UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

KLEBER HENRIQUE DOS SANTOS

A NEW DYNAMIC BETA PRIME MODEL WITH APPLICATION TO HYDRO-ENVIRONMENTAL DATA

Recife

2024

KLEBER HENRIQUE DOS SANTOS

A NEW DYNAMIC BETA PRIME MODEL WITH APPLICATION TO HYDRO-ENVIRONMENTAL DATA

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Área de Concentração: Estatística Aplicada

Orientador (a): Dr. Francisco Cribari Neto

Recife

2024

Catalogação na fonte
Bibliotecária: Mônica Uchôa, CRB4-1010

S237n Santos, Kleber Henrique dos.
A new dynamic beta prime model with application to hydro-environmental data / Kleber Henrique dos Santos.– 2024.
55 f.: il., fig.; tab.

Orientador: Francisco Cribari Neto.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. Centro de Ciências Exatas e da Natureza, Programa de Pós-graduação em Estatística. Recife, 2024.
Inclui referências.

1. Distribuição beta prime. 2. Modelo BPARMA generalizado. 3. Previsão. 4. Séries temporais. 5. Precisão variável. I. Cribari Neto, Francisco (orientador). II. Título.

310 CDD (23. ed.) UFPE - CCEN 2024 – 100

KLEBER HENRIQUE DOS SANTOS

A NEW DYNAMIC BETA PRIME MODEL WITH APPLICATION TO HYDRO-ENVIRONMENTAL DATA

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovado em: 26 de fevereiro de 2024.

BANCA EXAMINADORA

Prof. Dr Francisco Cribari Neto
Presidente/Orientador, UFPE

Prof. Dr Abraão David Costa do Nascimento
Examinador Interno, UFPE

Profª Drª Tarciana Liberal Pereira de Araújo
Examinadora Externa, UFPB

I dedicate this to my daughter, Ágatha, my fiancée, Vitória, and to my parents, Edgley and Tiane.

ACKNOWLEDGEMENTS

I express my gratitude to God for the strength to persevere and for providing me with what was necessary to complete this stage in my life.

To my fiancée and companion, Vitória, express my gratitude for all the love, dedication, and for always motivating me, even in moments when doubted my ability to complete this journey.

To my daughter, Ágatha, who came to captivate my life and became my greatest source of inspiration.

To my parents, Edgley and Tiane, thank immensely for all the encouragement and motivation. Without your support, would not have come this far. Thank you also for all the advice, patience, understanding, and, above all, for constant love. My eternal gratitude.

To my siblings, Keyla and Junior, thank you for enduring my worst moments and for always motivating me.

To the relatives who cheered for my success and prayed for me.

To my advisor, Professor Francisco Cribari Neto, for the competent guidance, excellent professionalism, and patience he had with me throughout this journey. Without your encouragement, support, and critical reviews, this work would never have been completed.

To the professors of the postgraduate program in statistics of the UFPE, thank you for the shared knowledge and encouragement, especially professors Francisco Cribari Neto, Gauss Cordeiro, Maria do Carmo Soares de Lima, Abraão David Costa do Nascimento, and Roberto Manghi.

To my postgraduate school colleagues, thank you for the shared moments and experiences that made this journey lighter and more enjoyable. Especially to Paula, Camila, Ewellyn, Jhon, Santiago, Ana Carla, Eric, Mari, Ana Bacca, Alan, Yessenia, Arthur, Ernando, and Jaime.

To the professors of the department of statistics of the UFPB, thank you for playing a fundamental role in my education. In particular, would like to express my gratitude to professors Tatiene, Ana Hermínia, Ana Flávia, Isabel, Everlane, Maria Lídia, Cláudio, Rodrigo, Marcelo, Hemílio, Luiz, Pedro, Eufrásio, Ulisses, Ronei, and Neir. A special thanks to professor Tarciana, who guided me during my undergraduate studies. Without her advice, teachings, and motivation, would not be completing this journey.

To the members of the examining committee, thank you for your willingness to participate

and for the valuable considerations.

To CAPES, for the financial support granted throughout this journey.

"Never lose sight of your starting point." - Saint Clare of Assisi

ABSTRACT

We introduce a dynamic model tailored for positively valued time series. It accommodates both autoregressive and moving average dynamics and allows for explanatory variables. The underlying assumption is that each random variable follows, conditional on the set of previous information, the beta prime distribution. A novel feature of the proposed model is that both the conditional mean and conditional precision evolve over time. The model thus comprises two dynamic submodels, one for each parameter. The proposed model for the conditional precision is parsimonious, incorporating first-order time dependence. Changes over time in the form of the distribution are determined by the time evolution of two parameters, and not just of the conditional mean. We present simple closed-form expressions for the model's conditional log-likelihood function, score vector and Fisher's information matrix. We also present Monte Carlo simulation results on the finite-sample performance of the conditional maximum likelihood estimators of the parameters that index the model. Finally, we use the proposed approach to model and forecast two seasonal water flow time series. Specifically, we model the inflow and outflow rates of the reservoirs of two hydroelectric power plants. Overall, the forecasts obtained using the proposed model are more accurate than those yielded by alternative models.

Keywords: Beta prime distribution; Generalized BPARMA model; Forecasting; Time series; Variable precision.

RESUMO

Apresentamos um modelo dinâmico para séries temporais que assumem apenas valores positivos. O modelo proposto acomoda dinâmicas autorregressivas e de médias móveis e permite a inclusão de variáveis explicativas. A suposição central é que cada variável aleatória segue, condicional ao conjunto de informações anteriores, distribuição beta prime. Uma característica inovadora do modelo proposto é que tanto a média condicional quanto a precisão condicional evoluem ao longo do tempo. O modelo compreende, portanto, dois submodelos dinâmicos, um para cada parâmetro. O modelo proposto para a precisão condicional é parcimonioso, incorporando dependência temporal de primeira ordem. Mudanças ao longo do tempo na forma da distribuição são determinadas pela evolução temporal dos dois parâmetros, e não apenas da média condicional. Apresentamos expressões simples em forma fechada para a função de log-verossimilhança condicional do modelo, vetor escore condicional e matriz de informação de Fisher condicional. Também apresentamos resultados de simulação de Monte Carlo sobre o desempenho em amostras finitas dos estimadores de máxima verossimilhança condicional dos parâmetros que indexam o modelo. Finalmente, usamos a abordagem proposta para modelar e prever duas séries temporais sazonias de fluxo de água. Especificamente, modelamos as vazões de entrada e saída dos reservatórios de duas usinas hidrelétricas. No geral, as previsões obtidas usando o modelo proposto são mais precisas do que aquelas geradas por modelos alternativos.

Palavras-chave: Distribuição beta prime; Modelo BPARMA generalizado; Previsão; Séries temporais; Precisão variável.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

A time series is a sequence of random variables that evolve sequentially over time. Time series analysis generally aims at describing the existing time dependence structure and to produce forecasts of future values. The Gaussian autoregressive moving average (ARMA) model, proposed by Box and Jenkins (1976), and its seasonal counterpart are commonly used to represent time series dynamics in a wide range of fields. It has, nonetheless, some limitations. For instance, it is tailored for random variables that assume values in $(-\infty, \infty)$ and is based on a distribution that displays no skewness and no excess kurtosis. Additionally, the model may yield improper (i.e., negative) forecasts when fitted to a time series that assumes values in $(0, \infty)$. Notably, positively-valued random variables typically display positive skewness and varying levels of kurtosis. It is thus important to develop time series models that are tailored for such processes.

In this research we introduce a class of time series models suitable for random variables that assume values in $\mathbb{R}_+$ and display positive skewness. It is based on the assumption that, conditional on the set of previous information, each variable in the stochastic process follows the beta prime distribution. The conditional mean and conditional precision are constantly evolving. The generalized model incorporates two submodels, one for the mean and another for the precision. The model allows autoregressive and moving average dynamics and also non-stochastic covariates, such as time trends and harmonic regressors used to describe seasonal movements. Notably, the beta prime distribution is related to the well known beta law, which is the most commonly used law for random variables with support in the standard unit interval. Indeed, if Z is beta distributed, then $Z/(1-Z)$ is beta prime distributed. Conversely, if W is a beta prime random variable, then $W/(1+W)$ is beta distributed.

More specifically, we introduce the class of generalized beta prime autoregressive moving average (generalized BPARMA) models, develop conditional likelihood inference for its parameters and show how to produce out-of-sample forecasts from a fitted model. The generalized BPARMA model is observation-driven and yields predictions that are guaranteed to be proper, that is, that to be in $(0, \infty)$. Its dynamic formulation for the mean submodel is similar to those of Generalized Autoregressive Moving Average (GARMA), Beta Autoregressive Moving Average (βARMA), Kumaraswamy Autoregressive Moving Average (KARMA), $\mathscr{G}_I^0$ Autoregressive Moving Average ($\mathscr{G}_I^0$ ARMA) and Chen Autoregressive Moving Average (CHARMA)

models. For details on these models, see, respectively, Benjamin, Rigby and Stasinopoulos (2003), Rocha and Cribari-Neto (2009), Rocha and Cribari-Neto (2017), Bayer, Bayer and Pumi (2017), Almeida-Junior and Nascimento (2021) and Stone et al. (2023). The generalized BPARMA model includes, however, a second submodel which represents the evolution of the precision parameter over time. Thus, both parameters of the beta prime law evolve over time. This adds a layer of flexibility to the model, since changes in the conditional density function over time are controlled by two parameters, and not by a single parameter. The dynamic formulation for the precision submodel is similar to that used in the generalized βARMA proposed by Scher, Cribari-Neto and Bayer (2024).

We use the proposed model to represent the monthly inflow of the Caconde water reservoir and the monthly outflow of the Guilman Amorim reservoir, both in Brazil. Harmonic regressors and an indicator variable are used to represent the existing seasonality in the time series. Out-of-sample forecasts are compared to those obtained from dynamic models based on the chi-squared, gamma and log-normal laws and also to predictions obtained from fixed precision BPARMA model, and the Gaussian SARIMA model. The results favor the generalized BPARMA model.

The main contributions of this research are as follows: (i) we introduce a dynamic model for positive time series based on the beta prime distribution; (ii) the proposed model contains two submodels, one for the conditional mean of the stochastic process and one for the conditional precision; (iii) we provide simple expressions in matrix form for the score function and for Fisher's information matrix; (iv) we present Monte Carlo simulation results on the behavior of the conditional maximum likelihood estimators in finite samples; (v) we forecast the future behavior of both the monthly inflow of the Caconde water reservoir and the monthly outflow of the Guilman Amorim water reservoir, each connected to a Brazilian hydroelectric power plant.

It is worth mentioning that we present expressions in matrix form that are easy to compute for the conditional score vector and the conditional expected information matrix (i.e., the conditional Fisher information matrix). These expressions can be used, for example, to carry out score tests (conditional score vector and conditional expected information) and Wald tests (conditional expected information), as detailed in Subsection 3.3. Similar results are not available for some of the alternative models. It should be noted that the use of observed information may result in a negative value for the score statistic; see Morgan, Palmer and Ridout (2007). In some cases, there is not even an expression provided for the conditional

observed information (negative of the Hessian matrix), which forces practitioners to resort to numerical approximations in order to obtain standard errors for the parameter estimates.

The remainder of the text is structured as follows. In Chapter 2, we introduce the class of generalized BPARMA models. Closed-form expressions for the model conditional log-likelihood function, conditional score function, and conditional Fisher's information matrix are presented in Chapter 3. We also briefly explain how to perform interval estimation and hypothesis tests on the model's parameters. In Chapter 4, we comment on model selection and diagnostic analysis. We also explain how to generate BPARMA sample paths and how to produce forecasts from a fitted BPARMA model. Monte Carlo simulation evidence is presented in Chapter 5. In Chapter 6 we present and discuss two hydro-environmental empirical applications. Finally, some concluding remarks are offered in Chapter 7.

2 THE GENERALIZED BPARMA MODEL

In this chapter, we will introduce a dynamic time series model in which the random component follows the beta prime marginal distribution and the systematic component that captures time dependence has an ARMA structure. This is a generalized class of dynamic models because it comprises two dynamic submodels, one for conditional mean and another for the conditional precision, both parameters evolving over time.

The beta prime (also known as inverse beta or type two beta) law was obtained by Keeping (1962) and McDonald (1984). It is indexed by two parameters, both positive. Notably, the beta law odds ratio follows the beta prime distribution, that is, if Z follows the beta distribution with parameters α and β, then $Y = Z/(1-Z)$ is beta prime distributed. Additionally, the ratio of independent random variables that follow standardized gamma distributions with unit scale parameter is also beta prime distributed. That is, if $U \sim GA(\alpha, 1)$, $V \sim GA(\beta, 1)$, and U and V are independent, then $Y = U/V$ follows the beta prime law.

The beta prime distribution is an attractive alternative to the more well-known Weibull, gamma and inverse Gaussian distributions. First, its hazard rate function may assume upside-down bathtub or increasing shapes depending on the parameter values whereas those of most classical two-parameter distributions, such as Weibull and gamma, are monotone. Second, the beta prime skewness and kurtosis may considerably exceed those of the gamma and inverse Gaussian laws. In particular, the ratio between the beta prime and gamma kurtosis coefficients exceeds one. This feature may be useful in some settings. For further details, see Bourguignon, Santos-Neto and Castro (2021).

Let Y be a random variable that follows the beta prime distribution, denoted $Y \sim BP(\alpha, \beta)$. The cumulative distribution and probability density functions of Y are, respectively,

$$F(y; \alpha, \beta) = I_{y/(1+y)}(\alpha, \beta)$$

and

$$f(y; \alpha, \beta) = \frac{1}{\mathscr{B}(\alpha, \beta)} y^{\alpha-1}(1+y)^{-(\alpha+\beta)},$$

$y > 0$. Here, $\alpha > 0$ and $\beta > 0$ are shape parameters, $I_x(\alpha, \beta) = \mathscr{B}_x(\alpha, \beta)/\mathscr{B}(\alpha, \beta)$ is the regularized beta function, $\mathscr{B}_x(\alpha, \beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt$ is the incomplete beta function, $\mathscr{B}(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ is the (complete) beta function and $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$ is

the gamma function. The mean and variance of Y are, respectively,

$$\mathbb{E}(Y) = \frac{\alpha}{\beta - 1}, \quad \beta > 1, \quad \text{and} \quad \mathrm{Var}(Y) = \frac{\alpha(\alpha + \beta - 1)}{(\beta - 2)(\beta - 1)^2}, \quad \beta > 2.$$

The distribution skewness is positive for all parameter values.

Bourguignon, Santos-Neto and Castro (2021) considered a new beta prime parameterization in terms of the distribution mean μ and a precision parameter ϕ. To that end, they set $\mu = \alpha/(\beta - 1)$ and $\phi = \beta - 2$. We write $Y \sim BP(\mu, \phi)$ to denote that Y follows the beta prime law under this parameterization. The distribution function becomes

$$F(y; \mu, \phi) = I_{y/(1+y)}(\mu(1 + \phi), \phi + 2),$$

$\mu, \phi > 0$. Here, $I_{y/(1+y)}(\mu(1 + \phi), \phi + 2) = \mathscr{B}_{y/(1+y)}(\mu(1 + \phi), \phi + 2)/\mathscr{B}(\mu(1 + \phi), \phi + 2)$. The probability density function is

$$f(y; \mu, \phi) = \frac{y^{\mu(1+\phi)-1}(1 + y)^{-[\mu(1+\phi)+\phi+2]}}{\mathscr{B}(\mu(1 + \phi), \phi + 2)}.$$

In Figure 1 we present beta prime density functions for different parameter values. The distribution is clearly quite flexible. Under this parameterization,

$$\mathbb{E}(Y) = \mu \quad \text{and} \quad \mathrm{Var}(Y) = \frac{\mu(1 + \mu)}{\phi}.$$

Using this setup, the authors introduced a regression model for beta prime distributed independent responses. Our interest, as noted earlier, resides on time series random variables with serial dependence. In what follows, we will use the beta prime law to construct a model for serially correlated random variables.

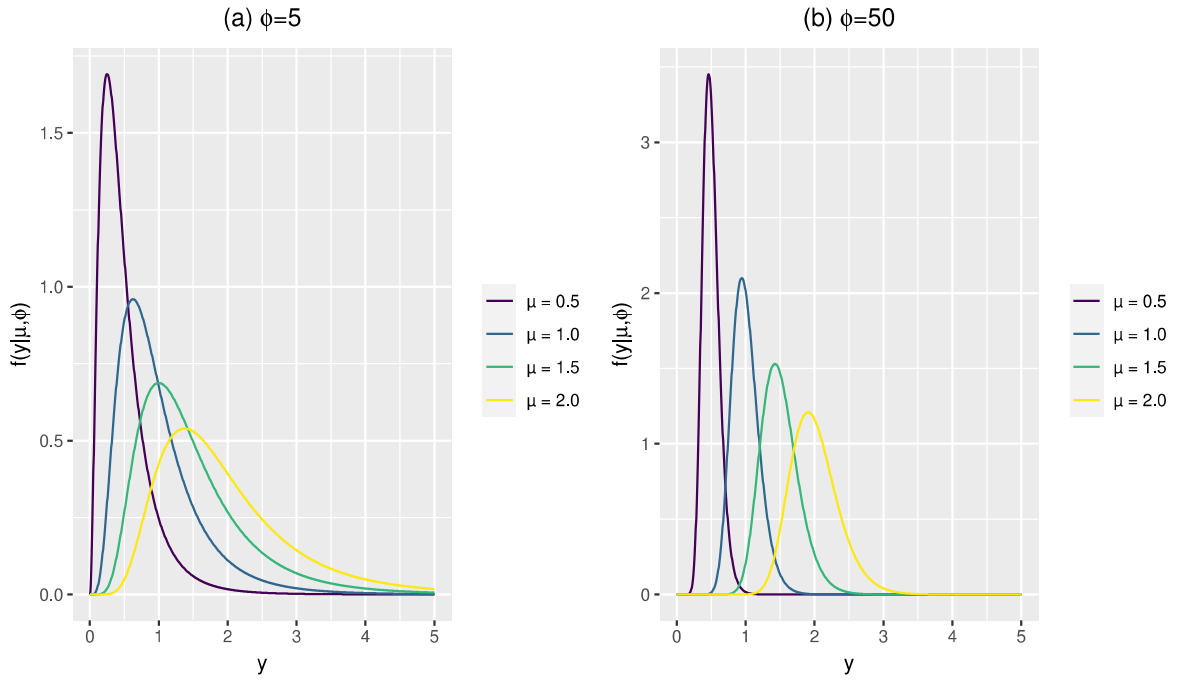
It is noteworthy that the beta prime skewness and kurtosis are impacted by the the distribution mean and precision. They are, respectively,

$$\psi_1 = \frac{2(1 + \phi)(1 + 2\mu)}{\phi - 1}\sqrt{\frac{\phi}{\mu(1 + \mu)(1 + \phi)^2}}, \quad \phi > 1,$$
$$\psi_2 = 6\left[\frac{5\phi - 1}{(\phi - 2)(\phi - 1)} + \frac{\phi}{\mu(1 + \mu)(\phi - 2)(\phi - 1)}\right], \quad \phi > 2.$$

By contrast, the skewness and kurtosis of the gamma law are only impacted by the value of the precision parameter, being given by $2\phi^{-1/2}$ and $6\phi^{-1}$, respectively.

Our aim is to introduce a dynamic model based on the beta prime law. In this model, the two parameters that index the distribution (mean and precision) evolve over time. The simultaneous temporal evolution of the two parameters controls the changes in the shape of

Figure 1 – Beta prime density functions for different parameter values.



the beta prime density over time, including changes to the asymmetry and kurtosis coefficients. Our model extends the formulation commonly used in dynamic models for positive random variables by including a separate dynamic submodel for the precision, i.e., by allowing the two parameters that index the distribution of reference to evolve over time.

Let $Y_1, \ldots, Y_n$ be time series random variables and let $\mathscr{F}_{t-1}$ be the smallest σ-algebra such that the variables $Y_1, \ldots, Y_{t-1}$ are measurable. Also, conditional on $\mathscr{F}_{t-1}$, Y_t is beta prime-distributed with mean μ_t and precision ϕ_t, i.e., $Y_t|\mathscr{F}_{t-1} \sim BP(\mu_t, \phi_t)$. Also, let $y_1, \ldots, y_n$ denote the observed values of $Y_1, \ldots, Y_n$. The conditional density function of $Y_t|\mathscr{F}_{t-1}$ is

$$f(y|\mathscr{F}_{t-1}; \mu_t, \phi_t) = \frac{y^{\mu_t(1+\phi_t)-1}(1+y)^{-[\mu_t(1+\phi_t)+\phi_t+2]}}{\mathscr{B}(\mu_t(1+\phi_t), \phi_t+2)}, \quad y > 0, \tag{2.1}$$

$\mu_t, \phi_t > 0$. It then follows that $\mathbb{E}(Y_t|\mathscr{F}_{t-1}) = \mu_t$ and $\mathrm{Var}(Y_t|\mathscr{F}_{t-1}) = \mu_t(1+\mu_t)/\phi_t$ are, respectively, the conditional mean and the conditional variance of Y_t. Note that not only the conditional mean, but also the conditional precision evolves over time.

The dynamic structure for μ_t is

$$\eta_{1t} = g_1(\mu_t) = \alpha_1 + \boldsymbol{x}_{t-i}^\top \boldsymbol{\beta} + \sum_{i=1}^{p} \varphi_i \Big[g_1(y_{t-i}) - \boldsymbol{x}_{t-i}^\top \boldsymbol{\beta} \Big] + \sum_{j=1}^{q} \theta_j r_{t-j}, \tag{2.2}$$

where $\alpha_1 \in \mathbb{R}$, $\boldsymbol{x}_t$ is an v-vector of non-stochastic regressors at time t, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_v)^\top$ is an v-vector of regression coefficients, $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)^\top$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top$ are autoregressive and moving average parameter vectors, respectively, and r_t is an error term which can be

defined in the original scale ($r_t = Y_t - \mu_t$) or in the predictor scale ($r_t = g_1(Y_t) - g_1(\mu_t)$). Also, $p, q \in \mathbb{N}$ are the autoregressive and moving average orders. Finally, $g_1 : \mathbb{R}_+ \to \mathbb{R}$ is a strictly monotone and twice differentiable link function. Its inverse must also be twice-differentiable. The dynamic formulation in (2.2) is similar to those used in GARMA (Benjamin, Rigby and Stasinopoulos (2003)), βARMA (Rocha and Cribari-Neto (2009) and (ROCHA; CRIBARI-NETO, 2017)) KARMA (Bayer, Bayer and Pumi (2017)), $\mathscr{G}_I^0$ ARMA (Almeida-Junior and Nascimento (2021)) and CHARMA (Stone et al. (2023)) models.

We propose a parsimonious dynamic structure for the temporal evolution of the precision parameter that adequately approximates past fluctuations in the level of uncertainty. To that end, we exploit an idea similar to that used by Scher, Cribari-Neto and Bayer (2024) in a dynamic model for doubly bounded random variables. We use, however, a different covariate for modeling the precision. In our case, $\mathrm{Var}(Y_t|\mathscr{F}_{t-1}) = \mu_t(1+\mu_t)/\phi_t$. Thus, for fixed ϕ_t, the conditional variance of the process grows with the conditional mean. The proposed formulation for the conditional precision submodel is

$$\eta_{2t} = g_2(\phi_t) = \alpha_2 + \delta z_{t-1}, \tag{2.3}$$

where $\alpha_2 \in \mathbb{R}$, $\delta \in \mathbb{R}$ and $z_{t-1} = y_{t-1}/y_{t-2}$, $g_2 : \mathbb{R}_+ \to \mathbb{R}$ being a strictly increasing and twice differentiable link function, g_2^{-1} also being twice differentiable. It is expected that $\alpha_2 > 0$ and $\delta < 0$. We use y_{t-1}/y_{t-2} as a proxy for the ratio between the levels of uncertainty (variability) at times $t-1$ and $t-2$. Notice that y_{t-1} is known at time t and is in $\mathscr{F}_{t-1}$. Note that the precision at t is given by $g_2^{-1}(\alpha_2 + \delta y_{t-1}/y_{t-2})$. With $\delta < 0$, the higher y_{t-1}/y_{t-2}, the lower ϕ_t. In other words, the greater the uncertainty in $t-1$ relative to $t-2$, the lower the precision in t. Notice that $y_{t-1}/y_{t-2} > 1$ signals an increase in uncertainty between times $t-2$ and $t-1$ leading to a decreased precision at time t. Conversely, a reduction in y_{t-1}/y_{t-2} signals a reduction in the conditional variance in the previous period and this leads to an increase in precision in the current period.

It is worth noticing that the standard, fixed precision BPARMA model is a particular case of our more general formulation, since it is obtained setting g_2 to be the identity function and $\delta = 0$.

The submodel for ϕ_t incorporates a dynamic structure with first-order time dependence since the conditional precision at time t is impacted by a variable that reflects the variance of the process at $t-1$. It should also be noted that the introduction of a second submodel adds a layer of flexibility to the dynamic model, since changes over time in the shape of the density

of $Y_t|\mathscr{F}_{t-1}$ are now controlled by the evolution of two parameters (μ_t and ϕ_t), rather than by the evolution of just one parameter (μ_t). In particular, the distribution skewness and kurtosis are impacted by the time evolution of the conditional mean and conditional precision.

The generalized $\mathrm{BPARMA}(p, q)$ is given by (2.1), (2.2) and (2.3). Whenever necessary, we will write $g_1(\mu_t) = \eta_{1t}$ and $g_2(\phi_t) = \eta_{2t}$, where η_{1t} and η_{2t} are the linear predictors of the conditional mean and precision submodels, respectively.

3 CONDITIONAL LIKELIHOOD INFERENCE

Let $Y_1, \ldots, Y_n$ be beta prime distributed time series random variables, each having generalized $\mathrm{BPARMA}(p,q)$ dynamics with parameter vector $\boldsymbol{\gamma} = (\alpha_1, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top, \alpha_2, \delta)^\top \in \Omega$, where $\Omega \subset \mathbb{R}^{p+q+v+3}$ is the parameter space. Let $m = \max(p,q)$. The generalized $\mathrm{BPARMA}(p,q)$ conditional log-likelihood function for $Y_1, \ldots, Y_n$ with observed values $y_1, \ldots, y_n$ is

$$\ell \equiv \ell(\boldsymbol{\gamma}) = \sum_{t=m+1}^{n} \ell_t(\mu_t, \phi_t), \tag{3.1}$$

where

$$\begin{aligned}\ell_t(\mu_t, \phi_t) = &[\mu_t(1+\phi_t) - 1]\log(y_t) - [\mu_t(1+\phi_t) + \phi_t + 2]\log(1+y_t) \\ &- \log(\Gamma(\mu_t(1+\phi_t))) - \log(\Gamma(\phi_t+2)) + \log(\Gamma(\mu_t(1+\phi_t) \\ &+ \phi_t + 2)).\end{aligned}$$

The conditional maximum likelihood estimator (CMLE) of $\boldsymbol{\gamma}$ is

$$\hat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}\in\Omega} \ell(\boldsymbol{\gamma}).$$

It cannot be expressed in closed-form. Conditional maximum likelihood estimates may be obtained by numerically maximizing the conditional log-likelihood function using a Newton or quasi-Newton algorithm; e.g., see Press et al. (1992) and Nocedal and Wright (2006). It is necessary to choose initial values for η_{1t}, η_{2t} and their recursive derivatives. We use $\eta_{1t} = g_1(Y_t)$ and set the derivatives of η_{1t} and η_{2t} with respect to the parameters in the mean and precision submodels, respectively, equal to zero for $t \in \{1, \ldots, m\}$. For more details, see Benjamin, Rigby and Stasinopoulos (2003). In the following, we will present the generalized BPARMA model conditional score function and conditional Fisher's information matrix. To that end, we partition the parameter vector $\boldsymbol{\gamma}$ as $\boldsymbol{\gamma} = (\dot{\boldsymbol{\gamma}}^\top, \ddot{\boldsymbol{\gamma}}^\top)^\top$, where $\dot{\boldsymbol{\gamma}} = (\alpha_1, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top)^\top$ and $\ddot{\boldsymbol{\gamma}} = (\alpha_2, \delta)^\top$. Here, $\dot{\boldsymbol{\gamma}}$ and $\ddot{\boldsymbol{\gamma}}$ are vectors of dimensions $(p+q+v+1)\times 1$ and 2×1, respectively. We consider errors in the prediction scale, i.e., $r_t = g_1(Y_t) - \eta_{1t}$.

3.1 CONDITIONAL SCORE VECTOR

The conditional score vector is obtained as the first-order derivative of the conditional log-likelihood function given in Equation (3.1) with respect to the elements of $\boldsymbol{\gamma}$. We can use the

chain rule to obtain the conditional log-likelihood derivative with respect to the j-th element of $\dot{\boldsymbol{\gamma}}$, for $j \in \{1, \ldots, (p+q+v+1)\}$, as

$$\frac{\partial \ell}{\partial \dot{\gamma}_j} = \sum_{t=m+1}^{n} \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \dot{\gamma}_j}.$$

Recall that $\eta_{1t} = g_1(\mu_t)$, which implies $d\mu_t/d\eta_{1t} = 1/g_1'(\mu_t)$, where prime denotes derivative. The derivative of $\ell_t(\mu_t, \phi_t)$ with respect to μ_t can be expressed as

$$\begin{aligned}
\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} &= (1+\phi_t)\log(Y_t) - (1+\phi_t)\log(1+Y_t) - (1+\phi_t) \\
&\quad \times \psi(\mu_t(1+\phi_t)) + (1+\phi_t)\psi(\mu_t(1+\phi_t)+\phi_t+2) \\
&= (1+\phi_t)\Bigg\{ \log\left(\frac{Y_t}{1+Y_t}\right) - [\psi(\mu_t(1+\phi_t)) \\
&\quad - \psi(\mu_t(1+\phi_t)+\phi_t+2)]\Bigg\} \\
&= (1+\phi_t)(Y_t^* - \mu_t^*), \qquad (3.2)
\end{aligned}$$

where $\psi(\cdot)$ is the digamma function, i.e., $\psi(x) = d\log(\Gamma(x))/dx$, $Y_t^* = \log(Y_t/(1+Y_t))$ and $\mu_t^* = \psi(\mu_t(1+\phi_t)) - \psi(\mu_t(1+\phi_t)+\phi_t+2)$. Let $\zeta_{1t} = (1+\phi_t)(Y_t^* - \mu_t^*)$. It follows that

$$\frac{\partial \ell}{\partial \dot{\gamma}_j} = \sum_{t=m+1}^{n} \frac{\zeta_{1t}}{g_1'(\mu_t)} \frac{\partial \eta_{1t}}{\partial \dot{\gamma}_j}.$$

The derivative of η_{1t} with respect to α_1 is

$$\frac{\partial \eta_{1t}}{\partial \alpha_1} = 1 + \sum_{j=1}^{q} \theta_j \frac{\partial r_{t-j}}{\partial \alpha_1} = 1 - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{1t-j}}{\partial \alpha_1}.$$

The derivative of η_{1t} with respect to a β_k, for $k \in \{1, \ldots, v\}$, is

$$\frac{\partial \eta_{1t}}{\partial \beta_k} = x_{tk} - \sum_{i=1}^{p} \varphi_i x_{(t-i)k} - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{1t-j}}{\partial \beta_k},$$

where x_{tk} is the k-th element of $\boldsymbol{x}_t$. The derivative of η_{1t} with respect to φ_l, for $l \in \{1, \ldots, p\}$, can be expressed as

$$\frac{\partial \eta_{1t}}{\partial \varphi_l} = g_1(Y_{t-1}) - \boldsymbol{x}_{t-1}'\boldsymbol{\beta} - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{1t-j}}{\partial \varphi_l}.$$

Finally, the derivative of η_{1t} with respect to θ_s, for $s \in \{1, \ldots, q\}$, is given by

$$\frac{\partial \eta_{1t}}{\partial \theta_s} = r_{t-s} - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{1t-j}}{\partial \theta_s}.$$

Note the recursion in the derivatives when there are moving average components in the model.

As for the parameters in the conditional precision submodel, we have that

$$\frac{\partial \ell}{\partial \ddot{\gamma}_i} = \sum_{t=m+1}^{n} \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{d\phi_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \ddot{\gamma}_i},$$

$i \in \{1, 2\}$. The first derivative of $\ell(\mu_t, \phi_t)$ with respect to ϕ_t is

$$\begin{aligned}
\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} =& \mu_t \log(Y_t) - (\mu_t + 1)\log(1 + Y_t) - \mu_t \psi(\mu_t(1 + \phi_t)) \\
& - \psi(\phi_t + 2) + (\mu_t + 1)\psi(\mu_t(1 + \phi_t) + \phi_t + 2) \\
=& \log\left(\frac{Y_t^{\mu_t}}{(1 + Y_t)^{1+\mu_t}}\right) - \mu_t[\psi(\mu_t(1 + \phi_t)) - \psi(\mu_t(1 + \phi_t) \\
& + \phi_t + 2)] + \psi(\mu_t(1 + \phi_t) + \phi_t + 2) - \psi(\phi_t + 2) \\
=& Y_t^{\dagger} - \mu_t\left(\mu_t^* - \frac{\Delta_t}{\mu_t}\right) = Y_t^{\dagger} - \mu_t^{\dagger}, \qquad (3.3)
\end{aligned}$$

where $Y_t^{\dagger} = \mu_t \log(Y_t) - (1 + \mu_t)\log(1 + Y_t)$ and $\mu_t^{\dagger} = \mu_t\,(\mu_t^* - \Delta_t/\mu_t)$, with $\Delta_t = \psi(\mu_t(1 + \phi_t) + \phi_t + 2) - \psi(\phi_t + 2)$.

Since $\eta_{2t} = g_2(\phi_t)$, it follows that $d\phi_t/d\eta_{2t} = 1/g_2'(\phi_t)$, and hence

$$\frac{\partial \ell}{\partial \ddot{\gamma}_i} = \sum_{t=m+1}^{n} \frac{\zeta_{2t}}{g_2'(\phi_t)} \frac{\partial \eta_{2t}}{\partial \ddot{\gamma}_i},$$

where $\zeta_{2t} = Y_t^{\dagger} - \mu_t^{\dagger}$.

The derivatives of η_{2t} with respect to the elements of $\ddot{\gamma}_i$ are

$$\frac{\partial \eta_{2t}}{\partial \alpha_2} = 1 \quad \text{and} \quad \frac{\partial \eta_{2t}}{\partial \delta} = z_{t-1}.$$

The conditional score vector is $\boldsymbol{U}(\boldsymbol{\gamma}) = (U_{\alpha_1}(\boldsymbol{\gamma}), \boldsymbol{U}_{\beta}(\boldsymbol{\gamma})^{\top}, \boldsymbol{U}_{\boldsymbol{\varphi}}(\boldsymbol{\gamma})^{\top}, \boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\gamma})^{\top}, U_{\alpha_2}(\boldsymbol{\gamma}), U_{\delta}(\boldsymbol{\gamma}))^{\top}$, where

$$\begin{aligned}
U_{\alpha_1}(\boldsymbol{\gamma}) =& \boldsymbol{\nu}^{\top}\boldsymbol{P_1}\boldsymbol{\zeta_1}, \quad \boldsymbol{U}_{\beta}(\boldsymbol{\gamma}) = \boldsymbol{G}^{\top}\boldsymbol{P_1}\boldsymbol{\zeta_1}, \quad \boldsymbol{U}_{\boldsymbol{\varphi}}(\boldsymbol{\gamma}) = \boldsymbol{H}^{\top}\boldsymbol{P_1}\boldsymbol{\zeta_1}, \\
\boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\gamma}) =& \boldsymbol{F}^{\top}\boldsymbol{P_1}\boldsymbol{\zeta_1}, \quad U_{\alpha_2}(\boldsymbol{\gamma}) = \boldsymbol{1_n}^{\top}\boldsymbol{P_2}\boldsymbol{\zeta_2} \quad \text{and} \quad U_{\delta}(\boldsymbol{\gamma}) = \boldsymbol{z}^{\top}\boldsymbol{P_2}\boldsymbol{\zeta_2}.
\end{aligned}$$

Here,

$$\begin{aligned}
&\boldsymbol{\zeta_1} = (\zeta_{1m+1}, \ldots, \zeta_{1n})^{\top}, \ \boldsymbol{\zeta_2} = (\zeta_{2m+1}, \ldots, \zeta_{2n})^{\top}, \ \boldsymbol{z} = (z_m, \ldots, z_{n-1})^{\top}, \\
&\boldsymbol{\nu} = \left(\frac{\partial \eta_{1m+1}}{\partial \alpha_1}, \ldots, \frac{\partial \eta_{1n}}{\partial \alpha_1}\right)^{\top}, \ \boldsymbol{P_1} = \operatorname{diag}\left\{\frac{1}{g_1'(\mu_{m+1})}, \ldots, \frac{1}{g_1'(\mu_n)}\right\}, \\
&\boldsymbol{P_2} = \operatorname{diag}\left\{\frac{1}{g_2'(\phi_{m+1})}, \ldots, \frac{1}{g_2'(\phi_n)}\right\}.
\end{aligned}$$

Additionally, $\boldsymbol{G}$ is an $(n-m)\times v$ matrix whose (a, j)-th element is $\partial \eta_{1a+m}/\partial \beta_j$, $\boldsymbol{H}$ is a matrix of order $(n - m) \times p$ with (a, j)-th element equal to $\partial \eta_{1a+m}/\partial \varphi_j$ and $\boldsymbol{F}$ is an $(n - m) \times q$ matrix with (a, j)-th element given by $\partial \eta_{1a+m}/\partial \theta_j$ and $\mathbf{1}_n$ is an $(n - m) \times 1$ vector of ones.

3.2 CONDITIONAL INFORMATION MATRIX

The conditional Fisher information matrix is obtained from the second-order log-likelihood derivatives. For the generalized BPARMA model, we have

$$\begin{aligned}\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\dot{\gamma}_i\partial\dot{\gamma}_j} &= \sum_{t=m+1}^{n}\frac{\partial}{\partial\mu_t}\left(\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{d\mu_t}{d\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\dot{\gamma}_j}\right)\frac{d\mu_t}{d\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\dot{\gamma}_i}\\ &= \sum_{t=m+1}^{n}\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\mu_t^2}\frac{d\mu_t}{d\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\dot{\gamma}_j}+\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial}{\partial\mu_t}\left(\frac{d\mu_t}{d\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\dot{\gamma}_j}\right)\\ &\quad\times\frac{d\mu_t}{d\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\dot{\gamma}_i},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\ddot{\gamma}_i\partial\ddot{\gamma}_j} &= \sum_{t=m+1}^{n}\frac{\partial}{\partial\phi_t}\left(\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{d\phi_t}{d\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\ddot{\gamma}_j}\right)\frac{d\phi_t}{d\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\ddot{\gamma}_i}\\ &= \sum_{t=m+1}^{n}\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\phi_t^2}\frac{d\phi_t}{d\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\ddot{\gamma}_j}+\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{\partial}{\partial\phi_t}\left(\frac{d\phi_t}{d\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\ddot{\gamma}_j}\right)\\ &\quad\times\frac{d\phi_t}{d\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\ddot{\gamma}_i}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\dot{\gamma}_i\partial\ddot{\gamma}_j} &= \sum_{t=m+1}^{n}\frac{\partial}{\partial\phi_t}\left(\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{d\mu_t}{d\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\ddot{\gamma}_j}\right)\frac{d\phi_t}{d\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\dot{\gamma}_i}\\ &= \sum_{t=m+1}^{n}\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\phi_t\partial\mu_t}\frac{d\mu_t}{d\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\ddot{\gamma}_j}+\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial}{\partial\phi_t}\left(\frac{d\mu_t}{d\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\ddot{\gamma}_j}\right)\\ &\quad\times\frac{d\phi_t}{d\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\dot{\gamma}_i}.\end{aligned}$$

Notice that $\partial\ell_t(\mu_t,\phi_t)/\partial\mu_t$ and $\partial\ell_t(\mu_t,\phi_t)/\partial\phi_t$ are given in the Equations (3.2) and (3.3), respectively. Since $\mathbb{E}(Y_t^*)=\mu_t^*$ and $\mathbb{E}(Y_t^\dagger)=\mu_t^\dagger$, then $\mathbb{E}\left(\partial\ell_t(\mu_t,\phi_t)/\partial\mu_t|\mathscr{F}_{t-1}\right)=0$ and $\mathbb{E}\left(\partial\ell_t(\mu_t,\phi_t)/\partial\phi_t|\mathscr{F}_{t-1}\right)=0$. Thus,

$$\begin{aligned}\mathbb{E}\left(\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\dot{\gamma}_i\partial\dot{\gamma}_j}\middle|\mathscr{F}_{t-1}\right) &= \sum_{t=m+1}^{n}\left[\mathbb{E}\left(\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\mu_t^2}\middle|\mathscr{F}_{t-1}\right)\right]\\ &\quad\times\left(\frac{d\mu_t}{d\eta_{1t}}\right)^2\frac{\partial\eta_{1t}}{\partial\dot{\gamma}_j}\frac{\partial\eta_{1t}}{\partial\dot{\gamma}_i},\end{aligned}\tag{3.4}$$

$$\begin{aligned}\mathbb{E}\left(\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\ddot{\gamma}_i\partial\ddot{\gamma}_j}\middle|\mathscr{F}_{t-1}\right) &= \sum_{t=m+1}^{n}\left[\mathbb{E}\left(\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\phi_t^2}\middle|\mathscr{F}_{t-1}\right)\right]\\ &\quad\times\left(\frac{d\phi_t}{d\eta_{2t}}\right)^2\frac{\partial\eta_{2t}}{\partial\ddot{\gamma}_j}\frac{\partial\eta_{2t}}{\partial\ddot{\gamma}_i}\end{aligned}\tag{3.5}$$

and

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\dot{\gamma}_i\partial\ddot{\gamma}_j}\Big|\mathscr{F}_{t-1}\right) = \sum_{t=m+1}^{n}\left[\mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\phi_t\partial\mu_t}\Big|\mathscr{F}_{t-1}\right)\right] \times \left(\frac{d\mu_t}{d\eta_{1t}}\right)\left(\frac{d\phi_t}{d\eta_{2t}}\right)\frac{\partial\eta_{1t}}{\partial\ddot{\gamma}_j}\frac{\partial\eta_{2t}}{\partial\dot{\gamma}_i}. \tag{3.6}$$

From (3.2) we obtain

$$\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\mu_t^2} = -(1+\phi_t)^2[\psi'(\mu_t(1+\phi_t)) - \psi'(\mu_t(1+\phi_t)+\phi_t+2)], \tag{3.7}$$

where $\psi'(\cdot)$ is the trigamma function, i.e., $\psi'(x) = d\psi(x)/dx$.

Finally, plugging (3.7) into (3.4) we obtain

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\dot{\gamma}_i\partial\dot{\gamma}_j}\Big|\mathscr{F}_{t-1}\right) = -\sum_{t=m+1}^{n}\frac{c_t}{(g_1'(\mu_t))^2}\frac{\partial\eta_{1t}}{\partial\dot{\gamma}_j}\frac{\partial\eta_{1t}}{\partial\dot{\gamma}_i},$$

where $c_t = (1+\phi_t)^2[\psi'(\mu_t(1+\phi_t)) - \psi'(\mu_t(1+\phi_t)+\phi_t+2)]$.

The second derivative of (3.3) with respect to ϕ_t can expressed as

$$\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\phi_t^2} = -\mu_t^2\psi'(\mu_t(1+\phi_t)) + (1+\mu_t)^2\psi'(\mu_t(1+\phi_t)+\phi_t+2) - \psi'(\phi_t+2). \tag{3.8}$$

Plugging (3.8) into (3.5), we obtain

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\ddot{\gamma}_i\partial\ddot{\gamma}_j}\Big|\mathscr{F}_{t-1}\right) = -\sum_{t=m+1}^{n}\frac{d_t}{(g_2'(\phi_t))^2}\frac{\partial\eta_{2t}}{\partial\ddot{\gamma}_j}\frac{\partial\eta_{2t}}{\partial\ddot{\gamma}_i},$$

where $d_t = \mu_t^2\psi'(\mu_t(1+\phi_t)) - (1+\mu_t)^2\psi'(\mu_t(1+\phi_t)+\phi_t+2) + \psi'(\phi_t+2)$.

The cross-derivative can be easily obtained by differentiating Equation (3.3) with respect to μ_t:

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

Notice that $Y_t^*|\mathscr{F}_{t-1}$ is beta-distributed, and hence $\mathbb{E}(Y_t^*|\mathscr{F}_{t-1}) = \mu_t^*$. It follows that

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\dot{\gamma}_i\partial\ddot{\gamma}_j}\Big|\mathscr{F}_{t-1}\right) = \sum_{t=m+1}^{n}\frac{t_t}{g_1'(\mu_t)g_2'(\phi_t)}\frac{\partial\eta_{1t}}{\partial\dot{\gamma}_j}\frac{\partial\eta_{2t}}{\partial\dot{\gamma}_i},$$

where $t_t = (1+\phi_t)\{\psi'(\mu_t(1+\phi_t)+\phi_t+2) + \mu_t[\psi'(\mu_t(1+\phi_t)+\phi_t+2) - \psi'(\mu_t(1+\phi_t))]\}$.

Let $\boldsymbol{C} = \mathrm{diag}\{c_{m+1}, \ldots, c_n\}$, $\boldsymbol{D} = \mathrm{diag}\{d_{m+1}, \ldots, d_n\}$ and $\boldsymbol{T} = \mathrm{diag}\{t_{m+1}, \ldots, t_n\}$. Fisher's information matrix for $\boldsymbol{\gamma}$ is

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

where $K_{\alpha_1,\alpha_1} = -\boldsymbol{\nu}^\top \boldsymbol{C}\boldsymbol{P}_1^2\boldsymbol{\nu}$, $\boldsymbol{K}_{\alpha_1,\boldsymbol{\beta}} = \boldsymbol{K}_{\boldsymbol{\beta},\alpha_1}^\top = -\boldsymbol{\nu}'\boldsymbol{C}\boldsymbol{P}_1^2\boldsymbol{G}$, $\boldsymbol{K}_{\alpha_1,\boldsymbol{\varphi}} = \boldsymbol{K}_{\boldsymbol{\varphi},\alpha_1}^\top = -\boldsymbol{\nu}^\top \boldsymbol{C}\boldsymbol{P}_1^2\boldsymbol{H}$, $\boldsymbol{K}_{\alpha_1,\boldsymbol{\theta}} = \boldsymbol{K}_{\boldsymbol{\theta},\alpha_1}^\top = -\boldsymbol{\nu}^\top \boldsymbol{C}\boldsymbol{P}_1^2\boldsymbol{F}$, $K_{\alpha_1,\alpha_2} = K_{\alpha_2,\alpha_1} = -\boldsymbol{\nu}^\top \boldsymbol{P}_2\boldsymbol{T}\boldsymbol{P}_1\mathbf{1}_n$, $\boldsymbol{K}_{\alpha_1,\boldsymbol{\delta}} = \boldsymbol{K}_{\delta,\alpha_1} = -\boldsymbol{\nu}^\top \boldsymbol{P}_2\boldsymbol{T}\boldsymbol{P}_1\boldsymbol{z}$, $\boldsymbol{K}_{\boldsymbol{\beta},\boldsymbol{\beta}} = -\boldsymbol{G}^\top \boldsymbol{C}\boldsymbol{P}_1^2\boldsymbol{G}$, $\boldsymbol{K}_{\boldsymbol{\beta},\boldsymbol{\varphi}} = \boldsymbol{K}_{\boldsymbol{\varphi},\boldsymbol{\beta}}^\top = -\boldsymbol{G}^\top \boldsymbol{C}\boldsymbol{P}_1^2\boldsymbol{H}$, $\boldsymbol{K}_{\boldsymbol{\beta},\boldsymbol{\theta}} = \boldsymbol{K}_{\boldsymbol{\theta},\boldsymbol{\beta}}^\top = -\boldsymbol{G}^\top \boldsymbol{C}\boldsymbol{P}_1^2\boldsymbol{F}$, $\boldsymbol{K}_{\boldsymbol{\beta},\alpha_2} = \boldsymbol{K}_{\alpha_2,\boldsymbol{\beta}}^\top = -\boldsymbol{G}^\top \boldsymbol{P}_2\boldsymbol{T}\boldsymbol{P}_1\mathbf{1}_n$, $\boldsymbol{K}_{\boldsymbol{\beta},\delta} = \boldsymbol{K}_{\delta,\boldsymbol{\beta}}^\top = -\boldsymbol{G}^\top \boldsymbol{P}_2\boldsymbol{T}\boldsymbol{P}_1\boldsymbol{z}$, $\boldsymbol{K}_{\boldsymbol{\varphi},\boldsymbol{\varphi}} = -\boldsymbol{H}^\top \boldsymbol{C}\boldsymbol{P}_1^2\boldsymbol{H}$, $\boldsymbol{K}_{\boldsymbol{\varphi},\boldsymbol{\theta}} = \boldsymbol{K}_{\boldsymbol{\theta},\boldsymbol{\varphi}}^\top = -\boldsymbol{H}^\top \boldsymbol{C}\boldsymbol{P}_1^2\boldsymbol{F}$, $\boldsymbol{K}_{\boldsymbol{\varphi},\alpha_2} = \boldsymbol{K}_{\alpha_2,\boldsymbol{\varphi}}^\top = -\boldsymbol{H}^\top \boldsymbol{P}_2\boldsymbol{T}\boldsymbol{P}_1\mathbf{1}_n$, $\boldsymbol{K}_{\boldsymbol{\varphi},\boldsymbol{\delta}} = \boldsymbol{K}_{\delta,\boldsymbol{\varphi}}^\top = -\boldsymbol{H}^\top \boldsymbol{P}_2\boldsymbol{T}\boldsymbol{P}_1\boldsymbol{z}$, $\boldsymbol{K}_{\boldsymbol{\theta},\boldsymbol{\theta}} = -\boldsymbol{F}^\top \boldsymbol{C}\boldsymbol{P}_1^2\boldsymbol{F}$, $\boldsymbol{K}_{\boldsymbol{\theta},\alpha_2} = \boldsymbol{K}_{\alpha_2,\boldsymbol{\theta}}^\top = -\boldsymbol{F}^\top \boldsymbol{P}_2\boldsymbol{T}\boldsymbol{P}_1\mathbf{1}_n$, $\boldsymbol{K}_{\boldsymbol{\theta},\delta} = \boldsymbol{K}_{\delta,\boldsymbol{\theta}}^\top = -\boldsymbol{F}^\top \boldsymbol{P}_2\boldsymbol{T}\boldsymbol{P}_1\boldsymbol{z}$, $K_{\alpha_2,\alpha_2} = -\mathbf{1}_n^\top \boldsymbol{D}\boldsymbol{P}_2^2\mathbf{1}_n$, $K_{\alpha_2,\delta} = K_{\delta,\alpha_2} = -\mathbf{1}_n^\top \boldsymbol{D}\boldsymbol{P}_2^2\boldsymbol{z}$ and $K_{\delta,\delta} = -\boldsymbol{z}^\top \boldsymbol{D}\boldsymbol{P}_2^2\boldsymbol{z}$, tr being the trace operator.

Under some regularity conditions, the CMLE is consistent and asymptotically normally distributed; for details, see Andersen (1970). That is, in large samples $\hat{\boldsymbol{\gamma}}$ is approximately normally distributed with mean $\boldsymbol{\gamma}$ and covariance matrix $\boldsymbol{K}^{-1}$:

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\varphi}} \\ \hat{\boldsymbol{\theta}} \\ \hat{\alpha}_2 \\ \hat{\delta} \end{pmatrix} \overset{\text{approx}}{\sim} \mathcal{N}_{p+q+v+3}\left(\begin{pmatrix} \alpha_1 \\ \boldsymbol{\beta} \\ \boldsymbol{\varphi} \\ \boldsymbol{\theta} \\ \alpha_2 \\ \delta \end{pmatrix}, \boldsymbol{K}^{-1}(\boldsymbol{\gamma}) \right),$$

where $\hat{\alpha}_1$, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\varphi}}$, $\hat{\boldsymbol{\theta}}$, $\hat{\alpha}_2$ and $\hat{\delta}$ are the CMLEs of α_1, $\boldsymbol{\beta}$, $\boldsymbol{\varphi}$, $\boldsymbol{\theta}$, α_2 and δ, respectively.

3.3 CONFIDENCE INTERVALS AND HYPOTHESIS TESTING

Let γ_k be the k-th element of the parameter vector $\boldsymbol{\gamma}$, $\hat{\gamma}_k$ be the k-th component of $\hat{\boldsymbol{\gamma}}$, computed using a sample path of size n from the generalized BPARMA(p, q) model, and $K(\hat{\boldsymbol{\gamma}})^{kk}$ be the k-th diagonal element of the inverse of the conditional Fisher information matrix evaluated at $\hat{\boldsymbol{\gamma}}$, $\boldsymbol{K}(\hat{\boldsymbol{\gamma}})^{-1}$. We have that

$$\frac{\hat{\gamma}_k - \gamma_k}{\sqrt{K(\hat{\boldsymbol{\gamma}})^{kk}}} \overset{\text{approx}}{\sim} \mathcal{N}(0, 1).$$

Using this result, it is possible to construct a $100(1-\varepsilon)\%$ asymptotic confidence interval for γ_k, $0 < \varepsilon < 1/2$, as

$$\left[\hat{\gamma}_k - z_{1-\varepsilon/2}\sqrt{K(\hat{\boldsymbol{\gamma}})^{kk}}, \hat{\gamma}_k + z_{1-\varepsilon/2}\sqrt{K(\hat{\boldsymbol{\gamma}})^{kk}}\right],$$

where $z_{1-\varepsilon/2}$ is the $1-\varepsilon/2$ standard normal quantile. The probability that this interval will contain γ_k approaches $1-\varepsilon$ as $n \to \infty$.

The asymptotic normality of $\hat{\boldsymbol{\gamma}}$ can also be used to perform test inferences on the model's parameters. Suppose we wish to test $\mathscr{H}_0 : \gamma_k = \gamma_k^{(0)}$ vs $\mathscr{H}_1 : \gamma_k \neq \gamma_0$, where $\gamma_k^{(0)}$ is a given scalar. We may use the z test statistic:

$$z = \frac{\hat{\gamma}_k - \gamma_k^{(0)}}{\sqrt{K(\hat{\boldsymbol{\gamma}})^{kk}}}.$$

Under the null hypothesis, the asymptotic distribution of z is standard normal. The null hypothesis is thus rejected at significance level ε if $|z| > z_{1-\varepsilon/2}$. The test's type I error probability will converge to ε as $n \to \infty$.

It is also possible to test more than one restriction using the likelihood ratio, score and Wald tests. The closed-form expressions we provide for the model's score function and Fisher's information matrix can be used for computing score and Wald test statistics. Let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \boldsymbol{\gamma}_2^\top)^\top$, where $\dim(\boldsymbol{\gamma}_1) = s$ and $\dim(\boldsymbol{\gamma}_2) = p + q + v + 3 - s$, and suppose we wish to test $\mathscr{H}_0 : \boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_1^{(0)}$ vs $\mathscr{H}_1 : \boldsymbol{\gamma}_1 \neq \boldsymbol{\gamma}_1^{(0)}$, where $\boldsymbol{\gamma}_1^{(0)} \in \mathbb{R}^s$ is a given vector. Also, let $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_1^\top, \hat{\boldsymbol{\gamma}}_2^\top)^\top$ and $\tilde{\boldsymbol{\gamma}} = (\boldsymbol{\gamma}_1^{(0)^\top}, \tilde{\boldsymbol{\gamma}}_2^\top)^\top$ be, respectively, the unrestricted and restricted conditional maximum likelihood estimators of $\boldsymbol{\gamma}$. The likelihood ratio test statistic can be easily computed as $\omega_1 = 2(\ell(\hat{\boldsymbol{\gamma}}) - \ell(\tilde{\boldsymbol{\gamma}}))$. Rao's score and Wald's test statistics are, respectively, $\omega_2 = \boldsymbol{U}(\tilde{\boldsymbol{\gamma}})^\top \boldsymbol{K}^{-1}(\tilde{\boldsymbol{\gamma}}) \boldsymbol{U}(\tilde{\boldsymbol{\gamma}})$ and $\omega_3 = (\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1^{(0)})^\top (\boldsymbol{K}^{\gamma_1\gamma_1}(\hat{\boldsymbol{\gamma}}))^{-1} (\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1^{(0)})$, where $K^{\gamma_1\gamma_1}$ is the $s \times s$ matrix formed out of the rows and columns of $\boldsymbol{K}^{-1}$ that correspond to $\boldsymbol{\gamma}_1$. Under $\mathscr{H}_0$, the three test statistics are asymptotically χ^2_s distributed. Test inferences are reached using approximate critical values from this distribution.

4 MODEL SELECTION, DIAGNOSTIC TESTS, SAMPLE PATHS AND FORECASTING

In what follows, we will outline generalized BPARMA model selection, diagnostic analysis and forecasting. In particular, the focus of diagnostic analysis is to determine whether the estimated model is capable of fully capturing the dynamics present in the data. Serial correlation in the residuals is taken as evidence of model misspecification.

4.1 MODEL SELECTION

Generalized BPARMA model selection can be carried out using information criteria. Such criteria are widely used to compare model fits and select a model from a set of candidate models. The goal is to find a parsimonious model, that is, a model that fits the data well and has a small number of parameters. Several information criteria are available in the literature, the most frequently used being Akaike's Information Criterion (AIC) (Akaike (1974)) and the Bayesian Information Criterion (BIC) (Schwarz (1978)). They are based on the maximized conditional log-likelihood function and incorporate a penalty term for model augmentation: AIC $= -2\ell(\hat{\boldsymbol{\gamma}}) + 2(p+q+v+3)$ and BIC $= -2\ell(\hat{\boldsymbol{\gamma}}) + \log(n)(p+q+v+3)$. The model with the lowest AIC or BIC is selected.

4.2 DEVIANCE

Goodness-of-fit assessment is an important step in diagnostic analysis. It can be based on the deviance statistic, which is defined as twice the difference between the conditional log-likelihood of the saturated model, where $\tilde{\mu}_t = Y_t$, and the fitted model:

$$D = 2(\tilde{\ell} - \hat{\ell}),$$

where $\tilde{\ell} = \sum_{t=m+1}^{n} \ell(Y_t, \hat{\phi}_t)$ and $\hat{\ell} = \sum_{t=m+1}^{n} \ell(\hat{\mu}_t, \hat{\phi}_t)$, $\hat{\mu}_t$ and $\hat{\phi}_t$ being obtained by evaluating Equations (2.2) and (2.3), respectively, at the conditional maximum likelihood estimates. Under correct model specification, D is asymptotically distributed as $\chi^2_{n-(p+q+v+m+3)}$; see Benjamin, Rigby and Stasinopoulos (2003) and Fokianos and Kedem (2004). A rule-of-thumb is to conclude that model specification is in error if $D/[n-(p+q+v+m+3)] > 1$; see Myers et al. (2012).

4.3 RESIDUAL ANALYSIS

Residual analysis is an important step in the validation of a statistical model. The main goal is to verify whether the estimated model yields a good data fit. There are several residuals that can be used to that end. We will use the Pearson residual; see Leiva et al. (2014) and Santos-Neto et al. (2016). The Pearson residual is based on the difference between the observed value and the fitted value. The t-th generalized $\mathrm{BPARMA}(p, q)$ Pearson residual is

$$\hat{r}_t^P = \frac{Y_t - \hat{\mu}_t}{\sqrt{\hat{\mu}_t(1+\hat{\mu}_t)/\hat{\phi}}}.$$

4.4 PORTMANTEAU TESTS

The goodness-of-fit of the estimated model can also be assessed using portmanteau-type tests. Here, one tests the joint nullity of the first k residual autocorrelations. The fitted model is considered to provide a good representation of the data if such autocorrelations are jointly negligible. Rejection of the null hypothesis implies rejection of the fitted model. The null and alternative hypotheses are

$$\mathscr{H}_0 : \rho_1 = \cdots = \rho_k = 0$$
$$\mathscr{H}_1 : \max(|\rho_1|, \ldots, |\rho_k|) > 0,$$

where ρ_l is the l-th residual autocorrelation.

The most commonly used test statistic was proposed by Ljung and Box (1978). It is given by

$$Q_{LB} = n(n+2)\sum_{l=1}^{k} \frac{\hat{\rho}_l^2}{n-l},$$

where

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^{n} \hat{r}_t^P \hat{r}_{t-l}^P}{\sum_{t=1}^{n} (\hat{r}_t^P)^2}, \quad l \in \{1, \ldots, k\}.$$

An alternative test statistic, denoted Q_M, was proposed by Monti (1994). It is as the above test statistic but uses partial residual autocorrelations instead of residual autocorrelations. Under the null hypothesis, both test statistics are asymptotically distributed as χ^2_{k-p-q}. The tests are carried out using critical values from this asymptotic null distribution.

4.5 SAMPLE PATHS

Generalized $\mathrm{BPARMA}(p,q)$ sample paths can be easily obtained using the algorithm that follows.

Algorithm 1.

1. Specify the values of the parameters α_1, $\boldsymbol{\beta}$, $\boldsymbol{\varphi}$, $\boldsymbol{\theta}$, α_2, δ and the sample size n;
2. Set $r_t = 0$ and $Y_t = 0$ for $t \in \{1, \ldots, m\}$;
3. For $t \in \{m+1, \ldots, n\}$, recursively compute

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

   and

$$\phi_t = g_2^{-1}\left(\alpha_2 + \delta z_{t-1}\right);$$

4. Randomly draw Y_t from $BP(\mu_t, \phi_t)$, for $t \in \{m+1, \ldots, n\}$.

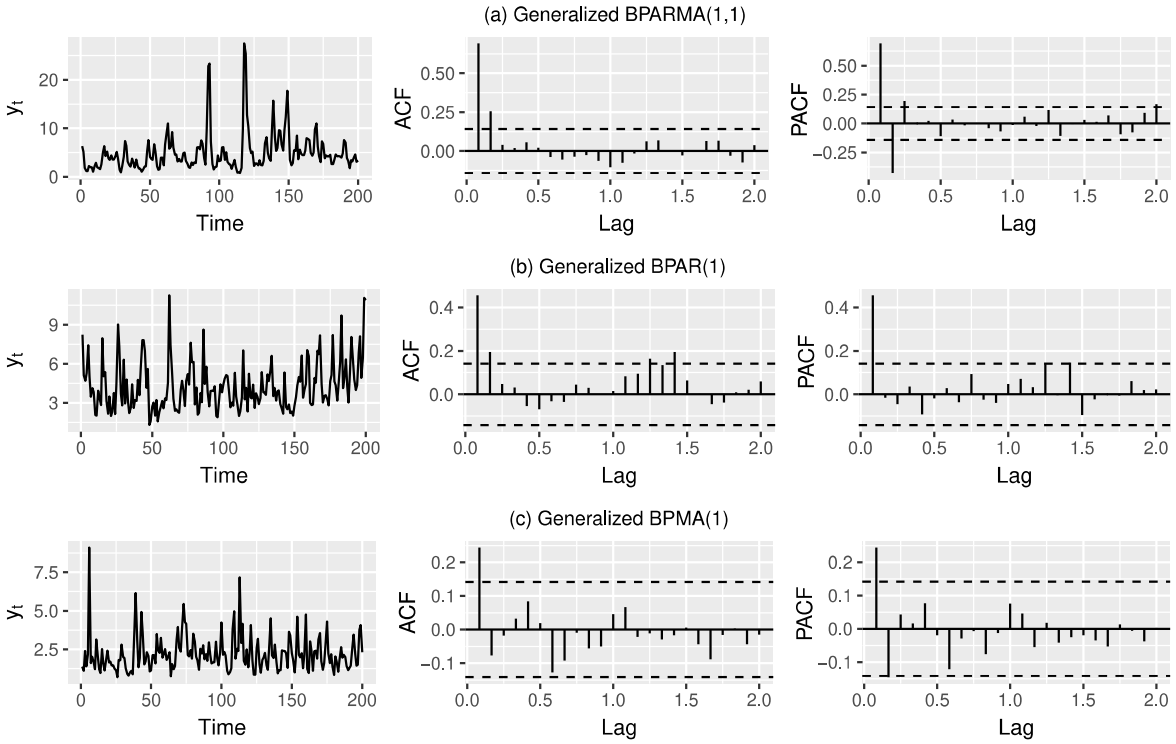
Figure 2 contains monthly sample paths of length 200 from three generalized BPARMA models. For each time series, we also present the corresponding correlogram (sample autocorrelation function – ACF) and partial correlogram (sample partial autocorrelation function – PACF). The sample paths were obtained from the following models: (i) generalized $\mathrm{BPARMA}(1,1)$ with $\alpha_1 = 0.8$, $\varphi = 0.5$, $\theta = 0.5$, $\alpha_2 = 2.0$ and $\delta = -0.1$, (ii) generalized $\mathrm{BPARMA}(1,0)$ with $\alpha_1 = 0.8$, $\varphi = 0.5$, $\alpha_2 = 2.0$ and $\delta = -0.1$, and (iii) generalized $\mathrm{BPARMA}(0,1)$ with $\alpha_1 = 0.8$, $\theta = 0.5$, $\alpha_2 = 2.0$ and $\delta = -0.1$. The dashed horizontal lines correspond to $\pm 1.96/\sqrt{n-m}$.

4.6 FORECASTING

Fitted values and out-of-sample forecasts can be easily obtained from a fitted generalized BPARMA(p,q) model. The fitted values are $\hat{\mu}_t$, $t \in \{m+1, \ldots, n\}$. They are obtained by replacing the elements of the parameter vector, $\boldsymbol{\gamma}$, with its CMLE, $\hat{\boldsymbol{\gamma}}$, and using the fitted model's residuals, which are taken to be zero for $t \in \{1, \ldots, m\}$ and $\hat{r}_t = g_1(Y_t) - g_1(\hat{\mu}_t)$ for $t \in \{m+1, \ldots, n\}$. The t-th fitted value is

$$\hat{\mu}_t = g_1^{-1}\left(\hat{\alpha}_1 + \boldsymbol{x_t}^\top\hat{\boldsymbol{\beta}} + \sum_{i=1}^{p}\hat{\varphi}_i(g_1(y_{t-i}) - \boldsymbol{x_{t-i}}^\top\hat{\boldsymbol{\beta}}) + \sum_{j=1}^{q}\hat{\theta}_j\hat{r}_{t-j}\right), \tag{4.1}$$

Figure 2 – Simulated BPARMA time series with correlograms and partial correlograms.



$t \in \{m+1, \ldots, n\}$.

For $h \in \{1, 2, \ldots\}$, the h-steps-ahead out-of-sample forecast, i.e., the forecast of Y_{n+h} made at time n, is given by

$$\hat{\mu}_{n+h} = g_1^{-1}\Bigg(\hat{\alpha}_1 + \boldsymbol{x}_{\boldsymbol{n+h}}^{\top}\hat{\boldsymbol{\beta}} + \sum_{i=1}^{p}\hat{\varphi}_i([g_1(y_{n+h-i})] - \boldsymbol{x}_{\boldsymbol{n+h-i}}^{\top}\hat{\boldsymbol{\beta}}) + \sum_{j=1}^{q}\hat{\theta}_j\hat{r}_{n+h-j}\Bigg), \tag{4.2}$$

where

$$[g_1(Y_t)] = \begin{cases} g_1(\hat{\mu}_t), & \text{if} \quad t > n, \\ g_1(Y_t), & \text{if} \quad t \leq n. \end{cases}$$

When the model includes covariates, it is necessary to use their next h values when computing the out-of-sample forecast in Equation (4.2).

5 SIMULATION EVIDENCE

In this chapter, we will present Monte Carlo simulation evidence on the finite sample performance of the conditional maximum likelihood estimators of the parameters that index the generalized BPARMA(p, q) model. The sample sizes are $n \in \{50, 100, 200, 500\}$. In each Monte Carlo replication, we generate a sample path of size $n+1000$ and then discard the first 1000 observations in order to minimize dependence on initial values, i.e., we perform a burn-in of size 1000. This is done to reduce dependence on initial values. Generalized BPARMA sample paths are obtained according to Algorithm 1. Beta prime random draws are obtained from beta random draws, i.e., we generate B_t from the beta distribution with parameters $\mu_t(1+\phi)$ and $\phi+2$, and then compute $Y_t = B_t/(1-B_t)$. Parameter estimation is carried out using the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm with box constraints and analytical first derivatives (L-BFGS-B algorithm). The following restrictions are imposed on the estimation process: $\alpha_2 > 0$ and $\delta < 0$. The starting values are as follows: $n^{-1}\sum_{t=1}^{n} g_1(Y_t)$ is used as starting value for α_1, all autoregressive and moving average parameters are set to zero, the initial value for α_2 is 3 and -1 is used as the initial value for δ. The number of Monte Carlo replications is 5,000. All simulations were performed using the statistical computing environment R; see R Core Team (2023). There was no convergence failure.

At the outset, data generation is carried out from the generalized $\mathrm{BPARMA}(1,1)$ model with log link function for the mean and conditional precision submodels and $\alpha_1 = 0.1$, $\varphi_1 = 0.7$, $\theta_1 = 0.2$, $\alpha_2 = 6.0$ and $\delta = -0.7$. Table 1 contains the mean estimates ('mean'), the standard derivation of all estimates ('SD'), and the estimated bias ('bias') and the mean squared error ('MSE') of each estimator. It is noteworthy that the biases and mean squared errors of all estimators diminish as the sample size increases, as expected. Overall, the estimates are quite accurate. For instance, when $n = 500$ the average estimate of φ_1 (θ_1) [δ] is 0.6930 (0.2036) [-0.7098], the true parameter value being 0.7 (0.2) [-0.7]. We note, however, that in all sample sizes there is slight underestimation of φ and slight overestimation of θ. Overall, we conclude that the generalized BPARMA parameters are reliably estimated by the conditional maximum likelihood method, especially when the sample size is not small.

Tables 2 and 3 contain, respectively, simulation results obtained using as the data generating process (i) the generalized BPARMA$(1,0)$ model with $\alpha_1 = 0.1$, $\varphi_1 = 0.7$, $\alpha_2 = 6.0$ and $\delta = -0.7$ and (ii) the generalized BPARMA$(0,1)$ model with $\alpha_1 = 0.1$, $\theta_1 = 0.2$, $\alpha_2 = 6.0$

and $\delta = -0.7$. As in previous simulations, we use the log link in the mean and precision submodels. Again, conditional maximum likelihood point estimation works as expected: the estimators' biases and mean squared errors approach zero as the sample size increases. The biases of $\hat{\varphi}_1$ and $\hat{\theta}_2$ in the autoregressive and moving average models, respectively, are slightly negative.

Table 1 – Mean value, standard deviation (SD), bias and mean squared error (MSE) of the conditional maximum likelihood estimators, generalized BPARMA$(1,1)$ model.

		Parameter				
	n	α_1	φ_1	θ_1	α_2	δ
		0.1	0.7	0.2	6.0	-0.7
Mean	50	0.1252	0.6204	0.2459	6.5262	-1.1101
	100	0.1130	0.6592	0.2241	6.2798	-0.9248
	200	0.1064	0.6798	0.2110	6.1101	-0.7861
	500	0.1022	0.6930	0.2036	6.0193	-0.7098
SD	50	0.0558	0.1577	0.1966	1.5205	1.4674
	100	0.0349	0.0996	0.1310	1.0993	1.0704
	200	0.0240	0.0684	0.0909	0.7893	0.7766
	500	0.0141	0.0406	0.0556	0.5595	0.5513
Bias	50	0.0252	-0.0796	0.0459	0.5262	-0.4101
	100	0.0130	-0.0408	0.0241	0.2798	-0.2248
	200	0.0064	-0.0202	0.0110	0.1101	-0.0861
	500	0.0022	-0.0070	0.0036	0.0193	-0.0098
MSE	50	0.0038	0.0312	0.0408	2.5887	2.3214
	100	0.0014	0.0116	0.0177	1.2866	1.1962
	200	0.0006	0.0051	0.0084	0.6351	0.6105
	500	0.0002	0.0017	0.0031	0.3134	0.3040

Table 2 – Mean value, standard deviation (SD), bias and mean squared error (MSE) of the conditional maximum likelihood estimators, generalized BPARMA$(1,0)$ model.

		Parameter			
	n	α_1	φ_1	α_2	δ
		0.1	0.7	6.0	-0.7
Mean	50	0.1204	0.6339	6.4669	-1.0763
	100	0.1102	0.6678	6.2573	-0.9111
	200	0.1048	0.6845	6.0981	-0.7787
	500	0.1018	0.6942	6.0199	-0.7122
SD	50	0.0410	0.1178	1.4562	1.4029
	100	0.0272	0.0780	1.0599	1.0313
	200	0.0184	0.0529	0.7758	0.7638
	500	0.0110	0.0321	0.5571	0.5481
Bias	50	0.0204	-0.0661	0.4669	-0.3763
	100	0.0102	-0.0322	0.2573	-0.2111
	200	0.0048	-0.0155	0.0981	-0.0787
	500	0.0018	-0.0058	0.0199	-0.0122
MSE	50	0.0021	0.0182	2.3386	2.1098
	100	0.0008	0.0071	1.1896	1.1081
	200	0.0004	0.0030	0.6115	0.5896
	500	0.0001	0.0011	0.3107	0.3005

Table 3 – Mean value, standard deviation (SD), bias and mean squared error (MSE) of the conditional maximum likelihood estimators, generalized BPARMA$(0,1)$ model.

		Parameter			
	n	α_1	θ_1	α_2	δ
		0.1	0.2	6.0	-0.7
Mean	50	0.0996	0.1890	6.5275	-1.1431
	100	0.0997	0.1956	6.2507	-0.9081
	200	0.1000	0.1969	6.0928	-0.7732
	500	0.1001	0.1986	6.0253	-0.7178
SD	50	0.0168	0.1676	1.4537	1.3962
	100	0.0122	0.1071	0.9694	0.9392
	200	0.0083	0.0711	0.6913	0.6750
	500	0.0052	0.0448	0.4775	0.4698
Bias	50	-0.0004	-0.0110	0.5275	-0.4431
	100	-0.0003	-0.0044	0.2507	-0.2081
	200	0.0000	-0.0031	0.0928	-0.0732
	500	0.0001	-0.0014	0.0253	-0.0178
MSE	50	0.0003	0.0282	2.3914	2.1459
	100	0.0001	0.0115	1.0025	0.9254
	200	0.0001	0.0051	0.4865	0.4610
	500	0.0000	0.0020	0.2286	0.2210

6 HYDROELECTRIC RESERVOIR WATER FLOW MODELING

Next, we will model water flows in reservoirs of two Brazilian hydroelectric power plants. We consider both inflow and outflow. The interest lies in forecasting future values of the two time series. Such forecasts assist the process of optimizing water use.

6.1 CACONDE WATER RESERVOIR INFLOW

In the following we will model and forecast a hydro-environmental time series. The interest lies in modeling the time evolution of the affluent flow of the Caconde reservoir. The affluent flow of a reservoir is the quantity of water that enters it over a specific period of time. This water input can originate from various sources, such as rivers, streams, precipitation (rain), among others. Affluent flow is an important indicator as it allows one to know the volume of water in the reservoir and thus properly manage the available water resources.

Formerly known as the Graminha Dam, the Caconde reservoir is owned by the Caconde Hydroelectric Power Plant and is currently operated by AES Brasil, a subsidiary of AES Corporation, one of the largest energy companies in the United States. Built to be a storage reservoir for the regulation of flow and the generation of electricity, it began operating in 1966. Located on the Pardo River in the municipality of Caconde, São Paulo, Brazil, the plant has an installed capacity of 80.4 megawatts. The reservoir of the hydroelectric plant, which operates at the river level, has an area of 31 square kilometers and a capacity of 555 billion liters of water. The useful volume of the reservoir is 504 billion liters.

We use data on the Caconde reservoir affluent flow from January 2000 to July 2023. The data are expressed in cubic meters per second (m^3/s). The sample size is $n = 283$ and the data were obtained through the Brazilian National Electric System Operator (Operador Nacional do Sistema Elétrico – ONS, <http://www.ons.org.br>). The final six observations (February 2023 through July 2023) were removed from the data prior to modeling. They were reserved for forecasting evaluation. The effective sample is thus $n = 277$.

Table 4 – Descriptive statistics, Caconde reservoir affluent flow.

min	max	median	mean	std. deviation	asymmetry	kurtosis
6.70	207.00	35.00	46.14	32.59	1.56	5.76

Table 4 contains some descriptive statistics on the time series. The average monthly affluent

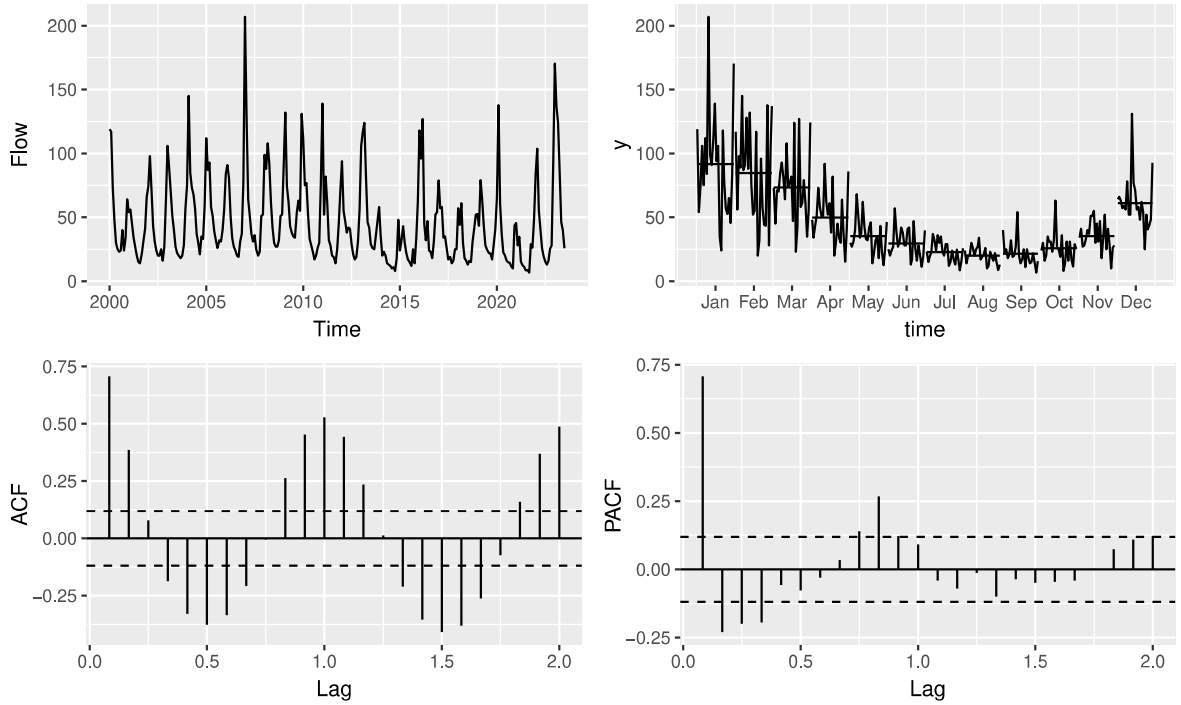
flow of the Caconde reservoir during the observed period is $46.14\,\mathrm{m}^3/\mathrm{s}$. September 2021 had the lowest inflow at $6.70\,\mathrm{m}^3/\mathrm{s}$. The month with the highest monthly inflow was January 2007, at $207.00\,\mathrm{m}^3/\mathrm{s}$. Figure 3 contains the observed time series (top left panel), the seasonal component over time (top right panel) and the correlogram and partial correlograms (bottom panels). In the top right panel, data for each season are plotted as a separate time series and the horizontal black line represents the mean of the observations within the season. The affluent flow increases from July to January, corresponding to the rainiest period, and decreases from February to June, which is the driest period. There is thus seasonality. We account for it using the approach recommended by Bloomfield (2004), i.e., we include in the model the following vector of harmonic covariates: $\boldsymbol{x}_t = (\sin(2\pi t/12), \cos(2\pi t/12))^\top$, $t \in \{1, \ldots, n\}$. Since seasonal fluctuations may not be fully captured by the harmonic regressors, we also considered dummy variables for the different months of the year and for groups of months. The indicator variable for January led to an improvement in the models' fit. We therefore included the following indicator variable in the fitted models: d_{1t}, which is equal to one for the months of January and equal to zero otherwise. Notably, January is the month of the year with the highest average affluent flow.

Notice from the upper right panel of Figure 3 that the variability of the series increases with the mean, with greater variability in the months when the mean affluent flow levels are higher. The beta prime law naturally accommodates this behavior. Recall that if $Y \sim BP(\mu, \phi)$, then $\mathrm{Var}(Y) = \mu(1+\mu)/\phi$. For a given precision, the variance of the distribution grows with the mean at a quadratic rate.

Parameter estimation is carried out as in the previous chapter, that is, using the L-BFGS-B nonlinear optimization algorithm with analytical first derivatives. We restrict α_2 and δ to be positive and negative, respectively, and the following starting values are used: $n^{-1}\sum_{t=1}^{n} g_1(Y_t)$ is used as starting value for α_1, all autoregressive and moving average parameters are set to zero, and the initial values for α_2 and δ are 3 and the -1, respectively.

At the outset, we select a generalized BPARMA model by considering all models such that $p = 0, \ldots, 4$ and $q = 0, \ldots, 4$. In all fitted models, g_1 and g_2 are the log link function. We also considered models with identity and square root precision link functions. However, slightly more accurate predictions were obtained using the log link. Model selection was carried out using information criteria (AIC and BIC), z tests, correlograms and partial correlograms of residuals, and portmanteau tests (Ljung-Box and Monti). When a given model had more than one autoregressive or moving average term without statistical significance at 5%, these terms

Figure 3 – Observed time series (top left), seasonal component (top right), correlogram (bottom left) and partial correlogram (bottom right), Caconde reservoir affluent flow, $n = 283$.



were individually excluded in descending order of p-values. The three models with the lowest information criteria values had dynamic terms that were not statistically significant. In all three cases, removing these terms led to the generalized BPARMA$(1, 0)$ model. We will then use the generalized BPARMA$(1, 0)$ model with harmonic regressors and an indicator variable for the month of January in our empirical analysis. The parameter estimates, standard errors, z test statistics, z tests, p-values, deviance value, and Ljung-Box and Monti test statistics and p-values for the selected model are presented in Table 5. The number of lags (k) in the two portmanteau test statistics is approximately equal to the square root of the sample size. These tests show no evidence of model misspecification at the usual significance levels. The coefficients β_1, β_2 and β_3 correspond, respectively, to $\sin(2\pi t/12)$, $\cos(2\pi t/12)$ and the dummy variable.

The estimates of α_2 and δ are 3.5766 and -1.1403, respectively. All estimated precisions are thus smaller than $\exp(3.5766) = 35.7529$. The smaller (larger) y_{t-1}/y_{t-2}, the closer (the farther) ϕ_t will be to (from) such a limiting value. Consider three situations, namely (i) $y_{t-1}/y_{t-2} = 1.4$ (sharp increase in the value of the process), (ii) $y_{t-1}/y_{t-2} = 1.0$ (no change in the value of the process), and (iii) $y_{t-1}/y_{t-2} = 0.6$ (sharp decrease in the value of the process). Using the parameter estimates and the inverse precision link, we obtain the

Table 5 – Generalized BPARMA model fit and diagnostic measures/tests, Caconde reservoir affluent flow.

parameter	estimate	std. error	z test statistic	z test p-value
α_1	0.8836	0.1416	6.2402	< 0.0001
φ_1	0.7689	0.0399	19.2533	< 0.0001
β_1	0.5811	0.0500	11.6106	< 0.0001
β_2	0.3636	0.0458	7.9440	< 0.0001
β_3	0.1927	0.0587	3.2818	0.0010
α_2	3.5766	0.2494	14.3410	< 0.0001
δ	-1.1403	0.2262	5.0414	< 0.0001
deviance $= 254.7287$				
Ljung-Box statistic ($k = 17$): $Q_{LB} = 16.5090$ (p-value $= 0.4180$)				
Monti statistic ($k = 17$): $Q_M = 15.8200$ (p-value $= 0.4656$)				

following corresponding estimated precisions at t (i.e, $\hat{\phi}_t$): 7.2440, 11.4307, and 18.0369. A sharp increase (decrease) in the value of y_{t-1} relative to y_{t-2} signals increased (decreased) uncertainty and the fitted model responds by reducing (increasing) the current precision relative to stability, i.e., relative to $y_{t-1} = y_{t-2}$.

To illustrate this more closely with the data, consider observations 230 and 231, whose observed values are, respectively, 43.37 and 79.08. The ratio between these two observations (y_{231}/y_{230}) is 1.8234, and the fitted precision value for the next period is $\hat{\phi}_{232} = 4.4706$. For reference, $\hat{\phi}_{231} = 13.9388$. By contrast, between time periods 242 and 243 the level of the series is reduced from 137.66 to 60.68. As a consequence, the estimated precision responds as follows: $\hat{\phi}_{244} = 21.6285$ vs $\hat{\phi}_{243} = 3.2157$. That is, the next estimated precision is increased relative to the current precision level given the current reduction in uncertainty.

Figure 4 contains diagnostic plots for the fitted generalized BPARMA model. The two panels are the residual correlogram (left) and the residual partial correlogram (right). In both panels, all autocorrelations and partial autocorrelations are within the asymptotic 95% bands (dashed blue lines). Hence, there is no clear evidence of serial correlation in the residuals which is in agreement with the conclusion drawn from the two portmanteau tests. The plots and tests validate the fitted model, indicating that it can be used for out-of-sample forecasting. Figure 5 shows the observed time series (black line) together with the fitted values (blue line). There is good agreement between observed and predicted values.

Based on the model selection procedure used previously, we arrived at the BPARMA$(1, 1)$ model with fixed precision, harmonic regressors and a dummy variable for the month of January (d_{1t}). In Table 6, we present the AIC and BIC values for the selected generalized BPARMA

Figure 4 – Residual correlogram (left) and residual partial correlogram (right), fitted generalized BPARMA model, Caconde reservoir affluent flow.

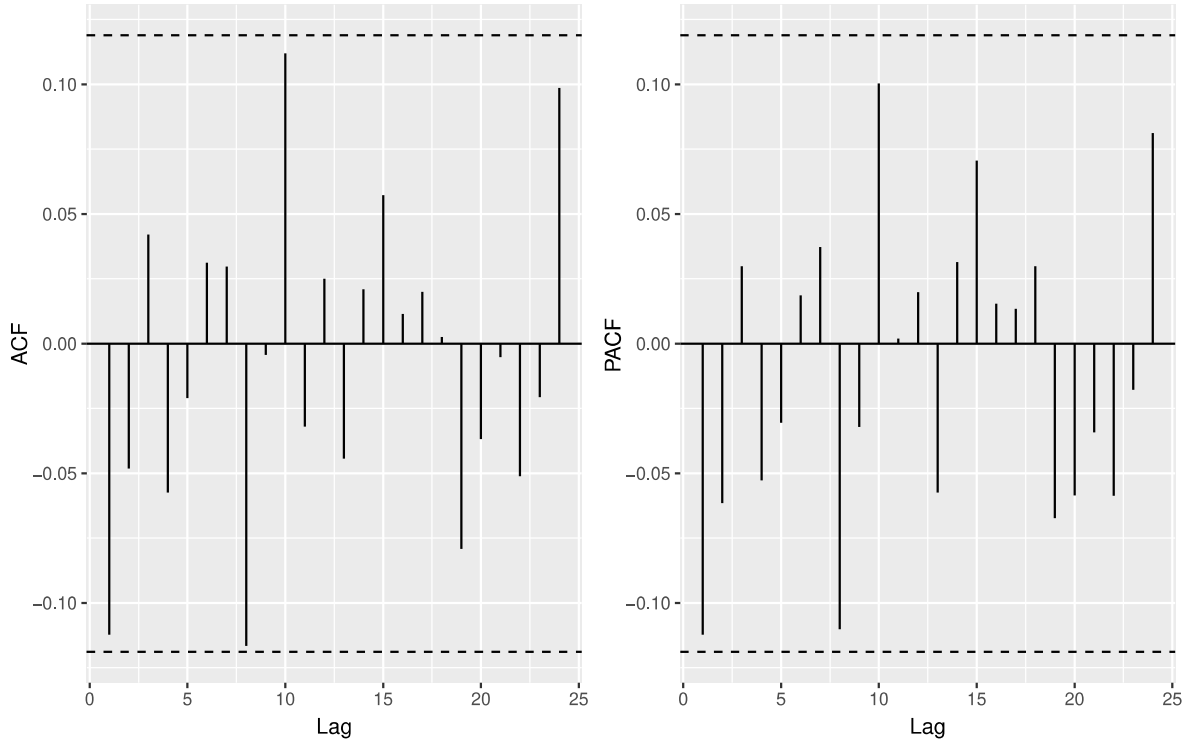
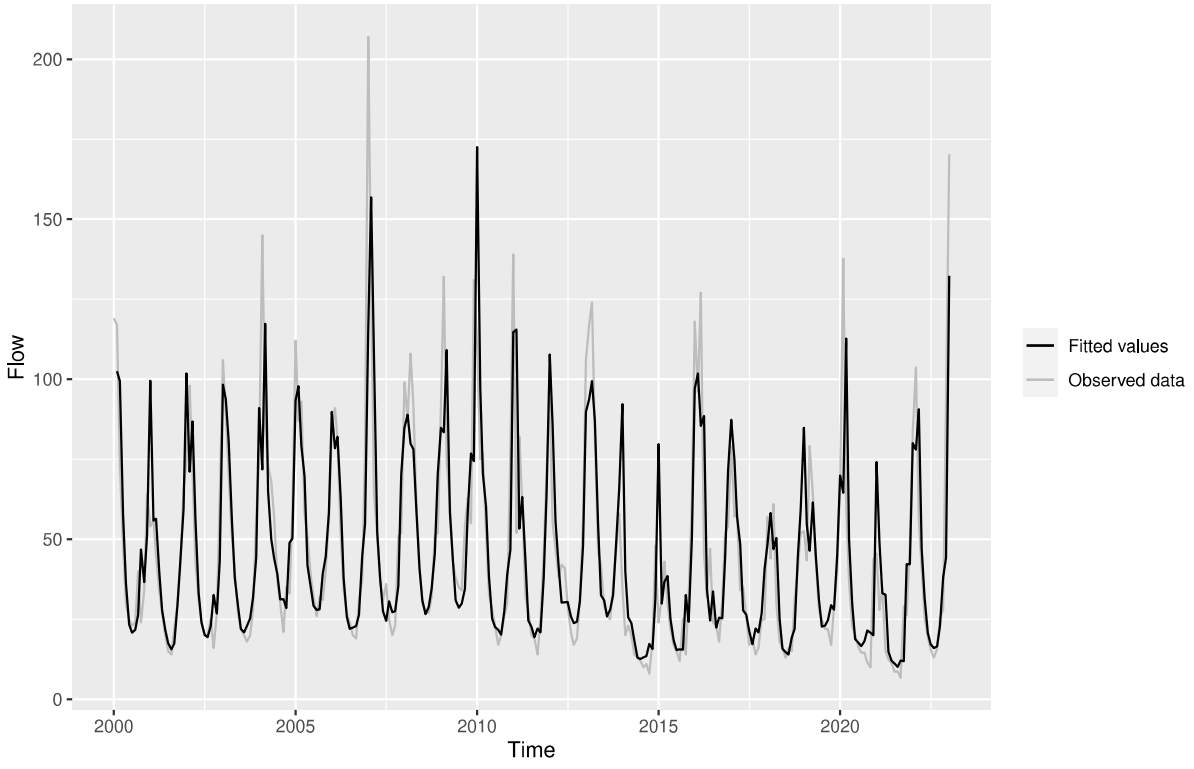


Figure 5 – Observed data and fitted values, Caconde reservoir affluent flow.



and standard BPARMA models. The smallest values are highlighted in boldface. Both criteria favor the generalized model over the standard model. We performed the likelihood ratio of

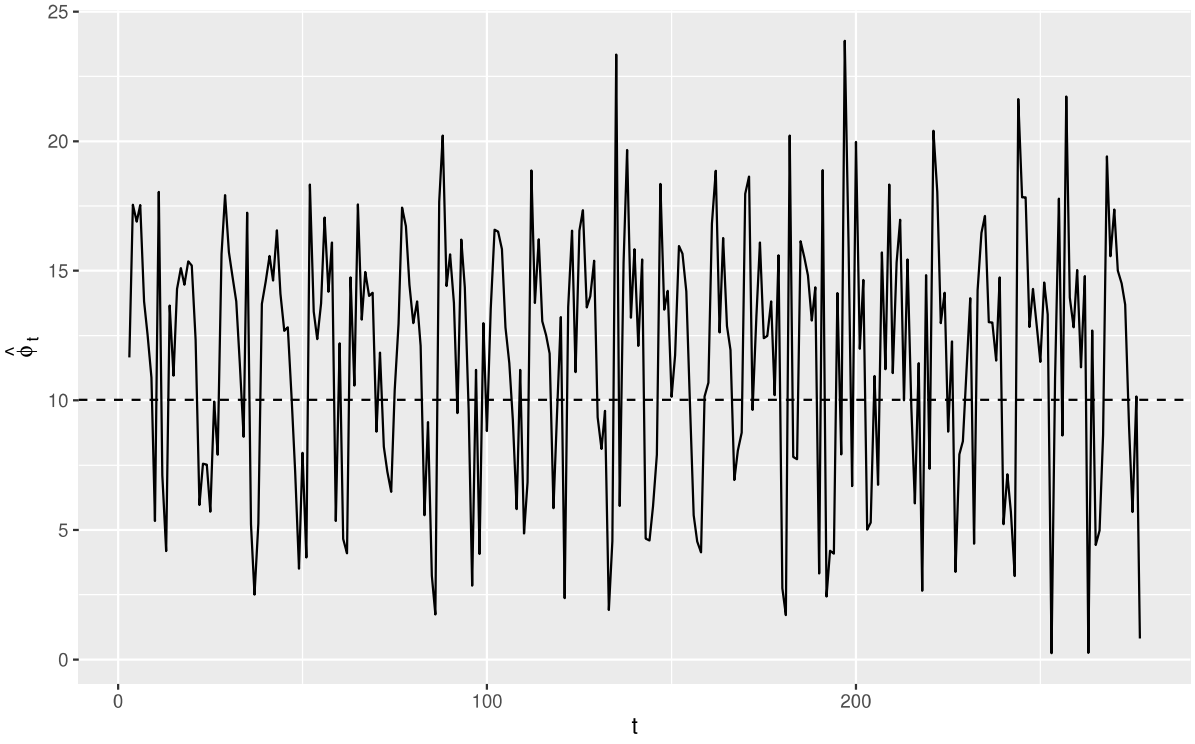
fixed vs varying precision, i.e., $\mathscr{H}_0 : \delta = 0$ versus $\mathscr{H}_1 : \delta \neq 0$. The likelihood ratio test p-value is smaller than 0.0001. There is thus evidence of varying precision.

Table 6 – Model selection criteria values for the generalized and standard models, Caconde reservoir affluent flow.

criterion	generalized BPARMA	BPARMA
AIC	**2099.2260**	2112.5320
BIC	**2128.2180**	2137.9000

Figure 6 contains the index plot of the estimated conditional precisions from the fitted generalized BPARMA model, i.e., it contains the plot of $\hat{\phi}_t$ vs t for $t \in \{m+1, \ldots, n\}$. The dashed horizontal line corresponds to the estimated precision from the fixed precision BPARMA model: $\hat{\phi} = 10.0174$. The average estimated precision from the generalized model is 11.7128. The estimated precisions from the generalized model fluctuate around the estimated (constant) precision of the standard model. The estimated precisions obtained from the generalized model are larger than the constant precision obtained from the standard model whenever $y_{t-1}/y_{t-2} < 1.1111$ and smaller otherwise.

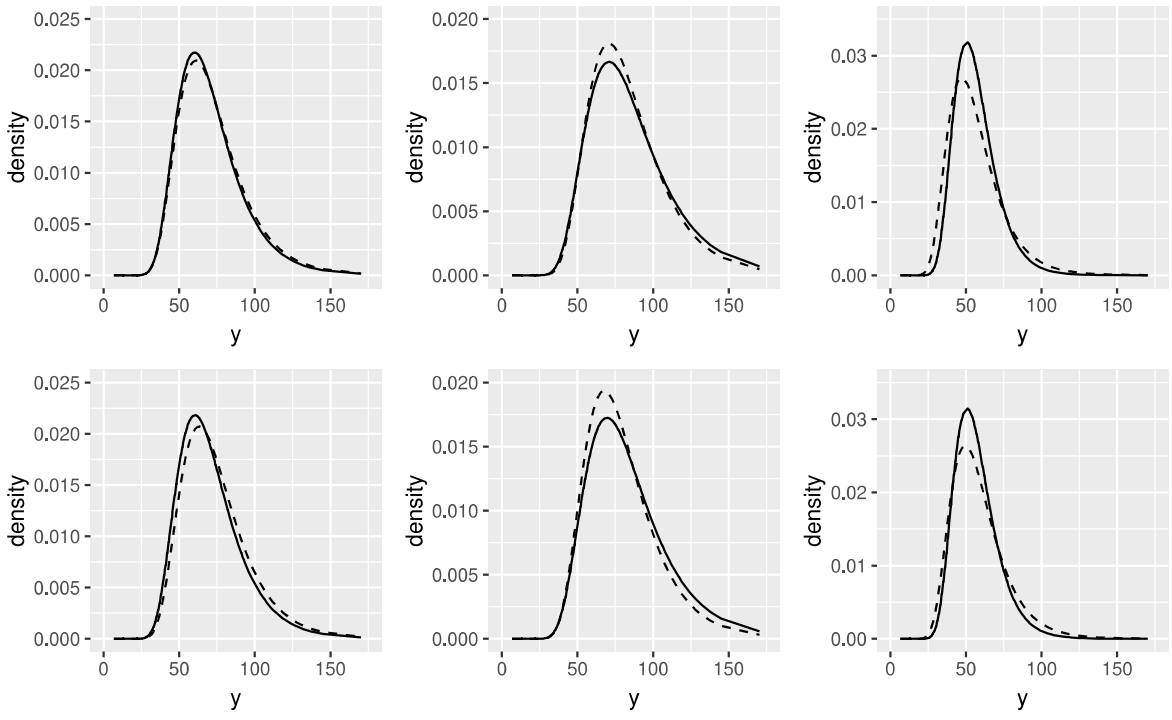
Figure 6 – Estimated conditional precisions (solid line) and fixed conditional precision estimate (dashed horizontal line).



As noted earlier, the introduction of a precision submodel adds an additional layer of flexibility to the modeling, as the shape of the beta prime density at time t is now determined

not only by the conditional mean at that time, but also by the conditional mean and precision at t. Given that the variance, skewness, and kurtosis of the beta prime distribution are functions of both the conditional mean and precision, the shape of the density over time can evolve more freely when both the conditional mean and precision vary with t. To illustrate, consider $t = 26, 27, 28$. In Figure 7, we present the beta prime densities corresponding to these three time instants (left column, middle column, right column, respectively) evaluated at estimates of the conditional mean and precision obtained from the generalized (solid line) and standard (dashed line) models. The top and bottom rows of panels correspond to the models of order $(1, 0)$ and $(1, 1)$, respectively. Note that the estimated beta prime densities exhibit greater variation over time when they are evaluated at the estimates of the conditional mean and precision obtained from the generalized model. In particular, there is more variability in the modes (peakedness) of the estimated beta prime densities obtained from the generalized model. As noted above, the variable precision in the beta prime model introduces an additional layer of flexibility by allowing the shape of the density to evolve more freely over time.

Figure 7 – Beta prime densities for $t = 26$ (left panels), $t = 27$ (middle panels) and $t = 28$ (right panels) evaluated at conditional means and conditional precisions estimated using the generalized (solid line) and standard (dashed line) models, $(1, 0)$ and $(1, 1)$ models (top row and bottom row, respectively).

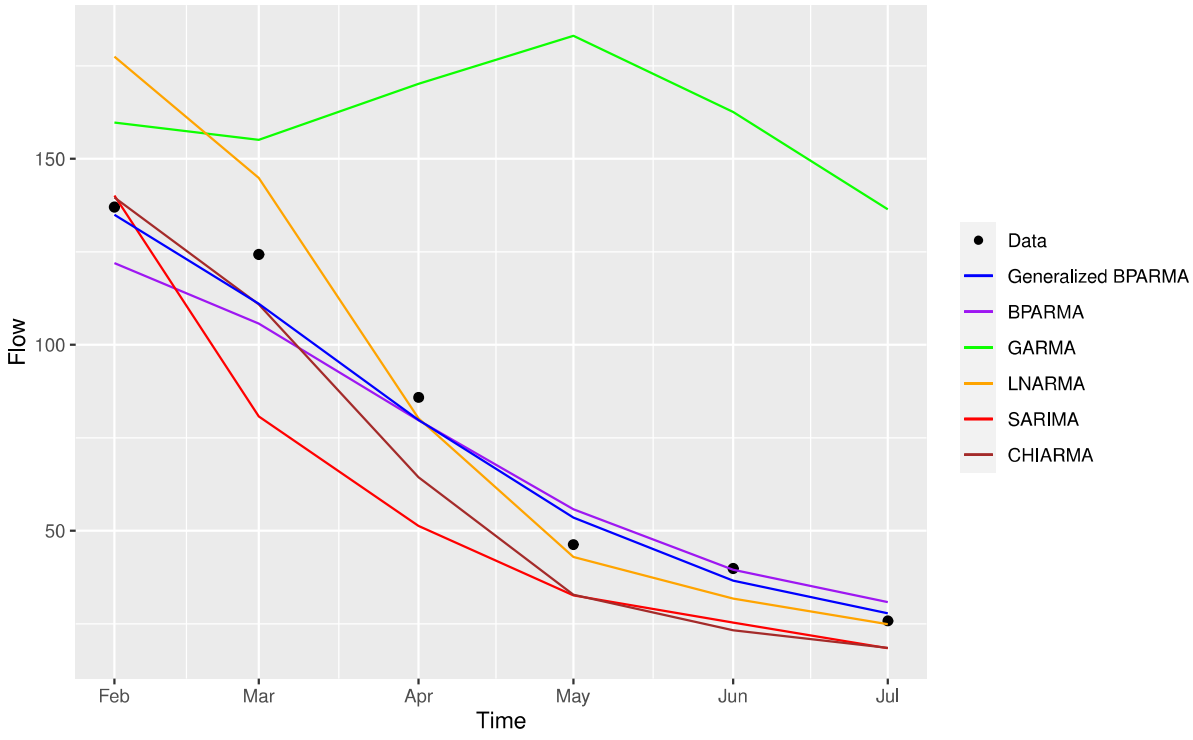


We will now move to out-of-sample forecasting. In addition to the generalized and standard BPARMA models, we will also use the CHIARMA, GARMA and LNARMA models, which are

based on the chi-squared, gamma and log-normal laws, respectively. The CHIARMA, GARMA and LNARMA models are defined using errors in original scale and are fitted using the PTSR package for R; see <https://cran.R-project.org/package=PTSR>. These models use the same regressors that were included in the BPARMA models, namely two harmonic covariates and an indicator variable for the month of January. Based on the AIC and BIC, we select the CHIARMA$(0, 3)$, GARMA$(4, 0)$ and LNARMA$(0, 3)$ models. We also consider the Gaussian SARIMA model, since it is widely used in empirical studies. The SARIMA model of order $(2, 0, 2)(2, 1, 0)$ was selected using the `auto.arima` function of the *forecast* package for R.

Figure 8 shows the six data points that had been reserved for forecasting evaluation together with the out-of-sample forecasts obtained from the five competing models. Notably, the generalized BPARMA forecasts are generally closer to the observed values than those yielded by the alternative dynamic models.

Figure 8 – Observed data and out-of-sample forecasts, Caconde reservoir affluent flow.



For each set of forecasts, we compute the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the symmetric mean absolute percentage error (sMAPE). These measures are defined as follows:

$$\text{MAE} = \frac{1}{h}\sum_{t=n+1}^{n+h}|Y_t - \hat{\mu}_t|, \quad \text{MAPE} = \frac{1}{h}\sum_{t=n+1}^{n+h}\frac{|Y_t - \hat{\mu}_t|}{|Y_t|} \times 100, \quad \text{and}$$

$$\text{sMAPE} = \frac{1}{h}\sum_{t=n+1}^{n+h}\frac{|Y_t - \hat{\mu}_t|}{(|Y_t| + |\hat{\mu}_t|)/2} \times 100,$$

where h is the forecasting horizon. (Here, $|Y_t| = Y_t$, since $Y_t > 0$.)

Table 7 contains measures of forecast accuracy from one to six steps ahead, that is, for $h \in \{1, \ldots, 6\}$. The generalized BPARMA model is the best performer according to all three forecast accuracy measures in all six forecasting horizons. For example, the MAPE of the three-steps-ahead forecast from the generalized BPARMA model is over 40% lower than that of the BPARMA model. The gains in forecasting accuracy for $h = 6$ according to the MAE of the generalized BPARMA model relative to the standard BPARMA, CHIARMA, GARMA, LNARMA an SARIMA models exceed 38%, 54%, 93%, 57% and 70%, respectively.

6.2 GUILMAN AMORIM RESERVOIR OUTFLOW

In the second empirical application, we present modeling and predictions for the hydro-environmental time series of outflow of the Guilman Amorim reservoir. The outflow of a reservoir is the amount of water released through gates, valves, or other control devices. In other words, it is the water that flows out of the reservoir towards the river. In reservoir management, outflow plays a crucial role, as it directly impacts the conditions of the river, water supply, hydroelectric power generation, flood control, and other aspects related to water resources.

The Guilman Amorim reservoir is located on the Piracicaba River, between the municipalities of Nova Era and Antônio Dias, in the state of Minas Gerais, Brazil. The Guilman Amorim hydroelectric power plant was inaugurated in 1997. Its operation is conducted by the UHE Guilman-Amorim consortium, formed by the companies ArcelorMittal Brasil and Samarco Mineração.

The data on the outflow of the Guilman Amorim reservoir were acquired from the Brazilian National Electric System Operator (Operador Nacional do Sistema Elétrico – ONS, <http://www.ons.org.br>), covering the period from January 2000 to September 2023. This variable is recorded in cubic meters per second (m^3/s). The sample size is $n = 285$. The last six observations (April to September 2023) were removed from the time series and will be used for accuracy evaluation of forecasts.

Table 8 contains some descriptive statistics of the data. The average outflow from the Guilman Amorim reservoir is $62.36\,\text{m}^3/\text{s}$. The month with the lowest outflow was October

Table 7 – Measures of forecasting accuracy for different forecasting horizons.

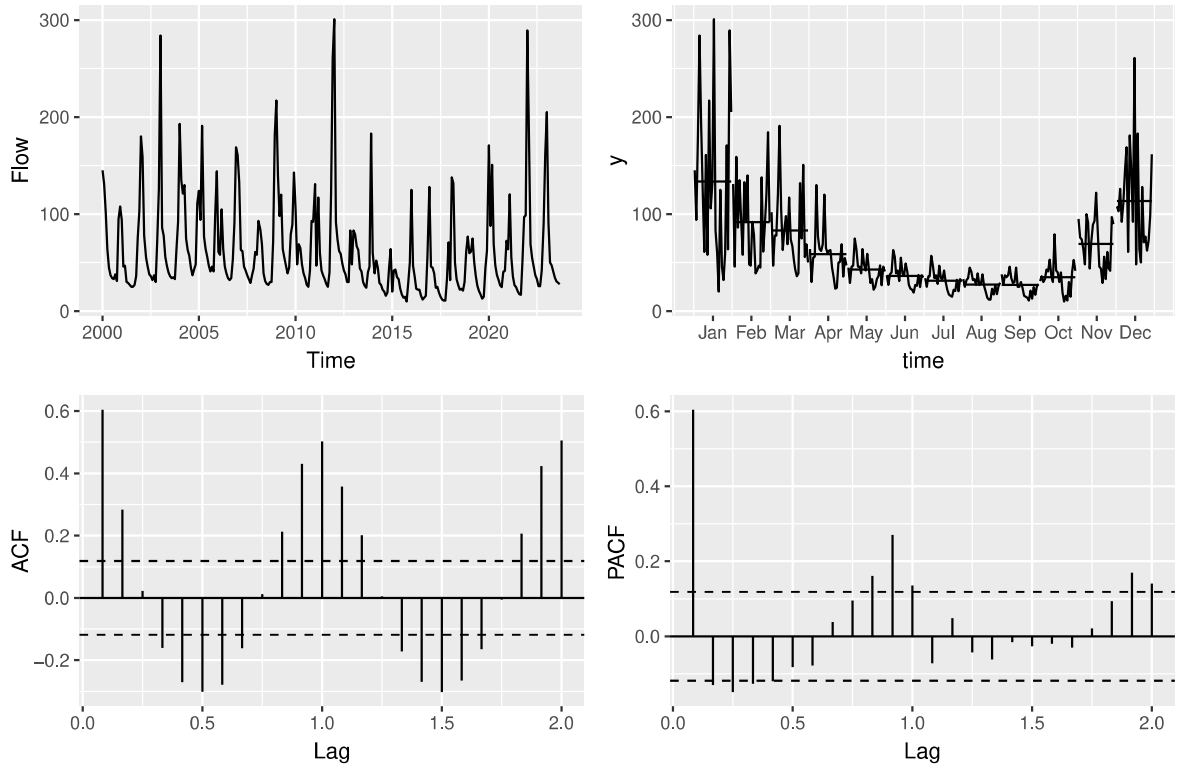
	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
			MAE			
Generalized BPARMA	**2.0339**	**7.6563**	**7.1194**	**7.1429**	**6.3712**	**5.6491**
BPARMA	15.0714	16.8482	13.2888	12.3405	9.9384	9.1211
CHIARMA	2.5756	8.0194	12.4961	12.7522	13.5209	12.4751
GARMA	22.7166	26.7406	45.9213	68.6486	79.4770	84.6782
LNARMA	40.4222	30.4963	22.2144	17.4958	15.6210	13.1655
SARIMA	3.0466	23.2878	27.0575	23.7187	21.8841	19.4578
			MAPE			
Generalized BPARMA	**1.4841**	**6.0830**	**6.4024**	**8.6976**	**8.6069**	**8.4913**
BPARMA	10.9970	12.9899	11.0552	13.4198	10.9016	12.3420
CHIARMA	1.8793	6.3548	12.5639	16.7250	21.7111	22.7807
GARMA	16.5754	20.6619	46.4956	108.7702	148.6581	195.4942
LNARMA	29.4945	23.0210	17.5411	14.9597	16.0449	13.9453
SARIMA	2.2230	18.6198	25.8447	26.7837	28.7289	28.6811
			sMAPE			
Generalized BPARMA	**1.4952**	**6.3899**	**6.6927**	**8.6337**	**8.6266**	**8.4575**
BPARMA	11.6369	13.9164	11.7622	13.4730	10.9448	12.0880
CHIARMA	1.8618	6.6561	13.9534	19.0159	25.7354	26.9016
GARMA	15.3068	18.6650	34.3917	55.6160	68.7512	80.0386
LNARMA	25.7039	20.4935	15.9308	13.8195	15.5955	13.5808
SARIMA	2.1985	22.3236	31.7027	32.4628	34.9028	34.6119

Table 8 – Descriptive statistics; Guilman Amorim reservoir outflow.

min	max	median	mean	std. deviation	asymmetry	kurtosis
10.00	301.00	46.00	62.36	48.94	2.10	8.37

2015, with $10.00\,\mathrm{m}^3/\mathrm{s}$. January 2012 had the highest outflow, reaching $301.00\,\mathrm{m}^3/\mathrm{s}$. Figure 9 shows the observed series (top-left), the seasonal component (top-right) and the correlogram (bottom-left) and partial correlogram (bottom-right). There is clearly seasonality. To model this seasonality, we will use harmonic regressors and a dummy variable, which equals one for the months of December and January (months with the highest outflows) and zero otherwise.

Figure 9 – Observed time series (top-left), seasonal component (top-right), correlogram (bottom-left) and partial correlogram (bottom-right); Guilman Amorim reservoir outflow.



We selected the generalized BPARMA model exploring all combinations of $p = 0, \ldots, 5$ and $q = 0, \ldots, 5$, except $p = q = 0$. We used the logarithmic link function in both submodels. As in the previous application, in the selection process, we used information criteria, z tests, correlograms and partial correlograms of the residuals, as well as portmanteau tests. We excluded autoregressive and moving average terms that were not significant at 5%. The model chosen was the generalized BPARMA(4,1) with harmonic regressors and a dummy variable for the months of December and January. Table 9 presents the parameter estimates, standard

errors, z test statistics and corresponding p-values, deviance value, and Ljung-Box and Monti test statistics and p-values. The portmanteau tests provide no evidence of model misspecification at the usual significance levels. As in the previous application, the coefficients β_1, β_2 and β_3 correspond, respectively, to $\sin(2\pi t/12)$, $\cos(2\pi t/12)$ and the dummy variable.

Table 9 – Generalized BPARMA model fit and diagnostic measures/tests; Guilman Amorim reservoir outflow.

parameter	estimate	std. error	z test statistic	z test p-value
α_1	2.4442	0.3521	6.9418	< 0.0001
φ_1	-0.2769	0.0487	5.6900	< 0.0001
φ_2	0.5520	0.0603	9.1466	< 0.0001
φ_4	0.1296	0.0518	2.5038	0.0123
θ_1	0.9990	0.0133	75.0996	< 0.0001
β_1	0.4778	0.0499	9.5735	< 0.0001
β_2	0.3674	0.0505	7.2808	< 0.0001
β_3	0.3457	0.0792	4.3670	< 0.0001
α_2	2.9450	0.2372	12.4159	< 0.0001
δ	-0.9279	0.2126	4.3645	< 0.0001
deviance $= 245.3111$				
Ljung-Box statistic ($k = 17$): $Q_{LB} = 17.3290$ (p-value $= 0.1847$)				
Monti statistic ($k = 17$): $Q_M = 17.6920$ (p-value $= 0.1696$)				

Figure 10 contains the correlogram (left panel) and partial correlogram (right panel) of the residuals from the estimated model. Nearly all the autocorrelations and partial autocorrelations are within the asymptotic confidence bands (level: 95%). There is therefore no evidence of model misspecification. Figure 11 shows the observed time series (gray line) together with the fitted values (black line). It can be seen that the model shows a satisfactory fit to the observed values.

Using the same selection procedure as for the generalized BPARMA model, we choose the standard BPARMA$(1,1)$ model with harmonic regressors and a dummy variable for the months of December and January. The AIC and BIC of the generalized BPARMA model, 2350.1304 and 2386.4425, are lower than those of the standard BPARMA model, 2397.4337 and 2422.8522. The p-value of the likelihood ratio test of $\mathscr{H}_0 : \delta = 0$ versus $\mathscr{H}_1 : \delta \neq 0$ is smaller than 0.0001. The information criteria and the likelihood ratio test thus favor the generalized BPARMA model over the standard BPARMA model.

As in the previous empirical investigation, in addition to the generalized and standard BPARMA models, we consider the CHIARMA, GARMA and LNARMA models, with the same

Figure 10 – Residual correlogram (left) and residual partial correlogram (right), fitted generalized BPARMA model; Guilman Amorim reservoir outflow.

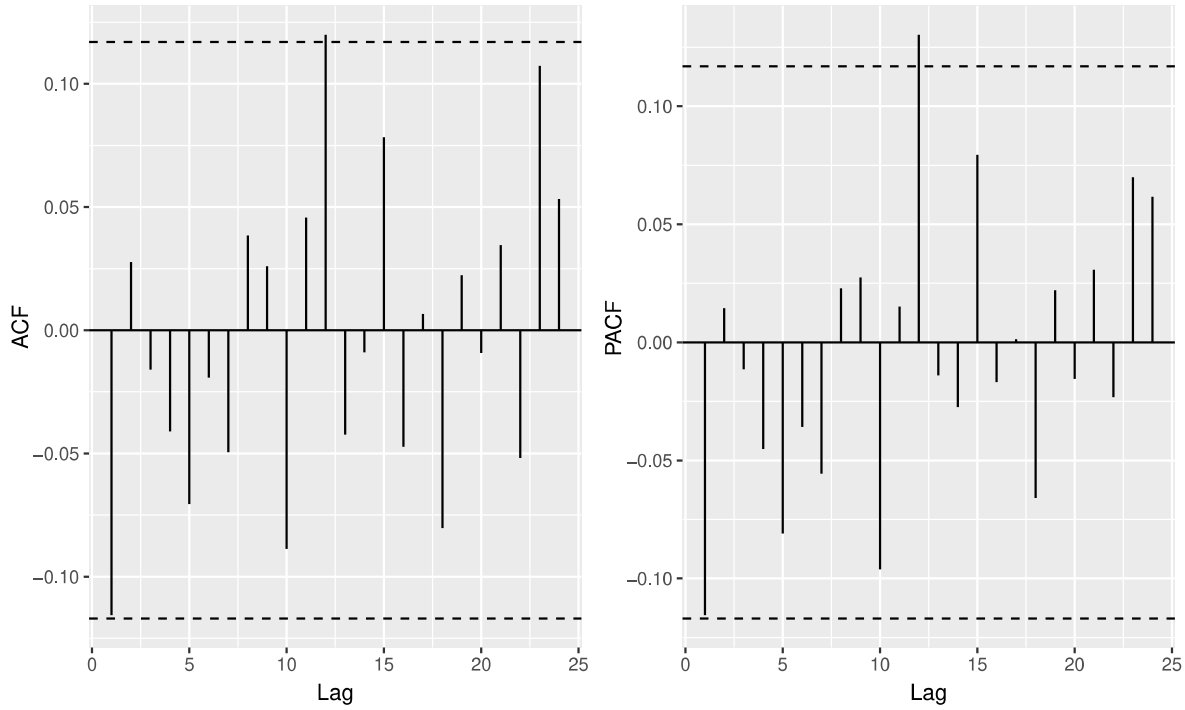
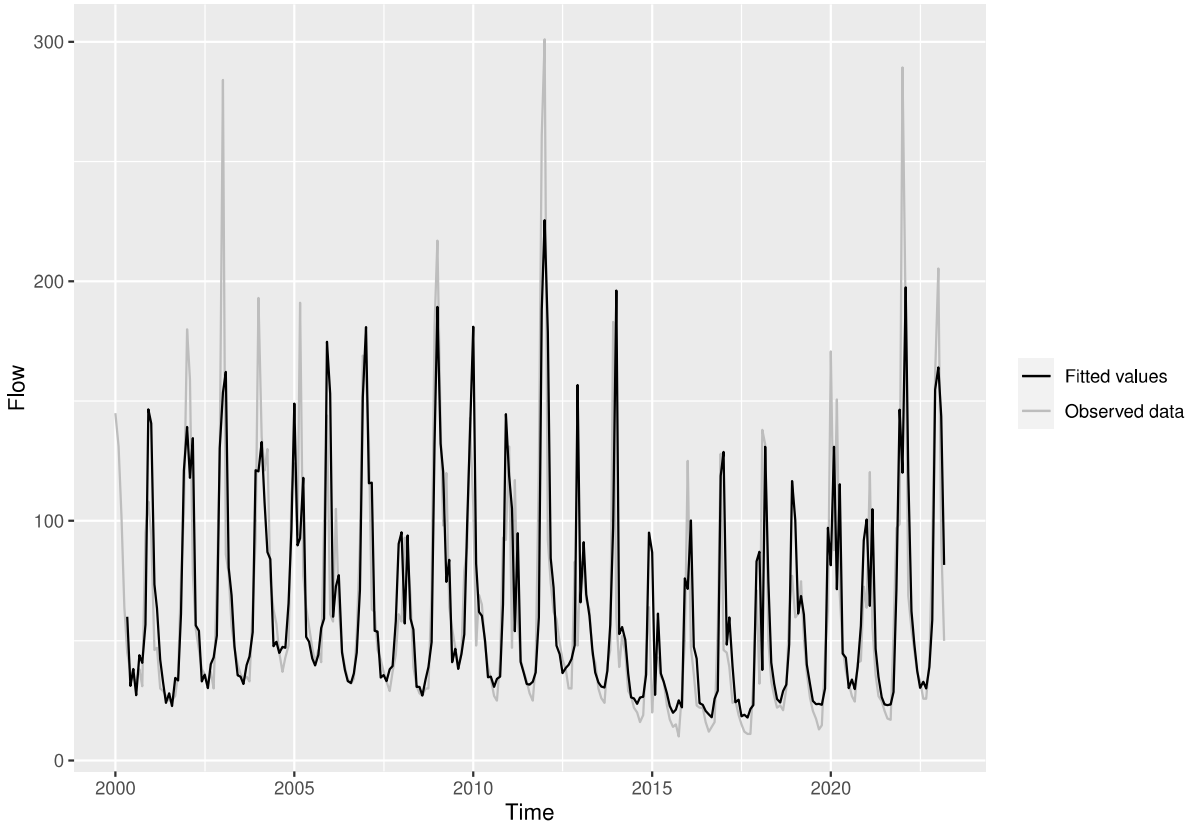
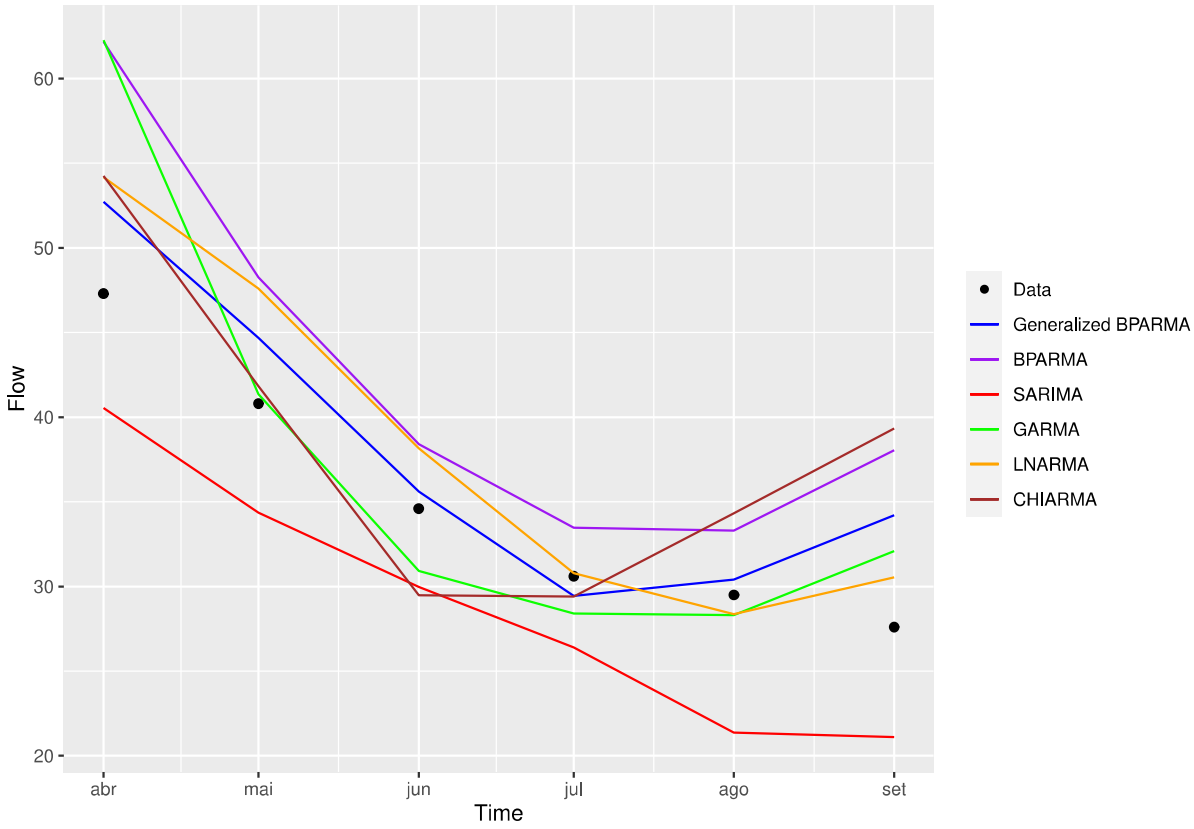


Figure 11 – Observed data and fitted values; Guilman Amorim reservoir outflow.



covariates. Based on the AIC and BIC, we select the CHIARMA$(4,0)$, GARMA$(3,0)$ and LNARMA$(0,2)$ models. We also include in the analysis the Gaussian SARIMA model of order $(1,0,0)(1,1,0)$, obtained using the `auto.arima` function of the *forecast* package for R. Figure 12 shows the six data points reserved for evaluating the forecasts, along with the out-of-sample forecasts from the generalized BPARMA model and the five competing models. It can be seen that, overall, the forecasts from the generalized BPARMA model are closer to the observed values than the forecasts obtained from the other models.

Figure 12 – Observed data and out-of-sample forecasts; Guilman Amorim reservoir outflow.



To evaluate the forecasts one to six steps ahead, i.e., for $h \in \{1,\ldots,6\}$, we use the same accuracy measures as in the previous empirical analysis. The results are shown in Table 10. According to the three measures, the generalized BPARMA model performs best for forecast horizons $h \in \{1,3,4,5,6\}$, while the CHIARMA model performs best for $h = 2$. For all forecast horizons, the forecast accuracy measures of the generalized BPARMA model are more than 50% lower than those of the standard BPARMA model.

Table 10 – Measures of forecasting accuracy for different forecasting horizons.

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
			MAE			
Generalized BPARMA	**5.4161**	4.6472	**3.4335**	**2.8661**	**2.4744**	**3.1637**
BPARMA	14.8694	11.1617	8.7068	7.2468	6.5577	7.2045
CHIARMA	6.9561	**3.9913**	4.3691	3.5772	3.8278	5.1467
GARMA	14.9738	7.7695	6.4095	5.3562	4.5216	4.5163
LNARMA	6.8760	6.8396	5.7494	4.3585	3.7122	3.5835
SARIMA	6.7483	6.5930	5.9375	5.5011	6.0275	6.1067
			MAPE			
Generalized BPARMA	**11.4505**	10.4781	**7.9547**	**6.9167**	**6.1488**	**9.1156**
BPARMA	31.4364	24.8529	20.2268	17.5123	16.5867	20.1259
CHIARMA	14.7062	**8.6112**	10.6777	8.9899	10.4666	15.8122
GARMA	31.6571	16.5212	14.5685	12.7209	10.9787	11.8602
LNARMA	14.5370	15.6058	13.8421	10.5335	9.1908	9.4344
SARIMA	14.2669	15.0228	14.4722	14.2789	16.9372	18.0413
			sMAPE			
Generalized BPARMA	**10.8305**	9.9525	**7.5903**	**6.6619**	**5.9356**	**8.5111**
BPARMA	27.1663	21.9533	18.1034	15.8150	15.0729	17.8619
CHIARMA	13.6989	**8.0920**	10.7264	9.0461	10.2637	14.3996
GARMA	27.3310	14.3533	13.3234	11.8540	10.3015	11.0919
LNARMA	13.5520	14.4717	12.9173	9.8395	8.6505	8.8943
SARIMA	15.3628	16.2464	15.6074	15.3821	18.7014	20.0358

7 CONCLUDING REMARKS

The beta prime distribution has received little attention in the literature, particularly for modeling random variables with time dependence. It is closely related to the beta distribution, which is the most widely used law with random phenomena that assume values in the standard unit interval. In fact, if Z is a beta distributed random variable, then $Z/(1-Z)$, the beta odds ratio, follows the beta prime law. In this research, we introduced a new class of dynamic time series models based on it with the precision parameter varying over time. The proposed generalized BPARMA model accommodates autoregressive and moving average dynamics and allows for the inclusion of non-stochastic regressors. It comprises two separate submodels, one for the conditional mean and another for the conditional precision. Hence, the two parameters that index the beta prime law are allowed to evolve over time. The variance, skewness, and kurtosis of the beta prime law are functions of the conditional mean and conditional precision. As such, the temporal evolution of these two parameters controls the shape of the conditional density, which evolves more freely than in models in which only the conditional mean changes over time.

We developed maximum likelihood parameter estimation for the new model. In particular, we obtained simple closed-form expressions for the model's conditional log-likelihood function, for the score vector, and for Fisher's information matrix that are presented in matrix form for ease of calculation. We also discussed diagnostic analysis for the proposed model and presented Monte Carlo simulation results on the finite-sample performance of the conditional maximum likelihood estimators of the parameters that index the model.

Two hydro-environmental empirical applications were presented and discussed. We modeled the monthly inflow of the Caconde water reservoir and the and the monthly outflow of the Guilman Amorim reservoir. We accounted for the existing seasonality by including two harmonic regressors and a dummy variable in the models. The generalized BPARMA model yielded a good data fit and out-of-sample forecasts that were more accurate than those obtained from competing models. Information criteria and the likelihood ratio test favored the generalized BPARMA model over its fixed precision counterpart. The proposed model also outperformed other alternative models when used for out of sample forecasting.

Two hydro-environmental empirical applications were presented and discussed. Specifically, we modeled the monthly inflow of the Caconde water reservoir and the and the monthly out-

flow of the Guilman Amorim reservoir. We accounted for the existing seasonality by including two harmonic regressors and a dummy variable in the model. The generalized BPARMA model yielded a god data fit and out-of-sample forecasts that were more accurate than those obtained from competing models. Information criteria and the likelihood ratio test favored the generalized BPARMA model over its fixed precision counterpart. The proposed model also outperformed other alternative models when used for out of sample forecasting.

REFERENCES

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716–723, 1974.

ALMEIDA-JUNIOR, P. M.; NASCIMENTO, A. D. G0I ARMA process for speckled data. *Journal of Statistical Computation and Simulation*, v. 91, n. 15, p. 3125–3153, 2021.

ANDERSEN, E. B. Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society B*, v. 32, n. 2, p. 283–301, 1970.

BAYER, F. M.; BAYER, D. M.; PUMI, G. Kumaraswamy autoregressive moving average models for double bounded environmental data. *Journal of Hydrology*, v. 555, n. 1-2, p. 385–396, 2017.

BENJAMIN, M. A.; RIGBY, R. A.; STASINOPOULOS, D. M. Generalized autoregressive moving average models. *Journal of the American Statistical Association*, v. 98, n. 461, p. 214–223, 2003.

BLOOMFIELD, P. *Fourier Analysis of Time Series: An Introduction*. New York: [s.n.], 2004.

BOURGUIGNON, M.; SANTOS-NETO, M.; CASTRO, M. de. A new regression model for positive random variables with skewed and long tail. *Metron*, v. 79, n. 1, p. 33–55, 2021.

BOX, G.; JENKINS, G. *Time Series Analysis: Forecasting and Control*. 2nd. ed. San Francisco: [s.n.], 1976.

FOKIANOS, K.; KEDEM, B. Partial likelihood inference for time series following generalized linear models. *Journal of Time Series Analysis*, v. 25, n. 2, p. 173–197, 2004.

KEEPING, E. S. *Introduction to Statistical Inference*. New York: [s.n.], 1962.

LEIVA, V.; SANTOS-NETO, M.; CYSNEIROS, F. J. A.; BARROS, M. Birnbaum-Saunders statistical modelling: a new approach. *Statistical Modelling*, v. 14, n. 1, p. 21–48, 2014.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. *Biometrika*, v. 65, n. 2, p. 297–303, 1978.

MCDONALD, J. B. Some generalized functions for the size distribution of income. *Econometrica*, v. 52, n. 3, p. 647–663, 1984.

MONTI, A. C. A proposal for a residual autocorrelation test in linear models. *Biometrika*, v. 81, n. 4, p. 776–780, 1994.

MORGAN, B. J. T.; PALMER, K. J.; RIDOUT, M. S. Negative score test statistic. *The American Statistician*, v. 61, n. 4, p. 285–288, 2007.

MYERS, R. H.; MONTGOMERY, D. C.; VINING, G. G.; ROBINSON, T. J. *Generalized Linear Models: With Applications in Engineering and the Sciences*. New York: [s.n.], 2012.

NOCEDAL, J.; WRIGHT, S. J. *Numerical Optimization*. 2nd. ed. New York: [s.n.], 2006.

PRESS, W. H.; TEUKOLSKY, S. A.; VETTERLING, W. T.; FLANNERY, B. P. *Numerical Recipes in C*. 2nd. ed. New York: [s.n.], 1992.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2023. Available at: <https://www.R-project.org/>.

ROCHA, A. V.; CRIBARI-NETO, F. Beta autoregressive moving average models. *Test*, v. 18, n. 3, p. 529, 2009.

ROCHA, A. V.; CRIBARI-NETO, F. Erratum to: Beta autoregressive moving average models. *Test*, v. 26, n. 2, p. 451–459, 2017.

SANTOS-NETO, M.; CYSNEIROS, F. J. A.; LEIVA, V.; BARROS, M. Reparameterized Birnbaum-Saunders regression models with varying precision. *Electronic Journal of Statistics*, v. 10, p. 2825–2855, 2016.

SCHER, V. T.; CRIBARI-NETO, F.; BAYER, F. M. Generalized βARMA model for double bounded time series forecasting. *International Journal of Forecasting*, v. 40, n. 2, p. 721–734, 2024.

SCHWARZ, G. Estimating the dimension of a model. *Annals of Statistics*, v. 6, n. 2, p. 461–464, 1978.

STONE, R. F.; LOOSE, L. H.; MELO, M. S.; BAYER, F. M. The Chen autoregressive moving average model for modeling asymmetric positive continuous time series. *Symmetry*, v. 15, n. 9, p. 1675, 2023.