Generalized Binary Time Series Models

Carsten Jentsch [1] and Lena Reichmann [1,2,*]

[1] Faculty of Statistics, TU Dortmund University, D-44221 Dortmund, Germany; jentsch@statistik.tu-dortmund.de
[2] Mathematical Institute, University of Mannheim, D-68131 Mannheim, Germany
[*] Correspondence: reichmann@uni-mannheim.de or reichmann@statistik.tu-dortmund.de

check for updates

Abstract: The serial dependence of categorical data is commonly described using Markovian models. Such models are very flexible, but they can suffer from a huge number of parameters if the state space or the model order becomes large. To address the problem of a large number of model parameters, the class of (new) discrete autoregressive moving-average (NDARMA) models has been proposed as a parsimonious alternative to Markov models. However, NDARMA models do not allow any negative model parameters, which might be a severe drawback in practical applications. In particular, this model class cannot capture any negative serial correlation. For the special case of binary data, we propose an extension of the NDARMA model class that allows for negative model parameters, and, hence, autocorrelations leading to the considerably larger and more flexible model class of generalized binary ARMA (gbARMA) processes. We provide stationary conditions, give the stationary solution, and derive stochastic properties of gbARMA processes. For the purely autoregressive case, classical Yule–Walker equations hold that facilitate parameter estimation of gbAR models. Yule–Walker type equations are also derived for gbARMA processes.

Keywords: binary time series; autoregressive-moving average; autocovariance structure; Yule–Walker equations; stationarity

1. Introduction

Categorical time series data are collected in many fields of applications and the statistical research focusing on such data structures evolved considerably over the last years. As an important special case, binary time series that correspond to categorical data with two categories, occur in many different contexts. Often, binary time series are obtained from binarization of observed real-valued data. Such processes are considered, e.g., in Kedem and Fokianos (2002). In Figure 1, we show three real data examples of binary time series from different fields of research. For example, in Figure 1a, the eruption duration of the Old Faithful Geyser in the Yellowstone National Park is binarized using a threshold. It is coded with a value of one if an eruption lasts for longer than three minutes and zero if it is shorter. In economics, the two states of recessions and economic growth are of interest, as discussed, e.g., in Bellégo (2009). One example of a recession/no-recession time series is shown in Figure 1b, where for every quarter it is shown if Italy is in a recession, indicated by zero, or not, indicated by one. Recently, there is great interest in the air pollution in European cities, where an exceedance of the threshold of 50 μg/m^3 PM$_{10}$ (fine dust) causes a fine dust alarm. The resulting sequence of states of no exceedance corresponding to zero and exceedance corresponding to one is shown in Figure 1c. Further examples can be found, e.g., in geography, where sequences with the two states of dry and wet days are considered, e.g., in Buishand (1978). In biomedical studies, binary time series occur in the case, where the participants keep daily diaries of their disease.

For example, in clinical trials, as in Fitzmaurice et al. (1995), the binary self assessment of participants of their arthritis is collected, where poor is indicated by zero and good by one. In natural language processing, the occurrence of vowels as a sequence can be of interest, as considered in Weiß (2009b), where a text is binarized by detecting a consonant or no consonant/vowel as the two states. The binarization of a time series by a threshold, as, e.g., in the PM_{10} example, or by categorizing the time series into two states, as, e.g., in dry and wet days, indeed simplifies the real valued time series to a binary version. As mentioned in Kedem (1980), nevertheless, the transformation keeps the random mechanism from which the data are generated. For the example of PM_{10} data, it might often be of more interest, whether a certain threshold is crossed (or not) instead of the actual amount. In general, the rhythm within the binarized time series contains a great amount of information of the original data.
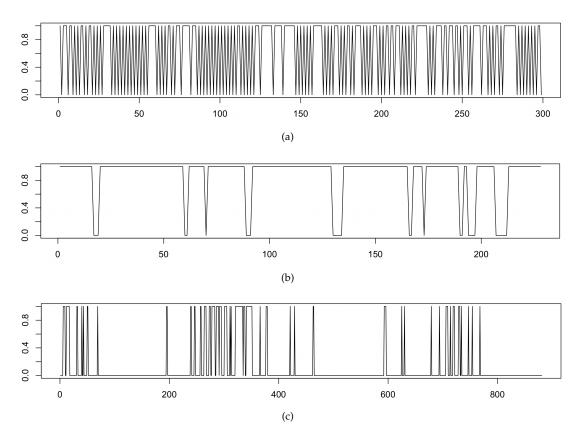

(a)

(b)

(c)

Figure 1. Three real data examples of binary time series: (**a**) binarized eruption duration of the Old Faithful Geyser over 299 eruptions; (**b**) quarterly detected binarized recession/no-recession time series of Italy from Quarter 1 in 1960 to Quarter 1 in 2017 (229 time points); and (**c**) binarized fine dust (PM_{10}) data from Stuttgart, Germany recorded daily from 3 March 2016 to 31 July 2018 over 881 days.

As discussed in Kedem (1980), binary Markov chains are typically used for modelling the dependence structure due to their great flexibility. However, the number of parameters to estimate from the data grows exponentially with the order of the Markov model leading to over-parametrization (see, e.g., McKenzie (2003)).

To avoid the estimation of a large number of parameters, Jacobs and Lewis (1983) proposed the class of (new) discrete autoregressive moving-average (NDARMA) models for categorical time series. More precisely, for processes with discrete and finite state space, a parsimonious model is suggested. The idea is to choose the current value for X_t randomly either from the past values of the time series $X_{t-1}, \ldots, X_{t-p}$ or

from one of the innovations $e_t, e_{t-1}, \ldots, e_{t-q}$ with certain probabilities, respectively. This random selection mechanism is described by independent and identically distributed (i.i.d) random vectors $(P_t, t \in \mathbb{Z})$ with

$$P_t := \left[a_t^{(1)}, \ldots, a_t^{(p)}, b_t^{(0)}, \ldots, b_t^{(q)} \right] \sim Mult\,(1; P),$$

where $Mult\,(1; P)$ denotes the multinomial distribution with parameter 1 and probability vector $P := \left[\alpha^{(1)}, \ldots, \alpha^{(p)}, \beta^{(0)}, \ldots, \beta^{(q)} \right]$ with $\alpha^{(1)}, \ldots, \alpha^{(p)} \in [0, 1)$, $\beta^{(0)} \in (0, 1]$ and $\beta^{(1)}, \ldots, \beta^{(q)} \in [0, 1)$ such that $\sum_{i=1}^{p} \alpha^{(i)} + \sum_{j=0}^{q} \beta^{(j)} = 1$. Then, the NDARMA(p,q) model equation is given by

$$X_t = \sum_{i=1}^{p} a_t^{(i)} X_{t-i} + \sum_{j=0}^{q} b_t^{(j)} e_{t-j}, \quad t \in \mathbb{Z}, \tag{1}$$

where $(e_t)_{t \in \mathbb{Z}}$ is an i.i.d. process taking values in a discrete and finite state space $\mathcal{S}$. Since for each time point t only one entry in the random vector P_t is realized to be one while all others become zero, the value of X_t takes either one of the values of X_s for $s \in \{t-1, \ldots, t-p\}$ or one of the error terms e_s for $s \in \{t, \ldots, t-q\}$. This sampling mechanism assures that the time series takes values in the state space $\mathcal{S}$, such that, e.g., for a binary time series with $\mathcal{S} = \{0, 1\}$, the process stays binary. In contrast to the real-valued ARMA model, the lagged time series values and errors are not weighted according to the model coefficients and summed-up since only one of them is actually multiplied with one and all the others with zero based on the realization of P_t.

The model parameters are the probabilities of the multinomial distribution, summarized in the parameter vector P, where all entries of P lie in the unit interval and sum-up to one. In comparison to Markov Chains, NDARMA models maintain the nice interpretable ARMA-type structure and have a parsimonious parameterization. Furthermore, NDARMA models fulfill certain Yule–Walker-type equations, as shown in Weiß and Göb (2008).

In Figure 2, one realization of an NDARMA(1,0) process, denoted by NDAR(1),

$$X_t = a_t X_{t-1} + b_t e_t, \quad [a_t, b_t] \sim Mult\,(1; \alpha, \beta), \quad \beta = 1 - \alpha \tag{2}$$

with binary state space is shown. NDAR(1) models are probably the simplest members of the NDARMA class, but Figure 2 nicely illustrates the limited flexibility of the whole NDARMA class. The sampling mechanism of choosing the predecessor with some probability α tends to generate long runs of the same value in particular when the parameter $\alpha \in (0, 1)$ is large. A switching from one state to the other, e.g., from $X_{t-1} = 0$ to $X_t = 1$, can only occur, e.g., if the error term e_t is selected (with probability $1 - \alpha$) and the error term takes the value $e_t = 1$. Hence, the NDARMA class does not allow *systematically* selecting the opposite value of X_{t-1} for X_t.
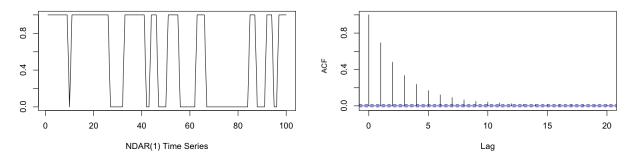

Figure 2. Realization of an NDAR(1) process (Equation (2)) with parameter vector $P = [0.7, 0.3]$ and error distribution $P\,(e_t = 1) = 0.5$ and corresponding autocorrelation function (ACF).

As for the NDARMA class all model parameters are restricted to be non-negative, which explains in particular why the NDARMA class can model exclusively non-negative autocorrelations in the data. For the example of a NDAR(1) process, the autocorrelation at lag one is equal to $\alpha \in [0,1)$, such that any alternating pattern that corresponds to negative model parameters as, e.g., observed in Figure 1a, cannot be captured. For a more detailed discussion of the properties of NDARMA models, we refer also to Jacobs and Lewis (1983) or Weiß (2009a). To increase its flexibility, Gouveia et al. (2018) proposed an extension of the NDARMA model class by using a variation function, but the resulting models do also not allow for negative model parameters and, hence, no negative dependence structure. Hence, whenever negative dependence structure is present in binary time series data, the NDARMA model is not suitable. In fact, in all three data examples of Figure 1, a straightforward estimation based on Yule–Walker estimators leads to at least some negative coefficients, such that NDAR models turn out to be not applicable.

To address this lacking flexibility of the NDARMA model class, we propose a simple and straightforward extension of the original idea of Jacobs and Lewis (1983) that allows also negative serial dependence. The resulting generalized binary ARMA (gbARMA) model class maintains the nicely interpretable model structure. Furthermore, no additional parameters are required to handle the negative dependence, preserving the parsimonious parameterization as well. In Figure 3, a realization of a gbARMA(1,0) process, denoted as gbAR(1), is shown. As a straightforward extension of an NDAR(1) model in Figure 2, gbAR(1) models allow for negative serial dependence. In fact, the range of the autocorrelation at lag one is extended from $[0,1)$ for NDAR(1) to $(-1,1)$ for gbAR(1) models.
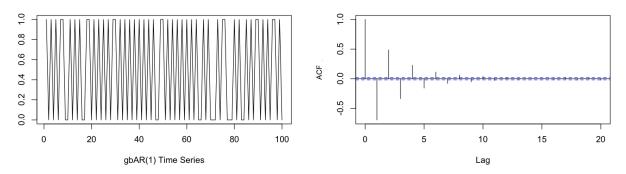

Figure 3. Realization of a gbAR(1) process (Equation (4)) with parameter vector $P = [-0.7, 0.3]$ and error distribution $P(e_t = 1) = 0.5$ and the corresponding autocorrelation function (ACF).

To allow for negative autocorrelation up to some limited extend, Kanter (1975) proposed the binary ARMA model class, where he applied the modulo 2 operator in an ARMA-type model equation. Using the modulo operation assures to stay in the binary state space, but the nice interpretability of the dependence structure in the model is lost since the past values of the time series are summed up prior to the modulo operation, see also McKenzie (1981). We follow a different path in this paper and propose a much simpler operation that enables modeling a systematic change of the state from one time point to the other.

The idea of allowing for negative serial dependence resulting in the gbARMA class is as follows: a negative model parameter $\alpha \in (-1,0)$ (and hence a negative autocorrelation $\alpha \in (-1,0)$) in binary time series data corresponds to the time series *systematically* changing from one state to the other over time. Hence, the natural idea to incorporate negative serial dependence in the *binary* NDAR(1) Model (Equation (2)) is to replace X_{t-1} by $1 - X_{t-1}$ as

$$(1 - X_{t-1}) = \begin{cases} 1 & \text{for } X_{t-1} = 0 \\ 0 & \text{for } X_{t-1} = 1 \end{cases} \tag{3}$$

holds. This leads to the model equation

$$X_t = a_t (1 - X_{t-1}) + b_t e_t, \quad [a_t, b_t] \sim Mult(1; |\alpha|, \beta).$$

This process has negative autocorrelation α at lag one. Note that, in comparison to Equation (2), as $\alpha \in (-1, 0)$ here, we have to use its absolute value $|\alpha|$ as the probability to select the $1 - X_{t-1}$. Altogether, for $\alpha \in (-1, 1)$, we can define the generalized binary AR(1) (gbAR(1)) process by the model equation

$$X_t = \begin{cases} a_t X_{t-1} + b_t e_t, & [a_t, b_t] \sim Mult(1; \alpha, \beta), & \alpha \in [0, 1) \\ a_t (1 - X_{t-1}) + b_t e_t, & [a_t, b_t] \sim Mult(1; |\alpha|, \beta), & \alpha \in (-1, 0) \end{cases}. \tag{4}$$

Note that Equation (4) extends the parameter space from $\alpha \in [0, 1)$ for NDAR(1) models to $\alpha \in (-1, 1)$ for gbAR(1) models. Further, note that, for identification of the model, we have to assume $\beta^{(0)} = \beta \in (0, 1]$. Using indicator variables, Equation (4) can be compactly written as

$$X_t = a_t \left(\mathbb{1}_{\{\alpha \geq 0\}} X_{t-1} + \mathbb{1}_{\{\alpha < 0\}} (1 - X_{t-1}) \right) + b_t e_t \tag{5}$$

$$= \left[a_t^{(+)} X_{t-1} + a_t^{(-)} \right] + b_t e_t \tag{6}$$

with $[a_t, b_t] \sim Mult(1; |\alpha|, \beta)$, $\beta = 1 - |\alpha|$, $a_t^{(+)} := a_t \left(\mathbb{1}_{\{\alpha \geq 0\}} - \mathbb{1}_{\{\alpha < 0\}} \right)$ and $a_t^{(-)} := a_t \mathbb{1}_{\{\alpha < 0\}}$.

In Figure 3, a realization of a gbAR(1) process with negative parameter $\alpha = -0.7$ is shown, where the time series tends to take systematically the opposite state of the predecessor. The corresponding autocorrelation plot reflects the negative serial dependence leading to an alternating pattern. Runs of the same state can only occur, when the error term e_t is selected (with probability $1 - |\alpha|$) and the error term e_t takes the same value as X_{t-1}, that is, $e_t = X_{t-1}$. The empirical autocorrelations for the Old Faithful Geyser data can be found in Figure 4a, where the pronounced alternating behavior clearly indicates negative linear dependence to be present in the data.

The idea of allowing for a negative model coefficient by replacing X_{t-1} by $1 - X_{t-1}$ in gbAR(1) processes (Equation (5)) can be also employed for each parameter in pth order gbAR processes, where each X_{t-i}, $i = 1, \ldots, p$ may be replaced by $1 - X_{t-i}$ in the model equation.

The paper is organized as follows. In Section 2, generalized binary AR processes of order $p \in \mathbb{N}$ are defined, where we also give stationarity conditions and state the stationary solution. Further, stochastic properties are derived that include formulas for the transition probabilities, the marginal distribution, and Yule–Walker equations. As a real data example, we illustrate the applicability of our model class to the geyser eruption data in Section 1. In Section 3, we present several simulation experiments. First, in Section 3.1, for the example of a gbAR(2) model, we illustrate the generality of the resulting gbAR model class in comparison to natural competitors including AR, NDAR, and Markov models of order two, respectively. In Section 3.2, we examine the estimation performance of Yule–Walker estimators in the gbAR models in Section 3.2.1. In Section 3.2.2, we investigate the benefit of using the parsimonious gbAR models in comparison to Markov models and their robustness in cases where the model is mis-specified. By adding a moving-average part to gbAR models in Section 4, ARMA-type extensions of gbAR models leading to gbARMA processes are discussed. We conclude in Section 5. All proofs are deferred to Appendix A.

2. The Generalized Binary Autoregressive (gbAR) Model Class

We define now generalized binary AR(p) (gbAR(p)) models for binary data based on the notation of NDAR(p) models by adopting the idea of replacing X_{t-1} by $1 - X_{t-1}$ for a negative parameter α as in Equations (5) and (6) separately for all or some of the lagged values $X_{t-1}, \ldots, X_{t-p}$. To be most flexible,

each parameter $\alpha^{(i)}$ corresponding to the lagged value X_{t-i}, $i = 1, \ldots, p$ is allowed to be either positive or negative, that is, $\alpha^{(i)} \in (-1, 1)$, respectively.

2.1. gbAR Models

The parameter vector $P := \left[\alpha^{(1)}, \ldots, \alpha^{(p)}, \beta^{(0)} \right]$ contains the probabilities of the multinomial distribution that controls the selection mechanism of NDAR models. As we allow now for $\alpha^{(i)} \in (-1, 1)$, $i = 1, \ldots, p$, i.e., the parameters can be negative, P has to be modified to serve again as a parameter vector of probabilities. This is achieved by taking entry-wise absolute values and we define

$$P_{|\cdot|} := \left[|\alpha^{(1)}|, \ldots, |\alpha^{(p)}|, \beta^{(0)} \right], \tag{7}$$

where $\beta^{(0)} \in (0, 1]$ such that $\sum_{i=1}^{p} |\alpha^{(i)}| + \beta^{(0)} = 1$. This enables us to give the definition of the generalized binary AR model of arbitrary order $p \in \mathbb{N}$.

Definition 1 (Generalized binary AR processes). *Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary process taking values in $\{0, 1\}$. Let $(e_t)_{t \in \mathbb{Z}}$ be a binary error process such that e_t is independent of $(X_s)_{s < t}$ with mean $\mu_e = E(e_t) = P(e_t = 1)$ and variance $\sigma_e^2 = Var(e_t) = P(e_t = 1)(1 - P(e_t = 1)) > 0$. Let $P := \left[\alpha^{(1)}, \ldots, \alpha^{(p)}, \beta^{(0)} \right]$ be the parameter vector with $P_{|\cdot|}$ as in Equation (7) such that $P_{|\cdot|} \mathbb{1}_{p+1} = 1$ with $\mathbb{1}_{p+1}$ the one vector of length $p + 1$. Further, let*

$$P_t := \left(a_t^{(1)}, \ldots, a_t^{(p)}, b_t^{(0)} \right) \sim Mult \left(1; P_{|\cdot|} \right), \quad t \in \mathbb{Z},$$

be iid random vectors, which are independent of $(e_t)_{t \in \mathbb{Z}}$ and $(X_t)_{s < t}$. Then, the process $(X_t)_{t \in \mathbb{Z}}$ is said to be a generalized binary AR process of order p (gbAR(p)), if it follows the recursion

$$X_t = \sum_{i=1}^{p} \left[a_t^{(+,i)} X_{t-i} + a_t^{(-,i)} \right] + b_t^{(0)} e_t \tag{8}$$

with $a_t^{(+,i)} := a_t^{(i)} \left(\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}} \right)$ and $a_t^{(-,i)} := a_t^{(i)} \mathbb{1}_{\{\alpha^{(i)} < 0\}}$ for $i = 1, \ldots, p$.

By rewriting the random variables $a_t^{(\cdot,i)}$, $\cdot \in \{-, +\}$ in the defining model (Equation (8)), the model can be represented in the spirit of Equation (5). However, the benefit of the representation in Equation (8) is that only one random variable is multiplied with the lagged value X_{t-i}, whereas $a_t^{(-,i)}$ is an additional random variable that accounts for the switching that leads to negative model coefficients.

2.2. Stochastic Properties of gbAR Models

Before calculating moments of the binary time series process $(X_t)_{t \in \mathbb{Z}}$ itself, we first consider the expectation of the random variables related to the multinomial selection mechanism. Noting that $E(a_t^{(i)}) = |\alpha^{(i)}|$, we have

$$E \left(a_t^{(-,i)} \right) = |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} < 0\}} =: \alpha^{(-,i)},$$
$$E \left(a_t^{(+,i)} \right) = \alpha^{(i)}.$$

This enables us to compute the stationary mean $\mu_X = E(X_t)$ of the process, directly leading to

$$\mu_X = \frac{\sum_{i=1}^p \alpha^{(-,i)} + \beta^{(0)} \mu_e}{1 - \sum_{i=1}^p \alpha^{(i)}}. \tag{9}$$

If all parameters $\alpha^{(1)}, \ldots, \alpha^{(p)}$ are non-negative, the above formula becomes $\mu_X = \mu_e$ due to $\sum_{i=1}^p \alpha^{(-,i)} = 0$ and $1 - \sum_{i=1}^p \alpha^{(i)} = \beta^{(0)}$, leading then to the well-known formula for the mean of NDAR(p) models. Otherwise, we have $\mu_X \neq \mu_e$ for gbAR(p) models in contrast to NDAR(p) models (see, e.g., Weiß (2009a)).

For the familiar stationary condition imposed on the model parameters $\alpha^{(1)}, \ldots, \alpha^{(p)}$ that all roots of the characteristic polynomial lie outside the unit circle, i.e., if

$$\left(1 - \alpha^{(1)}z - \ldots - \alpha^{(p)}z^p\right) \neq 0 \;\; \forall z \leq 1 \tag{10}$$

holds, the stationary solution of the gbAR(p) model can be derived. Note that the condition in (10) is equivalent to $\sum_{i=1}^p |\alpha^{(i)}| < 1$, such that the error has to be selected with strictly positive probability $\beta^{(0)} > 0$ by the multinomial distribution. If the stationarity condition in Equation (10) holds, a moving-average representation of the gbAR(p) process can be derived.

For constructing the stationary solution of the gbAR time series, we follow the common approach based on a multivariate representation of the model, as in (Lütkepohl 2005, Chap. 11.3.2). Precisely, the gbAR(p) model can be written as a p-dimensional gbVAR(1) process $(Y_t, t \in \mathbb{Z})$ with the following matrices and vectors, such that the first entry of $(Y_t, t \in \mathbb{Z})$ is equal to the gbAR(p) process. We define

$$Y_t := \begin{pmatrix} X_t \\ \vdots \\ X_{t-p+1} \end{pmatrix} \;\; (p \times 1) \quad \text{and} \quad U_t := \begin{pmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \;\; (p \times 1).$$

To obtain a vector autoregressive representation for Y_t, we have to define several matrices that contain the random variables of the multinomial distribution. Precisely, for $\cdot = \{-,+\}$, let

$$\widetilde{A}_t^{(\cdot)} := \begin{pmatrix} a_t^{(\cdot,1)} & \ldots & a_t^{(\cdot,p-1)} & a_t^{(\cdot,p)} \\ 1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & \ldots & 1 & 0 \end{pmatrix} \quad \text{and} \quad \widetilde{B}_t^{(1)} := \begin{pmatrix} b_t^{(0)} & 0 & \ldots & 0 \\ 0_{p-1\times 1} & 0_{p-1\times 1} & \ldots & 0_{p-1\times 1} \end{pmatrix}$$

be $p \times p$ matrices, where $0_{r\times s}$ denotes the $(r \times s)$-dimensional zero matrix. Based on the notation introduced above, gbVAR(p) processes can be represented as a vector-valued gbAR model of first order (gbVAR(1)) as follows

$$Y_t = \widetilde{A}_t^{(+)} Y_{t-1} + \widetilde{A}_t^{(-)} \mathbb{1}_p + \widetilde{B}_t^{(1)} U_t, \tag{11}$$

where $\mathbb{1}_p$ is the one vector of length p. The above notation enables us to state a moving-average representation of gbAR(p) processes as follows.

Theorem 1 (Moving-average representation of gbAR processes)**.** *Let* $(X_t)_{t\in\mathbb{Z}}$ *be a stationary gbAR(p) process, that is* $(X_t)_{t\in\mathbb{Z}}$ *fulfills Equation* (10)*. Then, we have*

(i) *For* $p = 1$*, the gbAR(1) model has a gbMA(∞)-type representation (in L_2-sense), that is,*

$$X_t = \sum_{i=0}^{\infty} \zeta_i a_{t-i}^{(-)} + \sum_{i=0}^{\infty} \zeta_i b_{t-i}^{(0)} e_{t-i}, \quad t \in \mathbb{Z}, \tag{12}$$

where $\zeta_0 := I_K$ *and* $\zeta_i := \prod_{j=0}^{i-1} a_{t-j}^{(+)}$ *since* $\lim_{k\to\infty} \prod_{i=0}^{k-1} a_{t-i}^{(+)} = 0$ *in* L_2.

(ii) *For* $p \in \mathbb{N}$*, the gbAR(p) model has a gbMA(∞)-type representation (in L_1-sense), that is,*

$$X_t = e_1^T \left(\sum_{i=0}^{\infty} \prod_{j=0}^{i-1} \widetilde{A}_{t-j}^{(+)} \widetilde{A}_{t-i}^{(-)} \mathbb{1}_p + \sum_{i=0}^{\infty} \prod_{j=0}^{i-1} \widetilde{A}_{t-j}^{(+)} \widetilde{B}_{t-i}^{(1)} U_{t-i} \right), \quad t \in \mathbb{Z}, \tag{13}$$

since $\lim_{k\to\infty} \prod_{i=0}^{k} \widetilde{A}_{t-i}^{(+)} = 0_{p\times p}$ *in* L_1*. Here,* e_1 *is the first unit vector and* $\mathbb{1}_p$ *is the one vector of length p. The notation used here is obtained from that used in Section* 4.2 *for the special case of* $q = 0$.

Hence, the process can be represented as an infinite weighted sum of the error terms. However, in comparison to classical AR or NDAR processes, an additional term appears that takes control of potential negative parameters and, consequently, allows for negative dependence to be modeled. This term vanishes if all parameters $\alpha_1, \ldots, \alpha_p$ are positive.

The second-order dependence structure of gbAR processes coincides with that of AR or NDAR processes in the sense that the same Yule–Walker equations for $h \neq 0$ hold. However, note again that the parameter space for gbAR models is considerably larger than for NDAR models allowing for negative parameters leading to more flexibility. The Yule–Walker equations link the model parameters to the autocovariances of the process. Hence, they can be used for estimating the model parameters by the same well-known Yule–Walker estimators. A link between the autocovariances, the model coefficients, and the mean and variance of the error terms is established by the Yule–Walker equation for $h = 0$, respectively.

Theorem 2 (Yule Walker Equations)**.** *Let* $(X_t)_{t\in\mathbb{Z}}$ *be a stationary gbAR(p) process.*

(i) *For all* $h \in \mathbb{N}$*, we have*

$$\gamma(h) = \sum_{i=1}^{p} \alpha^{(i)} \gamma(|h-i|). \tag{14}$$

(ii) *For* $h = 0$*, we have*

$$\gamma(0) = \sigma_e^2 + \frac{(1-2\mu_X)\sum_{i=1}^{p} \alpha^{(-,i)} + \left(\sum_{i=1}^{p} |\alpha^{(i)}| - 1\right)\mu_X^2 + \beta^{(0)}\mu_e^2}{\left(1 - \sum_{i=1}^{p} |\alpha^{(i)}|\right)}. \tag{15}$$

The next Lemma states some basic properties of the marginal distribution of gbAR processes and their transition probabilities. Since the time series has a binary state space, these conditional probabilities allow quantifying the probability to reach a certain state from the past values. In their derivation, the multinomial selection mechanism plays a crucial role and, in the stated formulas, the Kronecker delta δ_{ij} indicates if a past value has actually impact on the outcome of the time series or not.

Lemma 1 (Marginal, joint and transition probabilities of gbAR processes)**.** *Let* $(X_t)_{t \in \mathbb{Z}}$ *be a stationary gbAR(p) model and set* $p_i := P(e_t = i)$. *Then, the following properties hold:*

(i) $P(X_t = i_0 | X_{t-1} = i_1, \ldots, e_t = j_0)$
$= \sum_{l=1}^{p} |\alpha^{(l)}| \left[\mathbb{1}_{\{\alpha^{(l)} \geq 0\}} \delta_{i_0 i_l} + \mathbb{1}_{\{\alpha^{(l)} < 0\}} \delta_{i_0 (1-i_l)} \right] + \beta^{(0)} \delta_{i_0 j_0}$

(ii) $P(X_t = i_0 | X_{t-1} = i_1, \ldots)$
$= \sum_{l=1}^{p} |\alpha^{(l)}| \left[\mathbb{1}_{\{\alpha^{(l)} \geq 0\}} \delta_{i_0 i_l} + \mathbb{1}_{\{\alpha^{(l)} < 0\}} \delta_{i_0 (1-i_l)} \right] + \beta^{(0)} p_{i_0}$

(iii) $P(X_t = j) = \frac{\beta^{(0)}}{\left(1 - \sum_{i=1}^{p} \alpha^{(i)}\right)} p_j + \frac{\sum_{i=1}^{p} |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} < 0\}}}{\left(1 - \left[\sum_{i=1}^{p} \alpha^{(i)}\right]\right)}$

(iv) $P(X_t = i_0, e_t = j_0) = \beta^{(0)} \delta_{i_0 j_0} + \sum_{i=1}^{p} |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} < 0\}} \left(1 - p_{j_0} + \frac{p_{j_0}}{1 - \sum_{i=1}^{p} |\alpha^{(i)}| \left[\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}} \right]} \right)$

Comparing the results in Lemma 1 with (Weiß 2009a, Lemma 11.2.1.3) established for NDARMA(p,q) processes, the main difference is in Part (iii). The marginal distribution of the NDARMA process is equal to the marginal distribution of the error term process, but this does not hold for gbAR processes. Instead, the marginal distribution of gbAR processes depends on an additional term that results from the absolute values of the negative parameters.

In the following example, let us conclude this section with a more detailed look at the gbAR(1) model and a real data example.

Example 1 (gbAR(1) process)**.** *Let* $(X_t, t \in \mathbb{Z})$ *be a gbAR(1) process with parameter vector* $P := \left[\alpha^{(1)}, \beta^{(0)} \right]$, $\alpha^{(1)} \in (-1, 1)$ *and* $\beta^{(0)} = 1 - |\alpha^{(1)}|$. *The iid error term process* $(e_t, t \in \mathbb{Z})$ *follows the distribution* $P(e_t = 1) = p_1 \in (0, 1)$ *such that* $\mu_e = p_1$ *and* $\sigma_e^2 = p_1(1 - p_1) > 0$. *Then, the model equation equals*

$$X_t = a_t^{(1)} \left[\mathbb{1}_{\{\alpha^{(1)} \geq 0\}} X_{t-1} + \mathbb{1}_{\{\alpha^{(1)} < 0\}} (1 - X_{t-1}) \right] + b_t^{(0)} e_t,$$
$$\left[a_t^{(1)}, b_t^{(0)} \right] \sim Mult \left(1; \left[|\alpha^{(1)}|, \beta^{(0)} \right] \right).$$

At each time point t, if $\alpha^{(1)} \geq 0$, *either the predecessor* X_{t-1} *with probability* $\alpha^{(1)}$ *or the error term* e_t *with probability* $\beta^{(0)}$ *is selected by a multinomial distributed random variable to determine* X_t. *In the case of* $\alpha^{(1)} < 0$, *either* $1 - X_{t-1}$ *with probability* $|\alpha^{(1)}|$ *or the error term* e_t *with probability* $\beta^{(0)}$ *is selected. That is, as for each t, either* a_t *or* b_t *is equal to one and the other is zero, it holds*

$$X_t = \begin{cases} X_{t-1} & \text{if } a_t^{(1)} = 1 \text{ for } \alpha^{(1)} \geq 0 \\ 1 - X_{t-1} & \text{if } a_t^{(1)} = 1 \text{ for } \alpha^{(1)} < 0 \\ e_t & \text{if } b_t^{(0)} = 1 \end{cases}$$

For positive values of $\alpha^{(1)}$, *the gbAR(1) model coincides with the NDAR(1) model. A corresponding realization is shown in Figure 2, where for large values of* $\alpha^{(1)}$ *mainly the predecessor* X_{t-1} *is chosen and long runs of the same value occur. Figure 3 shows one realization of a gbAR(1) process with negative value of* $\alpha^{(1)}$. *The time series switches its states from zero to one and vice versa at most time points.*

The transition probability to move from state i_1 *at time* $t - 1$ *to state* i_0 *at time t is given by*

$$P(X_t = i_0 | X_{t-1} = i_1) = |\alpha^{(1)}| \left[\mathbb{1}_{\{\alpha^{(1)} \geq 0\}} \delta_{i_0 i_1} + \mathbb{1}_{\{\alpha^{(1)} < 0\}} \delta_{i_0 (1-i_1)} \right] + \left(1 - |\alpha^{(1)}| \right) p_{i_0}. \tag{16}$$

The probability of the process taking the value $i_0 = 1$ depends on two terms. First, the probability of choosing the error term is multiplied by the probability of the error term taking the same value as X_t, e.g., $P(e_t = i_0) = p_{i_0}$ with $i_0 = 1$. If the probability of choosing the predecessor is added, it depends on its value and the sign of α. If, for example $\alpha < 0$, then the probability of choosing X_{t-1} is just added if its value is the contrary of i_0, such that the Kronecker delta is equal to one. This leads to the representation of Equation (16) as

$$P\left(X_t = i_0 | X_{t-1} = i_1\right) = \begin{cases} |\alpha^{(1)}|\delta_{i_0 i_1} + \left(1 - |\alpha^{(1)}|\right) p_{i_0} & \text{if } \alpha \geq 0 \\ |\alpha^{(1)}|\delta_{i_0(1-i_1)} + \left(1 - |\alpha^{(1)}|\right) p_{i_0} & \text{if } \alpha < 0 \end{cases}.$$

Example 2 (Eruption duration of the Old Faithful Geyser). *The binarized eruption duration of the Old Faithful Geyser is illustrated in Figure 1a. Its empirical autocorrelation, as shown in Figure 4a, clearly indicates that there is negative serial dependence present in the data such that a gbAR(p) process appears to be appropriate. The order selection using the AIC criterion leads to a model of order $p = 2$ with $AIC = 159.83$. This selection is confirmed by an inspection of the partial autocorrelation in Figure 4b. Parameter estimation is based on the Yule–Walker Equation (14) leading to the estimated parameter vector $\widehat{P} = [-0.3949, 0.2659, 0.3393]$ and the fitted model*

$$X_t = \sum_{i=1}^{2} \left[a_t^{(+,i)} X_{t-i} + a_t^{(-,i)} \right] + b_t^{(0)} e_t,$$

$$P_t \sim Mult\left(1; [| - 0.3949|, 0.2659, 0.3393]\right).$$

The sample mean of the binary time series is equal to $\widehat{\mu}_X = 0.6488$ since long eruptions of the geyser arise more often than short eruption duration. The first parameter $\widehat{\alpha}_1$ is indeed estimated to be negative and the second one $\widehat{\alpha}_2$ to be positive. From $\widehat{\beta}^{(0)} = 1 - \sum_{i=1}^{2} |\widehat{\alpha}^{(i)}|$, an error term is chosen with probability $\widehat{\beta}^{(0)} = 0.3393$. In Figure 1a, a change from zero to one or vice versa can be observed in many time steps, whereas the run of ones in the time series correspond in most cases to choosing an error term. The error term distribution is calculated by Equation (9) with $\widehat{\mu}_e = P(e_t = 1) = 0.9953$.
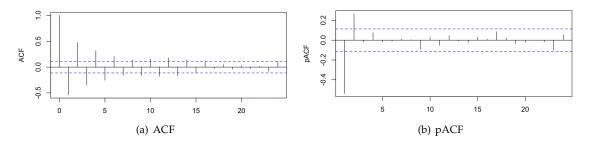

(a) ACF · (b) pACF

Figure 4. Autocorrelation (ACF) and partial autocorrelation (pACF) of the Old Faithful Geyser data

To measure the predictive power of the estimated model, we use ROC curves and the corresponding area under the curve (AUC). The ROC concept indicates a good predictive performance whenever the resulting curve is "far away" above the diagonal leading to an AUC larger than 0.5. Note that the diagonal corresponds to the case of independent observations, where no prediction based on past values is meaningful. For the one step ahead prediction, the transition probability of Lemma 1 (ii) is used by plugging in the estimated probabilities.

Comparing the predictor to the realized values in the sample leads to the ROC curve shown in Figure 5, where the corresponding AUC becomes 0.8317. Hence, as the ROC curve is "far away" above the diagonal and the AUC is larger than 0.5, the prediction performance of the gbAR model turns out to be considerably better than that of a model that relies on independent observations. By allowing for negative model parameters, gbAR models appear to be suitable for this real data example that shows negative serial dependence.
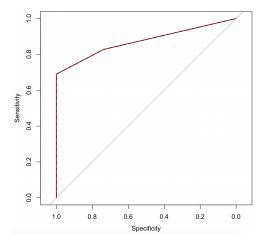
Figure 5. ROC curves based on Yule–Walker estimation (black) and MLE (red) of a gbAR(2) model fitted to the binarized Old Faithful Geyser eruption duration leading to an $AUC = 0.8317$ in both cases.

For further improvement, the Yule–Walker estimates might serve as starting values for a maximum likelihood estimation (MLE) based on the conditional log-likelihood function

$$\ell\left(\theta|x_p,\ldots,x_1\right) = \sum_{t=p+1}^{T} \log p\left(x_t|x_{t-1},\ldots,x_{t-p}\right),\tag{17}$$

where $p\left(x_t|x_{t-1},\ldots,x_{t-p}\right) := P\left(X_t = x_t|X_{t-1} = x_{t-1},\ldots,X_{t-p} = x_{t-p}\right)$ (see also (Weiß 2018, (B.6))). However, the resulting parameter estimates $\widehat{P}_{MLE} = [-0.3935, 0.2711, 0.3353]$ differ only slightly from the Yule–Walker estimates $\widehat{P} = [-0.3949, 0.2659, 0.3393]$, leading to virtually the same ROC and AUC.

To shed some light on the potential improvement of MLE in comparison to Yule–Walker estimation, we fit a gbAR(2) model to subsamples of length $T = 50, 100, 150, 200, 250$ of the binarized Old Faithful Geyser data. The parameter estimates for Yule–Walker estimates and MLE are shown in Table 1. The results differ only slightly and decrease with increasing subsample sizes.

Table 1. Comparison of (a) Yule–Walker and (b) MLE parameter estimates based on subsamples of length $T = 50, 100, 150, 200, 250$ of the binarized Old Faithful Geyser data.

(a) Yule–Walker			
T	$\widehat{\alpha}^{(1)}$	$\widehat{\alpha}^{(2)}$	$\widehat{\beta}^{(0)}$
50	-0.5819	0.1444	0.2738
100	-0.4610	0.2675	0.2715
150	-0.3748	0.3382	0.2871
200	-0.3738	0.3440	0.2822
250	-0.4048	0.2625	0.3328
299	-0.3949	0.2659	0.3393
(b) MLE			
T	$\widehat{\alpha}^{(1)}$	$\widehat{\alpha}^{(2)}$	$\widehat{\beta}^{(0)}$
50	-0.5556	0.1812	0.2632
100	-0.4546	0.2822	0.2632
150	-0.3658	0.3511	0.2830
200	-0.3706	0.3514	0.2780
250	-0.4004	0.2723	0.3259
299	-0.3935	0.2711	0.3353

3. Generality of the gbAR Model Class and Estimation Performance

In this section, we investigate the generality of the gbAR model class in comparison to obvious competitors in Section 3.1 and address the estimation performance in different setups and in comparison to parameter-intensive Markovian models in Section 3.2.

3.1. Illustration of the Generality of gbAR Models

By construction of the gbAR model and in contrast to NDAR models, negative parameters $\alpha^{(i)} \in (-1, 1), i = 1, \ldots, p$ are allowed such that negative autocorrelation is possible. Hence, the proposed gbAR model class clearly generalizes the NDAR model class. In this section, we aim to shed some light on the question how much more general the gbAR model actually is in comparison to other AR-type models such as AR, NDAR, and binAR models, as well as Markov models. For this purpose, we consider such models of order $p = 2$ and study their generality. That is, we compare the parameter ranges of these four model classes as well as the possible ranges of pairs of autocorrelation $(\rho(1), \rho(2))$. Precisely, we compare the flexibility of gbAR(2), NDAR(2), binAR(2), and AR(2) processes (even if they model continuous data) and second-order Markov chains. For all four autoregressive-type models, the autocorrelations depend on the model parameters as follows

$$\rho(1) = \frac{(1 - 2\mu)\alpha^{(1)}}{1 - (1 - 2\mu)\alpha^{(2)}},$$

$$\rho(2) = (1 - 2\mu)\frac{(1 - 2\mu)(\alpha^{(1)})^2}{1 - (1 - 2\mu)\alpha^{(2)}} + (1 - 2\mu)\alpha^{(2)}.$$

For the gbAR(2), AR(2), and NDAR(2) processes, it holds $\mu = 0$.

For a stationary AR(2) process, the range of possible coefficients is restricted to $\alpha^{(1)} \in (-2, 2)$ and $\alpha^{(2)} \in [-1, 1]$ such that $\alpha^{(1)} + \alpha^{(2)} < 1$ and $\alpha^{(2)} - \alpha^{(1)} < 1$. For a stationary NDAR(2) and binAR(2) process, the parameter range is restricted by $\alpha^{(1)}, \alpha^{(2)} \in [0, 1]$ with $\alpha^{(1)} + \alpha^{(2)} < 1$ and in the binAR(2)

process $P\left(e_t = 1\right) = \mu \in (0,1)$ (for further details, see Weiß (2009b)). For a gbAR(2) model, the restrictions read $|\alpha^{(1)}| + |\alpha^{(2)}| < 1$ with $\alpha^{(1)}, \alpha^{(2)} \in (-1,1)$.

The parameter ranges of AR, NDAR, binAR and gbAR models of order two, respectively, are illustrated and compared in Figure 6a and the corresponding range of pairs of autocorrelations $\left(\rho\left(1\right), \rho\left(2\right)\right)$ is shown in Figure 6b.
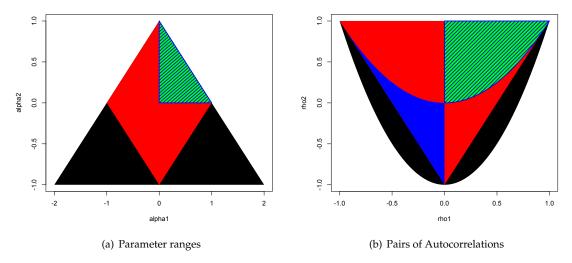

(a) Parameter ranges　　　　　　　　　　　　　　(b) Pairs of Autocorrelations

Figure 6. Comparison of (**a**) parameter ranges of $(\alpha^{(1)}, \alpha^{(2)})$ and (**b**) pairs of autocorrelations $\left(\rho\left(1\right), \rho\left(2\right)\right)$ for AR(2) (black), gbAR(2) (red), binAR(2) (blue), and NDAR(2) (green) models.

The parameter range as well as the range of autocorrelation pairs for the gbAR(2) model is considerably larger than those of an NDAR(2) model. The range of the classical AR model is again larger, but this is an unfair comparison as the AR model has been proposed for continuous data and is not suitable for binary data at all. In Figure 6b, the areas of AR(2), NDAR(2), and binAR(2) models are hyperboloid-shaped and, as shown in Jacobs and Lewis (1983), the autocorrelations of the NDAR(2) model take just positive values. In contrast to NDAR(2) processes, the binAR(2) process captures an additional area that corresponds to negative serial dependence. The range of autocorrelation pairs for the gbAR(2) is not hyperboloid-shaped, but forms a triangle. The range of this triangle comes actually close to the range of the AR(2) model, although the comparison with the AR(2) model is indeed unfair as the latter has been proposed for continuous data and the gbAR(2) for binary data. Compared to NDAR(2) processes, the extension allowing also for negative parameters leads to a much larger range of possible autocorrelation pairs than just the mirrored half parable. This is explained by the four times larger possible range for the model parameters of gbAR(2) processes in comparison to NDAR(2) processes, as shown in Figure 6a. In summary, by allowing for negative model parameters in gbAR models, we can get a considerably more flexible model class in comparison to NDAR models that is suitable to capture a wider range of dependence structures of binary time series data. In Figure 7, the possible range of autocorrelation pairs of gbAR(2) processes and Markov chains of order two are shown together. Recall that the gbAR(2) model is a parsimonious member of the class of Markov chains of order two and hence less flexible. Interestingly, with respect to pairs of autocorrelations of lags one and two, the possible range for the gbAR(2) model is only slightly smaller than that of a Markov chain of order two. Moreover, the largest range shown for the (continuous) AR(2) models in Figure 6b (in black) cannot be attained by Markov chains of order two. Hence, gbAR(2) models can to cover a large portion of the possible range of autocorrelation pairs of lags

one and two of second-order Markov chains. However, keep in mind that for a pth-order Markov chain, 2^p parameter have to be estimated. For example, $2^2 = 4$ parameters need to be specified for a second-order Markov chain, whereas gbAR(2) processes only require three parameters.
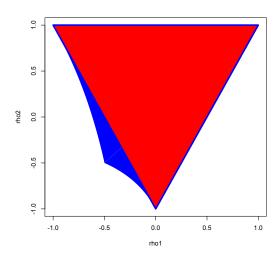

Figure 7. Autocorrelation pairs of the gbAR(2) process (red) compared to the pairs of autocorrelation of a second-order Markov chain (blue).

3.2. Simulations

In this simulation study, we addressed two things. First, as shown in Section 3.2.1, we investigated the estimation performance of Yule–Walker estimators for gbAR models of different orders and sample sizes. Second, as shown in Section 3.2.2, we studied the flexibility of the gbAR model class and compared the prediction performance to Markovian models in the case where the estimated model was correctly specified as a gbAR model and in the case where the underlying model was a Markovian model that does not belong to the class of gbAR models.

3.2.1. Estimation Performance

To study the estimation performance of Yule–Walker estimators in gbAR models, we considered three different specifications of gbAR(p) processes with $p = 1, 2, 3$ and sample sizes $T = 100, 200, 500, 1000$. Precisely, we considered the following gbAR data generating processes (DGPs):

(DGP1)　gbAR(1) with $\alpha^{(1)} = -0.85$, $mu_e = 0.3$, $\beta^{(0)} = 0.15$ and $\mu_X = 0.48378$.
(DGP2)　gbAR(2) with $\alpha^{(1)} = 0.42$, $\alpha^{(2)} = -0.38$, $mu_e = 0.3$, $\beta^{(0)} = 0.2$ and $\mu_X = 0.45833$.
(DGP3)　gbAR(3) with $\alpha^{(1)} = 0. - 0.294$, $\alpha^{(2)} = 0.382$, $\alpha^{(3)} = -0.2393$, $mu_e = 0.67$, $\beta^{(0)} = 0.0847$ and $\mu_X = 0.52140$.

The model parameters summarized in P were estimated based on Yule–Walker Equation (14) and the error term distribution using Equation (9). Note that, in all setups, we considered gbAR models that make use of the extended parameter space by including negative parameters $\alpha^{(i)}$ in the model. For each DGP, we simulated 1000 replications to calculate the mean squared error (MSE) to measure the estimation performance.

Table 2 summarizes the simulation results for all DGPs and all considered sample sizes. The estimation performance is generally good, as confirmed by rather small MSEs. It turns out that, as expected, in all considered setups, the estimation performance improves with increasing the sample size. It is interesting

to note that, relative to the estimation of the other quantities, the estimation of the mean of the error terms μ_e is generally less precise. This can be explained by the fact that the error terms e_t do enter the time series only in the case when it is actually selected which happens only with probabilities $\beta^{(0)}_{DGP1} = 0.15$, $\beta^{(0)}_{DGP2} = 0.2$ and $\beta^{(0)}_{DGP3} = 0.0874$ for the three DGPs, respectively. A comparison of the estimation performance of the gbAR models of different orders shows that the estimation performance declines with increasing order, which is of course plausible as the number of parameters gets larger, leading to more estimation uncertainty.

Table 2. Estimation performance for several Yule–Walker parameter estimates with respect to mean squared errors for three different DGPs over 1000 Monte Carlo replications.

	T	MSE of $\widehat{\alpha}^{(1)}$	MSE of $\widehat{\alpha}^{(2)}$	MSE of $\widehat{\alpha}^{(3)}$	MSE of $\widehat{\mu}_X$	MSE of $\widehat{\mu}_e$	MSE of $\widehat{\beta}^{(0)}$
DGP1	100	0.00271			0.00022	0.03140	0.00271
	200	0.00133			0.00011	0.01464	0.00133
	500	0.00051			0.00004	0.00543	0.00051
	1000	0.00025			0.00002	0.00262	0.00025
DGP2	100	0.00684	0.00588		0.00191	0.04877	0.01007
	200	0.00338	0.00314		0.00107	0.02689	0.00521
	500	0.00136	0.00124		0.00044	0.01169	0.00219
	1000	0.00068	0.00064		0.00020	0.00555	0.00108
DGP3	100	0.01079	0.01253	0.00900	0.00416	0.10187	0.02212
	200	0.00502	0.00525	0.00386	0.00200	0.08931	0.01044
	500	0.00194	0.00198	0.00143	0.00082	0.05859	0.00374
	1000	0.00097	0.00087	0.00065	0.00040	0.03913	0.000178

3.2.2. Robustness of gbAR Model Class

The class of gbAR(p) models form a parsimoniously parametrized subclass in the class of Markovian models of order p. To study the benefit of this newly proposed class of binary models, we wanted to compare the gbAR(p) model to Markov chains which are mostly used for binary data.

First, let us consider the case of an underlying gbAR model. Since gbAR(p) processes have a Markov chain representation, a comparison in terms of the transition probabilities becomes suitable. From Lemma 1 *(ii)*, the transition probabilities of gbAR models compute to

$$
P\left(X_t = i_0 | X_{t-1} = i_1, \ldots, X_{t-p} = i_p\right)
$$
$$
= \sum_{l=1}^{p} |\alpha^{(l)}| \left[\mathbb{1}_{\{\alpha^{(l)} \geq 0\}} \delta_{i_0 i_l} + \mathbb{1}_{\{\alpha^{(l)} < 0\}} \delta_{i_0(1-i_l)}\right] + \beta^{(0)} P\left(e_t = i_0\right). \tag{18}
$$

First, for $p = 1, 2, 3$, we simulated from the gbAR(p) models defined for DGP1-3 in Section 3.2.1 realizations of different sample sizes to estimate the transition probabilities of: (a) a pth-order Markov chain and (b) a gbAR(p) process. For gbAR models, the true transition probabilities are given by Equation (18) and can be estimated by replacing the model parameters by the corresponding estimators. Then, the MSE is calculated model-wise and over all transition probabilities. For all three DGPs with model orders $p = 1, 2, 3$, the simulation results are stated in Table 3. By "MSE gbAR(p)", we denote the mean squared error by evaluating the difference between the estimated transition probability and the truth over all possible transition probabilities from a gbAR(p) process. Equivalently, "MSE MC" denotes the corresponding difference between the estimated transition probability and the truth of a Markov model over all possible transition probabilities.

The MSEs are calculated over 1000 replications and show clearly that the estimated gbAR(p) transition probabilities have smaller MSEs for all sample sizes and orders in comparison to the MSEs of the Markov chain fits. This indicates that, in the case of an underlying gbAR(p) process, fitting the the more parsimonious model to the data leads to better estimation performance than fitting a Markov chain.

Next, we considered the situation where the underlying model is a Markov chain of order p that does *not* belong to the subclass of gbAR(p) models. In general, the $2^p \times 2$-dimensional transition probability matrix Q of a Markov chain of order p is defined by

$$Q = \left(p^{i_0}_{i_1 i_2 \ldots i_p} \right)_{i_j = 0,1, j=0,\ldots,p}, \tag{19}$$

where

$$P\left(X_t = i_0 | X_{t-1} = i_1, \ldots, X_{t-p} = i_p \right) =: p^{i_0}_{i_1 i_2 \ldots i_p}. \tag{20}$$

For the simulations, we had to make sure that the used specifications of Q are such that the resulting model is not a member of the gbAR class. For a set of specified transition probabilities, it is actually easy to check whether the resulting model is a gbAR model, by checking whether Equation (18) hold true.

It turns out that the class of gbAR(1) models and the class of binary Markov chains of order 1 coincide. Hence, for the simulations study, we chose transition probabilities such that Equation (18) does not hold for $p = 2, 3$. Precisely, we set

$$Q = \begin{pmatrix} 0.46 & 0.54 \\ 0.56 & 0.44 \end{pmatrix}, \quad Q = \begin{pmatrix} p^0_{00} & p^1_{00} \\ p^0_{01} & p^1_{01} \\ p^0_{10} & p^1_{10} \\ p^0_{11} & p^1_{11} \end{pmatrix} = \begin{pmatrix} 0.21 & 0.79 \\ 0.69 & 0.31 \\ 0.32 & 0.68 \\ 0.89 & 0.11 \end{pmatrix}, \quad Q = \begin{pmatrix} p^0_{000} & p^1_{000} \\ p^0_{001} & p^1_{001} \\ p^0_{010} & p^1_{010} \\ p^0_{100} & p^1_{100} \\ p^0_{011} & p^1_{011} \\ p^0_{101} & p^1_{101} \\ p^0_{110} & p^1_{110} \\ p^0_{111} & p^1_{111} \end{pmatrix} = \begin{pmatrix} 0.16 & 0.84 \\ 0.26 & 0.74 \\ 0.42 & 0.58 \\ 0.21 & 0.79 \\ 0.75 & 0.25 \\ 0.64 & 0.36 \\ 0.57 & 0.43 \\ 0.94 & 0.06 \end{pmatrix}.$$

Using such transition probabilities summarized in Q, binary time series were generated. Again, a gbAR(p) process and a pth order Markov chain were fitted. In Table 4, the MSE estimation performance for the different DGPs is summarized. Interestingly, although the corresponding gbAR fits (for $p = 2, 3$) do actually estimate the wrong models, with respect to MSE over all transition probabilities, their estimation performances for small sample sizes are superior to those of Markov chains that estimate the correct models. However, for large sample sizes, the estimated Markov models outperform the mis-specified gbAR model fits. As it is estimating the true model, this pattern was expected. In summary, for time series with small sample size, where the true underlying DGP is indeed a Markov chain and not a gbAR(p) process, the parsimonious gbAR model might be a good approximation, leading potentially to more precise estimates of the transition probabilities although the model is mis-specified.

Table 3. Comparison of the estimation performance of gbAR(p) model fits and Markov chain fits of order p for $p = 1, 2, 3$ for three different gbAR-DGP1-3 with respect to the mean squared difference of estimated transition probabilities to the truth over 1000 Monte Carlo replications.

T	MSE gbAR(1)	MSE MC $p = 1$	MSE gbAR(2)	MSE MC $p = 2$	MSE gbAR(3)	MSE MC $p = 3$
100	0.01887	0.02690	0.02647	0.02674	0.03302	0.06942
200	0.01288	0.02399	0.01889	0.01904	0.02333	0.04935
500	0.00788	0.02256	0.01125	0.01191	0.01536	0.02940
1000	0.00529	0.02168	0.00775	0.00852	0.01074	0.02024

Table 4. Comparison of a gbAR Model and Markov chain by its difference of the transition probabilities to the truth from an underlying Markov Process.

T	MSE gbAR(1)	MSE MC $p = 1$	MSE gbAR(2)	MSE MC $p = 2$	MSE gbAR(3)	MSE MC $p = 3$
100	0.03183	0.03310	0.02563	0.02784	0.02771	0.03216
200	0.02219	0.02365	0.01894	0.01951	0.02276	0.02202
500	0.01420	0.01610	0.01371	0.01242	0.01960	0.01380
1000	0.00980	0.01216	0.01132	0.00882	0.01838	0.00977

4. Further Extension: The Generalized Binary ARMA Class

In this section, we extend the gbAR model class and give a definition of generalized binary ARMA (gbARMA) models that additionally contain a moving average part in their model equations. In the spirit of the gbAR model as an extension of the NDAR model class, we allow also for negative parameters in the moving-average part of the model.

First, we provide the definition of the gbARMA(p,q) model, derive its stationary solution, and state some basic properties of marginal, joint, and transition probabilities of gbARMA(p,q) processes. We conclude this section with an example of a gbARMA(1,1) process.

4.1. gbARMA Models

To be most flexible, the gbARMA model class allows additionally for negative parameters to capture negative dependence structure also in the moving average part. As before, we assume $\beta^{(0)} \in (0, 1]$ for identification reasons. In the gbARMA(p,q) model class, the parameters $\alpha^{(i)}$ and $\beta^{(j)}$ are allowed to be either positive or negative, e.g., $\alpha^{(i)}, \beta^{(j)} \in (-1, 1)$ for $i = 1, \ldots, p$ and $j = 1, \ldots, q$. To modify the parameter vector $P := \left[\alpha^{(1)}, \ldots, \alpha^{(p)}, \beta^{(0)}, \ldots, \beta^{(q)} \right]$, again such that it contains the probabilities, we define

$$P_{|\cdot|} := \left[|\alpha^{(1)}|, \ldots, |\alpha^{(p)}|, \beta^{(0)}, |\beta^{(1)}| \ldots, |\beta^{(q)}| \right]. \tag{21}$$

Definition 2 (Generalized binary ARMA processes). *Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary process which takes values in $\{0, 1\}$. Let $(e_t)_{t \in \mathbb{Z}}$ be a binary error process such that e_t is independent of $(X_s)_{s < t}$ with mean μ_e and variance $\sigma_e^2 = Var(e_t) > 0$. Let $P := \left[\alpha^{(1)}, \ldots, \alpha^{(p)}, \beta^{(0)}, \ldots, \beta^{(q)} \right]$ be the parameter vector with $P_{|\cdot|}$ as in Equation (21) such that $P_{|\cdot|} \mathbb{1}_{p+q+1} = 1$. Further, let*

$$P_t := \left(a_t^{(1)}, \ldots, a_t^{(p)}, b_t^{(0)}, \ldots, b_t^{(q)} \right) \sim Mult\left(1; P_{|\cdot|} \right), \quad t \in \mathbb{Z}$$

be iid random vectors, which are independent of $(e_t)_{t \in \mathbb{Z}}$ and $(X_t)_{s<t}$. Then, the process $(X_t)_{t \in \mathbb{Z}}$ is said to be a generalized binary ARMA(p,q) process, if it follows the recursion

$$X_t = \sum_{i=1}^{p} \left[a_t^{(+,i)} X_{t-i} + a_t^{(-,i)} \right] + b_t^{(0)} e_t + \sum_{j=1}^{q} \left[b_t^{(+,j)} e_{t-j} + b_t^{(-,j)} \right] \tag{22}$$

with $a_t^{(+,i)} := a_t^{(i)} \left(\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}} \right)$, $a_t^{(-,i)} := a_t^{(i)} \mathbb{1}_{\{\alpha^{(i)} < 0\}}$ and analogous definitions for $b_t^{(+,\cdot)}$ and $b_t^{(-,\cdot)}$.

The model parameters are contained in the vector P with entries $\alpha^{(i)} \in (-1,1)$ for $i = 1, \ldots, p$, $\beta^{(0)} \in (0,1]$ and $\beta^{(j)} \in (-1,1)$ for $j = 1, \ldots, q$. Note that, as $\beta^{(0)} > 0$ holds, no random variable $b_t^{(-,0)}$ is contained in the model equation.

With probability $\sum_{i=1}^{p} |\alpha^{(i)}|$, a predecessor X_s, $s \in \{t-1, \ldots, t-p\}$ is chosen, whereas, with probability $\sum_{j=0}^{q} |\beta^{(j)}|$, the process takes the value of an error term e_s, $s \in \{t, \ldots, t-q\}$, where it follows that $\sum_{i=1}^{p} |\alpha^{(i)}| + \sum_{j=0}^{q} |\beta^{(j)}| = 1$.

4.2. Stochastic Properties of gbARMA Models

When dealing with possibly negative parameters also in the moving-average part of gbARMA models, the idea of Equation (4) is employed also for the lagged error terms. Hence, this allows modeling negative dependence in the moving average part as well. In the multinomial distribution, all values of the parameter vector P have to be considered in absolute value, thus we have to use $P_{|\cdot|}$ as defined in Equation (21). For the expectation of gbARMA processes, two additional sums show up in comparison to the NDARMA case. Precisely, we have

$$\mu_X = \frac{\sum_{i=1}^{p} \alpha^{(-,i)} + \sum_{j=1}^{q} \beta^{(-,j)} + \sum_{j=0}^{q} \beta^{(j)} \mu_e}{1 - \sum_{i=1}^{p} \alpha^{(i)}}.$$

The construction of the stationary solution of the gbARMA time series is similar to the construction of the gbAR(p) process introduced in Section 2.1 and (Lütkepohl 2005, Chap. 11.3.2). The vector representation of the process $(Y_t, t \in \mathbb{Z})$ is equipped with a moving average part and thus the dimension of the corresponding random matrices becomes $p + q \times p + q$. Precisely, the gbARMA(p,q) model can be written as a $(p+q)$-dimensional gbVAR(1) process $(Y_t, t \in \mathbb{Z})$ with the following matrices and vectors, such that the first entry of $(Y_t, t \in \mathbb{Z})$ is equal to the gbARMA(p,q) process. We define

$$Y_t := \begin{pmatrix} X_t \\ \vdots \\ X_{t-p+1} \\ e_t \\ \vdots \\ e_{t-q+1} \end{pmatrix} \quad ((p+q) \times 1) \quad \text{and} \quad U_t := \begin{pmatrix} e_t \\ 0 \\ \vdots \\ 0 \\ e_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad ((p+q) \times 1).$$

To obtain a vector autoregressive representation for Y_t, we define directly matrices that contain the random variables of the multinomial distribution. Precisely, for $\cdot = \{-, +\}$, let

$$\widetilde{A}_t^{(\cdot)} := \begin{pmatrix} A_{t,11}^{(\cdot)} & A_{t,12}^{(\cdot)} \\ A_{t,21}^{(\cdot)} & A_{t,22}^{(\cdot)} \end{pmatrix} \text{ and } \widetilde{B}_t^{(1)} := \begin{pmatrix} b_t^{(0)} & 0 & \cdots & 0 \\ 0_{p-1\times 1} & 0_{p-1\times 1} & \cdots & 0_{p-1\times 1} \\ 1 & 0 & \ddots & \vdots \\ 0_{q-1\times 1} & \cdots & & 0_{q-1\times 1} \end{pmatrix}$$

be $(p+q) \times (p+q)$ matrices, where

$$A_{t,11}^{(\cdot)} := \begin{pmatrix} a_t^{(\cdot,1)} & \cdots & a_t^{(\cdot,p-1)} & a_t^{(\cdot,p)} \\ 1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}, A_{t,12}^{(\cdot)} := \begin{pmatrix} b_t^{(\cdot,1)} & \cdots & b_t^{(\cdot,q)} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix},$$

$$A_{t,22}^{(\cdot)} := \begin{pmatrix} 0 & \cdots & 0 & 0 \\ 1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}$$

are $p \times p$, $p \times q$ and $q \times q$ matrices, respectively, and $A_{t,21} := 0_{q\times p}$. Based on the notation introduced above, gbARMA(p,q) processes can be represented as a vector-valued gbAR model of first order (gbVAR(1)) as follows

$$Y_t = \widetilde{A}_t^{(+)} Y_{t-1} + \widetilde{A}_t^{(-)} \mathbb{1}_{p+q} + \widetilde{B}_t^{(1)} U_t \tag{23}$$

with $\mathbb{1}_{p+q}$ being the one vector of length $p + q$.

To derive a suitable stationarity condition for the process, we know from Lütkepohl (2005) that it corresponds to the characteristic polynomial of the parameter matrix $\widetilde{\mathcal{A}} := E\left(\widetilde{A}_t^{(+)}\right)$.

$$det\left(I_{K(p+q)} - \widetilde{\mathcal{A}}z\right) \neq 0 \quad \forall \ |z| \leq 1.$$

From the block structure of $\widetilde{\mathcal{A}}$, the polynomial can be reduced to the determinant of the block matrices $\mathcal{A}_{11}^{(+)} := E\left(A_{t,11}^{(+)}\right)$ and $\mathcal{A}_{22}^{(+)} := E\left(A_{t,22}^{(+)}\right)$. Hence, a gbARMA(p,q) process is stationary if the roots of the characteristic polynomial of the autoregressive part lie outside the unit circle, that is, if

$$\left(1 - \alpha^{(1)}z - \ldots - \alpha^{(p)}z^p\right) \neq 0 \ \forall \ |z| \leq 1$$

holds. The assumption is fulfilled whenever an error term has a positive probability, such that there exists a $|\beta^{(j)}| > 0$ for some $j \in \{0, \ldots, q\}$. Therefore, the sum over all probabilities of choosing a predecessor fulfills $\sum_{i=1}^{p} |\alpha^{(i)}| < 1$. Without any restriction, we assume that $\beta^{(0)}$ is strictly positive for a stationary gbARMA process, i.e., $\beta^{(0)} \in (0, 1]$.

For a stationary gbARMA(p,q) process, a moving average representation can be derived using the above defined vectors and matrices.

Theorem 3 (Moving Average representation of gbARMA processes). *Let* $(X_t)_{t \in \mathbb{Z}}$ *be a stationary gbARMA(p,q) process with gbVAR(1) representation (Equation (23)). Then, it follows that*

$$X_t = e_1^T \left(\sum_{i=0}^{\infty} \prod_{j=0}^{i-1} \widetilde{A}_{t-j}^{(+)} \widetilde{A}_{t-i}^{(-)} \mathbb{1}_{p+q} + \sum_{i=0}^{\infty} \prod_{j=0}^{i-1} \widetilde{A}_{t-j}^{(+)} \widetilde{B}_{t-i}^{(1)} U_{t-i} \right),$$

where $\lim_{k \to \infty} \prod_{i=0}^{k} \widetilde{A}_{t-i}^{(+)} = 0_{(p+q) \times (p+q)}$ *in* L_1 *and* e_1 *is the first unit vector.*

The univariate moving average representation is obtained from the multivariate formula by multiplying it with the first unit vector because of $X_t = e_1^T Y_t$.

Considering the autocorrelation structure, Jacobs and Lewis (1983) and Weiß (2011) showed that the NDARMA(p,q) model fulfils a set of Yule–Walker type equations which was also derived by Möller and Weiß (2018) for the GenDARMA class of categorical processes. The following result shows that this property is maintained for the gbARMA class.

Theorem 4 (Yule–Walker-type equations). *Let* $(X_t)_{t \in \mathbb{Z}}$ *be a stationary gbARMA(p,q) process. Set* $\beta^{(k)} := 0$ *for* $k > q$. *Define coefficients* $(\phi_k)_{k \in \mathbb{Z}}$ *recursively by*

$$\phi_k = 0 \ \text{ for } k < 0, \ \ \phi_0 = \beta^{(0)}, \ \ \phi_k = \sum_{i=1}^{p} \alpha^{(i)} \phi_{k-i} + \beta^{(k)} \ \text{ for } k > 0.$$

Then, the autocovariance function for lags $k > 0$ *is obtained by*

$$\gamma(k) - \sum_{i=1}^{p} \alpha^{(i)} \gamma(|k-i|) = \sigma_e^2 \sum_{j=k}^{q} \beta^{(j)} \phi_{j-k}$$

The autocovariances of the NDARMA and GenDARMA processes can only be positive, whereas the Yule–Walker type equations of gbARMA processes allow for possibly negative model parameters $\alpha^{(i)}, \beta^{(j)} \in (-1,1)$ for $i = 1, \ldots, p$ and $j = 1, \ldots, q$.

For the generalized binary ARMA model, formulas for the marginal, joint and transition probabilities can be calculated, extending the results from Lemma 1.

Lemma 2 (Marginal, joint, and transition probability of gbARMA processes). *Let* $(X_t)_{t \in \mathbb{Z}}$ *be a stationary gbARMA(p,q) process. Then, the following properties hold:*

(i) $P(X_t = i_0 | X_{t-1} = i_1, \ldots, e_t = j_0 \ldots)$
$\quad = \sum_{l=1}^{p} |\alpha^{(l)}| \left[\mathbb{1}_{\{\alpha^{(l)} \geq 0\}} \delta_{i_0 i_l} + \mathbb{1}_{\{\alpha^{(l)} < 0\}} \delta_{i_0(1-i_l)} \right] + \beta^{(0)} \delta_{i_0 j_0}$
$\quad + \sum_{k=1}^{q} |\beta^{(k)}| \left[\mathbb{1}_{\{\beta^{(k)} \geq 0\}} \delta_{i_0 j_k} + \mathbb{1}_{\{\beta^{(k)} < 0\}} \delta_{i_0(1-j_k)} \right]$

(ii) *Defining* $p_i := P(e_t = i)$ *then it follows*
$\quad P(X_t = i_0 | X_{t-1} = i_1, \ldots, e_{t-1} = j_1 \ldots)$
$\quad = \sum_{l=1}^{p} |\alpha^{(l)}| \left[\mathbb{1}_{\{\alpha^{(l)} \geq 0\}} \delta_{i_0 i_l} + \mathbb{1}_{\{\alpha^{(l)} < 0\}} \delta_{i_0(1-i_l)} \right]$
$\quad + \sum_{k=1}^{q} |\beta^{(k)}| \left[\mathbb{1}_{\{\beta^{(k)} \geq 0\}} \delta_{i_0 j_k} + \mathbb{1}_{\{\beta^{(k)} < 0\}} \delta_{i_0(1-j_k)} \right] + \beta^{(0)} p_{i_0}$

(iii) $P(X_t = j) = \dfrac{\left[\sum_{l=1}^{q} |\beta^{(l)}| \left(\mathbb{1}_{\{\beta^{(l)} \geq 0\}} - \mathbb{1}_{\{\beta^{(l)} < 0\}} \right) \right] + \beta^{(0)}}{\left(1 - \sum_{i=1}^{p} \alpha^{(i)} \right)} p_j + \dfrac{\sum_{i=1}^{p} |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} < 0\}} + \sum_{l=1}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta^{(l)} < 0\}}}{\left(1 - \sum_{i=1}^{p} \alpha^{(i)} \right)}$

(iv) $P\left(X_t = i_0, e_t = j_0\right) =$

$$
p_{i_0} p_{j_0} \left[\frac{\sum_{l=1}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta^{(l)} < 0\}}}{1 - \sum_{i=1}^{p} |\alpha^{(i)}| \left[\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}} \right]} - \sum_{l=1}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta < 0\}} \right] + \beta^{(0)} \delta_{i_0 j_0}
$$

$$
+ \sum_{l=1}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta^{(l)} < 0\}} \left(1 - p_{j_0} + \frac{p_{j_0}}{1 - \sum_{i=1}^{p} |\alpha^{(i)}| \left[\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}} \right]} \right)
$$

$$
+ \sum_{i=1}^{p} |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} < 0\}} \left(1 - p_{j_0} + \frac{p_{j_0}}{1 - \sum_{i=1}^{p} |\alpha^{(i)}| \left[\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}} \right]} \right)
$$

The flexibility of gbARMA models obtained by allowing for negative parameters shows also in the transition probabilities and in the joint and marginal distributions. Hence, more complex structures can be captured since systematic changes in the error terms are allowed as well.

We conclude this section with an example of a gbARMA(1,1) model.

Example 3 (gbARMA(1,1) process). *Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary gbARMA(1,1) process. Then, the process follows the recursion*

$$
X_t = a_t^{(+,1)} X_{t-1} + a_t^{(-,1)} + b_t^{(0)} e_t + b_t^{(+,1)} e_{t-1} + b_t^{(-,1)}
$$

Four sign combinations of parameter pairs are possible and the corresponding model equations are given as follows:

$$
X_t = \begin{cases} a_t^{(1)} X_{t-1} + b_t^{(0)} e_t + b_t^{(1)} e_{t-1} & \text{for } \alpha^{(1)} \geq 0, \beta^{(1)} \geq 0 \\ a_t^{(1)} \left(1 - X_{t-1}\right) + b_t^{(0)} e_t + b_t^{(1)} e_{t-1} & \text{for } \alpha^{(1)} < 0, \beta^{(1)} \geq 0 \\ a_t^{(1)} X_{t-1} + b_t^{(0)} e_t + b_t^{(1)} \left(1 - e_{t-1}\right) & \text{for } \alpha^{(1)} \geq 0, \beta^{(1)} < 0 \\ a_t^{(1)} \left(1 - X_{t-1}\right) + b_t^{(0)} e_t + b_t^{(1)} \left(1 - e_{t-1}\right) & \text{for } \alpha^{(1)} < 0, \beta^{(1)} < 0 \end{cases}.
$$

Whereas for identification purposes $\beta^{(0)}$ only takes positive values, the predecessors X_{t-1} and e_{t-1} are systematically switched if the corresponding model parameters are negative, respectively.

For a stationary gbARMA(1,1) process, the moving average representation fulfills the following equation:

$$
X_t = \sum_{j=0}^{\infty} \prod_{i=0}^{j-1} a_{t-i}^{(+,1)} a_{t-j}^{(-,1)} + \sum_{j=0}^{\infty} \prod_{i=0}^{j-1} a_{t-i}^{(+,1)} b_{t-j}^{(0)} e_{t-j} + \sum_{j=0}^{\infty} \prod_{i=0}^{j-1} a_{t-i}^{(+,1)} \left[b_{t-j}^{(+,1)} e_{t-(j+1)} + b_{t-j}^{(-,1)} \right].
$$

From the stationarity assumption, we have $|\alpha^{(1)}| < 1$, $\beta^{(0)} \in (0, 1]$ and $|\beta^{(1)}| \in [0, 1)$.

The moving average representation consists of three parts. There first is a sum over all terms $a_{t-j}^{(-,1)}$ for the potential case of $\alpha^{(1)} < 0$. This part accounts for the choosing a predecessor and its switching. Since $\beta^{(0)}$ is strictly positive, the second is a sum over all error terms without any modification occurs. In the third sum, the random variable $b_t^{(-,1)}$ appears for controlling the case of $\beta^{(1)} < 0$.

5. Conclusions

By extending the NDARMA model class of Jacobs and Lewis (1983) to allow for negative parameters in the binary state space, the generalized binary ARMA model remains parsimonious, but it becomes more flexible to allow for negative model parameters and, hence, negative dependence structure in the

data. The extension of the model to a more general parameter space enables the application to real data without having that many restrictions as in the NDARMA model class. Although the extension leads to additional terms in the model equation, the Yule–Walker equations still provide a direct way to estimate the model parameters.

We discuss stationarity conditions for gbARMA models and derive the stationary solution. The resulting moving average representation shows an additional term, compared to most MA(∞)-type representations. These additional terms control for the switching of the states.

An illustration of autocorrelation pairs $(\rho\,(1)\,,\rho\,(2))$ of four different models of order 2 shows a comparison of the captured dependence structure of the time series models. It reveals that the proposed gbARMA model can capture a wide range of negative and positive dependence structures. A second-order Markov chain is shown to capture only a slightly larger range of negative dependence structure that gbAR(2) models. Hence, by allowing for negative parameters, the proposed extension of the NDARMA model class leads to a new model class that allows capturing a wide range of dependence structures in binary time series data, while maintaining a parsimonious parametrization. Moreover, in small sample sizes, parsimonious gbAR models might turn out to be beneficial in cases where the model is actually mis-specified as they may provide a sufficient approximation to the true model.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Proof of Theorem 1

Proof. (i) By recursively inserting the model equation, the process can be expressed as

$$X_t = \prod_{i=0}^{k-1} a_{t-i}^{(+)} X_{t-k} + \sum_{i=0}^{k-1}\prod_{j=0}^{i-1} a_{t-j}^{(+)} a_{t-i}^{(-)} + \sum_{i=0}^{k-1}\prod_{j=0}^{i-1} a_{t-j}^{(+)} b_{t-i}^{(0)} e_{t-i}.$$

Since the random variables $a_t^{(+)}$ take values in $\{-1,0,1\}$ and the process is also binary with mean $\mu_X \in (0,1)$, the convergence of the first part follows directly in quadratic mean (in L_2 sense), that is,

$$E\left(|\prod_{i=0}^{k-1}\left(a_{t-i}^{(+)}\right)X_{t-k}|^2\right) = E\left(\prod_{i=0}^{k-1}|a_{t-i}|^2|X_{t-k}|^2\right) = E\left(\prod_{i=0}^{k-1} a_{t-i}X_{t-k}\right)$$
$$= \alpha^k E\left(X_{t-k}\right) \longrightarrow_{k\to\infty} 0 \quad \text{for } \alpha \in (-1,1)\,.$$

Part (ii) follows from Theorem 3 by setting $q = 0$. Its proof can be found in Section A.2. $\square$

Appendix A.2. Proof of Theorem 3

Proof. The convergence is shown by using the p-norm $\|A\|_p$ for a matrix A, which is induced by the vector norm, that is,

$$\|A\|_p = \left(\sum_{j=1}^{K} \sum_{i=1}^{K} |a_{ij}|^p \right)^{\frac{1}{p}}.$$

For $p = 1$, we get

$$\lim_{k \to \infty} E \left\| \prod_{l=0}^{k-1} \widetilde{A}_{t-l}^{(+)} \right\|_1 = \lim_{k \to \infty} E \left(\sum_{i=1}^{p+q} \sum_{j=1}^{p+q} | \left(\prod_{l=0}^{k-1} \widetilde{A}_{t-l}^{(+)} \right) |_{ij} \right)$$

$$= \lim_{k \to \infty} E \left[\begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \left(|\prod_{l=0}^{k-1} \widetilde{A}_{t-l}^{(+)}| \right) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right] = \lim_{k \to \infty} \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \prod_{l=0}^{k-1} E \left[|\widetilde{A}_{t-l}^{(+)}| \right] \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$= \lim_{k \to \infty} \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \prod_{l=0}^{k-1} E \left[\widetilde{A}_{t-i} \right] \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \lim_{k \to \infty} \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \prod_{l=0}^{k-1} \widetilde{\mathcal{A}}_{|\cdot|} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$= \lim_{k \to \infty} \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \widetilde{\mathcal{A}}_{|\cdot|}^k \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \xrightarrow{k \to \infty} 0_{(p+q) \times (p+q)} \quad for |\alpha_{ij}| \in [0, 1).$$

Since the entries of $\widetilde{A}_{t-l}^{(+)}$ lie in $\{-1, 0, 1\}$, it follows that $|\widetilde{A}_{t-l}^{(+)}| := \widetilde{A}_{t-l} \in \{0, 1\}$ and the expectation is given by $E \left(\widetilde{A}_{t-l} \right) =: \widetilde{\mathcal{A}}_{|\cdot|}$ with only positive values in $(0, 1)$. Consequently, the term vanishes for $k \to \infty$ and as result we get the gbVMA(∞) representation. $\square$

Appendix A.3. Proof of Theorem 2 and 4

Proof. For a stationary gbARMA process, it follows

$$\gamma(k) = Cov(X_t, X_{t-k})$$

$$= Cov \left(\sum_{i=1}^{p} a_t^{(i)} \left[\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} X_{t-i} + \mathbb{1}_{\{\alpha^{(i)} < 0\}} (1 - X_{t-i}) \right] + \sum_{j=0}^{q} b_t^{(j)} \left[\mathbb{1}_{\{\beta^{(j)} \geq 0\}} e_{t-j} + \mathbb{1}_{\{\beta^{(j)} < 0\}} (1 - e_{t-j}) \right], X_{t-k} \right)$$

$$= \sum_{i=1}^{p} \alpha^{(i)} Cov(X_{t-i}, X_{t-k}) + \sum_{j=0}^{q} \beta^{(j)} Cov(e_{t-j}, X_{t-k})$$

The above equation leads in the Yule–Walker equations of Theorem 2 (*i*) since $q = 0$ and $Cov(e_t, X_{t-k}) = 0$ for $k > 0$.

We now have to consider the covariance of the error terms and the time series. Therefore, we define the variable $\varphi_k := Cov(X_t, e_{t-k})$. For $k < 0$, it follows $\varphi_k = 0$ and, for $k = 0$, we have

$$\varphi_0 = Cov(X_t, e_t) = \sum_{i=1}^{p} \alpha^{(i)} Cov(X_{t-i}, e_t) + \sum_{j=0}^{q} \beta^{(j)} Cov(e_{t-j}, e_t) = \sigma_e^2 \beta^{(0)}.$$

For $k \geq 1$, we get

$$\varphi_k = Cov\left(X_t, e_{t-k}\right) = \sum_{i=1}^{p} \alpha^{(i)} Cov\left(X_{t-i}, e_{t-k}\right) + \sum_{j=0}^{q} \beta^{(j)} Cov\left(e_{t-j}, e_{t-k}\right)$$

$$= \sum_{i=1}^{p} \alpha^{(i)} \varphi_{k-i} + \beta^{(k)} \sigma_e^2.$$

By defining $\phi_k := \frac{\varphi_k}{\sigma_e^2}$, the recursion of Theorem 4 follows.

Part (ii) of Theorem 2 is directly obtained by inserting the model equation and by using the property of multinomial choosing only one entry of P_t equal to one and all others to zero, such that

$$E\left(\left(a_t^{(+,i)}\right)^2\right) = |\alpha^{(i)}|, \quad E\left(\left(a_t^{(-,i)}\right)^2\right) = |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} < 0\}}.$$

□

Appendix A.4. Proof of Lemma 1 and 2

Proof. *(i)* The conditional probability is an immediate consequence from the model equation and multinomial distribution of the random variables a_t and b_t.

(ii) With the independence assumption on the error terms and Part (i), the conditional probability without conditioning on the current error term is:

$$P\left(X_t = i_0 | X_{t-1} = i_1, \ldots, e_{t-1} = j_1. \ldots\right)$$

$$= \sum_{j_0=0}^{1} P\left(X_t = i_0 | X_{t-1} = i_1, \ldots, e_t = j_0, \ldots\right) P\left(e_t = j_0 | X_{t-1} = i_1, \ldots, e_{t-1} = j_1, \ldots\right)$$

$$= \sum_{l=1}^{p} |\alpha^{(l)}| \left[\mathbb{1}_{\{\alpha^{(l)} \geq 0\}} \delta_{i_0 i_l} + \mathbb{1}_{\{\alpha^{(l)} < 0\}} \delta_{i_0(1-i_l)}\right] + \sum_{k=1}^{q} |\beta^{(k)}| \left[\mathbb{1}_{\{\beta^{(k)} \geq 0\}} \delta_{i_0 j_k} + \mathbb{1}_{\{\beta^{(k)} < 0\}} \delta_{i_0(1-j_k)}\right]$$

$$+ \beta^{(0)} p_{i_0}.$$

(iii) Consider the probability that the time series is in state $j \in \{0,1\}$ at time point t and note that $P\left(e_t = 1 - j\right) = 1 - P\left(e_t = j\right)$, $P\left(X_t = 1 - j\right) = 1 - P\left(X_t = j\right)$.

$$P\left(X_t = j\right) = \sum_{i=1}^{p} |\alpha^{(i)}| \left[\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} P\left(X_{t-i} = j\right) + \mathbb{1}_{\{\alpha^{(i)} < 0\}} P\left(X_{t-i} = 1 - j\right)\right]$$

$$+ \sum_{l=1}^{q} |\beta^{(l)}| \left[\mathbb{1}_{\{\beta^{(l)} \geq 0\}} P\left(e_{t-l} = j\right) + \mathbb{1}_{\{\beta^{(l)} < 0\}} P\left(e_{t-l} = 1 - j\right)\right] + \beta^{(0)} P\left(e_t = j\right)$$

$$= \sum_{i=1}^{p} |\alpha^{(i)}| \left[\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} P\left(X_t = j\right) + \mathbb{1}_{\{\alpha^{(i)} < 0\}} P\left(X_t = 1 - j\right)\right]$$

$$+ \sum_{l=1}^{q} |\beta^{(l)}| \left[\mathbb{1}_{\{\beta^{(l)} \geq 0\}} P\left(e_t = j\right) + \mathbb{1}_{\{\beta^{(l)} < 0\}} P\left(e_t = 1 - j\right)\right] + \beta^{(0)} P\left(e_t = j\right)$$

$$= \left(\sum_{i=1}^{p} |\alpha^{(i)}| \left[\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}}\right]\right) P\left(X_t = j\right) + \left(\sum_{l=1}^{q} |\beta^{(l)}| \left[\mathbb{1}_{\{\beta^{(l)} \geq 0\}} - \mathbb{1}_{\{\beta^{(l)} < 0\}}\right]\right) p_j$$

$$+ \beta^{(0)} p_j + \sum_{i=1}^{p} |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} < 0\}} + \sum_{l=1}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta^{(l)} < 0\}}.$$

Then, by rearranging the terms on the last right side, we get

$$
\begin{aligned}
&P\left(X_t = j\right) \\
&= \frac{\left[\sum_{l=1}^{q} |\beta^{(l)}| \left(\mathbb{1}_{\{\beta^{(l)} \geq 0\}} - \mathbb{1}_{\{\beta^{(l)} < 0\}}\right)\right] + \beta^{(0)}}{\left(1 - \sum_{i=1}^{p} \alpha^{(i)}\right)} p_j + \frac{\sum_{i=1}^{p} |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} < 0\}} + \sum_{l=0}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta^{(l)} < 0\}}}{\left(1 - \sum_{i=1}^{p} \alpha^{(i)}\right)}.
\end{aligned}
$$

(iv) Consider the joint probability of the error term and time series at time point t. We get

$$
\begin{aligned}
&P\left(X_t = i_0, e_t = j_0\right) \\
&= \sum_{i=1}^{p} |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} \geq 0\}} P\left(X_{t-i} = i_0, e_t = j_0\right) + \sum_{i=1}^{p} |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} < 0\}} P\left(X_{t-i} = 1 - i_0, e_t = j_0\right) \\
&\quad + \sum_{l=1}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta^{(l)} \geq 0\}} P\left(e_{t-l} = i_0, e_t = j_0\right) + \sum_{l=1}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta^{(l)} < 0\}} P\left(e_{t-l} = i_0, e_t = j_0\right) \\
&\quad + \beta^{(0)} P\left(e_t = i_0, e_t = j_0\right) \\
&= \sum_{i=1}^{p} |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} \geq 0\}} P\left(X_{t-i} = i_0\right) p_{j_0} + \sum_{i=1}^{p} |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} < 0\}} P\left(X_{t-i} = 1 - i_0\right) p_{j_0} \\
&\quad + \sum_{l=1}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta^{(l)} \geq 0\}} P\left(e_{t-l} = i_0\right) p_{j_0} + \sum_{l=1}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta^{(l)} < 0\}} P\left(e_{t-l} = 1 - i_0\right) p_{j_0} \\
&\quad + \beta^{(0)} P\left(e_t = i_0, e_t = j_0\right).
\end{aligned}
$$

By inserting Part *(iii)* into the equation above, we get

$$
\begin{aligned}
&\sum_{i=1}^{p} |\alpha^{(i)}| p_{j_0} \left[\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}}\right] \left(\frac{1 - \sum_{i=1}^{p} |\alpha^{(i)}| - \sum_{l=0}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta^{(l)} < 0\}}}{\left(1 - \left[\sum_{i=1}^{p} |\alpha^{(i)}| \left(\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}}\right)\right]\right)} p_{i_0} \right. \\
&\quad \left. + \frac{\sum_{i=1}^{p} |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} < 0\}} + \sum_{l=0}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta^{(l)} < 0\}}}{\left(1 - \left[\sum_{i=1}^{p} |\alpha^{(i)}| \left(\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}}\right)\right]\right)}\right) + \beta^{(0)} p_{j_0} \delta_{i_0, j_0} \\
&\quad + \sum_{i=1}^{p} |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} < 0\}} + \sum_{l=1}^{q} |\beta^{(l)}| p_{j_0} \left[\mathbb{1}_{\{\beta^{(l)} \geq 0\}} p_{i_0} + \mathbb{1}_{\{\beta^{(l)} < 0\}} \left(1 - p_{i_0}\right)\right].
\end{aligned}
$$

Using the properties of the parameters, the joint distribution of the time series and error term is given by

$$
\begin{aligned}
&P\left(X_t = i_0, e_t = j_0\right) \\
&= p_{i_0} p_{j_0} \left[\frac{\sum_{l=1}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta^{(l)} < 0\}}}{1 - \sum_{i=1}^{p} |\alpha^{(i)}| \left[\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}}\right]} - \sum_{l=1}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta < 0\}}\right] + \beta^{(0)} \delta_{i_0 j_0} \\
&\quad + \sum_{l=1}^{q} |\beta^{(l)}| \mathbb{1}_{\{\beta^{(l)} < 0\}} \left(1 - p_{j_0} + \frac{p_{j_0}}{1 - \sum_{i=1}^{p} |\alpha^{(i)}| \left[\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}}\right]}\right) \\
&\quad + \sum_{i=1}^{p} |\alpha^{(i)}| \mathbb{1}_{\{\alpha^{(i)} < 0\}} \left(1 - p_{j_0} + \frac{p_{j_0}}{1 - \sum_{i=1}^{p} |\alpha^{(i)}| \left[\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}}\right]}\right).
\end{aligned}
$$

□

References

Bellégo, Christophe, and Laurent Ferrara. 2009. Forecasting Euro-Area Recessions Using Time-Varying Binary Response Models for Financial. Working Paper, Banque de France, Paris, France, vol. 259.

Buishand, T. Adri. 1978. *The Binary DARMA(1,1) Process as a Model for Wet-Dry-Sequences*. Technical Note 78-01. Wageningen: Department of Mathematics, Statistics Division, Agricultural University Wageningen.

Fitzmaurice, Garrett M., and Stuart R. Lipsitz. 1995. A Model for Binary Time Series Data with Serial Odds Ratio Patterns. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 44: 51–61. [CrossRef]

Gouveia, Sónia, Tobias A. Möller, Christian H. Weiß, and Manuel G. Scotto. 2018. A full ARMA model for counts with bounded support and its application to rainy-days time series. *Stochastic Environmental Research and Risk Assessment* 32: 2495–514. [CrossRef]

Jacobs, Patricia A., and Peter A. W. Lewis. 1983. Stationary Discrete Autoregressive-Moving Average Time Series Generated By Mixtures. *Journal of Time Series Analysis* 4: 19–36. [CrossRef]

Kanter, Marek. 1975. Autoregression for discrete processes mod 2. *Journal of Applied Probability* 12: 371–75. [CrossRef]

Kedem, Benjamin. 1980. Binary Time Series. In *Lecture Notes in Pure and Applied Mathematics*. New York: Marcel Dekker, Inc.

Kedem, Benjamin, and Konstantinos Fokianos. 2002. *Regression Models for Time Series Analysis*. Hoboken: Whiley & Sons.

Lütkepohl, Helmut. 2005. *New Introduction to Multiple Time Series Analysis*. Berlin/Heidelberg: Springer.

McKenzie, Eddie. 1981. Extending the correlation structure of exponential autoregressive moving average processes. *Journal of Applied Probability* 18: 181–89. [CrossRef]

McKenzie, Eddie. 2009. Discrete variate time series. *Handbook of Statistics* 21: 573–606.

Möller, Tobias A., and Christian H. Weiß. 2018. Generalized Discrete ARMA Models. unpublished manuscript.

Weiß, Christian H., and Rainer Göb. 2008. Measuring serial dependence in categorical time series. *Advances in Statistical Analysis* 92: 71–89. [CrossRef]

Weiß, Christian H. 2009a. *Categorical Time Series Analysis and Application in Statistical Quality Control*. Berlin: de-Verlag im Internet GmbH.

Weiß, Christian H. 2009b. Properties of a class of binary ARMA models. *Statistics* 43: 131–38. [CrossRef]

Weiß, Christian H. 2011. Generalized choice models for categorical time series. *Journal of Statistical Planning and Inference* 141: 2849–62. [CrossRef]

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