CEFET/RJ
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO

MODELO AUTORREGRESSIVO MÉDIA-MÓVEL BASEADO EM JANELAS DESLIZANTES

Arthur Ronald Ferreira Diogenes Garcia

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Ciência da Computação, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca CEFET/RJ, como parte dos requisitos necessários à obtenção do grau de mestre.

Orientadores:
Eduardo Soares Ogasawara
Dayse Haime Pastore

Rio de Janeiro,

18 de Março de 2025

Modelo Autorregressivo Média-Móvel baseado em Janelas deslizantes

Dissertação de Mestrado em Ciência da Computação, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ.

Arthur Ronald Ferreira Diogenes Garcia

Aprovada por:

Presidente, Professor Eduardo Soares Ogasawara, D.Sc. (orientador)

Professora Dayse Haime Pastore, D.Sc. (coorientador)

Professor Jorge de Abreu Soares, D.Sc.

Professor Fábio André Machado Porto, D.Sc.

Rio de Janeiro,

18 de Março de 2025

Ficha catalográfica elaborada pela Biblioteca Central do CEFET/RJ

G216 Garcia, Arthur Ronald Ferreira Diogenes
Modelo autorregressivo média-móvel baseado em janelas deslizantes / Arthur Ronald Ferreira Diogenes Garcia. — 2025.
61f. : il. color. , enc.

Dissertação (Mestrado) Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, 2025.
Bibliografia : f. 56-61
Orientador: Eduardo Soares Ogasawara
Coorientadora: Dayse Haime Pastore

1. Análise de séries temporais. 2. Autorregressivo média-móvel. 3. Rede de computador - Protocolos. 4. Teoria da previsão. I. Ogasawara, Eduardo Soares. (Orient.). II. Pastore, Dayse Haime. (Coorient.). III. Título.

CDD 519.55

Elaborada pela bibliotecária Tania Mello – CRB/7 nº 5507/04

DEDICATÓRIA

A Deus, pois sem ele nada posso. A minha família, meus pais, meus irmãos e sobrinhos, a quem os considero como filhos. Aos meus avós paternos e maternos (*in memoriam*) pelo amor e carinho incondicional. Ao meu amor, Clarice Joana, com quem espero compartilhar nossas vidas. A Ana Clara e Ana Emília Veríssimo, pelos momentos únicos e especiais.

AGRADECIMENTOS

Agradeço ao CEFET/RJ, em especial ao corpo docente e discente. Ao meu orientador, Eduardo Soares Ogasawara, pelo apoio, compreensão e fé em um projeto que começou com um rascunho. À minha coorientadora, Dayse Haime Pastore, pelo apoio e parceria durante essa jornada do mestrado. Aos professores Kele Teixeira Belloze, Rafaelli de Carvalho Coutinho e Eduardo Bezerra da Silva. Aos colegas Rebecca Pontes Salles, Erica Carneira Queiroz da Silva e Diego Silva de Salles. Aos mentores Ian Monteiro Nunes, Francisco Pedrosa e Willian Santos de Albuquerque.

Por fim, agradeço ao Instituto Brasileiro de Geografia e Estatística (IBGE), que direta ou indiretamente contribuiu para que eu completasse essa jornada.

RESUMO

Modelo Autorregressivo Média-Móvel baseado em Janelas deslizantes

Arthur Ronald Ferreira Diogenes Garcia

Orientadores:
Eduardo Soares Ogasawara
Dayse Haime Pastore

Resumo da Dissertação submetida ao Programa de Pós-graduação em Ciência da Computação do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca CEFET/RJ como parte dos requisitos necessários à obtenção do grau de mestre.

Modelos de séries temporais vêm sendo propostos há décadas, sendo que a maioria desses modelos requer como pré-requisito a condição estacionária, i.e., que a série possua média, variância e covariância constantes. Diante disso, é necessário pré-processá-la a fim de obter uma série estacionária. Não obstante, é desejável para fins de análise que o modelo bem como as propriedades estatísticas da série possam ser interpretáveis, a fim de auxiliar na análise do usuário, facilitando sua aplicação. Esta dissertação propõe um modelo denominado WARMA, Autorregressivo Média-Móvel baseado em Janelas deslizantes, que combina uma fase de pré-processamento, visando obter a estacionariedade preservando as propriedades estatísticas da série original, e uma fase de modelagem por meio do ARMA (Autorregressivo Média-Móvel). Uma avaliação experimental utilizando séries econômico-financeiras mostra que o WARMA alcança desempenho comparável ao ARIMA (Autorregressivo Integrado Média-Móvel) em diferentes horizontes de previsão, com desempenhos complementares em domínios específicos. Isto é, enquanto um modelo se destaca em previsões de curto prazo, o outro demonstra melhor desempenho em predições de longo prazo. Adicionalmente, os resultados indicam que o WARMA fornece uma alternativa eficaz para análise de séries mantendo a interpretabilidade das propriedades da série original.

Palavras-chave: Séries Temporais, Estacionariedade, Autorregressivo Média-Móvel, Janelas Deslizantes, Normalização Adaptativa

Rio de Janeiro,

18 de Março de 2025

ABSTRACT

Modelo Autorregressivo Média-Móvel baseado em Janelas deslizantes

Arthur Ronald Ferreira Diogenes Garcia

Advisors:
Eduardo Soares Ogasawara
Dayse Haime Pastore

Abstract of dissertation submitted to Programa de Pós-graduação em Ciência da Computação - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca CEFET/RJ as partial fulfillment of the requirements for the degree of master.

Time series models have been proposed for decades, and most require a stationary condition as a prerequisite, i.e., the series has a constant mean, variance, and covariance. Given this requirement, preprocessing methods are necessary to obtain a stationary series. However, it is desirable for analysis purposes that the model and the statistical properties of the series remain interpretable to support user analysis and facilitate its application. This work proposes a model called WARMA, a Sliding Window-based Autoregressive Moving Average, which combines a preprocessing phase aimed at obtaining stationarity while preserving the statistical properties of the original series and a modeling phase using ARMA (Autoregressive Moving Average). An experimental evaluation using economic and financial series shows that WARMA achieves comparable performance to ARIMA (Autoregressive Integrated Moving-Average) over different forecast horizons, with complementary performance in specific domains. In other words, while one model excels in short-term forecasts, the other performs better in long-term predictions. In addition, the results indicate that WARMA provides an effective alternative for analyzing time series while maintaining the interpretability of the properties of the original series.

Key-words: Time Series, Stationarity, Autoregressive Moving-Average, Sliding Windows, Adaptive Normalization

Rio de Janeiro,

18 de Março de 2025

Sumário

I Introdução **15**

II Referencial teórico **19**

II.1 Funções ACF e PACF 22

II.2 Componentes de uma série 23

II.3 Modelos lineares 24

II.3.1 Regressão linear 24

II.3.2 Modelo Autorregressivo 25

II.3.3 Modelo Média Móvel 26

II.3.4 Autorregressivo Média-Móvel 28

II.4 Tratamento da não estacionariedade 29

II.4.1 Box-Cox 30

II.4.2 Integração 30

II.4.3 Integração fracionária 31

II.4.4 *Detrending* 33

II.4.5 Transformação por variação percentual 34

II.4.6 Transformada de Fourier 34

II.4.7 Transformada wavelets 35

II.4.8 Integração sazonal 36

III Trabalhos relacionados **38**

III.1 Autorregressivo Integrado Média-Móvel 38

III.2 Autorregressivo Fracionário Integrado Média-Móvel 38

III.3 Sazonal Autorregressivo (Fracionário) Integrado Média-Móvel 39

III.4 Autorregressivo Integrado Média-Móvel com variáveis exógenas 39

IV Metodologia **40**

IV.1 Pré-processamento 40

IV.1.1 Janelas deslizantes 40

IV.1.2 Normalização da média 42

IV.1.3 Processamento de *outliers* 42
IV.1.4 Normalização da variância 43
IV.1.5 Vetorizando a série pré-processada 44
IV.2 Modelo WARMA 45
IV.2.1 Notação WARMA 47

V Avaliação experimental 48
V.1 Metodologia 48
V.1.1 Identificação 48
V.1.2 Estimação 49
V.1.3 Diagnóstico 50
V.2 Conjunto de dados 51
V.3 Particionamento dos dados 51
V.4 Avaliação experimental I 52
V.4.1 Diagrama de cordas 52
V.4.2 Análise dos resultados 53
V.4.3 Análise de hiperparâmetros 54
V.5 Avaliação experimental II 55
V.5.1 Métricas 55
V.5.2 Análise por domínio 58
V.5.3 Análise global 65

VI Conclusões 66
VI.1 Considerações finais 66
VI.2 Produção bibliográfica 67
VI.3 Trabalhos futuros 67

Referências 70

Lista de Figuras

I.1 Série quantidade de passageiros em função do ano em (a), na qual foi aplicado o método Box-Cox, resultando na série em (b). 16

I.2 Série original não estacionária em (a), em que a linha tracejada delimita uma quebra estrutural (mudança de nível), comum em séries de natureza econômica, e a respectiva série estacionária por meio do WARMA em (b). 17

II.1 IPCA, índice oficial da inflação do Brasil, no qual adquiriu comportamento estacionário e estável alguns anos após a implantação do Plano real 20

II.2 Amostras de série estacionária, representada pelo ruído branco (a), e não estacionárias: passeio randômico (b), passeio randômico com *drift* (c) e tendência estacionário (d) 22

II.3 ACF de um ruído branco, caracterizado pela ausência de correlação serial 23

II.4 Decomposição da concentração atmosférica de CO_2 para o período de 1959 a 1997 24

II.5 ACF para diferentes valores de Integração fracionária 33

II.6 Tendência estacionária em (a) e a série resultante decorrente da remoção da tendência por meio de WT em (b) 36

IV.1 Visão geral do WARMA 40

IV.2 Série sintética e as respectivas subsequências obtidas a partir do procedimento de SW para w igual a 20 e τ igual a 1 em (a) e a mesma série as respectivas subsequências para w igual a 20 e τ igual a 20 em (b) 41

IV.3 Influência de *outlier* após a fase Normalização da média 43

IV.4 Fluxo do pré-processamento do WARMA 44

V.1 Tabela com valores randomicamente gerados em (a) e a respectiva representação gráfica por meio do diagrama de cordas em (b) 52

V.2 Diagrama de cordas relacionando o percentual de séries conforme a ordem de Integração e a quantidade de passos do WARMA 53

V.3 Quantidade de séries por domínio segundo a quantidade de passos 54

V.4 Quantidade de séries por domínio segundo a ordem de integração 54

V.5 Modelos WARMA selecionados segundo o tamanho das subsequências (w) em função do deslocamento entre elas (τ) 55
V.6 Diagrama de caixa aplicada ao domínio Mercado de ações para diferentes horizontes dos modelos ARIMA e WARMA 58
V.7 Diagrama de caixa aplicada ao domínio Câmbio monetário para diferentes horizontes dos modelos ARIMA e WARMA 59
V.8 Diagrama de caixa aplicada ao domínio Commodities para diferentes horizontes dos modelos ARIMA e WARMA 60
V.9 Diagrama de caixa aplicada ao domínio Atividade econômica para diferentes horizontes dos modelos ARIMA e WARMA 61
V.10 Diagrama de caixa aplicada ao domínio DLSP para diferentes horizontes dos modelos ARIMA e WARMA 62
V.11 Diagrama de caixa aplicada ao domínio IPCA para diferentes horizontes dos modelos ARIMA e WARMA 63
V.12 Diagrama de caixa aplicada ao domínio Taxa de juros para diferentes horizontes dos modelos ARIMA e WARMA 64
V.13 Diagrama de caixa aplicada diferentes horizontes de séries econômico-financeiras dos modelos ARIMA e WARMA 65

VI.1 Covariância positiva em (a), igual a zero em (b) e negativa em (c) 68
VI.2 *chirp* linear em (a) e a respectiva série estacionária por meio do processo *chirp* linear generalizado em (b) 69

Lista de Tabelas

II.1 Modelo ARMA e os respectivos modelos equivalentes AR e MA 28

V.1 Conjunto de séries econômico-financeiras 51

V.2 Sumário estatístico da métrica MASE aplicada ao domínio Mercado de ações para diferentes horizontes dos modelos ARIMA e WARMA 58

V.3 Sumário estatístico da métrica MASE aplicada ao domínio Câmbio monetário para diferentes horizontes dos modelos ARIMA e WARMA 59

V.4 Sumário estatístico da métrica MASE aplicada ao domínio Commodities para diferentes horizontes dos modelos ARIMA e WARMA 60

V.5 Sumário estatístico da métrica MASE aplicada ao domínio Atividade econômica para diferentes horizontes dos modelos ARIMA e WARMA 61

V.6 Sumário estatístico da métrica MASE aplicada ao domínio DLSP para diferentes horizontes dos modelos ARIMA e WARMA 62

V.7 Sumário estatístico da métrica MASE aplicada ao domínio IPCA para diferentes horizontes dos modelos ARIMA e WARMA 63

V.8 Sumário estatístico da métrica MASE aplicada ao domínio Taxa de juros para diferentes horizontes dos modelos ARIMA e WARMA 64

V.9 Sumário estatístico da métrica MASE para diferentes horizontes de séries econômico-financeiras dos modelos ARIMA e WARMA 65

Lista de Acrônimos

ACF	Função de Autocorrelação.
AIC	Critério da Informação de Akaike.
AICc	Critério da Informação de Akaike corrigido.
AN	Normalização Adaptativa.
AR	Autorregressivo.
ARFIMA	Autorregressivo Fracionário Integrado Média-Móvel.
ARIMA	Autorregressivo Integrado Média-Móvel.
ARIMAX	Autorregressivo Integrado Média-Móvel com variáveis eXógenas.
ARMA	Autorregressivo Média-Móvel.
BCB	Banco Central do Brasil.
BIC	Critério da Informação Baiesiano.
DFT	Transformada de Fourier discreta.
DLSP	Dívida Líquida do Setor Público.
FFT	Transformada rápida de Fourier.
FT	Transformada de Fourier.
IBGE	Instituto Brasileiro de Geografia e Estatística.
IPCA	Índice Nacional de Preços ao Consumidor Amplo.
IQR	intervalo interquartil.
LSE	Método dos Mínimos Quadrados.
LT	Transformada logarítmica.
MA	Média-Móvel.
MAE	Erro Absoluto Médio.
MAPE	Erro Percentual Absoluto Médio.
MASE	Erro Escalado Absoluto Médio.

MLE	Máxima Verossimilhança.
MoM	Método dos Momentos.
PACF	Função de Autocorrelação Parcial.
PCT	Transformação por variação percentual.
SARIMA	Sazonal Autorregressivo Integrado Média-Móvel.
sMAPE	Erro Percentual Absoluto Médio simétrico.
STFT	Transformada de Fourier Intervalar.
SW	Janelas deslizantes.
WARMA	Autorregressivo Média-Móvel baseado em Janelas deslizantes.
WT	Transformada wavelets.

Capítulo I Introdução

Série temporal (ou série, para fins de brevidade) é uma sequência cronologicamente ordenada de observações obtidas em intervalos igualmente espaçados, podendo ser classificada como estacionária ou não estacionária, sendo a primeira caracterizada por possuir média, variância e covariância constantes. Caso pelo menos uma dessas premissas seja inválida, a série é não estacionária. Do ponto de vista analítico, séries estacionárias permitem generalizar a análise de um período a outro da série, diferentemente do que ocorre com séries não estacionárias, nas quais somente o período em questão pode ser analisado [Gujarati and Porter, 2008]. Tal comportamento estacionário facilita a adoção de modelos sujeitos às restrições legais ou de negócio.

No entanto, para a maioria das séries, a não estacionariedade é a regra enquanto a estacionariedade é a exceção. Paralelamente, diversos modelos de séries requerem que a série a ser modelada seja estacionária. Como consequência, Box et al. [2015] popularizaram a Integração como método de pré-processamento de séries não estacionárias em estacionárias.

Posteriormente à Integração, diversos outros métodos surgiram ou foram adaptados para lidar com a não estacionariedade, sendo que, de modo geral, esses métodos focam em uma ou mais dessas propriedades estatísticas, i.e., média, variância e covariância [Salles et al., 2019; Yang and Zurbenko, 2010], de forma que a condição estacionária é alcançada pela aplicação de um ou mais métodos de pré-processamento.

Por outro lado, outros métodos agem coletivamente sobre a média, variância e covariância, como a própria Integração [Box et al., 2015], mesmo que a série requeira pré-processar somente uma ou duas dessas propriedades para obter a condição estacionária. Como resultado, o comportamento da série original é modificado, de forma que a interpretabilidade da série integrada fica comprometida, dificultando uma análise mais abrangente [Hyndman and Athanasopoulos, 2018].

Usando como exemplo a série passageiros aéreos (Figura I.1 (a)) [Box et al., 2015] e aplicando à mesma o método Box-Cox [Box and Cox, 1964], obtém-se uma série na qual somente a variância permaneceu constante (Figura I.1 (b)), devendo, portanto, ser aplicado em conjunto com outro método a fim de obter média constante e, consequentemente, uma série estacionária.

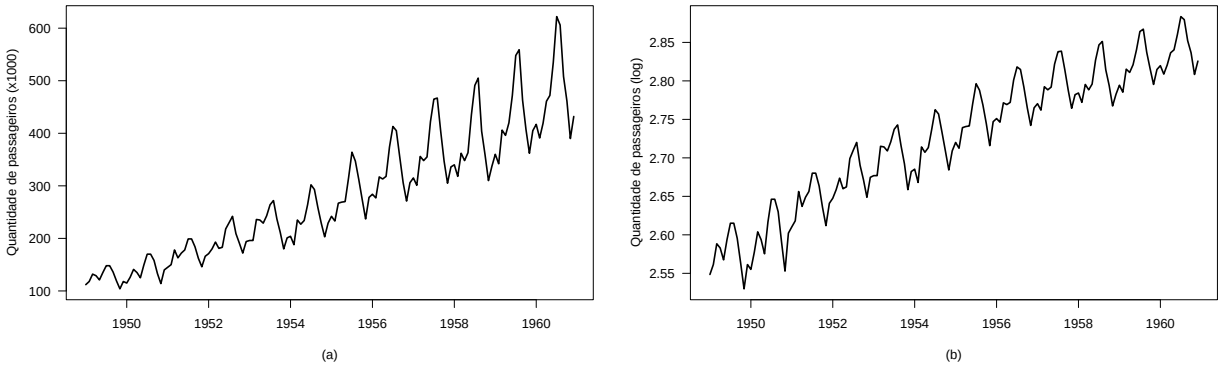


Figura I.1: Série quantidade de passageiros em função do ano em (a), na qual foi aplicado o método Box-Cox, resultando na série em (b).

Outro ponto é que, independentemente da série ser estacionária ou não, a eficácia de determinados modelos pode ser beneficiada conforme as características da série em questão, de forma que fica difícil distinguir se a eficácia do modelo foi em função unicamente dele ou da combinação entre o modelo e as características da série. Assim, pré-processar a série, visando obter uma série estacionária, permite fornecer uma base estável para que diferentes modelos possam ser comparados, por exemplo, para fins de predição [Pokorny, 1987; Salles et al., 2019].

Visando preencher essa lacuna, de forma preservar ao máximo o comportamento original da série, apresentamos um modelo denominado WARMA, que combina pré-processamento e modelagem, no qual o primeiro visa obter uma série estacionária através do processamento sequencial da média e variância, respeitando essa ordem.

Para isso, o WARMA introduz o conceito de passos, em que cada passo consiste na aplicação de um ou mais métodos de pré-processamento, visando obter o comportamento constante para uma, e somente uma, das propriedades estatísticas da série em questão. Assim, enquanto o passo um foca na média, o passo dois foca na variância. Seguindo essa lógica, uma série pode requerer de zero a dois passos para obter sua contrapartida estacionária, em que zero passo indica uma série estacionária e, portanto, não requer pré-processamento.

Para fins de exemplificação, uma quebra estrutural é simulada na Figura I.2 (a), caracterizada por uma mudança brusca no comportamento da série, usualmente devido à mudança de gestão, decisões políticas e choques externos, entre outros [Shrestha and Bhatta, 2018]. No instante de 0 a 100, a série apresenta um comportamento estacionário. Porém, devido à quebra estrutural no instante 100, ocorre uma mudança na média da série, variando de zero para cinco, fazendo com que a série seja não estacionária (Figura I.2 (a)).

Ao aplicar o WARMA, somente a média da série é pré-processada, de modo que a série resultante seja estacionária, além de preservar o comportamento de ambas subsequências à esquerda e à direita

do instante 100 (Figura I.2 (b)), quando comparadas com as respectivas subsequências da Figura I.2 (a). Como a série necessitou pré-processar somente a média, a quantidade de passos para obter a série estacionária é um.

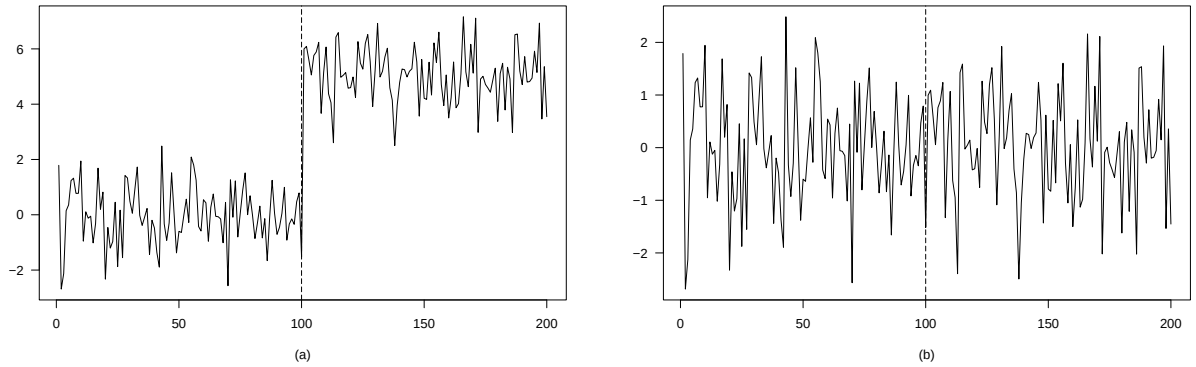


Figura I.2: Série original não estacionária em (a), em que a linha tracejada delimita uma quebra estrutural (mudança de nível), comum em séries de natureza econômica, e a respectiva série estacionária por meio do WARMA em (b).

Para obter séries estacionárias, o WARMA usa SW que, de forma geral, consiste em obter subsequências de tamanho fixo a partir da série original, às quais se aplicam uma ou mais funções matemáticas, visando alguma finalidade. No caso do WARMA, objetiva-se capturar o comportamento inercial das subsequências que compõem a série original [Ogasawara et al., 2010]. Isto é, subsequências nas quais as propriedades estatísticas permaneçam constantes, de forma obter uma série estacionária.

Uma vez de posse da série pré-processada e estacionária, passa-se para a fase de modelagem, na qual a série resultante do pré-processamento é modelada por meio do ARMA [Box et al., 2015]. Assim, a interpretabilidade do WARMA é alcançada através da preservação do comportamento das subsequências que compõem a série original, bem como do ARMA, por considerar no modelo resultante as observações e erros precedentes.

Além desta introdução, esta dissertação inclui mais seis capítulos. No capítulo II, revisamos os conceitos relacionados às séries estacionárias e não estacionárias, bem como os métodos de pré-processamento de séries estacionárias em não estacionárias. No capítulo III, modelos que, assim como o WARMA, combinam pré-processamento e modelagem são discorridos. No capítulo IV, os fundamentos que suportam o modelo WARMA são descritos, descrevendo inicialmente a fase de pré-processamento, responsável por obter a série estacionária, e posteriormente a incorporação dessa fase ao modelo ARMA, cujo resultado é o modelo WARMA.

Já a avaliação experimental é discutida no capítulo V, na qual séries econômico-financeiras de diferentes domínios, como commodities, juros, câmbio monetário, entre outros, são avaliadas em duas partes. Na primeira parte, relacionamos a quantidade de passos necessários para obter

a estacionariedade no WARMA com a ordem do modelo ARIMA [Box et al., 2015], que usa a Integração como método de pré-processamento para obter a série estacionária.

Nessa primeira parte, destaca-se que, para a maioria das séries econômico-financeiras avaliadas, a condição estacionária é satisfeita por meio do WARMA pré-processando somente a média e variância. Na segunda parte, a eficácia do modelo WARMA é avaliada para essas séries segundo o critério da previsão no curto e médio prazo, por meio de um benchmarking com o modelo ARIMA.

Por fim, as conclusões no capítulo VI, no qual as contribuições desta dissertação são resumidas. Além disso, avaliam-se a produção bibliográfica e trabalhos futuros. Neste contexto, a covariância é analisada, pois, conforme visto anteriormente, uma série estacionária requer média, variância e covariância constantes. Entretanto, o WARMA foca somente na média e variância. Diante disso, métodos que podem ser aplicados às séries de forma obter uma série com covariância constante são examinados, visando uma evolução do WARMA. Complementando a covariância, discute-se brevemente a sazonalidade, além de outras oportunidades de trabalhos futuros.

Capítulo II Referencial teórico

Uma série X é uma sequência de n observações aleatórias, $\langle \mathrm{x}_1, \ldots, \mathrm{x}_n \rangle$, em que x_1 corresponde ao valor assumido pela observação no primeiro instante e x_n ao valor assumido pela observação no instante mais recente. Já o tamanho da série é designado por |X| enquanto uma observação específica da série é referenciada por x_t, indexada no tempo $t = \langle 1, \ldots, n \rangle$ [Shumway and Stoffer, 2017; Salles et al., 2019].

Uma série é classificada como estritamente estacionária se o comportamento probabilístico de quaisquer sequência de valores $\langle \mathrm{x}_{t_1}, \ldots, \mathrm{x}_{t_k} \rangle$ for igual ao da respectiva sequência deslocada $\langle \mathrm{x}_{t_{1+\tau}}, \ldots, \mathrm{x}_{t_{k+\tau}} \rangle$, para qualquer $k = 1, 2, \ldots$ e para quaisquer deslocamento $\tau = 0, \pm 1, \pm 2, \ldots$ [Shumway and Stoffer, 2017]. Dado que, para a maioria das aplicações, as respectivas distribuições são desconhecidas a priori [Woodward et al., 2017], admite-se a definição covariância-estacionária, na qual a série apresenta média, variância e covariância constantes [Tsay, 2010; Enders, 2014].

Pela definição covariância-estacionária, as equações II.1a, II.1b e II.1c independem do instante t, em que E corresponde ao valor esperado, μ e σ^2_{x}, constantes, correspondem à média e variância da série, respectivamente, enquanto γ_l, constante, corresponde à autocovariância da série defasada por um inteiro l. Adicionalmente, para fins de simplificação, adotar-se-á ao longo desta dissertação a denominação estacionária para se referir à definição covariância-estacionária.

$$E[\mathrm{x}_t] = \mu \tag{II.1a}$$

$$E[(\mathrm{x}_t - \mu)^2] = \sigma^2_{\mathrm{x}} \tag{II.1b}$$

$$E[(\mathrm{x}_t - \mu)(\mathrm{x}_{t-l} - \mu)] = \gamma_l \tag{II.1c}$$

Embora não diretamente relacionada com a definição covariância-estacionária, mas de relativa importância, Pickup [2014] cita a propriedade estabilidade, através da qual a série é considerada estável se possuir um equilíbrio, para o qual converge no longo prazo. O conceito de equilíbrio pode ser visto como o valor médio esperado no longo prazo.

Dependendo dos requisitos do usuário, essa propriedade pode ser útil ao poder considerar a série integral, sem modificações, somente a porção estacionária ou ainda pré-processar a série inteira, visando obter uma série estacionária. Essa propriedade fica evidente na série do IPCA, aproxima-

damente a partir de 2004, quando o comportamento estacionário de longo prazo foi estabelecido (Figura II.1).

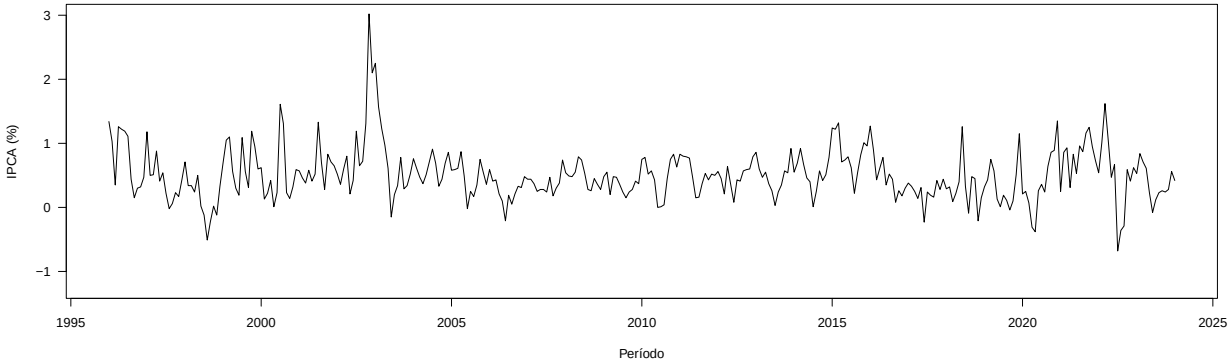


Figura II.1: IPCA, índice oficial da inflação do Brasil, no qual adquiriu comportamento estacionário e estável alguns anos após a implantação do Plano real

No campo das séries estacionárias, uma em especial se destaca, denominada ruído branco, caracterizada por ser independente e identicamente distribuída (iid) com média μ e variância σ^2. Caso essa distribuição seja normal e a média zero, o ruído branco é denominado gaussiano ou normal. Ciente disso, salvo disposição em contrário, considerar-se-á o gaussiano qualquer referência ao ruído branco.

Contrastando com o ruído branco, principal referência quando se faz menção às séries estacionárias, a literatura em geral elicita as séries passeio randômico e tendência estacionário, incluindo variações destas que, individualmente ou combinadas entre si, determinam os blocos constituintes de séries não estacionárias mais complexas.

O passeio randômico modela o comportamento de séries de diversas áreas de aplicação, como séries financeiras, por exemplo, sendo o mercado de ações, entre outros, um caso clássico [Tsay, 2010], além de ser um padrão contra o qual a eficiência de modelos mais complexos são avaliados [Metcalfe and Cowpertwait, 2009; Enders, 2014].

No passeio randômico, o valor corrente da série é igual ao valor anterior mais um valor obtido a partir de um ruído branco (Equação II.2), em que ε_t corresponde ao valor corrente do ruído branco, podendo ainda ser representado pelo somatório de ruídos brancos, conforme Equação II.3, em que x_0 corresponde ao valor inicial da série.

O passeio randômico é classificado como um processo de raiz unitária, pois a Equação (II.2) pode ser reescrita como $x_t = \rho x_{t-1} + \varepsilon_t$, sendo ρ igual a 1. Por essa razão, passeio randômico e raiz unitária podem ser considerados intercambiáveis [Gujarati and Porter, 2008]. Caso o valor absoluto de ρ fosse menor que 1, a série seria estacionária.

$$\mathrm{x}_t = \mathrm{x}_{t-1} + \varepsilon_t \tag{II.2}$$

$$= \mathrm{x}_0 + \sum_{i=1}^{t} \varepsilon_{t-i} \tag{II.3}$$

Ainda em relação ao passeio randômico, se adicionarmos à Equação II.2 uma constante γ, obtemos o que se conhece por passeio randômico com *drift*, segundo a Equação II.4, que se apresenta graficamente sob a forma de uma tendência, crescente ou decrescente, a depender do valor de γ.

$$\mathrm{x}_t = \mathrm{x}_{t-1} + \gamma + \epsilon_t \tag{II.4}$$

$$= \mathrm{x}_0 + \gamma t + \sum_{i=1}^{t} \epsilon_{t-i} \tag{II.5}$$

Tendência estacionária, por sua vez, é uma série caracterizada pela presença de tendência determinística e ausência de choques aleatórios, segundo Equação II.6, na qual o componente determinístico é representado pelo termo $\beta_1 t$, enquanto β_0 e ε_t correspondem, respectivamente, ao intercepto e a um ruído branco. De forma genérica, a Equação II.6 pode ser expandida para satisfazer polinômio de ordem m (Equação II.7).

$$\mathrm{x}_t = \beta_0 + \beta_1 t + \varepsilon_t \tag{II.6}$$

$$\mathrm{x}_t = \beta_0 + \beta_1 t + \cdots + \beta_m t^m + \varepsilon_t \tag{II.7}$$

Por fim, intercalando-se com os comportamentos não estacionários, pode ocorrer a presença de sinais quasi-estacionários, i.e., nos quais as propriedades estatísticas permanecem constantes durante um curto intervalo [Parchami et al., 2016]. Embora não seja um comportamento não estacionário em si, é importante destacá-lo devido à sua utilidade em algumas aplicações. Por exemplo, sinais de voz, embora não estacionários, são usualmente analisados sob a ótica de sinais quase estacionários de intervalos fixos [Tan and Kraljevski, 2014; Parchami et al., 2016].

Para fins de exemplificação, as Figuras II.2 (a), II.2 (b), II.2 (c) e II.2 (d) apresentam, respectivamente, as séries ruído branco, passeio randômico, passeio randômico com *drift* e tendência estacionário. Observe que as séries passeio randômico com *drift* e tendência estacionário possuem um comportamento similar, embora os modelos que as governam sejam diferentes, de forma que, muitas vezes, o usuário pode erroneamente considerar que há uma relação entre ambas, quando, na verdade, não há, caracterizando o que se conhece por correlação espúria [Pernagallo, 2023].

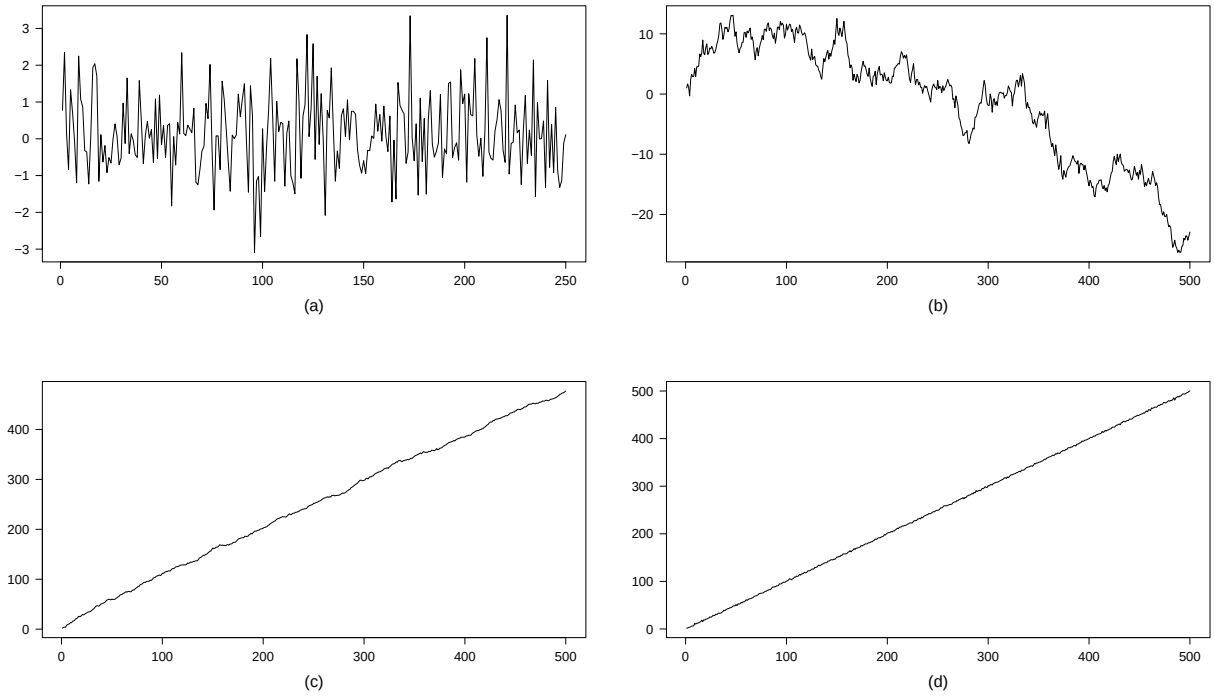


Figura II.2: Amostras de série estacionária, representada pelo ruído branco (a), e não estacionárias: passeio randômico (b), passeio randômico com *drift* (c) e tendência estacionário (d)

II.1 Funções ACF e PACF

A fim de auxiliar o usuário na análise de séries, uma ferramenta útil é a ACF, que mede a correlação entre a série em questão e uma versão defasada de si própria, i.e., para um dado valor $l = 1, 2, 3 \ldots$, a ACF mensura a dependência linear entre os pares de observações $(\mathrm{x}_1, \mathrm{x}_{1+l})$, $(\mathrm{x}_2, \mathrm{x}_{2+l}), \ldots, (\mathrm{x}_{|\mathrm{X}|-l}, \mathrm{x}_{|\mathrm{X}|})$, em que l e μ correspondem, respectivamente, à defasagem e à média da série em questão (Equação II.8).

$$\rho_l = \frac{\sum_{t=l+1}^{|\mathrm{X}|}(\mathrm{x}_t - \mu)(\mathrm{x}_{t-l} - \mu)}{\sum_{t=1}^{|\mathrm{X}|}(\mathrm{x}_t - \mu)^2} \tag{II.8}$$

Além da ACF, outra função é a PACF que, para mensurar a autocorrelação entre pares de observações $(\mathrm{x}_1, \mathrm{x}_{1+l}), (\mathrm{x}_2, \mathrm{x}_{2+l}), \ldots$, desconsidera a correlação indireta provocada pelas observações entre x_t e x_{t+l} [Enders, 2014]. Ambas as funções, individual ou conjuntamente, podem ser usadas, entre outras finalidades, para modelar e classificar séries, bem como identificar padrões repetitivos ou sazonais [Metcalfe and Cowpertwait, 2009; Box et al., 2015].

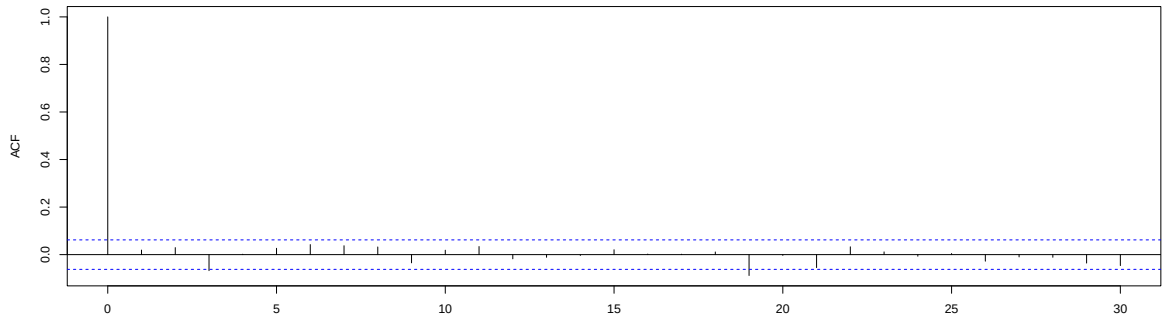


Figura II.3: ACF de um ruído branco, caracterizado pela ausência de correlação serial

II.2 Componentes de uma série

Apesar das funções ACF e PACF auxiliarem na análise do comportamento de séries, outros mecanismos são úteis. Um desses mecanismos consiste na decomposição de séries, facilitando a análise individual dos componentes delas resultantes, às vezes não observáveis na série original. Tais componentes incluem tendência, ciclos, sazonalidade e resíduos e foram originalmente relatados por Persons [1920]. Modelos de decomposição clássicos e contemporâneos são descritos por Liu et al. [2023].

Tendência apresenta um movimento sistemático durante um intervalo de longo prazo [Ord et al., 2017], podendo ser determinística ou estocástica, sendo que, em geral, o padrão apresentado pela tendência (como linear, crescente ou decrescente, quadrática e exponencial) dependerá da natureza do fator externo que a modela (como indicadores demográficos, mercado de ações e indicadores macroeconômicos) [Hoshmand, 2009].

Já a sazonalidade é caracterizada por um padrão que se repete no decorrer de um intervalo fixo e conhecido [Ord et al., 2017] que, a depender da intensidade da variação sazonal, pode ser necessário pré-processar a série de forma remover o efeito da sazonalidade, haja vista que ela pode mascarar os movimentos dos ciclos de negócios mais importantes [Neusser, 2016].

Ao contrário do que ocorre com a sazonalidade, o intervalo do padrão cíclico ao longo da série é variável e desconhecido, usualmente superior a um ano, caracterizado ainda por apresentar uma oscilação regular ao longo da tendência [Ord et al., 2017]. Ademais, os padrões cíclicos diferem em intensidade ou amplitude e estão relacionados com ciclos macroeconômicos e variações climáticas [Hoshmand, 2009; Metcalfe and Cowpertwait, 2009; Enders, 2014].

Os resíduos constituem a diferença entre a série e os demais componentes, i.e., tendência, ciclo e sazonalidade [Hyndman and Athanasopoulos, 2018], e usualmente contêm sinais ocultos não facilmente interpretáveis, de forma que entender e modelar esses sinais é frequentemente o objetivo da análise de séries [Sanchez, 2023].

Na Figura II.4, a decomposição da série CO_2, em partes por milhão (ppm), para o período de 1959 a 1997, obtida através do método *decompose*, incluído no pacote *stats*, da linguagem R. Observa-se ao longo da série uma tendência linear crescente e um padrão sazonal crescente e decrescente de mesma intensidade ao longo de cada ano. Um ponto adicional a ser observado na Figura II.4 é a ausência do componente ciclo. Isto ocorre porque a tendência e o ciclo são usualmente combinados através da tendência [Hyndman and Athanasopoulos, 2018].

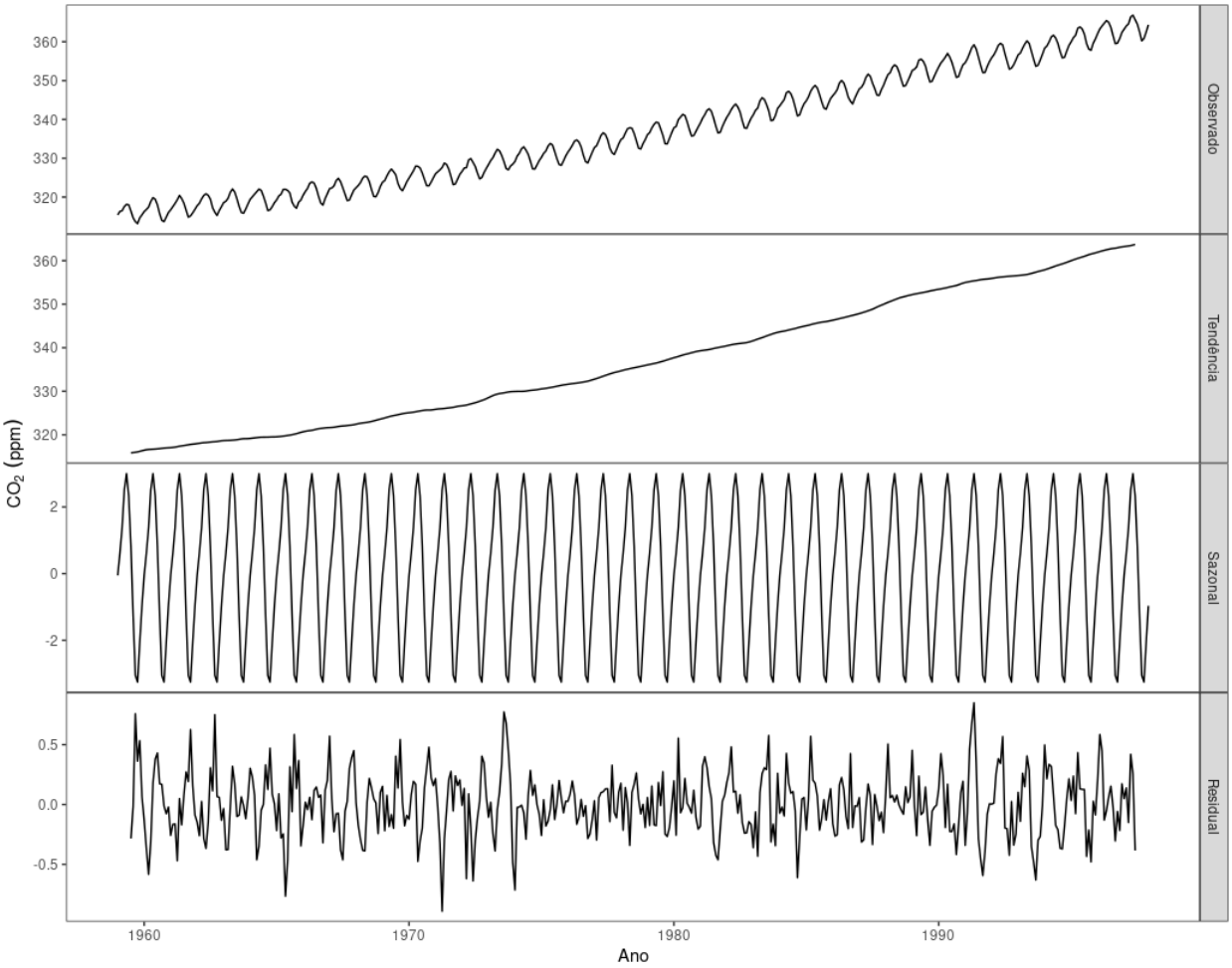


Figura II.4: Decomposição da concentração atmosférica de CO_2 para o período de 1959 a 1997

II.3 Modelos lineares

II.3.1 Regressão linear

Para séries univariadas com tendência determinística, a regressão linear é modelada relacionando as observações e os respectivos índices (Equação II.9), em que β_0 corresponde ao intercepto, β_1 representa a inclinação da reta do modelo resultante e ε_t o erro do modelo no instante t [Hyndman and Athanasopoulos, 2018].

$$\mathrm{x}_t = \beta_0 + \beta_1 t + \varepsilon_t, t = 1, 2, 3, \ldots \tag{II.9}$$

Já para séries com padrões sazonais, aplica-se um sistema de s equações generalizado pela Equação II.12, em que cada período do padrão sazonal é representado por um modelo linear com intercepto β_0 e inclinação β_1 próprios de cada período, sendo distinguidos pelo segundo subscrito, que vai de 1 a s, além do erro ε_t no instante t. Assim, para uma série na qual o padrão sazonal se repete a cada doze meses, a frequência da sazonalidade é doze, deve-se aplicar a Equação (II.10) quando os períodos forem os meses de janeiro, a Equação (II.11) quando os períodos forem os meses de fevereiro e assim por diante.

$$\mathrm{x}_{t,1} = \beta_{0,1} + \beta_{1,1} t + \varepsilon_{t,1}, t = 1, 1+s, 1+2s, \ldots \tag{II.10}$$

$$\mathrm{x}_{t,2} = \beta_{0,2} + \beta_{1,2} t + \varepsilon_{t,2}, t = 2, 2+s, 2+2s, \ldots \tag{II.11}$$

$$\vdots$$

$$\mathrm{x}_{t,s} = \beta_{0,s} + \beta_{1,s} t + \varepsilon_{t,s}, t = s, s+s, s+2s, \ldots \tag{II.12}$$

Um ponto a ser observado na regressão linear é a presença de *outliers* pois, dependendo da magnitude deste em relação às demais observações, a eficácia do modelo final ficará comprometida, devendo, portanto, ser adequadamente processado. Além disso, o usuário deve ficar atento à presença de heterocedasticidade na série. Neste caso, a série deve ser pré-processada usando, por exemplo, o método Box-Cox, a fim de estabilizar a variância.

Apesar da simplicidade e facilidade de interpretação, a regressão linear não considera correlação serial, quando uma variável se relaciona consigo mesma, comum em séries [Metcalfe and Cowpertwait, 2009].

II.3.2 Modelo Autorregressivo

No modelo AR, o valor corrente é função dos p valores precedentes, segundo a Equação II.13[1], em que ε_t ao erro, que idealmente deve se comportar como um ruído branco [Hyndman and Athanasopoulos, 2018]. O parâmetro p determina ainda a ordem do modelo AR, podendo ainda ser representado por AR(p), i.e., um modelo AR de ordem p.

[1]Na prática, pode haver a presença de uma constante. Porém, sem prejuízo dos conceitos discutidos.

$$\begin{aligned} \mathrm{x}_t &= \phi_1 \mathrm{x}_{t-1} + \cdots + \phi_p \mathrm{x}_{t-p} + \varepsilon_t \\ &= \sum_{i=1}^{p} \phi_i \mathrm{x}_{t-i} + \varepsilon_t \end{aligned} \tag{II.13}$$

Convém observar que a Equação II.13 pode ser simplificada, sem prejuízo das operações, através da introdução do operador de defasagem, denotado pela letra B, valendo a relação definida na Equação II.14. Na prática, o operador B retorna o valor observado k instantes anteriores a t.

$$B^k \mathrm{x}_t = B\mathrm{x}_{t-k} \tag{II.14}$$

De posse do operador de defasagem B, reescreve-se a Equação II.13 consoante Equação II.15, na qual a expressão $(1 - \phi_1 B - \cdots - \phi_p B^p)$ é denominada polinômio característico do modelo AR, sendo este polinômio denotado pelo operador autorregressivo $\phi(B)$ (Equação II.16).

$$\begin{aligned} \mathrm{x}_t &= \phi_1 \mathrm{x}_{t-1} + \cdots + \phi_p \mathrm{x}_{t-p} + \varepsilon_t \\ \mathrm{x}_t &= \phi_1 B\mathrm{x}_t + \cdots + \phi_p B^p \mathrm{x}_t + \varepsilon_t \\ \mathrm{x}_t - \phi_1 B\mathrm{x}_t - \cdots - \phi_p B^p \mathrm{x}_t &= \varepsilon_t \\ (1 - \phi_1 B - \cdots - \phi_p B^p)\mathrm{x}_t &= \varepsilon_t \qquad \text{(II.15)} \\ \phi(B)\mathrm{x}_t &= \varepsilon_t \qquad \text{(II.16)} \end{aligned}$$

Já a condição para que um modelo AR seja estacionário é que as raízes do polinômio característico do modelo AR residam fora do círculo unitário no plano complexo. Se pelo menos uma das raízes residir no interior do círculo unitário no plano complexo, o modelo AR é não estacionário [Box et al., 2015].

II.3.3 Modelo Média Móvel

Enquanto o modelo AR considera a relação entre o valor corrente e os valores precedentes, o MA define o valor corrente em função dos q valores precedentes gerados a partir de um ruído branco (Equação II.17). De forma análoga ao modelo AR, o parâmetro q determina a ordem do modelo MA, podendo ainda ser representado por MA(q), i.e., um modelo MA de ordem q.

O modelo MA é usado em muitas áreas do conhecimento, particularmente econometria, em que indicadores econômicos são afetados por uma variedade de eventos, como greves, decisões governa-

mentais, escassez de matéria-prima etc [Chatfield and Xing, 2019]. Ademais, embora o nome do modelo inclua a palavra média, o modelo em si é uma soma ponderada dos q valores precedentes obtidos a partir de um ruído branco. Entretanto, essa nomenclatura é bem estabelecida [Box et al., 2015].

$$\mathrm{x}_t = \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \tag{II.17}$$

O ruído branco usado na modelagem do MA é construído a partir da própria Equação II.17. Como exemplo, considere um modelo MA(1), segundo a Equação II.18.

$$\mathrm{x}_t = \theta_1 \varepsilon_{t-i} + \varepsilon_t \tag{II.18}$$

Assumindo ε_0 igual a 0, obtemos as Equações II.19, II.20, II.21 e assim sucessivamente. Por fim, sendo um modelo obtido a partir de um ruído branco, o MA é sempre estacionário, independentemente dos valores assumidos pelos coeficientes da Equação II.17 [Box et al., 2015].

$$\mathrm{x}_1 = \theta_1 \varepsilon_0 + \varepsilon_1 = \varepsilon_1 \Rightarrow \varepsilon_1 = \mathrm{x}_1 \tag{II.19}$$

$$\mathrm{x}_2 = \theta_1 \varepsilon_1 + \varepsilon_2 = \theta_1 \mathrm{x}_1 + \varepsilon_2 \Rightarrow \varepsilon_2 = \mathrm{x}_2 - \theta_1 \mathrm{x}_1 \tag{II.20}$$

$$\mathrm{x}_3 = \theta_1 \varepsilon_2 + \varepsilon_3 = \theta_1 (\mathrm{x}_2 - \theta_1 \mathrm{x}_1) + \varepsilon_3 \tag{II.21}$$

$$\vdots$$

Do mesmo modo que o modelo AR, a Equação II.17 pode ser simplificada por meio do operador de defasagem, seguindo o mesmo raciocínio aplicado ao modelo AR, resultando na Equação II.22, em que a expressão $(1 + \theta_1 B + \cdots + \theta_q B^q)$ é denominada polinômio característico do modelo MA, sendo este polinômio denotado pelo operador média-móvel $\theta(B)$ (Equação II.23).

$$\mathrm{x}_t = (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t \tag{II.22}$$

$$\mathrm{x}_t = \theta(B)\varepsilon_t \tag{II.23}$$

II.3.4 Autorregressivo Média-Móvel

O modelo ARMA combina os processos AR e MA a fim de obter um modelo mais robusto na modelagem de séries temporais. Sua forma genérica é dada pela Equação II.24, em que p corresponde à ordem do componente AR e q à ordem do componente MA, podendo ainda ser abreviada segundo a Equação II.25, em que $\phi(B)$ e $\theta(B)$ correspondem, respectivamente, aos operadores autorregressivos e média-móvel. Já a notação para o modelo ARMA segue a formato ARMA(p, q).

$$\mathrm{x}_t = \sum_{i=1}^{p} \phi_i \mathrm{x}_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{II.24}$$

$$\phi(B)\mathrm{x}_t = \theta(B)\varepsilon_t \tag{II.25}$$

Em relação à estacionariedade, assim como o modelo AR, o ARMA é estacionário se as raízes do polinômio característico do componente AR residirem fora do círculo unitário no plano complexo. Por outro lado, se pelo menos uma das raízes residir no interior do círculo unitário no plano complexo, o modelo ARMA é não estacionário [Box et al., 2015].

Vale a ressalva que o modelo ARMA pode generalizar, além dele próprio, o modelo AR e MA, se considerarmos a ordem do componente MA igual a zero, no caso do modelo AR, e a ordem do componente AR igual a zero, no caso do modelo MA (Tabela II.1).

Tabela II.1: Modelo ARMA e os respectivos modelos equivalentes AR e MA

ARMA	**Equivalência**
ARMA(p, 0)	AR(p)
ARMA(0, q)	MA(q)

Por fim, para a estimação dos coeficientes ϕ_i $(1 \leq i \leq p)$ e θ_j $(1 \leq j \leq q)$, conforme Equação II.24, o ARMA usa como base a premissa de que os erros são normalmente distribuídos com média μ e variância σ^2 (Equação II.26).

$$p(\varepsilon_t) = \frac{1}{\sqrt{2\pi{\sigma_\varepsilon}^2}} e^{-\frac{\varepsilon_t^2}{2{\sigma_\varepsilon}^2}}, t = 1, \ldots, n \tag{II.26}$$

Deseja-se, portanto, obter um modelo no qual a respectiva probabilidade satisfaça ao máximo a Equação II.27.

$$
\begin{aligned}
p(\varepsilon_1, ..., \varepsilon_n) &= p(\varepsilon_1) \times \cdots \times p(\varepsilon_n) && \text{(II.27)} \\
&= \frac{1}{\sqrt{2\pi {\sigma_\varepsilon}^2}} e^{-\frac{\varepsilon_1^2}{2{\sigma_\varepsilon}^2}} \times \cdots \times \frac{1}{\sqrt{2\pi {\sigma_\varepsilon}^2}} e^{-\frac{\varepsilon_n^2}{2{\sigma_\varepsilon}^2}} \\
&= \frac{1}{(\sqrt{2\pi {\sigma_\varepsilon}^2})^n} e^{[-\frac{\sum_{i=1}^n {\varepsilon_i}^2}{2{\sigma_\varepsilon}^2}]} \\
&= \left((2\pi {\sigma_\varepsilon}^2)^{(n/2)}\right)^{-1} e^{[-\frac{\sum_{i=1}^n {\varepsilon_i}^2}{2{\sigma_\varepsilon}^2}]} \\
&= (2\pi {\sigma_\varepsilon}^2)^{-(n/2)} e^{[-\frac{\sum_{i=1}^n {\varepsilon_i}^2}{2{\sigma_\varepsilon}^2}]} && \text{(II.28)} \\
&= L(\phi_i, \theta_j, \sigma_\varepsilon^2)
\end{aligned}
$$

Aplicando logaritmo à Equação II.28 visando simplificá-la, substituindo a operação do produto pela de adição, obtemos a Equação II.29, cuja decomposição resulta na Equação II.30, em que a porção variável corresponde à expressão $\sum_{i=1}^n {\varepsilon_i}^2$. Consequentemente, a Equação II.30 obterá seu valor máximo quando a expressão $\sum_{i=1}^n {\varepsilon_i}^2$ obter valor mínimo. Assim, parametrizam-se os coeficientes ϕ_i, θ_j e a variância σ_ε^2 de forma obter o valor mínimo dessa expressão.

$$
\begin{aligned}
\ln(L(\phi_i, \theta_j, \sigma_\varepsilon^2)) &= \ln((2\pi {\sigma_\varepsilon}^2)^{-(n/2)} e^{[-\frac{\sum_{i=1}^n {\varepsilon_i}^2}{2{\sigma_\varepsilon}^2}]}) && \text{(II.29)} \\
&= \ln((2\pi {\sigma_\varepsilon}^2)^{-(n/2)}) + \ln(e^{[-\frac{\sum_{i=1}^n {\varepsilon_i}^2}{2{\sigma_\varepsilon}^2}]}) \\
&= -\frac{n}{2}\ln(2\pi {\sigma_\varepsilon}^2) - \frac{\sum_{i=1}^n {\varepsilon_i}^2}{2{\sigma_\varepsilon}^2} && \text{(II.30)}
\end{aligned}
$$

II.4 Tratamento da não estacionariedade

A condição não estacionária de uma série pode ser causada por fatores como tendência, heterocedasticidade (variância não constante), raiz unitária, quebra estrutural, sazonalidade, ciclos, entre outros, sendo a tendência, determinística ou estocástica, o motivo mais comum pelo qual uma série não é estacionária [Enders, 2014].

A fim de lidar com esses diferentes fatores, diversos métodos de pré-processamento de séries não estacionárias em estacionárias foram criados ou adaptados para obter séries estacionárias. O próprio método da Integração foi originalmente aplicado em outros contextos diferentes daquele aplicado por Box et al. [2015] [Tintner, 1940; Rao and Tintner, 1963]. Nesta seção, esses métodos são discutidos, destacando, entre outros, suas características, aplicabilidade, limitações, etc.

II.4.1 Box-Cox

O método Box-Cox [Box and Cox, 1964] consiste na aplicação de uma família de funções exponenciais e logarítmicas (Equação II.31), visando obter linearidade, homoscedasticidade e normalidade, embora alguns estudos indiquem que essas características, em geral, não possam ser obtidas simultaneamente [Sakia, 1992], sendo que, no âmbito de modelos de séries lineares univariadas, é usualmente aplicado para estabilizar a variância, quando estas apresentam heterocedasticidade.

Originalmente, o método Box-Cox só admite valores positivos. Como solução de contorno, Box and Cox [1964] propuseram um parâmetro de deslocamento, λ_2, sendo λ_2 escolhido de forma que $\mathrm{x}_t > -\lambda_2$ [Sakia, 1992]. Caso a série contenha somente valores positivos, admite-se λ_2 igual a zero.

$$\mathrm{y}_t = \begin{cases} \frac{(\mathrm{x}_t+\lambda_2)^{\lambda_1}-1}{\lambda_1}, \lambda_1 \neq 0. \\ \log(\mathrm{x}_t + \lambda_2), \lambda_1 = 0. \end{cases} \tag{II.31}$$

Já para o caso no qual a série contenha valores negativos, a menos que o valor de λ_2 seja conhecido *a priori*, o parâmetro λ_1 pode não ser corretamente configurado por meio da MLE, em virtude de mínimos locais, dado que o parâmetro de deslocamento λ_2 influencia o intervalo dos valores ao qual é aplicada a MLE [Atkinson and Pericchi, 1991]. Como alternativa, além da extensão do método Box-Cox proposta por Yeo and Johnson [2000], há ainda o método proposto por Bickel and Doksum [1981] [Yeo and Johnson, 2000].

II.4.2 Integração

Box et al. [2015] introduziram a expressão Integrada para se referir ao método de pré-processamento que usa sucessivas diferenciações para obter uma série estacionária. Pelo método da diferenciação, uma nova série é obtida subtraindo observações consecutivas (Equação II.32). No entanto, embora seja comum associar a diferenciação como um método para se obter uma série estacionária, uma única diferenciação não garante obter uma série estacionária.

$$\begin{aligned} \mathrm{y}_t &= \mathrm{x}_t - \mathrm{x}_{t-1} \\ &= \mathrm{x}_t - \mathrm{x}_t B \\ &= (1-B)\mathrm{x}_t \end{aligned} \tag{II.32}$$

Assim, após aplicar uma diferenciação, caso a série resultante ainda não seja estacionária, aplica-se novamente mais uma diferenciação e assim sucessivamente, até obter uma série estacionária, de forma que a quantidade de diferenciações aplicada determina a ordem da Integração. Todavia, a fim

de generalizar a aplicação dessas sucessivas diferenciações, usa-se a expressão binominal $(1-B)^d$, em que d corresponde à ordem da Integração e B ao operador de defasagem (Equação II.14). Assim, a Integração pode ser expressa consoante Equação II.33.

Pela Equação II.33, observa-se que, a menos que a série em questão já seja estacionária, a Integração requer uma ou mais diferenciações para obter uma série estacionária. Na prática, porém, séries em geral requer até duas diferenciações [Box et al., 2015]. No entanto, para uma série modelada por um polinômio de grau m, Integração de ordem m é necessária para obter a respectiva série estacionária, de forma que há uma correspondência entre o grau do polinômio e a ordem da Integração [Metcalfe and Cowpertwait, 2009].

$$\mathrm{y}_t = (1-B)^d \mathrm{x}_t, d = 0, 1, 2, \ldots \tag{II.33}$$

Para fins de análise, embora os padrões presentes na série original sejam descaracterizados ao aplicar a Integração, para séries integradas de ordem um, a aplicação da Integração fornece uma visão da diferença entre observações adjacentes (Equação II.34). Já para séries integradas de ordem igual ou superior a dois, a série pré-processada passa a apresentar maior dificuldade de interpretação, como demonstrado na Equação II.35.

$$\begin{aligned}(1-B)^1 \mathrm{x}_t &= \mathrm{x}_t - \mathrm{x}_t B \\ &= \mathrm{x}_t - \mathrm{x}_{t-1}\end{aligned} \tag{II.34}$$

$$\begin{aligned}(1-B)^2 \mathrm{x}_t &= \mathrm{x}_t - 2\mathrm{x}_t B + \mathrm{x}_t B^2 \\ &= \mathrm{x}_t - 2\mathrm{x}_{t-1} + \mathrm{x}_{t-2}\end{aligned} \tag{II.35}$$

Um ponto a ser observado em séries integradas é que elas possuem memória de curto prazo, no sentido de que a função ACF decai exponencialmente a zero. Frente a isso, a aplicação subsequente de diferenciações a séries integradas, cenário conhecido como sobrediferenciação, produz correlação serial adicional, provocando um decaimento mais lento quando comparado com uma série integrada e, assim, aumentando a complexidade do modelo [Box et al., 2015].

II.4.3 Integração fracionária

Diferentemente de séries integradas, nas quais a função ACF decai exponencialmente a zero, séries de domínios como hidrologia, ciências geofísicas, finanças e economia podem apresentar memória de longo prazo [Box et al., 2015]. Neste caso, diferentemente do que ocorre com séries com

memória de curto prazo, a função ACF decresce hiperbolicamente [Hosking, 1981].

Pela definição de McLeod and Hipel [1978], uma série apresenta persistência de longo prazo se o somatório das autocorrelações resultar em um valor infinito (Equação II.36). Caso contrário, a série é classificada como de curto prazo. Na prática, se o gráfico da função ACF decrescer lentamente, pode indicar a presença de memória de longo prazo [Tsay, 2010].

$$\sum_{l=-\infty}^{\infty} |\rho_l| = \infty \tag{II.36}$$

Acontece que em séries com memória de longo prazo, a aplicação da Integração descarta essa correlação observada no longo prazo, sendo mais adequada para séries que apresentem memória de curto prazo [Liu et al., 2017]. Diante dessa limitação, Hosking [1981] e Granger and Joyeux [1980] expandiram a Equação II.33 para o domínio real, formalizando a Integração fracionária (Equação II.37).

$$\begin{aligned}(1-B)^d &= 1 - dB + \frac{d(d-1)}{2!}B^2 - \frac{d(d-1)(d-2)}{3!}B^3 + \cdots \\ &= \sum_{i=0}^{\infty} \frac{\prod_{j=0}^{i-1}(d-j)}{i!}(-B)^i\end{aligned} \tag{II.37}$$

Teoricamente, essa memória de longo prazo fica mais evidente à medida que o valor do parâmetro d se aproxima de 0.5. Por outro lado, à medida que o valor de d se aproxima de 0, o comportamento da função ACF se aproxima de uma série com memória de curto prazo. Na Figura II.5 a), a ACF de uma série integrada, *i.e.*, $d = 0$, na qual se observa o decaimento exponencial da ACF.

Já as Figuras II.5 (b), II.5 (c), II.5 (d), II.5 (e) e II.5 (f) apresentam a ACF de séries fracionárias integradas, nas quais o parâmetro d assume, respectivamente, os valores 0.5, 0.15, 0.25, 0.35 e 0.45. Caso a Integração fosse aplicada a qualquer uma delas, a ACF da série resultante seria equivalente à da Figura II.5 (a), descartando a correlação observada no longo prazo, apesar dessa correlação manifestar-se de forma mais enfática no intervalo $0.25 < d < 0.5$.

Em relação à estacionariedade, o intervalo do parâmetro d deve ser $-0.5 < d < 0.5$ [Hosking, 1981]. Para o intervalo $0 < d < 0.5$, a série apresenta memória de longo prazo, enquanto para $-0.5 < d < 0$, a soma dos valores absolutos da ACF limita-se a uma constante, de forma que a série é classificada como de curto prazo, segundo a definição da Equação II.36 [Baillie, 1996].

Concluindo, a Integração fracionária é aplicável somente quando a série for integrada de ordem um. Caso contrário, é necessário pré-processá-la a fim de obter uma série integrada de ordem um para, posteriormente, aplicar a Integração fracionária [Tsay, 2010].

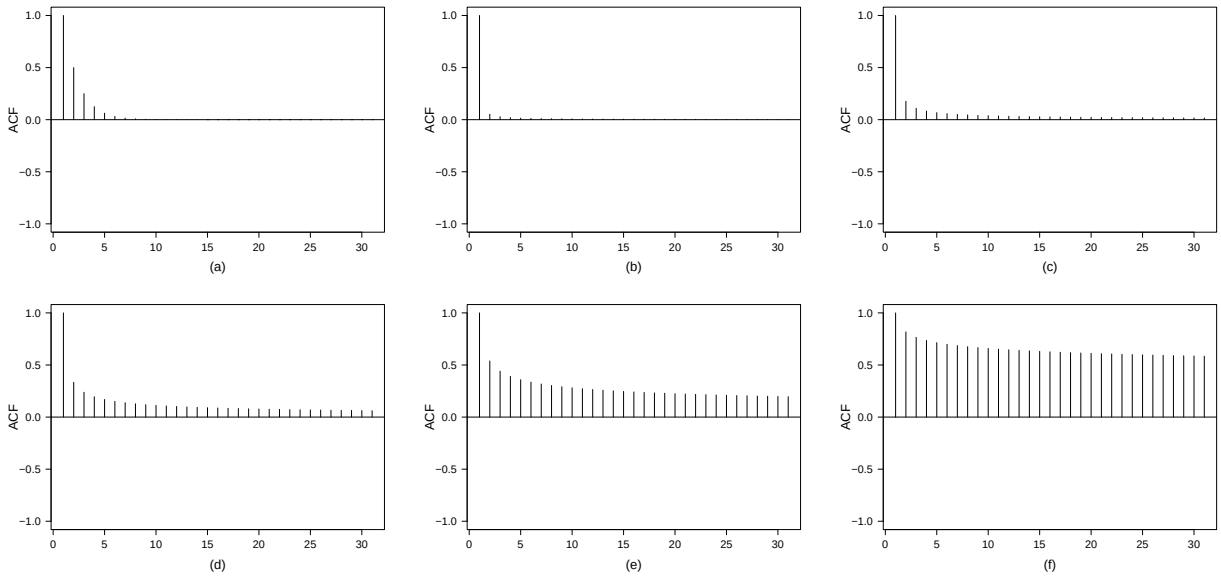


Figura II.5: ACF para diferentes valores de Integração fracionária

II.4.4 *Detrending*

O método *detrending* aplica-se quando a série segue um modelo tendência estacionário (Equação II.6) e consiste na remoção do componente determinístico $\beta_1 t$, responsável pela média variável da série, resultando na Equação II.38. Na prática, porém, haja vista que o componente determinístico é desconhecido *a priori*, ajusta-se um modelo regressivo à série em questão, sendo o resíduo desse modelo a série estacionária [Levendis, 2018].

$$\mathrm{x}_t = \beta_0 + \varepsilon_t \tag{II.38}$$

Assim como acontece com a Integração, o método *detrending* é aplicável a qualquer série cujo modelo subjacente seja um polinômio de grau m. Porém, vale a ressalva que, enquanto o método *detrending* preserva o ruído ε_t da série original (Equação II.38), a aplicação da Integração retorna um modelo MA de ordem um somado a uma constante, descaracterizando a série quando comparado com o método *detrending*, pois do ruído ε_t é subtraído o valor ε_{t-1} (Equação II.39).

$$\begin{aligned}(1-B)^1\mathrm{x}_t &= \mathrm{x}_t - \mathrm{x}_t B \\ &= \mathrm{x}_t - \mathrm{x}_{t-1} \\ &= \beta_0 + \beta_1 t + \varepsilon_t - (\beta_0 + \beta_1(t-1) + \varepsilon_{t-1}) \\ &= \beta_1 + \varepsilon_t - \varepsilon_{t-1}\end{aligned} \tag{II.39}$$

Já para o cenário no qual a série apresenta tendência estocástica, o método *detrending* não é aplicável [Enders, 2014], sendo necessário aplicar outro método como, por exemplo, integração. Nesse cenário, o componente determinístico da Equação II.6 é substituído por um componente estocástico, como pode ser observado no passeio randômico com *drift* (Equação II.5).

II.4.5 Transformação por variação percentual

Como demonstrado na Subseção II.4.2, a aplicação de Integração de ordem 1 fornece uma visão em relação à variação absoluta da série (Equação II.34). Em contrapartida, pode ser desejável, dependendo dos requisitos do usuário, obter uma variação relativa (Equação II.40), denominada PCT. Comparada com a variação absoluta, a série resultante da PCT fornece uma flutuação mais consistente quando comparada com aquela decorrente da Integração de ordem 1 [Ord et al., 2017].

$$\mathrm{y}_t = 100 \left(\frac{\mathrm{x}_t - \mathrm{x}_{t-1}}{\mathrm{x}_{t-1}} \right) \tag{II.40}$$

Outra opção à variação relativa é por meio da LT (Equação II.41). Porém, Cryer and Chan [2008] frisa que a LT assume que há uma relativa variação percentual entre observações adjacentes. Mais precisamente, a Equação II.41 é válida se a variação percentual for restrita ao intervalo $[-100\rho, 100\rho]$, no qual ρ $(0 \leq \rho \ \leq 1)$ assume um pequeno limiar aceitável [Salles et al., 2019].

$$\mathrm{y}_t \approx \log \left(\frac{\mathrm{x}_t}{\mathrm{x}_{t-1}} \right) \tag{II.41}$$

Ademais, a LT possui a vantagem secundária de estabilizar a variância, embora a PCT forneça esse mesmo diferencial, de forma que há uma propensão de que ambos os métodos produzam resultados similares [Ord et al., 2017].

II.4.6 Transformada de Fourier

A FT decompõe uma série em um conjunto de sinais sinusoidais de diferentes frequências, sendo cada de natureza estacionária. Dado que as séries abrangidas por esta dissertação residem no domínio discreto, aplica-se a DFT, na qual uma série contendo n observações é decomposta em n sinais sinusoidais (Equação II.42). Porém, na prática, devido ao custo computacional, usa-se a FFT [Song et al., 2021].

$$X(\omega) = \sum_{j=1}^{n} \mathrm{x}_j \cdot e^{-i2\pi\omega j/n}, \omega = 1, 2, 3, \ldots, n \tag{II.42}$$

Visto que o resultado da FT reside no domínio da frequência, a menos que a série em questão seja estacionária, dependendo da aplicação por parte do usuário, este fica impossibilitado de aplicá-la quando a aplicação requerer informações relacionadas a um intervalo específico. Como solução de contorno, aplica-se a STFT, na qual a FT é aplicada somente ao intervalo em questão [Guido, 2022].

Entretanto, a aplicação da STFT possui uma limitação: intervalos maiores podem capturar baixais frequências, mas não podem afirmar com precisão o instante de altas frequências; por outro lado, intervalos menores podem capturar altas frequências, mas podem não conseguir capturar baixas frequências [Guido, 2022]. Frente a esse paradoxo, o usuário deve considerar a aplicação e, assim, ajustar o intervalo segundo as necessidades da aplicação.

II.4.7 Transformada wavelets

A WT surgiu como uma alternativa à limitação da FT ao domínio da frequência. Assim como a FT, a WT usa duas funções-base, denominadas funções de escala e função wavelets, que satisfazem as Equações $\int \varphi(t)dx = 1$ e $\int \psi(t)dx = 0$, em que φ e ψ correspondem às funções de escala e wavelets, respectivamente. Outro diferencial da WT em relação à FT é que ela não se limita às funções de comportamento fixo, como as funções sinusoidais usadas na FT, diferencial que permite aplicações específicas conforme as características das funções de escala e wavelets.

Desde a sua concepção, principalmente na década de 80, dada a contribuição de diversos autores [Graps, 1995], a WT vem sendo continuamente empregada nos mais diversos campos como, por exemplo, síntese de imagens, análise e processamento de sinais, compressão de dados, reconhecimento de padrões etc [Rhif et al., 2019; Guido, 2022].

Através da WT, uma série $\mathrm{x}_t \in L^2$, i.e., o espaço das funções quadraticamente integráveis, pode ser decomposta segundo a Equação II.43, em que N corresponde ao tamanho da série e J ao nível de decomposição. Já $s_{j,k}$ representa o coeficiente de aproximação na escala j e localização k enquanto $d_{j,k}$ representa o coeficiente de detalhe na escala j e localização k [Joo and Kim, 2015].

$$\mathrm{x}_t = \sum_{k=1}^{N} s_{j,k}\varphi_{j,k}(t) + \sum_{j=1}^{J}\sum_{k=1}^{N} d_{j,k}\psi_{j,k}(t) \tag{II.43}$$

Na prática, porém, a decomposição da série conforme a Equação II.43 é implementada por meio

de um banco de filtros, no qual o filtro passa-alta é associado com a função wavelet, enquanto o filtro passa-baixa é associado com a função de escala [Hubbard, 1998]. Aplicadas a uma série, os wavelets podem capturar descontinuidades, rupturas e singularidades em altas frequências, enquanto em baixas frequências podem ser usados para identificar tendências de longo prazo [Lahmiri, 2014].

Além das propriedades já citadas, uma propriedade dos wavelets que pode ser considerada para obter uma série estacionária é os momentos nulos. Formalmente, uma função $f(x)$ possui n momentos nulos se a condição da Equação II.44 for satisfeita [Jarrah and Khanna, 2019]. Na prática, essa propriedade permite eliminar ou preservar da série algum comportamento que possa ser modelado segundo um polinômio [Fugal, 2009].

$$\int x^v f(x)dx = 0, 0 \leq v \leq k-1 \tag{II.44}$$

Usando como exemplo uma tendência determinística segundo a Equação II.6, a expressão $\beta_0+\beta_1 t$ corresponde a um polinômio de primeiro grau. Assim, aplicando-se à Equação II.6 um wavelet com 2 momentos nulos, obter-se-á uma série estacionária, em virtude do descarte do polinômio de primeiro grau (Figura II.6 (b)).

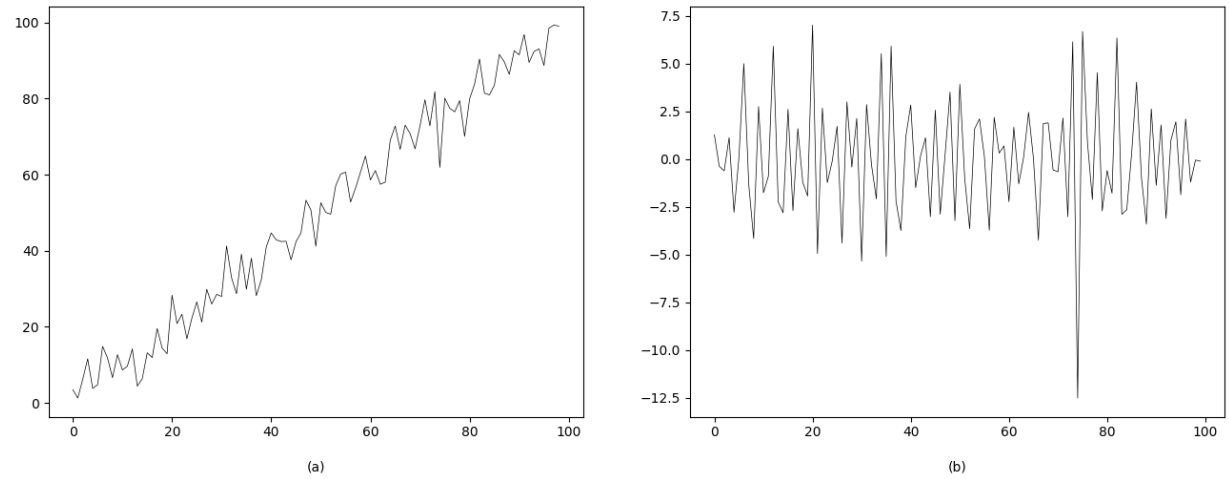


Figura II.6: Tendência estacionária em (a) e a série resultante decorrente da remoção da tendência por meio de WT em (b)

II.4.8 Integração sazonal

A Integração não considera a presença de sazonalidade, se houver, na série. Neste caso, aplica-se a Integração sazonal (Equação II.45), em que D corresponde à ordem da Integração sazonal, e s à frequência da sazonalidade. Por exemplo, no caso de uma série na qual a sazonalidade se repete a cada doze meses, o valor de s é doze.

$$(1 - B^s)^D \mathrm{y}_t = \mathrm{x}_t - \mathrm{x}_t B^s + \cdots \tag{II.45}$$
$$= \mathrm{x}_t - \mathrm{x}_{t-s}$$

Vale a ressalva de que Integração sazonal pode ou não ser sucedida de Integração, dependendo da série resultante ser ou não estacionária. Para o cenário no qual a Integração sazonal resulta em uma série não estacionária, a Integração sazonal pode ser combinada com a Integração em uma única equação, sem prejuízo do resultado, conforme Equação II.46.

$$(1 - B^s)^D (1 - B)^d \mathrm{x}_t = (1 - B^s)^D (\mathrm{x}_t - \mathrm{x}_t B + \cdots) \tag{II.46}$$
$$= (\mathrm{x}_t - \mathrm{x}_t B^s + \cdots)(\mathrm{x}_t - \mathrm{x}_t B + \cdots)$$
$$= (\mathrm{x}_t - \mathrm{x}_{t-s} + \cdots)(\mathrm{x}_t - \mathrm{x}_{t-1} + \cdots)$$

Ainda em relação à Integração sazonal, caso plausível para o usuário, uma alternativa é dessazonalizar a série, usando métodos como Integração sazonal ou removendo o componente sazonal por meio de métodos de decomposição de séries [Liu et al., 2023].

Capítulo III Trabalhos relacionados

Paralelamente à Integração e Integração sazonal, métodos de pré-processamento descritos no capítulo anterior, Box et al. [2015] inovaram ao combinar esses métodos com o modelo ARMA, resultando em uma categoria de modelos que combinam pré-processamento e modelagem. Posteriormente, principalmente na década de 80, outros modelos combinados surgiram, inclusive alguns dos quais a modelagem é não linear. Como o escopo desta dissertação considera somente modelos lineares, o presente capítulo foca em modelos combinados cuja modelagem seja de natureza linear.

III.1 Autorregressivo Integrado Média-Móvel

O método de Integração incorporado ao modelo ARMA dá origem ao ARIMA [Box et al., 2015] (Equação III.1). Um ponto a ser observado no ARIMA é que, embora Box et al. [2015] embarque somente a Integração como método para obter a série estacionária, em alguns casos, a aplicação conjunta do método Box-Cox e da Integração é necessária.

Isso ocorre porque a Integração possui limitação em séries que apresentem comportamento heterocedástico, devendo neste caso ser precedida pelo método Box-Cox, de forma que a distribuição dos erros do modelo ARIMA resultante apresente comportamento homocedástico.

$$\phi(B)(1-B)^d \mathrm{x}_t = \theta(B)\varepsilon_t \tag{III.1}$$

III.2 Autorregressivo Fracionário Integrado Média-Móvel

A Integração fracionária aplicada no contexto do modelo ARMA origina o ARFIMA, cuja equação característica é a mesma do ARIMA (Equação III.1), exceto pelo intervalo $0.5 < d < 0.5$ [Tsay, 2010], sendo, portanto, uma extensão do ARIMA. Adicionalmente, o usuário deve selecionar um método adequado para determinar os parâmetros do modelo ARFIMA pois, a depender do método escolhido, o modelo final pode ser constituído de diversos parâmetros, prejudicando por conseguinte a estimação deles [Brockwell and Davis, 1991].

III.3 Sazonal Autorregressivo (Fracionário) Integrado Média-Móvel

Para séries que apresentem padrão sazonal, a Integração sazonal aplicada ao modelo ARIMA dá origem ao modelo SARIMA, consoante Equação III.2, em que $(1 - B^s)^D$ corresponde à Integração sazonal, sendo s o padrão sazonal presente na série, enquanto $(1 - B)^d$ corresponde à Integração.

$$\phi(B)(1 - B^s)^D(1 - B)^d \mathrm{x}_t = \theta(B)\varepsilon_t \tag{III.2}$$

Convém ressaltar que o SARIMA, assim como a Integração sazonal, é aplicável somente quando houver um padrão sazonal. Por exemplo, caso a série apresente um padrão sazonal mensal em um padrão sazonal anual, nem o SARIMA nem a Integração sazonal são aplicáveis.

III.4 Autorregressivo Integrado Média-Móvel com variáveis exógenas

A incorporação de variáveis exógenas ao modelo ARIMA introduz o modelo ARIMAX (Equação III.3), em que k corresponde à quantidade de variáveis exógenas, sendo estas denotadas por y_{it} e os respectivos coeficientes por β_i. Assim como o modelo ARIMA, é usada como um modelo preditivo em vários campos [Kim et al., 2023].

$$\phi(B)(1 - B)^d \mathrm{x}_t = \theta(B)\varepsilon_t + \sum_{i=1}^{k} \beta_i \mathrm{y}_{it} \tag{III.3}$$

Capítulo IV Metodologia

Conforme visto no Capítulo I, o WARMA processa as propriedades estatísticas sequencialmente, baseado no conceito de passos, que indica a quantidade de passos necessários para obter a série estacionária. Assim, caso a série seja estacionária, nenhum passo ou 0 passos são necessários para obter a série estacionária. Caso contrário, processa-se primeiramente a média, indicando 1 passo. Persistindo a não estacionariedade, processa-se a variância, indicando 2 passos (média e variância).

No caso da média, o WARMA faz uso da AN, composto pelas seguintes fases, processadas sequencialmente: SW, normalização da média e processamento de *outliers*. A AN, além de obter uma série processada com média constante por meio da fase normalização da média, é responsável por processar os *outliers*. Seguindo a AN, a normalização da variância é responsável por obter a variância constante. Por fim, na fase de modelagem, aplica-se o ARMA para obter o modelo final. Na Figura IV.1, é mostrado o fluxo geral do WARMA.



Figura IV.1: Visão geral do WARMA

IV.1 Pré-processamento

IV.1.1 Janelas deslizantes

SW é um procedimento através do qual subsequências são obtidas a partir de uma série, às quais são usualmente aplicadas operações matemáticas, visando alguma finalidade. Formalmente, para uma série X de tamanho |X|, subsequências de tamanho w ($3 \leq w \leq |X|$) são organizadas sob a forma de uma matriz, de dimensão h por w, em que h corresponde à Equação IV.1, na qual o parâmetro τ ($1 < \tau \leq w$) corresponde ao deslocamento entre subsequências adjacentes (Figura

IV.2). Já a expressão $\lfloor \frac{(|\mathrm{X}|-w)}{\tau} \rfloor$ obtém o maior número inteiro menor ou igual a $\frac{(|\mathrm{X}|-w)}{\tau}$ [Xu et al., 2019].

$$h = \lfloor \frac{(|\mathrm{X}| - w)}{\tau} \rfloor + 1 \tag{IV.1}$$

Na Figura IV.2, aplicação do SW para subsequências de tamanho igual a vinte são obtidas configurando τ igual a um em (a) e τ igual a vinte em (b). Para $\tau < w$, as subsequências são denominadas sobrepostas, por compartilharem um dado intervalo da série da qual derivam. Por outro lado, para $\tau = w$, as subsequências são denominadas não sobrepostas.

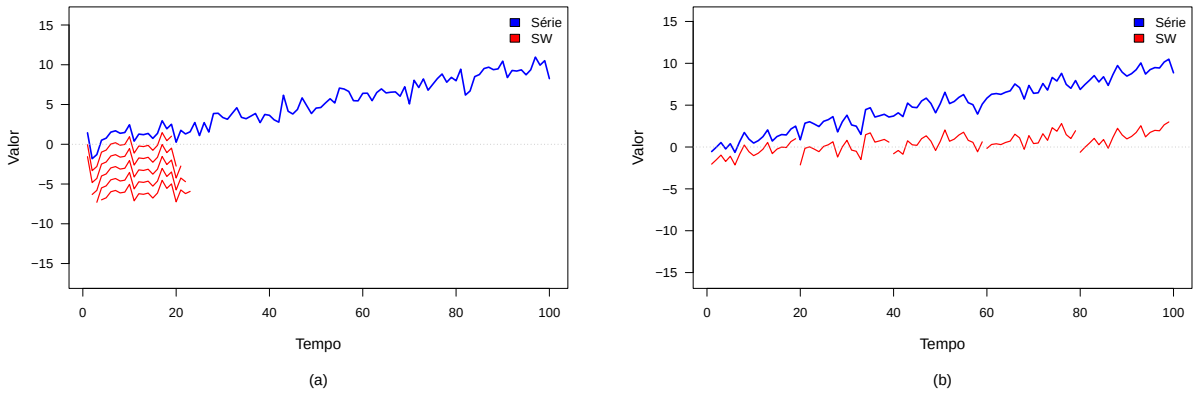


Figura IV.2: Série sintética e as respectivas subsequências obtidas a partir do procedimento de SW para w igual a 20 e τ igual a 1 em (a) e a mesma série as respectivas subsequências para w igual a 20 e τ igual a 20 em (b)

Uma consequência direta da Equação IV.1 é que, por envolver a divisão de inteiros, as primeiras $(|\mathrm{X}| - w) \bmod \tau$ observações da série X são desconsideradas. Com base nisso, definimos a i-ésima subsequência de X, denotada por $seq_{i,w,\tau}(\mathrm{X})$, conforme a Equação IV.2 [Ali et al., 2019]. Além disso, definimos, para fins de brevidade, o parâmetro o (Equação IV.3), pois esse mesmo parâmetro será usado na definição do modelo WARMA.

$$seq_{i,w,\tau}(\mathrm{X}) = \langle \mathrm{x}_{1+\mathrm{o}+(i-1)\tau}, \mathrm{x}_{2+\mathrm{o}+(i-1)\tau}, \ldots, \mathrm{x}_{w+\mathrm{o}+(i-1)\tau} \rangle, \tag{IV.2}$$

$$\mathrm{o} = (|\mathrm{X}| - w) \mod \tau \tag{IV.3}$$

De posse da Equação IV.2 e denotando por $\mathbf{A}$ a matriz resultante do SW, mapeamos o j-ésimo elemento da i-ésima subsequência, denotado por $(seq_{i,w,\tau}(\mathrm{X}))_j$, para o j-ésimo elemento da i-ésima linha da matriz $\mathbf{A}$, denotado por a_{ij}, segundo a Equação IV.4. Ademais, como cada linha de uma

matriz pode ser vista como um vetor, usaremos ao longo deste capítulo a expressão vetor-linha para se referir a uma linha da matriz.

$$a_{ij} = (seq_{i,w,\tau}(\mathrm{X}))_j, \forall a_i \in \mathbf{A}, \forall i \in [1,h], \forall j \in [1,w] \tag{IV.4}$$

No caso específico do WARMA, objetiva-se com o SW, para uma dada série, avaliar diferentes valores de w e τ de forma que as respectivas subsequências capturem o comportamento inercial da série, i.e., subsequências nas quais as propriedades estatísticas permaneçam constantes, de forma obter uma série estacionária.

IV.1.2 Normalização da média

Logo após o SW, aplica-se a normalização da média, cujo objetivo é fazer com que a média, medida de posição ou localização, dos elementos dos respectivos vetores-linha da matriz $\mathbf{A}$, resultante da fase SW, seja constante e igual a zero, satisfazendo a condição de média constante. Para isso, para cada vetor-linha da matriz $\mathbf{A}$, subtrai-se dos respectivos elementos do vetor-linha a respectiva média (Equação IV.5), em que $f(a_i)$ corresponde à média dos elementos do vetor-linha em questão (Equação IV.6), resultando em uma matriz $\mathbf{B}$, de mesma dimensão da matriz de entrada $\mathbf{A}$, cujas médias dos elementos dos respectivos vetores-linha são iguais a zero.

$$b_{ij} = a_{ij} - f(a_i), \forall a_i \in \mathbf{A}, \forall i \in [1,h], \forall j \in [1,w] \tag{IV.5}$$

$$f(v) = \frac{\sum_{j=1}^{|v|} v_j}{|v|} \tag{IV.6}$$

IV.1.3 Processamento de *outliers*

Assim como em outras fases do ciclo de vida de um projeto de análise de dados [Ogasawara et al., 2010; Cousineau and Chartier, 2010; Aguinis et al., 2013], o processamento inadequado de *outliers* pode levar a um modelo ineficaz, em que *outlier* é uma observação que se desvia consideravelmente das demais observações, levantando a suspeita de que foi gerada por um mecanismo diferente [Hawkins, 1980].

No pré-processamento do WARMA, as fases de Normalização da média e de Normalização da variância usam duas propriedades estatísticas sensíveis a *outliers*: média e variância. No caso da Normalização da média, uma subsequência associada a um dado vetor-linha, mesmo com média igual a zero após a aplicação da Normalização da média, pode estar distante das demais subsequências,

justamente pela presença de *outliers*.

Usando como exemplo a Figura IV.3 (a), a subsequência à esquerda da linha tracejada contém um *outlier*, medindo 250 no instante 50, enquanto a subsequência à direita da linha tracejada não contém *outliers*. Após a aplicação da Normalização da média (Figura IV.3 (b)), as observações da subsequência à esquerda da linha tracejada estão concentradas em volta do valor -2,50, embora possuam média igual a zero, devido à presença do *outlier*.

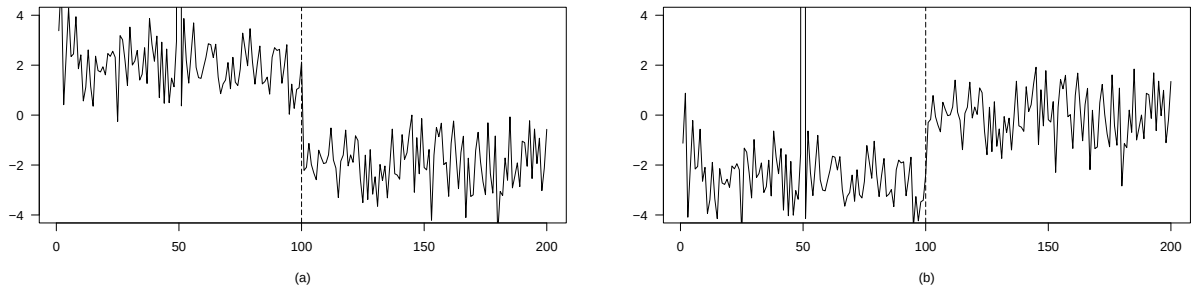


Figura IV.3: Influência de *outlier* após a fase Normalização da média

Diante dessa situação, na AN, quaisquer vetor-linha da matriz de entrada $\mathbf{B}$, no qual um ou mais elementos não satisfaçam o IQR, conforme Equação IV.7, na qual o IQR é definido considerando os elementos da matriz $\mathbf{B}$ (IQR($\mathbf{B}$)), são desconsiderados durante a fase de modelagem. Esse mesmo critério é empregado para identificar *outliers* no gráfico *boxplot* [Iglewicz and Hoaglin, 1993].

$$Q_1 - 1.5 \times \text{IQR}(\mathbf{B}) \leq b_{ij} \leq Q_3 + 1.5 \times \text{IQR}(\mathbf{B}), \forall b_i \in \mathbf{B}, \forall i \in [1, h], \forall j \in [1, w] \tag{IV.7}$$

Observa-se ainda que, embora originalmente a presente fase vise desconsiderar subsequências que contenham *outliers*, seu escopo é mais amplo, podendo ser usada para identificar mudança de nível, quebra estrutural ou ainda algum outro elemento que provoque não estacionariedade na série original.

IV.1.4 Normalização da variância

Aplicado o passo 1, caso a série permaneça não estacionária, aplica-se a normalização da variância, cujo objetivo é fazer com que a variância, medida de dispersão ou variabilidade, dos elementos dos respectivos vetores-linha que compõem a matriz $\mathbf{B}$, usada como entrada nesta fase, seja constante e igual a um, resultando em dois passos, um para a média e outro para a variância.

Neste passo, para cada vetor-linha da matriz $\mathbf{B}$, dividem-se os respectivos elementos do vetor-linha pelo respectivo desvio-padrão (Equação IV.8), em que $f(b_i)$ corresponde à média dos elementos do vetor-linha em questão (Equação IV.6), resultando em uma matriz $\mathbf{C}$, de mesma dimensão da

matriz de entrada **B**, cujas variâncias dos elementos dos respectivos vetores-linha são iguais a um.

Não obstante, para o caso em que o denominador da Equação IV.8, correspondente ao desvio-padrão, for igual a zero, os elementos resultantes do vetor-linha serão zero.

$$c_{ij} = \frac{b_{ij}}{\sqrt{\frac{1}{|b_i|}\sum_{k=1}^{|b_i|}\left(b_{ik} - f(b_i)\right)^2}}, \forall b_i \in \mathbf{B}, \forall i \in [1, h], \forall j \in [1, w] \tag{IV.8}$$

Por fim, cabe frisar que, embora a Equação IV.8 use o desvio-padrão em vez da variância no denominador, o nome da fase poderia ou deveria se chamar normalização do desvio-padrão. No entanto, quando se refere à série estacionária, a propriedade estatística usualmente referenciada é a variância. À vista disso e considerando que o desvio-padrão e a variância são relacionados matematicamente, optamos pela denominação normalização da variância.

Na Figura IV.4, revisitamos o pré-processamento do WARMA, descrito sob a forma de fluxograma. Inicialmente, a série é avaliada segundo a condição de estacionariedade. Caso estacionária, o fluxo se encerra. Caso contrário, incrementa o WARMA em um passo e aplicam-se os procedimentos associados ao passo um, i.e., janelas deslizantes, normalização da média e, por último, o processamento dos *outliers*.

Aplicado o passo um, avalia-se novamente a condição de estacionariedade. Caso estacionária, o fluxo se encerra. Caso contrário, incrementa o WARMA em mais um passo, resultando em dois passos. No passo dois, aplica-se a normalização da variância, encerrando o fluxo do pré-processamento do WARMA.

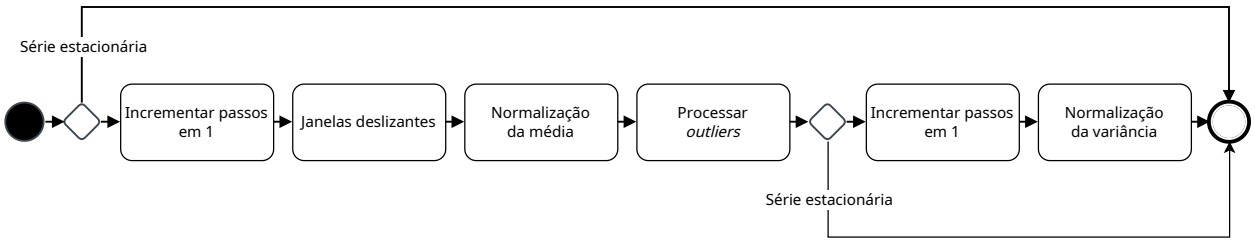


Figura IV.4: Fluxo do pré-processamento do WARMA

IV.1.5 Vetorizando a série pré-processada

Conforme visto nas fases que compõem o pré-processamento, as respectivas entradas e saídas são matrizes de mesmas dimensões. Entretanto, outros modelos de séries podem usar uma versão vetorial em vez de matricial. Diante desse cenário, descrevemos nesta subseção como obter uma forma vetorial a partir da matriz resultante do pré-processamento.

Para isso, usa-se dois mecanismos da álgebra linear: submatriz e vetorização matricial. Seja **A** uma matriz de dimensão m por n e p, q, r, s inteiros satisfazendo os intervalos $1 \leq p \leq q \leq m$ e

$1 \leq r \leq s \leq n$, a notação $\mathbf{A}_{p:q,r:s}$ define a submatriz representada pela Equação IV.9 [Boyd and Vandenberghe, 2018].

$$\mathbf{A}_{p:q,r:s} = \begin{bmatrix} a_{pr} & a_{p\ r+1} & \cdots & a_{ps} \\ a_{p+1\ r} & a_{p+1\ r+1} & \cdots & a_{p+1\ s} \\ \vdots & \vdots & \vdots & \vdots \\ a_{qr} & a_{q\ r+1} & \cdots & a_{qs} \end{bmatrix} \tag{IV.9}$$

Já na vetorização matricial (ou vetorização de uma matriz), aplicável em análise econométrica e diferenciação de vetores e matrizes, para uma matriz $\mathbf{A}$ de dimensão m por n, a respectiva vetorização, denotada por Vec($\mathbf{A}$), segue a Equação IV.10 [Dadkhah, 2011]. Na prática, as colunas da matriz $\mathbf{A}$ são empilhadas de forma obter uma representação vetorial.

$$\text{Vec}(\mathbf{A}) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \\ \vdots \\ a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} \tag{IV.10}$$

Dadas as definições acima e denotando por $\mathbf{D}$ a matriz gerada pelo pré-processamento do WARMA, de dimensão h por w, a série pré-processada, independentemente da quantidade de passos aplicados no pré-processamento, seguirá a Equação IV.11, na qual a vetorização matricial é aplicada à submatriz $\mathbf{D}_{1:h,((w-\tau)+1):w}$, em que τ corresponde ao deslocamento entre subsequências adjacentes.

$$\text{Vec}(\mathbf{A}_{1:m,((n-\tau)+1):n}) \tag{IV.11}$$

IV.2 Modelo WARMA

Nesta seção, incorporaremos o pré-processamento, apresentado na seção anterior, ao ARMA, resultando no modelo WARMA. Para facilitar a compreensão, primeiro resumiremos as equações necessárias para obter a série estacionária, considerando a quantidade de passos s. Estas equações são apresentadas em IV.13a, IV.13b e IV.13c, em que $\mu^{(\vec{a}^{(\mathrm{g}(t))})}$ e $\sigma^{(\vec{a}^{(\mathrm{g}(t))})}$ representam, respectivamente, a média e o desvio-padrão do i-ésimo vetor-linha da matriz $\mathbf{A}$, no qual o i-ésimo índice é

retornado pela função $\mathtt{g}(t)$ (Equação IV.12), que recebe como argumento o instante t de uma dada observação x_t ($o < t < |\mathrm{X}|$). Já a matriz $\mathbf{A}$ e o parâmetro o são definidos na fase SW.

$$\mathtt{g}(t) = \lfloor (t - o)/w \rfloor + 1 \tag{IV.12}$$

Na Equação IV.13a, a série já é estacionária e, portanto, não requer nenhum ou zero passo. Na Equação IV.13b, a série apresenta média não constante, requerendo um passo para obter a estacionariedade. E por último, na Equação IV.13c, a série apresenta média e variância não constantes, requerendo dois passos para obter a estacionariedade.

$$\mathrm{y}_t = \mathrm{x}_t, s = 0 \tag{IV.13a}$$

$$\mathrm{y}_t = \left(\mathrm{x}_t - \mu^{(\vec{a}^{(\mathtt{g}(t))})}\right), s = 1, \forall \vec{a}^{(\mathtt{g}(t))} \in \mathbf{A} \tag{IV.13b}$$

$$\mathrm{y}_t = \left(\frac{\mathrm{x}_t - \mu^{(\vec{a}^{(\mathtt{g}(t))})}}{\sigma^{(\vec{a}^{(\mathtt{g}(t))})}}\right), s = 2, \forall \vec{a}^{(\mathtt{g}(t))} \in \mathbf{A} \tag{IV.13c}$$

Para ser possível obter um modelo que combine o pré-processamento e o modelo ARMA, é necessário unificar as equações IV.13a, IV.13b e IV.13c em uma única equação, de forma que possa ser incorporada ao modelo ARMA. Para alcançar esse objetivo, introduzimos funções de controle, denotadas por $f_\mu(s)$ e $f_\sigma(s)$, cujas finalidades são, respectivamente, determinar a presença ou ausência dos parâmetros $\mu^{(\vec{a}^{(\mathtt{g}(t))})}$ e $\sigma^{(\vec{a}^{(\mathtt{g}(t))})}$, em que o argumento s das funções $f_\mu(s)$ e $f_\sigma(s)$ corresponde à quantidade de passos (Equação IV.14).

$$\mathrm{y}_t = \left(\frac{\mathrm{x}_t - f_\mu(s)\mu^{(\vec{a}^{(\mathtt{g}(t))})}}{f_\sigma(s)\sigma^{(\mu^{(\vec{a}^{(\mathtt{g}(t))})})}}\right), s = 0, 1, 2, \forall \vec{a}^{(\mathtt{g}(t))} \in \mathbf{A} \tag{IV.14}$$

Para a função de controle $f_\mu(s)$, a solução fornecida pela Equação IV.15 satisfaz as equações IV.13a, IV.13b e IV.13c. Já para a função de controle $f_\sigma(s)$, a Equação IV.16 satisfaz somente a Equação IV.13c, pois para $s \neq 2$, a Equação IV.14 resulta em divisão por zero. Uma solução de contorno para evitar a divisão por zero seria somar ao denominador da Equação IV.14 uma constante igual a um.

Todavia, essa solução de contorno não satisfaz a Equação IV.13c, por não somar uma constante ao denominador. Devido a isso, a Equação IV.14 é refatorada de modo que a função $f_\sigma(s)$ passe a

ser o expoente do parâmetro $\sigma^{(\vec{a}^{(\mathrm{g}(t))})}$, resultando na Equação IV.17. Assim, a Equação IV.16 passa a satisfazer as equações IV.13a, IV.13b e IV.13c.

$$f_{\mu}(s) = \begin{cases} 0, s = 0 \\ 1, s \neq 0 \end{cases} \tag{IV.15}$$

$$f_{\sigma}(s) = \begin{cases} 0, s \neq 2 \\ 1, s = 2 \end{cases} \tag{IV.16}$$

$$\mathrm{y}_t = \left(\frac{\mathrm{x}_t - f_{\mu}(s)\mu^{(\vec{a}^{(\mathrm{g}(t))})}}{(\sigma^{(\vec{a}^{(\mathrm{g}(t))})})^{f_{\sigma}(s)}} \right), s = 0, 1, 2, \forall \vec{a}^{(\mathrm{g}(t))} \in \mathbf{A} \tag{IV.17}$$

Aplicando as equações IV.15 e IV.16 ao modelo ARMA (Equação II.25), obtemos o modelo final do WARMA (Equação IV.18), que combina o pré-processamento para obter a estacionariedade, descrito na seção IV.1 deste capítulo, e o modelo ARMA.

$$\phi(B) \left(\frac{\mathrm{x}_t - f_{\mu}(s)\mu^{(\vec{a}^{(\mathrm{g}(t))})}}{(\sigma^{(\vec{a}^{(\mathrm{g}(t))})})^{f_{\sigma}(s)}} \right) = \theta(B)\varepsilon_t, s = 0, 1, 2, \forall \vec{a}^{(\mathrm{g}(t))} \in \mathbf{A} \tag{IV.18}$$

IV.2.1 Notação WARMA

Dado que a notação para o modelo ARIMA segue o formato ARIMA(p, d, q), o usuário pode pressupor que a notação para o modelo WARMA seja WARMA(p, s, q), sendo s a quantidade de passos. Embora válida, deve-se lembrar que o WARMA usa duas propriedades: w e τ, que correspondem ao tamanho das subsequências e ao deslocamento entre subsequências adjacentes, respectivamente.

Caso a notação WARMA(p, s, q) fosse usada, perguntas como "Qual o tamanho das subsequências usadas em um dado modelo WARMA ?" ou "As subsequências são sobrepostas, ou não?" ficariam sem respostas. Por essa razão, a notação para o modelo WARMA deve incluir as propriedades w e τ, resultando na notação WARMA(p, (w, s, τ), q). Não obstante, se o usuário desejar informação somente sobre o modelo ARMA, é suficiente remover o (w, s, τ) da notação WARMA, de forma que o resultado seja ARMA(p, q).

Todavia, para o cenário no qual o usuário deseje informar somente a quantidade de passos resultante do pré-processamento, é admissível a forma S(s), em que s determina a quantidade de passos. Porém, essa forma compacta deve ser aplicada de modo que não provoque interpretações dúbias e que fique clara para o público-alvo. No geral, deve-se adotar a forma completa S(w, s, τ).

Capítulo V Avaliação experimental

V.1 Metodologia

Simultaneamente à definição do modelo ARIMA, Box et al. [2015] definiram uma metodologia para a configuração do modelo ARIMA, denominada Box-Jenkins, composta por três fases processadas sequencialmente: identificação, estimação e diagnóstico. Para o nosso propósito, a metodologia Box-Jenkins é suficiente, visto que, assim como o ARIMA, o WARMA combina pré-processamento e modelagem, embora alguns autores proponham metodologias mais refinadas [Hewamalage et al., 2022; Salles et al., 2019].

V.1.1 Identificação

Box et al. [2015] propuseram um método para definir a ordem da Integração e a ordem do modelo ARMA, fundamentado no comportamento das funções ACF e PACF. Ainda que sirva para fins didáticos, devido à metodologia Box-Jenkins, a identificação do modelo ARMA por meio desse método é subjetiva, às vezes fornecendo diferentes modelos para uma mesma série integrada, dada a ausência de algoritmos à época para definir um único modelo [De Gooijer and Hyndman, 2006]. Todavia, a definição da ordem da Integração e da ordem do modelo ARMA são campos em contínua pesquisa [Box et al., 2015; Herranz, 2017; Tang and Röllin, 2021].

No caso do WARMA, a fim de definir a quantidade de passos, usaremos a função *ndiffs*, incluída no pacote *forecast* [Hyndman and Khandakar, 2008], implementado na linguagem R, cuja finalidade é determinar a ordem da Integração de uma série. Basicamente, a cada passo aplicado pelo WARMA, invoca-se a função *ndiffs*, passando como argumento a respectiva série pré-processada. Uma vez que a ordem de integração for zero, finaliza-se o pré-processamento do WARMA e passa-se à fase de modelagem.

Já para a definição da ordem do modelo ARMA, o consenso em geral entre autores é usar algum critério da informação como, por exemplo, AIC e BIC. Tais critérios da informação consideram fatores como o ajuste do modelo à série e a quantidade de parâmetros do modelo, em geral, diferenciando-se entre si no que diz respeito à penalidade imposta para a quantidade de parâmetros do modelo [Hyndman and Athanasopoulos, 2018].

V.1.2 Estimação

A base para a estimação do modelo ARMA foi demonstrada no Capítulo II, subseção II.3.4, na qual os erros são normalmente distribuídos com média μ e variância σ^2 (Equação II.26). A fim de calcular os parâmetros do modelo ARMA, são aplicados métodos como MLE, MoM e LSE [Shan et al., 2014]. Entretanto, visando automatizar o processo de estimação, usa-se a função *auto.arima*, disponível no pacote *forecast* [Hyndman and Khandakar, 2008], implementado na linguagem R.

Por padrão, a função *auto.arima* usa o critério da informação AICc, definido na Equação V.1, em que T representa a quantidade de observações usadas na modelagem, L denota a função MLE do modelo ARMA sob análise, refletindo a probabilidade associada à obtenção dessas observações, e k um parâmetro binário que assume valor 1 quando houver a presença de uma constante ($c \neq 0$) ou 0, caso contrário [Hyndman and Athanasopoulos, 2018].

Idealmente, quanto menor o valor da função AICc, melhor, de forma que o modelo se ajusta cada vez mais à série à medida que o valor da função se aproximar de $-\infty$ [Enders, 2014]. Adicionalmente, a função *auto.arima* usa internamente um algoritmo que otimiza a definição da ordem do modelo, além de aplicar o método MLE para a configuração dos respectivos coeficientes, selecionando como modelo resultante aquele que apresentar o menor valor para o AICc [Hyndman and Athanasopoulos, 2018].

$$\text{AICc} = \text{AIC} + \frac{2(p+q+k+1)(p+q+k+2)}{T-p-q-k-2} \tag{V.1}$$

$$\text{AIC} = -2\log(L) + 2(p+q+k+1) \tag{V.2}$$

O Algoritmo 1 implementa um processo de otimização automática para seleção do modelo WARMA, recebendo dois parâmetros como entrada: uma série X e o tamanho máximo das subsequências, *window.length*. Inicialmente, são estabelecidos valores iniciais para o critério AICc (∞) e para o modelo WARMA, definido como nulo.

O algoritmo opera por meio de um processo iterativo aninhado, no qual a iteração externa varia o tamanho das subsequências de três até o valor máximo especificado (*window.length*), enquanto a iteração interna varia o deslocamento da subsequência de um até o tamanho corrente da subsequência. A cada iteração, a série é pré-processada segundo a lógica do pré-processamento do WARMA (Figura IV.4) por meio da função *warma.preprocessing*. Posteriormente, um modelo WARMA é automaticamente ajustado aos dados processados.

A seleção do melhor modelo é baseada no critério da informação AICc. Se o modelo corrente

apresentar um AICc menor que o previamente armazenado, ele é configurado como o melhor modelo encontrado até o momento. Ao final do processo iterativo, o algoritmo retorna o modelo ARMA que apresentou o menor AICc dentre as combinações avaliadas durante o processo iterativo.

Por fim, a razão para usarmos a função *auto.arima* no Algoritmo 1 é garantir uma comparação imparcial entre os modelos ARIMA e WARMA, haja vista que a função *auto.arima* é altamente customizável. Dado que a série pré-processada pelo WARMA é estacionária, configuramos o parâmetro d da função *auto.arima* com o valor zero, prevenindo assim possíveis problemas como, por exemplo, a sobrediferenciação. Pela mesma razão, optamos por não aplicar a fase de processamento de *outliers*, presente no WARMA. Na prática, essa decisão metodológica fará com que os *outliers* presentes na série observada se destaquem visualmente quando comparados às respectivas observações na série modelada, facilitando sua detecção.

Algorithm 1: Algoritmo de seleção do modelo WARMA

Input:
Série temporal X
tamanho máximo da subsequência $window.length$
Output: Modelo ARMA otimizado

$AICc.model \longleftarrow \infty$;
$ARMA.model \longleftarrow$ NULL;
for $w \in \{3, \ldots, window.length\}$, $\tau \in \{1, \ldots, w\}$ **do**
    $\mathrm{Y} \longleftarrow warma.preprocessing(\mathrm{X}, w, \tau)$;
    $obj \longleftarrow auto.arima(\mathrm{Y})$;
    if $AICc(obj) < AICc.model$ **then**
        $AICc.model \longleftarrow AICc(obj)$;
        $ARMA.model \longleftarrow ARMA(obj)$;
return $ARMA.model$

V.1.3 Diagnóstico

Na prática, o gráfico da função ACF dos resíduos deve assemelhar-se ao de um ruído branco (Figura II.3). Caso contrário, pode indicar a presença de heterocedasticidade, sazonalidade, entre outros, sendo necessário nesses casos reavaliar a série a fim de sanar as causas que provocaram a não estacionariedade. Apesar de ser útil para analisar o comportamento da série residual, o gráfico da função ACF é limitado se considerarmos uma quantidade expressiva de séries para visualizá-las uma a uma.

Diante dessa realidade, o método comumente empregado para verificar correlação serial na série residual é o teste de hipótese Ljung-Box-Pierce [Ord et al., 2017], no qual não há correlação serial para qualquer $l = 1, 2, 3, \ldots$ na hipótese nula, enquanto há correlação serial para pelo menos um $l = 1, 2, 3, \ldots$ na hipótese alternativa, conforme Equações V.3a e V.3b, respectivamente, em que ρ corresponde à função ACF (Equação II.8).

$$\mathrm{H}_0 = \rho_l = 0, l = 1, 2, 3, \ldots \tag{V.3a}$$

$$\mathrm{H}_1 = \exists l \in \{1, 2, 3, \ldots\} : \rho_l \neq 0 \tag{V.3b}$$

V.2 Conjunto de dados

O conjunto de séries usadas nos experimentos é de natureza econômico-financeira com frequência mensal, dividido por domínios, a fim de facilitar a análise. Na Tabela V.1, descrevemos esses domínios e as respectivas quantidades de séries, bem como a(s) fonte(s).

Tabela V.1: Conjunto de séries econômico-financeiras

Domínio	**Quantidade de séries**	**Fonte**
Mercado de ações	11	B3[1] e Investing[2]
Câmbio monetário	17	Investing[3]
Commodities	21	Investing[4]
Atividade econômica	7	BCB
DLSP	9	BCB
IPCA	20	IBGE
Taxa de juros	7	BCB e IPEAData[5]

V.3 Particionamento dos dados

Para a definição do horizonte de previsão, usaremos como base a competição M4, considerada o padrão-ouro entre as competições de previsões de modelos de séries, cujos horizontes de previsão são determinados considerando a probabilidade de se aplicar um dado horizonte para uma dada frequência em uma empresa ou organização [Makridakis et al., 2020]. No caso de séries com frequência mensal, o horizonte de previsão adotado na competição M4 é 14, sendo este, portanto, o valor a ser adotado na avaliação experimental.

Assim, para uma série $\mathrm{X} = \langle \mathrm{x}_1, \ldots, \mathrm{x}_n \rangle$ de tamanho n e denotando por h o horizonte de previsão do modelo preditivo em questão, definimos o conjunto de treino e teste, respectivamente, em $\langle \mathrm{x}_1, \ldots, \mathrm{x}_T \rangle$ e $\langle \mathrm{x}_{T+1}, \ldots, \mathrm{x}_{T+h} \rangle$, em que $T = n - h$ determina o tamanho da série usado no conjunto de treino, enquanto h o tamanho da série usado no conjunto de teste.

[1]Disponível em: https://b3.com.br/
[2]Disponível em: https://www.investing.com/
[3]Disponível em: https://www.investing.com/
[4]Disponível em: https://www.investing.com/
[5]Disponível em: http://www.ipeadata.gov.br/

V.4 Avaliação experimental I

Teoricamente, conforme visto no Capítulo II, a ordem de Integração de uma série integrada pertence ao conjunto dos números inteiros não negativos (Equação II.33). Porém, séries em geral possuem ordem de Integração entre zero e dois, inclusive [Box et al., 2015]. Posto isso, consideraremos nesta seção somente séries nas quais as respectivas ordens de Integração sejam iguais a um ou dois. Caso uma ou mais séries apresentem ordem de Integração superior a dois, se houver, usaremos a notação I(∗).

Na caso do WARMA, a quantidade de passos de uma série varia entre zero e dois, inclusive. Caso, mesmo após aplicar dois passos e a não estacionariedade persistir, devido à covariância não constante, usaremos a notação S(∗). Outro ponto é que, independente da Integração ou do WARMA, só serão consideradas séries não estacionárias. Todavia, para fins de registro, caso haja séries estacionárias, forneceremos uma visão geral sob a forma de gráfico de barras, informando a quantidade de séries estacionárias e os respectivos domínios aos quais pertencem.

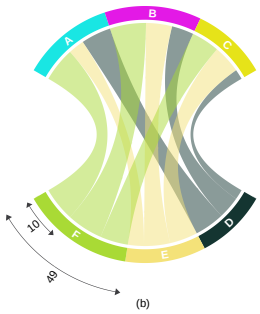
V.4.1 Diagrama de cordas

Um diagrama de cordas é uma ferramenta visual na qual entidades são dispostas radialmente como cordas geométricas e suas respectivas relações (ou relacionamentos) conectadas por arcos, na qual a largura de cada arco indica a significância da relação [Tintarev et al., 2018]. Dependendo da quantidade de entidades relacionadas, diferentes cores podem ser atribuídas aos arcos, de modo facilitar a visualização do relacionamento entre as entidades.

Na Figura V.1 (a), Tabela com valores randomicamente gerados e a respectiva representação gráfica por meio do diagrama de cordas (Figura V.1 (b)). As colunas A, B e C e as linhas D, E e F representam, para fins de ilustração, entidades genéricas, em que cada letra (ou entidade) é associada a um arco no diagrama de cordas, sendo o tamanho de cada arco correspondente à soma dos valores de cada linha ou coluna.

	A	B	C
D	13	11	16
E	7	13	15
F	10	16	16

(a)



(b)

Figura V.1: Tabela com valores randomicamente gerados em (a) e a respectiva representação gráfica por meio do diagrama de cordas em (b)

V.4.2 Análise dos resultados

Do conjunto de séries avaliadas, sob a ótica do WARMA e considerando somente as séries não estacionárias, observou-se que 100% das séries alcançaram a estacionariedade com no máximo dois passos. A maioria (82,5%) obteve a estacionariedade com somente um passo, enquanto 17,5% necessitaram de dois passos. É notável que as séries que demandaram dois passos apresentaram ordem de Integração igual a um (Figura V.2).

Quanto à ordem de Integração e, assim como o ainda WARMA, considerando somente séries não estacionárias, identificamos uma predominância significativa de séries com ordem de Integração igual a um (93,75%). As séries restantes (6,25%) apresentaram ordem de Integração igual a dois, sendo que estas exigiram dois passos para alcançar a estacionariedade (Figura V.2).

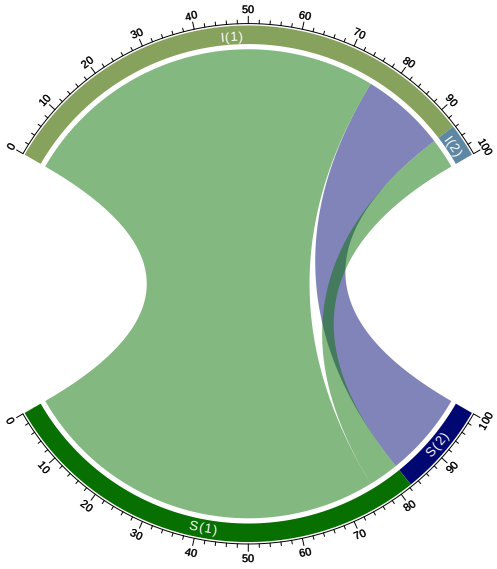


Figura V.2: Diagrama de cordas relacionando o percentual de séries conforme a ordem de Integração e a quantidade de passos do WARMA

Por ora, o diagrama de cordas apresentado na Figura V.2 revela um padrão: quando a ordem de Integração ou a quantidade de passos é igual a dois, existe necessariamente uma relação com um elemento de valor um, seja na ordem de Integração ou na quantidade de passos. No entanto, para uma análise mais abrangente e conclusiva, é necessário examinar como essa distribuição se comporta em cada domínio, considerando a ordem de Integração e a quantidade de passos.

No gráfico de barras da Figura V.3, agrupamos as séries por domínio segundo a quantidade de passos. Nesse gráfico, somente séries dos domínios commodities, atividade econômica e mercado de ações demandaram dois passos. Por outro lado, no gráfico de barras da Figura V.4, no qual agrupamos as séries por domínio segundo a ordem de Integração, somente o domínio DLSP requisitou ordem de Integração igual a dois.

Dado que as séries que apresentam dois passos ou ordem de integração igual a dois, bem como

seus respectivos domínios, constituem um percentual inferior em comparação às demais séries e domínios, podemos dizer, por ora, que a relação bidirecional observada no diagrama de cordas da Figura V.2 não justifica, no escopo desta dissertação, uma análise mais aprofundada.

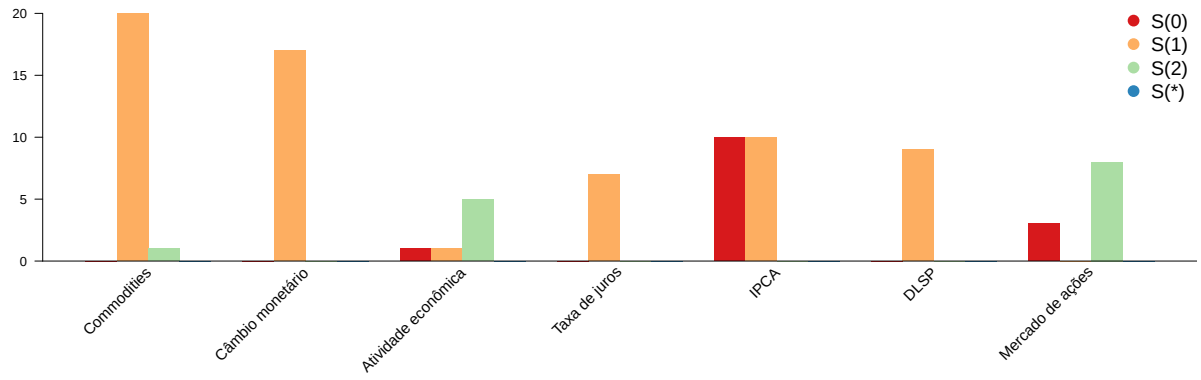


Figura V.3: Quantidade de séries por domínio segundo a quantidade de passos

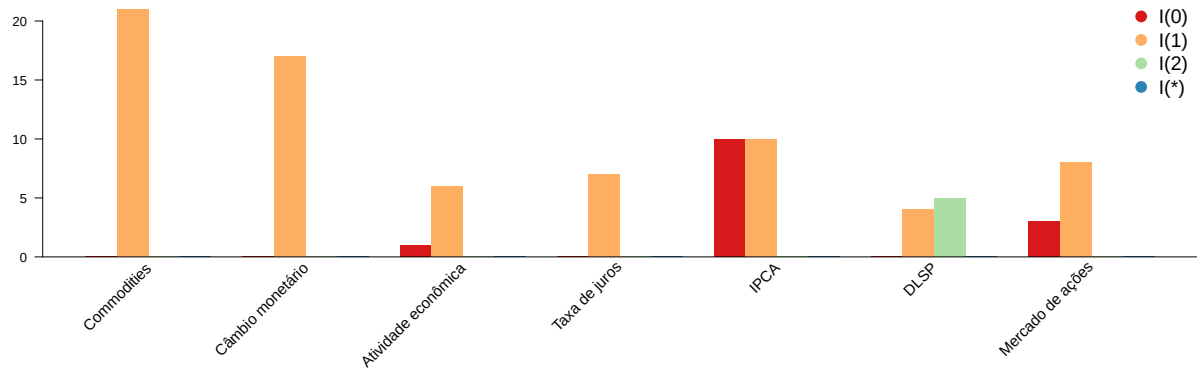


Figura V.4: Quantidade de séries por domínio segundo a ordem de integração

V.4.3 Análise de hiperparâmetros

Conforme vimos na Subseção V.1.2, a função *auto.arima*, por padrão, usa um subconjunto de valores para os parâmetros p e q de forma obter o modelo ARIMA. No WARMA, complementando esses parâmetros, necessitamos ainda definir os valores dos tamanhos das subsequências (w) e dos deslocamentos entre subsequências adjacentes (τ) que satisfaçam o critério da informação AICc.

Idealmente, é desejável que o tamanho das subsequências seja o maior possível, de forma que a maior subsequência possível corresponda ao tamanho da própria série, conservando assim a média, variância e covariância constantes. Todavia, para séries não estacionárias, à medida que o tamanho das subsequências aumenta, a probabilidade de uma ou mais das propriedades média, variância e covariância não serem constantes aumenta da mesma forma.

Ao analisar o gráfico de bolhas da Figura V.5, em que cada ponto (ou bolha) é proporcional à frequência, relacionamos os modelos WARMA selecionados pela função *auto.arima* segundo o tamanho das subsequências (w) em função do deslocamento entre elas (τ), observa-se uma predominância de modelos com valor de w igual a três, enquanto os respectivos valores de τ se distribuem entre um, dois e três. Além disso, vemos uma concentração de modelos, embora menor, para valores de w entre doze e quatorze.

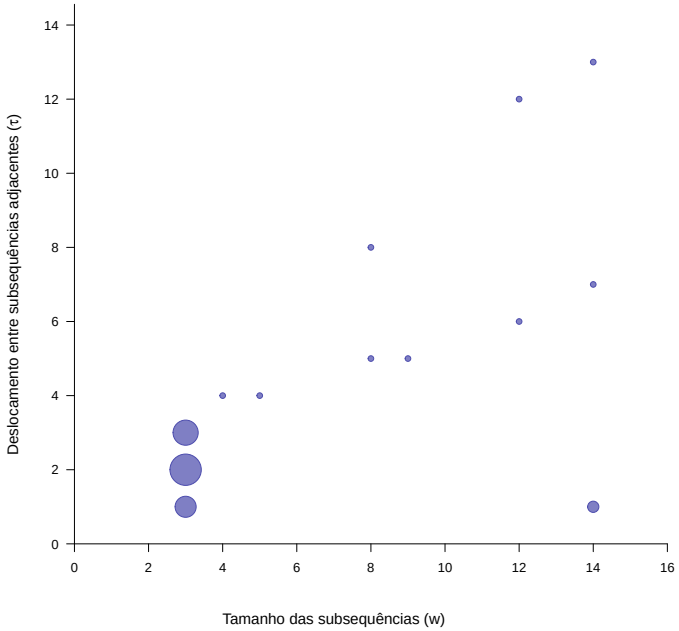


Figura V.5: Modelos WARMA selecionados segundo o tamanho das subsequências (w) em função do deslocamento entre elas (τ)

V.5 Avaliação experimental II

V.5.1 Métricas

Independente do domínio ao qual a série pertence, diferentes séries podem apresentar não só diferentes unidades de medida, mas ainda diferentes ordens de grandeza. À vista disso, é necessário adotar uma métrica que seja agnóstica às unidades de medida e às ordens de grandeza das séries, de maneira que seja possível compará-las segundo a finalidade desejada. Posto isso, avaliaremos as seguintes métricas: MAE, MAPE, sMAPE e MASE.

A métrica MAE fornece a média do erro absoluto para um horizonte de previsão h (Equação V.4), sendo o erro para um dado instante t definido pela Equação V.5, em que x_t e $\hat{\mathrm{x}}_t$ correspondem, respectivamente, ao valor observado e previsto no instante t [Hyndman and Athanasopoulos, 2018]. Apesar de ser dependente da escala, a métrica MAE é usada em diversas métricas agnósticas às grandezas das séries, como veremos mais adiante, daí a menção a essa métrica. Pela mesma razão,

definimos o modelo *naïve* (simples), no qual o valor previsto é igual ao valor anteriormente observado (Equação V.6). Adicionalmente, convenciona-se usar o sinal circunflexo para denotar um previsto em vez de observado, como em $\hat{\mathrm{x}}_t$.

$$\mathrm{MAE} = \frac{1}{\mathrm{h}} \sum_{t=1}^{\mathrm{h}} |e_t| \tag{V.4}$$

$$e_t = \mathrm{x}_t - \hat{\mathrm{x}}_t \tag{V.5}$$

$$\hat{\mathrm{x}}_{t+1} = \mathrm{x}_t \tag{V.6}$$

A métrica MAPE, por sua vez, fornece uma visão percentual do quão os valores previstos desviam dos valores observados (Equação V.7). Por exemplo, um MAPE de 5% significa que, em média, a previsão de um dado modelo sobrestima ou subestima os valores observados em 5% [Kolassa and Schütz, 2007]. Devido à sua popularidade na literatura e interpretabilidade, a métrica MAPE é o método preferido de analistas e profissionais de negócios [Kim and Kim, 2016].

Um problema com a métrica MAPE é a sensibilidade à simetria. Objetivamente, uma métrica é simétrica se dois valores equidistantes em relação a zero, como +50 e -50, por exemplo, produzirem resultados igualmente simétricos. Isso deve ser considerado quando o problema exigir simetria entre valores positivos e negativos, como, por exemplo, orçamento, no qual a métrica MAPE pode fornecer diferentes resultados para receitas e despesas [Makridakis, 1993].

A métrica MAPE possui ainda a desvantagem de gerar valores infinitos ou indefinidos se o valores reais da série forem iguais ou próximos a zero, cenário comum em áreas como varejo, biologia e finanças, entre outros. Além disso, se os valores reais da série forem muito pequenos, usualmente menores que um, a métrica MAPE produz erros percentuais extremamente grandes (*outliers*) [Kim and Kim, 2016].

$$\mathrm{MAPE} = \frac{1}{\mathrm{h}} \left(\sum_{t=1}^{\mathrm{h}} \left| \frac{e_t}{\mathrm{x}_t} \right| \right) \times 100\% \tag{V.7}$$

A fim de lidar com as limitações da métrica MAPE, Makridakis and Hibon [2000] propôs a métrica sMAPE, em que o erro absoluto é dividido pela média da soma dos valores absolutos do valor observado e previsto (Equação V.8). Assim como a métrica MAPE, é uma métrica percentual, porém com valores variando de -200% a +200% [Makridakis and Hibon, 2000].

Ainda que vise às limitações do MAPE, o sMAPE está sujeito à divisão por zero quando os

valores reais e previstos forem iguais a zero [Chen and Yang, 2004]. E apesar de incluir a palavra simétrico na denominação, não penaliza igualmente os erros positivos e negativos, particularmente quando o erro $e_t = \mathrm{x}_t - \hat{\mathrm{x}}_t$ for ou muito pequeno, ou muito grande [Hyndman and Koehler, 2006].

De qualquer forma, a métrica sMAPE foi usada como métrica na competição M4. Porém, para contornar as limitações desta métrica, a competição M4 define alguns requisitos como, por exemplo, somar uma constante às observações de cada série de modo que o valor mínimo de cada série pós-processada seja 10 [Makridakis et al., 2020].

$$\mathrm{sMAPE} = \frac{1}{\mathrm{h}} \left(\sum_{t=1}^{\mathrm{h}} \frac{|e_t|}{\frac{(|\mathrm{x}_t| + |\hat{\mathrm{x}}_t|)}{2}} \right) \times 100\% = \frac{1}{\mathrm{h}} \left(\sum_{t=1}^{\mathrm{h}} \frac{2|e_t|}{(|\mathrm{x}_t| + |\hat{\mathrm{x}}_t|)} \right) \times 100\% \tag{V.8}$$

Frente às limitações das métricas MAPE e sMAPE, Hyndman and Koehler [2006] definiram a métrica MASE, que dimensiona (ou escala, daí o nome da métrica) cada erro absoluto do modelo preditivo usando um fator de escala (Equação V.9), sendo este igual ao MAE do modelo naïve do conjunto de treino (Equação V.10). Uma vez dimensionado cada erro absoluto do método preditivo, calcula-se a MASE a partir da média definida pela Equação V.11.

$$\mathrm{SE}_i = \frac{|e_i|}{\mathrm{naïve}_{\mathrm{TREINO}}}, i = 1, \ldots, h \tag{V.9}$$

$$\mathrm{naïve}_{\mathrm{TREINO}} = \frac{1}{T-1} \sum_{j=2}^{T} |\mathrm{x}_j - \mathrm{x}_{j-1}| \tag{V.10}$$

$$\mathrm{MASE} = \frac{1}{h} \sum_{i=1}^{\mathrm{h}} |\mathrm{SE}_i| \tag{V.11}$$

Diferentemente das métricas MAPE e sMAPE, a MASE é uma métrica relativa, em que valores maiores que um indicam que a eficácia do MAE da previsão foi pior que o MAE do modelo naïve do conjunto de treino, enquanto valores menores que um indicam que a eficácia do MAE da previsão foi melhor que o MAE do modelo naïve do conjunto de treino. Adicionalmente, devido ao fato do numerador e o denominador da Equação V.11 usarem a mesma escala da série original, a métrica MASE é agnóstica às grandezas das séries [Hyndman and Athanasopoulos, 2018]. Por fim, a MASE e a sMAPE formam as métricas usadas na competição M4 [Makridakis et al., 2020].

V.5.2 Análise por domínio

Mercado de ações

O benchmarking entre os métodos ARIMA e WARMA aplicado ao domínio Mercado de ações revelou diferenças importantes no desempenho dos dois métodos ao longo dos horizontes analisados. De forma geral, o WARMA apresentou média da métrica MASE menor do que o ARIMA no horizonte mais curto, i.e., 1-3, no qual a média MASE do WARMA foi 0,66 menor (diferença percentual de 20,24%) (Tabela V.2). Para o horizonte de 1-6, o diagrama de caixa revela uma distribuição estatisticamente semelhante entre ambos os métodos, de forma que podem ser aplicados intercambiavelmente (Figura V.6).

A partir do horizonte de previsão 1-9, o método ARIMA apresenta média inferior ao WARMA. Porém, observa-se que as máximas do modelo WARMA para os horizontes de 1-9 a 1-14 aumentam consideravelmente se comparadas às máximas do ARIMA, o que pode justificar a diferença mais significativa entre as médias, quando comparadas com os horizontes de 1-3 a 1-6 (Tabela V.2).

Por fim, considerando os resultados apresentados, a escolha entre ARIMA e WARMA para o domínio Mercado de ações depende do horizonte de previsão, sendo o WARMA mais adequado para horizontes de 1-3 a 1-6, ao passo que o ARIMA é mais adequado para horizontes de 1-9 a 1-14.

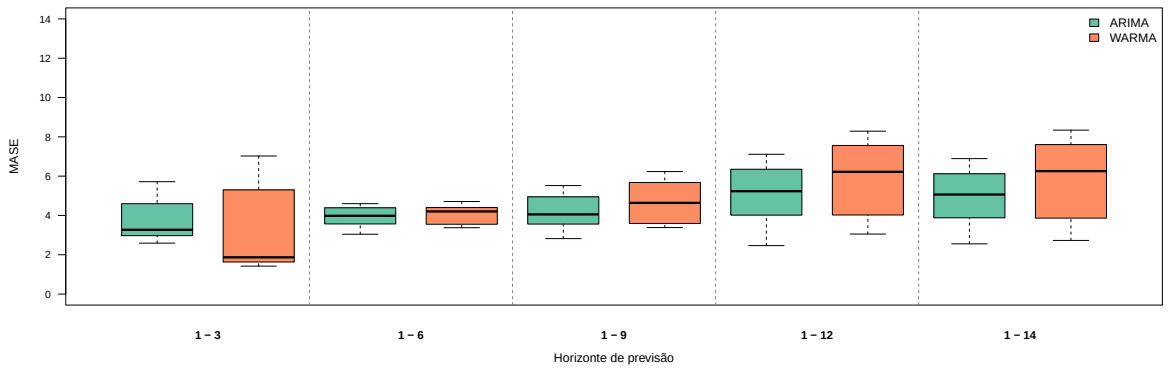


Figura V.6: Diagrama de caixa aplicada ao domínio Mercado de ações para diferentes horizontes dos modelos ARIMA e WARMA

Tabela V.2: Sumário estatístico da métrica MASE aplicada ao domínio Mercado de ações para diferentes horizontes dos modelos ARIMA e WARMA

	1 - 3		1 - 6		1 - 9		1 - 12		1 - 14	
	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA
Mínima	2.59	1.42	3.04	3.38	2.82	3.39	2.47	3.05	2.56	2.73
1º quartil	2.98	1.7	3.62	3.56	3.57	3.66	4.45	4.45	4.32	4.29
Mediana	3.27	1.87	3.98	4.2	4.05	4.64	5.23	6.22	5.06	6.26
Média	3.92	3.26	3.94	4.05	4.18	4.68	5.1	5.87	4.95	5.82
3º quartil	4.04	4.68	4.39	4.39	4.92	5.65	6.31	7.53	6.08	7.58
Máxima	7.1	7.03	4.6	4.71	5.52	6.23	7.11	8.29	6.89	8.34
Desvio-padrão	1.6	2.34	0.55	0.49	0.9	1.1	1.61	2	1.52	2.13

Câmbio monetário

Para o domínio Câmbio monetário, os resultados da Tabela V.3 mostram que, para qualquer horizonte, o método ARIMA apresenta valores MASE ligeiramente inferiores ao WARMA. A diferença é mais significativa no horizonte de 1-9, embora a diferença percentual neste horizonte seja 5,9%. Além disso, não há um padrão crescente da média MASE com o aumento do horizonte de previsão, como frequentemente é observado em análises de séries, sugerindo que ambos os métodos mantêm sua capacidade preditiva de forma relativamente estável, independentemente do horizonte considerado.

Outro ponto da Tabela V.3 revela que ambos os métodos apresentam comportamentos similares em relação ao desvio-padrão, indicando níveis comparáveis de consistência. Por exemplo, no horizonte 1-3, o ARIMA apresenta um desvio-padrão de 1,16 frente ao 1,03 do WARMA, sugerindo que ambos os métodos mantêm valores próximos de variabilidade ao longo dos diferentes horizontes. Porém, considerando somente as mínimas e as máximas, seja para o ARIMA ou WARMA, observa-se que as mínimas aumentam ao longo dos horizontes, enquanto que as máximas diminuem ao longo dos horizontes.

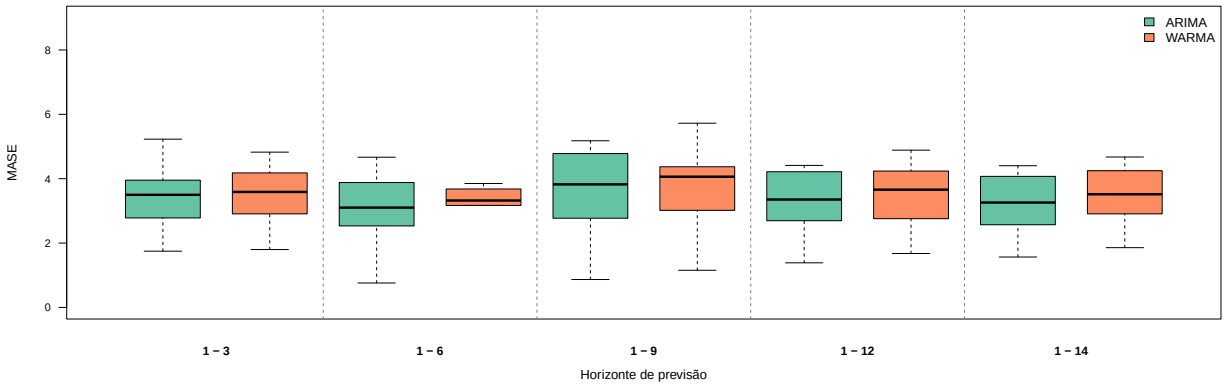


Figura V.7: Diagrama de caixa aplicada ao domínio Câmbio monetário para diferentes horizontes dos modelos ARIMA e WARMA

Tabela V.3: Sumário estatístico da métrica MASE aplicada ao domínio Câmbio monetário para diferentes horizontes dos modelos ARIMA e WARMA

	1 - 3		1 - 6		1 - 9		1 - 12		1 - 14	
	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA
Mínima	0.67	0.96	0.76	1.05	0.87	1.15	1.39	1.68	1.57	1.86
1º quartil	2.78	2.91	2.53	3.17	2.77	3.02	2.69	2.76	2.57	2.91
Mediana	3.5	3.59	3.1	3.32	3.82	4.06	3.35	3.66	3.26	3.52
Média	3.37	3.45	3.06	3.19	3.55	3.76	3.35	3.5	3.26	3.39
3º quartil	3.95	4.18	3.88	3.68	4.78	4.37	4.22	4.24	4.07	4.25
Máxima	5.23	4.83	4.67	4.63	5.18	5.72	4.41	4.89	4.4	4.67
Desvio-padrão	1.16	1.03	1.05	1.03	1.35	1.38	0.97	0.99	0.86	0.93

Commodities

Na comparação entre os métodos ARIMA e WARMA no domínio Commodities, o ARIMA mostrou-se superior nos horizontes avaliados, sendo a diferença mais notável no horizonte mais curto (1-3), onde o ARIMA apresenta um MASE médio de 1,95 comparado a 2,39 do WARMA, representando uma diferença percentual de 22,56%. Todavia, à medida que o horizonte de previsão aumenta, ocorre uma redução na diferença percentual, variando de 14,28% no horizonte 1-6 para 5,03% no horizonte 1-14 (Tabela V.4).

A partir do horizonte 1-6, as mínimas e máximas MASE de ambos os métodos são similares. A maior diferença percentual nos valores máximos é de aproximadamente 5% enquanto nos valores mínimos é menor que 10%. Essa similaridade sugere que os métodos possuem comportamentos parecidos, considerando seus melhores e piores cenários de previsão para horizontes mais longos.

Em relação ao comportamento ao longo dos horizontes, observa-se um padrão crescente da média MASE em ambos os métodos à medida que o horizonte de previsão aumenta. No caso do ARIMA, o MASE médio aumenta de 1,95 para 6,55, enquanto o WARMA apresenta um crescimento de 2,39 para 6,88 para os horizontes 1-3 e 1-14, respectivamente (Tabela V.4).

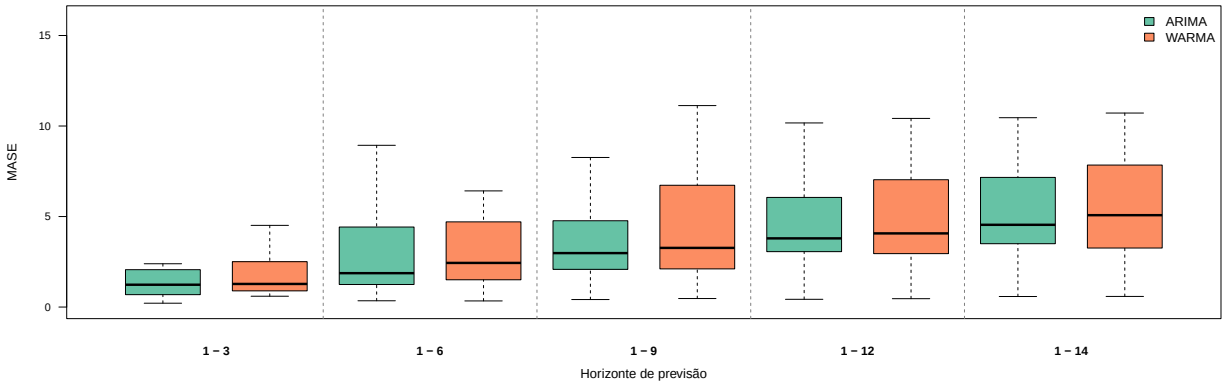


Figura V.8: Diagrama de caixa aplicada ao domínio Commodities para diferentes horizontes dos modelos ARIMA e WARMA

Tabela V.4: Sumário estatístico da métrica MASE aplicada ao domínio Commodities para diferentes horizontes dos modelos ARIMA e WARMA

	1 - 3		1 - 6		1 - 9		1 - 12		1 - 14	
	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA
Mínima	0.21	0.6	0.35	0.34	0.41	0.47	0.43	0.46	0.58	0.59
1º quartil	0.68	0.89	1.24	1.51	2.08	2.11	3.06	2.95	3.5	3.26
Mediana	1.23	1.28	1.87	2.44	2.98	3.27	3.79	4.07	4.54	5.08
Média	1.95	2.39	3.22	3.68	4.54	4.92	5.86	6.22	6.55	6.88
3º quartil	2.06	2.51	4.42	4.71	4.77	6.73	6.06	7.03	7.16	7.84
Máxima	9.56	9.92	15.57	16.33	20.98	21.99	24.08	25.3	24.93	26.38
Desvio-padrão	2.25	2.51	3.5	3.66	4.55	4.77	5.39	5.65	5.91	6.23

Atividade econômica

No domínio Atividade econômica, observa-se uma inversão de desempenho ao longo dos horizontes de previsão. O método WARMA demonstro desempenho superior nos horizontes mais curtos, com uma variação percentual de 29,00% no horizonte 1-3 (MASE médio de 1,37 contra 1,93 do ARIMA) e 20,58% no horizonte 1-6. No entanto, essa vantagem diminui progressivamente ao longo dos horizontes subsequentes (Tabela V.5).

Observa-se que o WARMA apresenta maior variabilidade nos horizontes mais longos, com desvios-padrão consistentemente maiores que o ARIMA a partir do horizonte 1-6, sugerindo que, embora as médias sejam similares, o ARIMA oferece previsões mais estáveis nesses horizontes. Todavia, verifica-se que as máximas do WARMA a partir do horizonte 1-6 indicam variações superiores às máximas do ARIMA, o que pode explicar essa variação no desvio-padrão.

Diferentemente dos padrões comumente observados em previsões de séries, não há um padrão decrescente do desempenho com o aumento do horizonte. As médias MASE do ARIMA apresentam um leve aumento até o horizonte 1-6, partindo de 1,93 a 1,62, e posteriormente se estabilizando por volta de 1,80 nos horizontes de 1-9 a 1-14. O WARMA, por sua vez, mostra um crescimento gradual no MASE médio, partindo de 1,37 no horizonte mais curto até 1,81 no mais longo.

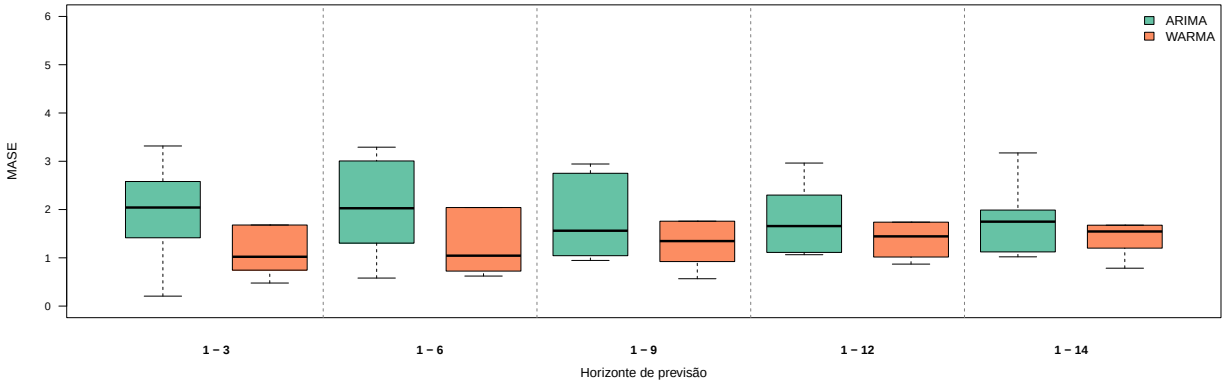


Figura V.9: Diagrama de caixa aplicada ao domínio Atividade econômica para diferentes horizontes dos modelos ARIMA e WARMA

Tabela V.5: Sumário estatístico da métrica MASE aplicada ao domínio Atividade econômica para diferentes horizontes dos modelos ARIMA e WARMA

	1 - 3		1 - 6		1 - 9		1 - 12		1 - 14	
	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA
Mínima	0.21	0.48	0.58	0.62	0.95	0.57	1.06	0.87	1.02	0.78
1º quartil	1.51	0.75	1.46	0.76	1.13	1.02	1.2	1.11	1.26	1.27
Mediana	2.04	1.02	2.03	1.04	1.56	1.35	1.66	1.44	1.75	1.55
Média	1.93	1.37	2.04	1.62	1.8	1.64	1.79	1.77	1.8	1.81
3º quartil	2.51	1.58	2.79	1.84	2.5	1.67	2.19	1.68	1.95	1.66
Máxima	3.32	3.28	3.29	4.22	2.94	3.91	2.96	4.08	3.17	4.1
Desvio-padrão	1.07	1.03	1.02	1.37	0.86	1.18	0.74	1.18	0.78	1.17

DLSP

No domínio DLSP, que retrata a Dívida Líquida do Setor Público, o método WARMA demonstra um desempenho superior consistente ao longo dos horizontes de previsão. A diferença no desempenho aumenta significativamente à medida que o horizonte de previsão aumenta, começando com uma vantagem de 6,66% no horizonte 1-3 e chegando a uma diferença de 23,73% no horizonte 1-14. Esse crescimento sugere que o WARMA é particularmente eficaz em previsões de longo prazo para este conjunto de dados (Tabela V.6).

Observa-se um padrão decrescente do desempenho em ambos os métodos conforme o horizonte de previsão aumenta, embora com variações de deterioração diferentes. O MASE médio do ARIMA cresce de 2,85 para 6,7, enquanto o WARMA apresenta um aumento mais controlado, de 2,66 para 5,11. Embora ambos os métodos mostrem aumento nos desvios-padrão com o horizonte mais longo, o WARMA mantém uma variabilidade ligeiramente maior ao longo dos horizontes, sugerindo que suas previsões, apesar de mais precisas, considerando a média, podem ser menos estáveis.

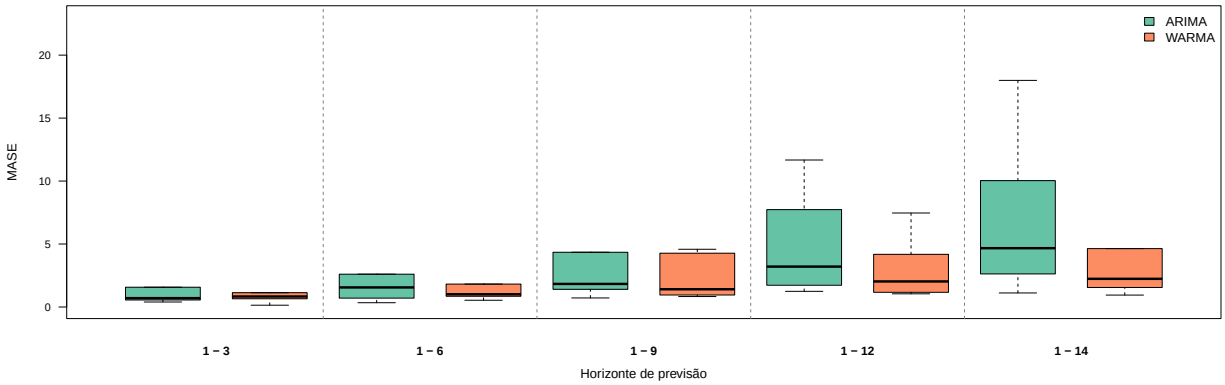


Figura V.10: Diagrama de caixa aplicada ao domínio DLSP para diferentes horizontes dos modelos ARIMA e WARMA

Tabela V.6: Sumário estatístico da métrica MASE aplicada ao domínio DLSP para diferentes horizontes dos modelos ARIMA e WARMA

	1 - 3		1 - 6		1 - 9		1 - 12		1 - 14	
	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA
Mínima	0.4	0.14	0.34	0.54	0.72	0.83	1.24	1.05	1.11	0.95
1º quartil	0.56	0.65	0.7	0.86	1.41	0.95	1.73	1.16	2.63	1.55
Mediana	0.7	0.83	1.55	1.01	1.83	1.41	3.21	2.03	4.67	2.24
Média	2.85	2.66	3.54	3.33	4.39	3.71	5.73	4.57	6.7	5.11
3º quartil	1.57	1.13	2.61	1.82	4.35	4.27	7.73	4.18	10.03	4.64
Máxima	13.62	15.71	15.8	17.86	15.65	17.73	17.42	19.5	17.99	20.08
Desvio-padrão	4.45	4.98	5.05	5.61	4.97	5.45	5.56	5.98	5.8	6.22

IPCA

Para o domínio IPCA, a análise comparativa entre os métodos ARIMA e WARMA, usando a métrica MASE, o ARIMA demonstra um desempenho superior ao longo dos horizontes de previsão. A diferença é mais pronunciada no horizonte curto (1-3), onde o ARIMA apresenta um MASE médio de 1,3 comparado a 1,5 do WARMA, representando uma vantagem de 15,55%. Essa superioridade do ARIMA, embora diminua nos horizontes mais longos, se mantém consistente, estabilizando-se em torno de 7,5% de vantagem nos horizontes 1-9, 1-12 e 1-14 (Tabela V.7).

Os desvios-padrão são ligeiramente menores para o ARIMA ao longo dos horizontes, indicando que além de mais preciso, o método oferece ainda maior consistência nas previsões. É interessante notar que ambos os métodos mostram valores mínimos e máximos similares, sugerindo que apresentam capacidades comparáveis em seus melhores e piores casos.

Um aspecto notável é que não se observa um padrão de deterioração do desempenho da média com o aumento do horizonte. O MASE médio do ARIMA varia entre 1,3 e 1,61, enquanto o WARMA varia entre 1,5 e 1,73. Isso sugere que ambos os métodos mantêm sua capacidade preditiva relativamente estável ao longo dos diferentes horizontes de previsão.

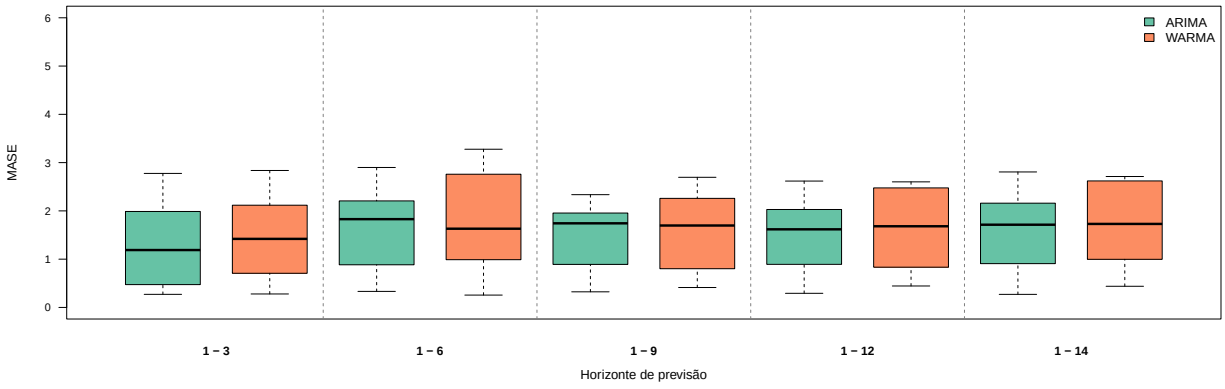


Figura V.11: Diagrama de caixa aplicada ao domínio IPCA para diferentes horizontes dos modelos ARIMA e WARMA

Tabela V.7: Sumário estatístico da métrica MASE aplicada ao domínio IPCA para diferentes horizontes dos modelos ARIMA e WARMA

	1 - 3		1 - 6		1 - 9		1 - 12		1 - 14	
	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA
Mínima	0.27	0.28	0.33	0.26	0.32	0.41	0.29	0.44	0.27	0.44
1º quartil	0.52	0.85	0.92	1.02	1.02	0.9	1.03	0.91	1.07	1.04
Mediana	1.19	1.42	1.83	1.63	1.74	1.7	1.62	1.68	1.71	1.73
Média	1.3	1.5	1.59	1.74	1.5	1.61	1.52	1.64	1.61	1.73
3º quartil	1.94	1.99	2.15	2.6	1.92	2.19	2	2.41	2.14	2.56
Máxima	2.78	2.84	2.9	3.28	2.34	2.7	2.62	2.6	2.81	2.71
Desvio-padrão	0.87	0.89	0.83	0.98	0.68	0.82	0.76	0.81	0.81	0.83

Taxa de juros

Para o domínio Taxa de juros, a análise comparativa entre os métodos ARIMA e WARMA observa-se um padrão de comportamento ao longo dos diferentes horizontes de previsão. O WARMA demonstra um desempenho notavelmente superior nos horizontes mais curtos, apresentando médias MASE consistentemente menores para os horizontes de 1-3 a 1-6 (Tabela V.8).

À medida que o horizonte de previsão se expande, ambos os métodos mostram um padrão de aumento no erro de previsão, apesar das variações de deterioração diferentes. Nota-se que a diferença de desempenho entre os métodos diminui nos horizontes mais longos (1-12 e 1-14), com ambos apresentando uma deterioração natural da acurácia. É particularmente interessante notar que, mesmo com o aumento do horizonte, o WARMA mantém uma vantagem na maioria das séries, embora essa vantagem seja menos pronunciada nos horizontes curtos.

Um aspecto notável é a variabilidade dos resultados entre as diferentes séries. Por exemplo, enquanto algumas séries fornecem um MASE relativamente baixo mesmo em horizontes longos, outras mostram um aumento mais acentuado no erro de previsão. Esta variabilidade sugere que as características específicas de cada série temporal podem influenciar significativamente o desempenho relativa dos métodos.

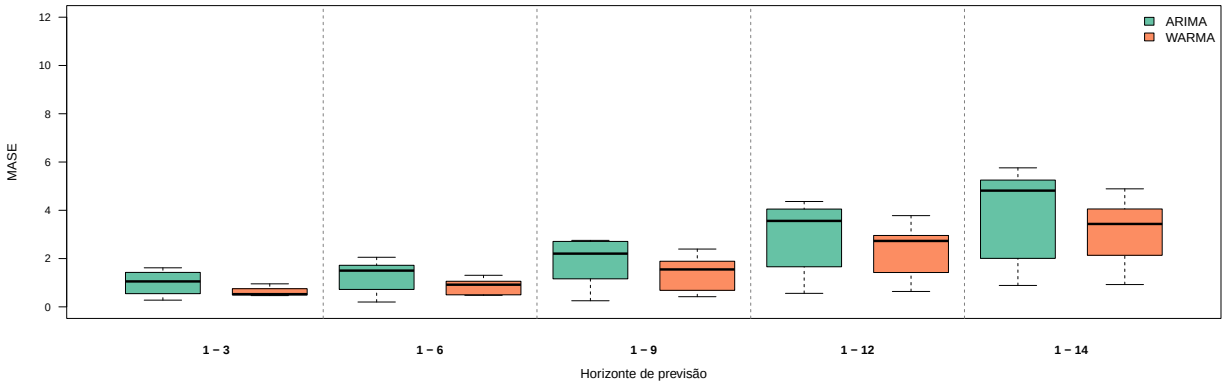


Figura V.12: Diagrama de caixa aplicada ao domínio Taxa de juros para diferentes horizontes dos modelos ARIMA e WARMA

Tabela V.8: Sumário estatístico da métrica MASE aplicada ao domínio Taxa de juros para diferentes horizontes dos modelos ARIMA e WARMA

	1 - 3		1 - 6		1 - 9		1 - 12		1 - 14	
	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA
Mínima	0.28	0.47	0.2	0.48	0.25	0.42	0.56	0.64	0.89	0.92
1º quartil	0.55	0.49	0.72	0.5	1.16	0.68	1.66	1.42	2.01	2.14
Mediana	1.05	0.53	1.5	0.92	2.2	1.55	3.56	2.73	4.81	3.43
Média	0.99	0.63	1.23	0.83	1.85	1.36	2.84	2.27	3.71	3.09
3º quartil	1.43	0.75	1.72	1.06	2.71	1.89	4.05	2.96	5.25	4.06
Máxima	1.62	0.95	2.05	1.31	2.75	2.39	4.36	3.78	5.76	4.89
Desvio-padrão	0.55	0.2	0.74	0.34	1.11	0.79	1.61	1.21	2.07	1.57

V.5.3 Análise global

Na análise global, descontando as séries estacionárias, restou um conjunto de 80 séries não estacionárias. Comparando os métodos ARIMA e WARMA, observa-se um desempenho similar entre os dois métodos, com uma ligeira vantagem para o ARIMA. As diferenças percentuais das médias do MASE são relativamente pequenas, variando de -3,35% no horizonte 1-3 a -0,7% no horizonte 1-14 (Tabela V.9).

Um padrão de deterioração do desempenho é observado em ambos os métodos à medida que o horizonte de previsão aumenta. O MASE médio do ARIMA cresce de 2,37 no horizonte 1-3 para 4,43 no horizonte 1-14, enquanto o WARMA mostra um aumento similar, de 2,45 para 4,46. Além disso, os desvios-padrão aumentam progressivamente com o horizonte em ambos os métodos, indicando maior variabilidade nas previsões de longo prazo.

Concluindo, desconsiderando as poucas exceções, as medidas estatísticas da Tabela V.9 apresentam pequenas variações quando comparadas umas às outras, de forma que se pode argumentar que ambos os métodos ARIMA e WARMA são praticamente equivalentes, considerando o desempenho geral. Essa conclusão pode ser observada ainda nos diagramas de caixa para os diferentes horizontes (Figura V.13).

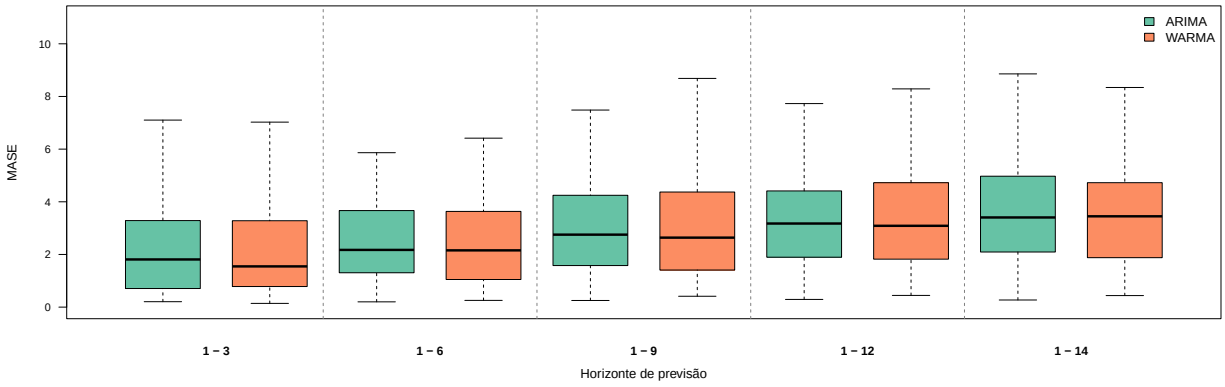


Figura V.13: Diagrama de caixa aplicada diferentes horizontes de séries econômico-financeiras dos modelos ARIMA e WARMA

Tabela V.9: Sumário estatístico da métrica MASE para diferentes horizontes de séries econômico-financeiras dos modelos ARIMA e WARMA

	1 - 3		1 - 6		1 - 9		1 - 12		1 - 14	
	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA	ARIMA	WARMA
Mínima	0.21	0.14	0.2	0.26	0.25	0.41	0.29	0.44	0.27	0.44
1º quartil	0.72	0.79	1.32	1.06	1.62	1.42	1.9	1.85	2.11	1.9
Mediana	1.81	1.55	2.17	2.16	2.75	2.64	3.17	3.09	3.41	3.45
Média	2.4	2.39	2.82	2.91	3.43	3.51	4.08	4.12	4.43	4.42
3º quartil	3.27	3.27	3.64	3.62	4.23	4.35	4.41	4.71	4.95	4.71
Máxima	13.62	15.71	15.8	17.86	20.98	21.99	24.08	25.3	24.93	26.38
Desvio-padrão	2.22	2.43	2.65	2.88	3.17	3.4	3.78	4	4.14	4.34

Capítulo VI Conclusões

As conclusões estão organizadas em três partes. Na Seção VI.1 são apresentadas as considerações finais. Na Seção VI.2 são apresentadas as produções bibliográficos alcançadas ao longo do mestrado. Finalmente, a Seção VI.3 aponta para os trabalhos futuros.

VI.1 Considerações finais

A condição estacionária de uma série é geralmente associada ao fato de que modelos de séries requerem que a série seja estacionária. Apesar disso, embora haja uma gama de métodos que, isoladamente ou combinados, obtém a série estacionária, pouco se discute sobre a preservação do comportamento e das propriedades estatísticas da série original ao aplicar algum método cujo resultado seja uma série estacionária. Ao obter a estacionariedade considerando a média, variância e covariância conjuntamente, fica difícil obter uma visão acerca da série. Além disso, mesmo que seja possível caracterizar essas propriedades por meio de uma avaliação visual, esta avaliação está sujeita ao ponto de vista do usuário em questão. Ademais, testes de hipótese avaliam somente a estacionariedade.

Nesta dissertação, foi apresentada a criação de um modelo que combina pré-processamento e modelagem, denominado Autorregressivo Média-Móvel baseado em Janelas deslizantes (WARMA), no qual o pré-processamento usa janelas deslizantes, visando obter uma série estacionária que preserva o comportamento e as propriedades estatísticas das subsequências que compõem a série original. Para a modelagem, o modelo Autorregressivo Média-Móvel (ARMA) foi selecionado por usar um componente determinístico, que modela a dependência entre as observações da série, e um componente estocástico, que incorpora a influência dos choques aleatórios passados.

Na Avaliação experimental I, verificou-se que, apesar do WARMA focar na média e variância não constantes, *i.e.*, séries de um ou dois passos, as séries avaliadas requereram somente um ou dois passos, sendo, portanto, suficiente para obter a estacionariedade. Todavia, convém observar que essa avaliação foi realizada em séries de natureza econômico-financeira e que, por essa razão, pode não se comportar igualmente em séries de outras naturezas.

Na Avaliação experimental II, os modelos ARIMA e WARMA apresentaram desempenhos complementares em alguns domínios, de modo que, enquanto um demonstrava desempenho superior no

curto prazo, o outro demonstrava desempenho superior no longo prazo. Analisando globalmente, observou-se um desempenho similar entre ambos os modelos para os diversos horizontes avaliados.

De modo geral, embora o WARMA consiga obter uma série estacionária, é recomendável que o WARMA seja aplicado com cautela no âmbito de séries heterocedásticas, necessitando estudos adicionais a fim de averiguar se o WARMA pode ser aplicável nesse contexto, visando estabilizar a variância e em que condições, se houver. Neste caso, assim como o ARIMA, aconselhamos o uso do método Box-Cox antes de proceder com o WARMA.

VI.2 Produção bibliográfica

O artigo que precedeu esta dissertação, denominado Autorregressivo de Integração Adaptativa [Ronald et al., 2019], foi publicado no XXXIV Simpósio Brasileiro de Banco de Dados (SBBD)[1], na categoria de artigos curtos, na qual o artigo submetido deve conter até seis páginas, sendo ainda agraciado como melhor artigo nessa categoria (*Best short paper*).

Neste artigo, o pré-processamento da série não estacionária é restrito à normalização da média, *i.e.*, aplica-se somente às séries que possuam somente média não constante. Além disso, a série resultante do pré-processamento foi modelada somente pelo AR. Na avaliação experimental, observou-se que o ARAI apresentou desempenho de previsão superior ao ARIMA, principalmente em horizontes de curto prazo.

Já para a submissão do artigo decorrente desta dissertação, visa-se o periódico *IEEE Latin America Transactions*[2], cuja previsão de submissão é o primeiro semestre de 2025. Além de explorar domínios de outras áreas, é esperado nesse artigo que outros modelos de séries estacionárias, além do ARIMA, sejam incluídos na avaliação experimental, objetivando fornecer um retorno mais abrangente para o usuário.

VI.3 Trabalhos futuros

Na Avaliação experimental II, os modelos ARIMA e WARMA apresentaram desempenhos complementares em alguns domínios. Diante disso, convém combinar o WARMA com outros modelos usados em séries, considerando que a combinação de múltiplos modelos geralmente fornece uma eficiência maior quando comparados com um único modelo [de O. Santos Júnior et al., 2019; Mariñas-Collado et al., 2022; Briki et al., 2024].

Outro ponto não discutido, que abre margem para trabalhos futuros, é a sazonalidade. Por ora, caso o usuário necessite aplicar o WARMA em séries sazonais, recomendamos aplicar algum método de dessazonalização para, posteriormente, aplicar o WARMA. Não obstante, caso mesmo

[1]Página do simpósio disponível em: https://sbbd.org.br/2019/
[2]Página do periódico disponível em: https://latamt.ieeer9.org/index.php/transactions

assim o usuário deseje aplicar o WARMA em séries sazonais, pode-se aplicar a mesma abordagem usada na regressão linear (Subseção II.3.1), *i.e.*, aplicar o WARMA separadamente para cada um dos períodos que compõem o padrão sazonal.

Apesar da definição de estacionariedade considerar média, variância e covariância, o WARMA obtém a estacionariedade considerando somente a média e a variância, sem abranger séries nas quais a covariância não seja constante. Diante disso, a presente subseção discute a covariância, de forma que possa servir de base para estudos futuros, visando um modelo completo do WARMA, *i.e.*, que considere a média, variância e covariância. Entretanto, não é objetivo desta subseção avaliar a covariância no contexto do WARMA, mas isoladamente

Enquanto o valor da covariância depende da unidade de medida em questão, o intervalo dos valores da correlação limita-se a -1 e 1, inclusive, facilitando dessa forma a interpretabilidade, quando comparada com a covariância, além de ser a razão para a popularidade dessa medida estatística [Diebold, 2007].

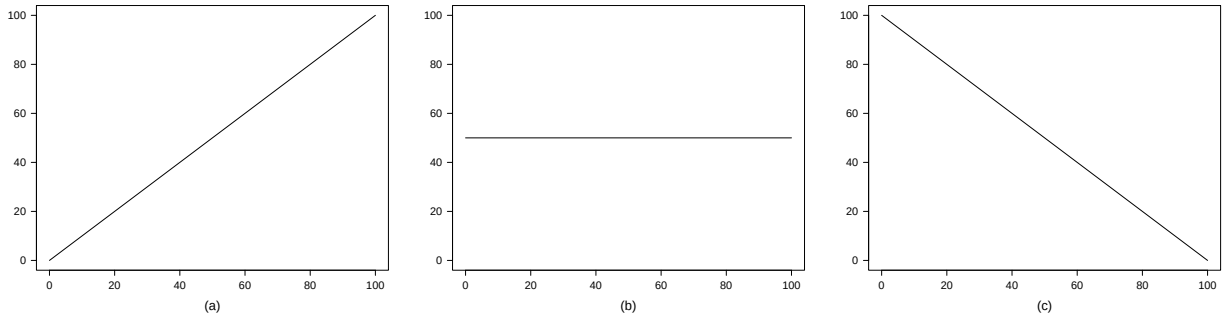


Figura VI.1: Covariância positiva em (a), igual a zero em (b) e negativa em (c)

Jiang et al. [2006] cita a frequência variável ao longo de uma série como um dos motivos para uma série ser não estacionária, uma vez que essa característica implica covariância não constante. Diante disso, Jiang et al. [2006] propõe o conceito de processo G-estacionário, no qual para uma série $\{\mathrm{X}(t), t \in (a, b)\}$, não estacionária, aplica-se ao eixo temporal desta série uma função $\mathsf{g}(t) = u$, visando estabilizar a frequência variável, resultando em uma série $\{\mathrm{Y}(u), u \in (-\infty, \infty)\}$, denominada série dual estacionária, cuja frequência permanece constante ao longo da série e, dessa forma, satisfaz a condição estacionária.

Como a função $\mathsf{g}(t)$ faz a ponte entre as funções $\mathrm{X}(t)$ e $\mathrm{Y}(u)$, um requisito para obter novamente a série $\mathrm{X}(t)$ é que a função $\mathsf{g}(t)$ admita inversa, *i.e.*, $\mathsf{g}^{-1}(u) = t, u \in (\mathsf{g}(a), \mathsf{g}(b))$

Para aplicar o processo G-estacionário, Xu et al. [2012] enumera os seguintes procedimentos

1. Transformar o eixo temporal a fim de obter uma série dual estacionária
2. Analisar a série dual estacionária obtida no item anterior. Isso inclui procedimentos como modelagem, previsão etc

3. Transformar a série dual estacionária de volta à escala original

Quando a frequência de uma série é ou monotonicamente crescente, ou decrescente, uma abordagem geral para obter a série estacionária é aplicar o método Box-Cox no eixo temporal. Tal procedimento, *i.e.*, no qual o método Box-Cox é usado para escalar o eixo temporal a fim de obter uma série estacionária, é denominado processo G(λ)-estacionário, em que λ corresponde ao valor usado pelo método Box-Cox [Jiang et al., 2006].

Outro caso no qual pode ser aplicável o processo G-estacionário é o *chirp* linear, caracterizado por frequência linearmente variável, crescente ou decrescente, ao longo da série. Neste caso, o processo G-estacionário é denominado processo *chirp* linear generalizado [Robertson et al., 2010]. Na Figura VI.2, aplica-se o processo *chirp* linear generalizado, sendo a Figura VI.2 (a) o *chirp* linear e a Figura VI.2 (b) a respectiva série resultante. Observe que o intervalo do *chirp* linear foi ampliado de 10 para 15, a fim de obter uma série com frequência constante.

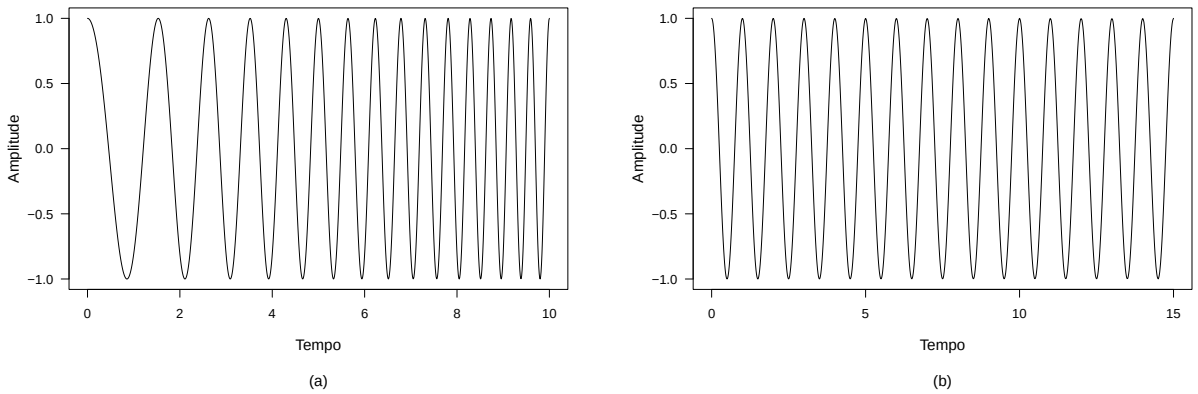


Figura VI.2: *chirp* linear em (a) e a respectiva série estacionária por meio do processo *chirp* linear generalizado em (b)

Apesar de ser um avanço, as aplicações de processos G-estacionário são limitadas a frequências variáveis bem características. No caso de frequências variáveis que se sobrepõem, as aplicações desses processos não conseguem extrair componentes de forma satisfatória. Diante disso, Xu et al. [2012] introduziu o conceito de Filtro-G, capaz de extrair dados de séries com frequências variáveis que se sobrepõem.

Referências

Aguinis, H., Gottfredson, R. K., and Joo, H. Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods*, 16(2):270–301, 2013.

Ali, M., Jones, M. W., Xie, X., and Williams, M. Timecluster: dimension reduction applied to temporal data for visual analytics. *The Visual Computer*, 35(6):1013–1026, 2019.

Atkinson, A. C. and Pericchi, L. R. Grouped Likelihood for the Shifted Power Transformation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):473–482, 1991.

Baillie, R. Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73(1):5–59, 1996.

Bickel, P. J. and Doksum, K. A. An Analysis of Transformations Revisited. *Journal of the American Statistical Association*, 76(374):296–311, 1981.

Box, G. E. P. and Cox, D. R. An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, 1964.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.

Boyd, S. and Vandenberghe, L. *Introduction to Applied Linear Algebra: : Vectors, Matrices, and Least Squares*. Cambridge University Press & Assessment, 2018.

Briki, S., Khabou, N., and Bouassida Rodriguez, I. A Comparative Analysis of Time Series Prediction Techniques a Systematic Literature Review (SLR). In Mosbah, M., Kechadi, T., Bellatreche, L., and Gargouri, F., editors, *Model and Data Engineering*, pages 3–14, Cham. Springer Nature Switzerland, 2024.

Brockwell, P. J. and Davis, R. A. *Time Series: Theory and Methods*. Springer New York, NY, 1991.

Chatfield, C. and Xing, H. *The Analysis of Time Series: An Introduction with R*. CRC Press, 2019.

Chen, Z. and Yang, Y. Assessing forecast accuracy measures, 2004.

Cousineau, D. and Chartier, S. Outliers detection and treatment: a review. *International Journal of Psychological Research*, 3(1):58–67, 2010.

Cryer, J. D. and Chan, K.-S. *Time Series Analysis: With Applications in R.* Springer Science & Business Media, 2008.

Dadkhah, K. *Foundations of Mathematical and Computational Economics.* Springer Berlin Heidelberg, 2011.

De Gooijer, J. G. and Hyndman, R. J. 25 years of time series forecasting. *International Journal of Forecasting*, 22(3):443–473, 2006.

de O. Santos Júnior, D. S., de Oliveira, J. F., and de Mattos Neto, P. S. An intelligent hybridization of ARIMA with machine learning models for time series forecasting. *Knowledge-Based Systems*, 175:72–86, 2019.

Diebold, F. X. *Elements of Forecasting.* Thomson/South-Western, 2007.

Enders, W. *Applied Econometric Time Series.* Wiley, 2014.

Fugal, D. L. *Conceptual Wavelets in Digital Signal Processing: An In-depth, Practical Approach for the Non-mathematician.* Space & Signals Technical Pub., 2009.

Granger, C. W. J. and Joyeux, R. An Introduction to Long-Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis*, 1(1):15–29, 1980.

Graps, A. An introduction to wavelets. *IEEE Computational Science and Engineering*, 2(2):50–61, 1995.

Guido, R. C. Wavelets behind the scenes: Practical aspects, insights, and perspectives. *Physics Reports*, 985:1–23, 2022.

Gujarati, D. N. and Porter, D. C. *Basic Econometrics.* McGraw-Hill Publishing, 2008.

Hawkins, D. M. *Identification of Outliers.* Springer Netherlands, 1980.

Herranz, E. Unit root tests. *WIREs Computational Statistics*, 9(3):e1396, 2017.

Hewamalage, H., Ackermann, K., and Bergmeir, C. Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*, 37(2):788–832, 2022.

Hoshmand, A. R. *Business Forecasting: A Practical Approach.* Routledge, 2009.

Hosking, J. Fractional differencing. *Biometrika*, 68(1):165–176, 1981.

Hubbard, B. B. *The World According to Wavelets The Story of a Mathematical Technique in the Making*. A K Peters/CRC Press, 1998.

Hyndman, R. and Khandakar, Y. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3):1–22, 2008.

Hyndman, R. J. and Athanasopoulos, G. *Forecasting: Principles and Practice*. OTexts, 2018.

Hyndman, R. J. and Koehler, A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.

Iglewicz, B. and Hoaglin, D. *How to Detect and Handle Outliers*. ASQC basic references in quality control. ASQC Quality Press, 1993.

Jarrah, A. and Khanna, N. Some results on vanishing moments of wavelet packets, convolution and cross-correlation of wavelets. *Arab Journal of Mathematical Sciences*, 25(2):169–179, 2019.

Jiang, H., Gray, H. L., and Woodward, W. A. Time-frequency analysis—G(λ)-stationary processes. *Computational Statistics & Data Analysis*, 51(3):1997–2028, 2006.

Joo, T. W. and Kim, S. B. Time series forecasting based on wavelet filtering. *Expert Systems with Applications*, 42(8):3868 – 3874, 2015.

Kim, H., Park, S., and Kim, S. Time-series clustering and forecasting household electricity demand using smart meter data. *Energy Reports*, 9:4111–4121, 2023.

Kim, S. and Kim, H. A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3):669–679, 2016.

Kolassa, S. and Schütz, W. Advantages of the MAD/MEAN ratio over the MAPE. *Foresight: The International Journal of Applied Forecasting*, 1(6):40–43, 2007.

Lahmiri, S. Wavelet low- and high-frequency components as features for predicting stock prices with backpropagation neural networks. *Journal of King Saud University - Computer and Information Sciences*, 26(2):218–227, 2014.

Levendis, J. D. *Time Series Econometrics: Learning Through Replication*. Springer International Publishing, 2018.

Liu, K., Chen, Y., and Zhang, X. An evaluation of ARFIMA (autoregressive fractional integral moving average) programs. *Axioms*, 6(2), 2017.

Liu, X., Liu, A., Chen, J. L., and Li, G. Impact of decomposition on time series bagging forecasting performance. *Tourism Management*, 97:104725, 2023.

Makridakis, S. Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, 9(4):527–529, 1993.

Makridakis, S. and Hibon, M. The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, 16(4):451–476, 2000.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. The m4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1):54–74, 2020.

Mariñas-Collado, I., Sipols, A. E., Santos-Martín, M. T., and Frutos-Bernal, E. Clustering and forecasting urban bus passenger demand with a combination of time series models. *Mathematics*, 10(15), 2022.

McLeod, A. and Hipel, K. Preservation of the rescaled adjusted range: 1. A reassessment of the Hurst Phenomenon. *Water Resources Research*, 14(3):491–508, 1978.

Metcalfe, A. V. and Cowpertwait, P. S. *Introductory Time Series with R*. Springer New York, 2009.

Neusser, K. *Time Series Econometrics*. Springer, 2016.

Ogasawara, E., Martinez, L., De Oliveira, D., Zimbrão, G., Pappa, G., and Mattoso, M. Adaptive Normalization: A novel data normalization approach for non-stationary time series. In *Proceedings of the International Joint Conference on Neural Networks*, 2010.

Ord, J. K., Fildes, R., and Kourentzes, N. *Principles of Business Forecasting*. wessex, Incorporated, 2017.

Parchami, M., Zhu, W.-P., Champagne, B., and Plourde, E. Recent Developments in Speech Enhancement in the Short-Time Fourier Transform Domain. *IEEE Circuits and Systems Magazine*, 16(3):45–77, 2016.

Pernagallo, G. An entropy-based measure of correlation for time series. *Information Sciences*, 643:119272, 2023.

Persons, W. M. A Non-Technical Explanation of the Index of General Business Conditions. *The Review of Economics and Statistics*, 2(2):39–48, 1920.

Pickup, M. *Introduction to Time Series Analysis*. SAGE Publications, 2014.

Pokorny, M. *An Introduction to Econometrics*. Basil Blackwell, 1987.

Rao, J. N. K. and Tintner, G. On THE VARIATE DIFFERENCE METHOD. *Australian Journal of Statistics*, 5(3):106–116, 1963.

Rhif, M., Ben Abbes, A., Farah, I. R., Martínez, B., and Sang, Y. Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review. *Applied Sciences*, 9(7), 2019.

Robertson, S. D., Gray, H. L., and Woodward, W. A. The generalized linear chirp process. *Journal of Statistical Planning and Inference*, 140(12):3676–3687, 2010.

Ronald, A., Salles, R., Belloze, K., Pastore, D., and Ogasawara, E. Modelo Autorregressivo de Integração Adaptativa. In *Anais do XXXIV Simpósio Brasileiro de Banco de Dados*, pages 175–180, Porto Alegre, RS, Brasil. SBC, 2019.

Sakia, R. M. The Box-Cox Transformation Technique: A Review. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 41(2):169–178, 1992.

Salles, R., Belloze, K., Porto, F., Gonzalez, P., and Ogasawara, E. Nonstationary time series transformation methods: An experimental review. *Knowledge-Based Systems*, 164:274–291, 2019.

Sanchez, J. *Time Series for Data Scientists: Data Management, Description, Modeling and Forecasting*. Cambridge University Press, 2023.

Shan, W., Li, L., and He, Q. Research on Weighted Iterative Stage Parameter Estimation Algorithm of Time Series Model. In *Manufacturing Technology, Electronics, Computer and Information Technology Applications*, volume 687 of *Applied Mechanics and Materials*, pages 3968–3971. Trans Tech Publications Ltd, 2014.

Shrestha, M. B. and Bhatta, G. R. Selecting appropriate methodological framework for time series data analysis. *The Journal of Finance and Data Science*, 4(2):71–89, 2018.

Shumway, R. H. and Stoffer, D. S. *Time Series Analysis and Its Applications: With R Examples*. Springer International Publishing, New York, NY, 4 edition, 2017.

Song, D., Chung Baek, A. M., and Kim, N. Forecasting Stock Market Indices Using Padding-Based Fourier Transform Denoising and Time Series Deep Learning Models. *IEEE Access*, 9:83786–83796, 2021.

Tan, Z.-H. and Kraljevski, I. Joint variable frame rate and length analysis for speech recognition under adverse conditions. *Computers & Electrical Engineering*, 40(7):2139–2149, 2014.

Tang, W. H. and Röllin, A. Model identification for ARMA time series through convolutional neural networks. *Decision Support Systems*, 146:113544, 2021.

Tintarev, N., Rostami, S., and Smyth, B. Knowing the unknown: visualising consumption blindspots in recommender systems. In *Proceedings of the 33rd Annual ACM Symposium on Applied*

Computing, SAC '18, page 1396–1399, New York, NY, USA. Association for Computing Machinery, 2018.

Tintner, G. *The Variate Difference Method*. Cowles Commission monographs. Principia Press, Incorporated, 1940.

Tsay, R. S. *Analysis of Financial Time Series*. Wiley, 2010.

Woodward, W. A., Gray, H. L., and Elliott, A. C. *Applied Time Series Analysis with R*. CRC Press, 2017.

Xu, L., Geng, X., He, X., Li, J., and Yu, J. Prediction in Autism by Deep Learning Short-Time Spontaneous Hemodynamic Fluctuations. *Frontiers in Neuroscience*, 13, 2019.

Xu, M., Cohlmia, K. B., Woodward, W. A., and Gray, H. L. G-Filtering Nonstationary Time Series. *Journal of Probability and Statistics*, 2012(1):738636, 2012.

Yang, W. and Zurbenko, I. Nonstationarity. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1):107–115, 2010.

Yeo, I. and Johnson, R. A. A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4):954–959, 2000.