

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Modelos para Séries Temporais com Dados Discretos

Guilherme de Oliveira Cherobim

Dissertação de Mestrado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Guilherme de Oliveira Cherobim

Modelos para Séries Temporais com Dados Discretos

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Marinho Gomes de Andrade Filho

USP – São Carlos
Março de 2026

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

C522m Cherobim, Guilherme de Oliveira
Modelos para Séries Temporais com Dados Discretos
/ Guilherme de Oliveira Cherobim; orientador Marinho
Gomes de Andrade Filho. -- São Carlos, 2026.
87 p.

Dissertação (Mestrado - Programa
Interinstitucional de Pós-graduação em Estatística)
-- Instituto de Ciências Matemáticas e de
Computação, Universidade de São Paulo, 2026.

1. Distribuições Discretas. 2. Zero-Inflação. 3.
Zero-Deflação. 4. Modelos ZMD-GARMA. 5.
Procedimentos MCMC. I. Gomes de Andrade Filho,
Marinho, orient. II. Título.

Guilherme de Oliveira Cherobim

Time Series Models for Discrete Data

Dissertation submitted to the Institute of Mathematics and Computer Science – ICMC-USP and to the Department of Statistics – DEs-UFSCar – in accordance with the requirements of the Statistics Interagency Graduate Program, for the degree of Master in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Marinho Gomes de Andrade Filho

USP – São Carlos
March 2026

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato Guilherme de Oliveira Cherobim, realizada em 19/03/2026.

Comissão Julgadora:

Prof. Dr. Marinho Gomes de Andrade Filho (ICMC-USP)

Prof(a). Dr(a). Thelma Sáfadi (UFLA)

Prof. Dr. Danilo Alvares da Silva (MRC Biosta)

Para meus pais.

AGRADECIMENTOS

Aos meus pais, Mauro e Simone, pelo apoio incondicional e por sempre acreditarem no meu potencial. Sem o incentivo e a paciência de vocês, esta caminhada não teria sido possível.

Ao professor Marinho Gomes de Andrade Filho, pela excelente orientação, pela paciência e pelos ensinamentos ao longo destes anos.

Aos amigos que dividiram momentos de estudo, a rotina do laboratório, repassando batidas, montando tabuleiros ou apenas compartilhando momentos de tranquilidade. Obrigado por deixarem essa jornada mais leve.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

*“A vida só pode ser comprendida olhando-se para trás;
mas só pode ser vivida olhando-se para a frente.”
(Søren Kierkegaard)*

RESUMO

CHEROBIM, G. DE O. **Modelos para Séries Temporais com Dados Discretos**. 2026. 87 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2026.

Séries temporais para dados discretos são frequentemente encontradas em diversas aplicações práticas. Em muitos casos, os dados apresentam modificações na frequência de zeros, como inflação ou deflação, e precisam ser devidamente tratados. Para acomodar tais características, este trabalho propõe a família de modelos Discreto Zero-Modificado GARMA. O grande diferencial metodológico é a incorporação de um parâmetro de modificação de zeros dinâmico e variável no tempo, permitindo a transição do modelo para diferentes modificações de zero a cada instante do tempo. Como principal contribuição, é apresentado o modelo Skellam Zero-Modificado GARMA, que estende a formulação para abranger séries temporais no conjunto dos números inteiros. Sob o paradigma Bayesiano, a inferência e a estimação dos parâmetros são feitas através do algoritmo Hamiltoniano de Monte Carlo com Gradiente Estocástico. A eficiência do amostrador foi avaliada em estudos de simulação para a recuperação de parâmetros. A aplicabilidade e versatilidade dos modelos são ilustradas na análise de três conjuntos de dados reais com comportamentos distintos. Os resultados demonstram que os modelos propostos capturam de forma adequada as mudanças estruturais e dinâmicas de modificação de zeros das séries, apresentando também boa capacidade preditiva.

Palavras-chave: Distribuições Discretas; Zero-Inflação; Zero-Deflação; Modelos ZMD-GARMA; Procedimentos MCMC.

ABSTRACT

CHEROBIM, G. DE O. **Time Series Models for Discrete Data**. 2026. 87 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2026.

Time series for discrete data are frequently encountered in various practical applications. In many cases, the data exhibit changes in the frequency of zeros, such as inflation or deflation, and need to be properly handled. To accommodate such characteristics, this work proposes the GARMA Zero-Modified Discrete model family. The major methodological difference is the incorporation of a dynamic and time-variable zero-modification parameter, allowing the model to transition to different zero-modifications at each instant in time. As a main contribution, the GARMA Zero-Modified Skellam model is presented, which extends the formulation to encompass time series in the set of integers. Under the Bayesian paradigm, inference and parameter estimation are performed using the Stochastic Gradient Monte Carlo Hamiltonian algorithm. Sampler efficiency was evaluated in simulation studies for parameter recovery. The applicability and versatility of the models are illustrated in the analysis of three real datasets with distinct behaviors. The results demonstrate that the proposed models adequately capture the structural and dynamic changes in the series' zero modifications, also showing good predictive capacity.

Keywords: Discrete Distributions; Zero-Inflation; Zero-Deflation; ZMD-GARMA Models; MCMC Methods.

LISTA DE ILUSTRAÇÕES

Figura 1 – Gráficos das séries ZMPS-GARMA(p,q)	57
Figura 2 – Gráficos das cadeias, histogramas, autocorrelações e convergências de cada parâmetro do modelo MPS1.	58
Figura 3 – Gráficos das cadeias, histogramas, autocorrelações e convergências de cada parâmetro do modelo MPS2.	59
Figura 4 – Gráficos das cadeias, histogramas, autocorrelações e convergências de cada parâmetro do modelo MPS3.	59
Figura 5 – Gráficos das séries ZMS-GARMA(p,q)	61
Figura 6 – Gráficos das cadeias, histogramas e autocorrelações de cada parâmetro do modelo MS1.	61
Figura 7 – Gráficos das cadeias, histogramas e autocorrelações de cada parâmetro do modelo MS2.	62
Figura 8 – Anomalias de temperatura na superfície terrestre entre janeiro de 1920 e dezembro de 2024	64
Figura 9 – Histograma, autocorrelação e convergência das cadeias para o modelo ajustado.	65
Figura 10 – Dados observados e a média ajustada das anomalias de temperatura de janeiro de 1920 a dezembro de 2019	67
Figura 11 – Parâmetro \hat{p}_t estimado de janeiro de 1920 a dezembro de 2019.	67
Figura 12 – Predições um passo adiante para as anomalias mensais de temperatura entre 2020 a 2024	68
Figura 13 – Diferença de oferta e procura de Bitcoin no intervalo 300 segundos	69
Figura 14 – Histograma, autocorrelação e convergência das cadeias para o modelo ajustado.	70
Figura 15 – Dados observados e a média ajustada da diferença da compra e venda de Bitcoin no intervalo.	71
Figura 16 – Parâmetro \hat{p}_t ajustado para o intervalo.	71
Figura 17 – Predições um passo adiante para as a diferença de compra e venda de Bitcoin.	72
Figura 18 – Notificações semanais de casos de Chikungunya em Ribeirão Preto entre janeiro de 2016 a dezembro de 2025	73
Figura 19 – Histograma, autocorrelação e convergência das cadeias para o modelo ajustado.	74
Figura 20 – Dados observados e média ajustada das notificações de casos de Chikungunya entre janeiro de 2016 a dezembro de 2024.	75
Figura 21 – Parâmetro \hat{p}_t estimado entre janeiro de 2016 a dezembro de 2024	76

Figura 22 – Predições um passo adiante para as notificações semanais de Chikungunya
no ano de 2025. 76

LISTA DE ALGORITMOS

Algoritmo 1 – Metropolis-Hastings	44
Algoritmo 2 – Hamiltoniano de Monte Carlo	47
Algoritmo 3 – Hamiltoniano de Monte Carlo com Gradiente Estocástico	49
Algoritmo 4 – Intervalo de Credibilidade	53

LISTA DE TABELAS

Tabela 1 – Algumas distribuições da família PS.	34
Tabela 2 – Testes de convergência para as cadeias do SGHMC.	57
Tabela 3 – Resultados de simulação para os modelos ZMPS-GARMA.	60
Tabela 4 – Testes de convergência para as cadeias do SGHMC.	60
Tabela 5 – Resultados de simulação para os modelos ZMS-GARMA.	62
Tabela 6 – Resultados de convergência para o modelo ajustado.	66
Tabela 7 – Estimativas <i>a posteriori</i> para os parâmetros do modelo ZMS-GARMA(2,1) ajustado aos dados.	66
Tabela 8 – Resultados de convergência para o modelo ajustado.	70
Tabela 9 – Estimativas <i>a posteriori</i> para os parâmetros do modelo ZMS-GARMA(2,1) ajustado aos dados.	70
Tabela 10 – Critérios de seleção para os modelos ajustados para as notificações de Chikungunya.	73
Tabela 11 – Resultados de convergência para o modelo ajustado.	74
Tabela 12 – Estimativas <i>a posteriori</i> para os parâmetros do modelo ZMNB-GARMA(3,0) ajustado aos dados.	75

LISTA DE ABREVIATURAS E SIGLAS

ACF	Função de Autocorrelação
AIC	Critério de Informação de Akaike
B-GARMA	Binomial GARMA
BIC	Critério de Informação Bayesiano
COM-P	COM-Poisson
COM-P-GARMA	COM-Poisson GARMA
CPO	Ordinada Preditiva Condicional
D-GARMA	Discretos GARMA
EBIC	Critério de Informação Bayesiano Esperado
GARMA	Generalizados Autorregressivos de Médias Móveis
GP	Poisson Generalizada
GP-GARMA	Poisson Generalizados GARMA
GR	Gelman-Rubin
GW	Geweke
HMC	Hamiltoniano de Monte Carlo
HPD	Maior Densidade a <i>Posteriori</i>
INAR	Autorregressivos Inteiros
INGARCH	Generalizados Autorregressivos de Heterocedasticidade Condicional a Valores Inteiros
MAAPE	Percentual Médio do Arcotangente do Erro Absoluto
MAB	Média do Viés Absoluto
MAE	Erro Absoluto Médio
MAEM	Erro Absoluto Médio da Mediana
MCMC	Monte Carlo por Cadeia de Markov
MH	Metropolis-Hastings
NB	Binomial Negativa
NB-GARMA	Binomial Negativo GARMA
NUTS	<i>No-U-Turn Sampler</i>
P-GARMA	Poisson GARMA
PIG	Poisson-Inversa-Gaussiana
PIG-GARMA	Poisson-Inversa-Gaussiana GARMA

PS Série de Potência
PS-GARMA Séries de Potência GARMA
RMSE Raiz do Erro Quadrático Médio
S-GARMA Skellam GARMA
SGHMC Hamiltoniano de Monte Carlo com Gradiente Estocástico
SINAR Autorregressivo de Valores Inteiros com Sinal
TINAR Autorregressivo de Inteiros Verdadeiros
WAIC Critério de Informação de Ampla Aplicação
ZDD Discreta Zero Deflacionada
ZID Discreta Zero Inflada
ZMB-GARMA Binomial Zero-Modificado GARMA
ZMCOM-P-GARMA COM-Poisson Zero-Modificados GARMA
ZMD Discretas Zero-Modificadas
ZMD-GARMA Discreto Zero-Modificado GARMA
ZMGP-GARMA Poisson Generalizados Zero-Modificados GARMA
ZMNB-GARMA Binomial Negativo Zero-Modificado GARMA
ZMP-GARMA Poisson Zero-Modificado GARMA
ZMPIG-GARMA Poisson-Inversa-Gaussiana Zero-Modificado GARMA
ZMPS Série de Potência Zero-Modificado
ZMPS-GARMA Séries de Potência Zero-Modificado GARMA
ZMS Skellam Zero-Modificado
ZMS-GARMA Skellam Zero-Modificado GARMA
ZTD Discreta Zero Truncada

SUMÁRIO

1	INTRODUÇÃO	25
2	MODELOS GARMA ZERO-MODIFICADOS PARA DADOS DIS- CRETOS	29
2.1	Os modelos D-GARMA	29
2.2	Os modelos ZMD-GARMA	30
2.2.1	<i>Propriedades dos modelos ZMD-GARMA</i>	30
2.2.2	<i>Modificação hurdle</i>	31
3	ALGUNS CASOS PARTICULARES	33
3.1	Os modelos PS-GARMA e ZMPS-GARMA	33
3.1.1	<i>Distribuição de Poisson</i>	34
3.1.2	<i>Distribuição Poisson Generalizada</i>	35
3.1.3	<i>Distribuição COM-Poisson</i>	35
3.1.4	<i>Distribuição Binomial</i>	36
3.1.5	<i>Distribuição Binomial Negativa</i>	37
3.2	Os modelos PIG-GARMA e ZMPIG-GARMA	37
3.3	Os modelos S-GARMA e ZMS-GARMA	38
4	INFERÊNCIA PARA OS MODELOS	39
4.1	Função log-verossimilhança	39
4.1.1	<i>Família Séries de Potência</i>	41
4.1.2	<i>Distribuição Skellam</i>	41
4.2	Abordagem Bayesiana	42
4.3	Métodos MCMC	43
4.3.1	<i>Metropolis-Hastings</i>	43
4.3.2	<i>Hamiltoniano de Monte Carlo</i>	44
4.3.3	<i>Hamiltoniano de Monte Carlo com Gradiente Estocástico</i>	46
4.4	Estimativas Posteriores de Monte Carlo	49
4.5	Seleção de Modelos	49
4.6	Distribuições Preditivas	50
4.7	Intervalos de Credibilidade	52
5	ESTUDOS DE SIMULAÇÃO	55

5.1	Processo de Simulação	55
5.2	ZMPS-GARMA(p,q)	56
5.3	ZMS-GARMA(p,q)	58
6	APLICAÇÕES EM DADOS REAIS	63
6.1	Anomalias na temperatura na superfície da terra	63
6.2	Compra e venda de Bitcoin	68
6.3	Notificações de casos de Chikungunya na cidade de Ribeirão Preto	71
7	CONSIDERAÇÕES FINAIS	77
	REFERÊNCIAS	79
	APÊNDICE A GRADIENTES DA FAMÍLIA PS	85
A.1	Poisson	85
A.2	Poisson Generalizada	85
A.3	COM-Poisson	86
A.4	Binomial	86
A.5	Binomial Negativa	87

INTRODUÇÃO

A modelagem de séries temporais para dados discretos tem recebido atenção crescente nos últimos anos. Diversos estudos se concentram nesse tipo de observação pela sua ampla aplicabilidade, como nas áreas de saúde, finanças e criminalidade. [Andrade, Conceição e Ravishanker \(2023\)](#) modelaram o número de mortes causadas pela influenza na cidade de São Paulo. [Andrews e Balakrishna \(2024\)](#) examinaram dados sobre Hepatite B e quantidades de transações de ativos financeiros. [Lastra, Pumi e Prass \(2025\)](#) estudaram o número da produção de automóveis e exportação de ônibus no Brasil. Crimes relacionados a drogas são estudados por [Alomani, Alzaid e Omair \(2018\)](#).

O caso mais comum de dados discretos em séries temporais é o dos inteiros não-negativos, comumente denominados dados de contagem. Modelos para dados de contagem seguindo uma formulação da família exponencial geral podem ser classificados em modelos de observações e modelos paramétricos ([COX *et al.*, 1981](#)), tal que a principal diferença está em como a estrutura de dependência é incorporada ao modelo. Em modelos de observações a distribuição da observação y_t é condicionada às observações passadas (y_{t-1}, \dots, y_1) , enquanto que nos modelos paramétricos a distribuição da observação é condicionada a um processo latente dinâmico.

Há uma vasta literatura sobre modelos de observações. Começando por [Zeger e Qaqish \(1988\)](#), que consideram uma classe de modelos de Markov nos quais a média e a variância condicionadas ao passado são funções explícitas de observações passadas, e [Zeger \(1988\)](#), que propõe um modelo regressivo para séries de contagem correlacionadas. [Li \(1994\)](#) estende o modelo autorregressivo proposto por [Zeger e Qaqish \(1988\)](#) adicionando uma estrutura de médias móveis à estrutura. [Shephard \(1995\)](#) considerou extensões dos modelos autorregressivos e de médias móveis para distribuições na família exponencial, tal conceito foi estendido e formalizado por [Benjamin, Rigby e Stasinopoulos \(2003\)](#) como os modelos Generalizados Autorregressivos de Médias Móveis (GARMA).

A família exponencial é amplamente utilizada para a modelagem de dados de contagem

no contexto regressivo e em séries temporais por englobar diversas distribuições discretas (ANDRADE; ANDRADE; EHLERS, 2015; FOKIANOS; KEDEM, 2004; LASTRA; PUMI; PRASS, 2025). A distribuição Poisson Inversa Gaussiana também pode ser utilizada para a modelagem de dados de contagem (PALA; CARVALHO; SÁFADI, 2023). Por reunir diversas distribuições discretas e ter suas propriedades bem estudadas, a família de distribuições Série de Potência é uma alternativa frequentemente utilizada para a modelagem de dados de contagem no contexto de regressão, como feito por Conceição *et al.* (2024), e para modelos GARMA, como em Andrade, Conceição e Ravishanker (2023).

Abordagens alternativas aos modelos GARMA foram propostos e utilizados ao longo dos anos, como o caso dos modelos Autorregressivos Inteiros (INAR) propostos por Alzaid e Al-osh (1990) e os modelos Generalizados Autorregressivos de Heterocedasticidade Condicional a Valores Inteiros (INGARCH) descritos por Ferland, Latour e Oraichi (2006).

Outro interesse para o estudo de dados discretos é o das observações no conjunto dos números inteiros. Várias abordagens para a modelagem de séries temporais com valores inteiros podem ser vistas em Karlis e Ntzoufras (2008) e Li, Chen e Zhu (2024). Em especial, dados inteiros podem ser modelados com o uso da distribuição Skellam. No contexto de regressão, Alzaid e Omair (2010), Karlis e Ntzoufras (2008) e Conceição, Suzuki e Andrade (2020) apresentam estudos de caso para dados inteiros utilizando a distribuição Skellam. Para séries temporais, alguns modelos foram propostos utilizando a distribuição Skellam: Freeland (2010) propôs o modelo Autorregressivo de Inteiros Verdadeiros (TINAR) de ordem 1 e Kachour e Truquet (2010) descreveu o modelo Autorregressivo de Valores Inteiros com Sinal (SINAR) de ordem p , ambas extensões do modelo INAR; Alomani, Alzaid e Omair (2018) propuseram um modelo INGARCH simétrico para a distribuição Skellam e Hu e Andrews (2021) exibiram um modelo INGARCH assimétrico.

Em muitas situações práticas, os dados de interesse apresentam excesso de zeros. Uma situação com menos ocorrência, mas igualmente importante, é a escassez de zeros. Tal comportamento é chamado de modificação de zeros. Umbach (1981) propôs a distribuição Poisson Zero-Inflacionada assumindo uma mistura entre a distribuição degenerada em zero e a distribuição tradicional de Poisson. Dietz e Böhning (2000) considerou também o caso de deflação de zeros na distribuição Poisson e definiu a distribuição Poisson Zero-Modificada. Conceição *et al.* (2017) estende a definição para os modelos Série de Potência Zero-Modificado (ZMPS) e Conceição, Suzuki e Andrade (2020) o faz para a distribuição Skellam Zero-Modificado (ZMS), ambos para o contexto de regressão. Andrade, Conceição e Ravishanker (2023) estendem a distribuição ZMPS para o modelo GARMA ao permitir que a razão de mistura variasse ao longo do tempo e Balakrishna, Anvar e Abraham (2024) exibem uma aplicação para séries temporais com intensidades de Markov.

O objetivo deste trabalho é a junção das distribuições Discretas Zero-Modificadas (ZMD) como descritas em Dietz e Böhning (2000), Conceição *et al.* (2017) e Conceição, Suzuki e

Andrade (2020) com os modelos GARMA, como feito por Andrade, Conceição e Ravishanker (2023), para propor os modelos Discretos GARMA (D-GARMA) para séries temporais e sua extensão para os modelos Discreto Zero-Modificado GARMA (ZMD-GARMA). Uma contribuição deste trabalho é a aplicação da distribuição Skellam nos modelos GARMA, preenchendo essa lacuna na literatura. O [Capítulo 2](#) apresenta a construção dos modelos D-GARMA e ZMD-GARMA. No [Capítulo 3](#) são apresentados casos particulares dos modelos.

A abordagem Bayesiana é amplamente utilizada na estimação de parâmetros dos modelos GARMA e diversos métodos são utilizados para a amostragem da distribuição posterior de interesse. O algoritmo de Metropolis-Hastings é utilizado por Andrade, Andrade e Ehlers (2015). Pala, Carvalho e Sáfadi (2023) e Andrade, Conceição e Ravishanker (2023) empregam o Hamiltoniano de Monte Carlo. Lastra, Pumi e Prass (2025) recorrem ao algoritmo de *Reversible Jump* para a estimação e seleção simultâneas de modelos.

Este trabalho utilizou o algoritmo Hamiltoniano de Monte Carlo com Gradiente Estocástico para a estimação dos parâmetros dos modelos ZMD-GARMA. A implementação do método e suas dependências foi feita utilizando a linguagem de programação R (R Core Team, 2023). Detalhes sobre a abordagem Bayesiana e os algoritmos utilizados são tratados no [Capítulo 4](#).

No [Capítulo 5](#) são exibidos estudos de simulação de alguns casos particulares do modelos ZMD-GARMA para a validação do método utilizado. No [Capítulo 6](#) são feitos estudos de caso para três conjuntos de dados reais para exemplificar a aplicabilidade do modelo. Por fim, no [Capítulo 7](#) são apresentadas as considerações sobre o estudo realizado e as perspectivas de trabalhos futuros.

MODELOS GARMA ZERO-MODIFICADOS PARA DADOS DISCRETOS

Neste capítulo são introduzidos os modelos D-GARMA, tratados na [Seção 2.1](#). Na [Seção 2.2](#) é feita a extensão para os modelos ZMD-GARMA, onde também são enunciadas suas principais propriedades.

2.1 Os modelos D-GARMA

Nos modelos para séries temporais com Dados Discretos $\{Y_t, t = 0, 1, 2, \dots\}$ é uma série temporal que assume valores inteiros. Seja $\mathbf{x}'_t = (x_{0t}, \dots, x_{mt})$ um vetor de dimensão $(m + 1) \times 1$ tal que $x_{0t} = 1$ e $x_{jt}, j = 1, \dots, m$ são variáveis exógenas conhecidas. A informação disponível até o tempo t é denotada por $\mathcal{F}_t = (y_{t-1}, \dots, y_1, \mathbf{x}_t, \dots, \mathbf{x}_1)$. Se Y_t tem distribuição condicional com parâmetros de média $\mu_t = E(Y_t | \mathcal{F}_t)$ e dispersão ν , então sua função massa de probabilidade condicional é definida por

$$\pi_D(y_t | \mu_t, \nu, \mathcal{F}_t) = \text{Prob}(Y_t = y_t | \mu_t, \nu, \mathcal{F}_t), \quad y_t \in \mathcal{A}_s, \quad (2.1)$$

em que o suporte \mathcal{A}_s é o subconjunto dos inteiros $\{s, s + 1, s + 2, \dots\}$, com $s \in \mathbb{Z}$.

O modelo GARMA para distribuições discretas D-GARMA considera a função massa de probabilidade dada em (2.1) e uma componente sistemática para a média μ_t que se relaciona aos dados passados \mathcal{F}_t através de uma função de ligação adequada $\psi(\mu_t)$ monótona e duplamente diferenciável da forma

$$\psi(\mu_t) = \eta_t = \mathbf{x}'_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j \{\psi(y_{t-j}^*) - \mathbf{x}'_{t-j} \boldsymbol{\beta}\} + \sum_{j=1}^q \theta_j \{\psi(y_{t-j}^*) - \eta_{t-j}\}, \quad (2.2)$$

tal que y_t^* é alguma acomodação de y_t para o domínio de $\psi(\cdot)$, $\psi(y_{t-j}^*) - \mathbf{x}'_{t-j} \boldsymbol{\beta}$ é o termo autorregressivo, $\psi(y_{t-j}^*) - \eta_{t-j}$ é o termo de desvio, ou de médias móveis, e os vetores $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ e $\boldsymbol{\theta}$ são os parâmetros de interesse do modelo.

2.2 Os modelos ZMD-GARMA

O modelo ZMD-GARMA surge da modificação da função massa de probabilidade em (2.1). A distribuição Zero-Modificada associada a uma distribuição discreta tem sua função massa de probabilidade condicional definida por

$$\pi_{ZMD}(y_t | \mu_t, \nu, p_t, \mathcal{F}_t) = (1 - p_t)I_{\{y_t=0\}} + p_t\pi_D(y_t | \mu_t, \nu, \mathcal{F}_t), \quad y \in \mathcal{A}_s, \quad (2.3)$$

em que $I_{\{y_t=0\}}$ é uma função indicadora com $I_{\{y_t=0\}} = 1$ se $y_t = 0$ e $I_{\{y_t=0\}} = 0$ caso contrário. Para que $\pi_D(y_t | \mu_t, \nu, \mathcal{F}_t)$ seja uma função de probabilidade válida é necessário que p_t satisfaça

$$0 \leq p_t \leq \frac{1}{1 - \pi_D(0 | \mu_t, \nu, \mathcal{F}_t)}. \quad (2.4)$$

O parâmetro p_t controla a frequência de zeros na série. Diferentes valores de p_t geram diferentes distribuições Zero-Modificadas (CONCEIÇÃO *et al.*, 2017; CONCEIÇÃO; SUZUKI; ANDRADE, 2020):

1. se $p_t = 0, \forall t$, (2.3) é a distribuição degenerada com toda a sua massa em zero.
2. para $0 < p_t < 1, \forall t$, (2.3) é a distribuição Discreta Zero Inflada (ZID).
3. quando $p_t = 1, \forall t$, (2.3) é a distribuição discreta tradicional.
4. para $1 < p_t < \frac{1}{1 - \pi_D(0; \mu_t, \nu)}, \forall t$, (2.3) é a distribuição Discreta Zero Deflacionada (ZDD).
5. se $p_t = \frac{1}{1 - \pi_D(0; \mu_t, \nu)}, \forall t$, (2.3) é a distribuição Discreta Zero Truncada (ZTD), que tem sua função massa de probabilidade da forma

$$\pi_{ZTD}(y_t | \mu_t, \nu) = \frac{\pi_D(y_t | \mu_t, \nu)}{1 - \pi_D(0 | \mu_t, \nu)}(1 - I_{y_t=0}), \quad y_t \in \mathcal{A}_s.$$

Em muitas situações práticas uma série temporal apresenta diferentes modificações de zeros em tempos distintos. O diferencial do modelo ZMD-GARMA está na natureza dinâmica do parâmetro p_t . Assumir uma modificação de zeros variante no tempo permite que o modelo capture naturalmente os cenários de inflação ou deflação de zeros da série. O parâmetro p_t é estruturado para possuir memória temporal, relacionando-se à média passada da série e a valores exógenos, como será detalhado na [Subseção 2.2.2](#), o que confere interpretabilidade prática a p_t .

2.2.1 Propriedades dos modelos ZMD-GARMA

A média condicional do modelo ZMD-GARMA é obtida diretamente de (2.3) como

$$\mu_t^{ZMD} = p_t \mu_t^D \quad (2.5)$$

e a variância da forma

$$v_t^{ZMD} = p_t \left[v_t^D + (1 - p_t) \left(\mu_t^D \right)^2 \right], \quad (2.6)$$

tais que μ_t^D e v_t^D são a média e a variância da distribuição discreta, respectivamente. No caso particular que $p_t = p$, a média e variância incondicionais ficam

$$E(Y_t) = E(E(Y_t | \mathcal{F}_t)) = pE(\mu_t^D)$$

e

$$\begin{aligned} \text{Var}(Y_t) &= \text{Var}(E(Y_t | \mathcal{F}_t)) + E(\text{Var}(Y_t | \mathcal{F}_t)) \\ &= p\text{Var}(\mu_t^D) + p \left[E(V(\mu_t^D, v)) + (1 - p) \left(E(\mu_t^D) \right)^2 \right]. \end{aligned}$$

A média μ_t e os valores defasados Y_{t-j} se relacionam pela função de ligação (2.2), que geralmente não são lineares e tornam difícil o cálculo da média e da variância incondicionais explicitamente. O mesmo pode ser dito da função de autocorrelação, dada por

$$\rho_j = \frac{\text{Cov}(Y_t, Y_{t-j})}{\sqrt{\text{Var}(Y_t) \text{Var}(Y_{t-j})}} = \frac{E(Y_t Y_{t-j}) - E(Y_t) E(Y_{t-j})}{\sqrt{\text{Var}(Y_t) \text{Var}(Y_{t-j})}}.$$

2.2.2 Modificação hurdle

Assim como em [Andrade, Conceição e Ravishanker \(2023\)](#), uma modificação *hurdle* pode ser obtida para o modelo ZMD-GARMA de forma a reduzir o custo computacional de estimações ([CONCEIÇÃO et al., 2017](#)). Combinando (2.1) com

$$w_t = p_t(1 - \pi_D(0 | \mu_t, v, \mathcal{F}_t)) \quad (2.7)$$

da forma

$$\begin{aligned} \pi_{ZMD}(y_t | \mu_t, v, p_t, \mathcal{F}_t) &= \pi_{ZMD}(0 | \mu_t, v, \mathcal{F}_t) I_{y=0} \\ &\quad + \pi_{ZMD}(y_t | \mu_t, v, \mathcal{F}_t) (1 - I_{y=0}) \\ &= I_{\{y=0\}} [(1 - p_t) + p_t \pi_D(0 | \mu_t, v, \mathcal{F}_t)] \\ &\quad + (1 - I_{\{y=0\}}) p_t \pi_D(p_t \pi_D(y_t | \mu_t, v, \mathcal{F}_t)) \\ &= [1 - p_t(1 - \pi_D(0 | \mu_t, v, \mathcal{F}_t))] I_{y=0} \\ &\quad + p_t(1 - \pi_D(0 | \mu_t, v, \mathcal{F}_t)) \frac{\pi_D(y_t | \mu_t, v, \mathcal{F}_t)}{1 - \pi_D(0 | \mu_t, v, \mathcal{F}_t)} (1 - I_{y=0}), \end{aligned}$$

como (2.4) pode ser reescrito da forma $0 \leq p_t(1 - \pi_D(0 | \mu_t, v, \mathcal{F}_t)) \leq 1$, a parametrização em (2.7) resulta em $0 \leq w_t \leq 1$. Substituindo (2.7) na equação acima resulta em

$$\begin{aligned} \pi_{ZMD}(y_t | \mu_t, v, \mathcal{F}_t) &= (1 - w_t) I_{\{y_t=0\}} + w_t \frac{\pi_D(y_t | \mu_t, v, \mathcal{F}_t)}{1 - \pi_D(0 | \mu_t, v, \mathcal{F}_t)} (1 - I_{\{y_t=0\}}) \\ &= (1 - w_t) I_{\{y_t=0\}} + w_t \pi_{ZTD}(y_t | \mu_t, v, \mathcal{F}_t). \end{aligned} \quad (2.8)$$

Assim como p_t , w_t controla a modificação de zeros na série da forma

1. se $w_t = 0, \forall t$, a (2.8) é a distribuição degenerada com toda sua massa em zero.
2. para $0 < w_t < 1 - \pi_D(0; \mu_t, \nu), \forall t$, (2.8) é a distribuição ZID.
3. se $w_t = 1 - \pi_D(0; \mu_t, \nu), \forall t$, (2.8) é a distribuição discreta tradicional.
4. para $1 - \pi_D(0; \mu_t, \nu) < \mu_t < 1, \forall t$, a expressão (2.8) é a distribuição ZDD.
5. se $w_t = 1, \forall t$, (2.8) é a distribuição ZTD.

A função de ligação logito é utilizada para w_t , dada por

$$\log \left(\frac{w_t}{1 - w_t} \right) = \mathbf{v}'_t \boldsymbol{\gamma} + \sum_{j=1}^s \lambda_j \eta_{t-j}, \quad (2.9)$$

com parâmetros $\boldsymbol{\gamma}' = (\gamma_0, \gamma_1, \dots, \gamma_r)$, $\boldsymbol{\lambda}' = (\lambda_1, \dots, \lambda_s)$, tal que $\mathbf{v}'_t = (v_{0t}, \dots, v_{rt})$ é um vetor de dimensão $(r+1) \times 1$ de variáveis exógenas, podendo ser um subconjunto de \mathbf{x}_t e a ordem de defasagem s é escolhida por critérios de seleção para um bom ajuste do modelo, como será visto no [Capítulo 4](#). A função logito não é aplicável para a distribuição degenerada com sua massa toda em zero ($w_t = 0$) ou à distribuição zero-truncada ($w_t = 1$).

ALGUNS CASOS PARTICULARES

Diferentes distribuições de probabilidade levam a diferentes modelos D-GARMA e ZMD-GARMA. A família de distribuições Série de Potência será abordada na [Seção 3.1](#) com suas principais propriedades e casos específicos. Duas distribuições provenientes da distribuição de Poisson também foram consideradas: a Poisson-Inversa-Gaussiana, para casos de sobredispersão, apresentada na [Seção 3.2](#) e a distribuição Skellam, que estende o estudo para dados discretos inteiros, discutida na [Seção 3.3](#).

3.1 Os modelos PS-GARMA e ZMPS-GARMA

A família de distribuições Série de Potência (PS) foi definida por [Noack \(1950\)](#) e reúne diversas distribuições discretas. Suas principais propriedades são demonstradas por [Noack \(1950\)](#) e [Patil \(1962\)](#).

Seja $\{Y_t, t = 0, 1, 2, \dots\}$ uma sequência definida nos inteiros não-negativos $\mathcal{A}_0 = \{0, 1, 2, \dots\}$, Y_t tem distribuição condicional pertencente à família PS com média $\mu_t > 0$ se sua distribuição massa de probabilidade condicional é da forma

$$\pi_{PS}(y_t | \mu_t, \nu, \mathcal{F}_t) = \frac{a(y_t, \nu)g(\mu_t, \nu)^{y_t}}{f(\mu_t, \nu)}, \quad y_t \in \mathcal{A}_0, \quad (3.1)$$

tal que as funções $a(y_t, \nu)$, $f(\mu_t, \nu)$ e $g(\mu_t, \nu)$ são positivas, finitas e duas vezes diferenciáveis, com

$$f(\mu_t, \nu) = \sum_{y_t \in \mathcal{A}_0} a(y_t, \nu)g(\mu_t, \nu)^{y_t}.$$

As distribuições de Poisson, COM-Poisson, Poisson Generalizada, Binomial e Binomial Negativa são todas pertencentes à família PS. A [Tabela 1](#) apresenta os componentes dessas distribuições.

Tabela 1 – Algumas distribuições da família PS.

PS	Distribuição	$f(\mu_t, \nu)$	$g(\mu_t, \nu)$	$a(y_t, \nu)$	\mathcal{A}_0
P	Poisson	e^{μ_t}	μ_t	$\frac{1}{y_t!}$	$\{0, 1, \dots\}$
GP	Poisson Generalizada	$e^{\mu_t(1+\mu_t\nu)^{-1}}$	$\frac{\mu_t e^{-\mu_t\nu(1+\mu_t\nu)^{-1}}}{1+\mu_t\nu}$	$\frac{(1+\nu y_t)^{y_t-1}}{y_t!}$	$\{0, 1, \dots\}$
COM-P	COM-Poisson	$\sum_{y_t=0}^{\infty} g(\mu_t, \nu) a(y_t, \nu)$	μ_t^ν	$\left(\frac{1}{y_t!}\right)^\nu$	$\{0, 1, \dots\}$
B	Binomial	$\binom{m}{m-\mu_t}^m$	$\frac{\mu_t}{m-\mu_t}$	$\binom{m}{y_t}$	$\{0, 1, \dots, m\}$
NB	Binomial Negativa	$\left(\frac{\nu}{\mu_t + \nu}\right)^{-\nu}$	$\frac{\mu_t}{\mu_t + \nu}$	$\frac{\Gamma(\nu + y_t)}{y_t! \Gamma(\nu)}$	$\{0, 1, \dots\}$

Fonte: Adaptada de [Andrade, Conceição e Ravishanker \(2023\)](#).

A média condicional de Y_t é dada por

$$\mu_t^{PS} = E(Y_t | \mathcal{F}_t) = \frac{f_\mu(\mu_t, \nu) g(\mu_t, \nu)}{f(\mu_t, \nu) g_\mu(\mu_t, \nu)} \quad (3.2)$$

e a respectiva variância condicional por

$$\nu_t^{PS} = \text{Var}(Y_t | \mathcal{F}_t) = \frac{g(\mu_t, \nu)}{g_\mu(\mu_t, \nu)}, \quad (3.3)$$

tais que $f_\mu(\mu_t, \nu)$ e $g_\mu(\mu_t, \nu)$ denotam as derivadas primeira de $f(\mu_t, \nu)$ e $g(\mu_t, \nu)$ em relação a μ , respectivamente, e são exibidas na [Seção 4.1](#). À exceção da COM-Poisson, a média condicional das distribuições apresentadas na [Tabela 1](#) é $\mu_t^{PS} = \mu_t$.

O modelo D-GARMA cuja distribuição dos dados Y_t pertence à família PS será chamado modelo Séries de Potência GARMA (PS-GARMA), assim como o modelo ZMD-GARMA receberá o nome de modelo Séries de Potência Zero-Modificado GARMA (ZMPS-GARMA).

3.1.1 Distribuição de Poisson

A distribuição de Poisson é uma distribuição de probabilidade discreta que descreve a probabilidade de um certo número de eventos ocorrer dentro de um intervalo fixo do tempo, supondo que a taxa média de ocorrência é conhecida e os eventos ocorrem de forma independente. Comumente utilizada para modelar a ocorrência de eventos raros em relação ao intervalo.

Substituindo seus componentes apresentados na [Tabela 1](#) em (3.1), a distribuição de Poisson assume a forma

$$\pi_P(y_t | \mu_t, \nu, \mathcal{F}_t) = \frac{e^{\mu_t} \mu_t^{y_t}}{y_t!},$$

a média condicional é dada por $\mu_t^P = \mu_t$ e sua variância condicional em (3.3) fica

$$\nu_t^P = \mu_t.$$

Os modelos PS-GARMA e ZMPS-GARMA cujos dados assumem distribuição de Poisson serão denominados Poisson GARMA (P-GARMA) e Poisson Zero-Modificado GARMA (ZMP-GARMA), respectivamente. Para os modelos ZMP-GARMA a média e a variância condicionais em (2.5) e (2.6) ficam, respectivamente, $\mu_t^{ZMP} = p_t \mu_t$ e

$$v_t^{ZMP} = p_t [\mu_t + (1 - p_t) \mu_t^2].$$

3.1.2 Distribuição Poisson Generalizada

A distribuição de Poisson Generalizada (GP) é uma extensão da distribuição de Poisson clássica que permite maior flexibilidade ao modelar dados de contagem por não considerar a média e a variância iguais. Em muitas ocasiões reais os dados podem apresentar variância maior do que a média, sobredispersão, ou variância menor do que a média, subdispersão.

Sua equação na forma (3.1) fica

$$\pi_{GP}(y_t | \mu_t, v, \mathcal{F}_t) = \frac{(1 + v y_t)^{y_t - 1} \left(\frac{\mu_t e^{-\mu_t v (1 + \mu_t v)^{-1}}}{1 + \mu_t v} \right)^{y_t}}{e^{\mu_t (1 + \mu_t v)^{-1}}},$$

com média condicional $\mu_t^{GP} = \mu_t$ e variância condicional (3.3) da forma

$$v_t^{GP} = \frac{\frac{\mu_t e^{-\mu_t v (1 + \mu_t v)^{-1}}}{1 + \mu_t v}}{e^{-\mu_t v (1 + \mu_t v)^{-1}}} = \mu_t (1 + \mu_t v)^2.$$

Os modelos PS-GARMA que assumem distribuição GP são designados Poisson Generalizados GARMA (GP-GARMA) e os respectivos modelos ZMPS-GARMA são Poisson Generalizados Zero-Modificados GARMA (ZMGP-GARMA). Para os modelos ZMGP-GARMA a média condicional é dada por $\mu_t^{ZMGP} = p_t \mu_t^{GP}$ e sua variância (2.6) por

$$v_t^{ZMGP} = p_t [\mu_t (1 + \mu_t v)^2 + (1 - p_t) (\mu_t)^2].$$

3.1.3 Distribuição COM-Poisson

A distribuição de Conway-Maxwell-Poisson, ou COM-Poisson (COM-P), também permite modelar dados com subdispersão ou sobredispersão. Os componentes apresentados na Tabela 1 são referentes à parametrização em Guikema e Goffelt (2008), que proporciona um parâmetro ligado à tendência central da COM-P para uma variedade de parâmetros v enquanto o mantém como parâmetro de forma. A distribuição massa de probabilidade (3.1) assume a forma

$$\pi_{COM-P}(y_t, | \mu_t, v, \mathcal{F}_t) = \frac{\left(\frac{1}{y_t!} \right)^v (\mu_t^v)^{y_t}}{\sum_{s=0}^{\infty} \left(\frac{\mu_t^s}{s!} \right)^s}.$$

Diferente das demais distribuições da família PS, sua média e variância condicionais são aproximadas assintoticamente por

$$\mu_t^{COM-P} \approx \mu_t + \frac{1-v}{2v}$$

e

$$v_t^{COM-P} \approx \frac{\mu_t}{v}.$$

Quando os dados observados Y_t assumem distribuição COM-Poisson seus modelos são designados COM-Poisson GARMA (COM-P-GARMA) e COM-Poisson Zero-Modificados GARMA (ZMCOM-P-GARMA). Para os modelos ZMCOM-P-GARMA, a média condicional é dada por

$$\mu_t^{ZMCOM-P} \approx p_t \left(\mu_t + \frac{1-v}{2v} \right)$$

e a variância condicional é da forma

$$v_t^{ZMCOM-P} \approx p_t \left[\frac{\mu_t}{v} + (1-p_t) \left(\mu_t + \frac{1-v}{2v} \right)^2 \right].$$

3.1.4 Distribuição Binomial

A distribuição Binomial é uma distribuição de probabilidade discreta que descreve o número de sucessos em m tentativas independentes, nas quais cada tentativa pode resultar apenas em sucesso ou fracasso.

A expressão da função massa de probabilidade em (3.1) é dada por

$$\pi_B(y_t | \mu_t, v, \mathcal{F}_t) = \frac{\binom{m}{y_t}^v \left(\frac{\mu_t}{m - \mu_t} \right)^{y_t}}{\left(\frac{m}{m - \mu_t} \right)^m},$$

sua média condicional é $\mu_t^B = \mu_t$ e sua variância (3.3) é

$$v_t^B = \frac{\frac{\mu_t}{m - \mu_t}}{m} = \frac{\mu_t(m - \mu_t)}{m(m - \mu_t)^2}.$$

O modelo PS-GARMA cujos dados Y_t seguem uma distribuição Binomial é denominado Binomial GARMA (B-GARMA) e seu equivalente zero-modificado é chamado Binomial Zero-Modificado GARMA (ZMB-GARMA). Para os modelos ZMB-GARMA a média condicional é da forma $\mu_t^{ZMB} = p_t \mu_t$ e a variância condicional e, (2.6) é

$$v_t^{ZMB} = p_t \left[\frac{\mu_t(m - \mu_t)}{m} + (1-p_t)(\mu_t)^2 \right].$$

3.1.5 Distribuição Binomial Negativa

A distribuição Binomial Negativa (NB) é uma distribuição discreta que modela o número de tentativas para obter um determinado número de sucessos em uma sequência de experimentos de Bernoulli independentes, todos com a mesma probabilidade de sucesso.

A forma da sua função massa de probabilidade em (3.1) fica

$$\pi_{NB}(y_t | \mu_t, \nu, \mathcal{F}_t) = \frac{\Gamma(\nu + y_t)}{y_t! \Gamma(\nu)} \left(\frac{\mu_t}{\mu_t + \nu} \right)^{y_t} \left(\frac{\nu}{\mu_t + \nu} \right)^{-\nu},$$

sua média condicional é $\mu_t^{NB} = \mu_t$ e sua variância condicional (3.3) fica

$$V_t^{NB} = \frac{\frac{\mu_t}{\mu_t + \nu}}{\frac{\nu}{(\mu_t + \nu)^2}} = \frac{\mu_t(\mu_t + \nu)}{\nu}.$$

Se os dados Y_t pertencem à distribuição NB, o modelo será denominado Binomial Negativo GARMA (NB-GARMA) e, no caso zero modificado, Binomial Negativo Zero-Modificado GARMA (ZMNB-GARMA). Para os modelos ZMNB-GARMA, a média condicional é dada por $\mu_t^{ZMNB} = p_t \mu_t$ e a variância condicional (2.6) fica

$$V_t^{ZMNB} = p_t \left[\frac{\mu_t(\mu_t + \nu)}{\nu} + (1 - p_t)(\mu_t)^2 \right].$$

3.2 Os modelos PIG-GARMA e ZMPIG-GARMA

A distribuição Poisson-Inversa-Gaussiana (PIG) foi proposta por [Holla \(1967\)](#) como uma alternativa para a distribuição de Poisson em casos de sobredispersão.

Se $Y_t | V$ segue uma distribuição de Poisson com média $\mu_t V$, tal que V segue uma distribuição Inversa Gaussiana com média 1 e parâmetro de dispersão $\frac{1}{\sigma}$, $\sigma > 0$, a função massa de probabilidade marginal para y_t é dada por

$$\pi_{PIG}(y_t | \mu_t, \sigma) = \frac{\mu_t^{y_t}}{y_t!} \sqrt{\frac{2}{\pi\sigma}} \exp\left(-\frac{1}{\sigma}\right) (1 + 2\sigma\mu_t)^{\frac{y_t}{2}} K_{S_t}(\Psi_t), \quad y_t \in \mathcal{A}_0, \quad (3.4)$$

tal que

$$S_t = y_t - \frac{1}{2},$$

$$\Psi_t = \frac{\sqrt{1 + 2\sigma\mu_t}}{\sigma}$$

e K_{S_t} é a função de Bessel modificada de segunda espécie dada por

$$K_{S_t}(\Psi_t) = \frac{1}{2} \int_0^\infty x^{S_t-1} \exp\left\{-\frac{\Psi_t}{2} \left(x + \frac{1}{x}\right)\right\} dx,$$

para $t = 1, 2, \dots, n$. Sua esperança é dada por $\mu_t^{PIG} = \mu_t$ e a variância por $v_t^{PIG} = \mu_t(\mu_t\sigma + 1)$.

Um modelo D-GARMA cujos dados Y_t seguem distribuição PIG é chamado Poisson-Inversa-Gaussiana GARMA (PIG-GARMA) e seu respectivo modelo ZMD-GARMA é o Poisson-Inversa-Gaussiana Zero-Modificado GARMA (ZMPIG-GARMA). Para os modelos ZMPIG-GARMA a média é dada por $\mu_t^{ZMPIG} = \mu_t$ e a variância em (2.6) por

$$v_t^{ZMPIG} = p_t \left[\mu_t(\mu_t\sigma + 1) + (1 - p_t)(\mu_t)^2 \right] = p_t\mu_t[1 + \mu_t(\sigma + 1 - p_t)].$$

3.3 Os modelos S-GARMA e ZMS-GARMA

A distribuição Skellam foi proposta por [Skellam \(1946\)](#), surgindo como a distribuição da diferença de duas variáveis aleatórias independentes que seguem distribuição Poisson com médias θ_1 e θ_2 , respectivamente. É uma generalização do problema considerado por [Irwin \(1937\)](#), que estudou a diferença de distribuições de Poisson com médias iguais.

Para um par de variáveis $(X_{1,t}, X_{2,t})$ tais que $X_{1,t} = W_{1,t} + W_{3,t}$ e $X_{2,t} = W_{2,t} + W_{3,t}$ tais que $W_{1,t}$ e $W_{2,t}$ são independentes, $W_{1,t}$ e $W_{2,t}$ tenham distribuição de Poisson com taxas $\theta_{1,t}$ e $\theta_{2,t}$, respectivamente, e $W_{3,t}$ siga uma distribuição qualquer, a função massa de probabilidade de $Y_t = X_{1,t} - X_{2,t}$ é dada por

$$\pi_s(y_t | \theta_{1,t}, \theta_{2,t}, \mathcal{F}_t) = e^{-(\theta_{1,t} + \theta_{2,t})} \left(\frac{\theta_{1,t}}{\theta_{2,t}} \right)^{\frac{y_t}{2}} I_{y_t} \left(2\sqrt{\theta_{1,t}\theta_{2,t}} \right), \quad y_t \in \mathbb{Z}, \quad (3.5)$$

tal que seu suporte é o conjunto dos números inteiros \mathbb{Z} , $I_{y_t}(x)$ é a função Bessel modificada de primeira espécie, dada por

$$I_{y_t}(x) = \left(\frac{x}{2} \right)^{y_t} \sum_{k=0}^{\infty} \frac{(x/2)^{2k}}{k!(y_t + k)!}, \quad y_t = 0, 1, 2, \dots,$$

para todo $y_t \in \mathbb{Z}$ tem-se $I_{-y_t}(x) = I_{y_t}(x)$.

Dizemos que Y_t segue uma distribuição Skellam com média $\mu_t = \theta_{1,t} - \theta_{2,t}$ e variância $v \geq \max_{t \in T} \mu_t$ se sua função massa de probabilidade é da forma (3.5). Uma parametrização em função da média μ_t e da variância $v \geq 0$ pode ser obtida como

$$\pi_s(y_t; \mu_t, v) = g(\mu_t, v)^{y_t} f(y_t, \mu_t, v), \quad (3.6)$$

tal que $f(y_t, \mu_t, v) = e^{-v} I_{y_t} \left(\sqrt{v^2 - \mu_t^2} \right)$ e $g(\mu_t, v) = \left(\frac{v + \mu_t}{v - \mu_t} \right)^{\frac{1}{2}}$. Da desigualdade triangular $|\theta_{1,t} - \theta_{2,t}| \leq \theta_{1,t} + \theta_{2,t}$ vem a restrição $|\mu_t| \leq v$.

Um modelo D-GARMA cujos dados Y_t seguem a distribuição Skellam será chamado Skellam GARMA (S-GARMA) e seu respectivo modelo ZMD-GARMA será o modelo Skellam Zero-Modificado GARMA (ZMS-GARMA). A média e a variância de Y_t são, respectivamente, $\mu_t^{ZMS} = p_t \mu_t^S$ e

$$v_t^{ZMS} = p_t \left[v_t^S + (1 - p_t) (\mu_t^S)^2 \right].$$

INFERÊNCIA PARA OS MODELOS

O paradigma Bayesiano é adotado para o ajuste dos modelos. A construção da função de verossimilhança é exibida na [Seção 4.1](#). A distribuição posterior é explicada na [Seção 4.2](#). A [Seção 4.3](#) detalha a construção do método Hamiltoniano de Monte Carlo com Gradiente Estocástico e a obtenção das respectivas estimativas é tratada na [Seção 4.4](#). Os critérios de seleção de modelos são explicados na [Seção 4.5](#). A [Seção 4.6](#) e a [Seção 4.7](#) mostram a construção das distribuições preditivas e dos intervalos de credibilidade, respectivamente.

4.1 Função log-verossimilhança

Seja $\mathbf{y}'_t = (y_1, y_2, \dots, y_t)$ uma amostra do processo aleatório Y_t . A informação sobre y_t disponível até o tempo t é denotada por $\mathcal{F}_t = (y_{t-1}, \dots, y_1, \mathbf{x}_t, \dots, \mathbf{x}_1, \mathbf{v}_t, \dots, \mathbf{v}_1), t = 1, \dots, n$, tal que $\mathbf{x}'_t = (x_{0t}, \dots, x_{mt})$ e $\mathbf{v}_t = (v_{0t}, \dots, v_{rt})$ são vetores de variáveis exógenas com $x_{0t} = v_{0t} = 1$. Seja $\mathcal{D}_n = (y_n, \dots, y_1, \mathbf{x}_n, \dots, \mathbf{x}_1, \mathbf{v}_n, \dots, \mathbf{v}_1)$ o conjunto de dados disponíveis no tempo n . Os parâmetros de interesse para a estimação são os vetores $\boldsymbol{\beta}' = (\beta_0, \beta_1, \dots, \beta_m), \boldsymbol{\phi}' = (\phi_1, \dots, \phi_p)$ e $\boldsymbol{\theta}' = (\theta_1, \dots, \theta_q)$ de (2.2), $\boldsymbol{\gamma}' = (\gamma_0, \gamma_1, \dots, \gamma_r)$ e $\boldsymbol{\lambda}' = (\lambda_1, \dots, \lambda_s)$ de (2.9) e o parâmetro de dispersão ν , dependendo da distribuição assumida.

O método da máxima verossimilhança é utilizado para obtenção das estimativas dos parâmetros dos modelos. A função de verossimilhança pode ser escrita como o produto das funções massa de probabilidade condicionais dadas em (2.8) (FOKIANOS; KEDEM, 2004), isso é,

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \nu \mid \mathcal{D}_n) = \prod_{t=p^*+1}^n (1 - w_t)^{I_{\{y_t=0\}}} \left[w_t \frac{\pi_D(y_t \mid \boldsymbol{\mu}_t \nu, \mathcal{F}_t)}{1 - \pi_D(0 \mid \boldsymbol{\mu}_t, \nu, \mathcal{F}_t)} \right]^{1 - I_{\{y_t=0\}}}, \quad (4.1)$$

tal que $0 < w_t < 1$, como definido em (2.9) e $p^* = \max\{p, q\}$. A função log-verossimilhança,

será, então

$$\begin{aligned}
\ell(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{v} \mid \mathcal{D}_n) &= \sum_{t=p^*+1}^n I_{\{y_t=0\}} \log(1-w_t) \\
&+ \sum_{t=p^*+1}^n (1-I_{\{y_t=0\}}) \log \left(w_t \frac{\pi_D(y_t \mid \boldsymbol{\mu}_t \mathbf{v}, \mathcal{F}_t)}{1-\pi_D(0 \mid \boldsymbol{\mu}_t, \mathbf{v}, \mathcal{F}_t)} \right) \\
&= \sum_{t=p^*+1}^n \left[I_{\{y_t=0\}} \log \left(\frac{1-w_t}{w_t} \right) + \log(w_t) \right] \\
&+ \sum_{t=p^*+1}^n (1-I_{\{y_t=0\}}) \log \left(\frac{\pi_D(y_t \mid \boldsymbol{\mu}_t \mathbf{v}, \mathcal{F}_t)}{1-\pi_D(0 \mid \boldsymbol{\mu}_t, \mathbf{v}, \mathcal{F}_t)} \right).
\end{aligned} \tag{4.2}$$

A log-verossimilhança pode ser separada em duas partes da forma

$$\ell_1(\boldsymbol{\gamma}, \boldsymbol{\lambda} \mid \mathcal{D}_n) = \sum_{t=p^*+1}^n \left[I_{\{y_t=0\}} \log \left(\frac{1-w_t}{w_t} \right) + \log(w_t) \right] \tag{4.3}$$

e

$$\ell_2(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{v} \mid \mathcal{D}_n) = \sum_{t=p^*+1}^n (1-I_{\{y_t=0\}}) \log \left(\frac{\pi_D(y_t \mid \boldsymbol{\mu}_t \mathbf{v}, \mathcal{F}_t)}{1-\pi_D(0 \mid \boldsymbol{\mu}_t, \mathbf{v}, \mathcal{F}_t)} \right). \tag{4.4}$$

De forma que a equação (4.2) pode ser reescrita como

$$\ell(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{v} \mid \mathcal{D}_n) = \ell_1(\boldsymbol{\gamma}, \boldsymbol{\lambda} \mid \mathcal{D}_n) + \ell_2(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{v} \mid \mathcal{D}_n).$$

As estimações dos parâmetros $(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ e $(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{v})$ podem ser obtidas de forma independente pelas equações (4.3) e (4.4). Dessa forma, dois vetores gradiente podem ser calculados, denotados

$$\mathbf{g}'_1 = \left(\frac{\partial \ell_1(\boldsymbol{\gamma}, \boldsymbol{\lambda} \mid \mathcal{D}_n)}{\partial \boldsymbol{\gamma}}, \frac{\partial \ell_1(\boldsymbol{\gamma}, \boldsymbol{\lambda} \mid \mathcal{D}_n)}{\partial \boldsymbol{\lambda}} \right)$$

e

$$\mathbf{g}'_2 = \left(\frac{\partial \ell_2(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{v} \mid \mathcal{D}_n)}{\partial \boldsymbol{\beta}}, \frac{\partial \ell_2(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{v} \mid \mathcal{D}_n)}{\partial \boldsymbol{\phi}}, \frac{\partial \ell_2(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{v} \mid \mathcal{D}_n)}{\partial \boldsymbol{\theta}}, \frac{\partial \ell_2(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{v} \mid \mathcal{D}_n)}{\partial \mathbf{v}} \right).$$

Os elementos de \mathbf{g}_1 podem ser calculados de forma geral para qualquer distribuição discreta como

$$\frac{\partial \ell_1(\cdot)}{\partial \gamma_j} = \sum_{t=p^*+1}^n [(1-I_{\{y_t=0\}}) \mathbf{v}_{tj} - w_t \mathbf{v}_{tj}], j = 0, \dots, r \tag{4.5}$$

e

$$\frac{\partial \ell_1(\cdot)}{\partial \lambda_j} = \sum_{t=p^*+1}^n [(1-I_{\{y_t=0\}}) \eta_{t-j} - w_t \eta_{t-j}], j = 0, \dots, r, \tag{4.6}$$

tal que $r, s \leq p^*$. Os elementos do vetor \mathbf{g}_2 dependem da distribuição adotada no modelo e são abordados a seguir.

4.1.1 Família Séries de Potência

Como mostrado em [Andrade, Conceição e Ravishanker \(2023\)](#) os elementos do vetor g_2 podem ser calculados de forma geral para qualquer parâmetro ξ_k por

$$\frac{\partial \ell_2(\cdot)}{\partial \xi_k} = \sum_{t=p^*+1}^n (1 - I_{\{y_t=0\}}) \left[y_t \frac{g_\mu(\mu_t, \nu)}{g(\mu_t, \nu)} - \frac{f_\mu(\mu_t, \nu)}{f(\mu_t, \nu) - a(0, \nu)} \right] \frac{\partial \mu_t}{\partial \xi_k}, \quad (4.7)$$

tal que $k = 1, \dots, K$, com $K = m + 1$ se ξ_k é um elemento do vetor $\boldsymbol{\beta}$, $K = p$ quando ξ_k é um elemento de $\boldsymbol{\phi}$, ou $K = q$ quando ξ_k pertence a $\boldsymbol{\theta}$. E, para o parâmetro ν ,

$$\frac{\partial \ell_2(\cdot)}{\partial \nu} = \sum_{t=p^*+1}^n (1 - I_{\{y_t=0\}}) \left[\frac{a_\nu(y_t, \nu)}{a(y_t, \nu)} + y_t \frac{g_\nu(\mu_t, \nu)}{g(\mu_t, \nu)} - \frac{f_\nu(\mu_t, \nu) - a_\nu(0, \nu)}{f(\mu_t, \nu) - a(0, \nu)} \right]. \quad (4.8)$$

Em (4.7) e (4.8), tem-se $g_\mu(\mu_t, \nu) = \frac{\partial g(\mu_t, \nu)}{\partial \mu_t}$, $f_\mu(\mu_t, \nu) = \frac{\partial f(\mu_t, \nu)}{\partial \mu_t}$, $a_\nu(y_t, \nu) = \frac{\partial a(y_t, \nu)}{\partial \nu}$, $g_\nu(\mu_t, \nu) = \frac{\partial g(\mu_t, \nu)}{\partial \nu}$ e $f_\nu(\mu_t, \nu) = \frac{\partial f(\mu_t, \nu)}{\partial \nu}$, com as funções exibidas na [Tabela 1](#). A forma explícita das derivadas e dos componentes do vetor gradiente g_2 para cada distribuição são apresentadas no [Apêndice A](#).

4.1.2 Distribuição Skellam

Quando os dados seguem a distribuição Skellam, os elementos do vetor g_2 podem ser calculados para qualquer parâmetro ξ_k de forma geral por

$$\frac{\partial \ell_2(\cdot)}{\partial \xi_k} = \sum_{t=p^*+1}^n (1 - I_{\{y_t=0\}}) \left[y_t \frac{g_\mu(\mu_t, \nu)}{g(\mu_t, \nu)} + \frac{f_\mu(y_t, \mu_t, \nu)}{f(y_t, \mu_t, \nu)} + \frac{f_\mu(0, \mu_t, \nu)}{1 - f(0, \mu_t, \nu)} \right] \frac{\partial \mu_t}{\partial \xi_k}, \quad (4.9)$$

tal que $k = 1, \dots, K$, com $K = m + 1$ se ξ_k é um elemento do vetor $\boldsymbol{\beta}$, $K = p$ quando ξ_k é um elemento de $\boldsymbol{\phi}$, ou $K = q$ quando ξ_k pertence a $\boldsymbol{\theta}$. E, para o parâmetro ν ,

$$\frac{\partial \ell_2(\cdot)}{\partial \nu} = \sum_{t=p^*+1}^n (1 - I_{\{y_t=0\}}) \left[y_t \frac{g_\nu(\mu_t, \nu)}{g(\mu_t, \nu)} + \frac{f_\nu(y_t, \mu_t, \nu)}{f(y_t, \mu_t, \nu)} + \frac{f_\nu(0, \mu_t, \nu)}{1 - f(0, \mu_t, \nu)} \right]. \quad (4.10)$$

As derivadas das funções $f(y_t, \mu_t, \nu)$ e $g(\mu_t, \nu)$ são exibidas em [Conceição, Suzuki e Andrade \(2020\)](#) da forma

$$\begin{aligned} f_\mu(y_t, \mu_t, \nu) &= -\frac{\mu_t e^{-\nu} I'_{y_t}(\sqrt{\nu^2 - \mu_t^2})}{\sqrt{\nu^2 - \mu_t^2}} \\ f_\nu(y_t, \mu_t, \nu) &= e^{-\nu} \left(\frac{\nu I'_{y_t}(\sqrt{\nu^2 + \mu_t^2})}{\sqrt{\nu^2 + \mu_t^2}} - I'_{y_t}(\sqrt{\nu^2 + \mu_t^2}) \right) \\ g_\mu(\mu_t, \nu) &= \frac{\nu}{\sqrt{(\nu + \mu_t)(\nu - \mu_t)^3}} e \\ g_\nu(\mu_t, \nu) &= -\frac{\mu_t}{\sqrt{(\nu + \mu_t)(\nu - \mu_t)^3}}, \end{aligned}$$

tal que $I'_{y_t}(x) = \frac{1}{2}(I_{y_t-1}(x) + I_{y_t+1}(x))$, sendo I_{y_t} a função de Bessel modificada, definida na Seção 3.3. Dessa forma, a expressão (4.9) fica

$$\frac{\partial \ell_2(\cdot)}{\partial \xi_k} = \sum_{t=p^*+1}^n \frac{1 - I_{\{y_t=0\}}}{\sqrt{v^2 - \mu_t^2}} \left[y_t \frac{v}{\sqrt{v^2 - \mu_t^2}} + \frac{-\mu_t I'_{y_t}(\sqrt{v^2 - \mu_t^2})}{I_{y_t}(\sqrt{v^2 - \mu_t^2})} \right. \\ \left. + \frac{-\mu_t e^{-v} I'_0(\sqrt{v^2 - \mu_t^2})}{1 - e^{-v} I_0(\sqrt{v^2 - \mu_t^2})} \right] \frac{\partial \mu_t}{\partial \xi_k}$$

e (4.10) é dada por

$$\frac{\partial \ell_2(\cdot)}{\partial v} = \sum_{t=p^*+1}^n (1 - I_{\{y_t=0\}}) \left[y_t \frac{-\mu_t}{v^2 - \mu_t^2} + \frac{v I'_{y_t}(\sqrt{v^2 - \mu_t^2})}{\sqrt{v^2 - \mu_t^2} I_{y_t}(\sqrt{v^2 - \mu_t^2})} - 1 \right. \\ \left. + \frac{e^{-v}}{1 - e^{-v} I_0(\sqrt{v^2 + \mu_t^2})} \left(\frac{v I'_0(\sqrt{v^2 - \mu_t^2})}{\sqrt{v^2 - \mu_t^2}} - I_0(\sqrt{v^2 - \mu_t^2}) \right) \right].$$

4.2 Abordagem Bayesiana

Sejam $\Theta' = (\beta', \phi', \theta', \gamma', \lambda', v)$ o vetor dos parâmetros de interesse e $\mathcal{L}(\Theta | \mathcal{D}_n)$ a função de verossimilhança condicionada aos dados \mathcal{D}_n para Θ como discutida na Seção 4.1. Seja $\pi_0(\Theta)$ a densidade a *priori* de Θ , pela regra de Bayes, sua densidade a *posteriori* condicional a \mathcal{D}_n é proporcional ao produto da *priori* com a função de verossimilhança (4.1), ou seja,

$$\pi(\Theta | \mathcal{D}_n) \propto \mathcal{L}(\Theta | \mathcal{D}_n) \pi_0(\Theta). \quad (4.11)$$

A distribuição a *priori* $\pi_0(\Theta)$ foi assumida independente entre os parâmetros de tal forma que a distribuição de cada parâmetro seja pouco informativa. Distribuições a *priori* normais multivariadas foram propostas para os parâmetros

$$\begin{aligned} \beta &\sim N(\mu_\beta, \sigma_\beta^2 \mathbf{I}_{(m+1)}), \\ \phi &\sim N(\mu_\phi, \sigma_\phi^2 \mathbf{I}_{(p)}), \\ \theta &\sim N(\mu_\theta, \sigma_\theta^2 \mathbf{I}_{(q)}), \\ \gamma &\sim N(\mu_\gamma, \sigma_\gamma^2 \mathbf{I}_{(r+1)}) \text{ e} \\ \lambda &\sim N(\mu_\lambda, \sigma_\lambda^2 \mathbf{I}_{(s)}), \end{aligned}$$

e uma distribuição log-normal para $v \sim LN(v_0, \sigma_v^2)$. \mathbf{I} denota a matriz identidade, $\mu_\beta, \mu_\phi, \mu_\theta, \mu_\gamma$ e μ_λ são vetores de hiperparâmetros de dimensões adequadas, v_0 é um escalar e $\sigma_\beta^2, \sigma_\phi^2, \sigma_\theta^2, \sigma_\gamma^2, \sigma_\lambda^2$ e σ_v^2 representam as variâncias das *priori*.

4.3 Métodos MCMC

A densidade *a posteriori* em (4.11) obtida da função de verossimilhança (4.1) discutida na Seção 4.1 e das distribuições *a priori* na Seção 4.2 não apresentam forma analítica convencional. Para modelos com distribuições *a posteriori* complexa, a amostragem é geralmente feita com a utilização de métodos Monte Carlo por Cadeia de Markov (MCMC).

4.3.1 Metropolis-Hastings

O algoritmo de Metropolis, proposto por [Metropolis et al. \(1953\)](#), foi o primeiro método MCMC amplamente utilizado para gerar cadeias de Markov que seguem uma distribuição $\pi(\Theta | \mathcal{D}_n)$. Uma extensão do algoritmo foi proposta por [Hastings \(1970\)](#), levando ao algoritmo de Metropolis-Hastings (MH).

O algoritmo MH gera uma sequência de valores de Θ que formam uma cadeia de Markov, de modo a aproximar a densidade *a posteriori* $\pi(\Theta | \mathcal{D}_n)$. Valores na cadeia $\Theta^{(t)}$ são indexados por $t = 0, 1, \dots, N$, tal que $\Theta^{(0)}$ é um valor inicial especificado.

O MH define uma probabilidade de transição que assegura que a cadeia de Markov seja ergódica e satisfaça condições de equilíbrio detalhado e reversibilidade. Tais condições são necessárias para garantir que a cadeia amostre de todo o suporte de Θ sem viés.

Valores de $\Theta^{(t)}$ na cadeia são definidos, em parte, por uma densidade proposta $q(\Theta' | \Theta^{(t-1)})$, na qual Θ' é o valor proposto para a próxima posição na cadeia. Várias funções proposta podem ser utilizadas no MH, sendo o passeio aleatório a mais utilizada.

Uma proposta é aceita com probabilidade

$$\alpha = \min \left\{ 1, \frac{\pi(\Theta' | \mathcal{D}_n) q(\Theta^{(t-1)} | \Theta')}{\pi(\Theta^{(t-1)} | \mathcal{D}_n) q(\Theta' | \Theta^{(t-1)})} \right\}, \quad (4.12)$$

se a proposta q é simétrica, a probabilidade simplifica para

$$\alpha = \min \left\{ 1, \frac{\pi(\Theta' | \mathcal{D}_n)}{\pi(\Theta^{(t-1)} | \mathcal{D}_n)} \right\},$$

que é a probabilidade de aceitação utilizada no algoritmo de Metropolis.

Como apontado por [Thomas e Tu \(2020\)](#), a probabilidade de aceitação α em (4.12) é um indicador importante da eficiência de um algoritmo MH. Se $\pi(\Theta' | \mathcal{D}_n) \geq \pi(\Theta^{(t-1)} | \mathcal{D}_n)$, o valor proposto $\pi(\Theta' | \mathcal{D}_n)$ é mais provável do que o valor anterior $\pi(\Theta^{(t-1)} | \mathcal{D}_n)$ pela função densidade e é sempre aceito. Quando $\pi(\Theta' | \mathcal{D}_n) < \pi(\Theta^{(t-1)} | \mathcal{D}_n)$, Θ' tem uma densidade menor do que $\Theta^{(t-1)}$ e é aceito aleatoriamente com probabilidade $\alpha \in (0, 1)$. Se o valor proposto Θ' não é aceito, ele é descartado e a cadeia continua no mesmo valor $\Theta^{(t)} \leftarrow \Theta^{(t-1)}$.

Dessa forma, o algoritmo se mantém em regiões de densidade *a posteriori* mais densas e passa por regiões de baixa densidade ocasionalmente. Dado que o algoritmo satisfaça condições

de ergodicidade e seja executado por um número suficiente de iterações, a distribuição empírica da cadeia MCMC é aproximadamente a verdadeira distribuição *a posteriori*, de forma que os valores gerados podem ser utilizados para estimação e inferência.

A estrutura geral do MH é exibida no [Algoritmo 1](#).

Algoritmo 1 – Metropolis-Hastings

```

1: procedimento METROPOLIS-HASTINGS( $\Theta^{(0)}, N$ ) ▷ Estado inicial e quantidade de iterações
2:   para  $t$  de 1 até  $N$  faça
3:      $\Theta^{(t)} \leftarrow q(\Theta' | \Theta^{(t-1)})$  ▷ Calcula o valor proposto
4:      $u \leftarrow U(0, 1)$ 
5:      $\alpha \leftarrow \min \left\{ 1, \frac{\pi(\Theta')q(\Theta^{(t-1)} | \Theta')}{\pi(\Theta^{(t-1)})q(\Theta' | \Theta^{(t-1)})} \right\}$  ▷ Calcula a probabilidade de aceitação
6:     se  $\alpha < u$  então
7:        $\Theta^{(t)} \leftarrow \Theta'$  ▷ O valor proposto é aceito na cadeia
8:     senão
9:        $\Theta^{(t)} \leftarrow \Theta^{(t-1)}$  ▷ A cadeia permanece no estado anterior
10:    fim se
11:  fim para
12:  retorna Cadeia com os valores estimados
13: fim procedimento

```

O MH apresenta poucos requisitos teóricos e suas limitações são principalmente computacionais. Com propostas geradas aleatoriamente, geralmente precisa de um número grande de iterações para chegar em áreas de maior densidade *a posteriori* (THOMAS; TU, 2020). Para propostas geradas por passeio aleatório, Gelman, Gilks e Roberts (1997) demonstram a necessidade de definir a variância da proposta para que a taxa de aceitação fique em torno de 0.25, por mais que mesmo os algoritmos eficientes fiquem abaixo desse valor (THOMAS; TU, 2020).

4.3.2 Hamiltoniano de Monte Carlo

O algoritmo Hamiltoniano de Monte Carlo (HMC), proposto por Duane *et al.* (1987), melhora a eficiência do MH utilizando um gerador de propostas guiado que se aproveita do gradiente da *log-posteriori*, exibido na [Seção 4.1](#), para direcionar a cadeia para áreas de maior densidade da distribuição *a posteriori*.

Em sistemas Hamiltonianos, as posições horizontal e vertical são dadas pelo par (Θ, \mathbf{p}) , com estimação interessada no parâmetro Θ . O parâmetro \mathbf{p} , comumente chamado de momento, é uma quantidade auxiliar utilizada na simulação de Θ sob as equações Hamiltonianas. O parâmetro Θ segue a distribuição *a posteriori* $\pi(\Theta | \mathcal{D}_n)$ definida em (4.11) e o momento \mathbf{p} é gerado por uma distribuição paramétrica. O momento corresponde a dimensionalidade de Θ como um vetor de tamanho k .

A equação Hamiltoniana é escrita como

$$\mathcal{H}(\Theta, \mathbf{p}) = U(\Theta) + K(\mathbf{p}), \quad (4.13)$$

tal que $U(\Theta)$ é a energia potencial do sistema e $K(\mathbf{p})$ a energia cinética, com $\Theta, \mathbf{p} \in \mathbb{R}^k$.

Adotando $U(\Theta) = -\log[\pi(\Theta | \mathcal{D}_n)]$ garante-se que Θ gerado pela função Hamiltoniana obedeça à distribuição de interesse. Para o momento, tipicamente é assumido $\mathbf{p} \sim N_k(0, \mathbf{M})$, tal que \mathbf{M} é uma matriz de covariância previamente especificada e k a dimensão do vetor de parâmetros Θ . [Betancourt \(2017\)](#) discute as vantagens de se adotar a distribuição gaussiana para \mathbf{p} , mesmo em diferentes métricas, sobre as demais distribuições. Com tais especificações, (4.13) fica

$$\mathcal{H}(\Theta, \mathbf{p}) = -\log[\pi(\Theta | \mathcal{D}_n)] + \frac{1}{2} \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p}.$$

O HMC percorre trajetórias governadas pelas equações diferenciais de primeira ordem conhecidas como Equações Hamiltonianas, dadas pelo sistema

$$\begin{aligned} \frac{\partial \mathbf{p}}{\partial t} &= -\frac{\partial \mathcal{H}(\Theta, \mathbf{p})}{\partial \Theta} = -\frac{\partial \ell(\Theta)}{\partial \Theta} = \nabla_{\Theta} \log[\pi(\Theta | \mathcal{D}_n)] \\ \frac{d\Theta}{dt} &= \frac{\partial \mathcal{H}(\Theta, \mathbf{p})}{\partial \mathbf{p}} = \frac{\partial K(\mathbf{p})}{\partial \mathbf{p}} = \mathbf{M}^{-1} \mathbf{p}, \end{aligned} \quad (4.14)$$

tal que $\nabla_{\Theta} \log[\pi(\Theta | \mathcal{D}_n)]$ é o gradiente da log-verossimilhança, discutido na [Seção 4.1](#). Uma solução das equações Hamiltonianas é uma função que define a forma como Θ pode ser amostrado de (Θ, \mathbf{p}) . A aleatoriedade do HMC vem do momento \mathbf{p} e do valor de Θ escolhido.

Dessa forma, um passo crítico do HMC é resolver as equações Hamiltonianas em (4.14). Uma abordagem costumeira para a solução de equações diferenciais é o método de Euler, que aproxima a solução da função em cada tempo t por uma função discreta. Porém, como apontado por [Brooks et al. \(2011\)](#), o método acumula erro entre as iterações, o que se torna um problema para um número grande de amostras.

Uma alternativa de resolução é o método *leapfrog*, que modifica o método de Euler usando um passo de tamanho discreto ε individual para \mathbf{p} e Θ , com um passo completo ε em Θ entre dois meio-passos $\frac{\varepsilon}{2}$ para \mathbf{p} ,

$$\begin{aligned} \mathbf{p}\left(t + \frac{\varepsilon}{2}\right) &= \mathbf{p}(t) + \frac{\varepsilon}{2} \log[\pi(\Theta(t) | \mathcal{D}_n)], \\ \Theta(t + \varepsilon) &= \Theta(t) + \varepsilon \mathbf{M}^{-1} \mathbf{p}\left(t + \frac{\varepsilon}{2}\right) \quad \text{e} \\ \mathbf{p}(t + \varepsilon) &= \mathbf{p}\left(t + \frac{\varepsilon}{2}\right) + \frac{\varepsilon}{2} \log[\pi(\Theta(t + \varepsilon) | \mathcal{D}_n)]. \end{aligned} \quad (4.15)$$

Para o HMC, múltiplos passos L são necessários para se deslocar a proposta suficientemente para o próximo estado. Apesar de ser um método com alta precisão, ainda introduz pequenos erros que levam a transições viesadas no Hamiltoniano. Uma estratégia amplamente utilizada é adotar a transição no Hamiltoniano como o valor proposto no MH.

Dado um estado inicial (Θ_0, \mathbf{p}_0) , após L passos de integração tem-se a proposta do estado final da trajetória (Θ_L, \mathbf{p}_L) . Uma proposta reversível é obtida tomando o oposto do momento final, $(\Theta_L, \mathbf{p}_L) \rightarrow (\Theta_L, -\mathbf{p}_L)$, que leva à proposta

$$q(\Theta', \mathbf{p}' | \Theta_L, \mathbf{p}_L) = \delta(\Theta' - \Theta_L) \delta(\mathbf{p}' + \mathbf{p}_L),$$

de modo que se obtém a probabilidade de aceitação do Metropolis-Hastings

$$\begin{aligned} \alpha &= \min \left\{ 1, \frac{q(\Theta_0, \mathbf{p}_0 | \Theta_L, -\mathbf{p}_L) \pi(\Theta_L, -\mathbf{p}_L)}{q(\Theta_L, -\mathbf{p}_L | \Theta_0, \mathbf{p}_0) \pi(\Theta_L, \mathbf{p}_0)} \right\} \\ &= \min \left\{ 1, \frac{\delta(\Theta_L - \Theta_0) \delta(-\mathbf{p}_L + \mathbf{p}_L) \pi(\Theta_L, -\mathbf{p}_L)}{\delta(\Theta_0 - \Theta_0) \delta(\mathbf{p}_0 + \mathbf{p}_0) \pi(\Theta_L, \mathbf{p}_0)} \right\} \\ &= \min \left\{ 1, \frac{\pi(\Theta_L, -\mathbf{p}_L)}{\pi(\Theta_L, \mathbf{p}_0)} \right\} \\ &= \min \left\{ 1, \frac{\exp(-\mathcal{H}(\Theta_L, -\mathbf{p}_L))}{\exp(-\mathcal{H}(\Theta_0, \mathbf{p}_0))} \right\} \\ &= \min \{ 1, \exp(-\mathcal{H}(\Theta_L, -\mathbf{p}_L) + \mathcal{H}(\Theta_0, \mathbf{p}_0)) \} \\ &= \min \{ 1, \exp(U(\Theta_0) - U(\Theta_L) + K(\mathbf{p}_0) - K(-\mathbf{p}_L)) \}. \end{aligned}$$

A estrutura geral do HMC é exibida no [Algoritmo 2](#).

A eficácia do HMC depende diretamente da escolha dos parâmetros ε e L . Um ε pequeno gera propostas mais próximas e maior taxa de aceitação, mas exige uma quantidade de passos L maior para garantir que a trajetória εL seja grande o suficiente para mover o parâmetro Θ para um ponto distante na distribuição. Em contrapartida, εL muito grande pode fazer com que a simulação dê uma volta completa. O ajuste dos parâmetros é feito ajustando ε e L e observando a taxa de aceitação da cadeia, [Brooks et al. \(2011\)](#) demonstram que a taxa de aceitação ideal é em torno de 0.65. Outro modo de observar a adequação dos parâmetros é pelo *traceplot* da cadeia gerada e sua autocorrelação, cadeias que se movem muito devagar e com alta autocorrelação indicam εL insuficiente ([THOMAS; TU, 2020](#)). Alterações na matriz de covariância \mathbf{M} também podem ser feitas.

Com o objetivo de eliminar a necessidade de ajuste do parâmetro L no HMC, [Hoffman e Gelman \(2011\)](#) propõem o algoritmo *No-U-Turn Sampler* (NUTS), que gera um conjunto de candidatos prováveis para o próximo estado da cadeia até começar a retroceder. O NUTS pode ser utilizado na linguagem de programação probabilística STAN como seu amostrador padrão ([CARPENTER et al., 2017](#)).

4.3.3 Hamiltoniano de Monte Carlo com Gradiente Estocástico

[Chen, Fox e Guestrin \(2014\)](#) estudaram a combinação da eficiência da exploração do espaço de estados do HMC com a eficiência computacional do gradiente estocástico para grandes volumes de dados e propuseram o algoritmo Hamiltoniano de Monte Carlo com Gradiente Estocástico (SGHMC).

Algoritmo 2 – Hamiltoniano de Monte Carlo

```

1: procedimento HAMILTONIANO DE MONTE CARLO( $\Theta^{(0)}, N, \varepsilon, L$ )      ▷ Estado inicial,
   quantidade de iterações, tamanho do passo e quantidade de passos.
2:   para  $t$  de 1 até  $N$  faça
3:      $\mathbf{p}(t) \leftarrow N(0, \mathbf{M})$ 
4:      $\Theta(t) \leftarrow \Theta^{(t-1)}$ 
5:     para  $i$  de 1 até  $L$  faça                                     ▷ Iterações do leapfrog
6:        $\mathbf{p}\left(t + \frac{\varepsilon}{2}\right) = \mathbf{p}(t) + \frac{\varepsilon}{2} \log[\pi(\Theta(t) | \mathcal{D}_n)]$ 
7:        $\Theta(t + \varepsilon) = \Theta(t) + \varepsilon \mathbf{M}^{-1} \mathbf{p}\left(t + \frac{\varepsilon}{2}\right)$ 
8:        $\mathbf{p}(t + \varepsilon) = \mathbf{p}\left(t + \frac{\varepsilon}{2}\right) + \frac{\varepsilon}{2} \log[\pi(\Theta(t + \varepsilon) | \mathcal{D}_n)]$ 
9:     fim para
10:     $\Theta' \leftarrow \Theta(t)$                                        ▷ A proposta recebe o valor ao final do leapfrog
11:     $\mathbf{p}' \leftarrow -\mathbf{p}(t)$                                        ▷ Toma o oposto do momento ao final da trajetória
12:     $u \leftarrow U(0, 1)$ 
13:     $\alpha \leftarrow \min\{1, \exp(U(\Theta^{(0)}) - U(\Theta') + K(\mathbf{p}(t)) - K(-\mathbf{p}'))\}$   ▷ Aceitação do HMC
14:    se  $u < \alpha$  então
15:       $\Theta^{(t)} \leftarrow \Theta'$                                        ▷ O valor proposto é aceito na cadeia
16:    senão
17:       $\Theta^{(t)} \leftarrow \Theta^{(t-1)}$                                    ▷ A cadeia permanece no valor anterior
18:    fim se
19:  fim para
20:  retorna Cadeia com os valores estimados
21: fim procedimento

```

Para não calcular o gradiente da energia potencial $\nabla U(\Theta)$, que precisa de todos os dados \mathcal{D}_n , uma estimação imprecisa é considerada, baseada em um subconjunto $\tilde{\mathcal{D}}_n$ amostrado uniformemente de \mathcal{D}_n , da forma

$$\nabla \tilde{U}(\Theta) = -\frac{|\mathcal{D}_n|}{|\tilde{\mathcal{D}}_n|} \sum_{y_i \in \tilde{\mathcal{D}}_n} \nabla \log[\pi(\Theta | \mathcal{D}_n)] - \nabla \log[\pi(\Theta)], \quad \tilde{\mathcal{D}}_n \subset \mathcal{D}_n, \quad (4.16)$$

que pode ser aproximado por

$$\nabla \tilde{U}(\Theta) \approx \nabla U(\Theta) + N(0, V(\Theta)),$$

tal que $V(\Theta)$ é a covariância do ruído do gradiente estocástico.

Substituindo $\nabla U(\Theta)$ por $\nabla \tilde{U}(\Theta)$ em (4.13), um ruído é adicionado à atualização do momento no sistema (4.14) da forma

$$\Delta \mathbf{p} = -\varepsilon \nabla(\Theta) = -\varepsilon \nabla U(\Theta) + N(0, \varepsilon^2 V(\Theta)),$$

de forma que o mesmo se torna

$$\begin{aligned} \frac{\partial \mathbf{p}}{\partial t} &= -\nabla_{\Theta} \log[\pi(\Theta | \mathcal{D}_n)] + N(0, 2\mathbf{B}(\Theta)) \\ \frac{d\Theta}{dt} &= \mathbf{M}^{-1} \mathbf{p}, \end{aligned}$$

com $\mathbf{B} = \frac{1}{2}V(\Theta)$. A modificação introduz um erro que pode ser corrigido com passos de MH, que geralmente é computacionalmente custoso. Para contornar tal problema, um componente de atrito é adicionado ao sistema, tornando-o

$$\begin{aligned}\frac{\partial \mathbf{p}}{\partial t} &= -\nabla_{\Theta} \log[\pi(\Theta | \mathcal{D}_n)] - \mathbf{B}(\Theta)\mathbf{M}^{-1}\mathbf{p} + N(0, 2\mathbf{B}(\Theta)) \\ \frac{d\Theta}{dt} &= \mathbf{M}^{-1}\mathbf{p}.\end{aligned}\tag{4.17}$$

As dinâmicas em (4.17) possuem propriedades de invariância similares ao do sistema Hamiltoniano original em (4.14), como demonstrado por [Chen, Fox e Guestrin \(2014\)](#).

O ruído \mathbf{B} do modelo é geralmente desconhecido e só se conhece uma estimativa $\hat{\mathbf{B}}$. Um fator de atrito \mathbf{C} arbitrário pode ser considerado, tal que $\mathbf{C} \succeq \hat{\mathbf{B}}$, e a dinâmica considerada para a simulação fica da forma

$$\begin{aligned}\frac{\partial \mathbf{p}}{\partial t} &= -\nabla_{\Theta} \log[\pi(\Theta | \mathcal{D}_n)] - \mathbf{C}\mathbf{M}^{-1}\mathbf{p} + N(0, 2(\mathbf{C} - \hat{\mathbf{B}})) + N(0, 2\mathbf{B}(\Theta)) \\ \frac{d\Theta}{dt} &= \mathbf{M}^{-1}\mathbf{p},\end{aligned}\tag{4.18}$$

removendo a necessidade de passos MH.

Em situações práticas, uma estimativa para $\hat{\mathbf{B}}$ é necessária. Sob uma escolha realista de $\hat{\mathbf{B}} = 0$, a atualização do momento é simplificada para

$$\frac{\partial \mathbf{p}}{\partial t} = -\nabla_{\Theta} \log[\pi(\Theta | \mathcal{D}_n)] - \mathbf{C}\mathbf{M}^{-1}\mathbf{p} + N(0, 2\mathbf{C}),$$

de forma que as dinâmicas são governadas pelo ruído controlado $N(0, 2\mathbf{C})$ e o termo de atrito $\mathbf{C}\mathbf{M}^{-1}$.

O sistema em (4.18) não pode ser resolvido analiticamente e requer métodos numéricos. Tomando o passo de integração de tamanho ε , as equações (4.18) podem ser reescritas de forma discretizada como

$$\begin{aligned}\Delta\Theta &= \varepsilon\mathbf{M}^{-1}\mathbf{p} \\ \Delta\mathbf{p} &= \varepsilon\nabla_{\Theta} \log[\pi(\Theta | \mathcal{D}_n)] - \alpha\mathbf{p} + N(0, 2\alpha\mathbf{M}),\end{aligned}\tag{4.19}$$

tal que $\alpha = \varepsilon\mathbf{C}\mathbf{M}^{-1}$. As constantes α e ε são termos de ajuste do algoritmo e α costuma ser fixado em valores pequenos. Assim como no HMC, o método *leapfrog* é utilizado para solucionar (4.19).

A estrutura do SGHMC com o integrador *leapfrog* é exibida no [Algoritmo 3](#).

Apesar de não precisar de passos MH para atingir a distribuição de interesse, a performance do SGHMC é sensível aos parâmetros ε e ao número de passos L , assim como o HMC.

Algoritmo 3 – Hamiltoniano de Monte Carlo com Gradiente Estocástico

-
- 1: **procedimento** HAMILTONIANO DE MONTE CARLO COM GRADIENTE ESTOCÁSTICO($\Theta^{(0)}, \varepsilon, \alpha, N, L$) \triangleright Estado inicial, tamanho do passo, constante de atrito, quantidade de iterações e quantidade de passos.
 - 2: **para** t de 1 até N **faça**
 - 3: $\mathbf{p}^{(t)} \leftarrow N(0, \mathbf{M})$ \triangleright Momento da amostra
 - 4: $\mathbf{p}_0 \leftarrow (1 - \alpha)\mathbf{p}^{(t)} + \frac{\varepsilon}{2}\nabla_{\Theta} \log[\pi(\Theta | \mathcal{D}_n)] + N(0, 2\alpha\mathbf{M})$
 - 5: $(\Theta_0, \mathbf{p}_0) \leftarrow (\Theta^{(t-1)}, \mathbf{p}^{(t)})$
 - 6: **para** i de 1 até L **faça** \triangleright Iterações da dinâmica em (4.19)
 - 7: $\Theta_i \leftarrow \Theta_{i-1} + \varepsilon_i \mathbf{M}^{-1} \mathbf{p}_{i-1}$
 - 8: **se** i diferente de L **então**
 - 9: $\mathbf{p}_i \leftarrow (1 - \alpha)\mathbf{p}_{i-1} + \varepsilon \nabla_{\Theta} \log[\pi(\Theta | \mathcal{D}_n)] + N(0, 2\alpha\mathbf{M})$
 - 10: **fim se**
 - 11: $\mathbf{p}_i \leftarrow (1 - \alpha)\mathbf{p}_{i-1} + \frac{\varepsilon}{2}\nabla_{\Theta} \log[\pi(\Theta | \mathcal{D}_n)] + N(0, 2\alpha\mathbf{M})$
 - 12: **fim para**
 - 13: $\Theta' \leftarrow \Theta_L$ \triangleright A proposta recebe o valor ao final da simulação
 - 14: **fim para**
 - 15: **retorna** Cadeia com os valores estimados
 - 16: **fim procedimento**
-

4.4 Estimativas Posteriores de Monte Carlo

Seja $\Theta' = (\Theta_1, \dots, \Theta_G)$ a cadeia simulada da distribuição posterior $\pi(\Theta | \mathcal{D}_n)$ pelo SGHMC. Seja $\Theta_j^{(g)} \in \Theta_g$, com $g = 1, \dots, G$, as amostras obtidas para cada elemento Θ_j do vetor de parâmetros do modelo. Como a esperança condicional para uma função $\psi(\Theta) | \mathcal{D}_n$ pode ser aproximada por

$$E(\psi(\Theta_j) | \mathcal{D}_n) \approx \frac{1}{G} \sum_{g=1}^G \psi(\Theta_j^{(g)}),$$

as estimativas para a média e a variância posterior de $\Theta_j, j = 1, \dots, k$, são calculadas por

$$E(\Theta_j | \mathcal{D}_n) \approx \hat{\Theta}_j = \frac{1}{G} \sum_{g=1}^G \Theta_j^{(g)}$$

e

$$V(\Theta_j | \mathcal{D}_n) \approx \sigma_{\Theta_j}^2 = \frac{1}{G} \sum_{g=1}^G \left(\Theta_j^{(g)} - \hat{\Theta}_j \right)^2$$

4.5 Seleção de Modelos

[Akaike \(1974\)](#) sugere o uso de critérios de informação no lugar de testes de hipóteses para a seleção de modelos, principalmente no contexto de séries temporais, e introduz o Critério de Informação de Akaike (AIC). No contexto Bayesiano, modelos estatísticos podem ser comparados de diversas formas, como por seu desvio, sua precisão preditiva e o valor da sua verossimilhança.

Três critérios foram adotados e comparados para a seleção de modelos ZMD-GARMA: o critério da Ordinada Preditiva Condicional (CPO) ligado à previsão preditiva, o Critério de Informação de Ampla Aplicação (WAIC) que se relaciona ao desvio e penalização por complexidade e o Critério de Informação Bayesiano Esperado (EBIC) referente à verossimilhança marginal aproximada.

O critério CPO, introduzido por [Gelfand, Dey e Chang \(1992\)](#), proporciona uma abordagem de validação cruzada para a seleção de modelos e pode ser estimado pelo resultado do MCMC da forma

$$\widehat{CPO}_t = \left[\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi_{ZMD}(y_t | \Theta^{(g)}, \mathcal{F}_t)} \right]^{-1}.$$

Em muitas situações, é numericamente vantajoso calcular o $\log(\widehat{CPO}_t)$ e nesse caso tem-se $\log(\widehat{CPO}) = \sum_{t=1}^T \log(\widehat{CPO}_t)$. O modelo com maior valor de $\log(\widehat{CPO})$, ou \widehat{CPO} , é tomado como o melhor.

O critério WAIC, introduzido por [Watanabe \(2012\)](#), é estimado utilizando a cadeia resultante do MCMC por

$$\widehat{WAIC} = \sum_{t=p^*+1}^n -2 \log \left(\frac{1}{G} \sum_{g=1}^G \pi_{ZMD}(y_t | \Theta^{(g)}, \mathcal{F}_t) \right) - p_{WAIC},$$

tal que

$$p_{WAIC} = \sum_{t=p^*+1}^n V_{g=1}^G \log \pi_{ZMD}(y_t | \Theta^{(g)}, \mathcal{F}_t),$$

com $V_{g=1}^G$ sendo a variância amostral, $V_{g=1}^G a_g = \frac{1}{G-1} \sum_{g=1}^G (a_g - \bar{a})^2$.

[Schwarz \(1978\)](#) propôs o Critério de Informação Bayesiano (BIC) como uma modificação do AIC para seleção de modelos no contexto Bayesiano. O critério EBIC, introduzido por [Carlin e Louis \(2000\)](#), é a esperança do BIC, que pode ser estimado por

$$\widehat{EBIC} = \frac{1}{G} \sum_{g=1}^G -2 \log(\mathcal{L}(\Theta^{(g)} | \mathcal{D}_n)) + p \log(n),$$

em que $\mathcal{L}(\Theta^{(g)} | \mathcal{D}_n)$ é a função de verossimilhança (4.1) avaliada utilizando a cadeia produzida no MCMC, p é o número de parâmetros do modelo e n é o número total de observações. Quanto menor for o \widehat{EBIC} melhor será o modelo.

4.6 Distribuições Preditivas

Seja $\mathcal{H}_{n+h} = \mathcal{D}_n \cup (\mathbf{x}_{n+1}, \dots, \mathbf{x}_{n+h}, \mathbf{v}_{n+1}, \dots, \mathbf{v}_{n+h})$, para fazer previsões de valores futuros Y_{n+h} , $h = 1, \dots, H$ dado \mathcal{H}_{n+h} , é necessário encontrar a função de densidade preditiva

$\pi_{ZMD}(y_{n+h} | \mathcal{H}_{n+h})$ e usar as funções de ligação (2.2) e (2.9) para prever as observações futuras $\hat{y}_{n+h} = E(Y_{n+h} | \mathcal{H}_{n+h})$. Tais previsões são calculadas recursivamente por

$$E(Y_{n+h} | \mathcal{H}_{n+h}) = \sum_{y_{n+h}=0}^{\infty} y_{n+h} \pi_{ZMD}(y_{n+h} | \mathcal{H}_{n+h})$$

tal que

$$\pi_{ZMD}(y_{n+h} | \mathcal{H}_{n+h}) = \int_{\Omega_{\Theta}} \pi_{ZMD}(y_{n+h} | \Theta, \mathcal{H}_{n+h}) \pi(\Theta | \mathcal{D}_n) d\Theta. \quad (4.20)$$

Combinando as equações anteriores e assumindo que $\pi_{ZMD}(y_{n+h} | \Theta, \mathcal{H}_{n+h})$ e $\pi(\Theta | \mathcal{D}_n)$ são ambas integráveis, vem

$$\begin{aligned} E(Y_{n+h} | \mathcal{H}_{n+h}) &= \sum_{y_{n+h}=0}^{\infty} y_{n+h} \int_{\Omega_{\Theta}} \pi_{ZMD}(y_{n+h} | \Theta, \mathcal{H}_{n+h}) \pi(\Theta | \mathcal{D}_n) d\Theta \\ &= \int_{\Omega_{\Theta}} \sum_{y_{n+h}=0}^{\infty} y_{n+h} \pi_{ZMD}(y_{n+h} | \Theta, \mathcal{H}_{n+h}) \pi(\Theta | \mathcal{D}_n) d\Theta \\ &= \int_{\Omega_{\Theta}} E(Y_{n+h} | \Theta, \mathcal{H}_{n+h}) \pi(\Theta | \mathcal{D}_n) d\Theta. \end{aligned} \quad (4.21)$$

Mais detalhes e propriedades dessa abordagem são discutidas em Krüger *et al.* (2020).

Utilizando as amostras do MCMC, $\{\Theta^{(g)}, g = 1, \dots, G\}$, a estimativa de Monte Carlo para a distribuição preditiva $\pi_{ZMD}(y_{n+h} | \mathcal{H}_{n+h})$ em (4.20) é

$$\hat{\pi}_{ZMD}(y_{n+h} | \mathcal{H}_{n+h}) = \frac{1}{G} \sum_{g=1}^G \pi_{ZMD}(y_{n+h} | \Theta^{(g)}, \mathcal{H}_{n+h}), \quad (4.22)$$

tal que

$$\begin{aligned} \pi_{ZMD}(y_{n+h} | \Theta^{(g)}, \mathcal{H}_{n+h}) &= (1 - w_{n+h}^{(g)}) I_{\{y_{n+h}=0\}} \\ &\quad + \frac{w_{n+h}^{(g)} \pi_D(y_{n+h} | \mu_{n+h}^{(g)}, \mathbf{v}^{(g)}, \mathcal{H}_{n+h})}{1 - \pi_D(0 | \mu_{n+h}^{(g)}, \mathbf{v}^{(g)}, \mathcal{H}_{n+h})} (1 - I_{\{y_{n+h}=0\}}). \end{aligned} \quad (4.23)$$

De forma semelhante, a estimativa de Monte Carlo para $E(Y_{n+h} | \mathcal{H}_{n+h})$ em (4.21) pode ser calculada por

$$\hat{y}_{n+h} = \frac{1}{G} \sum_{g=1}^G E(Y_{n+h} | \Theta^{(g)}, \mathcal{H}_{n+h}). \quad (4.24)$$

Considerando (2.5), (2.6) e (2.7), $E(Y_{n+h} | \Theta^{(g)}, \mathcal{H}_{n+h})$ pode ser calculado como

$$E(Y_{n+h} | \Theta^{(g)}, \mathcal{H}_{n+h}) = \frac{w_{n+h}^{(g)}}{1 - \pi_D(0 | \mu_{n+h}^{(g)}, \mathbf{v}^{(g)}, \mathcal{H}_{n+h})} \mu_t^D(\mu_{n+h}^{(g)}), \quad (4.25)$$

que pode ser calculado recursivamente considerando (2.2) e (2.9), respectivamente, como

$$\begin{aligned} \psi(\mu_{n+h}^{(g)}) &= \eta_{n+h}^{(g)} = \hat{\mathbf{x}}'_{n+h} \boldsymbol{\beta}^{(g)} \\ &\quad + \sum_{j=1}^p \phi_j^{(g)} \{ \psi(\hat{y}_{n+h-j}) - \hat{\mathbf{x}}'_{n+h-j} \boldsymbol{\beta} \} + \sum_{j=1}^q \theta_j^{(g)} \{ \psi(\hat{y}_{n+h-j}) - \eta_{n+h-j} \} \end{aligned} \quad (4.26)$$

e

$$w_{n+h}^{(g)} = \frac{\exp \left\{ \hat{\mathbf{v}}'_{n+h} \boldsymbol{\gamma}^{(g)} \sum_{j=1}^s \lambda_i^{(g)} \hat{y}_{n+h-j} \right\}}{1 + \exp \left\{ \hat{\mathbf{v}}'_{n+h} \boldsymbol{\gamma}^{(g)} \sum_{j=1}^s \lambda_i^{(g)} \hat{y}_{n+h-j} \right\}}, \quad (4.27)$$

em que $\hat{y}_{n+h-j} = y_{n+h-j}$ e $\hat{\mu}_{n+h-j} = \mu_{n+h-j}$ se $h \leq j$, enquanto que para $h \geq j$, \hat{y}_{n+h-j} é dado por (4.24) e

$$\hat{\mu}_{n+h-j} = \frac{1}{G} \sum_{g=1}^G \psi^{-1}(\eta_{n+h-j}^{(g)}).$$

A equação anterior é coerente, pois $\hat{\mu}_{n+h-j}$ só depende das estimativas calculadas para os tempos $t' \leq n+h-j-1$ (ANDRADE; CONCEIÇÃO; RAVISHANKER, 2023). As variáveis exógenas $(\hat{x}_{n+h}, \dots, \hat{x}_{n+1}, \hat{\mathbf{v}}_{n+h}, \dots, \hat{\mathbf{v}}_{n+1})$ são assumidas conhecidas ou possíveis de ser preditas de forma independente.

Outra abordagem para a predição de séries temporais para dados inteiros é o uso da mediana condicional, produzindo previsões coerentes que preservam a estrutura inteira dos dados (FREELAND; MCCABE, 2004; JUNG; TREMAYNE, 2006). A mediana condicional é estimada por \mathcal{M} , tal que \mathcal{M} é o menor número no suporte \mathcal{A}_s da variável aleatória y_{t+h} que satisfaz

$$\sum_{y_{n+h} \leq \mathcal{M}} \hat{\pi}_{ZMS}(y_{n+h} | \mathcal{H}_{n+h}) \geq \frac{1}{2}.$$

4.7 Intervalos de Credibilidade

Chen e Shao (1999) exibem um método para calcular os intervalos de credibilidade para as previsões y_{n+h} utilizando os quantis $100\alpha\%$ e $100(1-\alpha)\%$ de (4.20). A região $100(1-\alpha)\%$ de Maior Densidade a Posteriori (HPD) para y_{n+h} é o subconjunto $C \in \mathcal{A}_s$ definido como

$$C = \{y_{n+h} : \pi_{ZMD}(y_{n+h} | \mathcal{H}_{n+h}) \geq k\},$$

tal que k é o menor valor no suporte \mathcal{A}_s que satisfaz

$$\sum_{y_{n+h} \geq k} \hat{\pi}_{ZMD}(y_{n+h} | \mathcal{H}_{n+h}) \geq 1 - \alpha.$$

As estimativas de Monte Carlo são utilizadas para estimar (4.20) e calcular a região HPD $100(1-\alpha)\%$. Definindo a função de probabilidade acumulada como

$$S_{n+h}^{(k)} = P(y_{n+h} \leq y_{n+h}^{(k)} | \mathcal{H}_{n+h}) = \alpha,$$

o limite inferior, denotado por LB, como $LB(S_{n+h}^{(k+1)}) = 1$ se $S_{n+h}^{(k+1)} \geq \alpha$, zero caso contrário, e o limite superior, denotado UB, como $UB(S_{n+h}^{(k+1)}) = 1$ se $S_{n+h}^{(k+1)} \leq (1-\alpha)$, zero caso contrário.

Os percentis de $100\alpha\%$ e $100(1 - \alpha)\%$ são representados por

$$y_{n+h,\alpha} = \min \left\{ y_{n+h}^{(\rho)} \mid \sum_{k=1}^{\rho} \hat{\pi}_{ZMD}(y_{n+h}^{(k)} \mid \mathcal{H}_{n+h}) \geq \alpha \right\}$$

e

$$y_{n+h,(1-\alpha)} = \min \left\{ y_{n+h}^{(\rho)} \mid \sum_{k=1}^{\rho} \hat{\pi}_{ZMD}(y_{n+h}^{(k)} \mid \mathcal{H}_{n+h}) \geq 1 - \alpha \right\}.$$

O intervalo de credibilidade $100(1 - \alpha)\%$ para y_{n+h} é dado por

$$CI_{(1-\alpha)} = [y_{n+h,\alpha}; y_{n+h,(1-\alpha)}].$$

Um algoritmo para o cálculo do $100(1 - \alpha)\%$ HPD é exibido no [Algoritmo 4](#).

Algoritmo 4 – Intervalo de Credibilidade

- 1: **procedimento** INTERVALO DE CREDIBILIDADE($\{\hat{y}_{n+h}, h = 1, \dots, H\}, \Theta$) \triangleright Sequência de valores previstos, cadeia simulada pelo MCMC
 - 2: $h \leftarrow 1, k \leftarrow 0$
 - 3: $y_{n+h}^{(0)} \leftarrow \min\{10y_{\min}, 10y_{\max}\}, S_{n+h}^{(0)} \leftarrow 0, LB \leftarrow 0$ e $UB \leftarrow 0$ \triangleright Definindo valores iniciais.
 - 4: **enquanto** $LB \neq 1$ e $UB \neq 1$ **faça**
 - 5: Calcular $\mu_{n+h}^{(g)}$ e $w_{n+h}^{(g)}$ usando (4.26) e (4.27)
 - 6: $\hat{\pi}_{ZMD}(y_{n+h}^{(k)} \mid \mathcal{H}_{n+h}) \leftarrow \frac{1}{G} \sum_{g=1}^G \pi_{ZMD}(y_{n+h} \mid \Theta^{(g)}, \mathcal{H}_{n+h})$ utilizando (4.23)
 - 7: $S_{n+h}^{(k+1)} \leftarrow S_{n+h}^{(k)} + \hat{\pi}_{ZMD}(y_{n+h}^{(k)} \mid \mathcal{H}_{n+h})$
 - 8: **se** $LB = 0$ e $S_{n+h}^{(k+1)} \geq \alpha$ **então**
 - 9: $y_{n+h,\alpha} \leftarrow y_{n+h}^{(k)}$
 - 10: $LB \leftarrow 1$
 - 11: **senão se** $UB = 0$ e $S_{n+h}^{(k+1)} \leq (1 - \alpha)$ **então**
 - 12: $y_{n+h,(1-\alpha)} \leftarrow y_{n+h}^{(k)}$
 - 13: $UB \leftarrow 1$
 - 14: **senão se** $LB = 0$ ou $UB = 0$ **então**
 - 15: $k \leftarrow k + 1$
 - 16: $y_{n+h}^{(k)} \leftarrow y_{n+h}^{(k-1)} + 1$
 - 17: **fim se**
 - 18: **fim enquanto**
 - 19: **retorna** $y_{n+h,\alpha}$ e $y_{n+h,(1-\alpha)}$
 - 20: **fim procedimento**
-

ESTUDOS DE SIMULAÇÃO

Neste capítulo são apresentados os estudos de simulação para os modelos ZMD-GARMA. Na [Seção 5.1](#) é explicado o processo de simulação utilizado. A [Seção 5.2](#) apresenta os resultados de simulação para três séries dos modelos ZMPS-GARMA e a [Seção 5.3](#) apresenta os resultados para dois cenários dos modelos ZMS-GARMA.

5.1 Processo de Simulação

A eficácia da estimação Bayesiana pelo SGHMC foi avaliada por um estudo de simulação. Em cada caso, $Q = 100$ séries temporais foram simuladas, cada uma com tamanho $n = 200$. Em todos os casos, o componente de modificação *hurdle* (2.9) é da forma

$$\log\left(\frac{w_t}{1-w_t}\right) = \gamma_0 + \lambda_1 \eta_{t-1}. \quad (5.1)$$

Distribuições *a priori* normal padrão foram consideradas para os parâmetros β , ϕ , θ , γ e λ , como exibido na [Seção 4.2](#), e para a distribuição *a priori* log-normal de v foi definido $v_0 = 0$ e $\sigma_v^2 = 1$. De modo geral, o uso de variâncias maiores mostrou aumento no erro percentual dos parâmetros estimados.

Para fins de estimação, uma única cadeia de Markov foi gerada para cada uma das 100 séries simuladas. Dada a cadeia Θ resultante do método SGHMC, a média posterior $\hat{\Theta}_i$ é a estimativa de Θ_i como exibida em [Seção 4.4](#). Valores de *burn-in* e saltos são definidos para cada caso específico. A convergência individual das cadeias foi avaliada pelo critério de Geweke (GW) (GEWEKE, 1992) e a convergência geral do método MCMC utilizando três cadeias distintas através do critério de Gelman-Rubin (GR) (GELMAN; RUBIN, 1992), ambos disponíveis no R pela biblioteca CODA (PLUMMER *et al.*, 1999).

O desempenho das estimações foi avaliado pela Raiz do Erro Quadrático Médio (RMSE),

dado por

$$RMSE(\hat{\Theta}_i) = \sqrt{\frac{1}{Q} \sum_{j=1}^Q (\Theta_i - \hat{\Theta}_i^{(j)})^2}, \quad (5.2)$$

pela Média do Viés Absoluto (MAB), da forma

$$MAB(\hat{\Theta}_i) = \frac{1}{Q} \sum_{j=1}^Q |\Theta_i - \hat{\Theta}_i^{(j)}|, \quad (5.3)$$

e pelo Percentual Médio do Arcotangente do Erro Absoluto (MAAPE), proposto por [Kim e Kim \(2016\)](#) da forma

$$MAAPE(\hat{\Theta}_i) = \frac{100\%}{Q} \sum_{j=1}^Q \arctan \left| \frac{\Theta_i - \hat{\Theta}_i^{(j)}}{\Theta_i} \right|. \quad (5.4)$$

Devido à complexidade dos cálculos necessários e o consequente custo computacional do SGHMC, visto na [Seção 4.3](#), o processo de simulação foi paralelizado em 25 núcleos de processamento. Cada núcleo ficou, então, responsável por simular 4 séries temporais e fazer suas respectivas estimações via MCMC. O processo de paralelização no R foi feito utilizando as bibliotecas FOREACH ([WESTON, 2009](#)) e DOPARALLEL ([CORPORATION; WESTON, 2011](#)).

5.2 ZMPS-GARMA(p,q)

Todas as distribuições pertencentes à família PS podem utilizar a função de ligação logarítmica, isso é, $\log(\mu_t) = \eta_t$. Três séries foram simuladas para o estudo: MPS1 como um modelo Zero Deflacionado P-GARMA(1,0), MPS2 como Zero Inflacionada B-GARMA(1,1) e MPS3 sendo Zero Inflacionada NB-GARMA(2,0). O modelo MPS1 tem média da forma

$$\log(\mu_t) = \beta_0 + \beta_1 \log(t) + \phi_1 \log(y_{t-1}),$$

enquanto o modelo MPS2 tem a média dada por

$$\log(\mu_t) = \beta_0 + \phi_1 \log(y_{t-1}) + \theta_1 \log\left(\frac{y_{t-1}}{\mu_{t-1}}\right)$$

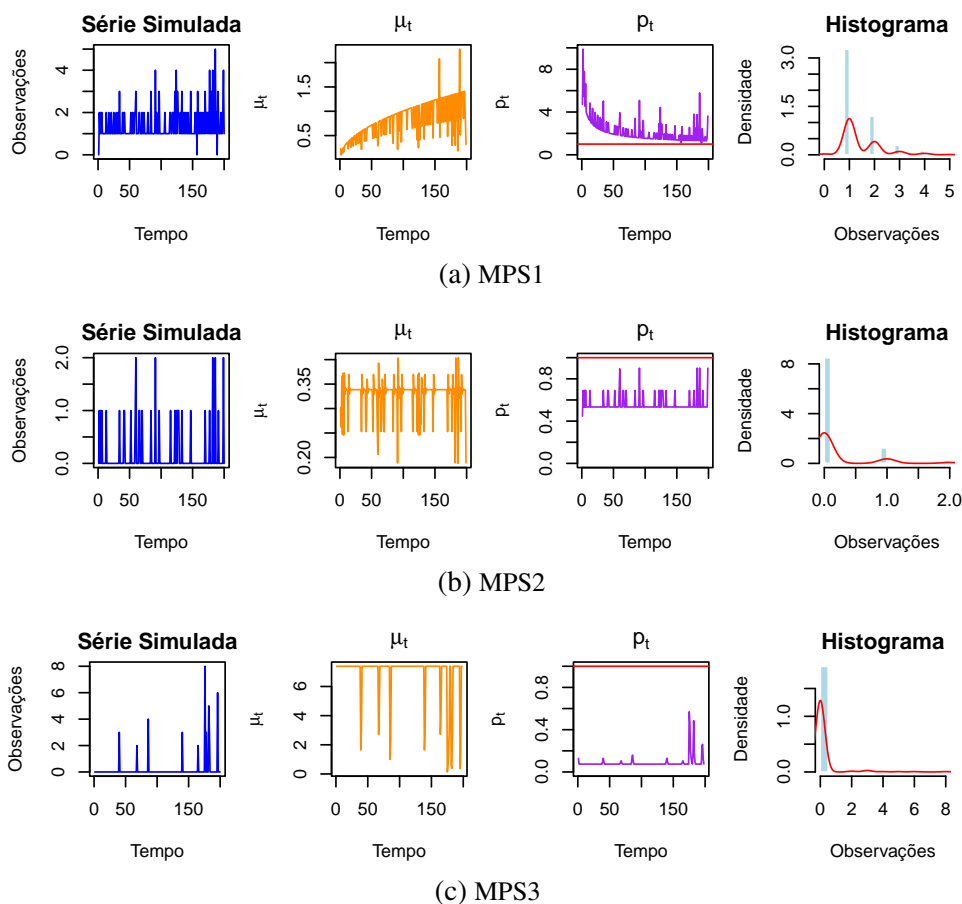
e a média do modelo MPS3 é

$$\log(\mu_t) = \beta_0 + \phi_1 \log(y_{t-1}) + \phi_2 \log(y_{t-2}).$$

Os gráficos da primeira repetição das séries simuladas, suas médias, valores de p_t e histogramas de cada modelo estão na [Figura 1](#). O número de tentativas m para o modelo MPS2 foi fixado em 5.

Em todos os casos, foram simuladas cadeias de 12000 iterações, das quais as 2000 primeiras foram descartadas em *burn-in* e saltos de 10 iterações foram feitos para diminuir a autocorrelação, resultando em 1000 amostras para o processo inferencial. A [Tabela 2](#) apresenta

Figura 1 – Gráficos das séries ZMPS-GARMA(p,q)



Fonte: Elaborada pelo autor.

Tabela 2 – Testes de convergência para as cadeias do SGHMC.

Θ	Valor z	\hat{R}	\hat{R}_{Θ_0}	Θ	Valor z	\hat{R}	\hat{R}_{Θ_0}	Θ	Valor z	\hat{R}	\hat{R}_{Θ_0}
β_0	0,76	1,00	1,00	β_0	0,63	1,00	1,00	β_0	2,10	1,00	1,00
β_1	-0,52	1,00	1,00	ϕ_1	-0,93	1,00	1,00	ϕ_1	0,42	1,00	1,00
ϕ_1	0,31	1,00	1,00	θ_1	0,14	1,00	1,01	θ_1	0,76	1,00	1,00
γ_0	1,03	0,99	1,00	γ_0	1,00	0,99	1,00	γ_0	1,68	1,00	1,00
λ_0	0,3	1,00	1,00	λ_0	0,46	0,99	1,00	λ_0	-0,23	0,99	1,00
								ν	0,73	1,00	1,00

(a) Modelo MPS1

(b) Modelo MPS2

(c) Modelo MPS3

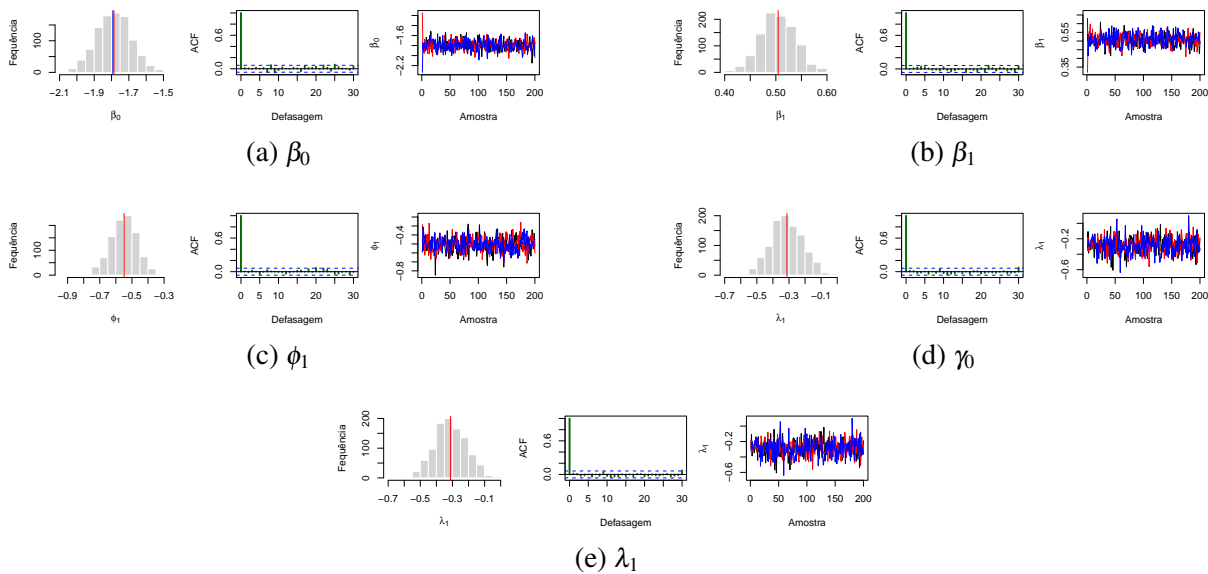
Fonte: Elaborada pelo autor.

os testes de convergência para as cadeias simuladas, todos os valores z para o teste GW são menores do que 2 em módulo, exceto para o parâmetro β_0 no modelo MPS3 (Tabela 2c), e todos os valores \hat{R} são próximos de 1, dando fortes indícios de convergência das cadeias.

Os gráficos do histograma e da autocorrelação após os processos de *burn-in* e espaçamento de uma repetição das cadeias para os modelos MPS1, MPS2 e MPS3 estão nas Figura 2, Figura 3 e Figura 4, respectivamente. Apesar de problemático no teste GW, o parâmetro β_0

aparenta ser convergente por análise gráfica (Figura 4a). Também foi conduzido um estudo de eficiência do amostrador tomando três valores diferentes para o chute inicial do SGHMC: o valor verdadeiro utilizado para gerar as amostras, Θ_0 , e $0.4\Theta_0$ e $1.4\Theta_0$. Todas as cadeias convergiram para a mesma região, como exibido no terceiro gráfico para cada parâmetro nas Figura 2, Figura 3 e Figura 4. Para fins de legibilidade e clareza gráfica, as cadeias apresentadas foram espaçadas sem *burn-in* para melhor representar sua trajetória e convergência. Todas as cadeias podem ser consideradas convergentes pelo critério GR, representado na Tabela 2 pelo valor \hat{R}_{Θ_0} .

Figura 2 – Gráficos das cadeias, histogramas, autocorrelações e convergências de cada parâmetro do modelo MPS1.



Fonte: Elaborada pelo autor.

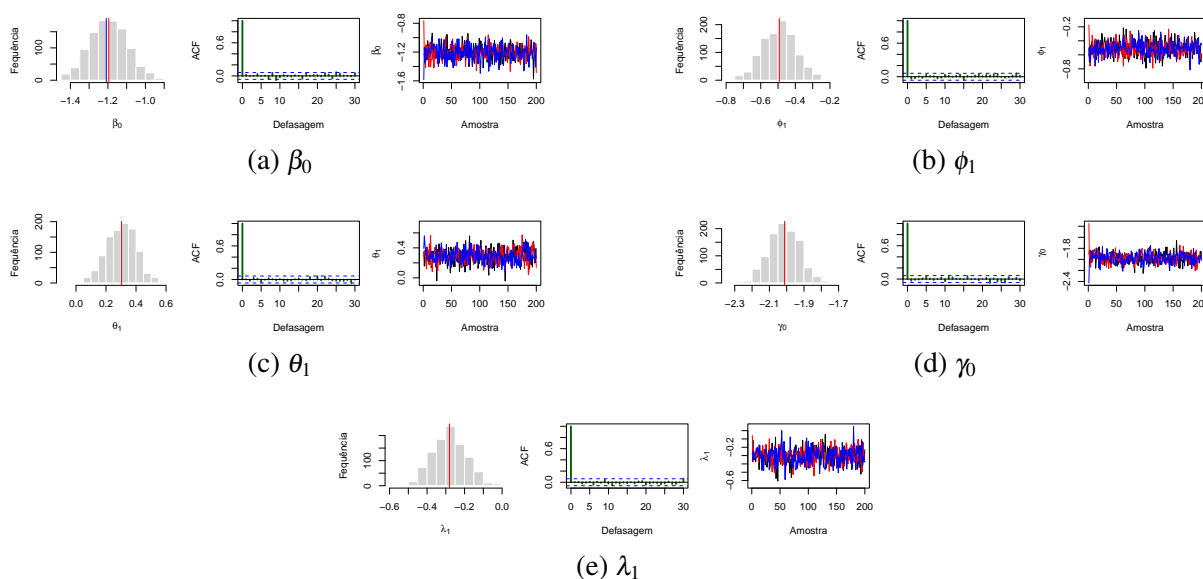
Verificadas a convergência das cadeias e a independência das amostras, o processo inferencial pôde ser realizado. Os resultados são exibidos na Tabela 3. Todos os valores estimados pela média da distribuição a *posteriori* são bem próximos aos valores verdadeiros com baixa incerteza, apresentando valores no desvio padrão de no máximo 0.05. Todos os intervalos de credibilidade contemplam o valor verdadeiro e todos os parâmetros têm probabilidade de cobertura de 100%. Apesar da menor precisão em alguns parâmetros apontado pelo MAAPE, os três modelos se mostraram robustos na estimação dos parâmetros.

5.3 ZMS-GARMA(p,q)

Como a função de ligação canônica da distribuição Skellam é a identidade, tem-se $\mu_t = \eta_t$. Duas séries foram simuladas para o estudo: MS1 como um modelo Zero Inflacionado S-GARMA(1,1) e MS2 como Zero Deflacionado S-GARMA(1,1). Ambos os modelos têm a média da forma

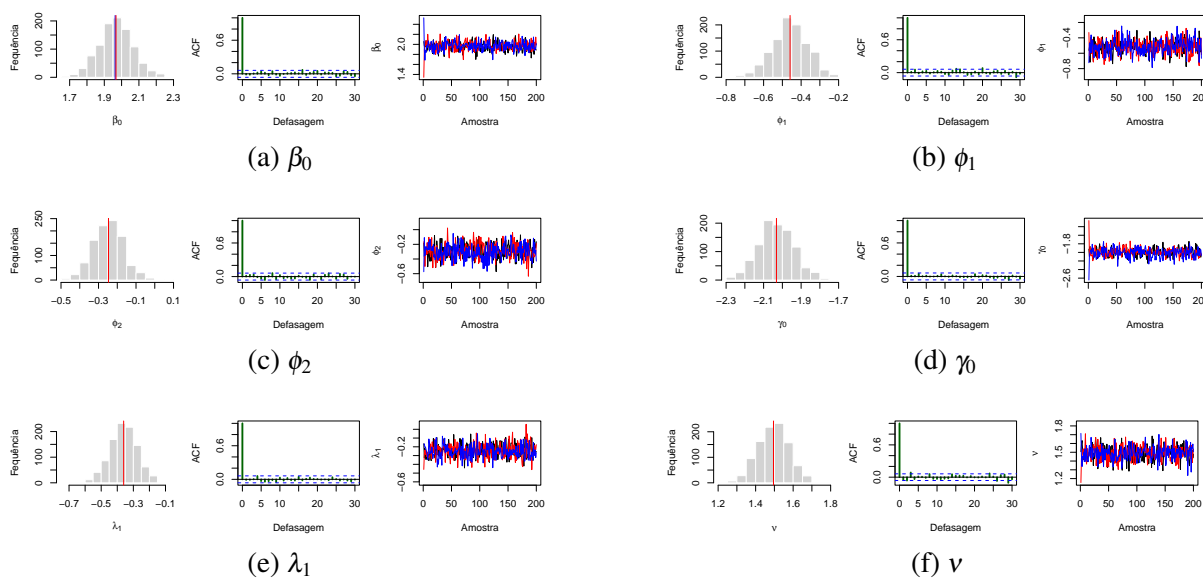
$$\mu_t = \beta_0 + \phi_1 y_{t-1} + \theta_1 \{y_{t-1} - \mu_{t-1}\}.$$

Figura 3 – Gráficos das cadeias, histogramas, autocorrelações e convergências de cada parâmetro do modelo MPS2.



Fonte: Elaborada pelo autor.

Figura 4 – Gráficos das cadeias, histogramas, autocorrelações e convergências de cada parâmetro do modelo MPS3.



Fonte: Elaborada pelo autor.

Os gráficos da primeira repetição das séries simuladas, suas médias, valores de p_t e histograma de cada modelo estão na [Figura 5](#).

Em ambos os casos, foram simuladas 42000 iterações, das quais as primeiras 2000 foram descartadas em *burn-in* e saltos de 40 iterações foram feitos para diminuir as autocorrelações, restando 1000 amostras para o processo de inferência. Os gráficos do histograma e a autocorrelação

Tabela 3 – Resultados de simulação para os modelos ZMPS-GARMA.

MPS1										
Θ	Valor	Média	DP	Mediana	2,5%	97,5%	MAB	RMSE	MAAPE	PC
β_0	-1,8	-1,797	0,015	-1,800	-1,985	-1,608	0,013	0,015	0,696	100
β_1	0,5	0,499	0,021	0,501	0,431	0,566	0,017	0,020	3,337	100
ϕ_1	-0,5	-0,508	0,042	-0,505	-0,673	-0,348	0,034	0,043	6,750	100
γ_0	4,0	4,000	0,019	4,002	3,808	4,190	0,016	0,019	0,389	100
λ_1	-0,3	-0,300	0,010	-0,301	-0,493	-0,107	0,007	0,010	2,390	100

MPS2										
Θ	Valor	Média	DP	Mediana	2,5%	97,5%	MAB	RMSE	MAAPE	PC
β_0	-1,2	-1,200	0,021	-1,201	-1,390	-1,010	0,017	0,021	1,396	100
ϕ_1	-0,5	-0,500	0,009	-0,502	-0,693	-0,307	0,007	0,009	1,319	100
θ_1	0,3	0,300	0,013	0,298	0,108	0,492	0,010	0,013	3,247	100
γ_0	-2,0	-2,002	0,031	-2,000	-2,181	-1,826	0,025	0,031	1,263	100
λ_1	-0,3	-0,295	0,034	-0,297	-0,467	-0,123	0,028	0,034	9,394	100

MPS3										
Θ	Valor	Média	DP	Mediana	2,5%	97,5%	MAB	RMSE	MAAPE	PC
β_0	2,0	1,998	0,031	1,995	1,819	2,178	0,024	0,031	1,218	100
ϕ_1	-0,5	-0,501	0,021	-0,504	-0,690	-0,314	0,014	0,021	2,808	100
ϕ_2	-0,3	-0,305	0,038	-0,312	-0,471	-0,151	0,030	0,038	9,908	100
γ_0	-2,0	-1,999	0,025	-1,999	-2,182	-1,817	0,020	0,025	1,002	100
λ_1	-0,3	-0,304	0,051	-0,300	-0,454	-0,159	0,042	0,051	13,610	100
ν	1,5	1,495	0,007	1,496	1,340	1,654	0,006	0,008	0,418	100

Fonte: Elaborada pelo autor.

ção após o processo de espaçamento entre as amostras para a primeira repetição das amostragens estão nas [Figura 6](#) e [Figura 7](#) para o MS1 e o MS2, respectivamente. A [Tabela 4](#) apresenta os testes de convergência para as cadeias simuladas, todos os valores z para o teste GW são menores do que 2 em módulo e todos os \hat{R} do teste GR são próximos de 1, indicando que todas as cadeias convergiram. O estudo de eficiência do amostrador para três valores diferentes do chute inicial, $0.4\Theta_0$ e $1.4\Theta_0$, também foi realizado para os modelos MS1 e MS2: todos os valores \hat{R}_{Θ_0} do teste GR foram próximos de 1, indicando que todas as cadeias convergiram para a mesma região.

Tabela 4 – Testes de convergência para as cadeias do SGHMC.

Θ	Valor z	\hat{R}	\hat{R}_{Θ_0}
β_0	0,75	1,01	1,00
ϕ_1	-1,29	1,00	1,00
θ_1	0,63	1,00	1,00
γ_0	-0,48	1,02	1,00
λ_1	0,17	1,00	1,00
ν	1,16	1,01	1,00

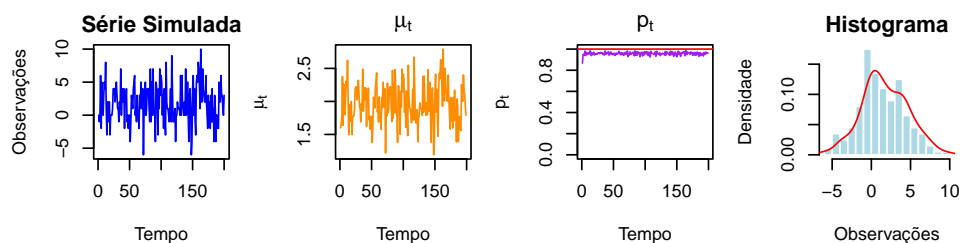
(a) Modelo MS1

Θ	Valor z	\hat{R}	\hat{R}_{Θ_0}
β_0	0,83	1,02	1,00
ϕ_1	-0,96	1,04	1,00
θ_1	1,41	1,03	1,00
γ_0	1,30	1,02	1,00
λ_1	-0,87	1,00	1,00
ν	-0,57	1,00	1,00

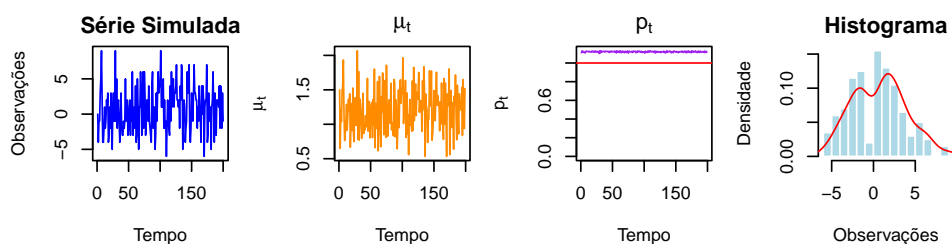
(b) Modelo MS2

Fonte: Elaborada pelo autor.

Figura 5 – Gráficos das séries ZMS-GARMA(p,q)



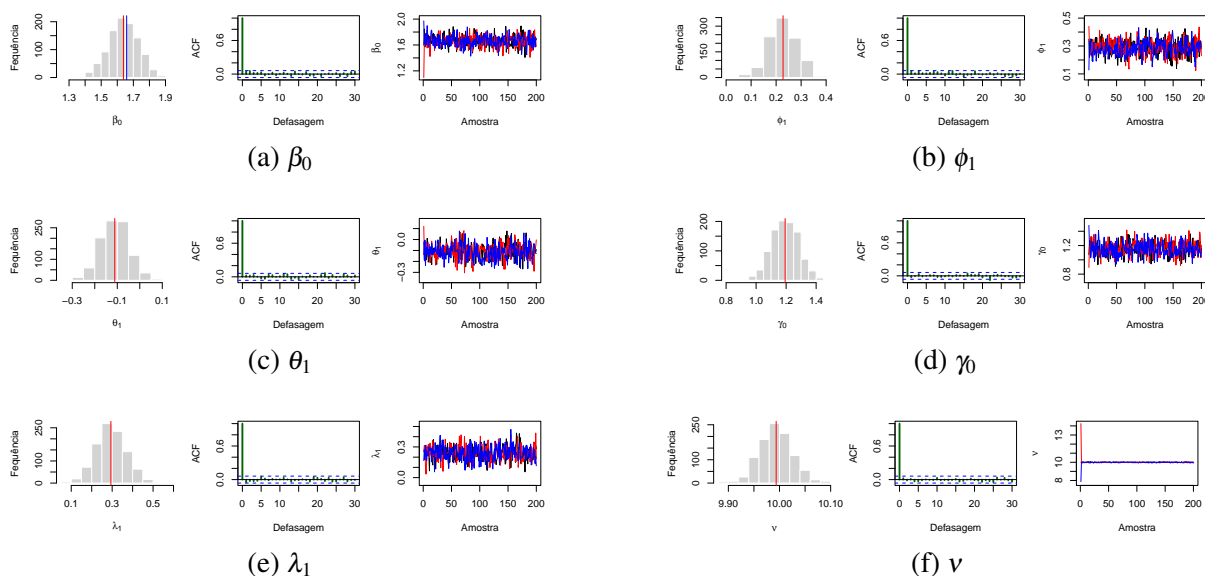
(a) MS1



(b) MS2

Fonte: Elaborada pelo autor.

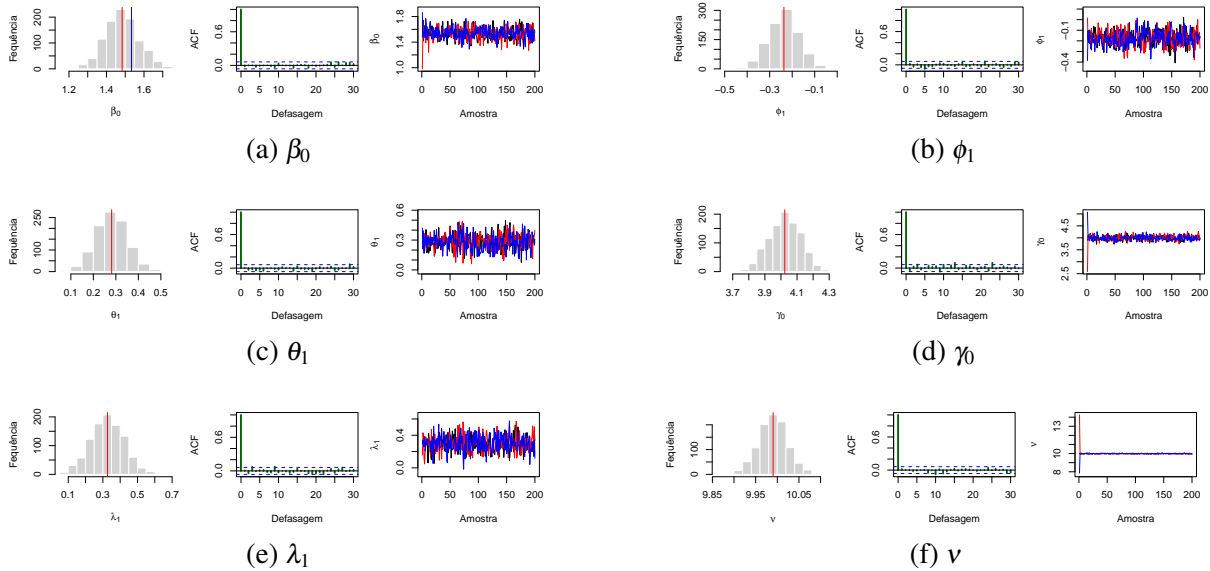
Figura 6 – Gráficos das cadeias, histogramas e autocorrelações de cada parâmetro do modelo MS1.



Fonte: Elaborada pelo autor.

Verificada a convergência das cadeias e obtida uma amostra com autocorrelação baixa, o processo de inferência pôde ser realizado. Os resultados são vistos na [Tabela 5](#). Em ambos os modelos a incerteza se mostra baixa, com desvio padrão de no máximo 0.05. Todos os intervalos de credibilidade contemplam o valor verdadeiro com no mínimo 96% de cobertura. Ambos os modelos apresentam bons resultados em termos de precisão, com valores baixos para o MAB, RMSE e MAAPE. O valor de MAAPE acentuado para θ_1 no MS2 pode ser atribuído ao valor verdadeiro absoluto muito pequeno.

Figura 7 – Gráficos das cadeias, histogramas e autocorrelações de cada parâmetro do modelo MS2.



Fonte: Elaborada pelo autor.

Tabela 5 – Resultados de simulação para os modelos ZMS-GARMA.

MS1										
Θ	Valor	Média	DP	Mediana	2,5%	97,5%	MAB	RMSE	MAAPE	PC
β_0	1,5	1,496	0,037	1,497	1,312	1,679	0,030	0,037	2,032	100
ϕ_1	-0,2	-0,196	0,046	-0,199	-0,332	-0,064	0,036	0,046	17,639	99
θ_1	0,3	0,309	0,040	0,309	0,164	0,458	0,033	0,041	10,843	100
γ_0	4,0	4,0004	0,015	4,001	3,810	4,192	0,012	0,015	0,290	100
λ_1	0,3	0,301	0,018	0,304	0,112	0,490	0,015	0,018	4,966	100
ν	10,0	9,991	0,001	9,991	9,928	10,053	0,009	0,009	0,093	100

MS2										
Θ	Valor	Média	DP	Mediana	2,5%	97,5%	MAB	RMSE	MAAPE	PC
β_0	1,6	1,611	0,041	1,608	1,434	1,788	0,033	0,042	2,066	100
ϕ_1	0,2	0,211	0,044	0,204	0,099	0,320	0,035	0,045	16,938	98
θ_1	-0,1	-0,108	0,057	-0,106	-0,236	0,025	0,044	0,057	37,661	96
γ_0	1,2	1,200	0,024	1,200	1,018	1,383	0,018	0,024	1,531	100
λ_1	0,3	0,298	0,046	0,299	0,154	0,445	0,034	0,046	11,173	100
ν	10,0	9,992	0,001	9,992	9,929	10,055	0,008	0,008	0,079	100

Fonte: Elaborada pelo autor.

APLICAÇÕES EM DADOS REAIS

Neste capítulo, a aplicabilidade prática e a flexibilidade dos modelos ZMD-GARMA são ilustradas em cenários reais que apresentam diferentes características de distribuição, dinâmica temporal e diferentes comportamentos de modificação de zeros. Três conjuntos de dados são analisados para testar a robustez da modelagem. Na [Seção 6.1](#) é feita a análise de anomalias de temperatura global, uma série que exibe mudanças estruturais e no comportamento de zeros. A [Seção 6.2](#) exibe a modelagem da diferença das ordens de compra e venda de Bitcoin, permitindo testar a ZMS-GARMA em um cenário de mudança estrutural de mercado. Finalmente, na [Seção 6.3](#), são analisados dados de notificações de Chikungunya, que apresentam forte inflação de zeros e surtos epidêmicos sazonais.

6.1 Anomalias na temperatura na superfície da terra

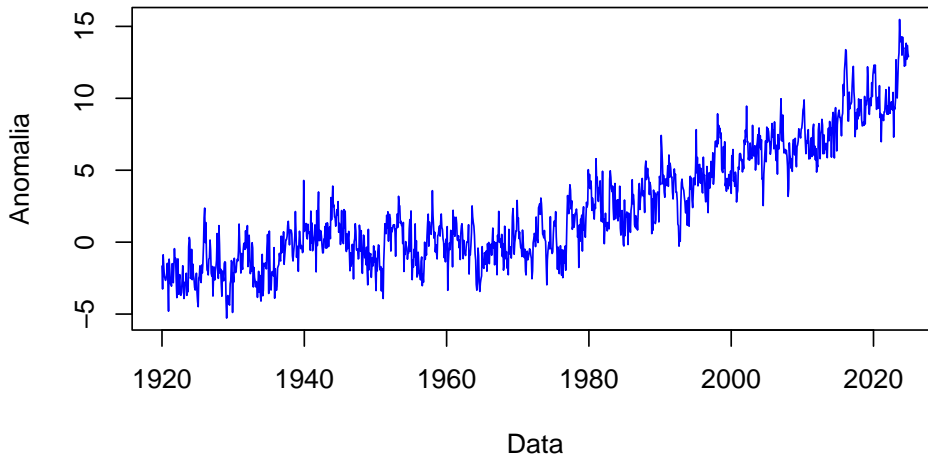
Anomalias na temperatura são uma métrica fundamental na climatologia, fornecendo uma compreensão das variações e tendências térmicas do planeta ao longo do tempo. Tal medida é definida pelo desvio do valor de temperatura observado em relação a um valor de referência, usualmente a média histórica calculada por um longo período de tempo. Desse modo, uma anomalia é uma flutuação extrema, que se desvia acentuadamente do padrão de variabilidade observado ao longo do tempo.

Os dados utilizados estão disponibilizados por [Rohde e Hausfather \(2020\)](#) e são referentes às estimativas de extensão do gelo marinho extrapoladas a partir das anomalias de temperatura do ar na superfície terrestre. O valor de referência para o cálculo é a média estimada de temperatura entre janeiro de 1951 a dezembro de 1980, calculada em 14,102 graus celsius. Os dados abrangem o período de janeiro de 1850 a dezembro de 2024.

Para o estudo o intervalo de janeiro de 1920 a dezembro de 2024 foi selecionado, resultando em 2508 observações. O período foi escolhido por apresentar uma abrangência

histórica suficiente para a modelagem e maior confiabilidade dos dados, evitando a maior incerteza das observações iniciais. Os valores das anomalias foram multiplicados por 10 para aumentar sua magnitude e apenas sua parte inteira foi considerada. A Figura 8 mostra a evolução histórica das anomalias ao longo do período analisado.

Figura 8 – Anomalias de temperatura na superfície terrestre entre janeiro de 1920 e dezembro de 2024



Fonte: Elaborada pelo autor.

Apesar de exibir regimes diferentes ao longo do tempo, o período de janeiro de 1920 a dezembro de 2019 foi utilizado para ajustar o modelo ZMS-GARMA(p,q) com uma função de ligação identidade, $\mu_t = \eta_t$, e uma periodicidade de 12 definida como

$$\eta_t = \beta_0 + \beta_1 \cos\left(\frac{2\pi t}{12}\right) + \sum_{j=1}^p \phi_j y_{t-j} + \sum_{j=1}^q \theta_j \{y_{t-j} - \mu_{t-j}\}, \quad (6.1)$$

foi ajustado aos dados. O termo trigonométrico foi incluído por mostrar melhoras nos ajustes do modelo e sua período definido após a análise da Função de Autocorrelação (ACF) da série e sua respectiva análise espectral pelo periodograma (BOX; JENKINS; REINSEL, 2008; SHUMWAY; STOFFER, 2017). A função de ligação para o parâmetro w_t foi a função de ligação logito, como em (2.9). Distribuições *a priori* normais padrão foram consideradas para os parâmetros β , ϕ , θ , γ e λ . Para o parâmetro de dispersão v , uma distribuição *a priori* log-normal com $v_0 = 0$ e $\sigma_v^2 = 1$ foi adotada.

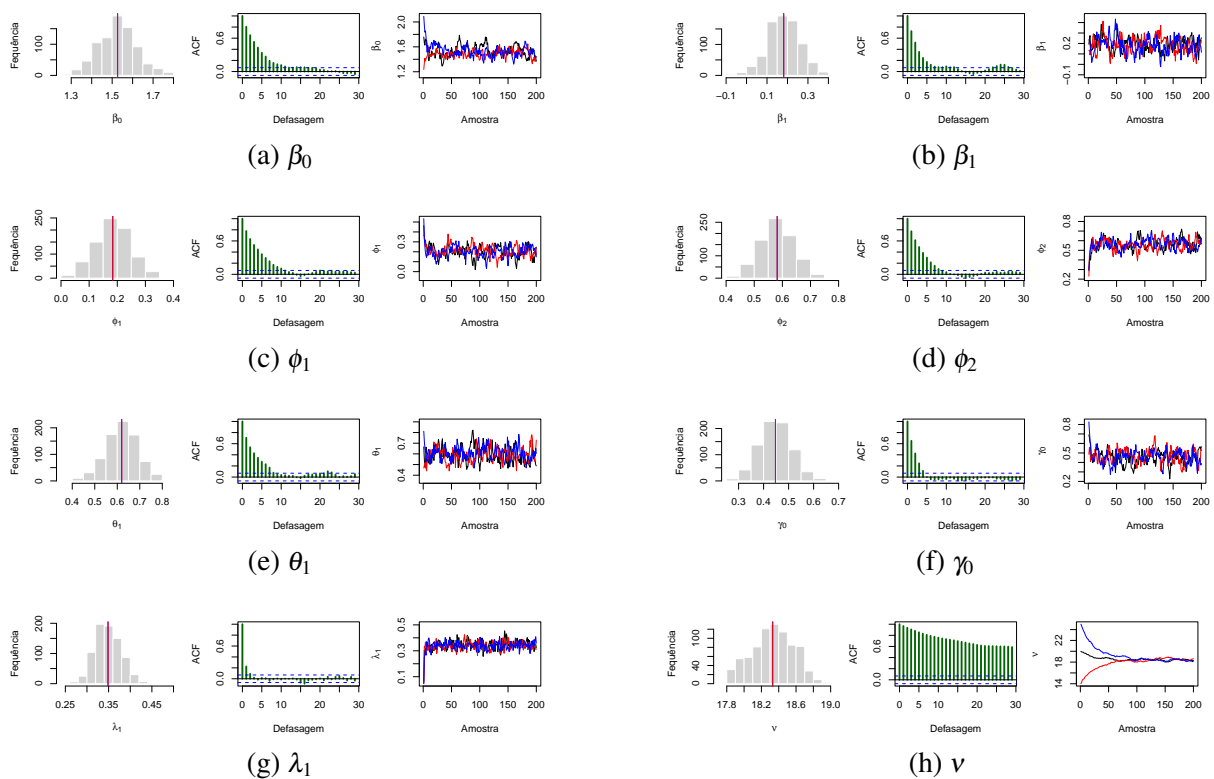
No algoritmo SGHMC, os valores iniciais para os parâmetros em (6.1) foram definidos pelas estimativas obtidas à partir de um ajuste ARMA(p,q) normal. Para os parâmetros em (2.9) foram utilizados os estimadores de mínimos quadrados considerando $1 - w_t$ igual à frequência de zeros observados. Para o parâmetro de dispersão v , a variância da série no intervalo de tempo, $v = 20$, foi adotada como valor inicial.

Todas as combinações de $p, q, r = 0, 1, \dots, 5$ foram ajustadas para os dados. Uma única cadeia multivariada de 50.000 iterações foi simulada da distribuição posterior, com passos de tamanho $\varepsilon = 0,01$, número de passos $L = 10$ e componente de fricção $\alpha = 0,1$. As 10.000

observações iniciais foram descartadas em processo de *burn-in* e saltos de 40 observações foram realizados no processo de espaçamento para reduzir a autocorrelação das amostras. As 1.000 observações restantes foram utilizadas Para calcular as estimativas *a posteriori* após as verificações de convergência das cadeias.

O modelo selecionado pelos critérios de seleção foi o ZMS-GARMA(2,1) com componente *hurdle* $s = 1$, com $\log(\widehat{CPO}) = -2666,73$, $\widehat{EBIC} = 5388,73$ e $\widehat{WAIC} = 5333,46$. O histograma das cadeias utilizadas para a estimação e sua respectiva autocorrelação são apresentados na Figura 9.

Figura 9 – Histograma, autocorrelação e convergência das cadeias para o modelo ajustado.



Fonte: Elaborada pelo autor.

A convergência das cadeias foi aferida pelo diagnóstico GW, exibido na Tabela 6, todos os valores z são menores que 2 em módulo, de modo que as cadeias são assumidas convergentes. Outro teste de convergência foi realizado ao iniciar o amostrador em valores distintos. O terceiro gráfico de cada parâmetro na Figura 9 mostra a evolução das cadeias em valores distintos contra a cadeia original utilizada para a estimação, em preto. Para melhor legibilidade e clareza gráfica, as cadeias apresentadas são apresentadas com espaçamento para representar sua trajetória e convergência. A convergência das cadeias com valores iniciais distintos também foi aferida pelo diagnóstico de GR, a Tabela 4 exhibe as estimativas pontuais, \hat{R} , e seu respectivo limite superior. Todas as estimativas pontuais são suficientemente próximas de 1, indicando convergência das cadeias.

Tabela 6 – Resultados de convergência para o modelo ajustado.

Θ	z	\hat{R}	Limite Superior para \hat{R}
β_0	0,7499	1,0068	1,0225
β_1	-1,0472	1,0043	1,0148
ϕ_1	-0,4513	1,0143	1,0450
ϕ_2	0,4781	1,0113	1,0353
θ_1	0,3484	1,0100	1,0348
γ_0	-0,4464	1,0163	1,0547
λ_1	0,4651	1,0106	1,0375
ν	-0,9138	1,0364	1,1089

Fonte: Elaborada pelo autor.

As medidas resumo *a posteriori* do modelo ajustado ZMS-GARMA(2,1) com $s = 1$ são exibidas na Tabela 7, incluindo os intervalos de credibilidade bayesianos. A Figura 10 exibe os valores observados junto com a média ajustada pelo modelo. O modelo ajustado captura a tendência da série e acompanha o comportamento dos dados.

Tabela 7 – Estimativas *a posteriori* para os parâmetros do modelo ZMS-GARMA(2,1) ajustado aos dados.

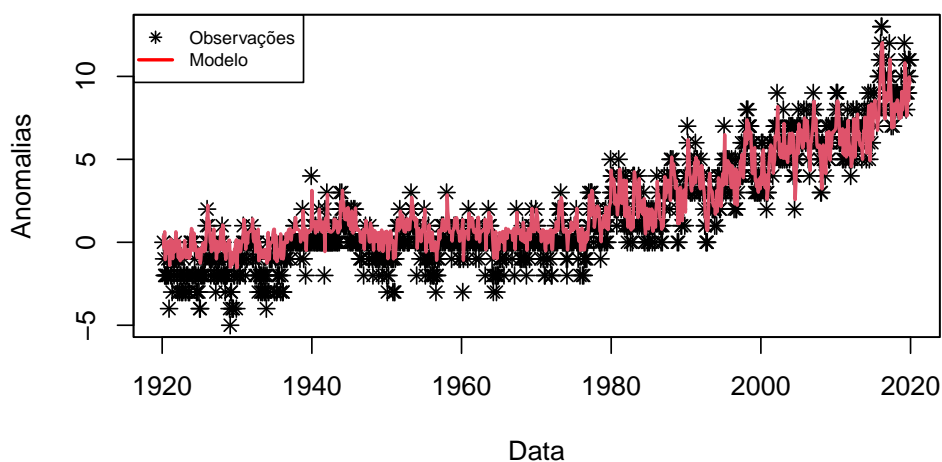
Θ	Média	Mediana	Desvio Padrão	2,5%	97,5%
β_0	1,53	1,53	0,09	1,35	1,72
β_1	0,18	0,18	0,08	0,01	0,32
ϕ_1	0,18	0,19	0,06	0,07	0,32
ϕ_2	0,58	0,58	0,06	0,47	0,70
θ_1	0,62	0,62	0,07	0,47	0,76
γ_0	0,45	0,45	0,06	0,33	0,57
λ_1	0,35	0,35	0,03	0,29	0,41
ν	18,33	18,34	0,24	17,85	18,72

Fonte: Elaborada pelo autor.

A Figura 11 exibe o parâmetro \hat{p}_t estimado entre janeiro de 1920 e dezembro de 2019. Analisando a figura, o parâmetro p_t mostra um crescimento gradual de valores menores que 1 para valores muito próximo a 1. Como discutido na Seção 2.2, valores de p_t menores do que 1 correspondem a uma frequência de zeros observados maior do que o esperado para a distribuição Skellam usual, enquanto valores próximos de 1 indicam o comportamento de zeros próximo da Skellam sem modificação. Dois cenários podem ser observados no comportamento estrutural da série. Na primeira porção da série, até 1980, as observações baixas e consistentes de p_t apontam para uma zero-inflação acentuada, compatível com o período de pouca variabilidade e alta concentração de observações em torno do zero observado nos dados. De 1980 a diante, p_t exibe crescimento gradual se aproximando de 1, indicando atenuação da inflação de zeros, que condiz com o aumento da variabilidade dos dados observados e o aumento das ocorrências diferentes de zero.

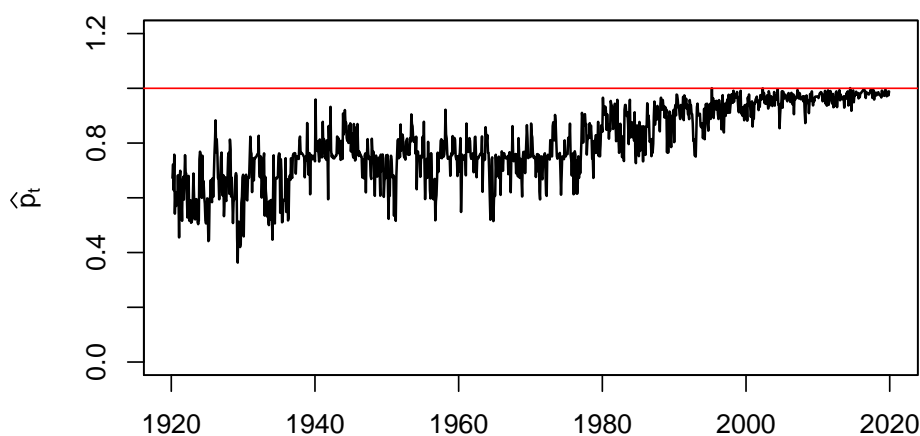
O modelo ZMS-GARMA(2,1) com $s = 1$ foi, então, utilizado para a predição dos

Figura 10 – Dados observados e a média ajustada das anomalias de temperatura de janeiro de 1920 a dezembro de 2019



Fonte: Elaborada pelo autor.

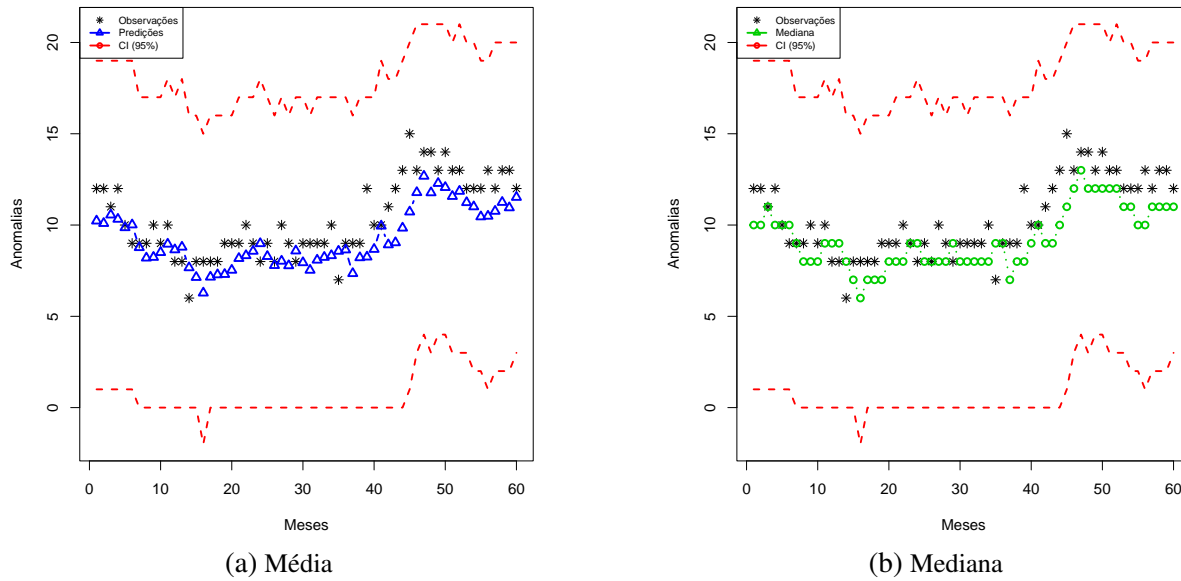
Figura 11 – Parâmetro $\hat{\rho}_t$ estimado de janeiro de 1920 a dezembro de 2019.



Fonte: Elaborada pelo autor.

valores de 60 meses entre janeiro de 2020 e dezembro de 2024. A [Figura 12](#) apresenta os dados observados, a predição um passo adiante pela média condicional, a mediana como predição coerente à estrutura dos dados e os intervalos de credibilidade de 95%. Os valores preditos são próximos dos valores observados e seguem a tendência dos dados. Para avaliar as predições, três métricas foram utilizadas: a RMSE, o Erro Absoluto Médio (MAE) e o Erro Absoluto Médio da Mediana (MAEM). As métricas resultantes foram RMSE = 1,56, MAE = 1,32 e MAEM = 1,38, indicando uma boa capacidade preditiva para o modelo.

Figura 12 – Predições um passo adiante para as anomalias mensais de temperatura entre 2020 a 2024



Fonte: Elaborada pelo autor.

6.2 Compra e venda de Bitcoin

A diferença entre os volumes de compra e venda de ativos ou criptomoedas, frequentemente denominado *Order Flow Imbalance*, é fundamental para compreender a microestrutura do mercado e a antecipação da movimentação de preços em curto prazo. Diferente dos preços de fechamento que costumam ser indicadores atrasados, a diferença de compra e venda reflete a pressão imediata de ordem e demanda.

Os dados utilizados estão disponibilizados pela OpenBinance (Binance, 2025) e referem-se às diferenças entre compra e venda de Bitcoin no intervalo de um segundo. Os dados compreendem cinco minutos, totalizando 300 observações. O intervalo de tempo foi escolhido para evitar o acúmulo de ruído introduzido pelas observações tomadas a um segundo e evitar muitas mudanças bruscas de regime, como conhecido pelas criptomoedas.

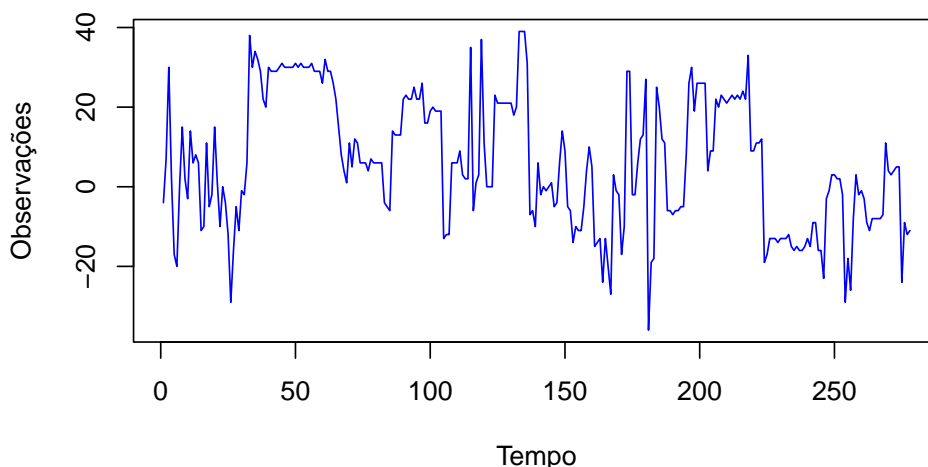
As primeiras 280 observações foram utilizadas para o ajuste do modelo ZMS-GARMA(p,q) e a função de ligação identidade, $\mu_t = \eta_t$, da forma

$$\eta_t = \beta_0 + \sum_{j=1}^p \phi_j y_{t-j} + \sum_{j=1}^q \theta_j \{y_{t-j} - \mu_{t-j}\}, \quad (6.2)$$

foi utilizada para a estimação do modelo. A função de ligação para o parâmetro w_t foi a função de ligação logito, como em (2.9). Distribuições *a priori* normais padrão foram consideradas para os parâmetros β , ϕ , θ , γ e λ . Para o parâmetro de dispersão v , uma distribuição *a priori* log-normal com $v_0 = 0$ e $\sigma_v^2 = 1$ foi adotada.

No algoritmo SGHMC, os valores iniciais para os parâmetros em (6.2) foram definidos pelas estimativas obtidas a partir de um ajuste ARMA(p,q) normal. Para os parâmetros em (2.9)

Figura 13 – Diferença de oferta e procura de Bitcoin no intervalo 300 segundos



Fonte: Elaborada pelo autor.

foram utilizados os estimadores de mínimos quadrados considerando $1 - w_t$ igual à frequência de zeros observados. Para o parâmetro de dispersão ν , o maior valor da média calculada no intervalo acrescido de 1, $\nu = 32$, foi adotada como valor inicial.

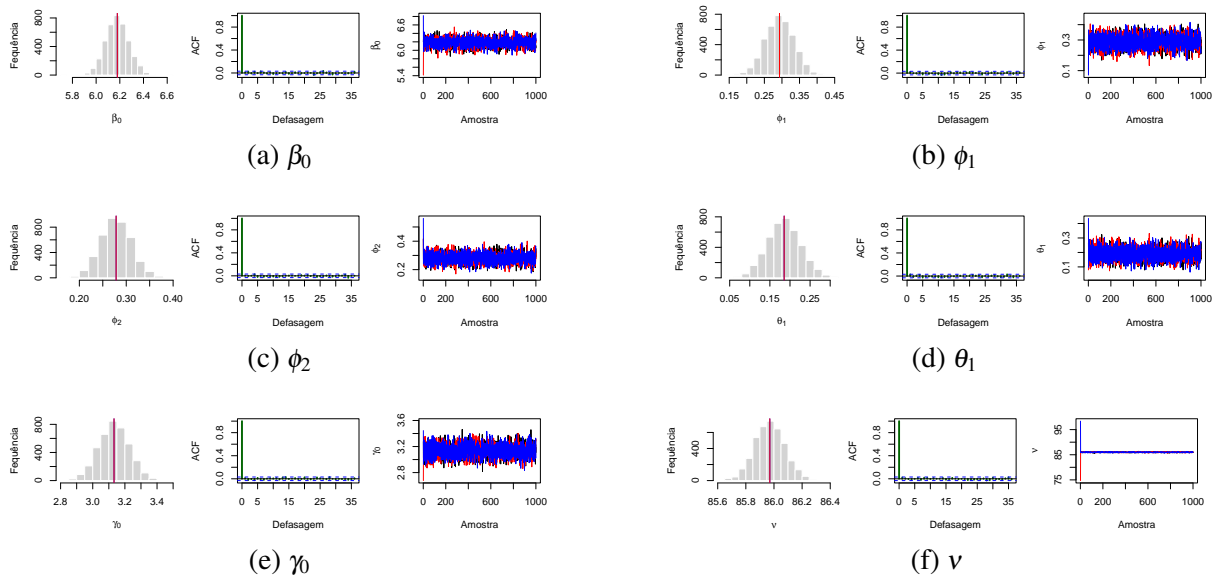
Combinações de $p, q, r = 0, 1, \dots, 5$ foram ajustadas para os dados. Uma única cadeia multivariada de 50.000 iterações foi simulada da distribuição posterior, com passos de tamanho $\varepsilon = 0,01$, número de passos $L = 10$ e componente de fricção $\alpha = 0,1$. As 10.000 observações iniciais foram descartadas em processo de *burn-in* e saltos de 10 observações foram realizados no processo de espaçamento para reduzir a autocorrelação das amostras. As 4.000 observações restantes foram utilizadas para calcular as estimativas *a posteriori* após as verificações de convergência das cadeias.

O modelo selecionado foi o modelo ZMS-GARMA(2,1) com componente *hurdle* $s = 0$, com os valores dos critérios de seleção $\log(\widehat{CPO}) = -895,09$, $\widehat{EBIC} = 1812,55$ e $\widehat{WAIC} = 1790,18$. Os histogramas das cadeias utilizadas para a estimação dos parâmetros e suas respectivas autocorrelações são exibidas na [Figura 14](#).

O critério GW foi utilizado para diagnosticar a convergência das cadeias resultantes, exibido na [Tabela 8](#). Todas as cadeias puderam ser assumidas convergentes, dado que todos os valores z são menores do que 2 em módulo. A convergência também foi testada iniciando o amostrador em valores distintos para cada parâmetro. O terceiro gráfico de cada parâmetro na [Figura 14](#) exibe a trajetória das cadeias simuladas para cada valor inicial contra a cadeia original em preto. Os gráficos exibem as cadeias espaçadas, para melhor legibilidade. O teste GR foi utilizado para verificar a convergência das cadeias, exibido na [Tabela 8](#). Como o valor $\hat{R} = 1$ para todos os parâmetros, as cadeias são assumidas convergentes.

As estimações *a posteriori* do modelo ajustado são exibidas na [Tabela 9](#) e a [Figura 15](#) exibe os dados observados da diferença de compra e venda de Bitcoin junto à média ajustada

Figura 14 – Histograma, autocorrelação e convergência das cadeias para o modelo ajustado.



Fonte: Elaborada pelo autor.

Tabela 8 – Resultados de convergência para o modelo ajustado.

Θ	z	\hat{R}	Limite Superior para \hat{R}
β_0	-0,50572	1	1
ϕ_1	0,68468	1	1
ϕ_2	0,22334	1	1
θ_1	-0,01037	1	1
γ_0	-1,35471	1	1
ν	0,35685	1	1

Fonte: Elaborada pelo autor.

pele modelo. A Figura 16 exibe os valores estimados de \hat{p}_t para o intervalo do ajuste. Os valores de $\hat{p}_t \approx 1$ indicam que a série exibe uma frequência de zeros próxima de uma distribuição sem modificações de zeros. O modelo ajustado acompanha a dinâmica geral da série e tenta se adaptar às aparentes mudanças de regime que ocorrem nos dados.

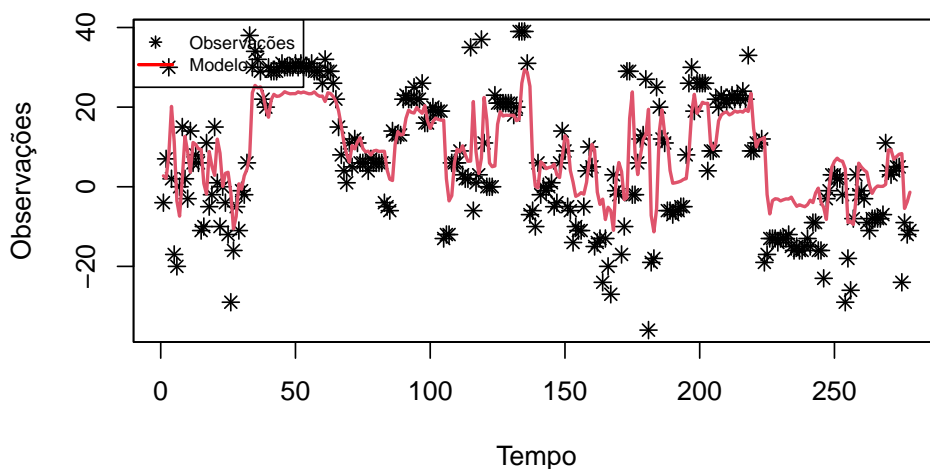
Tabela 9 – Estimativas *a posteriori* para os parâmetros do modelo ZMS-GARMA(2,1) ajustado aos dados.

Θ	Média	Mediana	Desvio Padrão	2,5%	97,5%
β_0	6,18	6,18	0,10	5,99	6,37
ϕ_1	0,29	0,29	0,04	0,21	0,37
ϕ_2	0,28	0,28	0,03	0,21	0,34
θ_1	0,19	0,19	0,04	0,10	0,27
γ_0	3,13	3,13	0,10	2,95	3,32
ν	85,97	85,97	0,11	85,77	86,18

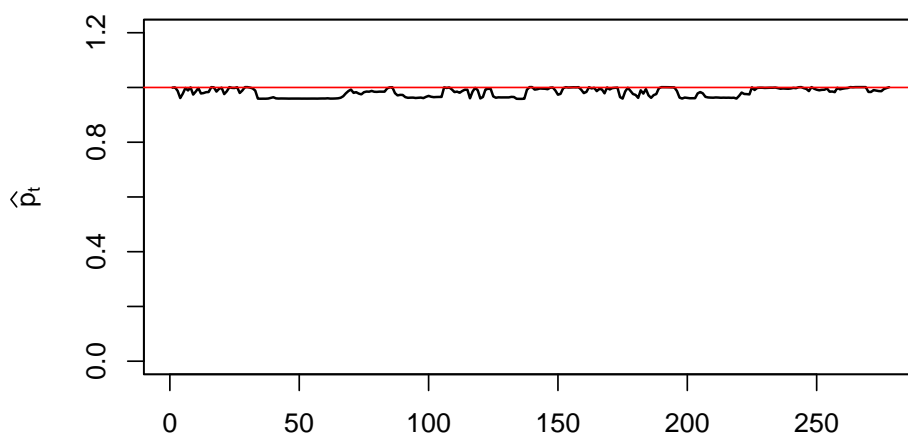
Fonte: Elaborada pelo autor.

Um estudo de predição foi realizado com as últimas 20 observações da série utilizando o

Figura 15 – Dados observados e a média ajustada da diferença da compra e venda de Bitcoin no intervalo.



Fonte: Elaborada pelo autor.

Figura 16 – Parâmetro \hat{p}_t ajustado para o intervalo.

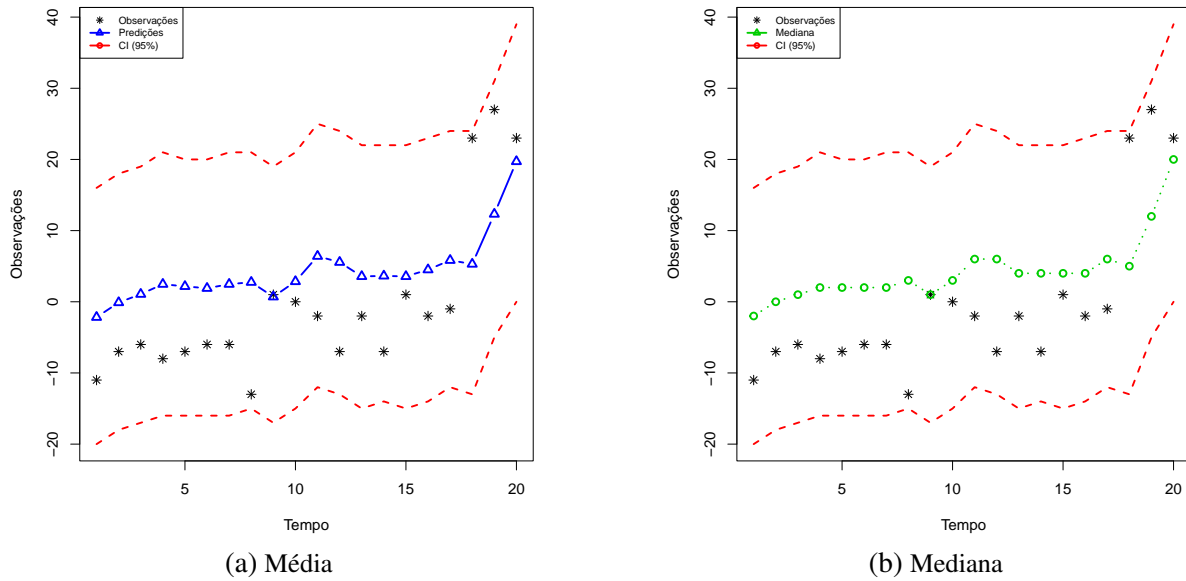
Fonte: Elaborada pelo autor.

modelo ZMS-GARMA(2,1) ajustado. A [Figura 17](#) apresenta as previsões pela média condicional e pela mediana, junto dos dados observados e do intervalo de credibilidade. Apesar do aparente viés positivo, o modelo ajustado captura a tendência da série e é capaz de prever o salto nas observações nos últimos instantes. As métricas RMSE = 9,4, MAE = 8,32 e MAEM = 8,35 indicam uma capacidade preditiva aceitável para o modelo.

6.3 Notificações de casos de Chikungunya na cidade de Ribeirão Preto

A Chikungunya é uma arbovirose transmitida para humanos principalmente por mosquitos *Aedes aegypti* e *Aedes albopictus* previamente infectados, sendo o *Aedes aegypti* o principal

Figura 17 – Predições um passo adiante para as a diferença de compra e venda de Bitcoin.



Fonte: Elaborada pelo autor.

vetor da doença no Brasil. A doença é conhecida principalmente por causar febre alta e dores intensas nas articulações, podendo também causar dores musculares, dor de cabeça, o aparecimento de manchas na pele e náusea. Como não há antiviral para a infecção, a maior medida adotada é a prevenção focada na eliminação de criadouros do mosquito, assim como a dengue.

Os dados utilizados estão disponibilizados no Datasus ([Brasil. Ministério da Saúde. DATASUS, 2026](#)) e acessados através da biblioteca Microdatasus ([SALDANHA; BASTOS; BARCELLOS, 2019](#)). Os dados de notificações semanais na cidade de Ribeirão Preto compreendem o intervalo de janeiro de 2016, quando iniciou-se o registro de notificações no sistema, até dezembro de 2025, totalizando 520 observações. A [Figura 18](#) exibe a evolução dos casos notificados ao longo do tempo. Uma observação importante é que os dados referentes ao ano de 2025 ainda constam como preliminares no sistema.

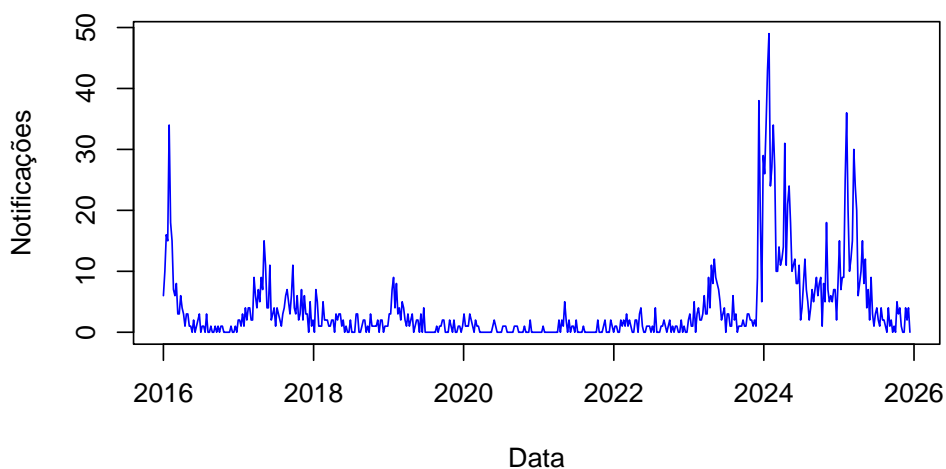
O período de janeiro de 2016 a dezembro de 2024 foi utilizado para ajustar o modelo ZMPS-GARMA(p,q) com uma função de ligação logaritmo $\log(\mu_t) = \eta_t$ e uma periodicidade de 52 definida como

$$\eta_t = \beta_0 + \beta_1 \cos\left(\frac{2\pi t}{52}\right) + \sum_{j=1}^p \phi_j \log(y_{t-j}) + \sum_{j=1}^q \theta_j \log\left(\frac{y_{t-j}}{\mu_{t-j}}\right). \quad (6.3)$$

O termo trigonométrico foi incluído por mostrar melhoras no ajuste do modelo e sua periodicidade 52 escolhida analisando o ACF da série e seu periodograma. Assim como nos casos anteriores, a função de ligação logito foi utilizada para o parâmetro w_t e distribuições normal padrão foram consideradas para β , ϕ , θ , γ e λ . Para o parâmetro de dispersão v , uma distribuição *a priori* log-normal com $v_0 = 0$ e $\sigma_v^2 = 1$ foi adotada, quando necessário à distribuição assumida.

Os valores iniciais para o algoritmo SGHMC para os parâmetros em (6.3) foram definidos

Figura 18 – Notificações semanais de casos de Chikungunya em Ribeirão Preto entre janeiro de 2016 a dezembro de 2025



Fonte: Elaborada pelo autor.

pelas estimativas obtidas pelo ajuste de um modelo ARMA(p,q) normal. Para os parâmetros em (2.9) foram utilizados os estimadores de mínimos quadrados para $1 - w_t$ igual à frequência de zeros nos dados. Para o parâmetro de dispersão v , a variância empírica de acordo com cada distribuição foi utilizada como valor inicial.

Combinações de $p, q, r = 0, 1, \dots, 5$ foram ajustadas para os dados assumindo as distribuições de Poisson, Binomial Negativa e a Poisson Generalizada. Uma única cadeia multivariada com 50.000 iterações foi simulada da distribuição posterior, com passos de tamanho $\varepsilon = 0,01$, número de passos $L = 10$ e componente de fricção $\alpha = 0,1$. As 10.000 observações iniciais foram descartadas em processo de *burn-in* e saltos de 20 observações foram realizados no processo de espaçamento para reduzir a autocorrelação das amostras. As 2.000 observações restantes foram utilizadas para calcular as estimativas *a posteriori* após as verificações de convergência das cadeias.

A Tabela 10 exibe os critérios de seleção para os melhores modelos assumindo cada distribuição. O modelo selecionado foi o ZMNB-GARMA(3,0), com componente *hurdle* $s = 1$. O histograma das cadeias utilizadas para a estimação dos parâmetros e sua respectiva autocorrelação são apresentados na Figura 19.

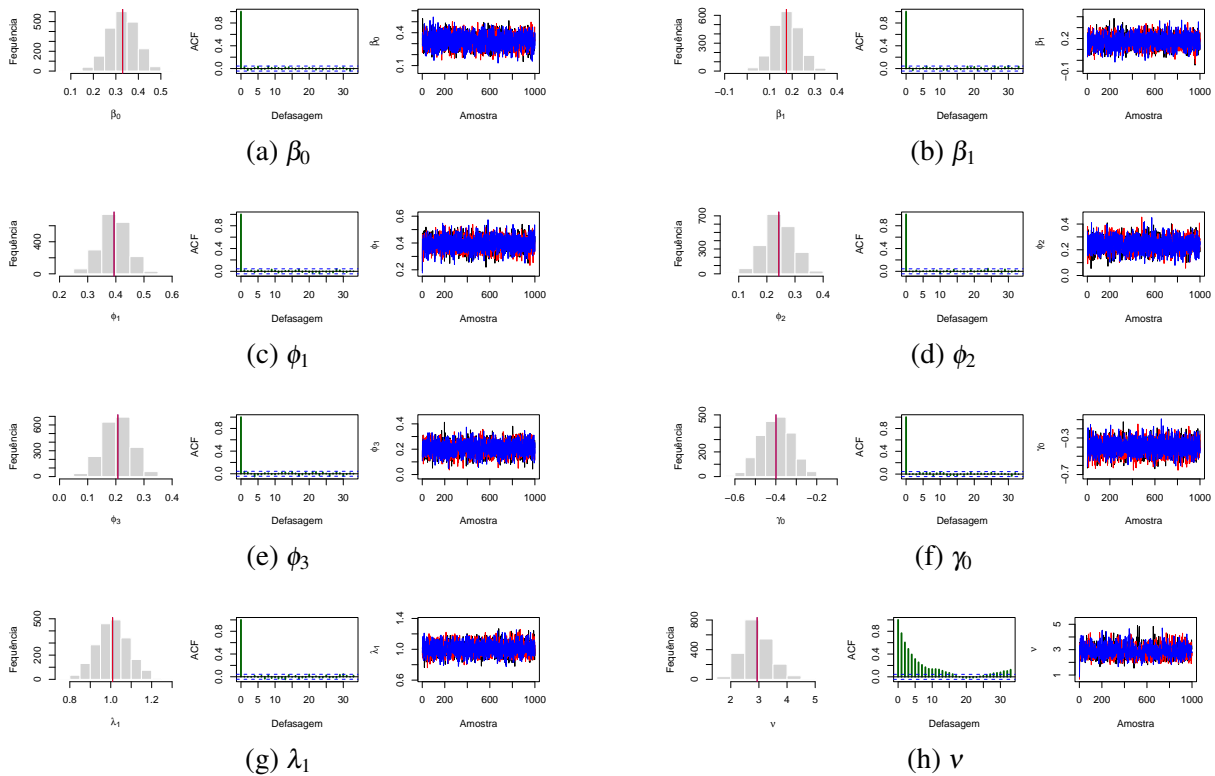
Tabela 10 – Critérios de seleção para os modelos ajustados para as notificações de Chikungunya.

Modelo	$-\log(\widehat{CPO})$	\widehat{EBIC}	\widehat{WAIC}
ZMP-GARMA(3,0), $s = 1$	-947,8	1973,02	1955,44
ZMNB-GARMA(3,0), $s = 1$	-873,86	1811,92	1766,89
ZMGP-GARMA(3,0), $s = 1$	-935,24	1934,54	1889,96

Fonte: Elaborada pelo autor.

A convergência das cadeias foi aferida pelo diagnóstico GW, exibido na Tabela 11. Todas

Figura 19 – Histograma, autocorrelação e convergência das cadeias para o modelo ajustado.



Fonte: Elaborada pelo autor.

as cadeias foram assumidas convergentes, dado que o valor z é menor que 2 em módulo para todas as cadeias. O teste de convergência foi realizado também para cadeias com valores iniciais distintos. O terceiro gráfico de cada parâmetro da Figura 19 exibe a evolução das cadeias para os diferentes valores iniciais contra a cadeia original utilizada para o processo inferencial (em preto). Os gráficos exibem as cadeias espaçadas para melhor interpretabilidade. Pelo teste de GR, exibido na Tabela 11, pode-se assumir que todas as cadeias convergiram para a mesma região, dado que $\hat{R} = 1$ em todos os casos.

Tabela 11 – Resultados de convergência para o modelo ajustado.

\ominus	z	\hat{R}	Limite Superior para \hat{R}
β_0	0,5398	1,00	1,00
β_1	-0,7858	1,00	1,00
ϕ_1	1,8303	1,00	1,00
ϕ_2	-1,9137	1,00	1,00
ϕ_3	0,1751	1,00	1,00
γ_0	1,0009	1,00	1,00
λ_1	-1,3114	1,00	1,00
ν	0,7348	1,00	1,01

Fonte: Elaborada pelo autor.

As medidas resumo *a posteriori* do modelo ajustado ZMNB-GARMA(3,0) com $s = 1$

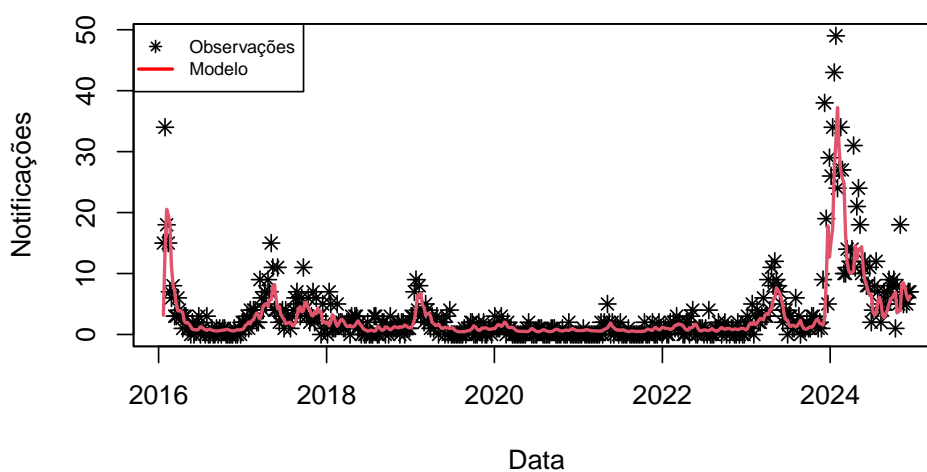
são exibidas na Tabela 12. A Figura 20 exibe os valores observados junto com a média ajustada pelo modelo. O modelo ajustado se mostrou capaz de capturar a tendência da série e acompanhar o comportamento dos dados.

Tabela 12 – Estimativas *a posteriori* para os parâmetros do modelo ZMNB-GARMA(3,0) ajustado aos dados.

Θ	Média	Mediana	Desvio Padrão	2,5%	97,5%
β_0	0,33	0,33	0,06	0,21	0,45
β_1	0,17	0,18	0,06	0,06	0,28
ϕ_1	0,39	0,39	0,05	0,29	0,49
ϕ_2	0,24	0,24	0,05	0,14	0,35
ϕ_3	0,21	0,21	0,05	0,10	0,30
γ_0	-0,40	-0,40	0,08	-0,55	-0,26
λ_1	1,01	1,01	0,08	0,87	1,18
ν	2,94	2,88	0,51	1,98	3,96

Fonte: Elaborada pelo autor.

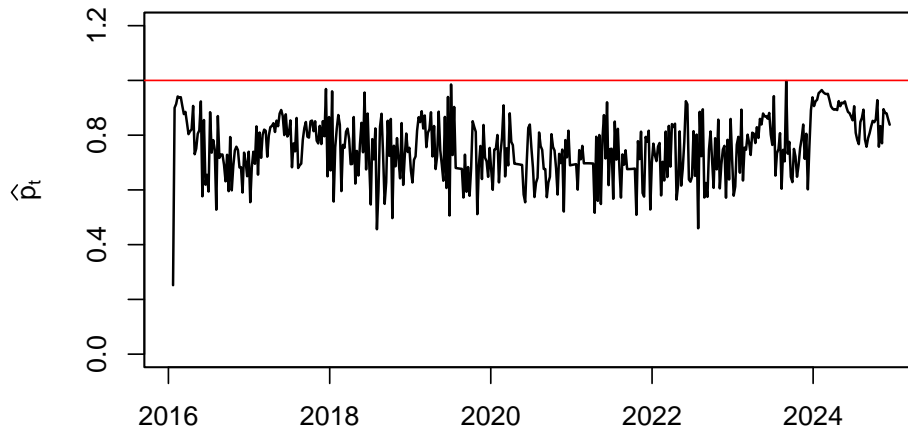
Figura 20 – Dados observados e média ajustada das notificações de casos de Chikungunya entre janeiro de 2016 a dezembro de 2024.



Fonte: Elaborada pelo autor.

A Figura 21 exibe o parâmetro \hat{p}_t estimado no período de 2016 a 2024. O valor estimado de \hat{p}_t se mantém menor que 1 de forma consistente, indicando uma frequência de zeros observados maior do que o esperado para uma distribuição Binomial Negativa comum, que condiz com os dados observados, que se mostram em zero ou próximos ao zero na maior parte das observações. Em alguns pontos, o valor de \hat{p}_t cresce para valores próximos de 1, coincidindo com momentos de concentração de notificações e picos epidêmicos registrados.

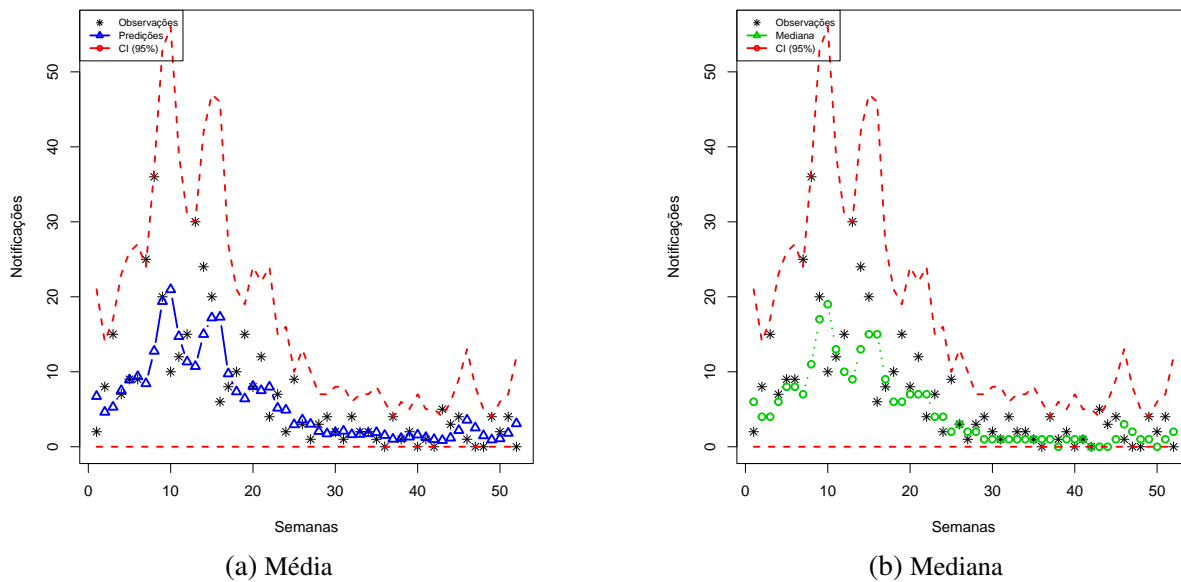
Por fim, o modelo ZMNB-GARMA(3,0) com $s = 1$ foi utilizado para a predição dos valores para as 52 semanas de 2025. A Figura 22 mostra os dados observados, a predição um passo adiante pela média condicional, a predição pela mediana e os intervalos de credibilidade a 95%. Em sua maioria, os valores preditos se mostram próximos dos valores observados e

Figura 21 – Parâmetro $\hat{\rho}_t$ estimado entre janeiro de 2016 a dezembro de 2024

Fonte: Elaborada pelo autor.

as previsões seguem a tendência dos dados. Apesar da presença de duas observações fora do intervalo de confiança, os valores de RMSE = 6,1, MAE = 3,72 e MAEM = 4 indicam uma capacidade preditiva aceitável para o modelo.

Figura 22 – Predições um passo adiante para as notificações semanais de Chikungunya no ano de 2025.



Fonte: Elaborada pelo autor.

CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo propor e validar uma classe de modelos para séries temporais com dados discretos, denominados modelos Discretos Zero-Modificados GARMA. A principal motivação para este desenvolvimento surgiu da necessidade de modelar dados de contagem e dados inteiros, presentes em diversas áreas de atuação. Outro fator importante foi a necessidade de modelos adequados para situações em que os dados apresentam excessos ou escassez de zeros, características frequentemente observadas em fenômenos reais que os modelos GARMA tradicionais não capturam adequadamente.

Ao longo do texto, foi apresentada a formulação dos modelos GARMA para distribuições de probabilidade discretas, nos modelos Discreto GARMA, e seus principais casos particulares. Também formalizou-se a extensão dos modelos para acomodar a dinâmica de modificações de zeros, permitindo que sua probabilidade de ocorrência variasse ao longo do tempo no modelo. Uma contribuição significativa deste trabalho foi o uso da distribuição Skellam nessa estrutura, permitindo a modelagem de dados no conjunto dos inteiros com comportamentos de modificação de zeros, preenchendo uma lacuna na literatura da modelagem de dados inteiros.

Sob um paradigma Bayesiano, utilizou-se o algoritmo Hamiltoniano de Monte Carlo com Gradiente Estocástico para a simulação de amostras da distribuição posterior dos modelos. Os estudos de simulação apresentados no [Capítulo 5](#) demonstraram a eficiência e robustez do método para a estimação dos parâmetros dos modelos. Em todos os casos, o método recuperou os valores verdadeiros dos parâmetros, com viés baixo e pouca incerteza sobre os valores estimados.

Três aplicações com dados reais e características distintas foram apresentadas no [Capítulo 6](#) para ilustrar a flexibilidade e aplicabilidade dos modelos ZMD-GARMA. Nas anomalias de temperatura na superfície terrestre o modelo ZMS-GARMA permitiu identificar mudanças estruturais na série ao longo de um século, capturando a mudança de um regime zero-inflacionado para uma distribuição Skellam padrão, refletindo o aumento da variabilidade climática. Na análise de *Order Flow Imbalance*, um conjunto de dados inteiros praticamente sem modifica-

ção de zeros, o modelo ZMS-GARMA adaptou-se bem às mudanças de regime presentes na série e aos ruídos característicos da mesma. Por fim, no contexto epidemiológico, o modelo ZMNB-GARMA capturou a inflação de zeros e ajustou-se bem aos surtos epidêmicos sazonais, fornecendo previsões coerentes.

Em suma, os modelos ZMD-GARMA se mostraram uma ferramenta versátil e poderosa para séries temporais com dados discretos, capazes de compreender a dinâmica de zeros e a estrutura temporal da série. Propostas de trabalhos futuros incluem a implementação de diferentes distribuições zero-modificadas, como o caso da ZMPIG-GARMA; a possível implementação de um pacote em R para lidar com os modelos ZMD-GARMA e a extensão de séries discretas com modificação de zeros para os modelos dinâmicos de séries temporais.

REFERÊNCIAS

AKAIKE, H. A New Look at the Statistical Model Identification. In: _____. **Selected Papers of Hirotugu Akaike**. Springer New York, 1974. p. 215–222. ISBN 9781461216940. Disponível em: http://dx.doi.org/10.1007/978-1-4612-1694-0_16. Citado na página 49.

ALOMANI, G. A.; ALZAID, A. A.; OMAIR, M. A. A Skellam GARCH model. **Brazilian Journal of Probability and Statistics**, Institute of Mathematical Statistics, v. 32, n. 1, fev. 2018. ISSN 0103-0752. Disponível em: <http://dx.doi.org/10.1214/16-BJPS338>. Citado nas páginas 25 e 26.

ALZAID, A.; OMAIR, M. On the Poisson difference distribution inference and applications. **Bulletin of the Malaysian Mathematical Sciences Society. Second Series**, Malaysian Mathematical Sciences Society, Bangi, v. 33, n. 1, p. 17–45, 2010. Disponível em: <http://eudml.org/doc/244475>. Citado na página 26.

ALZAID, A. A.; AL-OSH, M. An integer-valued p th-order autoregressive structure (INAR(p)) process. **Journal of Applied Probability**, Cambridge University Press (CUP), v. 27, n. 2, p. 314–324, jun. 1990. ISSN 1475-6072. Disponível em: <http://dx.doi.org/10.2307/3214650>. Citado na página 26.

ANDRADE, B. Silveira de; ANDRADE, M. G.; EHLERS, R. S. Bayesian GARMA models for count data. **Communications in Statistics: Case Studies, Data Analysis and Applications**, Informa UK Limited, v. 1, n. 4, p. 192–205, out. 2015. ISSN 2373-7484. Disponível em: <http://dx.doi.org/10.1080/23737484.2016.1190307>. Citado nas páginas 26 e 27.

ANDRADE, M. G.; CONCEIÇÃO, K. S.; RAVISHANKER, N. Zero-modified count time series modeling with an application to influenza cases. **AStA Advances in Statistical Analysis**, Springer Science and Business Media LLC, v. 108, n. 3, p. 611–637, nov. 2023. ISSN 1863-818X. Disponível em: <http://dx.doi.org/10.1007/s10182-023-00488-6>. Citado nas páginas 25, 26, 27, 31, 34, 41 e 52.

ANDREWS, D. K.; BALAKRISHNA, N. **Bayesian estimation for novel geometric INGARCH model**. arXiv, 2024. Disponível em: <https://arxiv.org/abs/2410.01283>. Citado na página 25.

BALAKRISHNA, N.; ANVAR, P. M.; ABRAHAM, B. Zero-modified count time series with Markovian intensities. **Journal of Statistical Planning and Inference**, Elsevier BV, v. 229, p. 106092, mar. 2024. ISSN 0378-3758. Disponível em: <http://dx.doi.org/10.1016/j.jspi.2023.07.006>. Citado na página 26.

BENJAMIN, M. A.; RIGBY, R. A.; STASINOPOULOS, D. M. Generalized Autoregressive Moving Average Models. **Journal of the American Statistical Association**, Informa UK Limited, v. 98, n. 461, p. 214–223, mar. 2003. ISSN 1537-274X. Disponível em: <http://dx.doi.org/10.1198/016214503388619238>. Citado na página 25.

BETANCOURT, M. A Conceptual Introduction to Hamiltonian Monte Carlo. arXiv, 2017. Disponível em: <https://arxiv.org/abs/1701.02434>. Citado na página 45.

Binance. **Binance Cryptocurrency Exchange API Docs**. [S.l.], 2025. Acessado em: 23 dez 2025. Disponível em: <<https://binance-docs.github.io/apidocs/>>. Citado na página 68.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time Series Analysis**. Wiley, 2008. ISSN 1940-6347. ISBN 9781118619193. Disponível em: <<http://dx.doi.org/10.1002/9781118619193>>. Citado na página 64.

Brasil. Ministério da Saúde. DATASUS. **TabNet: Febre de Chikungunya - Notificações registradas no Sistema de Informação de Agravos de Notificação (SINAN)**. Brasília, DF: Ministério da Saúde, 2026. Acesso em: 01 jan. 2026. Disponível em: <<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/cnv/chikunbr.def>>. Citado na página 72.

BROOKS, S.; GELMAN, A.; JONES, G.; MENG, X.-L. Handbook of Markov Chain Monte Carlo. Chapman and Hall/CRC, maio 2011. Disponível em: <<http://dx.doi.org/10.1201/b10905>>. Citado nas páginas 45 e 46.

CARLIN, B.; LOUIS, T. **Bayes and Empirical Bayes Methods for Data Analysis, Second Edition**. Chapman and Hall/CRC, 2000. ISSN 2154-8137. ISBN 9781420057669. Disponível em: <<http://dx.doi.org/10.1201/9781420057669>>. Citado na página 50.

CARPENTER, B.; GELMAN, A.; HOFFMAN, M. D.; LEE, D.; GOODRICH, B.; BETANCOURT, M.; BRUBAKER, M.; GUO, J.; LI, P.; RIDDELL, A. Stan: A Probabilistic Programming Language. **Journal of Statistical Software**, Foundation for Open Access Statistics, v. 76, n. 1, 2017. ISSN 1548-7660. Disponível em: <<http://dx.doi.org/10.18637/jss.v076.i01>>. Citado na página 46.

CHEN, M.-H.; SHAO, Q.-M. Monte Carlo Estimation of Bayesian Credible and HPD Intervals. **Journal of Computational and Graphical Statistics**, Informa UK Limited, v. 8, n. 1, p. 69–92, mar. 1999. ISSN 1537-2715. Disponível em: <<http://dx.doi.org/10.1080/10618600.1999.10474802>>. Citado na página 52.

CHEN, T.; FOX, E. B.; GUESTRIN, C. **Stochastic Gradient Hamiltonian Monte Carlo**. [S.l.], 2014. Disponível em: <<https://arxiv.org/abs/1402.4102>>. Citado nas páginas 46 e 48.

CONCEIÇÃO, K. S.; ANDRADE, M. G.; LACHOS, V. H.; RAVISHANKER, N. Bayesian Inference for Zero-Modified Power Series Regression Models. **Mathematics**, MDPI AG, v. 13, n. 1, p. 60, dez. 2024. ISSN 2227-7390. Disponível em: <<http://dx.doi.org/10.3390/math13010060>>. Citado na página 26.

CONCEIÇÃO, K. S.; LOUZADA, F.; ANDRADE, M. G.; HELOU, E. S. Zero-modified power series distribution and its Hurdle distribution version. **Journal of Statistical Computation and Simulation**, Informa UK Limited, v. 87, n. 9, p. 1842–1862, fev. 2017. ISSN 1563-5163. Disponível em: <<http://dx.doi.org/10.1080/00949655.2017.1289529>>. Citado nas páginas 26, 30 e 31.

CONCEIÇÃO, K. S.; SUZUKI, A. K.; ANDRADE, M. G. A Bayesian approach for zero-modified Skellam model with Hamiltonian MCMC. **Statistical Methods and Applications**, Springer Science and Business Media LLC, v. 30, n. 2, p. 747–765, jul. 2020. ISSN 1613-981X. Disponível em: <<http://dx.doi.org/10.1007/s10260-020-00541-7>>. Citado nas páginas 26, 27, 30 e 41.

CORPORATION, M.; WESTON, S. **doParallel: Foreach Parallel Adaptor for the “parallel” Package**. The R Foundation, 2011. Disponível em: <<http://dx.doi.org/10.32614/CRAN.package.doParallel>>. Citado na página 56.

COX, D. R.; GUDMUNDSSON, G.; LINDGREN, G.; BONDESSON, L.; HARSAAE, E.; LAAKE, P.; JUSELIUS, K.; LAURITZEN, S. L. Statistical Analysis of Time Series: Some Recent Developments [with Discussion and Reply]. **Scandinavian Journal of Statistics**, [Board of the Foundation of the Scandinavian Journal of Statistics, Wiley], v. 8, n. 2, p. 93–115, 1981. ISSN 03036898, 14679469. Disponível em: <<http://www.jstor.org/stable/4615819>>. Citado na página 25.

DIETZ, E.; BÖHNING, D. On estimation of the Poisson parameter in zero-modified Poisson models. **Computational Statistics & Data Analysis**, Elsevier BV, v. 34, n. 4, p. 441–459, out. 2000. ISSN 0167-9473. Disponível em: <[http://dx.doi.org/10.1016/S0167-9473\(99\)00111-5](http://dx.doi.org/10.1016/S0167-9473(99)00111-5)>. Citado na página 26.

DUANE, S.; KENNEDY, A.; PENDLETON, B. J.; ROWETH, D. Hybrid Monte Carlo. **Physics Letters B**, Elsevier BV, v. 195, n. 2, p. 216–222, set. 1987. ISSN 0370-2693. Disponível em: <[http://dx.doi.org/10.1016/0370-2693\(87\)91197-X](http://dx.doi.org/10.1016/0370-2693(87)91197-X)>. Citado na página 44.

FERLAND, R.; LATOUR, A.; ORAICHI, D. Integer-Valued GARCH Process. **Journal of Time Series Analysis**, Wiley, v. 27, n. 6, p. 923–942, ago. 2006. ISSN 1467-9892. Disponível em: <<http://dx.doi.org/10.1111/j.1467-9892.2006.00496.x>>. Citado na página 26.

FOKIANOS, K.; KEDEM, B. Partial Likelihood Inference For Time Series Following Generalized Linear Models. **Journal of Time Series Analysis**, Wiley, v. 25, n. 2, p. 173–197, mar. 2004. ISSN 1467-9892. Disponível em: <<http://dx.doi.org/10.1046/j.0143-9782.2003.00344.x>>. Citado nas páginas 26 e 39.

FREELAND, R.; MCCABE, B. Forecasting discrete valued low count time series. **International Journal of Forecasting**, Elsevier BV, v. 20, n. 3, p. 427–434, jul. 2004. ISSN 0169-2070. Disponível em: <[http://dx.doi.org/10.1016/S0169-2070\(03\)00014-1](http://dx.doi.org/10.1016/S0169-2070(03)00014-1)>. Citado na página 52.

FREELAND, R. K. True integer value time series. **AStA Advances in Statistical Analysis**, Springer Science and Business Media LLC, v. 94, n. 3, p. 217–229, ago. 2010. ISSN 1863-818X. Disponível em: <<http://dx.doi.org/10.1007/s10182-010-0135-0>>. Citado na página 26.

GELFAND, A. E.; DEY, D. K.; CHANG, H. Model Determination using Predictive Distributions with Implementation via Sampling-Based Methods. In: _____. **Bayesian Statistics 4**. Oxford University Press Oxford, 1992. p. 147–168. ISBN 9781383023756. Disponível em: <<http://dx.doi.org/10.1093/oso/9780198522669.003.0009>>. Citado na página 50.

GELMAN, A.; GILKS, W. R.; ROBERTS, G. O. Weak convergence and optimal scaling of random walk Metropolis algorithms. **The Annals of Applied Probability**, Institute of Mathematical Statistics, v. 7, n. 1, fev. 1997. ISSN 1050-5164. Disponível em: <<http://dx.doi.org/10.1214/aoap/1034625254>>. Citado na página 44.

GELMAN, A.; RUBIN, D. B. Inference from Iterative Simulation Using Multiple Sequences. **Statistical Science**, Institute of Mathematical Statistics, v. 7, n. 4, nov. 1992. ISSN 0883-4237. Disponível em: <<http://dx.doi.org/10.1214/ss/1177011136>>. Citado na página 55.

- GEWEKE, J. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In: _____. **Bayesian Statistics 4**. Oxford University Press Oxford, 1992. p. 169–194. ISBN 9781383023756. Disponível em: <<http://dx.doi.org/10.1093/oso/9780198522669.003.0010>>. Citado na página 55.
- GUIKEMA, S. D.; GOFFELT, J. P. A Flexible Count Data Regression Model for Risk Analysis. **Risk Analysis**, Wiley, v. 28, n. 1, p. 213–223, fev. 2008. ISSN 1539-6924. Disponível em: <<http://dx.doi.org/10.1111/j.1539-6924.2008.01014.x>>. Citado na página 35.
- HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, Oxford University Press (OUP), v. 57, n. 1, p. 97–109, abr. 1970. ISSN 0006-3444. Disponível em: <<http://dx.doi.org/10.1093/biomet/57.1.97>>. Citado na página 43.
- HOFFMAN, M. D.; GELMAN, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. arXiv, 2011. Disponível em: <<https://arxiv.org/abs/1111.4246>>. Citado na página 46.
- HOLLA, M. S. On a poisson-inverse gaussian distribution. **Metrika**, Springer Science and Business Media LLC, v. 11, n. 1, p. 115–121, dez. 1967. ISSN 1435-926X. Disponível em: <<http://dx.doi.org/10.1007/BF02613581>>. Citado na página 37.
- HU, X.; ANDREWS, B. Integer-valued asymmetric garch modeling. **Journal of Time Series Analysis**, Wiley, v. 42, n. 5–6, p. 737–751, jun. 2021. ISSN 1467-9892. Disponível em: <<http://dx.doi.org/10.1111/jtsa.12605>>. Citado na página 26.
- IRWIN, J. O. The frequency distribution of the difference between two independent variates following the same Poisson distribution. **Journal of the Royal Statistical Society**, v. 100, n. n. 3, p. 415–416, 1937. Citado na página 38.
- JUNG, R. C.; TREMAYNE, A. Coherent forecasting in integer time series models. **International Journal of Forecasting**, Elsevier BV, v. 22, n. 2, p. 223–238, abr. 2006. ISSN 0169-2070. Disponível em: <<http://dx.doi.org/10.1016/j.ijforecast.2005.07.001>>. Citado na página 52.
- KACHOUR, M.; TRUQUET, L. A p-Order signed integer-valued autoregressive (SINAR(p)) model: A SIGNED INTEGER-VALUED AUTOREGRESSIVE MODEL. **Journal of Time Series Analysis**, Wiley, v. 32, n. 3, p. 223–236, nov. 2010. ISSN 0143-9782. Disponível em: <<http://dx.doi.org/10.1111/j.1467-9892.2010.00694.x>>. Citado na página 26.
- KARLIS, D.; NTZOUFRAS, I. Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. **IMA Journal of Management Mathematics**, Oxford University Press (OUP), v. 20, n. 2, p. 133–145, ago. 2008. ISSN 1471-6798. Disponível em: <<http://dx.doi.org/10.1093/imaman/dpn026>>. Citado na página 26.
- KIM, S.; KIM, H. A new metric of absolute percentage error for intermittent demand forecasts. **International Journal of Forecasting**, Elsevier BV, v. 32, n. 3, p. 669–679, jul. 2016. ISSN 0169-2070. Disponível em: <<http://dx.doi.org/10.1016/j.ijforecast.2015.12.003>>. Citado na página 56.
- KRÜGER, F.; LERCH, S.; THORARINSDOTTIR, T.; GNEITING, T. Predictive Inference Based on Markov Chain Monte Carlo Output. **International Statistical Review**, Wiley, v. 89, n. 2, p. 274–301, set. 2020. ISSN 1751-5823. Disponível em: <<http://dx.doi.org/10.1111/insr.12405>>. Citado na página 51.

LASTRA, K. Z.; PUMI, G.; PRASS, T. S. Order selection in GARMA models for count time series: a Bayesian perspective. **Journal of Applied Statistics**, Informa UK Limited, p. 1–25, mar. 2025. ISSN 1360-0532. Disponível em: <<http://dx.doi.org/10.1080/02664763.2025.2483309>>. Citado nas páginas 25, 26 e 27.

LI, Q.; CHEN, H.; ZHU, F. \mathbb{Z} -valued time series: Models, properties and comparison. **Journal of Statistical Planning and Inference**, Elsevier BV, v. 230, p. 106099, maio 2024. ISSN 0378-3758. Disponível em: <<http://dx.doi.org/10.1016/j.jspi.2023.106099>>. Citado na página 26.

LI, W. K. Time Series Models Based on Generalized Linear Models: Some Further Results. **Biometrics**, JSTOR, v. 50, n. 2, p. 506, jun. 1994. ISSN 0006-341X. Disponível em: <<http://dx.doi.org/10.2307/2533393>>. Citado na página 25.

METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equation of State Calculations by Fast Computing Machines. **The Journal of Chemical Physics**, AIP Publishing, v. 21, n. 6, p. 1087–1092, jun. 1953. ISSN 1089-7690. Disponível em: <<http://dx.doi.org/10.1063/1.1699114>>. Citado na página 43.

NOACK, A. A Class of Random Variables with Discrete Distributions. **The Annals of Mathematical Statistics**, Institute of Mathematical Statistics, v. 21, n. 1, p. 127–132, mar. 1950. ISSN 0003-4851. Disponível em: <<http://dx.doi.org/10.1214/aoms/1177729894>>. Citado na página 33.

PALA, L. O. d. O.; CARVALHO, M. d. M.; SáFADI, T. Analysis of Count Time Series: A Bayesian GARMA(p, q) Approach. **Austrian Journal of Statistics**, Austrian Statistical Society, v. 52, n. 5, p. 131–151, set. 2023. ISSN 1026-597X. Disponível em: <<http://dx.doi.org/10.17713/ajs.v52i5.1568>>. Citado nas páginas 26 e 27.

PATIL, G. P. Certain properties of the generalized power series distribution II. **Annals of the Institute of Statistical Mathematics**, Springer Science and Business Media LLC, v. 14, n. 1, p. 179–182, dez. 1962. ISSN 1572-9052. Disponível em: <<http://dx.doi.org/10.1007/BF02868639>>. Citado na página 33.

PLUMMER, M.; BEST, N.; COWLES, K.; VINES, K.; SARKAR, D.; BATES, D.; ALMOND, R.; MAGNUSSON, A. **coda: Output Analysis and Diagnostics for MCMC**. The R Foundation, 1999. Disponível em: <<http://dx.doi.org/10.32614/CRAN.package.coda>>. Citado na página 55.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: [s.n.], 2023. Disponível em: <<https://www.R-project.org/>>. Citado na página 27.

ROHDE, R. A.; HAUSFATHER, Z. The Berkeley Earth Land/Ocean Temperature Record. **Earth System Science Data**, Copernicus GmbH, v. 12, n. 4, p. 3469–3479, dez. 2020. ISSN 1866-3516. Disponível em: <<http://dx.doi.org/10.5194/essd-12-3469-2020>>. Citado na página 63.

SALDANHA, R. d. F.; BASTOS, R. R.; BARCELLOS, C. Microdatasus: pacote para download e pré-processamento de microdados do Departamento de Informática do SUS (DATASUS). **Cadernos de Saúde Pública**, FapUNIFESP (SciELO), v. 35, n. 9, 2019. ISSN 0102-311X. Disponível em: <<http://dx.doi.org/10.1590/0102-311X00032419>>. Citado na página 72.

SCHWARZ, G. Estimating the Dimension of a Model. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 6, n. 2, mar 1978. ISSN 0090-5364. Disponível em: <<http://dx.doi.org/10.1214/aos/1176344136>>. Citado na página 50.

- SHEPHARD, N. Generalized linear autoregressions. 1995. Citado na página 25.
- SHUMWAY, R. H.; STOFFER, D. S. **Time Series Analysis and Its Applications: With R Examples**. Springer International Publishing, 2017. ISSN 2197-4136. ISBN 9783319524528. Disponível em: <<http://dx.doi.org/10.1007/978-3-319-52452-8>>. Citado na página 64.
- SKELLAM, J. G. The frequency distribution of the difference between two Poisson variates belonging to different populations. **Journal of the Royal Statistical Society**, v. 109, n. Series A, p. 296, 1946. Citado na página 38.
- THOMAS, S.; TU, W. Learning Hamiltonian Monte Carlo in R. arXiv, 2020. Disponível em: <<https://arxiv.org/abs/2006.16194>>. Citado nas páginas 43, 44 e 46.
- UMBACH, D. On inference for a mixture of a poisson and a degenerate distribution. **Communications in Statistics - Theory and Methods**, Informa UK Limited, v. 10, n. 3, p. 299–306, jan. 1981. ISSN 1532-415X. Disponível em: <<http://dx.doi.org/10.1080/03610928108828039>>. Citado na página 26.
- WATANABE, S. **A Widely Applicable Bayesian Information Criterion**. arXiv, 2012. Disponível em: <<https://arxiv.org/abs/1208.6338>>. Citado na página 50.
- WESTON, S. **foreach: Provides Foreach Looping Construct**. The R Foundation, 2009. Disponível em: <<http://dx.doi.org/10.32614/CRAN.package.foreach>>. Citado na página 56.
- ZEGER, S. L. A regression model for time series of counts. **Biometrika**, Oxford University Press (OUP), v. 75, n. 4, p. 621–629, 1988. ISSN 1464-3510. Disponível em: <<http://dx.doi.org/10.1093/biomet/75.4.621>>. Citado na página 25.
- ZEGER, S. L.; QAQISH, B. Markov Regression Models for Time Series: A Quasi-Likelihood Approach. **Biometrics**, JSTOR, v. 44, n. 4, p. 1019, dez. 1988. ISSN 0006-341X. Disponível em: <<http://dx.doi.org/10.2307/2531732>>. Citado na página 25.

GRADIENTES DA FAMÍLIA PS

A.1 Poisson

Para a distribuição de Poisson, tem-se as derivadas $f_\mu(\mu_t, \nu) = e^{\mu_t}$, $g_\mu(\mu_t, \nu) = 1$, $a_\nu(y_t, \nu) = 0$, $f_\nu(\mu_t, \nu) = 0$ e $g_\nu(\mu_t, \nu) = 0$, dessa forma, a equação (4.7) fica

$$\frac{\partial \ell_2(\cdot)}{\partial \xi_k} = \sum_{t=p^*+1}^n (1 - I_{\{y_t=0\}}) \left[y_t \frac{1}{\mu_t} - \frac{e^{\mu_t}}{e^{\mu_t} - 1} \right] \frac{\partial \mu_t}{\partial \xi_k}.$$

Como a distribuição de Poisson não possui parâmetro de dispersão, a expressão (4.8) não é utilizada.

A.2 Poisson Generalizada

Para a distribuição Poisson Generalizada, as derivadas ficam

$$\begin{aligned} f_\mu(\mu_t, \nu) &= \frac{e^{\mu_t(1+\mu_t\nu)^{-1}}}{(1+\mu_t\nu)^2}, \\ g_\mu(\mu_t, \nu) &= \frac{e^{-\mu_t(1+\mu_t\nu)^{-1}}(\mu_t(\nu-1)+1)}{(\mu_t\nu+1)^3}, \\ a_\nu(y_t, \nu) &= \mu_t \frac{(y_t-1)(\mu_t\nu+1)^{y_t-2}}{y_t!}, \\ f_\nu(\mu_t, \nu) &= -\mu_t^2 \frac{e^{\mu_t(1+\mu_t\nu)^{-1}}}{(1+\mu_t\nu)^2} e \\ g_\nu(\mu_t, \nu) &= -\mu_t \frac{e^{-\mu_t(1+\mu_t\nu)^{-1}}(\mu_t(\nu-1)+1)}{(\mu_t\nu+1)^3}. \end{aligned}$$

De modo que a equação (4.7) é dada por

$$\frac{\partial \ell_2(\cdot)}{\partial \xi_k} = \sum_{t=p^*+1}^n \frac{1 - I_{\{y_t=0\}}}{(1+\mu_t\nu)^2} \left[y_t \frac{(\mu_t(\nu-1)+1)e^{\frac{\mu_t(\nu-1)}{1+\mu_t\nu}}}{\mu_t} - \frac{1}{1 - e^{-\frac{\mu_t}{1+\mu_t\nu}}} \right] \frac{\partial \mu_t}{\partial \xi_k}$$

e (4.8) por

$$\frac{\partial \ell_2(\cdot)}{\partial \mathbf{v}} = \sum_{t=p^*+1}^n (1 - I_{\{y_t=0\}}) \left[\mu_t (y_t - 1) \frac{(\mu_t \mathbf{v} + 1)^{y_t - 2}}{(1 + \mathbf{v} y_t)^{y_t - 1}} - y_t \frac{(\mu_t (\mathbf{v} - 1) + 1)}{(\mu_t \mathbf{v} + 1)^2} e^{\frac{\mu_t (\mathbf{v} - 1)}{1 + \mu_t \mathbf{v}}} - \frac{\mu_t (1 - \mu_t e^{\mu_t (1 + \mu_t \mathbf{v})^{-1}})}{(1 + \mu_t \mathbf{v})^2 (e^{\mu_t (1 + \mu_t \mathbf{v})^{-1}} - 1)} \right].$$

A.3 COM-Poisson

As derivadas da COM-Poisson são dadas por

$$\begin{aligned} f_{\mu}(\mu_t, \mathbf{v}) &= \mathbf{v} \sum_{s=0}^{\infty} s \frac{\mu_t^{vs-1}}{s!^{\mathbf{v}}}, \\ g_{\mu}(\mu_t, \mathbf{v}) &= \mathbf{v} \mu_t^{\mathbf{v}-1}, \\ a_{\mathbf{v}}(y_t, \mathbf{v}) &= -\log(y_t!) (y_t!)^{-\mathbf{v}}, \\ f_{\mathbf{v}}(\mu_t, \mathbf{v}) &= \sum_{s=0}^{\infty} \left(\frac{\mu_t^s}{s!} \right)^{\mathbf{v}} (s \log(\mu_t) - \log(s!)) \text{ e} \\ g_{\mathbf{v}}(\mu_t, \mathbf{v}) &= \mu_t^{\mathbf{v}} \log(\mu_t). \end{aligned}$$

Dessa forma, a equação (4.7) fica

$$\frac{\partial \ell_2(\cdot)}{\partial \xi_k} = \sum_{t=p^*+1}^n (1 - I_{\{y_t=0\}}) \left[y_t \frac{\mathbf{v}}{\mu_t} - \frac{\mathbf{v} \sum_{s=0}^{\infty} s \frac{\mu_t^{vs-1}}{s!^{\mathbf{v}}}}{\sum_{s=0}^{\infty} \left(\frac{\mu_t^s}{s!} \right)^{\mathbf{v}} - 1^{\mathbf{v}}} \right] \frac{\partial \mu_t}{\partial \xi_k}$$

e (4.4) é

$$\frac{\partial \ell_2(\cdot)}{\partial \mathbf{v}} = \sum_{t=p^*+1}^n (1 - I_{\{y_t=0\}}) \left[-\log(y_t!) + y_t \log(\mu_t) - \frac{\sum_{s=0}^{\infty} \left(\frac{\mu_t^s}{s!} \right)^{\mathbf{v}} (s \log(\mu_t) - \log(s!))}{\sum_{s=0}^{\infty} \left(\frac{\mu_t^s}{s!} \right)^{\mathbf{v}} - 1^{\mathbf{v}}} \right].$$

A.4 Binomial

Para a distribuição Binomial, tem-se as derivadas

$$\begin{aligned} f_{\mu}(\mu_t, \mathbf{v}) &= \binom{m}{m - \mu_t}^{m-1}, \\ g_{\mu}(\mu_t, \mathbf{v}) &= \frac{m}{(m - \mu_t)^2}, \end{aligned}$$

$a_v(y_t, v) = 0$, $f_v(\mu_t, v) = 0$ e $g_v(\mu_t, v) = 0$, dessa forma, a equação (4.7) fica

$$\frac{\partial \ell_2(\cdot)}{\partial \xi_k} = \sum_{t=p^*+1}^n (1 - I_{\{y_t=0\}}) \left[y_t \frac{m}{\mu_t(m - \mu_t)} - \frac{\left(\frac{m}{m - \mu_t}\right)^{m-1}}{\left(\frac{m}{m - \mu_t}\right)^m - m} \right] \frac{\partial \mu_t}{\partial \xi_k}.$$

Como a distribuição Binomial não apresenta parâmetro de dispersão, a expressão (4.8) não é utilizada.

A.5 Binomial Negativa

Para a distribuição Binomial Negativa as derivadas ficam

$$\begin{aligned} f_\mu(\mu_t, v) &= \left(\frac{v}{\mu_t + v}\right)^{1-v}, \\ g_\mu(\mu_t, v) &= \frac{v}{(\mu_t + v)^2}, \\ a_v(y_t, v) &= \frac{\Gamma(y_t + v)(\psi^{(0)}(y_t + v) - \psi^{(0)}(v))}{y_t! \Gamma(v)}, \\ f_v(\mu_t, v) &= \left(\frac{v}{\mu_t + v}\right)^{-v} \frac{-(\mu_t + v) \log\left(\frac{v}{\mu_t + v}\right) - \mu_t}{\mu_t + v} e \\ g_v(\mu_t, v) &= -\frac{\mu_t}{(\mu_t + v)^2}. \end{aligned}$$

Dessa forma, a equação (4.7) fica

$$\frac{\partial \ell_2(\cdot)}{\partial \xi_k} = \sum_{t=p^*+1}^n (1 - I_{\{y_t=0\}}) \left[y_t \frac{v}{\mu_t(\mu_t + v)} - \frac{\left(\frac{v}{\mu_t + v}\right)^{1-v}}{\left(\frac{v}{\mu_t + v}\right)^{-v} - \frac{\Gamma(y_t)}{y_t!}} \right] \frac{\partial \mu_t}{\partial \xi_k}$$

e (4.8) é, então,

$$\begin{aligned} \frac{\partial \ell_2(\cdot)}{\partial v} &= \sum_{t=p^*+1}^n (1 - I_{\{y_t=0\}}) \left[\psi^{(0)}(y_t + v) - \psi^{(0)}(v) + \frac{y_t}{\mu_t + v} \right. \\ &\quad \left. - \frac{\left(\frac{v}{\mu_t + v}\right)^{-v} \frac{-(\mu_t + v) \log\left(\frac{v}{\mu_t + v}\right) - \mu_t}{\mu_t + v}}{\left(\frac{v}{\mu_t + v}\right)^{-v} - \frac{\Gamma(y_t)}{y_t!}} \right]. \end{aligned}$$

