



Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Programa de Pós-Graduação em Estatística

Bayesian Analysis of Beta Autoregressive Moving Average Models

Aline Foerster Grande

Porto Alegre, Março de 2023.

CIP - Catalogação na Publicação

Grande, Aline Foerster
Bayesian Analysis of Beta Autoregressive Moving
Average Models / Aline Foerster Grande. -- 2023.
47 f.
Orientador: Guilherme Pumi.

Coorientadora: Gabriela Bettella Cybis.

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. Modelos Beta ARMA. 2. Séries Temporais. 3.
Análise Bayesiana. 4. Modelos Não-Gaussianos. I. Pumi,
Guilherme, orient. II. Cybis, Gabriela Bettella,
coorient. III. Título.

Dissertação submetida por Aline Foerster Grande 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(a):

Prof. Dr. Guilherme Pumi

Co-orientador(a):

Prof. Dra. Gabriela Bettella Cybis

Comissão Examinadora:

Prof. Dra. Taiane Schaedler Prass (PPGEst - UFRGS)

Prof. Dr. Fábio Mariano Bayer (PPGEst - UFRGS)

Prof. Dr. Dani Gamerman (PPGE - UFRJ)

Data de Apresentação: 28 de Março de 2023

AGRADECIMENTOS

Gostaria de agradecer aos meus pais, Angela e Pedro, pelo apoio, incentivo e por estarem sempre do meu lado. À minha irmã, Amanda, pela amizade e confiança. Aos meus avós, tios, tias, primos e primas, por sempre estarem torcendo por mim. Aos meus amigos do colégio e faculdade, pelo aprendizado e por terem deixado essa jornada mais animada. Em especial a Daniela, Gabriela, Katerine, Mariana e Vitor, pela amizade sincera, por me ajudarem sempre quando precisei e pelos momentos divertidos que tivemos. À todos os professores que me ajudaram de alguma forma na minha formação. À professora Gabriela, minha coorientadora, e ao professor Guilherme, meu orientador, por toda a dedicação, por todos os ensinamentos, por sempre estarem dispostos a me ajudar e por sempre acreditarem na minha capacidade. Ao Humberto, pelo suporte, amor e companheirismo.

RESUMO

O presente trabalho propõe uma abordagem Bayesiana para a estimação dos parâmetros do modelo β ARMA(p, q), modelos de séries temporais para dados com suporte no intervalo $(0, 1)$. Para tanto, emprega-se a técnica de amostragem Monte Carlo Hamiltoniano, reconhecida por sua eficiência computacional na estimação de parâmetros em modelos mais complexos. O estudo conduz simulações de Monte Carlo considerando modelos β ARMA sob diversos cenários, bem como uma análise de sensibilidade com relação à escolha das prioris utilizadas e a detecção de raízes unitárias. Para ilustrar a aplicação da abordagem proposta, são utilizados dados de energia hidrelétrica como exemplo.

Palavras-chave: *Séries Temporais, Análise Bayesiana, Modelos Não-Gaussianos, Modelos β ARMA.*

ABSTRACT

The present work proposes a Bayesian approach to estimate the parameters of β ARMA(p, q) models, which are used for time series data with values in the interval $(0, 1)$. To achieve this goal, the study employs the Hamiltonian Monte Carlo sampling technique, which is known for its effectiveness in estimating parameters in complex models. The study also conducts Monte Carlo simulations to examine the performance of the proposed approach under different scenarios. Additionally, the sensitivity of the results to prior selection and unit roots detection is evaluated. To demonstrate the applicability of the proposed approach, the study provides an empirical analysis using hydroelectric energy data.

Keywords: *Time Series, Bayesian Analysis, Non-Gaussian Models, β ARMA Models.*

ÍNDICE

1	Introdução	2
2	Modelos βARMA	4
2.1	Modelos Autoregressivos (AR)	4
2.2	Modelos de Médias Móveis (MA)	5
2.3	Modelos Autoregressivos de Médias Móveis (ARMA)	5
2.4	Modelos β ARMA	6
3	Métodos Computacionais para Inferência Bayesiana	8
3.1	Inferência Bayesiana	8
3.2	Monte Carlo via Cadeias de Markov	9
3.2.1	Algoritmo Metropolis–Hastings	10
3.3	Monte Carlo Hamiltoniano	11
3.3.1	Algoritmo Monte Carlo Hamiltoniano	13
4	Considerações finais	15
	Apêndice	18
A	Artigo Grande et al. (2023)	19

CAPÍTULO 1

INTRODUÇÃO

A modelagem de dados tomando valores no intervalo $(0, 1)$, como razões e proporções, é uma área que vem crescendo rapidamente nos últimos anos. Diversas abordagens têm sido propostas na literatura para esse tipo de análise, incluindo trabalhos de referência ([Kumaraswamy, 1980](#); [Kieschnick and McCullough, 2003](#); [Ferrari and Cribari-Neto, 2004](#); [Melo et al., 2009](#); [Gamerman and Cepeda-Cuervo, 2013](#)). Uma área particularmente fértil de modelagem é a de séries temporais limitadas no suporte $(0, 1)$, que tem sido objeto de estudo de pesquisadores ([Zheng et al., 2015](#); [Bayer et al., 2017, 2018](#); [Pumi et al., 2019](#); [Scher et al., 2020](#)).

Modelos para dados limitados são não-Gaussianos por natureza e, embora várias abordagens foram propostas na literatura, como mencionado acima, algumas que vem se destacando são GLARMA (do inglês, Generalized Linear Autoregressive Moving Average Models) ([Franco et al., 2019](#)) e GARMA (Generalized Autoregressive Moving Average Models), uma abordagem mais geral, e sistematizada por [Benjamin et al. \(2003\)](#). A ideia da abordagem GARMA é combinar a simplicidade dos modelos ARMA com a flexibilidade dos modelos lineares generalizados ([Nelder and Wedderburn, 1972](#); [McCullagh and Nelder, 1989](#)), dentro de uma estrutura de modelos *observation driven* ([Cox et al., 1981](#)). Um dos modelos mais utilizados na prática, e que será o objeto de interesse neste estudo, são os modelos β ARMA. [McKenzie \(1985\)](#) introduziu o β AR, enquanto que [Rocha and Cribari-Neto \(2009\)](#) introduziu o modelo β ARMA completo. Nele, o componente aleatório é especificado através de uma distribuição beta enquanto o componente sistemático do modelo segue uma especificação ARMA. Em [Rocha and Cribari-Neto \(2009\)](#), uma abordagem frequentista baseada na verossimilhança condicional é utilizada para a estimação dos parâmetros do modelo. Modelos de regressão beta e modelos β ARMA são aplicados nas mais variadas áreas, por exemplo, em medicina ([Zou et al., 2010](#); [Rogers et al., 2012](#)), monitoramento online ([Guolo and Varin, 2014](#)), engenharia metalúrgica ([Pereira and Tavares, 2011](#)), neurociência ([Wang, 2012](#)) e pesquisa operacional ([Seifi et al., 2000](#)).

Quando falamos no contexto Bayesiano, [Amisano et al. \(2007\)](#); [Taddy \(2010\)](#); [Billio and Casarin \(2010, 2011\)](#) apresentam abordagens para estimação de processos beta autorregressivos (β AR) de primeira ordem. Em [Casarin et al. \(2012\)](#), os autores propõe uma abordagem Bayesiana para seleção de modelos no contexto de modelos β AR(p), baseada em cadeias de Markov de salto reversível (reversible-jump Markov chain Monte Carlo), uma extensão dos algoritmo de Markov chain Monte Carlo (MCMC) para o caso em que a dimensão do modelo pode variar ([Green, 1995](#)). Um ponto negativo em [Casarin et al. \(2012\)](#) é que os autores consideram uma abordagem diferente da geralmente utilizada na literatura de modelos β ARMA ao considerarem o link como sendo a identidade, o que

implica a necessidade de se impor limitações nos parâmetros do modelo para garantir que o modelo gere médias condicionais limitadas ao intervalo $(0, 1)$. Esta abordagem dificulta a sua utilização em problemas aplicados.

O principal produto deste trabalho é o artigo “Bayesian Analysis of Beta Autoregressive Moving Average Models”, apresentado no Apêndice. Nele propôs-se a utilização do método Monte Carlo Hamiltoniano (Duane et al., 1987; Mackay, 2003; Neal, 2011) para a estimação dos parâmetros do modelo β ARMA(p, q) como introduzido em Rocha and Cribari-Neto (2009). Uma série de simulações de Monte Carlo foram implementadas no RStan para avaliar a metodologia Bayesiana proposta. Em um primeiro momento avaliou-se o desempenho da abordagem bayesiana em termos de estimativa pontual e intervalar. Baseado na distribuição a priori dos parâmetros da parte autoregressiva, propôs-se um método para abordar o problema de identificação de raízes unitárias no componente sistemático do modelo, visto que as abordagens tradicionais utilizadas em modelos ARMA não são aplicáveis no contexto de modelos β ARMA. O artigo também inclui uma análise de sensibilidade com relação à escolha dos hiperparâmetros das prioris utilizadas. Por fim, uma aplicação empírica para exemplificar a utilidade do método é apresentada, na qual empregou-se Bayes Factor para seleção de modelos, e realizou-se previsões fora da amostra para explorar a posteriori preditiva e comparar as abordagens Bayesiana e frequentista. As simulações foram realizadas usando o software R (R Core Team, 2022). A geração de dados foi realizada usando o pacote BTSR (Prass and Pumi, 2022). A estimativa dos parâmetros foram realizadas com o auxílio do pacote RStan (Stan Development Team, 2022). Todos os dados e códigos utilizados no artigo estão disponíveis em github.com/AlineFoersterGrande/Bayesian.BARMA.

Este trabalho está organizado da seguinte forma. O Capítulo 2 revisa os modelos ARMA e os modelos β ARMA. O Capítulo 3 investiga os principais conceitos computacionais associados à inferência Bayesiana, como o algoritmo de Metropolis-Hastings e o Monte Carlo Hamiltoniano. Por fim, o Capítulo 4 apresenta as principais conclusões alcançadas neste trabalho, além de discutir ideias para futuras pesquisas. O artigo Grande et al. (2023), principal produto desta dissertação, encontra-se no Apêndice A.

CAPÍTULO 2

MODELOS β ARMA

No presente capítulo, revisamos alguns modelos para séries temporais. Iniciaremos com a discussão acerca dos modelos autoregressivos (AR) e médias móveis (MA), para posteriormente abordar os modelos autoregressivos de médias móveis (ARMA). Por último, discutiremos os modelos β ARMA.

2.1 Modelos Autoregressivos (AR)

Os modelos autoregressivos (AR) são uma classe de modelos estatísticos utilizados para modelar a dependência serial de uma série temporal. Neles o valor atual da variável em estudo é modelado como uma função linear de seus próprios valores passados, com um termo de erro aleatório adicionado. Dizemos que um processo $\{Y_t\}_{t \in \mathbb{Z}}$ é um $\text{AR}(p)$ se satisfaz

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t, \quad (2.1)$$

onde $\phi := (\phi_1, \dots, \phi_p)'$ denotam os coeficientes autoregressivos do modelo, ε_t é um ruído branco com $\mathbb{E}(\varepsilon_t) = 0$ e $\text{Var}(\varepsilon_t) = \sigma^2 < \infty$. Percebe-se que a variável resposta Y_t é modelada como uma relação linear dos seus p valores passados mais próximos, isto é, Y_{t-1}, \dots, Y_{t-p} . O modelo (2.1) pode ser escrito em função do operador lag $L^k(Y_t) := Y_{t-k}$ como

$$\phi(L)Y_t = (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)Y_t = \varepsilon_t.$$

Uma condição necessária e suficiente para a existência de uma solução estacionária para o sistema (2.1) é que as raízes do polinômio

$$\phi(z) := 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p, \quad z \in \mathbb{C}. \quad (2.2)$$

denominado polinômio característico associado ao componente $\text{AR}(p)$, sejam todas, em módulo, diferentes de 1, isto é,

$$\phi(z) = 0 \implies |z| \neq 1.$$

Quanto à causalidade, uma condição necessária e suficiente é que as raízes do polinômio característico associado (2.2) sejam, em módulo, maiores de 1, isto é,

$$\phi(z) = 0 \implies |z| > 1.$$

2.2 Modelos de Médias Móveis (MA)

Um modelo de médias móveis de ordem q , denominado por $MA(q)$, é visto como um modelo que apresenta apenas um termo de erro, ε_t , e as suas defasagens. Mais formalmente, dizemos que um processo $\{Y_t\}_{t \in \mathbb{Z}}$ é um $MA(q)$ se satisfaz

$$Y_t = \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t, \quad (2.3)$$

onde $\boldsymbol{\theta} := (\theta_1, \dots, \theta_q)'$ são os coeficientes do modelo e ε_t é um ruído branco com $\mathbb{E}(\varepsilon_t) = 0$ e $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$. Observe que um modelo MA é sempre estacionário, visto que é uma função mensurável de um processo estacionário. Um modelo $MA(q)$ tem média e variância constantes e autocovariâncias que podem ser diferentes de zero até o lag q , mas são iguais a zero depois. O modelo (2.3) pode ser reescrito como

$$Y_t = \theta(L)\varepsilon_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)\varepsilon_t.$$

O polinômio característico associado a um processo $MA(q)$ é dado por

$$\theta(z) := 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q, \quad z \in \mathbb{C}. \quad (2.4)$$

Um processo $MA(q)$ é invertível se pode ser escrito como um processo autoregressivo de ordem infinita, ou seja, um $AR(\infty)$. Logo, o modelo $AR(p)$ é sempre invertível. A invertibilidade é importante por três motivos: “primeiro, sem invertibilidade, a série não poderia ser estimada recursivamente, usando observações passadas; segundo, para haver unicidade de resultados; terceiro, para gerar a função de autocorrelação parcial” (Bueno, 2008).

Uma condição necessária e suficiente para a invertibilidade de um modelo $MA(q)$ é que as raízes do polinômio característico associado (2.4) sejam, em módulo, maiores que 1, isto é

$$\theta(z) = 0 \implies |z| > 1.$$

2.3 Modelos Autoregressivos de Médias Móveis (ARMA)

A combinação dos processos $AR(p)$ e $MA(q)$ resultam em processos denominados $ARMA(p, q)$, onde p se refere à ordem do componente AR e q se refere à ordem do componente MA. Um processo $\{Y_t\}_{t \in \mathbb{Z}}$ é dito ser um $ARMA(p, q)$ se satisfaz

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t, \quad (2.5)$$

onde $\boldsymbol{\phi} := (\phi_1, \dots, \phi_p)'$ e $\boldsymbol{\theta} := (\theta_1, \dots, \theta_q)'$ são os coeficientes do modelo e ε_t é um ruído branco com $\mathbb{E}(\varepsilon_t) = 0$ e $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$. O modelo (2.5) pode ser escrito como

$$\phi(L)Y_t = \theta(L)\varepsilon_t, \quad (2.6)$$

e $\phi(L)$ e $\theta(L)$ são polinômios das partes AR e MA, definidos em (2.2) e (2.4), respectivamente e que assumimos não terem raízes em comum. A existência de soluções estacionárias para (2.6) está

intimamente ligada à estacionariedade da parte AR. (2.6) possui solução estacionária se, e somente se,

$$\phi(z) = 0 \implies |z| \neq 1.$$

Quando existe solução estacionária para (2.6), ela é única. A causalidade de processos ARMA(p, q) se resume à causalidade da parte AR enquanto que a invertibilidade depende apenas da parte MA. Em resumo, existe solução causal em (2.6), se e somente se,

$$\phi(z) = 0 \implies |z| > 1,$$

e invertível, se e somente se,

$$\theta(z) = 0 \implies |z| > 1.$$

Mais detalhes relacionados aos modelos ARMA podem ser encontrados em [Brockwell and Davis \(1991\)](#)

2.4 Modelos β ARMA

Os modelos β ARMA combinam a simplicidade e flexibilidade dos modelos autoregressivos de médias móveis (ARMA), clássicos na literatura de séries temporais ([Brockwell and Davis, 1991](#)), com a flexibilidade dos modelos lineares generalizados ([Nelder and Wedderburn, 1972](#)), numa roupagem *observation driven*, especificado através de um componente aleatório e um sistemático. O componente aleatório modela a distribuição da variável resposta Y_t condicionada à informação observada até o tempo t , enquanto o componente sistemático modela a estrutura de dependência da média condicional de Y_t . Nesta seção, primeiramente, definiremos a parte do componente aleatório do modelo, e posteriormente, a parte do componente sistemático.

Seja $\{Y_t\}_{t \in \mathbb{Z}}$ uma série temporal de interesse assumindo valores em $(0, 1)$ e $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ um conjunto l -dimensional de covariáveis exógenas possivelmente aleatórias e possivelmente dependentes do tempo. Seja \mathcal{F}_t a σ -álgebra que representa o histórico observado do modelo até o momento t , ou seja, a σ -álgebra gerada por $(\mathbf{X}'_t, Y_t, \mathbf{X}'_{t-1}, Y_{t-1}, \dots)$. O modelo apresenta como componente aleatório uma distribuição beta, parametrizada como

$$p(y|\nu, \mu_t, \mathcal{F}_{t-1}) = \frac{\Gamma(\nu)}{\Gamma(\nu\mu_t)\Gamma(\nu(1-\mu_t))} y^{\nu\mu_t-1} (1-y)^{\nu(1-\mu_t)-1}, \quad (2.7)$$

onde $0 < y < 1$, $0 < \mu_t < 1$ e $\nu > 0$. O parâmetro ν pode ser interpretado como um parâmetro de precisão, quanto maior o seu valor, menor a variância condicional (para μ_t fixo). Observe que a densidade (2.7) está parametrizada em termos da média condicional μ_t e de um parâmetro de precisão ν , enquanto que, na parametrização tradicional, a densidade é dada por (2.8), onde a e b são denominados parâmetros de forma.

$$p(y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1} I(0 < y < 1) \quad (2.8)$$

Comparando (2.7) e (2.8) encontramos que $a = \mu_t\nu$ e $b = (1-\mu_t)\nu$, de forma que,

$$\mathbb{E}(Y_t|\mathcal{F}_{t-1}) = \frac{a}{a+b} = \mu_t,$$

$$\text{Var}(Y_t | \mathcal{F}_{t-1}) = \frac{ab}{(a+b)^2(a+b+1)} = \frac{\mu_t(1-\mu_t)}{1+\nu}.$$

O componente sistemático do modelo segue a ideia tradicional de um modelo linear generalizado com um termo aditivo extra, sendo expresso por

$$g(\mu_t) := \mathbf{X}'_t \boldsymbol{\beta} + \tau_t, \quad (2.9)$$

em que τ_t é um termo dinâmico adicional, responsável por modelar a dependência serial de μ_t . O termo τ_t pode assumir diversas formas e formas particulares são discutidas em vários artigos. Na sua ausência temos o modelo de regressão beta de [Ferrari and Cribari-Neto \(2004\)](#). Se τ_t satisfaz as equações de um ARFIMA, temos um modelo β ARFIMA ([Pumi et al., 2019](#)). Se τ_t satisfaz as equações de um ARMA, temos um modelo β ARMA(p, q) ([Rocha and Cribari-Neto, 2009](#)). Neste caso, a parte sistemática do modelo é especificada via

$$g(\mu_t) := \alpha + \mathbf{X}'_t \boldsymbol{\beta} + \sum_{i=1}^p \phi_i (g(Y_{t-i}) - \mathbf{X}'_{t-i} \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j r_{t-j}, \quad (2.10)$$

onde $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)'$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)'$ são coeficientes da estrutura ARMA, r_t denota um termo de erro, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_l)'$ é um vetor l -dimensional de parâmetros relacionados às covariáveis e $g : (0, 1) \rightarrow \mathbb{R}$ é uma função de ligação, bijetiva e duas vezes diferenciável. As funções de ligação mais comumente aplicadas são a logit, a probit e o log-log complementar. No caso da distribuição beta, a função de ligação usual é a função logit, definida como $g(\mu_t) = \log\left(\frac{\mu_t}{1-\mu_t}\right)$. O modelo β ARMA(p, q) é especificado por (2.7) e (2.10).

Estimação no contexto de β ARMA é tradicionalmente feita através do método de verossimilhança condicional ([Rocha and Cribari-Neto, 2009](#)) ou parcial ([Pumi et al., 2019](#)), contexto no qual a presente seção se baseia. Seja $\{y_t\}_{t=1}^n$ uma amostra de um processo β ARMA(p, q) e $\{\mathbf{x}_t'\}_{t=1}^n$ os valores observados das covariáveis associadas. Seja $\boldsymbol{\gamma}' = (\nu, \alpha, \boldsymbol{\beta}', \boldsymbol{\phi}', \boldsymbol{\theta}')$ $\in \Omega \subset \mathbb{R}^{l+p+q+2}$ o vetor de parâmetros e defina

$$\begin{aligned} \ell_t(\mu_t, \nu) &= \log(p(y_t | \mu_t, \nu, \mathcal{F}_{t-1})) \\ &= \log(\Gamma(\nu)) - \log(\Gamma(\mu_t \nu)) - \log(\Gamma(\nu(1-\mu_t))) + \\ &\quad + (\mu_t \nu - 1) \log(y_t) + (\nu(1-\mu_t) - 1) \log(1-y_t), \end{aligned}$$

de forma que a verossimilhança parcial é dada por

$$\ell(\boldsymbol{\gamma}) = \sum_{t=1}^n \ell_t(\mu_t, \nu), \quad (2.11)$$

e o estimador de máxima verossimilhança parcial de $\boldsymbol{\gamma}$ é dado por

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma} \in \Omega}{\text{argmax}} \{\ell(\boldsymbol{\gamma})\}. \quad (2.12)$$

CAPÍTULO 3

MÉTODOS COMPUTACIONAIS PARA INFERÊNCIA BAYESIANA

Neste capítulo, apresentaremos alguns métodos computacionais para inferência Bayesiana. Iniciaremos com uma breve introdução sobre a inferência Bayesiana. Em seguida, abordaremos o método de Monte Carlo via Cadeias de Markov (MCMC) e como funciona o seu algoritmo de amostragem para simulação de amostras da distribuição à posteriori. Por fim, apresentaremos o método de Monte Carlo Hamiltoniano e discutiremos em detalhes o funcionamento do seu algoritmo e como esse método pode ser usado para superar limitações do MCMC.

3.1 *Inferência Bayesiana*

Na inferência clássica, os parâmetros de um modelo, digamos γ , são considerados fixos e desconhecidos e o mecanismo probabilístico é determinado por $p(\mathbf{y}|\gamma)$, isto é, o foco é determinar a probabilidade de observar cada valor \mathbf{y} , sob diferentes valores de γ . Já a inferência Bayesiana é uma metodologia que utiliza variáveis aleatórias para representar os parâmetros desconhecidos dos modelos. Nela a inferência é baseada em $p(\gamma|\mathbf{y})$, chamada de distribuição a posteriori, que é a probabilidade do parâmetro condicional aos dados observados. Este é o fundamento de toda análise estatística baseada na inferência Bayesiana.

A posteriori é uma distribuição de probabilidades, obtida por meio da combinação da informação à priori sobre os parâmetros com a informação adquirida a partir dos dados amostrais (verossimilhança). É uma combinação de probabilidades e é obtida utilizando o Teorema de Bayes. Já a priori é uma distribuição de probabilidade que representa o conhecimento prévio sobre a distribuição dos parâmetros de um modelo antes de se considerar qualquer informação proveniente dos dados. Há vários tipos de priori, entre elas: não informativa (assume pouca ou nenhuma informação sobre o parâmetro), informativa (assume um grau maior de conhecimento acerca do parâmetro), própria (é integrável), imprópria (não é integrável) e conjugada (a priori e a posteriori vêm da mesma família de distribuições, mas os parâmetros normalmente são outros).

Seja $\mathbf{y} = (y_1, \dots, y_n)'$ uma amostra associado a um modelo com parâmetro $\gamma = (\gamma_1, \dots, \gamma_m)'$.

A distribuição conjunta à posteriori de γ dado \mathbf{y} é obtido utilizando-se o Teorema de Bayes, isto é,

$$p(\gamma|\mathbf{y}) = \frac{p(\gamma, \mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\gamma) \cdot p(\gamma)}{p(\mathbf{y})} \propto p(\mathbf{y}|\gamma)p(\gamma), \quad (3.1)$$

onde $p(\mathbf{y}|\gamma)$ é a função de verossimilhança, $p(\gamma)$ é a distribuição à priori de γ e $p(\mathbf{y})$ é uma constante normalizadora de $p(\gamma|\mathbf{y})$. Essa constante é importante para que a integral (no caso discreto, soma) da posteriori seja igual a 1. A equação (3.1) pode ser vista de outra forma

$$p(H|E) = \frac{p(E|H) \cdot p(H)}{p(E)}, \quad (3.2)$$

em que H denota a hipótese e E a evidência. Nota-se que $p(H)$ é a priori (probabilidade da hipótese), $p(E|H)$ é a verossimilhança (probabilidade da evidência dado a hipótese) e $p(H|E)$ é a posteriori (probabilidade da hipótese dado a evidência). Todavia, na prática, é frequentemente difícil ou impossível obter uma solução analítica para a distribuição à posteriori. Nesses casos, a solução pode ser obtida por meio de técnicas de simulação, sendo a simulação MCMC uma das abordagens mais utilizadas.

3.2 Monte Carlo via Cadeias de Markov

Nesta seção, introduziremos algumas definições e teoremas fundamentais sobre cadeias de Markov, tanto no caso discreto quanto no caso contínuo, e que são fundamentais para o entendimento do algoritmo Monte Carlo via Cadeias de Markov (MCMC). Em termos simples, uma cadeia de Markov é um processo estocástico em que dado o estado presente, o estado passado e o futuro são independentes (Gamerman and Lopes, 2006). De maneira formal, dizemos que $\{Y_t\}_{t \in T}$ é um processo de Markov com espaço de estados S se

$$\mathbb{P}(Y_t \in A | Y_{t_1} = y_1, Y_{t_2} = y_2, \dots, Y_{t_n} = y_n) = \mathbb{P}(Y_t \in A | Y_{t_n} = y_n), \quad (3.3)$$

para todo boreliano $A \subset S$, todos $t_1 < \dots < t_n < t \in T$ e todos $y_1, \dots, y_n \in S$.

Para melhor formalizar, definimos, no caso discreto, a função probabilidade de transição de n passos, como sendo

$$p_{ij}^{t,t+n} = \mathbb{P}(Y_{t+n} = j | Y_t = i), \quad (3.4)$$

para todos estados $i, j \in S$ e todos os tempos $t, n \geq 0$, a qual indica a probabilidade de Y_{t+n} estar no estado j dado que Y_t está no estado i . No caso em que o espaço de estados S é finito ou enumerável, mostra-se que as probabilidades de transição definidas em (3.4) definem, a menos da distribuição inicial, uma cadeia de Markov neste mesmo espaço de estados. Para aplicar com sucesso o método de MCMC, a cadeia de Markov gerada deve satisfazer certas condições, sejam elas:

- Homogeneidade: Uma cadeia de Markov é homogênea no tempo quando as probabilidades de transição definidas em (3.4) independem do tempo, isto é $p_{ij}^{t,t+n} = \mathbb{P}(Y_n = j | Y_0 = i)$ para todo $t \geq 0$. Neste caso, expressamos tais probabilidades de transição simplesmente por p_{ij}^n ;
- Aperiodicidade: O estado i é aperiódico se $p_{ii}^n > 0$ para todo n suficientemente grande;

- Irredutibilidade: Cada estado pode ser atingido a partir de qualquer outro em um número finito de iterações.

Caso as suposições estejam satisfeitas pode-se implementar o algoritmo de Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970), que examinaremos na próxima seção. Ele é muito utilizado em cenários onde a amostragem direta de valores da distribuição à posteriori não é possível.

No caso de distribuições contínuas, ao invés da probabilidade de transição p_{ij} , temos a densidade de transição $p(y, y')$, que faz a transição de $Y_n = y$ para $Y_{n+1} = y'$, onde $p(y, \cdot)$ é uma função de densidade. A densidade de transição pode ser definida como

$$p(y, y') = \begin{cases} q(y'|y)\alpha(y, y'), & \text{se } y \neq y'; \\ 1 - \int_{\mathcal{S}} q(y'|y)\alpha(y, y')dy', & \text{se } y = y', \end{cases} \quad (3.5)$$

em que $q(y'|y)$ é um densidade auxiliar, para propor y' de y , e $\alpha(y, y')$ é a probabilidade de aceitação que veremos mais adiante como calcular. A primeira expressão é a probabilidade de que a proposta y' seja aceita, e a segunda expressão é a probabilidade de que a proposta y' não seja aceita. Também temos no caso contínuo a densidade de transição de n passos, denotada por $p^{(n)}(y, y')$. Note que precisa satisfazer

$$\int_{\mathcal{S}} p(y, y')dy' = 1 \quad \text{e} \quad \int_{\mathcal{S}} p^{(n)}(y, y')dy' = 1, \quad (3.6)$$

para todo $n \geq 1$.

A distribuição estacionária é uma propriedade fundamental de uma cadeia de Markov, e no caso contínuo ela pode ser definida da seguinte forma: seja $\{Y_n\}_{n \geq 0}$ uma cadeia de Markov com espaço de estado contínuo \mathcal{S} com densidade de transição $p(y, y')$. Uma distribuição estacionária para $\{Y_n\}_{n \geq 0}$ em \mathcal{S} é a função de densidade de probabilidade $\pi(y)$ em \mathcal{S} que satisfaz

$$\pi(y') = \int_{\mathcal{S}} \pi(y)p(y, y')dy \quad \text{e} \quad \pi(y) = \int_{\mathcal{S}} \pi(y')p(y', y)dy'. \quad (3.7)$$

Quanto à reversibilidade, temos que uma cadeia de Markov em um espaço de estado contínuo \mathcal{S} com densidade de probabilidade de transição $p(y, y')$ é reversível em relação a uma densidade $\pi(y)$ se

$$\pi(y)p(y, y') = \pi(y')p(y', y), \quad (3.8)$$

para todo $y, y' \in \mathcal{S}$. Isso também é conhecido como condição do balanço detalhado.

3.2.1 Algoritmo Metropolis–Hastings

O algoritmo Metropolis-Hastings (MH), criado por Metropolis et al. (1953) e Hastings (1970), é um método de MCMC para gerar amostras de distribuições de probabilidade complexas. A ideia do algoritmo é propor novos valores para a cadeia de Markov com base em uma distribuição proposta e, em seguida, aceitar ou rejeitar essas propostas com base na probabilidade de aceitação. Ao repetir este processo, a cadeia de Markov explora o espaço da distribuição alvo e, após convergência gera

uma sequência de amostras correlacionadas da distribuição alvo que podem ser usadas para inferência ou simulação. Em [Gong and Flegal \(2016\)](#) é proposto um critério de parada para o algoritmo.

Suponha que o objetivo seja gerar n valores de uma distribuição conjunta $\pi(\boldsymbol{\gamma})$. Para isso, considere $q(\boldsymbol{\gamma}'|\boldsymbol{\gamma})$ uma densidade proposta (para propor $\boldsymbol{\gamma}'$ de $\boldsymbol{\gamma}$) e gere um valor inicial $\boldsymbol{\gamma}^{(0)}$ dessa distribuição. As etapas do algoritmo de Metropolis–Hastings são as seguintes:

- Passo 1. Inicialize o contador de iterações $i = 0$ e forneça um valor inicial $\boldsymbol{\gamma}^{(0)}$;
- Passo 2. Gere um novo valor $\boldsymbol{\gamma}'$ a partir da distribuição proposta $q(\cdot|\boldsymbol{\gamma})$;
- Passo 3. Gere $u \sim U(0, 1)$, e calcule a probabilidade de aceitação $\alpha(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$, dada por

$$\alpha(\boldsymbol{\gamma}, \boldsymbol{\gamma}') = \min \left\{ 1, \frac{\pi(\boldsymbol{\gamma}') q(\boldsymbol{\gamma}|\boldsymbol{\gamma}')}{\pi(\boldsymbol{\gamma}) q(\boldsymbol{\gamma}'|\boldsymbol{\gamma})} \right\} \quad (3.9)$$

em que $\pi(\boldsymbol{\gamma}) = p(\boldsymbol{\gamma}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})$ é a posteriori (alvo), $p(\mathbf{y}|\boldsymbol{\gamma})$ é a verossimilhança e $p(\boldsymbol{\gamma})$ é a priori;

- Passo 4. Se $u \leq \alpha$ então aceite o novo valor e defina

$$\boldsymbol{\gamma}^{(i+1)} = \boldsymbol{\gamma}'$$

senão, rejeite e defina

$$\boldsymbol{\gamma}^{(i+1)} = \boldsymbol{\gamma}^i$$

- Passo 5. Se $i < n$, faça $i = i + 1$ e repita os passos 2–5. Caso contrário, o algoritmo está finalizado.

3.3 Monte Carlo Hamiltoniano

Monte Carlo Hamiltoniano é uma técnica baseada na dinâmica Hamiltoniana concomitante a regra de transição de Metropolis-Hastings. É eficiente quanto a simulação de valores aleatórios de uma distribuição alvo, principalmente quando esta tem alta dimensão, visto que consegue percorrer rapidamente o suporte desta distribuição de interesse. Este algoritmo, também denominado por HMC, foi originalmente proposto por [Duane et al. \(1987\)](#), tendo sido intitulado por Monte Carlo Híbrido. Mais tarde, [Mackay \(2003\)](#) e [Neal \(2011\)](#) provaram a sua convergência para alguma distribuição alvo de interesse.

A ideia da dinâmica Hamiltoniana é transformar o espaço da densidade de interesse em uma superfície de energia potencial e após isso imaginar uma partícula fictícia se movendo sobre essa superfície sem atrito (Figura 3.1). Sabe-se que ao descer após o lançamento, a energia gravitacional potencial é convertida em energia cinética até o ponto mais baixo da rampa. À medida que a partícula sobe do outro lado da rampa, a energia cinética é convertida em energia potencial gravitacional. A partícula tende visitar com mais frequência regiões com baixa velocidade (menor energia cinética), que corresponde a regiões de densidade de menor probabilidade. O objetivo do algoritmo HMC é

seguir a trajetória dessa partícula e os eventuais valores que são traçados e usá-los como amostras da distribuição desejada (posteriori).

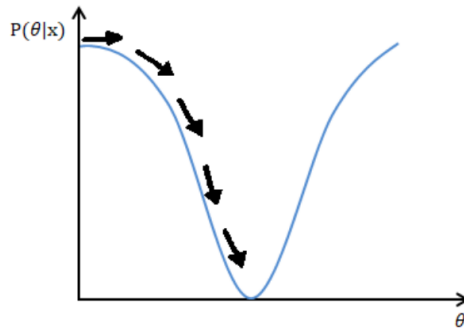


Figura 3.1: Figura Extraída de: [Reis \(2021\)](#)

Essas trajetórias são calculadas a partir das equações de Hamilton, utilizando o Hamiltoniano

$$H(\boldsymbol{\gamma}, \mathbf{p}) = U(\boldsymbol{\gamma}) + K(\mathbf{s}), \quad (3.10)$$

em que $U(\boldsymbol{\gamma}) = -\log(p(\boldsymbol{\gamma}))$ é a energia potencial, sendo $\boldsymbol{\gamma}$ o parâmetro de posição. Já $K(\mathbf{s})$ é a energia cinética, reescrita como $\frac{1}{2}\mathbf{s}'\mathbf{M}^{-1}\mathbf{s}$, em que \mathbf{s} é um parâmetro de momento e \mathbf{M} é a matriz de covariâncias de \mathbf{s} . Para determinar como $\boldsymbol{\gamma}$ e \mathbf{s} se comportam ao longo do tempo considera-se a dinâmica Hamiltoniana dada pelo sistema de equações diferenciais

$$\frac{\partial \mathbf{s}}{\partial t} = -\frac{\partial H(\boldsymbol{\gamma}, \mathbf{s})}{\partial \boldsymbol{\gamma}} = \nabla_{\boldsymbol{\gamma}} U(\boldsymbol{\gamma}), \quad (3.11)$$

$$\frac{\partial \boldsymbol{\gamma}}{\partial t} = \frac{\partial H(\boldsymbol{\gamma}, \mathbf{s})}{\partial \mathbf{s}} = \nabla_{\mathbf{s}} K(\mathbf{s}), \quad (3.12)$$

onde ∇_y denota o vetor escore para y , a Equação (3.11) indica a posição e a Equação (3.12) indica a velocidade de uma partícula. A ideia é que ao fixar os valores no primeiro tempo (valores iniciais), $(\boldsymbol{\gamma}(t_0), \mathbf{s}(t_0))$, a solução destas equações irá nos fornecer a posição e a velocidade da partícula no tempo seguinte $(\boldsymbol{\gamma}(t_1), \mathbf{s}(t_1))$. Essas derivadas são obtidas de forma automática através do pacote `RStan`, já mencionado anteriormente. A dinâmica Hamiltoniana possui algumas propriedades que garantem a convergência da cadeia: reversibilidade (importante para manter a estacionariedade da cadeia para cada atualização MCMC) e invariância (conservação da energia).

Todavia, é difícil obter uma solução analítica para as equações Hamiltonianas (3.11) e (3.12), logo é quase impossível simular a trajetória verdadeira de cada variável. Uma solução é utilizar o método numérico Störmer-Verlet (Leapfrog), que aproxima as equações por meio da discretização do tempo, usando um tamanho de passo ε ([Ruth, 1983](#)). A ideia do algoritmo é que ao começar com o estado no tempo t , pode-se iterativamente calcular os estados nos tempos $t + \varepsilon, t + 2\varepsilon, \dots, t + L\varepsilon$, onde L é o número de Leapfrogs (número de passos escolhidos), que representa o número de vezes para

resolver as equações de Störmer-Verlet, dada por

$$\mathbf{s}\left(t + \frac{\varepsilon}{2}\right) = \mathbf{s}(t) + \frac{\varepsilon}{2} \nabla_{\boldsymbol{\gamma}} \log(p(\boldsymbol{\gamma}(t))); \quad (3.13)$$

$$\boldsymbol{\gamma}(t + \varepsilon) = \boldsymbol{\gamma}(t) + \varepsilon \mathbf{M}^{-1} \mathbf{s}\left(t + \frac{\varepsilon}{2}\right); \quad (3.14)$$

$$\mathbf{s}(t + \varepsilon) = \mathbf{s}\left(t + \frac{\varepsilon}{2}\right) + \frac{\varepsilon}{2} \nabla_{\boldsymbol{\gamma}} \log(p(\boldsymbol{\gamma}(t + \varepsilon))), \quad (3.15)$$

em cada iteração. O processo de resolução começa em meio passo ($\varepsilon/2$) do momento \mathbf{s} em (3.13), em seguida, é realizado um passo completo para $\boldsymbol{\gamma}$ em (3.14), utilizando o valor de \mathbf{s} encontrado em (3.13). A mesma lógica segue depois em (3.15). Ao repetir esse processo iterativamente, consegue-se aproximar a trajetória de interesse.

Depois de entender a dinâmica Hamiltoniana e como ela se encaixa em casos discretos, podemos falar do algoritmo Monte Carlo Hamiltoniano. Ele apresenta alguns parâmetros de ajuste, como o tamanho do passo ε e o número L de passos escolhidos (Leapfrog), já vistos anteriormente. Sabe-se que com um ε pequeno, teremos uma boa amostragem, porém o custo computacional será muito grande. Já se ε for grande, o algoritmo irá rejeitar muitos pontos, e conseqüentemente perderemos eficiência computacional. Quanto ao L (comprimento da trajetória), se ele for muito curto, o algoritmo HMC se assemelha à um passeio aleatório ineficiente. Já se o comprimento for longo, o custo computacional será muito grande e o número de rejeições será muito alto. Uma forma automática de calibrar a escolha desses hiperparâmetros é através do algoritmo No U-Turn Sampler (NUTS), criado por Hoffman and Gelman (2014), que visa ajustar o tamanho de passos a cada iteração de amostragem durante o período de burn-in. Ele também auxilia na escolha da matriz de covariâncias \mathbf{M} , que é estimada no período de warm-up do algoritmo e geralmente é inicializada como a matriz identidade.

3.3.1 Algoritmo Monte Carlo Hamiltoniano

As etapas do algoritmo HMC seguem abaixo e vamos utilizá-lo para estimar os parâmetros $\boldsymbol{\gamma} = (\nu, \alpha, \beta_1, \dots, \beta_l, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$ do modelo β ARMA, mencionado anteriormente.

- Passo 1. Forneça um valor inicial para $\boldsymbol{\gamma}^{(0)}$ (especificado pelo usuário ou gerado automaticamente);
- Passo 2. Calcule $\log(p(\boldsymbol{\gamma}^{(0)}))$, onde $p(\cdot)$ é a distribuição alvo;
- Passo 3. Defina o número de amostras a serem geradas (N) e inicie o contador $i = 1, \dots, N$;
- Passo 4. Gere um número aleatório para o momento $\mathbf{s}^* \sim N_d(\mathbf{0}, \mathbf{M})$, onde \mathbf{M} é a matriz de covariância, e gere $u \sim U(0, 1)$;
- Passo 5. Faça $(\boldsymbol{\gamma}^I, \mathbf{s}^I) = (\boldsymbol{\gamma}^{(i-1)}, \mathbf{s}^*)$ e $H_0 = H(\boldsymbol{\gamma}^I, \mathbf{s}^I)$, ou seja, $H_0 = -\log(p(\boldsymbol{\gamma}^{(i-1)})) + \frac{1}{2} \mathbf{s}^{*'} \mathbf{M}^{-1} \mathbf{s}^*$;

- Passo 6. Execute o algoritmo de Leapfrog para atualizar os valores de γ e s , L vezes (onde L é o número de passos), e utilizando um tamanho de passo ε ;

$$\begin{aligned} \mathbf{s}^* &= \mathbf{s}^* + \frac{\varepsilon}{2} \nabla_{\gamma} \log(p(\gamma^{(i-1)})); \\ \gamma^{(i-1)} &= \gamma^{(i-1)} + \varepsilon \mathbf{M}^{-1} \mathbf{s}^*; \\ \mathbf{s}^* &= \mathbf{s}^* + \frac{\varepsilon}{2} \nabla_{\gamma} \log(p(\gamma^{(i-1)})). \end{aligned}$$

O algoritmo NUTS auxilia na escolha do número de passos L e no tamanho do passo ε ;

- Passo 7. Ao final da trajetória faça $(\gamma^D, \mathbf{s}^D) = (\gamma^{(i-1)}, \mathbf{s}^*)$ e $H_1 = H(\gamma^D, \mathbf{s}^D)$;
- Passo 8. Determine a taxa de aceitação: $\alpha[(\gamma^D, \mathbf{s}^D), (\gamma^I, \mathbf{s}^I)] = \min\{\exp(H_0 - H_1), 1\}$;
- Passo 9. Se $u < \alpha[(\gamma^D, \mathbf{s}^D), (\gamma^I, \mathbf{s}^I)]$ então

$$\gamma^{(i)} = \gamma^D$$

senão

$$\gamma^{(i)} = \gamma^I;$$

- Passo 10. Faça $i = i + 1$ e repita os passos 4–10.

O método HMC é um caso particular do MH, e ao comparar ambos algoritmos temos que o HMC apresenta uma técnica de amostragem mais elaborada e um maior custo computacional. No MH cada nova amostra é gerada a partir da amostra anterior, o que pode resultar em amostras altamente correlacionadas, enquanto que o HMC produz novas amostras simulando a dinâmica de um sistema Hamiltoniano, o que tende a gerar amostras menos correlacionadas e, portanto, obtendo um tamanho de amostra efetivo maior. Embora o HMC possa produzir amostras mais eficientes (menos correlacionadas), o gasto de recursos computacionais para gerar cada amostra é significativamente maior do que o do MH. Isso ocorre porque o HMC exige a simulação da dinâmica Hamiltoniana, a qual inclui uma etapa de integração numérica, elevando assim o custo computacional.

CAPÍTULO 4

CONSIDERAÇÕES FINAIS

Neste trabalho, exploramos a modelagem β ARMA sob o enfoque Bayesiano, fornecendo uma abordagem alternativa para a estimação dos parâmetros do modelo. Os resultados e detalhes encontram-se no artigo Grande et al. (2023) apresentado no Apêndice A

Em relação aos resultados encontrados no artigo, conclui-se que a abordagem Bayesiana proposta para estimar modelos β ARMA mostrou-se eficaz e robusta em diferentes cenários de simulação. O método demonstrou estimativa precisa de ponto e intervalo, detecção confiável de raízes quase-unitárias e robustez em relação às especificações a priori. Além disso, a aplicação dos modelos β ARMA à proporção média mensal de energia hidrelétrica armazenada no Sul do Brasil mostrou que a abordagem Bayesiana superou o modelo frequentista nos primeiros passos em termos de previsões fora da amostra.

Os resultados deste trabalho sugerem que a abordagem Bayesiana proposta tem potencial para avançar no campo da análise de séries temporais e contribuir para uma modelagem mais precisa de dados do mundo real. Como trabalho futuro, propõe-se a aplicação de métodos de seleção de modelos como Reversible Jump e Horseshoe, bem como o desenvolvimento de um pacote ou um script para compartilhar o método computacional implementado, tornando a abordagem mais acessível para pesquisadores e profissionais.

REFERÊNCIAS BIBLIOGRÁFICAS

- Amisano, G., Casarin, R., et al., 2007. Particle filters for Markov switching stochastic correlation models, in: Proceedings of the 2007 Intermediate Conference “Risk and Prediction”, Cleup, Padua, Citeseer. pp. 305–316.
- 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.
- Benjamin, M.A., Rigby, R.A., Stasinopoulos, D.M., 2003. Generalized autoregressive moving average models. *Journal of the American Statistical Association* 98, 214–223.
- Billio, M., Casarin, R., 2010. Identifying business cycle turning points with sequential monte carlo methods: an online and real-time application to the euro area. *Journal of Forecasting* 29, 145–167.
- Billio, M., Casarin, R., 2011. Beta autoregressive transition Markov-switching models for business cycle analysis. *Studies in Nonlinear Dynamics & Econometrics* 15.
- Brockwell, P.J., Davis, R.A., 1991. Time series: theory and methods. Springer Science & Business Media.
- Bueno, R.L.S., 2008. *Econometria de Séries Temporais*. Cengage Learning.
- Casarin, R., Valle, L.D., Leisen, F., 2012. Bayesian model selection for beta autoregressive processes. *Bayesian Analysis* 7, 385–410.
- 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.
- Duane, S., Kennedy, A., Pendleton, B.J., Roweth, D., 1987. Hybrid Monte Carlo. *Physics Letters B* 195, 216–222.
- Ferrari, S.L.P., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31, 799–815.
- Franco, G.C., Migon, H.S., Prates, M.O., 2019. Time series of count data. *Brazilian Journal of Probability and Statistics* 33, 756–781.
- Gamerman, D., Cepeda-Cuervo, E., 2013. Generalized spatial dispersion models. Universidade Federal do Rio de Janeiro .

- Gamerman, D., Lopes, H.F., 2006. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. CRC Press.
- Gong, L., Flegal, J.M., 2016. A practical sequential stopping rule for high-dimensional markov chain monte carlo. *Journal of Computational and Graphical Statistics* 25, 684–700.
- Green, P.J., 1995. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika* 82, 711–732.
- Guolo, A., Varin, C., 2014. Beta regression for time series analysis of bounded data, with application to Canada Google Flu Trends. *The Annals of Applied Statistics* 8, 74–88.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.
- Hoffman, M.D., Gelman, A., 2014. The No-U-Turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research* 15, 1593–1623.
- Kieschnick, R., McCullough, B.D., 2003. Regression analysis of variates observed on $(0, 1)$: percentages, proportions and fractions. *Statistical Modelling* 3, 193–213.
- Kumaraswamy, P., 1980. A generalized probability density function for double-bounded random processes. *Journal of Hydrology* 46, 79–88.
- Mackay, D.J., 2003. Information theory, inference and learning algorithms. Cambridge University Press.
- McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman & Hall / CRC.
- McKenzie, E., 1985. An autoregressive process for beta random variables. *Management Science* 31, 988–997.
- Melo, T.F., Vasconcellos, K.L., Lemonte, A.J., 2009. Some restriction tests in a new class of regression models for proportions. *Computational statistics & data analysis* 53, 3972–3979.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21, 1087–1092.
- Neal, R.M., 2011. Handbook of Markov chain Monte Carlo. Chapman and Hall CRC Press.
- Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)* 135, 370–384.
- Pereira, L.A., Tavares, M., 2011. Regressão beta para modelagem do rendimento metalúrgico na reciclagem de alumínio. *Anais da XII Escola de Modelos de Regressão*.
- Prass, T.S., Pumi, G., 2022. BTSR: Bounded Time Series Regression. R package version 0.1.1.
- 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, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.

- Reis, D.J.F., 2021. Análise Crítica do Monte Carlo Hamiltoniano na Avaliação de Afundamentos de Tensão. Master's thesis. Universidade Federal do Maranhão.
- Rocha, A.V., Cribari-Neto, F., 2009. Beta autoregressive moving average models. *TEST* 18, 529–545.
- Rogers, J.A., Polhamus, D., Gillespie, W.R., Ito, K., Romero, K., Qiu, R., Stephenson, D., Gastonguay, M.R., Corrigan, B., 2012. Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: a beta regression meta-analysis. *Journal of Pharmacokinetics and Pharmacodynamics* 39, 479 – 498.
- Ruth, R.D., 1983. A canonical integration technique. *IEEE Transactions on Nuclear Science* 30, 2669–2671.
- 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.
- Seifi, A., Ponnambalam, K., Vlach, J., 2000. Maximization of manufacturing yield of systems with arbitrary distributions of component values. *Annals of Operations Research* 99, 373–383.
- Stan Development Team, 2022. RStan: the R interface to Stan. URL: <https://mc-stan.org/>. r package version 2.21.5.
- Taddy, M.A., 2010. Autoregressive mixture models for dynamic spatial poisson processes: Application to tracking intensity of violent crime. *Journal of the American Statistical Association* 105, 1403–1417.
- Wang, X.F., 2012. Joint generalized models for multidimensional outcomes: A case study of neuroscience data from multimodalities. *Biometrical Journal* 54, 264–280.
- Zheng, T., Xiao, H., Chen, R., 2015. Generalized ARMA models with martingale difference errors. *Journal of Econometrics* 189, 492–506.
- Zou, K.H., Carlsson, M.O., Quinn, S.A., 2010. Beta-mapping and beta-regression for changes of ordinal-rating measurements on Likert scales: A comparison of the change scores among multiple treatment groups. *Statistics in Medicine* 29, 2486–2500.

APÊNDICE A

ARTIGO GRANDE ET AL. (2023)

Autores: Aline Foerster Grande, Guilherme Pumi e Gabriela Bettella Cybis

Título: Bayesian analysis of Beta Autoregressive Moving Average Models

Revista: a definir

Ano: 2023

Bayesian Analysis of Beta Autoregressive Moving Average Models

Aline Foerster Grande^a, Guilherme Pumi^{a,*} and Gabriela Bettella Cybis^a

Abstract

This work presents a Bayesian approach for the estimation of Beta Autoregressive Moving Average (β ARMA) models. We discuss standard choice for the prior distributions, and employ a Hamiltonian Monte Carlo algorithm to sample from the posterior. We propose a method to approach the problem of unit roots in the model's systematic component. We present a series of Monte Carlo simulations to evaluate the performance of the presented Bayesian approach. In addition to parameter estimation, we evaluate the proposed approach to verify the presence of unit roots in the model's systematic component and study prior sensitivity. An empirical application is presented to exemplify the usefulness of the method. In the application we compare the fitted Bayesian and frequentist approach in terms of their out-of-sample forecasting capabilities.

Keywords: dynamic models; time series analysis; generalized linear models, Bayesian models, unit roots.

1 Introduction

The literature related to regression models for time series has grown rapidly in the last few years. A particularly fertile area of research involves the theory of bounded time series regression. Typical examples are ratios and proportions observed over time. In these situations, it is well-known that ARMA models are not adequate. One of the most applied methods to deal with bounded time series is based on the GARMA (Generalized ARMA) approach, introduced under this name by Benjamin et al. (2003). The idea is to combine the flexibility of the generalized linear model (GLM) structure with the classical ARMA modeling approach. The GLM structure addresses the problem of bounds, still allowing for the presence of covariates, used to accommodate different structures, such as heteroscedasticity, trends, etc.; whereas the time series structure accommodates the presence of serial correlation in the process. Although simple, the approach is quite general.

After Benjamin et al. (2003), the GARMA structure has evolved in various forms, by assuming distributions outside the exponential family (Rocha and Cribari-Neto, 2009; Bayer et al., 2017, 2020; Prass et al., 2022; Pumi et al., 2022), including long-range dependence (Pumi et al., 2019) and even chaotic dependence structure (Pumi et al., 2021). Applications of such models are varied, covering several areas such as medicine (Zou et al., 2010; Rogers et al., 2012), online monitoring (Guolo and Varin, 2014), metallurgical engineering (Pereira and Tavares, 2011), neuroscience (Wang, 2012), operational research (Seifi et al., 2000), among many others.

Inference in the context of β ARMA models is based on conditional (Rocha and Cribari-Neto, 2009) or partial (Pumi et al., 2019) maximum likelihood, with a currently available package in R (R Core Team, 2022) implementing these methods, namely, the `BTSR` (Prass and Pumi, 2022). Small sample corrections for these methods are also available in the literature

^aPrograma de Pós Graduação em Estatística - Universidade Federal do Rio Grande do Sul.

*Corresponding author. This Version: May 18, 2023

E-mails: alinefoerstergrande@gmail.com (A.F. Grande); guilherme.pumi@ufrgs.br (G. Pumi); gcybis@gmail.com (G.B. Cybis)

(Palm and Bayer, 2018). The situation is different when we consider Bayesian estimation of β ARMA models. To the best of our knowledge, the only paper available is Casarin et al. (2012), where the authors consider a Bayesian approach to model selection in β AR models, without covariates, using a reversible jump Markov chain approach. The framework considered in the paper, however, is quite restrictive because the authors propose to use a linear link in the model's systematic component, requiring the imposition of conditions on the model's coefficients to keep the conditional mean within the interval $(0, 1)$. In addition, because it is non-standard, there are no statistical packages available to fit such a particular model, which hinders its application.

In this paper, we propose the study of a fully specified β ARMA model under the Bayesian framework. The model considers standard forms for the prior distributions, for which we make uninformative parameter choices. We employ the Hamiltonian Markov Chain Monte Carlo (MCMC) method to obtain the posterior distribution of the model parameters, and the No-U-Turn sampler (NUTS) for Hamiltonian parameter fine-tuning. We also use the model to assess the posterior probability that the model's systematic component contains a quasi-unit root. Furthermore, we consider a procedure consisting of a series of Bayes Factor (BF) tests to determine the order p and q of the β ARMA(p, q) model, and address the issue of prediction by assessing the posterior predictive distribution. The quality of parameter estimates and sensibility to prior specifications is assessed through simulation studies. The proposed framework allows for a comprehensive fully Bayesian analysis of bounded time series data under the β ARMA(p, q) model, complete with parameter estimation, sensitivity analysis, identification of unit roots and prediction.

We also present a Monte Carlo simulation study to verify the performance of the proposed approach. We evaluate point and interval estimation under various scenarios in the simulation, approach the problem of unit roots in β ARMA's systematic component, and provide a prior sensitivity analysis. We showcase the proposed methodology in an empirical application to hydroelectric energy storage. We perform model selection based on the proposed Bayes factor approach, comparing the selected model's out-of-sample forecasting capabilities with competitor models, including a model selected using the frequentist approach.

The following is a summary of the paper. Section 2 comprehensively reviews the class of β ARMA models. Section 3 delves into key concepts associated with the Bayesian inference, such as the Hamiltonian Monte Carlo algorithm, posterior predictive distribution, and quasi-unit root testing. In Section 4, we carry out a Monte Carlo simulation to assess finite sample behavior of the proposed approach. Finally, Section 5 provides an illustrative example to showcase the practical application of the proposed method.

2 β ARMA Models

Let $\{Y_t\}_{t \in \mathbb{Z}}$ be a time series of interest satisfying $Y_t \in (0, 1)$ for all $t \in \mathbb{Z}$, and let $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ denote a set of r -dimensional exogenous time dependent (possibly random) covariates. Let \mathcal{F}_t denote the σ -field representing the observed history of the model up to time t , that is, the sigma-field generated by $(\mathbf{X}'_t, Y_t, \mathbf{X}'_{t-1}, Y_{t-1}, \dots)$. The β ARMA(p, q) model introduced by Rocha and Cribari-Neto (2009) is an observation-driven model in which the random component follows a conditional beta distribution, parameterized in terms of its mean and a precision parameter as

$$p(y|\nu, \mu_t, \mathcal{F}_{t-1}) = \frac{\Gamma(\nu)}{\Gamma(\nu\mu_t)\Gamma(\nu(1-\mu_t))} y^{\nu\mu_t-1}(1-y)^{\nu(1-\mu_t)-1}, \quad (1)$$

for $0 < y < 1$, $\mu_t := \mathbb{E}(Y_t | \mathcal{F}_{t-1})$ and $\nu > 0$. Observe that $\text{Var}(Y_t | \mathcal{F}_{t-1}) = \frac{\mu_t(1-\mu_t)}{1+\nu}$, so that the model is conditionally heteroscedastic. Additionally, ν can be interpreted as a precision parameter in the sense that, the higher the ν , the smaller the conditional variance. Moreover, $\text{Var}(Y_t | \mathcal{F}_{t-1}) \leq \frac{1}{4\nu} \rightarrow 0$ as ν goes to infinity. Hence, in practice, very high values of ν account for (conditional) homoscedastic behavior. Let $g : (0, 1) \rightarrow \mathbb{R}$ be a twice-differentiable strictly monotonic link function. The most commonly applied link functions are the logit (applied here), probit, and complementary log-log, although parametric alternatives have been explored in the literature (Pumi et al., 2020). The systematic component in a $\beta\text{ARMA}(p, q)$ model follows an $\text{ARMA}(p, q)$ -like structure prescribed by

$$g(\mu_t) := \alpha + \mathbf{X}'_t \boldsymbol{\beta} + \sum_{i=1}^p \phi_i (g(Y_{t-i}) - \mathbf{X}'_{t-i} \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j r_{t-j}, \quad (2)$$

where $r_t := g(Y_t) - g(\mu_t)$ denotes an error term, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)'$ is the r -dimensional vector of parameters related to the covariates, $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)'$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)'$ are the AR and MA coefficients, respectively. The $\beta\text{ARMA}(p, q)$ model is defined by (1) and (2).

Parameter estimation in βARMA models is usually conducted using a frequentist approach, based on the conditional (Rocha and Cribari-Neto, 2009, 2017) or partial (Pumi et al., 2019) maximum likelihood, which is reflected in the notation here. In the next section we propose a fully Bayesian approach for parameter estimation in the context of $\beta\text{ARMA}(p, q)$ models.

3 Bayesian inference

Let y_1, \dots, y_n be a sample from a $\beta\text{ARMA}(p, q)$ model, and $\boldsymbol{\gamma} := (\nu, \alpha, \boldsymbol{\beta}', \boldsymbol{\phi}', \boldsymbol{\theta}')'$. The partial likelihood function is given by

$$L(\boldsymbol{\gamma} | \mathcal{F}_n) \propto \prod_{t=1}^n p(y_t; \mu_t, \nu | \mathcal{F}_{t-1}) \propto \prod_{t=1}^n \frac{y_t^{\nu\mu_t-1} (1-y_t)^{\nu(1-\mu_t)-1}}{\Gamma(\nu\mu_t)\Gamma(\nu(1-\mu_t))}.$$

For the nonnegative precision parameter ν we consider a gamma prior of the form $\nu \sim \text{Gamma}(a, b)$ (mean a/b). The link function in (2) guarantees that $\mu_t \in (0, 1)$ for all $\alpha, \boldsymbol{\beta} \in \mathbb{R}^r, \boldsymbol{\theta} \in \mathbb{R}^q$ and $\boldsymbol{\phi} \in \mathbb{R}^p$, thus we shall assume independent multivariate normal priors of the form

$$\alpha \sim N(0, \sigma_\alpha^2), \quad \boldsymbol{\beta} \sim N_r(\mathbf{0}, \sigma_\beta^2 I_r), \quad \boldsymbol{\phi} \sim N_p(\mathbf{0}, \sigma_\phi^2 I_p), \quad \boldsymbol{\theta} \sim N_q(\mathbf{0}, \sigma_\theta^2 I_q), \quad (3)$$

where $N_s(\mathbf{0}, K)$ denotes the s -variate normal distribution with mean $\mathbf{0} \in \mathbb{R}^s$ and variance-covariance matrix K , while I_s denotes the $s \times s$ identity matrix.

When no prior knowledge regarding the hyper-parameters $\sigma_\alpha^2, \sigma_\beta^2, \sigma_\phi^2$ and σ_θ^2 is available, we shall consider very high values of these parameters so that the prior will be non-informative. As for the hyper-parameters related to ν , in section 4.3 we shall conduct a sensitivity analysis regarding their choice.

Combining the joint prior density, denoted by $\pi_0(\boldsymbol{\gamma})$, with the likelihood function, we note that the posterior distribution is given by

$$\pi(\boldsymbol{\gamma} | \mathcal{F}_n) \propto L(\boldsymbol{\gamma} | \mathcal{F}_n) \pi_0(\boldsymbol{\gamma}). \quad (4)$$

Unfortunately, the joint posterior density (4) cannot be obtained in closed form. To sample from the posterior distribution we adopt a Hamiltonian Monte Carlo scheme.

3.1 Hamiltonian Monte Carlo (HMC)

Hamiltonian Monte Carlo is a state-of-the-art general purpose sampling technique based on the Hamiltonian dynamics concomitant with a Metropolis-Hastings transition rule (Duane et al., 1987; Mackay, 2003; Neal, 2011). It is generally an efficient alternative for simulating random values from a target distribution, and requires the evaluation of the log-density and its gradient. In order to sample from the joint distribution of $\boldsymbol{\gamma}$, a momentum variable $\boldsymbol{\kappa} \sim N_d(\mathbf{0}, I_d)$, independent of $\boldsymbol{\gamma}$, is added to the dynamics and sampled along with $\boldsymbol{\gamma}$. In this case, the Hamiltonian is defined by

$$H(\boldsymbol{\kappa}, \boldsymbol{\gamma}) := \log(\pi(\boldsymbol{\gamma}|\mathcal{F}_n)) + \frac{1}{2}\boldsymbol{\kappa}'\boldsymbol{\kappa},$$

and the sample space of $(\boldsymbol{\kappa}, \boldsymbol{\gamma})$ is explored according to a Hamiltonian dynamics, namely,

$$\frac{\partial H(\boldsymbol{\kappa}, \boldsymbol{\gamma})}{\partial \boldsymbol{\kappa}} = \boldsymbol{\kappa}' \quad \text{and} \quad \frac{\partial H(\boldsymbol{\kappa}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \Delta_{\boldsymbol{\gamma}} \log(\pi(\boldsymbol{\gamma}|\mathcal{F}_n)). \quad (5)$$

To approximate a solution of (5) and obtain an iteration of the HMC, the first step is to sample $\boldsymbol{\kappa}$ from a $N_d(\mathbf{0}, I_d)$. Next, dynamics (5) is approximated from time $t = 0$ to time $t = \tau$ by applying $\lfloor \tau/\varepsilon \rfloor$ Leapfrog steps of size ε , namely,

$$\boldsymbol{\kappa} \leftarrow \boldsymbol{\kappa} + \frac{\varepsilon}{2} \Delta_{\boldsymbol{\gamma}} \log(\pi(\boldsymbol{\gamma}|\mathcal{F}_n)), \quad \boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma} + \boldsymbol{\kappa} \varepsilon, \quad \boldsymbol{\kappa} \leftarrow \boldsymbol{\kappa} + \frac{\varepsilon}{2} \Delta_{\boldsymbol{\gamma}} \log(\pi(\boldsymbol{\gamma}|\mathcal{F}_n)). \quad (6)$$

It can be shown that (6) yields a valid Metropolis proposal which is accepted or rejected using standard acceptance probabilities. Finely tuned values of ε and τ are crucial for HMC's performance. An automatic way to fine-tune the choice of these hyper-parameters is by using the No U-Turn Sampler (NUTS) algorithm, created by Hoffman and Gelman (2014). We use package `RStan` (Stan Development Team, 2022) in R (R Core Team, 2022) to perform the HMC.

3.2 Posterior predictive distribution

Let $\tilde{\mathbf{Y}} = (Y_{n+1}, \dots, Y_{n+h})'$ represent the new observations for which we wish to make predictions. Then the posterior predictive distribution satisfies

$$p(\tilde{\mathbf{Y}}|\mathcal{F}_n) = \int_{\boldsymbol{\gamma}} p(\tilde{\mathbf{Y}}|\boldsymbol{\gamma}, \mathcal{F}_n) p(\boldsymbol{\gamma}|\mathcal{F}_n) d\boldsymbol{\gamma} = \mathbb{E}_{p(\boldsymbol{\gamma}|\mathcal{F}_n)}(p(\tilde{\mathbf{Y}}|\boldsymbol{\gamma}, \mathcal{F}_n)), \quad (7)$$

where $p(\boldsymbol{\gamma}|\mathcal{F}_n)$ is the posterior distribution of (4) and $p(\tilde{\mathbf{Y}}|\boldsymbol{\gamma}, \mathcal{F}_n)$ is the sampling density of the new data $\tilde{\mathbf{Y}}$, given parameters $\boldsymbol{\gamma}$ and \mathcal{F}_n . It can be computed iteratively by noting that

$$p(\tilde{\mathbf{Y}}|\boldsymbol{\gamma}, \mathcal{F}_n) = \prod_{k=1}^h p(Y_{n+k}|\boldsymbol{\gamma}, \mathcal{F}_{n+k-1}), \quad (8)$$

where

$$Y_{n+h}|\boldsymbol{\gamma}, \mathcal{F}_{n+h-1} \sim \text{Beta}(\mu_{n+h}, \nu), \quad (9)$$

and

$$\mu_t = g^{-1} \left(\alpha + \mathbf{x}'_t \boldsymbol{\beta} + \sum_{i=1}^p \phi_i(g(y_{t-i}) - \mathbf{x}'_{t-i} \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j r_{t-j} \right), \quad (10)$$

in which $r_t := (g(Y_t) - g(\mu_t))$ is computed from μ_t and denotes an error term. It follows from (7) that we can obtain samples from the posterior predictive distribution, if for each value of γ^* in a sample of the posterior distribution $p(\gamma|\mathcal{F}_n)$ we draw from $p(\tilde{Y}|\gamma, \mathcal{F}_n)$. This is done by recursively computing μ_t given previous data and γ^* , and then drawing a new value for Y_t^* from the beta distribution in (10), for $t \in \{n+1, \dots, n+h\}$.

3.3 Quasi-Unit roots testing

Recall the following conditions, endemic to the study of ARMA models:

1. The AR and MA characteristic polynomials, given respectively by

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \quad \text{and} \quad \theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q, \quad z \in \mathbb{C}, \quad (11)$$

do not have common roots.

2. The AR characteristic polynomial satisfy $\phi(z) \neq 0$ in the set $\mathcal{C}^* := \{z \in \mathbb{C} : |z| \neq 1\}$.

Condition 2 is a necessary and sufficient condition for the existence of a stationary solution for the ARMA equations and much attention is given to the presence of unit roots, that is, roots of the characteristic polynomial with absolute value equal to 1, which are usually associated with a random walk-like behavior. However, model coefficients are estimated from a time series and are subject to uncertainty implying that exact unit roots occur with probability 0. Hence, in practice we are concerned with the so-called quasi-unit roots, which are roots of the characteristic polynomial with modulus close to 1. Quasi-unit roots also appear when a short range dependence model as the ARMA is applied to data presenting long-range dependence.

In the frequentist framework, unit roots are approached by using traditional tests, such as the Dickey-Fuller (Dickey and Fuller, 1981), Phillips-Perron (Phillips and Perron, 1988), among others. However, these tests are notorious for failing to detect unit roots in many contexts, especially in more complex dependence structure. Hence, a good practice in the frequentist approach is to calculate the roots of the estimated characteristic polynomial obtained by plugging-in the estimated parameters, as a goodness-of-fit procedure.

Technically, since observation driven models are defined in a conditional fashion, there is no imposition of stationarity in the model. Nevertheless, since $g(\mu_t) = g(y_t) - r_t$, under 1 and 2 the systematic component of a β ARMA model, namely (2), can be written as

$$\phi(L)(g(Y_t) - \mathbf{x}'_t \boldsymbol{\beta} - \omega) = \theta(L)r_t, \quad (12)$$

where $\omega := \alpha/\phi(1)$, so that it satisfies the difference equations of an ARMA model with characteristic polynomials given in (11). Hence, it is of interest to study unit roots in this context. However, testing a double bounded time series for unit roots is not even a well-posed problem since random walk behavior in a bounded domain is not well defined. Besides, testing the original time series for unit roots is not the same as testing $g(\mu_t)$, since the latter depends on model parameters to be reconstructed. Hence, traditional unit root tests are useless in the context of β ARMA models. The only alternative under the frequentist framework is calculating the root of the approximated characteristic polynomial. But, there is no easy way to measure the uncertainty behind a given root of the approximated characteristic polynomial and no test to decide if a given quasi-unit root is or is not a consequence of a unit root in the underlying true model. The decision in this context ultimately falls under the user's experience.

Since the proposed Bayesian approach relies on obtaining samples for the joint posterior probability of the parameters, we can easily estimate the posterior probability that a unit root is present in the fitted model. We simply calculate the roots of the approximated characteristic polynomial for each sampled set of parameters in the posterior distribution, and consider the proportion of roots whose modulus falls below a certain threshold.

4 Monte Carlo Simulation Studies

In this section we present a Monte Carlo simulation study to assess the performance of the proposed approach in the context of β ARMA models. We perform three sets of simulations under a variety of scenarios to assess three different facets of the Bayesian analysis. In the first set, our interest lies in the accuracy of point estimation and the coverage of credible intervals. The second set analyzes the presence of quasi-unit roots in the model. Finally, we perform a sensitivity analysis on the prior parameter specification.

Simulation was performed using R (R Core Team, 2022). Data generation was performed using R package BTR (Prass and Pumi, 2022). A burn-in of 50 observations was applied in generating the time series, regardless the sample size and parameters. Estimation was performed using the RStan package (Stan Development Team, 2022). All data and codes are available at github.com/AlineFoersterGrande/Bayesian.BARMA.

4.1 Point and interval inference

In this section we examine point and interval estimation of the proposed Bayesian approach in the context of β ARMA(1,1) models. We simulate β ARMA(1,1) model with $(\phi, \theta) \in \{(0.4, 0.4), (0.4, 0.6), (0.6, 0.4), (0.6, 0.6), (-0.5, -0.3), (-0.5, -0.4), (-0.3, -0.3), (-0.3, -0.4)\}$, $\nu \in \{50, 100\}$, and sample sizes $n \in \{200, 500\}$. We fit a β ARMA(1,1) considering the following priors: $\nu \sim \text{Gamma}(5, 0.1)$, $\phi \sim N(\mu, \sigma^2)$ and $\theta \sim N(\mu, \sigma^2)$, where $\mu = 0$ and $\sigma^2 = 20,000^2$. For each simulated data set, we ran two MCMC chains consisting of 2,000 samples each with a warm up period of 50%. Each experiment was replicated 50 times.

The simulation results are shown in Table 1. For each parameter, the point estimates (average posterior mean) and average credible intervals (quantiles) are shown. From Table 1, we observe that the proposed Bayesian approach performs very well in all scenarios, with small biases for ϕ and θ even when $n = 200$. It is worth noting that the specific values of ϕ and θ do not appear to have a significant effect on the credible intervals or point estimates for ν . Although ν has the highest variability, it has a smaller overall ratio between the credible interval length and the true parameter value.

Considering the point estimation for sample size $n = 500$, the best scenario was $\nu = 50$, $\phi = -0.3$ and $\theta = -0.3$, while the worst scenario was $\nu = 100$, $\phi = 0.4$ and $\theta = 0.6$. As expected, credible intervals in the $n = 200$ scenario are considerably larger than those in the $n = 500$ scenario. From the results we conclude that the posterior estimates are close to the simulated values, which always fall within the mean credible interval.

Table 1: Monte Carlo simulation results for β ARMA(1,1) models. Presented are point and interval estimates obtained as the average posterior mean and average 95% credible intervals.

DGP			$n = 200$			$n = 500$		
ν	ϕ	θ	ν $CI_{0.95}(\nu)$	ϕ $CI_{0.95}(\phi)$	θ $CI_{0.95}(\theta)$	ν $CI_{0.95}(\nu)$	ϕ $CI_{0.95}(\phi)$	θ $CI_{0.95}(\theta)$
50	0.40	0.40	48.73 [39.98, 58.38]	0.41 [0.23, 0.58]	0.37 [0.18, 0.54]	49.57 [43.75, 55.79]	0.40 [0.28, 0.51]	0.40 [0.28, 0.51]
	0.40	0.60	47.91 [39.20, 57.47]	0.42 [0.26, 0.57]	0.55 [0.40, 0.68]	49.18 [43.35, 55.33]	0.40 [0.30, 0.50]	0.59 [0.50, 0.67]
	0.60	0.40	48.28 [39.52, 57.80]	0.60 [0.46, 0.73]	0.38 [0.22, 0.53]	49.27 [43.43, 55.44]	0.60 [0.51, 0.68]	0.40 [0.30, 0.49]
	0.60	0.60	47.55 [38.98, 57.03]	0.60 [0.47, 0.72]	0.57 [0.43, 0.69]	49.01 [43.25, 55.15]	0.60 [0.52, 0.68]	0.59 [0.51, 0.67]
	-0.50	-0.30	48.84 [40.08, 58.56]	-0.50 [-0.66, -0.32]	-0.29 [-0.46, -0.09]	49.57 [43.77, 55.75]	-0.49 [-0.60, -0.39]	-0.30 [-0.41, -0.18]
	-0.50	-0.40	47.98 [38.98, 58.03]	-0.50 [-0.65, -0.34]	-0.37 [-0.53, -0.18]	49.40 [43.60, 55.51]	-0.49 [-0.59, -0.39]	-0.40 [-0.50, -0.29]
	-0.30	-0.30	49.12 [40.18, 58.89]	-0.31 [-0.53, -0.08]	-0.27 [-0.49, -0.03]	49.77 [43.94, 55.96]	-0.30 [-0.44, -0.15]	-0.30 [-0.44, -0.14]
	-0.30	-0.40	49.03 [40.12, 58.75]	-0.32 [-0.52, -0.11]	-0.36 [-0.55, -0.16]	49.72 [43.90, 55.88]	-0.30 [-0.43, -0.17]	-0.39 [-0.51, -0.26]
100	0.40	0.40	93.11 [76.28, 111.53]	0.42 [0.24, 0.60]	0.36 [0.16, 0.53]	97.23 [85.81, 109.40]	0.40 [0.29, 0.52]	0.39 [0.27, 0.50]
	0.40	0.60	91.84 [75.20, 110.16]	0.43 [0.27, 0.58]	0.54 [0.38, 0.67]	96.53 [85.19, 108.60]	0.41 [0.31, 0.51]	0.58 [0.49, 0.66]
	0.60	0.40	92.39 [75.67, 110.73]	0.60 [0.46, 0.74]	0.37 [0.21, 0.52]	96.81 [85.40, 108.93]	0.60 [0.51, 0.68]	0.40 [0.29, 0.49]
	0.60	0.60	90.83 [74.43, 108.78]	0.61 [0.48, 0.73]	0.55 [0.41, 0.67]	96.07 [84.74, 108.12]	0.60 [0.52, 0.68]	0.58 [0.50, 0.66]
	-0.50	-0.30	91.45 [73.40, 111.82]	-0.50 [-0.67, -0.33]	-0.25 [-0.45, -0.04]	97.42 [85.88, 109.60]	-0.50 [-0.60, -0.39]	-0.30 [-0.41, -0.17]
	-0.50	-0.40	91.82 [74.33, 111.39]	-0.51 [-0.66, -0.35]	-0.35 [-0.52, -0.16]	97.07 [85.63, 109.25]	-0.50 [-0.60, -0.40]	-0.39 [-0.49, -0.28]
	-0.30	-0.30	93.68 [76.52, 112.28]	-0.32 [-0.54, -0.08]	-0.26 [-0.48, -0.02]	97.74 [86.19, 110.04]	-0.30 [-0.45, -0.15]	-0.29 [-0.44, -0.14]
	-0.30	-0.40	93.52 [76.69, 111.98]	-0.33 [-0.53, -0.12]	-0.35 [-0.54, -0.14]	97.67 [86.18, 109.92]	-0.31 [-0.43, -0.17]	-0.39 [-0.51, -0.26]

4.2 Quasi-unit roots

In this section we employ the posterior distribution to quantify the probability that the characteristic polynomial associated to (2) has a unit root. In this simulation we consider a model β ARMA(2,0) with $\phi = (\phi_1, \phi_2)$ given in the Table 2. The β ARMA(2,0) model was fitted considering $n = 500$, $\nu = 100$ and the following priors: $\nu \sim \text{Gamma}(5, 0.1)$, $\phi_i \sim N(\mu, \sigma^2)$, where $\mu = 0$ and $\sigma^2 = 20,000^2$ for $i \in \{1, 2\}$. For each simulated series, two MCMC chains consisting of 2,000 samples each with a warm up period of 50% were ran. The associated characteristic polynomial is given by $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$ whose roots are $\frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}$. The parameter values were chosen in order to provide characteristic polynomials with a variety of roots, some have smallest modulus very close to 1 while others are far from 1. The smallest modulus of the roots for each parameter ϕ is presented in Table 2's third column. The results show the probability that the smallest modulus of the roots is smaller than $\{1.01, 1.02, 1.03, 1.04, 1.05\}$, where 1.05 is frequently used as a practical rule-of-thumb to analyze unit-roots, in the sense that if the modulus is 1.05 or smaller, then we consider that a unit root is present in the true characteristic polynomial.

Point estimates and credible intervals in this simulation were omitted to save up space, but the results were similar in quality as those presented in section 4.1. From the results, we

observe that when the true modulus is < 1.05 , the posterior probability detects this with very high probability, over 95% in all cases. In the most difficult cases, when the true modulus is 1.05 and 1.057, the posterior probability that the modulus is < 1.05 is still very high. As the true modulus move away from 1.05, the posterior probability that the modulus is < 1.05 decreases very rapidly to values close to 0.

Table 2: Posterior probability that the characteristic polynomial's roots have modulus smaller than a given threshold. First two columns present the true parameter values with the smallest modulus of the characteristic polynomial's roots in the third column. Thresholds are presented in the last 5 columns.

ϕ_1	ϕ_2	root	< 1.01	< 1.02	< 1.03	< 1.04	< 1.05
-0.25	-0.95	1.026	0.050	0.463	0.929	0.992	0.996
-0.15	0.80	1.028	0.034	0.184	0.504	0.816	0.957
0.20	0.75	1.029	0.081	0.356	0.743	0.946	0.995
-0.30	-0.90	1.050	0.002	0.016	0.133	0.474	0.823
-0.10	0.80	1.057	0.030	0.164	0.483	0.809	0.954
0.40	0.50	1.070	0.003	0.018	0.084	0.234	0.469
0.10	0.80	1.100	0.004	0.005	0.015	0.057	0.159
-0.90	-0.60	1.300	0.000	0.000	0.000	0.000	0.000
0.10	-0.30	1.830	0.000	0.000	0.000	0.000	0.000
0.30	0.10	2.000	0.002	0.002	0.002	0.002	0.002
0.60	-0.10	3.160	0.001	0.001	0.001	0.001	0.001

4.3 Sensitivity Analysis

In this section we present a prior sensitivity analysis for parameter ν , which presented the highest bias among the parameter estimates. The other parameters were well estimated using non-informative priors. For the sensitivity analysis, we consider β ARMA(1, 1) models with $\nu \in \{50, 100\}$, $(\phi, \theta) \in \{(0.4, 0.6), (-0.3, -0.3)\}$ and sample size $n = 200$. The BTSR package was used for data generation. For each simulated series, the posterior was estimated from two MCMC chains with 2,000 samples each and a warmup period of 50%.

The priors of ϕ and θ are the same as in Section 4.1, namely, $\phi \sim \text{Normal}(\mu, \sigma^2)$ and $\theta \sim \text{Normal}(\mu, \sigma^2)$ with $\mu = 0$ and $\sigma^2 = 20,000^2$. As for ν , we consider $\nu \sim \text{Gamma}(\alpha, \beta)$ for a variety of (α, β) values, presented in Table 3 along with the respective induced prior mean and variance. Parameter choices induce priors that can be informative, as when $\nu = 50$ and prior has mean 50 and variance 25, very non-informative, when variance is 2,000, and even very informative but equivocated, such as when the prior has mean 1 and variance 25, in which case, the prior probability that $\nu \leq 50$ is 0.9979 and that $\nu \leq 20$ is 0.9872. Each scenario was replicated 50 times. The results are presented in Table 4, which shows the posterior mean and credible interval for each scenario, organized by the prior mean and the prior variance of ν .

Table 3: Values of (α, β) used as prior for ν in the simulation and respective values of the prior mean and the prior variance.

		Prior Mean		
		1	50	100
Prior Variance	2000	(0.0005, 0.0005)	(1.25, 0.025)	(5, 0.05)
	500	(0.002, 0.002)	(5, 0.1)	(20, 0.2)
	25	(0.04, 0.04)	(100, 2)	(400, 4)

Table 4: Results of the prior sensitivity analysis. Presented are the estimated values and credible intervals for each parameter, organized by prior mean and prior variance.

DGP	Prior Var	Prior Mean					
		1		50		100	
		$\hat{\nu}$	$CI_{0.95}(\nu)$	$\hat{\nu}$	$CI_{0.95}(\nu)$	$\hat{\nu}$	$CI_{0.95}(\nu)$
Scenario 1 $\nu = 50$ $\phi = 0.40$ $\theta = 0.60$	2000	47.85	[39.00, 57.69]	47.89	[39.11, 57.63]	49.06	[40.19, 58.83]
		0.42	[0.26, 0.57]	0.42	[0.26, 0.57]	0.42	[0.26, 0.57]
		0.55	[0.40, 0.68]	0.55	[0.40, 0.68]	0.55	[0.40, 0.68]
	500	47.80	[38.92, 57.60]	47.91	[39.20, 57.47]	52.31	[43.52, 61.99]
		0.42	[0.26, 0.57]	0.42	[0.26, 0.57]	0.42	[0.27, 0.56]
		0.55	[0.40, 0.68]	0.55	[0.40, 0.68]	0.56	[0.41, 0.68]
	25	46.97	[38.31, 56.56]	48.81	[42.28, 55.73]	81.90	[74.90, 89.20]
		0.42	[0.26, 0.57]	0.42	[0.26, 0.57]	0.42	[0.30, 0.54]
		0.55	[0.40, 0.68]	0.55	[0.41, 0.68]	0.56	[0.45, 0.66]
Scenario 2 $\nu = 50$ $\phi = -0.30$ $\theta = -0.30$	2000	49.12	[40.07, 59.09]	49.11	[40.04, 59.05]	50.31	[41.15, 60.34]
		-0.31	[-0.53, -0.08]	-0.32	[-0.53, -0.08]	-0.32	[-0.53, -0.08]
		-0.27	[-0.49, -0.03]	-0.27	[-0.49, -0.03]	-0.27	[-0.48, -0.03]
	500	49.06	[40.00, 58.95]	49.12	[40.18, 58.89]	53.60	[44.52, 63.61]
		-0.32	[-0.53, -0.08]	-0.31	[-0.53, -0.08]	-0.32	[-0.53, -0.09]
		-0.27	[-0.48, -0.03]	-0.27	[-0.49, -0.03]	-0.27	[-0.48, -0.04]
	25	48.22	[39.26, 58.01]	49.45	[42.85, 56.49]	82.62	[75.59, 89.94]
		-0.31	[-0.54, -0.08]	-0.32	[-0.53, -0.08]	-0.32	[-0.49, -0.14]
		-0.27	[-0.49, -0.03]	-0.27	[-0.48, -0.03]	-0.27	[-0.44, -0.09]
Scenario 3 $\nu = 100$ $\phi = 0.40$ $\theta = 0.60$	2000	95.84	[78.10, 115.39]	94.93	[77.41, 114.18]	96.12	[78.65, 115.25]
		0.43	[0.27, 0.58]	0.43	[0.27, 0.58]	0.43	[0.27, 0.58]
		0.54	[0.39, 0.67]	0.54	[0.39, 0.67]	0.54	[0.39, 0.67]
	500	95.79	[77.97, 115.45]	91.84	[75.20, 110.16]	96.55	[80.10, 114.43]
		0.43	[0.27, 0.58]	0.43	[0.27, 0.58]	0.43	[0.27, 0.58]
		0.54	[0.39, 0.67]	0.54	[0.38, 0.67]	0.54	[0.39, 0.67]
	25	92.34	[75.16, 111.26]	65.65	[56.90, 74.99]	99.04	[90.55, 107.86]
		0.43	[0.27, 0.58]	0.43	[0.24, 0.60]	0.43	[0.28, 0.57]
		0.54	[0.38, 0.67]	0.53	[0.35, 0.69]	0.54	[0.39, 0.66]
Scenario 4 $\nu = 100$ $\phi = -0.30$ $\theta = -0.30$	2000	97.99	[79.70, 118.09]	96.88	[79.04, 116.54]	98.06	[80.35, 117.52]
		-0.32	[-0.54, -0.09]	-0.32	[-0.54, -0.09]	-0.32	[-0.54, -0.09]
		-0.26	[-0.47, -0.02]	-0.26	[-0.47, -0.02]	-0.26	[-0.47, -0.02]
	500	97.83	[79.68, 117.84]	93.68	[76.52, 112.28]	98.28	[81.55, 116.39]
		-0.32	[-0.54, -0.09]	-0.32	[-0.54, -0.08]	-0.32	[-0.54, -0.09]
		-0.26	[-0.48, -0.02]	-0.26	[-0.48, -0.02]	-0.26	[-0.47, -0.02]
	25	94.34	[76.60, 113.72]	66.15	[57.32, 75.64]	99.48	[90.98, 108.41]
		-0.32	[-0.54, -0.08]	-0.32	[-0.58, -0.04]	-0.32	[-0.54, -0.09]
		-0.26	[-0.48, -0.02]	-0.26	[-0.51, 0.03]	-0.26	[-0.47, -0.03]

The changes in the prior for ν did not affect point estimates for ϕ and θ . Overall, point estimates and credible intervals for ν are on par with the ones presented in Section 4.1, with a few exceptions. In scenarios 1 and 2, where the true value is $\nu = 50$, when the prior mean is 100 and the prior variance is 25, point estimates are very biased and credible intervals do not contain 50. In this case, the prior probability that $\nu < 85$ is only 0.008, which flattens out the posterior in the vicinity of the true value of ν , 50, where the likelihood is likely to be higher. In scenarios 3 and 4, for which $\nu = 100$, when the prior variance is 25 and the prior mean is 50, essentially the same happens with prior probability that $\nu \leq 65$ being 0.9972.

Interestingly, the most extreme case where the prior mean is 1 and the prior variance is 25 yielded good overall results, despite the very small prior probability for values of ν distant from 1. This is so because 1 is very distant from the likelihood peaks, hence the likelihood values in the vicinity of 1 are very small, counteracting the effect of the prior. Near the likelihood peaks, the prior is basically flat so the prior probability is essentially dominated by the likelihood, allowing for the identification of the peaks in the posterior probability.

Comparing the results presented in Table 4 and Table 1, we observe that applying a more informative prior, for which the mean matches the data generating process, yielded slightly better results. The best results were obtained when the prior mean is 100 and the worst when the prior variance is 25. We conclude that, as expected, the parameters are well estimated when the prior is uninformative, but ν is sensible to poor choices of informative priors, which should in practice be avoided.

5 Empirical Application

In this section we present an application of the proposed methodology to a real data set. We consider the monthly average proportion of hydroelectric energy stored in southern Brazil. The time series consists of monthly averages ranging from January 2001 to October 2016, yielding a sample size of $n = 190$ observations (Figure 1). Data from November 2016 to April 2017 ($n = 6$), was reserved for forecasting purposes. The same time series was used in Scher et al. (2020), where the authors fit 5 different models to the data and present a comparison among them in terms of out-of-sample forecast. The best predictive model was a β ARMA(1,1), with parameters $\alpha = 0.3452$, $\nu = 11.7593$, $\phi = 0.5235$ and $\theta = 0.3588$, obtained using the conditional maximum likelihood approach of Rocha and Cribari-Neto (2009). In this section we focus on applying the proposed Bayesian approach to compare the results with those obtained by Scher et al. (2020).

Under the proposed Bayesian approach we fit a β ARMA(1,1) and a β ARMA(1,0) considering the following priors: $\nu \sim \text{Gamma}(5, 0.1)$, $\alpha \sim \text{Uniform}(-1, 1)$ (intercept), $\phi \sim N(\mu, \sigma^2)$ and $\theta \sim N(\mu, \sigma^2)$, where $\mu = 0$ and $\sigma^2 = 20,000^2$. For each model, we run four chains consisting of 2,000 samples each with a warm up period of 50%. In Table 5 we compile the results, presenting point estimates (posterior mean), 95% credible interval and effective sample size (ESS). From Table 5, point estimates obtained with the proposed approach are close to the ones obtained in Scher et al. (2020). In both models, none of the credible intervals contain 0, which bears evidence for the relevance of including these parameters in the model. The ESS of all parameters are high, above 2,000, indicating a very low correlation among the sampled values from the posterior, as expected from the HMC approach. To provide additional information about the posterior distributions of the model parameters, we present in Figures 2 and 3 the density plots for each parameter.

A Bayes Factor model selection approach was employed to define which model to consider

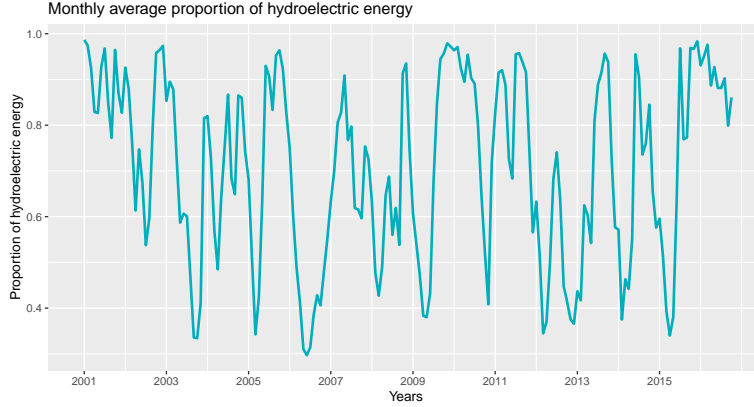


Figure 1: Monthly average proportion of hydroelectric energy stored in southern Brazil.

Table 5: Fitted β ARMA(1,1) and β ARMA(1,0) models for the proportion of hydroelectric energy stored in southern Brazil.

Parameter	α	ν	ϕ	θ
Estimates	0.36	10.73	0.52	0.35
CI _{0.95}	[0.19, 0.55]	[8.71, 12.94]	[0.37, 0.66]	[0.16, 0.51]
ESS	2094	3517	2021	2330
Estimates	0.23	10.01	0.65	—
CI _{0.95}	[0.12, 0.36]	[8.13, 12.02]	[0.57, 0.75]	—
ESS	2245	2887	2368	—

in this application. The idea is to compare several models in increasing order of complexity using a Bayes Factor approach to perform model selection. This is somewhat similar to model selection of stationary ARMA models in the Box and Jenkins framework. Table 6 presents the values of the (log) marginal likelihoods for each model. Log-Bayes Factor values, which measure how much more likely the data are under the null compared to the alternative model, are obtained as the difference in the log-marginal likelihood between the null and the alternative model, and can be computed directly from the table.

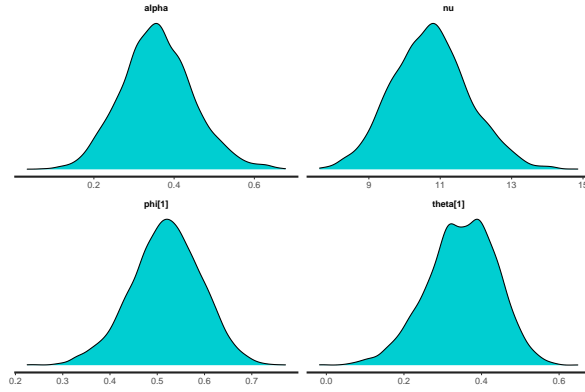


Figure 2: Posterior density plots of parameters α (top left), ν (top right), ϕ (bottom left) and θ (bottom right).

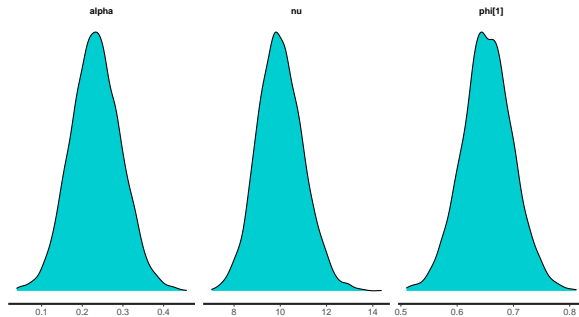


Figure 3: Posterior density plots of parameters α (left), ν (middle) and ϕ (right).

We note that the $\beta\text{ARMA}(1,0)$ model, with the largest log marginal likelihood, presents a log Bayes Factor of 6.19 when compared to the $\beta\text{ARMA}(1,1)$ with the second largest log marginal likelihood. This indicates decisive evidence in favor of the $\beta\text{ARMA}(1,0)$ model, and motivates its inclusion in this analysis. Furthermore, since the $\beta\text{ARMA}(1,1)$ performs well when compared to all other models considered (with a log Bayes Factor to the next best model of 4.46), and is of the same order used in Scher et al. (2020), allowing for a more direct comparison, it was also included.

We employ both models to generate out-of-sample forecasts and compare to the forecasts obtained for the $\beta\text{ARMA}(1,1)$ in Scher et al. (2020). As previously mentioned, the last six months (November 2016 to April 2017) were reserved for this purpose. Scher et al. (2020) only reports the mean absolute error (MAE) of forecast in their work, hence, we report the performance of the h -steps ahead forecasts obtained with the predictive posterior approach presented in Section 3.2 in terms of MAE. The results are presented in Table 7.

Table 6: Log Marginal Likelihood Results.

βARMA	(0, 1)	(1, 0)	(1, 1)	(1, 2)	(2, 1)
log-ML	107.26	117.91	111.72	99.69	101.47

Table 7: Results from the h -steps ahead forecasts for the Bayesian and frequentist approach. Presented are the MAE of forecasting for each model.

Model	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
Bayesian β ARMA(1, 1)	0.1184	0.1414	0.1371	0.1537	0.1778	0.1949
Bayesian β ARMA(1, 0)	0.0964	0.1331	0.1372	0.1572	0.1839	0.2020
Frequentist β ARMA(1, 1)	0.1244	0.1444	0.1364	0.1484	0.1694	0.1839

The results in Table 7 show that for $h \in \{1, 2\}$, the Bayesian β ARMA(1, 0) presented the best forecasted values in terms of MAE, followed by the Bayesian β ARMA(1, 1). For $h \in \{3, 4, 5, 6\}$, the frequentist β ARMA(1, 1) presented the best forecasted values. Comparing the Bayesian models, β ARMA(1, 0) is uniformly outperformed by the β ARMA(1, 1) for $h \in \{3, 4, 5, 6\}$, and only performs better for $h \in \{1, 2\}$, as mentioned before.

6 Discussion

In this paper we studied parameter estimation of β ARMA models under the Bayesian framework. The model considered standard forms for the prior distribution while the posterior distribution of the parameters was obtained applying the Hamiltonian Monte Carlo Markov Chain. We proposed the use of the posterior probability to approach the problem of unit roots in the model's systematic component. We considered a procedure consisting of a series of Bayes Factor tests to perform model selection and addressed the problem of forecasting using the posterior predictive distribution.

Monte Carlo simulations were conducted to evaluate the model's performance, which was divided into three parts. The first, presented in Section 4.1, provided evidence that the proposed Bayesian approach for β ARMA(1, 1) models performs well in terms of point and interval estimation. The method was able to accurately estimate the model parameters in a variety of scenarios, including different values of ν , ϕ , θ , and sample sizes. Even with a small sample size, the results showed that point estimation had only small biases for the parameters. The credible intervals were found to generally contain the true value, indicating good precision in the parameter estimates.

In Section 4.2, we put forth a Monte Carlo simulation to assess the proposed approach to unit roots in the context of β ARMA(2, 0) models. A variety of parameters were chosen to provide characteristic polynomials whose roots are near and far from one in modulus. The results show that the posterior probability can be successfully employed to quantify the uncertainty associated with a given quasi-unit root with high precision. Overall, the proposed Bayesian approach provides a reliable way of detecting unit roots in β ARMA models.

In Section 4.3, we performed a prior sensitivity analysis for parameter ν . We simulated a variety of β ARMA(1, 1) models, with different combinations of parameters, different sample sizes, and different priors for ν . The results showed that the Bayesian approach performs very well when the prior is non-informative. When the prior was informative and the information provided by the prior was correctly specified (low variance and correctly specified mean), the results showed a small improvement in the estimated values. However, in certain cases where the prior specification was informative, but the information provided was incorrect (low variance but misspecified mean, far from the true parameter), the effects in the posterior were evident, resulting in high bias in point estimates and distortions in the credible intervals. The best estimates were obtained when the prior mean was 100, while the worst were obtained

when the prior variance was 25. This suggests that caution should be used when specifying priors for ν , particularly when the prior variance is low.

In Section 5, we presented a real data application of the proposed Bayesian approach to the monthly average proportion of hydroelectric energy stored in southern Brazil. Two models, $\beta\text{ARMA}(1, 1)$ and $\beta\text{ARMA}(1, 0)$, were fitted using the Bayesian approach. We found that the $\beta\text{ARMA}(1, 0)$ model is the model selected by log-the Bayes factor approach, followed by a $\beta\text{ARMA}(1, 1)$. Both models were compared in terms of their out-of-sample forecasting capabilities to a $\beta\text{ARMA}(1, 1)$ selected with a frequentist approach in Scher et al. (2020). For most forecasting horizons, the Bayesian $\beta\text{ARMA}(1, 1)$ model generated more accurate forecasts than the $\beta\text{ARMA}(1, 0)$. Both Bayesian models outperform the frequentist model in the first few steps.

In summary, the proposed Bayesian approach provides a competitive, reliable and flexible tool for the estimation of βARMA models, with the potential to contribute to the dissemination of βARMA models and Bayesian analysis. One interesting topic for a future work is the development of an R package (or a user friendly script) capable of fitting the proposed Bayesian approach based on the Hamiltonian MCMC to further promote the method.

References

- Bayer, F. M., Bayer, D. M., Marinoni, A., and Gamba, P. (2020). A novel Rayleigh dynamical model for remote sensing data interpretation. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7):4989–4999.
- Bayer, F. M., Bayer, D. M., and Pumi, G. (2017). Kumaraswamy autoregressive moving average models for double bounded environmental data. *Journal of Hydrology*, 555:385–396.
- Benjamin, M. A., Rigby, R. A., and Stasinopoulos, D. M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association*, 98(461):214–223.
- Casarin, R., Valle, L. D., and Leisen, F. (2012). Bayesian model selection for beta autoregressive processes. *Bayesian Analysis*, 7(2):385–410.
- Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, 49(4):1057–1072.
- Duane, S., Kennedy, A., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195:216–222.
- Guolo, A. and Varin, C. (2014). Beta regression for time series analysis of bounded data, with application to Canada Google Flu Trends. *The Annals of Applied Statistics*, 8(1):74–88.
- Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.
- Mackay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge University Press.
- Neal, R. M. (2011). *Handbook of Markov chain Monte Carlo*. Chapman and Hall CRC Press.
- Palm, B. G. and Bayer, F. M. (2018). Bootstrap-based inferential improvements in beta autoregressive moving average model. *Communications in Statistics - Simulation and Computation*, 47(4):977–996.

- Pereira, L. A. and Tavares, M. (2011). Regressão beta para modelagem do rendimento metalúrgico na reciclagem de alumínio. *Anais da XII Escola de Modelos de Regressão*.
- Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.
- Prass, T. S., Carlos, J. H., Taufemback, C. G., and Pumi, G. (2022). Positive time series regression models.
- Prass, T. S. and Pumi, G. (2022). *BTSR: Bounded Time Series Regression*. R package version 0.1.1.
- Pumi, G., Prass, T. S., and Souza, R. R. a. (2021). A dynamic model for double-bounded time series with chaotic-driven conditional averages. *Scandinavian Journal of Statistics*, 48(1):68–86.
- Pumi, G., Prass, T. S., and Taufemback, C. G. (2022). Unit-Weibull autoregressive moving average models.
- Pumi, G., Rauber, C., and Bayer, F. M. (2020). Kumaraswamy regression model with Aranda-Ordaz link function. *TEST*, 29:1051–1071.
- Pumi, G., Valk, M., Bisognin, C., Bayer, F. M., and Prass, T. S. (2019). Beta autoregressive fractionally integrated moving average models. *Journal of Statistical Planning and Inference*, 200:196 – 212.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rocha, A. V. and Cribari-Neto, F. (2009). Beta autoregressive moving average models. *TEST*, 18(3):529–545.
- Rocha, A. V. and Cribari-Neto, F. (2017). Erratum to: Beta autoregressive moving average models. *TEST*, 26:451–459.
- Rogers, J. A., Polhamus, D., Gillespie, W. R., Ito, K., Romero, K., Qiu, R., Stephenson, D., Gastonguay, M. R., and Corrigan, B. (2012). Combining patient-level and summary-level data for Alzheimer’s disease modeling and simulation: a beta regression meta-analysis. *Journal of Pharmacokinetics and Pharmacodynamics*, 39(5):479 – 498.
- Scher, V. T., Cribari-Neto, F., Pumi, G., and Bayer, F. M. (2020). Goodness-of-fit tests for β ARMA hydrological time series modeling. *Environmetrics*, 31(3).
- Seifi, A., Ponnambalam, K., and Vlach, J. (2000). Maximization of manufacturing yield of systems with arbitrary distributions of component values. *Annals of Operations Research*, 99(1):373–383.
- Stan Development Team (2022). RStan: the R interface to Stan. R package version 2.21.5.
- Wang, X.-F. (2012). Joint generalized models for multidimensional outcomes: A case study of neuroscience data from multimodalities. *Biometrical Journal*, 54(2):264–280.
- Zou, K. H., Carlsson, M. O., and Quinn, S. A. (2010). Beta-mapping and beta-regression for changes of ordinal-rating measurements on Likert scales: A comparison of the change scores among multiple treatment groups. *Statistics in Medicine*, 29:2486–2500.