

PREVISIBILIDADE DE SÉRIES FINANCEIRAS

Laboratório de Modelagem, Análise e Controle de Sistemas Não Lineares
Departamento de Engenharia Eletrônica
Universidade Federal de Minas Gerais
Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, MG Brasil
Fone: +55 3499-4866 - Fax: +55 3499-4850



Previsibilidade de Séries Financeiras

Leonardo Teles de Carvalho

Dissertação submetida à banca examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Minas Gerais, como parte dos requisitos necessários à obtenção do grau de Mestre em Engenharia Elétrica.

Orientador: Prof. Luis Antonio Aguirre, PhD

Belo Horizonte, junho de 2011

Agradecimentos

Aos que foram sempre presentes e apoiadores desse projeto, além de corretores diligentes, minha família: Gabriela, Cascardo, Dona Beatriz (mamãe) e Lu. Dedico essas linhas, como agradecimento ao amor, compreensão pelas noites não dormidas e longas e enfadonhas conversas. Cada comentário, cada palavra, os *emails* trocados, tudo foi bastante significativo e cooperou para o resultado final, mas, mais importante, para a continuidade da manifestação dos laços dessa importante comunidade que é a minha família. Sou muito grato por ter como orientador o prof. Luis Aguirre, pessoa a qual estimo e tenho como exemplo de profissional, ser humano, discípulo. Em cada correção, pude ver o cuidado e dedicação no trabalho. Em cada conversa, a atenção e carinho com o próximo. Agradeço também ao engenheiro Francisco Lopes, e aos colegas da CBTU, cujo suporte permitiu que várias dessas linhas se tornassem possíveis. Ainda em tempo, agradeço o Jair, por estar também sempre presente e auxiliar nas decisões mais importantes nesses anos de estudo.

Sem Jesus, nada neste projeto, nesses anos passados fariam sentido. O propósito, em cada momento, é dado por nossa missão aqui, que é conhecer e servir ao Mestre. Obrigado Deus, por essas pessoas e pela minha vida, que tenho pelo Senhor, o amigo nas horas solitárias, Quem direciona os tantos erros de rota. Espero um dia conhecê-lo face a face.

Resumo

A baixa previsibilidade e os resultados controversos de previsão de séries financeiras colocam ao pesquisador dessa área uma importante questão: é possível modelar e prever tais séries? Uma análise prévia do determinismo e das dependências temporais existentes deve anteceder tentativas de previsão sob o risco de se tentar *prever o imprevisível*. O presente trabalho apresenta algumas técnicas que podem ser utilizadas no estudo da previsibilidade das séries financeiras. O uso de técnicas como a estimação de dimensões de correlação, o cálculo de entropias e medidas de informação é mostrado como uma forma de quantificar complexidade, regularidade e o fluxo de informações dos dados. Essas ferramentas são utilizadas para se detectar os tipos de determinismo presentes em séries financeiras. Embora essas ferramentas tipicamente dependam de uma grande amostra, o que é raro em séries desse tipo, a confiança estatística dos resultados pode ser levantada utilizando a análise de dados sub-rogados. Também é apresentado o teste BDS, cujo cálculo gera um índice relacionado a previsibilidade das séries financeiras. Verifica-se, no trabalho, a importância do tratamento da heteroscedasticidade nessas séries e como sua presença pode indicar uma falsa previsibilidade. Detectam-se dependências lineares e não lineares dinâmicas homoscedásticas em algumas séries. Essas estão relacionadas com a possível previsibilidade das séries. Esses resultados são validados por resultados de previsão baseados em redes RBF. Mostra-se que há uma correlação entre o resultado das ferramentas de previsibilidade apresentadas e o desempenho de previsão.

Palavras-chave: Previsibilidade, Determinismo, Previsão, Séries Financeiras, Dados Sub-rogados, Teste BDS.

Abstract

Poor forecasting performance and, sometimes, dubious results leads the financial time series researcher to ask if is it possible to really forecast anything from these series. Thus, it is of great relevance the study of the predictability of this kind of data. If it is, somehow, known that these series are to be unpredictable, then there is no point in trying to forecast such series.

Some statistical tools for the detection of time dependencies are presented and used on stock market series. Measures based on the estimation of dimension, entropy and information exchange are used as a way of discriminating between these series and some known models (e.g., white gaussian noise, AR and GARCH models and the logistic map) in terms of complexity, regularity and information flow from past to present values of the series. These measures are used to test the series against the null hipotesis of independence and linear, nonlinear and exclusive heteroscedastic dependence. The discriminating power of these measures was obtained by the use of surrogate data analysis. The presence of heteroscedasticity was seen to have significant undesirable effects on the results. Also, the BDS test was used as a way to quantify the predictability, by considering only the correlations not due to heteroscedasticity.

The ultimate test for predictability of a time series is its forecasting power. In this sense, the results of the predictability study were compared against the prediction errors obtained by RBF modelling of the data. It is shown that the results of both prediction and predictability are in agreement.

Palavras-chave: Predictability, Determinism, Forecasting, Financial Time Series, Surrogate Data, BDS Test.

Sumário

Resumo	iii
Abstract	v
Lista de Figuras	xi
Lista de Tabelas	xv
1 Introdução	1
1.1 Hipótese dos Mercados Eficientes (HME)	2
1.2 Previsão dos Retornos	4
1.3 Previsibilidade dos Retornos	8
1.4 Volatilidade, Risco e Previsibilidade	11
1.5 Algumas Considerações	14
1.6 Objetivos do Trabalho	15
1.7 Organização do Trabalho	16
2 Séries Financeiras e o Passeio Aleatório	19
2.1 As Séries Artificiais Utilizadas	23
2.2 O Teste Ljung-Box	25
2.3 Correlação e Esperança dos Retornos	30
2.4 Testes de Heteroscedasticidade Breusch-Pagan e White	32
2.5 Heteroscedasticidade dos Retornos	37
2.5.1 Teste Ljung-Box do Quadrado dos Retornos	39
2.5.2 Teste de White e Breusch-Pagan nas Séries de Retornos	39
2.6 Não Normalidade dos Retornos	40
2.7 Características Estatísticas das Séries Financeiras	42
3 Métodos e Técnicas	47

3.1	Previsão, Previsibilidade e Determinismo	47
3.2	Dimensão de Correlação	48
3.2.1	O Algoritmo de Judd	51
3.3	Entropia Aproximada e Entropia Amostral	54
3.3.1	Entropia Aproximada	54
3.3.2	Entropia Amostral	57
3.4	Informação Mútua Média	59
3.5	Teste BDS	61
3.6	Dados Sub-rogados	66
3.6.1	Algoritmo 0	68
3.6.2	Algoritmo 1	71
3.6.3	Algoritmo 2	73
3.6.4	Algoritmo PPS	80
4	Previsibilidade de Séries Financeiras	85
4.1	As Séries	85
4.1.1	<i>Dow Jones Industrial Average</i>	85
4.1.2	S&P 500	86
4.1.3	Ibovespa	86
4.1.4	Petrobras	89
4.2	Detecção de Determinismo e Quantificação da Previsibilidade - Abordagens	90
4.3	Independência dos Retornos	92
4.3.1	Teste BDS - Independência	93
4.3.2	Dados Sub-rogados - Algoritmo 0	95
4.4	Dependência Linear dos Retornos	97
4.4.1	Teste BDS - Resíduos de um Modelo Linear	97
4.4.2	Dados Sub-rogados - Algoritmo 1	99
4.5	Dependência Não Linear dos Retornos	101
4.5.1	Teste BDS - Resíduos de um Modelo AR-GARCH	101
4.5.2	Dados Sub-rogados - Algoritmo 2	105
4.5.3	Dados Sub-rogados - Algoritmo PPS	107
4.6	Quantificação da Previsibilidade	108
5	A Previsão como uma Validação das Técnicas de Previsibilidade	113
5.1	Modelo de Previsão	114
5.2	Cálculo dos Erros	118

5.3	Dimensionamento e Estrutura da Rede	120
5.3.1	A escolha de L e Q	121
5.4	Previsão das Séries Financeiras	122
5.4.1	Estimação das Estruturas das Redes	122
5.4.2	Desempenho das Redes	125
5.5	A Evolução de W_2 ao longo das Fases de Previsão	127
6	Resultados e Conclusões	131
6.1	Previsibilidade	132
6.1.1	Testes de Independência - A Inadequação do Passeio Aleatório .	133
6.1.2	Testes de Linearidade - A Detecção de Estruturas Não Lineares	133
6.1.3	Testes de Não Linearidade - A Detecção de Estruturas Não Lineares	134
6.1.4	Teste de Resíduos GARCH - A quantificação da previsibilidade	135
6.2	Previsão	136
6.3	Previsibilidade X Previsão	138
6.4	Conclusões	138
6.5	Sugestões de Trabalhos Futuros	139
	Apêndice A Rotinas Implementadas em MATLAB	141
A.1	geraRegressoresAR.m	141
A.2	calculaDimensaoGP.m	141
A.3	ljungbox_q.m	143
A.4	estimaModeloAR.m	144
A.5	dataSnooping.m	146
A.6	testaHeteroscedasticidade.m	150
A.7	calculaErros.m	154
A.8	analiseBDS.m	155
A.9	analisesurrogates.m	156
A.10	estimaModeloRBF.m	162
A.11	preparaDados.m	167
A.12	previsaoRBF.m	169
	Referências Bibliográficas	171

Lista de Figuras

1.1	As médias móveis são sempre lucrativas?	8
1.2	Comparação entre a distribuição gaussiana e a distribuição empírica dos retornos do índice <i>Dow Jones</i> de 27 de maio de 1896 a 27 de janeiro de 2011.	13
2.1	Não se consegue distinguir, de maneira clara e sem uma análise mais profunda, uma série financeira de séries geradas artificialmente.	20
2.2	Comparação de um passeio aleatório com tendência e o mesmo sem sua tendência.	22
2.3	Séries artificiais a serem trabalhadas.	26
2.4	Distribuição de frequências normalizada do valor de $q(10)$ e distribuição qui-quadrado com 10 graus de liberdade de um processo i.i.d. gaussiano.	27
2.5	$q(10)$ para processos AR(1), (a) com parâmetro 0,2 e (b) com parâmetro 0,4.	28
2.6	Função de autocorrelação dos retornos: (a) DJIA, (b) SP500, (c) IBOV.	31
2.7	Função de autocorrelação dos retornos: (a) PETR4O, (b) PETR4H, (c) PETR4L, (d) PETR4C.	32
2.8	PDF de uma distribuição qui-quadrado com 20 graus de liberdade.	36
2.9	Histograma e distribuição gaussiana estimada das séries: (a) DJIA, (b) SP500, (c) IBOV e (d) IBOV.	43
2.10	Histograma e distribuição gaussiana estimada das séries: (a) PETR4O, (b) PETR4H, (c) PETR4L, (d) PETR4C.	44
3.1	Espaço de fases de WGN e LOGMAP.	49
3.2	Estimativa da dimensão de correlação para o mapa logístico.	51
3.3	Dimensão de correlação estimada em função da dimensão de imersão para o mapa logístico (LOGMAP) e ruído branco gaussiano (WGN).	52
3.4	A distribuição de probabilidades de $P(\mathbf{x}_i - \mathbf{x}_j _\infty < \epsilon)$	53
3.5	A dimensão de correlação \hat{D}_2 em função de ϵ_0 e os erros de regressão na estimativa da distribuição.	55

3.6	Entropia aproximada $ApEn(2,\epsilon,3500)$ para mapas logísticos com valores do parâmetro de bifurcação λ tomados entre 3,5 e 4.	57
3.7	Entropia amostral $SampEn(2,\epsilon,3500)$ para mapas logísticos com valores do parâmetro de bifurcação λ entre 3,5 e 4.	59
3.8	Informação mútua média (AMI) calculada para séries AR(1).	61
3.9	Gráfico do logaritmo da integral de correlação $C(\epsilon)$ em função do logaritmo da distância ϵ para a série WGN.	63
3.10	Comparação entre $C_m(\epsilon)$ e $C_1(\epsilon)^m$ para a série WGN.	64
3.11	Algoritmo 0 - comparação de $D_2(\epsilon_0)$ entre as séries WGN e AR(1) e seus respectivos sub-rogados.	69
3.12	Algoritmo 0 - comparação de $SampEn(2,\epsilon,3500)$ entre as séries WGN e AR(1) e seus respectivos sub-rogados.	69
3.13	Algoritmo 0 - comparação da estatística AMI entre as séries WGN e AR(1) e seus respectivos sub-rogados.	70
3.14	Algoritmo 1 - $D_2(\epsilon_0)$ calculado para AR(1), MA(1) uniforme e ARCH(1) e seus respectivos sub-rogados.	73
3.15	Algoritmo 1 - Estatística discriminante $SampEn(2,\epsilon,3500)$ calculada para as séries AR(1), MA(1) uniforme e ARCH(1) e seus respectivos sub-rogados.	74
3.16	Algoritmo 1 - AMI.	75
3.17	Algoritmo 2 - $D_2(\epsilon_0)$ calculada para AR(1), AR(1) uniforme, AR-TENL(1) e ARCH(1) e seus respectivos sub-rogados.	77
3.18	Algoritmo 2 - $SampEn(2,\epsilon,3500)$ calculada para AR(1), AR(1) uniforme, AR-TENL(1) e ARCH(1) e seus respectivos sub-rogados.	78
3.19	Algoritmo 2 - estatística AMI calculada para AR(1), AR(1) uniforme, AR-TENL(1) e ARCH(1) e seus respectivos sub-rogados.	79
3.20	Algoritmo PPS - D_2	83
3.21	Algoritmo PPS - $SampEn(2,\epsilon,3500)$	83
3.22	Algoritmo PPS - AMI.	84
4.1	Índice <i>Dow Jones</i>	87
4.2	Índice S&P500.	88
4.3	Índice Ibovespa.	89
4.4	Ações preferenciais da Petrobras.	91
4.5	Fluxograma dos testes BDS.	92
4.6	Fluxograma da análise de dados sub-rogados.	93
4.7	Fluxograma do cálculo do índice de previsibilidade.	94
4.8	Teste BDS de independência.	95

4.9	Análise de dados sub-rogados - algoritmo 0.	96
4.10	Teste BDS de dependência linear.	99
4.11	Análise de dados sub-rogados - algoritmo 1.	101
4.12	Análise de dados sub-rogados - algoritmo 1.	102
4.13	Teste BDS de dependência não linear.	104
4.14	Análise de dados sub-rogados - algoritmo 2.	106
4.15	Análise de dados sub-rogados - algoritmo PPS.	107
4.16	Análise de dados sub-rogados - algoritmo PPS.	108
4.17	Análise de dados sub-rogados - algoritmo PPS.	109
4.18	Teste BDS de dependência exclusivamente heteroscedástica (H_{het}).	110
5.1	Dois padrões em um espaço de entrada de duas dimensões.	115
5.2	Mapeamento da camada escondida realizado por meio do treinamento dos dois neurônios.	117
5.3	Validação cruzada de DJIA, SP500, IBOV e PETO.	123
5.4	Validação cruzada de PETH, PETL e PETC.	124

Lista de Tabelas

2.1	Valores obtidos da estatística $q(10)$ para várias séries conhecidas e seus quadrados.	29
2.2	Estatística $q(10)$ para as séries trabalhadas.	33
2.3	Estatística $LM(5)$ para várias séries conhecidas.	35
2.4	Estatística $LM(20)$ para várias séries conhecidas.	37
2.5	Estatística $q(10)$ para o quadrado das séries trabalhadas.	39
2.6	Estatísticas $LM(5)$ e $LM(20)$ utilizadas nos testes de White e Breusch-Pagan, respectivamente.	40
2.7	Resultado do teste de normalidade Kolmogorov-Smirnov para as séries trabalhadas e para séries artificiais.	41
3.1	Estatística W_2 para várias séries conhecidas.	64
3.2	Índice de previsibilidade (Equação 3.12) das séries WGN e AR(1) para o teste de independência do algoritmo 0.	70
3.3	Índice de previsibilidade (Equação 3.12) das séries AR(1), MA(1) uniforme e ARCH(1) para o teste de linearidade do algoritmo 1.	73
3.4	Índice de previsibilidade (Equação 3.12) das séries AR(1), AR(1) uniforme, AR-TENL(1) e ARCH(1) para o teste de linearidade com transformação não linear estática do algoritmo 2.	80
4.1	W_2 e índice de previsibilidade (Equação 3.12) das séries para o teste de independência BDS.	94
4.2	Índice de previsibilidade (Equação 3.12) das séries para o algoritmo 0 utilizando AMI.	97
4.3	W_2 e índice de previsibilidade dos resíduos AR das séries para o teste BDS (Equação 3.12) de dependência linear.	100
4.4	Índice de previsibilidade (Equação 3.12) das séries para o algoritmo 1 utilizando AMI.	102

4.5	W_2 e índice de previsibilidade (Equação 3.12) dos resíduos AR-GARCH das séries para o teste BDS.	103
4.6	Índice de previsibilidade (Equação 3.12) das séries para o algoritmo 2 utilizando AMI.	105
4.7	W_2 e índice de previsibilidade das séries filtradas por GARCH para o teste BDS.	111
5.1	A quantidade de atrasos (L), o número de neurônios (Q) e o erro RMSE de cada rede nos dados de validação.	125
5.2	Desempenho da previsão um passo a frente das redes RBF em séries financeiras.	126
5.3	Estatística W_2 calculada para as séries e para os erros de previsão das mesmas em dados de treinamento (TR) e teste (TE) antes e depois do filtro GARCH.	128
6.1	W_2 e índice de previsibilidade dos resíduos GARCH das séries para o teste BDS.	136
6.2	Desempenho da previsão um passo a frente das redes RBF em séries financeiras.	137
6.3	Comparação entre o índice de previsibilidade e os desempenhos de previsão PB e DIR (do melhor para o pior desempenho).	138

Introdução

Há séculos, a previsibilidade dos mercados financeiros tem despertado o interesse de muitos pesquisadores e profissionais. As previsões nesse mercado têm graus bastante variados de sucesso e resultados frequentemente controversos, o que leva, muitas vezes, a questionar-se a validade desses resultados.

Isaac Newton presenciou uma das primeiras grandes bolhas especulativas do mercado, conhecida como a “Bolha de *South Sea*” (descrita em Malkiel [2007]). Durante esse evento, em 1720, *South Sea Company*, uma empresa inglesa de comércio marítimo obteve um monopólio de exploração de colônias espanholas. Isso levou a uma grande expectativa de ganhos futuros, o que gerou fortes especulações no preço de suas ações. Os preços saltaram de £128, em janeiro de 1720, ao patamar de £1000 em agosto do mesmo ano. Em setembro, a notícia da venda de ações da própria diretoria da empresa fez o preço cair rapidamente e retornar ao nível de £150. Diante desse fenômeno, o físico disse que conseguia descrever o movimento de corpos celestes, mas não a loucura das pessoas. A concepção de que sistemas complexos, como o movimento de planetas, poderiam ser descritos por equações matemáticas, mas o movimento de preços seria completamente aleatório era uma ideia que já acompanhava o mercado.

No final do século XIX, Bachelier comentou que, em um dado momento, compradores e vendedores sempre concordam com um determinado preço, contudo a análise que realizam para tanto chega obrigatoriamente a conclusões divergentes (Bachelier [1900]). Apesar de estarem sujeitos às mesmas informações disponíveis e possuírem um mesmo objetivo, a saber, maximizar o retorno de seu investimento, esses investidores têm ações completamente contrárias.

A argumentação de Bachelier não possui rigor matemático, todavia, faz uma consideração prática dos mercados:

“Opiniões contrárias a respeito dessas variações estão tão divididas que ao mesmo tempo em que compradores acreditam em um aumento do preço, vendedores acreditam em uma diminuição do mesmo (...) Ele (o especu-

lador) analisa as causas que podem influenciar a ascensão ou a queda dos mercados e a amplitude desses movimentos. Suas inferências são completamente pessoais, uma vez que sua contraparte necessariamente tem uma opinião contrária (...) Em um dado instante, o mercado não acredita nem em uma ascensão nem em uma queda do preço verdadeiro.”

Estatisticamente, porém, isso corresponde a considerar que a esperança matemática dos incrementos da série de preços é nula. Hoje esse é um processo conhecido como passeio aleatório (*Random Walk*). Uma das grandes contribuições de Bachelier foi exatamente aproximar a estatística e os processos estocásticos ao estudo de economia e finanças.

A consideração de Bachelier refere-se diretamente à maneira como as informações são absorvidas pelo mercado e refletidas no preço dos ativos, conhecida a partir de 1950 como o estudo da eficiência informacional dos mercados ou Hipótese do Mercado Eficiente (HME). Nessa época, o termo mercado eficiente e todas as suas implicações estavam ainda por surgir, porém a ideia central por trás da HME, que é a da aleatoriedade e não previsibilidade do mercado, já estava presente.

1.1 Hipótese dos Mercados Eficientes (HME)

Essa hipótese é importante por ter se tornado, ao longo dos anos, um fio condutor das pesquisas sobre previsão e previsibilidade de séries financeiras.

Eugene Fama deu forma à HME como é vista hoje. Qualquer informação que surja e possa ser relevante é rapidamente espalhada e tão logo se torne pública, seu valor é descontado do preço do ativo em questão. Em Fama [1970], o autor define como eficiente um mercado em que o preço reflete completamente as informações disponíveis. Por informação, entende-se também estudos e análises que possam justificar uma posição de compra ou venda. Assim, estudos de balanços e contabilidade de empresas, expectativas futuras de mercado, estudos da série histórica de preços, tentativas de identificação de tendências, entre outras ferramentas, já teriam seus efeitos completamente incluídos no preço fazendo com que se torne inútil o resultado dessas ferramentas. A consequência é a não previsibilidade dos preços.

Os testes de eficiência de mercado são divididos em:

1. *Weak Form* ou testes de previsibilidade dos retornos: considera-se apenas a série de preços e o retorno $r(t + 1)$ é definido como a diferença dos logaritmos dos preços p de um ativo entre o tempo $t + 1$ e o tempo t . São realizados testes

com alguns modelos matemáticos para os retornos com o objetivo de mostrar a eficiência. Um modelo comum para esse caso é testar $E[r(t+1)|\Phi(t)] = 0$. Ou seja, a esperança do retorno $r(t+1)$, dado que todo o conjunto de informações Φ esteja disponível no tempo t , é nulo. Um dos principais modelos de equilíbrio usado é *Random Walk*, ou passeio aleatório. Séries que possam ser modeladas como um passeio aleatório são consideradas eficientes;

2. *Semi-Strong Form* ou estudo de eventos: procura-se medir o impacto de eventos públicos, como a distribuição de dividendos, nos preços. Nesse caso, a HME tenta mostrar que, quando um desses eventos ocorre, os preços rapidamente se ajustam ao novo conjunto de informações. Assim, uma estratégia de investimento baseada no acompanhamento de tais eventos não será economicamente lucrativa;
3. *Strong Form* ou informações privilegiadas: o teste procura avaliar se o acesso a informações privilegiadas, que não estão disponíveis amplamente ao mercado, traria alguma vantagem ao investidor que a possuísse.

Em todas as formas anteriores, o conjunto de informações $\Phi(t)$ não está disponível em $t-1$. Além disso $\Phi(t)$ é totalmente incorporado ao preço no instante t , não tendo mais impacto no preço em instantes subsequentes. Estatisticamente, $E[r(t+1)|\Phi(t)] = E[r(t+1)]$. Esse processo, então, não possui dinâmica. Tipicamente $E[r(t+1)] = E[r(t)] = 0$ ou $E[r(t+1)] = E[r(t)] = \mu$ em que μ tem pequeno valor (*Random Walk with drift*). Como $p(t+1) = p(t) \cdot (1 + r(t+1))$ e a melhor previsão para o retorno segundo a HME é $E[r(t+1)] \approx 0$, $E[p(t+1)] = E[p(t)]$. Portanto, a eficiência implica não previsibilidade dos preços.

Pode-se dizer que grande parte da literatura e publicações concentram-se na primeira forma, *Weak Form*. Fama (Fama [1970]) traz uma revisão de vários resultados empíricos a favor disso. Há, entretanto, uma quantidade crescente de artigos contra essa forma de eficiência em determinadas séries. Como exemplo, tem-se Lo & MacKinlay [1988], que mostra como algumas séries semanais de retornos são estatisticamente diferentes de passeios aleatórios. Porém, não há consenso, como pode ser visto em Fama [1991].

Atualmente, inclui-se também nesses testes o uso de outras variáveis (exógenas) que não os valores passados da série. Logo, as previsões de séries financeiras podem ser consideradas como calcular $E[r(t+1)|\Phi(t)]$, em que $\Phi(t)$ são as informações disponíveis no tempo, incluindo-se outras séries ou dados que possam influenciar o retorno $r(t+1)$. Se $E[r(t+1)]$ varia ao longo do tempo (é diferente para cada t), essa variação pode ter origem na ineficiência ou ser racionalmente explicável e eficientemente incorporada ao

preço. Essa questão sobre a origem das variações no retorno esperado ainda está em discussão.

Dimson e Mussavian fazem um resumo da história da HME além de algumas contribuições mais recentes relacionadas a anomalias (ineficiências) de mercado (Dimson & Mussavian [1998]). Esses autores mostram evidências que revelam ineficiência informacional em estudos de sazonalidades, IPO's ¹ persistentemente negativos, variações de retorno relacionadas ao tamanho das empresas (*size effects*), entre outros.

1.2 Previsão dos Retornos

A previsão dos retornos de séries financeiras é, na atualidade, um tópico bastante recorrente. Preços de ativos são transformados em uma série temporal de retornos e diversas técnicas são utilizadas nas tentativas de se prever valores futuros dessa série. O uso de redes neurais, sistemas nebulosos e computação evolucionária a fim de se atingir esse objetivo está muito presente na literatura. Talvez isso se dê, atualmente, pela facilidade de implementação computacional dessas técnicas, somada à capacidade das mesmas de modelar dinâmicas complexas sem informação prévia do sistema (modelagem caixa preta).

Sitte e Sitte (Sitte & Sitte [2002]) mostram que previsões de um passo à frente usando redes *Multi-Layer Perceptron* (MLP) com arquitetura *time delay neural network* (TDNN) ou redes de Elman dão resultados próximos a um passeio aleatório. As previsões de preços em instantes de tempo $t+1$ aproximam-se muito do último valor do preço em t . É usada uma série de preços subtraída de uma curva exponencial de tendência. Os autores, porém, não ponderam se essa curva de tendência poderia ser incluída na série e prevista. Refenes (Refenes *et al.* [1997]) mostra algumas metodologias para o uso de redes neurais na previsão de volatilidade e seleção de variáveis de entrada para essa previsão. Em Pantazopoulos *et al.* [1998], são apresentadas técnicas de previsão utilizando modelos *neurofuzzy*. Embora os autores não comparem os resultados obtidos com o de outros modelos (como o passeio aleatório), esses são aparentemente muito bons, especialmente quando o modelo é utilizado para se prever a volatilidade em negociações de opções. Caetano e Yoneyama propõem um índice para a detecção de crises do mercado de ações baseado nos coeficientes de uma transformada Wavelet (Caetano & Yoneyama [2009]). Em Lee & Jo [1999], utiliza-se um sistema especialista baseado em gráficos *candlestick* para se prever retornos de ações. Kuo *et al.* [2001] traz a utilização de sistemas especialistas como fonte de dados qualitativos do mercado

¹*Initial Public Offering* ou IPO - Lançamento inicial de ações em bolsa por uma empresa.

para uma rede *neurofuzzy*. Além disso, são mostrados dados quantitativos gerados por meio de valores passados da série. Parte do treinamento da rede é feita por meio de algoritmos genéticos.

Deboeck (Deboeck [1994]) compila vários artigos sobre previsão de séries financeiras com o uso de redes neurais, sistemas nebulosos e sistemas não lineares. O autor mostra o uso de redes neurais na previsão de algumas séries como a do índice de preços de ações das bolsas de Tóquio (TOPIX) e Taiwan (TAIEX) e das notas do tesouro americano:

- o índice TOPIX é previsto utilizando-se redes MLP e, como entrada, algumas variáveis exógenas, como o volume de negociações, taxa de juros e o índice *Dow Jones* com um bom grau de sucesso. Embora o autor não detalhe a construção da rede nem a obtenção dos resultados, uma taxa de 62,1% de acerto na direção dos preços (retornos positivos ou negativos) foi obtida;
- a previsão do índice TAIEX também faz uso de uma rede MLP, mas as entradas utilizadas são baseadas em médias móveis e em outros indicadores utilizados na análise técnica². O resultado, medido em termos da quantidade de transações lucrativas utilizando-se as previsões, não é, porém, estatisticamente significativo devido ao baixo número de transações (apenas 4);
- para a previsão das séries de notas do tesouro americano, o autor analisou primeiramente o coeficiente de *Hurst*, usado como uma medida de previsibilidade. As séries consideradas mais previsíveis foram então utilizadas obtendo uma porcentagem de transações lucrativas de 61,1% e 72,9% nos dois períodos do estudo. O autor também não detalha a rede utilizada.

As técnicas derivadas da análise de sistemas dinâmicos não lineares também influenciam muitas das publicações encontradas atualmente. Como exemplo, a imersão de séries temporais e o cálculo da dimensão de correlação são bastante utilizadas para dimensionar modelos de previsão ou verificar a previsibilidade das séries. Saad e colaboradores comparam a previsibilidade obtida das redes TDNN, probabilística e com estrutura recorrente (a saída da rede é também usada como entrada), aplicadas a algumas séries de ações (Saad *et al.* [1998]). Os autores utilizam técnicas como a estimação da dimensão de correlação e expoentes de Lyapunov no dimensionamento das redes. Os

²A análise técnica é um conjunto de ferramentas utilizadas para se tentar identificar tendências de mercado baseada principalmente na série de preços. Há grande controvérsia a respeito da eficácia dessa metodologia. E há frequentemente divergências entre os próprios analistas técnicos a respeito de seus resultados. Uma referência sobre suas técnicas mais comuns pode ser vista em Matsura [2007].

resultados da previsão são positivos, mas semelhantes para todas as redes, sugerindo que a escolha da arquitetura de uma dessas redes seria indiferente em relação à sua capacidade de previsão. Cao e Soofi (Cao & Soofi [1999]) utilizam o método de modelagem linear local, baseada na reconstrução do espaço de fases para prever taxas do mercado de câmbio. O resultado obtido é semelhante à previsão do passeio aleatório, não sendo possível distinguir estatisticamente as duas previsões. Em Doeksen *et al.* [2005], são comparados modelos MLP e modelos nebulosos Takagi-Sugeno e Mamdani. Esses modelos são utilizados na previsão da direção dos retornos (prever se os retornos são negativos ou positivos) de algumas ações e obtém resultados semelhantes. Os autores mostram que a previsão da direção dos retornos pode obter resultados melhores do que a previsão do valor do retorno. Outra colocação relevante dos autores é a de que a escolha da série a ser prevista pode ser mais importante do que a escolha do modelo de previsão. Ou seja, há séries que podem ser mais previsíveis do que outras, o que justificaria um prévio estudo de previsibilidade dos dados.

Ang e Quek propõem um novo algoritmo para treinamento de redes *neurofuzzy* baseadas no modelo de Mamdani e o aplicam à previsão de preços de ações (Ang & Quek [2006]). Os resultados obtidos são muito bons mas podem não ter a confiança estatística mostrada pelos autores, pois o algoritmo proposto é utilizado juntamente com um modelo de médias móveis do preço cujos parâmetros são escolhidos sem justificativa aparente. Os parâmetros livres (escolhidos sem alguma estimativa, treinamento ou justificativa por parte de quem modela os dados) quase sempre influenciam em muito o desempenho dos algoritmos usados. O que chama atenção aqui é a presença de *data snooping biases*³, presentes em muitos artigos sobre previsão de séries financeiras. Brock e colegas colocam que poucos são os testes em séries financeiras que estão completamente livres dessa característica (Brock *et al.* [1992]). Segundo os autores, quanto mais profundos são os estudos em um conjunto de dados, mais provável é que padrões espúrios “interessantes” sejam observados. Séries de preços de ações (ou índices dos mesmos) são, provavelmente, as mais estudadas das séries financeiras. Consequentemente, essas são as mais susceptíveis à presença de *data snooping*. Um exemplo disso pode ser visto na busca de artigos científicos sobre o índice *Dow Jones*: são encontrados mais de 400 artigos sobre o índice em uma pesquisa apenas por títulos ou tópicos em uma conhecida base de dados⁴. Essa, que é mais longa série existente na área, tem

³*Data Snooping biases* são artefatos estatísticos espúrios devidos à aleatoriedade dos dados (Lo & MacKinlay [2002])

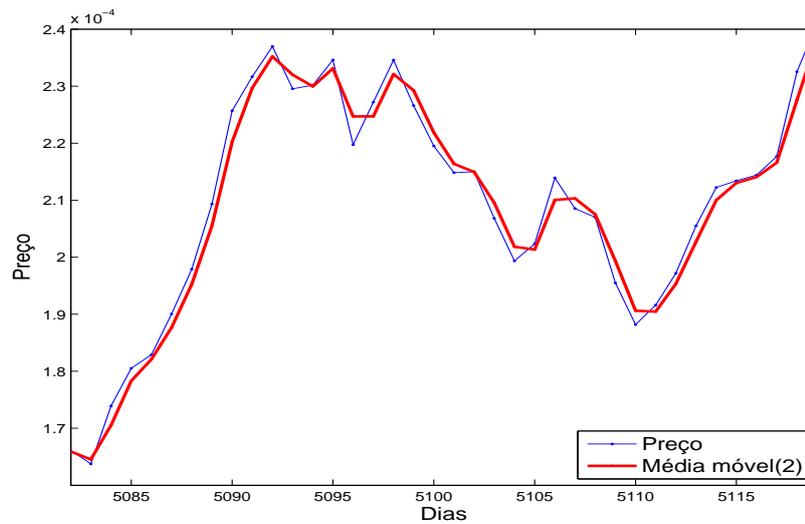
⁴Pesquisa realizada em 12 de maio de 2011, utilizando as palavras chave *DJIA* e *Dow Jones* em títulos e tópicos de artigos, na base *Isi Web of Knowledge* (www.isiknowledge.com). Foram encontrados 452 artigos. Considerando a série de 12386 observações, a média exata obtida é de 27,4 observações por artigo.

atualmente por volta de 28000 observações. Considerando-se a série a partir do ano de 1962 (data do primeiro artigo da pesquisa), são por volta de 12000 observações. Isso produz uma média de 1 publicação a cada 30 observações. Obviamente, esse raciocínio é extremamente simplista, mas chama atenção para o fato de que centenas de publicações, com diferentes objetivos, estudam praticamente os mesmos dados históricos. Como mostrado nesta seção, alguns desses artigos encontram subsídios para uma dependência temporal dos retornos, outros não. Em Lo & MacKinlay [2002], afirma-se que:

“(...) procurar em dados históricos por um desempenho superior em investimentos pode até produzir tal desempenho. Isso, porém, não necessariamente é evidência de que um desempenho superior verdadeiro exista. Ainda que esse desempenho *não exista*, com um conjunto suficientemente grande de dados e uma busca diligente, o desempenho superior, mas espúrio, pode quase sempre ser encontrado. Mesmo que essa busca tendenciosa seja inevitável em disciplinas não experimentais como economia e finanças, reconhecer a possibilidade da sua existência e entender suas propriedades estatísticas pode reduzir em muito seus efeitos.”

Um exemplo disso pode ser visto na utilização, bastante popular, de médias móveis como indicadores de momentos de compra ou venda em séries históricas de ações. A média móvel $MM_t(\tau)$ é calculada em cada instante t de uma série de preços $p(t)$ utilizando-se os τ últimos elementos da série de preços ($[y(t-1) \dots y(t-\tau)]$). Uma estratégia simples baseada em médias móveis ($MM(\tau)$) é comprar um ativo quando o preço histórico ultrapassar as médias móveis e vender caso o preço se torne inferior ao valor de $MM(\tau)$ (momento em que o gráfico de preço corta o de média móveis de cima para baixo). A Figura 1.1 mostra algumas observações do índice Ibovespa e a média móvel dos dois últimos valores do índice ($MM(2)$). A série é dividida em duas partes, cada uma com 5291 observações diárias do índice.

A Figura 1.1.(a) mostra, em um primeiro conjunto de dados que a estratégia anterior ($MM(2)$) sinaliza bons momentos de compra e venda, identificando corretamente as tendências de alta e baixa do índice Ibovespa. As Figuras 1.1.(b) e 1.1.(c) mostram o lucro obtido em cada um dos dois conjuntos de dados. Percebe-se que $MM(2)$ tem desempenho excelente na primeira parte dos dados mas, quando aplicada a um conjunto diferente de dados, não apresenta desempenho tão bom. Isso pode ser visto comparando-se esse lucro com o obtido com a simples estratégia *buy and hold* (Seção 1.5), cujo resultado também é mostrado na figura. Modelos possivelmente pro-



(a) Índice Ibovespa e MM(2) na primeira metade dos dados.

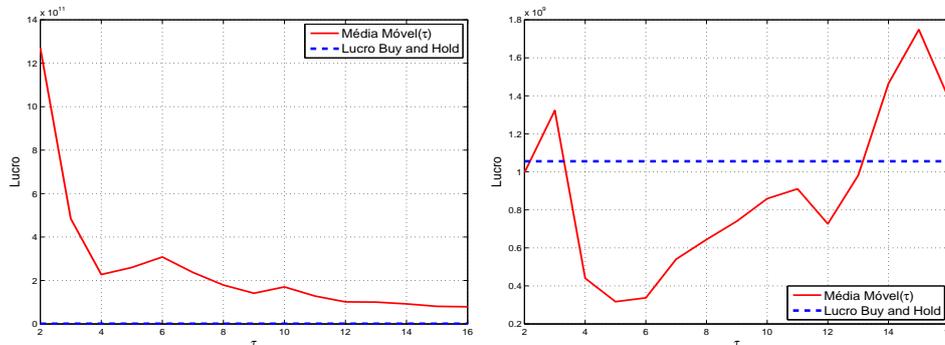
(b) Lucro $MM(\tau)$ na primeira metade dos dados. (c) Lucro $MM(\tau)$ na segunda metade dos dados.

Figura 1.1. As médias móveis são sempre lucrativas? O gráfico (a) mostra o índice Ibovespa e sua média móvel $MM(2)$. Pelo gráfico, percebe-se que a simples estratégia descrita utilizando-se apenas uma média dos dois últimos preços consegue identificar momentos muito bons de compra e venda. O gráfico (b) mostra o lucro obtido no primeiro conjunto de dados, em função de τ . O gráfico (c) mostra o lucro da estratégia $MM(\tau)$ em outro conjunto de dados, em função de τ . Percebe-se, com esse gráfico, que o bom desempenho de $MM(2)$ não é repetido em um conjunto de dados diferente, sendo inferior ao simples *buy and hold*.

missores, como os mostrados em Ang & Quek [2006], tem seus resultados muitas vezes comprometidos pelo *data snooping*.

1.3 Previsibilidade dos Retornos

A dinâmica presente em séries financeiras de retornos de ativos não é facilmente modelável. A própria existência de alguma estrutura determinística é questionada. Talvez a única exceção a essa afirmação seja a dependência temporal geralmente encontrada

no quadrado dos retornos (mostrada no Capítulo 2). Essa dependência, conhecida como heteroscedasticidade, todavia, não faz com que os valores da série de retornos sejam mais previsíveis. Isso significa que, embora a magnitude do retorno futuro possa ser prevista, não se poderia prever se o mesmo seria positivo ou negativo. Uma família de modelos comum para representar essa característica é a de modelos ARCH. Lima e Ohashi, por exemplo, sugerem que o mercado de *commodities* (açúcar) esteja tornando-se mais eficiente, devido à diminuição de coeficientes de um modelo ARGARCH variante no tempo, estimado por filtro de Kalman (Lima & Ohashi [1999]).

Algumas referências da seção anterior citam a hipótese de que as séries financeiras possuem comportamento caótico. Embora isso seja possível, as ferramentas utilizadas não são suficientes para se afirmar isso. Um exemplo dessa limitação pode ser vista em Kyrtsov & Terraza [2003]. Nesse artigo, os autores mostram a estimação da dimensão de correlação e dos expoentes de Lyapunov da série de retornos do índice do mercado francês CAC40. Os algoritmos para estimação dessas grandezas geralmente pressupõem uma grande quantidade de dados e a ausência de ruído. Essas são características praticamente impensáveis nesses tipos de séries. As dimensões encontradas são, tipicamente, de valor elevado. A estimação, tanto das dimensões quanto dos expoentes de Lyapunov apresentam elevada variância. Além disso, a previsibilidade em dados diferentes dos de treinamento dos modelos é sempre próxima à do passeio aleatório. De fato, o que o artigo destaca é a dificuldade em discernir, no caso dessas séries, se o processo que as origina é caótico e imerso em ruído ou puramente estocástico⁵.

Uma abordagem mais conservadora (e plausível) do que tentar detectar a presença de caos é a de se detectar não linearidades não triviais⁶. Discernir se os dados são puramente estocásticos ou apresentam alguma estrutura determinística é fundamental no estudo da previsibilidade. Grandezas como dimensões de correlação, entropias ou expoentes de Lyapunov (denominadas invariantes dinâmicos em Small [2005]) são bastante utilizadas para se detectar determinismo e não linearidades em dados experimentais.

É interessante notar que o cálculo da dimensão de correlação, embora idealizado para dados puramente determinísticos, é, na realidade, baseado no cálculo de uma probabilidade: a probabilidade de que pontos de uma série imersa em um espaço de estados sejam considerados próximos. Mesmo quando aplicados a séries determinísticas, os algoritmos de estimação dessa dimensão produzem resultados estocásticos. Esses

⁵É também, possível, como mostra Kyrtsov & Terraza [2003] que as séries financeiras sejam caóticas e imersas em ruído, mas também apresentem alguma estrutura estocástica.

⁶Considera-se que sejam não linearidades triviais dependências temporais facilmente percebidas nos dados, como é o caso de algumas oscilações, tendências ou sazonalidades. Essas dependências claramente percebidas ou conhecidas *a priori*, muitas vezes interferem na detecção de outras dependências não lineares mais sutis.

podem variar dependendo das observações da série e, principalmente, do comprimento da mesma (Wu *et al.* [1993]). Quando esse conceito é aplicado a séries ruidosas ou a processos estocásticos, os resultados obtidos são também estocásticos. Logo, a simples aplicação direta dessas ferramentas a séries financeiras pode trazer resultados errôneos. Juntamente com o cálculo dessa grandeza, deve existir algum procedimento estatístico que forneça a confiança estatística a essas grandezas estimadas. Como citado em Wu *et al.* [1993], não se sabe, por exemplo, a distribuição de probabilidades de expoentes de Lyapunov ou dimensões de correlação. Consequentemente, torna-se difícil determinar intervalos de confiança para essas grandezas fazendo com que a estimação das mesmas tenham pouca utilidade. No intuito de utilizar a dimensão de correlação para detectar determinismo, os autores sugerem uma estatística baseada na integral de correlação para se detectar independência de séries temporais. O procedimento sugerido quantifica a dependência temporal de $y(t)$ em função dos atrasos $[y(t-1) y(t-2) \dots]$ e gera uma estatística cuja distribuição é conhecida para dados independentes.

Como já mencionado, a falta de uma distribuição estatística para os invariantes dinâmicos pode ser contornada. Uma das técnicas conhecidas para esse fim é a análise de dados sub-rogados (Theiler *et al.* [1992]), semelhante a técnica estatística de *bootstrapping*. Kugiumtzis realiza a análise de dados sub-rogados a séries mensais, semanais e diárias da taxa de câmbio entre *dollar* americano e libra inglesa (USD/GBP) do mercado de câmbio (Kugiumtzis [2000]). Os testes são realizados para verificar a independência dos retornos das séries e a presença de não linearidades na volatilidade das mesmas (algoritmos AAFT, IAAFT e CAAFT). Nesse último teste, os resultados são inconclusivos, uma vez que apenas dois algoritmos utilizados (AAFT e IAAFT) rejeitam essa hipótese. O autor faz uma breve comparação entre os três algoritmos colocando o CAAFT como uma opção mais conservadora de teste. Hinich e colegas, também fazem uso de dados sub-rogados em séries financeiras, utilizando-os como uma alternativa mais robusta (em relação a quantidade de observações) de implementar os testes de não linearidade baseados no biespectro (Hinich *et al.* [2005]). Esses autores mostram que o teste proposto, quando aplicado em uma série de retornos do mercado de ações, indicam a presença de não linearidades que poderiam ser explicadas por modelos da família ARCH. Nakamura e Small aplicam também a análise de dados sub-rogados em algumas séries financeiras (Nakamura & Small [2006]). O algoritmo utilizado é chamado de *small shuffle* e realiza uma permutação aleatória local. Nesse caso, apenas os vizinhos de cada observação $y(t_1)$ (como $y(t_1 \pm 1), y(t_1 \pm 2), \dots, y(t_1 \pm \tau)$ em que $1 < t_1 < N$ e τ é pequeno) são permutados entre si. O algoritmo permite que a análise de dados sub-rogados seja aplicada a séries com tendências (não estacionárias em média), o que é uma grande vantagem, considerando-se a aplicação em séries finan-

ceiras. Os dados gerados apresentam a mesma tendência que os originais, mas tem sua dependência de curto prazo destruídas. Em contrapartida, não é claro se a dependência devida à não estacionariedade pode ocultar dependências sutis que possam existir nos dados.

Scheinkman e LeBaron analisam o determinismo de retornos diários e semanais de ações utilizando testes de independência (Scheinkman & LeBaron [1989]). Os autores aplicam testes BDS aos retornos e a resíduos de um modelo linear e ARCH e verificam que os resultados são compatíveis com a hipótese de presença de não linearidades nos retornos. Além disso, rejeitam a hipótese de independência para os mesmos.

Alguns métodos para detecção de determinismo são baseados em modelos de previsão. Em Small [2005], o erro de um modelo semelhante ao linear local, chamado *nonlinear prediction error*, é utilizado para se distinguir séries estudadas de séries substituídas. A resposta desse modelo é baseada em uma média das respostas de vetores próximos ao vetor de entrada. Os vetores são obtidos após imergir a série em um espaço de fases de determinada dimensão. A diferença dessa modelagem para a linear local é que essa última apresenta um modelo linear (por exemplo, baseado em mínimos quadrados) estimado para os vetores próximos ao vetor de entrada. Gautama e colegas desenvolveram um teste utilizando esse modelo (Gautama *et al.* [2004]). Nesse caso, são calculadas as variâncias de vetores próximos e é gerado um gráfico com as variâncias médias em função da distância.

Há ainda a tentativa de se detectar a previsibilidade de longo prazo das séries, por meio do cálculo do coeficiente de Hurst. Em Cajueiro & Tabak [2004], são mostrados resultados desse coeficiente para índices de ações de países emergentes. Esses resultados são ligeiramente favoráveis a rejeição da hipótese dos mercados eficientes (independência dos retornos). Em Carbone *et al.* [2004], as séries estudadas apresentam o valor do coeficiente de Hurst em torno do valor esperado para retornos independentes. Os cálculos do coeficiente, porém, encontram algumas dificuldades, como mostrado em Couillard & Davison [2005], que dificultam a sua interpretação e validade estatística.

1.4 Volatilidade, Risco e Previsibilidade

Uma das mais utilizadas medidas para o risco em aplicações financeiras é a volatilidade, medida por meio da variância da série em questão. A ideia é bastante intuitiva, uma vez que maiores variações nos preços podem levar a perdas de maior magnitude. Períodos de alta volatilidade, como crises financeiras, apresentariam ao investidor um grande risco. Essa ideia, porém, depende fortemente da suposição de que os retornos são

independentes e normais (gaussianos).

Supondo-se que o mercado seja um sistema complexo, mas determinístico, grandes variações de amplitude não terão nenhuma influência no “risco de investimento”. Como, nesse caso, os valores passados determinam valores futuros, maiores variações não trariam nenhuma aversão ao investidor⁷. Como exemplifica Pincus, se um investidor soubesse que o preço de um ativo seguisse um comportamento sempre senoidal, ainda que com grande amplitude, isso não o afetaria, uma vez que o preço futuro e as estratégias resultantes poderiam ser planejadas (Pincus & Kalman [2004]). Portanto, a previsibilidade, ou a falta da mesma, está bastante relacionada com o risco. Ferramentas que possam inferir e quantificar a previsibilidade podem alterar completamente a visão que os investidores possuem do risco por meio da volatilidade. Em Pincus & Kalman [2004], o autor também mostra a aplicação de uma dessas ferramentas, a entropia aproximada, como forma de se testar a hipótese dos mercados eficientes e de se prever momentos de crise.

A distribuição dos retornos também pode influenciar a percepção de risco. Um exemplo é mostrado na Figura 1.2, que compara uma distribuição gaussiana e a distribuição empírica (baseada em histogramas) dos retornos percentuais obtida do índice *Dow Jones*.

Pela figura, pode-se perceber que retornos acima de 3% ou abaixo de -3% são muito mais prováveis do que o seriam se a distribuição desses fosse gaussiana. Isso significa que a ocorrência de retornos de valor elevado é bastante subestimada quando se supõe que os retornos são gaussianos. Desenvolvendo-se um pouco mais o exemplo, poderia-se considerar, arbitrariamente, que retornos de alta volatilidade (maior risco) possam ser expressos por valores acima de 5% ou abaixo de -5%. Nesse caso, a probabilidade de alta volatilidade dos retornos do Índice *Dow Jones* é aproximadamente 422 vezes maior do que o considerado por uma distribuição gaussiana. Uma consequência disso é a de que a ocorrência de crises pode ser bastante subestimada, simplesmente por se considerar uma distribuição de retornos diferente da real. Voit [2003] desenvolve esse raciocínio e suas implicações para a precificação de opções e o mercado de crédito, fazendo uma analogia ao estudo de turbulências em hidrodinâmica.

Como poderá ser visto em mais detalhes no Capítulo 2, a correlação linear existente em algumas séries de retornos é praticamente nula. Entretanto, o mesmo não se pode dizer a respeito da correlação do quadrado dos retornos, intimamente relacionada com a volatilidade dos ativos. Modelos da família ARCH (*AutoRegressive Conditional Heteroscedasticity*) como o GARCH (*Generalized ARCH*), EGARCH (*Exponential Ge-*

⁷Baseado na hipótese de que o mercado é composto por investidores racionais que sempre procuram maximizar seu lucro, mas são aversos ao risco.

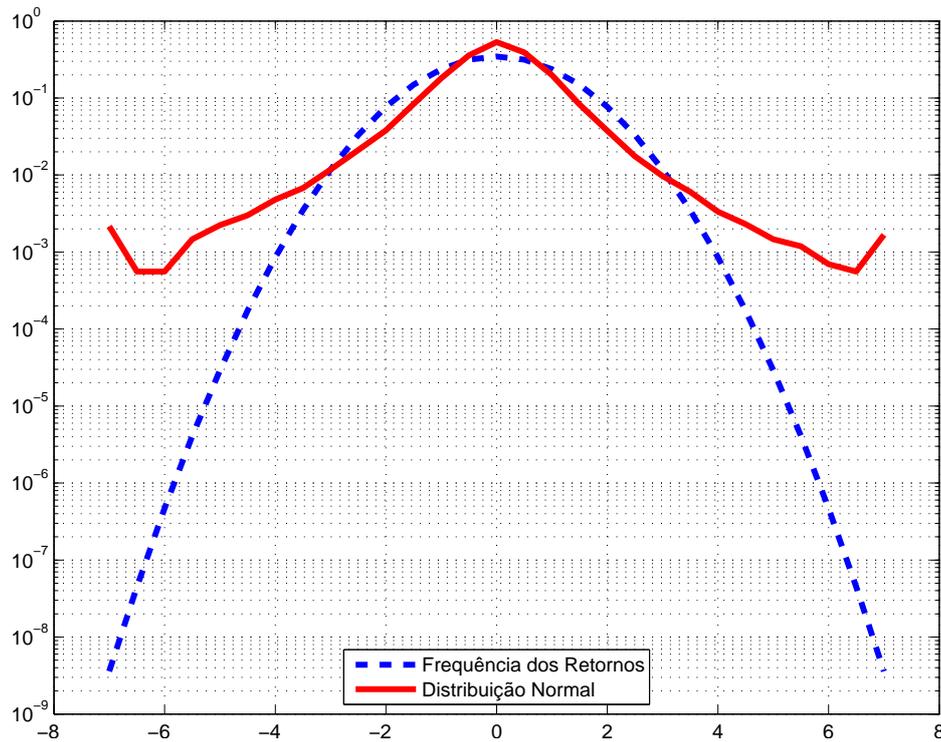


Figura 1.2. Comparação entre a distribuição gaussiana e a distribuição empírica dos retornos do índice *Dow Jones* de 27 de maio de 1896 a 27 de janeiro de 2011.

neralized ARCH), FIGARCH (*Fractionally Integrated ARCH*), entre outros, surgem para modelar essa correlação. Vários desses modelos são descritos em Tsay [2005] e outros livros na área de econometria. A previsão da volatilidade é um campo de estudo bastante útil, por exemplo, na precificação de opções. Exemplos desses estudos podem ser vistos em Burgess & Refenes [1999] e Ko [2009], em que o clássico modelo Black-Scholes (Black & Scholes [1973]) é utilizado em conjunto com a capacidade de mapeamento não linear das redes neurais.

A previsão de volatilidade não é o foco do presente trabalho. A dependência temporal do quadrado dos retornos não indica, por exemplo, se os retornos futuros serão positivos ou negativos, uma vez que traz informação apenas da amplitude e não do sinal dos valores futuros. Neste trabalho, considera-se, por esse motivo, que tal dependência não contribui para uma maior previsibilidade dos retornos⁸. Uma característica indesejável dessa dependência é o fato de as principais ferramentas que indicam previsibilidade serem influenciadas por ela. Consequentemente, a detecção de

⁸Embora a previsão da volatilidade possa representar um grande diferencial em termos de retorno financeiro como, por exemplo, no mercado de opções ou outros derivativos, o foco principal do trabalho é na previsão de preços ou retornos de ativos como, por exemplo, ações, moedas estrangeiras ou *commodities*. Nesses casos a previsão do sinal do retorno é, quando bem sucedida, é claramente, mais relevante.

estruturas não lineares que contribuam para uma melhor previsibilidade podem ficar comprometidas caso essa dependência (heteroscedástica) não seja filtrada ou tenham seu efeito contornado.

1.5 Algumas Considerações

Bachelier estudou processos estocásticos contínuos no tempo ainda em 1900. Além disso, tratou o movimento browniano como um limite de *random walks*, o que evidencia a importância de suas contribuições, resumidas em Balwin *et al.* [2006]. Essa referência apresenta também a tradução de sua tese para o inglês. Um histórico sobre o mercado financeiro, algumas críticas a respeito das tentativas de se prever o mercado bem como algumas discussões sobre análise técnica e fundamentalista, ferramentas amplamente utilizadas na análise de ativos são resumidas em Malkiel [2007].

O volume de artigos e pesquisas que envolvem a previsão ou previsibilidade de séries financeiras atualmente é imensurável. Há cinquenta anos, poder-se-ia fornecer um panorama mais geral do estudo das séries financeiras, como o realizado por Eugene Fama em seu doutorado (Fama [1965]), sob a ótica da eficiência do mercado. Hoje, porém, o estudo de séries financeiras é certamente muito mais abrangente e são muitas as contribuições de diferentes esferas de conhecimento além da economia.

Os estudos de estatística e matemática, indistinguíveis com a pesquisa na economia veem áreas como a identificação de sistemas e a inteligência computacional e até a psicologia como fontes prolíficas de contribuições. Torna-se então impossível abordar a área em um aspecto mais geral sem deixar de lado importantes colaborações.

Por fim, a previsão de séries financeiras deve ser relevante economicamente. Não basta que uma solução possua, estatisticamente, um erro mínimo para que seja viável. Um modelo de previsão que tenha relevância econômica deve superar todos os gastos típicos de operações no mercado, como o *bid-ask spread*⁹, taxas de corretagem, taxas de custódia, impostos e outros. Essa característica é bastante significativa para estratégias que envolvam um grande número de operações de compra e venda.

Além disso, essas estratégias devem também se mostrar mais lucrativas do que o simples *buy and hold*, técnica que minimiza os gastos anteriores. Supondo-se que a série de preços possa ser aproximada por um processo *submartingale*, ou seja, $E[p(t+1)] \geq E[p(t)]$, uma escolha natural de estratégia de investimento é a *buy and hold*. Essa suposição é bastante plausível, considerando-se que muitas séries históricas de longo

⁹*Bid-ask spread* é a diferença entre o preço que se deseja comprar ou vender e o que o mercado realmente oferece. Essa diferença tem um impacto negativo em qualquer estratégia de investimento, principalmente nas que ativamente realizam operações.

prazo como os principais índices de ações (*Dow Jones Industrial Average*, *Standard & Poor's 500*, Ibovespa, entre outros) possuem essa característica. *Buy and hold* significa comprar o quanto antes um ativo e permanecer com ele por um longo período de tempo. Isso seria equivalente a um investimento de longo prazo, pois $p(t+1)$ não seria necessariamente maior que $p(t)$, contudo a probabilidade $P(p(t+n) \geq p(t))$ aumenta com n .

1.6 Objetivos do Trabalho

Percebe-se que a previsibilidade é uma análise indispensável no estudo de séries financeiras e deve ser abordada para que se justifique modelagens para previsão. Não é objetivo do trabalho obter um modelo matemático extremamente lucrativo que, como no jargão de investidores e especuladores, consiga “ganhar do mercado”. Antes deve-se investigar se isso é possível, ou pelo menos, estatisticamente provável. A existência de dependências temporais em uma série financeira deve ser investigada antes de tentativas de previsão, sob o risco de se tentar *prever o imprevisível*. Talvez o número de artigos a respeito de previsibilidade, quando comparado com o de tentativas de previsão não evidencie essa importância. Entretanto, não há, atualmente, um modelo único que consiga capturar de maneira eficaz (para previsão) características como a dependência temporal dos retornos, a variação da volatilidade e uma distribuição de probabilidade adequada. Quando se soma a essas características as mudanças constantes a que os mercados estão submetidos (não estacionariedade), é factível que tal modelo não exista. Os muitos resultados controversos obtidos na área de previsão, de uma certa maneira, vão ao encontro dessa hipótese.

Nesse contexto, o presente trabalho busca apresentar algumas ferramentas que auxiliem na resposta de algumas perguntas fundamentais nessa análise:

- as séries de preços e seus retornos são completamente imprevisíveis, como um passeio aleatório?
- há algum determinismo ou alguma estrutura que possa ser identificada?
- pode-se classificar as estruturas existentes (se existentes) em classes conhecidas como dependências lineares ou não lineares (e de qual tipo)?
- as estruturas existentes (se existentes) contribuem para uma melhor previsibilidade da série ($E[p(t+1)|p(t), p(t-1), \cdot] \neq p(t)$)?

Este trabalho apresenta, inicialmente, uma comparação entre algumas séries financeiras e o modelo do passeio aleatório, com o intuito de motivar um estudo mais profundo a respeito do determinismo presente nessas séries. Mostra-se, neste trabalho, que os dados estudados não condizem com um passeio aleatório e apresentam alguma estrutura aparentemente oculta em uma simples inspeção visual dos mesmos. As ferramentas primeiramente utilizadas para mostrar a não adequação desses dados a um passeio aleatório fazem parte de um conjunto clássico de testes estatísticos.

O restante do trabalho apresenta técnicas que contornam algumas das limitações desses testes clássicos e fornecem resultados mais gerais. Esses resultados possibilitam identificar quais os tipos de dependências estão presentes nos dados. Para tanto, são utilizados os testes BDS de resíduos de alguns modelos das séries e a análise de dados sub-rogados. O procedimento para a escolha dos modelos e algoritmos desses testes é feito com o objetivo de detectar se há dependências lineares ou não lineares. No caso da existência de dependências não lineares, são realizados testes visando identificar se as mesmas contribuem para uma melhor previsibilidade dos retornos utilizando técnicas robustas à presença de heteroscedasticidade (algoritmo PPS) e filtros GARCH (na análise de resíduos BDS). Portanto, as análises realizadas neste trabalho possibilitam investigar de maneira mais geral a rejeição da hipótese de que séries financeiras são como passeios aleatórios.

Uma espécie de validação desses procedimentos de previsibilidade é feita por meio da tentativa de previsão das séries, usando redes RBF. Afinal, séries que possuem estruturas determinísticas previsíveis poderiam, em princípio, ser modeladas e estimadas. Os erros de previsão são, então, comparados com os resultados da análise de previsibilidade, a fim de verificar se séries mais previsíveis (segundo as técnicas estudadas) fornecem melhores resultados de previsão.

1.7 Organização do Trabalho

Este capítulo apresentou uma introdução a respeito da previsibilidade e previsão de séries financeiras, incluindo um breve histórico e uma revisão bibliográfica das técnicas existentes atualmente. O Capítulo 2 mostra uma comparação entre algumas séries financeiras e o modelo do passeio aleatório, amplamente utilizado para se justificar a independência dos retornos. Esse capítulo tem o intuito de evidenciar algumas limitações desse modelo e justificar a utilização de outras ferramentas mais avançadas. Essas ferramentas são apresentadas no Capítulo 3. Isso é feito por meio de técnicas não lineares e testes estatísticos que constituem uma base para se quantificar a previsibilidade

das séries financeiras. A aplicação dessas técnicas em séries financeiras é apresentada no Capítulo 4 e tem como resultado a geração de índices de previsibilidade, que tentam quantificar a capacidade de previsão que uma série pode fornecer. Como forma de validação das ferramentas utilizadas, o Capítulo 5 usa um modelo de previsão para essas séries. A ideia desse capítulo é a de comparar os resultados obtidos pelas previsões com os índices de previsibilidade obtidos nos capítulos anteriores. Séries cuja previsibilidade seja maior devem apresentar um menor erro de previsão, quando comparadas a outras de menor previsibilidade.

Séries Financeiras e o Passeio Aleatório

O presente capítulo apresenta alguns conceitos frequentemente utilizados em análise de séries temporais financeiras e realiza uma comparação entre séries financeiras e o modelo conhecido como passeio aleatório. Além disso, são apresentadas algumas séries geradas por meio de simulações, que são utilizadas para melhor visualizar os métodos apresentados. A comparação entre o passeio aleatório e essas séries, bem como as séries estudadas (apresentadas somente no Capítulo 4, Seção 4.1) mostra as limitações desse modelo e serve de base e motivação para o estudo realizado no restante deste trabalho.

É comum relacionar séries financeiras como preços de ações e índices de mercado a um processo estocástico chamado passeio aleatório. Atualmente, pode-se demonstrar (e isso será feito no presente capítulo) que o modelo do passeio aleatório não explica algumas características comumente encontradas nas séries em questão. Embora o passeio aleatório se mostre inadequado em alguns aspectos, esse modelo serve de base para a construção de uma análise mais profunda a respeito da previsibilidade das séries. A Figura 2.1 mostra a série de preços de uma determinada ação junto a três séries artificiais completamente aleatórias e geradas por meio de simulação. Não se pode, a princípio, identificar de maneira fácil qual é a série de preços de ações e quais são as séries geradas artificialmente.

O passeio aleatório, ou *Random Walk*, é um processo estocástico de tempo discreto que pode ser descrito pela seguinte equação:

$$p(t) = \sum_{i=1}^t \epsilon(i) \quad \epsilon(i) \sim N(0, \sigma^2) \quad (2.1)$$

em que $\epsilon(i)$ é gaussiano com média zero e variância σ^2 constante para todo i (passeio aleatório com incrementos gaussianos). Modelar uma série de preços de uma determinada ação ou outro ativo financeiro por um passeio aleatório traz várias consequências em termos de sua interpretação econômica. Pode-se perceber isso com mais clareza

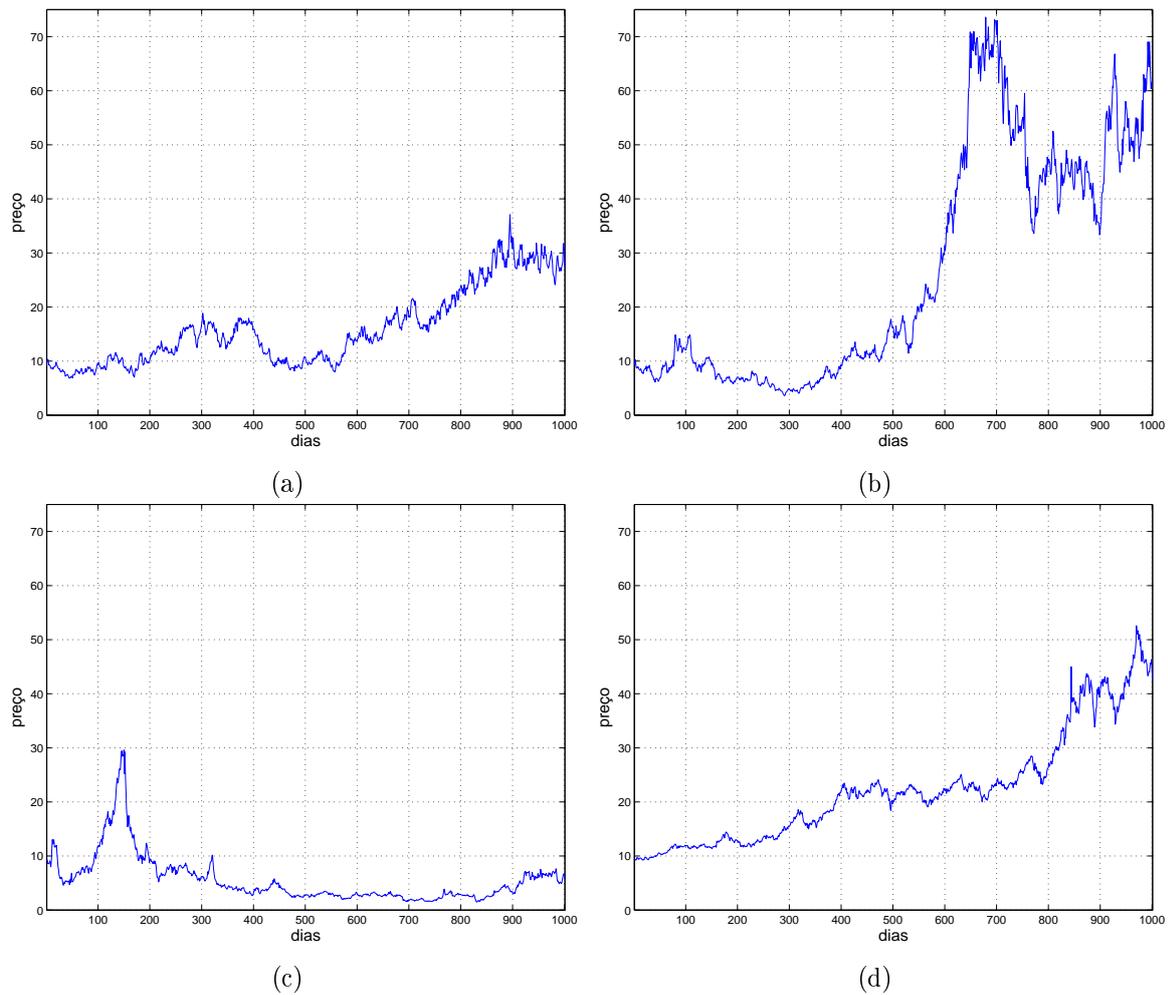


Figura 2.1. Um dos gráficos mostra a série de preços de uma ação e os outros mostram execuções de um séries artificiais como o passeio aleatório com valor inicial igual e variância das inovações semelhante à da série financeira. Não se consegue distinguir, de maneira clara e sem uma análise mais profunda, a série financeira das séries geradas artificialmente.

escrevendo-se a equação acima em uma forma iterativa, como:

$$p(t) = p(t - 1) + \epsilon(t). \quad (2.2)$$

$p(t)$ representa o logaritmo do preço de algum ativo no instante t . Como $\epsilon(t)$ tem esperança nula e dado que se conhece $p(t - 1)$ (valor da série $p(t)$ no instante anterior), tem-se que:

$$E[p(t)] = E[p(t - 1)] + E[\epsilon(t)], \quad (2.3)$$

$$E[p(t)] = E[p(t-1)] = p(t-1). \quad (2.4)$$

Ou seja, segundo o passeio aleatório, a melhor previsão que se pode alcançar a respeito do preço atual é o preço anterior.

Definindo-se o retorno financeiro em t sobre um investimento realizado em $t-1$ como a diferença dos logaritmos dos preços em t e $t-1$, tem-se:

$$r(t) = \log p(t) - \log p(t-1), \quad (2.5)$$

$$E[r(t)] = E[\log p(t) - \log p(t-1)] = 0. \quad (2.6)$$

O modelo do passeio aleatório de preços de ativos financeiros tem retorno esperado $E[r(t)]$ nulo. O logaritmo é utilizado para que a composição dos preços por meio de vários retornos seja simplificada (uma soma ao invés de uma multiplicação) e também para que não haja a possibilidade de haver preços negativos.

Uma pequena modificação geralmente feita no passeio aleatório é a inclusão de um termo de tendência (Campbell *et al.* [1997]). Com isso a Equação 2.2 torna-se:

$$p(t) = \mu + p(t-1) + \epsilon(t), \quad (2.7)$$

em que μ é uma constante. Esse é o passeio aleatório com tendência (*Random Walk with Drift*). O termo adicional generaliza o modelo para séries de retornos cujas médias sejam diferentes de zero. Pode-se então retornar ao modelo sem termo de tendência μ usando-se a média amostral (o passeio aleatório é ergódico quando se toma sua primeira diferença). Esse processo é mostrado na Figura 2.2. Subtrai-se essa média da série em questão chegando-se assim a um modelo sem tendência.

Uma consequência também bastante relevante é o fato da variância do incremento ser constante. O modelo não considera a existência dos chamados *clusters* de volatilidade, que são períodos em que se tem diferentes variâncias agrupadas ao longo da série. A volatilidade está ligada à amplitude dos movimentos dos preços e pode ser representada pela variância dos retornos. Em uma série temporal $y(t)$, esse efeito pode ser representado pela presença de heteroscedasticidade, que nada mais é que considerar variâncias diferentes ($\sigma^2(j) \neq \sigma^2(k)$ para algum $j \neq k$) e não constantes ao longo do tempo, como supõe o modelo. Ao longo do texto a variância da observação $y(t)$ será identificada na notação de função, como em $\sigma^2(t)$ evidenciando a dependência da mesma em relação a t , sempre que existir heteroscedasticidade.

Outro aspecto a ser citado é a consideração de que os incrementos são gaussianos.

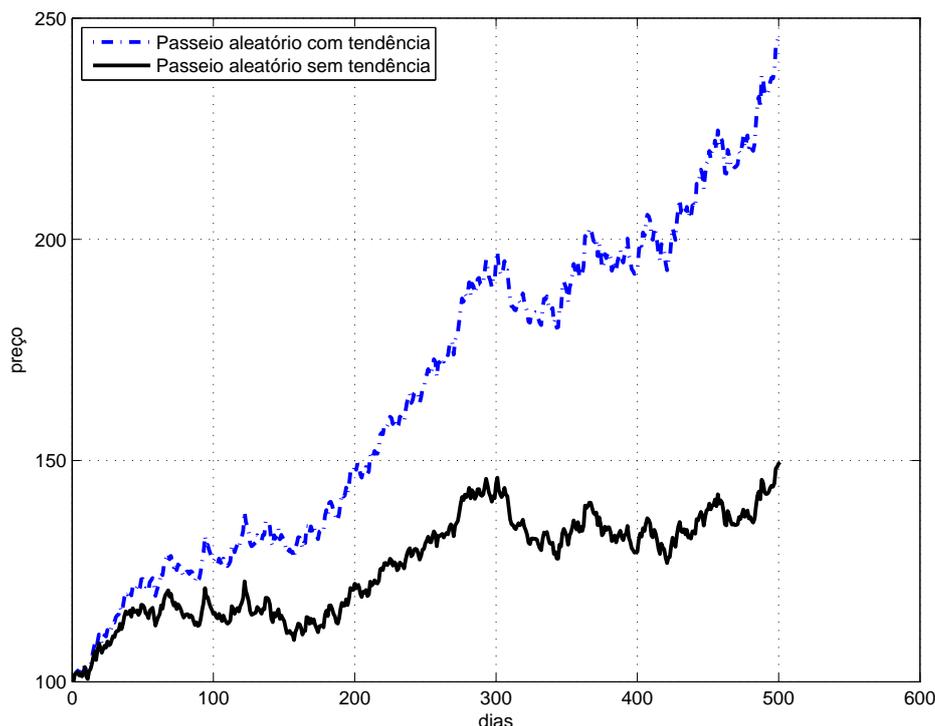


Figura 2.2. Comparação de um passeio aleatório com tendência e o mesmo sem sua tendência.

Pode-se, por meio de testes estatísticos, como o Kolmogorov-Smirnov, rejeitar ou não a normalidade dos retornos, o que afetaria a adequação do modelo a um passeio aleatório. Nesse caso, a variância não será mais uma estimativa confiável para o risco.

O fato de $r(t)$ ser independente e gaussiano e não haver heteroscedasticidade faz com que os incrementos $\epsilon(t)$ da série de retornos de um passeio aleatório sejam independentes e identicamente distribuídos (i.i.d.). Essa simplificação permite que a série seja considerada estacionária, o que simplifica muito a análise. Isso, porém, não necessariamente corresponde à realidade das séries financeiras.

As implicações principais do modelo do passeio aleatório podem ser resumidas como:

- os retornos são independentes e sua esperança é nula;
- não há heteroscedasticidade, ou variação da volatilidade;
- os retornos seguem uma distribuição gaussiana.

O restante do capítulo apresenta algumas séries utilizadas no restante do trabalho e compara o passeio aleatório e as séries financeiras estudadas, utilizando-se as técnicas clássicas para esse fim. Para tanto, são introduzidos testes frequentemente aplicados a

séries financeiras: o teste Ljung-Box, que possibilita verificar dependências temporais nos dados, e dois testes de heteroscedasticidade. A presença dessa última característica traz consequências tanto para a modelagem quanto para a validade dos testes aplicados aos dados. As séries estudadas serão descritas apenas na Seção 4.1, onde serão aplicadas outras técnicas que contornam as limitações apresentadas no presente capítulo. A descrição posterior dessas séries, porém, não interfere no objetivo principal do capítulo que é introduzir a importante ligação entre o passeio aleatório e a previsibilidade das séries financeiras.

2.1 As Séries Artificiais Utilizadas

No decorrer deste capítulo e do Capítulo 3, serão utilizadas algumas séries geradas por meio de simulações. Não são séries financeiras, mas auxiliam o entendimento de todas as técnicas apresentadas neste trabalho. São elas o ruído branco gaussiano (WGN), AR(1), ARCH(1), GARCH(1,1) e o mapa logístico (LOGMAP). O principal objetivo é obter uma base de comparação para os testes utilizados. A análise prévia dos testes em dados com características já conhecidas é fundamental quando se trabalha com séries financeiras. Esse passo anterior revela as limitações e potencialidades de cada método em uso além de se mostrar mais didático. Para todos os processos simulados são geradas três mil e quinhentas observações ($N = 3500$). A quantidade escolhida é compatível com a quantidade de dados geralmente disponível em séries financeiras.

A primeira série utilizada é o ruído branco gaussiano, ou WGN (*White Gaussian Noise*). Origina-se de um processo dado por:

$$y(t) = \epsilon(t), \quad \epsilon(t) \sim N(0, \sigma^2). \quad (2.8)$$

As variáveis $y(t)$ são independentes e identicamente distribuídas (i.i.d.). São, portanto, normais, descorrelacionadas e homoscedásticas (variância constante).

A série AR(1) (*AutoRegressive*) é simulada a partir do seguinte processo autor-regressivo:

$$y(t) = \theta_1 y(t-1) + \epsilon(t), \quad \epsilon(t) \sim N(0, \sigma^2). \quad (2.9)$$

O parâmetro θ_1 é constante e deve estar no intervalo $-1 < \theta_1 < 1$ para que o processo seja estacionário (no sentido amplo). Os incrementos, ou inovações, $\epsilon(t)$ têm as mesmas características do caso WGN; são i.i.d., gaussianos e têm esperança nula. O parâmetro de um processo AR(1) tem um papel importante na previsibilidade, uma vez que $\epsilon(t)$ tem esperança nula. Quanto maior seu valor em módulo, menor será a contribuição

da inovação $\epsilon(t)$ para a previsão de $y(t)$ baseado em $y(t-1)$. Consequentemente, a previsibilidade da série é proporcional ao parâmetro θ_1 . No presente trabalho, a série é gerada com parâmetro igual a 0,3 e variância σ^2 unitária. Em casos específicos em que o parâmetro possuir outro valor, o mesmo será explicitado.

O processo ARCH(1) (*AutoRegressive Conditional Heteroscedasticity*) é semelhante ao WGN. Entretanto, a variância das inovações possui uma dependência linear com o quadrado de valores atrasados da série. Esse processo é dado por:

$$\begin{aligned} y(t) &= \epsilon(t) & \epsilon(t) &\sim N(0, \sigma^2(t)) \\ \sigma^2(t) &= \alpha_0 + \alpha_1 y(t-1)^2. \end{aligned} \quad (2.10)$$

α_0 é uma constante que corresponde ao valor inicial da variância. O processo também tem esperança nula, mas não é mais i.i.d. como o WGN. A série ARCH(1) é gerada por meio de uma simulação com o parâmetro $\alpha_1 = 0,3$ e variância inicial α_0 unitária. Como a variância não é constante, a série é chamada heteroscedástica.

O processo GARCH(1,1) (*Generalized AutoRegressive Conditional Heteroscedasticity*), introduzido em Bollerslev [1986], inclui na equação de sua variância valores atrasados da própria variância. Pode ser visto como um processo ARMA (*AutoRegressive Moving-Average*) de $y(t)^2$. É dado por:

$$\begin{aligned} y(t) &= \epsilon(t) & \epsilon(t) &\sim N(0, \sigma^2(t)) \\ \sigma^2(t) &= \alpha_0 + \alpha_1 y(t-1)^2 + \alpha_2 \sigma^2(t-1). \end{aligned} \quad (2.11)$$

Tanto em ARCH(1) quanto em GARCH(1,1) há uma dependência temporal, mas isso não faz com que $y(t)$ desses processos seja mais previsível do que $y(t)$ de WGN. A esperança condicional desses processos é igual a esperança incondicional e tem valor nulo. A previsibilidade está presente apenas no processo da variância¹. O que torna ARCH e GARCH bastante relevantes neste trabalho é o fato de essa dependência poder ser interpretada erroneamente como alguma estrutura dinâmica que possibilite uma melhor estimativa para retornos futuros. As ferramentas utilizadas para se detectar determinismo e não linearidades devem ser capazes de diferenciar a presença de heteroscedasticidade da presença de não linearidades que influenciem $E[y(t)|y(t-1), y(t-2), \dots]$.

O mapa logístico (LOGMAP) é um mapa discreto que, embora bastante simples, apresenta comportamento caótico para determinados valores de seu parâmetro λ . Sua

¹Na literatura sobre econometria, encontra-se frequentemente a denominação de processo da média e processo da variância. O primeiro remete ao modelo de evolução dos retornos. O segundo, representa um modelo para a evolução da variância ao longo do tempo. Como a heteroscedasticidade é bastante comum em séries financeiras, essas denominações são muitas vezes necessárias para se distinguir o processo estudado.

equação é dada por:

$$y(t) = \lambda y(t-1) (1 - y(t-1)). \quad (2.12)$$

Para $\lambda \approx 4$ ou pouco menor que esse valor, $y(t)$ apresenta baixa correlação linear ou correlação linear insignificante (considerando intervalo de confiança de dois desvios padrão), semelhante a um ruído branco. A série LOGMAP utilizada neste trabalho é gerada por meio da equação anterior com $\lambda = 4$, y inicial $y(0) = 0,1$ e possui três mil e quinhentas observações².

A Figura 2.3 mostra as cinquenta primeiras observações das séries apresentadas anteriormente. Pela figura, as séries parecem ser semelhantes, mas apresentam características estatísticas bastante distintas, como será melhor visto no Capítulo 3.

2.2 O Teste Ljung-Box

O teste Ljung-Box, descrito em Ljung & Box [1978] é um teste estatístico que possibilita mensurar, por meio de um único valor, a autocorrelação de uma série em vários atrasos ao mesmo tempo. Esse valor é comparado a um valor limite, acima do qual se poderia rejeitar a hipótese de que determinada série se comporta como um ruído branco gaussiano. A função de autocorrelação (FAC) mostra se há correlações em um atraso específico da série. O teste considera a FAC em um conjunto de atrasos. Geralmente, o teste é usado para se verificar os resíduos de um modelo linear. Pode-se, todavia, usá-lo também, para testar a aleatoriedade.

O teste utiliza uma estatística $q(m)$, uma soma ponderada dos quadrados dos coeficientes da FAC que, até o atraso m , é dada por:

$$q(m) = N(N+2) \sum_{l=1}^m \frac{\hat{\rho}_l^2}{N-l}, \quad (2.13)$$

em que m é o número de atrasos usados, N é o número de observações e $\hat{\rho}_l$ é a FAC, sendo l seu atraso.

Supondo um processo i.i.d. gaussiano, o valor esperado da FAC é nulo em todos os atrasos diferentes de zero. Essa é a hipótese nula de que não há correlação linear na série. No caso de um número limitado de observações, $q(m)$ segue uma distribuição

²Para $\lambda = 4$, o mapa logístico apresenta uma crise de fronteira (destruição da bacia de atração). Nesse caso, eventualmente, a série gerada pode divergir, mas a probabilidade de isso acontecer é pequena. Na série gerada, com 3500 observações e condição inicial $y(0) = 0,1$, essa divergência não aconteceu. Isso, portanto, não interfere no uso da mesma como um exemplo de série não linear e sem dependência linear significativa.

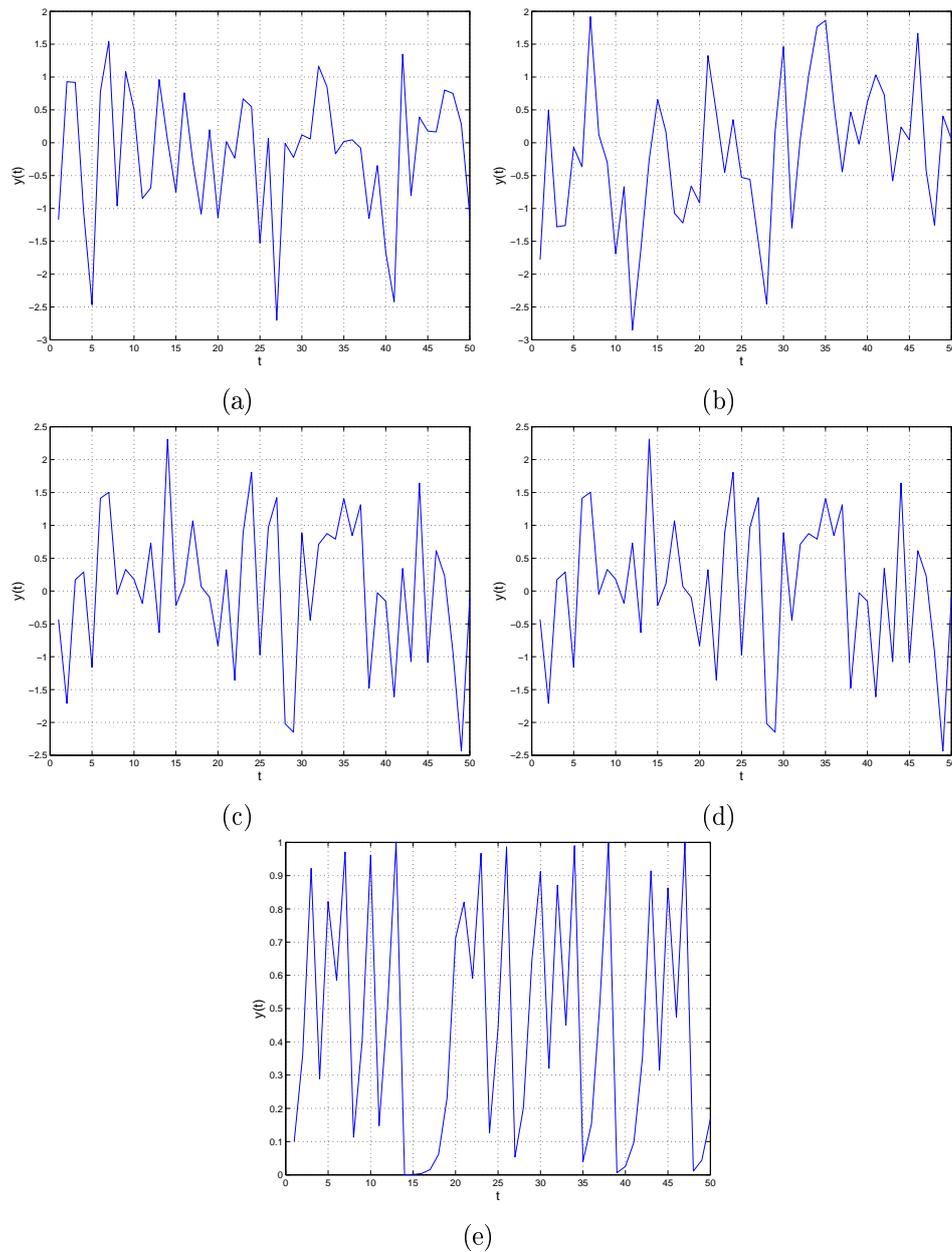


Figura 2.3. As figuras mostram as cinquenta primeiras observações das séries: (a) WGN, (b) AR(1), (c) ARCH(1), (d) GARCH(1,1) e (e) LOGMAP. Embora semelhantes, as séries apresentam características bastante diferentes.

qui-quadrado com m graus de liberdade ($\chi^2(m)$). É de se esperar que isso ocorra, uma vez que provém da soma do quadrado de variáveis gaussianas. Para um determinado número m de atrasos, define-se um percentil que seria um limite superior para o processo, acima do qual a hipótese nula seria rejeitada. Ou seja, para o caso de um intervalo de confiança de dois desvios padrão, o percentil 95 de uma distribuição $\chi^2(m)$ seria o limite superior acima do qual o processo não seria mais considerado i.i.d.

gaussiano. Nesse caso, 95% das séries i.i.d. gaussianas gerariam um $q(m)$ menor que esse número. Em contrapartida, se $q(m)$ da série for maior que o valor desse percentil, rejeita-se a hipótese nula de que não há correlação linear.

A Figura 2.4 mostra a distribuição de frequências normalizada de $q(10)$ de uma população de 10000 séries geradas a partir de uma variável gaussiana de variância unitária em um processo i.i.d..

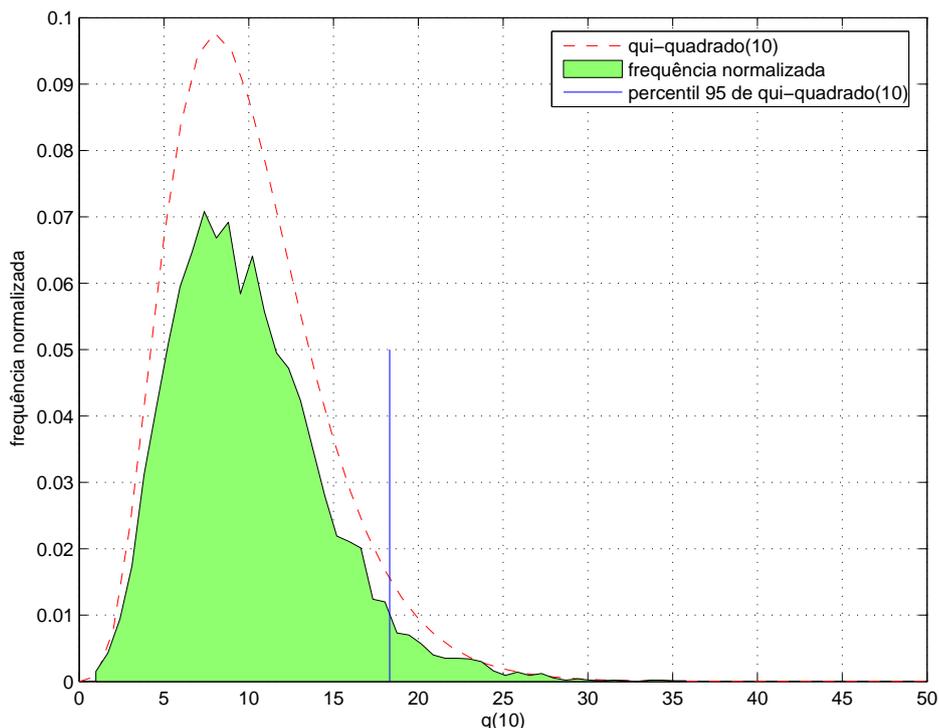
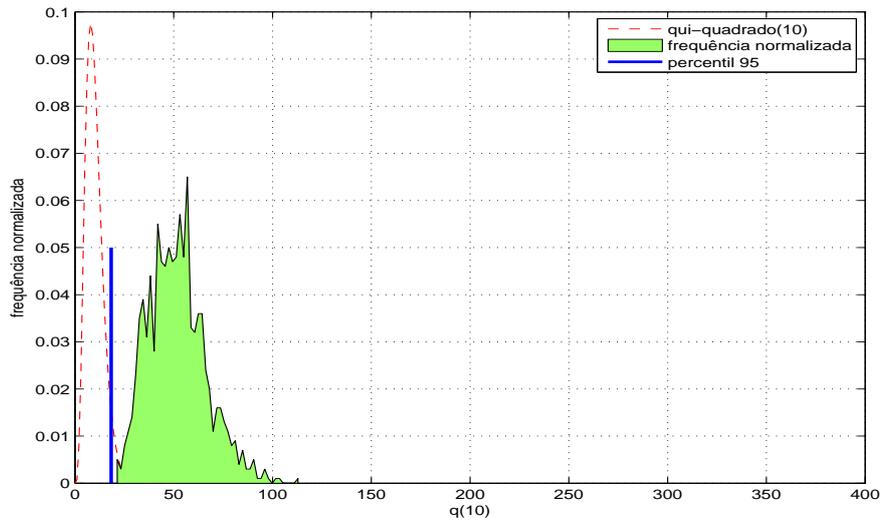


Figura 2.4. Distribuição de frequências normalizada do valor de $q(10)$ e distribuição qui-quadrado com 10 graus de liberdade de um processo i.i.d. gaussiano.

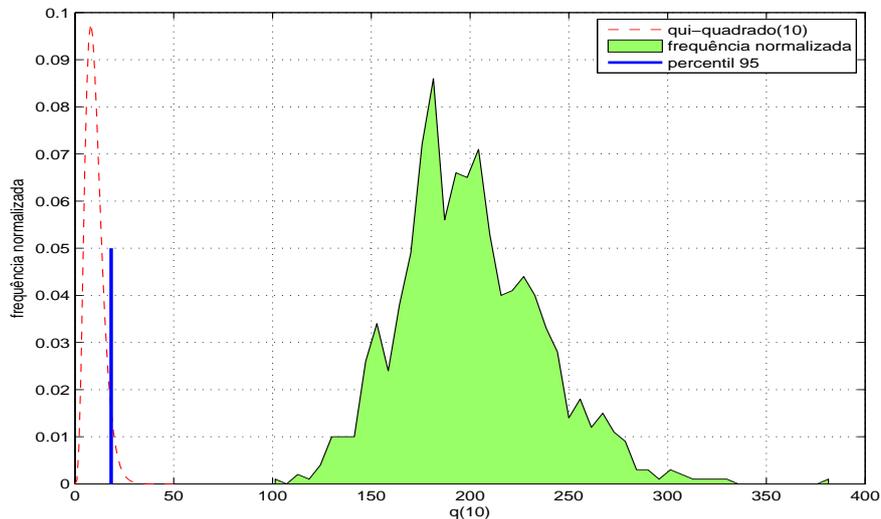
A estatística $q(m)$ segue assintoticamente uma distribuição qui-quadrado (Tsay [2005]) de m graus de liberdade. Pode-se perceber que a frequência normalizada de $q(10)$ se aproxima de $\chi^2(10)$. A figura também mostra o percentil 95 da distribuição. Esse é o valor limite (ou valor crítico) em que se define se as correlações são significativas ou não, a partir do qual $q(m)$ teria uma probabilidade de 0,95 de ter sido gerada por uma série correlacionada. Para 10 atrasos, esse valor é de 18,3. Portanto, para uma série que possua $q(10)$ maior que 18,3, é possível rejeitar a hipótese nula de que não há correlações lineares com 95% de confiança.

A Figura 2.5 mostra a distribuição de frequências de $q(10)$ para duas populações de 10000 processos AR(1) com inovações gaussianas de variância unitária.

A primeira população, Figura 2.5.(a), é formada por processos AR(1) com o parâmetro igual a 0,2. A segunda população, Figura 2.5.(b), é formada também por



(a) AR(1) com parâmetro 0,2.



(b) AR(1) com parâmetro 0,4.

Figura 2.5. $q(10)$ para processos AR(1), (a) com parâmetro 0,2 e (b) com parâmetro 0,4.

processos AR(1), porém com parâmetro igual a 0,4. O parâmetro pondera a correlação do processo e pode ser visto pela figura em que os processos de maior $q(10)$, no caso os de parâmetro 0,4, indicam uma maior correlação. Ambos os processos AR(1), porém, estão claramente acima do $q(10)$ limite (percentil 95 do $\chi^2(10)$, igual a 18,3), o que indica correlações significativas e, assim, presença de determinismo, segundo o teste. Esse fato também pode ser percebido na Tabela 2.1, que mostra também o ruído branco gaussiano (WGN) com estatística $q(10)$ inferior ao valor limite. Nesse caso, não se pode rejeitar a hipótese de que os dados sejam descorrelacionados, o que de fato é verdade.

Para o caso em que o processo não é gaussiano, os valores críticos para q (como

$q(10) = 18,3$), perdem sua utilidade pois não se pode mais aproximar $q(m)$ limite por qui-quadrado. Esse fato limita mas não invalida o teste. Isso porque $q(m)$ continua a quantificar o determinismo e pode-se utilizar alternativamente métodos não paramétricos como o de dados sub-rogados para se obter os valores limites da estatística.

Uma aplicação comum para o teste é a verificação de heteroscedasticidade. Uma série que apresente $q(m)$ abaixo do limite (sem determinismo), mas possua $q(m)$ acima do limite para o quadrado de suas observações, pode indicar a presença de heteroscedasticidade. A Tabela 2.1 mostra os valores de $q(m = 10)$ obtidos para observações das séries WGN, AR(1), ARCH(1), GARCH(1,2), LOGMAP e seus quadrados: WGN², AR(1)², ARCH(1)², GARCH(1,2)² e LOGMAP². A heteroscedasticidade de ARCH(1) e GARCH(1,1) pode ser detectada percebendo-se que o valor do quadrado das mesmas supera o valor limite de $q(10)$. Para o AR(1)², o valor de $q(10)$ não deve ser imediatamente considerado como heteroscedasticidade pois há correlação significativa de AR(1), como pode ser observado na primeira linha da mesma tabela. O recomendado nesses casos é, primeiramente, obter um modelo para a série e então aplicar o teste aos resíduos desse modelo.

Tabela 2.1. Valores obtidos da estatística $q(10)$ para várias séries conhecidas e seus quadrados. O valor limite de $q(10)$, acima do qual o teste indica correlações significativas, também é mostrado.

Série	$q(10)$	Valor limite de $q(10)$
WGN	12,7	18,3
WGN ²	6,01	18,3
AR(1)	373	18,3
AR(1) ²	24,5	18,3
ARCH(1)	13,7	18,3
ARCH(1) ²	336	18,3
GARCH(1,1)	14,0	18,3
GARCH(1,1) ²	411	18,3
LOGMAP	12,0	18,3
LOGMAP ²	247	18,3

Uma deficiência de $q(m)$ é medir apenas as correlações lineares. Um exemplo dessa limitação é caso do mapa logístico (LOGMAP), um sistema completamente determinístico, mas que não apresenta correlações lineares significativas para valores de seu parâmetro aproximadamente iguais a 4. A Tabela 2.1 também mostra o resultado do teste para esse mapa e seu quadrado. Nesse último caso não se pode falar em heteroscedasticidade pois a série é completamente determinística.

Essas limitações do teste não fazem com que seu resultado seja errado. Deve-se ressaltar que um valor de $q(m)$ acima do $q(m)$ limite rejeita a hipótese de que a série é um ruído branco gaussiano. Para valores de $q(m)$ abaixo do limite, não se pode rejeitar tal hipótese, o que é diferente de se afirmar ser a série um ruído branco. Esse é caso dos processos ARCH e GARCH e do mapa logístico, que no teste possuem valor $q(m)$ inferior ao $q(m)$ limite.

2.3 Correlação e Esperança dos Retornos

Uma primeira abordagem para a verificação da independência dos retornos é verificar sua autocorrelação linear. O fato de uma série apresentar correlações lineares nulas não indica independência. Todavia, a presença de correlações significativas nega a hipótese de independência³. Além disso, testes baseados em correlação são muito mais simples e diretos que testes de independência e, se a série for um processo gaussiano, a condição de decorrelação implica independência.

Campbell e colegas mostram vários testes da não adequação do passeio aleatório a dados financeiros (Campbell *et al.* [1997]). Alguns testes, como o da comparação entre a probabilidade de sequências (retornos consecutivos com mesmo sinal) e reversões (retornos consecutivos com sinais diferentes), têm apenas interesse histórico. Dentre os que são utilizados atualmente, pode-se citar o teste da razão das variâncias. Esse teste lança mão do fato de que, em um passeio aleatório, a variância da série de preços cresce linearmente. Nesse caso, a variância da soma dos retornos $\sum_{t=1}^N r(t)$ é igual a $r(t) \cdot N$ e a razão dessa variância por $\text{var}(r(t))$ é sempre $\frac{1}{N}$. Como mostrado por Campbell, esse teste pode ser deduzido por meio da função de autocorrelação. Outro teste descrito pelo mesmo autor é o teste Ljung-Box mostrado na Seção 2.2, que também é baseado em funções de autocorrelação. Para esta seção, esse último é o escolhido por também ser usado para se testar heteroscedasticidade. O objetivo desse teste é mostrar que a esperança dos retornos, condicionada ao conhecimento dos retornos anteriores, é igual à esperança incondicional dos retornos. Ou, supondo que os retornos são gaussianos (o teste é apenas para correlações lineares),

$$E[r(t)|r(t-1), r(t-2), r(t-3), \dots] = E[r(t)]. \quad (2.14)$$

As Figuras 2.6 e 2.7 mostram as funções de autocorrelação das séries de retorno.

³Vale ressaltar que a dependência temporal não institui uma relação de causalidade entre os retornos.

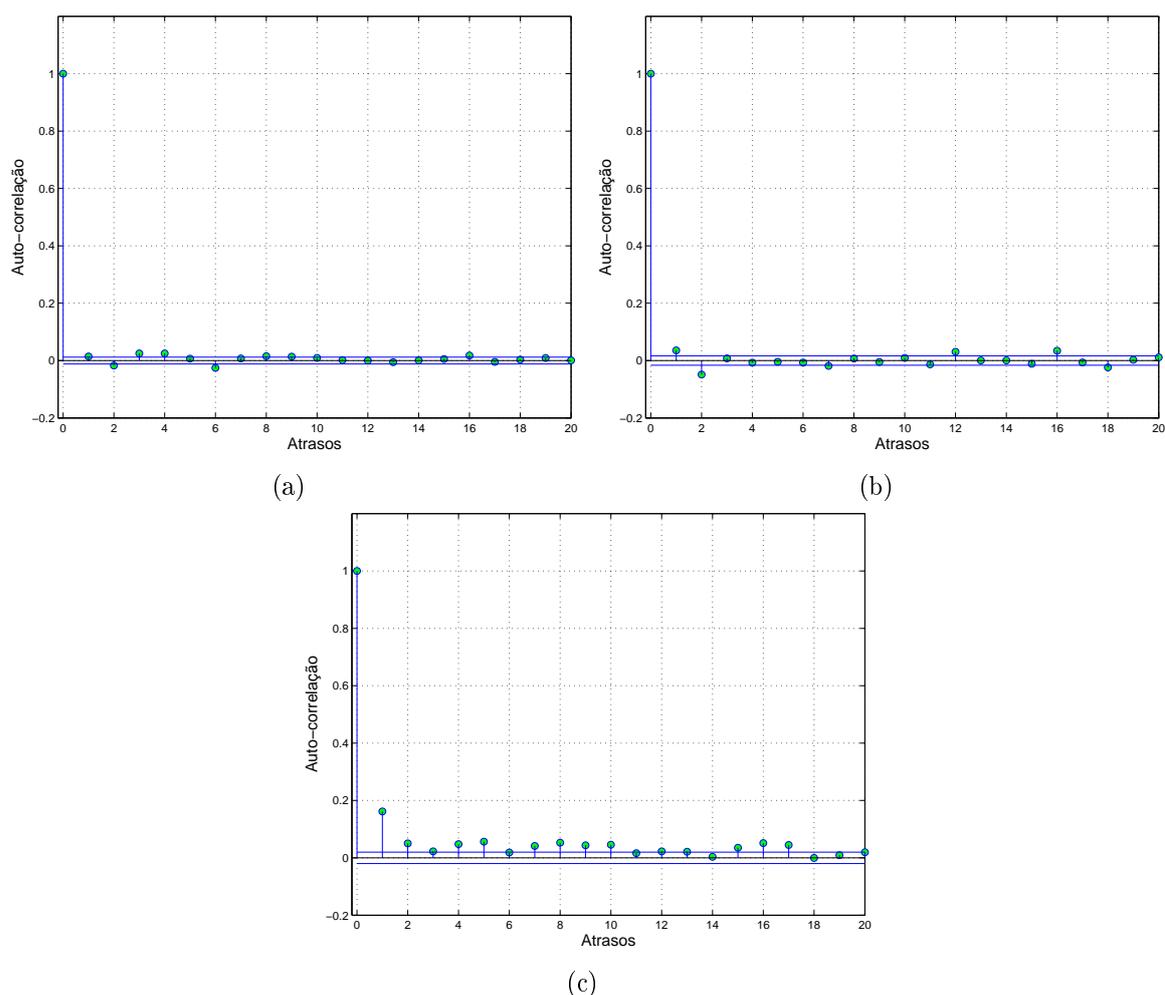


Figura 2.6. Função de autocorrelação dos retornos: (a) DJIA, (b) SP500, (c) IBOV. As linhas contínuas horizontais mostram o intervalo de confiança considerando-se dois desvios padrão.

Pelas figuras, os pontos da função acima do intervalo de confiança indicam que há correlação entre os valores presentes e passados dos retornos nos respectivos atrasos. Os resultados do teste Ljung-Box podem ser vistos na Tabela 2.2. Para todas as séries, o valor encontrado de $q(10)$ é superior ao valor limite, indicando que pode-se rejeitar a hipótese de que as séries se comportariam como um ruído branco gaussiano.

Pode-se perceber que, para as séries analisadas, a presença de correlações rejeita a hipótese do passeio aleatório com inovações gaussianas. Ou seja, a informação dos retornos passados permite uma melhor inferência da esperança de seus valores futuros. Isso não significa que essa inferência seja economicamente relevante, entretanto justifica um estudo mais aprofundado do assunto.

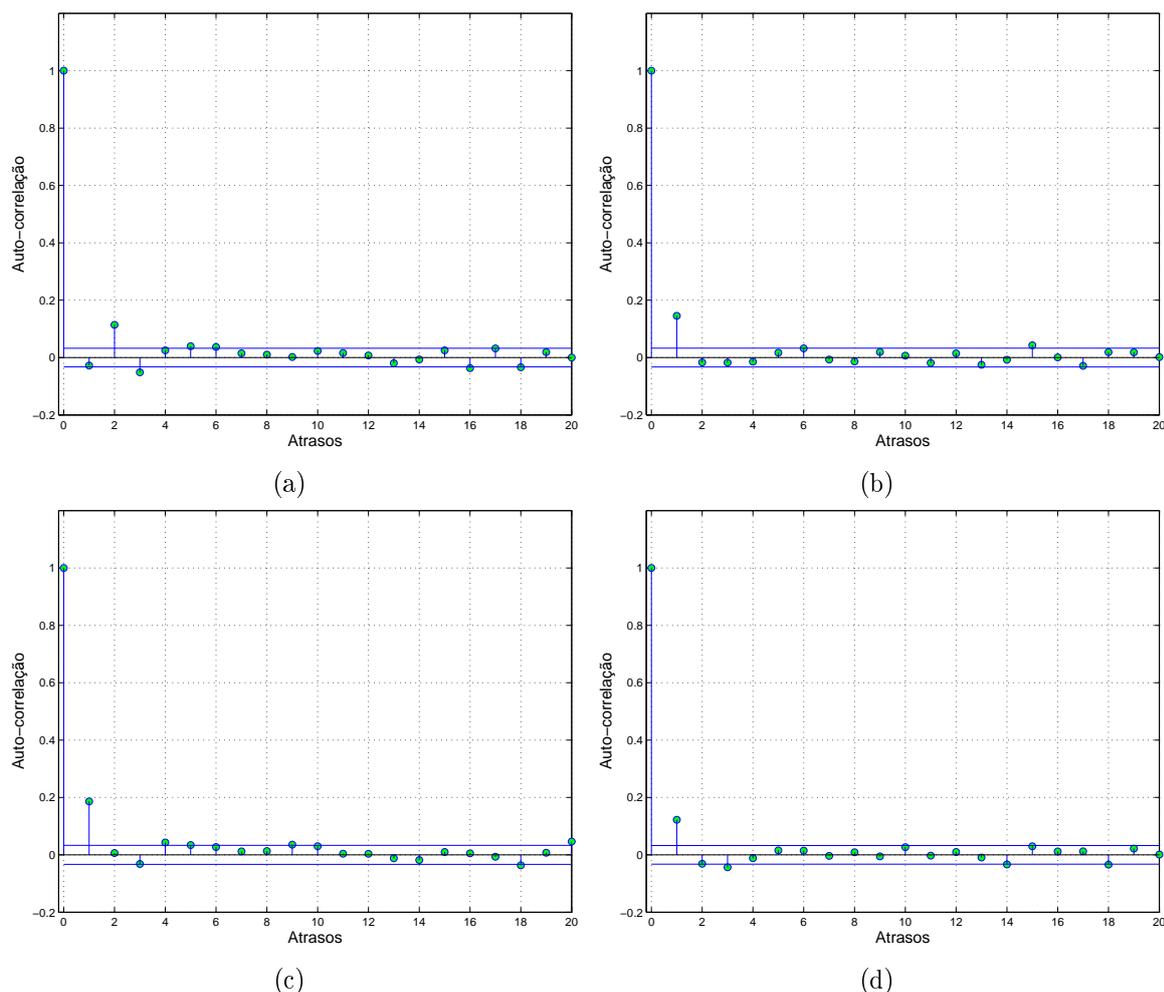


Figura 2.7. Função de autocorrelação dos retornos: (a) PETR4O, (b) PETR4H, (c) PETR4L, (d) PETR4C. As linhas contínuas horizontais mostram o intervalo de confiança considerando-se dois desvios padrão.

2.4 Testes de Heteroscedasticidade Breusch-Pagan e White

Um processo aleatório cuja variância não seja constante em todas as observações é dito heteroscedástico. Tal processo não é identicamente distribuído, o que traz várias consequências para sua análise. Uma delas é o fato de o estimador de mínimos quadrados, amplamente usado em identificação de sistemas, não ser o estimador mais eficiente (de menor variância assintótica) entre os estimadores lineares. Deve-se, então, modelar a heteroscedasticidade (e retirá-la dos dados, se possível) ou utilizar métodos que sejam robustos à mesma. Para isso, deve-se primeiro identificar se os dados possuem essa característica ou não. Seria desejável que se pudesse testar a seguinte hipótese:

Tabela 2.2. Estatística $q(10)$ para as séries trabalhadas. Valores de $q(10)$ acima do valor limite indicam presença de correlações lineares significativas.

Série	$q(10)$	Valor limite de $q(10)$
DJIA	92,8	18,3
SP500	70,5	18,3
IBOV	474	18,3
PETR4O	57,8	18,3
PETR4H	98,4	18,3
PETR4L	149	18,3
PETR4C	78,7	18,3

$$H_0 : \sigma^2(i) = \sigma^2 \quad \text{para todo } i$$

$$H_1 : \text{negar } H_0. \quad (\text{processo heteroscedástico})$$

O teste é, porém, bastante genérico. É possível conseguir, na prática, testar apenas condições mais específicas, partindo-se de algum modelo específico para a variância. Os testes propostos por Breusch e Pagan e por White, descritos em Wooldridge [2003], embora consigam detectar a heteroscedasticidade, não a testam diretamente, mas o fazem testando se os coeficientes de alguns modelos de variância são nulos. Partindo-se do seguinte modelo linear:

$$y(t) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k + \epsilon(t), \quad (2.15)$$

sendo x_1 , x_2 e x_k os regressores e θ_0 , θ_1 , θ_2 e θ_k os parâmetros a serem estimados, $\epsilon(t)$ é o erro cometido no instante t pelo modelo ao tentar explicar $y(t)$ por meio dos regressores e tem esperança nula. Esse modelo será chamado de modelo da média, para contrastar com o modelo da variância, descrito posteriormente. A hipótese nula de homoscedasticidade pode ser colocada da seguinte forma:

$$H_0 : \text{var}(\epsilon(t)|x_1, x_2, \dots, x_k) = \sigma^2.$$

Como $E[\epsilon(t)|x_1, x_2, \dots, x_k] = 0$, a hipótese anterior pode ser escrita como :

$$H_0 : \text{var}(\epsilon(t)|x_1, x_2, \dots, x_k) = E[\epsilon(t)^2|x_1, x_2, \dots, x_k] = \sigma^2.$$

Se $\epsilon(t)^2$, em cada instante, depender do valor da série em instantes anteriores, a hipótese acima é rejeitada. Pode-se, para isso, propor o seguinte modelo:

$$\epsilon(t)^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k + \nu, \quad (2.16)$$

em que δ_0 , δ_1 , δ_2 e δ_k são os parâmetros do modelo proposto para $\epsilon(t)^2$. Esse é o modelo da variância, que considera a correlação entre o quadrado do erro e os regressores do modelo da média. O teste Breusch-Pagan utiliza a Equação 2.16.

Adicionando-se à equação anterior os quadrados dos regressores e seus produtos cruzados chega-se ao modelo utilizado por White (White [1980]). Considerando $k = 3$, obtém-se a Equação 2.17 para o teste de White mostrada a seguir:

$$\begin{aligned} \epsilon(t)^2 = & \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \\ & \delta_4 x_1^2 + \delta_5 x_2^2 + \delta_6 x_3^2 + \\ & \delta_7 x_1 x_2 + \delta_8 x_1 x_3 + \delta_9 x_2 x_3 + \nu. \end{aligned} \quad (2.17)$$

Para as Equações 2.16 e 2.17, a hipótese nula de homoscedasticidade é:

$$H_0 : \delta_0 = \delta_1 = \delta_2 = \dots = \delta_k = 0. \quad (2.18)$$

Pode-se, então, rejeitar a homoscedasticidade se os coeficientes do modelo da variância estimado forem estatisticamente diferentes de zero. Caso os coeficientes sejam nulos, os dados podem ser realmente homoscedásticos ou pode ser que o modelo utilizado não seja adequado.

A princípio, não se conhece $\epsilon(t)^2$ para que se possa estimar os modelos acima. Entretanto os resíduos da Equação 2.15 podem ser usados como estimativa de $\epsilon(t)$, denominada $\hat{\epsilon}(t)$. O teste é então realizado utilizando-se a estatística *LM* (*Lagrange Multiplier*, descrita em Wooldridge [2003]), dada por:

$$LM(k) = N \cdot R_{\hat{\epsilon}(t)^2}^2 \quad (2.19)$$

em que N é a quantidade de observações e $R_{\hat{\epsilon}(t)^2}^2$ é o coeficiente de determinação da regressão realizada em $\hat{\epsilon}(t)^2$ (definida logo abaixo). A estatística *LM* é calculada em função da quantidade de parâmetros do modelo da Equação 2.17. Define-se a soma dos quadrados dos resíduos *SSR* (*residual sum of squares*), a soma dos quadrados do desvio da série (*total sum of squares*) *SST* e o cálculo de R^2 nas equações 2.20, 2.21 e 2.22, respectivamente.

$$SSR = \sum \text{resíduo}^2 = N\sigma_r^2, \quad (2.20)$$

$$SST = \sum (y(t) - \text{média}(y(t)))^2 = N\sigma^2, \quad (2.21)$$

$$R^2 = 1 - \frac{SSR}{SST}. \quad (2.22)$$

Para determinado N , SSR pode ser visto como uma medida da variância dos resíduos, σ_r^2 . Analogamente, SST representa a variância total, σ^2 , da série. Nesse sentido, R^2 é interpretado como a razão entre a variância que pôde ser explicada pelo modelo e a variância total.

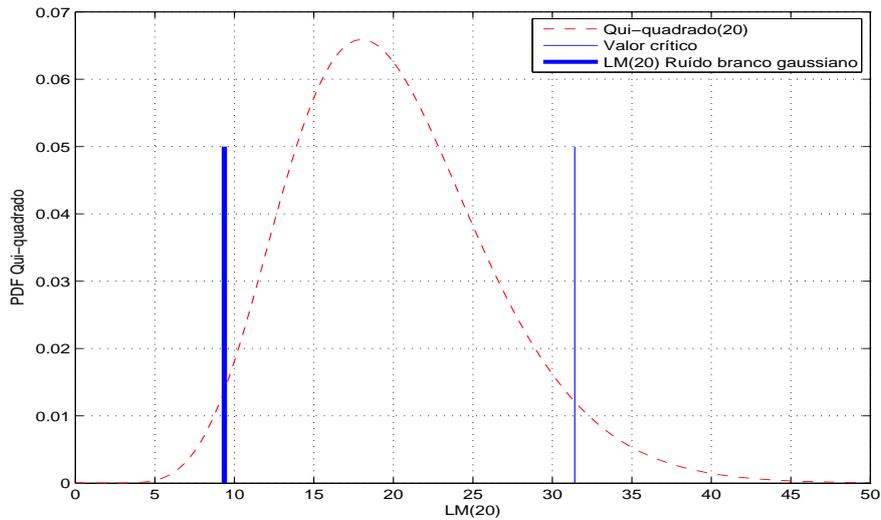
Segundo a hipótese da Equação 2.18, a estatística LM dos dados segue uma distribuição qui-quadrado com k graus de liberdade. k é a quantidade de variáveis independentes utilizadas nos modelos das equações 2.16 e 2.17. Quando se suspeita que a heteroscedasticidade depende de outras variáveis que não as descritas por Breusch-Pagan e White, pode-se incluí-las na estimação, apenas observando o fato de que k é alterado com a adição de novas variáveis independentes. No exemplo da Equação 2.17, $k = 9$, ou seja, δ_0 não é incluído para se obter o número de graus de liberdade da distribuição.

Pode-se finalmente calcular $LM(k)$ e comparar o valor obtido com um valor limite a partir do qual rejeitar-se-ia a hipótese nula de homoscedasticidade. Considerando-se um intervalo de confiança de dois desvios-padrão, o valor limite é aproximadamente o do percentil 95%. A Figura 2.8 mostra os resultados do Teste de White para o ruído branco gaussiano (homoscedástico) e um processo ARCH(1) com o parâmetro autorregressivo igual a 0,3. Valores de LM acima do limite indicam a rejeição da hipótese nula de homoscedasticidade. Na figura, também é mostrada a PDF (*probability density function*) da distribuição em questão.

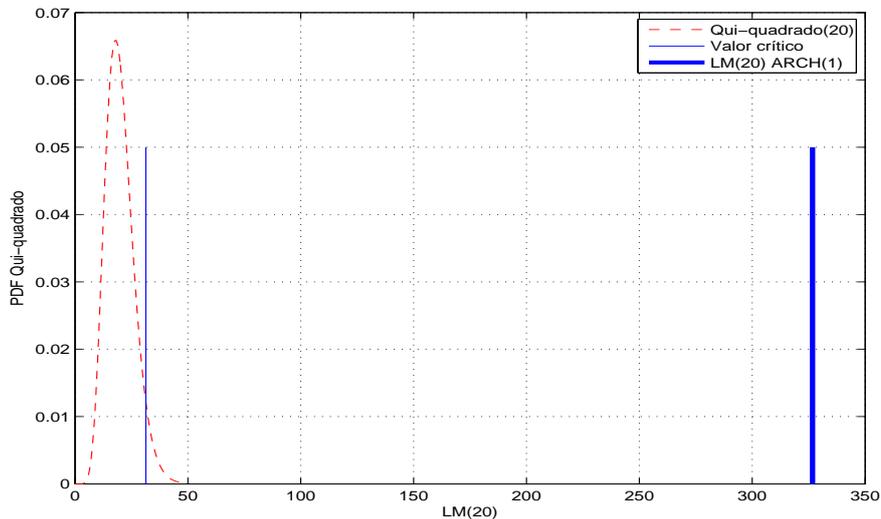
Pode-se perceber pelas figuras que a série heteroscedástica conseguiu ser detectada pelo teste de White. As Tabelas 2.3 e 2.4 mostram o resultado dos Testes Breusch-Pagan e White aplicadas a um ruído branco gaussiano (WGN), um ruído branco uniforme (WUN), processos ARCH(1), GARCH(1,1) e mapa logístico (LOGMAP).

Tabela 2.3. Estatística $LM(5)$ para várias séries conhecidas. Se $LM(5)$ é maior que seu valor limite, rejeita-se a hipótese nula de homoscedasticidade.

Série	$LM(5)$ Breusch-Pagan	- Valor limite de $LM(5)$
WGN	4,25	11,0
WUN	2,34	11,0
ARCH(1)	7,08	11,0
GARCH(1,1)	8,23	11,0
LOGMAP	6,71	11,0



(a) Caso homoscedástico.



(b) Caso heteroscedástico.

Figura 2.8. As figuras acima mostram a PDF de uma distribuição qui-quadrado com 20 graus de liberdade, o valor crítico ou limite, representado pelo percentil 95 e os valores da estatística LM para (a), um ruído branco gaussiano e (b), uma série ARCH(1). O $LM(20)$ para o caso do ruído está abaixo do valor limite enquanto o de ARCH(1) supera em muito esse valor.

Para o teste Breusch-Pagan, pode-se perceber que o teste falhou ao não detectar a heteroscedasticidade nas séries ARCH(1) e GARCH(1,1). Isso acontece pelo tipo de modelo. A Equação 2.16 utilizada não inclui os regressores que influenciam a variância de processos ARCH e GARCH. No caso ARCH, são os quadrados dos valores da série original. No caso GARCH, além dos quadrados da série original, há também como regressores os valores anteriores da série da variância. Nenhum desses é contemplado no modelo Breusch-Pagan.

Tabela 2.4. Estatística $LM(20)$ para várias séries conhecidas. Se $LM(20)$ é maior que seu valor limite, rejeita-se a hipótese nula de homoscedasticidade.

Série	$LM(20)$ White	- Valor limite de $LM(20)$
WGN	9,37	31,4
WUN	20,9	31,4
ARCH(1)	326	31,4
GARCH(1,1)	357	31,4
LOGMAP	3294	31,4

O resultado do teste de White rejeita corretamente a homoscedasticidade de ARCH(1) e GARCH(1,1). O mapa logístico, porém, é identificado como heteroscedástico. Isso se dá por que o teste não mede diretamente a heteroscedasticidade mas uma dependência temporal do quadrado da série. O mapa possui uma dependência não linear que é, de certa maneira, capturada pelo modelo de White. O que faz com que o teste de White obtenha resultados melhores que o Breusch-Pagan são o tipo de não linearidade considerada no modelo e não a quantidade de dados utilizados. Nos dois testes acima, foram utilizados os mesmos cinco valores atrasados de cada série: $y(t-1)$, $y(t-2)$, $y(t-3)$, $y(t-4)$, $y(t-5)$.

Esses resultados mostram que os testes de heteroscedasticidade, como os apresentados aqui, têm limites de desempenho e dependem do modelo de variância usado. Essa característica é ruim no aspecto de que nem sempre é possível saber quais serão os regressores a serem incluídos. Porém, uma vantagem é que esses testes acabam se tornando testes de especificação de modelos, o que auxilia no estudo da variância.

2.5 Heteroscedasticidade dos Retornos

Verificar a presença de heteroscedasticidade é, em outras palavras, verificar como se dá o processo da variância. Se a mesma é constante ao longo do tempo, os dados são homoscedásticos, como supõe o modelo do passeio aleatório. Ora, a variância não é obtida diretamente de cada observação mas estimada a partir de um conjunto representativo de dados. O problema de se estimar a variância em um processo heteroscedástico está exatamente em obter esses dados representativos. Em séries financeiras não há várias

observações em cada instante mas, no melhor caso, uma única observação⁴. Logo, uma estimativa precisa da variância é, muitas vezes, difícil de se obter.

Há várias técnicas para se estimar a variância de séries financeiras como o cálculo da volatilidade histórica (Tsay [2005]), o uso de cotações de abertura, fechamento, máxima e mínima (Garman & Klass [1980]) entre outros. Todos, porém, apresentam deficiências não havendo um único método que ofereça a melhor estimativa. Dado esse cenário, nesta seção serão apresentadas algumas técnicas para se testar a heteroscedasticidade não por meio da série de variâncias, mas pelos seu reflexo na série de retornos. O objetivo não é detalhar o assunto mas apenas utilizar alguns dos métodos conhecidos para se verificar a variabilidade da variância ao longo da série. Para isso serão utilizados o teste Ljung-Box do quadrado dos retornos e os testes de White e Breusch-Pagan, descritos na Seção 2.4.

É importante ressaltar que séries heteroscedásticas são não estacionárias. Essa característica, porém, não impõe alguma dependência temporal não linear nessas séries, desde que suas variâncias sejam independentes. A dependência da variância, porém, insere nos dados uma dependência temporal que pode ser percebida na série formada pelo quadrado dos valores originais, como visto na Seção 2.4. Em processos do tipo ARCH ou GARCH, por exemplo, as séries geradas são não estacionárias (a variância não é constante) e também não lineares (há dependência temporal no processo da variância). Há uma grande dificuldade em se distinguir os dois casos, uma vez que as técnicas utilizadas neste trabalho e várias outras existentes, de certa maneira⁵, pressupõe estacionariedade. Testes de heteroscedasticidade e dependência não linear são, portanto, de grande importância.

Ao longo deste trabalho, a heteroscedasticidade é tratada como uma forma de não linearidade. Embora exista uma clara diferença entre os dois termos, essa consideração é válida quando há uma dependência temporal no processo da variância (Tsay [2005]). Como mostra esta seção, essa dependência existe e, nesse contexto, o primeiro termo

⁴Há casos em que não se possui observações em todos os instantes considerados. Esse fenômeno é chamado de amostragem irregular e é recorrente no estudo de dados de alta frequência. Um exemplo é o de séries com dados intradiários com intervalo de cotações de ativos em minutos. Em alguns instantes, não há negociação, o que compromete a estimativa de variância. Esse efeito também é sentido em dados diários, em dias sem negociações como fins de semana e feriados. A escolha de ativos com grande liquidez (maior probabilidade de negociação), o aumento do intervalo de tempo ou a escolha de mercados sem interrupção de negociações são alternativas para se minimizar o problema. Uma abordagem interessante para o problema é a interpolação e previsão de cotações indisponíveis utilizando filtro de Kalman, em um modelo conhecido como *Local Trend Model* encontrado em Tsay [2005].

⁵Pode-se considerar, na hipótese nula desses testes, que os dados são estacionários. Nesse caso, a rejeição da hipótese indicaria também a possível presença de não estacionariedade. Essa indicação, entretanto, não faz parte dos objetivos dos testes, que são, testar independência, linearidade ou tipos de não linearidade.

será utilizado como um sinônimo para o segundo.

2.5.1 Teste Ljung-Box do Quadrado dos Retornos

O mesmo teste de dependência linear usado na seção anterior pode ser usado para se verificar a existência de heteroscedasticidade, como mostrado na Seção 2.2. O teste é aplicado ao quadrado da série de retornos $y(t)^2$. Se há correlação significativa, rejeita-se a hipótese de homoscedasticidade. A variância está intimamente relacionada com $y(t)^2$, supondo que a média seja nula. Uma dependência temporal em $y(t)^2$ sugere também uma dependência temporal do processo de variância. Logo, a variância de $y(t)$, dada por $\sigma^2(t)$ dependerá de valores anteriores da série:

$$\sigma^2(t) \approx F(y(t-1), y(t-2), \dots).$$

Consequentemente, $E[\sigma^2(t)] \neq E[\sigma^2(t)|y(t-1), y(t-2), \dots]$. Para valores diferentes de $y(t-1), y(t-2), \dots$, a variância, naturalmente, não será constante, indicando, portanto a heteroscedasticidade. A Tabela 2.5 mostra os resultados desse teste.

Tabela 2.5. Estatística $q(10)$ para o quadrado das séries trabalhadas. Valores de $q(10)$ acima do valor limite indicam presença de correlações lineares significativas.

Série	$q(10)$	Valor limite de $q(10)$
DJIA	3863	18,3
SP500	2501	18,3
IBOV	2976	18,3
PETR4O	1349	18,3
PETR4H	1588	18,3
PETR4L	1665	18,3
PETR4C	1800	18,3

Em todas as séries examinadas o valor de $q(10)$ é maior que o valor limite. Os resultados mostram a presença significativa de dependência linear no quadrado dos retornos, o que rejeita a hipótese de homoscedasticidade.

2.5.2 Teste de White e Breusch-Pagan nas Séries de Retornos

Os testes desta seção são aplicados nas séries estudadas utilizando-se cinco valores atrasados da série: $y(t-1)$, $y(t-2)$, $y(t-3)$, $y(t-4)$ e $y(t-5)$. Essa quantidade é escolhida arbitrariamente, pois não é objetivo do trabalho modelar precisamente a va-

riância, mas apenas detectar sua variação temporal. Julga-se que não haja correlações significativas em atrasos maiores que cinco. A Tabela 2.6 exhibe os resultados obtidos.

Tabela 2.6. Estatísticas $LM(5)$ e $LM(20)$ utilizadas nos testes de White e Breusch-Pagan, respectivamente. Também são mostrados os valores limite, a partir do qual a hipótese de homoscedasticidade é rejeitada.

Série	Teste de White		Teste Breusch-Pagan	
	$LM(5)$	Valor limite de $LM(5)$	$LM(20)$	Valor limite de $LM(20)$
DJIA	631	11,1	2371	31,4
SP500	340	11,1	1885	31,4
IBOV	36,5	11,1	1319	31,4
PETR4O	162	11,1	1130	31,4
PETR4H	85,6	11,1	744	31,4
PETR4L	334	11,1	1408	31,4
PETR4C	259	11,1	1102	31,4

No caso do mapa logístico, o teste de White resulta em $LM(5) = 31,9$ e o teste Breusch-Pagan, $LM(20) = 82,3$, para os mesmos valores limites mostrados na Tabela 2.6. A questão da distinção entre heteroscedasticidade e dependência não linear permanece em aberto. Isso porque o que é medido no teste é exatamente a dependência dos quadrados das observações em séries estocásticas. O teste não consegue, então, diferenciar uma série completamente determinística, mas não linear, de uma série que realmente apresente a variação de sua variância ao longo do tempo.

Os resultados obtidos indicam a presença de heteroscedasticidade para todas as séries estudadas. Como, em séries artificiais, os testes não diferenciam a não linearidade previsível da heteroscedasticidade, os resultados com séries financeiras podem ser pouco conclusivos em termos de identificar a correta estrutura presente nos dados. Em contrapartida, esses certamente apresentam uma outra ou as duas características, fazendo com que sejam diferentes de um passeio aleatório.

2.6 Não Normalidade dos Retornos

O estudo da distribuição de probabilidades dos retornos é relevante por vários motivos. Primeiramente, a distribuição poderá influenciar os métodos e testes a serem usados na análise dos dados. Outro motivo a ser citado é o fato de que se a distribuição não é gaussiana (ou normal), a variância pode não ser mais uma estimativa confiável do risco. De fato, o que se percebe em diversas séries financeiras é a presença de caudas grossas no histograma dos retornos (séries leptocúrticas) como mostrado na Figura 2.9.(d).

Esta é colocada em escala logarítmica para evidenciar que os valores mais afastados da média são mais prováveis do que em uma distribuição gaussiana correspondente (de mesma média e variância). As Figuras 2.9.(a), 2.9.(b) e 2.9.(c) não mostram claramente a presença das caudas grossas. Mas, ao se colocar esses histogramas em um gráfico logarítmico essa característica pode ser facilmente visualizada (gráfico 2.9.(d)). A maior probabilidade de se obter valores mais afastados da média faz com que o risco estimado apenas pela variância seja, de certa maneira, subdimensionado. Distribuições com essa característica, como a de *Lévy*, poderiam ser utilizadas, porém há outras alternativas (Nascimento Jr. *et al.* [2007] e Mandelbrot [1963]). Esses autores mostram que a origem desse fenômeno pode estar ligada a dependências temporais de longo prazo. Mandelbrot, por exemplo, defende que as caudas grossas seriam originadas de multifractais, o que faria com que as séries se tornassem previsíveis (Mandelbrot *et al.* [1997]). Portanto, a detecção de não normalidade pode até indicar, em alguns casos, a presença de determinismo.

A questão da escolha das distribuições não é o foco do presente trabalho. Esta seção apenas mostra que os resultados de testes de normalidade em séries indicam uma rejeição da hipótese de normalidade. Pelos motivos anteriormente citados, esses testes, servem de base para um estudo mais profundo a respeito da possibilidade de haver determinismo em séries financeiras, além de rejeitar o passeio aleatório gaussiano.

A Tabela 2.7 mostra o resultado do teste de Kolmogorov-Smirnov aplicado às séries estudadas no presente trabalho e também a algumas séries conhecidas, a saber, o ruído branco gaussiano (WGN) e o ruído branco uniforme (WUN). O tamanho da amostra (N) também é mostrado nessa tabela, a fim de se interpretar melhor a validade dos resultados dos testes.

Tabela 2.7. Resultado do teste de normalidade Kolmogorov-Smirnov para as séries trabalhadas e para séries artificiais.

Série	Resultado do teste ¹	Valor da Estatística	Valor limite da estatística	Tamanho da amostra
WGN	0	0,0150	0,0229	3500
WUN	1	0,500	0,0229	3500
DJIA	1	0,480	0,00820	28723
SP500	1	0,483	0,0110	15452
IBOV	1	0,463	0,0137	10581
PETR4O	1	0,460	0,0235	3903
PETR4H	1	0,463	0,0235	3903
PETR4L	1	0,463	0,0235	3903
PETR4C	1	0,462	0,0235	3903

¹0 - não rejeita hipótese; 1 - rejeita hipótese.

A segunda coluna mostra o resultado do teste. Nessa coluna, ‘1’ indica rejeição da hipótese nula de que a série é gaussiana e ‘0’ indica a não rejeição da hipótese. Em todas as séries, excetuando-se o ruído branco gaussiano, foi possível rejeitar a hipótese de que os dados têm distribuição gaussiana. O ruído branco uniforme é sabidamente não gaussiano e, juntamente com o ruído branco gaussiano, é usado para se testar a validade do método.

Os retornos das séries estudadas não são gaussianos. Esse resultado é coerente com o encontrado geralmente na literatura, como em Torres *et al.* [2002], que analisa ações comercializadas na Bovespa. A distribuição dos retornos é tipicamente leptocúrtica (caudas grossas). Isso é explicado no mercado pela alta frequência de ocorrência de movimentos de grande amplitude, como crises. Esses movimentos são mais frequentes do que se esperaria caso os retornos fossem gaussianos. As Figuras 2.9 e 2.10 mostram o histograma e a distribuição gaussiana correspondente estimada para cada uma das séries trabalhadas. O histograma foi normalizado em amplitude a fim de compará-lo com a distribuição. Percebe-se, pela figura, que há uma quantidade de observações mais afastadas da média, fora da linha que representa a distribuição (valores extremos). Essas observações são mais prováveis do que uma distribuição gaussiana poderia explicar. A Figura 2.9.(d) evidencia essa característica por meio do uso da escala logarítmica no eixo vertical. As outras séries apresentam resultado semelhante nessa escala. O teste de Kolmogorov-Smirnov realizado anteriormente valida essa suposição.

2.7 Características Estatísticas das Séries Financeiras

O presente capítulo mostrou a comparação de algumas séries financeiras com o modelo do passeio aleatório. Como a série de índices e preços é *não estacionária em média* (o primeiro momento estatístico varia ao longo do tempo), utilizou-se o logaritmo da primeira diferença dos índices e preços (série de retornos) visando à utilização de alguns testes estatísticos que pressupõem tal estacionariedade. Foram realizados alguns testes e seus resultados são sintetizados abaixo:

- há baixa autocorrelação linear nas séries de retornos. Considerando os custos inerentes aos processos de negociação desses ativos, essa característica faz com que modelos lineares do tipo ARMA possam ter aplicação bastante restrita;
- há significativa autocorrelação do quadrado da série de retornos. Essa característica pode ser associada a não estacionariedade em variância (heteroscedasti-

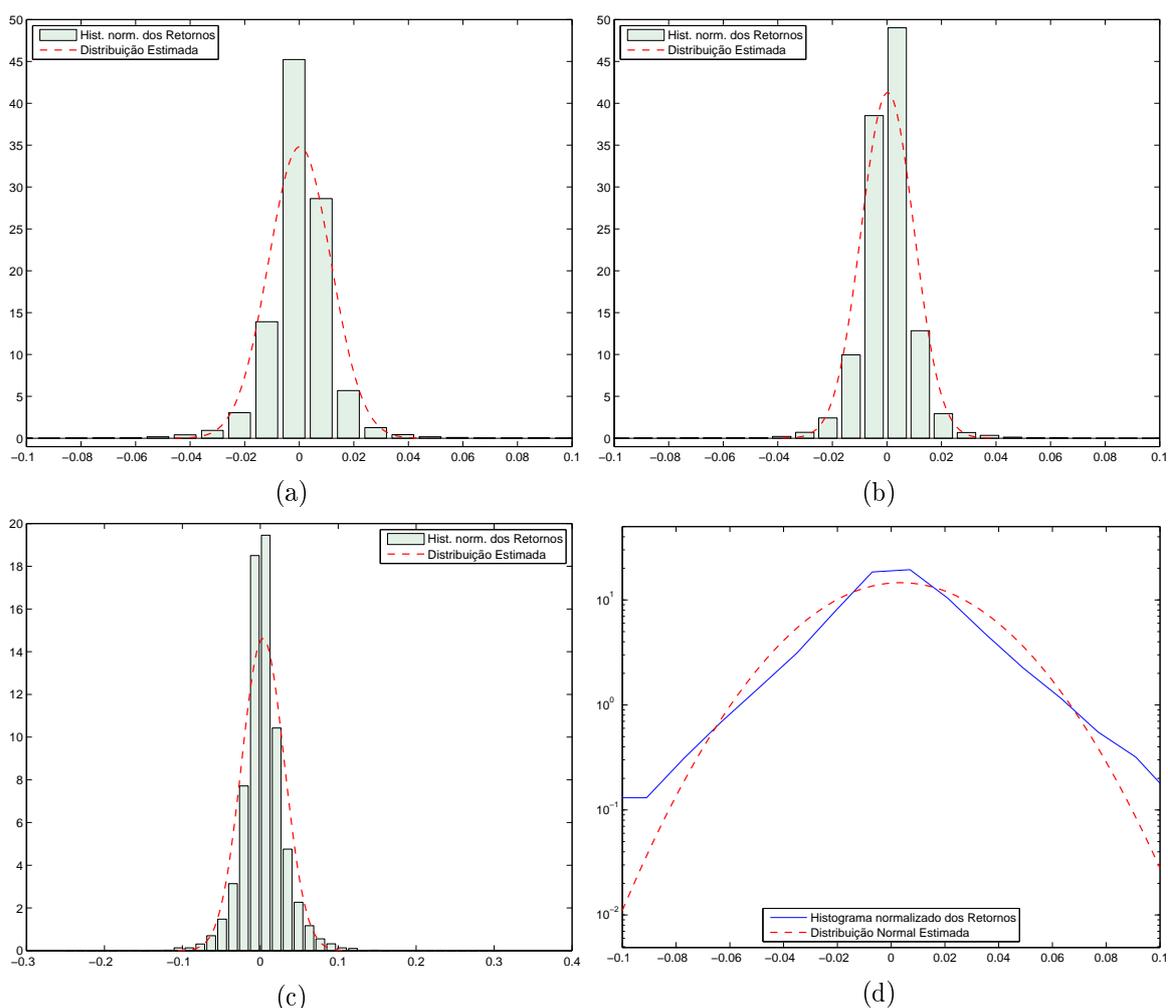


Figura 2.9. As figuras mostram o histograma e a distribuição gaussiana estimada (linha tracejada) das séries: (a) DJIA, (b) SP500, (c) IBOV e (d) IBOV em escala logarítmica. Observa-se a presença de algumas observações mais dispersas do que se poderia esperar se os retornos tivessem uma distribuição gaussiana. Isso se torna evidente quando se usa a escala logarítmica para se comparar a distribuição com o histograma.

cidade) e, conseqüentemente, variação da volatilidade. Alternativamente, essa correlação pode indicar estruturas não lineares previsíveis, mas os testes utilizados são inconclusivos em relação a essas duas alternativas;

- a distribuição dos retornos não é gaussiana.

Essas conclusões questionam a validade do passeio aleatório com inovações gaussianas como um modelo geral para séries financeiras como índices de mercados financeiros e preços de ações. O que merece atenção é, principalmente, o questionamento de que preços passados possam influenciar preços futuros. Isso justifica um estudo mais apro-

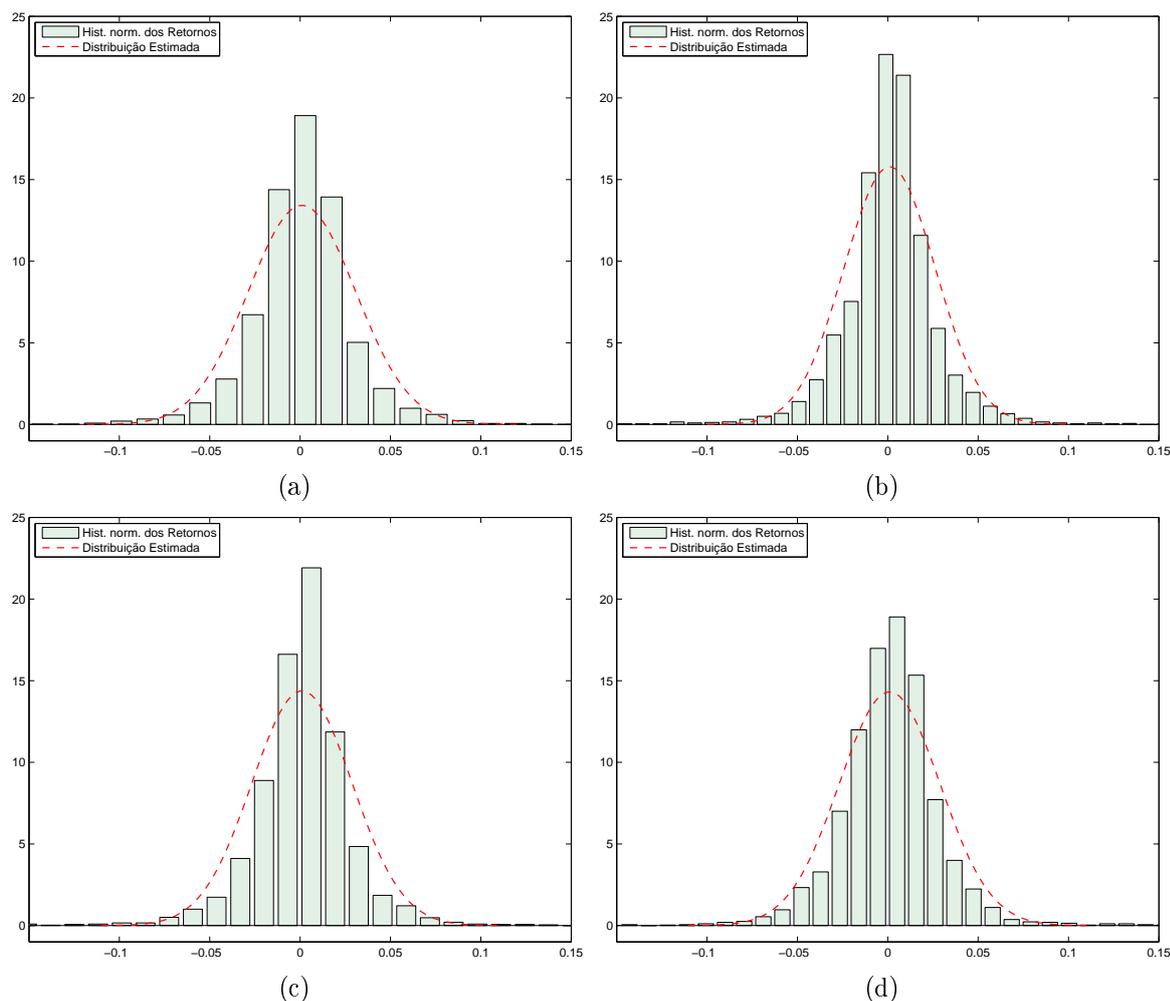


Figura 2.10. As figuras mostram o histograma e a distribuição gaussiana estimada (linha tracejada) das séries: (a) PETR4O, (b) PETR4H, (c) PETR4L, (d) PETR4C. Observa-se a presença de algumas observações mais dispersas do que se poderia esperar se os retornos tivessem uma distribuição gaussiana.

fundado a respeito da previsibilidade das séries mostradas visando explorar cada uma das três características mostradas acima. Os pontos principais a serem tratados neste trabalho nos próximos capítulos são:

- utilizar métodos de teste não paramétricos, para que não haja necessidade da suposição de que os retornos são gaussianos;
- estender os testes de previsibilidade para considerar dependências não lineares entre os retornos (por meio do uso de índices estatísticos mais abrangentes do que os baseados em autocorrelação linear);
- verificar se as dependências temporais existentes permitem uma melhor previsão

do valor futuro dos retornos ou contribuem apenas para a previsão de outras características como a volatilidade futura.

Métodos e Técnicas

O presente capítulo descreve as principais técnicas usadas ao longo deste trabalho. Inicia-se o capítulo revendo-se algumas nomenclaturas e definições. Posteriormente, são apresentadas algumas formas de se quantificar o determinismo. São utilizadas para esse fim a dimensão de correlação, as entropias aproximada e amostral, a informação mútua média e o teste BDS. Essas ferramentas serão denominadas índices de determinismo ou ainda, estatísticas discriminantes. Neste trabalho, não se fará distinção entre esses dois termos. O último, estatística discriminante, evidencia a função que essas ferramentas terão junto a utilização de dados sub-rogados. A análise de dados sub-rogados é também apresentada neste capítulo, finalizando a descrição das técnicas que serão usadas para se classificar as séries quanto a presença de diferentes tipos de determinismo.

3.1 Previsão, Previsibilidade e Determinismo

Antes de apresentar os métodos para se quantificar determinismo, serão mostradas algumas definições utilizadas no restante do trabalho, baseadas em Small & Tse [2003b].

Para uma série temporal $y(t)$ cuja esperança seja nula ($E[y(t)] = 0$):

- a série é dita previsível se:

$$E[y(t)|y(t-1),y(t-2),\dots] \neq 0;$$

- a série é dita determinística e previsível se existe uma função determinística (não estocástica) f , tal que:

$$y(t) = f(y(t-1),y(t-2),\dots);$$

- a série contém um componente determinístico se:

$$y(t) = f(y(t-1), y(t-2), \dots, \epsilon(t), \epsilon(t-1), \epsilon(t-2), \dots), \quad (3.1)$$

em que $\epsilon(t)$ é uma variável aleatória com esperança nula mas não necessariamente i.i.d.;

- o determinismo é dito não linear se a função f acima é uma função não linear.

Séries que apresentem dependência temporal da variância, como ARCH e GARCH, têm um componente determinístico. Esse componente, entretanto, não influencia sua previsibilidade, como definida acima. Uma série é previsível se há alguma dependência temporal entre seus valores que possa tornar sua esperança condicional não nula. Essa dependência, seja ela linear ou não linear, tem origem em um componente determinístico e é isso que diferencia tal série de uma completamente aleatória (independente, estocástica e sem determinismo).

3.2 Dimensão de Correlação

A dimensão de correlação¹ (D_2) é uma expansão do conceito de dimensão para objetos com dimensão fracionária (Small [2005]). No caso de uma série temporal, esses objetos surgem por meio da construção de figuras (espaço de fases) que representem a trajetória de um sistema dinâmico em um espaço definido pelas variáveis de estado. D_2 também mede a complexidade estrutural de uma série temporal (Small & Tse [2003b]) e pode ser utilizada como um indicador da presença de determinismo nos dados. Um ruído, por exemplo, tem dimensão de correlação infinita e ocupará todo o espaço de fases não importando a dimensão de imersão utilizada. D_2 , nesse caso, poderia ser usado como um indicador de previsibilidade. Isso, porém, não quer dizer que uma dimensão de correlação finita indique que uma série não seja estocástica. Pincus mostra um processo estocástico no qual a estimativa de D_2 converge para um número finito (Pincus [1991]). Como definido anteriormente, a série pode ser aleatória mas conter algum determinismo (*e.g.* na evolução da variância). Extrair esse determinismo é fundamental para se obter uma melhor previsibilidade. A Figura 3.1 mostra o contraste entre um ruído branco gaussiano e o mapa logístico em seus respectivos espaços de fases com

¹O índice identifica diferentes dimensões. Como mostra Monteiro [2006], D_0 é conhecida como dimensão de capacidade e D_1 é a dimensão de informação, relacionada com a entropia de Shannon. Dimensão e entropia são, portanto, conceitos correlatos e ambos remetem à complexidade dos dados.

dimensão de imersão 2. Não se pode distinguir o ruído do mapa utilizando-se funções de autocorrelação. Porém, a diferença entre os dois é clara no espaço de fases.

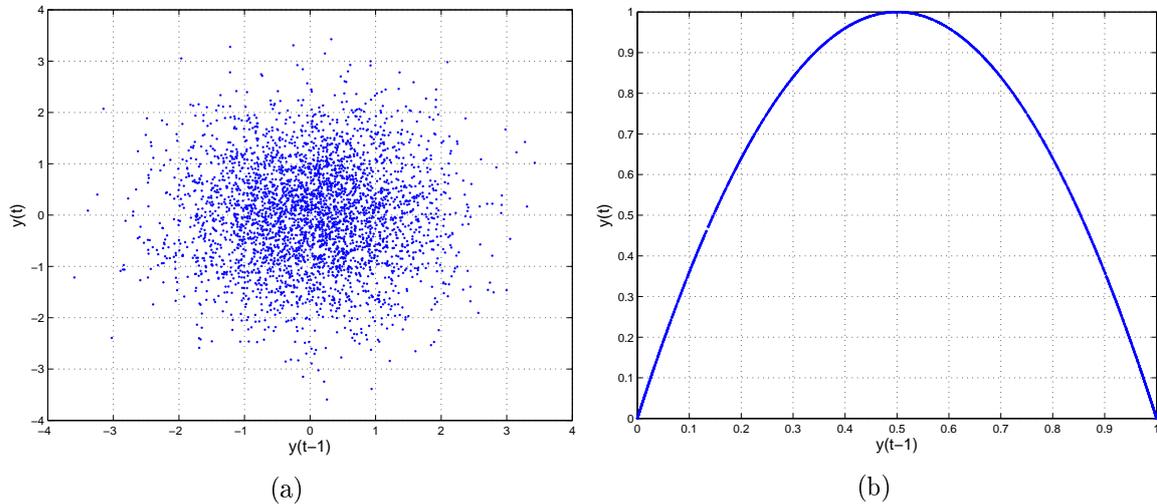


Figura 3.1. Espaço de fases de WGN e LOGMAP. A Figura (a), um ruído branco gaussiano (WGN), é claramente diferente da Figura (b), o mapa logístico (LOGMAP). Essa diferença não pode ser vista nas FAC das séries.

A dimensão de correlação quantifica essa diferença e tem a vantagem de identificar dimensões fracionárias ou de ordem maior que três. Um dos algoritmos mais conhecidos para se estimar D_2 é o desenvolvido por Grassberger e Procaccia (algoritmo GP) em Grassberger & Procaccia [1983]. Segue-se aqui a descrição desse algoritmo contida em Fiedler-Ferrara & Prado [1994], excetuando-se a definição de distância utilizada. Parte-se da fórmula da dimensão de correlação, definida por:

$$D_2 = \lim_{\epsilon \rightarrow 0} \frac{\log(C(\epsilon))}{\log(\epsilon)},$$

em que $C(\epsilon)$ é a integral de correlação e ϵ é distância máxima² entre dois pontos no espaço de fases. Para um conjunto infinito de pontos, $C(\epsilon)$ pode ser dada por:

$$C(\epsilon) = P(|\mathbf{x}_i - \mathbf{x}_j| < \epsilon). \quad (3.2)$$

$C(\epsilon)$ é, segundo Grassberger e Procaccia, a probabilidade de que a distância entre \mathbf{x}_i e \mathbf{x}_j seja menor que ϵ . Os vetores \mathbf{x}_i e \mathbf{x}_j têm tamanho m e ambos são definidos a partir

²Somente nesse contexto, de integral de correlação, ϵ será utilizado como uma distância. No restante do trabalho ϵ é uma variável aleatória que representa a inovação de uma série temporal.

da série temporal $y(t)$ como:

$$\begin{aligned}\mathbf{x}_i &= [y(i) \ y(i-1) \ y(i-2) \ \dots \ y(i-m+1)]^T \\ \mathbf{x}_j &= [y(j) \ y(j-1) \ y(j-2) \ \dots \ y(j-m+1)]^T,\end{aligned}\tag{3.3}$$

em que m é a dimensão de imersão da série, escolhida previamente. $C(\epsilon)$ poderia também ser definido como uma função de m , ou $C_m(\epsilon)$. Como a dimensão m é escolhida de antemão, subentende-se aqui $C(\epsilon)$ para um m constante a fim de simplificar a notação. A distância na Equação 3.2 pode ser definida de várias maneiras. A escolhida neste trabalho é a norma infinito, o valor máximo do módulo obtido após se subtrair o vetor \mathbf{x}_i por \mathbf{x}_j elemento a elemento:

$$|\mathbf{x}_i - \mathbf{x}_j|_\infty = \max(|\mathbf{x}_i - \mathbf{x}_j|),$$

onde

$$|\mathbf{x}_i - \mathbf{x}_j| = [|y(i) - y(j)| \ |y(i-1) - y(j-1)| \ \dots \ |y(i-m+1) - y(j-m+1)|]^T.$$

A probabilidade da Equação 3.2 é dada por:

$$C(\epsilon) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \{\text{número de pares } i, j \text{ tais que } |\mathbf{x}_i - \mathbf{x}_j|_\infty < \epsilon\}.$$

N é o número total de pontos. Para um número finito de pontos, $C(\epsilon)$ pode ser estimado pela frequência relativa de pares de pontos que atendam a $|\mathbf{x}_i - \mathbf{x}_j|_\infty < \epsilon$, dado um determinado ϵ . A dificuldade desse cálculo está no fato de que quanto menor for ϵ , menor será a quantidade de pontos da série que atendam a condição acima, o que traz problemas de subamostragem. O método usual para se estimar D_2 é, então, calcular $C(\epsilon)$ para vários valores pequenos de ϵ e estimar uma reta no gráfico $\log(\epsilon) \times \log(C(\epsilon))$. Para vários sistemas dinâmicos conhecidos, há uma região do gráfico cuja inclinação é constante, denominada *scaling range*. Essa é a região em que se ajusta a reta. Sua inclinação é a estimativa da dimensão de correlação D_2 . A Figura 3.2 mostra esse gráfico para a série LOGMAP com valores de m entre 2 e 8.

A Figura 3.3 mostra a comparação entre as estimativas de dimensão para LOGMAP e WGN. O gráfico exhibe D_2 em função da dimensão de imersão m . Pode-se perceber que o valor de D_2 para o ruído aumenta com a dimensão de imersão, diferentemente de LOGMAP, que se concentra em torno de $D_2 = 1, \forall m$. A série de ruído tem dimensão tão grande quanto a da de imersão (teoricamente, tem dimensão infinita). Como D_2 deve ser calculado para m suficientemente grande ($m \geq 2d + 1$, onde d é a

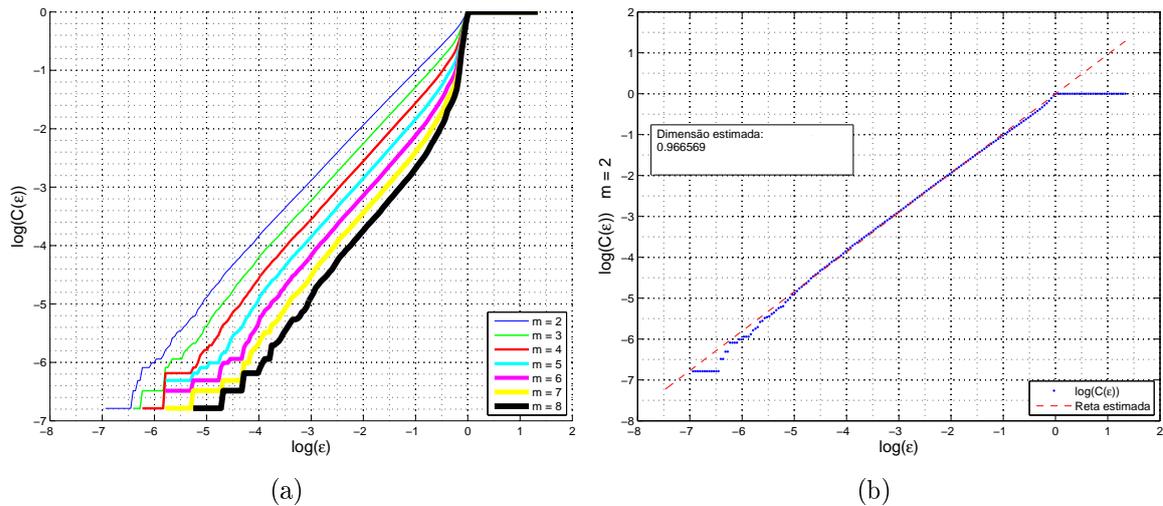


Figura 3.2. Estimativa da dimensão de correlação para o mapa logístico. A Figura (a) mostra o gráfico $\log(\epsilon) \times \log(C(\epsilon))$ para diversos valores de dimensão de imersão (m). Para $m = 2$, a Figura (b) exibe a reta estimada por meio dos valores de ϵ para os quais a inclinação do gráfico é aproximadamente constante.

dimensão de contagem de caixas, pelo Teorema de Takens), o algoritmo indica que essa série é possivelmente um ruído e não contém determinismo significativo. Ser significativo ou não é uma característica que será melhor abordada na descrição da técnica de dados sub-rogados na Seção 3.6.

O custo computacional do algoritmo de Grassberger e Procaccia é bastante elevado e necessita de séries longas para um cálculo razoavelmente preciso. Widman *et al.* [1998] contém várias otimizações do algoritmo identificando redundâncias no cálculo de várias distâncias devido a simetrias entre os vetores \mathbf{x}_i e \mathbf{x}_j (uma vez que todos vêm da mesma série temporal) e também redundâncias quando da utilização do algoritmo com diversos valores de m . Para resolver a questão da quantidade de observações, crítica em dados financeiros, pode-se utilizar a abordagem de Judd, descrita na próxima seção (e utilizada no restante do trabalho).

3.2.1 O Algoritmo de Judd

Judd propõe uma abordagem diferente do algoritmo original para que sejam necessárias menos observações para a convergência de D_2 , além de sugerir um método para calcular o intervalo de confiança de D_2 (Judd [1991] e Judd [1994]). Isso é feito por meio da estimação de um modelo polinomial para a distribuição de probabilidades das distâncias entre os vetores \mathbf{x}_i e \mathbf{x}_j (norma infinito, como definida anteriormente). A dimensão de correlação é um dos parâmetros a estimar desse modelo.

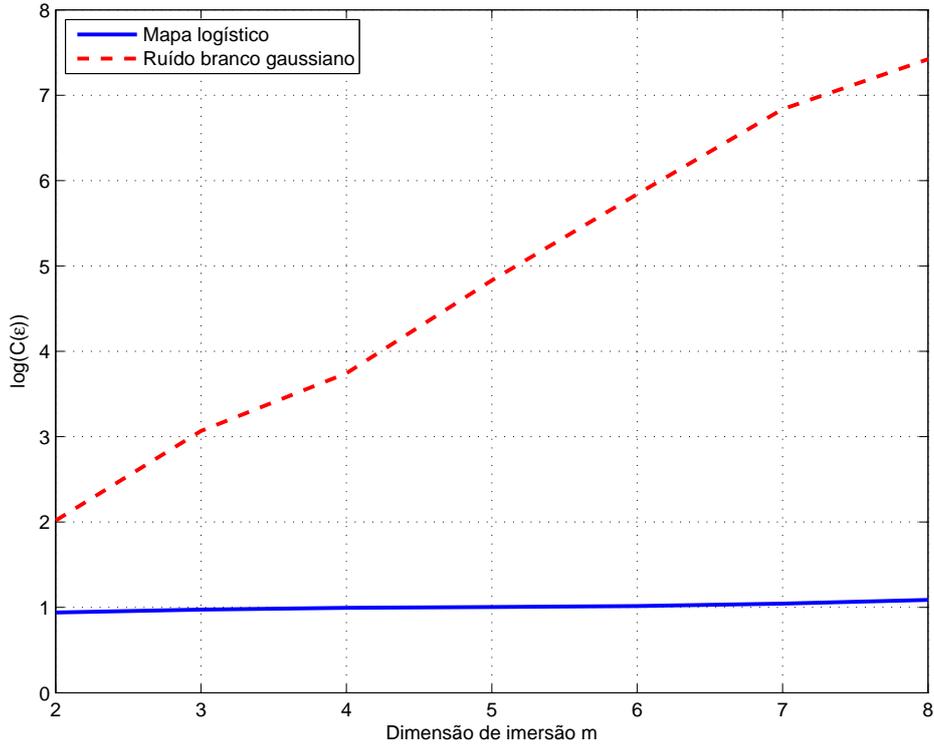


Figura 3.3. Dimensão de correlação estimada em função da dimensão de imersão para o mapa logístico (LOGMAP) e ruído branco gaussiano (WGN). Nota-se que, para o caso do ruído, a dimensão de correlação é sempre próxima do valor de imersão e não se estabiliza. Esse comportamento é condizente com o esperado que é um valor infinito. Como m limita o valor estimado de D_2 , esse último será sempre próximo a m . Para o caso da série LOGMAP, pode-se dizer que a dimensão está próxima da unidade, o que está correto considerando que esse é seu valor teórico.

O algoritmo de Judd parte da distribuição de probabilidades mostrada na Equação 3.2. Ela é chamada por Judd de distribuição de distâncias interpontos (*distribution of interpoint distances*), mas corresponde a integral de correlação. Sua abordagem consiste em estimar um polinômio para essa distribuição com a seguinte forma:

$$P(|\mathbf{x}_i - \mathbf{x}_j|_\infty < \epsilon) = \epsilon^{\hat{D}_2} (a_0 + a_1\epsilon + \dots + a_t\epsilon^p), \quad (3.4)$$

para todo $\epsilon < \epsilon_0$, em que \hat{D}_2 , a_0 , a_1 e a_t são os parâmetros do modelo e p é a ordem do polinômio usado.

A contribuição de Judd está em demonstrar que D_2 pode ser aproximado por \hat{D}_2 utilizando-se a Equação 3.4. Essa equação é válida para valores de *epsilon* suficientemente pequenos, menores que um valor limite, denominado ϵ_0 . Após levantar-se um histograma de distâncias entre os vetores, pode-se estimar o modelo descrito por essa

equação. A dimensão de correlação será um dos parâmetros estimados. A Figura 3.4 mostra a distribuição ajustada para determinado valor de ϵ_0 (o maior ϵ da figura) calculada para a série LOGMAP.

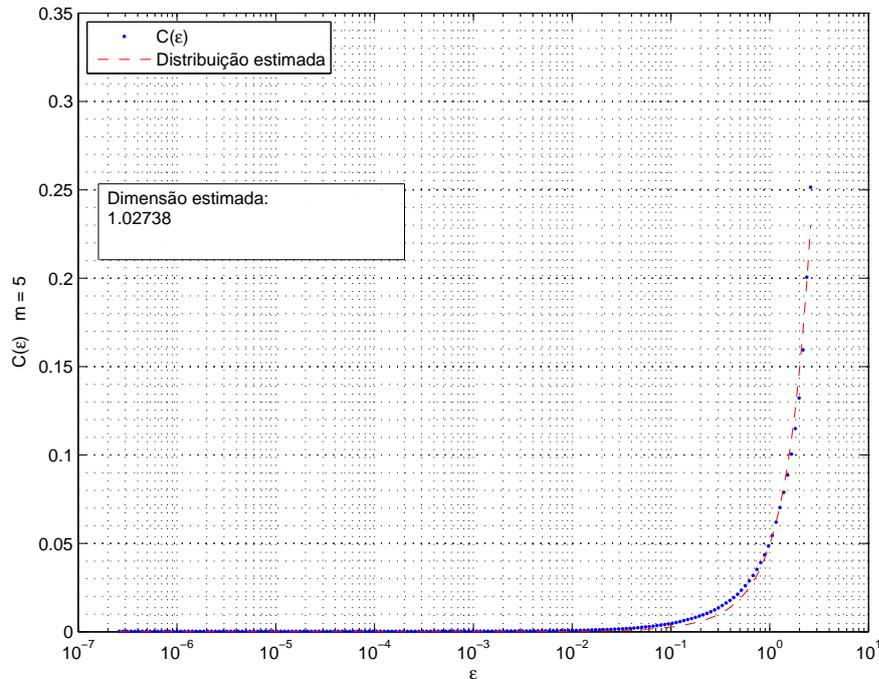


Figura 3.4. A distribuição de probabilidades de $P(|\mathbf{x}_i - \mathbf{x}_j|_\infty < \epsilon)$, ou $C(\epsilon)$, é estimada pelo polinômio da Equação 3.4 para $p = 2$ e $m = 5$. ϵ_0 é o maior valor de ϵ mostrado. A dimensão estimada está próxima da unidade, o que está correto considerando que esse é seu valor teórico.

A diferença principal entre Judd e GP está no fato de que o primeiro considera toda a distribuição interpontos e não apenas o intervalo definido pela *scaling range*, como no caso do algoritmo GP. Nesse último, pode-se perceber pela Figura 3.2 que à medida que o valor de ϵ diminui, a estimativa de $\log(C(\epsilon))$ torna-se menos precisa. Logo, seria necessário um grande número de observações para se contornar essa limitação. Também questiona-se o fato de não se conseguir definir exatamente (e de maneira sistemática) o início e o fim da região de *scaling range* do gráfico, o que faz a estimativa de D_2 variar bastante. O algoritmo de Judd, portanto, necessita de menos observações por usar toda a distribuição e não apenas uma parte dela. Além disso, não depende de uma escolha empírica da região de *scaling range* aumentando a precisão da estimativa de D_2 , denominada \hat{D}_2 .

Uma outra vantagem desse algoritmo é a possibilidade de se obter uma estimativa de erro da dimensão. O erro quadrático médio (RMSE) obtido ao se ajustar o

histograma a uma distribuição é também uma estimativa da dispersão de \hat{D}_2 , o que não acontece no caso do Algoritmo GP. Nesse último, não se pode falar que o erro de regressão da reta na região de *scaling range* é o erro de \hat{D}_2 porque o ajuste depende bastante da região escolhida. A Figura 3.5 mostra \hat{D}_2 para vários valores de ϵ_0 e seus respectivos erros de regressão da distribuição. Por meio dessa figura, pode-se obter um valor de ϵ_0 cujo erro seja mínimo. Deve-se porém, escolher esse parâmetro grande o suficiente para que o histograma seja representativo do atrator em questão. Isso não acontece para os valores estimados de dimensão que são próximos de zero.

Em Judd [1994], o autor trata ϵ_0 como um fator de escala e mostra que para alguns casos, a dimensão de correlação pode mudar de acordo com esse fator. Pode-se entender esse fenômeno como uma mudança na complexidade percebida do atrator em função da distância que se observa o mesmo. Isso faz com que a dimensão seja uma função do fator de escala ϵ_0 . Small e Judd utilizam a implementação de Judd mostrando que \hat{D}_2 pode ser usada como uma estatística para discriminar séries em dados sub-rogados (Small & Judd [1998]). Os autores citam que ϵ_0 influencia as estimativas da dimensão. Contudo, o gráfico $\log(\epsilon_0) \times D_2$ pode ser utilizado para se discriminar séries e detectar determinismo.

Dadas essas observações, o que será utilizado no restante do trabalho não é a estimativa de D_2 mas sim o gráfico com todos os valores do fator de escala $\log(\epsilon_0)$ e suas dimensões correspondentes. O gráfico será utilizado para se comparar séries semelhantes e com conteúdo determinístico diferente, por meio da análise de dados sub-rogados. O algoritmo implementado utiliza valores de ϵ espaçados logaritmicamente a fim de que valores menores sejam melhor amostrados. Além disso, o valor de p da Equação 3.4 é fixado em 2, como sugerido em Judd [1994].

3.3 Entropia Aproximada e Entropia Amostral

Nesta seção são introduzidos dois índices bastante semelhantes: a entropia aproximada e a entropia amostral. Essa última é apresentada como uma continuação do desenvolvimento da primeira.

3.3.1 Entropia Aproximada

A entropia aproximada é citada em Pincus [1991] como uma alternativa ao uso da dimensão de correlação para se quantificar a complexidade de dados experimentais. Segundo Pincus, a entropia aproximada (ApEn) é capaz de classificar sistemas, em função de sua complexidade, com poucas observações.

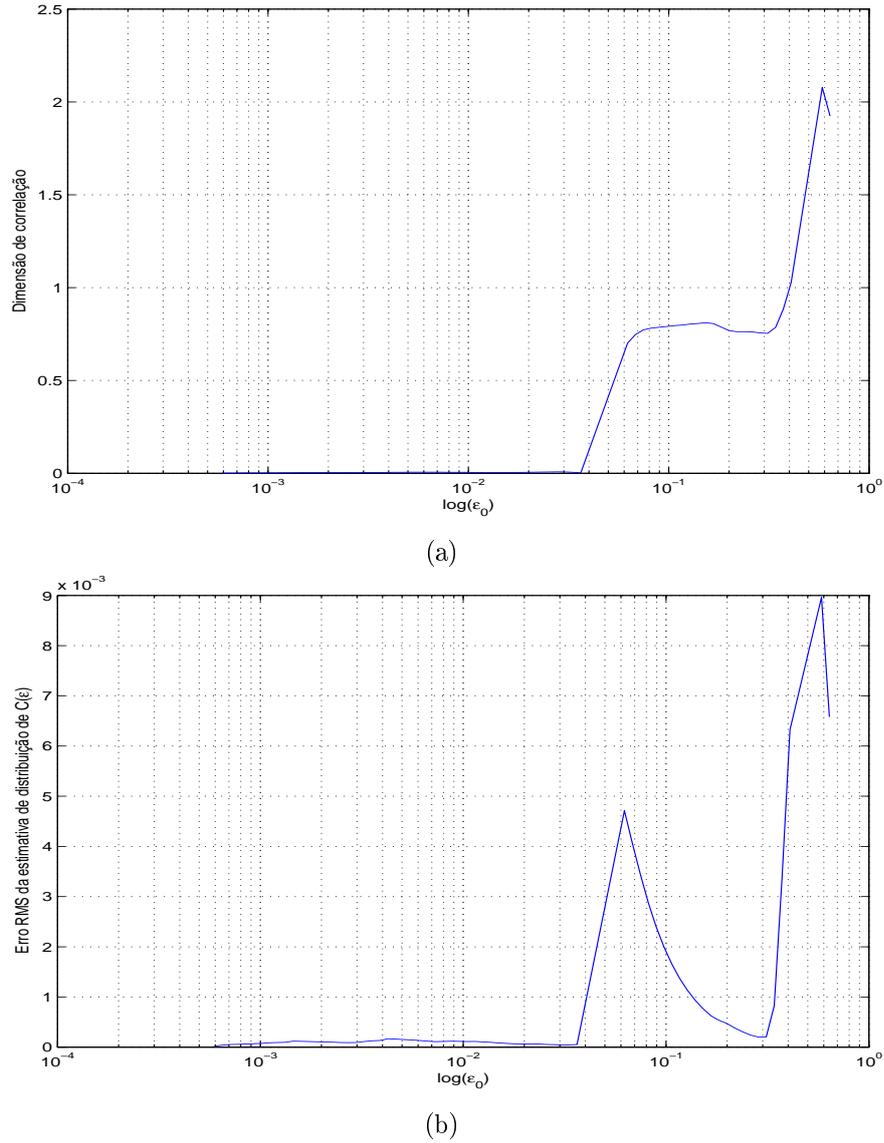


Figura 3.5. A Figura (a) mostra a dimensão de correlação \hat{D}_2 em função de ϵ_0 . Na mesma escala, a Figura (b) mostra os erros de regressão na estimativa da distribuição, que podem também ser vistos como a dispersão de \hat{D}_2 . O valor de ϵ_0 na figura 3.4 foi obtido escolhendo-se, na Figura (b) acima, o de menor erro cuja dimensão estimada fosse diferente de zero. Nessa faixa de valores em que ϵ_0 é pequeno, a dimensão é próxima de zero porque não há dados suficientes para se estimar a distribuição e o erro de regressão não representa a dispersão de $\hat{\epsilon}_0$.

Parte-se da estimativa da integral de correlação como definida na Equação 3.2 e desenvolvida na Seção 3.2, mas agora em função da dimensão de imersão m e, calculada para cada vetor \mathbf{x}_i . Considera-se:

$$C_{m,i}(\epsilon) = \frac{1}{N} \{\text{número de vetores } j \text{ tais que } |\mathbf{x}_i - \mathbf{x}_j|_\infty < \epsilon\} \quad (3.5)$$

Define-se, $\Phi_m(\epsilon)$ como:

$$\Phi_m(\epsilon) = \frac{1}{N} \sum_{i=1}^N \log C_{m,i}(\epsilon). \quad (3.6)$$

A definição da Equação 3.6 assemelha-se a $C_m(\epsilon)$ do cálculo da dimensão de correlação. Entretanto, para $\Phi_m(\epsilon)$, soma-se o logaritmo de $C_{m,i}$.

A entropia de Eckmann-Ruelle é, então, definida como:

$$\text{Entropia E-R} = \lim_{\epsilon \rightarrow 0} \lim_{m \rightarrow \infty} \lim_{N \rightarrow \infty} (\Phi_m(\epsilon) - \Phi_{m+1}(\epsilon)). \quad (3.7)$$

Essa entropia aparentemente não diferencia processos que diferem entre si em termos de complexidade. Além disso, seu valor é infinito para processos somados a ruído de qualquer amplitude (Pincus [1991]). Porém, ao se fixar m e ϵ na Equação 3.7 ao invés de se tomar seus limites, tem-se:

$$ApEn(m, \epsilon) = \lim_{N \rightarrow \infty} (\Phi_m(\epsilon) - \Phi_{m+1}(\epsilon)).$$

Essa fórmula, em um conjunto finito de dados, é implementada para determinado N como:

$$ApEn(m, \epsilon, N) = \Phi_m(\epsilon) - \Phi_{m+1}(\epsilon).$$

$ApEn$ mede a probabilidade de que padrões que estejam próximos permaneçam próximos ao incremento de m . Ou seja, $P(\text{padrões próximos em } m + 1 | \text{padrões próximos em } m)$. Considerando-se dois sistemas dinâmicos diferentes A e B, pode-se calcular essa probabilidade de ambos obtendo-se $ApEn_A(m, \epsilon, N)$ e $ApEn_B(m, \epsilon, N)$ para determinados m , ϵ e N . A motivação para o uso da $ApEn$ está na intuição de que se essas probabilidades têm valores distintos, então a distribuição de probabilidade marginal dos sistemas (considerando todos os m e ϵ e $N \rightarrow \infty$) deve também ser diferente. A vantagem desse procedimento é que são necessários muito menos pontos para se estimar a probabilidade condicional do que para se estimar a marginal. Como colocado em Pincus [1995], não é necessário reconstruir totalmente duas trajetórias no espaço de fases (ou dois atratores) para se distinguir uma do outra.

A Figura 3.6 mostra esse índice para o mapa logístico com valores do parâmetro de bifurcação λ tomados entre 3,5 e 4.

É de se esperar que, nesse intervalo de parâmetro, a complexidade do mapa (considerando-se sua topologia no espaço de fases) seja proporcional a λ . Um valor baixo de $ApEn$ indica uma maior regularidade, ou seja, menor complexidade da série. Para $\lambda = 3,5$, por exemplo, a série tem comportamento periódico, e portanto, elevada

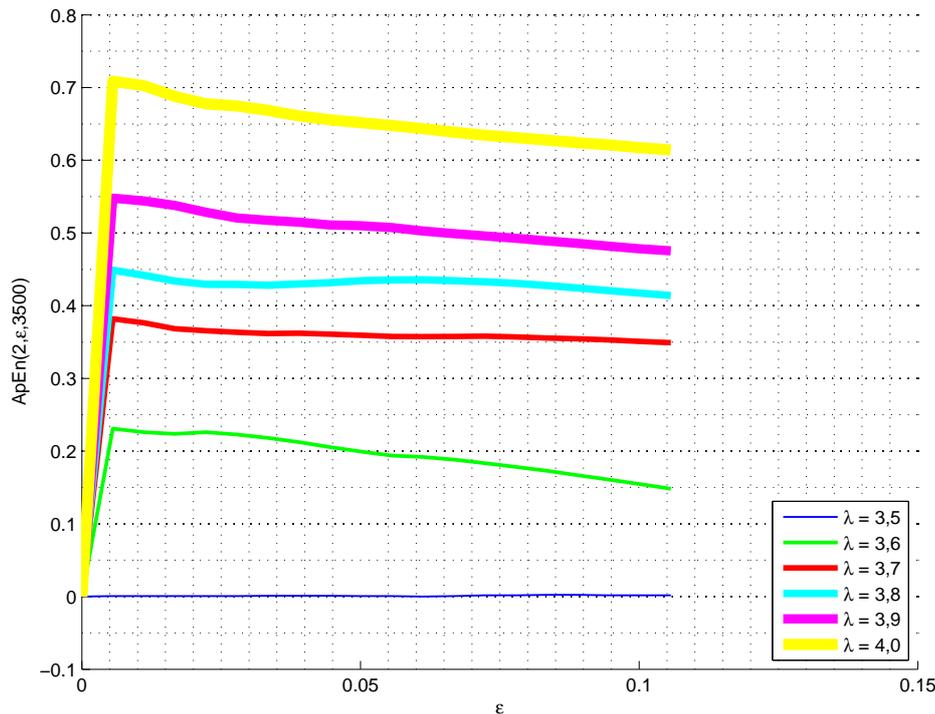


Figura 3.6. Entropia aproximada $ApEn(2, \epsilon, 3500)$ para mapas logísticos com valores do parâmetro de bifurcação λ tomados entre 3,5 e 4. ϵ é variado de 0 a aproximadamente 30% do valor do desvio padrão dessas séries. Observa-se que, para o intervalo de ϵ mostrado (excetuando-se o primeiro valor), a $ApEn$ consegue ordenar os mapas por complexidade (considerando que λ seja proporcional à complexidade, como em Pincus [1991]).

regularidade. Isso é refletido em um valor quase nulo de entropia. Percebe-se que a $ApEn$, para esse caso, consegue ordenar as séries de acordo com sua complexidade em todos os valores de ϵ calculados, como em Pincus [1991].

Na figura, ϵ é variado de 0 a aproximadamente 30% do valor do desvio padrão da série. O resultado da entropia para ϵ próximo de zero é também próximo de zero para todas as séries. Pode-se perceber isso pela Equação 3.5, uma vez que não há pontos a uma distância menor que zero além do próprio ponto i . Segundo Pincus [1991], deve-se utilizar valores de ϵ próximos de 10 a 20 % do desvio padrão da série. Como esse desvio padrão está próximo de 0,35, utilizou-se valores de ϵ entre 0,035 e 0,07 para ilustrar o cálculo da entropia aproximada.

3.3.2 Entropia Amostral

Uma limitação da entropia aproximada é considerar no cálculo de $C_{m,i}(\epsilon)$, na Equação 3.5, o caso $i = j$. Ou seja, considera-se que o próprio vetor seja próximo (distância

menor que ϵ) a ele mesmo, uma vez que $|\mathbf{x}_i - \mathbf{x}_j|_\infty$ é sempre menor que ϵ para $i = j$. Em Richman & Moorman [2000] esses vetores são chamados *self-matches* e fazem com que a estimativa da $ApEn(m, \epsilon)$ seja enviesada para N finito. Não se deve considerar a similaridade de um vetor com ele próprio como uma indicação de regularidade. Portanto, as estimativas da entropia aproximada tendem a indicar mais regularidade do que realmente há na série. Isso, porém, evita que $C_{m,i}$ seja nulo e tenha-se na Equação 3.6 o cálculo de $\log(0)$. Esse viés é, então, uma maneira de se garantir que $\Phi_m(\epsilon)$ seja sempre definido.

Richman (Richman & Moorman [2000]) contorna o problema do viés propondo uma família de estatísticas semelhante a $ApEn$ dada por:

$$SampEn(m, \epsilon) = \lim_{N \rightarrow \infty} -\log(C_{m+1}(\epsilon)/C_m(\epsilon)),$$

em que N é o tamanho da amostra. Essa fórmula é implementada para determinado N como:

$$SampEn(m, \epsilon, N) = -\log(C_{m+1}(\epsilon)/C_m(\epsilon)).$$

C_m é a estimativa da integral de correlação descrita na Seção 3.2. Nesse desenvolvimento, pode-se entender C_m como um cálculo de Φ_m , porém, sem o logaritmo e sem considerar os *self-matches* no número de vetores próximos. Isso resolve a questão do viés da entropia aproximada. Substituiu-se o somatório dos logaritmos da $ApEn$ pelo logaritmo da soma na $SampEn$.

A Figura 3.7 mostra esse índice para o mapa logístico com valores do parâmetro de bifurcação λ tomados entre 3,5 e 4, como feito para a entropia aproximada.

Um valor elevado de $SampEn$ também indica uma maior regularidade. Essa estatística consegue ordenar as séries da figura por complexidade para todos os valores de ϵ calculados. Para pequenos valores de ϵ , enquanto o valor da $ApEn$ tende a zero, o valor da $SampEn$ passa a não ser definido. Isso acontece porque, para ϵ suficientemente pequeno, a soma de todos os $C_{m,i}$ será nula o que resultará em $-\log(0)$. Porém, para os valores de ϵ próximos a 20% do desvio padrão, como recomendado em Richman & Moorman [2000], a entropia amostral é uma ferramenta bastante útil para se quantificar complexidade de séries. Os primeiros valores definidos de $SampEn$ são maiores que os correspondentes de $ApEn$ sugerindo o viés de $ApEn$, conforme explicado anteriormente. Como essa entropia é muito semelhante a anterior, e além disso, contorna o problema do viés, para o restante do trabalho apenas $SampEn$ será utilizada.

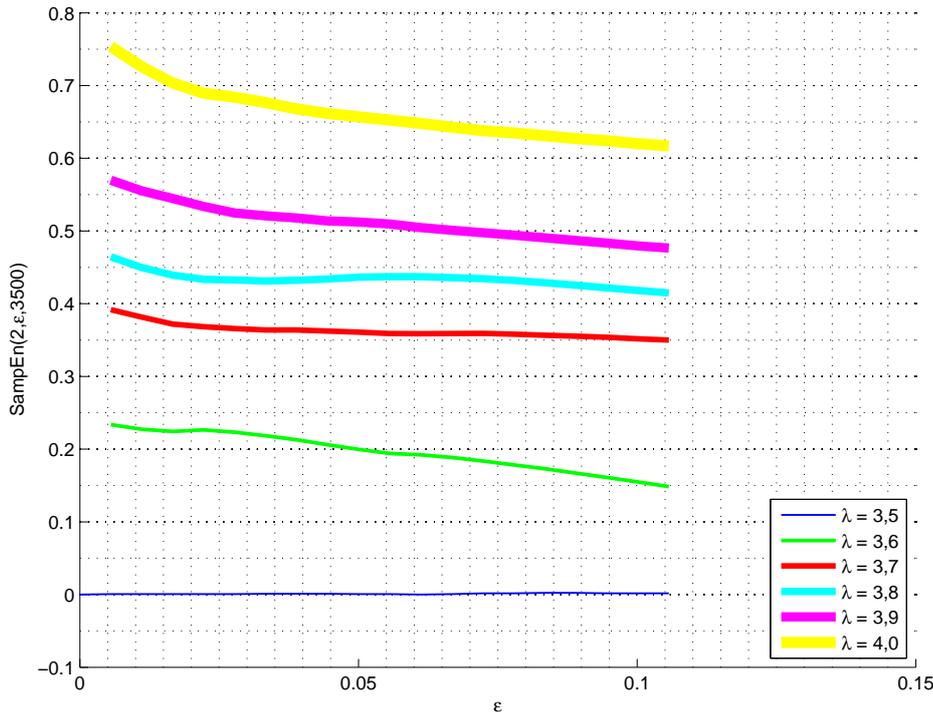


Figura 3.7. Entropia amostral $SampEn(2, \epsilon, 3500)$ para mapas logísticos com valores do parâmetro de bifurcação λ entre 3,5 e 4. ϵ é variado de 0 a aproximadamente 30% do valor do desvio padrão dessas séries (mesmos valores utilizados em $ApEn$). Para todo o intervalo de ϵ calculado, a $SampEn$ consegue ordenar os mapas por complexidade (considerando que λ seja proporcional a complexidade).

3.4 Informação Mútua Média

A entropia $H(X)$, como definida em Cover [2006], é a medida da incerteza de uma variável aleatória e também é conhecida como autoinformação. Para uma variável aleatória X , $H(X)$ é dada por:

$$H(X) = - \sum_{x \in X} p(x) \log(p(x)),$$

em que $p(x)$ é a probabilidade $p(X = x)$. A entropia, no contexto da teoria da informação, geralmente está associada à quantidade de informação que uma mensagem carrega, dado um determinado universo (alfabeto) de possíveis valores. $H(X)$, porém, não depende dos valores de X e sim de sua distribuição de probabilidades. Mas pode-se perceber, por exemplo, que um evento com probabilidade 1 contribui com uma informação nula, o que é intuitivo. Se o conteúdo de uma mensagem é sabido de antemão, não há ganho de informação ao se receber essa mensagem.

Pode-se definir, também, a entropia conjunta por:

$$H(X; Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log(p(x, y)).$$

A informação mútua média $I(X; Y)$ é definida como a entropia relativa entre a distribuição conjunta e o produto das distribuições de X e Y :

$$I(X; Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right).$$

Para que o conceito de $I(X; Y)$ se torne mais natural, parte-se da definição de entropia condicional, dada por:

$$H(Y|X) = - \sum_{x \in X} p(x) H(Y|X = x) = -E[\log(p(Y|X))].$$

A entropia condicional é o valor esperado da entropia conjunta ponderado pela distribuição de X . Pode ser mostrado, também, que $H(X; Y) = H(X) + H(Y|X)$. De maneira semelhante, $I(X; Y)$ pode ser escrita como:

$$I(X; Y) = H(Y) - H(Y|X). \quad (3.8)$$

Logo, a informação mútua média (AMI) é a redução da incerteza de Y dada a ocorrência de X . A AMI também pode ser entendida como a quantidade de informação que uma variável aleatória possui a respeito de outra variável. No caso de uma série temporal $y(t)$, tem-se interesse na informação que $y(t - \tau)$ possui a respeito de $y(t)$ para τ positivo. Ou seja, pretende-se calcular $I(y(t); y(t - \tau))$. É fácil perceber que se $y(t - \tau)$ e $y(t)$ são independentes, a AMI $I(y(t - \tau); y(t))$ é nula. A AMI não apresenta apenas a informação a respeito da dependência linear, como a função de autocorrelação, mas considera todos os momentos. Isso acontece porque a AMI não depende dessa correlação, mas da distribuição que deu origem aos dados. No entanto, apenas para ilustrar o uso da AMI, são utilizadas séries autorregressivas cujos valores são mostrados na Figura 3.8.

Essas séries AR(1) possuem 3500 valores e são da forma mostrada na Equação 2.9. A AMI é calculada para sete séries com valores de θ entre 0,2 e 0,7 para valores de atraso até dez observações ($\tau = 1$ a $\tau = 10$). Por essa figura, percebe-se que esse parâmetro é proporcional a AMI. Essa constatação é coerente com a Equação 2.9 e com o conceito atribuído à AMI. Um aumento de θ (maior dependência linear) faz com que a informação de valores anteriores da série a respeito do valor atual seja maior.

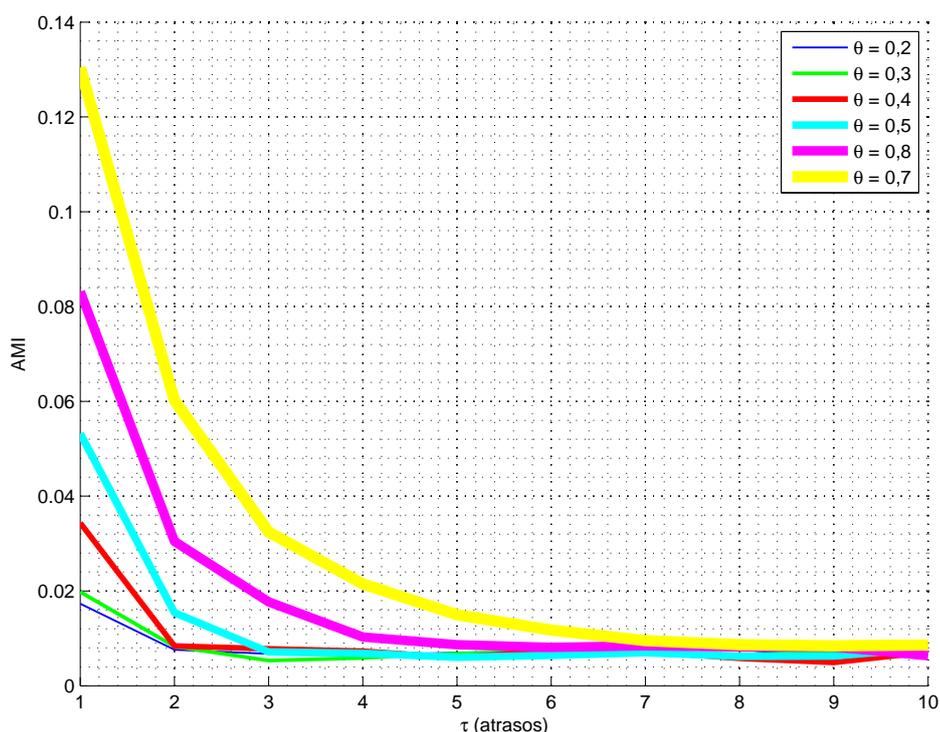


Figura 3.8. Informação mútua média (AMI) calculada para séries AR(1) (Equação 2.9) com valores de θ entre 0,2 e 0,7. Valores maiores desse parâmetro fornecem uma maior AMI. Esse resultado é esperado uma vez que um aumento de θ realmente faz com que a informação sobre o instante atual contida em instantes anteriores seja maior.

Um exemplo da utilização da AMI em dados financeiros pode ser vista em Darbellay & Wuertz [2000]. Os autores comparam a AMI dos retornos e seus módulos no mercado de câmbio, mas não comentam a respeito da validade estatística de seus cálculos. Evidentemente, o problema da estimação da AMI é o mesmo da estimação correta de distribuições baseada em amostras. Necessita-se de um grande número de dados e os mesmos devem ser representativos da distribuição conjunta. O tamanho da amostra N de cada série estudada é da ordem de milhares de observações. O resultado da estimação da AMI, porém, ganha bastante significância com o uso de dados sub-rogados. Com isso, tenta-se contornar o problema da quantidade de observações.

No presente trabalho, a AMI é utilizada considerando-se valores de τ de 1 a 10.

3.5 Teste BDS

O teste desenvolvido por W.A. Brock, W. Dechert e J. Scheinkman em 1987, conhecido como BDS, é um teste de independência usado em resíduos de séries temporais. Como

mostrado em Brock *et al.* [1996] o teste foi criado para testar a hipótese de que uma série de resíduos seja independente e identicamente distribuída (i.i.d.). O BDS é utilizado como critério de seleção de modelos e em testes de hipóteses. Como no caso do teste Ljung-Box, o BDS pode ser aplicado aos resíduos de um modelo linear ARMA da série. A diferença está no fato de que o primeiro teste verifica apenas a dependência linear, enquanto o segundo se propõe a identificar não linearidades dos resíduos. Testes como Breusch-Pagan e White, descritos na Seção 2.4 têm como limitação a necessidade de se saber *a priori* o tipo de não linearidade dos dados. Além disso, esses testes dependem do tipo de modelo não linear estimado. A Seção 2.4 mostra a necessidade de se modelar corretamente a variância a fim de se obter resultados satisfatórios. Ou seja, a correta detecção de heteroscedasticidade baseada em White e Breusch-Pagan depende de um modelo para a variância, que é, exatamente, o que não se conhece *a priori*. Nesse caso, a abordagem do teste BDS é mais flexível, por que indica (em tese) a presença de qualquer estrutura, linear ou não, presente nos dados. Supondo-se que a dependência temporal da variância seja a única presente, o teste BDS indicará a rejeição da hipótese de independência, não sendo necessário tal modelo.

O teste é construído por meio de uma estatística baseada na integral de correlação $C(\epsilon)$. A Equação 3.2 define essa integral por meio de uma probabilidade calculada em função da distância máxima ϵ . A Seção 3.2 cita o fato de o ruído i.i.d. ter dimensão de correlação infinita. Isso poderia ser visto no gráfico $\log(\epsilon) \times \log(C(\epsilon))$ por uma inclinação igual à dimensão de imersão m , para qualquer m . Na prática, essa relação é aproximada, mas ainda assim pode-se perceber que $C(\epsilon)$ estará relacionada a m . A Figura 3.9 exhibe esse gráfico para a série WGN, um ruído branco gaussiano.

Para cada valor de m , a curva $\log(C(\epsilon))$ se torna mais íngreme na *scaling range*. Pode-se considerar a integral de correlação como uma função de m , ou $C_m(\epsilon)$.

Brock, Dechert e Scheinkman mostraram que, para séries i.i.d., é válida a seguinte relação:

$$C_m(\epsilon) = C_1(\epsilon)^m. \quad (3.9)$$

Ou seja, a integral de correlação em qualquer dimensão m pode ser obtida por meio de uma função exponencial da mesma integral em dimensão 1. A Figura 3.10 mostra a comparação entre $C_m(\epsilon)$ e $C_1(\epsilon)^m$ para valores de m entre 2 e 5 para a série WGN. Não se consegue, na figura, distinguir os termos da Equação 3.9, o que está de acordo com as observações de Brock, Dechert e Scheinkman.

Nos casos em que a série não é i.i.d., a Equação 3.9 não se aplica. Entretanto, pode-se ainda estimar $C_m(\epsilon)$ e $C_1(\epsilon)$. A subtração do primeiro pelo segundo termo da

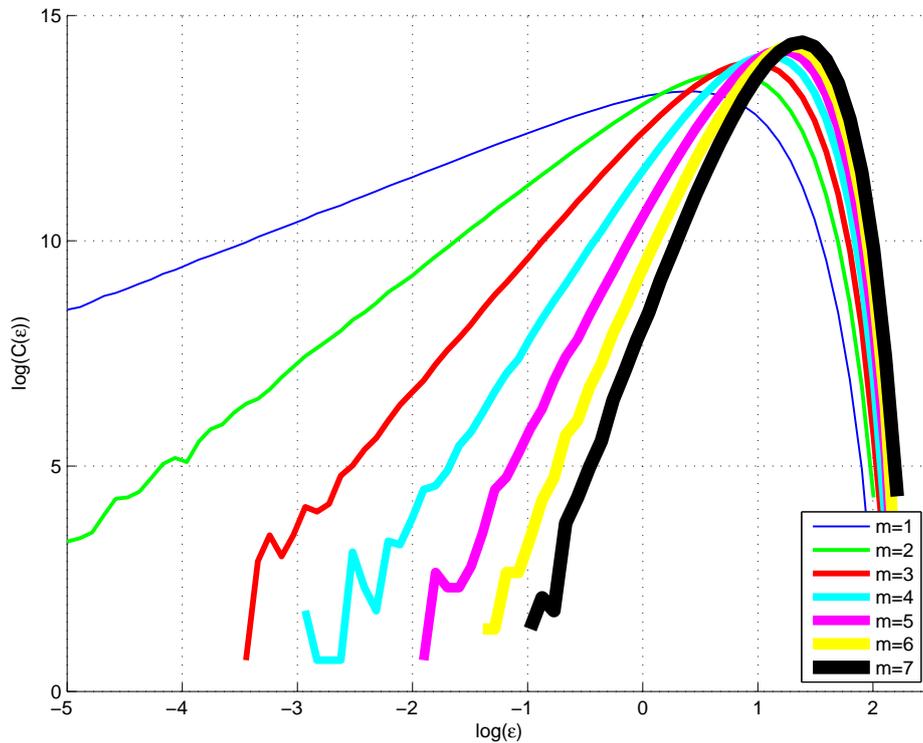


Figura 3.9. Gráfico do logaritmo da integral de correlação $C(\epsilon)$ em função do logaritmo da distância ϵ para a série WGN. A parte linear do gráfico se torna mais inclinada à medida que a dimensão de imersão m aumenta. Os pontos que aparecem são em função das descontinuidades

equação, caso seja nula, indicará a violação da hipótese de independência. Isso leva ao estudo das propriedades de:

$$T_m(\epsilon) = C_m(\epsilon) - C_1(\epsilon)^m. \quad (3.10)$$

Em Brock *et al.* [1996], é demonstrado que, se a série for independente, a distribuição de $T_m(\epsilon)$ é gaussiana com média zero e variância³ $V_m(\epsilon)$. Essa é a definição da estatística BDS:

$$W_m(\epsilon) = \sqrt{N} \frac{T_m(\epsilon)}{V_m(\epsilon)}, \quad (3.11)$$

em que N é a quantidade de observações. A estatística de teste BDS é definida como a diferença normalizada entre integral de correlação da série imersa em m e a mesma integral imersa em uma dimensão unitária. Consequentemente, se o módulo de $W_m(\epsilon)$ for maior que 1,96 (valor correspondente a um intervalo de confiança de $\alpha = 0,05$ para $N(0,1)$) pode-se rejeitar a hipótese de independência com 95% de certeza.

A Tabela 3.1 mostra os resultados do teste para as séries WGN, AR(1), ARCH(1),

³A dedução e a fórmula dessa variância está na referência indicada.

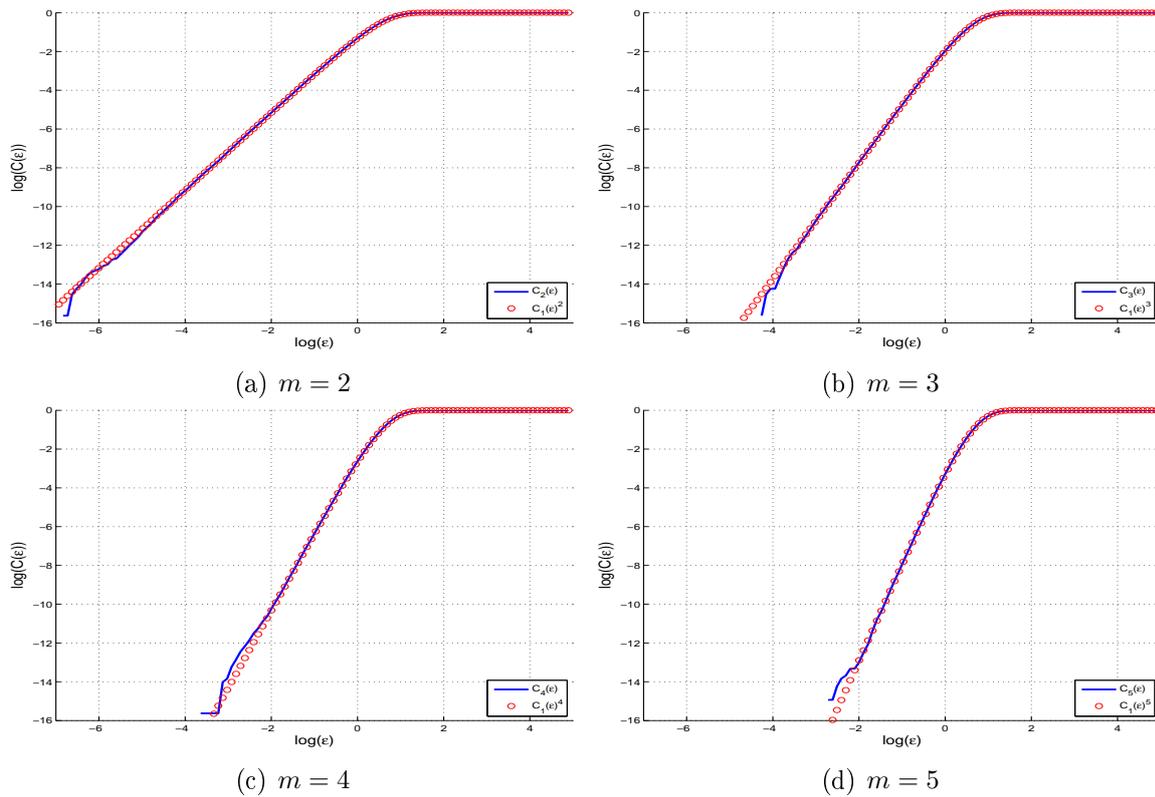


Figura 3.10. Comparação entre $C_m(\epsilon)$ e $C_1(\epsilon)^m$ para valores de $m = 2$ (a), $m = 3$ (b), $m = 4$ (c) e $m = 5$ (d) para a série WGN.

GARCH(1,1) e LOGMAP considerando $m = 2$. Esse será o valor de m utilizado no restante do trabalho. Além disso, ϵ é escolhido como $1,5\sigma$ em que σ é o desvio padrão da série, como definido na implementação descrita em Kanzler [1999]. Como simplificação da notação, $W_m(\epsilon)$ será indicado como W_2 , em que o valor de ϵ fica implícito como $1,5\sigma$. Na realidade W_2 é uma estatística simétrica cujo valor acima de 1,96 ou abaixo de -1,96 rejeita a hipótese de independência (*two tailed test*). Todavia, os resultados são sempre exibidos como o módulo de W_2 para que seja simplificada a comparação entre esse valor e o limite.

Tabela 3.1. Estatística W_2 para várias séries conhecidas. Se W_2 é maior que seu valor limite, rejeita-se a hipótese nula de independência.

Série	$ W_2 $	Valor limite de $ W_2 $
WGN	0,931	1,96
AR(1)	11,6	1,96
ARCH(1)	14,3	1,96
GARCH(1,1)	15,0	1,96
LOGMAP	16,6	1,96

Pela tabela, pode-se rejeitar a hipótese de independência para todas as séries, exceto WGN. Essa é, realmente, a única série cujas observações são independentes fazendo com que o teste seja bem-sucedido em todas as séries artificiais aqui testadas. Comparando-se o resultado da Tabela 3.1 com a Tabela 2.1 ou ainda, com a Tabela 2.3, pode-se perceber a vantagem do Teste BDS sobre Ljung-Box ou Breusch-Pagan. Esses últimos dois testes não conseguem rejeitar a hipótese de dependência temporal de ARCH, GARCH e LOGMAP. Ljung-Box não rejeita essa hipótese por perceber apenas dependências lineares e Breusch-Pagan não rejeita por não especificar corretamente o tipo de não linearidade.

A Tabela 3.1 não mostra o resultado de resíduos e sim das próprias séries a fim de que seja testada a independência das mesmas. Em Caporale *et al.* [2004], os autores mostram que o teste pode ser usado para testar a independência de resíduos de modelos GARCH sem que a distribuição de $W_m(\epsilon)$ seja afetada. A referência Brooks & Henry [2000], porém, traz um caso em que a estatística BDS falha ao não rejeitar a hipótese de independência para modelos GARCH assimétricos⁴ (que dependem do sinal ou magnitude das inovações). Essa falha, porém, não invalida o uso da estatística, mas apenas chama atenção para o fato de que podem existir outras hipóteses alternativas não modeladas e presentes nos ruídos a serem testados. Portanto, o teste BDS não deve ser considerado um teste geral para uma classe de modelos (por exemplo, a classe de todos os modelos lineares), mas para um modelo específico (como um modelo AR com parâmetros específicos). No caso de modelos GARCH, o teste BDS aplicado a um desses modelos não testa a presença de heteroscedasticidade de uma maneira geral, mas simplesmente indica que se os resíduos possuem alguma dependência. Generalizar esse resultado considerando-o como teste de heteroscedasticidade é uma extrapolação da capacidade do teste. Entretanto, essa extrapolação (mesmo não apresentando um rigor matemático ou estatístico) pode ocorrer, desde que haja confiança na representatividade e capacidade de generalização do modelo dos resíduos.

No restante do trabalho, o teste BDS será indicado pela estatística W_2 . Para as séries financeiras (Capítulo 4), o teste é empregado nos resíduos de modelos AR, GARCH e AR-GARCH além da própria série. Isso é feito a fim de investigar as seguintes hipóteses:

1. teste da própria série: há dependência temporal?
2. teste em resíduos de um modelo AR: se existir, a dependência é apenas linear?

⁴Tais modelos são semelhantes aos modelos Threshold Autoregressive (Tsay [2005]), mas são aplicados ao processo da variância.

3. teste em resíduos de um modelo AR-GARCH: se não for apenas linear, é devida a heteroscedasticidade?

Um resultado negativo para última pergunta indica uma possível presença de determinismo não linear previsível. Para quantificar o determinismo linear e não linear desconsiderando-se a heteroscedasticidade, calcula-se a estatística W_2 também em resíduos de um modelo GARCH das séries. Além disso, calcula-se o índice PREV, de W_2 , descrito na Seção 3.6.

3.6 Dados Sub-rogados

Se uma série é determinística, ou tem um componente determinístico, isso pode ser quantificado pelos índices mostrados anteriormente. Um dos problemas desses índices é a incerteza associada a suas estimativas. Como as séries a serem analisadas são estocásticas, os valores estimados também serão. Isso compromete a utilização direta de índices como, por exemplo, o de dimensão de correlação estimada pelo algoritmo de Grassberger e Procaccia. Nesse caso, a precisão da dimensão depende da existência de um modelo para sua distribuição de probabilidades, o que é difícil de se obter. No caso de séries financeiras, não é possível gerar mais dados que deem a confiança estatística necessária aos índices. Seus resultados, portanto, quando tratados de maneira isolada trazem pouca informação. Ou seja, não se pode discernir, pelo número gerado para o índice, se esse realmente indica determinismo ou é simplesmente um ruído. Outra questão importante é a origem do determinismo. A princípio, não se sabe se o valor do índice indica a presença de correlações lineares, não estacionariedade (como a heteroscedasticidade) ou algum determinismo não linear (e de qual tipo).

Nesse contexto, a técnica de dados sub-rogados (Theiler *et al.* [1992]) apresenta como alternativa a geração de outras séries baseadas na única série disponível. Essas têm características bastante semelhantes às da original e servem de comparação com a mesma. Pode-se assim rejeitar ou não uma hipótese a respeito da série original comparando-a com as séries geradas. Os testes de hipóteses baseados em dados sub-rogados originam-se na geração de séries que tenham quase as mesmas características estatísticas que a série original. Por quase, entende-se que são preservadas as características suficientes para que se possa dizer que são séries semelhantes, mas por construção, sejam coerentes com alguma hipótese nula. As séries substitutas (ou dados sub-rogados) são construídas utilizando-se a série original mas permitindo que alguma característica dessa seja aleatória. Pode-se, por exemplo, testar a dependência temporal gerando várias séries cujos elementos são os mesmos da série original mas em

ordem aleatória (algoritmo 0). Como os elementos são os mesmos, preserva-se todas as características relativas à distribuição dos dados, como sua média e variância. Todavia, a dependência temporal é destruída. Pode-se, então, calcular estatísticas como AMI, *SampEn* ou D_2 de todas as séries sub-rogadas. Obtém-se, em seguida, a distribuição de probabilidades dessas estatísticas e compara-se o valor da estatística da série original com o das séries sub-rogadas. Se o valor for muito diferente, rejeita-se, nesse exemplo, a hipótese de independência. Percebe-se, então, que há pelo menos duas fases distintas nesses testes. A primeira é o processo de geração de séries sub-rogadas, cujo método dependerá fundamentalmente da hipótese a ser testada. A segunda fase é a escolha das estatísticas discriminantes e comparação das séries sub-rogadas com a original em função dessas estatísticas. Theiler e colegas citam como possíveis escolhas para as estatísticas a estimação de expoentes de Lyapunov, erros de previsão (de um modelo linear local baseado nos pontos disponíveis da estrutura do atrator) e a dimensão de correlação (Theiler *et al.* [1992]). No presente trabalho, serão utilizados como estatísticas discriminantes a dimensão de correlação D_2 , calculada utilizando-se o algoritmo de Judd, a entropia amostral *SampEn* e a informação mútua média AMI.

Os três algoritmos mais comuns de geração de sub-rogados são os algoritmos 0, 1 e 2, como descrito em Theiler *et al.* [1992]. Utiliza-se também neste trabalho o algoritmo PPS a fim de se detectar não linearidades que não sejam devidas a heteroscedasticidade. A descrição e aplicação dos quatro algoritmos pode ser vista em Small [2005].

Cada algoritmo testa uma hipótese diferente e mais específica que a anterior, formando assim uma hierarquia de testes. Os testes são feitos primeiramente para se testar a independência temporal dos dados. Para isso, o algoritmo 0 gera séries substitutas permutando a ordem temporal da original de maneira aleatória. Tem-se assim séries que apresentam a mesma distribuição. Porém, qualquer potencial correlação originalmente existente na série em estudo é destruída nas séries sub-rogadas. Se a hipótese de independência puder ser rejeitada, testa-se a hipótese de que os dados são apenas linearmente correlacionados e têm incrementos gaussianos. O algoritmo 1 faz isso gerando séries cuja função de autocorrelação seja a mesma da original, porém sem correlações não lineares. Posteriormente, pode-se testar se a hipótese anterior é válida para incrementos não gaussianos. Isso é realizado por meio do algoritmo 2, que é uma generalização do algoritmo 1, fazendo com que os sub-rogados tenham também a mesma distribuição da série original. Se essa última hipótese for rejeitada, há dependência temporal não linear nos dados ou não estacionariedade. Em muitas séries, entretanto, essa dependência ou a falta de estacionariedade é óbvia. Esse é o caso de séries periódicas ou pseudo-periódicas e séries heteroscedásticas, respectivamente. Nes-

ses casos a rejeição de hipóteses dos algoritmos 0, 1 e 2 não traz muitas informações a respeito de outras dependências não lineares possivelmente existentes nos dados. Pode-se então utilizar o algoritmo PPS para testar se há nos dados alguma não linearidade que não seja devida a oscilações ou dependências heteroscedásticas. Esse algoritmo gera dados que possuam as mesmas não linearidades (devidas a oscilações) que a série original. São conservados nesse último caso, dependências como não linearidades periódicas ou a presença de heteroscedasticidade. Se uma série for rejeitada em todas as hipóteses anteriores, pode-se concluir que há dependências temporais não lineares que, em tese, poderiam ser previstas.

Chega-se, finalmente, a um ponto em que se pode estimar a previsibilidade das séries. Isso pode ser feito, como em Theiler *et al.* [1992] comparando-se a média dos índices das séries sub-rogadas e o índice da série original. Para tanto, pode ser usado o módulo da diferença desses índices discriminantes normalizados pelo seu desvio padrão:

$$PREV(\text{algoritmo}, \text{estatística}, \text{série}) = \frac{|I_D - \mu_H|}{\sigma_H}, \quad (3.12)$$

sendo I_D , o índice ou estatística utilizado no teste calculado para a série original. μ_H e σ_H são respectivamente, a média e o desvio padrão dos índices das séries sub-rogadas. Esse cálculo será realizado, em alguns algoritmos, para o primeiro atraso ($\tau = 1$) da estatística AMI e representa a distância, em desvios padrão, da série em relação aos sub-rogados. Pode-se comparar $PREV$ como definido anteriormente, dentro de cada algoritmo. Porém não há sentido em se comparar esse valor para séries geradas em diferentes algoritmos. Esse índice de previsibilidade quantifica a rejeição da hipótese de determinado algoritmo. Como a rejeição indica uma maior possibilidade de previsão, pode-se interpretar um elevado valor desse índice como uma forte propensão da série em ter um componente determinístico e previsível condizente com a hipótese rejeitada.

3.6.1 Algoritmo 0

São gerados conjuntos de dados i.i.d. com a mesma distribuição de probabilidades da série original. Isso é feito permutando-se os dados da série original de maneira aleatória, o que equivale a uma reamostragem dos dados originais sem reposição. Consegue-se dessa forma destruir qualquer dependência temporal da série mantendo-se sua distribuição. Conservam-se, conseqüentemente, a média, a variância e outros momentos de ordem superior, se existirem. Esse algoritmo é bastante semelhante à técnica estatística conhecida como *bootstrapping*. Se as estatísticas discriminantes utilizadas forem capazes de distinguir entre a série e seus sub-rogados, rejeita-se a hipótese nula de que

a série testada é i.i.d.. Caso contrário, não se pode afirmar que os dados sejam i.i.d.. É possível que sejam independentes e que as estatísticas utilizadas não tenham sido capazes de distinguir a mesma de seus sub-rogados. Esse raciocínio é válido para todos os testes: pode-se negar uma hipótese, mas não se pode provar a aderência a outra.

As Figuras 3.11, 3.12 e 3.13 mostram a aplicação do algoritmo 0 para a geração de dados sub-rogados das séries WGN e AR(1).

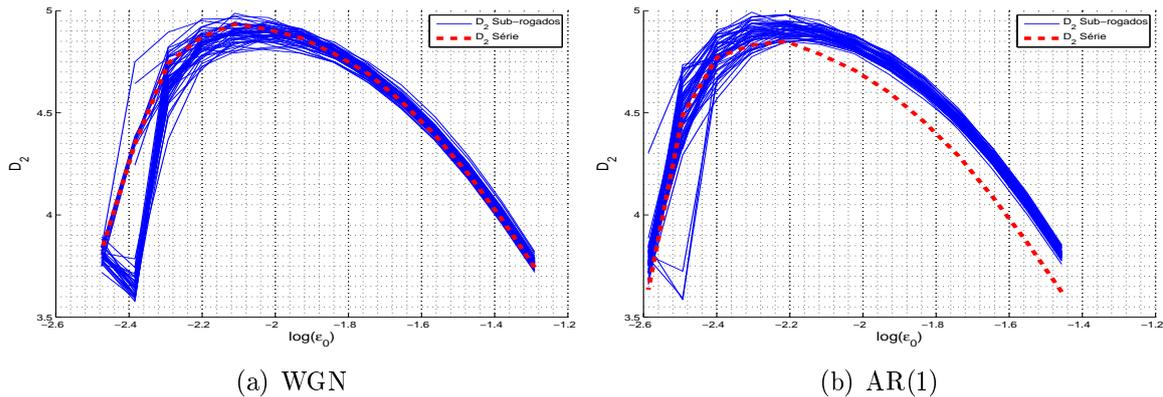


Figura 3.11. Algoritmo 0 - comparação de $D_2(\epsilon_0)$ entre as séries WGN e AR(1) e seus respectivos sub-rogados. A Figura (a) mostra essa comparação para a série WGN, enquanto a Figura (b) mostra o mesmo para série AR(1). Das duas séries, a segunda é a única cuja hipótese de independência pode ser rejeitada, uma vez que seu valor de $D_2(\epsilon_0)$ é claramente diferente dos de seus sub-rogados para vários ϵ_0 . O mesmo não se pode dizer da série WGN, cujo $D_2(\epsilon)$ encontra-se misturado entre $D_2(\epsilon)$ de seus sub-rogados.

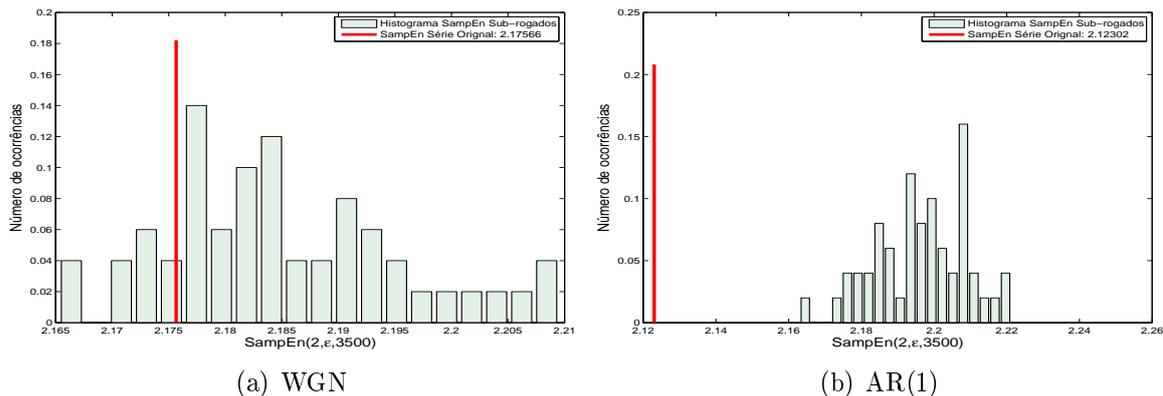


Figura 3.12. Algoritmo 0 - comparação de $SampEn(2, \epsilon, 3500)$ entre as séries WGN (a) e AR(1) (b) e seus respectivos sub-rogados. A segunda é a série cuja hipótese de independência pode ser rejeitada. Seu valor de $SampEn(2, \epsilon, 3500)$ encontra-se afastado dos de seus sub-rogados. A $SampEn(2, \epsilon, 3500)$ de WGN encontra-se misturado entre os intervalos da distribuição de valores para seus sub-rogados.

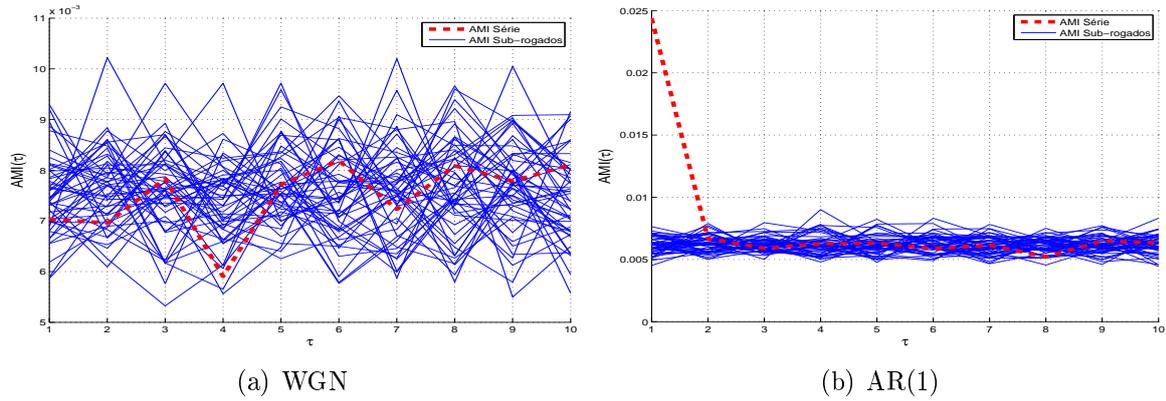


Figura 3.13. Algoritmo 0 - comparação da estatística AMI entre as séries WGN e AR(1) e seus respectivos sub-rogados. (a) mostra WGN e seus sub-rogados e (b) mostra o mesmo para série AR(1). Por meio da estatística AMI, pode-se rejeitar a hipótese de independência de AR(1), uma vez que seu valor é claramente diferente do de seus sub-rogados em $\tau = 1$. O mesmo não se pode dizer da série WGN, cuja AMI confunde-se entre os valores dessa estatística para seus sub-rogados. Esse resultado é esperado pois a dependência temporal mais significativa de AR(1) é exatamente no primeiro atraso ($y(t - \tau)$ para $\tau = 1$).

São geradas 50 séries sub-rogadas⁵. A dimensão de correlação, D_2 , é estimada considerando uma dimensão de imersão igual a cinco e vários valores do logaritmo de ϵ_0 . A entropia amostral $SampEn(2, \epsilon, 3500)$ é calculada para ϵ igual a 20% do desvio padrão da série e mostrada por meio de um histograma normalizado para que os 20 intervalos somados sejam iguais à unidade. A informação mútua média, AMI, é estimada para os 10 primeiros atrasos das séries (τ de 1 a 10). Para as três estatísticas discriminantes, pode-se rejeitar a hipótese de independência para a série AR(1), cujo coeficiente autoregressivo é 0,3. Já a série WGN, é sabidamente i.i.d. e o resultado da análise das figuras é coerente com esse fato. A Tabela 3.2 mostra os resultados do índice de previsibilidade dessas séries para o algoritmo 0, quantificando o que é observado nas figuras anteriores.

Tabela 3.2. Índice de previsibilidade (Equação 3.12) das séries WGN e AR(1) para o teste de independência do algoritmo 0.

Série	Índice de previsibilidade $PREV(\text{Alg. 0, AMI, série})$
WGN	0,594
AR(1)	26,5

Poderia-se questionar se séries com dependência de longo prazo, como as fra-

⁵Julga-se que 50 séries sub-rogadas são suficientes para se utilizar essa análise, uma vez que alguns trabalhos relevantes no assunto, como Small [2005], também utilizam essa quantidade.

cionalmente integradas (Campbell *et al.* [1997]), apresentariam uma não rejeição da hipótese de independência. Campbell faz a comparação entre um processo AR(1) e um processo fracionalmente integrado com mesma correlação linear no primeiro atraso (Campbell *et al.* [1997]). O autor mostra que a magnitude da correlação linear desse processo fracionalmente integrado no atraso 100 é maior que a do atraso 4 do processo AR(1). Como pode ser percebido nos dados colocados por Campbell, esses processos seriam rejeitados como independentes já que apresentam considerável correlação linear (maior do que o processo AR(1)). Poderia-se pensar em eliminar as dependências de curto prazo desses processos (processos com dependências apenas de longo prazo) e investigar o poder de rejeição dos testes de independência aqui utilizados nesses casos. Essa investigação, porém, foge do escopo deste trabalho e é deixado apenas como uma interessante sugestão para estudos futuros.

3.6.2 Algoritmo 1

O próximo teste verifica a hipótese de que a série seja um ruído branco gaussiano filtrado linearmente ou, equivalentemente, um processo ARMA. Considerando-se que a série analisada não seja i.i.d. (rejeição do algoritmo 0), a dependência temporal pode ser apenas linear. Nesse caso bastaria uma modelagem por meio de uma estrutura ARMA para capturar toda a dinâmica do processo. Quando suspeita-se da presença de não linearidades pode-se lançar mão da geração de dados sub-rogados utilizando o algoritmo 1. A comparação entre a série e seus sub-rogados gerados por esse algoritmo pode rejeitar a hipótese de que a série seja linear gaussiana e aponta para dependências não lineares dos dados, que não conseguiriam ser capturados por modelos ARMA. A geração de dados sub-rogados é feita tornando-se aleatórias as fases da transformada de Fourier da série. Uma das maneiras de se alcançar essa aleatoriedade é:

1. obter a transformada de Fourier da série;
2. somar a cada fase uma variável aleatória uniforme ($U(0,2\pi)$). O número complexo, resultado da transformada pode ser visto por meio de uma amplitude A e uma fase ϕ , como $A \exp(\phi)$. Multiplica-se, então, esse número por $1 \exp(\phi_u)$ em que ϕ_u é a variável aleatória uniforme. Assim a amplitude se mantém e a fase se torna aleatória;
3. obter a transformada inversa de Fourier e considerar como sub-rogado a parte real resultante.

Esse procedimento gera sub-rogados com a mesma sequência de correlação e mesma densidade espectral da série original. Depois de obtida a transformada de Fourier, a amplitude da transformada é preservada, mas sua fase se torna aleatória. O quadrado dessa amplitude é um estimador do periodograma e de sua densidade espectral. A função de auto-correlação, que só depende da amplitude, se conserva. Entretanto, qualquer dinâmica não linear que possa existir nos dados é destruída com esse processo. Os dados gerados são, portanto, lineares e têm a mesma dependência (linear) que a série em estudo. As estatísticas de teste, nesse caso, devem ser capazes de capturar dinâmicas não lineares que podem estar presentes na série e que certamente não estão nos sub-rogados. Percebe-se que, por esse motivo, não é possível usar como estatística lineares como $q(10)$ (teste Ljung-Box).

As Figuras 3.14, 3.15 e 3.16 mostram a aplicação do algoritmo 1 para a geração de dados sub-rogados das séries AR(1), MA(1) uniforme e ARCH(1). A série MA(1) uniforme é extraída de um processo MA(1) com parâmetro de média móvel da inovação igual a 0,3. Essa inovações não são gaussianas, mas uniformes, o que leva a série a estar em desacordo com a hipótese nula de linearidade gaussiana.

São geradas, como no algoritmo 0, 50 séries sub-rogadas. A dimensão de correlação, D_2 , é estimada considerando uma dimensão de imersão igual a cinco e vários valores do logaritmo de ϵ_0 . A entropia amostral $SampEn(2, \epsilon, 3500)$ é calculada para ϵ igual a 20% do desvio padrão de cada série e mostrada por meio de um histograma normalizado para que os 20 intervalos do histograma somados sejam iguais à unidade. O valor de ϵ muda em cada série, uma vez que a distribuição das amostras não se conserva de série para série, como no caso do algoritmo 0. A informação mútua média, AMI, é estimada para os 10 primeiros atrasos das séries (τ de 1 a 10).

Como mostrado nessas figuras, AR(1), que não apresenta nenhuma dinâmica não linear, se confunde em meio aos seus sub-rogados. Não se pode rejeitar a hipótese, segundo essa análise, de que essa série é um processo ARMA com inovações gaussianas. Em contrapartida, a série ARCH(1) cuja estrutura não linear já foi detectada por meio do teste de White na seção 2.4, tem sua hipótese rejeitada nessa análise. Ou seja, a análise de dados sub-rogados consegue detectar corretamente a presença de não linearidades em ARCH(1). A série MA(1) uniforme tem sua hipótese nula de linearidade gaussiana rejeitada. Apesar de ser gerada por um processo linear, seus incrementos não são gaussianos, o que não condiz com a hipótese nula. Portanto, essa série tem sua hipótese nula corretamente rejeitada.

Esse resultado também pode ser visto na Tabela 3.3, na qual confirma-se numericamente os resultados encontrados nas figuras anteriores.

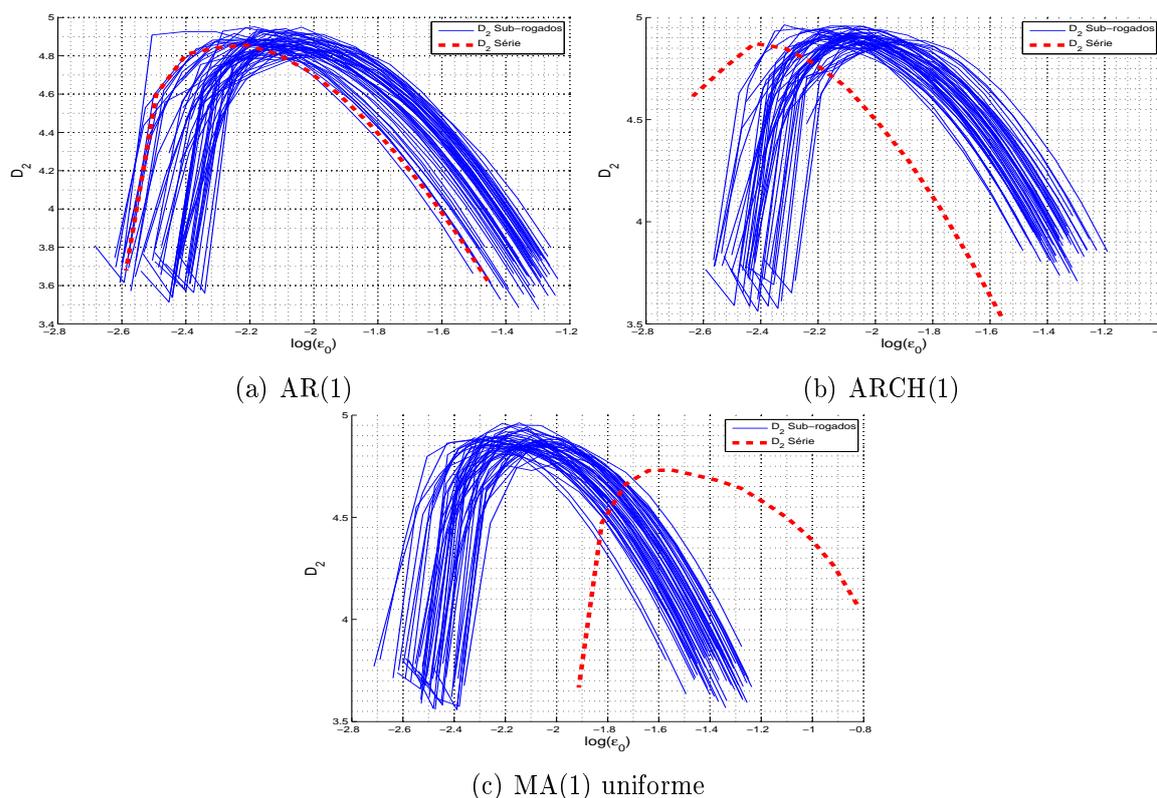


Figura 3.14. Algoritmo 1 - $D_2(\epsilon_0)$ calculado para AR(1), MA(1) uniforme e ARCH(1) e seus respectivos sub-rogados. No gráfico (a) a comparação entre a série AR(1) e seus sub-rogados, em (b) e em (c), o mesmo para as séries MA(1) uniforme e ARCH(1) respectivamente. O tracejado que indica $D_2(\epsilon_0)$ de AR(1) está bem próximo das linhas contínuas dos dados sub-rogados. Por outro lado, para ARCH(1), o tracejado tem valores bem distintos das linhas. Pode-se, então, rejeitar a hipótese de linearidade (com incrementos gaussianos) de ARCH(1).

Tabela 3.3. Índice de previsibilidade (Equação 3.12) das séries AR(1), MA(1) uniforme e ARCH(1) para o teste de linearidade do algoritmo 1.

Série	Índice de previsibilidade $PREV(\text{Alg. 1, AMI, série})$
AR(1)	0,521
MA(1) uniforme	10,4
ARCH(1)	7,01

3.6.3 Algoritmo 2

Uma das limitações do algoritmo anterior é o fato de os dados sub-rogados sempre terem uma distribuição gaussiana⁶. Uma possível mudança no algoritmo 1 é transformar os

⁶Os próprios dados sub-rogados são gaussianos. Os índices gerados por meio deles não necessariamente o serão.

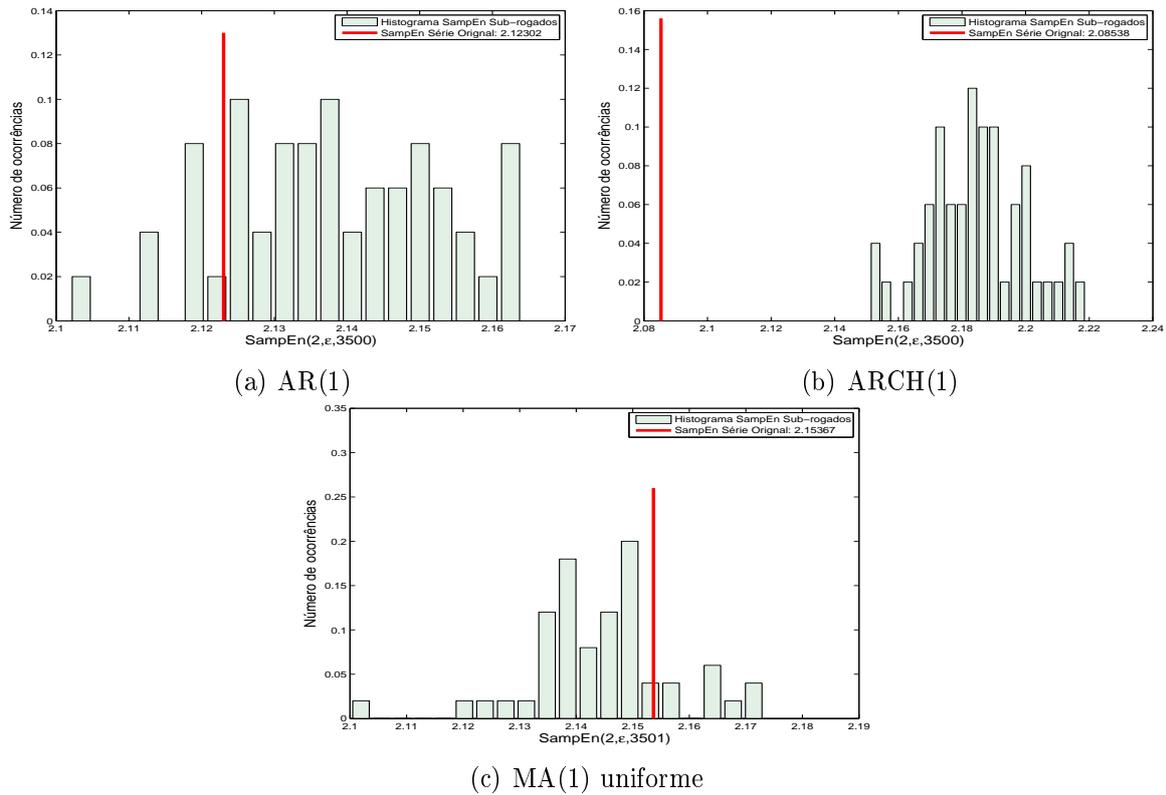


Figura 3.15. Algoritmo 1 - Estatística discriminante $SampEn(2, \epsilon, 3500)$ calculada para as séries AR(1), MA(1) uniforme e ARCH(1) e seus respectivos subrogados. Em (a), pode-se ver que a barra vertical que indica a $SampEn$ de AR(1) tem seu valor próximo ao das séries lineares geradas por meio da própria AR(1). (b) mostra esse resultado para a série MA(1) uniforme. A $SampEn$ não consegue rejeitar a hipótese para essa série. Em (c) pode-se ver os resultados de ARCH(1), caso em que se pode rejeitar a hipótese de linearidade (com incrementos gaussianos).

sub-rogados de maneira que sua distribuição de amplitudes seja equivalente à da série original. O algoritmo 2, também conhecido como AAFT (*Amplitude Adjusted Fourier Transform*), implementa essa mudança reorganizando as amplitudes dos dados criados. Pode-se descrever esse método como:

1. gerar uma sequência de dados gaussianos $y_g(t)$ de mesmo comprimento da série original $y_o(t)$;
2. reordenar $y_g(t)$ de modo que tenha a mesma distribuição temporal de amplitudes (mesmo *rank*) que $y_o(t)$. Por exemplo, se o maior elemento de $y_o(t)$ ocupa a décima posição em sua série, o maior elemento de $y_g(t)$ também deverá ocupar a décima posição dessa nova série reordenada. A série resultante terá os dados gaussianos gerados no primeiro passo, mas ordenados de tal forma que a

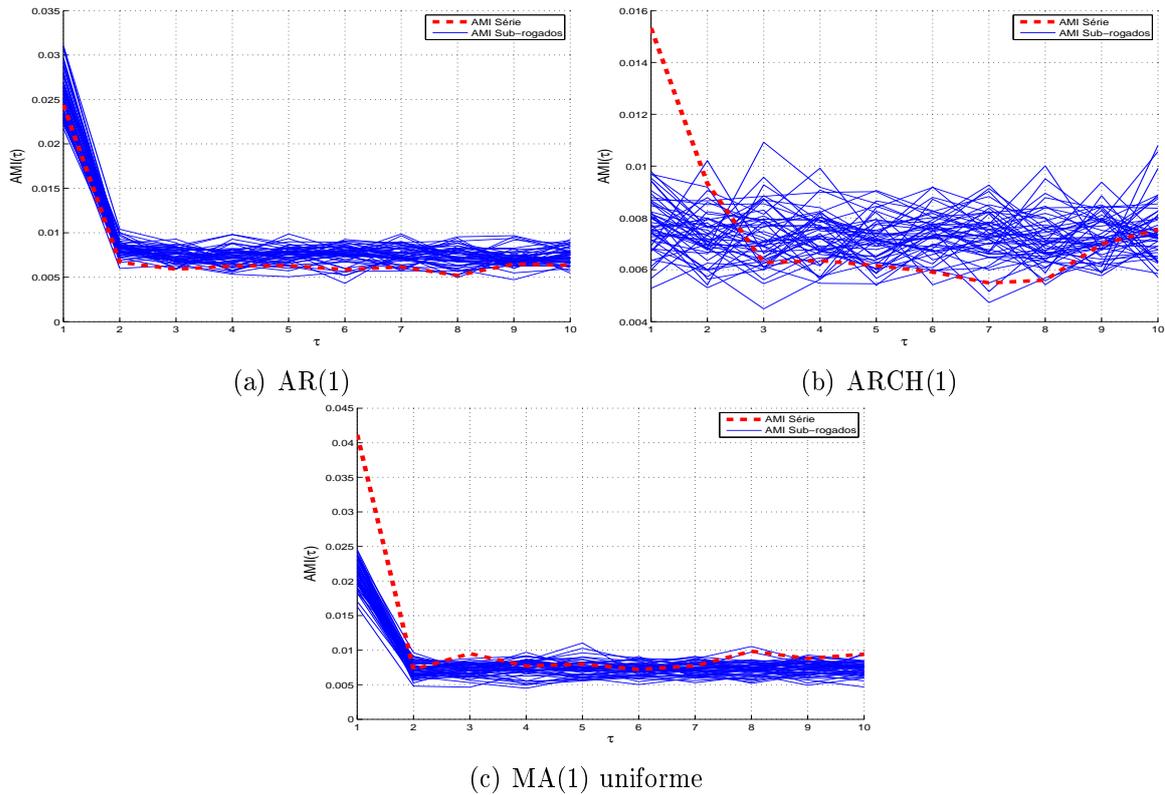


Figura 3.16. Algoritmo 1 - AMI. Em (a), a estatística AMI de AR(1) se confunde com a de seus sub-rogados. Em (b) e (c), para o primeiro atraso ($\tau = 1$), a figura aponta para a rejeição da hipótese de linearidade (com incrementos gaussianos) para as séries MA(1)uniforme e ARCH(1). Esse resultado é esperado, uma vez que ARCH(1) possui uma dependência não linear e MA(1) uniforme não possui incrementos gaussianos.

mesma variação temporal de amplitudes existente em $y_o(t)$ apareça nessa nova série, chamada de $y_{\text{rank}}(t)$. Logo, a função de autocorrelação de $y_{\text{rank}}(t)$ será, aproximadamente, a mesma de $y_o(t)$;

3. obter a transformada de Fourier da sequência $y_{\text{rank}}(t)$;
4. somar a cada fase de uma variável aleatória uniforme ($U(0,2\pi)$);
5. obter a parte real da transformada inversa de Fourier;
6. reordenar a série original $y_o(t)$ de modo que tenha a mesma distribuição temporal de amplitudes (mesmo *rank*) que a série obtida pelo passo anterior. Vale ressaltar que o resultado deste passo é uma permutação da série original. Conseqüentemente, a distribuição de probabilidade da série gerada é a mesma da original $y_o(t)$. Como o procedimento de reordenação de *rank* resulta em uma sé-

rie com aproximadamente a mesma função de autocorrelação, essa característica se conserva na série resultante.

Todo o procedimento anterior gera uma série $y_s(t)$ com as seguintes características:

- *mesma distribuição de probabilidades de $y_o(t)$.* Isso é alcançado pelo fato de $y_s(t)$ ser uma permutação de $y_o(t)$;
- *aproximadamente as mesmas dependências lineares de $y_o(t)$.* As dependências lineares originais são inseridas em $y_{rank}(t)$ no segundo passo e preservadas nos passos 3, 4 e 5, uma vez que essa informação está apenas no módulo e não nas fases da transformada de Fourier. Além disso, o último passo $y_s(t)$ com aproximadamente a mesma função de autocorrelação da série obtida pela transformada inversa;
- *dependências não lineares possivelmente existentes em $y_o(t)$ são destruídas.* Isso é feito tornando-se aleatórias as fases da transformada de Fourier do passo 4, semelhantemente ao que ocorre no algoritmo 1.

Esses passos geram dados sub-rogados que condizem com a seguinte hipótese: um processo gaussiano filtrado linearmente, mas que sofreu, em seguida, uma transformação não linear. Essa transformação altera sua distribuição para uma que não seja gaussiana (uma transformação linear geraria outra distribuição gaussiana). Portanto, uma série que se distingue de seus sub-rogados pelo algoritmo 2 não foi gerada por um processo gaussiano linearmente filtrado e estaticamente transformado (de maneira não linear). Um exemplo prático em que isso pode acontecer é quando há um processo linear experimental que é observado por meio de uma medição cuja distorção é não linear. Ou seja, o ruído de medição existente faz uma transformação não linear do processo observado, como em:

$$y_2(t) = f(y_1(t)),$$

em que $y_1(t)$ é a série linear, $y_2(t)$ é a série observada e f é uma função não linear. Séries como $y_2(t)$ não são rejeitadas pela hipótese do algoritmo 2. Se existir, porém, alguma não linearidade em $y_1(t)$, a hipótese deve ser rejeitada. Como exemplo, utiliza-se aqui, como $y_1(t)$ a série AR(1) e $y_2(t)$ é definido como:

$$y_2(t) = y_1(t)\sqrt{|y_1(t)|}.$$

Essa série é denominada AR-TENL (AR(1) com transformação estática não linear).

Além dessas duas séries, são utilizados como exemplo desse algoritmo as séries ARCH(1) e AR(1) uniforme, que pode ser descrita pela Equação 2.9, mas cujas inovações seguem a distribuição uniforme com valores entre $-0,5$ e $0,5$. As Figuras 3.17, 3.18 e 3.19 mostram as estatísticas calculadas para essas séries.

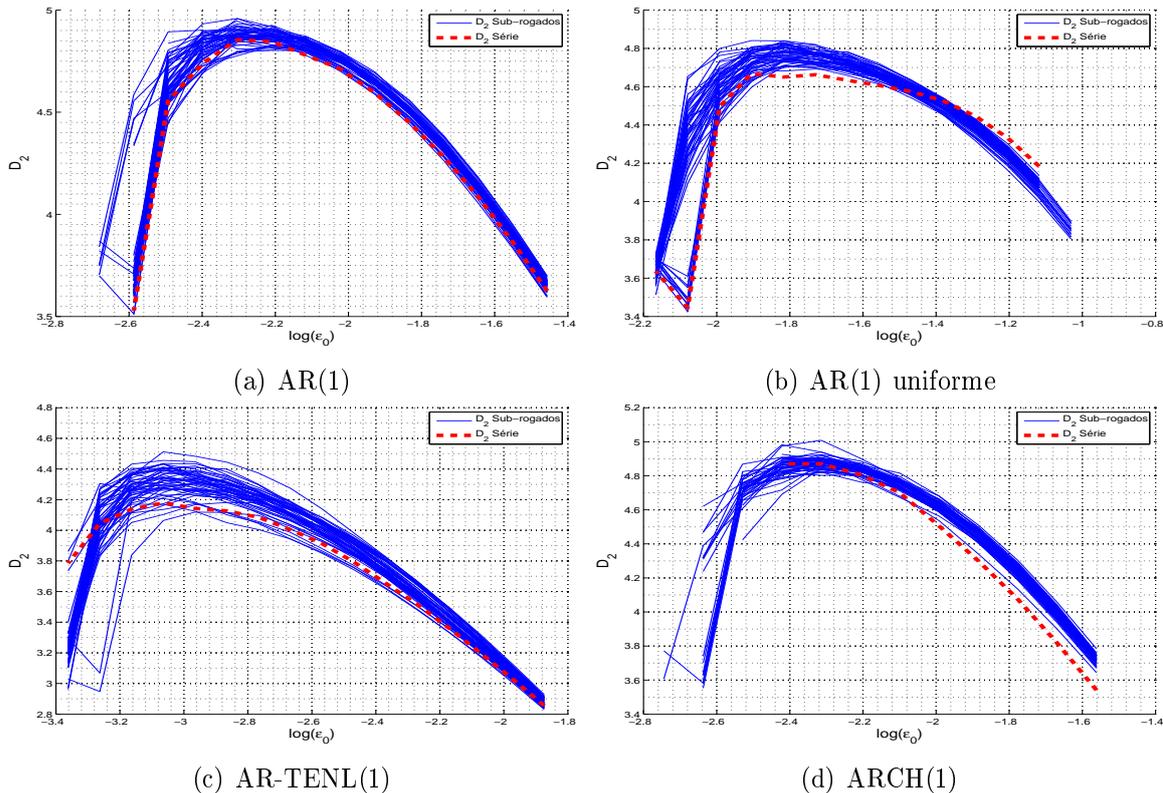


Figura 3.17. Algoritmo 2 - $D_2(\epsilon_0)$ calculada para AR(1), AR(1) uniforme, AR-TENL(1) e ARCH(1) e seus respectivos sub-rogados. Pode-se perceber que a única série que é rejeitada por essa estatística é ARCH(1). Realmente, essa é a única que possui “dinâmica” não linear.

Essas figuras mostram a rejeição da hipótese de não linearidade estática para a série ARCH(1), mas não rejeitam essa hipótese para AR(1) e AR-TENL(1). Esse é um resultado que está de acordo com o esperado, porque ARCH(1) é a única das séries possui dinâmica não linear. Essa hipótese para AR(1) uniforme é rejeitada apenas quando se utiliza a estatística AMI.

Um ruído branco gaussiano que passa por um filtro linear produz uma série gaussiana linear, uma vez que a transformação linear, apesar de modificar a distribuição, não faz com que a mesma deixe de ser gaussiana. Isso não acontece quando um ruído branco não gaussiano é passado por filtro linear. Um ruído branco uniforme, por exemplo, não produzirá uma série uniformemente distribuída, quando filtrado linearmente.

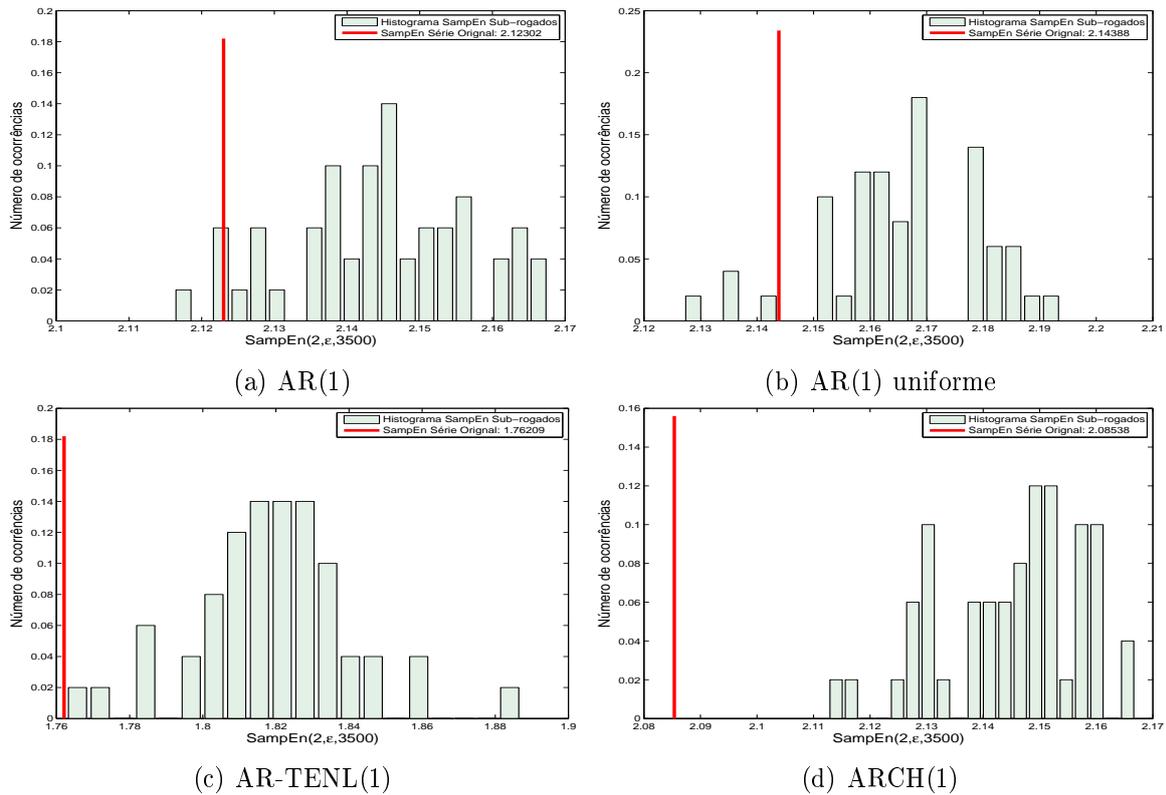


Figura 3.18. Algoritmo 2 - $SampEn(2, \epsilon, 3500)$ calculada para AR(1), AR(1) uniforme, AR-TENL(1) e ARCH(1) e seus respectivos sub-rogados. Como no caso da estatística D_2 , a única série que se afasta dos valores de $SampEn$ de seus sub-rogados é ARCH(1). Isso está de acordo com o esperado, uma vez que é a única que possui não linearidade em sua estrutura dinâmica.

O que se deve perceber é que o algoritmo 2 preserva a distribuição das séries e não das inovações de um possível modelo linear. Ou seja, um ruído branco uniforme filtrado linearmente é diferente de um ruído branco gaussiano filtrado linearmente e estaticamente transformado. Essa sutil diferença é o que justifica teoricamente a rejeição da hipótese para AR(1) uniforme. Apesar de não possuir dinâmica não linear essa série também não é uma transformação estática de um processo AR gaussiano. Ou seja, o resultado do teste não está errado. Entretanto, isso chama atenção para o fato de que podem existir hipóteses alternativas que não a de presença de não linearidades quando se rejeita a hipótese do algoritmo 2. Como D_2 e $SampEn$ não rejeitaram essa hipótese, pode-se conjecturar que tais estatísticas são mais robustas que AMI em relação à distribuição das amostras da série original.

Convém ressaltar também que, até esse ponto, rejeita-se a hipótese de independência, dependência linear ou dependência linear estaticamente não linear. Isso traz consequências bastante significativas para uma possível modelagem do processo repre-

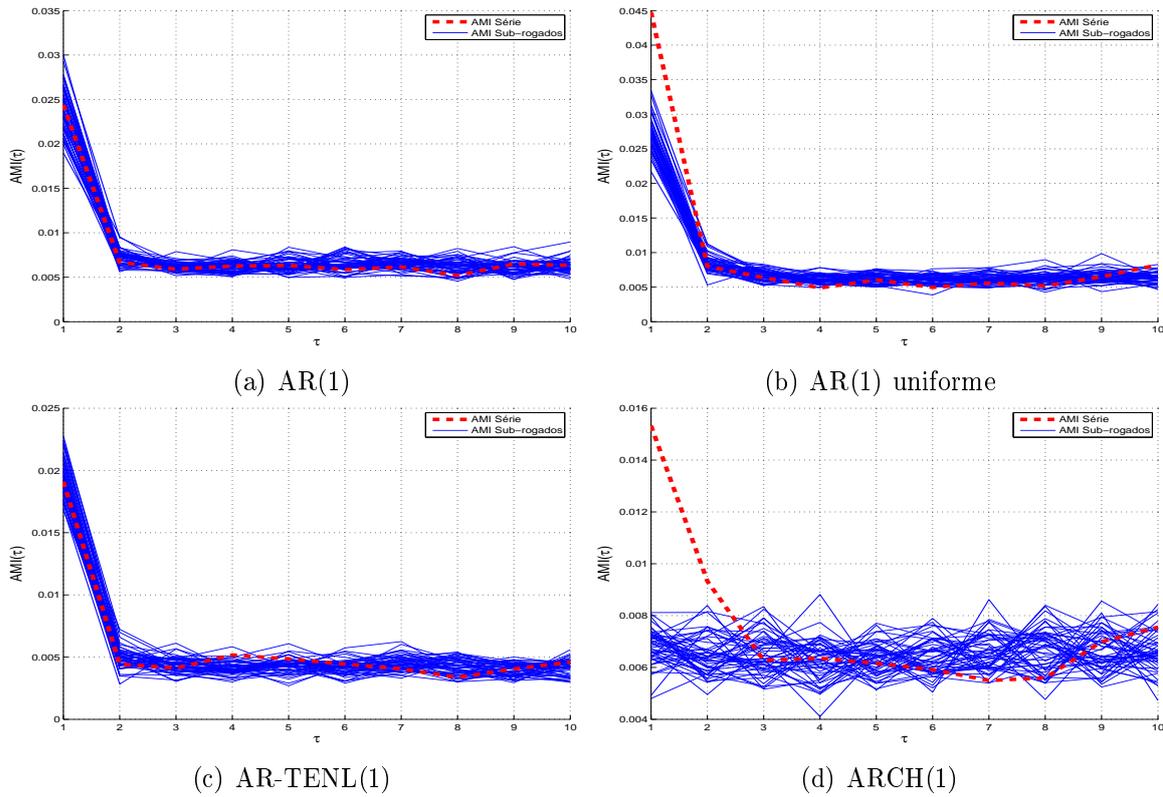


Figura 3.19. Algoritmo 2 - estatística AMI calculada para AR(1), AR(1) uniforme, AR-TENL(1) e ARCH(1) e seus respectivos sub-rogados. Nesse teste as séries AR(1) uniforme e ARCH(1) são rejeitadas como sendo lineares e estatisticamente transformadas. Esse resultado pode indicar que AMI é uma estatística mais sensível à distribuição dos dados que D_2 ou $SampEn$. A segunda série tem uma distribuição não gaussiana, o que influencia o resultado dessa estatística.

sentado pela série, fazendo com que a análise de dados sub-rogados seja uma ferramenta poderosa na etapa de identificação da estrutura (se existente) desse processo. Além disso, a importância de se discernir entre não linearidade estática e dinâmica é que a primeira não traz nenhuma informação a respeito dos valores futuros da série e é, portanto, irrelevante como contribuição para a previsibilidade.

Por fim, pode-se observar a Tabela 3.4, com os resultados numéricos do índice $PREV(2, AMI, série)$, descrito anteriormente. Essa tabela mostra, principalmente, o resultado da Figura 3.19, uma vez que AMI é a estatística utilizada para a geração desse índice. Pode-se então, desprezar a hipótese para AR(1) e AR-TENL(1), mas não para AR(1) uniforme e principalmente para ARCH(1).

Tabela 3.4. Índice de previsibilidade (Equação 3.12) das séries AR(1), AR(1) uniforme, AR-TENL(1) e ARCH(1) para o teste de linearidade com transformação não linear estática do algoritmo 2.

Série	Índice de previsibilidade $PREV(\text{Alg. 2, AMI, série})$
AR(1)	0,0819
AR(1) uniforme	7,53
AR-TENL(1)	0,168
ARCH(1)	12,3

3.6.4 Algoritmo PPS

Em uma grande quantidade de situações reais, os algoritmos anteriores não são suficientes para se determinar o tipo de dinâmica presente nos dados. É fácil perceber que em muitos casos, já se sabe de antemão se os dados são ou não independentes (algoritmo 0). Em outros, é clara a necessidade de uma modelagem linear e pode-se, eventualmente, tratar o caso da detecção de não linearidades por meio do resíduo desse modelo. Observa-se que a hipótese tratada pelo algoritmo 2 pode ser útil para separar não linearidades oriundas do processo de observação, que não contribuem para a previsibilidade. No caso de séries financeiras, no entanto, a não linearidade mais significativa acontece na forma de heteroscedasticidade. Essa não é estática, como a anterior, mas também não influencia a previsibilidade da série de retornos. Como mostrado na seção anterior, séries como ARCH(1) possuem a dinâmica necessária para que sejam rejeitadas pelo algoritmo 2. A dependência temporal é percebida no quadrado dessas séries, mostrada na Seção 2.4.

O algoritmo de geração de dados sub-rogados PPS (*Pseudo-Periodic Surrogates*) foi criado no intuito de capturar a dinâmica periódica do atrator (representada por meio de ciclos) mas tornar aleatória a dependência temporal dentro de cada período. Small coloca o algoritmo como uma alternativa para se detectar a dinâmica inter-ciclos, dada a tendência dos testes de não linearidade em perceber apenas a dinâmica do ciclo (Small *et al.* [2001]). Esse tipo de problema é comum em dados que apresentam sazonalidades, como o de consumo de energia elétrica (Rodrigues [2002]). São somadas à sazonalidade, flutuações que representam o consumo a curto prazo. Por serem mais significativas em termos de amplitude (como as variações devidas às estações do ano e a temperatura), as primeiras dominam a modelagem dos dados. Isso dificulta a previsão das flutuações. Nesse contexto, o objetivo do algoritmo é gerar dados que possuam as mesmas variações da série devidas a sazonalidade mas sem dependências de curto prazo. Ou seja, as flutuações dos dados gerados são independentes.

As séries financeiras, muitas vezes não apresentam ciclos claros, como os do exemplo anterior, mas possuem períodos cuja variância é elevada e períodos em que é significativamente menor (*clusters* de volatilidade). Em Small & Tse [2003b], Small mostra que o algoritmo PPS pode também ser utilizado para gerar séries que possuam essa característica de variância, mas sejam descorrelacionadas no curto prazo. Consegue-se, então, rejeitar a hipótese de heteroscedasticidade se a série original for diferente dessas séries criadas em termos de uma estatística discriminante. A rejeição indica que há nos dados algum determinismo passível de ser modelado.

O algoritmo é baseado em um modelo local linear, similar ao utilizado para se gerar o erro de previsão não linear (NLPE), em Small [2005] e pode ser descrito pelos seguintes passos:

1. escolher aleatoriamente um elemento T_1 da série original $y_o(t)$. Esse será o primeiro elemento da série sub-rogada, $y_s(t)$. Ou seja:

$$y_s(1) = y_o(T_1)$$

em que T_1 é um número aleatório tomado entre m (dimensão de imersão, utilizada no próximo passo) e N , tamanho da série. Durante esse passo, a série não está imersa em um espaço de fases, e a função de m é simplesmente, fornecer um limite inferior para a escolha de um elemento da série original;

2. realizar a imersão da série original $y_o(t)$ em um espaço de fases de dimensão m . A série será transformada em $N - m + 1$ vetores x_i (definidos na Equação 3.3);
3. localizar, entre os vetores gerados, o vetor x_{T_1+m} , que é aquele cujo primeiro elemento (identificado por $y(i)$ na equação) coincide com o elemento T_1 do passo anterior;
4. escolher, aleatoriamente, um vizinho de x_{T_1+m} no espaço de fases. Essa escolha deve se dar de acordo com a distribuição de probabilidades:

$$P(x_v = x_i) = \exp\left(\frac{-\|x_i - x_{T_1+m}\|}{\rho}\right),$$

sendo x_v o vetor vizinho escolhido e ρ um parâmetro denominado aqui como ruído de dispersão (*noise radius*). A probabilidade de um vizinho de x_{T_1+m} ser escolhido diminui exponencialmente com a distância entre eles. O próximo elemento da série $y_s(t)$ será o primeiro elemento do vetor x_v ;

5. repetir os passos anteriores (excluindo-se, obviamente, o primeiro passo) para obter $y_s(i)$ até que $i = N$ (a série gerada deve ter o mesmo tamanho que a original).

Os passos acima geram uma série $y_s(t)$ que, quando imersa em uma dimensão m , produz um atrator semelhante ao da série original. A diferença está nas pequenas distorções geradas (proporcionais a ρ), que modificam o atrator localmente, mas preservam seu aspecto global. Observa-se que, se ρ for pequeno, os vizinhos de $y_o(t)$, serão também os vizinhos de $y_s(t)$ e a série sub-rogada terá a tendência de ser uma cópia da original. Em contrapartida, se esse ruído de dispersão for muito grande, a tendência de $y_s(t)$ é ter elementos que são simplesmente amostrados aleatoriamente (com reposição) de $y_o(t)$. Esse último caso gerará uma série sub-rogada i.i.d., semelhante a do algoritmo 0.

Percebe-se que a escolha de ρ é importante porque o que se deseja é um ρ médio, que reproduza a dinâmica periódica “macroscópica” do atrator, mas que produza elementos vizinhos descorrelacionados. O procedimento seguido aqui é o mesmo descrito em Small & Tse [2003a]. Escolhe-se um valor de ρ que maximize a probabilidade de que sequências de dois elementos sejam idênticas tanto em $y_o(t)$ quanto em $y_s(t)$. Em outras palavras, para um determinado $y_s(T_i)$ que possua seu correspondente $y_o(T_j)$ na série original, $y_s(T_i + 1)$ é igual a $y_o(T_j + 1)$. Utiliza-se, para esse algoritmo, a implementação de Michael Small⁷ usada também em seu livro sobre o assunto (Small [2005]). A implementação faz uso do parâmetro de configuração igual a 0,6. Esse valor está próximo de 0,5, valor que maximiza a probabilidade acima. $p = 0,6$ distinguiu melhor as séries estudadas do que $p = 0,5$ e foi o escolhido para este trabalho.

Small & Tse [2003b] trazem a aplicação do algoritmo PPS como forma de se detectar dinâmicas presentes em séries financeiras que não tenham como origem a heteroscedasticidade e sejam, possivelmente, previsíveis. Como mostrado nessa referência, séries como ARCH(1) e GARCH(1,1) não devem ser rejeitadas, mas séries que possuam alguma dinâmica não linear previsível (como LOGMAP) devem ser rejeitadas. Como observado na Seção 2.5, a heteroscedasticidade não implica dependência não linear. Aqui, porém, supõe-se que não haja outra fonte de estacionariedade que não a produzida por processos ARCH ou GARCH. A hipótese nula é, portanto, voltada para testes desses processos.

As Figuras 3.20, 3.21 e 3.22 mostram os resultados obtidos por meio da aplicação do algoritmo PPS nas séries ARCH(1), GARCH(1,1) e LOGMAP. Embora o objetivo inicial do algoritmo seja o de detectar determinismo em séries pseudoperiódicas, seu

⁷Disponível em <http://www.eie.polyu.edu.hk/~ensmall/matlab/index.html>.

uso na detecção de “dinâmicas previsíveis” possivelmente presentes em séries heteroscedásticas é o principal objetivo desta seção. O procedimento é baseado em Small & Tse [2003b], em cuja aplicação do algoritmo PPS tem exatamente esse objetivo.

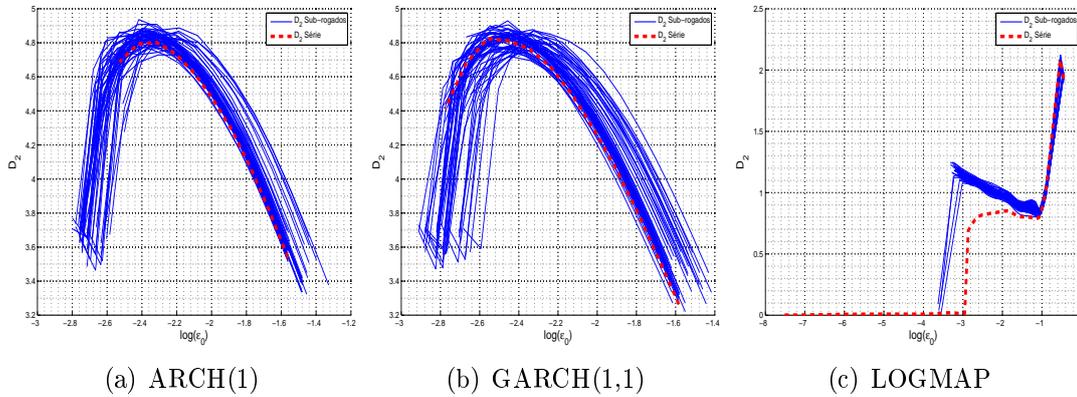


Figura 3.20. Algoritmo PPS - D_2 . Comparando-se a estatística D_2 dos dados, percebe-se que a série obtida do mapa logístico (c) é a única que consegue se diferenciar dos dados sub-rogados gerados. Isso indica que a série possui algum determinismo não linear diferente de, por exemplo, processos ARCH e GARCH.

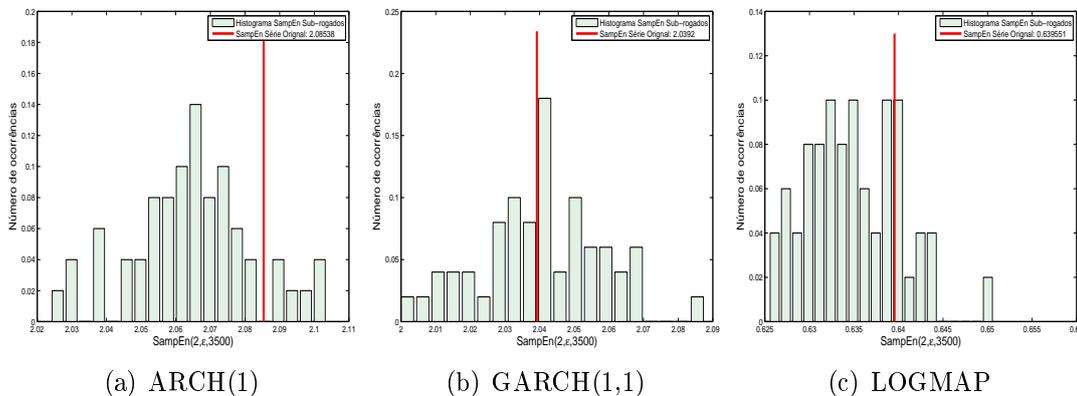


Figura 3.21. Algoritmo PPS - $SampEn(2, \epsilon, 3500)$. A estatística $SampEn(2, \epsilon, 3500)$ não é capaz de diferenciar as séries de seus sub-rogados. Esse resultado indica uma limitação dessa estatística ao ser aplicada em testes do algoritmo PPS.

As figuras mostram que a dimensão de correlação D_2 é a única estatística capaz de discriminar corretamente as séries de seus sub-rogados. LOGMAP é determinística e deve ser diferente de seus sub-rogados. A $SampEn(2, \epsilon, 3500)$ e a AMI, porém, não conseguem detectar essa diferença. Era de se esperar que ARCH(1) e GARCH(1,1) se confundissem com as séries geradas a partir delas, uma vez que seu único conteúdo determinístico está na determinação da variância de suas inovações, o que não influencia a previsibilidade. Esse resultado está de acordo com o descrito em Small & Tse [2003b].

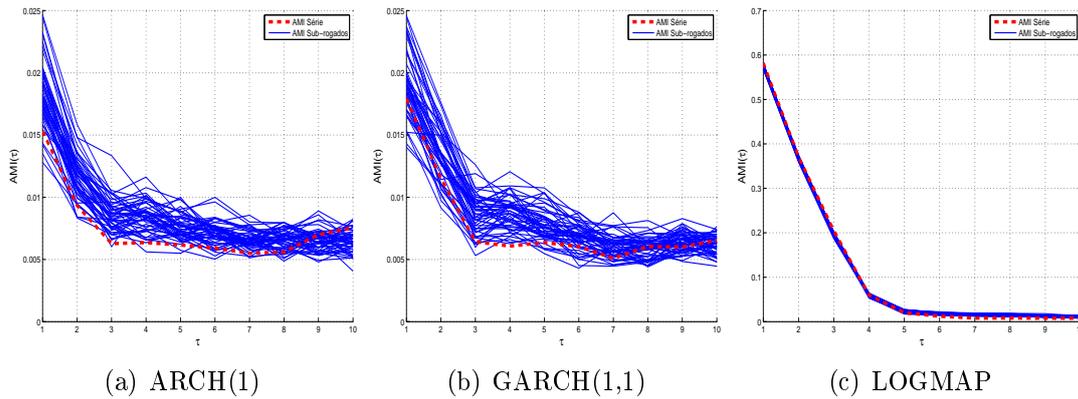


Figura 3.22. Algoritmo PPS - AMI. Como visto no caso da entropia amostral, a informação mútua média AMI, não distingue as séries ARCH(1), GARCH(1,1) e LOGMAP de seus sub-rogados. A série LOGMAP, como vista anteriormente (Seção 2.1), é claramente diferente de ARCH(1) e GARCH(1,1), o que mostra que a AMI não discrimina adequadamente as características testadas pelo algoritmo PPS.

Nesse caso, não se calcula o índice $PREV$ para esse algoritmo pelo fato de a estatística AMI, base para o cálculo do índice nos outros algoritmos, não ter o poder discriminatório adequado. A princípio, poderia-se utilizar o índice utilizando-se a estatística D_2 . Na prática, porém, utiliza-se um conjunto diferente de ϵ_0 para cada série, o que torna difícil o cálculo do índice.

Na prática, não se sabe, a princípio, qual será o discriminante que será o responsável pela rejeição das hipóteses. O importante, entretanto, é notar que o índice $PREV$ é uma forma de quantificar a diferença entre os dados sub-rogados e a série que os originou, fazendo com que esses testes não sejam simplesmente baseados em inspeções visuais.

No restante do trabalho, o índice $PREV$ será utilizado como medida de previsibilidade da série utilizando como estatística discriminante o W_2 baseado no teste BDS. Isso será feito com os resíduos de um modelo GARCH ajustado aos dados, a fim de desconsiderar o efeito das dependências heteroscedásticas na estimação da previsibilidade.

Previsibilidade de Séries Financeiras

O presente capítulo aplica algumas técnicas vistas no Capítulo 3 em séries financeiras. Essas séries são apresentadas na seção inicial e o restante do capítulo descreve os resultados obtidos, em seções organizadas não por técnica, mas pelo tipo de teste que é realizado: de independência, linearidade e detecção do tipo de não linearidade. A última seção do capítulo mostra uma maneira de se tentar quantificar a previsibilidade das séries.

4.1 As Séries

As séries financeiras escolhidas para a aplicação das técnicas de previsibilidade são obtidas nos mercados de capitais brasileiro e norte-americano. São elas: o índice *Dow Jones* industrial, o índice S&P 500, o índice Ibovespa e as séries de preços de abertura, máxima, mínima e fechamento das ações preferenciais da Petrobras. As séries não serão detalhadas, mas apenas apresentadas junto a uma breve justificativa da relevância da mesma para o trabalho. Metodologias de cálculo de índices e suas composições, bem como informações a respeito de empresas com capital em bolsa, podem ser facilmente obtidas em *sites* na Internet.

O estudo de previsibilidade é feito por meio do retorno das séries subtraídas de sua tendência, como mostrado no início do Capítulo 2. A fim de facilitar a leitura do trabalho, a apresentação dos resultados de cada série em outras seções seguirá preferencialmente a mesma ordem em que serão apresentadas nesta seção.

4.1.1 *Dow Jones Industrial Average*

A primeira série escolhida é retirada do índice *Dow Jones Industrial Average* (DJIA). É um índice formado pela média de preços de fechamento diário de 30 das principais empresas norte-americanas em vários setores da economia e tem grande importância nos estudos de séries financeiras. Charles Dow, seu criador, é considerado o pai da análise

técnica, que se baseia quase que exclusivamente na análise¹ da série temporal de ativos financeiros. Como o índice é calculado diariamente desde o final do século XIX, eventos históricos bastante relevantes podem ser observados ao longo de sua série. Além disso, o índice serviu de base para vários trabalhos importantes na área e, conseqüentemente, estudos sobre o mesmo podem ser diretamente comparados com outros trabalhos. Uma outra vantagem dessa série é ser bastante longa quando comparada com a maioria das séries diárias disponíveis no mercado. A duração da série DJIA utilizada neste trabalho compreende os retornos sem tendência de 27 de maio de 1896 a 27 de janeiro de 2011. A Figura 4.1 mostra o índice e seus retornos sem tendência (subtraindo-se a média dos retornos).

A série a ser estudada é mostrada na Figura 4.1.(b) e será denominada DJIA. Ela é obtida a partir da transformação $\log(p(t)) - \log(p(t-1))$ em que $p(t)$ é o valor do índice no dia t . Após a transformação, cada elemento é subtraído pela média amostral da série transformada. A série do índice foi obtida em *www.economagic.com*.

4.1.2 S&P 500

O *Standard & Poor's 500* (S&P 500) é um índice norte-americano composto de 500 empresas escolhidas entre as maiores de seu segmento e com alta liquidez no mercado de ações. Por ser composto de um número grande de empresas quando comparado ao DJIA (30 empresas), é considerado uma melhor representação do mercado. A Figura 4.2 mostra tanto a série do índice quanto seu retorno diário. O eixo das ordenadas é colocado em escala logarítmica para facilitar a visualização dos valores menores no início da série.

A transformação do índice em série de retornos é feita como na série DJIA e é mostrada na Figura 4.2. Essa será chamada SP500 neste trabalho e será também uma das séries nas quais será estudada a previsibilidade. A série do índice foi obtida em *www.economagic.com*.

4.1.3 Ibovespa

A bolsa de valores de São Paulo (BMF&Bovespa) tem como principal índice para seu mercado de ações o índice Bovespa, ou Ibovespa. Esse representa o comportamento médio das principais ações negociadas no país². O Ibovespa é composto por uma carteira teórica de ações das quais fazem parte empresas com grande número de negociações e

¹Apesar de também utilizar séries temporais, esse tipo de análise pouco se assemelha ao estudo desenvolvido neste trabalho.

²Extraído de <http://www.bmfbovespa.com.br>

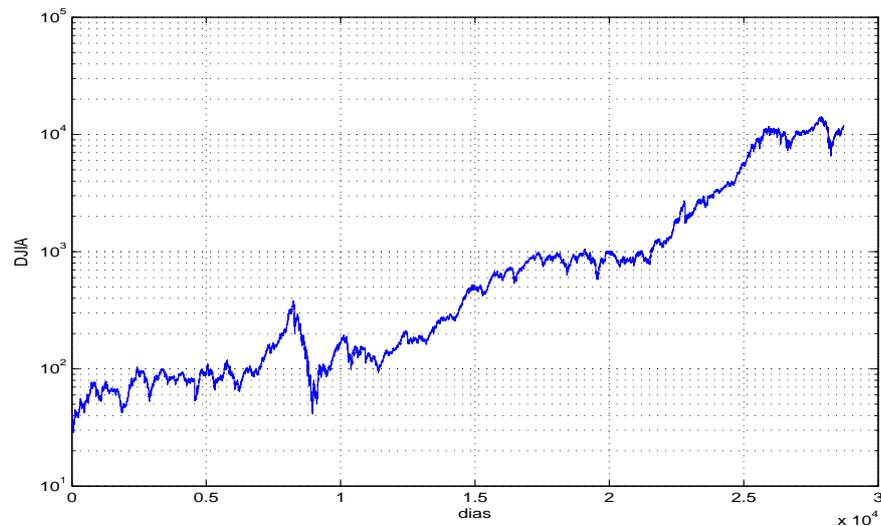
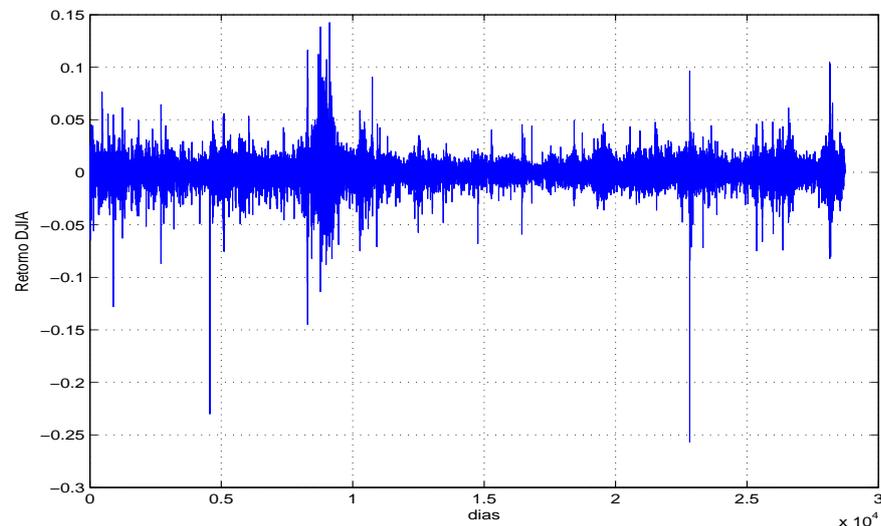
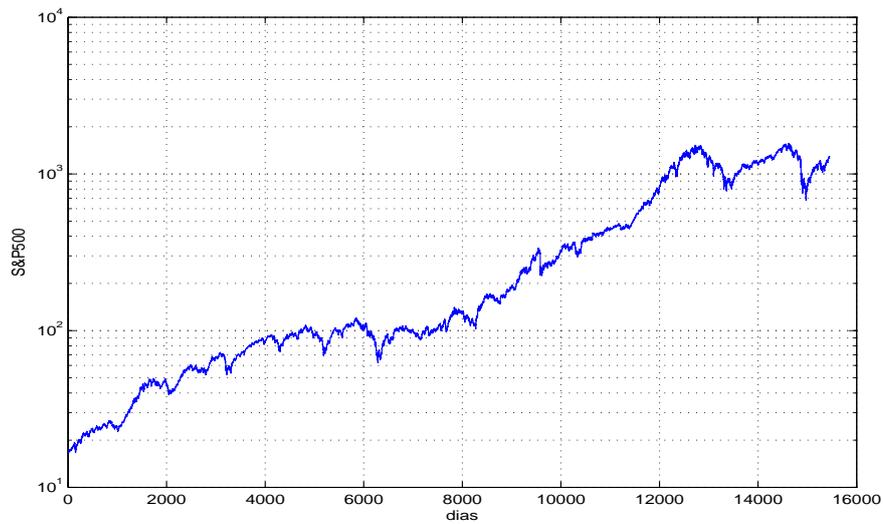
(a) Índice *Dow Jones*.(b) Retorno sobre o índice *Dow Jones*.

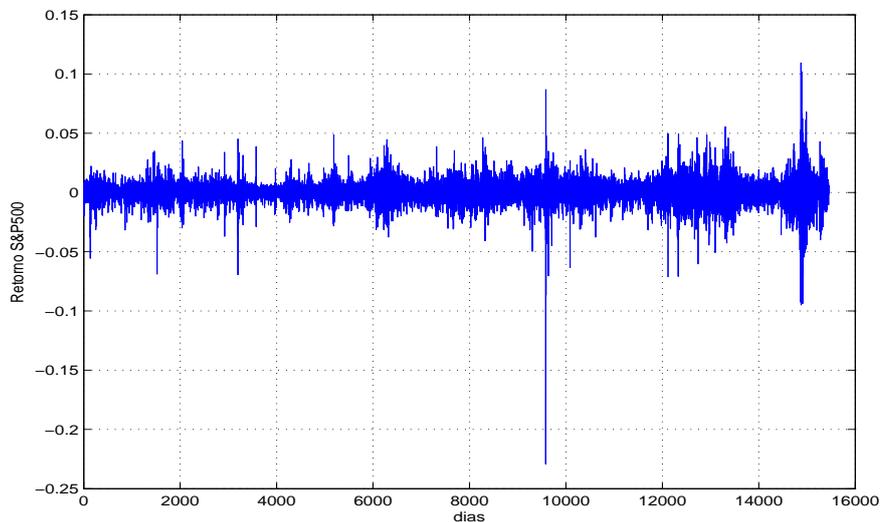
Figura 4.1. O gráfico (a), com eixo das ordenadas em escala logarítmica, mostra o índice *Dow Jones* durante 28.724 dias no período de 26 de maio de 1896 a 27 de janeiro de 2011. Após o primeiro dia, pode-se calcular o retorno como $\log(p(t)) - \log(p(t-1))$ em que $p(t)$ é o valor do índice (gráfico (a)) no instante t . O gráfico (b) mostra esse retorno subtraído de sua média amostral.

volume financeiro em bolsa. A metodologia de cálculo ordena todas as empresas em termos do produto dessas duas características. Escolhe-se as empresas que representarem juntas 80 % do valor total desse produto. Em abril de 2011, 67 empresas participavam do índice. A Figura 4.3 mostra tanto o índice quanto o seu retorno, ambos cotados diariamente.

A metodologia para cálculo dos retornos é a mesma que a utilizada no índice *Dow Jones* e gera a série IBOV a ser estudada no presente trabalho. A comparação entre



(a) Índice S&P500.

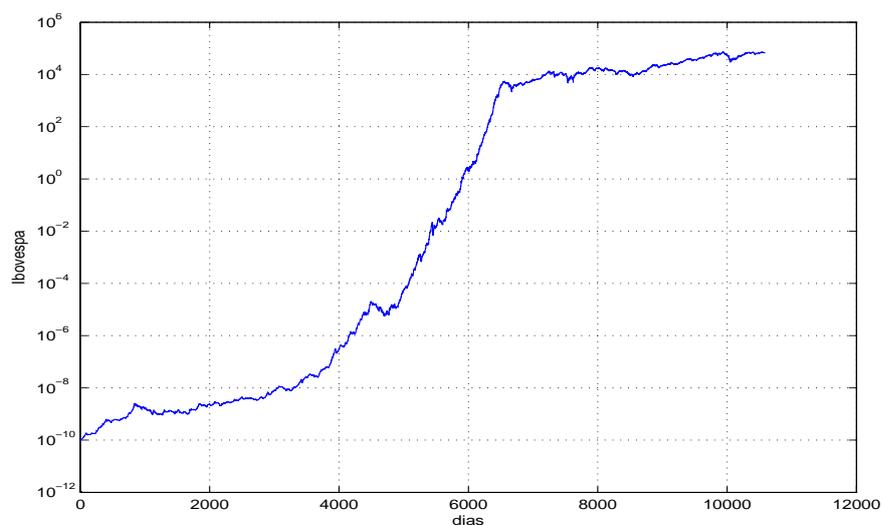


(b) Retorno sobre o índice S&P500.

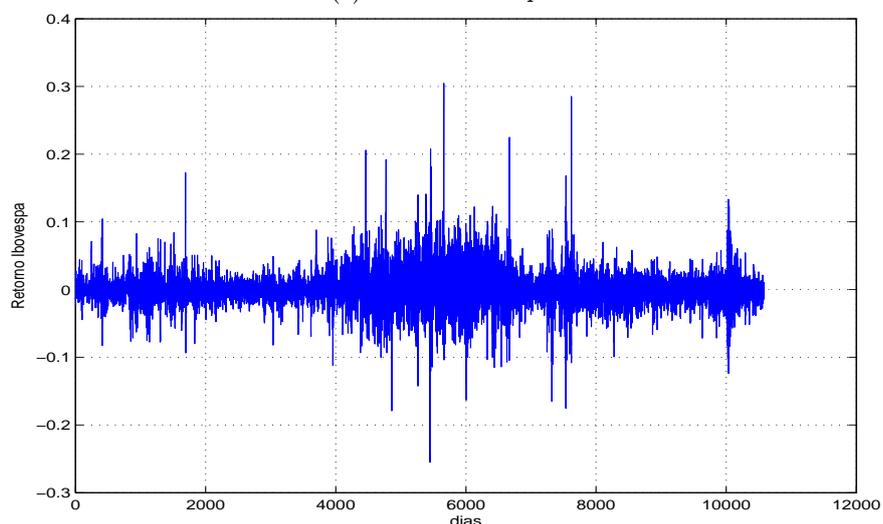
Figura 4.2. O gráfico (a), com eixo das ordenadas em escala logarítmica, mostra o índice S&P500 durante 15.453 dias no período de 4 de janeiro de 1950 a 27 de janeiro de 2011. Após o primeiro dia, pode-se calcular o retorno como $\log(p(t)) - \log(p(t-1))$ em que $p(t)$ é o valor do índice (gráfico (a)) no instante t . O gráfico (b) mostra esse retorno subtraído de sua média amostral.

a previsibilidade de IBOV e a de DJIA ou SP500 traz subsídios importantes para o estudo da eficiência³ dos mercados de países emergentes e desenvolvidos, considerando que Brasil e Estados Unidos são dois dos principais representantes desses dois mercados. A série do índice foi obtida em *www.bmfbovespa.com.br*.

³Eficiência no sentido informacional, como na hipótese dos mercados eficientes descrita no Capítulo 1.



(a) Índice Ibovespa.



(b) Retorno sobre o índice Ibovespa.

Figura 4.3. O gráfico (a), com eixo das ordenadas em escala logarítmica, mostra o índice Ibovespa durante 10.582 dias no período de 2 de janeiro de 1968 a 30 de dezembro de 2010. Após o primeiro dia, pode-se calcular o retorno como $\log(p(t)) - \log(p(t-1))$ em que $p(t)$ é o valor do índice (gráfico (a)) no instante t . O gráfico (b) mostra esse retorno subtraído de sua média amostral.

4.1.4 Petrobras

Das empresas que compõem o índice Ibovespa, a Petróleo Brasileiro S/A (Petrobras) é uma das mais importantes⁴. Isso porque suas ações ordinárias (PETR3) e preferenciais (PETR4) somadas contribuem com 14,1 % da composição do índice. Apenas a companhia mineradora Vale S/A, com 14,7% supera esse valor. As ações preferenciais

⁴Informações com base na carteira Ibovespa de janeiro, do período entre janeiro e abril de 2011

da Petrobras representam 11 % do Ibovespa e são as mais negociadas da empresa. Os preços de abertura, máxima, mínima e fechamento diários dessas ações preferenciais são os escolhidos para se estudar a previsibilidade de ações.

Analisar uma ação é, em princípio, diferente de se analisar um índice. A primeira diferença é no efeito de média que o índice apresenta, no intuito de representar o mercado como um todo. Uma possível diferença entre a previsibilidade de uma ação e a de um índice pode estar relacionada a esse efeito. Pode-se supor que a primeira está relacionada com os fatores internos e externos que influenciam o valor de uma empresa enquanto a segunda está relacionada à dinâmica do mercado de maneira mais abrangente.

Antes de analisar a previsibilidade dessas séries deve-se descontar das mesmas os efeitos da distribuição de capital aos acionistas. Quando há distribuições de dividendos e juros sobre capital próprio, esses são subtraídos do preço do ativo e deve-se ajustar a série resultante de modo a eliminar essa característica da análise. A Figura 4.4 mostra as cotações de abertura, máxima, mínima e de fechamento diárias (cotações *Open High Low Close*, ou OHLC) bem como seus retornos, como realizado no índice *Dow Jones*. Portanto, as séries mostradas na figura já são alteradas para desconsiderar esses efeitos.

Espera-se que as características individuais das cotações OHLC de uma mesma empresa possam refletir em seus índices de previsibilidade. Os instantes de abertura, fechamento, máxima e mínima das negociações representam momentos bastante distintos do dia, não só em termos do tempo, mas também das expectativas do mercado. Na abertura, por exemplo, há ordens de compra ou venda que começam a ser registradas desde o último fechamento e só são negociadas nos primeiros instantes de negociações do dia seguinte. Isso faz com que exista um grande atraso entre o momento das ordens e de suas negociações e certamente influi na dinâmica de abertura dos mercados.

Serão utilizadas as quatro séries de retorno da Figura 4.4, identificadas como PETO (abertura), PETH (máxima), PETL (mínima) e PETC (fechamento). Essas foram obtidas por meio do *site* www.bmfbovespa.com.br.

4.2 Detecção de Determinismo e Quantificação da Previsibilidade - Abordagens

Vários testes são realizados ao longo deste capítulo com o intuito de se detectar o determinismo das séries financeiras. A Figura 4.5 mostra como será aplicado o teste BDS nas séries a fim de se detectar os diferentes tipos de determinismo.

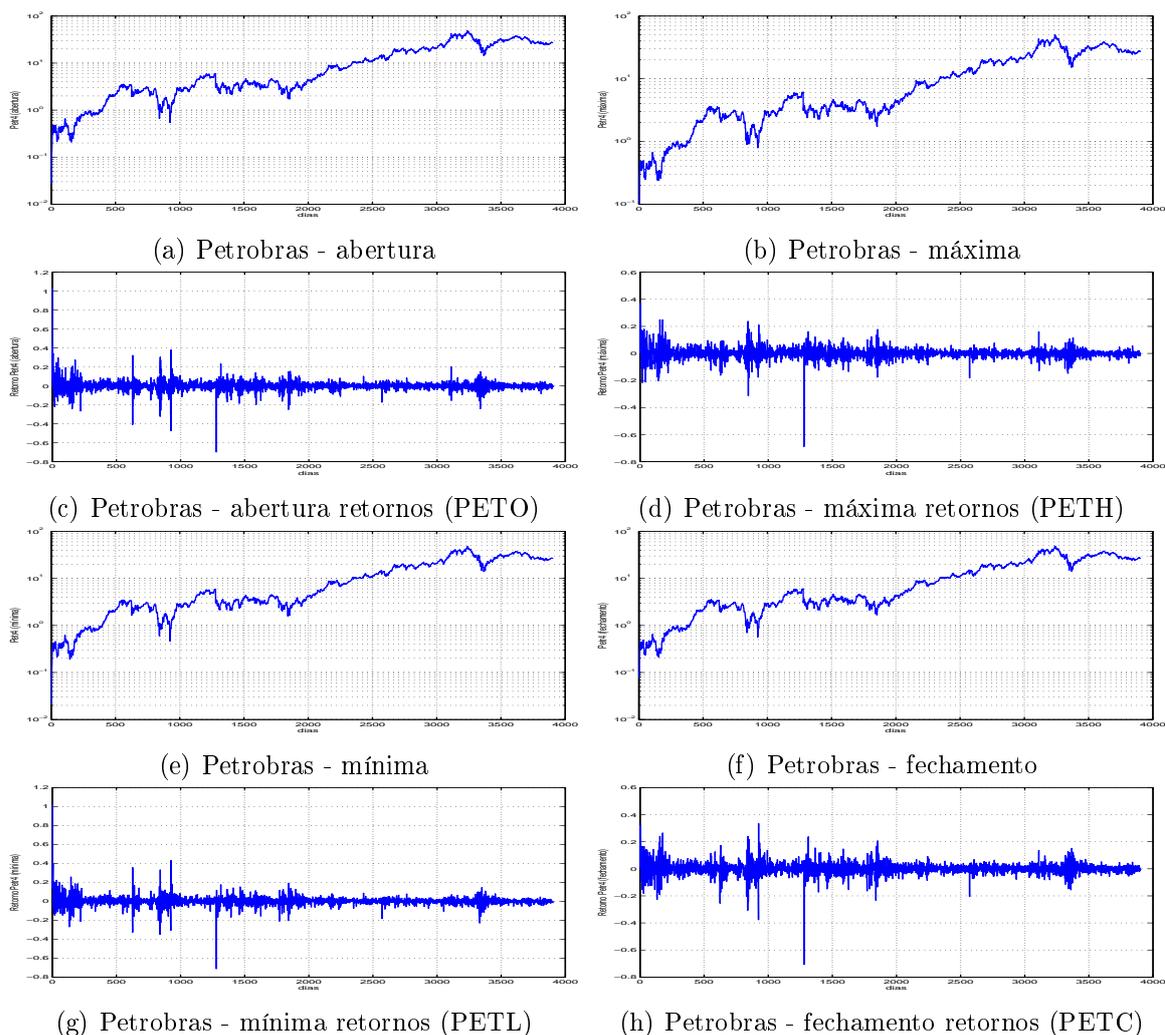


Figura 4.4. Os gráficos (a), (b), (e) e (f), com eixo das ordenadas em escala logarítmica, mostram os preços de abertura, máxima, mínima e fechamento diário das ações preferenciais da Petrobras durante 3.904 dias no período de 17 de abril de 1995 a 28 de janeiro de 2011. Após o primeiro dia, pode-se calcular o retorno como $\log(p(t)) - \log(p(t-1))$ em que $p(t)$ é o preço (gráficos (a), (b), (e) e (f)) no instante t . Os gráficos (c), (d), (g) e (h) mostram os respectivos retornos subtraídos de suas médias amostrais.

Os testes BDS a serem realizados são uma das duas abordagens utilizadas no presente trabalho. A outra é a análise de dados sub-rogados. A Figura 4.6 mostra o procedimento a ser seguida para essa abordagem.

Após as duas metodologias de teste (BDS e dados sub-rogados), se detectado algum determinismo, pode-se inferir o quanto desse determinismo é previsível desconsiderando do mesmo a parte devida a dependências heteroscedásticas. Isso é feito por meio do índice $PREV$ (Equação 3.12). A fim de tornar mais claro o procedimento de cálculo desse índice, a Figura 4.7 mostra as etapas principais a serem realizadas.

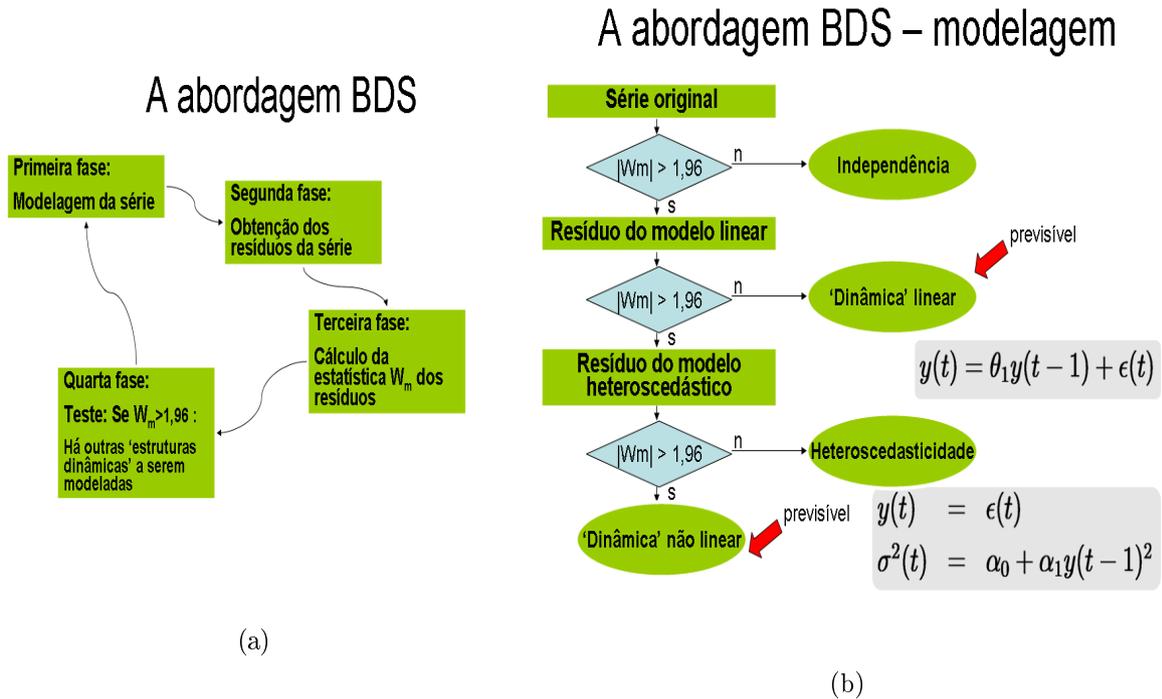


Figura 4.5. Fluxograma dos testes BDS. O gráfico (a) exhibe o procedimento realizado em cada teste. O gráfico (b) mostra a abordagem geral a ser seguida em cada série.

Esses procedimentos serão aplicados às séries financeiras estudadas

4.3 Independência dos Retornos

Esta seção mostra a aplicação do teste BDS e do algoritmo 0 (da análise de dados sub-rogados) nas séries DJIA, SP500, IBOV, PETO, PETH, PETL e PETC descritas na Seção 4.1. Os testes são aplicados diretamente às séries, no intuito de se considerar todas as dependências temporais existentes nas mesmas.

Uma característica bastante desejável desses testes é o fato de serem independentes da distribuição dos dados. Essa é uma característica essencial na análise de séries financeiras, contornando um dos problemas mostrados ao final do Capítulo 2. Naquele capítulo os resultados pressupunham a normalidade dos retornos. Com as ferramentas utilizadas neste capítulo, essa condição não precisa ser mais atendida.

A abordagem de Dados sub-rogados

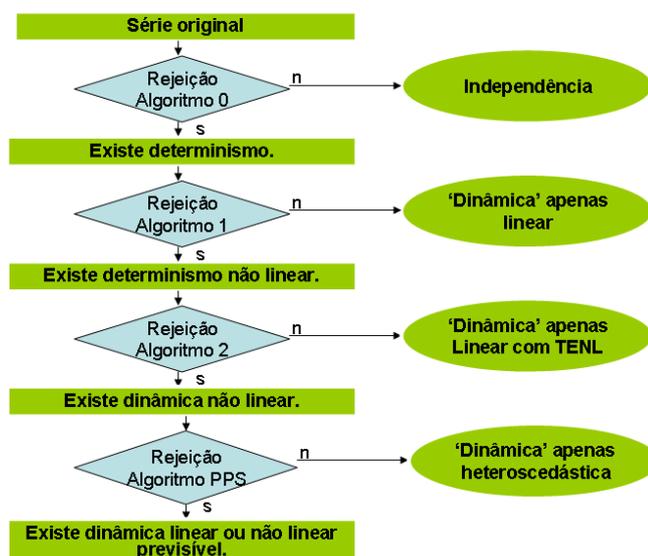


Figura 4.6. Fluxograma da análise de dados sub-rogados.

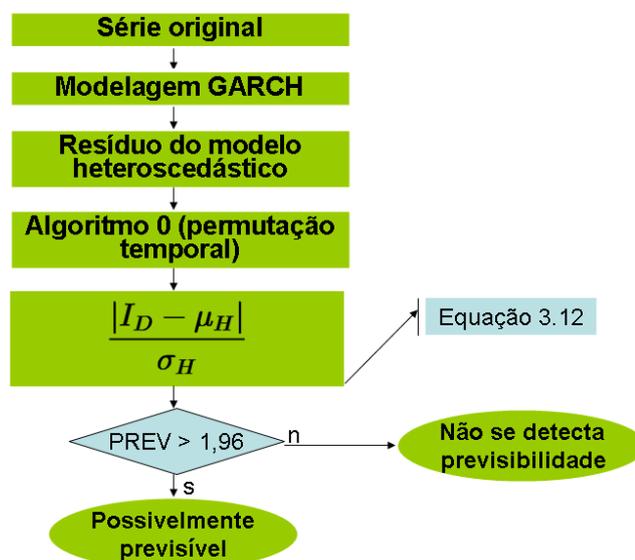
4.3.1 Teste BDS - Independência

O teste BDS, descrito na Seção 3.5, é aplicado a cada uma das séries. Dessa maneira pode-se testar a independência dos retornos. Por meio do teste realizado nesta seção não é possível se detectar o tipo de dependência existente. Pode-se porém, com o resultado obtido, comparar-se essas séries com o passeio aleatório de uma maneira mais geral, independente da distribuição dos retornos. Como foi visto no Capítulo 2, testes de independência que pressupõem distribuições gaussianas tem pouca aplicação em séries financeiras.

A Figura 4.8 exhibe a distribuição empírica (obtida por meio de histogramas) de W_2 para permutações das séries, além dos valores W_2 para as mesmas. As permutações são obtidas da mesma maneira que os dados sub-rogados gerados pelo algoritmo 0. Isso produz séries descorrelacionadas e permite comparar o valor de W_2 obtido com o de séries semelhantes (mesmos momentos estatísticos). Dessa forma, pode-se rejeitar a independência com mais segurança⁵.

⁵A estatística W_2 para dados independentes segue uma distribuição gaussiana com variância unitária e média nula. Consequentemente, não seria necessário obter a distribuição empírica desses dados, como realizado neste trabalho. Isso, entretanto, só é válido se atendidas certas condições em relação aos dados. Em Caporale *et al.* [2004] mostra-se que essas condições não são tão restritas e realizam-se alguns estudos de Monte Carlo com séries que contém heteroscedasticidade, como as trabalhadas aqui. Ainda assim, comparar W_2 com séries obtidas por meio de permutações da original é uma abordagem mais conservadora, porque elimina a necessidade de se atender a essas condições.

Cálculo do Índice *PREV*



$I_D = W_2$ dos resíduos; μ_H e σ_H = média e desvio-padrão de W_2 dos sub-rogados;

Figura 4.7. Fluxograma do cálculo do índice de previsibilidade.

A Tabela 4.1 lista os valores de W_2 obtidos bem como o índice *PREV*, calculado por meio da comparação das séries com as suas permutações. Mostra-se também o resultado do teste Kolmogorov-Smirnov, indicando que não se rejeita a hipótese de normalidade para nenhuma série, o que dá mais confiança ao índice *PREV*.

Tabela 4.1. W_2 e índice de previsibilidade (Equação 3.12) das séries para o teste de independência BDS. Valores de W_2 superiores a 1,96 indicam a rejeição da hipótese de independência. Mostra-se também o resultado do teste Kolmogorov-Smirnov (KS) da distribuição empírica de W_2 (valores superiores a 0,188 indicam a não normalidade da distribuição de W_2).

Série	$ W_2 $	Valor limite de W_2	Índice de previsibilidade <i>PREV</i> (BDS, série)	Teste KS
DJIA	42,8	1,96	53,2	0,101
SP500	26,7	1,96	22,7	0,105
IBOV	36,0	1,96	33,9	0,115
PETO	22,3	1,96	21,3	0,115
PETH	22,8	1,96	22,8	0,186
PETL	25,4	1,96	27,5	0,129
PETC	21,2	1,96	24,5	0,0730

Tanto na tabela, quanto nas figuras, pode-se perceber que há alguma dependência temporal nas séries. A DJIA apresenta o maior valor de *PREV*, o que não necessa-

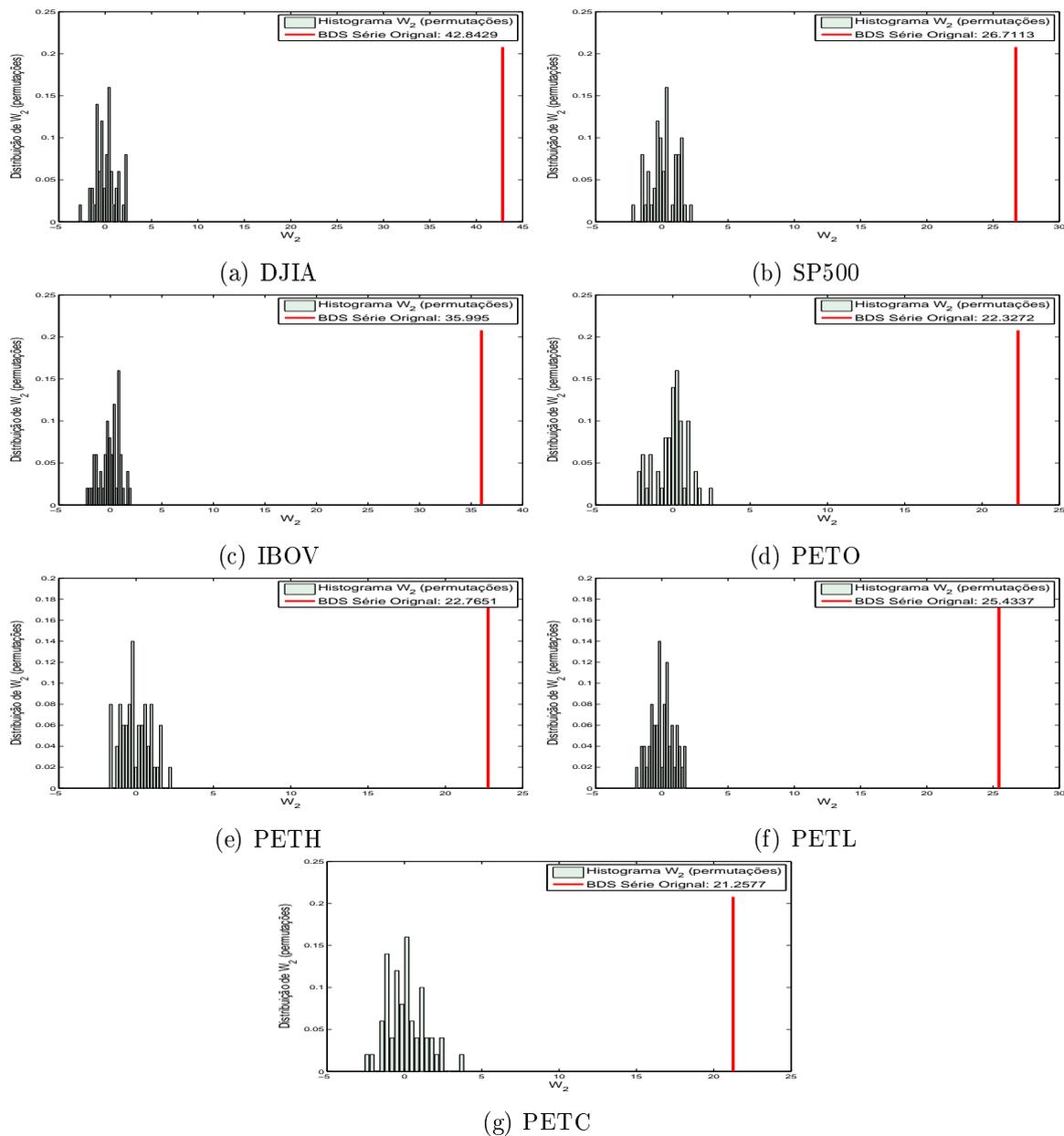


Figura 4.8. Teste BDS de independência. A estatística W_2 é calculada para as séries. Pode-se rejeitar a hipótese de independência para todas, uma vez que ultrapassam o valor limite de 1,96.

riamente indica uma maior previsibilidade, uma vez que esse teste ainda não indica o motivo do elevado valor desse índice em comparação com as outras séries.

4.3.2 Dados Sub-rogados - Algoritmo 0

O algoritmo 0 de geração de dados sub-rogados é utilizado para se rejeitar a independência dos dados. Nesta seção, o mesmo procedimento descrito na Seção 3.6 é aplicado

às séries. A Figura 4.9 apresenta a informação mútua média (AMI) das séries e dos dados sub-rogados criados.

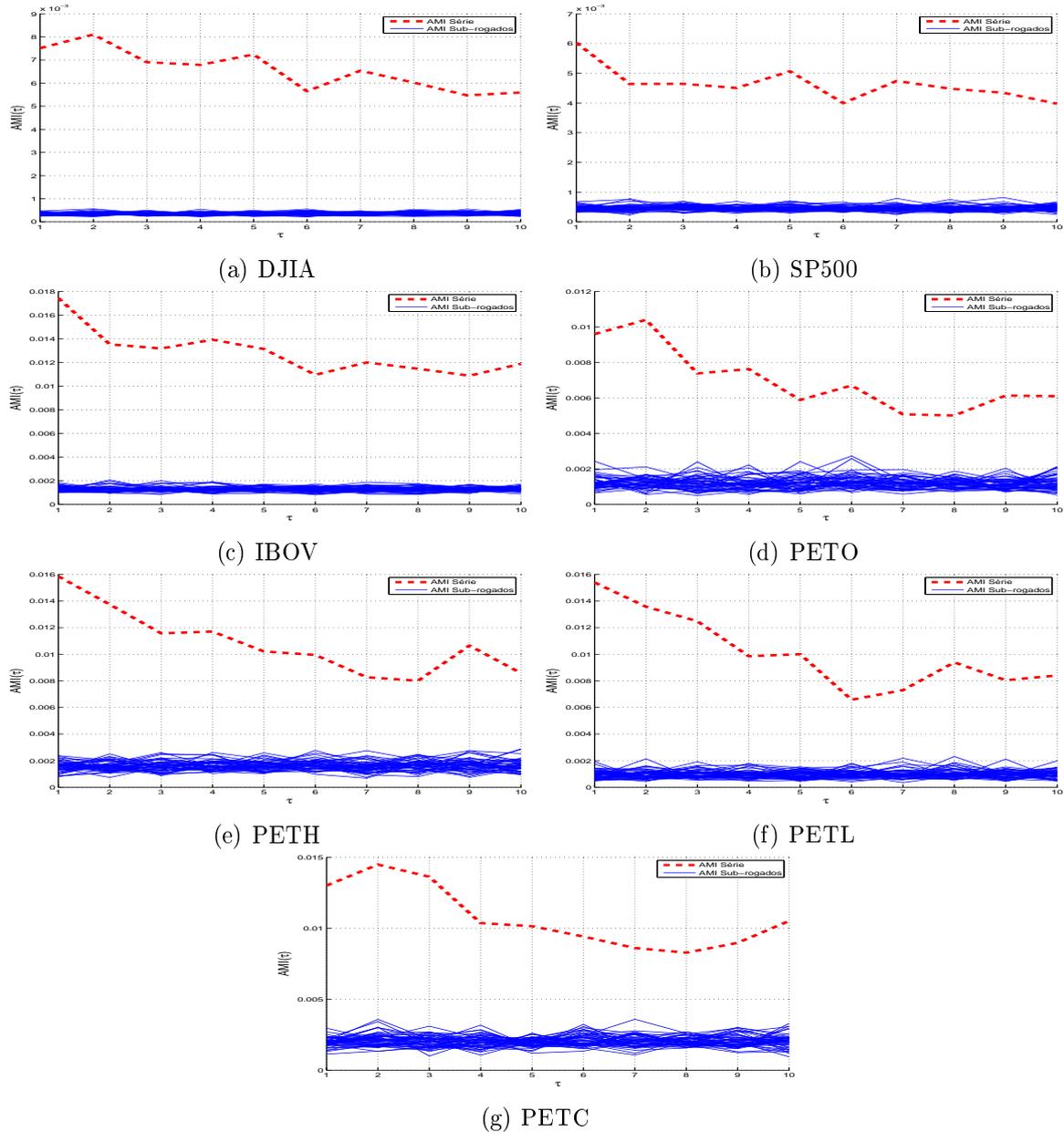


Figura 4.9. Análise de dados sub-rogados - algoritmo 0. A hipótese de independência pode ser rejeitada para todas as séries, utilizando a estatística AMI, mostrada nestas figuras.

Apenas o resultado obtido pela estatística AMI é mostrado, já que o mesmo é suficiente para rejeitar a hipótese de independência de todas as séries. A rejeição está de acordo com o resultado obtido pelo teste BDS e mostra que há algum determinismo nos dados.

Os valores do índice $PREV$ baseados na estatística AMI para esse algoritmo são mostrados na Tabela 4.2.

Tabela 4.2. Índice de previsibilidade (Equação 3.12) das séries para o algoritmo 0 utilizando AMI.

Série	Índice de previsibilidade $PREV(\text{Alg. 0, AMI, série})$
DJIA	122
SP500	68,0
IBOV	77,7
PETO	24,6
PETH	37,1
PETL	42,8
PETC	32,4

O tipo de determinismo não é identificado por esse algoritmo, mas o resultado já indica que uma modelagem dos retornos é válida e justificada. Isso está em oposição a HME e à utilização do modelo do passeio aleatório para essas séries. Esse resultado foi também obtido no Capítulo 2. Todavia, o resultado mostrado naquele capítulo considera que os retornos seguem uma distribuição gaussiana, o que coloca em dúvida a validade do seu resultado. Portanto, a rejeição da independência pela análise de dados sub-rogados, apresenta uma resposta mais geral (independente da distribuição) do que a obtida no Capítulo 2.

4.4 Dependência Linear dos Retornos

Na seção anterior pôde-se perceber a existência de dependências temporais em todas as séries estudadas. O próximo passo a ser dado é a investigação do tipo de dependência existente. Nesta seção, testa-se a hipótese de que as séries são originadas de um processo apenas linear. Sua rejeição direciona o estudo à identificação dos tipos de não linearidades que possam estar presentes. Isso porém, é feito em outra seção. O fato de as estatísticas utilizadas serem sensíveis às não linearidades é muito importante.

4.4.1 Teste BDS - Resíduos de um Modelo Linear

O teste BDS aplicado às séries indica a presença de algum determinismo mas não traz informações a respeito de seu tipo. Identificar se esse determinismo é linear ou não é bastante importante pois define caminhos completamente diferentes para uma

possível modelagem dos dados. Esta seção aplica o mesmo teste BDS nos resíduos de um modelo AR ajustado para cada uma das séries estudadas.

O modelo AR é estimado por meio do método de mínimos quadrados clássico. A ordem do modelo é escolhida como a menor ordem cujos resíduos não sejam rejeitados pela hipótese do teste Ljung-Box. Isso é feito da seguinte forma:

- inicia a escolha da ordem k com a estimação de um modelo de ordem unitária. Ou seja, AR(k) sendo $k = 1$;
- obtem-se a estatística $q(20)$ do teste Ljung-Box dos resíduos do modelo AR(k);
- compara-se o valor de $q(20)$ com o valor crítico do teste considerando um intervalo de confiança de dois desvios padrão. Se $q(20)$ for maior que o valor crítico aumenta-se em uma unidade a ordem e estima-se, então, o modelo AR($k + 1$);
- continua-se o procedimento acima até que $q(20)$ seja menor que o valor crítico. Quando isso acontecer, o k correspondente será a ordem suficiente para que os resíduos sejam descorrelacionados (com o intervalo de confiança acima).

O intuito dessa análise de resíduos é testar a presença de dependências temporais que não foram modeladas pelo modelo linear. Pode-se comparar, então, a série de resíduos com versões dessa série em que suas observações são colocadas de maneira aleatória. Esse é o mesmo procedimento realizado pelo algoritmo 0 de geração de dados sub-rogados, em que as séries criadas são permutações da série original, eliminando qualquer estrutura dinâmica que possa existir. A comparação é feita em termos da estatística W_2 . Se a série original for estatisticamente diferente das séries criadas, pode-se concluir que existe alguma dinâmica não linear passível de ser modelada.

A Figura 4.10 exhibe os resultados do procedimento acima comparando W_2 das séries com as séries geradas correspondentes.

O valor de W_2 de todas as séries ultrapassam o valor limite de 1,96. Esse valor é definido pelo percentil 95 da distribuição acumulada de W_2 em dados independentes. A Tabela 4.3 resume os resultados anteriores obtidos e exhibe também a ordem dos modelos AR criados bem como o índice $PREV$ calculado por meio de W_2 . O teste de normalidade indicou para todas as séries que a distribuição de W_2 podem ser gaussiana, o que melhora a interpretação do índice $PREV$.

A tabela não mostra valores muito diferentes do que os encontrados na Seção 4.3.1. Isso sugere que a “dinâmica” principal existente nos dados não foi capturada pelo modelo linear. Logo, justifica-se uma modelagem não linear dos retornos a fim de melhor mapear o determinismo existente.

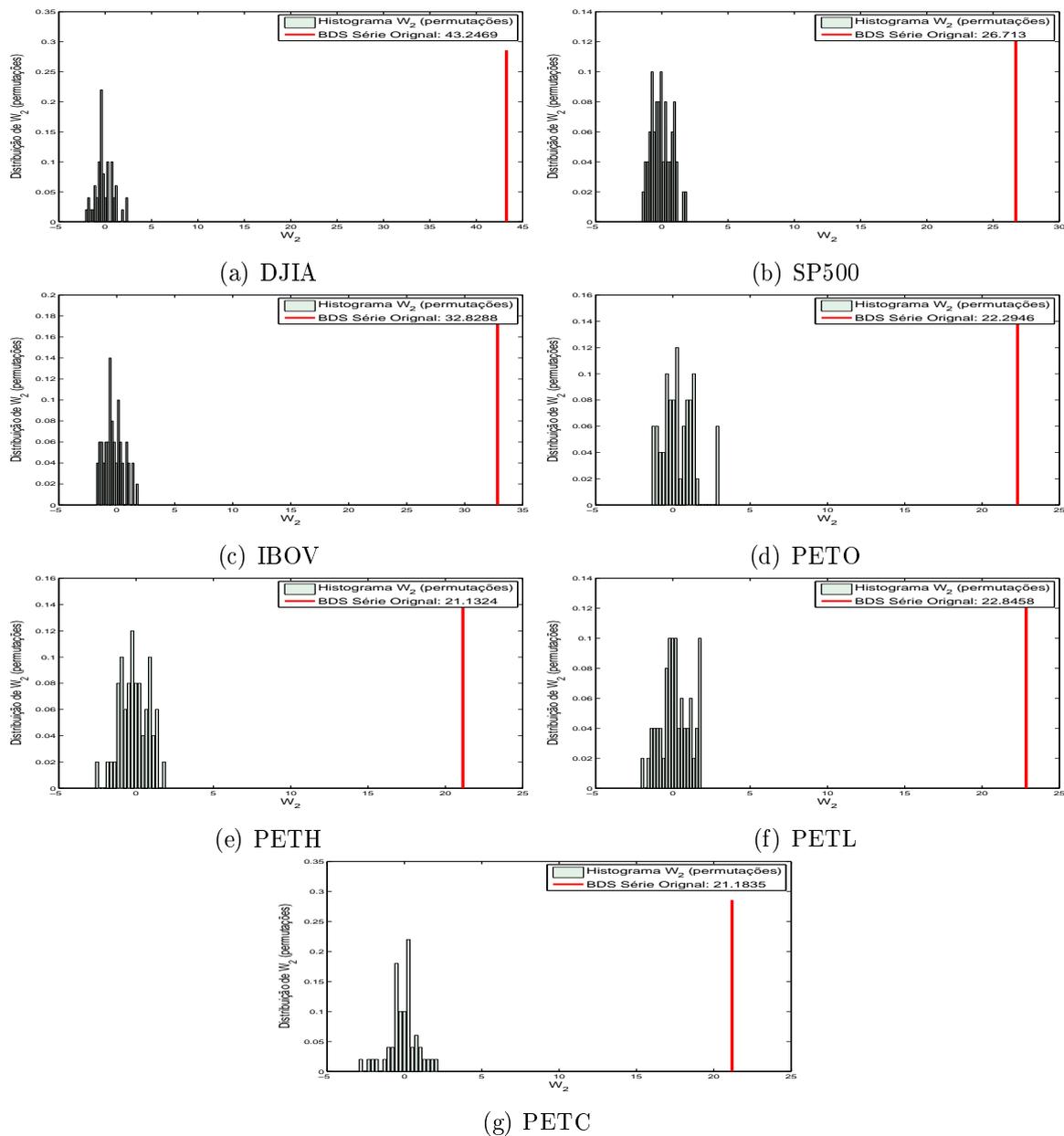


Figura 4.10. Teste BDS de dependência linear. A estatística W_2 é calculada para os resíduos de modelos AR das séries. Pode-se rejeitar a hipótese de independência dos resíduos para todas, já que ultrapassam o valor limite de 1,96.

4.4.2 Dados Sub-rogados - Algoritmo 1

O algoritmo 1 gera dados através das séries e semelhantes às mesmas (mesma sequência de correlação) mas sem dependências não lineares. Se as séries apresentarem tais dependências, suas estatísticas discriminantes se diferenciarão de seus sub-rogados. O teste é realizado seguindo o mesmo procedimento apresentado na Seção 3.6 e procura-se identificar se os dados são apenas lineares. Ressalta-se mais uma vez o fato de a

Tabela 4.3. W_2 e índice de previsibilidade dos resíduos AR das séries para o teste BDS (Equação 3.12) de dependência linear. Valores de W_2 superiores a 1,96 indicam a rejeição da hipótese. A tabela também mostra a ordem do modelo AR ajustado para cada série e o resultado do teste Kolmogorov-Smirnov (KS) da distribuição empírica de W_2 (valores superiores a 0,1884 indicam a não normalidade da distribuição de W_2).

Série	$ W_2 $	Valor limite de W_2	Índice de previsibilidade $PREV(BDS, \text{série})$	Ordem do modelo AR	Teste KS
DJIA	43,2	1,96	45,7	6	0,0767
SP500	26,7	1,96	27,4	15	0,108
IBOV	32,8	1,96	31,8	15	0,100
PETO	22,3	1,96	20,13	3	0,122
PETH	21,1	1,96	21,08	2	0,154
PETL	22,8	1,96	22,2	6	0,0887
PETC	21,2	1,96	21,4	3	0,0912

hipótese a ser testada é a de que há apenas dependências lineares gaussianas. Como frisado em Hinich *et al.* [2005], uma rejeição do teste não necessariamente indica a presença de não linearidades, uma vez que o teste indica rejeição em casos de algumas dependências lineares não gaussianas. Esse fato foi exemplificado na Seção 3.6.3 com uma série derivada de um processo linear com inovações uniformes (MA(1) uniforme).

Os resultados são mostrados por meio das Figuras 4.11 e 4.12. A primeira mostra a AMI de todas as séries exceto PETO. Essa série é mostrada em separado, na Figura 4.12, sendo exibidas as estatísticas AMI e $SampEn(2, \epsilon, 3903)$. 3903 é o número de dados da série de retorno PETO, ϵ corresponde a 20% de seu desvio padrão e 2 é a dimensão de imersão.

Pelas Figuras 4.11 e 4.12.b, pode-se rejeitar a hipótese de linearidade para todas as séries. Para PETO, pode-se perceber mais claramente a rejeição utilizando-se a entropia amostral. A Tabela 4.4 mostra os resultados quantitativos apresentados pelas figuras.

O valor baixo de $PREV$ para PETL pode ser melhor interpretado pela Figura 4.11.e. Nesse caso o primeiro atraso de AMI se mistura ao valor de seus subrogados. Mas os atrasos 2, 3, 4 e 5 evidenciam a rejeição da hipótese. Isso também acontece, embora em menor escala, para a série PETC. Com isso pode-se concluir que as séries estudadas *não são geradas por um processo gaussiano linear*.

A vantagem desse teste é detectar as não linearidades presentes sem necessitar de um modelo específico para esse fim. Uma rejeição utilizando o algoritmo 1 é, portanto, mais geral que, por exemplo, os testes de White e Breusch-Pagan e apontam para dependências que não puderam ser detectadas utilizando as técnicas do Capítulo 2.

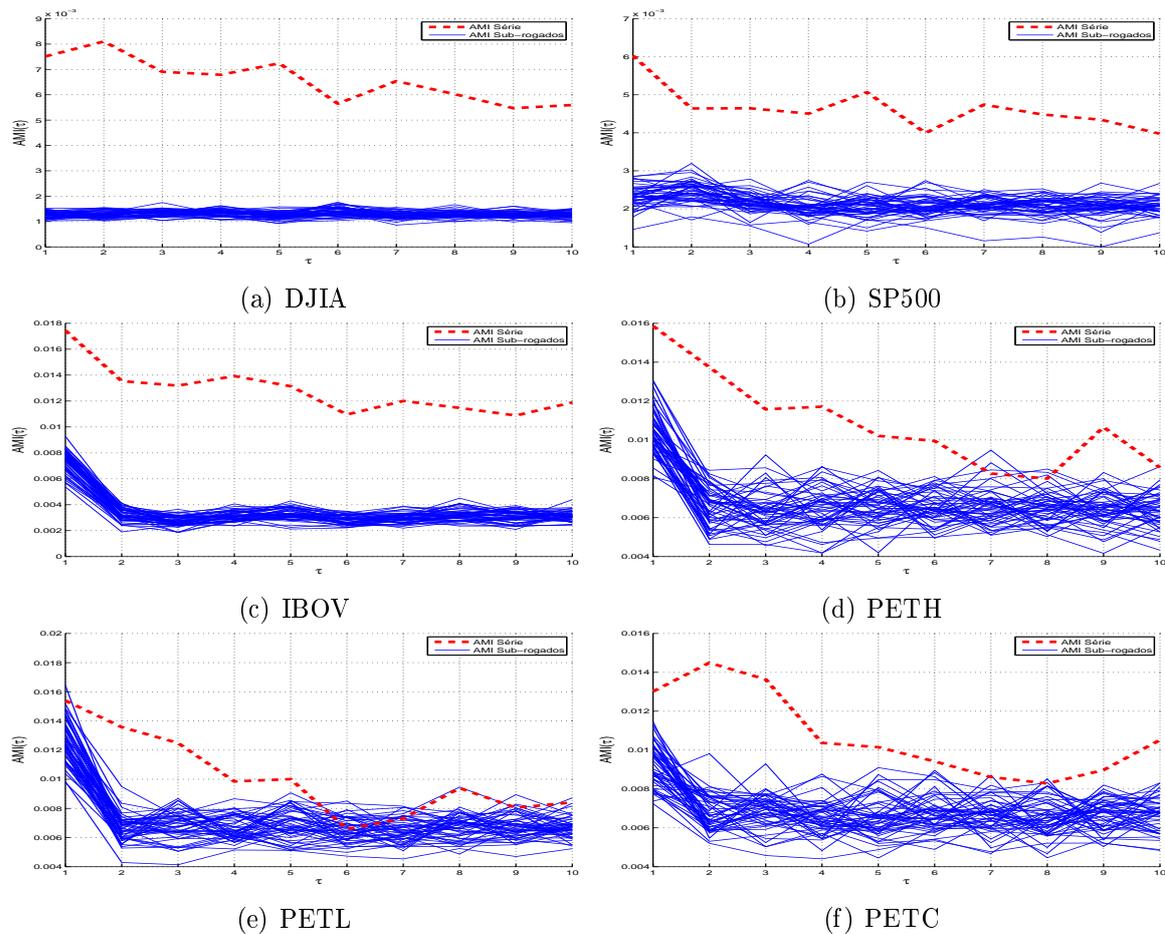


Figura 4.11. Análise de dados sub-rogados - algoritmo 1. Utilizando a AMI, a hipótese de linearidade é rejeitada em todos os casos mostrados.

4.5 Dependência Não Linear dos Retornos

Essa seção apresenta testes cujo objetivo principal é verificar a existência de heteroscedasticidade nos dados. Esses testes são de grande importância já que todo determinismo não linear detectado nas seções anteriores pode vir apenas dessa característica, o que, se confirmado, tornaria inútil uma modelagem não linear dos retornos. Ou seja, a única dependência temporal existente seria a dependência heteroscedástica. Essa característica pode mascarar os testes de previsibilidade indicando erroneamente a presença de dinâmicas previsíveis (no sentido colocado pela Seção 3.1).

4.5.1 Teste BDS - Resíduos de um Modelo AR-GARCH

Os resíduos obtidos por meio da modelagem AR realizada na Seção 4.4.1 mostraram uma dependência não linear significativa, como visto na Tabela 4.3. Baseado nos re-

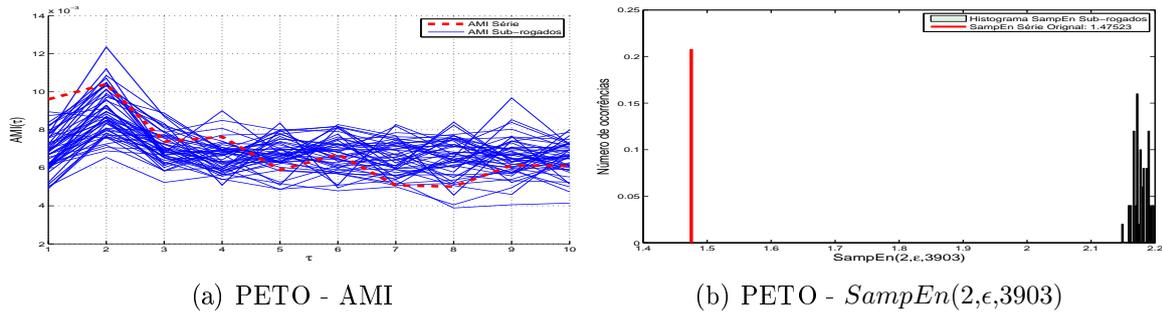


Figura 4.12. Análise de dados sub-rogados - algoritmo 1. AMI e $SampEn(2, \epsilon, 3903)$ são calculados para PETO e seus sub-rogados. Não se pode rejeitar a hipótese de linearidade de maneira clara utilizando-se apenas AMI. Em contrapartida, a $SampEn$ da série é bastante diferente dos dados gerados, rejeitando-se assim essa hipótese.

Tabela 4.4. Índice de previsibilidade (Equação 3.12) das séries para o algoritmo 1 utilizando AMI.

Série	Índice de previsibilidade $PREV(\text{Alg. 1, AMI, série})$
DJIA	45,3
SP500	13,2
IBOV	11,7
PETO	2,82
PETH	4,81
PETL	1,69
PETC	3,44

sultados obtidos pelos testes Breusch-Pagan e White do Capítulo 2, parte dessa dependência pode ser devida à presença da heteroscedasticidade. A modelagem da mesma é, portanto, um passo importante na direção de um melhor entendimento do determinismo presente nos retornos. Nesse contexto, a análise dos resíduos da modelagem GARCH das séries pode apontar para a existência de estruturas previsíveis nas séries.

A análise desta seção parte dos resíduos dos modelos AR. Esses são modelados por uma estrutura GARCH(1,1) - Ver Equação 2.11. Os resíduos dessa última modelagem são calculados da seguinte maneira:

$$r_{GARCH}(t) = \frac{r_{AR}(t)}{\sigma_{GARCH}(t)}. \quad (4.1)$$

Na equação anterior, o resíduo GARCH $r_{GARCH}(t)$ é obtido por meio da normalização da série de resíduos $r_{AR}(t)$ pelo desvio padrão $\sigma_{GARCH}(t)$. Esse desvio é o obtido pelo modelo GARCH de variância:

$$r_{AR}(t) = \epsilon(t) \quad \epsilon(t) \sim N(0, \sigma_{GARCH}(t)^2)$$

$$\sigma_{GARCH}(t)^2 = \alpha_0 + \alpha_1 r_{AR}(t-1)^2 + \alpha_2 \sigma_{GARCH}(t-1)^2.$$

Após a estimação dos parâmetros α_0 , α_1 e α_2 , pode-se obter $\sigma_{GARCH}(t)^2$ e efetuar o cálculo da Equação 4.1. A série $r_{GARCH}(t)$ terá variância constante e não apresentará dependência linear. Essa última série é denominada resíduo AR-GARCH(1,1) dos retornos.

A Figura 4.13 mostra os resultados do teste BDS realizados após o procedimento descrito acima.

A figura indica a rejeição das hipóteses de independência para os resíduos AR-GARCH(1,1) dos retornos da Petrobras (PETO, PETH, PETL e PETC). Entretanto, essa rejeição não é detectada nas séries dos índices DJIA, SP500 e IBOV. Esse resultado é mostrado numericamente na Tabela 4.5. A hipótese de independência dos resíduos é aproximadamente equivalente a uma hipótese de dependência linear heteroscedástica dos retornos. A tabela também mostra o resultado do teste de normalidade da distribuição de W_2 , que não é rejeitado para nenhuma das séries.

Tabela 4.5. W_2 e índice de previsibilidade (Equação 3.12) dos resíduos AR-GARCH das séries para o teste BDS. Valores de W_2 superiores a 1,96 indicam a rejeição da hipótese de dependência linear heteroscedástica. A tabela também mostra a ordem da parte AR do modelo AR-GARCH(1,1) ajustado para cada série e o resultado do teste Kolmogorov-Smirnov (KS) da distribuição empírica de W_2 (valores superiores a 0,1884 indicam a não normalidade da distribuição de W_2).

Série	$ W_2 $	Valor limite de W_2	Índice de previsibilidade $\hat{PREV}(BDS, \text{série})$	Ordem do modelo AR	Teste KS
DJIA	0,210	1,96	0,266	6	0,140
SP500	0,906	1,96	1,14	15	0,156
IBOV	0,148	1,96	0,367	15	0,148
PETO	3,51	1,96	3,32	3	0,101
PETH	6,27	1,96	5,55	2	0,180
PETL	3,95	1,96	4,58	6	0,0917
PETC	3,45	1,96	3,64	3	0,0729

Pode-se perceber, comparando as Tabelas 4.3 e 4.5, que o valor do índice reduziu muito. Isso pode indicar que parte do determinismo detectado na Tabela 4.3 foi capturado pela modelagem GARCH. Ou seja, poderia-se inferir que a principal contribuição para detecção de determinismo tem como origem a heteroscedasticidade, que,

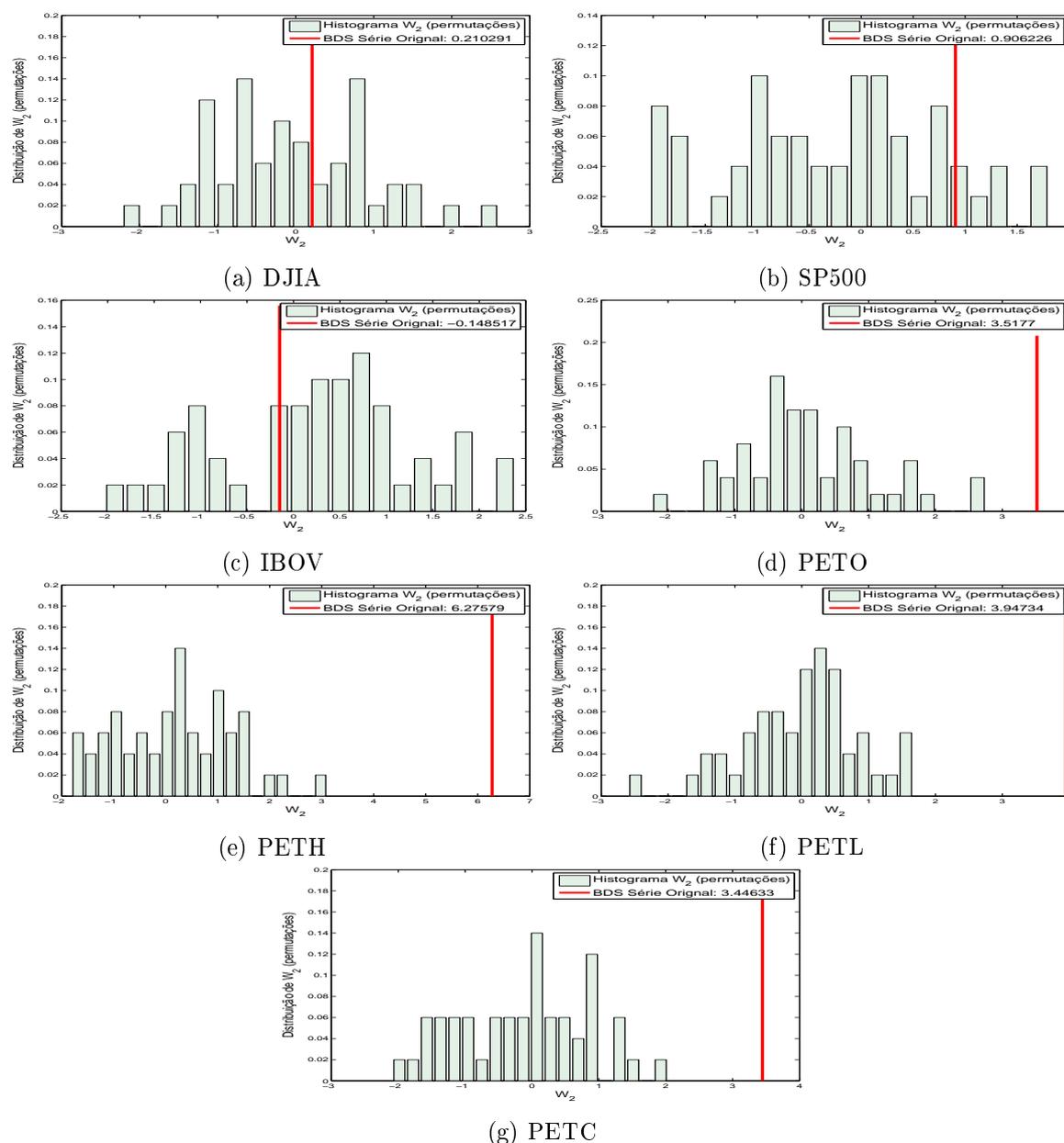


Figura 4.13. Teste BDS de dependência não linear. A estatística W_2 é calculada para os resíduos dos modelos AR-GARCH(1,1) das séries.

como ressaltada em vários pontos do trabalho, não contribui para a previsibilidade dos retornos.

Deve-se chamar atenção, porém, de que o teste não avalia uma classe de modelos como no caso da análise de dados sub-rogados. São considerados aqui os resíduos de um determinado modelo, e não se pode generalizar o resultado para toda uma classe. Não se pode afirmar que uma rejeição do teste indica a rejeição de modelos GARCH de maneira geral. O que se testa é a aderência das séries a um modelo GARCH(1,1)

específico com determinados parâmetros. Nesse sentido, a abordagem de dados sub-rogados traz hipóteses muito mais gerais. A análise de dados sub-rogados não lineares é, portanto, o próximo passo dos testes deste trabalho.

4.5.2 Dados Sub-rogados - Algoritmo 2

A primeira análise de dados sub-rogados a ser feita nesta seção é a do algoritmo 2. Esse gera dados lineares transformados estaticamente de maneira não linear. Pode-se considerar que essa análise verifica se a existência de não linearidades nos dados é de origem estática ou dinâmica. Como mostrado na Seção 3.6.3, a rejeição da hipótese testada por esse algoritmo aponta para a existência de estruturas dinâmicas não lineares ou a presença de não estacionariedade. Essa última hipótese alternativa pôde ser vista pela rejeição da série heteroscedástica ARCH(1) naquela seção⁶.

A distribuição da estatística AMI dos dados gerados pelo algoritmo bem como das séries pode ser vista na Figura 4.14.

Por meio da AMI, pode-se rejeitar a hipótese de que os dados são gerados por um processo linear e transformados estaticamente. Como foi possível rejeitar a hipótese utilizando-se a AMI a Tabela 4.6 mostra o índice *PREV* calculado baseado nessa estatística.

Tabela 4.6. Índice de previsibilidade (Equação 3.12) das séries para o algoritmo 2 utilizando AMI.

Série	Índice de previsibilidade <i>PREV</i> (Alg. 2, AMI, série)
DJIA	45,3
SP500	13,2
IBOV	11,7
PETO	2,82
PETH	4,81
PETL	1,69
PETC	3,44

Todas as séries apresentam alguma estrutura dinâmica não linear. Os valores da tabela são calculados utilizando-se o primeiro atraso da AMI. Embora alguns valores sejam pequenos, pode-se, com o auxílio das figuras, perceber que os outros atrasos contribuem para a rejeição da hipótese. Nesse caso, uma abordagem mais correta em termos de se quantificar a previsibilidade é o de considerar em *PREV* todos os atrasos.

⁶Como a não estacionariedade é consequência de uma dependência não linear (como visto na Equação 2.10), as séries heteroscedásticas como ARCH ou GARCH são consideradas como possuindo “dinâmica” não linear.

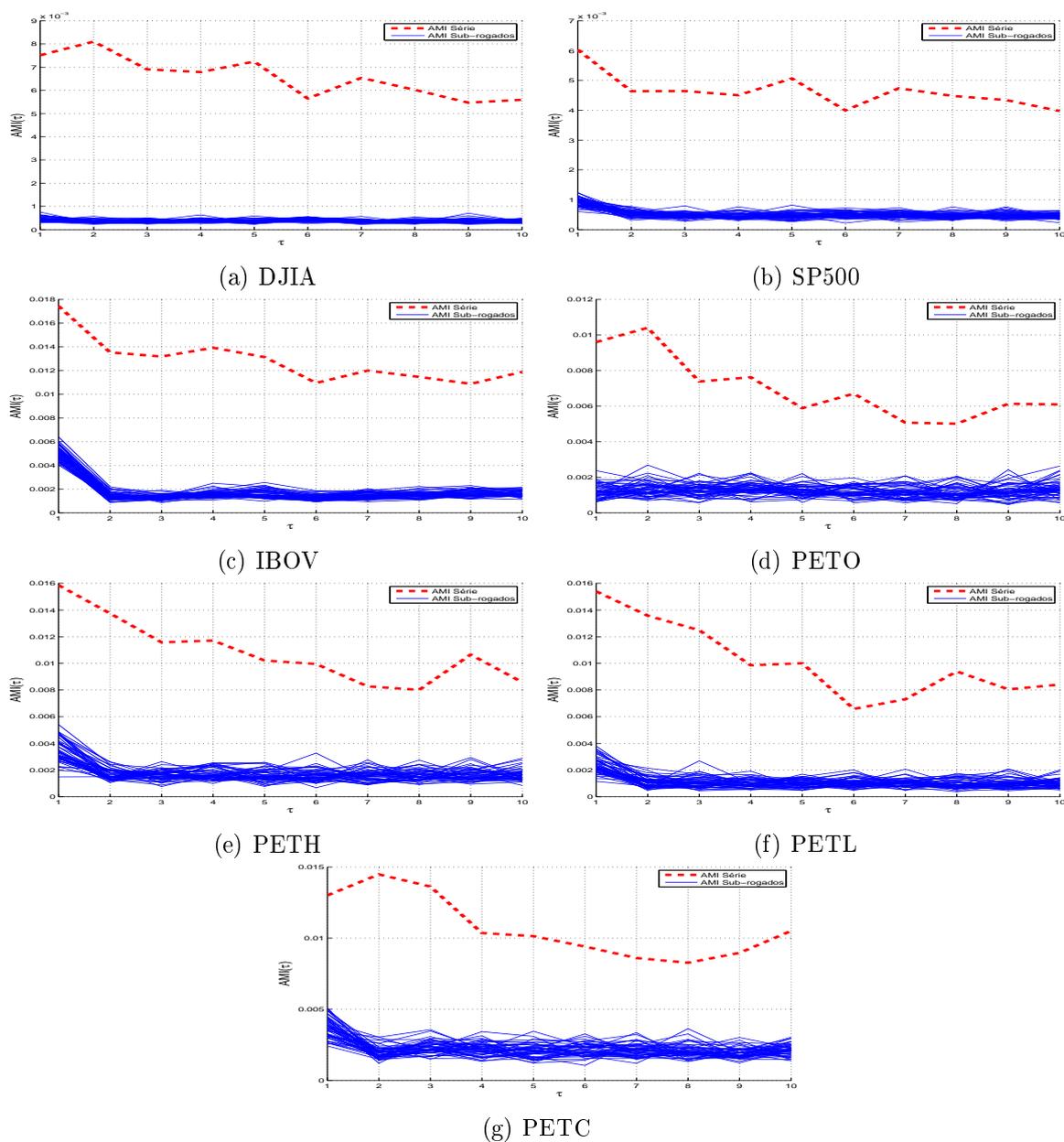


Figura 4.14. Análise de dados sub-rogados - algoritmo 2. As figuras (a) a (g) mostram que a estatística AMI das séries (traço pontilhado) difere dos dados gerados que atendem a hipótese nula. Essa hipótese é, portanto, rejeitada. Ou seja, Os dados não são lineares transformados estaticamente de maneira não linear.

Todavia, esse resultado ainda não corresponderia à previsibilidade considerado no trabalho, uma vez que leva em consideração “dinâmicas” devidas a heteroscedasticidade. A próxima seção mostra os resultados da análise de dados sub-rogados utilizando um algoritmo para tratar esse último caso.

4.5.3 Dados Sub-rogados - Algoritmo PPS

Essa seção trata da aplicação do algoritmo apresentado na Seção 3.6.4 às séries estudadas. O procedimento seguido é o mesmo das outras análises dessa seção, mas a hipótese a ser testada é mais específica: investiga-se aqui a presença de heteroscedasticidade, como feito em Small & Tse [2003b]. A hipótese PPS é considerada então como um teste de adequação das séries a processos GARCH.

As séries são comparadas com sub-rogados para que se obtenha a distribuição empírica das estatísticas discriminantes, como nos algoritmos 0, 1 e 2. As Figuras 4.15, 4.16 e 4.17 apresentam os resultados gráficos dessa comparação.

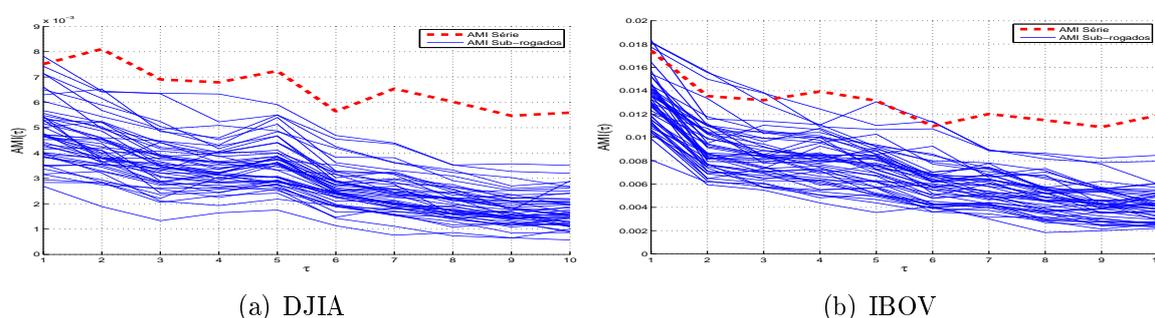


Figura 4.15. Análise de dados sub-rogados - algoritmo PPS. A hipótese testada pela geração de dados PPS pode ser rejeitada para DJIA e IBOV quando se considera principalmente os atrasos de 4 a 10 da estatística AMI. Isso indica presença de não linearidades que não sejam originadas por heteroscedasticidade.

O teste rejeita a hipótese PPS para DJIA e IBOV utilizando a estatística AMI. Pode-se perceber a rejeição da hipótese para valores de τ de 4 a 10. Nessa faixa de valores a AMI dessas duas séries é claramente diferente da AMI das séries geradas pelo algoritmo PPS.

As séries PETO e PETL são também rejeitadas e, conseqüentemente, apresentam dinâmicas não lineares diferentes de um processo GARCH. Para essas duas séries, porém, a rejeição se dá por meio da dimensão de correlação D_2 . Tanto AMI quanto *SampEn* não conseguem rejeitar a hipótese para essas séries.

Para as séries SP500, PETH e PETC não se pode rejeitar a hipótese PPS utilizando qualquer uma das três estatísticas apresentadas neste trabalho.

Os casos de não rejeição podem indicar que as séries são realmente originadas por um processo GARCH ou, apenas, que as estatísticas aplicadas não foram capazes de discernir as séries de seus sub-rogados. Nesse caso, seria necessária alguma outra estatística capaz de realizar essa distinção. Os testes baseados em dados sub-rogados sempre estarão limitados à capacidade das estatísticas utilizadas de discernir estruturas dinâmicas. Logo, a não rejeição traz poucas informações a respeito das características

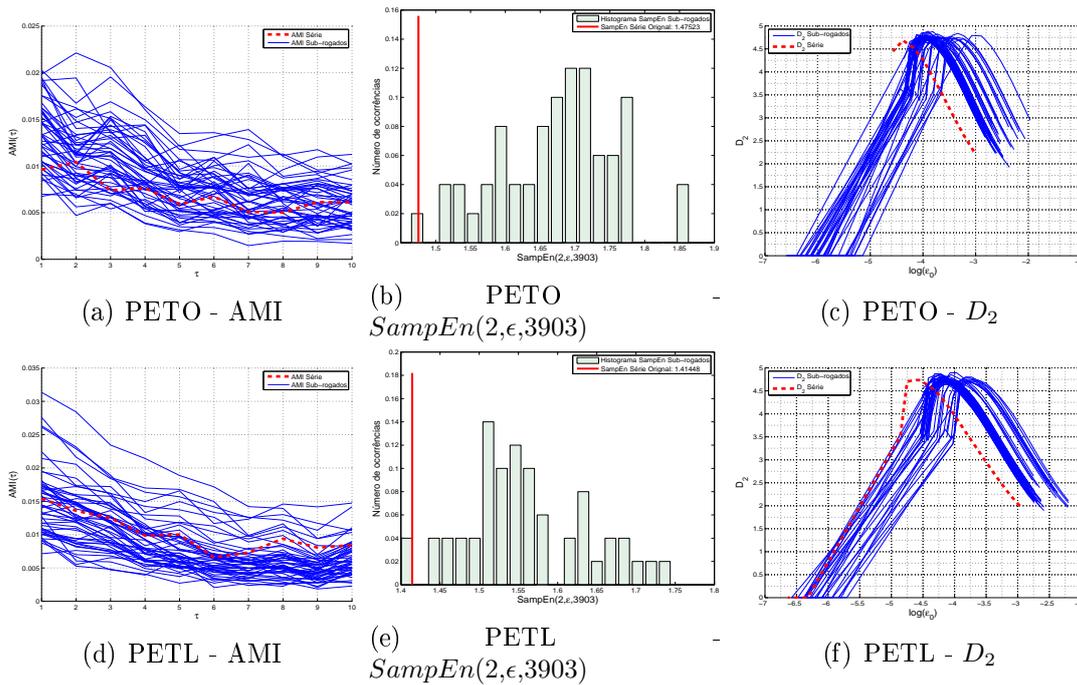


Figura 4.16. Análise de dados sub-rogados - algoritmo PPS. As estatísticas AMI e $SampEn$ não conseguem rejeitar a hipótese PPS para as séries PETO e PETL. Entretanto, a dimensão de correlação D_2 permite a rejeição indicando a presença de estruturas determinísticas diferentes de processos GARCH. Isso não significa a inexistência dos efeitos de heteroscedasticidade mas a presença de outros efeitos não lineares.

dinâmicas da série. Entretanto, a rejeição da hipótese, nos casos de DJIA, IBOV, PETO e PETC, indica a presença de estruturas que justificam uma modelagem não linear dos retornos.

4.6 Quantificação da Previsibilidade

Dos métodos apresentados anteriormente, o teste BDS de resíduos é o que mais facilmente pode ser traduzido em um índice de previsibilidade, uma vez que a análise de dados sub-rogados apresenta desempenho diferente para cada uma das estatísticas discriminantes. Além disso, comparar índices de previsibilidade baseados em estatísticas diferentes pode induzir a erros. O índice PREV, como mostrado na Equação 3.12, pode quantificar a previsibilidade das séries quando contabiliza apenas estruturas que contribuem para uma melhor previsão. Isso exclui as dependências identificadas pela modelagem GARCH. O que se evidencia nesta seção é que a existência de dependência temporal e determinismo é diferente da existência de estruturas previsíveis.

Esta seção aplica o teste BDS e calcula o índice PREV em séries filtradas por

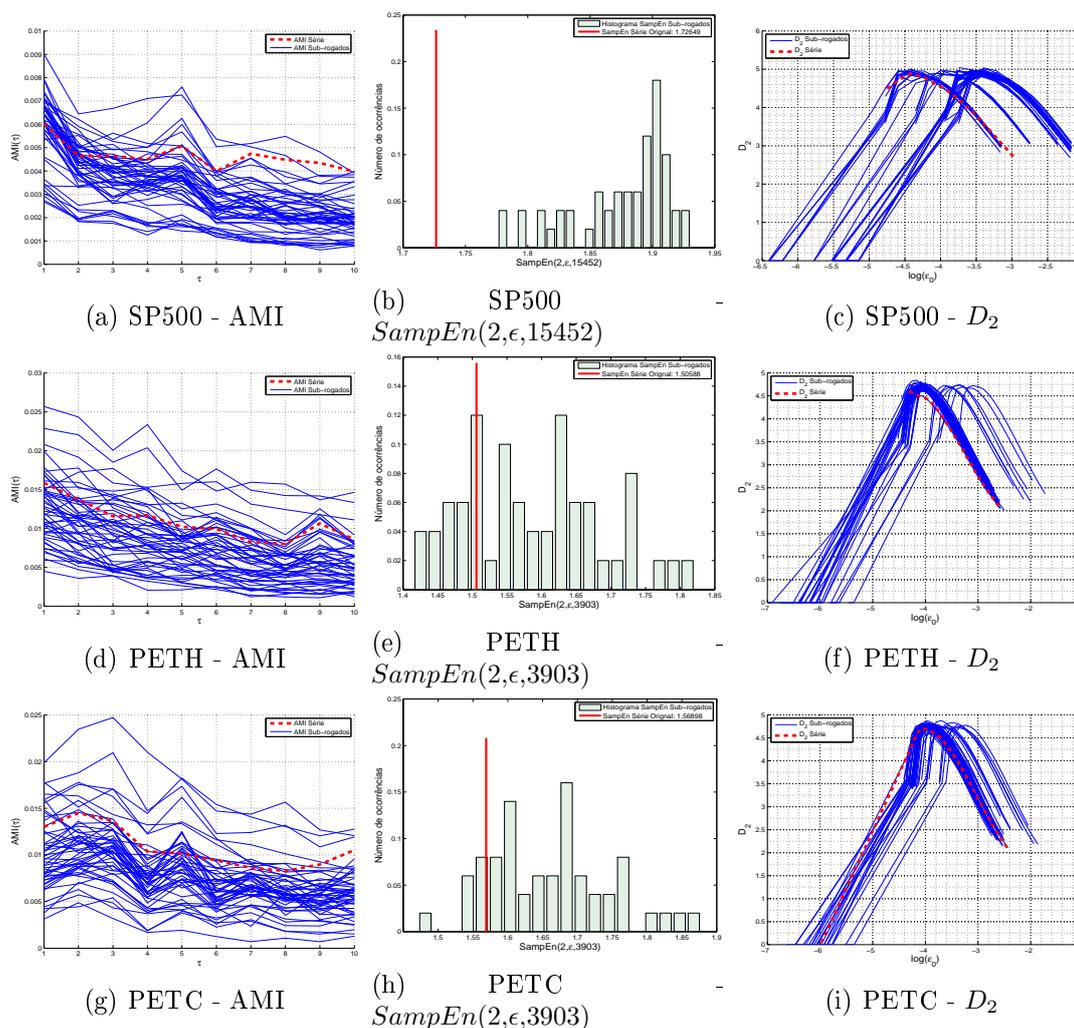


Figura 4.17. Análise de dados sub-rogados - algoritmo PPS. Para as séries SP500, PETH e PETC, não se pode rejeitar a hipótese de que tenham sido originadas por processos GARCH. Ou seja, por meio dessa análise de dados sub-rogados, o determinismo não linear presente nessas séries não contribui para a previsibilidade das mesmas.

modelos GARCH aplicados às séries estudadas. A série resultante mantém sua dependência temporal anterior mas é homoscedástica. O teste BDS dessas séries filtradas pode, então, ser entendido como um teste da hipótese de dependência heteroscedástica. Dados que tenham outra dependência temporal que não essa não estão de acordo com tal hipótese e isso justifica uma modelagem desses dados com o objetivo de se prever retornos futuros. Obtém-se, então uma medida de PREV que é robusta a tal não estacionariedade.

Se depois da modelagem GARCH, W_2 é tal que detecta determinismo, então tal determinismo não é devido à heteroscedasticidade. Ou seja, H_0 remete a séries com

dependência exclusivamente heteroscedástica e que, a parte dessa possível propriedade, eles não possuem determinismo. Essa hipótese será chamada H_{het} para evidenciar essa característica.

A Figura 4.18 mostra os resultados do teste BDS em séries filtradas por modelos GARCH(1,1).

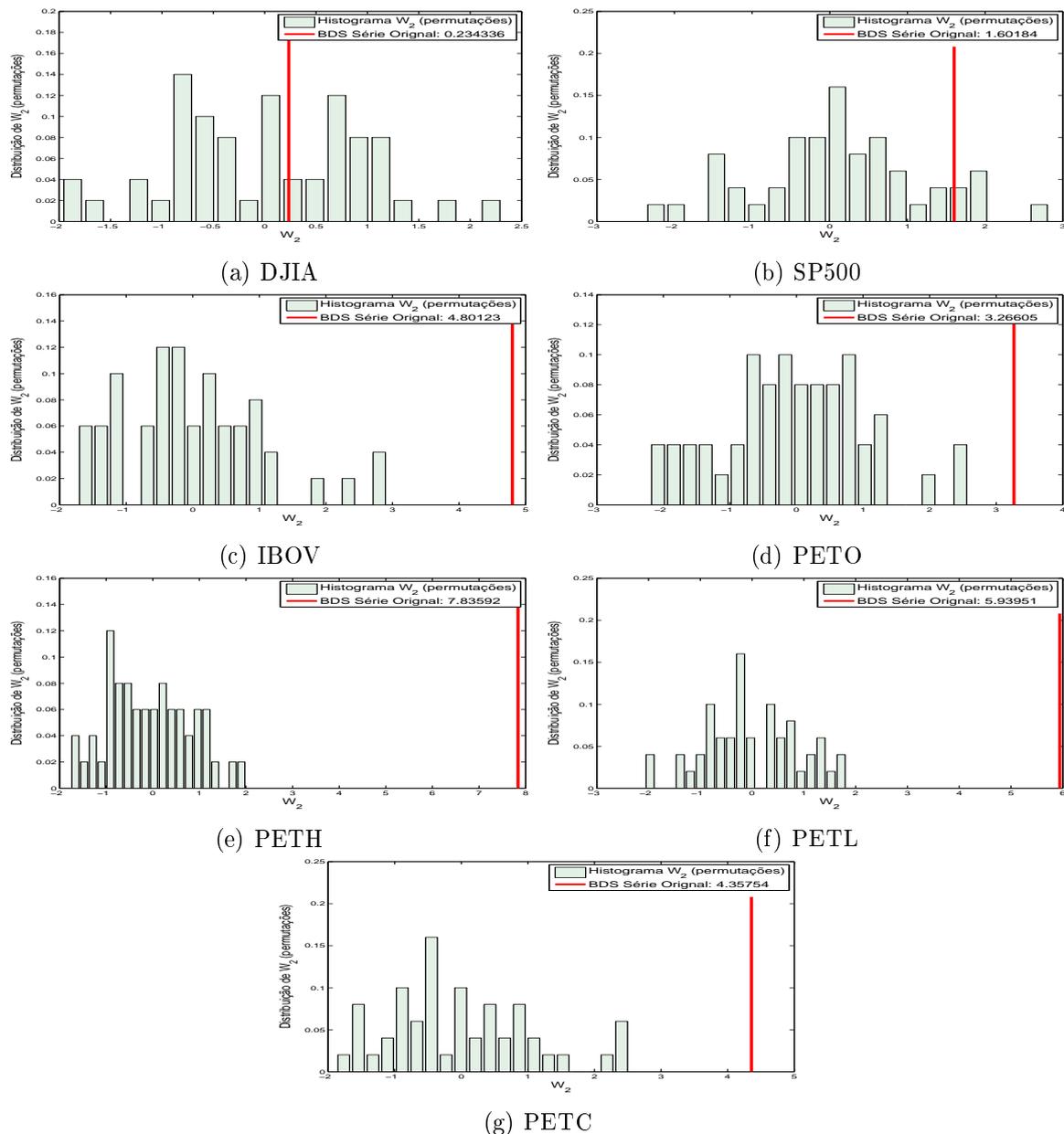


Figura 4.18. Teste BDS de dependência exclusivamente heteroscedástica (H_{het}). A estatística W_2 é calculada para as séries filtradas por GARCH(1,1).

As únicas séries das quais não se pode rejeitar H_{het} são DJIA e SP500. Isso não significa que essas séries não possuam outras dependências. Mas após a filtragem

GARCH, elas não são significativas, o que sugere que uma modelagem dos retornos para previsão seria inútil. As outras séries apresentam valores de $|W_2|$ mais elevados mas, comparando-se esses resultados com os da Tabela 4.1, percebe-se que estão próximos do valor limite. A maior parte do determinismo presente foi modelada pelo modelo GARCH.

A Tabela 4.7 mostra os resultados numéricos vistos nos gráficos anteriores.

Tabela 4.7. W_2 e índice de previsibilidade das séries filtradas por GARCH para o teste BDS. Valores de W_2 superiores a 1,96 indicam a rejeição de H_{het} . A tabela também mostra o resultado do teste Kolmogorov-Smirnov (KS) da distribuição empírica de W_2 (valores superiores a 0,1884 indicam a não normalidade da distribuição de W_2).

Série	$ W_2 $	Valor limite de W_2	Índice de previsibilidade $PREV(BDS, \text{série})$	Teste KS
DJIA	0,234	1,96	0,350	0,0950
SP500	1,60	1,96	1,40	0,123
IBOV	4,80	1,96	5,94	0,128
PETO	3,27	1,96	3,62	0,0887
PETH	7,84	1,96	8,61	0,190
PETL	5,94	1,96	5,65	0,0720
PETC	4,36	1,96	4,93	0,0770

O teste KS mostra a rejeição da hipótese de normalidade de PETH filtrada. Porém, esse resultado não invalida o uso do índice PREV, pois o gráfico da Figura 4.18 mostra W_2 da série filtrada bastante afastado da distribuição de W_2 das séries geradas por permutações. Pelos resultados, espera-se que essa série tenha um bom desempenho de previsão. Por outro lado, espera-se também que DJIA não apresente qualquer vantagem de previsão sobre um passeio aleatório. Mostrou-se em seções anteriores que DJIA e SP500 são diferentes desse modelo nulo. Mas essa diferença não necessariamente implica qualquer melhora na previsibilidade dessas séries, uma vez que o determinismo presente não se traduz em previsibilidade.

A Previsão como uma Validação das Técnicas de Previsibilidade

Os capítulos anteriores trataram da previsibilidade de séries financeiras. Várias técnicas foram apresentadas com o objetivo de quantificar essa característica. Pôde-se perceber a influência da heteroscedasticidade em todas as estatísticas discriminantes apresentadas. A fim de contornar esse problema, o teste BDS, por exemplo, foi aplicado aos resíduos de um modelo GARCH das séries (filtragem GARCH). Os valores de W_2 , obtidos para esse teste puderam ser resumidos em um índice de previsibilidade, cujo valor pode justificar ou não alguma tentativa de previsão. Esse valor indica um limite máximo para o desempenho de um modelo de previsão, o potencial que uma determinada série tem para ser prevista.

Os resultados das técnicas discutidas, no entanto, devem passar por algum tipo de validação. Esses devem, afinal, refletir a capacidade que um modelo tem de prever valores futuros. Ou seja, é de se esperar que, em um conjunto de dados no qual o índice de previsibilidade (descontados os efeitos da heteroscedasticidade) é elevado, um modelo de previsão seja capaz de obter, em média, um menor erro de previsão do que outro conjunto cujo índice seja pequeno.

Vários modelos baseados em diferentes estruturas têm sido propostos para a previsão de séries financeiras. Como exemplo, tem-se o *Threshold Autoregressive* e *Smooth Transition Autoregressive*, ambos baseados em variações de modelos lineares, encontrados em Tsay [2005]. Como visto na Seção 1.2, a utilização de alguns modelos baseados em lógica nebulosa, algoritmos genéticos, redes neurais ou modelos polinomiais não lineares tem se mostrado bem sucedida.

Como citado em Chen & Billings [1992], algumas redes neurais podem ser consideradas aproximadores universais de funções. Dentre essas, as redes baseadas em funções radiais conhecidas como *Radial Basis Function Networks*, ou redes RBF, apresentam rápida taxa de convergência no mapeamento não linear de funções. Essas qualidades mostram-se bastante interessantes na área de identificação de sistemas, principalmente

em problemas complexos em que não se tem muita informação a respeito da dinâmica presente nos dados. Esse é, precisamente, o caso das séries temporais financeiras.

O presente capítulo mostra o desenvolvimento de redes RBF para previsão de valores futuros das séries financeiras aqui estudadas. Uma pequena discussão a respeito das medidas de erro é feita visando uma melhor comparação entre o desempenho das redes em diferentes séries. A Seção 5.3 aponta as opções feitas em relação ao dimensionamento das redes, chamando atenção para a necessidade de se fazer com que os modelos estimados sejam bastante semelhantes, priorizando a avaliação da previsibilidade dos dados. Por fim, são apresentados os resultados da estimação dos modelos para as sete séries estudadas neste trabalho.

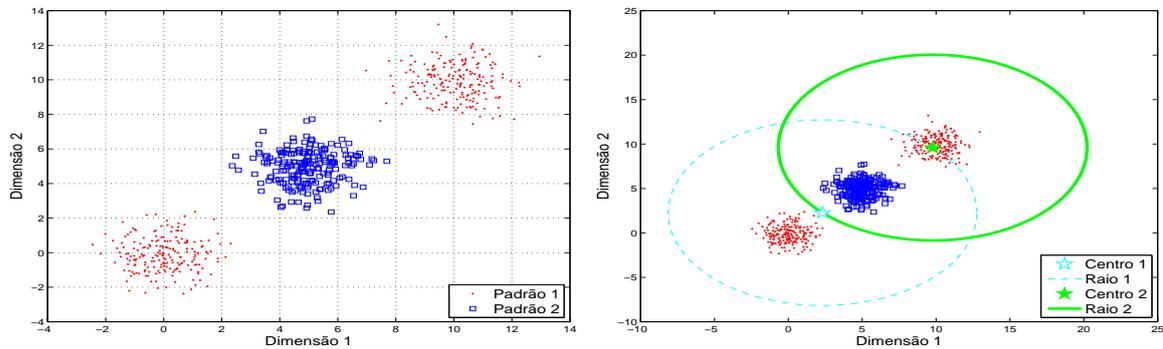
5.1 Modelo de Previsão

O objetivo do capítulo é o de desenvolver um modelo de previsão que seja capaz de capturar “dinâmicas não lineares” eventualmente presentes nos dados. Escolhe-se uma metodologia de treinamento comum para todas as sete séries a fim de compará-las em termos de previsibilidade. Para tanto, foi escolhido como modelo a rede neural baseada em funções de base radial.

As redes neurais *Radial Basis Function* (RBF) são aproximadores universais de funções (Haykin [2001]). Têm geralmente apenas uma camada escondida e uma camada de saída e funções bem definidas em cada uma delas, diferentemente de uma rede MLP. Seu treinamento é, tipicamente, mais rápido do que a MLP apesar de, na maioria das vezes necessitar de uma maior quantidade de neurônios. Outra vantagem das redes RBF é possuir menos parâmetros livres como, por exemplo, taxas de aprendizagem e quantidade de épocas de treinamento. Além disso, o treinamento e o papel desempenhado por cada uma de suas camadas é muito bem definido, o que facilita a construção de modelos que possam ser comparáveis entre si.

A camada escondida faz um mapeamento não linear do espaço de entrada por meio de funções radiais (na maioria dos casos, gaussianas), fazendo com que sua saída seja linearmente separável. Geralmente, um processo de aprendizado não supervisionado é aplicado a essa camada, com o objetivo de se ter neurônios que representem a distribuição dos dados no espaço de entrada. Isso se traduz na identificação da quantidade, localização e grau de abrangência dos neurônios. Esses mapeiam as regiões representativas de cada padrão no espaço de entrada. Como os neurônios dessa camada mapeiam regiões radiais, sua representação se dá pelo centro da região e seu campo receptivo, ou raio. A Figura 5.1 exibe um exemplo de como essa camada realiza

a identificação de classes no reconhecimento de 2 padrões não linearmente separáveis.



(a) Padrões não linearmente separáveis.

(b) Neurônios da camada escondida RBF.

Figura 5.1. O gráfico (a) mostra dois padrões em um espaço de entrada de duas dimensões. A camada escondida RBF consiste na alocação de neurônios (centros e raios) que mapeiam esse espaço de entrada. O gráfico (b) mostra os dois neurônios gerados para essa camada.

Na Figura 5.1, o espaço de entrada é mapeado por dois neurônios. Esses são, na realidade, pontos no espaço que representam o centro de uma função radial (gaussiana bidimensional), juntamente com o seu raio (seu desvio padrão).

No caso da previsão de séries temporais, é comum que se utilize como entrada valores anteriores da série. Essa será a abordagem usada neste trabalho, uma vez que o que se desejou medir nos capítulos anteriores foi exatamente a dependência temporal entre valores passados e futuros. O espaço de entrada pode ser visto como uma representação de cada uma das L entradas da rede em um espaço L -dimensional. O processo é semelhante à imersão de séries temporais em um espaço de fases. São gerados os vetores de entrada, por meio da série $y(t)$, para $L < t \leq N$:

$$\mathbf{x}_t = [y(t-1) \ y(t-2) \ \dots \ y(t-L)]^T. \quad (5.1)$$

A Equação 5.1 é semelhante a Equação 3.3 para $L = m + 1$. Nesse caso, a diferença está na ausência do elemento $y(t)$ no vetor \mathbf{x}_t . De fato, esse elemento não faz parte do vetor de entrada, mas é a saída desejada da rede. Utiliza-se, portanto, os L valores anteriores da série $[y(t-1) \ y(t-2) \ \dots \ y(t-L)]^T$ para se prever o valor atual $y(t)$. No exemplo da Figura 5.1, as dimensões 1 e 2 poderiam ser escolhidas como $y(t-1)$ e $y(t-2)$. Os padrões (pontos ou quadrados) seriam a resposta da rede, como, por exemplo, retornos futuros positivos ou negativos

O treinamento da camada escondida consiste na estimação dos centros e raios dos neurônios. Há diversos algoritmos para a estimação de centros. O utilizado neste trabalho é o k -médias, como descrito em Braga *et al.* [2007]. Esse separa o espaço de

entrada em L partições, cada uma com seu respectivo centro. Os dados de treinamento são alocados na partição mais próxima (de menor distância euclidiana em relação ao centro da partição). De maneira iterativa, o algoritmo modifica a posição dos centros, a fim de minimizar a soma das distâncias entre os dados de cada partição e seu centro. O raio de todos os neurônios é escolhido como o valor da menor distância euclidiana entre todos os centros. A escolha da quantidade de neurônios Q , bem como da quantidade de atrasos L , são descritas na Seção 5.3. Por hora, considera-se que esses números já são conhecidos, o que permite o devido treinamento de toda a rede.

A saída de cada neurônio da camada escondida é dada pela distância entre o mesmo e o vetor de entrada \mathbf{x} , ponderada pelo seu raio. Assim, a saída do n -ésimo neurônio com centro \mathbf{c}_n e raio σ_n , no instante t , é dada por:

$$y_{1,n}(t) = e^{-\frac{u_n(t)^2}{2\sigma_n^2}}, \quad (5.2)$$

em que

$$u_n(t) = \|\mathbf{x}_t - \mathbf{c}_n\|. \quad (5.3)$$

O subíndice 1 em $y_{1,n}(t)$ indica que essa é a saída de um neurônio n relativo à primeira camada (escondida) da rede. Logo, a saída dessa camada no instante t será um vetor $\mathbf{y}_1(t)$, tão longo quanto a quantidade de neurônios. As Equações 5.2 e 5.3 mostram que a saída do neurônio é inversamente proporcional a distância entre a entrada \mathbf{x}_t e seu centro. Para construir o modelo de previsão, será utilizado um mesmo valor de raio para todos os neurônios, o que, nas equações anteriores, pode ser representado por um σ_n constante para todo n .

Após o treinamento da camada escondida, a fim de se obter os centros e raios de seus neurônios, as saídas dessa camada são calculadas para todos os vetores de entrada (Equações 5.2 e 5.3). Cada vetor de entrada é mapeado em um vetor de distâncias. O vetor \mathbf{x}_t , de dimensão L produz uma saída de dimensão Q , sendo Q a quantidade de neurônios. Continuando o exemplo da Figura 5.1, os dois neurônios utilizados mapeiam o espaço dimensional mostrado naquela figura em um espaço também bidimensional, mostrado na Figura 5.2. Nessa figura, cada dimensão representa, portanto, a saída de um neurônio.

A Figura 5.2 também mostra que o espaço de saída da camada é linearmente separável, diferentemente do espaço de entrada. A reta de separação mostrada é calculada pela camada de saída. Essa camada recebe como entrada o vetor $\mathbf{y}_1(t)$. Sua saída é determinada pela soma ponderada dos elementos de $\mathbf{y}_1(t)$, acrescentada a um determinado limiar, ou *bias*. $\mathbf{y}_1(t)$ é o vetor formado pela saída de cada neurônio da camada escondida no instante t . O resultado da camada de saída e, conseqüentemente,

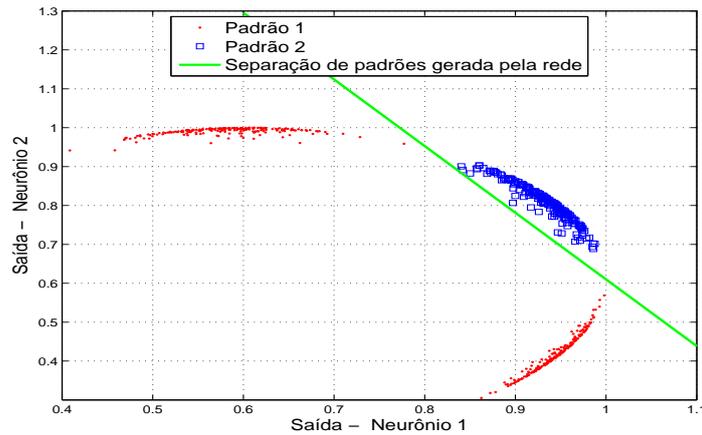


Figura 5.2. Mapeamento da camada escondida realizado por meio do treinamento dos dois neurônios mostrados na Figura 5.1. Nota-se que, após esse mapeamento, o espaço é linearmente separável (há um segmento de reta separando os padrões).

da rede, é dado, no instante t , por:

$$y_{\text{rbf}}(t) = \mathbf{w} \cdot \mathbf{y}_2(t), \quad (5.4)$$

em que

$$\begin{aligned} \mathbf{y}_2(t) &= [1 \ \mathbf{y}_1(t)^T]^T, \\ &= [1 \ y_{1,1}(t) \ y_{1,2}(t) \ \dots \ y_{1,Q}(t)]^T. \end{aligned} \quad (5.5)$$

O ponto na Equação 5.4 se refere ao produto interno entre \mathbf{w} e $\mathbf{y}_2(t)$. O vetor de pesos \mathbf{w} é estimado a fim de se minimizar a norma quadrática dos resíduos (Aguirre [2007]). A estimação de \mathbf{w} pode ser resolvida com o uso do método de mínimos quadrados clássico¹. Realiza-se isso minimizando o erro quadrático médio de $y_{\text{rbf}}(t)$ ao se tentar explicar $y(t)$ por meio de $\mathbf{w} \cdot \mathbf{y}_2(t)$ para todo t . É uma estimação de parâmetros utilizando-se o erro de previsão um passo a frente (Aguirre *et al.* [2010]), que é denominado resíduo por ser dado de treinamento. Utiliza-se a resolução de mínimos quadrados por meio da matriz pseudoinversa:

$$\mathbf{w} = (\Psi^T \Psi)^{-1} \Psi^T \mathbf{y}, \quad (5.6)$$

em que Ψ é a matriz de regressores formada pelos elementos de $\mathbf{y}_2(t)$ para todo t . Analogamente, \mathbf{y} é o vetor de $y(t)$ para todo t .

O treinamento da rede RBF, portanto, consiste na obtenção dos centros e raios

¹O uso do método de mínimos quadrados clássico supõe algumas condições como a de que o erro de estimação seja branco e aditivo, que as variáveis independentes sejam conhecidas e sem incerteza e que o modelo seja linear nos parâmetros.

da camada oculta, bem como do vetor \mathbf{w} da camada de saída. Pode-se, então, avaliar o desempenho da rede comparando-se os erros cometidos pelas previsões geradas $y_{\text{rbf}}(t)$ em relação à saída desejada $y(t)$.

5.2 Cálculo dos Erros

Uma das medidas mais comuns de desempenho de modelos é o erro quadrático médio (*Root Mean Square Error*, ou RMSE):

$$RMSE_{\text{modelo}} = \sqrt{\frac{1}{N} \sum_{t=1}^N (y(t) - \hat{y}(t)_{\text{modelo}})^2}, \quad (5.7)$$

em que N é a quantidade de observações utilizadas, $y(t)$ é a série a ser prevista, e o acento circunflexo em $y(t)$ denota a estimativa de $y(t)$ dada pelo modelo (como $y_{\text{rbf}}(t)$).

Entretanto, como mencionado em Armstrong & Collopy [1992], o RMSE não é adequado para se comparar resultados de previsão entre séries diferentes.

No caso de séries financeiras, é interessante se comparar o resultado da previsão (um passo a frente) com o do modelo do passeio aleatório (RW). A previsão do passeio aleatório na série de preços é igual ao preço no instante anterior. Conseqüentemente, a previsão desse modelo para a série de retornos é nula:

$$\begin{aligned} \hat{p}(t+1)_{\text{RW}} &= p(t), \\ \hat{r}(t+1)_{\text{RW}} &= \hat{p}(t+1)_{\text{RW}} - p(t) = 0. \end{aligned} \quad (5.8)$$

em que $\hat{p}(t+1)_{\text{RW}}$ indica a previsão do modelo do passeio para a série de preços e $\hat{r}(t+1)_{\text{RW}}$ é a mesma previsão, mas para a série de retornos $r(t)$. Ou seja, a previsão do RW sempre tem valor igual a zero.

A comparação entre os erros de previsão do modelo desenvolvido neste capítulo e o descrito na Equação 5.8 pode ser considerado um indicativo da previsibilidade da série. Se o modelo desenvolvido obtiver um desempenho inferior ao RW, pode-se concluir que não há nenhuma estrutura ou dependência temporal significativa a ser prevista. Supõe-se, com a afirmação anterior que o modelo tenha sido bem sucedido em “explicar a dinâmica” possivelmente presente nos dados.

Para se comparar a previsão de um modelo com o RW, pode-se utilizar diversas medidas de erro (Armstrong & Collopy [1992]). Uma delas é a razão conhecida como U_2 de Theil:

$$U_2 = \frac{RMSE_{\text{modelo}}}{RMSE_{\text{RW}}}. \quad (5.9)$$

Observe que a comparação entre os modelos é feita depois da totalização do erro por meio do RMSE. Isso é feito para que se evite a divisão entre os erros:

$$\text{Razão dos erros} = \frac{y(t) - \hat{y}(t)_{\text{modelo}}}{y(t) - \hat{y}(t)_{\text{RW}}}. \quad (5.10)$$

A razão dada pela Equação 5.10 tem um “comportamento explosivo” quando $y(t)$ tem valor próximo de zero. Isso acontece porque, como a previsão de RW é sempre zero, seu erro, nesse caso será muito pequeno. Consequentemente, essa razão terá um valor muito elevado, fazendo com que a média total dos erros relativos seja também muito elevada. E isso não necessariamente indica um modelo ruim ou uma série pouco previsível. Além disso, em casos de erro de RW nulo, a Equação 5.10 é indeterminada.

Outra medida bastante útil é a *percent better* (PB):

$$\text{PB} = 100 \frac{\sum_{t=1}^N j(t)}{N}$$

em que:

$$j(t) = \begin{cases} 1 & \text{se } |y(t) - \hat{y}(t)_{\text{modelo}}| < |y(t) - \hat{y}(t)_{\text{RW}}|, \\ 0 & \text{caso contrário.} \end{cases}$$

Uma das vantagens da medida PB está em sua fácil interpretação: é a porcentagem de vezes em que o erro do modelo em estudo foi menor que o erro do RW. Outra vantagem é a sua robustez em relação a *outliers* e a valores pequenos de erro do RW (como mencionado para medida U_2). A robustez a *outliers* é obtida pelo fato de não se considerar a magnitude do erro, mas simplesmente sua comparação qualitativa com o erro de RW. O comportamento de PB não é interessante em casos de estimação de parâmetros, uma vez que o erro do modelo seja menor que o de RW, PB se torna insensível a possíveis melhoras do modelo. Em contrapartida, em séries de baixa previsibilidade, como é o caso, PB é uma boa base de comparação do desempenho de um modelo em diferentes séries.

Além dessas medidas de erro, outro indicativo de bom desempenho em previsões de séries financeiras é a porcentagem de vezes em que a previsão acerta o sinal do retorno. Retornos positivos ou negativos indicam, respectivamente, a direção de alta ou baixa dos preços. Por esse motivo, essa última medida de desempenho é chamada de direção dos retornos (DIR).

5.3 Dimensionamento e Estrutura da Rede

Como o objetivo do trabalho não é avaliar a previsão das redes e sim a previsibilidade dos dados, as mesmas são construídas de maneira bastante semelhante. Um modelo é estimado para cada série e, idealmente, deseja-se que os modelos sejam equivalentes entre si. Optou-se, então, por simplificar a construção dos modelos fazendo com que os mesmos não possuam parâmetros livres, que possam fazer com que um modelo RBF seja melhor que outro. A finalidade, novamente, não é comparar os modelos e sim os dados.

Excetuando-se a escolha do número de atrasos L e da quantidade de neurônios Q , o treinamento da rede é descrito na Seção 5.1. A obtenção dos centros dos neurônios é baseada no algoritmo k -médias. Esse algoritmo não possui parâmetros a serem escolhidos e, portanto, seu desempenho deve depender, basicamente, dos dados. O mesmo pode ser dito a respeito da escolha do campo receptivo dos neurônios. O procedimento de escolha é o mesmo para todos os modelos: a menor distância entre os neurônios. O treinamento da camada de saída é feito pelo método de mínimos quadrados clássico e é também isento de parametrizações. Percebe-se, portanto, que, em todo o procedimento de treinamento descrito na Seção 5.1, a única fonte de variações para o desempenho da rede está nos dados utilizados.

Um aspecto de grande influência no desempenho da rede é a escolha da quantidade de atrasos L . O valor de L está relacionado à complexidade do processo que gerou a série. Não se conhece, a princípio, esse número, mas, atualmente, são usadas diferentes estratégias para sua escolha. Alguns autores utilizam técnicas de identificação de dimensões mínimas de imersão como a de falsos vizinhos. Algumas referências citadas na Seção 1.2 utilizam a estimativa de dimensão de correlação como L . Embora intuitivamente coerente, esse uso não possui base teórica. Tanto a técnica de falsos vizinhos quanto a estimação da dimensão de correlação são aplicáveis apenas em dados estacionários. Além disso, pôde-se perceber, ao longo da Seção 1.2 e do capítulo 4 que a estimação de dimensão em si já enfrenta vários problemas em séries desconhecidas e de tamanho finito. A escolha de L deve ser grande o suficiente para que o espaço de entrada possa representar corretamente a dinâmica do processo. Na prática, entretanto, à medida que L cresce, o treinamento da rede se torna mais complexo (computacionalmente) e há o risco de se incorrer em *overfitting*².

Outro parâmetro de grande influência é número de neurônios Q . Semelhantemente a L , Q está relacionado à complexidade. Esse último, todavia, não indica a

²Sobredimensionamento da rede. As consequências disso são um aparente melhor desempenho em dados de treinamento, mas uma menor capacidade de generalização do modelo.

complexidade dos dados e sim da própria rede, de sua capacidade de armazenamento de informações. Há, também, diversas maneiras de se escolher o número de neurônios. Para redes neurais (como RBF ou MLP) existem várias técnicas como a de poda (*pruning*) e a construtiva, baseada na adição de neurônios (Costa *et al.* [2003]). Grande parte das técnicas são baseadas em otimizações multiobjetivo (de Albuquerque Teixeira *et al.* [2000]), como, por exemplo, a regressão LASSO (Kokshenev & Braga [2008]) que utiliza a minimização da norma-1 (ao invés da norma euclideana, como no caso de mínimos quadrados).

É clara a influência de L e Q no treinamento da Seção 5.1. Um exemplo é da escolha do raio em função de L : quanto maior o número de neurônios, maior é a probabilidade se ter uma menor distância. A escolha de um L elevado, também sugere um elevado número de neurônios, já que uma quantidade de informação maior deve ser representada pela rede.

5.3.1 A escolha de L e Q

O dimensionamento da rede é baseado em uma validação cruzada. Inicia-se esse procedimento dividindo-se o conjunto de dados em três partes. A primeira parte (70% das observações) é destinada ao treinamento da rede. A segunda parte (15% das observações) é destinada à validação e a última parte dos dados (15% das observações) é reservada para o teste final da rede.

O conjunto de treinamento é utilizado para treinar a rede com determinados valores de L e Q . O desempenho dessa rede é medido por meio do erro de previsão (RMSE) cometido na segunda parte dos dados: os dados de validação. A rede é, consequentemente, treinada e validada em conjuntos diferentes de dados e seu desempenho é dado em função de L e Q . Os valores de L são inteiros tomados entre 1 e um valor máximo L_{\max} . Os valores de Q também são inteiros e tomados entre 1 e Q_{\max} . Esse procedimento é realizado para esses valores desses dois parâmetros e obtém-se uma matriz $L_{\max} \times Q_{\max}$ de desempenho das redes candidatas ($RMSE_{L,Q}$). É escolhida, a rede cujo conjunto L e Q apresenta melhor desempenho, ou, equivalentemente, menor $RMSE_{L,Q}$.

Como a melhor estrutura L e Q (de menor $RMSE_{L,Q}$) é escolhida tomando como base o desempenho sobre os dados de validação, esse conjunto de dados, torna-se, filosoficamente, de “treinamento”, ou seja, de escolha da melhor rede. A escolha, portanto, precisa ser testada em um conjunto diferente de dados: o conjunto de teste.

O desempenho final de previsão é obtido pelo erro cometido (U_2 , PB e DIR) por essa última rede nos dados de teste. Dessa maneira, consegue-se uma escolha

de L e Q com o mínimo possível de interferência de parâmetros livres (escolhidos heurísticamente).

!!!!

5.4 Previsão das Séries Financeiras

Os procedimentos descritos nas seções anteriores foram aplicados às sete séries estudadas. Primeiramente, são mostrados os resultados da escolha da estrutura (parâmetros L e Q). Em seguida, para a rede selecionada de cada série, é mostrado o desempenho das mesmas nos dados de teste.

5.4.1 Estimação das Estruturas das Redes

Para cada série, redes RBF, como as descritas na Seção 5.1 são treinadas com vários valores de atrasos L e neurônios Q . As 70% primeiras observações são utilizadas para se treinar as redes com valores de L variando num intervalo entre 1 e 10. Para cada L , são geradas redes com quantidade de neurônios tomados de 1 a 30. Após o treinamento, o RMSE de cada uma das 2100 redes (7 séries, 10 atrasos e 30 neurônios) é calculado para previsões *um passo à frente nos dados de validação*. Após esse procedimento, são geradas as Figuras 5.3 e 5.4. Essas mostram o desempenho das redes nos dados de validação em função de L e Q para cada série e são utilizadas como critério de seleção dos modelos.

Em cada um dos gráficos das Figuras 5.3 e 5.4, seleciona-se o par L e Q de menor erro em cada uma das séries. Nos gráficos 5.3.a, 5.3.b, 5.3.c e 5.3.d é possível perceber que, na maioria dos casos, o aumento de Q leva a um aumento do erro. De fato, os melhores desempenhos obtidos para DJIA, SP500, IBOV e PETO correspondem a valores pequenos de Q . Em relação a L , não se vê nenhuma tendência muito clara nesses gráficos. O gráfico de PETH (Figura 5.4.a) parece exibir uma melhora no desempenho com o aumento de Q . O gráfico 5.4.c se distingue dos demais por apresentar um vale na região entre $L = 4$ e $L = 6$. Embora essa região seja bem definida, o menor valor de erro para esse gráfico é obtido para $L = 1$.

Na maioria dos gráficos, percebe-se que a superfície dos erros é um tanto quanto irregular, indicando talvez, uma pouca previsibilidade causada pela aleatoriedade dos dados. Isso é um dos argumentos que pode justificar um desempenho fraco de previsão em todas as séries (quando se comparado ao passeio aleatório). Se as séries pudessem ser modeladas por um ruído branco (série de retornos de um passeio aleatório gaussiano), poderia-se esperar uma superfície $L \times Q$ aleatória, distribuída uniformemente. Isso é

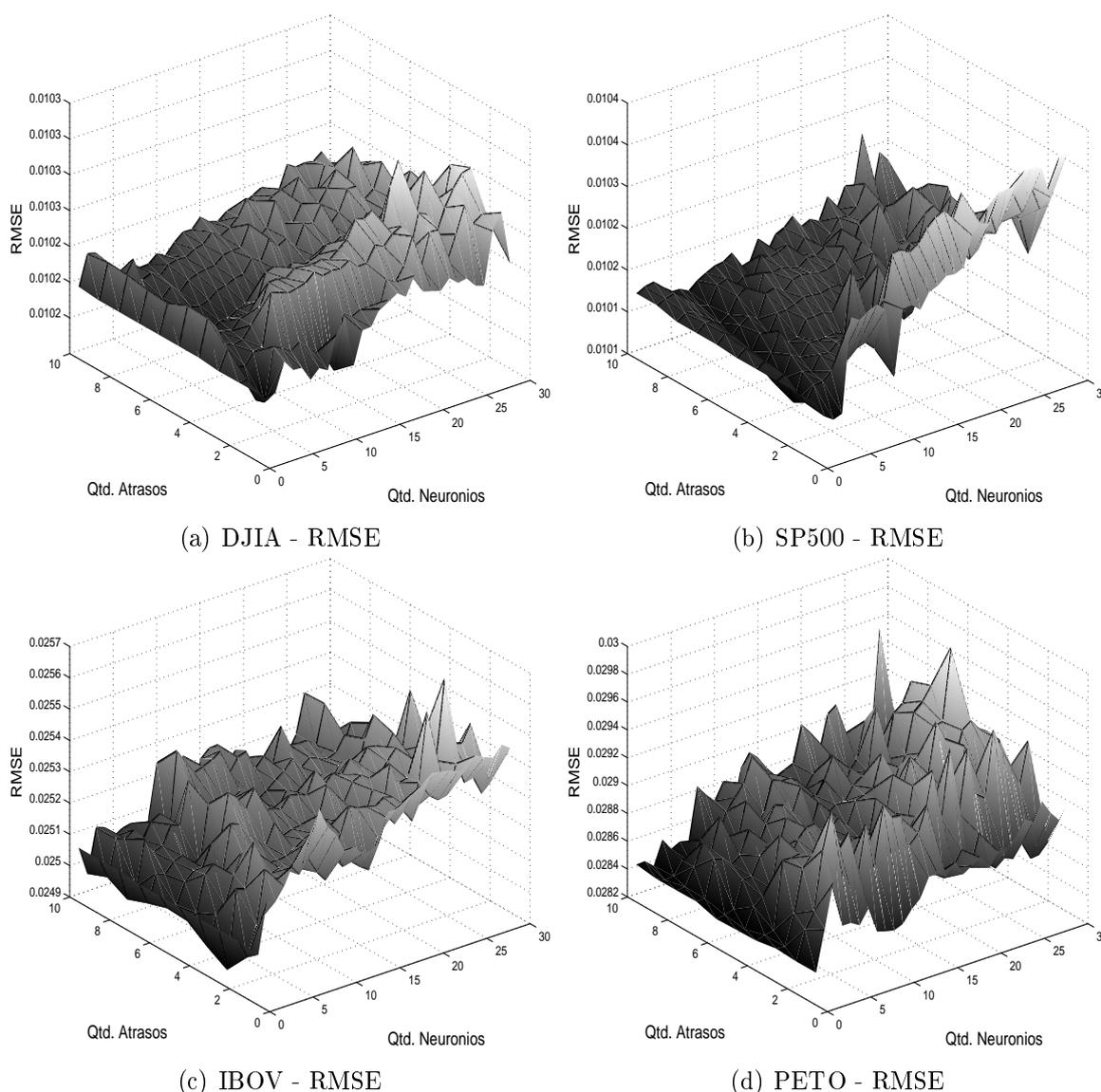


Figura 5.3. Validação cruzada de DJIA, SP500, IBOV e PETO. Os gráficos mostram o erro RMSE nos dados de validação em função da quantidade de atrasos L e de neurônios Q , obtido ao se prever cada uma das séries. Os pontos mais escuros mostram valores menores de erro. Para essas séries percebe-se um menor erro na região de menor Q .

esperado porque os valores presentes dessa série são independentes de valores passados. Portanto, o desempenho de previsão seria semelhante para todos os L , não importando também o valor de Q escolhido. Poderia-se, portanto, inferir que a ausência de regiões bem definidas (de menor erro RMSE), somada ao baixo desempenho de previsão (U_2 , PB e DIR) do modelo escolhido nos dados de teste indicaria baixa previsibilidade dos dados.

A Tabela 5.1 mostra os pares L e Q , escolhidos para gerar a rede que será usada

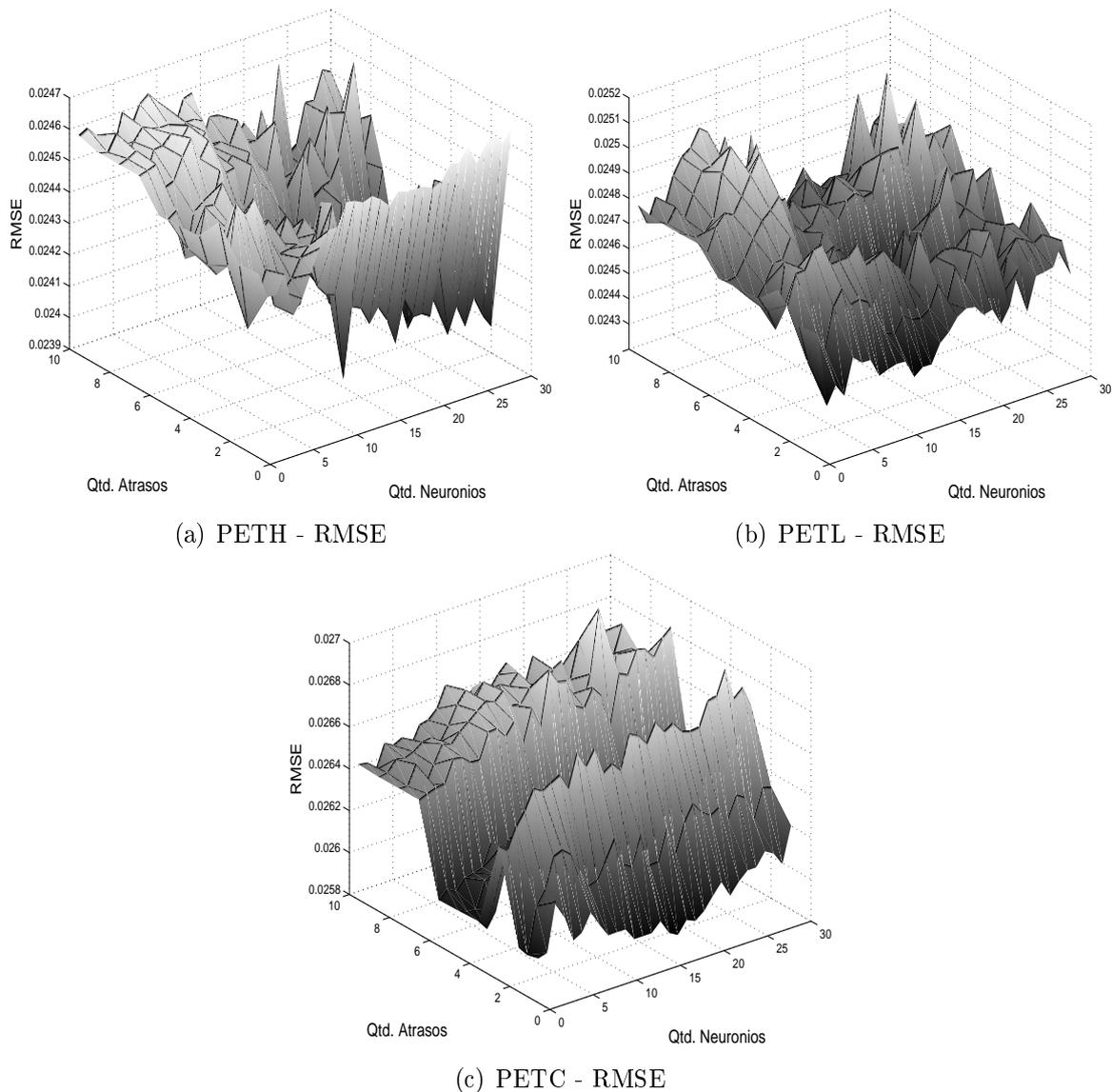


Figura 5.4. Validação cruzada de PETH, PETL e PETC. Os gráficos mostram o erro RMSE nos dados de validação em função da quantidade de atrasos L e de neurônios Q da estrutura da rede, obtido ao se prever cada uma das séries. Os pontos mais escuros mostram valores menores de erro.

nos dados de teste. Percebe-se que, na maioria das séries, o número de atrasos de melhor desempenho é pequeno, não sendo maior que 3 atrasos (exceto para PETO e PETH). Deve-se chamar a atenção para o fato de o erro RMSE não ser adequado para se comparar o desempenho *entre as séries*. Essa comparação será feita, por meio de U_2 , PB e DIR, com os dados de teste, mais adequados para esse fim.

A Tabela 5.1 resume a escolha das estruturas que serão usadas para se comparar a previsibilidade (obtida no Capítulo 4) e o desempenho de previsão. A obtenção dessas estruturas é importante para se obter modelos que evitem o *overfitting* de dados de

Tabela 5.1. Uma rede RBF é construída para cada série. A tabela mostra a quantidade de atrasos (L), o número de neurônios (Q) e o erro RMSE de cada rede nos dados de validação.

Série	L	Q	Erro RMSE
DJIA	2	4	0,0102
SP500	3	3	0,0101
IBOV	3	2	0,0249
PETO	5	2	0,0283
PETH	6	28	0,0239
PETL	1	12	0,0244
PETC	1	17	0,0259

treinamento.

Os dados de treinamento, validação e teste são, em ordem cronológica, a primeira, segunda e terceira parte de cada uma das séries, respectivamente. Como o procedimento de modelagem inclui tanto a estimação dos parâmetros da rede quanto a escolha da estrutura, pode-se entender os dados de validação também como um tipo de dado de treinamento, uma vez que foi utilizado na modelagem, na escolha da estrutura das melhores redes. Portanto, apenas a terceira parte dos dados não é usada na modelagem e essa será a escolhida para se avaliar a previsibilidade das séries em termos de erros de previsão. A fim de se reduzir a influência da não estacionariedade, as redes são treinadas utilizando os dados de validação também para a escolha de seus parâmetros, considerando-se que essa último está mais próximo cronologicamente dos dados de teste. A denominação dos conjuntos de dados nos restante do trabalho será dividida em dados de treinamento e teste considerando, que fazem parte do treinamento ambos os dados de treinamento e validação.

5.4.2 Desempenho das Redes

As redes obtidas por meio dos procedimentos apresentados nas Seções 5.1 e 5.3 são aplicadas ao conjunto de dados de teste (os últimos 15% dos dados) das séries. Os resultados podem ser vistos na Tabela 5.2. Considera-se que as redes têm capacidades bastante semelhantes de modelar a dinâmica presente nos dados. Portanto, o resultado da tabela deve indicar, além do desempenho do modelo, a previsibilidade das séries. Tanto RMSE quanto U_2 são medidas de erro e são, consequentemente, inversamente proporcionais ao desempenho de previsão. O *percent better* (PB) e o indicativo de porcentagem de acerto na direção dos retornos são diretamente proporcionais ao desempenho.

Tabela 5.2. Desempenho da previsão um passo a frente das redes RBF em séries financeiras.

Série	RMSE	U_2	PB (%)	DIR (%)
DJIA	0,0118	1,00	47,8	49,3
SP500	0,014	1,00	45,6	47,7
IBOV	0,0198	0,998	50,1	55,7
PETO	0,0324	0,999	52,6	53,1
PETH	0,0263	1,01	50,5	56,1
PETL	0,0317	0,997	53,2	55,3
PETC	0,0377	1,20	45,1	49,0

A segunda coluna da Tabela 5.2 mostra o erro RSME obtido. Por essa coluna, poder-se-ia pensar que o melhor desempenho é de DJIA. Essa série, entretanto, é a que tem um dos piores desempenhos em termos de PB, por exemplo.

PB mede a quantidade de vezes em que o erro de previsão é menor que o do modelo do passeio aleatório (RW). Nesse caso, DJIA tem desempenho pior do que o RW. SP500 e PETC também apresentam desempenho inferior ao RW. Apenas PETO e PETL superam esse modelo quando comparadas por meio do PB.

U_2 é a razão entre os erros de previsão (RMSE) do modelos RBF e RW. Como o erro de previsão RW é sempre o próprio valor da série, pode-se entender U_2 como um valor de erro RMSE normalizado pela variância dos dados. Valores maiores que a unidade indicam erros de previsão, em média, maiores que os cometidos pelo RW. Por outro lado, valores entre zero e um mostram previsões do modelo, em média, melhores que o RW. Todas as séries apresentam U_2 bastante próximo de 1, o que sugere baixa previsibilidade em todos os dados.

A última coluna da Tabela 5.2 mostra a porcentagem de vezes em que a previsão do modelo RBF acertou o sinal da observação futura (DIR). Essa medida é bastante utilizada em artigos cujo foco é a previsão de séries financeiras. Um baixo valor de erro não necessariamente implica bons resultados financeiros. Em contrapartida, saber se os retornos futuros serão positivos ou negativos tem consequências mais diretas e intuitivas em estratégias de investimentos. Um valor de direção de acerto igual a 50% é o esperado de uma variável aleatória cujo resultado, se positivo ou negativo, seja igualmente provável, como o lançamento de uma moeda³. De maneira geral, o resultado de todas as séries foi melhor no quesito direção do que no PB. Embora o desempenho dos modelos de várias séries em relação a PB seja considerado pior que o passeio aleatório, apenas DJIA, SP500 e PETC tiveram números inferiores a 50%

³O lançamento de uma moeda pode ser visto como uma variável aleatória de Bernoulli com $p = 0,5$, cujas faces são mapeadas nos valores 1 e -1 .

quando se avalia o desempenho da direção. Nesse último quesito, IBOV, PETH e PETL tiveram desempenho acima de 55%. Ainda assim, não se sabe se esses números de previsão são significativos economicamente, dado que mesmo os melhores resultados ainda ficam próximos de 50%.

5.5 A Evolução de W_2 ao longo das Fases de Previsão

O teste BDS também pode ser utilizado para se avaliar a adequação de modelos a dados. Nesse sentido a análise dos erros de previsão um passo a frente ($(y(t+1) - \hat{y}(t+1))$), por meio da estatística W_2 desse teste pode ser útil para mostrar se os modelos RBF explicam toda a “dinâmica” existente nos dados. Ou seja, se os erros dos modelos estimados forem independentes, não há mais o que ser modelado e o modelo pode ser considerado adequado para representar dinamicamente a evolução da série. Para tanto, pode-se calcular W_2 dos erros de previsão e, se o valor em módulo for inferior a 1,96, considera-se que esses erros são independentes, como um ruído branco. Caso contrário, há, possivelmente nos dados, estruturas ainda não modeladas.

Como visto na Seção 4.5.1, a dependência temporal da variância influencia bastante o resultado de W_2 . É possível que a rejeição da hipótese de independência dos erros de previsão se dê devido à presença dessa característica. Nesse caso, é desejável que se filtre esse efeito da variância, para uma correta interpretação de W_2 . Um filtro GARCH é, então, estimado a partir dos erros de previsão um passo a frente do modelo RBF e a estatística W_2 é calculada em cada passo de modelagem e filtragem. Isso é feito dividindo-se cada observação do erro pelo desvio padrão, estimado via GARCH, em cada instante. A fim de mostrar essa evolução de W_2 , a Tabela 5.3 apresenta essa estatística calculada para vários conjuntos de dados. Primeiramente, é mostrada a estatística para a série original ($|W_2|$ série). Em seguida, é mostrada a estatística para os erros de previsão em dados de treinamento ($|W_2|$ TR). A próxima coluna apresenta $|W_2|$ para os erros de previsão em dados de treinamento filtrados pelo modelo GARCH ($|W_2|$ TR homosc.). Em seguida são mostradas essas estatísticas para os dados de teste (TE) com e sem a filtragem GARCH respectivamente.

O valor elevado de W_2 na segunda coluna (w_2 série) da Tabela 5.3 mostra que as séries não são independentes, o que justifica uma modelagem dos dados. Após a modelagem, é gerada a terceira coluna (w_2 RV), contendo a mesma estatística, mas para os erros de previsão do modelo. Essa coluna foi calculada utilizando-se os dados de treinamento. Espera-se, a princípio, que esses erros sejam independentes. Percebe-

Tabela 5.3. Estatística W_2 calculada para as séries e para os erros de previsão das mesmas em dados de treinamento (TR) e teste (TE) antes e depois do filtro GARCH. Após a filtragem, os dados são denominados homoscedásticos (homosc.)

Série	$ W_2 $ série	$ W_2 $ TR	$ W_2 $ TR homosc.	$ W_2 $ TE	$ W_2 $ TE homosc.
DJIA	42,8	5,70	3,17	13,5	3,02
SP500	26,7	7,59	1,11	11,8	3,27
IBOV	36,0	9,54	1,70	7,74	2,43
PETO	22,3	5,12	0,0424	9,93	0,109
PETH	22,8	5,20	0,162	5,80	2,74
PETL	25,4	3,89	0,416	9,27	1,23
PETC	21,26	2,66	0,800	9,63	0,0843

se que grande parte da dependência temporal foi modelada pela rede RBF, dado que W_2 tem seu valor bastante reduzido, quando comparado à segunda coluna. Essa, entretanto, não apresenta valores inferiores a 1,96 para nenhuma das séries. Isso pode indicar uma modelagem inadequada ou a presença de heteroscedasticidade. Com o objetivo de se eliminar a hipótese de uma modelagem insuficiente, esses erros são tornados homoscedásticos pela utilização de um filtro GARCH.

Após a filtragem GARCH, é gerada a quarta coluna da tabela (w_2 RV homosc.), com valores ainda menores do que os da terceira coluna. Como W_2 gera uma distribuição simétrica (*two tailed test*) em torno de zero, deve-se rejeitar a hipótese de independência com valores menores que $-1,96$ ou maiores que $1,96$. A tabela 5.3 apresenta resultados em módulo para facilitar a comparação entre os mesmos. Apenas DJIA (comparando-se $|W_2|$ RV homosc.) tem sua hipótese de independência rejeitada com nível de confiança de 95%. Ainda assim, o valor de sua estatística é, em módulo, menor do que antes da filtragem. Esse resultado pode indicar que a modelagem de DJIA pela rede RBF não foi suficiente para capturar o determinismo existente (se ainda existente) nessa série. Não se realiza aqui uma nova modelagem a fim de manter o processo de modelagem comum a todas as séries, permitindo que ainda se possa compará-las em termos da capacidade de previsão. Portanto, o que se interpreta dessa rejeição é que DJIA possui um determinismo mais difícil de ser modelado. Pode haver, por exemplo, uma dependência de longo prazo ($L > 10$), desprezada no procedimento da Seção 5.3. Esse evento não invalida o procedimento apresentado. Essa dificuldade na modelagem pode, de certa maneira, ser entendida como uma menor previsibilidade.

A quinta e a sexta coluna da Tabela 5.3 mostram W_2 nos erros obtidos pela previsão dos dados de teste. Essas colunas mostram essa estatística antes (W_2 RT) e depois (W_2 RT homosc.) da filtragem GARCH desses erros. Os números apresentados são maiores que os da terceira e quarta colunas. Isso não é atípico, já que esses

dados não participaram do treinamento da rede e podem conter alguma “dinâmica” diferente da exibida nos primeiros conjuntos de dados. Sendo essa hipótese verdadeira, essa não estacionariedade pode impedir também que haja um bom desempenho de previsão. Percebe-se ainda que a última coluna apresenta números bem menores que a anterior. Conseqüentemente, parte da dependência ainda existente nos erros é devida à heteroscedasticidade. Quando essa característica é filtrada (sexta coluna), as séries PETO, PETL e PETC não são mais rejeitadas como independentes. Para essas séries, pode-se concluir que, estando válidos os procedimentos seguidos, não há mais dependência temporal a ser modelada. Os resultados de previsão dessas séries, seriam, portanto, os melhores possíveis. No caso de PETC, como sua previsão é inferior à previsão do passeio aleatório, poder-se-ia concluir que na série de retornos de fechamento da PETROBRAS, não há nada a ser previsto. Essa conclusão está, também, de acordo com os resultados da análise de dados sub-rogados PPS da Seção 4.5.3. É bastante interessante notar essa característica exatamente nos retornos de fechamento, uma vez que esses são, predominantemente, os mais estudados. Esses tipos de retorno são, via de regra, os mais modelados quando o objetivo de estudo é a previsão.

Além da possível não estacionariedade, os resultados dos dados de teste maiores que os de validação podem também indicar *overfitting*. O objetivo da validação cruzada, entretanto, é exatamente eliminar esse problema. Outra razão para se desconsiderar o *overfitting*, pelo menos em DJIA, SP500, IBOV e PETO é a baixa quantidade de neurônios utilizados no treinamento.

Em geral, pôde-se perceber uma baixa previsibilidade de todas as séries. As séries de melhor resultado de previsão coincidiram com as que obtiveram os melhores índices de previsibilidade PREV, o que condiz com os resultados obtidos no Capítulo 4. Pode-se dizer, que os estudos de previsibilidade estão, na maioria dos casos, de acordo com o desempenho de previsão das séries.

Resultados e Conclusões

O presente trabalho teve por objetivo apresentar algumas técnicas que podem ser utilizadas no estudo da previsibilidade das séries financeiras. Esse tema foi introduzido no Capítulo 1, que mostrou também uma breve revisão bibliográfica do assunto.

O Capítulo 2 apresentou algumas técnicas tradicionais para análise de séries financeiras. Pôde-se perceber algumas limitações dessas técnicas como a suposição de que os dados tenham uma distribuição gaussiana. O teste Ljung-Box não é capaz de detectar dependências não lineares e não pode ser utilizado como teste suficiente para a adequação de séries ao passeio aleatório. O teste de Breusch-Pagan não conseguiu identificar a heteroscedasticidade das séries ARCH e GARCH. O teste de White, da maneira como foi usado, não conseguiu distinguir dependências não lineares previsíveis da presença de heteroscedasticidade. Isso pôde ser visto na rejeição da hipótese de homoscedasticidade do mapa logístico, que nem sequer é estocástico. O teste de Kolmogorov-Smirnov rejeitou a hipótese de normalidade de todas as séries, o que mostra também, a rejeição do passeio aleatório gaussiano.

Diante das limitações dos testes do Capítulo 2, foram apresentadas várias alternativas com o objetivo de quantificar o determinismo dos dados. Algumas são baseadas na integral de correlação. Uma delas é a dimensão de correlação, que quando calculada por meio do Algoritmo de Judd, necessita de uma amostra menor do que o tradicional algoritmo de Grassberger e Proccacia, o que é bastante desejável em séries financeiras. Além disso, mostrou-se que esse algoritmo pode ser usado para se distinguir séries temporais com estruturas dinâmicas diferentes (e não lineares). Esse também foi o caso da entropia amostral, uma medida de regularidade dos dados. A informação mútua média, baseada na Teoria da Informação, derivada por meio da entropia de Shannon, é a medida da quantidade de informação que uma variável aleatória tem de outra. No contexto de séries temporais, a AMI foi utilizada como medida da quantidade de informação (previsibilidade) que valores passados na série possuem a respeito de valores futuros.

Todas essas técnicas foram utilizadas em conjunto com a análise de dados sub-

rogados. Essa análise tem sido uma alternativa promissora aos tradicionais testes de hipótese, por não necessitar de supor que os dados sigam certas distribuições estatísticas. Isso vai ao encontro da necessidade de se utilizar testes em séries financeiras que não suponham a normalidade. Foram apresentados algoritmos de geração de dados sub-rogados para hipóteses nulas de independência, linearidade gaussiana, linearidade (com transformação estática não linear) e dados periódicos. Esse último algoritmo, denominado PPS, não foi utilizado como teste para séries pseudoperiódicas, mas sim para a detecção de determinismo que seja robusta à presença de heteroscedasticidade. O PPS gera dados sub-rogados cuja hipótese nula inclui modelos da família GARCH e indicam, portanto, a presença de outras dependências nos dados (Small & Tse [2003b]). Chama-se a atenção aqui para a escolha do parâmetro de dispersão do algoritmo PPS. Uma escolha errada de tal parâmetro pode fazer com que o algoritmo sempre rejeite ou nunca rejeite a hipótese nula, independente dos dados utilizados. Valores próximos ao recomendado pelo autor não gerarão erros tão extremos. Todavia, falta um instrumento de validação para a escolha desse parâmetro, dado que a rejeição de hipóteses é bastante sensível ao mesmo.

Outro teste que foi apresentado e que obteve resultados mais confiáveis foi o BDS. Baseado também no cálculo da integral de correlação, o BDS é um teste de independência que compara uma estatística (W_2) calculada para os dados e a compara com um valor limite para dados independentes. Esse pôde ser aplicado em séries artificiais para mostrar algumas de suas características como a detecção de não linearidades e a robustez à não normalidade dos dados. Dados independentes e identicamente distribuídos sempre geram uma mesma distribuição de W_2 , o que facilita em muito os testes de dependência temporal.

Como foi visto, as séries financeiras estudadas aqui são não gaussianas (rejeição do teste KS), não lineares (dependência do quadrado dos retornos) e não estacionárias (heteroscedasticidade). Nesse contexto, a utilização das estatísticas discriminantes por meio da análise de dados sub-rogados e a análise de resíduos BDS de modelos AR, GARCH e AR-GARCH são bastante relevantes por serem robustas à distribuição, detectarem não linearidades e contornarem a questão da heteroscedasticidade (gerando dados sub-rogados com essa característica ou filtrando a mesma).

6.1 Previsibilidade

A análise de previsibilidade se dividiu em algumas fases nas quais tentou-se identificar os tipos de dependência presentes nos dados. Em cada teste, pôde-se perceber algumas

características peculiares dessas séries. As ferramentas utilizadas foram os algoritmos 0, 1, 2 e PPS de geração de dados sub-rogados e o teste BDS na análise de resíduos de alguns modelos. Para esse último teste, foram geradas também, permutações da série, de maneira semelhante ao algoritmo 0, a fim de que seus resultados fossem mais robustos. Dessa abordagem, gerou-se um índice de previsibilidade que quantifica a rejeição das hipóteses e, em certos testes, a previsibilidade dos dados.

6.1.1 Testes de Independência - A Inadequação do Passeio Aleatório

A primeira fase da análise é constituída de testes de independência. Fez-se uso do teste BDS das próprias séries e do algoritmo 0. Para todas as séries, em ambos os testes a hipótese de que os dados são i.i.d. foi rejeitada. Como esses testes são sensíveis a qualquer tipo de dependência, não fornecem informações a respeito da estrutura existente. Entretanto, os resultados são suficientes para se rejeitar a hipótese de que essas séries financeiras sejam como passeios aleatórios, o que contraria a HME. Ou seja, há valores passados dos retornos que estão relacionados a valores presentes ou futuros.

6.1.2 Testes de Linearidade - A Detecção de Estruturas Não Lineares

A segunda fase de testes analisa a presença de não linearidades. Foram usados o teste BDS em resíduos de um modelo AR e o algoritmo 1. Esse foi um primeiro passo no reconhecimento do tipo de dependências que poderiam existir. A hipótese testada foi a de que as dependências eram apenas lineares. Como colocado na Seção 4.4.1, os resultados dos testes de linearidade são bastante importantes, pois podem definir caminhos completamente diferentes para uma possível modelagem dos dados.

Modelos AR foram estimados com estrutura suficiente para que seus resíduos fossem considerados brancos, em relação ao teste Ljung-Box. Se esses resíduos não são independentes, existem não linearidades nos dados. O teste BDS rejeitou a hipótese de que esses resíduos são independentes para todas as séries. Analisando o valor de W_2 , pôde-se perceber, que, embora menores, são bastante próximos dos encontrados nos testes de independência das séries. Isso sugere que há uma pequena dependência linear (como no resultado da Seção 2.3), mas a “dinâmica principal” existente nos dados não foi capturada pelo modelo linear. O algoritmo 1 também rejeita a hipótese de linearidade de todas as séries. A interpretação do resultado desse algoritmo é sutilmente diferente do resultado do BDS. Isso porque as hipóteses não são exatamente iguais.

Esse comentário é válido para todos os resultados do teste BDS, que é baseado em modelos específicos dos dados. O BDS rejeita a hipótese de que os dados não podem ser modelados pelos respectivos modelos AR estimados (outros modelos AR poderiam fazer parte das hipóteses). A hipótese do algoritmo 1 é um pouco mais geral porque considera que os dados não são gerados por qualquer processo linear com incrementos gaussianos. Essa discussão, entretanto desvia-se um pouco do foco principal que é o de simplesmente identificar se há outras dependências nos dados além das lineares. A qualidade de uma modelagem depende da correta identificação da estrutura existente no processo em análise e o que se faz neste trabalho é exatamente essa identificação de estruturas.

Portanto, há em todas as séries uma pequena dependência linear, mas há também outras dependências, o que justificou a aplicação de outros testes de não linearidade.

6.1.3 Testes de Não Linearidade - A Detecção de Estruturas Não Lineares

Os testes de não linearidade têm por objetivo detectar estruturas não lineares que podem, eventualmente, estar presentes. Procurou-se testar dois tipos diferentes de não linearidade: a não linearidade estática e a dependência heteroscedástica. Ambos possuem a característica de não contribuírem para a previsibilidade, embora indiquem conteúdo determinístico (como definido na Seção 3.1).

O algoritmo 2 rejeitou a hipótese de transformação não linear estática de todas as séries. Esse teste pode ser visto como uma generalização do algoritmo 1 para dados não gaussianos. Portanto, a não normalidade é vista, pelo algoritmo 2 como um tipo de não linearidade. Essa, porém é estática, ou seja, não dinâmica. Como, pelos resultados obtidos, houve a rejeição dessa hipótese de não linearidade estática, conclui-se que os dados possuem “dinâmica” não linear.

O algoritmo PPS gera sub-rogados que possuem, em tese, a mesma dependência heteroscedástica que as séries originais (Small & Tse [2003b]). O uso desse algoritmo testou se havia nos dados alguma estrutura não linear que não fosse devida a essa dependência. O teste pode ser visto como um teste de não linearidade robusto à não estacionariedade da variância (heteroscedasticidade). Em DJIA, IBOV, PETH e PETL pôde-se rejeitar a hipótese de dependência heteroscedástica. Em SP500, PETH e PETC não se pôde rejeitar a hipótese. Para essas três últimas séries a não rejeição pode se dar pela limitação das estatísticas discriminantes em separar a série de seus sub-rogados. Portanto, os resultados mais conclusivos são os de rejeição. Deve-se lembrar que o algoritmo PPS possui um parâmetro livre que pode ter influenciado nos resultados.

O estudo de determinismo via algoritmo PPS necessita de outras fontes de validação de seus resultados. No presente trabalho, utilizou-se para esse fim, o teste BDS em resíduos AR-GARCH. Após a modelagem das partes linear e heteroscedástica, pôde-se estudar se seus resíduos eram independentes. Uma dependência nesses resíduos aponta para a presença de estruturas não lineares que possibilitam uma melhor previsão dos retornos futuros. Os resultados do BDS aplicados a esses resíduos mostraram resultados interessantes. Os índices DJIA, SP500 e IBOV não foram rejeitados, mas as quatro séries da PETROBRAS foram. Consequentemente, poderia-se supor que as únicas modelagens não lineares úteis seriam as de PETH, PETL e PETC e que um simples modelo linear seria o suficiente para mapear as dependências dos índices DJIA, SP500 e IBOV. Uma possível explicação para isso é o fato de o índice representar uma média de várias ações. Isso poderia fazer com que a dependência temporal presente em cada ação se perca no processo de cálculo da média. Supondo que os preços de ações diferentes não sejam correlacionados entre si, a soma das colaborações individuais geraria um índice temporalmente independente. Mesmo que as cotações não sejam completamente independentes, a média das mesmas pode reduzir a correlação temporal do índice. Além disso, os índices são mais sujeitos às diversas influências do mercado financeiro que as ações, quando consideradas individualmente. Essas fazem parte de um mercado específico em determinado setor da economia e a influência que sofrem é, de certo modo, mais restrita. Isso é mais uma conjectura que uma conclusão do trabalho, mas, certamente, sugere um estudo mais aprofundado dessas relações de previsibilidade entre ações e índices.

6.1.4 Teste de Resíduos GARCH - A quantificação da previsibilidade

Uma maneira de se tentar quantificar a previsibilidade é retirar as dependências heteroscedásticas dos dados e aplicar o teste BDS em seus resíduos. Para esse fim, foram estimados filtros GARCH para as séries, que geraram séries sem essas dependências. Os resultados mostrados na Seção 4.6 apontam, em geral, para uma baixa previsibilidade de todas as séries. Mas, pode-se perceber, que em alguns casos os valores de previsibilidade obtidos são bem mais baixos que outros. A tabela da Seção 4.6 é reproduzida novamente para evidenciar os resultados de previsibilidade obtidos. Esses são mostrados na Tabela 6.1.

Embora sejam diferentes de um passeio aleatório, séries como DJIA e SP500 não são mais previsíveis que esse modelo. Em algumas séries como PETH e IBOV, pode haver alguma estrutura dinâmica passível de ser modelada e prevista. Embora

Tabela 6.1. W_2 e índice de previsibilidade dos resíduos GARCH das séries para o teste BDS. Valores de W_2 superiores a 1,96 indicam a presença de estruturas possivelmente previsíveis.

Série	$ W_2 $	Valor limite de W_2	Índice de previsibilidade $PREV(BDS, \text{série})$
DJIA	0,234	1,96	0,350
SP500	1,60	1,96	1,40
IBOV	4,80	1,96	5,94
PETO	3,27	1,96	3,62
PETH	7,84	1,96	8,61
PETL	5,94	1,96	5,65
PETC	4,36	1,96	4,93

sua previsibilidade não seja significativamente elevada, esses resultados sugerem que modelos de PETH ou IBOV possam ter um desempenho melhor do que um passeio aleatório.

6.2 Previsão

Realizou-se uma modelagem dos dados visando mapear estruturas lineares e não lineares mas, ao mesmo tempo, minimizando a quantidade de parâmetros livres de treinamento. Esses poderiam influenciar o desempenho dos modelos fazendo com que a comparação dos dados se tornasse mais difícil. Portanto, priorizou-se a padronização dos modelos. Como enfatizado no Capítulo 5, o objetivo é comparar os dados e não os modelos. Foi feita também uma pequena discussão a respeito das medidas de desempenho de previsão.

Na escolha das estruturas das redes (quantidade de atrasos L e quantidade de neurônios Q), examinou-se a influência do desempenho das mesmas em função de L e Q . Em alguns casos, o aumento da quantidade de neurônios significou uma piora no desempenho. Em algumas séries, certos valores de atraso puderam ser identificados como melhores que outros. Mas esse comportamento não foi predominante no todo. Não se pôde perceber um padrão claro de desempenho em função de L e Q . Pelo contrário, a superfície de desempenho gerada em todos os casos foi bastante irregular. Essa característica pode indicar uma baixa previsibilidade dos dados, uma vez que o desempenho de previsão foi aparentemente aleatório e pouco dependente da estrutura escolhida.

O desempenho de previsão de alguns modelos foram melhores que os demais. Isso pode indicar que as séries cujos modelos obtiveram melhor desempenho têm um maior

nível de previsibilidade. Esse é o caso de PETH, IBOV, PETL e PETO. As séries com pior desempenho foram DJIA, SP500 e PETC. Esses resultados são reproduzidos novamente na Tabela 6.2.

Tabela 6.2. Desempenho da previsão um passo a frente das redes RBF em séries financeiras.

Série	U_2	PB (%)	Direção (%)
DJIA	1,00	47,8	49,3
SP500	1,00	45,6	47,7
IBOV	0,998	50,1	55,7
PETO	0,999	52,6	53,1
PETH	1,01	50,5	56,1
PETL	0,997	53,2	55,3
PETC	1,20	45,1	49,0

Os melhores resultados dessa tabela, em cada medida de desempenho, estão destacados. É interessante notar que as séries brasileiras (IBOV e PET) obtiveram melhores resultados que as demais (DJIA e SP500). Essa característica é estudada em Cajueiro & Tabak [2004], mas por meio do coeficiente de Hurst. Outra observação pertinente é o fato de o pior desempenho ser de uma séries de retornos de fechamento de mercados (PETC). Isso porque as séries de fechamento são, de longe, as mais estudadas, com a maior quantidade de tentativas de previsão. Pode-se pensar que esses estudos contribuíram para uma maior eficiência da série, como comentado em Malkiel [2007]: à medida em que o mercado descobre fontes de ineficiência (como, por exemplo, a existência de modelos de previsão lucrativos), essas são absorvidas no preço dos ativos. Pôde-se perceber também que, as tentativas de se prever a direção dos retornos foi mais bem sucedida do que a previsão do valor dos retornos. Essa constatação também é feita em Doeksen *et al.* [2005]. Pode-se pensar que é menos difícil separar classes de retornos em positivos ou negativos (problema de classificação) do que acertar o valor exato do retorno futuro (problema de regressão). A presença da heteroscedasticidade também pode influenciar negativamente a segunda abordagem.

A conclusão a que se chega, em vista dos resultados de previsão é que, em geral, há uma baixa previsibilidade das séries estudadas. Em contrapartida, pôde-se perceber que algumas séries apresentam desempenhos melhores que outras e ainda, melhores do que uma previsão por meio de um passeio aleatório. Ou seja, independente do modelo de previsão a ser usado, deve-se investigar se a série escolhida pode ser prevista ou não. Isso não só justifica o uso, como também demonstra a grande relevância do estudo da previsibilidade em séries financeiras.

6.3 Previsibilidade X Previsão

De maneira geral, os resultados de previsibilidade estão de acordo com o desempenho de previsão. As séries mais previsíveis em termos de W_2 e PREV (PETH, IBOV e PETL) foram também as que obtiveram melhores resultados na previsão (DIR). Além disso, o desempenho de SP500 e DJIA foi inferior ao de outras séries tanto na previsão como na previsibilidade. Essa comparação é resumida na Tabela 6.3.

Tabela 6.3. Comparação entre o índice de previsibilidade e os desempenhos de previsão PB e DIR (do melhor para o pior desempenho). Os valores de previsão melhores que o passeio aleatório ou cuja previsibilidade seja significativa estão destacados.

Ordem	Série	PREV	Série	PB	Série	DIR
1°	PETH	8,61	PETL	53,2	PETH	56,1
2°	IBOV	5,94	PETO	52,6	IBOV	55,7
3°	PETL	5,65	PETH	50,5	PETL	55,3
4°	PETC	4,93	IBOV	50,1	PETO	53,1
5°	PETO	3,62	DJIA	47,8	DJIA	49,3
6°	SP500	1,4	SP500	45,6	PETC	49
7°	DJIA	0,35	PETC	45,1	SP500	47,7

Os valores de previsibilidade estatisticamente significativos (maiores que 1,96) são colocados em destaque nessa tabela. Nela, também são destacados desempenhos de previsão melhores que o passeio aleatório (superiores a 50%, em relação a PB e DIR).

6.4 Conclusões

O presente trabalho estudou as séries financeiras em relação ao determinismo presente nas mesmas e sua previsibilidade baseada apenas em valores passados. Pôde-se perceber que todas as séries estudadas são diferentes do tradicional modelo do passeio aleatório, bastante comum na economia. Algumas das séries de retorno estudadas são não gaussianas, não lineares e não estacionárias. Foram utilizadas técnicas como a dimensão de correlação, a entropia amostral, a informação mútua média. Essas foram associadas à análise de dados sub-rogados. Com a finalidade de se quantificar as dependências temporais dos dados, o teste BDS de resíduos foi usado e por meio dele foram gerados índices de previsibilidade dos dados. Esses foram comparados a previsões das séries utilizando-se modelos RBF e percebeu-se que os resultados obtidos pelas técnicas de previsibilidade estão de acordo com os resultados de previsão, validando o uso dessas. Quanto ao estudo das estruturas determinísticas presentes nos dados,

percebeu-se uma fraca dependência linear em todas as séries. Em algumas, pôde-se identificar também uma dependência não linear previsível não muito significativa. Todas as séries apresentaram uma não estacionariedade bastante característica em seu processo de variância, detectada como uma dependência não linear heteroscedástica. A previsibilidade de algumas séries é equivalente à previsibilidade do passeio aleatório, entretanto foi possível identificar outras cuja previsibilidade é mais significativa. Esse último resultado pode ser validado pelo desempenho de previsão das séries.

6.5 Sugestões de Trabalhos Futuros

O presente trabalho fornece uma pequena amostra de alguns métodos para a detecção de determinismo e não linearidades em séries financeiras. Há vários outros métodos que não foram utilizados nesse trabalho, contudo parecem ser bastante promissores. Pode-se citar aqui o índice de Savit-Green. Em (Wu *et al.* [1993]), é derivada uma constante de normalização, que faz com que esse índice tenha média nula e variância unitária, como a estatística W_2 do teste BDS. Além disso, o índice quantifica a dependência temporal não linear em função dos atrasos da série utilizados, apontando para um melhor dimensionamento de modelos ou da escolha da dimensão de imersão.

Outras técnicas interessantes são mostradas em Gautama *et al.* [2004]. Os autores propõem uma técnica chamada *Delay Vector Variance* (DVV), baseada em outras duas, também mostradas no artigo: *deterministic versus stochastic*, ou DVS (Casdagli [1992]), e método $\delta - \epsilon$. A técnica proposta é baseada em um gráfico que compara a variância de respostas de um modelo linear local com a de respostas para o mesmo modelo, mas com dados sub-rogados (algoritmo IAAFT, semelhante ao algoritmo 2).

Em Nakamura & Small [2006], é apresentada a aplicação do algoritmo *small-shuffle* de dados sub-rogados em algumas séries financeiras. Sua principal qualidade é gerar dados que possuam a mesma tendência (de longo prazo) que os dados originais. Isso permite que as séries financeiras sejam trabalhadas diretamente, sem a necessidade de um pré-processamento, como a transformação de preços em retornos logarítmicos. Ou seja, a estacionariedade, em termos de média, não é exigida na análise. Isso é feito por meio de uma permutação aleatória local em que apenas os vizinhos de cada observação $y(t_1)$ (como $y(t_1 \pm 1)$, $y(t_1 \pm 2)$, \dots , $y(t_1 \pm \tau)$ em que $1 < t_1 < N$ e τ é pequeno) sejam reordenados. Embora esse procedimento possa ser aplicado em séries não estacionárias em média, o algoritmo não distingue dependências lineares de não lineares (como o caso de dependências da família ARCH). Ainda assim, o teste pode certamente ser incluído em trabalhos futuros sobre determinismo em séries financeiras,

a fim de se comparar a influência do pré-processamento quase sempre presente nesses tipos de dados.

Dimensões e entropias estimadas em uma série temporal são ferramentas bastante úteis quando se estuda a previsibilidade das mesmas. O estudo de dependências de longo prazo que utilizam o coeficiente de Hurst, como em Carbone *et al.* [2004], Cajueiro & Tabak [2004] e Couillard & Davison [2005], apresentam uma abordagem diferente da baseada naquelas estimativas. Conseqüentemente, sua utilização, em conjunto com dados sub-rogados ou a análise de resíduos (como feito neste trabalho, com o teste BDS) pode indicar rejeições de hipóteses para as quais dimensões e entropias são insensíveis.

Rotinas Implementadas em MATLAB

Neste apêndice são mostradas as principais funções implementadas no presente trabalho em linguagem MATLAB.

A.1 geraRegressoresAR.m

```
%Gera a matriz de regressores para uma estrutura AR(n) em que n é
% passado como parâmetro, além da série. O protótipo da função é:
% function [y,psi] = geraRegressoresAR(serie,n)
function [y,psi] = geraRegressoresAR(serie, n)

y = serie(n+1:end);

for i = 1:n
    psi(:,i) = serie(n+1-i:end-i);
end
end
```

A.2 calculaDimensaoGP.m

```
%function [m,logCm,logr] = calculaDimensaoGP(serie,vetor_m)
%Implementação do algoritmo-GP
%Algumas idéias de A fast general purpose algorithm for the computation
%of auto- and cross-correlation integrals from single channel data
function [m,logCm,logr] = calculaDimensaoGP(serie,vetor_m)
%m - número de elementos dos vetores os quais serão calculadas as
%distâncias. Se a série é imersa em uma dimensão 2, por exemplo,
%os vetores serão [x(t) x(t-1)] e isso é feito fazendo m=2;
```

```

if ~exist('vetor_m','var')
    %Varia m de 1 a 10
    vetor_m = 1:10;
end

%Calculo do vetor r
%vetor_r = linspace(0,std(serie),20);
%vetor_r = 0.1 * (max(serie)-min(serie)) * linspace(0.05,1,20)';
nbins=200;
%Seleção do r baseado no algoritmo do Small corrint.m
%get bins : distributed logarithmically
binl=log(min(diff(unique(serie))))-1; %smallest diff
%if isempty(binl),binl=eps;end; %just in-case
binh=log(max(vetor_m)*(max(serie)-min(serie)))+1;%seems to work
%if isnan(binh),binh=log(max(de)*max(y))+1;end; %just in-case
binstep=(binh-binl)./(nbins-1);
bins=binl:binstep:binh;
bins=exp(bins);
vetor_r = bins;

for im = 1:length(vetor_m)
    %im
    %Calcula C(m) para determinado m e vários r(vetor_r)
    Cm(im,:) = calculaCm(serie,vetor_m(im),vetor_r);
end
%Cada vetor linha i possui uma curva Cm com um 'm' e vários 'r'
logCm = log10(Cm);
m = vetor_m;
logr = log10(vetor_r);

%Calcula C(m) para determinado m
function [Cm] = calculaCm(serie,m,vetor_r)

%A série é o vetor u
u = serie(:);

```

```

%Comprimento do vetor u
N = length(u);
%Número de vetores
nVec = N-m+1;

%Define-se uma sequencia de vetores x(1),x(2),...x(N-m+1) em R^m
%em que x(i) = [u(i),u(i+1),...,u(i+m-1)] e nVec = N-m+1
for i = 1:nVec
    x(:,i) = u(i:i+m-1);
end

Cmr = zeros(1,length(vetor_r));%inicializa Cmr

%Para todos os vetores
for i = 1:nVec
    for j = i+1:nVec%Só considera distância entre pontos diferentes

        %Calcula a distância d entre cada vetor,...
        d = max(abs(x(:,i) - x(:,j)));
        %Incrementa os C(m,r);Se r>d == d<=r incrementa indices r (ir)
        ir = find(vetor_r > d);
        %Considera o dobro da quantidade de pontos encontrados, já que
        %tanto d(i,j) quanto d(j,i) devem ser contados
        Cmr(ir) = Cmr(ir) + 2;
    end
end

%Calculo do Cm
%Divide (Cim / nVec) pelo número de vetores ([N-m+1] = nVec)
Cm = (Cmr / nVec) / nVec;
%size(Cm)%Deve ser vetor linha

```

A.3 ljungbox_q.m

```

%Calcula a estatística q modificada por Ljung e Box (1978) para se realizar
%o teste de portmanteau. Recebe como parâmetro a série.

```

```

%O resultado é um vetor q(m) em que m é o número de atrasos até 25
%q(end) é a estatística q calculada com o número de atrasos fornecido
%para a função.
%function q = ljungbox_q(vetor)
function q = ljungbox_q(vetor)

m = 25; %Numero de atrasos a serem calculados
nSTDs = 2;%Numero de desvios padrao para o intervalo de confiança

%Gera FAC do vetor criado. A FAC é normalizada (ex. FAC(0) = 1)
[ACF,Lags,Bounds] = autocorr(vetor,m,nSTDs);
ACF = ACF(2:end);%Desconsidera-se ACF(1), equivalente a FAC(0), ou atraso=0

%Calcula Q(m)
T = length(vetor);
N = [T - [1:m]]';%vetor de valores do denominador do somatorio N=T-1,T-2,...
somatorio = cumsum( (ACF.^2)./N );%somatorio calculado até o atraso m
q = T * (T + 2) * somatorio;

```

A.4 estimaModeloAR.m

```

%function modelo = estimaModeloAR(serie,k)
%serie deve ser vetor coluna
%k é o número de parâmetros
%Se k não foi definido, detecta k por meio do teste Ljung-box, válido para
% k até 50;
%
%modelo é uma estrutura com:
%modelo.parametros - vetor de parametros na ordem y(k-1) y(k-2) etc.
%modelo.y - vetor usado para se calcular o modelo (y desejado)
%modelo.x - vetor usado para se calcular o modelo (regressores)
%modelo.yhat - vetor estimado por meio de x e parametros
%modelo.residuos - residuos (y - yhat)
%modelo.SSR - soma dos quadrados dos residuos (representa a variancia
% dos residuos)
%modelo.SST - soma dos quadrados dos desvios da série em relação a sua

```

```

%média (representa a variancia da série)
%modelo.SSE - soma dos quadrados dos desvios do modelo em relação a sua
%média (representa a variancia explicada pelo modelo)
%modelo.r2 - R^2 encontrado na regressão
%modelo.AIC - valor do critério de informação de akaike do modelo
function modelo = estimaModeloAR(serie,k)

%Se k não foi definido, detecta k por meio do teste Ljung-box,
% válido até k = 50;
if nargin <2
    for k = 1:50
        modelo = estimaModeloARfixo(serie,k);
        [h,pValue,stat,cValue] = lbqtest(modelo.residuos);
        if h == 0
            break;
        end
    end
    if k == 50
        disp('impossível modelar AR(50)')
    end
else
    modelo = estimaModeloARfixo(serie,k);
end

end
function modelo = estimaModeloARfixo(serie,k)

serie = serie(:);%serie deve ser vetor coluna

[y,x] = geraRegressoresAR(serie,k);

%Mínimos quadrados
teta = inv(x'*x)*x'*y;%Parametros estimados

yhat = x * teta;
u = y - yhat;

```

```
SSR = sum(u.^2);%Soma dos quadrados dos resíduos(variancia dos residuos)
SST = sum((y-mean(y)).^2);%Soma dos desvios da série(variancia da série)
SSE = sum((yhat-mean(y)).^2);%mean(y) == mean(yhat); (variancia
                                     %explicitada pelo modelo)

%R^2 do modelo
rsquare = SSE / SST; %ou 1 - (SSR/SST)

%Critério de informação de Akaike
n = length(y);
AIC = 2*k + n*(log(SSR));

%escreve resultados na estrutura
modelo.parametros = teta;
modelo.y = y;
modelo.x = x;
modelo.yhat = yhat;
modelo.residuos = u;
modelo.SSR = SSR;
modelo.SSE = SSE;
modelo.SST = SST;
modelo.r2 = rsquare;
modelo.AIC = AIC;
modelo.k = k;
end
```

A.5 dataSnooping.m

```
close all;
clear all;
clc;

carrega_series;
serie = ibov;
```

```
clearvars -except serie;

N = 1000;%Últimos N valores da série
%serie = serie(end-N+1:end);
Max_lag = 30;%Atraso máxima para ser calculada a média móvel. O min. é 2.

metade = floor(length(serie)/2);
serie1 = serie(1:metade);
serie2 = serie(metade+1:end);

%Otimiza mm para a primeira metade da série
for lag = 2:Max_lag

    [short,mm]= movavg(serie1,2,lag,1);
    mm = mm(lag:end);
    y = serie1(lag:end);

    sinais = [0;diff(y-mm>0)];

    portfolio_inicial = 0;
    dinheiro_inicial = 1000;

    portfolio = portfolio_inicial;
    dinheiro = dinheiro_inicial;
    for i = 1:length(sinais)
        if sinais(i) == 1 && dinheiro~=0%Compra
            portfolio = dinheiro / y(i);%Investe todo o dinheiro
            dinheiro = 0;
        end
        if sinais(i) == -1 && portfolio~=0%Vende
            dinheiro = portfolio * y(i);%"Liquida";
            portfolio = 0;
        end
    end
end
```

```

dinheiro = dinheiro + portfolio * y(end);%"Liquida";

lucro(lag) = 100 * (dinheiro - dinheiro_inicial) / dinheiro_inicial;

end
lucro = lucro(2:end);
lags = 2:Max_lag;

lucro_buy_and_hold = 100*(y(end)-y(1))/y(1)

figure;
plot(lags(1:15),lucro(1:15),'-r','LineWidth',2);
grid on;
ylabel('Lucro','fontSize',14);
xlabel('\tau','fontSize',14);
hold on;
line([lags(1) lags(15)],[lucro_buy_and_hold lucro_buy_and_hold],...
      'LineStyle','--','Color','b','LineWidth',3);
hold off;
lh = legend('Média Móvel(\tau)','Lucro Buy and Hold','location','northeast')
set(lh,'fontSize',14)
print('-depsc','fig_mm_lucro1');

[melhor_lucro indice] = max(lucro);
melhor_lucro
melhor_lag = lags(indice)

%Mostra resultado com melhor lag na serie1
[short,melhor_mm]= movavg(serie1,2,melhor_lag,1);
melhor_mm = melhor_mm(lag:end);
y = serie1(lag:end)';

figure;
plot(y,'.-b');
hold on;

```

```

plot(melhor_mm, '-r', 'LineWidth', 2);
ylabel('Preço', 'fontSize', 14);
xlabel('Dias', 'fontSize', 14);
str = sprintf('Média móvel(%d)', melhor_lag);
lh = legend('Preço', str, 'location', 'southeast')
set(lh, 'fontSize', 14)
hold off;
axis([5082 5119 1.6e-4 2.4e-4]);
print('-depsc', 'fig_mm_ibov');

%Otimiza mm para a segunda metade da série
for lag = 2:Max_lag

    [short,mm]= movavg(serie2,2,lag,1);
    mm = mm(lag:end);
    y = serie2(lag:end);

    sinais = [0;diff(y-mm>0)];

    portfolio_inicial = 0;
    dinheiro_inicial = 1000;

    portfolio = portfolio_inicial;
    dinheiro = dinheiro_inicial;
    for i = 1:length(sinais)
        if sinais(i) == 1 && dinheiro~=0%Compra
            portfolio = dinheiro / y(i);%Investe todo o dinheiro
            dinheiro = 0;
        end
        if sinais(i) == -1 && portfolio~=0%Vende
            dinheiro = portfolio * y(i);%"Liquida";
            portfolio = 0;
        end
    end
    dinheiro = dinheiro + portfolio * y(end);%"Liquida";

```

```

    lucro(lag) = 100 * (dinheiro - dinheiro_inicial) / dinheiro_inicial;

end

lucro = lucro(2:end);
lags = 2:Max_lag;

figure;
plot(lags(1:15),lucro(1:15),'-r','LineWidth',2);
grid on;
ylabel('Lucro','fontSize',14);
xlabel('\tau','fontSize',14);
hold on;
line([lags(1) lags(15)],[lucro_buy_and_hold lucro_buy_and_hold],...
      'LineStyle','--','Color','b','LineWidth',3);
hold off;
lh = legend('Média Móvel(\tau)','Lucro Buy and Hold','location','southeast')
set(lh,'fontSize',14)
print('-depsc','fig_mm_lucro2');

```

A.6 testaHeteroscedasticidade.m

```

%Teste de heteroscedasticidade
%function [H ,pValor,v,vCritico,grausLiberdade]
%           = testaHeteroscedasticidade(serie,teste)
%Testa a série de resíduos quanto a presença de heteroscedasticidade.
%Supõe-se que a série seja originada de um processo homocedástico(H=0).
%Considerando-se, essa hipótese, a estatística LM calculada pelo teste deve
%ser inferior ao valor limite (vCritico) em 95% dos casos.
%
%A função recebe como parâmetro a série a ser testada (deve ser um
%vetor coluna) e a variável teste, que pode possuir os valores:
% 'bp'      = realiza o teste segundo Breusch-Pagan
% 'white'   = realiza o teste segundo White.

```

```

%A função retorna H:
%H:
% H = 1 - A série em questão possui heteroscedasticidade. O teste
% foi capaz de rejeitar a hipótese de homoscedasticidade. Correponde ao
% resultado: v >= vCritico
% H = 0 - Não se pode dizer se a série possui heteroscedasticidade. O
% teste não foi capaz de rejeitar a hipótese. Corresponde ao
% resultado: v < vCritico
%
% Isso é feito considerando que o modelo usado nos teste foi o adequado.
%
%pValor: Pode ser visto como a probabilidade de o resultado de o teste
%falhar, dado que a hipótese tenha sido rejeitada.
%
%v: Valor da estatística LM, usada no teste. Se esse valor é maior que o
%valor crítico, a hipótese de homoscedasticidade é rejeitada.
%
%vCritico: Valor da CDF (qui-quadrado com K graus de liberdade) que
%corresponde a 95% da área de sua PDF. É o limite de rejeição, a partir do
%qual se pode rejeitar a hipótese de homoscedasticidade
%
%grausLiberdade: a quantidade de graus de liberdade do qui-quadrado
%Os testes são implementados seguindo o livro:
%Wooldridge, J.M. (2 ed.) Introductory Econometrics: A Modern Approach -
% Thomson - cap. 8 (Heteroskedasticity)

function [H ,pValor,v,vCritico,grausLiberdade] = ...
    testaHeteroscedasticidade(serie,teste)
K = 5;%número de regressores (desconsiderando o delta0 abaixo)
%K tem relação com o número de graus de liberdade do qui-quadrado limite

%Constroi-se o modelo (White):
%  $u^2(k) = \text{delta0} +$ 
%  $\text{delta1} * xu2(k) + \text{delta2} * xu2(k-1) + \dots +$ 
%  $\text{delta3} * xu(k) + \text{delta4} * xu(k-1) + \dots +$ 

```

```

%          delta5 * xu(k) * xu2(k) + delta6 * xu(k-1) + xu2(k-1) + ...
%Em que xu2(k),xu2(k-1),... são os regressores de u^2.
%Para o modelo de Breusch-Pagan utiliza-se apenas as duas primeiras linhas
%da equação acima.
%ps:A equação acima poderia ser modificada para qualquer conjunto de
%variáveis independentes, como por exemplo, escolher apenas a primeira e
%terceira linha da equação. A única mudança a ser feita se refere a
%modificar o número de graus de liberdade para um valor igual a quantidade
%de variáveis independentes dessa equação como descrito no final da pg. 258

u = serie;

%Quadrado dos resíduos
u2 = u.^2;

%Regressores
[u,xu] = geraRegressoresAR(u, K);
[u2,xu2] = geraRegressoresAR(u2, K);

%Monta matriz de regressores para cada um dos testes
if strcmp(teste,'bp')%breush-pagan
    X = [ones(length(u2),1) xu];%coluna com ones representa o
        %regressor para delta0
    grausLiberdade = size(X,2) - 1; %Número de regressores
        %(desconsiderando delta0)
end
if strcmp(teste,'white')%white
    X = [ones(length(u2),1) xu];%coluna com ones representa o
        % regressor para delta0
    %X = [X xu2];%Regressores xu2(k),xu2(k-1),etc.
    %Inclui regressores cruzados: u(k-1)*u2(k-1),u(k-1)*u2(k-2),etc.
    for i = 1:K%Para cada regressor u(k-i)
        for j = 1:K%Para cada regressor u(k-j)
            if (j>=i)
                X = [X xu(:,i).* xu(:,j)];
            end
        end
    end
end

```

```

end
    grausLiberdade = size(X,2) - 1; %Número de regressores
                                %(desconsiderando delta0)
end

%Mínimos quadrados
teta = (X'*X)\X'*u2;%Parametros estimados

%Encontra resíduos do modelo
u2hat = X * teta;
residuo_u2 = u2 - u2hat;

%Funções para cálculo do R^2
SSR = sum(residuo_u2.^2);%Soma dos quadrados dos resíduos(variancia
                        % dos resíduos)
SST = sum((u2-mean(u2)).^2);%Soma dos desvios da série(variancia
                        % da série)
SSE = sum((u2hat-mean(u2)).^2);%mean(u2) == mean(u2hat); (variancia
                        %explicada pelo modelo)

%R^2 do modelo
r2 = SSE / SST; %ou 1 - (SSR/SST)

%LM statistic
LM = r2 * length(u2);
v = LM;

%p-valor
pValor = 1-chi2cdf(LM,grausLiberdade);

%Valor limite para o intervalo de confiança de 95%
vCritico = chi2inv(0.95,grausLiberdade);

%Se o modelo usado nos teste for o adequado,
if (LM >= vCritico)%Rejeita a hipótese de homoscedasticidade
    H = 1;%A série possui heteroscedasticidade
else%O teste não foi capaz de rejeitar a hipótese

```

```
H = 0;%Não se pode dizer se a série possui heteroscedasticidade
end
```

A.7 calculaErros.m

```
function erros = calculaErros(y,yd)

erro = yd - y;
erro_rw = yd - 0;

%Root mean Square Error
erros.rmse = sqrt(mean(erro.^2));
%Root mean Square Error cometido pelo Random Walk
erros.rmse_rw = sqrt(mean(erro_rw.^2));
%Mean absolute Error
erros.mae = mean(abs(erro));
%Percent better
erros.pb = 100 * sum((abs(erro)-abs(erro_rw))<0) / length(yd);
%U2 de Theil
erros.u2 = erros.rmse/erros.rmse_rw;
%Mean Relative Absolute Error
erros.mrae = mean(abs(erro) ./ abs(erro_rw));
%Median Relative Absolute Error
erros.mdrae = median(abs(erro) ./ abs(erro_rw));
%Geometric Mean Relative Absolute Error
erros.gmrae = geomean(abs(erro) ./ abs(erro_rw));

%Acertou a direção (em porcentagem)
erros.direcao = 100 * (sum((y>=0)&(yd>=0))+sum((y<0)&(yd<0))) / length(yd);

%erros.erros = erro;
end
```

A.8 analiseBDS.m

```
function [estrutura, prev] = analiseBDS(serie,Nsur)

%Parametros para a funcao BDS
maxdim = 2;
distance = 1.5;
flag = 0;
maxram = 500;%Importante ajustar maxram para a quantidade disponivel
                % de memoria para o MATLAB, senao, trava

z = geraSurrogates(serie,'alg0',Nsur);

%Calcula BDS para os Surrogates
for n = 1:Nsur
    %Mostra status atual do calculo
    if (~mod(n,10) || n==Nsur)
        str=sprintf('Calculando BDS para sub-rogados... %g%% concluido',...
                    (100*n)/Nsur);

        disp(str);
    end

    y = z(:,n);
    %Calcula a estatística BDS para a série n
    [W, SIG, C, C1, K] = bds (y, maxdim, distance, flag, maxram);
    w_sur(n) = W;
end

%Calcula a estatística BDS para a série original
disp('Calculando BDS para série original. ');
y = serie;
[W, SIG, C, C1, K] = bds (y, maxdim, distance, flag, maxram);
w_serie = W;

if Nsur > 0
    %Compara bds da serie e surrogates
```

```

Nbins = 20;
[n,x] = hist(w_sur(:),Nbins);
n = n / (sum(n)); %Normaliza para que fique em função
                  % da probabilidade

figure;
bar(x,n,'FaceColor',[.89 .94 .9],'EdgeColor','k');
hold on;
line([w_serie w_serie],[0 max(n)*1.3],'Color','r','LineWidth',3);%Linha
                                     %que indica bds para a série original
str = sprintf('BDS Série Original: %g', w_serie);
legend('Histograma W_2 (permutações)',str,'location','northeast');
xlabel('W_2');
ylabel('Distribuição de W_2 (permutações)');
hold off;

media_sur = mean(w_sur);
std_sur = std(w_sur);

prev = abs((w_serie - media_sur)) ./ std_sur;

estrutura.w_sur = w_sur;

else
    prev = NaN;
end

estrutura.w_serie = w_serie;

end

```

A.9 analisesurrogates.m

```

function [estrutura, prev] =
    analiseSurrogates(serie,Nsur,algoritmo,est,msg,fig)
%function [] = analiseSurrogates(serie,Nsur,algoritmo,est,msg,fig)

```

```

%-serie é a série a ser analisada
%-Nsur é o número de séries sub-rogadas a serem geradas para a análise
%-algoritmo é a seleção do alg. de geração de dados sub-rogados.
%Pode ser 'alg0' para o algoritmo 0, 'alg1' para o algoritmo 1,
%estrutura retorna os resultados dos cálculos das estatísticas na série
%original e nos sub-rogados
%-est é a estatística usada. Pode ser igual a 'todas', 'd2', 'sampen'
                                     % ou 'ami'

%-msg é um flag que, se igual a 1 mostra as mensagens de progressao dos
%cálculos. Se for igual a 0, oculta as mensagens.
%-img é um flag que ativa a exibição das figuras da análise.

%Dimensão de imersão
m = 5;

%Gera surrogates
%Se for PPS,...
if strcmp(algoritmo,'PPS')
    z = geraSurrogatesPPS(serie,Nsur,m);
else
    if strcmp(algoritmo,'sinal')%Se for sinal,...
        z = sinalAleatorio(serie,Nsur);
    else
        if strcmp(algoritmo,'ss')%Se for small-shuffle
            disp('gerando small-shuffle');
            z = smallshuffle(serie,Nsur);
        else%Se for alg. 0, 1 ou 2,...
            z = geraSurrogates(serie,algoritmo,Nsur);
        end
    end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%Início calculo D_2%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Calcula a estatística D2(dim. de corr.) para os Surrogates
if strcmp(est,'todas') || strcmp(est,'d2')

```

```

for n = 1:Nsur
    %Mostra status atual do calculo
    if (~mod(n,5) || n==Nsur) && msg == 1
        str = sprintf(...
            'Calculando D2 para sub-rogados... %g%% concluido',...
            (100*n)/Nsur);
        disp(str);
    end

    [M,DC,EPS0,CI] = calculaDimCorJudd(z(:,n),m);

    dimcor_sur{n}.epsilon0 = DC{1}(1,:);
    dimcor_sur{n}.d2 = DC{1}(2,:);
end

%Calcula a estatistica D2 para a serie original
if msg == 1
    disp('Calculando D2 para série original. ');
end
[M,DC,EPS0,CI] = calculaDimCorJudd(serie,m);
dimcor_serie.epsilon0 = DC{1}(1,:);
dimcor_serie.d2 = DC{1}(2,:);

if fig
    figure;
    hold on;
    for i = 1:Nsur
        hs = plot(log(dimcor_sur{i}.epsilon0),dimcor_sur{i}.d2,'-b');
    end
    h = plot(log(dimcor_serie.epsilon0),dimcor_serie.d2,'--r',...
        'LineWidth',4);

    hold off;
    grid minor;
    xlabel('log(\epsilon_0)', 'fontsize',14);
    ylabel('D_2', 'fontsize',14);
    legend([hs h], 'D_2 Sub-rogados', 'D_2 Série', 'Location', 'NorthWest');
end

```

```

end
estrutura.d2.sur = dimcor_sur;
estrutura.d2.serie = dimcor_serie;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%Fim calculo D_2%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%Início calculo SampEn%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Calcula a SampEn para os Surrogates
if strcmp(est,'todas') || strcmp(est,'sampen')
    %epsilon = 0.2 * std(serie);
    m = 2;

    for n = 1:Nsur
        %Mostra status atual do calculo
        if (~mod(n,10) || n==Nsur) && msg == 1
            str = sprintf(...
                'Calculando SampEn para sub-rogados... %g%% concluido',...
                (100*n)/Nsur);
            disp(str);
        end
        %É importante que cada Sampen seja calculada em relação ao desvio
        %padrão da sua série
        epsilon = 0.2 * std(z(:,n));
        %Como o resultado do sampen é calculado para m=0,1,...,m-1,
        %deve-se chamar a função com m+1 e pegar o ultimo elemento
        % do resultado. Esse será o sampen(m)
        SampEn = calculaSampEn(z(:,n),m+1,epsilon);
        SampEn_sur(n) = SampEn(end);
    end

    %Calcula a estatística SampEn para a série original

```



```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%%%Início calculo AMI%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%Calcula a estatística AMI para os Surrogates
```

```
if strcmp(est,'todas') || strcmp(est,'ami')
```

```
    lags = 1:10;
```

```
    for n = 1:Nsur
```

```
        %Mostra status atual do calculo
```

```
        if (~mod(n,10) || n==Nsur) && msg == 1
```

```
            str = sprintf(...
```

```
                'Calculando AMI para sub-rogados... %g%% concluido',...
```

```
                (100*n)/Nsur);
```

```
            disp(str);
```

```
        end
```

```
        [AMI,lag]=calculaAMI(z(:,n),z(:,n),lags);
```

```
        AMI_sur(:,n) = AMI';%vetores coluna para cada surrogate
```

```
    end
```

```
%Calcula a estatística AMI para a serie original
```

```
if msg == 1
```

```
    disp('Calculando AMI para série original.');
```

```
end
```

```
[AMI,lag]=calculaAMI(serie,serie,lags);
```

```
AMI_serie = AMI';%vetor coluna
```

```
%Compara a estatística AMI da serie e surrogates
```

```
if fig
```

```
    figure;
```

```
    hold on;
```

```
    for n = 1:Nsur
```

```
        hs = plot(lags,AMI_sur(:,n),'-b');
```

```
    end
```

```

h = plot(lags,AMI_serie,'--r','LineWidth',4);
legend([h hs],'AMI Série','AMI Sub-rogados');
xlabel('\tau','fontsize',14);
ylabel('AMI(\tau)','fontsize',14);
grid on;
hold off;

end

estrutura.ami.sur = AMI_sur;
estrutura.ami.serie = AMI_serie;

end

%PREV será a distância entre ami da série e da média dos sub-rogados
%normalizado pelo desvio padrão dos sub-rogados
for i = 1:length(lags)
    media_AMI_sur(i,1) = mean(AMI_sur(i,:));
    std_AMI_sur(i,1) = std(AMI_sur(i,:));
end

end

% AMI_serie
% media_AMI_sur
% std_AMI_sur
prev = abs((AMI_serie - media_AMI_sur)) ./ std_AMI_sur;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%Fim calculo AMI%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
end

```

A.10 estimaModeloRBF.m

```
function modelo = estimaModeloRBF(xtr,ydtr,Nneuronios,sigma)
```

```
N = size(xtr,1);%Número de "vetores linha" de entrada.
```

```
%Define os centros
```

```
%Treinamento k-médias
```

```

opts = statset('Display','off');
[idx, ctrs] = kmeans(xtr, Nneuronios, ...
    'Distance', 'sqEuclidean', ... '%city', ...
    'Replicates', 5, ...
    'Options', opts);
c = ctrs;

%Define os raios
if nargin < 4
    sigma = calcula_Raios(c, 'menor');
end

%Pesos da primeira camada
% c
% sigma

%Calcula entradas e saida da camada de saida da rede
for i=1:Nneuronios%Para todos os neuronios
    for j=1:N%Para todos os dados de entrada
        %Distancia do vetor de entrada j ao centro i
        u = norm(xtr(j,:) - c(i,:));
        %Saida da camada escondida GAUSSIANA
        %Cada linha j tem a resposta de cada
        %estimulo (dado de entrada) j ao neuronio i
        y1(j,i) = exp( -(u^2) / (2*sigma(i)^2) );
    end
end

y1 = [ones(N,1) y1]; %K+1 elementos, por causa do bias
    % (o bias é como um neuronio cuja saída é sempre um)

%Estimação dos parametros da segunda camada por MQ classico
psi = y1;%Lembrar que os dados devem estar em linhas e nao em colunas
yd = ydtr;
% size(psi)

```

```

% size(yd)
%tetaMQ = inv(psi'*psi)*psi'*yd;
%tetaMQ = (psi'*psi)\psi'*yd;
%tetaMQ = inv(psi'*psi)*psi'*yd;
tetaMQ = pinv(psi'*psi)*(psi'*yd);

% yMQ = psi * tetaMQ;

w2 = tetaMQ;

modelo.c = c;%Centros
modelo.sigma = sigma;%Raios
modelo.w2 = w2;%Pesos da segunda camada

end

function sigma = calcula_Raios(c,alg)

Nneu = size(c,1);

%Calculo dos raios
for i=1:Nneu%Calcula a distancia entre os centros

    %Inicializa a variável "menor distancia" com a distância ao próx. vetor
    if (i~=Nneu)%Quando o vetor i for o último, inicializa com a
        % dist. entre ele e o primeiro
        menor(i) = norm(c(i,:)-c(i+1,:));
    else
        menor(i) = norm(c(i,:)-c(1,:));
    end

    %Encontra menor distancia do vetor i

```

```

    for j=1:Nneu%roda em todos os vetor j,
        if (i~=j)%em que j é diferente de i
            dist = norm(c(i,:)-c(j,:));
            if (dist<menor(i))
                menor(i) = dist;
            end
        end
    end
end

menor = menor(:);

if strcmp(alg,'mediamenor')
    %Procedimento 1: Todos os raios iguais
    %Calcula sigma, o raio dos neuronios
    sigma = ones(Nneu,1) .* mean(menor);
end
if strcmp(alg,'menor')
    %Procedimento 2: Raios diferentes, menor distancia entre os centros
    %Calcula sigma, o raio dos neuronios
    sigma = menor;
end
end

%function [w,y] = mykmeans(dados,K)
%Calcula os K pesos(centros dos padroes) para o vetor de dados fornecido
%Cada dado está em uma coluna
%w são os K centros obtidos. Cada coluna de w representa um centro.
%y é a identificação de qual dados pertence a qual centro. O valor de y
%identifica uma coluna de w.
function [w,y] = mykmeans(dados,K)

N_dados = size(dados,2);
N_dimensoes = size(dados,1);

dados = dados(:,randperm(size(dados,2)));%Vetor de dados em

```

```

                                % ordem aleatoria

%Os primeiros K dados ja foram atribuidos aos centros w
for i=1:K
    w(:,i) = dados(:,i);
end

% figure;
% plot(dados(1,:),dados(2,:),'.r');
% hold on;
% plot(w(1,:),w(2,:),'*b');
% hold off;
% pause;

mudou = 1;%Flag que indica a mudança dos pesos
while (mudou)
    for i=1:N_dados
        %Associa o novo dado ao centro mais próximo
        for j=1:K
            %Calcula as distâncias entre o dado i e o centro j
            yi(j) = norm(dados(:,i)-w(:,j));
        end
        [C,I] = min(yi);%Encontra o centro mais proximo
        %Atribui ao dado i o centro do grupo mais proximo I
        y(i) = I;
    end

    %Calcula os novos centros

    soma = zeros(N_dimensoes,K);
    for i=1:N_dados
        %Soma os dados relativos a cada centro
        soma(:,y(i)) = soma(:,y(i)) + dados(:,i);
    end
    %O novo centro será o vetor da média dos vetores pertencentes a cada
    %centro
    for i=1:K

```

```

        %Calcula a média
        w_novo(:,i) = soma(:,i)./length(find(y==i));
    end
    if (isequal(w,w_novo))
        mudou = 0;
    else
        w = w_novo;
    end

%   plot(dados(1,:),dados(2,:),'.r');
%   hold on;
%   plot(w(1,:),w(2,:),'*b');
%   hold off;
%   pause;
end
end

```

A.11 preparaDados.m

```

%function [xtr,ydtr,xval,ydval,xtest,ydtest]
%           = preparaDados(serie,Ndelays,Nval,Ntest)
%serie - a série temporal a ser transformada em vetores
%        de entrada e saída para a rede TDNN
%Ndelays - Número de atrasos (AR) a serem considerados para
%         a criação do vetor de entrada TDNN
%Nval - 'porcentagem' de dados da série a serem destinados
%       ao treinamento.
%   Ex.: Nval = 0.2 -> 20% (finais) dos dados destinados
%         a validação
%Ntest - 'porcentagem' de dados da série a serem destinados
%       ao teste final.
%   Ex.: Ntest = 0.2 -> 20% (após os dados de validação)
%         destinados ao teste final. Nesse caso os dados
%         de validação representam a validação cruzada como
%         em um treinamento mobj que também minimiza a
%         complexidade da rede

```

```
function [xtr,ydtr,xval,ydval,xtest,ydtest] = ...
    preparaDados(serie,Ndelays,Nval,Ntest)

%Vetor serie sempre é vetor coluna
serie = serie(:);

xtr = [];
ydtr = [];
xval = [];
ydval = [];
xtest = [];
ydtest = [];

if nargin < 2
    Ndelays = 1
    Nval = 0;
    Ntest = 0;

end
if nargin < 3
    Nval = 0;
    Ntest = 0;
end
if nargin < 4
    Ntest = 0;
end

Ntr = 1 - Nval - Ntest;

%Gera vetores de entrada e saída
[yd,x] = geraRegressoresAR(serie, Ndelays);

%Comprimento da série
N = length(yd);

%Último valor da série de treinamento
fim_tr = floor(Ntr*N);
```

```

%Último valor da série de validação
fim_val = fim_tr + floor(Nval*N);

%Dados de treinamento
xtr = x(1:fim_tr,:);
ydtr = yd(1:fim_tr);
%Dados de Validação
xval = x(fim_tr+1:fim_val,:);
ydval = yd(fim_tr+1:fim_val);
%Dados de teste
xtest = x(fim_val+1:end,:);
ydtest = yd(fim_val+1:end);

end

```

A.12 previsaoRBF.m

```

function y = previsaoRBF(modelo,x)

c = modelo.c;
sigma = modelo.sigma;
w2 = modelo.w2;

Nneu = size(c,1);%Numero de neuronios

for i=1:Nneu
    %Distancia do vetor de entrada ao centro i
    u = norm(x-c(i,:));
    %Saida da camada escondida GAUSSIANA
    yG(i) = exp( -(u^2) / (2*sigma(i)^2) );
end
%A saida da camada escondida é a entrada para a camada de saída
x2 = yG;

%Adiciona o bias

```

```
x2 = [1 x2];
```

```
y2 = x2 * w2;
```

```
y = y2;
```

```
end
```

Referências Bibliográficas

- Aguirre, L. A. (2007). *Introdução à identificação de sistemas: técnicas lineares e não-lineares aplicadas a sistemas reais*. Editora UFMG, Belo Horizonte, 3 edição.
- Aguirre, L. A.; Barbosa, B. H. & Braga, A. P. (2010). Prediction and simulation errors in parameter estimation for nonlinear systems. *Mechanical Systems and Signal Processing*, 24:2855–2867.
- Ang, K. K. & Quek, C. (2006). Stock trading using RSPOP: a novel rough set-based neuro-fuzzy approach. *IEEE Transactions on Neural Networks*, 17(5):1301–1315.
- Armstrong, S. & Collopy, F. (1992). Error measures for generalizing about forecasting methods: empirical comparisons. *International Journal of Forecasting*, 8:69–80.
- Bachelier, L. (1900). *Théorie de la spéculation*. Tese de doutorado, Annales scientifiques de l'École Normale Supérieure, Paris.
- Balwin, T. T.; Bommer, W. H. & Rubin, R. S. (2006). *Louis Bachelier's theory of speculation: the origins of modern finance*, pp. 15–79. Princeton University Press, Princeton.
- Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.
- Braga, A. P.; Carvalho, A. C. P. L. F. & Ludermir, T. B. (2007). *Redes Neurais Artificiais: teoria e aplicações*. LTC, Rio de Janeiro, 2 edição.
- Brock, W.; Lakonishok, J. & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance*, 47(5):1731–1764.

- Brock, W. A.; Scheinkman, J. A.; Dechert, W. D. & LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, 15(3):197–235.
- Brooks, C. & Henry, O. T. (2000). Can portmanteau nonlinearity tests serve as general mis-specification tests? evidence from symmetric and asymmetric GARCH models. *Economics Letters*, 67:245–251.
- Burgess, A. & Refenes, A.-P. (1999). Modelling non-linear moving average processes using neural networks with error feedback: an application to implied volatility forecasting. *Signal Processing*, 74:89–99.
- Caetano, M. A. L. & Yoneyama, T. (2009). A new indicator of imminent occurrence of drawdown in the stock market. *Physica A*, 388:3563–3571.
- Cajueiro, D. O. & Tabak, B. M. (2004). The Hurst exponent over time testing the assertion that emerging markets are becoming more efficient. *Physica A*, 336:521–537.
- Campbell, J. Y.; Lo, A. W. & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton University Press, New Jersey.
- Cao, L. & Soofi, A. S. (1999). Nonlinear deterministic forecasting of daily dollar exchange rates. *International Journal of Forecasting*, 15:421–430.
- Caporale, G. M.; Ntantamis, C.; Pantelidis, T. & Pittis, N. (2004). The BDS test as a test for the adequacy of a GARCH specification - a Monte Carlo study. *Economic Series - Institute for Advanced Studies*, 156.
- Carbone, A.; Castelli, G. & Stanley, H. (2004). Time-dependent Hurst exponent in financial time series. *Physica A*, 344:267–271.
- Casdagli, M. (1992). Chaos and deterministic versus stochastic non-linear modelling. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54(2):303–328.
- Chen, S. & Billings, S. A. (1992). Neural networks for nonlinear dynamic system modelling and identification. *International Journal of Control*, 56(2):319–346.
- Costa, M. A.; Braga, A. P. & de Menezes, B. R. (2003). Improving neural networks generalization with new constructive and pruning methods. *Journal of Intelligent & Fuzzy Systems*, 13:75–83.

- Couillard, M. & Davison, M. (2005). A comment on measuring the Hurst exponent of financial time series. *Physica A*, 348:404–418.
- Cover, T. M. (2006). *Elements of information theory*. John Wiley & Sons Inc., New Jersey.
- Darbellay, G. A. & Wuertzc, D. (2000). The entropy as a tool for analysing statistical dependences in financial time series. *Physica A*, 287:429–439.
- de Albuquerque Teixeira, R.; Braga, A. P.; Takahashi, R. H. & Saldanha, R. R. (2000). Improving generalization of MLPs with multi-objective optimization. *Neurocomputing Letters*, 35:189–194.
- Deboeck, G. J. (1994). *Trading on the edge: neural, genetic and fuzzy systems for chaotic financial markets*. John Wiley & Sons, Inc., New York.
- Dimson, E. & Mussavian, M. (1998). A brief history of market efficiency. *European Financial Management*, 4(1):91–193.
- Doeksen, B.; Abraham, A.; Thomas, J. & Paprzycki, M. (2005). Real stock trading using soft computing models. *Proceedings of the International Conference on Information Technology: Coding and Computing*.
- Fama, E. (1965). *The behaviour of stock market prices*. Tese de doutorado, Journal of Business.
- Fama, E. F. (1970). Efficient capital markets: a review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.
- Fama, E. F. (1991). Efficient capital markets: II. *The Journal of Finance*, 46(5):1575–1617.
- Fiedler-Ferrara, N. & Prado, C. P. C. (1994). *Caos - uma Introdução*. Editora Edgard Blücher Ltda., São Paulo.
- Garman, M. B. & Klass, M. J. (1980). On the estimation of security price volatility from historical data. *Journal of Business*, 53:67–78.
- Gautama, T.; Mandic, D. P. & Hulle, M. M. V. (2004). The delay vector variance method for detecting determinism and nonlinearity in time series. *Physica D*, 190:167–176.

- Grassberger, P. & Procaccia, I. (1983). Characterization of strange attractors. *Physical Review Letters*, 50:346–349.
- Haykin, S. (2001). *Redes neurais: princípios e práticas*. Bookman, Porto Alegre, 2 edição.
- Hinich, M. J.; Mendes, E. M. & Stone, L. (2005). Detecting nonlinearity in time series: surrogate and bootstrap approaches. *Studies in Nonlinear Dynamics & Econometrics*, 9(4).
- Judd, K. (1991). An improved estimator of dimension and some comments on providing confidence intervals. *Physica D*, 56:216–228.
- Judd, K. (1994). Estimating dimension from small samples. *Physica D*, 71:421–429.
- Kanzler, L. (1999). Very fast and correctly sized estimation of the BDS statistic. Disponível em SSRN: <http://ssrn.com/abstract=151669>.
- Ko, P.-C. (2009). Option valuation based on the neural regression model. *Expert Systems with Applications*, 36:464–471.
- Kokshenev, I. & Braga, A. P. (2008). A multi-objective approach to RBF network learning. *Neurocomputing*, 71:1203–1209.
- Kugiumtzis, D. (2000). Surrogate data test on time series. *Contribuição para o livro Nonlinear Deterministic Modelling and Forecasting of Economic and Financial Time Series*.
- Kuo, R.; Chen, C. & Hwang, Y. (2001). An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. *Fuzzy Sets and Systems*, 118:21–45.
- Kyrtsou, C. & Terraza, M. (2003). Is it possible to study chaotic and ARCH behaviour jointly? application of a noisy Mackey-Glass equation with heteroskedastic errors to the Paris stock exchange returns series. *Computational Economics*, 21:257–279.
- Lee, K. & Jo, G. (1999). Expert system for predicting stock market timing using a candlestick chart. *Expert Systems with Applications*, 16:357–364.
- Lima, R. C. & Ohashi, A. (1999). The efficient market hypothesis and the dynamic behavior of future sugar prices. *Revista Econômica do Nordeste*, 30:484–493.

- Ljung, G. M. & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.
- Lo, A. W. & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: evidence from a simple specification test. *The Review of Financial Studies*, 1(1):41–66.
- Lo, A. W. & MacKinlay, A. C. (2002). *A non-random walk down wall street*. Princeton University Press, New Jersey.
- Malkiel, B. G. (2007). *A random walk down wall street: the time-tested strategy for successful investing*. W. W. Norton & Company, New York.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419.
- Mandelbrot, B.; Fisher, A. & Calvet, L. (1997). A multifractal model of asset returns. *Cowles Foundation for Research in Economics*.
- Matsura, E. (2007). *Comprar ou vender: como investir na bolsa usando análise gráfica*. Saraiva, São Paulo.
- Monteiro, L. H. A. (2006). *Sistemas dinâmicos*. Editora Livraria da Física, São Paulo.
- Nakamura, T. & Small, M. (2006). Testing for dynamics in the irregular fluctuations of financial data. *Physica A*, 366:377–386.
- Nascimento Jr., H. B.; Fulco, U.; M.L.Lyra; Serva, M. & Viswanathan, G. G. (2007). Porque as bolsas de valores quebram: a origem das caudas grossas nas distribuições de retornos. *Revista Brasileira de Ensino de Física*, 29(3):341–346.
- Pantazopoulos, K. N.; Tsoukalas, L. H.; Bourbakis, N. G.; Brün, M. J. & Houstis, E. N. (1998). Financial prediction and trading strategies using neurofuzzy approaches. *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, 28(4):520–531.
- Pincus, S. (1995). Approximate entropy (ApEn) as a complexity measure. *American Institute of Physics - Chaos*, 5(1):110–117.
- Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 88:2297–2301.

- Pincus, S. M. & Kalman, R. E. (2004). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 101(38):13709–13714.
- Refenes, A. P. N.; Burgess, A. N. & Bentz, Y. (1997). Neural networks in financial engineering: a study in methodology. *IEEE Transactions on Neural Networks*, 8(6):1222–1267.
- Richman, J. S. & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *AJP - Heart and Circulatory Physiology*, 278:2039–2049.
- Rodrigues, D. D. (2002). Previsão de fornecimento de energia elétrica: estudo de caso do ponto de vista do barramento. Dissertação de mestrado, Universidade Federal de Minas Gerais, Belo Horizonte.
- Saad, E. W.; Prokhorov, D. V. & Wunsch, D. C. (1998). Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. *IEEE Transactions on neural networks*, 9(6):1456–1470.
- Scheinkman, J. A. & LeBaron, B. (1989). Nonlinear dynamics and stock returns. *The Journal of Business*, 62(3):311–337.
- Sitte, R. & Sitte, J. (2002). Neural networks approach to the random walk dilemma of financial time series. *Applied Intelligence*, 16(1):163–1171.
- Small, M. (2005). *Applied nonlinear time series analysis - applications in physics, physiology and finance*. World Scientific, Singapore.
- Small, M. & Judd, K. (1998). Correlation dimension: a pivotal statistic for non-constrained realizations of composite hypotheses in surrogate data analysis. *Physica D*, 120:386–400.
- Small, M. & Tse, C. K. (2003a). Detecting determinism in time series: the method of surrogate data. *IEEE Transactions On Circuits And Systems I: Fundamental Theory And Applications*, 50(5):663–672.
- Small, M. & Tse, C. K. (2003b). Determinism in financial time series. *Studies in Nonlinear Dynamics & Econometrics*, 7(3).
- Small, M.; Yu, D. & Harrison, R. G. (2001). Surrogate test for pseudoperiodic time series data. *Physical Review Letters*, 87(18).

- Theiler, J.; Eubank, S.; Longtin, A.; Galdrikian, B. & Farmer, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58:77–94.
- Torres, R.; Bonomo, M. & Fernandes, C. (2002). A aleatoriedade do passeio na Bovespa: testando a eficiência do mercado acionário brasileiro. *Revista Brasileira de Economia*, 56(2):199–247.
- Tsay, R. S. (2005). *Analysis of financial time series*. Wiley Interscience, 2 edição.
- Voit, J. (2003). From brownian motion to operational risk: statistical physics and financial markets. *Physica A*, 321:286–299.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838.
- Widman, G.; Lehnertz, K.; Jansen, P.; Meyer, W.; Burr, W. & Elger, C. (1998). A fast general purpose algorithm for the computation of auto- and cross-correlation integrals from single channel data. *Physica D*, 121:65–74.
- Wooldridge, J. M. (2003). *Introductory econometrics: a modern approach*. Mason: Thomson South-Western, 2 edição.
- Wu, K.; Savit, R. & Brock, W. (1993). Statistical tests for deterministic effects in broad band time series. *Physica D*, 69:172–188.

