{\beta}_l-\beta_0|\\ &\quad+n\left|\widetilde{\alpha}_{t-1}(\boldsymbol{\eta}_0)\right||\hat{\tau}_l-\tau_0|+\left(|\hat{\beta}_l|+\frac{n|\hat{\tau}_l|}{4}\right)\left|\mathrm{logit}\big(\widetilde{\alpha}_{t-1}(\hat{\boldsymbol{\eta}}_l)\big)-\mathrm{logit}\big(\widetilde{\alpha}_{t-1}(\boldsymbol{\eta}_0)\big)\right|\\ &\leq c|\hat{\boldsymbol{\eta}}_l-\boldsymbol{\eta}_0|+\left(|\hat{\beta}_l|+\frac{n|\hat{\tau}_l|}{4}\right)\left|\mathrm{logit}\big(\widetilde{\alpha}_{t-1}(\hat{\boldsymbol{\eta}}_l)\big)-\mathrm{logit}\big(\widetilde{\alpha}_{t-1}(\boldsymbol{\eta}_0)\big)\right|\\ &\leq\cdots\leq tc|\hat{\boldsymbol{\eta}}_l-\boldsymbol{\eta}_0|+\left(|\hat{\beta}_l|+\frac{n|\hat{\tau}_l|}{4}\right)^t\left|\mathrm{logit}\big(\widetilde{\alpha}_0(\hat{\boldsymbol{\eta}}_l)\big)-\mathrm{logit}\big(\widetilde{\alpha}_0(\boldsymbol{\eta}_0)\big)\right|\xrightarrow{a.s} 0,\end{aligned}$$

as $l\to\infty$, $t\to\infty$. Hence, (A.1) $\xrightarrow{a.s} 0$, as $l\to\infty$, $t\to\infty$. Then we prove the result:

$$\left|P_n\big(x|\alpha_t(\hat{\boldsymbol{\eta}}_l),\mathcal{D}_{t-1}\big)-P_n\big(x|\widetilde{\alpha}_t(\boldsymbol{\eta}_0),\mathcal{D}_{t-1}\big)\right|\xrightarrow{a.s.} 0,\quad \text{as } l\to\infty, t\to\infty.$$

Using the mean value theorem, there exists at least one $\alpha_t(\boldsymbol{\eta})$ lying in between $\alpha_t(\hat{\boldsymbol{\eta}}_l)$ and $\widetilde{\alpha}_t(\boldsymbol{\eta}_0)$ such that, as $l \to \infty$, $t \to \infty$,

$$\begin{aligned}&\left|P_n\big(x|\alpha_t(\hat{\boldsymbol{\eta}}_l), \mathcal{D}_{t-1}\big) - P_n\big(x|\widetilde{\alpha}_t(\boldsymbol{\eta}_0), \mathcal{D}_{t-1}\big)\right| \\ &\quad \le \left|X_t - n\alpha_t(\boldsymbol{\eta})\right|\left|\operatorname{logit}\big(\alpha_t(\hat{\boldsymbol{\eta}}_l)\big) - \operatorname{logit}\big(\widetilde{\alpha}_t(\boldsymbol{\eta}_0)\big)\right| \xrightarrow{a.s} 0,\end{aligned}$$

by the first result of (3) in Corollary 1 and Lemma 4. □

Proof of Theorem 3. According to Klimko and Nelson (1978), we split the proof into several intermediate results:

(i) Let $f(x) = \frac{\exp(x)}{1+\exp(x)}$. Using Lemma 4, we have $g_t(\boldsymbol{\theta}_0) = E(X_t|\mathcal{F}_{t-1}) = nf(\boldsymbol{Y}_t^\top \boldsymbol{\theta})$. Note that $f'(x) = f(x)(1 - f(x)) < f(x) \le 1$, $f''(x) = f'(x)(1 - 2f(x)) < f(x) \le 1$ and $f'''(x) = f'(x)[(1 - 2f(x))^2 - 2f'(x)] = f'(x)[1 + 6f(x)(1 - f(x))] < 7f(x) \le 7$. Then we have $\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial r_i}$, $\frac{\partial^2 g_t(\boldsymbol{\theta}_0)}{\partial r_i \partial r_j}$ and $\frac{\partial^3 g_t(\boldsymbol{\theta}_0)}{\partial r_i \partial r_j \partial r_k}$ exist and are continuous, $i, j, k = 0, 1, 2, \ldots, p$;

(ii) Let $u_t(\boldsymbol{\theta}_0) = X_t - g_t(\boldsymbol{\theta}_0)$, $\forall i, j = 0, 1, 2, \ldots, p$, $E|u_t(\boldsymbol{\theta}_0)\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial r_i}| < \infty$, $E|u_t(\boldsymbol{\theta}_0)\frac{\partial^2 g_t(\boldsymbol{\theta}_0)}{\partial r_i \partial r_j}| < \infty$ and $E|(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial r_i}\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial r_j})| < \infty$;

(iii) $\forall i, j, k = 0, 1, 2, \ldots, p$, there exist functions $H^{(0)}(X_0, X_1, \ldots, X_{t-p})$, $H_i^{(1)}(X_0, X_1, \ldots, X_{t-p})$, $H_{ij}^{(2)}(X_0, X_1, \ldots, X_{t-p})$, and $H_{ijk}^{(3)}(X_0, X_1, \ldots, X_{t-p})$ such that $|g_t(\boldsymbol{\theta}_0)| \le H^{(0)}$, $|\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial r_i}| \le H_i^{(1)}$, $|\frac{\partial^2 g_t(\boldsymbol{\theta}_0)}{\partial r_i \partial r_j}| \le H_{ij}^{(2)}$, $|\frac{\partial^3 g_t(\boldsymbol{\theta}_0)}{\partial r_i \partial r_j \partial r_k}| \le H_{ijk}^{(3)}$ and $E|X_t H_{ijk}^{(3)}| < \infty$, $E|H^{(0)} H_{ijk}^{(3)}| < \infty$, $E|H_i^{(1)} H_{ij}^{(2)}| < \infty$;

(iv) $E(u_t^2|\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial r_i}\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial r_j}|) < \infty$ and $E(X_t|X_0, X_1, \ldots, X_{t-p}, \ldots, X_{t-1}) \overset{a.s.}{=} E(X_t|\mathcal{F}_{t-1})$.

Theorem 3 holds if the above conditions are satisfied.

Obviously, $\frac{\partial E(X_t|\mathcal{F}_{t-1})}{\partial r_i}$, $\frac{\partial^2 E(X_t|\mathcal{F}_{t-1})}{\partial r_i \partial r_j}$ and $\frac{\partial^3 E(X_t|\mathcal{F}_{t-1})}{\partial r_i \partial r_j \partial r_k}$, $i, j, k = 0, 1, 2, \ldots, p$, exist and are continuous, that is, (i) holds. We also have

$$\begin{aligned}
\left|\frac{\partial E(X_t|\mathcal{F}_{t-1})}{\partial r_0}\right| &= nf(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big(1 - f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big) < n, \\
\left|\frac{\partial E(X_t|\mathcal{F}_{t-1})}{\partial r_j}\right| &= nX_{t-j} f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big(1 - f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big) < nX_{t-j}, \\
\left|\frac{\partial^2 E(X_t|\mathcal{F}_{t-1})}{\partial r_0^2}\right| &= \left|nf'(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big(1 - 2f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big)\right| < 3n, \\
\left|\frac{\partial^2 E(X_t|\mathcal{F}_{t-1})}{\partial r_0 \partial r_k}\right| &= \left|nX_{t-k} f'(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big(1 - 2f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big)\right| \le 3nX_{t-k}, \\
\left|\frac{\partial^2 E(X_t|\mathcal{F}_{t-1})}{\partial r_j \partial r_k}\right| &= \left|nX_{t-j}X_{t-k} f'(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big(1 - 2f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big)\right| \le 3nX_{t-j}X_{t-k}, \\
\left|\frac{\partial^3 E(X_t|\mathcal{F}_{t-1})}{\partial r_0^3}\right| &= \left|nf'(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big[1 + 6f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big(1 - f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big)\big]\right| < 7n, \\
\left|\frac{\partial^3 E(X_t|\mathcal{F}_{t-1})}{\partial r_0^2 \partial r_k}\right| &= \left|nX_{t-k} f'(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big[1 + 6f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big(1 - f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big)\big]\right| < 7nX_{t-k}, \\
\left|\frac{\partial^3 E(X_t|\mathcal{F}_{t-1})}{\partial r_0 \partial r_j \partial r_k}\right| &= \left|nX_{t-j}X_{t-k} f'(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big[1 + 6f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big(1 - f(\boldsymbol{Y}_t^\top \boldsymbol{\theta})\big)\big]\right| \\
&< 7nX_{t-j}X_{t-k},
\end{aligned}$$

$$\left|\frac{\partial^3 E(X_t|\mathcal{F}_{t-1})}{\partial r_i \partial r_j \partial r_k}\right| = \left|nX_{t-i}X_{t-j}X_{t-k}f'(\boldsymbol{Y}_t^\top\boldsymbol{\theta})[1+6f(\boldsymbol{Y}_t^\top\boldsymbol{\theta})(1-f(\boldsymbol{Y}_t^\top\boldsymbol{\theta}))]\right|$$
$$< 7nX_{t-i}X_{t-j}X_{t-k}.$$

Since the range of X_t is finite, any moments exist. Thus, conditions (ii)–(iv) are satisfied. Hence, Theorem 3 holds. □

Proof of Theorem 4. For convenience, we denote $l_t(\boldsymbol{\theta}) = \log P_n(X_t|\mathcal{F}_{t-1})$.

(1) We observe that $l_t(\boldsymbol{\theta})$ is a measurable function of X_t for all $\boldsymbol{\theta} \in \Theta$, and is continuous in an open and convex neighbourhood $N(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$, then there at least exists a point $\overline{\boldsymbol{\theta}} \in N(\boldsymbol{\theta}_0)$ such that $l_t(\boldsymbol{\theta})$ attains the maximum value at $\overline{\boldsymbol{\theta}}$. Thus,

$$E\Big(\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} l_t(\boldsymbol{\theta})\Big) = E\big(\log P_n(X_t|\mathcal{F}_{t-1})\big)_{\overline{\boldsymbol{\theta}}} \le \log E\big(P_n(X_t|\mathcal{F}_{t-1})\big)_{\overline{\boldsymbol{\theta}}} < \infty.$$

Note that $\{X_t\}$ is stationary and ergodic and in terms of Theorem 4.2.1 in Amemiya (1985), $\frac{1}{T}\sum_{t=1}^{T} l_t(\boldsymbol{\theta}) \to El_t(\boldsymbol{\theta})$ in probability as $T \to \infty$. By Jensen's inequality, we have

$$E\big(l_t(\boldsymbol{\theta})\big) - E\big(l_t(\boldsymbol{\theta}_0)\big) = E\log\frac{P_n(X_t|\mathcal{F}_{t-1})_{\boldsymbol{\theta}}}{P_n(X_t|\mathcal{F}_{t-1})_{\boldsymbol{\theta}_0}} \le 0. \tag{A.2}$$

Thus, $El_t(\boldsymbol{\theta})$ attains a strict local maximum at $\boldsymbol{\theta}_0$ and Proposition 2.

Hence, the conditions of Theorem 4.1.2 of Amemiya (1985) are fulfilled, thus part (1) holds.

(2) The proof of part 2 rests on the Taylor series expansion of the score vector around $\boldsymbol{\theta}_0$. We have

$$\begin{aligned}\mathbf{0} &= T^{-1/2}\frac{\partial\ell(\hat{\boldsymbol{\theta}}_T^{\mathrm{cml}})}{\partial\boldsymbol{\theta}}\\ &= T^{-1/2}\frac{\partial\ell(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \left(\frac{1}{T}\frac{\partial^2\ell(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right)\sqrt{T}\big(\hat{\boldsymbol{\theta}}_T^{\mathrm{cml}} - \boldsymbol{\theta}_0\big),\end{aligned}$$

where $\boldsymbol{\theta}^\star$ lies in between $\hat{\boldsymbol{\theta}}_T^{\mathrm{cml}}$ and $\boldsymbol{\theta}_0$. According to Theorem 4.1.3 of Amemiya (1985), we need to show that

$$T^{-1/2}\frac{\partial\ell(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \xrightarrow{d} N\big(\mathbf{0}, \boldsymbol{I}(\boldsymbol{\theta}_0)\big), \tag{A.3}$$

$$\frac{1}{T}\frac{\partial^2\ell(\boldsymbol{\theta}^\star)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \to -\boldsymbol{I}(\boldsymbol{\theta}_0) \quad \text{in probability}, \tag{A.4}$$

where $\boldsymbol{I}(\boldsymbol{\theta}_0) = E[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^\top}]$. The theorem will straightforwardly follow. Again, we will split the proof into several intermediate results:

(i) It is easy to see $E\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = \mathbf{0}$, $\mathrm{Cov}(\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}) = E(\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}})(\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}})^\top$.

Using the ergodic theorem, $\frac{1}{T}\frac{\partial\ell(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \to E(\frac{1}{P_n(X_t|\mathcal{F}_{t-1})}\frac{\partial P_n(X_t|\mathcal{F}_{t-1})}{\partial\boldsymbol{\theta}})_{\boldsymbol{\theta}_0}$ in probability one.

Using the martingale central limit theorem and the Cramér–Wold device, it is direct to obtain (A.3), i.e., $\frac{1}{\sqrt{T}}\frac{\partial\ell(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{I}(\boldsymbol{\theta}_0))$.

(ii) According to Lemma 1 and Lemma 4 in Appendix B, we obtain that all the partial derivatives $\frac{\partial l_t(\boldsymbol{\theta})}{\partial r_i}$ exist and three times continuous differentiable in Θ, thus $\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial r_i\partial r_j}$ exists and is continuous in $N(\boldsymbol{\theta}_0)$, $\forall i, j = 0, 1, 2, \ldots, p$. Thus, there at least exists a point $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0)$ such that $\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial r_i\partial r_j}$ attains the maximum value at $\boldsymbol{\theta}$. Hence, $E\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial r_i\partial r_j} = E\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial r_i\partial r_j} < \infty$.

For convenience, we denote $\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \boldsymbol{G}(\boldsymbol{X}_t, \boldsymbol{\theta}) = (g_{ij}(\boldsymbol{X}_t, \boldsymbol{\theta}))$ and $E \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \boldsymbol{G}(\boldsymbol{\theta}) = (g_{ij}(\boldsymbol{X}_t, \boldsymbol{\theta}))$. We only need to prove $g_{ij}(\boldsymbol{X}_t, \boldsymbol{\theta})$ converges to a finite and non-stochastic function $g_{ij}(\boldsymbol{\theta}) = E(g_{ij}(\boldsymbol{X}_t, \boldsymbol{\theta}))$. Let $h(\boldsymbol{X}_t, \boldsymbol{\theta}) = g_{ij}(\boldsymbol{X}_t, \boldsymbol{\theta}) - E[g_{ij}(\boldsymbol{X}_t, \boldsymbol{\theta})]$, then $Eh(\boldsymbol{X}_t, \boldsymbol{\theta}) = 0$, i.e.,

$$\sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - E \frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right\| = o_p(1).$$

Note that $E(\frac{1}{P_n(X_t|\mathcal{F}_{t-1})} \frac{\partial^2 P_n(X_t|\mathcal{F}_{t-1})}{\partial r_i \partial r_j}) = 0$, $\forall i, j = 0, 1, 2, \ldots, p$, thus $E(\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}) = -E(\frac{l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}) = -\boldsymbol{I}(\boldsymbol{\theta}_0)$, i.e. (A.4) holds. The proof is completed. □

Proof of Theorem 5. Similar to Theorem 3, we obtain the consistency and asymptotic normality of $\tilde{\boldsymbol{\eta}}_T^{\text{cls}}$. To illustrate the consistency and asymptotic normality of $\hat{\boldsymbol{\eta}}_T^{\text{cls}}$, we define

$$Q_T(\boldsymbol{\eta}) = \frac{1}{T} \sum_{t=2}^{T} (X_t - g_{\boldsymbol{\eta}}(u_t))^2 := \frac{1}{T} \sum_{t=2}^{T} q_t(\boldsymbol{\eta}),$$

$$\widetilde{Q}_T(\boldsymbol{\eta}) = \frac{1}{T} \sum_{t=2}^{T} (X_t - g_{\boldsymbol{\eta}}(\widetilde{u}_t))^2 := \frac{1}{T} \sum_{t=2}^{T} \widetilde{q}_t(\boldsymbol{\eta}),$$

and $\overline{\widetilde{Q}}(\boldsymbol{\eta}) = E(X_t - g_{\boldsymbol{\eta}}(\widetilde{u}_t))^2$.

If the assumptions in Theorem 5 hold, then the following results are satisfied. As $T \to \infty$,

(i) $\|Q_T(\boldsymbol{\eta}) - \overline{\widetilde{Q}}(\boldsymbol{\eta})\|_\Omega \xrightarrow{a.s.} 0$.

(ii) $\|\boldsymbol{W}(\boldsymbol{\eta}) - \widetilde{\boldsymbol{W}}(\boldsymbol{\eta})\|_\Omega \xrightarrow{a.s.} 0$, where $\widetilde{\boldsymbol{W}} = E(\widetilde{U}_t(\boldsymbol{\eta})^2 \frac{\partial g_{\boldsymbol{\eta}}(\widetilde{u}_t)}{\partial \boldsymbol{\eta}} \frac{\partial g_{\boldsymbol{\eta}}(\widetilde{u}_t)}{\partial \boldsymbol{\eta}^\top})$ with $\widetilde{U}_t(\boldsymbol{\eta}) = X_t - g_{\boldsymbol{\eta}}(\widetilde{u}_t)$.

(iii) $\|\boldsymbol{V}(\boldsymbol{\eta}) - \widetilde{\boldsymbol{V}}(\boldsymbol{\eta})\|_\Omega \xrightarrow{a.s.} 0$, where $\widetilde{\boldsymbol{V}} = E(\frac{\partial g_{\boldsymbol{\eta}}(\widetilde{u}_t)}{\partial \boldsymbol{\eta}} \frac{\partial g_{\boldsymbol{\eta}}(\widetilde{u}_t)}{\partial \boldsymbol{\eta}^\top} - \widetilde{U}_t(\boldsymbol{\eta}) \frac{\partial^2 g_{\boldsymbol{\eta}}(\widetilde{u}_t)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top})$.

Note that

$$\|Q_T(\boldsymbol{\eta}) - \overline{\widetilde{Q}}(\boldsymbol{\eta})\|_\Omega \le \underbrace{\|Q_T(\boldsymbol{\eta}) - \widetilde{Q}_T(\boldsymbol{\eta})\|_\Omega}_{\text{(I)}} + \underbrace{\|\widetilde{Q}_T(\boldsymbol{\eta}) - \overline{\widetilde{Q}}(\boldsymbol{\eta})\|_\Omega}_{\text{(II)}}. \tag{A.5}$$

Therefore, the uniform convergence in (i) follows if both terms, that is, (I) and (II), on the right-hand side of the inequality in equation (A.5) converge almost surely to zero. Note that

$$\begin{aligned} \text{(I)} = \|Q_T(\boldsymbol{\eta}) - \widetilde{Q}_T(\boldsymbol{\eta})\|_\Omega &\le \frac{1}{T} \left\| \sum_{t=2}^{T} (X_t - g_{\boldsymbol{\eta}}(u_t))^2 - \sum_{t=2}^{T} (X_t - g_{\boldsymbol{\eta}}(\widetilde{u}_t))^2 \right\|_\Omega \\ &\le \frac{1}{T} \sum_{t=2}^{T} \|U_t(\boldsymbol{\eta})^2 - \widetilde{U}_t(\boldsymbol{\eta})^2\|_\Omega = \frac{1}{T} \sum_{t=2}^{T} \sup_{\boldsymbol{\eta} \in \Omega} \|U_t(\boldsymbol{\eta})^2 - \widetilde{U}_t(\boldsymbol{\eta})^2\|_\Omega. \end{aligned}$$

Using the mean value theorem, there exists an $\alpha_t(\boldsymbol{\eta}^*)$ lies between in $\alpha_t(\hat{\boldsymbol{\eta}}_T^{\text{cls}})$ and $\alpha_t(\hat{\boldsymbol{\eta}}_T^{\text{cls}})$ such that $|U_t(\boldsymbol{\eta})^2 - \widetilde{U}_t(\boldsymbol{\eta})^2| \le 2n|X_t - n\alpha_t(\boldsymbol{\eta}^*)||\alpha_t(\hat{\boldsymbol{\eta}}_T^{\text{cls}}) - \alpha_t(\tilde{\boldsymbol{\eta}}_T^{\text{cls}})| \le 2n(X_t + n)|\alpha_t(\hat{\boldsymbol{\eta}}_T^{\text{cls}}) - \alpha_t(\tilde{\boldsymbol{\eta}}_T^{\text{cls}})| \xrightarrow{a.s.} 0$ by Theorem 2. Hence, (I) $\xrightarrow{a.s.} 0$, $T \to \infty$.

The item (II) $\xrightarrow{a.s.} 0$, $T \to \infty$ follows by $E(\widetilde{U}_t(\boldsymbol{\eta})^2) = \operatorname{Var}(X_t|\widetilde{\alpha}_t, \mathcal{D}_{t-1}) < \infty$ and the ergodic theorem. Therefore, (i) holds.

According to Theorem 2, we have $\|s_t(\alpha_{t-1}) - \widetilde{s}_t(\widetilde{\alpha}_{t-1})\|_\Omega = n\|\alpha_t(\boldsymbol{\eta}) - \widetilde{\alpha}_t(\boldsymbol{\eta})\|_\Omega \xrightarrow{a.s.} 0$, $T \to \infty$, Using the Proposition 6.1.3 in Brockwell and Davis (1991), we have, $T \to \infty$,

$$\left\| \frac{\partial q_t(\boldsymbol{\eta})}{\partial w} - \frac{\partial \widetilde{q}_t(\boldsymbol{\eta})}{\partial w} \right\|_\Omega \le 2n^2 (\|\alpha_t - \widetilde{\alpha}_t\|_\Omega + 2\|\alpha_t^2 - \widetilde{\alpha}_t^2\|_\Omega + \|\alpha_t^3 - \widetilde{\alpha}_t^3\|_\Omega) \xrightarrow{a.s.} 0,$$

$$\left\| \frac{\partial q_t(\boldsymbol{\eta})}{\partial \beta} - \frac{\partial \widetilde{q}_t(\boldsymbol{\eta})}{\partial \beta} \right\|_{\Omega} \leq 2n^2 \big(\|u_{t-1}\alpha_t - \widetilde{u}_{t-1}\widetilde{\alpha}_t\|_{\Omega} + 2\|u_{t-1}\alpha_t^2 - \widetilde{u}_{t-1}\widetilde{\alpha}_t^2\|_{\Omega} + \|u_{t-1}\alpha_t^3 - \widetilde{u}_{t-1}\widetilde{\alpha}_t^3\|_{\Omega} \big) \xrightarrow{a.s.} 0,$$

$$\left\| \frac{\partial q_t(\boldsymbol{\eta})}{\partial \tau} - \frac{\partial \widetilde{q}_t(\boldsymbol{\eta})}{\partial \tau} \right\|_{\Omega} \leq 2n^2 \big(\|s_t\alpha_t - \widetilde{s}_t\widetilde{\alpha}_t\|_{\Omega} + 2\|s_t\alpha_t^2 - \widetilde{s}_t\widetilde{\alpha}_t^2\|_{\Omega} + \|s_t\alpha_t^3 - \widetilde{s}_t\widetilde{\alpha}_t^3\|_{\Omega} \big) \xrightarrow{a.s.} 0,$$

Hence, we obtain $\|\frac{1}{T}\sum_{t=2}^{T} \frac{\partial q_t(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} - E\frac{\partial \widetilde{q}_t(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\|_{\Omega} \xrightarrow{a.s.} 0$ and $q_t(\boldsymbol{\eta}) \to \widetilde{q}_t(\boldsymbol{\eta})$ in probability one. Thus the results of (ii) hold. Similarly, the item (iii) holds. Then the proof of Theorem 5 is end. □

Proof of Theorem 6. Similar to Theorem 4, we obtain the consistency and asymptotic normality of $\widetilde{\boldsymbol{\eta}}_T^{\mathrm{cml}}$. To illustrate the consistency and asymptotic normality of $\hat{\boldsymbol{\eta}}_T^{\mathrm{cml}}$, we define

$$\ell_T(\boldsymbol{\eta}) = \frac{1}{T}\sum_{t=2}^{T} l_t(\boldsymbol{\eta}) = \frac{1}{T}\sum_{t=2}^{T} \log P_n\big(X_t|\alpha_t(\boldsymbol{\eta}), \mathcal{D}_{t-1}\big),$$

$$\widetilde{\ell}_T(\boldsymbol{\eta}) = \frac{1}{T}\sum_{t=2}^{T} \widetilde{l}_t(\boldsymbol{\eta}) = \frac{1}{T}\sum_{t=2}^{T} \log P_n\big(X_t|\widetilde{\alpha}_t(\boldsymbol{\eta}), \mathcal{D}_{t-1}\big) \quad \text{and} \quad \overline{\widetilde{\ell}}(\boldsymbol{\eta}) = E\widetilde{l}_t(\boldsymbol{\eta}).$$

If the assumptions in Theorem 6 hold, we have

$$T^{-1/2}\frac{\partial \widetilde{\ell}(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}} \xrightarrow{d} N(\mathbf{0}, \widetilde{\boldsymbol{I}}(\boldsymbol{\eta}_0)) \quad \text{and} \quad \sqrt{T}(\widetilde{\boldsymbol{\eta}}_T^{\mathrm{cml}} - \boldsymbol{\eta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \widetilde{\boldsymbol{I}}^{-1}(\boldsymbol{\eta}_0))$$

with $\widetilde{\boldsymbol{I}}(\boldsymbol{\eta}_0) = E(\frac{\partial \widetilde{l}_t(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}} \frac{\partial \widetilde{l}_t(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}^{\top}})$.

To prove Theorem 6, we also need illustrate the following results:

(i) Similar to Theorem 3.1 in Gorgi (2018), we have $\|\ell_T(\boldsymbol{\eta}) - \overline{\widetilde{\ell}}(\boldsymbol{\eta})\|_{\Omega} \xrightarrow{a.s.} 0$, $T \to \infty$. Hence, the CML estimator defined in equation (3.4) is strongly consistent with respect to $\boldsymbol{\eta}_0$.

(ii) According to the third item of Corollary 1, we have $|\operatorname{logit}(\alpha_t(\hat{\boldsymbol{\eta}}_T)) - \operatorname{logit}(\widetilde{\alpha}_t(\boldsymbol{\eta}_0))| \xrightarrow{a.s.} 0$ and $|P_n(x|\alpha_t(\hat{\boldsymbol{\eta}}_T), \mathcal{D}_{t-1}) - P_n(x|\widetilde{\alpha}_t(\boldsymbol{\eta}_0), \mathcal{D}_{t-1})| \xrightarrow{p} 0$, if $\hat{\boldsymbol{\eta}}_T \xrightarrow{a.s.} \boldsymbol{\eta}_0$, $T \to \infty$. Hence, we obtain, as $T \to \infty$,

$$\begin{aligned} &\left\| \frac{\partial \ell_t(\hat{\boldsymbol{\eta}}_T^{\mathrm{cml}})}{\partial \boldsymbol{\eta}} - E\frac{\partial \widetilde{l}_t(\widetilde{\boldsymbol{\eta}}_T^{\mathrm{cml}})}{\partial \boldsymbol{\eta}} \right\|_{\Omega} \\ &\leq \left\| \frac{\partial \ell_t(\hat{\boldsymbol{\eta}}_T^{\mathrm{cml}})}{\partial \boldsymbol{\eta}} - \frac{\partial \widetilde{\ell}_t(\hat{\boldsymbol{\eta}}_T^{\mathrm{cml}})}{\partial \boldsymbol{\eta}} \right\|_{\Omega} + \left\| \frac{\partial \widetilde{\ell}_t(\hat{\boldsymbol{\eta}}_T^{\mathrm{cml}})}{\partial \boldsymbol{\eta}} - \frac{\partial \widetilde{\ell}_t(\widetilde{\boldsymbol{\eta}}_T^{\mathrm{cml}})}{\partial \boldsymbol{\eta}} \right\|_{\Omega} \\ &\quad + \left\| \frac{\partial \widetilde{\ell}_t(\widetilde{\boldsymbol{\eta}}_T^{\mathrm{cml}})}{\partial \boldsymbol{\eta}} - E\frac{\partial \widetilde{l}_t(\widetilde{\boldsymbol{\eta}}_T^{\mathrm{cml}})}{\partial \boldsymbol{\eta}} \right\|_{\Omega} \\ &\xrightarrow{a.s.} 0. \end{aligned}$$

Similarly, $\|\frac{1}{T}\sum_{t=1}^{T} \frac{\partial l_t(\hat{\boldsymbol{\eta}}_T^{\mathrm{cml}})}{\partial \boldsymbol{\eta}} \frac{\partial l_t(\hat{\boldsymbol{\eta}}_T^{\mathrm{cml}})}{\partial \boldsymbol{\eta}^{\top}} - E\frac{\partial \widetilde{l}_t(\widetilde{\boldsymbol{\eta}}_T^{\mathrm{cml}})}{\partial \boldsymbol{\eta}} \frac{\partial \widetilde{l}_t(\widetilde{\boldsymbol{\eta}}_T^{\mathrm{cml}})}{\partial \boldsymbol{\eta}^{\top}}\|_{\Omega} \xrightarrow{a.s.} 0$, $T \to \infty$. Therefore, $\|\boldsymbol{I}(\boldsymbol{\eta}_0) - \widetilde{\boldsymbol{I}}(\boldsymbol{\eta}_0)\| \xrightarrow{a.s.} 0$ and $T^{-1/2}\frac{\partial \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}} \xrightarrow{d} N(\mathbf{0}, \widetilde{\boldsymbol{I}}(\boldsymbol{\eta}_0))$, where $\boldsymbol{I}(\boldsymbol{\eta}_0) = E[\frac{\partial l_t(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \frac{\partial l_t(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^{\top}}]_{\boldsymbol{\eta}_0}$.

(iii) For convenience, we denote $\boldsymbol{\eta} = (w, \beta, \tau)^\top = (\eta_1, \eta_2, \eta_3)^\top$.

Similar to (ii), $\forall i, j, k = 1, 2, 3$, we obtain $\|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top} - E\frac{\partial^2 \widetilde{l}_t(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top}\|_\Omega \xrightarrow{a.s.} 0$, $\|E(\frac{1}{P_n(x|\alpha_t,\mathcal{D}_{t-1})}\frac{\partial^2 P_n(x|\alpha_t,\mathcal{D}_{t-1})}{\partial\eta_i\partial\eta_j}) - E(\frac{1}{P_n(x|\widetilde{\alpha}_t,\mathcal{D}_{t-1})}\frac{\partial^2 P_n(x|\widetilde{\alpha}_t,\mathcal{D}_{t-1})}{\partial\eta_i\partial\eta_j})\|_\Omega \xrightarrow{a.s.} 0$ and $\|E|\frac{\partial^3 l_t(\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j\partial\eta_k}| - E|\frac{\partial^3 \widetilde{l}_t(\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j\partial\eta_k}|\|_\Omega \xrightarrow{a.s.} 0$, as $T \to \infty$. Thus, the asymptotic normality of $\hat{\boldsymbol{\eta}}_T^{\mathrm{cml}}$ holds. □

Appendix B: Auxiliary results

Lemma 1. *Let X follow a binomial distribution with parameter α, i.e., $P_n(X = x) = \binom{n}{x}\alpha^x(1-\alpha)^{n-x}$. Then we have $\frac{dP_n(X=x)}{d\alpha} = \frac{n}{1-\alpha}[P_{n-1}(x-1) - P_n(x)]$.*

The proof of Lemma 1 can be seen from that of the Proposition 1 in Freeland and McCabe (2004).

Lemma 2. *If $\{X_t\}$ is the strictly stationary and ergodic solution of model (2.1) and Assumption 3 holds, then model (2.1) is identifiable.*

Proof. According to (2.1), we have $\alpha_t(\boldsymbol{\theta}) = \alpha_t(\boldsymbol{\theta}_0)$, that is, $\mathrm{logit}(\alpha_t(\boldsymbol{\theta})) = \mathrm{logit}(\alpha_t(\boldsymbol{\theta}_0))$ by Lemma 4, if $P_n(X_t|\mathcal{F}_{t-1})_{\boldsymbol{\theta}} = P_n(X_t|\mathcal{F}_{t-1})_{\boldsymbol{\theta}_0}$. Thus, we have $E(\mathrm{logit}\,\alpha_t(\boldsymbol{\theta})) = E(\mathrm{logit}\,\alpha_t(\boldsymbol{\theta}_0))$, i.e., $r_0 + \mu\sum_{j=1}^{p} r_k = r_0^0 + \mu\sum_{j=1}^{p} r_k^0$. Hence, $r_0 = r_0^0$ and $r_j = r_j^0$, $j = 1, 2, \ldots, p$. □

Lemma 3. *If $\{X_t\}$ is the strictly stationary and ergodic process and Assumption 5 hold, then model (2.5) is identifiable.*

Proof. According to (2.5), we have $\widetilde{\alpha}_t(\boldsymbol{\eta}) = \widetilde{\alpha}_t(\boldsymbol{\eta}_0)$, if $P_n(X_t|\widetilde{\alpha}_t, \mathcal{D}_{t-1})_{\boldsymbol{\eta}} = P_n(X_t|\widetilde{\alpha}_t, \mathcal{D}_{t-1})_{\boldsymbol{\eta}_0}$. Because $\mathrm{logit}(x)$ is a strictly monotone increasing function, then $\mathrm{logit}(\widetilde{\alpha}_t(\boldsymbol{\eta})) = \mathrm{logit}(\widetilde{\alpha}_t(\boldsymbol{\eta}_0))$, if $P_n(X_t|\widetilde{\alpha}_t, \mathcal{D}_{t-1})_{\boldsymbol{\eta}} = P_n(X_t|\widetilde{\alpha}_t, \mathcal{D}_{t-1})_{\boldsymbol{\eta}_0}$. Thus, $E(\mathrm{logit}(\widetilde{\alpha}_t(\boldsymbol{\eta}))) = E(\mathrm{logit}(\widetilde{\alpha}_t(\boldsymbol{\eta}_0)))$, i.e., $\frac{w}{1-\beta} = \frac{w_0}{1-\beta_0}$, then we obtain $w = w_0$ and $\beta = \beta_0$. Hence, $\tau = \tau_0$ by $\mathrm{logit}(\widetilde{\alpha}_t(\boldsymbol{\eta})) = \mathrm{logit}(\widetilde{\alpha}_t(\boldsymbol{\eta}_0))$. □

Lemma 4. *Let $f(x) = \frac{\exp(x)}{1+\exp(x)}$, $x \in (-\infty, +\infty)$, then*

(1) $f'(x) = f(x)(1 - f(x))$, $f''(x) = f'(x)(1 - 2f(x))$ *and* $f'''(x) = f'(x)[1 + 6f(x)(1 - f(x))]$.
(2) $|f(x_1) - f(x_2)| \le \frac{1}{4}|x_1 - x_2|$, $\forall x_1 < x_2$.

Proof. The proof of (i) is easy to deduce and we omit it. By the assumption, we have

$$f'(x) = \frac{e^x}{(1+e^x)^2} = \frac{e^x}{1 + 2e^x + e^{2x}} = \frac{1}{e^{-x} + e^x + 2} \le 1/4, \quad \forall x \in \mathbb{R},$$

thus, $0 < f'(x) \le 1/4$. Let $F_1(x) = x/4 + f(x)$ and $F_2(x) = x/4 - f(x)$. Then $F_1'(x) = 1/4 + f'(x) > \frac{1}{4}$ and $F_2'(x) = 1/4 - f'(x) \ge 1/4 + 1/4 = 1/2$, hence, $F_1(x)$ and $F_2(x)$ are strictly monotone increasing functions. If $x_1 > x_2$, we have $F_1(x_2) > F_1(x_1)$ and $F_2(x_2) >

$F_2(x_1)$, i.e.,

$$\begin{cases} \dfrac{1}{4}x_2 + f(x_2) > \dfrac{1}{4}x_1 + f(x_1), \\ \dfrac{1}{4}x_2 - f(x_2) > \dfrac{1}{4}x_1 - f(x_1), \end{cases} \quad \Rightarrow \quad \begin{cases} \dfrac{1}{4}(x_2 - x_1) > f(x_1) - f(x_2), \\ \dfrac{1}{4}(x_2 - x_1) > -\big(f(x_1) - f(x_2)\big). \end{cases}$$

Hence, $|f(x_1) - f(x_2)| < \frac{1}{4}|x_1 - x_2|$. □

Lemma 5. *Let* $\{X_t, t \in \mathbb{Z}\}$ *is a stationary and ergodic random variable sequence and* $s_t(\alpha_{t-1}) = X_{t-1} - n\alpha_{t-1}$. *Then the following inequalities are satisfied with probability* 1:

(1) $|s_t(\alpha_{t-1})| \le \max(X_{t-1}, n - X_{t-1})$.
(2) $-\frac{n}{4} \le \frac{\partial s_t(\alpha_{t-1})}{\partial \mathrm{logit}(\alpha_{t-1})} \le 0$.

Proof. According to (2.4), we have that $s_t(\alpha_{t-1}) = X_{t-1} - n\alpha_{t-1} = X_{t-1} - n\frac{\exp(\mathrm{logit}(\alpha_{t-1}))}{1+\exp(\mathrm{logit}(\alpha_{t-1}))}$ and $\frac{\partial s_t(\alpha_{t-1})}{\partial \mathrm{logit}(\alpha_{t-1})} = -n\alpha_{t-1}(1 - \alpha_{t-1})$. Hence, $|s_t(\alpha_{t-1})| \le \max(X_{t-1}, n - X_{t-1})$ and $-\frac{n}{4} \le \frac{\partial s_t(\alpha_{t-1})}{\partial \mathrm{logit}(\alpha_{t-1})} \le 0$. □

Acknowledgments

We are very grateful to anonymous referees for providing several exceptionally helpful comments which led to a significant improvement of the paper. We acknowledge the constructive comments from Professor Christian H. Weiß of Helmut Schmidt University on the earlier versions. Zhu's work is supported by National Natural Science Foundation of China (Nos. 11871027, 11731015), Science and Technology Developing Plan of Jilin Province (No. 20170101057JC), and Cultivation Plan for Excellent Young Scholar Candidates of Jilin University. Li's work is supported by Natural Science Foundation of Changchun Normal University (No. 2018-004). Zhu is the corresponding author.

References

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge: Harvard University Press.

Bazzi, M., Blasques, F., Koopman, S. J. and Lucas, A. (2017). Time-varying transition probabilities for Markov regime switching models. *Journal of Time Series Analysis* **38**, 458–478. MR3637726

Blasques, F., Gorgi, P., Koopman, S. J. and Wintenberger, O. (2018a). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics* **12**, 1019–1052. MR3776279

Blasques, F., Koopman, S. J. and Lucas, A. (2014). Stationarity and ergodicity of univariate generalized autoregressive score processes. *Electronic Journal of Statistics* **8**, 1088–1112. MR3263114

Blasques, F., Koopman, S. J. and Lucas, A. (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika* **102**, 325–343. MR3371007

Blasques, F., Lucas, A. and Silde, E. (2018b). A stochastic recurrence equations approach for score driven correlation models. *Econometric Reviews* **37**, 166–181. MR3750653

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. New York: Springer. MR2839251

Creal, D., Koopman, S. J. and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics* **28**, 777–795. MR3083971

Cui, Y. and Lund, R. (2010). Inference in binomial AR(1) models. *Statistics & Probability Letters* **80**, 1985–1990. MR2734270

Current, W. L. and Garcia, L. S. (1991). Cryptosporidiosis. *Clinical Microbiology Reviews* **4**, 325–358.

Freeland, R. K. and McCabe, B. P. M. (2004). Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis* **25**, 701–722. MR2089191

Gorgi, P. (2018). Integer-valued autoregressive models with survival probability driven by a stochastic recurrence equation. *Journal of Time Series Analysis* **39**, 150–171. MR3767058

Harvey, A. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. New York: Cambridge University Press. MR3135260

Höhle, M. and Paul, M. (2008). Count data regression charts for the monitoring of surveillance time series. *Computational Statistics & Data Analysis* **52**, 4357–4368. MR2432467

Kim, H. Y. and Park, Y. (2010a). Coherent forecasting in binomial AR(p) model. *Communications of the Korean Statistical Society* **17**, 27–37. https://doi.org/10.5351/CKSS.2010.17.1.027.

Kim, H. Y. and Park, Y. (2010b). Markov chain approach to forecast in the binomial autoregressive models. *Communications of the Korean Statistical Society* **17**, 441–450. https://doi.org/10.5351/CKSS.2010.17.3.441.

Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics* **6**, 629–642. MR0494770

Lee, Y. and Lee, S. (2019). CUSUM test for general nonlinear integer-valued GARCH models: Comparison study. *Annals of the Institute of Statistical Mathematics* **71**, 1033–1057. MR3993524 https://doi.org/10.1007/s10463-018-0676-7

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resour Bulletin* **21**, 645–650. https://doi.org/10.1111/j.1752-1688.1985.tb05379.x.

Möller, T. A., Silva, M. E., Weiß, C. H., Scotto, M. G. and Pereira, I. (2016). Self-exciting threshold binomial autoregressive processes. *AStA Advances in Statistical Analysis* **100**, 369–400. MR3551305

Ristić, M. M. and Popović, B. V. (2019). A new bivariate binomial time series model. *Markov Processes and Related Fields* **25**, 301–328.

Ristić, M. M., Weiß, C. H. and Janjić, A. D. (2016). A binomial integer-valued ARCH model. *The International Journal of Biostatistics* **12**, 20150051. MR3581286

Scotto, M. G., Weiß, C. H., Silva, M. E. and Pereira, I. (2014). Bivariate binomial autoregressive models. *Journal of Multivariate Analysis* **125**, 233–251. MR3163841

Steutel, F. W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability* **7**, 893–899. MR0542141

Tsay, R. S. (1992). Model checking via parametric bootstraps in time series analysis. *Journal of the Royal Statistical Society Series C Applied Statistics* **41**, 1–15. https://doi.org/10.2307/2347612.

Weiß, C. H. (2009a). Monitoring correlated processes with binomial marginals. *Journal of Applied Statistics* **36**, 399–414. MR2668770

Weiß, C. H. (2009b). Jumps in binomial AR(1) processes. *Statistics & Probability Letters* **79**, 2012–2019. MR2750180

Weiß, C. H. (2009c). A new class of autoregressive models for time series of binomial counts. *Communications in Statistics Theory and Methods* **38**, 447–460. MR2510797

Weiß, C. H. (2015). A Poisson INAR(1) model with serially dependent innovations. *Metrika* **78**, 829–851. MR3383933

Weiß, C. H. (2018). *An Introduction to Discrete-Valued Time Series*. Chichester: John Wiley & Sons.

Weiß, C. H. and Feld, M. H. J. M. (2019). On the performance of information criteria for model identification of count time series. *Studies in Nonlinear Dynamics and Econometrics*. To appear. https://doi.org/10.1515/snde-2018-0012.

Weiß, C. H. and Kim, H. Y. (2014). Diagnosing and modeling extra-binomial variation for time-dependent counts. *Applied Stochastic Models in Business and Industry* **30**, 588–608. MR3271077

Weiß, C. H. and Pollett, P. K. (2012). Chain binomial models and binomial autoregressive processes. *Biometrics* **68**, 815–824. MR3055186

Weiß, C. H. and Pollett, P. K. (2014). Binomial autoregressive processes with density-dependent thinning. *Journal of Time Series Analysis* **35**, 115–132. MR3166349

Wintenberger, O. (2013). Continuous invertibility and stable QML estimation of the EGARCH(1, 1) model. *Scandinavian Journal of Statistics* **40**, 846–867. MR3145121

Yang, K., Wang, D. and Li, H. (2018). Threshold autoregression analysis for finite-range time series of counts with an application on measles data. *Journal of Statistical Computation and Simulation* **88**, 597–614. MR3764838

Zhang, H., Wang, D. and Zhu, F. (2011a). Empirical likelihood for first-order random coefficient integer-valued autoregressive processes. *Communications in Statistics Theory and Methods* **40**, 492–509. MR2765843

Zhang, H., Wang, D. and Zhu, F. (2011b). Empirical likelihood inference for random coefficient INAR(p) process. *Journal of Time Series Analysis* **32**, 195–203. MR2808249

Zhang, H., Wang, D. and Zhu, F. (2012). Generalized RCINAR(1) process with signed thinning operator. *Communications in Statistics Theory and Methods* **41**, 1750–1770. MR2913009

Zheng, H. and Basawa, I. V. (2008). First-order observation-driven integer-valued autoregressive processes. *Statistics & Probability Letters* **78**, 1–9. MR2381267

H. Chen
F. Zhu
School of Mathematics
Jilin University
Changchun 130012
China
E-mail: chenhp0107@163.com
zfk8010@163.com

Q. Li
College of Mathematics
Changchun Normal University
Changchun 130032
China
E-mail: 46968158@qq.com