ARTICLE IN PRESS

International Journal of Forecasting xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Generalized βARMA model for double bounded time series forecasting

Vinícius T. Scher [a], Francisco Cribari-Neto [a,*], Fábio M. Bayer [b]

[a] Departamento de Estatística, Centro de Ciências Exatas e da Natureza, Universidade Federal de Pernambuco, Recife/PE, Brazil
[b] Departamento de Estatística and LACESM, Universidade Federal de Santa Maria, Santa Maria/RS, Brazil

ARTICLE INFO

Keywords:
Beta distribution
βARMA model
Forecasting
Stored hydroelectric energy
Variable precision
Useful volume

ABSTRACT

The βARMA model is tailored for use with time series that assume values in (0, 1). We generalize the model in which both the conditional mean and conditional precision evolve over time. The standard βARMA model, in which precision is constant, is a particular case of our model. The more general model formulation includes a parsimonious submodel for the precision parameter. We present the model conditional log-likelihood function, the conditional score function, and the conditional Fisher information matrix. We use the proposed model to forecast future levels of stored hydroelectric energy and the useful volume of a water reservoir in the South of Brazil. Our results show that more accurate forecasts are typically obtained by allowing the precision parameter to evolve over time.

1. Introduction

There is often interest in modeling and predicting future values of time series that assume values in the standard unit interval, (0, 1). Forecasts must be restricted to such an interval. Some authors have focused on the case where the interest lies in forecasting compositional time series, that is, a set of time series that assume values in (0, 1) and that, at each instant of time, sum to one. For example, AL-Dhurafi, Masseran, and Zamzuri (2018) propose to apply the additive logistic transformation to the data and then model the transformed data using a vector autoregressive model and Snyder, Ord, Koehler, McLaren, and Beaumont (2017) use state space models on log-ratio transformed shares. Zheng and Chen (2017) introduce the class of Dirichlet autoregressive moving average (DARMA) models for use with compositional data. For more details on compositional time series modeling, see Brunsdon, Smith, and T (1998) and Mills (2010).

There are situations, however, where the interest lies in modeling an individual time series that assumes values in (0, 1). For example, modeling the time-series behavior of income concentration indices, rates, and proportions is typically desirable. For instance, Lu and Meyer (2020) use a dynamic formulation of the beta regression model introduced by Ferrari and Cribari-Neto (2004) to investigate the usefulness of the endemic-epidemic beta model as a forecasting tool. Using this and competing models, they predict the United States' short-term and seasonal flu activity. Melchior, Zanini, Guerra, and Rockenbach (2021) use different approaches to model and forecast mortality rates due to occupational accidents in southern Brazil. The authors use a conventional time series model and two models that consider the fractional nature of the data; the latter two are based on the beta and Kumaraswamy laws. In their analysis, the former displays superior performance. It is known in the literature as the beta autoregressive moving average (βARMA) model.

The βARMA model was introduced by Rocha and Cribari-Neto (2009) and Rocha and Cribari-Neto (2017) as a dynamic extension of the class of beta regression models proposed by Ferrari and Cribari-Neto (2004); see Cribari-Neto and Zeileis (2010) and Douma

* Corresponding author.
E-mail address: cribari@de.ufpe.br (F. Cribari-Neto).

https://doi.org/10.1016/j.ijforecast.2023.05.005
0169-2070/

Please cite this article as: V.T. Scher, F. Cribari-Neto and F.M. Bayer, Generalized βARMA model for double bounded time series forecasting. International Journal of Forecasting (2023), https://doi.org/10.1016/j.ijforecast.2023.05.005.

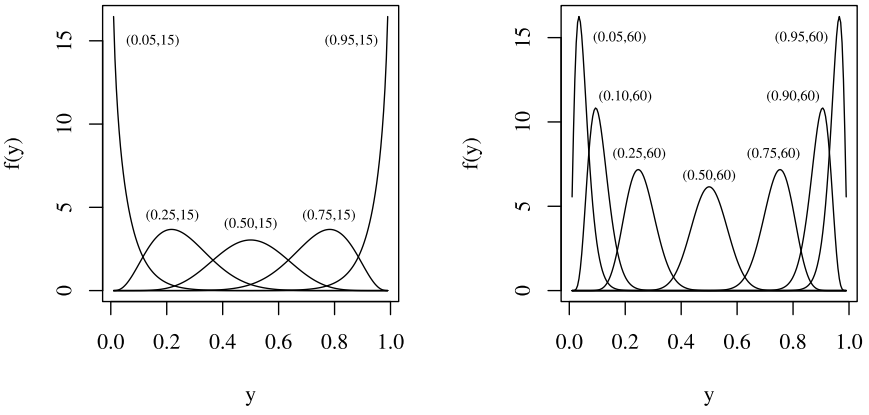

Fig. 1. Beta densities for different values of (μ, ϕ).

and Weedon (2019). Both models are tailored for use with double-bounded random variables, i.e., random variables that assume values in (a, b), where a and b are known and finite. Focus is typically placed on random variables that assume values in the standard unit interval, $(0, 1)$, since when $a \neq 0$ and/or $b \neq 1$, one can subtract a from the variable of interest and then divide that by $b - a$ to obtain a random variable with support in the standard unit interval. The βARMA model is useful for modeling the behavior of double bounded random variables that evolve over time and predicting their future levels. It incorporates autoregressive and moving average dynamics and allows for the inclusion of fixed covariates. A novel feature of the model is that it will never yield improper forecasts, i.e., forecasts that lie outside $(0, 1)$. We note that a model for count time series with a dynamic structure similar to that of the βARMA process was recently introduced by Sales, Alencar, and Ho (2022).

In the βARMA model, it is assumed that the variable of interest follows the beta law parameterized in terms of its mean, μ, and a precision parameter, ϕ; see the expression for the beta density in Ferrari and Cribari-Neto (2004), Equation (4). The beta distribution is quite flexible since its density can assume many shapes depending on the parameter values. The uniform distribution is a special case obtained with $\mu = 0.5$ and $\phi = 2$. The two panels in Fig. 1 contain plots of beta densities for different (μ, ϕ) values. The density is symmetric when $\mu = 0.5$ and asymmetric otherwise. It can be J-shaped and inverted J-shaped. It can also be skewed to the right or the left. Additionally, since the beta variance is $\mu(1 - \mu)/(1 + \phi)$ and the mean parameter evolves over time, the βARMA model is naturally heteroskedastic. The density shape and the distribution variance change over time, thus making the model quite flexible.

βARMA diagnostic analysis was developed by Scher, Cribari-Neto, Pumi, and Bayer (2020). The authors introduce a portmanteau test statistic, Q_4, that works well to test whether a fitted βARMA model is correctly specified. The test statistic uses partial residual autocorrelations, and rejection of the null hypothesis is taken as evidence of model misspecification. The authors prove that, under the null hypothesis of correct model specification, the limiting distribution of their test statistic is χ^2_{m-p-q}, where m is the number of lags (number of partial autocorrelations), and p and q are the autoregressive and moving average orders, respectively.

Model selection was investigated by Cribari-Neto, Scher, and Bayer (2023). They consider different βARMA model selection strategies, including using information criteria and forecasting accuracy measures such as that proposed by Poler and Mula (2011). Their results indicate that the best strategy employs an information criterion based on data (bootstrap) resampling. The authors also present an empirical analysis in which the main goal is to model the proportion of stored hydroelectric energy in South Brazil. They show that more accurate model selection typically translates into more accurate forecasts. Bayesian model selection for beta autoregressive processes was developed by Casarin, Valle, and Leisen (2012). Aknouche and Dimitrakopoulos (2023) develop a multiplicative autoregressive conditional proportion model for time series that assume values in $(0, 1)$ in the spirit of generalized autoregressive conditional heteroskedastic and autoregressive conditional duration models. A dynamic beta model with an iteration of a map that can present chaotic behavior was proposed by Pumi, Prass, and Souza (2021). Guolo and Varin (2014) introduce a dynamic beta regression model in which the serial dependence is modeled by a Gaussian copula, with a correlation matrix corresponding to a stationary autoregressive and moving average process. A dynamic model for double bounded random variables based on the Kumaraswamy law was introduced by Bayer, Bayer, and Pumi (2017). Like the βARMA model, it incorporates autoregressive and moving average dynamics.

This paper aims to introduce a generalized, more flexible formulation of the βARMA model. In the standard formulation of the model, the beta law mean (conditional on previous information) evolves over time, and the precision parameter is assumed to be globally fixed. By contrast, in the generalized formulation of the model, both the conditional mean and the conditional precision are allowed to evolve dynamically. The generalized model contains two submodels, one for the mean and another for the precision. We use a parsimonious formulation for the dynamic structure that drives the precision over time. Notably, the proposed model has an additional layer of

flexibility since it allows the beta density shape to evolve more freely over time.

We use the generalized βARMA model to obtain out-of-sample forecasts of the time series modeled by Cribari-Neto et al. (2023) and an updated series that includes more recent data. Out-of-sample forecasts are also produced from models fitted to subsets of the data that range from 45 to 255 data points in steps of five observations, totaling 43 sample sizes. We compare such forecasts to those yielded by standard, fixed precision βARMA models. The results show that the generalized model's forecasts are typically more accurate than those from the standard βARMA model. Additionally, we use different criteria to evaluate short-term probabilistic forecasts obtained from the generalized and standard βARMA models. All criteria favor the former. We also model the useful volume (expressed as a proportion) of an accumulation reservoir in the South of Brazil. This time series displays more variability than those of stored energy. Again, the generalized model yields more accurate forecasts. Indeed, its one-step-ahead forecast is nearly 22% more accurate than that obtained under fixed precision.

The remainder of the paper is structured as follows. In Section 2, we introduce the generalized βARMA model, and in Section 3 we develop conditional maximum likelihood inference for the model's parameters. In particular, we present the conditional log-likelihood function, the conditional score function, and the conditional Fisher information matrix. Section 4 contains two empirical analyses in which we model and forecast the future levels of stored hydroelectric energy and the useful volume of a water reservoir using generalized and standard βARMA models. The results show that out-of-sample forecasts obtained by allowing for varying precision are typically more accurate than those obtained under the assumption of fixed precision. Also, the generalized model is favored by information criteria relative to the standard model, and testing inferences indicate the existence of varying precision at the usual significance levels. Concluding remarks are offered in Section 5.

2. The generalized βARMA model

The βARMA model proposed by Rocha and Cribari-Neto (2009) and Rocha and Cribari-Neto (2017) is a dynamic time series model for use with double bounded random variables, notably with variables that assume values in (0, 1), such as rates and proportions. In the following, we will introduce a generalization of the model. The new model comprises two sub-models, one for the conditional mean of the process and one for the conditional precision.

Let $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ be an n-vector of time series random variables such that each y_t, $t = 1, \ldots, n$, conditionally on the set of previous information $\mathcal{F}_{t-1}$, follows the beta law with mean μ_t and precision ϕ_t. Here, $\mathcal{F}_{t-1}$ is the smallest σ-algebra such that the variables $y_1, \ldots, y_{t-1}$ are measurable. The conditional density of y_t, given $\mathcal{F}_{t-1}$, is

$$f(y_t; \mu_t, \phi_t|\mathcal{F}_{t-1}) = \frac{\Gamma(\phi_t)}{\Gamma(\mu_t\phi_t)\Gamma((1-\mu_t)\phi_t)} y_t^{\mu_t\phi_t-1}(1-y_t)^{(1-\mu_t)\phi_t-1},$$

$0 < y_t < 1$, $0 < \mu_t < 1$ and $\phi_t > 0$, where $\Gamma(\cdot)$ is the gamma function. The conditional mean and the conditional variance of y_t are, respectively, $\mathbb{E}(y_t|\mathcal{F}_{t-1}) = \mu_t$ and $\text{var}(y_t|\mathcal{F}_{t-1}) = \mu_t(1-\mu_t)/(1+\phi_t)$.

Let $g_1 : (0, 1) \mapsto \mathbb{R}$ be a strictly increasing and twice differentiable link function, such as the logit, probit, cauchit, log–log, and complementary log–log functions. In the βARMA model, the dynamic structure for μ_t is

$$g_1(\mu_t) = \alpha_1 + \boldsymbol{x}_t^\top\boldsymbol{\beta} + \sum_{i=1}^{p}\varphi_i\left[g_1(y_{t-i}) - \boldsymbol{x}_{t-i}^\top\boldsymbol{\beta}\right] + \sum_{j=1}^{q}\theta_j r_{t-j}, \tag{1}$$

where $\alpha_1 \in \mathbb{R}$ and $p, q \in \mathbb{N}$ are the autoregressive and moving average orders. Here, $\boldsymbol{x}_t \in \mathbb{R}^c$ is a set of non-random covariates at time t and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_c)^\top \in \mathbb{R}^c$ is a vector of parameters. Also, r_t is an error term that can be specified in the original scale, $y_t - \mu_t$, or in the predictor scale, $g_1(y_t) - g_1(\mu_t)$; in what follows, we will consider the latter. The dynamic formulation in Eq. (1) is similar to that used in the class of generalized autoregressive moving average (GARMA) models introduced by Benjamim, Rigby, and Stasinopoulos (2003).

In the standard formulation of the βARMA model, the precision parameter is assumed to be constant for all observations, i.e., $\phi_t = \phi\ \forall t$. We generalize the model by allowing the precision parameter to evolve over time. To that end, we postulate a simple and parsimonious dynamic structure for the precision parameter. The main idea we use is that the precision at time t is impacted by a variable that approximates the variance of the process at time $t-1$. Recall that such a variance increases with $\mu_{t-1}(1-\mu_{t-1})$. We use $y_{t-1}(1-y_{t-1})$, which is known at time t conditional of $\mathcal{F}_{t-1}$, as a proxy for $\mu_{t-1}(1-\mu_{t-1})$. Let $g_2 : \mathbb{R}_+ \mapsto \mathbb{R}$ be a strictly increasing and twice-differentiable link function, such as the logarithm function. Following Cribari-Neto and Zeileis (2010), we also consider the square root and identity functions. The dynamic submodel for the precision parameter is specified as

$$g_2(\phi_t) = \alpha_2 + \delta z_{t-1}, \tag{2}$$

where $\alpha_2 \in \mathbb{R}$, $\delta \in \mathbb{R}$ and $z_t = y_t(1-y_t)$. The tth precision is $\phi_t = \exp(\alpha_2 + \delta z_{t-1})$, $\phi_t = (\alpha_2 + \delta z_{t-1})^2$ and $\phi_t = \alpha_2 + \delta z_{t-1}$ for the log, square root and identity link functions, respectively. The standard βARMA model is a particular case of our model when g_2 is the identity link and $\delta = 0$.

The rationale for the proposed extension of the βARMA model is as follows. As noted above, for a given precision value, the variance of y_t increases with $\mu_t(1-\mu_t)$, being maximal at $\mu_t = 0.5$ and approaching zero as μ_t approaches zero or one. The precision submodel of the generalized βARMA model includes $y_{t-1}(1-y_{t-1})$ as an explanatory variable. When its value increases, there is evidence of a variability increase in the previous period, and the model responds by decreasing the value of the current precision. It is expected that $\alpha_2 > 0$ and $\delta < 0$. The intercept (α_2) determines the maximal

precision level, which is given by $g_2^{-1}(\alpha_2)$, and δz_{t-1} determines how the individual precisions fluctuate below it. Notice that $\delta < 0$ implies $\phi_t < g_2^{-1}(\alpha_2)\ \forall t$. Consider, e.g., (i) $y_{t-1} = 0.5$ and (ii) $y_{t-1} = 0.05$ or 0.95; then, $z_{t-1} = 0.25$ and 0.0475, respectively. As long as $\delta < 0$, the current precision decreases whenever the previous value of the process moves towards the middle of the standard unit interval and increases otherwise. The value of δ determines the magnitude of the changes in the precision levels between consecutive time periods. As an example, suppose $\alpha_2 = 20$, $\delta = -39$ and the link function is identity. (These values are close to the estimates obtained in the next section for one of the time series we model.) When $y_{t-1} = 0.6$, we get $\phi_t = 10.64$, whereas when $y_{t-1} = 0.95$ we obtain $\phi_t = 18.1475$. The current precision is higher in the latter case, i.e., when the previous value of the process is close to one.

Finally, in the standard βARMA model, changes to the beta density shape over time are only driven by changes in μ_t. In contrast, in the more general formulation of the model, they follow from changes in μ_t and ϕ_t. Thus, the generalized βARMA framework has greater flexibility since the beta density shape may evolve more freely over time.

3. Parameter estimation

Parameter estimation for the generalized βARMA model given in (1) and (2) is performed by conditional maximum likelihood. The model can be expressed more concisely as $g_1(\mu_t) = \eta_{1t}$ and $g_2(\phi_t) = \eta_{2t}$, where η_{1t} and η_{2t} are, respectively, the mean and precision linear predictors. Let $\boldsymbol{\nu} = (\alpha_1, \boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top, \boldsymbol{\beta}^\top, \alpha_2, \delta)^\top$ be the k-dimensional parameter vector, where $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)^\top$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_c)^\top$, with $k = p + q + c + 3 < n$. The total conditional log-likelihood function for the parameter vector $\boldsymbol{\nu}$, given the first $a = \max\{p, q\}$ observations, is

$$\ell \equiv \ell(\mu_t, \phi_t) = \sum_{t=a+1}^{n} \ell_t,$$

where $\ell_t \equiv \ell_t(\mu_t, \phi_t) = \log f(y_t; \mu_t, \phi_t | \mathcal{F}_{t-1})$, i.e.,

$$\begin{aligned}\ell_t(\mu_t, \phi_t) = {} & \log\left(\Gamma(\phi_t)\right) - \log\left(\Gamma(\mu_t\phi_t)\right) \\ & - \log\left(\Gamma((1-\mu_t)\phi_t)\right) \\ & + (\mu_t\phi_t - 1)\log(y_t) \\ & + [(1-\mu_t)\phi_t - 1]\log(1-y_t).\end{aligned}$$

The conditional maximum likelihood estimators of the model parameters cannot be expressed in closed form. Point estimates can be obtained by numerically maximizing ℓ using a Newton or quasi-Newton optimization algorithm. In what follows, we will use the BFGS quasi-Newton algorithm with analytical first derivatives; for details, see Nocedal and Wright (2006). When the model has moving average components, it is necessary to account for the recursive structure of the derivatives of ℓ; see Rocha and Cribari-Neto (2017).

We will present closed-form expressions for the conditional score function and the conditional expected information matrix in the following. The latter is useful, for instance, for obtaining standard errors for the maximum likelihood point estimates, interval estimation, and hypothesis testing inferences.

Let $\boldsymbol{U} = (U_{\alpha_1}, \boldsymbol{U}_{\boldsymbol{\varphi}}^\top, \boldsymbol{U}_{\boldsymbol{\theta}}^\top, \boldsymbol{U}_{\boldsymbol{\beta}}^\top, U_{\alpha_2}, U_\delta)^\top$ be the conditional score vector. To obtain a closed-form expression, we need to obtain the derivatives of ℓ with respect to each parameter. Using the results presented in Appendix A, the elements of the score vector $\boldsymbol{U}$ can be expressed in matrix form as

$$\begin{aligned}&U_{\alpha_1} = \boldsymbol{s}^\top \Phi T_1(\boldsymbol{y}^* - \boldsymbol{\mu}^*), \quad \boldsymbol{U}_{\boldsymbol{\beta}} = M^\top \Phi T_1(\boldsymbol{y}^* - \boldsymbol{\mu}^*),\\ &\boldsymbol{U}_{\boldsymbol{\varphi}} = P^\top \Phi T_1(\boldsymbol{y}^* - \boldsymbol{\mu}^*), \quad \boldsymbol{U}_{\boldsymbol{\theta}} = R^\top \Phi T_1(\boldsymbol{y}^* - \boldsymbol{\mu}^*),\\ &U_{\alpha_2} = \mathbf{1}_n^\top H T_2 \mathbf{1}_n, \quad \text{and} \quad U_\delta = \boldsymbol{\omega}^\top H T_2 \mathbf{1}_n.\end{aligned}$$

The conditional maximum likelihood estimator of $\boldsymbol{\nu}$ is obtained as the solution to the system of equations for $\boldsymbol{U} = \mathbf{0}_k$, where $\mathbf{0}_k$ is the $k \times 1$ vector of zeros. As noted earlier, it cannot be expressed in closed form, and estimates can be obtained by numerically maximizing the conditional log-likelihood function. Starting values for the parameters can be selected as follows: (i) all moving average parameters are set equal to zero, (ii) the values for the autoregressive parameters and α_1 are selected by regressing $g_1(y_t)$ on a constant and $g_1(y_{t-1}), \ldots, g_1(y_{t-p})$ using ordinary least squares, (iii) δ is set equal to zero, and (iv) α_2 is set equal to $g_2^{-1}(\phi^0)$, where ϕ^0 is computed as described on page 805 of Ferrari and Cribari-Neto (2004).

We need to compute the expected values of the second order log-likelihood derivatives to obtain the conditional Fisher information matrix. Using the results presented in Appendix B, the joint conditional Fisher information matrix can be expressed as

$$K \equiv K(\boldsymbol{\nu}) = \begin{bmatrix} K_{\alpha_1\alpha_1} & K_{\alpha_1\boldsymbol{\beta}} & K_{\alpha_1\boldsymbol{\varphi}} & K_{\alpha_1\boldsymbol{\theta}} & K_{\alpha_1\alpha_2} & K_{\alpha_1\delta} \\ K_{\boldsymbol{\beta}\alpha_1} & K_{\boldsymbol{\beta}\boldsymbol{\beta}} & K_{\boldsymbol{\beta}\boldsymbol{\varphi}} & K_{\boldsymbol{\beta}\boldsymbol{\theta}} & K_{\boldsymbol{\beta}\alpha_2} & K_{\boldsymbol{\beta}\delta} \\ K_{\boldsymbol{\varphi}\alpha_1} & K_{\boldsymbol{\varphi}\boldsymbol{\beta}} & K_{\boldsymbol{\varphi}\boldsymbol{\varphi}} & K_{\boldsymbol{\varphi}\boldsymbol{\theta}} & K_{\boldsymbol{\varphi}\alpha_2} & K_{\boldsymbol{\varphi}\delta} \\ K_{\boldsymbol{\theta}\alpha_1} & K_{\boldsymbol{\theta}\boldsymbol{\beta}} & K_{\boldsymbol{\theta}\boldsymbol{\varphi}} & K_{\boldsymbol{\theta}\boldsymbol{\theta}} & K_{\boldsymbol{\theta}\alpha_2} & K_{\boldsymbol{\theta}\delta} \\ K_{\alpha_2\alpha_1} & K_{\alpha_2\boldsymbol{\beta}} & K_{\alpha_2\boldsymbol{\varphi}} & K_{\alpha_2\boldsymbol{\theta}} & K_{\alpha_2\alpha_2} & K_{\alpha_2\delta} \\ K_{\delta\alpha_1} & K_{\delta\boldsymbol{\beta}} & K_{\delta\boldsymbol{\varphi}} & K_{\delta\boldsymbol{\theta}} & K_{\delta\alpha_2} & K_{\delta\delta} \end{bmatrix},$$

where $K_{\alpha_1\alpha_1} = \boldsymbol{s}^\top W_1 \boldsymbol{s}$, $K_{\alpha_1\boldsymbol{\beta}} = K_{\boldsymbol{\beta}\alpha_1}^\top = M^\top W_1 \boldsymbol{s}$, $K_{\alpha_1\boldsymbol{\varphi}} = K_{\boldsymbol{\varphi}\alpha_1}^\top = P^\top W_1 \boldsymbol{s}$, $K_{\alpha_1\boldsymbol{\theta}} = K_{\boldsymbol{\theta}\alpha_1}^\top = R^\top W_1 \boldsymbol{s}$, $K_{\alpha_1\alpha_2} = K_{\alpha_2\alpha_1}^\top = \boldsymbol{s}^\top W_3 \mathbf{1}_n$, $K_{\alpha_1\delta} = K_{\delta\alpha_1}^\top = \boldsymbol{s}^\top W_3 \boldsymbol{\omega}$, $K_{\boldsymbol{\beta}\boldsymbol{\beta}} = M^\top W_1 M$, $K_{\boldsymbol{\beta}\boldsymbol{\varphi}} = K_{\boldsymbol{\varphi}\boldsymbol{\beta}}^\top = M^\top W_1 P$, $K_{\boldsymbol{\beta}\boldsymbol{\theta}} = K_{\boldsymbol{\theta}\boldsymbol{\beta}}^\top = M^\top W_1 R$, $K_{\boldsymbol{\beta}\alpha_2} = K_{\alpha_2\boldsymbol{\beta}}^\top = M^\top W_3 \mathbf{1}_n$, $K_{\boldsymbol{\beta}\delta} = K_{\delta\boldsymbol{\beta}}^\top = M^\top W_3 \boldsymbol{\omega}$, $K_{\boldsymbol{\varphi}\boldsymbol{\varphi}} = P^\top W_1 P$, $K_{\boldsymbol{\varphi}\boldsymbol{\theta}} = K_{\boldsymbol{\theta}\boldsymbol{\varphi}}^\top = R^\top W_1 P$, $K_{\boldsymbol{\varphi}\alpha_2} = K_{\alpha_2\boldsymbol{\varphi}}^\top = P^\top W_3 \mathbf{1}_n$, $K_{\boldsymbol{\varphi}\delta} = K_{\delta\boldsymbol{\varphi}}^\top = P^\top W_3 \boldsymbol{\omega}$, $K_{\boldsymbol{\theta}\boldsymbol{\theta}} = R^\top W_1 R$, $K_{\boldsymbol{\theta}\alpha_2} = K_{\alpha_2\boldsymbol{\theta}}^\top = R^\top W_3 \mathbf{1}_n$, $K_{\boldsymbol{\theta}\delta} = K_{\delta\boldsymbol{\theta}}^\top = R^\top W_3 \boldsymbol{\omega}$, $K_{\alpha_2\alpha_2} = \mathbf{1}_n^\top W_2 \mathbf{1}_n$, $K_{\alpha_2\delta} = K_{\delta\alpha_2}^\top = \boldsymbol{\omega}^\top W_2 \mathbf{1}_n$, and $K_{\delta\delta} = \boldsymbol{\omega}^\top W_2 \boldsymbol{\omega}$.

The parameter vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ are not orthogonal (i.e., Fisher's information matrix is not block diagonal). When n is large, $\hat{\boldsymbol{\nu}}$, the conditional maximum likelihood estimator of $\boldsymbol{\nu}$, is approximately distributed as $\mathcal{N}_k(\boldsymbol{\nu}, K^{-1}(\boldsymbol{\nu}))$.

4. Out-of-sample forecasting evaluation

The generalized βARMA model adds an additional layer of flexibility to the standard formulation of the model since it allows the precision parameter to evolve

over time. To what extent such additional flexibility translates into more accurate short-term out-of-sample forecasts? We will answer this question using the data analyzed by Cribari-Neto et al. (2023) and an updated version of the time series. The interest lies in modeling and forecasting stored hydroelectric energy, and their focus was on (fixed precision) βARMA model selection. Their results indicate that model selection based on an Empirical Information Criterion (EIC), which uses parametric bootstrap resampling, typically outperforms model identification based on alternative strategies. We will also model and forecast a different time series: the useful volume of a water reservoir used for hydroelectric power generation. In both cases, the source of the data is the Brazilian National Electric System Operator (Operador Nacional do Sistema Elétrico – ONS, http://www.ons.org.br).

4.1. Stored hydroelectric energy

At the outset, our interest lies in modeling and forecasting the proportion of stored hydroelectric energy in South Brazil. Stored energy is the energy value of the accumulated water, i.e., how much energy (in Megawatt monthly) can be generated from the stored volume of water expressed as a proportion of the total hydroelectric power plant capacity. Our analysis is structured in four parts: (i) we use the monthly averages of stored energy from July 2000 to April 2018, totaling 214 observations, with the final six observations reserved for forecasts evaluation; the complete data contain $n = 208$ observations corresponding to the July 2000 to October 2017 period; (ii) we use data from July 2000 to May 2022, totaling 263 observations, with the final six data points reserved for forecasts evaluation, hence $n = 257$; (iii) we use 43 sample sizes in a sequential forecasting analysis; the sample sizes are $n \in \{45, 50, \ldots, 255\}$; (iv) we consider 211 samples of sizes $n \in \{45, 46, \ldots, 255\}$. The latter part is devoted to the evaluation of probabilistic forecasts. The data in the first part of our forecasting exercise are the same as used by Cribari-Neto et al. (2023). In the second, third, and fourth parts of the analysis, we work with an updated version of the time series. In all cases, our goal is to perform a comparative analysis between standard (fixed precision) and generalized (variable precision) βARMA forecasts. All estimations, descriptive analyses, and graphical analyses were carried out using the R statistical computing environment (R Core Team, 2023).

In all empirical analyses that follow, model selection was performed using three information criteria, namely AIC (Akaike), SIC (Schwarz), and EIC (empirical, bootstrap-based). For details on the former two criteria, see Burnham and Anderson (2004) and Choi (1992); for more information on the EIC, see Cavanaugh and Shumway (1997) and Cribari-Neto et al. (2023). EIC-based model selection was performed using 1,000 bootstrap replications. We search for the best model by considering all combinations of p and q, such as $p, q = 0, \ldots, 4$, except for the $(0, 0)$ model. The Q_4 portmanteau test proposed by Scher et al. (2020) was performed using the residuals from the selected model. Following Cribari-Neto et al. (2023), models for which the correct model specification is rejected by the Q_4 test at the 5% significance level are discarded. When that happens, the next best fitting model is selected according to the model selection criterion. In all fitted models, g_1 is the logit link, and in all fitted generalized βARMA models, g_2 is the identity link. We also considered the log and square root precision link functions; these results will not be shown for brevity. Slightly more accurate forecasts were obtained using the identity precision link.

Table 1
Descriptive statistics, stored hydroelectric energy in South Brazil.

n	min	max	median	mean	variance	skewness	kurtosis
214	0.30	0.99	0.73	0.70	0.04	−0.27	−1.22
263	0.15	0.99	0.64	0.66	0.05	−0.19	−1.04

Table 1 presents descriptive statistics for the two time series ($n = 214$ and $n = 263$). We report the maximal and minimal values and the means, medians, variances, coefficients of skewness, and coefficients of excess kurtosis. The longer time series displays the smaller minimal value, mean, and median. It also shows less skewness and excess kurtosis.

In the first part of our study, we follow Cribari-Neto et al. (2023) and split the series into three subsets of distinct sample sizes, namely $n = 75$ (Sample I), $n = 150$ (Sample II), and $n = 208$ (Sample III). The Q_4 test statistic was computed using 9, 13, and 14 lags for 75, 150, and 208 observations, respectively. We performed model selection for the generalized βARMA model. When $n = 75$ (Sample I), the EIC selected the βAR(3) model whereas the AIC and SIC selected the βARMA(1, 1) model. When $n = 150$ (Sample II), the EIC selected the βAR(3) model, and the βARMA(1, 1) model was selected by the other information criteria; the latter differs from the βARMA(2, 1) model selected by the same criteria under fixed precision in Cribari-Neto et al. (2023). Finally, when $n = 208$, the βARMA(2, 3) model was selected by the EIC, whereas the remaining criteria selected the βARMA(1, 1) model. Table 2 presents the values of the AIC, SIC, and EIC for the βARMA models with fixed and variable precision (denoted by ϕ and ϕ_t, respectively) selected by these criteria. The smallest values for each criterion are denoted in boldface. Notably, the three model selection criteria favor the generalized models over the standard models in all three samples.

Table 3 contains the conditional maximum likelihood estimates (standard errors in parentheses) of the parameters that index the selected generalized βARMA models. For each sample, the models above and below the horizontal line are those selected by the EIC and AIC/SIC, respectively. As expected, for all models $\hat{\alpha}_2 > 0$ and $\hat{\delta} < 0$.

Next, we test the null hypothesis of constant precision versus the alternative hypothesis of variable precision, i.e., we test $\mathcal{H}_0 : \delta = 0$ against $\mathcal{H}_1 : \delta \neq 0$. The likelihood ratio test p-values for the selected generalized βARMA models in the three sample sizes are presented in Table 4. All p-values are expressed as percentages. In all cases, fixed precision is rejected at the 5% significance

Table 2
Model selection criteria values for the selected standard and generalized models.

Sample	Criterion	ϕ	ϕ_t
I	AIC	−113.81	**−118.42**
	SIC	−104.54	**−106.83**
	EIC	−124.00	**−129.90**
II	AIC	−247.53	**−255.62**
	SIC	−235.49	**−240.57**
	EIC	−260.16	**−267.27**
III	AIC	−322.41	**−332.09**
	SIC	−309.06	**−315.40**
	EIC	−340.03	**−349.99**

level, and in some cases, rejection takes place at 1%. There is thus evidence in favor of variable dispersion.

We will now move to forecasting accuracy evaluation. Using the selected fixed and variable precision models, we produced, for each sample, forecasts of y_{n+h}, $h \in \{1, \ldots, 6\}$. That is, we produced forecasts for the next six observations. Table 5 contains the mean absolute prediction errors (MAPEs) of the forecasts computed as $\text{MAPE}(h) = h^{-1} \sum_{j=1}^{h} |y_{n+j} - \hat{y}_n(j)|$, with $\hat{y}_n(j)$ denoting the forecast of y_{n+j} made at time n. All figures are multiplied by 100, and the best result for each forecasting horizon is displayed in boldface.

The figures in Table 5 convey important information. First, more accurate forecasts are obtained using the generalized βARMA model in all three samples. Second, in Samples I and II, the models selected by the EIC yield the best results. In Sample III, the best results for $h \in \{1, 2, 3\}$ are obtained using the EIC for generalized βARMA model selection, whereas the most accurate forecasts for $h \in \{4, 5, 6\}$ are yielded by the variable precision model selected by the AIC and SIC. Third, in some cases, the gains in forecasting accuracy achieved by allowing variable precision are large. Consider, e.g., Sample I, $h = 2$ and βAR(3) (model selected by the EIC). The MAPE of the forecasts from the generalized model is over 30% smaller than that of the standard model's forecasts; in Sample III, $h = 1$ and βARMA(2, 3), the gain in MAPE is nearly 58%.

We will now move to the second part of our empirical investigation, in which more recent data are used. Here, the data range from July 2000 to May 2022, totaling 263 observations, and the final six observations are reserved for forecasting accuracy evaluation. Thus, the effective sample size is $n = 257$. The Q_4 portmanteau test statistic used to assess model misspecification employs 16 lags. The βARMA(1, 1) model was selected by all three criteria under fixed precision, with AIC $= -391.81$, SIC $= -377.62$ and EIC $= -392.63$. Considering variable precision, the βAR(3) model was selected by AIC, SIC, and EIC, with values −397.95, −378.66, and −399.25, respectively. The three criteria favor the generalized βARMA model. The p-values of the likelihood ratio test of $\mathcal{H}_0 : \delta = 0$ for the βARMA(1, 1) and βAR(3) models are 0.19% and 1.98%, respectively. We thus reject the null hypothesis of fixed precision. We only report precision estimates (standard errors in parentheses) for brevity. For the standard model, $\hat{\phi} = 10.96$ (0.95). For the generalized model, $\hat{\alpha}_2 = 19.70$ (3.78) and $\hat{\delta} = -36.73$ (17.94); again, as expected, these estimates are positive and negative, respectively. Fig. 2 contains the index plot of $\hat{\phi}_t$ with a dashed horizontal line at the precision parameter estimate from the fitted standard model (10.96). The generalized model's minimal and maximal estimated precisions are 10.52 and 19.20, respectively, with the average estimated precision being 13.18. As expected, the largest estimated precisions are associated with observations that are close to an endpoint of the standard unit interval. For instance, there are 56 estimated precisions in excess of 16, and they all coincide with observations that are somewhat close to one; the minimum value of these 56 data points is 0.89.

As noted earlier, changes in the beta density shape across observations are only driven by μ_t in the standard βARMA model; by contrast, in the more general formulation of the model, they are driven by μ_t and ϕ_t. The generalized model thus has an additional layer of flexibility since it allows the beta density shape to evolve more freely over time. To exemplify that, we present in Fig. 3 the beta density functions evaluated at the mean and precision estimates obtained from the two models ($\hat{\mu}_t$ and $\hat{\phi}_t$ for the generalized model and $\hat{\mu}_t$ and $\hat{\phi}$ for the standard model) for observations 48 (left panels), 49 (middle panels) and 50 (right panels). The top row panels are for the βARMA(1, 1) model, and the bottom row panels are for the βAR(3) model. The former was selected under fixed precision and the latter under variable precision. Each model was fitted assuming fixed and variable precision. Note that the shape of the beta density changes more intensely over the three time periods when we consider the mean and precision estimates obtained from the generalized model. In particular, notice that the fitted densities drawn in solid lines (generalized model) peak more intensely than those drawn in dashed lines (standard model) in the two middle panels (observation 49) while transitioning between two less skewed densities (which correspond to observations 48 and 50).

Again, forecasts of y_{n+h} for $h \in \{1, \ldots, 6\}$ were produced using the generalized and standard βARMA models. The MAPEs ($\times 100$) are presented in Table 6. The forecasts obtained under variable precision are more accurate for all forecasting horizons. When $h = 1$, the gain in accuracy from using the generalized model exceeds 12%.

In the third part of our empirical analysis, we consider 43 sample sizes that range from $n = 45$ to $n = 255$ in steps of five observations (i.e., $n \in \{45, 50, \ldots, 255\}$). For each sample size, a generalized and a standard βARMA model are selected using the EIC; as before, the Q_4 portmanteau test of the correct model specification is performed on the residuals from both models. For each sample size, forecasts of the next six observations ($h \in \{1, \ldots, 6\}$) are produced using each model, MAPEs are computed for each set of forecasts, and the ratios between the MAPEs of the forecasts from the generalized and standard models are computed. That is, we compute, for each n, $\text{MAPE}_r(h) = \text{MAPE}_g(h)/\text{MAPE}_s(h)$, where the subscripts '$g$' and '$s$' stand for 'generalized' and 'standard,' respectively. Values of $\text{MAPE}_r(h)$ smaller than one (greater than one) favor the generalized (standard) model. Overall, the results favor the generalized model, especially when the interest lies in short-term forecasting. For instance,

Table 3
Parameter estimates (standard errors in parentheses), generalized models.

Sample	Model	$\hat{\alpha}_1$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\alpha}_2$	$\hat{\delta}$
I	βAR(3)	0.19	0.93	−0.37	0.20	–	–	–	22.41	−51.51
		(0.11)	(0.10)	(0.13)	(0.09)	–	–	–	(6.88)	(33.95)
	βARMA(1, 1)	0.40	0.56	–	–	0.40	–	–	22.26	−54.73
		(0.15)	(0.10)	–	–	(0.11)	–	–	(6.51)	(31.11)
II	βAR(3)	0.21	0.96	−0.33	0.11	–	–	–	23.74	−52.16
		(0.07)	(0.08)	(0.10)	(0.07)	–	–	–	(5.33)	(25.29)
	βARMA(1, 1)	0.32	0.61	–	–	0.35	–	–	24.50	−58.02
		(0.10)	(0.07)	–	–	(0.08)	–	–	(5.32)	(24.71)
III	βARMA(2, 3)	0.15	0.40	−0.59	–	−0.54	−0.17	0.24	21.44	−47.76
		(0.06)	(0.15)	(0.12)	–	(0.16)	(0.09)	(0.09)	(4.15)	(19.56)
	βARMA(1, 1)	0.35	0.58	–	–	0.31	–	–	20.11	−43.84
		(0.08)	(0.06)	–	–	(0.07)	–	–	(3.78)	(17.70)

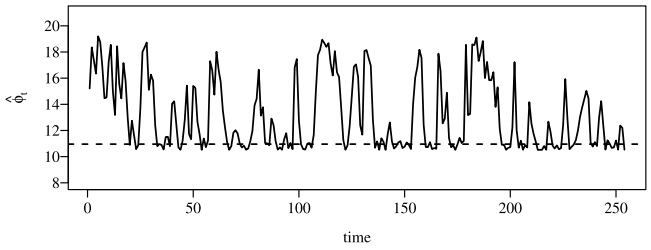

Fig. 2. Estimated precisions from the generalized model (solid line) and fixed precision estimate from the standard model (dashed line).

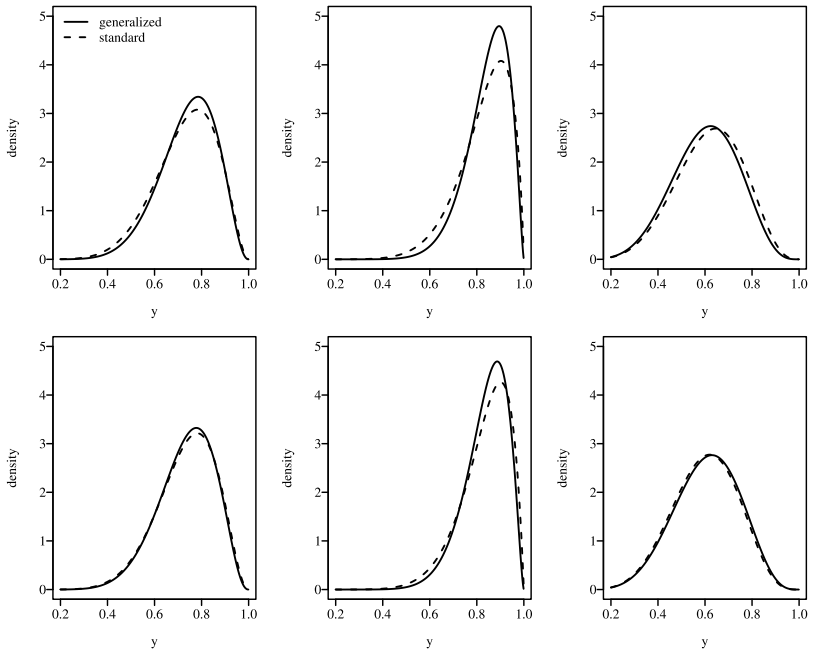

Fig. 3. Fitted beta densities for observations 48 (left panels), 49 (center panels), 50 (right panels) from the generalized (solid lines) and standard (dashed lines) βARMA(1, 1) (top row) and βAR(3) (bottom row) models.

for $h = 1$ ($h = 2$) [$h = 3$], the ratio mentioned above was smaller than one in 79.07% (76.04%) [74.42%] of the 43 sample sizes. In Fig. 4, we present, in six panels, plots of $\text{MAPE}_r(h)$ against the sample size, each panel corresponding to a forecasting horizon. Points that lie below (above) the horizontal line drawn at 1.0 indicate better (worse) forecasting accuracy of the generalized βARMA model relative to the standard model. The former outperforms the latter, especially for $h \in \{1, 2, 3\}$.

We also computed the mean values of $\text{MAPE}_r(h)$ (i.e., average over the 43 sample sizes) for each h. The figure for $h = 1$ ($h = 2$) [$h = 3$] is 0.83 (0.85) [0.88]. It is thus clear

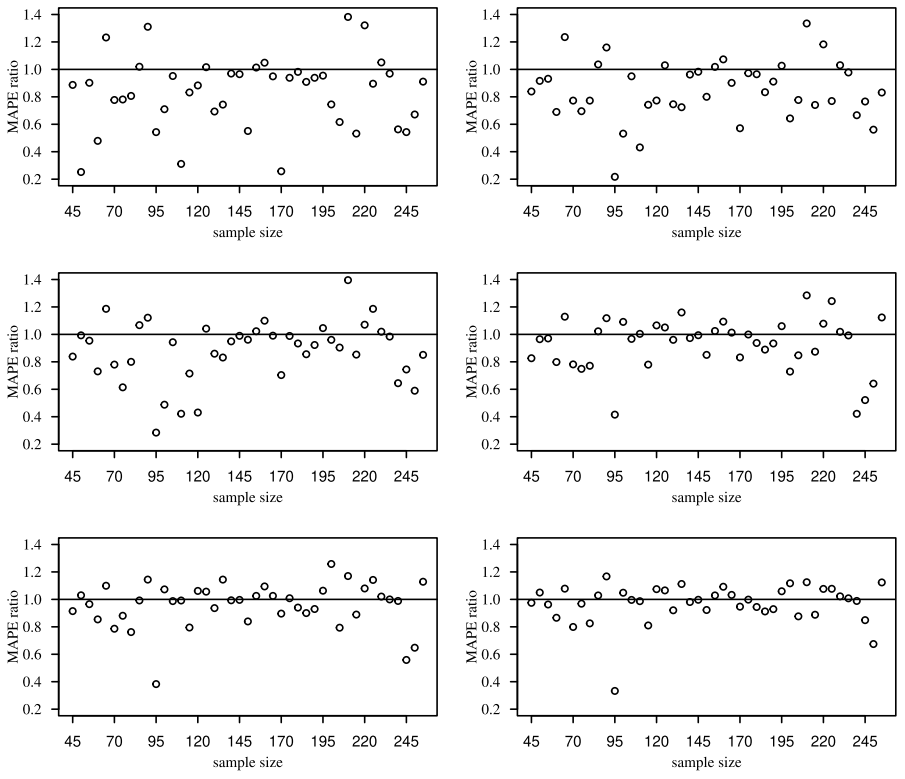

Fig. 4. $\text{MAPE}_r(h)$ vs $n \in \{45, 50, \ldots, 255\}$: $h = 1$: top left; $h = 2$: top right; $h = 3$: middle left; $h = 4$: middle right; $h = 5$: bottom left; $h = 6$: bottom right.

Table 4
p-values (%) of the likelihood ratio test of constant precision ($\mathcal{H}_0 : \delta = 0$).

Sample	Model	p-value
I	βAR(3)	4.63%
	βARMA(1, 1)	1.35%
II	βAR(3)	2.26%
	βARMA(1, 1)	0.02%
III	βARMA(2, 3)	0.02%
	βARMA(1, 1)	<0.01%

that the generalized model yielded forecasts that were, on average, considerably more accurate than those from the standard model. It also outperformed the standard model for larger forecasting horizons ($h \in \{4, 5, 6\}$), but by smaller margins; e.g., the mean MAPE ratio for $h = 4$ ($h = 5$) [$h = 6$] is 0.93 (0.96) [0.97]. It is natural for the gains in forecasting accuracy achieved by the generalized model to be more pronounced for small forecasting horizons since the two sets of forecasts (generalized and standard) converge to $\bar{y}$, the average value of y_t's, as h increases.

The generalized model is favored by the AIC, SIC, and EIC in nearly all 43 samples used in the analysis above. In 42 (38) [40] out of the 43 samples, the EIC (AIC) [SIC] of the selected generalized model is smaller than that of the selected standard model.

The fourth part of our empirical analysis focuses on probabilistic forecasting; see, e.g., Gneiting and Katzfuss (2014). A probabilistic forecast is a predictive probability distribution over future quantities or events of interest. We consider 211 sample sizes that range from $n = 45$ to $n = 255$ in steps of one observation (i.e., $n \in \{45, 46, \ldots, 255\}$). We select a generalized and standard model for each sample size based on the EIC with 500 bootstrap replications. Again, the Q_4 portmanteau test is carried out on the residuals from the two models. For each sample, one-step-ahead forecasts are obtained from the two competing models. At the outset, we note that the generalized βARMA forecasts correlate more strongly with the observed values than those obtained from the standard model: 0.84 vs. 0.80. Both sets of forecasts are used to compute the probabilities of occurrence of three events related to the next value of the series: (i) $y_{n+1} < 0.2$, (ii) $0.2 \leq y_{n+1} \leq 0.8$, and (iii) $y_{n+1} > 0.8$. Such probabilities are computed using beta distributions with parameters $(\hat{\mu}_{n+1}, \hat{\phi}_{n+1})$ and $(\hat{\mu}_{n+1}, \hat{\phi})$ according to the generalized and standard models, respectively. Using these probabilities and the realized values of the series $(y_{46}, \ldots, y_{256})$, we constructed receiver operating characteristic (ROC) plots for the three events of interest. They are presented in Fig. 5. The left, middle, and right panels are, respectively, for low ($y_{n+1} < 0.2$), normal ($0.2 \leq y_{n+1} \leq 0.8$), and high ($y_{n+1} > 0.8$) levels of stored hydroelectric energy. The black and red curves are for predicted probabilities obtained from the generalized and standard models. We also include in the plots the corresponding figures for the areas under the curve (AUC).

Table 5
Mean absolute prediction errors ($\times 100$), first empirical analysis.

Sample	Model	Precision	MAPE					
			$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
I	βAR(3)	ϕ_t	**7.73**	**6.53**	**4.92**	**4.65**	**5.38**	**7.37**
		ϕ	9.90	9.39	8.01	6.21	6.11	7.61
	βARMA(1, 1)	ϕ_t	12.02	12.61	11.97	10.27	8.23	8.48
		ϕ	12.96	13.39	12.22	9.95	8.53	9.30
II	βAR(3)	ϕ_t	**1.40**	**6.69**	**5.68**	**4.61**	**5.59**	**7.27**
		ϕ	2.54	8.36	5.91	5.42	6.66	7.88
	βARMA(1, 1)	ϕ_t	2.48	8.23	5.83	5.49	6.80	7.91
	βARMA(2, 1)	ϕ	3.16	8.41	6.47	5.47	6.42	7.92
III	βARMA(2, 3)	ϕ_t	**0.27**	**1.33**	**7.09**	7.13	5.75	5.94
		ϕ	0.64	2.28	7.38	7.34	6.04	5.85
	βARMA(1, 1)	ϕ_t	0.51	3.70	7.50	**7.01**	**5.67**	**5.68**
		ϕ	0.88	3.69	7.75	7.48	6.21	5.86

Table 6
Mean absolute prediction errors ($\times 100$), second empirical analysis.

Sample	Model	Precision	MAPE					
			$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
$n=257$	βAR(3)	ϕ_t	**11.72**	**16.90**	**21.82**	**20.37**	**17.17**	**18.85**
	βARMA(1, 1)	ϕ	13.32	18.65	23.42	21.79	18.19	19.65

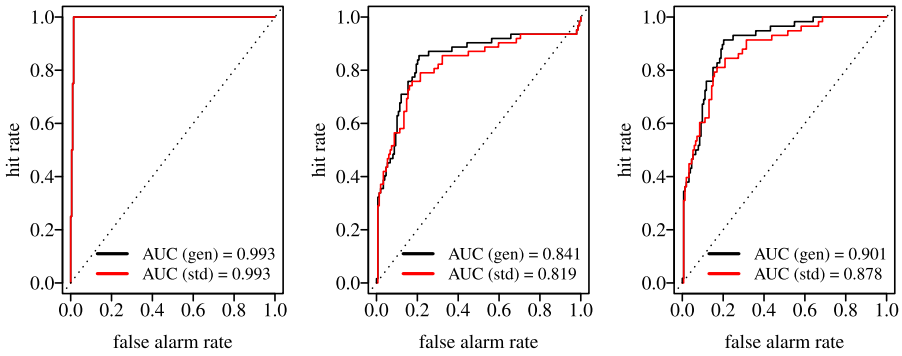

Fig. 5. ROC curves obtained using one-step ahead forecasts from the generalized (black curve) and standard (red curve) models; the left, middle, and right panels correspond, respectively, to the events $y_{n+1} < 0.2$, $0.2 \leq y_{n+1} \leq 0.8$, and $y_{n+1} > 0.8$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The higher the curve and the larger the AUC value, the more accurate the forecasts. For details on ROC curves and AUC measures, see Marzban (2004). Fig. 5 shows that the two models yield accurate detection of next-period low stored energy, with AUC equal to 0.99 for both sets of forecasts. The generalized model is the best performer for the other two events (normal and high stored energy in the next time period). For instance, in the right panel, the AUC values for the generalized and standard models are 0.90 and 0.88, respectively.

Using one-step ahead forecasts obtained from the fitted models, we obtain the corresponding predicted probability distributions and use them to construct a reliability diagram; see Hsu and Murphy (1986). It is presented in Fig. 6. Overall, there is better agreement between forecast probabilities and observed frequencies for the generalized model. The agreement is excellent for forecast probabilities in (0, 0.2) and (0.8, 1). The agreement is reasonably good even in the middle range of probabilities, where there is more discrepancy between the observed frequencies and forecast probabilities. For instance, when the forecast probability is 0.5, the corresponding observed frequency is 0.57. Notably, the curve that relates observed frequencies and forecast probabilities for the standard model is S-shaped.

Using the one-step ahead probabilistic predictions and the observed values, we compute the continuous ranked probability scores (CRPSs) of the forecasts yielded by the generalized and standard models. For details on such scores, see Gneiting and Raftery (2007) and Gneiting and Katzfuss (2014). We use the expression for the CRPS of beta predictive distributions given in Taillardat, Mestre, Zamo, and Naveau (2016). We obtained the following CRPS score values ($\times 100$) for the forecasts from the generalized and standard models, respectively: 6.74 and 8.89. This result favors the generalized model. The skill score of the generalized model relative to the standard model is $(\text{CRPS}_s - \text{CRPS}_g)/\text{CRPS}_s = 0.24$. The generalized model is

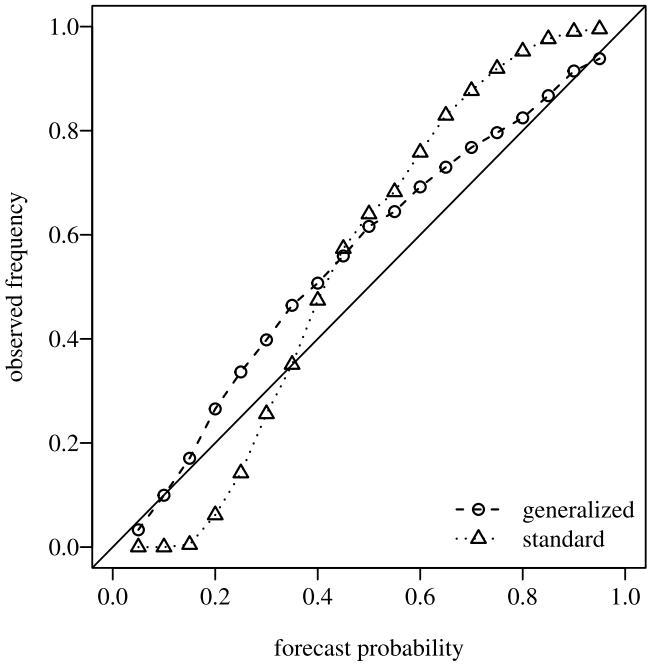

Fig. 6. Reliability diagram for one-step ahead forecasts obtained from the generalized (circles and dashed line) and standard (triangles and dotted line) models.

thus 24% better than the standard model when it comes to short-run forecasting.

4.2. Useful volume of a water reservoir

Our next empirical analysis also deals with hydrological data related to power generation. Hydropower plants cannot utilize all the water stored in their reservoirs for power generation. The 'useful volume' represents the volume of water (in m^3) in the reservoir that can be effectively used for power generation. The reservoir's useful volume is typically expressed as the ratio between (i) the current level of the reservoir and (ii) the difference between maximum and minimum normal operating levels.

In what follows, we will use data on the useful volume (expressed as proportions) of an accumulation reservoir of the Iguaçu basin, which accounts for 51.19% of the power generation in the Southern region of Brazil. More specifically, we focus on the Governador Bento Munhoz da Rocha Netto reservoir, located in the Pinhão county (the state of Paraná). The data cover the period from July 2000 to May 2022, totaling 263 observations, with the last six data points reserved for forecasting accuracy evaluation. Thus, $n = 257$. Table 7 presents descriptive statistics on the monthly useful volume series. (The exact figure for the maximal useful volume is 0.9971.) Notably, the time series sample variance is nearly 60% larger than that of the stored energy series for the same time period.

Under fixed precision, the AIC, SIC, and EIC select the βARMA(1, 1) model, with corresponding values −251.98, −237.78, and −259.62. Under variable precision, the βAR(3) model is selected by the three information criteria, with values −253.99, −238.70, and −261.47, respectively. It is important to note that all three criteria favor the generalized model. The parameter estimates (standard errors in parentheses) for the fixed precision βARMA(1, 1) model are $\hat{\alpha} = 0.29$ (0.07), $\hat{\varphi}_1 = 0.37$ (0.04), $\hat{\theta}_1 = 0.17$ (0.05) and $\hat{\phi} = 4.42$ (0.37). For the variable precision βAR(3) model, we obtain $\hat{\alpha}_1 = 0.22$ (0.07), $\hat{\varphi}_1 = 0.57$ (0.04), $\hat{\varphi}_2 = -0.10$ (0.04), $\hat{\varphi}_3 = 0.16$ (0.04), $\hat{\alpha}_2 = 5.87$ (0.98) and $\hat{\delta} = -9.03$ (3.09).

Table 7
Descriptive statistics, Iguaçu basin useful volume.

min	max	median	mean	variance	skewness	kurtosis
0.03	∼1.00	0.69	0.64	0.07	−0.26	−1.19

Table 8
Mean absolute prediction errors (×100), useful volume forecasting.

Model	Precision	MAPE					
		$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
βAR(3)	ϕ_t	**11.64**	**16.71**	**21.75**	**16.91**	**18.52**	**21.47**
βARMA(1, 1)	ϕ	14.85	19.42	23.84	18.38	19.64	22.39

Table 8 contains the MAPEs (×100) of the forecasts of y_{n+h} with $h \in \{1, \ldots, 6\}$. The generalized model outperforms the standard model in all forecasting horizons. The gains in forecasting accuracy from allowing for variable precision are especially large for $h = 1$ and $h = 2$ (21.62% and 13.95%, respectively).

5. Concluding remarks

The βARMA model proposed by Rocha and Cribari-Neto (2009) and Rocha and Cribari-Neto (2017) is useful for modeling random variables that assume values in (0, 1) and evolve over time. The model extends the class of beta regressions because it can be used with non-independent, serially correlated random variables. Like its regression counterpart, it accommodates distributional asymmetries, accounts for heteroskedasticity, and does not yield improper predictions. Both models are based on the assumption that the variable of interest is beta-distributed. In the dynamic model, such an assumption is made for the variable of interest at each time period conditional on the previous information set. As is well known, the beta law is very flexible since its density can assume many shapes depending on the parameter values. Also, both models are based on the beta parametrization proposed by Ferrari and Cribari-Neto (2004), according to which the beta density is indexed by the distribution mean and a precision parameter.

The standard formulation of the βARMA model allows the mean parameter to evolve over time. Still, it imposes that the precision is globally fixed, i.e., the precision parameter has the same value for all observations. In this paper, we introduced a more general formulation of the model, which allows the two parameters that index the beta law (the conditional mean and the conditional precision) to vary over time. We developed conditional maximum likelihood inference on the proposed model's parameters. In particular, we presented closed-form expressions for the model's conditional log-likelihood function, conditional score function, and conditional expected information matrix.

The proposed model was used to forecast future levels of stored hydroelectric energy in the South of Brazil. Different sample sizes were considered, and, for each sample size, generalized and standard βARMA models were selected using three different information criteria. One of the model selection criteria we used employs bootstrap resampling. In most configurations, more accurate out-of-sample forecasts were obtained using the model's more general formulation. In some cases, the gains in forecasting accuracy were large. Different criteria were used to evaluate probabilistic forecasts obtained from the best fitting generalized and standard models. They all favored the former. We also predicted the future values of the useful volume of a water reservoir used for hydroelectric power generation located in the South region of Brazil. Again, the general model proposed in this paper outperformed the standard model. Notably, the generalized model was favored over the standard model by the three model selection criteria in nearly all empirical analyses presented and discussed in the previous section. For instance, the former was favored by the EIC (a bootstrap-based information criterion) in 42 out of the 43 samples used in a sequential analysis of stored energy data.

The general formulation of the model is more flexible than the fixed precision model in that changes in the beta density shape over time are driven by two parameters (mean and precision), not only by a single parameter (mean) like in the standard model formulation. We encourage practitioners who wish to model and forecast double bounded time series to use the generalized βARMA model proposed in this paper.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Francisco Cribari-Neto reports financial support was provided by National Council for Scientific and Technological Development. Fabio Mariano Bayer reports financial support was provided by National Council for Scientfic and Technological Research. Vinicius Teodoro Scher reports financial support was provided by Foundation for Support of Science and Technology of Pernambuco State.

Acknowledgment

We gratefully acknowledge financial support from Fundação de Apoio à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), Brazil, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, and Centro Nacional de Supercomputação da Universidade Federal do Rio Grande do Sul (CESUP/UFRGS), Brazil. We also thank two anonymous referees for comments and suggestions that led to a much improved manuscript.

Appendix A. Conditional score function

We will now derive the components of the conditional score vector. The derivative of ℓ with respect to α_1 is

$$\frac{\partial \ell}{\partial \alpha_1} = \sum_{t=a+1}^{n} \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{\partial \mu_t}{\partial \eta_{1t}} \frac{\partial \eta_{1t}}{\partial \alpha_1}.$$

It is important to note that $\partial \mu_t / \partial \eta_{1t} = 1/g_1'(\mu_t)$ and

$$\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} = \phi_t \left\{ \log\left(\frac{y_t}{1-y_t}\right) - [\psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)] \right\},$$

where $\psi(\cdot)$ is the digamma function. Let $y_t^* = \log(y_t/(1-y_t))$ and $\mu_t^* = \psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)$. Hence,

$$\frac{\partial \ell}{\partial \alpha_1} = \sum_{t=a+1}^{n} \phi_t (y_t^* - \mu_t^*) \frac{1}{g_1'(\mu_t)} \left(1 + \sum_{j=1}^{q} \theta_j \frac{\partial r_{t-j}}{\partial \alpha_1}\right).$$

Let $\boldsymbol{s}$ be the $(n-a)$-dimensional vector with ith element given by

$$\frac{\partial \eta_{1(i+a)}}{\partial \alpha_1} = 1 + \sum_{j=1}^{q} \theta_j \frac{\partial r_{i+a-j}}{\partial \alpha_1},$$

$\boldsymbol{y}^* = (y_{a+1}^*, \ldots, y_n^*)^\top$ and $\boldsymbol{\mu}^* = (\mu_{a+1}^*, \ldots, \mu_n^*)^\top$. It thus follows that

$$U_{\alpha_1} = \boldsymbol{s}^\top \Phi T_1 (\boldsymbol{y}^* - \boldsymbol{\mu}^*),$$

where $\Phi = \text{diag}\{\phi_{a+1}, \ldots, \phi_n\}$ and $T_1 = \text{diag}\{1/g_1'(\mu_{a+1}), \ldots, 1/g_1'(\mu_n)\}$.

Additionally, for $l \in \{1, \ldots, c\}$,

$$\frac{\partial \ell}{\partial \beta_l} = \sum_{t=a+1}^{n} \phi_t (y_t^* - \mu_t^*) \frac{1}{g_1'(\mu_t)}$$

$$\times \left(x_{tl} - \sum_{i=1}^{p} \varphi_i x_{(t-i)l} + \sum_{j=1}^{q} \theta_j \frac{\partial r_{t-j}}{\partial \beta_l} \right).$$

Let M be the $(n-a) \times c$ matrix whose ith row is

$$\frac{\partial \eta_{1(i+a)}}{\partial \boldsymbol{\beta}} = \boldsymbol{x}_{i+a} - \sum_{i=1}^{p} \varphi_i \boldsymbol{x}_{i+a} + \sum_{j=1}^{q} \theta_j \frac{\partial r_{i+a-j}}{\partial \boldsymbol{\beta}}.$$

We obtain

$$\boldsymbol{U}_{\boldsymbol{\beta}} = M^\top \Phi T_1(\boldsymbol{y}^* - \boldsymbol{\mu}^*).$$

For $i \in \{1, \ldots, p\}$, we have

$$\frac{\partial \ell}{\partial \varphi_i} = \sum_{t=a+1}^{n} \phi_t (y_t^* - \mu_t^*) \frac{1}{g_1'(\mu_t)} \times \left(g_1(y_{t-i}) - \boldsymbol{x}_{t-i}^\top \boldsymbol{\beta} + \sum_{j=1}^{q} \theta_j \frac{\partial r_{t-j}}{\partial \varphi_i} \right).$$

Let P be the $(n-a) \times p$ matrix whose (i,j)th element is

$$\frac{\partial \eta_{1(i+a)}}{\partial \varphi_j} = g_1(y_{i+a-j}) - \boldsymbol{x}_{i+a-j}^\top \boldsymbol{\beta} + \sum_{l=1}^{q} \theta_l \frac{\partial r_{i+a-l}}{\partial \varphi_j}.$$

Thus,

$$\boldsymbol{U}_{\boldsymbol{\varphi}} = P^\top \Phi T_1(\boldsymbol{y}^* - \boldsymbol{\mu}^*).$$

The derivative of ℓ with respect to θ_j, for $j \in \{1, \ldots, q\}$, is

$$\frac{\partial \ell}{\partial \theta_j} = \sum_{t=a+1}^{n} \phi_t (y_t^* - \mu_t^*) \frac{1}{g_1'(\mu_t)} \left(r_{t-j} + \sum_{j=1}^{q} \theta_j \frac{\partial r_{t-j}}{\partial \theta_j} \right).$$

Let R be the $(n-a) \times q$ matrix whose (i,j)th element is

$$\frac{\partial \eta_{1(i+a)}}{\partial \theta_j} = r_{i+a-j} + \sum_{l=1}^{q} \theta_l \frac{\partial r_{i+a-l}}{\partial \theta_j}.$$

Therefore,

$$\boldsymbol{U}_{\boldsymbol{\theta}} = R^\top \Phi T_1(\boldsymbol{y}^* - \boldsymbol{\mu}^*).$$

The conditional score function for α_2 is

$$\frac{\partial \ell}{\partial \alpha_2} = \sum_{t=a+1}^{n} \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \alpha_2}.$$

Here, $\partial \phi_t / \partial \eta_{2t} = 1/g_2'(\phi_t)$ and

$$\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} = \mu_t(y_t^* - \mu_t^*) + \log(1 - y_t) - \psi((1-\mu_t)\phi_t) + \psi(\phi_t).$$

Note that $g_2'(\phi_t) = 1$ when the identity precision link function is used. Let

$$H = \operatorname{diag}\{\mu_{a+1}(y_{a+1}^* - \mu_{a+1}^*) + \log(1 - y_{a+1}) - \psi((1-\mu_{a+1})\phi_{a+1}) + \psi(\phi_{a+1}), \ldots, \mu_n(y_n^* - \mu_n^*) + \log(1 - y_n) - \psi((1-\mu_n)\phi_n) + \psi(\phi_n)\}$$

and $T_2 = \operatorname{diag}\{1/g_2'(\phi_{a+1}), \ldots, 1/g_2'(\phi_n)\}$. Thus,

$$U_{\alpha_2} = \mathbf{1}_n^\top H T_2 \mathbf{1}_n,$$

where $\mathbf{1}_n$ is an $(n-a) \times 1$ vector of ones. When g_2 is the identity link, T_2 is the $(n-a)$-dimensional identity matrix, and hence $U_{\alpha_2} = \mathbf{1}_n^\top H \mathbf{1}_n$.

Finally,

$$\frac{\partial \ell}{\partial \delta} = \sum_{t=a+1}^{n} [\mu_t(y_t^* - \mu_t^*) + \log(1 - y_t) - \psi((1-\mu_t)\phi_t) + \psi(\phi_t)] \times \frac{1}{g_2'(\phi_t)} z_{t-1}.$$

Let $\boldsymbol{\omega}$ is the $(n-a)$-dimensional vector given by $\boldsymbol{\omega} = (z_a, \ldots, z_{n-1})^\top$. Therefore,

$$U_\delta = \boldsymbol{\omega}^\top H T_2 \mathbf{1}_n.$$

When the precision link function is identity, $U_\delta = \boldsymbol{\omega}^\top H \mathbf{1}_n$.

Appendix B. Conditional information matrix

We will now derive the elements of the conditional Fisher information matrix. To that end, we must obtain closed-form expressions for the conditional log-likelihood second-order derivatives. Let $\boldsymbol{\lambda} = (\alpha_1, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top)^\top$ and $\boldsymbol{\gamma} = (\alpha_2, \delta)^\top$. We have

$$\frac{\partial^2 \ell}{\partial \lambda_i \partial \lambda_j} = \sum_{t=a+1}^{n} \frac{\partial}{\partial \mu_t} \left(\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{\partial \mu_t}{\partial \eta_{1t}} \frac{\partial \eta_{1t}}{\partial \lambda_j} \right) \frac{\partial \mu_t}{\partial \eta_{1t}} \frac{\partial \eta_{1t}}{\partial \lambda_i}$$
$$= \sum_{t=a+1}^{n} \left[\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \mu_t^2} \frac{\partial \mu_t}{\partial \eta_{1t}} \frac{\partial \eta_{1t}}{\partial \lambda_j} + \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{\partial}{\partial \mu_t} \left(\frac{\partial \mu_t}{\partial \eta_{1t}} \frac{\partial \eta_{1t}}{\partial \lambda_j} \right) \right] \times \frac{\partial \mu_t}{\partial \eta_{1t}} \frac{\partial \eta_{1t}}{\partial \lambda_i},$$

$$\frac{\partial^2 \ell}{\partial \gamma_i \partial \gamma_j} = \sum_{t=a+1}^{n} \frac{\partial}{\partial \phi_t} \left(\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_j} \right) \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_i}$$
$$= \sum_{t=a+1}^{n} \left[\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \phi_t^2} \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_j} + \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \times \frac{\partial}{\partial \phi_t} \left(\frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_j} \right) \right] \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_i},$$

$$\frac{\partial^2 \ell}{\partial \lambda_i \partial \gamma_j} = \sum_{t=a+1}^{n} \frac{\partial}{\partial \phi_t} \left(\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{\partial \mu_t}{\partial \eta_{1t}} \frac{\partial \eta_{1t}}{\partial \gamma_j} \right) \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \lambda_i}$$
$$= \sum_{t=a+1}^{n} \left[\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \phi_t \partial \mu_t} \frac{\partial \mu_t}{\partial \eta_{1t}} \frac{\partial \eta_{1t}}{\partial \gamma_j} + \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \times \frac{\partial}{\partial \phi_t} \left(\frac{\partial \mu_t}{\partial \eta_{1t}} \frac{\partial \eta_{1t}}{\partial \gamma_j} \right) \right] \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \lambda_i}.$$

Since $\mathbb{E}(\partial \ell_t(\mu_t, \phi_t)/\partial \mu_t | \mathcal{F}_{t-1}) = \mathbb{E}(\partial \ell_t(\mu_t, \phi_t)/\partial \phi_t | \mathcal{F}_{t-1}) = 0$, we have

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \lambda_i \partial \lambda_j} \middle| \mathcal{F}_{t-1} \right) = \sum_{t=a+1}^{n} \mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1} \right)$$

$$\times \left(\frac{\partial \mu_t}{\partial \eta_{1t}}\right)^2 \frac{\partial \eta_{1t}}{\partial \lambda_j}\frac{\partial \eta_{1t}}{\partial \lambda_i},$$

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \gamma_i \partial \gamma_j}\middle|\mathcal{F}_{t-1}\right) = \sum_{t=a+1}^{n} \mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial \phi_t^2}\middle|\mathcal{F}_{t-1}\right)\left(\frac{\partial \phi_t}{\partial \eta_{2t}}\right)^2 \times \frac{\partial \eta_{2t}}{\partial \gamma_j}\frac{\partial \eta_{2t}}{\partial \gamma_i},$$

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \lambda_i \partial \gamma_j}\middle|\mathcal{F}_{t-1}\right) = \sum_{t=a+1}^{n} \mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial \phi_t \partial \mu_t}\middle|\mathcal{F}_{t-1}\right) \times \left(\frac{\partial \mu_t}{\partial \eta_{1t}}\right)\left(\frac{\partial \phi_t}{\partial \eta_{2t}}\right) \times \frac{\partial \eta_{1t}}{\partial \lambda_j}\frac{\partial \eta_{2t}}{\partial \gamma_i}.$$

Using

$$\begin{aligned}
\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial \mu_t^2} &= -\phi_t^2[\psi'(\mu_t\phi_t)+\psi'((1-\mu_t)\phi_t)],\\
\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial \phi_t^2} &= -\mu_t^2(\psi'(\mu_t\phi_t)) - (1-\mu_t)^2 \times [\psi'((1-\mu_t)\phi_t)] + \psi'(\phi_t),\\
\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial \phi_t \partial \mu_t} &= \psi((1-\mu_t)\phi_t) - \psi(\mu_t\phi_t) + (1-\mu_t)\phi_t\psi'((1-\mu_t)\phi_t) - \mu_t\phi_t\psi'(\mu_t\phi_t) + \log\left(\frac{y_t}{1-y_t}\right),
\end{aligned}$$

we obtain

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \lambda_i \partial \lambda_j}\middle|\mathcal{F}_{t-1}\right) = -\sum_{t=a+1}^{n} \frac{A_t}{g_1'(\mu_t)^2}\frac{\partial \eta_{1t}}{\partial \lambda_j}\frac{\partial \eta_{1t}}{\partial \lambda_i},$$

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \gamma_i \partial \gamma_j}\middle|\mathcal{F}_{t-1}\right) = -\sum_{t=a+1}^{n} \frac{B_t}{g_2'(\phi_t)^2}\frac{\partial \eta_{2t}}{\partial \gamma_j}\frac{\partial \eta_{2t}}{\partial \gamma_i},$$

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \lambda_i \partial \gamma_j}\middle|\mathcal{F}_{t-1}\right) = -\sum_{t=a+1}^{n} \frac{C_t}{g_1'(\mu_t)g_2'(\phi_t)}\frac{\partial \eta_{1t}}{\partial \lambda_j}\frac{\partial \eta_{2t}}{\partial \gamma_i},$$

where $A_t = \phi_t^2[\psi'(\mu_t\phi_t)+\psi'((1-\mu_t)\phi_t)]$, $B_t = \mu_t^2(\psi'(\mu_t\phi_t))+(1-\mu_t)^2[\psi'((1-\mu_t)\phi_t)]-\psi'(\phi_t)$, $C_t = \psi(\mu_t\phi_t)-\psi((1-\mu_t)\phi_t)-(1-\mu_t)\phi_t\psi'((1-\mu_t)\phi_t)+\mu_t\phi_t\psi'(\mu_t\phi_t)-\log(y_t/(1-y_t))$ and $\psi'(\cdot)$ is the trigamma function.

Let $W_1 = \text{diag}\{w_{1(a+1)},\ldots,w_{1(n)}\}$, $W_2 = \text{diag}\{w_{2(a+1)},\ldots,w_{2(n)}\}$ and $W_3 = \text{diag}\{w_{3(a+1)},\ldots,w_{3(n)}\}$, with

$$w_{1(t)} = \frac{A_t}{g_1'(\mu_t)^2}, \quad w_{2(t)} = \frac{B_t}{g_2'(\phi_t)^2} \quad \text{and} \quad w_{3(t)} = \frac{C_t}{g_1'(\mu_t)g_2'(\phi_t)}.$$

Thus,

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \alpha_1^2}\middle|\mathcal{F}_{t-1}\right) = -\boldsymbol{s}^\top W_1 \boldsymbol{s}, \quad \mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \alpha_1}\middle|\mathcal{F}_{t-1}\right) = -M^\top W_1 \boldsymbol{s},$$

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\varphi}\partial \alpha_1}\middle|\mathcal{F}_{t-1}\right) = -P^\top W_1 \boldsymbol{s}, \quad \mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\theta}\partial \alpha_1}\middle|\mathcal{F}_{t-1}\right) = -R^\top W_1 \boldsymbol{s},$$

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \alpha_1 \partial \alpha_2}\middle|\mathcal{F}_{t-1}\right) = -\boldsymbol{s}^\top W_3 \mathbf{1}_n, \quad \mathbb{E}\left(\frac{\partial^2 \ell}{\partial \alpha_1 \partial \delta}\middle|\mathcal{F}_{t-1}\right) = -\boldsymbol{s}^\top W_3 \boldsymbol{\omega},$$

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top}\middle|\mathcal{F}_{t-1}\right) = -M^\top W_1 M, \quad \mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \boldsymbol{\varphi}^\top}\middle|\mathcal{F}_{t-1}\right) = -M^\top W_1 P,$$

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \boldsymbol{\theta}^\top}\middle|\mathcal{F}_{t-1}\right) = -M^\top W_1 R, \quad \mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \alpha_2}\middle|\mathcal{F}_{t-1}\right) = -M^\top W_3 \mathbf{1}_n,$$

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \delta}\middle|\mathcal{F}_{t-1}\right) = -M^\top W_3 \boldsymbol{\omega}, \quad \mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}^\top}\middle|\mathcal{F}_{t-1}\right) = -P^\top W_1 P,$$

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\theta}^\top}\middle|\mathcal{F}_{t-1}\right) = -P^\top W_1 R, \quad \mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\varphi}\partial \alpha_2}\middle|\mathcal{F}_{t-1}\right) = -P^\top W_3 \mathbf{1}_n,$$

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\varphi}\partial \delta}\middle|\mathcal{F}_{t-1}\right) = -P^\top W_3 \boldsymbol{\omega}, \quad \mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\middle|\mathcal{F}_{t-1}\right) = -R^\top W_1 R,$$

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\theta}\partial \alpha_2}\middle|\mathcal{F}_{t-1}\right) = -R^\top W_3 \mathbf{1}_n, \quad \mathbb{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\theta}\partial \delta}\middle|\mathcal{F}_{t-1}\right) = -R^\top W_3 \boldsymbol{\omega},$$

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \delta \partial \alpha_2}\middle|\mathcal{F}_{t-1}\right) = -\boldsymbol{\omega}^\top W_2 \mathbf{1}_n, \quad \mathbb{E}\left(\frac{\partial^2 \ell}{\partial \delta^2}\middle|\mathcal{F}_{t-1}\right) = -\boldsymbol{\omega}^\top W_2 \boldsymbol{\omega}.$$

References

Aknouche, A., & Dimitrakopoulos, S. (2023). Autoregressive conditional proportion: A multiplicative-error model for (0, 1)-valued time series. *Journal of Time Series Analysis*, *44*(4), 393–417. http://dx.doi.org/10.1111/jtsa.12679.

AL-Dhurafi, N. A., Masseran, N., & Zamzuri, Z. H. (2018). Compositional time series analysis for air pollution index data. *Stochastic Environmental Research and Risk Assessment*, *32*, 2903–2911. http://dx.doi.org/10.1007/s00477-018-1542-0.

Bayer, F. M., Bayer, D. M., & Pumi, G. (2017). Kumaraswamy autoregressive moving average models for double bounded environmental data. *Journal of Hydrology*, *555*, 385–396. http://dx.doi.org/10.1016/j.jhydrol.2017.10.006.

Benjamim, M. A., Rigby, R. A., & Stasinopoulos, M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association*, *98*(461), 214–223. http://dx.doi.org/10.1198/016214503388619238.

Brunsdon, T. M., Smith, M. F., & T (1998). The time series analysis of compositional data. *Journal of Official Statistics*, *14*(3), 237–253.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods Research*, *33*(2), 261–304. http://dx.doi.org/10.1177/0049124104268644.

Casarin, R., Valle, L. D., & Leisen, F. (2012). Bayesian model selection for beta autoregressive processes. *Bayesian Analysis*, *7*(2), 385–410. http://dx.doi.org/10.1214/12-BA713.

Cavanaugh, J. E., & Shumway, R. H. (1997). A bootstrap variant of AIC for state-space model selection. *Statistica Sinica*, *7*(2), 473–496.

Choi, B. S. (1992). *ARMA model identification*. New York: Springer-Verlag.

Cribari-Neto, F., Scher, V. T., & Bayer, F. M. (2023). Beta autoregressive moving average model selection with application to modeling and forecasting stored hydroelectric energy. *International Journal of Forecasting*, *39*(1), 98–109. http://dx.doi.org/10.1016/j.ijforecast.2021.09.004.

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, *34*, 1–24. http://dx.doi.org/10.18637/jss.v034.i02.

Douma, J. C., & Weedon, J. T. (2019). Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. *Methods in Ecology and Evolution*, *10*(9), 1412–1430. http://dx.doi.org/10.1111/2041-210X.13234.

Ferrari, S. L. P., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, *31*(7), 799–815. http://dx.doi.org/10.1080/0266476042000214501.

Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, *1*(1), 125–151. http://dx.doi.org/10.1146/annurev-statistics-062713-085831.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. http://dx.doi.org/10.1198/016214506000001437.

Guolo, A., & Varin, C. (2014). Beta regression for time series analysis of bounded data, with application to Canada Google Flu trends. *Annals of Applied Statistics*, *8*(1), 74–88. http://dx.doi.org/10.1214/13-AOAS684.

Hsu, W.-r., & Murphy, A. H. (1986). The attributes diagram: A geometrical framework for assessing the quality of probability forecasts. *International Journal of Forecasting*, *2*(3), 285–293. http://dx.doi.org/10.1016/0169-2070(86)90048-8.

Lu, J., & Meyer, S. (2020). Forecasting flu activity in the United States: Benchmarking an endemic-epidemic beta model. *International Journal of Environmental Research and Public Health*, *17*(4), 1381. http://dx.doi.org/10.3390/ijerph17041381.

Marzban, C. (2004). The ROC curve and the area under it as performance measures. *Weather and Forecasting*, *19*(2), 1106–1114. http://dx.doi.org/10.1175/825.1.

Melchior, C., Zanini, R. R., Guerra, R. R., & Rockenbach, D. A. (2021). Forecasting Brazilian mortality rates due to occupational accidents using autoregressive moving average approaches. *International Journal of Forecasting*, *37*(2), 825–837. http://dx.doi.org/10.1016/j.ijforecast.2020.09.010.

Mills, T. C. (2010). Forecasting compositional time series. *Quality & Quantity*, *44*, 673–690. http://dx.doi.org/10.1007/s11135-009-9229-8.

Nocedal, J., & Wright, S. J. (2006). *Numerical optimization* (2nd ed.). New York: Springer.

Poler, R., & Mula, J. (2011). Forecasting model selection through out-of-sample rolling horizon weighted error. *Expert Systems with Applications*, *38*(12), 14778–14785. http://dx.doi.org/10.1016/j.eswa.2011.05.072.

Pumi, G., Prass, T. S., & Souza, R. R. (2021). A dynamic model for double bounded time series with chaotic-driven conditional averages. *Scandinavian Journal of Statistics*, *41*(1), 68–86. http://dx.doi.org/10.1111/sjos.12439.

R Core Team (2023). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, URL https://www.R-project.org/.

Rocha, A. V., & Cribari-Neto, F. (2009). Beta autoregressive moving average models. *TEST*, *18*(3), 529–545. http://dx.doi.org/10.1007/s11749-008-0112-z.

Rocha, A. V., & Cribari-Neto, F. (2017). Erratum to: Beta autoregressive moving average models. *TEST*, *26*(2), 451–459. http://dx.doi.org/10.1007/s11749-017-0528-4.

Sales, L. O. F., Alencar, A. P., & Ho, L. L. (2022). The BerG generalized autoregressive moving average model for count time series. *Computers & Industrial Engineering*, *168*, Article 108104. http://dx.doi.org/10.1016/j.cie.2022.108104.

Scher, V. T., Cribari-Neto, F., Pumi, G., & Bayer, F. M. (2020). Goodness-of-fit tests for βARMA hydrological time series modeling. *Environmetrics*, *31*(3), Article e2607. http://dx.doi.org/10.1002/env.2607.

Snyder, R. D., Ord, J. K., Koehler, A. B., McLaren, K. R., & Beaumont, A. N. (2017). Forecasting compositional time series: A state space approach. *International Journal of Forecasting*, *33*(2), 502–512. http://dx.doi.org/10.1016/j.ijforecast.2016.11.008.

Taillardat, M., Mestre, O., Zamo, M., & Naveau, P. (2016). Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. *Monthly Weather Review*, *144*(6), 2375–2393. http://dx.doi.org/10.1175/MWR-D-15-0260.1.

Zheng, T., & Chen, R. (2017). Dirichlet ARMA models for compositional time series. *Journal of Multivariate Analysis*, *158*, 31–46. http://dx.doi.org/10.1016/j.jmva.2017.03.006.