Computational and Applied Mathematics (2023) 42:183
https://doi.org/10.1007/s40314-023-02322-w

Inflated beta autoregressive moving average models

Fábio M. Bayer[1,2] · Guilherme Pumi[2,3] · Tarciana Liberal Pereira[4] · Tatiene C. Souza[4]

Received: 29 August 2022 / Revised: 19 December 2022 / Accepted: 1 May 2023 /
Published online: 17 May 2023

Abstract
In this paper, we introduce the inflated beta autoregressive moving average (IβARMA) models for modeling and forecasting time series data that assume values in the intervals (0,1], [0,1) or [0,1]. The proposed model considers a set of regressors, an autoregressive moving average structure and a link function to model the conditional mean of inflated beta conditionally distributed variable observed over the time. We develop partial likelihood estimation and derive closed-form expressions for the score vector and the cumulative partial information matrix. Hypotheses testing, confidence interval, some diagnostic tools and forecasting are also proposed. We evaluate the finite sample performances of partial maximum likelihood estimators and confidence interval using Monte Carlo simulations. Two empirical applications related to forecasting hydro-environmental data are presented and discussed.

Keywords Inflated beta distribution · Forecasts · Rates and proportions · Time series

Mathematics Subject Classification 62F03 · 62F10 · 62F12 · 62M10 · 62P12

Communicated by Clémentine Prieur.

✉ Fábio M. Bayer
bayer@ufsm.br

Guilherme Pumi
guilherme.pumi@ufrgs.br

Tarciana Liberal Pereira
tarcianalp@gmail.com

Tatiene C. Souza
tatitex@gmail.com

1 Departamento de Estatística and LACESM, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil

2 Programa de Pós-Graduação em Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

3 Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

4 Departamento de Estatística, Universidade Federal da Paraíba, João Pessoa, PB, Brazil

1 Introduction

The most commonly assumption made in applications of time series is, by far, Gaussianity (Chuang and Yu 2007; Box et al. 2015). This should be no surprise as Gaussianity allows for simpler derivation of theoretical results, often leading to simpler models. However, this assumption can be too restrictive for many applications (Tiku et al. 2000; Bayer et al. 2020). As a consequence, there has been an increase in interest in non-Gaussian time series models (Zheng et al. 2015). In this field, some general models based on generalized linear models (GLM) (McCullagh and Nelder 1989) are considered in Li (1991), Li (1994), Benjamin et al. (2003), and Fokianos and Kedem (2004). A comprehensive reference on general models for time series analysis is Kedem and Fokianos (2002).

When dealing with data that is naturally bounded, for example, proportion of votes in an election, relative humidity and administrative efficiency score, Gaussian models are not adequate. This problem is stressed from a forecasting point of view where, due to the nature of Gaussian models, one can obtain forecasted values outside the natural bounds of the data. Recent works on non-Gaussian time series modeling are interested in modeling the behavior of double bounded variables, such as in Rocha and Cribari-Neto (2009), da-Silva et al. (2011), Guolo and Varin (2014), Bayer et al. (2017), Bayer et al. (2018), Pumi et al. (2019), Pumi et al. (2021) and Melchior et al. (2021). These models are based on the assumption that the variable of interest follows a continuous distribution. However, there are numerous instances where the data present one or more values that appear more frequently than they should, even in the case of discrete data. This phenomenon is called inflation. There are several time series applications and forecasting problems for which a continuous distribution is not suitable for data modeling because of the presence of point masses. For instance, in very humid places such as the Amazon, relative air humidity data often present excess of 1's and it is also common to observe excess of 0's in percentage of useful volume in water reservoirs.

In these situations, continuous distributions, such as the beta distribution, are not suitable for data modeling, since the beta-based log-likelihood function becomes unbounded and, thus, requiring a new strategy for modeling this type of data. A common and simple approach to handle with the problem of inflation is to model the underlying distribution as a mixture of a hypothesized distribution, responsible for the overall probabilistic behavior, and a discrete (possible point mass) distribution to account for the inflation itself. This idea is easily extended to consider covariates in the model. An early work in this direction is the zero-inflated Poisson regression of Lambert (1992). In the context of double bounded continuous data, Ospina and Ferrari (2012) and Bayes and Valdivieso (2016) generalized the beta regression model (Ferrari and Cribari-Neto 2004) to account for the presence of 0's and/or 1's, possibly in excess. To this end, the authors considered the inflated beta distribution which is a convex combination of the beta distribution and two point masses located at 0 and 1.

This paper examines the problem of modeling and forecasting time series assuming values in the intervals $(0, 1]$, $[0, 1)$ or $[0, 1]$. We consider an autoregressive moving average (ARMA) structure for predicting the conditional mean of an inflated beta distributed dependent variable. The proposed model, called inflated beta autoregressive moving average (IβARMA), naturally accommodates occurrences of values 0 and/or 1 with positive probability, extending the applicability of the beta autoregressive moving average (βARMA) models (Rocha and Cribari-Neto 2009). Another generalization we propose is the use of the so-called partial likelihood for inference in the introduced IβARMA model. This framework generalizes both, the conditional likelihood applied in Rocha and Cribari-Neto (2009) and the standard likelihood by allowing the presence of time-dependent random covariates in the model. Finally, since the distribution of rates and proportions are typically asymmetric, the proposed model naturally accommodates asymmetries also avoiding transforming the data and any of its undesirable effects.

The paper unfolds as follows. Section 2 introduces the proposed model and partial likelihood method for parameter estimation. Confidence intervals and hypothesis testing strategies are presented in Section 3. Section 4 focuses on model selection criteria, residuals and forecasting. Section 5 contains Monte Carlo simulation results on parameter estimation. Section 6 illustrates the methodology by applying the model to relative humidity data and the percentage of useful volume in a water reservoir. Concluding remarks are given in Sect. 7. Finally, closed-form expressions for the first derivatives of partial log-likelihood function (score function) and for the partial cumulative information matrix are provided in the Appendix.

2 Proposed model

Let $\{\boldsymbol{x}_t\}_{t\in\mathbb{Z}}$ denote a set of r-dimensional possibly random covariates for a stochastic process $\{y_t\}_{t\in\mathbb{Z}}$ taking values in [0, 1] and let $\mathscr{F}_{t-1}$ denote the σ-field generated by past observations of the response variable and the past and present values (when known) of the covariates available up to time t. In practical terms, $\mathscr{F}_{t-1}$ can be viewed as the set containing all information regarding the data available to the researcher at time t. In technical terms, we have $\mathscr{F}_{t-1} = \sigma\{y_{t-1}, y_{t-2}, \ldots, \boldsymbol{x}_t^\top, \boldsymbol{x}_{t-1}^\top, \boldsymbol{x}_{t-2}^\top \ldots\}$.

In the terminology of Cox (1981), the proposed model is an observation-driven model in which the random component y_t follows, conditionally on $\mathscr{F}_{t-1}$, the inflated beta distribution with probability density function given by Bayes and Valdivieso (2016)

$$f(y_t;\alpha_0, \alpha_1, \mu_t, \phi \mid \mathscr{F}_{t-1}) = \big[\alpha_0(1-\mu_t)\big]^{I_0(y_t)}\big[\alpha_1\mu_t\big]^{I_1(y_t)}\big[c_t\mathtt{b}(y_t;\nu_t,\phi)\big]^{I_{(0,1)}(y_t)}, \quad (1)$$

where $I_A(\cdot)$ is the indicator function of the set A, $\mu_t = \mathbb{E}(y_t \mid \mathscr{F}_{t-1}) \in (0, 1)$, $\alpha_0, \alpha_1 \in [0, 1]$,

$$c_t := 1 - \alpha_0(1-\mu_t) - \alpha_1\mu_t \quad \text{and} \quad \nu_t := \frac{(1-\alpha_1)\mu_t}{c_t}, \quad (2)$$

and $\mathtt{b}(y;\nu,\phi)$ is the beta density function parameterized as in Ferrari and Cribari-Neto (2004):

$$\mathtt{b}(y;\nu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\nu\phi)\Gamma\big((1-\nu)\phi\big)} y^{\nu\phi-1}(1-y)^{(1-\nu)\phi-1}, \; 0 < y < 1,$$

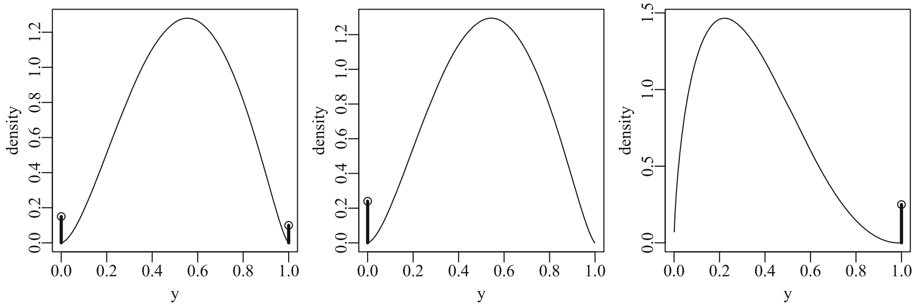
where $0 < \nu < 1$ and $\phi > 0$. We say that, conditionally on $\mathscr{F}_{t-1}$, $y_t \sim \mathcal{BI}(\alpha_0, \alpha_1, \mu_t, \phi)$. The respective conditional cumulative distribution function is given by

$$F(y_t;\alpha_0, \alpha_1, \mu_t, \phi \mid \mathscr{F}_{t-1}) = \alpha_0(1-\mu_t) + \alpha_1\mu_t I_1(y_t) + c_t\mathtt{B}(y_t;\nu_t,\phi)I_{(0,1]}(y_t), \quad (3)$$

where $\mathtt{B}(y;\nu,\phi)$ is the cumulative distribution function of a beta distribution with mean parameter ν and precision ϕ. Observe that, if $y_t \sim \mathcal{BI}(\alpha_0, \alpha_1, \mu_t, \phi)$, then

$$\mathbb{E}(y_t \mid \mathscr{F}_{t-1}) = \mu_t \;\text{ and }\; \mathrm{Var}(y_t \mid \mathscr{F}_{t-1}) = \frac{(1+\alpha_1\phi)}{1+\phi}\mu_t + \left(\frac{(1-\alpha_1)^2\phi}{c_t(1+\phi)} - 1\right)\mu_t^2.$$

When $\alpha_0 > 0$ and $\alpha_1 = 0$, a random variable y_t following (1) is said to have a zero-inflated beta distribution. If $\alpha_0 = 0$ and $\alpha_1 > 0$, the random variable y_t is said to have an one-inflated beta distribution. The former case occurs with probability $P(y_t = 0 \mid \mathscr{F}_{t-1}) = \alpha_0(1-\mu_t)$



(a) $\alpha_0 = 0.3,\ \alpha_1 = 0.2,\ \mu = 0.5$ (b) $\alpha_0 = 0.4,\ \alpha_1 = 0.0,\ \mu = 0.4$ (c) $\alpha_0 = 0.0,\ \alpha_1 = 0.5,\ \mu = 0.5$

Fig. 1 Mixed probability function (1) for $y \sim \mathcal{BI}(\alpha_0, \alpha_1, \mu, \phi)$, with $\phi = 5$ and different parameter values of α_0, α_1, and μ

and the second one with probability $P(y_t = 1 \mid \mathscr{F}_{t-1}) = \alpha_1 \mu_t$. Note that this makes sense, since the higher the conditional mean, the greater the probability of having y_t occurrences equal to one (closer to the upper bound). Similarly, the probability of occurrence of values equal to zero is greater when the conditional mean is close to the lower bound of the interval. Of course, if $\alpha_0 = 0$ and $\alpha_1 = 0$, (1) is simply the density of a beta distributed random variable. Figure 1 presents the graphs of some inflated beta mixed probability functions.

Let $g : (0, 1) \to \mathbb{R}$ be a twice differentiable monotonic one-to-one link function for which the inverse link is of class $\mathcal{C}^2(\mathbb{R})$ (the class of twice continuously differentiable functions in $\mathbb{R}$). The model's systematic component specify the conditional mean of y_t through a GLM-like structure (McCullagh and Nelder 1989) with the addition of an extra term responsible to capture a possible serial dependence in the conditional mean which, conditionally on $\mathscr{F}_{t-1}$, is assumed to follow the ARMA(p, q)-like structure as

$$g(\mu_t) = \eta_t = \alpha + \boldsymbol{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{j=1}^{q} \theta_j r_{t-j}, \tag{4}$$

where η_t is the linear predictor, α is an intercept, $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_r)^\top \in \mathbb{R}^r$ is the vector of parameters related to the covariates, $\boldsymbol{\varphi} := (\varphi_1, \ldots, \varphi_p)^\top \in \mathbb{R}^p$ and $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_q)^\top \in \mathbb{R}^q$ are the autoregressive and moving average parameters, respectively, and $r_t := y_t - \mu_t$ is the moving average error term. The proposed IβARMA(p, q) is defined by (1) and (4).

Observe that (4) is similar to the systematic component in Rocha and Cribari-Neto (2009), but with the error and autoregressive term in the response's level. With this specification we avoid problems with $g(y_t)$, when y_t equals 0 or 1, which occurs with positive probability in the IβARMA(p, q) processes, otherwise inevitable when the deterministic structure is defined on the predictor level, i.e., when $r_t = g(y_t) - g(\mu_t)$. This framework, using autoregressive terms in the response's level, is widely considered in other classes of models, such as, for instance, for binary outcomes in Markov regression models for times series (Zeger and Qaqish 1988), time series following the canonical form of the exponential family of distributions (Kedem and Fokianos 2002; Fokianos and Kedem 2004), and for discrete double bounded time series using beta binomial distribution (Palm et al. 2021). One interesting consequence of using the error term in the response's level is that the ARMA-like structure does not follow the usual "rules" of a classical ARMA model. This is so because, although (4) resembles an ARMA structure, it cannot actually be rewritten as an ARMA-like difference equation since the left hand side of (4) is on the predictive level while the autoregressive structure and error term are not. Some consequences of this fact is that high values of the autoregressive coefficients are not necessarily associated to non-stationary or explosive behavior, as in usual ARMA models.

Remark 1 A clarification about the notation. Although we are denoting random variables as small letters, in (4) $\boldsymbol{x}_t^\top$ may contain non-random terms (such as a trend) as well as random potentially time-dependent components. Also, the term $\boldsymbol{x}_t^\top$ contains the covariates that are known up to time t. If $\boldsymbol{x}_t^\top$ is deterministic, then we typically know $\boldsymbol{x}_t^\top$ at time t (and all subsequent ones). However, if the covariates are random, its values are typically known only up to time $t-1$. As a convention, the notation in (4) is understood in this fashion.

2.1 Partial likelihood estimation

Parameter estimation in the proposed IβARMA model is carried on via partial maximum likelihood estimation (PMLE) (Cox 1975). The PMLE is most useful when, at each time step, the knowledge regarding a given model can be updated in a sequential way. GLM-like models for time series typically present a natural nested structure such as the one provided by (1) and (4), where the knowledge regarding model is updated in a sequential way, quantified by $\mathscr{F}_t$, via the inclusion $\mathscr{F}_{t-1} \subset \mathscr{F}_t$, $\forall t$. In this situation, PMLE allows for sequential conditional inference, naturally accommodating autoregressive components, time-dependent random covariates and any type of interaction between them (Kedem and Fokianos 2002). For practical purposes, PMLE is very similar to conditional likelihood and the computational implementation of the both methods is almost identical.

Let $y_1, \ldots, y_n$ be a sample from an IβARMA(p, q) process, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be the observed covariates, and $\boldsymbol{\gamma} := \big(\alpha_0, \alpha_1, \phi, \alpha, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top\big)^\top \in \Omega := \Psi \times (0, \infty) \times \mathbb{R}^{p+q+r+1}$ denote the κ-dimensional parameter vector, where $\Psi := \{x, y \in [0, 1] : 0 \leq x + y \leq 1\} \subset [0, 1]^2$ and $\kappa := (p + q + r + 4)$ is the number of model parameters. We shall denote (1) by $f(y_t; \boldsymbol{\gamma} \mid \mathscr{F}_{t-1})$. We observe that the partial log-likelihood function for $\boldsymbol{\gamma}$, conditionally to $\mathscr{F}_{t-1}$, is null for the first $m = \max(p, q)$ values of t, and hence we have

$$\ell(\boldsymbol{\gamma}) := \sum_{t=m+1}^{n} \log\big(f(y_t; \boldsymbol{\gamma} \mid \mathscr{F}_{t-1})\big) = \sum_{t=m+1}^{n} \ell_t(\boldsymbol{\gamma}), \tag{5}$$

where

$$\begin{aligned}\ell_t(\boldsymbol{\gamma}) : = &\big[\log(\alpha_0) + \log(1-\mu_t)\big]I_0(y_t) + \big[\log(\alpha_1) + \log(\mu_t)\big]I_1(y_t)\\ &+ \Big[\log(c_t) + \log\big(\Gamma(\phi)\big) - \log\big(\Gamma(\nu_t\phi)\big) - \log\Big(\Gamma\big([1-\nu_t]\phi\big)\Big)\\ &+ (\nu_t\phi - 1)\log(y_t) + \big([1-\nu_t]\phi - 1\big)\log(1-y_t)\Big]I_{(0,1)}(y_t),\end{aligned}$$

with c_t and ν_t given by (2). As a convention, the indicator function takes priority in evaluating any expression. Observe that if either α_0 or α_1 are zero, then the respective term in (1) is dropped from the likelihood. The PMLE $\widehat{\boldsymbol{\gamma}}$ is obtained upon maximizing (5) with respect to $\boldsymbol{\gamma}$, that is, $\widehat{\boldsymbol{\gamma}} := \sup_{\boldsymbol{\gamma}\in\Omega}\{\ell(\boldsymbol{\gamma})\}$. More details about numerical optimization, the associated score vector and cumulative partial information matrix are derived in the Appendix.

3 Asymptotic theory and large sample inference

In this section, we briefly discuss large sample properties of the PMLE in the context of IβARMA processes. Hypothesis test and confidence intervals are also discussed.

3.1 Large sample theory

Rigorous asymptotic theory for the maximum likelihood estimator in the context of generalized linear models has a relatively long history. For non-canonical links the pioneer work of Fahrmeir and Kaufmann (1985) set grounds for latter development of the theory. For GLM-like dynamic models, a general theory for PMLE is presented in the works of Fokianos and Kedem (1998, 2004); Kedem and Fokianos (2002). In the particular case of the βARMA models, Rocha and Cribari-Neto (2009, 2017) present a central limit theorem for the conditional maximum likelihood, but without formally stating assumptions nor providing a proof for the result. The asymptotic theory of the PMLE for IβARMA processes is closely related to the general theory provided in Fokianos and Kedem (1998, 2004), but not exactly so because, even though the inflated beta is a member of the exponential family, it cannot be put in canonical form. However, we conjecture that under suitable conditions, closely related to the ones presented in Fokianos and Kedem (2004), the PMLE is consistent and asymptotically normally distributed, so that, for large n,

$$\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \sim \mathcal{N}_\kappa\big(\mathbf{0}, K(\boldsymbol{\gamma}_0)^{-1}\big), \tag{6}$$

where $\boldsymbol{\gamma}_0$ denotes the true parameter and $K(\boldsymbol{\gamma}_0)$ is a suitable positive definite invertible matrix, $\mathcal{N}_\kappa(\mathbf{0}, \boldsymbol{\Sigma})$ denotes the κ-variate normal distribution with null mean vector $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. In finite sample, $K(\boldsymbol{\gamma})$, is approximated by the cumulative partial information matrix presented in the Appendix A.2.

3.2 Confidence interval and hypothesis test

Let γ_j and $\widehat{\gamma}_j$ denote the jth component of $\boldsymbol{\gamma}$ and of the PMLE $\widehat{\boldsymbol{\gamma}}$ based on a sample of size n of an IβARMA(p, q), respectively. From approximation (6),

$$\{K_n(\widehat{\boldsymbol{\gamma}})^{jj}\}^{-1/2}(\widehat{\gamma}_j - \gamma_j) \sim \mathcal{N}(0, 1),$$

holds for large n, where $K_n(\widehat{\boldsymbol{\gamma}})^{jj}$ is the jth diagonal element of $K_n(\widehat{\boldsymbol{\gamma}})^{-1}$ and $K_n(\widehat{\boldsymbol{\gamma}})$ is the cumulative partial information matrix presented in the Appendix A.2 evaluated at the PMLE $\widehat{\boldsymbol{\gamma}}$. Let z_δ represent the δ standard normal quantile. An asymptotic approximate $100(1-\delta)\%$, $0 < \delta < 1/2$, confidence interval for γ_j, $j = 1, \ldots, \kappa$, is given by

$$\Big[\widehat{\gamma}_j - z_{1-\delta/2}\big(K_n(\widehat{\boldsymbol{\gamma}})^{jj}\big)^{1/2},\ \widehat{\gamma}_j + z_{1-\delta/2}\big(K_n(\widehat{\boldsymbol{\gamma}})^{jj}\big)^{1/2}\Big].$$

The asymptotic normality of the PMLE provides the means to conduct hypothesis testing based on asymptotic versions of well-known statistics. We shall consider hypothesis of the form: $\mathcal{H}_0 : \gamma_j = \gamma_j^0$ against $\mathcal{H}_1 : \gamma_j \neq \gamma_j^0$. The traditional z statistic is given by Pawitan (2001)

$$z = \frac{\widehat{\gamma}_j - \gamma_j^0}{\sqrt{K_n(\widehat{\boldsymbol{\gamma}})^{jj}}}.$$

Under $\mathcal{H}_0$, the limiting distribution of z is standard normal. It is also possible to perform more general hypothesis testing inference using the likelihood ratio (Neyman and Pearson 1928), Rao's score (Rao 1948), Wald (Wald 1943), and gradient (Terrell 2002) statistics. As long as a normal approximation like (6) holds, the technique presented in Fahrmeir (1987) can be applied to shown that, under the null hypothesis, such test statistics are asymptotically chi-squared distributed, with the same degree of freedoms as their counterparts under independence.

4 Diagnostic analysis and forecasting

In this section, we shall discuss results related to model selection criteria, residual analysis as well as in-sample and out-of-sample predictions for the proposed IβARMA models. Diagnostic checks can be applied to a fitted model to determine whether it captures the true data dynamics. A fitted model that passes all diagnostic checks can then be used for forecasting.

4.1 Model selection criteria

Model selection criteria in the class of IβARMA model can be based on Akaike's information criterion (AIC) (Akaike 1974). Following the idea of modified AIC (MAIC) proposed in Bayer et al. (2018), we consider the following criterion:

$$\text{MAIC} = 2\kappa - 2\hat{\ell}_*, \tag{7}$$

where $\hat{\ell}_* := \frac{n}{n-m}\ell(\hat{\boldsymbol{\gamma}})$ and κ is the dimension of $\boldsymbol{\gamma}$. Notice that when comparing models of different orders (with different values of m), $\hat{\ell}_*$ can be interpreted as the sum of n terms. Therefore, the MAIC does not incorrectly penalize models with larger values of m (Bayer et al. 2018). Among a set of competitor models, we favor the model with smallest MAIC. Replacing the term 2κ in (7) by $\log(n)\kappa$ and $\log\big(\log(n)\big)\kappa$ we obtain modified versions of the Schwarz information criterion (MSIC) (Schwarz 1978) and Hannan and Quinn information criterion (HQ) (Hannan and Quinn 1979), respectively.

4.2 Residuals

Residual analysis is important for checking whether the selected model provides a good fit to the data (Kedem and Fokianos 2002). Visual inspection of a time series residuals plot is an indispensable first step when assessing goodness-of-fit (Box et al. 2015). In the context of GLM-like dynamic models there are several ways to define residuals. For the proposed IβARMA(p, q), following Ospina and Ferrari (2012) for inflated beta regression, we will consider the randomized quantile residual given by

$$r_t^q = \Phi^{-1}(u_t),$$

with $\Phi(\cdot)$ being the standard normal distribution function, u_t a uniform random variable on the interval $(a_t, b_t]$, where $a_t = \lim_{y\uparrow y_t} F(y; \widehat{\boldsymbol{\gamma}} \mid \mathscr{F}_{t-1})$, $b_t = F(y_t; \widehat{\boldsymbol{\gamma}} \mid \mathscr{F}_{t-1})$, and $F(y_t; \boldsymbol{\gamma} \mid \mathscr{F}_{t-1})$ is the cumulative distribution function of an inflated beta distribution conditional to $\mathscr{F}_{t-1}$, given in (3).

Under the correct model specification, the residuals are expected to behave as white noise, i.e., they should be serially uncorrelated and follow a zero mean and constant variance process (Kedem and Fokianos 2002; Bayer et al. 2018). Hence, it is expected that 95% of residuals autocorrelations lie inside the (asymptotic) interval $\left[-1.96/\sqrt{n-m}, 1.96/\sqrt{n-m}\right]$. Plots of the residuals autocorrelation function (ACF) with horizontal lines at $\pm 1.96/\sqrt{n-m}$ can be used for assessing whether the residuals display white noise behavior (Kedem and Fokianos 2002). It is also possible to test if the residual autocorrelations are equal to zero using, for example, adapted versions of tests given by Ljung and Box (1978) and Monti (1994) (see also Scher et al. 2020).

4.3 Prediction and forecasting

For $t = m+1, \ldots, n$, in-sample estimates of μ_t (denoted $\widehat{\mu}_t$ and called predicted values), are obtained by replacing $\boldsymbol{\gamma}$ with its PMLE, $\widehat{\boldsymbol{\gamma}}$, and r_t by $y_t - \widehat{\mu}_t$ in (4). For $h = 1, 2, \ldots$, h-step ahead forecasts can be computed as

$$\widehat{\mu}_{n+h} = g^{-1}\left(\widehat{\alpha} + \boldsymbol{x}_t^{\top}\widehat{\boldsymbol{\beta}} + \sum_{i=1}^{p} \widehat{\varphi}_i \left[y_{n+h-i}\right] + \sum_{j=1}^{q} \widehat{\theta}_j \left[r_{n+h-j}\right]\right),$$

where $[y_t] := y_t I_{(-\infty,n]}(t) + \widehat{\mu}_t I_{(n,\infty)}(t)$ and $[r_t] := (y_t - \widehat{\mu}_t) I_{(-\infty,n]}(t)$.

5 Monte Carlo simulation study

In this section, we investigate the finite sample performance of the proposed PMLE approach in the context of zero, one, and zero-one IβARMA model through a Monte Carlo simulation study. In all simulations, we consider an IβARMA model with systematic component given by

$$g(\mu_t) = \alpha + \beta_1 x_t + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \theta_1 r_{t-1} + \theta_2 r_{t-2},$$

with $g(\cdot)$ as the logit link and x_t are randomly drawn from a uniform (0, 1) distribution, kept fixed during the simulations. We consider three different scenarios: (i) the zero IβARMA(1, 1), in which we take $\alpha = -1.5$, $\beta_1 = 1.0$, $\varphi_1 = 1.5$, $\varphi_2 = 0$, $\theta_1 = -1$, $\theta_2 = 0$, $\phi = 30$, and $\alpha_0 = 0.07$, (ii) the one IβARMA(2,2) with $\alpha = 9.675$, $\beta_1 = 0.636$, $\varphi_1 = 0$, $\varphi_2 = -6.929$, $\theta_1 = 0$, $\theta_2 = 9.304$, $\phi = 72.991$, and $\alpha_1 = 0.017$, and (iii) the zero-one IβARMA(1,1) for which $\alpha = -0.5$, $\beta_1 = 1.0$, $\varphi_1 = 1.5$, $\varphi_2 = 0$, $\theta_1 = -1$, $\theta_2 = 0$, $\phi = 20$, $\alpha_0 = 0.07$, and $\alpha_1 = 0.08$. For the one IβARMA(2, 2) scenario, the parameter values were selected in order to match the ones found in the real data application considered in the Sect. 6.1. As we explained before, since the error term is in the response's level, although (4) is inspired by and resembles an ARMA-like structure, it does not follow the usual "rules" followed by ARMA models, so that it is natural to have high values of φ_1. We consider sample sizes $n \in \{100, 300, 500\}$. All codes were written in R version 3.5.2 (R Core Team 2020) by the authors and are available at https://github.com/fabiobayer/IBARMA.

Tables 1, 2, and 3 present the results based on R = 10,000 replications for the zero IβARMA(1, 1), one IβARMA(2, 2), and zero-one IβARMA(1, 1) scenarios, respectively. Presented in the tables are the mean, standard error (SE), and relative bias (RB) of the point estimators and 95% approximate confidence interval coverage rate (CR). For arbitrary estimates $\hat{\varrho}_1, \ldots, \hat{\varrho}_R$ of a quantity ϱ, RB is defined as $\text{RB} = \left[\frac{1}{R}\sum_{i=1}^{R}\frac{\hat{\varrho}_i - \varrho}{\varrho}\right] \times 100\%$, while CR is defined as the proportion of times that the interval contained the true parameter value.

Table 1 Monte Carlo simulation results for point estimation and coverage rate (CR) for the zero IβARMA(1,1)

n	Measures	α −1.5	β_1 1	φ_1 1.5	θ_1 −1	ϕ 30	α_0 0.07
100	Mean	−1.489	1.001	1.469	−1.032	32.078	0.070
	SE	0.189	0.142	0.496	0.645	4.794	0.032
	RB	−0.703	0.069	−2.036	3.225	6.927	0.132
	CR	0.914	0.937	0.911	0.895	0.948	0.919
300	Mean	−1.496	1.002	1.487	−1.004	30.667	0.070
	SE	0.108	0.074	0.262	0.330	2.543	0.019
	RB	−0.270	0.216	−0.839	0.411	2.225	−0.355
	CR	0.941	0.946	0.939	0.932	0.951	0.925
500	Mean	−1.497	1.001	1.493	−1.002	30.375	0.070
	SE	0.086	0.058	0.206	0.255	1.957	0.014
	RB	−0.170	0.079	−0.494	0.162	1.248	−0.311
	CR	0.945	0.948	0.946	0.940	0.951	0.941

Table 2 Monte Carlo simulation results for point estimation and coverage rate (CR) for the one IβARMA(2,2)

n	Measures	α 9.675	β_1 0.636	φ_2 −6.929	θ_2 9.304	ϕ 72.991	α_1 0.017
100	Mean	9.54	0.528	−6.666	8.902	77.765	0.021
	SE	3.352	1.902	3.723	4.786	12.051	0.012
	RB	−1.399	−16.987	−3.802	−4.326	6.541	25.987
	CR	0.883	0.924	0.878	0.86	0.949	0.998
300	Mean	9.668	0.609	−6.891	9.212	74.547	0.017
	SE	2.324	1.273	2.518	3.154	6.378	0.008
	RB	−0.068	−4.173	−0.544	−0.989	2.132	0.684
	CR	0.917	0.94	0.916	0.909	0.952	0.866
500	Mean	9.746	0.596	−6.961	9.302	73.936	0.017
	SE	1.975	1.598	2.208	2.711	5.206	0.006
	RB	0.732	−6.216	0.462	−0.026	1.294	1.155
	CR	0.927	0.943	0.926	0.925	0.953	0.900

In all scenarios, both SE and RB decrease as the sample size increases, which provides evidence of the consistency of the PMLE. For instance, for β_1 estimator in Table 1, the SE is equal to 0.142 for $n = 100$ and equal to 0.058 for $n = 500$, while for $\hat{\phi}$ we have a RB of 6.927% for $n = 100$ and 1.248% for $n = 500$. Table 2 shows similar results for the one-inflated scenario. Table 3 also shows a similar scenario for α_0 and α_1 estimators. For instance, the relative bias also decreases considerably as the sample size increases. For example, with $n = 100$, the estimator of α_0 has RB $= -8.936\%$ and the estimator of α_1 presents RB $= -0.140\%$. For $n = 500$, the bias of $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are reduced to -0.522% and -0.055%, respectively. In general, coverage rates are close to the nominal value of 95%. For instance, for $n = 500$ the empirical coverage of the confidence intervals for φ_1 are,

Table 3 Monte Carlo simulation results for point estimation and coverage rate (CR) for the zero-one IβARMA(1,1)

n	Measures	α −0.5	β_1 1	φ_1 1.5	θ_1 −1	ϕ 20	α_0 0.070	α_1 0.080
100	Mean	−0.430	1.008	1.405	−0.964	21.413	0.064	0.080
	SE	0.585	0.195	0.800	0.942	3.309	0.446	0.031
	RB	−13.990	0.829	−6.300	−3.621	7.063	−8.936	−0.140
	CR	0.891	0.933	0.889	0.874	0.948	0.833	0.931
300	Mean	−0.471	1.002	1.462	−0.976	20.473	0.069	0.080
	SE	0.336	0.105	0.458	0.515	1.753	0.029	0.018
	RB	−5.722	0.239	−2.533	−2.397	2.367	−0.921	−0.013
	CR	0.931	0.941	0.928	0.927	0.950	0.887	0.940
500	Mean	−0.483	1.002	1.477	−0.982	20.285	0.070	0.080
	SE	0.266	0.079	0.360	0.399	1.323	0.022	0.014
	RB	−3.426	0.187	−1.546	−1.760	1.423	−0.522	−0.055
	CR	0.937	0.949	0.937	0.936	0.950	0.929	0.939

respectively, 94.6% (zero inflated—Table 1) and 93.7% (zero-one inflated—Table 3), and CR = 92.6% for φ_2 in the one-inflated scenario (Table 2). As for the precision parameter estimator, in all scenarios the estimation is very good, improving as the sample size grows. In general, the numerical results show good properties of the PMLE approach even in moderate sample sizes.

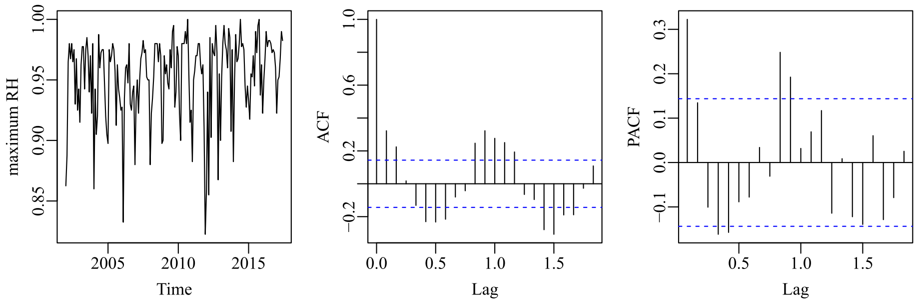
6 Practical examples

This section presents two empirical applications related to hydro-environmental time series. The first one considers relative air humidity data, while the second one is about the percentage of useful volume of a hydroelectric plant's water reservoir.

6.1 Relative air humidity application

The relative air humidity (RH) is an important meteorological characteristic to public health, irrigation scheduling design, and hydrological studies. The importance of modeling and forecasting environmental variables such as RH is largely discussed in literature (Grassly and Fraser 2006; Bayer et al. 2017). As the humidity variability has been linked to several infectious diseases (Tamerius et al. 2013), it is important to model properly in order to make accurate forecasts useful to policymakers.

To showcase the proposed IβARMA model, we present an application to the maximum RH in Santa Maria, RS, Brazil. The actual time series consists of monthly maximum values from January 2002 through June 2018, yielding a total sample size $n = 198$, and were obtained from Brazilian National Institute of Meteorology (INMET) (2018). The last 12 observations were reserved for out-of-sample forecast purposes, so that only the first $n = 186$ observations were used for model identification and estimation. Figure 2 presents the maximum RH time series plot as well as the plots of the sample ACF and partial ACF (PACF). Figure 2a shows



(a) Time series (b) Sample ACF (c) Sample PACF

Fig. 2 Time series, ACF and PACF plots for the maximum RH in Santa Maria, Brazil

Table 4 Fitted IβARMA model for the maximum RH in Santa Maria, Brazil

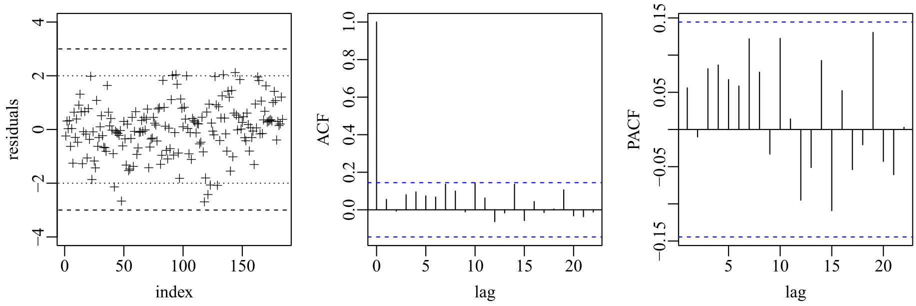
Parameter	Estimate	Std. error	z stat.	Pr(>| z |)
α	9.6750	2.7559	3.5107	0.0004
φ_2	−6.9288	2.8901	−2.3974	0.0165
θ_2	9.3044	3.0828	3.0182	0.0025
β_1	0.6361	0.0837	7.6000	< 0.0001
ϕ	72.9910	7.8699	9.2747	< 0.0001
α_1	0.0170	0.0098	1.7452	0.0810

MAIC = −839.0387 MSIC = −816.4585
Ljung–Box (DF = 18): $Q = 24.195$ (p−value = 0.149)

that we have three observed values equal to 1 in the observations 105, 150, and 166. The RH reaches 1 (100%) when the air is saturated. The visible sinusoidal pattern in the ACF and PACF (Fig. 2b and c) indicates the presence of a seasonal dynamic. The yearly variability is handled in the sense of harmonic regression approach (Bloomfield 2013), by introducing seasonal covariates $\big(\sin(2\pi t/12), \cos(2\pi t/12)\big)$, for $t \in \{1, \ldots, n\}$. However, only one of them was significant and, thus, the final fitted model only contains $x_t = \cos(2\pi t/12)$ as covariate.

In order to select an adequate IβARMA model for the data, we systematically tried different orders p and q and select the one whose parameter were all significant and whose residual did not reject the null hypothesis in the Ljung–Box test. All tests were conducted at 10% significance level. Table 4 presents parameter estimates, standard errors, z statistics and p-values for the selected IβARMA model for the maximum RH. An IβARMA(2,2) was selected, but the parameters φ_1 and θ_1 were not statistically significant. The MAIC and MSIC criteria and the Ljung–Box test (using 20 lags, with corrected degrees of freedom in view of Scher et al. (2020)) considering the randomized quantile residuals are also presented. Anyone familiar with ARMA modeling might be surprised with the magnitude of φ_2, but, as we explained before, this is so because the particular form (4) assumes given that the error and AR terms are in the response's level.

Figure 3 contains some diagnostic plots based on the residual. Figure 3a shows a random displacement with constant variance and no other visible pattern in the residuals. Figure 3b and c evidences that the residual autocorrelations are not significantly different from zero. Hence, all plots and tests indicate that the fitted model is adequate and can be used for out-of-



(a) Residual (b) Residual ACF (c) Residual PACF

Fig. 3 Diagnostic plots for the fitted IβARMA model to the RH data based on randomized quantile residuals

sample forecasting. In Fig. 4a, we plot the data (solid line) along with in-sample and 12-step ahead forecast (dashed red line).

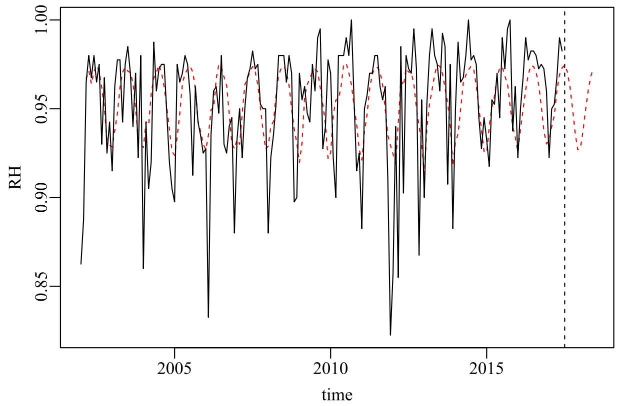
For comparison purposes, we also fit an additive Holt–Winters with automatic adjustment for smoothing constants and a βARMA model. Observe that values equal to 1 imply an asymptote in the likelihood function of a βARMA model. To solve this problem, we replaced them with 0.999. The βARMA model is then fitted to this new time series (with ones replaced by 0.999) by using the same procedure as described above for the IβARMA model. It is important to note that while this approach solves the numerical problem in the likelihood function, it does not solve the inflation problem. When replacing values equal to one for 0.999, an inflation in this value is induced. This reflects on the estimates of the parameters of the dynamic structure, since the probability mass at 0.999 exceeds what is allowed by the beta distribution, which is an absolutely continuous distribution. Figure 4b shows the out-of-sample forecasts for the fitted model. We can see that the forecast values by the IβARMA model are closer to the actual values than the competing models.

In what follows, we consider an out-of-sample forecasting analysis based on a cross-validation study using a rolling window approach. We consider a training data window with n_T observations moving over the time. For each training window $y_i, \ldots, y_{i+n_T}$, with $i \in \{1, \ldots, N\}$, a test window of size h, $y_{i+n_T+1}, \ldots, y_{i+n_T+h}$, is considered. Then, the forecasting accuracy measures for the h-steps-ahead forecasts produced, with $h \in \{1, \ldots, H\}$. For each h, we have N forecast values that are used to calculate the following out-of-sample accuracy measures: mean square error (MSE), mean absolute error (MAE), and mean directional accuracy (MDA). The first two measures (MSE and MAE) are widely used quantitative measures for forecasting evaluation, while MDA is a popular metric in economics and finance field, providing the probability that the model detect the correct direction (upward or downward) of the time series. The forecast measures are given by

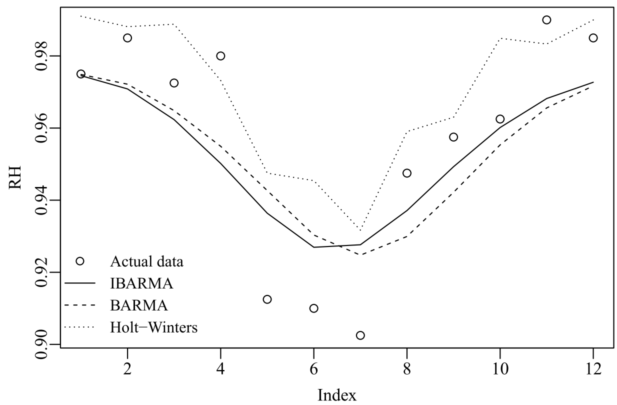
$$\mathrm{MSE}_h := \frac{1}{N}\sum_{i=1}^{N}(y_{i+n_T+h} - \widehat{\mu}_{i+n_T}^{h})^2,$$

$$\mathrm{MAE}_h := \frac{1}{N}\sum_{i=1}^{N} \mid y_{i+n_T+h} - \widehat{\mu}_{i+n_T}^{h} \mid,$$

$$\mathrm{MDA}_h := \frac{1}{N}\sum_{i=1}^{N} I\big((y_{i+n_T+h} - y_{i+n_T})(\widehat{\mu}_{i+n_T}^{h} - y_{i+n_T}) > 0\big),$$



(a) Fitted values



(b) Out-of-sample comparison

Fig. 4 In-sample and out-of-sample forecasts for maximum RH in Santa Maria, Brazil

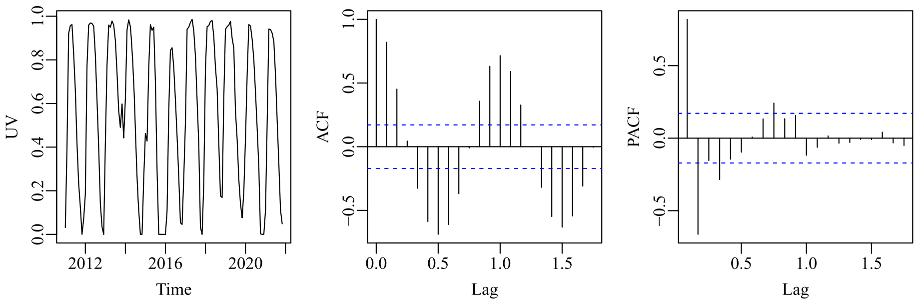
where $I(\cdot)$ is the indicator function and $\widehat{\mu}_t^h$ is the forecast for y_{t+h} using the information available up to time t. For the RH time series with 198 observations, we fixed $n_T = 110$, $N = 76$, and $H = 12$.

6.2 Percentage of useful volume application

In this section, we present an empirical analysis of the monthly UV (percentage of useful volume) time series of the Samuel water reservoir in the Rondônia State, Brazil. The UV is defined as the perceptual volume of water between the maximum and minimum normal operating levels (Operador Nacional do Sistema Elétrico 2022; Sagrillo et al. 2021). The hydropower is the primary source of electric energy in Brazil, with an installed hydroelectric capacity representing about 64% of the National Interconnected System (SIN) in 2021 Sagrillo et al. (2021). Thus, the water resource management is important for the generation of electricity in the country. All code were written in `R` and can be found at https://github.com/fabiobayer/IBARMA.

Table 5 Out-of-sample forecasting measures for the models to RH data in Santa Maria, Brazil. The best results are highlighted in bold

Measure	Model	h				
		1	3	6	9	12
$\text{MSE}_h\,(\times 10^4)$	IβARMA	**8.60**	**8.74**	**9.03**	10.22	**7.31**
	βARMA	8.93	9.15	9.44	**9.73**	7.65
	Holt–Winters	9.77	9.43	10.96	11.70	10.41
$\text{MAE}_h\,(\times 10^4)$	IβARMA	**210.79**	**214.81**	**220.85**	234.49	**208.30**
	βARMA	214.80	222.84	229.81	**234.26**	212.30
	Holt–Winters	234.81	224.54	243.34	247.67	230.81
MDA_h	IβARMA	35.53	**60.53**	**69.74**	**59.21**	38.16
	βARMA	**38.16**	**60.53**	67.11	53.95	32.90
	Holt–Winters	**38.16**	57.90	**69.74**	**59.21**	**40.79**



(a) Time series (b) Sample ACF (c) Sample PACF

Fig. 5 Time series, ACF and PACF plots for the UV in Rondônia, Brazil

The times series under analysis consists of 143 monthly UV observations from January 2011 to November 2022 obtained from Operador Nacional do Sistema Elétrico (2022). As in the first application, the last 12 observations were set aside for forecasting purposes. In this case, $n = 131$ observations are used for fitting the model. Among these 131 observations, 12 of them are equal to zero, representing about 9.16% of the sample size. UV values equal to zero occurs when the amount of water left in a reservoir cannot be used to make hydroelectric power, known as dead pool volume.

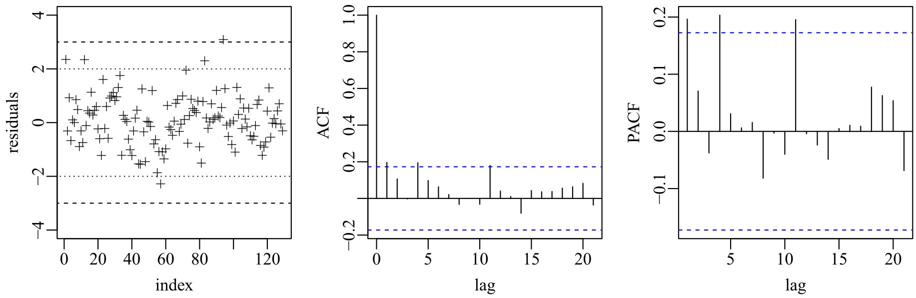
Figure 5 shows the UV times series, the sample ACF and PACF. The sinusoidal pattern suggests the presence of a deterministic seasonal component, which is modeled introducing two seasonal covariates $\big(\sin(2\pi t/12), \cos(2\pi t/12)\big)$, for $t \in \{1, \ldots, n\}$. Table 6 presents the fitted model for the UV time series, selected by the same methodology described in the previous application. An IβARMA(1,2) with $\theta_1 = 0$ (not statistically significant) was selected. All tests were conducted at 10% significance.

The Ljung–Box test in Table 6 and diagnostic plots in Fig. 6 are evidence that the model is well fitted. Figure 7a shows the actual UV times series and in-sample and 12-step ahead forecast, while Fig. 7b presents the out-of-sample forecast values for IβARMA, βARMA, and Holt–Winters models. For fitting βARMA model the zeros were replaced by 0.001. The predicted and forecast values of the IβARMA show that the proposed model was able to

Table 6 Fitted IβARMA model for the UV in Rondônia, Brazil

Parameter	Estimate	Std. error	z stat.	Pr(>| z |)
α	−2.3997	0.2127	−11.2841	< 0.0001
φ_1	4.7892	0.3966	12.0749	< 0.0001
θ_2	−1.9773	0.7553	−2.6181	0.0088
β_1	−0.8690	0.1273	−6.8289	< 0.0001
β_2	−0.9641	0.1371	−7.0334	< 0.0001
ϕ	16.9173	2.2730	7.4429	< 0.0001
α_0	0.2082	0.0507	4.1096	< 0.0001

MAIC = −196.4671 MSIC = −173.4655
Ljung–Box (DF = 18): $Q = 22.615$ (p − value = 0.2058)



(a) Residual (b) Residual ACF (c) Residual PACF

Fig. 6 Diagnostic plots for the fitted IβARMA model to the UV data based on randomized quantile residuals

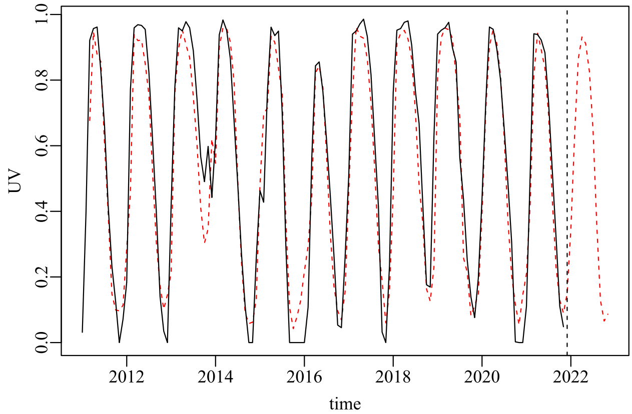
accommodate the general pattern of the times series, being useful for inflated double bounded times series modeling.

Finally, we perform an out-of-sample forecasting analysis using a cross-validation procedure in the same lines as the one presented in Sect. 6.1. We set $n_T = 100$, $N = 31$, and $H = 12$, where $n_T + N + H = 143$. The MSE$_h$, MAE$_h$, and MDA$_h$, with $h \in \{1, 3, 6, 9, 12\}$ are presented in Table 7. We can see that the IβARMA model outperforms the competitors in almost all measures. Particularly, considering the MDA, the proposed model outperforms the other models for almost all forecast horizons, being much superior for larger h values.

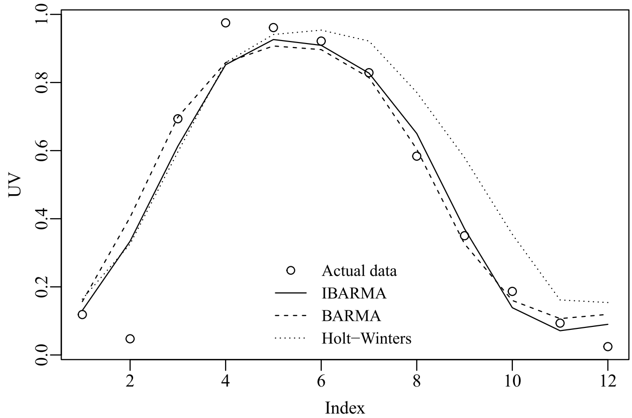
7 Conclusion

In this article, we have proposed a new dynamical model for modeling and forecasting time series that are double bounded in the intervals (0,1], [0,1) or [0,1]. The IβARMA model is built upon the assumption that the conditional distribution of the response variable, given its past behavior, is the inflated beta (Bayes and Valdivieso 2016). This distribution allows for modeling double bounded processes assuming values in the unit interval containing excess of zeros and/or ones.

The contribution of the paper is twofold. First, we proposed an observation-driven model for which the random component follows an inflated beta distribution over time while the deterministic part considers an autoregressive moving average structure and a set of regressors connected to the conditional mean through a link function. Second, we propose a partial like-



(a) Fitted values



(b) Out-of-sample comparison

Fig. 7 In-sample and out-of-sample forecasts for the UV data in Rondônia, Brazil

Table 7 Out-of-sample forecasting measures for the UV data in Rondônia, Brazil. The best results are highlighted in bold

Measure	Model	h				
		1	3	6	9	12
$MSE_h\,(\times 10^4)$	IβARMA	65.86	**165.65**	**161.89**	**147.48**	**133.36**
	βARMA	**55.24**	246.23	241.46	238.22	218.14
	Holt–Winters	130.72	539.15	623.68	377.44	216.62
$MAE_h\,(\times 10^4)$	IβARMA	661.88	**1053.56**	**999.60**	**963.03**	**892.65**
	βARMA	**586.42**	1170.42	1158.41	1149.54	1096.73
	Holt–Winters	865.33	1914.81	2056.95	1546.09	1194.14
MDA_h	IβARMA	**90.32**	**93.55**	**93.55**	**80.64**	**35.48**
	βARMA	77.42	87.10	93.55	74.19	9.68
	Holt–Winters	64.52	74.19	93.55	80.64	9.68

lihood approach for parameter estimation presenting closed-form expressions for the score vector and Fisher information matrix. We briefly discuss large sample inference and consider the problem of interval estimation and hypotheses testing. Diagnostic tools and forecasting method are developed and discussed. In addition, we provided Monte Carlo evidence on the finite sample accuracy of point estimation and confidence interval. Finally, two empirical illustrations are presented and discussed to evidence the applicability of the proposed model for forecasting inflated double bounded data. The results highlight the usefulness of the proposed model as well as its competitiveness against other established models.

Acknowledgements We gratefully acknowledge partial financial support from CNPq and FAPERGS, Brazil. The comments and suggestions of the anonymous referee and the Associated Editor are gratefully acknowledged.

Funding Funding was provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (310617/2020-0) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (21/2551-0002048-2).

Data availability The RH time series is publicly available on the Brazilian National Institute of Meteorology (INMET) website and the UV data is available on the Operador Nacional do Sistema Elétrico (ONS) website.

Appendix A Score vector and cumulative partial information matrix

In this appendix, we shall derive the partial score vector and the cumulative partial information matrix from (5). These are useful for the asymptotic theory and inference as well as numerical considerations.

A.1 Score vector and optimization algorithm

To obtain the partial score vector, we shall need to obtain the derivative of the log-likelihood $\ell(\boldsymbol{\gamma})$ given in (5) with respect to each coordinate γ_j, with $j \in 1, \ldots, \kappa$, of the parameter $\boldsymbol{\gamma}$. To obtain the derivative of $\ell(\boldsymbol{\gamma})$ with respect to α_i, $i = 0, 1$, observe that, in view of (5),

$$\frac{\partial c_t}{\partial \alpha_i} = (\mu_t - 1)I_0(i) - \mu_t I_1(i) \quad \text{and} \quad \frac{\partial \nu_t}{\partial \alpha_i} = \frac{(-1)^i(\alpha_{1-i} - 1)(\mu_t - 1)\mu_t}{c_t^2}.$$

Now, for $i = 0, 1$, it is straightforward to show that

$$\frac{\partial \ell(\boldsymbol{\gamma})}{\partial \alpha_i} = \sum_{t=m+1}^{n} \frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \alpha_i} = \frac{1}{\alpha_i} I_i(y_t) + \left(\frac{1}{c_t} \left[\frac{\partial c_t}{\partial \alpha_i} \right] + \phi(y_t^* - \mu_t^*) \left[\frac{\partial \nu_t}{\partial \alpha_i} \right] \right) I_{(0,1)}(y_t),$$

where $y_t^* := \log\left(\frac{y_t}{1-y_t}\right)$, $\mu_t^* := \psi(\nu_t \phi) - \psi\big((1-\nu_t)\phi\big)$ and $\psi : (0, \infty) \to \mathbb{R}$ is the digamma function defined as $\psi(z) = \frac{d}{dz} \log\big(\Gamma(z)\big)$. The derivative with respect to ϕ is easy to obtain:

$$\frac{\partial \ell(\boldsymbol{\gamma})}{\partial \phi} = \sum_{t=m+1}^{n} \Big[\nu_t(y_t^* - \mu_t^*) + \log(1 - y_t) - \psi\big((1-\nu_t)\phi\big) + \psi(\phi) \Big] I_{(0,1)}(y_t).$$

For the remaining parameters, i.e., for $j \in 4, \ldots, \kappa$, by the chain rule, and since $\eta_t = g(\mu_t)$, $\dfrac{d\mu_t}{d\eta_t} = \dfrac{1}{g'(\mu_t)}$, so that

$$\frac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_j} = \sum_{t=m+1}^{n} \frac{1}{g'(\mu_t)} \frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \mu_t} \frac{\partial \eta_t}{\partial \gamma_j}. \tag{A1}$$

Observe that $\frac{\partial v_t}{\partial \mu_t} = (\alpha_0 - 1)(\alpha_1 - 1)c_t^{-2}$ and

$$\frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \mu_t} = \left(\frac{\alpha_0 - \alpha_1}{c_t} + \phi(y_t^* - \mu_t^*)\left[\frac{(\alpha_0 - 1)(\alpha_1 - 1)}{c_t^2}\right]\right) I_{(0,1)}(y_t) + \frac{1}{\mu_t} I_1(y_t) - \frac{1}{1-\mu_t} I_0(y_t). \tag{A2}$$

Substituting (A2) into expression (A1), we obtain a simple formula that allows the computation of $\partial\ell(\boldsymbol{\gamma})/\partial\gamma_j$ for each remaining coordinate γ_j, by determining the derivatives $\partial\eta_t/\partial\gamma_j$, a much simpler task. We have

$$\frac{\partial \eta_t}{\partial \alpha} = 1 - \sum_{k=1}^{q} \theta_k \frac{1}{g'(\mu_{t-k})} \frac{\partial \eta_{t-k}}{\partial \alpha} \quad \text{and} \quad \frac{\partial \eta_t}{\partial \beta_l} = x_{tl} - \sum_{k=1}^{q} \theta_k \frac{1}{g'(\mu_{t-k})} \frac{\partial \eta_{t-k}}{\partial \beta_l},$$

where x_{tl} denotes the lth element of $\boldsymbol{x}_t$, for $l = 1, \ldots, r$. We also have, for $l = 1, \ldots, p$, and $j = 1, \ldots, q$,

$$\frac{\partial \eta_t}{\partial \varphi_l} = y_{t-l} - \sum_{k=1}^{q} \theta_k \frac{1}{g'(\mu_{t-k})} \frac{\partial \eta_{t-k}}{\partial \varphi_l} \quad \text{and} \quad \frac{\partial \eta_t}{\partial \theta_j} = r_{t-j} - \sum_{k=1}^{q} \theta_k \frac{1}{g'(\mu_{t-k})} \frac{\partial \eta_{t-k}}{\partial \theta_j}.$$

Let $T = \operatorname{diag}\{1/g'(\mu_{m+1}), \ldots, 1/g'(\mu_n)\}$, $\mathbf{a} = \left(\frac{\partial \eta_{m+1}}{\partial \alpha}, \ldots, \frac{\partial \eta_n}{\partial \alpha}\right)^\top$ and $\boldsymbol{v} = \left(\frac{\partial \ell_{m+1}(\boldsymbol{\gamma})}{\partial \mu_{m+1}}, \ldots, \frac{\partial \ell_n(\boldsymbol{\gamma})}{\partial \mu_n}\right)^\top$. Finally, let R, P, Q be the matrices with dimension $(n-m) \times r$, $(n-m) \times p$ and $(n-m) \times q$, respectively, for which the (i, j)th elements are given by

$$R_{i,j} = \frac{\partial \eta_{i+m}}{\partial \beta_j}, \quad P_{i,j} = \frac{\partial \eta_{i+m}}{\partial \varphi_j}, \quad \text{and} \quad Q_{i,j} = \frac{\partial \eta_{i+m}}{\partial \theta_j}$$

and set $U_\alpha(\boldsymbol{\gamma}) := \mathbf{a}^\top T \boldsymbol{v}$, $U_{\boldsymbol{\beta}}(\boldsymbol{\gamma}) := R^\top T \boldsymbol{v}$, $U_{\boldsymbol{\varphi}}(\boldsymbol{\gamma}) := P^\top T \boldsymbol{v}$ and $U_{\boldsymbol{\theta}}(\boldsymbol{\gamma}) := Q^\top T \boldsymbol{v}$.

For $U_{\alpha_j}(\boldsymbol{\gamma}) := \frac{\partial \ell(\boldsymbol{\gamma})}{\partial \alpha_j}$, and $U_\phi(\boldsymbol{\gamma}) := \frac{\partial \ell(\boldsymbol{\gamma})}{\partial \phi}$, then the partial score vector is given by

$$U(\boldsymbol{\gamma}) = \left(U_{\alpha_0}(\boldsymbol{\gamma}), U_{\alpha_1}(\boldsymbol{\gamma}), U_\phi(\boldsymbol{\gamma}), U_\alpha(\boldsymbol{\gamma}), U_{\boldsymbol{\beta}}(\boldsymbol{\gamma})^\top U_{\boldsymbol{\varphi}}(\boldsymbol{\gamma})^\top, U_{\boldsymbol{\theta}}(\boldsymbol{\gamma})^\top\right)^\top.$$

The PMLE of $\boldsymbol{\gamma}$, $\widehat{\boldsymbol{\gamma}}$, is obtained as a solution of the non-linear system $U(\boldsymbol{\gamma}) = \mathbf{0}$, where $\mathbf{0}$ is the null vector in $\mathbb{R}^\kappa$. There is no closed form solution for such a system and, hence, PMLE must be obtained numerically (Nocedal and Wright 1999). In this work, we use the so-called Broyden–Fletcher–Goldfarb–Shanno (BFGS) method (Press et al. 1992). In practice, to calculate $\widehat{\boldsymbol{\gamma}}$ from a sample, we initialize $r_t = 0$ and $\mu_t = 0$ for $t \le \max\{p, q\}$ and calculate μ_t and r_t for $t > \max\{p, q\}$ recursively from the data using (4). The BFGS algorithm also requires initialization of the parameters. The starting values of α, $\boldsymbol{\beta}$ and $\boldsymbol{\varphi}$ were set as the OLS estimate of

$$g(y_t) = \alpha + \boldsymbol{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^{p} \varphi_i y_{t-i} + \text{error term},$$

restricted to the observations where $y \in (0, 1)$. The vector parameter $\boldsymbol{\theta}$ is initialized as a null vector, as in Bayer et al. (2017), while inflation parameters α_0 and α_1 were initialized as the sample proportion of zeroes and ones, respectively.

A.2 Cumulative partial information matrix

In this appendix we derive the cumulative partial information matrix, given by

$$K_n(\boldsymbol{\gamma}) = -\sum_{t=m+1}^{n} \mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top} \;\Big|\; \mathscr{F}_{t-1}\right).$$

Since direct knowledge of the unconditional distribution of the proposed model is not obtainable, K_n will be the first step toward finding the asymptotic variance-covariance matrix related to the PMLE. In this case, under suitable assumptions (Fokianos and Kedem 2004), there exists a non-random information matrix, denoted by $K(\boldsymbol{\gamma})$, such that the weak convergence

$$\frac{K_n(\boldsymbol{\gamma})}{n} \underset{n\to\infty}{\longrightarrow} K(\boldsymbol{\gamma}),$$

holds, where $K(\boldsymbol{\gamma})$ is a positive definite and invertible matrix. The matrix $K(\boldsymbol{\gamma})^{-1}$ is the asymptotic variance-covariance matrix related to the PMLE, presented in (6).

For $i, j \in \{4, \ldots, \kappa\}$ (that is, $\gamma_j \notin \{\alpha_0, \alpha_1, \phi\}$), it can be shown that

$$\begin{aligned}\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j} &= \frac{\partial}{\partial \mu_t}\left(\frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \mu_t}\frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \gamma_j}\right)\frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \gamma_i}\\ &= \left[\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \mu_t^2}\frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \gamma_j} + \frac{\partial \ell_t(\boldsymbol{\gamma})}{\partial \mu_t}\frac{\partial}{\partial \mu_t}\left(\frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \gamma_j}\right)\right]\frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \gamma_i}.\end{aligned}$$

Since by Lemma A.1, $\mathbb{E}\big(\partial \ell_t(\boldsymbol{\gamma})/\partial \mu_t \mid \mathscr{F}_{t-1}\big) = 0$, we arrive at

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j} \;\Big|\; \mathscr{F}_{t-1}\right) = \frac{1}{g'(\mu_t)^2}\mathbb{E}\left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \mu_t^2} \;\Big|\; \mathscr{F}_{t-1}\right)\frac{\partial \eta_t}{\partial \gamma_j}\frac{\partial \eta_t}{\partial \gamma_i}.$$

The second-order derivatives of $\ell_t(\boldsymbol{\gamma})$ with respect to μ_t is given by

$$\begin{aligned}\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \mu_t^2} &= \left(\frac{-(\alpha_0 - \alpha_1)^2}{c_t^2} + \phi(\alpha_0 - 1)(\alpha_1 - 1)\frac{\partial}{\partial \mu_t}\left[\frac{(y_t^* - \mu_t^*)}{c_t^2}\right]\right) I_{(0,1)}(y_t)\\ &\quad - \frac{1}{\mu_t^2} I_1(y_t) - \frac{1}{(1-\mu_t)^2} I_0(y_t).\end{aligned}$$

Observe that

$$\frac{\partial \mu_t^*}{\partial \mu_t} = \frac{\partial \mu_t^*}{\partial \nu_t}\frac{\partial \nu_t}{\partial \mu_t} = \frac{\phi(\alpha_0 - 1)(\alpha_1 - 1)}{c_t^2}\Big[\psi'(\nu_t \phi) + \psi'\big([1-\nu_t]\phi\big)\Big].$$

We have, for $y_t \in (0, 1)$,

$$\begin{aligned}\frac{\partial}{\partial \mu_t}\left[\frac{y_t^* - \mu_t^*}{c_t^2}\right] &= -\frac{2(\alpha_0 - \alpha_1)}{c_t^4}(y_t^* - \mu_t^*)\\ &\quad - \frac{\phi(\alpha_0 - 1)(\alpha_1 - 1)}{c_t^4}\Big[\psi'(\nu_t \phi) + \psi'\big([1-\nu_t]\phi\big)\Big],\end{aligned}$$

hence

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \mu_t^2} \;\middle|\; \mathscr{F}_{t-1}\right) = \frac{\alpha_0\mu_t + \alpha_1(1-\mu_t)}{\mu_t(\mu_t - 1)} - \frac{(\alpha_0-\alpha_1)^2}{c_t} - \frac{\left[\phi(\alpha_0-1)(\alpha_1-1)\right]^2}{c_t^3}\left[\psi'(\nu_t\phi) + \psi'\big([1-\nu_t]\phi\big)\right].$$

Second mixed derivatives related to α_0 and α_1 are obtained through direct differentiation of the log-likelihood. We have, for $i \in \{0, 1\}$ and $j \in \{4, \ldots, \kappa\}$

$$\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \gamma_j \partial \alpha_i} = \left[\frac{\partial}{\partial \gamma_j}\left(\frac{1}{c_t}\left[\frac{\partial c_t}{\partial \alpha_i}\right]\right) + \frac{\partial}{\partial \gamma_j}\left(\phi(y_t^* - \mu_t^*)\left[\frac{\partial \nu_t}{\partial \alpha_i}\right]\right)\right] I_{(0,1)}(y_t),$$

which, by Lemma (A.1), yields

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \gamma_j \partial \alpha_i} \;\middle|\; \mathscr{F}_{t-1}\right) = \left[\frac{\partial}{\partial \gamma_j}\left(\frac{1}{c_t}\left[\frac{\partial c_t}{\partial \alpha_i}\right]\right) - \phi \frac{\partial \nu_t}{\partial \alpha_i}\frac{\partial \mu_t^*}{\partial \mu_t}\frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \gamma_j}\right] c_t.$$

Writing $\dfrac{\partial c_t}{\partial \gamma_j} = \dfrac{\partial c_t}{\partial \mu_t}\dfrac{\partial \mu_t}{\partial \eta_t}\dfrac{\partial \eta_t}{\partial \gamma_j} = \dfrac{\alpha_0 - \alpha_1}{g'(\mu_t)}\dfrac{\partial \eta_t}{\partial \gamma_j}$, we have

$$\frac{\partial}{\partial \gamma_j}\left(\frac{1}{c_t}\left[\frac{\partial c_t}{\partial \alpha_i}\right]\right) = \frac{1}{c_t g'(\mu_t)}\left[(-1)^i + \big[(\mu_t - 1)I_0(i) - \mu_t I_1(i)\big]\frac{\alpha_0 - \alpha_1}{c_t}\right]\frac{\partial \eta_t}{\partial \gamma_j},$$

and thus $\mathbb{E}\left(\dfrac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \gamma_j \partial \alpha_i} \;\middle|\; \mathscr{F}_{t-1}\right) = \dfrac{s_t^{(i)}}{g'(\mu_t)}\dfrac{\partial \eta_t}{\partial \gamma_j}$, where

$$s_t^{(i)} := \frac{(-1)^i\phi^2(\alpha_0-1)^{i+1}(\alpha_1-1)^{2-i}(1-\mu_t)\mu_t}{c_t^3}\left[\psi'(\nu_t\phi) + \psi'\big([1-\nu_t]\phi\big)\right] + (-1)^i + +\big[(\mu_t-1)I_0(i) - \mu_t I_1(i)\big]\frac{\alpha_0-\alpha_1}{c_t}. \tag{A3}$$

For $j \in \{4, \ldots, \kappa\}$, it is easy to show that

$$\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \gamma_j \partial \phi} = -\frac{\partial}{\partial \gamma_j}\left[\nu_t\mu_t^* + \psi\big([1-\nu_t]\phi\big)\right] I_{(0,1)}(y_t)$$
$$= -\left(\frac{(1-\alpha_0)(1-\alpha_1)}{c_t^3 g'(\mu_t)}\left[c_t\mu_t^* + \phi(1-\alpha_1)\psi'(\nu_t\phi) - \phi\alpha_1\psi'\big([1-\nu_t]\phi\big)\right]\frac{\partial \eta_t}{\partial \gamma_j}\right) I_{(0,1)}(y_t). \tag{A4}$$

Observe that, except for the indicator function, all terms in (A4) are $\mathscr{F}_{t-1}$-measurable, so that

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \gamma_j \partial \phi} \;\middle|\; \mathscr{F}_{t-1}\right) = \frac{d_t}{g'(\mu_t)}\frac{\partial \eta_t}{\partial \gamma_j},$$

where

$$d_t := \frac{(1-\alpha_0)(1-\alpha_1)\phi}{c_t^2}\left[(1-\nu_t)\psi'\big([1-\nu_t]\phi\big) - \nu_t\psi'(\nu_t\phi)\right]. \tag{A5}$$

For $i \in \{0, 1\}$,

$$\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \alpha_i \partial \phi} = \left((y_t^* - \mu_t^*)\frac{\partial \nu_t}{\partial \alpha_i} + \left[\phi\psi'\big([1-\nu_t]\phi\big) - \nu_t\frac{\partial \mu_t^*}{\partial \nu_t}\right]\frac{\partial \nu_t}{\partial \alpha_i}\right) I_{(0,1)}(y_t).$$

The first term has conditional expectation 0 (Lemma A.1), so that

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \alpha_i \partial \phi} \;\middle|\; \mathscr{F}_{t-1}\right) = \frac{(-1)^i(\alpha_{1-i}-1)(\mu_t-1)\mu_t\phi}{c_t^2}\Big[(1-\nu_t)\psi'\big([1-\nu_t]\phi\big) - \nu_t\psi'(\nu_t\phi)\Big].$$

Since $\dfrac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \phi^2} = \Big[\psi'(\phi) - \nu_t^2\psi'(\nu_t\phi) - (1-\nu_t)^2\psi'\big([1-\nu_t]\phi\big)\Big] I_{(0,1)}(y_t)$, we have

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \phi^2} \;\middle|\; \mathscr{F}_{t-1}\right) = c_t\Big[\psi'(\phi) - \nu_t^2\psi'(\nu_t\phi) - (1-\nu_t)^2\psi'\big([1-\nu_t]\phi\big)\Big].$$

Finally, for $i, j \in \{0, 1\}$,

$$\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \alpha_j \partial \alpha_i} = \left[\phi(y_t^* - \mu_t^*)\frac{\partial^2 \nu_t}{\partial \alpha_j \partial \alpha_i} - \phi\frac{\partial \nu_t}{\partial \alpha_i}\frac{\partial \mu_t^*}{\partial \alpha_j} - \frac{1}{c_t^2}\frac{\partial c_t}{\partial \alpha_i}\frac{\partial c_t}{\partial \alpha_j}\right] I_{(0,1)}(y_t) - \frac{1}{\alpha_i^2} I_i(j) I_i(y_t).$$

Upon observing that $P(y_t = i) = \alpha_i(1 - i + (-1)^i\mu_t)$, it follows that

$$\begin{aligned}\mathbb{E}\left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma})}{\partial \alpha_j \partial \alpha_i} \;\middle|\; \mathscr{F}_{t-1}\right) &= \frac{I_i(j)(i-1+(-1)^i\mu_t)}{\alpha_i} - c_t\left[\phi\frac{\partial \nu_t}{\partial \alpha_i}\frac{\partial \mu_t^*}{\partial \alpha_j} + \frac{1}{c_t^2}\frac{\partial c_t}{\partial \alpha_i}\frac{\partial c_t}{\partial \alpha_j}\right] \\ &= \frac{I_i(j)(i-1+(-1)^i\mu_t)}{\alpha_i} \\ &\quad - c_t\phi^2\Big[\psi'(\nu_t\phi) + \psi'\big([1-\nu_t]\phi\big)\Big]\frac{\partial \nu_t}{\partial \alpha_i}\frac{\partial \nu_t}{\partial \alpha_j} - \frac{1}{c_t}\frac{\partial c_t}{\partial \alpha_i}\frac{\partial c_t}{\partial \alpha_j}.\end{aligned}$$

For $i, j \in \{0, 1\}$, let

$$\begin{aligned}A_{\{i,j\}} &:= \operatorname{diag}\left\{\mathbb{E}\left(\frac{\partial \ell_{m+1}(\mu_{m+1},\varphi)}{\partial \alpha_i \partial \alpha_j} \;\middle|\; \mathscr{F}_m\right), \ldots, \mathbb{E}\left(\frac{\partial \ell_n(\mu_n,\varphi)}{\partial \alpha_i \partial \alpha_j} \;\middle|\; \mathscr{F}_{n-1}\right)\right\}, \\ B_i &:= \operatorname{diag}\left\{\mathbb{E}\left(\frac{\partial^2 \ell_{m+1}(\boldsymbol{\gamma})}{\partial \alpha_i \partial \phi} \;\middle|\; \mathscr{F}_m\right), \ldots, \mathbb{E}\left(\frac{\partial^2 \ell_n(\boldsymbol{\gamma})}{\partial \alpha_i \partial \phi} \;\middle|\; \mathscr{F}_{n-1}\right)\right\}, \\ C &:= \operatorname{diag}\left\{\mathbb{E}\left(\frac{\partial^2 \ell_{m+1}(\boldsymbol{\gamma})}{\partial \phi^2} \;\middle|\; \mathscr{F}_m\right), \ldots, \mathbb{E}\left(\frac{\partial^2 \ell_n(\boldsymbol{\gamma})}{\partial \phi^2} \;\middle|\; \mathscr{F}_{n-1}\right)\right\}, \\ V &:= \operatorname{diag}\left\{\mathbb{E}\left(\frac{\partial^2 \ell_{m+1}(\boldsymbol{\gamma})}{\partial \mu_{m+1}^2} \;\middle|\; \mathscr{F}_m\right), \ldots, \mathbb{E}\left(\frac{\partial^2 \ell_n(\boldsymbol{\gamma})}{\partial \mu_n^2} \;\middle|\; \mathscr{F}_{n-1}\right)\right\},\end{aligned}$$

$\boldsymbol{s}_i := (s_{m+1}^{(i)}, \ldots, s_n^{(i)})^\top$ and $\boldsymbol{d} := (d_{m+1}, \ldots, d_n)^\top$, where $s_t^{(i)}$ and d_t are given in (A3) and (A5), respectively. Thus, the joint cumulative partial information matrix for $\boldsymbol{\gamma}$ based on a sample of size n is

$$K_n(\boldsymbol{\gamma}) := -\begin{pmatrix} K_{(\alpha_0,\alpha_0)} & K_{(\alpha_0,\alpha_1)} & K_{(\alpha_0,\phi)} & K_{(\alpha_0,\alpha)} & K_{(\alpha_0,\boldsymbol{\beta})} & K_{(\alpha_0,\boldsymbol{\varphi})} & K_{(\alpha_0,\boldsymbol{\theta})} \\ K_{(\alpha_1,\alpha_0)} & K_{(\alpha_1,\alpha_1)} & K_{(\alpha_1,\phi)} & K_{(\alpha_1,\alpha)} & K_{(\alpha_1,\boldsymbol{\beta})} & K_{(\alpha_1,\boldsymbol{\varphi})} & K_{(\alpha_1,\boldsymbol{\theta})} \\ K_{(\phi,\alpha_0)} & K_{(\phi,\alpha_1)} & K_{(\phi,\phi)} & K_{(\phi,\alpha)} & K_{(\phi,\boldsymbol{\beta})} & K_{(\phi,\boldsymbol{\varphi})} & K_{(\phi,\boldsymbol{\theta})} \\ K_{(\alpha,\alpha_0)} & K_{(\alpha,\alpha_1)} & K_{(\alpha,\phi)} & K_{(\alpha,\alpha)} & K_{(\alpha,\boldsymbol{\beta})} & K_{(\alpha,\boldsymbol{\varphi})} & K_{(\alpha,\boldsymbol{\theta})} \\ K_{(\boldsymbol{\beta},\alpha_0)} & K_{(\boldsymbol{\beta},\alpha_1)} & K_{(\boldsymbol{\beta},\phi)} & K_{(\boldsymbol{\beta},\alpha)} & K_{(\boldsymbol{\beta},\boldsymbol{\beta})} & K_{(\boldsymbol{\beta},\boldsymbol{\varphi})} & K_{(\boldsymbol{\beta},\boldsymbol{\theta})} \\ K_{(\boldsymbol{\varphi},\alpha_0)} & K_{(\boldsymbol{\varphi},\alpha_1)} & K_{(\boldsymbol{\varphi},\phi)} & K_{(\boldsymbol{\varphi},\alpha)} & K_{(\boldsymbol{\varphi},\boldsymbol{\beta})} & K_{(\boldsymbol{\varphi},\boldsymbol{\varphi})} & K_{(\boldsymbol{\varphi},\boldsymbol{\theta})} \\ K_{(\boldsymbol{\theta},\alpha_0)} & K_{(\boldsymbol{\theta},\alpha_1)} & K_{(\boldsymbol{\theta},\phi)} & K_{(\boldsymbol{\theta},\alpha)} & K_{(\boldsymbol{\theta},\boldsymbol{\beta})} & K_{(\boldsymbol{\theta},\boldsymbol{\varphi})} & K_{(\boldsymbol{\theta},\boldsymbol{\theta})} \end{pmatrix},$$

where, for $i, j \in \{0, 1\}$, $K_{(\alpha_i,\alpha_j)} = \operatorname{tr}(A_{\{i,j\}})$, $K_{(\alpha_i,\phi)} = K_{(\phi,\alpha_i)}^\top = \operatorname{tr}(B_i)$, $K_{(\alpha_i,\alpha)} = K_{(\alpha,\alpha_i)} = \boldsymbol{s}_i^\top T\boldsymbol{a}$, $K_{(\alpha_i,\boldsymbol{\beta})} = K_{(\boldsymbol{\beta},\alpha_j)}^\top = \boldsymbol{s}_i^\top TR$, $K_{(\alpha_i,\boldsymbol{\varphi})} = K_{(\boldsymbol{\varphi},\alpha_i)}^\top = \boldsymbol{s}_i^\top TP$, $K_{(\alpha_i,\boldsymbol{\theta})} =$

$K^{\top}_{(\boldsymbol{\theta},\alpha_i)} = \boldsymbol{s}_i^{\top} T Q$, $K_{(\phi,\phi)} = \mathrm{tr}(C)$, $K_{(\phi,\alpha)} = K_{(\alpha,\phi)} = \boldsymbol{d}^{\top} T \boldsymbol{a}$, $K_{(\phi,\boldsymbol{\beta})} = K^{\top}_{(\boldsymbol{\beta},\phi)} = \boldsymbol{d}^{\top} T R$, $K_{(\phi,\boldsymbol{\varphi})} = K^{\top}_{(\boldsymbol{\varphi},\phi)} = \boldsymbol{d}^{\top} T P$, $K_{(\phi,\boldsymbol{\theta})} = K^{\top}_{(\boldsymbol{\theta},\phi)} = \boldsymbol{d}^{\top} T Q$, $K_{(\alpha,\alpha)} = \boldsymbol{a}^{\top} T^2 V \boldsymbol{a}$, $K_{(\alpha,\beta)} = K^{\top}_{(\beta,\alpha)} = \boldsymbol{a}^{\top} T^2 V R$, $K_{(\alpha,\varphi)} = K^{\top}_{(\varphi,\alpha)} = \boldsymbol{a}^{\top} T^2 V P$, $K_{(\alpha,\theta)} = K^{\top}_{(\theta,\alpha)} = \boldsymbol{a}^{\top} T^2 V Q$, $K_{(\boldsymbol{\beta},\boldsymbol{\beta})} = R^{\top} T^2 V R$, $K_{(\boldsymbol{\beta},\boldsymbol{\varphi})} = K^{\top}_{(\boldsymbol{\varphi},\boldsymbol{\beta})} = R^{\top} T^2 V P$, $K_{(\boldsymbol{\beta},\boldsymbol{\theta})} = K^{\top}_{(\boldsymbol{\theta},\boldsymbol{\beta})} = R^{\top} T^2 V Q$, $K_{(\boldsymbol{\varphi},\boldsymbol{\varphi})} = P^{\top} T^2 V P$, $K_{(\boldsymbol{\varphi},\boldsymbol{\theta})} = K^{\top}_{(\boldsymbol{\theta},\boldsymbol{\varphi})} = P^{\top} T^2 V Q$, $K_{(\boldsymbol{\theta},\boldsymbol{\theta})} = Q^{\top} T^2 V Q$.

Lemma A.1 *With the notation in A.1,*

$$\mathbb{E}\big((y_t^* - \mu_t^*) I_{(0,1)}(y_t) \mid \mathscr{F}_{t-1}\big) = 0 \quad \text{and} \quad \mathbb{E}\bigg(\frac{\partial \ell_t(\mu_t, \nu)}{\partial \mu_t} \;\Big|\; \mathscr{F}_{t-1}\bigg) = 0.$$

Proof Observe that

$$\mathbb{E}\big((y_t^* - \mu_t^*) I_{(0,1)}(y_t) \mid \mathscr{F}_{t-1}\big) = c_t \int_0^1 \left(\frac{\log(x)}{\log(1-x)} - \mu_t^*\right) \mathrm{b}(x; \nu_t, \phi) dx = 0,$$

by standard results on the beta distribution. Hence

$$\begin{aligned}
\mathbb{E}\bigg(\frac{\partial \ell_t(\mu_t, \nu)}{\partial \mu_t} \;\Big|\; \mathscr{F}_{t-1}\bigg) &= \left(\frac{1}{\mu_t}\right) P(y_t = 1 \mid \mathscr{F}_{t-1}) - \left(\frac{1}{1-\mu_t}\right) P(y_t = 0 \mid \mathscr{F}_{t-1}) \\
&\quad + \left(\frac{\alpha_0 - \alpha_1}{c_t}\right) P\big(y_t \in (0,1) \mid \mathscr{F}_{t-1}\big) \\
&= \left(\frac{\alpha_0 - \alpha_1}{c_t}\right) c_t - \left(\frac{1}{1-\mu_t}\right) \alpha_0 (1-\mu_t) + \left(\frac{1}{\mu_t}\right) \alpha_1 \mu_t = 0,
\end{aligned}$$

as asserted. □

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19(6):716–723

Bayer FM, Bayer DM, Pumi G (2017) Kumaraswamy autoregressive moving average models for doubled bounded environmental data. J Hydrol 555:385–396

Bayer FM, Cintra RJ, Cribari-Neto F (2018) Beta seasonal autoregressive moving average models. J Stat Comput Simul 88(15):2961–2981

Bayer FM, Bayer DM, Marinoni A, Gamba P (2020) A novel Rayleigh dynamical model for remote sensing data interpretation. IEEE Trans Geosci Remote Sens 58(7):4989–4999

Bayes CL, Valdivieso L (2016) A beta inflated mean regression model for fractional response variables. J Appl Stat 43(10):1814–1830

Benjamin MA, Rigby RA, Stasinopoulos DM (2003) Generalized autoregressive moving average models. J Am Stat Assoc 98(461):214–223

Bloomfield P (2013) Fourier analysis of time series: an introduction, 2nd edn. Wiley-Interscience, New Jersey, p 288

Box G, Jenkins GM, Reinsel G, Ljung GM (2015) Time series analysis: forecasting and control, 5th edn. Wiley, Hardcover

Brazilian National Institute of Meteorology (INMET) (2018) Meteorological database for research and teaching. http://www.inmet.gov.br/projetos/rede/pesquisa. Accessed Oct 2018

Chuang M-D, Yu G-H (2007) Order series method for forecasting non-Gaussian time series. J Forecast 26(4):239–250

Cox DR (1975) Partial likelihood. Biometrika 62(2):69–76

Cox DR (1981) Statistical analysis of time series: some recent developments. Scand J Stat 8:93–115

da-Silva CQ, Migon HS, Correia LT (2011) Dynamic bayesian beta models. Comput Stat Data Anal 55(6):2074–2089

Fahrmeir L (1987) Asymptotic testing theory for generalized linear models. Statistics 18(1):65–76

Fahrmeir L, Kaufmann H (1985) Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. Ann Stat 13(1):342–368

Ferrari SLP, Cribari-Neto F (2004) Beta regression for modelling rates and proportions. J Appl Stat 31(7):799–815
Fokianos K, Kedem B (1998) Prediction and classification of non-stationary categorical time series. J Multivar Anal 67(2):277–296
Fokianos K, Kedem B (2004) Partial likelihood inference for time series following generalized linear models. J Time Ser Anal 25(2):173–197
Grassly NC, Fraser C (2006) Seasonal infectious disease epidemiology. Proc R Soc Lond B: Biol Sci 273(1600):2541–2550
Guolo A, Varin C (2014) Beta regression for time series analysis of bounded data, with application to Canada Google Flu Trends. Ann Appl Stat 8(1):74–88
Hannan EJ, Quinn BG (1979) The determination of the order of an autoregression. J R Stat Soc Ser B 41(2):190–195
Kedem B, Fokianos K (2002) Regression models for time series analysis. Wiley, New Jersey
Lambert D (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34(1):1–14
Li WK (1991) Testing model adequacy for some Markov regression models for time series. Biometrika 78(1):83–89
Li WK (1994) Time series models based on generalized linear models: some further results. Biometrics 50(2):506–511
Ljung GM, Box GEP (1978) On a measure of lack of fit in time series models. Biometrika 65(2):297–303
McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman and Hall, Boca Raton
Melchior C, Zanini RR, Guerra RR, Rockenbach DA (2021) Forecasting Brazilian mortality rates due to occupational accidents using autoregressive moving average approaches. Int J Forecast 37(2):825–837
Monti AC (1994) A proposal for a residual autocorrelation test in linear models. Biometrika 81(4):776–780
Neyman J, Pearson ES (1928) On the use and interpretation of certain test criteria for purposes of statistical inference. Biometrika 20A(1/2):175–240
Nocedal J, Wright SJ (1999) Numerical optimization. Springer, New York
Operador Nacional do Sistema Elétrico (2022) Dados Hidrológicos. http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/dados_hidrologicos_volumes.aspx. Accessed Dec 2022
Ospina R, Ferrari SLP (2012) A general class of zero-or-one inflated beta regression models. Comput Stat Data Anal 56(6):1609–1623
Palm BG, Bayer FM, Cintra RJ (2021) Signal detection and inference based on the beta binomial autoregressive moving average model. Digital Signal Process 109:102911
Pawitan Y (2001) In all likelihood: statistical modelling and inference using likelihood. Oxford Science publications, New York
Press W, Teukolsky S, Vetterling W, Flannery B (1992) Numerical recipes in C: the art of scientific computing, 2nd edn. Cambridge University Press, New York
Pumi G, Valk M, Bisognin C, Bayer FM, Prass TS (2019) Beta autoregressive fractionally integrated moving average models. J Stat Plann Inference 200:196–202
Pumi G, Prass TS, Souza RR (2021) A dynamic model for double-bounded time series with chaotic-driven conditional averages. Scand J Stat 48(1):68–86
R Core Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2020). R Foundation for Statistical Computing
Rao CR (1948) Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. Math Proc Cambridge Philos Soc 44(1):50–57
Rocha AV, Cribari-Neto F (2009) Beta autoregressive moving average models. TEST 18(3):529–545
Rocha AV, Cribari-Neto F (2017) Erratum to: beta autoregressive moving average models. TEST 26(2):451–459
Sagrillo M, Guerra RR, Bayer FM (2021) Modified Kumaraswamy distributions for double bounded hydro-environmental data. J Hydrol 603:127021
Scher VT, Cribari-Neto F, Pumi G, Bayer FM (2020) Goodness-of-fit tests for βARMA hydrological time series modeling. Environmetrics 31(3):2607
Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464
Tamerius JD, Shaman J, Alonso WJ, Bloom-Feshbach K, Uejio CK, Comrie A, Viboud C (2013) Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. PLoS Pathog 9(3):1003194
Terrell GR (2002) The gradient statistic. Comput Sci Stat 34:206–215
Tiku ML, Wong W-K, Vaughan DC, Bian G (2000) Time series models in non-normal situations: symmetric innovations. J Time Ser Anal 21(5):571–596

Wald A (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans Am Math Soc 54:426–482

Zeger SL, Qaqish B (1988) Markov regression models for time series: a quasi-likelihood approach. Biometrics 44(4):1019–1031

Zheng T, Xiao H, Chen R (2015) Generalized ARMA models with martingale difference errors. J Econometr 189(2):492–506