



Models for autoregressive processes of bounded counts: How different are they?

Hee-Young Kim¹ · Christian H. Weiß² · Tobias A. Möller²

Received: 25 September 2019 / Accepted: 11 March 2020 / Published online: 27 March 2020
© The Author(s) 2020

Abstract

We focus on purely autoregressive (AR)-type models defined on the bounded range $\{0, 1, \dots, n\}$ with a fixed upper limit $n \in \mathbb{N}$. These include the binomial AR model, binomial AR conditional heteroscedasticity (ARCH) model, binomial-variation AR model with their linear conditional mean, nonlinear max-binomial AR model, and binomial logit-ARCH model. We consider the key problem of identifying which of these AR-type models is the true data-generating process. Despite the volume of the literature on model selection, little is known about this procedure in the context of nonnested and nonlinear time series models for counts. We consider the most popular approaches used for model identification, Akaike's information criterion and the Bayesian information criterion, and compare them using extensive Monte Carlo simulations. Furthermore, we investigate the properties of the fitted models (both the correct and wrong models) obtained using maximum likelihood estimation. A real-data example demonstrates our findings.

Keywords Binomial autoregressive models · Count time series · Model adequacy · Model selection · Parameter estimation

1 Introduction

Count time series occur in various fields, including investigations of natural phenomena (e. g., rare disease occurrences, animal sightings, and severe weather events) and in economic contexts (e. g., monitoring the number of transactions). Significant progress

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00180-020-00980-6>) contains supplementary material, which is available to authorized users.

✉ Christian H. Weiß
weissc@hsu-hh.de

¹ Division of Economics and Statistics, National Statistics, Korea University, Sejong, South Korea

² Department of Mathematics and Statistics, Helmut Schmidt University, PO box 700822, 22008 Hamburg, Germany

has been made in modeling count time series over the last 20 years; see Weiß (2018) for a recent survey. The most popular stationary count time series models are arguably the integer-valued autoregressive moving-average (INARMA) models, dating back to McKenzie (1985) and Al-Osh and Alzaid (1987), which use a probabilistic operator called *binomial thinning*. The simplest INARMA model is the integer-valued autoregressive model of order one, abbreviated as INAR(1). The model is defined as $X_t = \alpha \circ X_{t-1} + \varepsilon_t$, with $t \in \mathbb{N} = \{1, 2, \dots\}$, where $\{\varepsilon_t\}$ is an innovation process composed of independent and identically distributed (i. i. d.) non-negative integer-valued random variables (i. e., having the range $\mathbb{N}_0 = \{0, 1, \dots\}$) with finite mean and variance. The symbol “ \circ ” denotes the binomial thinning operator (Steutel and van Harn 1979), defined by $\alpha \circ X = \sum_{i=1}^X Y_i$, where $\{Y_j\}$ are i. i. d. Bernoulli random variables with $P(Y_j = 1) = \alpha = 1 - P(Y_j = 0)$. By construction, the observations X_t generated by the INAR(1) process can attain any value from \mathbb{N}_0 ; that is, the INAR(1) model applies to time series composed of unbounded counts.

INARMA models with unbounded support appear to provide an adequate framework for modeling dependent count data in many applications. However, these models cannot be applied if the counts exhibit a finite upper bound $n \in \mathbb{N}$ that needs to be preserved by the data-generating mechanism. Examples of bounded-counts time series are the weekly number of rainy days, which has a fixed upper limit of $n = 7$ (Cui and Lund 2009; Gouveia et al. 2018), the number of transactions among n companies (Weiß and Kim 2013), and the number of metapopulations on n patches (Weiß and Pollett 2012); a data example is presented in Sect. 7, in which we examine access counts to a web server.

Thus, the question is how to construct stationary and integer-valued time series models defined on the bounded range $\{0, 1, \dots, n\}$. Modeling time series of bounded counts is challenging. In the present work, we focus on purely autoregressive (AR)-type models, because these can be traced back to finite Markov chains and are particularly attractive in practice (e. g., in estimation, forecasting, asymptotic theory); see Weiß (2018) for more details. These AR-type models include, on the one hand, three members from the class of conditional linear AR (CLAR) models, because their conditional mean $E[X_t | X_{t-1}, \dots]$ is a linear expression in X_{t-1}, \dots, X_{t-p} (see Grunwald et al. 2000). These CLAR models are the p th-order binomial AR model (BAR(p)) based on the binomial thinning operator and a random mixture mechanism (Weiß 2009); the p th-order binomial AR conditional heteroscedasticity model (BARCH(p)) being a type of conditional regression model (Ristić et al. 2016); and the bvAR(p) model using the binomial variation (bv) operator (Gouveia et al. 2018). On the other hand, we also consider two types of models with a nonlinear conditional mean, the max-BAR(p) model as a counterpart to the conventional max-AR(p) model (Weiß et al. 2018); and the binomial logit-ARCH(p) model (logit-BARCH(p)) as another conditional regression model (Chen et al. 2019). A detailed description of these AR-type models is provided in Sects. 2 and 6, respectively.

All these AR-type models are Markovian of order p . Furthermore, the three types of CLAR(p) models also satisfy the *Yule–Walker equations*

$$\rho(k) = \sum_{i=1}^p \alpha_i \rho(|k-i|) \quad \text{for } k = 1, 2, \dots, \quad (1.1)$$

where $\rho(k) = \text{Corr} X_t, X_{t-k}$ is the autocorrelation function (ACF). Equivalently, if we consider the partial ACF (PACF) $\rho_{\text{part}}(k)$, which is defined for stationary processes as a function of the ACF, then $\rho_{\text{part}}(k) = 0$ for all $k > p$. Nevertheless, the models do vary in terms of their dispersion behavior, for example, when comparing the actual variance to the mean. For bounded counts, this comparison is commonly based on the binomial index of dispersion (BID), defined as

$$\text{BID} = \frac{n \text{Var}(X_t)}{E(X_t)(n - E(X_t))}. \quad (1.2)$$

Here a binomial distribution always leads to $\text{BID} = 1$, and a distribution with $\text{BID} > 1$ is said to exhibit extra-binomial variation (overdispersion). Conversely, a distribution with $\text{BID} < 1$ expresses underdispersion with respect to a binomial distribution.

Thus, a natural question arises as to how different these models are. From a practical perspective, we would like to know how to identify the true model underlying the data-generating process (DGP) given an AR-type time series of bounded counts. The goal of this study is to compare and investigate these AR-type models for bounded-counts time series. In particular, we focus on identifying the model behind the DGP. This is a crucial part of any time series analysis, because fitting a misspecified model to the data will lead to biased estimates and, subsequently, misleading inferences (e. g., forecasting) based on the previous results. The most popular approach for model identification is to apply an information criterion (Burnham and Anderson 2002), such as Akaike's information criterion (AIC) (Akaike 1974) or the Bayesian information criterion (BIC) (Schwarz 1978). These criteria penalize the likelihoods in order to determine a model providing both a good fit and a parsimonious parametrization. Selecting this model gives the minimal value for the considered criterion.

The remainder of this paper is organized as follows. We start our analyses by considering the CLAR-type models: the BAR(p), bvAR(p), and BARCH(p) models. In Sect. 2, we provide their definition and essential properties. Section 3 describes the design of our simulation study and analyzes the performance of the maximum likelihood (ML) estimation. In Sect. 4, we first summarize the AIC and BIC and then evaluate their performance in terms of model selection based on our Monte Carlo simulation results. Section 5 investigates the consequences of fitting the wrong model to the DGP. Section 6 is composed of further analyses of specific research questions. We take a more detailed look at the task of order selection among BARCH(p) models, and we compare the conditionally linear BARCH(p) model with the two nonlinear AR-type models, the max-BAR(p) and the logit-BARCH(p) model. We illustrate our findings with a real-data example in Sects. 7 and 8 concludes the paper.

2 CLAR-type models for time series of bounded counts

In what follows, we examine three CLAR-type models for bounded-counts time series. For model order $p \in \mathbb{N}$, all of them are Markovian of order p ; that is, the conditional distributions $P(X_t = x \mid X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \dots) = P(X_t = x \mid X_{t-1} = x_{t-1}, \dots, X_{t-p} = x_{t-p})$ depend only on the last p observations. These conditional distributions are invariant in time t ; thus, we abbreviate the transition probabilities as $p(x \mid x_1, \dots, x_p) = P(X_t = x \mid X_{t-1} = x_1, \dots, X_{t-p} = x_p)$. The three CLAR-type models satisfy the Yule–Walker equations (1.1), which constitutes the main difference to the nonlinear max-BAR and logit-BARCH models discussed in Sect. 6.

2.1 Binomial AR(p) model

The first CLAR-type model for bounded counts $(X_t)_{\mathbb{N}}$ with range $\{0, 1, \dots, n\}$ dates back to McKenzie (1985). The model is defined by an AR(1)-like recursion, where the multiplication “ \cdot ” is substituted by the binomial thinning operator “ \circ ” (recall Sect. 1, $\alpha \circ X \mid X \sim \text{Bin}(X, \alpha)$). Using this operator, the *first-order binomial AR model* (BAR(1)) (McKenzie 1985) is defined by

$$X_t = \alpha \circ X_{t-1} + \beta \circ (n - X_{t-1}) \quad \text{for } t \geq 1, \quad X_0 \sim \text{Bin}(n, \pi), \quad (2.1)$$

where $\pi \in (0, 1)$, $\rho \in (\max\{-\frac{\pi}{1-\pi}, -\frac{1-\pi}{\pi}\}; 1)$, and the thinning probabilities satisfy $\beta = \pi(1 - \rho)$ and $\alpha = \beta + \rho$.

Weiß (2009) extended the BAR(1) model given in (2.1) to a *p*th-order binomial AR model (BAR(p)) using an additional probabilistic mixture mechanism. Assuming $(D_{t,1}, \dots, D_{t,p})$ to be independent and multinomially distributed according to $\text{Mult}(1; \phi_1, \dots, \phi_p)$, with $\phi_1 + \dots + \phi_p = 1$, the BAR(p) model recursion is given by

$$X_t = \sum_{i=1}^p D_{t,i} \left(\alpha \circ X_{t-i} + \beta \circ (n - X_{t-i}) \right), \quad (2.2)$$

where all thinnings are performed independently. The ACF of the stationary BAR(p) process (2.2) satisfies the Yule–Walker equations (1.1), with $\alpha_i := \rho \phi_i$. The marginal distribution is the binomial distribution $\text{Bin}(n, \pi)$. The transition probabilities are as follows:

$$\begin{aligned} & p(x \mid x_{t-1}, \dots, x_{t-p}) \\ &= \sum_{i=1}^p \phi_i \sum_{y=0}^x \binom{x_{t-i}}{y} \alpha^y (1 - \alpha)^{x_{t-i}-y} \binom{n - x_{t-i}}{x - y} \beta^{x-y} (1 - \beta)^{n - x_{t-i} - x + y}. \end{aligned} \quad (2.3)$$

2.2 Binomial-variation AR(p) model

A second CLAR-type model for bounded counts was proposed by Gouveia et al. (2018). Its definition uses the binomial variation operator $bv_n(X)$, given by $bv_n(X)|X \sim \text{Bin}(n, X/n)$, and it is thus referred to as the *bvAR(p) model*. Similar to (2.2), it is based on a random mixture generated by $(D_{t,0}, \dots, D_{t,p}) \sim \text{Mult}(1; \phi_0, \dots, \phi_p)$, with $\phi_0 + \dots + \phi_p = 1$. The model recursion is given by

$$X_t = \sum_{i=1}^p D_{t,i} bv_n(X_{t-i}) + D_{t,0} \epsilon_t, \tag{2.4}$$

where all bv operators are performed independently, and the innovations ϵ_t are i.i.d. bounded counts. The ACF of the stationary $bvAR(p)$ process (2.4) satisfies the Yule–Walker equations (1.1), with $\alpha_i := \phi_i$. If the innovations ϵ_t are binomially distributed following $\text{Bin}(n, \pi)$, then the observations X_t are not binomial, but instead exhibit extra-binomial variation (Gouveia et al. 2018). The transition probabilities are as follows:

$$p(x|x_{t-1}, \dots, x_{t-p}) = \phi_0 \binom{n}{x} \pi^x (1 - \pi)^{n-x} + \sum_{i=1}^p \phi_i \binom{n}{x} \left(\frac{x_{t-i}}{n}\right)^x \left(1 - \frac{x_{t-i}}{n}\right)^{n-x}. \tag{2.5}$$

2.3 Binomial ARCH(p) model

Another CLAR-type model for bounded counts is given by the *p*th-order binomial ARCH model (BARCH(p)) (Ristić et al. 2016). This model is a type of conditional regression model, which requires that $X_t|X_{t-1}, \dots$ be conditionally binomially distributed according to $\text{Bin}(n, \pi_t)$, where

$$\pi_t = \alpha_0 + \frac{1}{n} \sum_{i=1}^p \alpha_i X_{t-i}. \tag{2.6}$$

Here $\alpha_0 > 0$, $\alpha_1, \dots, \alpha_p \geq 0$, and $\alpha_0 + \sum_{i=1}^p \alpha_i < 1$. As before, the ACF of the stationary BARCH(p) process (2.6) satisfies the Yule–Walker equations (1.1). The stationary marginal distribution is not binomial, but instead exhibits extra-binomial variation (Ristić et al. 2016). The transition probabilities are given by

$$p(x|x_{t-1}, \dots, x_{t-p}) = \binom{n}{x} \pi_t^x (1 - \pi_t)^{n-x}. \tag{2.7}$$

3 ML estimation for AR-type models

The CLAR-type models for bounded-counts time series summarized in Sect. 2 are Markovian of order p , and we provided closed-form formulae for calculating their

transition probabilities $p(x|x_1, \dots, x_p)$. Using these formulae, a conditional ML estimation is easily implemented by numerically maximizing the respective log-likelihood function,

$$\ell(\theta | x_p, \dots, x_1) = \sum_{t=p+1}^T \ln p(x_t | x_{t-1}, \dots, x_{t-p}),$$

where θ represents the vector of model parameters. This implementation was used in the simulation study described in Sect. 3.1, and for the data example in Sect. 7. The performance of the ML estimation (if the correct type of model has been chosen for the given time series data) is analyzed briefly in Sect. 3.2.

Finally, we note that ML estimation is possible in exactly the same way for the nonlinear AR-type models discussed in Sect. 6. We also provide closed-form formulae for the transition probabilities $p(x|x_1, \dots, x_p)$ and investigate the estimation performance.

3.1 Design of simulation study

For this research, we conduct an extensive simulation study (using the SAS/IML 9.4 procedure) where various DGP scenarios with $n = 10$ are used. The first and main part of our study focusses on the CLAR-type models introduced in Sect. 2; the corresponding DGP scenarios are summarized in Tables 1 and 2. The data are generated from the BAR(1) model for scenarios 1–4, from the bvAR(1) model for scenarios 5–8, from the BARCH(1) model for scenarios 9–12, and an analogous allocation holds for the second-order models (scenarios 13–24) in Table 2. Tables 1 and 2 summarize the parameters of the respective DGP in the column “Parameters.” We set the parameters for the BAR(1), bvAR(1), and BARCH(1) DGPs to ensure that the mean $E(X_t) = 1.5$, 4 and the ACF satisfies $\rho(1) = 0.3, 0.6$ (see the highlighted numbers in Table 1). The model properties shown in Table 1 are $E(X_t)$, $Var(X_t)$, $P(X_t = 0)$, the ACF, and the PACF.

To explore the effect of higher model orders, we simulate the AR(2)-type models; the results are shown in Table 2. The parameters of the BAR(2), bvAR(2), and BARCH(2) DGPs are chosen to ensure that the mean and $\rho(1)$ of each DGP are equal to those of BAR(1), bvAR(1), and BARCH(1), respectively. Additionally, the second-order autoregression is chosen to yield $\rho_{\text{part}}(2) = 0.2$ (see the highlighted numbers in Table 2). For all 24 scenarios, we use the sample sizes $T = 100, 250, 500, 1000$. The number of replications is 1000. The data are generated with a pre-run of length 250 to ensure stationarity.

In Sect. 6, we consider further DGP scenarios regarding third-order models as well as nonlinear AR-type models. For these DGPs, we work with 1000 replications and a pre-run of length 250. Details on their parametrization are presented in Sect. 6.

Table 1 The AR(1)-type models with $n = 10$ used in the simulation study and their properties, where (P)ACF $_k$ refers to the (P)ACF at lag k

Scenario	DGP	Parameters	Mean	Var	$P(X = 0)$	ACF1	ACF2	ACF3	ACF4	PACF1	PACF2	PACF3	PACF4
	BAR(1)	α	β										
1		0.405	0.105	1.500	1.275	0.197	0.300	0.090	0.027	0.008	0.300	0.000	0.000
2		0.660	0.060	1.500	1.275	0.197	0.600	0.360	0.216	0.130	0.600	0.000	0.000
3		0.580	0.280	4.000	2.400	0.006	0.300	0.090	0.027	0.008	0.300	0.000	0.000
4		0.760	0.160	4.000	2.400	0.006	0.600	0.360	0.216	0.130	0.600	0.000	0.000
	bvAR(1)	π	ϕ_1										
5		0.15	0.3	1.500	1.747	0.255	0.300	0.090	0.027	0.008	0.300	0.000	0.000
6		0.15	0.6	1.500	2.772	0.355	0.600	0.360	0.216	0.130	0.600	0.000	0.000
7		0.40	0.3	4.000	3.288	0.023	0.300	0.090	0.027	0.008	0.300	0.000	0.000
8		0.40	0.6	4.000	5.217	0.070	0.600	0.360	0.216	0.130	0.600	0.000	0.000
	BARCH(1)	α_0	α_1										
9		0.105	0.3	1.500	1.387	0.211	0.300	0.090	0.027	0.008	0.300	0.000	0.000
10		0.060	0.6	1.500	1.886	0.270	0.600	0.360	0.216	0.130	0.600	0.000	0.000
11		0.280	0.3	4.000	2.612	0.008	0.300	0.090	0.027	0.008	0.300	0.000	0.000
12		0.160	0.6	4.000	3.550	0.019	0.600	0.360	0.216	0.130	0.600	0.000	0.000

Numbers in italic font provide a characterization of the respective scenario

Table 2 The AR(2)-type models with $n = 10$ used in the simulation study and their properties, where (P)ACF $_k$ refers to the (P)ACF at lag k

Scenario	DGP	Parameters		Mean	Var	$P(X = 0)$	ACF1	ACF2	ACF3	ACF4	PACF1	PACF2	PACF3	PACF4
		α	β	ϕ_1										
BAR(2)														
13		0.524	0.084	0.545	1.500	1.275	0.197	0.300	0.272	0.125	0.084	0.300	0.200	0.000
14		0.728	0.048	0.706	1.500	1.275	0.197	0.600	0.488	0.354	0.268	0.600	0.200	0.000
15		0.664	0.224	0.545	4.000	2.400	0.006	0.300	0.272	0.125	0.084	0.300	0.200	0.000
16		0.808	0.128	0.706	4.000	2.400	0.006	0.600	0.488	0.354	0.268	0.600	0.200	0.000
bvAR(2)														
		π	ϕ_1	ϕ_2										
17		0.15	0.24	0.2	1.500	2.111	0.294	0.300	0.272	0.125	0.084	0.300	0.200	0.000
18		0.15	0.48	0.2	1.500	3.286	0.397	0.600	0.488	0.354	0.268	0.600	0.200	0.000
19		0.40	0.24	0.2	4.000	3.974	0.039	0.300	0.272	0.125	0.084	0.300	0.200	0.000
20		0.40	0.48	0.2	4.000	6.186	0.096	0.600	0.488	0.354	0.268	0.600	0.200	0.000
BARCh(2)														
		α_0	α_1	α_2										
21		0.084	0.24	0.2	1.500	1.439	0.218	0.300	0.272	0.125	0.084	0.300	0.200	0.000
22		0.048	0.48	0.2	1.500	1.953	0.277	0.600	0.488	0.354	0.268	0.600	0.200	0.000
23		0.224	0.24	0.2	4.000	2.708	0.009	0.300	0.272	0.125	0.084	0.300	0.200	0.000
24		0.128	0.48	0.2	4.000	3.676	0.021	0.600	0.488	0.354	0.268	0.600	0.200	0.000

Numbers in italic font provide a characterization of the respective scenario

3.2 Estimation performance for adequate models

As the first task of our simulation study, we examine the performance of the ML estimation under the premise that the correct type of model is fitted to the given time series data; the full estimation results are summarized in Supplement S.2. This corresponds to the ideal case. We analyze the performance when determining the true type of model underlying the DGP in Sect. 4 and the consequences of fitting the wrong model to the DGP in Sect. 5. As shown in the boxplots in Supplement S.2, the conditional ML estimation works well for all the considered models. There might be some biases and a notable dispersion for the small sample size, $T = 100$ (particularly for the second-order models), but both decrease quickly with increasing sample size. Thus, provided an adequate model has been chosen (and that the time series is sufficiently large), an ML estimation is a reliable approach for model fitting for any of the considered AR-type models.

4 Using information criteria for model selection

In this section, we focus on the model selection considering the six types of candidate models presented in Sect. 3.1. Since these models are not nested, traditional likelihood ratio tests cannot be used for this purpose. In contrast, “a substantial advantage in using information-theoretic criteria is that they are valid for nonnested models” (Burnham and Anderson 2002, p. 88). We consider the AIC and the BIC, which are the most widely known and used model selection tools in statistical practice (Burnham and Anderson 2002; Cavanaugh and Neath 2019; Dziak et al. 2019). Both criteria penalize the log-likelihood values by adding a term that increases with the number of parameters (and for the BIC, with the number of observations). Therefore, the criteria balance the model fit against the model complexity. If these criteria are applied to *nested* candidate models, they will essentially “act like a hypothesis test with a particular α level” (Dziak et al. 2019, p. 5). But also in the case of *nonnested* models, they control “the tradeoff between the likelihood term and the penalty on the number of parameters, hence the tradeoff between good fit to the observed data and parsimony” (Dziak et al. 2019, p. 6).

The AIC is twice the negative maximized log-likelihood plus a penalty term, which is equal to twice the number of model parameters; that is,

$$\text{AIC} = -2\ell_{\max} + 2n_{\text{model}}, \quad (4.1)$$

where ℓ_{\max} is the value of the maximized log-likelihood and n_{model} is the number of model parameters.

Similar to the AIC, the BIC adjusts the log-likelihood by adding a penalty term that grows proportionally with the number of model parameters, but this also depends on the number of observations in the sample, T :

$$\text{BIC} = -2\ell_{\max} + n_{\text{model}} \ln T. \quad (4.2)$$

The AIC and BIC can be applied to the conditional log-likelihood and to the full log-likelihood (for the first approach, see Sect. 3). Because the number of terms in $\ell(\boldsymbol{\theta} | x_p, \dots, x_1)$, the conditional log-likelihood, varies with p , we insert the corrective factor $T/(T - p)$ before ℓ_{\max} in (4.1) and (4.2) to account for this distortion (Weiß 2018).

As noted by Burnham and Anderson (2002) and Cavanaugh and Neath (2019), the AIC and BIC were originally designed for different objectives. The AIC provides an asymptotically unbiased estimator of the expected Kullback–Leibler discrepancy, whereas the BIC serves as an asymptotic approximation to a transformation of the Bayesian posterior probability of a candidate model. In practice, however, they are usually treated as competitors in the context of model selection. With increasing sample size T , the BIC tends to choose fitted models that are more parsimonious than those favored by the AIC, because its penalty is larger than that of the AIC if $\ln T > 2$ (or $T \geq 8$). Thus, the choice between AIC and BIC might be guided by “the relative importance one assigns to sensitivity versus specificity” (Dziak et al. 2019, p.1).

One of the crucial questions examined in this research is how well the AIC and BIC perform in terms of selecting an AR-type model for a bounded-counts time series. This question has yet to be investigated in the literature. For real-valued time series, Rinke and Sibbertsen (2016) evaluated the ability of information criteria to distinguish between linear and nonlinear models, and Psaradakis et al. (2009) considered the problem of selecting from among competing nonlinear time series models that belong to different families. Information criteria are generally recommended for selecting from nonnested time series models (Burnham and Anderson 2002, Sect. 2.12.4); some asymptotic considerations regarding regression models can be found in Sect. 8.7 of Claeskens and Hjort (2008). However, little is known about such an ability in nonnested and nonlinear time series models *for counts* (Jung et al. 2016). Two recent studies, Weiß and Feld (2020) and Weiß et al. (2019), partially follow the direction outlined by Jung et al. (2016), but concentrate on the case of *unbounded counts*. Diop and Kenge (2020) propose a penalized criterion relying on a Poisson quasi-likelihood approach for some INGARCH-type processes of counts, and they prove its consistency under certain regularity conditions. We also point out the work by Katz (1981), but this cannot be applied here, because it focuses on the order selection of general finite Markov processes, that is, without the parametric relations implied by the models in Sect. 2. Therefore, in this study, we evaluate the performance of various information criteria for simultaneous model class and lag order selection using the three CLAR-type models for bounded-counts time series. In Sect. 6, we also investigate the ability of the criteria to distinguish between models that have a linear or nonlinear autocorrelation structure.

The performance of the model selection criteria (AIC and BIC) is assessed using the Monte Carlo simulations described in Sect. 3.1. Once a time series has been simulated, the six time series models listed in Tables 1 and 2 (i. e., BAR(1), BAR(2), bvAR(1), bvAR(2), BARCH(1), BARCH(2)) are fitted. Then, we store the CML estimates and the resulting maximized conditional log-likelihood values (ℓ_{\max}). Next, we calculate the AIC and BIC values for each model according to (4.1) and (4.2), respectively (together with the corrective factor $T/(T - p)$), and determine the model that yields the lowest AIC or BIC value. This process is repeated 1000 times to compute the

number of times each criterion selected each of the six possible models. The full results are provided in Supplement S.1.

The first part of each table in Supplement S.1 shows the number of times each candidate model was selected for a given type of DGP based on the AIC or BIC. Here, the number of correct identifications is in italics. The second part of each table shows the average of the deviations $AIC_i - AIC_{\min}$ and $BIC_i - BIC_{\min}$, where the subscript “*i*” denotes the model under consideration and the subscript “min” denotes the minimal AIC (or BIC) value for the respective time series. In routine model selection applications, the optimal fitted model is that which minimizes the selected information criterion. However, models with similar values should receive the same ranking by a criterion. A common rule of thumb is to treat any fitted model as a viable candidate if it yields an AIC value within two units of the minimal AIC (Burnham and Anderson 2002, Sect. 2.6); an analogous procedure is recommended by Neath and Cavanaugh (2012) for the BIC. These refined model selections can be evaluated using the respective second parts of the tables in Supplement S.1.

For all scenarios, the rates of successful identification generally increase with T . For the largest sample size $T = 1000$, these rates range from 70 to 100% for the AIC and from 80 to 100% for the BIC. In general, for the model order $p = 1$, the performance of the BIC is similar or superior to that of the AIC. For the model order $p = 2$, however, the reverse is true; that is, the performance of the AIC is superior to that of the BIC. For illustration, the results for scenarios 4, 8, 12, 16, 20, and 24, with mean 4 and $\rho(1) = 0.6$, are discussed in more detail. For the DGP scenario 4 (BAR(1)), the AIC’s rates for identifying the correct model are 83.7%, 83.8%, 87.8% and 88.5% for $T = 100, 250, 500, 1000$, respectively, and the BIC’s success rates quickly tend to 100%. Analogous identification rates are observed for the bvAR(1) scenario 8, whereas scenario 12 (BARCH(1)) leads to slightly lower success rates for small T . It is also interesting to investigate the means of $AIC_i - AIC_{\min}$ and $BIC_i - BIC_{\min}$. For scenario 4, we have a mean $AIC_{\text{BAR}(2)} - AIC_{\min} \leq 2$ for all T , but a mean $BIC_{\text{BAR}(2)} - BIC_{\min} > 4$ for all T (Supplement S.1). That is, following the two-units rule, the AIC also considers the BAR(2) model a substantial candidate (in the mean) if the DGP is BAR(1) according to scenario 4, whereas the BIC does not make such a mistake. These results indicate that the BIC leads to better results regarding the model order for scenario 4. Analogous conclusions hold for scenarios 8 and 12 and, more generally, for model order $p = 1$. Note, however, that for some BAR(1) scenarios, there is a notable rate of BARCH(1) misidentifications (up to 20–30%); that is, a model from outside the DGP’s model family is chosen. This time, these rates are slightly larger for the BIC. The same phenomenon is observed in the opposite direction; that is, a BARCH(1) DGP is often misidentified as BAR(1).

Next, we consider the model order $p = 2$. For scenario 16 (BAR(2), with mean 4 and $\rho(1) = 0.6$), and a small sample size $T = 100$, we observe somewhat lower rates of successful identification, 69.4% (AIC) and 45.9% (BIC). Furthermore, the rates of falsely voting for BAR(1) are 29.1% (AIC) and 51.9% (BIC). For $T \geq 250$, the correct rates exceed 94.8% (AIC) and 83.7% (BIC). Therefore, the AIC now demonstrates better model selection than the BIC, which is very different from scenario 4, where the BIC leads to better results. The means of both $AIC_{\text{BAR}(1)} - AIC_{\min}$ and $BIC_{\text{BAR}(1)} - BIC_{\min}$ are larger than two; that is, based on the two-units rule, we do not have

misidentifications such as BAR(1) in the mean. Analogous conclusions are drawn for scenario 20 (bvAR(2)).

In scenario 24 (BARC(2)), the situation is more complicated. The rates of correct identifications are lower than before, 54.6% (AIC) and 29.6% (BIC) for $T = 100$. In particular, the AIC misidentifies as BARC(1) with 27.6%, and the BIC falsely selects BARC(1) in most cases (53.2%). Additionally, the mean of $\text{AIC}_{\text{BARC}(1)} - \text{AIC}_{\min} = 2.92 > 2$, but the mean of $\text{BIC}_{\text{BARC}(1)} - \text{BIC}_{\min} = 1.68 < 2$. For $T \geq 250$, the rates of correct identifications are above 90.5% (AIC) and 72.1% (BIC). This implies that the BIC tends to choose a model order that is too low for $T = 100$, but its performance improves for $T \geq 250$. Finally, also for the second-order DGPs, there may be misidentifications across different model families, particularly between the BAR and BARC models.

5 Properties of (Mis)fitted models

Our analyses in Sect. 4 showed that, depending on the sample size and the type of DGP, there is considerable risk of selecting the wrong model. Most often, the wrong model order within the correct model family is chosen, but we observed a large proportion of cases for which an incorrect model family was selected. Therefore, we next examine the effect of an incorrect model selection, as the selected model might later be applied to, for example, the forecasting of future observations. More precisely, we investigate this effect by comparing relevant moment properties (mean, BID, and ACF) of the fitted model with the same properties of the true DGP. In this context, we analyze the influence of estimation uncertainty when the correct model is selected. Our analyses rely on the Monte Carlo simulations described in Sect. 3.1. We calculate the properties from the estimated parameters for each replication and each model fit. The means across all replications are tabulated in Sect. S.3 of the supplementary material for all considered scenarios and all time series lengths T . In the following, we present a summary of our findings. We focus on the BID and ACF, because the fitted models' mean is quite close to the true mean in most cases.

If the DGP is BAR(1) (scenarios 1–4), then the fitted BAR models have a BID of one (by definition), and they capture the ACF structure very well, even for $T = 100$. In contrast, we observe additional dispersion for the fitted bvAR and BARC models. These models exhibit extra-binomial variation, by construction, and they cannot match a BID of one. However, in contrast to the bvAR models, the BARC models have good fit to the true ACF (in the mean). The bvAR models only reach an acceptable fit for the ACF for Scenario 4 (mean $\mu = 4$, $\rho(1) = 0.6$), whereas the BID for these fitted models is larger than two.

For the case of a bvAR(1) DGP (scenarios 5–8), for small sample sizes, we observe somewhat larger discrepancies when we fit the correct models compared to the previous scenarios. For the misfitted models, the BAR models underestimate the true ACF and BID. The BARC models again fit the ACF quite well, but they underestimate the BID.

If the DGP is BARC(1) (scenarios 9–12), we again find some discrepancy between the true and fitted BARC models' ACF if T is small. The (mis)fitted BAR models

underestimate the BID and ACF, particularly for higher autocorrelation levels. For the fitted bvAR models, we observe a visible underestimation of the mean. The estimated BID is, in general, higher than the true BID, whereas the opposite holds for the ACF. The underestimation of the ACF decreases with a higher autocorrelation level.

For the second-order DGPs, the ACF at lag one is well captured, in general, by the fitted first-order models from the same family. The other results for the first-order DGPs are transferable to the second-order DGPs. The BARCH and BAR models are closer to each other than they are to the other competitor: the bvAR models have unique features in their structure that cannot be mimicked by the other models.

Overall, it appears that the fitted BARCH models match the ACF quite well for most models and scenarios, whereas the BID is not captured well. The BAR and bvAR models cannot mimic the ACF of different DGPs. More flexibility for the BID can be obtained by switching from the basic binomial distribution to a distribution with an additional dispersion parameter, such as the beta-binomial distribution if extra-binomial variation is required. Returning to our initial question of how different these models are, we find that the three CLAR-type models for bounded counts differ notably in terms of their model properties. However, the BARCH models are relatively flexible and mimic at least the true ACF structure quite well, even if the underlying DGP follows another CLAR-type model.

6 Further analyses

In what follows, we conduct further analyses regarding the choice of the model order (Sect. 6.1) and regarding nonlinear AR-type DGPs (Sects. 6.2, 6.3). To keep our analyses manageable, we limit to the BARCH model among the CLAR models for bounded counts because this model is the most flexible in imitating the true DGP's autocorrelation structure.

6.1 More details on order selection

As shown in Sect. 4, the most common type of misidentification is the selection of the wrong model order within the true model class. For this reason, we conduct a further simulation experiment focusing on model order selection among BARCH(p) models. Extending the previous scenarios 9 & 10 ($p = 1$) and 21 & 22 ($p = 2$), we consider two 3rd-order DGPs (numbered 25 & 26) having the marginal mean $\mu = 1.5$; see Table 3 for details. Then we simulate each of the six DGPs 9, 10, 21, 22, 25, and 26, as described in Sect. 3.1, where the set of candidate models is given by BARCH(1–3) in each case. The full model selection results are summarized in Supplement S.4.

The selection results for the first-order BARCH DGP are completely analogous to our finding in Sect. 4: the wrong model order is rarely selected (in particular, the BIC shows excellent performance), and in the case of a misidentification, mainly the BARCH(2) model is selected. The reverse scenario is given by a BARCH(3) DGP (with BARCH(1–2) being further candidates). Now, the rates of successful order selection are much lower, particularly for $T \leq 250$. The BIC performs particularly

Table 3 The BARCH(3) models with $n = 10$ used in the simulation study and their properties, where (P)ACF k refers to the (P)ACF at lag k

Scenario (Sc.) 25:	$(\alpha_0, \alpha_1, \alpha_2, \alpha_3) = (0.068, 0.194, 0.2, 0.15)$									
Sc. 26:	$(\alpha_0, \alpha_1, \alpha_2, \alpha_3) = (0.037, 0.400, 0.2, 0.15)$									
Sc.	Mean	Var	$P(X = 0)$	(P)ACF1	ACF2	ACF3	ACF4	PACF2	PACF3	PACF4
25	1.500	1.488	0.224	0.300	0.303	0.269	0.158	0.234	0.150	0.000
26	1.500	2.046	0.287	0.600	0.530	0.482	0.389	0.266	0.150	0.000

badly, choosing the wrong model order more often than the true model order (for DGP 25, it votes for the first-order model in most cases). Even for sample size $T = 500$, the rate of successfully identifying the model order $p = 3$ is still below 80%.

A particularly interesting case is that of the BARCH(2) DGP, because here, a model-order misidentification is possible in both directions, toward a model order that is too large or too small. The results in Supplement S.4 show that the most common misidentification is choosing p too small. For the BIC, it is nearly the only type of misidentification, whereas the AIC shows a more complex pattern: choosing a too small p mainly occurs for short time series ($T = 100$), while choosing p too large happens for all T in about 10–15% of all cases, with slightly increasing values for increasing T . In summary, the AIC always has a moderate tendency to choose the model order p too large. But for small sample sizes, and particularly for the BIC, there is a much larger risk of choosing a model order that is too small. This tendency should be kept in mind when interpreting the actual decision of AIC or BIC for a given data example.

6.2 Nonlinear AR-type models: max-BAR(p) model

In this and the subsequent section, we consider the use of AIC and BIC when deciding between types of linear and nonlinear AR models. More precisely, in both sections, we use the BARCH model as the representative CLAR-type model, but we consider two different types of nonlinear AR-type models for bounded counts. In the present section, it is the *max-BAR(p) model* (Weiß et al. 2018), which is defined by

$$X_t = \max \{ \alpha_1 \circ X_{t-1}, \dots, \alpha_p \circ X_{t-p}, \epsilon_t \}, \tag{6.1}$$

where all thinnings are performed independently, and the innovations ϵ_t are i. i. d. bounded counts (here, we assume $\epsilon_t \sim \text{Bin}(n, \pi)$). Analytic formulae for the mean, variance, and (P)ACF are not available, but these properties can be computed numerically because (6.1) defines a p th-order Markov process (Weiß et al. 2018). The conditional cumulative distribution function (CDF) is given by

$$F_{X_t | x_{t-1}, \dots, x_{t-p}}(x) = F_{\alpha_1 \circ x_{t-1}}(x) \cdots F_{\alpha_p \circ x_{t-p}}(x) \cdot F_\epsilon(x), \tag{6.2}$$

Table 4 The max-BAR(1) and max-BAR(2) models with $n = 10$ used in the simulation study and their properties, where (P)ACF k refers to the (P)ACF at lag k

Scenario (Sc.) 27:			$(\pi, \alpha_1) = (0.15, 0.3)$	Sc. 31:			$(\pi, \alpha_1, \alpha_2) = (0.15, 0.24, 0.2)$			
Sc. 28:			$(\pi, \alpha_1) = (0.15, 0.6)$	Sc. 32:			$(\pi, \alpha_1, \alpha_2) = (0.15, 0.48, 0.2)$			
Sc. 29:			$(\pi, \alpha_1) = (0.40, 0.3)$	Sc. 33:			$(\pi, \alpha_1, \alpha_2) = (0.40, 0.24, 0.2)$			
Sc. 30:			$(\pi, \alpha_1) = (0.40, 0.6)$	Sc. 34:			$(\pi, \alpha_1, \alpha_2) = (0.40, 0.48, 0.2)$			
Sc.	Mean	Var	$P(X = 0)$	(P)ACF1	ACF2	ACF3	ACF4	PACF2	PACF3	PACF4
27	1.629	1.117	0.118	0.103	0.009	0.001	0.000	-0.002	0.000	0.000
28	1.879	0.998	0.050	0.307	0.087	0.023	0.006	-0.008	-0.002	0.000
29	4.051	2.225	0.002	0.029	0.000	0.000	0.000	0.000	0.000	0.000
30	4.275	1.870	0.000	0.171	0.023	0.002	0.000	-0.007	0.000	0.000
31	1.658	1.055	0.092	0.070	0.052	0.005	0.002	0.047	-0.002	-0.001
32	1.805	0.971	0.051	0.208	0.071	0.017	0.005	0.029	-0.003	0.000
33	4.048	2.215	0.001	0.016	0.010	0.000	0.000	0.010	0.000	0.000
34	4.165	1.987	0.000	0.093	0.013	0.001	0.000	0.004	-0.001	0.000

where $F_{\alpha ol}(k) = \sum_{m=0}^{\min\{k,l\}} \binom{l}{m} \alpha^m (1 - \alpha)^{l-m}$ is the CDF of the binomial thinning operation, and $F_\epsilon(x)$ is the CDF of the innovation term. (6.2) is used to compute the transition probabilities, as follows:

$$p(x|x_{t-1}, \dots, x_{t-p}) = F_{X_t|x_{t-1}, \dots, x_{t-p}}(x) - F_{X_t|x_{t-1}, \dots, x_{t-p}}(x - 1). \tag{6.3}$$

For our simulation experiments, we choose the max-BAR parameters (scenarios 27–34 in Table 4) to be equal to those of the bvAR DGPs in Tables 1 and 2. The obtained simulation results are summarized in Supplement S.5.

If the true DGP is BARCH(1–2), it is hardly misidentified as a max-BAR process. Misidentifications mainly happen within the BARCH family, in the way described in Sects. 4 and 6.1. An analogous statement holds for the reverse situation (slightly weakened regarding DGPs 29 and 33), so these types of linear and nonlinear models are well separated by AIC and BIC. But another problem occurs. While the rates of successfully identifying a max-BAR(1) DGP behave analogously to the other AR(1)-type DGPs in Sect. 4 (quickly increasing to around 80% for the AIC and nearly 100% for the BIC with increasing T), the identification of a max-BAR(2) DGP might fail. For scenarios 33 and 34, the success rate is small even for $T = 1000$: 52% and 37%, respectively, for the AIC, and below 1% for the BIC in both cases. These low rates are essentially caused by falsely choosing the max-BAR(1) model. Furthermore, in scenarios 33 and 34, we find that $AIC_{\max\text{-BAR}(2)} - AIC_{\min} \leq 2$ in the mean, but $BIC_{\max\text{-BAR}(2)} - BIC_{\min} > 2$. Thus, the BIC does not even consider the max-BAR(2) model a reasonable candidate based on the two-units rule. Even if the correct model is selected, an ML estimation shows relatively large dispersion (see Supplement S.5), although the dispersion, as well as the bias, decrease with increasing T .

Thus, it is natural to ask why we have such problems with the max-BAR DGPs. For ordinary max-AR models for real-valued time series (Davis and Resnick 1989), it is

known that higher-order autoregressions might not be uniquely determined. Consider the max-AR(2) model as an example, defined by $Y_t = \max \{a_1 \cdot Y_{t-1}, a_2 \cdot Y_{t-2}, \varepsilon_t\}$. If $a_1^2 \geq a_2$, this model can be reduced to a max-AR(1) model (Davis and Resnick 1989, pp. 790–791); thus, the model order is not unique. This ambiguity is intuitively clear: because $Y_{t-1} \geq a_1 \cdot Y_{t-2}$, by model recursion, we have $a_1 \cdot Y_{t-1} \geq a_1^2 \cdot Y_{t-2}$, which is $\geq a_2 \cdot Y_{t-2}$. Therefore, the max operator at time t will never select $a_2 \cdot Y_{t-2}$. This type of non-identifiability causes problems in parameter estimation; see Zhang and Smith (2010) and Kunihama et al. (2012).

For the max-BAR(2) model in (6.1), we do not have such an exact reducibility because we use probabilistic thinning operators instead of multiplications. Nevertheless, we have an analogous tendency: if α_2 is much smaller than α_1^2 , the probability $P(\alpha_1 \circ X_{t-1} \geq \alpha_2 \circ X_{t-2})$ will be quite large, but there is always a nonzero probability that $\alpha_1 \circ X_{t-1} < \alpha_2 \circ X_{t-2}$ (e. g., if $\alpha_1 \circ X_{t-1} = 0$, but $\alpha_2 \circ X_{t-2} > 0$). However, if the innovation term ε_t has a large mean, then the max operator will most often select the innovation term. Thus, both AR components are rarely activated, which makes model identification difficult. This result is also evident for models 33 and 34 in Table 4. The mean parameter π is rather large and, as a result, the serial dependence is extremely weak (e. g., $\rho(1) = 0.016$ or $\rho(1) = 0.093$). Thus, it is extremely difficult to correctly identify the second-order structure.

6.3 Nonlinear AR-type models: logit-BARCH(p) model

As the second type of nonlinear AR model for bounded counts, we consider the logit-BARCH(p) model developed by Chen et al. (2019). The model's definition is actually closely related to that of the BARCH(p) model in Sect. 2.3, but it uses a logit link function instead of a linear one. More precisely, defining $\text{logit}(x) = \ln\left(\frac{x}{1-x}\right)$ (so its inverse equals $\text{logit}^{-1}(x) = (1 + \exp(-x))^{-1}$), the logit-BARCH(p) model requires that $X_t|X_{t-1}, \dots$ be conditionally binomially distributed according to $\text{Bin}(n, \pi_t)$, where

$$\text{logit}(\pi_t) = r_0 + \sum_{i=1}^p r_i X_{t-i} \quad (6.4)$$

with $r_j \in \mathbb{R}$ for $j = 0, \dots, p$. Thus, the model parameters r_j are not constrained, which is attractive for parameter estimation. In fact, the transition probabilities as required for ML estimation are simply given by

$$p(x|x_{t-1}, \dots, x_{t-p}) = \binom{n}{x} \pi_t^x (1 - \pi_t)^{n-x} \quad (6.5)$$

$$\text{with } \pi_t = \text{logit}^{-1}\left(r_0 + \sum_{i=1}^p r_i x_{t-i}\right).$$

On the other hand, analytic formulae for the mean, variance, and (P)ACF are not available. These properties have to be computed numerically using the Markov property (Weiß et al. 2018), as an analogy to Sect. 6.2. This is done for the simulation scenarios

Table 5 The logit-BARCH(1) and logit-BARCH(2) models with $n = 10$ used in the simulation study and their properties, where (P)ACF k refers to the (P)ACF at lag k

Scenario (Sc.) 35:			$(r_0, r_1) = (-2.14, 0.20)$				Sc. 39: $(r_0, r_1, r_2) = (-2.39, 0.20, 0.1)$			
Sc. 36:			$(r_0, r_1) = (-2.75, 0.40)$				Sc. 40: $(r_0, r_1, r_2) = (-2.99, 0.35, 0.1)$			
Sc. 37:			$(r_0, r_1) = (-0.94, 0.13)$				Sc. 41: $(r_0, r_1, r_2) = (-1.24, 0.11, 0.1)$			
Sc. 38:			$(r_0, r_1) = (-1.66, 0.30)$				Sc. 42: $(r_0, r_1, r_2) = (-1.92, 0.25, 0.1)$			
Sc.	Mean	Var	$P(X = 0)$	(P)ACF1	ACF2	ACF3	ACF4	PACF2	PACF3	PACF4
35	1.354	1.240	0.243	0.247	0.062	0.015	0.004	0.001	0.000	0.000
36	0.886	0.938	0.420	0.382	0.155	0.065	0.028	0.010	0.003	0.001
37	3.950	2.614	0.009	0.309	0.095	0.029	0.009	0.000	0.000	0.000
38	3.851	4.026	0.028	0.675	0.458	0.311	0.211	0.004	0.001	0.000
39	1.184	1.122	0.296	0.253	0.169	0.067	0.035	0.112	0.001	0.000
40	0.667	0.676	0.515	0.280	0.147	0.062	0.029	0.075	0.003	0.001
41	4.051	2.838	0.009	0.343	0.327	0.167	0.122	0.238	0.000	0.000
42	3.327	4.024	0.052	0.684	0.585	0.466	0.381	0.220	0.007	0.004

35–42 shown in Table 5. The choice of the parameter values is motivated as follows. As shown in Sect. 6.2, if the nonlinear model behaves very differently from the considered linear candidate model, the information criteria can well discriminate between these model families. Therefore, we attempt to choose the logit-BARCH DGPs such that the shape of the actual inverse link function is similar to that of the respective BARCH(p) model in Tables 1 and 2. We expect that the model selection will be more difficult this time.

The obtained simulation results are summarized in Supplement S.6. As in the previous sections, the results are composed of boxplots of the simulated estimates for the case of fitting the adequate model. As an analogy to the other DGPs considered in this paper, the boxplots confirm the asymptotic unbiasedness and consistency of the ML estimators, although the dispersion might be quite large for the small sample size $T = 100$ (particularly for the estimators \hat{r}_1 and \hat{r}_2 of the second-order model). In addition, Supplement S.6 again provides tables with selection numbers and average distances to the minimal AIC or BIC, respectively. As conjectured, the model selection is quite problematic in some cases. For the BARCH(1) DGP 11, the AIC more often selects the logit-BARCH family, without notable improvement for increasing T . The BIC does slightly better because it rarely misidentifies as a second-order model. The reverse is observed for the logit-BARCH(1) DGP 37, where we have a stable number of BARCH(1) false selections. For the remaining scenarios (including all second-order DGPs), the fraction of correct model identification increases with increasing T , although sometimes quite slowly. Thus, we conclude that the BARCH and logit-BARCH models might sometimes behave so similarly that a reliable discrimination by AIC and BIC is not possible. This underlines the necessity to not rely solely on information criteria for model choice, but to also consider further diagnostic tools.

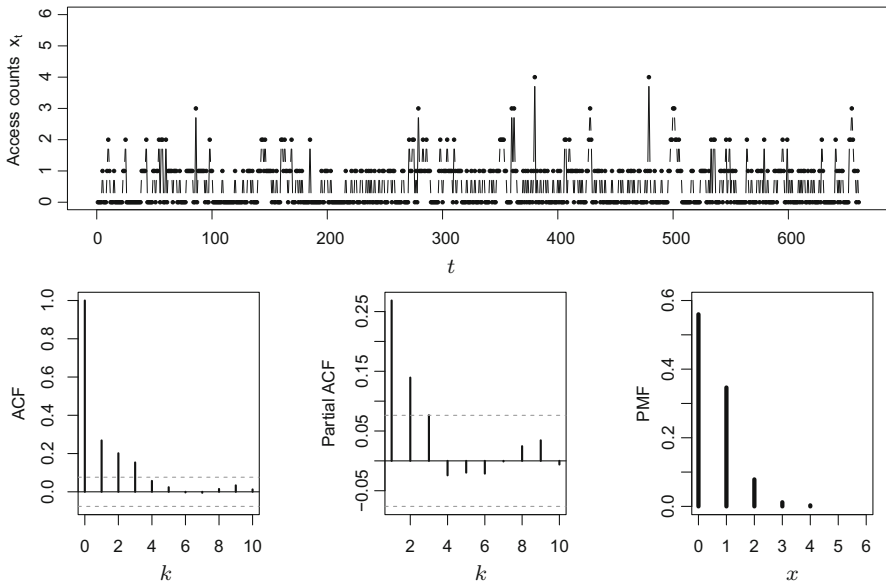


Fig. 1 Access counts time series discussed in Sect. 7: plots of time series, sample ACF, sample PACF, and sample probability mass function (PMF)

7 Real-data application: access counts

As a real-data application of model selection, we consider the data set introduced by Weiß (2009) that records access to the websites of $n = 6$ staff members from a statistics department. The counts x_1, \dots, x_{661} provide the number of staff members whose home directory was accessed for each minute between 10 AM and 9 PM on November 29, 2005. As shown by Weiß (2009) (see also the (P)ACF plot in Fig. 1), this bounded-counts time series has an AR-type autocorrelation structure, and a BAR(p) model (Sect. 2.1) of order $p = 2$ or $p = 3$ works reasonably well. At the time of the data analysis, the four other AR-type models for bounded-counts time series (those from Sects. 2.2, 2.3, 6.2 and 6.3) were not known; thus, Weiß (2009) had to limit the set of candidate models to BAR(p) models. In what follows, we perform a new model selection based on a much larger set of candidate models: all AR(p)-type models discussed in Sects. 2 and 6, with model orders $p = 1, 2, 3$. The full estimation results are provided in Supplement S.7.

The obtained values for the AIC and BIC are summarized in Table 6(a). The models that minimize the AIC and BIC (in bold font) are the BAR(3) or BAR(2) models, respectively, as in Weiß (2009). However, if we consider the two-units rule (see the values in italic font), it becomes clear that the corresponding BARCH models also appear to be a reasonable choice, whereas the max-BAR and bvAR models clearly perform worse. Among the logit-BARCH models, the second-order model should be further considered. This appears plausible in view of our previous findings: the max-BAR and bvAR models clearly differ from the BAR model, whereas the BARCH model is relatively flexible, with the logit-BARCH model sometimes coming close.

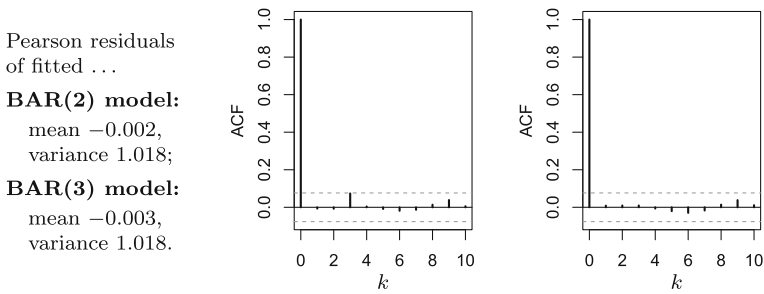


Fig. 2 Access counts time series discussed in Sect. 7: analysis for Pearson residuals for the fitted BAR(2) and BAR(3) model. Summary with their means and variances, plots of their sample ACF (left: $p = 2$; right: $p = 3$)

On the other hand, this time, we do not know the true model underlying the DGP. Thus, it is important to verify and explain these decisions and distinctions, including whether they are perhaps misleading. Therefore, we next compare the properties of the fitted models in Table 6(b) with those of the sample.

From Table 6(b), it becomes clear that the BAR and BARCH models outperform the bvAR and max-BAR models in terms of the marginal properties (particularly the variance and zero probability) and the ACF. The bvAR models, and especially the max-BAR models, exhibit much lower autocorrelation values than are observed in the time series itself. Therefore, these types of models are not adequate for the data, confirming the decisions based on the AIC and BIC. For the logit-BARCH models, we observe that the third-order model does not show clear improvement over the second-order model, and both have a lower lag-three ACF than is observed in the data. Therefore, the models do not match as well as the BAR(3) and BARCH(3) models. But if selecting among the second-order models, the logit-BARCH(2) model performs quite well. If we compare the BAR and BARCH models, we find similar performance for both $p = 2$ and $p = 3$. The BAR models' variance and zero probability are slightly closer to the corresponding sample properties than those of the BARCH models, supporting the slight preference for the BAR models of both the AIC and BIC. The closeness between the ACFs of the models and the sample is effected more by the actual model order than by the model type. Weiß (2009) noted that a “third-order model should be preferred” for the serial dependence structure, despite the BIC’s preference for order $p = 2$. This conclusion is confirmed by the results in Table 6(b) for both the BAR and BARCH models: the ACFs at lags three and four are well approximated only if $p = 3$, with a slight advantage for the BAR model.

Thus, depending on the considered information criterion, model selection leads to either the BAR(2) or BAR(3) model. In any case, the result for model selection must be checked for model adequacy. A widely used tool for this purpose is the (standardized) Pearson residuals (Jung et al. 2016; Weiß 2018). For an adequate model, the Pearson residuals should have a sample mean close to 0, a variance close to 1, and ACF values close to 0. For both the fitted BAR(2) or BAR(3) model, we compute the resulting Pearson residuals and conduct the analyses summarized in Fig. 2. We find that the mean and variance satisfy the described requirements quite well, so both models seem to adequately describe the data’s (conditional) mean and variance. However, for the

Table 6 Access counts time series discussed in Sect. 7: AIC and BIC in (a), sample properties and fitted models' properties in (b)

P	AIC		BIC		Mean	Var	$P(X = 0)$	ACF1	ACF2	ACF3	ACF4	
	BAR(p)	logit-BARCH(p)	max-BAR(p)	BAR(p)								
<i>(a)</i>												
1	1235.4	1236.4	1255.2	1244.4	1251.9	1236.4	1255.2	1236.8	1244.4	1260.9	1245.4	1264.2
2	<i>1228.8</i>	<i>1229.6</i>	1255.9	1242.3	1252.4	<i>1229.6</i>	1255.9	<i>1228.9</i>	1242.3	1265.9	<i>1243.1</i>	1269.4
3	1228.3	<i>1230.3</i>	1258.9	1246.3	1255.4	<i>1230.3</i>	1258.9	1232.0	1246.3	1273.4	1248.3	1276.8
<i>(b)</i>												
Data	0.552	0.514	0.560	0.269	0.560	0.514	0.560	0.269	0.201	0.153	0.057	0.057
BAR(1)	0.553	0.502	0.560	0.261	0.560	0.502	0.560	0.261	0.068	0.018	0.005	0.005
bvAR(1)	0.562	0.624	0.585	0.219	0.562	0.624	0.585	0.219	0.048	0.011	0.002	0.002
BARCH(1)	0.553	0.533	0.570	0.263	0.553	0.533	0.570	0.263	0.069	0.018	0.005	0.005
max-BAR(1)	0.556	0.440	0.532	0.164	0.556	0.440	0.532	0.164	0.026	0.004	0.001	0.001
logit-BARCH(1)	0.554	0.537	0.569	0.269	0.554	0.537	0.569	0.269	0.075	0.021	0.006	0.006
BAR(2)	0.555	0.504	0.559	0.250	0.555	0.504	0.559	0.250	0.179	0.070	0.038	0.038
bvAR(2)	0.569	0.659	0.588	0.201	0.569	0.659	0.588	0.201	0.113	0.036	0.015	0.015
BARCH(2)	0.555	0.540	0.570	0.257	0.555	0.540	0.570	0.257	0.185	0.074	0.040	0.040
max-BAR(2)	0.559	0.426	0.523	0.153	0.559	0.426	0.523	0.153	0.068	0.017	0.005	0.005
logit-BARCH(2)	0.558	0.533	0.571	0.289	0.558	0.533	0.571	0.289	0.222	0.101	0.060	0.060
BAR(3)	0.559	0.507	0.556	0.245	0.559	0.507	0.556	0.245	0.178	0.141	0.067	0.067
bvAR(3)	0.571	0.666	0.588	0.196	0.571	0.666	0.588	0.196	0.110	0.051	0.020	0.020
BARCH(3)	0.556	0.541	0.570	0.250	0.556	0.541	0.570	0.250	0.182	0.132	0.065	0.065
max-BAR(3)	0.559	0.426	0.523	0.153	0.559	0.426	0.523	0.153	0.067	0.016	0.005	0.005
logit-BARCH(3)	0.559	0.554	0.570	0.277	0.559	0.554	0.570	0.277	0.212	0.096	0.057	0.057

(a) Minimal value of AIC (BIC) in bold font; AIC (BIC) values deviating by ≤ 2 from minimal AIC (BIC) in italic font

(b) Sample properties highlighted in italic font

fitted BAR(2) model as selected by the BIC, there is a notably large ACF value at lag three (≈ 0.074), whereas all other ACF values are reasonably close to 0. This indicates that the chosen model order might have been too small. In fact, the AIC's choice, the BAR(3) model, has ACF values close to 0 throughout. Therefore, together with our previous findings, this leads us to conclude that the BAR(3) model should be selected for the access counts time series.

8 Conclusion

We compared five AR-type models for bounded-counts time series and investigated the differences between them. Three of them, the BAR, bvAR, and BARCH models, have a conditionally linear mean, whereas the max-BAR and logit-BARCH models are nonlinear. We found that the linear BAR and bvAR models behave quite differently from the BARCH model, and none of these two models can mimic the properties of the time series generated by either of the other models. In contrast, the BARCH model proved to be sufficiently flexible to imitate the ACF structure of its competitors. Thus, the consequences of erroneously fitting a BARCH model to the data are less severe than in the case of the two other linear models. Regarding the non-CLAR models, the max-BAR model with its highly nonlinear data-generating mechanism is markedly different from all other models. In contrast, the logit-BARCH model, although being nonlinear in a formal sense, is sometimes hard to distinguish from the linear BARCH model. Generally, for the model choice, the AIC is a more reliable advisor for higher-order AR-type DGPs, whereas the BIC is better suited to avoid the overfitting of a first-order AR-type DGP (also see the Discussion in Dziak et al. 2019). However, since the actual model order is not known in advance, it is crucial to not blindly trust any of these criteria. Instead, the model choice should always be accompanied by a careful adequacy check, for example, by examining the moment properties or Pearson residuals considered in this paper.

Acknowledgements Open Access funding provided by Projekt DEAL. The authors thank the associate editor and the two referees for their careful reading of this paper and for providing valuable comments that have helped strengthen this manuscript. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07045707).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19(6):717–723

- Al-Osh MA, Alzaid AA (1987) First-order integer-valued autoregressive (INAR(1)) process. *J Time Ser Anal* 8(3):261–275
- Burnham KP, Anderson DR (2002) *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edn. Springer, New York
- Cavanaugh JE, Neath AA (2019) The Akaike information criterion: background, derivation, properties, application, interpretation, and refinements. *WIREs Comput Stat* 11(3):e1460
- Chen H, Li Q, Zhu F (2019) Two classes of dynamic binomial integer-valued ARCH models. *Braz J Probab Stat*, forthcoming
- Claeskens G, Hjort NL (2008) *Model selection and model averaging*. Cambridge University Press, Cambridge
- Cui Y, Lund R (2009) A new look at time series of counts. *Biometrika* 96(4):781–792
- Davis RA, Resnick SI (1989) Basic properties and prediction of max-ARMA processes. *Adv Appl Probab* 21(4):781–803
- Diop ML, Kenge W (2020) Consistent model selection procedure for general integer-valued time series. [arXiv:2002.08789](https://arxiv.org/abs/2002.08789)
- Dziak JJ, Coffman DL, Lanza ST, Li R, Jermiin LS (2019) Sensitivity and specificity of information criteria. *Brief Bioinform*, bbz016
- Gouveia S, Möller TA, Weiß CH, Scotto MG (2018) A full ARMA model for counts with bounded support and its application to rainy-days time series. *Stoch Environ Res Risk Assess* 32(9):2495–2514
- Grunwald G, Hyndman RJ, Tedesco L, Tweedie RL (2000) Non-Gaussian conditional linear AR(1) models. *Aust N Z J Stat* 42(4):479–495
- Jung RC, McCabe BPM, Tremayne AR (2016) Model validation and diagnostics. In: Davis RA et al (eds) *Handbook of discrete-valued time series*. CRC Press, Boca Raton, pp 189–218
- Katz RW (1981) On some criteria for estimating the order of a Markov chain. *Technometrics* 23(3):243–249
- Kunihama T, Omori Y, Zhang Z (2012) Efficient estimation and particle filter for max-stable processes. *J Time Ser Anal* 33(1):61–80
- McKenzie E (1985) Some simple models for discrete variate time series. *Water Res Bull* 21(4):645–650
- Neath AA, Cavanaugh JE (2012) The Bayesian information criterion: background, derivation, and applications. *WIREs Comput Stat* 4(2):199–203
- Psaradakis Z, Sola M, Spagnoloy F (2009) Selecting nonlinear time series models using information criteria. *J Time Ser Anal* 30(4):369–394
- Rinke S, Sibbertsen P (2016) Information criteria for nonlinear time series models. *Stud Nonlinear Dyn Econ* 20(3):325–341
- Ristić MM, Weiß CH, Janjić AD (2016) A binomial integer-valued ARCH model. *Int J Biostat* 12(2):20150051
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6(2):461–464
- Steutel FW, van Harn K (1979) Discrete analogues of self-decomposability and stability. *Ann Probab* 7(5):893–899
- Weiß CH (2009) A new class of autoregressive models for time series of binomial counts. *Commun Stat Theory Methods* 38(4):447–460
- Weiß CH (2018) *An introduction to discrete-valued time series*. Wiley, Chichester
- Weiß CH, Feld MH-JM (2020) On the performance of information criteria for model identification of count time series. *Stud Nonlinear Dyn Econom* 24(1):20180012
- Weiß CH, Feld MH-JM, Mamode Khan N, Sunecher Y (2019) INARMA modeling of count time series. *Stats* 2(2):284–320
- Weiß CH, Kim H-Y (2013) Parameter estimation for binomial AR(1) models with applications in finance and industry. *Stat Pap* 54(3):563–590
- Weiß CH, Pollett PK (2012) Chain binomial models and binomial autoregressive processes. *Biometrics* 68(3):815–824
- Weiß CH, Scotto MG, Möller TA, Gouveia S (2018) The max-BARMA models for counts with bounded support. *Stat Probab Lett* 143:28–36
- Zhang Z, Smith RL (2010) On the estimation and application of max-stable processes. *J Stat Plan Inference* 140(5):1135–1153