Computational and Applied Mathematics (2024) 43:27
https://doi.org/10.1007/s40314-023-02513-5

Forecasting the proportion of stored energy using the unit Burr XII quantile autoregressive moving average model

Tatiane Fontana Ribeiro[1,3] · Fernando A. Peña-Ramírez[2] · Renata Rojas Guerra[3] · Airlane P. Alencar[1] · Gauss M. Cordeiro[4]

Received: 3 March 2023 / Revised: 10 October 2023 / Accepted: 2 November 2023

Abstract
This paper defines the unit Burr XII autoregressive moving average (UBXII-ARMA) model for continuous random variables in the unit interval, where any quantile can be modeled by a dynamic structure including autoregressive and moving average terms, time-varying regressors, and a link function. Our main motivation is to analyze the time series of the proportion of stored hydroelectric energy in Southeast Brazil and even identify a crisis period with lower water levels. We consider the conditional maximum likelihood method for parameter estimation, obtain closed-form expressions for the conditional score function, and conduct simulation studies to evaluate the accuracy of the estimators and estimated coverage rates of the parameters' asymptotic confidence intervals. We discuss the goodness-of-fit assessment and forecasting for the new model. Our forecasts of the proportion of the stored energy outperformed those obtained from the Kumaraswamy autoregressive moving average and beta autoregressive moving average models. Furthermore, only the UBXII-ARMA detected a significant effect of lower water levels before 2002 and after 2013.

Keywords Burr XII distribution · Unit interval · βARMA · KARMA · Forecasting · Quantile regression model

Mathematics Subject Classification 60E05 · 62F99 · 62J99

Communicated by Kelly Cristina Poldi.

✉ Tatiane Fontana Ribeiro
tatianefr@ime.usp.br

1 Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, SP, Brazil

2 Departamento de Estadística, Universidad Nacional de Colombia, Bogotá, Colombia

3 Departamento de Estatística, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil

4 Departamento de Estatística, Universidade Federal de Pernambuco, Recife, PE, Brazil

Published online: 23 December 2023

1 Introduction

Hydro-environmental resources are fundamental for human activities such as energy generation, agricultural irrigation, navigation, tourism, and watershed management. They play an important role in the functioning of society and in controlling global-scale hydrological processes (Lehner et al. 2022; Sagrillo et al. 2021). Hence, when analyzing these data, it is necessary to consider models that properly accommodate their characteristics. In the case of variables serially dependent over time, we can refer to time series models based on Gaussian assumptions, as are those more widely employed in hydrology and related areas. Classic examples are the autoregressive moving average (ARMA) models and autoregressive integrated moving average (ARIMA) models (Box et al. 2011). On the other hand, real-world often data do not present Gaussian distributions, making this assumption quite restrictive for various applications (Bayer et al. 2017; Palm et al. 2021).

Several models have been developed to overcome the issue of analyzing non-Gaussian time series. Benjamin et al. (2003) extended the Gaussian ARMA time series models to a non-Gaussian framework by developing dynamic models for random variables in the canonical form of the exponential family, giving rise to the generalized autoregressive moving average (GARMA) models. Recently, several time series models following a construction similar to the GARMA model have been introduced. Melo and Alencar (2020) proposed the Conway–Maxwell–Poisson autoregressive moving average model, which can be used for modeling time series of counts with equidispersion, underdispersion, and overdispersion. Palm et al. (2021) introduced the beta-binomial autoregressive moving average model for modeling quantized amplitude data and bounded count data. Other recent advances were developed by Bayer et al. (2020a, b), Almeida-Junior and Nascimento (2021), and Palm et al. (2022) in the context of positive outcomes.

Alternative models for double-bounded time series are also needed since they avoid data transformation before modeling. Moreover, models with support in $\mathbb{R}$ or $\mathbb{R}^+$ are not suitable for analyzing time series of this type since these models can produce forecasting that extends beyond the natural bounds of the data. We refer to the beta autoregressive moving average (βARMA) (Rocha and Cribari-Neto 2009) and the Kumaraswamy autoregressive moving average (KARMA) (Bayer et al. 2017) as pioneering models to time series taking values in the double-bounded interval (a, b), such that $a < b$. These models have been widely applied to hydrological problems. See Palm and Bayer (2017), Scher et al. (2020), and Bayer et al. (2017), for instance. Recently, extensions were proposed for the βARMA model. Bayer et al. (2023) introduced the inflated βARMA model, tailored for use with time series data that assume values in the inflated intervals of zero or one. Additionally, Scher et al. (2023) generalized the βARMA model, including a dynamic submodel for the precision parameter. This approach allows the conditional precision to evolve over time, similar to the conditional mean.

The βARMA and KARMA models employ parameterizations based on the conditional mean and median, respectively. To the best of our knowledge, ARMA models for analyzing other conditional quantiles, besides the median, of double-bounded responses are absent in the literature. Thus, we attempt to fulfill this gap by proposing a dynamic model based on the unit Burr XII (UBXII) quantile regression (Korkmaz and Chesneau 2021; Ribeiro et al. 2022). The so-called UBXII autoregressive moving average model (UBXII-ARMA) is obtained by adding ARMA terms to the systematic component of the UBXII regression.

Quantile regression approaches to model double-bounded response variables to the unit interval have been receiving attention in the literature. Several new unit distributions and

their quantile regression-associated models have been defined and applied in many areas. These new distributions are based on different transformations of positive random variables, such as inverted exponential function and hyperbolic secant transformation; see, for example, Korkmaz and Korkmaz (2021), Korkmaz et al. (2021a, 2022a, 2023) and Mazucheli et al. (2023). Transformations of random variables with support in the unit interval and real line have also been applied to generate new unit distributions and quantile regression models. (Korkmaz et al. 2021b) applied the quadratic transmutation scheme on the unit Rayleigh random variable and defined the transmuted unit Rayleigh quantile regression model. A transformation of a random variable with a normal distribution involving the hyperbolic secant function and its inverse was considered by Korkmaz et al. (2021c) to introduce the arcsecant hyperbolic normal distribution and quantile regression. Similarly, Korkmaz et al. (2022b) proposed the unit folded normal distribution and quantile regression by applying the absolute hyperbolic tangent transformation to a normal random variable. Furthermore, two regression models based on the Vasicek distribution, suitable for modeling the mean or quantiles of the bounded responses in the unit interval, were presented by Mazucheli et al. (2022).

The primary motivation of our proposal is that it allows estimating the conditional quantiles of the distribution of the response variable of interest with support on the (0, 1), providing a complete view of possible relationships between predictor variables (Cade and Noon 2003; Lima et al. 2022) and past values of the response variable, rather than analyzing only central tendency effects. The quantile-based approach also allows for verifying the impact of the quantiles on the model's parameter values (Mazucheli et al. 2020). Our methodology offers the advantage of assessing differences in the magnitude of extreme values instead of solely analyzing mean or median impact estimations, as is common in other statistical approaches. Specifically, we evaluate the tail values representing flood events (upper percentiles, when the proportion of stored hydroelectric energy is very high). Another advantage is that quantile estimates exhibit lower sensitivity to outliers compared to conditional mean estimations. For instance, parametrizations in terms of the median can be more useful in the presence of atypical observations in the conditional response or when the data present quite asymmetric distribution (Bayer et al. 2017). Other recent studies that consider a similar approach using quantile modeling can be found in Lima et al. (2022) and Mohsenipour et al. (2020).

A second motivation is the ability to accommodate hydro-environmental data that assume values in the double-bounded interval. This usefulness is illustrated by the proportion of stored hydroelectric energy in Southeast Brazil. Forecasting future levels of this variable can contribute to the management of hydropower generation, such as projections of the marginal cost and make-decision of companies in the electricity sector (http://www.ons.org.br/). This information is increasingly important due to climate change, which has added uncertainty to hydropower generation (Cribari-Neto et al. 2021). In this respect, time series of the stored hydroelectric energy of other Brazilian regions have been studied by Palm and Bayer (2017), Scher et al. (2020, 2023), and Cribari-Neto et al. (2021), using the βARMA and its extensions. However, the UBXII density has shapes not assumed by the beta and Kumaraswamy distributions, such as the reverse tilde shaped (Ribeiro et al. 2022). Thus, the use of the βARMA and KARMA can be compromised for modeling hydro-environmental data with these characteristics.

The remainder of this paper is outlined as follows. Section 2 introduces the new time series model for conditional variables restricted to the interval (0, 1). Section 3 discusses conditional maximum likelihood estimation for the UBXII-ARMA models and provides closed-form expressions for the conditional score vector. Furthermore, we present asymptotic confidence intervals for the model's parameters based on the conditional maximum likeli-

hood estimator (CMLE) properties. Some diagnostic measures and forecasting methods are presented in Sect. 5. Monte Carlo simulation studies to assess the finite-sample performance of the CMLEs are conducted in Sect. 6, where we evaluate the point estimates and the estimated coverage probability from the asymptotic confidence intervals. Still, in this section, we assess the performance of the proposed model's CMLEs under the misspecification of the random component. In Sect. 7, the analysis of the proportion of stored hydroelectric energy is carried out to provide empirical evidence of the proposed model's potentiality. Finally, some conclusions are discussed in Sect. 8.

2 The proposed quantile UBXII-ARMA model

In this section, we introduce a dynamic time series model based on the UBXII distribution. This distribution is obtained by an exponential transformation of a Burr XII (BXII) random variable and can be reparameterized in two ways; see Korkmaz and Chesneau (2021) and Ribeiro et al. (2022). In particular, the BXII distribution has been receiving attention in regression modeling (Silva et al. 2008; de Araújo et al. 2022), and generalized distributions (Qin and Gui 2020; Bhatti et al. 2021; Guerra et al. 2021), among other related areas. Our proposal is to incorporate the serial correlation into the UBXII regression model, introducing autoregressive and moving average terms in the linear predictor. We shall consider a similar approach to that employed in the construction of the GARMA (Benjamin et al. 2003), βARMA (Rocha and Cribari-Neto 2009), and KARMA (Bayer et al. 2017) models.

Let $\{y_t\}_{t\in\mathbb{Z}}$ be the discrete-time observations, where each y_t assumes values in the unit interval and $\mathcal{F}_{t-1}$ be the set of observations up to time $t-1$. Notice that, if $\tilde{y} \in (a, b)$, we can apply the transformation $y_t = (\tilde{y} - a)/(b - a)$ to bring this double-bounded variable to the unit interval. Let us assume that each y_t follows a UBXII distribution conditionally to $\mathcal{F}_{t-1}$ with conditional quantile q_t and shape parameter c, say $y_t|\mathcal{F}_{t-1} \sim \text{UBXII}(q_t, c)$. The conditional probability density function of y_t, with the quantile-based parametrization proposed by Ribeiro et al. (2022), is

$$f(y_t|\mathcal{F}_{t-1}) = \frac{\log\tau^{-c}\log^{c-1}y_t^{-1}}{y_t\log\left(1+\log^c q_t^{-1}\right)}\left(1+\log^c y_t^{-1}\right)^{\log\tau/\log\left(1+\log^c q_t^{-1}\right)-1}, \tag{1}$$

where $0 < y_t < 1$, $0 < q_t < 1$ is the conditional quantile, $c > 0$ is a shape parameter, and $0 < \tau < 1$ indicates the order $(\tau \times 100)$th of the conditional quantile q_t of Y_t. For example, if $\tau = 0.5$, q_t is the median of Y_t; or if $\tau = 0.1$, q_t is the 10th conditional percentile of Y_t. Moreover, the conditional cumulative distribution function (cdf) and conditional quantile function (cqf) are

$$F(y_t|\mathcal{F}_{t-1}) = \left(1+\log^c y_t^{-1}\right)^{\log\tau/\log\left(1+\log^c q_t^{-1}\right)}, \quad 0 < y_t < 1, \tag{2}$$

and

$$Q(u|\mathcal{F}_{t-1}) = \exp\left\{-[u^{\log(1+\log^c q_t^{-1})/\log\tau} - 1]^{1/c}\right\}, \quad 0 < \tau < 1, \tag{3}$$

respectively. Values of the UBXII distribution may be easily simulated by the inversion method since the cqf has a simple closed-form expression.

The UBXII distribution is quite versatile since its density can assume many different shapes according to the selected parameter values combination. We develop an open web application in `Shiny` (R Core Team 2023) available at https://unitati.shinyapps.io/UBXII/ to visualize dynamic graphics of the UBXII density. In this application, it is possible to note

Table 1 Sample mean of variances of $y \sim \text{UBXII}(q, c)$ with $q \in \{0.25, 0.50, 0.75\}$ and $c \in \{1.00, 5.00, 10.00, 15.00, 20.00\}$ for $\tau \in \{0.1, 0.5, 0.9\}$

τ	c				
	1.00	5.00	10.00	15.00	20.00
$q = 0.25$					
0.1	0.0707	0.0133	0.0056	0.0039	0.0033
0.5	0.1130	0.0236	0.0149	0.0131	0.0125
0.9	0.0375	0.0140	0.0117	0.0111	0.0109
$q = 0.50$					
0.1	0.0382	0.0057	0.0017	0.0008	0.0005
0.5	0.1096	0.0078	0.0019	0.0008	0.0005
0.9	0.0562	0.0183	0.0024	0.0009	0.0005
$q = 0.75$					
0.1	0.0111	0.0017	0.0006	0.0003	0.0002
0.5	0.0683	0.0025	0.0007	0.0003	0.0002
0.9	0.0905	0.0041	0.0009	0.0004	0.0002

that c is related to the variance of the distribution UBXII. To analyze this relationship, we conducted a Monte Carlo simulation study. We generate 10,000 samples of $y \sim \text{UBXII}(q, c)$ of size $n = 200$ with $q \in \{0.25, 0.50, 0.75\}$ and $c \in \{1.00, 5.00, 10.00, 15.00, 20.00\}$ for $\tau \in \{0.1, 0.5, 0.9\}$. We present the mean of the variances in each scenario in Table 1. Notice that, for all considered scenarios, the variance decreases when c increases. These results provide numerical evidence that c is a precision parameter.

To define the dynamic component of the model, we propose the following specification for the conditional quantile q_t

$$\eta_t = g(q_t) = \alpha + \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^{p} \phi_i \left[g(y_{t-i}) - \mathbf{x}_{t-i}^\top \boldsymbol{\beta} \right] + \sum_{j=1}^{q} \theta_j \, r_{t-j}, \tag{4}$$

where η_t is the linear predictor, $g(\cdot)$ is a twice-differentiable one-to-one monotonic link function, $\alpha \in \mathbb{R}$ is the model intercept, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^\top \in \mathbb{R}^k$ is a k-dimensional vector of unknown parameters corresponding to the covariates, $\boldsymbol{x}_t = (x_{t1}, \ldots, x_{tk})^\top$ is a non-random covariates vector, being $k < n$, and ϕ_i $(i = 1, \ldots, p)$ and θ_j $(j = 1, \ldots, q)$ are the autoregressive (AR) and moving average (MA) parameters, respectively. Thus, they are the parameters of the ARMA structure with $p, q \in \mathbb{N}$, say ARMA(p,q). The term r_{t-j} corresponds to a random error. Several types of residuals can be used for this moving average error term (Benjamin et al. 2003). Generally, they are measured on the predictor scale as $r_t = g(y_t) - g(q_t)$; see Rocha and Cribari-Neto (2009), Bayer et al. (2017, 2020a), Almeida-Junior and Nascimento (2021), and Palm et al. (2021). It is only required that r_t must be $\mathcal{F}_{t-1}$-measurable and assumed that $\mathbb{E}(r_t|\mathcal{F}_{t-1} = 0)$; see Rocha and Cribari-Neto (2009) for more details. Here, we propose to use the residuals in the predictor scale when $\tau = 0.5$, i.e., for q_t being the conditional median. When $\tau \neq 0.5$, we propose to consider the quantile residuals (Dunn and Smyth 1996) for r_t. That is,

$$r_t = \Phi^{-1}\left[F(y_t|\mathcal{F}_{t-1})\right], \tag{5}$$

where $\Phi^{-1}(\cdot)$ is the standard normal quantile function and $F(y_t|\mathcal{F}_{t-1})$ is the conditional cumulative distribution function given in (2). For the link function $g(\cdot)$, there are various functions available. Some examples are the logit, probit, and complementary log–log

(cloglog) links. Note that the conditional quantile of y_t is a function of the past observations y_{t-i}, covariates $\boldsymbol{x}_t$, and MA error terms.

In this way, from (1) and (4), we define the so-called UBXII-ARMA(p, q) dynamic model. In a similar manner to classical ARMA models, we require non-common factors between the AR and MA characteristic polynomials. Further, the AR polynomial must have no unit characteristic root; see Brockwell and Davis (2009) for more details.

The dynamic component (see Eq. (4)) is similar to that one proposed by Rocha and Cribari-Neto (2009) and later by Bayer et al. (2017). However, there are two remarkable differences. First, the random component is entirely distinct from both proposals since the response variable here has a UBXII distribution. Second, in the class of UBXII-ARMA(p, q) models, it is possible to model any quantile of the response instead of the mean or only the median. Thus, it is a more general time series model and a new alternative that allows analyzing a range of double-bounded conditional responses on the interval $(0, 1)$.

3 Parameter estimation

The model-fitting procedure described herein is performed using the conditional maximum likelihood method. Let $\boldsymbol{\gamma} = (\alpha, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top, c)^\top$ be the $(p+q+k+2)$-dimensional parameter vector that index the UBXII-ARMA(p, q) model in a sample $(y_1, \boldsymbol{x_1}), \ldots, (y_n, \boldsymbol{x_n})$, satisfying the specification given in (1) and (4). As in Rocha and Cribari-Neto (2009), Bayer et al. (2017, 2020a), Almeida-Junior and Nascimento (2021), and Palm et al. (2021), we consider the conditional maximum likelihood method to estimate $\boldsymbol{\gamma}$. The conditional log-likelihood function can be expressed as

$$\ell(\boldsymbol{\gamma}) = \sum_{t=m+1}^{n} \log f(y_t \mathcal{F}_{t-1}) = \sum_{t=m+1}^{n} \ell_t(q_t, c), \tag{6}$$

where

$$\ell_t(q_t, c) = \log(\log \tau^{-c}) - \log y_t + (c-1)\log(\log y_t^{-1}) - \log[t(y_t)] - \log\{\log[t(q_t)]\} - \frac{\log \tau^{-1} \log[t(y_t)]}{\log[t(q_t)]},$$

$m = \max\{p, q\} < n$ and $t(x) = 1 + \log^c(x^{-1})$.

The conditional maximum likelihood estimators (CMLEs), $\hat{\boldsymbol{\gamma}}$, are obtained maximizing (6). We can obtain the score vector, set its components equal to zero, and solve the resulting non-linear equation system. In what follows, we compute the score vector by differentiating (6) concerning each component of the unknown parameter vector $\boldsymbol{\gamma}$.

3.1 Conditional score vector

The conditional score vector, denoted by $U(\boldsymbol{\gamma})$, is composed of the partial derivatives of ℓ with respect to each component of $\boldsymbol{\gamma}$. That is, $U(\boldsymbol{\gamma}) := \partial\ell/\partial\boldsymbol{\gamma} = [U_\alpha(\boldsymbol{\gamma}), U_{\boldsymbol{\beta}}(\boldsymbol{\gamma})^\top, U_{\boldsymbol{\phi}}(\boldsymbol{\gamma})^\top, U_{\boldsymbol{\theta}}(\boldsymbol{\gamma})^\top, U_c(\boldsymbol{\gamma})]^\top$. Let γ_v be the vth component of $\boldsymbol{\gamma}$. Then, the $(k + p + q + 1)$ first components of the conditional score vector are obtained using the

chain rule as

$$U_{\gamma_j}(\boldsymbol{\gamma}) := \frac{\partial \ell}{\partial \gamma_j} = \sum_{t=m+1}^{n} \frac{\partial \ell_t(q_t, c)}{\partial q_t} \frac{\mathrm{d}q_t}{\mathrm{d}\eta_t} \frac{\partial \eta_t}{\partial \gamma_j}. \tag{7}$$

Thus, defining the quantities

$$y_t^\star := \log[t(y_t)], \quad q_t^\star := \frac{c \log^{c-1} q_t^{-1}}{q_t\, t(q_t) \log[t(q_t)]}, \quad \text{and} \quad q_t^\dagger := \frac{\log \tau^{-c} \log^{c-1} q_t^{-1}}{q_t\, t(q_t) \log^2[t(q_t)]},$$

the two first derivatives in (7) reduce to

$$\frac{\partial \ell_t(q_t, c)}{\partial q_t} = q_t^\star - q_t^\dagger\, y_t^\star \quad \text{and} \quad \frac{\mathrm{d}q_t}{\mathrm{d}\eta_t} = \frac{1}{g'(q_t)}.$$

The partial derivatives $\partial \eta_t / \partial \gamma_v$ depend on the moving average error term (r_t). Here, we compute the derivatives for $\tau = 0.5$, when r_t is on the predictor scale. Thus, the derivative $\partial \eta_t / \partial \gamma_v$ is computed recursively as

$$\begin{aligned}
\frac{\partial \eta_t}{\partial \alpha} &= 1 - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \alpha}, \\
\frac{\partial \eta_t}{\partial \beta_l} &= x_{tl} - \sum_{i=1}^{p} \phi_i x_{(t-i)l} - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \beta_l}, \quad \text{for } l = 1, \ldots, k, \\
\frac{\partial \eta_t}{\partial \phi_i} &= g(y_{t-i}) - \boldsymbol{x}_{t-i}^\top \boldsymbol{\beta} - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \phi_i}, \quad \text{for } i = 1, \ldots, p.
\end{aligned}$$

and

$$\frac{\partial \eta_t}{\partial \theta_j} = r_{t-j} - \sum_{w=1}^{q} \theta_w \frac{\partial \eta_{t-w}}{\partial \theta_j}, \quad \text{for } j = 1, \ldots, q.$$

The last component of the conditional score vector, $U_c(\gamma)$, follows from direct differentiation of (6)

$$\frac{\partial \ell}{\partial c} = \sum_{t=m+1}^{n} \frac{\partial \ell_t(q_t, c)}{\partial c} = \sum_{t=m+1}^{n} y_t^\sharp,$$

where

$$\begin{aligned}
y_t^\sharp = \frac{1}{c} &+ \log(\log y_t^{-1}) - \frac{\log(\log q_t^{-1})[t(q_i) - 1]}{t(q_t) \log[t(q_t)]} - \frac{[t(y_t) - 1] \log(\log y_t^{-1})}{t(y_t)} \\
&- \frac{\log \tau^{-1} \log[t(q_t)][t(y_t)]^{-1}[t(y_t) - 1] \log(\log y_t^{-1})}{\log^2[t(q_t)]} \\
&+ \frac{\log \tau^{-1}[t(q_t) - 1] \log(\log q_t^{-1}) \log[t(y_t)]}{t(q_t) \log^2[t(q_t)]}.
\end{aligned}$$

Let $\boldsymbol{M}$, $\boldsymbol{P}$, $\boldsymbol{R}$ be matrices with dimensions $(n-m)\times k$, $(n-m)\times p$ and $(n-m)\times q$, respectively. The (i,j)-th element of those matrices are given by

$$\boldsymbol{M}_{i,j}=\frac{\partial\eta_{i+m}}{\partial\beta_j},\quad P_{i,j}=\frac{\partial\eta_{i+m}}{\partial\phi_j},\quad \text{and}\quad R_{i,j}=\frac{\partial\eta_{i+m}}{\partial\theta_j},$$

respectively. Then, we can compactly write the score vector's components of $\boldsymbol{\gamma}$ as

$$\begin{aligned}
U_{\alpha}(\boldsymbol{\gamma}) &= \boldsymbol{a}^{\top}\boldsymbol{T}\,(\boldsymbol{q}^{\star}-\boldsymbol{q}^{\dagger}\,\boldsymbol{y}^{\star})\\
U_{\boldsymbol{\beta}}(\boldsymbol{\gamma}) &= \boldsymbol{M}^{\top}\boldsymbol{T}\,(\boldsymbol{q}^{\star}-\boldsymbol{q}^{\dagger}\,\boldsymbol{y}^{\star})\\
U_{\boldsymbol{\phi}}(\boldsymbol{\gamma}) &= \boldsymbol{P}^{\top}\boldsymbol{T}\,(\boldsymbol{q}^{\star}-\boldsymbol{q}^{\dagger}\,\boldsymbol{y}^{\star})\\
U_{\boldsymbol{\theta}}(\boldsymbol{\gamma}) &= \boldsymbol{R}^{\top}\boldsymbol{T}\,(\boldsymbol{q}^{\star}-\boldsymbol{q}^{\dagger}\,\boldsymbol{y}^{\star})\\
U_{c}(\boldsymbol{\gamma}) &= \boldsymbol{y}^{\sharp\top}\mathbf{1},
\end{aligned}$$

where $\boldsymbol{a}=(\partial\eta_{m+1}/\partial\alpha,\ldots,\partial\eta_n/\partial\alpha)^{\top}$, $\boldsymbol{T}$ is a diagonal matrix defined as $\boldsymbol{T}=\operatorname{diag}\{1/g'(q_{m+1}),\ldots,1/g'(q_n)\}$, $\boldsymbol{q}^{\star}=(q^{\star}_{m+1},\ldots,q^{\star}_n)^{\top}$, $\boldsymbol{q}^{\ddagger}=(q\dagger_{m+1},\ldots,q^{\dagger}_n)^{\top}$, $\boldsymbol{y}^{\star}=(y^{\star}_{m+1},\ldots,y^{\star}_n)^{\top}$, $\boldsymbol{y}^{\sharp}=(y^{\sharp}_{m+1},\ldots,y^{\sharp}_n)^{\top}$, and $\mathbf{1}$ is an $(n-m)$-dimensional vector of ones.

By setting each $U(\boldsymbol{\gamma})$ component equal to zero, i.e., $U_{\alpha}(\boldsymbol{\gamma})=0$, $U_{\boldsymbol{\beta}}(\boldsymbol{\gamma})=\mathbf{0}$, $U_{\boldsymbol{\phi}}(\boldsymbol{\gamma})=\mathbf{0}$, $U_{\boldsymbol{\theta}}(\boldsymbol{\gamma})=\mathbf{0}$, $U_c(\boldsymbol{\gamma})=0$, and solving these equations simultaneously, the CMLE $\hat{\boldsymbol{\gamma}}=(\hat{\alpha},\hat{\boldsymbol{\beta}}^{\top},\hat{\boldsymbol{\phi}}^{\top},\hat{\boldsymbol{\theta}}^{\top},\hat{c})^{\top}$ of $\boldsymbol{\gamma}$ is obtained. However, since this system is non-linear and cannot be solved explicitly, we may maximize the log-likelihood function in (6) through non-linear optimization methods such as Newton–Raphson or quasi-Newton-type algorithms. We consider the quasi-Newton algorithm the so-called Broyden– Fletcher–Goldfarb–Shanno (BFGS) method (Press et al. 1992), including the conditional score function $U(\boldsymbol{\gamma})$. This method is an iterative optimization algorithm, and thus, it requires initialization. We compute the starting values for α, $\boldsymbol{\beta}$, and $\boldsymbol{\phi}$ from an ordinary least squares estimate by considering a linear regression, where the response is $\boldsymbol{Y}=(g(y_{m+1}),\ldots,g(y_n))^{\top}$, and the covariates matrix is expressed as

$$\boldsymbol{X}=\begin{bmatrix}
1 & x_{(m+1)1} & \cdots & x_{(m+1)r} & g(y_m) & g(y_{m-1}) & \cdots & g(y_{m-p+1})\\
1 & x_{(m+2)1} & \cdots & x_{(m+2)r} & g(y_{m+1}) & g(y_m) & \cdots & g(y_{m-p+2})\\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\
1 & x_{n1} & \cdots & x_{nr} & g(y_{n-1}) & g(y_{n-2}) & \cdots & g(y_{n-p})
\end{bmatrix}.$$

For the moving average parameters $\boldsymbol{\theta}$, the starting values are set to zero.

4 Large sample inference

From the asymptotic properties of likelihood estimators, under the usual regularity conditions (Sen et al. 2009), we have that

$$\hat{\boldsymbol{\gamma}}\sim\mathcal{N}_{(k+p+q+2)}\left(\boldsymbol{\gamma},\boldsymbol{K}^{-1}(\boldsymbol{\gamma})\right),$$

where $\boldsymbol{K}^{-1}(\boldsymbol{\gamma})$ is the inverse of the expected information matrix defined as $\boldsymbol{K}(\boldsymbol{\gamma})=\mathbb{E}[-\partial\ell^2(\boldsymbol{\gamma})/(\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^{\top})]$ and $\mathcal{N}_r$ denotes a r-dimensional normal distribution. This means that the CMLE of $\boldsymbol{\gamma}$, $\hat{\boldsymbol{\gamma}}$, is asymptotically unbiased and normally distributed with covariance matrix equal to the inverse of the Fisher's information matrix. The matrix $\boldsymbol{K}(\boldsymbol{\gamma})$ can be consistently estimated using the observed information matrix evaluated at the CMLE $\hat{\boldsymbol{\gamma}}$, denoted by $\boldsymbol{J}(\hat{\boldsymbol{\gamma}})$ (Lindsay and Li 1997).

From the asymptotic normality of $\hat{\boldsymbol{\gamma}}$, it is possible to construct a $100(1-\xi)\%$ approximate confidence interval with $\xi \in (0, 1/2)$ for each element of $\boldsymbol{\gamma}$, i.e., for γ_i $(i = 1, \ldots, p+q+k+2)$ as follows

$$\left[\hat{\gamma}_i - z_{1-\xi/2}\sqrt{J(\hat{\boldsymbol{\gamma}})^{ii}}; \hat{\gamma}_i + z_{1-\xi/2}\sqrt{J(\hat{\boldsymbol{\gamma}})^{ii}}\right], \tag{8}$$

where $z_{1-\xi/2}$ is the standard normal quantile, such that $\Phi(z) = 1-\xi/2$, and $J(\hat{\boldsymbol{\gamma}})^{ii}$ is the (i,i)th element of the $\boldsymbol{J}^{-1}(\hat{\boldsymbol{\gamma}})$.

Analogously, based on the asymptotic distribution of the CMLE, we can construct asymptotic test statistics for testing the null hypothesis $\mathcal{H}_0 : \gamma_i = \gamma_i^0$ against $\mathcal{H}_1 : \gamma_i \neq \gamma_i^0$. Based on the Wald test (Wald 1943), the test statistics may be defined as (Pawitan 2001):

$$Z = \frac{\hat{\gamma}_i - \gamma_i^0}{\sqrt{J(\hat{\boldsymbol{\gamma}})^{ii}}}.$$

Under $\mathcal{H}_0$, the Z statistic has approximately a standard normal distribution. Thus, at the $\xi\%$-significance level, we reject the null hypothesis, whether the absolute observed value of Z, denoted by $|z|$, exceeds the quantile $z_{1-\xi/2}$.

5 Diagnostic analysis and forecasting

After fitting a model, it is important to perform adequacy tests to check whether it fully captures the data dynamics. Since a fitted time series model passes all diagnostic checks, we may use it for out-of-sample forecasting. In what follows, we introduce and discuss the residual analysis that is useful to identify whether the assumptions of the model are satisfied. Besides, we present criteria that can be used for model selection and discuss how to obtain out in-sample and out-of-sample forecasts from UBXII-ARMA(p, q) model.

5.1 Residual analysis and model selection

We consider the quantile residuals (Dunn and Smyth 1996) since they have several advantages over other residuals (Pereira 2019). For the UBXII-ARMA model, the quantile residuals are defined in (5), being $F(y_t|\mathcal{F}_{t-1})$ (given in (2)) evaluated at the CMLEs, specifically at the predicted quantiles $\hat{q}_t$ that corresponds to fitted values and at the estimate $\hat{c}$.

The quantile residuals are supposed to be roughly non-correlated and normally distributed with zero mean and unit variance when the model is suitable for the data. Furthermore, their sample autocorrelations and partial autocorrelations functions must be non-significant.

To choose the most suitable model among several competitive models, the Akaike information criterion (AIC) (Akaike 1973) or, alternatively, the Bayesian information criterion (BIC) (Akaike 1978) can be considered for models selection. Both criteria are based on the conditional maximum log-likelihood function, $\hat{\ell}$, namely

$$\text{AIC}(\nu) = 2\,(\nu - \hat{\ell}) \quad \text{and} \quad \text{BIC}(\nu) = \nu \log n - 2\hat{\ell},$$

where ν is the number of estimated parameters and n is the sample size. Among the candidate models set, we selected the model that provides the smallest AIC and BIC values. For detailed properties of these criteria, the reader is referred to Choi (2012).

5.2 Forecasting

Suppose that we are interested in forecasting the quantile at time s, q_s, using all the available information at time n ($s > n$). Then, the forecast horizon is $h_0 = s - n$. Initially, we compute the predicted quantiles $\{\hat{q}_t\}_{t=m+1}^{n}$ based on the CMLE $\hat{\boldsymbol{\gamma}}$, starting at $t = m + 1$, sequentially calculating

$$\hat{q}_t = g^{-1}\left(\hat{\alpha} + \mathbf{x}_t^\top \hat{\boldsymbol{\beta}} + \sum_{i=1}^{p} \hat{\phi}_i \left[\, g(y_{t-i}) - \mathbf{x}_{t-i}^\top \hat{\boldsymbol{\beta}} \,\right] + \sum_{j=1}^{q} \hat{\theta}_j \, \hat{r}_{t-j}\right),$$

where

$$\hat{r}_t = \begin{cases} 0 & \text{if } t \leq m. \\ g(y_t) - g(\hat{q}_t) & \text{if } m < t \leq n \text{ and } \tau = 0.5 \\ \Phi^{-1}\left[F(y_t|\mathcal{F}_{t-1})\right], & \text{if } m < t \leq n \text{ and } \tau \neq 0.5. \end{cases}$$

Then, for $t = n + 1, \ldots, s$, we need to assume that the observations from the covariates $\boldsymbol{x_t}$ are known. Hence, the forecast of the conditional quantiles , for $h = 1, \ldots, h_0$, are obtained sequentially from

$$\hat{q}_{n+h} = g^{-1}\left(\hat{\alpha} + \mathbf{x}_{n+h}^\top \hat{\boldsymbol{\beta}} + \sum_{i=1}^{p} \hat{\phi}_i \left[\, g(y_{n+h-i}) - \mathbf{x}_{n+h-i}^\top \hat{\boldsymbol{\beta}} \,\right] + \sum_{j=1}^{q} \hat{\theta}_j \, \hat{r}_{n+h-j}\right),$$

where for $t > n$, $\hat{r}_t = 0$ and

$$g(y_t) = \begin{cases} g(\hat{q}_t) & \text{if } t > n, \\ g(y_t) & \text{if } t \leq n. \end{cases}$$

To empirically evaluate the forecasting performance of the UBXII-ARMA model and compare it to other fitted models, we consider two measures of forecast accuracy: the mean square error (MSE) and the mean absolute percentage error (MAPE). These measures allow for assessing the difference between the observed and the predicted value. The MSE is largely used due to its theoretical relevance in statistical modeling. However, when the observations assume real positive values, the use of the MAPE is indicated; see Hyndman and Koehler (2006) for a detailed discussion about available measures of univariate time series forecast accuracy. The MSE and MAPE are, respectively, defined by

$$\text{MSE} = \frac{1}{h_0}\sum_{h=1}^{h_0}(y_h - \hat{q}_h)^2 \quad \text{and} \quad \text{MAPE} = \left(\frac{1}{h_0}\sum_{h=1}^{h_0}\frac{|y_h - \hat{q}_h|}{|y_h|}\right) \times 100,$$

where the y_h is each observed value and $\hat{q}_h$ is the corresponding predicted value for the forecast horizon $h = 1, \ldots, h_0$. Low values for MSE and MAPE indicate more accurate predictions.

6 Simulation study

We conduct a Monte Carlo simulation study to assess the finite sample performance of the CMLEs and asymptotic confidence intervals of the UBXII-ARMA(p, q) model's parameters. Samples of sizes $n \in \{75, 125, 200, 300, 500, 1000\}$ are considered in four distinct scenarios.

For each scenario and sample size, we compute 10,000 times the CMLEs and the confidence intervals of the model parameters.

Several dynamic specifications and different parameter values are considered. For all settings, $\tau = 0.5$ was chosen, implying that q_t is the conditional median of Y_t. We consider the logit for $g(\cdot)$ in (4). The simulation schemes are

- Scenario 1: UBXII-ARMA(2, 2) with parametric values $\alpha = 0.5$, $\phi_1 = 0.6$, $\phi_2 = -0.4$, $\theta_1 = 0.4$, $\theta_2 = 0.1$, and $c = 5.6$.
- Scenario 2: UBXII-ARMA(1, 1) with parametric values $\alpha = 0.2$, $\phi_1 = 0.6$, $\theta_1 = 0.1$, and $c = 3.8$.
- Scenario 3: UBXII-ARMA(2, 1) with parametric values $\alpha = 0.4$, $\phi_1 = 0.6$, $\phi_2 = -0.4$, $\theta_1 = 0.3$, and $c = 4.5$.
- Scenario 4: UBXII-ARMA(1, 2) with parametric values $\alpha = 0.7$, $\phi_1 = -0.7$, $\theta_1 = 0.4$, $\theta_2 = 0.6$, and $c = 3.5$.

For Scenario 2, we also consider the probit and cloglog link functions. The results are given in Appendix A (see Tables 10 and 11) and are very similar to those with logit link function. For the simulation of samples from a UBXII-ARMA(p, q) process, we use the same algorithm employed by Bayer et al. (2017). The random component was generated through the inversion method, replacing $u \sim U(0, 1)$ in (3). Besides, we assume the dynamic structure given in (4). All Monte Carlo simulations are performed using the `R` programming language (R Core Team 2023). Maximization of the conditional log-likelihood function in (6) is carried out using the BFGS quasi-Newton non-linear optimization algorithm implemented at the `optim` function, including the conditional score function $U(\boldsymbol{\gamma})$.

To numerically evaluate the behavior of the CMLEs, we compute the absolute values of the percentage relative bias (RB%) and mean squared error (MSE). Monte Carlo results for the different structures are reported in Tables 2 and 3.

The largest RB%s are around 14, 16, and 30 for the moving average parameters' CMLEs (θ_1, θ_2) in Scenarios 1 and 2, respectively, when the sample size is $n \in \{75, 125, 200\}$. All the remaining RB%s are lower than 11. Note that the RB%s and MSEs are small in most scenarios only for $n \geq 200$. The RB% and MSE values of Scenario 2 are close to the Monte Carlo simulation results for point estimation in the βARMA(1, 1) model carried out by Palm and Bayer (2017) (see Table 3, no corrected estimates). Roughly, it is noteworthy that the UBXII-ARMA(1, 1) model yields smaller MSE values by comparing the sample size $n \in \{75, 125\}$ with $n \in \{50, 100\}$.

Overall, Bayer et al. (2017) also find that the estimates of the other parameters tend to present better performance compared to the parameter estimates of the moving average terms. The MSEs decrease for all scenarios from $n > 200$, implying that the performance and accuracy of the CMLEs improve when the sample size increases. Therefore, the numerical evaluation indicates that the properties of the CMLEs (asymptotically unbiased and consistent) remained.

In Scenarios 1, 3, and 4, the RB% exhibits minor oscillations for sample sizes $n < 200$. Although these oscillations are particularly noticeable within the context of time series models, they also can occur in regression analysis or probabilities distribution (see Figure 4 in Korkmaz et al. 2021a, which shows oscillations of the bias and MSE). This behavior may vary depending on the specific scenario and chosen sample sizes. For example, comparable findings were obtained for the βARMA model (Palm and Bayer 2017), where it is possible to observe that the total relative bias oscillates in some instances. Similar behavior to RB% is also kept for the mean squared error (MSE) when $n < 200$ in Scenario 4, as MSE depends on the bias. However, as the sample size n increases beyond 200, the RB% and MSE converge

Table 2 Simulation results on point estimation of the UBXII-ARMA(p, q) model under ARMA structures from Scenario 1 and 2

n	Measure	$\hat{\alpha}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{c}$
			Scenario 1				
75	RB%	0.2339	3.3881	3.7045	6.0854	5.8984	5.3870
	MSE	0.0299	0.1264	0.0279	0.1465	0.0779	0.4080
125	RB%	0.9537	2.4038	1.7548	3.9443	7.2060	3.1812
	MSE	0.0158	0.0719	0.0153	0.0802	0.0422	0.2021
200	RB%	1.7290	3.7209	2.1092	5.9806	14.3574	1.9005
	MSE	0.0097	0.0400	0.0077	0.0447	0.0244	0.1098
300	RB%	1.1835	2.5614	1.4466	4.1709	10.6165	1.2470
	MSE	0.0060	0.0247	0.0049	0.0273	0.0149	0.0682
500	RB%	0.5497	1.2143	0.7469	1.9129	4.4785	0.7564
	MSE	0.0032	0.0133	0.0029	0.0145	0.0078	0.0388
1000	RB%	0.2463	0.4814	0.2780	0.7722	1.7194	0.3847
	MSE	0.0015	0.0063	0.0014	0.0068	0.0037	0.0187
			Scenario 2				
75	RB%	10.2923	5.8411	–	29.2570	–	3.2664
	MSE	0.0104	0.0172	–	0.0219	–	0.1377
125	RB%	5.2377	3.0889	–	15.6317	–	1.9586
	MSE	0.0054	0.0087	–	0.0119	–	0.0749
200	RB%	3.4119	2.0115	–	9.7231	–	1.1877
	MSE	0.0031	0.0051	–	0.0071	–	0.0431
300	RB%	2.4575	1.4414	–	6.7811	–	0.8037
	MSE	0.0020	0.0033	–	0.0045	–	0.0278
500	RB%	1.5339	0.8680	–	4.1243	–	0.4907
	MSE	0.0012	0.0019	–	0.0027	–	0.0162
1000	RB%	0.8564	0.4645	–	1.9345	–	0.2597
	MSE	0.0006	0.0009	–	0.0013	–	0.0080

to zero. This convergence is depicted in Fig. 1, which illustrates the total RB% (TRB) and total MSE (TMSE) of the CMLEs obtained from the UBXII-ARMA(p, q) model with $n \in \{75, 125, 200, 300, 1000\}$. The TRB and TMSE are computed from the sum of the RB% and MSE of the parameters' CMLEs for each sample size and scenario, respectively. Notably, both measures show an evident decay to zero for $n > 200$ in all scenarios.

Table 4 presents the estimated coverage probability from the asymptotic confidence intervals for the parameters $\alpha, \phi_1, \phi_2, \theta_1, \theta_2, c$. The limits of the confidence intervals are computed from (8) with a significance level (ξ) of 5%. Overall, the coverage probability of the 95% pointwise confidence intervals of all the parameters is close to the considered nominal level when $n > 200$. Scenario 1 has the lowest coverage probabilities, mainly for $n = 75$.

A similar simulation study is carried out to analyze the performance of the parameters' CMLEs of the UBXII-ARMA model when τ assumes extreme values, such as $\tau \in \{0.1, 0.9\}$. In these cases, q_t is the 10th and the 90th conditional percentile of Y_t, respectively. We select Scenario 2 with the link function logit to perform the analysis. However, we choose $c = 5.0$ for $\tau = 0.1$ and $c = 13.5$ for $\tau = 0.9$ to keep a similar variance of the simulated data when changing the value of τ. Tables 12 and 13 (see Appendix A) display the RB%, MSE, and

Table 3 Simulation results on point estimation of the UBXII-ARMA(p, q) model under ARMA structures from Scenario 3 and 4

n	Measure	$\hat{\alpha}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{c}$
			Scenario 3				
75	RB%	0.0138	3.3712	4.8744	8.2591	–	4.2183
	MSE	0.0071	0.0327	0.0173	0.0422	–	0.2193
125	RB%	0.1561	1.4443	2.1207	3.1087	–	2.4920
	MSE	0.0038	0.0168	0.0095	0.0194	–	0.1136
200	RB%	0.1379	0.8969	1.2951	2.0487	–	1.5053
	MSE	0.0023	0.0096	0.0056	0.0110	–	0.0640
300	RB%	0.0407	0.5242	0.8567	1.3184	–	0.9965
	MSE	0.0015	0.0060	0.0036	0.0068	–	0.0405
500	RB%	0.0732	0.3426	0.4869	0.7702	–	0.6109
	MSE	0.0009	0.0036	0.0022	0.0039	–	0.0234
1000	RB%	0.0462	0.1147	0.1634	0.3227	–	0.3162
	MSE	0.0005	0.0017	0.0010	0.0019	–	0.0113
			Scenario 4				
75	RB%	1.4759	0.9546	–	0.2882	0.6091	3.2104
	MSE	0.0240	0.0327	–	0.0220	0.0448	0.1405
125	RB%	0.7866	1.1885	–	0.7399	0.6516	1.8932
	MSE	0.0092	0.0047	–	0.0096	0.0122	0.0750
200	RB%	0.5730	0.7696	–	0.8358	0.9922	1.1757
	MSE	0.0432	0.0213	–	0.0058	0.0074	0.0499
300	RB%	0.0628	0.6146	–	0.4544	0.6662	0.8681
	MSE	0.0034	0.0132	–	0.0023	0.0033	0.0330
500	RB%	0.1184	0.9338	–	0.4968	0.6420	0.5608
	MSE	0.0017	0.0009	–	0.0009	0.0016	0.0180
1000	RB%	0.1984	0.8428	–	0.5452	0.5517	0.3950
	MSE	0.0008	0.0005	–	0.0005	0.0012	0.0104

estimated coverage probability from the asymptotic confidence intervals for the parameters $\alpha, \phi_1, \theta_1, c$ and sample sizes considered. The results are quite similar to the previous when $\tau = 0.5$. In general, the RB% and MSE tend to zero when n increases. For $\tau = 0.1$ the α estimates have the biggest bias. However, when $\tau = 0.9$, this does not occur. The estimated coverage probabilities with $\tau = 0.1$ are similar to those for $\tau = 0.5$, and when $n > 125$, they closely approach 95% for all parameters. On the other hand, if $\tau = 0.9$, these coverage probabilities tend to approach 95% at a slower rate for the parameters ϕ_1, θ_1, and c.

We also assess the performance of the UBXII-ARMA model's CMLEs under the misspecification of the random component. For this, we conduct a simulation study to compare the proposed model with the KARMA and βARMA models, which are well known in the analysis of limited time series. Another goal is to evaluate the behavior of the AIC and BIC as selection criteria for different models and analyze the performance of the forecasting measures MSE and MAPE in-sample and out-of-sample for the different schemes of data generation.

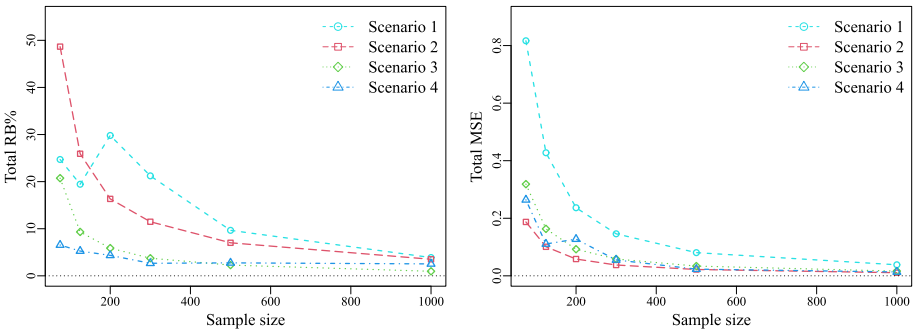


Fig. 1 Total RB%s and total MSE of the CMLEs from UBXII-ARMA(p, q) model with $n \in \{75, 125, 200, 300, 1000\}$

The KARMA and βARMA models are defined similarly to the UBXII-ARMA model. The random component in the KARMA model has Kumaraswamy (Kw) distribution, that is, $Y_t|\mathcal{F}_{t-1} \sim \text{Kw}(\omega_t, \varphi)$. The conditional density of Y_t given $\mathcal{F}_{t-1}$ by considering the standard unit interval is

$$f(y_t|\mathcal{F}_{t-1}) = \frac{\varphi \log 0.5}{\log(1-\omega_t^{\varphi})} y_t^{\varphi}(1-y_t^{\varphi})^{\log 0.5/\log(1-\omega_t^{\varphi})-1}, \quad y_t \in (0,1),$$

where $0 < \omega_t < 1$ is the conditional median of $Y_t|\mathcal{F}_{t-1}$ and $\varphi > 0$ is a precision parameter. The dynamic component is similar to the given by (4). It is obtained by replacing the quantile q_t by the median ω_t in (4) with r_t measured on the predictor scale.

In the βARMA model, the random component has a beta distribution. Then, in this case, $Y_t|\mathcal{F}_{t-1} \sim \text{beta}(\mu_t, \varphi)$ with conditional distribution given by

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

where $0 < \mu_t < 1$ is the conditional mean of $Y_t|\mathcal{F}_{t-1}$, $\delta > 0$ is a precision parameter and $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}\text{e}^{-x}\text{d}x$ is the complete gamma function. Similarly, to get the dynamic component of this model, q_t must be replaced by μ_t in (4) with r_t measured on the predictor scale.

In what follows, we describe how the simulation was performed.

1. We generated a sample of size $n = 200$ observations for each one of the models with ARMA(1, 1) structures and true parameters: $\alpha = 0.1$, $\phi_1 = 0.4$, $\theta_1 = 0.3$, and $c = 6.6$ for the UBXII-ARMA model (with $\tau = 0.5$), $\varphi = 9.8$ for the KARMA model, and $\delta = 62$ for the βARMA model.
2. We fit the true model and the UBXII-ARMA model for each generating scheme. When the UBXII-ARMA is the true model, we fit the others two competing models.
3. We compute the MSE, MAPE (in-sample and out-of-sample), AIC, and BIC for the three fitted models in each setting.
4. To obtain the out-of-sample forecastings, we consider a forecasting horizon with $h = 6$. That is, the last six observations were separated from the sample, and the model is fitted with the 194 remaining observations.
5. Each scheme is repeated $R = 10,000$ times.
6. After, we then calculated the frequencies at which the AIC (AIC(%)) and BIC (BIC(%)) criteria selected the corrected model.

Table 4 Estimated coverage probability from the asymptotic confidence intervals for $\boldsymbol{\gamma}$

Parameter	α	ϕ_1	ϕ_2	θ_1	θ_2	c
n	Scenario 1					
75	0.8244	0.7556	0.8683	0.7463	0.7770	0.9271
125	0.8777	0.8316	0.9028	0.8313	0.8449	0.9376
200	0.9037	0.8763	0.9270	0.8781	0.8845	0.9412
300	0.9233	0.9038	0.9356	0.9042	0.9086	0.9450
500	0.9354	0.9307	0.9426	0.9282	0.9322	0.9498
1000	0.9455	0.9425	0.9463	0.9427	0.9398	0.9510
n	Scenario 2					
75	0.9486	0.9491	–	0.9275	–	0.9437
125	0.9521	0.9509	–	0.9377	–	0.9439
200	0.9541	0.9497	–	0.9405	–	0.9485
300	0.9533	0.9532	–	0.9442	–	0.9482
500	0.9523	0.9501	–	0.9443	–	0.9507
1000	0.9501	0.9477	–	0.9445	–	0.9501
n	Scenario 3					
75	0.9249	0.8967	0.9020	0.8867	–	0.9369
125	0.9396	0.9201	0.9205	0.9180	–	0.9427
200	0.9461	0.9364	0.9337	0.9342	–	0.9459
300	0.9488	0.9431	0.9394	0.9412	–	0.9460
500	0.9472	0.9464	0.9452	0.9456	–	0.9498
1000	0.9479	0.9502	0.9500	0.9493	–	0.9511
n	Scenario 4					
75	0.9277	0.9300	–	0.9161	0.9070	0.9303
125	0.9387	0.9416	–	0.9278	0.9292	0.9376
200	0.9439	0.9385	–	0.9359	0.9368	0.9404
300	0.9429	0.9442	–	0.9442	0.9457	0.9448
500	0.9464	0.9445	–	0.9418	0.9431	0.9397
1000	0.9510	0.9392	–	0.9442	0.9483	0.9420

7. We also obtain the mean of MSEs and MAPEs (in-sample and out-of-sample) for each setting.

Table 5 displays the performance of the UBXII-ARMA model when compared with the well-known βARMA and KARMA models used for analyzing bounded time series. One can note that the estimates obtained from the βARMA model and, mainly, from the KARMA model differ from those using the proposed UBXII-ARMA model. The βARMA model is the most competitive but still presents a worse performance for fitting UBXII random variables. The selection percentage of AIC and BIC is the same for all cases. However, the results indicate that in this study both information criteria are reliable for model selection among the considered competitive models since they selected the correct model in most cases. The mean MSEs (in-sample and out-of-sample) were very close in each scheme. On the other hand, the mean MAPEs obtained for out-of-sample forecasting favor the UBXII-ARMA model in all the cases.

Table 5 Parameter estimates and standard errors (in parentheses) from Monte Carlo simulations for ARMA(1,1) structures with parameters $\alpha = 0.1$, $\phi_1 = 0.4$, $\theta_1 = 0.3$, and $\nu \in \{6.6, 9.8, 62\}$

Simulation	UBXII-ARMA(1,1)			KARMA(1,1)		βARMA(1,1)	
Par./Meas	UBXII-ARMA	KARMA	βARMA	UBXII-ARMA	KARMA	UBXII-ARMA	βARMA
$\hat{\alpha}$	0.1015	0.1917	0.1440	0.0591	0.1008	0.0821	0.1019
	(0.0306)	(0.0568)	(0.0364)	(0.0346)	(0.0261)	(0.0328)	(0.0302)
$\hat{\phi}_1$	0.3943	0.2699	0.3791	0.2306	0.3939	0.3296	0.3865
	(0.0863)	(0.1886)	(0.1071)	(0.1850)	(0.0888)	(0.1281)	(0.1065)
$\hat{\theta}_1$	0.3044	0.3087	0.3082	0.3127	0.3051	0.3050	0.3077
	(0.0896)	(0.1955)	(0.1104)	(0.1895)	(0.0894)	(0.1344)	(0.1107)
$\hat{\nu}$	6.6811	8.7562	65.1748	5.4559	9.9379	5.9402	63.5678
	(0.3578)	(0.7221)	(8.2662)	(0.3537)	(0.5758)	(0.3316)	(6.4870)
AIC(%)	99.2700	0.0000	0.7300	0.0100	99.9900	4.9000	95.1000
BIC(%)	99.2700	0.0000	0.7300	0.0100	99.9900	4.9000	95.1000
$\overline{\text{MSE}}$ (in)	0.0037	0.0038	0.0037	0.0043	0.0042	0.0039	0.0038
$\overline{\text{MSE}}$ (out)	0.0057	0.0057	0.0057	0.0063	0.0063	0.0060	0.0059
$\overline{\text{MAPE}}$ (in)	8.48	8.98	8.66	10.42	10.33	9.53	9.53
$\overline{\text{MAPE}}$ (out)	10.67	11.17	10.87	12.62	13.04	11.67	11.75

$\hat{\nu}$ is $\hat{c}$ for UBXII-ARMA model, $\hat{\varphi}$ for KARMA model, and $\hat{\delta}$ for βARMA model

Table 6 Descriptive statistics of the monthly average proportions of stocked energy in the Southeast of Brazil

Min	1st Qu.	Med	Mean	3rd Qu.	Max	Var	Asym	Exc. Kurt
0.1582	0.3733	0.5547	0.5464	0.7278	0.8782	0.0411	− 0.0515	− 1.2480

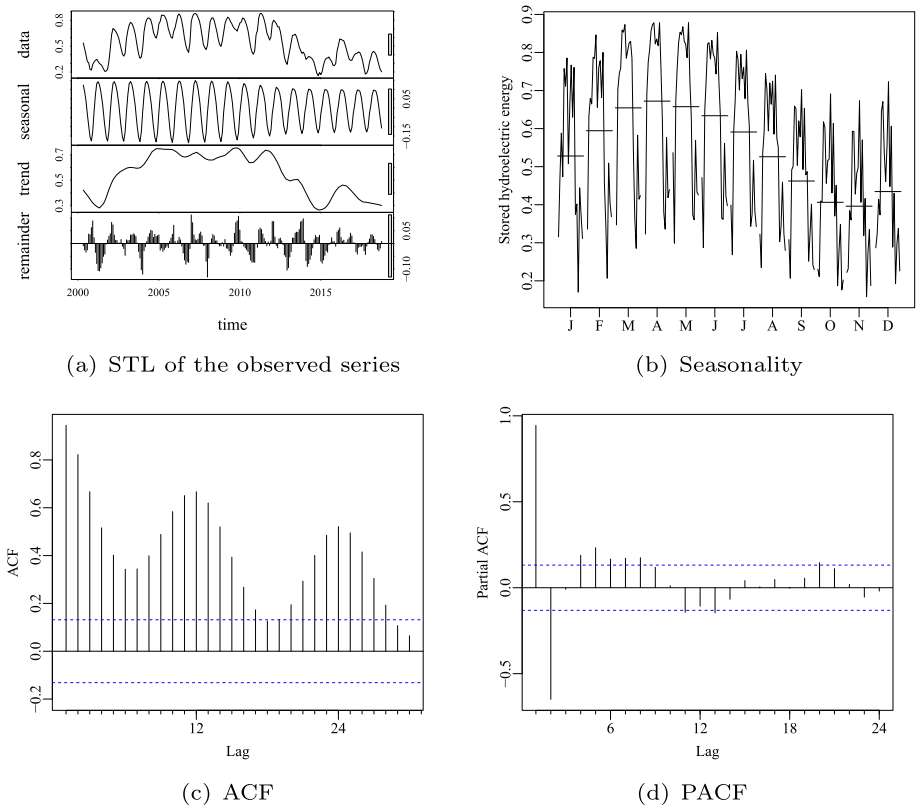
7 Application

This section presents an empirical application study of the UBXII-ARMA(p, q) model. The data refer to the monthly proportion of stored hydroelectric energy in Southeast Brazil and are available at [32]. The dynamics of the proportion of stored hydroelectric energy in South Brazil was studied by Palm and Bayer (2017) and, more recently, it was also analyzed by Scher et al. (2020). Studying the stored energy is essential to best managing water resources and ensuring the energy supply in Brazil. Energy forecasting is useful to predict problems of lack or accumulation of energy and could avoid waste. Consequently, one can also attend to energy demand satisfactorily and reduce costs and environmental consequences (Hong et al. 2014; Palm and Bayer 2017; Shaqsi et al. 2020).

In this application, we consider the time series of the proportion of stored hydroelectric energy in Southeast Brazil in the period from May 2000 to August 2019, thus covering 232 months. The last ten observations were considered only to assess the forecasting performance of the model. Hence, we fitted the model for the time series from May 2000 to April 2019, totaling 222 months. Our interest is to model the median; hence, we set $\tau = 0.5$. Modeling the median instead of the mean in time series offers the advantage of increased robustness against extreme values or outliers. While the mean can be heavily influenced by outliers, which may not be representative of the overall data pattern, the median is less affected by such extreme values. We used the programming language `R` (R Core Team 2023) to carry out the whole analysis. The computer code used in this application is available at https://github.com/tatianefribeiro/ubxiiarma.

Table 6 brings some descriptive statistic measures. Notice that these data present approximately symmetric distribution since the mean and median are quite close and the asymmetry coefficient is close to zero. This indicates that it is more suitable to model the conditional median of this time series than another quantile. Figure 2a shows the seasonal trend decomposition using LOESS (STL) (Cleveland et al. 1990) for time series of the proportion of stored energy. We may identify the annual seasonality and that the proportions of stored energy were lower before 2002 and after 2013. Figure 2b also shows the seasonality of the time series. The seasonality is obtained using the `monthplot` function, which extracts subseries from a time series and plots them all in one frame (R Core Team 2023). The proportions of stored energy grow until April, then decrease from April to November. This seasonal pattern may be modeled using trigonometric functions as covariates as a simple harmonic regression approach (Bloomfield 2004). Figure 2c shows the sample autocorrelation function (ACF) of the time series, whereas Fig. 2d brings the sample partial autocorrelation function (PACF). Note that there is a cut-off after lag 2 in the PACF's plot. This indicates that this series can be well modeled by a structure of the type AR(2).

In addition to the UBXII-ARMA model, we also fitted the βARMA and KARMA models for comparison purposes. The final models were selected according to the AIC criterion. All considered models have autoregressive and moving average orders up to three and logit link function. The smallest AIC in each class was obtained by the UBXII-AR(2), βAR(2), and KAR(2) models. The three models agree with what was observed in the PACF'S plot.



(a) STL of the observed series (b) Seasonality

(c) ACF (d) PACF

Fig. 2 Observed proportions of stored hydroelectric energy time series in Southeast of Brazil, its Seasonal-Trend decomposition, autocorrelation, and partial autocorrelation plots

The proposed UBXII-AR(2) model assumes that the monthly proportion Y_t has UBXII distribution with conditional median q_t and shape parameter c, with the dynamic component

$$\log\left(\frac{q_t}{1-q_t}\right) = \alpha + \boldsymbol{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^{2} \phi_i \left[\log\left(\frac{y_{t-i}}{1-y_{t-i}}\right) - \boldsymbol{x}_{t-i}^\top \boldsymbol{\beta}\right], \tag{9}$$

where $\alpha \in \mathbb{R}$ is a parameter to be estimated (intercept of the model), $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top \in \mathbb{R}^3$ is the parameter vector associated to the vector of covariates $\boldsymbol{x}_t^\top = (\cos(2\pi t/12), \sin(2\pi t/12), crisis_t)$, and ϕ_1 and ϕ_2 are the AR coefficients. The covariate $crisis_t$ is an indicator variable equal to 1 for t before 2002 or after 2013 ($t = 1, \ldots, 20, 153, \ldots, 222$) and zero otherwise. The βAR(2) model has assumes that $Y_t|\mathcal{F}_{t-1}$ follows a beta distribution ($Y_t|\mathcal{F}_{t-1} \sim \text{beta}(\mu_t, \delta)$) in that the dynamic component is obtained from Eq. (9) by replacing q_t by μ_t. Finally, in the KAR(2) model is supposed that $Y_t|\mathcal{F}_{t-1}$ follows a Kumaraswamy distribution ($Y_t|\mathcal{F}_{t-1} \sim \text{Kw}(\omega_t, \varphi)$) and the dynamic component also is given in (9), by replacing q_t by ω_t.

Table 7 gives the parameter estimates, standard errors, z statistic value, and p value of each fitted model. The residual analysis indicated that all these models were appropriate to describe the monthly proportion of stored energy. Note that, as expected, both covariates

Table 7 Estimates, standard errors, z values and p values of the parameters α (intercept), β_j ($j = 1, 2, 3$) and ϕ_k ($k = 1, 2$) (autoregressive coefficients), and c of the UBXII-AR(2) model for the proportion of stored hydroelectric energy in Southeast Brazil

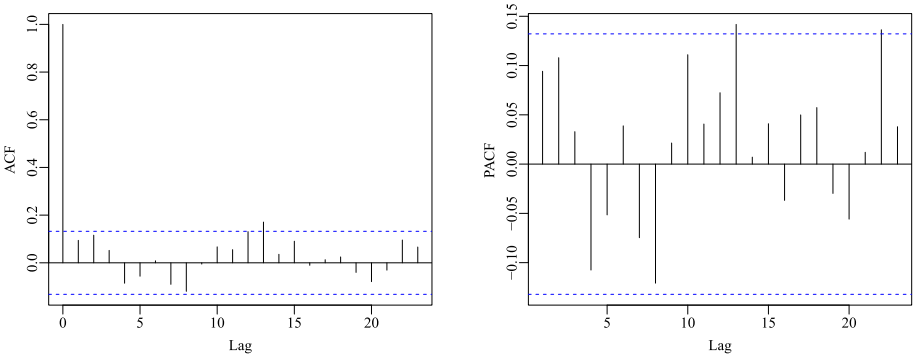
	Estimate	SE	z value	p value
		UBXII-AR(2)		
α	0.0206	0.0156	1.3260	0.1848
β_1	0.4034	0.0472	8.5449	< 0.0001
β_2	0.1138	0.0419	2.7172	0.0066
β_3	-0.2630	0.1316	-1.9988	0.0456
ϕ_1	1.3222	0.0432	30.5828	< 0.0001
ϕ_2	-0.4072	0.0430	-9.4752	< 0.0001
c	11.3464	0.6468	17.5428	< 0.0001
		βAR(2)		
α	0.0071	0.0097	0.7267	0.4674
β_1	0.6172	0.0402	15.3585	< 0.0001
β_2	0.1791	0.0395	4.5336	< 0.0001
β_3	0.0155	0.0948	0.1632	0.8704
ϕ_1	1.3797	0.0504	27.3683	< 0.0001
ϕ_2	-0.4170	0.0506	-8.2343	< 0.0001
δ	200.7800	19.1045	10.5096	< 0.0001
		KAR(2)		
α	0.0304	0.0132	2.3106	0.0209
β_1	0.8756	0.0637	13.7389	< 0.0001
β_2	0.3578	0.0869	4.1184	< 0.0001
β_3	0.0912	0.0746	1.2232	0.2213
ϕ_1	1.6120	0.0644	25.0486	< 0.0001
ϕ_1	-0.6674	0.0621	-10.7487	< 0.0001
φ	14.6954	0.7177	20.4769	< 0.0001

considered to account for this monthly seasonal component were significant at the usual significance level of 5% in all the classes of models.

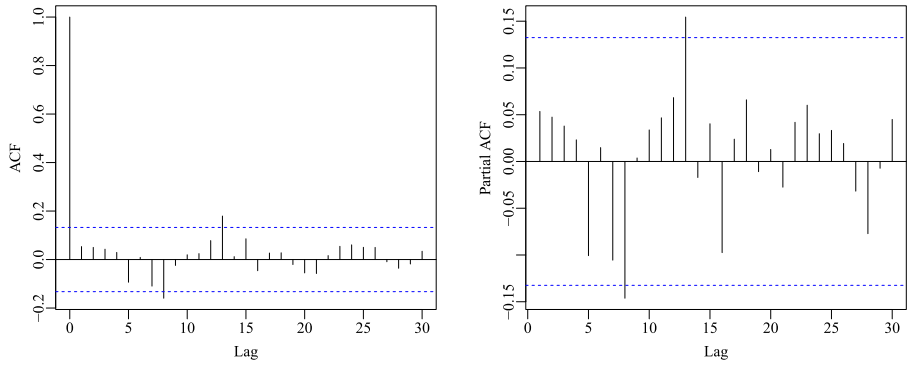
The crisis effect was significant only for the UBXII-AR(2) ($p = 0.0456$). The crisis effect estimate is -0.2630, resulting in $\exp(-0.2630) = 0.769$, indicating that during the crisis the odds ratio decreased 23.1%. For example, if the proportion of stored hydroelectric energy is 50% in the non-crisis period, the stored energy is only 43.5% during the crisis period. For a proportion of 40% before de crisis, this proportion reaches 33.9% in the crisis, for the same month of the year.

Figure 3a, b shows the residual ACF and PACF from the fitted UBXII-AR(2) model. Both the plots indicate that the residuals do not present significant autocorrelation. The sample ACFs and PACFs of the residuals from βAR(2) and KAR(2) are similar those UBXII-AR(2) model. Consequently, the models appear suitable for out-of-sample forecasting.

Figure 4a gives a plot of the observed (solid lines) and predicted values (blue dashed lines) from the fitted UBXII-AR(2) model. Note that the proposed model provides accurate forecasts since the fitted values are quite close to observed data over time. That is, the UBXII-AR(2) model is suitable to capture the proportion of stored energy dynamics. Similarly, in the out-of-sample forecasting comparison, the new dynamic model has the best performance; see Table 8. Moreover, Fig. 4a also shows naive prediction which is the predicted value not including the autoregressive and moving average terms (red dashed lines). It measures the

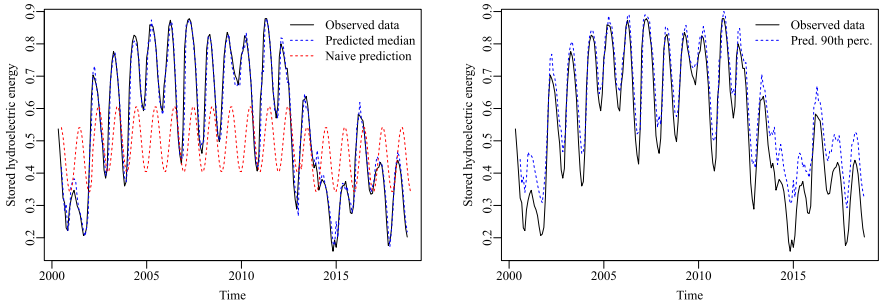


(a) Residual ACF ($\tau = 0.5$) (b) Residual PACF ($\tau = 0.5$)



(c) Residual ACF ($\tau = 0.9$) (d) Residual PACF ($\tau = 0.9$)

Fig. 3 Residual diagnostic plots of the fitted UBXII-AR(2) and UBXII-ARMA(2, 3) model for the proportion of stored hydroelectric energy in the Southeast of Brazil



(a) Observed stored energy proportion and fitted values. (b) Observed stored energy proportion and predicted 90th percentiles.

Fig. 4 In-sample forecasting performance plots from the UBXII-AR(2) and UBXII-ARMA(2, 3) models and naive prediction

effect of the covariates (cosine, sine, and crisis) on the median proportion of stored energy, excluding the AR and MA terms. That is, we plot the estimated quantiles according to

$$\log\left(\frac{\hat{q}_t}{1-\hat{q}_t}\right) = \hat{\alpha} + \hat{\beta}_1\cos(2\pi t/12) + \hat{\beta}_2\sin(2\pi t/12) + \hat{\beta}_3\text{crisis}_t,$$

where $\hat{\alpha}$ and $\hat{\beta}_j$ $(j = 1, 2, 3)$ are given in Table 7 (UBXII-AR(2) model). The plot displays the effect of the water crisis since the time series changes of the level before 2002 and after 2013.

Using the fitted models in Table 7, we calculate out-of-sample forecasts for different forecast horizons (h) ranging from 1 to 10 months. Table 8 provides values of the accuracy measures (defined in Sect. 5) for the three fitted models for each $h \in \{1, \ldots, 10\}$. According to the results, the UBXII-AR(2) provides the best out-of-sample forecasting since the MSE and MAPE presented the smallest values for all forecast horizons h considered. Thus, the UBXII-AR(2) forecasts are closest to the observed proportions for the ten months. Although the UBXII-AR(2) and βAR(2) model's MSEs and MAPEs are close for some forecasting horizons h, the βAR(2) model does not capture the crisis effect. To Cordeiro et al. (2021), when the choice of the distribution for the response variable is appropriate, generally, there is an improvement in the standard errors of the estimated regression coefficients. Therefore, we conclude that the UBXII-AR(2) model is more suitable for these data.

For this same data set, we also perform an application of the UBXII-ARMA(p, q) considering $\tau = 0.9$. Our objective is to exemplify the use of the UBXII-ARMA(p, q) model for other quantiles besides the median. Therefore, we model the 90th conditional percentile q_t of the proportion of stored hydroelectric energy in Southeast Brazil from May 2000 to April 2019. The best model selected according to the AIC criterion was the UBXII-ARMA(2, 3), given by

$$\log\left(\frac{q_t}{1-q_t}\right) = \alpha + \boldsymbol{x}_t^\top\boldsymbol{\beta} + \sum_{i=1}^{2}\phi_i\left[\log\left(\frac{y_{t-i}}{1-y_{t-i}}\right) - \boldsymbol{x}_{t-i}^\top\boldsymbol{\beta}\right] + \sum_{j=1}^{3}\theta_j r_{t-j},$$

where $\boldsymbol{x}_t^\top = (\cos(2\pi t/12), \cos(4\pi t/12), \sin(8\pi t/12))$ and θ_j $(j = 1, 2, 3)$ are the MA coefficients. Table 9 presents the parameter's estimates, standard errors, value of z statistic, and p values associated. The estimates of β_3 and ϕ_2 are not statistically significant at a significance level of 5%. Nevertheless, including these coefficients has led to an improvement in the fit of the model. The $crisis$ covariable was not considered since it was not statistically significant.

Figure 4b displays the observed stored hydroelectric energy proportion in Southeast Brazil and the predicted 90th conditional percentile. We observe that the predicted values (blue dashed lines) are higher than those observed values (solid lines) since the 90th percentile represents the value below which 90% of the observed data points lie. However, they are quite close to observed data.

The sample ACF and PACF plots of the residuals from the fitted UBXII-ARMA(2, 3) model are given in Fig. 3c, d. In both plots, just two lags are outside the 95% confidence interval. Thus, there is no evidence that the residuals exhibit significant autocorrelation or partial autocorrelation. This suggests that the model has adequately captured the temporal dependencies in the data.

Table 8 Out-of-sample forecasting performance comparison for the best fitted models in each class

Measures	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$	$h = 7$	$h = 8$	$h = 9$	$h = 10$
UBXII-AR(2)										
MSE	0.0008	0.0008	0.0010	0.0026	0.0025	0.0024	0.0022	0.0020	0.0018	0.0017
MAPE	11.8809	10.7594	12.0357	16.2349	15.2815	14.2899	13.4488	12.4404	11.4297	10.9559
βAR(2)										
MSE	0.0011	0.0010	0.0013	0.0038	0.0045	0.0050	0.0053	0.0049	0.0044	0.0039
MAPE	13.5198	12.1962	13.6820	19.3421	19.6867	19.6652	19.3141	18.2176	16.5105	14.9722
KAR(2)										
MSE	0.0005	0.0003	0.0063	0.0221	0.0345	0.0456	0.0540	0.0586	0.0598	0.0586
MAPE	9.4649	6.1112	20.8785	38.0784	44.9699	49.2189	51.9913	53.4618	54.0139	54.1628

Table 9 Fitted UBXII-ARMA(2, 3) model for the 90th conditional percentile of the proportion of stored hydroelectric energy in Southeast Brazil

	Estimate	SE	z value	p value
α	0.3401	0.0323	10.5460	< 0.0001
β_1	0.4341	0.0433	10.0162	< 0.0001
β_2	-0.0561	0.0172	-3.2635	0.0011
β_3	-0.0116	0.0065	-1.7914	0.0732
ϕ_1	0.8524	0.1275	6.6867	< 0.0001
ϕ_2	-0.0635	0.1132	-0.5610	0.5748
θ_1	0.0776	0.0225	3.4455	0.0006
θ_2	0.0568	0.0138	4.1293	< 0.0001
θ_3	0.0297	0.0096	3.0857	0.0020
c	10.5225	0.6325	16.6355	< 0.0001

8 Conclusion

This paper proposes a new class of dynamic models: the unit Burr XII quantile autoregressive moving average (UBXII-ARMA) model. The new UBXII-ARMA model is defined by incorporating autoregressive and moving average terms additively into the linear predictor of the UBXII regression proposed by Ribeiro et al. (2022). This class is appropriate for modeling and forecasting continuous dependent variables over time in the interval $(0, 1)$, such as rates, proportions, and indexes. The new model is versatile for modeling time series proportional data since the UBXII density has different shapes depending on the parameter values. Furthermore, our proposed model depends on the quantile of the response variable over time, which differentiates this model from the usual generalized autoregressive moving average (GARMA) models or other type-GARMA models that depend on the mean or median response.

We presented the conditional maximum likelihood estimation, derived closed-form expressions for the score function, and presented interval estimation based on the asymptotic properties of the conditional maximum likelihood estimators (CMLEs). We assessed the finite-sample performance of the CMLEs in the UBXII-ARMA framework through Monte Carlo simulations. Moreover, we present some diagnostic analysis and forecasting tools to check whether the proposed model captures the data dynamics.

To illustrate our proposed model and methodology, we analyzed the proportion of stored hydroelectric energy in Southeast Brazil over time. This is an important issue since we already had water and energy shortages in 2013 (Leahy 2015). According to some goodness-of-fit measures, the UBXII-ARMA model outperformed the KARMA and βARMA models for this data set. The forecasts of our model correspond to the forecast median and outperformed the KARMA and βARMA models, even being close to the βARMA forecasts. Also, only the UBXII-ARMA model captured a significant effect of lower water levels before 2002 and after 2013, as discussed in Leahy (2015). In future studies, we intend to use the UBXII-ARMA model for analyzing other hydro-environmental variables, such as relative air humidity and the proportion of the useful volume of water reservoirs. We also aim to develop prediction intervals for the UBXII-ARMA model and propose control charts based on it.

Acknowledgements We thank the three referees and Associate Editor for their valuable comments and suggestions. We also gratefully acknowledge partial financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). The author Renata Rojas Guerra acknowledges the support of Ser-

rapilheira Institute/Serra - 2211-41692; FAPERGS/23/2551-0001595-1, FAPERGS/23/2551-0000851-3; and CNPq/306274/2022-1. The author Airlane P. Alencar acknowledges FAPESP/23/02538-0.

Data Availability The data supporting this research is publicly available and can be accessed at http://www.ons.org.br/. It is also provided in the following repository https://github.com/tatianefribeiro/ubxiiarma, with all the computer codes used in the application.

Appendix A Simulation results for other link functions and quantiles

Tables 10 and 11 display simulation results on point and interval estimation of the parameters that index the UBXII-ARMA(1, 1) model for $\tau = 0.5$ with probit and cloglog link functions.

Table 10 Simulation results on point estimation of the UBXII-ARMA(1, 1) model considering link functions probit and cloglog with $\tau = 0.5$

n	Measure	$\hat{\alpha}$	$\hat{\phi}_1$	$\hat{\theta}_1$	$\hat{c}$
Link: probit					
75	RB%	10.2971	6.2068	30.9502	3.3949
	MSE	0.0073	0.0191	0.0238	0.1431
125	RB%	5.0587	3.1361	15.6065	2.0489
	MSE	0.0035	0.0090	0.0124	0.0779
200	RB%	3.1434	1.9642	9.4577	1.2377
	MSE	0.0021	0.0053	0.0074	0.0448
300	RB%	2.1849	1.3618	6.3538	0.8316
	MSE	0.0013	0.0034	0.0047	0.0287
500	RB%	1.2354	0.7832	3.8106	0.5156
	MSE	0.0008	0.0020	0.0028	0.0167
1000	RB%	0.5507	0.3558	1.6452	0.2652
	MSE	0.0004	0.0010	0.0014	0.0082
Link: cloglog					
75	RB%	10.0274	6.2765	30.9295	3.5839
	MSE	0.0064	0.0188	0.0234	0.1551
125	RB%	5.0627	3.2281	16.1747	2.1495
	MSE	0.0032	0.0093	0.0126	0.0838
200	RB%	3.1703	2.0320	9.8872	1.3007
	MSE	0.0018	0.0054	0.0075	0.0481
300	RB%	2.1963	1.4049	6.6352	0.8742
	MSE	0.0012	0.0035	0.0048	0.0308
500	RB%	1.2448	0.8077	3.9867	0.5444
	MSE	0.0007	0.0021	0.0029	0.0179
1000	RB%	0.5519	0.3633	1.7040	0.2800
	MSE	0.0004	0.0010	0.0014	0.0088

Table 11 Estimated coverage probability from the asymptotic confidence intervals for UBXII-ARMA(1, 1) model's parameters with link functions probit and cloglog ($\tau = 0.5$)

Parameter	α	ϕ_1	θ_1	c
n	Link: probit			
75	0.9444	0.9449	0.9247	0.9437
125	0.9511	0.9498	0.9387	0.9442
200	0.9529	0.9495	0.9407	0.9486
300	0.9525	0.9520	0.9459	0.9481
500	0.9527	0.9501	0.9462	0.9513
1000	0.9475	0.9479	0.9478	0.9511
n	Link: cloglog			
75	0.9464	0.9453	0.9253	0.9431
125	0.9490	0.9482	0.9378	0.9442
200	0.9526	0.9495	0.9396	0.9487
300	0.9531	0.9512	0.9461	0.9470
500	0.9529	0.9502	0.9456	0.9515
1000	0.9475	0.9477	0.9472	0.9513

Tables 12 and 13 display simulation results on point and interval estimation of the parameters that index the UBXII-ARMA(1, 1) model for $\tau \in \{0.1, 0.9\}$ with logit link function.

Table 12 Simulation results on point estimation of the UBXII-ARMA(p, q) model for $\tau \in \{0.1, 0.9\}$

n	Measure	$\hat{\alpha}$	$\hat{\phi}_1$	$\hat{\theta}_1$	$\hat{c}$
$\tau = 0.1$					
75	RB%	20.0169	4.1472	0.8748	3.6643
	MSE	0.0251	0.0132	0.0013	0.2672
125	RB%	10.2803	2.0369	0.0158	2.0998
	MSE	0.0130	0.0069	0.0007	0.1433
200	RB%	6.2271	1.2264	0.0158	1.2762
	MSE	0.0077	0.0041	0.0004	0.0832
300	RB%	4.4273	0.8718	0.0889	0.8710
	MSE	0.0050	0.0027	0.0003	0.0540
500	RB%	2.6866	0.5314	0.1067	0.5013
	MSE	0.0029	0.0015	0.0002	0.0315
1000	RB%	1.3512	0.2706	0.0227	0.2729
	MSE	0.0014	0.0008	0.0001	0.0155
$\tau = 0.9$					
75	RB%	1.0868	1.2901	2.7443	0.8696
	MSE	0.0016	0.0064	0.0001	1.8188
125	RB%	0.7281	0.4999	1.6539	0.3898
	MSE	0.0012	0.0034	0.0001	1.0897
200	RB%	0.6090	0.4699	1.0103	0.1927
	MSE	0.0012	0.0019	0.0001	0.6335
300	RB%	0.7301	0.3666	0.7788	0.0578
	MSE	0.0004	0.0012	< 0.0001	0.4130

Table 12 continued

n	Measure	$\hat{\alpha}$	$\hat{\phi}_1$	$\hat{\theta}_1$	$\hat{c}$
500	RB%	0.8259	0.3097	0.5511	0.0435
	MSE	0.0002	0.0006	< 0.0001	0.2405
1000	RB%	0.8343	0.3628	0.4038	0.0504
	MSE	0.0001	0.0003	< 0.0001	0.1197

Table 13 Estimated coverage probability from the asymptotic confidence intervals for the parameters that index the UBXII-ARMA(1, 1) model with $\tau \in \{0.1, 0.9\}$

Parameter	α	ϕ_1	θ_1	c
n	$\tau = 0.1$			
75	0.9259	0.9277	0.9222	0.9447
125	0.9399	0.9372	0.9339	0.9463
200	0.9425	0.9415	0.9424	0.9495
300	0.9461	0.9458	0.9428	0.9505
500	0.9486	0.9497	0.9480	0.9473
1000	0.9486	0.9491	0.9486	0.9470
n	$\tau = 0.9$			
75	0.9254	0.8962	0.8826	0.9014
125	0.9341	0.9011	0.9105	0.9098
200	0.9371	0.9096	0.9226	0.9201
300	0.9400	0.9172	0.9288	0.9267
500	0.9432	0.9201	0.9358	0.9219
1000	0.9437	0.9277	0.9382	0.9249

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: 2nd international symposium on information theory, vol 1973. Akademiai Kaido, pp 267–281

Akaike H (1978) A Bayesian analysis of the minimum AIC procedure. Ann Inst Stat Math 30(1):9–14

Almeida-Junior PM, Nascimento AD (2021) ARMA process for speckled data. J Stat Comput Simul 91(15):3125–3153

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

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

Bayer DM, Bayer FM, Gamba P (2020b) A 3-D spatiotemporal model for remote sensing data cubes. IEEE Trans Geosci Remote Sens 59(2):1082–1093

Bayer FM, Pumi G, Pereira TL, Souza TC (2023) Inflated beta autoregressive moving average models. Comput Appl Math 42(4):183

Benjamin MA, Rigby RA, Stasinopoulos DM (2003) Generalized autoregressive moving average models. J Am Stat Assoc 98(1):214–223

Bhatti FA, Ali A, Hamedani G, Korkmaz MÇ, Ahmad M (2021) The unit generalized log Burr XII distribution: properties and application. AIMS Math 6:10222–10252

Bloomfield P (2004) Fourier analysis of time series: an introduction. Wiley, New York

Box GE, Jenkins GM, Reinsel GC (2011) Time series analysis: forecasting and control. Wiley, New York

Brockwell Peter J, Davis Richard A (2009) Time series: theory and methods. Springer, New York

Cade BS, Noon BR (2003) A gentle introduction to quantile regression for ecologists. Front Ecol Environ 1(8):412–420

Choi B (2012) ARMA model identification. Springer, New York

Cleveland RB, Cleveland JE, William S, McRae Terpenning I (1990) Stl: a seasonal-trend decomposition procedure based on loess. J Off Stat 6:3–73

Cordeiro GM, Figueiredo D, Silva L, Ortega EM, Prataviera F (2021) Explaining COVID-19 mortality rates in the first wave in Europe. Model Assist Stat Appl 16(3):211–221
Cribari-Neto F, Scher VT, Bayer FM (2021) Beta autoregressive moving average model selection with application to modeling and forecasting stored hydroelectric energy. Int J Forecast 39:98–109
de Araújo FJM, Guerra RR, Peña-Ramírez FA (2022) The Burr XII quantile regression for salary-performance models with applications in the sports economy. Comput Appl Math 41(6):1–20
Dunn PK, Smyth GK (1996) Randomized quantile residuals. J Comput Graph Stat 5(3):236–244
Guerra RR, Peña-Ramírez FA, Cordeiro GM (2021) The Weibull Burr XII distribution in lifetime and income analysis. Anais da Academia Brasileira de Ciências 93:e20190961
Hong T, Pinson P, Fan S (2014) Global energy forecasting competition 2012. Elsevier, Amsterdam
Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. Int J Forecast 22(4):679–688
Korkmaz MÇ, Chesneau C (2021) On the unit Burr-XII distribution with the quantile regression modeling and applications. Comput Appl Math 40(1):1–26
Korkmaz MÇ, Korkmaz ZS (2021) The unit log–log distribution: a new unit distribution with alternative quantile regression modeling and educational measurements applications. J Appl Stat 50(4):889–908
Korkmaz MÇ, Altun E, Alizadeh M, El-Morshedy M (2021a) The log exponential-power distribution: properties, estimations and quantile regression model. Mathematics 9(21):2634
Korkmaz MÇ, Chesneau C, Korkmaz ZS (2021b) Transmuted unit Rayleigh quantile regression model: Alternative to beta and Kumaraswamy quantile regression models. Univ Politeh Buchar Sci Bull Ser Appl Math Phys 83(3):149–158
Korkmaz MÇ, Chesneau C, Korkmaz ZS (2021c) On the arcsecant hyperbolic normal distribution. Properties, quantile regression modeling and applications. Symmetry 13(1):117
Korkmaz MÇ, Altun E, Chesneau C, Yousof HM (2022a) On the unit-Chen distribution with associated quantile regression and applications. Math Slovaca 72(3):765–786
Korkmaz MC, Chesneau C, Korkmaz ZS (2022b) The unit folded normal distribution: A new unit probability distribution with the estimation procedures, quantile regression modeling and educational attainment applications. J Reliab Stat Stud 15(01):261–298
Korkmaz MÇ, Chesneau C, Korkmaz ZS (2023) A new alternative quantile regression model for the bounded response with educational measurements applications of OECD countries. J Appl Stat 50(1):131–154
Leahy J (2015) São Paulo drought raises fears of Brazil energy crisis. Financial times. https://www.ft.com/content/a140a1e6-b14e-11e4-a830-00144feab7de. Accessed 13 September 2021
Lehner B, Messager ML, Korver MC, Linke S (2022) Global hydro-environmental lake characteristics at high spatial resolution. Sci Data 9(1):1–19
Lima LB, Cribari-Neto F, Lima-Junior DP (2022) Dynamic quantile regression for trend analysis of streamflow time series. River Res Appl 38(6):1051–1060
Lindsay BG, Li B (1997) On second-order optimality of the observed Fisher information. Ann Stat 25(5):2172–2199
Mazucheli J, Menezes AFB, Fernandes LB, de Oliveira RP, Ghitany ME (2020) The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates. J Appl Stat 47(6):954–974
Mazucheli J, Alves B, Korkmaz MÇ, Leiva V (2022) Vasicek quantile and mean regression models for bounded data: new formulation, mathematical derivations, and numerical applications. Mathematics 10(9):1389
Mazucheli J, Korkmaz MÇ, Menezes AF, Leiva V (2023) The unit generalized half-normal quantile regression model: formulation, estimation, diagnostics, and numerical applications. Soft Comput 27(1):279–295
Melo M, Alencar A (2020) Conway–Maxwell–Poisson autoregressive moving average model for equidispersed, underdispersed, and overdispersed count data. J Time Ser Anal 41(6):830–857
Mohsenipour M, Shahid S, Ziarh GF, Yaseen ZM (2020) Changes in monsoon rainfall distribution of Bangladesh using quantile regression model. Theor Appl Climatol 142:1329–1342
Operador Nacional do Sistema Elétrico (2023). http://www.ons.org.br/. Accessed 22 July 2023
Palm BG, Bayer FM (2017) Bootstrap-based inferential improvements in beta autoregressive moving average model. Commun Stat Simul Comput 47(4):977–996
Palm BG, Bayer FM, Cintra RJ (2021) Signal detection and inference based on the beta binomial autoregressive moving average model. Digit Signal Proc 109:102911
Palm BG, Bayer FM, Cintra RJ (2022) 2-D Rayleigh autoregressive moving average model for SAR image modeling. Comput Stat Data Anal 171:107–453
Pawitan Y (2001) In all likelihood: statistical modelling and inference using likelihood. Oxford University Press, Sweden
Pereira GHA (2019) On quantile residuals in beta regression. Commun Stat Simul Comput 48(1):302–316
Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical recipes in C: the art of scientific computing, 2nd edn. Cambridge University Press, New York

Qin X, Gui W (2020) Statistical inference of Burr-XII distribution under progressive Type-II censored competing risks data with binomial removals. J Comput Appl Math 378:112–922
R Core Team (2023) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
Ribeiro TF, Peña-Ramírez FA, Guerra RR, Cordeiro GM (2022) Another unit Burr XII quantile regression model based on the different reparameterization applied to dropout in Brazilian undergraduate courses. PLoS ONE 17(11):1–25
Rocha AV, Cribari-Neto F (2009) Beta autoregressive moving average models. TEST 18(3):529–545
Sagrillo M, Guerra RR, Bayer FM (2021) Modified Kumaraswamy distributions for double bounded hydroenvironmental 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
Scher VT, Cribari-Neto F, Bayer FM (2023) Generalized βARMA model for double bounded time series forecasting. Int J Forecast. https://www.sciencedirect.com/science/article/abs/pii/S0169207023000493
Sen PK, Singer JM, de Lima ACP (2009) From finite sample to asymptotic methods. Cambridge University Press, New York
Shaqsi AZA, Sopian K, Al-Hinai A (2020) Review of energy storage services, applications, limitations, and benefits. Energy Rep 6:288–306
Silva GO, Ortega EMM, Cancho VG, Barreto ML (2008) Log-Burr XII regression models with censored data. Comput Stat Data Anal 52(7):3820–3842
Wald A (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans Am Math Soc 54(3):426–482