

ORIGINAL ARTICLE

GENERALIZED BINARY VECTOR AUTOREGRESSIVE PROCESSES

CARSTEN JENTSCH^{a*}  AND LENA REICHMANN^{a,b}^a*Department of Statistics, TU Dortmund University, Dortmund, Germany*^b*Mathematical Institute, University of Mannheim, Mannheim, Germany*

Vector-valued extensions of univariate generalized binary auto-regressive (gbAR) processes are proposed that enable the joint modeling of serial and cross-sectional dependence of multi-variate binary data. The resulting class of generalized binary vector auto-regressive (gbVAR) models is parsimonious, nicely interpretable and allows also to model negative dependence. We provide stationarity conditions and derive moving-average-type representations that allow to prove geometric mixing properties. Furthermore, we derive general stochastic properties of gbVAR processes, including formulae for transition probabilities. In particular, classical Yule–Walker equations hold that facilitate parameter estimation in gbVAR models. In simulations, we investigate the estimation performance, and for illustration, we apply gbVAR models to particulate matter (PM₁₀, ‘fine dust’) alarm data observed at six monitoring stations in Stuttgart, Germany.

Received 26 November 2020; Accepted 21 July 2021

Keywords: Binary data, mixing properties, multi-variate time series, stationarity conditions, transition probabilities, Yule–Walker equations

MOS subject classification: 62M10; 62H12; 62F40; 62M20.

1. INTRODUCTION

Categorical data are collected in many fields of applications. When such data are observed over time, serial dependence is often present that has to be taken into account, for example, for modeling purposes or for statistical inference. Hence, the statistical research focusing on such data structures evolved considerably over the last years. With the collection of huge amounts of data nowadays, this leads particularly to a growing interest in statistical methods for the analysis of multi-variate categorical time series. As an important special case, multi-variate binary time series, that correspond to categorical time series data with just two categories, occur in many different contexts such as agriculture, biology, economy, engineering, environmetrics, genetics, geography, geology, medical science, natural language processing or sports; see, for example, Jentsch and Reichmann (2019) for some univariate examples. For instance, recent related literature addresses the detection of dependent Bernoulli sequences in Ritzwoller and Romano (2020) or the efficient generation of high-dimensional binary data with specified correlation structures in Jiang *et al.* (2020). Often, binary time series are obtained from binarization of observed real-valued data, when, for example, the interest is, whether some event occurs (or not) or a certain threshold is crossed (or not) instead of the actual value. Although simplified, this transformation will generally contain a great amount of the information and the dynamics of the original data.

Multi-variate binary time series obtained from a suitable thresholding procedure are for instance of much interest in economics, where periods of recession and of economic growth (no-recession) are considered; see for example, Bellégo and Ferrara (2009) and Startz (2006) on forecasting recessions in the Euro area and in the United States respectively, and Candelon *et al.* (2012), who use multi-variate dynamic probit models to study predictive relationships in binary processes with applications to financial crises.

* Correspondence to: Carsten Jentsch, Department of Statistics, TU Dortmund University, D-44221 Dortmund, Germany. E-mail: jentsch@statistik.tu-dortmund.de

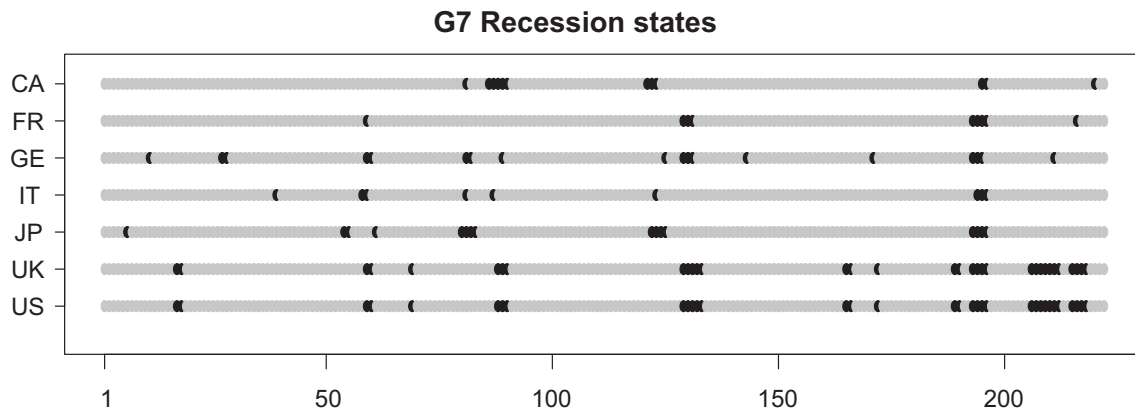


Figure 1. Quarterly time series indicating periods of recessions (black dots) and of economic growth (no dots) for the G7 countries Canada (CA), France (FR), Germany (DE), Italy (IT), Japan (JP), United Kingdom (GB), and United States (US) from Q2/1960 to Q1/2017

Considering jointly such recession time series of several countries, a multi-variate analysis allows to study not only the serial, but also the cross-sectional dependence in the data. In turn, this allows to investigate the spillover effects between several countries, that is, how a recession in one country will affect the economy in other countries in the future. In Figure 1, we show quarterly time series indicating periods of recessions and of economic growth for the G7 countries Canada, France, Germany, Italy, Japan, United Kingdom, and the United States from Q2/1960 to Q1/2017.

In signal processing, large numbers of nodes, that is, inexpensive sensors, are employed to make binary decisions whenever a signal is above a certain threshold or not, see for example, Cheng *et al.* (2013). Hence, multiple two-state time series with states ‘detection’ and ‘no detection’ are observed. In such applications, binarization of the original signal is particularly beneficial as binary data are inexpensive to store.

In recent years, there is increasing interest in air pollution in European cities and metropolitan areas. The EU established the *European emission standards*, which include limits for particulates in the air. Whenever the amount of PM_{10} (coarse particles with a diameter between 2.5 and $10\ \mu\text{m}$, ‘fine dust’) exceeds the threshold of $50\ \mu\text{g}/\text{m}^3$ at a monitoring station, this causes a ‘fine dust alarm’. Hence, for each monitoring station, this results in a binary sequence with states ‘exceedance’ and ‘no exceedance’. In fact, the current public discourse centers to a large extent around whether the threshold is exceeded or not, and less about the actual amount of fine dust measured. Typically, several monitoring stations in one city allow for a *joint* analysis of the fine dust pollution. In Figure 2,

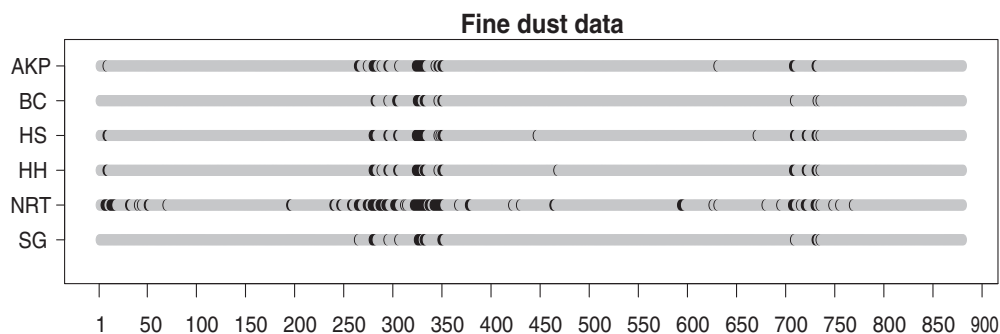


Figure 2. Daily time series indicating fine dust alarms (black dots) and no fine dust alarms (no dots) at six monitoring stations: Arnulf-Klett Platz (AKP), Bad Cannstatt (BC), Hauptstätter Straße (HS), Hohenheimer Straße (HH), Neckartor (NRT), Stadtgarten (SG) in Stuttgart, Germany from 1 March 2016 to 31 July 2018

we show the recorded fine dust alarms at six monitoring stations in the urban area of Stuttgart, Germany from 1 March 2016 to 31 July 2018. The occurrences of alarms tend to cluster, but a closer inspection reveals that the station Neckartor often shows an alarm before the others. Hence, a multi-variate analysis of this pattern might be helpful to allow for an improved prediction of future exceedances. In Section 4, we discuss this data set in more detail.

Typically, Markovian models are used to describe the dependence structure of categorical time series, see for example, Kedem (1980). Such models are very flexible and allow to capture a broad range of serial dependence, but the number of parameters grows exponentially with the order of the Markov model. As indicated by McKenzie (2003), already in the univariate case, this likely leads to over-parametrization. For a K -dimensional multi-variate binary time series, this problem is even much more intricate as the fitting of an unrestricted Markov model of order p requires the estimation of 2^{Kp} parameters. Hence, Markov models are not feasible, when the time series dimension or the model order become large. Alternatively, regression-based approaches such as logit and probit models are useful to study binary data; see for example Fokianos and Kedem (2003). Multi-variate regression analysis of panel data with binary outcomes has been studied in Czado (2000). Using nonlinear vector autoregressive models for latent variables associated with correlated binary time-series data, multi-variate dynamic probit models have been considered in Candelson *et al.* (2012). Estimation inference and identification issues in multi-variate dynamic panel data logit models are considered by Honoré and Kyriazidou (2019a) and Honoré and Kyriazidou (2019b). Such models allow for parsimonious and flexible modeling of serial dependence in binary time series. However, in contrast to classical autoregressive models, estimation in multi-variate dynamic panel models is no longer explicit and may cause identification issues in small samples as discussed, for example, in Honoré Kyriazidou (2019a, 2019b).

In the following, we develop a flexible and nicely interpretable autoregressive model framework for binary time series data which satisfies classical Yule–Walker equations such that an explicit estimation of the model parameters via Yule–Walker estimation becomes possible.

1.1. The Univariate Case: NDARMA versus gbARMA

In the univariate case, 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. To make sure that the process $(X_t, t \in \mathbb{Z})$ takes only values contained in a *discrete* state space S , their idea is to choose X_t randomly either from the past values of the time series X_{t-1}, \dots, X_{t-p} or from one of the innovations $e_t, e_{t-1}, \dots, e_{t-q}$ with certain probabilities respectively. This random selection mechanism is described by i.i.d. random vectors $(P_t, t \in \mathbb{Z})$ with

$$P_t := \left[a_t^{(1)}, \dots, a_t^{(p)}, b_t^{(0)}, \dots, b_t^{(q)} \right] \sim Mult(1; \mathcal{P}), \quad t \in \mathbb{Z}, \tag{1.1}$$

where $Mult(1; \mathcal{P})$ denotes the multinomial distribution with parameter 1 and probability vector $\mathcal{P} := \left[\alpha^{(1)}, \dots, \alpha^{(p)}, \beta^{(0)}, \dots, \beta^{(q)} \right]$ with $\alpha^{(1)}, \dots, \alpha^{(p)} \in [0, 1)$, $\beta^{(0)} \in (0, 1]$ and $\beta^{(1)}, \dots, \beta^{(q)} \in [0, 1)$ 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

$$X_t = \sum_{i=1}^p a_t^{(i)} X_{t-i} + \sum_{j=0}^q b_t^{(j)} e_{t-j}, \quad t \in \mathbb{Z}, \tag{1.2}$$

where $(e_t, t \in \mathbb{Z})$ is an i.i.d. innovation process taking values in the state space S .

NDARMA models are contained as special cases in the broad class of Markov models, but are considerably more parsimonious and still nicely interpretable due to their ARMA-type structure. In this spirit, Weiß and Göb (2008) showed that Yule–Walker-type equations hold and Weiß (2009a) discussed the connection of NDARMA models to general Markov chains.

It is important to note that the probability vector \mathcal{P} of the multinomial distribution in (1.1) contains the NDARMA model parameters, which are naturally restricted to satisfy two conditions: all entries of \mathcal{P} have to lie in the unit interval and they have to sum-up to one. Hence, in contrast to general Markov chains, NDARMA models are particularly restricted to capture exclusively *non-negative* serial dependence. To address this lacking flexibility of the NDARMA model class, Jentsch and Reichmann (2019) proposed a simple and straightforward extension of the original idea of Jacobs and Lewis (1983) that allows also to capture negative serial dependence in univariate binary time series. In the resulting generalized binary ARMA (gbARMA) model class, in contrast to NDARMA models, the parameters $\alpha^{(i)}$ and $\beta^{(j)}$ are allowed to be either positive or negative. Precisely, gbARMA models allow for $\alpha^{(1)}, \dots, \alpha^{(p)} \in (-1, 1)$ and $\beta^{(1)}, \dots, \beta^{(q)} \in (-1, 1)$, with $\beta^{(0)} \in (0, 1]$, such that $\sum_{i=1}^p |\alpha^{(i)}| + \beta^{(0)} + \sum_{j=1}^q |\beta^{(j)}| = 1$. The parameter vector $\mathcal{P} = [\alpha^{(1)}, \dots, \alpha^{(p)}, \beta^{(0)}, \dots, \beta^{(q)}]$ has to be modified to contain valid probabilities in $[0, 1]$ for the selection mechanism. We define

$$\mathcal{P}_{|\cdot|} := [|\alpha^{(1)}|, \dots, |\alpha^{(p)}|, \beta^{(0)}, |\beta^{(1)}|, \dots, |\beta^{(q)}|]. \tag{1.3}$$

As in NDARMA models, the random selection mechanism for gbARMA models is again described by i.i.d. multinomial random vectors $(P_t, t \in \mathbb{Z})$ with

$$P_t := \left(a_t^{(1)}, \dots, a_t^{(p)}, b_t^{(0)}, \dots, b_t^{(q)} \right) \sim \text{Mult} \left(1; \mathcal{P}_{|\cdot|} \right), \quad t \in \mathbb{Z}.$$

Then, the gbARMA(p,q) process $(X_t, t \in \mathbb{Z})$ follows the model equation

$$X_t = \sum_{i=1}^p \left[a_t^{(+,i)} X_{t-i} + a_t^{(-,i)} \right] + b_t^{(0)} e_t + \sum_{j=1}^q \left[b_t^{(+,j)} e_{t-j} + b_t^{(-,j)} \right], \quad t \in \mathbb{Z}, \tag{1.4}$$

where $(e_t, t \in \mathbb{Z})$ is an i.i.d. innovation process which takes values in $\{0, 1\}$. In the above, we set $a_t^{(+,i)} := a_t^{(i)} (\mathbb{1}_{\{\alpha^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha^{(i)} < 0\}})$ and $a_t^{(-,i)} := a_t^{(i)} \mathbb{1}_{\{\alpha^{(i)} < 0\}}$, $i = 1, \dots, p$, with analogous definitions for $b_t^{(+,j)}$ and $b_t^{(-,j)}$, $j = 1, \dots, q$. A detailed description of the gbARMA model class and its properties can be found in Jentsch and Reichmann (2019).

1.2. An Example: NDAR(1) versus gbAR(1)

In a nutshell, the benefit of a gbARMA model in comparison to an NDARMA model for binary data, is that it allows to pick *systematically* the opposite value of a predecessor if the corresponding model parameter is negative. To illustrate this, let us consider the simplest case of a gbAR(1) model with $\mathcal{P} = [\alpha, \beta]$ following

$$X_t = a_t^{(+)} X_{t-1} + a_t^{(-)} + b_t e_t, \quad t \in \mathbb{Z}, \tag{1.5}$$

where $a_t^{(+)} := a_t (\mathbb{1}_{\{\alpha \geq 0\}} - \mathbb{1}_{\{\alpha < 0\}})$ and $a_t^{(-)} := a_t \mathbb{1}_{\{\alpha < 0\}}$ such that $P_t := (a_t, b_t) \sim \text{Mult} (1; \mathcal{P}_{|\cdot|})$ with $\mathcal{P}_{|\cdot|} = [|\alpha|, \beta]$ and $|\alpha| + \beta = 1$. Equation (1.5) can be rewritten to get

$$X_t = \begin{cases} a_t X_{t-1} + b_t e_t, & [a_t, b_t] \sim \text{Mult} (1; \alpha, \beta), \quad \alpha \in [0, 1) \\ a_t (1 - X_{t-1}) + b_t e_t, & [a_t, b_t] \sim \text{Mult} (1; |\alpha|, \beta), \quad \alpha \in (-1, 0) \end{cases}. \tag{1.6}$$

Hence, depending on the sign of the parameter $\alpha \in (-1, 1)$, either the predecessor X_{t-1} or its opposite value $1 - X_{t-1}$ can be systematically picked by the random selection mechanism. Actually, the first equation in (1.6), that allows for $\alpha \in [0, 1)$, corresponds to an NDAR(1) model with a binary state space, which is contained in a gbAR(1) model as a special case. Hence, the gbAR(1) model is way more flexible and captures the full set

of possible parametrizations of a binary Markov chain (of first order). As derived by Jentsch and Reichmann (2019), the formula for the stationary mean $\mu_X = E(X_t)$ of a gbAR(1) process fulfills $\mu_X = \mu_e$ if $\alpha \in [0, 1)$ and $\mu_X = (-\alpha + (1 + \alpha)\mu_e)/(1 - \alpha)$ if $\alpha \in (-1, 0)$. Hence, when α is parametrized in the marginal success probability μ_X , it can be shown that the possible range of α is $\left[\max\left(\frac{-\mu_X}{1-\mu_X}, -\frac{1-\mu_X}{\mu_X}\right); 1 \right)$; see also Weiß (2009a, Remark 12.2.1.2). In particular, for $\mu_X = 1/2$, the lower bound of the possible range for α becomes -1 .

Similarly, in view of the operations $a_t X_{t-1}$ and $a_t (1 - X_{t-1})$ in (1.6), that take the value of the time series from the time point before or its opposite value, respectively, it seems plausible to combine them and allow both at the same time. That is, we could think of a model equation of the form

$$X_t = a_{t,=}X_{t-1} + a_{t,\neq}(1 - X_{t-1}) + b_{t,=}\epsilon_t, \quad t \in \mathbb{Z}, \tag{1.7}$$

where $([a_{t,=}, a_{t,\neq}, b_{t,=}], t \in \mathbb{Z})$ are i.i.d. with $[a_{t,=}, a_{t,\neq}, b_{t,=}] \sim Mult(1; \alpha_=: \alpha_+, \alpha_=: \alpha_-, \beta_=: \beta_+)$ such that $\alpha_+, \alpha_- \in [0, 1]$, $\beta_+ \in (0, 1]$ and $\alpha_+ + \alpha_- + \beta_+ = 1$ as well as $E(\epsilon_t) = \mu_e \in (0, 1)$. However, the model (1.7) is *not identified* as it is indistinguishable from the gbAR(1) model (1.6) with

$$\alpha = \alpha_+ - \alpha_-, \quad \beta = 1 - |\alpha_+ - \alpha_-| \tag{1.8}$$

and

$$\mu_e = \frac{\alpha_- \mathbb{1}_{\{\alpha_+ - \alpha_- \geq 0\}} + \alpha_+ \mathbb{1}_{\{\alpha_+ - \alpha_- < 0\}} + \beta_+ \mu_e}{1 - |\alpha_+ - \alpha_-|}. \tag{1.9}$$

We refer to the Supporting Information for a proof of (1.8) and (1.9).

In Figure 3, we show realizations and corresponding autocorrelation functions (ACRs) of a gbAR(1) process with positive parameter $\alpha \in [0, 1)$ (i.e. an NDAR(1) process) and a gbAR(1) process with negative parameter $\alpha \in (-1, 0)$. Whereas the NDAR(1) process with positive α shows long runs of the same value and a non-negative ACR, the gbAR(1) process with negative α shows an oscillating pattern and an alternating ACR.

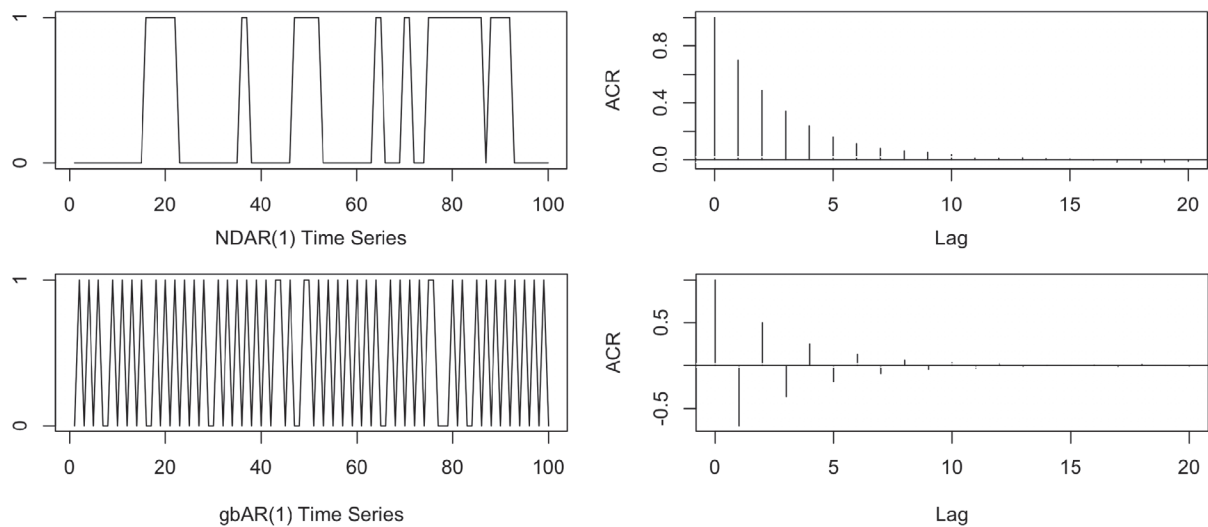


Figure 3. NDAR(1) vs. gbAR(1): Realizations and ACR of an NDAR(1) process (i.e. of a gbAR(1) process with positive α in the first line of (1.6)) (upper panel) with $\mathcal{P} = [0.7, 0.3]$ and of a gbAR(1) process (1.5) (lower panel) with $\mathcal{P} = [-0.7, 0.3]$. In both cases, we used $P(e_t = 1) = 0.5$

1.3. Toward a Multi-variate Analysis: A Bivariate gbVAR(1)

In the case, when more than just one binary time series is observed, as, for example, for the G7 recession data in Figure 1 or the fine dust alarm data in Figure 2, a multi-variate (i.e. joint) analysis is desirable.

For categorical time series data, Möller and Weiß (2020) proposed a multi-variate extension of the NDARMA class with (non-negative) *scalar* model parameters that control the multinomial selection mechanism. This approach restricts the flexibility of the resulting Generalized DARMA (GDARMA) class to model *joint* dependence to some large extent. Instead, GDARMA models make use of a variation function applied to lagged observations and innovations to increase the entry-wise variation over time. However, the proposed variation function does not modify the past time series values in a systematical way, neither is the resulting process suitable to capture any negative dependence structure.

To achieve more model flexibility, let us first consider the case of two *independent* gbAR(1) processes $(X_t, t \in \mathbb{Z})$ and $(Y_t, t \in \mathbb{Z})$. By stacking them, we get a bivariate process

$$\begin{aligned} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} &= \begin{pmatrix} a_{t,X}^{(+)}X_{t-1} + a_{t,X}^{(-)} + b_{t,X}e_{t,X} \\ a_{t,Y}^{(+)}Y_{t-1} + a_{t,Y}^{(-)} + b_{t,Y}e_{t,Y} \end{pmatrix} \\ &= \begin{pmatrix} a_{t,X}^{(+)} & 0 \\ 0 & a_{t,Y}^{(+)} \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} a_{t,X}^{(-)} & 0 \\ 0 & a_{t,Y}^{(-)} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} b_{t,X} & 0 \\ 0 & b_{t,Y} \end{pmatrix} \begin{pmatrix} e_{t,X} \\ e_{t,Y} \end{pmatrix}. \end{aligned} \tag{1.10}$$

However, due to the diagonal structure of the (random) coefficient matrices, such a model in (1.10) is not yet sufficient to study the cross-sectional dependence between two binary time series. Naturally, this can be achieved by allowing the off-diagonal elements of the coefficient matrices in (1.10) to be non-zero. This leads to the bivariate gbVAR(1) model

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} a_{t,11}^{(+)} & a_{t,12}^{(+)} \\ a_{t,21}^{(+)} & a_{t,22}^{(+)} \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} + \begin{pmatrix} a_{t,11}^{(-)} & a_{t,12}^{(-)} \\ a_{t,21}^{(-)} & a_{t,22}^{(-)} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} b_{t,11} & 0 \\ 0 & b_{t,22} \end{pmatrix} \begin{pmatrix} e_{t,1} \\ e_{t,2} \end{pmatrix}, \tag{1.11}$$

which can be compactly written as

$$X_t = A_t^{(+)}X_{t-1} + A_t^{(-)}\mathbb{1}_2 + B_t e_t, \tag{1.12}$$

where $(e_t, t \in \mathbb{Z})$ is an i.i.d. innovation process taking values in $\{0, 1\}^2$. Note that $Cov(e_t) = \Sigma_e$ is allowed to be non-diagonal, whereas B_t is imposed to be diagonal for identification reasons; compare also Remark 2.13. For comprehensive discussions of multi-variate Bernoulli distributions allowing for dependence leading to non-diagonal Σ_e , we refer to the classical article by Bahadur (1961) and the more recent work by Dai *et al.* (2013).

The model parameters of the process (1.12) are summarized in the matrix $\mathcal{P} := [\mathcal{A}, \mathcal{B}]$ with $\mathcal{A} := (\alpha_{kl})_{k,l=1,2}$ and $\mathcal{B} := \text{diag}(\beta_{11}, \beta_{22})$ and $\mathcal{P}_{|\cdot|} := [\mathcal{A}_{|\cdot|}, \mathcal{B}]$ with $\mathcal{A}_{|\cdot|} := (|\alpha_{kl}|)_{k,l=1,2}$, where $\alpha_{kl} \in (-1, 1)$, $\beta_{kk} \in (0, 1]$ for all $k, l = 1, 2$, such that $|\alpha_{k1}| + |\alpha_{k2}| + \beta_{kk} = 1$ for $k = 1, 2$.

The natural approach is to adopt row-wise the random (multinomial) selection mechanism leading to mutually *independent* i.i.d. vector-valued processes $(P_{t,1\bullet}, t \in \mathbb{Z})$ and $(P_{t,2\bullet}, t \in \mathbb{Z})$, where

$$P_{t,1\bullet} := [a_{t,11}, a_{t,12}, b_{t,11}, 0] \sim \text{Mult}(1; \mathcal{P}_{|\cdot|,1\bullet}), \tag{1.13}$$

$$P_{t,2\bullet} := [a_{t,21}, a_{t,22}, 0, b_{t,22}] \sim \text{Mult}(1; \mathcal{P}_{|\cdot|,2\bullet}). \tag{1.14}$$

Here, $\mathcal{P}_{|\cdot|}$ is defined to contain entry-wise the absolute values of \mathcal{P} with row-sums equal to one such that $\mathcal{P}_{|\cdot|,k\bullet}$ become valid arguments for the multinomial distributions in (1.13) and (1.14). Note also the desirable redundancy

in the above notation obtained by including the off-diagonal zeros of B_t in $P_{t,k\bullet}$, but this allows to use the whole rows $\mathcal{P}_{|\cdot|,k\bullet}$ as arguments of the multinomial distributions.

1.4. Outline

In the spirit of the gbVAR(1) model (1.11) as a natural extension of the univariate gbAR(1) model in (1.5), we provide a full description and investigation of the corresponding generalized binary vector AR (gbVAR) model class in this article. In Section 2, we define generalized binary VAR processes of order $p \in \mathbb{N}$ and derive general stochastic properties including formulae for the mean vector, stationarity conditions, moving-average representations, geometric mixing properties, Yule–Walker equations and transition probabilities. Furthermore, we discuss possible extensions and identification issues and address parameter estimation in gbVAR models based on Yule–Walker estimation. In Section 3, we examine the finite sample performance of these parameter estimators by means of different criteria in simulations. For illustration, we use gbVAR models to analyze a six-dimensional binary time series that indicates fine dust alarms in the urban area of Stuttgart, Germany, in Section 4. We conclude in Section 5. All proofs, additional simulation results and an extension to gbVARMA processes by adding a moving-average part are deferred to the Supporting Information.

2. THE GBVAR MODEL CLASS

In Section 2.1, we define gbVAR processes as multi-variate extensions of (univariate) gbAR models as introduced by Jentsch and Reichmann (2019). The definition naturally extends the bivariate gbVAR(1) model (1.11) to arbitrary order $p \in \mathbb{N}$ and dimension $K \in \mathbb{N}$. Stochastic properties are derived in Section 2.2. Identification issues and parameter estimation are discussed in Sections 2.3 and 2.4 respectively.

2.1. gbVAR Models

For a K -dimensional binary time series $(X_t, t \in \mathbb{Z})$, let the matrix

$$\mathcal{P} := [\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}, \mathcal{B}] \tag{2.1}$$

contain the autoregressive coefficient matrices $\mathcal{A}^{(i)} = (\alpha_{kl}^{(i)})_{k,l=1,\dots,K}$, $i = 1, \dots, p$ and $\mathcal{B} = \text{diag}(\beta_{11}, \dots, \beta_{KK})$ of a gbVAR(p) model. As the gbVAR model allows for $\alpha_{kl}^{(i)} \in (-1, 1)$, the entries of \mathcal{P} have to satisfy $\sum_{i=1}^p \sum_{l=1}^K |\alpha_{kl}^{(i)}| + \beta_{kk} = 1$, $k = 1, \dots, K$, and \mathcal{P} has to be modified to serve as a parameter matrix containing (row-wise) valid probabilities of multinomial distributions. This is achieved by taking entry-wise absolute values in \mathcal{P} and we define

$$\begin{aligned} \mathcal{P}_{|\cdot|} &:= \left[\left(|\alpha_{kl}^{(1)}| \right)_{k,l=1,\dots,K}, \dots, \left(|\alpha_{kl}^{(p)}| \right)_{k,l=1,\dots,K}, \text{diag}(\beta_{11}, \dots, \beta_{KK}) \right] \\ &=: [\mathcal{A}_{|\cdot|}^{(1)}, \dots, \mathcal{A}_{|\cdot|}^{(p)}, \mathcal{B}]. \end{aligned} \tag{2.2}$$

These prerequisites enable us to give the definition of the generalized binary vector AR model of order $p \in \mathbb{N}$.

Definition 2.1 (gbVAR(p) processes). Let $(X_t, t \in \mathbb{Z})$ be a stationary K -dimensional process taking values in $\{0, 1\}^K$. Let $(e_t, t \in \mathbb{Z})$ be an i.i.d. K -dimensional binary innovation process, such that e_t is independent of $(X_s, s < t)$ with mean vector $\mu_e = (\mu_{e,1}, \dots, \mu_{e,K})' = E(e_t)$, where $\mu_{e,i} = P(e_{t,i} = 1)$, $i = 1, \dots, K$, and variance–covariance matrix $\Sigma_e = (\sigma_{e,kl})_{k,l=1,\dots,K} = \text{Cov}(e_t)$, where $\sigma_{e,ii} = \mu_{e,i}(1 - \mu_{e,i})$. For $p \in \mathbb{N}$ let \mathcal{P} be the parameter matrix as in (2.1) with diagonal matrix \mathcal{B} , $\mathcal{A}^{(p)} \neq O_{K \times K}$, and $\mathcal{P}_{|\cdot|}$ as in (2.2) such that $\mathcal{P}_{|\cdot|} \mathbb{1}_{K(p+1)} = \mathbb{1}_K$, where $\mathbb{1}_M = (1, \dots, 1)'$ denotes the M -dimensional vector of ones. Furthermore, let $(P_t, t \in \mathbb{Z})$ with $P_t = [A_t^{(1)}, \dots, A_t^{(p)}, B_t]$

be a $K \times K(p + 1)$ -dimensional i.i.d. process with mutually independent rows $(P_{t,k\bullet}, t \in \mathbb{Z}), k = 1, \dots, K$, such that

$$P_{t,k\bullet} := \left[a_{t,k\bullet}^{(1)}, \dots, a_{t,k\bullet}^{(p)}, b_{t,k\bullet} \right] \sim \text{Mult} \left(1; \mathcal{P}_{|\cdot|,k\bullet} \right),$$

which are independent of $(e_t, t \in \mathbb{Z})$ and $(X_s, s < t)$. Here, we set $A_t^{(i)} = (a_{t,kl}^{(i)})_{k,l=1,\dots,K}$ and $B_t = (b_{t,kl})_{k,l=1,\dots,K} = \text{diag}(b_{t,11}, \dots, b_{t,KK})$ with $a_{t,k\bullet}^{(i)} = (a_{t,k1}^{(i)}, \dots, a_{t,kK}^{(i)})$ for $i = 1, \dots, p$ and $b_{t,k\bullet}$ defined similarly.

Then the process $(X_t, t \in \mathbb{Z})$ is said to be a generalized binary vector AR process of order p (gbVAR(p)), if it follows the recursion

$$X_t = \sum_{i=1}^p \left[A_t^{(+,i)} X_{t-i} + A_t^{(-,i)} \mathbb{1}_K \right] + B_t e_t, \quad t \in \mathbb{Z}, \tag{2.3}$$

with

$$A_t^{(+,i)} := \left\{ \begin{array}{ll} a_{t,kl}^{(i)}, & \alpha_{kl}^{(i)} \geq 0 \\ -a_{t,kl}^{(i)}, & \alpha_{kl}^{(i)} < 0 \end{array} \right\}_{k,l=1,\dots,K} = \left(a_{t,kl}^{(i)} \left(\mathbb{1}_{\{\alpha_{kl}^{(i)} \geq 0\}} - \mathbb{1}_{\{\alpha_{kl}^{(i)} < 0\}} \right) \right)_{k,l=1,\dots,K},$$

$$A_t^{(-,i)} := \left\{ \begin{array}{ll} 0, & \alpha_{kl}^{(i)} \geq 0 \\ a_{t,kl}^{(i)}, & \alpha_{kl}^{(i)} < 0 \end{array} \right\}_{k,l=1,\dots,K} = \left(a_{t,kl}^{(i)} \mathbb{1}_{\{\alpha_{kl}^{(i)} < 0\}} \right)_{k,l=1,\dots,K}$$

for $i = 1, \dots, p$.

It is worth noting that the Definition 2.1 does not restrict the innovation distribution and explicitly allows for degenerate distributions of e_t with singular Σ_e caused, for example, by $\text{Var}(e_{t,k}) = 0$ for some $k \in \{1, \dots, K\}$ or deterministic relationships between the $e_{t,k}, k = 1, \dots, K$. In the same way, it does not exclude degenerate (joint) distributions of $(X_t, t \in \mathbb{Z})$. Merely, it only imposes stationarity of the process $(X_t, t \in \mathbb{Z})$. A necessary and sufficient conditions ensuring stationarity will be discussed below in Section 2.2. Nevertheless, by not excluding degenerate cases that correspond to parameters lying on the boundary of the parameter space, we encounter identification issues that will be discussed in detail in Section 2.3.

It is possible to rewrite the gbVAR(p) model according to the alternative presentation of the univariate gbAR(1) model in (1.6). However, for the multi-variate case, this becomes cumbersome and the main benefit of the presentation of the gbVAR model in (2.3) is the closed-form expression using an autoregressive-type model equation. Note that X_t is equal to the sum of lagged time series observations X_{t-i} multiplied in a familiar fashion with (random) matrices $A_t^{(+,\cdot)}$, that is, $\sum_{i=1}^p A_t^{(+,i)} X_{t-i}$ (as in VAR-type models), plus an additional term $\sum_{i=1}^p A_t^{(-,i)} \mathbb{1}_K$ (related to negative coefficients) plus an innovation-type term $B_t e_t$.

In the following, we pick up the introductory example from Section 1.2 and consider a bivariate gbVAR(1) model in more detail to illustrate the class of gbVAR models.

Example 2.2 (Bivariate gbVAR(1) model). Let $(X_t, t \in \mathbb{Z})$ follow a two-dimensional gbVAR(1) model

$$X_t = A_t^{(+)} X_{t-1} + A_t^{(-)} \mathbb{1}_K + B_t e_t, \tag{2.4}$$

with parameter matrix $\mathcal{P} = [\mathcal{A}, \mathcal{B}]$ and $\mathcal{P}_{|\cdot|} = [\mathcal{A}_{|\cdot|}, \mathcal{B}]$, where

$$\mathcal{A} = \begin{pmatrix} 0.49 & 0.35 \\ -0.43 & -0.39 \end{pmatrix}, \quad \mathcal{A}_{|\cdot|} = \begin{pmatrix} |0.49| & |0.35| \\ |-0.43| & |-0.39| \end{pmatrix}, \quad \text{and} \quad \mathcal{B} = \begin{pmatrix} 0.16 & 0.00 \\ 0.00 & 0.18 \end{pmatrix},$$

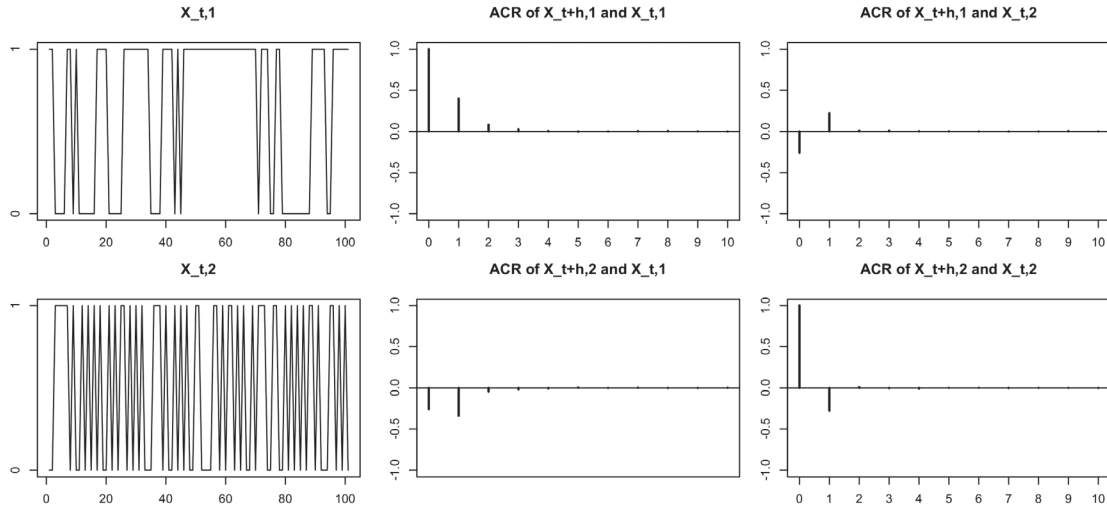


Figure 4. Realization and ACR of the bivariate gbVAR(1) process as specified in Example 2.2

such that $\mathcal{P}_{\cdot|1}\mathbb{1}_4 = \mathbb{1}_2$ holds. Hence, for the (mutually independent) multinomial selection mechanisms, we have

$$P_{t,1\bullet} = [a_{t,11}, a_{t,12}, b_{t,11}, 0] \sim \text{Mult}(1; ([0.49], [0.35], 0.16, 0)),$$

$$P_{t,2\bullet} = [a_{t,21}, a_{t,22}, 0, b_{t,22}] \sim \text{Mult}(1; ([-0.43], [-0.39], 0, 0.18)).$$

Taking the negative signs of the entries in \mathcal{A} into account, the gbVAR(1) process follows the model equation

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} a_{t,11}X_{t-1,1} & +a_{t,12}X_{t-1,2} & +b_{t,11}e_{t,1} \\ a_{t,21}(1 - X_{t-1,1}) & +a_{t,22}(1 - X_{t-1,2}) & +b_{t,22}e_{t,2} \end{pmatrix}.$$

Hence, in the second dimension the opposite values of the predecessors $X_{t-1,1}$ or $X_{t-1,2}$ are selected, whenever $a_{t,21}$ or $a_{t,22}$ become 1 respectively.

Suppose the innovation process $(e_t, t \in \mathbb{Z})$ consists of two independent Bernoulli processes $(e_{t,1}, t \in \mathbb{Z})$ and $(e_{t,2}, t \in \mathbb{Z})$ with $\mu_{e,1} = P(e_{t,1} = 1) = 0.4$ and $\mu_{e,2} = P(e_{t,2} = 1) = 0.8$ leading to $\Sigma_e = \text{diag}(0.24, 0.16)$. In Figure 4, we show a realization of the bivariate gbVAR(1) process with the above specification together with the corresponding serial and cross-sectional autocorrelation structure. By allowing for positive as well as negative entries in the non-diagonal parameter matrix \mathcal{A} , the gbVAR(1) model becomes rather flexible and allows to describe diverse serial and cross-sectional dependence structures.

2.2. Stochastic Properties of gbVAR Models

First, we consider the expectation of the random matrices $A_t^{(+,i)}$ and $A_t^{(-,i)}$, $i = 1, \dots, p$. Note that, by construction, we have $E(A_t^{(i)}) = \mathcal{A}_{\cdot|i}^{(i)}$ and $E(B_t) = \mathcal{B}$. Hence, from the definitions of $A_t^{(+,i)}$ and $A_t^{(-,i)}$, we get

$$E(A_t^{(+,i)}) = [\alpha_{kl}^{(i)}]_{k,l=1,\dots,K} = \mathcal{A}^{(i)}, \tag{2.5}$$

$$E(A_t^{(-,i)}) = [|\alpha_{kl}^{(i)}| \mathbb{1}_{\{\alpha_{kl}^{(i)} < 0\}}]_{k,l=1,\dots,K} =: \mathcal{A}^{(-,i)} \tag{2.6}$$

such that $\mathcal{A}_{\cdot|i}^{(i)} = \mathcal{A}^{(i)} + 2\mathcal{A}^{(-,i)}$. This enables us to compute the stationary mean vector $\mu_X := E(X_t)$ of a gbVAR(p) process.

Lemma 2.3 (Stationary mean of gbVAR processes). Let $(X_t, t \in \mathbb{Z})$ be a stationary K -dimensional gbVAR(p) process. Then, we have

$$\mu_X = \left(I - \sum_{i=1}^p \mathcal{A}^{(i)} \right)^{-1} \left(\sum_{j=1}^p \mathcal{A}^{(-j)} \mathbb{1}_K + B \mu_e \right). \tag{2.7}$$

The latter result reflects the relationship between the mean vector of the time series μ_X , the mean vector μ_e of the innovation process and the autoregressive parameters $\mathcal{P} = [\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}, B]$. In comparison to the univariate NDAR(p) process (see e.g. Weiß (2009a)), additional matrices $\mathcal{A}^{(-\cdot)}$ appear in the formula for the mean that correspond to potentially negative model parameters. Further note that, in contrast to univariate gbAR(p) processes discussed in Jentsch and Reichmann (2019), we do not get $\mu_X = \mu_e$ in the special case when the parameter matrices in \mathcal{P} contain exclusively non-negative entries, such that all $\mathcal{A}^{(-j)}$ vanish. This is due to the diagonal structure of B leading to $I - \sum_{i=1}^p \mathcal{A}^{(i)} \neq B$ in general.

In view of Definition 2.1, which supposes the gbVAR(p) process $(X_t, t \in \mathbb{Z})$ to be *stationary* and to *fulfill the gbVAR recursion* (2.3), its stationary solution can be derived in form of a moving-average-type gbVMA(∞) process. As for classical AR processes, the case $p = 1$ allows for a direct approach to construct the moving-average representation by recursively plugging-in the gbVAR(1) model equation. For all $d \in \mathbb{N}_0$, by recursively plugging-in (2.3), we get

$$\begin{aligned} X_t &= A_t^{(+)} X_{t-1} + A_t^{(-)} \mathbb{1}_K + B_t e_t \\ &= A_t^{(+)} \left(A_{t-1}^{(+)} X_{t-2} + A_{t-1}^{(-)} \mathbb{1}_K + B_{t-1} e_{t-1} \right) + A_t^{(-)} \mathbb{1}_K + B_t e_t \\ &\vdots \\ &= \prod_{j=0}^d A_{t-j}^{(+)} X_{t-(d+1)} + \sum_{i=0}^d \left(\prod_{j=0}^{i-1} A_{t-j}^{(+)} \right) \left(A_{t-i}^{(-)} \mathbb{1}_K + B_{t-i} e_{t-i} \right) \\ &= \zeta_{t,d} X_{t-(d+1)} + \sum_{i=0}^d \zeta_{t,i-1} \eta_{t-i}, \\ &= \zeta_{t,d} X_{t-(d+1)} + X_t^{(d)} \end{aligned} \tag{2.8}$$

with an obvious notation for $X_t^{(d)}$, where $\zeta_{t,i} := \prod_{j=0}^i A_{t-j}^{(+)}$, $i \in \mathbb{N}_0$, $\zeta_{t,-1} = I_K$, and $\eta_{t-i} := A_{t-i}^{(-)} \mathbb{1}_K + B_{t-i} e_{t-i}$. For gbVAR(p) processes of general order $p \in \mathbb{N}$, we follow the common approach described, for example, in Lütkepohl (2005, Chap. 11.3.2) to rewrite the K -dimensional gbVAR(p) process as a Kp -dimensional gbVAR(1) process $(\tilde{X}_t, t \in \mathbb{Z})$. Precisely, by defining the Kp -dimensional vectors

$$\tilde{X}_t := (X_t', \dots, X_{t-p+1}')' \quad \text{and} \quad \tilde{e}_t := (e_t', 0, \dots, 0)' \tag{2.9}$$

and the $(Kp \times Kp)$ -dimensional matrices

$$\begin{aligned} \tilde{A}_t^{(+)} &:= \begin{pmatrix} A_t^{(+,1)} & \dots & A_t^{(+,p-1)} & A_t^{(+,p)} \\ I_K & & 0_{K \times K} & 0_{K \times K} \\ & \ddots & & \vdots \\ 0_{K \times K} & \dots & I_K & 0_{K \times K} \end{pmatrix}, & \tilde{A}_t^{(-)} &:= \begin{pmatrix} A_t^{(-,1)} & \dots & A_t^{(-,p-1)} & A_t^{(-,p)} \\ 0_{K \times K} & & 0_{K \times K} & 0_{K \times K} \\ & \ddots & & \vdots \\ 0_{K \times K} & \dots & 0_{K \times K} & 0_{K \times K} \end{pmatrix}, \\ \tilde{B}_t &:= \begin{pmatrix} B_t & 0_{K \times K(p-1)} \\ 0_{K(p-1) \times K} & 0_{K(p-1) \times K(p-1)} \end{pmatrix}, \end{aligned}$$

where I_K denotes the K -dimensional unity matrix and $0_{r \times s}$ the $(r \times s)$ -dimensional zero matrix, we get an autoregressive representation for the process $(\tilde{X}_t, t \in \mathbb{Z})$. That is, the K -dimensional gbVAR(p) process $(X_t, t \in \mathbb{Z})$ can be represented as a Kp -dimensional gbVAR(1) process $(\tilde{X}_t, t \in \mathbb{Z})$ as follows

$$\tilde{X}_t = \tilde{A}_t^{(+)}\tilde{X}_{t-1} + \tilde{A}_t^{(-)}\mathbb{1}_{Kp} + \tilde{B}_t\tilde{e}_t, \quad t \in \mathbb{Z}, \tag{2.10}$$

where $\mathbb{1}_{Kp}$ is the one vector of length Kp . Note that the first K entries of $(\tilde{X}_t, t \in \mathbb{Z})$ equal the gbVAR(p) process $(X_t, t \in \mathbb{Z})$. By exploiting the above representation (2.10) of \tilde{X}_t as a gbVAR(1) process, analogous to (2.8), we get

$$\tilde{X}_t = \tilde{\zeta}_{t,d}\tilde{X}_{t-(d+1)} + \sum_{i=0}^d \tilde{\zeta}_{t,i-1}\tilde{\eta}_{t-i} = \tilde{\zeta}_{t,d}\tilde{X}_{t-(d+1)} + \tilde{X}_t^{(d)}, \tag{2.11}$$

with $\tilde{\zeta}_{t,i} := \prod_{j=0}^i \tilde{A}_{t-j}^{(+)}$, $i \in \mathbb{N}_0$, $\tilde{\zeta}_{t,-1} = I_{Kp}$, and $\tilde{\eta}_t := \tilde{A}_{t-1}^{(-)}\mathbb{1}_{Kp} + \tilde{B}_{t-1}\tilde{e}_{t-1}$. Now, to establish moving-average-type representations of gbVAR processes, we have to assure that $\tilde{\zeta}_{t,d} \xrightarrow{P} O_{K \times K}$ and $\tilde{\zeta}_{t,d} \xrightarrow{P} O_{Kp \times Kp}$ holds, if we let $d \rightarrow \infty$ on the right-hand sides of (2.8) and (2.11) respectively. For $p = 1$, a sufficient condition familiar from (causal) vector-valued autoregressive processes is

$$\det(I_K - \mathcal{A}_{|\cdot|}z) \neq 0 \quad \forall z \in \mathbb{C} : |z| \leq 1, \tag{2.12}$$

and, for general $p \in \mathbb{N}$,

$$\det(I_{Kp} - \tilde{\mathcal{A}}_{|\cdot|}z) \neq 0 \quad \forall z \in \mathbb{C} : |z| \leq 1, \tag{2.13}$$

where $\tilde{\mathcal{A}}_{|\cdot|} = E(\tilde{A}_t)$ and \tilde{A}_t defined as $\tilde{A}_t^{(+)}$ with $A_t^{(+,i)}$ replaced by $A_t^{(i)}$, $i = 1, \dots, p$. Note that (2.13) is equivalent to the condition that all eigenvalues of $\tilde{\mathcal{A}}_{|\cdot|}$ have modulus smaller than one, and to the condition that all roots of the characteristic matrix polynomial lie outside the unit circle, that is,

$$\det(I_K - \mathcal{A}_{|\cdot|}^{(1)}z - \dots - \mathcal{A}_{|\cdot|}^{(p)}z^p) \neq 0 \quad \forall z \in \mathbb{C} : |z| \leq 1. \tag{2.14}$$

The following result extends Theorem 1 in Jentsch and Reichmann (2019) to the multi-variate case.

Theorem 2.4 (Moving-average representation of gbVAR processes). Let $(X_t, t \in \mathbb{Z})$ be a K -dimensional gbVAR(p) process that fulfills (2.3) for all $t \in \mathbb{Z}$ which is equivalent to $(\tilde{X}_t, t \in \mathbb{Z})$ fulfilling (2.10) for all $t \in \mathbb{Z}$.

- (i) If $p = 1$ and condition (2.12) holds, the gbVAR(1) model has a gbVMA(∞)-type representation

$$X_t = \sum_{i=0}^{\infty} \zeta_{t,i-1}\eta_{t-i}, \quad t \in \mathbb{Z}, \tag{2.15}$$

converging in L_1 .

- (ii) If $p \in \mathbb{N}$ and condition (2.14) holds, the gbVAR(p) model has a gbVMA(∞)-type representation

$$X_t = J\tilde{X}_t = J \left(\sum_{i=0}^{\infty} \tilde{\zeta}_{t,i-1}\tilde{\eta}_{t-i} \right), \quad t \in \mathbb{Z}, \tag{2.16}$$

converging in L_1 , where $J := [I_K, 0_{K \times Kp}]$.

In comparison to classical vector autoregressive processes, the moving-average representation of gbVAR processes contains an additional term $\tilde{A}_{t-i}^{(-)}\mathbb{1}_{Kp}$ in $\tilde{\eta}_{t-i} = \tilde{A}_{t-i}^{(-)}\mathbb{1}_{Kp} + \tilde{B}_{t-i}\tilde{e}_{t-i}$ that takes control of potential negative

parameters. If all entries in the parameter matrix $\mathcal{P} = [\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}, \mathcal{B}]$ are non-negative, this additional term vanishes. It is also important to note that the conditions in (2.12)–(2.14) are sufficient, but not necessary for the gbVAR process to be stationary as they rely on the modified coefficient matrices $\mathcal{A}_{|\cdot|}^{(1)}, \dots, \mathcal{A}_{|\cdot|}^{(p)}$ instead of $\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}$.

The stationarity condition (2.12) for gbVAR(1) processes is illustrated in the following example.

Example 2.5 (Stationarity of gbVAR(1) models). Let $K = 2$ and consider the bivariate gbVAR(1) process with parameter matrix $\mathcal{P} = [\mathcal{A}, \mathcal{B}]$, where

$$\mathcal{A} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \quad \text{and} \quad \mathcal{B} = \begin{pmatrix} \beta_{11} & 0 \\ 0 & \beta_{22} \end{pmatrix}$$

such that $|\alpha_{11}| + |\alpha_{12}| + \beta_{11} = 1$ and $|\alpha_{21}| + |\alpha_{22}| + \beta_{22} = 1$. Let us consider four cases:

- (i) If $\beta_{11}, \beta_{22} > 0$, both innovations $e_{t,1}$ and $e_{t,2}$ are allowed to enter the gbVAR(1) model. From $|\alpha_{11}| + |\alpha_{12}| < 1$ and $|\alpha_{21}| + |\alpha_{22}| < 1$, we get that all eigenvalues of $\mathcal{A}_{|\cdot|}$ and of \mathcal{A} have modulus smaller than one. Hence, condition (2.12) holds.
- (ii) If $\beta_{11} = 0, \beta_{22} > 0$ and $\alpha_{12} \neq 0$, only the innovation $e_{t,2}$ is allowed to enter the gbVAR(1) model, but as $\alpha_{12} \neq 0$, it may reach $X_{t,1}$ after one time step as well. Indeed, in this case, it can be checked that all eigenvalues of $\mathcal{A}_{|\cdot|}$ and of \mathcal{A} have modulus smaller than one and condition (2.12) holds.
- (iii) If $\beta_{11} = 0, \beta_{22} > 0$ and $\alpha_{12} = 0$, only the innovation $e_{t,2}$ is allowed to enter the gbVAR(1) model, but it will never reach $\{X_{t,1}\}$. In this case, where $|\alpha_{11}| = 1 - |\alpha_{12}| - \beta_{11} = 1$, the largest (in modulus) eigenvalues of both matrices $\mathcal{A}_{|\cdot|}$ and \mathcal{A} are equal to one such that condition (2.12) does not hold. Indeed, this scenario leads to a reducible Markov chain (of order one) with equivalence classes $\{(0, 0), (0, 1)\}$ and $\{(1, 0), (1, 1)\}$.
- (iv) If $\beta_{11} = \beta_{22} = 0$ and $\alpha_{ij} \neq 0, i, j = 1, 2$, such that

$$\mathcal{A} = \begin{pmatrix} \alpha_{11} & 1 - \alpha_{11} \\ \alpha_{22} & -(1 - \alpha_{21}) \end{pmatrix}, \quad \mathcal{A}_{|\cdot|} = \begin{pmatrix} \alpha_{11} & 1 - \alpha_{11} \\ \alpha_{21} & 1 - \alpha_{21} \end{pmatrix}, \quad \text{and} \quad \mathcal{B} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

no innovations at all are allowed to enter the gbVAR(1) model. In this case, the eigenvalues of \mathcal{A} are (in modulus) smaller than one and this bivariate gbVAR(1) model describes an irreducible, aperiodic Markov chain (of order one) with states $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$. For example, if $\alpha_{11} = \alpha_{21} = 0$, the eigenvalues of \mathcal{A} compute to $\lambda_{1,2} = \pm 0.7071$ leading to a Markov chain with transition matrix P and stationary distribution π according to

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix} \quad \text{and} \quad \pi = \begin{pmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{pmatrix}.$$

Nevertheless, the largest eigenvalue (in modulus) of $\mathcal{A}_{|\cdot|}$ is always one, such that condition (2.12) does not hold.

The latter example illustrates that $\beta_{kk} > 0$ for all $k = 1, \dots, K$ is a sufficient, but not a necessary condition for (2.12) (and also for (2.14)) to hold.

For the derivation of asymptotic theory, (strong) mixing concepts are very helpful to quantify the serial dependence structure of time series processes; see, for example Doukhan (1994), Dedecker *et al.* (2007), or Bradley (2005) for overviews. As given in Billingsley (1968), when defined on a suitable probability space (Ω, \mathcal{A}, P) , a process $(Z_t, t \in \mathbb{Z})$ is called ψ -mixing, if for all subsets $\mathcal{E}_1 \in \sigma(Z_t, Z_{t-1}, \dots)$ and $\mathcal{E}_2 \in \sigma(Z_{t+h}, Z_{t+h+1}, \dots)$ of the

induced σ -fields, and a non-negative sequence $(f_h, h \in \mathbb{N})$ with $f_h \rightarrow 0$ for $h \rightarrow \infty$, we have

$$\left| P(\mathcal{E}_1 \cap \mathcal{E}_2) - P(\mathcal{E}_1)P(\mathcal{E}_2) \right| \leq f_h P(\mathcal{E}_1)P(\mathcal{E}_2). \tag{2.17}$$

If the right-hand side of the inequality (2.17) is replaced by $f_h P(\mathcal{E}_1)$, we get the definition of the φ -mixing property. If $f_h \leq \text{const.}\rho^h$ for all $h \in \mathbb{N}$, the process $(Z_t, t \in \mathbb{Z})$ is called geometrically ψ -mixing or φ -mixing respectively.

In contrast to Jacobs and Lewis (1983) or Weiß (2009a), who employ a suitable Markov chain representation and take a detour by showing primitivity first to deduce (strong) mixing properties, we can make direct use of the moving-average-type representation derived in Theorem 2.4 for gbVAR(p) processes to prove (geometric) ψ - and φ -mixing.

Theorem 2.6 (Mixing properties of gbVAR processes). Let $(X_t, t \in \mathbb{Z})$ be a K -dimensional gbVAR(p) process that fulfills (2.3) for all $t \in \mathbb{Z}$ which is equivalent to $(\tilde{X}_t, t \in \mathbb{Z})$ fulfilling (2.10) for all $t \in \mathbb{Z}$ and suppose that (2.13) holds. Then, the processes $(\tilde{X}_t, t \in \mathbb{Z})$ and $(X_t, t \in \mathbb{Z})$ are ergodic and geometrically ψ - and φ -mixing.

Note that gbVAR processes are of autoregressive-type, but they are *nonlinear* due to the *random* coefficient matrices. Nevertheless, we can show that the autocovariance structure of gbVAR processes coincides with that of classical VAR processes in the sense that the same Yule–Walker equations hold. For this purpose, we denote by $\Gamma_X(h) = \text{Cov}(X_{t+h}, X_t)$, $h \in \mathbb{Z}$, the corresponding autocovariance matrices of the gbVAR(p) process $(X_t, t \in \mathbb{Z})$.

Theorem 2.7 (Yule–Walker equations for gbVAR(p) models, $h > 0$). Let $(X_t, t \in \mathbb{Z})$ be a K -dimensional gbVAR(p) process that fulfills (2.14). Then, for all $h \in \mathbb{N}$ (with $h \neq 0$), we have

$$\Gamma_X(h) = \sum_{i=1}^p \mathcal{A}^{(i)} \Gamma_X(h-i) \tag{2.18}$$

leading to the system of Yule–Walker equations

$$[\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}] \begin{pmatrix} \Gamma_X(i-j) \\ i, j = 1, \dots, p \end{pmatrix} = [\Gamma_X(1), \dots, \Gamma_X(p)]. \tag{2.19}$$

By replacing the autocovariances $\Gamma_X(h)$ by sample versions $\hat{\Gamma}_X(h)$, Yule–Walker equations can be used for parameter estimation using the well-known Yule–Walker estimators. Before these will be addressed in Section 2.4, we will discuss possible identification issues in Section 2.3.

The derivation of a Yule–Walker-type equation for $h = 0$ is much more intricate. For this purpose, we will use Hadamard products denoted by ‘ \circ ’, where $A \circ B := (a_{ij}b_{ij})_{i,j}$ for two matrices A and B of the same dimensions.

Theorem 2.8 (Yule–Walker equation for gbVAR(p) models, $h = 0$). Let $(X_t, t \in \mathbb{Z})$ be a K -dimensional gbVAR(p) process that fulfills (2.14). Then, for $h = 0$, we have

$$\begin{aligned} \Gamma_X(0) &= \sum_{i,j=1}^p \left[I_K \circ \left\{ \left(\mathcal{A}_{\cdot i}^{(i)} \mu_X \mathbb{1}'_K \right) \mathbb{1}_{\{i=j\}} - \left(\mathcal{A}^{(i)} (\Gamma_X(j-i) + \mu_X \mu_X') \mathcal{A}^{(j)'} \right) \right\} + \left(\mathcal{A}^{(i)} (\Gamma_X(j-i)) \mathcal{A}^{(j)'} \right) \right] \\ &+ \sum_{i,j=1}^p \left[I_K \circ \left\{ \left(-\mathcal{A}^{(-i)} \mu_X \mathbb{1}'_K \right) \mathbb{1}_{\{i=j\}} - \left(\mathcal{A}^{(i)} \mu_X \mathbb{1}'_K \mathcal{A}^{(-j)'} \right) \right\} \right] \\ &- \sum_{i=1}^p \left[I_K \circ \left\{ \left(\mathcal{A}^{(i)} \mu_X \mu_e' \mathcal{B}' \right) + \left(\mathcal{A}^{(-i)} \mathbb{1}_K \mu_e' \mathcal{B}' \right) \right\} \right] \\ &+ \sum_{i,j=1}^p \left[I_K \circ \left\{ \left(-\mathcal{A}^{(-i)} \mu_X \mathbb{1}'_K \right) \mathbb{1}_{\{i=j\}} - \left(\mathcal{A}^{(-i)} \mathbb{1}_K \mu_X' \mathcal{A}^{(j)'} \right) \right\} \right] \end{aligned}$$

$$\begin{aligned}
 & + \sum_{i,j=1}^p \left[I_K \circ \left\{ \left(\mathcal{A}^{(-i)} \mathbb{1}_K \mathbb{1}'_K \right) \mathbb{1}_{\{i=j\}} - \left(\mathcal{A}^{(-i)} \mathbb{1}_K \mathbb{1}'_K \mathcal{A}^{(-j)'} \right) \right\} \right] \\
 & - \sum_{j=1}^p \left[I_K \circ \left\{ \left(\mathcal{B} \mu_e \mu'_X \mathcal{A}^{(j)'} \right) + \left(\mathcal{B} \mu_e \mathbb{1}'_K \mathcal{A}^{(-j)'} \right) \right\} \right] \\
 & + I_K \circ \left\{ \mathcal{B} \mu_e \mathbb{1}'_K - \mathcal{B} \mu_e \mathbb{1}'_K \mathcal{B} \right\} + \mathcal{B} \Sigma_e \mathcal{B}.
 \end{aligned} \tag{2.20}$$

Note that the formula derived in Theorem 2.8 contains the expression

$$\Gamma_X(0) = \sum_{i,j=1}^p \left(\mathcal{A}^{(i)} \left(\Gamma_X(j-i) \right) \mathcal{A}^{(j)'} \right) + \mathcal{B} \Sigma_e \mathcal{B},$$

which is similar to the classical formula for the Yule–Walker equation for $h = 0$ of VAR(p) processes (see e.g. Lütkepohl, 2005, Section 11.4, here only $\mathcal{B} \Sigma_e \mathcal{B}$ is replaced by Σ_e) plus some additional terms that contain “ $I_K \circ \dots$ ”. Note that these additional terms *only* adjust the diagonal entries. They show up due to the random coefficients, that control the selection mechanism in the gbVAR model.

Alternatively, the last line in (2.20) can be rearranged to get

$$\begin{aligned}
 & I_K \circ \left\{ \mathcal{B} \mu_e \mathbb{1}'_K + \mathcal{B} (\Sigma_e - \mu_e \mathbb{1}'_K) \mathcal{B} \right\} + \left(\mathbb{1}_{K \times K} - I_K \right) \circ \left\{ \mathcal{B} \Sigma_e \mathcal{B} \right\} \\
 & = I_K \circ \left\{ \mathcal{B} \mu_e \mathbb{1}'_K - \mathcal{B} \left(\mu_e \mu'_e \right) \mathcal{B} \right\} + \mathcal{B} \left\{ \left(\mathbb{1}_{K \times K} - I_K \right) \circ \Sigma_e \right\} \mathcal{B}.
 \end{aligned} \tag{2.21}$$

Note that the Hadamard multiplication in the last term on the right-hand side sets the diagonal of Σ_e (which is already determined by μ_e due to $\sigma_{e,ii} = \mu_{e,i}(1 - \mu_{e,i})$) to zero and that the first term on the right-hand side of (2.21) contains only \mathcal{B} and μ_e , but not Σ_e . This is particularly useful to identify the off-diagonal elements of Σ_e , which can be used for estimation purposes as discussed in Section 2.4.

Example 2.9 (Special cases of Theorem 2.8). For $p = 1$, the formula derived in Theorem 2.8 simplifies to become

$$\begin{aligned}
 \Gamma_X(0) & = I_K \circ \left\{ \left(\mathcal{A}_{|\cdot|}^{(1)} \mu_X \mathbb{1}'_K \right) - \left(\mathcal{A}^{(1)} \left(\Gamma_X(0) + \mu_X \mu'_X \right) \mathcal{A}^{(1)'} \right) \right\} + \left(\mathcal{A}^{(1)} \Gamma_X(0) \mathcal{A}^{(1)'} \right) \\
 & + \left[I_K \circ \left\{ \left(- \mathcal{A}^{(-,1)} \mu_X \mathbb{1}'_K \right) - \left(\mathcal{A}^{(1)} \mu_X \mathbb{1}'_K \mathcal{A}^{(-,1)'} \right) \right\} \right] - \left[I_K \circ \left\{ \mathcal{A}^{(1)} \mu_X \mu'_e \mathcal{B}' \right\} \right] \\
 & + \left[I_K \circ \left\{ - \mathcal{A}^{(-,1)} \mu_X \mathbb{1}'_K - \left(\mathcal{A}^{(-,1)} \mathbb{1}_K \mu'_X \mathcal{A}^{(1)'} \right) \right\} \right] \\
 & + \left[I_K \circ \left\{ \left(\mathcal{A}^{(-,1)} \mathbb{1}_K \mathbb{1}'_K \right) - \mathcal{A}^{(-,1)} \mathbb{1}_K \mathbb{1}'_K \mathcal{A}^{(-,1)'} \right\} \right] \\
 & - \left[I_K \circ \left\{ \mathcal{A}^{(-,1)} \mathbb{1}_K \mu'_e \mathcal{B}' \right\} \right] - \left[I_K \circ \left\{ \mathcal{B} \mu_e \mu'_X \mathcal{A}^{(1)'} \right\} \right] - \left[I_K \circ \left\{ \mathcal{B} \mu_e \mathbb{1}'_K \mathcal{A}^{(-,1)'} \right\} \right] \\
 & + I_K \circ \left\{ \mathcal{B} \mu_e \mathbb{1}'_K - \mathcal{B} \left(\mu_e \mathbb{1}'_K \right) \mathcal{B} \right\} + \mathcal{B} \Sigma_e \mathcal{B}.
 \end{aligned}$$

If additionally all entries in $\mathcal{A}^{(1)}$ are non-negative, all terms containing $\mathcal{A}^{(-,1)} = 0_{K \times K}$ vanish and we can replace $\mathcal{A}_{|\cdot|}^{(1)}$ by $\mathcal{A}^{(1)}$. This leads to

$$\begin{aligned}
 \Gamma_X(0) & = I_K \circ \left\{ \left(\mathcal{A}^{(1)} \mu_X \mathbb{1}'_K \right) - \left(\mathcal{A}^{(1)} \left(\Gamma_X(0) + \mu_X \mu'_X \right) \mathcal{A}^{(1)'} \right) \right\} + \left(\mathcal{A}^{(1)} \Gamma_X(0) \mathcal{A}^{(1)'} \right) \\
 & - \left[I_K \circ \left\{ \mathcal{A}^{(1)} \mu_X \mu'_e \mathcal{B}' \right\} \right] - \left[I_K \circ \left\{ \mathcal{B} \mu_e \mu'_X \mathcal{A}^{(1)'} \right\} \right] \\
 & + I_K \circ \left\{ \mathcal{B} \mu_e \mathbb{1}'_K - \mathcal{B} \left(\mu_e \mathbb{1}'_K \right) \mathcal{B} \right\} + \mathcal{B} \Sigma_e \mathcal{B}.
 \end{aligned} \tag{2.22}$$

In the following result, we derive expressions for one step ahead transition probabilities for gbVAR processes to reach a certain state in $\{0, 1\}^K$ given the past values of the time series.

Lemma 2.10 (Transition probabilities of gbVAR processes). Let $(X_t, t \in \mathbb{Z})$ be a K -dimensional gbVAR(p) process. Furthermore, for $k = 0, 1, \dots, p$, let $s_k, r_0 \in \{0, 1\}^K$ with $s_k := (s_{k,1}, \dots, s_{k,K})'$ and $r_0 := (r_{0,1}, \dots, r_{0,K})'$. Denote by $\delta_{ij} = \mathbb{1}_{\{i=j\}}$ the Kronecker delta and set $p_{r_0} := P(e_t = r_0)$. Then, the transition probability given the past p values of the time series becomes

$$\begin{aligned}
 &P(X_t = s_0 | X_{t-1} = s_1, \dots, X_{t-p} = s_p) \tag{2.23} \\
 &= \sum_{r_0 \in \{0,1\}^K} p_{r_0} \prod_{k=1}^K \left[\sum_{i=1}^p \sum_{l=1}^K |\alpha_{kl}^{(i)}| \left[\mathbb{1}_{\{\alpha_{kl}^{(i)} \geq 0\}} \delta_{s_{0,k} s_{i,l}} + \mathbb{1}_{\{\alpha_{kl}^{(i)} < 0\}} \delta_{s_{0,k} (1-s_{i,l})} \right] + \beta_{kk} \delta_{s_{0,k} r_{0,k}} \right].
 \end{aligned}$$

Remark 2.11 (Dependent multinomial selection). In view of the mutually independent multinomial selection mechanisms $P_{t,k}, k = 1, \dots, K$, used to define gbVAR processes in Definition 2.1, a possible extension would be to allow for dependence between the multinomial distributions. Such an extension would not affect the Yule–Walker equations for $h > 0$ in Theorem 2.7, but those in Theorem 2.8 for $h = 0$. However, we do not follow this path as it would complicate things considerably and the benefit of such an extension of the gbVAR model class would be comparatively small. Furthermore, multinomial distributions allowing for dependence seem to be less developed; see, for example Johnson *et al.* (1997, chap. 36) for a rather restrictive attempt in this direction.

Note that the scalar selection mechanism used for the GDARMA approach proposed in Möller and Weiß (2020) coincides with a setup of mutually dependent multinomial selection mechanisms with perfectly correlated coefficient matrices’ diagonal entries.

2.3. Identification of gbVAR Models

As discussed in Theorem 2.4, a gbVAR(p) process that fulfills condition (2.14) has a (causal) L_1 -converging gbVMA(∞)-type representation (2.16). Moreover, Theorem 2.7 shows that standard Yule–Walker equations hold for $h \in \mathbb{N}$, that relate $(\Gamma_X(h), h \in \mathbb{Z})$ and $\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}$, and Theorem 2.8 indicates a Yule–Walker-type equation for $h = 0$ that also involves $\mu_X, \mu_e, \mathcal{B}$ and Σ_e . However, the stationarity condition (2.14) alone does not guarantee identifiability of the gbVAR(p) model, that is, identification of model parameters $\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}, \mathcal{B}$ and Σ_e , from the autocovariance function $(\Gamma(h), h \in \mathbb{Z})$ (and the mean vector μ_X) via Yule–Walker equations.

For given $p \in \mathbb{N}$, using Theorem 2.7, the gbVAR(p) parameter matrices $\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}$ are identified from the autocovariance function (ACF) if and only if the Yule–Walker matrix

$$\Gamma_{\tilde{X},p} := Cov(\tilde{X}_t) = \begin{pmatrix} \Gamma_X(i-j) \\ i, j = 1, \dots, p \end{pmatrix} \tag{2.24}$$

is *non-singular*, where \tilde{X}_t is defined in (2.9), such that we can re-arrange (2.19) to get

$$[\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}] = [\Gamma_X(1), \dots, \Gamma_X(p)] \begin{pmatrix} \Gamma_X(i-j) \\ i, j = 1, \dots, p \end{pmatrix}^{-1}. \tag{2.25}$$

Consequently, \mathcal{B} is also identified by imposed diagonality of \mathcal{B} and the requirement that $\mathcal{P} \mathbb{1}_{K(p+1)} = \mathbb{1}_K$ has to hold. Similarly, using Theorem 2.8, the innovation covariance matrix Σ_e is identified if and only if \mathcal{B} is *non-singular* which is equivalent to $\beta_{kk} > 0$ for all $k = 1, \dots, K$. In this case, we can re-arrange (2.20) to isolate Σ_e on one side of the equation by making use of \mathcal{B}^{-1} .

Now, let us discuss the case of a *singular* Yule–Walker matrix $\Gamma_{\tilde{X}_p}$. It is singular if and only if there exists vectors $g \in \mathbb{R}^{Kp}$, $g \neq 0_{Kp}$, such that $g' \tilde{X}_t$ is constant. As \tilde{X}_t takes values in $\{0, 1\}^{Kp}$, such a vector g can always be chosen from $\{-1, 0, 1\}^{Kp} \setminus \{0_{Kp}\}$ leading to $g' \tilde{X}_t$ either equal to 0 or equal to 1 for all t . For example, if one component $(X_{t,k}, t \in \mathbb{Z})$ is constant over time, we have that $e'_k X_t$ is constant in $\{0, 1\}$, where e_k is the k th K -dimensional unit vector. Hence, for general $p \in \mathbb{N}$ and as \tilde{X}_{t-1} contains $X_{t-1,k}, \dots, X_{t-p,k}$, we can find p many linearly independent vectors g that have only one entry equal to one and all others equal to zero leading to $g' \tilde{X}_t$ constant in $\{0, 1\}$. If, for some $h = 0, 1, \dots, p-1$, we have $X_{t,i} = X_{t-h,j}$ or $X_{t,i} = 1 - X_{t-h,j}$ for some $i, j = 1, \dots, K$, we can find a vector g having two non-zero entries such that $g' \tilde{X}_t$ is constant equal to 0 or 1 for all t respectively.

Suppose, we can find $L \leq Kp$ linearly independent vectors g_1, \dots, g_L with the property $g' \tilde{X}_t$ constant. If $L = Kp$, we are in the purely deterministic case. If $L < Kp$, the underlying gbVAR(p) model is partially identified. After having identified L such linearly independent vectors g_1, \dots, g_L , we can find also vectors $g_{L+1}, \dots, g_{Kp} \in \{-1, 0, 1\}^{Kp}$ such that the $Kp \times Kp$ matrix G with $G' = [G'_1, G'_2] = [g_1, \dots, g_L, g_{L+1}, \dots, g_{Kp}]$ is of full rank Kp . Moreover, the g_{L+1}, \dots, g_{Kp} can be chosen such that they have only one non-zero entry equal to 1 respectively. Together, we have

$$G\tilde{X}_t = \begin{pmatrix} v_L \\ \tilde{Y}_t \end{pmatrix} \quad \text{and} \quad \tilde{X}_t = G^{-1} \begin{pmatrix} v_L \\ \tilde{Y}_t \end{pmatrix} \tag{2.26}$$

where v_L is a constant vector with 0 and 1 entries and \tilde{Y}_t is $(Kp - L)$ -dimensional with $Cov(\tilde{Y}_t)$ having full rank. Plugging-in for \tilde{X}_t in the Yule–Walker matrix in (2.19) and multiplication from the right with $((G^{-1}H')')^+$, where $H = (0_{(Kp-L) \times (L)}, I_{(Kp-L) \times (Kp-L)})$ and $((G^{-1}H')')^+$ denotes the $(Kp - L) \times Kp$ Moore–Penrose inverse of $(G^{-1}H')'$, leads to

$$\begin{aligned} [\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}]Cov\left(\tilde{X}_t\right)((G^{-1}H')')^+ &= [\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}]G^{-1}H'Cov\left(\tilde{Y}_t\right)(G^{-1}H')'((G^{-1}H')')^+ \\ &= [\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}]G^{-1}H'Cov\left(\tilde{Y}_t\right) \\ &= [\Gamma_X(1), \dots, \Gamma_X(p)]((G^{-1}H')')^+. \end{aligned}$$

By defining $\Gamma_{\tilde{Y}_p} = Cov(\tilde{Y}_t)$ and $\mathbf{A}_{\tilde{Y}_p} = [\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}]G^{-1}H'$, the dimension-reduced $K \times (Kp - L)$ parameter matrix $\mathbf{A}_{\tilde{Y}_p}$ is identified by

$$\mathbf{A}_{\tilde{Y}_p} = [\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}]G^{-1}H' = [\Gamma_X(1), \dots, \Gamma_X(p)]((G^{-1}H')')^+ \Gamma_{\tilde{Y}_p}^{-1}. \tag{2.27}$$

Finally, we can obtain one candidate of $[\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}]$ by right multiplying $\mathbf{A}_{\tilde{Y}_p}$ with $(G^{-1}H')'$ to get a $K \times Kp$ parameter matrix.

In the case, where the matrix \mathcal{B} is not invertible, which corresponds to at least one diagonal entry $\beta_{kk} = 0$ for some $k = 1, \dots, K$, the innovation distribution of those $e_{t,k}$ is not identifiable. However, we can rearrange (2.20) using (2.21) to isolate $\mathcal{B}\{(\mathbb{1}_{K \times K} - I_K) \circ \Sigma_e\} \mathcal{B}$ on one side of the equation. Now, let $M \leq K$ be the number of zero diagonal elements of \mathcal{B} . If $M = K$, we are in the special case (see e.g. Example 2.5(iv)), where no innovations enter the model at all and the distribution of e_t is clearly not identifiable.

If $M < K$, we define a $(K - M) \times K$ matrix F that picks those components $k \in \{1, \dots, K\}$, where the diagonal elements β_{kk} of \mathcal{B} are non-zero such that FBF' has full rank $K - M$. Then, isolating $\mathcal{B}\mu_e$ on one side of the equation in (2.7) and left-multiplication with F and inserting $F'F$ leads to

$$F \left(\left(I - \sum_{i=1}^p \mathcal{A}^{(i)} \right) \mu_X - \sum_{j=1}^p \mathcal{A}^{(-j)} \mathbb{1}_K \right) = (FBF')(F\mu_e) = \mathcal{B}_e \mu_e \tag{2.28}$$

with an obvious notation for $\mathcal{B}_{\tilde{e}}$ and $\mu_{\tilde{e}}$. Since $\mathcal{B}_{\tilde{e}}$ has full rank, we can left-multiply the whole equation with $\mathcal{B}_{\tilde{e}}^{-1}$ to identify $\mu_{\tilde{e}}$, which is the mean vector of \tilde{e} which is defined to consist of those $e_{t,k}$'s, where the corresponding β_{kk} 's are non-zero. This allows also to identify the diagonal elements of the covariance matrix $\Sigma_{\tilde{e}}$ of \tilde{e} . Furthermore, after re-arranging (2.20) using (2.21) to isolate $\mathcal{B}\{(\mathbb{1}_{K \times K} - I_K) \circ \Sigma_e\} \mathcal{B}$ on one side of the equation, left-multiplication with F , right-multiplication with F' and inclusion of $F'F$ twice, we can get on one side of the equation

$$\begin{aligned} (FBF')(F\{(\mathbb{1}_{K \times K} - I_K) \circ \Sigma_e\}F')(FBF') &= (FBF')((\mathbb{1}_{(K-M) \times K} \circ F)\{(\mathbb{1}_{K \times K} - I_K) \circ \Sigma_e\}(\mathbb{1}_{K \times (K-M)} \circ F'))(FBF') \\ &= (FBF')(\{\mathbb{1}_{(K-M) \times K}(\mathbb{1}_{K \times K} - I_K)\mathbb{1}_{K \times (K-M)} \circ F \Sigma_e F'\})(FBF') \\ &= (FBF')(\{\mathbb{1}_{(K-M) \times (K-M)} - I_{K-M}\} \circ F \Sigma_e F')(FBF') \\ &= \mathcal{B}_{\tilde{e}}(\{\mathbb{1}_{(K-M) \times (K-M)} - I_{K-M}\} \circ \Sigma_{\tilde{e}}) \mathcal{B}_{\tilde{e}}. \end{aligned} \tag{2.29}$$

Again, since $\mathcal{B}_{\tilde{e}}$ has full rank, we can left- and right-multiply the whole resulting equation with $\mathcal{B}_{\tilde{e}}^{-1}$ to identify also the non-diagonal elements of $\Sigma_{\tilde{e}}$ of those $e_{t,k}$. Note that this identification is indeed sufficient, as the innovations $e_{t,k}$ that are not identified, do not enter the gbVAR(p) at all.

The identification issues discussed above are illustrated in the following example.

Example 2.12 (gbVAR(2) with singular Yule–Walker matrix). Suppose $(X_t, t \in \mathbb{Z})$ is a bivariate gbVAR(2) process with

$$\mathcal{A}_1 = \begin{pmatrix} \alpha_{11}^{(1)} & \alpha_{12}^{(1)} \\ 1 & 0 \end{pmatrix}, \quad \mathcal{A}_2 = \begin{pmatrix} \alpha_{11}^{(2)} & \alpha_{12}^{(2)} \\ 0 & 0 \end{pmatrix}, \quad \text{and} \quad \mathcal{B} = \begin{pmatrix} \beta_{11} & 0 \\ 0 & 0 \end{pmatrix}$$

where $\alpha_{11}^{(1)}, \alpha_{12}^{(1)}, \alpha_{11}^{(2)}, \alpha_{12}^{(2)} \geq 0$ and $\beta_{11} > 0$ such that $\alpha_{11}^{(1)} + \alpha_{12}^{(1)} + \alpha_{11}^{(2)} + \alpha_{12}^{(2)} + \beta_{11} = 1$ and with $E(e_t) = \mu_e$ and $\text{Var}(e_t) = \Sigma_e$. Then, we are facing two identification issues. First, $e_{t,2}$ does not enter the model at all such that only the distribution of the innovation $e_{t,1}$, that is, $E(e_{t,1}) = \mu_{e,1}$ and $\text{Var}(e_{t,1}) = \Sigma_{e,11} = \mu_{e,1}(1 - \mu_{e,1})$ are identifiable as discussed in (2.28) and (2.29). Second, by construction, the process $(X_t, t \in \mathbb{Z})$ fulfills $X_{t,2} = X_{t-1,1}$ for all $t \in \mathbb{Z}$. Hence, we can choose $g_1 = (0, 1, -1, 0)'$ leading to $g_1' \tilde{X}_t = 0$, where $\tilde{X}_t = (X_t', X_{t-1}')' = (X_{t,1}, X_{t,2}, X_{t-1,1}, X_{t-1,2})'$. As this is the only vector g_1 (up to scale) with this property, we have $L = 1$ in the notation described above. Defining also $g_2 = (1, 0, 0, 0)'$, $g_3 = (0, 1, 0, 0)'$ and $g_4 = (0, 0, 0, 1)$ and $G' = [g_1, g_2, g_3, g_4]$ we get a lower-dimensional process \tilde{Y}_t of dimension $Kp - L = 3$ as in (2.26) with non-singular covariance matrix $\text{Cov}(\tilde{Y}_t)$ leading to a (2×3) -dimensional coefficient matrix $\mathbf{A}_{\tilde{Y},p}$ as in (2.27), where $H = [0_{3 \times 1}, I_3]$ and $p = 2$. This leads to

$$(G^{-1}H')' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad ((G^{-1}H')')^+ = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and $\mathbf{A}_{\tilde{Y},p} = [\mathcal{A}_1, \mathcal{A}_2]G^{-1}H'$.

Now we address why diagonality of \mathcal{B} has to be imposed to achieve identifiability of Σ_e .

Remark 2.13 (Non-diagonal \mathcal{B} matrix). The parameter matrices $\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}$ can be easily estimated using Yule–Walker estimators as will be described in Section 2.4 such that, by imposing a diagonal structure of \mathcal{B} , that is, $\mathcal{B} = \text{diag}(\beta_{11}, \dots, \beta_{KK})$, it is straightforward to define $\beta_{kk} := 1 - \sum_{i=1}^p \sum_{l=1}^K |\alpha_{kl}^{(i)}|$, for $k = 1, \dots, K$. If the diagonality of \mathcal{B} is not enforced, it is unclear how to allocate $1 - \sum_{i=1}^p \sum_{l=1}^K |\alpha_{kl}^{(i)}|$ to K free parameters $\beta_{k1}, \dots, \beta_{kK}$. Hence,

allowing \mathcal{B} to be non-diagonal leads to identification issues. For example, let $K = 2$ and suppose \mathcal{B} is potentially non-diagonal with non-negative entries. Then, the bivariate process $\{B_t e_t, t \in \mathbb{Z}\}$ is i.i.d. taking values $s_0 \in \{0, 1\}^2$ and we get

$$\begin{aligned}
 P(B_t e_t = s_0) &= \sum_{r_0 \in \{0,1\}^2} P(B_t e_t = s_0 | e_t = r_0) P(e_t = r_0) \\
 &= \sum_{r_0 \in \{0,1\}^2} \left(\beta_{11} \delta_{r_{0,1} s_{0,1}} + \beta_{12} \delta_{r_{0,2} s_{0,1}} \right) \left(\beta_{21} \delta_{r_{0,1} s_{0,2}} + \beta_{22} \delta_{r_{0,2} s_{0,2}} \right) p_{r_0}.
 \end{aligned}
 \tag{2.30}$$

As the distribution of $B_t e_t$ takes values in $\{0, 1\}^2$, it is fully specified by three parameters. If \mathcal{B} is imposed to be diagonal such that $\beta_{12} = \beta_{21} = 0$ and β_{11} and β_{22} are predetermined (by $\beta_{kk} := 1 - \sum_{i=1}^p \sum_{l=1}^K |\alpha_{kl}^{(i)}|$), the distribution of e_t specified by, for example, $p_{(0,0)}, p_{(0,1)}$ and $p_{(1,0)}$ (with $p_{(1,1)} = 1 - p_{(0,0)} - p_{(0,1)} - p_{(1,0)}$), is identified. If \mathcal{B} is not restricted to be diagonal, this is not the case.

2.4. Parameter Estimation in gbVAR Models

The joint distribution of a gbVAR process is fully determined by the marginal distribution of the i.i.d. innovations $(e_t, t \in \mathbb{Z})$ and by the model parameters in \mathcal{P} . In view of the second-order dependence structure of gbVAR processes, the Yule–Walker equations derived in Theorems 2.7 and 2.8 constitute an important link between the mean vector μ_X , the autocovariance function Γ_X , the gbVAR coefficients in \mathcal{P} , which are all population quantities of the process $(X_t, t \in \mathbb{Z})$, and mean vector μ_e and covariance matrix Σ_e , which are population quantities of the innovation process. Here, μ_X and $\Gamma_X(h)$ are easily estimable from a gbVAR data sample X_1, \dots, X_n by their sample versions

$$\hat{\mu}_X := \bar{X} = \frac{1}{n} \sum_{t=1}^n X_t,
 \tag{2.31}$$

$$\hat{\Gamma}_X(h) := \begin{cases} \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X})(X_t - \bar{X})', & 0 \leq h < n \\ 0, & h \geq n \end{cases}
 \tag{2.32}$$

and $\Gamma_X(-h) := \Gamma_X'(h)$ for $h < 0$. If $\tilde{X}_t = (X_t', X_{t-1}', \dots, X_{t-p+1}')'$ constructed from $(X_t, t \in \mathbb{Z})$ has a non-singular covariance matrix, we can use the Yule–Walker equation system (2.19) to construct the well-known Yule–Walker estimator $[\hat{\mathcal{A}}^{(1)}, \dots, \hat{\mathcal{A}}^{(p)}]$ for $[\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}]$ by replacing the ACF Γ_X by the sample ACF $\hat{\Gamma}_X$. Otherwise, we have to apply the Yule–Walker estimation to the corresponding lower-dimensional process $(\tilde{Y}_t, t \in \mathbb{Z})$ obtained as in Section 2.3.

In the following, we assume that $(X_t, t \in \mathbb{Z})$ has a non-singular Yule–Walker matrix $\Gamma_{\tilde{X},p}$. Then, Yule–Walker estimators (see e.g. Lütkepohl, 2005, eq. (3.3.17)), which are consistent for $[\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}]$, are defined by

$$[\hat{\mathcal{A}}^{(1)}, \dots, \hat{\mathcal{A}}^{(p)}] = [\hat{\Gamma}_X(1), \dots, \hat{\Gamma}_X(p)] \begin{pmatrix} \hat{\Gamma}_X(i-j) \\ i, j = 1, \dots, p \end{pmatrix}^{-1},
 \tag{2.33}$$

where $\hat{\mathcal{A}}^{(i)} = (\hat{\alpha}_{kl}^{(i)})_{k,l=1,\dots,K}$. As we imposed diagonality of \mathcal{B} to achieve identification of Σ_e , by using the natural restriction $\beta_{kk} = 1 - \sum_{i=1}^p \sum_{l=1}^K |\alpha_{kl}^{(i)}|$ for $k = 1, \dots, K$, we get immediately the estimator $\hat{\mathcal{B}}$ defined by

$$\hat{\mathcal{B}} := I_K - \text{diag} \left([\hat{\mathcal{A}}^{(1)}, \dots, \hat{\mathcal{A}}^{(p)}] \mathbb{1}_{Kp} \right).
 \tag{2.34}$$

From Lemma 2.3, we get an estimator $\hat{\mu}_e$ for the innovation mean μ_e by rearranging (2.7) and plugging-in the sample versions of $\mathcal{A}^{(i)}$, $\mathcal{A}^{(-j)}$, \mathcal{B} and μ_X to get

$$\hat{\mu}_e = \hat{B}^{-1} \left(\left[I_K - \sum_{i=1}^p \hat{\mathcal{A}}^{(i)} \right] \hat{\mu}_X - \sum_{j=1}^p \hat{\mathcal{A}}^{(-j)} \mathbb{1}_K \right), \tag{2.35}$$

where $\hat{\mathcal{A}}^{(-i)} := (|\hat{\alpha}_{kl}^{(i)}| \mathbb{1}_{\{\hat{\alpha}_{kl}^{(i)} < 0\}})_{k,l=1,\dots,K}$, $i = 1, \dots, p$ analogous to (2.6). In scenarios where (some) diagonal elements $\hat{\beta}_{kk}$ equal zero such that \hat{B} is no longer invertible, the corresponding $\mu_{e,k}$'s are not identified, but the remaining mean parameters are still identified via (2.28); see also Remark 2.14. The *diagonal* of Σ_e , that is, $I_K \circ \Sigma_e$ can be estimated by

$$\widehat{I_K \circ \Sigma_e} := \text{diag}(\hat{\mu}_{e,i}(1 - \hat{\mu}_{e,i}), i = 1, \dots, K). \tag{2.36}$$

Moreover, by using the Yule–Walker equation (2.20) from Theorem 2.8, it is also possible to construct an estimator for the *non-diagonal* elements of Σ_e , that is, for $(\mathbb{1}_{K \times K} - I_K) \circ \Sigma_e$. Such an estimator is obtained by replacing the last line of (2.20) by (2.21), separating $(\mathbb{1}_{K \times K} - I_K) \circ \Sigma_e$ on one side of the equation (achieved by left- and right-multiplication with B^{-1}) and replacing all population quantities on the other side of the equation by their sample versions proposed above. In cases, where B is not invertible due to (some) zero diagonal entries, the non-diagonal elements of Σ_e are still identified; see the discussion around (2.29).

Finally, transition probabilities derived in Lemma 2.10 can be estimated in a similar fashion by replacing population quantities by the corresponding estimators to get

$$\begin{aligned} \hat{P}(X_t = s_0 | X_{t-1} = s_1, \dots, X_{t-p} = s_p) &= \hat{p}_{s_0 | s_1, \dots, s_p} \\ &= \sum_{r_0 \in \{0,1\}^K} \hat{p}_{r_0} \prod_{k=1}^K \left[\sum_{i=1}^p \sum_{l=1}^K |\hat{\alpha}_{kl}^{(i)}| \left[\mathbb{1}_{\{\hat{\alpha}_{kl}^{(i)} \geq 0\}} \delta_{s_{0,k} s_{i,l}} + \mathbb{1}_{\{\hat{\alpha}_{kl}^{(i)} < 0\}} \delta_{s_{0,k}(1-s_{i,l})} \right] + \hat{\beta}_{kk} \delta_{s_{0,k} r_{0,k}} \right], \end{aligned} \tag{2.37}$$

where

$$\hat{p}_{r_0} = \left(\prod_{k=1, r_{0,k}=1}^K \hat{\mu}_{e,k} \right) \left(\prod_{k=1, r_{0,k}=0}^K (1 - \hat{\mu}_{e,k}) \right) \tag{2.38}$$

in the case where the random vectors e_t consist of mutually independent Bernoulli random variables. In the dependent case, the estimator for off-diagonal elements of Σ_e , that is, $(\mathbb{1}_{K \times K} - I_K) \circ \Sigma_e$, derived above can be used to allow also for dependent Bernoulli random variables to incorporate linear dependence.

Remark 2.14 (Estimation outside of the parameter space). Estimation of the gbVAR parameters $\mathbf{A} := [\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}]$ using the Yule–Walker estimator $\hat{\mathbf{A}} := [\hat{\mathcal{A}}^{(1)}, \dots, \hat{\mathcal{A}}^{(p)}]$ can lead to invalid parameters with

$$\left[\hat{\mathcal{A}}_{|\cdot|}^{(1)}, \dots, \hat{\mathcal{A}}_{|\cdot|}^{(p)} \right] \mathbb{1}_{K^p} \notin [0, 1]^{K^p} \tag{2.39}$$

such that (at least) for one row $k_0 \in \{1, \dots, K\}$, we have

$$\left(|\hat{\alpha}_{k_0, \cdot}^{(1)}|, \dots, |\hat{\alpha}_{k_0, \cdot}^{(p)}| \right) \mathbb{1}_{K^p} = \sum_{i=1}^p \sum_{l=1}^K |\hat{\alpha}_{k_0,l}^{(i)}| > 1. \tag{2.40}$$

Let K^* denote the number of such rows and by \mathbf{K} the set of all indices $k_0 \in \{1, \dots, K\}$ satisfying (2.40). Then, we propose to estimate a constraint model that assures valid parameters $[\hat{\mathcal{A}}_{|\cdot|}^{(1)}, \dots, \hat{\mathcal{A}}_{|\cdot|}^{(p)}] \mathbb{1}_{K^p} \in [0, 1]^{K^*}$; see Lütkepohl,

2005, Sect. 5.2). For this purpose, we construct a $K^* \times (K^2p + K)$ matrix C of rank K^* by first defining the auxiliary matrix

$$\tilde{c}_{k_0} = \left(\left(\mathbb{1}_{\{A_{k,m} \geq 0\}} - \mathbb{1}_{\{A_{k,m} < 0\}} \right) \delta_{\{k_0=k\}} \right)_{k=1, \dots, K, m=1, \dots, Kp}$$

of dimension $K \times Kp$ for each $k_0 \in \mathbf{K}$. These matrices \tilde{c}_{k_0} have only entries of -1 and 1 in row $k = k_0$ and zero otherwise. Then, the matrix C combines the vectorized auxiliary matrices for every $k_0 \in \mathbf{K} := \{k_{0,1}, \dots, k_{0,K^*}\}$ by

$$C := \begin{pmatrix} \text{vec}(\tilde{c}_{k_{0,1}})' \\ \vdots \\ \text{vec}(\tilde{c}_{k_{0,K^*}})' \end{pmatrix}.$$

Furthermore, we define $Z = [Z_{(p-1)}, \dots, Z_T]$ to contain the variables $Z_t := \text{vec} [X_t, \dots, X_{t-p+1}]$ and the parameters $\beta = \text{vec} ([\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(p)}])$, thus $\hat{\beta}$ is the vectorized estimated parameter matrix from the Yule–Walker estimator (2.33). As we restrict the rows to have row sums (of the absolutes) equal to one, we set $c = \mathbb{1}_{K^*}$, which leads to the constraint estimator defined by

$$\hat{\beta}_{\text{valid}} = \hat{\beta} + \left[(ZZ')^{-1} \right] C' \left[C (ZZ')^{-1} C' \right]^{-1} (c - C\hat{\beta}). \tag{2.41}$$

Revectorizing $\hat{\beta}_{\text{valid}}$ leads to the constraint parameter estimator \hat{A}_{valid} and the corresponding \hat{B}_{valid} matrix is obtained as described in (2.34). Since \hat{B}_{valid} contains now K^* zero diagonal entries, the innovation process $(e_t, t \in \mathbb{Z})$ is not identified for those rows. The remaining entries of μ_e can be estimated by a reduced system of dimension $K - K^*$ consisting of all rows with $\hat{\beta}_{\text{valid},kk} \neq 0$; see also the discussion after (2.35).

3. SIMULATION STUDY

We investigate the performance of Yule–Walker-based estimators in gbVAR models as described in Section 2.4 by Monte Carlo simulations. To illustrate the estimation performance in several gbVAR model setups, we consider a) the (average) mean squared error (MSE) of different parameter estimators and b) the (average) mean absolute deviation error (MADE) of transition probability estimators.

For this purpose, we consider three different gbVAR(p) setups with orders $p = 1, 2$ and of dimensions $K = 3, 4$ for sample sizes $n = 100, 500, 1000$ to be able to judge the performance of parameter estimation in several gbVAR model specifications. Precisely, we consider data generating processes (DGPs) with the following specifications:

(DGP1) gbVAR(1) with $K = 3, \mu_e = (0.48, 0.52, 0.47)'$,

$$\mathcal{A}^{(1)} = \begin{pmatrix} 0.15 & -0.25 & 0.49 \\ -0.19 & 0.27 & 0.28 \\ 0.17 & -0.39 & 0.21 \end{pmatrix} \text{ and } B = \text{diag}(0.11, 0.26, 0.23)$$

(DGP2) gbVAR(1) with $K = 4, \mu_e = (0.48, 0.52, 0.47, 0.33)'$,

$$\mathcal{A}^{(1)} = \begin{pmatrix} -0.18 & 0.25 & -0.19 & -0.15 \\ 0.33 & -0.23 & 0.18 & -0.18 \\ -0.27 & -0.29 & 0.21 & -0.11 \\ 0.08 & 0.15 & -0.21 & -0.32 \end{pmatrix} \text{ and } B = \text{diag}(0.23, 0.08, 0.12, 0.24)$$

(DGP3) gbVAR(2) with $K = 3, \mu_e = (0.48, 0.52, 0.47)'$,

$$\mathcal{A}^{(1)} = \begin{pmatrix} -0.09 & 0.15 & -0.13 \\ 0.13 & -0.11 & 0.28 \\ 0.13 & -0.19 & -0.18 \end{pmatrix}, \quad \mathcal{A}^{(2)} = \begin{pmatrix} -0.18 & 0.07 & -0.19 \\ -0.09 & -0.17 & 0.15 \\ -0.17 & -0.09 & 0.14 \end{pmatrix},$$

and $B = \text{diag}(0.19, 0.07, 0.10)$.

For all DGPs, the corresponding innovation process $(e_t, t \in \mathbb{Z})$ consists of K independent Bernoulli processes $(e_{t,k}, t \in \mathbb{Z}), k = 1, \dots, K$ with $\mu_{e,k} = P(e_{t,k} = 1)$ leading to diagonal Σ_e matrices with diagonal entries $\mu_{e,k}(1 - \mu_{e,k}), k = 1, \dots, K$. Note that we make use of positive as well as negative entries in $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$. Hence, these coefficient matrices are related to the diagonal matrix B via $\beta_{kk} = 1 - \sum_{i=1}^p \sum_{l=1}^K |\alpha_{kl}^{(i)}|$. In Section 3.2, we address also the estimation of the off-diagonal elements of a non-diagonal variance–covariance matrix Σ_e .

3.1. Average MSE Estimation Performance

To measure the estimation performance, we calculate averages of the entry-wise mean squared errors (MSE) of the estimators $\hat{\mathcal{A}}^{(1)}$ and $\hat{\mathcal{A}}^{(2)}, \hat{\mu}_X, \hat{\mu}_e$ and \hat{B} , respectively, based on 1000 Monte Carlo replications for each DGP and each sample size. The simulation results are presented in Table I. It can be seen that the estimation performance improves with increasing sample size for all estimators and all DGPs. In comparison, the estimation of the mean innovation vector μ_e is least precise with an average mean squared error around 10 percent. This phenomenon can be explained by formula (2.35), which requires the inversion of the diagonal matrix \hat{B} . Due to rather small diagonal entries of B , already small deviations in \hat{B} might lead to a less stable estimation of μ_e and to a larger MSE.

3.2. Average MSE Estimation of Non-diagonal Σ_e

As discussed in Remark 2.13, the imposed diagonality of $B^{(0)}$ does generally allow to identify also the non-diagonal entries of Σ_e . These can be estimated by using the Yule–Walker equation for $h = 0$ from Theorem 2.8 in conjunction with (2.21), where the corresponding estimator is obtained by replacing all population quantities by their sample analogs as described in Section 2.4.

For illustration, we consider a bivariate gbVAR(1) process, where we used the four entries $\alpha_{12}, \alpha_{13}, \alpha_{32}$ and α_{33} in $\mathcal{A}^{(1)}$ of DGP1 leading to $B = \text{diag}(0.26, 0.4)$. The bivariate marginal distribution of the innovations $(e_t, t \in \mathbb{Z})$ is fully specified by $\mu_{e,1} = P(e_{t,1} = 1) = 0.260, \mu_{e,2} = P(e_{t,2} = 1) = 0.382$ and

$$\Sigma_e = \begin{pmatrix} 0.19240 & 0.134680 \\ 0.13468 & 0.236076 \end{pmatrix},$$

Table I. Average MSE estimation performance for different parameter estimators $\hat{\mathcal{A}}^{(1)}, \hat{\mathcal{A}}^{(2)}, \hat{\mu}_X, \hat{\mu}_e$ and \hat{B} for three different parameter specifications DGP1, DGP2, and DGP3, respectively

	n	MSE of $\hat{\mathcal{A}}^{(1)}$	MSE of $\hat{\mathcal{A}}^{(2)}$	MSE of $\hat{\mu}_e$	MSE of $\hat{\mu}_X$	MSE of \hat{B}
DGP1	100	0.0085		0.0626	0.0046	0.0214
	500	0.0017		0.0152	0.0009	0.0034
	1000	0.0008		0.0070	0.0005	0.0015
DGP2	100	0.0085		0.0794	0.0022	0.0426
	500	0.0017		0.0388	0.0004	0.0085
	1000	0.0008		0.0208	0.0002	0.0035
DGP3	100	0.0084	0.0083	0.1041	0.0015	0.0821
	500	0.0018	0.0018	0.0701	0.0003	0.0374
	1000	0.0009	0.0009	0.0502	0.0002	0.0198

Table II. Mean squared error of diagonal and non-diagonal elements of $\widehat{\Sigma}_e$

n	$\widehat{\sigma}_{11}$	$\widehat{\sigma}_{12}, \widehat{\sigma}_{21}$	$\widehat{\sigma}_{22}$
100	0.0100	16.7383	0.0028
500	0.0025	0.0101	0.0002
1000	0.0009	0.0045	0.0001

where Σ_e determines the joint probabilities $P((e_{t,1}, e_{t,2}) = (i, j))$ for $i, j \in \{0, 1\}$. Practically, such an innovation process can be generated by first sampling the innovation $e_{t,1}$ from a Bernoulli distribution with $P(e_{t,1} = 1) = 0.26$. Then, we generate $e_{t,2}$ conditional on the outcome of $e_{t,1}$ such that $P(e_{t,2} = 1 | e_{t,1} = 1) = 0.9$ and $P(e_{t,2} = 1 | e_{t,1} = 0) = 0.2$. This leads to a marginal Bernoulli distribution of $e_{t,2}$ with $P(e_{t,2} = 1) = 0.9 \cdot 0.26 + 0.2 \cdot (1 - 0.26) = 0.382$. For a comprehensive discussion of multi-variate Bernoulli distributions allowing for dependence also beyond $K = 2$, we refer to Dai *et al.* (2013).

In Table II, we report the MSE for the diagonal and non-diagonal elements of $\widehat{\Sigma}_e$ for different sample sizes. It can be seen that the MSE decays for increasing sample size. However, for a small sample size of $n = 100$, the MSE of the off-diagonal elements $\widehat{\sigma}_{e,12}$ is huge with 16.7383. This value is caused by a mis-estimation for some few Monte Carlo replications, where the estimated parameters already show a large MSE. Due to a matrix inversion, this leads to unstable and unreliable estimates. Nevertheless, this issue disappears for larger sample sizes such that the joint distribution of the innovation process in form of the non-diagonal entries of Σ_e can be consistently estimated. However, as \mathcal{B} is imposed to be diagonal for identification reasons, the non-diagonal entries of Σ_e do not have large effects on the stochastic properties of the model. Hence, in practice, it seems to be recommendable to avoid the estimation of a non-diagonal Σ_e due to potentially unstable estimation results.

3.3. Average MADE Estimation Performance

An alternative concept to measure the estimation performance in gbVAR models is based on (average) mean absolute deviation error (MADE) of transition probability estimators. Note that a direct comparison of prediction probabilities in $[0, 1]$ and outcomes in $\{0, 1\}$ are not straightforward and might be misleading to judge the parameter estimation performance. Hence, we compare the (one step ahead) population transition probabilities with the corresponding estimated transition probabilities and consider

$$|p_{s_0|s_1, \dots, s_p} - \widehat{p}_{s_0|s_1, \dots, s_p}| = \left| P(X_t = s_0 | X_{t-1} = s_1, \dots, X_{t-p} = s_p) - \widehat{P}(X_t = s_0 | X_{t-1} = s_1, \dots, X_{t-p} = s_p) \right| \quad (3.1)$$

with $p_{s_0|s_1, \dots, s_p} = P(X_t = s_0 | X_{t-1} = s_1, \dots, X_{t-p} = s_p)$ as obtained in Theorem 2.10 and $\widehat{p}_{s_0|s_1, \dots, s_p} = \widehat{P}(X_t = s_0 | X_{t-1} = s_1, \dots, X_{t-p} = s_p)$ as constructed in (2.37) for the special case, where $e_{t,1}, \dots, e_{t,K}$ are independent such that $p_{r_0} = P(e_t = r_0) = \prod_{k=1}^K P(e_{t,k} = r_{0,k})$. Mainly, there are two possibilities to use (3.1) to judge the average estimation accuracy in gbVAR models. The first one, considers the absolute deviation of the transition probabilities according to their actual appearances in the Monte Carlo sample under consideration. That is, given $(X_1, \dots, X_n) = (x_1, \dots, x_n)$, we compute

$$\frac{1}{n-p} \sum_{t=p+1}^n |p_{x_t|x_{t-1}, \dots, x_{t-p}} - \widehat{p}_{x_t|x_{t-1}, \dots, x_{t-p}}|. \quad (3.2)$$

The second option is to calculate the absolute deviate of the transition probabilities over all possible states of s_0, \dots, s_p in the state space $\{0, 1\}^K$ leading to

$$\frac{1}{2^{K(p+1)}} \sum_{s_0, s_1, \dots, s_p \in \{0, 1\}^K} |p_{s_0|s_1, \dots, s_p} - \widehat{p}_{s_0|s_1, \dots, s_p}|. \quad (3.3)$$

Table III. Estimation performance based on both versions of average MADE in (3.2) and (3.3) for DGP1, DGP2 and DGP3, respectively

n	Average MADE (3.2)			Average MADE (3.3)		
	DGP1	DGP2	DGP3	DGP1	DGP2	DGP3
100	0.0341	0.0358	0.0196	0.0338	0.0169	0.0179
500	0.0154	0.0151	0.0082	0.0149	0.0077	0.0076
1000	0.0108	0.0106	0.0054	0.0104	0.0054	0.0050

Based on 1000 Monte Carlo samples, we report the estimation performance using both versions (3.2) and (3.3) of average MADE for each DGP and each sample size in Table III.

4. REAL DATA EXAMPLE: PM10 DATA

With increasing environmental awareness, there has been great interest in collecting and analyzing data describing the extent of the pollution of the environment and its impact on the health of the population. In recent years, there is particularly growing interest in air pollution with particulate matter in European cities and metropolitan areas. The European Union established the *European emission standards*, which include limits for particulates in the air. In 2008, the European Parliament made a policy on critical values in the rule 2008/50/EG for air pollution substances. In particular, it is by law not allowed to exceed the threshold on 35 or more days per year. For particulate matter PM_{10} (coarse particles with a diameter between 2.5 and $10\ \mu\text{m}$), the liability has a threshold of $50\ \mu\text{g}/\text{m}^3$. Hence, whenever the amount of PM_{10} exceeds the threshold of $50\ \mu\text{g}/\text{m}^3$ at a certain monitoring station, this will cause a ‘*fine dust alarm*’. Hence, for each such monitoring station, this results in a binary sequence with states ‘exceedance’ and ‘no exceedance’. In fact, the current public discourse centers to a large extent around whether the threshold is exceeded or not, and less about the actual amount of fine dust measured.

In view of these EU regulations, Stuttgart, Germany is one poorly prominent city, where air pollution generally is a major problem. The reasons for these problems are essentially twofold. On the one hand, they can be explained by its geographic location in a valley leading to a poor air exchange in the city area. On the other hand, the main industry such as automobile companies and suppliers as well as financial industry is located near to the city center. Due to the restricted space in a valley to expand, many people live in suburbs of Stuttgart and have to commute to their work places. The commuting traffic concentrates on few main traffic routes, which are highly frequented during rush hours. Hence, large portions of particulate matter in the air in and around Stuttgart are caused by individual mobility.

In Figure 2, we show the recorded fine dust alarms at six monitoring stations in Stuttgart, Germany for 886 consecutive days from 1 March 2016 to 31 July 2018. Precisely, with $Y_{t,k}$, $k = 1, \dots, 6$, representing the day-wise mean of the PM_{10} values for each station, Figure 2 shows binarized time series data ($X_t = (X_{t,1}, \dots, X_{t,6})$, $t \in \mathbb{Z}$), where $X_{t,k} = 1$, if $Y_{t,k} \geq 50$ and $X_{t,k} = 0$ otherwise. The locations of the six monitoring stations Arnulf-Klett Platz (AKP), Bad Cannstatt (BC), Hauptstätter Straße (HS), Hohenheimer Straße (HH), Neckartor (NRT) and Stadtgarten (SG) are illustrated in Figure 5.

A first inspection of the data in Figure 2 shows that fine dust alarms tend to occur in clusters indicating serial and cross-sectional dependence. However, all sequences do not show long runs of fine dust alarms, but rather long runs without any alarm. Moreover, fine dust alarms tend to show more likely in winter. This is due to the fact that the topological influence of stationary temperature inversion hinders vertical air exchange. Hence, after the analysis of the whole data set, we will also run the analysis on the subsample of the data collected in the *fine dust alarm period* (October 15–March 15) afterwards.

First, when considering the whole data set, one station (NRT, $k = 5$) shows considerable more exceedances in comparison to the other stations with fine dust alarms occurring in about 13% of the days in the considered time

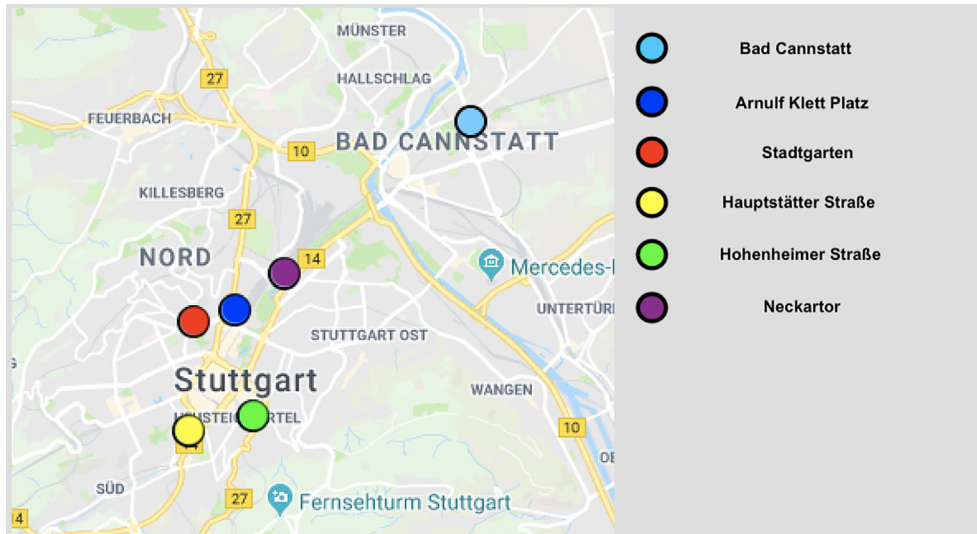


Figure 5. Locations of six major PM₁₀ monitoring stations in Stuttgart, Germany

Table IV. AUC values for a fitted gbVAR(*p*) model for *p* ∈ {1, 2} componentwise for each station and the overall mean

	AKP	BC	HS	HH	NRT	SG	mean
<i>p</i> = 1	0.9368	0.9500	0.9337	0.9459	0.8022	0.9305	0.9164
<i>p</i> = 2	0.9388	0.9572	0.9352	0.9479	0.8111	0.9306	0.9201

period. This is captured by the sample mean vector

$$\hat{\mu}_x = (0.0420, 0.0227, 0.0397, 0.0386, 0.1283, 0.0249)'$$

The mean of all the other stations lie around 3% indicating that only few fine dust alarms are detected. Overall, this is not surprising, as NRT is located at one of the most frequented roads of Stuttgart, where high buildings on one side of the road hinder the air exchange and favor air pollution. In contrast, BC with the smallest value of 0.0227 is located at an accommodation route to the city outside of the city center. Now, to study the serial dependence in the data, we aim to fit a gbVAR(*p*) model. As gbVAR processes satisfy standard Yule–Walker equations and can be estimated by Yule–Walker estimators as described in (2.33), we can make use of classical order selection criteria such as Hannan–Quinn (HQ) or BIC to determine an appropriate order *p* of the fitted gbVAR process. Whereas HQ selects a more parsimonious model with *p* = 1, BIC leads to *p* = 2. To make a choice which model fits best in terms of prediction performance, we use the receiver operating characteristic (ROC) curve and the corresponding area under the curve (AUC), where an AUC near to one indicates good prediction performance. For this purpose, similar to Section 3.3, we make use of transition probability estimators $\hat{P}(X_t = s_0 | X_{t-1} = s_1, \dots, X_{t-p} = s_p)$ as constructed in (2.37) to estimate the transition probabilities for each station, which allows to compute the ROC curves and AUC values. In Table IV, we show the resulting component-by-component AUC values and their overall means for model orders *p* ∈ {1, 2}. Both models show a good prediction performance with AUC values near to one. However, the additional benefit of fitting a gbVAR(2) model in comparison to a more parsimonious gbVAR(1) model is minor. Hence, we make use of a gbVAR(1) model in the following to further analyze the PM₁₀ data set.

In this case, Yule–Walker estimation leads to an estimated parameter matrix having $\sum_{k=1}^K |\hat{\alpha}_{kl}| = 1.1451 > 1$ for *k* = 5 corresponding to NRT. Hence, in view of Remark 2.14, we have to use constraint estimation leading to the

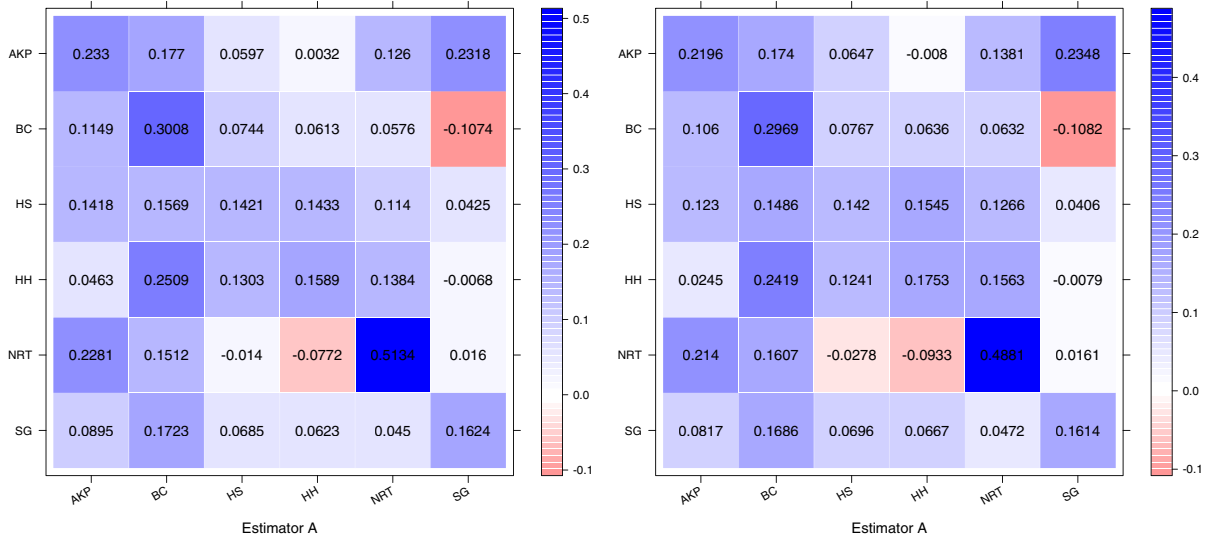


Figure 6. Heatmaps of estimated coefficient matrices for fitted gbVAR(1) processes on the whole PM₁₀ data sample (\hat{A} , left panel) and on the fine dust alarm period (\hat{A}_{FDAP} , right panel)

estimated parameter matrix \hat{A} as shown in the left panel of Figure 6. The absolute eigenvalues of \hat{A} compute to $\{0.7092, 0.2933, 0.2933, 0.1143, 0.1143, 0.0019\}$ such that the fitted gbVAR(1) model is stationary.

In each row of \hat{A} , we can see which past state at time $t - 1$ of the six monitoring stations (fine dust alarm or not) does affect the state at time t . For example, with 51.33% probability, NRT takes the same value as the day before. In contrast, the largest entry in row $k = 4$ (HS) is 0.2509 in the second column corresponding to BC such that it takes its value of the day before with probability of about 25%. Note also, as can be seen in the left panel of Figure 6, that the fitting of a gbVAR model leads to some negative coefficients in \hat{A} . From a modeling perspective, this naturally leads to more flexibility in comparison to models that do allow only for non-negative coefficients.

For identification purposes, we impose B to be diagonal; see also Remark 2.13. Since we constrained the estimation to achieve $\sum_{i=1}^K |\hat{\alpha}_{5i}| = 1$ and set $\hat{\beta}_{55} = 0$ for NRT, the effective innovation process is of reduced dimension $\tilde{K} = 6 - 1 = 5$. Hence, as described in Section 2.4, we get

$$diag(\hat{B}) = (0.1694, 0.2837, 0.2594, 0.2684, 0.0000, 0.3999).$$

The diagonal entries of \hat{B} indicate how often the corresponding innovation terms are selected. For example, for SG, in about 40% of the days, the innovation term enters the gbVAR model, whereas this happens only in about 17% for AKP.

By using formulae (2.35) and (2.36), we can estimate the mean vector and the variances of the innovations. This leads to

$$\hat{\mu}_e = (0.1176, 0.4084, 0.0693, 0.0993, 0.0000, 0.0203)' \tag{4.1}$$

$$(\hat{\sigma}_{e,ii}, i = 1, \dots, 6) = (0.1038, 0.2416, 0.0645, 0.0895, 0.0000, 0.0199). \tag{4.2}$$

Hence, for PM₁₀ data, this indicates that the innovation terms generally take zero values with high probability. Note that $\hat{\mu}_{e,5}$ and $\hat{\sigma}_{e,55}$ are not identified due to $\hat{\beta}_{55} = 0$ and we set $\hat{\mu}_{e,5} = \hat{\sigma}_{e,55} = 0$ for convenience.

Now, we conduct the same analysis as above, but only for the data collected in the *fine dust alarm period* (FDAP) (October 15–March 15). First, we obtain the sample mean vector

$$\hat{\mu}_{X,FDAP} = (0.08981, 0.04854, 0.08252, 0.08010, 0.25243, 0.05340)'$$

As there are only very few alarms outside of the FDAP, the sample vector $\hat{\mu}_{X,FDAP}$ has entries that are about twice as large as those of $\hat{\mu}_X$. Now, in contrast to the analysis of the whole sample, both order selection criteria HQ and BIC select the order $p = 1$ and a corresponding AUC analysis (not reported) leads again to only minor improvements when using order $p = 2$ in comparison to the more parsimonious gbVAR(1) model. For the FDAP, again a constraint estimation has to be executed leading to the estimated parameter matrix $\hat{\mathcal{A}}_{FDAP}$ as shown in the right panel of Figure 6. The absolute eigenvalues of $\hat{\mathcal{A}}_{FDAP}$ compute to $\{0.6845, 0.3070, 0.3070, 0.1147, 0.1147, 0.0069\}$. Finally, based on $\hat{\mathcal{A}}_{FDAP}$, we get

$$\text{diag}(\hat{\mathcal{B}}_{FDAP}) = (0.1609, 0.2853, 0.2646, 0.2700, 0.0000, 0.4050)$$

leading to

$$\hat{\mu}_{e,FDAP} = (0.3255, 0.4457, 0.1436, 0.1858, 0.0000, 0.0449)' \quad (4.3)$$

$$(\hat{\sigma}_{e,FDAP,ii}, i = 1, \dots, 6) = (0.2195, 0.2471, 0.1230, 0.1513, 0.0000, 0.0429). \quad (4.4)$$

In comparison to the analysis of the whole data sample, there is only a pronounced change in the estimated mean vectors, when restricting the analysis to the FDAP. In particular, $\hat{\mu}_{X,FDAP}$ is estimated about twice as large as $\hat{\mu}_X$. In contrast, the estimation of the coefficient matrix leading to $\hat{\mathcal{A}}$ and $\hat{\mathcal{A}}_{FDAP}$ turns out to be very similar.

5. CONCLUSION

We consider vector-valued extensions of gbAR processes introduced by Jentsch and Reichmann (2019) to model multi-variate binary time series data with potentially negative model parameters. We define generalized binary vector AR (gbVAR) models of order $p \in \mathbb{N}$ and provide a full description of the gbVAR model class. We derive stochastic properties of gbVAR processes including formulae for the mean vector, stationarity conditions, moving-average representations, geometric mixing properties, Yule–Walker equations and transition probabilities. Possible identification issues and parameter estimation in gbVAR models based on Yule–Walker estimators are discussed.

In a simulation study, the estimation performance of Yule–Walker estimators and related estimators is analyzed in several regards indicating good finite sample properties. In a real data application, we fit gbVAR processes to binarized PM₁₀ data from Stuttgart, Germany. The estimated gbVAR(1) model contains positive as well as negative coefficients to capture the serial dependence in the data and proves to yield accurate predictions.

ACKNOWLEDGEMENTS

This work was partly supported by the German Research Foundation DFG [SFB 884, Project B6 and INST 35/1134-1 FUGG]; by the Eliteprogramme for Postdocs of the Baden Württemberg Stiftung and the German Federal State of Baden-Württemberg through the <https://www.bwhpc.de/>. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data used in Section 4 are available in the Supporting Information of this article. It is openly available from the webpage of the ‘Landesanstalt für Umwelt Baden Württemberg’ at <https://udo.lubw.baden-wuerttemberg.de/public/>.

SUPPORTING INFORMATION

Additional supporting information may be found online in the supporting information tab for this article.

REFERENCES

- Bahadur RR. 1961. A representation of the joint distribution of responses to n dichotomous items. In *Studies in Item Analysis and Prediction*. Stanford Mathematical Studies in the Social Sciences VI, Solomon H. (ed.) Stanford, CA: Stanford University Press; 158–168.
- Bellégo C, Ferrara L. 2009. *Forecasting Euro-area recessions using time-varying binary response models for financial variables*. Working papers, Banque de France, 259.
- Billingsley P. 1968. *Convergence of Probability Measures*, 1st ed. New York: John Wiley and Sons.
- Bradley RC. 2005. Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* **2**: 107–144.
- Candelon B, Dumitrescu E-I, Hurlin C, Palm FC. 2012. Multivariate dynamic probit models: an application to financial crises mutation. *Advances in Econometrics* **32**: 395–427.
- Cheng L, Wang Y, Wu C, Wu H, Zhang Y. 2013. Signal processing for a positioning system with binary sensory outputs. *Sensors and Actuators A: Physical* **201**: 86–92.
- Czado C. 2000. Multivariate regression analysis of panel data with binary outcomes applied to unemployment data. *Statistical Papers* **41**: 281–304.
- Dai B, Ding S, Wahba G. 2013. Multivariate Bernoulli distribution. *Bernoulli* **19**: 1465–1483.
- Dedecker J., Doukhan P., Lang G., Leon J.R., Louhichi S., Prieur C. 2007. Weak dependence: with examples and applications. In *Lecture Notes in Statistics*: Springer.
- Doukhan P. 1994. Mixing – properties and examples. In *Lecture Notes in Statistics*: Springer.
- Fokianos K, Kedem B. 2003. Regression theory for categorical time series. *Statistical Science* **18**(3): 357–376.
- 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. *Stochastic Environmental Research and Risk Assessment* **32**: 2495–2514.
- Honoré BE, Kyriazidou E. 2019a. Chapter 8 – Panel vector autoregressions with binary data. In *Panel Data Econometrics*, Tsionas M. (ed.): Academic Press; 197–223.
- Honoré BE, Kyriazidou E. 2019b. Identification in binary response panel data models: Is point-identification more common than we thought?. *Annals of Economics and Statistics* **134**: 207–226.
- Jacobs PA, Lewis PAW. 1983. Stationary discrete autoregressive – Moving average time series generated by mixtures. *Journal of Time Series Analysis* **4**: 19–36.
- Jiang W, Song S, Hou L, Zhao H. 2020. *A set of efficient methods to generate high-dimensional binary data with specified correlation structures*, ArXiv:2007.14080v1.
- Jentsch C, Reichmann L. 2019. Generalized binary time series models. *Econometrics* **7**: 49.
- Johnson N, Kotz S, Balakrishnan N. 1997. *Discrete Multivariate Distributions*. New York: Wiley & Sons.
- Kedem B. 1980. Binary time series. In *Lecture Notes in Pure and Applied Mathematics* New York: Marcel Dekker.
- Lütkepohl H. 2005. *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.
- McKenzie E. 2003. Discrete variate time series. *Handbook of Statistics* **21**: 573–606.
- Möller TA, Weiß CH. 2020. Generalized Discrete autoregressive moving-average models. *Applied Stochastic Models in Business and Industry*: 1–19.
- Ritzwoller DM, Romano JP. 2020. *Uncertainty in the hot hand fallacy: detecting streaky alternatives to random Bernoulli sequences*, Technical Report No. 2020-02, February 2020, Department of Statistics, Stanford University.
- Startz R. 2006. *Binomial Autoregressive Moving Average Models with an Application to U.S. Recessions*. Working Papers 56, Center for Statistics in the Social Sciences, University of Washington.
- Weiß CH, Göb R. 2008. Measuring serial dependence in categorical time series. *Advances in Statistical Analysis* **92**: 71–89.
- Weiß CH. 2009a. *Categorical time series analysis and application in statistical quality control*, Dissertation, de-Verlag im Internet GmbH.
- Weiß CH. 2009b. Properties of a class of binary ARMA models. *Statistics* **43**: 131–138.
- Weiß CH. 2011a. Empirical measures of signed serial dependence in categorical time series. *Journal of Statistical Computation and Simulation* **81**: 411–429.
- Weiß CH. 2011b. Generalized choice models for categorical time series. *Journal of Statistical Planning and Inference* **141**: 2849–2862.