



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

Aplicações de modelos autorregressivos com
memória variável à dendrocronologia

por

Leandro Siller Loureiro

Orientador: Prof. Dr. Jhames Matos Sampaio
Coorientador: Prof. Dr. Lucas Moreira

Brasília
2018

Aplicações de modelos autorregressivos com memória variável à dendrocronologia

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília como parte dos requisitos necessários à obtenção do título de Mestre em Estatística.

Comissão Julgadora:

- Prof. Dr. Jhames Matos Sampaio (orientador) - UnB
- Prof. Dr. Eduardo de Oliveira Horta - UFRGS
- Prof. Dr. José Augusto Fiorucci - UnB

Dedicatória

*À minha amada esposa,
Priscila.
Ao meu querido filho,
Miguel.*

Agradecimentos

Em primeiro lugar, a Deus, por ter me concedido, em abundância, bênçãos, saúde e paz fundamentais para que eu pudesse concluir o curso.

À Priscila, minha esposa, por todo apoio e paciência; e a Miguel, meu filho, que em breve virá ao mundo e, que desde já, está sendo motivo de grande alegria para toda a família.

A meus pais, Mario e Izabel, responsáveis pela minha educação e pela formação do meu caráter.

A meus irmãos, Luciano e Leonardo, pelo companheirismo em todos os momentos, principalmente, no período em que vivemos juntos.

Aos meus chefes imediatos no Centro de Comunicação Social da Marinha, em especial ao Almirante Rocha, por ter autorizado a minha ausência nos horários de aulas presenciais, e aos Comandantes Glatthardt, Adriano, Marcos Simas, Chaves e Serrano pela compreensão nos momentos em que não estive presente.

Aos meus companheiros de trabalho, por sempre serem solícitos, nos momentos em que necessitei efetuar uma troca de serviço para assistir às aulas, e por terem, ocasionalmente, sofrido sobrecarga de trabalho, em função de minha ausência.

Ao professor Jhames, meu orientador, pela confiança em mim depositada e pelo seu contagiante otimismo, fator relevante para que eu pudesse dar andamento, de maneira tranquila, ao trabalho.

Ao professor Lucas, meu coorientador, pelas excelentes sugestões e colocações em relação à dissertação, extremamente importantes para o resultado final do trabalho.

Aos professores pertencentes ao Programa de Pós-Graduação em Estatística da UnB (PPGEST/UnB), em especial aos professores Bernardo, Cira e Juliana por todos os conselhos e ensinamentos, e aos professores Antônio Eduardo, Cibele, Gustavo e Raul pela excelência na ministração de suas respectivas disciplinas.

Ao professor do Instituto de Matemática e Estatística da Universidade do Estado do Rio de Janeiro (IME/UERJ), José Francisco, meu orientador de graduação, pessoa extremamente importante na minha formação em Estatística, exemplo de profissional, dedicado e motivado no desempenho de suas atividades acadêmicas.

Ao professor do IME/UERJ, André Luiz, pela sua importância em minha formação acadêmica e por ter confiado em minha capacidade intelectual ao me recomendar ao PPGEST/UnB.

Aos meus amigos, há quase 8 anos, Jonathan, Lívia e Wander, pelo aprendizado, apoio

e incentivo que sempre me ofereceram durante todo esse período.

Aos meus amigos, Damião e Elisângela, pessoas agradabilíssimas, e grandes companheiros durante o mestrado, por toda a ajuda e motivação constantemente ofertadas.

Aos demais alunos do PPGEST/UnB, no período de 2016 a 2017, pela disponibilidade e auxílio, especialmente, nas reuniões em grupos de estudo que eventualmente realizávamos.

À Désirée, autora da principal referência desta dissertação, por sua generosa ajuda, ao disponibilizar seus códigos computacionais, o que foi fundamental para a realização deste trabalho.

Ao PPGEST/UnB, pela excelência no programa de pós-graduação oferecido, o que consequentemente, aperfeiçoou meu nível teórico/prático na área de Estatística.

Ao IME/UERJ, instituto de excelência onde me graduei, e pelo qual tenho enorme carinho, por ser responsável pela minha ascensão profissional e acadêmica.

Aos professores Eduardo de Oliveira Horta (UFRGS) e José Augusto Fiorucci (UnB) por contribuírem com o trabalho, fazendo sugestões e críticas construtivas.

A todas as pessoas que, direta ou indiretamente, ajudaram-me nessa difícil, porém, gratificante jornada.

Resumo

LOUREIRO, L. S. **Aplicações de modelos autorregressivos com memória variável à dendrocronologia**. 2018. Dissertação (Mestrado) - Departamento de Pós-Graduação em Estatística, Universidade de Brasília, Brasília, 2018.

Neste trabalho generalizamos o estudo realizado por Fadel (2012) sobre os modelos autorregressivos com memória variável (AR-MV). Mais precisamente, verificamos a manutenção das propriedades observadas nos AR-MV de segunda e terceira ordens em modelos AR-MV com ordens superiores (quatro e cinco). Por intermédio de estudos simulados, avaliamos o comportamento gráfico dos modelos AR-MV. Comparamos o desempenho e a capacidade preditiva dos modelos AR-MV de ordens quatro e cinco com modelos autorregressivos (AR) de ordens idênticas. Analisamos as propriedades assintóticas dos modelos AR-MV de quarta e quinta ordens. Além disso, confrontamos os modelos AR-MV e AR em quatro aplicações a dados dendrocronológicos reais.

Palavras-chave: Análise de séries temporais; Dendrocronologia; Modelos autorregressivos; Modelos autorregressivos com memória variável.

Abstract

LOUREIRO, L. S. **Applications of autoregressive models with variable memory to dendrochronology**. 2018. Dissertação (Mestrado) - Departamento de Pós-Graduação em Estatística, Universidade de Brasília, Brasília, 2018.

In this work we generalize the study carried out by Fadel (2012) on autoregressive models with variable memory (AR-MV). More precisely, we observed the maintenance of the properties observed in second and third orders AR-MV in AR-MV models with higher orders (four and five). By means of simulated studies, we evaluated the graphical behavior of the AR-MV models. We compared the performance and predictive capacity of AR-MV models of orders four and five with autoregressive (AR) models of identical orders. We analyzed the asymptotic properties of fourth and fifth order AR-MV models. In addition, we compared the AR-MV and AR models in four applications with real dendrochronological data.

Keywords: Autoregressive models; Autoregressive models with variable memory; Dendrochronology; Time series analysis.

Sumário

Lista de Abreviaturas	xiii
Lista de Figuras	xv
Lista de Tabelas	xxi
1 Introdução	1
2 Revisão de literatura	5
2.1 Processos estocásticos e séries temporais	5
2.1.1 Processos estocásticos estacionários	6
2.1.2 Processos estocásticos ergódicos	7
2.1.3 Ruído branco	8
2.2 Abordagem de Box e Jenkins	8
2.2.1 Modelos ARMA	9
2.2.2 Funções de autocorrelação e autocorrelação parcial	11
2.2.3 Análise de estacionariedade	15
2.2.4 Identificação do modelo	16
2.2.5 Diagnóstico do modelo	18
2.3 Séries Temporais Lineares e Não-Lineares	19
2.4 Modelos autorregressivos com limiar	22
2.4.1 Modelos autorregressivos com limiar autoexcitantes	22
3 Modelos autorregressivos com memória variável	27
3.1 Ergodicidade geométrica para um modelo AR-MV	27
3.2 Ilustração de propriedades do modelo por meio de simulações	29
3.2.1 AR-MV(2)	30
3.2.2 AR-MV(3)	32
3.2.3 AR-MV(4)	35
3.2.4 AR-MV(5)	37
3.3 Estimação dos parâmetros	40
3.4 Previsão	41
3.4.1 Método do Esqueleto	42

3.4.2	Método de Monte Carlo	43
3.4.3	Erro de previsão	44
3.5	Estudos simulados	44
3.5.1	AR-MV(4)	45
3.5.2	AR-MV(5)	50
3.6	Propriedades assintóticas	60
3.6.1	AR-MV(4)	60
3.6.2	AR-MV(5)	67
4	Aplicações	75
4.1	Aplicação I - AR-MV(2)	76
4.1.1	Banco de dados	76
4.1.2	Análise de normalidade dos dados	77
4.1.3	Análise de estacionariedade dos dados	78
4.1.4	Identificação do modelo ARIMA	78
4.1.5	Diagnóstico do modelo ARIMA selecionado	79
4.1.6	Estimação e diagnóstico do modelo AR-MV	81
4.1.7	Comparação dos modelos ARIMA e AR-MV ajustados	83
4.1.8	Previsão	83
4.2	Aplicação II - AR-MV(3)	84
4.2.1	Banco de dados	84
4.2.2	Análise de normalidade dos dados	85
4.2.3	Análise de estacionariedade dos dados	86
4.2.4	Identificação do modelo ARIMA	86
4.2.5	Diagnóstico do modelo ARIMA selecionado	88
4.2.6	Estimação e diagnóstico do modelo AR-MV	89
4.2.7	Comparação dos modelos ARIMA e AR-MV ajustados	91
4.2.8	Previsão	92
4.3	Aplicação III - AR-MV(4)	93
4.3.1	Banco de dados	93
4.3.2	Análise de normalidade dos dados	93
4.3.3	Análise de estacionariedade dos dados	94
4.3.4	Identificação do modelo ARIMA	95
4.3.5	Diagnóstico do modelo ARIMA selecionado	97
4.3.6	Estimação e diagnóstico do modelo AR-MV	98
4.3.7	Comparação dos modelos ARIMA e AR-MV ajustados	100
4.3.8	Previsão	101
4.4	Aplicação IV - AR-MV(5)	102
4.4.1	Banco de dados	102
4.4.2	Análise de normalidade dos dados	102

4.4.3	Análise de estacionariedade dos dados	103
4.4.4	Identificação do modelo ARIMA	104
4.4.5	Diagnóstico do modelo ARIMA selecionado	105
4.4.6	Estimação e diagnóstico do modelo AR-MV	106
4.4.7	Comparação dos modelos ARIMA e AR-MV ajustados	108
4.4.8	Previsão	109
5	Considerações finais	111
	Referências Bibliográficas	113
	Apêndice A – Testes para detecção de raízes unitárias	117
A.1	Teste de Dickey-Fuller aumentado	117
A.2	Teste de Phillips-Perron	118
A.3	Teste KPSS	118
	Apêndice B – Testes para detecção de não linearidade	121
B.1	Teste de Keenan	121
B.2	Teste de Tsay	121
B.3	Teste da não linearidade de limiar	122
	Apêndice C – Testes para detecção de normalidade	123
C.1	Teste de Kolmogorov-Smirnov	123
C.2	Teste de Lilliefors	123
C.3	Teste de Cramer-von Mises	124
C.4	Teste de Shapiro-Wilk	124
C.5	Teste de Shapiro-Francia	124
C.6	Teste de Anderson-Darling	125

Lista de Abreviaturas

ADF:	Dickey-Fuller aumentado (<i>Dickey-Fuller augmented</i>).
AIC:	Critério de Informação de Akaike (<i>Akaike information criterion</i>).
AR(p):	Autorregressivo de ordem p (<i>Autoregressive of order p</i>).
ARIMA(p, d, q):	Autorregressivo integrado de médias móveis de ordem p, d, q (<i>Integrated autoregressive of moving averages of order p, d, q</i>).
ARMA(p, q):	Autorregressivo de médias móveis de ordem p, q (<i>Autoregressive of moving averages of order p, q</i>).
AR-MV(p):	Autorregressivo com memória variável de ordem (p) (<i>Autoregressive with variable memory of order p</i>).
BIC:	Critério de Informação Bayesiano (<i>Bayesian information criterion</i>).
ESQ:	Esqueleto (<i>Skeleton</i>).
FAC:	Função de autocorrelação (<i>Autocorrelation function</i>).
FACP:	Função de autocorrelação parcial (<i>Partial autocorrelation function</i>).
FACV:	Função de autocovariância (<i>Autocovariance function</i>).
fda:	Função de distribuição acumulada (<i>Cumulative distribution function</i>).
iid:	Independente e identicamente distribuída (<i>Independently and identically distributed</i>).
KPSS:	Kwiatkowski, Phillips, Schmidt e Shin (<i>Kwiatkowski, Phillips, Schmidt e Shin</i>).
MA(q):	Média móvel de ordem q (<i>Moving Average of order q</i>).
MAE:	Média absoluta dos erros de previsão (<i>Mean absolute prediction error</i>).
MC:	Monte Carlo (<i>Monte Carlo</i>).
MQE:	Média quadrática dos erros de previsão (<i>Mean square prediction error</i>).
NLAR:	Autorregressivo não linear (<i>Autoregressive non-linear</i>).
PP:	Phillips-Perron (<i>Phillips-Perron</i>).
Q-Q:	Quantil-quantil (<i>Quantile-quantile</i>).
RB:	Ruído branco (<i>White noise</i>).
TAR:	Autorregressivo com limiar (<i>Threshold autoregressive</i>).
SETAR:	Autorregressivo com limiar autoexcitante (<i>Self-exciting threshold autoregressive model</i>).
SQR:	Soma dos quadrados dos resíduos (<i>Sum of squared residuals</i>).

Lista de Figuras

2.1	Simulações da FAC e da FACP de um processo AR(1), com parâmetros $\phi_1 = (0.5, -0.5)$ e $n = 1000$	16
2.2	Simulações da FAC e da FACP de um processo MA(1), com parâmetros $\theta_1 = (0.5, -0.5)$ e $n = 1000$	17
2.3	Simulações da FAC e da FACP de um processo ARMA(1,1), com parâmetros $\phi_1 = (0.5, -0.5), \theta_1 = (0.5, -0.5)$ e $n = 1000$	17
2.4	Diagramas de dispersão de Z_t contra os <i>lags</i> 1 a 6 de um processo AR(1) simulado com $\phi_1 = 0.9$ e $n = 200$	20
2.5	Diagramas de dispersão de Z_t contra os <i>lags</i> 1 a 6 de um processo TAR (2,1,1) simulado com $\phi_{1,1} = 0.9, \phi_{2,1} = -1.8$ e $n = 200$	21
3.1	Série temporal dos dados simulados para o Exemplo 3.1.	30
3.2	(a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.1.	31
3.3	Série temporal dos dados simulados para o Exemplo 3.2.	31
3.4	(a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.2.	32
3.5	Série temporal dos dados simulados para o Exemplo 3.3.	33
3.6	(a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.3.	33
3.7	Série temporal dos dados simulados para o Exemplo 3.4.	34
3.8	(a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.4.	34
3.9	Série temporal dos dados simulados para o Exemplo 3.5.	35
3.10	(a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.5.	36
3.11	Série temporal dos dados simulados para o Exemplo 3.6.	37
3.12	(a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.6.	37
3.13	Série temporal dos dados simulados para o Exemplo 3.7.	38

3.14	(a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.7.	39
3.15	Série temporal dos dados simulados para o Exemplo 3.8.	40
3.16	(a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.8.	40
3.17	Série temporal dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n=500$	46
3.18	Série temporal dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n=10000$	46
3.19	(a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n = 500$	48
3.20	(a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n = 10000$	48
3.21	(a) Histograma e (b) gráfico Q-Q dos resíduos dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n = 500$	49
3.22	(a) Histograma e (b) gráfico Q-Q dos resíduos dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n = 10000$	49
3.23	(a) Dispersão, (b) FAC e (c) FACP dos resíduos dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n = 500$	50
3.24	(a) Dispersão, (b) FAC e (c) FACP dos resíduos dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n = 10000$	50
3.25	Valores observados e previsões pontuais, sob os modelos ajustados para os dados simulados no Exemplo 3.9, com origem em 500 para $n=500$	51
3.26	Valores observados e previsões pontuais, sob os modelos ajustados para os dados simulados no Exemplo 3.9, com origem em 10000 para $n=10000$	51
3.27	Série temporal dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n=500$	53
3.28	Série temporal dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n=10000$	53
3.29	(a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n = 500$	55
3.30	(a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP amostrais dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n = 10000$	55
3.31	(a) Histograma e (b) gráfico Q-Q dos resíduos dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n = 500$	56
3.32	(a) Histograma e (b) gráfico Q-Q dos resíduos dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n = 10000$	56
3.33	(a) Dispersão, (b) FAC e (c) FACP dos resíduos dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n = 500$	57

3.34 (a) Dispersão, (b) FAC e (c) FACP dos resíduos dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n = 10000$ 57

3.35 Valores observados e previsões pontuais, sob os modelos ajustados para os dados simulados no Exemplo 3.10, com origem em 500 para $n=500$ 59

3.36 Valores observados e previsões pontuais, sob os modelos ajustados para os dados simulados no Exemplo 3.10, com origem em 10000 para $n=10000$ 59

3.37 (a), (b) e (c) Histogramas de $\hat{\alpha}_1, \hat{\alpha}_2$ e $\hat{\alpha}_3$ e (d), (e) e (f) dispersões de $\hat{\alpha}_1, \hat{\alpha}_2$ e $\hat{\alpha}_3$ para o Exemplo 3.11 com $n=500$ 62

3.38 (a), (b), (c) e (d) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$, (e), (f), (g) e (h) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ e (i), (j), (k) e (l) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ para o Exemplo 3.11 com $n=500$, quando α é conhecido. 63

3.39 (a), (b), (c) e (d) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$, (e), (f), (g) e (h) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ e (i), (j), (k) e (l) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ para o Exemplo 3.11 com $n=500$, quando α é desconhecido. 63

3.40 (a), (b) e (c) Histogramas de $\hat{\alpha}_1, \hat{\alpha}_2$ e $\hat{\alpha}_3$ e (d), (e) e (f) dispersões de $\hat{\alpha}_1, \hat{\alpha}_2$ e $\hat{\alpha}_3$ para o Exemplo 3.11 com $n=1000$ 64

3.41 (a), (b), (c) e (d) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$, (e), (f), (g) e (h) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ e (i), (j), (k) e (l) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ para o Exemplo 3.11 com $n=1000$, quando α é conhecido. 64

3.42 (a), (b), (c) e (d) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$, (e), (f), (g) e (h) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ e (i), (j), (k) e (l) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ para o Exemplo 3.11 com $n=1000$, considerando α desconhecido. 65

3.43 (a), (b) e (c) Histogramas de $\hat{\alpha}_1, \hat{\alpha}_2$ e $\hat{\alpha}_3$ e (d), (e) e (f) dispersões de $\hat{\alpha}_1, \hat{\alpha}_2$ e $\hat{\alpha}_3$ para o Exemplo 3.11 com $n=5000$ 66

3.44 (a), (b), (c) e (d) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$, (e), (f), (g) e (h) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ e (i), (j), (k) e (l) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ para o Exemplo 3.11 com $n=5000$, quando α é conhecido. 66

3.45 (a), (b), (c) e (d) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$, (e), (f), (g) e (h) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ e (i), (j), (k) e (l) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ para o Exemplo 3.11 com $n=5000$, quando α é desconhecido. 67

3.46 (a), (b), (c) e (d) Histogramas de $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ e $\hat{\alpha}_4$ e (e), (f), (g) e (h) dispersões de $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ e $\hat{\alpha}_4$ para o Exemplo 3.12 com $n=500$ 70

3.47 (a), (b), (c), (d) e (e) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$, (f), (g), (h), (i) e (j) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ e (k), (l), (m), (n) e (o) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ para o Exemplo 3.12 com $n=500$, quando α é conhecido. 70

3.48 (a), (b), (c), (d) e (e) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$, (f), (g), (h), (i) e (j) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ e (k), (l), (m), (n) e (o) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ para o Exemplo 3.12 com $n=500$, quando α é desconhecido. 71

3.49 (a), (b), (c) e (d) Histogramas de $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ e $\hat{\alpha}_4$ e (e), (f), (g) e (h) dispersões de $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ e $\hat{\alpha}_4$ para o Exemplo 3.12 com $n=1000$ 71

3.50	(a), (b), (c), (d) e (e) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$, (f), (g), (h), (i) e (j) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ e (k), (l), (m), (n) e (o) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ para o Exemplo 3.12 com $n=1000$, quando α é conhecido.	72
3.51	(a), (b), (c), (d) e (e) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$, (f), (g), (h), (i) e (j) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ e (k), (l), (m), (n) e (o) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ para o Exemplo 3.12 com $n=1000$, quando α é desconhecido.	72
3.52	(a), (b), (c) e (d) Histogramas de $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ e $\hat{\alpha}_4$ e (e), (f), (g) e (h) dispersões de $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ e $\hat{\alpha}_4$ para o Exemplo 3.12 com $n=5000$	73
3.53	(a), (b), (c), (d) e (e) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$, (f), (g), (h), (i) e (j) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ e (k), (l), (m), (n) e (o) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ para o Exemplo 3.12 com $n=5000$, quando α é conhecido.	74
3.54	(a), (b), (c), (d) e (e) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$, (f), (g), (h), (i) e (j) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ e (k), (l), (m), (n) e (o) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ para o Exemplo 3.12 com $n=5000$, quando α é desconhecido.	74
4.1	Ilustração do corte de uma árvore (CCVFloresta (2018)).	75
4.2	(a) Histograma e (b) gráfico Q-Q para a série temporal gerada para as medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).	77
4.3	Série temporal gerada para as medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).	78
4.4	(a) FAC e (b) FACP para a série temporal das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).	79
4.5	(a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo AR(2) para a série temporal padronizada das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).	80
4.6	(a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo AR-MV(2) para a série temporal padronizada das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).	81
4.7	(a) Histograma e (b) gráfico Q-Q para a série temporal gerada para as medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).	82
4.8	Valores observados e previsões pontuais dos modelos AR(2) e AR-MV(2) para a série temporal padronizada das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA, para o período de 1961 a 1965.	84
4.9	(a) Histograma e (b) gráfico Q-Q para a série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).	85
4.10	Série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).	86

4.11 (a) FAC e (b) FACP amostrais para as medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978). 87

4.12 (a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo AR(3) para as medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978). 88

4.13 (a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo AR-MV(3) para a série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978). 90

4.14 (a) Histograma e (b) gráfico Q-Q para a série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978). 91

4.15 Valores observados e previsões pontuais dos modelos AR(3) e AR-MV(3) da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA, para o período de 1979 a 1983. 93

4.16 (a) Histograma e (b) gráfico Q-Q para a série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980). 94

4.17 Série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980). 95

4.18 (a) FAC e (b) FACP amostrais para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980). 96

4.19 (a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo AR(4) para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980). 97

4.20 (a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo AR-MV(4) para a série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980). 99

4.21 (a) Histograma e (b) gráfico Q-Q para a série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980). 100

4.22 Valores observados e previsões pontuais dos modelos AR(4) e AR-MV(4) da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA, para o período de 1981 a 1985. 102

4.23 (a) Histograma e (b) gráfico Q-Q para a série temporal gerada para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.). 103

4.24 Série temporal gerada para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.). 104

4.25 (a) FAC e (b) FACP amostrais para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.). 105

4.26 (a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo AR(5) para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.). 106

4.27 (a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo AR-MV(5) para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.) 107

4.28 (a) Histograma e (b) gráfico Q-Q para a série temporal de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.). 108

4.29 Valores observados e previsões pontuais dos modelos AR(5) e AR-MV(5) da série temporal padronizada de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA, para o período de 615 a.C. a 611 a.C. 110

Lista de Tabelas

2.1	Comportamento geral das FAC e FACP dos processos $AR(p)$, $MA(q)$ e $ARMA(p, q)$.	18
3.1	Valores estimados dos parâmetros dos modelos ajustados para os dados simulados do Exemplo 3.9. (O valor entre parênteses indica o erro-padrão de cada estimativa).	47
3.2	Estimativas do MAE, MQE e os erros de previsão a cada passo, sob os modelos $AR(4)$ (com e sem intercepto) e $AR-MV(4)$ para os dados simulados do Exemplo 3.9.	52
3.3	Valores estimados dos parâmetros dos modelos ajustados para os dados simulados do Exemplo 3.10. (O valor entre parênteses indica o erro-padrão de cada estimativa).	54
3.4	Estimativas do MAE, MQE e os erros de previsão a cada passo, sob os modelos $AR(5)$ (com e sem intercepto) e $AR-MV(5)$ para os dados simulados no Exemplo 3.10.	58
3.5	Estimativas das estatísticas descritivas dos estimadores do modelo do Exemplo 3.11, para os casos α conhecido e desconhecido para $n=500, 1000$ e 5000 .	61
3.6	Estimativas do MAE, MQE e a média dos quadrados dos erros a cada passo para o Exemplo 3.11.	62
3.7	Estimativas das estatísticas descritivas dos estimadores do modelo do Exemplo 3.12, para os casos α conhecido e desconhecido para $n=500, 1000$ e 5000 .	67
3.8	Estimativas do MAE, MQE e a média dos quadrados dos erros a cada passo para o Exemplo 3.12.	69
4.1	Testes para a detecção de normalidade na série temporal das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).	77
4.2	Testes para a detecção de estacionariedade na série de medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).	78
4.3	Estimativas dos parâmetros dos modelos $AR(2)$ e $AR(9)$ das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).	79

4.4	Estatísticas qui-quadrado do teste de Box-Pierce modificado para os <i>lags</i> 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo AR(2) da série temporal padronizada das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).	80
4.5	Estatísticas qui-quadrado do teste de Box-Pierce modificado para os <i>lags</i> 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo AR-MV(2) da série temporal padronizada das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).	82
4.6	Testes para a detecção de normalidade na série temporal das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).	82
4.7	Estimativas dos parâmetros e critérios de comparação, sob os modelos AR(2) e AR-MV(2), para as medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960). (O valor entre parênteses indica o erro-padrão de cada estimativa).	83
4.8	Estimativas do MAE, MQE e os erros de previsão a cada passo, sob os modelos AR(2) e AR-MV(2), para o período de 1961 a 1965, da série temporal padronizada das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA.	84
4.9	Testes para a detecção de normalidade na série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).	85
4.10	Testes para a detecção de estacionariedade na série de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).	86
4.11	Estimativas dos parâmetros dos modelos AR(3) e AR(13) para as medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).	87
4.12	Estatísticas qui-quadrado do teste de Box-Pierce modificado para os <i>lags</i> 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo AR(3) da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).	89
4.13	Estatísticas qui-quadrado do teste de Box-Pierce modificado para os <i>lags</i> 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo AR-MV(3) da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).	90
4.14	Testes para a detecção de normalidade na série temporal de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).	91

4.15	Estimativas dos parâmetros e critérios de comparação, sob os modelos AR(3) e AR-MV(3), para a série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978). (O valor entre parênteses indica o erro-padrão de cada estimativa).	92
4.16	Estimativas do MAE, MQE e os erros de previsão a cada passo, sob os modelos AR(3) e AR-MV(3), para o período de 1979 a 1983, da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA.	92
4.17	Testes para a detecção de normalidade na série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).	94
4.18	Testes para a detecção de estacionariedade para a série de medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).	94
4.19	Estimativas dos parâmetros dos modelos AR(4), AR(6) e AR(24) para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).	96
4.20	Estimativas dos parâmetros do modelo AR(6), sem os coeficientes ϕ_3 e ϕ_4 , para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).	96
4.21	Crítérios de comparação entre os modelos AR(4) e AR(6) para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).	96
4.22	Estatísticas qui-quadrado do teste de Box-Pierce modificado para os <i>lags</i> 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo AR(4) da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).	98
4.23	Estatísticas qui-quadrado do teste de Box-Pierce modificado para os <i>lags</i> 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo AR-MV(4) da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).	99
4.24	Testes para a detecção de normalidade na série temporal das medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).	100
4.25	Estimativas dos parâmetros e critérios de comparação, sob os modelos AR(4) e AR-MV(4), para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980). (O valor entre parênteses indica o erro-padrão de cada estimativa).	101

4.26	Estimativas do MAE, MQE e os erros de previsão a cada passo, sob os modelos AR(4) e AR-MV(4), para o período de 1981 a 1985, da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA.	101
4.27	Testes para a detecção de normalidade na série temporal gerada para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).	103
4.28	Testes para a detecção de estacionariedade na série de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).	104
4.29	Estimativas dos parâmetros do modelo AR(5), sem os coeficientes ϕ_2 , ϕ_3 e ϕ_4 , para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).	104
4.30	Estatísticas qui-quadrado do teste de Box-Pierce modificado para os <i>lags</i> 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo AR(5) da série temporal padronizada de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).	105
4.31	Estatísticas qui-quadrado do teste de Box-Pierce modificado para os <i>lags</i> 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo AR-MV(5) da série temporal padronizada de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).	107
4.32	Testes para a detecção de normalidade na série temporal de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).	108
4.33	Estimativas dos parâmetros e critérios de comparação, sob os modelos AR(5) e AR-MV(5), para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.). (O valor entre parênteses indica o erro-padrão de cada estimativa).	109
4.34	Estimativas do MAE, MQE e os erros de previsão a cada passo, sob os modelos AR(5) e AR-MV(5), para o período de 615 a.C. a 611 a.C., da série temporal padronizada de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA.	109

Capítulo 1

Introdução

Um dos principais objetivos da análise de séries temporais é encontrar um modelo que se ajuste adequadamente a um hipotético conjunto de dados, o qual seja objeto de investigação. Atualmente, dispomos de uma volumosa quantidade de modelos formulados, com o intuito de atender às mais diversas especificidades que se possam encontrar num conjunto de observações ordenado no tempo.

Nesse contexto, existe a consolidada metodologia de Box e Jenkins (1976), cujo objetivo é ajustar modelos autorregressivos integrados de médias móveis (ARIMA) a séries temporais lineares. Os modelos ARIMA possuem alguns casos particulares. Contudo, neste trabalho, falaremos apenas do modelo autorregressivo de médias móveis (ARMA), do modelo média móvel (MA) e, sobretudo, do modelo autorregressivo (AR).

Sob outra perspectiva, existem os modelos de séries temporais não lineares, que começaram a ser investigados sistematicamente e com maior profundidade a partir do ano de 1977, perante diversas tentativas mal sucedidas de se analisar um conjunto de dados clássico sobre a quantidade de lincas capturados na região noroeste do Canadá, no período de 1821 a 1934, que vinha sendo examinado por diversos pesquisadores, durante muito tempo, por meio de modelos apropriados a séries temporais lineares (Tong, 2007).

Nesse contexto, surgiu o modelo autorregressivo com limiar (TAR, do inglês *threshold autoregressive model*), que foi mencionado pela primeira vez em Tong (1977) e, cuja ideia geral, é a segmentação de modelos não lineares em partes lineares, por meio de uma variável politômica, I_t , denominada variável limiar.

O modelo TAR possui um caso particular de grande importância, denominado modelo autorregressivo com limiar autoexcitante (SETAR, do inglês *self-exciting threshold autoregressive model*) que possui como característica o fato de que a variável limiar I_t é função de uma defasagem dela própria.

Fadel (2012) introduziu então, o modelo autorregressivo com memória variável (AR-MV), proveniente do modelo SETAR, com a restrição de que alguns parâmetros não se modificam, o parâmetro de defasagem é sempre igual a 1 e os erros são iguais. Em seu trabalho, a autora adaptou a teoria existente para os modelos SETAR ao caso particular

AR-MV, desenvolvido por ela, discutindo sobre as condições para a ergodicidade geométrica do modelo e sobre a estimação dos parâmetros; realizando estudos simulados para o modelo AR-MV; construindo intervalos de confiança para as estimativas; e realizando previsões. Além disso, a referida autora desenvolveu estudos simulados para os modelos AR-MV(2) e AR-MV(3) e os comparou com modelos AR.

Veremos mais à frente que a principal etapa no processo de identificação de um modelo é a análise das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) estimadas para os dados e, por meio dos estudos realizados por Fadel (2012), percebemos que os comportamentos apresentados pelas FAC e FACP dos modelos AR-MV(2) e AR-MV(3) são similares aos padrões conhecidos dessas funções para os modelos AR(2) e AR(3), respectivamente. Dessa forma, decidimos investigar essa propriedade em modelos AR-MV de ordens superiores.

Analisaremos neste trabalho quatro conjuntos de dados dendrocronológicos reais, que nada mais são do que medições de anéis de crescimento de árvores, para fins de estimação da idade desses vegetais. As árvores que serão objeto de estudo neste trabalho podem viver durante milhares de anos e, intuitivamente, é razoável pensarmos que, durante todo esse período, tenha ocorrido algum fenômeno climático de grande vulto, o que poderia ocasionar uma estagnação no crescimento ou um crescimento acelerado das árvores, ou seja, uma mudança abrupta na evolução dos dados, sugerindo um processo não linear. Por outro lado, os gráficos da FAC e FACP, para esses quatro conjuntos de dados possuem comportamentos similares aos gráficos dessas mesmas funções para os modelos AR, que é uma classe de modelos apropriada para a análise de séries temporais lineares. Nesse sentido, resolvemos comparar as classes AR contra AR-MV e investigar o potencial desta classe não linear para a modelagem desses dados, embora tais modelos tenham naturezas distintas.

Neste trabalho, avaliamos ainda o comportamento dos modelos AR-MV, a fim de observar graficamente a propriedade de ergodicidade geométrica desses modelos, provada por Fadel (2012). Conforme já havia sido demonstrado no trabalho dessa autora, se o somatório dos parâmetros autorregressivos do modelo é maior ou igual a 1 ($\sum_{i=1}^p |\phi_i| \geq 1$), a série temporal gerada não é estacionária, o que foi mostrado também para os AR-MV de quarta e quinta ordens (Seção 3.2).

Simulamos um caso particular dos modelos AR-MV(4) e AR-MV(5), verificando as qualidades das estimativas geradas pelos modelos para dois tamanhos de amostra ($n = 500$ e $n = 10000$). Os resíduos desses modelos foram analisados, de forma a conferir a adequabilidade aos dados. Ademais, as capacidades preditivas e a ajustabilidade dos modelos foram testadas em comparação a modelos AR(4) e AR(5). Concluímos que, para ambos os casos, os modelos AR-MV adequaram-se melhor aos dados quando comparados aos modelos AR, independentemente do tamanho da amostra. Entretanto, em relação à previsão, constatamos a importância do tamanho da amostra, pois, no caso onde $n = 500$, o modelo AR sem intercepto mostrou uma capacidade preditiva mais apropriada. Todavia, quando $n = 10000$, o AR-MV foi preferível (Seção 3.5).

Examinamos as propriedades assintóticas dos estimadores dos modelos AR-MV(4) e AR-MV(5), por meio de estudo simulado, para três tamanhos de amostra distintos. Ao final do estudo pudemos concluir que o fato de α (denominado parâmetro limiar) ser conhecido é muito relevante no processo de estimação dos parâmetros autorregressivos (ϕ), pois quando o valor verdadeiro do α é desconhecido, não observamos a normalidade nas estimativas de alguns dos parâmetros autorregressivos, especialmente, nos parâmetros ϕ de maior ordem. Apesar disso, o aumento do tamanho da amostra mostrou favorecer as estimativas dos parâmetros dos modelos (Seção 3.6).

Comparamos o desempenho dos modelos AR-MV contra modelos AR em quatro conjuntos de dados reais. Após a análise dessas aplicações pudemos concluir que, em relação ao ajuste dos modelos, houve um empate, pois, nas aplicações I e II, o AR foi melhor, mas, nas aplicações III e IV, o AR-MV foi preferível. É importante ressaltar que o AR-MV obteve um melhor ajuste nos casos onde os dados exigiam modelos de maior ordem autorregressiva, ou seja, quarta e quinta. No que diz respeito às previsões, também houve um emparelhamento dos modelos. Nas aplicações I e IV os métodos referentes ao AR-MV mostraram uma capacidade preditiva mais adequada, enquanto que nas aplicações II e III, o AR foi superior nesse quesito (Capítulo 4).

Os seguintes trabalhos, posteriores à dissertação de Fadel (2012), trouxeram algumas extensões ao assunto “memória variável”. Barbosa (2015), em seu trabalho de iniciação científica, estendeu a abordagem feita em Fadel (2012), para o modelo AR-MV(p). Nesse modelo, o estado atual da série depende dos estados anteriores, mas a memória do processo depende, através de um valor α , do estado atual do processo. Ou seja, quanto maior o valor atual do processo, maior a ordem do modelo autorregressivo. O trabalho de Barbosa (2015) teve como objetivo generalizar a condição para permitir que o limiar entre as ordens do processo seja dado não por valores constantes, mas por uma família de funções suaves obtidas como expansões em bases de *splines*. Pinto (2016), em sua tese, apresentou uma nova classe de modelos, os processos de saltos com memória de alcance variável. O autor desenvolveu um novo estimador para a árvore de contexto imersa nesse processo, levando em conta mais parâmetros fornecidos pela amostra de observações disponível. Pinto (2016) obteve ainda uma cota superior da taxa de convergência da árvore estimada para a árvore real, provando a convergência quase certa do estimador.

Como contribuição, neste trabalho generalizamos os estudos simulados desenvolvidos por Fadel (2012), restringidos aos modelos AR-MV(2) e AR-MV(3), para os modelos AR-MV(4) e AR-MV(5) e, com o intuito de averiguar a utilidade prática do modelo AR-MV e os benefícios na seleção desse modelo em relação ao modelo AR, foram realizadas quatro aplicações a dados reais. Cabe ressaltar que todo o estudo computacional deste trabalho foi realizado com o auxílio do *software* R, versão 3.2.5, disponível em R Core Team (2017).

Em suma, os objetivos desta dissertação são desenvolver estudos simulados para os modelos AR-MV(4) e AR-MV(5), no intuito principal de investigar se a propriedade de semelhança entre as FAC e FACP dos modelos AR-MV e AR são mantidas quando estendemos

os modelos para ordens superiores; e realizar aplicações com os modelos AR-MV de ordens 2, 3, 4 e 5, utilizando conjuntos de dados reais, comparando assim seus desempenhos, com os de modelos AR de ordens idênticas.

O presente trabalho está dividido da seguinte forma. No Capítulo 2, realizamos uma revisão de literatura. No Capítulo 3, fazemos uma análise detalhada dos modelos AR-MV. No Capítulo 4, apresentamos aplicações dos modelos AR-MV a quatro conjuntos de dados reais. O Capítulo 5 é dedicado às considerações finais.

Capítulo 2

Revisão de literatura

Neste capítulo faremos uma breve revisão de conceitos, necessários para uma melhor compreensão deste trabalho. Na Seção 2.1 apresentamos os conceitos de processos estocásticos e séries temporais. Na Seção 2.2 trazemos os principais aspectos da abordagem de Box e Jenkins. Na Seção 2.3 explicamos as diferenças entre séries temporais lineares e não lineares. Na Seção 2.4 abordamos os modelos autorregressivos com limiar.

2.1 Processos estocásticos e séries temporais

Um processo estocástico pode ser entendido como um fenômeno que progride sequencialmente no tempo a partir de leis probabilísticas (Box e Jenkins, 1976).

A definição para processos estocásticos, a seguir, é dada por Hoel *et al.* (1972).

Definição 2.1. *De forma geral, um processo estocástico é definido como uma coleção de variáveis aleatórias $Z_t, t \in T$, definida em um mesmo espaço de probabilidades, onde T é um subconjunto de $(-\infty, \infty)$ e comumente é entendido como o conjunto de parâmetros relativos ao tempo.*

Suponha um processo estocástico como sendo uma família $\mathbf{Z} = \{Z_t, t \in T\}$ tal que, para cada t , Z_t é uma variável aleatória. O conjunto T normalmente é tomado como o conjunto dos inteiros ($\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$) ou o conjunto dos reais (\mathbb{R}) (Morettin e Tolo, 2006, p.19).

Admitindo que Z_t tenha origem em um experimento que seja repetido sob condições idênticas, a cada experimento obtemos um registro dos valores de Z_t ao longo do tempo. Cada registro particular é uma realização do processo estocástico e uma série temporal é uma realização amostral do processo estocástico, isto é, é uma amostra finita do conjunto de todas as trajetórias possíveis que podem ser geradas pelo processo estocástico.

Uma série temporal então pode ser entendida como um conjunto de observações ordenadas no tempo. As séries temporais podem ser classificadas em contínuas e discretas, caso o conjunto de dados seja contínuo ou discreto, respectivamente. Na prática, para que seja possível analisar uma série contínua, é necessário discretizá-la, ou seja, convertê-la em uma série discreta fracionada em intervalos de tempo iguais.

De acordo com Morettin e Tolo (2006), os propósitos da análise de séries temporais podem ser: (i) averiguar de que maneira se dá a geração da série temporal; (ii) realizar predições para futuras observações; (iii) entender o comportamento da série temporal; e (iv) buscar periodicidades expressivas nas observações.

Por exemplo, uma série temporal com m observações sucessivas pode ser suposta como uma realização amostral entre todas as m -uplas que poderiam ser geradas por um mesmo processo gerador dos dados ou processo estocástico.

Conforme Morettin e Tolo (2006), um processo estocástico está determinado quando conhecemos as distribuições finito-dimensionais dadas por

$$F(z_1, \dots, z_n; t_1, \dots, t_n) = P\{Z_{t_1} \leq z_1, \dots, Z_{t_n} \leq z_n\}, \quad n \geq 1, \quad (2.1)$$

para $z_1, \dots, z_n \in \mathbb{R}$ e $t_1, \dots, t_n \in T$.

Dessa forma, para cada escolha finita de índices $\{t_1, \dots, t_n\} \subset T$, o vetor aleatório $(X_{t_1}, \dots, X_{t_n})$ tem uma certa função de distribuição acumulada (fda) conjunta. As “distribuições finito-dimensionais” referem-se à coleção de todas as fda obtidas ao percorrermos todas as possíveis escolhas de subconjuntos de T com tamanho n , para $n = 1, 2, \dots$.

Todavia, na prática, essas funções de distribuição não são conhecidas e dispomos de apenas uma amostra do processo estocástico (a série temporal observada). Dessa maneira, se for admissível supor a estacionariedade e a ergodicidade do processo estocástico, é possível analisar algumas propriedades da série temporal de maneira adequada. Além disso, a confirmação dessas propriedades para a série analisada permitirá proceder inferências estatísticas sobre os parâmetros estimados (Bueno, 2011).

2.1.1 Processos estocásticos estacionários

Genericamente, um processo estocástico é estacionário se os seus aspectos estatísticos/probabilísticos permanecem invariantes ao longo do tempo. A seguir, serão definidos os conceitos de processos estritamente e fracamente estacionários, utilizando como referência Morettin e Tolo (2006).

Definição 2.2. *Um processo estocástico $\mathbf{Z} = \{Z_t, t \in T\}$ é dito estacionário em sentido estrito, se todas as distribuições de probabilidade, equação (2.1), mantêm-se idênticas sob translações no tempo, ou seja,*

$$F(z_1, \dots, z_n; t_1 + k, \dots, t_n + k) = F(z_1, \dots, z_n; t_1, \dots, t_n), \quad (2.2)$$

para quaisquer $t_1, \dots, t_n, k \in T$.

A estacionariedade estrita implica que as variáveis aleatórias Z_t e Z_{t+k} possuem as mesmas distribuições de probabilidade qualquer que seja k . Portanto, se estiverem bem definidas,

a média e a variância do processo serão constantes, ou seja,

$$\mu_t = \mu, \forall t; \quad (2.3)$$

$$\text{Var}_t = \sigma^2, \forall t. \quad (2.4)$$

Definição 2.3. Um processo estocástico $\mathbf{Z} = \{Z_t, t \in T\}$ é dito fracamente estacionário ou estacionário de segunda ordem se, e somente se,

$$(i) E(Z_t) = \mu_t = \mu, \text{ constante}, \forall t \in T;$$

$$(ii) E(Z_t^2) < \infty, \forall t \in T;$$

$$(iii) \text{Cov}(Z_t, Z_{t+k}) = E\{(Z_t - \mu)(Z_{t+k} - \mu)\} = \gamma_k, \text{ é uma função de } |k| \text{ apenas.}$$

Assim, a estacionariedade fraca ou de segunda ordem é um processo que possui média e variância constantes, o segundo momento finito e as suas autocovariâncias dependendo apenas do intervalo de tempo (*lag*) k entre as observações.

Além disso, se o processo estocástico for Gaussiano (ou seja, quando as distribuições finito-dimensionais (equação (2.1)) são todas normais multivariadas) e estacionário de segunda ordem, ele será estritamente estacionário, pois a distribuição normal é determinada unicamente em termos do primeiro e do segundo momentos (Morettin e Tolo, 2006).

2.1.2 Processos estocásticos ergódicos

Intuitivamente, um processo estocástico é ergódico quando suas estatísticas podem ser estimadas com base em uma realização suficientemente grande da trajetória desse processo. Todo processo ergódico também é estacionário, pois uma realização de um processo não estacionário não poderá conter todas as informações necessárias para a especificação do processo.

No que segue, tomaremos como definição de ergodicidade o seguinte teorema de Parzen (1958)

Teorema 2.1. Sejam Z_t variáveis aleatórias definidas para t em $T = \{0, \pm 1, 2, \dots\}$. Suponha a existência de uma função característica $\varphi(u_1, \dots, u_k)$, de modo que, para todo u_1, \dots, u_k ,

$$\varphi(u_1, \dots, u_k) = E[\exp i\{u_1 Z_{t_1+h} + \dots + u_k Z_{t_k+h}\}], \quad (2.5)$$

$\forall h$ em T .

Suponha também que, para cada $\tau \in T$, exista uma função característica $\varphi(u_1, \dots, u_k; \tau)$, tal que,

$$\varphi(u_1, \dots, u_k; \tau) = E[\exp i\{u_1(Z_{t_1+h} - Z_{t_1+h+\tau}) + \dots + u_k(Z_{t_k+h} - Z_{t_k+h+\tau})\}], \quad (2.6)$$

$\forall h$ em T .

Assim, um processo estocástico $(Z_t : t \in \mathbb{Z})$ é dito ergódico se, e somente se, existirem funções (2.5) e (2.6) tais que vale (2.7)

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{\tau=0}^n \varphi(u_1, \dots, u_k; \tau) = |\varphi(u_1, \dots, u_k)|^2, \quad (2.7)$$

para qualquer escolha de k pontos t_1, \dots, t_k , $k \in \mathbb{Z}$, e para quaisquer números reais u_1, \dots, u_k .

Além disso, de acordo com Bueno (2011), um processo fracamente estacionário é ergódico para o primeiro momento quando a média amostral convergir em probabilidade para a média populacional, ou seja,

$$\bar{Z} = \frac{1}{n} \sum_{t=1}^n Z_t \xrightarrow{P} \mu. \quad (2.8)$$

Ainda segundo Bueno (2011), um processo fracamente estacionário é ergódico para o segundo momento quando a função de autocovariância amostral convergir em probabilidade para a função de autocovariância populacional, ou seja,

$$\frac{1}{n-k} \sum_{t=1}^n \{(Z_t - \mu)(Z_{t+k} - \mu)\} \xrightarrow{P} \gamma_k, \text{ para todo } k. \quad (2.9)$$

2.1.3 Ruído branco

A sequência $\{\varepsilon_t, t \in \mathbb{Z}\}$ é um ruído branco discreto se

$$E(\varepsilon_t) = \mu_\varepsilon, \forall t; \quad (2.10)$$

$$E(\varepsilon_t^2) = \sigma_\varepsilon^2, \forall t; \quad (2.11)$$

$$\text{Cov}(\varepsilon_t, \varepsilon_{t+k}) = 0, k \neq 0. \quad (2.12)$$

Sem perda de generalidade, $\mu_\varepsilon = 0$. Assim, é convencional representar um processo ruído branco da seguinte maneira

$$\varepsilon_t \sim \text{RB}(0, \sigma_\varepsilon^2). \quad (2.13)$$

2.2 Abordagem de Box e Jenkins

A abordagem de Box e Jenkins (1976), também conhecida como metodologia ARIMA, é uma classe de modelos lineares que visa ajustar modelos autorregressivos integrados de médias móveis, ARIMA(p, d, q), a um determinado conjunto de observações.

Conforme Morettin e Tolo (2006), a estruturação do modelo é baseada em um ciclo iterativo composto pelas seguintes etapas:

- (a) *Especificação* - Esta fase consiste na triagem de uma classe geral de modelos para análise;
- (b) *Identificação* - Esta fase consiste na seleção de um modelo adequado, levando em consideração, principalmente, a análise das funções de autocorrelações e autocorrelações parciais dos dados;
- (c) *Estimação* - Esta fase consiste na estimação dos parâmetros do modelo selecionado;
- (d) *Diagnóstico* ou *Verificação* - Esta fase consiste em analisar os resíduos do modelo ajustado para verificação da sua adequabilidade.

2.2.1 Modelos ARMA

Seguindo a metodologia criada por Box e Jenkins (1976), podemos escrever a seguinte definição.

Definição 2.4. Um processo estacionário $\mathbf{Z} = \{Z_t, t \in T\}$ é dito um processo ARMA(p, q), $p, q \in \mathbb{N}$ se $\forall t \in \mathbb{Z}$

$$Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \cdots - \phi_p Z_{t-p} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}, \quad (2.14)$$

onde $a_t \sim RB(0, \sigma^2)$.

Além do mais, a seguinte definição é extraída de Brockwell e Davis (1996)

Definição 2.5. O processo $\{Z_t\}$ é dito ser um ARMA(p, q) com média μ se $\{Z_t - \mu\}$ é um processo ARMA(p, q).

As definições de causalidade e invertibilidade de um processo ARMA(p, q) são dadas a seguir, de acordo com Brockwell e Davis (1996)

Definição 2.6. Um processo ARMA(p, q) $\{Z_t\}$ é causal ou uma função causal de $\{a_t\}$ se houver constantes $\{\psi_j\}$ tal que $\sum_{j=0}^{\infty} |\psi_j| < \infty$ e

$$Z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}, \forall t. \quad (2.15)$$

Definição 2.7. Um processo ARMA(p, q) $\{Z_t\}$ é invertível se houver constantes $\{\pi_j\}$ tal que $\sum_{j=0}^{\infty} |\pi_j| < \infty$ e

$$a_t = \sum_{j=0}^{\infty} \pi_j Z_{t-j}, \forall t. \quad (2.16)$$

A equação (2.14) pode ser representada por

$$\Phi(B)Z_t = \Theta(B)a_t, \quad (2.17)$$

onde $\Phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$, denominado operador autorregressivo, e $\Theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$, denominado operador médias móveis, denotam os parâmetros do modelo; B representa o operador retroativo, tal que, $B^k Z_t = Z_{t-k}$, $k \in \mathbb{Z}$; e a_t é um ruído branco. Portanto, podemos escrever

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)Z_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)a_t. \quad (2.18)$$

Na equação (2.18), as defasagens da variável Z_t no polinômio do lado esquerdo representam a parte autorregressiva (AR) do modelo e as defasagens dos choques aleatórios a_t no polinômio do lado direito representam a parte média móvel (MA).

A equação representa uma ampla classe de modelos denominados ARMA(p, q), onde p representa a ordem de defasagem do termo autorregressivo e q a ordem de defasagem do termo média móvel.

Exemplo 2.1. Fazendo $p = 1$ e $q = 0$ temos o modelo AR(1), denominado modelo autorregressivo de primeira ordem, no qual o valor da série no instante t depende somente do valor da série no instante $t - 1$ e da inovação a_t

$$Z_t - \phi_1 Z_{t-1} = a_t \iff Z_t - \phi_1 B Z_t = a_t \iff (1 - \phi_1 B)Z_t = a_t. \quad (2.19)$$

Exemplo 2.2. Fazendo $p = 0$ e $q = 1$ temos o modelo MA(1), denominado modelo média móvel de primeira ordem, no qual o valor da série no instante t é a combinação linear dos ruídos brancos em t e $t - 1$

$$Z_t = (1 - \theta_1 B)a_t. \quad (2.20)$$

Exemplo 2.3. Fazendo $p = 1$ e $q = 1$ obtemos o modelo ARMA(1,1), denominado modelo autorregressivo média móvel de primeira ordem, cuja equação é apresentada a seguir

$$(1 - \phi_1 B)Z_t = (1 - \theta_1 B)a_t. \quad (2.21)$$

Os modelos supracitados são apenas exemplos da ampla família de modelos ARMA(p, q).

Modelos ARIMA

Há situações em que a série temporal em estudo apresenta uma componente de tendência, dessa maneira, o processo estocástico gerador da série é não estacionário. Neste caso a série deve passar por d diferenças simples para que se torne estacionária, condição básica para a aplicação da metodologia de Box e Jenkins (1976). Por exemplo, para remover uma tendência linear basta tomar a primeira diferença da série ($d = 1$)

$$\Delta Z_t = Z_t - Z_{t-1}. \quad (2.22)$$

Em algumas situações pode ser necessário tomar a segunda diferença ($d = 2$)

$$\Delta^2 Z_t = Z_t - 2Z_{t-1} + Z_{t-2}. \quad (2.23)$$

Neste caso, a metodologia de Box e Jenkins (1976) é aplicada na série resultante das diferenças e o modelo é denominado autorregressivo integrado de médias móveis ou ARIMA(p, d, q), onde d representa o número de diferenças necessárias para tornar a série estacionária

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d Z_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t, \quad (2.24)$$

e o termo $(1 - B)^d$ caracteriza a parte integrada (I) do modelo.

Brockwell e Davis (1996) definem um processo ARIMA(p, d, q) da seguinte forma

Definição 2.8. *Se d é um inteiro não negativo, então $\{Z_t\}$ é um processo ARIMA(p, d, q) se $Y_t := (1 - B)^d Z_t$ é um processo ARMA(p, q) causal.*

2.2.2 Funções de autocorrelação e autocorrelação parcial

A seguir apresentamos os conceitos de funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) dos modelos ARIMA. Essas funções são úteis, especialmente, na identificação dos modelos a serem ajustados aos dados.

Função de autocorrelação de um AR(p)

Seja o modelo AR(p) dado por

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t,$$

onde $\tilde{Z}_t = Z_t - \mu$. Multiplicando em ambos lados da equação anterior o termo \tilde{Z}_{t-k} e tomando o valor esperado, temos

$$E[\tilde{Z}_t \tilde{Z}_{t-k}] = \phi_1 E[\tilde{Z}_{t-1} \tilde{Z}_{t-k}] + \phi_2 E[\tilde{Z}_{t-2} \tilde{Z}_{t-k}] + \dots + \phi_p E[\tilde{Z}_{t-p} \tilde{Z}_{t-k}] + E[\tilde{Z}_{t-k} a_t].$$

O termo $E[\tilde{Z}_{t-k} a_t] = 0$, para $k > 0$, porque os erros são contemporâneos à observações. Logo,

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}, \text{ para } k > 0,$$

que é a função de autocovariância (FACV). A função de autocorrelação (FAC) é então obtida ao dividirmos a equação anterior por $\gamma_0 = \text{Var}(Z_t)$. Assim,

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}, \text{ para } k > 0. \quad (2.25)$$

A FAC de um processo AR(p) apresenta decaimento exponencial a zero, à medida que k aumenta. Os valores dos *lags* podem ser positivos, quando as raízes da equação característica

$\Phi(B) = 0$ forem positivas; podem ser negativos, quando as raízes forem negativas; ou ainda apresentarem um padrão senoidal, no caso de raízes complexas. Para mais detalhes, ver Box e Jenkins (1976).

Fazendo $k = 1, 2, \dots, p$ em (2.25), e como $\rho_{-k} = \rho_k$, obtemos

$$\begin{cases} \rho_1 = \phi_1 + \phi_2\rho_1 + \dots + \phi_p\rho_{p-1}, \\ \rho_2 = \phi_1\rho_1 + \phi_2 + \dots + \phi_p\rho_{p-2}, \\ \vdots \\ \rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \dots + \phi_p, \end{cases}$$

denominadas equações de Yule-Walker. Em forma matricial,

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & & \ddots & \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}. \quad (2.26)$$

Dessa forma, é possível utilizar as equações de Yule-Walker para estimar os coeficientes ϕ_1, \dots, ϕ_p do modelo $AR(p)$ substituindo as FAC populacionais por suas estimativas amostrais (Morettin e Tolo, 2006).

Função de autocorrelação de um $MA(q)$

Seja o modelo $MA(q)$ dado por

$$Z_t = \mu + a_t - \theta_1 a_{t-1} + \theta_2 a_{t-2} - \dots - \theta_q a_{t-q},$$

e fazendo $\tilde{Z}_t = Z_t - \mu$, temos

$$\tilde{Z}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t.$$

A FACV é calculada da seguinte maneira

$$\gamma_k = E\{\tilde{Z}_t \tilde{Z}_{t-k}\} = E\left\{ \left[a_t - \sum_{j=1}^q \theta_j a_{t-j} \right] \left[a_{t-k} - \sum_{l=1}^q \theta_l a_{t-k-l} \right] \right\} \quad (2.27)$$

$$\begin{aligned} &= E(a_t a_{t-k}) - \sum_{j=1}^q \theta_j E(a_{t-k} a_{t-j}) - \sum_{l=1}^q \theta_l E(a_t a_{t-k-l}) \\ &+ \sum_{j=1}^q \sum_{l=1}^q \theta_j \theta_l E(a_{t-j} a_{t-k-l}). \end{aligned} \quad (2.28)$$

Lembramos que

$$\text{Cov}(a_t, a_{t-k}) = \gamma_a(k) = E(a_t a_{t-k}) = \begin{cases} \sigma_a^2, & k = 0, \\ 0, & k \neq 0. \end{cases} \quad (2.29)$$

A variância é então dada por

$$\gamma_0 = \text{Var}(Z_t) = \sigma_Z^2 = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_a^2. \quad (2.30)$$

Explicitando a FACV em termos de $\gamma_a(k)$ temos que

$$\gamma_k = \gamma_a(k) - \sum_{j=1}^q \theta_j \gamma_a(j-k) - \sum_{l=1}^q \theta_l \gamma_a(k+l) + \sum_{j=1}^q \sum_{l=1}^q \theta_j \theta_l \gamma_a(k+l-j),$$

que resulta em

$$\gamma_k = \left(-\theta_k + \sum_{l=1}^{q-k} \theta_l \theta_{k+l} \right) \sigma_a^2, \quad k = 0 \quad (2.31)$$

$$= (-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \cdots + \theta_q \theta_{q-k}) \sigma_a^2, \quad k = 1, \dots, q, \quad (2.32)$$

$$= 0, \quad k > q. \quad (2.33)$$

A função de autocorrelação (FAC) é então obtida ao dividirmos a equação (2.32) pela (2.30). Assim,

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \cdots + \theta_q \theta_{q-k}}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2}, & k = 1, \dots, q \\ 0, & k > q \end{cases} \quad (2.34)$$

Assim, a FAC de um processo MA(q) é igual a zero para *lags* maiores que a ordem q do processo média móvel.

Função de autocorrelação de um ARMA(p, q)

Seja o modelo ARMA(p, q) dado por

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \cdots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q},$$

onde $\tilde{Z}_t = Z_t - \mu$.

Multiplicando em ambos lados da equação anterior o termo \tilde{Z}_{t-k} e aplicando o valor esperado, temos

$$E[\tilde{Z}_t \tilde{Z}_{t-k}] = \phi_1 E[\tilde{Z}_{t-1} \tilde{Z}_{t-k}] + \cdots + \phi_p E[\tilde{Z}_{t-p} \tilde{Z}_{t-k}] + E[\tilde{Z}_{t-k} a_t] - \theta_1 E[a_{t-1} \tilde{Z}_{t-k}] - \cdots - \theta_q E[a_{t-q} \tilde{Z}_{t-k}].$$

Assim,

$$\gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_p \gamma_{k-p} + \gamma_{za}(k) - \theta_1 \gamma_{za}(k-1) - \cdots - \theta_q \gamma_{za}(k-q),$$

onde $\gamma_{za}(k) = E[\tilde{Z}_{t-k} a_t]$ é a covariância cruzada entre Z_t e a_t .

Como \tilde{Z}_{t-k} só depende dos erros a_t ocorridos até o instante $t - k$, temos

$$\gamma_{za}(k) \begin{cases} = 0, k > 0, \\ \neq 0, k \leq 0. \end{cases} \quad (2.35)$$

Então, a função de autocovariância pode ser escrita da seguinte forma

$$\gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_p \gamma_{k-p}, \quad k > q.$$

Portanto, a FAC é dada por

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, \quad \text{para } k > q. \quad (2.36)$$

A FAC de um processo ARMA(p, q) é infinita em extensão, decaindo para zero exponencialmente e/ou em forma senoidal após o *lag* $q - p$.

Função de autocorrelação parcial

A função de autocorrelação parcial (FACP) é definida apenas para modelos autorregressivos. Dessa forma, para um modelo média móvel, somente é possível a construção da FACP, caso o MA(q) seja invertível. Analogamente, para um modelo ARMA(p, q), é necessário que a parte média móvel do modelo seja invertível.

Seja ϕ_{kj} o j -ésimo coeficiente de um modelo AR(k), de maneira que ϕ_{kk} seja o último coeficiente autorregressivo. De acordo com (2.25), a FAC desse modelo é dada por

$$\rho_j = \phi_{k1} \rho_{j-1} + \phi_{k2} \rho_{j-2} + \cdots + \phi_{kk} \rho_{j-k}, \quad \text{para } j = 1, \dots, k,$$

de onde obtemos as equações de Yule-Walker a seguir

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}.$$

Resolvendo as equações continuamente para $k = 1, 2, 3, \dots$ chegamos a

$$\phi_{11} = \rho_1,$$

$$\phi_{22} = \frac{\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{bmatrix}}{\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}},$$

$$\phi_{33} = \frac{\begin{bmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{bmatrix}}{\begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix}},$$

e, de forma geral,

$$\phi_{kk} = \frac{\begin{bmatrix} \mathbf{P}_k^* \end{bmatrix}}{\begin{bmatrix} \mathbf{P}_k \end{bmatrix}}, \quad (2.37)$$

onde \mathbf{P}_k é a matriz de autocorrelações e \mathbf{P}_k^* é a matriz de autocorrelações com a última coluna substituída pelo vetor de autocorrelações.

Assim, o último coeficiente de um modelo $\text{AR}(k)$, ϕ_{kk} , é denominado função de autocorrelação parcial (FACP) (Morettin e Tolo, 2006).

A FACP de um processo $\text{AR}(p)$ apresenta $\phi_{kk} \neq 0$, para $k \leq p$ e $\phi_{kk} = 0$, para $k > p$.

2.2.3 Análise de estacionariedade

Como dito anteriormente, para se aplicar a metodologia de Box e Jenkins (1976), a série em estudo deve ser pelo menos estacionária de segunda ordem, ou seja, a série tem que ter média, variância e covariância finitas e constantes.

O exame da estacionariedade pode ser realizado observando a função de autocorrelação (FAC) amostral da série temporal investigada, de modo que o decrescimento rápido da FAC para zero é um comportamento típico de uma série estacionária. Caso essa propriedade não seja constatada é necessário tomar diferenças (d) na série, a fim de que a FAC exiba esse comportamento (Morettin e Tolo, 2006). A referida função é estimada por

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^n (Z_t - \bar{Z})(Z_{t-k} - \bar{Z})}{\sum_{t=1}^n (Z_t - \bar{Z})^2}, \quad (2.38)$$

onde n é o número de observações da série temporal investigada, \bar{Z} é a média da série temporal e k é a defasagem (*lag*) da autocorrelação.

Conforme (Brockwell e Davis, 1996, p. 94), para um tamanho de amostra n grande, os valores da FAC amostral para $k > q$, têm distribuição aproximadamente normal, com média zero e variância dada por

$$\text{Var}(\hat{\rho}_k) \approx \frac{1}{n} \left[1 + 2 \sum_{v=1}^q \rho_v^2 \right], \quad k > q. \quad (2.39)$$

Dessa forma, é possível construir um intervalo com 95% de confiança para $\hat{\rho}_k$, dado a seguir

$$\hat{\rho}_k \pm 1.96 \sqrt{\text{Var}(\hat{\rho}_k)}. \quad (2.40)$$

A condição de estacionariedade para um modelo autorregressivo é que tenha as raízes da equação característica $\Phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p) = 0$ fora do círculo unitário. Dessa forma, existem alguns testes que podem ser utilizados para detecção de raízes unitárias, os quais podem ser consultados no Apêndice A.

2.2.4 Identificação do modelo

A identificação é uma das fases mais complexas entre todos os estágios de estruturação de um modelo ARIMA. A seleção de modelos é baseada principalmente na análise das funções de autocorrelação (FAC) e funções de autocorrelação parcial (FACP) amostrais do conjunto de observações sob análise.

Nas Figuras 2.1, 2.2 e 2.3 expomos, por meio de simulações, as características teóricas das FAC e FACP para um processo AR(1), um MA(1) e um ARMA(1,1), respectivamente, com variações de sinais nos parâmetros ϕ e θ .

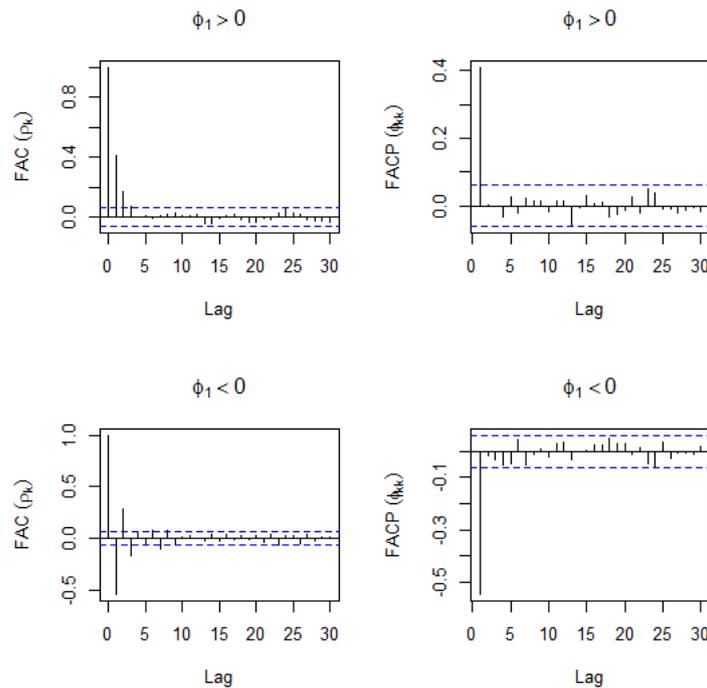


Figura 2.1: Simulações da FAC e da FACP de um processo AR(1), com parâmetros $\phi_1 = (0.5, -0.5)$ e $n = 1000$.

Comparando as características teóricas simuladas da FAC e da FACP, ilustradas nas Figuras 2.1, 2.2 e 2.3, com as estimativas destas funções, obtidas com base na série temporal investigada, podemos identificar o padrão teórico mais semelhante ao observado e assim determinar a ordem do modelo que melhor explica a dinâmica da série temporal investigada.

A Tabela 2.1, adaptada de Morettin e Toloi (2006) e de Bueno (2011), apresenta o padrão genérico dos modelos.

A FAC e a FACP têm comportamentos teóricos conhecidos, entretanto na prática, como

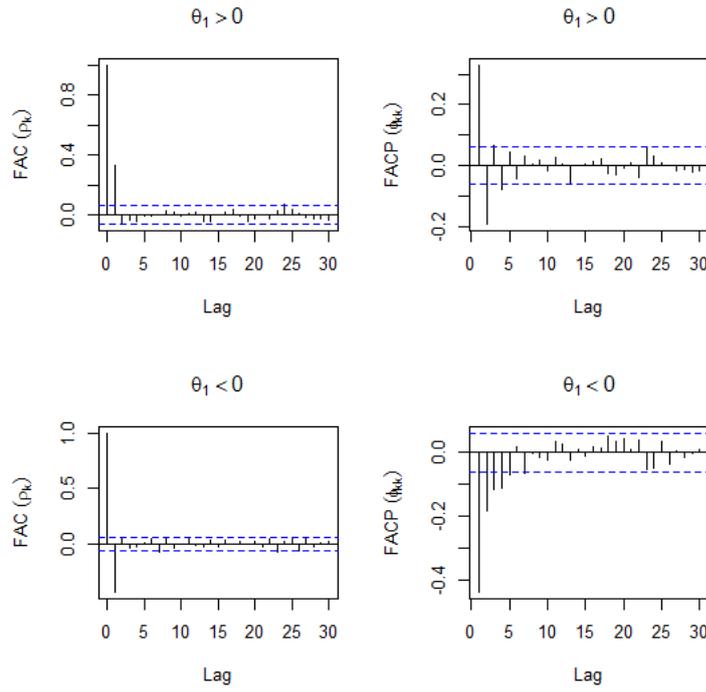


Figura 2.2: Simulações da FAC e da FACP de um processo $MA(1)$, com parâmetros $\theta_1 = (0.5, -0.5)$ e $n = 1000$.

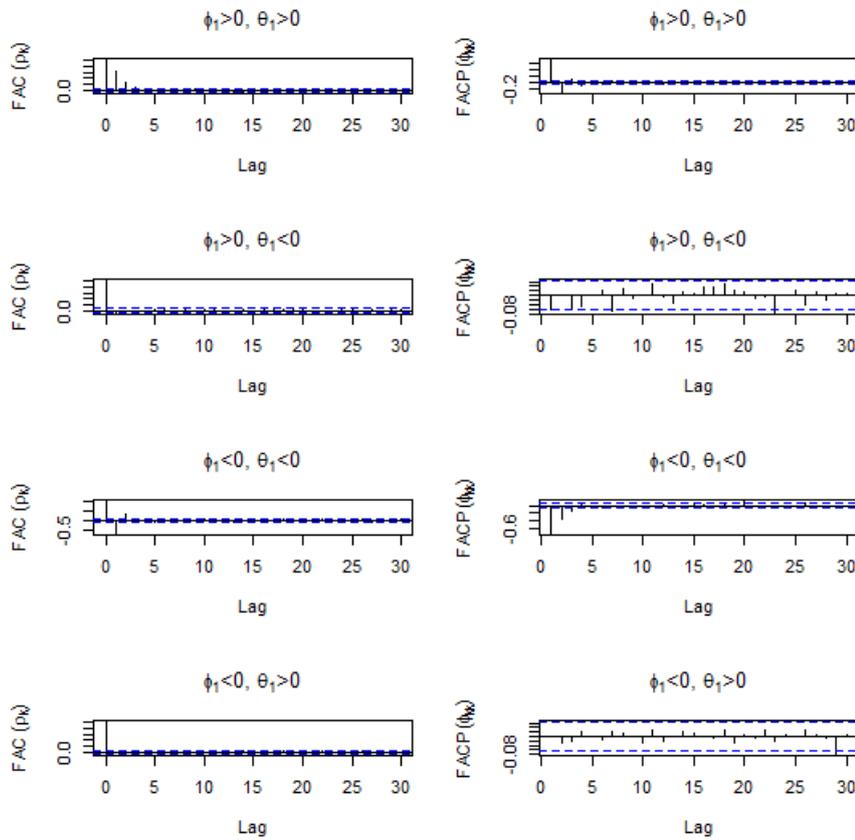


Figura 2.3: Simulações da FAC e da FACP de um processo $ARMA(1,1)$, com parâmetros $\phi_1 = (0.5, -0.5), \theta_1 = (0.5, -0.5)$ e $n = 1000$.

Tabela 2.1: *Comportamento geral das FAC e FACP dos processos AR(p), MA(q) e ARMA(p, q).*

Modelo	FAC (ρ_k)	FACP (ϕ_{kk})
AR(p)	decai exponencialmente e/ou em um padrão senoidal	significativa até o lag $\phi_{p,p}$, ou seja, truncada na defasagem p
MA(q)	significativa até o lag ρ_q , ou seja, truncada na defasagem q	decai exponencialmente e/ou em um padrão senoidal
ARMA(p, q)	decai exponencialmente e/ou em um padrão senoidal se $k > q$	decai exponencialmente e/ou em um padrão senoidal se $k > q$

estas funções são estimadas, a análise da FAC e da FACP amostrais pode ser uma tarefa complexa, o que dificulta a identificação da ordem do modelo.

Assim, para auxiliar na seleção do modelo, existem formas alternativas de identificação. Neste trabalho serão utilizados como métodos de seleção alternativos os critérios de informação de Akaike (AIC), proposto por Akaike (1973) e Akaike (1974), e Bayesiano (BIC), proposto por Schwarz (1978), cujas fórmulas são dadas por

$$\text{AIC} = -2\ln L(\hat{\vartheta}) + 2m, \quad (2.41)$$

$$\text{BIC} = -2\ln L(\hat{\vartheta}) + m\ln(n), \quad (2.42)$$

onde m é o número de parâmetros estimados, n é o número de observações da série temporal e $L(\hat{\vartheta})$ é a estimativa da função de verossimilhança do modelo.

Além desses critérios, a soma dos quadrados dos resíduos (SQR) e a variância residual dos modelos serão avaliadas. Para todos esses métodos de comparação, o melhor modelo será o que apresentar os menores valores nessas estatísticas.

2.2.5 Diagnóstico do modelo

A seleção do melhor modelo não deve se basear apenas nos critérios apresentados anteriormente. A análise dos resíduos de modelos concorrentes ajustados é de extrema importância na escolha final do modelo que melhor representa os dados.

Se os resíduos são autocorrelacionados, então a suposição de independência dos erros é violada. Dessa forma, devemos excluir do processo de escolha modelos com essa característica.

Para um modelo adequado deverão ser confirmados, na ocasião da análise dos resíduos, um comportamento aleatório, a inexistência de padrões identificáveis, bem como a ausência de observações discrepantes (*outliers*), o que pode ser feito por meio do gráfico de dispersão dos resíduos *versus* tempo. Deverá ser ratificada ainda a não existência de significância nas defasagens da FAC e da FACP dos resíduos.

O teste de Box-Pierce modificado, proposto por Box e Pierce (1970) e alterado por Ljung e Box (1978), é utilizado para verificar a existência de autocorrelação serial nos resí-

duos e, é representado formalmente por

$$Q(K) = n(n+2) \sum_{j=1}^K \frac{\hat{r}_j^2}{n-j}, \quad (2.43)$$

onde K é o *lag* analisado, n é o número de observações residuais e \hat{r}_j é a autocorrelação de ordem j dos resíduos do modelo estimado (e_t)

$$\hat{r}_j = \frac{\sum_{t=j+1}^n e_t e_{t-j}}{\sum_{t=1}^n e_t^2},$$

onde $e_t = \hat{\Theta}^{-1}(B)\hat{\Phi}^{-1}(B)\Delta^d Z_t$.

A estatística $Q(K)$ terá aproximadamente distribuição χ^2 com $K - p - q$ graus de liberdade. Este teste é utilizado para verificar se um conjunto de autocorrelações dos resíduos até a ordem K é (ou não) estatisticamente diferente de zero. Observamos que valores altos das autocorrelações dos resíduos implicam em valores altos de $Q(K)$. Por outro lado, em um processo ruído branco todas as autocorrelações são nulas e $Q(K)$ é nulo. As suposições acima permitem testar as seguintes hipóteses com base na estatística $Q(K)$

$$\left\{ \begin{array}{l} H_0 : \text{as } K \text{ primeiras autocorrelações são nulas (os resíduos são ruído branco);} \\ H_1 : \text{pelo menos uma autocorrelação, } \hat{r}_j, \text{ é estatisticamente diferente de zero (os} \\ \text{resíduos não são ruído branco).} \end{array} \right.$$

2.3 Séries Temporais Lineares e Não-Lineares

Em termos intuitivos, uma série temporal linear é aquela que pode ser expressa por um modelo linear, ou seja, que pode ser escrita por uma combinação linear de suas observações passadas adicionada a um termo de erro.

Adotaremos a seguinte definição de série temporal linear dada por Tsay (2005).

Definição 2.9. *Uma série temporal Z_t é dita linear se for possível escrevê-la da seguinte maneira*

$$\tilde{Z}_t = \sum_{i=0}^{\infty} \psi_i a_{t-i}, \quad (2.44)$$

onde $\tilde{Z}_t = Z_t - \mu$; ψ_i são os pesos de Z_t com $\psi_0 = 1$ e a_t é uma sequência de variáveis aleatórias iid que possuem uma distribuição de probabilidade bem definida. Além disso, supomos que a sequência a_t é contínua e que $E(a_t) = 0$ e $\text{Var}(a_t) = \sigma_a^2$, ou seja, a_t é ruído branco.

Salientamos que uma função f de \mathbb{R}^p em \mathbb{R} é dita linear se para vetores $x, y \in \mathbb{R}^p$ e qualquer escalar real α , $f(\alpha x + y) = \alpha f(x) + f(y)$.

Apesar de modelos lineares terem se mostrado adequados para descrever uma grande gama de fenômenos, existem séries temporais que apresentam, empiricamente, características que indicam violações de certas propriedades implicadas por modelos lineares, como

mudanças abruptas, variância condicional evoluindo no tempo (volatilidade) e irreversibilidade no tempo (Morettin e Toloi, 2006, p.37).

Morettin e Toloi (2006) mostram que a série da equação (2.44) pode ser escrita de maneira alternativa em função dos valores passados do processo $(\tilde{Z}_{t-1}, \tilde{Z}_{t-2}, \dots)$ adicionado a um ruído a_t

$$\tilde{Z}_t = \sum_{j=1}^{\infty} \pi_j \tilde{Z}_{t-j} + a_t, \quad (2.45)$$

onde $\pi_j, j = \{1, \dots, \infty\}$ é uma sequência de pesos.

Por exemplo, o modelo AR(1), dado por

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + a_t, \quad (2.46)$$

segue o padrão especificado em (2.45) e, por isso, é dito linear.

Graficamente também é possível avaliar a linearidade de uma série de tempo, observando o diagrama de dispersão da série Z_t contra as suas defasagens mais próximas $(Z_{t-1}, Z_{t-2}, \dots)$. Para ilustrar, suponha uma série temporal simulada de tamanho $n = 200$, gerada pelo modelo linear AR(1) com $\phi_1 = 0.9$ e $a_t \sim N(0, 4)$. Notamos, por meio da Figura 2.4, que há

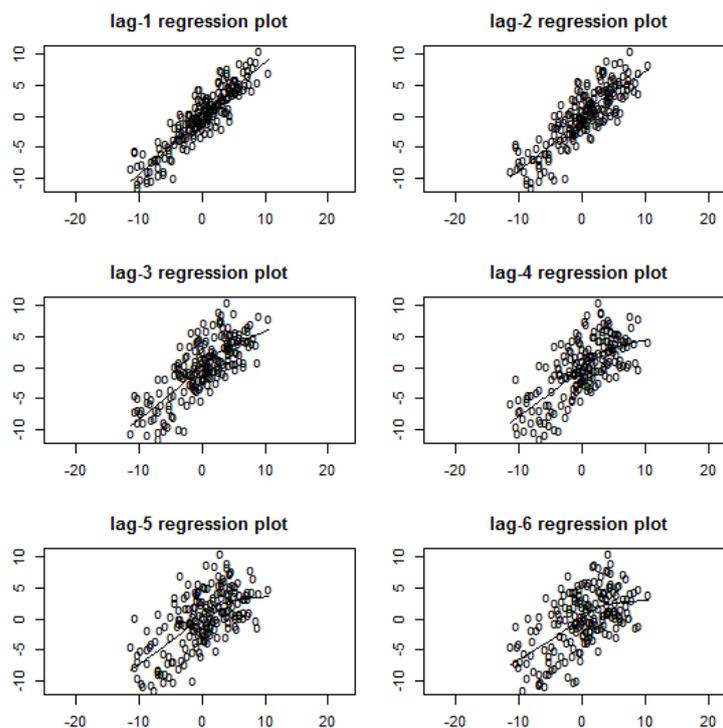


Figura 2.4: Diagramas de dispersão de Z_t contra os lags 1 a 6 de um processo AR(1) simulado com $\phi_1 = 0.9$ e $n = 200$.

um padrão linear crescente, que se desfaz lentamente à medida que as defasagens aumentam.

Por outro lado, admitindo um processo simulado de tamanho $n = 200$, gerado por um modelo não linear TAR (2,1,1), com $\phi_0^{(1)} = \phi_0^{(2)} = 0$, $\phi_1^{(1)} = 0.9$, $\phi_1^{(2)} = -1.8$ e os erros de cada regime com distribuição $N(0, 4)$. Temos os seguintes diagramas de dispersão, apresentados

na Figura 2.5. Percebemos que não há um padrão aparente nos diagramas e sim, nuvens de pontos espalhadas em cada um dos gráficos. Os modelos TAR são utilizados para modelagem de séries temporais não-lineares e serão vistos com detalhes na próxima seção.

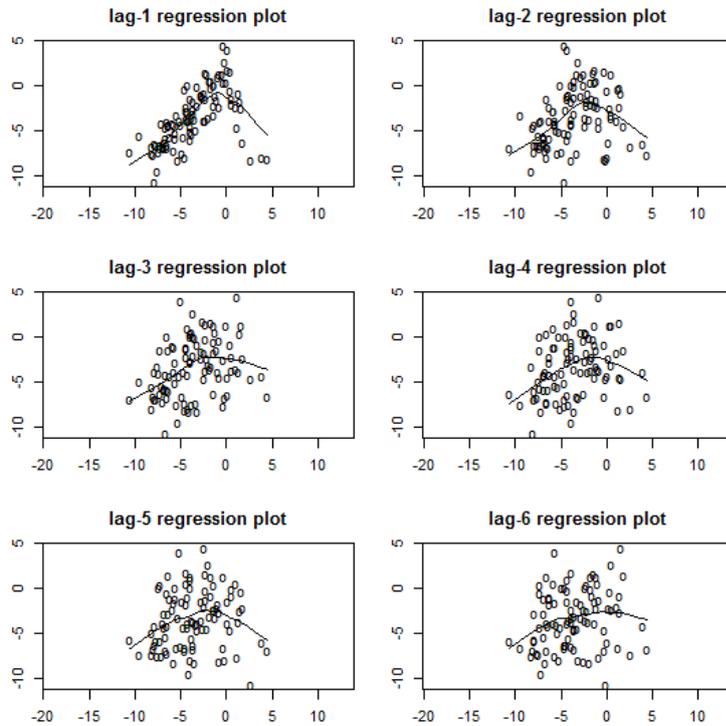


Figura 2.5: Diagramas de dispersão de Z_t contra os lags 1 a 6 de um processo TAR (2,1,1) simulado com $\phi_{1,1} = 0.9$, $\phi_{2,1} = -1.8$ e $n = 200$.

Existem testes utilizados para a detecção de não linearidades em uma dada série temporal, como por exemplo, o teste de Keenan, o de Tsay e o da não linearidade de limiar, os quais encontram-se detalhados no Apêndice B.

2.4 Modelos autorregressivos com limiar

Utilizando como referência principal Tong e Lim (1980), serão apresentados os conceitos básicos de um modelo autorregressivo com limiar (TAR), *threshold autoregressive model*.

Suponha $\{\mathbf{Z}_t\}$ uma série temporal p -dimensional e, para cada t , seja \mathbf{I}_t , uma variável aleatória politômica, tomando valores inteiros em $\{1, 2, \dots, l\}$.

Definição 2.10. Um processo $\{\mathbf{Z}_t; \mathbf{I}_t\}$ é dito ser um TAR geral se

$$\mathbf{Z}_t = \mathbf{B}^{(I_t)} \mathbf{Z}_t + \mathbf{A}^{(I_t)} \mathbf{Z}_{t-1} + \varepsilon_t^{(I_t)} + \mathbf{C}^{(I_t)}, \quad (2.47)$$

onde, para $I_t = i$, $\mathbf{A}^{(i)}$ e $\mathbf{B}^{(i)}$ são matrizes não aleatórias de coeficientes $p \times p$, $\mathbf{C}^{(i)}$ é um vetor $p \times 1$ de constantes e $\varepsilon_t^{(i)}$ é uma sequência ruído branco p -dimensional de vetores aleatórios independentes com matriz de covariância diagonal. Suponha também que $\varepsilon_t^{(i)}$ e $\varepsilon_t^{(i')}$ são independentes para $i \neq i'$.

Um modelo TAR com s regimes pode ser representado da seguinte maneira

$$\mathbf{Z}_t = \phi_0^{(I_t)} + \sum_{i=1}^k \phi_i^{(I_t)} \mathbf{Z}_{t-i} + a_t^{(I_t)}, \quad (2.48)$$

onde os a_t são independentes e identicamente distribuídos (iid), com média 0 e variância $\sigma_{I_t}^2$, e ϕ_i^j , $i = 0, \dots, k$, $j = 1, \dots, s$, são parâmetros reais.

Note que (2.48) pode ainda ser escrita da seguinte forma

$$\mathbf{Z}_t = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^k \phi_i^{(1)} \mathbf{Z}_{t-i} + a_t^{(1)}, & \text{se } I_t = 1 \\ \phi_0^{(2)} + \sum_{i=1}^k \phi_i^{(2)} \mathbf{Z}_{t-i} + a_t^{(2)}, & \text{se } I_t = 2 \\ \vdots \\ \phi_0^{(s)} + \sum_{i=1}^k \phi_i^{(s)} \mathbf{Z}_{t-i} + a_t^{(s)}, & \text{se } I_t = s \end{cases}. \quad (2.49)$$

2.4.1 Modelos autorregressivos com limiar autoexcitantes

Os modelos autorregressivos com limiar autoexcitantes (SETAR), *self-exciting threshold autoregressive model*, são um caso especial dos modelos TAR quando a variável politômica $\{I_t\}$ é função de uma defasagem dela própria.

Seja então $\alpha_0, \alpha_1, \dots, \alpha_l$ denotando um subconjunto ordenado linearmente dos números reais, tais que $\alpha_0 < \alpha_1 < \dots < \alpha_l$, onde α_0 e α_l representam $-\infty$ e $+\infty$, respectivamente. Esse subconjunto define uma partição do conjunto dos números reais \mathbb{R} , isto é, $\mathbb{R} = R_1 \cup R_2 \cup \dots \cup R_l$, onde $R_i = (\alpha_{i-1}, \alpha_i]$.

Suponha $\mathbf{Z}_t = (Z_t, Z_{t-1}, \dots, Z_{t-p+1})^t$, $\mathbf{A}^{(i)} = \left[\begin{array}{cccc|c} a_1^{(i)} & a_2^{(i)} & \dots & a_{p-1}^{(i)} & a_p^{(i)} \\ \hline & & & \mathbf{I}_{p-1} & \mathbf{0}_{(p-1) \times 1} \end{array} \right]$, $\mathbf{B}^{(i)} = \mathbf{0}$, $\varepsilon_t^{(i)} = (\varepsilon_t^{(i)}, 0, \dots, 0)$, $\mathbf{C}^{(i)} = (a_0^{(i)}, 0, \dots, 0)$ e $R_i^{(p)}$ é o conjunto de cilindros no produto cartesiano de p retas reais, no intervalo R_i com d -ésimo espaço de coordenadas, para algum d inteiro fixo pertencente a $1, 2, \dots, p$. Além disso, $I_t = i$ se $\mathbf{Z}_{t-1} \in R_i^{(p)}$.

A substituição desses elementos em (2.47), resulta na seguinte equação

$$Z_t = a_0^{(i)} + \sum_{j=1}^p a_j^{(i)} Z_{t-j} + \varepsilon_t^{(i)}, \text{ se } Z_{t-d} \in R_i, i = 1, 2, \dots, l. \quad (2.50)$$

O parâmetro d é denominado defasagem do modelo e, comumente, $1 < d \leq p$, onde p representa a ordem do modelo autorregressivo.

Uma vez que $\{I_t\}$ é agora uma função de $\{Z_t\}$ denominamos a série temporal univariada $\{Z_t\}$, dada pela equação (2.50), de modelo autorregressivo com limiar autoexcitante de ordem $(l; p, \dots, p)$ ou SETAR($l; p, \dots, p$) onde p é repetido l vezes. Se, para $i = 1, 2, \dots, l$, $a_j^{(i)} = 0$, para $j = p_i + 1, p_i + 2, \dots, p$, então denominamos $\{Z_t\}$ de SETAR($l; p_1, p_2, \dots, p_l$). Denominamos ainda $\alpha_1, \dots, \alpha_{l-1}$, os parâmetros limiares. Observe que um SETAR($1; p$) é apenas um modelo AR linear de ordem p .

Para exemplificar os modelos SETAR, suponha $s = 2$, $k = 2$ e tome $I_t = 1$, se $Z_{t-d} \leq \alpha$ e $I_t = 2$, se $Z_{t-d} > \alpha$, onde d e α representam, respectivamente, os parâmetros de defasagem e limiar. Dessa forma, podemos escrever o SETAR com 2 regimes de ordem 2, da seguinte forma

$$Z_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} Z_{t-1} + \phi_2^{(1)} Z_{t-2} + a_t^{(1)}, & \text{se } Z_{t-d} \leq \alpha \\ \phi_0^{(2)} + \phi_1^{(2)} Z_{t-1} + \phi_2^{(2)} Z_{t-2} + a_t^{(2)}, & \text{se } Z_{t-d} > \alpha \end{cases}. \quad (2.51)$$

O sistema de equações anterior indica que se o valor de defasagem d do processo não for maior que o parâmetro limiar α , a distribuição condicional de Z_t segue um modelo AR(2) com intercepto $\phi_0^{(1)}$ e parâmetros autorregressivos $\phi_1^{(1)}$ e $\phi_2^{(1)}$. Por outro lado, se Z_{t-d} for maior que α , o processo segue a distribuição dada pelo outro regime, ou seja, AR(2) com $(\phi_0^{(2)}, \phi_1^{(2)}$ e $\phi_2^{(2)})$.

Assim, o processo se alterna entre dois mecanismos lineares dependentes da posição do valor de defasagem d do processo. Quando Z_{t-d} não excede o limiar, dizemos que o processo está no regime inferior e, caso contrário, está no regime superior (Cryer e Chan, 2008).

Ergodicidade geométrica

Na análise de séries temporais é interessante verificar em que condições é possível se obter estacionariedade. Tong e Lim (1980) mencionam que uma condição suficiente para o SETAR, descrito em (2.50), ser ergódico no sentido de Tweedie (1975), é que o autovalor máximo de $C = A^{(i)T} A^{(i)}$, onde T representa a transposta da matriz, seja estritamente menor que 1, para $i = 1, \dots, l$ e que os $\varepsilon_t^{(i)}$ tenham distribuições absolutamente contínuas.

An e Huang (1996) também argumentaram sobre as condições suficientes para ergodicidade geométrica de modelos autorregressivos não lineares (NLAR) e, as definições expostas pelos autores, serão utilizadas para a apresentação dos seguintes conceitos.

O modelo NLAR(p) pode ser escrito da seguinte maneira

$$Z_t = \phi(Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}) + a_t, \quad t \geq 1, \quad (2.52)$$

em que a_t , $t \geq 1$ são variáveis aleatórias iid com densidades positivas em quase todo ponto e primeiro momento finito. Além disso, a_t é independente de Z_{t-s} , $s \geq 1$, $E(a_t)=0$ e a função ϕ , de \mathbb{R}^p a \mathbb{R} , é mensurável.

A fim de escrever o modelo (2.52) vetorialmente, definimos $\mathbf{Z}_t = (Z_t, Z_{t-1}, \dots, Z_{t-p+1})^t$, $\mathbf{T} : \mathbb{R}^p \rightarrow \mathbb{R}^p$, $\mathbf{T}(\mathbf{Z}_t) = (\phi(\mathbf{Z}_t), Z_t, Z_{t-1}, \dots, Z_{t-p+2})^t$ e $\mathbf{a}_t = (a_t, 0, \dots, 0)^t$, então (2.52) pode ser reescrito como

$$\mathbf{Z}_t = \mathbf{T}(\mathbf{Z}_{t-1}) + \mathbf{a}_t, \quad t \geq 1. \quad (2.53)$$

Dessa forma, \mathbf{Z}_t é uma cadeia de Markov e podemos verificar sob que condições o processo (2.52) é geometricamente ergódico. Para isso, serão utilizados alguns lemas e teoremas descritos a seguir. De acordo com (Fadel, 2012, p.6), o estudo da ergodicidade geométrica é realizado, porque se a cadeia de Markov definida em (2.53) for geometricamente ergódica, então o processo gerado pelo modelo (2.52) é estritamente estacionário.

Inicialmente, será apresentado o Lema 2.1, de autoria de Chan e Tong (1985).

Lema 2.1. *Suponha que a função autorregressiva não linear ϕ dada no modelo expresso pela equação (2.52) seja delimitada por conjuntos limitados. Então $\{Z_t\}$ satisfazendo (2.53) é uma cadeia de Markov aperiódica e μ -irredutível, onde μ representa a medida de Lebesgue. Além disso, os conjuntos μ -compactos não nulos são conjuntos pequenos.*

Definição 2.11. *Um conjunto $K \in \mathbb{R}$ é dito pequeno se para todo conjunto $A \in \mathbb{R}$*

$$\inf_{y \in K} \sum_{n=1}^m P^n(y, A) > 0, \quad (2.54)$$

para algum $m = m(A) > 0$, onde $P^n(y, \cdot)$ é a probabilidade de transição a n passos de \mathbf{Z}_t .

Para determinar a ergodicidade geométrica de $\{Z_t\}$ será utilizado o seguinte lema:

Lema 2.2. *(Critério de Tweedie (1975)) Seja $\{Z_t\}$ uma cadeia de Markov aperiódica e irredutível. Suponha que exista um pequeno conjunto C , uma função mensurável não negativa g e constantes positivas c_1 , c_2 e $\rho < 1$, tal que*

$$E\{g(Z_{t+1})|Z_t = z\} \leq \rho g(z) - c_1, \text{ para qualquer } z \notin C, \quad (2.55)$$

e

$$E\{g(Z_{t+1})|Z_t = z\} \leq c_2, \text{ para qualquer } z \in C. \quad (2.56)$$

Então $\{Z_t\}$ é geometricamente ergódica.

Esse critério é utilizado para obtermos a ergodicidade geométrica de cadeias de Markov.

Ao comprovar a ergodicidade geométrica das cadeias de Markov, outra lema muito útil é o seguinte

Lema 2.3. *(Critério de Tjostheim (1990) h-passos) Se houver um inteiro positivo h tal que $\{Z_{kh}\}$ seja geometricamente ergódica, então $\{Z_t\}$ é geometricamente ergódica.*

Prova. A prova desse lema pode ser encontrada em Tjostheim (1990). \square

Combinando os Lemas 2.1 e 2.3, o seguinte lema, proposto por Tjostheim (1990), pode tornar mais fácil e clara a obtenção da ergodicidade geométrica dos modelos NLAR.

Lema 2.4. *Suponha que Z_t satisfaça (2.53) e a função $\phi(\cdot)$ satisfaça o Lema 2.1. Então, se existir um inteiro positivo q , constantes positivas c_1, c_2 e $\rho < 1$ e um conjunto limitado $C_K = \{z : \|z\| \leq K\}$ tal que (2.55) e (2.56) valham quando Z_t for substituído por Z_{qt} , segue-se que $\{Z_t\}$ é geometricamente ergódico.*

Suponha agora o caso em que a função autorregressiva não linear $\phi(\cdot)$, atinente ao modelo dado na equação (2.52), satisfaz as seguintes condições

$$\sup_{\|z\| \leq K} \|\phi(z)\| < \infty, \text{ para cada } K > 0, \quad (2.57)$$

e

$$\lim_{\|z\| \rightarrow \infty} \frac{|\phi(z) - \alpha^\tau z|}{\|z\|} = 0, \quad (2.58)$$

onde $\alpha = (\alpha_1, \dots, \alpha_p)^\tau$ satisfaz

$$u_p - \alpha_1 u^{p-1} - \dots - \alpha_{p-1} u - \alpha_p \neq 0, \forall |u| \geq 1 \quad (2.59)$$

e $\|\cdot\|$ denota a norma Euclideana em \mathbb{R}^p .

Assim, o seguinte teorema foi apresentado por An e Huang (1996).

Teorema 2.2. *Se a função autorregressiva não-linear ϕ , cujo modelo é encontrado na equação (2.52), satisfaz (2.57) e (2.58), então o modelo (2.53), e, conseqüentemente, o modelo (2.52), são geometricamente ergódicos.*

Prova. A prova deste Teorema pode ser encontrada em (An e Huang, 1996, p. 948). \square

No entanto, para a classe de modelos autorregressivos com limiar, o Teorema 2.2 é inadequado na obtenção da ergodicidade geométrica. Para estudar a ergodicidade geométrica dessa classe de modelos, An e Huang (1996) propuseram o seguinte teorema

Teorema 2.3. *Se existe um número positivo $\lambda < 1$ e uma constante c tal que*

$$|\phi(Z_1, Z_2, \dots, Z_p)| \leq \lambda \max\{|Z_1|, \dots, |Z_p|\} + c \quad (2.60)$$

então o modelo dado pela equação (2.52) é geometricamente ergódico.

Prova. A demonstração completa do Teorema 2.3 pode ser encontrada em (An e Huang, 1996, p. 949-950). \square

No próximo capítulo serão detalhados os modelos autorregressivos com memória variável.

Os valores $\{\alpha_0, \alpha_1, \dots, \alpha_p\}$ são chamados de parâmetros limiares do modelo AR-MV(p). Eles formam uma partição dos números reais de forma que $\alpha_0 = -\infty$ e $\alpha_p = \infty$, $\mathbb{R} = \bigcup_{i=0}^{p-1} A_i$ e $A_i = (\alpha_i, \alpha_{i+1}]$, $i = 0, 1, 2, \dots, p-1$.

Como exemplo, suponha o modelo AR-MV(3), onde 3 representa a maior ordem autorregressiva entre todos os regimes, que pode ser escrito da seguinte maneira

$$Z_t = \begin{cases} \phi_1^{(1)} Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq \alpha_1 \\ \phi_1^{(1)} Z_{t-1} + \phi_2^{(1)} Z_{t-2} + a_t, & \text{se } \alpha_1 < Z_{t-1} \leq \alpha_2 \\ \phi_1^{(1)} Z_{t-1} + \phi_2^{(1)} Z_{t-2} + \phi_3^{(1)} Z_{t-3} + a_t, & \text{se } Z_{t-1} > \alpha_2 \end{cases} \quad (3.2)$$

Percebemos que o critério de seleção da variável Z_t , para cada um dos três regimes, é baseado no valor da variável imediatamente anterior (Z_{t-1}). Ademais, o parâmetro autorregressivo $\phi_1^{(1)}$ é o mesmo para os três regimes, assim como o parâmetro $\phi_2^{(1)}$ é o mesmo para os regimes 2 e 3. Além disso, os erros nos três regimes são idênticos.

Com o intuito de encontrar condições suficientes para garantir a ergodicidade geométrica de (3.1), Fadel (2012) utilizou o Teorema 2.3 (An e Huang, 1996), para que pudesse apresentar a seguinte proposição.

Proposição 3.1. *Se $\sum_{i=1}^p |\phi_i| < 1$, então o modelo AR-MV(p) é geometricamente ergódico.*

Prova (Fadel, 2012). Primeiramente, podemos escrever a equação (3.1) da seguinte forma

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} \mathbb{I}(Z_{t-1} > \alpha_1) + \dots + \phi_p Z_{t-p} \mathbb{I}(Z_{t-1} > \alpha_{p-1}) + a_t, \quad (3.3)$$

em que

$$\begin{aligned} \phi(Z_{t-1}, \dots, Z_{t-p}) &= \phi_1 Z_{t-1} + \phi_2 Z_{t-2} \mathbb{I}(Z_{t-1} > \alpha_1) + \dots + \phi_p Z_{t-p} \mathbb{I}(Z_{t-1} > \alpha_{p-1}) \\ &= \phi_1 Z_{t-1} + \sum_{i=2}^p \phi_i Z_{t-i} \mathbb{I}(Z_{t-1} > \alpha_{i-1}). \end{aligned} \quad (3.4)$$

Assim,

$$\begin{aligned}
|\phi(Z_{t-1}, \dots, Z_{t-p})| &= |\phi_1 Z_{t-1} + \sum_{i=2}^p \phi_i Z_{t-i} \mathbb{I}(Z_{t-1} > \alpha_{i-1})| \\
&\leq |\phi_1 Z_{t-1}| + |\sum_{i=2}^p \phi_i Z_{t-i} \mathbb{I}(Z_{t-1} > \alpha_{i-1})| \\
&\leq |\phi_1 Z_{t-1}| + \sum_{i=2}^p |\phi_i Z_{t-i}| \mathbb{I}(Z_{t-1} > \alpha_{i-1}) \\
&\leq |\phi_1 Z_{t-1}| + \sum_{i=2}^p |\phi_i Z_{t-i}| \\
&= \sum_{i=1}^p |\phi_i Z_{t-i}| \\
&= \sum_{i=1}^p |\phi_i| |Z_{t-i}| \\
&\leq \max\{|Z_{t-1}|, \dots, |Z_{t-p}|\} \sum_{i=1}^p |\phi_i|.
\end{aligned} \tag{3.5}$$

□

Dessa forma, é suficiente que $\sum_{i=1}^p |\phi_i| < 1$ para garantir que (3.1) seja geometricamente ergódico. De fato, basta conhecer os coeficientes do modelo relativo ao j -ésimo regime ($\phi_i^{(j)}$). Além disso, como os primeiros coeficientes se repetem, $\phi_1^{(1)} = \phi_1^{(2)} = \phi_1^{(3)} = \dots = \phi_1^{(j)}$, e $\phi_i^{(j)} = 0$, para $i > j$, o máximo será atingido no último regime.

3.2 Ilustração de propriedades do modelo por meio de simulações

Nesta seção serão apresentados alguns exemplos, contendo dados simulados, com o objetivo de estudar o comportamento do modelo AR-MV.

Cabe salientar que, em sua dissertação, (Fadel, 2012, p. 13-17) analisou o comportamento dos modelos AR-MV de ordens 2 e 3 e que um dos objetivos deste trabalho, consiste em avaliar modelos AR-MV de maiores ordens. Contudo, a fim de apresentar esse tipo de modelo ao leitor iniciante, os padrões comportamentais dos modelos AR-MV(2) e AR-MV(3) serão reexibidos. Ademais, a análise será estendida aos modelos AR-MV(4) e AR-MV(5).

Para isso, serão supostos vários parâmetros distintos (α e ϕ) e diferentes valores para a função L_t . Em todos os casos, o tamanho da amostra será $n = 1000$ e $a_t \sim N(0, 4)$.

Após a análise dos exemplos, foi possível registrar que o comportamento das séries temporais está associado ao valor atribuído para os parâmetros autorregressivos do modelo, de maneira que quando a condição para a ergodicidade geométrica ($\sum_{i=1}^p |\phi_i| < 1$) não é satisfeita, a série temporal não é estacionária e os dados não são normais, conforme podemos

visualizar nos exemplos 3.2, 3.4, 3.6 e 3.8.

Além disso, visualizamos que a FAC dos dados gerados nas simulações onde a condição de ergodicidade geométrica é observada, apresenta um padrão semelhante ao esperado para um modelo autorregressivo. Por outro lado, quando a condição de ergodicidade geométrica não é satisfeita, o padrão apresentado pela FAC é de um decaimento excessivamente lento, padrão esperado para séries não estacionárias.

Percebemos ainda que a FACP apresenta o mesmo comportamento, independentemente da condição de ergodicidade geométrica. Além do mais, a FACP mostra o padrão esperado para um modelo autorregressivo, ou seja, apenas os p primeiros coeficientes de um modelo AR-MV(p) são significativos.

Assim, podemos concluir, empiricamente, que o comportamento da FAC e da FACP dos modelos AR-MV, para os exemplos aqui descritos, são os mesmos esperados para os modelos AR convencionais, com exceção dos casos em que $\sum_{i=1}^p |\phi_i| \geq 1$, onde a FAC apresenta um decaimento lento, em função da não estacionariedade da série temporal gerada.

3.2.1 AR-MV(2)

Exemplo 3.1. *Os dados deste exemplo foram simulados a partir de um modelo AR-MV(2), com parâmetros $p = 2$, $\phi_1 = 0.2$, $\phi_2 = 0.6$ e $\alpha = 0.5$. Note que $\sum_{i=1}^p |\phi_i| < 1$. O seguinte sistema de equações resume essas informações*

$$Z_t = \begin{cases} 0.2Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq 0.5 \\ 0.2Z_{t-1} + 0.6Z_{t-2} + a_t, & \text{se } Z_{t-1} > 0.5 \end{cases} . \quad (3.6)$$

A Figura 3.1 apresenta a série temporal gerada pelo modelo, possibilitando a percepção de um padrão estacionário.

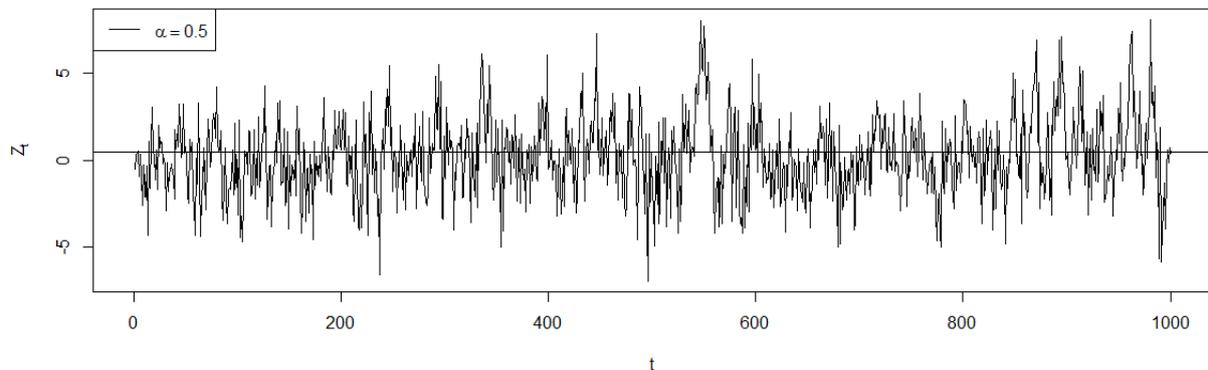


Figura 3.1: *Série temporal dos dados simulados para o Exemplo 3.1.*

A Figura 3.2 exhibe o histograma e o gráfico Q-Q dos dados. Percebemos que o ajuste segue adequadamente uma distribuição normal de probabilidade. Além disso, nessa figura são expostos também os gráficos da FAC e FACP dos dados. Em relação à FAC, observamos

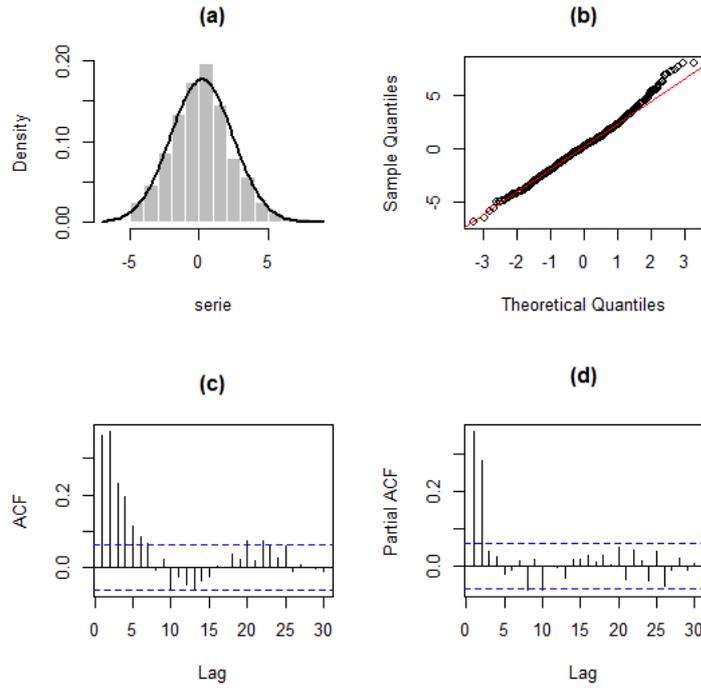


Figura 3.2: (a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.1.

um decaimento rápido para zero em formato senoidal; e a FACP apresenta os *lags* 1 e 2 altamente significativos. Observe que esses são os comportamentos esperados para um modelo AR(2).

Exemplo 3.2. Neste exemplo simulamos um modelo AR-MV(2), com parâmetros $p = 2$, $\phi_1 = 0.8$, $\phi_2 = 0.2$ e $\alpha = 1$. Perceba que $\sum_{i=1}^p |\phi_i| \geq 1$. O seguinte sistema de equações sumariza essas informações

$$Z_t = \begin{cases} 0.8Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq 1 \\ 0.8Z_{t-1} + 0.2Z_{t-2} + a_t, & \text{se } Z_{t-1} > 1 \end{cases}. \quad (3.7)$$

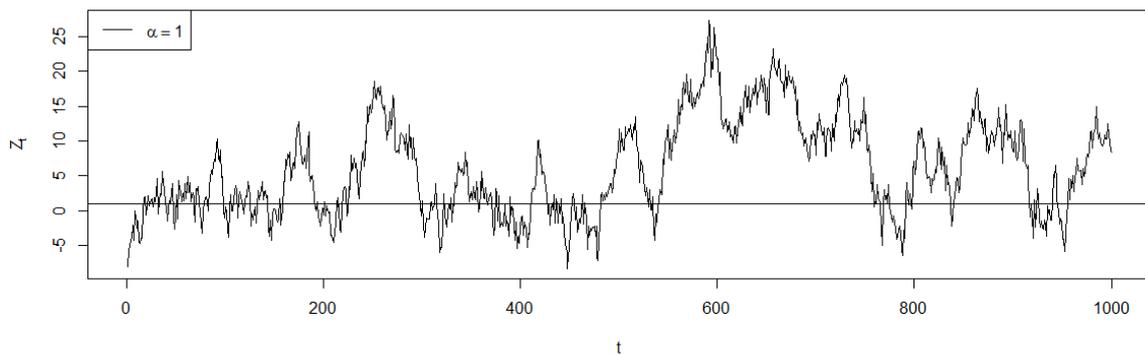


Figura 3.3: Série temporal dos dados simulados para o Exemplo 3.2.

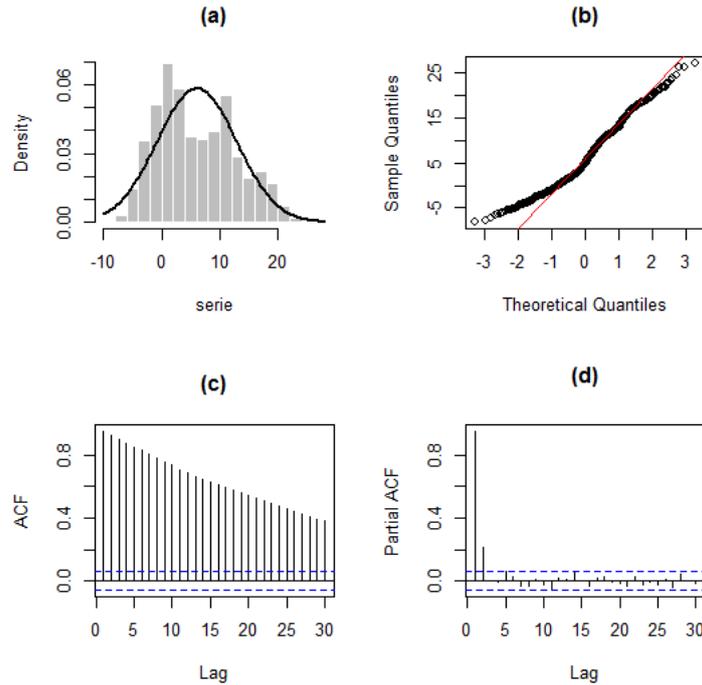


Figura 3.4: (a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.2.

A Figura 3.3 mostra a série temporal gerada pelo modelo e é possível observar um padrão não estacionário, com mudanças aparentes no nível e na variância da série temporal.

A Figura 3.4 expõe o histograma e o gráfico Q-Q dos dados. Visualizamos que o ajuste não segue convenientemente uma distribuição normal de probabilidade. Ademais, nessa figura também são exibidos os gráficos da FAC e FACP das observações. No que concerne à FAC, reparamos um decaimento para zero excessivamente lento, padrão geralmente observado em séries não estacionárias; e a FACP retrata os *lags* 1 e 2 significativos.

3.2.2 AR-MV(3)

Exemplo 3.3. As observações deste exemplo foram simuladas a partir de um modelo AR-MV(3), com parâmetros $p = 3$, $\phi_1 = 0.3$, $\phi_2 = 0.4$, $\phi_3 = 0.2$, $\alpha_1 = -0.5$ e $\alpha_2 = 0.5$. Repare que $\sum_{i=1}^p |\phi_i| < 1$. O sistema de equações a seguir consolida essas informações

$$Z_t = \begin{cases} 0.3Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq -0.5 \\ 0.3Z_{t-1} + 0.4Z_{t-2} + a_t, & \text{se } -0.5 < Z_{t-1} \leq 0.5 \\ 0.3Z_{t-1} + 0.4Z_{t-2} + 0.2Z_{t-3} + a_t, & \text{se } Z_{t-1} > 0.5 \end{cases} \quad (3.8)$$

A série temporal gerada pelo modelo, disposta na Figura 3.5, se desenvolve em torno da média, apresentando assim, características de um processo estacionário.

O histograma e o gráfico Q-Q dos dados são mostrados na Figura 3.6. Vemos que o ajuste segue de maneira satisfatória uma distribuição normal de probabilidade. Além do mais, nessa figura são revelados também os gráficos da FAC e FACP dos dados. No que se refere à FAC,

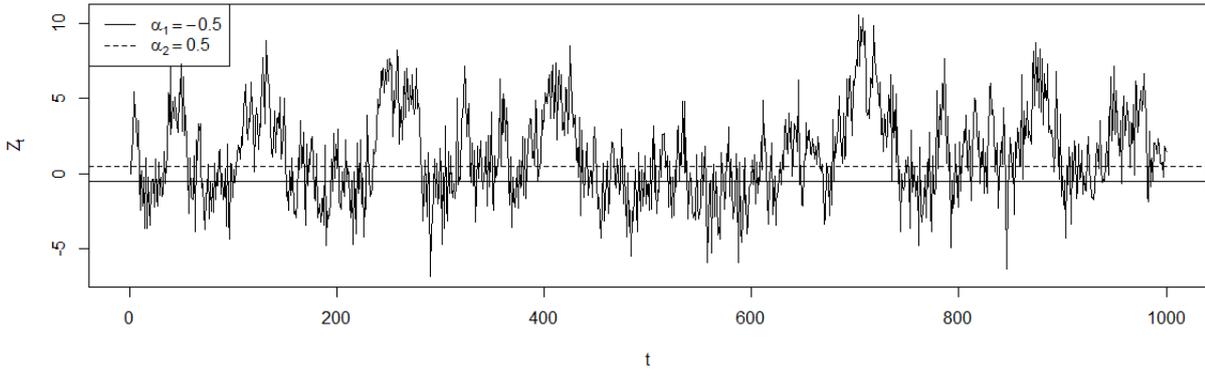


Figura 3.5: Série temporal dos dados simulados para o Exemplo 3.3.

um decaimento para zero é constatado; e a FACP exprime os lags 1, 2 e 3 significativos. Salientamos que esses são os comportamentos esperados para um modelo AR(3).

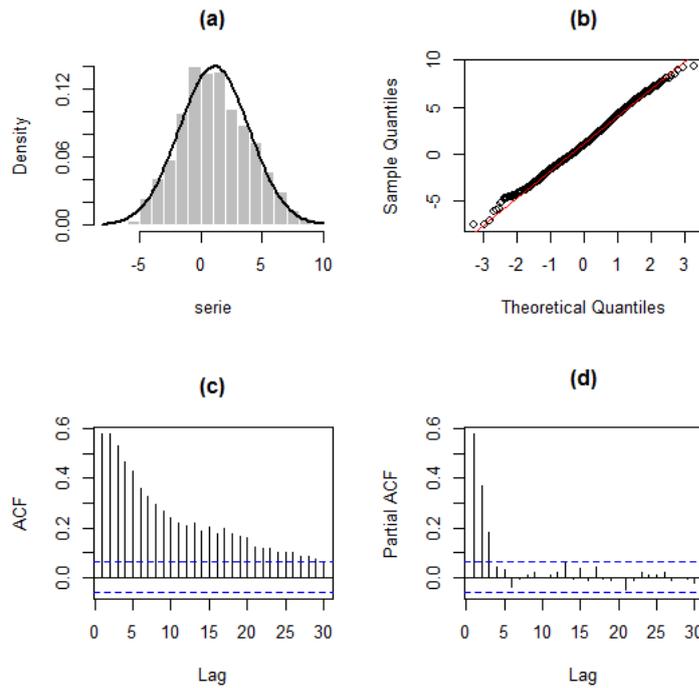


Figura 3.6: (a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.3.

Exemplo 3.4. As observações utilizadas neste exemplo foram simuladas a partir de um modelo AR-MV(3), com parâmetros $p = 3$, $\phi_1 = 0.2$, $\phi_2 = 0.5$, $\phi_3 = 0.3$, $\alpha_1 = -2$, $\alpha_2 = 2$. Veja que $\sum_{i=1}^p |\phi_i| \geq 1$. O seguinte sistema de equações sintetiza essas informações

$$Z_t = \begin{cases} 0.2Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq -2 \\ 0.2Z_{t-1} + 0.5Z_{t-2} + a_t, & \text{se } -2 < Z_{t-1} \leq 2 \\ 0.2Z_{t-1} + 0.5Z_{t-2} + 0.3Z_{t-3} + a_t, & \text{se } Z_{t-1} > 2 \end{cases} \quad (3.9)$$

A Figura 3.7 exterioriza a série temporal gerada pelo modelo, possibilitando-nos notar um padrão não estacionário, com modificações notórias no nível e na variância da série.

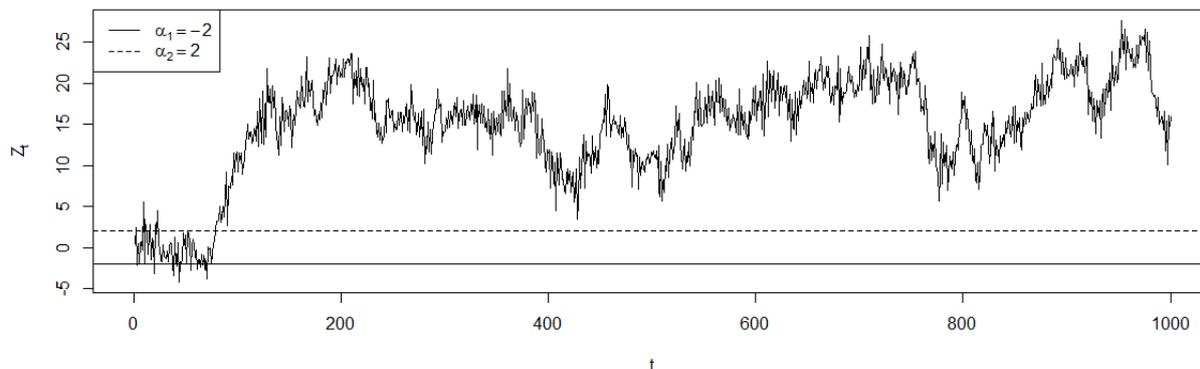


Figura 3.7: *Série temporal dos dados simulados para o Exemplo 3.4.*

A Figura 3.8 mostra o histograma e o gráfico Q-Q dos dados. Apreendemos que não há uma ajustabilidade conveniente a uma distribuição normal. Outrossim, nessa figura são vistos os gráficos da FAC e FACP dos dados. No que diz respeito à FAC, visualizamos um decaimento para zero demasiadamente prolongado, padrão habitualmente observado em séries não estacionárias; e a FACP exibe os lags 1, 2 e 3 significativos.

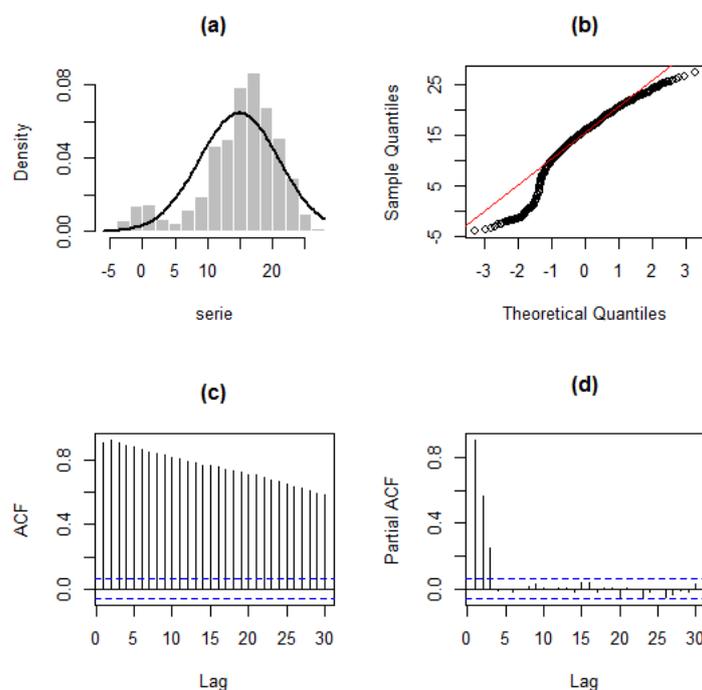


Figura 3.8: (a) *Histograma*, (b) *gráfico Q-Q*, (c) *FAC* e (d) *FACP* dos dados simulados para o Exemplo 3.4.

3.2.3 AR-MV(4)

Exemplo 3.5. Os dados deste exemplo foram simulados a partir de um modelo AR-MV(4), com parâmetros $p = 4$, $\phi_1 = 0.15$, $\phi_2 = 0.2$, $\phi_3 = 0.25$, $\phi_4 = 0.3$, $\alpha_1 = -2$, $\alpha_2 = 0$, $\alpha_3 = 2$. Note que $\sum_{i=1}^p |\phi_i| < 1$. O seguinte sistema de equações resume essas informações

$$Z_t = \begin{cases} 0.15Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq -2 \\ 0.15Z_{t-1} + 0.2Z_{t-2} + a_t, & \text{se } -2 < Z_{t-1} \leq 0 \\ 0.15Z_{t-1} + 0.2Z_{t-2} + 0.25Z_{t-3} + a_t, & \text{se } 0 < Z_{t-1} \leq 2 \\ 0.15Z_{t-1} + 0.2Z_{t-2} + 0.25Z_{t-3} + 0.3Z_{t-4} + a_t, & \text{se } Z_{t-1} > 2 \end{cases} \quad (3.10)$$

A Figura 3.9 apresenta a série temporal gerada pelo modelo, o que nos faz perceber um padrão estacionário.

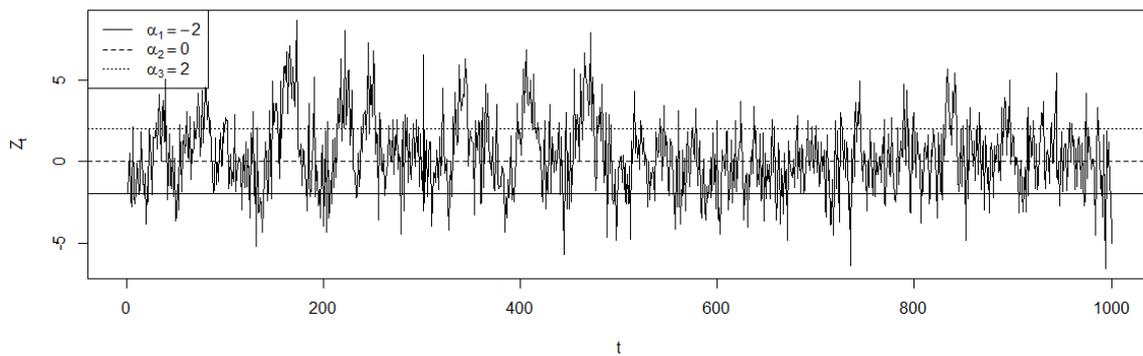


Figura 3.9: Série temporal dos dados simulados para o Exemplo 3.5.

A seguir, a Figura 3.10 exibe o histograma e o gráfico Q-Q dos dados. Percebemos que o ajuste segue adequadamente uma distribuição normal de probabilidade. Além disso, nessa figura são expostos também os gráficos da FAC e FACP dos dados. Em relação à FAC, observamos um decaimento veloz para zero; e a FACP apresenta os *lags* 1, 2, 3 e 4 significativos. Reparemos que esses são os comportamentos esperados para um modelo AR(4).

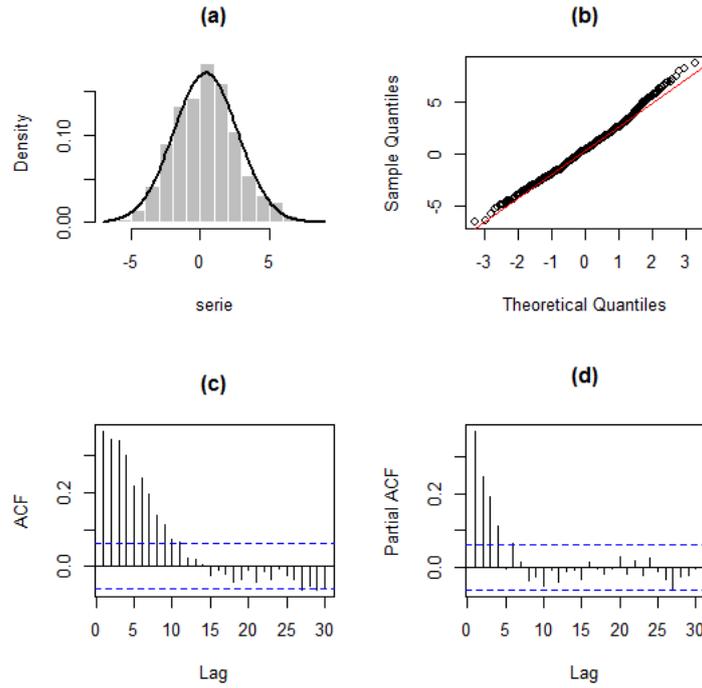


Figura 3.10: (a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.5.

Exemplo 3.6. Neste exemplo foram simulados dados a partir de um modelo $AR-MV(4)$, com parâmetros $p = 4$, $\phi_1 = 0.15$, $\phi_2 = 0.2$, $\phi_3 = 0.25$, $\phi_4 = 0.4$, $\alpha_1 = -2$, $\alpha_2 = 0$, $\alpha_3 = 2$. Repare que $\sum_{i=1}^p |\phi_i| \geq 1$. O seguinte sistema de equações abrevia essas informações

$$Z_t = \begin{cases} 0.15Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq -2 \\ 0.15Z_{t-1} + 0.2Z_{t-2} + a_t, & \text{se } -2 < Z_{t-1} \leq 0 \\ 0.15Z_{t-1} + 0.2Z_{t-2} + 0.25Z_{t-3} + a_t, & \text{se } 0 < Z_{t-1} \leq 2 \\ 0.15Z_{t-1} + 0.2Z_{t-2} + 0.25Z_{t-3} + 0.4Z_{t-4} + a_t, & \text{se } Z_{t-1} > 2 \end{cases} . \quad (3.11)$$

A Figura 3.11 expõe a série temporal gerada pelo modelo, possibilitando-nos identificar um padrão não estacionário, com mudanças aparentes no nível e na variância da série.

Por meio da Figura 3.12, visualizamos o histograma e o gráfico Q-Q dos dados. Notamos que o ajuste é inadequado a uma distribuição normal de probabilidade. Além de tudo, nessa figura observamos os gráficos da FAC e FACP dos dados. No que concerne à FAC, um decaimento para zero bastante lento é visto, padrão comum em séries não estacionárias; e por meio da FACP notamos que os lags 1, 2, 3 e 4 são significativos.

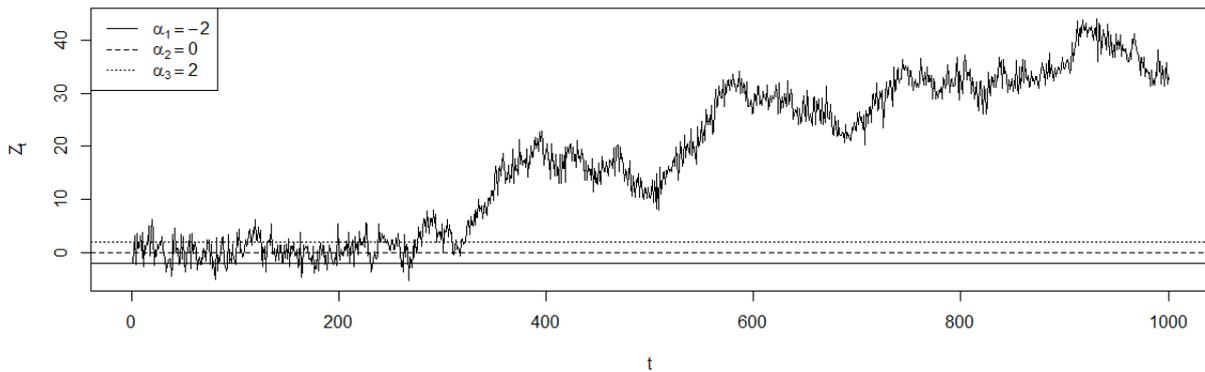


Figura 3.11: Série temporal dos dados simulados para o Exemplo 3.6.

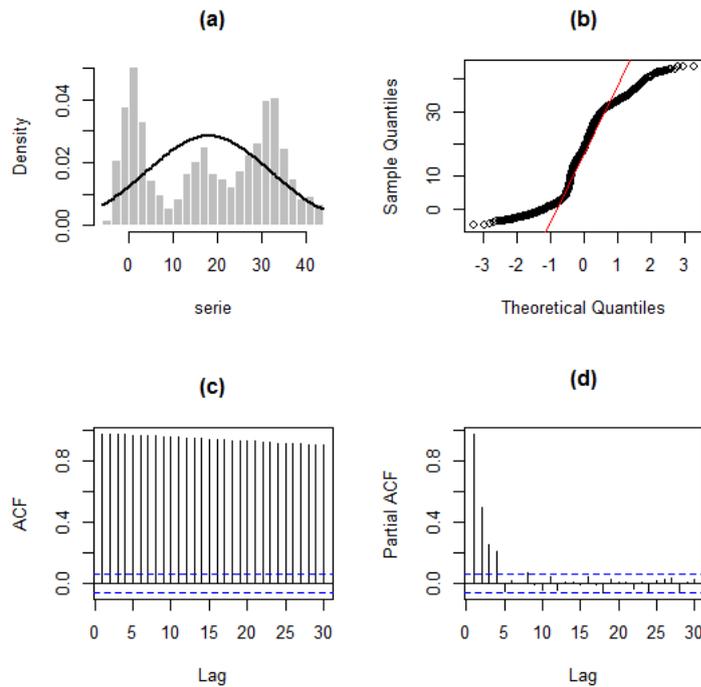


Figura 3.12: (a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.6.

3.2.4 AR-MV(5)

Exemplo 3.7. Estes dados foram simulados a partir de um modelo AR-MV(5), com parâmetros $p = 5$, $\phi_1 = 0.05$, $\phi_2 = 0.15$, $\phi_3 = 0.2$, $\phi_4 = 0.25$, $\phi_5 = 0.3$, $\alpha_1 = -2$, $\alpha_2 = -1$, $\alpha_3 = 1$, $\alpha_4 = 2$. Note que $\sum_{i=1}^p |\phi_i| < 1$. O seguinte sistema de equações reúne essas informações

$$Z_t = \begin{cases} 0.05Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq -2 \\ 0.05Z_{t-1} + 0.15Z_{t-2} + a_t, & \text{se } -2 < Z_{t-1} \leq -1 \\ 0.05Z_{t-1} + 0.15Z_{t-2} + 0.2Z_{t-3} + a_t, & \text{se } -1 < Z_{t-1} \leq 1 \\ 0.05Z_{t-1} + 0.15Z_{t-2} + 0.2Z_{t-3} + 0.25Z_{t-4} + a_t, & \text{se } 1 < Z_{t-1} \leq 2 \\ 0.05Z_{t-1} + 0.15Z_{t-2} + 0.2Z_{t-3} + 0.25Z_{t-4} + 0.3Z_{t-5} + a_t, & \text{se } Z_{t-1} > 2 \end{cases} \quad (3.12)$$

A Figura 3.13 mostra a série temporal gerada pelo modelo, pela qual podemos constatar um padrão estacionário.

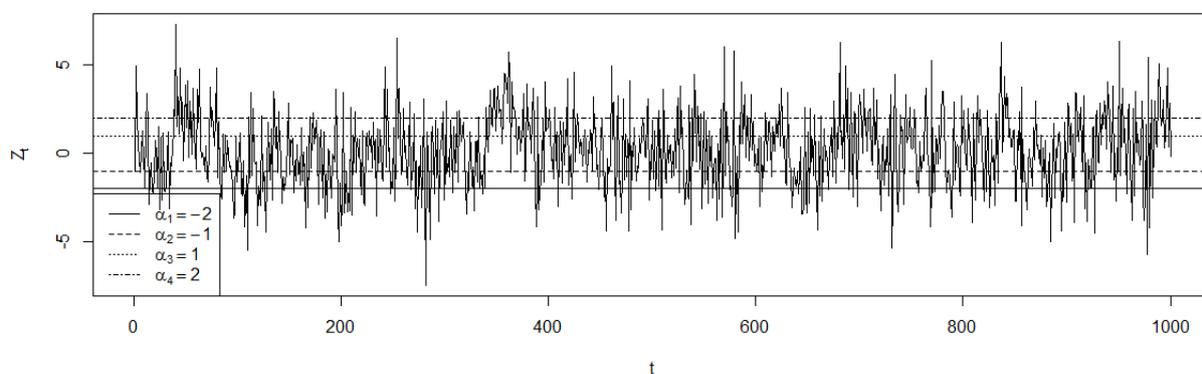


Figura 3.13: Série temporal dos dados simulados para o Exemplo 3.7.

Em seguida, a Figura 3.14 apresenta o histograma e o gráfico Q-Q dos dados. Notamos que o ajuste segue de forma adequada uma distribuição normal de probabilidade. Ademais, nessa figura são observados os gráficos da FAC e FACP dos dados. Com relação à FAC, vemos um decaimento rápido para zero; e a FACP apresenta os *lags* 1, 2, 3, 4 e 5 significativos. Lembre que esses são os comportamentos esperados para um modelo AR(5).

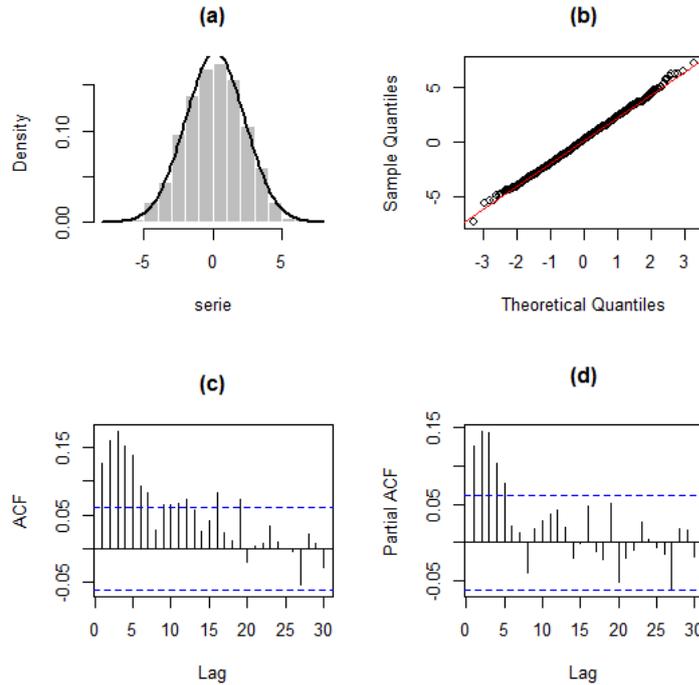


Figura 3.14: (a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.7.

Exemplo 3.8. Simulamos os dados deste exemplo a partir de um modelo AR-MV(5), com parâmetros $p = 5$, $\phi_1 = 0.1$, $\phi_2 = 0.15$, $\phi_3 = 0.2$, $\phi_4 = 0.25$, $\phi_5 = 0.3$, $\alpha_1 = -2$, $\alpha_2 = -1$, $\alpha_3 = 1$, $\alpha_4 = 2$. Observe que $\sum_{i=1}^p |\phi_i| \geq 1$. O seguinte sistema de equações condensa essas informações:

$$Z_t = \begin{cases} 0.1Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq -2 \\ 0.1Z_{t-1} + 0.15Z_{t-2} + a_t, & \text{se } -2 < Z_{t-1} \leq -1 \\ 0.1Z_{t-1} + 0.15Z_{t-2} + 0.2Z_{t-3} + a_t, & \text{se } -1 < Z_{t-1} \leq 1 \\ 0.1Z_{t-1} + 0.15Z_{t-2} + 0.2Z_{t-3} + 0.25Z_{t-4} + a_t, & \text{se } 1 < Z_{t-1} \leq 2 \\ 0.1Z_{t-1} + 0.15Z_{t-2} + 0.2Z_{t-3} + 0.25Z_{t-4} + 0.3Z_{t-5} + a_t, & \text{se } Z_{t-1} > 2 \end{cases} \quad (3.13)$$

A Figura 3.15 exibe a série temporal gerada pelo modelo, o que nos permite notar um padrão não estacionário, com mudanças visíveis no nível e na variância da série.

A Figura 3.16 mostra o histograma e o gráfico Q-Q dos dados. Verificamos que o ajuste não é adequado a uma distribuição normal de probabilidade. Ainda, nessa figura são expostos os gráficos da FAC e FACP dos dados. Em relação à FAC, notamos um decaimento para zero com excessiva lentidão, padrão regularmente observado em séries não estacionárias; e a FACP mostra os lags 1, 2, 3, 4 e 5 significativos.

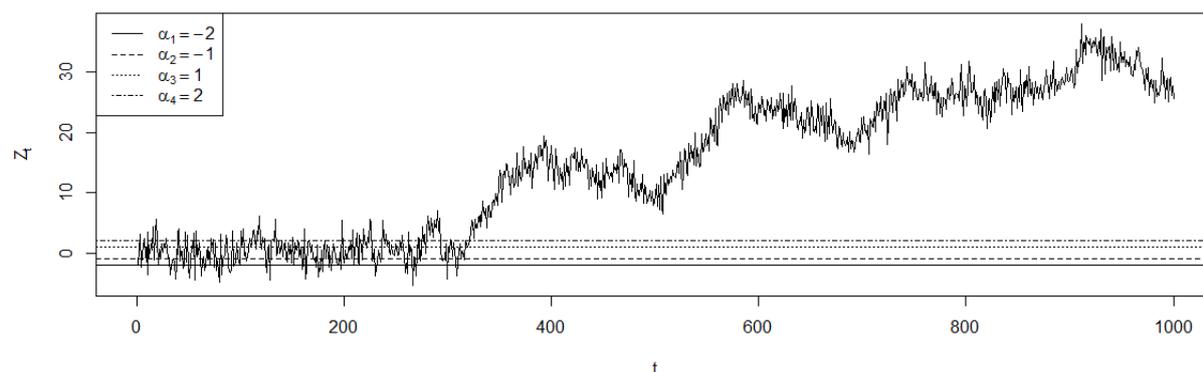


Figura 3.15: Série temporal dos dados simulados para o Exemplo 3.8.

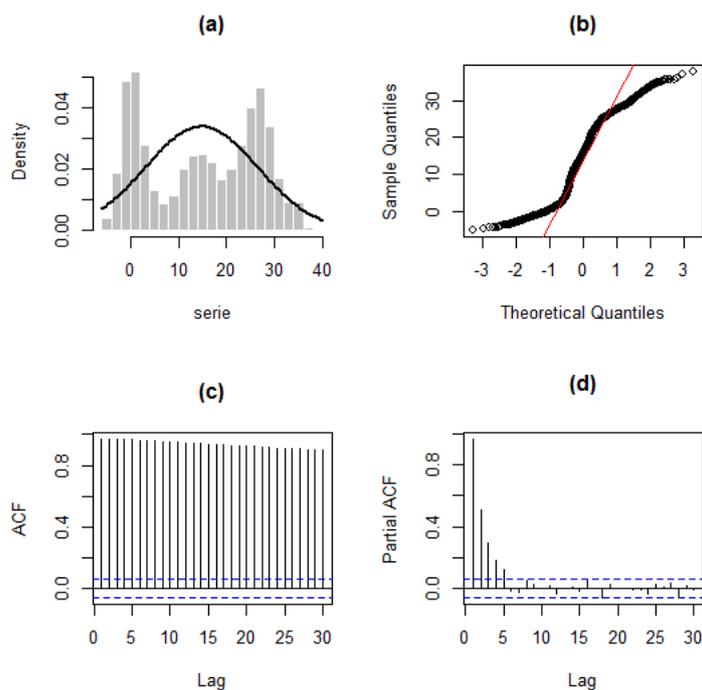


Figura 3.16: (a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.8.

3.3 Estimação dos parâmetros

Nesta seção serão apresentados os estimadores condicionais dos parâmetros do modelo AR-MV, a depender dos casos em que o parâmetro limiar (α) é conhecido e quando é desconhecido. O processo de obtenção desses estimadores foi baseado nos métodos dos mínimos quadrados condicionais e de máxima verossimilhança. Mais detalhes poderão ser encontrados em Fadel (2012).

Caso α conhecido

No caso em que α é conhecido, sob o modelo AR-MV(p), Fadel (2012) mostra que os estimadores condicionais de mínimos quadrados e de máxima verossimilhança para ϕ são idênticos e dados por

$$\hat{\phi}(\alpha) = [(\mathbf{Z}_\alpha^*)^T \mathbf{Z}_\alpha^*]^{-1} (\mathbf{Z}_\alpha^*)^t \mathbf{Z}, \quad (3.14)$$

em que $\mathbf{Z}_\alpha^* = \begin{bmatrix} Z_p & Z_{p+1_1}(\alpha_1) & \dots & Z_{p+1_{p-1}}(\alpha_{p-1}) \\ \vdots & \vdots & \vdots & \vdots \\ Z_{n-1} & Z_{n_1}(\alpha_1) & \dots & Z_{n_{p-1}}(\alpha_{p-1}) \end{bmatrix}$ e $Z_{t_i}(\alpha_i) = Z_{t-(i+1)} \mathbb{I}(Z_{t-1} > \alpha_i)$,
supondo $\mathbb{I}(\cdot)$

Além disso,

$$\mathbf{Z} = \begin{bmatrix} Z_{p+1} \\ \vdots \\ Z_n \end{bmatrix} = \begin{bmatrix} Z_p & Z_{p+1_1}(\alpha_1) & \dots & Z_{p+1_{p-1}}(\alpha_{p-1}) \\ \vdots & \vdots & \vdots & \vdots \\ Z_{n-1} & Z_{n_1}(\alpha_1) & \dots & Z_{n_{p-1}}(\alpha_{p-1}) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_p \end{bmatrix} + \begin{bmatrix} a_{p+1} \\ \vdots \\ a_n \end{bmatrix}. \quad (3.15)$$

A matriz de covariância dos parâmetros é dada por

$$\hat{V}(\hat{\beta}(\alpha)) = \hat{\sigma}^2 [(\mathbf{Z}_\alpha^*)^T \mathbf{Z}_\alpha^*]^{-1}, \quad (3.16)$$

onde $\hat{\sigma}^2 = \frac{\sum_{t=p+1}^n \{[Z_t - E(Z_t)]^2\}}{n - p}$.

Caso α desconhecido

No caso em que α é desconhecido, Fadel (2012) mostra que, para se obter o estimador condicional de mínimos quadrados para α , sob o modelo AR-MV(p), é necessário minimizar

$$S(\alpha) = \mathbf{Z}^T \mathbf{Z} - \mathbf{Z}^T \mathbf{Z}_\alpha^* [(\mathbf{Z}_\alpha^*)^T \mathbf{Z}_\alpha^*]^{-1} (\mathbf{Z}_\alpha^*)^T \mathbf{Z}. \quad (3.17)$$

Assim, de posse das estimativas dos parâmetros lineares, procedemos como se α fosse conhecido, e estimamos ϕ por meio da equação (3.14).

3.4 Previsão

Nesta seção serão mostrados os métodos utilizados para a realização de previsões do modelo AR-MV, quais sejam, Esqueleto e Monte Carlo. Maiores detalhes poderão ser encontrados em Fadel (2012).

3.4.1 Método do Esqueleto

O nome “esqueleto” deste método tem relação com a estrutura do modelo ao desprezarmos os erros, ou seja, quando os erros são supostos nulos.

Suponha então, o seguinte modelo AR-MV(p)

$$Z_t = \phi(Z_{t-1}, \dots, Z_{t-p}) + a_t, t \geq p + 1. \quad (3.18)$$

A previsão um passo à frente é dada por

$$\hat{Z}_t(1) = E[Z_{t+1}|Z_t, \dots, Z_1] = \phi(Z_t, \dots, Z_{t-p+1}). \quad (3.19)$$

A dois passos é dada por

$$\hat{Z}_t(2) = \phi(\hat{Z}_t(1), Z_t, \dots, Z_{t+2-p}). \quad (3.20)$$

A três passos é dada por

$$\hat{Z}_t(3) = \phi(\hat{Z}_t(2), \hat{Z}_t(1), Z_t, \dots, Z_{t+3-p}). \quad (3.21)$$

Portanto, para $h > p$, temos a seguinte fórmula

$$\hat{Z}_t(h) = \phi(\hat{Z}_t(h-1), \dots, \hat{Z}_t(h-p)). \quad (3.22)$$

Exemplo 3.9. *Suponha que queiramos estimar as previsões a um, dois e três passos à frente de um modelo AR-MV(3) hipotético pelo método do esqueleto. Assim, a previsão a um passo é dada por*

$$\hat{Z}_t(1) = \begin{cases} \hat{\phi}_1 Z_t, & \text{se } Z_t \leq \hat{\alpha}_1 \\ \hat{\phi}_1 Z_t + \hat{\phi}_2 Z_{t-1}, & \text{se } \hat{\alpha}_1 < Z_t \leq \hat{\alpha}_2 \\ \hat{\phi}_1 Z_t + \hat{\phi}_2 Z_{t-1} + \hat{\phi}_3 Z_{t-2}, & \text{se } Z_t > \hat{\alpha}_2 \end{cases},$$

onde Z_t representa a última observação da série temporal analisada.

A dois passos é dada por

$$\hat{Z}_t(2) = \begin{cases} \hat{\phi}_1 \hat{Z}_t(1), & \text{se } \hat{Z}_t(1) \leq \hat{\alpha}_1 \\ \hat{\phi}_1 \hat{Z}_t(1) + \hat{\phi}_2 Z_t, & \text{se } \hat{\alpha}_1 < \hat{Z}_t(1) \leq \hat{\alpha}_2 \\ \hat{\phi}_1 \hat{Z}_t(1) + \hat{\phi}_2 Z_t + \hat{\phi}_3 Z_{t-1}, & \text{se } \hat{Z}_t(1) > \hat{\alpha}_2 \end{cases}.$$

A três passos é dada por

$$\hat{Z}_t(3) = \begin{cases} \hat{\phi}_1 \hat{Z}_t(2), & \text{se } \hat{Z}_t(2) \leq \hat{\alpha}_1 \\ \hat{\phi}_1 \hat{Z}_t(2) + \hat{\phi}_2 \hat{Z}_t(1), & \text{se } \hat{\alpha}_1 < \hat{Z}_t(2) \leq \hat{\alpha}_2 \\ \hat{\phi}_1 \hat{Z}_t(2) + \hat{\phi}_2 \hat{Z}_t(1) + \hat{\phi}_3 Z_t, & \text{se } \hat{Z}_t(2) > \hat{\alpha}_2 \end{cases}.$$

3.4.2 Método de Monte Carlo

De maneira geral, o método de Monte Carlo consiste na simulação de n amostras de uma variável aleatória, que segue uma determinada distribuição de probabilidades. Neste caso, especificamente, serão geradas n amostras para a variável $a_t \sim N(0, \sigma^2)$, que corresponde ao termo de erro do modelo AR-MV. Dessa forma, teremos n previsões para cada um dos horizontes de previsão (h) fixados. Ao final, calculamos a média desses valores, encontrando assim, a previsão h passos à frente desejada.

Novamente, suponha o seguinte modelo AR-MV(p)

$$Z_t = \phi(Z_{t-1}, \dots, Z_{t-p}) + a_t, t \geq p + 1. \quad (3.23)$$

A previsão um passo à frente é a mesma dada para o método Esqueleto, ou seja

$$\hat{Z}_t^{MC}(1) = E[Z_{t+1}|Z_t, \dots, Z_1] = \phi(Z_t, \dots, Z_{t-p+1}). \quad (3.24)$$

A dois passos temos que

$$\hat{Z}_t^{MC_j}(2) = \phi(\hat{Z}_t^{MC}(1), Z_t, \dots, Z_{t+2-p}) + a_{2,j}, \quad (3.25)$$

onde $a_{2,j} \sim N(0, \sigma_2^2)$.

A três passos temos que

$$\hat{Z}_t^{MC_j}(3) = \phi(\hat{Z}_t^{MC_j}(2), \hat{Z}_t^{MC}(1), Z_t, \dots, Z_{t+3-p}) + a_{3,j}, \quad (3.26)$$

onde $a_{3,j} \sim N(0, \sigma_3^2)$.

Para $h > p$, temos que

$$\hat{Z}_t^{MC_j}(h) = \phi(\hat{Z}_t^{MC_j}(h-1), \dots, \hat{Z}_t^{MC_j}(h-p)) + a_{h,j}, \quad (3.27)$$

onde $a_{h,j} \sim N(0, \sigma_h^2)$.

Portanto, tomando a média das j iterações do método, $j = 1, \dots, n$, obtemos

$$\hat{Z}_t^{MC}(h) = \frac{1}{n} \sum_{j=1}^n \hat{Z}_t^{MC_j}(h). \quad (3.28)$$

Exemplo 3.10. *Suponha que queiramos estimar as previsões para um, dois e três passos à frente de um modelo AR-MV(3) hipotético, pelo método de Monte Carlo. A previsão um passo à frente é dada da mesma maneira do método Esqueleto. A dois passos é dada por*

$$\hat{Z}_t^{MC_j}(2) = \begin{cases} \hat{\phi}_1 \hat{Z}_t^{MC_j}(1) + a_{2,j}, & \text{se } \hat{Z}_t^{MC_j}(1) \leq \hat{\alpha}_1 \\ \hat{\phi}_1 \hat{Z}_t^{MC_j}(1) + \hat{\phi}_2 Z_t + a_{2,j}, & \text{se } \hat{\alpha}_1 < \hat{Z}_t^{MC_j}(1) \leq \hat{\alpha}_2 \\ \hat{\phi}_1 \hat{Z}_t^{MC_j}(1) + \hat{\phi}_2 Z_t + \hat{\phi}_3 Z_{t-1} + a_{2,j}, & \text{se } \hat{Z}_t^{MC_j}(1) > \hat{\alpha}_2 \end{cases} .$$

A três passos é dada por

$$\hat{Z}_t^{MC_j}(3) = \begin{cases} \hat{\phi}_1 \hat{Z}_t^{MC_j}(2) + a_{3,j}, & \text{se } \hat{Z}_t^{MC_j}(2) \leq \hat{\alpha}_1 \\ \hat{\phi}_1 \hat{Z}_t^{MC_j}(2) + \hat{\phi}_2 \hat{Z}_t^{MC_j}(1) + a_{3,j}, & \text{se } \hat{\alpha}_1 < \hat{Z}_t^{MC_j}(2) \leq \hat{\alpha}_2 \\ \hat{\phi}_1 \hat{Z}_t^{MC_j}(2) + \hat{\phi}_2 \hat{Z}_t^{MC_j}(1) + \hat{\phi}_3 Z_t + a_{3,j}, & \text{se } \hat{Z}_t^{MC_j}(2) > \hat{\alpha}_2 \end{cases} .$$

3.4.3 Erro de previsão

O erro de previsão m passos à frente é dado por

$$e_m = Z_{t+m} - \hat{Z}_t(m). \quad (3.29)$$

Neste trabalho serão tomadas duas medidas de comparação de previsão entre os modelos, quais sejam, a média absoluta dos erros de previsão (MAE) e a média quadrática dos erros de previsão (MQE), respectivamente, dadas por

$$MAE = \frac{1}{h} \sum_{m=n+1}^{n+h} |e_m| \quad \text{e} \quad MQE = \frac{1}{h} \sum_{m=n+1}^{n+h} e_m^2. \quad (3.30)$$

3.5 Estudos simulados

Nesta seção realizamos estudos simulados dos modelos AR-MV(4) e AR-MV(5). Verificamos as qualidades das estimativas geradas pelos modelos, variando o tamanho da amostra, e comparando-as com os valores verdadeiros; os resíduos desses modelos serão analisados, de forma a conferir a adequabilidade do modelo aos dados; e, por fim, as capacidades preditivas e a ajustabilidade dos modelos serão testadas em comparação com modelos autorregressivos de mesma ordem, ou seja, AR(4) e AR(5).

Inicialmente foram selecionados valores para os parâmetros α e ϕ . Para o parâmetro α foram escolhidos valores inteiros no intervalo $(-2, 2)$. Para o parâmetro ϕ a escolha dos valores foi realizada de modo a validar a condição de ergodicidade geométrica, ou seja, $\sum_{i=1}^p |\phi_i| < 1$, conforme demonstrado por intermédio da equação (3.5). Além disso, $a_t \sim N(0, 4)$.

O algoritmo para a estimação, utilizado por Fadel (2012), basicamente, funciona da seguinte maneira: são indicados como valores iniciais para os parâmetros α , as estimativas do ajuste de um modelo AR, de mesma ordem do modelo AR-MV, aos dados e, então, são estimados os parâmetros ϕ . Em seguida, por meio dos parâmetros ϕ , são estimados

novamente os parâmetros α , e assim, sucessivamente, sendo fixada como critério de parada, uma variação nos parâmetros de 10^{-5} .

Nos Exemplos 3.9 e 3.10 foram geradas amostras de tamanho $n + 5$, sendo que os cinco últimos dados gerados foram reservados para se testar a capacidade preditiva dos modelos. Cabe ressaltar que esses dados não foram utilizados na fase da modelagem.

Em cada exemplo comparamos o ajuste e a capacidade preditiva dos modelos AR-MV em relação à modelos AR de mesma ordem.

Após examinarmos os dois exemplos, pudemos concluir que, para ambos os casos, os modelos AR-MV adequaram-se melhor aos dados, quando comparados aos modelos AR, independentemente do tamanho da amostra. Também nos dois casos, verificamos um ajuste satisfatório dos dados à distribuição normal, que apresentou uma melhora significativa quando utilizamos $n = 10000$. Da mesma forma, os resíduos dos modelos AR-MV se ajustaram bem à normal, mostrando clara melhora quando aumentamos o tamanho da amostra. Ainda em relação aos resíduos, os gráficos de dispersão apresentaram pontos aleatórios e livres de valores discrepantes e as FAC e FACP não expuseram defasagens significativas, confirmando a pertinência desses modelos. No que concerne à previsão, constatamos a importância do tamanho da amostra, pois, no caso onde $n = 500$, para os dois exemplos, o modelo AR sem intercepto exibiu uma capacidade preditiva mais adequada. Entretanto, quando $n = 10000$, o AR-MV foi melhor em ambos os exemplos.

3.5.1 AR-MV(4)

Exemplo 3.11. *Os dados deste exemplo foram simulados a partir de um modelo AR-MV(4), com parâmetros $p = 4$, $\phi_1 = 0.15$, $\phi_2 = 0.2$, $\phi_3 = 0.25$, $\phi_4 = 0.3$, $\alpha_1 = -2$, $\alpha_2 = 0$, $\alpha_3 = 2$ e $\sigma^2 = 4$. Note que $\sum_{i=1}^p |\phi_i| < 1$. O seguinte sistema de equações resume essas informações*

$$Z_t = \begin{cases} 0.15Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq -2 \\ 0.15Z_{t-1} + 0.2Z_{t-2} + a_t, & \text{se } -2 < Z_{t-1} \leq 0 \\ 0.15Z_{t-1} + 0.2Z_{t-2} + 0.25Z_{t-3} + a_t, & \text{se } 0 < Z_{t-1} \leq 2 \\ 0.15Z_{t-1} + 0.2Z_{t-2} + 0.25Z_{t-3} + 0.3Z_{t-4} + a_t, & \text{se } Z_{t-1} > 2 \end{cases} \quad (3.31)$$

As Figuras 3.17 e 3.18 apresentam as séries temporais geradas pelo modelo com tamanho de amostra $n = 500$ e $n = 10000$, respectivamente. Podemos perceber um padrão estacionário nas duas séries, com melhores evidências para a Figura 3.18, em função do tamanho da amostra.

Na Tabela 3.1 estão descritos os valores das estimativas dos parâmetros dos modelos AR-MV(4) e AR(4) (com e sem intercepto), para dois tamanhos de amostra ($n = 500$) e ($n = 10000$). Como critérios de comparação entre os modelos estipulamos a soma dos quadrados dos resíduos (SQR), o AIC e o BIC. Notamos que em ambos os tamanhos de amostra, o modelo AR-MV foi superior em todos os critérios, mostrando sua adequabilidade

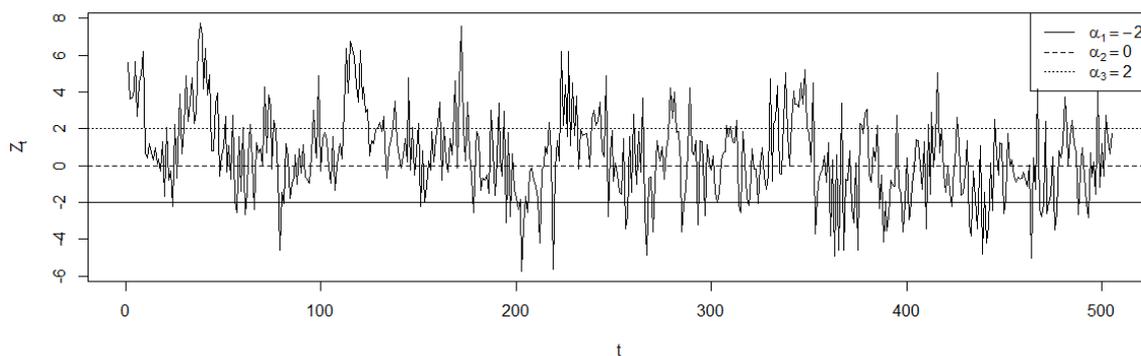


Figura 3.17: *Série temporal dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n=500$.*

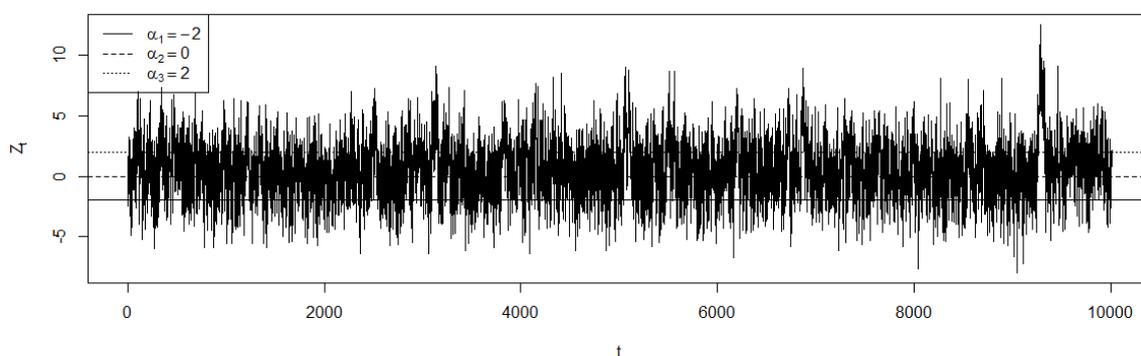


Figura 3.18: *Série temporal dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n=10000$.*

aos dados. Podemos perceber também que à medida em que aumentamos o tamanho da amostra, as estimativas aproximam-se dos valores reais dos parâmetros.

Tabela 3.1: Valores estimados dos parâmetros dos modelos ajustados para os dados simulados do Exemplo 3.9. (O valor entre parênteses indica o erro-padrão de cada estimativa).

n	Parâmetros	AR(4) (com intercepto)	AR(4) (sem intercepto)	AR-MV(4)
500	$\hat{\alpha}_1$	-	-	-1.98908
	$\hat{\alpha}_2$	-	-	-0.22113
	$\hat{\alpha}_3$	-	-	1.41812
	intercepto	0.59388 (0.29334)	- -	- -
	$\hat{\phi}_1$	0.19804 (0.04454)	0.20614 (0.04451)	0.15432 (0.04460)
	$\hat{\phi}_2$	0.23178 (0.04504)	0.23939 (0.04503)	0.26829 (0.04798)
	$\hat{\phi}_3$	0.13602 (0.04533)	0.14293 (0.04535)	0.22191 (0.05576)
	$\hat{\phi}_4$	0.12677 (0.04491)	0.13406 (0.04491)	0.20097 (0.07424)
	$\hat{\sigma}^2$	4.11069	4.14124	3.89693
	SQR	2038.90	2054.06	1932.88
	AIC	2121.24	2122.99	2092.24
	BIC	2140.27	2144.02	2113.27
10000	$\hat{\alpha}_1$	-	-	-1.99980
	$\hat{\alpha}_2$	-	-	0.00574
	$\hat{\alpha}_3$	-	-	2.00242
	intercepto	0.45013 (0.06496)	- -	- -
	$\hat{\phi}_1$	0.19151 (0.00994)	0.19647 (0.00994)	0.13913 (0.00981)
	$\hat{\phi}_2$	0.21024 (0.00997)	0.21514 (0.00997)	0.20565 (0.01044)
	$\hat{\phi}_3$	0.17583 (0.00997)	0.18073 (0.00997)	0.25897 (0.01246)
	$\hat{\phi}_4$	0.10577 (0.00994)	0.11074 (0.00994)	0.30878 (0.01653)
	$\hat{\sigma}^2$	4.23365	4.25266	3.95561
	SQR	42319.59	42509.58	39540.30
	AIC	42804.79	42847.61	42123.27
	BIC	42838.84	42883.66	42159.32

Em seguida, por meio das Figuras 3.19 e 3.20, verificamos o ajuste satisfatório dos dados a uma distribuição normal e percebemos que o aumento do tamanho da amostra melhora sobremaneira esse ajuste. Além disso, para os gráficos da FAC e FACP dos dados, constatamos um decaimento exponencial para zero na FAC, padrão esperado para um modelo autorregressivo; e, repararamos também, que os quatro primeiros *lags* da FACP são significativos, comportamento esperado para um modelo autorregressivo de ordem 4.

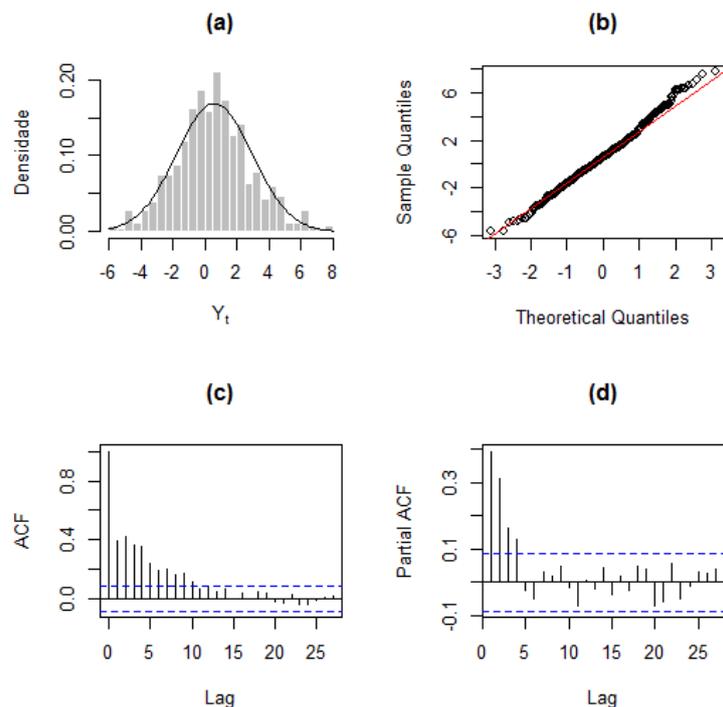


Figura 3.19: (a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n = 500$.

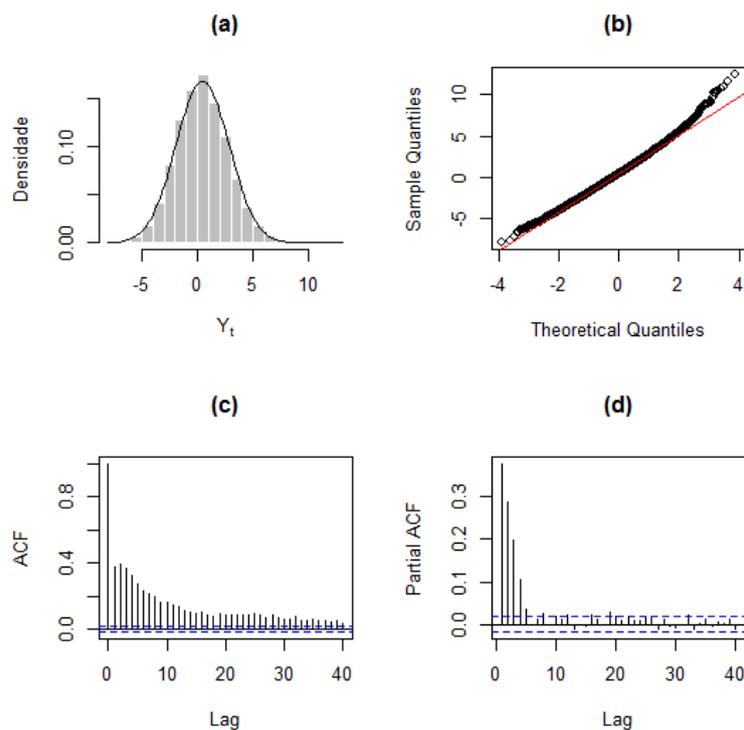


Figura 3.20: (a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n = 10000$.

Nas Figuras 3.21 e 3.22 constatamos os ajustes satisfatórios dos resíduos do modelo AR-MV(4), para $n = 500$ e $n = 10000$, à distribuição normal e, podemos notar também que o

aumento do tamanho da amostra melhora sobremaneira esse ajuste.

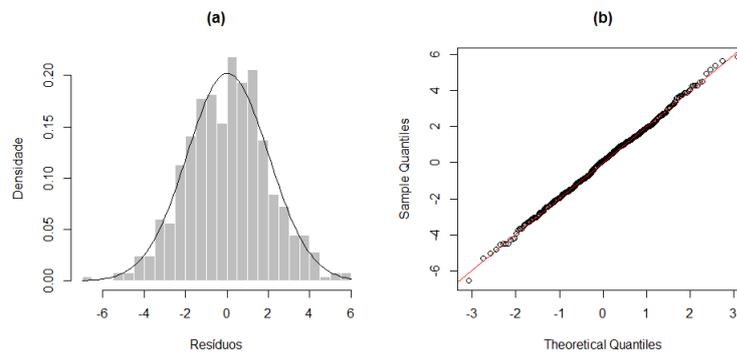


Figura 3.21: (a) Histograma e (b) gráfico Q-Q dos resíduos dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n = 500$.

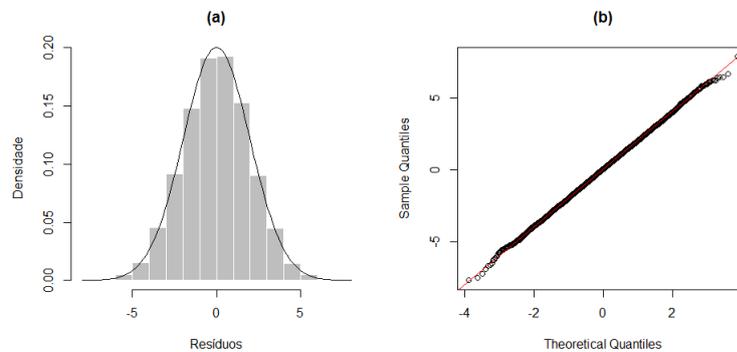


Figura 3.22: (a) Histograma e (b) gráfico Q-Q dos resíduos dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n = 10000$.

Ainda na fase de diagnóstico do modelo, as Figuras 3.23 e 3.24 apresentam os gráficos de dispersão, a FAC e a FACP dos resíduos do modelo AR-MV(4) para os dois tamanhos de amostra fixados. Por meio dos gráficos de dispersão, validamos a aleatoriedade dos resíduos e a inexistência de observações discrepantes. Além disso, os gráficos da FAC e FACP dos resíduos não apresentam defasagens altamente significativas, mostrando assim, a adequabilidade do ajuste do modelo.

A Tabela 3.2 mostra as previsões sob os modelos AR(4) (com e sem intercepto) e AR-MV(4) (métodos Monte Carlo e Esqueleto) a cinco passos à frente, para os dois tamanhos de amostra fixados. Notamos que, para o caso em que $n = 500$, o modelo AR(4) com intercepto mostrou uma capacidade preditiva superior ao AR-MV(4). No entanto, com o aumento do tamanho da amostra, o método de Monte Carlo, referente ao AR-MV(4), foi melhor.

Os gráficos constantes das Figuras 3.25 e 3.26 mostram os valores verdadeiros e as previsões pontuais dos modelos, para os dois tamanhos de amostra.

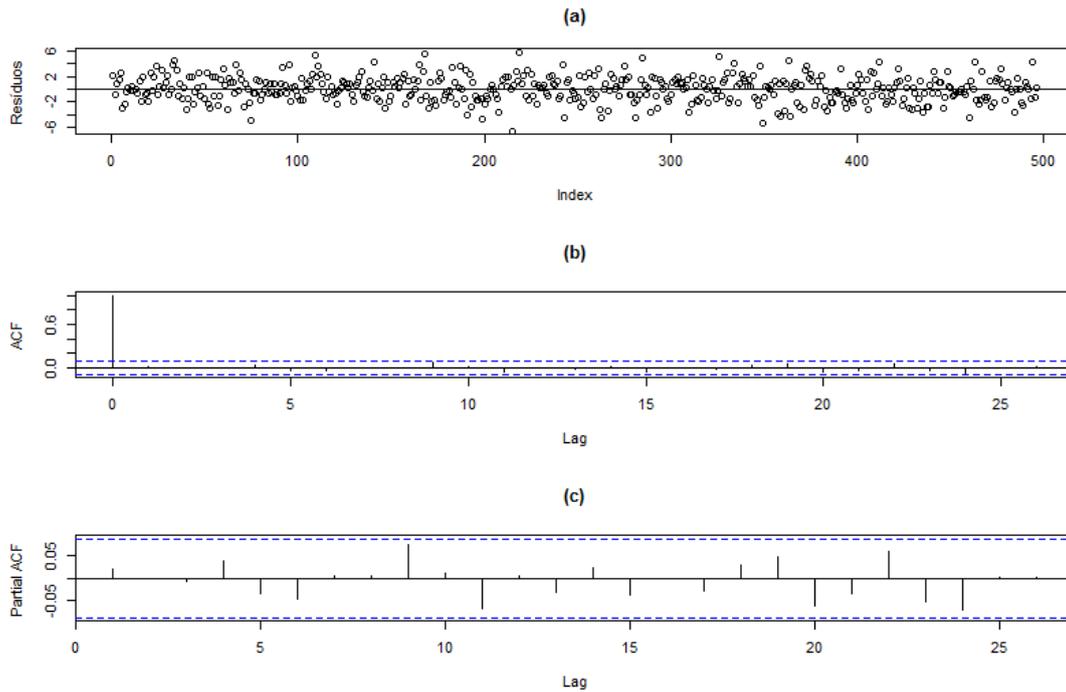


Figura 3.23: (a) *Dispersão*, (b) *FAC* e (c) *FACP* dos resíduos dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n = 500$.

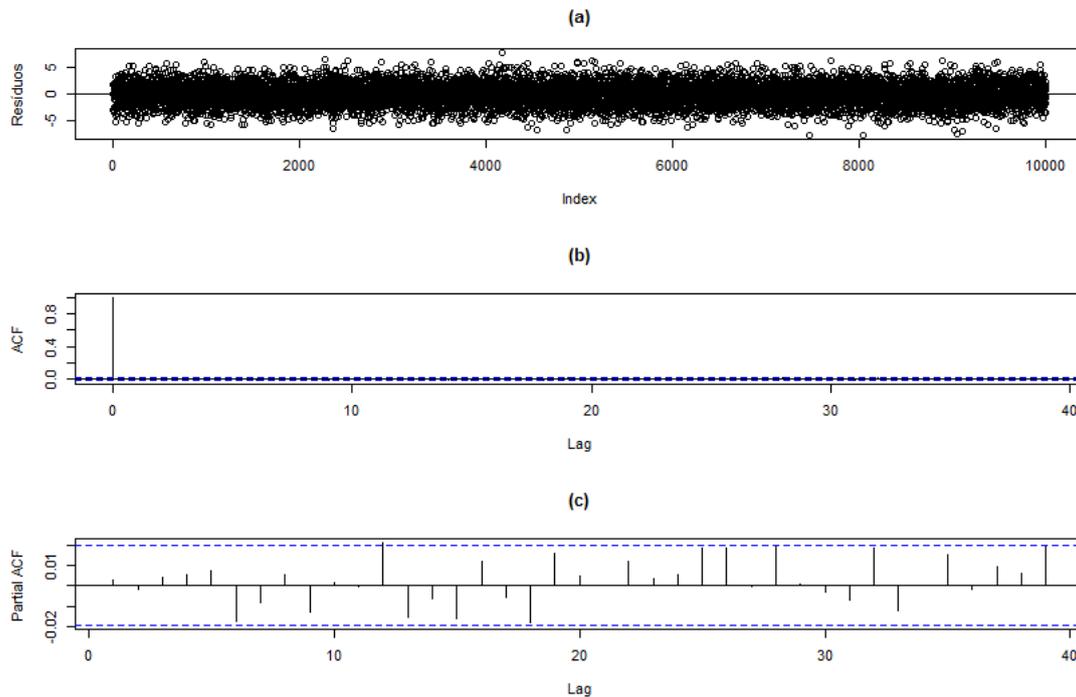


Figura 3.24: (a) *Dispersão*, (b) *FAC* e (c) *FACP* dos resíduos dos dados simulados para o Exemplo 3.9, com tamanho de amostra $n = 10000$.

3.5.2 AR-MV(5)

Exemplo 3.12. Os dados deste exemplo foram simulados a partir de um modelo $AR-MV(5)$, com parâmetros $p = 5$, $\phi_1 = 0.05$, $\phi_2 = 0.15$, $\phi_3 = 0.2$, $\phi_4 = 0.25$, $\phi_5 = 0.3$, $\alpha_1 = -2$,

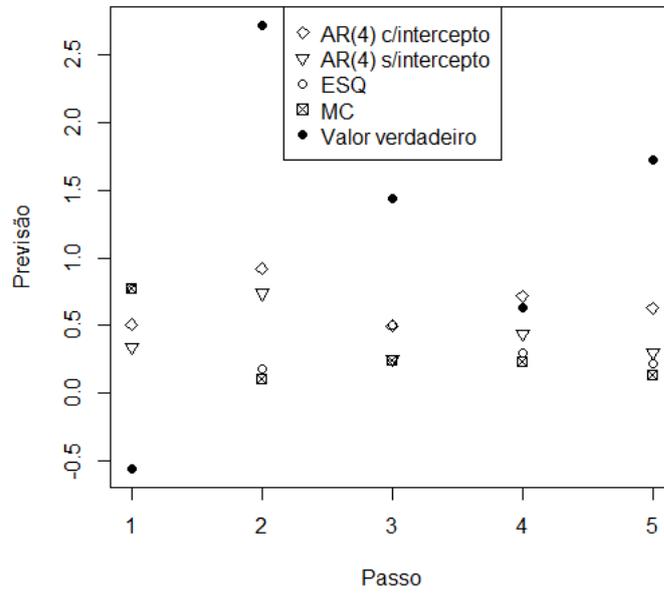


Figura 3.25: Valores observados e previsões pontuais, sob os modelos ajustados para os dados simulados no Exemplo 3.9, com origem em 500 para $n=500$.

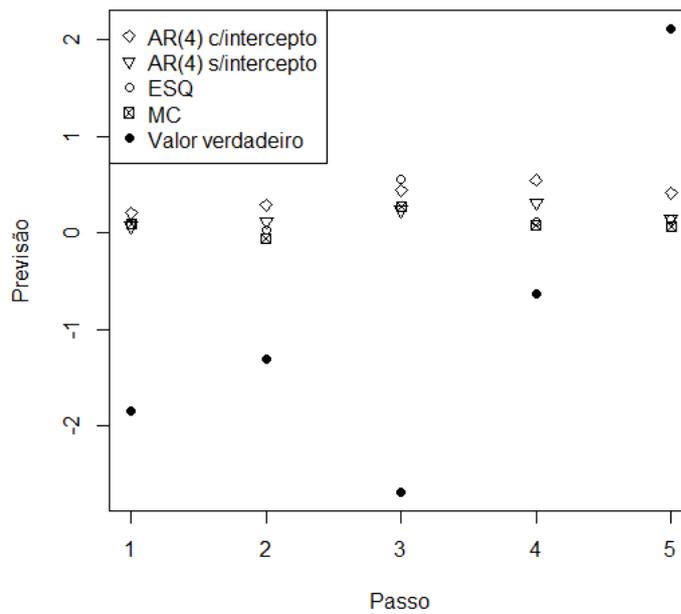


Figura 3.26: Valores observados e previsões pontuais, sob os modelos ajustados para os dados simulados no Exemplo 3.9, com origem em 10000 para $n=10000$.

Tabela 3.2: Estimativas do MAE, MQE e os erros de previsão a cada passo, sob os modelos $AR(4)$ (com e sem intercepto) e $AR-MV(4)$ para os dados simulados do Exemplo 3.9.

n	Passo	Valor verdadeiro	AR(4) (com intercepto)	AR(4) (sem intercepto)	AR-MV(4) (Esqueleto)	AR-MV(4) (Monte Carlo)
500	1	-0.56063	-1.06856	-0.90376	-1.33232	-1.33232
	2	2.72020	1.80067	1.98236	2.53729	2.61398
	3	1.43493	0.93712	1.18549	0.92916	1.18798
	4	0.62817	-0.08972	0.18764	0.32980	0.39209
	5	1.72403	1.09446	1.42205	1.50170	1.58905
	MAE	-	0.99811	1.13626	1.32605	1.42309
	MQE	-	1.29366	1.64186	2.28802	2.53962
10000	1	-1.84339	-2.05042	-1.91203	-1.93046	-1.93046
	2	-1.30561	-1.59347	-1.42482	-1.33108	-1.24971
	3	-2.68057	-3.11621	-2.91543	-3.23497	-2.95253
	4	-0.63859	-1.17915	-0.95069	-0.74351	-0.71122
	5	2.10783	1.69768	1.96684	1.97262	2.04700
	MAE	-	1.92738	1.83396	1.84253	1.77818
	MQE	-	4.14532	3.79159	4.08151	3.74038

$\alpha_2 = -1$, $\alpha_3 = 1$, $\alpha_4 = 2$ e $\sigma^2 = 4$. Note que $\sum_{i=1}^p |\phi_i| < 1$. O seguinte sistema de equações resume essas informações

$$Z_t = \begin{cases} 0.05Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq -2 \\ 0.05Z_{t-1} + 0.15Z_{t-2} + a_t, & \text{se } -2 < Z_{t-1} \leq -1 \\ 0.05Z_{t-1} + 0.15Z_{t-2} + 0.2Z_{t-3} + a_t, & \text{se } -1 < Z_{t-1} \leq 1 \\ 0.05Z_{t-1} + 0.15Z_{t-2} + 0.2Z_{t-3} + 0.25Z_{t-4} + a_t, & \text{se } 1 < Z_{t-1} \leq 2 \\ 0.05Z_{t-1} + 0.15Z_{t-2} + 0.2Z_{t-3} + 0.25Z_{t-4} + 0.3Z_{t-5} + a_t, & \text{se } Z_{t-1} > 2 \end{cases} \quad (3.32)$$

As Figuras 3.27 e 3.28 apresentam as séries temporais geradas pelo modelo com tamanho de amostra $n = 500$ e $n = 10000$, respectivamente. Podemos reparar um padrão estacionário nas duas séries, com melhores evidências para a Figura 3.28, em função do tamanho da amostra.

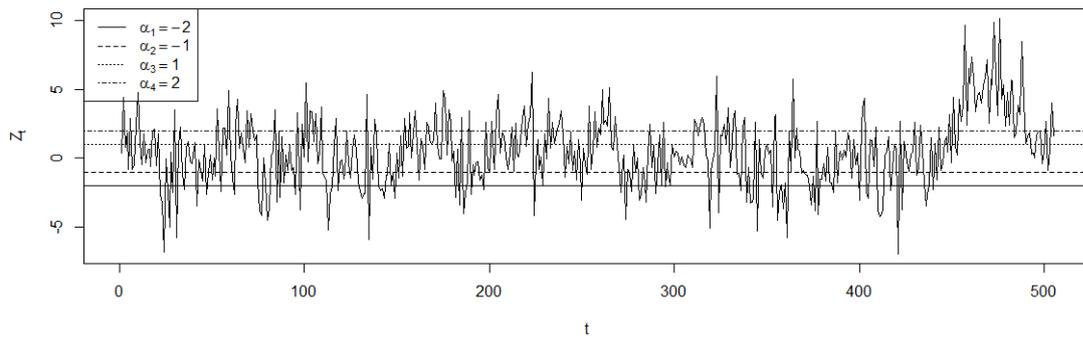


Figura 3.27: *Série temporal dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n=500$.*

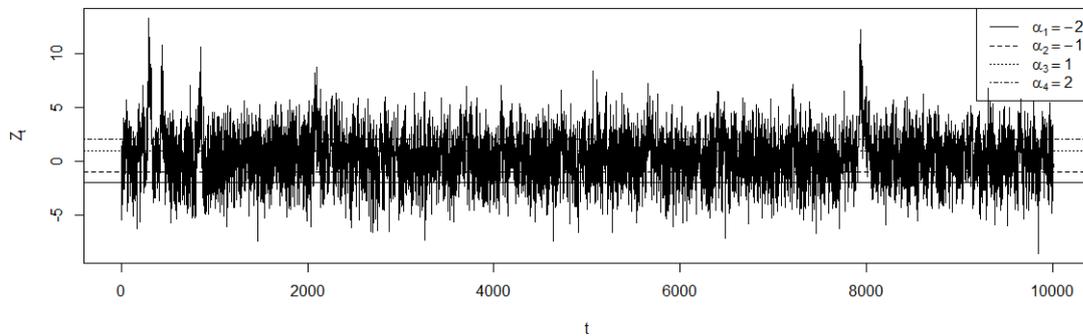


Figura 3.28: *Série temporal dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n=10000$.*

Na Tabela 3.3 estão descritos os valores das estimativas dos parâmetros dos modelos AR-MV(5) e AR(5) (com e sem intercepto), para dois tamanhos de amostra ($n = 500$) e ($n = 10000$). Como critérios de comparação entre os modelos fixamos a SQR, o AIC e o BIC e, em ambos os tamanhos de amostra, o modelo AR-MV foi superior em todos os critérios, mostrando sua adequabilidade aos dados. Podemos perceber também que à medida em que aumentamos o tamanho da amostra, as estimativas aproximam-se dos valores reais dos parâmetros.

Em seguida, por meio das Figuras 3.29 e 3.30, reconhecemos o ajuste satisfatório dos dados a uma distribuição normal e constatamos que o aumento do tamanho da amostra melhora sobremodo esse ajuste. Além disso, para os gráficos da FAC e FACP dos dados notamos um decaimento exponencial para zero na FAC, padrão esperado para um modelo autorregressivo; e observamos o comportamento esperado para um modelo autorregressivo de ordem 5 na FACP, porque os cinco primeiros *lags* são significativos.

Tabela 3.3: Valores estimados dos parâmetros dos modelos ajustados para os dados simulados do Exemplo 3.10. (O valor entre parênteses indica o erro-padrão de cada estimativa).

n	Parâmetros	AR(5)	AR(5)	AR-MV(5)
		(com intercepto)	(sem intercepto)	
500	$\hat{\alpha}_1$	-	-	-2.27636
	$\hat{\alpha}_2$	-	-	-0.92212
	$\hat{\alpha}_3$	-	-	1.26016
	$\hat{\alpha}_4$	-	-	2.23387
	intercepto	0.46823 (0.34153)	- -	- -
	$\hat{\phi}_1$	0.10630 (0.04468)	0.11001 (0.04467)	0.04267 (0.04244)
	$\hat{\phi}_2$	0.24305 (0.04448)	0.24698 (0.04445)	0.18393 (0.04808)
	$\hat{\phi}_3$	0.11114 (0.04559)	0.11460 (0.04560)	0.16586 (0.05059)
	$\hat{\phi}_4$	0.14946 (0.04444)	0.15308 (0.044435)	0.27436 (0.05826)
	$\hat{\phi}_5$	0.09788 (0.04482)	0.10105 (0.04483)	0.32487 (0.07223)
	$\hat{\sigma}^2$	5.06485	5.08247	4.56579
	SQR	2507.10	2515.82	2260.07
	AIC	2222.37	2222.14	2168.45
BIC	2245.60	2247.37	2193.68	
10000	$\hat{\alpha}_1$	-	-	-2.00034
	$\hat{\alpha}_2$	-	-	-1.01206
	$\hat{\alpha}_3$	-	-	1.00832
	$\hat{\alpha}_4$	-	-	1.98961
	intercepto	0.33066 (0.06193)	- -	- -
	$\hat{\phi}_1$	0.11571 (0.00995)	0.11868 (0.00995)	0.06976 (0.00979)
	$\hat{\phi}_2$	0.18279 (0.00996)	0.18574 (0.00996)	0.16940 (0.01057)
	$\hat{\phi}_3$	0.16037 (0.01000)	0.16327 (0.01000)	0.19000 (0.01128)
	$\hat{\phi}_4$	0.11066 (0.00996)	0.11363 (0.00996)	0.22772 (0.01516)
	$\hat{\phi}_5$	0.09652 (0.00996)	0.09950 (0.00996)	0.26799 (0.01780)
	$\hat{\sigma}^2$	4.28029	4.29176	4.01651
	SQR	42781.52	42896.19	40145.00
	AIC	42911.98	42936.77	42273.76
BIC	42953.24	42980.03	42317.02	

Nas Figuras 3.31 e 3.32 observamos os ajustes satisfatórios dos resíduos do modelo AR-MV(5), para $n = 500$ e $n = 10000$, à distribuição normal e notamos que o aumento do tamanho da amostra melhora esse ajuste.

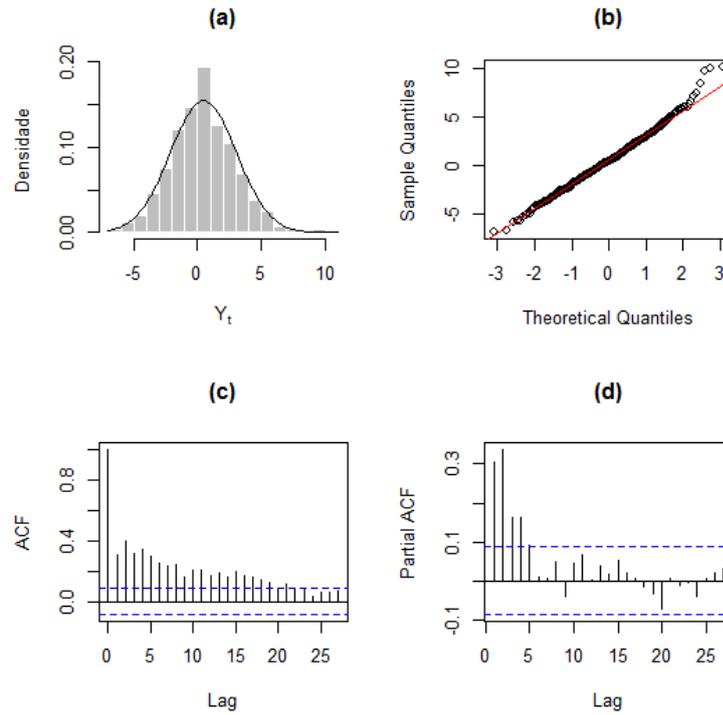


Figura 3.29: (a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n = 500$.

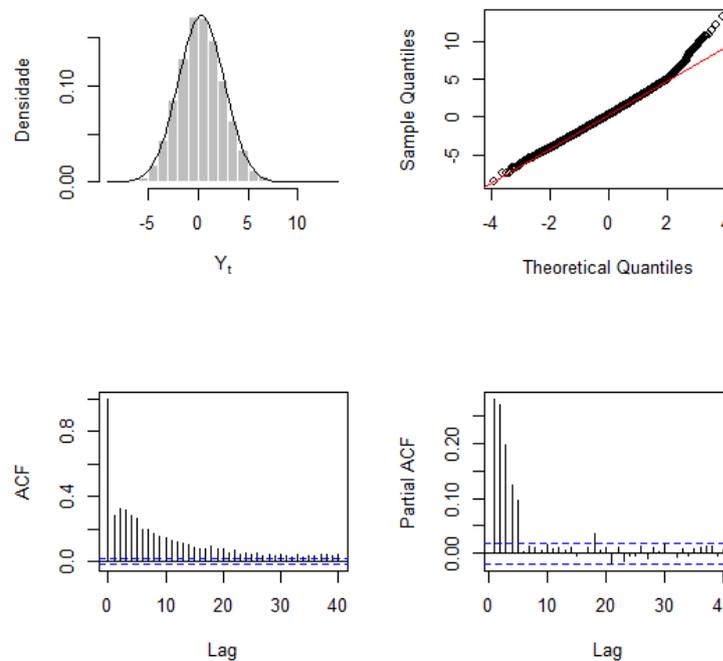


Figura 3.30: (a) Histograma, (b) gráfico Q-Q, (c) FAC e (d) FACP amostrais dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n = 10000$.

Ainda na fase de diagnóstico do modelo, as Figuras 3.33 e 3.34 apresentam os gráficos de dispersão, a FAC e a FACP dos resíduos do modelo AR-MV(5) para os dois tamanhos

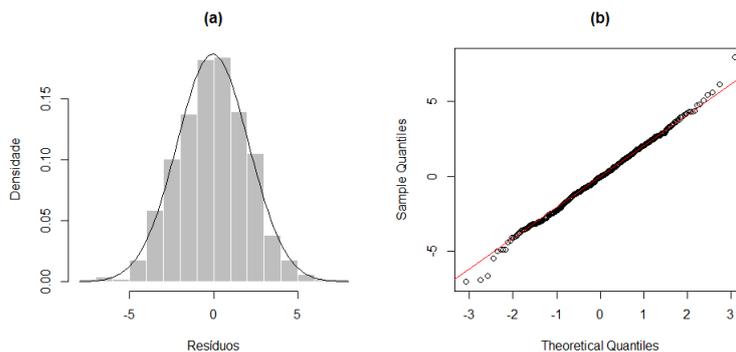


Figura 3.31: (a) Histograma e (b) gráfico Q-Q dos resíduos dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n = 500$.

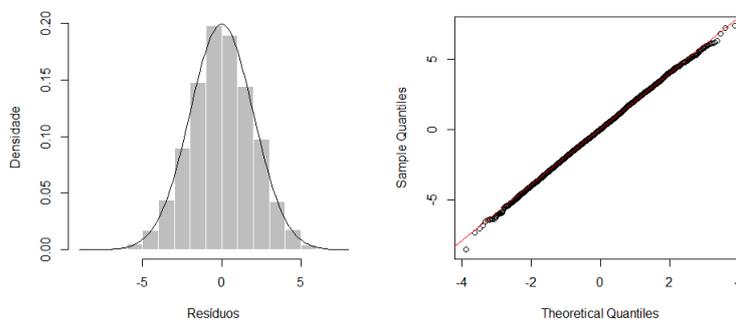


Figura 3.32: (a) Histograma e (b) gráfico Q-Q dos resíduos dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n = 10000$.

de amostra fixados. Por meio dos gráficos de dispersão, a aleatoriedade dos resíduos e a inexistência de observações discrepantes são constatadas. Além disso, os gráficos da FAC e FACP dos resíduos não apresentam defasagens altamente significativas, mostrando assim, a adequabilidade do ajuste do modelo.

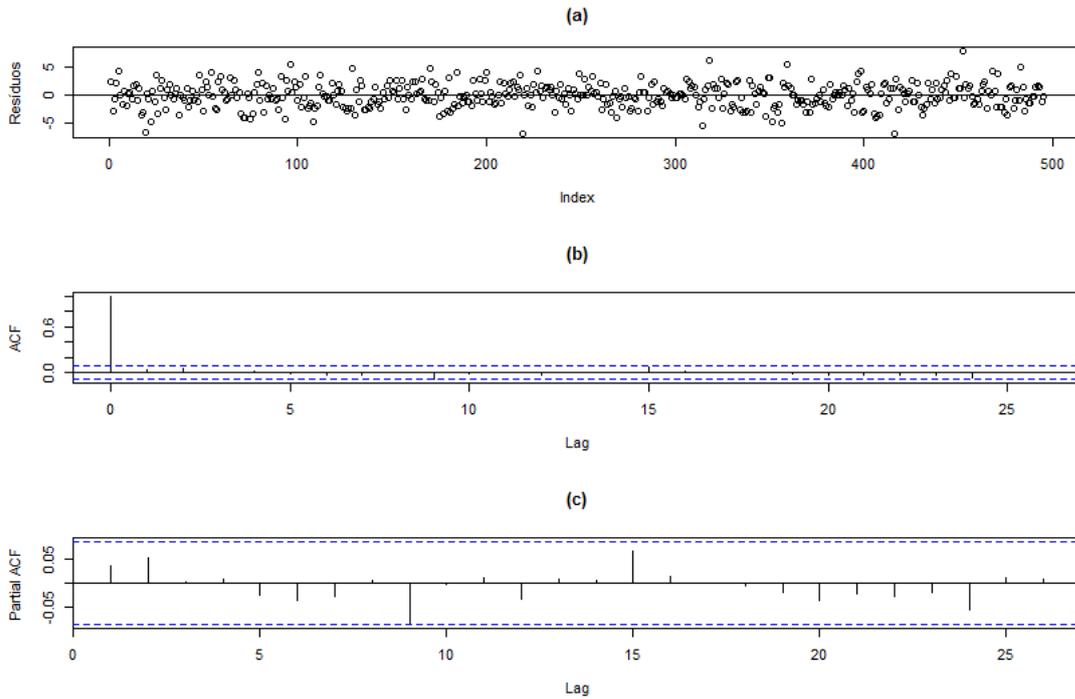


Figura 3.33: (a) *Dispersão*, (b) *FAC* e (c) *FACP* dos resíduos dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n = 500$.

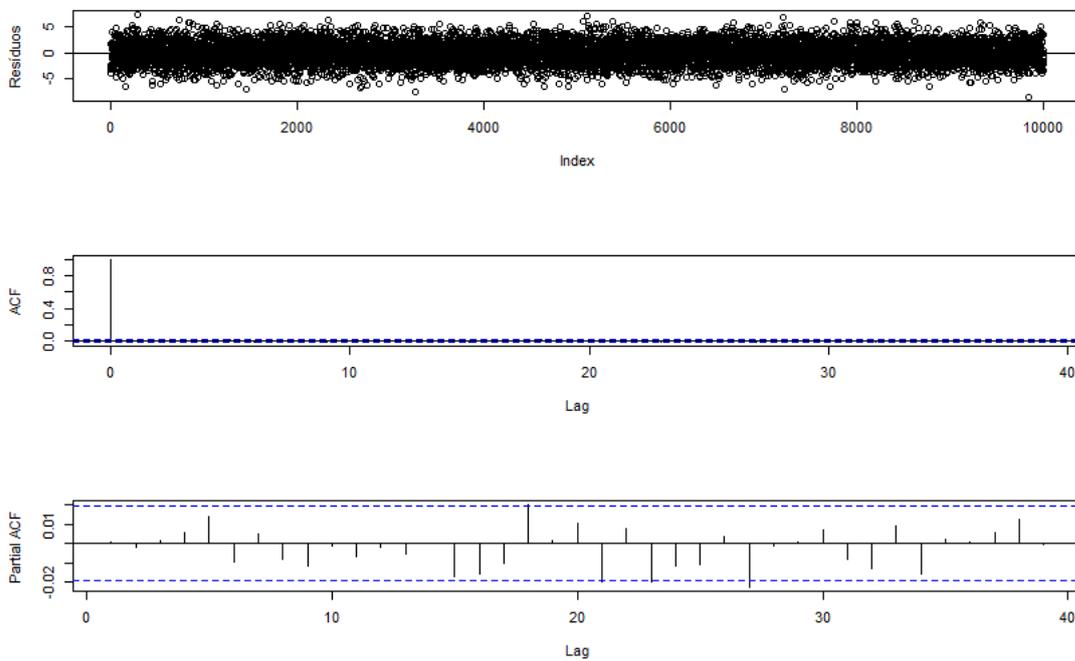


Figura 3.34: (a) *Dispersão*, (b) *FAC* e (c) *FACP* dos resíduos dos dados simulados para o Exemplo 3.10, com tamanho de amostra $n = 10000$.

A Tabela 3.4 mostra as previsões sob os modelos AR(5) (com e sem intercepto) e AR-MV(5) (métodos Monte Carlo e Esqueleto) a cinco passos à frente, para os dois tamanhos de amostra. Para o caso em que $n = 500$, o modelo AR(5) com intercepto mostra uma capacidade preditiva superior ao AR-MV(5). No entanto, com o aumento do tamanho da amostra, os métodos referentes ao AR-MV(5), foram melhores, em especial, o esqueleto.

Tabela 3.4: *Estimativas do MAE, MQE e os erros de previsão a cada passo, sob os modelos AR(5) (com e sem intercepto) e AR-MV(5) para os dados simulados no Exemplo 3.10.*

n	Passo	Valor verdadeiro	AR(5) (com intercepto)	AR(5) (sem intercepto)	AR-MV(5) (Esqueleto)	AR-MV(5) (Monte Carlo)
500	1	2.69395	1.93764	2.05536	2.41029	2.41029
	2	-0.90096	-1.66797	-1.53475	-0.92930	-0.86625
	3	0.69013	0.09916	0.26735	0.56560	0.38221
	4	4.02892	3.53169	3.72521	3.97135	4.05985
	5	1.66781	1.09425	1.31624	1.63775	1.47055
	MAE	-	1.66614	1.77978	1.90286	1.83783
	MQE	-	4.04332	4.45222	5.08937	5.07018
10000	1	-3.67153	-2.86802	-2.74186	-2.75186	-2.75186
	2	1.98416	2.78742	2.93027	2.86239	2.88574
	3	0.83983	1.64094	1.81375	1.23971	1.31096
	4	0.19482	0.88999	1.08822	0.54623	0.59550
	5	-0.57631	-0.14925	0.06647	-0.31718	-0.34086
	MAE	-	1.66712	1.72811	1.54348	1.57698
	MQE	-	3.90046	4.11651	3.54037	3.61792

Os gráficos constantes das Figuras 3.35 e 3.36 mostram os valores verdadeiros e as previsões pontuais dos modelos, para os dois tamanhos de amostra fixados.

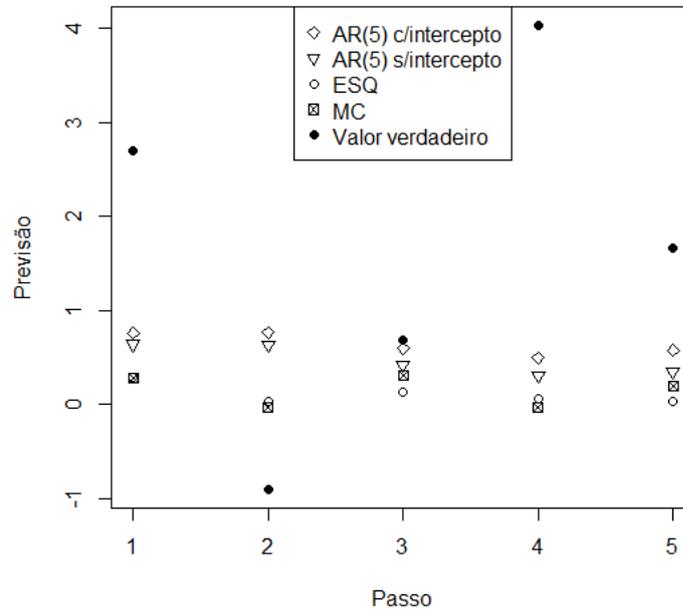


Figura 3.35: Valores observados e previsões pontuais, sob os modelos ajustados para os dados simulados no Exemplo 3.10, com origem em 500 para $n=500$.

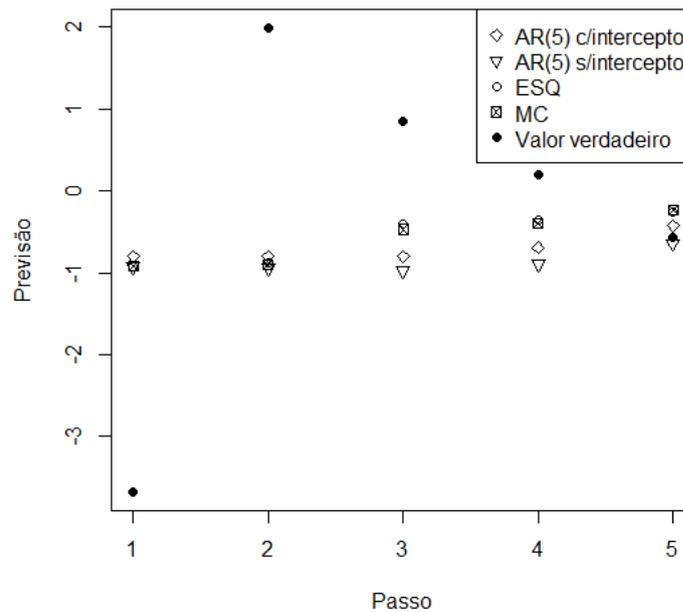


Figura 3.36: Valores observados e previsões pontuais, sob os modelos ajustados para os dados simulados no Exemplo 3.10, com origem em 10000 para $n=10000$.

3.6 Propriedades assintóticas

Nosso objetivo nesta seção é examinar as propriedades assintóticas dos estimadores dos modelos AR-MV(4) e AR-MV(5) por meio de estudo simulado. Suporemos os cenários em que α é conhecido e desconhecido e três tamanhos de amostra distintos $n = 500$, $n = 1000$ e $n = 5000$.

Em cada replicação será simulado um conjunto de observações de tamanho $n + 5$. No entanto, somente as n primeiras serão usadas para estimar os parâmetros dos modelos AR-MV e AR. As previsões são então geradas pelos dois modelos e os erros de previsão são reservados, além de serem calculadas as medidas MAE e MQE. No caso do método de Monte Carlo, utilizamos 1000 réplicas para cada iteração. Refizemos o procedimento para cada 1000 iterações e supomos as respectivas médias como estimativas das medidas MAE e MQE. Supomos ainda como estimativa para cada passo da previsão, a média dos quadrados dos erros de previsão para as 1000 iterações. Além disso, $a_t \sim N(0, 4)$.

Ao final do estudo pudemos concluir que o fato de α ser conhecido é muito relevante no processo de estimação de ϕ , pois quando o valor verdadeiro do parâmetro limiar é desconhecido, não observamos a normalidade nas estimativas de alguns dos parâmetros autorregressivos, especialmente, nos parâmetros ϕ de maior ordem. Apesar disso, a elevação do tamanho da amostra mostrou favorecer as estimativas dos parâmetros dos modelos.

3.6.1 AR-MV(4)

Exemplo 3.13. *Os dados deste exemplo foram simulados a partir de um modelo AR-MV(4), com parâmetros $p = 4$, $\phi_1 = 0.15$, $\phi_2 = 0.2$, $\phi_3 = 0.25$, $\phi_4 = 0.3$, $\alpha_1 = -2$, $\alpha_2 = 0$, $\alpha_3 = 2$ e $\sigma^2 = 4$.*

Na Tabela 3.5 estão dispostas as principais estatísticas descritivas referentes às estimativas dos parâmetros do modelo AR-MV(4) sob análise, dando importância ao fato de termos o conhecimento prévio ou não de α , e tomando três tamanhos de amostra distintos, quais sejam, 500, 1000 e 5000.

Ao observarmos as estatísticas média e mediana das estimativas, percebemos que o aumento do tamanho da amostra interfere diretamente na qualidade da estimativa do parâmetro em questão, ou seja, à medida em que aumentamos o tamanho da amostra, obtemos estimativas mais próximas dos valores verdadeiros estipulados na descrição deste exemplo. Além do mais, o aumento do tamanho da amostra tem influência direta na redução das estimativas dos desvios-padrão.

Um outro fator que interfere na qualidade das estimativas de ϕ é o conhecimento prévio de α . Notamos que há significativa melhora nas estimativas dos parâmetros autorregressivos do modelo, quando conhecemos de antemão o valor verdadeiro de α , em especial, a estimativa do parâmetro ϕ_4 (parâmetro autorregressivo de maior ordem).

A Tabela 3.6 descreve os erros de previsão dos modelos sob comparação, quais sejam,

Tabela 3.5: Estimativas das estatísticas descritivas dos estimadores do modelo do Exemplo 3.11, para os casos α conhecido e desconhecido para $n=500, 1000$ e 5000 .

n	Estimativa	Mediana	Média	Desvio-padrão
500	$\hat{\alpha}_1$	-1.9410	-1.8870	1.6757
	$\hat{\alpha}_2$	0.0331	-0.2031	1.5804
	$\hat{\alpha}_3$	0.6342	0.1967	2.1587
	$\hat{\phi}_1$	0.1570	0.1573	0.0462
	$\hat{\phi}_1 \alpha$	0.1495	0.1497	0.0437
	$\hat{\phi}_2$	0.2074	0.2085	0.0542
	$\hat{\phi}_2 \alpha$	0.1973	0.1941	0.0467
	$\hat{\phi}_3$	0.2558	0.2538	0.0681
	$\hat{\phi}_3 \alpha$	0.2468	0.2448	0.0564
	$\hat{\phi}_4$	0.2033	0.1952	0.3315
	$\hat{\phi}_4 \alpha$	0.2855	0.2815	0.0810
	1000	$\hat{\alpha}_1$	-1.9680	-1.8760
$\hat{\alpha}_2$		0.0429	-0.0010	0.8098
$\hat{\alpha}_3$		1.0400	0.5944	1.5989
$\hat{\phi}_1$		0.1603	0.1599	0.0344
$\hat{\phi}_1 \alpha$		0.1495	0.1499	0.0317
$\hat{\phi}_2$		0.2065	0.2079	0.0351
$\hat{\phi}_2 \alpha$		0.1989	0.1987	0.0320
$\hat{\phi}_3$		0.2576	0.2580	0.0499
$\hat{\phi}_3 \alpha$		0.2476	0.2478	0.0397
$\hat{\phi}_4$		0.2242	0.2219	0.0851
$\hat{\phi}_4 \alpha$		0.2901	0.2895	0.0534
5000		$\hat{\alpha}_1$	-2.0050	-2.0450
	$\hat{\alpha}_2$	-0.0003	0.0171	0.1984
	$\hat{\alpha}_3$	1.9730	1.4320	1.0432
	$\hat{\phi}_1$	0.1541	0.1554	0.0182
	$\hat{\phi}_1 \alpha$	0.1490	0.1491	0.0139
	$\hat{\phi}_2$	0.2029	0.2027	0.0163
	$\hat{\phi}_2 \alpha$	0.2011	0.2003	0.0151
	$\hat{\phi}_3$	0.2532	0.2545	0.0208
	$\hat{\phi}_3 \alpha$	0.2491	0.2493	0.0176
	$\hat{\phi}_4$	0.2892	0.2639	0.0667
	$\hat{\phi}_4 \alpha$	0.2977	0.2981	0.0233

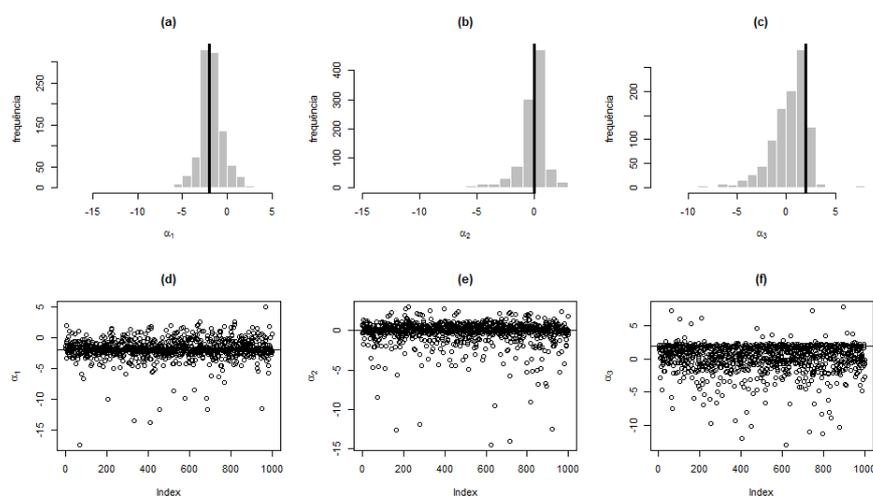
AR(4) e AR-MV(4). Percebemos que a ação de aumentar o tamanho da amostra, neste caso, não parece reduzir os erros de previsão dos modelos, pelo contrário, erros menores podem ser observados quando $n = 500$. Além disso, por meio das estimativas das MAE e MQE, notamos que, nos três casos, os métodos atinentes ao modelo AR-MV(4) foram superiores ao modelo AR(4), conforme destacado em negrito.

A seguir, por meio das Figuras 3.37 a 3.45, ilustramos os dados apresentados na Tabela 3.5, para os três tamanhos de amostra fixados, respectivamente. Nas Figuras 3.37, 3.40 e 3.43 visualizamos os histogramas e dispersões para os parâmetros limiares do modelo,

Tabela 3.6: Estimativas do MAE, MQE e a média dos quadrados dos erros a cada passo para o Exemplo 3.11.

n	Passo	AR(4) (com intercepto)	AR(4) (sem intercepto)	AR-MV(4) (Esqueleto)	AR-MV(4) (Monte Carlo)
500	1	4.18675	4.19439	4.05291	4.05291
	2	4.30619	4.31654	4.22833	4.22819
	3	4.53727	4.56197	4.43443	4.39868
	4	4.81408	4.88278	4.70448	4.68878
	5	5.24102	5.13211	5.04961	5.04177
	MAE	1.70419	1.70258	1.67921	1.67950
	MQE	4.61706	4.61756	4.49395	4.48207
1000	1	4.37704	4.39913	4.13484	4.13484
	2	4.56335	4.61401	4.55485	4.55655
	3	4.98556	5.04358	5.00539	4.91526
	4	4.92419	5.01280	4.77308	4.69812
	5	5.07636	5.03651	5.16835	5.05050
	MAE	1.74233	1.74886	1.72880	1.72107
	MQE	4.78530	4.82120	4.72730	4.67105
5000	1	4.39665	4.40142	4.17880	4.17880
	2	4.52244	4.53447	4.38893	4.39916
	3	4.84504	4.86300	4.73502	4.71018
	4	5.53123	5.55019	5.26470	5.34573
	5	5.62923	5.35310	5.46249	5.45138
	MAE	1.77113	1.76190	1.73782	1.74116
	MQE	4.98492	4.94044	4.80599	4.81705

variando o tamanho da amostra. Reparamos que, com o aumento da amostra, as estimativas aproximam-se do valor real dos parâmetros e que o número de *outliers* é reduzido. Contudo, para o parâmetro α_3 , há uma excessiva quantidade de estimativas menores que o valor real do parâmetro, ou seja, o referido parâmetro é demasiadamente subestimado.

**Figura 3.37:** (a), (b) e (c) Histogramas de $\hat{\alpha}_1$, $\hat{\alpha}_2$ e $\hat{\alpha}_3$ e (d), (e) e (f) dispersões de $\hat{\alpha}_1$, $\hat{\alpha}_2$ e $\hat{\alpha}_3$ para o Exemplo 3.11 com $n=500$.

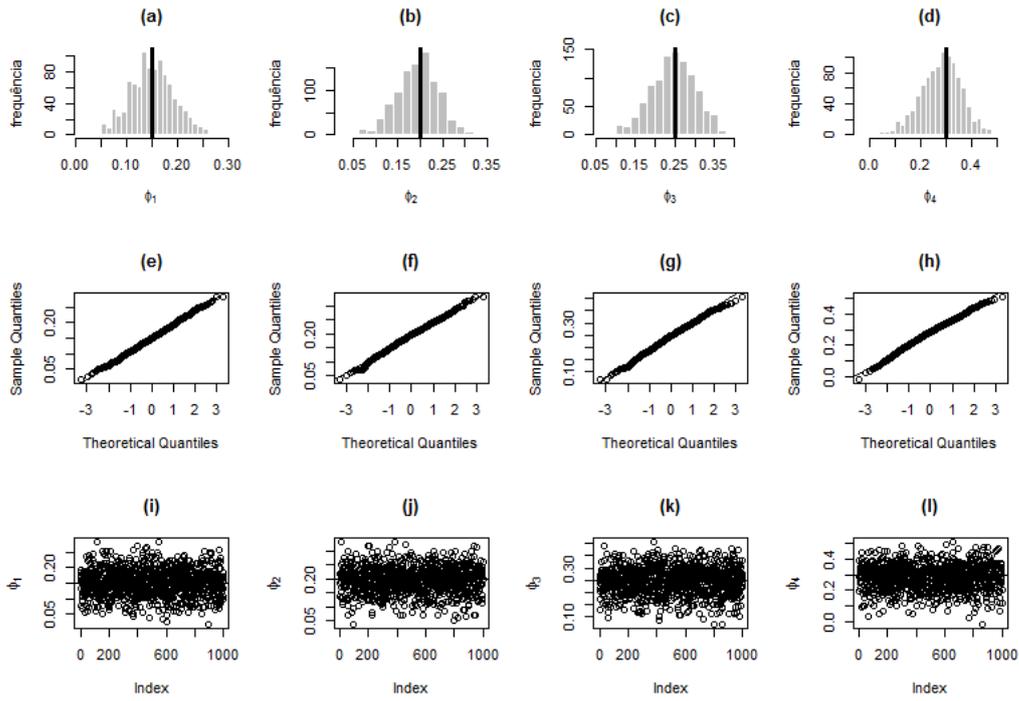


Figura 3.38: (a), (b), (c) e (d) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$, (e), (f), (g) e (h) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ e (i), (j), (k) e (l) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ para o Exemplo 3.11 com $n=500$, quando α é conhecido.

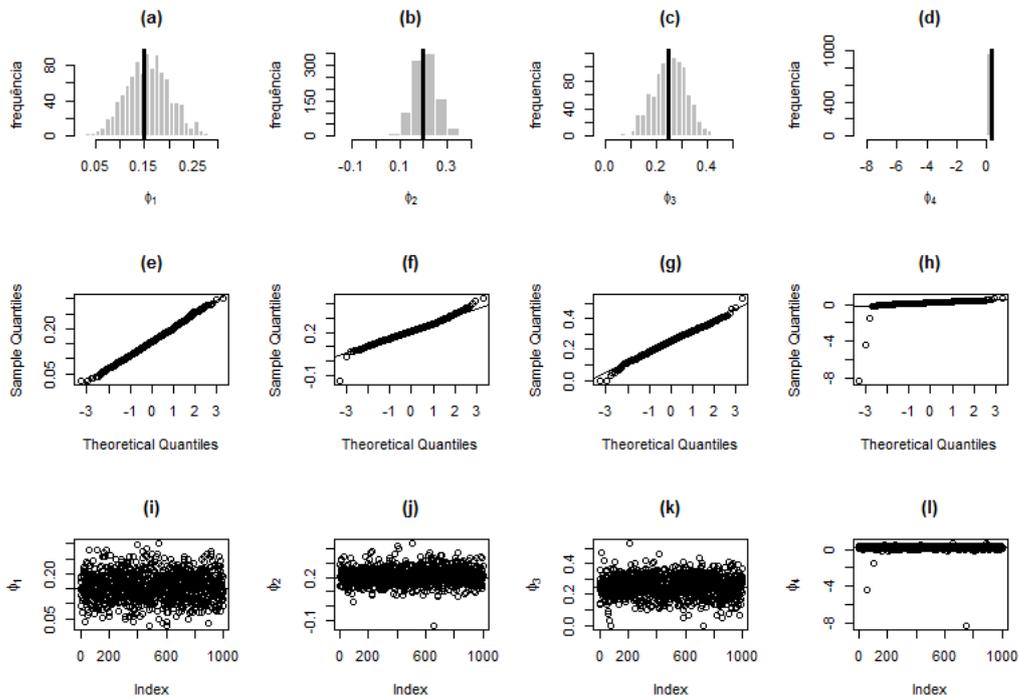


Figura 3.39: (a), (b), (c) e (d) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$, (e), (f), (g) e (h) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ e (i), (j), (k) e (l) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ e $\hat{\phi}_4$ para o Exemplo 3.11 com $n=500$, quando α é desconhecido.

Em relação às Figuras 3.38, 3.41 e 3.44, visualizamos os histogramas, os gráficos Q-Q e as dispersões para os parâmetros autorregressivos do modelo, quando o parâmetro α é conhecido, variando o tamanho da amostra. Observamos o bom ajuste das estimativas do modelo à distribuição normal, já no caso onde $n = 500$.

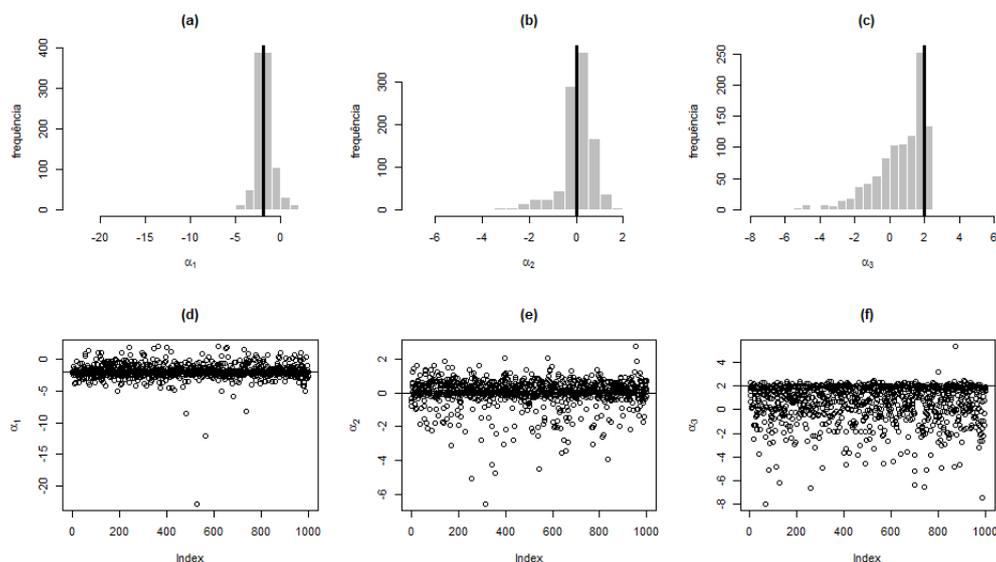


Figura 3.40: (a), (b) e (c) Histogramas de $\hat{\alpha}_1$, $\hat{\alpha}_2$ e $\hat{\alpha}_3$ e (d), (e) e (f) dispersões de $\hat{\alpha}_1$, $\hat{\alpha}_2$ e $\hat{\alpha}_3$ para o Exemplo 3.11 com $n=1000$.

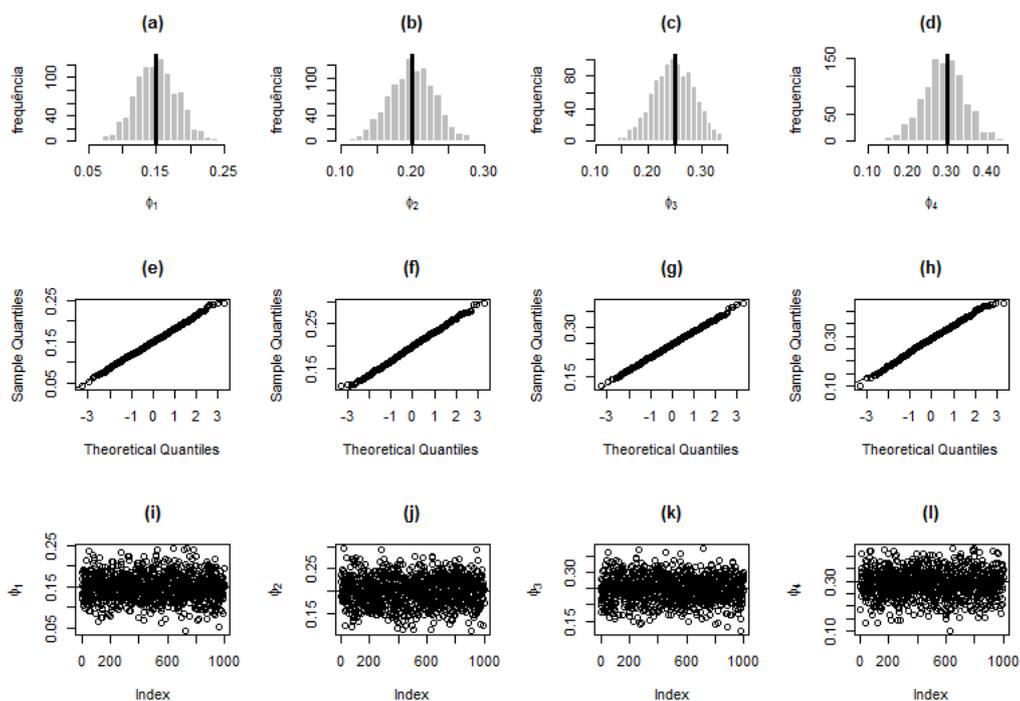


Figura 3.41: (a), (b), (c) e (d) Histogramas de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ e $\hat{\phi}_4$, (e), (f), (g) e (h) gráficos Q-Q de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ e $\hat{\phi}_4$ e (i), (j), (k) e (l) dispersões de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ e $\hat{\phi}_4$ para o Exemplo 3.11 com $n=1000$, quando α é conhecido.

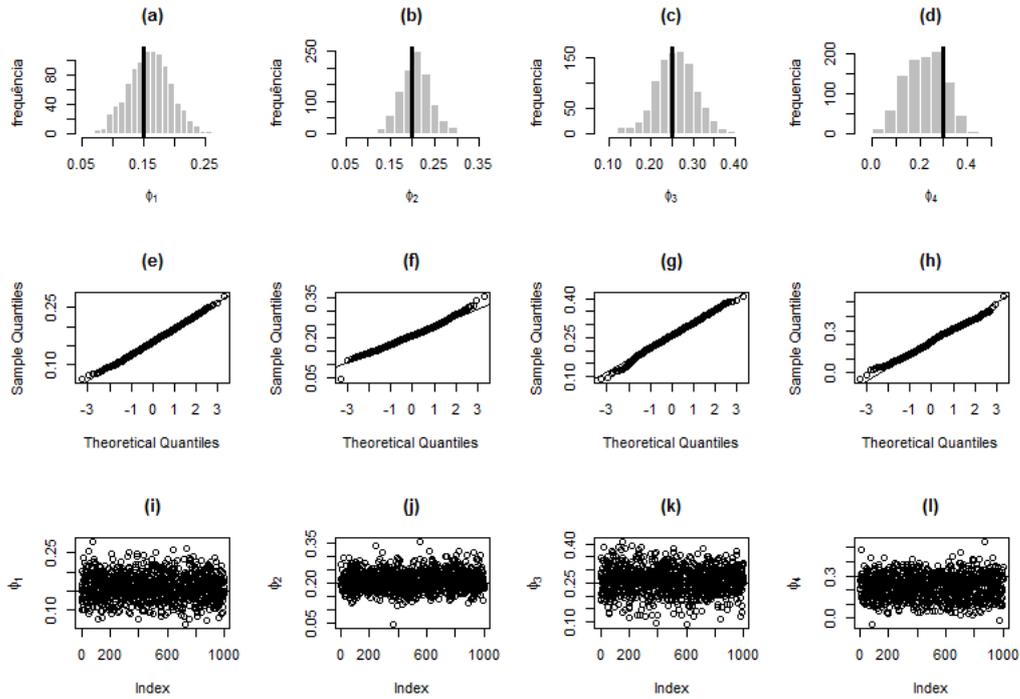


Figura 3.42: (a), (b), (c) e (d) Histogramas de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ e $\hat{\phi}_4$, (e), (f), (g) e (h) gráficos Q-Q de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ e $\hat{\phi}_4$ e (i), (j), (k) e (l) dispersões de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ e $\hat{\phi}_4$ para o Exemplo 3.11 com $n=1000$, considerando α desconhecido.

No que diz respeito às Figuras 3.39, 3.42 e 3.45, visualizamos os histogramas, os gráficos Q-Q e as dispersões para os parâmetros autorregressivos do modelo, quando o parâmetro α é desconhecido, variando o tamanho da amostra. Percebemos que à proporção em que se eleva o tamanho da amostra, os gráficos apresentam uma forma mais adequada à distribuição normal. Todavia, as estimativas para o parâmetro autorregressivo de maior ordem (ϕ_4) não se mostram apropriadas à distribuição normal, mesmo no caso onde $n = 5000$.

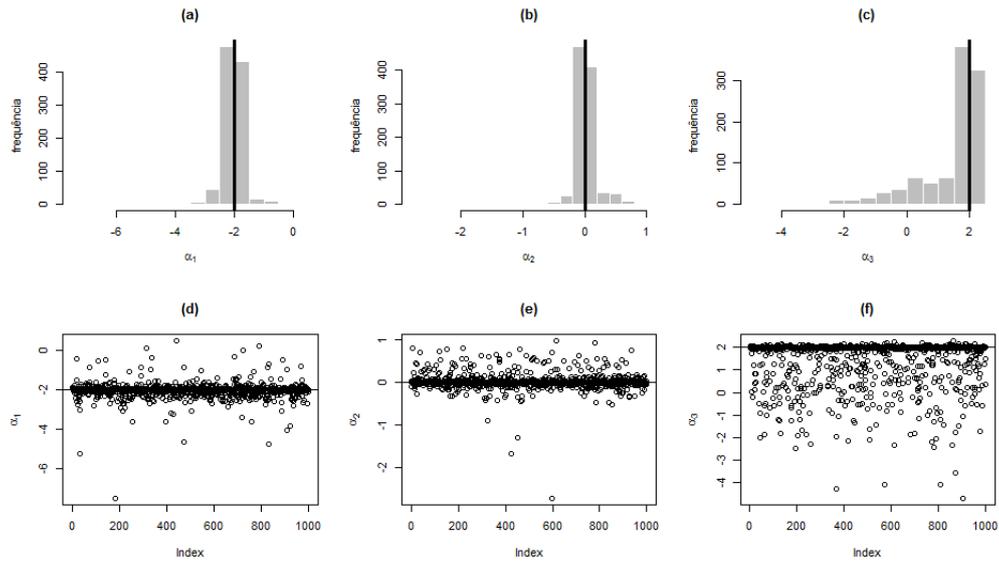


Figura 3.43: (a), (b) e (c) Histogramas de $\hat{\alpha}_1$, $\hat{\alpha}_2$ e $\hat{\alpha}_3$ e (d), (e) e (f) dispersões de $\hat{\alpha}_1$, $\hat{\alpha}_2$ e $\hat{\alpha}_3$ para o Exemplo 3.11 com $n=5000$.

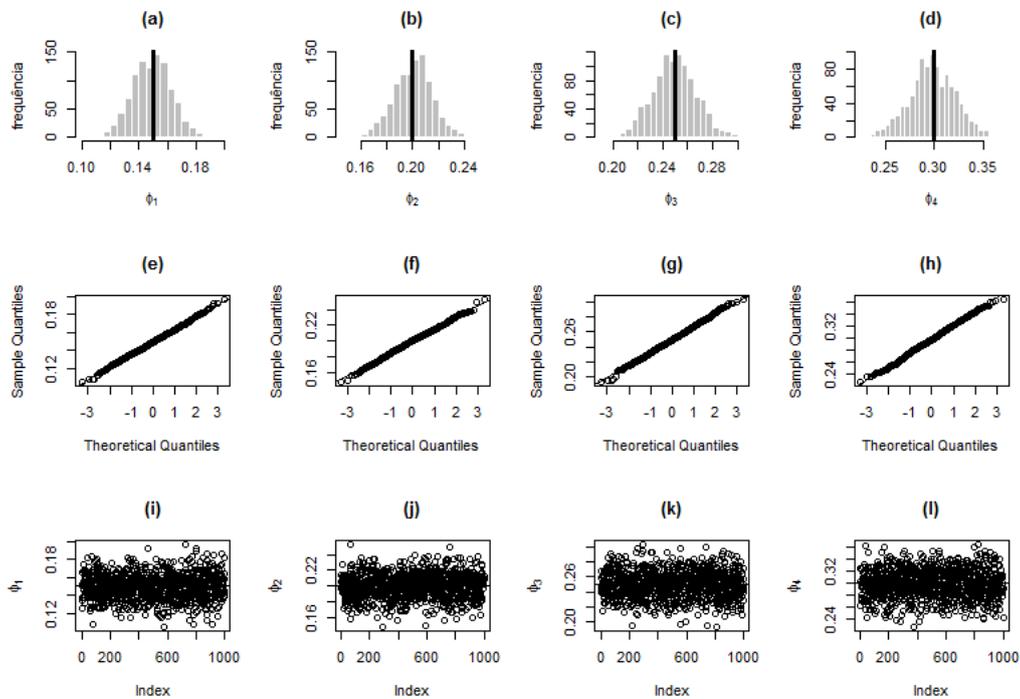


Figura 3.44: (a), (b), (c) e (d) Histogramas de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ e $\hat{\phi}_4$, (e), (f), (g) e (h) gráficos Q-Q de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ e $\hat{\phi}_4$ e (i), (j), (k) e (l) dispersões de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ e $\hat{\phi}_4$ para o Exemplo 3.11 com $n=5000$, quando α é conhecido.

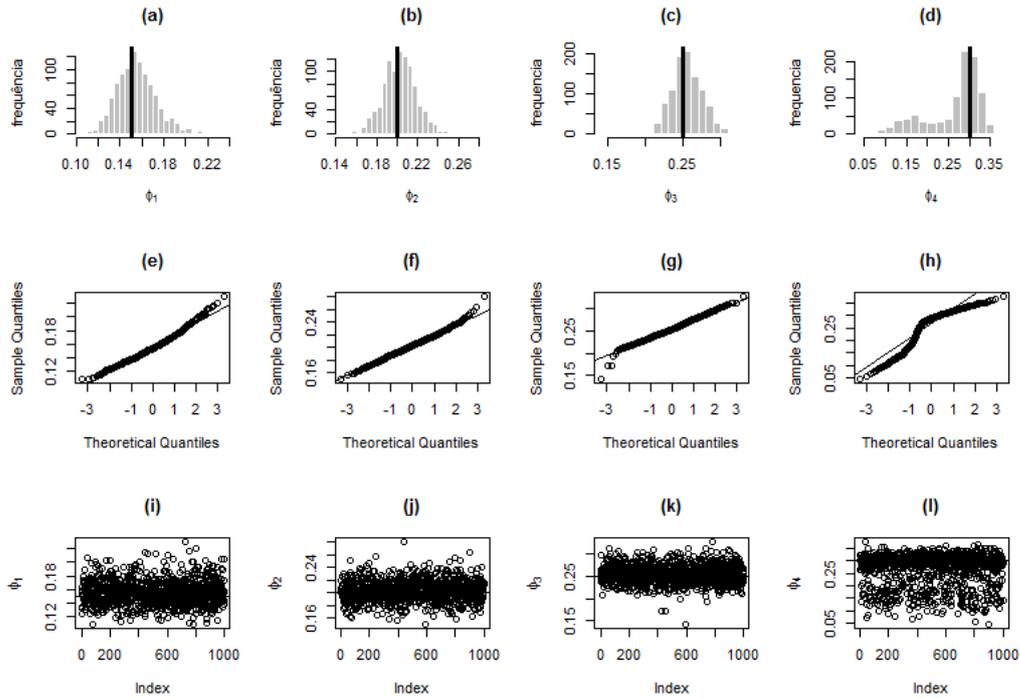


Figura 3.45: (a), (b), (c) e (d) Histogramas de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ e $\hat{\phi}_4$, (e), (f), (g) e (h) gráficos Q-Q de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ e $\hat{\phi}_4$ e (i), (j), (k) e (l) dispersões de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ e $\hat{\phi}_4$ para o Exemplo 3.11 com $n=5000$, quando α é desconhecido.

3.6.2 AR-MV(5)

Exemplo 3.14. Os dados deste exemplo foram simulados a partir de um modelo AR-MV(5), com parâmetros $p = 5$, $\phi_1 = 0.05$, $\phi_2 = 0.15$, $\phi_3 = 0.2$, $\phi_4 = 0.25$, $\phi_5 = 0.3$, $\alpha_1 = -2$, $\alpha_2 = -1$, $\alpha_3 = 1$, $\alpha_4 = 2$ e $\sigma^2 = 4$.

Na Tabela 3.7 estão dispostas as principais estatísticas descritivas referentes às estimativas dos parâmetros do modelo AR-MV(5) sob análise, ressaltando a questão do conhecimento prévio ou não de α , e fixando três tamanhos de amostra distintos, quais sejam, 500, 1000 e 5000.

Tabela 3.7: Estimativas das estatísticas descritivas dos estimadores do modelo do Exemplo 3.12, para os casos α conhecido e desconhecido para $n=500$, 1000 e 5000.

n	Estimativa	Mediana	Média	Desvio-padrão
	$\hat{\alpha}_1$	-1.5530	-1.4350	2.0048
	$\hat{\alpha}_2$	-1.0980	-1.2860	1.7655
	$\hat{\alpha}_3$	0.9523	0.3631	1.8308
	$\hat{\alpha}_4$	1.2200	0.4206	2.5309
	$\hat{\phi}_1$	0.0506	0.0501	0.0466
	$\hat{\phi}_1 \alpha$	0.0485	0.0475	0.0457

Continua na próxima página

Tabela 3.7 – *Continuação da página anterior*

n	Estimativa	Mediana	Média	Desvio-padrão
500	$\hat{\phi}_2$	0.1642	0.1668	0.0586
	$\hat{\phi}_2 \alpha$	0.1446	0.1442	0.0469
	$\hat{\phi}_3$	0.2024	0.2067	0.1114
	$\hat{\phi}_3 \alpha$	0.1949	0.1946	0.0494
	$\hat{\phi}_4$	0.2366	0.2259	0.1619
	$\hat{\phi}_4 \alpha$	0.2388	0.2374	0.0662
500	$\hat{\phi}_5$	0.2242	0.2258	0.1330
	$\hat{\phi}_5 \alpha$	0.2827	0.2859	0.0846
	$\hat{\alpha}_1$	-1.8260	-1.5340	1.5017
	$\hat{\alpha}_2$	-1.0940	-1.1970	0.9372
	$\hat{\alpha}_3$	0.9967	0.6067	1.3284
	$\hat{\alpha}_4$	1.7780	0.9701	1.7018
1000	$\hat{\phi}_1$	0.0565	0.0561	0.0322
	$\hat{\phi}_1 \alpha$	0.0525	0.0513	0.0306
	$\hat{\phi}_2$	0.1633	0.1642	0.0401
	$\hat{\phi}_2 \alpha$	0.1484	0.1483	0.0344
	$\hat{\phi}_3$	0.1997	0.2002	0.0382
	$\hat{\phi}_3 \alpha$	0.1974	0.1972	0.0357
1000	$\hat{\phi}_4$	0.2419	0.2360	0.0707
	$\hat{\phi}_4 \alpha$	0.2396	0.2404	0.0480
	$\hat{\phi}_5$	0.2492	0.2363	0.0957
	$\hat{\phi}_5 \alpha$	0.2827	0.2847	0.0537
	$\hat{\alpha}_1$	-1.9470	-1.7690	0.8160
	$\hat{\alpha}_2$	-1.0090	-1.0730	0.3023
5000	$\hat{\alpha}_3$	1.0160	1.0200	0.4992
	$\hat{\alpha}_4$	1.9980	1.7110	0.9734
	$\hat{\phi}_1$	0.0520	0.0523	0.0155
	$\hat{\phi}_1 \alpha$	0.0505	0.0502	0.0139
	$\hat{\phi}_2$	0.1546	0.1543	0.0161
	$\hat{\phi}_2 \alpha$	0.1501	0.1498	0.0144
5000	$\hat{\phi}_3$	0.2013	0.2007	0.0182
	$\hat{\phi}_3 \alpha$	0.2005	0.1997	0.0160
	$\hat{\phi}_4$	0.2523	0.2516	0.0330
	$\hat{\phi}_4 \alpha$	0.2474	0.2479	0.0205
	$\hat{\phi}_5$	0.2938	0.2785	0.0591
	$\hat{\phi}_5 \alpha$	0.2971	0.2964	0.0241

Ao analisarmos as estatísticas média e mediana das estimativas, reparamos que o aumento do tamanho da amostra interfere diretamente na qualidade da estimativa do parâmetro em questão, ou seja, à proporção em que aumentamos o tamanho da amostra, obtemos estimativas mais próximas dos valores reais descritos neste exemplo. Além do que, o aumento do tamanho da amostra tem influência direta na redução das estimativas dos desvios-padrão.

Uma outra questão com ingerência na qualidade das estimativas de ϕ é o conhecimento antecipado de α . Constatamos que há significativa melhora nas estimativas dos parâmetros autorregressivos do modelo quando conhecemos de antemão o valor verdadeiro de α , principalmente, a estimativa do parâmetro autorregressivo de maior ordem (ϕ_5).

A Tabela 3.8 descreve os erros de previsão dos modelos equiparados, quais sejam, AR(5) e AR-MV(5). Percebemos que o aumento do tamanho da amostra, neste caso, não parece ser um fator de redução dos erros de previsão dos modelos, pois reparamos que os menores erros são encontrados quando $n = 1000$. Ademais, por meio das estimativas das MAE e MQE, notamos que nos três casos, os métodos atinentes ao modelo AR-MV(5) foram superiores ao modelo AR(5), com destaque para o método de Monte Carlo, melhor nos três cenários analisados, conforme destacado em negrito.

Tabela 3.8: Estimativas do MAE, MQE e a média dos quadrados dos erros a cada passo para o Exemplo 3.12.

n	Passo	AR(5) (com intercepto)	AR(5) (sem intercepto)	AR-MV(5) (Esqueleto)	AR-MV(5) (Monte Carlo)
500	1	4.62192	4.65190	4.49878	4.49878
	2	4.09433	4.11146	4.08435	4.09984
	3	4.44681	4.46461	4.53875	4.42293
	4	4.74480	4.78849	4.98735	4.87385
	5	5.17991	5.10122	5.11930	5.08935
	MAE	1.71440	1.71370	1.71191	1.70530
	MQE	4.61756	4.62354	4.64571	4.59695
1000	1	4.02013	4.04193	3.86189	3.86189
	2	4.09053	4.12122	4.09407	4.10059
	3	4.78921	4.79865	4.72979	4.66276
	4	4.81806	4.87485	4.81518	4.68691
	5	4.70216	4.65424	4.59501	4.56885
	MAE	1.69082	1.69294	1.67826	1.66978
	MQE	4.48402	4.49818	4.41919	4.37620
5000	1	4.40165	4.44728	4.19095	4.19095
	2	4.50166	4.50606	4.45559	4.46943
	3	4.61671	4.67040	4.62636	4.54021
	4	4.84170	4.89401	4.74224	4.62208
	5	4.70012	4.61037	4.79648	4.67542
	MAE	1.71608	1.72020	1.69913	1.69054
	MQE	4.61237	4.62562	4.56233	4.49962

Em seguida, por meio das Figuras 3.46 a 3.54, ilustramos os dados apresentados na Ta-

bela 3.7, para os três tamanhos de amostra fixados, respectivamente. Nas Figuras 3.46, 3.49 e 3.52 registramos os histogramas e dispersões para os parâmetros lineares do modelo, diversificando o tamanho da amostra. Notamos que, com o aumento da amostra, as estimativas aproximam-se do valor real dos parâmetros e que a quantidade de *outliers* é reduzida.

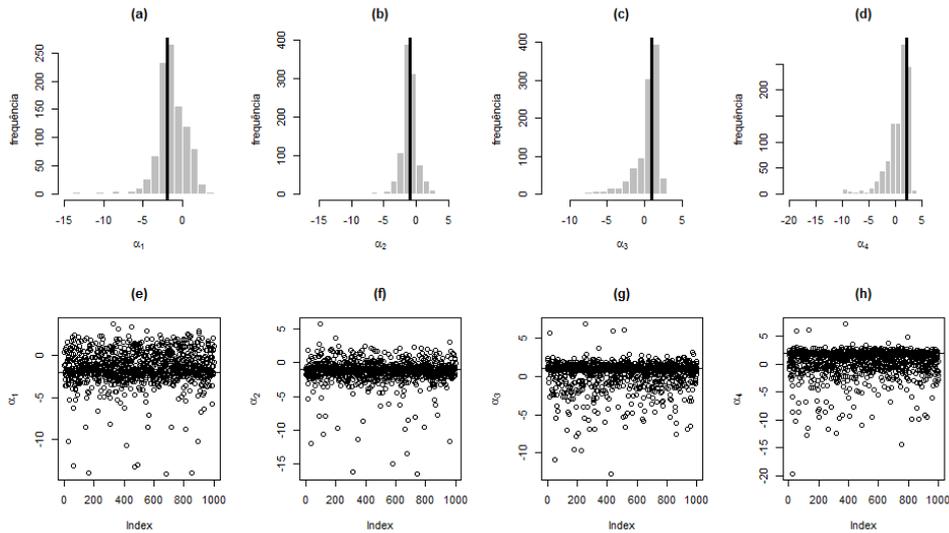


Figura 3.46: (a), (b), (c) e (d) Histogramas de $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$ e $\hat{\alpha}_4$ e (e), (f), (g) e (h) dispersões de $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$ e $\hat{\alpha}_4$ para o Exemplo 3.12 com $n=500$.

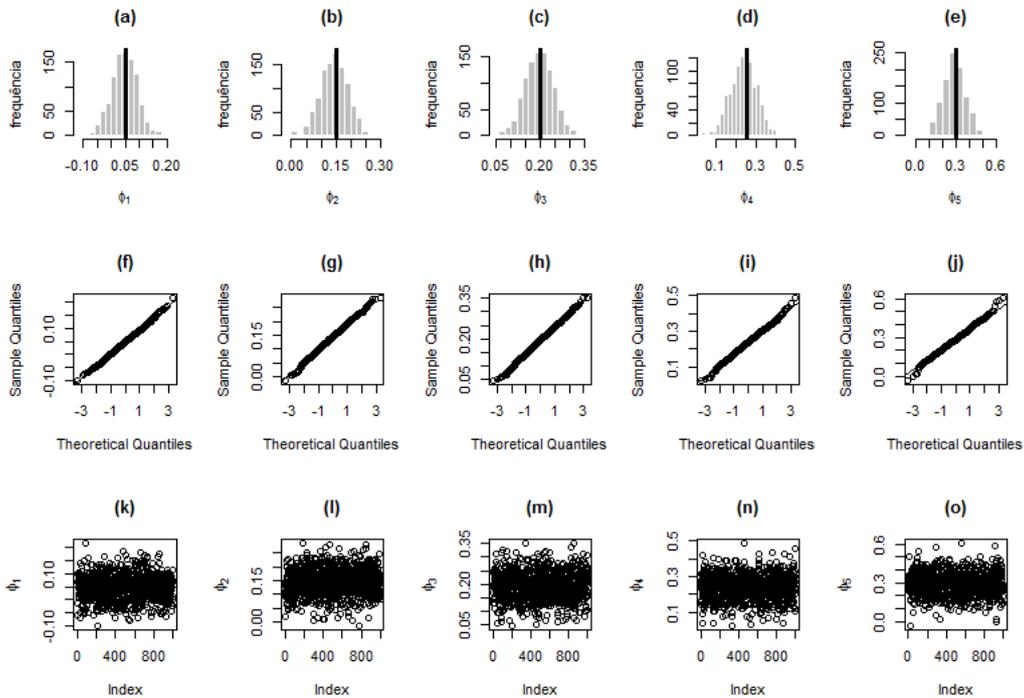


Figura 3.47: (a), (b), (c), (d) e (e) Histogramas de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ e $\hat{\phi}_5$, (f), (g), (h), (i) e (j) gráficos Q-Q de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ e $\hat{\phi}_5$ e (k), (l), (m), (n) e (o) dispersões de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ e $\hat{\phi}_5$ para o Exemplo 3.12 com $n=500$, quando α é conhecido.

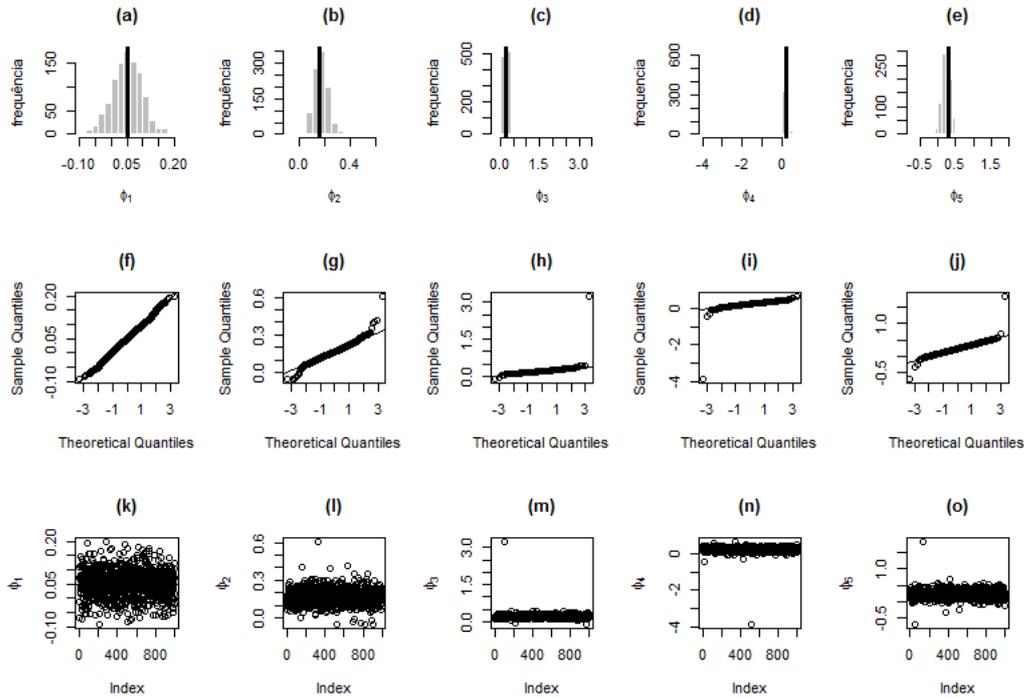


Figura 3.48: (a), (b), (c), (d) e (e) Histogramas de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ e $\hat{\phi}_5$, (f), (g), (h), (i) e (j) gráficos Q-Q de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ e $\hat{\phi}_5$ e (k), (l), (m), (n) e (o) dispersões de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ e $\hat{\phi}_5$ para o Exemplo 3.12 com $n=500$, quando α é desconhecido.

Em relação às Figuras 3.47, 3.50 e 3.53, observamos os histogramas, os gráficos Q-Q e as dispersões para os parâmetros autorregressivos do modelo, quando o parâmetro α é conhecido, diversificando o tamanho da amostra. Reparamos o bom ajuste das estimativas do modelo à distribuição normal, já no caso onde $n = 500$.

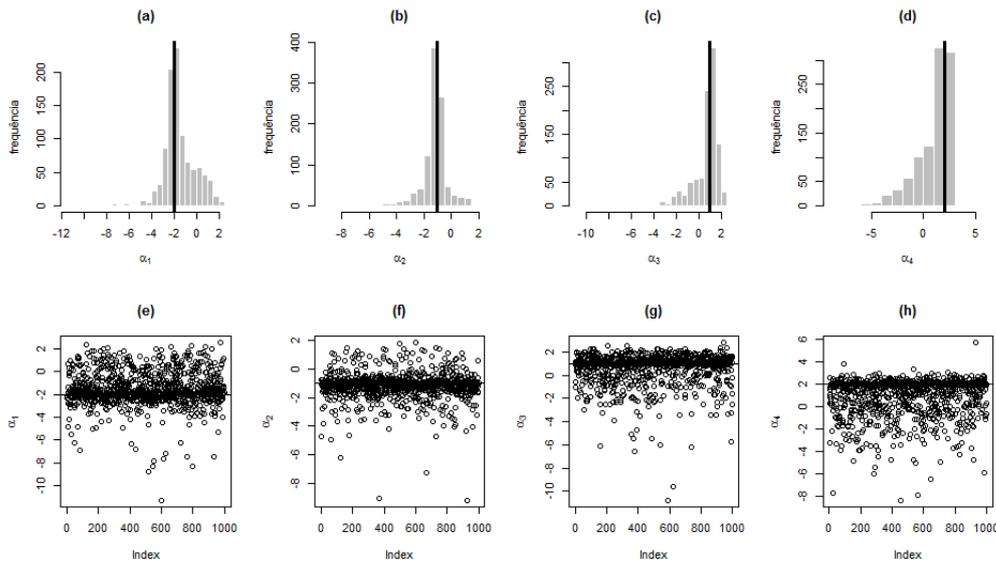


Figura 3.49: (a), (b), (c) e (d) Histogramas de $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$ e $\hat{\alpha}_4$ e (e), (f), (g) e (h) dispersões de $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$ e $\hat{\alpha}_4$ para o Exemplo 3.12 com $n=1000$.

No que diz respeito às Figuras 3.48, 3.51 e 3.54, visualizamos os histogramas, os gráficos

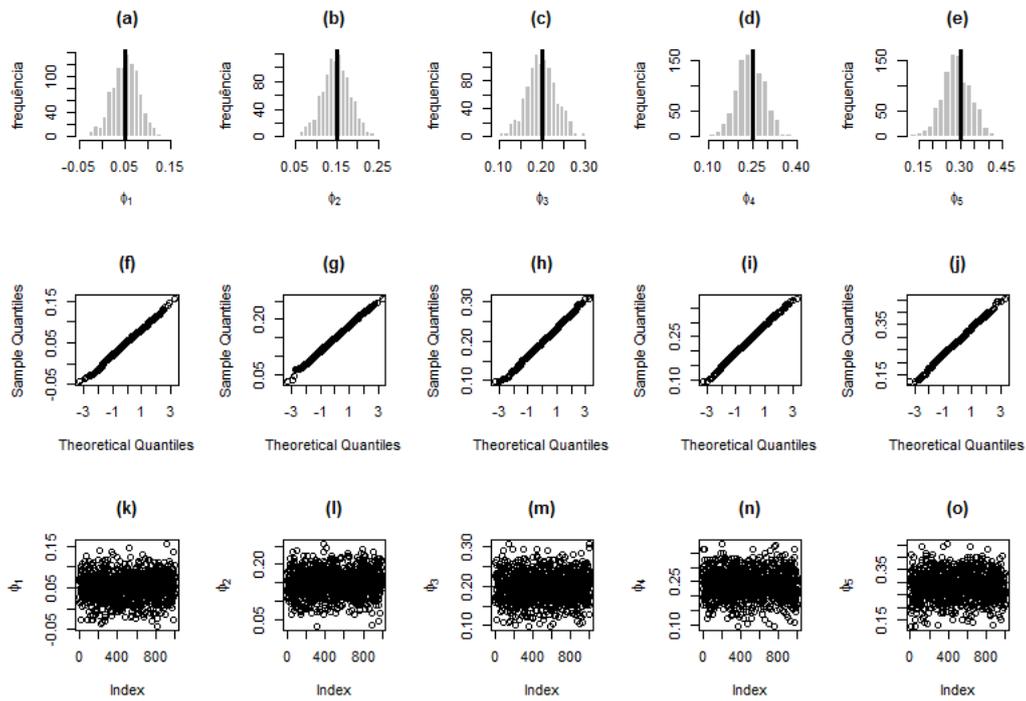


Figura 3.50: (a), (b), (c), (d) e (e) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$, (f), (g), (h), (i) e (j) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ e (k), (l), (m), (n) e (o) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ para o Exemplo 3.12 com $n=1000$, quando α é conhecido.

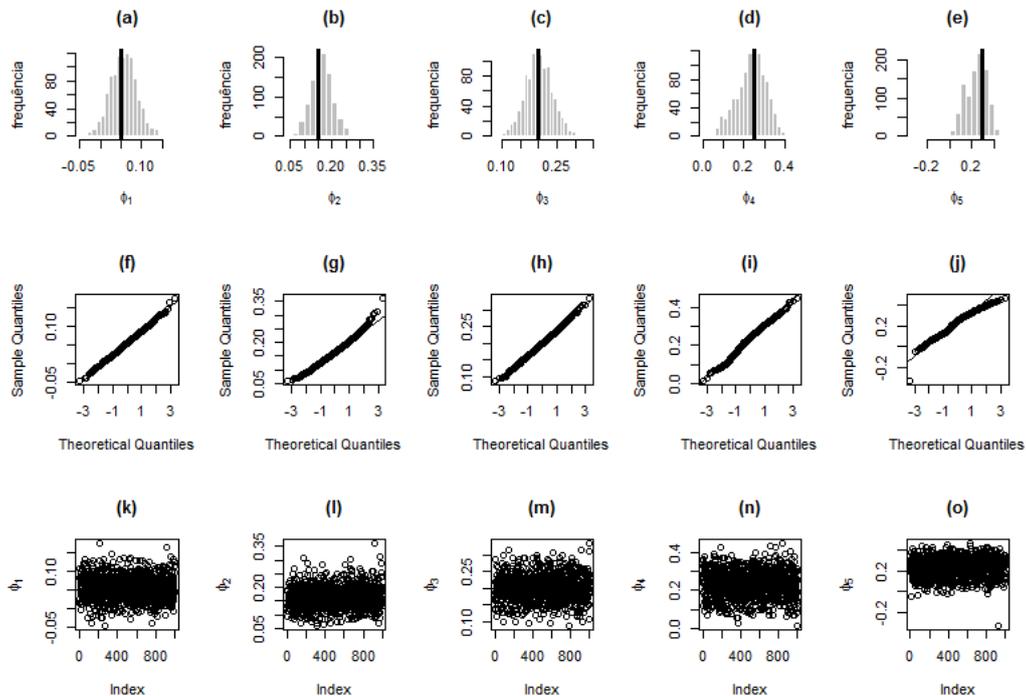


Figura 3.51: (a), (b), (c), (d) e (e) Histogramas de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$, (f), (g), (h), (i) e (j) gráficos Q-Q de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ e (k), (l), (m), (n) e (o) dispersões de $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_5$ para o Exemplo 3.12 com $n=1000$, quando α é desconhecido.

Q-Q e as dispersões para os parâmetros autorregressivos do modelo, quando o parâmetro α é desconhecido, diversificando o tamanho da amostra. Percebemos que à proporção em que se eleva o tamanho da amostra, os gráficos apresentam uma forma mais adequada à distribuição normal. No caso onde $n = 500$ (Figura 3.48), com exceção de ϕ_1 , todas as estimativas para os parâmetros ϕ não se mostram apropriadas à distribuição normal. No caso em que $n = 1000$ (Figura 3.51), todas as estimativas se mostram adequadas para os parâmetros ϕ à distribuição normal, com exceção de ϕ_5 . Já no caso onde $n = 5000$ (Figura 3.54), as estimativas para os parâmetros ϕ_4 e ϕ_5 foram inapropriadas à distribuição normal. Fadel (2012) sugere que o comportamento do gráfico Q-Q para esses parâmetros seja causado pelo número insuficiente de observações dentro desses regimes.

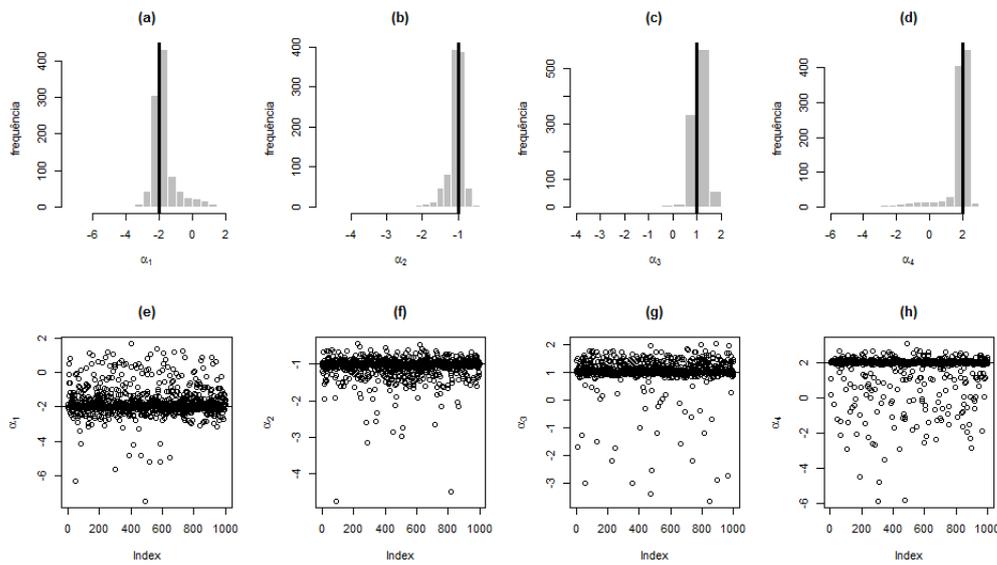


Figura 3.52: (a), (b), (c) e (d) Histogramas de $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$ e $\hat{\alpha}_4$ e (e), (f), (g) e (h) dispersões de $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$ e $\hat{\alpha}_4$ para o Exemplo 3.12 com $n=5000$.

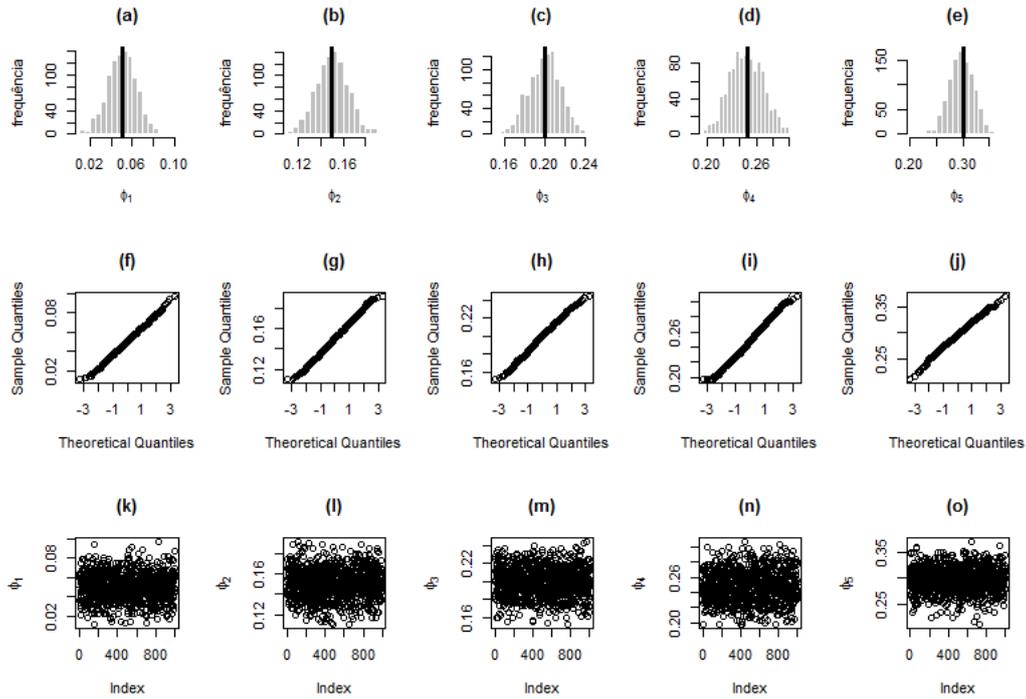


Figura 3.53: (a), (b), (c), (d) e (e) Histogramas de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ e $\hat{\phi}_5$, (f), (g), (h), (i) e (j) gráficos Q-Q de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ e $\hat{\phi}_5$ e (k), (l), (m), (n) e (o) dispersões de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ e $\hat{\phi}_5$ para o Exemplo 3.12 com $n=5000$, quando α é conhecido.

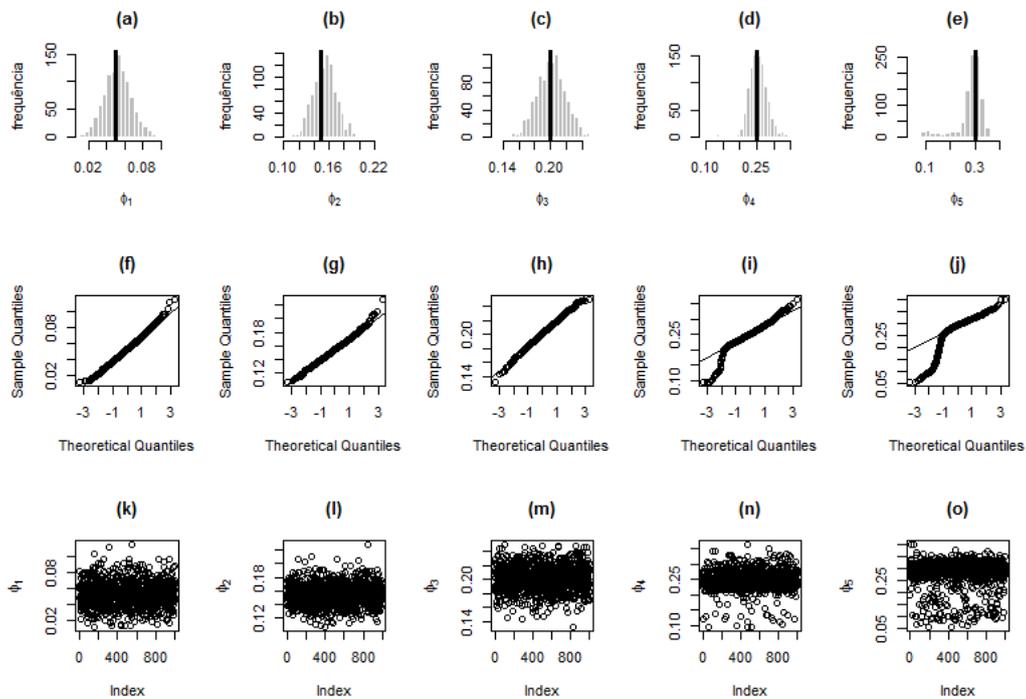


Figura 3.54: (a), (b), (c), (d) e (e) Histogramas de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ e $\hat{\phi}_5$, (f), (g), (h), (i) e (j) gráficos Q-Q de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ e $\hat{\phi}_5$ e (k), (l), (m), (n) e (o) dispersões de $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ e $\hat{\phi}_5$ para o Exemplo 3.12 com $n=5000$, quando α é desconhecido.

Capítulo 4

Aplicações

Neste capítulo aplicamos os modelos AR-MV em quatro conjuntos de dados dendrocronológicos reais, extraídos do *site* Datamarket (2017). De acordo com o *site* CCVFloresta (2018), a dendrocronologia é uma técnica de datação que se baseia nos anéis de crescimento das árvores.

Para ilustrar, a Figura 4.1 apresenta uma ilustração retirada do *site* CCVFloresta (2018). O número 1 representa um nó originado por um ramo antigo. O número 2 retrata o 1º ano de crescimento. Os números 3 e 4 simbolizam os crescimentos de primavera e outono, respectivamente. O número 5 é uma cicatriz resultante de um incêndio. O número 6 mostra um anel de crescimento anual.

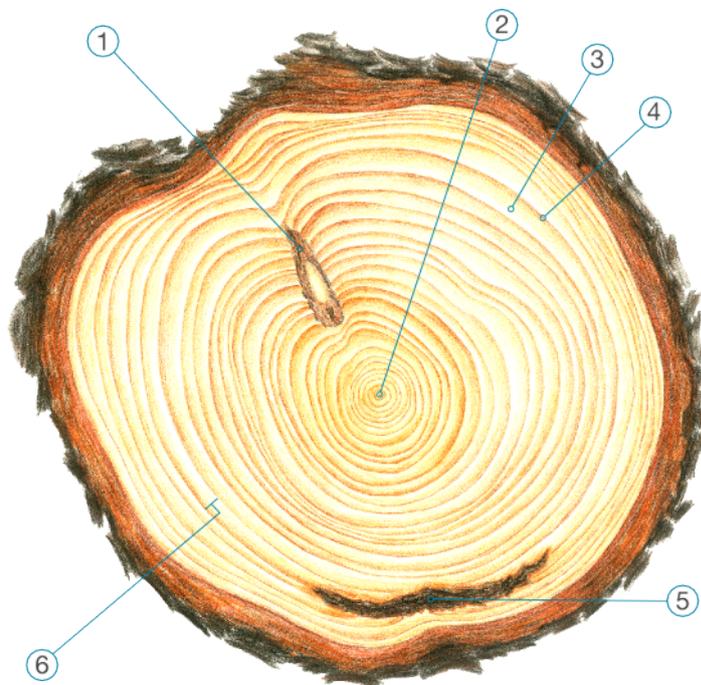


Figura 4.1: Ilustração do corte de uma árvore (CCVFloresta (2018)).

Identificamos que os modelos autorregressivos de ordens 2, 3, 4 e 5, ou seja, AR(2), AR(3), AR(4) e AR(5) se ajustaram de maneira adequada a esses conjuntos de observações

e, com a finalidade de realizarmos a comparação de modelos, ajustamos modelos AR-MV de ordens idênticas, aos dados.

O objetivo desta aplicação é comparar o desempenho do ajuste de um modelo AR(p), $p = \{2, \dots, 5\}$, ao de um modelo AR-MV(p), aos dados, utilizando para isso, alguns critérios bem conhecidos na literatura, quais sejam, a variância residual ($\hat{\sigma}^2$), a SQR, bem como os critérios AIC (Akaike, 1973), (Akaike, 1974) e BIC (Schwarz, 1978).

É necessário salientar que as cinco últimas observações de todos os bancos de dados analisados não serão utilizadas na fase da modelagem, no intuito de empregá-las num momento posterior, para testar a capacidade preditiva dos modelos selecionados.

Vimos na Seção 2.3 que os modelos AR são utilizados para a modelagem de séries temporais lineares, enquanto o modelo AR-MV, que é oriundo da classe de modelos TAR, é empregado na modelagem de séries não lineares. Embora tais modelos tenham naturezas distintas, suas FAC e FACP se comportam de maneira semelhante (como verificamos nos estudos simulados). Além disso, as árvores que serão objeto de estudo neste trabalho, podem viver durante milhares de anos e, intuitivamente, é razoável pensarmos que, durante todo esse período, tenha ocorrido algum fenômeno climático de grande vulto, o que poderia ocasionar temporariamente uma estagnação no crescimento ou um crescimento acelerado das árvores. Dito isso, podemos pensar neste conjunto de dados de uma forma especial, porque pequenas variações podem ter modificado o comportamento de crescimento e causado assim pequenas mudanças de regime. Nesse sentido, decidimos comparar as classes AR contra AR-MV e investigar o potencial desta classe não linear para a modelagem dos dados.

Após a análise dessas aplicações pudemos concluir que, em relação ao ajuste dos modelos, houve um empate, pois, nas aplicações I e II, o AR foi melhor, mas, nas aplicações III e IV, o AR-MV foi preferível. É importante ressaltar que o AR-MV obteve um melhor ajuste nos casos onde os dados exigiam modelos de maior ordem autorregressiva, ou seja, quarta e quinta. No que diz respeito às previsões, também houve um emparelhamento. Nas aplicações I e IV os métodos referentes ao AR-MV mostraram uma capacidade preditiva mais adequada, enquanto que nas aplicações II e III, o AR foi superior nesse quesito.

4.1 Aplicação I - AR-MV(2)

4.1.1 Banco de dados

Este banco de dados é composto por 655 observações, atinentes ao período de 1311 a 1965, contendo medições de anéis de crescimento de pinheiros da espécie “Limber” (nome científico: *Pinus flexilis*), na região de Dell, Montana, EUA.

4.1.2 Análise de normalidade dos dados

Nesta subseção será avaliada a aderência dos dados, apresentados na subseção anterior, a uma distribuição normal, tendo em vista que esse conjunto de observações será padronizado, no intuito de viabilizar a estimação do parâmetro limiar (α) pelo modelo AR-MV a ser ajustado posteriormente.

Inicialmente, serão analisados o histograma e o gráfico Q-Q dos dados originais. Percebemos, por meio da Figura 4.2, que o ajuste segue satisfatoriamente uma distribuição normal de probabilidade. Contudo, para ratificar essa percepção visual, serão utilizados os testes mostrados na Tabela 4.1, os quais podem ser verificados de maneira mais detalhada no Apêndice C.

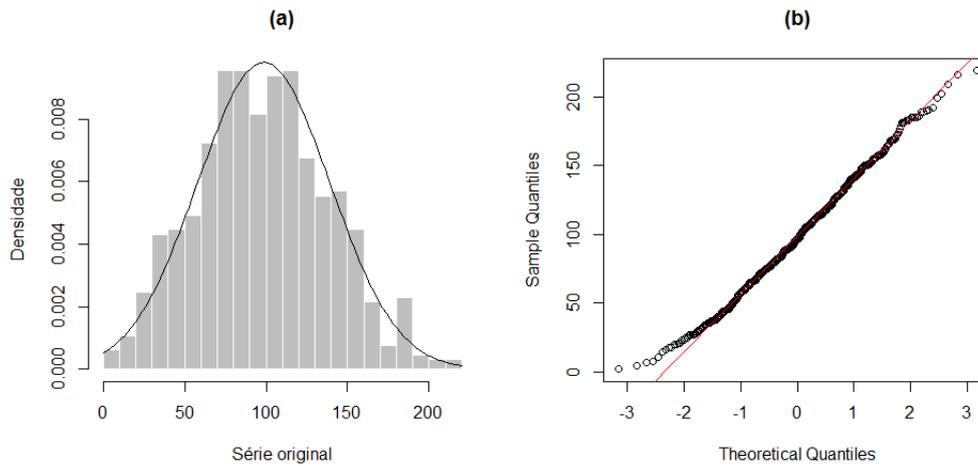


Figura 4.2: (a) Histograma e (b) gráfico Q-Q para a série temporal gerada para as medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).

Tabela 4.1: Testes para a detecção de normalidade na série temporal das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).

Teste	Estatística	P-valor
Kolmogorov-Smirnov	0.03369	0.45186
Lilliefors	0.03369	0.07658
Cramer-von Mises	0.08876	0.15932
Shapiro-Wilk	0.99456	0.02041
Shapiro-Francia	0.99534	0.04672
Anderson-Darling	0.61726	0.10783

De acordo com a Tabela 4.1, com exceção dos testes Shapiro-Wilk e Shapiro-Francia, ao nível de significância de 5%, não rejeitamos a hipótese nula de normalidade dos dados.

Portanto, tomando como referência, além dos testes exibidos na Tabela 4.1, a Figura 4.2, é razoável supor normalidade para os dados sob análise.

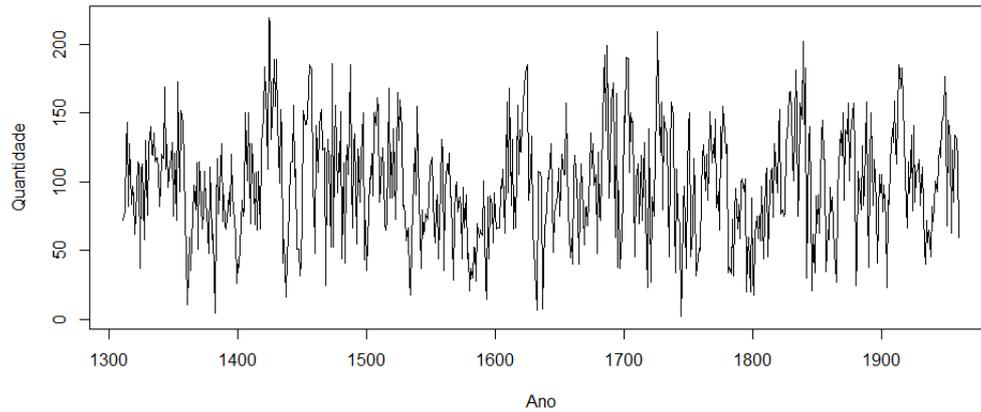


Figura 4.3: *Série temporal gerada para as medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).*

4.1.3 Análise de estacionariedade dos dados

A Figura 4.3 exhibe a série temporal dos dados em questão, a qual nos faz perceber um padrão estacionário. No entanto, para ratificar essa percepção visual, serão utilizados os testes mostrados na Tabela 4.2.

Tabela 4.2: *Testes para a detecção de estacionariedade na série de medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).*

	ADF	PP	KPSS
Estatística	-6.67	-419.51	0.09
P-valor	<0.01	<0.01	>0.10

Com base nos testes exibidos na Tabela 4.2, ao nível de significância de 5%, confirmamos a estacionariedade dos dados. Dessa forma, é possível analisar a FAC e a FACP da série com o intuito de identificar um modelo ARIMA adequado para este conjunto de dados.

4.1.4 Identificação do modelo ARIMA

A Figura 4.4 expõe os gráficos da FAC e FACP dos dados. Percebemos que a FAC apresenta um decaimento exponencial veloz para zero, comportamento próprio de um modelo autorregressivo; e a FACP possui os *lags* 1 e 2 bastante expressivos, indicando que o AR(2) pode ser um modelo adequado aos dados. Assim, podemos postular o seguinte modelo

$$(1 - \phi_1 B - \phi_2 B^2)Z_t = a_t. \quad (4.1)$$

Contudo, percebemos também que o *lag* 9 da FACP possui uma pequena significância e, por esse motivo, será testada a adequabilidade da inclusão do parâmetro ϕ_9 ao modelo apresentado na equação (4.1).

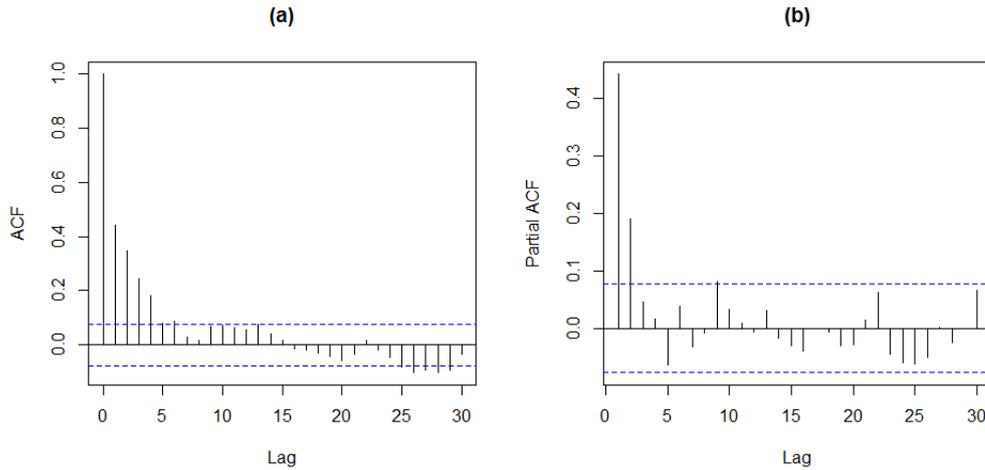


Figura 4.4: (a) FAC e (b) FACP para a série temporal das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).

Na Tabela 4.3 são apresentadas as estimativas dos coeficientes autorregressivos referentes aos modelos AR(2) e AR(9), bem como seus respectivos erros-padrão e p-valores. Cabe ressaltar que os modelos com interceptos não serão mostrados, pois esses coeficientes não foram significativos para os dois casos. Observamos que, para o modelo AR(2), os coeficientes autorregressivos são altamente significativos, mostrando a conveniência desse modelo aos dados. Em relação ao modelo AR(9), constatamos que os dois primeiros coeficientes autorregressivos são significativos. No entanto, ao nível de significância de 5%, o coeficiente ϕ_9 não é expressivo, mostrando que sua inclusão ao modelo apresentado na equação (4.1) não é pertinente. Logo, essa opção será desprezada.

Tabela 4.3: Estimativas dos parâmetros dos modelos AR(2) e AR(9) das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).

	AR(2)			AR(9)		
	Coefficiente	Erro-padrão	P-valor	Coefficiente	Erro-padrão	P-valor
ϕ_1	0.35783	0.0385	<0.01	0.35753	0.03843	<0.01
ϕ_2	0.19016	0.0385	<0.01	0.18875	0.03843	<0.01
ϕ_9	-	-	-	0.05539	0.03446	0.10805

Diante do exposto, decidimos por selecionar o AR(2) como o modelo mais adequado para ajustamento dos dados em questão, cujas estimativas encontram-se sumarizadas na seguinte equação

$$(1 - 0.35783B - 0.19016B^2)Z_t = a_t. \quad (4.2)$$

4.1.5 Diagnóstico do modelo ARIMA selecionado

Inicialmente, na Figura 4.5 são apresentados os gráficos de dispersão, FAC e FACP dos resíduos. Observamos que os dados se distribuem aleatoriamente em torno de zero no

gráfico de dispersão. Ademais, os valores das autocorrelações e autocorrelações parciais não indicam alteração no comportamento de ruído branco dos resíduos, dado que não há *lags* significativos.

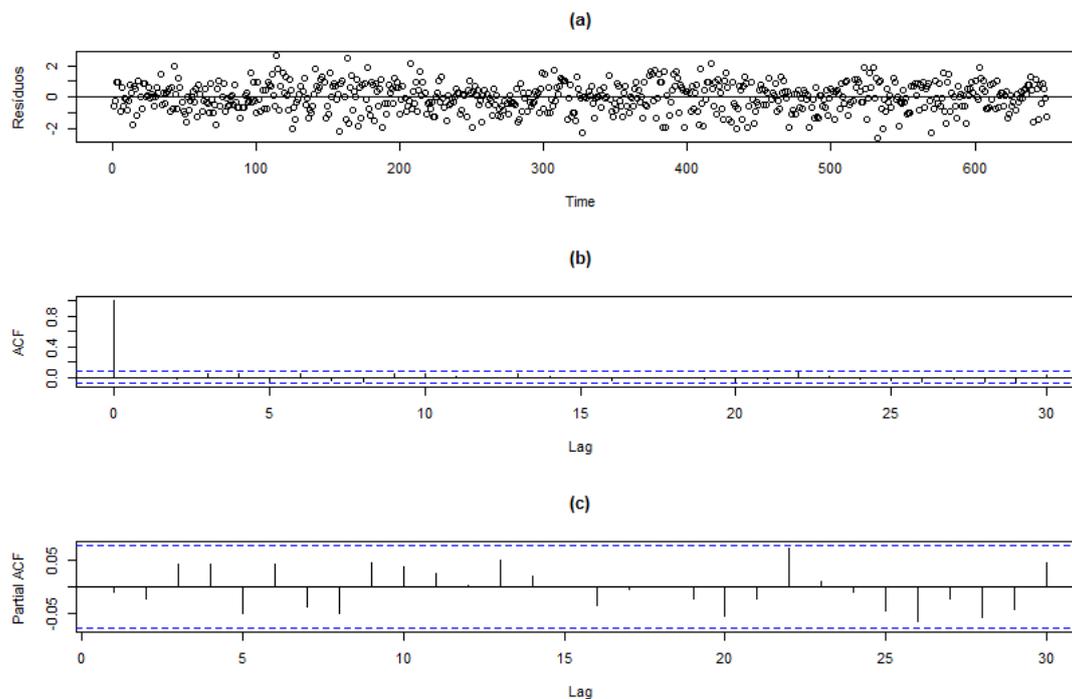


Figura 4.5: (a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo $AR(2)$ para a série temporal padronizada das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).

A seguir, a Tabela 4.4 apresenta os valores das estatísticas qui-quadrado do teste de Box-Pierce modificado para os *lags* 1 a 30, e seus respectivos p-valores.

Tabela 4.4: Estatísticas qui-quadrado do teste de Box-Pierce modificado para os *lags* 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo $AR(2)$ da série temporal padronizada das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).

Lag	Estatística	P-valor	Lag	Estatística	P-valor	Lag	Estatística	P-valor
1	0.04588	0.83040	11	10.59168	0.47808	21	16.03878	0.76747
2	0.36068	0.83499	12	10.60373	0.56315	22	19.50595	0.61391
3	1.55978	0.66854	13	12.72699	0.46911	23	19.59892	0.66597
4	2.74500	0.60136	14	12.90151	0.53429	24	19.67217	0.71526
5	4.39215	0.49444	15	12.90615	0.60955	25	20.65141	0.71188
6	5.56088	0.47413	16	13.45639	0.63914	26	22.67992	0.65099
7	6.24759	0.51116	17	13.46378	0.70463	27	22.95028	0.68768
8	8.00697	0.43279	18	13.47293	0.76275	28	24.37847	0.66141
9	9.30947	0.40921	19	13.76314	0.79732	29	26.68143	0.58889
10	10.48605	0.39893	20	15.73046	0.73320	30	27.11797	0.61707

Como não há valores significantes, podemos afirmar que os resíduos do modelo $AR(2)$ ajustado a estes dados são ruído branco.

4.1.6 Estimação e diagnóstico do modelo AR-MV

Dado que o ajuste de um modelo AR(2) é adequado aos dados, o modelo AR-MV(2) também será ajustado. Além disso, para verificar a adequação do modelo AR-MV(2) aos dados em estudo, os resíduos desse modelo serão avaliados.

As estimativas dos parâmetros do modelo AR-MV(2) ajustado encontram-se resumidas no seguinte sistema de equações

$$Z_t = \begin{cases} 0.408Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq -0.436 \\ 0.408Z_{t-1} + 0.136Z_{t-2} + a_t, & \text{se } Z_{t-1} > -0.436 \end{cases}. \quad (4.3)$$

Na Figura 4.6 são apresentados os gráficos de dispersão, FAC e FACP dos resíduos. Reparamos que os dados se distribuem aleatoriamente em torno de zero no gráfico de dispersão. Ademais, os valores das autocorrelações e autocorrelações parciais não indicam alteração no comportamento de ruído branco dos resíduos, dado que não há *lags* significativos.

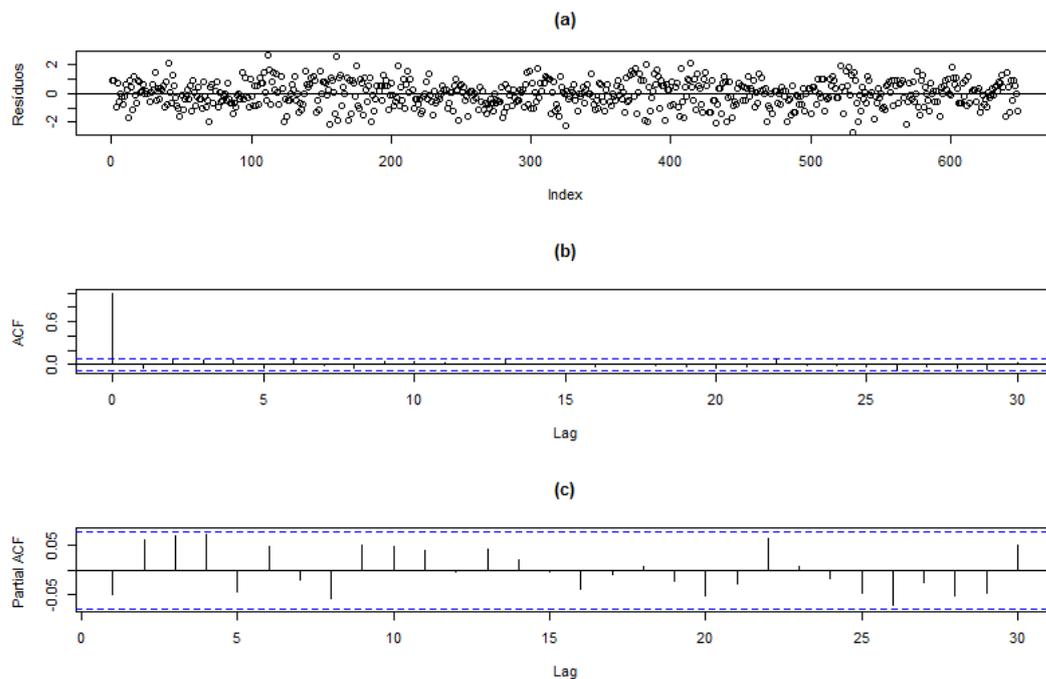


Figura 4.6: (a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo AR-MV(2) para a série temporal padronizada das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).

A seguir, a Tabela 4.5 apresenta os valores das estatísticas qui-quadrado do teste de Box-Pierce modificado para os *lags* 1 a 30, e seus respectivos p-valores. Percebemos que os *lags* 4, 6 e 9 são significativos ao nível de 5%. No entanto, como os gráficos da FAC e FACP (Figura 4.6) não apresentaram *lags* significativos, podemos supor que os resíduos do modelo AR-MV(2) ajustado aos dados são ruído branco.

Por intermédio da Figura 4.7, percebemos um ajuste satisfatório dos resíduos à distribuição normal. Contudo, para ratificar essa percepção visual, aplicamos os testes exibidos na

Tabela 4.5: Estatísticas qui-quadrado do teste de Box-Pierce modificado para os lags 1 a 30 e respectivos p -valores dos resíduos do ajuste pelo modelo AR-MV(2) da série temporal padronizada das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).

Lag	Estatística	P-valor	Lag	Estatística	P-valor	Lag	Estatística	P-valor
1	1.49678	0.22117	11	18.50544	0.07057	21	23.80617	0.30254
2	4.00281	0.13515	12	18.51347	0.10097	22	25.87479	0.25710
3	6.52316	0.08875	13	20.86926	0.07556	23	25.88407	0.30637
4	9.54513	0.04883	14	21.01574	0.10123	24	25.91937	0.35727
5	10.63941	0.05902	15	21.09597	0.13379	25	26.92022	0.35988
6	13.35265	0.03777	16	21.58727	0.15702	26	29.27738	0.29862
7	13.60707	0.05863	17	21.58797	0.20109	27	29.76629	0.32476
8	15.09712	0.05728	18	21.59057	0.25067	28	31.43773	0.29799
9	16.92137	0.04996	19	21.87937	0.29030	29	34.22917	0.23097
10	18.04366	0.05423	20	23.39269	0.26993	30	34.54916	0.25941

Tabela 4.6, e notamos que em todos os testes, não rejeitamos a hipótese nula de normalidade dos dados.

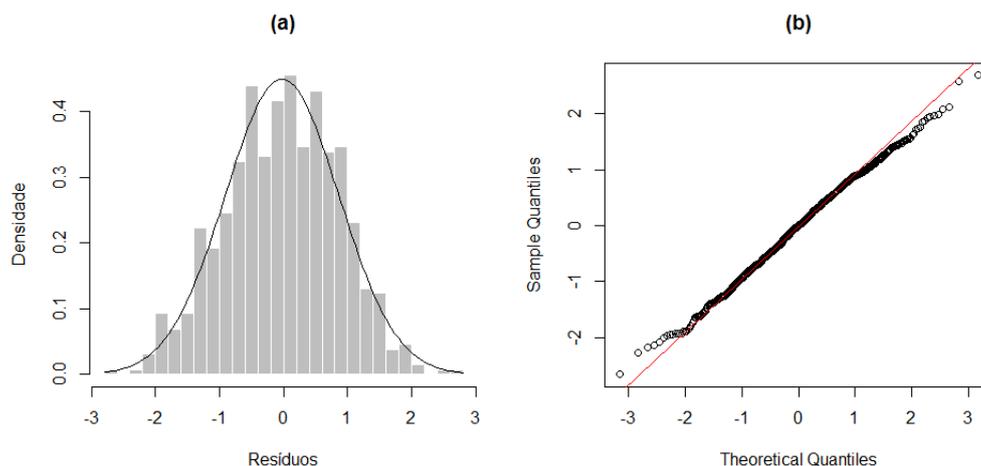


Figura 4.7: (a) Histograma e (b) gráfico Q-Q para a série temporal gerada para as medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).

Tabela 4.6: Testes para a detecção de normalidade na série temporal das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960).

Teste	Estatística	P-valor
Kolmogorov-Smirnov	0.02845	0.67081
Lilliefors	0.02845	0.22979
Cramer-von Mises	0.09015	0.15290
Shapiro-Wilk	0.99702	0.28319
Shapiro-Francia	0.99730	0.32355
Anderson-Darling	0.56648	0.14186

4.1.7 Comparação dos modelos ARIMA e AR-MV ajustados

Como os modelos mostraram-se adequados, nesta seção realizaremos a comparação desses modelos, sob o ponto de vista da ajustabilidade aos dados. Na Tabela 4.7 estão descritos os valores estimados para os parâmetros indicados. Além disso, nessa tabela encontramos os valores estabelecidos como critérios de comparação. Observamos que, conforme todos os critérios de comparação utilizados, o modelo AR(2) é superior.

Tabela 4.7: Estimativas dos parâmetros e critérios de comparação, sob os modelos AR(2) e AR-MV(2), para as medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA (1311-1960). (O valor entre parênteses indica o erro-padrão de cada estimativa).

Estimativas/Critérios	AR-MV(2)	AR(2)
$\hat{\alpha}$	-0.43628	-
$\hat{\phi}_1$	0.40824 (0.03686)	0.35783 (0.03850)
$\hat{\phi}_2$	0.13556 (0.04567)	0.19016 (0.03850)
$\hat{\sigma}^2$	0.78532	0.76535
SQR	508.89	497.47
AIC	1688.35	1677.08
BIC	1701.77	1690.51

4.1.8 Previsão

Nesta etapa serão realizadas as previsões, sob os modelos AR(2) e AR-MV(2), para o período de 1961 a 1965, cujas observações verdadeiras (-1.36641, 0.46745, 1.49441, 1.56777 e 2.32576) não foram utilizadas na fase da estimação, justamente com o objetivo de serem comparadas as capacidades preditivas dos dois modelos selecionados.

A Tabela 4.8 exibe as estimativas do MAE, MQE e os erros em cada uma das cinco etapas de previsões para os modelos em teste. Notamos que os valores previstos pelos métodos referentes ao AR-MV(2) foram superiores às previsões geradas pelo modelo AR(2), sendo que o método de Monte Carlo, apresentou os menores erros de previsão.

Por fim, a Figura 4.8 expõe o gráfico contendo os valores verdadeiros observados para o período de 1961 a 1965 e as previsões geradas pelos dois modelos. Percebemos que apesar de os métodos de previsão do modelo AR-MV(2) serem superiores ao AR(2), as predições de ambos os modelos são bastante próximas.

Tabela 4.8: Estimativas do MAE, MQE e os erros de previsão a cada passo, sob os modelos $AR(2)$ e $AR-MV(2)$, para o período de 1961 a 1965, da série temporal padronizada das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA.

Passo	AR(2) (sem intercepto)	AR-MV(2) (Esqueleto)	AR-MV(2) (Monte Carlo)
1	-1.10115	-0.97828	-0.97828
2	0.74316	0.75478	0.78192
3	1.64351	1.66433	1.68071
4	1.67355	1.67608	1.61293
5	2.39197	2.39301	2.32134
MAE	1.51067	1.49330	1.47504
MQE	2.59764	2.56650	2.47668

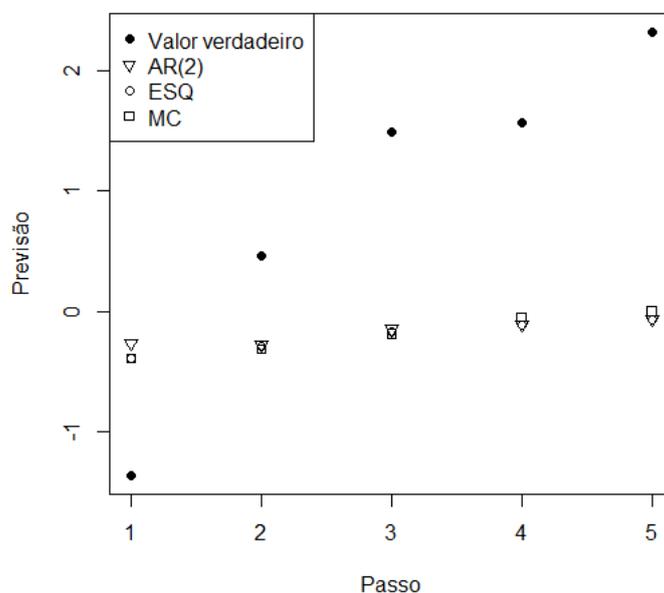


Figura 4.8: Valores observados e previsões pontuais dos modelos $AR(2)$ e $AR-MV(2)$ para a série temporal padronizada das medições de anéis de crescimento de pinheiros “Limber” em Dell, Montana, EUA, para o período de 1961 a 1965.

4.2 Aplicação II - AR-MV(3)

4.2.1 Banco de dados

Este banco de dados é composto por 787 observações, referentes ao período de 1197 a 1983, contendo medições de anéis de crescimento de pinheiros “Bristlecone” (nome científico: *Pinus longaeva*), na região de San Francisco Peaks, Arizona, EUA.

4.2.2 Análise de normalidade dos dados

Nesta subseção será avaliada a aderência dos dados, apresentados na subseção anterior, a uma distribuição normal, tendo em vista que esse conjunto de observações será padronizado, no intuito de viabilizar a estimação dos parâmetros limiares (α_1 e α_2) pelo modelo AR-MV a ser ajustado posteriormente.

Primeiramente, serão analisados o histograma e o gráfico Q-Q dos dados originais. Constatamos, por meio da Figura 4.9, que o ajuste segue, de maneira apropriada, uma distribuição normal de probabilidade. No entanto, para confirmar esse entendimento inicial, serão utilizados os testes mostrados na Tabela 4.9, os quais encontramos detalhados no Apêndice C.

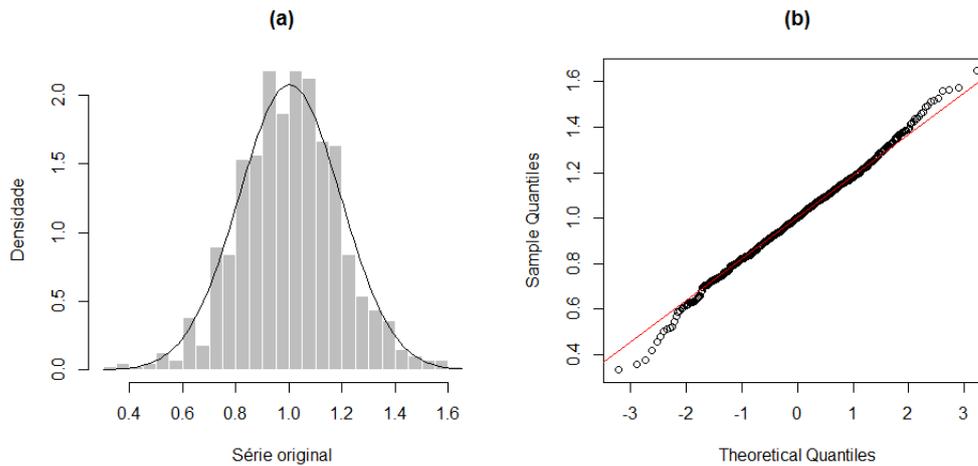


Figura 4.9: (a) Histograma e (b) gráfico Q-Q para a série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).

Tabela 4.9: Testes para a detecção de normalidade na série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).

Teste	Estatística	P-valor
Kolmogorov-Smirnov	0.02340	0.78523
Lilliefors	0.02340	0.37551
Cramer-von Mises	0.05923	0.38680
Shapiro-Wilk	0.99652	0.08360
Shapiro-Francia	0.99630	0.06022
Anderson-Darling	0.50013	0.20800

Segundo todos os testes expostos na Tabela 4.9, ao nível de 5% de significância, não rejeitamos a hipótese nula de normalidade dos dados.

Logo, é razoável afirmar que os dados sob avaliação seguem uma distribuição aproximadamente normal.

4.2.3 Análise de estacionariedade dos dados

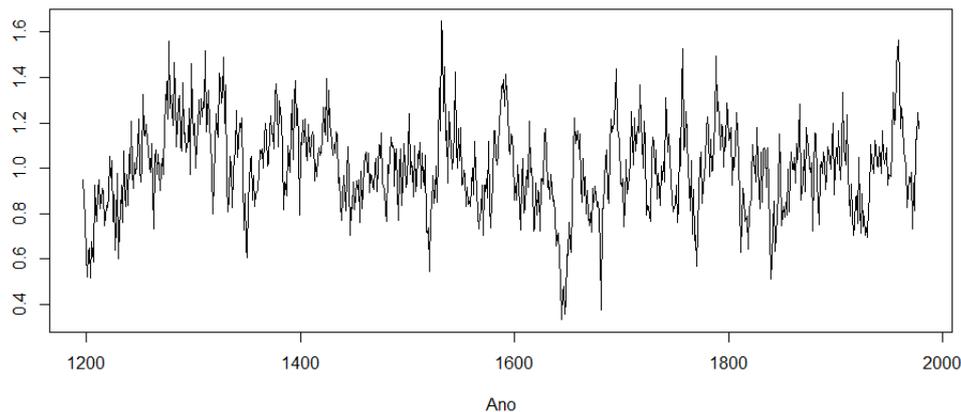


Figura 4.10: Série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).

A Figura 4.10 externa a série temporal das observações em questão, por meio da qual verificamos um padrão estacionário nos dados. Entretanto, para atestar essa concepção visual, serão utilizados os testes mostrados na Tabela 4.10.

Tabela 4.10: Testes para a detecção de estacionariedade na série de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).

	ADF	PP	KPSS
Estatística	-6.70	-233.78	0.3647
P-valor	<0.01	<0.01	0.09237

Tomando como base os testes apresentados na Tabela 4.10, ao nível de significância de 5%, confirmamos a estacionariedade dos dados. Assim, é possível investigar a FAC e a FACP da série a fim de identificar um modelo ARIMA coerente, para este conjunto de observações.

4.2.4 Identificação do modelo ARIMA

A Figura 4.11 mostra os gráficos da FAC e FACP dos dados. Notamos que a FAC exibe um decaimento exponencial veloz para zero, comportamento próprio de um modelo autorregressivo; e a FACP revela as três primeiras defasagens significativas, sugerindo a hipótese de que o AR(3) pode ser um modelo apropriado aos dados. Logo, postulamos o seguinte modelo

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)Z_t = a_t. \quad (4.4)$$

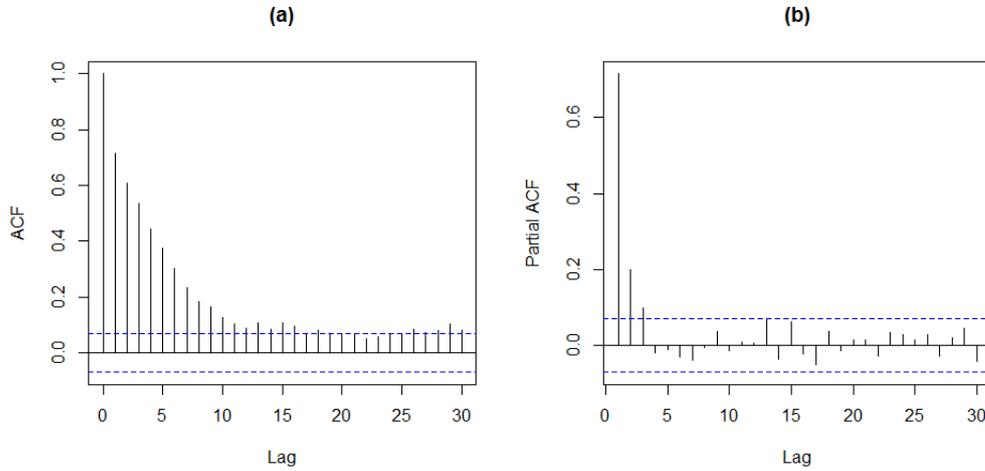


Figura 4.11: (a) FAC e (b) FACP amostrais para as medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).

Todavia, visualizamos também que o *lag* 13 da FACP possui uma ligeira significância e, desse modo, será averiguada a adequabilidade da inclusão do parâmetro ϕ_{13} ao modelo exposto na equação (4.4).

Na Tabela 4.11 são mostradas as estimativas dos coeficientes autorregressivos atinentes aos modelos AR(3) e AR(13), assim como seus respectivos erros-padrão e p-valores. Salientamos que esses modelos, com o intercepto, não serão apresentados, porque esses coeficientes não são significativos para os dois casos. Visualizamos que, para o modelo AR(3), os coeficientes autorregressivos são altamente significativos, revelando a pertinência do ajuste desse modelo aos dados. No que concerne ao modelo AR(13), verificamos que os três primeiros coeficientes autorregressivos são significativos. Todavia, ao nível de significância de 5%, o coeficiente ϕ_{13} não é expressivo, deixando claro que sua inclusão ao modelo exposto na equação (4.4) não é adequada. Portanto, essa opção será desprezada.

Tabela 4.11: Estimativas dos parâmetros dos modelos AR(3) e AR(13) para as medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).

	AR(3)			AR(13)		
	Coefficiente	Erro-padrão	P-valor	Coefficiente	Erro-padrão	P-valor
ϕ_1	0.54920	0.03555	<0.01	0.54884	0.03551	<0.01
ϕ_2	0.14625	0.04027	<0.01	0.14535	0.04023	<0.01
ϕ_3	0.09898	0.03558	<0.01	0.09565	0.03560	<0.01
ϕ_{13}	-	-	-	0.03254	0.02501	0.19326

Diante disso, resolvemos optar pelo AR(3) como o modelo mais adequado para ajustamento dos dados em questão, cujas estimativas encontram-se sumarizadas na seguinte equação

$$(1 - 0.54920B - 0.14625B^2 - 0.09898B^3)Z_t = a_t. \quad (4.5)$$

4.2.5 Diagnóstico do modelo ARIMA selecionado

Na Figura 4.12 são expostos os gráficos de dispersão, FAC e FACP dos resíduos. Podemos reparar que os dados são distribuídos aleatoriamente em torno de zero no gráfico de dispersão. Além do mais, os valores das autocorrelações e autocorrelações parciais não sugerem alteração no comportamento de ruído branco dos resíduos, pelo fato de não haver defasagens significativas.

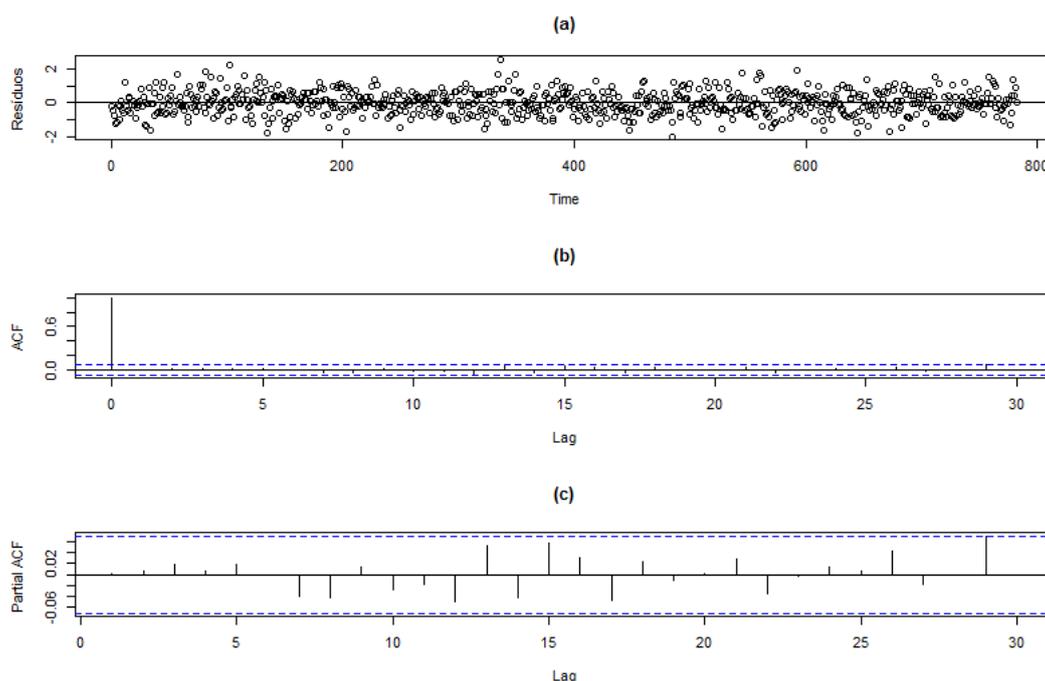


Figura 4.12: (a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo $AR(3)$ para as medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).

Em seguida, a Tabela 4.12 expõe os valores das estatísticas qui-quadrado do teste de Box-Pierce modificado para os *lags* 1 a 30, e seus respectivos p-valores. Como não há estatísticas significantes, ao nível de 5%, podemos confirmar que os resíduos são ruído branco.

Tabela 4.12: Estatísticas qui-quadrado do teste de Box-Pierce modificado para os lags 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo AR(3) da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).

Lag	Estatística	P-valor	Lag	Estatística	P-valor	Lag	Estatística	P-valor
1	0.00556	0.94058	11	4.36669	0.95794	21	16.28219	0.75359
2	0.03102	0.98461	12	6.41881	0.89352	22	17.11729	0.75686
3	0.30437	0.95920	13	8.28405	0.82462	23	17.21039	0.79880
4	0.34477	0.98674	14	9.58300	0.79199	24	17.52023	0.82564
5	0.61054	0.98752	15	12.16819	0.66626	25	17.52215	0.86199
6	0.61066	0.99622	16	13.04458	0.66950	26	18.89611	0.84079
7	1.79542	0.97029	17	14.86408	0.60527	27	19.33392	0.85751
8	3.19941	0.92123	18	15.62394	0.61877	28	19.33674	0.88742
9	3.33239	0.94965	19	15.65662	0.68005	29	23.74928	0.74119
10	4.02524	0.94620	20	15.66659	0.73708	30	23.75478	0.78300

4.2.6 Estimação e diagnóstico do modelo AR-MV

Dado que o ajuste de um modelo AR(3) é adequado aos dados, o modelo AR-MV(3) também será ajustado. Para averiguar a pertinência do modelo AR-MV(3) aos dados sob avaliação, os resíduos desse modelo serão estudados.

As estimativas dos parâmetros do modelo AR-MV(3) ajustado encontram-se sumarizadas no seguinte sistema de equações

$$Z_t = \begin{cases} 0.659Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq -1.12 \\ 0.659Z_{t-1} + 0.151Z_{t-2} + a_t, & \text{se } -1.12 < Z_{t-1} \leq 1.13 \\ 0.659Z_{t-1} + 0.151Z_{t-2} - 0.06Z_{t-3} + a_t, & \text{se } Z_{t-1} > 1.13 \end{cases} \quad (4.6)$$

A Figura 4.13 apresenta os gráficos de dispersão, FAC e FACP dos resíduos. Notamos que os dados se distribuem aleatoriamente em torno de zero no gráfico de dispersão. Visualizamos ainda que os lags 3 e 29 da função de autocorrelação, bem como o lag 3 da função autocorrelação parcial são significativos. Esse tipo de comportamento pode indicar que os resíduos não são ruído branco.

A seguir, a Tabela 4.13 expressa os valores das estatísticas qui-quadrado do teste de Box-Pierce modificado para os lags 1 a 30, e seus respectivos p-valores. Constatamos que os lags 3, 4, 5 e 6 são significativos ao nível de 5%. Logo, os resíduos do modelo AR-MV(3) para estes dados não são ruído branco, indicando que o modelo pode não ser o mais adequado aos dados.

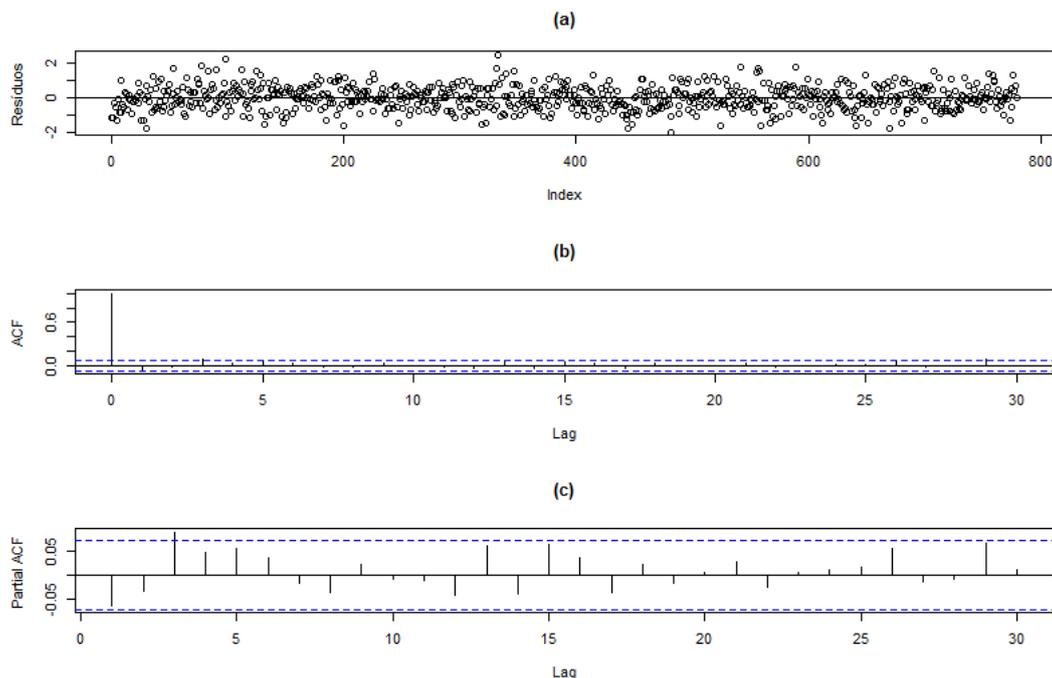


Figura 4.13: (a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo $AR-MV(3)$ para a série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).

Tabela 4.13: Estatísticas qui-quadrado do teste de Box-Pierce modificado para os lags 1 a 30 e respectivos p -valores dos resíduos do ajuste pelo modelo $AR-MV(3)$ da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).

Lag	Estatística	P-valor	Lag	Estatística	P-valor	Lag	Estatística	P-valor
1	3.06744	0.07988	11	15.44716	0.16293	21	28.67308	0.12207
2	3.69410	0.15770	12	16.49576	0.16957	22	29.14193	0.14085
3	10.24461	0.01660	13	19.48684	0.10877	23	29.14286	0.17562
4	11.21100	0.02429	14	21.07861	0.09963	24	29.51814	0.20124
5	12.66628	0.02672	15	23.70434	0.07029	25	29.58743	0.24011
6	13.53489	0.03529	16	24.72706	0.07478	26	31.67312	0.20415
7	13.70685	0.05665	17	26.76530	0.06160	27	32.23059	0.22365
8	14.21592	0.07631	18	27.84703	0.06442	28	32.23148	0.26519
9	15.25466	0.08417	19	27.86440	0.08607	29	37.38607	0.13660
10	15.29012	0.12184	20	27.88188	0.11223	30	37.39057	0.16603

Por meio da Figura 4.14, percebemos um ajuste satisfatório dos resíduos à distribuição normal. Todavia, para confirmar essa percepção visual, aplicamos os testes exibidos na Tabela 4.14, e notamos que em todos os testes, não rejeitamos a hipótese nula de normalidade dos dados.

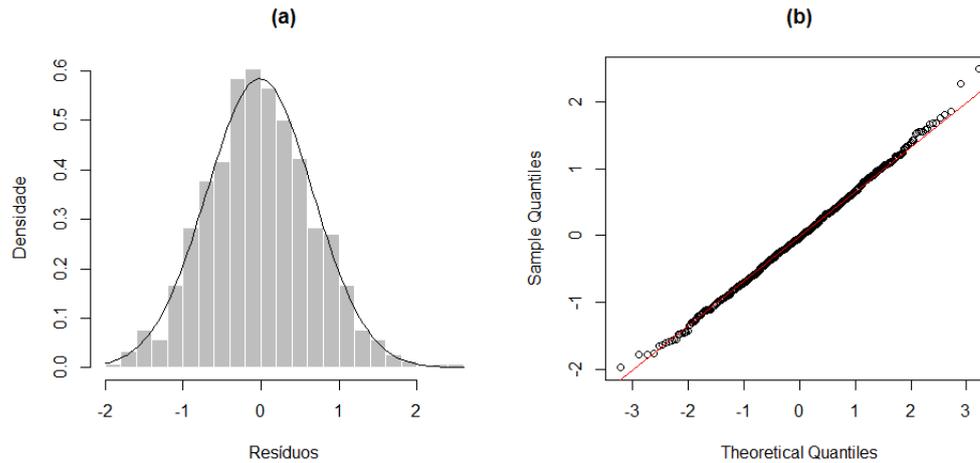


Figura 4.14: (a) Histograma e (b) gráfico Q-Q para a série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).

Tabela 4.14: Testes para a detecção de normalidade na série temporal de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978).

Teste	Estatística	P-valor
Kolmogorov-Smirnov	0.01932	0.93312
Lilliefors	0.01932	0.68433
Cramer-von Mises	0.03228	0.81443
Shapiro-Wilk	0.99840	0.70552
Shapiro-Francia	0.99840	0.63579
Anderson-Darling	0.20340	0.87623

4.2.7 Comparação dos modelos ARIMA e AR-MV ajustados

Como os modelos mostraram-se adequados, nesta seção realizaremos a comparação desses modelos, sob o ponto de vista da ajustabilidade aos dados. Na Tabela 4.15 estão reproduzidos os valores estimados para os parâmetros indicados. Além do que, nessa tabela estão os valores estabelecidos como critérios de comparação. Reparamos que, de acordo com todos os critérios de comparação utilizados, o modelo AR(3) é superior.

Tabela 4.15: Estimativas dos parâmetros e critérios de comparação, sob os modelos AR(3) e AR-MV(3), para a série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA (1197-1978). (O valor entre parênteses indica o erro-padrão de cada estimativa).

Estimativas/Critérios	AR-MV(3)	AR(3)
$\hat{\alpha}_1$	-1.12421	-
$\hat{\alpha}_2$	1.12674	-
$\hat{\phi}_1$	0.65916 (0.03024)	0.54920 (0.03555)
$\hat{\phi}_2$	0.15069 (0.03446)	0.14625 (0.04027)
$\hat{\phi}_3$	-0.05854 (0.06627)	0.09898 (0.03558)
$\hat{\sigma}^2$	0.46587	0.45241
SQR	362.91	353.78
AIC	1623.67	1607.79
BIC	1642.30	1626.43

4.2.8 Previsão

Nesta fase serão realizadas as previsões, sob os modelos AR(3) e AR-MV(3), para o período de 1979 a 1983, cujas observações verdadeiras (0.66027, 1.90207, 2.30745, 3.24649 e 2.20995) não foram usadas na fase da estimação, precisamente, na intenção de serem comparadas as capacidades preditivas dos dois modelos elegidos.

A Tabela 4.16 mostra as estimativas do MAE, MQE e os erros em cada uma das cinco etapas de previsões para os modelos em teste. Percebemos que os valores previstos pelo modelo AR(3) foram ligeiramente melhores em relação aos métodos de previsão referentes ao AR-MV(3).

Tabela 4.16: Estimativas do MAE, MQE e os erros de previsão a cada passo, sob os modelos AR(3) e AR-MV(3), para o período de 1979 a 1983, da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA.

Passo	AR(3) (sem intercepto)	AR-MV(3) (Esqueleto)	AR-MV(3) (Monte Carlo)
1	-0.05575	-0.09403	-0.09403
2	1.26153	1.27444	1.30993
3	1.76527	1.78007	1.79107
4	2.78418	2.80429	2.81070
5	1.81335	1.83900	1.86480
MAE	1.53602	1.55837	1.57411
MQE	3.15012	3.20953	3.26205

Finalmente, a Figura 4.15 mostra o gráfico que contém os valores verdadeiros observados para o período de 1979 a 1983 e as previsões geradas pelos dois modelos. Observamos que, embora o modelo AR(3) apresente as melhores predições para o período em questão, todos

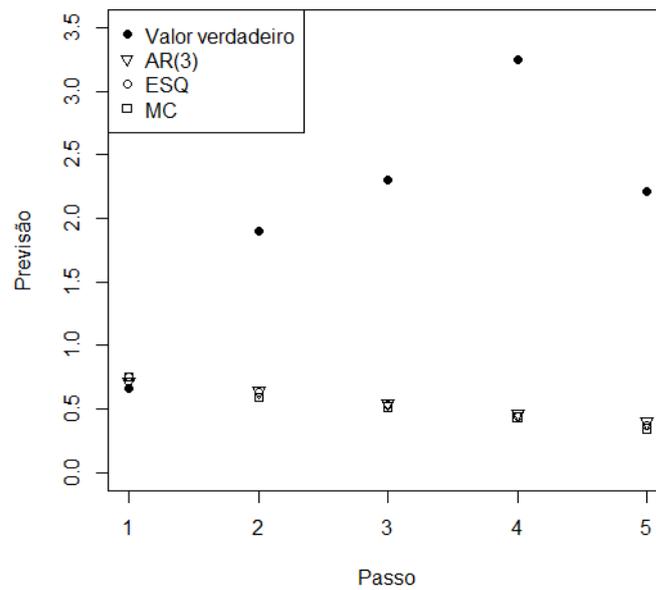


Figura 4.15: Valores observados e previsões pontuais dos modelos $AR(3)$ e $AR-MV(3)$ da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em San Francisco Peaks, Arizona, EUA, para o período de 1979 a 1983.

os métodos geraram estimativas muito próximas.

4.3 Aplicação III - AR-MV(4)

4.3.1 Banco de dados

Este banco de dados é composto por 1700 observações, atinentes ao período de 286 a 1985, contendo medições de anéis de crescimento de pinheiros “Bristlecone” (nome científico: *Pinus longaeva*), na região de Wild Horse Ridge, Utah, EUA.

4.3.2 Análise de normalidade dos dados

Nesta subseção será verificado o ajuste dos dados, apresentados na subseção anterior, a uma distribuição normal, com o objetivo de normalizar esse conjunto de observações, buscando-se viabilizar a estimação dos parâmetros limiares (α_1 , α_2 e α_3) pelo modelo AR-MV a ser ajustado posteriormente.

Em primeiro lugar, serão examinados o histograma e o gráfico Q-Q dos dados originais. A Figura 4.16 exhibe um ajuste satisfatoriamente apropriado a uma distribuição normal de probabilidade. Contudo, para ratificar essa percepção visual, serão utilizados os testes mostrados na Tabela 4.17, os quais podem ser verificados de maneira mais detalhada no Apêndice C.

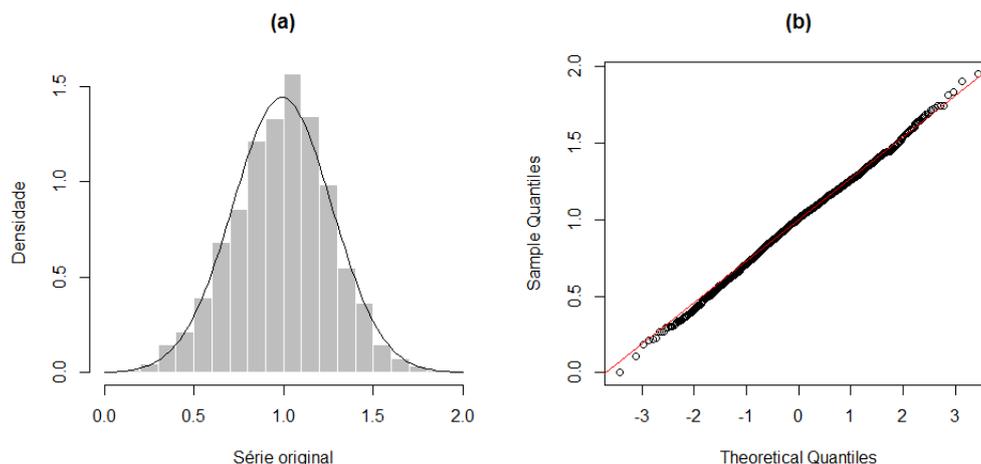


Figura 4.16: (a) Histograma e (b) gráfico Q-Q para a série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).

Tabela 4.17: Testes para a detecção de normalidade na série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).

Teste	Estatística	P-valor
Kolmogorov-Smirnov	0.02516	0.23335
Lilliefors	0.02516	0.01413
Cramer-von Mises	0.21050	0.00396
Shapiro-Wilk	0.99826	0.07362
Shapiro-Francia	0.99823	0.06224
Anderson-Darling	1.15827	0.00501

Segundo os dados mostrados na Tabela 4.17, há um empate. Dos seis testes efetivados, concluímos que em três os dados não são normais e que nos outros três, seguem uma distribuição normal, ao nível de significância de 5%.

Sendo assim, tomando como referência, além dos testes exibidos na Tabela 4.17, a Figura 4.16, é aceitável admitir normalidade para os dados sob análise.

4.3.3 Análise de estacionariedade dos dados

A Figura 4.17 exterioriza a série temporal dos dados em foco, permitindo-nos notar um padrão estacionário nos dados. Apesar disso, a fim de autenticar essa apreensão, serão empregados os testes mostrados na Tabela 4.18.

Tabela 4.18: Testes para a detecção de estacionariedade para a série de medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).

	ADF	PP	KPSS
Estatística	-9.344	-1230.4	0.036
P-valor	<0.01	<0.01	>0.10

Tomando como alicerce os testes mostrados na Tabela 4.18, ao nível de significância de

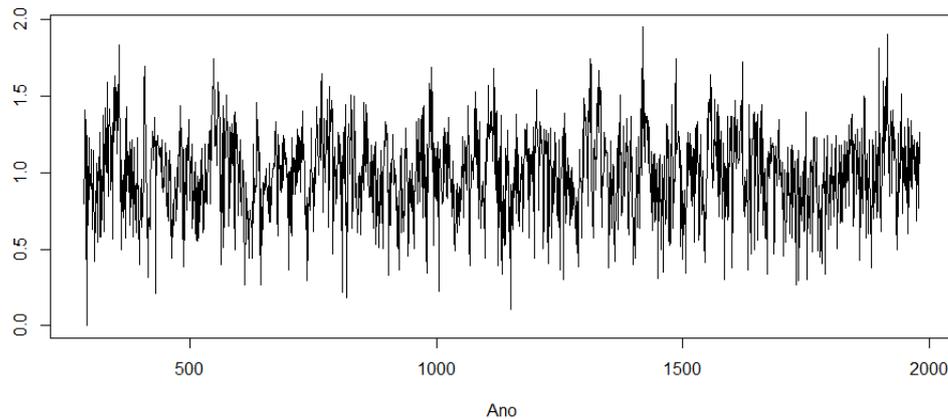


Figura 4.17: *Série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).*

5%, a estacionariedade dos dados é reafirmada. Nesse sentido, é possível estudar a FAC e a FACP da série com o intuito de identificar um modelo ARIMA coerente, para este conjunto de dados.

4.3.4 Identificação do modelo ARIMA

A Figura 4.18 exibe os gráficos da FAC e FACP dos dados. Observamos que a FAC ostenta um decaimento exponencial veloz para zero, padrão conhecido de um modelo autorregressivo; e a FACP exibe os quatro primeiros *lags* significativos, sugerindo que o AR(4) pode ser um modelo apropriado aos dados. Portanto, podemos postular o seguinte modelo

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4)Z_t = a_t. \quad (4.7)$$

Percebemos também que os *lags* 6 e 24 possuem uma modesta significância e, por essa razão, a adequabilidade da inserção dos parâmetros ϕ_6 e ϕ_{24} ao modelo dado pela equação (4.7) será investigada.

Na tabela 4.19 são expostas as estimativas dos coeficientes autorregressivos alusivos aos modelos AR(4), AR(6) e AR(24), assim como seus respectivos erros-padrão e p-valores. Salientamos que esses modelos, com o intercepto, não serão apresentados, porque esses coeficientes não são significativos para os dois casos. Reparamos que, sob o modelo AR(4), todos os coeficientes são significativos ao nível de 5%, atestando a conveniência desse modelo. Avaliando o modelo AR(6), observamos que os coeficientes ϕ_3 e ϕ_4 são não significantes à 5%, o que sugere o descarte desses parâmetros. No que diz respeito ao AR(24), o coeficiente ϕ_{24} não foi expressivo ao nível de 5% e, por isso, esse modelo será excluído da próxima análise.

Na Tabela 4.20 estão dispostos os valores dos coeficientes do modelo AR(6) estimado sem os coeficientes ϕ_3 e ϕ_4 , insignificantes, conforme a Tabela 4.19. Destacamos que os coeficientes são altamente significantes.

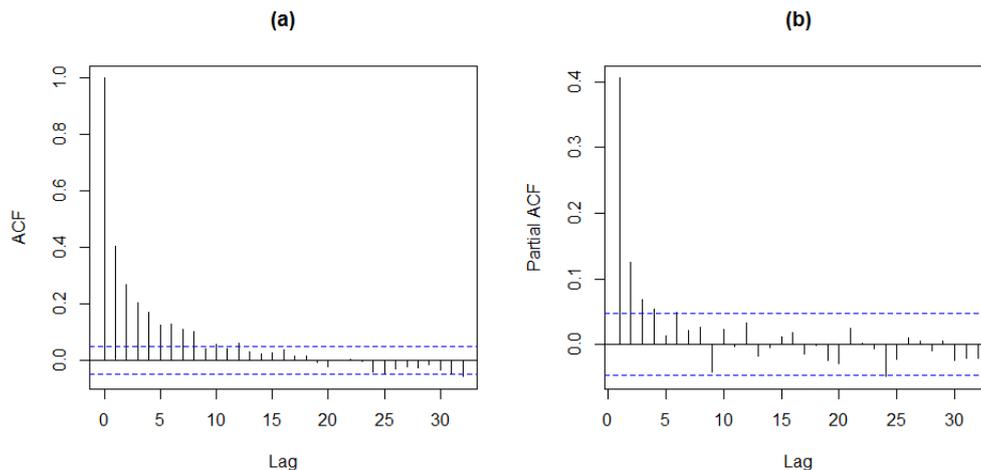


Figura 4.18: (a) FAC e (b) FACP amostrais para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).

Tabela 4.19: Estimativas dos parâmetros dos modelos $AR(4)$, $AR(6)$ e $AR(24)$ para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).

	AR(4)			AR(6)			AR(24)		
	Coefficiente	Erro-padrão	P-valor	Coefficiente	Erro-padrão	P-valor	Coefficiente	Erro-padrão	P-valor
ϕ_1	0.34202	0.02426	<0.01	0.34064	0.02424	<0.01	0.34176	0.02423	<0.01
ϕ_2	0.09566	0.02559	<0.01	0.09244	0.02566	<0.01	0.09593	0.02556	<0.01
ϕ_3	0.05020	0.02547	0.04872	0.04604	0.02554	0.07139	0.05059	0.02547	0.04703
ϕ_4	0.05408	0.02421	0.02555	0.04368	0.02470	0.07706	0.05302	0.02418	0.02831
ϕ_6	-	-	-	0.04837	0.02312	0.03644	-	-	-
ϕ_{24}	-	-	-	-	-	-	-0.03939	0.02203	0.07394

Tabela 4.20: Estimativas dos parâmetros do modelo $AR(6)$, sem os coeficientes ϕ_3 e ϕ_4 , para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).

AR(6)	Coefficiente	Erro-padrão	P-valor
ϕ_1	0.35039	0.02408	<0.01
ϕ_2	0.11617	0.02425	<0.01
ϕ_6	0.06475	0.02244	<0.01

Como ferramenta de apoio à decisão de escolha do modelo mais apropriado aos dados, a Tabela 4.21 exprime alguns critérios de comparação entre os dois modelos sob análise. Percebemos que o $AR(4)$ é superior ao $AR(6)$ em três dos quatro critérios averiguados, conforme destacado em negrito.

Tabela 4.21: Critérios de comparação entre os modelos $AR(4)$ e $AR(6)$ para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).

Critérios	AR(4)	AR(6)
$\hat{\sigma}^2$	0.80984	0.81208
SQR	1372.67	1376.48
AIC	4462.92	4465.62
BIC	4490.10	4487.36

Diante do que foi visto, resolvemos escolher o $AR(4)$ como o modelo mais adequado

para o ajustamento dos dados em questão, cujas estimativas encontram-se sumarizadas na seguinte equação

$$(1 - 0.34202B - 0.09566B^2 - 0.05020B^3 - 0.05408B^4)Z_t = a_t. \quad (4.8)$$

4.3.5 Diagnóstico do modelo ARIMA selecionado

Primeiramente, na Figura 4.19 são exibidos os gráficos de dispersão, FAC e FACP dos resíduos. Visualizamos que os dados se distribuem aleatoriamente em torno de zero no gráfico de dispersão. Além do mais, os valores das autocorrelações e autocorrelações parciais não indicam nenhuma alteração no comportamento de ruído branco dos resíduos, uma vez que não há defasagens significativas.

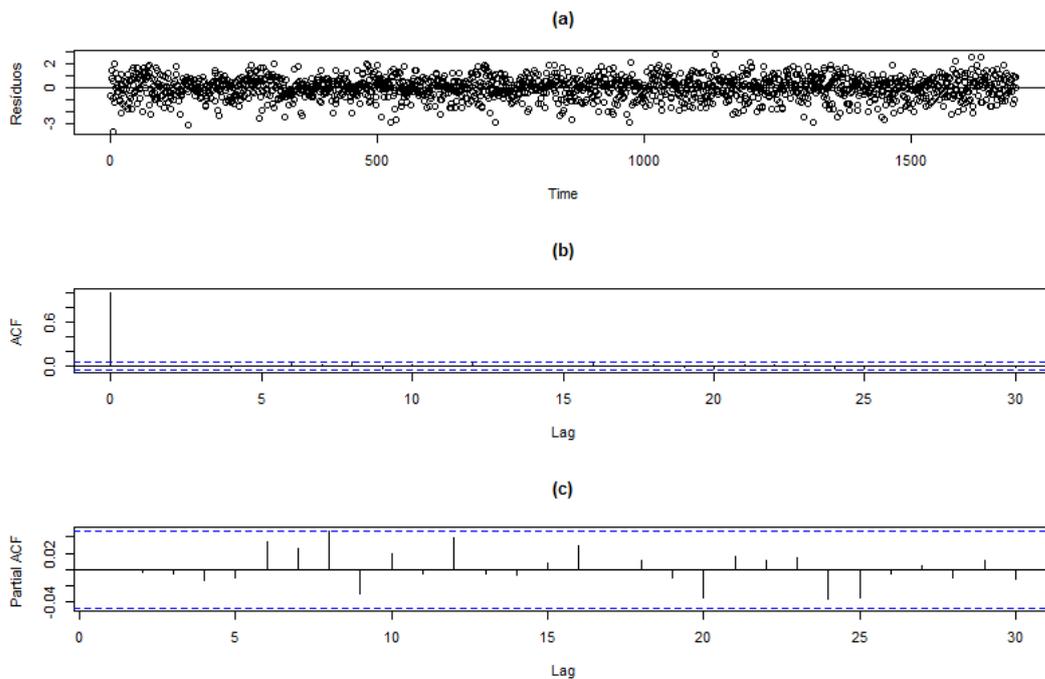


Figura 4.19: (a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo $AR(4)$ para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).

Seguidamente, a Tabela 4.22 retrata os valores das estatísticas qui-quadrado do teste de Box-Pierce modificado para os lags 1 a 30, e seus respectivos p-valores. Como não há estatísticas significantes, ao nível de 5%, afirmamos que os resíduos são ruído branco.

Tabela 4.22: Estatísticas qui-quadrado do teste de Box-Pierce modificado para os lags 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo AR(4) da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).

Lag	Estatística	P-valor	Lag	Estatística	P-valor	Lag	Estatística	P-valor
1	0.00045	0.98299	11	9.72874	0.55494	21	16.59853	0.73514
2	0.01890	0.99060	12	12.27754	0.42366	22	16.94586	0.76634
3	0.07958	0.99417	13	12.30243	0.50301	23	17.41152	0.78858
4	0.41990	0.98082	14	12.32167	0.58049	24	19.04256	0.74972
5	0.58973	0.98847	15	12.40727	0.64798	25	21.40928	0.66958
6	2.64130	0.85233	16	13.93976	0.60320	26	21.47074	0.71728
7	3.74175	0.80900	17	13.96561	0.66953	27	21.47185	0.76371
8	7.56186	0.47739	18	14.35772	0.70549	28	21.56114	0.80119
9	9.09531	0.42852	19	14.55930	0.75017	29	22.00702	0.81990
10	9.62427	0.47405	20	16.35970	0.69407	30	22.26042	0.84440

4.3.6 Estimação e diagnóstico do modelo AR-MV

Dado que o ajuste de um modelo AR(4) é adequado aos dados, o modelo AR-MV(4) também será ajustado. Ademais, para examinar a adequação do modelo AR-MV(4) aos dados em estudo, os resíduos desse modelo serão averiguados.

As estimativas dos parâmetros do modelo AR-MV(4) ajustado encontram-se resumidas no seguinte sistema de equações

$$Z_t = \begin{cases} 0.338Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq -1.26 \\ 0.338Z_{t-1} + 0.130Z_{t-2} + a_t, & \text{se } -1.26 < Z_{t-1} \leq -0.05 \\ 0.338Z_{t-1} + 0.130Z_{t-2} + 0.02Z_{t-3} + a_t, & \text{se } -0.05 < Z_{t-1} \leq 1.27 \\ 0.338Z_{t-1} + 0.130Z_{t-2} + 0.02Z_{t-3} + 0.230Z_{t-4} + a_t, & \text{se } Z_{t-1} > 1.27 \end{cases} \quad (4.9)$$

Na Figura 4.20 são visualizados os gráficos de dispersão, FAC e FACP dos resíduos. Notamos que os dados se distribuem aleatoriamente em torno de zero no gráfico de dispersão. Além do mais, os valores das autocorrelações e autocorrelações parciais não sugerem nenhuma alteração no comportamento de ruído branco dos resíduos, uma vez que não há lags expressivos.

Em seguida, a Tabela 4.23 exibe os valores das estatísticas qui-quadrado do teste de Box-Pierce modificado para os lags 1 a 30, e seus respectivos p-valores. Nenhuma defasagem é significativa ao nível de 5%, o que reafirma a impressão obtida ao avaliar a Figura 4.20, ou seja, os resíduos do modelo AR-MV(4) para estes dados são ruído branco.

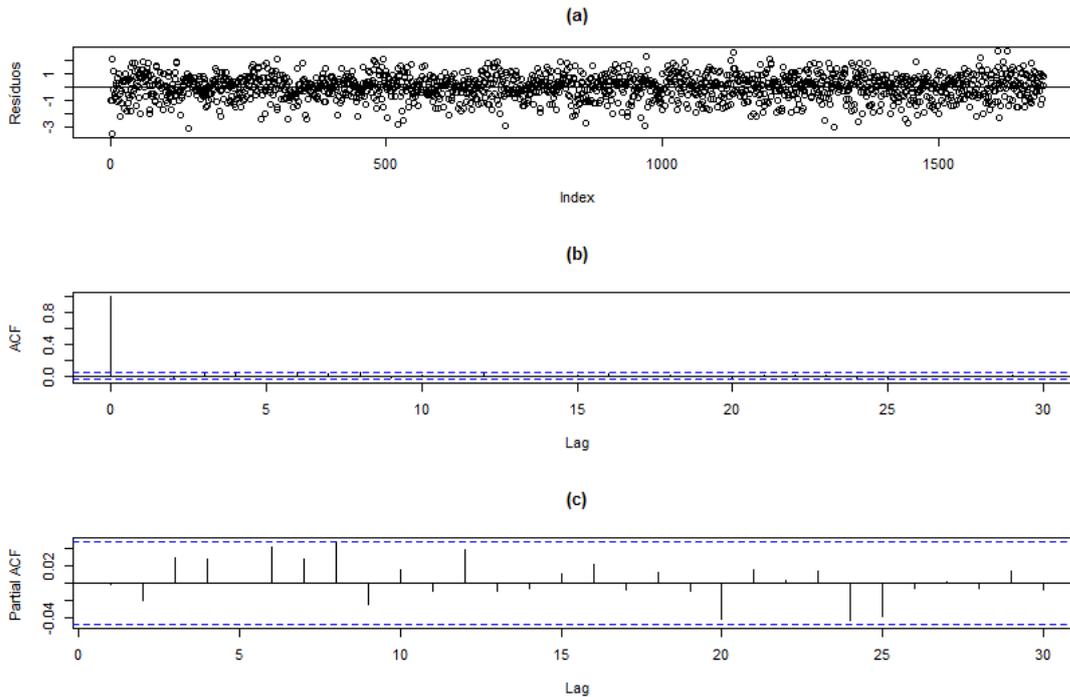


Figura 4.20: (a) Dispersão, (b) FAC e (c) FACP amostrais dos resíduos do ajuste pelo modelo AR-MV(4) para a série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).

Tabela 4.23: Estatísticas qui-quadrado do teste de Box-Pierce modificado para os lags 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo AR-MV(4) da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).

Lag	Estatística	P-valor	Lag	Estatística	P-valor	Lag	Estatística	P-valor
1	0.00320	0.95491	11	12.91744	0.29876	21	20.57463	0.48517
2	0.66065	0.71869	12	15.76762	0.20211	22	20.69997	0.53934
3	2.12547	0.54678	13	15.83873	0.25795	23	20.90824	0.58669
4	3.49475	0.47868	14	15.84808	0.32274	24	23.59995	0.48465
5	3.49960	0.62345	15	16.15168	0.37204	25	26.13701	0.40032
6	6.45792	0.37389	16	17.21907	0.37156	26	26.16028	0.45432
7	7.92454	0.33930	17	17.40181	0.42748	27	26.16085	0.50969
8	11.51757	0.17406	18	17.91002	0.46159	28	26.32488	0.55519
9	12.39558	0.19192	19	17.98053	0.52374	29	26.68055	0.58894
10	12.88944	0.22992	20	20.26228	0.44163	30	26.82521	0.63243

Por meio da Figura 4.21, percebemos um ajuste inadequado dos resíduos à distribuição normal. Todavia, para confirmar essa percepção visual, aplicamos os testes exibidos na Tabela 4.24, e notamos que em todos os testes, rejeitamos a hipótese nula de normalidade dos dados.

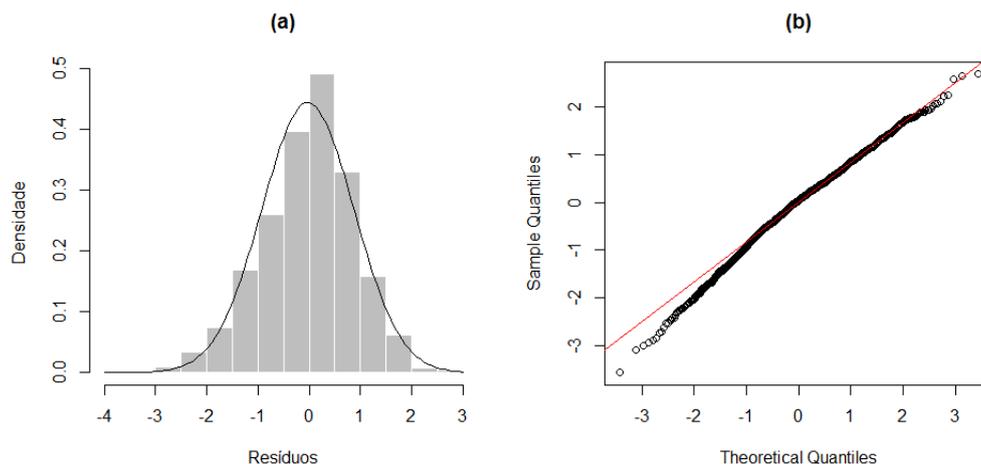


Figura 4.21: (a) Histograma e (b) gráfico Q-Q para a série temporal gerada para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).

Tabela 4.24: Testes para a detecção de normalidade na série temporal das medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980).

Teste	Estatística	P-valor
Kolmogorov-Smirnov	0.04054	<0.01
Lilliefors	0.04054	<0.01
Cramer-von Mises	0.63470	<0.01
Shapiro-Wilk	0.99280	<0.01
Shapiro-Francia	0.99279	<0.01
Anderson-Darling	3.55373	<0.01

4.3.7 Comparação dos modelos ARIMA e AR-MV ajustados

Como os modelos mostraram-se adequados, nesta seção realizaremos a comparação desses modelos, sob o ponto de vista da ajustabilidade aos dados. Na Tabela 4.25 estão descritos os valores estimados para os parâmetros indicados. Outrossim, nessa tabela encontramos os valores estabelecidos como critérios de comparação. Notamos que, diferentemente dos cenários anteriores, nesta aplicação o AR-MV(4) foi superior ao AR(4) em todos os critérios de comparação utilizados.

Tabela 4.25: Estimativas dos parâmetros e critérios de comparação, sob os modelos $AR(4)$ e $AR-MV(4)$, para as medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA (286-1980). (O valor entre parênteses indica o erro-padrão de cada estimativa).

Estimativas/Critérios	AR-MV(4)	AR(4)
$\hat{\alpha}_1$	-1.26299	-
$\hat{\alpha}_2$	-0.05258	-
$\hat{\alpha}_3$	1.26677	-
$\hat{\phi}_1$	0.33776 (0.02413)	0.34202 (0.02426)
$\hat{\phi}_2$	0.12993 (0.02601)	0.09566 (0.02559)
$\hat{\phi}_3$	0.02124 (0.03200)	0.05020 (0.02547)
$\hat{\phi}_4$	0.22887 (0.07001)	0.05408 (0.02421)
$\hat{\sigma}^2$	0.80711	0.80984
SQR	1364.82	1372.67
AIC	4446.47	4462.92
BIC	4473.63	4490.10

4.3.8 Previsão

Neste ponto serão executadas as previsões, sob os modelos $AR(4)$ e $AR-MV(4)$, para o período de 1981 a 1985, cujas observações verdadeiras (1.76265, 1.79868, 1.27975, 1.95365 e 2.25636) não foram utilizadas na fase da estimação, exatamente no intuito de serem comparadas as capacidades preditivas dos dois modelos escolhidos.

A Tabela 4.26 mostra as estimativas do MAE, MQE e os erros em cada uma das cinco etapas de previsões para os modelos em teste. Reparamos que os valores previstos pelo modelo $AR(4)$ foram superiores às previsões geradas pelos métodos concernentes ao modelo $AR-MV(4)$.

Tabela 4.26: Estimativas do MAE, MQE e os erros de previsão a cada passo, sob os modelos $AR(4)$ e $AR-MV(4)$, para o período de 1981 a 1985, da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA.

Passo	AR(4) (sem intercepto)	AR-MV(4) (Esqueleto)	AR-MV(4) (Monte Carlo)
1	1.42152	1.39201	1.39201
2	1.54369	1.54290	1.54985
3	1.09851	1.12467	1.13657
4	1.79789	1.86016	1.79777
5	2.15450	2.19920	2.19770
MAE	1.60322	1.62379	1.61478
MQE	2.69694	2.77596	2.73868

Por último, a Figura 4.22 expõe o gráfico que contém os valores verdadeiros observados para o período de 1981 a 1985 e as previsões geradas pelos dois modelos. Apesar de o modelo AR(4) ser superior ao AR-MV(4) neste quesito, os valores previstos para ambos os modelos foram muito próximos, subestimando os valores verdadeiros das observações para o período em questão.

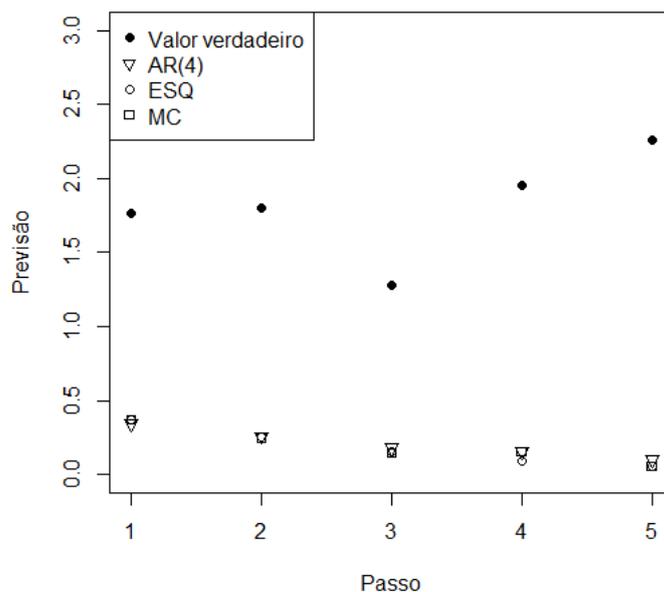


Figura 4.22: Valores observados e previsões pontuais dos modelos AR(4) e AR-MV(4) da série temporal padronizada de medições de anéis de crescimento de pinheiros “Bristlecone” em Wild Horse Ridge, Utah, EUA, para o período de 1981 a 1985.

4.4 Aplicação IV - AR-MV(5)

4.4.1 Banco de dados

Este banco de dados é composto por 411 observações, referentes ao período de 1021 a.C. a 611 a.C, contendo medições de anéis de crescimento do pinheiro “Bristlecone” (nome científico: *Pinus longaeva*), na região de Indian Garden, Nevada, EUA.

4.4.2 Análise de normalidade dos dados

Nesta subseção será avaliada a aderência dos dados, apresentados na subseção anterior, a uma distribuição normal, tendo em vista que esse conjunto de observações será padronizado, no intuito de viabilizar a estimação dos parâmetros limiares (α_1 , α_2 , α_3 e α_4) pelo modelo AR-MV a ser ajustado posteriormente.

Inicialmente, serão analisados o histograma e o gráfico Q-Q dos dados originais. Percebemos, por meio da Figura 4.23, que o ajuste segue satisfatoriamente uma distribuição

normal de probabilidade. Contudo, para ratificar essa percepção visual, serão utilizados os testes mostrados na Tabela 4.27, os quais podem ser verificados de maneira mais detalhada no Apêndice C.

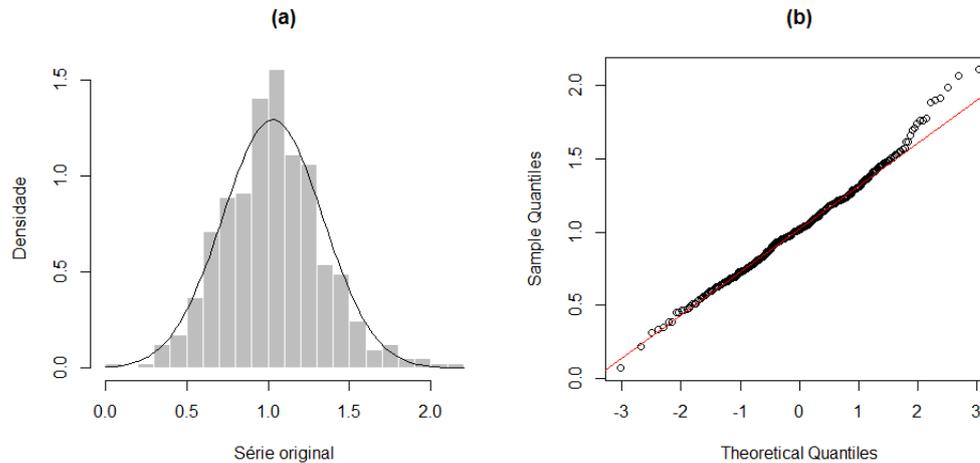


Figura 4.23: (a) Histograma e (b) gráfico Q-Q para a série temporal gerada para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).

Tabela 4.27: Testes para a detecção de normalidade na série temporal gerada para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).

Teste	Estatística	P-valor
Kolmogorov-Smirnov	0.04054	0.51686
Lilliefors	0.04054	0.10774
Cramer-von Mises	0.09876	0.11653
Shapiro-Wilk	0.99113	0.01551
Shapiro-Francia	0.99050	0.01158
Anderson-Darling	0.68834	0.07176

De acordo com a Tabela 4.27, com exceção dos testes Shapiro-Wilk e Shapiro-Francia, ao nível de significância de 5%, não rejeitamos a hipótese nula de normalidade dos dados.

Portanto, tomando como referência, além dos testes exibidos na Tabela 4.27, a Figura 4.23, é razoável supor normalidade para os dados sob análise.

4.4.3 Análise de estacionariedade dos dados

A Figura 4.24 exibe a série temporal dos dados em questão, fazendo-nos perceber um padrão estacionário nos dados. Contudo, para ratificar essa percepção visual, serão utilizados os testes mostrados na Tabela 4.28.

Com base nos testes exibidos na Tabela 4.28, ao nível de significância de 5%, confirmamos a estacionariedade dos dados. Dessa forma, é possível analisar a FAC e a FACP da série com o intuito de identificar um modelo ARIMA adequado, para este conjunto de dados.

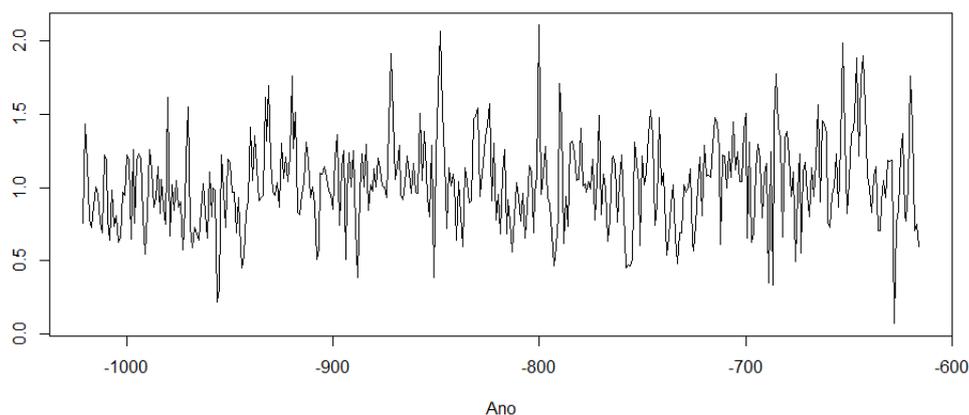


Figura 4.24: *Série temporal gerada para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).*

Tabela 4.28: *Testes para a detecção de estacionariedade na série de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).*

	ADF	PP	KPSS
Estatística	-6.13	-281.59	0.456
P-valor	<0.01	<0.01	0.053

4.4.4 Identificação do modelo ARIMA

A Figura 4.25 expõe os gráficos da FAC e FACP dos dados. Percebemos que a FAC apresenta um decaimento exponencial veloz para zero, comportamento próprio de um modelo autorregressivo; e a FACP possui o primeiro e o quinto *lags* significativos, indicando que o AR(5) incompleto, sem os coeficientes ϕ_2 , ϕ_3 e ϕ_4 , pode ser um modelo adequado para os dados. Assim, podemos postular o seguinte modelo

$$(1 - \phi_1 B - \phi_5 B^5) Z_t = a_t. \quad (4.10)$$

Na Tabela 4.29 estão dispostos os valores dos coeficientes do modelo AR(5) estimado sem os coeficientes ϕ_2 , ϕ_3 e ϕ_4 , não significativos, conforme a Figura 4.25. Notamos que os coeficientes ϕ_1 e ϕ_5 são significantes ao nível de 5%.

Tabela 4.29: *Estimativas dos parâmetros do modelo AR(5), sem os coeficientes ϕ_2 , ϕ_3 e ϕ_4 , para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).*

AR(5)	Coefficiente	Erro-padrão	P-valor
ϕ_1	0.30269	0.04705	<0.01
ϕ_5	0.10802	0.04755	0.02309

Diante do exposto, decidimos por selecionar o AR(5) como o modelo mais adequado para ajustamento dos dados em questão, cujas estimativas encontram-se sumarizadas na seguinte equação

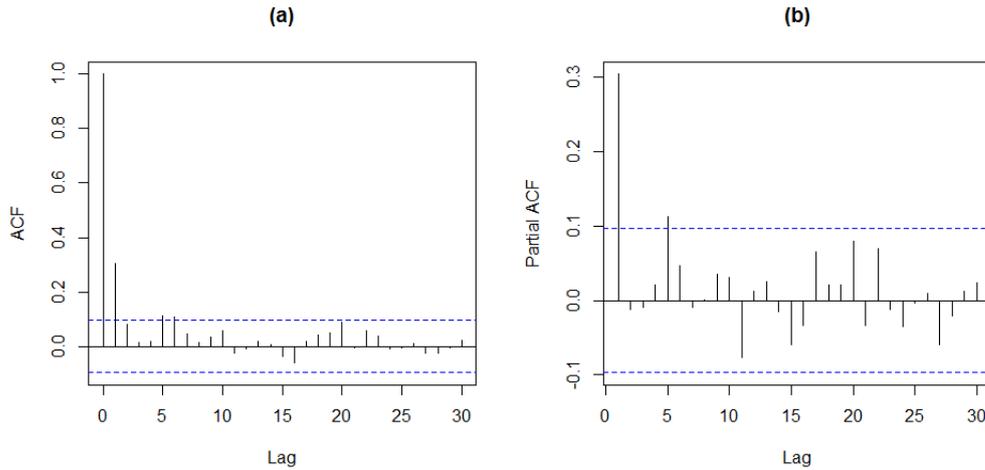


Figura 4.25: (a) FAC e (b) FACP amostrais para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).

$$(1 - 0.30269B - 0.10802B^5)Z_t = a_t. \quad (4.11)$$

4.4.5 Diagnóstico do modelo ARIMA selecionado

Inicialmente, na Figura 4.26 são apresentados os gráficos de dispersão, FAC e FACP dos resíduos. Observamos que os dados se distribuem aleatoriamente em torno de zero no gráfico de dispersão. Ademais, os valores das autocorrelações e autocorrelações parciais não indicam nenhuma alteração no comportamento de ruído branco dos resíduos, dado que não há *lags* significativos.

A seguir, a Tabela 4.30 apresenta os valores das estatísticas qui-quadrado do teste de Box-Pierce modificado para os *lags* 1 a 30, e seus respectivos p-valores. Como não há estatísticas significantes, ao nível de 5%, podemos afirmar que os resíduos são ruído branco.

Tabela 4.30: Estatísticas qui-quadrado do teste de Box-Pierce modificado para os *lags* 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo AR(5) da série temporal padronizada de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).

Lag	Estatística	P-valor	Lag	Estatística	P-valor	Lag	Estatística	P-valor
1	0.00040	0.98414	11	4.71656	0.94411	21	12.73045	0.91777
2	0.02505	0.98755	12	4.83303	0.96332	22	14.29516	0.89067
3	0.12724	0.98838	13	4.99202	0.97537	23	14.66847	0.90612
4	0.34679	0.98660	14	4.99992	0.98581	24	14.96631	0.92174
5	0.49822	0.99219	15	5.80872	0.98278	25	15.11527	0.93865
6	1.66562	0.94774	16	7.28351	0.96740	26	15.47546	0.94801
7	1.67511	0.97562	17	7.51522	0.97585	27	15.99549	0.95303
8	1.75191	0.98770	18	7.74413	0.98221	28	16.13985	0.96370
9	1.90607	0.99287	19	7.90658	0.98759	29	16.15211	0.97377
10	3.53578	0.96587	20	11.71182	0.92564	30	16.68751	0.97614

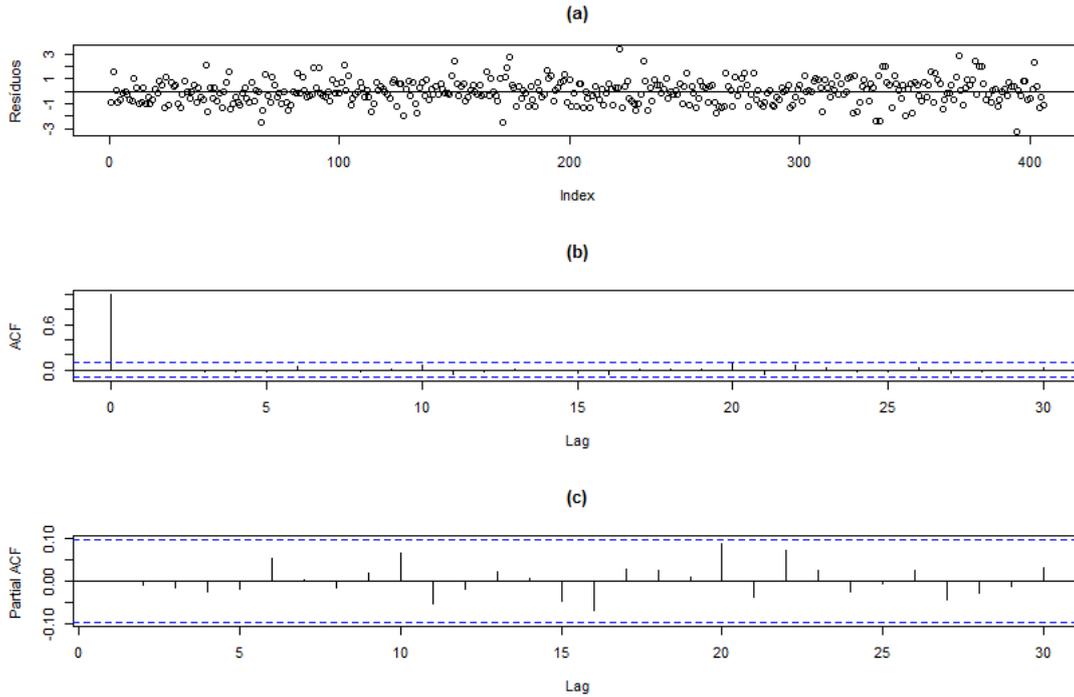


Figura 4.26: (a) *Dispersão*, (b) *FAC* e (c) *FACP amostrais* dos resíduos do ajuste pelo modelo $AR(5)$ para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.-616 a.C.).

4.4.6 Estimação e diagnóstico do modelo AR-MV

Dado que o ajuste de um modelo $AR(5)$ é adequado aos dados, o modelo $AR-MV(5)$ também será ajustado. Ademais, para verificar a adequação do modelo $AR-MV(5)$ aos dados em estudo, os resíduos desse modelo serão avaliados.

As estimativas dos parâmetros do modelo $AR-MV(5)$ ajustado encontram-se resumidas no seguinte sistema de equações

$$Z_t = \begin{cases} 0.314Z_{t-1} + a_t, & \text{se } Z_{t-1} \leq -0.230 \\ 0.314Z_{t-1} - 0.040Z_{t-2} + a_t, & \text{se } -0.230 < Z_{t-1} \leq 0.458 \\ 0.314Z_{t-1} - 0.040Z_{t-2} - 0.021Z_{t-3} + a_t, & \text{se } 0.458 < Z_{t-1} \leq 0.655 \\ 0.314Z_{t-1} - 0.040Z_{t-2} - 0.021Z_{t-3} - 0.114Z_{t-4} + a_t, & \text{se } 0.655 < Z_{t-1} \leq 0.964 \\ 0.314Z_{t-1} - 0.040Z_{t-2} - 0.021Z_{t-3} - 0.114Z_{t-4} + 0.227Z_{t-5} + a_t, & \text{se } Z_{t-1} > 0.964 \end{cases} \quad (4.12)$$

Na Figura 4.27 são apresentados os gráficos de dispersão, FAC e FACP dos resíduos. Observamos que os dados se distribuem aleatoriamente em torno de zero no gráfico de dispersão. Notamos também que os valores das autocorrelações e autocorrelações parciais não indicam alteração no comportamento de ruído branco dos resíduos, com exceção de uma pequena significância no *lag* 20.

A seguir, a Tabela 4.31 apresenta os valores das estatísticas qui-quadrado do teste de Box-Pierce modificado para os *lags* 1 a 30, e seus respectivos p-valores. Percebemos que nenhum *lag* é significativo ao nível de 5%, eliminando a dúvida que surgiu ao analisar a Figura 4.27.

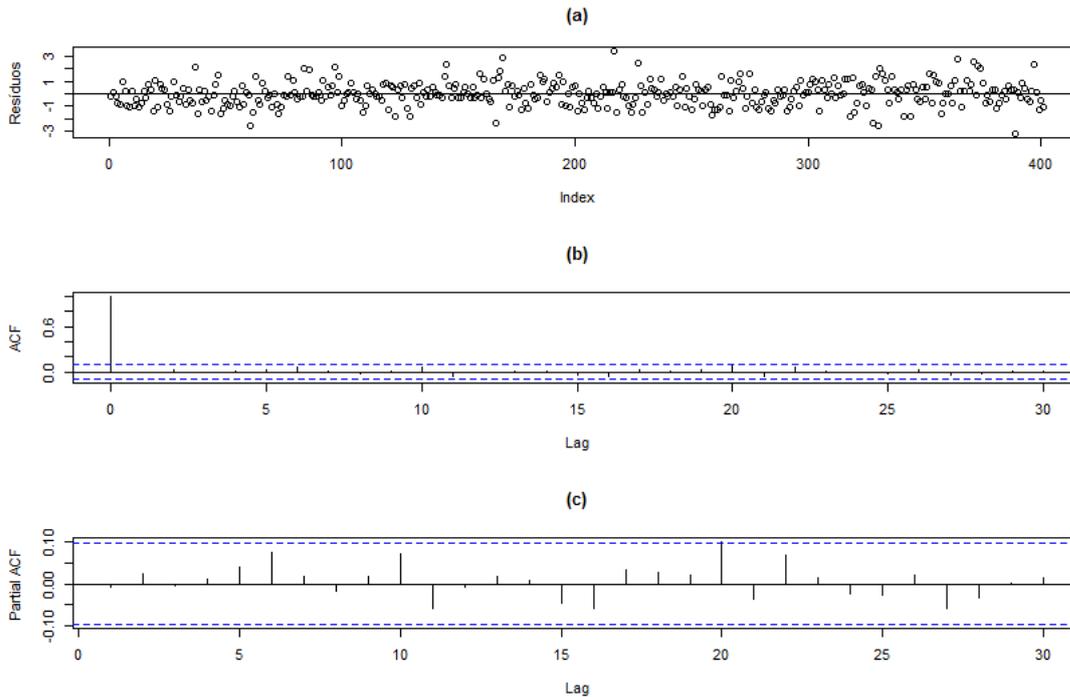


Figura 4.27: (a) *Dispersão*, (b) *FAC* e (c) *FACP amostrais* dos resíduos do ajuste pelo modelo AR-MV(5) para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.)

Assim, podemos afirmar que os resíduos do modelo AR-MV(5), para estes dados, são ruído branco.

Tabela 4.31: *Estatísticas qui-quadrado do teste de Box-Pierce modificado para os lags 1 a 30 e respectivos p-valores dos resíduos do ajuste pelo modelo AR-MV(5) da série temporal padronizada de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).*

Lag	Estatística	P-valor	Lag	Estatística	P-valor	Lag	Estatística	P-valor
1	0.01844	0.89198	11	7.04559	0.79538	21	15.07909	0.81896
2	0.24842	0.88319	12	7.05031	0.85427	22	16.97911	0.76451
3	0.25850	0.96763	13	7.14494	0.89453	23	17.12598	0.80302
4	0.31645	0.98873	14	7.18210	0.92746	24	17.18271	0.84077
5	0.91349	0.96924	15	7.74129	0.93374	25	17.39636	0.86692
6	3.24818	0.77711	16	8.72755	0.92421	26	17.76958	0.88399
7	3.40203	0.84549	17	8.96208	0.94144	27	18.59026	0.88445
8	3.48066	0.90069	18	9.14401	0.95631	28	18.79944	0.90421
9	3.61604	0.93482	19	9.55529	0.96308	29	18.83722	0.92539
10	5.85291	0.82747	20	13.78247	0.84135	30	18.95255	0.94100

Por meio da Figura 4.28, percebemos um ajuste satisfatório dos resíduos à distribuição normal. Todavia, para confirmar essa percepção visual, aplicamos os testes exibidos na Tabela 4.32, e notamos que houve empate. Dos seis testes aplicados, três rejeitam a hipótese de normalidade dos dados ao nível de 5%. Com base nesses resultados é razoável aceitar que os dados seguem uma distribuição normal de probabilidade.

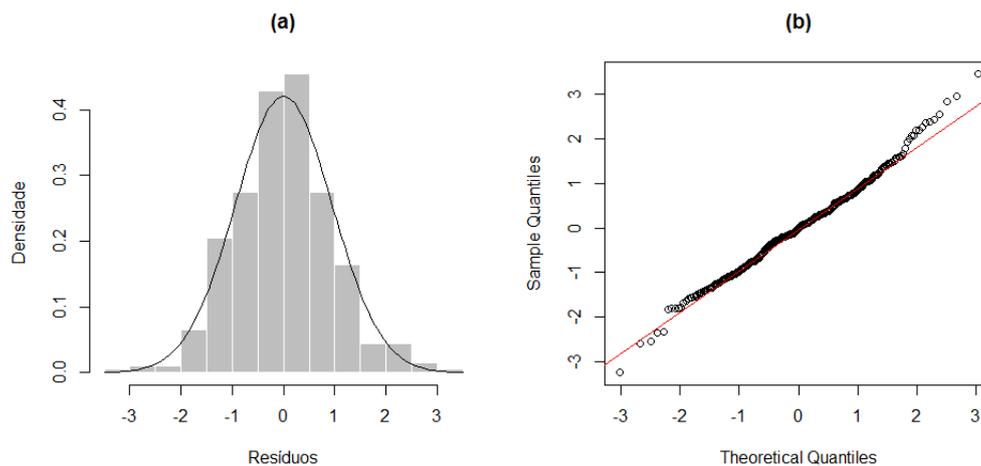


Figura 4.28: (a) Histograma e (b) gráfico Q-Q para a série temporal de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).

Tabela 4.32: Testes para a detecção de normalidade na série temporal de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.).

Teste	Estatística	P-valor
Kolmogorov-Smirnov	0.04171	0.48805
Lilliefors	0.04171	0.09106
Cramer-von Mises	0.11452	0.07090
Shapiro-Wilk	0.99190	0.02791
Shapiro-Francia	0.99077	0.01454
Anderson-Darling	0.75956	0.04783

4.4.7 Comparação dos modelos ARIMA e AR-MV ajustados

Como os modelos mostraram-se adequados, nesta seção realizaremos a comparação desses modelos, sob o ponto de vista da ajustabilidade aos dados. Na Tabela 4.33 estão descritos os valores estimados para os parâmetros indicados. Além disso, nessa tabela encontramos os valores estabelecidos como critérios de comparação. Observamos que, apenas sob o critério BIC, o modelo AR(5) é melhor. Nos outros critérios de comparação, o modelo AR-MV(5) é superior.

Tabela 4.33: Estimativas dos parâmetros e critérios de comparação, sob os modelos AR(5) e AR-MV(5), para as medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA (1021 a.C.- 616 a.C.). (O valor entre parênteses indica o erro-padrão de cada estimativa).

Estimativas/Critérios	AR-MV(5)	AR(5)
$\hat{\alpha}_1$	-0.22987	-
$\hat{\alpha}_2$	0.45788	-
$\hat{\alpha}_3$	0.65514	-
$\hat{\alpha}_4$	0.96424	-
$\hat{\phi}_1$	0.31396 (0.04887)	0.30269 (0.04705)
$\hat{\phi}_2$	-0.04037 (0.06577)	- (-)
$\hat{\phi}_3$	-0.02132 (0.08204)	- (-)
$\hat{\phi}_4$	-0.11430 (0.09947)	- (-)
$\hat{\phi}_5$	0.22698 (0.10705)	0.10802 (0.04755)
$\hat{\sigma}^2$	0.89885	0.90077
SQR	360.44	365.71
AIC	1107.23	1115.91
BIC	1131.19	1127.93

4.4.8 Previsão

Nesta etapa serão realizadas as previsões, sob os modelos AR(5) e AR-MV(5), para o período de 615 a.C. a 611 a.C., cujas observações verdadeiras (-0.91348, 0.05836, 0.11036, 0.81243 e 0.13636) não foram utilizadas na fase da estimação, justamente, com o objetivo de serem comparadas as capacidades preditivas dos dois modelos selecionados.

A Tabela 4.34 exibe as estimativas do MAE, MQE e os erros em cada uma das cinco etapas de previsões para os modelos em teste. Notamos que os valores previstos pelos métodos referentes ao modelo AR-MV(5) foram superiores às previsões geradas pelo modelo AR(5).

Tabela 4.34: Estimativas do MAE, MQE e os erros de previsão a cada passo, sob os modelos AR(5) e AR-MV(5), para o período de 615 a.C. a 611 a.C., da série temporal padronizada de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA.

Passo	AR(5) (sem intercepto)	AR-MV(5) (Esqueleto)	AR-MV(5) (Monte Carlo)
1	-0.74773	-0.47464	-0.47464
2	-0.02663	0.19614	0.21896
3	0.19876	0.13590	0.21888
4	0.93680	0.81488	0.81347
5	0.32500	0.13610	0.17149
MAE	0.44698	0.35153	0.37949
MQE	0.31651	0.19296	0.20246

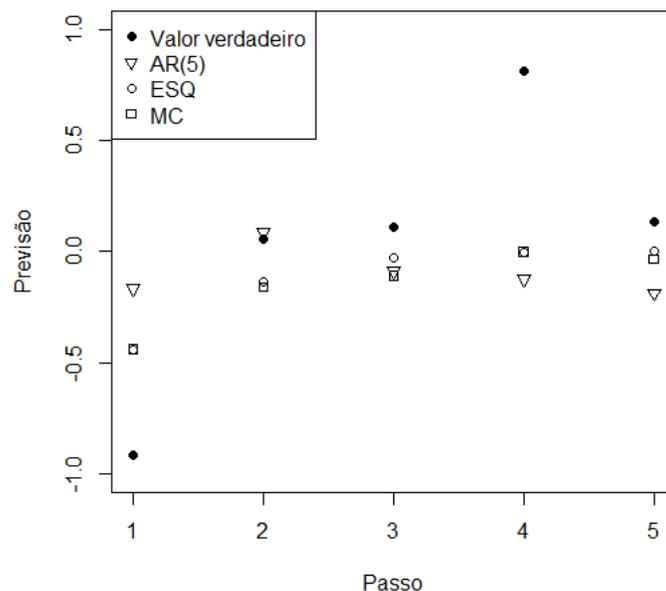


Figura 4.29: Valores observados e previsões pontuais dos modelos $AR(5)$ e $AR-MV(5)$ da série temporal padronizada de medições de anéis de crescimento do pinheiro “Bristlecone” em Indian Garden, Nevada, EUA, para o período de 615 a.C. a 611 a.C.

Por fim, a Figura 4.29 expõe o gráfico contendo os valores verdadeiros observados para o período de 615 a.C. a 611 a.C. e as previsões geradas pelos dois modelos. Notamos que os métodos de predição atinentes ao modelo $AR-MV(5)$ foram superiores ao $AR(5)$, sendo que o Esqueleto gerou as melhores previsões.

Capítulo 5

Considerações finais

Neste trabalho generalizamos os estudos do modelo autorregressivo com memória variável (AR-MV) introduzidos por Fadel (2012), que restringiu suas avaliações aos modelos de ordens mais baixas, quais sejam, 2 e 3. Foi de grande interesse investigar o fato de que o comportamento das FAC e FACP dos modelos AR-MV possuem padrões similares a essas mesmas funções para os modelos AR. Dessa forma, estudamos de maneira aprofundada, principalmente por meio de estudos simulados, modelos de ordens superiores, especificamente, os modelos AR-MV(4) e AR-MV(5) e constatamos a manutenção dessa propriedade também nesses modelos, o que sugere que essa deva ser uma particularidade dos AR-MV.

Na Seção 3.2 avaliamos o comportamento dos modelos AR-MV, tendo em vista a demonstração gráfica desses modelos nos cenários onde a propriedade de ergodicidade geométrica desses modelos é confirmada e quando é violada. Vimos que quando o somatório dos parâmetros autorregressivos do modelo é menor que 1 ($\sum_{i=1}^p |\phi_i| < 1$), a série temporal gerada é estacionária e os dados são normais, o que não acontece na situação oposta.

No campo dos estudos simulados (Seção 3.5), os modelos AR-MV(4) e AR-MV(5) foram melhores que os seus rivais, AR(4) e AR(5), respectivamente, na questão da ajustabilidade dos modelos aos dados. No que concerne à capacidade preditiva dos modelos, constatamos que os modelos AR-MV foram melhores quando tínhamos à disposição uma grande amostra. No caso em que a amostra era pequena, o AR foi superior.

Na Seção 3.6 verificamos as propriedades assintóticas do modelo AR-MV e pudemos perceber que o prévio conhecimento do parâmetro limiar (α) e o tamanho da amostra refletem diretamente na qualidade da estimação dos parâmetros autorregressivos (ϕ), sobretudo, nos parâmetros de maiores ordens.

Outra questão que buscamos nesta dissertação foi investigar as condições de aplicabilidade do AR-MV, visto que um ponto relevante não abordado por Fadel (2012) é a utilidade prática do modelo para descrever séries temporais reais. Reparamos que é possível sua comparação com os modelos AR, pelo motivo exposto no primeiro parágrafo e, por esse motivo, foram realizadas quatro aplicações do modelo a dados reais, no Capítulo 4. Nessas aplicações, almejamos principalmente mostrar as vantagens do modelo AR-MV frente a um

modelo consolidado nos meios acadêmico e profissional, há mais de quatro décadas. Após a análise dessas aplicações pudemos concluir que, em relação ao ajuste dos modelos, houve um emparelhamento pois, nas aplicações I e II, o AR foi melhor, mas nas aplicações III e IV, o AR-MV foi mais adequado. É relevante salientar que o AR-MV obteve um melhor ajuste nos casos onde os dados exigiam modelos de maior ordem autorregressiva, ou seja, quarta e quinta ordens. Com relação às previsões, os modelos também foram equivalentes. Nas aplicações I e IV os métodos referentes ao AR-MV mostraram uma capacidade preditiva mais adequada, enquanto que nas aplicações II e III, o AR foi superior.

Sendo assim, recomendamos um aprofundamento ainda maior no estudo do modelo AR-MV, especialmente, em relação aos aspectos empíricos perceptíveis, mas não provados teoricamente. Dessa forma, Assim, propomos testar o modelo AR-MV nos casos em que seja pertinente o ajuste do modelo AR aos dados sob análise, de forma a compará-los. É possível que o modelo AR-MV ajuste-se de maneira mais adequada, gere melhores estimativas e realize previsões mais oportunas que o modelo AR em questão.

Por fim, sugerimos que sejam pesquisados os seguintes tópicos, no caso de uma futura exploração do modelo AR-MV:

- (a) Estudar o caso onde a ordem autorregressiva de cada regime ou submodelo dependa somente dos valores passados dos próprios dados sob análise. Por exemplo, supondo que p representa a ordem de um modelo AR-MV hipotético, não necessariamente, o primeiro regime será regido por um AR(1), o segundo regime por um AR(2), o terceiro regime por um AR(3) e assim, sucessivamente. Ocorrerão situações em que a ordem autorregressiva de um regime inferior será maior que a ordem autorregressiva de um regime superior e vice-versa;
- (b) Averiguar a possibilidade de os erros, que compõem o modelo AR-MV, seguirem outra distribuição de probabilidade, que não seja a normal e, caso afirmativa, analisar o modelo detalhadamente;
- (c) Investigar a questão levantada neste trabalho de que o modelo AR-MV pode ser superior, em qualidade, em relação ao AR, quando pretendemos analisar dados apropriados à modelagem autorregressiva de grandes ordens;
- (d) Demonstrar teoricamente que o comportamento das FAC e FACP de um modelo AR(k) é idêntico ao comportamento apresentado por essas funções, para um modelo AR-MV(k), $\forall k, k > 1$, conforme indicam nossos estudos.

Referências Bibliográficas

- Akaike(1973)** H. Akaike. Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, 60, p. 255-265. Citado na pág. 18, 76
- Akaike(1974)** H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, p. 716-723. Citado na pág. 18, 76
- An e Huang(1996)** H. Z. An e F. C. Huang. The geometrical ergodicity of nonlinear autoregressive models. *Statistica Sinica* 6, p. 943-956. Citado na pág. 23, 25, 28
- Anderson e Darling(1952)** T. W. Anderson e D. A. Darling. Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23, p. 193-212. Citado na pág. 125
- Barbosa(2015)** E. K. P. Barbosa. Modelos autorregressivos de memória variável. *XXIII Congresso de Iniciação Científica da Unicamp*. Citado na pág. 3
- Box e Jenkins(1976)** G. E. P. Box e G. M. Jenkins. *Time series analysis: forecasting and control*. Holden-Day. Citado na pág. 1, 5, 8, 9, 10, 11, 12, 15
- Box e Pierce(1970)** G. E. P. Box e D. A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 64, p. 1509-1526. Citado na pág. 18
- Brockwell e Davis(1996)** P. J. Brockwell e R. A. Davis. *Introduction to Time Series and Forecasting*. Springer, 2 edição. Citado na pág. 9, 11, 15
- Bueno(2011)** R. L. S. Bueno. *Econometria de Séries Temporais*. Cengage Learning, 2 edição. Citado na pág. 6, 8, 16, 118
- CCVFloresta(2018)** CCVFloresta, 2018. URL <https://www.ccvfloresta.com/dendrocronologia>. Citado na pág. xviii, 75
- Chan(1991)** K. S. Chan. Percentage points of likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society B*, 53, 3, p. 691-696. Citado na pág. 122
- Chan e Tong(1985)** K. S. Chan e H. Tong. On the use of the deterministic lyapunov function for the ergodicity of stochastic difference equations. *Advances in Applied Probability*, 17, p. 666-678. Citado na pág. 24
- Cramér(1928)** H. Cramér. On the composition of elementary errors. *Scandinavian Actuarial Journal*, 1, p. 13-74. Citado na pág. 124
- Cryer e Chan(2008)** J.D. Cryer e K.S. Chan. *Time Series Analysis: With Applications in R*. New York: Springer Verlag. Citado na pág. 23, 121, 122

- Datamarket(2017)** Datamarket, 2017. URL <https://datamarket.com/data/list/?q=>. Citado na pág. 75
- Dickey e Fuller(1979)** D. A. Dickey e W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, Vol. 74, 366, p. 427-431. Citado na pág. 117, 118
- Dickey e Fuller(1981)** D. A. Dickey e W. A. Fuller. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, Vol. 49, 4, p. 1057-1073. Citado na pág. 117
- Fadel(2012)** D. F. Fadel. Modelos autorregressivos com memória variável, 2012. Citado na pág. v, vii, 1, 2, 3, 24, 27, 28, 29, 40, 41, 44, 73, 111
- Gross e Ligges(2015)** J. Gross e U. Ligges. *Package Nortest - tests for normality*, 2015. Citado na pág. 124, 125
- Hoel et al.(1972)** P. G. Hoel, S. C. Port e C. J. Stone. *Introduction to stochastic processes*. Boston: Houghton Mifflin Company. Citado na pág. 5
- Keenan(1985)** D. Keenan. A tukey nonlinear type test for time series nonlinearities. *Biometrika*, 72, p. 39-44. Citado na pág. 121
- Kolmogorov(1933)** A. N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Instituto Italiano degli Attuari*, 4, p. 83-91. Citado na pág. 123
- Kwiatkowski et al.(1992)** D. Kwiatkowski, P. C. B. Phillips, P. Schmidt e Y. Shin. Testing the null hypothesis of stationary against the alternative of a unit root. *Journal of Econometrics*, Vol. 54, p. 159-178. Citado na pág. 118
- Lilliefors(1967)** H. Lilliefors. On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62, p. 399-402. Citado na pág. 123
- Ljung e Box(1978)** G. M. Ljung e G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65, p. 297-303. Citado na pág. 18
- Morettin e Tolo(2006)** P. A. Morettin e C. M. C. Tolo. *Análise de Séries Temporais*. Blucher, 2 edição. Citado na pág. 5, 6, 7, 8, 12, 15, 16, 20, 117
- Parzen(1958)** E. Parzen. Conditions that a stochastic process be ergodic. *Annals of Mathematical Statistics*, 29, p. 299-301. Citado na pág. 7
- Phillips e Perron(1988)** P. C. B. Phillips e P. Perron. Testing for a unit root in times series regression. *Biometrika*, Vol. 75, 2, p. 335-346. Citado na pág. 118
- Pinto(2016)** D. R. Pinto. Processos de salto com memória de alcance variável, 2016. Citado na pág. 3
- Portal Action(2017)** Portal Action. Estatcamp - Consultoria Estatística e Qualidade e por DIGUP - Desenvolvimento de Sistemas e Consultoria Estatística, São Carlos, São Paulo, Brasil, 2017. URL <http://www.portalaction.com.br/>. Citado na pág. 123, 124

- R Core Team(2017)** R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL <http://www.R-project.org/>. Citado na pág. 3
- Schwarz(1978)** G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6, p. 461-464. Citado na pág. 18, 76
- Shapiro e Francia(1972)** S. S. Shapiro e R. S. Francia. An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67, p. 215-216. Citado na pág. 124
- Shapiro e Wilk(1965)** S. S. Shapiro e M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52 (3-4), p. 591-611. Citado na pág. 124
- Smirnov(1939)** N. V. Smirnov. Estimate of deviation between empirical distribution functions in two independent samples. *Bulletin Moscow University*, 2, p. 3-16. Citado na pág. 123
- Tjostheim(1990)** D. Tjostheim. Non-linear time series and markov chains. *Advances in Applied Probability*, 22, p. 587-611. Citado na pág. 24, 25
- Tong(1977)** H. Tong. Discussion of a paper by a. j. lawrance and n. t. kottegoda. *J. R. Statist. Soc. A*, 140, p. 34-35. Citado na pág. 1
- Tong(2007)** H. Tong. Birth of the threshold time series model. *Statistica Sinica* 17, p. 8-14. Citado na pág. 1
- Tong e Lim(1980)** H. Tong e K. S. Lim. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 42, No. 3., p. 245-292. Citado na pág. 22, 23
- Tsay(1986)** R. S. Tsay. Nonlinearity tests for time series. *Biometrika*, 73, p. 461-466. Citado na pág. 121
- Tsay(2005)** R.S. Tsay. *Analysis of Financial Time Series*. John Wiley Sons, Inc., Hoboken, New Jersey. Citado na pág. 19
- Tweedie(1975)** R. L. Tweedie. Sufficient conditions for ergodicity and recurrence of markov chains on a general state space. *Stochastic Processes and Their Applications*, 3, p. 383-403. Citado na pág. 23, 24
- von Mises(1928)** R. E. von Mises. *Wahrscheinlichkeit, Statistik und Wahrheit*. Julius Springer. Citado na pág. 124

Apêndice A – Testes para detecção de raízes unitárias

A.1 Teste de Dickey-Fuller aumentado

O teste proposto por Dickey e Fuller (1981) denominado Dickey-Fuller aumentado (do inglês, *augmented Dickey-Fuller test*), abreviadamente, ADF, é uma extensão do teste original também proposto por Dickey e Fuller (1979). No entanto, neste teste, supomos o erro um processo estacionário qualquer em vez de um ruído branco, como no primeiro artigo.

Guiando-nos por Morettin e Toloi (2006), suponha que a série possa ser representada por um AR(p)

$$Z_t = \theta_0 + \sum_{i=1}^p \phi_i Z_{t-i} + a_t, \quad (5.1)$$

que pode ser escrito como

$$\Delta Z_t = \theta_0 + \phi_1^* Z_{t-1} + \phi_2^* \Delta Z_{t-1} + \cdots + \phi_p^* \Delta Z_{t-p+1} + a_t, \quad (5.2)$$

onde θ_0 representa o intercepto e $\phi_1^* = \sum_{i=1}^p \phi_i - 1$, $\phi_j^* = -\sum_{i=j+1}^p \phi_i$, $j = 2, \dots, p$.

Caso o polinômio autorregressivo $\Phi(B)$ tenha uma raiz unitária, logo $\phi(1) = 0$ ou $1 - \phi_1 - \phi_2 - \cdots - \phi_p = 0$ ou então $\sum_{i=1}^p \phi_i = 1$ e, dessa forma, $\phi_1^* = 0$. Assim, testar a hipótese de que o polinômio autorregressivo possui raiz unitária é análogo a testar que $\phi_1^* = 0$.

Notamos que ϕ_1^* pode ser estimado como coeficiente do termo Z_{t-1} na regressão de mínimos quadrados de ΔZ_t sobre $1, Z_{t-1}, \Delta Z_{t-1}, \dots, \Delta Z_{t-p+1}$. Assim, temos a seguinte estatística

$$\frac{\hat{\phi}_1^*}{e.\hat{p}.(\hat{\phi}_1^*)}, \quad (5.3)$$

onde o termo $e.\hat{p}.(\hat{\phi}_1^*)$ representa o erro-padrão do coeficiente $\hat{\phi}_1^*$.

As hipóteses para o teste são as seguintes

$$\begin{cases} H_0 : \text{o processo possui raiz unitária (não é estacionário);} \\ H_1 : \text{o processo não possui raiz unitária (é estacionário).} \end{cases}$$

Os valores críticos foram tabelados por Dickey e Fuller (1981), por meio de simulações de Monte Carlo.

A.2 Teste de Phillips-Perron

O teste proposto por Phillips e Perron (1988) denominado Phillips-Perron, abreviadamente, PP, também é uma extensão do teste original proposto por Dickey e Fuller (1979). Entretanto, neste teste, é realizada uma correção não paramétrica, permitindo que o mesmo seja consistente ainda que não haja variáveis defasadas dependentes e correlação serial nos erros (Bueno, 2011).

Suponha a seguinte regressão

$$\Delta Z_t = \theta_0 + \theta_1 t + \delta Z_{t-1} + \sum_{i=1}^p \phi_i \Delta Z_{t-i} + a_t, \quad (5.4)$$

onde θ_0 representa o intercepto e θ_1 a tendência da regressão.

A estatística z é calculada por

$$z = n\hat{\delta}_n - \frac{n^2\hat{\sigma}^2}{2s_n^2} \left(\hat{\lambda}_n^2 - \hat{\gamma}_{0,n} \right), \quad (5.5)$$

onde

$$\begin{aligned} \hat{\gamma}_{j,n} &= \frac{1}{n} \sum_{i=1+j}^n r_i r_{i-j}, \\ \hat{\lambda}_n^2 &= \hat{\gamma}_{0,n} + 2 \sum_{j=1}^q \left(1 - \frac{j}{q+1} \right) \hat{\gamma}_{j,n}, \\ s_n^2 &= \frac{1}{n-k} \sum_{i=1}^n r_i^2, \end{aligned}$$

em que r_i representa o resíduo em z_i utilizando estimadores de mínimos quadrados, k é o número de covariáveis na regressão e, q é o número de defasagens utilizadas para calcular $\hat{\lambda}_n^2$.

Caso o processo seja não correlacionado teremos covariâncias nulas e neste caso, $\hat{\lambda}_n^2 = \hat{\gamma}_{0,n}$. Se o processo for homocedástico teremos que $d.p.(\delta) = 1/n$ e, então, z é dada por

$$z = n\hat{\delta} = \frac{\hat{\delta}}{se(\hat{\delta})}. \quad (5.6)$$

Cabe ressaltar que z é a estatística de Dickey-Fuller e, portanto, tem a mesma distribuição da estatística do teste ADF.

As hipóteses para o teste são as mesmas do teste anterior, ou seja

$$\begin{cases} H_0 : \text{o processo possui raiz unitária (não é estacionário);} \\ H_1 : \text{o processo não possui raiz unitária (é estacionário).} \end{cases}$$

A.3 Teste KPSS

O teste proposto por Kwiatkowski *et al.* (1992) homenageia os seus criadores Denis Kwiatkowski, Peter C. B. Phillips, Peter Schmidt e Yongcheol Shin.

Suponha como $Z_t, t = 1, 2, \dots, n$ as observações de uma série temporal qualquer a qual queremos testar sua estacionariedade. Suponha agora que é possível realizar a decomposição dessa série em componentes de tendência, passeio aleatório e erro, ou seja,

$$Z_t = \theta_1 t + r_t + a_t,$$

onde r_t é o passeio aleatório dado por

$$r_t = r_{t-1} + \mu_t,$$

com μ_t iid com média zero e variância σ_μ^2 .

Supondo agora, $e_t, t = 1, 2, \dots, n$ os resíduos de uma regressão em Z_t explicada pelas componentes de tendência, passeio aleatório e intercepto.

Seja $\hat{\sigma}_\varepsilon^2$ um estimador para a variância dos erros nesta regressão, definimos a soma parcial dos resíduos por

$$S_t = \sum_{i=1}^t e_i, \quad t = 1, 2, \dots, T.$$

A estatística do teste é dada por

$$\text{KPSS} = \sum_{t=1}^n \frac{S_t^2}{n^2 \hat{\sigma}_\varepsilon^2}. \quad (5.7)$$

Os autores demonstram que a estatística KPSS tem distribuição que converge assintoticamente para um Movimento Browniano e seus valores críticos são tabelados.

As hipóteses para este teste são as seguintes

$$\begin{cases} H_0 : \text{o processo não possui raiz unitária (é estacionário);} \\ H_1 : \text{o processo possui raiz unitária (não é estacionário).} \end{cases}$$

Devemos notar que a hipótese nula deste teste é igual a alternativa dos testes anteriores e vice-versa.

Apêndice B – Testes para detecção de não linearidade

B.1 Teste de Keenan

Seja Z_1, \dots, Z_n o conjunto de observações a ser analisado. O teste proposto por Keenan (1985) pode ser implementado da seguinte forma: (i) fazer a regressão de Z_t em Z_{t-1}, \dots, Z_{t-m} , incluindo o termo intercepto, onde m é algum inteiro positivo pré-especificado. Em seguida, calcular os valores ajustados $\{\hat{Z}_t\}$ e os resíduos $\{\hat{\varepsilon}_t\}$, para $t = m + 1, \dots, n$, e a soma dos quadrados dos resíduos (SQR); (ii) fazer a regressão de \hat{Z}_t^2 em Z_{t-1}, \dots, Z_{t-m} , incluindo o intercepto, e calcular os resíduos $\{\hat{\varepsilon}_t\}$ para $t = m + 1, \dots, n$; e (iii) fazer a regressão de $\hat{\varepsilon}_t$ nos resíduos $\hat{\varepsilon}_t$ sem o intercepto para $t = m + 1, \dots, n$, e a estatística teste de Keenan, dada em (5.8), é obtida ao multiplicar $(n - 2m - 2)/(n - m - 1)$ à estatística F para testar se a última função de regressão é identicamente zero. Assim,

$$\hat{F} = \frac{\eta^2 (n - 2m - 2)}{\text{SQR} - \eta^2}, \quad (5.8)$$

onde, $\eta = \eta_0 \sqrt{\sum_{t=m+1}^n \hat{\varepsilon}_t^2}$ e η_0 é o coeficiente de regressão.

Sob a hipótese nula de linearidade, a estatística de teste \hat{F} é distribuída aproximadamente como uma distribuição F com 1 e $n - 2m - 2$ graus de liberdade (Cryer e Chan, 2008).

B.2 Teste de Tsay

Tsay (1986) estendeu a abordagem de Keenan supondo alternativas não-lineares mais gerais. O teste de Keenan pode ser derivado heurísticamente como segue

$$Z_t = \phi_0 + \phi_1 Z_{t-1} + \dots + \phi_m Z_{t-m} + \exp \left\{ \eta \left(\sum_{j=1}^m \phi_j Z_{t-j} \right)^2 \right\} + a_t, \quad (5.9)$$

onde $\{a_t\}$ são iid com média zero e variância finita.

Uma alternativa mais geral à não linearidade pode ser formulada substituindo o termo

$$\exp \left\{ \eta \left(\sum_{j=1}^m \phi_j Z_{t-j} \right)^2 \right\} \quad (5.10)$$

por

$$\left. \begin{aligned} & \exp(\delta_{1,1}Z_{t-1}^2 + \delta_{1,2}Z_{t-1}Z_{t-2} + \dots + \delta_{1,m}Z_{t-1}Z_{t-m} \\ & + \delta_{2,2}Z_{t-2}^2 + \delta_{2,3}Z_{t-2}Z_{t-3} + \dots + \delta_{2,m}Z_{t-2}Z_{t-m} + \dots \\ & + \delta_{m-1,m-1}Z_{t-m+1}^2 + \delta_{m-1,m}Z_{t-m+1}Z_{t-m} + \delta_{m,m}Z_{t-m}^2) + a_t \end{aligned} \right\}. \quad (5.11)$$

Usando a aproximação $\exp(x) \approx 1+x$, vemos que o modelo não-linear é aproximadamente um modelo AR quadrático. Todavia, os coeficientes dos termos quadráticos são agora sem restrições. O teste de Tsay é equivalente a supor o seguinte modelo de regressão quadrática

$$\left. \begin{aligned} & Z_t = \phi_0 + \phi_1 Z_{t-1} + \dots + \phi_m Z_{t-m} \\ & + \delta_{1,1}Z_{t-1}^2 + \delta_{1,2}Z_{t-1}Z_{t-2} + \dots + \delta_{1,m}Z_{t-1}Z_{t-m} \\ & + \delta_{2,2}Z_{t-2}^2 + \delta_{2,3}Z_{t-2}Z_{t-3} + \dots + \delta_{2,m}Z_{t-2}Z_{t-m} + \dots \\ & + \delta_{m-1,m-1}Z_{t-m+1}^2 + \delta_{m-1,m}Z_{t-m+1}Z_{t-m} + \delta_{m,m}Z_{t-m}^2 + a_t \end{aligned} \right\}, \quad (5.12)$$

e testando se ou não todos os $m(m+1)/2$ coeficientes $\delta_{i,j}$ são zero. Novamente, isto pode ser realizado por um teste F que todos os $\delta_{i,j}$'s são zero na equação (Cryer e Chan, 2008).

B.3 Teste da não linearidade de limiar

Segundo Cryer e Chan (2008), pelo fato de os testes de Keenan e Tsay serem projetados para detecção de não linearidade quadrática, eles podem não ser sensíveis à não linearidade de limiar. Para esse fim específico, pode ser utilizado um teste de razão de verossimilhança. A hipótese nula para o teste é um modelo $AR(p)$ contra a hipótese alternativa de um modelo TAR de dois regimes de ordem p e com variância do erro constante, ou seja; $\sigma_1 = \sigma_2 = \sigma$. Com essas premissas, o modelo geral pode ser reescrito

$$Z_t = \phi_{1,0} + \phi_{1,1}Z_{t-1} + \dots + \phi_{1,p}Z_{t-p} + \{\phi_{2,0} + \phi_{2,1}Z_{t-1} + \dots + \phi_{2,p}Z_{t-p}\}I(Z_{t-d} > \alpha) + \sigma e_t, \quad (5.13)$$

onde a notação $I(\cdot)$ é uma variável indicadora que é igual a 1 se e somente se a expressão incluída for verdadeira. Além disso, nessa formulação, o coeficiente $\phi_{2,0}$ representa a mudança no intercepto no regime superior em relação ao regime inferior. Os coeficientes $\phi_{2,1}, \dots, \phi_{2,p}$ são interpretados de maneira semelhante. A hipótese nula afirma que $\phi_{2,0} = \phi_{2,1} = \dots = \phi_{2,p} = 0$. Embora a defasagem possa ser teoricamente maior do que a ordem autorregressiva, isso raramente acontece na prática. Assim, supondo que $d \leq p$ e sob a validade da linearidade, a distribuição amostral do teste não depende de d .

Na prática, o teste é realizado com p e d fixos. A estatística teste de razão de verossimilhança é dada então por

$$T_n = (n-p) \ln \left\{ \frac{\hat{\sigma}^2(H_0)}{\hat{\sigma}^2(H_1)} \right\}, \quad (5.14)$$

onde $n-p$ é o tamanho efetivo da amostra, $\hat{\sigma}^2(H_0)$ é o estimador de máxima verossimilhança da variância do erro, a partir do ajuste linear de $AR(p)$ e $\hat{\sigma}^2(H_1)$ do ajuste TAR com o limiar pesquisado em algum intervalo finito.

Sob a hipótese nula de que $\phi_{2,0} = \phi_{2,1} = \dots = \phi_{2,p} = 0$, o parâmetro α está ausente. Portanto, a distribuição amostral do teste de razão de verossimilhança em H_0 não é mais aproximadamente χ^2 com p graus de liberdade. Em vez disso, tem uma distribuição amostral não padronizada. Chan (1991) derivou um método de aproximação para calcular os p-valores do teste que são altamente precisos para pequenos p-valores. O teste depende de um intervalo de busca para o parâmetro limiar (α). A escolha desse intervalo deve assegurar a existência de dados adequados pertencentes a cada um dos dois regimes.

Apêndice C – Testes para detecção de normalidade

A seguir serão apresentados alguns testes que podem ser realizados para verificação da normalidade dos dados. Cabe ressaltar que as hipóteses para todos os testes a seguir são as seguintes

$$\begin{cases} H_0 : \text{os dados provêm de uma população normal;} \\ H_1 : \text{os dados não provêm de uma população normal.} \end{cases}$$

C.1 Teste de Kolmogorov-Smirnov

A ideia do teste proposto por Kolmogorov (1933) e Smirnov (1939) é verificar a máxima diferença absoluta entre a função de distribuição acumulada (fda) suposta para os dados e a sua função de distribuição empírica.

Seja uma amostra aleatória simples x_1, x_2, \dots, x_n de uma população com fda contínua F_x desconhecida. A estatística teste é a seguinte

$$D_n = \sup_x |F(x) - F_n(x)|, \quad (5.15)$$

onde $F(x)$ representa a fda suposta para os dados e $F_n(x)$ representa a fda empírica dos dados.

Sob H_0 , a distribuição assintótica da estatística de Kolmogorov-Smirnov é dada por

$$\lim_{n \rightarrow \infty} P[\sqrt{n}D_n \leq x] = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j-1} \exp^{-2j^2 x^2}.$$

Entretanto, essa distribuição somente é válida quando a distribuição de H_0 é conhecida, o que, na prática, dificilmente ocorrerá.

Assim, comparamos o valor da estatística teste calculada, por meio da equação (5.15), com o valor crítico tabelado, para um determinado nível de significância (Portal Action, 2017).

C.2 Teste de Lilliefors

O teste proposto por Lilliefors (1967) é derivado do teste de Kolmogorov-Smirnov. Assim, seja uma amostra aleatória simples x_1, x_2, \dots, x_n de uma população com fda contínua F_x desconhecida. A estatística teste é a seguinte

$$D^+ = \max_{i=1, \dots, n} \{i/n - p_{(i)}\}, D^- = \max_{i=1, \dots, n} \{p_{(i)} - (i-1)/n\}, \quad (5.16)$$

onde $p_{(i)} = \Phi([x_i - \bar{x}]/s)$, Φ é a fda da distribuição normal padrão e \bar{x} e s são, respectivamente, a média e o desvio-padrão dos dados.

Assim, comparamos o valor da estatística teste calculada, por meio da equação (5.16), com o valor crítico tabelado, para um determinado nível de significância (Gross e Ligges, 2015).

C.3 Teste de Cramer-von Mises

Seja uma amostra aleatória simples x_1, x_2, \dots, x_n de uma população com fda contínua F_x desconhecida. A estatística para o teste, proposto por Cramér (1928) e von Mises (1928), é a seguinte

$$W = \frac{1}{12n} + \sum_{i=1}^n \left(p_{(i)} - \frac{2i-1}{2n} \right)^2, \quad (5.17)$$

onde $p_{(i)} = \Phi([x_i - \bar{x}]/s)$, Φ é a fda da distribuição normal padrão e \bar{x} e s são, respectivamente, a média e o desvio-padrão dos dados.

Assim, comparamos o valor da estatística teste calculada, por meio da equação (5.17), com o valor crítico tabelado, para um determinado nível de significância (Gross e Ligges, 2015).

C.4 Teste de Shapiro-Wilk

Suponha uma amostra aleatória x_1, x_2, \dots, x_n . O teste, proposto por Shapiro e Wilk (1965), é baseado na estatística W dada por

$$W = \frac{b^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2}, \quad (5.18)$$

em que $x_{(i)}$ são os valores da amostra ordenados, onde $x_{(1)}$ é o menor e $x_{(n)}$ é o maior, e a constante b é determinada da seguinte forma

$$b = \begin{cases} \sum_{i=1}^{n/2} a_{n-i+1} (x_{(n-i+1)} - x_{(i)}), & \text{se } n \text{ é par,} \\ \sum_{i=1}^{(n+1)/2} a_{n-i+1} (x_{(n-i+1)} - x_{(i)}), & \text{se } n \text{ é ímpar,} \end{cases}$$

em que a_{n-i+1} são constantes geradas pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho n de uma distribuição normal. Assim, comparamos o valor da estatística teste calculada, por meio da equação (5.18), com o valor crítico tabelado, para um determinado nível de significância (Portal Action, 2017).

C.5 Teste de Shapiro-Francia

O teste de Shapiro-Francia, proposto por Shapiro e Francia (1972), é uma simplificação do teste de Shapiro-Wilk, e é baseado na correlação entre os valores da amostra ordenada e os quantis ordenados esperados da distribuição normal padrão (Gross e Ligges, 2015).

Suponha uma amostra aleatória x_1, x_2, \dots, x_n . Seja $x_{(i)}$ o i -ésimo valor ordenado de uma amostra de tamanho n . Seja $m_{i:n}$ a média da i -ésima estatística de ordem ao tomar n amostras independentes de uma distribuição normal.

A estatística teste W' é dada pelo coeficiente de correlação de Pearson entre x e m , ou seja

$$W' = \frac{\text{cov}(x, m)}{\sigma_x \sigma_m}. \quad (5.19)$$

Sob a hipótese nula de que os dados são normais, essa correlação será forte, então os valores de W' ficarão próximos de -1 (correlação negativa) ou 1 (correlação positiva).

C.6 Teste de Anderson-Darling

Seja uma amostra aleatória simples x_1, x_2, \dots, x_n de uma população com fda contínua F_x desconhecida.

A estatística para o teste proposto por Anderson e Darling (1952) é a seguinte

$$A = -n - \frac{1}{n} \sum_{i=1}^n [2i - 1] [\ln(p_{(i)}) + \ln(1 - p_{(n-i+1)})], \quad (5.20)$$

onde $p_{(i)} = \Phi([x_i - \bar{x}]/s)$, Φ é a fda da distribuição normal padrão e \bar{x} e s são, respectivamente, a média e o desvio-padrão dos dados.

Assim, comparamos o valor da estatística teste calculada, por meio da equação (5.20), com o valor crítico tabelado, para um determinado nível de significância (Gross e Ligges, 2015).