

O Uso de Quase U-Estatísticas para Séries Temporais Unie Multivariadas

Marcio Valk

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO
CIENTÍFICA
DA
UNIVERSIDADE ESTADUAL DE CAMPINAS
PARA
OBTENÇÃO DO TÍTULO
DE
DOUTOR EM ESTATÍSTICA

Programa: Estatística

Orientador: Aluísio de Souza Pinheiro

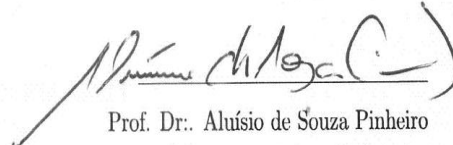
Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da FAPESP

Campinas, março de 2011

O Uso de Quase U-Estatísticas para Séries Temporais Uni e Multivariadas

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por Marcio Valk e aprovada pela Comissão Julgadora.

Campinas 22 de fevereiro de 2011



Prof. Dr. Aluísio de Souza Pinheiro
Orientador

Banca Examinadora:

- Prof. Dr. Aluísio de Souza Pinheiro (IMECC-Unicamp) - Orientador
- Prof. Dr. Luiz Koodi Hotta (IMECC-Unicamp)
- Prof. Dr. Francisco Cribari Neto (DE-UFPE)
- Prof. Dr. Juvêncio Santos Nobre (DEMA-UFC)
- Prof. Dr. Pedro Alberto Morettin (IME-USP)

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de Doutor em Estatística.

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**

Bibliotecária: Maria Fabiana Bezerra Müller – CRB8 / 6162

Valk, Marcio

V239u O uso de quase U-Estatísticas para séries temporais uni e multivariadas/Marcio Valk-- Campinas, [S.P. : s.n.], 2011.

Orientador : Aluísio de Souza Pinheiro

Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1.Análise de séries temporais. 2.Séries temporais. 3.Estatística não paramétrica. 4.Testes de hipóteses estatísticas. 5.Valores estranhos (Estatística). 6.Teoria da previsão. I. Pinheiro, Aluísio de Souza. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Título em inglês: The use of quasi U-Statistics for univariate and multivariate time series

Palavras-chave em inglês (Keywords): 1.Time-series analysis. 2.Time-series. 3.Nonparametric statistics. 4.Statistical hypothesis testing. 5.Outliers (Statistics). 6.Prediction theory.

Área de concentração: Estatística

Titulação: Doutor em Estatística

Banca examinadora: Prof. Dr. Aluísio de Souza Pinheiro (IMECC – UNICAMP)
Prof. Dr. Luiz Koodi Hotta (IMECC – UNICAMP)
Prof. Dr. Francisco Cribari Neto (DE- UFPE)
Prof. Dr. Juvêncio Santos Nobre (DEMA – UFC)
Prof. Dr. Pedro Alberto Morettin (IME - USP)

Data da defesa: 22/02/2011

Programa de Pós-Graduação: Doutorado em Estatística

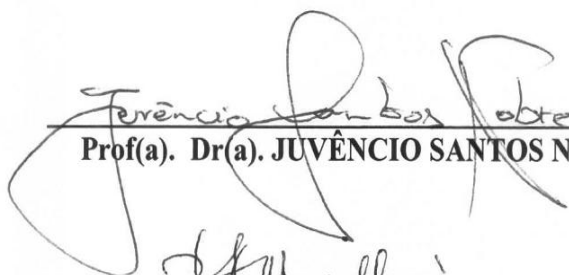
Tese de Doutorado defendida em 22 de fevereiro de 2011 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.


Prof(a). Dr(a). **ALUÍSIO DE SOUZA PINHEIRO**


Prof(a). Dr(a). **LUIZ KOODI HOTTA**


Prof(a). Dr(a). **FRANCISCO CRIBARI NETO**


Prof(a). Dr(a). **JUVÊNCIO SANTOS NOBRE**


Prof(a). Dr(a). **PEDRO ALBERTO MORETTIN**

Agradecimentos

Finalizada a etapa mais importante da minha vida, não poderia deixar de expressar o mais profundo agradecimento a todos aqueles que me apoiaram nesta longa caminhada e contribuíram para a realização deste trabalho.

À minha esposa Karen, o meu maior agradecimento pelo apoio incondicional, compreensão e carinho compartilhados durante os últimos dez anos.

Ao Professor Aluísio Pinheiro, pela disponibilidade, orientação e amizade.

Aos meus amigos Airton, Denise e Lucas, pela amizade e principalmente pelo apoio.

A toda minha família, especialmente, aos meus pais, pelas palavras de incentivo, ao meu sogro e à minha sogra, pelo apoio e ao meu padrinho Pedro, por ter me apresentado a Matemática.

Aos meus colegas e amigos do IMECC. Especialmente às amigas Lorena e Bia, e aos amigos Rodrigo, Márcio Diniz, Márcio Rodrigues, Lucas, Diego, Cícero, Marcus, Lino, Welington, Luiz e José, pela amizade e companhia.

Aos professores do Departamento de Estatística da Unicamp, pela disponibilidade e atenção.

À professora Sílvia R.C. Lopes, pela amizade e apoio no início da minha carreira.

À FAPESP, pelo suporte financeiro.

Resumo

Classificação e agrupamento de séries temporais são problemas bastante explorados na literatura atual. Muitas técnicas são apresentadas para resolver estes problemas. No entanto, as restrições necessárias, em geral, tornam os procedimentos específicos e aplicáveis somente a uma determinada classe de séries temporais. Além disso, muitas dessas abordagens são empíricas. Neste trabalho, propomos métodos para classificação e agrupamento de séries temporais baseados em quase U-estatísticas (Pinheiro et al. (2009) e Pinheiro et al. (2010)). Como núcleos das U-estatísticas são utilizadas métricas baseadas em ferramentas bem conhecidas na literatura de séries temporais, entre as quais o periodograma e a autocorrelação amostral. Três situações principais são consideradas: séries univariadas; séries multivariadas; e séries com valores aberrantes. É demonstrada a normalidade assintótica dos testes propostos para uma ampla classe de métricas e modelos. Os métodos são estudados também por simulação e ilustrados por aplicação em dados reais.

Palavras-chave: Agrupamento de Séries Temporais, Classificação de Séries Temporais, Testes Não-Paramétricos, U-estatísticas, Valores Aberrantes

Abstract

Classification and clustering of time series are problems widely explored in the current literature. Many techniques are presented to solve these problems. However, the necessary restrictions in general, make the procedures specific and applicable only to a certain class of time series. Moreover, many of these approaches are empirical. We present methods for classification and clustering of time series based on Quasi U-statistics (Pinheiro et al. (2009) and Pinheiro et al. (2010)). As kernel of U-statistics are used metrics based on tools well known in the literature of time series, including the sample autocorrelation and periodogram. Three main situations are considered: univariate time series, multivariate time series, and time series with outliers. It is demonstrated the asymptotic normality of the proposed tests for a wide class of metrics and models. The methods are also studied by simulation and applied in a real data set.

Keywords: Non-Parametric Tests, Outliers, Time Series Classification, Time Series Clustering, U-Statistics.

Sumário

1	Introdução	1
2	U-Estatísticas	5
2.1	Noções Preliminares	5
2.1.1	Decomposição de Hoeffding	6
2.1.2	Normalidade Assintótica	8
2.1.3	Algumas Variações de U-Estatísticas	9
2.2	Uma Classe de Quase U-Estatísticas	10
2.2.1	Medidas de Diversidade	11
2.2.2	Propriedades	15
3	Quase U-Estatísticas em Séries Temporais	21
3.1	Uma Classe de Funções-Núcleos	21
3.2	Independência Assintótica de Periodogramas	29
3.3	Núcleos Baseados em Métricas para Séries Temporais	33
3.3.1	Idéia Geral	35
3.3.2	Estatística de Teste Baseada no Periodograma	36
3.3.3	Estatística de Teste Baseada na Autocorrelação Amostral	38
3.3.4	Estatística de Teste Baseada no Log do Periodograma Normalizado	39
4	Classificação e Agrupamento de Séries Temporais	43
4.1	Método de Classificação e Agrupamento	43
4.1.1	Comportamento de B_n com Heterogeneidade Dentro dos Grupos	44
4.1.2	Tamanhos dos Grupos Iguais ao Número de Amostras i.d.	44
4.1.3	Todos os Possíveis Agrupamentos	49
4.2	Estudo da Estatística de Teste através de Simulação	53
4.2.1	Comparação das Métricas através de Simulação	53
4.2.2	Simulação com a Métrica d_{LNP}	53
4.2.3	Simulação com a Métrica d_{ACF}	59
4.2.4	O Caso Não Linear vs Não Linear	67
4.2.5	Estudo de Simulação para Classificação	67

4.2.6	Estudos de Simulação com Três Grupos	74
4.3	Estudos de Simulação para o Método de Agrupamento	77
4.3.1	Simulação sob H_0	80
4.3.2	Correção de p-Valores	81
4.4	Aplicação	82
4.4.1	Procedimento “Leave-One-Out” Utilizando a Métrica d_{LNP}	84
4.4.2	Classificação	84
4.4.3	Procedimento “Leave-One-Out” Utilizando a Métrica d_{ACF}	85
4.4.4	Procedimento “Leave-One-Out” Combinando as Fases S e P	86
5	Caso Multivariado	89
5.1	Séries Temporais Multivariadas	89
5.1.1	Modelo VARMA	90
5.1.2	Modelo VARFIMA(p,d,q)	92
5.1.3	Modelo VGARCH	94
5.2	Métricas para Séries Temporais Multivariadas	96
5.2.1	Normalidade Assintótica no Caso Multivariado	97
5.3	Estudo de Simulação para o Caso Multivariado	100
5.3.1	Poder do Teste	100
5.3.2	Simulação para Comparar as Métricas d_{LNP} , d_P e d_{ACF}	101
5.3.3	Agrupamento de Séries Temporais Bivariadas Correlacionadas	103
5.3.4	Agrupamento de Séries Temporais Bivariadas	103
6	Séries Temporais Com Outliers	105
6.1	Outliers	106
6.2	Outliers Aditivos e Modelos Estacionários	106
6.3	Simulação com Outliers Aditivos	108
7	Discussão	113
	Referências Bibliográficas	115
	Apêndice	121

Lista de Figuras

2.1	Exemplo: Pontos de \mathbb{R}^2	16
4.1	Gráfico das Séries Não Lineares	67
4.2	Gráfico das Séries Utilizadas para Classificação	69
4.3	Séries Estacionárias e Não-Estacionárias Simuladas	78
4.4	O Menor p-Valor entre todos os Possíveis Agrupamentos	80
4.5	O Menor p-Valor Corrigido entre todos os Possíveis Agrupamentos	82
4.6	Sinais Sismológicos	82
4.7	Fase P e Fase S de Terremotos e Explosões	83
4.8	Fase S dos Terremotos	86
4.9	Fase S das Explosões	87
4.10	Procedimento “Leave-One-Out” com Séries de Terremotos	87
4.11	Procedimento “Leave-One-Out” com Séries de Explosões	88
6.1	Efeito de Outliers Aditivos e de Inovação	107
6.2	Os p-Valores do Teste de 6 contra 6 Séries com Diferentes Magnitudes de Outliers	110

Lista de Tabelas

4.1	Valores de θ em Função de D_{WB} , com $N_1 = 6$ e $N_2 = 6$	51
4.2	Valores de θ em Função de D_{WB} , com $N_1 = 5$ e $N_2 = 7$	52
4.3	Comparação do Poder do Teste com Diferentes Métricas	54
4.4	Média e Desvio Padrão da Estatística de Teste da Tabela 4.3	54
4.5	Probabilidades de Erro do Tipo I Empíricos a 5%: Modelo AR(1) e Métrica \mathbf{d}_{LNP}	55
4.6	Média e Desvio Padrão da Estatística de Teste da Tabela 4.5	56
4.7	Probabilidades de Erro do Tipo I Empíricos a 5%: Modelo MA(1) e Métrica \mathbf{d}_{LNP}	56
4.8	Média e Desvio Padrão da Estatística de Teste da Tabela 4.7	57
4.9	Poder Empírico do Teste a 5%: Modelos AR(1) vs MA(1) e Métrica \mathbf{d}_{LNP}	57
4.10	Média e Desvio Padrão da Estatística de Teste da Tabela 4.9	58
4.11	Poder Empírico do Teste a 5%: Modelos MA(1) vs MA(2) e Métrica \mathbf{d}_{LNP}	58
4.12	Poder Empírico a 5%: Modelo MA(1) vs MA(2), Métrica \mathbf{d}_{ACF} e $L = [0.1K]$	59
4.13	Média e Desvio Padrão da Estatística de Teste da Tabela 4.12	59
4.14	Probabilidades de Erro do Tipo I Empíricos a 5%: Modelo MA(1) e Métrica \mathbf{d}_{ACF}	60
4.15	Média e Desvio Padrão da Estatística de Teste da Tabela 4.14	60
4.16	Poder Empírico a 5%: Modelo AR(1) vs MA(2), Métrica \mathbf{d}_{ACF} e $L = 10 + [0.02K]$	61
4.17	Média e Desvio Padrão da Estatística de Teste da Tabela 4.16	61
4.18	Poder Empírico a 5%: Modelo MA(1) vs MA(2), Métrica \mathbf{d}_{ACF} e $L = 10 + [0.02K]$	62
4.19	Média e Desvio Padrão da Estatística de Teste da Tabela 4.18	62
4.20	Probabilidades de Erro do Tipo I Empíricos a 5%: Modelo AR(1) e Métrica \mathbf{d}_{ACF}	63
4.21	Média e Desvio Padrão da Estatística de Teste da Tabela 4.20	63
4.22	Probabilidades de Erro do Tipo I Empíricos a 5%: Modelo MA(1) e Métrica \mathbf{d}_{ACF}	64
4.23	Média e Desvio Padrão da Estatística de Teste da Tabela 4.22	64
4.24	Poder Empírico a 5%: Modelo AR(1) vs MA(1), Métrica \mathbf{d}_{ACF} e $L = [0.1K]$	65
4.25	Média e Desvio Padrão da Estatística de Teste da Tabela 4.24	65
4.26	Poder Empírico a 5%: Modelo MA(1) vs MA(2), Métrica \mathbf{d}_{ACF} e $L = [0.1K]$	66
4.27	Média e Desvio Padrão da Estatística de Teste da Tabela 4.26	66
4.28	Poder Empírico a 5%. Caso Não Linear vs Não Linear	68
4.29	Classificação: Estacionário vs Estacionário.	70
4.30	Classificação: Estacionário vs Não Estacionário.	71

4.31	Classificação: Não Estacionário vs Estacionário.	72
4.32	Classificação: Não Estacionário vs Não Estacionário.	73
4.33	Probabilidades de Erro do Tipo I Empíricos a 5%: Modelo AR(1) com 3 Grupos . . .	74
4.34	Média e Desvio Padrão da Estatística de Teste da Tabela 4.33	75
4.35	Poder Empírico do Teste a 5% com 3 Grupos: 2 AR(1) e 1 MA(1)	75
4.36	Média e Desvio Padrão da Estatística de Teste da Tabela 4.35	76
4.37	Poder Empírico do Teste a 5% com 3 Grupos: AR(1), MA(1) e MA(2)	76
4.38	Média e Desvio Padrão da Estatística de Teste da Tabela 4.37	77
4.39	Modelos Utilizados no Estudo de Simulação para Agrupamento	79
4.40	Porcentagem de Sucesso em Agrupamento de Séries Temporais	80
4.41	“Leave-One-Out” da Fase S Utilizando d_{LNP}	85
4.42	Classificação do Evento de Origem Desconhecida Utilizando a Fase S	85
4.43	“Leave-One-Out” da Fase S Utilizando d_{ACF}	85
5.1	Poder do Teste a 5% para Séries Temporais Bivariadas	101
5.2	Comparação de Métricas no Caso Bivariado.	102
5.3	Percentual de Acerto para Agrupamento de Séries Correlacionadas	103
5.4	Percentual de Acerto para Agrupamento de Séries Multivariadas	104
6.1	Comparação das Métricas d_{LNP} e d_P na Presença de Outliers	109
6.2	Poder do Teste em Séries Temporais com Outliers e Inovações t (1000rep)	110
6.3	Poder do Teste em Séries Temporais com Outliers e Inovações t (10000rep)	111
6.4	Agrupamento de Séries Temporais com Outliers e Inovações t	111
6.5	Poder do Teste para Séries Não-Estacionárias com Outliers e Inovações t	112

Capítulo 1

Introdução

A classificação e o agrupamento de séries temporais são problemas de interesse em diversas áreas, tais como, econômica, negócios, finanças, medicina, biologia, geoinformática (Hirano e Tsumoto (2005), Gvishiani et al. (2008), Savvides et al. (2008), Mezer et al. (2009), Otranto (2010)). Do ponto de vista conjuntural, um fenômeno muito comum é o aparecimento de observações atípicas, chamados “outliers”. Fajardo et al. (2009) estudam o impacto de outliers aditivos nas estimativas de parâmetros de modelos estacionários, especialmente o modelo ARFIMA. Por ser um campo tão diversificado, a identificação de semelhanças ou diferenças em séries temporais é útil para tomar decisões e fazer previsões. Suponha que temos um número de séries temporais e queremos fazer previsões. Como resultado dos testes para diferenças entre os processos subjacentes, grupos de séries temporais semelhantes podem ser identificados. Então, ao invés de ajustar modelos para todas as séries e fazer previsão de cada uma delas, um modelo único pode ser ajustado sendo um representante de cada grupo e, em seguida, a previsão pode ser realizada neste representante. Isto é especialmente útil quando o problema é fazer previsão de um grande número de séries temporais, como pode ser o caso do controle de estoque. Em termos de tempo e custos reduzidos isso certamente seria mais vantajoso. Também é sabido que melhores estimativas são obtidas por agregação de conjuntos de dados similares (Aigner e Goldfeld (1974)). Estes dados semelhantes podem ser identificados com base em técnicas para a diferenciação entre eles.

Na área da economia, a globalização tem acelerado a integração dos mercados financeiros mundiais nos últimos anos, tornando-os mais dependentes do que anteriormente. É preciso considerá-los conjuntamente para compreender melhor a estrutura dinâmica das finanças globais. Além disso, a criação de bancos de dados e os avanços na área da informática permitem o armazenamento e o processamento de um grande volume de dados. Por estes motivos, torna-se imprescindível desenvolver técnicas que reduzam a quantidade de dados sem redução significativa de informação.

O problema da identificação de semelhanças ou diferenças entre séries temporais tem sido estudado por vários autores, tais como Galeano e Peña (2000), Piccolo (1990) e Caiado et al. (2006). Um problema fundamental neste tipo de análise de séries temporais é a escolha de

uma métrica relevante. A abordagem empírica que é utilizada é inadequada porque não tem uma estatística de teste e sempre encontra, pelo menos, dois grupos, mesmo que as séries temporais sejam identicamente distribuídas. Maharaj (2000) apresenta um teste de hipótese na comparação de duas séries temporais estacionárias, baseada nos parâmetros autorregressivos e propôs um método de classificação usando o p-valor deste teste como uma medida de similaridade. Maharaj (2002) compara duas séries temporais não-estacionárias utilizando uma abordagem de espectros evolutivos, levando em consideração as alterações estruturais ao longo do tempo. Essas técnicas de comparação de séries temporais são aplicáveis tanto a séries que são estacionárias como a séries não estacionárias.

Com o objetivo de encontrar similaridades ou dissimilaridades em grupos de n vetores aleatórios, cada um de comprimento K , Pinheiro et al. (2009), generalizando o trabalho de Pinheiro et al. (2005), introduzem uma classe de U-estatísticas generalizadas e as incorpora em uma MANOVA convencional para dados de alta dimensionalidade, com ênfase em dados categóricos. Essa abordagem baseia-se na formulação geral das distâncias de Hamming, cujas contrapartes amostrais são U-estatísticas generalizadas. Isso é especialmente útil, pois na análise de dados qualitativos de alta dimensionalidade, ferramentas convencionais, como a análise multivariada discreta, são ineficientes ou inapropriadas. Também é comum a exigência de homoscedasticidade e normalidade dos dados para inferência. Para encontrar similaridades ou dissimilaridades em grupos de n vetores aleatórios, cada um de comprimento K , os esquemas amostrais estudados são: (i) um grande número de séries curtas, isto é, $n \gg K$; (ii) um pequeno (ou médio) número de séries longas, isto é, $n \ll K$; e (iii) um grande número de séries longas, isto é, n e K grandes.

Com relação à separação de um conjunto de vetores aleatórios em grupos, a situação (i) caracteriza a ANOVA clássica, para a qual procedimentos de decomposição de variância podem ser implementados com algum grau de sucesso, principalmente para variáveis aleatórias contínuas. Os trabalhos de Galeano e Peña (2000), Piccolo (1990) e Caiado et al. (2006), citados anteriormente se encaixam na situação (ii).

O que propomos neste trabalho é a adaptação e generalização dos resultados trabalho de Pinheiro et al. (2009) para séries temporais uni e multivariadas, criando assim ferramentas que podem ser utilizadas para resolver os problemas descritos nos primeiros parágrafos deste capítulo. Tal adaptação envolve a obtenção de resultados assintóticos para U-estatísticas baseadas em sequências de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) e sequências com alguma estrutura de dependência. Podemos utilizar os resultados de Pinheiro et al. (2009) para qualquer um dos casos (i)-(iii). A normalidade assintótica de U-estatísticas baseadas em sequências i.i.d. é obtida por Hoeffding (1948). Representações martingais, tais como, a representação martingal de Hoeffding (1961) e a representação por martingais reversos de Berk (1966), permitem demonstrar a Lei Forte dos Grandes Números para essa classe U-estatísticas. No caso em que a sequência é fracamente estacionária, as

propriedades assintóticas são desenvolvidas sob condições “mixing”. Yoshihara (1976) obtém normalidade assintótica com a condição de regularidade absoluta. Outros exemplos de propriedades assintóticas, que são desenvolvidas sob condições “mixing”, podem ser encontrados em Denker e Keller (1983). A Lei dos grandes números para sequências estacionárias e ergódicas é estudada por Aaronson et al. (1996). Para processos com longa dependência, U-estatísticas são consideradas por Dehling e Taqqu (1989, 1991).

Rifi e Utzet (2000), Major (1994) e O’Neil e Redner (1993) estudam as propriedades assintóticas de U-estatísticas ponderadas baseadas em sequências i.i.d.. Hsing e Wu (2004) demonstram a normalidade assintótica de U-estatísticas ponderadas quando as observações provêm de uma série temporal não linear e também de processos lineares. No entanto, esses resultados são mais restritivos que os resultados de Pinheiro et al. (2009) pois, neste último, permite-se que o núcleo da U-estatística seja função de mais de uma amostra, sendo portanto uma U-estatística generalizada. Pinheiro et al. (2010) consideram uma classe de U-estatísticas baseadas em um núcleo de ordem $m \geq 3$, estacionário de ordem r e que pode ser aplicado a quaisquer vetores aleatórios i.i.d. de dimensão arbitrária K (mesmo que $K \rightarrow \infty$). Um teste para componentes de variância em modelos de efeitos aleatórios baseado nos resultados de Pinheiro et al. (2009) é apresentado por Nobre et al. (2008).

No Capítulo 2, introduzimos as U-estatísticas e apresentamos os resultados do trabalho de Pinheiro et al. (2009) que são a base para o desenvolvimento deste trabalho.

Para testar a homogeneidade de grupos de séries temporais, adaptamos os resultados de Pinheiro et al. (2009), obtendo uma classe de quase U-estatísticas e uma classe de séries temporais, para os quais a normalidade assintótica da estatística de teste ocorre. Esses resultados são apresentados no Capítulo 3. Para a proposta quase U-estatísticas, sob hipótese de homogeneidade de grupos, a normalidade assintótica ocorre com a taxa de n em vez de \sqrt{n} .

No Capítulo 4, generalizamos este resultado, provando que a normalidade assintótica ocorre mesmo se a hipótese de homogeneidade de cada um dos grupos não for satisfeita. A distribuição da estatística do teste é alterada estatisticamente a medida em que a configuração de grupos testados é alterada. No entanto, o poder do teste é maior quando a hipótese de mesma distribuição dentro dos grupos é satisfeita. Isso nos permite agrupar séries temporais executando vários testes utilizando um procedimento bootstrap.

No Capítulo 5, generalizamos os resultados obtidos no caso univariado, para séries temporais multivariadas e no Capítulo 6, testamos os métodos propostos em séries temporais com outliers. Finalmente, no Capítulo 7 apresentamos uma breve discussão sobre o trabalho e no Apêndice apresentamos alguns tópicos sobre séries temporais.

Capítulo 2

U-Estatísticas

U-estatísticas foram introduzidas por Hoeffding (1948) e formam uma classe muito ampla de funcionais especialmente importante na teoria de estimação. Em estatística básica, elas surgem naturalmente na construção de estimadores não-viesados de mínima variância. A teoria relacionada a U-estatísticas proporciona uma estrutura teórica simples, que pode ser utilizada em estatística não-paramétrica para provar resultados assintóticos de uma ampla classe de testes estatísticos e estimadores. Exposições detalhadas sobre o tema podem ser encontradas em Denker (1985), Lee (1990), Sen e Singer (1993) e Lehmann (1999).

Neste capítulo, apresentamos a classe de U-estatísticas. Na Seção 2.1, introduzimos uma série de conceitos sobre U-estatísticas e reproduzimos resultados clássicos, como a normalidade assintótica para U-estatísticas (Hoeffding (1948)). Na Seção 2.2, apresentamos uma classe de quase U-estatísticas em que se baseiam os métodos propostos nessa tese (Pinheiro et al. (2009, 2010)).

2.1 Noções Preliminares

Classes de U-estatísticas são importantes por várias razões. Nelas, estão incluídas muitas estatísticas usuais, como por exemplo, os estimadores de média e variância \bar{X} e S^2 . A estrutura simples das U-estatísticas é ideal para estudar processos de estimação em geral, tais como bootstrap e jackknife, e as aplicações da teoria frequentemente geram novas estatísticas úteis para problemas práticos inferenciais.

As U-estatísticas surgem a partir da representação de uma característica populacional de interesse como funcional da função de distribuição. Para isto, considera-se um funcional definido sobre um conjunto \mathcal{D} de funções de distribuição em \mathbb{R} , i.e.,

$$\theta \equiv \theta(F), \quad F \in \mathcal{D}. \tag{2.1}$$

O objetivo é estimar $\theta(F)$ através de uma amostra i.i.d., X_1, \dots, X_n , com distribuição F , assumindo-se que \mathcal{D} seja conhecido e $F \in \mathcal{D}$ seja desconhecida. Halmos (1946) mostra que

um funcional $\theta(\cdot)$ pode ser estimado sem viés se, e somente se, existe uma função $\phi : \mathbb{R}^k \rightarrow \mathbb{R}$, para algum $k \leq n$, tal que

$$\theta(F) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(x_1, \dots, x_k) dF(x_1) \dots dF(x_k), \quad \forall F \in \mathcal{D}. \quad (2.2)$$

Todo funcional que satisfaz a condição (2.2) para alguma função $\phi(\cdot, \dots, \cdot)$ é chamado *funcional estatístico regular de grau k* . A função $\phi(\cdot, \dots, \cdot)$ associada é chamada de núcleo de θ .

Diz-se que $\theta(F)$ é um *parâmetro estimável* em \mathcal{D} se, para algum inteiro positivo k , existe um estimador não-viesado de $\theta(F)$ baseado em n variáveis aleatórias (v.a.'s) i.i.d. com distribuição F , isto é, se existe um núcleo $f : \mathbb{R}^k \rightarrow \mathbb{R}$, tal que

$$\mathbb{E}_F f(X_1, \dots, X_k) = \theta(F), \quad \forall F \in \mathcal{D}.$$

O menor inteiro k com esta propriedade é chamado de *grau de $\theta(\cdot)$* .

O resultado que garante a existência e a unicidade de um estimador simétrico, sem viés e de mínima variância para θ foi provado por Halmos (1946).

Definição 2.1.1. (Hoeffding, 1948). Sejam X_1, \dots, X_n variáveis aleatórias *i.i.d.* com distribuição F . Seja $\theta \equiv \theta(F)$ uma característica populacional que pode ser escrita como

$$\theta = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(x_1, \dots, x_k) dF(x_1) \dots dF(x_k) = \mathbb{E}\phi(X_1, \dots, X_k),$$

em que o núcleo ϕ é tomado, sem perda de generalidade, como simétrico em seus argumentos. Assumindo-se $n \geq k$, a U-estatística de grau k para a estimação de θ é dada por

$$U_n = \binom{n}{k}^{-1} \sum_{(n,k)} \phi(X_{i_1}, \dots, X_{i_k}), \quad (2.3)$$

em que $\sum_{(n,k)}$ denota a soma sob as permutações distintas de tamanho k de $\{1, \dots, n\}$.

2.1.1 Decomposição de Hoeffding

Sejam

$$\begin{aligned} \phi_c(x_1, \dots, x_c) &= \mathbb{E}[\phi(X_1, \dots, X_k) | X_1 = x_1, \dots, X_c = x_c], \\ \sigma_c^2 &= \text{Var}\phi_c(X_1, \dots, X_c), \quad \text{para } c = 1, \dots, k \end{aligned} \quad (2.4)$$

e $\sigma_0^2 = 0$. Se $\sigma_1^2 = 0$, diz-se que a U-estatística é *degenerada*. Se $\sigma_1^2 = \dots = \sigma_{c-1}^2 = 0$ e $\sigma_c^2 > 0$, diz-se que a U-estatística é *degenerada de ordem c*. A mesma U-estatística é denotada por Hoeffding (1948), como *estacionária de ordem c-1*. Neste trabalho, adotamos essa última notação.

Note que

$$\phi_c(x_1, \dots, x_c) = \mathbb{E}[\phi_d(X_1, \dots, X_d) | X_1 = x_1, \dots, X_c = x_c], \quad (2.5)$$

para $1 \leq c < d \leq k$ e

$$\mathbb{E}\phi_c(x_1, \dots, x_c) = \mathbb{E}\phi(X_1, \dots, X_k), \quad 1 \leq c \leq k. \quad (2.6)$$

Para a variância da U-estatística, uma expressão útil pode ser desenvolvida em termos de σ_c^2 . Considere U_n uma U-estatística com núcleo ϕ de grau k . Então, (Hoeffding (1948))

$$\text{Var}U_n = \binom{n}{k}^{-1} \sum_{c=1}^k \binom{k}{c} \binom{n-k}{k-c} \sigma_c^2. \quad (2.7)$$

Uma das propriedades mais elegantes e úteis de toda a teoria de U-estatísticas é a decomposição de Hoeffding, que demonstra que toda U-estatística de grau k pode ser escrita como uma combinação linear de U-estatísticas não-correlacionadas de graus $1, \dots, k$ (Hoeffding (1948)). Esta decomposição é baseada nas seguintes projeções ortogonais de ϕ :

$$\begin{aligned} \psi_1(x_1) &= \phi_1(x_1) - \theta \\ \psi_2(x_1, x_2) &= \phi_2(x_1, x_2) - \phi_1(x_1) - \phi_1(x_2) - \theta \\ &\vdots \\ \psi_c(x_1, \dots, x_c) &= \phi_c(x_1, \dots, x_c) - \sum_{j=1}^{c-1} \sum_{(c,j)} \psi_j(x_{i_1}, \dots, x_{i_j}) - \theta, \end{aligned} \quad (2.8)$$

para $c = 3, 4, \dots, k$, em que $\sum_{(c,j)}$ é tomada para todos os subconjuntos de cardinalidade j de $\{x_1, \dots, x_c\}$.

Teorema 2.1.1. (Hoeffding, 1948). *Seja U_n uma U-estatística de grau k baseada no núcleo ϕ . Então,*

$$U_n = \theta + \sum_{j=1}^k \binom{k}{j} H_n^{(j)}, \quad (2.9)$$

em que $H_n^{(j)} = \binom{n}{j}^{-1} \sum_{(n,j)} \psi_j(X_{\nu_1}, \dots, X_{\nu_j})$ e os ψ_j 's são definidos em (2.8).

Demonstração. Utilizando a relação

$$\sum_{(n,k)} S_{j\{i_1, \dots, i_k\}} = \binom{n-j}{k-j} \sum_{(n,j)} \psi_j(x_{\nu_1}, \dots, x_{\nu_j}),$$

em que $S_{j\{i_1, \dots, i_k\}} = \sum \psi_j(x_{\nu_1}, \dots, x_{\nu_j})$, com a soma sendo realizada sobre todos os subconjuntos de cardinalidade j , $\{\nu_1, \dots, \nu_j\}$, de $\{i_1, \dots, i_k\}$; a identidade

$$\binom{n}{k}^{-1} \binom{n-j}{k-j} = \binom{k}{j} \binom{n}{j}^{-1};$$

e a relação (2.8) para $c = k$, segue que

$$\begin{aligned} U_n &= \binom{n}{k}^{-1} \sum_{(n,k)} \phi(x_{i_1}, \dots, x_{i_k}) = \binom{n}{k}^{-1} \sum_{(n,k)} \left(\sum_{j=1}^k S_{j\{i_1, \dots, i_k\}} + \theta \right) \\ &= \theta + \binom{n}{k}^{-1} \sum_{j=1}^k \binom{n-j}{k-j} \sum_{(n,j)} \psi_j(x_{\nu_1}, \dots, x_{\nu_j}) \\ &= \theta + \sum_{j=1}^k \binom{k}{j} H_n^{(j)}, \end{aligned}$$

em que $H_n^{(j)}$ pode ser interpretada como uma U-estatística de grau j baseada no núcleo ψ_j . \square

Truncando-se a soma (2.9) em $c \leq k$, tem-se

$$U_n = \theta + \sum_{j=1}^c \binom{k}{j} H_n^{(j)} + R_n^{(c)}, \quad (2.10)$$

em que $R_n^{(c)}$ é uma U-estatística com núcleo $\sum_{j=c+1}^k g_j(x_1, \dots, x_k)$, para $g_j(x_1, \dots, x_k) = S_{j\{1, \dots, k\}}$.

A decomposição de Hoeffding é a ferramenta básica para demonstrar vários resultados para U-estatísticas como, por exemplo, a normalidade assintótica que é tratada na subseção seguinte.

2.1.2 Normalidade Assintótica

Hoeffding (1948) demonstra que as U-estatísticas com $\sigma_1^2 > 0$ são assintoticamente normais.

Teorema 2.1.2. (Hoeffding, 1948). *Se $0 < \sigma_1^2 < \infty$, então, quando $n \rightarrow \infty$,*

$$\sqrt{n}(U - \theta) \xrightarrow{D} N(0, k^2 \sigma_1^2). \quad (2.11)$$

Demonstração. Segue da decomposição de Hoeffding que

$$U - \theta = kH_n^{(1)} + R_n.$$

Logo,

$$\begin{aligned} \sqrt{n}(U - \theta) &= \sqrt{n}kH_n^{(1)} + \sqrt{n}R_n \\ &= \frac{k}{\sqrt{n}} \sum_{i=1}^n \psi_1(X_i) + \sqrt{n}R_n, \end{aligned}$$

em que

$$R_n = \sum_{j=2}^k \binom{k}{j} H_n^{(j)}.$$

A variância de $\sqrt{n}R_n$ é $n \sum_{j=2}^k \binom{k}{j}^2 \binom{n}{j}^{-1} \delta_j^2$, em que $\binom{n}{j} \delta_j = \text{Var}H_n^{(j)}$, o que implica que

$$\text{Var}\sqrt{n}R_n = n\text{Var}R_n = O(n^{-1}).$$

Então, pelo Teorema de Slutsky, o comportamento assintótico de $\sqrt{n}(U_n - \theta)$ é o mesmo de $n^{-1/2}k \sum_{i=1}^n \psi_1(X_i)$. Como $\mathbb{E}\psi_1(X_1) = 0$, $\text{Var}\psi_1(X_1) = \sigma_1^2 > 0$ e $\psi_1(X_1), \dots, \psi_1(X_n)$ é uma sequência de v.a.'s i.i.d., pode-se utilizar o Teorema Central do Limite (TCL) clássico, o que demonstra o teorema. \square

2.1.3 Algumas Variações de U-Estatísticas

Algumas generalizações da classe de U-estatísticas levam a outros testes e estatísticas clássicos. Uma classe de U-estatísticas baseadas em núcleos de ordem $m \geq 3$, estacionário de ordem r e que pode ser aplicado a quaisquer vetores aleatórios i.i.d. de dimensão arbitrária K (mesmo que $K \rightarrow \infty$), é considerada por Pinheiro et al. (2010). Neste trabalho, considera-se que a estatística T_n seja uma combinação linear definida por

$$T_n = \sum_{i_1, \dots, i_m}^{1, n} \eta_{n, i_1, \dots, i_m} \phi(\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_m}), \quad (2.12)$$

em que $\mathbf{X}_1, \dots, \mathbf{X}_n$ são vetores aleatórios K -dimensionais i.i.d., não necessariamente quantitativos como em Pinheiro et al. (2005)); $\eta_{n, i_1, \dots, i_m}$ é um conjunto de pesos, possivelmente aleatório, que satisfazem

$$\sum_{i_1, \dots, i_m}^{1, n} \eta_{n, i_1, \dots, i_m} = 0; \quad (2.13)$$

$\sum_{i_1, \dots, i_m}^{1, n}$ varia sobre todas as permutações estritamente ordenadas de $1, \dots, n$ e ϕ é um núcleo de grau m , estacionário de ordem r , para o qual $\mathbb{E}\phi(X_1, \dots, X_m) = \theta$. Por (2.12), pode-se assumir sem perda de generalidade que $\theta = 0$.

Algumas configurações de $\eta_{n, i_1, \dots, i_m}$ conduzem a classes especiais de U-estatísticas generalizadas. Por exemplo, se $\eta_{n, i_1, \dots, i_m} \equiv m!(n-m)!/n!$ e $r \geq 1$, T_n é uma U-estatística estacionária de ordem r , para a qual as variâncias condicionais são tais que, $0 = \sigma_1^2 = \dots = \sigma_r^2 < \sigma_{r+1}^2$. Neste caso, $n^{(r+1)/2}(T_n - \theta)$ converge para uma (possivelmente) combinação linear infinita de variáveis aleatórias independentes, cada uma destas distribuídas conforme uma integral de Wiener $(r-1)$ -dimensional (Dynkin e Mandelbaum (1983)).

Se $K = 1$ e $\eta_{n, i_1, \dots, i_m}$ assume somente valores 0 ou 1, T_n é dita ser uma U-estatística *incompleta*. A distribuição assintótica de T_n será uma combinação linear de integrais de Wiener independentes ou uma mistura de tal distribuição com uma v.a. normal e independente, sob suaves condições amostrais (Janson (1984)). Para uma classe de U-estatísticas condicionais, em que os pesos podem ser decompostos da forma $\eta_{n, i_1, \dots, i_m} = e(i_1) \dots e(i_k)$, sendo $e(\cdot)$ a função peso marginal, a normalidade assintótica segue de Stute (1991). Além disso, a natureza condicional da classe provém do fato de que os pesos são definidos como funções aleatórias de um outro conjunto de v.a.'s i.i.d..

Para $K = 1$, O'Neil e Redner (1993) e Major (1994) apresentam resultados assintóticos em uma configuração mais geral para a classe de U-estatísticas ponderadas, definidas por (2.13). No caso em que $m = 2$, O'Neil e Redner (1993) discutem o uso da técnica "moment matching" para determinar a distribuição assintótica de T_n . Sob algumas condições de regularidade para $\eta_{n, i_1, \dots, i_m}$, um limite não-normal é provado para $r = 1$ ou $r = 0$. Para $r = 0$, prova-se que a classe de U-estatísticas ponderadas é assintoticamente normal sob o segundo conjunto de condições nos pesos. Normalidade assintótica também é demonstrada para $r = 1$ no contexto de U-estatísticas *incompletas*. A idéia comum no contexto de U-estatísticas ponderadas é a ortogonalidade do conjunto de pesos (possivelmente aleatórios). Major (1994) destaca que a abordagem acima mencionada não pode ser adotada para $m \geq 3$. Uma aproximação Poisson é utilizada para procurar o comportamento assintótico de T_n . Quatro resultados principais estabelecem a distribuição assintótica para U-estatísticas ponderadas, sob diferentes esquemas de ponderamento. Em três situações, $r \leq m$, e, para todos os quatro casos, supõem-se que X_1, \dots, X_n são v.a.'s i.i.d. uniformemente distribuídas.

Considerando $K \geq 2$, impossibilita-se a utilização dos resultados de Major (1994). No entanto, a classe proposta por Pinheiro et al. (2009) é construída de tal forma que, embora ϕ seja degenerada, a escolha dos pesos garantem a normalidade assintótica sob suaves condições. Tais condições podem ser facilmente interpretadas em termos de normas l^p , $p \geq 2$.

2.2 Uma Classe de Quase U-Estatísticas

Em muitas situações X_1, \dots, X_n são vetores K -dimensionais, os quais serão denotados aqui por $\mathbf{X}_1, \dots, \mathbf{X}_n$, quantitativos, ou até mesmo, qualitativos. Estas condições surgem, por exemplo, no estudo de seqüências de DNA, cuja cardinalidade K , ou seja, o número de posições, é bem maior que o tamanho amostral n ($K \gg n$) e, em séries financeiras. Em economia e finanças, podemos estar interessados em um grande número de ativos financeiros. O valor de cada um destes ativos varia ao longo do tempo gerando uma série temporal. A cada dia surgem novos ativos no mercado e seria interessante se pudéssemos colocá-los em grupos cujo comportamento seja similar. Para superar as limitações da análise de variância clássica (ANOVA), como a exigência de homogeneidade e normalidade da fonte de variação, e da ANOVA multivariada (MANOVA), para a qual um grande tamanho amostral é necessário, Pinheiro et al. (2009) introduzem medidas de diversidade adequadas e apresentam uma decomposição para dados de alta dimensionalidade, cujas contrapartes amostrais são U-estatísticas generalizadas. Tal decomposição é essencial para o desenvolvimento deste trabalho e será apresentada nesta seção.

2.2.1 Medidas de Diversidade

Sejam X_1, \dots, X_n v.a.'s i.i.d., não necessariamente contínuas ou sequer quantitativas, seguindo distribuição F . Sendo as X_i 's quantitativas, sua diversidade ou dispersão é medida em termos da variabilidade de F ; uma medida usual é dada pelo desvio-padrão σ_F , tal que $\sigma_F^2 = \mathbb{E}(X - \mathbb{E}X)^2$ pode ser reescrita como

$$\sigma_F^2 = \mathbb{E}d(X_1, X_2), \text{ em que } d(x, y) = \frac{1}{2}(x - y)^2. \quad (2.14)$$

Suponha também que tenhamos uma segunda amostra Y_1, \dots, Y_m , independentemente obtida de uma distribuição H , de forma que $\sigma_H^2 = \mathbb{E}d(Y_1, Y_2) = \frac{1}{2}\mathbb{E}(Y_1 - Y_2)^2$. Interessamo-nos então trabalhar com $\gamma(F, H) = \mathbb{E}[d(X, Y)]$; podemos notar que

$$\gamma(F, H) = \frac{1}{2}(\sigma_F^2 + \sigma_H^2) + \frac{1}{2}(\mathbb{E}X - \mathbb{E}Y)^2 \geq \frac{1}{2}(\sigma_F^2 + \sigma_H^2), \quad \forall F, H, \quad (2.15)$$

em que a igualdade vale se, e somente se, $\mathbb{E}X = \mathbb{E}Y$. Esta simples igualdade pode ser estendida para o caso multiamostrais sem que seja necessária a existência de homoscedasticidade. Considere em seguida uma extensão para vetores aleatórios $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})'$ e $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ip})'$, cujas médias são denotadas pelos vetores $\boldsymbol{\mu}_1$ e $\boldsymbol{\mu}_2$ e matrizes de dispersão $\boldsymbol{\Sigma}_1$ e $\boldsymbol{\Sigma}_2$, respectivamente. Novamente, teremos

$$\Gamma(F, H) = \frac{1}{2}\mathbb{E}\{(\mathbf{X} - \mathbf{Y})(\mathbf{X} - \mathbf{Y})'\} = \frac{1}{2}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) + \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)', \quad (2.16)$$

de forma que para todo $\boldsymbol{\lambda} \in \mathbb{R}^p$,

$$\begin{aligned} \boldsymbol{\lambda}'\Gamma(F, H)\boldsymbol{\lambda} &= \frac{1}{2}\{\boldsymbol{\lambda}'(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\boldsymbol{\lambda}\} + \frac{1}{2}\|\boldsymbol{\lambda}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\|^2 \\ &\geq \boldsymbol{\lambda}'\left\{\frac{1}{2}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\right\}\boldsymbol{\lambda}, \quad \forall F, H, \end{aligned} \quad (2.17)$$

sendo a igualdade verdadeira se, e somente se, $\boldsymbol{\lambda}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \mathbf{0}$, $\forall \boldsymbol{\lambda} \in \mathbb{R}^p$, i.e., $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. Se escolhermos uma função real $\phi(\mathbf{A})$, para uma matriz positiva-definida \mathbf{A} , como uma norma, gostaríamos de restringir nossa atenção a uma classe de $\phi(\cdot)$'s tal que

$$\phi(\Gamma(F, H)) \geq \frac{1}{2}\{\phi(\boldsymbol{\Sigma}_1) + \phi(\boldsymbol{\Sigma}_2)\}, \quad \forall \phi \in \Phi, \quad (2.18)$$

sendo a igualdade verdadeira se, e somente se, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. Por ambas as versões empíricas de $\boldsymbol{\Sigma}_1$ e $\boldsymbol{\Sigma}_2$ serem U-estatísticas, e sendo $\Gamma(F, H)$ uma U-estatística, (2.18) pode ser incorporada a um problema de teste MANOVA (para a homogeneidade dos vetores de média) sem que seja fundamental a homogeneidade de $\boldsymbol{\Sigma}_1$ e $\boldsymbol{\Sigma}_2$.

Entre as possíveis escolhas de $\phi(\cdot)$, as mais comumente utilizadas são (i) o critério de variância generalizada $\phi(\mathbf{A}) = |\mathbf{A}|$, o determinante de \mathbf{A} (usualmente elevado a potência $\frac{1}{p}$), (ii) o critério do traço $\phi(\mathbf{A}) = \text{tr}(\mathbf{A})$, e (iii) o máximo autovalor de Roy (1953), $\phi(\mathbf{A}) = \text{ch}_{\max}(\mathbf{A})$. Note que

$$\text{ch}_{\max}(\mathbf{A}) \geq \frac{1}{p}\text{tr}(\mathbf{A}) \geq |\mathbf{A}|^{\frac{1}{p}} \geq 0, \quad (2.19)$$

em que $|\mathbf{A}|$ só pode ser nulo se \mathbf{A} for deficiente em posto. Para o critério do traço, (2.17) nos leva a (2.18) de uma forma aditiva e o mesmo vale para o critério de Roy de uma forma subaditiva. Note também que se \mathbf{A} e \mathbf{B} são matrizes positivas semidefinidas simétricas de ordem $p \times p$, então, por transformações ortogonais, ambas podem ser reduzidas a formas diagonais com elementos não-negativos. Portanto, pela invariância dos autovalores sob tais transformações, temos que $\text{ch}_j(\mathbf{A} + \mathbf{B}) \geq \text{ch}_j(\mathbf{A})$, $\forall j = 1, \dots, p$. Logo, se $\phi(\mathbf{A})$ é escolhida como uma função de $\text{ch}_j(\mathbf{A})$, (2.18) é verdadeira para uma classe geral, sem sequer a necessidade de assumir-se que todas as matrizes sejam positivas-definidas. Em geral, o critério de variância generalizada não pertencerá a esta classe. O teste clássico de razão de verossimilhanças para igualdade de $\boldsymbol{\mu}_1$ e $\boldsymbol{\mu}_2$ se baseia na homogeneidade $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$; portanto, sempre que $\boldsymbol{\Sigma}$ for positiva-definida (p.d.),

$$\boldsymbol{\Sigma}^{-1}\Gamma(F, H) = \mathbf{I}_p + \frac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)', \quad (2.20)$$

e o critério de variância generalizada terá a mesma propriedade. Suponha agora um modelo qualitativo, em que X_i possa assumir uma entre C respostas qualitativas, a saber $1, \dots, C$ (\geq

2), com respectivas probabilidades π_1, \dots, π_C . Dessa forma, F é uma distribuição multinomial representada por $\boldsymbol{\pi}_F = (\pi_1^{(F)}, \pi_2^{(F)}, \dots, \pi_C^{(F)})'$, definida no simplex $\mathcal{S}_{C-1} = \{\boldsymbol{x} \in [0, 1]^C : \boldsymbol{x}'\mathbf{1} = \mathbf{1}\}$. Medidas quantitativas de tendência central e dispersão não têm sentido físico nesse contexto. No entanto, medidas de diversidade podem ser construídas em termos do vetor $\boldsymbol{\pi}_F$. Gini (1912) e Simpson (1949), este aparentemente sem conhecer o trabalho de Gini, propuseram a medida

$$I(\boldsymbol{\pi}_F) = 1 - \boldsymbol{\pi}'_F \boldsymbol{\pi}_F = \sum_{c=1}^C \pi_c^{(F)}(1 - \pi_c^{(F)}), \quad (2.21)$$

conhecida como índice de biodiversidade de Gini-Simpson. Sua interpretação é bem simples: se tomarmos $d(X_1, X_2) = I(X_1 \neq X_2)$, então

$$I(\boldsymbol{\pi}_F) = \mathbb{E}_F[d(X_1, X_2)] = P\{X_1 \neq X_2\} = 1 - \boldsymbol{\pi}'_F \boldsymbol{\pi}_F. \quad (2.22)$$

De forma similar, para a distribuição multinomial H associada a um vetor de probabilidades $\boldsymbol{\pi}_H$, define-se o índice de Gini-Simpson por $I(\boldsymbol{\pi}_H)$. Tomemos então

$$I(F, H) = P\{X \neq Y\} = 1 - P\{X = Y\} = 1 - \boldsymbol{\pi}'_F \boldsymbol{\pi}_H. \quad (2.23)$$

É fácil ver que

$$I(F, H) \geq \frac{1}{2} \{I(\boldsymbol{\pi}_F) + I(\boldsymbol{\pi}_H)\} \quad \forall F, H, \quad (2.24)$$

em que a igualdade vale se e somente se $F \stackrel{\mathcal{D}}{=} H$ (ou $\boldsymbol{\pi}_F = \boldsymbol{\pi}_H$), ambas definidas em um simplex comum \mathcal{S}_{C-1} . Logo, se estivermos interessados em testar a homogeneidade de $\boldsymbol{\pi}_F$ e $\boldsymbol{\pi}_H$ em contraponto a uma alternativa de diversidade, parece desejável incorporar as medidas $I(F, H)$, $I(\boldsymbol{\pi}_F)$ e $I(\boldsymbol{\pi}_H)$ à formulação do teste. Suas versões empíricas são U-estatísticas generalizadas, o que nos possibilita utilizar resultados assintóticos consolidados para conclusões estatísticas coerentes.

Consideremos a seguir um modelo qualitativo multidimensional em que $\mathbf{X}_i = (X_{i1}, \dots, X_{iK})'$ são K -vetores, cada X_{ik} assumindo uma entre $C (\geq 2)$ respostas qualitativas, $1, 2, \dots, C$. Nenhuma dificuldade técnica adicional se impõe se permitirmos que o número de categorias varie em $k (= 1, 2, \dots, K)$. Por simplicidade notacional, no entanto, mantemos C fixo.

Seja $\boldsymbol{c} = (c_1, \dots, c_k)'$, em que cada c_k assume um dos valores $1, \dots, C$. Seja ainda

$$\mathcal{C}_K = \{\boldsymbol{c} : c_k = 1, \dots, C; k = 1, \dots, K\} \quad (2.25)$$

de forma que a cardinalidade de \mathcal{C}_K é igual a C^K . Tomemos

$$\boldsymbol{\pi} = \{\pi(\mathbf{c}) : \mathbf{c} \in \mathcal{C}_K = 1\}; \quad \pi(\mathbf{c}) = P\{\mathbf{X} = \mathbf{c}\}. \quad (2.26)$$

A distribuição de \mathbf{X} , F , associa essa distribuição multidimensional ao vetor de probabilidades $\boldsymbol{\pi}$ que iremos denotar $\boldsymbol{\pi}_F$. Similarmente, para \mathbf{Y} , com uma distribuição H , teremos $\boldsymbol{\pi}_H$. Nossa intenção é comparar $\boldsymbol{\pi}_F$ e $\boldsymbol{\pi}_H$, com ênfase em seus aspectos de diversidade, especialmente quando K é muito grande, $K \gg n$.

Em muitas aplicações, existe uma ênfase nas medidas de diversidade marginais combinadas em uma única medida de diversidade, incorporando-se os k índices de Gini-Simpson de maneira análoga ao critério de traço no caso quantitativo. Motivado por isso, introduzimos a distância de Hamming como

$$\begin{aligned} d_H(\mathbf{X}_i, \mathbf{X}_j) &= K^{-1} \sum_{k=1}^K I(X_{ik} \neq X_{jk}) \\ &= K^{-1} \sum_{k=1}^K d(X_{ik}, X_{jk}), \end{aligned} \quad (2.27)$$

em que $d(x, y) = I(x \neq y)$ é a função-distância basal do índice de Gini-Simpson. Tomemos as marginais de X_{ik} , $F^{(k)}$, $k = 1, \dots, K$, com seus correspondentes índices de Gini-Simpson, $I(F^{(k)})$. Temos, então, a distância de Hamming populacional, para F , dada por \mathcal{H}_F :

$$\mathcal{H}_F = \mathbb{E}_F[d_H(\mathbf{X}_i, \mathbf{X}_j)] = K^{-1} \sum_{k=1}^K I(X_{ik} \neq X_{jk}) = K^{-1} \sum_{k=1}^K I(F^{(k)}). \quad (2.28)$$

Portanto, \mathcal{H}_F é uma medida de diversidade para F baseada na distância de Hamming em (2.27). Define-se \mathcal{H}_H de forma análoga. Mais ainda, definimos a distância de Hamming entre F e H por $\mathcal{H}(F, H) = \mathbb{E}[d_H(\mathbf{X}, \mathbf{Y})]$, em que \mathbf{X} e \mathbf{Y} têm respectivas distribuições F e H . Aplicando-se (2.24) para cada índice marginal, obtemos

$$\begin{aligned} \mathcal{H}(F, H) &= \frac{1}{K} \sum_{k=1}^K I(F^{(k)}, H^{(k)}) \geq \frac{1}{2} \left\{ \frac{1}{K} \sum_{k=1}^K I(F^{(k)}) + \frac{1}{K} \sum_{k=1}^K I(H^{(k)}) \right\} \\ &= \frac{1}{2} \{\mathcal{H}_F + \mathcal{H}_H\}, \end{aligned} \quad (2.29)$$

em que a igualdade vale se e somente se $F^{(k)} \stackrel{\mathcal{D}}{=} H^{(k)}$, $\forall k = 1, \dots, K$.

Para modelos quantitativos ou qualitativos (ou qualquer mistura), consideremos que os n vetores aleatórios independentes, $\mathbf{X}_1, \dots, \mathbf{X}_n$, seguem distribuição F . Defina-se o parâmetro

funcional $\delta(F)$. Seja $\delta(F^{(k)})$ um funcional análogo para a k -ésima marginal $F^{(k)}$, $k = 1, \dots, K$, $\boldsymbol{\delta}(\cdot) = (\delta(F^{(1)}), \dots, \delta(F^{(K)}))'$. Assumimos que $\delta(F)$ seja uma combinação convexa de $\boldsymbol{\delta}(\cdot)$. Por exemplo, podemos utilizar uma função linear convexa:

$$\delta(F) = \boldsymbol{\lambda}'\boldsymbol{\delta}(\cdot) : \boldsymbol{\lambda} \in \mathbb{R}^{+K} \text{ com } \boldsymbol{\lambda}'\boldsymbol{\lambda} = 1. \quad (2.30)$$

Assume-se que $\delta(F^{(k)})$ são parâmetros estimáveis ou funcionais regulares, no sentido de Hoeffding (1948). Levando-se em consideração que $\delta(F^{(k)})$ são funções-distância que, tipicamente, aplicam-se a um par de pontos, assumimos que existe um núcleo simétrico de grau 2 tal que

$$\delta(F^{(k)}) = \int \int \phi(x_1, x_2) dF^{(k)}(x_1) dF^{(k)}(x_2), \quad (2.31)$$

em que $\phi(\cdot)$ é não-negativo. Na realidade, em vista de (2.32), $\delta(F^{(k)}) = \delta(F^{(k)}, F^{(k)})$. Assumimos ainda que

$$\delta(F^{(k)}, H^{(k)}) = \int \int \phi(x_1, x_2) dF^{(k)}(x_1) dH^{(k)}(x_2) \quad (2.32)$$

satisfaz, para as distribuições F e H ,

$$\delta(F^{(k)}, H^{(k)}) \geq \frac{1}{2}(\delta(F^{(k)}) + \delta(H^{(k)})), \quad \forall k = 1, \dots, K. \quad (2.33)$$

Isso implica por sua vez que $\delta(F, H) \geq \{\delta(F) + \delta(H)\}/2 \forall F, H$. O teste proposto para a homogeneidade de H grupos em relação a suas medidas de diversidade se baseia em $\delta(\cdot, \cdot)$ e suas versões empíricas, que são todas U -estatísticas generalizadas.

2.2.2 Propriedades

Para a construção da estatística de teste, considere G (≥ 2) grupos independentes de amostras provenientes de distribuições F_1, \dots, F_G de respectivos tamanhos n_1, \dots, n_G em que cada F_g é K -dimensional e definida em um espaço de probabilidade comum. Considere um núcleo não-negativo $\phi(x, y)$ (que pode ser expresso como uma combinação linear convexa dos núcleos $\phi(x_k, y_k)$, $k = 1, \dots, K$). Denotemos os vetores das observações no g -ésimo grupo por $\mathbf{X}_{g1}, \dots, \mathbf{X}_{gn_g}$, $g = 1, \dots, G$, e tomemos

$$U_{n,gg} = \binom{n_g}{2}^{-1} \sum_{1 \leq i < j \leq n_g} \phi(\mathbf{X}_{gi}, \mathbf{X}_{gj}), \quad g = 1, \dots, G. \quad (2.34)$$

Note que $U_{n,gg}$, para $1 \leq g \leq G$, são U -estatísticas e, portanto, estimadores não viesados de $\delta(F_g)$. Similarmente, sejam as U -estatísticas generalizadas

$$U_{n,gg'} = \frac{1}{n_g n_{g'}} \sum_{i=1}^{n_g} \sum_{j=1}^{n_{g'}} \phi(\mathbf{X}_{gi}, \mathbf{X}_{g'j}), \quad 1 \leq g < g' \leq G. \quad (2.35)$$

Estas U-estatísticas são estimadores não viesados de $\delta(F_g, F_{g'})$ que satisfazem (2.33). Tomemos $n = n_1 + \dots + n_G$. Portanto, a U-estatística para a amostra combinada é

$$U_n = \sum_{g=1}^G \frac{n_g(n_g - 1)}{n(n-1)} U_{n,gg} + 2 \sum_{1 \leq g < g' \leq G} \frac{n_g n_{g'}}{n(n-1)} U_{n,gg'}, \quad (2.36)$$

que corresponde às componentes dentro dos grupos e entre os grupos. Note que

$$\begin{aligned} U_n &= \sum_{g=1}^G \frac{n_g}{n} U_{n,gg} + 2 \sum_{1 \leq g < g' \leq G} \frac{n_g n_{g'}}{n(n-1)} (2U_{n,gg'} - U_{n,gg} - U_{n,g'g'}) \\ &= W_n + B_n. \end{aligned} \quad (2.37)$$

O último termo de (2.36) é não negativo, enquanto em (2.37), B_n pode ser tanto positivo ou negativo. Sob a hipótese de homogeneidade dos G grupos, $\mathbb{E}B_n = 0$, mas $\mathbb{E}B_n \geq 0$ sob hipóteses alternativas.

Exemplo 2.2.1. Para ilustrar o que acontece quando procedemos com a decomposição (2.37), considere o conjunto de pontos do \mathbb{R}^2 , representados na Figura 2.1, como sendo vetores de tamanho $K = 2$.

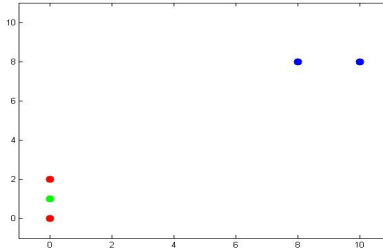


Figura 2.1: Exemplo: Pontos de \mathbb{R}^2 .

Os pontos são $X_1 = (0,0)$, $X_2 = (0,2)$, $Z = (0,1)$, $Y_1 = (8,8)$ e $Y_2 = (10,8)$. Seja $D = K^{-1} \sum_{t=1}^K (X_t - Y_t)^2$. Assim,

$$2D = \begin{pmatrix} & X_1 & X_2 & Z & Y_1 & Y_2 \\ X_1 & 0 & 4 & 1 & 128 & 164 \\ X_2 & & 0 & 1 & 100 & 128 \\ Z & & & 0 & 113 & 145 \\ Y_1 & & & & 0 & 4 \\ Y_2 & & & & & 0 \end{pmatrix}.$$

Portanto,

$$U_n = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} D_{ij} = \frac{1}{10} \{788\}.$$

Se colocarmos Z num grupo com X_1 e X_2 , temos

$$W_n^{(3,2)} = \frac{3}{5} \frac{1}{3} (1 + 1 + 4) + \frac{2}{5} (4) = \frac{28}{10}$$

e

$$B_n^{(3,2)} = \frac{3.2}{5.4} \left[2 \cdot \frac{1}{3.2} (2 \cdot 128 + 164 + 100 + 113 + 145) - \frac{1}{3} (1 + 1 + 4) - 4 \right] = \frac{760}{10}.$$

Se colocamos Z num grupo com Y_1 e Y_2 , então teremos

$$W_n^{(2,3)} = \frac{2}{5} (4) + \frac{3}{5} \frac{1}{3} (113 + 145 + 4) = \frac{540}{10},$$

$$B_n^{(2,3)} = \frac{3.2}{5.4} \left[2 \cdot \frac{1}{3.2} (2 \cdot 1 + 2 \cdot 128 + 164 + 100) - 4 - \frac{1}{3} (113 + 145 + 4) \right] = \frac{248}{10}.$$

Claramente, a medida dentro do grupo ($W_n^{(3,2)}$), para o grupo X_1, X_2, Z , é bem menor que para o grupo Y_1, Y_2, Z , ($W_n^{(2,3)}$), o que é esperado pois Z está mais próximo de X_1, X_2 do que Y_1, Y_2 . Enquanto B_n , que é a medida entre grupos, é maior quando consideramos o agrupamento X_1, X_2, Z contra o agrupamento Y_1, Y_2 (denotamos essa medida por $B_n^{(3,2)}$), do que quando consideramos o agrupamento X_1, X_2 contra o Y_1, Y_2, Z (a medida entre estes grupos é denotada por $B_n^{(2,3)}$). Isto mostra que os grupos X_1, X_2, Z e Y_1, Y_2 são mais dissimilares.

Se F_1, \dots, F_G não são todas iguais, $\mathbb{E}B_n > 0$. Além disso, sendo uma combinação linear de U-estatísticas generalizadas, B_n segue as leis de atração do TCL, isto é, $n^{1/2}(B_n - \mathbb{E}B_n)$ é assintoticamente normal. A situação se modifica se $F_1 \equiv \dots \equiv F_G$.

O interesse se concentra nessa situação não convencional. Primeiramente, $\mathbb{E}B_n = 0$ e B_n pode assumir valores negativos e positivos. Em segundo lugar, sendo essas U-estatísticas generalizadas estacionárias de ordem 1, elas têm, em geral, distribuições assintóticas não

normais.

Para G grupos homogêneos X_{gi} , $1 \leq i \leq n_g$, $1 \leq g \leq G$, de vetores aleatórios i.i.d. com distribuição F , a decomposição de Hoeffding tem propriedades interessantes:

- i) Reescrevendo-se \mathbf{X}_{gi} , $1 \leq i \leq n_g$, $1 \leq g \leq G$, como $\mathbf{X}_1, \dots, \mathbf{X}_n$, com a convenção de que os primeiros n_1 índices se relacionam ao grupo 1, os próximos n_2 ao grupo 2, etc., segue que

$$\begin{aligned}
B_n &= \frac{n_1 n_2}{n(n-1)} (2U_{n12} - U_{n1} - U_{n2}) \\
&= \frac{n_1 n_2}{n(n-1)} \left\{ \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \phi(X_{1i}, X_{2j}) - \frac{2}{n_1(n_1-1)} \sum_{1 \leq i < j \leq n_1} \phi(X_{1i}, X_{1j}) \right. \\
&\quad \left. - \frac{2}{n_2(n_2-1)} \sum_{1 \leq i < j \leq n_2} \phi(X_{2i}, X_{2j}) \right\} \\
&= \frac{n_1 n_2}{n(n-1)} \left\{ \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} [\phi_0 + \psi_1(X_{1i}) + \psi_1(X_{2j}) + \psi_2(X_{1i}, X_{2j})] - \right. \\
&\quad \left. - \frac{2}{n_1(n_1-1)} \sum_{i < j \leq n_1} [\phi_0 + \psi_1(X_{1i}) + \psi_1(X_{1j}) + \psi_2(X_{1i}, X_{1j})] \right. \\
&\quad \left. - \frac{2}{n_2(n_2-1)} \sum_{i < j \leq n_2} [\phi_0 + \psi_1(X_{2i}) + \psi_1(X_{2j}) + \psi_2(X_{2i}, X_{2j})] \right\} \\
&= \frac{n_1 n_2}{n(n-1)} \left\{ 2\phi_0 - 2 \frac{n_1}{n_1} \frac{n_1-1}{2} \phi_0 - (n_2-1)\phi_0 + \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} [\psi_1(X_{1i}) + \psi_1(X_{2j})] \right. \\
&\quad \left. - \frac{2}{n_1(n_1-1)} \sum_{1 < j \leq n_1} [\psi_1(X_{1i}) + \psi_1(X_{1j})] - \frac{2}{n_2(n_2-1)} \sum_{1 < j \leq n_2} [\psi_1(X_{2i}) + \psi_1(X_{2j})] \right. \\
&\quad \left. + \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \phi_2(X_{1i}, X_{2j}) - \frac{2}{n_1(n_1-1)} \sum_{1 < j \leq n_1} \psi_2(X_{1i}, X_{1j}) \right. \\
&\quad \left. - \frac{2}{n_2(n_2-1)} \sum_{1 < j \leq n_2} \psi_2(X_{2i}, X_{2j}) \right\} \\
&= \frac{n_1 n_2}{n(n-1)} \left\{ \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \psi_2(X_{1i}, X_{2j}) - \frac{2}{n_1(n_1-1)} \sum_{i < j \leq n_1} \psi_2(X_{1i}, X_{1j}) \right. \\
&\quad \left. - \frac{2}{n_2(n_2-1)} \sum_{1 \leq i < j \leq n_2} \psi_2(X_{2i}, X_{2j}) \right\}
\end{aligned}$$

e, portanto, podemos reescrever B_n equivalentemente como

$$\binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \eta_{nij} \psi_2(\mathbf{X}_1, \mathbf{X}_2), \quad (2.38)$$

em que

$$\eta_{nij} = \begin{cases} 1, & \text{se } i \text{ e } j \text{ pertencem a grupos diferentes;} \\ -\frac{(n-n_g)}{(n_g-1)}, & \text{se } i \text{ e } j \text{ vêm ambos do grupo } g, 1 \leq g \leq G. \end{cases} \quad (2.39)$$

Assim,

$$\begin{aligned} \sum_{1 \leq i < j \leq n} \eta_{nij} &= 0, \\ \sum_{1 \leq i < j \leq n} \eta_{nij}^2 &= \binom{n}{2} (G-1) \left\{ 1 + \frac{1}{n} \sum_{g=1}^G \frac{n-n_g}{(n_g-1)(G-1)} \right\}. \end{aligned} \quad (2.40)$$

ii) $\psi_1(\mathbf{X}_i)$ são v.a.'s i.i.d. centradas em 0;

iii) $\psi_2(\mathbf{X}_i, \mathbf{X}_j)$ são ortogonais e também centradas em 0, de forma que $\mathbb{E}\psi_2(\mathbf{X}_1, \mathbf{X}_2)\psi_2(\mathbf{X}_3, \mathbf{X}_4) = 0$ e

$$\mathbb{E}\psi_2(\mathbf{X}_1, \mathbf{X}_2)\psi_2(\mathbf{X}_1, \mathbf{X}_3) = 0. \quad (2.41)$$

Portanto, sob H_0 , que é a hipótese de homogeneidade dos grupos, B_n é equivalente a uma U-estatística ponderada degenerada que pode ser representada de maneira geral por

$$T_n = \sum_{1 \leq i < j \leq n} \eta_{nij} \phi(\mathbf{X}_i, \mathbf{X}_j), \quad n \geq 4, \quad (2.42)$$

em que $\phi(x, y)$ é um núcleo estacionário de ordem 1, centrado em zero e forma um sistema ortogonal, para o qual $\mathbb{E}[\phi(\mathbf{X}_1, \mathbf{X}_2)|\mathbf{X}_1] = \phi_1(\mathbf{X}_1) = 0$, $\mathbb{E}\phi(\mathbf{X}_1, \mathbf{X}_2)\phi(\mathbf{X}_3, \mathbf{X}_4) = 0$ e $\mathbb{E}\phi^2(\mathbf{X}_1, \mathbf{X}_2) < \infty$. Em geral, assume-se que, para a sequência i.i.d. $\mathbf{X}_1, \dots, \mathbf{X}_n$ em que cada vetor \mathbf{X}_i é K -dimensional, o núcleo $\phi(\cdot, \cdot)$ é escrito da forma

$$\phi(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{K} \sum_{\ell=1}^K \phi^*(X_{i\ell}, X_{j\ell}), \quad (2.43)$$

para algum núcleo estacionário de ordem 1, $\phi^*(\cdot, \cdot)$. Pinheiro et al. (2009) mostram que, sob convenientes condições de regularidade, vale a normalidade assintótica de T_n para $n \rightarrow \infty$

e/ou $K \rightarrow \infty$. Esse resultado vale sem que seja necessário controlar os crescimentos relativos de n e K e é apresentado no seguinte teorema.

Teorema 2.2.1. *Seja $\mathbf{X}_1, \dots, \mathbf{X}_n$ uma sequência de vetores aleatórios i.i.d. de dimensão $K \times 1$ com distribuição F . Seja $\phi(\cdot, \cdot)$ um núcleo de ordem 2, tal que*

$$\phi(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{K} \sum_{\ell=1}^K \phi^*(X_{i\ell}, X_{j\ell}) \quad (2.44)$$

para algum núcleo $\phi^*(\cdot, \cdot)$ estacionário de ordem 1. Seja T_n definida por

$$T_n = \sum_{1 \leq i, j \leq n} \eta_{nij} \phi(\mathbf{X}_i, \mathbf{X}_j), \quad n \geq 4. \quad (2.45)$$

Assuma que $\phi(\mathbf{x}, \mathbf{y})$ forma um sistema ortogonal, para o qual

$$\mathbb{E}[\phi_2(\mathbf{X}_1, \mathbf{X}_2) | \mathbf{X}_1] = \phi_1(\mathbf{X}_1) = 0 \quad (2.46)$$

e

$$\mathbb{E}\phi_2(\mathbf{X}_1, \mathbf{X}_2)\phi_2(\mathbf{X}_1, \mathbf{X}_3) = 0, \quad \mathbb{E}\phi^2(\mathbf{X}_1, \mathbf{X}_2) < \infty. \quad (2.47)$$

Suponha que $\{\eta_{nij}, 1 \leq i < j \leq n, n \geq 1\}$ é um arranjo triangular de variáveis aleatórias independentes $\{\mathbf{X}_1, \dots, \mathbf{X}_n, n \geq 1\}$ e

$$\sum_{1 \leq i < j \leq K} \eta_{nij}^2 - M_n = o_p(M_n) \quad \text{quando } n \rightarrow \infty. \quad (2.48)$$

em que $M_n = \sum_{1 \leq i < j \leq n} \eta_{nij}^2$. Suponha também que

$$\sum_{1 \leq \ell < m \leq K} \mathbb{E}\phi^*(X_{i\ell}, X_{j\ell})\phi^*(X_{im}, X_{jm}) = O(K) \quad \text{quando } K \rightarrow \infty. \quad (2.49)$$

Então,

$$(M_n V_n^*)^{-\frac{1}{2}} T_n \rightarrow N(0, 1) \quad \text{quando } n \rightarrow \infty \text{ e } K \rightarrow \infty, \quad (2.50)$$

em que $V_n^* = U_n^{(2,2)} - U_n^{(3)}$ e $U_n^{(2,2)}, U_n^{(3)}$ são definidas, respectivamente, por

$$U_n^{(3)} = \frac{1}{n(n-1)(n-2)} \sum_{1 \leq i \neq j \neq \ell \leq n} \phi(\mathbf{X}_i, \mathbf{X}_j)\phi(\mathbf{X}_i, \mathbf{X}_\ell) \quad (2.51)$$

e

$$U_n^{(2,2)} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \phi^2(\mathbf{X}_i, \mathbf{X}_j). \quad (2.52)$$

Este resultado é a base para toda a pesquisa desenvolvida neste trabalho e sua demonstração se baseia no Teorema 4.1 de Pinheiro et al. (2009).

Capítulo 3

Quase U-Estatísticas em Séries Temporais

Neste capítulo estão os principais resultados que obtivemos durante o desenvolvimento deste trabalho. Entre os principais, destacamos os Teoremas 3.1.1 e 3.3.1 que sintetizam todo o esforço realizado na busca de uma classe de funções que pudesse ser utilizada como núcleo da estatística de teste B_n , definida em (2.37), atendendo a todas as condições para que a normalidade assintótica seja válida. O primeiro resultado, embora não tenha muita utilidade prática, é de relevância teórica, pois direciona a busca dos núcleos para funções baseadas no periodograma e na função de autocovariância amostral. Essas funções têm propriedades assintóticas conhecidas e que independem dos momentos da série temporal, o que é fundamental para a obtenção dos resultados assintóticos.

Na Seção 3.1, apresentamos uma classe de funções-núcleos, para as quais vale a normalidade assintótica da estatística B_n . Na Seção 3.2, apresentamos alguns resultados assintóticos para o periodograma e na Seção 3.3 apresentamos métricas para séries temporais que podem ser utilizadas como núcleos da estatística de teste.

3.1 Uma Classe de Funções-Núcleos

Pinheiro et al. (2009) mostram que, sob condições de regularidade apropriadas, B_n é assintoticamente normal para $n \rightarrow \infty$ e/ou $K \rightarrow \infty$. Essas condições surgem, por exemplo, no estudo de sequências de DNA, para as quais a cardinalidade $K < n$, isto é, o número de sequências é maior que o tamanho da própria sequência.

Nosso primeiro objetivo é obter a normalidade assintótica da estatística de teste sob a hipótese de homogeneidade dos G grupos de séries temporais. De fato, em Pinheiro et al. (2009) nenhuma condição específica é imposta com relação à dependência em cada vetor, somente a condição de regularidade (2.49). Nesta seção, apresentamos uma classe de núcleos para a estatística T_n , definida em (2.45), e uma classe de séries temporais para as quais a condição (2.49) é satisfeita, garantindo a normalidade assintótica. Para apresentar esses resultados, consideremos as seguintes suposições:

(A1) $\mathbf{X}_i = \{X_{i1}, \dots, X_{iK}\}'$, $i = 1, \dots, n$, séries temporais lineares de tamanho K .

(A2) O p -ésimo momento de $X_{i,\ell}$, $1 \leq i \leq n$ e $1 \leq \ell \leq K$, é finito.

(A3) Os vetores $\mathbf{X}_1, \dots, \mathbf{X}_n$ são i.i.d.

Considere a classe das funções

$$\phi^{**}(X_{i\ell}, X_{j\ell}) = c(X_{i\ell} + X_{j\ell})^p - c\mathbb{E}(X_{i\ell} + X_{j\ell})^p, \quad (3.1)$$

em que $c \in \mathbb{R}$, $p \in \mathbb{N}$ e é finito. Seja $\psi_2^{**}(\cdot, \cdot)$ o termo de ordem 2 da decomposição de Hoeffding do núcleo $\phi^{**}(\cdot, \cdot)$,

$$T_n = \sum_{1 \leq i, j \leq n} \eta_{nij} \phi(\mathbf{X}_i, \mathbf{X}_j), \quad n \geq 4, \quad (3.2)$$

em que

$$\phi(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{K} \sum_{\ell=1}^K \psi_2^{**}(X_{i\ell}, X_{j\ell}), \quad (3.3)$$

e $\{\eta_{nij}, 1 \leq i < j \leq n, n \geq 1\}$ é um arranjo triangular de variáveis aleatórias independentes de $\{X_1, \dots, X_n, n \geq 1\}$, satisfazendo

$$\sum_{1 \leq i < j \leq K} \eta_{nij}^2 - M_n = o_p(M_n), \quad \text{quando } n \rightarrow \infty, \quad (3.4)$$

em que $M_n = \sum_{1 \leq i < j \leq n} \eta_{nij}^2$.

Teorema 3.1.1. *Seja T_n definida por (3.2). Então, sob (A1)-(A3),*

$$(M_n V_n^*)^{-\frac{1}{2}} T_n \rightarrow N(0, 1), \quad \text{quando } n \text{ e/ou } K \rightarrow \infty, \quad (3.5)$$

em que $V_n^* = U_n^{(2,2)} - U_n^{(3)}$,

$$U_n^{(3)} = \frac{1}{n(n-1)(n-2)} \sum_{1 \leq i \neq j \neq \ell \leq n} \phi(X_i, X_j) \phi(X_i, X_\ell) \quad (3.6)$$

e

$$U_n^{(2,2)} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \phi^2(X_i, X_j). \quad (3.7)$$

Para demonstrar esse teorema, primeiro mostramos os dois seguintes resultados.

Lema 3.1.1. *Sejam $\phi^{**}(X_{i\ell}, X_{j\ell})$ um núcleo centrado em zero e $\psi_2^{**}(X_{i\ell}, X_{j\ell})$ o termo de segunda ordem da decomposição de Hoeffding de $\phi^{**}(X_{i\ell}, X_{j\ell})$. Então, $\phi^*(X_{i\ell}, X_{j\ell}) = \psi_2^{**}(X_{i\ell}, X_{j\ell})$ é um núcleo de alguma U-estatística estacionária de ordem 1.*

Demonstração. Pela decomposição de Hoeffding, segue que

$$\psi_2^{**}(X_1, X_2) = \phi^{**}(X_1, X_2) - \{\psi_1^{**}(X_1) - \phi_0^{**}\} - \{\psi_1^{**}(X_2) - \phi_0^{**}\} - \phi_0^{**},$$

em que $\phi_0^{**} = \mathbb{E}\phi^{**}(X_1, X_2)$.

Seja $\phi^*(X_1, X_2) = \psi_2^{**}(X_1, X_2)$. Para que $\phi^*(\cdot, \cdot)$ seja estacionário de primeira ordem, $\psi_1^*(\cdot)$ deve ser nulo, com probabilidade 1 (c.p.1). Observe que

$$\phi_0^* = \mathbb{E}\phi^*(X_1, X_2) = \mathbb{E}\psi_2^{**}(X_1, X_2) = 0,$$

e

$$\begin{aligned} \psi_1^*(X_1) &= \mathbb{E}[\phi^*(X_1, X_2)|X_1] - \phi_0^* = \mathbb{E}[\psi_2^{**}(X_1, X_2)|X_1] \\ &= \mathbb{E}\{[\phi^{**}(X_1, X_2) - \psi_1^{**}(X_1) - \psi_1^{**}(X_2) + \phi_0^{**}]|X_1\} = \mathbb{E}[\phi^{**}(X_1, X_2)/X_1] \\ &\quad - \mathbb{E}\{\mathbb{E}[\phi^{**}(X_1, X_2)|X_1]/X_1\} - \mathbb{E}\{\mathbb{E}[\phi^{**}(X_1, X_2)/X_2]|X_1\} + \phi_0^{**} \\ &= \psi_1^{**}(X_1) - \psi_1^{**}(X_1) - c_1 + \phi_0^{**} = c_2, \end{aligned}$$

em que c_1 e c_2 são constantes, e

$$\mathbb{E}\psi_1^*(X_1) = \mathbb{E}\{\mathbb{E}[\phi^*(X_1, X_2)|X_1]\} = \mathbb{E}\phi^*(X_1, X_2) = 0.$$

Portanto, $\psi_1^*(X_1)$ é constante c.p.1 e tem média nula o que implica que $\psi_1^*(X_1) = 0$ c.p.1. Analogamente, $\psi_1^*(X_2) = 0$ c.p.1. □

Lema 3.1.2. *Seja $\phi^{**}(X_{i\ell}, X_{j\ell})$ um núcleo centrado em zero. Seja $\phi^*(X_{i\ell}, X_{j\ell})$ o termo de segunda ordem da decomposição de Hoeffding do núcleo $\phi^{**}(X_{i\ell}, X_{j\ell})$. Então*

$$\sum_{1 \leq \ell < m \leq K} \mathbb{E}\phi^{**}(X_{i\ell}, X_{j\ell})\phi^{**}(X_{im}, X_{jm}) = O(K) \text{ quando } K \rightarrow \infty, \quad (3.8)$$

se, e somente se,

$$\begin{aligned} \sum_{1 \leq \ell < m \leq K} \mathbb{E}\phi^*(X_{i\ell}, X_{j\ell})\phi^*(X_{im}, X_{jm}) &= \sum_{1 \leq \ell < m \leq K} \mathbb{E}\psi_2^{**}(X_{i\ell}, X_{j\ell})\psi_2^{**}(X_{im}, X_{jm}) \\ &= O(K) \text{ quando } K \rightarrow \infty. \end{aligned}$$

Demonstração. Supondo que $\phi^{**}(\cdot, \cdot)$ tem média zero e sendo $\phi^*(X_{i\ell}, X_{j\ell}) = \psi_2^{**}(X_{i\ell}, X_{j\ell})$,

segue que

$$\begin{aligned} \phi^*(X_{i\ell}, X_{j\ell})\phi^*(X_{im}, X_{jm}) &= \phi^{**}(X_{i\ell}, X_{j\ell})\phi^{**}(X_{im}, X_{jm}) - \phi^{**}(X_{i\ell}, X_{j\ell})\psi_1^{**}(X_{im}) \\ &\quad - \phi^{**}(X_{i\ell}, X_{j\ell})\psi_1^{**}(X_{jm}) - \phi^{**}(X_{im}, X_{jm})\psi_1^{**}(X_{i\ell}) - \phi^{**}(X_{im}, X_{jm})\psi_1^{**}(X_{j\ell}) \\ &\quad + \psi_1^{**}(X_{i\ell})\psi_1^{**}(X_{im}) + \psi_1^{**}(X_{j\ell})\psi_1^{**}(X_{jm}) + \psi_1^{**}(X_{j\ell})\psi_1^{**}(X_{im}) + \psi_1^{**}(X_{i\ell})\psi_1^{**}(X_{jm}). \end{aligned}$$

Observe que o valor esperado dos dois últimos termos da soma é zero. Portanto,

$$\begin{aligned} \mathbb{E}\phi^{**}(X_{im}, X_{jm})\psi_1^{**}(X_{i\ell}) &= \mathbb{E}\{\phi^{**}(X_{im}, X_{jm})\mathbb{E}[\phi^{**}(X_{i\ell}, X_{j\ell})|X_{i\ell}]\} \\ &= \mathbb{E}\{\mathbb{E}[\phi^{**}(X_{im}, X_{jm})\mathbb{E}(\phi^{**}(X_{i\ell}, X_{j\ell})|X_{i\ell})|\mathbf{X}_j]\} = \mathbb{E}\{\mathbb{E}[\phi^{**}(X_{i\ell}, X_{j\ell})|X_{i\ell}] \times \\ &\quad \mathbb{E}[\phi^{**}(X_{im}, X_{jm})|\mathbf{X}_j]\} = \mathbb{E}\{\mathbb{E}[\phi^{**}(X_{i\ell}, X_{j\ell})|X_{i\ell}]\} \mathbb{E}\{\mathbb{E}[\phi^{**}(X_{im}, X_{jm})|\mathbf{X}_j]\} \\ &= 0 \end{aligned}$$

e

$$\begin{aligned} \mathbb{E}[\psi_1^{**}(X_{i\ell})\psi_1^{**}(X_{im})] &= \mathbb{E}\{\mathbb{E}[\phi^{**}(X_{i\ell}, X_{j\ell})|X_{i\ell}]\mathbb{E}[\phi^{**}(X_{im}, X_{jm})|X_{im}]\} \\ &= \mathbb{E}\{\mathbb{E}[\mathbb{E}(\phi^{**}(X_{i\ell}, X_{j\ell})|X_{i\ell})\mathbb{E}(\phi^{**}(X_{im}, X_{jm})|X_{im})|\mathbf{X}_j]\} = \mathbb{E}\{\mathbb{E}[\phi^{**}(X_{i\ell}, X_{j\ell})|X_{i\ell}] \times \\ &\quad \mathbb{E}[\mathbb{E}(\phi^{**}(X_{im}, X_{jm})|X_{im})|\mathbf{X}_j]\} = \mathbb{E}\{\mathbb{E}[\phi^{**}(X_{i\ell}, X_{j\ell})|X_{i\ell}]\} \mathbb{E}[\mathbb{E}(\phi^{**}(X_{im}, X_{jm})|X_{im})|\mathbf{X}_j] \\ &= 0. \end{aligned}$$

Logo, para processos que podem ser linearizados,

$$\sum_{1 \leq \ell < m \leq K} \mathbb{E}\phi^{**}(X_{i\ell}, X_{j\ell})\phi^{**}(X_{im}, X_{jm}) = O(K), \text{ quando } K \rightarrow \infty,$$

se, e somente se,

$$\sum_{1 \leq \ell < m \leq K} \mathbb{E}\phi^*(X_{i\ell}, X_{j\ell})\phi^*(X_{im}, X_{jm}) = O(K), \text{ quando } K \rightarrow \infty.$$

□

Demonstração do Teorema 3.1.1. Pelos Lemas 3.1.1 e 3.1.2 sabemos que o termo de segunda ordem da decomposição de Hoeffding de um núcleo centrado em zero é um núcleo estacionário de ordem 1, e esse termo atende a condição de regularidade (3.8) se, e somente se, o próprio núcleo satisfaz a condição. Portanto, basta mostrar que o núcleo

$$\phi^{**}(X_{il}, X_{j\ell}) = (X_{il} + X_{j\ell})^p - \mathbb{E}(X_{il} + X_{j\ell})^p, \quad (3.9)$$

em que $p \in \mathbb{N}$ e é finito, atende a condição (3.8) para que tenhamos determinado uma classe de núcleos para os quais a normalidade assintótica de T_n ocorre.

Utilizando a expansão binomial

$$(x + y)^p = \sum_{j=0}^p \binom{p}{j} x^j y^{p-j}, \quad (3.10)$$

segue que

$$\phi^{**}(X_{il}, X_{j\ell}) = (X_{il} + X_{j\ell})^p - \mathbb{E}(X_{il} + X_{j\ell})^p = \sum_{k=0}^p \binom{p}{k} X_{il}^k X_{j\ell}^{p-k} - \sum_{k=0}^p \binom{p}{k} \mathbb{E}X_{il}^k \mathbb{E}X_{j\ell}^{p-k}.$$

Portanto,

$$\begin{aligned} & \phi^{**}(X_{il}, X_{j\ell}) \phi^{**}(X_{im}, X_{jm}) \\ &= \sum_{k_1=0}^p \binom{p}{k_1} \left(X_{il}^{k_1} X_{j\ell}^{p-k_1} - \mathbb{E}X_{il}^{k_1} \mathbb{E}X_{j\ell}^{p-k_1} \right) \sum_{k_2=0}^p \binom{p}{k_2} \left(X_{im}^{k_2} X_{jm}^{p-k_2} - \mathbb{E}X_{im}^{k_2} \mathbb{E}X_{jm}^{p-k_2} \right) \\ &= \sum_{k_1=0}^p \sum_{k_2=0}^p \binom{p}{k_1} \binom{p}{k_2} \left(X_{il}^{k_1} X_{j\ell}^{p-k_1} X_{im}^{k_2} X_{jm}^{p-k_2} - X_{il}^{k_1} X_{j\ell}^{p-k_1} \mathbb{E}X_{im}^{k_2} \mathbb{E}X_{jm}^{p-k_2} \right. \\ & \quad \left. - \mathbb{E}X_{il}^{k_1} \mathbb{E}X_{j\ell}^{p-k_1} X_{im}^{k_2} X_{jm}^{p-k_2} + \mathbb{E}X_{il}^{k_1} \mathbb{E}X_{j\ell}^{p-k_1} \mathbb{E}X_{im}^{k_2} \mathbb{E}X_{jm}^{p-k_2} \right) \end{aligned}$$

e

$$\begin{aligned} & \mathbb{E} \phi^{**}(X_{il}, X_{j\ell}) \phi^{**}(X_{im}, X_{jm}) \\ &= \sum_{k_1=0}^p \sum_{k_2=0}^p \binom{p}{k_1} \binom{p}{k_2} \left\{ \mathbb{E}X_{il}^{k_1} X_{im}^{k_2} \mathbb{E}X_{j\ell}^{p-k_1} X_{jm}^{p-k_2} - \mathbb{E}X_{il}^{k_1} \mathbb{E}X_{j\ell}^{p-k_1} \mathbb{E}X_{im}^{k_2} \mathbb{E}X_{jm}^{p-k_2} \right\}. \end{aligned}$$

Estamos supondo que as variáveis aleatórias são lineares, i.e., $X_{im} = \sum_{r=0}^{\infty} \gamma_{ir} \varepsilon_{i,m-r}$ e $X_{il} = \sum_{k=0}^{\infty} \gamma_{ik} \varepsilon_{i,\ell-k}$, em que $m > \ell$ e $\sum_{r=0}^{\infty} \gamma_{ir}^2 < \infty$. Logo,

$$\begin{aligned}
X_{im} &= \sum_{r=0}^{\infty} \gamma_{ir} \varepsilon_{i,m-r} = \gamma_{i0} \varepsilon_{i,m} + \gamma_{i1} \varepsilon_{i,m-1} + \cdots + \gamma_{i,m-\ell-1} \varepsilon_{i,m-\{m-\ell-1\}} + \gamma_{i,m-\ell} \varepsilon_{i,m-\{m-\ell\}} \\
&\quad + \gamma_{i,m-\ell+1} \varepsilon_{i,\ell-1} + \gamma_{i,m-\ell+2} \varepsilon_{i,\ell-2} + \cdots = \sum_{r_3=0}^{m-\ell-1} \gamma_{ir_3} \varepsilon_{i,m-r_3} + \sum_{r_2=0}^{\infty} \gamma_{i,m-\ell+r_2} \varepsilon_{i,\ell-r_2}. \quad (3.11)
\end{aligned}$$

Defina $\gamma_{ir_2}^* = \gamma_{i,m-\ell+r_2} / \alpha_i^{m-\ell}$, em que $\alpha_i \in (0, 1)$. Sabemos que

$$\sum_{r=0}^{\infty} \gamma_{ir}^2 < \infty, \quad (3.12)$$

logo $\sum_{r_2=0}^{\infty} \gamma_{i,m-\ell+r_2}^2 < \infty$ e, portanto,

$$\sum_{r_2=0}^{\infty} (\gamma_{ir_2}^*)^2 < \infty. \quad (3.13)$$

Como consequência imediata de (3.12) e (3.13), temos que

$$\sum_{r=0}^{\infty} |\gamma_{ir} \gamma_{ir}^*| < \infty. \quad (3.14)$$

Assim podemos escrever (3.11) da forma

$$X_{im} = \sum_{r_3=0}^{m-\ell-1} \gamma_{ir_3} \varepsilon_{i,m-r_3} + \alpha_i^{m-\ell} \sum_{r_2=0}^{\infty} \gamma_{ir_2}^* \varepsilon_{i,\ell-r_2}. \quad (3.15)$$

Supondo $k_1, k_2 \geq 1$,

$$\begin{aligned}
\mathbb{E} X_{i,\ell}^{k_1} X_{i,m}^{k_2} &= \mathbb{E} \left(\sum_{r_1=0}^{\infty} \gamma_{ir_1} \varepsilon_{i,\ell-r_1} \right)^{k_1} \left(\sum_{r_2=0}^{\infty} \gamma_{ir_2} \varepsilon_{i,m-r_2} \right)^{k_2} \\
&= \mathbb{E} \left(\sum_{r_1=0}^{\infty} \gamma_{ir_1} \varepsilon_{i,\ell-r_1} \right)^{k_1} \left(\alpha_i^{m-\ell} \sum_{r_2=0}^{\infty} \gamma_{ir_2}^* \varepsilon_{i,\ell-r_2} + \sum_{r_3=0}^{m-\ell-1} \gamma_{ir_3} \varepsilon_{i,m-r_3} \right)^{k_2} \\
&= \mathbb{E} \left(\sum_{r_1=0}^{\infty} \gamma_{ir_1} \varepsilon_{i,\ell-r_1} \right)^{k_1} \sum_{s=0}^{k_2} \binom{k_2}{s} \left(\sum_{r_3=0}^{m-\ell-1} \gamma_{ir_3} \varepsilon_{i,m-r_3} \right)^s \left(\alpha_i^{m-\ell} \sum_{r_2=0}^{\infty} \gamma_{ir_2}^* \varepsilon_{i,\ell-r_2} \right)^{k_2-s} \\
&= \mathbb{E} \sum_{r_{11}, r_{12}, \dots, r_{1k_1}=0}^{\infty} \gamma_{ir_{11}} \cdots \gamma_{ir_{1k_1}} \varepsilon_{i,\ell-r_{11}} \cdots \varepsilon_{i,\ell-r_{1k_1}} \sum_{s=0}^{k_2} \binom{k_2}{s} \sum_{r_{31}, r_{32}, \dots, r_{3s}=0}^{m-\ell-1} \gamma_{ir_{31}} \cdots \gamma_{ir_{3s}} \times
\end{aligned}$$

$$\begin{aligned}
& \varepsilon_{i,\ell-r_{31}} \cdots \varepsilon_{i,\ell-r_{3s}} \alpha_i^{(k_2-s)(m-\ell)} \sum_{r_{21}, r_{22}, \dots, r_{2k_2-s}=0}^{\infty} \gamma_{ir_{21}} \cdots \gamma_{ir_{2k_2-s}} \varepsilon_{i,\ell-r_{21}} \cdots \varepsilon_{i,\ell-r_{2k_2-s}} \\
= & \mathbb{E} \sum_{r_{11}, r_{12}, \dots, r_{1k_1}=0}^{\infty} \gamma_{ir_{11}} \cdots \gamma_{ir_{1k_1}} \varepsilon_{i,\ell-r_{11}} \cdots \varepsilon_{i,\ell-r_{1k_1}} \sum_{s=1}^{k_2-1} \binom{k_2}{s} \sum_{r_{31}, r_{32}, \dots, r_{3s}=0}^{m-\ell-1} \gamma_{ir_{31}} \cdots \gamma_{ir_{3s}} \times \\
& \varepsilon_{i,\ell-r_{31}} \cdots \varepsilon_{i,\ell-r_{3s}} \times \alpha_i^{(k_2-s)(m-\ell)} \sum_{r_{21}, r_{22}, \dots, r_{2k_2-s}=0}^{\infty} \gamma_{ir_{21}}^* \cdots \gamma_{ir_{2k_2-s}}^* \varepsilon_{i,\ell-r_{21}} \cdots \varepsilon_{i,\ell-r_{2k_2-s}} \\
+ & \sum_{r_{11}, r_{12}, \dots, r_{1k_1}=0}^{\infty} \gamma_{ir_{11}} \cdots \gamma_{ir_{1k_1}} \varepsilon_{i,\ell-r_{11}} \cdots \varepsilon_{i,\ell-r_{1k_1}} \binom{k_2}{0} \alpha_i^{k_2(m-\ell)} \sum_{r_{21}, r_{22}, \dots, r_{2k_2}=0}^{\infty} \gamma_{ir_{21}}^* \cdots \gamma_{ir_{2k_2}}^* \times \\
& \varepsilon_{i,\ell-r_{21}} \cdots \varepsilon_{i,\ell-r_{2k_2}} + \sum_{r_{11}, r_{12}, \dots, r_{1k_1}=0}^{\infty} \gamma_{ir_{11}} \cdots \gamma_{ir_{1k_1}} \times \varepsilon_{i,\ell-r_{11}} \cdots \varepsilon_{i,\ell-r_{1k_1}} \binom{k_2}{k_2} \times \\
& \sum_{r_{31}, r_{32}, \dots, r_{3k_2}=0}^{m-\ell-1} \gamma_{ir_{31}} \cdots \gamma_{ir_{3k_2}} \varepsilon_{i,m-r_{31}} \cdots \varepsilon_{i,m-r_{3k_2}} = C_1 \alpha_i^{m-\ell} + \sum_{r_{11}, r_{12}, \dots, r_{1k_1}=0}^{\infty} \\
& \sum_{r_{31}, r_{32}, \dots, r_{3k_2}=0}^{m-\ell-1} \gamma_{ir_{11}} \cdots \gamma_{ir_{1k_1}} \gamma_{ir_{31}} \cdots \gamma_{ir_{3k_2}} \mathbb{E} \varepsilon_{i,\ell-r_{11}} \cdots \varepsilon_{i,\ell-r_{1k_1}} \varepsilon_{i,m-r_{31}} \cdots \varepsilon_{i,m-r_{3k_2}}
\end{aligned}$$

e, repetindo o procedimento anterior,

$$\begin{aligned}
\mathbb{E} X_{i,\ell}^{k_1} \mathbb{E} X_{i,m}^{k_2} &= C_2 \alpha_i^{m-\ell} + \sum_{r_{11}, r_{12}, \dots, r_{1k_1}=0}^{\infty} \sum_{r_{31}, r_{32}, \dots, r_{3k_2}=0}^{m-\ell-1} \gamma_{ir_{11}} \cdots \gamma_{ir_{1k_1}} \gamma_{ir_{31}} \cdots \gamma_{ir_{3k_2}} \times \\
& \mathbb{E} \varepsilon_{i,\ell-r_{11}} \cdots \varepsilon_{i,\ell-r_{1k_1}} \mathbb{E} \varepsilon_{i,m-r_{31}} \cdots \varepsilon_{i,m-r_{3k_2}}.
\end{aligned}$$

Observe que $\ell - r_{1r} \neq m - r_{3s}$ para todo $1 \leq r \leq k_1$ e $1 \leq s \leq k_2$. De fato, se para algum $1 \leq r \leq k_1$ e $1 \leq s \leq k_2$ valer $\ell - r_{1r} = m - r_{3s}$, então $m - \ell + r_{1r} = r_{3s}$. Assim, como $r_{3s} \in \{0, \dots, m - \ell - 1\}$, segue que, se $r_{3s} = 0 \Rightarrow r_{1r} = -(m - \ell) < 0$ e $r_{3s} = m - \ell - 1$, então $r_{1r} = -1 < 0$. Logo,

$$\mathbb{E} \varepsilon_{i,\ell-r_{11}} \cdots \varepsilon_{i,\ell-r_{1k_1}} \varepsilon_{i,m-r_{31}} \cdots \varepsilon_{i,m-r_{3k_2}} = \mathbb{E} \varepsilon_{i,\ell-r_{11}} \cdots \varepsilon_{i,\ell-r_{1k_1}} \mathbb{E} \varepsilon_{i,m-r_{31}} \cdots \varepsilon_{i,m-r_{3k_2}}$$

e

$$\mathbb{E} X_{i,\ell}^{k_1} X_{i,m}^{k_2} - \mathbb{E} X_{i,\ell}^{k_1} \mathbb{E} X_{i,m}^{k_2} = C_3 \alpha_i^{m-\ell},$$

em que C_1, C_2 e $C_3 \in \mathbb{R}$. Ainda,

$$\begin{aligned}
& \mathbb{E}X_{i,\ell}^{k_1} X_{i,m}^{k_2} X_{j,\ell}^{k_3} X_{j,m}^{k_4} - \mathbb{E}X_{i,\ell}^{k_1} \mathbb{E}X_{i,m}^{k_2} \mathbb{E}X_{j,\ell}^{k_3} \mathbb{E}X_{j,m}^{k_4} = \mathbb{E}X_{i,\ell}^{k_1} X_{i,m}^{k_2} \mathbb{E}X_{j,\ell}^{k_3} X_{j,m}^{k_4} - \mathbb{E}X_{i,\ell}^{k_1} \mathbb{E}X_{i,m}^{k_2} \mathbb{E}X_{j,\ell}^{k_3} \mathbb{E}X_{j,m}^{k_4} \\
& = \left\{ C_{i1} \alpha_i^{m-\ell} + \sum_{r_{11}, \dots, r_{1k_1}=0}^{\infty} \sum_{r_{21}, \dots, r_{2k_2}=0}^{m-\ell-1} \gamma_{ir_{11}} \cdots \gamma_{ir_{1k_1}} \gamma_{ir_{21}} \cdots \gamma_{ir_{2k_2}} \mathbb{E}\varepsilon_{i,\ell-r_{11}} \cdots \varepsilon_{i,\ell-r_{1k_1}} \varepsilon_{i,m-r_{21}} \times \right. \\
& \quad \left. \cdots \varepsilon_{i,m-r_{2k_2}} \right\} \left\{ C_{j1} \alpha_j^{m-\ell} + \sum_{r_{31}, \dots, r_{3k_3}=0}^{\infty} \sum_{r_{41}, \dots, r_{4k_4}=0}^{m-\ell-1} \gamma_{jr_{31}} \cdots \gamma_{jr_{3k_3}} \gamma_{jr_{41}} \cdots \gamma_{jr_{4k_4}} \mathbb{E}\varepsilon_{j,\ell-r_{31}} \times \right. \\
& \quad \left. \cdots \varepsilon_{j,\ell-r_{3k_3}} \varepsilon_{j,m-r_{41}} \cdots \varepsilon_{j,m-r_{4k_4}} \right\} - \left\{ C_{i1}^* \alpha_i^{m-\ell} + \sum_{r_{11}, \dots, r_{1k_1}=0}^{\infty} \sum_{r_{21}, \dots, r_{2k_2}=0}^{m-\ell-1} \gamma_{ir_{11}} \cdots \gamma_{ir_{1k_1}} \times \right. \\
& \quad \left. \gamma_{ir_{21}} \cdots \gamma_{ir_{2k_2}} \mathbb{E}\varepsilon_{i,\ell-r_{11}} \cdots \varepsilon_{i,\ell-r_{1k_1}} \mathbb{E}\varepsilon_{i,m-r_{21}} \cdots \varepsilon_{i,m-r_{2k_2}} \right\} \left\{ C_{j1}^* \alpha_j^{m-\ell} + \sum_{r_{31}, \dots, r_{3k_3}=0}^{\infty} \right. \\
& \quad \left. \sum_{r_{41}, \dots, r_{4k_4}=0}^{m-\ell-1} \gamma_{jr_{31}} \cdots \gamma_{jr_{3k_3}} \gamma_{jr_{41}} \cdots \gamma_{jr_{4k_4}} \mathbb{E}\varepsilon_{j,\ell-r_{31}} \cdots \varepsilon_{j,\ell-r_{3k_3}} \mathbb{E}\varepsilon_{j,m-r_{41}} \cdots \varepsilon_{j,m-r_{4k_4}} \right\} \\
& = C_{i1} C_{j1} \alpha_i^{m-\ell} \alpha_j^{m-\ell} + C_{i1} \alpha_i^{m-\ell} C_4 + C_{j1} \alpha_j^{m-\ell} C_5 \\
& + \sum_{r_{11}, \dots, r_{1k_1}=0}^{\infty} \sum_{r_{21}, \dots, r_{2k_2}=0}^{m-\ell-1} \sum_{r_{31}, \dots, r_{3k_3}=0}^{\infty} \sum_{r_{41}, \dots, r_{4k_4}=0}^{m-\ell-1} \gamma_{ir_{11}} \cdots \gamma_{ir_{1k_1}} \gamma_{ir_{21}} \cdots \gamma_{ir_{2k_2}} \gamma_{jr_{31}} \cdots \gamma_{jr_{3k_3}} \times \\
& \quad \gamma_{jr_{41}} \cdots \gamma_{jr_{4k_4}} \mathbb{E}\varepsilon_{i,\ell-r_{11}} \cdots \varepsilon_{i,\ell-r_{1k_1}} \varepsilon_{i,m-r_{21}} \cdots \varepsilon_{i,m-r_{2k_2}} \mathbb{E}\varepsilon_{j,\ell-r_{31}} \cdots \varepsilon_{j,\ell-r_{3k_3}} \varepsilon_{j,m-r_{41}} \times \\
& \quad \cdots \varepsilon_{j,m-r_{4k_4}} \mathbb{E}\varepsilon_{i,\ell-r_{11}} \cdots \varepsilon_{i,\ell-r_{1k_1}} \mathbb{E}\varepsilon_{i,m-r_{21}} \cdots \varepsilon_{i,m-r_{2k_2}} \mathbb{E}\varepsilon_{j,\ell-r_{31}} \cdots \varepsilon_{j,\ell-r_{3k_3}} \times \\
& \quad \mathbb{E}\varepsilon_{j,m-r_{41}} \cdots \varepsilon_{j,m-r_{4k_4}} - C_{i1}^* C_{j1}^* \alpha_i^{m-\ell} \alpha_j^{m-\ell} - C_{i1}^* \alpha_i^{m-\ell} C_4^* - C_{j1}^* \alpha_j^{m-\ell} C_5^* \\
& = C_6 \alpha_i^{m-\ell} \alpha_j^{m-\ell} + C_7 \alpha_i^{m-\ell} + C_8 \alpha_j^{m-\ell},
\end{aligned}$$

em que $C_{i1}, C_{j1}, C_{i1}^*, C_{j1}^*, C_4, C_5, C_6, C_7$ e $C_8 \in \mathbb{R}$. O resultado também vale se $k_1 = 0$ e/ou $k_2 = 0$ ou, $k_3 = 0$ e/ou $k_4 = 0$. Portanto,

$$\mathbb{E}\phi^{**}(X_{i\ell}, X_{j\ell})\phi^{**}(X_{im}, X_{jm}) = C_6 \alpha_i^{m-\ell} \alpha_j^{m-\ell} + C_7 \alpha_i^{m-\ell} + C_8 \alpha_j^{m-\ell},$$

o que garante que a soma em ℓ e m , dada por (3.8), é no máximo $O(K)$. □

Para evitar qualquer tipo de confusão, é importante ressaltar que a classe de núcleos

que encontramos nesta seção atende às condições de regularidade para que a estatística T_n , definida por (2.45), seja assintoticamente normal. Embora o resultado teórico apresentado no Teorema 3.1.1 determine um avanço na busca por funções que possam ser utilizadas como núcleo da estatística de teste, este ainda é um pouco restritivo para fins práticos, uma vez que depende do conhecimento prévio de alguns momentos da série temporal. Uma maneira de contornar esse problema é utilizar núcleos baseados em ferramentas como o periodograma cujas propriedades assintóticas são bem conhecidas e independem do conhecimento dos momentos da série temporal.

3.2 Independência Assintótica de Periodogramas

Um dos principais objetivos da análise espectral é a estimação da densidade espectral e o estimador natural é o periodograma. Para obter esse estimador, considere um conjunto arbitrário de observações X_1, \dots, X_K (possivelmente a valor complexo) observadas nos tempos $1, \dots, K$. O vetor $\mathbf{X} := (X_1, \dots, X_K)'$ pertence ao espaço \mathbb{C}^K . Supondo que X_1, \dots, X_K são os valores de uma função periódica de período K , podemos esperar que cada X_t , para $t = 1, \dots, K$, pode ser expresso como uma combinação linear de harmônicos,

$$X_t = \frac{1}{\sqrt{K}} \sum_{-\pi < \omega_j \leq \pi} a_j e^{it\omega_j}, \quad t = 1, \dots, K, \quad (3.16)$$

em que as frequências $\omega_j = 2\pi j/K$ são os inteiros múltiplos das frequências fundamentais $2\pi/K$, as quais pertencem ao intervalo $(-\pi, \pi]$. As frequências $2\pi j/K$, $-\pi < \omega_j \leq \pi$, são as frequências de Fourier da série $\mathbf{X} = \{X_1, \dots, X_K\}'$. A representação (3.16) pode ser reescrita em forma de vetor como

$$\mathbf{X} = \sum_{j \in F_K} a_j \mathbf{e}_j, \quad (3.17)$$

em que

$$\mathbf{e}_j = \frac{1}{\sqrt{K}} (e^{i\omega_j}, e^{i2\omega_j}, \dots, e^{iK\omega_j})', \quad j \in F_K, \quad (3.18)$$

$$F_K = \{j \in \mathbb{Z} : -\pi < \omega_j \equiv 2\pi j/K \leq \pi\} = \{-\lfloor (K-1)/2 \rfloor, \dots, \lfloor K/2 \rfloor\}, \quad (3.19)$$

em que $\lfloor x \rfloor$ denota a parte inteira de x . Note que F_K contém K inteiros. A validade e unicidade da representação (3.17) e os valores dos coeficientes a_j são consequências simples do resultado de Brockwell e Davis (1991), o qual mostra que os vetores \mathbf{e}_j , $j \in F_K$, definidos em (3.18) constituem uma base ortonormal para \mathbb{C}^K . Uma consequência imediata desse resultado é que para qualquer $\mathbf{X} \in \mathbb{C}^K$,

$$\mathbf{X} = \sum_{j \in F_K} a_j \mathbf{e}_j, \quad (3.20)$$

em que

$$a_j = \langle \mathbf{X}, \mathbf{e}_j \rangle = \frac{1}{\sqrt{K}} \sum_{t=1}^K X_t e^{-it\omega_j}. \quad (3.21)$$

O valor do periodograma de \mathbf{X} na frequência $\omega_j = 2\pi j/K$, $j \in F_K$, denotado por $I_K(\omega_j)$, é definido em termos da transformada discreta de Fourier $\{a_j\}$ de \mathbf{X} ,

$$I_K(\omega_j) = |a_j|^2 = |\langle \mathbf{X}, \mathbf{e}_j \rangle|^2 = \frac{1}{K} \left| \sum_{t=1}^K X_t e^{-it\omega_j} \right|^2. \quad (3.22)$$

Analiticamente, o periodograma na frequência ω_j pode ser interpretado como o quadrado da norma da projeção de \mathbf{X} sobre \mathbf{e}_j , isto é, $I_K(\omega_j) = \|\text{proj}_{\mathbf{e}_j} \mathbf{X}\|^2$.

Na prática, existem muitas vantagens quando consideramos os vetores definidos em (3.18) como uma base para \mathbb{C}^K (usualmente referida como domínio da frequência). A principal delas é que os coeficientes a_j , que podem ser facilmente interpretados neste contexto, definem o periodograma. Entre as várias propriedades do periodograma, interessa-nos conhecer o comportamento assintótico de uma sequência de periodogramas, mais precisamente, com relação à dependência de uma sequência $I_K(\omega_1), \dots, I_K(\omega_K)$, quando K cresce. Observe que os coeficientes a_j são funções de ω_j e de K . Assim, denotaremos $a_j = S_K(\omega_j)$.

Para obter propriedades assintóticas de $S_K(\omega)$ e $I_K(\omega)$, é comum impor condições bastante restritivas, tais como estacionariedade e linearidade dos processos (Brockwell e Davis (1991)). No entanto, Shao e Wu (2007) demonstram que $S_K(\omega)$ e $I_K(\omega)$ são assintoticamente normais, impondo condições que envolvem somente momentos (condicionais), que podem ser facilmente verificadas para uma variedade de séries temporais não lineares. Para obter esses resultados, é necessário introduzir algumas notações. Seja ξ um vetor aleatório. Então, $\xi \in \mathcal{L}^p$ ($p > 0$) se $\|\xi\|_p := (\mathbb{E}|\xi|^p)^{1/p} < \infty$. Seja $\|\cdot\| = \|\cdot\|_2$. Para $\xi \in \mathcal{L}^1$, defina a projeção

$$\mathcal{P}_t \xi = \mathbb{E}[\xi | \mathcal{F}_t] - \mathbb{E}[\xi | \mathcal{F}_{t-1}], \quad t \in \mathbb{Z}, \quad (3.23)$$

em que $\mathcal{F}_t = (\dots, \varepsilon_{t-1}, \varepsilon_t)$ denota a σ -álgebra gerada por $\{\varepsilon_u\}_{u \leq t}$. Sejam $\omega_j = 2\pi j/K$, $1 \leq j \leq K$, as frequências de Fourier. Sob suaves condições, Shao e Wu (2007) demonstram que $S_K(\omega_j)$ são assintoticamente normais independentes. Denote a parte real e a parte imaginária de $S_K(\omega_j)/\sqrt{\pi K f(\omega_j)}$, respectivamente, por

$$Z_j = \frac{\sum_{t=1}^K X_t \cos(t\omega_j)}{\sqrt{\pi K f(\omega_j)}}, \quad Z_{j+m} = \frac{\sum_{t=1}^K X_t \sin(t\omega_j)}{\sqrt{\pi K f(\omega_j)}}, \quad j = 1, \dots, m, \quad (3.24)$$

em que $m = m_K := \lfloor (K-1)/2 \rfloor$. Seja $\Omega_p = \{c \in \mathbb{R}^p : |c| = 1\}$ a esfera unitária. Considere o conjunto $J = \{j_1, \dots, j_p\}$, com $1 \leq j_1 < \dots < j_p \leq 2m$, e defina $Z_J = (Z_{j_1}, \dots, Z_{j_p})'$. Seja $\Xi_{m,p} = \{J \subset \{1, \dots, 2m\} : \#J = p\}$, em que $\#$ é a cardinalidade de J .

Teorema 3.2.1. (Shao and Wu, 2007). *Assuma que $\mathbf{X}_t \in \mathcal{L}^2$,*

$$\kappa = \sum_{t=0}^{\infty} \|\mathcal{P}_0 X_t\| < \infty \quad (3.25)$$

e $f_* = \min_{\omega \in \mathbb{R}} f(\omega) > 0$. Então, para qualquer $p \in \mathbb{N}$ fixo, tem-se

$$\sup_{J \in \Xi} \sup_{c \in \Omega_p} \sup_{x \in \mathbb{R}} |\mathbb{P}(Z'_J c \leq x) - \Phi(x)| = o(1) \quad \text{quando } n \rightarrow \infty. \quad (3.26)$$

Corolário 3.2.1. (Do Teorema 3.2.1). *Seja $q \in \mathbb{N}$. Sob as condições do Teorema 3.2.1, tem-se*

$$\left\{ \frac{S_K(\omega_{\ell_k})}{\sqrt{\pi K f(\omega_{\ell_k})}}, 1 \leq j \leq q \right\} \Rightarrow \{Y_{2j-1} + iY_{2j}, 1 \leq j \leq q\} \quad (3.27)$$

para inteiros $1 \leq \ell_1 < \dots < \ell_q \leq m$, em que os índices l_j podem depender de K , e Y_k , $1 \leq k \leq 2q$, são normais padrão i.i.d.. Consequentemente, para $I_K^*(\omega) = I_K(\omega)/f(\omega)$,

$$\{I_K^*(\omega), 1 \leq j \leq q\} \Rightarrow \{E_j, 1 \leq j \leq q\}, \quad (3.28)$$

em que E_j são v.a.'s i.i.d. exponenciais padrão.

Nos exemplos seguintes, apresentamos alguns processos que atendem as condições do Teorema 3.2.1.

Exemplo 3.2.1. Suponha que $\{X_t\}_{t \in \mathbb{Z}}$ seja um processo linear, tal que $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, em que ε_t são i.i.d. com média zero, variância finita e $\sum_{j=0}^{\infty} a_j^2 < \infty$. Então,

$$\begin{aligned} \mathcal{P}_0 X_k &= \mathbb{E}[X_k | \mathcal{F}_0] - \mathbb{E}[X_k | \mathcal{F}_{-1}] = \mathbb{E} \left[\sum_{j=0}^{\infty} a_j \varepsilon_{k-j} \middle| \mathcal{F}_0 \right] - \mathbb{E} \left[\sum_{j=0}^{\infty} a_j \varepsilon_{k-j} \middle| \mathcal{F}_{-1} \right] \\ &= \mathbb{E} \sum_{j=0}^{k-1} a_j \varepsilon_{k-j} + a_k \varepsilon_0 + \sum_{j=k+1}^{\infty} a_j \varepsilon_{k-j} - \mathbb{E} \sum_{j=0}^{k-1} a_j \varepsilon_{k-j} - \mathbb{E} a_k \varepsilon_0 - \sum_{j=k+1}^{\infty} a_j \varepsilon_{k-j} \\ &= a_k \varepsilon_0 - \mathbb{E} a_k \varepsilon_0. \end{aligned}$$

Segue que

$$\|\mathcal{P}_0 X_k\| = (\mathbb{E}|a_k \varepsilon_0 - \mathbb{E} a_k \varepsilon_0|)^{1/2} = |a_k| \sigma_\varepsilon,$$

portanto, a condição (3.25) é equivalente a $\sum_{j=0}^{\infty} |a_j| < \infty$. Dentre os modelos lineares mais conhecidos destacamos o modelo ARMA(p, q), definido no Apêndice A. Pelo Teorema 4.4.2 de Brockwell e Davis (1991), a densidade espectral desse modelo é dada por

$$f_X(\lambda) = \frac{\sigma^2 |\theta(e^{-i\lambda})|}{2\pi |\phi(e^{-i\lambda})|},$$

em que $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ e $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ não têm nenhuma raiz no círculo unitário. Assim, $f_X(\lambda) > 0$ para todo $\lambda \in (-\pi, \pi]$. Logo, essa classe de modelos ARMA satisfaz as condições do Teorema 3.2.1.

Exemplo 3.2.2. Um dos primeiros modelos de séries temporais para heteroscedasticidade é o modelo ARCH. Na sua forma mais simples, um modelo ARCH expressa a série dos retornos como

$$X_t = \sigma_t \varepsilon_t, \quad (3.29)$$

em que usualmente se assume que $\varepsilon_t \sim \mathcal{N}(0, 1)$, i.i.d. com σ_t satisfazendo

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2. \quad (3.30)$$

Para garantir que $\sigma_t^2 > 0$ e para que o modelo seja bem definido, é suficiente que $\alpha_0 > 0$, $\alpha_i \geq 0$, para $i = 1, \dots, p$ e $\alpha_1 + \dots + \alpha_p < 1$. No caso mais simples, em que $p = 1$, obtém-se

$$\mathbb{E}X_t = \mathbb{E}\mathbb{E}[X_t | \mathcal{F}_{t-1}] = \mathbb{E}\mathbb{E}[\sigma_t \varepsilon_t | \mathcal{F}_{t-1}] = 0,$$

$$\gamma(h) = \mathbb{E}X_{t+h}X_t = \mathbb{E}\mathbb{E}[X_{t+h}X_t | \mathcal{F}_{t+h-1}] = \mathbb{E}X_t \mathbb{E}[\sigma_{t+h} \varepsilon_{t+h} | \mathcal{F}_{t+h-1}] = 0,$$

para $h > 0$. Além disso, $\{X_t\}_{t \in \mathbb{Z}}$ é causal mas não é linear. Segue que

$$\begin{aligned} \mathcal{P}_0 X_k &= \mathbb{E}[\sigma_k \varepsilon_k | \mathcal{F}_0] - \mathbb{E}[\sigma_k \varepsilon_k | \mathcal{F}_{-1}] = \mathbb{E}[\sigma_k | \mathcal{F}_0] \mathbb{E}[\varepsilon_k | \mathcal{F}_0] - \mathbb{E}[\sigma_k | \mathcal{F}_{-1}] \mathbb{E}[\varepsilon_k | \mathcal{F}_{-1}] \\ &= 0, \end{aligned}$$

portanto, a condição (3.25) é satisfeita. Além disso, como $\gamma(h) = 0$, para todo $h > 0$, $f(\lambda) = \sigma^2 > 0$ para todo $\lambda \in (-\pi, \pi]$. Logo, um processo ARCH(1) satisfaz as condições do Teorema 3.2.1.

Exemplo 3.2.3. Similarmente à extensão de um modelo AR para um modelo ARMA, pode-se estender a noção de um modelo ARCH para um modelo ARCH *generalizado*, (GARCH). Especificamente, um modelo GARCH pode ser expresso da forma

$$X_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1), \quad (3.31)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 + \sum_{j=1}^q \alpha_j X_{t-j}^2.$$

Condições sobre os α_i 's e β_j 's são necessárias para que a equação (3.31) seja bem definida. O caso mais simples, para o modelo GARCH(p, q), ocorre quando $p = q = 1$. Nesse caso, a condição para α_1 e β_1 é $\alpha_1 + \beta_1 < 1$. Segue que

$$\mathbb{E}X_t = \mathbb{E}\mathbb{E}[X_t|\mathcal{F}_{t-1}] = \mathbb{E}\mathbb{E}[\sigma_t \varepsilon_t|\mathcal{F}_{t-1}] = 0,$$

$$\gamma(h) = \mathbb{E}X_{t+h}X_t = \mathbb{E}\mathbb{E}[X_{t+h}X_t|\mathcal{F}_{t+h-1}] = \mathbb{E}X_t\mathbb{E}[\sigma_{t+h}\varepsilon_{t+h}|\mathcal{F}_{t+h-1}] = 0,$$

para $h > 0$.

Analogamente ao caso ARCH,

$$\begin{aligned} \mathcal{P}_0 X_k &= \mathbb{E}[\sigma_k \varepsilon_k | \mathcal{F}_0] - \mathbb{E}[\sigma_k \varepsilon_k | \mathcal{F}_{-1}] = \mathbb{E}[\sigma_k | \mathcal{F}_0] \mathbb{E}[\varepsilon_k | \mathcal{F}_0] - \mathbb{E}[\sigma_k | \mathcal{F}_{-1}] \mathbb{E}[\varepsilon_k | \mathcal{F}_{-1}] \\ &= 0, \end{aligned}$$

e portanto a condição (3.25) é satisfeita. Novamente, $\gamma(h) = 0$ para todo $h > 0$ e $f(\lambda) = \sigma^2 > 0$ para todo $\lambda \in (-\pi, \pi]$. Logo, um processo GARCH satisfaz as condições do Teorema 3.2.1.

Observação 3.2.1. *Existem várias generalizações do modelo GARCH, por exemplo, os modelos e-GARCH e t-GARCH, para os quais os cálculos acima são análogos e portanto também satisfazem as condições do Teorema 3.2.1.*

3.3 Núcleos Baseados em Métricas para Séries Temporais

As técnicas mais conhecidas para agrupar e classificar séries temporais se baseiam em métricas, que são utilizadas para medir a distância ou dissimilaridade entre séries temporais. Apesar de utilizar a nomenclatura métrica em muitas situações, as funções assim denominadas não necessariamente satisfazem todas as propriedades para que sejam uma métrica, no sentido matemático. No nosso contexto, uma métrica consiste em uma função, que a cada vetor associa um número real não-negativo. A métrica mais comum é a métrica Euclidiana clássica definida

por

$$d_{EUCL}(\mathbf{X}, \mathbf{Y}) = \frac{1}{K} \sum_{t=1}^K (X_t - Y_t)^2, \quad (3.32)$$

em que \mathbf{X} e \mathbf{Y} são séries temporais de tamanho K .

A métrica sugerida por Galeano e Peña (2000) é baseada na função de autocorrelação amostral (ACF). Suponha que $\hat{\boldsymbol{\rho}}_r = (\hat{\rho}_{r1}, \dots, \hat{\rho}_{rm})$ é o vetor dos coeficientes de autocorrelação estimados da série temporal r , para algum m , tal que, $\hat{\rho}_s \cong 0$, para $s > m$. Esta distância entre as séries temporais \mathbf{X} e \mathbf{Y} é definida por

$$d_{ACF}(\mathbf{X}, \mathbf{Y}) = (\hat{\boldsymbol{\rho}}_X - \hat{\boldsymbol{\rho}}_Y)' \boldsymbol{\Omega} (\hat{\boldsymbol{\rho}}_X - \hat{\boldsymbol{\rho}}_Y), \quad (3.33)$$

em que $\boldsymbol{\Omega}$ é uma matriz de ponderação. Quando $\boldsymbol{\Omega} = \mathbf{I}_K$ (matriz identidade), obtém-se a métrica Euclidiana entre os coeficientes de autocorrelação das séries \mathbf{X} e \mathbf{Y} . Quando \mathbf{X} e \mathbf{Y} têm mesma distribuição, $\text{cov}(\hat{\boldsymbol{\rho}}_X) = \text{cov}(\hat{\boldsymbol{\rho}}_Y) = \text{cov}(\hat{\boldsymbol{\rho}})$. Se $\boldsymbol{\Omega} = [\text{cov}(\hat{\boldsymbol{\rho}})]^{-1}$ é a inversa matriz de covariância das autocorrelações, obtém-se a distância de Mahalanobis entre as autocorrelações. É também comum o uso de pesos que diminuem com retardo das autocorrelações.

Caiado et al. (2006) propõem uma distância baseada no periodograma normalizado. Sejam $I_X(w_j)$ e $I_Y(w_j)$ periodogramas das séries temporais \mathbf{X} e \mathbf{Y} , respectivamente, na frequência $w_j = 2\pi j/K$, para $j = 1, \dots, \lfloor K/2 \rfloor$, variando de 0 a π . A distância entre \mathbf{X} e \mathbf{Y} pode ser definida por

$$d_P(\mathbf{X}, \mathbf{Y}) = \frac{1}{K} \sum_{j=1}^{\lfloor K/2 \rfloor} \left[I_X(w_j) - I_Y(w_j) \right]^2. \quad (3.34)$$

Para levar em consideração apenas a estrutura de correlação, é melhor utilizar o periodograma normalizado (ou periodograma reescalado) substituindo-se $I(w_j)$ em (3.34) por $\tilde{I}(w_j) = I(w_j)/\hat{\gamma}(0)$, em que $\hat{\gamma}(0)$ é a variância amostral da série temporal. Esta métrica fica definida por

$$d_{NP}(\mathbf{X}, \mathbf{Y}) = \frac{1}{K} \sum_{j=1}^{\lfloor K/2 \rfloor} \left[\tilde{I}_X(w_j) - \tilde{I}_Y(w_j) \right]^2. \quad (3.35)$$

Como a variância do periodograma ordinário é proporcional ao valor do espectro quadrado na correspondente frequência, Caiado et al. (2006) definem o logaritmo do periodograma normalizado:

$$d_{LNP}(\mathbf{X}, \mathbf{Y}) = \frac{1}{K} \sum_{j=1}^{\lfloor K/2 \rfloor} \left[\log \tilde{I}_X(w_j) - \log \tilde{I}_Y(w_j) \right]^2. \quad (3.36)$$

Outra medida útil no domínio do tempo, baseada na distância de Kullback-Leibler (Caiado et al. (2006)), é definida por

$$\mathbf{d}_{KLT D}(\mathbf{X}, \mathbf{Y}) = \frac{1}{K} \text{tr} (R_X R_Y^{-1}) - \log(|R_X|/|R_Y|) - n, \quad (3.37)$$

em que R_X e R_Y são matrizes $L \times L$ das autocorrelações das séries temporais \mathbf{X} e \mathbf{Y} , respectivamente, calculadas em L tempos sucessivos. A distância de Kullback-Leibler no domínio da frequência é assintoticamente equivalente a (3.37), mas muito mais fácil de calcular. Esta medida é definida por

$$\mathbf{d}_{KLF D}(\mathbf{X}, \mathbf{Y}) = \frac{1}{K} \sum_{j=1}^{\lfloor K/2 \rfloor} \left[\frac{\tilde{I}_X(w_j)}{\tilde{I}_Y(w_j)} - \log \frac{\tilde{I}_X(w_j)}{\tilde{I}_Y(w_j)} - 1 \right]^2. \quad (3.38)$$

3.3.1 Idéia Geral

Desde o princípio deste trabalho, um dos principais objetivos é encontrar uma classe de funções núcleos cujas distâncias funcionais baseadas nestes núcleos satisfaçam a desigualdade (2.33) e, além disso, atendam às condições do Teorema 2.2.1. Assim, elas podem ser utilizadas no nosso teste para homogeneidade de grupos, ou seja, podem ser um núcleo da estatística B_n em (2.37), para uma determinada classe de séries temporais. Nesta seção mostraremos que as métricas \mathbf{d}_P , \mathbf{d}_{LNP} e \mathbf{d}_{ACF} atendem todas as condições necessárias para que possam ser incorporadas a esta desejada classe de núcleos.

Sejam $\mathbf{X}_1, \dots, \mathbf{X}_n$ séries temporais de tamanho K . Uma métrica $\mathbf{d} : \mathbb{R}^{2K} \rightarrow \mathbb{R}^+$ pode ser utilizada como núcleo de B_n se atende as seguintes condições:

- i) Simetria: $\mathbf{d}(\mathbf{X}_i, \mathbf{X}_j) = \mathbf{d}(\mathbf{X}_j, \mathbf{X}_i)$. Se a métrica $\mathbf{d} : \mathbb{R}^{2K} \rightarrow \mathbb{R}$ pode ser escrita da forma

$$\mathbf{d}(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{K} \sum_{\ell=1}^{g(K)} d(Z_{i\ell}, Z_{j\ell}), \quad (3.39)$$

em que $\mathbf{d} : \mathbb{R}^2 \rightarrow \mathbb{R}$, $g(K) \rightarrow \infty$, quando $K \rightarrow \infty$, e $Z_{i\ell}$ relaciona a série temporal \mathbf{X}_i com o índice ℓ (exemplos: $X_{i\ell}$, $\hat{\rho}_{X_i}(\ell)$ e $I_{X_i}(\omega_\ell)$), então $d(x, y) = d(y, x)$ é uma condição suficiente;

- ii) Convexidade: A medida funcional entre grupos $\delta(F, H)$ baseada no núcleo $\mathbf{d}(\cdot, \cdot)$, definida em (2.33), deve ser maior ou igual à média das medidas dentro dos grupos $\delta(F)$ e $\delta(H)$ definida em (2.31), ou seja, a condição

$$\delta(F, H) \geq \{\delta(F) + \delta(H)\}/2, \quad \forall F, H, \quad (3.40)$$

deve ser satisfeita, em que a igualdade é verdadeira se, e somente se, $F \equiv H$. Como B_n é um estimador de $\{2\delta(F, H) - \delta(F) - \delta(H)\}$, a condição (3.40) garante que sob H_0 , $\mathbb{E}B_n = 0$ e $\mathbb{E}B_n \geq 0$ sob alternativas.

A maneira mais simples de verificar a convexidade é decompondo o núcleo $\mathbf{d}(\cdot, \cdot)$ no sentido de Hoeffding (1948) e verificar se

$$2\mathbf{d}_{0FH} - \mathbf{d}_{0FF} - \mathbf{d}_{0HH} \geq 0, \quad \forall F, H, \quad (3.41)$$

em que $\mathbf{d}_{0FH} = \mathbb{E}\mathbf{d}(\mathbf{X}_F, \mathbf{Y}_H)$ e \mathbf{X}_F indica que $\mathbf{X} \sim F$.

Exemplo 3.3.1. Mostraremos que a tradicional métrica Euclidiana \mathbf{d}_{EUCL} não satisfaz à condição ii). Suponha que numa amostra de tamanho $n = 12$ existam dois grupos de séries, em que $\mathbf{X}_1, \dots, \mathbf{X}_6$ correspondem ao grupo 1 com distribuição F , e $\mathbf{X}_7, \dots, \mathbf{X}_{12}$ pertencem ao grupo 2 com distribuição H , isto é, $n_1 = n_2 = 6$ e dentro do grupo as séries são i.i.d. temos

$$\begin{aligned} \mathbf{d}_{0FH} &= \mathbb{E}\mathbf{d}(\mathbf{X}_1, \mathbf{X}_7) = \frac{1}{2}\mathbb{E}(\mathbf{X}_1 - \mathbf{X}_7)^2 = \frac{1}{2}\mathbb{E}\mathbf{X}_1^2 - \mathbb{E}\mathbf{X}_1\mathbf{X}_7 + \frac{1}{2}\mathbb{E}\mathbf{X}_7^2 \\ &= \frac{1}{2}\sigma_{FF}^2 + \frac{1}{2}\sigma_{HH}^2, \end{aligned} \quad (3.42)$$

em que $\sigma_{FF}^2 = \mathbb{E}\mathbf{X}_1^2$ e $\sigma_{HH}^2 = \mathbb{E}\mathbf{X}_7^2$. Ainda,

$$\begin{aligned} \mathbf{d}_{0FF} &= \mathbb{E}\mathbf{d}(\mathbf{X}_1, \mathbf{X}_2) = \frac{1}{2}\mathbb{E}(\mathbf{X}_1 - \mathbf{X}_2)^2 = \frac{1}{2}\mathbb{E}\mathbf{X}_1^2 - \mathbb{E}\mathbf{X}_1\mathbf{X}_2 + \frac{1}{2}\mathbb{E}\mathbf{X}_2^2 \\ &= \sigma_{FF}^2, \end{aligned}$$

$$\begin{aligned} \mathbf{d}_{0HH} &= \mathbb{E}\mathbf{d}(\mathbf{X}_7, \mathbf{X}_8) = \frac{1}{2}\mathbb{E}(\mathbf{X}_7 - \mathbf{X}_8)^2 = \frac{1}{2}\mathbb{E}\mathbf{X}_7^2 - \mathbb{E}\mathbf{X}_7\mathbf{X}_8 + \frac{1}{2}\mathbb{E}\mathbf{X}_8^2 \\ &= \sigma_{HH}^2. \end{aligned}$$

Portanto, $2\mathbf{d}_{0FH} = \mathbf{d}_{0FF} + \mathbf{d}_{0HH}$, para quaisquer que sejam F e H . Logo, $\mathbb{E}B_n = 0$ sempre que as séries tiverem média zero.

Para uma métrica ser utilizada no teste para homogeneidade de grupos, ela deve atender também as condições do Teorema 2.2.1. Basicamente, se uma métrica atende as condições i) e ii) então, basta que o termo de segunda ordem da decomposição de Hoeffding desta métrica atenda a condição de regularidade (2.49). Embora a condição (2.49) dependa da distribuição das séries temporais, uma classe ampla de modelos pode ser incluída.

3.3.2 Estatística de Teste Baseada no Periodograma

Embora seja difícil encontrar, existem funções que podem ser utilizadas para testar a homogeneidade dos grupos de séries temporais. Um exemplo é a métrica baseada no periodograma definida em (3.34). Pela definição (3.22) podemos ver facilmente que $\mathbb{E}I_X(w_\ell) \geq 0$. Logo, definindo-se

$$\begin{aligned}\mu_{2_{IF}} &= \mathbb{E}I_{X_i}(w_\ell)^2 \text{ e } \mu_{IF} = \mathbb{E}I_{X_i}(w_\ell), \text{ para } \mathbf{X}_i \sim F; \\ \mu_{2_{IH}} &= \mathbb{E}I_{X_j}(w_\ell)^2 \text{ e } \mu_{IH} = \mathbb{E}I_{X_j}(w_\ell), \text{ para } \mathbf{X}_j \sim H,\end{aligned}$$

tem-se

$$\begin{aligned}\mathbf{d}_{0FH} &= \mu_{2_{IF}} + \mu_{2_{IH}} - 2\mu_{IF}\mu_{IH}; \\ \mathbf{d}_{0FF} &= 2\mu_{2_{IF}} - 2\mu_{IF}^2; \\ \mathbf{d}_{0HH} &= 2\mu_{2_{IH}} - 2\mu_{IH}^2,\end{aligned}$$

em que \mathbf{d}_{0FH} é calculada como em (3.42). Assim,

$$\mathbf{d}_{0FF} + \mathbf{d}_{0HH} - 2\mathbf{d}_{0FH} = -2\mu_{IF}^2 - 2\mu_{IH}^2 + 4\mu_{IF}\mu_{IH} = -2(\mu_{IF} - \mu_{IH})^2 \leq 0.$$

Portanto, $2\mathbf{d}_{0FH} > \mathbf{d}_{0FF} + \mathbf{d}_{0HH}$ para qualquer que sejam as distribuições $F \neq H$. Logo, $EB_n > 0$ mesmo que as séries tenham média zero. Além disso, a métrica baseada no periodograma é simétrica e, portanto, atende as condições i) e ii).

Resta mostrar que o periodograma satisfaz a condição de regularidade (2.49). Para a métrica \mathbf{d}_P em (3.34), segue que

$$d(Z_{i\ell}, Z_{j\ell}) = [I_{X_i}(w_\ell) - I_{X_j}(w_\ell)]^2, \quad (3.43)$$

em que $g(K) = \lfloor K/2 \rfloor$. O termo de segunda ordem da decomposição de Hoeffding do núcleo $\phi(X, Y) = (X - Y)^2$ é dado por

$$\begin{aligned}\psi_2(X, Y) &= (X - Y)^2 - \mathbb{E}[(X - Y)^2/X] - \mathbb{E}[(X - Y)^2/Y] + \mathbb{E}[(X - Y)^2] \\ &= (X - Y)^2 - I^2 + 2X\mathbb{E}Y - \mathbb{E}Y^2 - Y^2 + 2Y\mathbb{E}X - \mathbb{E}X^2 + \mathbb{E}X^2 - 2\mathbb{E}XY + \mathbb{E}Y^2 \\ &= X^2 + Y^2 - 2XY + 2X\mathbb{E}Y - X^2 - 2\mathbb{E}XY + 2X\mathbb{E}X - Y^2 \\ &= 2[-XY + X\mathbb{E}X + Y\mathbb{E}X - \mathbb{E}XY].\end{aligned} \quad (3.44)$$

Assim sendo, a condição de regularidade (2.49) resume-se a

$$\sum_{1 \leq \ell < m \leq g(K)} \mathbb{E}\phi^*(Z_{i\ell}, Z_{j\ell})\phi^*(Z_{im}, Z_{jm}) = O(K), \quad (3.45)$$

em que $Z_{i\ell} = I_{X_i}(w_\ell)$ e $\phi^*(\cdot, \cdot) \equiv \psi_2(\cdot, \cdot)$ dado em (3.44).

O periodograma atende a condição (3.45), e a demonstração é uma consequência do corolário 3.2.1, que garante, sob as condições impostas, que a sequência $\{I_X(\omega_1), \dots, I_X(\omega_m)\}$ é assintoticamente independente, mesmo que o tamanho da sequência m dependa de n .

Assim, o periodograma atende todas as condições necessárias para ser utilizado como um núcleo para a estatística de teste B_n .

3.3.3 Estatística de Teste Baseada na Autocorrelação Amostral

Outra métrica com propriedades interessantes é a métrica \mathbf{d}_{ACF} , sugerida por Galeano e Peña (2000), baseada na estimativa da função de autocorrelação e definida em (3.33). Mostraremos que essa métrica, quando $\mathbf{\Omega} = \mathbf{I}_K$, pode ser utilizada como núcleo da estatística de teste B_n . Iniciamos mostrando que \mathbf{d}_{ACF} atende as condições i) e ii). A condição de simetria i) é claramente satisfeita. Para a condição ii), basta verificar que

$$\mathbf{d}_{ACF}(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{L} \sum_{\ell=1}^L \left[\hat{\rho}_{X_i}(\ell) - \hat{\rho}_{X_j}(\ell) \right]^2 \quad (3.46)$$

e, portanto, $\mathbf{d}_{ACF}(\cdot, \cdot)$ pode ser reescrita da forma (3.39), em que $d(Z_{i\ell}, Z_{j\ell}) = (Z_{i\ell} - Z_{j\ell})^2$, $g(K) = L$ e $Z_{i\ell} = \hat{\rho}_{X_i}(\ell)$. Logo,

$$d(Z_{i\ell}, Z_{j\ell}) = \left[\hat{\rho}_{X_i}(\ell) - \hat{\rho}_{X_j}(\ell) \right]^2 = \hat{\rho}_{X_i}(\ell)^2 - 2\hat{\rho}_{X_i}(\ell)\hat{\rho}_{X_j}(\ell) + \hat{\rho}_{X_j}(\ell)^2$$

e

$$\begin{aligned} 2d_{0FH} - d_{0FF} - d_{0HH} &= 2\mathbb{E}\hat{\rho}_{iF}(\ell)^2 - 4\mathbb{E}\hat{\rho}_{iF}(\ell)\hat{\rho}_{jH}(\ell) + 2\mathbb{E}\hat{\rho}_{jH}(\ell)^2 \\ &\quad - \mathbb{E}\hat{\rho}_{iF}(\ell)^2 + 2\mathbb{E}\hat{\rho}_{iF}(\ell)\hat{\rho}_{jF}(\ell) - \mathbb{E}\hat{\rho}_{jF}(\ell)^2 - \mathbb{E}\hat{\rho}_{iH}(\ell)^2 + 2\mathbb{E}\hat{\rho}_{iH}(\ell)\hat{\rho}_{jH}(\ell) - \mathbb{E}\hat{\rho}_{jH}(\ell)^2 \\ &= -4\mathbb{E}\hat{\rho}_{iF}(\ell)\hat{\rho}_{jH}(\ell) + 2\mathbb{E}\hat{\rho}_{iH}(\ell)\hat{\rho}_{jH}(\ell) + 2\mathbb{E}\hat{\rho}_{iF}(\ell)\hat{\rho}_{jF}(\ell) \\ &= 2[\mathbb{E}\hat{\rho}_{iF}(\ell) - \mathbb{E}\hat{\rho}_{iH}(\ell)]^2 \geq 0, \end{aligned}$$

em que $\hat{\rho}_{iF}(\ell)$ é o estimador da autocorrelação da série \mathbf{X}_i no “lag” ℓ sendo que $\mathbf{X}_i \sim F$, $\hat{\rho}_{jH}(\ell)$ é o estimador da autocorrelação da série \mathbf{X}_j no “lag” ℓ sendo que $\mathbf{X}_j \sim H$ e é igualdade vale se, e somente se, $F \equiv H$. É importante ressaltar que \mathbf{X}_i e \mathbf{X}_j são independentes.

Mostramos que \mathbf{d}_{ACF} atende as condições i) e ii). Resta mostrar que a condição de regularidade (3.45) é satisfeita para $Z_{i\ell} = \hat{\rho}_{X_i}(\ell)$. O termo de segunda ordem da decomposição

de Hoeffding do núcleo \mathbf{d}_{ACF} é dado em (3.44) e, portanto,

$$\begin{aligned} \psi_2 \left(\hat{\rho}_{X_i}(\ell), \hat{\rho}_{X_j}(\ell) \right) &= \phi^* \left(\hat{\rho}_{X_i}(\ell), \hat{\rho}_{X_j}(\ell) \right) \\ &= -2\hat{\rho}_{X_i}(\ell)\hat{\rho}_{X_j}(\ell) + 2\hat{\rho}_{X_i}(\ell)\mathbb{E}\hat{\rho}_{X_j}(\ell) + 2\hat{\rho}_{X_j}(\ell)\mathbb{E}\hat{\rho}_{X_i}(\ell) - 2\mathbb{E}\hat{\rho}_{X_i}(\ell)\hat{\rho}_{X_j}(\ell) \\ &= 2\mathbb{E}\hat{\rho}_{X_i}(\ell) \left[\hat{\rho}_{X_j}(\ell) - \mathbb{E}\hat{\rho}_{X_j}(\ell) \right] - 2\hat{\rho}_{X_i}(\ell) \left[\hat{\rho}_{X_j}(\ell) - \mathbb{E}\hat{\rho}_{X_j}(\ell) \right] \\ &= -2 \left[\hat{\rho}_{X_i}(\ell) - \mathbb{E}\hat{\rho}_{X_i}(\ell) \right] \left[\hat{\rho}_{X_j}(\ell) - \mathbb{E}\hat{\rho}_{X_j}(\ell) \right]. \end{aligned}$$

Segue que

$$\begin{aligned} &\mathbb{E}\phi^* \left(\hat{\rho}_{X_i}(\ell), \hat{\rho}_{X_j}(\ell) \right) \phi^* \left(\hat{\rho}_{X_i}(m), \hat{\rho}_{X_j}(m) \right) \\ &= 4\mathbb{E} \left(\hat{\rho}_{X_i}(\ell) - \mathbb{E}\hat{\rho}_{X_i}(\ell) \right) \left(\hat{\rho}_{X_i}(m) - \mathbb{E}\hat{\rho}_{X_i}(m) \right) \mathbb{E} \left(\hat{\rho}_{X_j}(\ell) - \mathbb{E}\hat{\rho}_{X_j}(\ell) \right) \left(\hat{\rho}_{X_j}(m) - \mathbb{E}\hat{\rho}_{X_j}(m) \right) \\ &= 4\text{Cov} \left(\hat{\rho}_{X_i}(\ell), \hat{\rho}_{X_i}(m) \right) \text{Cov} \left(\hat{\rho}_{X_j}(\ell), \hat{\rho}_{X_j}(m) \right). \end{aligned} \quad (3.47)$$

Pelo Teorema 7.2.1 de Brockwell e Davis (1991), segue que, para todo processo linear estacionário, a covariância $\text{Cov} \left(\hat{\rho}_{X_i}(\ell), \hat{\rho}_{X_i}(m) \right) = O(K^{-1})$. Portanto, (3.47) é $O(K^{-2})$ e isto é mais do que suficiente para garantir a condição (3.45).

Acabamos de demonstrar que para qualquer processo linear estacionário, a métrica \mathbf{d}_{ACF} pode ser utilizada como núcleo da estatística de teste B_n , com normalidade assintótica garantida.

3.3.4 Estatística de Teste Baseada no Log do Periodograma Normalizado

Nesta subseção, mostraremos que a métrica \mathbf{d}_{LNP} , apresentada por Caiado et al. (2006) e definida em (3.36), pode ser utilizada como um núcleo para a estatística B_n , ou seja, mostraremos que essa métrica atende as condições i), ii) e a condição de regularidade (3.45). Seguindo a notação em (3.39), temos que $\mathbf{d}(\mathbf{X}_i, \mathbf{X}_j) = K^{-1} \sum_{\ell=1}^{\lfloor K/2 \rfloor} d(Z_{i\ell}, X_{j\ell})$, em que $Z_{i\ell} = \tilde{I}_X(\omega_\ell)$ e $d(x, y) = (x - y)^2$.

As condições i) e ii) seguem, naturalmente, da demonstração feita para o periodograma na subseção 3.3.2. Para a condição de regularidade, consideremos a métrica baseada somente no periodograma normalizado

$$\mathbf{d}_{NP}(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{K} \sum_{\ell=1}^{\lfloor K/2 \rfloor} \left(\tilde{I}_{X_i}(\omega_\ell) - \tilde{I}_{X_j}(\omega_\ell) \right)^2, \quad (3.48)$$

em que

$$\tilde{I}_{X_i}(\omega_\ell) = \frac{I_{X_i}(\omega_\ell)}{\hat{\gamma}_{X_i}(0)}. \quad (3.49)$$

Observe que

$$\begin{aligned}
\mathbf{d}_{NP}(\mathbf{X}_i, \mathbf{X}_j) &= \frac{1}{K} \sum_{\ell=1}^{\lfloor K/2 \rfloor} \left[\tilde{I}_{X_i}(\omega_\ell) - \tilde{I}_{X_j}(\omega_\ell) \right]^2 \\
&= \frac{1}{K} \sum_{\ell=1}^{\lfloor K/2 \rfloor} \left[\left(2 + 4 \sum_{m=1}^{K-1} \hat{\rho}_{X_i}(m) \cos(m\omega_\ell) \right) - \left(2 + 4 \sum_{m=1}^{K-1} \hat{\rho}_{X_j}(m) \cos(m\omega_\ell) \right) \right]^2 \\
&= \frac{1}{K} \sum_{\ell=1}^{\lfloor K/2 \rfloor} \left[4 \sum_{m=1}^{K-1} \cos(m\omega_\ell) \left[\hat{\rho}_{X_i}(m) - \hat{\rho}_{X_j}(m) \right] \right]^2.
\end{aligned}$$

Pelas propriedades de ortogonalidade da função cosseno,

$$\begin{aligned}
\sum_{\ell=1}^{\lfloor K/2 \rfloor} \cos(m\omega_\ell) &= \frac{K}{4}, \text{ se } K \text{ é par e} \\
\sum_{\ell=1}^{\lfloor K/2 \rfloor} \cos(m\omega_\ell) \cos(s\omega_\ell) &= 0, \text{ para } m \neq s,
\end{aligned}$$

(Caiado et al. (2006)), segue que

$$\begin{aligned}
\mathbf{d}_{NP}(\mathbf{X}_i, \mathbf{X}_j) &= \frac{16}{K} \left\{ \frac{K}{4} \left[\hat{\rho}_{X_i}(1) - \hat{\rho}_{X_j}(1) \right]^2 + \cdots + \frac{K}{4} \left[\hat{\rho}_{X_i}(K-1) - \hat{\rho}_{X_j}(K-1) \right]^2 \right\} \\
&= 2 \sum_{\ell=1}^{K-1} \left[\hat{\rho}_{X_i}(\ell) - \hat{\rho}_{X_j}(\ell) \right]^2 \\
&= 2K \mathbf{d}_{ACF}(\mathbf{X}_i, \mathbf{X}_j).
\end{aligned}$$

Portanto, para $L = K - 1$ em (3.46), e $g(K) = \lfloor K/2 \rfloor$ para a métrica \mathbf{d}_{NP} , obtemos a relação $\mathbf{d}_{NP}(\mathbf{X}_i, \mathbf{X}_j) = 2K \mathbf{d}_{ACF}(\mathbf{X}_i, \mathbf{X}_j)$. A condição de regularidade (2.49) é necessária para garantir que a ordem da covariância da soma destas métricas não seja maior que a ordem da variância. A métrica \mathbf{d}_{ACF} satisfaz tal condição para a classe de séries temporais lineares e estacionárias. Logo, a métrica baseada no periodograma normalizado \mathbf{d}_{NP} também atende tal condição, uma vez que o fator K altera a ordem da covariância na mesma proporção que altera a ordem da variância.

Mostramos que, para $\phi^*(\cdot, \cdot) = \psi_2(\cdot, \cdot)$, dado em (3.44) e $\tilde{I}_{X_i}(\omega_\ell)$, dado em (3.49),

$$\sum_{1 \leq l < m \leq \lfloor K/2 \rfloor} \mathbb{E} \phi^* \left(\tilde{I}_{X_i}(\omega_\ell), \tilde{I}_{X_j}(\omega_\ell) \right) \phi^* \left(\tilde{I}_{X_i}(\omega_m), \tilde{I}_{X_j}(\omega_m) \right), \quad (3.50)$$

é de ordem menor que a soma das variâncias, quando $K \rightarrow \infty$.

Queremos mostrar que \mathbf{d}_{LNP} atende as condições para normalidade assintótica, ou seja, que o termo de segunda ordem da decomposição de Hoeffding satisfaz a condição de regula-

ridade (2.49) para $\log \tilde{I}_{X_i}(\omega)$. Outra maneira de incluir a métrica \mathbf{d}_{LNP} na classe de funções núcleos cuja normalidade assintótica de B_n ocorre é mostrando que \mathbf{d}_{LNP} se relaciona com alguma outra métrica com propriedades já conhecidas, como foi o caso $\mathbf{d}_{NP} = K\mathbf{d}_{ACF}$.

Embora não tenhamos conseguido demonstrar formalmente que a métrica \mathbf{d}_{LNP} atende a condição (2.49), ela é a que apresenta melhor desempenho nos estudos de simulação do Capítulo 4. Portanto, \mathbf{d}_{LNP} deve atender a condição (2.49), mas a demonstração formal fica em aberto.

O que demonstramos nesta seção pode ser resumido no seguinte resultado.

Teorema 3.3.1. *Sejam $\mathbf{X}_1, \dots, \mathbf{X}_n$ séries temporais i.i.d., estacionárias e*

$$T_n = \sum_{1 \leq i < j \leq n} \eta_{nij} \phi(\mathbf{X}_i, \mathbf{X}_j), \quad (3.51)$$

em que $\phi(\mathbf{X}_i, \mathbf{X}_j)$ é uma das métricas definidas nesta seção (\mathbf{d}_{NP} , \mathbf{d}_{LNP} e \mathbf{d}_{ACF}). Defina $M_n = \sum_{1 \leq i < j \leq n} \eta_{nij}^2$, em que η_{nij} é definido em (4.18). Então,

$$(M_n V_n^*)^{-\frac{1}{2}} T_n \rightarrow N(0, 1) \text{ quando } n \rightarrow \infty \text{ e/ou } K \rightarrow \infty, \quad (3.52)$$

em que $V_n^* = U_n^{(2,2)} - U_n^{(3)}$,

$$U_n^{(3)} = \frac{1}{n(n-1)(n-2)} \sum_{1 \leq i \neq j \neq \ell \leq n} \phi(\mathbf{X}_i, \mathbf{X}_j) \phi(\mathbf{X}_i, \mathbf{X}_\ell)$$

e

$$U_n^{(2,2)} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \phi^2(\mathbf{X}_i, \mathbf{X}_j).$$

No Teorema 3.3.1, encontramos uma classe de funções núcleos que, para uma determinada classe de séries temporais, atende todas as condições do Teorema 2.2.1. Essa classe é definida de maneira geral na seguinte definição.

Definição 3.3.1. Sejam \mathbf{X}_i e \mathbf{X}_j séries temporais estacionárias. Considere a função $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ com a propriedade

$$g(K) = O(K) \text{ quando } K \rightarrow \infty. \quad (3.53)$$

Defina,

$$\phi(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{K} \sum_{m=1}^{g(K)} \phi(Z_{im}, Z_{jm}), \quad (3.54)$$

em que $\phi(\cdot, \cdot) : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ é dada por $\phi(x, y) = (x - y)^2$ e Z_{im} associa a série temporal X_i à posição m , $m = 1, \dots, g(K)$. As possibilidades para a variável Z_{im} são:

- a) $Z_{im} = \tilde{I}_{X_i}(\omega_m)$, com $g(K) = \lfloor \frac{K}{2} \rfloor$;
- b) $Z_{im} = \log \tilde{I}_{X_i}(\omega_m)$ com, $g(K) = \lfloor \frac{K}{2} \rfloor$;
- c) e $Z_{im} = \hat{\rho}_{X_i}(m)$, em que $g(K) < \frac{K}{4}$ é recomendado.

Capítulo 4

Classificação e Agrupamento de Séries Temporais

Classificação e agrupamento de séries temporais são problemas bastante explorados na literatura atual. Muitas técnicas são apresentadas para resolver esses problemas. No entanto, os procedimentos são específicos, aplicáveis somente a uma determinada classe de séries temporais. Maharaj (2000) apresenta um procedimento para agrupar séries temporais estacionárias, a partir das estimativas dos parâmetros autorregressivos. Huang et al. (2004) propõem um método para discriminação e agrupamento de séries temporais não estacionárias. Discriminação entre séries temporais estacionárias e não estacionárias pode ser feita através do procedimento apresentado por Caiado et al. (2006). Para modelos heteroscedásticos, Otranto (2008) apresenta uma técnica baseada em métricas autorregressivas, as quais são utilizadas para medir distância entre processos GARCH.

Propomos, neste capítulo, um procedimento que permite agrupar séries temporais. A idéia basal é utilizar a mais forte similaridade entre séries dentro dos grupos para agrupá-las e classificá-las. O procedimento é baseado nos resultados assintóticos para a estatística B_n , e, portanto, depende da escolha de umas das métricas \mathbf{d}_P , \mathbf{d}_{NP} , \mathbf{d}_{LNP} ou \mathbf{d}_{ACF} . Os resultados assintóticos valem para um classe ampla de séries temporais estacionárias lineares e não lineares. Na Seção 4.1, apresentamos um método de classificação e agrupamento. Na Seção 4.2, realizamos estudos de simulação para comparação do desempenho das estatísticas propostas sob várias métricas. Na Seção 4.3, realizamos um estudo de simulação para testar o desempenho do método de classificação e agrupamento. Uma aplicação a dados sismológicos é feita na Seção 4.4.

4.1 Método de Classificação e Agrupamento

Nesta seção, desenvolvemos um método para classificação e agrupamento de séries temporais baseado nos resultados assintóticos obtidos para a estatística B_n , sob a hipótese de homogeneidade dos grupos, e baseadas em resultados assintóticos para qualquer mistura de duas distribuições dentro dos grupos. Esses resultados são apresentados na sequência.

4.1.1 Comportamento de B_n com Heterogeneidade Dentro dos Grupos

Uma condição necessária até o momento era que, dentro do grupo g , os elementos deveriam ter a mesma distribuição, isto é, $X_{1g}, X_{2g}, \dots, X_{n_gg}$ têm distribuição F_g . Dessa forma, desenvolvemos um teste para a homogeneidade dos grupos, cuja hipótese nula de interesse é $H_0 : F_g \equiv F_{g'}$, para $1 \leq g < g' \leq G$, e a hipótese alternativa é $H_1 : \exists g \neq g'$, tal que $F_g \neq F_{g'}$. No entanto, na prática, a suposição de distribuição idêntica dentro de cada grupo não é razoável, pelo fato de não conhecermos a distribuição dos elementos da amostra. Por exemplo, considere um conjunto com $n = 12$ séries temporais, sendo 6 delas com distribuição A e 6 com distribuição B . Suponha que a “mistura” em cada grupo seja máxima, ou seja, suponha que o tamanho de cada grupo seja 6 e que o teste será realizado sobre a seguinte configuração: Grupo 1: $AAABBB$ e Grupo 2: $BBBAAA$. Dentro do grupo $g = 1$, U_{n1} será calculada a partir de pares de séries com distribuição AA , BB e AB com a seguinte proporção: $1/5AA$, $1/5BB$ e $3/5AB$. Dentro do grupo $g' = 2$, a proporção será a mesma. Entre os grupos 1 e 2, U_{n12} terá a seguinte proporção: $1/4AA$, $1/4BB$ e $1/2AB$. Essa configuração não satisfaz a suposição de homogeneidade dentro dos grupos, que é necessária na demonstração da normalidade assintótica de B_n .

Para possibilitar testes de similaridades e dissimilaridades entre grupos, mostraremos que quando não temos distribuição idêntica dentro de cada grupo, (isto é, temos pelo menos duas distribuições), a homogeneidade dos grupos sempre é rejeitada. Além disso, o poder do teste, em caso de mistura, é menor do que quando temos duas distribuições, e as variáveis são separadas em grupos com mesma distribuição. Isto permite-nos agrupar séries, testando todos os agrupamentos possíveis e escolhendo o que tem o menor p-valor.

4.1.2 Tamanhos dos Grupos Iguais ao Número de Amostras i.d.

Sem perda de generalidade, suponhamos que $G = 2$. Suponha que o número de amostras em cada grupo é denotado, respectivamente, por N_1 e N_2 , com $N_1 + N_2 = n$, e que todas as n séries temporais são de tamanho K , sendo que N_1 dessas séries têm distribuição A , e N_2 delas têm distribuição B .

Gostaríamos de agrupar estas séries de tal forma que aquelas com mesma distribuição fossem classificadas no mesmo grupo. Para isso, teremos que testar todos os agrupamentos possíveis. Mas, para simplificar um pouco os cálculos, considere n séries temporais de tamanho K , separadas em dois grupos de tamanho N_1 e N_2 (que são exatamente o número de séries temporais com distribuição A e B , respectivamente), com $N_1 + N_2 = n$ e ordenadas da forma

$$\underbrace{X_1^{(A)}, \dots, X_{n_{1A}}^{(A)}}_{n_{1A}}, \underbrace{X_{n_{1A}+1}^{(B)}, \dots, X_{N_1}^{(B)}}_{n_{1B}}, \underbrace{X_{N_1+1}^{(A)}, \dots, X_{N_1+1+n_{2A}}^{(A)}}_{n_{2A}}, \underbrace{X_{N_1+n_{2A}+2}^{(B)}, \dots, X_n^{(B)}}_{n_{2B}},$$

em que a notação $X_i^{(A)}$ indica que a amostra X_i tem distribuição A , e $X_j^{(B)}$ indica que X_j tem distribuição B .

Sejam $I_1 = \{1, \dots, N_1\}$, $I_2 = \{N_1+1, \dots, n\}$, $I_{1A} = \{1, \dots, n_{1A}\}$, $I_{1B} = \{n_{1A}+1, \dots, N_1\}$, $I_{2A} = \{N_1+1, \dots, N_1+1+n_{2A}\}$ e $I_{2B} = \{N_1+2+n_{2A}, \dots, n\}$. Temos que testar a similaridade desse agrupamento. A hipótese nula neste caso é: $H_0^* : \mathcal{L}(X_i) \equiv \mathcal{L}(X_j)$, para todo $i \in I_1$ e $j \in I_2$. A hipótese alternativa é: $H_1^* : \mathcal{L}(X_i) \neq \mathcal{L}(X_j)$ para algum $i \in I_1$ e $j \in I_2$.

Basicamente, o que iremos mostrar aqui é que, se $A = B$, então não rejeitamos H_0^* , para quaisquer que sejam N_1 e N_2 . No entanto, se $A \neq B$, então sempre rejeitamos H_0^* . Além disso, o maior poder ocorre quando as séries que têm distribuição A estão em um grupo, e as que têm distribuição B estão em outro.

Para cada agrupamento de tamanhos (N_1, N_2) , $B_n = N_1 N_2 [n(n-1)]^{-1} \{2U_{n12} - U_{n1} - U_{n2}\}$. Portanto, podemos fazer a decomposição de Hoeffding para cada uma das seguintes U-estatísticas U_{n12} , U_{n1} e U_{n2} . Para $g = 1, 2$,

$$\begin{aligned} U_{ng} &= \binom{n_g}{2}^{-1} \sum_{1 \leq i < j \leq n_g} \phi(X_{ig}, X_{jg}) \\ &= \binom{n_g}{2}^{-1} \sum_{1 \leq i < j \leq n_g} \{\phi_{0ij} + \psi_{1ij}(X_{ig}) + \psi_{1ij}(X_{jg}) + \psi_{2ij}(X_{ig}, X_{jg})\}, \end{aligned} \quad (4.1)$$

em que $\phi_{0ij} = \mathbb{E}\phi(X_i, X_j)$, $\psi_{1ij}(X_i) = \phi_{1ij}(X_i) - \phi_{0ij}$, com $\phi_{1ij}(X_i) = \mathbb{E}[\phi(X_i, X_j)|X_i]$ e $\psi_{2ij}(X_i, X_j) = \phi(X_i, X_j) - \psi_{1ij}(X_i) - \psi_{1ij}(X_j) - \phi_{0ij}$. Para a última U-estatística segue

$$\begin{aligned} U_{n12} &= \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \phi(X_{i1}, X_{j2}) \\ &= \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \{\phi_{0ij} + \psi_{1ij}(X_{i1}) + \psi_{1ij}(X_{j2}) + \psi_{2ij}(X_{i1}, X_{j2})\}. \end{aligned} \quad (4.2)$$

Sejam n_{gA} e n_{gB} o número de amostras, respectivamente, com distribuição A e B , dentro do grupo $g = 1, 2$. Com esta configuração dentro dos grupos teremos três situações possíveis para a distribuição de um par de amostras, o que implica três distintos valores para ϕ_0 , a saber: $\theta_{AA} = \mathbb{E}\phi(X_{ig}^{(A)}, X_{jg}^{(A)})$, $\theta_{BB} = \mathbb{E}\phi(X_{ig}^{(B)}, X_{jg}^{(B)})$ e $\theta_{AB} = \mathbb{E}\phi(X_{ig}^{(A)}, X_{jg}^{(B)})$.

Com estas definições, podemos escrever U_{ng} , para $g = 1, 2$, e U_{n12} da seguinte forma:

$$U_{ng} = \binom{n_g}{2}^{-1} \sum_{1 \leq i < j \leq n_g} \{\phi_{0ij} + \psi_{1ij}(X_{ig}) + \psi_{1ij}(X_{jg}) + \psi_{2ij}(X_{ig}, X_{jg})\},$$

$$\begin{aligned}
&= \binom{n_g}{2}^{-1} \left\{ \frac{n_{gA}(n_{gA}-1)}{2} \theta_{AA} + \frac{n_{gB}(n_{gB}-1)}{2} \theta_{BB} + n_{gA}n_{gB} \theta_{AB} \right\} \\
&+ \binom{n_g}{2}^{-1} \sum_{1 \leq i < j \leq n_g} \{ \psi_{1ij}(X_{ig}) + \psi_{1ij}(X_{jg}) \} + \binom{n_g}{2}^{-1} \sum_{1 \leq i < j \leq n_g} \psi_{2ij}(X_{ig}, X_{jg}),
\end{aligned} \tag{4.3}$$

$$\begin{aligned}
U_{n12} &= \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \{ \phi_{0ij} + \psi_{1ij}(X_{i1}) + \psi_{1ij}(X_{j2}) + \psi_{2ij}(X_{i1}, X_{j2}) \}, \\
&= \frac{1}{N_1 N_2} \{ n_{1A}n_{2A} \theta_{AA} + n_{1B}n_{2B} \theta_{BB} + (n_{1A}n_{2B} + n_{2A}n_{1B}) \theta_{AB} \} \\
&+ \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \{ \psi_{1ij}(X_{i1}) + \psi_{1ij}(X_{j2}) \} + \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \psi_{2ij}(X_{i1}, X_{j2}).
\end{aligned} \tag{4.4}$$

Denote

$$\psi_{1ij}(X_i) = \begin{cases} \psi_{1AA}(X_i), & \text{se } X_i \sim A \text{ e } X_j \sim A; \\ \psi_{1BB}(X_i), & \text{se } X_i \sim B \text{ e } X_j \sim B; \\ \psi_{1AB}(X_i), & \text{se } X_i \sim A \text{ e } X_j \sim B; \\ \psi_{1BA}(X_i), & \text{se } X_i \sim B \text{ e } X_j \sim A. \end{cases} \tag{4.5}$$

De fato, para os termos com AB , o que importa é a distribuição da variável condicional. Logo, podemos simplificar a notação destes termos

$$\begin{aligned}
\psi_{1AB}(X_i) &= \psi_{1BA}(X_i) = \psi_{1A}(X_i), \text{ se } X_i \sim A \text{ e} \\
\psi_{1AB}(X_i) &= \psi_{1BA}(X_i) = \psi_{1B}(X_i), \text{ se } X_i \sim B.
\end{aligned} \tag{4.6}$$

Com estas notações, podemos reescrever as somas dos termos de ordem 1 da decomposição de Hoeffding nas equações (4.3) e (4.4) como

$$\begin{aligned}
&\sum_{i,j \in I_g} \{ \psi_{1ij}(X_{ig}) + \psi_{1ij}(X_{jg}) \} \\
&= \sum_{i,j \in I_{gA}} \{ \psi_{1AA}(X_{ig}) + \psi_{1AA}(X_{jg}) \} + \sum_{i,j \in I_{gB}} \{ \psi_{1BB}(X_{ig}) + \psi_{1BB}(X_{jg}) \} \\
&+ \sum_{i,j \in I_{gA}} \sum_{i,j \in I_{gB}} \{ \psi_{1A}(X_{ig}) + \psi_{1B}(X_{jg}) \},
\end{aligned} \tag{4.7}$$

para $g = 1, 2$, e

$$\begin{aligned}
& \sum_{i \in I_1} \sum_{j \in I_2} \{\psi_{1ij}(X_i) + \psi_{1ij}(X_j)\} \\
&= \sum_{i \in I_{1A}} \sum_{j \in I_{2A}} \{\psi_{1AA}(X_i) + \psi_{1AA}(X_j)\} + \sum_{i \in I_{1B}} \sum_{j \in I_{2B}} \{\psi_{1BB}(X_i) + \psi_{1BB}(X_j)\} \\
&+ \sum_{i \in I_{1A}} \sum_{j \in I_{2B}} \{\psi_{1A}(X_i) + \psi_{1B}(X_j)\} + \sum_{i \in I_{1B}} \sum_{j \in I_{2A}} \{\psi_{1B}(X_i) + \psi_{1A}(X_j)\} \quad (4.8)
\end{aligned}$$

para U_{n12} .

Segue que os termos de ordem 1 da decomposição de Hoeffding de $2U_{n12} - U_{n1} - U_{n2}$ são os seguintes:

$$\begin{aligned}
& \frac{2}{N_1 N_2} \left\{ \sum_{i \in I_{1A}} \sum_{j \in I_{2A}} \{\psi_{1AA}(X_i) + \psi_{1AA}(X_j)\} + \sum_{i \in I_{1B}} \sum_{j \in I_{2B}} \{\psi_{1BB}(X_i) + \psi_{1BB}(X_j)\} \right\} \\
&+ \frac{2}{N_1 N_2} \left\{ \sum_{i \in I_{1A}} \sum_{j \in I_{2B}} \{\psi_{1A}(X_i) + \psi_{1B}(X_j)\} + \sum_{i \in I_{1B}} \sum_{j \in I_{2A}} \{\psi_{1B}(X_i) + \psi_{1A}(X_j)\} \right\} \\
&- \binom{N_1}{2}^{-1} \left\{ \sum_{i, j \in I_{1A}} \{\psi_{1AA}(X_i) + \psi_{1AA}(X_j)\} + \sum_{i, j \in I_{1B}} \{\psi_{1BB}(X_i) + \psi_{1BB}(X_j)\} \right\} \\
&- \binom{N_1}{2}^{-1} \sum_{i, j \in I_{1A}} \sum_{i, j \in I_{1B}} \{\psi_{1A}(X_i) + \psi_{1B}(X_j)\}. \\
&- \binom{N_2}{2}^{-1} \left\{ \sum_{i, j \in I_{2A}} \{\psi_{1AA}(X_i) + \psi_{1AA}(X_j)\} + \sum_{i, j \in I_{2B}} \{\psi_{1BB}(X_i) + \psi_{1BB}(X_j)\} \right\} \\
&- \binom{N_2}{2}^{-1} \sum_{i, j \in I_{2A}} \sum_{i, j \in I_{2B}} \{\psi_{1A}(X_i) + \psi_{1B}(X_j)\}.
\end{aligned}$$

Podemos reescrever (4.9) como

$$\begin{aligned}
& \sum_{i \in I_{1A}} \left[\frac{2n_{2A}}{N_1 N_2} - \frac{2(n_{1A} - 1)}{N_1(N_1 - 1)} \right] \psi_{1AA}(X_i) + \sum_{j \in I_{2A}} \left[\frac{2n_{1A}}{N_1 N_2} - \frac{2(n_{2A} - 1)}{N_2(N_2 - 1)} \right] \psi_{1AA}(X_j) \\
&+ \sum_{i \in I_{1B}} \left[\frac{2n_{2B}}{N_1 N_2} - \frac{2(n_{1B} - 1)}{N_1(N_1 - 1)} \right] \psi_{1BB}(X_i) + \sum_{j \in I_{2B}} \left[\frac{2n_{1B}}{N_1 N_2} - \frac{2(n_{2B} - 1)}{N_2(N_2 - 1)} \right] \psi_{1BB}(X_j) \\
&+ \sum_{i \in I_{1A}} \left[\frac{2n_{2B}}{N_1 N_2} - \frac{2n_{1B}}{N_1(N_1 - 1)} \right] \psi_{1A}(X_i) + \sum_{i \in I_{1B}} \left[\frac{2n_{2A}}{N_1 N_2} - \frac{2n_{1A}}{N_1(N_1 - 1)} \right] \psi_{1B}(X_i) \quad (4.9)
\end{aligned}$$

$$\begin{aligned}
& + \sum_{i \in I_{2A}} \left[\frac{2n_{2A}}{N_1 N_2} - \frac{2n_{1A}}{N_2(N_2 - 1)} \right] \psi_{1A}(X_i) + \sum_{i \in I_{2B}} \left[\frac{2n_{1A}}{N_1 N_2} - \frac{2n_{2A}}{N_2(N_2 - 1)} \right] \psi_{1B}(X_i) \\
& = \sum_{i \in I_A} Y_i^* + \sum_{i \in I_B} W_i^* + \sum_{i \in I_A} Z_i^* + \sum_{i \in I_B} S_i^*,
\end{aligned}$$

em que

$$\begin{aligned}
Y_i^* &= 2 \left(\frac{n_{\xi_i^* A}}{N_1 N_2} - \frac{1}{N_{\xi_i}} \right) \psi_{1AA}(X_i), \quad W_i^* = 2 \left(\frac{n_{\xi_i^* B}}{N_1 N_2} - \frac{1}{N_{\xi_i}} \right) \psi_{1BB}(X_i), \\
Z_i^* &= 2 \left(\frac{n_{\xi_i^* B}}{N_1 N_2} - \frac{n_{\xi_i B}}{N_{\xi_i}(N_{\xi_i} - 1)} \right) \psi_{1A}(X_i), \\
S_i^* &= 2 \left(\frac{n_{\xi_i^* A}}{N_1 N_2} - \frac{n_{\xi_i A}}{N_{\xi_i}(N_{\xi_i} - 1)} \right) \psi_{1B}(X_i),
\end{aligned} \tag{4.10}$$

$$\xi_i = \begin{cases} 1, & \text{se } X_i \in \text{grupo 1} \\ 2, & \text{se } X_i \in \text{grupo 2} \end{cases} \quad \text{e} \quad \xi_i^* = \begin{cases} 2, & \text{se } X_i \in \text{grupo 1} \\ 1, & \text{se } X_i \in \text{grupo 2} \end{cases}. \tag{4.11}$$

O ponto principal é que, devido à propriedade de ortogonalidade da decomposição de Hoeffding, Y_i^* , W_i^* , Z_i^* e S_i^* são independentes e, portanto, são não correlacionados. Além disso, cada uma das somas em (4.10) é uma soma de variáveis i.i.d., para as quais resultados assintóticos, como TCL, podem ser provados. Com esses resultados, estamos em condições de estudar o comportamento assintótico de B_n , que pode ser escrito como

$$\begin{aligned}
B_n &= \frac{N_1 N_2}{n(n-1)} \{2U_{n12} - U_{n1} - U_{n2}\} \\
&= \left\{ \theta + \sum_{i \in I_A} Y_i + \sum_{i \in I_B} W_i + \sum_{i \in I_A} Z_i + \sum_{i \in I_B} S_i + T_n \right\},
\end{aligned} \tag{4.12}$$

em que

$$\begin{aligned}
\theta &= \frac{N_1 N_2}{n(n-1)} \left\{ \left(\frac{2n_{1A}n_{2A}}{N_1 N_2} - \frac{n_{1A}(n_{1A} - 1)}{N_1(N_1 - 1)} - \frac{n_{2A}(n_{2A} - 1)}{N_2(N_2 - 1)} \right) \theta_{AA} \right. \\
&\quad + \left(\frac{2n_{1B}n_{2B}}{N_1 N_2} - \frac{n_{1B}(n_{1B} - 1)}{N_1(N_1 - 1)} - \frac{n_{2B}(n_{2B} - 1)}{N_2(N_2 - 1)} \right) \theta_{BB} \\
&\quad \left. + \left(\frac{2(n_{1A}n_{2B} + n_{2A}n_{1B})}{N_1 N_2} - \frac{2n_{1A}n_{1B}}{N_1(N_1 - 1)} - \frac{2n_{2A}n_{2B}}{N_2(N_2 - 1)} \right) \theta_{AB} \right\},
\end{aligned} \tag{4.13}$$

$Y_i = \frac{N_1 N_2}{n(n-1)} Y_i^*$, $W_i = \frac{N_1 N_2}{n(n-1)} W_i^*$, $Z_i = \frac{N_1 N_2}{n(n-1)} Z_i^*$, $S_i = \frac{N_1 N_2}{n(n-1)} S_i^*$ e

$$T_n = \frac{N_1 N_2}{n(n-1)} \left\{ \frac{2}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \psi_2(X_i, X_j) - \binom{N_1}{2}^{-1} \sum_{1 \leq i < j \leq N_1} \psi_2(X_{i1}, X_{j1}) - \binom{N_2}{2}^{-1} \sum_{1 \leq i < j \leq N_2} \psi_2(X_{i2}, X_{j2}) \right\}. \quad (4.14)$$

Definindo-se as variáveis $\sigma_{AA}^2 = \text{Var}\psi_{1AA}(X_i)$, $\sigma_{BB}^2 = \text{Var}\psi_{1BB}(X_i)$, $\sigma_A^2 = \text{Var}\psi_{1A}(X_i)$ e $\sigma_B^2 = \text{Var}\psi_{1B}(X_i)$. As variâncias das somas $\sum_{i \in I_A} Y_i$, $\sum_{i \in I_B} W_i$, $\sum_{i \in I_A} Z_i$ e $\sum_{i \in I_B} S_i$ são todas da ordem de n^{-1} . Segue do Teorema 4 da Seção 1.6 de Lee (1990) que

$$\text{Var}\sqrt{n}T_n = O(n^{-1}).$$

Portanto,

$$\sqrt{n}(B_n - \theta) \xrightarrow{D} \mathcal{N}(0, \eta^2), \quad (4.15)$$

em que θ é dado por (4.13) e

$$\begin{aligned} \eta^2 = & \left(\frac{2N_1 N_2}{\sqrt{n}(n-1)} \right)^2 \left\{ \left[n_{1A} \left(\frac{n_{2A}}{N_1 N_2} - \frac{1}{N_1} \right)^2 + n_{2A} \left(\frac{n_{1A}}{N_1 N_2} - \frac{1}{N_2} \right)^2 \right] \sigma_{AA}^2 \right. \\ & + \left[n_{1B} \left(\frac{n_{2B}}{N_1 N_2} - \frac{1}{N_1} \right)^2 + n_{2B} \left(\frac{n_{1B}}{N_1 N_2} - \frac{1}{N_2} \right)^2 \right] \sigma_{BB}^2 \\ & + \left[n_{1A} \left(\frac{n_{2B}}{N_1 N_2} - \frac{n_{1B}}{N_1(N_1-1)} \right)^2 + n_{2A} \left(\frac{n_{1B}}{N_1 N_2} - \frac{n_{2B}}{N_2(N_2-1)} \right)^2 \right] \sigma_A^2 \\ & \left. + \left[n_{1B} \left(\frac{n_{2A}}{N_1 N_2} - \frac{n_{1A}}{N_1(N_1-1)} \right)^2 + n_{2B} \left(\frac{n_{1A}}{N_1 N_2} - \frac{n_{2A}}{N_2(N_2-1)} \right)^2 \right] \sigma_B^2 \right\}. \quad (4.16) \end{aligned}$$

4.1.3 Todos os Possíveis Agrupamentos

A suposição até aqui é que N_1 e N_2 são os tamanhos dos agrupamentos e, também, o número de séries com distribuição A e B , respectivamente. No entanto, precisamos testar os agrupamentos sem supor N_1 e N_2 conhecidos. Vamos mostrar que para quaisquer que sejam N_1 e N_2 , denotando o número de séries com distribuição A e B , respectivamente, o menor p-valor resultante de testes de similaridade entre grupos de tamanhos n_1 e n_2 é aquele com $n_1 = N_1$ e $n_2 = N_2$.

Suponha que N_1 e N_2 denotam o número de séries com distribuição A e B , respectivamente, com $N_1 + N_2 = n$. Considere um agrupamento em que os tamanhos dos grupos são n_1 e n_2 com $n_1 + n_2 = n$ e $0 \leq n_1, n_2 \leq n - 2$. Defina n_{1A}, n_{1B}, n_{2A} e n_{2B} como na seção anterior. Observe que $n_{1A} + n_{2A} = N_1$, o que implica que $n_{2A} = N_1 - n_{1A}$, mas com a restrição de que $0 \leq n_{1A} \leq \min\{n_1, N_1\}$. Analogamente, $n_{1B} + n_{2B} = N_2$, o que implica que $n_{2B} = N_2 - n_{1B}$, com $0 \leq n_{1B} \leq \min\{n_2, N_2\}$. Portanto, os resultados encontrados na seção anterior seguem substituindo-se n_{2A} por $N_1 - n_{1A}$ e n_{2B} por $N_2 - n_{1B}$, respeitando-se as respectivas restrições.

Assim,

$$\begin{aligned} \theta = & \frac{n_1 n_2}{n(n-1)} \left\{ \left(\frac{2n_{1A}(N_1 - n_{1A})}{n_1 n_2} - \frac{n_{1A}(n_{1A} - 1)}{n_1(n_1 - 1)} - \frac{(N_1 - n_{1A})(N_1 - n_{1A} - 1)}{n_2(n_2 - 1)} \right) \theta_{AA} \right. \\ & + \left(\frac{2n_{1B}(N_2 - n_{1B})}{n_1 n_2} - \frac{n_{1B}(n_{1B} - 1)}{n_1(n_1 - 1)} - \frac{(N_2 - n_{1B})(N_2 - n_{1B} - 1)}{n_2(n_2 - 1)} \right) \theta_{BB} \\ & \left. + \left(\frac{2(n_{1A}(N_2 - n_{1B}) + (N_1 - n_{1A})n_{1B})}{n_1 n_2} - \frac{2n_{1A}n_{1B}}{n_1(n_1 - 1)} - \frac{2(N_1 - n_{1A})(N_2 - n_{1B})}{n_2(n_2 - 1)} \right) \theta_{AB} \right\} \end{aligned} \quad (4.17)$$

e

$$\begin{aligned} \eta^2 = & \frac{1}{n} \left(\frac{2n_1 n_2}{n-1} \right)^2 \left\{ \left[n_{1A} \left(\frac{N_1 - n_{1A}}{n_1 n_2} - \frac{1}{n_1} \right)^2 + (N_1 - n_{1A}) \left(\frac{n_{1A}}{n_1 n_2} - \frac{1}{n_2} \right)^2 \right] \sigma_{AA}^2 \right. \\ & + \left[n_{1B} \left(\frac{N_2 - n_{1B}}{n_1 n_2} - \frac{1}{n_1} \right)^2 + (N_2 - n_{1B}) \left(\frac{n_{1B}}{n_1 n_2} - \frac{1}{n_2} \right)^2 \right] \sigma_{BB}^2 \\ & + \left[n_{1A} \left(\frac{N_2 - n_{1B}}{n_1 n_2} - \frac{n_{1B}}{n_1(n_1 - 1)} \right)^2 + (N_1 - n_{1A}) \left(\frac{n_{1B}}{n_1 n_2} - \frac{N_2 - n_{1B}}{n_2(n_2 - 1)} \right)^2 \right] \sigma_A^2 \\ & \left. + \left[n_{1B} \left(\frac{N_1 - n_{1A}}{n_1 n_2} - \frac{n_{1A}}{n_1(n_1 - 1)} \right)^2 + (N_2 - n_{1B}) \left(\frac{n_{1A}}{n_1 n_2} - \frac{N_1 - n_{1A}}{n_2(n_2 - 1)} \right)^2 \right] \sigma_B^2 \right\}. \end{aligned} \quad (4.18)$$

Observe que trabalhar com a expressão de θ , em (4.17), é complicado devido ao grande número de variáveis envolvidas. No entanto, podemos supor $n = 12$, por exemplo, e mostrar o comportamento de θ para todas as configurações possíveis, confirmando a afirmação do início da seção, de que o menor p-valor ocorre quando A 's e B 's estão em grupos distintos. Para tanto, basta verificar o valor da média θ nas diferentes configurações. Para isso, consideremos $D_{WB} = 2\theta_{AB} - \theta_{AA} - \theta_{BB}$.

Nas Tabelas 4.1 e 4.2 temos alguns valores de θ em função de D_{WB} . Observe que basta conhecermos n_1 e n_{1A} para conhecermos o restante das variáveis. Por exemplo, se $n = 12$ e $N_1 = 6 \Rightarrow N_2 = 6$. Se $n_{1A} = 0 \Rightarrow n_{1B} = 6$, o que já determina θ .

Tabela 4.1: Valores de θ em função de $D_{WB} = 2\theta_{AB} - \theta_A - \theta_B$, com $N_1 = 6$ e $N_2 = 6$ em todos os possíveis valores de n_1 e n_{1A} .

$n = 12$	$N_1 = 6, N_2 = 6$		
$n_1 = 6$	θ	$n_1 = 5$	θ
$n_{1A} = 0$	$\frac{6*6}{12*11} * 1 * D_{WB} = \mathbf{0.2727}D_{WB}$	$n_{1A} = 0$	$\frac{5*7}{12*11} * \frac{5}{7} * D_{WB} = 0.1894D_{WB}$
$n_{1A} = 1$	$\frac{6*6}{12*11} \frac{7}{18} D_{WB} = 0.1061D_{WB}$	$n_{1A} = 1$	$\frac{5*7}{12*11} \frac{4}{21} D_{WB} = 0.0505D_{WB}$
$n_{1A} = 2$	$\frac{6*6}{12*11} \frac{1}{45} D_{WB} = 0.0061D_{WB}$	$n_{1A} = 2$	$\frac{5*7}{12*11} \frac{1}{14} D_{WB} = 0.0189D_{WB}$
$n_{1A} = 3$	$\frac{6*6}{12*11} \frac{-1}{10} D_{WB} = -0.0273D_{WB}$	$n_{1A} = 3$	$\frac{5*7}{12*11} \frac{1}{14} D_{WB} = 0.0189D_{WB}$
$n_{1A} = 4$	$\frac{6*6}{12*11} \frac{1}{45} D_{WB} = 0.0061D_{WB}$	$n_{1A} = 4$	$\frac{5*7}{12*11} \frac{4}{21} D_{WB} = 0.0505D_{WB}$
$n_{1A} = 5$	$\frac{6*6}{12*11} \frac{7}{18} D_{WB} = 0.1061D_{WB}$	$n_{1A} = 5$	$\frac{5*7}{12*11} * \frac{5}{7} * D_{WB} = 0.1894D_{WB}$
$n_{1A} = 6$	$\frac{6*6}{12*11} * 1 * D_{WB} = 0.2727D_{WB}$		
$n_1 = 4$	θ	$n_1 = 3$	θ
$n_{1A} = 0$	$\frac{4*8}{12*11} \frac{15}{28} D_{WB} = 0.1299D_{WB}$	$n_{1A} = 0$	$\frac{3*9}{12*11} * \frac{5}{12} * D_{WB} = 0.0852D_{WB}$
$n_{1A} = 1$	$\frac{4*8}{12*11} \frac{5}{112} D_{WB} = 0.0108D_{WB}$	$n_{1A} = 1$	$\frac{3*9}{12*11} \frac{5}{54} D_{WB} = 0.0189D_{WB}$
$n_{1A} = 2$	$\frac{4*8}{12*11} \frac{5}{42} D_{WB} = 0.0289D_{WB}$		
$n_1 = 2$	θ		
$n_{1A} = 0$	$\frac{2*10}{12*11} \frac{1}{3} D_{WB} = 0.0505D_{WB}$		
$n_{1A} = 1$	$\frac{2*10}{12*11} \frac{5}{18} D_{WB} = 0.0421D_{WB}$		

Por depender de D_{WB} , não é possível montar tabelas para todos os possíveis valores de N_1 e N_2 , pois para alguns desses valores, como $N_1 = 4$ e $N_2 = 8$, nem todos os resultados podem ser fatorados e escritos em função de D_{WB} . No entanto, o comportamento é semelhante aos casos considerados, em que θ , em valor absoluto, é sempre maior quando $n_{1A} = n_1 = N_1$.

Embora sejam pouco formais, os resultados apresentados nas Tabelas 4.1 e 4.2 indicam que a afirmação feita no início da seção, de que o menor p-valor está associado diretamente à maior dissimilaridade entre os grupos e maior similaridade dentro do grupos, é, possivelmente, verdadeira.

Essa característica do teste é o nosso ponto de partida para o desenvolvimento do método de agrupamento, que consiste em testar todos os possíveis agrupamentos via procedimento bootstrap e escolher o que tem p-valor menor.

Tabela 4.2: Valores de θ em função de $D_{WB} = 2\theta_{AB} - \theta_A - \theta_B$, com $N_1 = 5$ e $N_2 = 7$ em todos os possíveis valores de n_1 e n_{1A} .

$n = 12$	$N_1 = 5, N_2 = 7$		
$n_1 = 6$	θ	$n_1 = 5$	θ
$n_{1A} = 0$	$\frac{6*6}{12*11} * \frac{2}{3} * D_{WB} = 0.1818D_{WB}$	$n_{1A} = 0$	$\frac{5*7}{12*11} \frac{10}{21} D_{WB} = 0.1263D_{WB}$
$n_{1A} = 1$	$\frac{6*6}{12*11} \frac{8}{45} D_{WB} = 0.0485D_{WB}$	$n_{1A} = 1$	$\frac{5*7}{12*11} \frac{2}{35} D_{WB} = 0.0152D_{WB}$
$n_{1A} = 2$	$\frac{6*6}{12*11} \frac{-1}{15} D_{WB} = -0.0182D_{WB}$	$n_{1A} = 2$	$\frac{5*7}{12*11} \frac{-1}{10} D_{WB} = -0.0265D_{WB}$
$n_{1A} = 3$	$\frac{6*6}{12*11} \frac{-1}{15} D_{WB} = -0.0182D_{WB}$	$n_{1A} = 3$	$\frac{5*7}{12*11} 0.0048D_{WB} = 0.0130D_{WB}$
$n_{1A} = 4$	$\frac{6*6}{12*11} \frac{8}{45} D_{WB} = 0.0485D_{WB}$	$n_{1A} = 4$	$\frac{5*7}{12*11} \frac{13}{35} D_{WB} = 0.0985D_{WB}$
$n_{1A} = 5$	$\frac{6*6}{12*11} \frac{2}{3} D_{WB} = 0.1818D_{WB}$	$n_{1A} = 5$	$\frac{5*7}{12*11} * 1 * D_{WB} = \mathbf{0.2652D_{WB}}$
$n_1 = 4$	θ	$n_1 = 3$	θ
$n_{1A} = 0$	$\frac{4*8}{12*11} * \frac{5}{14} * D_{WB} = 0.0866D_{WB}$	$n_{1A} = 0$	$\frac{3*9}{12*11} \frac{5}{18} D_{WB} = 0.0568D_{WB}$
$n_{1A} = 1$	$\frac{4*8}{12*11} \frac{-1}{28} D_{WB} = -0.0087D_{WB}$	$n_{1A} = 1$	$\frac{3*9}{12*11} \frac{7}{54} D_{WB} = 0.0265D_{WB}$
$n_{1A} = 2$	$\frac{4*8}{12*11} \frac{-17}{168} D_{WB} = -0.0245D_{WB}$	$n_{1A} = 2$	$\frac{3*9}{12*11} \frac{-1}{36} D_{WB} = -0.0057D_{WB}$
$n_{1A} = 3$	$\frac{4*8}{12*11} \frac{9}{56} D_{WB} = 0.0390D_{WB}$	$n_{1A} = 3$	$\frac{3*9}{12*11} \frac{7}{12} D_{WB} = 0.1193D_{WB}$
$n_{1A} = 4$	$\frac{4*8}{12*11} \frac{3}{4} D_{WB} = 0.1818D_{WB}$		
$n_1 = 2$	θ		
$n_{1A} = 0$	$\frac{2*10}{12*11} * \frac{2}{9} * D_{WB} = 0.0337D_{WB}$		
$n_{1A} = 1$	$\frac{2*10}{12*11} \frac{-4}{15} D_{WB} = -0.0404D_{WB}$		
$n_{1A} = 0$	$\frac{2*10}{12*11} \frac{7}{15} D_{WB} = 0.0707D_{WB}$		

4.2 Estudo da Estatística de Teste através de Simulação

No Capítulo 3, estudamos as propriedades assintóticas da estatística de teste B_n . Sob a hipótese de homogeneidade dos grupos, B_n é assintoticamente normal e este resultado independe quantitativamente da escolha da métrica. No entanto, é preciso um estudo de simulação para comparar as métricas em diferentes situações. Nesta seção, realizamos algumas simulações para comparar as métricas \mathbf{d}_P , \mathbf{d}_{NP} , \mathbf{d}_{LNP} e \mathbf{d}_{ACF} .

4.2.1 Comparação das Métricas através de Simulação

Realizamos um breve estudo de simulação para comparar as métricas \mathbf{d}_P , \mathbf{d}_{NP} , \mathbf{d}_{LNP} e \mathbf{d}_{ACF} e escolher a melhor delas, ou seja, a métrica que apresenta o melhor desempenho nas situações testadas, para ser utilizada como núcleo no nosso teste. Simulamos 6, 12 e 50 séries de tamanhos 50, 100, 200, 500, 1000 e 10000, sendo metade destas séries geradas a partir de um modelo AR(1), definido no Apêndice, com parâmetro autoregressivo $\phi_1 = 0.5$, e a outra metade das séries geradas a partir de um modelo MA(1), definido no Apêndice, com parâmetro $\theta_1 = 1$. Ambos os modelos são gerados a partir de inovações gaussianas de média 0 e variância 1. Foram realizadas 10000 replicações, exceto para $K = 10000$, em que realizamos 500 repetições. Para a métrica \mathbf{d}_{ACF} utilizamos $L = 0.1K$, e para as métricas baseadas no periodograma utilizamos todas as frequências ω_ℓ , $\ell = 1 \dots, \lfloor K/2 \rfloor$. O poder do teste é apresentado na Tabela 4.3 e mostra que a métrica \mathbf{d}_{LNP} é a que tem o melhor desempenho. A média e o desvio padrão da estatística são apresentados na Tabela 4.4, na qual pode-se observar que o desvio padrão diminui, a medida em que aumentamos n e K e a média é diferente de zero. Esse comportamento é observado para todas as métricas.

Embora a métrica \mathbf{d}_P tenha apresentado um desempenho superior a métrica \mathbf{d}_{ACF} , acreditamos que um estudo para determinar o valor ótimo de L contribua para que essa métrica tenha um desempenho melhor. Além disso, a métrica \mathbf{d}_{LNP} é visivelmente superior à \mathbf{d}_P e, portanto, justifica a escolha de \mathbf{d}_{ACF} e \mathbf{d}_{LNP} para que sejam exploradas em diferentes configurações.

4.2.2 Simulação com a Métrica \mathbf{d}_{LNP}

Nesta subseção pretendemos estudar, através de simulações, o comportamento da métrica \mathbf{d}_{LNP} , em diferentes configurações e distribuições dos grupos de séries temporais. Em primeiro lugar, simulamos 4, 12 e 50 séries independentes a partir de um processo AR(1) gaussianas, de tamanhos 64, 256, 512 e 1024. O tamanho da série é tomado como uma potência de 2 para que o método proposto possa ser comparado com outros métodos baseados em wavelets. Todas as séries têm o mesmo parâmetro autorregressivo ϕ , e este parâmetro varia dentro da gama de modelos causais em que $|\phi| < 1$. As inovações (erros) são normalmente distribuídas com variância unitária. Assim, todas as séries têm mesma distribuição. A Tabela 4.5 reporta a

Tabela 4.3: Comparação do poder do teste com nível nominal de 5%, utilizando as métricas d_{ACF} , d_P , d_{NP} e d_{LNP} , com $n/2$ séries geradas a partir de um modelo AR(1) e $n/2$ séries geradas a partir de um modelo MA(1). Os resultados são baseados em 10000 replicações.

n	Métrica	K					
		50	100	200	500	1000	10000
6	d_{ACF}	0.0908	0.1131	0.1715	0.3390	0.6324	1.0000
	d_P	0.0912	0.1249	0.2104	0.4594	0.7632	1.0000
	d_{NP}	0.0600	0.0691	0.0926	0.1394	0.2428	0.9940
	d_{LNP}	0.2215	0.5128	0.8926	0.9999	1.0000	1.0000
12	d_{ACF}	0.1409	0.2197	0.3732	0.7875	0.9864	1.0000
	d_P	0.1898	0.3405	0.6118	0.9578	0.9999	1.0000
	d_{NP}	0.0911	0.1247	0.1921	0.4098	0.7135	1.0000
	d_{LNP}	0.4746	0.8886	0.9998	1.0000	1.0000	1.0000
50	d_{ACF}	0.5397	0.8562	0.9946	1.0000	1.0000	1.0000
	d_P	0.9238	0.9997	1.0000	1.0000	1.0000	1.0000
	d_{NP}	0.3628	0.6625	0.9429	1.0000	1.0000	1.0000
	d_{LNP}	0.9974	1.0000	1.0000	1.0000	1.0000	1.0000

Tabela 4.4: Média e desvio padrão da estatística de teste, referente ao estudo de simulação apresentado na Tabela 4.3.

n	Métrica	K					
		50	100	200	500	1000	10000
6	d_{ACF}	0.0059 0.0103	0.0038 0.0049	0.0022 0.0020	0.0009 0.0006	0.0005 0.0002	0.0000 0.0000
	d_P	0.0146 0.0236	0.0144 0.0171	0.0146 0.0121	0.0146 0.0076	0.0146 0.0055	0.0147 0.0018
	d_{NP}	0.0020 0.0071	0.0022 0.0054	0.0024 0.0041	0.0025 0.0027	0.0025 0.0019	0.0025 0.0006
	d_{LNP}	0.5968 0.4590	0.7137 0.3446	0.8201 0.2598	0.9403 0.1822	1.0191 0.1373	1.1829 0.0551
12	d_{ACF}	0.0054 0.0057	0.0034 0.0027	0.0020 0.0011	0.0009 0.0003	0.0004 0.0001	0.0000 0.0000
	d_P	0.0135 0.0116	0.0133 0.0080	0.0133 0.0058	0.0132 0.0037	0.0133 0.0026	0.0133 0.0008
	d_{NP}	0.0019 0.0033	0.0020 0.0026	0.0022 0.0019	0.0022 0.0013	0.0023 0.0009	0.0023 0.0003
	d_{LNP}	0.5356 0.2664	0.6468 0.2075	0.7479 0.1567	0.8567 0.1128	0.9253 0.0869	1.0771 0.0334
50	d_{ACF}	0.0050 0.0023	0.0032 0.0011	0.0019 0.0005	0.0008 0.0001	0.0004 0.0000	0.0000 0.0000
	d_P	0.0126 0.0038	0.0125 0.0027	0.0125 0.0020	0.0124 0.0012	0.0124 0.0009	0.0124 0.0003
	d_{NP}	0.0018 0.0010	0.0019 0.0008	0.0021 0.0006	0.0021 0.0004	0.0021 0.0003	0.0022 0.0001
	d_{LNP}	0.5052 0.1158	0.6073 0.0910	0.6975 0.0700	0.8011 0.0499	0.8656 0.0383	1.0077 0.0159

proporção de vezes que a estatística B_n rejeita a hipótese nula ao nível nominal de 5%. Foram realizadas 10000 simulações independentes e, com isso, os percentuais empíricos relatados são

razoavelmente precisos, no sentido de que os valores estão próximos de 5%. Nessa tabela também mostramos que os testes com a estatística B_n rejeitam H_0 a taxas próximas de 5% (por vezes um pouco maior). A média e o desvio padrão de B_n , para cada caso, são relatados na Tabela 4.6. Podemos observar que a média e o desvio padrão diminuem com o aumento de n e/ou K .

Tabela 4.5: Probabilidades de erro do tipo I empíricos a 5% utilizando-se a métrica \mathbf{d}_{LNP} . As n séries foram geradas a partir de um modelo AR(1). Os resultados são baseados em 10000 replicações.

K	$n \setminus \phi$	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75
64	4	0.0501	0.0524	0.0524	0.0498	0.0526	0.0493	0.0520
	12	0.0493	0.0476	0.0477	0.0487	0.0500	0.0478	0.0463
	50	0.0462	0.0473	0.0471	0.0461	0.0471	0.0446	0.0474
256	4	0.0501	0.0486	0.0491	0.0508	0.0499	0.0494	0.0526
	12	0.0480	0.0483	0.0473	0.0485	0.0473	0.0497	0.0503
	50	0.0500	0.0507	0.0471	0.0496	0.0469	0.0483	0.0502
512	4	0.0492	0.0533	0.0498	0.0510	0.0470	0.0487	0.0495
	12	0.0485	0.0495	0.0474	0.0478	0.0523	0.0490	0.0497
	50	0.0486	0.0499	0.0496	0.0476	0.0516	0.0493	0.0504
1024	4	0.0521	0.0504	0.0487	0.0494	0.0524	0.0491	0.0498
	12	0.0473	0.0479	0.0496	0.0484	0.0498	0.0483	0.0512
	50	0.0466	0.0499	0.0522	0.0535	0.0500	0.0509	0.0491

Na Tabela 4.7 analisamos os erros do tipo I para modelos MA(1). Simulamos 4, 12 e 50 séries MA(1) independentes com o mesmo parâmetro média móvel, θ , e erros gaussianos de média zero e variância unitária. As outras especificações são as mesmas usadas para fazer a simulação da Tabela 4.5. A média e o desvio padrão da estatística de teste referentes à simulação da Tabela 4.7 são apresentados na Tabela 4.8. Os resultados podem ser interpretados semelhantemente aos das Tabelas 4.5 e 4.6.

Na Tabela 4.9 apresentamos resultados de um estudo de simulação para o poder do teste em que a métrica \mathbf{d}_{LNP} é utilizada como núcleo da estatística de teste. Os números de séries simuladas são 4, 12 e 50, sendo que a primeira metade das séries são provenientes de um processo AR(1), com parâmetro ϕ e com erros gaussianos de média zero e variância unitária. A segunda metade das séries são independentes, geradas a partir de um processo MA(1), com o parâmetro de média móvel $\theta = \pm\sqrt{\phi^2/(1-\phi^2)}$ e os erros gaussianos de média zero e variância unitária. O sinal + é escolhido se $\phi > 0$, e o sinal - é escolhido se $\phi < 0$. Essa escolha de θ faz com que as variâncias das séries sejam todas idênticas. A escolha dos parâmetros foi feita dessa forma para seguir a mesma ideia de Lund et al. (2009), embora não seja nossa intenção comparar os resultados, porque as técnicas são aplicáveis a situações diferentes. A média e o desvio padrão da estatística de teste são apresentados na Tabela 4.10.

Tabela 4.6: Média e desvio padrão da estatística de teste, referentes ao estudo de simulação apresentado na Tabela 4.5.

K	$n \setminus \phi$	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75
64	4	0.0037 0.3910	-0.0018 0.3730	-0.0064 0.3748	-0.0030 0.3755	-0.0065 0.3698	-0.0006 0.3776	-0.0009 0.3876
	12	-0.0007 0.0915	-0.0010 0.0871	-0.0003 0.0868	0.0012 0.0872	0.0009 0.0873	0.0010 0.0887	0.0017 0.0924
	50	0.0002 0.0197	0.0001 0.0188	0.0000 0.0187	0.0001 0.0187	0.0002 0.0186	-0.0001 0.0187	0.0000 0.0197
256	4	0.0008 0.1762	0.0015 0.1714	-0.0009 0.1737	-0.0008 0.1730	-0.0003 0.1731	-0.0007 0.1743	0.0000 0.1768
	12	-0.0005 0.0405	0.0001 0.0401	-0.0001 0.0402	-0.0002 0.0400	0.0002 0.0403	0.0005 0.0404	0.0005 0.0413
	50	-0.0000 0.0088	0.0001 0.0087	-0.0000 0.0087	0.0001 0.0088	0.0000 0.0088	0.0000 0.0088	-0.0001 0.0089
512	4	-0.0000 0.1222	-0.0009 0.1219	-0.0008 0.1208	0.0016 0.1217	0.0005 0.1188	0.0008 0.1227	-0.0002 0.1207
	12	-0.0002 0.0285	0.0001 0.0280	0.0000 0.0285	0.0001 0.0282	0.0000 0.0277	0.0002 0.0282	0.0000 0.0283
	50	-0.0000 0.0061	0.0000 0.0061	-0.0001 0.0060	-0.0000 0.0060	0.0000 0.0060	0.0000 0.0061	0.0000 0.0060
1024	4	0.0002 0.0841	-0.0017 0.0849	0.0007 0.0846	0.0009 0.0846	0.0005 0.0839	-0.0004 0.0847	0.0001 0.0850
	12	-0.0000 0.0198	0.0000 0.0194	-0.0000 0.0196	0.0000 0.0198	-0.0003 0.0197	-0.0002 0.0198	-0.0002 0.0197
	50	-0.0000 0.0043	0.0000 0.0042	-0.0000 0.0042	-0.0000 0.0043	-0.0000 0.0042	-0.0000 0.0042	0.0000 0.0043

Tabela 4.7: Probabilidades de erro do tipo I empíricos a 5% utilizando-se a métrica d_{LNP} . As n séries foram geradas a partir de um modelo MA(1). Os resultados são baseados em 10000 replicações.

K	$n \setminus \theta$	-2.00	-1.00	0.00	0.25	0.50	1.00	2.00
64	4	0.0492	0.0507	0.0476	0.0520	0.0501	0.0515	0.0493
	12	0.0477	0.0487	0.0484	0.0505	0.0466	0.0473	0.0501
	50	0.0461	0.0485	0.0463	0.0467	0.0462	0.0457	0.0468
256	4	0.0497	0.0509	0.0515	0.0485	0.0531	0.0496	0.0510
	12	0.0493	0.0497	0.0499	0.0506	0.0470	0.0510	0.0490
	50	0.0499	0.0490	0.0504	0.0473	0.0496	0.0449	0.0513
512	4	0.0504	0.0492	0.0498	0.0532	0.0515	0.0499	0.0546
	12	0.0512	0.0483	0.0477	0.0509	0.0502	0.0504	0.0508
	50	0.0503	0.0504	0.0480	0.0468	0.0496	0.0488	0.0494
1024	4	0.0501	0.0494	0.0528	0.0501	0.0494	0.0489	0.0489
	12	0.0494	0.0503	0.0516	0.0507	0.0492	0.0509	0.0498
	50	0.0494	0.0480	0.0520	0.0475	0.0483	0.0499	0.0503

Podemos observar que o poder aumenta quando aumentamos n e/ ou K , e o desvio padrão de B_n diminui. Outra característica importante é que a média de B_n é diferente de zero.

A Tabela 4.11 reporta o poder do teste em situações em que testamos os modelos MA(1)

Tabela 4.8: Média e desvio padrão da estatística de teste, referentes ao estudo de simulação apresentado na Tabela 4.7.

K	$n \setminus \theta$	-2.00	-1.00	0.00	0.25	0.50	1.00	2.00
64	4	0.0024 0.3749	0.0010 0.3727	0.0104 0.3681	0.0032 0.3764	0.0051 0.3708	-0.0037 0.3926	-0.0065 0.3777
	12	-0.0007 0.0863	-0.0009 0.0866	0.0007 0.0873	-0.0009 0.0862	-0.0001 0.0872	0.0006 0.0929	-0.0001 0.0855
	50	0.0001 0.0189	0.0002 0.0187	0.0000 0.0189	0.0001 0.0183	0.0001 0.0189	0.0005 0.0200	0.0000 0.0190
256	4	0.0013 0.1718	-0.0007 0.1723	0.0015 0.1201	-0.0014 0.1711	-0.0003 0.1720	0.0019 0.1845	-0.0002 0.1698
	12	0.0004 0.0410	0.0001 0.0401	0.0002 0.0281	-0.0002 0.0401	0.0002 0.0402	-0.0000 0.0429	-0.0007 0.0403
	50	0.0000 0.0088	-0.0000 0.0086	-0.0000 0.0061	0.0001 0.0086	0.0000 0.0088	0.0001 0.0092	0.0000 0.0086
512	4	-0.0005 0.1206	-0.0015 0.1205	0.0000 0.1728	-0.0002 0.1207	-0.0003 0.1193	0.0002 0.1302	0.0001 0.1209
	12	-0.0003 0.0279	-0.0002 0.0278	-0.0009 0.0395	0.0001 0.0280	0.0001 0.0285	-0.0002 0.0300	0.0006 0.0283
	50	0.0000 0.0061	0.0000 0.0060	0.0000 0.0087	-0.0000 0.0060	0.0000 0.0060	-0.0000 0.0065	0.0000 0.0060
1024	4	-0.0014 0.0841	0.0002 0.0844	-0.0006 0.0859	0.0006 0.0839	-0.0020 0.0838	-0.0008 0.0906	0.0010 0.0847
	12	0.0001 0.0195	-0.0001 0.0195	-0.0002 0.0194	0.0000 0.0200	0.0000 0.0197	0.0000 0.0211	0.0004 0.0197
	50	0.0000 0.0042	0.0000 0.0042	0.0000 0.0042	-0.0000 0.0043	-0.0000 0.0042	0.0000 0.0045	0.0000 0.0042

Tabela 4.9: Poder empírico do teste a 5% utilizando-se a métrica d_{LNP} . Foram geradas $n/2$ séries a partir de um modelo AR(1) e $n/2$ séries a partir de um modelo MA(1). Os resultados são baseados em 10000 replicações.

K	$n \setminus \phi$	-0.75	-0.50	-0.25	0.00	0.25	0.375	0.50	0.75
64	4	0.1967	0.0605	0.0499	0.0537	0.0519	0.0517	0.0599	0.1885
	12	0.7473	0.1217	0.0491	0.0477	0.0511	0.0586	0.1340	0.7275
	50	1.0000	0.6681	0.0558	0.0471	0.0573	0.1563	0.6741	1.0000
256	4	0.8269	0.0958	0.0472	0.0502	0.0521	0.0507	0.0997	0.8243
	12	1.0000	0.4801	0.0505	0.0495	0.0467	0.0940	0.4733	1.0000
	50	1.0000	0.9999	0.0758	0.0488	0.0740	0.5378	0.9999	1.0000
512	4	0.9919	0.1549	0.0519	0.0506	0.0520	0.0597	0.1603	0.9915
	12	1.0000	0.8009	0.0509	0.0506	0.0548	0.1380	0.8116	1.0000
	50	1.0000	1.0000	0.1052	0.0494	0.1000	0.8735	1.0000	1.0000
1024	4	1.0000	0.2779	0.0499	0.0488	0.0516	0.0646	0.2738	1.0000
	12	1.0000	0.9840	0.0552	0.0531	0.0566	0.2345	0.9865	1.0000
	50	1.0000	1.0000	0.1588	0.0498	0.1601	0.9938	1.0000	1.0000

contra MA(2), com vários coeficientes MA. A configuração das séries é como na Tabela 4.9. Os erros do modelo são novamente ruído branco gaussiano com média zero e variância unitária. A primeira metade das séries em cada comparação segue um processo MA(2) com coeficientes

Tabela 4.10: Média e desvio padrão da estatística de teste, referentes ao estudo de simulação apresentado na Tabela 4.9.

K	$n \setminus \phi$	-0.75	-0.50	-0.25	0.00	0.25	0.375	0.50	0.75
64	4	0.5692 0.4982	0.1070 0.3927	0.0056 0.3742	0.0030 0.3695	-0.0024 0.3729	0.0262 0.3797	0.1216 0.4003	0.5761 0.5091
	12	0.4601 0.1733	0.0879 0.1077	0.0056 0.0875	-0.0011 0.0856	0.0046 0.0876	0.0268 0.0933	0.0969 0.1103	0.4823 0.1856
	50	0.4327 0.0705	0.0817 0.0333	0.0041 0.0198	0.0001 0.0188	0.0041 0.0199	0.0234 0.0242	0.0898 0.0363	0.4496 0.0765
256	4	0.6972 0.2403	0.1181 0.1859	0.0032 0.1738	0.0002 0.1720	0.0038 0.1724	0.0300 0.1777	0.1223 0.1862	0.7136 0.2475
	12	0.5709 0.0890	0.0990 0.0508	0.0044 0.0407	0.0003 0.0400	0.0034 0.0406	0.0253 0.0435	0.1004 0.0514	0.5858 0.0922
	50	0.5332 0.0373	0.0918 0.0164	0.0041 0.0091	0.0000 0.0086	0.0043 0.0092	0.0241 0.0114	0.0940 0.0171	0.5478 0.0396
512	4	0.7297 0.1701	0.1213 0.1284	0.0064 0.1219	0.0016 0.1197	0.0050 0.1207	0.0311 0.1223	0.1243 0.1291	0.7435 0.1737
	12	0.5998 0.0635	0.0997 0.0354	0.0042 0.0280	0.0000 0.0282	0.0048 0.0284	0.0264 0.0303	0.1011 0.0356	0.6081 0.0639
	50	0.5606 0.0271	0.0935 0.0117	0.0043 0.0064	0.0000 0.0061	0.0042 0.0064	0.0244 0.0079	0.0948 0.0118	0.5688 0.0274
1024	4	0.7545 0.1211	0.1243 0.0905	0.0052 0.0847	0.0007 0.0835	0.0054 0.0843	0.0324 0.0848	0.1245 0.0905	0.7616 0.1208
	12	0.6157 0.0439	0.1008 0.0248	0.0045 0.0201	-0.0001 0.0196	0.0047 0.0202	0.0263 0.0212	0.1014 0.0247	0.6215 0.0452
	50	0.5763 0.0188	0.0946 0.0083	0.0043 0.0045	-0.0000 0.0043	0.0043 0.0045	0.0245 0.0056	0.0951 0.0083	0.5812 0.0194

de θ_1 e θ_2 constantes. A segunda metade da série segue um processo MA(1), cujo coeficiente MA é também θ_1 . Assim, os dois modelos diferem apenas nos seus coeficientes MA de segunda ordem. O restante das especificações da simulação é igual aos da Tabela 4.9. O comportamento do teste é semelhante ao reparado na Tabela 4.9.

Tabela 4.11: Poder empírico do teste a 5%, com $n/2$ séries geradas a partir de um modelo MA(1) e $n/2$ séries geradas a partir de um modelo MA(2) com θ_1 fixo igual a 1. A estatística de teste é baseada na métrica d_{LNP} e os resultados são baseados em 10000 replicações.

θ_2	$n \setminus K$	64	256	512	1024
0.10	4	0.0488	0.0513	0.0614	0.0935
	12	0.0518	0.0825	0.1557	0.3573
	50	0.0751	0.3459	0.7733	0.9928
0.25	4	0.0568	0.1479	0.3216	0.6568
	12	0.1219	0.6401	0.9514	0.9998
	50	0.6106	1.0000	1.0000	1.0000
0.50	4	0.1592	0.7301	0.9715	0.9999
	12	0.5903	1.0000	1.0000	1.0000
	50	0.9999	1.0000	1.0000	1.0000

4.2.3 Simulação com a Métrica d_{ACF}

Realizamos um estudo de simulação semelhante ao anterior, mas utilizando a métrica d_{ACF} (3.33). Note que essa medida depende do número de autocorrelações L . Teoricamente, recomenda-se que $L < K/4$. Testamos alguns valores de L , variando-os no intervalo $[[0.01K], [0.1K]]$. Não obtemos um L ótimo, no entanto qualquer valor escolhido nesse intervalo conduz a resultados comparáveis aos obtidos com a medida d_{LNP} . Assintoticamente, d_{ACF} parece ser inferior a d_{LNP} , por ter convergência aparentemente mais lenta.

Nas Tabelas 4.12, 4.14, 4.16, 4.18, 4.20, 4.22, 4.24 e 4.26 apresentamos os resultados do estudo de simulação realizado com os mesmos parâmetros daquelas apresentadas nas Tabelas 4.5, 4.7, 4.9 e 4.11, mas utilizando a métrica d_{ACF} para diferentes valores de L . Para cada uma das situações consideradas, a média e o desvio padrão da estatística de teste são apresentados, respectivamente, nas Tabelas 4.13, 4.15, 4.17, 4.19, 4.21, 4.23, 4.25 e 4.27.

Tabela 4.12: Poder empírico do teste a 5%, com $n/2$ séries geradas a partir de um modelo MA(1) e $n/2$ séries geradas a partir de um modelo MA(2) com θ_1 fixo igual a 1. A estatística de teste é baseada na métrica d_{ACF} , com $L = [0.1K]$. Os resultados são baseados em 1000 replicações.

θ_2	$n \setminus K$	64	256	512	1024
0.10	4	0.0650	0.0530	0.0430	0.0470
	12	0.0620	0.0560	0.0660	0.0800
	50	0.0790	0.1230	0.2310	0.3590
0.25	4	0.0610	0.0680	0.0800	0.1090
	12	0.0900	0.1570	0.2500	0.4640
	50	0.2750	0.8110	0.9920	1.0000
0.50	4	0.0810	0.1300	0.1880	0.3680
	12	0.1860	0.5580	0.8870	0.9970
	50	0.8240	1.0000	1.0000	1.0000

Tabela 4.13: Média e desvio padrão da estatística de teste, referentes ao estudo de simulação apresentado na Tabela 4.12.

θ_2	$n \setminus K$	64	256	512	1024
0.10	4	0.0004 0.0097	0.0001 0.0017	0.0000 0.0006	0.0000 0.0002
	12	0.0003 0.0024	0.0000 0.0004	0.0000 0.0001	0.0000 0.0000
	50	0.0002 0.0006	0.0000 0.0001	0.0000 0.0000	0.0000 0.0000
0.25	4	0.0025 0.0110	0.0006 0.0018	0.0003 0.0007	0.0001 0.0002
	12	0.0015 0.0030	0.0005 0.0005	0.0002 0.0001	0.0001 0.0000
	50	0.0016 0.0010	0.0004 0.0001	0.0002 0.0000	0.0001 0.0000
0.50	4	0.0054 0.0126	0.0018 0.0022	0.0009 0.0008	0.0005 0.0003
	12	0.0052 0.0044	0.0015 0.0007	0.0008 0.0002	0.0004 0.0000
	50	0.0048 0.0016	0.0014 0.0002	0.0007 0.0000	0.0003 0.0000

Tabela 4.14: Probabilidades de erro do tipo I empíricas a 5%, com as n séries geradas a partir de um modelo MA(1). A estatística de teste é baseada na métrica d_{ACF} , com $L = 10$. Os resultados são baseados em 1000 replicações.

K	$n \setminus \theta$	-2.00	-1.00	0.00	0.25	0.50	1.00	2.00
64	4	0.0511	0.0499	0.0538	0.0527	0.0526	0.0539	0.0540
	12	0.0461	0.0445	0.0459	0.0479	0.0513	0.0448	0.0478
	50	0.0455	0.0438	0.0443	0.0431	0.0441	0.0479	0.0473
256	4	0.0528	0.0535	0.0546	0.0530	0.0519	0.0523	0.0529
	12	0.0436	0.0470	0.0486	0.0466	0.0474	0.0466	0.0462
	50	0.0437	0.0428	0.0503	0.0440	0.0451	0.0472	0.0468
512	4	0.0549	0.0539	0.0534	0.0514	0.0553	0.0563	0.0521
	12	0.0481	0.0464	0.0482	0.0465	0.0462	0.0501	0.0467
	50	0.0480	0.0444	0.0482	0.0433	0.0472	0.0474	0.0487
1024	4	0.0513	0.0527	0.0511	0.0544	0.0555	0.0532	0.0534
	12	0.0463	0.0459	0.0474	0.0478	0.0466	0.0449	0.0467
	50	0.0458	0.0440	0.0469	0.0439	0.0496	0.0500	0.0486

Tabela 4.15: Média e desvio padrão da estatística de teste, referentes ao estudo de simulação apresentado na Tabela 4.14.

K	$n \setminus \theta$	-2.00	-1.00	0.00	0.25	0.50	1.00	2.00
64	4	0.0001 0.0053	-0.0000 0.0046	-0.0000 0.0071	-0.0000 0.0044	0.0000 0.0067	-0.0000 0.0080	-0.0000 0.0068
	12	0.0000 0.0012	0.0000 0.0011	0.0000 0.0016	-0.0000 0.0010	-0.0000 0.0015	0.0000 0.0018	0.0000 0.0016
	50	0.0000 0.0002	-0.0000 0.0002	-0.0000 0.0003	-0.0000 0.0002	-0.0000 0.0003	-0.0000 0.0003	-0.0000 0.0003
256	4	0.0000 0.0014	-0.0000 0.0013	-0.0000 0.0020	-0.0000 0.0012	0.0000 0.0020	0.0000 0.0024	0.0000 0.0020
	12	-0.0000 0.0003	-0.0000 0.0003	-0.0000 0.0004	-0.0000 0.0002	-0.0000 0.0004	-0.0000 0.0005	0.0000 0.0004
	50	-0.0000 0.0000	0.0000 0.0000	0.0000 0.0001	-0.0000 0.0000	-0.0000 0.0001	-0.0000 0.0001	0.0000 0.0001
512	4	0.0000 0.0007	-0.0000 0.0006	-0.0000 0.0010	0.0000 0.0006	0.0000 0.0010	0.0000 0.0012	-0.0000 0.0010
	12	-0.0000 0.0001	-0.0000 0.0001	-0.0000 0.0002	0.0000 0.0001	0.0000 0.0002	0.0000 0.0002	0.0000 0.0002
	50	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	-0.0000 0.0000	0.0000 0.0000	-0.0000 0.0000	0.0000 0.0000
1024	4	-0.0000 0.0003	0.0000 0.0003	0.0000 0.0005	0.0000 0.0003	-0.0000 0.0005	0.0000 0.0006	-0.0000 0.0005
	12	0.0000 0.0000	-0.0000 0.0000	-0.0000 0.0001	0.0000 0.0000	0.0000 0.0001	-0.0000 0.0001	-0.0000 0.0001
	50	-0.0000 0.0000	0.0000 0.0000	-0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	-0.0000 0.0000

Tabela 4.16: Poder empírico do teste a 5%, com $n/2$ séries geradas a partir de um modelo AR(1) e $n/2$ séries geradas a partir de um modelo MA(1) com $\theta = \pm\sqrt{\phi^2/(1-\phi^2)}$. A estatística de teste é baseada na métrica \mathbf{d}_{ACF} , com $L = 10 + \lfloor 0.02K \rfloor$. Os resultados são baseados em 1000 replicações.

K	$n \setminus \phi$	-0.75	-0.50	-0.25	0.00	0.25	0.375	0.50	0.75
64	4	0.1800	0.0850	0.0440	0.0500	0.0530	0.0540	0.0670	0.1890
	12	0.5940	0.1460	0.0580	0.0430	0.0500	0.0780	0.1570	0.5770
	50	0.9990	0.5490	0.0790	0.0460	0.0850	0.2180	0.5870	1.0000
256	4	0.6470	0.1510	0.0570	0.0520	0.0550	0.0960	0.1610	0.6700
	12	1.0000	0.5190	0.0690	0.0490	0.0750	0.2210	0.5510	0.9970
	50	1.0000	0.9980	0.2130	0.0410	0.1860	0.8340	1.0000	1.0000
512	4	0.9450	0.3260	0.0590	0.0510	0.0560	0.1040	0.3270	0.9520
	12	1.0000	0.9060	0.0940	0.0460	0.1050	0.4210	0.8760	1.0000
	50	1.0000	1.0000	0.3970	0.0440	0.3860	0.9920	1.0000	1.0000
1024	4	1.0000	0.6040	0.0620	0.0420	0.0620	0.1990	0.6060	1.0000
	12	1.0000	1.0000	0.1550	0.0410	0.1490	0.7260	1.0000	1.0000
	50	1.0000	1.0000	0.7180	0.0530	0.7130	1.0000	1.0000	1.0000

Tabela 4.17: Média e desvio padrão da estatística de teste, referentes ao estudo de simulação apresentado na Tabela 4.16.

K	$n \setminus \phi$	-0.75	-0.50	-0.25	0.00	0.25	0.375	0.50	0.75
64	4	0.0337 0.0285	0.0048 0.0107	0.0002 0.0054	0.0000 0.0042	0.0000 0.0051	0.0010 0.0066	0.0037 0.0092	0.0261 0.0230
	12	0.0262 0.0113	0.0033 0.0034	0.0001 0.0013	0.0000 0.0009	0.0001 0.0012	0.0008 0.0019	0.0029 0.0030	0.0210 0.0091
	50	0.0249 0.0053	0.0030 0.0013	0.0001 0.0003	0.0000 0.0002	0.0001 0.0003	0.0007 0.0006	0.0027 0.0012	0.0198 0.0040
256	4	0.0307 0.0128	0.0034 0.0034	0.0001 0.0013	0.0000 0.0010	0.0001 0.0013	0.0010 0.0020	0.0033 0.0034	0.0288 0.0116
	12	0.0250 0.0055	0.0027 0.0012	0.0001 0.0003	0.0000 0.0002	0.0001 0.0003	0.0007 0.0006	0.0027 0.0012	0.0233 0.0051
	50	0.0231 0.0024	0.0026 0.0005	0.0001 0.0001	0.0000 0.0000	0.0001 0.0001	0.0007 0.0002	0.0025 0.0005	0.0219 0.0022
512	4	0.0239 0.0069	0.0026 0.0017	0.0001 0.0006	0.0000 0.0004	0.0001 0.0006	0.0007 0.0010	0.0026 0.0016	0.0233 0.0066
	12	0.0195 0.0031	0.0021 0.0006	0.0001 0.0001	0.0000 0.0001	0.0001 0.0001	0.0005 0.0003	0.0021 0.0007	0.0191 0.0030
	50	0.0183 0.0013	0.0020 0.0003	0.0001 0.0000	0.0000 0.0000	0.0000 0.0000	0.0005 0.0001	0.0020 0.0003	0.0178 0.0014
1024	4	0.0164 0.0033	0.0018 0.0008	0.0000 0.0002	0.0000 0.0001	0.0000 0.0002	0.0005 0.0004	0.0018 0.0008	0.0164 0.0034
	12	0.0135 0.0014	0.0015 0.0003	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	0.0004 0.0001	0.0015 0.0003	0.0134 0.0014
	50	0.0127 0.0006	0.0014 0.0001	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	0.0003 0.0000	0.0014 0.0001	0.0125 0.0006

Tabela 4.18: Poder empírico do teste a 5%, com $n/2$ séries geradas a partir de um modelo MA(1) e $n/2$ séries geradas a partir de um modelo MA(2) com θ_1 fixo igual a 1. A estatística de teste é baseada na métrica d_{ACF} , com $L = 10 + \lfloor 0.02K \rfloor$. Os resultados são baseados em 1000 replicações.

θ_2	$n \setminus K$	64	256	512	1024
0.10	4	0.0570	0.0560	0.0510	0.0590
	12	0.0440	0.0750	0.0680	0.1010
	50	0.0810	0.1540	0.3180	0.5580
0.25	4	0.0660	0.0700	0.1050	0.1600
	12	0.0820	0.1990	0.3880	0.7390
	50	0.2320	0.8820	0.9980	1.0000
0.50	4	0.0750	0.1610	0.3300	0.6300
	12	0.1700	0.6840	0.9610	1.0000
	50	0.7670	1.0000	1.0000	1.0000

Tabela 4.19: Média e desvio padrão da estatística de teste, referentes ao estudo de simulação apresentado na Tabela 4.18.

θ_2	$n \setminus K$	64	256	512	1024
0.10	4	0.0005 0.0080	0.0001 0.0021	0.0000 0.0010	0.0001 0.0004
	12	0.0002 0.0019	0.0001 0.0005	0.0001 0.0002	0.0000 0.0001
	50	0.0001 0.0004	0.0001 0.0001	0.0001 0.0000	0.0000 0.0000
0.25	4	0.0016 0.0093	0.0010 0.0025	0.0008 0.0012	0.0005 0.0005
	12	0.0009 0.0022	0.0008 0.0007	0.0006 0.0003	0.0004 0.0001
	50	0.0009 0.0007	0.0008 0.0002	0.0006 0.0001	0.0004 0.0000
0.50	4	0.0035 0.0103	0.0031 0.0031	0.0024 0.0015	0.0016 0.0006
	12	0.0031 0.0030	0.0025 0.0010	0.0019 0.0005	0.0013 0.0002
	50	0.0028 0.0010	0.0023 0.0004	0.0018 0.0002	0.0012 0.0001

Tabela 4.20: Probabilidades de erro do tipo I empíricos a 5%, com as n séries geradas a partir de um modelo AR(1). A estatística de teste é baseada na métrica d_{ACF} , com $[0.1K]$. Os resultados são baseados em 10000 replicações.

K	$n \setminus \phi$	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75
64	4	0.0599	0.0588	0.0572	0.0518	0.0542	0.0559	0.0611
	12	0.0491	0.0469	0.0472	0.0467	0.0459	0.0493	0.0502
	50	0.0530	0.0495	0.0494	0.0467	0.0498	0.0491	0.0513
256	4	0.0553	0.0560	0.0520	0.0505	0.0531	0.0530	0.0589
	12	0.0485	0.0467	0.0454	0.0478	0.0488	0.0470	0.0474
	50	0.0496	0.0467	0.0469	0.0463	0.0451	0.0441	0.0494
512	4	0.0548	0.0534	0.0486	0.0505	0.0475	0.0525	0.0560
	12	0.0472	0.0490	0.0446	0.0491	0.0471	0.0492	0.0488
	50	0.0473	0.0487	0.0476	0.0490	0.0441	0.0453	0.0475
1024	4	0.0554	0.0553	0.0505	0.0496	0.0508	0.0522	0.0524
	12	0.0478	0.0473	0.0484	0.0477	0.0491	0.0461	0.0467
	50	0.0454	0.0485	0.0506	0.0474	0.0484	0.0494	0.0448

Tabela 4.21: Média e desvio padrão da estatística de teste, referentes ao estudo de simulação apresentado na Tabela 4.20.

K	$n \setminus \phi$	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75
64	4	-0.0002 0.0210	0.0000 0.0120	-0.0000 0.0072	0.0000 0.0057	0.0000 0.0068	-0.0000 0.0110	-0.0002 0.0200
	12	-0.0000 0.0048	0.0000 0.0028	-0.0000 0.0016	0.0000 0.0013	0.0000 0.0016	0.0000 0.0026	0.0000 0.0048
	50	-0.0000 0.0010	-0.0000 0.0006	-0.0000 0.0004	0.0000 0.0003	0.0000 0.0003	-0.0000 0.0005	-0.0000 0.0010
256	4	0.0001 0.0067	-0.0000 0.0022	0.0000 0.0011	-0.0000 0.0008	-0.0000 0.0011	0.0000 0.0021	0.0000 0.0060
	12	0.0000 0.0015	0.0000 0.0005	0.0000 0.0003	0.0000 0.0002	0.0000 0.0003	0.0000 0.0005	-0.0000 0.0014
	50	0.0000 0.0003	-0.0000 0.0001	0.0000 0.0001	0.0000 0.0000	0.0000 0.0001	-0.0000 0.0001	0.0000 0.0003
512	4	0.0000 0.0027	-0.0000 0.0009	-0.0000 0.0004	-0.0000 0.0003	0.0000 0.0004	0.0000 0.0008	-0.0000 0.0025
	12	-0.0000 0.0006	-0.0000 0.0002	-0.0000 0.0001	0.0000 0.0001	-0.0000 0.0001	0.0000 0.0002	-0.0000 0.0006
	50	0.0000 0.0001	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	0.0000 0.0001
1024	4	-0.0000 0.0010	0.0000 0.0003	0.0000 0.0001	-0.0000 0.0001	-0.0000 0.0001	0.0000 0.0003	0.0000 0.0010
	12	0.0000 0.0002	-0.0000 0.0001	0.0000 0.0000	-0.0000 0.0000	-0.0000 0.0000	0.0000 0.0001	-0.0000 0.0002
	50	0.0000 0.0001	-0.0000 0.0000	0.0000 0.0000	-0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	0.0000 0.0001

Tabela 4.22: Probabilidades de erro do tipo I empíricas a 5%, com as n séries geradas a partir de um modelo MA(1). A estatística de teste é baseada na métrica d_{ACF} , com $[0.1K]$. Os resultados são baseados em 10000 replicações.

K	$n \setminus \theta$	-2.00	-1.00	0.00	0.25	0.50	1.00	2.00
64	4	0.0542	0.0531	0.0573	0.0518	0.0547	0.0531	0.0561
	12	0.0452	0.0456	0.0479	0.0467	0.0475	0.0475	0.0476
	50	0.0483	0.0461	0.0493	0.0467	0.0502	0.0489	0.0507
256	4	0.0533	0.0504	0.0537	0.0505	0.0516	0.0514	0.0492
	12	0.0469	0.0495	0.0461	0.0478	0.0486	0.0463	0.0475
	50	0.0478	0.0462	0.0454	0.0463	0.0433	0.0453	0.0444
512	4	0.0493	0.0490	0.0488	0.0505	0.0496	0.0495	0.0507
	12	0.0460	0.0469	0.0459	0.0491	0.0480	0.0515	0.0484
	50	0.0477	0.0478	0.0471	0.0490	0.0444	0.0475	0.0465
1024	4	0.0505	0.0493	0.0491	0.0496	0.0516	0.0497	0.0516
	12	0.0486	0.0490	0.0493	0.0477	0.0489	0.0495	0.0483
	50	0.0500	0.0478	0.0508	0.0474	0.0486	0.0503	0.0490

Tabela 4.23: Média e desvio padrão da estatística de teste, referentes ao estudo de simulação apresentado na Tabela 4.22.

K	$n \setminus \theta$	-2.00	-1.00	0.00	0.25	0.50	1.00	2.00
64	4	-0.0000 0.0065	0.0000 0.0059	-0.0001 0.0086	0.0000 0.0057	0.0000 0.0082	-0.0000 0.0095	-0.0000 0.0082
	12	-0.0000 0.0015	0.0000 0.0014	-0.0000 0.0020	0.0000 0.0013	0.0000 0.0019	0.0000 0.0022	-0.0000 0.0019
	50	-0.0000 0.0003	0.0000 0.0003	-0.0000 0.0004	0.0000 0.0003	0.0000 0.0004	-0.0000 0.0005	0.0000 0.0004
256	4	-0.0000 0.0010	-0.0000 0.0008	0.0000 0.0014	-0.0000 0.0008	-0.0000 0.0014	-0.0000 0.0016	-0.0000 0.0013
	12	0.0000 0.0002	0.0000 0.0002	0.0000 0.0003	0.0000 0.0002	0.0000 0.0003	-0.0000 0.0004	0.0000 0.0003
	50	0.0000 0.0000	0.0000 0.0000	0.0000 0.0001	0.0000 0.0000	0.0000 0.0001	-0.0000 0.0001	0.0000 0.0001
512	4	0.0000 0.0004	-0.0000 0.0003	-0.0000 0.0005	-0.0000 0.0003	0.0000 0.0005	0.0000 0.0006	-0.0000 0.0005
	12	-0.0000 0.0001	0.0000 0.0001	-0.0000 0.0001	0.0000 0.0001	-0.0000 0.0001	0.0000 0.0001	0.0000 0.0001
	50	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000
1024	4	0.0000 0.0001	0.0000 0.0001	0.0000 0.0002	-0.0000 0.0001	-0.0000 0.0002	0.0000 0.0002	-0.0000 0.0002
	12	0.0000 0.0000	-0.0000 0.0000	0.0000 0.0000	-0.0000 0.0000	-0.0000 0.0000	0.0000 0.0001	0.0000 0.0000
	50	0.0000 0.0000	-0.0000 0.0000	0.0000 0.0000	-0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000

Tabela 4.24: Poder empírico do teste a 5%, com $n/2$ séries geradas a partir de um modelo AR(1) e $n/2$ séries geradas a partir de um modelo MA(1) com $\theta = \pm\sqrt{\phi^2/(1-\phi^2)}$. A estatística de teste é baseada na métrica d_{ACF} , com $L = \lfloor 0.1K \rfloor$. Os resultados são baseados em 10000 replicações.

K	$n \setminus \phi$	-0.75	-0.50	-0.25	0.00	0.25	0.375	0.50	0.75
64	4	0.2384	0.0772	0.0555	0.0539	0.0554	0.0597	0.0777	0.2116
	12	0.6616	0.1640	0.0560	0.0441	0.0553	0.0892	0.1561	0.6094
	50	1.0000	0.6052	0.0893	0.0433	0.0903	0.2339	0.5986	0.9999
256	4	0.5801	0.1346	0.0540	0.0497	0.0547	0.0718	0.1426	0.6000
	12	0.9987	0.5036	0.0640	0.0482	0.0614	0.1671	0.4994	0.9989
	50	1.0000	0.9993	0.1789	0.0465	0.1683	0.7730	0.9996	1.0000
512	4	0.8847	0.2143	0.0523	0.0524	0.0521	0.0843	0.2201	0.9026
	12	1.0000	0.8350	0.0719	0.0482	0.0739	0.2858	0.8252	1.0000
	50	1.0000	1.0000	0.2960	0.0496	0.2967	0.9783	1.0000	1.0000
1024	4	0.9982	0.4054	0.0544	0.0472	0.0538	0.1197	0.3896	0.9962
	12	1.0000	0.9915	0.0979	0.0512	0.0975	0.5214	0.9909	1.0000
	50	1.0000	1.0000	0.5259	0.0497	0.5332	1.0000	1.0000	1.0000

Tabela 4.25: Média e desvio padrão da estatística de teste, referentes ao estudo de simulação apresentado na Tabela 4.24.

K	$n \setminus \theta$	-0.75	-0.50	-0.25	0.00	0.25	0.375	0.50	0.75
64	4	0.0542 0.0405	0.0068 0.0144	0.0003 0.0072	-0.0000 0.0056	0.0003 0.0069	0.0016 0.0088	0.0061 0.0133	0.0439 0.0353
	12	0.0443 0.0179	0.0055 0.0053	0.0003 0.0018	-0.0000 0.0013	0.0003 0.0017	0.0014 0.0027	0.0049 0.0048	0.0361 0.0155
	50	0.0416 0.0081	0.0052 0.0022	0.0003 0.0005	-0.0000 0.0003	0.0002 0.0005	0.0013 0.0010	0.0046 0.0020	0.0339 0.0070
256	4	0.0187 0.0082	0.0021 0.0024	0.0001 0.0011	-0.0000 0.0008	0.0001 0.0011	0.0006 0.0015	0.0021 0.0024	0.0177 0.0077
	12	0.0153 0.0034	0.0017 0.0008	0.0001 0.0003	-0.0000 0.0002	0.0001 0.0003	0.0005 0.0005	0.0017 0.0008	0.0145 0.0032
	50	0.0143 0.0015	0.0016 0.0004	0.0001 0.0001	0.0000 0.0000	0.0001 0.0001	0.0004 0.0002	0.0016 0.0003	0.0135 0.0014
512	4	0.0097 0.0032	0.0011 0.0009	0.0001 0.0004	0.0000 0.0003	0.0001 0.0004	0.0003 0.0006	0.0011 0.0009	0.0094 0.0030
	12	0.0080 0.0013	0.0009 0.0003	0.0000 0.0001	0.0000 0.0001	0.0000 0.0001	0.0002 0.0002	0.0009 0.0003	0.0078 0.0013
	50	0.0074 0.0006	0.0008 0.0001	0.0000 0.0000	-0.0000 0.0000	0.0000 0.0000	0.0002 0.0001	0.0008 0.0001	0.0072 0.0006
1024	4	0.0050 0.0012	0.0006 0.0003	0.0000 0.0001	-0.0000 0.0001	0.0000 0.0001	0.0002 0.0002	0.0006 0.0003	0.0049 0.0011
	12	0.0041 0.0005	0.0005 0.0001	0.0000 0.0000	0.0000 0.0000	0.0000 0.0000	0.0001 0.0001	0.0005 0.0001	0.0040 0.0005
	50	0.0038 0.0002	0.0004 0.0000	0.0000 0.0000	-0.0000 0.0000	0.0000 0.0000	0.0001 0.0000	0.0004 0.0000	0.0038 0.0002

Tabela 4.26: Poder empírico do teste a 5%, com $n/2$ séries geradas a partir de um modelo MA(1) e $n/2$ séries geradas a partir de um modelo MA(2) com θ_1 fixo igual a 1. A estatística de teste é baseada na métrica d_{ACF} , com $L = \lfloor 0.1K \rfloor$. Os resultados são baseados em 10000 replicações.

θ_2	n \ K	64	256	512	1024
0.10	4	0.0551	0.0547	0.0534	0.0532
	12	0.0586	0.0584	0.0616	0.0738
	50	0.0812	0.1339	0.2064	0.3574
0.25	4	0.0627	0.0657	0.0753	0.0969
	12	0.0898	0.1619	0.2498	0.4687
	50	0.2776	0.8352	0.9897	1.0000
0.50	4	0.0796	0.1278	0.1868	0.3227
	12	0.1952	0.5604	0.8663	0.9953
	50	0.8299	1.0000	1.0000	1.0000

Tabela 4.27: Média e desvio padrão da estatística de teste, referentes ao estudo de simulação apresentado na Tabela 4.26.

θ_2	n \ K	64	256	512	1024
0.10	4	0.0002 0.0100	0.0001 0.0017	0.0001 0.0006	0.0000 0.0002
	12	0.0003 0.0025	0.0001 0.0004	0.0000 0.0002	0.0000 0.0001
	50	0.0003 0.0006	0.0001 0.0001	0.0000 0.0000	0.0000 0.0000
0.25	4	0.0023 0.0113	0.0006 0.0020	0.0003 0.0007	0.0002 0.0003
	12	0.0018 0.0033	0.0005 0.0005	0.0003 0.0002	0.0001 0.0001
	50	0.0016 0.0011	0.0005 0.0002	0.0003 0.0001	0.0001 0.0000
0.50	4	0.0065 0.0137	0.0019 0.0023	0.0010 0.0009	0.0005 0.0003
	12	0.0053 0.0044	0.0016 0.0007	0.0008 0.0003	0.0004 0.0001
	50	0.0049 0.0017	0.0015 0.0003	0.0007 0.0001	0.0004 0.0000

4.2.4 O Caso Não Linear vs Não Linear

Consideramos também o caso não linear contra não linear. Na teoria, as métricas \mathbf{d}_{LNP} , \mathbf{d}_{NP} e \mathbf{d}_{ACF} dependem da condição de estacionariedade e linearidade para que aconteça a normalidade assintótica. No entanto, a métrica \mathbf{d}_P pode ser utilizada desde que a série atenda as condições do Teorema 3.2.1. Na Tabela 4.28 estão os resultados do poder da estatística de teste quando testamos dois modelos não lineares. Neste caso, consideramos o modelo ARCH(1). Um dos modelos é fixo, com coeficiente autorregressivo $\phi_1 = 0.25$. Para o segundo modelo, variamos o coeficiente para valores $\phi_1^* = 0.00, 0.10, 0.25, 0.50, 0.75$ e 0.95 . Diferentes tamanhos amostrais são considerados. Na Figura 4.1 apresentamos séries simuladas de tamanho 256 que caracterizam os modelos estudados.

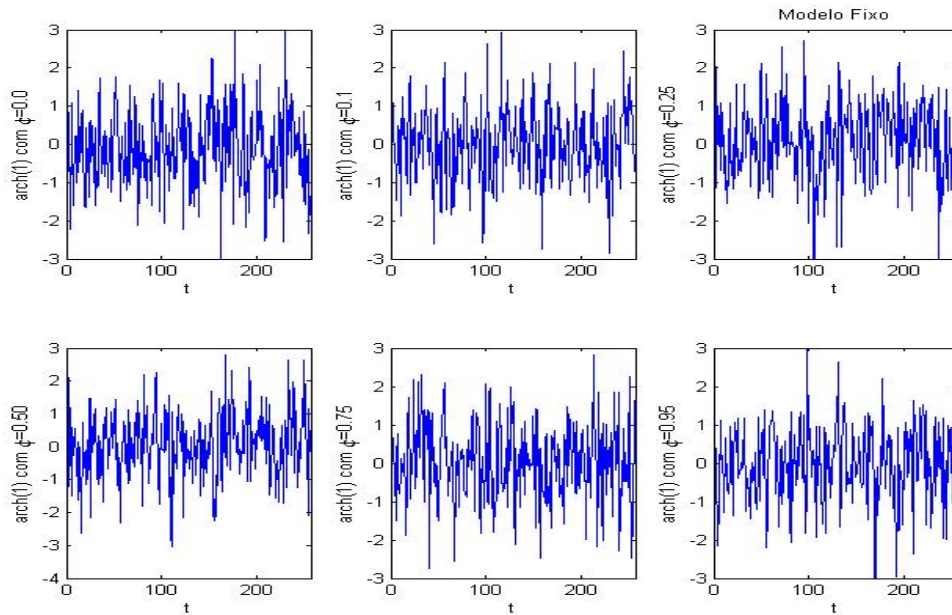


Figura 4.1: Gráfico das séries não lineares utilizadas no estudo da métrica \mathbf{d}_{LNP} e \mathbf{d}_P . O modelo ARCH(1) é considerado com coeficiente assumindo valores 0.0, 0.1, 0.25, 0.50, 0.75, 0.95.

Como previsto na teoria, a métrica \mathbf{d}_P é capaz de diferenciar esses modelos não lineares, enquanto métricas baseadas no periodograma normalizado não têm essa propriedade.

4.2.5 Estudo de Simulação para Classificação

A classificação é baseada somente nos valores da estatística de teste. A série a ser classificada é colocada em ambos os grupos e testada. O procedimento que resultar em um maior

Tabela 4.28: Poder empírico a 5%. Caso não linear vs não linear com as métricas d_{LNP} e d_P . Os modelos considerados são ARCH(1). Metade das n séries são geradas a partir de um modelo ARCH(1) fixo, em que o coeficiente $\phi_1 = 0.25$. A segunda metade das n séries são geradas a partir de um modelo ARCH(1) com coeficiente ϕ_1^* .

		d_{LNP}				d_P			
ϕ_1^*	$n \setminus K$	64	256	512	1024	64	256	512	1024
0.00	4	0.052	0.049	0.047	0.052	0.061	0.071	0.103	0.166
	12	0.053	0.045	0.044	0.054	0.099	0.275	0.471	0.800
	50	0.043	0.044	0.053	0.047	0.440	0.979	1.000	1.000
0.10	4	0.051	0.047	0.048	0.052	0.057	0.062	0.069	0.076
	12	0.051	0.050	0.046	0.046	0.064	0.094	0.141	0.272
	50	0.037	0.042	0.051	0.047	0.165	0.527	0.817	0.995
0.25	4	0.049	0.055	0.046	0.052	0.050	0.053	0.040	0.046
	12	0.053	0.057	0.043	0.046	0.052	0.051	0.053	0.056
	50	0.044	0.050	0.048	0.050	0.054	0.044	0.050	0.043
0.50	4	0.044	0.042	0.052	0.050	0.056	0.052	0.025	0.218
	12	0.050	0.045	0.043	0.043	0.042	0.155	0.454	0.750
	50	0.052	0.061	0.046	0.058	0.240	0.895	0.997	1.000
0.75	4	0.056	0.045	0.048	0.047	0.043	0.023	0.011	0.025
	12	0.046	0.044	0.049	0.051	0.045	0.066	0.072	0.003
	50	0.041	0.038	0.043	0.046	0.038	0.007	0.058	0.110
0.95	4	0.057	0.057	0.047	0.050	0.005	0.007	0.032	0.001
	12	0.045	0.045	0.045	0.053	0.008	0.002	0.013	0.007
	50	0.046	0.043	0.053	0.047	0.001	0.020	0.022	0.007

valor da estatística indica que aquela configuração de grupo é mais dissimilar. Basicamente, a série é mais similar ao grupo em que ela foi colocada e menos similar ao outro grupo. Nesse estudo, consideramos a métrica d_{LNP} e diferentes configurações de modelos. Na Tabela 4.29 consideramos o caso Estacionário contra Estacionário. As n_1 primeiras séries foram geradas a partir de um modelo AR(1), com $\phi = 0.3$, e as outras n_2 séries foram geradas a partir de um modelo AR(1), com $\phi = 0.8$. A série classificada foi gerada a partir do primeiro modelo. A séries simuladas e apresentadas na Figura 4.2 representam os modelos utilizados para o estudo de classificação.

Na Tabela 4.30 consideramos o caso Estacionário contra Não-Estacionário. As n_1 primeiras séries foram geradas a partir de um modelo AR(1), com $\phi = 0.95$, e as outras n_2 séries foram geradas a partir de um modelo ARIMA(0,1,0). A série classificada foi gerada a partir do segundo modelo.

Na Tabela 4.31 consideramos o caso Não Estacionário contra Estacionário. As n_1 primeiras séries foram geradas a partir de um modelo ARIMA(0,1,0) e as outras n_2 séries foram geradas a partir de um modelo AR(1), com $\phi = 0.95$. A série classificada foi gerada a partir do

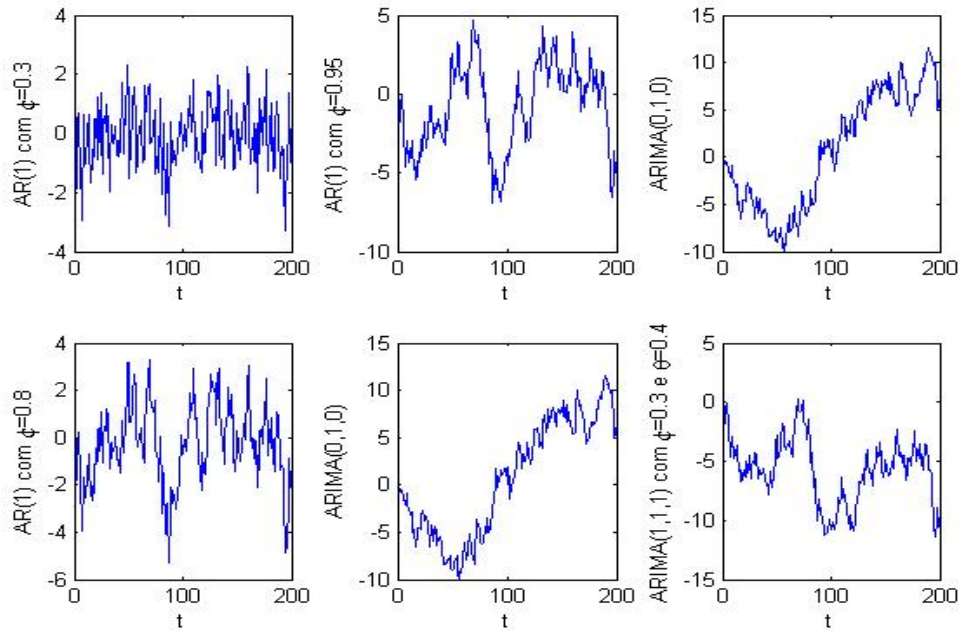


Figura 4.2: Gráfico das séries utilizadas para classificação. Na primeira coluna estão representadas as séries utilizadas para o estudo de classificação na Tabela 4.29, na qual a primeira série é gerada a partir de um modelo AR(1) com $\phi = 0.3$ e a segunda série é gerada a partir de um modelo AR(1) com $\phi = 0.8$. Na segunda coluna estão representadas as séries utilizadas para o estudo de classificação na Tabela 4.30, na qual a primeira série é gerada a partir de um modelo AR(1) com $\phi = 0.95$ e a segunda série é gerada a partir de um modelo ARIMA(0,1,0). Na terceira coluna estão representadas as séries utilizadas para o estudo de classificação na Tabela 4.31, na qual a primeira série é gerada a partir de um modelo ARIMA(0,1,0) e a segunda série é gerada a partir de um modelo ARIMA(1,1,1), com $\phi = 0.3$ e $\theta = -0.4$.

segundo modelo.

Na Tabela 4.32 consideramos o caso Não Estacionário contra Não Estacionário. As n_1 primeiras séries foram geradas a partir de um modelo ARIMA(1,1,1), com $\phi = 0.3$ e $\theta = -0.4$, e n_2 séries foram geradas a partir de um modelo ARIMA(0,1,0). A série classificada foi gerada a partir do segundo modelo. Em todas as situações foram realizadas 1000 replicações.

Podemos observar que os resultados para classificação não são assintóticos em n o que já era esperado, a medida que acrescentamos somente uma série. No entanto, em todos os casos, os resultados são satisfatórios, exceto o caso Não Estacionário contra Não Estacionário.

Tabela 4.29: Percentual de acerto do método de classificação: Estacionário vs Estacionário. As n_1 primeiras séries foram geradas a partir de um modelo AR(1), com $\phi = 0.3$, e n_2 séries foram geradas a partir de um modelo AR(1), com $\phi = 0.8$. A série classificada foi gerada a partir do primeiro modelo. Foram realizadas 1000 replicações.

K	$n_1 \setminus n_2$	4	12	20	50	100
25	4	74.80	79.30	81.60	82.50	83.30
	12	78.60	84.00	84.50	85.20	86.00
	20	79.70	84.20	82.90	89.30	86.20
	50	79.00	83.00	85.80	86.70	86.60
	100	78.80	85.80	84.50	86.30	88.40
50	4	90.80	90.50	93.80	93.90	94.10
	12	91.20	94.20	94.20	94.70	95.80
	20	91.90	94.40	96.40	95.60	95.10
	50	93.00	94.70	96.00	95.70	96.60
	100	92.60	95.10	95.60	96.80	95.40
100	4	98.50	98.60	99.40	98.90	99.20
	12	98.40	99.10	99.10	99.80	99.50
	20	98.60	99.40	99.60	100	99.90
	50	98.60	99.30	99.60	99.90	99.70
	100	98.80	99.50	99.40	99.60	99.50
200	4	100	100	100	100	100
	12	99.90	100	100	100	100
	20	100	100	100	100	100
	50	100	99.90	100	100	100
	100	100	100	100	100	100
500	4	100	100	100	100	100
	12	100	100	100	100	100
	20	100	100	100	100	100
	50	100	100	100	100	100
	100	100	100	100	100	100
1000	4	100	100	100	100	100
	12	100	100	100	100	100
	20	100	100	100	100	100
	50	100	100	100	100	100
	100	100	100	100	100	100

Tabela 4.30: Percentual de acerto do método de classificação: Estacionário vs Não Estacionário. As n_1 primeiras séries foram geradas a partir de um modelo AR(1), com $\phi = 0.95$, e n_2 séries foram geradas a partir de um modelo ARIMA(0,1,0). A série classificada foi gerada a partir do segundo modelo. Foram realizadas 1000 replicações.

K	$n_1 \setminus n_2$	4	12	20	50	100
25	4	50.2	53.80	53.50	53.80	50.60
	12	46.4	50.40	49.50	50.80	47.20
	20	49.8	52.00	51.60	51.80	49.90
	50	48.4	46.30	47.40	48.00	50.10
	100	45.6	52.40	49.90	48.60	50.30
50	4	50.9	50.70	50.60	53.30	51.20
	12	47.7	48.50	51.20	50.30	55.20
	20	46.5	53.10	49.40	48.80	53.50
	50	47.1	49.70	50.50	52.10	49.40
	100	47.5	50.00	48.00	50.50	50.10
100	4	53.1	60.10	60.80	56.70	61.10
	12	55.2	55.00	56.80	59.00	60.50
	20	50.2	56.10	60.50	63.00	60.90
	50	50.1	59.40	55.90	61.60	61.30
	100	54.4	55.70	58.80	63.00	59.00
200	4	72.2	77.20	79.20	80.50	77.60
	12	72.7	75.50	76.00	80.30	78.70
	20	71.6	76.60	78.50	77.00	81.20
	50	72.6	77.50	78.40	82.10	80.70
	100	73.3	76.80	76.00	78.80	80.30
500	4	97.5	99.10	98.90	98.70	99.20
	12	98.3	98.50	98.90	99.30	99.30
	20	98.2	99.20	98.70	98.90	99.30
	50	98.4	98.60	98.40	99.40	99.20
	100	98.6	99.20	98.90	99.00	99.30
1000	4	100	100	100	100	100
	12	100	100	100	100	100
	20	100	100	100	100	100
	50	100	100	100	100	100
	100	100	100	100	100	100
2000	4	100	100	100	100	100
	12	100	100	100	100	100
	20	100	100	100	100	100
	50	100	100	100	100	100
	100	100	100	100	100	100

Tabela 4.31: Percentual de acerto do método de classificação: Não Estacionário vs Estacionário. As n_1 primeiras séries foram geradas a partir de um modelo ARIMA(0,1,0) e n_2 séries foram geradas a partir de um modelo AR(1), com $\phi = 0.95$. A série classificada foi gerada a partir do segundo modelo. Foram realizadas 1000 replicações.

K	$n_1 \setminus n_2$	4	12	20	50	100
25	4	52.00	50.20	49.80	49.30	49.90
	12	52.70	49.40	47.70	47.80	46.60
	20	51.90	50.50	49.10	47.50	47.60
	50	53.50	49.20	50.70	48.00	50.00
	100	53.30	49.10	51.30	48.20	49.20
50	4	52.70	50.90	50.30	47.80	47.60
	12	53.30	50.30	47.10	45.30	45.20
	20	55.40	53.90	50.00	48.50	47.80
	50	57.10	53.60	51.60	49.00	48.60
	100	55.50	54.90	53.50	50.50	49.00
100	4	56.10	55.00	53.70	53.90	53.80
	12	57.90	56.20	54.70	55.70	53.90
	20	58.90	57.10	56.20	57.20	54.70
	50	59.10	58.80	58.80	57.10	56.40
	100	60.00	59.30	59.10	58.30	57.90
200	4	67.90	67.90	67.70	67.10	66.90
	12	71.50	71.80	72.10	71.50	70.90
	20	72.70	73.50	72.20	72.50	71.90
	50	73.40	73.30	72.50	71.80	72.10
	100	73.80	73.40	72.90	71.60	71.20
500	4	88.30	87.50	87.60	88.20	87.80
	12	88.30	88.00	88.40	88.00	88.00
	20	88.60	87.80	87.90	87.60	87.60
	50	88.90	88.00	88.50	88.00	88.00
	100	88.50	87.70	88.20	87.60	87.60
1000	4	96.60	97.00	96.80	96.90	96.70
	12	96.60	97.30	97.00	97.00	97.00
	20	96.50	97.20	97.20	97.00	97.20
	50	96.50	97.10	97.00	97.10	97.10
	100	96.50	97.10	97.00	97.10	97.10
2000	4	99.00	99.30	99.40	99.40	99.30
	12	99.00	99.40	99.40	99.30	99.50
	20	99.00	99.40	99.40	99.40	99.50
	50	99.10	99.40	99.40	99.40	99.60
	100	99.00	99.40	99.40	99.50	99.50

Tabela 4.32: Percentual de acerto do método de classificação: Não Estacionário vs Não Estacionário. As n_1 primeiras séries foram geradas a partir de um modelo ARIMA(1,1,1), com $\phi = 0.3$ e $\theta = -0.4$, e n_2 séries foram geradas a partir de um modelo ARIMA(0,1,0). A série classificada foi gerada a partir do segundo modelo. Foram realizadas 1000 replicações.

K	$n_1 \setminus n_2$	4	12	20	50	100
25	4	53.10	50.30	51.20	50.50	49.90
	12	54.00	50.80	51.70	50.70	51.60
	20	53.20	53.00	53.20	52.50	52.80
	50	54.20	51.30	52.20	51.20	52.50
	100	53.70	52.80	53.30	53.50	53.30
50	4	53.70	51.20	50.20	48.30	48.50
	12	56.10	53.40	50.70	48.40	48.50
	20	56.80	54.60	53.70	50.90	51.50
	50	56.50	55.60	55.50	52.80	52.20
	100	56.60	57.10	56.30	52.90	53.40
100	4	53.40	51.80	49.80	49.30	48.00
	12	54.80	52.70	50.60	50.10	48.80
	20	55.10	54.50	53.10	51.70	50.10
	50	56.20	56.40	56.80	54.20	52.20
	100	56.80	58.10	56.70	54.30	54.80
200	4	52.00	51.00	50.10	48.30	47.80
	12	56.20	53.70	54.00	51.40	52.20
	20	57.60	55.90	56.40	53.80	52.20
	50	57.60	57.50	57.70	55.10	55.10
	100	58.20	58.40	59.20	56.10	56.60
500	4	52.50	47.80	47.80	47.20	47.60
	12	58.80	55.60	53.90	55.00	54.40
	20	56.80	54.90	54.40	54.10	54.30
	50	59.60	56.90	54.60	56.60	56.90
	100	59.70	56.10	56.10	58.10	57.90
1000	4	53.30	48.00	47.00	47.60	47.40
	12	56.70	52.50	53.10	51.80	51.10
	20	57.20	54.70	55.10	52.50	53.00
	50	58.10	57.60	57.50	57.90	58.10
	100	59.80	58.10	59.40	58.50	58.90
2000	4	50.50	46.10	45.70	45.60	44.50
	12	56.80	53.00	52.90	51.60	51.30
	20	59.00	56.80	55.60	54.90	54.70
	50	60.50	58.20	59.20	58.10	56.90
	100	59.90	57.80	58.40	56.90	56.10
5000	4	51.50	48.00	46.30	43.90	44.60
	12	57.80	56.90	54.40	51.40	50.70
	20	58.80	57.60	56.30	54.00	53.60
	50	59.60	60.30	59.50	57.80	57.00
	100	59.30	62.00	60.80	59.40	58.60

4.2.6 Estudos de Simulação com Três Grupos

Realizamos um estudo de simulação para analisar o comportamento da estatística de teste na presença de 3 grupos, cada um deles de tamanho $n_g = n/3$, para $g = 1, 2, 3$. Primeiramente, para verificar o comportamento de B_n sob H_0 , consideramos 3 grupos de séries AR(1), cujo parâmetro autorregressivo ϕ varia de -0.75 á 0.75 . Podemos observar na Tabela 4.33 que, quando n e/ou K crescem, a média e o desvio padrão diminuem e as probabilidades de erro do Tipo I, para testes a 5%, estão bem próximos de 0.05. A média e o desvio padrão da estatística de teste, para essa situação, são apresentados na Tabela 4.34;

Tabela 4.33: Probabilidades de erro do tipo I a 5% quando consideramos 3 grupos iguais, utilizando-se a métrica \mathbf{d}_{LNP} . As séries foram geradas a partir de um modelo AR(1). Os resultados são baseados em 1000 replicações.

K	$n \setminus \phi$	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75
64	6	0.0530	0.0460	0.0490	0.0620	0.0430	0.0440	0.0530
	12	0.0490	0.0520	0.0480	0.0540	0.0560	0.0550	0.0460
	45	0.0480	0.0460	0.0450	0.0530	0.0480	0.0550	0.0560
256	6	0.0480	0.0440	0.0490	0.0540	0.0530	0.0480	0.0480
	12	0.0500	0.0510	0.0450	0.0580	0.0570	0.0470	0.0460
	45	0.0480	0.0540	0.0450	0.0490	0.0490	0.0480	0.0450
512	6	0.0550	0.0510	0.0510	0.0510	0.0500	0.0490	0.0480
	12	0.0480	0.0460	0.0490	0.0500	0.0540	0.0460	0.0540
	45	0.0500	0.0470	0.0480	0.0450	0.0500	0.0560	0.0450
1024	6	0.0480	0.0510	0.0540	0.0500	0.0520	0.0570	0.0460
	12	0.0600	0.0500	0.0510	0.0450	0.0590	0.0520	0.0470
	45	0.0410	0.0510	0.0520	0.0480	0.0560	0.0470	0.0510

Na Tabela 4.35, consideramos 3 grupos, sendo dois deles com a mesma distribuição com séries AR(1), e o terceiro grupo com séries MA(1). Para cada parâmetro autorregressivo ϕ , tomamos $\theta = \pm \sqrt{\phi^2/(1 - \phi^2)}$ em que o sinal de θ é o mesmo de ϕ . Com parâmetros escolhidos dessa forma, teremos séries com a mesma variância em todos os grupos, exatamente como feito na subseção anterior. Nessa configuração, o que devemos observar é que o comportamento assintótico da estatística está dentro do esperado, com o aumento do poder à medida em que aumentamos n e/ou K . A média e o desvio padrão da estatística de teste, para essa situação, são apresentados na Tabela 4.36.

Na Tabela 4.37, observamos que o poder cresce muito mais rapidamente devido ao fato em que os 3 grupos considerados seguem distribuições distintas. O primeiro grupo de séries seguem processo AR(1), no segundo grupo as séries seguem processo MA(1), com o parâmetro média móvel θ definido como na Tabela 4.35, e o terceiro grupo com séries MA(2), sendo que o parâmetro média móvel de primeira ordem $\theta_1 = 1$ para todos os casos, e o parâmetro média

Tabela 4.34: Média e desvio padrão da estatística de teste, referentes ao estudo de simulação apresentado na Tabela 4.33.

K	$n \setminus \phi$	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75
64	6	-0.0189 0.3424	0.0065 0.3409	0.0219 0.3381	0.0052 0.3298	-0.0007 0.3194	0.0178 0.3519	-0.0126 0.3409
	12	0.0051 0.1330	-0.0016 0.1329	0.0040 0.1346	0.0017 0.1288	-0.0027 0.1277	-0.0028 0.1332	0.0045 0.1420
	45	0.0015 0.0313	0.0004 0.0307	0.0012 0.0308	-0.0005 0.0301	0.0003 0.0309	-0.0002 0.0312	0.0011 0.0326
256	6	-0.0052 0.1567	0.0005 0.1503	0.0042 0.1561	0.0039 0.1522	0.0037 0.1618	0.0001 0.1535	0.0037 0.1521
	12	0.0023 0.0666	0.0013 0.0582	-0.0003 0.0610	-0.0035 0.0584	-0.0023 0.0585	0.0039 0.0603	0.0012 0.0602
	45	-0.0003 0.0140	-0.0014 0.0137	0.0006 0.0143	-0.0003 0.0141	-0.0003 0.0138	-0.0003 0.0144	0.0008 0.0141
512	6	0.0020 0.1099	-0.0018 0.1086	-0.0004 0.1065	0.0018 0.1046	-0.0058 0.1055	-0.0005 0.1133	-0.0027 0.1024
	12	0.0019 0.0432	-0.0007 0.0419	-0.0010 0.0412	-0.0013 0.0406	-0.0008 0.0398	-0.0009 0.0442	0.0023 0.0425
	45	-0.0002 0.0095	0.0001 0.0102	0.0004 0.0097	-0.0001 0.0094	0.0002 0.0096	0.0001 0.0097	-0.0002 0.0092
1024	6	-0.0011 0.0766	0.0011 0.0768	0.0023 0.0755	0.0019 0.0746	-0.0069 0.0750	-0.0020 0.0769	-0.0027 0.0751
	12	-0.0001 0.0301	0.0009 0.0292	0.0021 0.0286	0.0019 0.0293	0.0014 0.0285	0.0016 0.0300	0.0007 0.0290
	45	0.0002 0.0069	-0.0002 0.0066	0.0004 0.0067	0.0001 0.0068	-0.0002 0.0069	0.0001 0.0068	-0.0002 0.0069

Tabela 4.35: Poder do teste a 5% quando consideramos 3 grupos. Dois grupos iguais, sendo as séries geradas a partir de um modelo AR(1), e um modelo MA(1). A estatística de teste é baseada na métrica d_{LNP} . Os resultados são baseados em 1000 replicações.

K	$n \setminus \phi$	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75
64	4	0.1700	0.0650	0.0510	0.0630	0.0430	0.0480	0.2140
	12	0.5900	0.0850	0.0550	0.0480	0.0520	0.0860	0.6250
	50	1.0000	0.3990	0.0560	0.0500	0.0560	0.4530	1.0000
256	4	0.8320	0.0880	0.0500	0.0590	0.0500	0.0980	0.8590
	12	1.0000	0.2360	0.0520	0.0550	0.0520	0.2830	1.0000
	50	1.0000	0.9940	0.0520	0.0540	0.0650	0.9900	1.0000
512	4	0.9940	0.1450	0.0450	0.0460	0.0580	0.1590	0.9940
	12	1.0000	0.4320	0.0480	0.0480	0.0460	0.4630	1.0000
	50	1.0000	1.0000	0.0650	0.0520	0.0680	1.0000	1.0000
1024	4	1.0000	0.2310	0.0480	0.0450	0.0460	0.2320	1.0000
	12	1.0000	0.7530	0.0590	0.0450	0.0540	0.8180	1.0000
	50	1.0000	1.0000	0.0900	0.0530	0.0930	1.0000	1.0000

móvel de segunda ordem θ_2 é definido como na Tabela 4.35. A média e o desvio padrão da estatística de teste são apresentados na Tabela 4.38.

Tabela 4.38: Média e desvio padrão da estatística de teste, referentes ao estudo de simulação apresentado na Tabela 4.37.

K	$n \setminus \phi$	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75
64	6	1.2610 0.5025	0.9918 0.4538	1.0921 0.4885	1.5135 0.6333	0.8448 0.4720	0.5026 0.4276	1.2172 0.4919
	12	1.1292 0.2669	0.8930 0.2367	0.9957 0.2594	1.3300 0.3510	0.7741 0.2474	0.4802 0.2111	1.1288 0.2674
	45	1.0513 0.1203	0.8473 0.1071	0.9318 0.1158	1.2549 0.1597	0.7284 0.1067	0.4555 0.0811	1.0487 0.1203
256	6	1.3275 0.2358	0.9726 0.2174	1.0658 0.2172	1.6656 0.3225	0.8816 0.2310	0.5927 0.2041	1.4243 0.2543
	12	1.2016 0.1342	0.8868 0.1143	0.9593 0.1149	1.5055 0.1874	0.8051 0.1152	0.5365 0.1014	1.3068 0.1380
	45	1.1279 0.0607	0.8300 0.0480	0.9061 0.0531	1.4117 0.0846	0.7542 0.0514	0.5019 0.0420	1.2228 0.0602
512	6	1.3326 0.1742	0.9746 0.1480	1.0461 0.1487	1.7237 0.2317	0.8907 0.1564	0.5977 0.1421	1.4806 0.1709
	12	1.2205 0.0941	0.8839 0.0767	0.9566 0.0826	1.5668 0.1427	0.8073 0.0799	0.5399 0.0715	1.3507 0.0951
	45	1.1482 0.0418	0.8277 0.0331	0.8989 0.0352	1.4734 0.0643	0.7602 0.0362	0.5093 0.0289	1.2661 0.0403
1024	6	1.3563 0.1184	0.9715 0.1046	1.0596 0.1044	1.7814 0.1769	0.8938 0.1074	0.6018 0.0977	1.5173 0.1256
	12	1.2334 0.0661	0.8818 0.0541	0.9519 0.0562	1.6245 0.1024	0.8153 0.0588	0.5466 0.0512	1.3729 0.0658
	45	1.1555 0.0290	0.8276 0.0244	0.8954 0.0252	1.5198 0.0468	0.7624 0.0252	0.5145 0.0203	1.2874 0.0304

4.3 Estudos de Simulação para o Método de Agrupamento

O procedimento para encontrar o melhor agrupamento de n séries temporais pode ser descrito da seguinte maneira. Primeiramente, consideramos todos os agrupamentos possíveis das séries. Em seguida, testamos todos os agrupamentos via procedimento bootstrap realizado seguindo os passos descritos em Pinheiro e Pinheiro (2007). Então, verificamos qual o menor p-valor de todos os testes. Se o p-valor for pequeno, há indicação de uma grande dissimilaridade dos grupos testados. Se o p-valor for grande, então os grupos são mais similares. Se queremos dissimilaridade, olhamos para o menor p-valor. Muitas vezes dois agrupamentos distintos apresentam mesmo p-valor ou, até mesmo, diferenças ínfimas. Nesses casos, consideramos não somente o menor e sim um grupo dos menores p-valores. Então, verificamos quais são as séries que aparecem juntas com maior frequência, determinando os grupos.

O estudo de simulação foi realizado utilizando-se a métrica com melhor desempenho nas situações testadas, d_{LNP} e duas configurações: séries temporais estacionárias contra estacionárias e estacionárias contra não estacionárias. Em ambos os procedimentos, dois grupos de 6 séries são combinadas em uma amostra composta de 12 séries simuladas. Foram considerados os tamanhos amostrais 50, 100, 200, 500, 1000 e 10000. Cada caso (distribuição \times comprimento da série) é replicado 1000 vezes.

A Tabela 4.39 apresenta os modelos utilizados no estudo de simulação. O primeiro e segundo grupos do caso estacionário contra não estacionário foram compostos por $n_1 = 6$

modelos ARMA (Tabela 4.39 (b)-(g)) e $n_2 = 6$ modelos ARIMA (Tabela 4.39 (h)-(m)), respectivamente. A simulação estacionária contra estacionária é baseado em conjuntos de $n_1 = 6$ séries AR(1) com $\phi_1 = 0.8$ (Tabela 4.39 (a)), e $n_2 = 6$ séries AR(1) com $\phi_1 = 0.3$ (Tabela 4.39 (n)). Todas as séries são geradas a partir de um ruído normal com média zero e variância unitária. A Figura 4.3 mostra a forma típica destas séries.

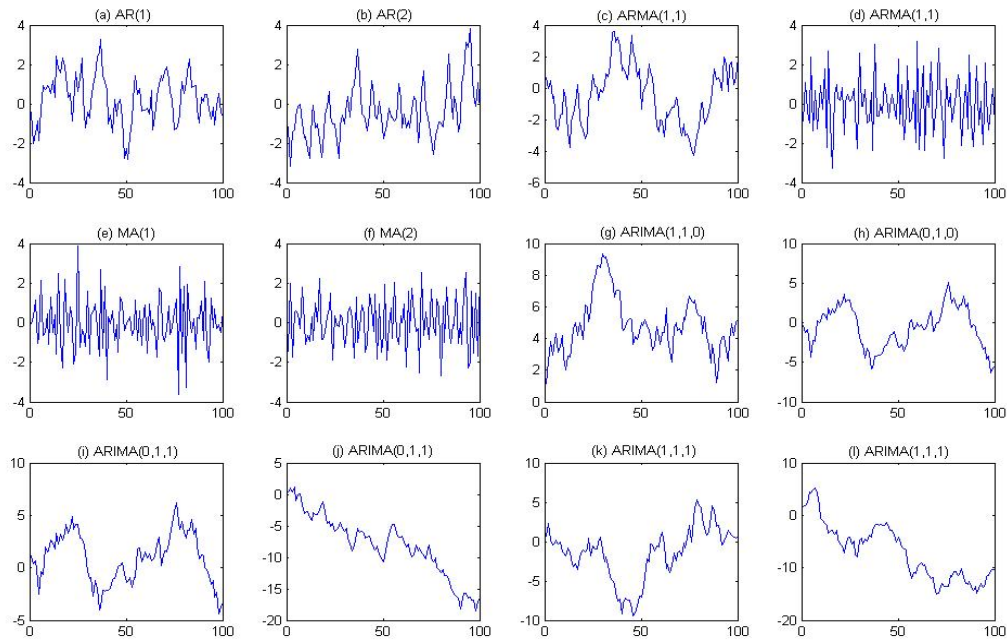


Figura 4.3: Séries estacionárias e não estacionárias simuladas.

Escolhemos todos os modelos com parâmetros altos, próximos da não estacionariedade dos modelos AR e ARMA, para dificultar a tarefa de classificar as séries em estacionárias e não estacionárias e também porque os modelos foram estudados por Caiado et al. (2006). Assim, escolhemos as distâncias que apresentaram melhor desempenho nas simulações deste trabalho. Para as funções que dependem de outros parâmetros, como é o caso do periodograma e da autocorrelação, que dependem do número de frequências, e de L , respectivamente, utilizamos os resultados de Caiado et al. (2006), em que os melhores resultados simulados foram obtidos com frequências altas ($\lfloor \sqrt{n} + 1 \rfloor$ até $\lfloor n/2 \rfloor$) e $L = K/10$.

A estatística de teste proposta, B_n , é calculada para o núcleo \mathbf{d}_{LNP} , dado pelo quadrado da distância euclidiana entre os log's dos periodogramas normalizados, com todas as frequências (de 1 a $\lfloor K/2 \rfloor$). As mesmas séries também são agrupadas utilizando “complete linkage hierarchical clustering algorithm” com as estatísticas LNP , $EUCL$, ACF , $ACFG$ e $KLFD$. As associações empíricas de cada estatística são então comparadas com as verdadeiras, e as proporções de séries corretamente associadas são apresentadas na Tabela 4.40.

Tabela 4.39: Modelos utilizados no estudo de simulação. Todas as séries são geradas a partir de um ruído com média zero e variância unitária. Os tamanhos das séries são 50, 100, 200, 500, 1000 e 10000.

Modelo	Distribuição	Parâmetros	Modelo	Distribuição	Parâmetros
(a)	AR(1)	$\phi_1 = 0.8$	(b)	AR(2)	$\phi_1 = 0.8; \phi_2 = -0.1$
(c)	ARMA(1,1)	$\phi_1 = 0.8; \theta_1 = 0.1$	(d)	ARMA(1,1)	$\phi_1 = -0.1; \theta_1 = -0.95$
(e)	ARMA(1,1)	$\phi_1 = -0.1; \theta_1 = -0.95$	(f)	MA(1)	$\theta_1 = -0.9$
(g)	MA(2)	$\theta_1 = -0.95; \theta_2 = -0.1$	(h)	ARIMA(1,1,0)	$\phi_1 = -0.1$
(i)	ARIMA(0,1,0)		(j)	ARIMA(0,1,1)	$\theta_1 = 0.1$
(k)	ARIMA(0,1,1)	$\theta_1 = -0.1;$	(l)	ARIMA(1,1,1)	$\phi_1 = 0.1; \theta_1 = -0.1$
(m)	ARIMA(1,1,1)	$\phi_1 = 0.05; \theta_1 = -0.05$	(n)	AR(1)	$\phi = 0.3$

No caso de séries temporais estacionárias contra não estacionário, \mathbf{d}_{LNP} tem melhor desempenho que *EUCL* e *ACF*, para todos os comprimentos de série. Também supera *LNP* para todas as séries, exceto para $K = 10000$, na qual ambos têm desempenho perfeito. \mathbf{d}_{LNP} ultrapassa *ACFG* para tamanhos de amostra menor ($K = 50, 100$ e 200) em uma diferença de até 5%, é superado para $K = 500$ e 1000 em até 1, 11%, e empata para $K = 10000$ com 100% de acerto. \mathbf{d}_{LNP} supera *KLFD* por quase 10%, 1% e 0, 77%, para $K = 50, 500$ e 1000 , respectivamente. É superado para $K = 100$ e 200 por 0, 70% e 1, 58%, respectivamente. Para $K = 10000$, ambas obtêm 100% de acertos.

No caso de séries estacionárias contra estacionárias, \mathbf{d}_{LNP} supera todos os outros métodos, exceto para os empates com *LNP* e *ACF*, para $K = 10000$, e com *ACFG* para $K = 1000$ e 10000 .

A métrica \mathbf{d}_{LNP} é tão eficaz quanto qualquer um dos métodos competidores, superando todos para as séries temporais de tamanhos até $K = 1000$ (à exceção de um empate com *ACFG* para $K = 1000$). Note que \mathbf{d}_{LNP} alcança agrupamento perfeito para $K = 500$ e mais de 99% para $K = 200$. Além disso, seu pior desempenho é 88, 49%, quase 10% melhor do que seu mais próximo concorrente. Essa métrica também é muito bem sucedida para discriminação de séries temporais estacionárias contra não estacionária, sendo muito mais poderosa que *LNP*, *EUCL* e *ACF*, sendo marginalmente melhor do que *ACFG* e *KLFD*. Resumindo, \mathbf{d}_{LNP} é mais bem sucedida do que qualquer um dos outros métodos para agregação de dois grupos de séries temporais, especialmente se não conhecemos se a série é estacionária ou não estacionária. Além disso, é consistentemente mais bem sucedida do que *LNP*, para ambos os casos, estacionário contra não estacionário ou estacionário contra estacionário. Essa situação é especialmente interessante, porque \mathbf{d}_{LNP} foi construída baseada no mesmo núcleo de *LNP*, e seu desempenho superior, ilustra a vantagens estatísticas da decomposabilidade em subgrupos com sua normalidade assintótica associada para a estatística de teste.

Tabela 4.40: Porcentagem de sucesso em agrupamento de séries temporais. O estimador d_{LNP} é baseado na métrica d_{LNP} , com todas as frequências (de 1 à $\lfloor K/2 \rfloor$). Para LNP e KL são utilizadas “high frequencies” ($(\sqrt{K} + 1) \cdot \dots \cdot K/2$). $L = K/10$ é utilizado para ACF e $ACFG$, e $p = 0.05$, para $ACFG$. 1000 replicações são realizadas com $K = 50, 100, 200, 500, 1000$ e 500, para $K = 10000$.

K	50		100		200		500		1000		10000	
	<i>caso₁</i>	<i>caso₂</i>	<i>caso₁</i>	<i>caso₂</i>	<i>caso₁</i>	<i>caso₂</i>	<i>caso₁</i>	<i>caso₂</i>	<i>caso₁</i>	<i>caso₂</i>	<i>caso₁</i>	<i>caso₂</i>
d_{LNP}	76.45	88.49	80.12	96.87	89.77	99.65	98.42	100	99.70	100	100	100
LNP	68.65	59.87	73.38	62.33	75.60	74.03	81.45	96.32	95.47	99.92	100	100
$EUCL$	63.25	58.82	63.42	58.55	62.88	58.35	63.11	58.33	63.33	58.33	64.10	58.30
ACF	66.03	67.38	68.40	71.23	67.62	77.60	67.89	88.48	68.96	95.92	68.50	100
$ACFG$	75.05	79.58	75.43	89.13	84.58	95.02	99.53	99.75	100	100	100	100
$KLFG$	66.54	74.13	80.82	82.47	91.35	87.85	97.42	91.52	98.93	93.42	100	95.90

caso₁: Estacionário vs Não-Estacionário. Modelos dos grupos 1 e 2 são dados na Tabela 4.39 (b)-(g) e (h)-(m), respectivamente.

caso₂: Estacionário vs Estacionário. Modelos dos grupos 1 e 2 são dados na Tabela 4.39 (a) e (n), respectivamente.

4.3.1 Simulação sob H_0

Neste caso, simulamos 12 séries $AR(1)$, com coeficiente $\phi = 0.8$, para diferentes tamanhos amostrais e testamos todos os possíveis agrupamentos. Os p-valores encontrados são altos, como o esperado, e os menores de cada replicação estão na figura seguinte, para os diferentes valores de K .

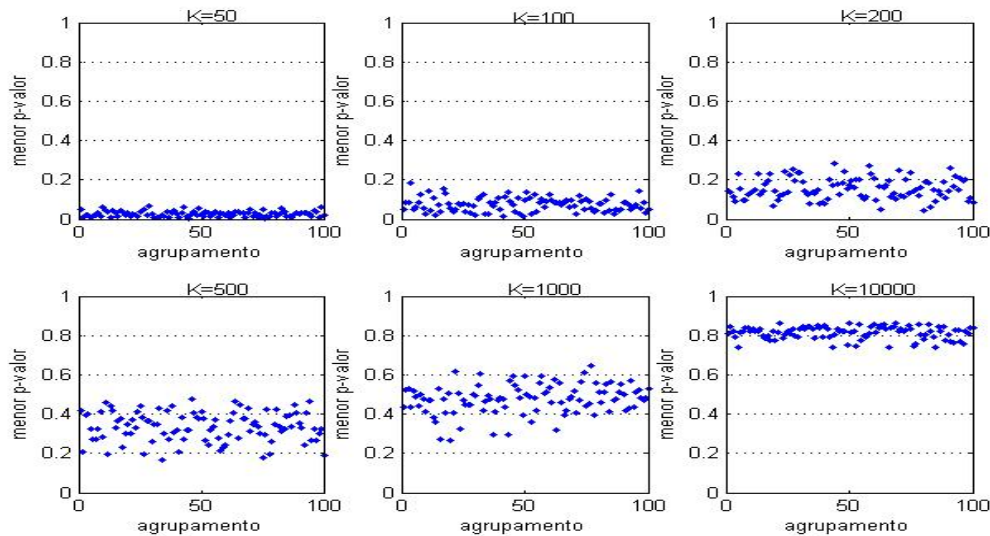


Figura 4.4: Cada ponto no gráfico representa o menor p-valor entre todos os possíveis agrupamentos de 12 séries temporais $AR(1)$, com coeficiente $\phi = 0.8$, simuladas. 100 replicações são realizadas para cada tamanho amostral K e a métrica utilizada é d_{LNP} , com todas as frequências.

Podemos observar na Figura 4.4 que estes p-valores, que na prática indicam os grupos mais

dissimilares, são altos principalmente para $K > 200$. Mesmo para $K < 200$, os valores são relativamente altos se compararmos com os que utilizamos para agrupar as séries na subseção anterior, que são diferente de zero na terceira casa decimal. Não é possível comparar as técnicas neste caso, pois as técnicas de agrupamento baseadas em distâncias sempre separam em dois grupos.

4.3.2 Correção de p-Valores

A técnica para agrupamento e classificação de séries que estamos propondo baseia-se em múltiplos testes. Quando um procedimento é baseado em múltiplos testes é necessário preocupar-se com a possibilidade de rejeitar H_0 sendo H_0 verdadeira (erro tipo I). A probabilidade de ter no mínimo um falso positivo é denotada “Family-wise error rate” (FWER). O controle do crescimento do erro tipo I ao testar, simultaneamente, uma família de hipóteses é uma questão central na área de comparações múltiplas. Vários métodos estatísticos têm sido desenvolvidos para evitar conclusões falsas positivas. O método de correção mais conhecido na literatura é o *Método de Bonferroni*, em que os p-valores são multiplicados pelo número de comparações. Correções menos conservadoras são apresentadas por Holm (1979), Hochberg (1988), Hommel (1988), Benjamini e Hochberg (1995) e Benjamini e Yekutieli (2001), respectivamente. Todos estes métodos já estão implementados no software R.

Os primeiros quatro métodos são concebidos para dar um forte controle da FWER. Hochberg e o método de Hommel são válidos quando os testes de hipóteses são independentes ou quando eles são não negativamente associados (Sarkar e Chang (1997)). O método de Hommel é mais poderoso do que o método de Hochberg, mas a diferença é geralmente pequena e os p-valores de Hochberg são mais rápidos para calcular.

Os métodos de Benjamini e Hochberg, e Benjamini e Yekutieli controlam a “false discovery rate” (FDR), que é a proporção esperada de descobertas falsas entre as hipóteses rejeitadas. A FDR é uma condição menos rigorosa do que a FWER. Assim, esses métodos são mais poderosos que os anteriores. Por esse motivo, consideramos aqui o método de Benjamini e Hochberg (BH), que pode ser sintetizado da seguinte forma:

- (i) Dispor os p-valores P_1, \dots, P_m , correspondentes às hipóteses H_1, \dots, H_m em ordem crescente.
- (ii) Para uma desejada FDR α , comparar o p-valor P_i com o valor crítico $\alpha \times i/m$.
- (iii) Seja $k = \max(i : P_i < \alpha \times i/m)$. Se k existe, então rejeite H_1, \dots, H_k .

Na Figura 4.5, apresentamos os p-valores da Figura 4.4, corrigidos pelo método Benjamini e Hochberg. Em nenhum caso a hipótese nula (de que os grupos testados são homogêneos) é rejeitada.

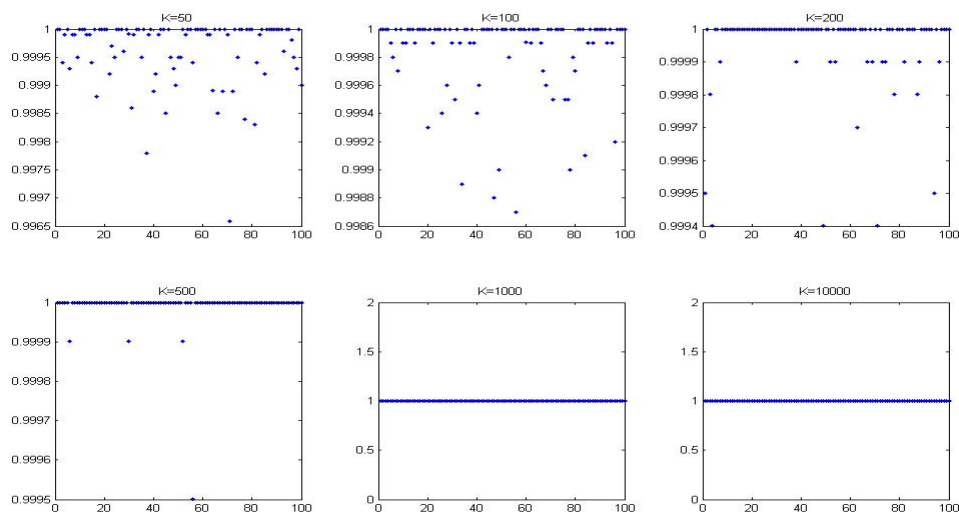


Figura 4.5: Cada ponto no gráfico representa o menor p-valor, corrigido pelo método Benjamini e Hochberg, entre todos os possíveis agrupamentos das 12 séries temporais $AR(1)$, com coeficiente $\phi = 0.8$, da Figura 4.4.

4.4 Aplicação

Para ilustrar o método de agrupamento e classificação que estamos propondo, trabalhamos com dados de terremotos e explosões. São consideradas 8 séries provenientes de terremotos e 8 séries provenientes de explosões, além de uma série com fonte geradora desconhecida.

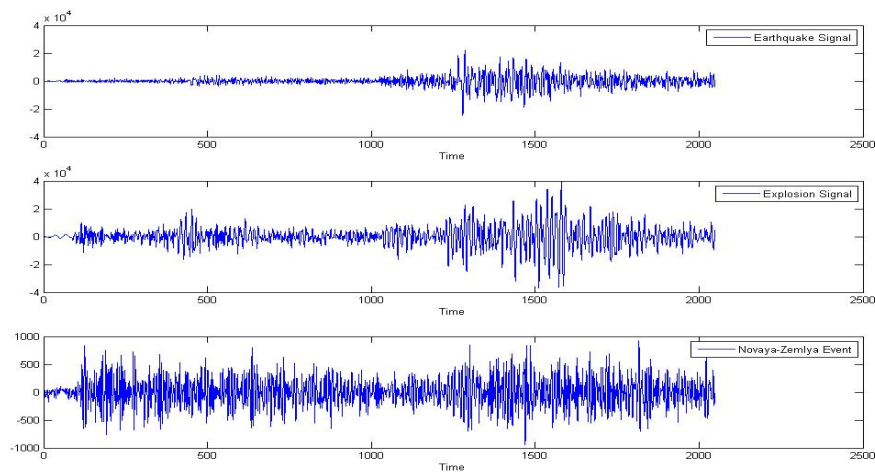


Figura 4.6: Sinais sísmológicos. Nos dois primeiros painéis temos os gráficos de sinais provenientes de um terremoto e uma explosão, respectivamente. O último painel mostra um evento ocorrido em Novaya Zemlya (a origem é desconhecida). Fonte dos dados: O arquivo eq+exp.dat de Shumway e Stoffer (2006) disponível em <http://lib.stat.cmu.edu/general/tsa2/>.

A Figura 4.6 mostra um sinal proveniente de um terremoto, um sinal proveniente de uma explosão de mineração na Escandinávia, medidos por uma estação local, e um evento em Novaya Zemlya, região da Rússia, de origem desconhecida. Este conjunto de dados reais é objeto de estudo de vários pesquisadores. Kakizawa et al. (1998) estudaram se estas séries sismológicas são provenientes de terremotos ou testes nucleares. Esse mesmo conjunto de dados reais foi considerado por Jeng e Huang (2008) e Fryzlewicz e Ombao (2009) para ilustrar suas abordagens. Apesar das diferentes abordagens terem bons desempenhos, nenhum deles pode testar se o agrupamento ou a classificação são estatisticamente confiáveis.

O fato de estarmos aptos a testar se dois grupos de séries temporais independentes têm a mesma distribuição, realmente faz diferença para classificação. Conhecendo-se dois grupos de séries temporais, somos capazes de decidir a qual grupo se assemelha mais uma série temporal qualquer.

O terremoto e a explosão descritos graficamente na Figura 4.6 representam, tipicamente, as séries do conjunto tomado como base de dados para este estudo, compilado por Kakizawa et al. (1998). Todos os eventos escolhidos foram em terra, ou bem próximos, e foram distribuídos uniformemente sobre a Escandinávia, de modo a minimizar a possibilidade de que discriminadores poderiam ser mais concentrados na localização ou nas diferenças entre mar e terra.

Kakizawa et al. (1998) adicionaram um evento de origem incerta observado em Novaya Zemlya. Todos os eventos, exceto o evento russo, ocorreram na Península Escandinava e foram gravados por matrizes sísmicas localizadas na Noruega e na Finlândia. As observações na Noruega foram feitas pelas estações sísmicas experimentais Norwegian and Arctic e, na Finlândia, pela estação sísmica experimental. Cada registro mostrado na Figura 4.6 é, na verdade, composto de duas fases: a fase P de primeiras chegadas e a fase S de chegadas tardias. A Figura 4.7 mostra os traços característicos das diferentes fases dos terremotos e explosões.

As chegadas são sinais realmente diferentes, tomando diferentes caminhos através da crosta terrestre e são separados por um intervalo de tempo razoavelmente longo. Para os sismólogos, separar as fases P e S em dois sinais distintos é muito importante. Em geral, os pontos de partida da fase inicial P e da fase S podem ser identificados diretamente a partir dos registros, como na Figura 4.6. Kakizawa et al. (1998) consideraram também ambas as fases conjuntamente, como um vetor bivariado com 1024 pontos em cada vetor.

Queremos testar se o grupo formado por uma série de terremotos e do grupo formado por uma série de explosões são sinais provenientes da mesma fonte geradora. Obviamente, conhecemos preliminarmente que não são, por isso temos de rejeitar a hipótese nula, de que grupos têm uma mesma fonte geradora. Utilizando as fases S, testamos a hipótese nula utilizando um procedimento bootstrap duplo, o qual foi realizado seguindo os passos de McCullough e Vinod (1998), e obtemos o p-valor 0,03, o que nos leva a rejeitar a hipótese nula, como deveríamos.

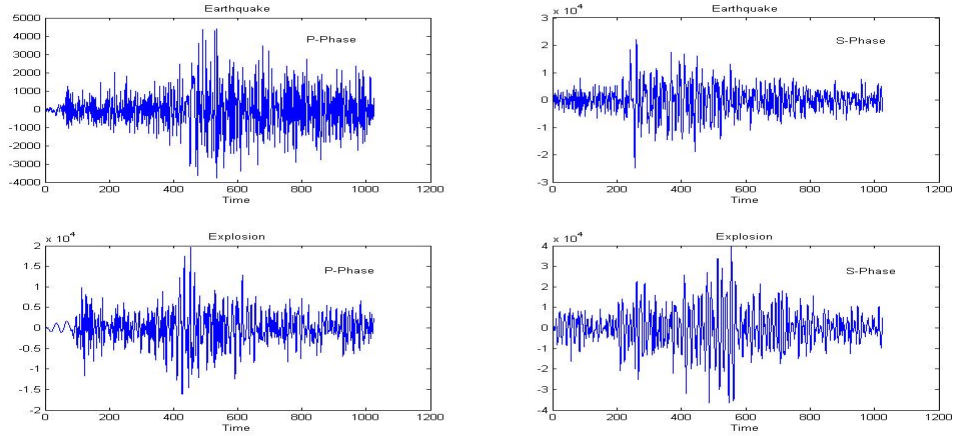


Figura 4.7: Fase P e fase S de terremotos e explosões.

4.4.1 Procedimento “Leave-One-Out” Utilizando a Métrica d_{LNP}

Realizamos um procedimento “leave-one-out” para verificar a confiabilidade do nosso teste. A Tabela 4.41 mostra o valor da estatística $B_{n_{8 \times 8}}$, que é a estatística para testar a hipótese nula de que a fase S das 8 séries de terremotos e a fase S das 8 séries de explosões têm a mesma fonte geradora. O valor da estatística $B_{n_{8 \times 8}}$ é 0.4664, com p-valor 0.03, rejeitando a hipótese nula. Para o procedimento “leave-one-out”, primeiramente retiramos, uma a uma, as séries de terremotos, indicando a série removida por L_{EQ} , colocando-a no grupo de explosões e, então, calcula-se a estatística B_n , a qual é denotada por $B_{n_{7 \times 9}}$, e compara-se com a estatística $B_{n_{8 \times 8}}$. O mesmo procedimento é realizado através da remoção de um série de explosões e colocando-a no grupo de terremotos. A série removida é indicada por L_{EX} e a estatística B_n , para este grupo, é denotada por $B_{n_{9 \times 7}}$. Para casos em que $B_{n_{7 \times 9}}$ ou $B_{n_{9 \times 7}}$ são maiores que $B_{n_{8 \times 8}}$, indicando que o agrupamento 7×9 ou 9×7 é mais divergente que 8×8 , testa-se a hipótese nula. Quando removemos do grupo de terremotos as séries 1 e 4, $B_{n_{7 \times 9}}$ é maior que $B_{n_{8 \times 8}}$, mas somente um p-valor é menor que no caso 8×8 , indicando que somente em um caso teríamos um agrupamento melhor que 8×8 . Quando retiramos das explosões, para as séries 9, 11 e 16, $B_{n_{9 \times 7}}$ é maior que $B_{n_{8 \times 8}}$, mas o p-valor da série 11 é maior que o p-valor obtido testando-se 8×8 . Em resumo, temos três situações indicando um agrupamento melhor que o agrupamento correto 8×8 .

4.4.2 Classificação

O evento desconhecido, NZ, ocorreu em Novaya Zemlyanas nas proximidades de uma região utilizada pela Rússia para realização de testes nucleares. O objetivo aqui é encontrar um bom classificador que possa distinguir entre terremotos e explosões, e determinar a classe do evento desconhecido NZ. A Tabela 4.42 mostra os resultados dos testes. Quando colocamos

Tabela 4.41: “Leave-one-out” da fase S dos terremotos (L_{EQ}) e explosões (L_{EX}) utilizando a métrica d_{LNP} .

$B_{n_{8 \times 8}}$	p-valor	L_{EQ}	$B_{n_{7 \times 9}}$	p-valor	L_{EX}	$B_{n_{9 \times 7}}$	p-valor		
0.4664	0.0300	1	0.9289*	0.0100	9	0.6548*	0.0210		
		2	0.3787			10		0.1308	
		3	0.1334			11		0.5470*	0.0420
		4	0.5081*			12		0.0271	
		5	0.1506			13		0.0528	
		6	0.2018			14		0.0955	
		7	0.3097			15		0.2832	
		8	0.2456			16		1.0656*	

a fase S da série NZ no grupo de explosões, EX e testamos se o grupo EX+NZ e o grupo de terremotos, EQ, têm a mesma fonte geradora, temos um p-valor pequeno, indicando uma grande divergência entre os grupos. No entanto, para a fase S de NZ no grupo EQ, obtemos um p-valor grande, indicando que para essa situação, os grupos não são tão divergentes. Isso nos leva a concluir que NZ é uma explosão.

Tabela 4.42: Classificação do evento de origem desconhecida em Novaya Zemlya utilizando a fase S.

	EQ+NZ vs. EX	EX+NZ vs. EQ
p-valor	0.510	0.017

4.4.3 Procedimento “Leave-One-Out” Utilizando a Métrica d_{ACF}

Também consideramos um procedimento “leave-one-out” utilizando a métrica d_{ACF} . Aqui, o procedimento bootstrap duplo não é necessário porque em nada altera o resultado obtido pelo bootstrap simples.

Tabela 4.43: “Leave-one-out” da fase S dos terremotos (L_{EQ}) e explosões (L_{EX}) utilizando a medida a métrica d_{ACF} .

$B_{n_{8 \times 8}}$	p-valor	L_{EQ}	$B_{n_{7 \times 9}}$	p-valor	L_{EX}	$B_{n_{9 \times 7}}$	p-valor	
0.0263	0.0180	1	0.0376*	0.0100	9	0.0266*	0.0270	
		2	0.0204			10		0.0129
		3	0.0224			11		0.0244
		4	0.0158			12		0.0122
		5	0.0115			13		0.0109
		6	0.0132			14		0.0082
		7	0.0256			15		0.0165
		8	0.0144			16		0.0491*

Como podemos ver na Tabela 4.43, a primeira série de EQ, quando colocada no grupo EX, indica uma maior divergência entre os grupos, quando comparados com o agrupamento padrão EX contra EQ. O mesmo acontece quando colocamos a série 8 do grupo EX no grupo EQ. Note que, utilizando diferentes medidas de divergência, podemos concluir que a primeira série de EQ e a oitava série de EX parecem estar trocadas. Por isso, fizemos os gráficos das fases S de todas as séries. Nas Figuras 4.8 e 4.9 podemos ver que a primeira série de EQ tem grande semelhança com as séries do grupo EX, enquanto a oitava série do grupo EX possui grande similaridade com o grupo EQ. Essa pode ser uma explicação por estas séries parecerem trocadas.

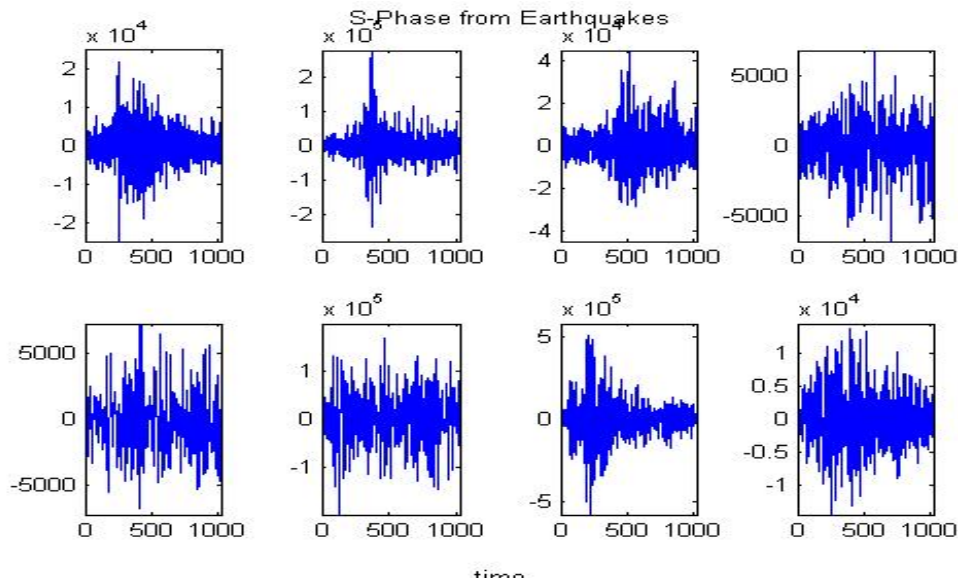


Figura 4.8: Fase S dos terremotos.

4.4.4 Procedimento “Leave-One-Out” Combinando as Fases S e P

Todos os estudos anteriores são baseados na fase S das séries. A partir das fases P não é possível diferenciar terremotos de explosões. Realizamos um procedimento “leave-one-out” considerando as fases S e P conjuntamente, em uma proporção $\alpha S_p + (1 - \alpha) S_s$, em que S_p e S_s denotam, respectivamente, as fases P e S de todas as séries. Calculamos todas as estatísticas $B_{n_7 \times 9}$ e $B_{n_9 \times 7}$, mas somente consideramos os casos em que elas foram maiores que $B_{n_8 \times 8}$, exatamente como feito quando consideramos somente a fase S. As séries para as quais $B_{n_7 \times 9}$ e $B_{n_9 \times 7}$ são maiores que $B_{n_8 \times 8}$, quando consideramos as fases P e S conjuntamente, coincidem com aquelas onde só a fase S é considerada. Observe nas Figuras 4.10 e 4.11 que quando α cresce, isto é, quando mais peso é dado à fase P, os p-valores dos testes crescem, indicando menor divergência dos grupos. Quando α é maior que 0.7, a influência da fase P não permite mais que a hipótese de homogeneidade dos grupos seja rejeitada. Isso mostra que o estudo deve ser realizado com a fase S.

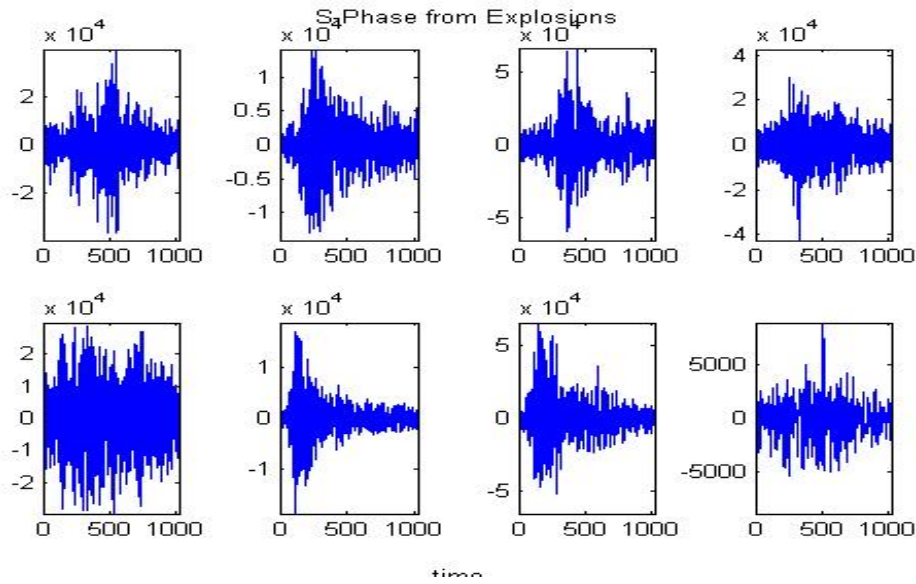


Figura 4.9: Fase S das explosões.

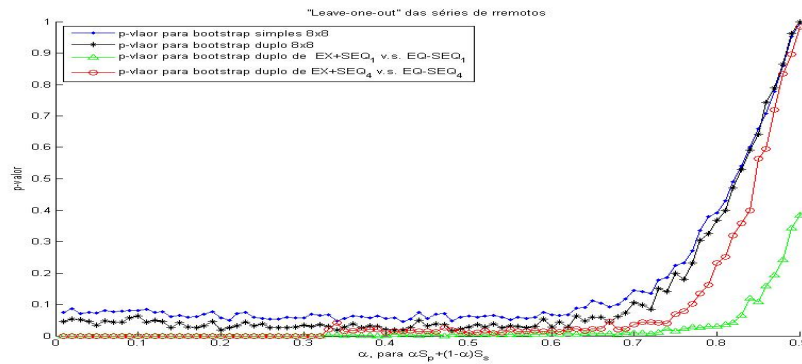


Figura 4.10: Procedimento “Leave-one-out” com séries sendo retiradas do grupo de terremotos (EQ) e colocando-as no grupo de explosões (EX), considerando-se as fases P e S de cada série em uma proporção $\alpha S_p + (1 - \alpha) S_s$, em que S_p e S_s denotam, respectivamente, as fases P e S das 16 séries. Para cada α , pontos indicam p-valores obtidos, através de um bootstrap simples, testando homogeneidade dos grupos EQ e EX, em que EQ e EX denotam, respectivamente, as fases S e P combinadas de terremotos e explosões. Asteriscos indicam p-valores obtidos, através de um procedimento bootstrap duplo, testando a homogeneidade de EQ e EX. Círculos denotam p-valores obtidos, através de um procedimento bootstrap duplo, testando a homogeneidade de $EQ - SEQ_1$ e $EX + SEQ_1$, em que $EQ - SEQ_1$ significa que a série 1 foi retirada do grupo EQ e $EX + SEQ_1$ significa que a série 1 de EQ foi adicionada ao grupo EX. Triângulos denotam p-valores obtidos através de um procedimento bootstrap duplo, testando a homogeneidade de $EQ - SEQ_8$ e $EX + SEQ_8$.

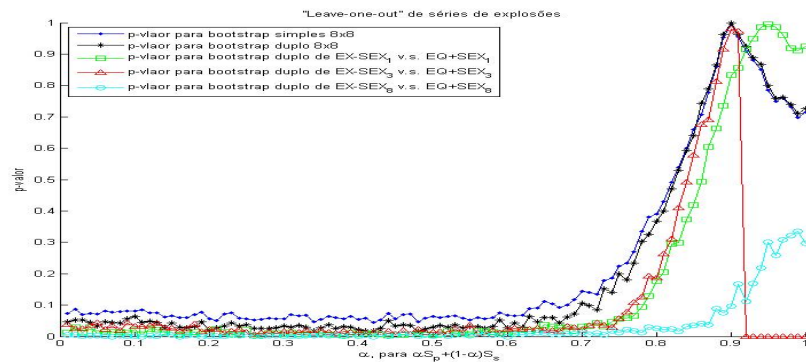


Figura 4.11: Procedimento “Leave-one-out” com séries sendo retiradas do grupo de explosões (EX) e colocando-as no grupo de terremotos (EQ), considerando-se as fases P e S de cada série em uma proporção $\alpha S_p + (1-\alpha)S_s$, em que S_p e S_s denotam, respectivamente, as fases P e S das 16 séries. Para cada α , pontos indicam p-valores obtidos através de um bootstrap simples, testando homogeneidade dos grupos EQ e EX, em que EQ e EX denotam, respectivamente, as fases S e P combinadas de terremotos e explosões. Asteriscos indicam p-valores obtidos através de um procedimento bootstrap duplo, testando a homogeneidade de EQ e EX. Asteriscos indicam p-valores obtidos através de um procedimento bootstrap duplo, testando a homogeneidade de EQ e EX. Quadrados indicam p-valores obtidos através de um procedimento bootstrap duplo, testando a homogeneidade de $EX - SEX_1$ e $EQ + SEX_1$, em que $EX - SEX_1$ significa que a série 1 foi retirada do grupo EX e $EQ + SEX_1$ significa que a série 1 de EX foi adicionada ao grupo EQ. Triângulos denotam p-valores obtidos, através de um procedimento bootstrap duplo, testando a homogeneidade de $EX - SEX_3$ e $EQ + SEX_3$. Círculos denotam p-valores obtidos através de um procedimento bootstrap duplo, testando a homogeneidade de $EX - SEX_8$ e $EQ + SEX_8$.

Capítulo 5

Caso Multivariado

A globalização econômica e de comunicação na Internet tem acelerado a integração dos mercados financeiros mundiais nos últimos anos. Os movimentos de preço em um mercado podem se espalhar fácil e imediatamente a outro mercado. Por essa razão, os mercados financeiros estão mais dependentes uns dos outros do que nunca e é preciso considerá-los conjuntamente para compreender melhor a estrutura dinâmica das finanças globais. Saber como os mercados estão inter relacionadas é de grande importância em finanças. Da mesma forma, para um investidor ou uma instituição financeira detentora de ativos múltiplos, as relações dinâmicas entre os retornos dos ativos desempenham um papel importante na tomada de decisão. Um método para discriminar e agrupar séries temporais multivariadas não gaussianas, baseado na distância de Kullback-Leibler, é proposto por Kakizawa et al. (1998).

Na Seção 5.1, apresentamos alguns dos modelos multivariados mais conhecidos na literatura de séries temporais. Na Seção 5.2, generalizamos os resultados do Capítulo 3 e na Seção 5.3, realizamos estudos de simulação para estudar o comportamento da estatística de teste B_n com séries temporais multivariadas.

5.1 Séries Temporais Multivariadas

Neste capítulo e, particularmente, nesta seção, somente o caso bivariado será considerado, embora as definições seguintes sejam válidas para qualquer dimensão e a notação *negrito* é utilizada para denotar séries temporais bivariadas (multivariadas).

Considere uma sequência de séries temporais bivariadas $\mathbf{X}_1, \dots, \mathbf{X}_n$ independentes, isto é, $\mathbf{X}_i = (X_{i1}, X_{i2})$, para $i = 1, \dots, n$, em que

$$\mathbf{X}_{i1} = \begin{pmatrix} X_{i11} \\ \vdots \\ X_{i1K} \end{pmatrix} \text{ e } \mathbf{X}_{i2} = \begin{pmatrix} X_{i21} \\ \vdots \\ X_{i2K} \end{pmatrix}. \quad (5.1)$$

Cada \mathbf{X}_i , para $i = 1, \dots, n$, é uma série temporal bivariada. Para uma série temporal bivariada

$\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$, o vetor de médias é definido por

$$\boldsymbol{\mu}_t = \mathbb{E}\mathbf{X}_t = (\mathbb{E}X_{1t}, \mathbb{E}X_{2t})$$

e a matriz de covariâncias é definida por

$$\boldsymbol{\Gamma}(t+h, t) = \text{Cov}(\mathbf{X}_{t+h}, \mathbf{X}_t) = \begin{pmatrix} \gamma_{11}(t+h, t) & \gamma_{12}(t+h, t) \\ \gamma_{21}(t+h, t) & \gamma_{22}(t+h, t) \end{pmatrix},$$

em que $\gamma_{ij}(t+h, t) = \text{Cov}(X_{i,t+h}, X_{j,t})$. Em notação matricial,

$$\boldsymbol{\Gamma}(t+h, t) = \mathbb{E}(\mathbf{X}_{t+h} - \boldsymbol{\mu}_{t+h})(\mathbf{X}_t - \boldsymbol{\mu}_t)'$$

A série temporal bivariada (ou multivariada), $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$, é dita ser *estacionária* se os momentos $\boldsymbol{\mu}_t$ e $\boldsymbol{\Gamma}(t+h, t)$ são ambos independentes de t . Neste caso, podemos utilizar a notação

$$\boldsymbol{\mu} = \mathbb{E}X_t$$

e

$$\boldsymbol{\Gamma}(h) = \text{Cov}(\mathbf{X}_{t+h}, \mathbf{X}_t).$$

A matriz de variâncias e covariâncias de \mathbf{X}_t é dada por $\boldsymbol{\Gamma}(0)$, isto é, $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}(0)$.

Nenhuma condição é imposta sobre a matriz $\boldsymbol{\Sigma}$. Logo, essa matriz pode ser não diagonal, o que significa que estamos permitindo alguma estrutura de dependência entre as séries temporais. Esta generalização é um avanço pois a suposição de independência era necessária até o momento.

5.1.1 Modelo VARMA

Em um contexto univariado, uma importante classe de processos são os *autorregressivos de médias móveis* (ARMA), os quais foram introduzidos por Box e Jenkins, definidos em termos de equações de diferenças lineares com coeficientes constantes. A classe dos *processos autorregressivos de médias móveis multivariados* de ordem p e q , denotados aqui por VARMA(p, q), é amplamente discutida em Brockwell e Davis (1991), Lütkepohl (1991) e Reinsel (1993). Para definir essa classe de processos será necessário definir os processos chamados ruído branco multivariados.

Um processo estocástico m -variado, $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$, é dito ser um *ruído branco multivariado* com média zero e matriz de variâncias e covariâncias $\boldsymbol{\Sigma}$, denotado por

$$\boldsymbol{\varepsilon}_t \sim RB(\mathbf{0}, \boldsymbol{\Sigma}), \tag{5.2}$$

se, e somente se, $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ é um processo estacionário multivariado com média $\mathbf{0}$ e matriz de variâncias e covariâncias dada por

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m^2 \end{pmatrix}, \quad (5.3)$$

com componentes $\sigma_j^2 = \text{Var}\varepsilon_{j,t} = \gamma_{jj}$ e $\gamma_{jk} = \text{Cov}(\varepsilon_{j,t}, \varepsilon_{k,t}) = 0$, para todo $j \neq k \in \{1, \dots, m\}$.

Um exemplo muito utilizado é o *ruído branco multivariado gaussiano*, em que $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Também usaremos a notação

$$\boldsymbol{\varepsilon}_t \sim IID(\mathbf{0}, \boldsymbol{\Sigma}), \text{ para todo } t \in \mathbb{Z}, \quad (5.4)$$

para indicar que as componentes de $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ são i.i.d. com média $\mathbf{0}$ e matriz de variâncias e covariâncias $\boldsymbol{\Sigma}$, dada na equação (5.3).

Diz-se que um processo estocástico m -variado, $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$, é um *processo autorregressivo de médias móveis multivariado*, aqui denotado por VARMA(p, q), se ele for uma solução estacionária da equação de diferenças

$$\mathbf{X}_t - \phi_1 \mathbf{X}_{t-1} - \cdots - \phi_p \mathbf{X}_{t-p} = \boldsymbol{\varepsilon}_t + \boldsymbol{\theta}_1 \boldsymbol{\varepsilon}_{t-1} + \cdots + \boldsymbol{\theta}_q \boldsymbol{\varepsilon}_{t-q}, \quad (5.5)$$

em que $\phi_1, \dots, \phi_p, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_q$ são matrizes reais $m \times m$ e $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ é o processo ruído branco multivariado. A equação (5.5) pode ser escrita na forma compacta

$$\boldsymbol{\Phi}(B)\mathbf{X}_t = \boldsymbol{\Theta}(B)\boldsymbol{\varepsilon}_t, \quad (5.6)$$

em que B denota o operador de defasagem usual,

$$\boldsymbol{\Phi}(B) = \phi_0 - \sum_{\ell=1}^p \phi_\ell B^\ell \text{ e } \boldsymbol{\Theta}(B) = \boldsymbol{\theta}_0 + \sum_{\ell=1}^q \boldsymbol{\theta}_\ell B^\ell, \quad (5.7)$$

com $\phi_1, \dots, \phi_p, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_q$ matrizes reais m e $\phi_0 = \mathbf{I}_m = \boldsymbol{\theta}_0$, em que \mathbf{I}_m é a matriz identidade de tamanho $m \times m$.

Exemplo 5.1.1. Quando a ordem do polinômio $\boldsymbol{\Theta}(B)$, na equação (5.7), for zero, ou seja $q = 0$, obtemos o *processo autorregressivo m -variado* de ordem p , denotado por VAR(p). Da mesma forma, se $p = 0$ na equação (5.7), obtemos o *processo de média móvel multivariado* de ordem q , denotado por VMA(q).

No contexto univariado, um enfoque maior é dado aos processos estocásticos ARMA(p, q)

causais e invertíveis. Os critérios de causalidade e invertibilidade para os processos VARMA(p, q) podem ser encontrados em Brockwell e Davis (1991), por exemplo.

Os processos VARMA(p, q) não são muito utilizados devido às dificuldades computacionais e aos problemas de identificabilidade. A matriz de variâncias e covariâncias (ou a matriz das funções de densidades espectrais) não determina unicamente Φ , Θ e Σ , a menos que condições mais restritivas sejam impostas. A não identificabilidade implica que a superfície de verossimilhança não tem um único máximo.

5.1.2 Modelo VARFIMA(p, \mathbf{d}, q)

Granger e Joyeux (1980) e Hosking (1981) apresentam a classe de processos chamada *processos autorregressivos de médias móveis com integração fracionária*, (ARFIMA(p, d, q)). A generalização do processo ARFIMA(p, \mathbf{d}, q), para o caso multivariado são os chamados processos VARFIMA(p, \mathbf{d}, q). Um processo m -dimensional, $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$, é um *processo autorregressivo fracionariamente integrado de média móvel multivariado* ou VARFIMA(p, \mathbf{d}, q), se ele é uma solução estacionária da equação de diferenças

$$\Phi(B)(1 - B)^{\mathbf{d}}(\mathbf{X}_t - \boldsymbol{\mu}) = \Theta(B)\boldsymbol{\varepsilon}_t, \quad (5.8)$$

em que B é o operador de defasagem, $\boldsymbol{\mu}$ é o vetor das médias, $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ é o processo ruído branco multivariado e $\Phi(B)$ e $\Theta(B)$ são os polinômios em B , dados na equação (5.7).

Assumimos que os coeficientes polinomiais $\Phi(B)$ e $\Theta(B)$ satisfazem as condições de causalidade e invertibilidade. O operador $\nabla^{\mathbf{d}}(B) = (1 - B)^{\mathbf{d}}$ é uma matriz diagonal $m \times m$ caracterizada pelos parâmetros d_1, \dots, d_m , dada por

$$\nabla^{\mathbf{d}}(B) = \begin{pmatrix} (1 - B)^{d_1} & 0 & \cdots & 0 \\ 0 & (1 - B)^{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (1 - B)^{d_m} \end{pmatrix}, \quad (5.9)$$

em que $d_j \in (-0.5, 0.5)$, para todo $j \in \{1, \dots, m\}$. O termo $(1 - B)^{d_j}$ é definido pela expansão

$$(1 - B)^{-d_j} = \sum_{\ell=0}^{\infty} \psi_{j,\ell} B^\ell \equiv \psi_{j,\ell}(B), \quad (5.10)$$

em que

$$\psi_{j,\ell} = \frac{\Gamma(d_j + \ell)}{\Gamma(\ell + 1)\Gamma(d_j)}, \quad \psi_0^0 = 1 \text{ e } \psi_\ell^0 = 0, \quad (5.11)$$

para todo $j \in \{1, \dots, m\}$ e $\ell \neq 1$. Consequentemente, temos que

$$\begin{aligned} \nabla^{\mathbf{d}}(B)^{-1} &\equiv \begin{pmatrix} (1-B)^{-d_1} & 0 & \cdots & 0 \\ 0 & (1-B)^{-d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (1-B)^{-d_m} \end{pmatrix} \\ &\equiv \Psi^{\mathbf{d}}(B) = \mathbf{I}_m + \sum_{\ell=1}^{\infty} \Psi_{\ell}^{\mathbf{d}} B^{\ell}, \end{aligned} \quad (5.12)$$

em que cada $\Psi_{\ell}^{\mathbf{d}}$, para $\ell \in \mathbb{N}$, é uma matriz diagonal $m \times m$ com $\Psi_{\ell}^{d_j}$ sendo o ℓ -ésimo elemento da diagonal.

Um caso particular do processo estacionário VARFIMA(p, \mathbf{d}, q) ocorre quando os graus p e q dos polinômios $\Phi(B)$ e $\Theta(B)$ são iguais a zero, ou seja, quando $\Phi(B) \equiv \mathbf{I} \equiv \Theta(B)$. Formalmente, diz-se que este processo estocástico m -dimensional, $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$, é um *processo fracionariamente integrado multivariado*, aqui denotado por VARFIMA($0, \mathbf{d}, 0$), com vetor de parâmetros de integração $\mathbf{d} = (d_1, \dots, d_m)$, com $d_j \in (-0.5, 0.5)$, para todo $j \in \{1, \dots, m\}$, se ele é uma solução estacionária da equação de diferenças

$$\nabla^{\mathbf{d}}(B)\mathbf{X}_t = \varepsilon_t, \quad (5.13)$$

em que B é o operador de defasagem, $\nabla^{\mathbf{d}}(B)$ é dado pela equação (5.9) e $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é o processo ruído branco multivariado.

Exemplo 5.1.2. Seja $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ um processo VARFIMA($0, \mathbf{d}, 0$) bivariado, ou seja, com dimensão $m = 2$. Então, $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ satisfaz a equação

$$\begin{pmatrix} (1-B)^{d_1} & 0 \\ 0 & (1-B)^{d_2} \end{pmatrix} \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}, \quad (5.14)$$

para todo $t \in \mathbb{Z}$, em que $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t})'$ é um processo ruído branco bivariado com média $\mathbf{0}$ e matriz de variâncias e covariâncias Σ . Nesse caso, decorre da teoria para processos univariados, que para cada $d_j > -1$ real, $j = 1, 2$,

$$\nabla^{d_j}(B) = (1-B)^{d_j} = \sum_{\ell=0}^{\infty} \pi_{j,\ell} B^{\ell}, \quad (5.15)$$

em que

$$\pi_{j,\ell} = \frac{\Gamma(\ell - d_j)}{\Gamma(\ell + 1)\Gamma(-d_j)} = \prod_{0 < \kappa \leq \ell} \frac{\kappa - 1 - d_j}{\kappa}, \quad (5.16)$$

para $\ell \in \mathbb{N} - \{0\}$, $\pi_0 = 1$ e

$$\Gamma(x) = \begin{cases} \int_0^\infty t^{x-1} e^{-t} dt, & \text{se } x > 0; \\ \infty, & \text{se } x = 0; \\ x^{-1} \Gamma(1+x), & \text{se } x < 0. \end{cases} \quad (5.17)$$

5.1.3 Modelo VGARCH

É natural considerar modelagem de processos com volatilidade em situações de maior dimensionalidade. No caso de um GARCH univariado, a equação básica é regida por

$$X_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim RB(0, 1),$$

em que $\sigma_t^2 = \mathbb{E}(X_t^2 | \mathcal{F}_{t-1})$ denota a variância condicional, que satisfaz a equação

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i X_{t-1}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2$$

e \mathcal{F}_{t-1} é a σ -álgebra gerada pelos elementos da série até o tempo $t - 1$.

Em geral, podemos estar interessados em um portfólio que consiste de um vetor de retornos, no qual a matriz de covariância evolui com o tempo. Suponha que, após filtrar esta série multivariada através de um modelo ARMA, chegamos a um *portfólio* que consiste de K retornos de ativos de inovação $X_{i,t}$, $i = 1, \dots, m$. Colocando-se essas inovações em um vetor \mathbf{X}_t , define-se $\sigma_{ii,t} = \text{Var}[X_{i,t} | \mathcal{F}_{t-1}]$ e $\sigma_{ij,t} = \text{Cov}(X_{i,t}, X_{j,t} | \mathcal{F}_{t-1})$. Nesse caso, $\Sigma_t = [\sigma_{ij,t}]$ denota a matriz de variância e covariância condicional de todos os retornos. A simples generalização de um modelo GARCH(1,1) relaciona a matriz de variância e covariância condicional Σ_t a $X_t X_t'$ como segue:

$$X_t = \Sigma_t^{1/2} Z_t$$

e

$$\mathbb{E}[X_t X_t' | \mathcal{F}_{t-1}] = \Sigma_t.$$

Uma das dificuldades encontradas nessa equação consiste em encontrar um sistema adequado que descreva a dinâmica de Σ_t , levando em consideração o fato de que Σ_t deve ser positiva definida.

Para descrever o modelo GARCH multivariado de forma geral, considere o operador

$vech(A)$, em que A é uma matriz quadrada, definido por

$$vech \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{22} \\ a_{32} \\ a_{33} \end{pmatrix}.$$

Em geral, se A é uma matriz $m \times m$, $vech(A)$ é um vetor de dimensão $m(m+1)/2$. Esse operador é aplicado em matrizes simétricas para separar os elementos distintos. Com essa notação, pode-se modelar Σ_t como segue:

$$vech(\Sigma_t) = \omega + \Psi vech(\Sigma_{t-1}) + \Lambda vech(X_{t-1}X_{t-1}). \quad (5.18)$$

Note que \mathbf{X}_t e \mathbf{Z}_t são vetores 3×1 , ω é 6×1 e Ψ e Λ são matrizes quadradas 6×6 .

Exemplo 5.1.3. Para ter uma ideia de como as dinâmicas da volatilidade são descritas em (5.18), considere o caso especial em que $\omega = 0$ e $\Psi = \mathbf{0}$:

$$vech(\Sigma_t) = \begin{pmatrix} \sigma_{11,t} \\ \sigma_{21,t} \\ \sigma_{22,t} \\ \sigma_{31,t} \\ \sigma_{32,t} \\ \sigma_{33,t} \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{15} & \lambda_{16} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{25} & \lambda_{26} \\ \lambda_{31} & \lambda_{32} & \cdots & \lambda_{35} & \lambda_{36} \\ \lambda_{41} & \lambda_{42} & \cdots & \lambda_{45} & \lambda_{46} \\ \lambda_{51} & \lambda_{52} & \cdots & \lambda_{55} & \lambda_{56} \\ \lambda_{61} & \lambda_{62} & \cdots & \lambda_{65} & \lambda_{66} \end{pmatrix} \begin{pmatrix} X_{1,t-1}X_{1,t-1} \\ X_{2,t-1}X_{1,t-1} \\ X_{2,t-1}X_{2,t-1} \\ X_{3,t-1}X_{1,t-1} \\ X_{3,t-1}X_{2,t-1} \\ X_{3,t-1}X_{3,t-1} \end{pmatrix}.$$

Quando equacionamos os elementos nos dois lados da equação anterior, vemos que cada volatilidade está relacionada com o passado do quadrado dos retornos de uma forma bastante complicada. Por exemplo, escrevendo a primeira equação, temos a seguinte expressão

$$\begin{aligned} \sigma_{11,t} = & \lambda_{11}X_{1,t-1}^2 + \lambda_{12}X_{2,t-1}X_{1,t-1} + \lambda_{13}X_{2,t-1}^2 \\ & + \lambda_{14}X_{3,t-1}X_{1,t-1} + \lambda_{15}X_{3,t-1}X_{2,t-1} + \lambda_{16}X_{3,t-1}^2. \end{aligned}$$

Mesmo sob a hipótese simplificadora de que Σ_t depende apenas de X_t , mas não de ω ou Σ_{t-1} (isto é, $\omega = 0$ e $\Psi = \mathbf{0}$), esta equação está longe de ser simples.

O exemplo mais simples para a equação (5.18) ocorre quando Ψ e Λ são matrizes diagonais. Neste caso, o (i, j) -ésimo elemento de Σ_t é

$$\sigma_{ij,t} = \omega_{ij} + \beta_{ij}\sigma_{ij,t-1} + \alpha_{ij}X_{i,t-1}X_{j,t-1}.$$

Para esse modelo, cada elemento de Σ_t segue um modelo GARCH(1,1) univariado.

Um modelo denominado modelo BEKK por Engle e Kroner (1995) trabalha com formas quadráticas no lugar dos elementos individuais de Σ_t . Neste caso,

$$\Sigma_t = C'C + B'\Sigma_{t-1} + A'X_{t-1}X'_{t-1}A, \quad (5.19)$$

em que C é uma matriz triangular com $k(k+1)/2$ parâmetros e, B e C são matrizes quadradas $k \times k$.

Para a unicidade da representação BEKK, $b_{11} > 0$ e $a_{11} > 0$ (elementos de A e B , respectivamente) é assumido. Σ_t é positiva definida sob a condição (suficiente) de que no mínimo uma das matrizes C ou A tenha posto completo. Esta representação para um modelo GARCH(1,1) multivariado pode ser encontrada em Engle e Kroner (1995).

5.2 Métricas para Séries Temporais Multivariadas

Uma questão importante no contexto multivariado é como medir a distância entre duas séries temporais bivariadas (multivariadas). Uma alternativa simples é generalizar a ideia descrita no Capítulo 3. Sejam $\mathbf{X}_{it} = (X_{i1t}, X_{i2t})$, com distribuição \mathbf{F} , e $\mathbf{X}_{jt} = (X_{j1t}, X_{j2t})$, com distribuição \mathbf{H} . Os vetores de média e matrizes de dispersão são, respectivamente $\boldsymbol{\mu}_i$, $\boldsymbol{\mu}_j$, $\boldsymbol{\Sigma}_i$ e $\boldsymbol{\Sigma}_j$. Teremos então, no presente contexto, que cada série temporal bivariada é um vetor de tamanho 2. Para definir uma métrica para séries temporais bivariadas, preservando as propriedades do caso univariado, procedemos da seguinte maneira. Para a série temporal X_t , $t = 1, \dots, K$, denote por $I_X(\omega) \equiv I_{K,X}(\omega)$ o periodograma de X_t na frequência ω . Para as séries temporais bivariadas $\mathbf{X}_{it} = (X_{i1t}, X_{i2t})$ e $\mathbf{X}_{jt} = (X_{j1t}, X_{j2t})$, que para simplificar a notação serão denotadas daqui em diante por $\mathbf{X}_i = (X_{i1}, X_{i2})$ e $\mathbf{X}_j = (X_{j1}, X_{j2})$. Seja $I_{X_{ir}}(\omega)$ o periodograma da componente $r \in \{1, 2\}$ da série temporal bivariada $i \in \{1, \dots, n\}$. Para calcular a distância entre duas séries bivariadas, podemos calcular a distância, utilizando uma medida como, por exemplo, o log do periodograma normalizado, entre todos os seus periodogramas. Essa maneira intuitiva resulta na matriz

$$\mathbf{M}_{ij} = \begin{pmatrix} \frac{1}{K} \sum_{\ell=1}^{\lfloor \frac{K}{2} \rfloor} (\tilde{I}_{X_{i1}}(\omega_\ell) - \tilde{I}_{X_{j1}}(\omega_\ell))^2 & \frac{1}{K} \sum_{\ell=1}^{\lfloor \frac{K}{2} \rfloor} (\tilde{I}_{X_{i1}}(\omega_\ell) - \tilde{I}_{X_{j2}}(\omega_\ell))^2 \\ \frac{1}{K} \sum_{\ell=1}^{\lfloor \frac{K}{2} \rfloor} (\tilde{I}_{X_{i2}}(\omega_\ell) - \tilde{I}_{X_{j1}}(\omega_\ell))^2 & \frac{1}{K} \sum_{\ell=1}^{\lfloor \frac{K}{2} \rfloor} (\tilde{I}_{X_{i2}}(\omega_\ell) - \tilde{I}_{X_{j2}}(\omega_\ell))^2 \end{pmatrix}, \quad (5.20)$$

em que $\tilde{I}_X(\omega) = \log(I_X(\omega)/\hat{\gamma}_X(0))$.

Para efeito de comparação, é necessária a escolha de uma norma para a matriz \mathbf{M}_{ij} . Entre as possíveis escolhas de uma função real como uma norma está o critério do traço. Assim, fica estabelecida uma medida de similaridade entre duas séries temporais bivariadas, \mathbf{X}_i e \mathbf{X}_j , dada por

$$\Phi(\mathbf{X}_i, \mathbf{X}_j) = \text{traço}(\mathbf{M}_{ij}). \quad (5.21)$$

Para escrever (5.20) de forma mais geral, assim como no caso univariado, estabelecemos a seguinte definição.

Definição 5.2.1. Sejam X_i e X_j séries temporais univariadas. Considere a função $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ com a propriedade

$$g(K) = O(K) \text{ quando } K \rightarrow \infty. \quad (5.22)$$

Defina

$$\phi(X_i, X_j) = \frac{1}{K} \sum_{m=1}^{g(K)} \phi(Z_{im}, Z_{jm}), \quad (5.23)$$

em que $\phi(\cdot, \cdot) : \mathbb{R}^2 \rightarrow \mathbb{R}$ e Z_{im} associa a série temporal univariada X_i à posição m , $m = 1, \dots, g(K)$.

Os exemplos mais comuns para a notação Z_{im} são:

- a) $Z_{im} = X_{im}$, com $g(K) = K$;
- b) $Z_{im} = \tilde{I}_{X_i}(\omega_m)$ em que, geralmente, $g(K) = \lfloor K/2 \rfloor$;
- c) e $Z_{im} = \hat{\rho}_{X_i}(m)$, em que $g(K) < K/4$ é recomendado.

Exemplo 5.2.1. Observe que, na equação (5.20), $\phi(X_{i1}, X_{j1}) = K^{-1} \sum_{\ell=1}^{\lfloor \frac{K}{2} \rfloor} (\tilde{I}_{X_{i1}}(\omega_\ell) - \tilde{I}_{X_{j1}}(\omega_\ell))^2$. Nesse caso, $\phi(x, y) = (x - y)^2$ e $Z_{i1m} = \tilde{I}_{X_{i1}}(\omega_m)$, com $g(K) = \lfloor K/2 \rfloor$.

A definição a seguir determina uma classe de métricas para séries temporais bivariadas.

Definição 5.2.2. Sejam \mathbf{X}_i e \mathbf{X}_j séries temporais bivariadas. Uma classe de funções para medir a distância entre séries temporais bivariadas é dada por

$$\Phi(\mathbf{X}_i, \mathbf{X}_j) = \text{traço}(\mathbf{M}_{ij}), \quad (5.24)$$

em que

$$\mathbf{M}_{ij} = \begin{pmatrix} \phi(X_{i1}, X_{j1}) & \phi(X_{i1}, X_{j2}) \\ \phi(X_{i2}, X_{j1}) & \phi(X_{i2}, X_{j2}) \end{pmatrix} \quad (5.25)$$

e $\phi(\cdot, \cdot)$ é dada na Definição 5.2.1.

5.2.1 Normalidade Assintótica no Caso Multivariado

Ao considerarmos esta abordagem multivariada, pretendemos comportar estruturas de correlação entre as séries sem perder as excelentes propriedades advindas da suposição de independência. Por serem $\mathbf{X}_1, \dots, \mathbf{X}_n$ séries temporais bivariadas independentes, U_n pode ser decomposta e escrita como uma soma $U_n = W_n + B_n$, analogamente ao caso univariado dado

em (2.37). O resultado a seguir refere-se à normalidade assintótica de B_n , sob H_0 , para o caso bivariado. Esse resultado é uma extensão natural do caso univariado.

Teorema 5.2.1. *Sejam $\mathbf{X}_1, \dots, \mathbf{X}_n$ séries temporais bivariadas independentes e*

$$T_n = \sum_{1 \leq i < j \leq n} \eta_{nij} \Phi(\mathbf{X}_i, \mathbf{X}_j), \quad (5.26)$$

em que $\Phi(\mathbf{X}_i, \mathbf{X}_j)$ é dado em (5.24), e $\phi(\cdot, \cdot)$ é dado na Definição 5.2.1. Assuma que $\Phi(\mathbf{x}, \mathbf{y})$ forma um sistema ortogonal. Suponha que $\{\eta_{nij}, 1 \leq i < j \leq n, n \geq 1\}$ é um arranjo triangular de variáveis aleatórias independentes de $\{\mathbf{X}_1, \dots, \mathbf{X}_n, n \geq 1\}$ e

$$\sum_{1 \leq i < j \leq K} \eta_{nij}^2 - M_n = o_p(M_n) \text{ quando } n \rightarrow \infty, \quad (5.27)$$

em que $M_n = \sum_{1 \leq i < j \leq n} \eta_{nij}^2$. Suponha também que

$$\sum_{1 \leq \ell < m \leq g(K)} \mathbb{E} \phi(Z_{i\ell}, Z_{j\ell}) \phi(Z_{im}, Z_{jm}) = O(K), \text{ para } s = 1, 2 \quad (5.28)$$

e

$$\sum_{\ell, m=1}^{g(K)} \mathbb{E} \phi(Z_{i\ell}, Z_{j\ell}) \phi(Z_{im}, Z_{jm}) = O(K). \quad (5.29)$$

Então,

$$(M_n V_n^*)^{-\frac{1}{2}} T_n \rightarrow N(0, 1) \text{ quando } n \rightarrow \infty \text{ e } K \rightarrow \infty, \quad (5.30)$$

em que $V_n^* = U_n^{(2,2)} - U_n^{(3)}$,

$$U_n^{(3)} = \frac{1}{n(n-1)(n-2)} \sum_{1 \leq i \neq j \neq \ell \leq n} \Phi(\mathbf{X}_i, \mathbf{X}_j) \Phi(\mathbf{X}_i, \mathbf{X}_\ell) \quad (5.31)$$

e

$$U_n^{(2,2)} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \Phi^2(\mathbf{X}_i, \mathbf{X}_j). \quad (5.32)$$

Demonstração. Utilizando a norma do traço, segue que

$$\Phi(\mathbf{X}_i, \mathbf{X}_j) = \phi(X_{i1}, X_{i2}) + \phi(X_{j1}, X_{j2})$$

e, portanto,

$$\text{Var}\Phi(\mathbf{X}_i, \mathbf{X}_j) = \text{Var}\phi(X_{i1}, X_{i2}) + \text{Var}\phi(X_{j1}, X_{j2}) + 2\text{Cov}(\phi(X_{i1}, X_{i2}), \phi(X_{j1}, X_{j2})). \quad (5.33)$$

Os dois primeiros termos da soma (5.33) remetem ao caso univariado, para os quais a condição (5.28) deve ser satisfeita. No último termo da soma (5.33) temos

$$\begin{aligned} \text{Cov}(\phi(X_{i1}, X_{i2}), \phi(X_{j1}, X_{j2})) &= \text{Cov} \left\{ \frac{1}{K} \sum_{\ell=1}^{g(K)} \phi(Z_{i1\ell}, Z_{j1\ell}), \frac{1}{K} \sum_{m=1}^{g(K)} \phi(Z_{i2m}, Z_{j2m}) \right\} \\ &= \frac{1}{K^2} \left\{ \sum_{l,m=1}^{g(K)} \mathbb{E} \phi(Z_{i1l}, Z_{j1l}) \phi(Z_{i2m}, Z_{j2m}) \right\}. \end{aligned} \quad (5.34)$$

Portanto, para que o resultado ainda seja válido, (5.34) deve ser no máximo da ordem das variâncias, ou seja, no máximo $O(K)$. Logo, se a condição (5.29) também é atendida, temos a normalidade assintótica para T_n em K e/ou n . \square

O Teorema 5.2.1 é restrito à norma do traço, mas é bem geral em relação à escolha da métrica $\phi(\cdot, \cdot)$. Para fins práticos, ainda é necessário encontrar uma classe de funções $\phi(\cdot, \cdot)$ que atendam as condições do Teorema 5.2.1. Essa tarefa é relativamente simples, pela linearidade da generalização proposta e é desenvolvida no teorema seguinte.

Teorema 5.2.2. *Sejam $\mathbf{X}_1, \dots, \mathbf{X}_n$ séries temporais bivariadas i.i.d., estacionárias e seja*

$$T_n = \sum_{1 \leq i < j \leq n} \eta_{nij} \Phi(\mathbf{X}_i, \mathbf{X}_j), \quad (5.35)$$

em que $\Phi(\mathbf{X}_i, \mathbf{X}_j)$, dado em (5.24), é definida a partir das métricas \mathbf{d}_{NP} , \mathbf{d}_{LNP} ou \mathbf{d}_{ACF} . Defina

$$M_n = \sum_{1 \leq i < j \leq n} \eta_{nij}^2,$$

em que η_{nij} é definido em (4.18). Então,

$$(M_n V_n^*)^{-\frac{1}{2}} T_n \rightarrow N(0, 1) \text{ quando } n \rightarrow \infty \text{ e/ou } K \rightarrow \infty, \quad (5.36)$$

em que $V_n^* = U_n^{(2,2)} - U_n^{(3)}$ e, $U_n^{(2,2)}$ e $U_n^{(3)}$ são definidas, respectivamente, por (5.31) e (5.32).

Demonstração. A condição (5.28) é equivalente ao caso univariado. Resta mostrar que o termo de segunda ordem da decomposição de Hoeffding das métricas \mathbf{d}_{NP} e \mathbf{d}_{ACF} atendem a

condição (5.28). Segue do caso univariado que, para X e Y v.a.'s independentes,

$$\phi(X, Y) = -2XY + X\mathbb{E}Y + Y\mathbb{E}X + \mathbb{E}XY = -2(X - \mathbb{E}X)(Y - \mathbb{E}Y).$$

Portanto, em (5.28)

$$\phi(Z_{i1\ell}, Z_{j1\ell}) = -2(Z_{i1\ell} - \mathbb{E}Z_{i1\ell})(Z_{j1\ell} - \mathbb{E}Z_{j1\ell})$$

e

$$\begin{aligned} \mathbb{E}\phi(Z_{i1\ell}, Z_{j1\ell})\phi(Z_{i2m}, Z_{j2m}) &= 4\mathbb{E}(Z_{i1\ell} - \mathbb{E}Z_{i1\ell})(Z_{j1\ell} - \mathbb{E}Z_{j1\ell})(Z_{i2m} - \mathbb{E}Z_{i2m})(Z_{j2m} - \mathbb{E}Z_{j2m}) \\ &= 4\text{Cov}(Z_{i1\ell}, Z_{i2m})\text{Cov}(Z_{j1\ell}, Z_{j2m}). \end{aligned}$$

Assim, a condição (5.28) é verdadeira para a métrica \mathbf{d}_{ACF} , pois se $Z_{i1\ell} = \hat{\rho}_{i1}(\ell)$, então

$$\text{Cov}(\hat{\rho}_{i1}(\ell), \hat{\rho}_{i2}(m)) \leq \text{Cov}(\hat{\rho}_{i1}(\ell), \hat{\rho}_{i1}(m)) = O(K^{-1}),$$

para séries temporais lineares e estacionárias.

Assim, como no caso univariado, utilizando-se da relação $\mathbf{d}_{ACF} = K\mathbf{d}_{NP}$, o resultado segue para a métrica \mathbf{d}_{NP} . □

Para a métrica \mathbf{d}_{LNP} a extensão do resultado será natural, desde que seja demonstrado o caso univariado.

5.3 Estudo de Simulação para o Caso Multivariado

Esta seção é dedicada ao estudo, através de simulações, do comportamento da estatística de teste B_n para séries temporais bivariadas e também para séries com componentes correlacionadas. Consideramos o caso linear estacionário, através dos modelos VAR, VMA, VARMA e VARFIMA, definidos na Seção 5.1, e também alguns dos modelos não lineares mais conhecidos, como VARCH e VGARCH, também definidos na Seção 5.1.

5.3.1 Poder do Teste

Para estudar o comportamento da estatística de teste no caso bivariado, primeiramente realizamos uma simulação para o poder do teste ao nível nominal de 5%. Para isso, consideramos a seguinte configuração dos grupos.

i) Para $n = 12, 24$ e 100 , geramos metade dessas séries, para diferentes valores de K , a partir de um modelo VAR(1), com matriz dos coeficientes autorregressivos $\phi_1^{(1)}$, e a outra metade das séries são geradas a partir de um modelo VARMA(1,1), tendo como matriz dos

coeficientes autorregressivos $\phi_1^{(2)}$, e como matriz dos coeficientes de médias móveis θ_1 , as quais são dadas por

$$\phi_1^{(1)} = \begin{pmatrix} 0.5 & 0.1 \\ 0.3 & 0.4 \end{pmatrix}, \quad \phi_1^{(2)} = \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.1 \end{pmatrix} \quad \text{e} \quad \theta_1 = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{pmatrix}.$$

Tabela 5.1: Poder do teste a 5% para séries temporais bivariadas. Modelo VAR(1) contra modelo VARMA(1,1) definidos em i).

$n \setminus K$	50	100	200	500	1000	10000
12	0.850	0.998	1.000	1.000	1.000	1.000
24	0.998	1.000	1.000	1.000	1.000	1.000
50	1.000	1.000	1.000	1.000	1.000	1.000

5.3.2 Simulação para Comparar as Métricas d_{LNP} , d_P e d_{ACF}

Na Tabela 5.2 apresentamos um estudo de simulação para comparar as métricas d_{LNP} , d_P e d_{ACF} no caso bivariado, nas situações descritas a seguir.

Caso 1: As n séries foram geradas a partir de um modelo VAR(1), em que a matriz de coeficientes autorregressivos é dada por

$$\phi = \begin{pmatrix} 0.5 & 0.1 \\ 0.3 & 0.4 \end{pmatrix}. \quad (5.37)$$

Caso 2: As n séries foram geradas a partir de um modelo VARMA(1,1), em que a matriz de coeficientes autorregressivos é dada por (5.37) e a matriz de coeficientes médias móveis é dada por

$$\theta = \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.1 \end{pmatrix}. \quad (5.38)$$

Caso 3: Metade das n séries foi gerada a partir de um modelo VAR(1), em que a matriz de coeficientes autorregressivos é dada por (5.37) e a outra metade das séries foi gerada a partir de um modelo VARMA(1,1) em que a matriz de coeficientes autorregressivos é dada por (5.38) e a matriz de coeficientes médias móveis é dada por

$$\theta = \begin{pmatrix} 0.5 & 0.2 \\ 0.2 & 0.1 \end{pmatrix}. \quad (5.39)$$

Caso 4: Metade das n séries foi gerada a partir de um modelo VAR(1), em que a matriz

de coeficientes autorregressivos é dada por (5.37) e a outra metade das séries foi gerada a partir de um modelo VGARCH(1,1) em que em que os parâmetros são dados pelas matrizes

$$C = \begin{pmatrix} 1 & 0 \\ 0.02 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0.2 & 0.3 \\ 0.1 & 0.5 \end{pmatrix} \quad \text{e} \quad A = \begin{pmatrix} 0.3 & 0.4 \\ 0.3 & 0.5 \end{pmatrix}. \quad (5.40)$$

Caso 5: Metade das n séries foi gerada a partir de um modelo VARMA(1,1), em que a matriz de coeficientes autorregressivos é dada por (5.37) e a matriz de coeficientes autorregressivos é dada por (5.38). A outra metade das séries foi gerada a partir de um modelo VGARCH(1,1) em que em que os parâmetros são dados em (5.40).

Caso 6: Metade das n séries foi gerada a partir de um modelo VGARCH(1,1), em que os parâmetros são dados pelas matrizes em (5.40). A outra metade das séries foi gerada a partir de um modelo VGARCH(1,1) em que em que os parâmetros são dados pelas matrizes

$$C^* = \begin{pmatrix} 1 & 0.01 \\ 0.2 & 1 \end{pmatrix}, \quad B^* = \begin{pmatrix} 0.5 & 0.0 \\ 0.1 & 0.5 \end{pmatrix} \quad \text{e} \quad A^* = \begin{pmatrix} 0.3 & 0.04 \\ 0.6 & 0.15 \end{pmatrix}. \quad (5.41)$$

Para cada um dos casos consideramos $n = 4, 12$ e 50 , com $K = 100, 300$ e 1000 . Realizamos 10000 replicações para cada um dos casos. Nos casos 3,4 e 5 podemos observar que a métrica d_{LNP} obteve melhor desempenho e, entre as métricas testadas, ela é a melhor opção.

5.3.3 Agrupamento de Séries Temporais Bivariadas Correlacionadas

Nesta subseção, consideramos duas séries temporais correlacionadas como uma série temporal bivariada. Essa configuração permite testar alguma estrutura de correlação entre as séries. Na Tabela 5.3, apresentamos os resultados de três diferentes configurações de modelos. Nos três casos, 12 séries bivariadas são simuladas, sendo divididas em dois grupos de 6. Para cada série bivariada, são simuladas duas séries univariadas componentes. As componentes têm a mesma distribuição e as inovações são correlacionadas. O parâmetro de correlação ρ assume valores em $\{0.0, 0.2, 0.5, 0.9\}$. Para o Caso 1, no primeiro grupo, 6 séries bivariadas são simuladas, em que cada componente segue um processo AR(1) com parâmetro autorregressivo $\phi = 0.3$. O segundo grupo é composto de 6 séries bivariadas, cujas componentes seguem um processo AR(1) com parâmetro autorregressivo $\phi = 0.8$. No Caso 2, o primeiro modelo é um AR(1) com parâmetro autorregressivo $\phi = 0.5$, e o segundo modelo é um ARCH(1) com parâmetro $\phi = 0.5$. Para o Caso 3, são considerados os modelos ARMA(1,1), com parâmetros $\phi = 0.3$ e $\theta = 0.5$, e o modelo GARCH(1,1) com parâmetros $\phi_G = 0.3$ e $\phi_A = 0.5$. Em cada um dos casos são realizadas 1000 replicações.

Podemos observar que o aumento da correlação afeta um pouco a eficácia da técnica de agrupamento, principalmente para K pequeno.

Tabela 5.2: Comparação das métricas d_{LNP} , d_P e d_{ACF} no caso bivariado em diferentes situações. Os Casos 1 e 2 reportam os resultados para os Erros do Tipo I. O restante do Casos reportam o poder do teste.

		d_{LNP}			d_P			d_{ACF}		
Casos	$n \setminus K$	100	300	1000	100	300	1000	100	300	1000
Caso 1	4	0.051	0.050	0.048	0.054	0.054	0.048	0.057	0.051	0.051
	12	0.045	0.052	0.050	0.050	0.050	0.047	0.045	0.045	0.046
	50	0.048	0.050	0.048	0.046	0.048	0.050	0.047	0.048	0.046
Caso 2	4	0.050	0.052	0.049	0.055	0.052	0.050	0.055	0.055	0.051
	12	0.045	0.049	0.050	0.046	0.050	0.051	0.047	0.045	0.044
	50	0.049	0.051	0.048	0.044	0.048	0.050	0.051	0.048	0.047
Caso 3	4	0.601	0.998	1.000	0.073	0.124	0.368	0.429	0.646	0.983
	12	0.995	1.000	1.000	0.195	0.485	0.996	0.915	0.999	1.000
	50	1.000	1.000	1.000	0.833	1.000	1.000	1.000	1.000	1.000
Caso 4	4	0.238	0.884	0.541	0.005	0.004	0.007	0.447	0.788	0.966
	12	0.819	0.999	0.994	0.005	0.010	0.010	0.956	0.999	1.000
	50	1.000	1.000	1.000	0.003	0.005	0.001	1.000	1.000	1.000
Caso 5	4	0.941	1.000	1.000	0.000	0.054	0.037	0.811	0.937	1.000
	12	1.000	1.000	1.000	0.004	0.328	0.047	0.999	1.000	1.000
	50	1.000	1.000	1.000	0.000	0.552	0.143	1.000	1.000	1.000
Caso 6	4	0.049	0.049	0.050	0.002	0.000	0.007	0.048	0.048	0.049
	12	0.050	0.050	0.050	0.005	0.003	0.003	0.046	0.048	0.051
	50	0.047	0.048	0.047	0.001	0.008	0.000	0.043	0.047	0.048

Tabela 5.3: Percentual de acerto para agrupamento de séries temporais correlacionadas.

Casos	$\rho \setminus K$	50	100	200	500	1000	10000
Caso 1	0.0	94.88	99.57	99.92	100	100	100
	0.2	95.30	99.60	99.93	99.98	100	100
	0.5	94.47	99.53	99.90	99.98	100	100
	0.9	90.50	98.33	99.77	99.97	100	100
Caso 2	0.0	85.35	95.05	99.13	99.95	100	100
	0.2	85.00	94.32	99.10	99.93	100	100
	0.5	83.27	94.50	98.95	99.97	100	100
	0.9	83.62	93.00	98.38	99.95	100	100
Caso 3	0.0	97.90	99.77	99.93	100	100	100
	0.2	98.02	99.72	99.98	100	99.98	100
	0.5	97.83	99.48	99.97	99.98	99.98	100
	0.9	95.82	99.17	99.73	99.98	100	100

5.3.4 Agrupamento de Séries Temporais Bivariadas

Nesta seção, avaliamos a performance do nosso método de agrupamento em processos multivariados. Consideramos casos em que, sob H_0 , a normalidade assintótica da estatística de teste B_n ocorre, quando utilizamos as métricas d_{ACF} , d_{NP} e d_{LNP} . Também consideramos situações em que não sabemos se essa normalidade assintótica da estatística ocorre, que é o caso dos modelos não lineares. As três situações simuladas são descritas a seguir:

Caso 1: VAR(1)×VAR(1). Consideramos 12 séries temporais bivariadas geradas a partir de um modelo VAR(1), o qual é definido em (5.5). As primeiras 6 séries têm como matriz dos coeficientes autorregressivos $\phi_1^{(1)}$, e as 6 séries restantes tem matriz dos coeficientes autorregressivos $\phi_1^{(2)}$, as quais são dadas por

$$\phi_1^{(1)} = \begin{pmatrix} 0.5 & 0.1 \\ 0.3 & 0.4 \end{pmatrix} \quad \text{e} \quad \phi_1^{(2)} = \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.1 \end{pmatrix}.$$

Caso 2: VAR(1)×VARMA(1,1). Novamente consideramos 12 séries temporais bivariadas, sendo 6 destas séries geradas a partir de um modelo VAR(1) com matriz dos coeficientes autorregressivos $\phi_1^{(1)}$, e as 6 séries restantes geradas a partir de um modelo VARMA(1,1), com matrizes de coeficientes $\phi_1^{(2)}$ e θ_1 . $\phi_1^{(1)}$ e $\phi_1^{(2)}$ são iguais ao Caso 1 e θ_1 é dado por

$$\theta_1 = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{pmatrix}.$$

Caso 3: VAR(1)×VGARCH(1,1). Neste caso, consideramos um modelo não-estacionário. As 6 primeiras séries são geradas a partir de um modelo VAR(1) com matriz dos coeficientes autorregressivos $\phi_1^{(1)}$, dada no Caso 1. As outras 6 séries são geradas a partir de um modelo VGARCH(1,1), o qual é definido na Subseção 5.1.3. Através da representação BEKK, a matriz de variância depende dos parâmetros A , B e C , os quais foram escolhidos como sendo

$$A = \begin{pmatrix} 0.3 & 0.4 \\ 0.3 & 0.5 \end{pmatrix}, \quad B = \begin{pmatrix} 0.2 & 0.3 \\ 0.1 & 0.5 \end{pmatrix} \quad \text{e} \quad C = \begin{pmatrix} 1 & 0 \\ 0.02 & 1 \end{pmatrix}.$$

Para cada um dos casos, foram realizadas 500 replicações.

Tabela 5.4: Percentual de acerto para agrupamento de séries temporais multivariadas.

Modelo	K					
	50	100	200	500	1000	10000
Caso 1	81.87	94.00	98.20	99.67	99.93	100
Caso 2	93.50	97.55	98.68	99.60	100	100
Caso 3	82.15	90.70	96.25	98.78	99.28	99.73

Podemos observar na Tabela 5.4 que, para o Caso 3, o método tem maior dificuldade para chegar aos 100% de acerto. Um estudo mais aprofundado é necessário para descobrir se essa dificuldade tem relação com a não linearidade do modelo VGARCH. No Caso 2, o percentual de acerto é bem razoável, mesmo para amostras de tamanho pequeno.

Capítulo 6

Séries Temporais Com Outliers

As observações que desviam significativamente da média são usualmente referidas na literatura como *outliers* ou *valores atípicos*. Por muito tempo essas observações foram eliminadas dos conjuntos de dados, pois pensava-se que eram impurezas ou ruídos devido a erros de registro. No entanto, a partir de estudos como de Anscombe (1960), os outliers ganharam maior atenção e atualmente existe uma variedade de aplicações, como monitoramento de condições médicas, detecção de fraudes, detecção de invasão de sistemas, diagnósticos de falhas (Jain (2006), Ferdousi e Maeda (2006), Ueno (2003) e Maruyama e Matsuoka (2010)).

Em séries temporais, a relevância dos outliers aumentou consideravelmente a partir do trabalho de Fox (1972), que introduziu os conceitos de outlier do tipo I e do tipo II, também denominados por outliers aditivos e outliers de inovação. Mais tarde, os outliers também começaram a ser considerados com o desenvolvimento da teoria denominada análise de intervenção. Essa teoria é constituída de métodos que lidam com séries temporais que foram afetadas por acontecimentos inesperados e incontrolláveis, como, por exemplo, greves, mudanças políticas, guerras e catástrofes ambientais, como os derramamentos de petróleo e secas. Essa área de estudo foi desenvolvida primeiramente por Box e Tiao (1975) e, posteriormente, por Tiao (1985), Chang et al. (1988) e Abraham e Chuang (1993), entre outros.

Um dos problemas mais comuns na análise de séries temporais com outliers é o comprometimento dos procedimentos usuais de modelagem, que podem ser induzidos à identificação incorreta do modelo e à estimação viesada dos parâmetros do modelo. Fajardo et al. (2009) estudam o impacto de outliers aditivos nas estimativas de parâmetros de modelos estacionários, especialmente o modelo ARFIMA.

Como visto no Capítulo 3, a estacionariedade é necessária para que, sob H_0 , a estatística de teste com o núcleo baseado em métricas como \mathbf{d}_{LNP} , \mathbf{d}_{NP} e \mathbf{d}_{ACF} , seja assintoticamente normal. No entanto, forças exógenas ou intervenções podem fazer com que as séries se tornem não estacionárias (Chareka et al. (2006)). Por esse motivo, desenvolvemos este capítulo para estudar o comportamento da estatística de teste B_n em séries temporais com outliers. Na Seção 6.1, apresentamos os outliers aditivos e os outliers de inovação. Na Seção 6.2, uma atenção especial é dada ao trabalho de Fajardo et al. (2009), o qual apresenta uma decom-

posição do periodograma para séries temporais com outliers. Na Seção 6.3 são apresentadas diversas simulações com o objetivo de estudar o comportamento do teste em séries temporais com outliers aditivos.

6.1 Outliers

Box e Tiao (1975) propõem um modelo de intervenção mais geral que inclui três tipos de intervenções importantes em séries temporais, que são: mudanças de nível, outliers de inovação e outliers aditivos. No entanto, para este trabalho consideramos somente o modelo desenvolvido por Fox (1972), o qual abrange os efeitos dos outliers aditivos e de inovação.

O *outlier aditivo* é caracterizado por ser ele a única observação afetada. O modelo para este tipo de outlier, descrito por Fox (1972), é

$$U_t = \sum_{i=1}^p \phi_i U_{t-i} + \varepsilon_t, \text{ para } t = 1, \dots, n, \quad (6.1)$$

em que ϕ_i , para $i = 1, \dots, p$, são os parâmetros autorregressivos e ε_t são i.i.d. $N(0, \sigma_\varepsilon^2)$. As observações Y_t são tais que

$$Y_t = \begin{cases} U_t, & \text{se } t \neq T, \\ U_T + \Delta, & \text{se } t = T. \end{cases} \quad (6.2)$$

Assume-se que a série não tem tendência e os ϕ_i 's são tomados de tal forma que o processo $\{U_t\}_{t \in \mathbb{Z}}$ seja estacionário. A ordem p da regressão é assumida conhecida. É possível testar se Y_T , para um determinado valor de T , é um outlier ou testar se existe algum outlier na amostra \mathbf{Y} . Estes procedimentos podem ser encontrados em Fox (1972).

Outro tipo de outlier comum é o *outlier de inovação*. Fox (1972) definiu o seguinte modelo para estes outliers

$$Y_t = \sum_{i=1}^p Y_{t-i} + \Delta_t + \varepsilon_t, \quad (6.3)$$

em que

$$\Delta_t = \begin{cases} 0, & \text{se } t \neq T, \\ \Delta, & \text{se } t = T, \end{cases} \quad (6.4)$$

e os ϕ_i 's e ε_t são definidos como em (6.1). Então, o outlier Δ afeta, não somente a observação Y_T , mas também as observações subsequentes Y_{T+1}, \dots, Y_n .

Na Figura 6.1, apresentamos os efeitos de outliers aditivos e de outliers de inovação em um modelo ARFIMA(1, d , 0), com parâmetro autorregressivo $\phi_1 = 0.5$ e com parâmetro de integração fracionária $d = 0.3$. Adicionamos outliers de magnitude 0, 5 e 10, com p_j definidos em (6.6), dado por $p_j = 0.05$ e com inovações normalmente distribuídas.

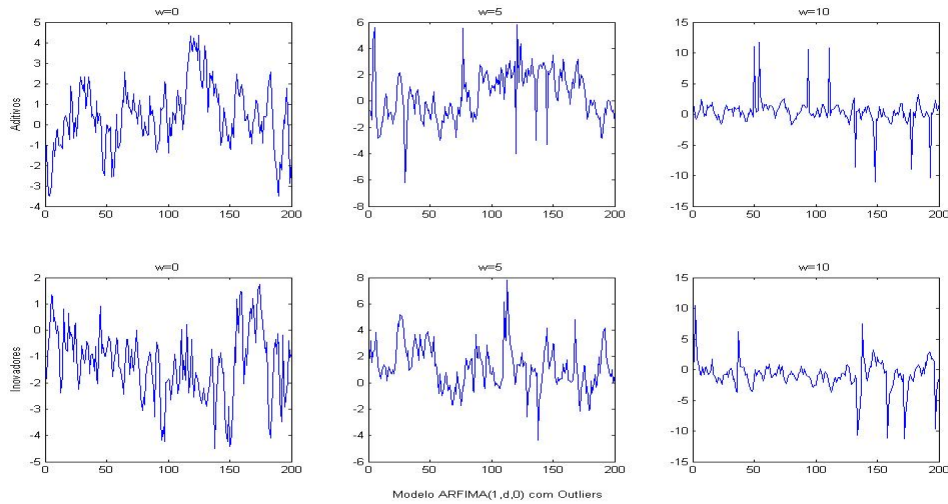


Figura 6.1: Efeito de outliers aditivos e de inovação em um processo ARFIMA(1, d , 0), com tamanho $K = 200$ e para magnitudes ω igual a 0, 5 e 10.

6.2 Outliers Aditivos e Modelos Estacionários

Observações com valores extremos podem afetar a modelagem e por este motivo a detecção de outliers é muito importante. Além disso, é importante saber o grau de impacto de outliers sobre as estimativas dos parâmetros. Fajardo et al. (2009) estudam o impacto de outliers aditivos nas estimativas de parâmetros de modelos estacionários, especialmente no modelo ARFIMA e, em especial, apresentam uma decomposição do periodograma para séries temporais com outliers. Essa decomposição é interessante pela alta dependência da magnitude do outlier em um de seus termos. No trabalho de Fajardo et al. (2009) são considerados processos $\{U_t\}_{t \in \mathbb{Z}}$ fracamente estacionários e processos $\{Y_t\}_{t \in \mathbb{Z}}$ contaminados por outliers aditivos, os quais podem ser descritos por

$$Y_t = U_t + \sum_{j=1}^m \omega_j X_{j,t}, \quad (6.5)$$

em que m é o número máximo de outliers, o parâmetro ω_j representa a magnitude do j -ésimo outlier e $X_{j,t} (\equiv X_j)$ é uma variável aleatória com distribuição de probabilidade

$$Pr(X_j = -1) = Pr(X_j = 1) = p_j/2 \quad \text{e} \quad Pr(X_j = 0) = 1 - p_j, \quad (6.6)$$

em que $\mathbb{E}X_j = 0$ e $\mathbb{E}X_j^2 = \text{Var}X_j = p_j$. Esses outliers afetam a função de densidade espectral, as autocorrelações e, conseqüentemente, o periodograma, que pode ser escrito como

$$I_Y(\lambda) = I_U(\lambda) + \Delta(\omega), \quad \lambda \in (-\pi, \pi], \quad (6.7)$$

em que $I_Y(\lambda)$ é o periodograma da série Y ,

$$\Delta(\omega) = \frac{\omega^2}{2\pi n} \pm \frac{\omega}{\pi n} \left\{ (U_T - \bar{U}) + \sum_{h=1}^{K-1} (U_{T-h} + U_{T+h} - 2\bar{U}) \cos(h\lambda) \right\} + o_p(n^{-1}),$$

e T é o tempo em que ocorre o outlier de magnitude ω .

Pretendemos utilizar a propriedade do periodograma (6.7), que é a forte dependência da magnitude ω (da ordem de ω^2), para agrupamento e classificação de séries. Supondo que as séries $\mathbf{X}_1, \dots, \mathbf{X}_n$, $n \geq 4$, são divididas em dois grupos, em que as primeiras $n_1 \geq 2$ séries têm outliers e as $n_2 \geq 2$, com $n_1 + n_2 = n$, séries não têm outliers, espera-se que a discrepância entre os periodogramas de séries com outliers e de séries sem outliers contribua para um valor grande das medidas entre grupos e um valor pequeno das medidas dentro dos grupos. Assim, seria possível separar séries com outliers de séries sem outliers e, além disso, se a diferença entre as magnitudes for razoavelmente grande, poderíamos separar as séries em grupos com outliers de mesma magnitude. Este estudo de séries com outliers deve ser dividido em duas etapas. Primeiramente considerando outliers aditivos e na sequência outliers de inovação.

6.3 Simulação com Outliers Aditivos

Primeiramente, realizamos uma simulação para comparar as métricas \mathbf{d}_{LNP} , \mathbf{d}_P e \mathbf{d}_{ACF} na presença de outliers aditivos. Simulamos séries ARFIMA(1, d , 0), em que o parâmetro autorregressivo é $\phi_1 = 0.5$ e o parâmetro de integração fracionária é $d = 0.3$. Adicionamos outliers aditivos de magnitude 3, 5 e 10, com p_j definido em (6.6), dado por $p_j = 0.05$ e com inovações normalmente distribuídas. Testamos a hipótese de homogeneidade H_0 para grupos de tamanho n_g igual a 2, 6 e 25, com K igual a 100, 300 e 1000. Os 4 grupos homogêneos gerados são: sem outliers (ou com outlier de magnitude 0); com um outlier de magnitude 3; com um outlier de magnitude 5 e com um outlier de magnitude 10. As configurações testadas são 3×0 , 5×0 , 10×0 , 5×3 , 10×3 e 10×5 . Foram realizadas 1000 replicações.

A partir dos resultados da Tabela 6.1, notamos que a métrica \mathbf{d}_{LNP} é a melhor opção para ser utilizada como núcleo da estatística de teste.

Realizamos uma simulação para verificar se o método de agrupamento é apresenta bom desempenho quando o objetivo é agrupar as séries com diferentes magnitudes de outliers. Na Figura 6.2, apresentamos o resultado de simulações com séries temporais com diferentes magnitudes de outliers. Cada grupo tem tamanho $n_g = 6$, para $g = 1, 2$, e os casos considerados foram: amplitude $w_1 = 3$ contra amplitude $w_2 = 0$; amplitude $w_1 = 5$ contra amplitude

Tabela 6.1: Comparação da métricas d_{LNP} , d_P e d_{ACF} na presença de outliers aditivos, através do poder empírico.

1000 Replicações		d_{LNP}			d_P			d_{ACF}		
$w_1 \times w_2$	$n \setminus K$	100	300	1000	100	300	1000	100	300	1000
3×0	4	0.670	0.154	0.422	0.480	0.450	0.570	0.550	0.630	0.480
	12	0.234	0.557	0.998	0.540	0.600	0.440	0.790	0.840	0.640
	50	0.825	1.000	1.000	0.530	0.570	0.460	0.169	0.158	0.166
5×0	4	0.171	0.479	0.971	0.590	0.300	0.560	0.810	0.780	0.730
	12	0.658	0.996	1.000	0.640	0.650	0.650	0.176	0.211	0.220
	50	0.999	1.000	1.000	0.276	0.226	0.171	0.748	0.893	0.893
10×0	4	0.423	0.927	1.000	0.174	0.214	0.156	0.221	0.234	0.272
	12	0.973	1.000	1.000	0.557	0.948	0.979	0.732	0.886	0.966
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5×3	4	0.570	0.710	0.125	0.590	0.510	0.540	0.610	0.500	0.610
	12	0.104	0.222	0.590	0.570	0.470	0.570	0.770	0.670	0.870
	50	0.375	0.903	1.000	0.950	0.900	0.630	0.175	0.241	0.276
10×3	4	0.119	0.340	0.874	0.134	0.134	0.128	0.114	0.158	0.171
	12	0.422	0.961	1.000	0.474	0.832	0.893	0.345	0.613	0.738
	50	0.998	1.000	1.000	0.999	1.000	1.000	0.973	1.000	1.000
10×5	4	0.570	0.940	0.187	0.830	0.730	0.750	0.840	0.930	0.121
	12	0.114	0.372	0.891	0.281	0.510	0.615	0.153	0.218	0.390
	50	0.559	0.989	1.000	0.963	1.000	1.000	0.563	0.904	0.992

$w_2 = 0$ e amplitude $w_1 = 10$ contra amplitude $w_2 = 3$. Devemos rejeitar H_0 , ou seja, os p-valores devem ser pequenos. Podemos observar uma certa dificuldade do método em rejeitar a hipótese nula, embora tenha funcionado bem para o caso 5×0 , com $K = 1000$.

Na Tabela 6.2, apresentamos o poder do teste para séries temporais com outliers aditivos de magnitude w e inovações independentes com distribuição t . Diferentes configurações são testadas, incluindo dois valores distintos para os graus de liberdade, gl , da distribuição t . São considerados os graus de liberdade igual a 8 e 12. A característica mais importante dos resultados obtidos é que, mesmo na pior situação simulada em que as magnitudes são próximas (caso 5×3), aumentando-se o tamanho amostral n e/ou o tamanho das séries K , o poder de rejeição do teste chega a 100%.

Como podemos observar nas Tabelas 6.2 e 6.3, as simulações com 1000 replicações apresentam resultado bem similares com as simulações com 10000 replicações, e os graus de liberdade não afetam visivelmente os resultados.

Realizamos também uma simulação com o método de agrupamento para diferentes magnitudes de outliers em séries temporais com inovações t . São considerados graus de liberdade igual a 8 e 12 para os tamanhos K igual a 100, 300 e 1000. Cada grupo de 6 séries é simulado

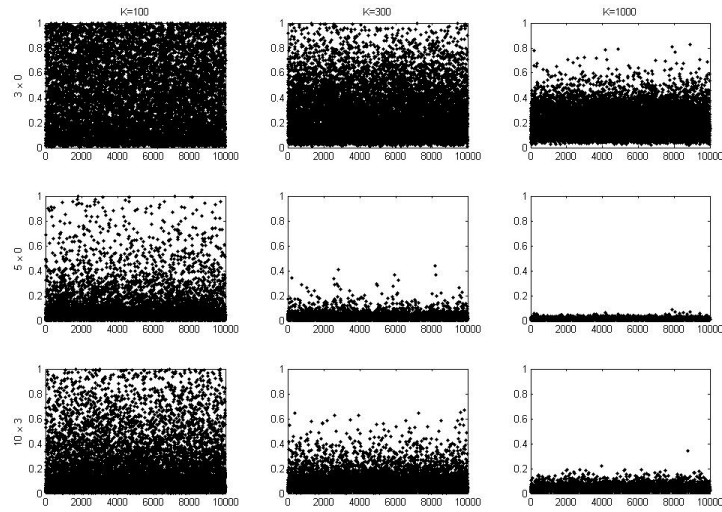


Figura 6.2: Os p-valores do teste de 6 contra 6 séries com diferentes magnitudes de outliers.

Tabela 6.2: Poder do teste para diferentes magnitudes de outliers em séries temporais com inovações t .

1000 Replicações		$gl=8$			$gl=12$		
$w_1 \times w_2$	$n \setminus K$	100	300	1000	100	300	1000
3×0	12	0.132	0.408	0.929	0.178	0.475	0.948
	24	0.309	0.774	0.999	0.367	0.906	1.000
	100	0.982	1.000	1.000	0.990	1.000	1.000
5×0	12	0.480	0.963	1.000	0.519	0.976	1.000
	24	0.845	1.000	1.000	0.908	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000
10×0	12	0.955	1.000	1.000	0.966	1.000	1.000
	24	0.999	1.000	1.000	0.999	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000
5×3	12	0.084	0.192	0.567	0.099	0.211	0.536
	24	0.150	0.422	0.945	0.158	0.472	0.967
	100	0.677	0.996	0.100	0.688	0.998	1.000
10×3	12	0.447	0.973	1.000	0.464	0.972	1.000
	24	0.836	1.000	1.000	0.855	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000
10×5	12	0.139	0.395	0.903	0.130	0.393	0.899
	24	0.274	0.796	1.000	0.289	0.773	1.000
	100	0.957	1.000	1.000	0.962	1.000	1.000

com uma magnitude w igual a 0, 3, 5 ou 10. O modelo considerado é o mesmo utilizado no estudo apresentado na Figura 6.2. Os percentuais de acerto do método de agrupamento, uti-

Tabela 6.3: Poder do teste para diferentes magnitudes de outliers em séries temporais com inovações t .

10000 Replicações		$gl=8$			$gl=12$		
$w_1 \times w_2$	$n \setminus K$	100	300	1000	100	300	1000
3×0	12	0.155	0.408	0.914	0.179	0.468	0.950
	24	0.320	0.813	0.999	0.365	0.875	1.000
	100	0.968	1.000	1.000	0.985	1.000	1.000
5×0	12	0.491	0.963	1.000	0.541	0.979	1.000
	24	0.861	1.000	1.000	0.90.0	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000
10×0	12	0.952	1.000	1.000	0.958	1.000	1.000
	24	0.999	1.000	1.000	0.999	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000
5×3	12	0.093	0.190	0.553	0.9.40	0.200	0.573
	24	0.163	0.438	0.943	0.166	0.457	0.954
	100	0.690	0.998	0.100	0.716	0.998	1.000
10×3	12	0.475	0.964	1.000	0.480	0.967	1.000
	24	0.854	1.000	1.000	0.860	0.100	1.000
	100	1.000	1.000	1.000	1.000	0.100	1.000
10×5	12	0.139	0.383	0.908	0.134	0.370	0.894
	24	0.308	0.800	0.999	0.290	0.790	0.999
	100	0.960	1.000	1.000	0.946	0.100	1.000

lizando a métrica d_{LNP} , com todas as possíveis combinações de magnitudes, estão na Tabela 6.4.

Tabela 6.4: Percentual de acerto para o método de agrupamento, para diferentes magnitudes de outliers em séries temporais com inovações t .

500 Replicações		$gl=8$			$gl=12$		
$w_1 \times w_2$	$\setminus K$	100	300	1000	100	300	1000
3×0		68.63	70.63	85.45	69.10	73.08	88.85
5×0		83.90	96.13	99.88	84.42	96.85	99.90
10×0		96.10	99.60	99.98	96.38	99.70	99.97
5×3		65.52	62.85	66.27	66.30	65.10	67.47
10×3		81.15	95.80	99.93	82.67	95.07	99.97
10×5		67.95	67.80	77.68	66.22	67.40	75.35

Tabela 6.5: Poder do teste para séries não estacionárias com outliers e inovações t .

500 Replicações		d_{LNP}			d_P		
$w_1 \times w_2$	$n \setminus K$	100	300	1000	100	300	1000
3×0	4	0.046	0.044	0.046	0.052	0.052	0.066
	12	0.054	0.054	0.056	0.044	0.036	0.054
	25	0.050	0.072	0.098	0.030	0.040	0.046
5×0	4	0.066	0.076	0.096	0.066	0.054	0.042
	12	0.064	0.094	0.124	0.050	0.050	0.054
	25	0.132	0.276	0.464	0.044	0.052	0.050
10×0	4	0.126	0.182	0.266	0.044	0.066	0.044
	12	0.210	0.422	0.460	0.048	0.052	0.054
	25	0.782	0.952	0.998	0.046	0.052	0.048
5×3	4	0.056	0.060	0.048	0.070	0.062	0.064
	12	0.054	0.062	0.074	0.054	0.062	0.050
	25	0.074	0.130	0.164	0.050	0.054	0.054
10×3	4	0.096	0.170	0.218	0.044	0.054	0.054
	12	0.154	0.266	0.330	0.046	0.054	0.042
	25	0.728	0.932	0.980	0.052	0.052	0.038
10×5	4	0.088	0.114	0.094	0.060	0.040	0.060
	12	0.134	0.222	0.198	0.040	0.050	0.058
	25	0.440	0.656	0.694	0.050	0.050	0.052

Capítulo 7

Discussão

Um dos objetivos iniciais deste trabalho foi a adaptação dos resultados de Pinheiro et al. (2009) para séries temporais. Nestes resultados inclui-se a normalidade assintótica da estatística de teste, para a homogeneidade de G grupos de séries temporais. Como esta estatística depende da escolha de um núcleo e esse é aplicado em séries temporais, a convergência da estatística de teste está vinculada à escolha do núcleo e, conseqüentemente, a uma classe dessas séries temporais. No Capítulo 3 apresentamos uma classe de núcleos em que, sob a hipótese de homogeneidade dos grupos e para séries temporais lineares, as estatísticas de teste associados são assintoticamente normais. No entanto, essa classe depende do conhecimento prévio dos momentos da série, o que é um problema para fins práticos.

A utilização de métricas para séries temporais como núcleos da estatística de teste é fundamental. Por serem baseadas em ferramentas conhecidas e bastante estudadas na literatura de séries temporais, como, por exemplo, o periodograma e a autocorrelação amostral, é possível utilizar suas propriedades para obter a normalidade assintótica das estatísticas de teste baseadas nesses núcleos. Um característica destes resultados é a razão de convergência que, sob a hipótese de homogeneidade dos grupos, é da ordem de $n\sqrt{K}$.

Outro resultado importante desse trabalho é a obtenção da distribuição assintótica da estatística de teste sob a hipótese de heterogeneidade dentro dos grupos. Observamos que a média da estatística de teste depende diretamente da configuração dos grupos. Basicamente, a média aumenta juntamente com a similaridade dentro dos grupos e dissimilaridade entre os grupos, e diminui à medida a similaridade dentro dos grupos e dissimilaridade entre os grupos diminuem. Isso se reflete nos p-valores, proporcionando métodos para agrupamento e classificação. Esses métodos são baseados no fato de que o menor p-valor indica uma maior dissimilaridade dos grupos testados. Assim, testamos todos os agrupamentos possíveis e escolhemos o que tem o menor p-valor. Esse método de agrupamento é comparado, através de simulações, com alguns dos métodos mais conhecidos na literatura e, nas situações testadas, apresenta um desempenho superior ou, no mínimo, igual.

Um estudo de simulação indica que, para as situações testadas, a métrica \mathbf{d}_{LNP} é superior a \mathbf{d}_{NP} , \mathbf{d}_P e \mathbf{d}_{ACF} . No entanto, a classe de séries temporais na qual a métrica \mathbf{d}_P é aplicável

é, possivelmente, maior que as demais citadas, possibilitando inclusive diferenciar séries não estacionárias de séries não estacionárias.

A definição de uma métrica para séries temporais bivariadas e a obtenção da normalidade assintótica da estatística de teste para essas séries, sob a hipótese nula, é um importante avanço para a incorporação de estruturas de dependência entre as séries. Um estudo de simulação indica que o método de agrupamento tem desempenho semelhante ao caso univariado.

A estatística de teste e o método de agrupamento também são estudados, através de simulação, utilizando séries temporais com outliers. Nesse caso, consideramos a magnitude dos outliers como sendo a característica diferencial dos grupos. Pode-se observar que, com magnitudes com diferenças maiores que 2, já é possível agrupar. No entanto, para obtermos 100% de sucesso no agrupamento é necessário diferenças maiores.

Para encontrar os dois grupos com maior similaridade dos elementos dentro do grupo e com maior dissimilaridade entre os grupos, são necessários um número de testes da ordem de 2^{n-1} . Todos esses testes demandam um procedimento bootstrap e, portanto, são computacionalmente intensivos. Se o objetivo é agrupar em 3 diferentes grupos, o número de testes será da ordem de 3^{n-1} . Portanto, um desafio deste trabalho será encontrar um maneira de tornar essa técnica mais eficiente computacionalmente, embora a classificação com três grupos seja igualmente eficiente e rápida. Uma questão em aberto é a demonstração da normalidade assintótica da estatística de teste para a métrica d_{LNP} .

A ampliação e generalização das métricas e a generalização das U-estatísticas de grau 2 para U-estatísticas de grau p são objetivos do projeto de pós-doutorado, o qual foi submetido a Fapesp sob a supervisão do Professor Dr. Pedro Alberto Morettin.

Referências Bibliográficas

- Aaronson, J., Burton, R., Dehling, H., Gilat, D., Hill, T., e Weiss, B. (1996). Strong laws for L - and U -statistics. *Transactions of the American Mathematical Society*, 348(7):2845–2866.
- Abraham, B. e Chuang, A. (1993). Expectation-maximization algorithms and the estimation of time series models in the presence of outliers. *J. Time Ser. Anal.*, 14(3):221–234.
- Aigner, D. J. e Goldfeld, S. M. (1974). Estimation and prediction from aggregate data when aggregates are measured more accurately than their components. *Econometrica*, 42:113–134.
- Anscombe, F. J. (1960). Rejection of outliers. *Technometrics*, 2:123–147.
- Benjamini, Y. e Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B. Methodological*, 57(1):289–300.
- Benjamini, Y. e Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188.
- Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist.* 37 (1966), 51–58; correction, *ibid*, 37:745–746.
- Box, G. E. P. e Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, Calif., revised edition. Holden-Day Series in Time Series Analysis.
- Box, G. E. P., Jenkins, G. M., e Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Prentice Hall Inc., Englewood Cliffs, NJ, third edition. Holden-Day Series in Time Series Analysis.
- Box, G. E. P. e Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *J. Amer. Statist. Assoc.*, 70:70–79.
- Brockwell, P. J. e Davis, R. A. (1991). *Time series: theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition.
- Caiado, J., Crato, N., e Peña, D. (2006). A periodogram-based metric for time series classification. *Comput. Statist. Data Anal.*, 50(10):2668–2684.
- Chang, I., Tiao, G. C., e Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, 30(2):193–204.

- Chareka, P., Matarise, F., e Turner, R. (2006). A test for additive outliers applicable to long-memory time series. *J. Econom. Dynam. Control*, 30(4):595–621.
- Dehling, H. e Taqqu, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U -statistics. *Ann. Statist.*, 17(4):1767–1783.
- Dehling, H. e Taqqu, M. S. (1991). Bivariate symmetric statistics of long-range dependent observations. *J. Statist. Plann. Inference*, 28(2):153–165.
- Denker, M. (1985). *Asymptotic distribution theory in nonparametric statistics*. Advanced Lectures in Mathematics. Friedr. Vieweg & Sohn, Braunschweig.
- Denker, M. e Keller, G. (1983). On U -statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahrsch. Verw. Gebiete*, 64(4):505–522.
- Dynkin, E. B. e Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Statist.*, 11(3):739–745.
- Engle, R. F. e Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric Theory*, 11(1):122–150.
- Fajardo, F., Reisen, V., e Cribari-Neto, F. (2009). Robust estimation in long-memory processes under additive outliers. *J. Statist. Plann. Inference*, 139(8):2511–2525.
- Ferdousi, Z. e Maeda, A. (2006). Unsupervised outlier detection in time series data. In *Proceedings of the 22nd International Conference on Data Engineering Workshops, ICDEW '06*, pages 121–, Washington, DC, USA. IEEE Computer Society.
- Fox, A. J. (1972). Outliers in time series. *J. Roy. Statist. Soc. Ser. B*, 34:350–363.
- Fryzlewicz, P. e Ombao, H. (2009). Consistent classification of nonstationary time series using stochastic wavelet representations. *J. Amer. Statist. Assoc.*, 104(485):299–312.
- Galeano, P. e Peña, D. (2000). Multivariate analysis in vector time series. *Resenhas*, (4):383–404.
- Geweke, J. e Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.*, 4(4):221–238.
- Gini, C. (1912). Variabilita e mutabilita. *Studi Economico-Giuridici della R. Universita di Cagliari*, 3(2):3–59.
- Granger, C. W. J. e Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.*, 1(1):15–29.
- Gvishiani, A. D., Agayan, S. M., Bogoutdinov, S. R., Zlotniki, Z., e Bonnin, Z. (2008). Mathematical methods of geoinformatics. III. Fuzzy comparisons and the detection of anomalies in time series. *Cybernet. Systems Anal.*, 44(3):309–323.
- Halmos, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statistics*, 17:34–43.

- Hirano, S. e Tsumoto, S. (2005). *Clustering Time-Series Medical Databases Based on the Improved Multiscale Matching*. Lecture Notes in Comput. Sci. Springer, New York.
- Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4):800–802.
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, 19:293–325.
- Hoeffding, W. (1961). On sequences of sums of independent random vectors. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. II*, pages 213–226. Univ. California Press, Berkeley, Calif.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist.*, 6(2):65–70.
- Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75.
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.
- Hsing, T. e Wu, W. B. (2004). On weighted U -statistics for stationary processes. *Ann. Probab.*, 32(2):1600–1631.
- Huang, H.-Y., Ombao, H., e Stoffer, D. S. (2004). Discrimination and classification of non-stationary time series using the SLEX model. *J. Amer. Statist. Assoc.*, 99(467):763–774.
- Jain, G. (2006). Monitoring health by detecting drifts and outliers in patterns of an inhabitant in a smart home. Master's thesis.
- Janson, S. (1984). The asymptotic distributions of incomplete U -statistics. *Z. Wahrsch. Verw. Gebiete*, 66(4):495–505.
- Jeng, S.-L. e Huang, Y.-T. (2008). Time series classification based on spectral analysis. *Comm. Statist. Simulation Comput.*, 37(1-2):132–142.
- Kakizawa, Y., Shumway, R. H., e Taniguchi, M. (1998). Discrimination and clustering for multivariate time series. *J. Amer. Statist. Assoc.*, 93(441):328–340.
- Lee, A. J. (1990). *U-statistics*, volume 110 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York. Theory and Practice.
- Lehmann, E. L. (1999). *Elements of large-sample theory*. Springer Texts in Statistics. Springer-Verlag, New York.
- Lund, R., Bassily, H., e Vidakovic, B. (2009). Testing equality of stationary autocovariances. *J. Time Series Anal.*, 30(3):332–348.
- Lütkepohl, H. (1991). *Introduction to multiple time series analysis*. Springer-Verlag, Berlin.
- Maharaj, E. A. (2000). Clusters of time series. *J. Classification*, 17(2):297–314.

- Maharaj, E. A. (2002). Comparison of non-stationary time series in the frequency domain. *Comput. Statist. Data Anal.*, 40(1):131–141.
- Major, P. (1994). Asymptotic distributions for weighted U -statistics. *Ann. Probab.*, 22(3):1514–1535.
- Mandelbrot, B. (1965). Une classe processus stochastiques homothétiques à soi; application à la loi climatologique H. E. Hurst. *C. R. Acad. Sci. Paris*, 260:3274–3277.
- Mandelbrot, B. e Wallis, J. (1969). Computer experiments with fractional gaussian noises. part one, averages and variances. 5(1):228–241.
- Mandelbrot, B. B. e Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437.
- Maruyama, N. e Matsuoka, S. (2010). Model-based fault localization: Finding behavioral outliers in large-scale computing systems. *New Generation Computing*, 28:237–255. 10.1007/s00354-009-0088-6.
- McCullough, B. e Vinod, H. (1998). Implementing the double bootstrap. *Computational Economics*, 12:79–95. 10.1023/A:1008637230094.
- Mezer, A., Yovel, Y., Pasternack, O., Gorne, T., e Assaf, Y. (2009). Cluster analysis of resting-state fmri time series. *Neuroimage*, 45:1117–1125.
- Nobre, J., Singer, J., e Silvapulle, M. (2008). U -tests for variance components in one-way random effects models. *Institute of Mathematical Statistical Collections*, 1:197–210.
- O’Neil, K. A. e Redner, R. A. (1993). Asymptotic distributions of weighted U -statistics of degree 2. *Ann. Probab.*, 21(2):1159–1169.
- Otranto, E. (2008). Clustering heteroskedastic time series by model-based procedures. *Comput. Stat. Data Anal.*, 52:4685–4698.
- Otranto, E. (2010). Identifying financial time series with similar dynamic conditional correlation. *Comput. Statist. Data Anal.*, 54(1):1–15.
- Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*. Wiley Series in Probability and Statistics.
- Piccolo, D. (1990). A distance measure for classifying arima models. *J. Time Series Anal.*, (11):152–164.
- Pinheiro, A. e Pinheiro, H. (2007). *Métodos Estatísticos Não-Paramétricos e suas Aplicações*. Impa.
- Pinheiro, A., Sen, P. K., e Pinheiro, H. P. (2005). Comparison of genomic sequences using the Hamming distance. *J. Statist. Plann. Inference*, 130(1-2):325–339.
- Pinheiro, A., Sen, P. K., e Pinheiro, H. P. (2009). Decomposability of high-dimensional diversity measures: Quasi- u -statistics, martingales and nonstandard asymptotics. *J. Multivar. Anal.*, 100:1645–1656.

- Pinheiro, A., Sen, P. K., e Pinheiro, H. P. (2010). A class of asymptotically normal degenerate quasi U -statistics. *Ann. Inst. of Stat. Math.*, pages 1–18.
- Priestley, M. B. (1988). *Nonlinear and nonstationary time series analysis*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London.
- Reinsel, G. C. (1993). *Elements of multivariate time series analysis*. Springer Series in Statistics. Springer-Verlag, New York.
- Reisen, V. A. (1994). Estimation of the fractional difference parameter in the ARIMA(p, d, q) model using the smoothed periodogram. *J. Time Ser. Anal.*, 15(3):335–350.
- Rifi, M. e Utzet, F. (2000). On the asymptotic behavior of weighted U -statistics. *J. Theoret. Probab.*, 13(1):141–167.
- Robinson, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.*, 23(5):1630–1661.
- Robinson, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *Ann. Statist.*, 23(3):1048–1072.
- Roy, P. M. (1953). A note on the unreduced balanced incomplete block designs. *Sankhyā*, 13:11–16.
- Sarkar, S. K. e Chang, C.-K. (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. *J. Amer. Statist. Assoc.*, 92(440):1601–1608.
- Savvides, A., Promponas, V., e Fokianos, K. (2008). Clustering of biological time series by cepstral coefficients based distances. *Pattern Recognition*, 41:2398–2412.
- Sen, P. K. e Singer, J. M. (1993). *Large Sample Methods in Statistics*. Chapman & Hall, New York. An introduction with applications.
- Shao, X. e Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.*, 35(4):1773–1801.
- Shumway, R. H. e Stoffer, D. S. (2006). *Time series analysis and its applications*. Springer Texts in Statistics. Springer, New York, second edition. With R examples.
- Simpson, E. (1949). The measurement of diversity. *Nature*, 163:688.
- Stute, W. (1991). Conditional U -statistics. *Ann. Probab.*, 19(2):812–825.
- Tiao, G. C. (1985). Autoregressive moving average models, intervention problems and outlier detection in time series. In *Time series in the time domain*, volume 5 of *Handbook of Statist.*, pages 85–118. North-Holland, Amsterdam.
- Ueno, M. (2003). On-line statistical outlier detection of irregular learning processes for e-learning. In Lassner, D. e McNaught, C., editors, *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2003*, pages 227–234, Honolulu, Hawaii, USA. AACE.
- Yoshihara, K.-i. (1976). Limiting behavior of U -statistics for stationary, absolutely regular processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 35(3):237–252.

Apêndice

Alguns Tópicos sobre Séries Temporais

Uma série temporal é um registro de valores de qualquer quantidade variável medida em diferentes pontos do tempo. De uma maneira mais formal, diz que uma série temporal é uma amostra de um processo estocástico. Neste capítulo, reunimos alguns conceitos básicos de séries temporais.

Propriedades Básicas

Na análise clássica de séries temporais, duas suposições são comuns para inferência: (i) estacionariedade e (ii) linearidade. Nesta seção, apresentamos algumas das principais propriedades de séries temporais.

Estacionariedade

A série X_t é denominada *estacionária* se suas propriedades estatísticas não mudam com o tempo. Mais precisamente, X_t é dita ser *completamente estacionária* ou *fortemente estacionária* se, para qualquer conjunto $\{t_1, \dots, t_n\}$, e qualquer inteiro k , a distribuição de probabilidade conjunta de $\{X_{t_1}, \dots, X_{t_n}\}$ é idêntica à distribuição de probabilidade conjunta de $\{X_{t_1+k}, \dots, X_{t_n+k}\}$. Mais estritamente, diz-se que X_t é *fracamente estacionária* ou *estacionária de ordem 2* se somente os momentos comuns até ordem 2 das distribuições de probabilidade acima existem e são idênticos. As séries que são fracamente estacionárias são usualmente chamadas de *estacionárias*, e neste caso tem-se

$$\mathbb{E}X_t = \mu \quad \text{e} \quad \mathbb{E}X_t^2 = \mu_2,$$

em que μ e μ_2 independem de t . Conseqüentemente, $\text{Var}X_t = \sigma^2$ independe de t . Além disso, $\mathbb{E}(X_t X_s)$ será uma função de $(t - s)$. Assim, para uma série estacionária, a média e a covariância de X_t permanece constante ao longo do tempo, e a covariância entre quaisquer dois valores X_t, X_s , dependem somente da separação no tempo e não da localização individual.

Função de Auto-Covariância e Auto-Correlação

Se considerarmos $\text{Cov}(X_t, X_{t+k})$, podemos ver que esta quantidade depende somente de k e não de t . Dessa forma, podemos reescrever

$$\begin{aligned}\text{Cov}(X_t, X_{t+k}) &= \mathbb{E}(X_t - \mu)(X_{t+k} - \mu) \\ &= \gamma(k), \quad r \in \mathbb{Z}.\end{aligned}\tag{1}$$

A função $\Gamma(\cdot)$ é chamado de *função de autocovariância* de X_t . A função de autocorrelação é definida por

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}.\tag{3}$$

Processos Gaussianos

Um processo $\{X_t\}_{t \in \mathbb{Z}}$ é dito ser *Gaussiano* se, para todo t_1, \dots, t_n , o conjunto de variáveis aleatórias $\{X_{t_1}, \dots, X_{t_n}\}$ tem distribuição normal multivariada. Como a distribuição normal multivariada é completamente determinada pela média e a matriz de variâncias e covariâncias, segue que, se $\{X_t\}_{t \in \mathbb{Z}}$ é Gaussiano e estacionário, então $\{X_t\}_{t \in \mathbb{Z}}$ é fortemente estacionário.

Função Densidade Espectral

Quando $\gamma(k)$ decai para zero suficientemente rápido, de tal forma que $\sum_{k=-\infty}^{\infty} \gamma(k) < \infty$, pode-se introduzir a *Transformada Discreta de Fourier* de $\gamma(\cdot)$, a saber

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-i\omega k}, \quad -\pi < \omega \leq \pi.\tag{4}$$

A função $f(\omega)$ é chamado de *função densidade espectral* de X_t . Invertendo-se (4), tem-se

$$\gamma(k) = \int_{-\pi}^{\pi} f(\omega) e^{i\omega k}, \quad k \in \mathbb{Z}.\tag{5}$$

Pode ser mostrado que, mesmo se $\gamma(\cdot)$ não decai suficientemente rápido para que a função $f(\cdot)$ exista, pode-se ainda representar $\gamma(\cdot)$ de forma similar a (5), utilizando a *transformada generalizada de Fourier*, a saber

$$\gamma(k) = \int_{-\pi}^{\pi} e^{i\omega k} dF(\omega),\tag{6}$$

em que $F(\omega)$ é uma função não decrescente, com $F(-\pi) = 0$, $F(\pi) = \sigma^2$, denominada *função*

de distribuição espectral. Quando $f(\cdot)$ existe, tem-se $dF(\omega) = f(\omega)d\omega$ e

$$F(\omega) = \int_{-\pi}^{\omega} f(\omega)d\omega. \quad (7)$$

Correspondente a (5), existe uma representação análoga para X_t . Suponha que $\mathbb{E}X_t = 0$, então

$$X_t = \int_{-\pi}^{\pi} e^{i\omega t} dZ(\omega), \quad (8)$$

em que $Z(\omega)$ é um processo estocástico (a valor complexo) com incrementos ortogonais, isto é, $\mathbb{E}dZ(\omega)dZ^*(\omega') = 0$, para $\omega \neq \omega'$ e é relacionado com $F(\omega)$ por

$$\mathbb{E}|dZ(\omega)|^2 = dH(\omega). \quad (9)$$

A equação (9) é chamada *representação espectral* de X_t e significa, na prática, que qualquer série temporal, a tempo discreto, estacionária, pode ser representada como uma soma de senos e cossenos envolvendo frequências, que variam continuamente no intervalo $(-\pi, \pi)$, amplitudes aleatórias denotadas por $|dZ(\omega)|$ e fases aleatórias denotadas por $\arg\{dZ(\omega)\}$.

Representação Linear para Processos Estacionários

Uma representação linear alternativa para um processo estacionário que também envolve variáveis ortogonais é a seguinte. Considere um processo estacionário de média zero, X_t , com representação espectral

$$X_t = \int_{-\pi}^{\pi} e^{it\omega} dZ(\omega). \quad (10)$$

Priestley (1988) mostra que se X_t tem espectro absolutamente contínuo, isto é, a função densidade espectral, $f_X(\omega)$, existe para todo ω , então este pode ser representado da forma

$$X_t = \sum_{u=-\infty}^{\infty} \eta_u \varepsilon_{t-u}, \quad (11)$$

que é uma combinação linear do passado, presente e futuro de um processo não-correlacionado ε_t . Se uma condição mais forte é imposta para $f_X(\omega)$, então X_t pode ser representado por uma combinação linear somente do passado e do presente, de um processo não-correlacionado. Esta condição extra é que $\log f_X(\omega)$ seja integrável, isto é,

$$\int_{-\pi}^{\pi} \log f_X(\omega)d\omega > -\infty. \quad (12)$$

Sob tal condição, pode-se encontrar uma função $h(\omega)$ satisfazendo

$$f_X(\omega) = |h(\omega)|^2 = h(\omega)h^*(\omega), \quad (13)$$

e tendo série de fourier unilateral

$$h(\omega) = \sum_{u=0}^{\infty} \eta_u e^{-i\omega u}. \quad (14)$$

Dessa forma, obtém-se uma representação linear unilateral para X_t , que é dada por

$$X_t = \sum_{u=0}^{\infty} \eta_u \varepsilon_{t-u}. \quad (15)$$

Revisão em Modelos Lineares

A forma mais geral de escrever um modelo para X_t é

$$h(\dots, X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2}, \dots) = \varepsilon_t, \quad (16)$$

em que ε_t é processo ruído branco e h alguma função pré-estabelecida. Na prática, assume-se que h é uma função linear de $\dots, X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2}, \dots$ de tal maneira que (16) pode ser escrita da forma

$$\sum_{u=0}^{\infty} \alpha_u X_{t-u} = \varepsilon_t, \quad (17)$$

em que $\{\alpha_u\}$ é uma sequência de constantes.

Introduzindo o operador “backward shift”, B , definido por $BX_t = X_{t-1}$, $B^2X_t = X_{t-2}$, \dots , (17) pode ser expressa da forma

$$H(B)X_t = \varepsilon_t, \quad (18)$$

em que

$$H(z) = \sum_{u=0}^{\infty} h_u z^u. \quad (19)$$

A função H pode ser invertida permitindo-se escrever X_t como uma função linear do presente e do passado de ε_t , como

$$X_t = H^{-1}(B)\varepsilon_t, \quad (20)$$

e se $H^{-1}(z)$ é analítica (isto é, se $H(z)$ não possui raízes dentro do círculo unitário), ela pode

ser escrita da forma

$$H^{-1}(z) = g_0 + g_1z + g_2z^2 + \dots \quad (21)$$

Assim X_t pode ser reescrito como

$$X_t = \sum_{u=0}^{\infty} g_u \varepsilon_{t-u}, \quad (22)$$

ou

$$X_t = G(B)\varepsilon_t, \quad (23)$$

em que

$$G(B) = \sum_{u=0}^{\infty} g_u z^u. \quad (24)$$

A relação entre $H(z)$ e $G(z)$ é dada por $G(z) \equiv H^{-1}(z)$ ou $G(z)H^{-1}(z) \equiv 1$.

Modelos AR, MA e ARMA

Mandelbrot (1965) e Mandelbrot e Van Ness (1968) definem o *Movimento Browniano Fracionário*, que é um processo estocástico estacionário com longa dependência a tempo contínuo, para explicar o efeito Hurst. O nome de *efeito Hurst* foi dado para explicar o comportamento de dependência entre as observações, mesmo distantes, proposto primeiramente pelo hidrólogo Harold E. Hurst, em 1951, enquanto investigava a série temporal dos níveis do rio Nilo. Em seguida, Mandelbrot e Wallis (1969) define o *Ruído Gaussiano Fracionário* que é uma versão do Movimento Browniano Fracionário a tempo discreto, mostrando que este processo também exibe o efeito Hurst. Posteriormente, Box e Jenkins (1976) introduziram a classe dos processos *autorregressivos de médias móveis* (ARMA).

Esses modelos lineares podem ser definidos a partir de suposições para a função G . Assumindo-se que $G^{-1}(z)(\equiv H(z))$ pode ser aproximada por um polinômio de ordem finita,

$$G^{-1}(z) = 1 + a_1z + \dots + a_pz^p, \quad (25)$$

então (22) reduz-se a um processo AR(p)

$$X_t + a_1X_{t-1} + \dots + a_pX_{t-p} = \varepsilon_t. \quad (26)$$

Se $G(z)$ pode ser aproximada por polinômio de ordem finita da forma,

$$G(z) = 1 + b_1z + \cdots + b_qz^q, \quad (27)$$

então (22) reduz-se a um processo MA(q)

$$X_t = \varepsilon_t + b_1\varepsilon_{t-1} + \cdots + b_q\varepsilon_{t-q}. \quad (28)$$

Se $G(z)$ pode ser escrita da forma

$$G(z) = \frac{1 + b_1z + \cdots + b_qz^q}{1 + a_1z + \cdots + a_pz^p}, \quad (29)$$

então, (22) reduz-se a um processo ARMA(p, q), e pode ser escrito da forma

$$X_t + a_1X_{t-1} + \cdots + a_pX_{t-p} = \varepsilon_t + b_1\varepsilon_{t-1} + \cdots + b_q\varepsilon_{t-q}. \quad (30)$$

Granger e Joyeux (1980) e Hosking (1981) estendem a classe de processos ARIMA, definidos em Box et al. (1994), para os chamados *processos autorregressivos de médias móveis com integração fracionária*, denotados por ARFIMA(p, d, q). Esses processos são conhecidos por possuírem longa dependência e são estacionários quando $d \in (0, 0.5)$. Para a situação univariada, considere $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico satisfazendo a equação

$$\Phi(\beta)\nabla^d(\beta)(X_t - \mu) = \Theta(\beta)\varepsilon_t, \quad \text{para todo } t \in \mathbb{Z}, \quad (31)$$

em que μ é a média do processo, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é o processo ruído branco, $\Phi(\cdot)$ e $\Theta(\cdot)$ são os polinômios em β , sendo β o operador de defasagem e $\nabla^d(\beta)$ o operador diferença. Então, $\{X_t\}_{t \in \mathbb{Z}}$ é chamado de *processo autorregressivo fracionariamente integrado de média móvel de ordem* (p, d, q), denotado por ARFIMA(p, d, q), em que d é o grau ou parâmetro de diferenciação, p é o grau do polinômio $\Phi(\cdot)$ e q é o grau do polinômio $\Theta(\cdot)$.

Para a estimação do parâmetro d , um grande número de procedimentos é sugerido na literatura, como pode ser visto em Geweke e Porter-Hudak (1983), Reisen (1994), Robinson (1995a) e Robinson (1995b).

Modelos Lineares e Estacionariedade

Todo processo estacionário que tenha *espectro puramente contínuo* e densidade espectral $f_X(\omega)$ pode ser representado de forma linear

$$X_t = \sum_{u=-\infty}^{\infty} \eta_u \varepsilon_{t-u}, \quad (32)$$

e se $f_X(\omega)$ satisfaz a condição (12), a equação (32) pode ser reescrita da forma

$$X_t = \sum_{u=0}^{\infty} \eta_u \varepsilon_{t-u}, \quad (33)$$

desde que $\{\varepsilon_t\}$ seja uma sequência de variáveis estritamente independentes. Esta propriedade é chamada *causalidade*. Se $\{X_t\}_{t \in \mathbb{Z}}$ é um processo Gaussiano, então podemos relaxar a hipótese de independência e assumir somente que $\{\varepsilon_t\}$ seja uma sequência de variáveis não-correlacionadas.

Seja $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ uma sequência de variáveis aleatórias i.i.d. e seja

$$X_t = G(\dots, \varepsilon_{t-1}, \varepsilon_t), \quad (34)$$

em que G é uma função mensurável, tal que X_t é a própria variável aleatória. Então, o processo $\{X_t\}_{t \in \mathbb{Z}}$ é causal no sentido de que depende somente de $\mathcal{F}_t = (\dots, \varepsilon_{t-1}, \varepsilon_t)$ e não das inovações futuras $\varepsilon_{t+1}, \varepsilon_{t+2}, \dots$. A classe de processos (34) é bem ampla.

Estacionariedade e causalidade são dois conceitos relacionados, mas não necessariamente equivalentes. Para ilustrar este ponto considere $\{\varepsilon_t\}$ um ruído branco com $\text{Var} \varepsilon_t < \infty$ e considere, por exemplo, o processo $y_t = \varepsilon_t + \varepsilon_{t+1}$, esse processo é estacionário mas não é causal, uma vez que y_t depende dos valores futuros da sequência $\{\varepsilon_t\}$. Por outro lado, o processo $y_t = \varepsilon_t + \varepsilon_{t-1}$ é causal e estacionário (Palma (2007)).

Assuma que $\{X_t\}_{t \in \mathbb{Z}}$ tem média zero e função de variância finita. Se $\{X_t\}_{t \in \mathbb{Z}}$ é “short-range dependent”, a saber

$$\sum_{k=0}^{\infty} |\gamma(k)| < \infty, \quad (35)$$

então a densidade spectral

$$f(\omega) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma(k) e^{ik\omega}, \quad \omega \in \mathbb{R}, \quad (36)$$

é contínua e limitada.