



**Universidade de Brasília**  
**Departamento de Estatística**

**Métodos de Estimação para Modelos Autorregressivos de Memória Variável  
Aplicados a Séries Temporais Climáticas**

**Davi Wentrick Feijó**

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília  
2025**

**Davi Wentrick Feijó**

**Métodos de Estimação para Modelos Autorregressivos de Memória Variável  
Aplicados a Séries Temporais Climáticas**

Orientador(a): Prof.Lucas Moreira

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília  
2025**

# Sumário

<b>1 Resumo</b>	5
<b>2 Abstract</b>	6
<b>3 Introdução</b>	7
<b>4 Fundamentação Teórica</b>	9
4.1 Processos Estocásticos	9
4.1.1 Processos Estacionários	9
4.1.2 Ergodicidade	10
4.1.3 Ruído Branco	11
4.2 Modelos de Séries Temporais	11
4.2.1 Modelos Auto-Regressivos (AR)	12
4.2.2 Modelos Auto-Regressivos de Médias Móveis (ARMA)	12
4.2.3 Modelos Auto-Regressivos Integrados de Médias Móveis (ARIMA)	13
4.2.4 Função de Autocovariância (FAC)	14
4.2.5 Função de Autocorrelação	15
4.2.6 Função de Autocorrelação do modelo AR( $p$ )	15
4.2.7 Função de Autocorrelação do modelo ARMA	16
4.2.8 Função de Autocorrelação Parcial (FACP)	17
4.2.9 Método de Box e Jenkins	18
4.3 Modelo TAR (Threshold Autoregressive)	19
4.4 Modelo SETAR (Self-Exciting Threshold Autoregressive)	20
4.5 Modelo AR-MV( $p$ )	20
4.5.1 Ergodicidade do Modelo AR-MV	21
4.5.2 Estimação do Modelo AR-MV com $\alpha$ conhecido	22
4.5.3 Estimação do Modelo AR-MV com $\alpha$ desconhecido	22
4.5.4 Algoritmo $k$ -means	23
4.6 Função de Verossimilhança	24
4.7 Previsão no Modelo AR-MV( $p$ )	25

<b>5 Metodologia</b> . . . . .	26
5.1 Geração de Dados Simulados do Modelo ARMV( $p$ ) . . . . .	26
5.2 Estimação dos Parâmetros e Seleção da Ordem do Modelo AR-MV( $p$ ) . . . . .	27
5.2.1 Iteração Alternada entre $\alpha$ e $\phi$ . . . . .	27
5.2.2 Seleção da Ordem $p$ via Teste de Razão de Verossimilhança . . . . .	28
5.3 Obtenção dos Dados Climáticos . . . . .	29
<b>6 Simulações</b> . . . . .	30
6.1 Simulação ARMV(6) . . . . .	31
6.1.1 Estimação dos parâmetros . . . . .	33
6.1.2 Estudo do ajuste do modelo ARMV(6) . . . . .	35
6.1.3 Estudo dos resíduos do modelo ARMV(6) . . . . .	36
6.1.4 Estudo preditivo do modelo ARMV(6) . . . . .	37
6.2 Simulação ARMV(7) . . . . .	39
6.2.1 Estimação dos parâmetros . . . . .	42
6.2.2 Estudo do ajuste do modelo . . . . .	45
6.2.3 Estudo dos resíduos do modelo ARMV(7) . . . . .	46
6.2.4 Estudo preditivo do modelo ARMV(7) . . . . .	47
<b>7 Aplicação do Modelo</b> . . . . .	49
7.1 Banco de Dados . . . . .	49
7.2 Serie de dados climáticos . . . . .	49
7.2.1 Estimação dos parâmetros . . . . .	52
7.2.2 Avaliação de ajuste do modelo . . . . .	54
7.2.3 Estudo dos resíduos do modelo ARMV(6) aplicado aos Dados Climáticos	55
7.2.4 Estudo preditivo do modelo . . . . .	57
<b>8 Conclusão</b> . . . . .	59

## 1 Resumo

Este trabalho tem como objetivo investigar e aplicar métodos de estimação para o modelo autorregressivo com memória variável (AR-MV), que permite que a ordem do modelo dependa do valor anterior da série. A partir de uma formulação geral do modelo AR-MV( $p$ ), são propostas abordagens para a estimação dos limiares e coeficientes por meio de um algoritmo iterativo que combina o método de  $k$ -médias com mínimos quadrados condicionais. As propriedades do processo de estimação são avaliadas por meio de estudos de simulação e medidas de desempenho, considerando critérios como AIC, BIC, MAE e MAPE. Adicionalmente, realiza-se uma comparação com modelos AR tradicionais. Por fim, o método proposto é aplicado a uma série temporal real de temperaturas médias mensais de Brasília (2001–2024), evidenciando sua capacidade de recuperar parâmetros adequados. Os resultados mostram que os modelos AR-MV ajustados a partir do processo de estimação desenvolvido apresentam desempenho superior aos modelos AR tradicionais, destacando a eficácia da metodologia proposta.

**Palavras-chave:** modelos autorregressivos; memória variável; estimação de limiares; séries temporais; modelos de regime.

## 2 Abstract

### Resumo

This work aims to investigate and apply estimation methods for the autoregressive model with variable memory (AR-MV), which allows the model order to depend on the previous value of the series. Based on a general formulation of the AR-MV( $p$ ) model, we propose approaches for estimating thresholds and coefficients through an iterative algorithm that combines the  $k$ -means method with conditional least squares. The estimation procedure's properties are evaluated through simulation studies and performance metrics, using criteria such as AIC, BIC, MAE, and MAPE. Additionally, a comparison with traditional AR models is performed. Finally, the proposed method is applied to a real monthly time series of average temperatures in Brasília (2001–2024), demonstrating its ability to recover suitable parameters. The results show that AR-MV models fitted using the proposed estimation process outperform traditional AR models, highlighting the effectiveness of the proposed methodology.

**Keywords:** autoregressive models; variable memory; threshold estimation; time series; regime-switching models.

### 3 Introdução

O estudo de séries temporais tem como principal objetivo modelar o comportamento de uma variável ao longo do tempo, diferenciando o que é aleatoriedade do que constitui informação útil para o modelo. Isso permite realizar inferências sobre a variável em estudo.

Em 1970, o método de Box e Jenkins formalizou técnicas para modelar, diagnosticar e prever séries temporais. Esse método introduziu os modelos autorregressivos integrados de médias móveis (ARIMA), que são lineares e deram origem a modelos derivados, como o Autorregressivo (AR) e o de Médias Móveis (MA).

No entanto, com o tempo, percebeu-se que esses modelos lineares frequentemente não eram suficientes para descrever todos os comportamentos observados, especialmente em séries financeiras. Essas séries apresentavam variância não constante ao longo do tempo (heterocedasticidade), o que comprometia os resultados obtidos pelos modelos lineares. Para lidar com essa limitação, surgiram os modelos não lineares ARCH e GARCH, cujo objetivo é modelar dados considerando a variância condicional.

Com o avanço dos modelos não lineares, surgiu o modelo TAR (Threshold Autoregressive), mencionado por Tong (1977) pela primeira vez. Esse modelo segmenta problemas não lineares em partes lineares, utilizando uma variável indicadora chamada limiar. A motivação para o desenvolvimento do TAR veio da necessidade de modelar fenômenos complexos que ainda não podiam ser descritos adequadamente por outros métodos, como no caso estudado por Tong, envolvendo a quantidade de linces capturados no noroeste do Canadá. O modelo SETAR (Self-Exciting Threshold Autoregressive Model), derivado dos modelos TAR, foi estudado por Tong e Lim (1980). Nesse modelo, a ideia central é que a variável que define o limiar seja uma função da defasagem da própria variável.

Posteriormente, o modelo AR-MV (Autorregressivo de Memória Variável) foi proposto por Fadel (2012) como um caso particular do modelo SETAR. Nesse modelo, a ordem da cadeia autorregressiva depende do intervalo em que o primeiro antecessor  $Y_{t-1}$  se encontra, de acordo com um parâmetro  $\alpha$  que define esses intervalos. Em Fadel (2012), a autora discute propriedades importantes, como a ergodicidade do modelo, a estimativa de parâmetros, intervalos de confiança e métodos de previsão. Tais aspectos foram avaliados por meio de estudos simulados, nos quais o modelo AR-MV foi comparado a outros modelos da classe AR. O foco principal foi na análise dos modelos AR-MV(2) e AR-MV(3).

Essa classe de modelos também foi estudada por Loureiro (2018), que analisou os anéis de crescimento de árvores como forma de estimar a idade das plantas. O autor buscou avaliar a propriedade de ergodicidade identificada por Fadel (2012) e conduziu

estudos simulados com os modelos AR-MV(4) e AR-MV(5).

Na dissertação de Rangel (2020), o autor apresenta um estudo sistemático dos modelos AR-MV de ordem 2 a 5. Além disso ele calcula a log-verossimilhança máxima de cada ajuste e compara os critérios AIC e BIC, identificando o modelo que melhor se ajusta aos dados.

Este trabalho tem como objetivo principal o estudo aprofundado dos modelos autorregressivos com mudanças de regime (AR-MV), com ênfase na estimação conjunta da ordem do modelo, dos limiares que definem os regimes e dos coeficientes autorregressivos associados, a partir exclusivamente dos dados observados. Além disso, investiga-se a forma de realizar essa estimação conjunta e a eficiência do procedimento adotado.

Para validar a abordagem proposta, realizam-se estudos de simulação nos quais os parâmetros verdadeiros são conhecidos, permitindo verificar a capacidade do método de recuperá-los adequadamente. Em seguida, os modelos são aplicados a uma série real de temperatura média diária registrada em Brasília no período de 2001 a 2024, proveniente de dados meteorológicos do INMET (Instituto Nacional de Meteorologia). Essa série apresenta características de variabilidade que tornam particularmente relevante a consideração de mudanças de regime, como padrões sazonais, tendências de longo prazo e flutuações de curta duração.

A motivação para a aplicação dos modelos AR-MV em dados climáticos reside na necessidade crescente de ferramentas estatísticas capazes de capturar e prever comportamentos complexos e não lineares, especialmente em um contexto de mudanças climáticas globais. Compreender como diferentes regimes de variabilidade térmica se estruturam ao longo do tempo pode fornecer conhecimentos valiosos para tomada de decisão em áreas como agricultura, energia, saúde e gestão de recursos hídricos. Assim, este trabalho visa contribuir para o desenvolvimento metodológico dos modelos AR-MV e para sua aplicação em contextos práticos de grande relevância.

## 4 Fundamentação Teórica

### 4.1 Processos Estocásticos

Os modelos utilizados para descrever séries temporais são processos estocásticos, ou seja, processos controlados por leis probabilísticas, conforme descrito por Morettin e Toloi (2006). A seguir, apresentamos a definição de processos estocásticos proposta pelos autores.

**Definição 1** *Seja  $T$  um conjunto arbitrário. Um processo estocástico é uma família  $Y = \{Y(t), t \in T\}$  tal que, para cada  $t \in T$ ,  $Y(t)$  é uma variável aleatória.*

Um processo estocástico pode ser definido como uma família de variáveis aleatórias (v.a.) estabelecidas em um mesmo espaço de probabilidade  $(\Omega, \mathcal{A}, \mathcal{P})$ . O conjunto de índices  $T$  geralmente é considerado como o conjunto dos inteiros  $T = \{0, \pm 1, \pm 2, \dots\}$  ou o subconjunto dos reais  $\mathbb{R}$ . Para cada  $t \in T$ , a variável  $Y(t)$  é uma v.a. real, e o conjunto de valores  $\{Y(t), t \in T\}$  define o espaço dos estados  $S$  do processo, sendo os valores individuais de  $Y(t)$  denominados estados. Além disso, o processo  $Y = \{Y(t), t \in T\}$  está completamente caracterizado quando são conhecidas todas as distribuições finito-dimensionais associadas, descritas por

$$F(y_1, \dots, y_n; t_1, \dots, t_n) = P\{Y(t_1) \leq y_1, \dots, Y(t_n) \leq y_n\}, \quad \text{com } y_i \in \mathbb{R}, t_i \in T.$$

#### 4.1.1 Processos Estacionários

Os conceitos de estacionariedade estrita e estacionariedade fraca serão apresentados conforme a definição de Morettin e Toloi (2006).

**Definição 2** *Um processo estocástico  $Y = \{Y(t), t \in T\}$  diz-se estritamente estacionário se todas as distribuições finito-dimensionais permanecem as mesmas sob translações no tempo, ou seja,*

$$F(y_1, \dots, y_n; t_1 + \tau, \dots, t_n + \tau) = F(y_1, \dots, y_n; t_1, \dots, t_n)$$

*para quaisquer  $t_1, \dots, t_n, \tau \in T$  e  $y_i \in \mathbb{R}$*

Nesse contexto, o parâmetro  $\tau$  representa um deslocamento (ou translação) no eixo temporal. A definição afirma que a distribuição conjunta das variáveis  $Y(t_1), \dots, Y(t_n)$

não se altera quando todos os instantes de tempo  $t_i$  são deslocados por uma mesma quantidade  $\tau$ . Isso implica, em particular, que todas as distribuições unidimensionais são invariantes sob translações no tempo. Consequentemente, a média  $\mu(t)$  e a variância  $V(t)$  são constantes, isto é,

$$\mu(t) = \mu, \quad V(t) = \sigma^2, \quad \text{para todo } t \in T.$$

**Definição 3** Um processo estocástico  $Y = \{Y(t), t \in T\}$  diz-se fracamente estacionário, ou estacionário de segunda ordem (ou em sentido amplo), se e somente se

- (i)  $\mathbb{E}[Y(t)] = \mu(t) = \mu$ , constante, para todo  $t \in T$ ;
- (ii)  $\mathbb{E}[Y^2(t)] < \infty$ , para todo  $t \in T$ ;
- (iii)  $\gamma(t_1, t_2) = \text{Cov}(Y(t_1), Y(t_2))$  depende apenas da defasagem  $|t_1 - t_2|$  para todo  $t_i \in T$

#### 4.1.2 Ergodicidade

Conforme definido por Bueno (2011), um processo ergódico pode ser descrito da seguinte forma:

**Definição 4** Um processo fracamente estacionário é ergódico para o primeiro momento se

$$\mathbb{E}(\bar{Y}^{(s)}) = p \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T Y_t^{(s)} = p \lim_{S \rightarrow \infty} \frac{1}{S} \sum_{s=1}^S Y_t^{(s)} = \mathbb{E}(Y_t),$$

em que  $\bar{Y}^{(s)}$  representa a média temporal da  $s$ -ésima realização do processo, e  $p\lim$  indica convergência em probabilidade.

Isso significa que a média amostral converge em probabilidade para o valor esperado populacional. Dessa forma, mesmo com apenas uma realização do processo, é possível obter uma estimativa consistente da média.

**Definição 5** Um processo fracamente estacionário é ergódico para o segundo momento se

$$\frac{1}{T-j} \sum_{t=1}^T (Y_t - \mu)(Y_{t-j} - \mu) \xrightarrow{P} \gamma_j, \quad \text{para todo } j,$$

em que  $\xrightarrow{P}$  denota convergência em probabilidade.

A função  $\gamma_j$  representa a autocovariância teórica do processo com defasagem  $j$ , definida como  $\gamma_j = \text{Cov}(Y_t, Y_{t-j}) = \mathbb{E}[(Y_t - \mu)(Y_{t-j} - \mu)]$ . Em um processo fracamente

estacionário, essa função depende apenas da defasagem  $j$  e não do tempo absoluto  $t$ , refletindo a estrutura de dependência temporal do processo. A ergodicidade para o segundo momento garante que essa autocovariância possa ser estimada de forma consistente com base em uma única realização do processo.

#### 4.1.3 Ruído Branco

**Definição 6** (MORETTIN; TOLOI, 2006) Dizemos que  $\{\varepsilon_t, t \in \mathbb{Z}\}$  é um ruído branco discreto se as variáveis aleatórias  $\varepsilon_t$  são não correlacionadas, isto é,

$$\text{Cov}(\varepsilon_t, \varepsilon_s) = 0, \quad \text{para } t \neq s.$$

Um processo de ruído branco será estacionário se satisfizer

$$\mathbb{E}[\varepsilon_t] = \mu_\varepsilon \quad \text{e} \quad \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, \quad \text{para todo } t \in \mathbb{Z}.$$

Neste trabalho, assumiremos sem perda de generalidade que  $\mu_\varepsilon = 0$ . Assim, indicamos brevemente

$$\varepsilon_t \sim \text{RB}(0, \sigma_\varepsilon^2).$$

Além disso, se as variáveis  $\varepsilon_t$  forem independentes e identicamente distribuídas (i.i.d.), então o processo é denominado puramente aleatório. Neste caso, escrevemos

$$\varepsilon_t \sim \text{i.i.d. } (0, \sigma_\varepsilon^2).$$

## 4.2 Modelos de Séries Temporais

Nesta seção, revisamos os principais modelos de séries temporais clássicos, cujas definições e propriedades seguem de forma detalhada em Morettin e Toloi (2006). Inicialmente, definimos o modelo AR( $p$ ), em que a observação presente é expressa como combinação linear de  $p$  defasagens do processo mais um termo de erro não correlacionado. Na sequência, introduzimos o modelo MA( $q$ ), no qual a série é representada como soma ponderada de  $q$  choques passados. A combinação destes dois paradigmas dá origem ao modelo ARMA( $p, q$ ), que capta simultaneamente dependências autorregressivas e de média móvel. Por fim, discutimos a generalização ARIMA( $p, d, q$ ), que incorpora diferenciação de ordem  $d$  para lidar com não-estacionariedade na média. Em seguida, apresentam-se as funções de autocorrelação (FAC) e de autocorrelação parcial (FACP), instrumentos

fundamentais para identificar padrões de dependência em séries temporais, e introduz-se o método Box–Jenkins, que estrutura em três etapas — identificação, estimativa e diagnóstico — o processo de modelagem ARIMA.

#### 4.2.1 Modelos Auto-Regressivos (AR)

(MORETTIN; TOLOI, 2006) Considere a representação geral de um processo auto-regressivo infinito, dada por

$$Y_t = \sum_{j=1}^{\infty} \pi_j Y_{t-j} + a_t, \quad (4.2.1)$$

em que  $\{a_t\}$  é um ruído branco com média zero e variância constante, e  $\pi_j$  são os coeficientes da representação auto-regressiva.

Dizemos que o processo é um *modelo auto-regressivo de ordem p*, denotado por AR( $p$ ), quando  $\pi_j = 0$  para todo  $j > p$ . Nesse caso, a representação infinita se reduz a uma forma finita

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + a_t, \quad (4.2.2)$$

em que renomeamos os coeficientes  $\pi_j$  como  $\phi_j$ , para  $j = 1, \dots, p$ , segundo a notação usual

$$\pi_j = \begin{cases} \phi_j, & \text{se } 1 \leq j \leq p, \\ 0, & \text{se } j > p. \end{cases}$$

Se definirmos o *operador auto-regressivo estacionário de ordem p* como

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \quad (4.2.3)$$

então o modelo pode ser reescrito de forma compacta como

$$\phi(B)Y_t = a_t, \quad (4.2.4)$$

em que  $B$  é o operador defasagem (backshift), tal que  $BY_t = Y_{t-1}$ .

#### 4.2.2 Modelos Auto-Regressivos de Médias Móveis (ARMA)

(MORETTIN; TOLOI, 2006) Para muitas séries temporais encontradas na prática, o uso exclusivo de modelos auto-regressivos ou de médias móveis pode exigir um número elevado de parâmetros. Nesse contexto, a combinação de ambos os componentes representa uma alternativa eficaz para modelagem, resultando nos chamados modelos ARMA( $p, q$ ).

A forma geral de um modelo ARMA de ordem  $(p, q)$  é dada por

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}, \quad (4.2.5)$$

- $\phi_j \in \mathbb{R}$ , para  $j = 1, \dots, p$ : coeficientes auto-regressivos;
- $\theta_j \in \mathbb{R}$ , para  $j = 1, \dots, q$ : coeficientes de média móvel;
- $\{a_t\}$  é um ruído branco com  $\mathbb{E}(a_t) = 0$  e  $\text{Var}(a_t) = \sigma_a^2$ .

Utilizando os operadores de defasagem, definimos

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \quad (4.2.6)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q. \quad (4.2.7)$$

Dessa forma, o modelo ARMA( $p, q$ ) pode ser escrito de maneira compacta como

$$\phi(B) Y_t = \theta(B) a_t. \quad (4.2.8)$$

#### 4.2.3 Modelos Auto-Regressivos Integrados de Médias Móveis (ARIMA)

Considere um processo  $\{Y_t\}$  que não seja estacionário. Definimos o operador de diferença como

$$\Delta Y_t = Y_t - Y_{t-1} \quad \text{e, de modo recursivo,} \quad \Delta^d Y_t = \Delta(\Delta^{d-1} Y_t) \quad (d \geq 1).$$

Se a  $d$ -ésima diferença do processo, denotada por  $W_t = \Delta^d Y_t$ , for estacionária, então  $W_t$  pode ser representado por um modelo ARMA( $p, q$ ), isto é

$$\phi(B) W_t = \theta(B) a_t, \quad (4.2.9)$$

com  $\phi(B)$  e  $\theta(B)$  sendo os polinômios auto-regressivo e de médias móveis em  $B$ , e  $\{a_t\} \sim \text{RB}(0, \sigma_a^2)$ .

Como  $W_t = \Delta^d Y_t$ , o processo  $Y_t$  pode ser descrito por um modelo autorregressivo integrado de médias móveis, ou ARIMA( $p, d, q$ ), cuja equação é:

$$\phi(B) \Delta^d Y_t = \theta(B) a_t, \quad (4.2.10)$$

ou, equivalentemente,

$$\varphi(B) Y_t = \theta(B) a_t, \quad (4.2.11)$$

em que

$$\varphi(B) = \phi(B)(1 - B)^d$$

é um polinômio auto-regressivo de ordem  $p+d$  com  $d$  raízes unitárias (no círculo unitário).

Diz-se que  $Y_t$  segue um modelo ARIMA( $p, d, q$ ) se

- $p$  é a ordem da parte autorregressiva;
- $d$  é o número de diferenciações necessárias para tornar o processo estacionário;
- $q$  é a ordem da parte de média móvel.

Na prática, costuma-se usar  $d = 0, 1$  ou  $2$ , suficientes para capturar os principais tipos de não-estacionariedade

- (a) *Não-estacionariedade em nível*: oscilações em torno de níveis médios distintos ao longo do tempo;
- (b) *Não-estacionariedade em inclinação*: presença de tendência persistente de alta ou baixa.

Quando  $d = 0$ , o ARIMA( $p, 0, q$ ) reduz-se ao processo estacionário ARMA( $p, q$ ).

#### 4.2.4 Função de Autocovariância (FAC)

Seja  $\{Y_t, t \in \mathbb{Z}\}$  um processo estocástico real, discreto e estritamente estacionário, com média nula, isto é,  $\mathbb{E}[Y_t] = 0$  para todo  $t \in \mathbb{Z}$ . A função de autocovariância (FACV) é definida como

$$\gamma_\tau = \mathbb{E}[Y_t Y_{t+\tau}], \quad \tau \in \mathbb{Z}.$$

A FACV mede o grau de dependência linear entre valores do processo separados por uma defasagem  $\tau$ , e satisfaz as seguintes propriedades

1.  $\gamma_0 \geq 0$ ;
2.  $\gamma_{-\tau} = \gamma_\tau$  (simetria);
3.  $|\gamma_\tau| \leq \gamma_0$ ;
4.  $\gamma_\tau$  é não negativamente definida, no sentido de que, para quaisquer inteiros  $\tau_1, \dots, \tau_n \in \mathbb{Z}$  e quaisquer números reais  $a_1, \dots, a_n, n \geq 1$ , vale:

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma_{\tau_j - \tau_k} \geq 0.$$

Além disso, é comum que a FACV de um processo estacionário converja para zero quando  $|\tau| \rightarrow \infty$ , refletindo a perda de dependência linear ao longo do tempo.

#### 4.2.5 Função de Autocorrelação

A função de autocorrelação (FAC) do processo é obtida pela normalização da autocovariância pela variância do processo, sendo definida como

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}, \quad \tau \in \mathbb{Z}.$$

A FAC possui as mesmas propriedades estruturais da FACV, com a diferença de que agora

$$\rho_0 = 1, \quad \text{e} \quad |\rho_\tau| \leq 1 \quad \text{para todo } \tau \in \mathbb{Z}.$$

#### 4.2.6 Função de Autocorrelação do modelo AR(p)

Multiplicando ambos os lados da equação do modelo AR(p) (4.2.2),

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + a_t,$$

por  $Y_{t-j}$  e tomando a esperança, obtemos

$$E(Y_t Y_{t-j}) = \phi_1 E(Y_{t-1} Y_{t-j}) + \phi_2 E(Y_{t-2} Y_{t-j}) + \cdots + \phi_p E(Y_{t-p} Y_{t-j}) + E(a_t Y_{t-j}).$$

Como  $a_t$  é ruído branco e não está correlacionado com  $Y_{t-j}$  para  $j > 0$ , tem-se

$$E(a_t Y_{t-j}) = 0.$$

Logo, a função de autocovariância  $\gamma_j = \text{Cov}(Y_t, Y_{t-j})$  satisfaaz a equação de re-corrência

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \cdots + \phi_p \gamma_{j-p}, \quad \text{para } j > 0. \quad (4.2.12)$$

Dividindo por  $\gamma_0 = \text{Var}(Y_t)$ , obtemos a função de autocorrelação

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \cdots + \phi_p \rho_{j-p}, \quad \text{para } j > 0. \quad (4.2.13)$$

Para  $j = 0$ , temos:

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \cdots + \phi_p \gamma_p + \sigma_a^2,$$

e como  $\gamma_{-j} = \gamma_j$ , resulta que:

$$1 = \phi_1\rho_1 + \phi_2\rho_2 + \cdots + \phi_p\rho_p + \frac{\sigma_a^2}{\gamma_0}, \quad (4.2.14)$$

ou, equivalentemente:

$$\text{Var}(Y_t) = \gamma_0 = \frac{\sigma_a^2}{1 - \phi_1\rho_1 - \cdots - \phi_p\rho_p}. \quad (4.2.15)$$

As equações da FAC para  $j = 1, 2, \dots, p$ , conhecidas como equações de Yule-Walker, são

$$\rho_1 = \phi_1 + \phi_2\rho_1 + \cdots + \phi_p\rho_{p-1}, \quad (4.2.16)$$

$$\rho_2 = \phi_1\rho_1 + \phi_2 + \cdots + \phi_p\rho_{p-2}, \quad (4.2.17)$$

$$\vdots \quad (4.2.18)$$

$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \cdots + \phi_p. \quad (4.2.19)$$

Em forma matricial, podemos escrevê-las como

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}$$

Os coeficientes  $\phi_1, \dots, \phi_p$  do modelo AR(p) podem ser estimados substituindo-se as autocorrelações  $\rho_j$  por suas estimativas amostrais  $r_j$ , conforme definido anteriormente.

#### 4.2.7 Função de Autocorrelação do modelo ARMA

Multiplicando ambos os membros da equação do modelo ARMA(p, q) (4.2.5)

$$\tilde{Z}_t = \phi_1\tilde{Z}_{t-1} + \cdots + \phi_p\tilde{Z}_{t-p} + a_t - \theta_1a_{t-1} - \cdots - \theta_qa_{t-q},$$

por  $\tilde{Z}_{t-j}$  e tomindo a esperança, obtemos

$$\gamma_j = E(\tilde{Z}_t\tilde{Z}_{t-j}) = E \left[ \left( \sum_{i=1}^p \phi_i \tilde{Z}_{t-i} + a_t - \sum_{k=1}^q \theta_k a_{t-k} \right) \tilde{Z}_{t-j} \right].$$

Desenvolvendo a expressão, temos

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \cdots + \phi_p \gamma_{j-p} + \gamma_{Za}(j) - \theta_1 \gamma_{Za}(j-1) - \cdots - \theta_q \gamma_{Za}(j-q),$$

em que  $\gamma_{Za}(j) = E(a_t \tilde{Z}_{t-j})$  é a covariância cruzada entre o ruído branco  $a_t$  e o processo  $\tilde{Z}_{t-j}$ .

Sabemos que  $\tilde{Z}_{t-j}$  depende apenas de choques anteriores a  $a_t$  até o tempo  $t-j$ .

Assim

$$\gamma_{Za}(j) = \begin{cases} 0, & j > 0, \\ \neq 0, & j \leq 0. \end{cases}$$

Portanto, para  $j > q$ , a equação da função de autocovariância se reduz a

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \cdots + \phi_p \gamma_{j-p}, \quad j > q,$$

o que mostra que, a partir da defasagem  $j = q+1$ , a função de autocovariância segue a mesma estrutura de um modelo AR(p).

A função de autocorrelação (FAC) é então obtida como

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \cdots + \phi_p \rho_{j-p}, \quad j > q,$$

mostrando que as autocorrelações de defasagens  $j = 1, \dots, q$  são diretamente influenciadas pelos parâmetros de médias móveis. Para  $j > q$ , as autocorrelações comportam-se como nos modelos auto-regressivos, caracterizando a natureza mista do modelo ARMA.

#### 4.2.8 Função de Autocorrelação Parcial (FACP)

Box e Jenkins (1970) propõem a utilização de um instrumento adicional no procedimento de identificação de modelos de séries temporais: a *função de autocorrelação parcial* (FACP). Essa função é útil para determinar a ordem apropriada de modelos autoregressivos (AR), de médias móveis (MA) ou ARMA, a partir do comportamento das autocorrelações observadas.

Seja  $\hat{\phi}_{kj}$  o  $j$ -ésimo coeficiente estimado por mínimos quadrados de um modelo AR( $k$ ), e  $\hat{\phi}_{kk}$  o último coeficiente estimado. Então, a função de autocorrelação parcial é definida como:

$$\hat{\phi}_{kk} = \text{último coeficiente do modelo AR}(k).$$

Sabemos que os coeficientes  $\hat{\phi}_{kj}$  satisfazem as equações de Yule–Walker:

$$\rho_j = \hat{\phi}_{k1} \rho_{j-1} + \hat{\phi}_{k2} \rho_{j-2} + \cdots + \hat{\phi}_{kk} \rho_{j-k}, \quad j = 1, \dots, k,$$

das quais obtemos o seguinte sistema linear:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \hat{\phi}_{k1} \\ \hat{\phi}_{k2} \\ \vdots \\ \hat{\phi}_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}$$

Resolvendo esse sistema sucessivamente para  $k = 1, 2, 3, \dots$ , obtemos

$$\begin{aligned} \hat{\phi}_{11} &= \rho_1, \\ \hat{\phi}_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \\ \hat{\phi}_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}. \end{aligned}$$

De forma geral, tem-se

$$\hat{\phi}_{kk} = \frac{|\mathbf{P}_k^*|}{|\mathbf{P}_k|},$$

onde  $\mathbf{P}_k$  é a matriz de autocorrelações de ordem  $k$  e  $\mathbf{P}_k^*$  é a matriz obtida substituindo a última coluna de  $\mathbf{P}_k$  pelo vetor  $[\rho_1, \rho_2, \dots, \rho_k]^\top$ .

A função  $\hat{\phi}_{kk}$ , encarada como função de  $k$ , é a chamada *função de autocorrelação parcial*.

#### 4.2.9 Método de Box e Jenkins

Segundo Morettin e Toloi (2006) uma metodologia bastante utilizada na análise de modelos paramétricos é conhecida como abordagem de Box e Jenkins (1970). Tal metodologia consiste em ajustar modelos auto-regressivos integrados de médias móveis,

ARIMA( $p, d, q$ ), a um conjunto de dados.

A estratégia para a construção do modelo será baseada em um ciclo iterativo, no qual a escolha da estrutura do modelo é baseada nos próprios dados. Os estágios do ciclo iterativo são

- (a) uma classe geral de modelos é considerada para a análise (especificação);
- (b) há identificação de um modelo, com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
- (c) a seguir vem a fase de estimação, na qual os parâmetros do modelo identificado são estimados;
- (d) finalmente, há a verificação ou diagnóstico do modelo ajustado, através de uma análise de resíduos, para se saber se este é adequado para os fins em vista (previsão, por exemplo).

Caso o modelo não seja adequado, o ciclo é repetido, voltando-se à fase de identificação. Um procedimento que muitas vezes é utilizado é identificar não só um único modelo, mas alguns modelos que serão então estimados e verificados. Se o propósito é previsão, escolher-se-á entre os modelos ajustados o melhor, por exemplo, no sentido de fornecer o menor erro quadrático médio de previsão.

### 4.3 Modelo TAR (Threshold Autoregressive)

Os modelos autorregressivos com limiar (TAR) surgiram da necessidade de descrever séries temporais que exibem comportamentos não lineares e mudanças de regime, eventos em que a dinâmica do processo muda de acordo com o nível da própria série. Essa abordagem permite capturar assimetrias, ciclos limites e respostas distintas a inovações de acordo com a região do espaço de estados em que o sistema se encontra (TONG, 1978).

Seja  $\{Y_t\}_{t \in \mathbb{Z}}$  uma série temporal de variáveis aleatórias com valores em  $\mathbb{R}$ . Definimos (TONG; LIM, 1980):

- $\ell \in \mathbb{N}$ , número de regimes.
- Limiares  $\{\alpha_j\}_{j=0}^{\ell}$  tais que

$$-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_{\ell-1} < \alpha_{\ell} = +\infty, \quad \alpha_j \in \mathbb{R}.$$

- Atraso de transição  $d \in \mathbb{N}_+$ , indicando que o regime em  $t$  é determinado por  $Y_{t-d}$ .
- Ordens autorregressivas por regime  $p_j \in \mathbb{N}_0$ , para  $j = 1, \dots, \ell$ .

- Coeficientes regime-dependentes  $\phi_{j,0} \in \mathbb{R}$  (intercepto) e  $\phi_{j,i} \in \mathbb{R}$ ,  $i = 1, \dots, p_j$ .
- Ruído branco  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  com  $\mathbb{E}[\varepsilon_t] = 0$ ,  $\text{Var}(\varepsilon_t) = \sigma_j^2 < \infty$  quando no regime  $j$ .

O modelo TAR com  $\ell$  regimes escreve-se, para todo  $t \in \mathbb{Z}$ , como (??):

$$Y_t = \phi_{j,0} + \sum_{i=1}^{p_j} \phi_{j,i} Y_{t-i} + \varepsilon_t, \quad \text{se } \alpha_{j-1} < Y_{t-d} \leq \alpha_j, \quad j = 1, \dots, \ell.$$

Aqui cada regime  $j$  aplica-se sempre que  $Y_{t-d}$  cai no intervalo  $(\alpha_{j-1}, \alpha_j]$ .

#### 4.4 Modelo SETAR (Self-Exciting Threshold Autoregressive)

O modelo SETAR constitui uma instância particular de TAR em que a própria série determina automaticamente as transições de regime, tornando explícita a relação entre inovações passadas e mudanças na dinâmica do processo. Graças à sua formulação simples e interpretável, o SETAR é amplamente adotado para capturar regimes auto-induzidos (TSAY, 1989).

Neste caso, consideramos um único limiar  $\alpha \in \mathbb{R}$  e dois regimes ( $\ell = 2$ ). Definimos:

- Série  $\{Y_t\}_{t \in \mathbb{Z}} \subset \mathbb{R}$ .
- Limiar único  $\alpha \in \mathbb{R}$  e atraso  $d \in \mathbb{N}_+$ .
- Ordens  $p_1, p_2 \in \mathbb{N}_0$  para regimes baixo ( $j = 1$ ) e alto ( $j = 2$ ).
- Coeficientes  $\phi_{j,0} \in \mathbb{R}$  e  $\phi_{j,i} \in \mathbb{R}$ ,  $i = 1, \dots, p_j$ .
- Ruídos  $\{\varepsilon_t^{(j)}\}$  regime-dependentes, com  $\mathbb{E}[\varepsilon_t^{(j)}] = 0$ ,  $\text{Var}(\varepsilon_t^{(j)}) = \sigma_j^2 < \infty$ .

No SETAR de dois regimes e atraso  $d$ , para cada  $t$  (??):

$$Y_t = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^{p_1} \phi_i^{(1)} Y_{t-i} + \varepsilon_t^{(1)}, & \text{se } Y_{t-d} \leq \alpha, \\ \phi_0^{(2)} + \sum_{i=1}^{p_2} \phi_i^{(2)} Y_{t-i} + \varepsilon_t^{(2)}, & \text{se } Y_{t-d} > \alpha. \end{cases}$$

#### 4.5 Modelo AR-MV(p)

O modelo AR-MV(p), também conhecido como modelo autorregressivo com memória variável, proposto por Fadel (2012) é uma extensão do modelo SETAR com profundidade

de defasagem  $d = 1$  e variação de coeficientes dependente de estados definidos por limiares. Nesse modelo, os coeficientes autorregressivos variam conforme o valor da observação passada  $Y_{t-1}$ , o que confere flexibilidade para capturar diferentes dinâmicas em distintos regimes.

A forma geral do modelo AR-MV(p) é dada por

$$Y_t = \sum_{i=1}^{L_t} \phi_i Y_{t-i} + \epsilon_t,$$

em que  $\{\epsilon_t\}$  é uma sequência de ruídos brancos independentes e identicamente distribuídos com distribuição Normal( $0, \sigma^2$ ), e  $\epsilon_t$  é independente de  $Y_{t-s}$  para todo  $s \geq 1$ .

A variável  $L_t$ , que determina a quantidade de defasagens em cada instante  $t$ , é definida conforme a seguinte regra

$$L_t = \begin{cases} 1 & \text{se } \alpha_0 < Y_{t-1} \leq \alpha_1, \\ 2 & \text{se } \alpha_1 < Y_{t-1} \leq \alpha_2, \\ \vdots & \\ p & \text{se } \alpha_{p-1} < Y_{t-1} \leq \alpha_p, \end{cases}$$

com os limiares  $\{\alpha_0, \alpha_1, \dots, \alpha_p\}$  obedecendo as condições

$$\alpha_0 = -\infty, \quad \alpha_p = \infty.$$

Esses limiares definem uma partição do conjunto dos números reais

$$\mathbb{R} = \bigcup_{j=0}^{p-1} A_j, \quad \text{com } A_j = (\alpha_j, \alpha_{j+1}], \quad j = 0, 1, \dots, p-1.$$

Esse modelo permite que a estrutura autorregressiva se adapte dinamicamente ao comportamento da série temporal, tornando-o particularmente útil para capturar mudanças de regime.

#### 4.5.1 Ergodicidade do Modelo AR-MV

Segundo Fadel (2012), o modelo AR-MV( $p$ ) será geometricamente ergódico sob a seguinte condição

**Proposição 4.1** *Se  $\sum_{i=1}^p |\phi_i| < 1$ , então o modelo AR-MV( $p$ ) é geometricamente ergódico.*

A demonstração dessa proposição baseia-se na reescrita do modelo em função de uma transformação da equação (3.1) e na aplicação de resultados clássicos sobre processos estocásticos com memória. Para o desenvolvimento completo da prova e argumentos formais, consulte Fadel (2012, Capítulo 3).

#### 4.5.2 Estimação do Modelo AR-MV com $\alpha$ conhecido

No caso em que  $\alpha$  é conhecido, sob o modelo AR-MV( $p$ ), os estimadores de mínimos quadrados condicionais (CLS) e de máxima verossimilhança para  $\phi = (\phi_1, \dots, \phi_p)^\top$  são equivalentes e dados por:

$$\hat{\phi}(\alpha) = (\mathbf{Y}_\alpha^{*\top} \mathbf{Y}_\alpha^*)^{-1} \mathbf{Y}_\alpha^{*\top} \mathbf{Y} \quad (4.5.1)$$

em que  $\mathbf{Y}_\alpha^*$  é a matriz de defasagens ponderadas pelo limiar  $\alpha$ , definida como

$$\mathbf{Y}_\alpha^* = \begin{bmatrix} Y_p & Y_{p+1}(\alpha_1) & \dots & Y_{p+1}(\alpha_{p-1}) \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n-1} & Y_n(\alpha_1) & \dots & Y_n(\alpha_{p-1}) \end{bmatrix}, \quad (4.5.2)$$

onde  $Y_t(\alpha_j) = Y_{t-(j+1)} \cdot \mathbb{I}(Y_{t-1} > \alpha_j)$ , com  $\mathbb{I}(\cdot)$  representando a função indicadora.

O modelo AR-MV também pode ser representado matricialmente da seguinte forma

$$\mathbf{Y} = \begin{bmatrix} Y_{p+1} \\ \vdots \\ Y_n \end{bmatrix} = \mathbf{Y}_\alpha^* \phi + \begin{bmatrix} \epsilon_{p+1} \\ \vdots \\ \epsilon_n \end{bmatrix} \quad (4.5.3)$$

#### 4.5.3 Estimação do Modelo AR-MV com $\alpha$ desconhecido

Quando os limiares  $\alpha$  são desconhecidos, a estimação de  $\alpha$  é feita via mínimos quadrados condicionais, minimizando a soma dos quadrados dos resíduos

$$S(\alpha) = \mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{Y}_\alpha^* (\mathbf{Y}_\alpha^{*\top} \mathbf{Y}_\alpha^*)^{-1} \mathbf{Y}_\alpha^{*\top} \mathbf{Y} \quad (4.5.4)$$

Neste caso,  $\phi(\alpha)$  é substituído na função de erro, tornando  $S(\alpha)$  uma função apenas dos parâmetros limiares. A minimização de  $S(\alpha)$  fornece os estimadores  $\hat{\alpha}$ .

A variância dos resíduos é estimada por

$$\hat{\sigma}^2 = \frac{S(\hat{\alpha})}{n-p} \quad (4.5.5)$$

A matriz de covariância de  $\hat{\phi}(\hat{\alpha})$  pode ser estimada por

$$\hat{V}(\hat{\phi}(\hat{\alpha})) = \hat{\sigma}^2 (\mathbf{Y}_{\hat{\alpha}}^{*\top} \mathbf{Y}_{\hat{\alpha}}^*)^{-1} \quad (4.5.6)$$

Nos casos acima, assume-se que os estimadores  $\hat{\alpha}$  e  $\hat{\phi}$  são consistentes e que a distribuição assintótica continua válida. Para mais detalhes, ver Fadel (2012).

#### 4.5.4 Algoritmo $k$ -means

O algoritmo  $k$ -means, introduzido formalmente por MacQueen (1967), é um método de particionamento que visa agrupar um conjunto de  $n$  observações multivariadas  $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\} \subset \mathbb{R}^d$  em  $k$  grupos (ou clusters), de forma a minimizar a variabilidade intra-cluster. O problema de otimização correspondente pode ser formulado como

$$\min_{\{C_1, \dots, C_k\}} \sum_{j=1}^k \sum_{\mathbf{y}_i \in C_j} \|\mathbf{y}_i - \boldsymbol{\mu}_j\|^2,$$

em que:

- $\mathbf{y}_i \in \mathbb{R}^d$  representa a  $i$ -ésima observação;
- $C_j \subset \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$  é o conjunto de observações atribuídas ao  $j$ -ésimo cluster;
- $\boldsymbol{\mu}_j = \frac{1}{|C_j|} \sum_{\mathbf{y}_i \in C_j} \mathbf{y}_i$  é o centróide (média vetorial) do cluster  $C_j$ .

O algoritmo segue um procedimento iterativo clássico (também conhecido como algoritmo de Lloyd–Forgy), descrito da seguinte forma

1. Inicializa-se aleatoriamente um conjunto de  $k$  centróides  $\{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k\}$ ;
2. Enquanto não houver convergência
  - (a) **Atribuição:** cada observação  $\mathbf{y}_i$  é atribuída ao cluster  $C_j$  cujo centróide  $\boldsymbol{\mu}_j$  minimiza  $\|\mathbf{y}_i - \boldsymbol{\mu}_j\|$ ;
  - (b) **Atualização:** cada centróide  $\boldsymbol{\mu}_j$  é atualizado como a média das observações pertencentes a  $C_j$ .

Esse processo garante uma redução monótona da soma total dos quadrados intra-cluster e converge para um ponto de mínimo local em um número finito de iterações. No entanto, o resultado final pode depender da inicialização dos centróides, não havendo garantia de que o mínimo global seja atingido.

## 4.6 Função de Verossimilhança

A função de verossimilhança é o núcleo da inferência estatística paramétrica, pois expressa o grau de compatibilidade entre os dados observados e um valor do parâmetro do modelo. A partir dela, definem-se estimadores (como o de máxima verossimilhança) e procedimentos de teste.

Considere uma amostra de tamanho  $n$ ,

$$Y_1, Y_2, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} f(x | \theta),$$

onde:

- $Y_i \in \mathbb{R}$  são as observações;
- $\theta \in \Theta \subseteq \mathbb{R}^p$  é o vetor de parâmetros do modelo;
- $f(y | \theta)$  é a função de massa (ou densidade) de  $X_i$  sob  $\theta$ .

A *função de verossimilhança* é a função de probabilidade conjunta dos dados, vista como função de  $\theta$ :

$$L(\theta; Y_1, \dots, Y_n) = \prod_{i=1}^n f(Y_i | \theta).$$

É comum trabalhar com a *log-verossimilhança*

$$\ell(\theta; Y_1, \dots, Y_n) = \log L(\theta; Y_1, \dots, Y_n) = \sum_{i=1}^n \log f(Y_i | \theta).$$

### Teste da Razão de Verossimilhança

Seja  $L(\boldsymbol{\theta} | \mathbf{Y})$  a função de verossimilhança baseada em uma amostra  $\mathbf{Y}$  proveniente de uma família de distribuições parametrizada por  $\boldsymbol{\theta} \in \Theta$ . Suponha-se que se deseja testar as hipóteses:

$$H_0: \boldsymbol{\theta} \in \Theta_0 \quad \text{versus} \quad H_1: \boldsymbol{\theta} \in \Theta \setminus \Theta_0,$$

em que  $\Theta_0 \subset \Theta$ , ou seja, trata-se de hipóteses aninhadas. A estatística do *teste da razão de verossimilhança* (Likelihood Ratio Test, LRT) é definida por:

$$\lambda(\mathbf{Y}) = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} L(\boldsymbol{\theta} | \mathbf{Y})}{\sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta} | \mathbf{Y})}.$$

Como  $\lambda(\mathbf{Y}) \in [0, 1]$ , valores pequenos de  $\lambda(\mathbf{Y})$  indicam maior evidência contra  $H_0$ . Em geral, utiliza-se a estatística transformada:

$$-2 \log \lambda(\mathbf{Y}),$$

a qual, sob certas condições de regularidade e assumindo  $H_0$  verdadeira, converge em distribuição assintótica para uma distribuição qui-quadrado com  $k$  graus de liberdade, onde  $k = \dim(\Theta) - \dim(\Theta_0)$ , conforme estabelecido pelo Teorema de Wilks. (CASELLA; BERGER, 2002)

## 4.7 Previsão no Modelo AR-MV( $p$ )

A previsão no modelo AR-MV( $p$ ) é baseada na esperança condicional  $\hat{y}_t(m) = \mathbb{E}[y_{t+m} | \mathcal{B}_t]$ , onde  $\mathcal{B}_t$  denota a  $\sigma$ -álgebra gerada pelos valores observados da série até o instante  $t$ , ou seja,  $\mathcal{B}_t = \sigma(y_s : s \leq t)$ . Essa  $\sigma$ -álgebra representa formalmente toda a informação disponível no tempo  $t$ , incluindo o histórico completo da série  $(y_1, y_2, \dots, y_t)$ .

A dinâmica do processo depende do intervalo em que o valor passado  $y_{t-1}$  se encontra. Para  $m = 1$ , a previsão é obtida aplicando a equação do modelo no regime identificado, ou seja, com os coeficientes correspondentes ao intervalo  $A_j$  tal que  $y_{t-1} \in A_j$ :

$$\hat{y}_t(1) = \sum_{i=1}^p \phi_i y_{t+1-i}.$$

Para horizontes de previsão maiores ( $m > 1$ ), são comumente utilizados dois métodos

- **Esqueleto determinístico (Skeleton):** considera que os choques futuros  $e_{t+h}$  são nulos. As previsões são obtidas de forma recursiva, substituindo valores passados por previsões anteriores, sempre determinando o regime a partir do último valor disponível (real ou previsto):

$$\hat{y}_t(h) = \sum_{i=1}^p \phi_i \cdot \hat{y}_t(h-i).$$

- **Simulação Monte Carlo:** incorpora a aleatoriedade dos choques  $e_{t+h} \sim \mathcal{N}(0, \hat{\sigma}^2)$ , gerando  $N$  trajetórias futuras da série. A previsão final é dada pela média dessas trajetórias simuladas

$$\hat{y}_t^{MC}(h) = \frac{1}{N} \sum_{j=1}^N \hat{y}_t^{(j)}(h).$$

Ambas as abordagens permitem avaliar a tendência e a incerteza das previsões, sendo amplamente utilizadas em modelos com regimes, conforme discutido por Fadel (2012).

## 5 Metodologia

O presente estudo foi conduzido integralmente no ambiente R (versão 4.5.0) utilizando-se scripts desenvolvidos para implementar as funções básicas necessárias à geração de dados, estimativa e aplicação do modelo ARMV( $p$ )

### 5.1 Geração de Dados Simulados do Modelo ARMV( $p$ )

Para fins de avaliação e validação dos métodos propostos, foi realizada a simulação de séries temporais com base na estrutura do modelo autorregressivo com limiares variáveis, denotado por ARMV( $p$ ). Esse modelo é uma extensão do modelo autorregressivo clássico, permitindo que os coeficientes variem de acordo com o regime identificado por um conjunto de limiares sobre os valores passados da série.

O processo de simulação consiste nas seguintes etapas:

1. **Definição dos parâmetros do modelo:** Define-se um vetor de limiares  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{p-1})$ , que partitiona a reta real em  $p$  regimes. Define-se também um vetor de coeficientes autorregressivos  $\phi = (\phi_1, \phi_2, \dots, \phi_p)$ , que determina o comportamento da série em cada regime.
2. **Configuração da simulação:** Especifica-se o número total de observações a serem geradas, considerando um período de *burn-in* (aquecimento), um conjunto de observações úteis para análise e, opcionalmente, um número de valores adicionais reservados para previsão. Denotamos o total como  $T = n + n_{\text{pred}} + \text{burn-in}$ .
3. **Inicialização da série temporal:** Os  $p$  primeiros valores da série  $\{Y_t\}_{t=1}^p$  são gerados a partir de uma distribuição inicial, normalmente uma normal padrão  $\mathcal{N}(0, 1)$  ou uma normal com média zero e variância pré-definida, a fim de fornecer condições iniciais para o processo. Na simulação utilizada usamos uma  $\mathcal{N}(0, 4)$ .
4. **Geração do termo de ruído:** Uma sequência de erros  $\{a_t\}_{t=1}^T$  é gerada de forma independente e identicamente distribuída, usualmente  $a_t \sim \mathcal{N}(0, \sigma^2)$ , representando o componente estocástico do modelo. Novamente utilizamos uma  $\mathcal{N}(0, 4)$ .

5. **Simulação do processo AR-MV( $p$ ):** A série temporal é gerada iterativamente para  $t = p + 1, \dots, T$ , com base na seguinte lógica:

- Determina-se o regime ativo no tempo  $t$  com base na observação anterior  $Y_{t-1}$ , utilizando os limiares definidos. O regime  $L_t$  é tal que  $Y_{t-1} \in (\alpha_{L_{t-1}}, \alpha_{L_t}]$ , com a convenção de que  $\alpha_0 = -\infty$  e  $\alpha_p = \infty$ .
- Define-se que, neste regime  $L_t$ , serão utilizados os  $L_t$  primeiros coeficientes de  $\phi$  para gerar  $Y_t$  como uma combinação linear dos últimos  $L_t$  termos da série:

$$Y_t = \sum_{j=1}^{L_t} \phi_j Y_{t-j} + a_t$$

Essa formulação permite que o modelo capture diferentes dinâmicas ao longo do tempo, dependendo do regime em que a série se encontra. Assim, o modelo ARMV( $p$ ) é capaz de representar estruturas de dependência não lineares e adapta-se a mudanças no comportamento da série temporal.

## 5.2 Estimação dos Parâmetros e Seleção da Ordem do Modelo AR-MV( $p$ )

A estimação dos parâmetros do modelo AR-MV( $p$ ), com limiares  $\alpha$  e coeficientes autorregressivos  $\phi$ , pode ser realizada por meio de um procedimento iterativo que alterna entre a estimação de  $\alpha$  e  $\phi$ , até atingir convergência.

### 5.2.1 Iteração Alternada entre $\alpha$ e $\phi$

Inicialmente, fixa-se um valor de  $p$  e seguem-se os seguintes passos:

1. **Chute inicial de  $\phi$ :** estima-se um modelo autorregressivo tradicional AR( $p$ ), sem intercepto, por máxima verossimilhança, a fim de obter um vetor inicial de coeficientes  $\phi^{(0)}$ .
2. **Chute inicial de  $\alpha$ :** aplica-se o algoritmo *k-means* com  $p$  centros à série  $\{y_t\}$ , desconsiderando os 5% valores mais extremos, e define-se os limiares como os pontos médios entre os centróides consecutivos.
3. **Iteração:** dados os vetores  $\phi^{(k)}$  e  $\alpha^{(k)}$  da iteração  $k$ :
  - Estima-se  $\alpha^{(k+1)}$ , mantendo  $\phi^{(k)}$  fixo, pela minimização da variância dos resíduos;

- Estima-se  $\phi^{(k+1)}$ , mantendo  $\alpha^{(k+1)}$  fixo, via mínimos quadrados condicionais;
- O processo é repetido até que as variações em  $\alpha$  e  $\phi$  entre iterações consecutivas sejam menores que uma tolerância pré-definida.
- A função `optim` do R, com o método BFGS, é utilizada para minimizar a função objetivo

4. **Log-verossimilhança:** com os valores finais  $\hat{\phi}$  e  $\hat{\alpha}$ , estima-se a variância dos resíduos  $\hat{\sigma}^2$  e calcula-se a log-verossimilhança sob normalidade:

$$\ell = -\frac{n}{2} [\log(2\pi\hat{\sigma}^2) + 1],$$

onde  $n$  representa o número de observações utilizadas na estimação.

### 5.2.2 Seleção da Ordem $p$ via Teste de Razão de Verossimilhança

A seleção da ordem  $p$  do modelo AR-MV pode ser feita utilizando o teste da razão de verossimilhança, comparando modelos com ordens consecutivas. O procedimento segue os seguintes passos:

1. Ajusta-se o modelo AR-MV( $p$ ) para um valor inicial  $p_{\min}$  e calcula-se a log-verossimilhança  $\ell_{p_{\min}}$ ;
2. Para  $p = p_{\min} + 1, \dots, p_{\max}$ , ajusta-se o modelo AR-MV( $p$ ) e calcula-se a log-verossimilhança  $\ell_p$ ;
3. A estatística de teste da razão de verossimilhança é dada por:

$$LR = 2(\ell_p - \ell_{p-1}),$$

que, sob a hipótese nula de que o modelo com ordem  $p - 1$  é suficiente, segue uma distribuição qui-quadrado com 2 graus de liberdade (um novo coeficiente  $\phi$  e um novo limiar  $\alpha$ );

4. Compara-se o valor- $p$  com um nível de significância  $\alpha$  (geralmente,  $\alpha = 0,05$ ). Se o valor- $p$  for inferior a esse nível, aceita-se o modelo de ordem maior e continua-se o processo. Caso contrário, interrompe-se a seleção e define-se a ordem ótima como sendo aquela do último modelo aceito.

Esse procedimento é especialmente útil para evitar o superajuste, selecionando automaticamente a complexidade adequada ao modelo com base em evidência estatística.

### 5.3 Obtenção dos Dados Climáticos

Os dados foram obtidos diretamente no portal do INMET, onde é possível escolher as estações meteorológicas de interesse — tanto as convencionais quanto as automáticas. Após selecionar a estação desejada, pode-se especificar quais variáveis serão extraídas (por exemplo, temperaturas máxima, mínima e média, índices pluviométricos etc.) e a frequência temporal das observações (diária, mensal ou em intervalos de horas por dia, disponibilizados pelas estações automáticas). Em seguida, o INMET processa a requisição e envia os arquivos de dados por e-mail.

## 6 Simulações

Com o objetivo de avaliar a capacidade do procedimento de estimação proposto para o modelo AR-MV( $p$ ), conduzimos um estudo de simulação em que séries temporais são geradas a partir de configurações conhecidas de limiares  $\alpha$  e coeficientes autorregressivos  $\phi$ . A ideia central é verificar se, a partir apenas dos dados simulados, o algoritmo de estimação é capaz de recuperar adequadamente os parâmetros  $\alpha$  e  $\phi$ , bem como a ordem  $p$  do modelo.

Para isso, procedemos da seguinte forma:

- Geramos séries temporais sintéticas sob a estrutura do modelo AR-MV( $p$ ), com  $p = 6$  e  $p = 7$ , utilizando valores conhecidos de  $\alpha$ ,  $\phi$  e  $\sigma^2$ ;
- Aplicamos, sobre cada série gerada, a função de estimação descrita anteriormente, a qual realiza a estimação conjunta de  $\alpha$  e  $\phi$  por meio de um procedimento iterativo;
- Utilizamos também o teste de razão de verossimilhança para avaliar a seleção correta da ordem  $p$  do modelo;
- Comparamos os parâmetros estimados com os valores reais utilizados na geração das séries, analisando o erro de estimação, a estabilidade do processo e o ajuste do modelo;
- Por fim, coletamos medidas descritivas e representações gráficas que permitam visualizar o comportamento da série gerada e a aderência do modelo estimado.

Este estudo é fundamental para verificar a robustez do método proposto em contextos controlados e compreender suas limitações e pontos fortes antes de aplicá-lo em dados reais.

## 6.1 Simulação ARMV(6)

Na simulação do modelo ARMV(6) foi utilizado os vetores de parâmetros  $\alpha = (-2, -1, 0, 1, 2)$  e  $\beta = (0,05, 0,10, 0,15, 0,20, 0,20, 0,25)$ , além de ruído aleatório gerado a partir de uma distribuição Normal com média  $\mu = 0$  e desvio-padrão  $\sigma = 2$ . O seguinte sistema de equações define nosso modelo.

$$Y_t = \begin{cases} 0.05 Y_{t-1} + \epsilon_t, & Y_{t-1} \leq -2, \\ 0.05 Y_{t-1} + 0.10 Y_{t-2} + \epsilon_t, & -2 < Y_{t-1} \leq -1, \\ 0.05 Y_{t-1} + 0.10 Y_{t-2} + 0.15 Y_{t-3} + \epsilon_t, & -1 < Y_{t-1} \leq 0, \\ 0.05 Y_{t-1} + 0.10 Y_{t-2} + 0.15 Y_{t-3} + 0.20 Y_{t-4} + \epsilon_t, & 0 < Y_{t-1} \leq 1, \\ 0.05 Y_{t-1} + 0.10 Y_{t-2} + 0.15 Y_{t-3} + 0.20 Y_{t-4} + 0.20 Y_{t-5} + \epsilon_t, & 1 < Y_{t-1} \leq 2, \\ 0.05 Y_{t-1} + 0.10 Y_{t-2} + 0.15 Y_{t-3} + 0.20 Y_{t-4} + 0.20 Y_{t-5} + 0.25 Y_{t-6} + \epsilon_t, & Y_{t-1} > 2. \end{cases}$$

Ao todo, serão produzidos 10110 valores, dos quais 100 são descartados como “burn-in” e os ultimos 10 ficam reservados para avaliar a capacidade preditiva do modelo, resultando em uma série final de 10000 observações. Antes de iniciar a geração de dados, criamos um vetor de ruídos aleatórios de comprimento 10110 e um vetor de observações inicial contendo os 6 primeiros valores, sorteados aleatoriamente a partir de uma Normal; com isso, obtemos os valores do modelo ARMV(6) representados abaixo.

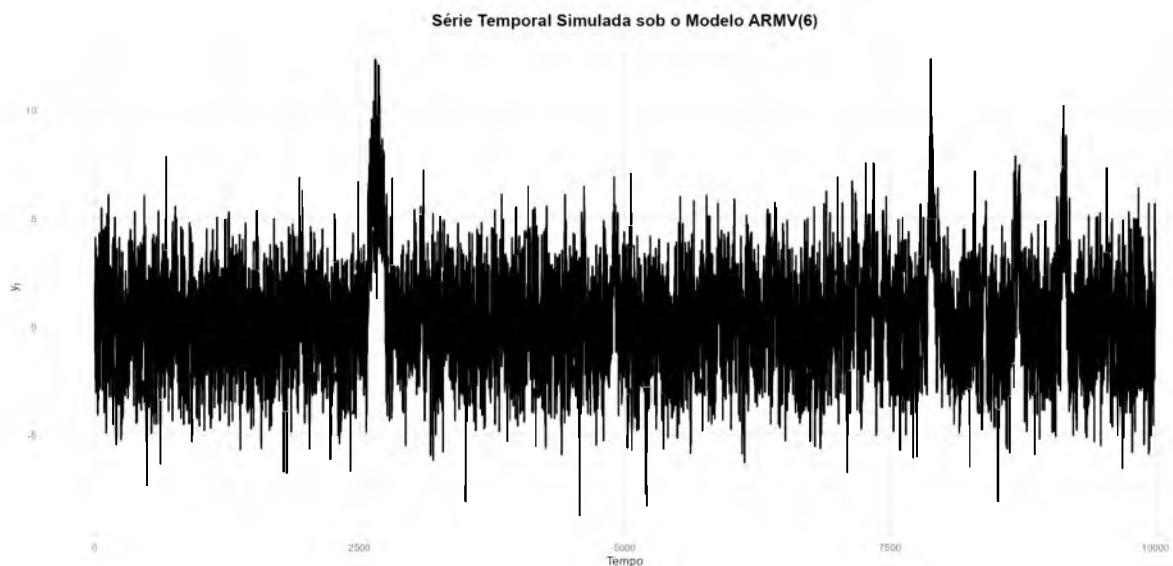


Figura 1: Serie temporal dos dados simulados para um ARMV(6)

Podemos observar que a série apresenta alguns picos nos quais os valores se mantêm elevados por determinados períodos. Nota-se um pico acentuado no início da

série, seguido por outros dois, de menor duração, localizados mais ao final da série.

Tabela 1: Resumo dos resultados da simulação ARMV(6)

Estatística	Valor	Detalhes
Média	0,2971	—
Variância	5,4233	—
Número de observações	10000	—
<b>Teste ADF</b>		
Estatística Dickey-Fuller	-11,643	Lag = 21
<i>p</i> -valor	0,01	Hipótese alternativa: estacionária
<b>Teste PP</b>		
Z( $\alpha$ ) de Dickey-Fuller	-14368	Lag truncado = 12
<i>p</i> -valor	0,01	Hipótese alternativa: estacionária

A Tabela 1 apresenta um resumo estatístico da série simulada sob o modelo ARMV(6). Os resultados dos testes de raiz unitária — ADF (Augmented Dickey-Fuller) e PP (Phillips-Perron) — indicam, com forte evidência, a rejeição da hipótese nula de não estacionariedade, dado que ambos os testes apresentam estatísticas significativamente negativas e valores-*p* inferiores a 0,01. Esses resultados reforçam a hipótese de que a série simulada é estacionária, como esperado pelo fato da soma dos coeficientes ser menor que 1.

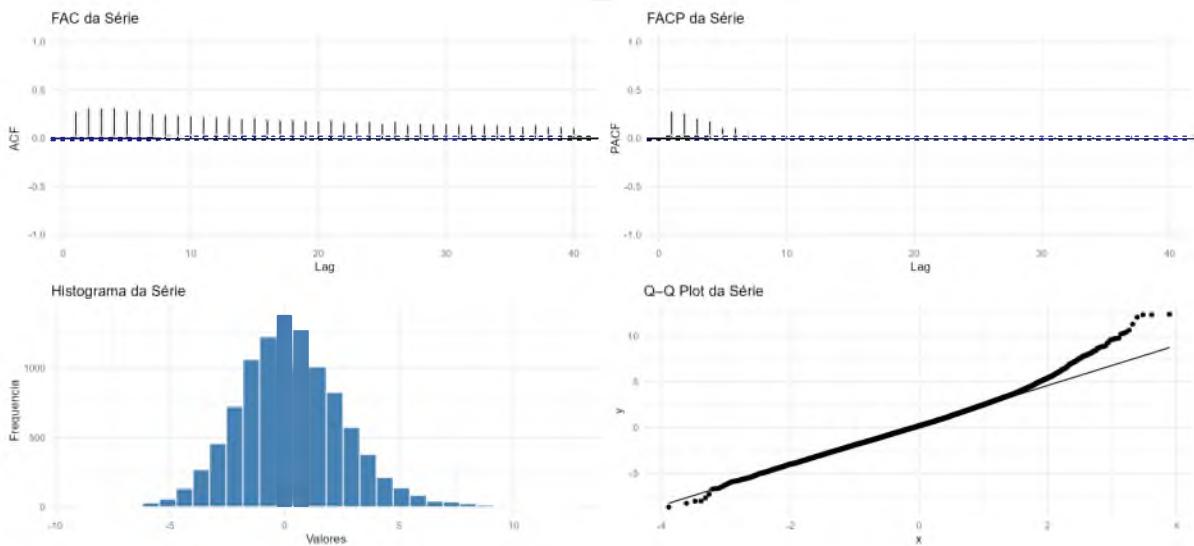


Figura 2: Gráficos de diagnóstico da série simulada AR-MV(6): FAC, FACP, histograma e Q-Q plot

A Figura 2 mostra os principais diagnósticos da série simulada ARMV(6). A FAC apresenta um decaimento gradual, indicando persistência temporal, enquanto a FACP

exibe um corte claro após a sexta defasagem, confirmando a ordem do modelo. O histograma sugere distribuição aproximadamente normal, e o Q–Q plot indica boa aderência à normalidade, com desvios leves nas caudas. Esses resultados reforçam a estacionariedade da série, a ordem autoregressiva e a adequação da hipótese de erro gaussiano.

### 6.1.1 Estimação dos parâmetros

Nesta etapa, buscamos estimar os parâmetros do modelo AR-MV — limiares  $\alpha$  e coeficientes autorregressivos  $\phi$  — a partir da série simulada.

O procedimento de estimação, baseado em iteração alternada entre  $\alpha$  e  $\phi$ , convergiu em apenas duas iterações. O valor final selecionado para a ordem do modelo foi  $p = 6$ , conforme definido pelo critério da razão de verossimilhança.

Tabela 2: Comparação entre Parâmetros Estimados e Verdadeiros

Parâmetro	Valor Verdadeiro	Valor Estimado
$\alpha_1$	-2.0000	-1.9323
$\alpha_2$	-1.0000	-0.7620
$\alpha_3$	0.0000	-0.1347
$\alpha_4$	1.0000	1.3053
$\alpha_5$	2.0000	2.0113
$\phi_1$	0.0500	0.0298
$\phi_2$	0.1000	0.1020
$\phi_3$	0.1500	0.1455
$\phi_4$	0.2000	0.1961
$\phi_5$	0.2000	0.1720
$\phi_6$	0.2500	0.3099

Os resultados da Tabela 9 mostram que os parâmetros estimados para o modelo AR-MV(6) apresentam boa proximidade em relação aos valores verdadeiros utilizados na simulação. Os limiares  $\alpha_j$  foram recuperados com pequenas variações, e a separação entre regimes. Os coeficientes autorregressivos  $\phi_j$  também foram bem estimados, com desvios modestos, especialmente para os primeiros lags. A maior diferença foi observada em  $\phi_6$ , cuja estimativa foi ligeiramente superior ao valor real, o que pode ser atribuído à variabilidade amostral. De forma geral, os resultados indicam que o procedimento de estimação adotado foi eficaz para recuperar a estrutura do modelo simulado.

Tabela 3: Estimativas dos Coeficientes  $\phi$ , Erros Padrão e Intervalos de Confiança

Coef.	Valor Verdadeiro	Estimativa	Erro Padrão	IC 95%	Cobertura
$\phi_1$	0,05	0,0298	0,0098	0,0106 – 0,0490	Não
$\phi_2$	0,10	0,1020	0,0106	0,0812 – 0,1228	Sim
$\phi_3$	0,15	0,1455	0,0115	0,1230 – 0,1679	Sim
$\phi_4$	0,20	0,1961	0,0127	0,1711 – 0,2210	Sim
$\phi_5$	0,20	0,1720	0,0162	0,1403 – 0,2037	Sim
$\phi_6$	0,25	0,3099	0,0183	0,2741 – 0,3458	Não

Os resultados apresentados na Tabela 17 indicam boa acurácia na recuperação dos coeficientes  $\phi$  do modelo, com cinco dos seis valores verdadeiros contidos nos respectivos intervalos de confiança de 95%. Observa-se, no entanto, subestimação de  $\phi_1$  e superestimação de  $\phi_6$ , ambos fora dos intervalos de cobertura, o que sugere maior incerteza nas extremidades da estrutura autorregressiva.

Uma possível explicação para a discrepância em  $\phi_6$  está relacionada à própria dinâmica da série simulada: como observamos anteriormente, há trechos em que a série permanece por longos períodos em um mesmo regime, especialmente no regime associado ao último limiar. Esse padrão implica um número desproporcionalmente maior de observações nesse regime, o que pode enviesar as estimativas dos coeficientes correspondentes.

Para uma análise mais clara da segmentação induzida pelos limiares estimados, será apresentada, a seguir, a decomposição da série conforme os regimes identificados.

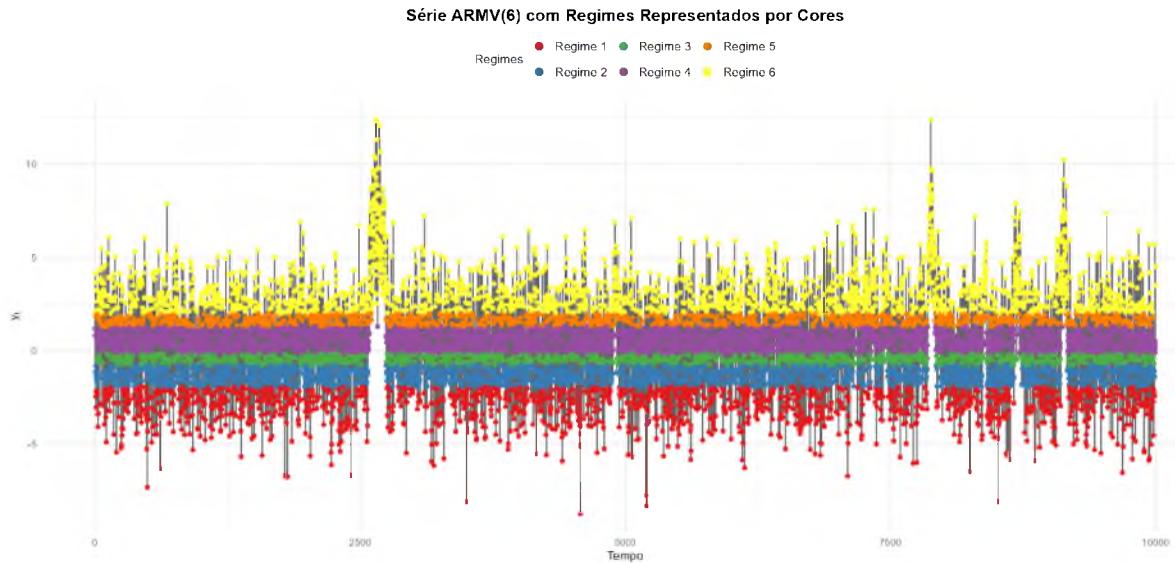


Figura 3: Série simulada de um modelo AR-MV(6), com observações representadas por cores conforme os regimes estimados.

Na Figura 3, é possível observar quatro picos distintos na série simulada: um mais prolongado em torno do tempo 2500 e três menores entre aos tempos 7500 e 10000. Esses episódios prolongados em níveis elevados fazem com que haja uma concentração maior de observações no regime associado ao último limiar ( $\alpha_5$ ), o que pode explicar a superestimação de  $\phi_6$ . Por outro lado, a menor frequência de permanência em valores baixos reduz o número de observações no regime do primeiro limiar ( $\alpha_1$ ), o que pode justificar a subestimação de  $\phi_1$ . Essa assimetria na distribuição das observações entre os regimes afeta diretamente a qualidade das estimativas nos extremos do modelo.

Tabela 4: Resumo Estatístico por Regime do AR-MV(6)

Regime	N de Observações	Média	Mediana	Variância
1	1557	-3,0612	-2,8011	0,9423
2	1712	-1,3213	-1,3111	0,1117
3	1139	-0,4423	-0,4436	0,0329
4	2528	0,5486	0,5294	0,1651
5	956	1,6462	1,6303	0,0411
6	2108	3,5780	3,1380	2,2871

A Tabela 4 mostra que o Regime 6 concentra um número significativamente maior de observações do que os regimes mais extremos, como o 1, 2 e 5. Esse desequilíbrio pode ter impactado a qualidade das estimativas nos limites do modelo, contribuindo para desvios maiores em alguns parâmetros em relação aos seus valores verdadeiros.

### 6.1.2 Estudo do ajuste do modelo ARMV(6)

Tabela 5: Métricas de Ajuste — ARMV(6) vs AR

Modelo	AIC	AICc	BIC	MAE	MSE	RMSE	MAPE
ARMV	42107,03	42107,05	42157,50	1,59	3,95	1,99	187,41
AR6c	42851,32	42851,33	42909,00	4,25	19,11	4,37	93,73
AR6	42865,42	42865,42	42915,89	4,33	19,83	4,45	95,50
AR7c	42829,99	42830,00	42894,88	4,25	19,12	4,37	93,77
AR7	42842,61	42842,63	42900,30	4,32	19,80	4,45	95,45

A Tabela 19 apresenta as métricas de ajuste dos diferentes modelos aos dados simulados. Observa-se que o modelo ARMV(6), que corresponde ao verdadeiro gerador da série, atinge os melhores resultados em todas as métricas de ajuste (menores valores

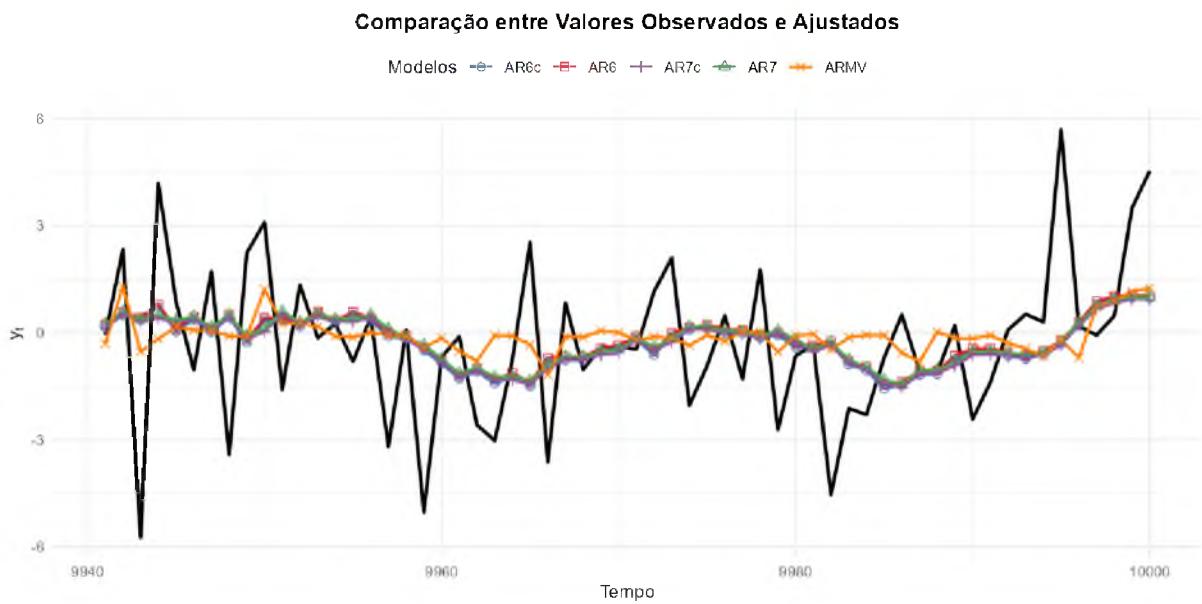


Figura 4: Ajuste dos modelos AR-MV(6), AR(6) e AR(7) às últimas 60 observações da série simulada.

de variância residual, MAE, MSE, RMSE, AIC, AICc, BIC e MAPE) em comparação com os modelos AR tradicionais. Esse desempenho superior era esperado, já que os dados foram simulados a partir de um processo ARMV(6).

### 6.1.3 Estudo dos resíduos do modelo ARMV(6)

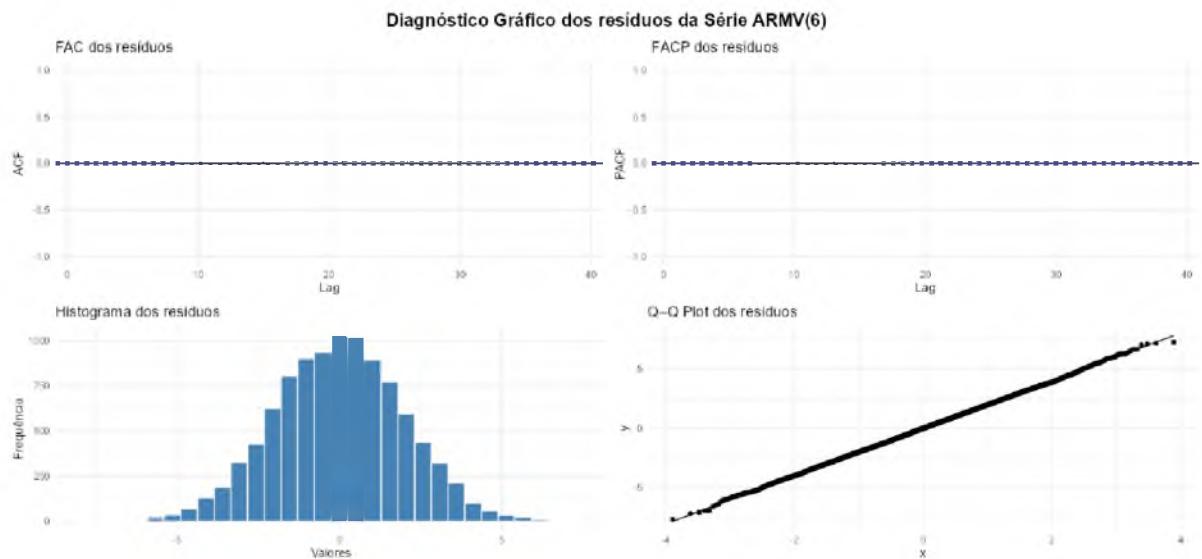


Figura 5: Gráficos de diagnóstico dos resíduos do modelo AR-MV(6): FAC, FACP, histograma e Q-Q plot

A Figura 5 exibe os diagnósticos dos resíduos do modelo AR-MV(6). Os gráficos de FAC e FACP indicam ausência de autocorrelação significativa, sugerindo que os resíduos são essencialmente ruído branco. O histograma apresenta uma distribuição simétrica e em

forma de sino, enquanto o Q-Q plot mostra boa aderência à linha teórica, com pequenos desvios apenas nas caudas. Esses resultados confirmam que o modelo ajustado capturou bem a estrutura da série, gerando resíduos não correlacionados e aproximadamente normais.

#### 6.1.4 Estudo preditivo do modelo ARMV(6)

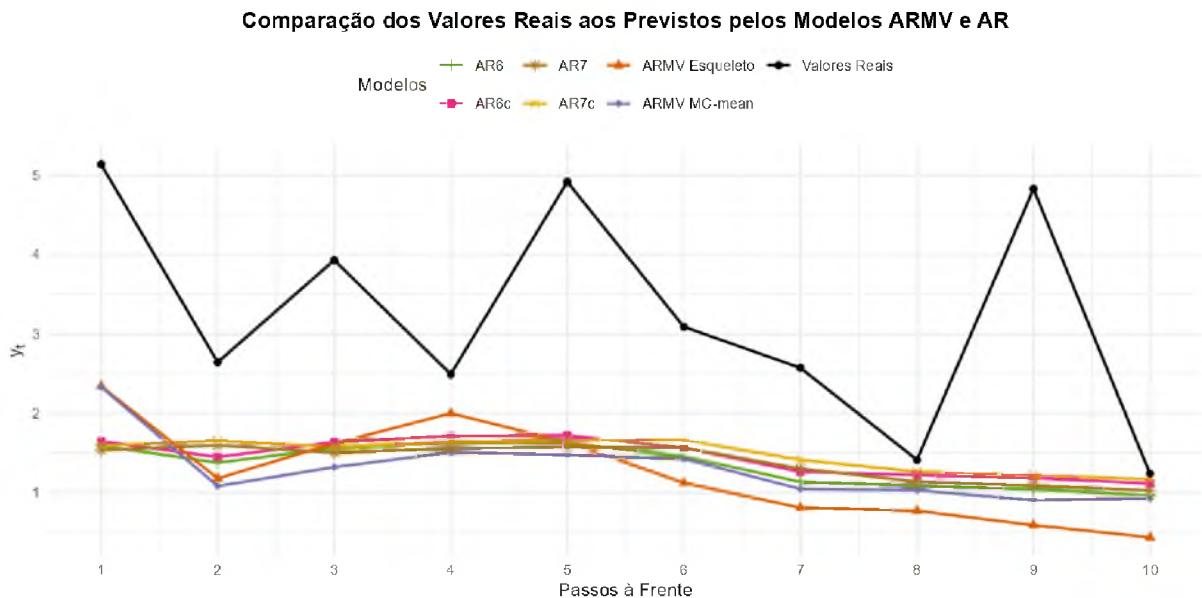


Figura 6: Previsão a 10 passos dos modelos AR-MV(6), AR(6) e AR(7) aplicada à série simulada.

Tabela 6: Métricas de Previsão a 1–10 Passos à Frente

Modelo	MAE	MSE	RMSE	MAPE
ARMV Esqueleto	1,9764	5,2331	2,2876	58,58
ARMV MC-mean	1,9221	5,0731	2,2524	53,66
AR6	1,8804	5,0690	2,2515	51,14
AR6c	1,7753	4,7106	2,1704	46,70
AR7	1,8384	5,0112	2,2386	48,99
AR7c	1,7413	4,6859	2,1647	44,90

Na Figura 6 são apresentadas as previsões a 1–10 passos à frente de seis modelos (ARMV Esqueleto, ARMV MC-mean, AR6, AR6c, AR7 e AR7c) em comparação aos valores reais (linha preta). Nota-se que todos os modelos geram previsões estáveis e próximas entre si, com pouca sensibilidade aos grandes picos e oscilações presentes na série real, especialmente para horizontes mais longos. Entre os modelos avaliados, AR7c e AR6c apresentaram o melhor desempenho preditivo, registrando os menores valores

de MAE, MSE, RMSE e MAPE. No geral, os modelos AR com média (indicados por “c”) superaram os demais em todas as métricas de erro, indicando maior acurácia para previsões nesse cenário simulado.

Tabela 7: Coeficientes Estimados — ARMV vs AR

Modelo	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\phi}_5$	$\hat{\phi}_6$	$\hat{\phi}_7$
AR6	0,08518	0,14396	0,14477	0,14819	0,09822	0,10948	—
AR6c	0,08342	0,14217	0,14298	0,14639	0,09644	0,10774	—
AR7	0,07974	0,13909	0,13740	0,14097	0,09102	0,10524	0,04979
AR7c	0,07822	0,13753	0,13592	0,13948	0,08955	0,10371	0,04830
ARMV	0,02980	0,10198	0,14545	0,19606	0,17201	0,30993	—

A Tabela 7 permite comparar os coeficientes autorregressivos estimados entre os modelos AR tradicionais e o modelo AR-MV(6). Nota-se que os modelos AR (AR6, AR7 e suas versões com intercepto) apresentam um padrão decrescente e suavemente distribuído entre os lags, com coeficientes em torno de 0,08 a 0,14. Por outro lado, o modelo AR-MV(6) exibe um padrão mais assimétrico: os coeficientes iniciais são menores (especialmente  $\hat{\phi}_1 = 0,0298$ ), enquanto os de ordem mais alta, como  $\hat{\phi}_6 = 0,3099$ , são substancialmente maiores.

Esse comportamento reflete a natureza do modelo AR-MV, que permite regimes distintos e, portanto, capta dinâmicas mais complexas. Em contraste, os modelos AR impõem estrutura homogênea, o que pode limitar a capacidade de modelar variações bruscas ou não lineares. A presença de coeficientes mais elevados nos últimos lags do ARMV sugere que esses termos são mais relevantes em regimes específicos, o que é coerente com os picos e mudanças de comportamento observados na série simulada.

## 6.2 Simulação ARMV(7)

Na simulação do modelo AR-MV(7) foram usados os limiares

$$\alpha = (-2.5, -1.5, -0.5, 0.5, 1.5, 2.5)$$

e o vetor de coeficientes autorregressivos

$$\phi = (0.05, 0.05, 0.10, 0.10, 0.20, 0.20, 0.25),$$

além de um termo de erro  $\epsilon_t \sim \mathcal{N}(0, 2^2)$ . Geraram-se 10.110 observações no total, sendo as primeiras 100 destinadas ao *burn-in* e as 10 finais reservadas para teste da capacidade preditiva. O sistema de equações que define o modelo AR-MV(7) é

$$Y_t = \begin{cases} \phi_1 Y_{t-1} + \varepsilon_t, & Y_{t-1} \leq -2.5, \\ \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, & -2.5 < Y_{t-1} \leq -1.5, \\ \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \varepsilon_t, & -1.5 < Y_{t-1} \leq -0.5, \\ \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \phi_4 Y_{t-4} + \varepsilon_t, & -0.5 < Y_{t-1} \leq 0.5, \\ \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \phi_4 Y_{t-4} + \phi_5 Y_{t-5} + \varepsilon_t, & 0.5 < Y_{t-1} \leq 1.5, \\ \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \phi_4 Y_{t-4} + \phi_5 Y_{t-5} + \phi_6 Y_{t-6} + \varepsilon_t, & 1.5 < Y_{t-1} \leq 2.5, \\ \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \phi_4 Y_{t-4} + \phi_5 Y_{t-5} + \phi_6 Y_{t-6} + \phi_7 Y_{t-7} + \varepsilon_t, & Y_{t-1} > 2.5. \end{cases}$$

A seguir, apresentamos a série gerada juntamente com as estatísticas descritivas e os resultados dos testes ADF e PP.

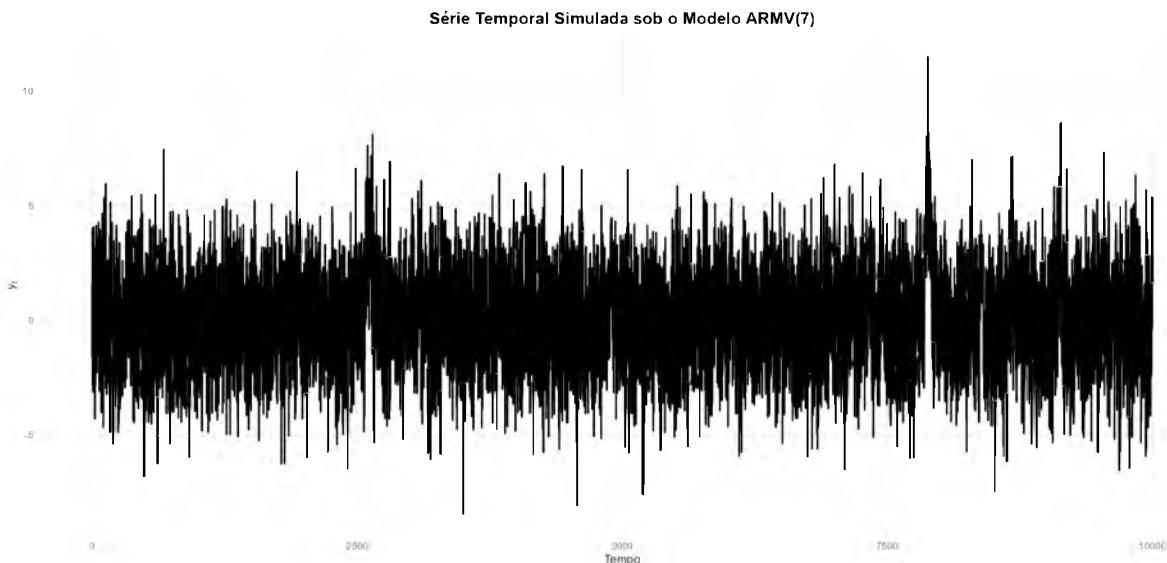


Figura 7: Série temporal dos dados simulados para um ARMV(7)

Tabela 8: Resumo dos resultados da simulação ARMV(7)

Estatística	Valor	Detalhes
Média	0,0988	—
Variância	4,4621	—
Número de observações	10000	
<b>Teste ADF</b>		
Estatística Dickey-Fuller	-15,139	Lag = 21
<i>p</i> -valor	0,01	Hipótese alternativa: estacionária
<b>Teste PP</b>		
Z( $\alpha$ ) de Dickey-Fuller	-13559	Lag truncado = 12
<i>p</i> -valor	0,01	Hipótese alternativa: estacionária

A Tabela 8 apresenta as estatísticas descritivas e os testes de raiz unitária aplicados à série simulada do modelo ARMV(7). A média próxima de zero e a variância moderada indicam uma série centrada e com dispersão compatível com o processo gerador. Os testes ADF e PP rejeitam a hipótese nula de não estacionariedade com elevada margem ( $p$ -valor  $< 0,01$ ), confirmando que a série simulada é estacionária, como esperado pelo modelo. Esses resultados reforçam a validade do processo ARMV gerado e indicam adequação para análises subsequentes de estimação e previsão.

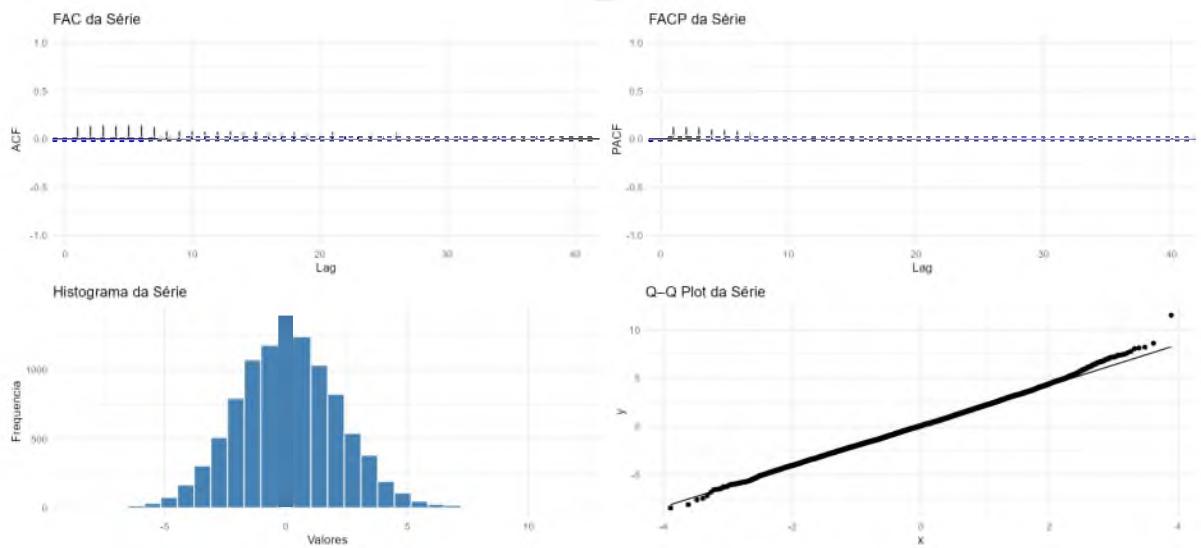


Figura 8: Gráficos de diagnóstico da série simulada AR-MV(7): FAC, FACP, histograma e Q–Q plot

A Figura 8 apresenta os diagnósticos da série simulada para o modelo ARMV(7). Os gráficos de FAC e FACP indicam ausência de autocorrelação significativa, sugerindo que a série gerada é estacionária. O histograma na figura 8 revela uma distribuição aproximadamente simétrica e com formato semelhante ao de uma Normal, o que está de acordo com os pressupostos do modelo gerador. Já o gráfico Q–Q reforça essa interpretação, mostrando boa aderência dos quantis amostrais à reta teórica. Esses resultados indicam que a série simulada apresenta propriedades compatíveis com um processo ARMV estacionário com erros gaussianos.

### 6.2.1 Estimação dos parâmetros

Nesta etapa, nosso objetivo é estimar, a partir da série simulada, os limiares  $\alpha$  e os coeficientes autorregressivos  $\phi$ . O algoritmo de estimação convergiu em apenas duas iterações, fornecendo os seguintes valores estimados para  $\alpha$ :

Tabela 9: Comparação entre Parâmetros Estimados e Verdadeiros — ARMV(7)

Parâmetro	Valor Verdadeiro	Valor Estimado
$\alpha_1$	-2,5	-1,9451
$\alpha_2$	-1,5	-1,2640
$\alpha_3$	-0,5	-0,3448
$\alpha_4$	0,5	0,4313
$\alpha_5$	1,5	1,3308
$\alpha_6$	2,5	2,4719
$\phi_1$	0,0500	0,0304
$\phi_2$	0,0500	0,0510
$\phi_3$	0,1000	0,0972
$\phi_4$	0,1000	0,0998
$\phi_5$	0,2000	0,1800
$\phi_6$	0,2000	0,2185
$\phi_7$	0,2500	0,2467

Os resultados apresentados na Tabela 9 mostram que os limiares estimados  $\hat{\alpha}_j$  e os coeficientes autorregressivos  $\hat{\phi}_j$  se aproximam dos valores verdadeiros utilizados na geração da série ARMV(7). Apesar de pequenas discrepâncias, principalmente nos primeiros limiares, os valores estimados mantêm a ordem e espaçamento aproximado em relação aos verdadeiros, o que sugere uma boa recuperação da estrutura de mudança de regime do modelo.

No que se refere aos coeficientes  $\phi_j$ , nota-se que os valores estimados seguem de forma bastante próxima os coeficientes verdadeiros, com desvios pequenos e sem inversões de magnitude, o que indica que o procedimento de estimação foi eficaz em identificar a estrutura de dependência temporal.

Esses resultados corroboram a capacidade do estimador em recuperar os parâmetros de um processo ARMV(7) a partir dos dados simulados, reforçando a consistência do método aplicado mesmo em modelos com maior número de defasagens e múltiplos limiares.

Tabela 10: Estimativas dos Coeficientes  $\phi$ , Erros Padrão e Intervalos de Confiança

Coef.	Valor Verdadeiro	Estimativa	Erro Padrão	IC 95%	Cobertura
$\phi_1$	0,05	0,0304	0,0099	0,0111 – 0,0497	Não
$\phi_2$	0,05	0,0510	0,0108	0,0299 – 0,0721	Sim
$\phi_3$	0,10	0,0972	0,0112	0,0752 – 0,1192	Sim
$\phi_4$	0,10	0,0998	0,0128	0,0748 – 0,1249	Sim
$\phi_5$	0,20	0,1800	0,0146	0,1514 – 0,2086	Sim
$\phi_6$	0,20	0,2185	0,0179	0,1834 – 0,2537	Sim
$\phi_7$	0,25	0,2467	0,0247	0,1984 – 0,2950	Sim

A Tabela 10 apresenta as estimativas dos coeficientes  $\phi_j$  do modelo ARMV(7), juntamente com seus erros padrão, intervalos de confiança de 95% e a indicação de cobertura, isto é, se o valor verdadeiro encontra-se dentro do respectivo intervalo.

Observa-se que 6 dos 7 coeficientes tiveram seus valores verdadeiros contidos nos intervalos de confiança, o que indica boa precisão das estimativas obtidas pelo método de estimação. A única exceção é o coeficiente  $\phi_1$ , cuja estimativa foi subestimada e cujo intervalo [0,0111, 0,0497] não inclui o valor verdadeiro de 0,05. Isso pode estar associado à menor influência ou menor frequência do regime correspondente nos dados simulados, especialmente considerando que coeficientes iniciais tendem a ser mais difíceis de estimar em modelos com múltiplos regimes e ordem elevada.

No geral, os coeficientes mais centrais e dominantes ( $\phi_3$  a  $\phi_7$ ) foram bem estimados, com valores próximos aos verdadeiros e intervalos estreitos, refletindo a eficiência do estimador no contexto simulado. Esse resultado sugere que o modelo foi capaz de capturar adequadamente a estrutura autoregressiva da série ARMV(7) simulada.



Figura 9: Série simulada de um modelo AR-MV(7), com observações representadas por cores conforme os regimes estimados.

Tabela 11: Resumo Estatístico por Regime do AR-MV(7)

Regime	N de Observações	Média	Mediana	Variância
1	1623	-3,0317	-2,7799	0,8994
2	1023	-1,5822	-1,5746	0,0381
3	1520	-0,7882	-0,7819	0,0683
4	1556	0,0413	0,0431	0,0504
5	1543	0,8555	0,8419	0,0678
6	1450	1,8579	1,8486	0,1060
7	1285	3,6161	3,3078	1,1009

A Tabela 11 apresenta um resumo estatístico das observações segmentadas por regime na série simulada a partir do modelo AR-MV(7). Observa-se que os regimes 1 e 7 concentram os valores mais extremos da série, com médias de  $-3,03$  e  $3,62$ , respectivamente, e também maiores variâncias, o que reflete maior dispersão nos extremos. Já os regimes centrais, especialmente o 4, apresentam médias próximas de zero e baixa variabilidade, compatíveis com um comportamento mais estável.

A mediana próxima à média em todos os regimes indica simetria nas distribuições dentro de cada faixa. Essa segmentação ajuda a explicar a natureza da série simulada, pois cada regime contribui com características distintas à dinâmica geral do processo. Além disso, a distribuição relativamente equilibrada de observações entre os regimes garante diversidade suficiente para uma estimativa robusta dos parâmetros.

### 6.2.2 Estudo do ajuste do modelo

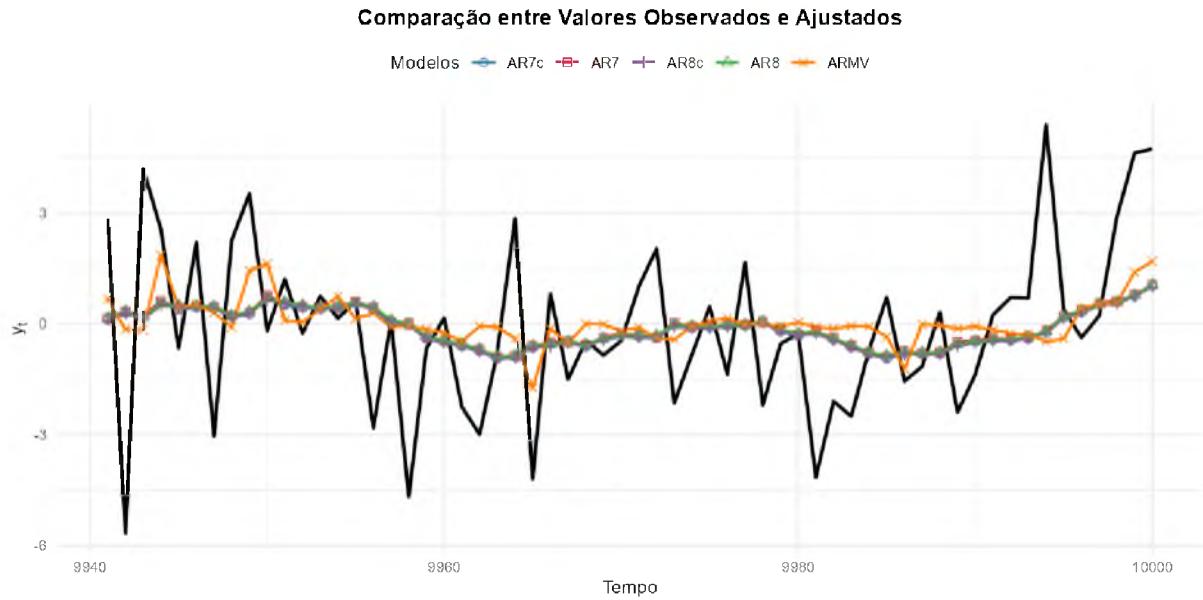


Figura 10: Ajuste dos modelos AR-MV(7), AR(7) e AR(8) às últimas 60 observações da série simulada.

Tabela 12: Métricas de Ajuste — ARMV(7) vs AR

Modelo	AIC	AICc	BIC	MAE	MSE	RMSE	MAPE
ARMV	42096,88	42096,89	42154,56	1,5894	3,9478	1,9869	167,88
AR6	42632,99	42633,00	42683,47	4,6794	22,2123	4,7130	98,96
AR6c	42628,96	42628,97	42686,64	4,6302	21,7491	4,6636	97,92
AR7	42597,07	42597,08	42654,76	4,6765	22,1999	4,7117	98,90
AR7c	42593,71	42593,73	42658,61	4,6303	21,7651	4,6653	97,92
AR8	42597,22	42597,24	42662,12	4,6759	22,1949	4,7111	98,89
AR8c	42594,00	42594,02	42666,10	4,6303	21,7659	4,6654	97,92

A Tabela 12 apresenta as métricas de ajuste calculadas sobre uma série simulada a partir de um modelo AR-MV(7). Observa-se que o próprio modelo AR-MV(7) apresenta desempenho superior em todas as métricas consideradas, como era esperado. Especificamente, obteve a menor variância residual ( $\hat{\sigma}^2 = 3,9478$ ) e os menores valores de AIC, AICc e BIC, indicando melhor ajuste segundo os critérios de informação.

### 6.2.3 Estudo dos resíduos do modelo ARMV(7)

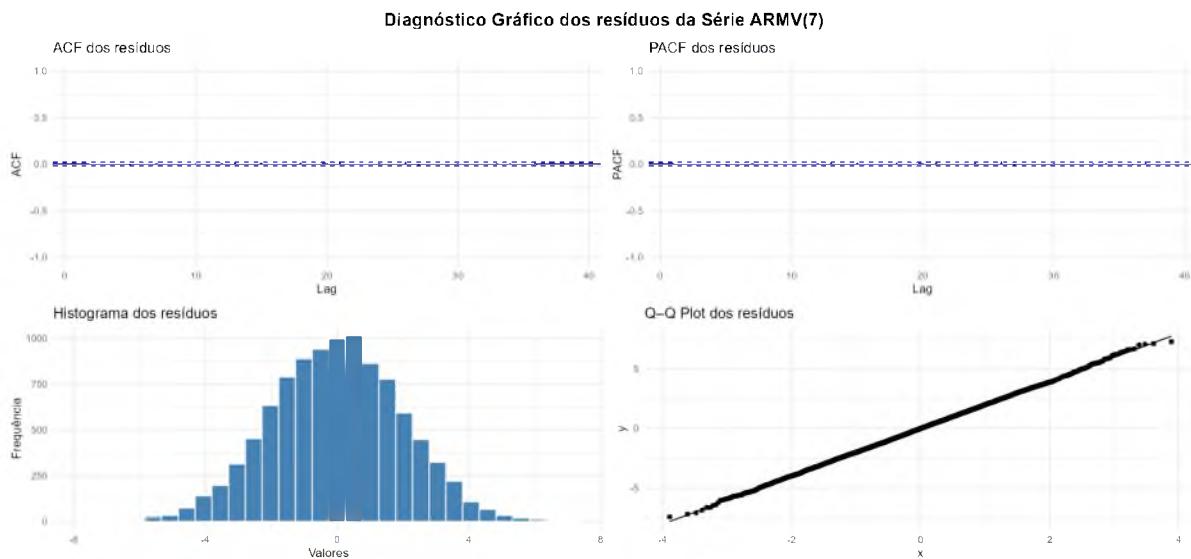


Figura 11: Gráficos de diagnóstico dos resíduos do modelo AR-MV(7): FAC, FACP, histograma e Q-Q plot

A Figura 11 apresenta os diagnósticos gráficos dos resíduos do modelo AR-MV(7). Os gráficos de autocorrelação (FAC) e autocorrelação parcial (FACP) mostram que não há valores significativos fora dos limites de confiança, sugerindo ausência de dependência serial nos resíduos. O histograma indica que os resíduos seguem uma distribuição aproximadamente simétrica e semelhante à normal. Essa suposição é reforçada pelo gráfico Q-Q, no qual os quantis empíricos aderem bem à linha de referência teórica. No conjunto, os resultados apontam para resíduos não autocorrelacionados e aproximadamente gaussianos, confirmando a adequação do modelo estimado à estrutura dos dados simulados.

### 6.2.4 Estudo preditivo do modelo ARMV(7)

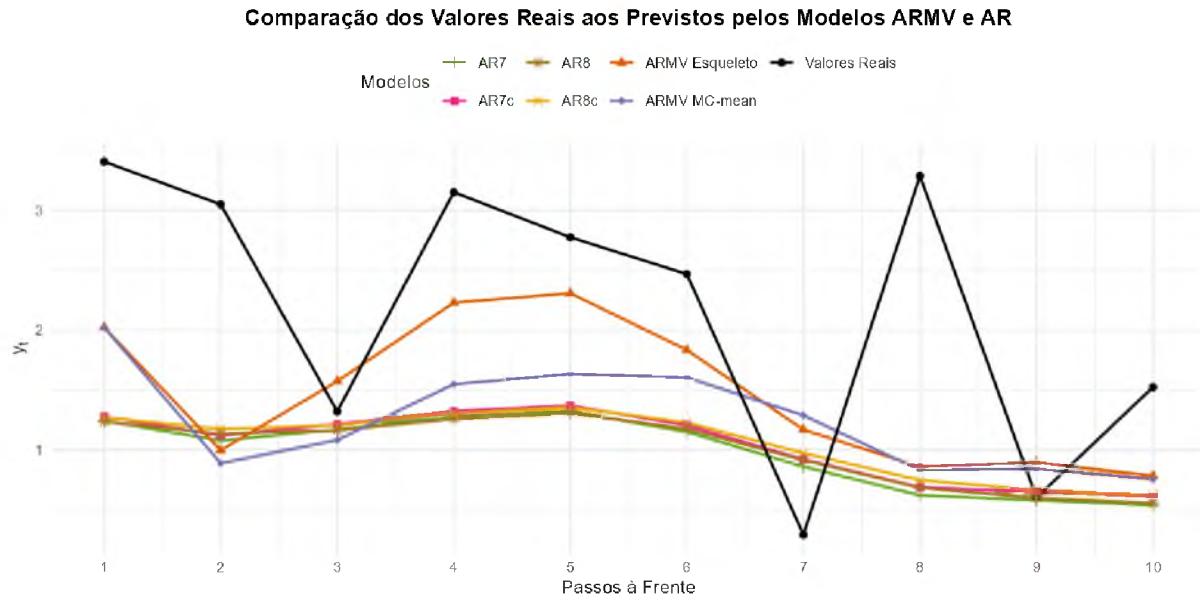


Figura 12: Previsão a 10 passos dos modelos AR-MV(7), AR(7) e AR(8) aplicada à série simulada.

A Figura 12 apresenta as previsões a 1–10 passos à frente dos modelos AR7, AR8, AR7 com constante (AR7c), AR8 com constante (AR8c), bem como dos métodos ARMV Esqueleto e média de Monte Carlo (ARMV MC-mean), em comparação com os valores reais da série simulada (linha preta). Observa-se que, embora todos os modelos apresentem séries de previsões suavizadas, os métodos baseados em ARMV foram capazes de captar parte da estrutura de oscilação da série original, sobretudo o modelo ARMV Esqueleto, que reproduz com mais fidelidade os picos e vales.

Tabela 13: Métricas de Previsão a 1–10 Passos à Frente

Modelo	MAE	MSE	RMSE	MAPE (%)
ARMV Esqueleto	1,00473	1,49483	1,22263	67,17015
ARMV MC-mean	1,18476	1,89232	1,37561	76,40377
AR7c	1,28441	2,31409	1,52121	65,71092
AR7	1,31686	2,43379	1,56006	64,99339
AR8c	1,28074	2,27535	1,50843	67,37590
AR8	1,30876	2,38935	1,54575	66,11036

Essa impressão é corroborada pelos resultados quantitativos. O modelo ARMV Esqueleto obteve os melhores desempenhos em termos de MAE (1,00473), MSE (1,49483) e RMSE (1,22263), destacando-se como o mais preciso nos erros absolutos e quadráticos médios. No entanto, o modelo AR7 apresentou o menor erro percentual médio (MAPE =

64,99%), seguido de AR7c (65,71%) e AR8 (66,11%), indicando que os modelos lineares com constante também oferecem boa capacidade preditiva em termos relativos.

Por outro lado, o método ARMV MC-mean, embora ainda tenha tido um desempenho superior aos modelos AR nas outras métricas, apresentou desempenho inferior, especialmente no MAPE (76,40%), sugerindo que, neste cenário, a variabilidade introduzida pelas simulações não resultou em ganho preditivo.

Tabela 14: Coeficientes Estimados — ARMV vs AR

Modelo	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\phi}_5$	$\hat{\phi}_6$	$\hat{\phi}_7$	$\hat{\phi}_8$
AR7	0,06082	0,07740	0,08617	0,07891	0,08376	0,08387	0,06157	—
AR7c	0,06026	0,07683	0,08560	0,07835	0,08320	0,08330	0,06102	—
AR8	0,05998	0,07626	0,08503	0,07784	0,08259	0,08281	0,06075	0,01361
AR8c	0,05945	0,07574	0,08451	0,07733	0,08208	0,08230	0,06025	0,01310
ARMV	0,03039	0,05097	0,09716	0,09984	0,17999	0,21852	0,24670	—

A Tabela 14 apresenta os coeficientes estimados para os modelos ARMV(7) e AR de ordens 7 e 8, com e sem constante. Observa-se que os modelos AR apresentam coeficientes de menor magnitude e relativamente estáveis entre si, refletindo a estrutura linear e homogênea típica desses modelos. Em contrapartida, os coeficientes estimados para o modelo ARMV(7) variam significativamente em magnitude, com destaque para os valores mais elevados associados aos defasagens superiores ( $\hat{\phi}_5$ ,  $\hat{\phi}_6$  e  $\hat{\phi}_7$ ).

## 7 Aplicação do Modelo

### 7.1 Banco de Dados

O conjunto de dados utilizado foi obtido no portal do Instituto Nacional de Meteorologia (INMET) (2024) para a região de Brasília. Escolheu-se a estação OM-BRA (Observatório Meteorológico de Brasília), mantida pelo INMET, e foram extraídas medições diárias no período de 01/01/2001 a 31/12/2024. Ao todo, a série contém 8 767 observações em quatro variáveis: data, temperatura mínima, temperatura média e temperatura máxima. O objetivo desta base é servir como exemplo para a aplicação do método proposto, visando a obtenção de estimativas consistentes dos parâmetros e a construção de um modelo com desempenho competitivo em relação aos modelos autorregressivos tradicionais (AR).

### 7.2 Serie de dados climáticos

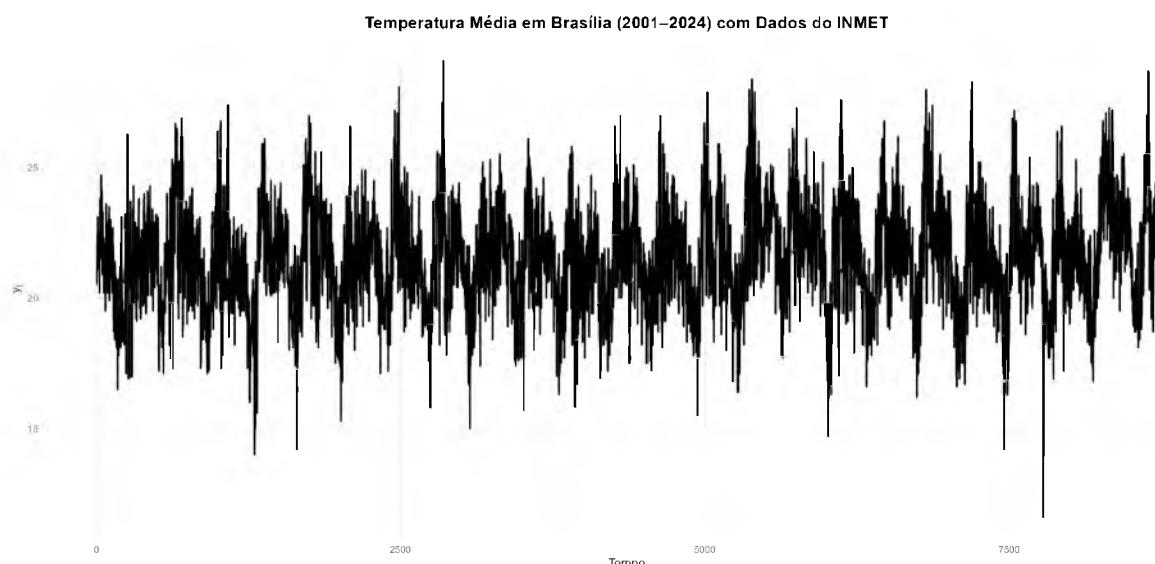


Figura 13: Serie temporal dos dados simulados para um ARMV(6)

A Figura 13 apresenta a série temporal da temperatura média diária registrada em Brasília entre 2001 e 2024. Visualmente, demonstra comportamento estacionário, oscilando em torno de uma média estável ao longo dos anos, sem indicação clara de tendência de aumento ou queda. Observa-se um padrão de flutuação recorrente, compatível com a presença de sazonalidade anual (típica em séries climáticas) com picos e vales ao longo do tempo, refletindo as variações entre estações do ano.

Tabela 15: Resumo dos resultados da série de dados climáticos

Estatística	Valor	Detalhes
Média	21,5861	—
Variância	3,9395	—
Número de observações	8719	—
<b>Teste Augmented Dickey-Fuller</b>		
Estatística Dickey-Fuller	-9,8365	Lag = 20
<i>p</i> -valor	0,01	Hipótese alternativa: estacionária
<b>Teste Phillips-Perron</b>		
Z( $\alpha$ ) de Dickey-Fuller	-1409,8	Lag truncado = 12
<i>p</i> -valor	0,01	Hipótese alternativa: estacionária
<b>Teste KPSS (nível)</b>		
Estatística KPSS	1,2454	Lag truncado = 12
<i>p</i> -valor	0,01	Hipótese nula: estacionariedade em nível

A Tabela 15 apresenta um resumo estatístico da série de temperatura média diária em Brasília no período de 2001 a 2024. A média observada foi de aproximadamente 21,59°C, com variância de 3,94, indicando uma dispersão moderada em torno da média ao longo do tempo. O número total de observações foi de 8.719, caracterizando uma série extensa e adequada para análise de séries temporais.

Para avaliar a estacionariedade da série, foram aplicados três testes formais: o teste de Dickey-Fuller Aumentado (ADF), o teste de Phillips-Perron (PP) e o teste KPSS (nível). Tanto o ADF quanto o PP apresentaram estatísticas de teste significativamente negativas (ADF = -9,84; PP = -1409,8), com *p*-valores iguais a 0,01, rejeitando a hipótese nula de raiz unitária. Esses resultados sugerem que a série é estacionária.

Entretanto, o teste KPSS, que adota a hipótese nula oposta, apresentou um valor estatístico elevado (KPSS = 1,25) com *p*-valor também igual a 0,01, rejeitando essa hipótese. Esse resultado indica possível não estacionariedade em nível, revelando um aparente conflito com os testes ADF e PP.

Essa divergência pode indicar a presença de uma tendência leve ou de uma estrutura quase integrada na série, o que é comum em dados climáticos. Ainda assim, prosseguiremos com a análise para ilustrar a aplicação do método proposto.

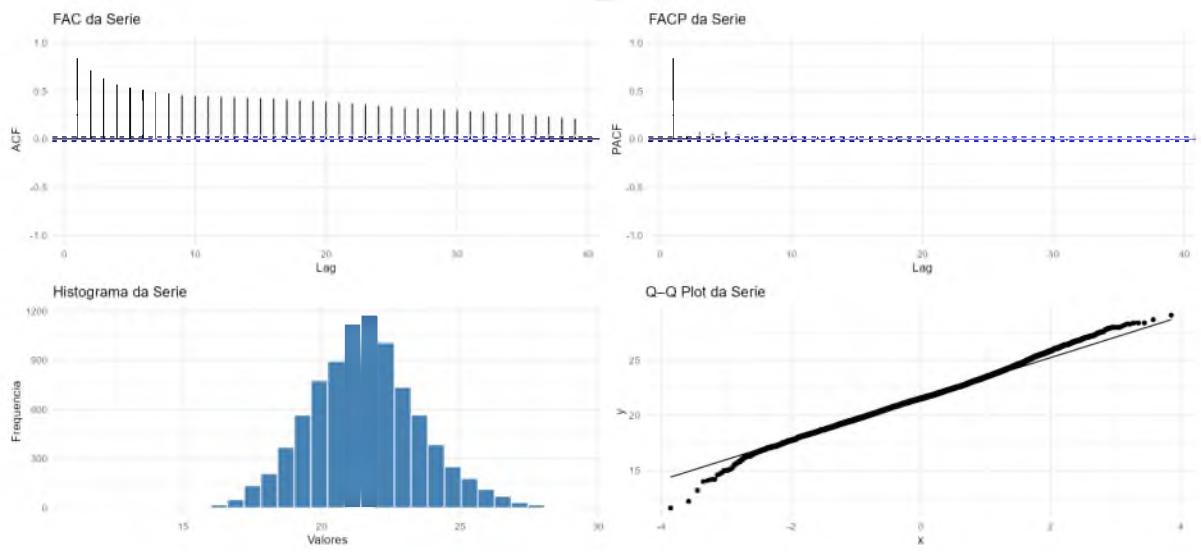


Figura 14: Serie temporal dos dados simulados para um ARMV(6)

A Figura 14 exibe os diagnósticos da série climática, evidenciando uma forte dependência temporal, com a FAC apresentando decaimento lento e a FACP sinalizando corte no primeiro defasagem, característica típica de processos autoregressivos. O histograma revela uma distribuição aproximadamente simétrica, com leve assimetria à esquerda, enquanto o gráfico Q–Q indica aderência moderada à normalidade, com desvios nas extremidades. Esses indícios, em conjunto, reforçam a hipótese de que a série seja estacionária.

### 7.2.1 Estimação dos parâmetros

Nessa etapa utilizamos a função que foi descrita na etapa da metodologia e os resultados obtidos indicam que o modelo de ordem  $p = 6$  fornece o melhor ajuste. A seguir, apresenta-se o resumo do modelo obtido:

Tabela 16: Parâmetros Estimados do Modelo ARMV(6) para os Dados Climáticos

(a) Limiares Estimados		(b) Coeficientes Estimados	
Limiares	Estimativa	Coef.	Estimativa
$\hat{\alpha}_1$	19,5592	$\hat{\phi}_1$	1,0227
$\hat{\alpha}_2$	20,5775	$\hat{\phi}_2$	-0,0091
$\hat{\alpha}_3$	21,5777	$\hat{\phi}_3$	-0,0053
$\hat{\alpha}_4$	22,6182	$\hat{\phi}_4$	-0,0141
$\hat{\alpha}_5$	23,7281	$\hat{\phi}_5$	-0,0091
		$\hat{\phi}_6$	-0,0079

A análise dos parâmetros estimados do modelo AR-MV(6) aplicado à série de dados climáticos revela uma segmentação bem distribuída dos limiares  $\hat{\alpha}_j$ , com valores crescentes entre aproximadamente 19,56 e 23,73. Isso sugere que os regimes foram adequadamente identificados ao longo da faixa de variação da temperatura.

Quanto aos coeficientes  $\hat{\phi}_i$ , observa-se que apenas o primeiro apresenta valor expressivo (1,0227), indicando forte dependência com a defasagem imediata, enquanto os demais coeficientes são próximos de zero, o que aponta para uma estrutura autoregressiva essencialmente de ordem 1.

Tabela 17: Estimativas dos Coeficientes  $\phi$ , Erros Padrão e Intervalos de Confiança

Coef.	Estimativa	Erro Padrão	IC 95%
$\phi_1$	1,0227	0,0016	[1,0195 ; 1,0259]
$\phi_2$	-0,0091	0,0022	[-0,0133 ; -0,0048]
$\phi_3$	-0,0053	0,0019	[-0,0090 ; -0,0016]
$\phi_4$	-0,0141	0,0016	[-0,0173 ; -0,0109]
$\phi_5$	-0,0091	0,0018	[-0,0126 ; -0,0056]
$\phi_6$	-0,0079	0,0020	[-0,0118 ; -0,0041]

Os resultados apresentados na Tabela 17 mostram as estimativas dos coeficientes  $\phi_j$  para o modelo ajustado aos dados climáticos, juntamente com seus respectivos erros

padrão e intervalos de confiança de 95%. Observa-se que o primeiro coeficiente é significativamente diferente de zero e assume valor próximo de 1, sugerindo forte persistência na série. Os demais coeficientes apresentam magnitudes reduzidas e negativas, embora estatisticamente significativas, o que pode indicar correções de curto prazo em relação ao valor anterior. O conjunto das estimativas sugere uma dinâmica autoregressiva com forte dependência temporal e componentes de ajustamento finos nos "lags" subsequentes.

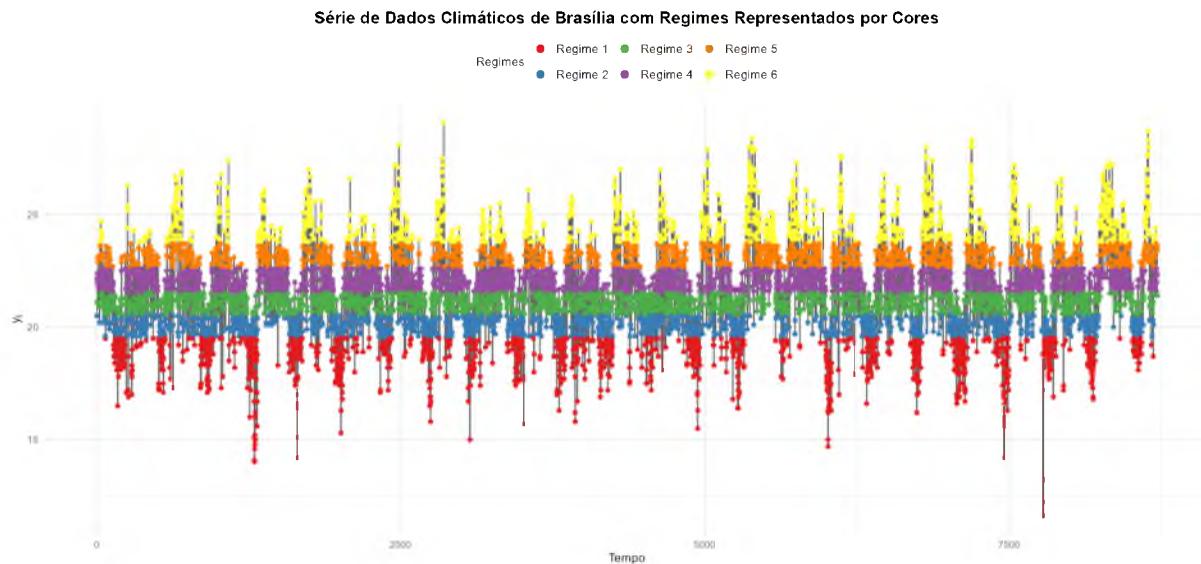


Figura 15: Serie temporal dos dados simulados para um ARMV(6)

A Figura 15 apresenta a série de temperatura média diária em Brasília segmentada em regimes distintos, identificados por diferentes cores, com base nos limiares estimados pelo modelo AR-MV(6). Observa-se que os regimes capturam variações sazonais e oscilações de amplitude ao longo do tempo.

Tabela 18: Resumo Estatístico por Regime da Série Climática com Segmentação AR-MV(6)

Regime	N de Observações	Média	Mediana	Variância
1	1276	18,5593	18,8	0,8753
2	1297	20,0744	20,1	0,0801
3	1803	21,0806	21,1	0,0807
4	1974	22,0748	22,1	0,1025
5	1221	23,1616	23,2	0,0974
6	1148	24,9362	24,7	1,0080

A Tabela 18 apresenta um resumo estatístico da série climática segmentada pelos regimes identificados via modelo AR-MV(6). Observa-se uma clara ordenação crescente nos valores médios e medianos conforme o regime, indicando que os limiares estimados

conseguiram identificar faixas distintas de temperatura.

### 7.2.2 Avaliação de ajuste do modelo

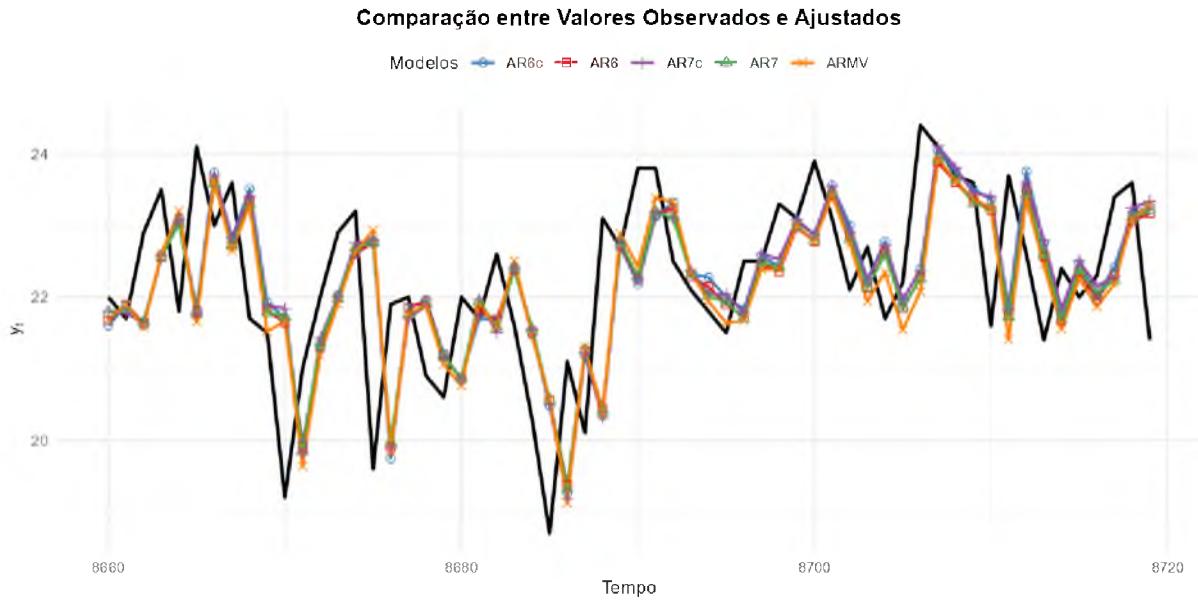


Figura 16: Serie temporal dos dados simulados para um ARMV(6)

Tabela 19: Métricas de Ajuste — ARMV vs AR

Modelo	AIC	AICc	BIC	MAE	MSE	RMSE	MAPE
ARMV	26262,74	26262,75	26312,25	0,83	1,19	1,09	3,90
AR6c	26046,41	26046,42	26102,99	1,30	2,81	1,67	6,09
AR6	26349,79	26349,80	26399,30	1,44	3,45	1,85	6,75
AR7c	26030,82	26030,84	26094,48	1,30	2,81	1,67	6,10
AR7	26299,78	26299,79	26356,36	1,43	3,40	1,84	6,71

A Tabela 19 apresenta as métricas de ajuste dos modelos AR e ARMV aplicados à série climática. Observa-se que o modelo ARMV obteve o menor valor de  $\hat{\sigma}^2$ , menor erro médio absoluto (MAE), menor erro quadrático médio (MSE) e o menor erro percentual médio absoluto (MAPE), evidenciando melhor desempenho geral na modelagem da série.

Apesar de os modelos AR com constante (AR6c e AR7c) apresentarem valores ligeiramente menores de AIC, o ARMV se destaca por capturar mais eficientemente a estrutura dos dados, entregando previsões com menor erro absoluto e percentual. Isso sugere que a modelagem por regimes com limiares oferece uma representação mais precisa para a série de temperatura.

Tabela 20: Coeficientes Estimados — ARMV vs AR

Modelo	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\phi}_5$	$\hat{\phi}_6$	$\hat{\phi}_7$
AR6	0,81729	0,00361	0,03065	0,00866	0,04458	0,09448	—
AR6c	0,78023	-0,00728	0,01958	-0,00239	0,03370	0,05751	—
AR7	0,81001	0,00014	0,02997	0,00631	0,04429	0,03144	0,07714
AR7c	0,77766	-0,00881	0,01969	-0,00326	0,03402	0,02249	0,04489
ARMV	1,02271	-0,00909	-0,00529	-0,01409	-0,00914	-0,00792	—

A Tabela 20 apresenta os coeficientes estimados para os modelos AR e ARMV aplicados à série de dados climáticos. Observa-se que, no modelo ARMV, o coeficiente associado ao primeiro defasado ( $\hat{\phi}_1$ ) é significativamente maior do que os demais, os quais apresentam valores bastante reduzidos. Isso sugere que o modelo ARMV dá maior peso ao valor imediatamente anterior da série, o que tende a estabilizar as previsões em torno do último valor observado.

Por outro lado, os modelos AR tradicionais apresentam coeficientes mais distribuídos a partir de  $\hat{\phi}_2$ , indicando que esses modelos incorporam de maneira mais ativa informações de múltiplas defasagens no processo preditivo, o que pode torná-los mais flexíveis, porém potencialmente mais sensíveis a variações na série.

### 7.2.3 Estudo dos resíduos do modelo ARMV(6) aplicado aos Dados Climáticos

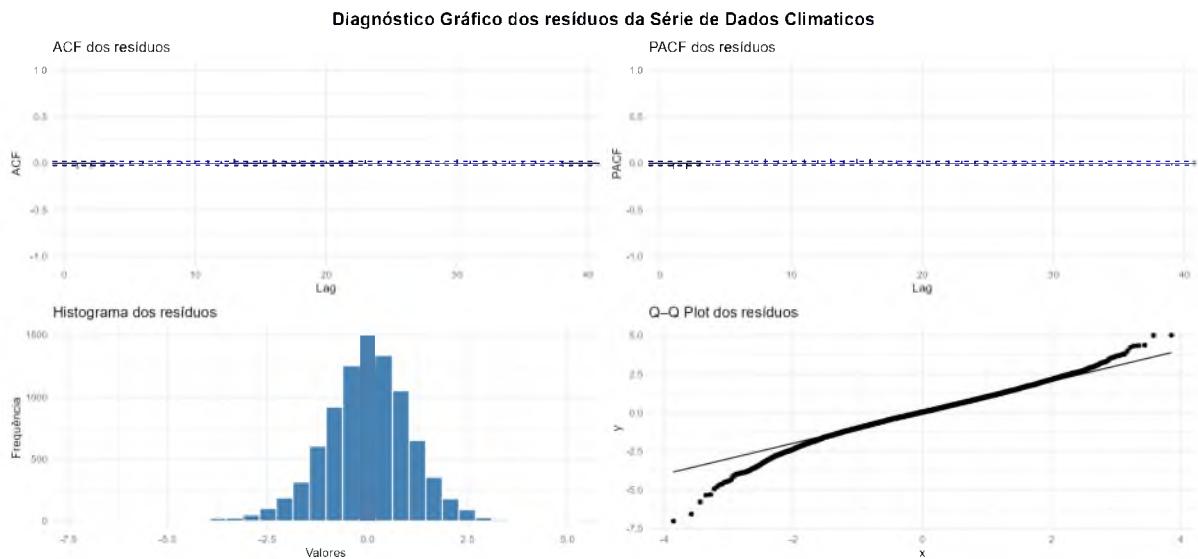


Figura 17: Serie temporal dos dados simulados para um ARMV(6)

Na Figura 17, são apresentados quatro gráficos de diagnóstico dos resíduos do modelo AR-MV(6) ajustado à série climática. A função de autocorrelação (FAC) e a

função de autocorrelação parcial (FACP) mostram que todos os coeficientes se mantêm dentro dos limites de significância, indicando ausência de dependência serial nos resíduos. O histograma revela uma distribuição aproximadamente simétrica e centrada em zero, sugerindo normalidade. Já o gráfico Q–Q exibe alinhamento satisfatório dos resíduos à reta teórica, com pequenas discrepâncias nas caudas. Esses resultados atestam a adequação do modelo ajustado, uma vez que os resíduos se comportam como ruído branco, não autocorrelacionados e aproximadamente gaussianos.

### 7.2.4 Estudo preditivo do modelo

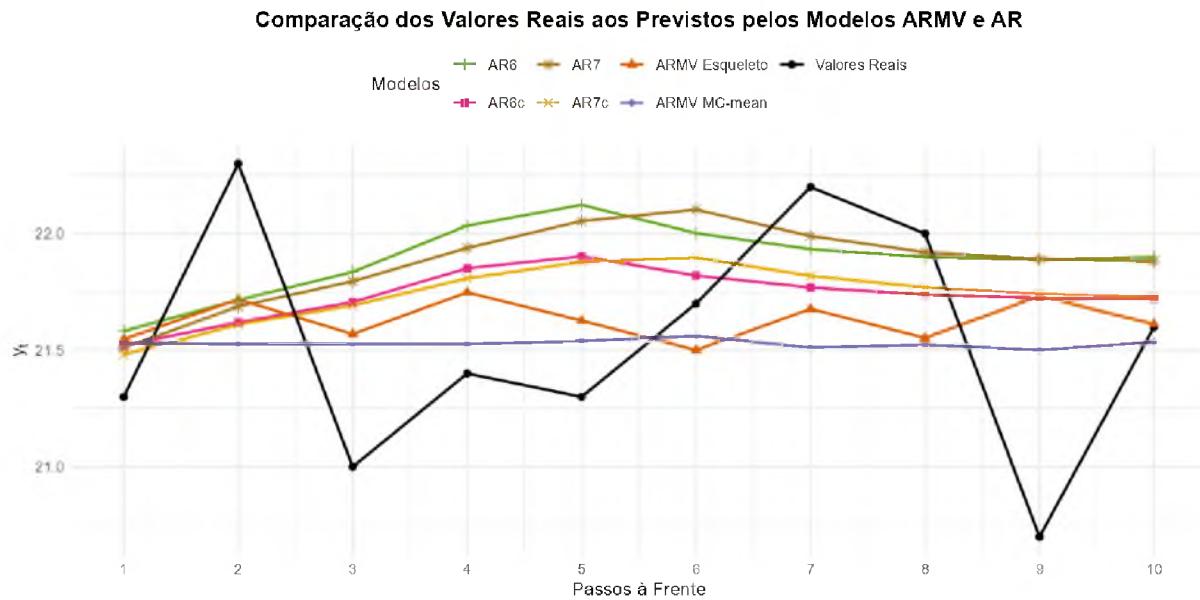


Figura 18: Serie temporal dos dados simulados para um ARMV(6)

O Gráfico 18 apresenta a comparação entre os valores reais da série de temperatura média (linha preta) e as previsões geradas por diferentes modelos para os 10 passos à frente. Observa-se que os modelos ARMV Esqueleto e MC-mean mantêm previsões mais estáveis e próximas da média da série, ajustando-se melhor aos valores reais nesse horizonte. Em contraste, os modelos AR tendem a se afastar dos dados observados, especialmente nos tempos 3 a 6. Esse bom desempenho do ARMV está associado à maior influência do coeficiente associado ao primeiro defasado (AR1), o que contribui para a preservação da tendência local recente da série, promovendo maior estabilidade nas previsões.

Tabela 21: Métricas de Previsão a 1–10 Passos à Frente

Modelo	MAE	MSE	RMSE	MAPE
ARMV Esqueleto	0,42871	0,25302	0,50301	1,99914
ARMV MC-mean	0,40637	0,23678	0,48660	1,88378
AR6c	0,46183	0,29076	0,53922	2,15764
AR6	0,53167	0,38732	0,62235	2,49304
AR7c	0,45296	0,28291	0,53189	2,11673
AR7	0,50731	0,36183	0,60152	2,37869

A Tabela 21 apresenta métricas sobre a qualidade preditiva dos modelos analisados. Os dois modelos baseados na estrutura ARMV obtiveram os menores erros MAE,

MSE, RMSE e MAPE. O modelo ARMV MC-mean apresentou o melhor desempenho geral, com  $MAE = 0,40637$  e  $MAPE = 1,88\%$ , indicando uma boa capacidade preditiva. Por outro lado, os modelos AR6 e AR7, que não consideram regimes, apresentaram os maiores erros .

## 8 Conclusão

Para desenvolver este trabalho, foi necessário elaborar uma nova estratégia de estimação de todos os parâmetros do modelo AR-MV( $p$ ) exclusivamente a partir dos dados observados. Embora Fadel (2012) já tenha proposto métodos para estimar os limiares  $\alpha$  e os coeficientes  $\phi$ , a minimização direta da função de resíduos mostrava-se altamente dependente dos chutes iniciais, devido à complexidade de sua superfície de otimização Fadel (2012). Para superar esse desafio e garantir consistência e escalabilidade, adotamos o algoritmo  $k$ -means na etapa de identificação dos limiares: sua simplicidade e eficiência facilitam a replicação do procedimento, um dos objetivos centrais desta pesquisa.

Além disso, desenvolvemos em R um conjunto de funções modulares e dinâmicas que automatizam todo o fluxo de trabalho com modelos AR-MV( $p$ ): desde a simulação de séries, passando pelo ajuste completo, tanto pelo procedimento “esqueleto” quanto por Monte Carlo, até a geração de previsões e o cálculo de métricas de desempenho. Dentre essas funções, destacam-se uma função genérica que estima simultaneamente a ordem  $p$ , os limiares  $\alpha$  e os coeficientes  $\phi$  de qualquer AR-MV( $p$ ), e uma função de seleção de modelo que, com base no teste da razão de verossimilhança, escolhe automaticamente o modelo mais adequado.

Nos estudos de simulação com AR-MV(6) e AR-MV(7), simulamos séries com parâmetros conhecidos e estimamos simultaneamente a ordem, os limiares e os coeficientes. Os resultados mostraram que o procedimento proposto recupera com boa precisão os valores verdadeiros e geram modelos com desempenho igual ou melhor na maior parte das métricas, validando a combinação de agrupamento e critérios de informação para definir a estrutura do modelo unicamente a partir dos dados.

Na aplicação empírica à série diária de temperatura média de Brasília (2001–2024), o modelo AR-MV apresentou melhor desempenho de ajuste e previsão comparado aos modelos AR lineares tradicionais. Observou-se ainda que, no AR-MV(6), apenas o primeiro coeficiente teve impacto relevante na previsão, enquanto os demais atuaram como ajustes finos, sem prejudicar a acurácia global.

Reconhecem-se, porém, limitações importantes. A estimativa conjunta continua sensível aos valores iniciais: chutes mal calibrados podem comprometer a convergência e gerar limiares distantes dos reais, especialmente quando os regimes têm tamanhos muito desiguais. Ademais, o método ainda não dispõe de instrumentos formais de inferência para os limiares, restringindo a quantificação de sua incerteza.

Em suma, este trabalho avança a metodologia de estimativa de modelos AR-MV, oferecendo uma abordagem prática, replicável e eficiente para parametrização completa a partir de dados. Pesquisas futuras podem focar em otimizações computacionais dos

algoritmos, extensão a componentes sazonais e desenvolvimento de métodos de inferência formal para os limiares.

## Referências

- BOX, G. E. P.; JENKINS, G. M. *Time Series Analysis: Forecasting and Control*. Revised edition, 1976. San Francisco: Holden-Day, 1970. 575 p. (Holden-Day series in time series analysis and digital processing). Original from the University of Michigan, digitized on August 27, 2007. ISBN 0816211043, 9780816211043.
- BUENO, R. L. S. *Econometria de Séries Temporais*. 2. ed. [S.l.]: Cengage Learning, 2011.
- CASELLA, G.; BERGER, R. L. *Statistical Inference*. 2. ed. Pacific Grove, CA: Duxbury Press, 2002.
- FADEL, D. F. *Modelos Autorregressivos com Memória Variável*. Dissertação (Dissertação de Mestrado) — Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica, Departamento de Estatística, 2012. Orientadora: Profa. Dra. Nancy Lopes Garcia.
- Instituto Nacional de Meteorologia (INMET). *Portal de Acesso a Dados Meteorológicos*. 2024. <https://portal.inmet.gov.br>.
- LOUREIRO, L. S. *Aplicações de modelos autorregressivos com memória variável à dendrocronologia*. Dissertação (Dissertação de Mestrado) — Universidade de Brasília, Instituto de Ciências Exatas, Departamento de Estatística, 2018. Orientador: Prof. Dr. Jhames Matos Sampaio. Coorientador: Prof. Dr. Lucas Moreira.
- MACQUEEN, J. B. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume-1*. [S.l.]: University of California Press, 1967. p. 281–297.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de séries temporais*. 2. ed. São Paulo: Edgard Blucher, 2006. ISBN 978-85-212-0389-6.
- RANGEL, P. A. *Modelos Autorregressivos com Memória Variável*. Dissertação (Dissertação (Mestrado)) — Universidade de Brasília, Departamento de Estatística, Brasília, Brasil, 2020.
- TONG, H. Discussion of a paper by a. j. lawrance and n. t. kottegoda. *J. R. Statist. Soc. A*, v. 140, p. 34–35, 1977.
- TONG, H. On a threshold model. *Research Memorandum*, 1978. Disponível em: <https://warwick.ac.uk/fac/sci/statistics/staff/academic-research/tong/memoranda/tong78.pdf>.
- TONG, H.; LIM, K. S. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 42, n. 3, p. 245–292, 1980.
- TSAY, R. S. Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 1989. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1989.10478882>.