



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Modelo Kumaraswamy Inflacionado Autorregressivo de Médias Móveis com Aplicações em Dados Hidroambientais

Camila Malu da Rosa

Porto Alegre - RS, abril de 2023

CIP - Catalogação na Publicação

da Rosa, Camila Malu
Modelo Kumaraswamy Inflacionado Autorregressivo de
Médias Móveis com Aplicações em Dados Hidroambientais
/ Camila Malu da Rosa. -- 2023.
59 f.
Orientador: Fábio Mariano Bayer.

Coorientador: Francisco Cribari-Neto.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Instituto de Matemática e
Estatística, Programa de Pós-Graduação em Estatística,
Porto Alegre, BR-RS, 2023.

1. Distribuição Kumaraswamy Inflacionada. 2.
Modelagem ARMA. 3. Modelo dinâmico. 4. Previsões. 5.
Taxas e proporções. I. Bayer, Fábio Mariano, orient.
II. Cribari-Neto, Francisco, coorient. III. Título.

Dissertação submetida por Camila Malu da Rosa¹ como requisito parcial para a obtenção do título de Mestre em Estatística pelo Programa de Pós-Graduação em Estatística da Universidade Federal do Rio Grande do Sul.

Orientador:

Fábio Mariano Bayer (UFSM)

Coorientador:

Francisco Cribari-Neto (UFPE)

Comissão Examinadora:

Airlane Alencar (USP)

Cleiton Guollo Taufemback (UFRGS)

Klaus Leite Pinto Vasconcellos (UFPE)

Data de Apresentação: 5 de abril de 2023

¹Bolsista do Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES

Agradecimentos

À minha mãe, Janinha, e meu pai, João Pedro, que me deram todo o apoio e incentivo para que eu chegasse até esta etapa de minha vida.

Ao meu companheiro, Vagner, por todo amor, carinho e paciência nesse período.

Ao meu orientador, Dr. Fábio Mariano Bayer, pelos conhecimentos passados durante o mestrado, pela paciência e dedicação no trabalho.

Ao meu coorientador, Dr. Francisco Cribari-Neto, pelas sábias sugestões e comentários que foram de fundamental importância para o desenvolvimento da pesquisa.

E à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de Mestrado.

Resumo

Neste trabalho, propomos o modelo Kumaraswamy inflacionado autorregressivo e de médias móveis (IKARMA), uma classe de modelos dinâmicos para séries temporais que assumem valores em $[0, 1)$ ou $(0, 1]$. Para isso, supõe-se que a variável dependente assume, condicionalmente ao conjunto de observações passadas, distribuição Kumaraswamy inflacionada (KI). A distribuição KI é obtida a partir de uma mistura de distribuições, encampanando uma distribuição discreta (Bernoulli) e outra contínua (Kumaraswamy). Como é amplamente conhecido na literatura, a família de distribuição Kumaraswamy é muito útil para modelar dados hidroambientais e de áreas afins. No modelo proposto, o componente aleatório segue distribuição KI, enquanto a parte determinística considera duas estruturas dinâmicas, uma para a mediana condicional e outra para o parâmetro de mistura, sendo esta última simples e parcimoniosa. A estrutura dinâmica utilizada para mediana condicional abriga dinâmicas autorregressiva e de médias móveis e permite a inclusão de um conjunto de regressores. Após revisar a literatura e apresentar o modelo proposto, discutimos procedimentos inferenciais sobre os parâmetros do modelo via máxima verossimilhança condicional. Expressões explícitas para o vetor escore condicional e a matriz de informação de Fisher condicional são apresentadas. Algumas ferramentas de diagnóstico e previsão também são propostas. Resultados de simulações de Monte Carlo realizadas para avaliar o desempenho das inferências em amostras de tamanho finito são apresentados. Por fim, duas aplicações empíricas relacionadas a dados hidroambientais reais são apresentadas e discutidas.

Palavras-chave: Distribuição Kumaraswamy inflacionada; Modelagem ARMA; Modelo dinâmico; Previsões; Taxas e proporções.

Abstract

In this work, we propose the inflated Kumaraswamy autoregressive and moving average model (IKARMA), a class of dynamic models for time series that assume values in $[0, 1)$ or $(0, 1]$. It is supposed that the dependent variable assumes, conditionally to the set of past observations, an inflated Kumaraswamy distribution (KI). The KI distribution is obtained from a mixture of distributions, embracing a discrete distribution (Bernoulli) and a continuous one (Kumaraswamy). As it is widely known in the literature, the Kumaraswamy distribution family is very useful for modeling hydro environmental and related data. In the proposed model, the random component follows the KI distribution, while the deterministic part considers two dynamic structures, one for the conditional median and another for the mixture parameter, the latter being simple and parsimonious. The dynamic structure used for conditional median accommodates autoregressive and moving average dynamics and allows the inclusion of a set of regressors. After reviewing the literature and presenting the proposed model, we discuss inferential procedures about the model parameters via conditional maximum likelihood. Closed-form expressions for the conditional score vector and the conditional Fisher information matrix are presented. Some diagnostic and forecasting tools are also proposed. Results of Monte Carlo simulations performed to evaluate the performance of inferences in finite sample sizes are presented. Finally, two empirical applications related to real hydro environmental data are presented and discussed.

Keywords: ARMA modeling; Dynamic model; Forecasts; Inflated Kumaraswamy distribution; Rates and proportions.

Lista de Figuras

3.1	Função densidade de probabilidade $ki(\lambda_t, b, \mu_t, \varphi \mathcal{F}_{t-1})$, com $\varphi = 10$ e diferentes valores dos parâmetros λ_t e μ_t	18
5.1	Série temporal de UR observadas em Caxias do Sul, RS, Brasil: série, componente sazonal, correlograma e correlograma parcial.	37
5.2	Gráficos de diagnóstico do modelo IKARMA ajustado para dados de UR máxima mensal na cidade de Caxias do Sul, RS, com base nos resíduos quantílicos aleatorizados: gráfico dos resíduos contra os índices das observações, QQ-plot dos resíduos, correlograma e correlograma parcial dos resíduos.	39
5.3	Predições e previsões obtidas a partir dos diferentes modelos ajustados para a UR máxima mensal na cidade de Caxias do Sul, RS.	41
5.4	Série temporal de VU de São Simão, MG: série, componente sazonal, correlograma e correlograma parcial.	43
5.5	Gráficos de diagnóstico do modelo IKARMA ajustado para dados VU de São Simão, MG, com base nos resíduos quantílicos aleatorizados: gráfico dos resíduos contra os índices das observações, QQ-plot dos resíduos, correlograma e correlograma parcial dos resíduos.	44
5.6	Predições e previsões obtidas a partir dos diferentes modelos ajustados para o VU de São Simão, MG.	46

Lista de Tabelas

4.1	Resultados da simulação de Monte Carlo, modelo IKARMA(1,1) inflacionado em zero ($b = 0$).	31
4.2	Resultados da simulação de Monte Carlo, modelo IKARMA(2,2) inflacionado em um ($b = 1$).	32
4.3	Resultados da simulação de Monte Carlo, modelo IKARMA(1,1) inflacionado em um ($b = 1$).	33
5.1	Medidas descritivas da variável UR máxima mensal na cidade de Caxias do Sul, RS.	36
5.2	Modelo IKARMA(1,1) ajustado para dados de UR máxima mensal na cidade de Caxias do Sul, RS: estimativas, erros-padrão, estatísticas z , p -valores z , AIC e SIC, estatística de Ljung-Box e p -valor do teste de Ljung-Box.	38
5.3	Medidas de acurácia de predições para os diferentes modelos ajustados para a UR máxima mensal na cidade de Caxias do Sul, RS.	40
5.4	Medidas de acurácia de previsões para os diferentes modelos ajustados para a UR máxima mensal na cidade de Caxias do Sul, RS.	40
5.5	Medidas descritivas da variável VU de São Simão, MG.	42
5.6	Modelo IKARMA(1,1) ajustado para dados de VU de São Simão, MG: estimativas, erros-padrão, estatísticas z , p -valores z , AIC e SIC, estatística de Ljung-Box e p -valor do teste de Ljung-Box.	43
5.7	Medidas de acurácia de predições para os diferentes modelos ajustados para o VU de São Simão, MG.	45
5.8	Medidas de acurácia de previsões para os diferentes modelos ajustados para o VU de São Simão, MG.	45

ÍNDICE

1	INTRODUÇÃO	4
1.1	Objetivos	7
1.1.1	Objetivo Geral	7
1.1.2	Objetivos Específicos	7
1.2	Suporte Computacional	8
1.3	Organização do Texto	8
2	REVISÃO BIBLIOGRÁFICA	9
2.1	Modelo GARMA	9
2.2	Modelo β ARMA	11
2.3	Modelo $I\beta$ ARMA	13
2.4	Modelo KARMA	14
3	O MODELO PROPOSTO	17
3.1	Inferência	20
3.1.1	Estimação Pontual	20
3.1.2	Matriz de Informação de Fisher Condicional	23
3.1.3	Intervalos de Confiança e Testes de Hipóteses	26
3.2	Análise de Diagnóstico e Previsão	27
4	AVALIAÇÃO NUMÉRICA	30
5	APLICAÇÕES	35
5.1	Aplicação de Umidade Relativa do Ar	35
5.2	Aplicação do Percentual de Volume Útil	40
6	CONCLUSÕES	47

Capítulo 1

INTRODUÇÃO

Mudanças no clima global vêm gerando discussões sobre como realizar gestão sustentável dos recursos hídricos, envolvendo questões hidroambientais, com vistas muitas vezes ao processo de tomada de decisões. Tais recursos são essenciais em diversas esferas produtivas, como, por exemplo, na agricultura, criação de animais, sistema industrial, produção de alimentos, geração de energia e principalmente para o consumo humano (Sagrillo et al., 2021).

Os dados hidroambientais englobam uma ampla gama de possibilidades, envolvendo variáveis com diferentes características. Neste cenário, dados duplamente limitados, como taxas e proporções, são usualmente observados e a proposição de modelos específicos para modelá-los se torna útil para uma gestão eficiente dos recursos hídricos. Diversos modelos para dados no intervalo unitário padrão $(0, 1)$ têm recebido destaque na literatura, não somente para modelar dados relativos a recursos hídricos, mas com vistas à modelagem de séries temporais não gaussianas, de forma mais geral. Novos modelos vêm sendo propostos na busca de maior flexibilidade e maior generalidade, possibilitando a modelagem de conjuntos de dados nas mais diversas áreas do conhecimento. Como exemplo, podemos citar os trabalhos desenvolvidos por Fokianos and Kedem (1998); Benjamin et al. (2003); Cribari-Neto and Zeileis (2010); Cribari-Neto and Santos (2019); Bayer et al. (2017); Pumi et al. (2019, 2021), entre outros.

Uma das abordagens que tem recebido destaque nesse contexto se baseia na ideia de introduzir componentes autorregressivas e de médias móveis (ARMA) em uma estrutura mais geral, semelhante àquela utilizada em modelos lineares generalizados (GLM) (McCullagh and Nelder, 1989). Com base nesses modelos, é possível representar adequadamente séries temporais de naturezas diversas sob diferentes distribuições probabilísticas, sendo estas postuladas para a variável de interesse. Estruturas do tipo GLM acomodam a distribuição da série temporal (componente aleatório), permitindo ainda a presença de covariáveis, tendências e padrões heterocedásticos, entre outros; a estrutura ARMA acomoda a presença de correlação serial no processo.

Esses modelos são do tipo *observation-driven*, segundo nomenclatura introduzida em Cox et al. (1981). Eles abrangem dois componentes, um aleatório e um sistemático. O componente aleatório é responsável pelas características distribucionais do modelo, incluindo uma medida

de locação, como a média (condicional) ou a mediana (condicional). O componente sistemático é definido com uma estrutura de dependência que conduz a dinâmica da variável resposta ao longo do tempo. Esses modelos são relativamente simples e têm sido amplamente utilizados na modelagem de séries temporais não gaussianas. São tipicamente baseados na abordagem que hoje é conhecida como ARMA generalizado (GARMA) (Benjamin et al., 2003).

A classe de modelos GARMA estende a de GLMs através da adição de um termo extra ao preditor linear que incorpora componentes autorregressivos (AR) e de médias móveis (MA). Apesar dos modelos GARMA possuírem estrutura bastante geral, eles não conseguem acomodar algumas importantes distribuições para dados duplamente limitados, pois as mesmas não pertencem à família exponencial canônica. Para situações como essas, Rocha and Cribari-Neto (2009) propuseram um modelo em que a variável resposta segue, condicionalmente a um conjunto de informações prévias, distribuição beta, parametrizada em termos da média condicional. Tal modelo é conhecido como modelo beta ARMA (β ARMA). Bayer et al. (2017) introduziram um modelo alternativo considerando a distribuição Kumaraswamy, parametrizada em termos da mediana condicional, sendo tal modelo conhecido como modelo Kumaraswamy ARMA (KARMA). Ambos os modelos são ditos ser tipo-GARMA, não se encaixando especificamente na classe GARMA, pois as distribuições postuladas para a variável de interesse não pertencem à família exponencial canônica, mas as estruturas dinâmicas utilizadas são semelhantes à classe GARMA.

Particularmente, na modelagem de dados hidroambientais duplamente limitados, a distribuição Kumaraswamy mostra-se promissora. Ela foi introduzida por Kumaraswamy (1980), tendo o autor notado que a distribuição beta muitas vezes não apresenta desempenho satisfatório quando usada na modelagem de dados hidrológicos. Adicionalmente, a distribuição Kumaraswamy possui algumas vantagens analíticas em relação à distribuição beta. Por exemplo, suas funções de distribuição e quantílica podem ser expressas em forma fechada, sem dependerem de funções especiais. Assim, sequências de números aleatórios da distribuição Kumaraswamy podem ser facilmente geradas com base no método da inversão. Em contraste, a geração de ocorrências beta requer o uso de algoritmos de aceitação-rejeição, que costumam ser mais custosos computacionalmente. De forma semelhante à distribuição beta, a distribuição Kumaraswamy é bastante flexível para representar o comportamento de taxas e proporções, uma vez que sua função densidade pode assumir muitas formas diferentes dependendo dos valores dos parâmetros. Wang et al. (2017) afirma que a distribuição Kumaraswamy é particularmente útil para modelar variáveis que descrevem fenômenos naturais, biológicos e epidemiológicos duplamente limitados.

Muitos trabalhos envolvendo a distribuição Kumaraswamy vêm sendo desenvolvidos na literatura. Cordeiro and de Castro (2011), com base na distribuição Kumaraswamy, propuseram uma nova família de distribuições generalizadas para estender as distribuições normal, Weibull, gama, Gumbel e gaussiana inversa, entre várias distribuições conhecidas. Lemonte (2011) obteve estimadores aproximadamente não viesados para os parâmetros que indexam a distribuição Kumaraswamy. Um método de seleção entre as distribuições beta e

Kumaraswamy foi proposto por Silva and Barreto-Souza (2014). Sagrillo et al. (2021) transformaram a distribuição Kumaraswamy, obtendo duas novas distribuições com o mesmo número de parâmetros. Mais recentemente, Alduais et al. (2022) estimaram os parâmetros da distribuição Kumaraswamy usando inferência sob uma perspectiva bayesiana.

Outra situação comumente observada envolve dados duplamente limitados no intervalo unitário padrão, ou seja, dados em que há observações iguais a zeros e/ou uns. Tais valores possuem probabilidade não nula de ocorrência e dados que os contêm são ditos ser inflacionados. Pode haver inflacionamento simples (dados em $[0, 1)$ ou $(0, 1]$) ou inflacionamento duplo (dados em $[0, 1]$). Para modelar dados com essas características, as distribuições descritas anteriormente não são adequadas. Para contornar esse problema, modelos que acomodam inflacionamento têm sido propostos. Eles combinam distribuições discretas e contínuas para acomodar dados nos intervalos $[0, 1)$, $(0, 1]$ e/ou $[0, 1]$. Por exemplo, Ospina and Ferrari (2010) introduzem distribuições beta inflacionadas para modelar dados duplamente limitados que contêm zeros e/ou uns. O inflacionamento duplo de dados para o modelo beta em uma perspectiva Bayesiana foi considerado por Galvis et al. (2014); Liu and Kong (2015); Mohsenkhani et al. (2019); Nogarotto et al. (2020). Por outro lado, com foco em modelos de regressão para a média da variável inflacionada, Bayes and Valdivieso (2016) introduziram um modelo de regressão beta inflacionado baseado em uma reparametrização da distribuição beta inflacionada que permite a modelagem da média da variável inflacionada diretamente por meio de variáveis explicativas. Liu and Eugenio (2018) comparam inferências frequentistas e Bayesianas em regressões beta sujeitas a dados com inflacionamento duplo, mas não apresentam expressões em forma fechada para quantidades importantes do modelo, como a matriz de informação de Fisher. Bayer et al. (2023) propuseram o modelo beta inflacionado ARMA ($I\beta$ ARMA) para modelagem e previsão de dados de séries temporais que assumem valores nos intervalos $(0, 1]$, $[0, 1)$ ou $[0, 1]$, ampliando a aplicabilidade dos modelos β ARMA.

Especificamente na área hidroambiental, generalizações da distribuição Kumaraswamy têm sido propostas para acomodar dados inflacionados. Por exemplo, nos trabalhos de Cribari-Neto and Santos (2019) e Bayer et al. (2021) são apresentados a distribuição Kumaraswamy inflacionada e o modelo de regressão Kumaraswamy inflacionado, respectivamente, supondo observações independentes. Esses modelos podem ser úteis, por exemplo, para modelar dados de umidade relativa do ar (UR). Por vezes, o registro da UR pode chegar a 100%, portanto, nesse conjunto de dados haverá inflacionamento simples em um. Outro exemplo importante que podemos citar é o percentual de volume útil (VU) de reservatórios de água de uma usina hidrelétrica. Muitas vezes esse reservatório pode chegar na capacidade mínima, quando há incidência de secas na região, ou ultrapassar sua capacidade de armazenamento, período com grande volume de precipitação. Essas situações extremas podem resultar em dados com inflacionamento em zero e/ou um. Contudo, os modelos baseados na distribuição Kumaraswamy atualmente disponíveis não são adequados para a modelagem de dados inflacionados que são observados temporalmente, com correlação serial.

Neste sentido, o objetivo do presente trabalho é propor uma classe de modelos dinâmicos

para variáveis aleatórias hidroambientais duplamente limitadas sujeitas a inflacionamento simples. Isso se dará assumindo distribuição Kumaraswamy inflacionada para o componente aleatório e considerando um modelo de mistura de distribuições discreta e contínua. No modelo proposto, serão modelados, ao longo do tempo, a mediana condicional da parte contínua e a probabilidade de inflação (parâmetro de mistura). Ou seja, o modelo proposto pode ser utilizado com variáveis aleatórias com evolução temporal que assumem valores em $[0, 1)$ ou em $(0, 1]$. É denominado modelo Kumaraswamy inflacionado autorregressivo de médias móveis (IKARMA) estendendo os modelos KARMA (Bayer et al., 2017), pois uma distribuição discreta é incorporada para acomodar inflacionamento simples. Nosso modelo também estende os modelos inflacionados Kumaraswamy (Cribari-Neto and Santos, 2019; Bayer et al., 2021), pois inclui uma estrutura dinâmica ARMA adotada para acomodar dependência temporal. Ao longo do trabalho, aspectos de inferência, diagnóstico, seleção de modelos e predição serão discutidos. O modelo proposto será aplicado a dados hidroambientais.

1.1 Objetivos

1.1.1 Objetivo Geral

A presente dissertação tem por objetivo principal contribuir na área de modelagem estatística de séries temporais hidroambientais duplamente limitadas e sujeitas a inflacionamento.

1.1.2 Objetivos Específicos

- Estender os trabalhos de Bayer et al. (2017), Cribari-Neto and Santos (2019) e Bayer et al. (2021) através da introdução do modelo IKARMA, com inflação simples em zero ou em um;
- Desenvolver aspectos de estimação pontual via máxima verossimilhança condicional e inferências sobre os parâmetros do modelo proposto;
- Derivar formas matriciais fechadas para o vetor escore condicional e para a matriz de informação de Fisher condicional;
- Implementar os aspectos metodológicos desenvolvidos no ambiente computacional R (R Core Team, 2021) com o intuito de viabilizar aplicações;
- Realizar simulações de Monte Carlo para avaliar o desempenho das inferências em amostras de tamanho finito;
- Aplicar o modelo proposto em diferentes bases de dados hidroambientais, comparando os resultados obtidos àqueles oriundos de modelos disponíveis na literatura, evidenciando a aplicabilidade do modelo proposto.

1.2 Suporte Computacional

Todas as implementações e avaliações numéricas, assim como a elaboração de gráficos usados ao longo da dissertação, foram feitas no ambiente computacional R na versão 4.1.3 (R Core Team, 2021). Tal ambiente computacional encontra-se disponível gratuitamente no site <https://www.r-project.org/>.

O código computacional utilizado para a obtenção de estimativas dos parâmetros que indexam o modelo IKARMA(p, q) no ambiente computacional R encontra-se disponibilizado pelos autores em <https://github.com/fabiobayer/IKARMA>.

1.3 Organização do Texto

A presente dissertação está organizada da seguinte forma. No Capítulo 2, é apresentada uma revisão bibliográfica para o problema de interesse, trazendo-se à tona conceitos relevantes sobre os modelos GARMA, β ARMA, $I\beta$ ARMA e KARMA. Também remetemos o leitor a trabalhos relevantes que se encontram disponíveis na literatura, onde aplicações e generalizações destes modelos são evidenciadas.

No Capítulo 3, o modelo IKARMA é proposto com métodos inferenciais baseados em máxima verossimilhança condicional. Apresentamos expressões para o vetor escore condicional e para a matriz de informação de Fisher condicional. Comentamos também sobre a construção dos intervalos de confiança e acerca de testes de hipóteses. Ainda nesse capítulo, são apresentadas medidas para análise de diagnóstico e seleção do modelo.

No Capítulo 4, são apresentados os resultados de um estudo de simulação de Monte Carlo realizado para avaliar o desempenho dos estimadores de máxima verossimilhança condicional em amostras de tamanho finito.

O Capítulo 5 contém duas aplicações do modelo proposto através da análise da série temporal de umidade relativa do ar, observada em Caxias do Sul, RS (5.1) e outra aplicação sobre o percentual de volume útil da Usina Hidrelétrica São Simão, localizada no estado de Minas Gerais, Brasil (5.2).

Por fim, o Capítulo 6 reúne algumas considerações finais e propostas para pesquisas futuras a serem realizadas como continuação desta dissertação.

Capítulo 2

REVISÃO BIBLIOGRÁFICA

Neste capítulo, serão apresentados alguns estudos que embasam o presente trabalho, visando facilitar a compreensão do leitor. Serão revistas algumas informações fundamentais sobre os modelos GARMA, β ARMA, I β ARMA e KARMA, os quais são nomeados aqui como modelos GARMA ou tipo-GARMA. Os modelos GARMA se referem à proposta de Benjamin et al. (2003), onde é introduzida uma ampla classe de modelos dinâmicos para variáveis aleatórias com distribuição condicional pertencente à família exponencial canônica. Os demais modelos são denotados de forma conjunta como modelos do tipo-GARMA, pois possuem estrutura dinâmica semelhante à dos modelos GARMA, mas a variável aleatória não pertence à família exponencial canônica.

Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo estocástico, $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ um vetor r -dimensional de covariáveis evoluindo no tempo e $\mathcal{F}_{t-1} = \sigma\{Y_{t-1}, Y_{t-2}, \dots, \mathbf{X}_t, \mathbf{X}_{t-1}, \dots\}$ uma sigma-álgebra definida por todas as informações observadas até o instante t . Dessa forma, assume-se que

$$Y_t | \mathcal{F}_{t-1} \sim f(y; \cdot | \mathcal{F}_{t-1}), \quad \mu_t = g(\mathbf{X}_t, \mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-l}, Y_{t-1}, \dots, Y_{t-m}), \quad (2.1)$$

em que $l, m \geq 0$ informam um conjunto de observações defasadas, $f(y; \cdot | \mathcal{F}_{t-1})$ é a densidade condicional de Y_t dado \mathcal{F}_{t-1} , μ_t é a medida de locação condicional de interesse da variável Y_t e $g(\cdot)$ é uma função de ligação monótona e duas vezes diferenciável. Nos modelos descritos a seguir, inclusive no modelo proposto no Capítulo 3, assumem-se densidades condicionais $f(y; \cdot | \mathcal{F}_{t-1})$, assim como a estrutura dinâmica para μ_t pode variar a depender das características dos dados em questão.

2.1 Modelo GARMA

A classe de modelos GARMA, introduzida por Benjamin et al. (2003), é uma extensão dos GLM, no sentido de que introduz uma estrutura dinâmica para modelar dados observados ao longo do tempo. Também é uma extensão dos clássicos modelos da classe ARMA (Box et al., 2015) com distribuição condicional Gaussiana, acomodando distribuições condicionais não Gaussianas, particularmente, da família exponencial canônica. O modelo GARMA pode

ser utilizado para representar a evolução temporal de inúmeras séries temporais, como as que apresentam apenas valores nos reais positivos, séries binárias ou ainda séries que provêm de contagem, sendo possível também adicionar na modelagem covariáveis dependentes do tempo.

Os modelos GARMA são extensões naturais dos modelos não Gaussianos desenvolvidos por Zeger (1988) e aprimorados por Li (1994). Zeger (1988) introduziu os modelos autorregressivos exponenciais, sendo casos particulares do uso das distribuições condicionais Poisson e gama. Mais tarde, Li (1994) retratou uma forma de média móvel ao modelo de Zeger (1988). Esta nova classe de modelos abrangeu o uso de estudos propostos anteriormente às distribuições da família exponencial, trazendo assim novas perspectivas na análise de séries temporais. Seu surgimento possibilitou uma parametrização mais parcimoniosa em relação àquela usada em modelos puramente autorregressivos ou de médias móveis, ao considerar uma combinação desses componentes e, ainda, podendo acomodar comportamentos determinísticos devido à influência de variáveis exógenas.

No modelo GARMA (Benjamin et al., 2003), assume-se que a distribuição condicional é dependente no tempo t , onde a variável resposta, condicional às informações passadas \mathcal{F}_{t-1} , segue distribuição pertencente à família exponencial canônica, com função densidade dada por

$$f(y_t; \nu_t, \varphi | \mathcal{F}_{t-1}) = \exp \left\{ \frac{y_t \nu_t - b(\nu_t)}{\varphi} + d(y_t, \varphi) \right\}, \quad (2.2)$$

em que ν_t e φ são os parâmetros canônico e de dispersão, respectivamente, e as funções $b(\cdot)$ e $d(\cdot)$ definem uma família exponencial específica. Por exemplo, se $Y_t | \mathcal{F}_{t-1}$ tem distribuição Poisson, então $\varphi = 1$, $\nu_t = \log(\mu_t)$, $b(\nu_t) = \exp(\nu_t)$, $d(y_t, \varphi) = -\log(y_t!)$. Já para a distribuição gama, temos $\varphi = 1/\alpha$, $\nu_t = -1/\mu_t$, $b(\nu_t) = -\log(\mu_t)$, $d(y_t, \varphi) = \alpha \log(\alpha) + (\alpha - 1) \log(y_t) - \alpha y_t / \mu_t - \log(\Gamma(\alpha))$.

McCullagh and Nelder (1989) apresentam os termos $\mu_t = b'(\nu_t) = \mathbb{E}(Y_t)$ e $V(\mu_t)$ que representam a média e a função de variância de Y_t . No GLM padrão, a média μ_t apresenta relação com o preditor η_t , que envolve variáveis regressoras e parâmetros desconhecidos. Essa relação ocorre através da função de ligação, a qual é monótona e duas vezes diferenciável, tal que $g(\mu_t) = \eta_t$.

Nos modelos GARMA, o preditor linear possui um componente adicional, τ_t , que permite que dinâmicas AR e MA sejam incorporadas, além de acrescentar no preditor linear valores passados das variáveis explicativas (Benjamin et al., 2003). Assim, o GARMA possui a seguinte estrutura dinâmica, onde μ_t é a média condicional de Y_t :

$$g(\mu_t) = \eta_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \tau_t, \quad t = i, \dots, n, \quad (2.3)$$

em que

$$\tau_t = \sum_{i=1}^p \phi_i A(y_{t-i}, \mathbf{x}_{t-i}, \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j M(y_{t-j}, \mu_{t-j}), \quad (2.4)$$

com $\mathbf{x}_t \in \mathbb{R}^k$, A e M sendo funções referentes aos termos AR e MA, respectivamente, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^\top$, $k < n$, p e $q \in \mathbb{N}$ sendo as ordens AR e MA, cujos vetores paramétricos são $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^\top$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^\top$, respectivamente.

Benjamin et al. (2003) destacam que o modelo descrito na Equação (2.4) é bastante geral e flexível, mas que, na prática, deve-se considerar um modelo mais parcimonioso, como

$$g(\mu_t) = \eta_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \phi_i [g(y_{t-i}) - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j [g(y_{t-j}) - \eta_{t-j}]. \quad (2.5)$$

Portanto, o modelo GARMA(p, q) é definido pelas Equações (2.2) e (2.5), onde o preditor linear está relacionado à média condicional μ_t através de alguma função de ligação. A depender da distribuição assumida para a variável dependente e do respectivo espaço paramétrico da média, diferentes funções de ligação podem ser consideradas. A função de ligação mais usual para as distribuições Poisson e gama é a log ($\eta = \log(\mu)$), já para a distribuição binomial teríamos as ligações logit ($\eta = \log(\mu/(1 - \mu))$), probit ($\eta = \phi^{-1}(\mu)$), e a complemento log-log ($\eta = \log(-\log(1 - \mu))$). Para a distribuição normal a ligação mais usual é a identidade ($\eta = \mu$) e para a normal inversa seria a inversa-quadrática ($\eta = \mu^{-2}$), a qual é a ligação canônica.

Em vista disso, a estrutura apresentada em (2.1), (2.3) e (2.4) é bastante versátil, pois permite a incorporação de certas particularidades de uma distribuição de interesse no componente aleatório do modelo, mantendo a estrutura de dependência na resposta média condicional. Adicionalmente, o componente dinâmico pode conter qualquer estrutura de dependência desejada na variável resposta. Com isso, tem-se uma ferramenta inferencial muito simples e poderosa, com os modelos GARMA apresentando vantagens sobre os modelos clássicos de séries temporais, tais como os modelos da classe ARIMA (Box et al., 2015).

Apesar dos modelos GARMA serem bastante flexíveis e apresentarem muitas vantagens, como descrito anteriormente, quando nos deparamos com variáveis aleatórias duplamente limitadas, estes modelos não são os mais adequados. Nesses casos, as distribuições beta e Kumaraswamy são as mais utilizadas na literatura, porém, elas não pertencem à família exponencial canônica de distribuições, dada na Equação (2.2). Dessa forma, focando na modelagem de séries temporais de dados duplamente limitados, os modelos β ARMA e KARMA foram propostos. Rocha and Cribari-Neto (2009) propuseram um modelo em que a variável de interesse segue distribuição beta, parametrizada em termos da média condicional, e Bayer et al. (2017) apresentaram um modelo considerando a distribuição Kumaraswamy, parametrizada em termos da mediana condicional. Nesses estudos, foi considerada uma estrutura do tipo ARMA para o componente sistemático, de forma semelhante à apresentada na Equação (2.5). A seguir são descritos esses e outros modelos no estado da arte, úteis para modelagem de séries temporais duplamente limitadas.

2.2 Modelo β ARMA

Rocha and Cribari-Neto (2009) propuseram o modelo β ARMA como um modelo dinâmico

de séries temporais duplamente limitadas, onde a variável resposta está restrita ao intervalo $(0, 1)$. O modelo proposto assume que a variável de interesse Y_t segue distribuição beta condicionalmente às observações passadas, $Y_t | \mathcal{F}_{t-1} \sim \text{Beta}(\mu_t, \phi)$, com função densidade dada por

$$f(y_t; \mu_t, \phi | \mathcal{F}_{t-1}) = \frac{\Gamma(\phi)}{\Gamma(\mu_t \phi) \Gamma((1 - \mu_t) \phi)} y_t^{(\mu_t \phi - 1)} (1 - y_t)^{(1 - \mu_t) \phi - 1}, \quad 0 < y_t < 1, \quad (2.6)$$

em que a média e a variância condicionais são, respectivamente,

$$\mathbb{E}(Y_t | \mathcal{F}_{t-1}) = \mu_t$$

e

$$\text{Var}(Y_t | \mathcal{F}_{t-1}) = \frac{V(\mu_t)}{1 + \phi},$$

$V(\mu_t) = \mu_t(1 + \mu_t)$ sendo a função de variância. Quanto maior for o valor do parâmetro de precisão (ϕ) , menor será a variância de Y_t fixando-se μ_t (Ferrari and Cribari-Neto, 2004).

A distribuição beta talvez seja a distribuição mais importante e mais comumente utilizada na literatura para dados com suporte no intervalo unitário $(0, 1)$. Mais detalhes sobre essa distribuição podem ser encontradas em Gupta and Nadarajah (2004). A distribuição beta é bastante flexível, pois sua função de densidade pode assumir formatos variados, conforme os valores dos parâmetros. Tal densidade pode assumir formas simétricas, assimétricas à esquerda ou à direita (Ferrari and Cribari-Neto, 2004).

Seja Y_t , $t = 1, \dots, n$, em que cada Y_t possui densidade condicional definida em (2.6). O modelo βARMA tem a seguinte estrutura dinâmica:

$$g(\mu_t) = \alpha + \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \varphi_i [g(y_{t-i}) - \mathbf{x}_t^\top \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j r_{t-j}, \quad (2.7)$$

em que α é o intercepto, $\mathbf{x}_t \in \mathbb{R}^k$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^\top$, $k < n$, p e $q \in \mathbb{N}$ são as ordens AR e MA, cujos vetores paramétricos são $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_p)^\top$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^\top$, respectivamente. Adicionalmente, r_t é o termo de erro, que pode ser definido na escala original $(y_t - \mu_t)$ ou na escala do preditor $(g(y_t) - \eta_t)$. É possível mostrar que $\mathbb{E}(y_t - \mu_t | \mathcal{F}_{t-1}) = 0$ e $\mathbb{E}(g(y_t) - \eta_t | \mathcal{F}_{t-1}) \approx 0$.

Apesar de o modelo proposto assumir que a variável resposta possui distribuição beta, restrita ao intervalo unitário $(0, 1)$, é possível usar o modelo βARMA com variáveis duplamente limitadas que assumem, mais geralmente, valores em (a, b) , em que a e b são escalares conhecidos e $a < b$. Nesses casos, modela-se $(Y_t - a)/(b - a)$, que assume valores em $(0, 1)$ (Rocha and Cribari-Neto, 2009).

Diversos trabalhos foram desenvolvidos embasados no modelo βARMA . Por exemplo, podemos citar o trabalho desenvolvido por Bayer et al. (2018), onde os autores introduzem o modelo beta sazonal autorregressivo e de médias móveis (βSARMA), para modelagem e previsão de dados de séries temporais que assumem valores no intervalo unitário,

incorporando uma dinâmica sazonal à estrutura dinâmica do modelo. Em Palm and Bayer (2018), são apresentadas correções nos estimadores pontuais para melhorias em intervalos de confiança quando o tamanho da amostra é pequeno. Pumi et al. (2019) propuseram o modelo beta autorregressivo de médias móveis com integração fracionária (β ARFIMA), estendendo o modelo β ARMA ao considerar uma estrutura capaz de acomodar longa dependência serial. Testes de adequabilidade de ajuste no modelo β ARMA são investigados em Scher et al. (2020). Por sua vez, Pumi et al. (2021) substituem a estrutura de médias móveis por um processo caótico de forma que se alterando a transformação utilizada é possível mudar substancialmente as propriedades dos caminhos e da estrutura de dependência do processo resultante. Em Palm et al. (2021a), os autores propõem cinco intervalos de predição para o modelo β ARMA. Mecanismos de seleção de modelos β ARMA são extensivamente investigados em Cribari-Neto et al. (2023).

2.3 Modelo $I\beta$ ARMA

Em diversas situações práticas, dados restritos ao intervalo unitário padrão, como taxas e proporções, apresentam inflacionamento em zeros e/ou uns. Modelos como o descrito anteriormente (β ARMA) não são apropriados nesses casos, pois valores iguais a zero ou um não estão no suporte da distribuição. No caso da distribuição beta, passa-se a ter funções de log-verossimilhanças ilimitadas. A fim de acomodar inflacionamento de dados, Bayer et al. (2023) propuseram o modelo beta inflacionado autorregressivo e de médias móveis ($I\beta$ ARMA), útil para modelagem e previsão de dados de séries temporais que assumem valores nos intervalos $(0, 1]$, $[0, 1)$ ou $[0, 1]$. Tal modelo amplia a aplicabilidade dos modelos β ARMA.

O modelo proposto em Bayer et al. (2023) assume que a variável aleatória $Y_t|\mathcal{F}_{t-1}$ segue distribuição beta inflacionada, segundo a parametrização introduzida em Bayes and Valdivieso (2016), com função densidade de probabilidade dada por

$$f(y_t; \alpha_0, \alpha_1, \mu_t, \phi|\mathcal{F}_{t-1}) = [\alpha_0(1 - \mu_t)]^{I_{\{0\}}(y_t)} [\alpha_1\mu_t]^{I_{\{1\}}(y_t)} [c_t h(y_t; \nu_t, \phi|\mathcal{F}_{t-1})]^{I_{\{0,1\}}(y_t)}, \quad (2.8)$$

em que $\mathbb{I}_A(\cdot)$ é a função indicadora do conjunto A , $\mu_t = \mathbb{E}(Y_t|\mathcal{F}_{t-1}) \in (0, 1)$, $\alpha_0, \alpha_1 \in [0, 1]$,

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

$h(y_t; \nu_t, \phi|\mathcal{F}_{t-1})$ sendo a função de densidade beta parametrizada como na Equação (2.6). Se $Y_t|\mathcal{F}_{t-1} \sim BI(\alpha_0, \alpha_1, \mu_t, \phi)$, em que BI é a distribuição beta inflacionada, então

$$\mathbb{E}(Y_t|\mathcal{F}_{t-1}) = \mu_t \quad \text{e} \quad \text{Var}(Y_t|\mathcal{F}_{t-1}) = \frac{(1 + \alpha_1\phi)}{1 + \phi} \mu_t + \left(\frac{(1 - \alpha_1)^2\phi}{c_t(1 + \phi)} - 1 \right) \mu_t^2.$$

O componente sistemático do modelo especifica a média condicional de Y_t através de uma estrutura do tipo GLM com a adição de um termo extra responsável por capturar uma possível dependência serial na média condicional que, condicional em \mathcal{F}_{t-1} , é assumido ter estrutura

ARMA(p, q):

$$\eta_t = g(\mu_t) = \alpha + \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \varphi_i y_{t-i} + \sum_{j=1}^q \theta_j r_{t-j}, \quad (2.9)$$

em que η_t é preditor linear, α é o intercepto, $\mathbf{x}_t \in \mathbb{R}^k$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)^\top$ é o vetor r -dimensional de parâmetros relacionados às covariáveis, $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_p)^\top$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^\top$ são os vetores de coeficientes AR e MA, respectivamente, e $r_t = y_t - \mu_t$ é o termo de erro.

Destaca-se que, nesse modelo, μ_t é a média condicional da variável aleatória inflacionada e não somente a média da parte contínua. Essa abordagem/parametrização segue a proposta feita em Bayes and Valdivieso (2016) e difere da abordagem introduzida em Ospina and Ferrari (2012). Quando $\alpha_0 > 0$ e $\alpha_1 = 0$, $Y_t | \mathcal{F}_{t-1}$ segue distribuição beta inflacionada em zero. Se $\alpha_0 = 0$ e $\alpha_1 > 0$, a variável aleatória $Y_t | \mathcal{F}_{t-1}$ segue distribuição beta inflacionada em um. Nessa parametrização, as probabilidades de $Y_t | \mathcal{F}_{t-1}$ assumir valores iguais a zero ou um variam com μ_t . O primeiro caso ocorre com probabilidade $\Pr(Y_t = 0 | \mathcal{F}_{t-1}) = \alpha_0(1 - \mu_t)$ e o segundo caso, com probabilidade $\Pr(Y_t = 1 | \mathcal{F}_{t-1}) = \alpha_1 \mu_t$. Logo, quanto mais próximo das bordas do intervalo unitário padrão estiver a média condicional, maior será a probabilidade de inflacionamento (Bayer et al., 2023).

A Equação (2.9) é semelhante ao componente sistemático definido em Rocha and Cribari-Neto (2009), mas com os termos AR e MA na escala da variável resposta. Com essa especificação, problemas ao computar $g(y_t)$ são evitados, quando y_t é igual a 0 ou 1, o que ocorre com probabilidade positiva em processos $I\beta$ ARMA(p, q) (Bayer et al., 2023). Essa abordagem, onde termos autorregressivos são empregados no nível da variável resposta, é muito usada em outras classes de modelos, como em modelos para séries temporais seguindo a forma canônica da família exponencial de distribuições (Fokianos and Kedem, 2004; Kedem and Fokianos, 2005) e, mais recentemente, em modelos para séries temporais discretas duplamente limitadas usando distribuição beta binomial (Palm et al., 2021b). Uma das decorrências deste fato é que valores elevados dos coeficientes AR não estão necessariamente associados a comportamento não estacionário ou explosivo, como nos modelos ARMA usuais (Bayer et al., 2023).

2.4 Modelo KARMA

A flexibilidade da distribuição beta a torna muito versátil, podendo ser usada em uma ampla gama de aplicações (Jones, 2009; Lemonte et al., 2013). Entretanto, ela muitas vezes não apresenta ajuste satisfatório quando ajusta a dados hidroambientais (Kumaraswamy, 1980; Lemonte et al., 2013).

Para modelagens hidroambientais e em áreas afins, a distribuição Kumaraswamy é geralmente preferida à distribuição beta (Nadarajah, 2008; Lemonte et al., 2013). Diversos trabalhos que aplicam a distribuição Kumaraswamy a dados dessa natureza podem ser encontrados na literatura (Nadarajah, 2008). Nessa perspectiva, os modelos β ARMA e

$I\beta$ ARMA podem não ser os mais adequados para análises empíricas que envolvem dados hidroambientais. Visando a preencher essa lacuna na literatura, Bayer et al. (2017) propuseram uma alternativa ao modelo β ARMA para séries hidroambientais.

O modelo KARMA (Bayer et al., 2017) é um modelo dinâmico de séries temporais para variáveis aleatórias duplamente limitadas com suporte em $(0, 1)$. Assume-se que a variável de interesse Y_t , condicional a informações pretéritas, possui distribuição Kumaraswamy, i.e., $Y_t|\mathcal{F}_{t-1} \sim K(\mu_t, \varphi)$, em que μ_t é a mediana condicional. As funções densidade de probabilidade e de distribuição cumulativa são dadas, respectivamente, por (Mitnik and Baek, 2013)

$$f(y_t; \mu_t, \varphi | \mathcal{F}_{t-1}) = \frac{\varphi \log(0.5)}{\log(1 - \mu_t^\varphi)} y_t^{\varphi-1} (1 - y_t^\varphi)^{\frac{\log(0.5)}{\log(1 - \mu_t^\varphi)} - 1}, \quad (2.10)$$

$$F(y_t; \mu_t, \varphi | \mathcal{F}_{t-1}) = 1 - (1 - y_t^\varphi)^{\frac{\log(0.5)}{\log(1 - \mu_t^\varphi)}}, \quad (2.11)$$

para $0 < y_t < 1$, com $0 < \mu_t < 1$ e $\varphi > 0$. A função quantílica de Y_t pode ser facilmente obtida invertendo $F(y_t; \mu_t, \varphi | \mathcal{F}_{t-1})$, uma vez que esta é parametrizada através da mediana. A função quantílica é dada por

$$F^{-1}(u; \mu_t, \varphi | \mathcal{F}_{t-1}) = \left[1 - (1 - u)^{\frac{\log(1 - \mu_t^\varphi)}{\log(0.5)}} \right]^{\frac{1}{\varphi}}, \quad (2.12)$$

em que $\mu_t = \left(1 - 0.5^{\frac{\log(1 - \mu_t^\varphi)}{\log(0.5)}} \right)^{\frac{1}{\varphi}}$ é a mediana condicional. A média e a variância condicionais são dadas, respectivamente, por

$$\mathbb{E}(Y_t | \mathcal{F}_{t-1}) = \frac{\log(0.5)}{\log(1 - \mu_t^\varphi)} B \left(1 + \frac{1}{\varphi}, \frac{\log(0.5)}{\log(1 - \mu_t^\varphi)} \right),$$

$$\text{Var}(Y_t | \mathcal{F}_{t-1}) = \left\{ \frac{\log(0.5)}{\log(1 - \mu_t^\varphi)} B \left(1 + \frac{2}{\varphi}, \frac{\log(0.5)}{\log(1 - \mu_t^\varphi)} \right) - \left[\frac{\log(0.5)}{\log(1 - \mu_t^\varphi)} B \left(1 + \frac{1}{\varphi}, \frac{\log(0.5)}{\log(1 - \mu_t^\varphi)} \right) \right]^2 \right\},$$

em que $B(\cdot, \cdot)$ é a função beta (Gupta and Nadarajah, 2004). Pode-se observar que a avaliação da mediana condicional é mais simples que as da esperança e variância condicionais de Y_t . Então, os autores propuseram a seguinte especificação dinâmica para a mediana condicional da variável de interesse:

$$\eta_t = g(\mu_t) = \alpha + \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \phi_i [g(y_{t-i}) - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j r_{t-j}, \quad (2.13)$$

em que η_t é o preditor linear, \mathbf{x}_t é o vetor r -dimensional de covariáveis conhecidas no tempo t , $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)^\top$ é o vetor r -dimensional de parâmetros relacionados às covariáveis e $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^\top$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^\top$ são os vetores de coeficientes AR e MA, respectivamente.

Pode-se observar que a parte dinâmica do modelo KARMA, dada na Equação (2.13), é a mesma utilizada por Rocha and Cribari-Neto (2009). No entanto, a componente aleatória do modelo, na Equação (2.10), é diferente e está parametrizada em função da mediana condicional.

Conforme observam Lemonte and Bazán (2016), métodos de regressão baseados na mediana são tipicamente robustos contra observações atípicas. Em comparação com os modelos baseados na média, aqueles especificados pela mediana tendem a apresentar melhor desempenho quando a distribuição da população é assimétrica. Como vantagem adicional, no modelo KARMA as funções de distribuição e quantílica da distribuição da variável de interesse podem ser expressas em forma fechada, sendo parametrizadas em termos da mediana.

Especificamente na área hidroambiental, o modelo KARMA é adequado para dados que assumem valores no intervalo padrão unitário $(0,1)$. Em situações extremas, esses dados podem apresentar inflacionamentos, contendo observações iguais a zero e/ou um. Logo, o modelo KARMA não é adequado para a modelagem desses dados inflacionados observados temporalmente. Neste sentido, proporemos a seguir uma classe de modelos dinâmicos para variáveis hidroambientais duplamente limitadas sujeitas a inflacionamento simples.

Capítulo 3

O MODELO PROPOSTO

Questões hidroambientais têm gerado muitas discussões em relação a mudanças no clima global e sobre como promover uma gestão sustentável dos recursos hídricos. Na modelagem de dados hidroambientais duplamente limitados, a distribuição Kumaraswamy tipicamente apresenta desempenho satisfatório comparativamente à distribuição beta (Kumaraswamy, 1980). Neste sentido, os modelos β ARMA e $I\beta$ ARMA podem não ser os mais adequados para análises de dados envolvendo questões hidroambientais. O modelo KARMA foi introduzido por Bayer et al. (2017) para modelagem de dados hidroambientais, onde os autores propuseram um modelo dinâmico de séries temporais para variáveis duplamente limitadas com suporte em $(0, 1)$ assumindo que a variável resposta possui distribuição condicional Kumaraswamy.

Em diversas situações práticas, dados hidroambientais podem apresentar observações iguais a zero e/ou um. Nesses casos, o modelo KARMA não pode ser usado, por ser definido apenas no intervalo aberto entre 0 e 1. Para contornar esse problema, propomos um modelo dinâmico, onde uma combinação de uma distribuição discreta e outra contínua será considerada para acomodar ocorrências de dados nos intervalos $[0, 1)$ ou $(0, 1]$. O modelo proposto será denominado modelo Kumaraswamy inflacionado autorregressivo de médias móveis (IKARMA).

Suponha que $\{Y_t\}_{t \in \mathbb{Z}}$ é um processo estocástico, onde $Y_t \in [0, 1)$ ou $(0, 1]$ com probabilidade 1, $\forall t \in \mathbb{Z}$, $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ é um vetor r -dimensional de covariáveis evoluindo no tempo e $\mathcal{F}_{t-1} = \sigma\{Y_{t-1}, Y_{t-2}, \dots, \mathbf{X}_t, \mathbf{X}_{t-1}, \dots\}$ é uma sigma-álgebra gerada por todas as informações observadas até o instante t . Suponha que, condicional ao conjunto de informações passadas \mathcal{F}_{t-1} , Y_t segue distribuição Kumaraswamy inflacionada em zero ou um (Cribari-Neto and Santos, 2019). A função densidade de probabilidade é expressa por

$$ki(y_t; \lambda_t, b, \mu_t, \varphi | \mathcal{F}_{t-1}) = \begin{cases} \lambda_t(1 - b), & \text{se } y_t = 0, \\ \lambda_t b, & \text{se } y_t = 1, \\ (1 - \lambda_t)f(y_t; \mu_t, \varphi | \mathcal{F}_{t-1}), & \text{se } y_t \in (0, 1), \end{cases}$$

em que $0 < \lambda_t < 1$ é o parâmetro de mistura que assumimos variar ao longo das observações, b é a probabilidade de que uma variável aleatória com distribuição Bernoulli seja igual a um ($b = 1$ ou $b = 0$), e $f(y_t; \mu_t, \varphi | \mathcal{F}_{t-1})$ é a função densidade de probabilidade Kumaraswamy dada

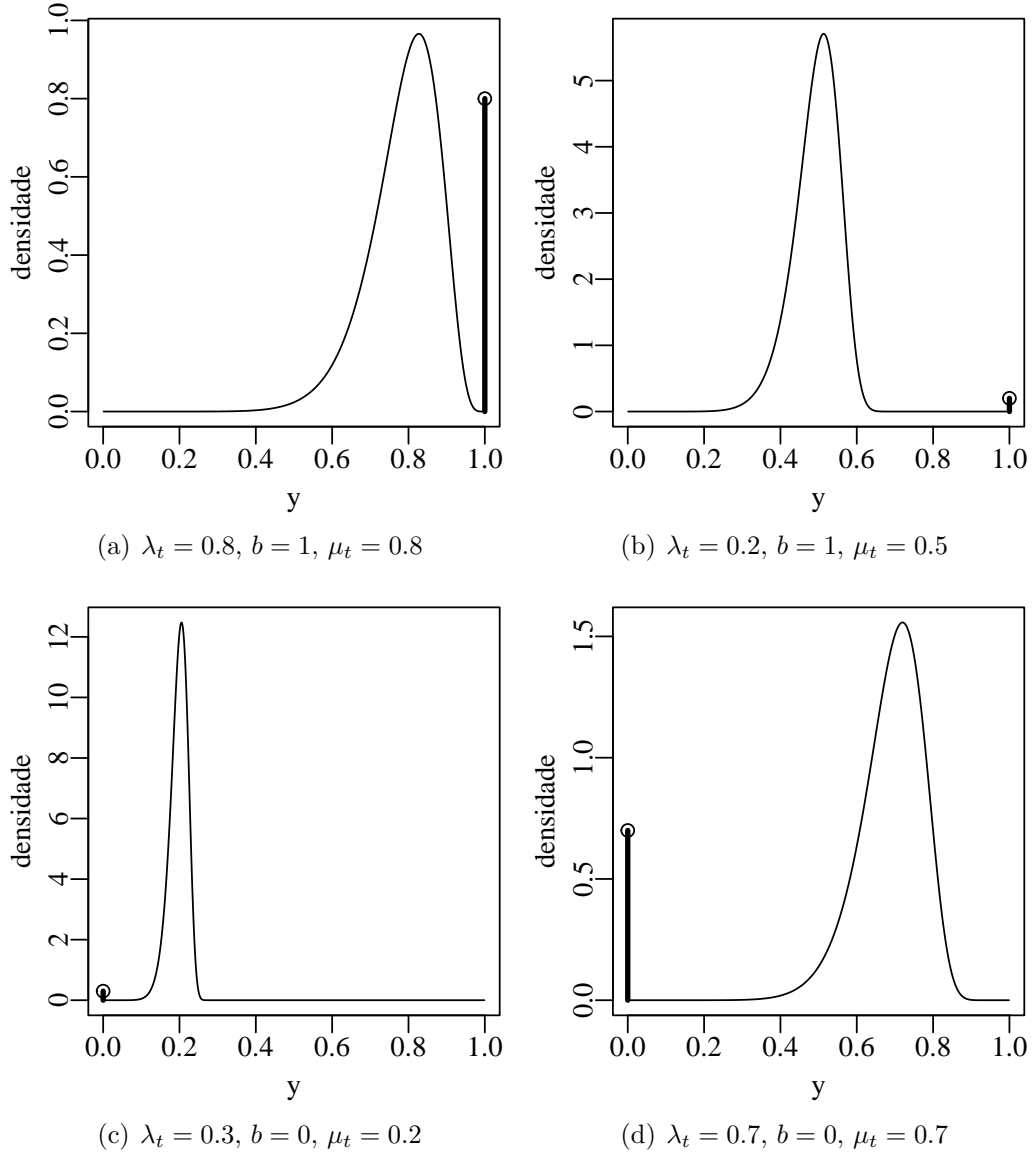


Figura 3.1: Função densidade de probabilidade $ki(\lambda_t, b, \mu_t, \varphi | \mathcal{F}_{t-1})$, com $\varphi = 10$ e diferentes valores dos parâmetros λ_t e μ_t .

na Equação (2.10), indexada pela mediana μ_t , onde $0 < \mu_t < 1$, $\varphi > 0$ sendo o parâmetro de precisão. Podemos reescrever a expressão acima como

$$ki(y_t; \lambda_t, b, \mu_t, \varphi | \mathcal{F}_{t-1}) = [\lambda_t(1-b)]^{\mathbb{I}_{\{0\}}(y_t)} \times [\lambda_t b]^{\mathbb{I}_{\{1\}}(y_t)} \times [(1-\lambda_t)f(y_t; \mu_t, \varphi | \mathcal{F}_{t-1})]^{\mathbb{I}_{\{0,1\}}(y_t)}, \quad (3.1)$$

$0 \leq y_t \leq 1$, em que $\mathbb{I}_A(y)$ é a função indicadora, sendo igual a um se $y \in A$ e igual a zero caso contrário. A Figura 3 apresenta gráficos de algumas funções densidade de probabilidade Kumaraswamy inflacionada.

Dizemos que $Y_t | \mathcal{F}_{t-1}$ segue distribuição Kumaraswamy inflacionada em zero ou em um e denotamos por $Y_t | \mathcal{F}_{t-1} \sim ki(\lambda_t, b, \mu_t, \varphi)$. Observe que $\Pr(Y_t = 1 | \mathcal{F}_{t-1}) = \lambda_t b$ e $\Pr(Y_t = 0 | \mathcal{F}_{t-1}) = \lambda_t(1-b)$. Então, como casos particulares, temos que quando $b = 0$, obtemos a distribuição Kumaraswamy inflacionada em zero, $Y_t | \mathcal{F}_{t-1} \sim ki_0(\lambda_t, \mu_t, \varphi)$; quando

$b = 1$ obtemos a distribuição Kumaraswamy inflacionada em um, $Y_t|\mathcal{F}_{t-1} \sim ki_1(\lambda_t, \mu_t, \varphi)$. Adicionalmente, quando $b = \lambda_t = 0$, $Y_t|\mathcal{F}_{t-1}$ segue distribuição Kumaraswamy conforme a parametrização introduzida por Mitnik (2013), apresentada na Equação (2.10).

As funções de distribuição acumulada e quantílica de Y_t , condicionalmente em \mathcal{F}_{t-1} , são dadas, respectivamente, por

$$KI(y_t; \lambda_t, b, \mu_t, \varphi|\mathcal{F}_{t-1}) = \lambda_t(1-b)\mathbb{I}_{\{0\}}(y_t) + (\lambda_t b)\mathbb{I}_{\{1\}}(y_t) + (1-\lambda_t)F(y_t; \mu_t, \varphi|\mathcal{F}_{t-1}), \quad 0 \leq y_t \leq 1,$$

$$KI^{-1}(u; \lambda_t, b, \mu_t, \varphi|\mathcal{F}_{t-1}) = \begin{cases} 0, & \text{se } u \leq \lambda_t(1-b), \\ F^{-1}\left(\frac{u-\lambda_t(1-b)}{1-\lambda_t}; \mu_t, \varphi|\mathcal{F}_{t-1}\right), & \text{caso contrário,} \\ 1, & \text{se } u \geq 1-\lambda_t b, \end{cases}$$

em que $0 < u < 1$, $F(y_t; \mu_t, \varphi|\mathcal{F}_{t-1})$ e $F^{-1}\left(\frac{u-\lambda_t(1-b)}{1-\lambda_t}; \mu_t, \varphi|\mathcal{F}_{t-1}\right)$ são as funções acumulada e quantílica da distribuição Kumaraswamy, respectivamente, dadas pelas Equações (2.11) e (2.12).

Considerando b fixo/conhecido quando há inflacionamento simples ($b = 0$ ou $b = 1$) e sendo $g_1 : (0, 1) \rightarrow \mathbb{R}$ e $g_2 : (0, 1) \rightarrow \mathbb{R}$ duas funções de ligação estritamente monótonas e duas vezes diferenciáveis, tais que suas funções inversas $g_1^{-1} : \mathbb{R} \rightarrow (0, 1)$ e $g_2^{-1} : \mathbb{R} \rightarrow (0, 1)$ também são duas vezes diferenciáveis, propomos as seguintes estruturas dinâmicas para o parâmetro de mistura e para a mediana condicional, λ_t e μ_t , respectivamente:

$$\eta_{1t} = g_1(\lambda_t) = \omega_1 + \omega_2 g_1(s_{t-1}^*), \quad (3.2)$$

$$\eta_{2t} = g_2(\mu_t) = \alpha + \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \phi_i [g_2(y_{t-i}^*) - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j r_{t-j}, \quad (3.3)$$

em que η_{1t} e η_{2t} são os preditores lineares; ω_1 e ω_2 são os parâmetros da estrutura dinâmica de λ_t ; $s_{t-1}^* = \min[\max(s_{t-1}; 0.5/n); (n-0.5)/n]$, com $s_{t-1} = |y_t - (1-b)| \in [0, 1]$; α é o intercepto da estrutura dinâmica da mediana; \mathbf{x}_t é o vetor r -dimensional de covariáveis conhecidas no tempo t ; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)^\top$ é o vetor r -dimensional de parâmetros relacionados às covariáveis; $y_{t-1}^* = \min[\max(y_{t-1}; 0.5/n); (n-0.5)/n]$; $r_t = g_2(y_t^*) - g_2(\mu_t)$ é o termo de erro; $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^\top$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^\top$ são os vetores de coeficientes AR e MA, respectivamente. Todas as funções de ligação tradicionais, como logit, probit, loglog, cloglog e Cauchit, podem ser utilizadas para $g_1(\cdot)$ ou $g_2(\cdot)$. As quantidades s_{t-1}^* e y_{t-1}^* são assim consideradas para contornar problemas no cômputo de $g_1(s_t)$ e $g_2(y_t)$, quando s_t e y_t são iguais a zero ou um, o que ocorre com probabilidade não nula em processos IKARMA. Esta é uma solução apontada por Benjamin et al. (2003) nos modelos GARMA para evitar inexistência de $g(y_{t-j})$ em (2.5) para certos valores de y_{t-j} , a depender da distribuição e da função de ligação assumidas.

O modelo IKARMA diferencia-se de qualquer outro modelo na literatura, pois apresenta duas estruturas dinâmicas, a saber: uma para λ_t , com uma estrutura simples e parcimoniosa, e uma estrutura do tipo ARMA para a mediana condicional (μ_t). Isso permite que o modelo

proposto seja bastante flexível e versátil, podendo ser aplicado em diferentes cenários.

3.1 Inferência

Nesta seção, apresentaremos a função de log-verossimilhança condicional utilizada na estimação dos parâmetros do modelo, desenvolveremos as teorias de estimação pontual e intervalar e apresentaremos testes de hipóteses. Serão apresentadas formas fechadas para o vetor escore e para a matriz de informação de Fisher condicionais.

Seja y_1, \dots, y_n uma amostra do modelo IKARMA(p, q), definido nas Equações (3.1), (3.2) e (3.3), e seja $\boldsymbol{\gamma} = (\omega_1, \omega_2, b, \alpha, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top, \varphi)^\top$ o vetor de parâmetros do modelo. Os estimadores de máxima verossimilhança condicional (EMVC) são obtidos maximizando-se o logaritmo da função de verossimilhança condicional. Observamos que a função de log-verossimilhança para $\boldsymbol{\gamma}$, condicionalmente em \mathcal{F}_{t-1} , é nula para os primeiros $m = \max(p, q)$ valores de t e, portanto, temos

$$\ell \equiv \ell(\boldsymbol{\gamma}; \mathbf{y}) = \sum_{t=m+1}^n \log [ki(y_t; \lambda_t, b, \mu_t, \varphi | \mathcal{F}_{t-1})] = \sum_{t=m+1}^n [\ell_{1t}(\lambda_t) + \ell_{2t}(\mu_t, \varphi) + \ell_{3t}(b)], \quad (3.4)$$

em que

$$\begin{aligned} \ell_{1t}(\lambda_t) &= \mathbb{I}_{\{0,1\}}(y_t) \log(\lambda_t) + [1 - \mathbb{I}_{\{0,1\}}(y_t)] \log(1 - \lambda_t), \\ \ell_{2t}(\mu_t, \varphi) &= [1 - \mathbb{I}_{\{0,1\}}(y_t)] \left[\log(\varphi) + \log\left(\frac{\log(0.5)}{\log(1 - \mu_t^\varphi)}\right) + (\varphi - 1) \log(y_t) + \left(\frac{\log(0.5)}{\log(1 - \mu_t^\varphi)} - 1\right) \log(1 - y_t^\varphi) \right], \\ \ell_{3t}(b) &= \mathbb{I}_{\{1\}}(y_t) \log(b) + \mathbb{I}_{\{0\}}(y_t) \log(1 - b). \end{aligned}$$

Os estimadores dos parâmetros do modelo IKARMA(p, q) são obtidos através da maximização da função dada em (3.4).

3.1.1 Estimação Pontual

O vetor escore condicional é o gradiente (ou seja, o vetor de primeiras derivadas) da função de log-verossimilhança condicional em relação aos parâmetros que estão sendo estimados. Derivando a função de log-verossimilhança condicional, dada na Equação (3.4), em relação ao j -ésimo elemento do vetor de parâmetros $\boldsymbol{\gamma}$, temos os seguintes componentes do vetor escore:

$$\begin{aligned} U_{\omega_1}(\boldsymbol{\gamma}) &= \frac{\partial \ell(\boldsymbol{\gamma}; \mathbf{y})}{\partial \omega_1} = \sum_{t=m+1}^n \frac{\partial \ell_{1t}(\lambda_t)}{\partial \lambda_t} \frac{d\lambda_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \omega_1}, \\ U_{\omega_2}(\boldsymbol{\gamma}) &= \frac{\partial \ell(\boldsymbol{\gamma}; \mathbf{y})}{\partial \omega_2} = \sum_{t=m+1}^n \frac{\partial \ell_{1t}(\lambda_t)}{\partial \lambda_t} \frac{d\lambda_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \omega_2}, \\ U_{\alpha}(\boldsymbol{\gamma}) &= \frac{\partial \ell(\boldsymbol{\gamma}; \mathbf{y})}{\partial \alpha} = \sum_{t=m+1}^n \frac{\partial \ell_{2t}(\mu_t, \varphi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \alpha}, \end{aligned}$$

$$U_{\beta_l}(\boldsymbol{\gamma}) = \frac{\partial \ell(\boldsymbol{\gamma}; \mathbf{y})}{\partial \beta_l} = \sum_{t=m+1}^n \frac{\partial \ell_{2t}(\mu_t, \varphi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \beta_l},$$

$$U_{\phi_i}(\boldsymbol{\gamma}) = \frac{\partial \ell(\boldsymbol{\gamma}; \mathbf{y})}{\partial \phi_i} = \sum_{t=m+1}^n \frac{\partial \ell_{2t}(\mu_t, \varphi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \phi_i},$$

$$U_{\theta_j}(\boldsymbol{\gamma}) = \frac{\partial \ell(\boldsymbol{\gamma}; \mathbf{y})}{\partial \theta_j} = \sum_{t=m+1}^n \frac{\partial \ell_{2t}(\mu_t, \varphi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \theta_j},$$

$$U_{\varphi}(\boldsymbol{\gamma}) = \frac{\partial \ell(\boldsymbol{\gamma}; \mathbf{y})}{\partial \varphi} = \sum_{t=m+1}^n \frac{\partial \ell_{2t}(\mu_t, \varphi)}{\partial \varphi}.$$

Podemos escrever

$$\frac{d\lambda_t}{d\eta_{1t}} = \frac{1}{g'_1(\lambda_t)},$$

$$\frac{d\mu_t}{d\eta_{2t}} = \frac{1}{g'_2(\mu_t)},$$

$$\frac{\partial \ell_{1t}(\lambda_t)}{\partial \lambda_t} = \frac{\mathbb{I}_{\{0,1\}}(y_t) - \lambda_t}{\lambda_t(1 - \lambda_t)} = a_t,$$

$$\frac{\partial \ell_{2t}(\mu_t, \varphi)}{\partial \mu_t} = [1 - \mathbb{I}_{\{0,1\}}(y_t)] \left[\varphi \frac{\mu_t^{\varphi-1}}{(1 - \mu_t^{\varphi}) \log(1 - \mu_t^{\varphi})} [\delta_t \log(1 - y_t^{\varphi}) + 1] \right] = [1 - \mathbb{I}_{\{0,1\}}(y_t)] \varphi c_t,$$

em que

$$c_t = \frac{\mu_t^{\varphi-1}}{(1 - \mu_t^{\varphi}) \log(1 - \mu_t^{\varphi})} [\delta_t \log(1 - y_t^{\varphi}) + 1] \quad \text{e} \quad \delta_t = \frac{\log(0.5)}{\log(1 - \mu_t^{\varphi})}.$$

As derivadas de $\ell(\boldsymbol{\gamma}; \mathbf{y})$ em relação a ω_1 e ω_2 envolvem as derivadas de η_{1t} com relação as essas quantidades, que são dadas por

$$\frac{\partial \eta_{1t}}{\partial \omega_1} = 1 \quad \text{e} \quad \frac{\partial \eta_{1t}}{\partial \omega_2} = g_1(s_{t-1}^*).$$

A derivada de $\ell(\boldsymbol{\gamma}; \mathbf{y})$ em relação a α , sendo $r_t = g(y_t^*) - g(\mu_t)$ o termo de erro, envolve

$$\frac{\partial \eta_{2t}}{\partial \alpha} = 1 + \sum_{j=1}^q \theta_j \frac{\partial r_{t-j}}{\partial \alpha} = 1 - \sum_{j=1}^q \theta_j \frac{\partial \eta_{2t-j}}{\partial \alpha}.$$

A derivada de $\ell(\boldsymbol{\gamma}; \mathbf{y})$ em relação a β_l , para $l = 1, \dots, r$, envolve

$$\frac{\partial \eta_{2t}}{\partial \beta_l} = \mathbf{x}_{tl} - \sum_{i=1}^p \phi_i \mathbf{x}_{(t-i)l} - \sum_{j=1}^q \theta_j \frac{\partial \eta_{2t-j}}{\partial \beta_l},$$

em que \mathbf{x}_{tl} é o l -ésimo elemento de \mathbf{x}_t . A derivada de $\ell(\boldsymbol{\gamma}; \mathbf{y})$ em relação a ϕ_i , para $i = 1, \dots, p$, envolve

$$\frac{\partial \eta_{2t}}{\partial \phi_i} = g(y_{t-i}^*) - \mathbf{x}_{t-i}^{\top} \boldsymbol{\beta} - \sum_{j=1}^q \theta_j \frac{\partial \eta_{2t-j}}{\partial \phi_i},$$

enquanto tal derivada em relação a θ_j , para $j = 1, \dots, q$, envolve

$$\frac{\partial \eta_{2t}}{\partial \theta_j} = r_{t-j} - \sum_{j=1}^q \theta_j \frac{\partial \eta_{2t-j}}{\partial \theta_j}.$$

Para a derivada de $\ell(\boldsymbol{\gamma}; \mathbf{y})$ em relação a φ , podemos calcular a derivada de (3.4) diretamente, de tal forma que

$$\frac{\partial \ell_{2t}(\mu_t, \varphi)}{\partial \varphi} = [1 - \mathbb{I}_{\{0,1\}}(y_t)] \left[\frac{1}{\varphi} + \log(y_t) + c_t \mu_t \log(\mu_t) - (\delta_t - 1) \frac{y_t^\varphi \log(y_t)}{(1 - y_t^\varphi)} \right].$$

Em forma matricial, o vetor escore

$$\mathbf{U}(\boldsymbol{\gamma}) = (U_{\omega_1}(\boldsymbol{\gamma}), U_{\omega_2}(\boldsymbol{\gamma}), U_\alpha(\boldsymbol{\gamma}), \mathbf{U}_\beta(\boldsymbol{\gamma})^\top, \mathbf{U}_\phi(\boldsymbol{\gamma})^\top, \mathbf{U}_\theta(\boldsymbol{\gamma})^\top, U_\varphi(\boldsymbol{\gamma}))^\top$$

pode ser expresso pelos seguintes componentes:

$$U_{\omega_1}(\boldsymbol{\gamma}) = \mathbf{a}^\top \mathbf{T}_1 \boldsymbol{\iota}, \quad U_{\omega_2}(\boldsymbol{\gamma}) = \mathbf{a}^\top \mathbf{T}_1 \mathbf{m}, \quad U_\alpha(\boldsymbol{\gamma}) = \boldsymbol{\nu}^\top \mathbf{T}_2 \mathbf{c},$$

$$\mathbf{U}_\beta(\boldsymbol{\gamma}) = \mathbf{M}^\top \mathbf{T}_2 \mathbf{c}, \quad \mathbf{U}_\phi(\boldsymbol{\gamma}) = \mathbf{P}^\top \mathbf{T}_2 \mathbf{c}, \quad \mathbf{U}_\theta(\boldsymbol{\gamma}) = \mathbf{R}^\top \mathbf{T}_2 \mathbf{c},$$

$$\text{e } U_\varphi(\boldsymbol{\gamma}) = \frac{n-m}{\varphi} + \sum_{t=m+1}^n \left[\log(y_t) + c_t \mu_t \log(\mu_t) - (\delta_t - 1) \frac{y_t^\varphi \log(y_t)}{(1 - y_t^\varphi)} \right],$$

com $\mathbf{a} = (a_{m+1}, \dots, a_n)^\top$, $\boldsymbol{\iota} = (1, \dots, 1)^\top$, $\mathbf{m} = (g_1(s_m^*), \dots, g_1(s_{n-1}^*))^\top$, $\boldsymbol{\nu} = \left(\frac{\partial \eta_{2m+1}}{\partial \alpha}, \dots, \frac{\partial \eta_{2n}}{\partial \alpha} \right)^\top$, $\mathbf{c} = ([1 - \mathbb{I}_{\{0,1\}}(y_{m+1})] \varphi c_{m+1}, \dots, [1 - \mathbb{I}_{\{0,1\}}(y_n)] \varphi c_n)^\top$, $\mathbf{T}_1 = \text{diag} \{1/g'_1(\lambda_{m+1}), \dots, 1/g'_1(\lambda_n)\}$, $\mathbf{T}_2 = \text{diag} \{1/g'_2(\mu_{m+1}), \dots, 1/g'_2(\mu_n)\}$ e $\mathbf{M}, \mathbf{P}, \mathbf{R}$ são matrizes de dimensões $(n-m) \times r$, $(n-m) \times p$ e $(m-n) \times q$, respectivamente, com elementos (i, j) dados por

$$M_{i,j} = \frac{\partial \eta_{2i+m}}{\partial \beta_j}, \quad P_{i,j} = \frac{\partial \eta_{2i+m}}{\partial \phi_j} \quad \text{e} \quad R_{i,j} = \frac{\partial \eta_{2i+m}}{\partial \theta_j}.$$

O EMVC de $\boldsymbol{\gamma}$, se existir, é obtido como solução do sistema $\mathbf{U}(\boldsymbol{\gamma}) = \mathbf{0}$, em que $\mathbf{0}$ é o vetor nulo em $\mathbb{R}^{r+p+q+5}$. Tal sistema não tem uma solução de forma fechada. As estimativas de máxima verossimilhança condicional são obtidas numericamente usando um algoritmo de otimização não linear como método de Newton ou quasi-Newton. Neste trabalho será utilizado um algoritmo quasi-Newton conhecido como método de Broyden-Fletcher-Goldfarb-Shanno (BFGS) (Press et al., 2007), com primeiras derivadas analíticas. Tal algoritmo exige que valores iniciais sejam fornecidos. O valor inicial para ω_1 foi definido como $g_1(\bar{p})$, em que \bar{p} é a proporção amostral de valores iguais a zero ou um. Para ω_2 e os parâmetros MA, θ_j , com $j = 1, \dots, q$, foram atribuídos valores iniciais iguais a zero. O intercepto α , os parâmetros dos regressores, β_j , com $j = 1, \dots, r$, e os parâmetros autorregressivos, (ϕ_j) , com $j = 1, \dots, p$, foram inicializados a partir de uma estimativa de mínimos quadrados ordinários de uma regressão linear, onde

$[g_2(y_{m+1}), g_2(y_{m+2}), \dots, g_2(y_n)]^\top$ é o vetor de respostas e

$$\begin{bmatrix} 1 & x_{m1} & x_{m2} & \dots & x_{mr} & g_2(y_m) & g_2(y_{m-1}) & \dots & g_2(y_{m-p+1}) \\ 1 & x_{(m+1)1} & x_{(m+1)2} & \dots & x_{(m+1)r} & g_2(y_{m+1}) & g_2(y_m) & \dots & g_2(y_{m-p+2}) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nr} & g_2(y_{n-1}) & g_2(y_{n-2}) & \dots & g_2(y_{n-p}) \end{bmatrix}$$

é a matriz de regressores.

O parâmetro de precisão, φ , é inicializado em $0.2 \times \check{\varphi}$, onde $\check{\varphi}$ é igual ao valor inicial para o parâmetro de precisão no modelo de regressão beta, como em Ferrari and Cribari-Neto (2004).

3.1.2 Matriz de Informação de Fisher Condicional

Nesta seção, serão derivadas as quantidades necessárias para se obter uma expressão em forma fechada para a matriz de informação de Fisher condicional. Esta matriz é usada para se obter erros-padrão para os estimadores, na construção de intervalos de confiança aproximados e na realização de testes de hipóteses. Para isto, é necessário calcular os valores esperados de todas as derivadas de segunda ordem da função de log-verossimilhança condicional em relação a cada parâmetro do modelo.

Sejam $\boldsymbol{\omega} = (\omega_1, \omega_2)^\top$, $\boldsymbol{\gamma}_i \in \{\alpha, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top\}$ e $\boldsymbol{\gamma}_j \in \{\alpha, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top\}$. É possível mostrar que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\gamma}; \mathbf{y})}{\partial \omega_i \partial \omega_j} &= \sum_{t=m+1}^n \frac{\partial}{\partial \lambda_t} \left(\frac{\partial \ell_{1t}(\lambda_t)}{\partial \lambda_t} \frac{d\lambda_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \omega_i} \right) \frac{d\lambda_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \omega_j} \\ &= \sum_{t=m+1}^n \left[\frac{\partial^2 \ell_{1t}(\lambda_t)}{\partial \lambda_t^2} \frac{d\lambda_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \omega_i} + \frac{\partial \ell_{1t}(\lambda_t)}{\partial \lambda_t} \frac{\partial}{\partial \lambda_t} \left(\frac{d\lambda_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \omega_i} \right) \right] \frac{d\lambda_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \omega_j}. \end{aligned}$$

Como $\mathbb{E} \left(\frac{\partial \ell_{1t}(\lambda_t)}{\partial \lambda_t} \middle| \mathcal{F}_{t-1} \right) = 0$, segue que

$$\mathbb{E} \left(\frac{\partial^2 \ell(\boldsymbol{\gamma}; \mathbf{y})}{\partial \omega_i \partial \omega_j} \middle| \mathcal{F}_{t-1} \right) = \sum_{t=m+1}^n \mathbb{E} \left(\frac{\partial^2 \ell_{1t}(\lambda_t)}{\partial \lambda_t^2} \middle| \mathcal{F}_{t-1} \right) \left(\frac{d\lambda_t}{d\eta_{1t}} \right)^2 \frac{\partial \eta_{1t}}{\partial \omega_i} \frac{\partial \eta_{1t}}{\partial \omega_j}.$$

A segunda derivada de $\ell_{1t}(\lambda_t)$ em relação a λ_t é

$$\frac{\partial^2 \ell_{1t}(\lambda_t)}{\partial \lambda_t^2} = -\frac{\mathbb{I}_{\{0,1\}}(y_t)}{\lambda_t^2} - \frac{(1 - \mathbb{I}_{\{0,1\}}(y_t))}{(1 - \lambda_t)^2}.$$

Como $\mathbb{E}(\mathbb{I}_{\{0,1\}}(y_t) | \mathcal{F}_{t-1}) = \lambda_t$, temos que $\mathbb{E} \left(\frac{\partial^2 \ell_{1t}(\lambda_t)}{\partial \lambda_t^2} \middle| \mathcal{F}_{t-1} \right) = -\frac{1}{\lambda_t(1-\lambda_t)}$.

Adicionalmente,

$$\frac{\partial^2 \ell_{2t}(\mu_t, \varphi)}{\partial \gamma_i \partial \gamma_j} = \sum_{t=m+1}^n \frac{\partial}{\partial \mu_t} \left(\frac{\partial \ell_{2t}(\mu_t, \varphi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_j} \right) \frac{d\mu_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_i}$$

$$= \sum_{t=m+1}^n \left[\frac{\partial^2 \ell_{2t}(\mu_t, \varphi)}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_j} + \frac{\partial \ell_{2t}(\mu_t, \varphi)}{\partial \mu_t} \frac{\partial}{\partial \mu_t} \left(\frac{d\mu_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_j} \right) \right] \frac{d\mu_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_i}.$$

Do Lema 1 de Bayer et al. (2017), temos que $\mathbb{E} \left(\frac{\partial \ell_2(\mu_t, \varphi)}{\partial \mu_t} \middle| \mathcal{F}_{t-1} \right) = 0$. Assim,

$$\mathbb{E} \left(\frac{\partial^2 \ell_2(\mu_t, \varphi)}{\partial \gamma_i \partial \gamma_j} \middle| \mathcal{F}_{t-1} \right) = \sum_{t=m+1}^n \mathbb{E} \left(\frac{\partial^2 \ell_2(\mu_t, \varphi)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1} \right) \left(\frac{d\mu_t}{d\eta_{2t}} \right)^2 \frac{\partial \eta_{2t}}{\partial \gamma_i} \frac{\partial \eta_{2t}}{\partial \gamma_j}. \quad (3.5)$$

Agora, sejam

$$\zeta_{tk} = \frac{\mu_t^{k\varphi-2}}{(1 - \mu_t^\varphi)^k \log(1 - \mu_t^\varphi)^k},$$

$$A_t = \frac{\partial}{\partial \mu_t} \left(\frac{\mu_t^{\varphi-1}}{(1 - \mu_t^\varphi) \log(1 - \mu_t^\varphi)} \right) = \varphi \zeta_{2k} [1 + \log(1 - \mu_t^\varphi)] + (\varphi - 1) \zeta_{1k},$$

de modo que, pela regra da multiplicação, obtemos

$$\begin{aligned} \frac{\partial^2 \ell_{2t}(\mu_t, \varphi)}{\partial \mu_t^2} &= [1 - \mathbb{I}_{\{0,1\}}(y_t)] \left[\varphi A_t \left(\frac{\log(0.5)}{\log(1 - \mu_t^\varphi)} \log(1 - y_t^\varphi) + 1 \right) + \frac{\varphi^2 \log(0.5)}{\log(1 - \mu_t^\varphi)} \log(1 - y_t^\varphi) \right] \\ &= [1 - \mathbb{I}_{\{0,1\}}(y_t)] [\varphi A_t + \varphi \delta_t (A_t + \varphi \zeta_{2k}) \log(1 - y_t^\varphi)]. \end{aligned} \quad (3.6)$$

Pelo Lema 1 de Bayer et al. (2017), temos $\mathbb{E}(\log(1 - Y_t^\varphi) | \mathcal{F}_{t-1}) = -1/\delta_t$. Assim,

$$\mathbb{E} \left(\frac{\partial^2 \ell_2(\mu_t, \varphi)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1} \right) = (\lambda_t - 1) \varphi^2 \zeta_{2k} = w_t.$$

Tomando a esperança condicional e usando o Lema 2 de Bayer et al. (2017), substituindo o resultado do Lema 1 em (3.6), de (3.5) obtemos

$$\mathbb{E} \left(\frac{\partial^2 \ell(\gamma; \mathbf{y})}{\partial \gamma_i \partial \gamma_j} \middle| \mathcal{F}_{t-1} \right) = \sum_{t=m+1}^n \frac{w_t}{(g_2'(\mu_t))^2} \frac{\partial \eta_{2t}}{\partial \gamma_i} \frac{\partial \eta_{2t}}{\partial \gamma_j}. \quad (3.7)$$

Para a segunda derivada de ℓ_{3t} em relação a φ , temos que

$$\frac{\partial c_t}{\partial \varphi} = \delta_t \mu_t \zeta_{t2} \log(\mu_t) \log(1 - y_t^\varphi) - \delta_t \mu_t \zeta_{t1} \left(\frac{y_t^\varphi \log(y_t)}{1 - y_t^\varphi} \right) + c_t \log(\mu_t) \left(\frac{\mu_t}{(1 - \mu_t^\varphi) \log(1 - \mu_t^\varphi)} + \frac{1}{1 - \mu_t^\varphi} \right)$$

e

$$\frac{\partial}{\partial \varphi} \left(\frac{y_t^\varphi \log(y_t)}{1 - y_t^\varphi} (\delta_t - 1) \right) = (\delta_t - 1) \frac{y_t^\varphi \log(y_t)^2}{(1 - y_t^\varphi)^2} + \frac{\delta_t \mu_t \log(\mu_t)}{(1 - \mu_t^\varphi) \log(1 - \mu_t^\varphi)} \left(\frac{y_t^\varphi \log(y_t)}{1 - y_t^\varphi} \right),$$

de modo que

$$\begin{aligned} \frac{\partial^2 \ell_{2t}(\mu_t, \varphi)}{\partial \varphi^2} &= [1 - \mathbb{I}_{\{0,1\}}(y_t)] \left[-\frac{1}{\varphi^2} + \delta_t \mu_t^2 \zeta_{t2} \log(\mu_t)^2 \log(1 - y_t^\varphi) - 2\delta_t \mu_t^2 \log(\mu_t) \zeta_{t1} \left(\frac{y_t^\varphi \log(y_t)}{1 - y_t^\varphi} \right) \right. \\ &\quad \left. - (\delta_t - 1) \frac{y_t^\varphi \log(y_t)^2}{(1 - y_t^\varphi)^2} + c_t \mu_t \log(\mu_t)^2 \left(\zeta_{t1} \mu_t^2 + \frac{1}{1 - \mu_t^\varphi} \right) \right]. \end{aligned} \quad (3.8)$$

Tomando a esperança condicional em (3.8) e substituindo os resultados do Lema 2 de Bayer et al. (2017), segue que

$$\mathbb{E} \left(\frac{\partial^2 \ell(\boldsymbol{\gamma}; \mathbf{y})}{\partial \varphi^2} \middle| \mathcal{F}_{t-1} \right) = \sum_{t=m+1}^n \left\{ (\lambda_t - 1) \left[\frac{1}{\varphi^2} + \mu_t^2 \zeta_{t2} \log(\mu_t)^2 + 2\delta_t \mu_t^2 \log(\mu_t) \zeta_{t1} \left(\frac{1 - \psi(\delta_t + 1) - k}{(\delta_t - 1)\varphi} \right) + \frac{\delta_t \{ \psi(\delta_t) [\psi(\delta_t) + 2(k-1)] - \psi'(\delta_t) + k_0 \}}{(\delta_t - 2)\varphi^2} \right] \right\},$$

em que $\psi : (0, \infty) \rightarrow \mathbb{R}$ é a função digamma definida como $\psi(z) = \frac{d}{dz} \log(\Gamma(z))$, $\psi'(z) = \frac{d}{dz} \psi(z)$ é a função trigamma, $k = 0.5772156649\dots$ é a contante de Euler-Mascheroni e $k_0 = \pi^2/(6 + k^2 - 2k)$. Quanto à derivada em relação a $\boldsymbol{\gamma}_j$ e φ , temos

$$\frac{\partial^2 \ell_{2t}(\mu_t, \varphi)}{\partial \varphi \partial \boldsymbol{\gamma}_j} = [1 - \mathbb{I}_{\{0,1\}}(y_t)] \left[c_t \frac{\partial [\mu_t \log(\mu_t)]}{\partial \boldsymbol{\gamma}_j} + \mu_t \log(\mu_t) \frac{\partial c_t}{\partial \boldsymbol{\gamma}_j} - \frac{y_t^\varphi \log(y_t)}{1 - y_t^\varphi} \frac{\partial \delta_t}{\partial \boldsymbol{\gamma}_j} \right].$$

Lembre que $c_t = [1 - \mathbb{I}_{\{0,1\}}(y_t)] \frac{1}{\varphi} \frac{\partial \ell_{2t}(\mu_t, \varphi)}{\partial \mu_t}$, de modo que

$$\frac{\partial c_t}{\partial \boldsymbol{\gamma}_j} = \frac{\partial c_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \boldsymbol{\gamma}_j} = \frac{1}{\varphi g_2'(\mu_t)} \frac{\partial^2 \ell_{2t}(\mu_t, \varphi)}{\partial \mu_t^2} \frac{\partial \eta_{2t}}{\partial \boldsymbol{\gamma}_j},$$

$$\frac{\partial [\mu_t \log(\mu_t)]}{\partial \boldsymbol{\gamma}_j} = \frac{\partial [\mu_t \log(\mu_t)]}{\partial \mu_t} \frac{\partial \mu_t}{\partial \boldsymbol{\gamma}_j} = \left(\frac{\log(\mu_t) + 1}{g_2'(\mu_t)} \right) \frac{\partial \eta_{2t}}{\partial \boldsymbol{\gamma}_j},$$

e

$$\frac{\partial \delta_t}{\partial \boldsymbol{\gamma}_j} = \frac{\partial \delta_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \boldsymbol{\gamma}_j} = \frac{1}{g_2'(\mu_t)} \frac{\partial \eta_{2t}}{\partial \boldsymbol{\gamma}_j} \left(\frac{\log(0.5)\varphi \mu_t^{\varphi-1}}{(1 - \mu_t^\varphi) \log(1 - \mu_t^\varphi)^2} \right) = \left(\frac{\varphi \mu_t \delta_t \zeta_{t1}}{g_2'(\mu_t)} \right) \frac{\partial \eta_{2t}}{\partial \boldsymbol{\gamma}_j}.$$

Por isso,

$$\frac{\partial^2 \ell_{2t}(\mu_t, \varphi)}{\partial \varphi \partial \boldsymbol{\gamma}_j} = [1 - \mathbb{I}_{\{0,1\}}(y_t)] \left\{ \left[\frac{1}{g_2'(\mu_t)} \left(c_t (\log(\mu_t) + 1) + \frac{\mu_t \log(\mu_t)}{\varphi} \frac{\partial^2 \ell_{2t}(\mu_t, \varphi)}{\partial \mu_t^2} - \varphi \mu_t \delta_t \zeta_{t1} \frac{y_t^\varphi \log(y_t)}{1 - y_t^\varphi} \right) \right] \frac{\partial \eta_{2t}}{\partial \boldsymbol{\gamma}_j} \right\},$$

e tomando a esperança condicional, segue que

$$\mathbb{E} \left(\frac{\partial^2 \ell(\boldsymbol{\gamma}; \mathbf{y})}{\partial \varphi \partial \boldsymbol{\gamma}_j} \middle| \mathcal{F}_{t-1} \right) = \sum_{t=m+1}^n (\lambda_t - 1) \frac{d_t}{g_2'(\mu_t)} \frac{\partial \eta_{2t}}{\partial \boldsymbol{\gamma}_j},$$

em que $d_t = -\varphi \mu_t \log(\mu_t) \zeta_{t2} - \varphi \mu_t \delta_t \zeta_{t1} \left(\frac{1 - \psi(\delta_t + 1) - k}{(\delta_t - 1)\varphi} \right)$.

Agora, sejam $\mathbf{L} = \text{diag}\{-1/[\lambda_1(1 - \lambda_1)], \dots, -1/[\lambda_n(1 - \lambda_n)]\}$, $\mathbf{N} = \text{diag}\{[\partial^2 \ell_{2m+1}(\mu_{m+1}, \varphi)]/[\partial \varphi^2], \dots, [\partial^2 \ell_{2n}(\mu_n, \varphi)]/[\partial \varphi^2]\}$, $\mathbf{W} = \text{diag}\{w_{m+1}, \dots, w_n\}$, $\mathbf{D} = \text{diag}\{d_{m+1}, \dots, d_n\}$. Ainda, seja \mathbf{Q} a matriz de dimensão $(n - m) \times 2$, cujo elemento (i, j) é dado por $Q_{i,j} = \partial \eta_{1i+m} / \partial \omega_j$. Dessa forma, a matriz de informação de Fisher

condicional para γ é

$$\mathbf{K} = \mathbf{K}(\gamma) = \begin{bmatrix} \mathbf{K}_{(\omega,\omega)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{(\alpha,\alpha)} & \mathbf{K}_{(\alpha,\beta)} & \mathbf{K}_{(\alpha,\phi)} & \mathbf{K}_{(\alpha,\theta)} & \mathbf{K}_{(\alpha,\varphi)} \\ \mathbf{0} & \mathbf{K}_{(\beta,\alpha)} & \mathbf{K}_{(\beta,\beta)} & \mathbf{K}_{(\beta,\phi)} & \mathbf{K}_{(\beta,\theta)} & \mathbf{K}_{(\beta,\varphi)} \\ \mathbf{0} & \mathbf{K}_{(\phi,\alpha)} & \mathbf{K}_{(\phi,\beta)} & \mathbf{K}_{(\phi,\phi)} & \mathbf{K}_{(\phi,\theta)} & \mathbf{K}_{(\phi,\varphi)} \\ \mathbf{0} & \mathbf{K}_{(\theta,\alpha)} & \mathbf{K}_{(\theta,\beta)} & \mathbf{K}_{(\theta,\phi)} & \mathbf{K}_{(\theta,\theta)} & \mathbf{K}_{(\theta,\varphi)} \\ \mathbf{0} & \mathbf{K}_{(\varphi,\alpha)} & \mathbf{K}_{(\varphi,\beta)} & \mathbf{K}_{(\varphi,\phi)} & \mathbf{K}_{(\varphi,\theta)} & \mathbf{K}_{(\varphi,\varphi)} \end{bmatrix}, \quad (3.9)$$

em que $\mathbf{K}_{(\omega,\omega)} = \mathbf{Q}^\top \mathbf{L} \mathbf{T}_1^2 \mathbf{Q}$, $\mathbf{K}_{(\alpha,\alpha)} = -\boldsymbol{\nu}^\top \mathbf{W} \mathbf{T}_2^2 \boldsymbol{\nu}$, $\mathbf{K}_{(\alpha,\beta)} = \mathbf{K}_{(\beta,\alpha)}^\top = -\boldsymbol{\nu}^\top \mathbf{W} \mathbf{T}_2^2 \mathbf{M}$, $\mathbf{K}_{(\alpha,\phi)} = \mathbf{K}_{(\phi,\alpha)}^\top = -\boldsymbol{\nu}^\top \mathbf{W} \mathbf{T}_2^2 \mathbf{P}$, $\mathbf{K}_{(\alpha,\theta)} = \mathbf{K}_{(\theta,\alpha)}^\top = -\boldsymbol{\nu}^\top \mathbf{W} \mathbf{T}_2^2 \mathbf{R}$, $\mathbf{K}_{(\alpha,\varphi)} = \mathbf{K}_{(\varphi,\alpha)}^\top = -\boldsymbol{\nu}^\top \mathbf{D} \mathbf{T}_2 \mathbf{1}$, $\mathbf{K}_{(\beta,\beta)} = -\mathbf{M}^\top \mathbf{W} \mathbf{T}_2^2 \mathbf{M}$, $\mathbf{K}_{(\beta,\phi)} = \mathbf{K}_{(\phi,\beta)}^\top = -\mathbf{M}^\top \mathbf{W} \mathbf{T}_2^2 \mathbf{P}$, $\mathbf{K}_{(\beta,\theta)} = \mathbf{K}_{(\theta,\beta)}^\top = -\mathbf{M}^\top \mathbf{W} \mathbf{T}_2^2 \mathbf{R}$, $\mathbf{K}_{(\beta,\varphi)} = \mathbf{K}_{(\varphi,\beta)}^\top = -\mathbf{M}^\top \mathbf{D} \mathbf{T}_2 \mathbf{1}$, $\mathbf{K}_{(\phi,\phi)} = -\mathbf{P}^\top \mathbf{W} \mathbf{T}_2^2 \mathbf{P}$, $\mathbf{K}_{(\phi,\theta)} = \mathbf{K}_{(\theta,\phi)}^\top = -\mathbf{P}^\top \mathbf{W} \mathbf{T}_2^2 \mathbf{R}$, $\mathbf{K}_{(\phi,\varphi)} = \mathbf{K}_{(\varphi,\phi)}^\top = -\mathbf{P}^\top \mathbf{D} \mathbf{T}_2 \mathbf{1}$, $\mathbf{K}_{(\theta,\theta)} = -\mathbf{R}^\top \mathbf{W} \mathbf{T}_2^2 \mathbf{R}$, $\mathbf{K}_{(\theta,\varphi)} = \mathbf{K}_{(\varphi,\theta)}^\top = -\mathbf{R}^\top \mathbf{D} \mathbf{T}_2 \mathbf{1}$, $\mathbf{K}_{(\varphi,\varphi)} = -tr(\mathbf{N})$, $\mathbf{1}$ é um vetor de uns de dimensão $(n - m) \times 1$ e $tr(\cdot)$ é a função traço.

3.1.3 Intervalos de Confiança e Testes de Hipóteses

Em grandes amostras e sob condições usuais de regularidade (Andersen, 1970; Fahrmeir and Kaufmann, 1985), $\hat{\gamma}$ tem distribuição aproximadamente normal, com vetor de média γ e matriz de covariância $\mathbf{K}^{-1}(\gamma)$. Esse fato pode ser usado para se obter intervalos de confiança aproximados e para se realizar inferências via testes de hipóteses com base nas distribuições nulas assintóticas das estatísticas de teste.

Sejam γ_i o i -ésimo componente do vetor de parâmetros γ e $K(\hat{\gamma}_i)^{ij}$ o elemento (i, j) da inversa da matriz de informação de Fisher condicional em (3.9) avaliada em $\hat{\gamma} \in \mathbb{R}^{p+q+r+5}$, em que $\hat{\gamma}_i$ é o EMVC de γ_i . Pela normalidade assintótica dos EMVC e pelo Teorema 3.1 de Bayer et al. (2017) temos que

$$\frac{\hat{\gamma}_i - \gamma_i}{\sqrt{K(\hat{\gamma})^{ii}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Com base nesse resultado, podemos construir intervalos de confiança assintóticos para os parâmetros do modelo. Mais especificamente, seja z_δ o quantil δ da distribuição normal padrão e seja $0 < \alpha < 1$. O intervalo de confiança assintótico para γ_i , $i = 1, \dots, (p + q + r + 5)$, de nível $100(1 - \alpha)\%$ é

$$\left[\hat{\gamma}_i - z_{1-\alpha/2} \sqrt{K(\hat{\gamma})^{ii}}; \hat{\gamma}_i + z_{1-\alpha/2} \sqrt{K(\hat{\gamma})^{ii}} \right].$$

Passemos agora para testes de hipóteses. Suponha que o interesse reside em testar a hipótese nula $\mathcal{H}_0 : \gamma_i = \gamma_i^0$ contra a hipótese alternativa $\mathcal{H}_1 : \gamma_i \neq \gamma_i^0$, em que γ_i^0 é um dado valor. Uma estatística de teste é a raiz quadrada da estatística Wald (Wald, 1943), dada por Pawitan (2001)

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

Sob \mathcal{H}_0 , Z tem aproximadamente distribuição $\mathcal{N}(0, 1)$. Assim, o teste é realizado

comparando a estatística Z a quantis da distribuição normal padrão. A hipótese nula é rejeitada ao nível de significância $\alpha \in (0, 1)$ se $|z| > z_{1-\alpha/2}$.

Hipóteses mais gerais sobre o vetor paramétrico podem ser testadas usando as estatísticas da razão de verossimilhanças (Neyman and Pearson, 1928), escore de Rao (Rao, 1948) e Wald (Wald, 1943). Essas estatísticas de teste são, sob \mathcal{H}_0 , assintoticamente distribuídas como χ_q^2 , sendo q o número de restrições impostas em \mathcal{H}_0 . A hipótese nula é rejeitada no nível de significância α se a estatística de teste exceder $\chi_{q;1-\alpha}^2$.

3.2 Análise de Diagnóstico e Previsão

Nesta seção, serão apresentadas algumas medidas de diagnóstico e métodos de previsão. A análise de diagnóstico é aplicada a um modelo ajustado para determinar se ele captura adequadamente a dinâmica presente nos dados. Em geral, o interesse reside na detecção de especificação incorreta ou de outliers. Um modelo ajustado que passa em todas as verificações de diagnóstico pode ser usado para a geração de previsões (Bayer et al., 2017).

Os critérios de informação são ferramentas importantes para a comparação e seleção de modelos. Eles balanceiam a complexidade do modelo e sua capacidade explicativa. Os critérios de informação de Akaike (AIC) (Akaike, 1974), Schwartz (SIC) (Schwarz, 1978) e Hannan e Quinn (HQ) (Hannan and Quinn, 1979) são definidos como

$$\text{AIC} = -2\ell(\hat{\gamma}) + 2(k), \quad \text{SIC} = -2\ell(\hat{\gamma}) + k \log(n) \quad \text{e} \quad \text{HQ} = -2\ell(\hat{\gamma}) + 2k \log(\log(n)),$$

em que k é o número de parâmetros a serem estimados no modelo e $\ell(\hat{\gamma})$ é a função de log-verossimilhança condicional maximizada. Dado um conjunto de modelos ajustados a uma mesma base de dados, deve-se calcular o valor do critério de informação escolhido para cada modelo e selecionar o modelo que apresenta o menor valor de tal critério. Critérios de seleção são amplamente utilizados nas mais diferentes classes de modelos, inclusive no modelo β ARMA, como amplamente discutido em Cribari-Neto et al. (2023).

A análise dos resíduos é utilizada para determinar se o modelo ajustado fornece um bom ajuste aos dados (Kedem and Fokianos, 2005). Para o modelo proposto IKARMA, sugerimos usar o resíduo quantílico aleatorizado (Dunn and Smyth, 1996). Trata-se de uma versão aleatorizada do resíduo de Cox and Snell (1968), sendo dado por

$$r_t^q = \Phi^{-1}(u_t), \quad t = 1, \dots, n, \quad (3.10)$$

em que $\Phi(\cdot)$ denota a função de distribuição normal padrão, u_t é uma variável aleatória uniforme no intervalo $(a_t, c_t]$, com $a_t = \lim_{y \uparrow y_t} KI(y, \hat{\gamma} | \mathcal{F}_{t-1})$ e $c_t = KI(y_t, \hat{\gamma} | \mathcal{F}_{t-1})$. Aqui, $KI(y_t, \gamma | \mathcal{F}_{t-1}) = [\lambda_t(1-b)]^{\mathbb{I}_{\{0\}}(y_t)} + [\lambda_t b]^{\mathbb{I}_{\{1\}}(y_t)} + [(1-\lambda_t)F(y_t; \mu_t, \varphi | \mathcal{F}_{t-1})]^{\mathbb{I}_{\{0,1\}}(y_t)}$, $F(\cdot)$ sendo a função de distribuição cumulativa Kumaraswamy $k(\mu_t, \varphi)$ (2.11). No modelo IKARMA inflacionado em zero, u_t é uma variável aleatória uniforme em $(0, \lambda_t]$ se $y_t = 0$ e $u_t = KI_0(y_t, \gamma | \mathcal{F}_{t-1})$ se $y_t \in (0, 1)$. Por outro lado, no modelo IKARMA inflacionado em um,

u_t é uma variável aleatória uniforme em $[\lambda_t, 1)$ se $y_t = 1$ e $u_t = KI_1(y_t, \hat{\gamma} | \mathcal{F}_{t-1})$ se $y_t \in (0, 1)$.

Quando o modelo ajustado está corretamente especificado, os resíduos devem apresentar comportamento de ruído branco, ou seja, devem seguir um processo não correlacionado de média zero e variância constante (Kedem and Fokianos, 2005). Uma tendência detectável no gráfico dos resíduos contra os preditores pode decorrer de especificação incorreta da função de ligação (Ospina and Ferrari, 2012). Uma possibilidade para testar a adequação de um modelo é executar um teste do tipo portmanteau, como, por exemplo, o teste de Ljung-Box (Ljung and Box, 1978) nos resíduos do modelo ajustado, como discutido em Scher et al. (2020) para o modelo β ARMA.

Previsões baseadas no modelo IKARMA(p, q) podem ser produzidas usando a teoria de previsão de séries temporais para modelos ARMA (Brockwell and Davis, 2009; Box et al., 2015). Seja h_0 o horizonte de previsão. Assumiremos que os valores das covariáveis x_t , para $t = n + 1, \dots, n + h_0$, são conhecidos ou podem ser obtidos. Se as covariáveis são funções determinísticas de t , como por exemplo, senos e cossenos em análise harmônica, tendências polinomiais etc., elas podem ser determinadas para valores de $t > n$.

O primeiro passo reside em obter as estimativas $\hat{\mu}_{m+1}, \dots, \hat{\mu}_n$ para a mediana condicional da parte contínua da distribuição com base no EMVC. Para fazer isso, precisamos recompor o termo de erro $\{r_t\}_{t-1}^n$, que denotaremos por \hat{r}_t . Começamos por definir $\hat{r}_t = \mathbb{E}(r_t)$, que geralmente é igual a 0, para $t \in \{1, \dots, m\}$. Começando em $t = m + 1$, definimos

$$\hat{\mu}_t = g_2^{-1} \left(\hat{\alpha} + \mathbf{x}_t^\top \hat{\boldsymbol{\beta}} + \sum_{i=1}^p \hat{\phi}_i [g_2(y_{t-i}^*) - \mathbf{x}_{t-i}^\top \hat{\boldsymbol{\beta}}] + \sum_{j=1}^q \hat{\theta}_j \hat{r}_{t-j} \right),$$

em que $r_t = g_2(y_t^*) - g_2(\hat{\mu}_t)$, para $t \in \{m + 1, \dots, n\}$. Para $h = 1, 2, \dots, h_0$ passos à frente, os valores previstos de μ_{n+h} são dados por

$$\hat{\mu}_{n+h} = g_2^{-1} \left(\hat{\alpha} + \mathbf{x}_{n+h}^\top \hat{\boldsymbol{\beta}} + \sum_{i=1}^p \hat{\phi}_i [g_2(y_{n+h-i}^*) - \mathbf{x}_{n+h-i}^\top \hat{\boldsymbol{\beta}}] + \sum_{j=1}^q \hat{\theta}_j \hat{r}_{n+h-j} \right),$$

em que $r_t = 0$, para $t > n$, e

$$[g_2(y_t^*)] = \begin{cases} g_2(\hat{\mu}_t), & \text{se } t > n, \\ g_2(y_t^*), & \text{se } t \leq n. \end{cases}$$

A forma apresentada acima é a forma usual de predição encontrada em modelos dinâmicos gerais, inclusive o modelo I β ARMA, que contempla séries com inflacionamento. Contudo, proporemos outra forma de realizar predição, para poder obter valores preditos iguais a zero ou um, algo relevante na modelagem e previsão de séries hidroambientais duplamente limitadas inflacionadas. Para isso, ao invés de considerar a mediana apenas da parte contínua, propomos obter estimativas da mediana da variável inflacionada. Considerando b fixo/conhecido, os valores preditos (dentro da amostra) e previstos (fora da amostra) são dados, respectivamente,

por

$$\hat{y}_t = KI^{-1}(0.5; \hat{\lambda}_t, b, \hat{\mu}_t, \hat{\varphi} | \mathcal{F}_{t-1}),$$

$$\hat{y}_{n+h} = KI^{-1}(0.5; \hat{\lambda}_{n+h}, b, \hat{\mu}_{n+h}, \hat{\varphi} | \mathcal{F}_{t-1}),$$

em que $\hat{\varphi}$ é o EMVC de φ , $\hat{\mu}_t$ e $\hat{\mu}_{n+h}$ são dados como nas equações de predição anteriormente introduzidas e $\hat{\lambda}_t$ e $\hat{\lambda}_{n+h}$ são predições e previsões do parâmetro de mistura λ_t , respectivamente. Para obter as estimativas de $\hat{\lambda}_{m+1}, \dots, \hat{\lambda}_n$ para λ_t , começando em $t = m + 1$, definimos

$$\hat{\lambda}_t = g_1^{-1} [\hat{\omega}_1 + \hat{\omega}_2 g_1(s_{t-1}^*)],$$

em que $s_{t-1}^* = \min[\max(s_{t-1}; 0.5/n)$, com $s_{t-1} = |y_t - (1 - b)|$. Agora, para $h = 1, 2, \dots, h_0$, o valor previsto h passos à frente de λ_{n+h} é dado por

$$\hat{\lambda}_{n+h} = g_1^{-1} [\hat{\omega}_1 + \hat{\omega}_2 g_1(s_{n+h-1}^*)],$$

em que

$$s_{t-1} = \begin{cases} |\hat{\mu}_t - (1 - b)|, & \text{se } t > n, \\ |y_t - (1 - b)|, & \text{se } t \leq n. \end{cases}$$

Considerando os valores preditos \hat{y}_t , usamos o critério de que se a estimativa do parâmetro de mistura (λ_t) for maior que 0.5 o valor predito será igual a zero (ou um). Ou seja, se a probabilidade estimada do processo ser igual a zero (um) num dado momento for maior que a probabilidade estimada de estar em $(0, 1)$, então o valor predito é zero (um). Logo, as predições podem assumir valores iguais a zeros ou uns. Essa forma de predição diferencia-se das predições no modelo $I\beta$ ARMA, onde nunca haverá valores preditos inflacionados.

Capítulo 4

AVALIAÇÃO NUMÉRICA

Serão apresentados a seguir resultados de um estudo de simulação de Monte Carlo realizado para avaliar o desempenho em amostras de tamanho finito dos EMVCs para os parâmetros que indexam o modelo IKARMA(p, q) proposto.

Na avaliação numérica, consideramos três cenários distintos, a saber:

- (i) IKARMA(1, 1) com inflacionamento em zero ($b = 0$), com $\alpha = -1.00$, $\phi_1 = -0.45$, $\theta_1 = 0.30$, $\varphi = 5.00$, $\omega_1 = -3.00$ e $\omega_2 = 0.50$;
- (ii) IKARMA(2, 2) com inflacionamento em um ($b = 1$), com $\alpha = 1.00$, $\phi_1 = 0.50$, $\phi_2 = -0.20$, $\theta_1 = -0.40$, $\theta_2 = -0.15$, $\varphi = 10.00$, $\omega_1 = -3.00$ e $\omega_2 = 0.40$;
- (iii) IKARMA(1, 1) com inflacionamento em um ($b = 1$) e a covariável $x_t = [\sin(2\pi t/12)]$, para $t \in \{1, \dots, n\}$, com $\alpha = 6.24$, $\phi_1 = -0.60$, $\theta_1 = 0.81$, $\varphi = 41.00$, $\omega_1 = -3.38$, $\omega_2 = 0.51$ e $\beta = -0.30$.

Para os cenários descritos acima, os valores dos parâmetros foram selecionados baseados em modelos ajustados com dados de aplicações reais. Consideramos $n \in \{50, 100, 200, 500\}$ e 10000 réplicas de Monte Carlo. Todos os códigos foram escritos no ambiente computacional R na versão 4.1.3 (R Core Team, 2021) pelos autores e estão disponíveis em <https://github.com/fabiobayer/IKARMA>.

Amostras de tamanho n do processo IKARMA(p, q) foram geradas como descrito a seguir. O primeiro passo é definir $r_t = 0$ e $\mu_t = g_2^{-1}(\alpha)$ para $t = 1, \dots, m$. Segundo passo: para $t = m + 1$, obtemos η_{1t} (3.2) e η_{2t} (3.3) e então definimos $\lambda_t = g_1^{-1}(\eta_{1t})$ e $\mu_t = g_2^{-1}(\eta_{2t})$. Em seguida, y_t é gerado pelo método da inversão a partir da Equação (3.1). Para isso, geramos $u \sim U(0, 1)$ e definiremos $y_t = \left(1 - (1 - u)^{\frac{\log(1 - \mu_t^\varphi)}{\log(0.5)}}\right)^{\frac{1}{\varphi}}$. Iteramos o segundo passo para $t = m + 1, \dots, n$. A amostra final é y_1, \dots, y_n .

As Tabelas 4.1, 4.2 e 4.3 apresentam os resultados das simulações. As estatísticas de desempenho apresentadas são a média, a mediana, o erro-padrão (SE), o viés e o viés relativo (RB) das estimativas pontuais; apresentamos ainda as taxas de cobertura (CR) dos intervalos de confiança assintóticos de nível 95%. O viés relativo percentual é definido como $100 \times$ a razão entre o viés e o valor verdadeiro do parâmetro.

Tabela 4.1: Resultados da simulação de Monte Carlo, modelo IKARMA(1, 1) inflacionado em zero ($b = 0$).

Parâmetros	α	ϕ_1	θ_1	φ	ω_1	ω_2
	-1.000	-0.450	0.300	5.000	-3.000	0.500
$n = 50$						
Média	-0.955	-0.386	0.243	5.314	-2.899	0.156
Mediana	-0.979	-0.433	0.271	5.247	-2.841	0.387
SE	0.218	0.310	0.309	0.686	0.919	1.107
Viés	0.045	0.064	-0.057	0.314	0.101	-0.344
RB	-4.496	-14.137	-18.941	6.274	-3.374	-68.822
CR	0.853	0.807	0.809	0.952	0.951	0.987
$n = 100$						
Média	-0.990	-0.437	0.291	5.148	-3.027	0.429
Mediana	-0.994	-0.447	0.293	5.122	-2.975	0.471
SE	0.124	0.176	0.180	0.453	0.551	0.347
Viés	0.010	0.013	-0.009	0.148	-0.027	-0.071
RB	-1.012	-2.942	-3.018	2.967	0.900	-14.199
CR	0.909	0.882	0.884	0.956	0.960	0.973
$n = 300$						
Média	-0.999	-0.448	0.299	5.050	-3.021	0.496
Mediana	-0.997	-0.449	0.298	5.041	-3.004	0.495
SE	0.054	0.074	0.078	0.252	0.292	0.075
Viés	0.001	0.002	-0.001	0.050	-0.021	-0.004
RB	-0.121	-0.398	-0.304	0.996	0.691	-0.738
CR	0.945	0.932	0.932	0.965	0.953	0.953
$n = 500$						
Média	-1.000	-0.450	0.301	5.028	-3.011	0.498
Mediana	-1.000	-0.451	0.301	5.022	-3.000	0.497
SE	0.040	0.053	0.056	0.195	0.226	0.052
Viés	0.000	0.000	0.001	0.028	-0.011	-0.002
RB	0.001	0.099	0.365	0.551	0.354	-0.440
CR	0.953	0.940	0.942	0.968	0.950	0.948

Tabela 4.2: Resultados da simulação de Monte Carlo, modelo IKARMA(2, 2) inflacionado em um ($b = 1$).

Parâmetros	α	ϕ_1	ϕ_2	θ_1	θ_2	φ	ω_1	ω_2
	1.000	0.500	-0.200	-0.400	-0.150	10.000	-3.000	0.400
$n = 50$								
Média	1.172	0.454	-0.275	-0.362	-0.072	11.131	-2.950	0.234
Mediana	1.089	0.486	-0.269	-0.390	-0.098	10.982	-2.831	0.324
SE	0.438	0.274	0.229	0.307	0.278	1.704	1.188	0.714
Viés	0.172	-0.046	-0.075	0.038	0.078	1.131	0.050	-0.166
RB	17.184	-9.253	37.666	-9.592	-51.906	11.310	-1.659	-41.585
RC	0.748	0.776	0.713	0.776	0.716	0.929	0.953	0.983
$n = 100$								
Média	1.047	0.492	-0.225	-0.395	-0.127	10.477	-3.010	0.347
Mediana	1.025	0.499	-0.221	-0.401	-0.136	10.417	-2.945	0.371
SE	0.214	0.136	0.118	0.152	0.135	1.103	0.647	0.294
Viés	0.047	-0.008	-0.025	0.005	0.023	0.477	-0.010	-0.053
RB	4.713	-1.656	12.268	-1.316	-15.178	4.769	0.320	-13.136
CR	0.853	0.884	0.836	0.886	0.839	0.940	0.952	0.968
$n = 300$								
Média	1.005	0.501	-0.204	-0.402	-0.147	10.138	-3.009	0.394
Mediana	1.005	0.501	-0.203	-0.401	-0.148	10.119	-2.990	0.396
SE	0.045	0.031	0.037	0.033	0.038	0.590	0.314	0.085
Viés	0.005	0.001	-0.004	-0.002	0.003	0.138	-0.009	-0.006
RB	0.497	0.205	1.920	0.453	-1.834	1.384	0.308	-1.542
CR	0.926	0.934	0.916	0.938	0.922	0.954	0.950	0.955
$n = 500$								
Média	1.002	0.500	-0.201	-0.400	-0.150	10.089	-3.010	0.397
Mediana	1.002	0.500	-0.201	-0.400	-0.150	10.078	-2.999	0.396
SE	0.027	0.020	0.023	0.021	0.023	0.457	0.235	0.057
Viés	0.002	0.000	-0.001	0.000	0.000	0.089	-0.010	-0.003
RB	0.209	0.039	0.607	0.090	-0.330	0.886	0.341	-0.659
CR	0.946	0.940	0.933	0.944	0.934	0.954	0.949	0.952

Tabela 4.3: Resultados da simulação de Monte Carlo, modelo IKARMA(1, 1) inflacionado em um ($b = 1$).

Parâmetros	α	ϕ_1	θ_1	φ	ω_1	ω_2	β
	6.240	-0.600	0.810	41.000	-3.380	0.510	-0.300
$n = 50$							
Média	5.396	-0.414	0.555	41.941	-3.983	0.631	-0.276
Mediana	6.125	-0.640	0.822	40.556	-3.349	0.503	-0.264
SE	1.896	0.548	0.543	21.617	5.737	1.280	0.277
Viés	-0.844	0.186	-0.255	0.941	-0.603	0.121	0.024
RB	-13.531	-30.968	-31.483	2.294	17.839	23.729	-7.902
RC	0.799	0.637	0.692	0.870	0.961	0.964	0.888
$n = 100$							
Média	5.877	-0.519	0.712	42.329	-3.488	0.526	-0.289
Mediana	6.313	-0.645	0.850	41.337	-3.372	0.509	-0.285
SE	1.374	0.408	0.377	24.123	1.341	0.288	0.181
Viés	-0.363	0.081	-0.098	1.329	-0.108	0.016	0.011
RB	-5.822	-13.461	-12.115	3.242	3.183	3.115	-3.561
RC	0.814	0.716	0.688	0.893	0.953	0.955	0.920
$n = 300$							
Média	6.147	-0.577	0.788	42.182	-3.383	0.509	-0.307
Mediana	6.200	-0.592	0.807	41.267	-3.362	0.506	-0.305
SE	0.513	0.129	0.116	30.895	0.546	0.102	0.104
Viés	-0.093	0.023	-0.022	1.182	-0.003	-0.001	-0.007
RB	-1.492	-3.916	-2.703	2.883	0.089	-0.146	2.215
RC	0.907	0.889	0.878	0.913	0.951	0.952	0.937
$n = 500$							
Média	6.146	-0.575	0.790	41.441	-3.382	0.510	-0.308
Mediana	6.171	-0.581	0.798	41.168	-3.370	0.508	-0.307
SE	0.339	0.083	0.069	15.455	0.381	0.067	0.083
Viés	-0.094	0.025	-0.020	0.441	-0.002	0.000	-0.008
RB	-1.505	-4.189	-2.482	1.077	0.070	-0.037	2.602
RC	0.912	0.888	0.899	0.931	0.953	0.953	0.937

Observando os resultados, em todos os cenários, percebemos que SE e RB diminuem à medida que o tamanho amostral aumenta. Os valores RB são menores que 1%, em módulo, a partir do tamanho de amostra $n = 300$ no primeiro cenário (4.1), $n = 500$, no segundo cenário (4.2). Assim, temos indicativo numérico de que os EMVC são consistentes e de que a teoria desenvolvida, assim como as implementações computacionais, estão corretas. Por exemplo, para o estimador $\hat{\phi}_1$ na Tabela 4.1 com o modelo IKARMA(1, 1) inflacionado em zero, SE é igual a 0.310 para $n = 50$ e igual a 0.053 para $n = 500$. Para o mesmo cenário e parâmetro, temos RB igual a -14.137% para $n = 50$ e 0.099% para $n = 500$. A Tabela 4.2 mostra resultados semelhantes para o cenário com o modelo IKARMA(2, 2) inflacionado em um. Por exemplo, para o estimador $\hat{\theta}_2$, RB é igual a -51.906% para $n = 50$ e já diminui consideravelmente quando $n = 300$, com RB = -1.834% . O mesmo acontece para SE, que é igual a 0.278 e 0.038, para $n = 50$ e $n = 300$, respectivamente. A Tabela 4.3 também mostra panorama similar para os estimadores $\hat{\phi}_1$ e $\hat{\theta}_1$. O viés relativo de ambos os estimadores diminui consideravelmente à medida que o tamanho amostral aumenta. Por exemplo, com $n = 50$, o estimador $\hat{\phi}_1$ tem RB = -30.968% e o estimador $\hat{\theta}_1$ apresenta RB = -31.483% . Para $n = 500$, os vieses relativos de $\hat{\phi}_1$ e $\hat{\theta}_1$ são reduzidos para -4.189% e -2.482% , respectivamente. Quanto ao estimador do parâmetro de precisão ($\hat{\varphi}$), em todos os cenários a estimação é muito boa, melhorando conforme o tamanho amostral aumenta.

Adicionalmente, em geral, as taxas de cobertura dos intervalos de confiança estão próximas ao valor nominal de 95%. Por exemplo, no primeiro, segundo e terceiro cenários, para $n = 500$, as coberturas empíricas dos intervalos de confiança para ω_1 são, respectivamente, 95.0% (Tabela 4.1), 94.9% (Tabela 4.2) e 95.3% (Tabela 4.3).

No geral, os resultados numéricos apresentados acima mostram boas propriedades das inferências via máxima verossimilhança condicional no modelo IKARMA, mesmo em tamanhos amostrais moderados.

Capítulo 5

APLICAÇÕES

Este capítulo objetiva evidenciar a aplicabilidade prática do modelo proposto na modelagem de séries temporais hidroambientais duplamente limitadas e inflacionadas. Para isso, serão considerados dados de umidade relativa do ar (UR) e o percentual de volume útil (UV) do reservatório de água de uma hidrelétrica.

5.1 Aplicação de Umidade Relativa do Ar

A UR é definida pela relação entre a quantidade de vapor de água disponível no ar e a quantidade máxima de vapor de água que pode ser mantida no ar na mesma temperatura. Ou seja, a UR é definida como a razão entre a pressão parcial de umidade no ar e a pressão de saturação do vapor de água na mesma temperatura (Ozbek et al., 2022). Se a UR do ar chegar a 100% em uma determinada temperatura, isso indica que o ar atingiu a saturação e, se mais umidade for adicionada, condensação será formada no ar. Além de humanos, plantas e animais são afetados pela UR do ar.

Estudos para monitorar a UR são de suma importância no gerenciamento de recursos hídricos, na área agrícola, em estudos climáticos, assim como para a gestão da saúde pública (Bayer and Bayer, 2015). Entender características meteorológicas, dentre elas a umidade relativa do ar, e determinar previsões para o futuro pode auxiliar em medidas preventivas na área da saúde, assim como em previsões climáticas e no gerenciamento dos recursos hídricos (Bayer and Bayer, 2015). Por exemplo, usando dados meteorológicos de várias cidades da Índia, em Parishwad et al. (1998) foi desenvolvido um modelo para estimar UR, temperatura do ar e velocidade do vento. As taxas de erro para temperatura do ar, UR e velocidade do vento nas estimativas fornecidas pelo modelo desenvolvido foram de 10.5%, 14.6% e 26.7%, respectivamente. Sousa et al. (2007) destaca que o aumento da temperatura e da UR podem estar relacionados a uma maior incidência na disseminação de vírus e microrganismos causadores de doenças como meningite, dengue e pneumonia. Segundo Barcellos et al. (2009), características climáticas como a temperatura, UR e precipitação, associadas a características físicas e químicas dos poluentes presentes na atmosfera, podem potencializar os efeitos causados por mudanças climáticas. Liu et al. (2021) analisaram como a temperatura e a

Tabela 5.1: Medidas descritivas da variável UR máxima mensal na cidade de Caxias do Sul, RS.

Média	Mediana	Desvio-Padrão	Coefficiente de Variação	Mínimo	Máximo
0.9701	0.9750	0.0279	2.8737	0.8825	1.0000

umidade relativa afetaram o desempenho do aprendizado em alunos de graduação, onde os mesmos foram submetidos a diferentes temperaturas do ar e níveis de UR, em que o grau de desconforto, fadiga, distração e eficiência de aprendizagem foram medidos e analisados sob perspectivas psicológicas e fisiológicas. Foi determinado que a UR é um fator mais importante no desempenho do aprendizado do que a temperatura do ambiente.

Além das questões destacadas acima sobre a UR, notamos que sua modelagem e seu monitoramento adequados são importantes para diversas áreas da atividade humana. Por exemplo, previsões adequadas da UR podem auxiliar o Estado a adotar medidas preventivas referentes à saúde pública, gestão dos recursos hídricos ou também na previsão do clima (Bayer and Bayer, 2015; Bayer et al., 2017).

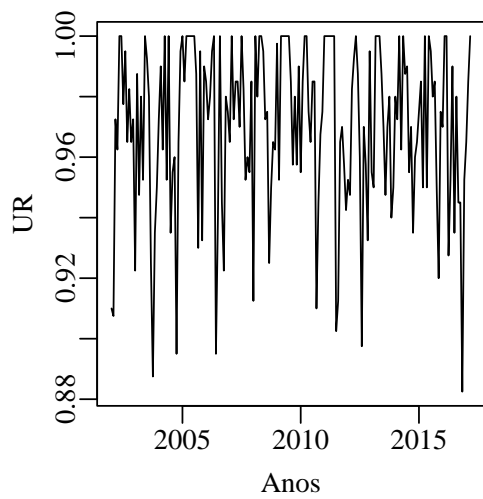
Apresentaremos a seguir uma aplicação para a UR máxima mensal em Caxias do Sul, RS, Brasil. A série temporal real consiste em valores máximos mensais de janeiro de 2002 a abril de 2017, resultando em um tamanho amostral $n = 183$, onde 22.95% das observações são inflacionadas em um. Os dados foram obtidos em Brazilian National Institute of Meteorology (INMET) (2018). As últimas 12 observações foram reservadas para comparação de previsões. A Figura 5.1 apresenta (a) o gráfico da série temporal, (b) o componente sazonal nos dados, (c) a função de autocorrelação amostral (ACF) e (d) a função de autocorrelação parcial amostral (PACF). A Tabela 5.1 contém algumas medidas descritivas da série da UR.

Analisando-se a Figura 5.1 e a Tabela 5.1, percebe-se que no período analisado o valor médio da variável UR mensal máxima é de 97.01%, com coeficiente de variação de 2.87%. A variabilidade da UR pode ser verificada na Figura 5.1 (a), em que o valor mínimo registrado foi de 88.25% e o valor máximo foi de 100%. Podemos observar na Figura 5.1 (b) que há flutuações sazonais na UR. Há uma variabilidade intra-anual, alcançando, em média, os menores valores entre agosto e novembro e os maiores valores nos meses de março a junho. A umidade do ar está relacionada ao movimento das massas de ar e à transição da água do estado líquido para gasoso devido à evaporação das plantas e superfícies evaporantes (Bayer and Bayer, 2015), características afetadas pelas épocas do ano. Para representar esse componente sazonal mensal, consideramos uma abordagem de regressão harmônica simples (Bloomfield, 2004), introduzindo a seguinte covariável ao modelo:

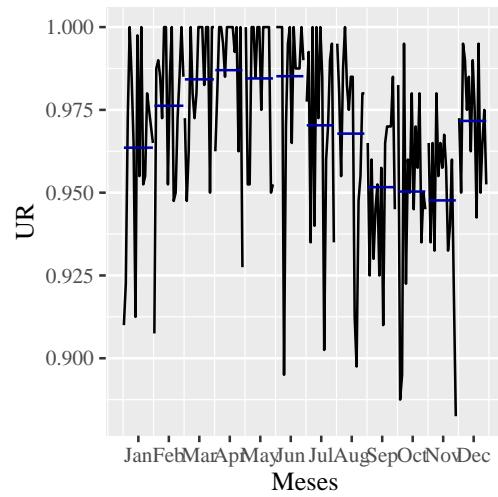
$$x_t = [\sin(2\pi t/12)], \quad \text{para } t \in \{6, \dots, n + 5\},$$

em que t inicia em 6 para ajustar a fase da função seno com a série temporal observada.

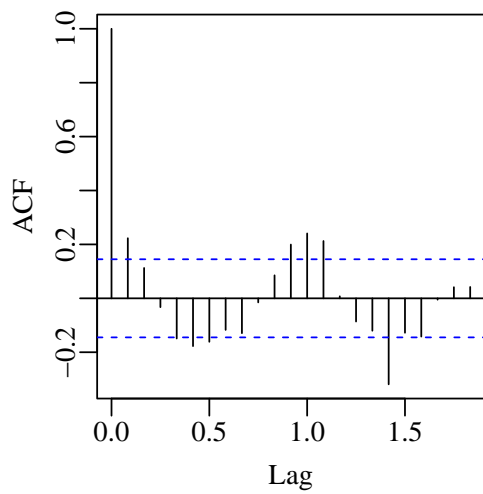
Utilizaremos a função de ligação logit. A seleção do melhor modelo será feita com base na metodologia de Box-Jenkins (Box et al., 2015). Em particular, usaremos correlogramas e correlogramas parciais e os critérios de informação AIC, SIC e HQ, descritos na Seção 3.2. A



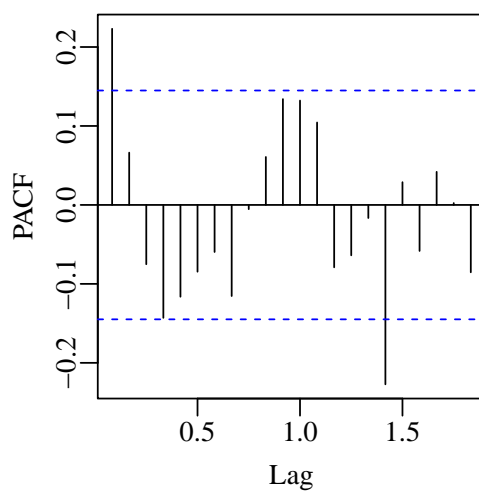
(a) Dados observados



(b) Sazonalidade



(c) ACF



(d) PACF

Figura 5.1: Série temporal de UR observadas em Caxias do Sul, RS, Brasil: série, componente sazonal, correlograma e correlograma parcial.

Tabela 5.2: Modelo IKARMA(1, 1) ajustado para dados de UR máxima mensal na cidade de Caxias do Sul, RS: estimativas, erros-padrão, estatísticas z , p -valores z , AIC e SIC, estatística de Ljung-Box e p -valor do teste de Ljung-Box.

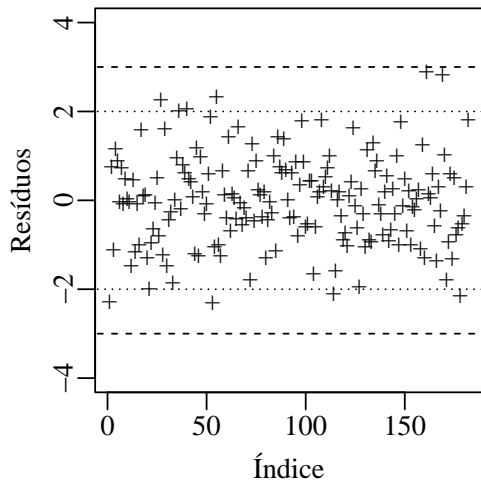
Parâmetro	Estimativa	Erro-Padrão	z estat.	$P(> z)$
α	6.2378	0.3933	15.8603	< 0.0001
ϕ_1	-0.8028	0.1033	7.7727	< 0.0001
θ_1	0.8119	0.0937	8.6629	< 0.0001
φ	41.0066	3.7173	11.0314	< 0.0001
ω_1	-3.3764	0.6691	5.0475	< 0.0001
ω_2	0.5153	0.1459	3.5332	0.0004
β_1	-0.2967	0.0980	3.0263	0.0025
Ljung-Box (lag = 20): $Q = 25.622$ (p -valor = 0.1787)				

estimação dos parâmetros do modelo será realizada por máxima verossimilhança condicional, como detalhado na Seção 3.1. Por fim, a avaliação de diagnóstico será feita através da análise residual descrita na Seção 3.2, com base no resíduo quantílico aleatorizado. A Tabela 5.2 contém as estimativas dos parâmetros do modelo IKARMA(1, 1) com covariável selecionado e demais quantidades relevantes, enquanto a Figura 5.1 contém gráficos de diagnósticos residuais: gráfico dos resíduos versus os índices das observações, gráfico QQ normal dos resíduos, correlograma e correlograma parcial dos resíduos.

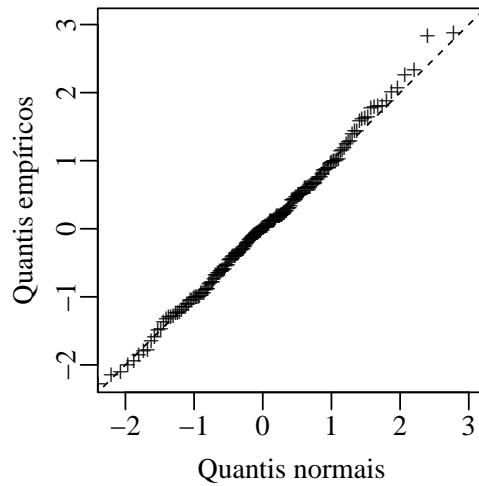
A Figura 5.1 (a) apresenta os resíduos quantílicos aleatorizados contra seus índices. Observando este gráfico, não se percebe qualquer padrão temporal, havendo distribuição aparentemente aleatória dos resíduos em torno de zero e dentro do intervalo que se estende de -3 a 3 . Deve-se notar que não há resíduo fora do intervalo $[-3, 3]$. A Figura 5.1 (b) apresenta o gráfico residual quantil-quantil normal (QQ-plot), de onde se percebe que há boa coincidência entre quantis empíricos e quantis normais. Trata-se de bom indicativo de que o modelo está corretamente especificado. Por fim, o correlograma e o correlograma parcial dos resíduos apresentados nas Figuras 5.1 (c) e (d), respectivamente, não apresentam evidência substantiva contra a boa especificação do modelo; deve-se lembrar que os intervalos indicados nesses gráficos são de natureza assintótica. Lembre também que o teste de Ljung-Box (Tabela 5.2) não rejeitou a hipótese nula de ausência de correlação serial nos resíduos ao nível de significância de 10%.

Predições dentro e fora da amostra são apresentadas nas Figuras 5.1 (a) e (b), respectivamente. As predições IKARMA foram obtidas a partir das duas estratégias introduzidas na Seção 3.2. Apresentamos também as correspondentes predições obtidas a partir do modelo I β ARMA ajustado de mesma ordem e com a mesma covariável. Nesses gráficos, a legenda IKARMA(1, 1) denota os valores preditos por meio da mediana condicional da parte contínua ($\hat{\mu}_t$), enquanto a legenda IKARMA*(1, 1) se refere aos valores preditos pela mediana da variável inflacionada (\hat{y}_t).

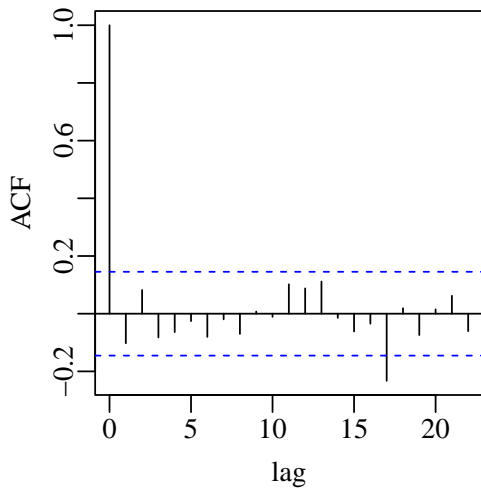
Podemos observar que as predições dentro da amostra do modelo IKARMA*(1, 1) destacam-se entre os métodos de predição considerados. Percebe-se que os valores preditos \hat{y}_t estão muito mais próximos dos valores observados iguais a um. Esse comportamento não é observado para



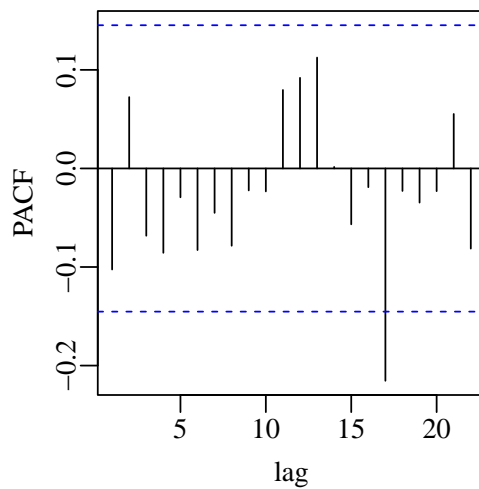
(a) Resíduos quantílicos vs índices



(b) QQ-plot



(c) ACF residual



(d) PACF residual

Figura 5.2: Gráficos de diagnóstico do modelo IKARMA ajustado para dados de UR máxima mensal na cidade de Caxias do Sul, RS, com base nos resíduos quantílicos aleatorizados: gráfico dos resíduos contra os índices das observações, QQ-plot dos resíduos, correlograma e correlograma parcial dos resíduos.

Tabela 5.3: Medidas de acurácia de predições para os diferentes modelos ajustados para a UR máxima mensal na cidade de Caxias do Sul, RS.

	IKARMA(1, 1)	IKARMA*(1, 1)	IBARMA(1, 1)
MSE	0.0006	0.0007	0.0006
RMSE	0.0255	0.0261	0.0251
MAPE	2.1495	2.0050	2.0753
* Previsão da mediana da variável inflacionada			

Tabela 5.4: Medidas de acurácia de previsões para os diferentes modelos ajustados para a UR máxima mensal na cidade de Caxias do Sul, RS.

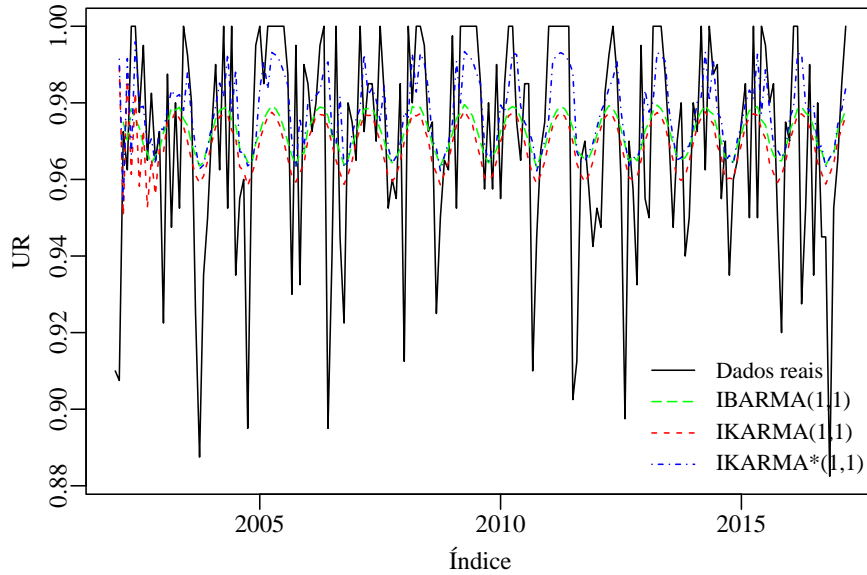
	IKARMA(1, 1)	IKARMA*(1, 1)	IBARMA(1, 1)
MSE	0.0010	0.0013	0.0011
RMSE	0.0322	0.0366	0.0328
MAPE	2.1405	2.1947	2.0994
* Previsão da mediana da variável inflacionada			

as predições fornecidas pelo modelo $I\beta$ ARMA. As predições conseguiram capturar o padrão sazonal presente nos dados. A Figura 5.1 (b) mostra os valores previstos fora da amostra para os modelos ajustados. Visualmente não é possível identificar a superioridade de um conjunto de previsões. Para uma melhor comparação, algumas medidas de qualidade de predição são consideradas. O erro quadrático médio (MSE), a raiz quadrada do erro quadrático médio (RMSE) e o erro percentual absoluto médio (MAPE) entre os dados reais observados e cada uma das formas de predição são apresentados nas Tabelas 5.3 e 5.4, para predição dentro da amostra e previsão, respectivamente. Notamos que o modelo proposto supera o modelo $I\beta$ ARMA em praticamente todas as figuras de mérito elencadas. Destaca-se ainda que o modelo $I\beta$ ARMA não consegue fornecer valores preditos de UR iguais a 0% ou 100%, o que é uma limitação do modelo em aplicações práticas.

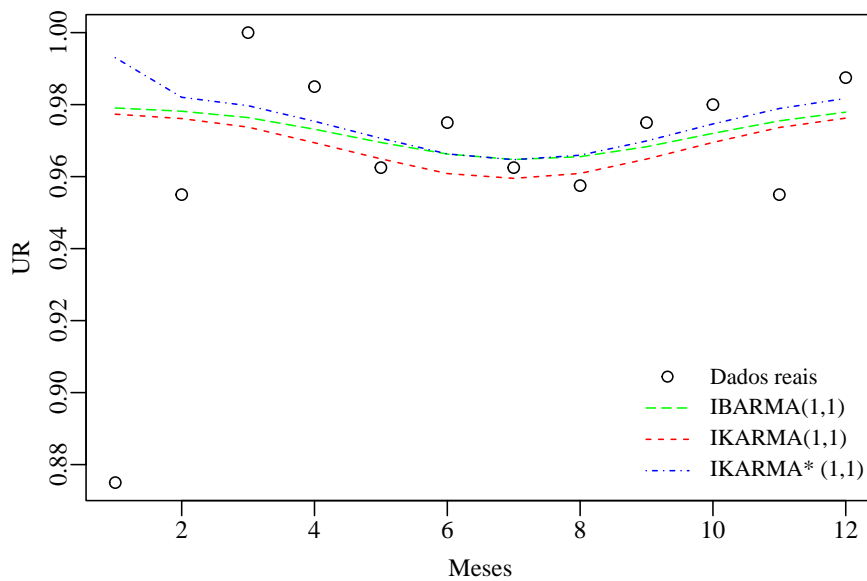
5.2 Aplicação do Percentual de Volume Útil

Em um reservatório, o volume útil (VU) é definido como o volume de água entre os níveis máximo e mínimo normal de operação (Operador Nacional do Sistema Elétrico: Dados Hidrológicos, 2023), assumindo valores no intervalo unitário padrão. Ou seja, representa a parcela de água do reservatório que pode efetivamente ser usada para geração de energia (Eletrobras Furnas, 2023). A modelagem do VU é importante para alcançar uma gestão eficaz da quantidade e da qualidade da água (Ferreira et al., 2021; Nozari et al., 2021; Lindim et al., 2011).

A água é um recurso fundamental para a sobrevivência humana e para o desenvolvimento socioeconômico, sendo indispensável nas mais variadas atividades humanas. Ao considerar o crescimento da população registrado nas últimas décadas, a importância da água torna-se ainda maior. Tal aumento da população ocasionou um aumento da demanda de água em qualidade adequada aos diversos usos (Fontenele et al., 2011). Além disso, o aumento de



(a) Predição dentro da amostra.



(b) Previsão fora da amostra.

Figura 5.3: Predições e previsões obtidas a partir dos diferentes modelos ajustados para a UR máxima mensal na cidade de Caxias do Sul, RS.

Tabela 5.5: Medidas descritivas da variável VU de São Simão, MG.

Média	Mediana	Desvio-Padrão	Coefficiente de Variação	Mínimo	Máximo
0.6162	0.6621	0.2730	0.4430	0.1043	1.0000

outras interferências sobre os recursos hídricos tem causado deterioração da qualidade das águas superficiais, tais como a poluição proveniente do uso de fertilizantes e agrotóxicos na agricultura, os despejos de efluentes domésticos e industriais, entre outras (Alves et al., 2008; Von Sperling, 2005). Dessa forma, a busca pela conservação dos recursos hídricos e da avaliação de sua qualidade e disponibilidade é de suma importância para a população (Von Sperling, 2005).

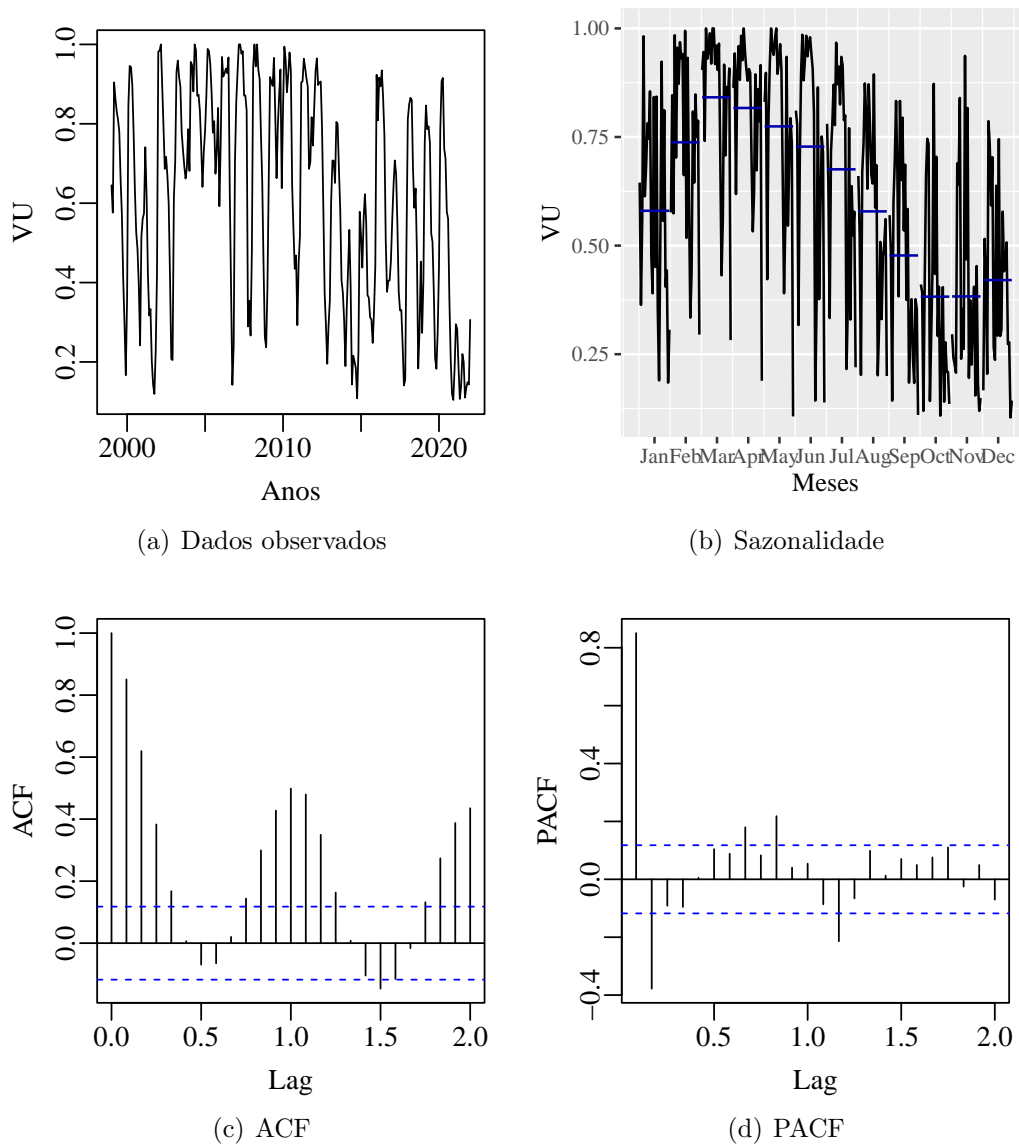
Apresentaremos a seguir uma análise empírica do VU mensal da Usina Hidrelétrica São Simão, localizada no estado de Minas Gerais, Brasil. A série temporal contém 289 observações de VU mensais de janeiro de 1999 a janeiro de 2023 obtidas em Operador Nacional do Sistema Elétrico: Dados Hidrológicos (2023). Como na primeira aplicação, as últimas 12 observações foram reservadas para fins de previsão. A modelagem é realizada, assim, com base em $n = 277$ observações, sendo seis observações iguais a um. Dessa forma, cerca de 2.17% das observações são inflacionadas. Valores de VU iguais a um ocorrem quando a capacidade de armazenamento do reservatório está em seu limite máximo. A Figura 5.4 apresenta a série temporal de VU (a), o componente sazonal presente nos dados (b), o correlograma (c) e o correlograma parcial (d). A Tabela 5.5 contém algumas medidas descritivas da série de VU.

Analisando-se a Figura 5.4 e a Tabela 5.5, percebe-se que no período analisado o valor médio da variável VU mensal é de 0.6162, com coeficiente de variação de 0.4430. A variabilidade do VU pode ser verificada na Figura 5.4 (a), em que o valor mínimo registrado foi de 0.1043, no mês de dezembro de 2020 e o valor máximo foi de 1.0. Podemos observar na Figura 5.4 (b) que há variabilidade intra-anual, alcançando, em média, os menores valores entre outubro e novembro e os maiores valores nos meses de março a abril. Para a seleção, estimação e avaliação de diagnóstico do modelo ajustado, foi usada a mesma abordagem descrita na Seção 5.1. A seguinte covariável, com ajuste de fase, foi introduzida no modelo para representar o componente sazonal mensal:

$$x_t = [\sin(2\pi t/12)], \quad \text{para } t \in \{6, \dots, n + 5\}.$$

A Tabela 5.6 contém as estimativas dos parâmetros do modelo IKARMA(1,1) com covariável selecionada e demais quantidades relevantes, enquanto a Figura 5.2 contém gráficos de diagnósticos residuais: gráfico dos resíduos versus os índices das observações, gráfico QQ normal dos resíduos, correlograma e correlograma parcial dos resíduos.

A Figura 5.2 (a) apresenta os resíduos quantílicos aleatorizados contra seus índices. Observando este gráfico, não se percebe qualquer padrão temporal, havendo distribuição aparentemente aleatória dos resíduos em torno de zero e dentro do intervalo que se estende de -3 a 3 . Deve-se notar que não há resíduo fora do intervalo $[-3, 3]$. A Figura 5.2 (b) apresenta o QQ-plot, de onde se percebe que há boa coincidência entre quantis empíricos residuais e quantis teóricos normais. Trata-se de bom indicativo de que o modelo está corretamente



(a) Dados observados

(b) Sazonalidade

(c) ACF

(d) PACF

Figura 5.4: Série temporal de VU de São Simão, MG: série, componente sazonal, correlograma e correlograma parcial.

Tabela 5.6: Modelo IKARMA(1, 1) ajustado para dados de VU de São Simão, MG: estimativas, erros-padrão, estatísticas z , p -valores z , AIC e SIC, estatística de Ljung-Box e p -valor do teste de Ljung-Box.

Parâmetro	Estimativa	Erro-Padrão	z estat.	$P(> z)$
α	0.1792	0.0536	3.3407	0.0008
ϕ_1	0.7075	0.0400	17.6939	< 0.0001
θ_1	0.1748	0.0599	2.9165	0.0035
φ	3.6852	0.2217	16.6194	< 0.0001
ω_1	-5.1184	0.7662	6.6805	< 0.0001
ω_2	0.6721	0.1996	3.3669	0.0008
β_1	-1.3785	0.1184	11.6447	< 0.0001
AIC = -341.9289, SIC = -323.8088				
Ljung-Box (lag = 20): $Q = 28.049$ (p -valor = 0.1082)				

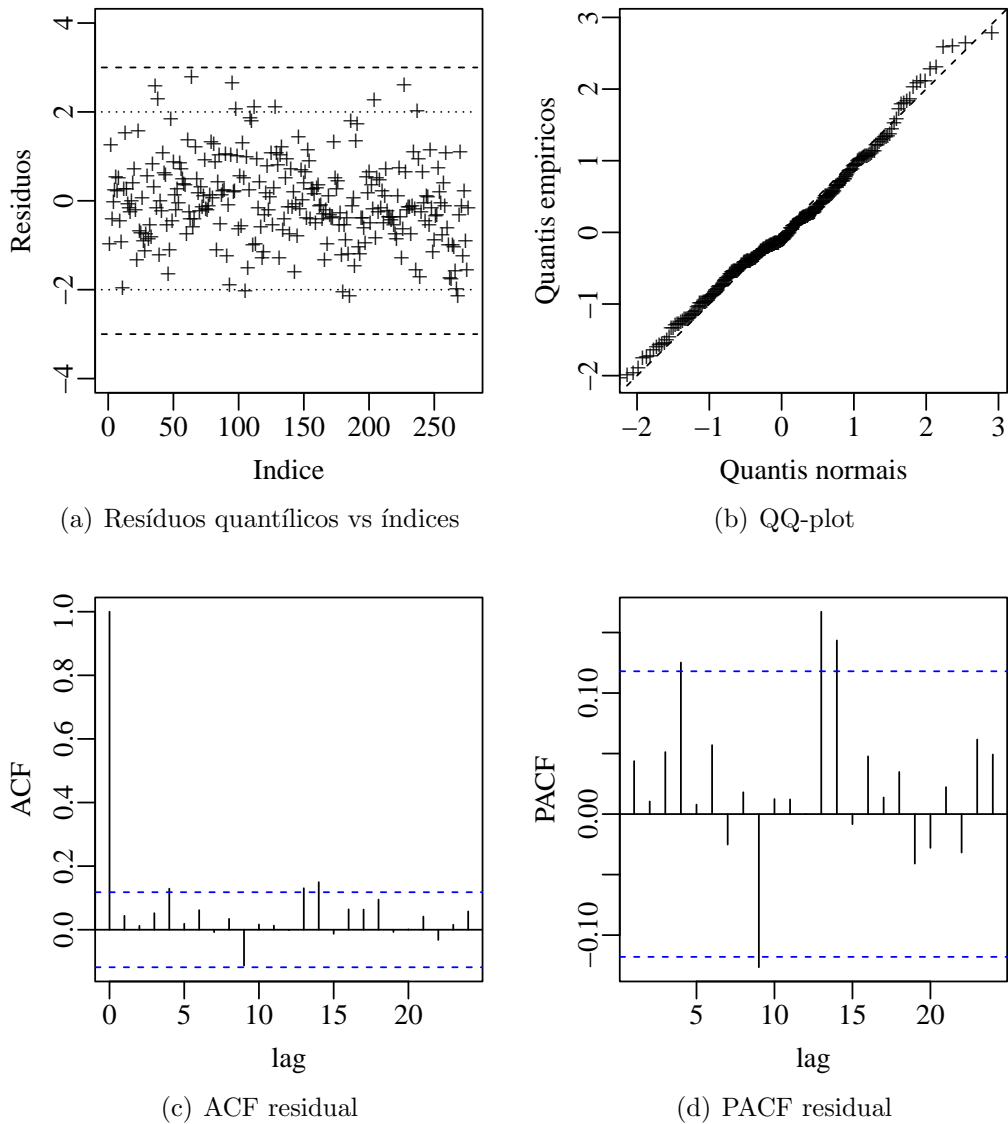


Figura 5.5: Gráficos de diagnóstico do modelo IKARMA ajustado para dados VU de São Simão, MG, com base nos resíduos quantílicos aleatorizados: gráfico dos resíduos contra os índices das observações, QQ-plot dos resíduos, correlograma e correlograma parcial dos resíduos.

especificado. Por fim, o correlograma e o correlograma parcial dos resíduos apresentados nas Figuras 5.2 (c) e (d), respectivamente, não apresentam evidência substantiva contra a boa especificação do modelo; note-se que os intervalos indicados nesses gráficos são assintóticos. Lembre também que o teste de Ljung-Box (Tabela 5.6) não rejeitou a hipótese nula de ausência de correlação serial nos resíduos ao nível de significância de 10%.

Predições dentro e fora da amostra são apresentadas nas Figuras 5.2 (a) e (b), respectivamente. As predições IKARMA foram obtidas a partir das duas estratégias introduzidas na Seção 3.2. Apresentamos também as correspondentes predições obtidas a partir do modelo $I\beta$ ARMA ajustado de mesma ordem e com a mesma covariável. Nesses gráficos, a legenda IKARMA(1, 1) denota os valores preditos por meio da mediana condicional da parte contínua ($\hat{\mu}_t$), enquanto a legenda IKARMA*(1, 1) se refere aos valores preditos pela

Tabela 5.7: Medidas de acurácia de predições para os diferentes modelos ajustados para o VU de São Simão, MG.

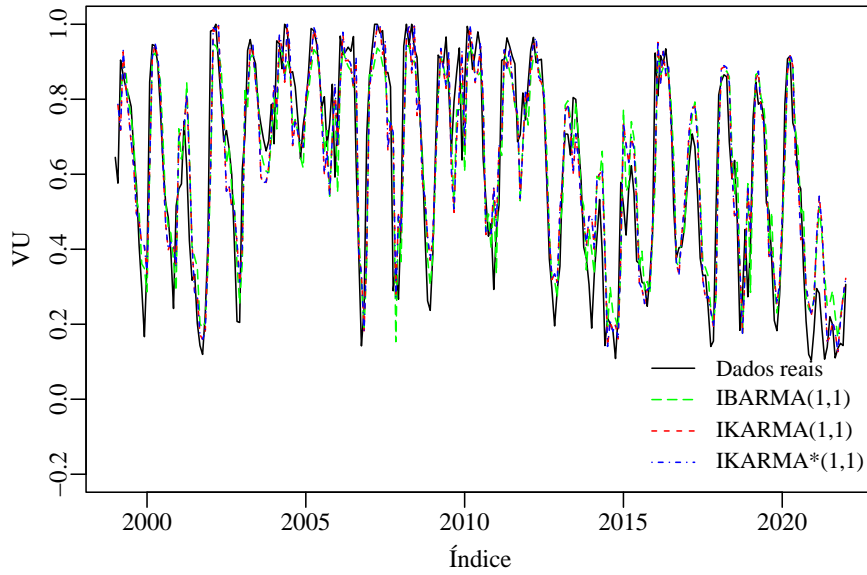
	IKARMA(1, 1)	IKARMA*(1, 1)	IBARMA(1, 1)
MSE	0.0143	0.0143	0.0142
RMSE	0.1196	0.1197	0.1192
MAPE	22.4447	22.5053	22.8158
* Previsão da mediana da variável inflacionada			

Tabela 5.8: Medidas de acurácia de previsões para os diferentes modelos ajustados para o VU de São Simão, MG.

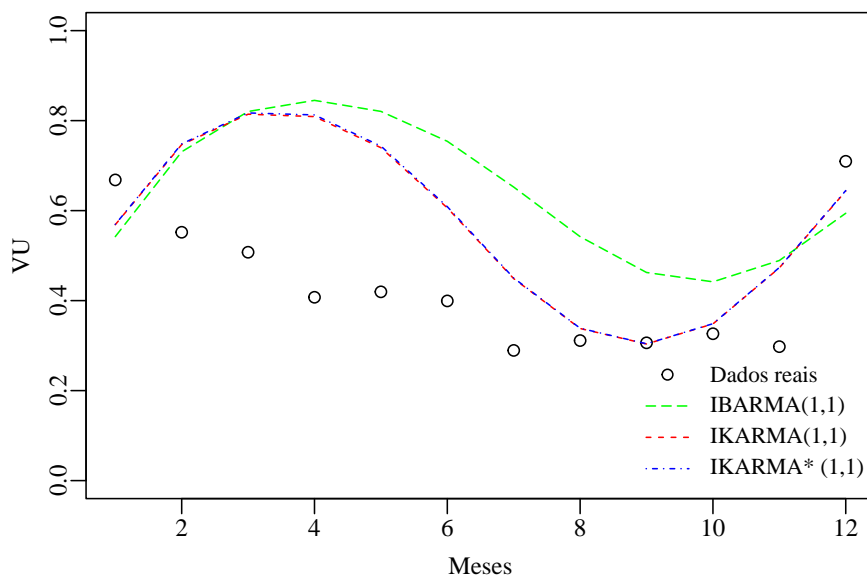
	IKARMA(1, 1)	IKARMA*(1, 1)	IBARMA(1, 1)
MSE	0.0425	0.0432	0.0747
RMSE	0.2062	0.2080	0.2733
MAPE	39.7324	40.0614	64.2817
* Previsão da mediana da variável inflacionada			

mediana da variável inflacionada (\hat{y}_t).

Como na análise anterior (Seção 5.1), observamos que as predições dentro da amostra do modelo $IKARMA^*(1, 1)$ se destacam entre os métodos de predição considerados. Percebe-se que os valores preditos \hat{y}_t estão muito mais próximos dos valores observados iguais a um, capturando adequadamente o padrão sazonal dos dados. A Figura 5.2 (b) mostra os valores previstos fora da amostra para os modelos ajustados. Contudo, visualmente é difícil identificar a superioridade de um dado conjunto de previsões. Para melhor comparação, algumas medidas de qualidade de predição são consideradas, como descritas na Seção 5.1. As Tabelas 5.7 e 5.8 mostram a predição dentro da amostra e previsão, respectivamente. Notamos que o modelo proposto supera o modelo $I\beta ARMA$ em praticamente todas as figuras elencadas, em que destacamos novamente que o modelo $I\beta ARMA$ não consegue fornecer valores preditos de VU iguais a 1, o que pode ser entendido como uma limitação do modelo em aplicações práticas.



(a) Predição dentro da amostra.



(b) Previsão fora da amostra.

Figura 5.6: Predições e previsões obtidas a partir dos diferentes modelos ajustados para o VU de São Simão, MG.

Capítulo 6

CONCLUSÕES

Neste trabalho, foi proposto um novo modelo dinâmico para modelagem e previsão de séries temporais duplamente limitadas e sujeitas a inflacionamento que assumem valores em $[0, 1)$ ou $(0, 1]$. Modelamos a mediana condicional de uma variável aleatória com distribuição condicional Kumaraswamy inflacionada, que envolve mistura de distribuições contínua e discreta. Tal distribuição permite a modelagem de dados duplamente limitados com valores no intervalo unitário sujeitos a inflacionamento em zero ou em um. Além de uma estrutura dinâmica para a mediana condicional, é proposta uma estrutura dinâmica parcimoniosa para o parâmetro de mistura, permitindo-se assim que a probabilidade de inflacionamento varie ao longo do tempo.

Além da proposição do modelo, mostramos como realizar inferência sobre os parâmetros que indexam o modelo IKARMA sob a perspectiva de máxima verossimilhança condicional. Expressões matriciais em forma fechada para o vetor escore condicional e para a matriz de informação de Fisher condicional foram obtidas. Com base em resultados assintóticos, foram abordadas a construção de intervalos de confiança e a realização de testes de hipóteses. Também propusemos um conjunto de técnicas para análise de diagnóstico que podem ser empregadas para identificar má especificação do modelo ajustado e a presença de observações atípicas e influentes nos dados. Essas ferramentas incluem análise de resíduos e procedimentos de seleção de modelos. Adicionalmente, ferramentas de previsão também foram discutidas, tendo sido propostas duas formas para obtenção de valores preditos.

Após a formalização da teoria, experimentos computacionais foram realizados. Para avaliar o desempenho dos estimadores de máxima verossimilhança condicionais (EMVC) em amostras de tamanho finito, um estudo de simulação de Monte Carlo foi executado. O estudo da simulação evidenciou que os estimadores dos parâmetros do modelo possuem boas propriedades, mesmo em tamanhos amostrais relativamente pequenos. Por fim, duas ilustrações empíricas foram apresentadas e discutidas para evidenciar a aplicabilidade do modelo proposto IKARMA. Na primeira análise foram utilizados dados de umidade relativa do ar máxima mensal na cidade de Caxias do Sul, RS, inflacionados em um. Já na segunda análise foram utilizados dados do volume útil da Usina Hidrelétrica São Simão, localizada no estado de Minas Gerais, Brasil, também inflacionados em um. Nas duas aplicações modelamos os dados por meio do modelo IKARMA e, comparativamente, pelo modelo $I\beta$ ARMA. Os resultados das aplicações

evidenciaram superioridade do modelo proposto na predição de umidade relativa e volume útil dentro e fora da amostra.

De forma pontual, destacamos as seguintes diferenças entre o modelo proposto no presente trabalho e aqueles disponíveis na literatura:

- O modelo proposto considera dados duplamente limitados e inflacionados, diferentemente dos modelos KARMA e β ARMA.
- Consideramos duas estruturas dinâmicas no modelo, uma para a mediana condicional μ_t e outra para o parâmetro de mistura λ_t . Essa abordagem com duas estruturas dinâmicas não é encontrada em qualquer outro modelo da literatura.
- Em uma das formas de predição propostas, os valores preditos podem assumir valores iguais a zeros e uns, a depender do valor estimado para o parâmetro de mistura λ_t , em cada instante t . Não é possível obter valores preditos iguais a zero e/ou um a partir do modelo concorrente I β ARMA.
- Uma vez que o modelo IKARMA se baseia na família de distribuição Kumaraswamy, o mesmo se torna importante alternativa para modelagem de dados duplamente limitados e inflacionados em áreas hidroambientais.

Como trabalhos futuros, identificamos os seguintes possíveis avanços:

- Proposição de um novo modelo onde dados duplamente limitados serão considerados com inflacionamento duplo, ou seja, com suporte em $[0, 1)$, $(0, 1]$ ou $[0, 1]$. Assim, a aplicabilidade do modelo se tornaria mais abrangente.
- Desenvolvimento de um novo modelo dinâmico bidimensional, como uma extensão do modelo IKARMA (1D) para imagens (2D), contemplando aspectos metodológicos e aplicações relacionadas à modelagem de imagens digitais advindas de sensoriamento remoto, em abordagem semelhante à considerada em Palm et al. (2022).

Bibliografia

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723.
- Alduais, F.S., Yassen, M.F., Almazah, M.M., Khan, Z., 2022. Estimation of the Kumaraswamy distribution parameters using the E-Bayesian method. *Alexandria Engineering Journal* 61, 11099–11110.
- Alves, E.C., da Silva, C.F., Cossich, E.S., Tavares, C.R.G., de Souza Filho, E.E., Carniel, A., 2008. Avaliação da qualidade da água da bacia do rio Pirapó–Maringá, Estado do Paraná, por meio de parâmetros físicos, químicos e microbiológicos. *Acta Scientiarum. Technology* 30, 39–48.
- Andersen, B.A., 1970. Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistic Society Serie B* 32, 283–301.
- Barcellos, C.C., Monteiro, A.M.V., Corvalán, C., Gurgel, H.C., Carvalho, M.S., Artaxo, P., Hacon, S., Ragoni, V., et al., 2009. Mudanças climáticas e ambientais e as doenças infecciosas: cenários e incertezas para o Brasil. *Epidemiologia e Serviços de Saúde* 18, 285–304.
- Bayer, D.M., Bayer, F., 2015. Relative air humidity forecasting through beta autoregressive moving average model in Brasília. *Revista Brasileira de Meteorologia* 30, 319–326.
- Bayer, F.M., Bayer, D.M., Pumi, G., 2017. Kumaraswamy autoregressive moving average models for double bounded environmental data. *Journal of Hydrology* 555, 385–396.
- Bayer, F.M., Cintra, R.J., Cribari-Neto, F., 2018. Beta seasonal autoregressive moving average models. *Journal of Statistical Computation and Simulation* 88, 2961–2981.
- Bayer, F.M., Cribari-Neto, F., Santos, J., 2021. Inflated Kumaraswamy regressions with application to water supply and sanitation in Brazil. *Statistica Neerlandica* 75, 453–481.
- Bayer, F.M., Pumi, G., Pereira, T.L., Souza, T.C., 2023. Inflated beta autoregressive moving average models. *Computational and Applied Mathematics* 42.
- Bayes, C.L., Valdivieso, L., 2016. A beta inflated mean regression model for fractional response variables. *Journal of Applied Statistics* 43, 1814–1830.

- Benjamin, M.A., Rigby, R.A., Stasinopoulos, D.M., 2003. Generalized autoregressive moving average models. *Journal of the American Statistical Association* 98, 214–223.
- Bloomfield, P., 2004. *Fourier analysis of time series: an introduction*. John Wiley & Sons.
- Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.
- Brazilian National Institute of Meteorology (INMET), 2018. Meteorological database for research and teaching. Available in: <http://www.inmet.gov.br/projetos/rede/pesquisa>. Access: October 2018.
- Brockwell, P.J., Davis, R.A., 2009. *Time series: theory and methods*. Springer Science & Business Media.
- Cordeiro, G.M., de Castro, M., 2011. A new family of generalized distributions. *Journal of Statistical Computation and Simulation* 81, 883–898.
- Cox, D.R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., Lauritzen, S.L., 1981. Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics* , 93–115.
- Cox, D.R., Snell, E.J., 1968. A general definition of residuals. *Journal of the Royal Statistical Society: Series B* 30, 248–265.
- Cribari-Neto, F., Santos, J., 2019. Inflated Kumaraswamy distributions. *Anais da Academia Brasileira de Ciências* 91.
- Cribari-Neto, F., Scher, V.T., Bayer, F.M., 2023. Beta autoregressive moving average model selection with application to modeling and forecasting stored hydroelectric energy. *International Journal of Forecasting* 39, 98–109.
- Cribari-Neto, F., Zeileis, A., 2010. Beta regression in R. *Journal of Statistical Software* 34, 1–24.
- Dunn, P.K., Smyth, G.K., 1996. Randomized quantile residuals. *Journal of Computational and Graphical Statistics* 5, 236–244.
- Eletrobras Furnas, 2023. Available in: <https://www.furnas.com.br/>. Access: March 2023.
- Fahrmeir, L., Kaufmann, H., 1985. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics* 13, 342–368.
- Ferrari, S., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31, 799–815.

- Ferreira, D.B., Barroso, G.R., Dantas, M.S., Oliveira, K.L.d., Christofaro, C., Oliveira, S.C., 2021. Pluviometric patterns in the São Francisco River basin in Minas Gerais, Brazil. *RBRH* 26.
- Fokianos, K., Kedem, B., 1998. Prediction and classification of non-stationary categorical time series. *Journal of Multivariate Analysis* 67, 277–296.
- Fokianos, K., Kedem, B., 2004. Partial likelihood inference for time series following generalized linear models. *Journal of Time Series Analysis* 25, 173–197.
- Fontenele, S.B., de Andrade, E.M., Salgado, E.V., Meireles, A.C.M., Sabiá, R.J., 2011. Análise espaço-temporal da qualidade da água na parte alta da bacia do rio Salgado, Ceará. *Revista Caatinga* 24, 102–109.
- Galvis, D.M., Bandyopadhyay, D., Lachos, V.H., 2014. Augmented mixed beta regression models for periodontal proportion data. *Statistics in Medicine* 33, 3759–3771.
- Gupta, A.K., Nadarajah, S., 2004. Handbook of beta distribution and its applications. CRC press.
- Hannan, E.J., Quinn, B.G., 1979. The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B* 41, 190–195.
- Jones, M., 2009. Kumaraswamy’s distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology* 6, 70–81.
- Kedem, B., Fokianos, K., 2005. Regression models for time series analysis. John Wiley & Sons.
- Kumaraswamy, P., 1980. A generalized probability density function for double-bounded random processes. *Journal of Hydrology* 46, 79–88.
- Lemonte, A.J., 2011. Improved point estimation for the Kumaraswamy distribution. *Journal of Statistical Computation and Simulation* 81, 1971–1982.
- Lemonte, A.J., Barreto-Souza, W., Cordeiro, G.M., 2013. The exponentiated Kumaraswamy distribution and its log-transform. *Brazilian Journal of Probability and Statistics* 27, 31–53.
- Lemonte, A.J., Bazán, J.L., 2016. New class of Johnson distributions and its associated regression model for rates and proportions. *Biometrical Journal* 58, 727–746.
- Li, W.K., 1994. Time series models based on generalized linear models: some further results. *Biometrics* , 506–511.
- Lindim, C., Pinho, J., Vieira, J., 2011. Analysis of spatial and temporal patterns in a large reservoir using water quality and hydrodynamic modeling. *Ecological Modelling* 222, 2485–2494.

- Liu, C., Zhang, Y., Sun, L., Gao, W., Jing, X., Ye, W., 2021. Influence of indoor air temperature and relative humidity on learning performance of undergraduates. *Case Studies in Thermal Engineering* 28, 101458.
- Liu, F., Eugenio, E.C., 2018. A review and comparison of Bayesian and likelihood-based inferences in beta regression and zero-or-one-inflated beta regression. *Statistical Methods in Medical Research* 27, 1024–1044.
- Liu, F., Kong, Y., 2015. zoib: An R package for Bayesian inference for beta regression and zero/one inflated beta regression. *R Jornal* 7, 34.
- Ljung, G.M., Box, G.E., 1978. On a measure of lack of fit in time series models. *Biometrika* 65, 297–303.
- McCullagh, P., Nelder, J., 1989. *Generalized Linear Models.* (Chapman & Hall/CRC: London.).
- Mitnik, P.A., 2013. New properties of the Kumaraswamy distribution. *Communications in Statistics-Theory and Methods* 42, 741–755.
- Mitnik, P.A., Baek, S., 2013. The Kumaraswamy distribution: median-dispersion reparameterizations for regression modeling and simulation-based estimation. *Statistical Papers* 54, 177–192.
- Mohsenkhani, Z.F., Mohhammadzadeh, M., Baghfalaki, T., 2019. Augmented mixed beta regression models with skew-normal independent distributions: Bayesian analysis of labor force data. *Communications in Statistics-Simulation and Computation* 48, 2147–2164.
- Nadarajah, S., 2008. On the distribution of Kumaraswamy. *Journal of Hydrology* 348, 568–569.
- Neyman, J., Pearson, E.S., 1928. On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. *Biometrika* 20A, 175–240.
- Nogarotto, D.C., Azevedo, C.L.N., Bazán, J.L., 2020. Bayesian modeling and prior sensitivity analysis for zero–one augmented beta regression models with an application to psychometric data. *Brazilian Journal of Probability and Statistics* 34, 304–322.
- Nozari, H., Moradi, P., Godarzi, E., 2021. Simulation and optimization of control system operation and surface water allocation based on system dynamics modeling. *Journal of Hydroinformatics* 23, 211–230.
- Operador Nacional do Sistema Elétrico: Dados Hidrológicos, 2023. Available in: https://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/dados_hidrologicos_volumes.aspx. Access: March 2023.
- Ospina, R., Ferrari, S.L., 2010. Inflated beta distributions. *Statistical Papers* 51, 111–126.

- Ospina, R., Ferrari, S.L., 2012. A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis* 56, 1609–1623.
- Ozbek, A., Ünal, Ş., Bilgili, M., 2022. Daily average relative humidity forecasting with LSTM neural network and ANFIS approaches. *Theoretical and Applied Climatology* 150, 697–714.
- Palm, B.G., Bayer, F.M., 2018. Bootstrap-based inferential improvements in beta autoregressive moving average model. *Communications in Statistics-Simulation and Computation* 47, 977–996.
- Palm, B.G., Bayer, F.M., Cintra, R.J., 2021a. Prediction intervals in the beta autoregressive moving average model. *Communications in Statistics-Simulation and Computation* Forthcoming, 1–22.
- Palm, B.G., Bayer, F.M., Cintra, R.J., 2021b. Signal detection and inference based on the beta binomial autoregressive moving average model. *Digital Signal Processing* 109, 102911.
- Palm, B.G., Bayer, F.M., Cintra, R.J., 2022. 2-D Rayleigh autoregressive moving average model for SAR image modeling. *Computational Statistics & Data Analysis* 171, 107453.
- Parishwad, G., Bhardwaj, R., Nema, V., 1998. Prediction of monthly-mean hourly relative humidity, ambient temperature, and wind velocity for India. *Renewable Energy* 13, 363–380.
- Pawitan, Y., 2001. *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge University Press.
- Pumi, G., Prass, T.S., Souza, R.R., 2021. A dynamic model for double-bounded time series with chaotic-driven conditional averages. *Scandinavian Journal of Statistics* 48, 68–86.
- Pumi, G., Valk, M., Bisognin, C., Bayer, F.M., Prass, T.S., 2019. Beta autoregressive fractionally integrated moving average models. *Journal of Statistical Planning and Inference* 200, 196–212.
- R Core Team, R., 2021. *R: A language and environment for statistical computing* .
- Rao, C.R., 1948. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, in: *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press. pp. 50–57.
- Rocha, A.V., Cribari-Neto, F., 2009. Beta autoregressive moving average models. *Test* 18, 529–545.
- Sagrillo, M., Guerra, R.R., Bayer, F.M., 2021. Modified Kumaraswamy distributions for double bounded hydro-environmental data. *Journal of Hydrology* 603, 127021.

- Scher, V.T., Cribari-Neto, F., Pumi, G., Bayer, F.M., 2020. Goodness-of-fit tests for β ARMA hydrological time series modeling. *Environmetrics* 31, e2607.
- Schwarz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics* , 461–464.
- Silva, R.B., Barreto-Souza, W., 2014. Beta and Kumaraswamy distributions as non-nested hypotheses in the modeling of continuous bounded data. arXiv preprint arXiv:1406.1941 .
- Sousa, N.M.N., Dantas, R.T., Limeira, R.C., 2007. Influência de variáveis meteorológicas sobre a incidência do dengue, meningite e pneumonia em João Pessoa-PB. *Revista Brasileira de Meteorologia* 22, 183–192.
- Von Sperling, M., 2005. Introdução à qualidade das águas e ao tratamento de esgotos: princípios do tratamento biológico de águas residuárias. Belo Horizonte: UFMG 1, 452.
- Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* 54, 426–482.
- Wang, B.X., Wang, X.K., Yu, K., 2017. Inference on the Kumaraswamy distribution. *Communications in Statistics-Theory and Methods* 46, 2079–2090.
- Zeger, S.L., 1988. A regression model for time series of counts. *Biometrika* 75, 621–629.