

UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES  
PÓS-GRADUAÇÃO EM MODELAGEM DE SISTEMAS COMPLEXOS

FREDERICO ALEXANDRE DE SOUSA FRIAS

**Aprendizado supervisionado para modelagem da tendência do preço  
intradiário de ações brasileiras baseado em multifractalidade e causalidade**

São Paulo

2021

FREDERICO ALEXANDRE DE SOUSA FRIAS

**Aprendizado supervisionado para modelagem da tendência do preço  
intradiário de ações brasileiras baseado em multifractalidade e causalidade**

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Modelagem de Sistemas Complexos.

Área de concentração: Sistemas Complexos

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 22 de março de 2021. A versão original encontra-se em acervo reservado na Biblioteca da EACH-USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Orientador: Prof. Dr. Camilo Rodrigues Neto

São Paulo

2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

### CATALOGAÇÃO-NA-PUBLICAÇÃO

(Universidade de São Paulo. Escola de Artes, Ciências e Humanidades. Biblioteca)

CRB-8 4936

Frias, Frederico Alexandre de Sousa

Aprendizado supervisionado para modelagem da tendência do preço intradiário de ações brasileiras baseado em multifractalidade e causalidade / Frederico Alexandre de Sousa Frias ; orientador, Camilo Rodrigues Neto. – 2021.

119 f : il.

Dissertação (Mestrado em Ciências) - Programa de Pós-Graduação em Modelagem de Sistemas Complexos, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo.  
Versão corrigida

1. Aprendizado computacional. 2. Análise de séries temporais. 3. Sistemas dinâmicos. 4. Fractais. 5. Causalidade. 6. Algoritmos genéticos. 7. Ações. 8. Bolsa de valores. I. Rodrigues Neto, Camilo, orient. II. Título.

CDD 22.ed.– 006.31

Dissertação de autoria de Frederico Alexandre de Sousa Frias, sob o título “**Aprendizado supervisionado para modelagem da tendência do preço intradiário de ações brasileiras baseado em multifractalidade e causalidade**”, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Modelagem de Sistemas Complexos, na área de concentração Sistemas Complexos, aprovada em 22 de março de 2021 pela comissão julgadora constituída pelos doutores:

**Prof. Dr. Camilo Rodrigues Neto**

Escola de Artes, Ciências e Humanidades da Universidade de São Paulo  
Presidente

**Prof. Dr. Alexandre Ferreira Ramos**

Escola de Artes, Ciências e Humanidades da Universidade de São Paulo

**Prof. Dr. Birajara Soares Machado**

Instituto do Cérebro da Sociedade Beneficente Israelita Brasileira Hospital Albert Einstein

**Prof. Dr. Juan Carlos Ruilova Terán**

Escola de Economia de São Paulo da Fundação Getúlio Vargas

*Dedico este trabalho a Isabel e Diamantino.*

*“Kodoish, Kodoish, Kodoish, Adonai Tsebayoth”*

*(Mantra)*

## Resumo

FRIAS, Frederico Alexandre de Sousa. **Aprendizado supervisionado para modelagem da tendência do preço intradiário de ações brasileiras baseado em multifractalidade e causalidade**. 2021. 119 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2021.

Com a intensa atividade provocada pelos algoritmos de negociação, há cada vez mais informações referentes à evolução temporal da oferta e da demanda de uma ação. Os dados em alta frequência fornecem um panorama quantitativo para a criação de mercado e execução de operações. Compreender e modelar a dinâmica de um instrumento financeiro é de extrema importância para uma gestão mais eficiente de portfólio, de alocação de recursos bem como de risco. Este trabalho, portanto, tem como principal motivação propor um modelo híbrido não paramétrico capaz de caracterizar a tendência de queda ou de subida do preço de uma dada ação, utilizando, para tal, uma carteira de ativos integrantes do índice Bovespa com forte relação causal com o ativo modelado. A combinação do método de máquina de vetor de suporte com um algoritmo genético otimizou o classificador binário, produzindo resultados com uma acurácia superior em comparação com o algoritmo de aprendizado de máquina supervisionado puro. Além disso, a análise do comportamento histórico da estrutura fractal apresentada pela série temporal da ação alvo corroborou com os resultados obtidos na predição da dinâmica do movimento no período de tempo estudado.

Palavras-chaves: Sistemas complexos, Aprendizado de máquina, Análise de séries temporais, Sistemas dinâmicos, Fractalidade, Causalidade, Algoritmos genéticos, Mercado de ações.

## Abstract

FRIAS, Frederico Alexandre de Sousa. **Supervised learning for modeling the trend of Brazilian stocks intraday price based on multifractality and causality**. 2021. 119 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2021.

With the intense activity caused by the algorithmic trading, there is increasingly information regarding the temporal evolution of the supply and demand of a stock. High-frequency data provide a quantitative viewpoint for market creation and execution of operations. Understanding and modeling the dynamics of a financial instrument is extremely important for a more efficient management of a portfolio, of a resource allocation as well as risk. This work, therefore, has as its main motivation to propose a hybrid non-parametric model capable of characterizing the tendency of price to fall or rise for a given stock, using, for this purpose, a portfolio of assets that are part of the Bovespa's index with a strong causal relationship to the modeled asset. The combination of the support vector machine method with a genetic algorithm has optimized the binary classifier, producing results with higher accuracy than to them produced by the pure supervised machine learning algorithm. Furthermore, the analysis of the historical behavior of the fractal structure presented by the time series of the target stock corroborated with the results obtained in the prediction of the dynamics of the movement in the studied period.

Keywords: Complex systems, Machine learning, Time series analysis, Dynamical systems, Fractality, Causality, Genetic algorithms, Stock market.



## Lista de Figuras

Figura 1 – Valor de mercado e representatividade das ações integrantes do Ibovespa e do BOVA11 em novembro de 2018. . . . .	27
Figura 2 – Exemplos de fractais: (a) geométricos e (b) recursivos. . . . .	34
Figura 3 – Tipos de auto-semelhança: (a) auto-similar e (b) auto-afim. . . . .	35
Figura 4 – CYRE3: (a) $\ln(F_q(s))$ versus $\ln(s)$ ( $q = 2$ ), (b) espectro multifractal $f(\alpha)$ , (c) $\tau(q)$ versus $q$ e (d) $h(q)$ versus $q$ . . . . .	41
Figura 5 – EMBR3: (a) $\ln(F_q(s))$ versus $\ln(s)$ ( $q = 2$ ), (b) espectro multifractal $f(\alpha)$ , (c) $\tau(q)$ versus $q$ e (d) $h(q)$ versus $q$ . . . . .	42
Figura 6 – ITUB4: (a) $\ln(F_q(s))$ versus $\ln(s)$ ( $q = 2$ ), (b) espectro multifractal $f(\alpha)$ , (c) $\tau(q)$ versus $q$ e (d) $h(q)$ versus $q$ . . . . .	43
Figura 7 – PETR4: (a) $\ln(F_q(s))$ versus $\ln(s)$ ( $q = 2$ ), (b) espectro multifractal $f(\alpha)$ , (c) $\tau(q)$ versus $q$ e (d) $h(q)$ versus $q$ . . . . .	44
Figura 8 – Gráficos comparativos dos expoentes de Hurst das séries temporais originais e "embaralhadas" ( <i>shuffle</i> ) das ações alvo. . . . .	46
Figura 9 – Gráficos comparativos dos espectros multifractais $f(\alpha)$ das séries temporais originais e "embaralhadas" ( <i>shuffle</i> ) das ações alvo. . . . .	47
Figura 10 – Expoente de Hurst ( $q = 2$ ) entre o período de maio de 2018 a agosto de 2019 para cada uma das ações alvo. . . . .	49
Figura 11 – Atrator de Lorenz simulado com os seguintes parâmetros hidrodinâmicos: $\sigma = 10$ , $\rho = 28$ e $\beta = 8/3$ . . . . .	56
Figura 12 – Séries temporais obtidas pela projeção de cada uma das variáveis do sistema de equações de Lorenz. . . . .	56
Figura 13 – Reconstrução do atrator de Lorenz utilizando cada uma das variáveis do sistema de equações 27 e seus atrasos temporais ( $\tau = 6$ ). . . . .	58
Figura 14 – Exemplificação do mapeamento cruzado convergente entre as variedades-sombras $M_x$ e $M_y$ . . . . .	59
Figura 15 – Mapa de calor do coeficiente de Pearson do <i>log</i> -retorno, com frequência diária, das ações do Ibovespa entre janeiro a novembro de 2018. Destacado em vermelho estão as <i>small caps</i> e, em azul, as <i>blue chips</i> . O ETF BOVA11 encontra-se em negrito. . . . .	61

Figura 16 – Gráficos da dimensão de <i>embedding</i> (E), calculada pela projeção simplex, para cada ação alvo. O ponto preenchido denota o máximo da habilidade de predição ( $\rho$ ). . . . .	62
Figura 17 – Gráficos do decaimento do tempo de predição ( <i>tp</i> ), calculado pela projeção simplex, para cada ação alvo. O ponto preenchido denota o máximo da habilidade de predição ( $\rho$ ), sendo igual a 1 dia para todos os ativos. . . . .	63
Figura 18 – Gráficos do $\rho_{DCCA}$ e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo CYRE3. O eixo das abscissas está em dias. . . . .	64
Figura 19 – Gráficos do $\rho_{DCCA}$ e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo EMBR3. O eixo das abscissas está em dias. . . . .	66
Figura 20 – Gráficos do $\rho_{DCCA}$ e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo ITUB4. O eixo das abscissas está em dias. . . . .	67
Figura 21 – Gráficos do $\rho_{DCCA}$ e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo PETR4. O eixo das abscissas está em dias. . . . .	69
Figura 22 – Esquema simplificado do processo de indução de um classificador em um aprendizado supervisionado. . . . .	73
Figura 23 – Exemplificação do treinamento de uma classificação binária com: (a) subajustamento ( <i>underfitting</i> ); (b) ajustamento apropriado; (c) superajustamento ( <i>overfitting</i> ). . . . .	74
Figura 24 – Ilustração do princípio de minimização do risco estrutural (SRM). . . . .	75
Figura 25 – Ilustração do hiperplano ótimo e dos vetores de suporte numa classificação binária com a presença de erros. . . . .	81
Figura 26 – Exemplificação do SVM não linear: fronteira não linear no espaço de entradas (figura à esquerda); fronteira linear no espaço de características (figura à direita). . . . .	82
Figura 27 – Acurácia ao longo das cem mil gerações do algoritmo genético com população de tamanho 100, 10% de mutação e número de pais igual a 4 para cada ação alvo entre as datas 12 a 14 de novembro de 2018. . . . .	90

Figura 28 – Acurácia ao longo das cem mil gerações do algoritmo genético com população de tamanho 100, 10% de mutação e número de pais igual a 4 para cada ação alvo entre as datas 13 a 16 de novembro de 2018. . . . .	93
Figura 29 – Comparativo das matrizes de confusão do modelo híbrido (SVM+GA) com o modelo individual (SVM) para cada ação alvo para a data de 16 de novembro de 2018. . . . .	95
Figura 30 – Comparativo das matrizes de confusão do modelo híbrido (SVM+GA) com o modelo individual (SVM) para cada ação alvo para a data de 19 de novembro de 2018. . . . .	96
Figura 31 – Gráficos da autocorrelação dos <i>log</i> -retornos, com frequência diária, das ações alvo para diferentes atrasos ( $0 \leq k \leq 20$ ): (a) CYRE3, (b) EMBR3, (c) ITUB4 e (d) PETR4. . . . .	108
Figura 32 – Gráficos semi- <i>log</i> da função de densidade de probabilidade (PDF) dos <i>log</i> -retornos normalizados ( $Z$ ), com frequência diária, das ações alvo para diferentes intervalos de tempo ( $\Delta t$ 1 a 9 dias): (a) CYRE3, (b) EMBR3, (c) ITUB4 e (d) PETR4. . . . .	112
Figura 33 – Gráficos dos <i>log</i> -retornos, com frequência diária, das ações alvo comparados com um sinal de ruído branco gaussiano: (a) CYRE3, (b) EMBR3, (c) ITUB4 e (d) PETR4. . . . .	115
Figura 34 – Gráficos do efeito de alavancagem dos <i>log</i> -retornos, com frequência diária, das ações alvo ajustados por uma curva exponencial: (a) CYRE3, (b) EMBR3, (c) ITUB4 e (d) PETR4. . . . .	118

## Lista de Tabelas

Tabela 1 – Hierarquia de escalas de tempo (em segundos). . . . .	24
Tabela 2 – Ações integrantes do Ibovespa e do BOVA11 em novembro de 2018. . .	28
Tabela 3 – Resultados dos ajustes, para $q = 2$ , da regressão linear por mínimos quadrados de cada uma das ações alvo. . . . .	45
Tabela 4 – Expoentes $h(q)$ e $h_s(q)$ das séries temporais originais e "embaralhadas", respectivamente, das ações alvo <i>small caps</i> . . . . .	48
Tabela 5 – Expoentes $h(q)$ e $h_s(q)$ das séries temporais originais e "embaralhadas", respectivamente, das ações alvo <i>blue chips</i> . . . . .	48
Tabela 6 – Características da dependência da série temporal em relação ao expoente $h$ . . . . .	50
Tabela 7 – Valores do $\rho_{DCCA}$ e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo CYRE3. . . . .	65
Tabela 8 – Valores do $\rho_{DCCA}$ e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo EMBR3. . . . .	67
Tabela 9 – Valores do $\rho_{DCCA}$ e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo ITUB4. . . . .	68
Tabela 10 – Valores do $\rho_{DCCA}$ e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo PETR4. . . . .	69
Tabela 11 – Valores- $p$ de significância medidos pelo teste de fase aleatória para os ativos integrantes da carteira de cada ação alvo. . . . .	71
Tabela 12 – Funções de <i>kernel</i> mais comuns com seus respectivos parâmetros. . . .	83
Tabela 13 – Indicadores técnicos com suas respectivas classificações. . . . .	88
Tabela 14 – Indicadores técnicos selecionados pelo algoritmo genético para cada ação alvo entre as datas 12 a 14 de novembro de 2018. . . . .	91
Tabela 15 – Indicadores técnicos selecionados pelo algoritmo genético para cada ação alvo entre as datas 13 a 16 de novembro de 2018. . . . .	94
Tabela 16 – Comparativo das acurácias do modelo híbrido (SVM+GA) com o modelo individual (SVM) para cada ação alvo utilizando o valores preditos para as datas de 16 e 19 de novembro de 2018. . . . .	97

Tabela 17 – Resumo com os principais resultados obtidos em cada um dos métodos utilizados na modelagem da tendência do preço das ações alvo. . . . .	101
Tabela 18 – Valores de curtose dos <i>log</i> -retornos, com frequência diária, das ações alvo.	110
Tabela 19 – Valores de assimetria dos <i>log</i> -retornos, com frequência diária, das ações alvo. . . . .	111

## Lista de abreviaturas e siglas

BM&F	<i>Bolsa de Mercadorias e Futuros</i>
BOVESPA	<i>Bolsa de Valores de São Paulo</i>
CCM	<i>Convergent Cross Mapping</i>
CETIP	<i>Central de Custódia e de Liquidação Financeira de Títulos</i>
CVM	<i>Comissão de Valores Mobiliários</i>
DCCA	<i>Detrended Cross-Correlation Analysis</i>
DFA	<i>Detrended Fluctuation Analysis</i>
EMH	<i>Efficient Market Hypothesis</i>
ETF	<i>Exchange Traded Funds</i>
FMH	<i>Fractal Market Hypothesis</i>
KKT	<i>Karush-Kuhn-Tucker</i>
MFDFA	<i>Multifractal Detrended Fluctuation Analysis</i>
NASDAQ	<i>National Association of Securities Dealers Automated Quotations</i>
S&P 500	<i>Standard &amp; Poor's 500</i>
SRM	<i>Structural Risk Minimization</i>
SVM	<i>Support Vector Machine</i>

## Sumário

<b>1</b>	<b>Introdução</b>	16
1.1	<i>Mercado de ações no Brasil</i>	18
1.2	<i>Modelo quantitativo proposto</i>	20
1.3	<i>Dados empíricos</i>	23
<b>2</b>	<b>Hipótese sobre o mercado financeiro</b>	30
2.1	<i>Hipótese do mercado fractal</i>	31
2.2	<i>Breve descrição de fractalidade</i>	34
2.3	<i>Análise fractal e multifractal</i>	37
<b>3</b>	<b>Seleção de ativos para a construção do portfólio</b>	51
3.1	<i>Interdependência entre ativos</i>	51
3.1.1	Quantificando correlação cruzada	52
3.1.2	Busca por relação causal	54
3.2	<i>Composição da carteira correlacionada</i>	60
<b>4</b>	<b>Modelagem da tendência de séries temporais financeiras</b>	72
4.1	<i>Máquina de vetor de suporte</i>	72
4.1.1	Princípio de minimização do risco estrutural	74
4.1.2	Hiperplano ótimo	77
4.1.3	SVM não linear	81
4.2	<i>Redução de dimensionalidade com algoritmo genético</i>	83
4.3	<i>Predição da dinâmica do movimento</i>	87
<b>5</b>	<b>Conclusão</b>	99
5.1	<i>Considerações finais</i>	99
5.2	<i>Sugestões futuras</i>	101
	<b>Referências Bibliográficas</b>	102
	<b>Apêndice A – Fluxograma - Modelo Quantitativo</b>	106
	<b>Apêndice B – Fatos Estilizados</b>	107

<i>B.1</i>	<i>Propriedades do retorno</i> . . . . .	107
B.1.1	Ausência de autocorrelação (linear) . . . . .	107
B.1.2	Caudas gordas . . . . .	110
B.1.3	Assimetria ganho/perda . . . . .	110
B.1.4	Gaussianidade agregativa . . . . .	111
<i>B.2</i>	<i>Propriedades da volatilidade</i> . . . . .	114
B.2.1	Agrupamento de volatilidade / Intermitência . . . . .	114
B.2.2	Efeito de alavancagem . . . . .	117



## 1 Introdução

Nos últimos anos, as mudanças tecnológicas vem revolucionando a forma como os ativos financeiros são negociados. Desde sua criação, a Bolsa de Valores de São Paulo sofreu grandes mudanças na maneira como as transações são efetuadas no pregão, ambiente onde ocorre o encontro entre os compradores e os vendedores, possibilitando a realização do fechamento das operações. Por muito tempo, o pregão era feito em um espaço físico nas instalações da bolsa. Conhecido como pregão viva-voz, era um ambiente onde corretores se encontravam pessoalmente com o intuito de negociar com as contrapartes ativos e contratos cujas ordens de compra ou de venda eram recebidas por telefone. Nos anos de 2005 (BOVESPA) e de 2009 (BM&F), o antigo pregão foi substituído por um pregão eletrônico, onde as negociações deixaram de ser presenciais para serem registradas em sistemas eletrônicos por meio de terminais instalados nas corretoras de valores [1]. Esta mudança trouxe maior velocidade de execução das ordens e, conseqüentemente, maior volume de negócios que podem ser feitos diariamente.

Recentemente, o mercado de ações passou a ser automatizado. O chamado algoritmo de negociação, do inglês, *algorithmic trading* (AT), possibilitou que os investidores, por meio de plataformas eletrônicas, empreguem algoritmos no gerenciamento das ordens de compra ou de venda. Nesse sistema, um algoritmo executa instruções de negociação pré-programadas por tempo (data e hora), preço e quantidade em um único estoque, em pares de ações ou em cestas de estoques. Esta forma de operar vem sendo amplamente utilizada por instituições financeiras e tenciona dividir as negociações em partes menores de modo a reduzir o impacto sobre o mercado de forma otimizada bem como o risco.

Essas plataformas eletrônicas reúnem todas as ordens pendentes em um "livro" de ordens, chamado no jargão financeiro de *book* de ofertas. As ordens são executadas mecanicamente contra os melhores preços disponíveis naquele instante de tempo. Com a intensa atividade provocada por esses algoritmos, a frequência de envio de pedidos aumentou, diminuindo o tempo de execução das ordens. Com isso, há cada vez mais dados referentes à evolução temporal da oferta e da demanda bem como do comportamento dos preços das ações. A análise dessa informação acarreta um grande desafio por sua complexidade e por seu volume.

Os dados das ações em alta frequência fornecem um panorama quantitativo para a criação de mercado e execução de operações, além de possibilitar a modelagem estatística afim de prever o comportamento dos preços, do volume e das ofertas de negócios num curto espaço de tempo. Recentes eventos no mercado financeiro mostram que a gestão de riscos tem de levar em conta horizontes de tempo mais curtos do que normalmente considera. Em geral, o risco de operações financeiras é mensurado com janelas de tempo diárias ou superiores. No evento de 2010, conhecido como *Flash Crash*, os principais índices norte-americanos (S&P 500, Dow Jones e NASDAQ) caíram bruscamente e tornaram a subir ao mesmo patamar num intervalo de minutos. A dinâmica dos preços em escalas de tempo de alta frequência pode ser propagada para escalas de baixa frequência, levando a possíveis perturbações no mercado.

Análises estatísticas de dados de mercado são essenciais, tanto pela razão fundamental de entender a dinâmica de mercado quanto por razões relacionadas aos principais problemas de precificação e de administração de carteiras. Compreender e modelar a dinâmica de uma ação é de extrema importância para uma gestão mais eficiente de portfólio, de alocação de recursos financeiros bem como de risco. Entretanto, a modelagem do mercado de ações é uma tarefa desafiadora na previsão de séries temporais financeiras em virtude do alto grau de incerteza envolvida no movimento deste mercado, uma vez que são muitos os fatores interagentes, incluindo eventos políticos, condições econômicas gerais além das expectativas dos *traders* [2]. Desta forma, o mercado acionário pode ser considerado um sistema complexo adaptativo, sistema composto por uma grande quantidade de agentes heterogêneos conectados, interagindo entre si, com adaptação via evolução ou aprendizagem e com comportamento emergente, ou seja, a sua dinâmica global é resultante da ação das partes que o constituem e não imposta por um controle central [3]. A interação dos seus agentes (investidores, *traders*, instituições, entre outros) exhibe uma tendência não linear, produzindo uma série de preços e volumes negociados. Assim, torna-se crítico investigar e descobrir os padrões ocultos nos mercados de ações em uma estrutura não paramétrica.

### 1.1 Mercado de ações no Brasil

Atualmente, a bolsa de valores oficial do Brasil é a B3<sup>1</sup>. Sediada na cidade de São Paulo, é resultado da fusão, aprovada em março de 2017 pelo Conselho Administrativo de Defesa Econômica (CADE) e pela Comissão de Valores Mobiliários (CVM), entre a Bolsa de Valores, Mercadorias e Futuros de São Paulo (BM&FBOVESPA) com a Central de Custódia e de Liquidação Financeira de Títulos (CETIP). A BM&FBOVESPA (SA:BVMF3) era uma companhia de capital aberta formada pela integração das operações da antiga Bolsa de Valores de São Paulo (BOVESPA) e da Bolsa de Mercadorias e Futuros (BM&F). Já a CETIP (SA:CTIP3), diferentemente de uma bolsa, tinha como principal atividade a administração do mercado de balcão organizado, do inglês, *over-the-counter* (OTC).

A B3 é uma associação civil sem fins lucrativos, de capital aberto, com ação negociada em bolsa (SA:B3SA3), sendo considerada uma das principais empresas de infraestrutura de mercado financeiro no mundo. Suas atividades vão desde a elaboração e administração de sistemas de negociação, compensação, liquidação, depósito e registro de ativos financeiros até a atuação como contraparte central garantidora das operações realizadas [4]. Com a fusão, passou a atuar em dois segmentos: BM&FBOVESPA e CETIP UTVM.

No segmento BM&FBOVESPA, as negociações dos ativos são realizadas em uma plataforma única, o PUMA *Trading System*. Este segmento se originou da incorporação das operações da BOVESPA com as da BM&F. No primeiro, as transações são feitas das seguintes formas: à vista, a termo, no mercado futuro ou de opções. Entre os ativos negociados estão: ativos mercantis societários (ações), BDRs (do inglês, *Brazilian Depositary Receipts*), ETFs (fundos de índices), entre outros. Já no segundo, operam-se derivativos de mercadorias e futuros, tais como: derivativos agropecuários (açúcar, boi gordo, café arábica, etanol, soja) e derivativos financeiros (ouro, índices, taxa de câmbio, taxa de juros, títulos da dívida externa).

O segmento CETIP UTVM trouxe maior agilidade e flexibilidade para a negociação e o registro de títulos e valores mobiliários de renda fixa, com destaque para os títulos públicos, as debêntures, os CRAs (Certificados de Recebíveis Agrícolas) e os CRIs (Certificados de Recebíveis Imobiliários). É através deste segmento que a B3 administra o mercado

<sup>1</sup> Simbolizada por [B]<sup>3</sup> em alusão às letras iniciais de Brasil, Bolsa e Balcão.

de balcão organizado por meio da Plataforma Eletrônica, sistema integrado por dois subsistemas: Cetip | Trader e CetipNet.

Dentre os ativos financeiros de uma bolsa, temos os índices acionários. Estes são tidos como termômetros que indicam o desempenho de um dado conjunto de ações. Em suma, um índice de ações representa a evolução temporal da valorização de uma carteira teórica de papéis em um único indicador. O desempenho do índice é uma média ponderada, de acordo com os fatores definidos em sua regulamentação, do desempenho das suas ações constituintes.

Na B3, existem diversas famílias de índices de ações: índices amplos, índices de sustentabilidade, índices de governança, índices de segmentos e setoriais. Dos índices amplos, podemos destacar o índice Bovespa, mais conhecido como Ibovespa. Criado em janeiro de 1968, é considerado o principal índice do mercado de ações no Brasil, cujo objetivo é ser um indicador do desempenho médio das cotações das principais ações negociadas na B3. Para manter a representatividade do índice, a composição da sua carteira teórica é reavaliada quadrimestralmente de acordo com a metodologia de ponderação por liquidez, ou seja, quanto maior a liquidez da ação, maior será o seu peso no índice. A apuração do valor para cada ativo é feita pelo índice de negociabilidade, medida que relaciona o valor e a quantidade negociada de uma ação com o valor e a quantidade total das ações em circulação no mercado acionário (*free float*<sup>2</sup>). Desta forma, a determinação de quais ativos e em que quantidades serão integrantes da carteira teórica do Ibovespa tem como base a negociabilidade e a representatividade de cada ação no mercado acionário bem como outros critérios de seleção, descritos em maiores detalhes em [1]. Além disso, também é denominado um índice de retorno total uma vez que reflete o impacto que a distribuição de proventos por parte das companhias emissoras terá no retorno do mesmo.

Segundo dados de dezembro de 2019, o índice Bovespa tinha um valor de mercado, do inglês, *market cap*, de US\$ 1.01 trilhão. Composto por 65 ações de 62 empresas [5], representa cerca de 80% do número de negócios e do volume financeiro do mercado acionário. Uma ação é um valor mobiliário correspondente à menor parcela do capital social de uma empresa. Somente as ações das companhias que são registradas na CVM é que podem ser negociadas no mercado de valores mobiliários. A maioria das ações é escritural, ou seja, não necessita de emissão de certificado físico, são apenas um registro em

---

<sup>2</sup> Percentual das ações disponíveis para negociação, ou seja, que não pertencem ao controlador, nem aos administradores e que não estão em tesouraria.

uma conta de depósito. Elas concedem direitos e deveres de um sócio conforme a espécie da ação: ordinária (ON) ou preferencial (PN). A primeira confere ao seu detentor o direito a voto nas assembleias de acionistas. A segunda, prioridade na distribuição de dividendos ou no reembolso de capital, mas retira do titular o direito ao voto [6].

Apesar do índice ser um instrumento financeiro, não é possível ao investidor negociá-lo no mercado. Uma alternativa é adquirir um índice subjacente, mais conhecido como ETF (do inglês, *Exchange Traded Fund*), que é um fundo de investimento cujas cotas, semelhantes a uma ação, são negociadas na bolsa, pretendendo espelhar a rentabilidade e as variações de um índice de referência. Com isso, o investidor detem uma carteira com todas as ações constituintes do índice referenciado sem ter de comprá-las separadamente, proporcionando praticidade, rapidez no investimento e facilidade em acompanhar o seu desempenho. No caso do Ibovespa, o índice que o referencia é o BOVA11.

## 1.2 Modelo quantitativo proposto

Diversas investigações científicas contestam a hipótese do mercado eficiente que sugere que os movimentos do mercado acionário seguem um processo do tipo passeio aleatório. Alguns desses estudos apontam para presença de correlações de longo alcance em séries temporais financeiras, sugerindo que os ativos exibem uma tendência persistente [7–10]. Sendo assim, se as informações dos preços das ações forem pré-processadas de forma eficiente e forem aplicados algoritmos apropriados, as tendências e os padrões poderão ser identificados, sugerindo como a dinâmica da ação se comportará no futuro.

O campo de predição financeira é caracterizado por intensidade de dados, ruídos, natureza não estacionária, alto grau de incerteza e relações ocultas. A capacidade de prever a direção, de modo que o investidor possa tomar uma decisão de compra ou de venda, e não o valor exato do preço futuro de uma ação é o fator mais importante para ganhar dinheiro usando a predição financeira. Estudos mostraram que a predição do movimento em comparação ao valor pode gerar lucros mais elevados [2].

Investimentos quantitativos têm sido propostos para tal fim. Este tipo de investimento combina dados financeiros com matemática e tecnologia de computadores, sendo um método de gerenciamento de ativos amplamente utilizado na comunidade internacional de investimentos nos últimos dez anos [11]. A tecnologia de investimento quantitativo

cobre quase todo o processo de investimento, incluindo seleção quantitativa, arbitragem de futuros de índices de ações e de futuros de *commodities*, algoritmos de negociação, alocação de ativos e controle de riscos. Seu desempenho de investimento e estabilidade continuam a crescer e a obter cada vez mais o reconhecimento dos investidores.

Este trabalho, portanto, se insere no contexto destes investimentos quantitativos. Sua principal motivação está em buscar um modelo capaz de caracterizar a tendência do movimento do preço de uma ação. Na literatura existente, há dois segmentos: o da investigação empírica da dinâmica dos mercados de ações, inter-relações causais entre vários ativos financeiros, empregando métodos paramétricos convencionais e não paramétricos avançados; o de estruturas de modelagem de previsão dos movimentos futuros. Neste estudo, tentou-se combinar as duas diretrizes de pesquisa, apresentando um quadro integrado.

Inicialmente, foram escolhidas as ações alvo, instrumentos estes que se pretende modelar a tendência do movimento do preço. Dentre os ativos selecionados, dois métodos, o DFA e o MFDFA, foram empregados para caracterizar o espectro fractal e multifractal, respectivamente, das séries temporais financeiras. A análise fractal é uma ferramenta não paramétrica, sendo eficaz em compreender a natureza da dinâmica temporal de séries temporais não estacionárias e não lineares. Este tipo de técnica leva em conta a relação do tipo lei de potência, uma das propriedades de um sistema complexo. Sendo assim, entender a natureza da dinâmica do ativo financeiro é crucial antes de aplicar o algoritmo de modelagem preditiva.

O próximo passo consiste em selecionar as ações que serão usadas como referência na caracterização da dinâmica da ação alvo. Primeiramente, uma medida de correlação chamada de coeficiente DCCA de correlação cruzada ( $\rho_{DCCA}$ ) foi empregada de modo a capturar a dependência entre o ativo alvo e as demais ações constituintes do Ibovespa. Um segundo método, designado de mapeamento cruzado convergente, foi aplicado às ações com maior correlação, mensurada pela técnica anterior, em relação à modelada. Esta última ferramenta tem sido utilizada para capturar a relação de causa e efeito entre duas séries temporais não lineares.

Para os ativos que apresentaram forte dependência com a ação estudada, determinados na última etapa, foram calculados alguns indicadores técnicos. Estes indicadores serão os parâmetros que o algoritmo de previsão fará uso afim de classificar a tendência do preço da ação alvo. Entretanto, se faz necessário o emprego de uma técnica para selecionar estas características de modo que a classificação seja otimizada. Para tal fim, um algoritmo

genético foi utilizado em conjunto com o algoritmo preditivo. Os algoritmos genéticos são uma classe particular de algoritmos evolucionários que usam técnicas inspiradas na biologia evolutiva, como seleção, cruzamento, mutação e herança. Quando problemas apresentam um grande número de variáveis, os mesmos tornam-se quase que impossíveis de serem resolvidos analiticamente. Nestes casos, algoritmos genéticos podem ser utilizados para "evoluir" até alcançar a solução que melhor se aproxime da resposta desejada.

Por último, uma técnica de aprendizado de máquina supervisionada não linear foi empregada na modelagem do movimento do preço da ação alvo. A aprendizagem supervisionada usada foi a máquina de vetor de suporte. Este método é caracterizado pelo controle de capacidade da função de decisão, pelo uso das funções de *kernel* além de estar estabelecido no princípio de minimização do risco estrutural, sendo robusto a problemas de sobreajuste<sup>3</sup>, tendo um elevado desempenho de generalização. Outra propriedade importante é de que este algoritmo de aprendizado é equivalente a resolver um problema de programação quadrática linearmente restrito de modo que a solução seja sempre única e globalmente ideal, diferente do treinamento de redes neurais que requer otimização não linear com o risco de ficar preso em mínimos locais.

O trabalho está dividido da seguinte maneira: continuando neste capítulo, tem-se uma descrição dos dados empíricos utilizados neste estudo. O capítulo 2 aborda uma possível hipótese para descrever o mercado financeiro, além de evidenciar a análise fractal das séries temporais dos ativos alvo. No capítulo 3, as correlações, causais e não causais, são quantificadas entre as ações modeladas e as demais do Ibovespa. O algoritmo de predição proposto neste trabalho é apresentado no capítulo 4 que, em conjunto com alguns indicadores técnicos, é usado na caracterização da tendência do movimento do preço das ações alvo. Nesse mesmo capítulo, o algoritmo genético é empregado na otimização das variáveis usadas na predição. No capítulo 5, é feita uma discussão acerca dos resultados obtidos além de sugestões para futuros trabalhos. Por fim, o apêndice A detalha o fluxograma do modelo quantitativo proposto e o apêndice B evidencia a verificação de algumas das propriedades estatísticas, denominadas de fatos estilizados, com o intuito de refutar a hipótese de que o mercado acionário é aleatório o que, conseqüentemente, inviabilizaria a modelagem da tendência do preço de uma dada ação com o objetivo de obter lucros anormais.

---

<sup>3</sup> Do inglês, *overfitting*. Termo usado para descrever quando um modelo se ajusta muito bem ao conjunto de dados anteriormente observado, mas se mostra ineficaz para prever novos resultados.

### 1.3 Dados empíricos

No mercado de ações, as transações possuem aleatoriedade nas suas intensidades bem como no tempo em que se sucedem. As flutuações do preço de uma ação definem uma escala que é dada em unidades monetárias do país onde é negociada. Como esta escala varia no tempo, é de grande importância a determinação da variável mais apropriada a ser estudada.

Seja o valor de uma ação em um dado tempo  $t$  dado por  $P(t)$ . Na escala de preços, o incremento do valor do ativo, definido como  $r(t)$ , pode ser abordado como a variável aleatória de estudo. A representação da escala de preços pela variação absoluta, dada por

$$r_{abs}(t) = P(t + \Delta t) - P(t), \quad (1)$$

apresenta dois problemas: diferentes ações possuem preços em patamares distintos o que leva a variações absolutas desiguais em seus preços; variações do preço de uma dada ação em grandes escalas de tempo tendem a ser proporcionais ao preço.

De maneira alternativa, é possível apresentar a variação absoluta dos preços na forma deflacionada:

$$r_{abs}^{def}(t) = [P(t + \Delta t) - P(t)] \cdot D(t), \quad (2)$$

onde  $D(t)$  é o fator de deflação. Uma vez que a moeda de um país sofre mudanças em seu valor ao longo do tempo, sendo uma das causas a inflação, esta abordagem considera que o preço da ação é dado em termos de moeda constante. Contudo, a predição deste fator é difícil a longo prazo e, por existirem diferentes índices inflacionários, sua escolha não é única.

Em contrapartida à variação absoluta, a escala de preços pode ser analisada em termos de variação relativa:

$$r_{rel}^{simp}(t) = \frac{P(t + \Delta t) - P(t)}{P(t)}. \quad (3)$$

O incremento dado pela equação 3 mostra o percentual de ganho ou perda, dito retorno simples, independente do nível do preço da ação, ou seja, ativos distintos com variações absolutas de preço diferentes podem apresentar variações relativas similares. Entretanto, para longos horizontes de tempo, esta forma de representação de escala dos preços é sensível a variações de escala.



Ao analisar o retorno de uma dada ação utilizando o incremento dos logaritmos neperianos de seus preços em sucessivos intervalos de tempo, conhecido como retorno contínuo,

$$r_{rel}^{cont}(t) = \ln(P(t + \Delta t)) - \ln(P(t)), \quad (4)$$

não é necessário incorporar fatores de deflação ao modelo, se os mesmos forem constantes ao longo do tempo. Entretanto, esta não é uma boa aproximação dado que a taxa de crescimento da economia varia. Além disso, o retorno passa a ser não linear o que afeta as propriedades estatísticas de um processo estocástico.

Tabela 1 – Hierarquia de escalas de tempo (em segundos).

Regime	Escala de tempo
Ultra alta frequência (UHF)	$10^{-3}$ - 0.1
Alta frequência (HF)	1 - 100
Diário	$10^3$ - $10^4$

Fonte: R. Cont, 2011 [12]

Para dados intradiários (dados com alta ou ultra alta frequência, ver tabela 1), e para períodos de tempo suficientemente curtos de baixa inflação, os retornos das equações 1, 2, 3 e 4 são aproximadamente iguais:

$$r_{abs}(t) \approx r_{abs}^{def}(t) \approx r_{rel}^{simp}(t) \approx r_{rel}^{cont}(t). \quad (5)$$

Já a escala de tempo possui três possíveis formas de escala: o tempo físico, o tempo do mercado e o número de transações. Na primeira delas, existe a restrição imposta pelos períodos em que os mercados estão fechados (noite, finais de semana e feriados). Mesmo naqueles em que o mercado está ativo 24 horas por dia, há limitações temporais impostas pelos ciclos biológicos.

O tempo do mercado, período no qual o mercado de um dado ativo está aberto, é a escolha mais utilizada. Estudos empíricos mostraram que a variância dos retornos dos preços de fechamento entre dias consecutivos é cerca de 20% menor que a variância dos retornos considerando os valores de fechamento dos finais de semana [13]. Contudo, esta escala possui problemas como: informações afetarem o preço do ativo enquanto o mercado está fechado e a negociação do ativo ser considerada uniforme durante o tempo do mercado, sendo que a volatilidade é maior na abertura e no fechamento.

O tempo também poder ser medido em relação à quantidade de transações de um ativo, ou seja, transação-a-transação (do inglês, *tick-by-tick*). Neste caso, elimina-se uma

das fontes de aleatoriedade que é o tempo decorrido entre transações, mas a aleatoriedade do volume permanece.

As possíveis escalas mencionadas anteriormente, tanto de tempo como de preços, possuem suas vantagens e limitações. Na escala de preços, as análises realizadas neste trabalho consideraram o retorno contínuo ou *log* retorno, dado pela equação 4. As escalas de tempo foram definidas respeitando a frequência dos dados financeiros estudados, ou seja, transação-a-transação.

Entretanto, no que concerne a escala de tempo, é importante salientar que os dados intradiários das ações possuem diferentes frequências ao longo de suas séries temporais. Isto implica dizer que seus intervalos de tempo são irregulares (não constantes). A grande maioria dos métodos de análise de séries temporais só consegue lidar com dados regulares e não se estendem facilmente a dados de amostragem irregular. Estes métodos foram desenvolvidos quando as limitações de recursos computacionais favoreciam a análise de dados com intervalos iguais. Nestes casos, rotinas de álgebra linear eficientes podem ser empregadas e muitos desses problemas têm soluções explícitas.

Ao tratar de forma heurística esta questão, ignorando os horários de negociação, tratando-os como se fossem igualmente espaçados, pode-se introduzir um viés significativo, levando a predições incorretas. Um exemplo deste problema é considerar o retorno de uma ação que muda lentamente, mas de forma consistente, registrado com muita frequência no início e, em seguida, significativamente menos frequentemente. Neste caso, pareceria que o ativo mudou mais rapidamente, enquanto na verdade as propriedades estatísticas não se alteraram.

Talvez a abordagem mais utilizada seja transformar os dados irregulares em dados com intervalos regulares, empregando algum tipo de técnica de redimensionamento (por exemplo, interpolação) e, em seguida, aplicar os métodos existentes desenvolvidos para tais dados. Contudo, este procedimento possui desvantagens significativas, das quais citam-se algumas a seguir:

- em um movimento *browniano* padrão, a amostragem com interpolação linear reduz a distribuição condicional a um único valor determinístico, ou seja, ignora a estocasticidade em torno da média condicional. Sendo assim, estimativas de segundo momento como volatilidades, autocorrelações e covariâncias podem estar sujeitas a um viés significativo;

- redução ou diluição do conteúdo de informações de um conjunto de dados. No primeiro caso, dados são omitidos se observações consecutivas estiverem próximas, fazendo com que a inferência estatística seja menos eficiente. No segundo, dados redundantes são introduzidos se observações consecutivas estiverem distantes, enviesando as estimativas de significância estatística.

Evidentemente, muitas informações sobre a dinâmica do sistema são perdidas caso haja estrutura e irregularidade suficientes nos tempos de amostragem. Na literatura, existem poucos métodos de análise de dados de séries temporais com espaçamento desigual. Alguns autores sugeriram o uso de processos de difusão em tempo contínuo para resolver este problema. A ênfase tem sido principalmente na modelagem de processos de média móvel autoregressiva, do inglês, *autoregressive moving average* (ARMA). No entanto, estes modelos possuem limitações práticas [14].

Outro problema encontrado na literatura ao lidar com dados irregulares é o fenômeno conhecido como efeito Epps, do inglês, *Epps effect*. Reportado em 1979 em seu artigo [15], Epps mostrou que as correlações de retorno de ações diminuem a medida que a frequência de amostragem de dados aumenta. Os dois principais fatores que causam este efeito são: o possível efeito de *lead-lag*<sup>4</sup> entre os retornos das ações e a assincronissade das transações de diferentes ações [16]. O primeiro ocorre principalmente entre ativos com capitalização muito diferente. Neste caso, a correlação cruzada possui um máximo em um intervalo de tempo diferente de zero. Conforme a escala de tempo da amostra fica na mesma ordem de magnitude do retardo característico, as correlações cruzadas aumentam. No segundo fator, o efeito Epps foi reduzido ao considerar somente transações síncronas.

Isto posto, para evitar a influência dos efeitos encontrados por Epps em dados de alta frequência devidos, em sua maior parte, pela assincronicidade bem como o retardo temporal, apenas os valores de fechamento foram utilizados em ambos os métodos de mensuração da dependência entre os ativos bem como para caracterizar a fractalidade das séries temporais das ações alvo deste estudo. Uma vez que o algoritmo de aprendizado supervisionado bem como o algoritmo evolutivo não fazem nenhuma premissa acerca dos dados, é possível redimensionar as séries temporais de modo que fiquem síncronas e com igual espaçamento sem que, assim, comprometam a modelagem da dinâmica da ação. Desta forma, realizou-se a sincronização temporal entre as séries temporais dos ativos

---

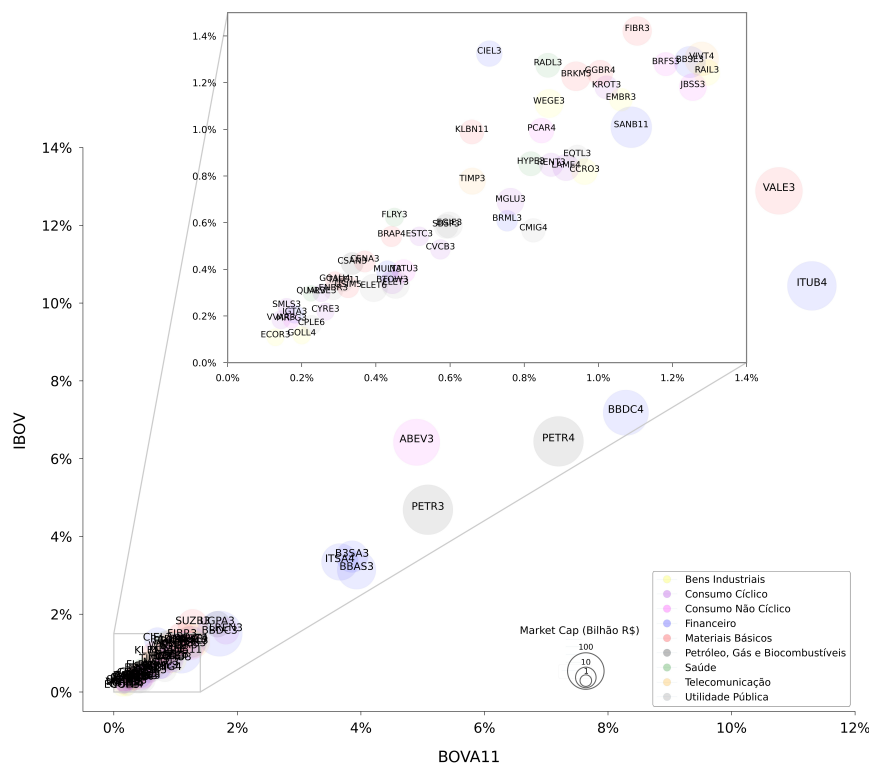
<sup>4</sup> Caso onde a variável principal é correlacionada com os valores de outra variável atrasada temporalmente.

usados como referência e a da ação alvo. A escala de tempo foi padronizada de modo que os intervalos consecutivos entre as amostras fossem constantes com uma frequência de cinco minutos entre si. No caso de não haver dado disponível, ocasionado pela não negociação do ativo naquele dado instante, considerou-se um valor igual ao último registro que o antecede, sem a realização de nenhum tipo de interpolação.

Os dados das ações são oriundos do *book* de ofertas da B3, refletindo cada uma das transações realizadas (*tick-by-tick*). Seus preços não estão ajustados a proventos (dividendos, entre outros). Para evitar as descontinuidades provocadas por estes tipos de eventos, a análise não foi realizada entre os pregões, somente nos dados de um mesmo dia. O período de tempo considerado na predição do movimento da ação alvo compreende os dias 12 a 19 de novembro de 2018.

A seguir, temos um gráfico com o valor de mercado e a representatividade de todas as ações, integrantes do Ibovespa bem como do BOVA11, que foram utilizadas neste trabalho. Os valores são de novembro de 2018, sendo fornecidos pela B3, no caso do índice, e pela *BlackRock*, no caso do ETF.

Figura 1 – Valor de mercado e representatividade das ações integrantes do Ibovespa e do BOVA11 em novembro de 2018.



Fonte: Frederico Frias, 2021

O detalhamento dos dados ilustrados na figura 1 encontra-se na tabela abaixo.

Tabela 2 – Ações integrantes do Ibovespa e do BOVALL em novembro de 2018.

Código	Nome	Tipo	Sector	Subsetor	Segmento	Valor de Mercado (MM R\$)	IBOV (Peso %)	BOVALL (Peso %)
ABEV3	AMBEV S.A.	ON	Consumo Não Cíclico	Bebidas	Cervejas e Refrigerantes	266.018,73	6,4211	4,0020
BSA3	B3 S.A. - BRASIL BOLSA BALCÃO	ON NM	Financeiro	Serviços Financeiros Diversos	Serviços Financeiros Diversos	58.067,71	3,4819	3,8560
BBAS3	ECO BRASIL S.A.	ON NM	Financeiro	Intermediários Financeiros	Bancos	127.883,56	3,1403	3,9300
BDDC3	BCO BRADESCO S.A.	ON NI	Financeiro	Intermediários Financeiros	Bancos	242.082,89	1,5007	1,7120
BDDC4	BCO BRADESCO S.A.	PN NI	Financeiro	Intermediários Financeiros	Bancos	242.082,89	7,1807	8,2890
BBSE3	BB SEGURIDADE PARTICIPAÇÕES S.A.	ON EDR NM	Financeiro	Previdência e Seguros	Seguradoras	56.640,00	1,2881	1,2470
BRAP4	BRADESPAR S.A.	PN NI	Materiais Básicos	Mineração	Minerais Metálicos	10.641,47	0,5407	0,4420
BRFS3	BRF S.A.	ON NM	Consumo Não Cíclico	Alimentos Processados	Carnes e Derivados	18.012,53	1,2793	1,1820
BRKM5	BRASKEM S.A.	PNA NI	Materiais Básicos	Químicos	Petroquímicos	42.024,90	1,2272	0,9400
BRML3	BR MALLS PARTICIPAÇÕES S.A.	ON NM	Financeiro	Exploração de Imóveis	Exploração de Imóveis	11.942,98	0,6079	0,7540
BTOW3	BTOW - COMPANHIA DIGITAL	ON NM	Consumo Cíclico	Comércio	Produtos Diversos	17.499,49	0,3449	0,4440
CCO3	CCB S.A.	ON NM	Bens Industriais	Transporte	Exploração de Rodovias	26.239,50	0,8182	0,9630
CIEB3	CIELO S.A.	ON NM	Financeiro	Serviços Financeiros Diversos	Serviços Financeiros Diversos	25.864,08	1,3243	0,7060
CMIG4	CIA ENERGÉTICA DE MINAS GERAIS - CEMIG	PN NI	Utilidade Pública	Energia Elétrica	Energia Elétrica	18.545,62	0,5668	0,8250
CPLE6	CIA PARANAENSE DE ENERGIA - COPEL	PNB NI	Utilidade Pública	Energia Elétrica	Energia Elétrica	8.747,13	0,1013	0,2260
CSAN3	COSAN S.A.	ON NM	Petróleo, Gás e Biocombustíveis	Petróleo, Gás e Biocombustíveis	Exploração, Refino e Distribuição	14.056,38	0,4247	0,3360
CSNA3	CIA SIDERURGICA NACIONAL	ON	Materiais Básicos	Siderurgia e Metalurgia	Siderurgia	12.307,34	0,4335	0,3700
CVGB3	CVC BRASIL OPERADORA E AGÊNCIA DE VIAGENS S.A.	ON NM	Consumo Cíclico	Viagens e Lazer	Viagens e Turismo	9.059,66	0,4852	0,5740
CYRE3	CYRELA BRAZIL REALTY S.A. EMPREEND E PART	ON NM	Consumo Cíclico	Construção Civil	Incorporações	5.692,34	0,2185	0,2040
ECOR3	ECORODOVIAS INFRAESTRUTURA E LOGÍSTICA S.A.	ON NM	Bens Industriais	Transporte	Exploração de Rodovias	5.195,90	0,1084	0,0980
EGE3	ENGE BRASIL ENERGIA S.A.	ON NM	Utilidade Pública	Energia Elétrica	Energia Elétrica	28.074,44	0,5893	0,5980
ELET3	CENTRAIS ELET BRAS S.A. - ELETROBRÁS	ON NI	Utilidade Pública	Energia Elétrica	Energia Elétrica	34.078,22	0,3342	0,4530
ELET6	CENTRAIS ELET BRAS S.A. - ELETROBRÁS	PNB NI	Utilidade Pública	Energia Elétrica	Energia Elétrica	34.078,22	0,3204	0,3940
EMR3	EMBRER S.A.	ON NM	Bens Industriais	Material de Transporte	Material Aeronáutico e de Defesa	16.142,14	1,1267	1,0610
ENR3	EDP - ENERGIAS DO BRASIL S.A.	ON NM	Utilidade Pública	Energia Elétrica	Energia Elétrica	8.307,78	0,3103	0,2850
EQTL3	EQUATORIAL ENERGIA S.A.	ON NM	Utilidade Pública	Energia Elétrica	Energia Elétrica	14.438,63	0,8850	0,9430
ESTC3	ESTÁCIO PARTICIPAÇÕES S.A.	ON NM	Consumo Cíclico	Diversos	Serviços Educacionais	7.649,95	0,5412	0,5170
FIBR3	FIBRIA CELULOSE S.A.	ON NM	Materiais Básicos	Madeira e Papel	Madeira e Celulose	40.559,09	1,4211	1,1050
FLRY3	FLURY S.A.	ON NM	Saúde	Serviços Médicos Hospitalares, Análises e Diagnósticos	Serviços Médicos Hospitalares, Análises e Diagnósticos	6.822,21	0,6232	0,4500
GBR4	GERDAU S.A.	PN NI	Materiais Básicos	Siderurgia e Metalurgia	Siderurgia	24.721,61	1,2417	1,0050
GOAU4	METALURGICA GERDAU S.A.	PN NI	Materiais Básicos	Siderurgia e Metalurgia	Siderurgia	7.153,41	0,3502	0,2890
GOLL4	GOL LINHAS AÉREAS INTELIGENTES S.A.	PN N2	Bens Industriais	Transporte	Transporte Aéreo	5.766,21	0,1169	0,2000
HYPE3	HYPERA S.A.	ON NM	Saúde	Comércio e Distribuição	Medicamentos e Outros Produtos	20.294,84	0,8522	0,8180
IGTA3	IGUATEMI EMPRESA DE SHOPPING CENTERS S.A.	ON NM	Financeiro	Exploração de Imóveis	Exploração de Imóveis	6.910,81	0,2081	0,1810
ITSA4	ITAUSA INVESTIMENTOS ITAÚ S.A.	PN NI	Financeiro	Intermediários Financeiros	Bancos	105.117,34	3,3436	3,6600
ITUB4	ITAÚ UNIBANCO HOLDING S.A.	PN NI	Financeiro	Intermediários Financeiros	Bancos	327.500,18	10,4427	11,3010
JBS3	JBS S.A.	ON NM	Consumo Não Cíclico	Alimentos Processados	Carnes e Derivados	32.117,36	1,1808	1,2550
KLBN11	KLBIN S.A.	UNT ED N2	Materiais Básicos	Madeira e Papel	Papel e Celulose	20.155,38	0,9886	0,6590
KROT3	KROTON EDUCACIONAL S.A.	ON NM	Consumo Cíclico	Diversos	Serviços Educacionais	17.264,61	1,1797	1,0220
LAMB4	LOJAS AMERICANAS S.A.	PN NI	Consumo Cíclico	Comércio	Produtos Diversos	28.735,78	0,8363	0,9130
LEN3	LOJAS RENNEN S.A.	ON NM	Consumo Cíclico	Comércio	Têxtil, Vestuário e Calçados	28.152,96	1,5672	1,8130
MGLU3	MAGAZINE LUIZA S.A.	ON NM	Consumo Cíclico	Comércio	Eletrodomésticos	31.877,07	0,6898	0,7630
MRF3	MARFRIG GLOBAL FOODS S.A.	ON NM	Consumo Não Cíclico	Alimentos Processados	Carnes e Derivados	3.851,93	0,1804	0,1720
MRVE3	MRV ENGENHARIA E PARTICIPAÇÕES S.A.	ON NM	Consumo Cíclico	Construção Civil	Incorporações	5.387,41	0,2974	0,2630
MULT3	MULTIPLAN - EMPREEND IMOBILIÁRIOS S.A.	ON N2	Financeiro	Exploração de Imóveis	Exploração de Imóveis	13.117,96	0,3904	0,4310
NATU3	NATURA COSMÉTICOS S.A.	ON NM	Consumo Não Cíclico	Produtos de Uso Pessoal e de Limpeza	Produtos de Uso Pessoal	17.642,00	0,3016	0,4750
PCAR4	CIA BRASILEIRA DE DISTRIBUIÇÃO	PN NI	Consumo Não Cíclico	Comércio e Distribuição	Alimentos	22.414,93	0,9943	0,8470
PETR3	PETRÓLEO BRASILEIRO S.A. PETROBRÁS	ON N2	Petróleo, Gás e Biocombustíveis	Petróleo, Gás e Biocombustíveis	Exploração, Refino e Distribuição	351.984,24	4,6862	5,0850
PETR4	PETRÓLEO BRASILEIRO S.A. PETROBRÁS	PN N2	Petróleo, Gás e Biocombustíveis	Petróleo, Gás e Biocombustíveis	Exploração, Refino e Distribuição	351.984,24	6,4443	7,1900
QUAL3	QUALICORP CONSULTORIA E CORRETORA DE SEGUROS S.A.	ON NM	Saúde	Serviços Médicos Hospitalares, Análises e Diagnósticos	Serviços Médicos Hospitalares, Análises e Diagnósticos	3.924,83	0,2970	0,2250
RADL3	RAIA DROGASIL S.A.	ON NM	Saúde	Comércio e Distribuição	Medicamentos e Outros Produtos	20.526,88	1,2743	0,8640
RALL3	RUMO S.A.	ON NM	Bens Industriais	Transporte	Transporte Ferroviário	27.189,24	1,2433	1,2930
RENT3	LOCALIZA RENT A CAR S.A.	ON NM	Consumo Cíclico	Diversos	Aluguel de carros	17.952,99	0,8478	0,8730
SANB11	BCO SANTANDER (BRASIL) S.A.	UNT	Financeiro	Intermediários Financeiros	Bancos	161.036,51	1,0088	1,0890
SBSF3	CIA SANEAMENTO BÁSICO EST SÃO PAULO	ON NM	Utilidade Pública	Água e Saneamento	Água e Saneamento	19.295,48	0,5838	0,5900
SMLS3	SMILES FIDELIDADE S.A.	ON NM	Consumo Cíclico	Diversos	Programas de Fidelização	5.481,15	0,2387	0,1590
SUZB3	SUZANO PAPEL E CELULOSE S.A.	ON NM	Materiais Básicos	Madeira e Papel	Papel e Celulose	45.217,23	1,7439	1,2770
TABE11	TRANSMISSORA ALIANÇA DE ENERGIA ELÉTRICA S.A.	UNT N2	Utilidade Pública	Energia Elétrica	Energia Elétrica	7.744,34	0,3442	0,3130
TIMP3	TIM PARTICIPAÇÕES S.A.	ON NM	Telecomunicação	Telecomunicação	Telecomunicação	28.713,45	0,7777	0,6600
UGPA3	ULTRAPAR PARTICIPAÇÕES S.A.	ON NM	Petróleo, Gás e Biocombustíveis	Petróleo, Gás e Biocombustíveis	Exploração, Refino e Distribuição	26.290,14	1,7435	1,6750
USIM5	USINAS SID DE MINAS GERAIS S.A. USIMINAS	PNA NI	Materiais Básicos	Siderurgia e Metalurgia	Siderurgia	13.514,56	0,3245	0,3250
VALE3	VALE S.A.	ON NM	Materiais Básicos	Mineração	Minerais Metálicos	279.020,27	12,8862	10,7880
VIVT4	TELEFÔNICA BRASIL S.A.	PN	Telecomunicação	Telecomunicação	Telecomunicação	76.230,19	1,3017	1,2790
VVAR3	VIA VAREJO S.A.	ON NM	Consumo Cíclico	Comércio	Eletrodomésticos	6.703,99	0,1834	0,1440
WEGE3	WEG S.A.	ON NM	Bens Industriais	Máquinas e Equipamentos	Motores, Compressores e Outros	37.398,10	1,1104	0,8670
<b>Total</b>						<b>2.977.960,37</b>	<b>100,00</b>	<b>98,84</b>

No próximo capítulo, discute-se uma hipótese sobre o mercado financeiro. Além disso, é feita uma análise fractal nas ações alvo. Foram escolhidas duas *blue chips* e duas *small caps*, termos utilizados no mercado financeiro para designar, respectivamente, ações de empresas mais bem cotadas na bolsa de valores e de empresas com baixa capitalização. O ativo é considerado *blue chip* quando é de uma empresa de grande porte com comprovada lucratividade e poucas obrigações. São instrumentos ditos de "primeira linha", ou seja, com maior liquidez de mercado uma vez que têm uma alta procura pelos investidores.

Neste caso, as duas ações analisadas foram a ITUB4, ação do setor financeiro, e a PETR4, de petróleo, gás e biocombustíveis. Estas ações estão entre as quatro de maior representatividade no índice Bovespa, que juntas totalizam um peso de 16.89%, e com os maiores valores de mercado, dois critérios utilizados em conjunto na seleção (ver figura 1).

As *small caps* escolhidas foram a CYRE3 e a EMBR3. As duas são integrantes do índice SMLL, cujo objetivo é ser o indicador do desempenho médio das cotações dos ativos de uma carteira composta pelas empresas de menor capitalização. A primeira é de uma incorporadora e a segunda, uma empresa do ramo da aeronáutica. Pela tabela 2, é possível observar que ambas representam juntas um pouco mais de 1.3% do Ibovespa.

## 2 Hipótese sobre o mercado financeiro

Os mercados financeiros são sistemas complexos onde investidores interagem entre si e lidam com as informações afim de determinar o melhor valor de um instrumento financeiro. O preço de uma ação é o resultado de diferentes conjuntos e origens de informação. Cada investidor tem sua preferência por um dado ativo bem como possui uma maneira particular de o analisar. O risco e retorno mensurados são uma estimativa que difere de um analista para outro. Além do mais, as predições acerca do valor deste ativo sofrerão alterações a medida em que o analista receber novos insumos a respeito do mesmo.

Durante algum tempo, existiu a crença de que os mercados financeiros obedeciam à hipótese do mercado eficiente, do inglês, *efficient market hypothesis* (EMH). Esta hipótese, originalmente formulada na década de 60 por Fama [17], defende que os mercados são extremamente eficientes em relação à informação, ou seja, toda a informação é instantaneamente processada, refletindo imediatamente no preço do ativo negociado.

A hipótese do mercado eficiente está associada à idéia de que o preço de um ativo no mercado especulativo é descrito por um processo estocástico dito passeio aleatório. Este processo é uma martingale, sequência de variáveis aleatórias no qual as informações de eventos passados não ajudam a prever eventos futuros, importando apenas o evento atual. Neste modelo, a esperança do próximo valor na sequência é igual ao último valor observado, independentemente do conhecimento do histórico passado. Conseqüentemente, os preços são independentes no tempo.

Em sua forma estrita, um mercado eficiente é um sistema idealizado, onde só é possível obter ganhos financeiros havendo arbitragem, lucro auferido pela discrepância de preços de um dado ativo. Com essa premissa, o investidor não conseguiria obter retornos superiores à média do mercado, dado um determinado nível de risco. Sendo assim, um mercado eficiente é aquele onde o preço de uma ação sempre reflete todas as informações disponíveis e é igual ao valor justo de mercado da empresa. Além disso, nenhum investidor teria condições de prever o futuro de maneira acurada, de forma a bater o mercado sistematicamente.

A hipótese do mercado eficiente foi aceita e utilizada em muitos modelos econômicos. Entretanto, diversos estudos empíricos realizados em mais de meio século comprovam

a existência de propriedades estatísticas que contradizem esta hipótese, evidenciando a existência dos chamados fatos estilizados [13, 18, 19].

Analisando os dados empíricos dos ativos financeiros de um ponto de vista estatístico, é possível notar o compartilhamento de algumas propriedades não triviais entre os mesmos. Elas são comuns entre uma ampla variedade de instrumentos financeiros, mercados e períodos de tempo. Tais observações confirmam a compreensão de que os preços não são independentes e, por isso, não deveriam ser modelados por um passeio aleatório. Estes fatos estilizados são considerados em termos das propriedades qualitativas do retorno e da volatilidade.

Algumas destas propriedades foram analisadas nas séries temporais dos ativos alvo aqui estudados. A descrição de cada uma delas bem como os seus respectivos resultados encontram-se no apêndice B deste trabalho.

### 2.1 *Hipótese do mercado fractal*

A hipótese do mercado eficiente foi bem sucedida em levar o universo matemático para o ambiente de mercado de capitais. Entretanto, dados empíricos analisados ao longo de uma dezena de anos contestam esta hipótese, atentando para a presença de propriedades estatísticas, denominadas de fatos estilizados, comuns a diversos instrumentos financeiros. Como visto no apêndice B, algumas destas propriedades foram comprovadas ao examinar os dados de ações, com frequência diária, negociadas no mercado brasileiro. Fatos como o comportamento da distribuição de frequência negativamente assimétrica bem como a constatação de caudas pesadas na distribuição dos retornos implicam que os dados financeiros não são normalmente distribuídos. Como consequência disso, a concepção de que os preços das ações podem ser descritos por um passeio aleatório também fica enfraquecida.

Em virtude da hipótese do mercado eficiente não estar de acordo com o que é observado nos dados financeiros reais, uma hipótese de mercado alternativa foi proposta por Peters [20]. Chamada de hipótese do mercado fractal, do inglês, *fractal market hypothesis* (FMH), veio no seguimento de suas críticas anteriores à EMH [21], fazendo uso da teoria do caos e dos objetos fractais de modo a fornecer uma nova estrutura para uma modelagem mais exata da turbulência, descontinuidade e não-periodicidade que



caracterizam os mercados financeiros atuais. Peters não restringiu sua análise com nenhum requisito estatístico uma vez que o propósito desta hipótese era o de expor um modelo de comportamento dos investidores bem como dos movimentos dos preços dos ativos que fossem compatíveis com os dados empíricos.

A hipótese do mercado fractal tem como principal alicerce a heterogeneidade dos investidores, nomeadamente no que concerne aos seus horizontes de investimento, isto é, o prazo que pretendem manter o capital aplicado. De acordo com a mesma, o mercado financeiro é composto por investidores com horizontes de investimento distintos que variam suas análises de informação mediante seus horizontes de tempo individuais. Participantes com horizontes de curto prazo (segundos, minutos), como os *market makers*<sup>1</sup> ou *noise-traders*<sup>2</sup>, focam na informação técnica e no comportamento de grupos de outros investidores do mercado, enquanto os de horizontes de longo prazo (vários anos), como os fundos de pensão, tomam suas decisões em bases fundamentalistas, sem se preocupar com o comportamento da multidão [22]. Uma dada informação pode acarretar oportunidades distintas entre os investidores com diferentes horizontes de investimento. Por exemplo, a mesma pode ser um sinal de compra/venda para o investidor de curto/longo prazo e vice-versa.

A coexistência dos investidores com os mais variados perfis de horizonte de tempo é o que garante a estabilidade do mercado. Quando um desses investidores começa a predominar face ao outro, o horizonte de investimento deste passa a ser dominante no mercado, levando a uma crise nos preços uma vez que as ordens de compra ou de venda do horizonte predominante não serão atendidas com uma ordem contrária dos restantes horizontes. Portanto, o mercado se torna instável quando há hegemonia de um dos horizontes de investimento. O inverso, segundo esta hipótese, também é válido. Ou seja, quando existe uma ampla gama de horizontes de investimento atuando no mercado, o mesmo funciona de forma estável.

Neste ponto, a hipótese do mercado fractal se diferencia da hipótese do mercado eficiente já que a última se baseia em modelos de equilíbrio, funcionando bem em períodos de calma no mercado e inútil durante instabilidades, pelo modelo não saber lidar com a transição para uma fase turbulenta. Ao contrário da EMH, a FMH pressupõe que a

<sup>1</sup> Do português, "criador" de mercado. Pessoa jurídica, cadastrada na bolsa de valores, que se compromete a manter regularmente e continuamente ofertas de compra e venda afim de fomentar a liquidez no mercado.

<sup>2</sup> Participante do mercado que compra ou vende um ativo mobiliário de forma irracional e errática.

instabilidade no mercado tem como causa a existência de um horizonte de investimento característico, implicando na perda de sua estrutura fractal e, conseqüentemente, provocando grandes e improváveis eventos como crises no mercado de ações. Sendo assim, a hipótese do mercado fractal sugere que todos os horizontes de tempo são representados nos períodos de estabilidade para que a oferta e a demanda sejam suavemente equilibradas.

Esta hipótese também enfatiza o impacto da informação no comportamento dos investidores. Na ótica do mercado eficiente, a informação assim como o investidor são tratados genericamente. Um investidor é qualquer um que seja um tomador de preços racional, que saiba avaliar as informações disponíveis no mercado no momento atual, maximizando o seu retorno. Esta abordagem tem como consequência a de que todos os tipos de informação impactam homogeneamente todos os investidores.

Segundo Peters [20], os mercados existem para prover um ambiente de estabilidade e liquidez que favoreça as negociações dos agentes. Além disso, os investidores almejam conseguir um ativo por um bom preço que não é obrigatoriamente o preço justo no sentido econômico, como ocorre nas operações intradiárias em geral.

Em suma, a hipótese do mercado fractal tem como principais suposições [23]:

- a liquidez está disponível quando o mercado possui investidores com os mais diversos horizontes de investimento. A mesma é a principal razão da estabilidade do mercado (equilíbrio entre oferta e demanda);
- a informação tem um impacto distinto nos diferentes horizontes de investimento;
- os preços são reflexo da combinação de negociações técnicas de curto prazo e avaliações fundamentalistas de longo prazo.

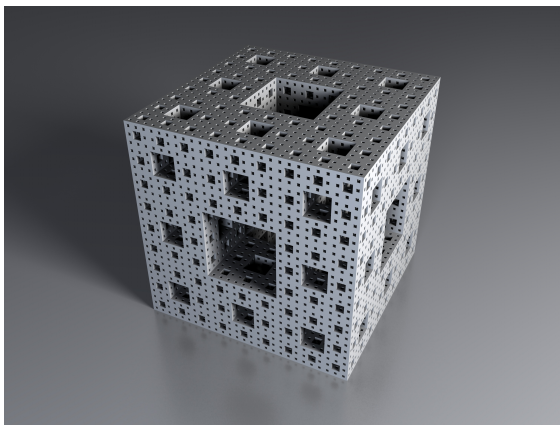
Esta hipótese tem grande valia ao introduzir a noção de fractalidade e de dependência de longo alcance nas séries temporais financeiras, com sua crença de que o mercado é formado por processos estocásticos de curto prazo com processos determinísticos de longo prazo. Nesta perspectiva, torna-se viável o emprego da análise fractal com o intuito de distinguir séries temporais aleatórias de séries temporais fractais. A seguir, será apresentado, de maneira sucinta, o conceito de fractal.

## 2.2 Breve descrição de fractalidade

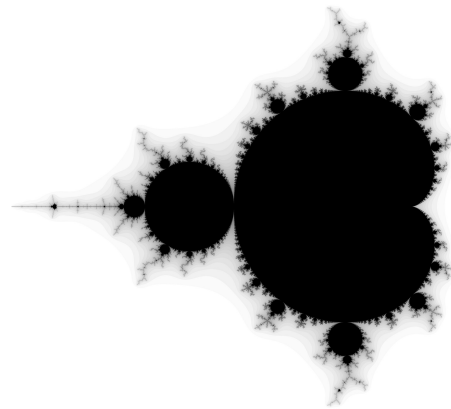
Fractal é uma palavra derivada do latim do adjetivo *fractus*, que significa "fração", "fragmento" ou "irregular". Foi cunhada por Benoit Mandelbrot em 1975 em seu trabalho [24]. Embora tenha nomeado os fractais, Benoit não foi o primeiro a estudá-los. Matemáticos, como Cantor<sup>3</sup>, Koch<sup>4</sup>, Hilbert<sup>5</sup>, entre outros, investigavam objetos que desafiavam as noções comuns de infinito, para os quais não existia uma explicação objetiva. Os chamados "monstros matemáticos" colocavam em prova algumas das bases matemáticas relacionadas à análise, à álgebra e à geometria daquela época, entre o final do século XIX e início do século XX. Esses objetos foram criados com o intuito de descrever as reais formas da natureza que, ao contrário da geometria euclidiana, não podem ser representadas por formas geométricas simples, como quadrados ou circunferências.

Em suma, um fractal é um objeto com invariância geométrica quando observado em escalas diferentes, onde suas partes se assemelham ao todo. São objetos que podem ser obtidos geometricamente, como a esponja de Menger (figura 2a), ou aleatoriamente, por processos recursivos (figura 2b), possuindo propriedades específicas que os caracterizam, sendo as principais delas: auto-semelhança, complexidade, dimensão e escala.

Figura 2 – Exemplos de fractais: (a) geométricos e (b) recursivos.



(a) Esponja de Menger



(b) Conjunto de Mandelbrot

Um objeto auto-semelhante é aquele que possui o mesmo aspecto visual independentemente da escala observada. Existem dois tipos de auto-semelhança: a exata (ou determinística) e a aproximada (ou estatística). A auto-semelhança exata só existe em

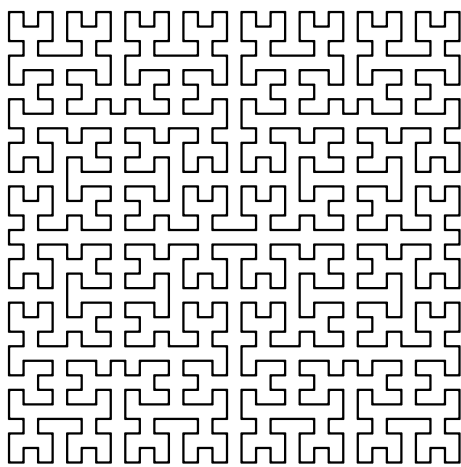
<sup>3</sup> Georg Cantor (1845-1918), matemático russo.

<sup>4</sup> Helge von Koch (1870-1924), matemático sueco.

<sup>5</sup> David Hilbert (1862-1943), matemático alemão.

processos matemáticos, sendo gerada pela iteração da mesma regra de construção indefinidamente. Também chamada de auto-similaridade, é uma propriedade de simetria com invariância por transformação geométrica isotrópica, isto é, a alteração do comprimento de escala, sob o qual o fractal é analisado, leva a sucessivas réplicas idênticas dele mesmo independentemente da direção considerada. A curva de Hilbert (figura 3a), construída por um processo recursivo, é um exemplo de fractal auto-similar.

Figura 3 – Tipos de auto-semelhança: (a) auto-similar e (b) auto-afim.



(a) Curva de Hilbert



(b) Couve-flor

O fractal com auto-semelhança aproximada é aquele constituído por cópias anisotrópicas, onde a transformação geométrica de sua escala leva a partes que não mantêm fixas as proporções originais do mesmo. Entretanto, as cópias possuem a mesma estrutura ou uma distribuição estatística idêntica. Neste caso, os fractais são chamados de auto-afins. Exemplos deste tipo de fractal estão presentes em abundância na natureza, entre eles está a couve-flor (figura 3b).

Outra característica dos fractais é a complexidade. Por serem gerados de maneira contínua por um processo iterativo, sucessivas transformações de escala destes objetos levam, indefinidamente, a cópias que são semelhantes ao todo. Desta forma, nunca será possível obter uma "figura final" do mesmo, sendo considerados objetos de complexidade infinita.

Esta propriedade pode ser observada ao analisar a curva de Koch, objeto fractal construído por um processo iterativo no qual um segmento de reta é dividido em três segmentos iguais, onde o terço médio é substituído por um triângulo equilátero sem base. Esta regra é aplicada em cada um dos segmentos de reta resultantes. A cada iteração,

três segmentos de reta são substituídos por quatro de igual comprimento. Desta forma, a cada passo sucessivo, o comprimento total é multiplicado por  $4/3$ . A sequência numérica produzida por este processo recursivo é uma progressão geométrica de razão  $4/3$ , cujo limite tende a infinito. Isto implica que a figura resultante deste processo iterativo terá comprimento infinito. Este limite foi definido como "limite interno" por Mandelbrot.

A curva de Koch, se observada em um plano bidimensional, "ocupa" um espaço maior que uma reta. Pela geometria euclidiana, geometria baseada nos postulados de Euclides descritos em sua famosa obra chamada *Os Elementos*, a dimensão é um número inteiro positivo que expressa a quantidade de coordenadas necessárias para caracterizar uma forma. Por exemplo, uma reta tem dimensão 1, pois precisa de uma coordenada para representá-la. Um plano, que possui comprimento e largura, necessita de duas coordenadas, portanto, é um objeto bidimensional na geometria euclidiana (dimensão 2). Sendo assim, a curva de Koch possui uma dimensão maior que 1 e menor que 2, uma vez que esta curva não "ocupa" totalmente a faixa de espaço que a contém. Esta constatação evidencia que a geometria euclidiana é insuficiente e até mesmo grosseira para explicar e descrever os fractais, derrubando a crença de Euclides de que todas as formas da natureza podiam ser representadas por sua geometria.

A noção de dimensão fractal possibilita descrever contornos irregulares ou superfícies rugosas. Esta dimensão está relacionada com a estrutura de ocupação espacial de uma forma. Sua noção introduz o conceito de dimensão fracionária (não inteira), viabilizando a existência de objetos com dimensões desde zero a infinito.

A existência de formas irregulares na natureza foi alvo de estudo de Mandelbrot em seu trabalho sobre a extensão territorial do litoral da Grã-Bretanha [25]. Ele foi além das dimensões inteiras, abordando a questão do ponto de vista de dimensão fracionária, mostrando que esta poderia ser traduzida como o grau de fragmentação da forma. Além disso, ele descobriu que esse grau de irregularidade permanecia constante qualquer que fosse a escala utilizada, ou seja, os padrões de forma continuavam inalterados. Isto implica dizer que o sistema não possui uma escala característica, podendo ser descrito por uma lei de potência. Também chamada de lei de escala, ganhou considerável relevância na mecânica estatística pela constatação de que diversas grandezas termodinâmicas, tais como a compressibilidade e susceptibilidade magnética, possuem caráter universal na região crítica, caracterizado pelos mesmos expoentes críticos. Com isto, surgiu o conceito de

classe de universalidade, no qual sistemas com os mesmos expoentes pertencem à mesma classe.

A lei de potência pode ser descrita pela seguinte relação matemática:

$$f(\lambda x) = \lambda^\alpha \cdot f(x) = \lambda^\alpha \cdot A \cdot x^\alpha, \quad (6)$$

onde  $x$  é a variável,  $A$  é a constante de normalização,  $\alpha$  é o expoente de escala e  $\lambda$  é uma constante que simboliza a variação de escala. Pela equação 6, é possível notar que a mudança de escala pode ser incorporada na constante de normalização, o que implica que sistemas caracterizados por esta lei são livres de escala.

### 2.3 Análise fractal e multifractal

De acordo com a hipótese proposta por Peters, os investidores não reagem instantaneamente às informações que recebem nem as interpretam da mesma forma, sendo possível compreender o motivo de estruturas estatísticas auto-semelhantes existirem no mercado financeiro e como o risco é compartilhado entre os investidores.

Como visto anteriormente, Mandelbrot introduziu o modelo fractal para descrever uma determinada classe de objetos que exibem um comportamento complexo. Em [26], ele aplicou o conceito de fractalidade em séries temporais financeiras. Usar a noção de fractal em um objeto significa analisá-lo em diferentes escalas e resoluções, inter-relacionando os resultados. No caso de séries temporais financeiras, um estudo estatístico do ponto de vista fractal é baseado na análise de intervalos de tempo de diferentes tamanhos. Uma série temporal possui uma característica fractal quando em um tempo específico  $t$  a mesma se comporta de maneira singular, ou seja, existe um componente com potência não inteira, caracterizado por um degrau ou cume na série. Quando esta série possui comportamento escalar com um único expoente de escala, a série é dita monofractal. Porém, há séries temporais caracterizadas pela presença de diferentes expoentes de correlação observados em subconjuntos distintos da série. Nestes casos, a multiplicidade de expoentes de escala é o que define uma série temporal multifractal.

Existem alguns métodos para a análise multifractal de séries não estacionárias, entre eles o *wavelet transform modulus maxima* (WTMM). Este modelo, proposto por Muzy [27],

utiliza a transformada de wavelet contínua<sup>6</sup> afim de detectar singularidades (descontinuidades) na série. Contudo, o mesmo é de implementação complexa e computacionalmente dispendioso. Uma solução alternativa mais simples a este método é o *multifractal detrended fluctuation analysis* (MFDFA), apresentado por Kantelhardt em [28]. Esta metodologia é uma generalização do método *detrended fluctuation analysis* (DFA), técnica que tem sido amplamente utilizada em diversos campos da ciência [29–33], sendo inicialmente proposta em [34] ao tentar constatar correlações de longo alcance em sequências de DNA. Os resultados produzidos pelo método MFDFA tendem a ser mais confiáveis do que os obtidos por WTMM, com vantagem para valores negativos de  $q$  (ordem da função de flutuação) e para séries temporais curtas.

Seja uma série temporal de preços  $\{P_t\}_{t=1}^T$ . Esta série será tratada do ponto de vista dos seus retornos, por motivos explicados anteriormente neste trabalho. Sendo assim, a equação 4 deve ser aplicada à série de preços, transformando  $P_t$  na série de *log*-retornos  $\{r_t\}_{t=1}^{T-1}$ .

O método MFDFA consiste em cinco passos.

- Passo 1: calcular a série  $y(k)$ :

$$y(k) = \sum_{i=1}^k [r(i) - \bar{r}], \quad k = 1, 2, \dots, T - 1. \quad (7)$$

Segundo [28], a subtração da média  $\bar{r}$  na equação 7 não é mandatária, desde que esta seja excluída no terceiro passo, onde é feito o tratamento das tendências locais.

- Passo 2: dividir  $y(k)$  em  $N$  segmentos não sobrepostos de tamanho  $s$ , onde  $N = \text{int}(\frac{T-1}{s})$ . Cada segmento é denominado de  $\nu$ , onde  $\nu = 1, 2, \dots, N$ . Nos casos em que  $N$  não for múltiplo de  $s$ , uma parte final da série temporal será desconsiderada. Para que isto não ocorra, deve-se repetir este processo a partir do extremo oposto, originando  $2N$  segmentos.
- Passo 3: calcular a função de flutuação para cada segmento  $\nu$ :

$$\mathcal{F}^2(\nu, s) = \frac{1}{s} \sum_{i=1}^s \left[ y([\nu - i]s + i) - y_\nu^m(i) \right]^2, \quad \nu = 1, 2, \dots, N, \quad (8)$$

$$\mathcal{F}^2(\nu, s) = \frac{1}{s} \sum_{i=1}^s \left[ y([T - 1] - [\nu - N]s + i) - y_\nu^m(i) \right]^2, \quad \nu = N + 1, \dots, 2N, \quad (9)$$

<sup>6</sup> Do português, onduleta. Método muito utilizado em processamento de sinais, capaz de decompor e representar uma série temporal em um novo domínio, como o espaço, possibilitando a análise em diferentes escalas de frequência e de tempo.

onde  $y_\nu^m(i)$  é um ajuste polinomial de ordem  $m$  em cada segmento  $\nu$ . Podem ser usados polinômios lineares ( $m = 1$ ), quadráticos ( $m = 2$ ), cúbicos ( $m = 3$ ) ou de ordem superior, obtendo respectivamente os métodos *DFA1*, *DFA2*, *DFA3*, *DFA $m$* . Este ajuste elimina tendências locais de ordem  $m$  da série  $y(k)$  (ou, equivalentemente, à ordem  $m - 1$  da série temporal original).

- Passo 4: calcular a função de flutuação de ordem  $q$ :

$$\mathcal{F}_q(s) = \left( \frac{1}{2N} \sum_{\nu=1}^{2N} [\mathcal{F}^2(\nu, s)]^{q/2} \right)^{\frac{1}{q}}. \quad (10)$$

A variável  $q$  pode assumir qualquer valor real exceto zero. O método DFA é definido quando  $q = 2$ . No caso em que  $q = 0$ , a função  $\mathcal{F}_q(s)$  não pode ser determinada diretamente pela equação 10 em razão do seu expoente divergente. Em vez disso, uma média logarítmica tem de ser utilizada para se obter  $\mathcal{F}_0(s)$ .

$$\mathcal{F}_0(s) = \exp\left(\frac{1}{4N} \sum_{\nu=1}^{2N} \ln[\mathcal{F}^2(\nu, s)]\right). \quad (11)$$

Por construção,  $\mathcal{F}_q(s)$  só é definido para  $s \geq m + 2$ .

- Passo 5: repetir os passos 2 a 4 utilizando diferentes janelas de tempo  $s$  para cada ordem  $q$ , de modo a determinar o comportamento escalar da função de flutuação. Se a série temporal analisada tiver correlações de longo alcance, a medida em que a escala de tempo  $s$  aumentar, a função  $\mathcal{F}_q(s)$  irá aumentar conforme uma lei de potência:

$$\mathcal{F}_q(s) \sim s^{h(q)}. \quad (12)$$

O expoente  $h(q)$  é conhecido como o expoente de Hurst generalizado uma vez que o mesmo, para  $q = 2$ , é idêntico ao expoente  $H$  de Hurst no caso de séries temporais estacionárias. O expoente de Hurst é utilizado na análise *R/S*, proposta por Henry Hurst [35], afim de detectar correlações de longo alcance em séries temporais e é determinado ao se comparar o comportamento escalar dos desvios cumulativos reescalados pela média com a hipótese nula de que a série temporal é aleatória. Este expoente retrata as propriedades de correlação da série temporal, sendo interpretado do seguinte modo:

- $0 < h(q) < 0.5 \rightarrow$  série antipersistente com correlação negativa;
- $h(q) = 0.5 \rightarrow$  série sem correlação de longo alcance;
- $0.5 < h(q) < 1 \rightarrow$  série persistente com correlação positiva.



Ao analisar o gráfico *log-log* de  $\mathcal{F}_q(s)$  *versus*  $q$ , é possível notar a relação linear de  $\mathcal{F}_q(s)$  com a ordem  $q$ , cujo coeficiente angular é dado por  $h(q)$ . Contudo, a função  $\mathcal{F}_q(s)$  apresenta comportamento não confiável para escalas de tempo  $s$  relativamente grandes em comparação com o tamanho da série temporal estudada uma vez que o número de segmentos  $N$  utilizados no passo 4 se torna muito pequeno. Neste caso, estas janelas de tempo não são consideradas no cálculo da função de flutuação. Já para tamanhos de segmentos muito pequenos ( $s \approx 10$ ), o expoente  $h(q)$ , na maioria das vezes, pode não depender linearmente de  $q$  [28].

O parâmetro  $h(q)$  é independente de  $q$  no caso de séries temporais monofractais, ou seja, a equação 12 terá o mesmo comportamento escalar para todos os valores de  $q$ . Entretanto, para valores positivos de  $q$ , os segmentos  $\nu$  com grandes desvios em relação ao ajuste polinomial exercerão maior influência sobre  $\mathcal{F}_q(s)$ . Já para valores negativos de  $q$ , os segmentos  $\nu$  com menor variância irão dominar a função  $\mathcal{F}_q(s)$ . Em outras palavras,  $h(q)$  será dependente de  $q$  nestas duas situações.

Nas séries multifractais, em geral, grandes flutuações produzem expoentes de escala menores do que pequenas flutuações neste tipo de série. Isto ocorre, pois para escalas de tempo muito pequenas ( $s \ll T - 1$ ), os segmentos com maior/menor flutuação irão ter maior peso na função  $\mathcal{F}_q(s)$  se  $q$  for maior/menor do que zero. Portanto,  $\mathcal{F}_q(s)$  ( $q < 0$ ) será menor do que  $\mathcal{F}_q(s)$  ( $q > 0$ ) [36].

Utilizando a analogia da multifractalidade com a termodinâmica, é possível relacionar analiticamente o expoente de escala multifractal  $h(q)$  com o expoente de escala  $\tau(q)$ :

$$\tau(q) = qh(q) - 1. \quad (13)$$

O parâmetro  $\tau(q)$  é o equivalente à energia livre, energia total disponível de um sistema termodinâmico, definido a partir da função de partição  $Z(q)$ , função que descreve as propriedades estatísticas, como entropia, de um sistema em equilíbrio termodinâmico.

Outra forma de caracterizar o comportamento multifractal de séries temporais é através do espectro de singularidade  $f(\alpha)$ , relacionado com  $h(q)$  via transformada de Legendre:

$$\alpha(q) = \frac{d\tau(q)}{dq} \quad (14)$$

e

$$f(\alpha(q)) = q\alpha(q) - \tau(q), \quad (15)$$

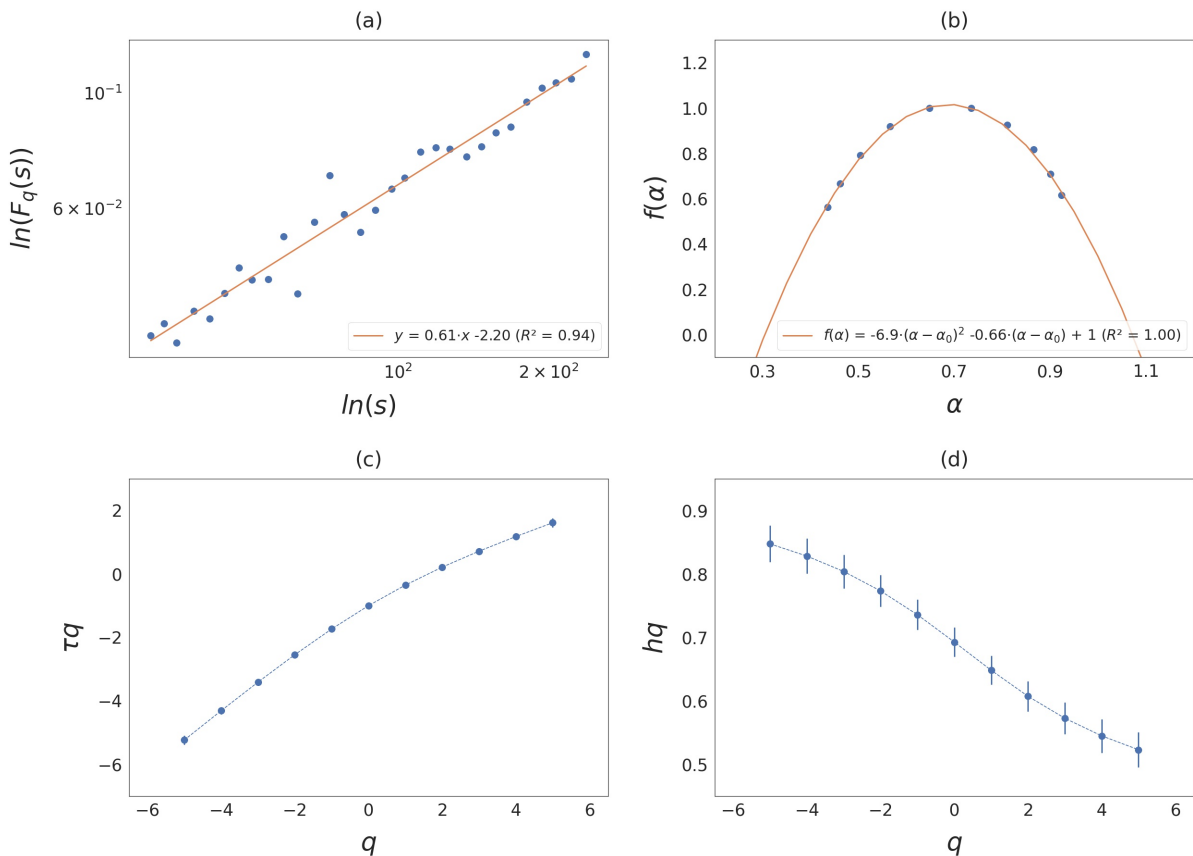
onde  $\alpha$  é chamado de expoente de *Hölder*, análogo à energia, e  $f(\alpha)$  denota a dimensão do subconjunto da série que é caracterizada por  $\alpha$ , correlato à entropia de Shannon. A partir destas variáveis, pode ser determinada a dimensão multifractal generalizada:

$$D(q) = \frac{\tau(q)}{q-1} = \frac{qh(q) - 1}{q-1}, \quad \forall q \neq 1. \quad (16)$$

O gráfico de  $D(q)$  *versus*  $q$  representa a "força" da multifractalidade, ou seja, quanto mais constante for a relação entre  $D(q)$  e  $q$ , mais fraca será a multifractalidade da série temporal [27]. Além disso, para valores extremos de  $q$ , o método MFDFA torna-se menos acurado. Para selecionar um intervalo apropriado de valores de  $q$ , é necessário garantir que o espectro de singularidade seja maior que zero ( $f(\alpha(q)) > 0$ ) e que a dimensão multifractal generalizada seja menor ou igual à dimensão topológica  $d$  ( $D(q) \leq d$ ).

Para cada um dos ativos estudados, foram calculadas as funções de flutuação  $\mathcal{F}_q(s)$ , para  $q = 2$ , o espectro multifractal  $f(\alpha)$  além dos expoentes  $\tau(q)$  e  $h(q)$  em função de  $q$ .

Figura 4 – CYRE3: (a)  $\ln(F_q(s))$  *versus*  $\ln(s)$  ( $q = 2$ ), (b) espectro multifractal  $f(\alpha)$ , (c)  $\tau(q)$  *versus*  $q$  e (d)  $h(q)$  *versus*  $q$ .

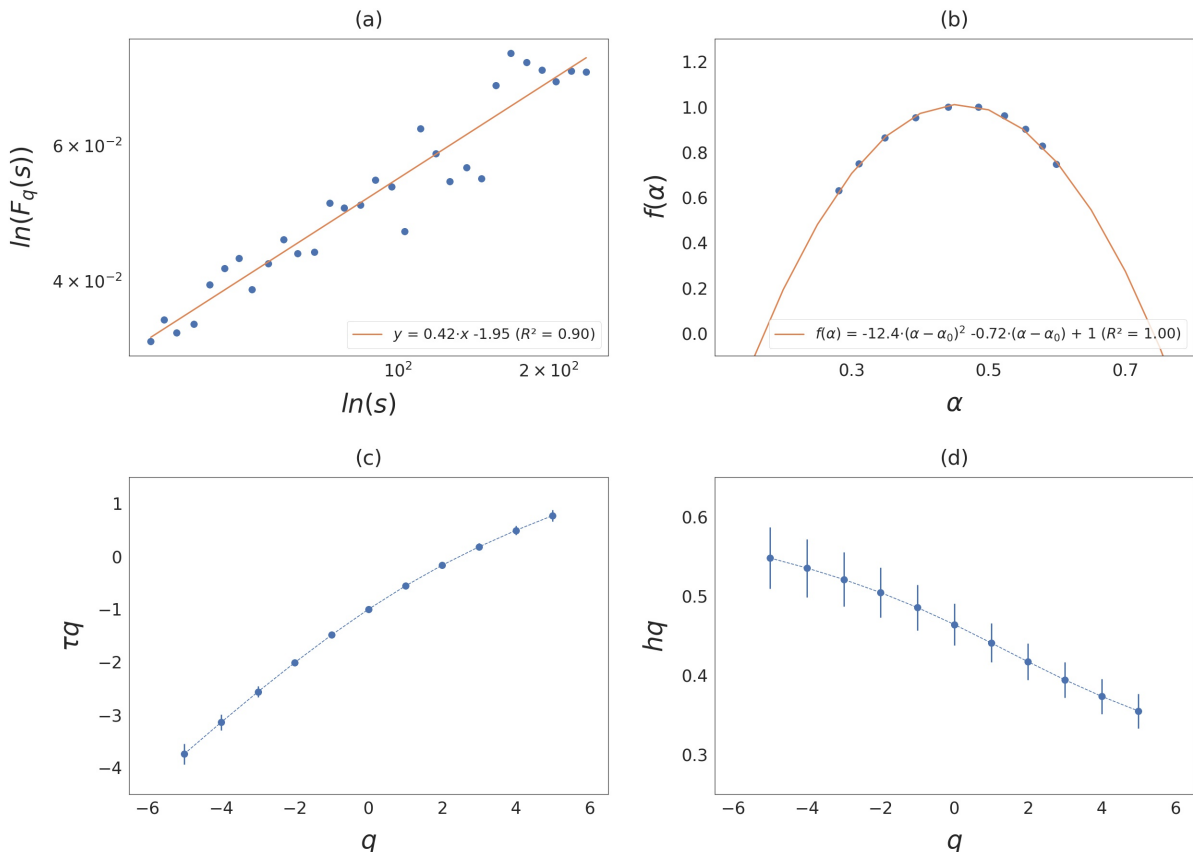


Analisando a figura 4d, pode-se observar que os valores de  $h(q)$  diminuem com o aumento de  $q$ , evidenciando a escala multifractal da ação CYRE3. Os espectros podem ser caracterizados quantitativamente pelo ajuste de uma função quadrática, utilizando o método dos mínimos quadrados, na vizinhança do máximo  $\alpha_0$ , como pode ser visto na figura 4b. O ajuste é do tipo:

$$f(\alpha) = A \cdot (\alpha - \alpha_0)^2 + B \cdot (\alpha - \alpha_0) + C, \quad (17)$$

onde  $C$  é uma constante dada por  $f(\alpha_0) = 1$  e  $B$  é uma medida de assimetria do espectro, assumindo valor igual a zero para um espectro perfeitamente simétrico [31].

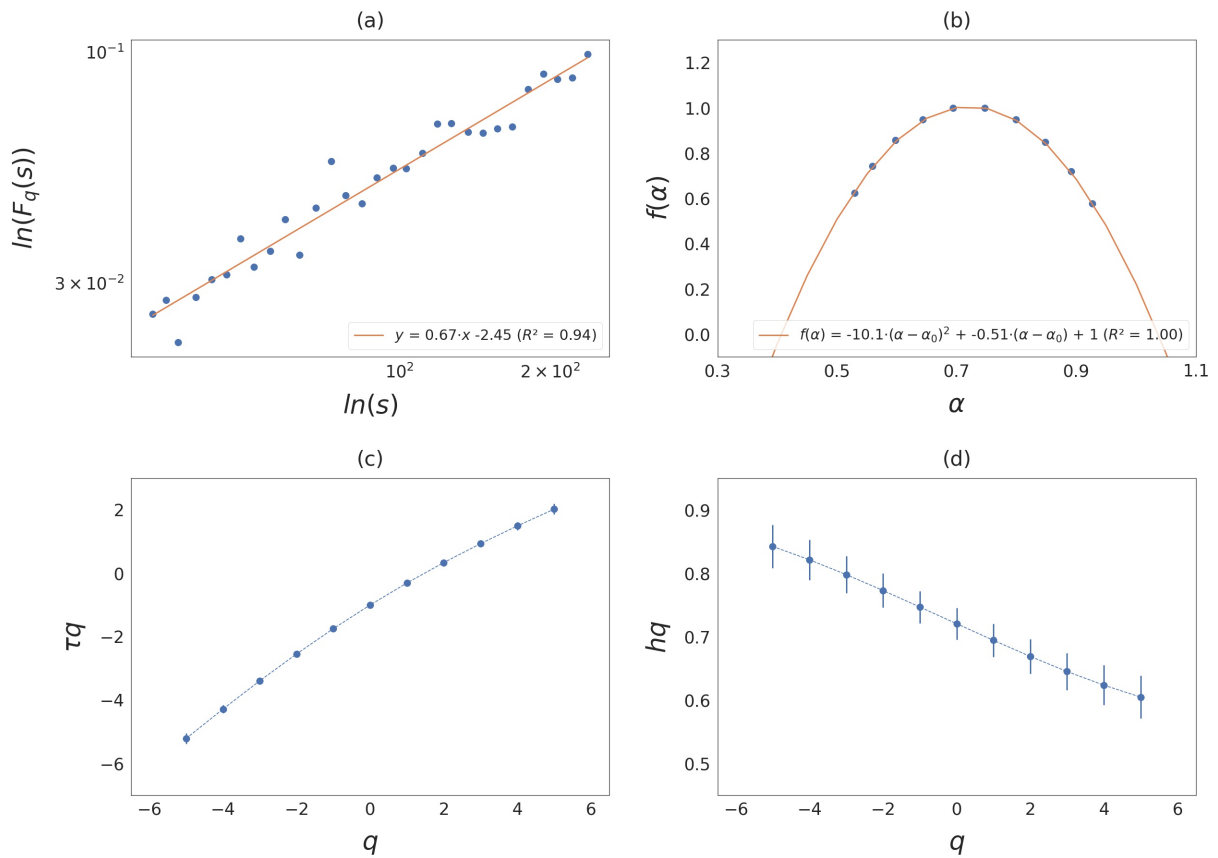
Figura 5 – EMBR3: (a)  $\ln(F_q(s))$  versus  $\ln(s)$  ( $q = 2$ ), (b) espectro multifractal  $f(\alpha)$ , (c)  $\tau(q)$  versus  $q$  e (d)  $h(q)$  versus  $q$ .



Fonte: Frederico Frias, 2021

A largura do espectro multifractal denota essencialmente a faixa de expoentes. A mesma pode ser obtida pela extrapolação da curva quadrática ajustada para zero, isto é, definida por  $\alpha_1 - \alpha_2$ , com  $f(\alpha_1) = f(\alpha_2) = 0$ . Quanto maior o valor da largura maior será a multifractalidade do espectro, sendo igual a zero para uma série temporal monofractal.

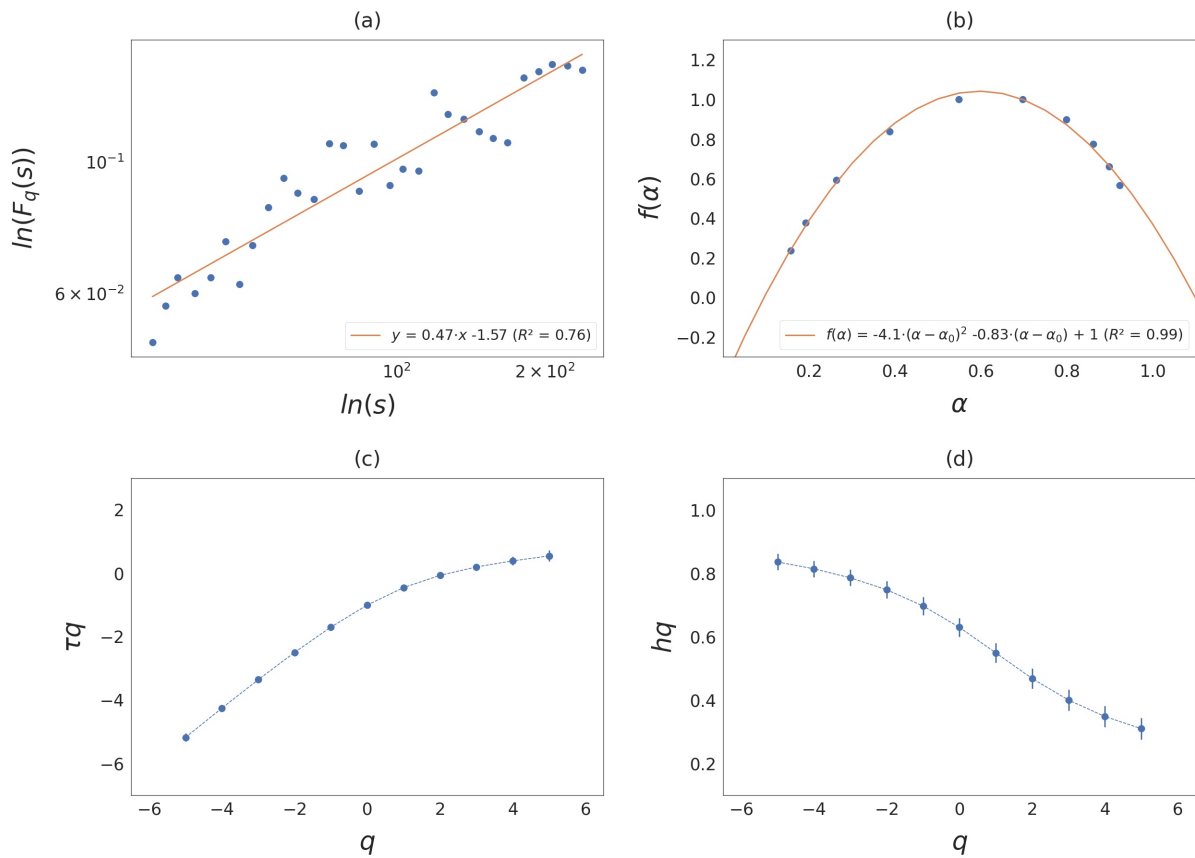
Figura 6 – ITUB4: (a)  $\ln(F_q(s))$  versus  $\ln(s)$  ( $q = 2$ ), (b) espectro multifractal  $f(\alpha)$ , (c)  $\tau(q)$  versus  $q$  e (d)  $h(q)$  versus  $q$ .



Fonte: Frederico Frias, 2021

O gráfico *log-log* de  $\mathcal{F}_q(s)$  versus  $q$  pode ser analisado na figura 4a. A reta em vermelho é o ajuste da regressão linear, cujo coeficiente angular é o expoente de Hurst  $h(q)$  para  $q = 2$ . Nota-se que a série temporal é persistente, uma vez que  $h(q) > 0.5$ . Na figura 4c, temos o expoente de escala  $\tau(q)$ . O mesmo possui um caráter não linear, caracterizado em séries temporais multifractais. Se a série fosse monofractal, o gráfico de  $\tau(q)$  versus  $q$  seria configurado por uma reta.

Figura 7 – PETR4: (a)  $\ln(F_q(s))$  versus  $\ln(s)$  ( $q = 2$ ), (b) espectro multifractal  $f(\alpha)$ , (c)  $\tau(q)$  versus  $q$  e (d)  $h(q)$  versus  $q$ .



Fonte: Frederico Frias, 2021

Analisando as figuras 5, 6 e 7, também pode-se notar o caráter multifractal das séries temporais das demais ações alvo. É possível observar os valores dos ajustes das regressões lineares de cada um dos ativos modelados na tabela 3. O coeficiente angular é o próprio expoente de Hurst  $h(q)$ , como já mencionado anteriormente. Os resultados obtidos pelo método de mínimos quadrados estão bem ajustados aos dados empíricos, dado que os resíduos estão dispersos aleatoriamente em torno de zero e com variância constante. A confirmação de que o modelo está bem ajustado é evidenciada pelos valores de  $R^2$ <sup>7</sup> e de  $R^2$  ajustado<sup>8</sup>.

<sup>7</sup> É uma medida de ajuste de um modelo estatístico aos valores observados de uma variável aleatória. Varia entre 0 e 1 e expressa a quantidade da variância dos dados que é explicada pelo modelo. Quanto maior, mais explicativo é o mesmo.

<sup>8</sup> Versão modificada do  $R^2$  que penaliza a inclusão de regressores pouco explicativos.

Tabela 3 – Resultados dos ajustes, para  $q = 2$ , da regressão linear por mínimos quadrados de cada uma das ações alvo.

Ação	$h(q)$	$R^2$	$R^2$ ajustado
CYRE3	$0.607 \pm 0.024$	0.959	0.957
EMBR3	$0.417 \pm 0.023$	0.920	0.918
ITUB4	$0.669 \pm 0.027$	0.955	0.953
PETRA	$0.468 \pm 0.032$	0.883	0.879

Fonte: Frederico Frias, 2021

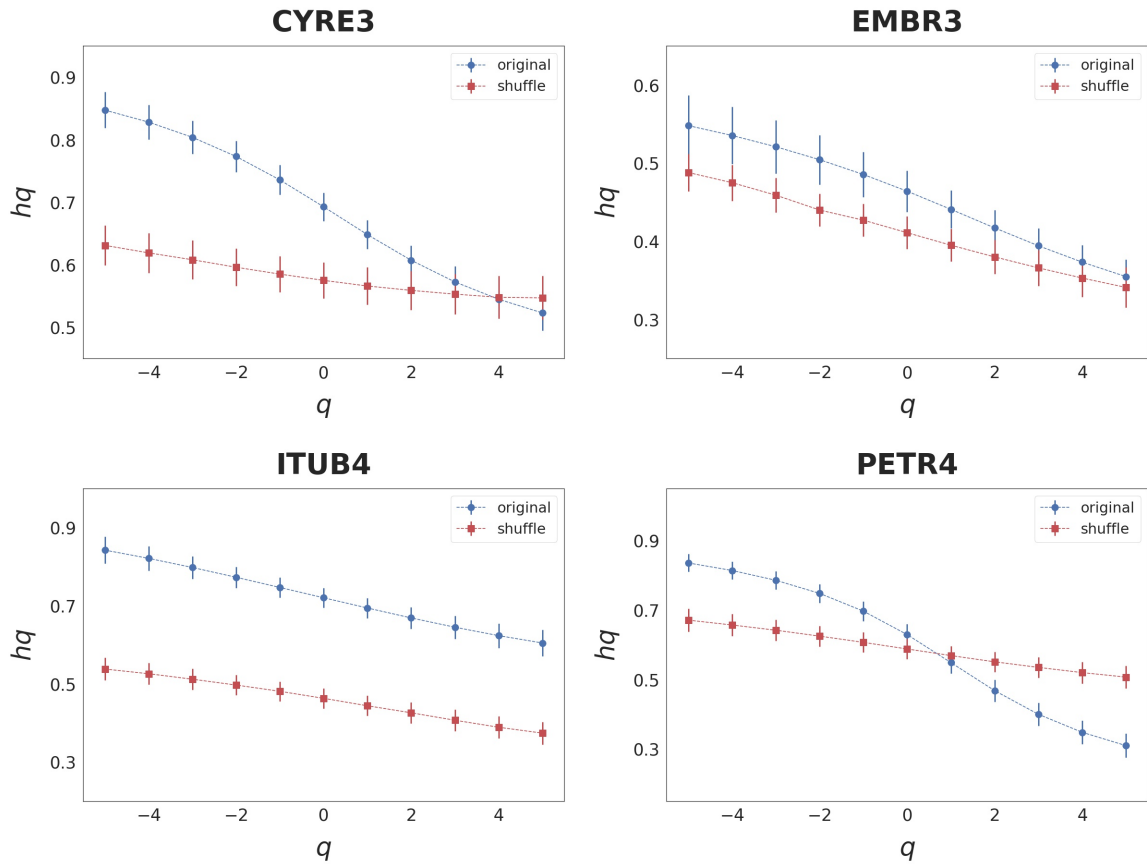
A origem da multifractalidade em uma série temporal pode ser verificada no processo de "embaralhamento", onde os valores da série original são colocados em ordem aleatória. Em geral, a multifractalidade pode ser explicada por dois fatores distintos:

- devido a uma ampla função de densidade de probabilidade para os valores da série temporal, isto é, uma distribuição de probabilidade de caudas pesadas. Neste caso, a multifractalidade não pode ser removida por "embaralhamento" aleatório e os dados "embaralhados" têm a mesma variação de  $h(q)$  que os dados originais;
- devido a uma variedade de correlações de longo alcance provocado por pequenas e grandes flutuações. Neste cenário, a série "embaralhada" não exibirá uma escala multifractal, dado que todas as correlações de longo alcance serão destruídas pelo processo de "embaralhamento", restando uma sequência totalmente não correlacionada.

Sendo assim, se uma série temporal tiver multifractalidade devido à correlação de longo alcance e também devido à forma da distribuição de probabilidade, a série "embaralhada" terá largura do espectro menor e, portanto, multifractalidade mais fraca do que a série original. Caso a multifractalidade seja devido somente à correlação de longo alcance, os dados "embaralhados" mostrarão uma escala não fractal.

Os valores das séries temporais de cada uma das ações modeladas neste trabalho foram "embaralhados" aleatoriamente para destruir todas as correlações de longo alcance presentes nas amostras. Os expoentes de Hurst, obtidos pela regressão linear que melhor se ajustou aos dados, são mostrados na figura 8 para diferentes valores de  $q$  (-5 a 5) para ambas as séries: original e "embaralhada" (*shuffle*).

Figura 8 – Gráficos comparativos dos expoentes de Hurst das séries temporais originais e "embaralhadas" (*shuffle*) das ações alvo.



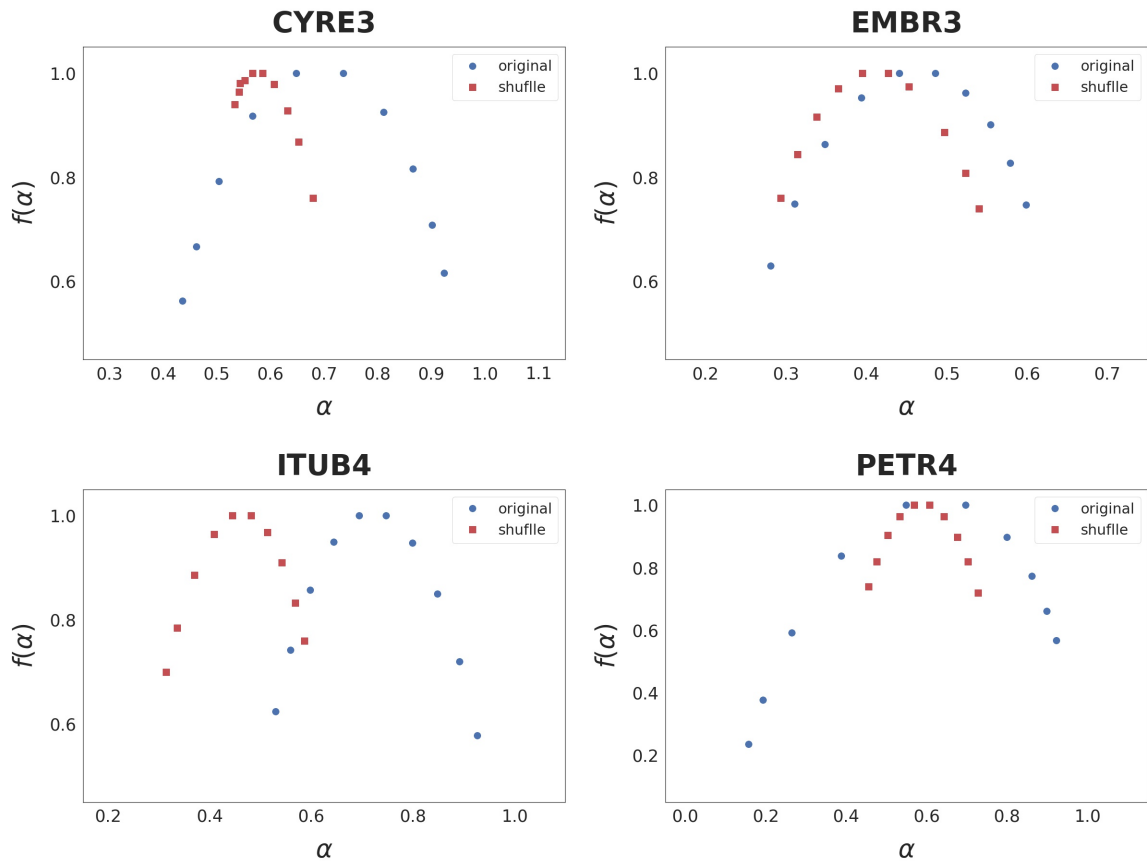
Fonte: Frederico Frias, 2021

Observa-se que a série temporal "embaralhada" das ações CYRE3 e PETR4 tem uma inclinação menos acentuada comparativamente com a da série original, com os expoentes  $h_s(q)$  aproximadamente constantes ao longo dos diferentes valores de  $q$ , o que denotaria a perda de sua estrutura multifractal. Isto implica dizer que todas as correlações foram destruídas. Contudo, pelos gráficos da figura 9, é possível evidenciar que o espectro multifractal para estes ativos possui máximos diferentes de  $\alpha_0 = 0.5$ . No caso ideal, os dados "embaralhados" devem se comportar como um sinal sem escala multifractal, onde o gráfico do expoente de Hurst não muda em geral com  $q$  e o gráfico de  $f(\alpha)$  apresenta um pico próximo a 0.5.

Para todas as ações alvo, a "força" da multifractalidade, determinada pela largura do espectro multifractal, é menor que o de suas séries temporais originais correspondentes, como pode ser observado na figura 9. Isso confirma o fato de que a multifractalidade é origi-

nada das correlações de longo alcance bem como da função de densidade de probabilidade ampla.

Figura 9 – Gráficos comparativos dos espectros multifractais  $f(\alpha)$  das séries temporais originais e "embaralhadas" (*shuffle*) das ações alvo.



Fonte: Frederico Frias, 2021

Os resultados dos expoentes de Hurst, obtidos pelos ajustes das regressões lineares, das *small caps* estão na tabela 4. A série original é determinada pelo expoente  $h(q)$  e a série "embaralhada", pelo  $h_s(q)$ .



Tabela 4 – Expoentes  $h(q)$  e  $h_s(q)$  das séries temporais originais e "embaralhadas", respectivamente, das ações alvo *small caps*.

$q$	CYRE3		EMBR3	
	$h(q)$	$h_s(q)$	$h(q)$	$h_s(q)$
-5	$0.847 \pm 0.029$	$0.631 \pm 0.032$	$0.548 \pm 0.039$	$0.488 \pm 0.024$
-4	$0.828 \pm 0.028$	$0.619 \pm 0.032$	$0.535 \pm 0.037$	$0.475 \pm 0.023$
-3	$0.804 \pm 0.027$	$0.608 \pm 0.031$	$0.521 \pm 0.034$	$0.459 \pm 0.022$
-2	$0.773 \pm 0.025$	$0.596 \pm 0.030$	$0.504 \pm 0.032$	$0.440 \pm 0.021$
-1	$0.736 \pm 0.029$	$0.585 \pm 0.029$	$0.486 \pm 0.029$	$0.427 \pm 0.021$
0	$0.693 \pm 0.023$	$0.575 \pm 0.029$	$0.464 \pm 0.026$	$0.411 \pm 0.021$
1	$0.648 \pm 0.023$	$0.566 \pm 0.030$	$0.441 \pm 0.025$	$0.395 \pm 0.021$
2	$0.607 \pm 0.024$	$0.559 \pm 0.031$	$0.417 \pm 0.023$	$0.380 \pm 0.022$
3	$0.573 \pm 0.025$	$0.553 \pm 0.032$	$0.394 \pm 0.022$	$0.366 \pm 0.023$
4	$0.545 \pm 0.027$	$0.548 \pm 0.034$	$0.373 \pm 0.022$	$0.353 \pm 0.024$
5	$0.523 \pm 0.028$	$0.547 \pm 0.035$	$0.355 \pm 0.022$	$0.341 \pm 0.026$

Fonte: Frederico Frias, 2021

A tabela 5 evidencia o comparativo entre os expoentes  $h(q)$  e  $h_s(q)$  das ações alvo consideradas *blue chips*.

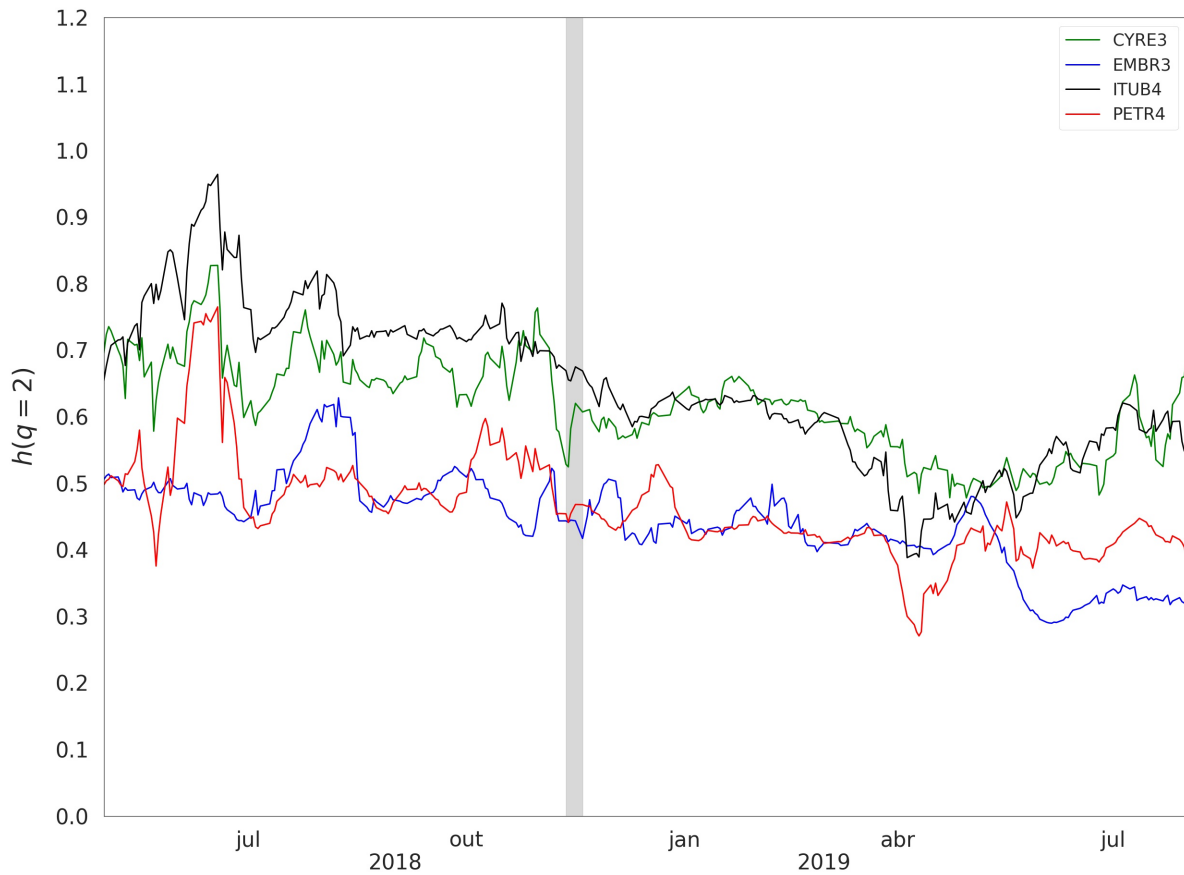
Tabela 5 – Expoentes  $h(q)$  e  $h_s(q)$  das séries temporais originais e "embaralhadas", respectivamente, das ações alvo *blue chips*.

$q$	ITUB4		PETR4	
	$h(q)$	$h_s(q)$	$h(q)$	$h_s(q)$
-5	$0.842 \pm 0.034$	$0.538 \pm 0.029$	$0.836 \pm 0.026$	$0.671 \pm 0.033$
-4	$0.821 \pm 0.032$	$0.526 \pm 0.028$	$0.814 \pm 0.026$	$0.657 \pm 0.032$
-3	$0.798 \pm 0.029$	$0.512 \pm 0.027$	$0.786 \pm 0.026$	$0.642 \pm 0.031$
-2	$0.773 \pm 0.027$	$0.497 \pm 0.026$	$0.748 \pm 0.027$	$0.625 \pm 0.030$
-1	$0.747 \pm 0.025$	$0.481 \pm 0.025$	$0.697 \pm 0.028$	$0.607 \pm 0.029$
0	$0.720 \pm 0.025$	$0.463 \pm 0.026$	$0.629 \pm 0.029$	$0.588 \pm 0.029$
1	$0.694 \pm 0.026$	$0.444 \pm 0.026$	$0.549 \pm 0.031$	$0.569 \pm 0.027$
2	$0.669 \pm 0.027$	$0.426 \pm 0.027$	$0.468 \pm 0.032$	$0.551 \pm 0.029$
3	$0.645 \pm 0.029$	$0.407 \pm 0.028$	$0.399 \pm 0.033$	$0.535 \pm 0.030$
4	$0.624 \pm 0.032$	$0.389 \pm 0.028$	$0.348 \pm 0.034$	$0.520 \pm 0.031$
5	$0.605 \pm 0.034$	$0.374 \pm 0.029$	$0.309 \pm 0.034$	$0.507 \pm 0.032$

Fonte: Frederico Frias, 2021

Uma análise útil para os diferentes participantes do mercado financeiro é saber como o expoente de Hurst muda ao longo do tempo: economistas tensionam determinar a natureza predominante dos mercados (persistente ou reversão à média); estrategistas governamentais pretendem estabelecer a posição atual da economia no ciclo de negócios; investidores de curto prazo querem explorar condições médias de reversão [37].

Figura 10 – Expoente de Hurst ( $q = 2$ ) entre o período de maio de 2018 a agosto de 2019 para cada uma das ações alvo.



Fonte: Frederico Frias, 2021



Sendo assim, determinou-se o expoente de Hurst ( $q = 2$ ) histórico para cada uma das ações modeladas. O intervalo de tempo considerado no cálculo do ajuste da regressão linear foi de onze meses. A cada medição de um expoente, a amostra de dados foi rolada em um dia para que o próximo fosse computado. Este procedimento foi realizado no período de maio de 2018 a agosto de 2019 com amostras com frequência diária.

Como observado na figura 10, os instrumentos EMBR3 e PETR4 possuem  $h < 0.5$ , ou seja, comportamento anti-persistente durante a janela de tempo utilizada na predição da tendência do movimento do preço da ação (área em cinza no gráfico). Séries temporais com expoentes de Hurst nesta faixa tem maior probabilidade de terem declínios (crescimentos) após uma alta (baixa) em sua tendência.

No caso da PETR4, a dinâmica do ativo pode ser considerada análoga à de um movimento *browniano*, apresentando uma característica aleatória, uma vez que os valores dos seus expoentes estão bem próximos de 0.5. Neste cenário, decrementos/incrementos são

igualmente prováveis (ver tabela 6), tornando a modelagem deste ativo pouco interessante durante este intervalo de tempo.

Tabela 6 – Características da dependência da série temporal em relação ao expoente  $h$ .

	$h \in [0.0, 0.5)$	$h \approx 0.5$	$h \in (0.5, 1.0]$
	0.0  0.5	$< 0.5 >$	0.5  1.0
Autocovariância	$< 0 \forall$ atraso	$= 0 \forall$ atraso	$> 0 \forall$ atraso
Comportamento	Anti-persistente	<i>Browniano</i>	Persistente
Interpretação Estatística	Decrementos (incrementos) com maior probabilidade de serem precedidos por incrementos	Decrementos / incrementos igualmente prováveis	Incrementos (decrementos) com maior probabilidade de serem precedidos por incrementos
	(decrementos)		(decrementos)
Característica	Reverte para a média com mais frequência do que para uma aleatória	Movimento aleatório	Exibem memória longa e "tendências" e "ciclos" de duração variável

Fonte: A. Karp, 2019 [37]

Já as ações CYRE3 e ITUB4 tem um comportamento persistente ( $h > 0.5$ ), caracterizado por memórias de longo prazo. Este tipo de série temporal, em geral, têm crescimentos (declínios) precedidos com maior probabilidade por incrementos (decrementos). Em particular, a ação CYRE3 possui uma subida abrupta em seu expoente de Hurst durante o período de modelagem. É importante ter este tipo de dinâmica ao se predizer a tendência do ativo uma vez que o preço do mesmo tende a ter uma continuidade no movimento, seja ela de alta ou de queda. Neste cenário, seria mais viável obter maiores retornos financeiros.

Após a caracterização da fractalidade dos ativos alvo, o próximo passo é construir um portfólio para cada um deles composto por ações com forte correlação com o ativo modelado. Isto é feito no capítulo a seguir, onde dois métodos são utilizados para se determinar a dependência entre duas séries temporais.

### 3 Seleção de ativos para a construção do portfólio

No capítulo anterior, caracterizou-se a dinâmica de séries temporais financeiras mediante a identificação de leis de escala. Dados empíricos diários das ações alvo foram analisados por intermédio do método MFDFA, onde constatou-se a presença de um espectro multifractal nesses ativos financeiros. A verificação do comportamento fractal é de relevante importância neste trabalho, pois, a partir desta propriedade, é possível embasar o conceito da metodologia utilizada na previsão do movimento dos preços das ações aqui estudadas.

A etapa seguinte do modelo quantitativo proposto consiste em selecionar as ações integrantes do Ibovespa que tenham uma dinâmica similar ao do ativo modelado. Para isso, é necessário caracterizar a relação entre a ação alvo e cada um dos demais instrumentos financeiros. Neste capítulo, são apresentados dois métodos que visam capturar a dependência entre duas séries temporais. Um desses modelos é o coeficiente DCCA de correlação cruzada que mensura, por meio do emprego dos métodos DFA e DCCA, a associação entre duas séries não estacionárias. O outro método, denominado mapeamento cruzado convergente, infere causalidade entre duas séries temporais ao mapear mutuamente os seus respectivos atratores. Este último visa confirmar a dependência, quantificada pelo primeiro método, de um ativo com o outro.

#### 3.1 Interdependência entre ativos

Um sistema complexo é composto por um conjunto de partes conectadas por alguma forma de inter-relacionamento entre as mesmas. Muito mais do que conhecer individualmente cada um dos seus constituintes, é necessário compreender os modos de relação entre os elementos do sistema afim de caracterizá-lo, pois o comportamento global de um sistema complexo é distinto daquele obtido considerando-se unicamente a soma de suas partes<sup>1</sup>. Em outras palavras, as propriedades desse sistema só podem ser identificadas durante o seu comportamento coletivo, sendo irreduzíveis no nível de seus elementos.

Entretanto, ao analisar o sistema como um todo, um fluxo de informações não triviais para investigar vem à tona, com uma série de consequências e propriedades emergentes. Emergência é uma característica essencial de um sistema complexo, que define

---

<sup>1</sup> Aristóteles: *O todo é maior do que a simples soma das suas partes.*

o *feedback* entre o fenômeno e o agente, cujo efeito, aparentemente, não apresenta uma causa conhecida [38].

O mercado acionário pode ser considerado um sistema complexo. Adicionalmente, as ações que o compõe são os seus constituintes, podendo ter uma relação de dependência estatística, causal ou não causal, a medida em que são simultaneamente negociadas. Com o intuito de capturar esta dependência entre os elementos do sistema que é o mercado de ações brasileiro, mais especificamente os que são integrantes do Ibovespa, foram empregados os métodos  $\rho_{DCCA}$  e CCM, descritos a seguir.

### 3.1.1 Quantificando correlação cruzada

As séries temporais financeiras podem ser estudadas no domínio da frequência ou no domínio do tempo. Neste último, a maioria dos métodos estatísticos envolve algum tipo de modelo linear, sendo o mais comum deles o coeficiente de correlação de Pearson [39]:

$$\rho_{xy} = \frac{Cov(x, y)}{\sigma_x \sigma_y}, \quad (18)$$

onde  $Cov$  é a covariância e  $\sigma$  é o desvio padrão. Este coeficiente mensura a associação linear entre duas séries temporais estacionárias, falhando no caso de séries não estacionárias ou quando a relação entre ambas é não linear.

As séries não estacionárias não conseguem manter o seu valor médio, a sua variância ou os seus momentos de ordem superior invariantes no tempo. Para analisar este tipo de série temporal, existem pelo menos duas abordagens possíveis: por diferenciação matemática ou por remoção de tendência [40]. Nesta última, um ajuste polinomial de ordem  $m$  pode eliminar uma tendência externa que é a responsável pela não estacionariedade na maior parte das vezes.

O método DCCA, do inglês, *detrended cross-correlation analysis*, foi proposto com o intuito de minimizar os efeitos dessas tendências externas, tendo a habilidade de mensurar as correlações cruzadas de longo alcance em diversas escalas de tempo entre duas séries temporais não estacionárias de mesmo período de tempo e de igual comprimento, por meio da função de covariância  $\mathcal{F}_{DCCA}^2(s)$ . Os autores do DCCA [41] encontraram correlações cruzadas de lei de potência entre os *log*-retornos dos índices norte-americanos, Dow Jones e NASDAQ, constatando que  $\mathcal{F}_{DCCA} \sim s^{\lambda_{DCCA}}$ , onde  $\lambda_{DCCA} \approx (\alpha_x + \alpha_y)/2$ ,

com  $\alpha = \alpha_{DFA}$ . Caso a série temporal fosse correlacionada com ela própria, a função de covariância proposta pelos autores se reduziria à função de variância  $\mathcal{F}_{DFA}^2(s)$ .

Contudo, o expoente  $\lambda_{DCCA}$  não quantifica o nível de correlação cruzada entre as séries temporais. Com o propósito de suprir esta necessidade, Zebende propôs o coeficiente DCCA de correlação cruzada ( $\rho_{DCCA}$ ) [42]. Além de identificar as forças de acoplamento em diferentes escalas de tempo, este coeficiente é mais robusto do que o coeficiente de Pearson ao analisar o grau de dependência entre duas de séries temporais não estacionárias [40, 43].

O coeficiente  $\rho_{DCCA}$  é um método obtido em 5 passos [30], sendo os dois primeiros idênticos aos passos 1 e 2 do método MFDFA, ambos descritos na seção 2.3.

No terceiro passo, a função de flutuação é calculada para cada segmento  $\nu$ :

$$\mathcal{F}_{xy}^2(\nu, s) = \frac{1}{s} \sum_{i=1}^s \left[ x([\nu-i]s+i) - x_{\nu}^m(i) \right] \left[ y([\nu-i]s+i) - y_{\nu}^m(i) \right], \quad \nu = 1, 2, \dots, N, \quad (19)$$

$$\mathcal{F}^2(\nu, s) = \frac{1}{s} \sum_{i=1}^s \left[ x([T-1]-[\nu-N]s+i) - x_{\nu}^m(i) \right] \left[ y([T-1]-[\nu-N]s+i) - y_{\nu}^m(i) \right], \quad \nu = N+1, \dots, 2N, \quad (20)$$

onde  $x$  e  $y$  são, respectivamente, as séries  $x(k)$  e  $y(k)$  de mesmo período de tempo e de igual comprimento e  $x_{\nu}^m$  e  $y_{\nu}^m$  são os seus correspondentes polinômios de ajuste de ordem  $m$  em cada segmento  $\nu$ .

No passo seguinte, é calculada a função de flutuação de ordem  $q$ :

$$\mathcal{F}_{xy}(q, s) = \left( \frac{1}{2N} \sum_{\nu=1}^{2N} \left[ \mathcal{F}_{xy}^2(\nu, s) \right]^{q/2} \right)^{\frac{1}{q}}. \quad (21)$$

Em princípio,  $q$  pode assumir qualquer valor real, exceto zero. Para  $q = 2$ , temos o DCCA padrão:

$$\mathcal{F}_{xy}^2(s) = \frac{1}{2N} \sum_{\nu=1}^{2N} \mathcal{F}_{xy}^2(\nu, s). \quad (22)$$

No quinto e último passo, o coeficiente DCCA de correlação cruzada é dado pela seguinte relação:

$$\rho_{DCCA}(s) = \frac{\mathcal{F}_{xy}^2(s)}{\sqrt{\mathcal{F}_{xx}^2(s)\mathcal{F}_{yy}^2(s)}}. \quad (23)$$

Analogamente ao coeficiente de correlação de Pearson, o coeficiente  $\rho_{DCCA}$  pode assumir valores entre -1 e 1, sendo:

$$\sigma_{DCCA} = \begin{cases} +1 \rightarrow \text{correlação perfeita} \\ 0 \rightarrow \text{sem correlação} \\ -1 \rightarrow \text{anti-correlação} \end{cases} . \quad (24)$$

Segundo os autores do artigo [40], o coeficiente DCCA de correlação cruzada é capaz de mensurar a relação linear entre duas séries temporais não estacionárias em múltiplas escalas de tempo. Além disso, o  $\rho_{DCCA}$  é equivalente ao coeficiente de correlação de Pearson quando este último é aplicado às séries derivadas das séries originais após a eliminação das tendências externas. Nesse mesmo estudo, o coeficiente DCCA de correlação cruzada se mostrou mais sensível do que o de Pearson a mudanças do parâmetro que determina a "força" de correlação cruzada entre duas séries temporais do modelo de média móvel autorregressiva fracionada integrada, do inglês, *autoregressive fractionally integrated moving average* (ARFIMA), sendo, portanto, mais eficiente ao detectar o efeito de memória dos dados e as correlações cruzadas reais.

### 3.1.2 Busca por relação causal

Do ponto de vista da complexidade, o mercado acionário é um sistema dinâmico aberto que evolui com o tempo, composto por múltiplos agentes interagindo entre si, apresentando geometria fractal e vários estados de equilíbrio conhecidos como atratores. São sistemas que estão em um estado entre a ordem e a desordem.

De forma sucinta, um sistema dinâmico é um processo em constante evolução temporal, podendo ser descrito genericamente por uma equação do tipo:

$$x_n = \Omega_t(x_0), \quad (25)$$

onde  $t$  é o tempo contínuo,  $x_n$  é o sistema no estado  $n$  e  $\Omega$  é a função que leva o sistema de um estado a outro. Como observado na equação 25, este tipo de sistema é dependente do seu estado inicial  $x_0$ .

Em um sistema composto por um conjunto de sistemas dinâmicos interagindo mutuamente, como é o caso do sistema acionário com suas diferentes ações, o sistema dinâmico pode ser representado por:

$$\sum_{i=1}^N x_n^i = \Omega_t^i(x_0^i, x_0^j), \quad i \neq j. \quad (26)$$

Em 1963, Edward Lorenz, ao estudar a dinâmica atmosférica de convecção [44], sugeriu um modelo de equações diferenciais semelhantes à equação 26:

$$\begin{cases} \dot{x} &= \sigma(y - x) \\ \dot{y} &= x(\rho - z) - y \\ \dot{z} &= xy - \beta z \end{cases} \quad (27)$$

O sistema de equações de Lorenz é um sistema não linear que tem como intuito descrever uma camada atmosférica através de uma representação aproximada de um fluido uniforme sendo aquecido na parte inferior e resfriado na parte superior, onde  $x$  é a intensidade do movimento convectivo,  $y$  é a diferença de temperatura entre as correntes ascendentes e descendentes,  $z$  é a distorção vertical do perfil de temperatura e  $\sigma$ ,  $\rho$  e  $\beta$  são parâmetros hidrodinâmicos. Ao analisar este modelo, Lorenz constatou que só era possível prever a dinâmica do sistema para um curto intervalo de tempo, uma vez que o mesmo é bastante dependente das condições iniciais, evoluindo para estados muito diferentes apesar dos seus estados iniciais serem bem próximos entre si. Apesar de ser um sistema determinístico, o mesmo exibia um comportamento caótico.

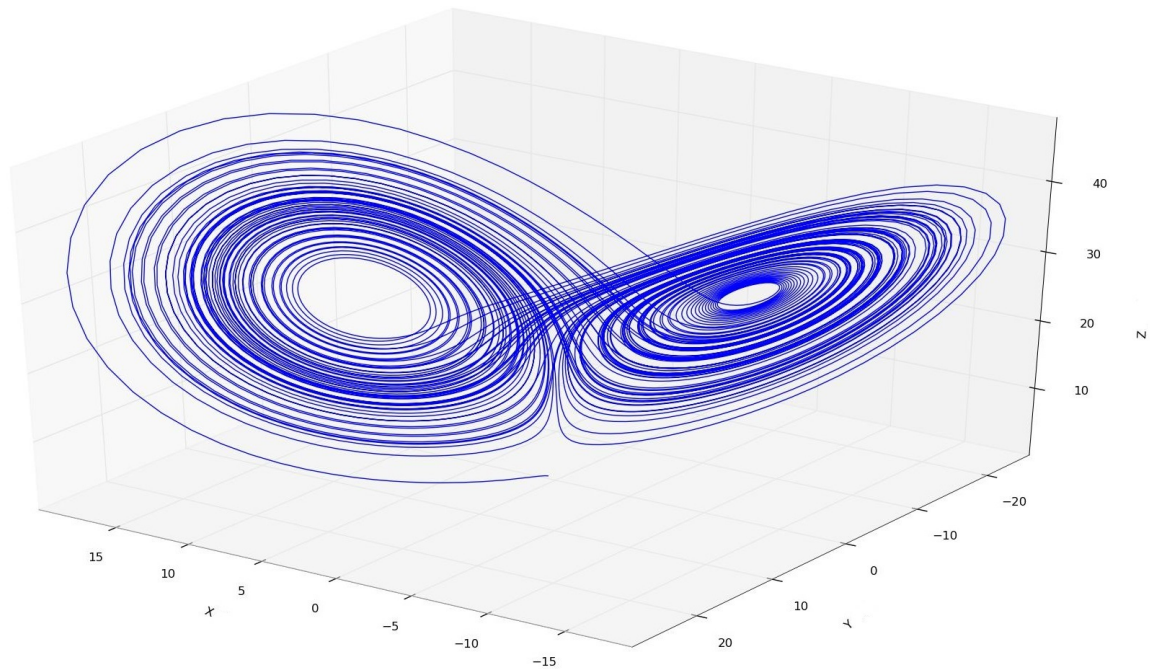
Embora não fosse possível realizar previsões a longo prazo, o sistema estava preso a uma região limitada do espaço-de-fase<sup>2</sup> denominada, neste caso, por "atrator estranho" [45]. Um atrator é um conjunto de comportamentos característicos para o qual um sistema dinâmico evoluiu. Como exemplo, temos o movimento de um pêndulo que converge para uma oscilação de período constante (atrator). A denominação "estranho" veio do fato do atrator possuir uma dimensão fractal (igual a 2.3).

O atrator de Lorenz é obtido ao integrar numericamente as equações 27, colocando a evolução temporal de cada uma das variáveis do sistema em um espaço tridimensional, como mostra a figura abaixo. É possível notar que as trajetórias nunca passam mais de uma vez pelo mesmo ponto como também nunca se cruzam.

<sup>2</sup> Representação das variáveis dinâmicas relevantes de um sistema num mesmo espaço euclidiano para uma série de parâmetros e condições iniciais.



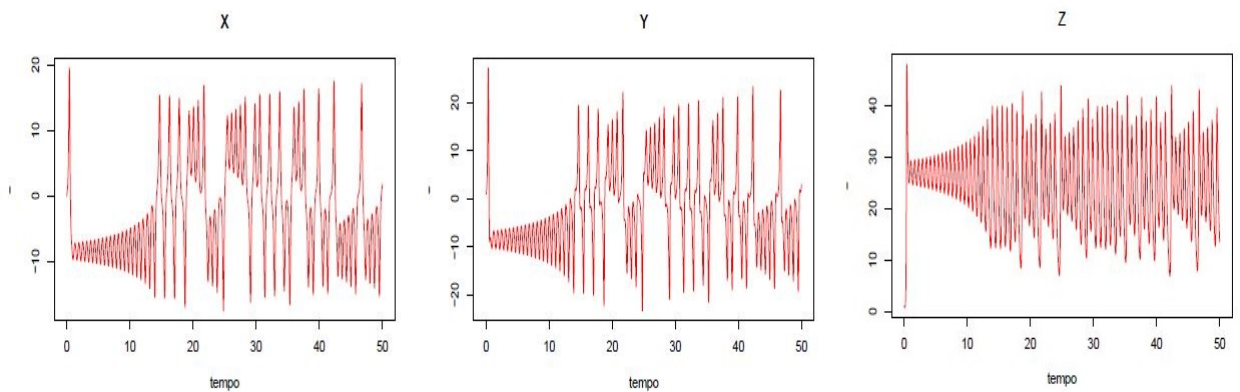
Figura 11 – Atrator de Lorenz simulado com os seguintes parâmetros hidrodinâmicos:  $\sigma = 10$ ,  $\rho = 28$  e  $\beta = 8/3$ .



Fonte: Frederico Frias, 2021

Ao projetar a dinâmica do atrator de Lorenz sobre os eixos do espaço tridimensional euclidiano, cada uma das variáveis das equações 27 pode ser representada por um série temporal, conforme pode ser observado na figura 12. A projeção de cada uma dessas séries num único espaço-de-fase é o que gera o atrator de Lorenz.

Figura 12 – Séries temporais obtidas pela projeção de cada uma das variáveis do sistema de equações de Lorenz.



Fonte: Frederico Frias, 2021

Em 1981, Floris Takens formulou um teorema capaz de reconstruir um atrator com base em apenas umas das séries temporais que o constitui [46]. Em suma, este teorema, definido abaixo em uma de suas formas simplificadas, afirma que através das séries temporais do atrator em  $\mathbf{M}$  e seus atrasos temporais ( $\tau$ ), chamadas de funções observações da variedade, é possível reconstruir seja qual for o atrator descrito no mesmo. Este novo atrator, denominado variedade-sombra ( $\mathbf{M}_s$ ), é um "mergulho" da variedade original, possuindo todas as propriedades topológicas da mesma, além da dinâmica do atrator reconstruído [47].

**Teorema 1 (Teorema de Takens)** *Seja  $\mathbf{M}$  uma variedade diferenciável compacta de dimensão  $d$  e  $\Omega: \mathbf{M} \rightarrow \mathbf{M}$  um difeomorfismo  $C^2$  com as seguintes propriedades:*

1. *Existe um número finito de pontos periódicos de  $\Omega$  com período menor ou igual a  $2d$ ;*
2. *Se  $\mathbf{x}$  é um ponto periódico com período  $k \leq 2d$ , então os valores próprios da derivada de  $\Omega^k$  são todos distintos.*

*Então, para  $\mathbf{y}$  genérico, a aplicação*

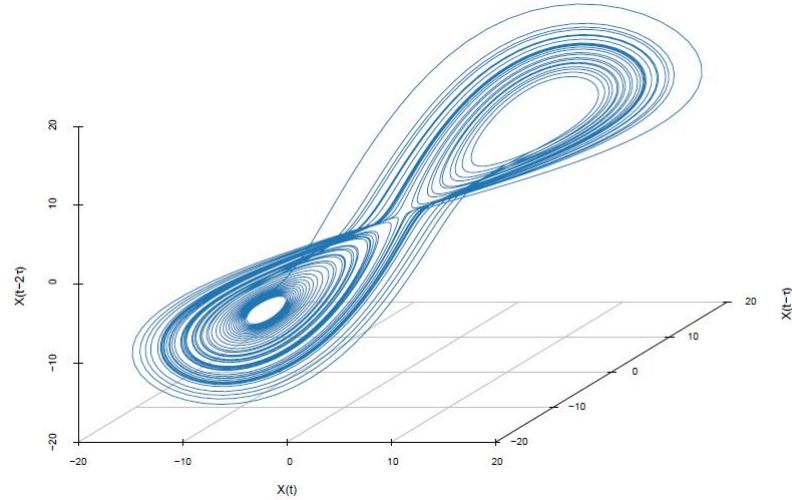
$$\Omega_{(\Omega, \mathbf{y})}(\mathbf{x}) = (\mathbf{y}(\mathbf{x}), \mathbf{y}(\Omega(\mathbf{x})), \mathbf{y}(\Omega^2(\mathbf{x})), \dots, \mathbf{y}(\Omega^{2d}(\mathbf{x})))$$

*é um mergulho.*

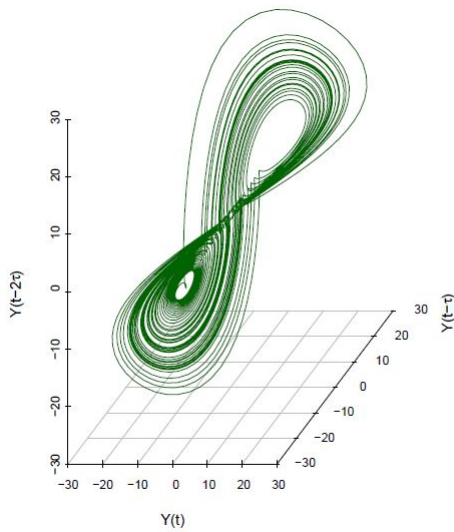
O teorema 1 é de extrema importância no estudo de sistemas dinâmicos, uma vez que nem sempre se tem ciência acerca de todas as variáveis que constituem um dado sistema. Desta forma, pode-se determinar as principais características que o regem a partir da reconstrução do atrator original mediante a utilização de algumas de suas variáveis.

A figura 13 exemplifica a reconstrução do atrator de Lorenz a partir de cada uma das variáveis do sistema de equações 27.

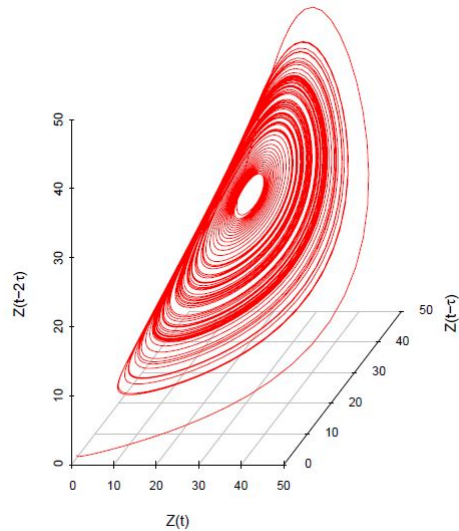
Figura 13 – Reconstrução do atrator de Lorenz utilizando cada uma das variáveis do sistema de equações 27 e seus atrasos temporais ( $\tau = 6$ ).



(a) Variedade-sombra de  $x$



(b) Variedade-sombra de  $y$



(c) Variedade-sombra de  $z$

Fonte: Frederico Frias, 2021

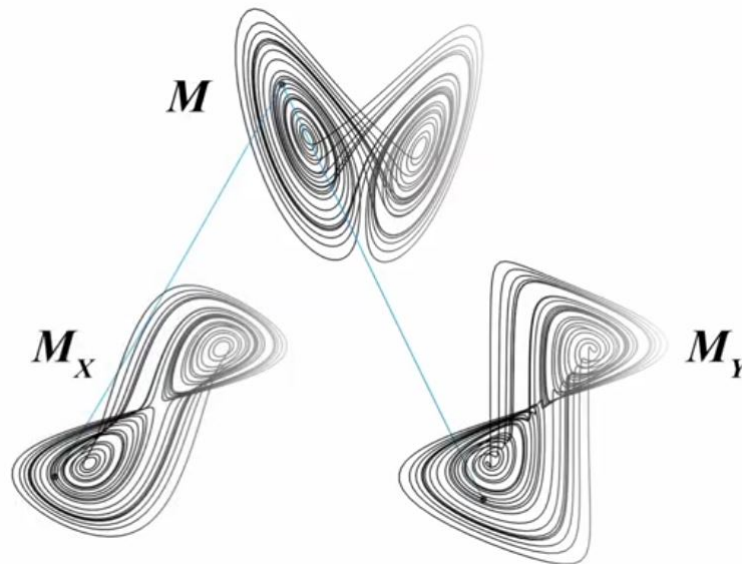
É importante salientar que a variedade-sombra da variável  $z$ , conforme pode ser constatado pela figura 13c, não reproduz o atrator original uma vez que esta série temporal não é um mapeamento de  $\mathbf{M}$  em  $\mathbb{R}$ , violando uma premissa do teorema de Takens que afirma que as funções observações  $\mathbf{M}_s$  mapeiam 1:1 a variedade original  $\mathbf{M}$ .

Uma consequência deste teorema é que é possível inferir se duas séries temporais distintas tem como origem o mesmo sistema dinâmico ao mapear os seus respectivos espaços-de-fase reconstruídos entre si. Se cada variedade-sombra,  $\mathbf{M}_x$  e  $\mathbf{M}_y$ , mapeia na

variedade original  $\mathbf{M}$ , então é possível mapear as variedades-sombra uma na outra. Sendo assim, ambas as séries temporais participam do mesmo sistema dinâmico.

No artigo [48], Sugihara propôs uma metodologia com intuito de capturar a relação de causa e efeito entre duas séries temporais ecológicas e/ou biológicas. Batizado de mapeamento cruzado convergente, do inglês, *convergent cross mapping* (CCM), este método assume que todas as variáveis do sistema dinâmico são regidas por fatores determinísticos, sendo possível inferir que as mesmas têm em comum o mesmo atrator no espaço-de-fase.

Figura 14 – Exemplificação do mapeamento cruzado convergente entre as variedades-sombras  $M_x$  e  $M_y$ .



Fonte: Sugihara Lab

O conceito deste procedimento é tentar mapear mutuamente os espaços-de-fase reconstruídos a partir das séries temporais estudadas. Caso seja possível identificar uma variedade-sombra a partir de outra variedade-sombra, então, as duas séries estão dinamicamente ligadas e, conseqüentemente, há uma relação de causalidade entre ambas. Se os pontos de uma dada variedade-sombra  $\mathbf{M}_x$  estimam os pontos de uma outra variedade-sombra  $\mathbf{M}_y$ , então,  $\mathbf{x}$  mapeia  $\mathbf{y}$  e, portanto,  $\mathbf{y}$  é a causa de  $\mathbf{x}$ . A explicação para isso é que o efeito guarda informação a respeito de sua causa.

Ao contrário do método de Granger [49, 50], o mapeamento cruzado convergente pode ser utilizado para detectar a causalidade entre séries temporais que possuem uma dependência não linear entre si.

Abaixo, temos o passo-a-passo do método CCM.

- Passo 1: sejam duas séries temporais de preço de mesmo tamanho  $N$  e período de tempo:  $x(t)$  e  $y(t)$ .
- Passo 2: construir o vetor de atraso  $Y_i(t)$  com dimensão de incorporação  $d$  e tempo de atraso  $\tau$ :

$$Y_i = [y(i), y(i - \tau), y(i - 2\tau), \dots, y(i - [d - 1]\tau)], \quad (28)$$

onde  $i = (d - 1)\tau + 1, (d - 1)\tau + 2, \dots, N$ .

- Passo 3: para cada  $Y_i$ , encontrar os  $d + 1$  vizinhos mais próximos de  $M_y$  denotados por  $n_j \forall j = 1, \dots, d + 1$ .
- Passo 4: usando os vizinhos mais próximos  $n_j$  de  $M_y$ , localize os valores correspondentes em  $x(t)$ :  $x_j, x_{j+1}, \dots, x_{d+1}$ .
- Passo 5: gere uma estimativa cruzada de  $x(t)$  dada por  $\hat{x}|M_y$  da média ponderada de  $x_j$ 's:

$$\hat{x} = \sum_{j=1}^{d+1} w_j \cdot x_j, \quad (29)$$

$$w_j = \frac{u_j}{\sum u_k}, \quad k = 1, \dots, d + 1, \quad u_j = \exp\left(-\frac{\|Y - Y_j\|}{\|Y - Y_1\|}\right). \quad (30)$$

- Passo 6: calcular o coeficiente de correlação:

$$\rho_{\hat{x}|M_y} = \frac{\sum \hat{x}_i \cdot x_i - N \cdot \bar{\hat{x}} \cdot \bar{x}_i}{\sqrt{\sum x_i^2 - N \cdot \bar{x}^2} \cdot \sqrt{\sum \hat{x}_i^2 - N \cdot \bar{\hat{x}}^2}}. \quad (31)$$

- Passo 7: repetir os passos de 1 a 6 para diferentes comprimentos  $N$ .

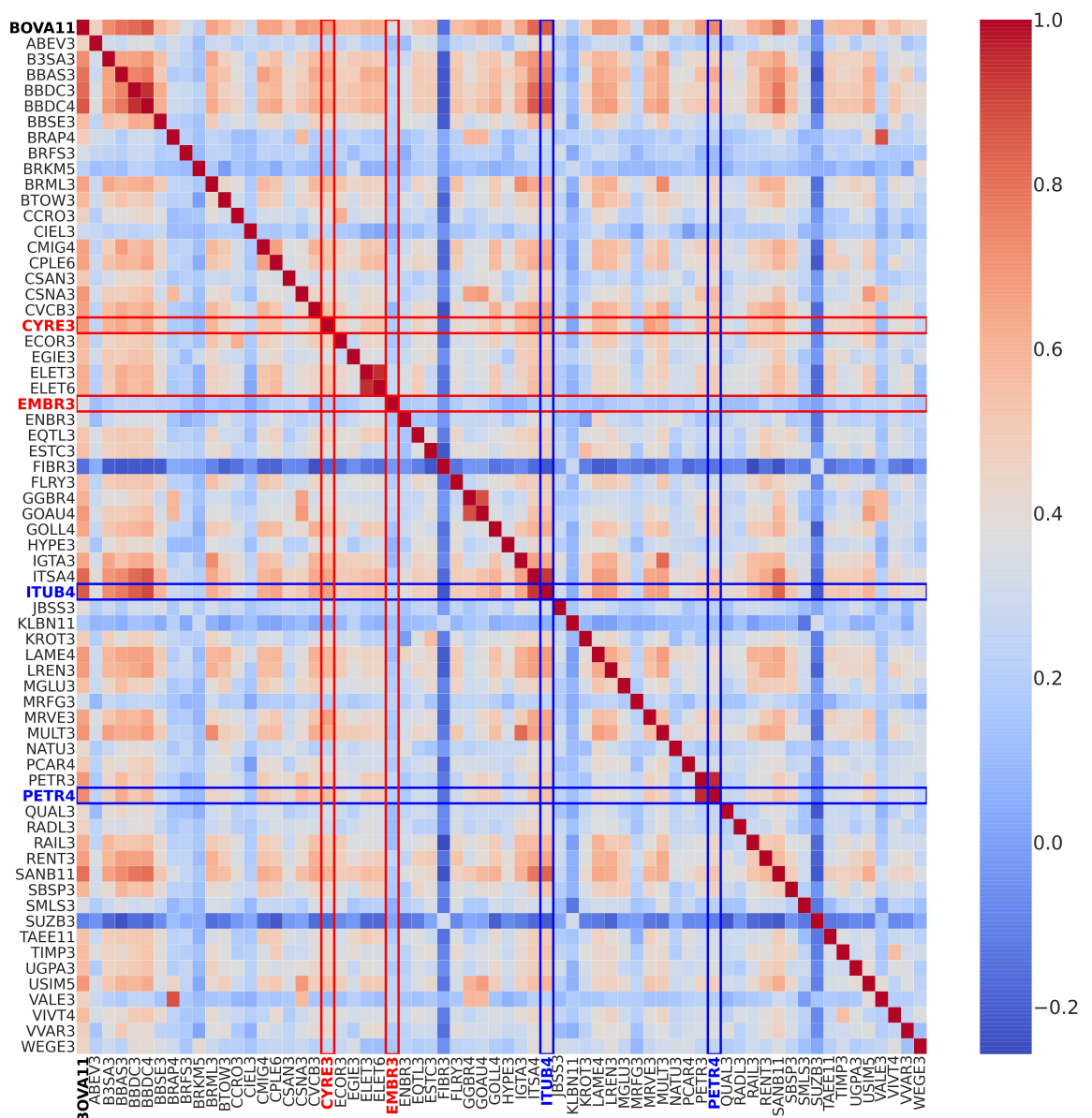
### 3.2 Composição da carteira correlacionada

Nesta seção, serão evidenciados os resultados dos dois métodos utilizados para capturar a relação de dependência entre a ação modelada e as demais constituintes do Ibovespa. Os dados possuem frequência diária e o período de tempo usado para mensurar a interdependência entre os ativos foi de onze meses (219 observações), entre janeiro a novembro de 2018, terminando na semana utilizada na modelagem da tendência do ativo alvo.

Para cada ação estudada, foram elencadas as ações com os maiores coeficientes  $\rho_{DCCA}$ . O critério de seleção foi baseado em um coeficiente positivo com valor acima do

coeficiente de referência. Neste caso, usou-se como *benchmark* o coeficiente de Pearson, dado pela equação 18, onde os resultados para cada um dos ativos integrantes do índice Bovespa são apresentados na figura 15. Esta medida de correlação é bastante difundida na literatura, sendo capaz de quantificar a dependência linear entre duas séries temporais estacionárias, como mencionado anteriormente.

Figura 15 – Mapa de calor do coeficiente de Pearson do *log*-retorno, com frequência diária, das ações do Ibovespa entre janeiro a novembro de 2018. Destacado em vermelho estão as *small caps* e, em azul, as *blue chips*. O ETF BOVA11 encontra-se em negrito.

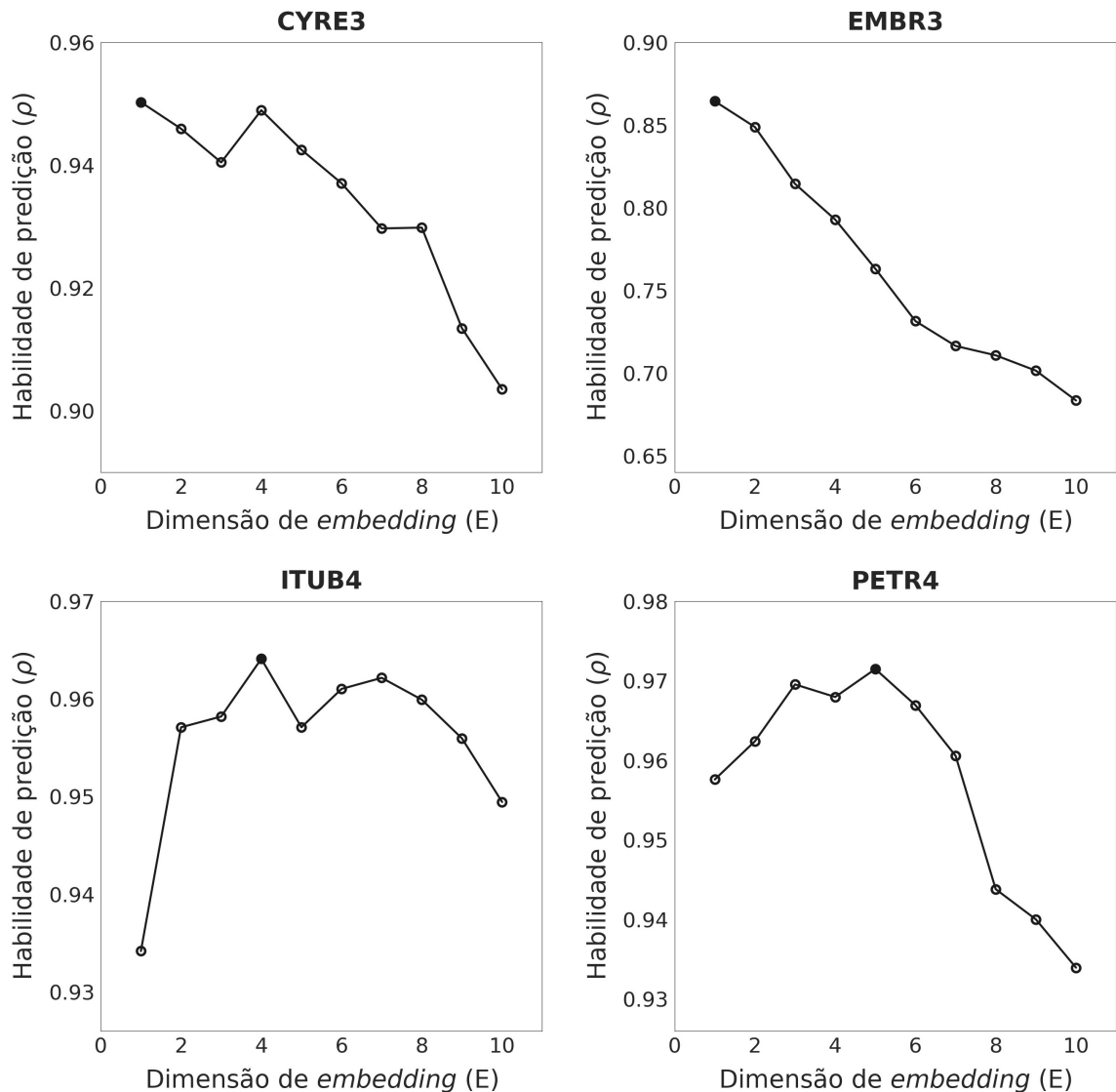


Fonte: Frederico Frias, 2021

Entretanto, correlação não implica necessariamente que haja uma relação de causa e efeito entre os ativos. Isto posto, foi aplicada a técnica do CCM com o intuito de confirmar que as ações listadas pelo método anterior possuem uma ligação causal com a ação estudada.

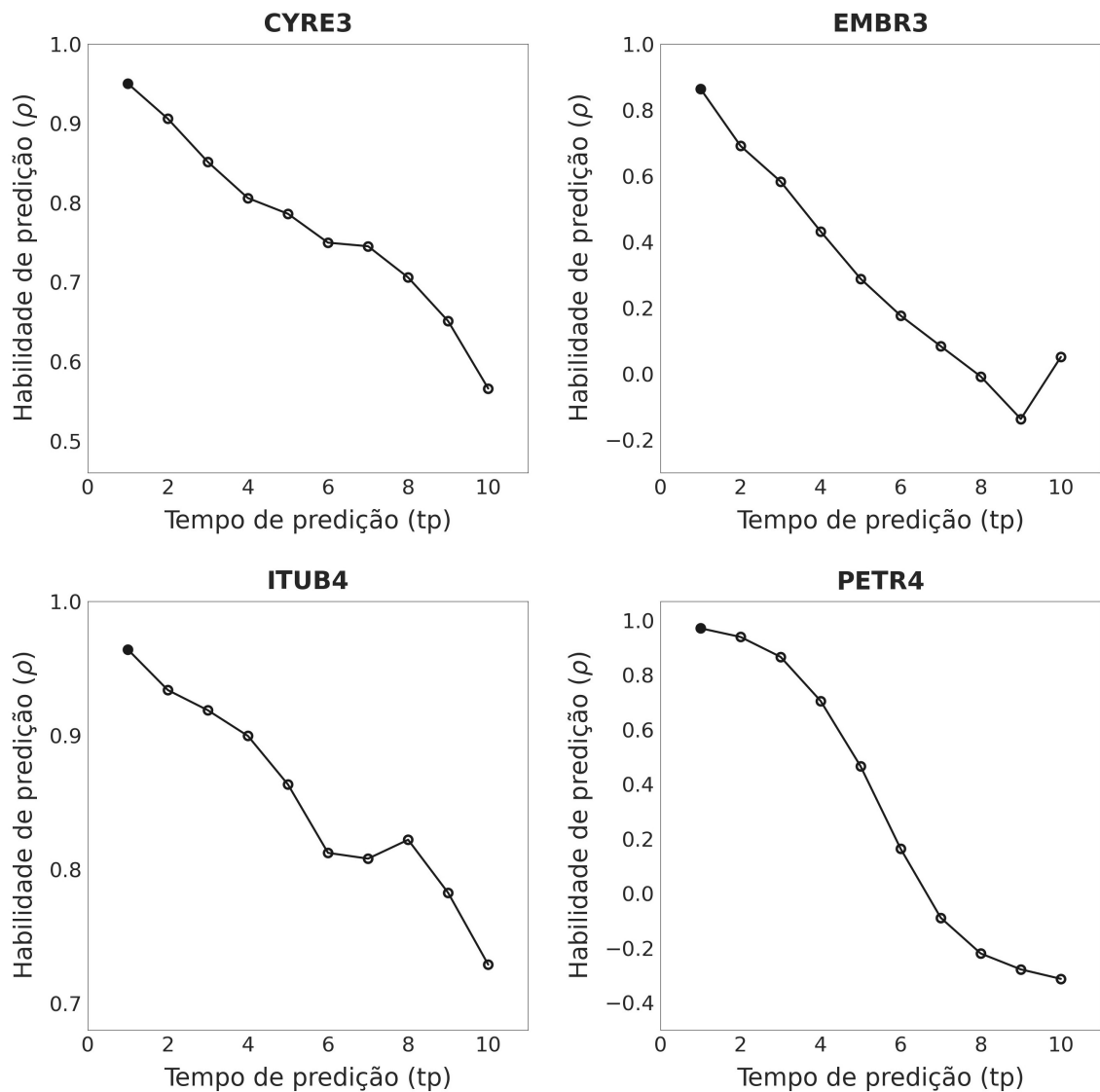
O primeiro passo é encontrar a dimensão de incorporação, conhecida como dimensão de *embedding*. Esta dimensão é calculada por um método de predição do vizinho mais próximo, projeção *simplex*, onde a habilidade de predição é mensurada pela correlação entre os valores observados e os preditos.

Figura 16 – Gráficos da dimensão de *embedding* (E), calculada pela projeção simplex, para cada ação alvo. O ponto preenchido denota o máximo da habilidade de predição ( $\rho$ ).



A habilidade de predição pode ser usada como um indicador para a dimensão de incorporação ideal. Se a mesma for muito pequena, a reconstrução do atrator terá singularidades, ou seja, pontos que correspondem a diferentes estados do sistema, mas se sobrepõem na reconstrução, prejudicando o desempenho da predição. A melhor dimensão de *embedding* será aquela em que a reconstrução será mapeada um a um para o atrator original. Na figura 16, evidencia-se a dimensão de incorporação utilizada no cálculo do CCM de cada uma das ações modeladas neste trabalho.

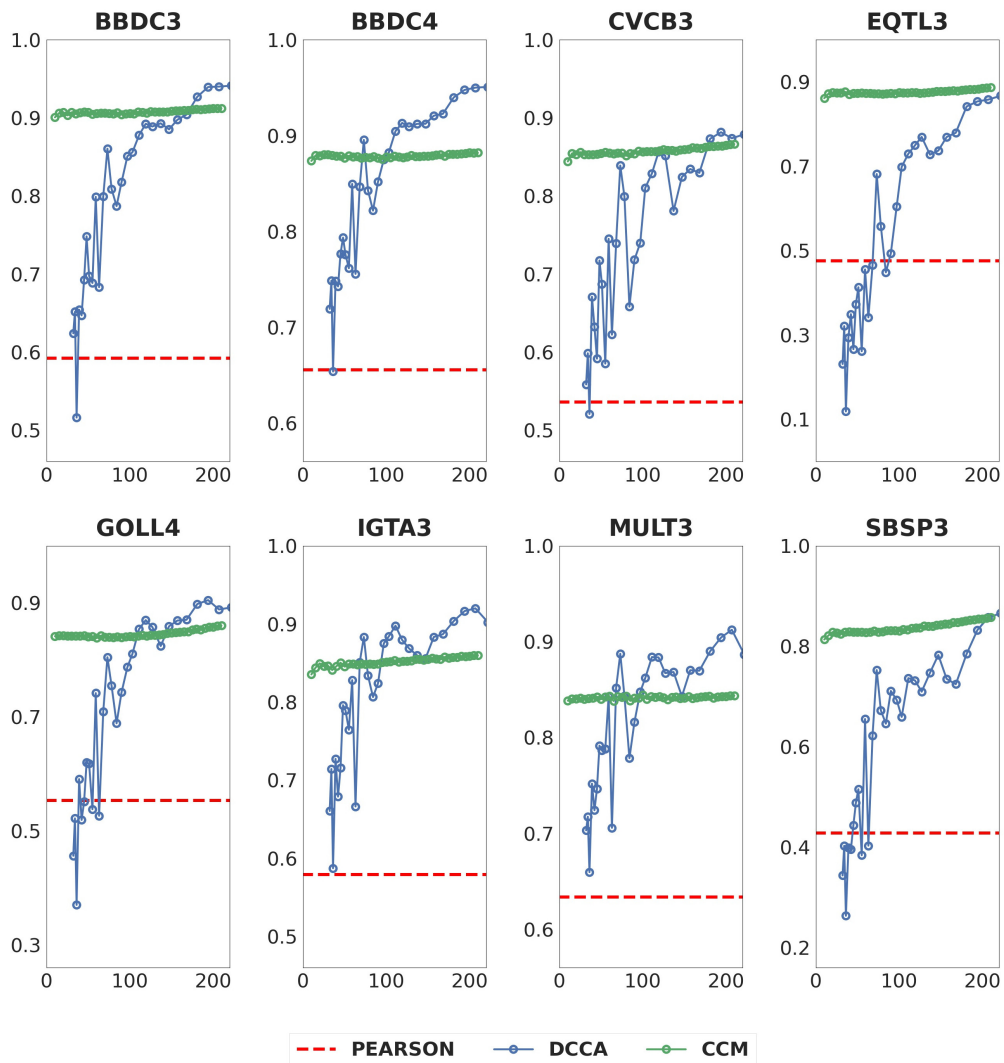
Figura 17 – Gráficos do decaimento do tempo de predição ( $tp$ ), calculado pela projeção simplex, para cada ação alvo. O ponto preenchido denota o máximo da habilidade de predição ( $\rho$ ), sendo igual a 1 dia para todos os ativos.





Outra variável importante a se determinar antes de aplicar a técnica do mapeamento cruzado convergente é o tempo de predição  $tp$ , isto é, o número de passos de tempo no futuro em que as predições são realizadas. Sabe-se que em muitos sistemas naturais as trajetórias próximas podem divergir ao longo do tempo, acarretando perda de informações sobre o estado do sistema e, conseqüentemente, dificultando a predição de longo prazo. Afim de demonstrar esse efeito, é possível examinar a habilidade de predição a medida em que  $tp$  aumenta. A figura 17 mostra o tempo, em dias, obtido para cada uma das ações alvo, lembrando que foi utilizada a melhor dimensão de *embedding* em cada caso (figura 16).

Figura 18 – Gráficos do  $\rho_{DCCA}$  e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo CYRE3. O eixo das abscissas está em dias.



Fonte: Frederico Frias, 2021

Por fim, o CCM pode ser aplicado para testar as relações causais entre os ativos. Para cada teste individual, utilizou-se a dimensão de incorporação correspondente à melhor dimensão de *embedding* da variável usada na reconstrução como também o melhor tempo de predição  $tp$ . É importante salientar que o interesse neste estudo é tentar inferir que uma determinada ação afeta a dinâmica da ação modelada, isto é, que o ativo alvo é o efeito cuja causa é o ativo integrante do portfólio.

Das ações listadas anteriormente pelo coeficiente DCCA de correlação cruzada, verificou-se, pelo método do mapeamento cruzado convergente, quais delas tem uma influência causal sobre a ação modelada. As oito que apresentaram os maiores valores de  $\rho_{DCCA}$  e do CCM foram selecionadas para a construção de uma carteira de ativos que será utilizada no capítulo subsequente como base para a classificação da tendência do preço da ação alvo.

Na figura 18, temos os ativos que apresentaram uma forte dependência com o ativo CYRE3. Em azul, evidencia-se o perfil do coeficiente  $\rho_{DCCA}$  ao longo do tempo. A linha em verde representa a série temporal do preço da ação CYRE3 mapeando a da ação específica, o que implica dizer que essa dada ação é a causa de CYRE3. O coeficiente de Pearson, que foi usado como referência, é a linha tracejada em vermelho. Os valores de cada uma das três técnicas empregadas está na tabela 7.

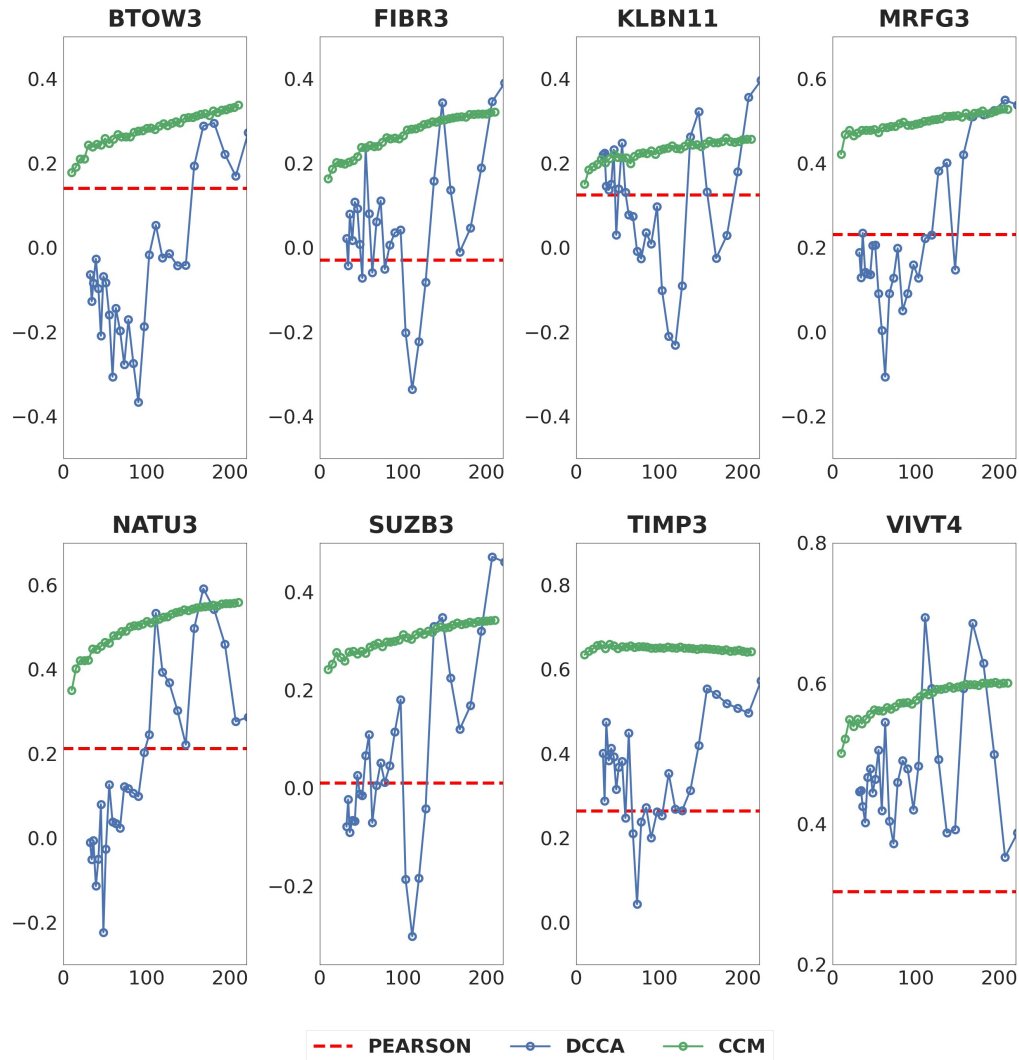
Tabela 7 – Valores do  $\rho_{DCCA}$  e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo CYRE3.

Ação	Pearson	$\rho_{DCCA}$	CCM
BBDC3	0.5923	0.9502	0.9122
BBDC4	0.6556	0.9584	0.8822
CVCB3	0.5363	0.9036	0.8665
EQTL3	0.4761	0.8942	0.8871
GOLL4	0.5531	0.9151	0.8602
IGTA3	0.5794	0.9268	0.8598
MULT3	0.6338	0.9115	0.8439
SBSP3	0.4284	0.8734	0.8577

Fonte: Frederico Frias, 2021

O portfólio da ação EMBR3 é constituído pelos ativos presentes na figura 19.

Figura 19 – Gráficos do  $\rho_{DCCA}$  e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo EMBR3. O eixo das abscissas está em dias.



Fonte: Frederico Frias, 2021

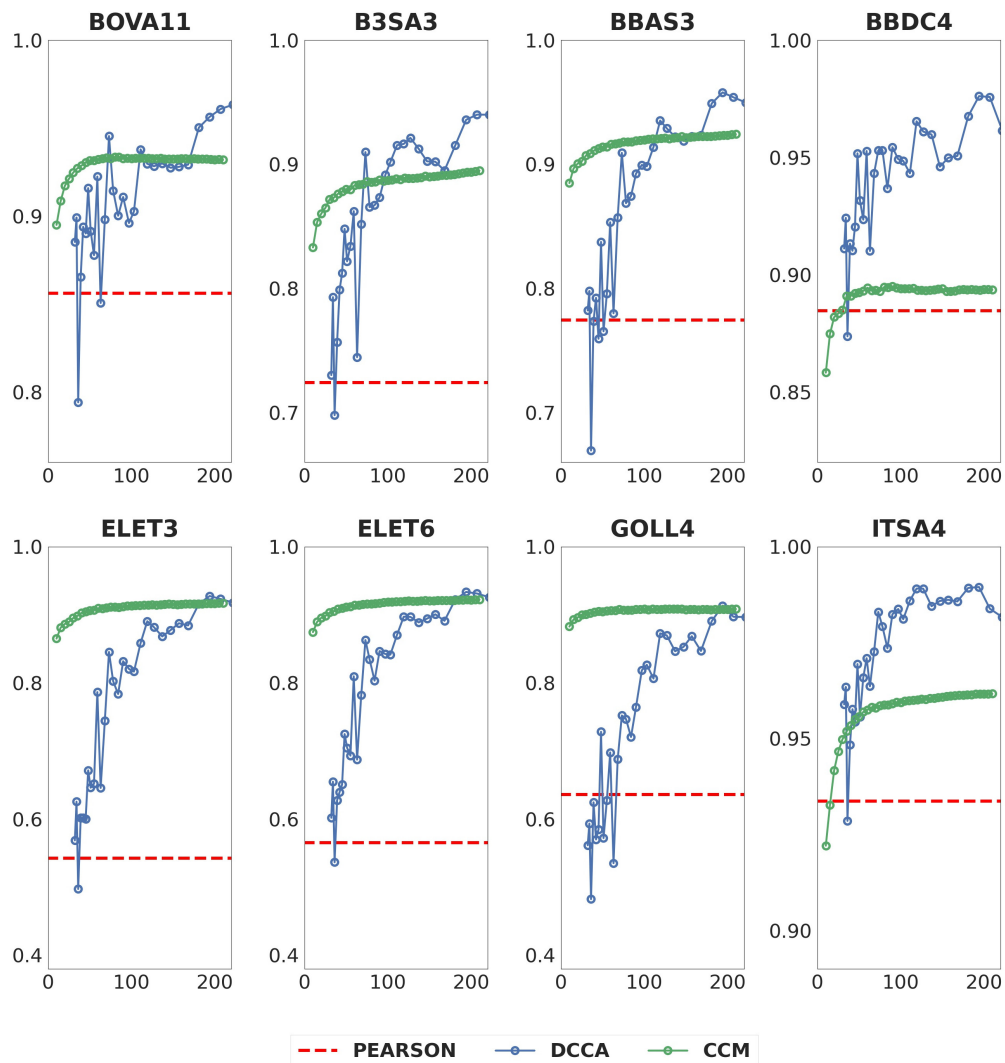
Como pode ser observado na tabela 8, os valores obtidos nos três métodos não são muito expressivos. Esta ação é ilíquida e com pouca representatividade no Ibovespa (cerca de 1.12%), assim como a CYRE3. Entretanto, sua diferença em relação ao ativo anterior é que a empresa desta ação faz parte do segmento aeronáutico e de defesa, segmento bem específico, com características singulares em relação aos demais presentes no índice. Possivelmente, esta particularidade faz com que o preço deste ativo não sofra tanta influência por parte das demais ações do Ibovespa.

Tabela 8 – Valores do  $\rho_{DCCA}$  e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo EMBR3.

Ação	Pearson	$\rho_{DCCA}$	CCM
BTOW3	0.1406	0.2545	0.3385
FIBR3	-0.0299	0.2999	0.3213
KLBN11	0.1247	0.2787	0.2574
MRFG3	0.2313	0.5255	0.5282
NATU3	0.2119	0.2581	0.5588
SUZB3	0.0101	0.3513	0.3426
TIMP3	0.2645	0.5948	0.6419
VIVT4	0.3039	0.3523	0.6006

Fonte: Frederico Frias, 2021

Figura 20 – Gráficos do  $\rho_{DCCA}$  e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo ITUB4. O eixo das abscissas está em dias.



Fonte: Frederico Frias, 2021

No caso da ação ITUB4, das oito ações que compõem a carteira (ver figura 20), quatro pertencem ao mesmo setor do ativo alvo, que é o setor financeiro, sendo uma delas a ITSA4, *holding* do conglomerado que abarca o banco Itaú. Além disso, o BOVA11 também faz parte do portfólio deste ativo com uma forte correlação com o mesmo, como observado na tabela 9. Esta ação é a que possui maior peso no ETF do índice (acima de 11%).

A princípio, seria de se esperar o contrário: a ação alvo causando o BOVA11. Entretanto, é possível que a relação causal entre ambos seja recíproca. Uma vez que o ETF é negociável, diferentemente do Ibovespa, pode existir um cenário em que, ao influenciar a dinâmica do BOVA11, a ação ITUB4 também seja influenciada pelo mesmo.

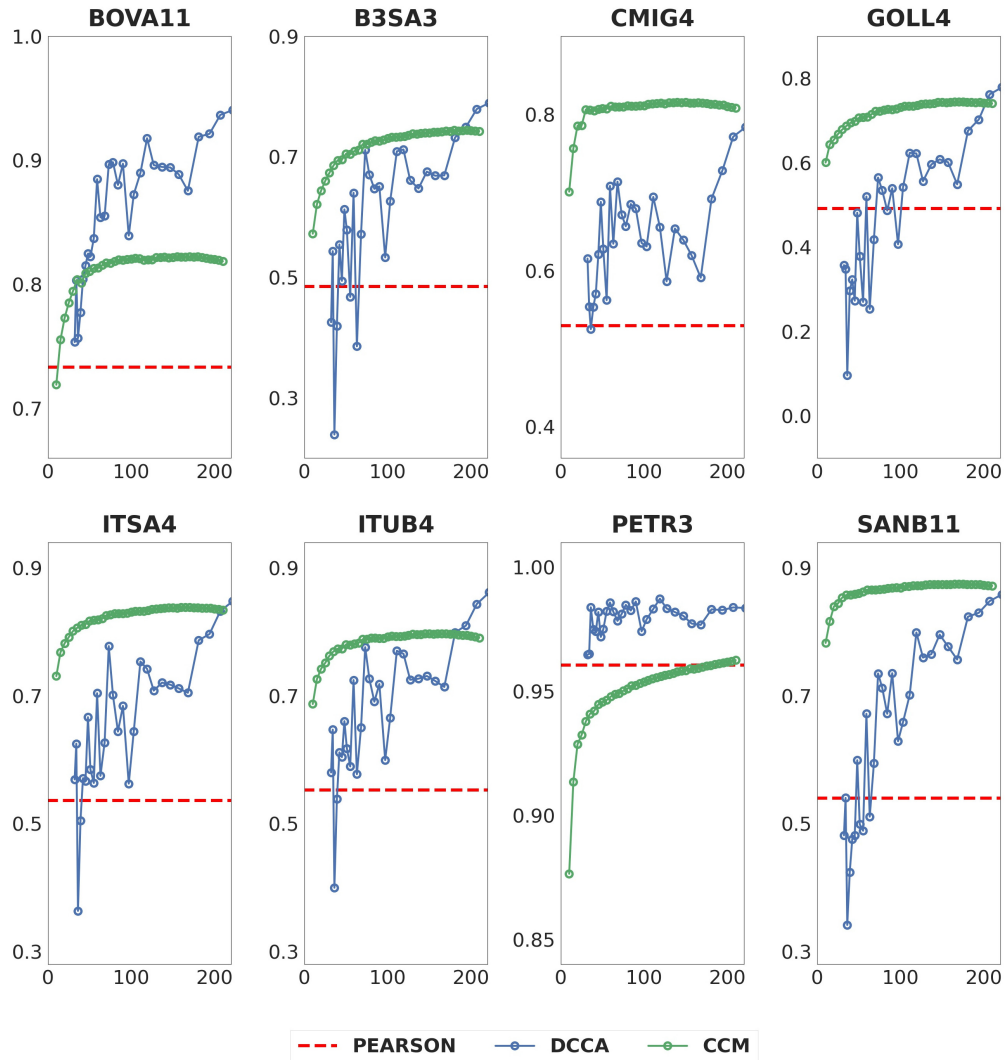
Tabela 9 – Valores do  $\rho_{DCCA}$  e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo ITUB4.

Ação	Pearson	$\rho_{DCCA}$	CCM
BOVA11	0.8562	0.9604	0.9319
B3SA3	0.7244	0.9411	0.8949
BBAS3	0.7744	0.9601	0.9242
BBDC4	0.8847	0.9706	0.8936
ELET3	0.5428	0.9350	0.9174
ELET6	0.5655	0.9401	0.9225
GOLL4	0.6359	0.9233	0.9085
ITSA4	0.9337	0.9864	0.9617

Fonte: Frederico Frias, 2021

A ação PETR4 também possui o ETF em seu portfólio, provavelmente, pela mesma razão elucidada anteriormente. Como seu peso no BOVA11 é menor do que o do ativo ITUB4 (cerca de 7%), sendo o quarto de maior percentual global, os valores de correlação serão, naturalmente, inferiores ao da ação do segmento bancário em todos os métodos. Esta premissa pode ser constatada ao analisar os dados das tabelas 9 e 10. Outra evidência direta, é a presença da PETR3 na carteira deste ativo. Esta é a ação ordinária da empresa Petrobrás que, evidentemente, terá uma forte relação de dependência com a sua ação preferencial. Ao analisar o coeficiente  $\rho_{DCCA}$  deste mesmo ativo na figura 21, nota-se que o mesmo possui uma variação aproximadamente constante ao longo do período analisado, ou seja, pode-se inferir que a ação PETR3 certamente apresentará uma correlação quase perfeita com a ação PETR4 independentemente do intervalo de tempo estudado. Além disso, observa-se na tabela 10 que os valores deste ativo são os mais altos obtidos em comparação com os demais ativos presentes no portfólio da PETR4, como era de se esperar.

Figura 21 – Gráficos do  $\rho_{DCCA}$  e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo PETR4. O eixo das abscissas está em dias.



Fonte: Frederico Frias, 2021

Tabela 10 – Valores do  $\rho_{DCCA}$  e do CCM dos ativos com maior correlação e que apresentaram uma relação causal com a ação alvo PETR4.

Ação	Pearson	$\rho_{DCCA}$	CCM
BOVA11	0.7333	0.9196	0.8186
B3SA3	0.4847	0.7817	0.7425
CMIG4	0.5296	0.7702	0.8083
GOLL4	0.4914	0.7484	0.7406
ITSA4	0.5359	0.8074	0.8344
ITUB4	0.5528	0.8165	0.7908
PETR3	0.9605	0.9758	0.9626
SANB11	0.5399	0.8547	0.8721

Fonte: Frederico Frias, 2021

Ao inferir causalidade, é importante investigar se a mesma poderia ser gerada na ausência de uma relação causal real, ou seja, é necessário mensurar a importância da relação causal entre as séries temporais das ações integrantes de cada um dos portfólios acima descritos com a ação alvo correspondente. A causalidade falsamente inferida pode às vezes ser mitigada por testes de dados chamados de *surrogates*. Esta técnica cria um conjunto de dados sintéticos de controle que retêm certos aspectos do conjunto de dados original, mas seriam improváveis se houvesse uma relação causal. O teste de causalidade é repetido nos dados oriundos do método de *surrogate* para constatar se os mesmos fornecem um sinal tão forte quanto os originais. Neste caso, provavelmente existe uma relação causal.

O ideal é que o *surrogate* preserve aspectos importantes dos dados que não dependem de uma relação causal ao mesmo tempo em que remove os aspectos que dependem. Neste trabalho, foi utilizado o teste de fase aleatória. Esta técnica consiste no emprego da transformada de Fourier para representar uma série temporal como uma soma de ondas senoidais, deslocando aleatoriamente cada componente dessas ondas no tempo. Após isso, os componentes deslocados de fase são somados. Este procedimento preserva o espectro de potência em vez da função de massa de probabilidade de uma série temporal, assumindo que o mesmo não depende da presença de uma relação causal [51].

O intuito é propor a hipótese nula ( $H_0$ ) de que a ação integrante da carteira de ativos não está causando a ação alvo correspondente. Para isso, a técnica de randomização de fase é empregada para gerar os *surrogates* condizentes à hipótese nula. Em suma, a série temporal original é transformada no domínio de Fourier, randomiza-se a fase em cada frequência sem perturbar a magnitude e, por fim, calcula-se a transformada inversa de Fourier. A fase possui as informações sobre a relação temporal de uma determinada variável com todas as outras. Na randomização, toda a sequência de fases é destruída.

Sendo assim, a habilidade de predição do método do mapeamento cruzado convergente dos dados originais deve ser maior do que a habilidade dos dados gerados pelo teste de *surrogate*. Para demonstrar esta hipótese, o CCM foi calculado repetidamente (100 vezes), empregando os dados gerados pela técnica da fase aleatória. O valor- $p$  de significância foi mensurado como a fração dos valores de habilidade de predição do mapeamento cruzado convergente obtidos com as séries temporais oriundas do *surrogate* que eram iguais ou maiores do que o valor de habilidade da série original. Uma ligação causal é comprovada neste critério se o valor- $p$  for menor ou igual ao nível de significância de 0.05.

Tabela 11 – Valores- $p$  de significância medidos pelo teste de fase aleatória para os ativos integrantes da carteira de cada ação alvo.

CYRE3		EMBR3		ITUB4		PETR4	
BBDC3	0.0099	BTOW3	0.4752	BOVA11	0.0099	BOVA11	0.0099
BBDC4	0.0099	FIBR3	0.8217	B3SA3	0.0099	B3SA3	0.0297
CVCB3	0.0297	KLBN11	0.8613	BBAS3	0.0099	CMIG4	0.0198
EQTL3	0.0099	MRFG3	0.2079	BBDC4	0.0099	GOLL4	0.0693
GOLL4	0.0099	NATU3	0.1089	ELET3	0.0099	ITSA4	0.0099
IGTA3	0.0198	SUZB3	0.8415	ELET6	0.0099	ITUB4	0.0198
MULT3	0.0693	TIMP3	0.1485	GOLL4	0.0099	PETR3	0.0099
SBSP3	0.0198	VIVT4	0.1881	ITSA4	0.0099	SANB11	0.0099

Fonte: Frederico Frias, 2021

Conforme observa-se na tabela 11, todos os ativos da carteira do ITUB4 apresentaram valores- $p$  abaixo do limite estipulado. Com isso, a hipótese nula de que a ação do portfólio não está causando a ação alvo foi rejeitada para todas as ações desta carteira. Em contrapartida, o inverso ocorre para o ativo EMBR3, em que todos os ativos tiveram valores- $p$  acima do nível de significância, implicando que  $H_0$  é verdadeira, isto é, nenhuma das ações do portfólio causa a ação modelada. Nos casos dos ativos CYRE3 e PETR4, não houve rejeição da hipótese nula apenas para a ação MULT3 e GOLL4 respectivamente. É importante mencionar que há a possibilidade de ocorrência de erros do tipo I<sup>3</sup> e/ou do tipo II<sup>4</sup> dentre os resultados obtidos pelo teste de significância. Possivelmente, os resultados obtidos neste teste, que confirma a existência ou não de causalidade entre o ativo modelado com os ativos constituintes da sua carteira, serão postos em prova pelo algoritmo utilizado na predição da tendência do movimento do preço. Afinal, quanto mais forte for a dependência causal entre duas séries temporais, mais fácil deveria ser prever a tendência de uma baseado no comportamento da outra.

No próximo capítulo, será aplicado um algoritmo de aprendizado de máquina para caracterizar, de forma binária (descida ou subida), a tendência do movimento do preço de cada uma das ações aqui estudadas. Esta técnica utilizará como referência na sua classificação os ativos, selecionados pelos métodos descritos no presente capítulo, que apresentaram forte dependência com a ação a ser modelada. Juntamente com este método, é empregado um algoritmo genético com o intuito de melhorar a acurácia da predição da dinâmica da ação.

<sup>3</sup> Rejeita-se  $H_0$ , sendo que  $H_0$  é verdadeira.

<sup>4</sup> Não rejeita-se  $H_0$ , sendo que  $H_0$  é falsa.



## 4 Modelagem da tendência de séries temporais financeiras

O presente capítulo descreve o modelo utilizado na predição do movimento do preço de uma dada ação. Para tal, utilizou-se um algoritmo de aprendizado de máquina supervisionado que gera uma classificação binária (tendência de queda ou de subida) a partir de um conjunto de parâmetros de entrada, composto por certos indicadores técnicos das ações que possuem forte correlação com aquela que se está analisando. A principal vantagem deste método está na capacidade de modelar processos complexos, possivelmente não lineares, sem assumir nenhum conhecimento prévio sobre o processo de geração de dados subjacentes.

No entanto, séries temporais financeiras frequentemente contêm padrões bastante complexos, o que dificulta este tipo de algoritmo a alcançar o melhor desempenho em todas as situações. Com o intuito de refinar a acurácia das predições, tem-se tornado uma prática comum combinar diferentes métodos. Os resultados destes modelos híbridos possuem, em geral, uma performance superior em relação ao modelo individual. A principal razão disto ocorrer é que esta combinação promove uma complementaridade ao fazer uso do recurso exclusivo de cada técnica na captura de padrões dos conjuntos de dados. Para aumentar a acurácia do algoritmo de aprendizado de máquina, empregou-se, neste trabalho, um algoritmo genético.

### 4.1 Máquina de vetor de suporte

Máquina de vetor de suporte, do inglês, *support vector machine* (SVM), é uma técnica de aprendizado de máquina desenvolvida por Vapnik [52]. Este método tem como alicerce a teoria de aprendizado estatístico, sendo empregado na predição de séries temporais financeiras [9–11, 53]. Seu conceito é baseado no princípio de minimização do risco estrutural, do inglês, *structural risk minimization* (SRM), com desempenho de generalização superior ao habitual princípio utilizado em redes neurais convencionais, o da minimização do risco empírico, do inglês, *empirical risk minimization* (ERM).

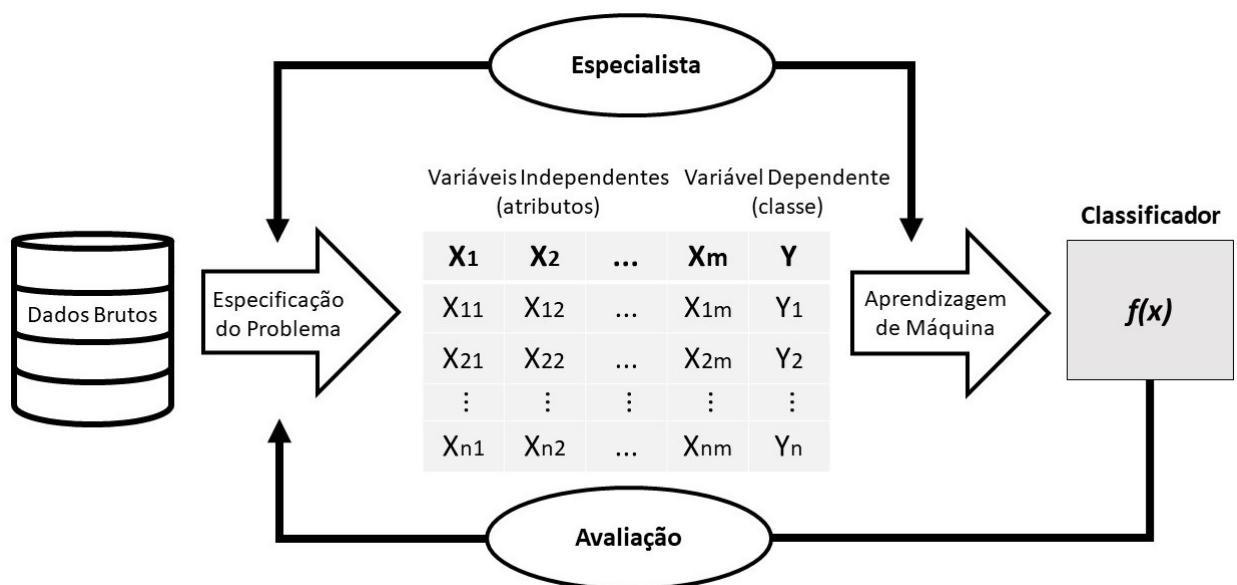
As técnicas de aprendizado de máquina, do inglês, *machine learning*, utilizam da indução com o intuito de obter soluções genéricas a partir de um conjunto de exemplos. Os principais tipos de aprendizado são o supervisionado e o não supervisionado. No

supervisionado, o algoritmo, através de uma coleção de exemplos do tipo entrada e saída esperada, interpreta a informação de modo a gerar saídas precisas ao receber entradas não apresentadas anteriormente. Já no não supervisionado, a interpretação ocorre mediante uma medida de qualidade, sem que hajam exemplos classificados [54].

O SVM é um algoritmo de aprendizado supervisionado. Neste tipo de aprendizado, ocorre um processo de inferência chamado de treinamento, onde é gerado um classificador, função  $f$  que, dada uma entrada de um dado  $\mathbf{x}$ , produz uma previsão de um rótulo  $y$ . Este classificador é aquele, entre todas as hipóteses testadas pelo algoritmo, com melhor eficácia em caracterizar o conjunto de exemplos apresentados no treinamento. A partir desta indução, o algoritmo tem de ser capaz de prever de forma precisa o rótulo de novos dados.

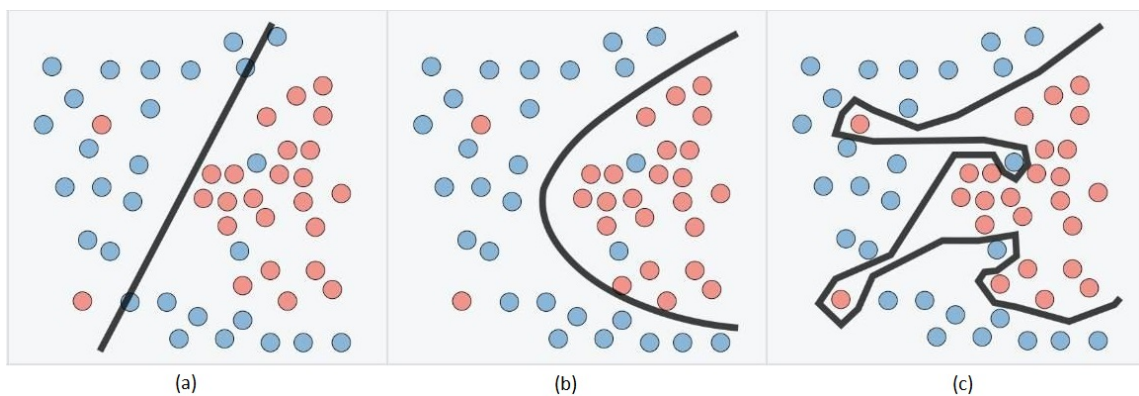
Cada dado  $\mathbf{x}$  possui um vetor de atributos ou características que apresenta um aspecto específico do exemplo. Os rótulos podem ser discretos ou contínuos. No caso dos discretos, tem-se um problema de classificação que pode ser multiclasse, quando há mais de dois rótulos envolvidos, ou binário, quando só existem dois rótulos. Na figura 22, é representado um esquema simplificado do processo de inferência de um classificador em um aprendizado supervisionado, onde, a partir de um grupo de  $n$  dados, cada um com  $m$  atributos, e suas respectivas classes  $y$ , é criado um classificador  $f(\mathbf{x})$ .

Figura 22 – Esquema simplificado do processo de indução de um classificador em um aprendizado supervisionado.



Após a geração do classificador, deve ser realizada uma previsão utilizando um grupo de dados, denominado de teste, apartado do conjunto de dados do treinamento, de modo a estimar o grau de efetividade do conceito aprendido [55]. Se o algoritmo for demasiado especializado com elevada suscetibilidade a cometer erros ao ser deparado com novos dados, ocorrerá um superajustamento (figura 23c). Caso essa baixa eficácia ocorra ainda no processo de treinamento, haverá um subajustamento (figura 23a). Sendo assim, o classificador deve ser estimado o mais genérico possível (figura 23b), afim de alcançar o ajuste mais apropriado que rotule corretamente grande parte dos dados além de minimizar a influência de eventuais ruídos na classificação.

Figura 23 – Exemplificação do treinamento de uma classificação binária com: (a) subajustamento (*underfitting*); (b) ajustamento apropriado; (c) superajustamento (*overfitting*).

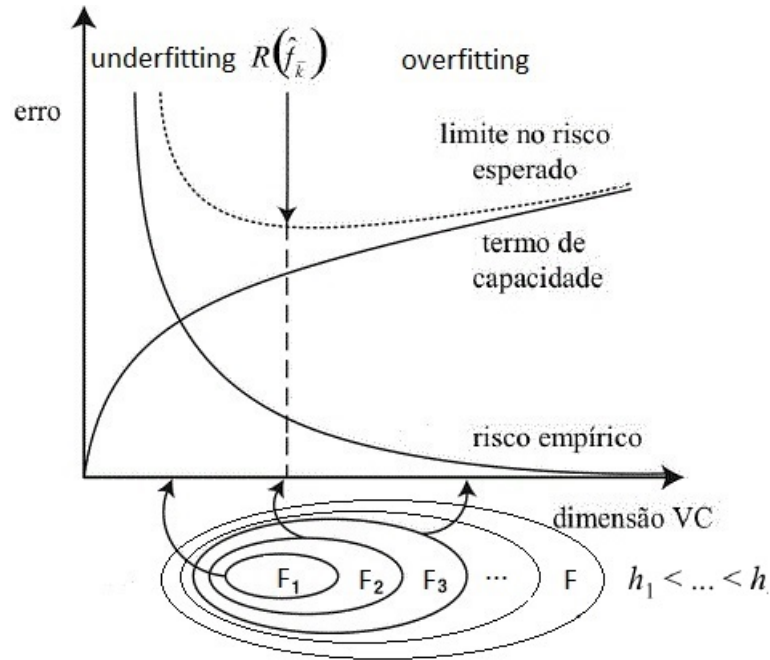


Fonte: Frederico Frias, 2021

#### 4.1.1 Princípio de minimização do risco estrutural

Para gerar um classificador  $\hat{f}$  que minimize o erro sobre os dados de treinamento, denominado de risco empírico, como também leve a um menor erro sobre os dados de teste, o método SVM, conforme mencionado anteriormente, faz uso do princípio de minimização do risco estrutural. Este princípio envolve a minimização de um limite superior sobre o erro de generalização. Ao contrário da minimização do risco empírico, que minimiza somente o risco dos dados de treinamento, o SRM apresenta uma maior habilidade de generalização face a dados ainda não observados.

Figura 24 – Ilustração do princípio de minimização do risco estrutural (SRM).



Fonte: Frederico Frias, 2021

A figura 24 mostra uma ilustração deste princípio. O SRM restringe o grupo de funções da qual  $\hat{f}$  é deduzida, impondo um limite no risco esperado (linha tracejada da figura acima), ao considerar a complexidade (ou capacidade) das mesmas. Com isso, o risco esperado de um classificador  $f$  para dados de teste pode ser obtido, com uma garantia de probabilidade  $1 - \theta$ , por:

$$R(f) \leq R_{emp}(f) + \sqrt{\frac{h[\ln(2n/h) + 1] - \ln(\theta/4)}{n}}, \quad (32)$$

onde  $\theta \in [0, 1]$ ,  $h$  é a dimensão de Vapnik-Chervonenkis (VC),  $n$  é a quantidade de exemplos do treinamento, a parcela da raiz é o termo de capacidade e  $R_{emp}$  é o risco empírico, definido a seguir:

$$R_{emp}(f) = \frac{1}{n} \sum_{i=1}^n c(f(x_i), y_i), \quad (33)$$

sendo  $c(f(\mathbf{x}), y)$  uma função de custo que compara a previsão  $f(\mathbf{x})$  com a saída desejada  $y$ .

A dimensão VC mede a capacidade do grupo de funções  $F$  à qual  $f$  pertence [56]. Quanto maior for seu valor, maior será a complexidade das funções de classificação que podem ser inferidas a partir de  $F$ . Em outras palavras, a dimensão VC é o número máximo de exemplos de treinamento que pode ser aprendido pela máquina sem erro, para todas as rotulações possíveis das funções de classificação.

A inequação 32 mostra que o risco esperado pode ser minimizado ao escolher um classificador  $\hat{f}$  que diminua o risco empírico e que faça parte de um grupo de funções  $F$  com baixa dimensão  $VC$ . Para que isso ocorra, é necessário dividir  $F$  em subconjuntos de funções com dimensão  $VC$  crescente. Seja  $\hat{f}_k$  o classificador com menor risco empírico de um subconjunto específico  $F_k$ . A medida em que  $k$  aumenta, o risco empírico desse classificador diminui, mas sua capacidade cresce. Deve existir um  $\bar{k}$  tal que o risco esperado seja mínimo ( $R(\hat{f}_{\bar{k}})$  expresso na figura 24), isto é, a soma do risco empírico e do termo de capacidade (lado direito da inequação 32) seja o menor possível. Embora este conceito tenha sido útil no SRM, existe o problema da dimensão  $VC$  ser não trivial de computar ou até mesmo de ter um valor infinito.

Uma alternativa a esse método é limitar o erro esperado por uma margem de confiança, vinculada à distância da fronteira de decisão induzida. Esta margem determina o quão bem duas classes podem ser separadas e é definida por:

$$\rho = \min_i [y_i f(x_i)]. \quad (34)$$

Com isso, o erro marginal de uma função  $f$  é dado pela equação 35. Esse erro fornece a proporção de exemplos de treinamento cuja margem de confiança é inferior a uma determinada constante  $\rho > 0$ :

$$R_\rho(f) = \frac{1}{n} \sum_{i=1}^n I(y_i f(x_i) < \rho), \quad (35)$$

onde  $I(p) = 1$  se  $p$  é verdadeiro e  $I(p) = 0$ , caso contrário.

Adotando esta nova abordagem, o risco esperado de um classificador  $f$ , com uma garantia de probabilidade  $1 - \theta$ , é dado pelo limite da soma do erro marginal do conjunto de treinamento com um termo de capacidade, conforme definido abaixo:

$$R(f) \leq R_\rho(f) + \sqrt{\frac{c}{n} \left[ \frac{R^2}{\rho^2} \log^2 \left( \frac{n}{\rho} \right) + \log \left( \frac{1}{\theta} \right) \right]}, \quad (36)$$

sendo  $c$  uma constante,  $R > 0$  e  $\|x\| \leq R$ . Um  $\rho$  elevado implica um termo de capacidade menor. É possível observar pela inequação 36 que, para alcançar o risco esperado mínimo, deve-se buscar um hiperplano, denominado por ótimo, com a maior margem  $\rho$  e que gere poucos erros marginais.

### 4.1.2 Hiperplano ótimo

Seja um problema de classificação binária de um conjunto de vetores de treinamento  $T \in \mathbb{R}^n$  com  $N$  dados,  $T = \{(\mathbf{x}_i, y_i), i = 1, 2, \dots, N\}$ , onde  $\mathbf{x}_i \in \mathbf{x}$  é o  $i$ -ésimo vetor de dados de entrada e  $y_i \in \{-1, +1\}$ , os seus respectivos rótulos. O grupo  $T$  é linearmente classificável se os dados das classes  $-1$  e  $+1$  puderem ser separados por um hiperplano dado por:

$$f(\mathbf{x}) = \mathbf{w}^T \cdot \mathbf{x} + b = 0, \quad (37)$$

onde  $\mathbf{w} \in \mathbb{R}^n$  é um vetor normal ao hiperplano descrito,  $\cdot$  denota o produto escalar e  $\frac{b}{\|\mathbf{x}\|}$  é a distância do hiperplano à origem, com  $b \in \mathbb{R}$ .

O hiperplano  $H$ , definido na equação 37, divide o espaço dos dados  $\mathbf{x}$  em duas regiões:  $\mathbf{w}^T \cdot \mathbf{x} + b > 0$  e  $\mathbf{w}^T \cdot \mathbf{x} + b < 0$ . Uma função sinal  $g(\mathbf{x}) = \text{sgn}(f(\mathbf{x})) = \text{sgn}(\mathbf{w}^T \cdot \mathbf{x} + b)$  pode ser empregada na obtenção das classificações:

$$g(\mathbf{x}) = \text{sgn}(f(\mathbf{x})) = \begin{cases} +1, & \text{se } \mathbf{w}^T \cdot \mathbf{x} + b > 0 \\ -1, & \text{se } \mathbf{w}^T \cdot \mathbf{x} + b < 0 \end{cases}. \quad (38)$$

Logo, um conjunto de treinamento é linearmente separável se for possível determinar pelo menos um par  $(\mathbf{w}, b)$  tal que a função sinal  $g(x)$  consiga classificar corretamente todos os exemplos contidos neste grupo. Um número infinito de hiperplanos equivalentes pode ser obtido ao multiplicar  $\mathbf{w}$  e  $b$  por uma mesma constante. Sendo assim, o SVM define dois hiperplanos paralelos,  $H_-$  e  $H_+$ , de modo que os rótulos mais próximos ao hiperplano  $H$  satisfaçam a equação objetivo:

$$y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) - 1 \geq 0, \quad \forall (\mathbf{x}_i, y_i) \in T. \quad (39)$$

A distância geométrica entre um dado  $\mathbf{x}_i$  e o hiperplano  $H$  é calculada por:

$$d_i(\mathbf{w}, \mathbf{x}_i, b) = \frac{|\mathbf{w}^T \cdot \mathbf{x}_i + b|}{\|\mathbf{w}\|} = \frac{y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b)}{\|\mathbf{w}\|}, \quad (40)$$

em que  $\|\cdot\|$  representa a norma de um vetor. Levando em conta a restrição imposta pela inequação 39, obtém-se:

$$d_i(\mathbf{w}, \mathbf{x}_i, b) \geq \frac{1}{\|\mathbf{w}\|}. \quad (41)$$

A inequação 41 identifica o limite inferior da distância entre os hiperplanos  $H_-$  e  $H_+$  com o hiperplano de separação  $H$ . Portanto, a distância mínima, definida como

margem geométrica do classificador linear, entre o hiperplano  $H$  e o conjunto de dados de treinamento é:

$$d_- = d_+ = \frac{1}{\|\mathbf{w}\|}. \quad (42)$$

Conseqüentemente, a distância entre ambos os hiperplanos é obtida pela seguinte relação:

$$d = d_- + d_+ = \frac{2}{\|\mathbf{w}\|}. \quad (43)$$

Como o objetivo do algoritmo SVM é maximizar a margem de separação dos dados, então deve-se minimizar  $\|\mathbf{w}\|$ , sujeito à restrição imposta pela inequação 39, através de:

$$\epsilon(\mathbf{w}, b) = \min_{\mathbf{w}, b} \left[ \frac{1}{2} \|\mathbf{w}\|^2 \right]. \quad (44)$$

O problema de otimização expresso em 44 é denominado programação quadrática e pode ser resolvido com o método dos multiplicadores de Lagrange:

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i [y_i (\mathbf{w}^T \cdot \mathbf{x}_i + b) - 1], \quad (45)$$

sendo o multiplicador de Lagrange  $\alpha > 0$ .

A solução é dada pelo ponto de sela ao igualar as derivadas parciais de  $\mathcal{L}$  a zero:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial w} = 0 = w - \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \\ \frac{\partial \mathcal{L}}{\partial b} = 0 = - \sum_{i=1}^N \alpha_i y_i \end{cases} \Rightarrow \begin{cases} \alpha_i \geq 0, \quad i = 1, 2, \dots, N \\ \sum_{i=1}^N \alpha_i y_i = 0 \end{cases}. \quad (46)$$

Substituindo o conjunto de equações 46 (lado esquerdo) em 45, tem-se o seguinte problema de otimização, sujeito às restrições 46 (lado direito):

$$\max_{\alpha} \left[ \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)) \right]. \quad (47)$$

O objetivo é determinar os valores ótimos do par  $(\mathbf{w}, b)$ , representados por  $(\mathbf{w}^*, b^*)$ . Essa formulação é denominada forma dual, enquanto o problema original é referenciado como forma primal. A forma dual é mais interessante, pois apresenta restrições mais simples, tem a representação do problema de otimização em termos de produtos internos, útil no caso de SVMs não lineares, e utiliza apenas os dados de treinamento e os seus rótulos.

Seja  $\alpha_i^*$  solução do problema de otimização 47. Então,  $\mathbf{w}^*$  pode ser obtido pela relação:

$$w^* = \sum_{i=1}^N \alpha_i^* y_i \mathbf{x}_i. \quad (48)$$

O parâmetro  $b$  pode ser obtido utilizando as equações de Karush-Kuhn-Tucker (KKT), provenientes da teoria de otimização com restrições:

$$\alpha_i^*(y_i[(\mathbf{w}^*)^T \cdot \mathbf{x}_i + b^*] - 1) = 0, \quad \forall i = 1, \dots, N. \quad (49)$$

Pela equação 49, tem-se que  $\alpha_i^* \neq 0$  somente para os dados que se encontram sobre os hiperplanos  $H_-$  e  $H_+$ , sendo os exemplos que se situam mais próximos ao hiperplano separador, exatamente sobre as margens. Sendo assim, os dados que possuem  $\alpha_i^* > 0$  são denominados vetores de suporte, do inglês, *support vectors* (SVs). Como consequência, o hiperplano ótimo é determinado unicamente pelos mesmos, o que faz desses dados os mais informativos do conjunto de treinamento.

Dado um vetor de suporte  $\mathbf{x}_i$ , podemos obter  $b^*$  por meio da condição de KKT:

$$b^* = y_i - [(\mathbf{w}^*)^T \cdot \mathbf{x}_i]. \quad (50)$$

Por último, o classificador  $g(x)$  é dado por:

$$g(\mathbf{x}) = \text{sgn}(f(\mathbf{x})) = \text{sgn}\left(\sum_{\mathbf{x}_i \in SV} y_i \alpha_i^* \mathbf{x}_i \cdot \mathbf{x} + b^*\right). \quad (51)$$

A equação 51 representa o hiperplano que separa os dados com maior margem, considerado aquele com melhor capacidade de generalização.

Contudo, em certas situações reais, os dados não são linearmente separáveis. Isto pode acontecer devido a fatores como a presença de ruídos e *outliers* nos dados ou até mesmo à própria natureza do problema. Sendo assim, a restrição dada pela inequação 39 deve ser estendida para lidar com conjuntos de treinamento mais gerais:

$$y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \forall (\mathbf{x}_i, y_i) \in T, \quad (52)$$

sendo  $\xi_i \geq 0$  a variável de folga, variável essa que permite lidar com ruídos e *outliers* nos dados, isto é, suaviza as margens do classificador linear, permitindo que alguns dados permaneçam entre os hiperplanos  $H_-$  e  $H_+$  ( $0 < \xi_i \leq 1$ ) bem como a ocorrência de alguns erros de classificação ( $\xi_i > 1$ ).

Para maximizar a margem de separação dos dados, deve-se minimizar  $\epsilon(\mathbf{w}, b, \xi)$ , sujeito à restrição imposta pela inequação 52, através de:

$$\epsilon(\mathbf{w}, b, \xi) = \min_{\mathbf{w}, b, \xi} \left[ \frac{1}{2} \|\mathbf{w}\|^2 + C \left( \sum_{i=1}^N \xi_i \right) \right], \quad (53)$$



onde o termo da somatória é um limite superior no número de erros de treinamento e  $C$  é um parâmetro constante positivo usado para controlar a compensação entre o erro de treinamento e a margem, impondo um peso à minimização dos erros no conjunto de treinamento em relação à minimização da complexidade do modelo.

Novamente, o problema de otimização gerado é quadrático, com as restrições lineares dadas por 52. Sua solução é semelhante à apresentada anteriormente, com a introdução de uma função Lagrangiana:

$$\mathcal{L}(\mathbf{w}, b, \alpha, \mu, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \left( \sum_{i=1}^N \xi_i \right) - \sum_{i=1}^N \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^N \mu_i \xi_i, \quad (54)$$

sendo os multiplicadores de Lagrange  $\alpha, \mu > 0$ .

O problema dual é idêntico ao apresentado para margens rígidas (equação 47), exceto pela restrição em  $\alpha$  ( $0 \leq \alpha_i \leq C, \forall i = 1, \dots, N$ ). O vetor  $\mathbf{w}^*$  continua sendo determinado por 48. O parâmetro de folga pode ser determinado pela seguinte relação:

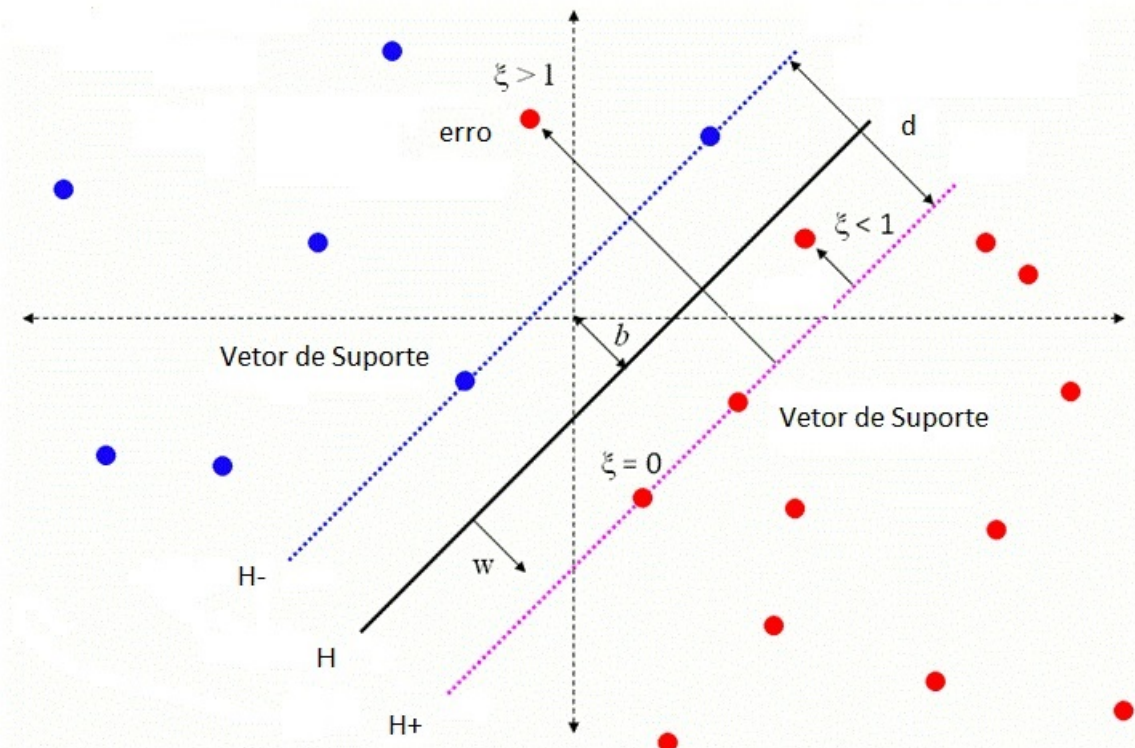
$$\xi_i^* = \max \left( 0, 1 - y_i \sum_{j=1}^N \alpha_j^* y_j (\mathbf{x}_i \cdot \mathbf{x}_j) + b^* \right). \quad (55)$$

A variável  $b^*$  é obtida por  $\alpha^*$  e das seguintes condições KKT:

$$\begin{cases} \alpha_i^* (y_i [(\mathbf{w}^*)^T \cdot \mathbf{x}_i + b^*] - 1 + \xi_i) = 0 \\ (C - \alpha_i^*) \xi_i = 0 \end{cases}. \quad (56)$$

Assim como nas SVMs de margens rígidas, os pontos para os quais  $\alpha_i^* > 0$  são chamados de vetores de suporte. Entretanto, para  $\alpha_i^* < C$ , tem-se que  $\xi_i^* = 0$  (segunda equação de 56), o que implica que os mesmos estão sobre as margens. Os vetores de suporte com  $\alpha_i^* = C$  podem caracterizar três casos: erros ( $\xi_i^* > 1$ ); pontos classificados corretamente, mas que estão entre as margens ( $0 < \xi_i^* \leq 1$ ); pontos sobre as margens ( $\xi_i^* = 0$ ). Na Figura 25, são ilustrados os possíveis tipos de vetores de suporte. Todos os pontos classificados corretamente que estão fora das margens têm  $\xi_i^* = 0$  e  $\alpha_i^* = 0$ .

Figura 25 – Ilustração do hiperplano ótimo e dos vetores de suporte numa classificação binária com a presença de erros.



Fonte: Frederico Frias, 2021

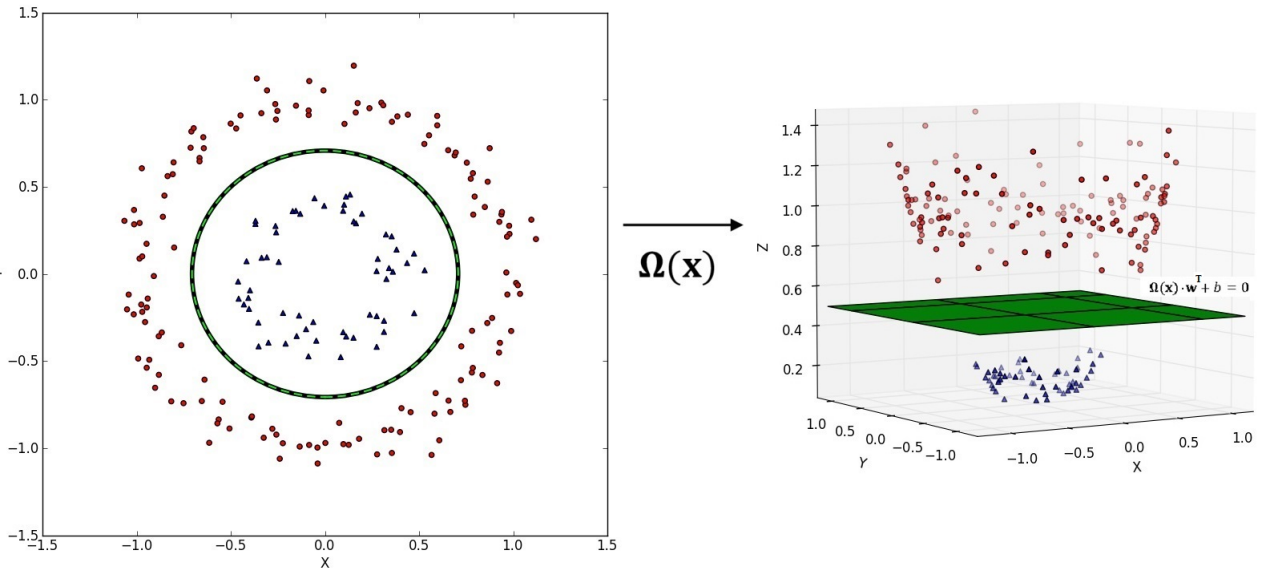
#### 4.1.3 SVM não linear

Se o conjunto de dados não for linearmente classificável, como é o caso de estudo deste trabalho, não será possível separar os rótulos por um hiperplano ótimo. Para uma divisão binária arbitrária, é necessário recorrer a funções de complexidade superior, ditas funções de *kernel*. O SVM não linear mapeia o conjunto de treinamento do espaço original, também chamado de espaço de entradas, para um novo espaço de maior dimensão, referenciado como espaço de características, do inglês, *feature space*.

Esta transformação é fundada no teorema de Cover que afirma que um problema não linear tem maior probabilidade de ser linearmente separável em um espaço de mais alta dimensionalidade. Com isso, o mapeamento  $\Omega : \mathbb{R}^n \rightarrow \mathbb{R}^m$  possibilita a classificação linear por um hiperplano ótimo de um conjunto de dados não linear em  $\mathbb{R}^n$  em um novo espaço  $\mathbb{R}^m$  de dimensão superior. A figura 26 mostra um exemplo de um conjunto de treinamento constituído de rótulos binários (círculos e triângulos) não lineares que passa a

ser linearmente separável após a aplicação de uma função  $\Omega$  que transforma o espaço de entradas ( $\mathbb{R}^2$ ) em um espaço de características ( $\mathbb{R}^3$ ).

Figura 26 – Exemplificação do SVM não linear: fronteira não linear no espaço de entradas (figura à esquerda); fronteira linear no espaço de características (figura à direita).



Fonte: Frederico Frias, 2021

Em suma, os dados são mapeados para um espaço de maior dimensão utilizando  $\Omega$  e, logo após, aplica-se a SVM linear com margens suaves, permitindo lidar com ruídos e *outliers* presentes nos dados, sobre este novo espaço, de modo a encontrar o hiperplano com maior margem de separação afim de garantir uma boa generalização.

O problema de otimização é dado por:

$$\max_{\alpha} \left[ \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left( \alpha_i \alpha_j y_i y_j (\Omega(\mathbf{x}_i) \cdot \Omega(\mathbf{x}_j)) \right) \right], \quad (57)$$

com as restrições:

$$\begin{cases} 0 \leq \alpha_i \leq C, \quad \forall i = 1, \dots, N \\ \sum_{i=1}^N \alpha_i y_i = 0 \end{cases}. \quad (58)$$

Com isso, o classificador  $g(x)$  pode ser obtido da relação:

$$g(\mathbf{x}) = \text{sgn}(f(\mathbf{x})) = \text{sgn} \left( \sum_{\mathbf{x}_i \in SV} y_i \alpha_i^* \Omega(\mathbf{x}_i) \cdot \Omega(\mathbf{x}) + b^* \right). \quad (59)$$

A dimensão do espaço de características pode ser muito alta, até mesmo infinita, tornando a computação de  $\Omega$  extremamente custosa e inviável. Entretanto, pela equação

59, é possível observar que a única informação necessária sobre o mapeamento é de como realizar o cálculo de produtos escalares, pois tem-se sempre  $\Omega(\mathbf{x}_i) \cdot \Omega(\mathbf{x}_j)$ . Isso é obtido com o uso de funções denominadas *kernels*.

Um *kernel*  $K$  é uma função que recebe dois pontos  $\mathbf{x}_i$  e  $\mathbf{x}_j$  do espaço de entradas e computa o produto escalar desses dados no espaço de características [56]:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Omega(\mathbf{x}_i) \cdot \Omega(\mathbf{x}_j). \quad (60)$$

Algumas das funções de *kernel* mais utilizadas estão listadas na tabela 12. Cada uma apresenta parâmetros que devem ser determinados pelo usuário.

Tabela 12 – Funções de *kernel* mais comuns com seus respectivos parâmetros.

Tipo de <i>kernel</i>	Função $K(\mathbf{x}_i, \mathbf{x}_j)$	Parâmetros
Linear	$(\mathbf{x}_i \cdot \mathbf{x}_j)$	—
Polinomial	$(\delta(\mathbf{x}_i \cdot \mathbf{x}_j) + \kappa)^d$	$\delta, \kappa$ e $d$
Base Radial (RBF)	$\exp(-\gamma\ \mathbf{x}_i - \mathbf{x}_j\ ^2)$	$\gamma$
Sigmoidal	$\tanh(\delta(\mathbf{x}_i \cdot \mathbf{x}_j) + \kappa)$	$\delta$ e $\kappa$

Fonte: Frederico Frias, 2021

#### 4.2 Redução de dimensionalidade com algoritmo genético

Em geral, tarefas de classificação de padrões envolvem um espaço de entradas dimensional muito elevado. Dados de alta dimensão são problemáticos para algoritmos de classificação pelo seu alto custo computacional. Ademais, a grande quantidade de características amplia o tamanho do espaço de pesquisa e, conseqüentemente, dificulta o processamento dos dados, comprometendo a eficiência e a eficácia do aprendizado de máquina. Sendo assim, é de suma importância a escolha da combinação das características de entrada originais que mais contribuam para a classificação do problema. Entretanto, a máquina de vetor de suporte não possui uma detecção interna automatizada para tal fim.

A redução da dimensionalidade como uma etapa de pré-processamento para o algoritmo de classificação é eficaz na remoção de dados irrelevantes e até mesmo redundantes, maximizando a precisão do aprendizado [57]. Esta técnica possui algumas vantagens, dentre elas:

- reduz a dimensionalidade do espaço de características, aumentando a velocidade do algoritmo;

- remove os dados redundantes, irrelevantes e ruidosos;
- aumenta a precisão do modelo resultante.

Existem duas técnicas de redução de dimensionalidade: seleção e extração de características. Na primeira delas, seleciona-se o melhor subconjunto de características que será aquele que possui o menor número de dimensões que contêm os elementos significativos para a resolução do problema específico. A mesma tem a vantagem de que nenhuma informação sobre a relevância de uma dada característica é perdida. Entretanto, se as características originais forem muito diversas e um conjunto pequeno das mesmas for necessário, há a possibilidade de perda de informação, uma vez que algumas dessas características poderão ser omitidas durante o processo de seleção.

A extração de características é um método de transformação do espaço de entradas para um subespaço de baixa dimensão que preserva a maior parte das informações importantes. Uma vantagem desta técnica é que o tamanho do espaço das características pode ser reduzido sem perda de muitas informações em relação ao espaço original. A desvantagem é que as informações sobre a contribuição de uma característica original muitas vezes são perdidas.

No caso da técnica de seleção de características, existem duas abordagens distintas: a de filtragem e a de *wrapper*. No método de filtragem, a seleção ocorre antes de aplicar o classificador no subconjunto de características escolhidas. A mesma é realizada de acordo com uma medida de relevância predefinida, independente do desempenho de generalização do algoritmo de aprendizagem. Já a técnica de *wrapper* treina o classificador com um determinado subconjunto de características, retornando o desempenho de generalização estimado da máquina de aprendizado como uma avaliação desse subconjunto de entrada. Repete-se esta etapa para cada subconjunto de características considerado.

O método de *wrapper* funciona melhor que o de filtragem dado que o processo de seleção é otimizado para o classificador do algoritmo de aprendizado de máquina. Contudo, quando o espaço de características é grande, o mesmo tem um alto custo computacional além de ter o processo de seleção lento. Apesar da desvantagem computacional citada, utilizou-se neste trabalho a técnica de *wrapper*, fazendo uso de um algoritmo genético na seleção do subconjunto de características de entrada mais relevante entre o conjunto de indicadores técnicos utilizados.

Algoritmo genético, do inglês, *genetic algorithm* (GA), é uma classe particular dos algoritmos evolutivos. Baseado no princípio de sobrevivência do mais apto, este tipo de método emprega técnicas inspiradas na biologia evolutiva, como seleção, mutação, herança e cruzamento. Tem sido muito utilizado em diversos domínios, como engenharia, física, ciência da computação, biologia, entre outros, com o intuito de encontrar soluções ideais, sejam elas exatas ou aproximadas, em problemas de otimização e de busca.

É considerado um método de pesquisa heurística adaptativa, com busca probabilística baseada na seleção e na genética natural. O algoritmo genético pode ser empregado para encontrar a solução, muitas vezes aproximada, mais apta para um problema em que a representação entre as variáveis de entrada e os dados esperados não está bem definida. Embora randomizados, os algoritmos genéticos não são aleatórios. Em vez disso, exploram informações históricas para direcionar a busca para a região de melhor desempenho dentro do espaço de pesquisa.

Um algoritmo genético é iniciado com um conjunto de soluções. Fazendo um paralelo com a terminologia clássica, uma solução para um problema é denominada indivíduo. O conjunto de indivíduos é chamado de população. Cada indivíduo possui uma cadeia de cromossomos que codifica as suas características. O cromossomo é uma sequência de alelos que representam uma quantidade de informação, como um *bit* [58].

No decorrer de cada geração, os cromossomos são selecionados probabilisticamente de acordo com os seus valores de aptidão, medidos por uma função objetivo relacionada ao problema de busca e de otimização. Quanto mais apta for a solução, maior será a chance de se reproduzir. Desses cromossomos, alguns acasalam-se aleatoriamente, produzindo descendentes. A próxima geração será selecionada por meio dos operadores genéticos de recombinação, do inglês, *crossover*, e/ou de mutação. Como os cromossomos com altos valores de aptidão possuem grande probabilidade de serem selecionados, os da geração subsequente poderão ter um valor médio de aptidão maior do que os da geração anterior. O processo de evolução é repetido até que a condição final, como o número de populações ou a melhoria da solução, seja satisfeita.

A seguir, são apresentadas por ordem de execução cada uma das etapas da metodologia de um algoritmo genético.

- Inicialização: geração aleatória de muitas soluções individuais de forma a cobrir todo o espaço de busca. O conjunto de soluções inicial irá compor a população que

poderá conter centenas ou milhares de soluções possíveis, dependendo da natureza do problema;

- Avaliação: cada membro da população é avaliado de acordo com uma função objetivo, comumente chamada de *fitness*. A avaliação mede a aptidão do indivíduo, ou seja, o quão o mesmo está adequado em cumprir os requisitos do problema. Em geral, as funções *fitness* são criadas para que uma parcela reduzida de soluções menos adequadas seja escolhida, proporcionando uma maior diversidade da população afim de evitar convergência precoce para uma solução ruim;
- Seleção: a cada iteração, uma parcela da população é selecionada para dar continuidade a uma nova geração. A seleção visa aumentar a aptidão geral da população, escolhendo os indivíduos mais adequados e descartando aqueles que não estão bem adaptados para resolver o problema alvo. A idéia principal é aumentar as chances dos indivíduos mais aptos afim de preservá-los para a próxima geração;
- Reprodução: após a seleção de indivíduos, o algoritmo genético simula o mecanismo natural de reprodução sexual. A idéia é criar uma população constituída de descendentes com maior aptidão. Ao combinar certas características de dois indivíduos (*crossover*) já aptos, a prole possivelmente herdará as melhores características de cada um dos seus progenitores, elevando sua aptidão em relação à de seus pais. Em geral, a aptidão média da população terá aumentado com este procedimento, dado que o melhor da primeira geração é selecionado para procriação, junto com uma pequena parcela de soluções menos adequadas, selecionadas pelos motivos mencionados anteriormente;
- Mutação: esta etapa do processo, apesar de relativamente pequena, desempenha um papel importante na alteração da genética da população. Se não houvesse mutação, é bem provável que todas as combinações atingidas durante as sucessivas gerações já estariam presentes na população inicial. A mutação faz pequenas alterações aleatórias no genoma do indivíduo que, se proporciona uma vantagem ao mesmo, aumentará sua aptidão. Caso contrário, a seleção se encarregará de remover o gene mutado;
- Finalização: ao término da mutação, o *pool* genético estará completo. O processo é reiniciado para a próxima geração, até que seja alcançada a condição final. Há diversos motivos para finalizar um algoritmo genético, dentre eles estão: o processo encontrou uma solução muito boa; a aptidão não aumentou mais que um limite predefinido nas últimas  $n$  iterações; restrições em relação ao tempo disponível.

Apresentado o algoritmo utilizado na seleção dos melhores indicadores técnicos bem como o de aprendizado supervisionado de classificação binária utilizado na modelagem da tendência do preço de uma dada ação, na próxima seção serão evidenciados os resultados obtidos com este modelo híbrido.

#### 4.3 *Predição da dinâmica do movimento*

Como mencionado na seção 1.3, os dados considerados nesta etapa do trabalho foram redimensionados de modo que o intervalo de tempo entre os mesmos fosse constante, espaçamento de cinco minutos, e sincronizados entre si. No caso de falta de informação, por falta de negociação do ativo no período de tempo em questão, utilizou-se o último dado disponível no preenchimento.

Primeiramente, o algoritmo genético foi usado como mecanismo de seleção, do tipo *wrapper*, das características que serão utilizadas na classificação binária da tendência do preço da ação alvo pelo método de aprendizado de máquina. Essas características foram calculadas, para cada um dos ativos da carteira, a partir dos indicadores técnicos apresentados na tabela 13. Todos eles se baseiam no preço da ação e são classificados, segundo terminologias do mercado financeiro, em dois tipos: de tendência e de *momentum*.

Indicadores de tendência são utilizados para medir a direção de uma tendência usando alguma forma de cálculo de preço médio para estabelecer uma linha de base. A medida que os preços se movem acima da média, pode ser considerada uma tendência de alta. Quando os preços caem abaixo da média, é uma sinalização de baixa.

Já os indicadores de *momentum* ajudam a identificar a velocidade do movimento dos preços. Eles indicam os pontos onde o mercado pode reverter. Normalmente, isso aparece como uma linha abaixo de um gráfico de preços que oscila conforme o *momentum* muda. Quando há uma divergência entre o preço e o indicador, pode sinalizar uma mudança nos preços futuros.

Apesar da eficácia da análise técnica não estar devidamente comprovada estatisticamente, estes indicadores foram empregados neste estudo por serem amplamente difundidos no mercado financeiro. Sendo assim, presume-se que os mesmos tenham um peso relevante nas negociações dos ativos e, conseqüentemente, forte influência na dinâmica de seus movimentos.



Tabela 13 – Indicadores técnicos com suas respectivas classificações.

Indicador técnico	Equação	Classificação
MOM <sup>1</sup>	$p_t - p_{t-n}$	$> 0 \uparrow$ $< 0 \downarrow$
ROC <sup>2</sup>	$\left(\frac{p_t}{p_{t-n}} - 1\right) \cdot 100$	$> 0 \uparrow$ $< 0 \downarrow$
RSI <sup>3</sup>	$\frac{100 \cdot RS}{1 + RS}, RS = \frac{SMA_n^\uparrow}{SMA_n^\downarrow}$	$< 30 \uparrow$ $30 \leq x_i \leq 70 \ \& \ x_i > x_{i-1} \uparrow$ $> 70 \downarrow$ $30 \leq x_i \leq 70 \ \& \ x_i < x_{i-1} \downarrow$
EMA <sup>4</sup>	$\frac{\sum_{i=0}^{n-1} (1 - \alpha)^i \cdot p_{t-i}}{\sum_{i=0}^{n-1} (1 - \alpha)^i}$	$\leq p_t \uparrow$ $> p_t \downarrow$
SMA <sup>5</sup>	$\frac{\sum_{i=0}^{n-1} p_{t-i}}{n}$	$\leq p_t \uparrow$ $> p_t \downarrow$
SMM <sup>6</sup>	$\frac{p \left\lfloor \frac{(n+1)}{2} \right\rfloor + p \left\lceil \frac{(n+1)}{2} \right\rceil}{2}$	$\leq p_t \uparrow$ $> p_t \downarrow$
WMA <sup>7</sup>	$\frac{\sum_{i=0}^{n-1} (n-i) \cdot p_{t-i}}{n + (n-1) + \dots + 1}$	$\leq p_t \uparrow$ $> p_t \downarrow$

<sup>1</sup> *Momentum.*<sup>2</sup> Taxa de variação, do inglês, *rate-of-change.*<sup>3</sup> Índice de força relativa, do inglês, *relative strength index.*<sup>4</sup> Média móvel exponencial, do inglês, *exponential moving average.*<sup>5</sup> Média móvel simples, do inglês, *simple moving average.*<sup>6</sup> Mediana móvel simples, do inglês, *simple moving median.*<sup>7</sup> Média móvel ponderada, do inglês, *weighted moving average.*

Os indicadores técnicos de *momentum* são: MOM, ROC e RSI. Já os de tendência são: EMA, SMA, SMM e WMA. Estes indicadores podem ser calculados com diferentes janelas de tempo, por exemplo, é possível calcular o *rate-of-change* com 1, 2, ..., ou  $n$  intervalos, o que implica dizer que a taxa de variação pode ser observada a partir de 1, 2, ..., ou  $n$  passos atrás, respectivamente, em relação ao tempo atual. Na análise técnica empregada no mercado de ações, existem estratégias que combinam diferentes indicadores técnicos com janelas de tempo específicas. Contudo, neste trabalho, múltiplos intervalos de tempo foram considerados para cada um dos indicadores, deixando a cargo do algoritmo genético selecionar quais os melhores para descrever o ativo alvo no período de tempo analisado. As janelas de tempo utilizadas foram: de 1 a 10 para o MOM e o ROC; de 5 a 10 para os de média móvel; 5 e 10 para o RSI.

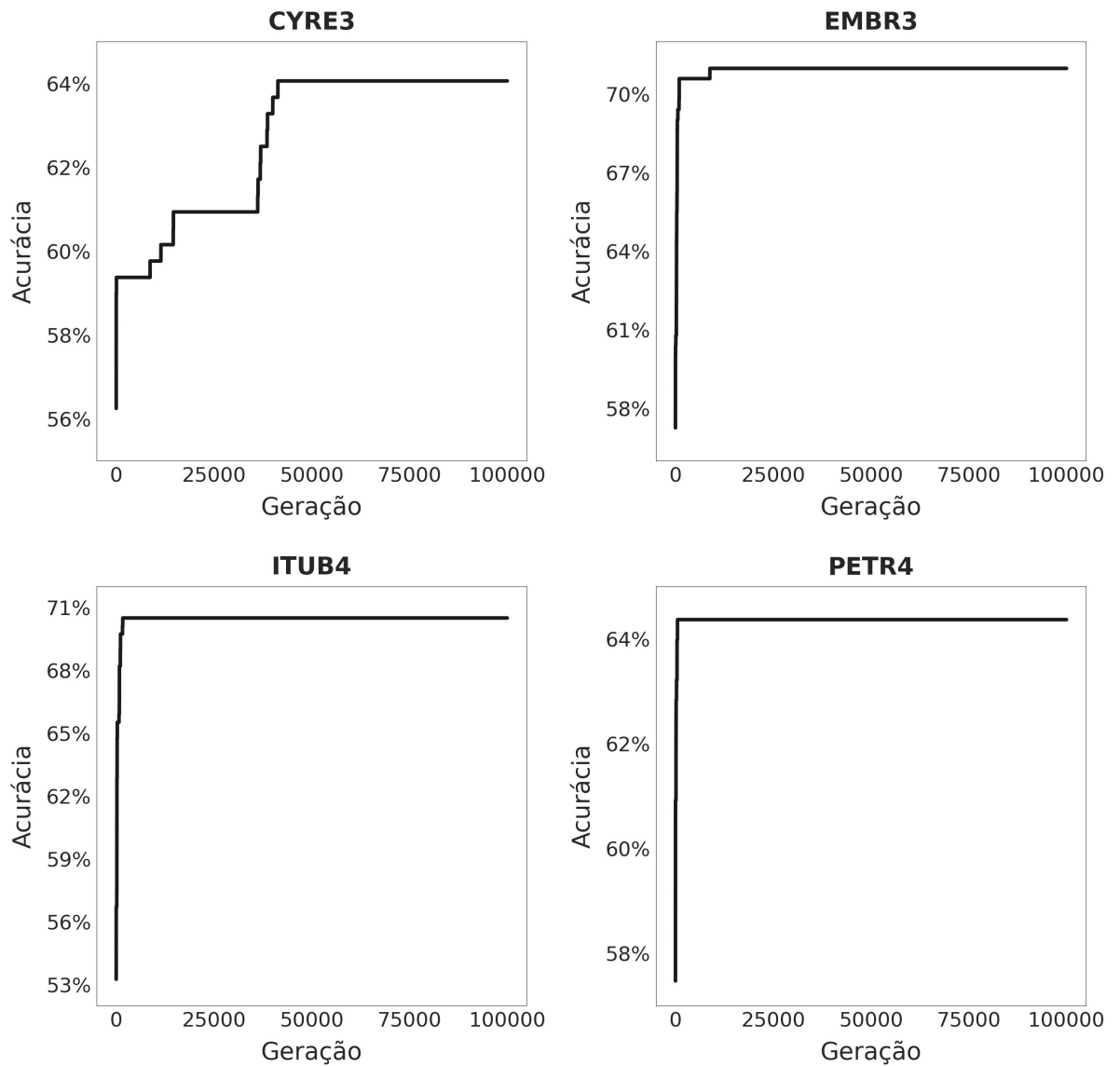
As séries temporais das ações foram divididas em dois grupos de três dias úteis cada: de 12 a 14; de 13 a 16. Ambos correspondentes ao mês de novembro de 2018. Estes serão os grupos de treinamento que o algoritmo genético, em conjunto com o método de aprendizado de máquina, fará uso para selecionar o melhor conjunto de características a ser utilizada na predição da tendência da ação alvo nos dias 16 e 19 respectivamente.

A população do algoritmo genético é composta por 100 indivíduos. Cada indivíduo possui 368 cromossomos: 7 indicadores técnicos (tabela 13) com diferentes janelas de tempo, totalizando 48, para cada um dos 8 ativos integrantes da carteira. A população foi inicializada por uma função geradora de números inteiros pseudo-aleatórios: 0 (característica desativada) ou 1 (característica ativada). Para avaliar os seus membros, utilizou-se a acurácia, fração dos valores preditos classificados corretamente pelo algoritmo de aprendizado de máquina que, neste estudo, foi o SVM não linear com a função RBF, *kernel* que obteve os melhores resultados dentre os demais descritos na tabela 12.

A cada iteração do algoritmo genético, os quatro indivíduos mais aptos foram selecionados. Por aptidão, entenda-se os melhores avaliados pela função objetivo, isto é, os que obtiveram os maiores valores de acurácia dentre os indivíduos da mesma geração. Após a seleção, foi realizada a reprodução por meio do método denominado *single-point crossover*, onde um ponto nos cromossomos de ambos os pais é escolhido aleatoriamente e os *bits* à direita desse ponto são trocados entre os dois cromossomos pais, resultando em dois filhos em que cada um carrega algumas informações genéticas dos seus progenitores. Por fim, ocorreu a mutação com 10% de probabilidade de que um *bit* arbitrário em uma sequência genética fosse invertido aleatoriamente, por uma função geradora de números inteiros

pseudo-aleatórios (0 ou 1), de seu estado original. O algoritmo genético foi finalizado ao alcançar cem mil gerações.

Figura 27 – Acurácia ao longo das cem mil gerações do algoritmo genético com população de tamanho 100, 10% de mutação e número de pais igual a 4 para cada ação alvo entre as datas 12 a 14 de novembro de 2018.



Fonte: Frederico Frias, 2021

O gráfico da acurácia ao longo das gerações para cada ação alvo usando os dados de 12 a 14 de novembro de 2018 encontra-se na figura 27, enquanto que, na tabela 14, estão evidenciados os indicadores técnicos selecionados pelo algoritmo genético para o mesmo período de tempo.

Tabela 14 – Indicadores técnicos selecionados pelo algoritmo genético para cada ação alvo entre as datas 12 a 14 de novembro de 2018.

Ação	Indicadores técnicos
CYRE3	BBDC3_MOM_2, BBDC3_MOM_3, BBDC3_MOM_5, BBDC3_ROC_2, BBDC3_ROC_3, BBDC3_ROC_4, BBDC3_ROC_6, BBDC3_ROC_10, BBDC3_EMA_5, BBDC3_WMA_5, BBDC3_WMA_6, BBDC4_MOM_7, BBDC4_ROC_7, BBDC4_EMA_5, BBDC4_SMA_6, BBDC4_SMM_9, BBDC4_SMM_10, BBDC4_WMA_7, BBDC4_RSI_10, CVCB3_MOM_3, CVCB3_MOM_7, CVCB3_ROC_3, CVCB3_ROC_5, CVCB3_ROC_9, CVCB3_WMA_5, CVCB3_WMA_8, CVCB3_WMA_9, CVCB3_RSI_10, EQTL3_MOM_4, EQTL3_MOM_6, EQTL3_ROC_1, EQTL3_EMA_7, EQTL3_SMA_8, EQTL3_SMM_10, EQTL3_WMA_7, GOLL4_MOM_8, GOLL4_ROC_1, GOLL4_ROC_4, GOLL4_SMM_7, GOLL4_SMM_8, GOLL4_WMA_8, GOLL4_WMA_9, IGTA3_MOM_5, IGTA3_MOM_8, IGTA3_MOM_10, IGTA3_SMM_5, IGTA3_WMA_7, IGTA3_WMA_10, MULT3_MOM_2, MULT3_MOM_4, MULT3_MOM_8, MULT3_ROC_2, MULT3_ROC_4, MULT3_ROC_8, MULT3_ROC_9, MULT3_SMA_9, MULT3_SMM_5, MULT3_SMM_6, MULT3_SMM_7, MULT3_WMA_10, SBSP3_MOM_2, SBSP3_ROC_5, SBSP3_ROC_6, SBSP3_EMA_5, SBSP3_SMA_6, SBSP3_SMA_9, SBSP3_SMM_6
EMBR3	BTOW3_MOM_6, BTOW3_MOM_9, BTOW3_ROC_6, BTOW3_EMA_5, BTOW3_SMM_7, BTOW3_RSI_5, FIBR3_MOM_3, FIBR3_SMM_10, KLBN11_ROC_4, KLBN11_SMM_8, KLBN11_SMM_9, MRFG3_ROC_6, MRFG3_ROC_7, MRFG3_EMA_8, MRFG3_SMM_6, MRFG3_RSI_10, NATU3_MOM_10, NATU3_SMA_10, SUZB3_MOM_3, SUZB3_ROC_3, SUZB3_ROC_8, SUZB3_ROC_10, TIMP3_MOM_6, TIMP3_ROC_6, TIMP3_EMA_10, TIMP3_WMA_10, VIVT4_ROC_6, VIVT4_EMA_5, VIVT4_SMA_9, VIVT4_WMA_6
ITUB4	BOVA11_SMM_6, BOVA11_SMM_8, B3SA3_ROC_9, B3SA3_WMA_7, BBDC4_MOM_10, BBDC4_ROC_8, BBDC4_EMA_10, ELET3_MOM_7, ELET6_MOM_6, ELET6_ROC_9, GOLL4_MOM_9, GOLL4_MOM_10, GOLL4_ROC_2, GOLL4_ROC_3, GOLL4_ROC_4, GOLL4_ROC_6, GOLL4_WMA_5, ITSA4_ROC_2, ITSA4_ROC_10, ITSA4_EMA_8, ITSA4_EMA_10, ITSA4_WMA_5
PETRA4	BOVA11_MOM_3, BOVA11_ROC_2, BOVA11_ROC_4, BOVA11_ROC_8, BOVA11_WMA_8, BOVA11_WMA_9, BOVA11_RSI_5, BOVA11_RSI_10, B3SA3_MOM_9, B3SA3_ROC_1, B3SA3_ROC_4, B3SA3_ROC_7, B3SA3_SMA_7, B3SA3_SMA_10, B3SA3_SMM_7, B3SA3_WMA_5, B3SA3_WMA_8, CMIG4_MOM_2, CMIG4_MOM_3, CMIG4_MOM_10, CMIG4_ROC_1, CMIG4_ROC_3, CMIG4_ROC_7, CMIG4_ROC_8, CMIG4_ROC_10, CMIG4_EMA_10, CMIG4_SMA_8, CMIG4_SMA_9, CMIG4_SMM_6, CMIG4_SMM_9, CMIG4_SMM_10, CMIG4_WMA_6, CMIG4_WMA_9, CMIG4_WMA_10, GOLL4_MOM_5, GOLL4_ROC_4, GOLL4_EMA_5, GOLL4_EMA_7, GOLL4_SMA_8, GOLL4_SMM_6, GOLL4_SMM_7, GOLL4_SMM_8, GOLL4_SMM_10, GOLL4_WMA_9, GOLL4_WMA_10, ITSA4_MOM_10, ITSA4_ROC_1, ITSA4_ROC_3, ITSA4_ROC_9, ITSA4_ROC_10, ITSA4_EMA_5, ITSA4_EMA_9, ITSA4_SMA_6, ITSA4_SMA_7, ITSA4_SMA_10, ITSA4_SMM_8, ITSA4_SMM_10, ITSA4_WMA_8, ITSA4_RSI_5, ITSA4_RSI_10, ITUB4_MOM_1, ITUB4_MOM_4, ITUB4_MOM_10, ITUB4_ROC_7, ITUB4_EMA_6, ITUB4_EMA_8, ITUB4_SMM_7, ITUB4_WMA_10, ITUB4_RSI_10, PETR3_MOM_4, PETR3_MOM_9, PETR3_EMA_6, PETR3_SMA_7, PETR3_SMA_9, PETR3_SMM_10, PETR3_WMA_5, SANB11_MOM_1, SANB11_MOM_4, SANB11_MOM_8, SANB11_MOM_9, SANB11_MOM_10, SANB11_ROC_1, SANB11_ROC_4, SANB11_ROC_8, SANB11_SMA_5, SANB11_SMM_9, SANB11_WMA_5, SANB11_WMA_7, SANB11_RSI_10

É importante mencionar que os valores oriundos dos indicadores técnicos foram normalizados de acordo com suas respectivas classificações descritas na tabela 13. O principal objetivo da normalização é garantir a qualidade dos dados antes de serem usados no algoritmo de aprendizagem. Para isso, os dados foram dimensionados na mesma faixa de valores para cada recurso de entrada, assumindo o valor de 1 para uma tendência de subida e -1, para descida. Deste modo, o viés no algoritmo genético é minimizado de um recurso para outro além de ter a possibilidade de acelerar o tempo de treinamento [2]. Este recurso é especialmente útil para a aplicação de modelagem onde as entradas estão geralmente em escalas muito diferentes, como é o caso deste estudo.

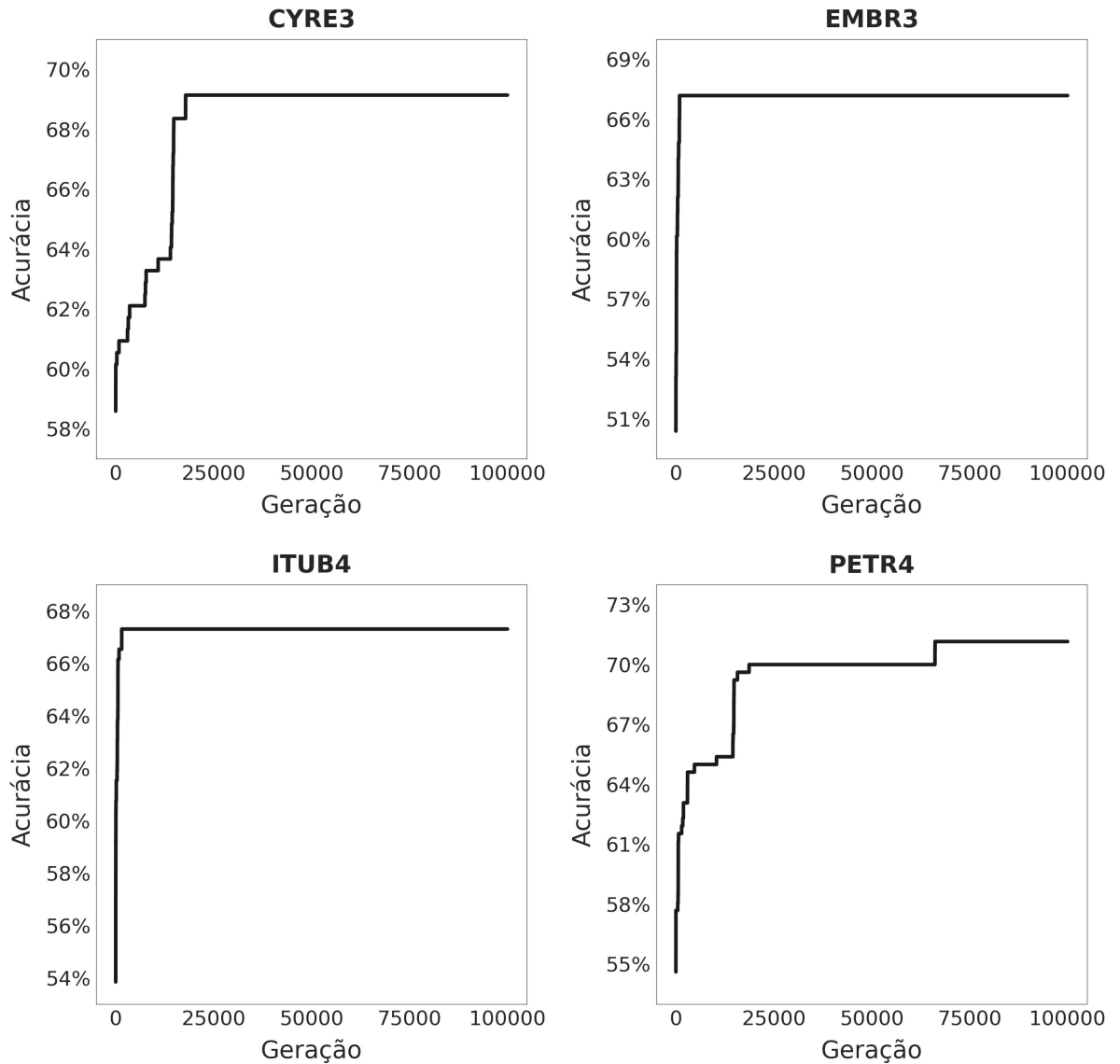
Para além da normalização, os valores dos indicadores técnicos foram deslocados no tempo, ou seja, como o intuito deste trabalho é tentar prever o movimento do preço da ação alvo baseado no conjunto de indicadores de *momentum* e de tendência das ações que constituem o seu portfólio, então, para modelar o movimento da mesma no tempo  $t$ , se faz necessário utilizar o conjunto de indicadores técnicos presentes em  $t - n$  (neste caso,  $n = 300s$ ) como entrada do algoritmo de aprendizado de máquina.

Afim de estimar o desempenho do classificador gerado pelo modelo do SVM, empregou-se uma técnica para avaliar a capacidade de generalização do algoritmo de aprendizado a partir de um conjunto de dados denominada validação cruzada. O conceito deste método é particionar o grupo de dados em subconjuntos mutuamente exclusivos e, posteriormente, o uso de alguns destes subconjuntos para a estimação dos parâmetros do modelo (dados de treinamento), sendo os subconjuntos restantes (dados de validação ou de teste) utilizados na validação do modelo [59].

A técnica de validação cruzada empregada foi a *k-fold*. Este método consiste em dividir aleatoriamente o conjunto total de dados em  $k$  subconjuntos de tamanho aproximadamente igual. Em seguida, as  $k - 1$  subamostras são usadas no treinamento do classificador, enquanto a subamostra restante, como dados de validação para teste. Este processo é realizado  $k$  vezes, alternando de forma circular o subconjunto de teste. Ao final das  $k$  iterações, calcula-se a acurácia sobre os erros encontrados, obtendo uma medida mais confiável sobre a capacidade do modelo de representar o processo gerador dos dados.

Neste trabalho, o conjunto de dados foi dividido em *5-folds* ( $k = 5$ ), dos quais quatro (80%) foram usados como conjunto de treinamento para a construção do algoritmo de predição e o restante (20%), como conjunto de teste para justificar o desempenho de generalização do modelo.

Figura 28 – Acurácia ao longo das cem mil gerações do algoritmo genético com população de tamanho 100, 10% de mutação e número de pais igual a 4 para cada ação alvo entre as datas 13 a 16 de novembro de 2018.



Fonte: Frederico Frias, 2021

Na figura 28, evidencia-se o gráfico da acurácia ao longo das gerações para cada ação alvo utilizando os dados de 13 a 16 de novembro de 2018. Os indicadores técnicos escolhidos pelo algoritmo genético para o período em questão podem ser observados na tabela 15.

Tabela 15 – Indicadores técnicos selecionados pelo algoritmo genético para cada ação alvo entre as datas 13 a 16 de novembro de 2018.

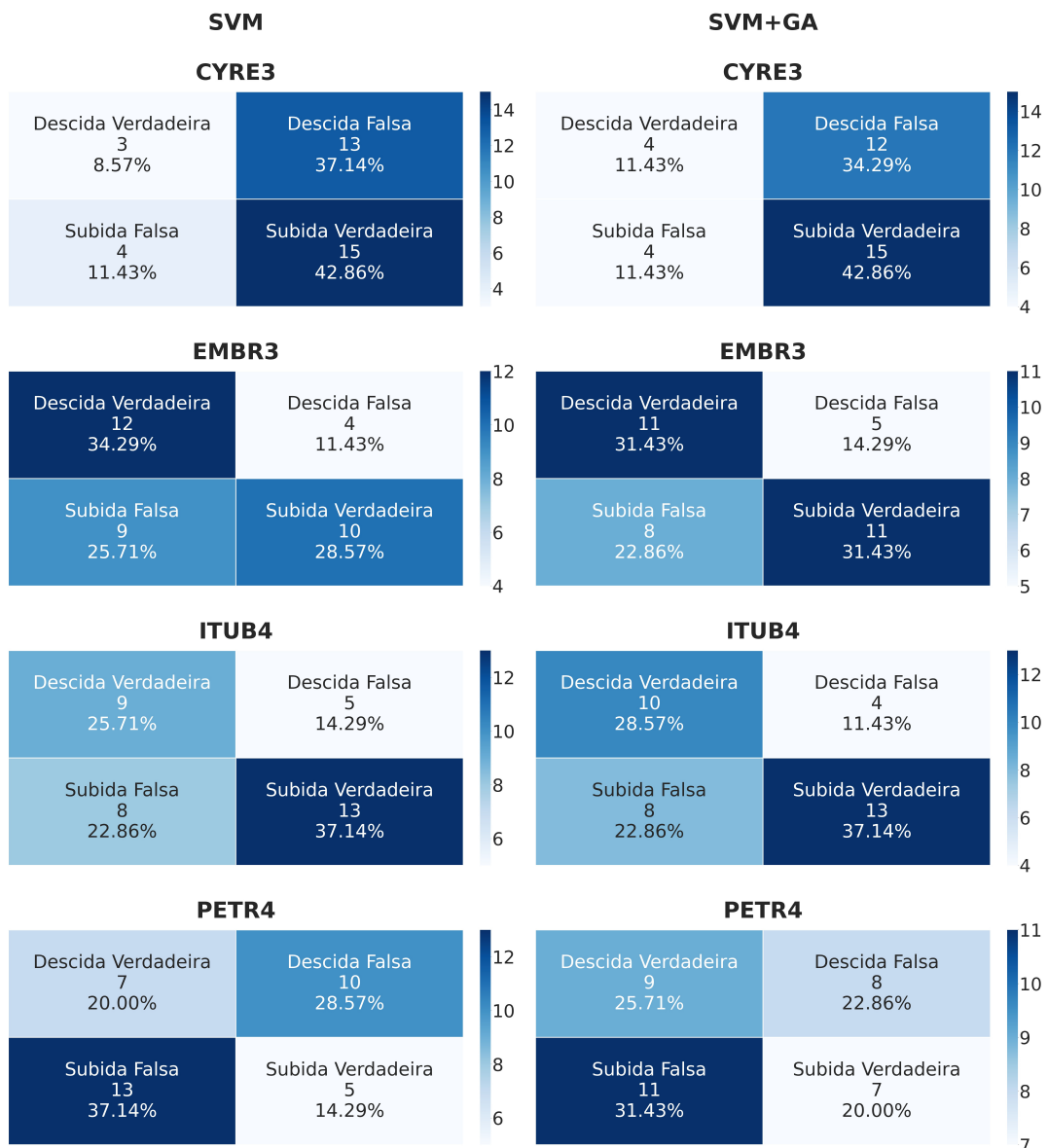
Ação	Indicadores técnicos
CYRE3	BBDC3_MOM_1, BBDC3_SMM_8, BBDC4_ROC_7, BBDC4_EMA_6, BBDC4_SMM_8, BBDC4_RSI_10, CVCB3_MOM_8, CVCB3_MOM_10, EQTL3_MOM_7, EQTL3_ROC_3, EQTL3_ROC_4, EQTL3_SMA_8, EQTL3_SMM_10, GOLL4_MOM_4, GOLL4_ROC_5, IGTA3_MOM_4, IGTA3_SMM_7, MULT3_MOM_3, MULT3_EMA_6, MULT3_SMA_5, MULT3_SMA_8, MULT3_SMM_5, MULT3_SMM_8, MULT3_WMA_8, MULT3_WMA_10
EMBR3	BTOW3_MOM_1, BTOW3_MOM_3, BTOW3_MOM_4, BTOW3_ROC_9, BTOW3_EMA_7, BTOW3_EMA_9, BTOW3_SMA_6, BTOW3_SMM_8, BTOW3_WMA_5, FIBR3_MOM_9, FIBR3_ROC_1, FIBR3_ROC_6, FIBR3_SMM_6, FIBR3_SMM_7, KLBN11_MOM_4, KLBN11_MOM_8, KLBN11_ROC_3, KLBN11_WMA_7, KLBN11_WMA_8, KLBN11_RSI_5, MRFG3_SMA_8, NATU3_ROC_2, NATU3_ROC_7, NATU3_EMA_8, NATU3_WMA_10, SUZB3_ROC_1, SUZB3_ROC_2, SUZB3_EMA_9, SUZB3_SMA_9, SUZB3_WMA_7, TIMP3_WMA_10, VIVT4_ROC_2, VIVT4_EMA_8
ITUB4	BOVA11_SMA_8, BOVA11_SMM_9, BOVA11_WMA_9, B3SA3_ROC_9, B3SA3_EMA_8, B3SA3_SMA_5, B3SA3_SMM_5, B3SA3_SMM_6, B3SA3_SMM_8, B3SA3_SMM_10, B3SA3_WMA_5, B3SA3_WMA_7, BBAS3_MOM_10, BBAS3_ROC_3, BBAS3_ROC_8, BBAS3_SMM_9, BBAS3_WMA_6, ELET3_MOM_8, ELET3_RSI_10, ELET6_ROC_7, ELET6_SMA_5, ELET6_SMA_10, ELET6_SMM_9, GOLL4_SMM_6, GOLL4_RSI_5, ITSA4_ROC_6, ITSA4_ROC_9, ITSA4_EMA_6, ITSA4_SMA_6, ITSA4_SMM_5, ITSA4_SMM_7, ITSA4_WMA_5
PETR4	BOVA11_MOM_8, BOVA11_SMM_10, BOVA11_WMA_9, B3SA3_ROC_3, B3SA3_EMA_9, B3SA3_SMA_9, B3SA3_SMA_10, B3SA3_RSI_10, CMIG4_ROC_2, CMIG4_ROC_8, CMIG4_ROC_10, CMIG4_WMA_8, GOLL4_MOM_4, GOLL4_WMA_7, ITSA4_ROC_5, ITUB4_ROC_6, ITUB4_EMA_10, ITUB4_RSI_5, PETR3_MOM_10, PETR3_SMA_5, SANB11_MOM_1, SANB11_MOM_2, SANB11_MOM_10, SANB11_ROC_5, SANB11_ROC_7, SANB11_ROC_10, SANB11_SMA_8, SANB11_SMA_10, SANB11_SMM_7

Fonte: Frederico Frias, 2021

Após a seleção das melhores características pelo algoritmo genético, fez-se a predição do movimento do preço da ação alvo utilizando o SVM não linear com o *kernel* RBF. Apenas os primeiros registros dos dias 16 e 19 de novembro de 2018 foram utilizados como dados de teste pelo algoritmo de aprendizado, garantindo que os mesmos não excedessem o percentual de 15% do total de dados (treinamento + teste). Na função objetivo (tabela 12), o  $\gamma$  considerado foi inversamente proporcional ao número de características.

As matrizes de confusão<sup>1</sup> dos dados preditos nos dias 16 e 19 para cada ação modelada encontram-se, respectivamente, nas figuras 29 e 30. Nessas figuras, é possível comparar as classificações, em quantidade e em percentual, de subida bem como de descida do movimento do preço do ativo feitas corretamente pelo modelo híbrido proposto (SVM+GA) como também pelo modelo individual (SVM).

Figura 29 – Comparativo das matrizes de confusão do modelo híbrido (SVM+GA) com o modelo individual (SVM) para cada ação alvo para a data de 16 de novembro de 2018.

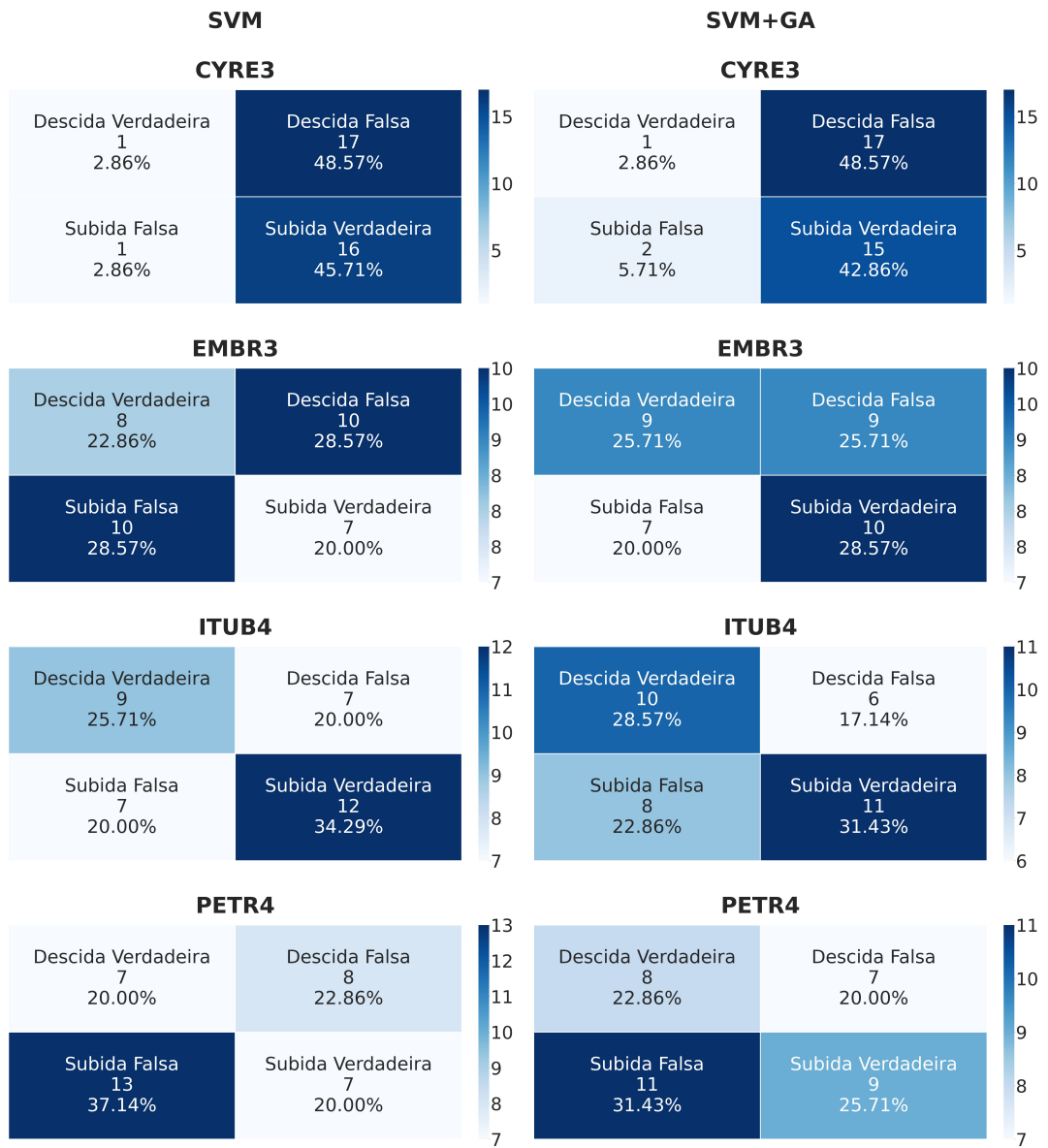


Fonte: Frederico Frias, 2021

<sup>1</sup> Tabela que permite uma visualização detalhada do desempenho de um algoritmo de classificação. Com duas dimensões, possui duas linhas (valores reais) e duas colunas (valores preditos) que relatam o número de falsos positivos, falsos negativos, verdadeiros positivos e verdadeiros negativos.



Figura 30 – Comparativo das matrizes de confusão do modelo híbrido (SVM+GA) com o modelo individual (SVM) para cada ação alvo para a data de 19 de novembro de 2018.



Fonte: Frederico Frias, 2021

Na tabela 16, pode-se observar um comparativo dos valores das acurácias obtidas por ambos os modelos nos dois dias utilizados na predição do movimento do preço de cada uma das ações alvo.

Tabela 16 – Comparativo das acurácias do modelo híbrido (SVM+GA) com o modelo individual (SVM) para cada ação alvo utilizando o valores preditos para as datas de 16 e 19 de novembro de 2018.

Ação	16/11/2018		19/11/2018	
	SVM (%)	SVM+GA (%)	SVM (%)	SVM+GA (%)
CYRE3	51.42	54.29	48.57	45.71
EMBR3	62.86	62.86	42.85	54.29
ITUB4	62.85	65.71	60.00	60.00
PETRA	34.29	45.71	40.00	48.57

Fonte: Frederico Frias, 2021

Em todas as situações, exceto no dia 19 para o ativo CYRE3, o modelo híbrido teve desempenho melhor ou igual ao algoritmo de aprendizado isoladamente, evidenciando que o algoritmo genético conseguiu selecionar os indicadores técnicos mais representativos para descrever o movimento de cada ação alvo em ambos os períodos analisados. No caso da CYRE3, é possível notar uma queda acentuada do dia 16 para o dia 19 na quantidade de características escolhidas ao analisar as tabelas 14 e 15. Esta ação é a mais ilíquida dentre as quatro modeladas, com negociações esparsas e com pouco volume ao longo do dia. Em particular, o dia 19 teve menos liquidez do que o dia 16 para o ativo em questão. Desta forma, é provável que o método evolutivo tenha encontrado uma solução ótima, ao escolher indicadores específicos que caracterizaram bem os dados usados na etapa de treinamento, mas que pouco representaram os dados usados na fase de teste, ocasionando a piora da acurácia em relação à obtida pelo algoritmo individual que considerou todos eles.

Comparando somente os resultados obtidos pelo modelo proposto, as ações que tiveram os melhores desempenhos nos dois dias foram a EMBR3 e a ITUB4, sendo que esta última alcançou os maiores valores de acurácia. Considerando este ativo, o algoritmo genético teve um desempenho alto no dia 16, convergindo relativamente rápido para a solução final (figura 27), o que pode explicar em partes a performance do método híbrido ser ligeiramente superior ao do SVM usado individualmente. Já no dia 19, os resultados preditos atingiram uma acurácia igual em ambos os modelos, indicando, possivelmente, a redundância dos indicadores técnicos filtrados pelo algoritmo evolutivo.

Além do mais, o expoente de Hurst desta ação foi o maior dentre as demais ações alvo, permanecendo acima de 0.5 em todo o período de tempo considerado na modelagem da tendência, como observado na área em cinza da figura 10. Segundo as características

apontadas na tabela 6, o expoente acima deste valor indica que a ação possui uma persistência em seu movimento, exibindo memória de longo prazo. Existem estudos que apontam a relação deste expoente com a habilidade de prever uma série temporal, evidenciando que expoentes de Hurst altos podem ser preditos com mais precisão do que aqueles com valores próximos ao de séries aleatórias, uma vez que períodos com uma estrutura de tendência forte podem ser utilizados no aprendizado da rede neural afim de beneficiar a predição [60].

Analisando os resultados de acurácia obtidos para as ações ITUB4 e PETR4 deste ponto de vista, pode-se inferir que o primeiro ativo mencionado terá naturalmente um desempenho maior do que os restantes ativos modelados e que o segundo terá o seu movimento mais difícil de ser predito dado que o seu comportamento é análogo ao de um movimento aleatório, com expoente de Hurst contíguo a 0.5 (tabela 3). Apesar da predição da tendência do preço da ação alvo ser feita de forma indireta, isto é, utiliza-se da dinâmica dos ativos de seu portfólio como referência na sua modelagem, ainda assim pode-se fazer uso desta premissa uma vez que estas ações são fortemente correlacionadas e possuem uma relação causal, como comprova-se pelas tabelas 9 e 10, com o ativo em questão.

Por último, temos a ação EMBR3 que apresentou um desempenho atípico no dia 16. Este caso pode ser similar ao do ITUB4 no dia 19, onde é muito provável que os indicadores técnicos removidos pelo algoritmo genético fossem irrelevantes ou redundantes para a modelagem do seu movimento. Mas o que torna este caso singular é a alta acurácia na predição da tendência de um ativo que é ilíquido, com um comportamento anti-persistente e cujos ativos de seu portfólio possuem baixa correlação (tabela 8) e, inclusive, sem comprovada dependência causal, como é possível observar pela tabela 11. Mesmo com todos estes fatores adversos, a performance obtida para esta ação neste dia em específico foi superior a 60%. Uma análise mais detalhada precisaria ser feita para chegar a uma conclusão acerca dos resultados obtidos neste caso.

## 5 Conclusão

### 5.1 Considerações finais

Neste trabalho, procurou-se modelar a tendência do movimento (descida ou subida) do preço de uma dada ação do mercado brasileiro a partir de indicadores técnicos gerados pelos dados intradiários de ações com forte correlação e relação causal com a mesma. Para isso, foi necessário comprovar a existência das propriedades estatísticas denominadas de fatos estilizados afim de refutar a crença de que o mercado financeiro obedece à hipótese do mercado eficiente. Desta forma, algumas das propriedades de retorno bem como de volatilidade, utilizando dados diários dos quatro ativos modelados neste estudo (CYRE3, EMBR3, ITUB4 e PETR4), foram comprovadas, refutando a idéia de que os preços das ações podem ser descritos por um passeio aleatório e, conseqüentemente, inviabilizando sua predição afim de auferir lucros não apenas por arbitragem.

Como alternativa à EMH, utilizou-se uma nova hipótese, a do mercado fractal. A heterogeneidade dos investidores, no que concerne os seus horizontes de investimento, é seu principal fundamento, sendo a principal razão de haver estabilidade no mercado. Em suma, a uniformidade leva à instabilidade, enquanto a presença de uma ampla gama de horizontes torna o mercado estável. Sendo assim, mensurou-se a fractalidade das ações alvo empregando o método MF DFA. Todos os quatro ativos apresentaram um comportamento multifractal, o que, segundo esta hipótese, implica dizer que os diferentes tipos de investidores estão ali representados, garantindo sua estabilidade. Baseando-se nesta premissa da FMH, a presença de uma estrutura fractal na série temporal da ação modelada é uma condição necessária do modelo proposto, pois, assim, evita-se tanto períodos de alta volatilidade quanto a dificuldade de negociar o ativo ocasionada pela homogeneidade de horizontes de investimento presentes naquele mercado.

Em seguida, construiu-se uma carteira composta pelas ações que apresentaram uma forte correlação com o ativo modelado. Nesta etapa, foram utilizados dois métodos de captura da dependência entre duas séries temporais:  $\rho_{DCCA}$  e o CCM. O primeiro deles quantifica a correlação cruzada entre dois ativos em diferentes escalas de tempo, sendo capaz de mensurar a relação não linear entre ambos. No segundo deles, tenta-se mapear mutuamente os espaços-de-fase reconstruídos a partir de suas séries temporais, identificando se ambas estão dinamicamente ligadas, o que implica dizer que existe uma

relação de causalidade. Cada ação alvo teve pelo menos oito ativos com correlação alta e dependência de causa e efeito com a mesma, exceto a EMBR3. Esta, por sua vez, além de não ter obtido resultados expressivos de correlação com as demais ações do Ibovespa, também não teve a comprovação, por meio da técnica de *surrogates*, da existência de causalidade com os ativos de seu portfólio.

Após a construção da carteira de ativos, diferentes indicadores técnicos, de *momentum* e de tendência, foram calculados a partir do preço intradiário dos mesmos. Estes indicadores foram redimensionados para uma escala de -1 (tendência de descida) e +1 (tendência de subida) afim de minimizar o viés da classificação além de ter a possibilidade de acelerar o tempo de processamento. Com o intuito de filtrar as características mais representativas da dinâmica da ação alvo, empregou-se um algoritmo genético.

Por último, o algoritmo de aprendizado de máquina SVM com a função de *kernel* RBF foi usado para prever o movimento do preço do ativo modelado baseando-se nos indicadores selecionados pelo método evolutivo. Como forma de avaliar o desempenho do mesmo, utilizou-se a medida de acurácia. O modelo híbrido se mostrou mais eficaz do que o algoritmo de aprendizado isoladamente. Ademais, os dados preditos para cada uma das ações alvo, com exceção da EMBR3, estão de acordo com os seus expoentes de Hurst assim como com os resultados de dependência, causal ou não, com os ativos de suas respectivas carteiras.

Na tabela 17, tem-se uma síntese dos principais resultados obtidos em cada um dos métodos empregados neste trabalho com o intuito de modelar a tendência do preço intradiário de cada uma das ações estudadas.

Tabela 17 – Resumo com os principais resultados obtidos em cada um dos métodos utilizados na modelagem da tendência do preço das ações alvo.

Ação	$h(q = 2)$	Portfólio	16/11/2018		19/11/2018	
			SVM	SVM+GA	SVM	SVM+GA
CYRE3	$0.607 \pm 0.024$ (Persistente)	Correlações acima de 0.87 com 7 ativos com dependência causal	51.42%	54.29%	48.57%	45.71%
EMBR3	$0.417 \pm 0.023$ (Anti-persistente)	Correlações abaixo de 0.53 sem nenhum ativo com dependência causal	62.86%	62.86%	42.85%	54.29%
ITUB4	$0.669 \pm 0.027$ (Persistente)	Correlações acima de 0.92 com todos os ativos com dependência causal	62.85%	65.71%	60.00%	60.00%
PETR4	$0.468 \pm 0.032$ ( $\approx$ Browniano)	Correlações acima de 0.77 com 7 ativos com dependência causal	34.29%	45.71%	40.00%	48.57%

Fonte: Frederico Frias, 2021

## 5.2 Sugestões futuras

Como sugestões para futuras pesquisas, seria interessante inserir a dimensão fractal como ferramenta adicional na classificação do modelo híbrido, fazendo uso do expoente de Hurst calculado em diferentes intervalos de tempo. Além disso, poderiam se empregar indicadores técnicos de outros tipos, como de volume e de volatilidade, como também adicionar mais dos tipos já utilizados.

Outra medida importante seria investigar algum método de mensuração da dependência que utilize os dados dos ativos em sua frequência de tempo própria, evitando as desvantagens decorridas pelo redimensionamento das suas séries temporais. Por fim, construir alguma estratégia de negociação intradiária, considerando os custos de transação, em comparação com alguma do tipo *buy and hold* com o objetivo de analisar a viabilidade financeira do modelo híbrido proposto.

## Referências Bibliográficas

- 1 BM&FBOVESPA. *Por dentro da BM&FBOVESPA: guia prático de uma das maiores bolsas de valores e derivativos do mundo*. [S.l.]: BM&FBOVESPA, 2017. Citado 2 vezes nas páginas 16 e 19.
- 2 CHOUDHRY, R.; GARG, K. A hybrid machine learning system for stock market forecasting. *International Journal of Computer and Information Engineering*, v. 2, n. 3, p. 689–692, 2008. Citado 3 vezes nas páginas 17, 20 e 92.
- 3 BOCCARA, N. *Modeling complex systems*. 2. ed. [S.l.]: Springer New York Dordrecht Heidelberg London, 2010. ISBN 9781441965615. Citado na página 17.
- 4 B3 S.A. - Brasil, B. e. B. *B3 Institucional*. 2019. Disponível em: [http://www.b3.com.br/pt\\_br/b3/institucional/quem-somos/](http://www.b3.com.br/pt_br/b3/institucional/quem-somos/). Citado na página 18.
- 5 B3 S.A. - Brasil, B. e. B. *B3 Ibovespa*. 2019. Disponível em: [http://www.b3.com.br/pt\\_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-composicao-da-carteira.htm](http://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-composicao-da-carteira.htm). Citado na página 19.
- 6 CVM. *Mercado de valores mobiliários brasileiro*. 3. ed. Rio de Janeiro: Comissão de Valores Mobiliários, 2014. ISBN 9788567896007. Citado na página 20.
- 7 COSTA, L.; VASCONCELOS, G. L. Long-range correlations and nonstationarity in the Brazilian stock market. *Physica A*, v. 329, p. 231–248, 2003. Citado na página 20.
- 8 MORALES, R.; MATTEO, T. D.; ASTE, T. Non-stationary multifractality in stock returns. *Physica A*, Elsevier B.V., v. 392, n. 24, p. 6470–6483, 2013. ISSN 0378-4371. Citado na página 20.
- 9 GHOSH, I.; JANA, M. K. S. R. K. Fractal inspection and machine learning-based predictive modelling framework for financial markets. *Arabian Journal for Science and Engineering*, Springer Berlin Heidelberg, v. 43, n. 8, p. 4273–4287, 2018. ISSN 2191-4281. Citado 2 vezes nas páginas 20 e 72.
- 10 PATEL, J. et al. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, Elsevier Ltd, v. 42, n. 1, p. 259–268, 2015. ISSN 0957-4174. Citado 2 vezes nas páginas 20 e 72.
- 11 FAN, X.; LI, H.; ZHU, Z. A quantitative investment model based on multifractal theory and support vector machine. *Institute of Electrical and Electronic Engineers*, p. 239–244, 2014. Citado 2 vezes nas páginas 20 e 72.
- 12 CONT, R. Statistical modeling of high-frequency financial data. *IEEE Signal Processing Magazine*, IEEE, v. 28, p. 16–25, 2011. Citado na página 24.
- 13 STANLEY, H. E.; MANTEGNA, R. N. *An introduction to econophysics: correlations and complexity in finance*. [S.l.]: Cambridge University Press, 2000. ISBN 0 521 62008 2. Citado 3 vezes nas páginas 24, 31 e 110.

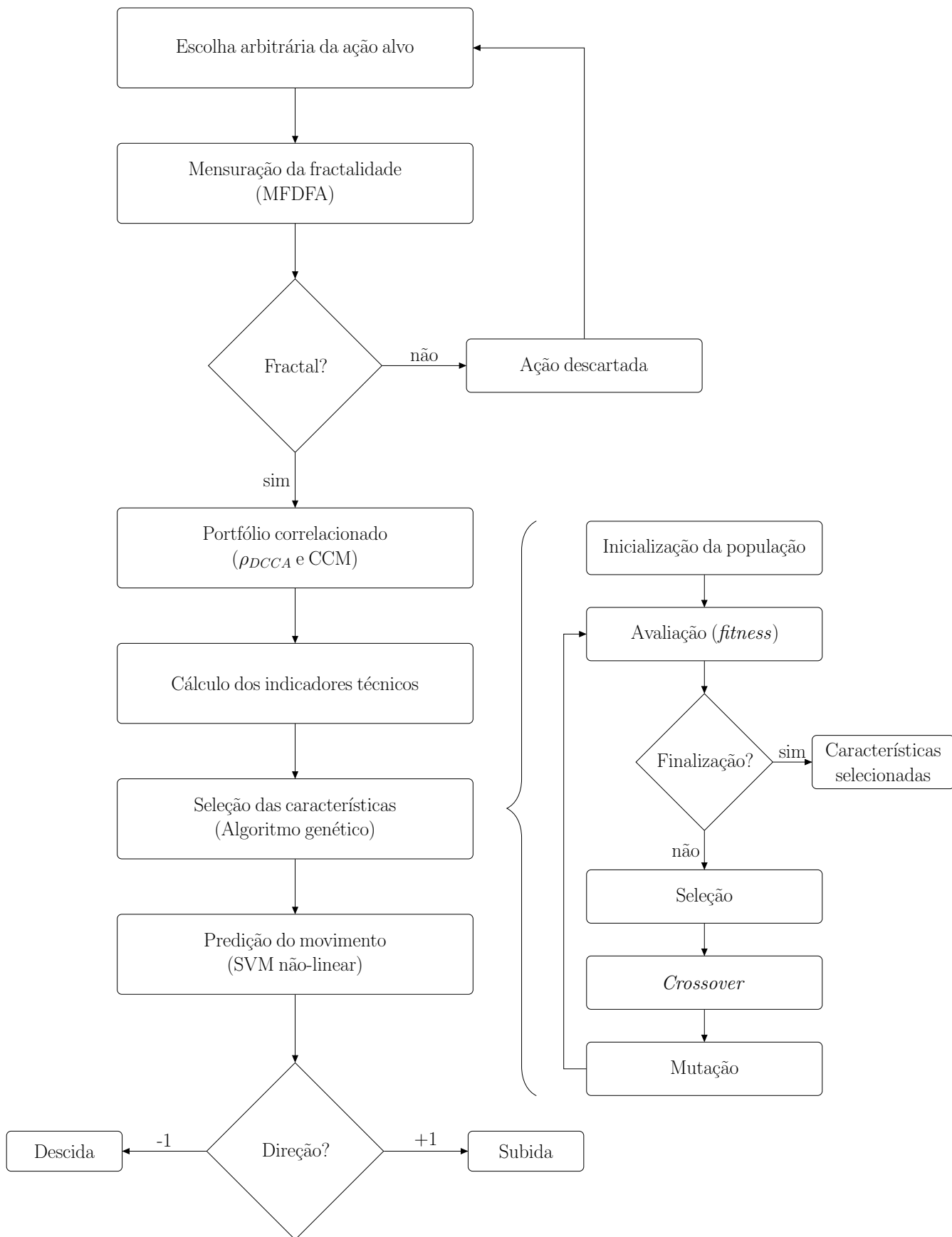
- 14 ERDOGAN, E. et al. Statistical models for unequally spaced time series. *Proceedings of the 2005 SIAM International Conference on Data Mining, SDM 2005*, n. April 2005, p. 626–630, 2005. Citado na página 26.
- 15 EPPS, T. W. Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, v. 74, n. 366, p. 291, 1979. ISSN 01621459. Citado na página 26.
- 16 TÓTH, B.; KERTÉSZ, J. The Epps effect revisited. *Quantitative Finance*, v. 9, n. 7, p. 793–802, 2009. ISSN 14697696. Citado na página 26.
- 17 FAMA, E. F. Efficient capital markets: a review of theory and empirical work. *The Journal of Finance*, v. 25, p. 383–417, 1969. Citado na página 30.
- 18 BOUCHAUD, J.-P.; POTTERS, M. *Theory of financial risks: from statistical physics to risk management*. [S.l.]: Cambridge University Press, 2000. 218 p. ISBN 0521782325. Citado 3 vezes nas páginas 31, 111 e 115.
- 19 CONT, R. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, v. 1, p. 223–236, 2001. Citado 5 vezes nas páginas 31, 108, 112, 115 e 117.
- 20 PETERS, E. E. *Fractal market analysis applying chaos theory to investment and economics*. [S.l.]: John Wiley & Sons, Inc., 1994. Citado 2 vezes nas páginas 31 e 33.
- 21 PETERS, E. E. *Chaos and order in the capital markets — a new view of cycles, prices, and market volatility*. [S.l.]: John Wiley & Sons, Inc., 1991. Citado na página 31.
- 22 KRISTOUFEK, L. Detrending moving-average cross-correlation coefficient: measuring cross-correlations between non-stationary series. *Physica A*, Elsevier B.V., v. 406, p. 169–175, 2014. ISSN 0378-4371. Citado na página 32.
- 23 WERON, A.; WERON, R. Fractal market hypothesis and two power-laws. v. 11, p. 289–296, 2000. Citado na página 33.
- 24 MANDELBROT, B. *The fractal geometry of nature*. [S.l.]: W.H. Freeman and Company, 1975. Citado na página 34.
- 25 MANDELBROT, B. How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science*, v. 156, n. 3775, p. 636–638, 1967. Citado na página 36.
- 26 MANDELBROT, B. New methods in statistical economics. *Journal of Political Economy*, v. 71, p. 421±440, 1963. Citado na página 37.
- 27 SALAT, H.; MURCIO, R.; ARCAUTE, E. Multifractal methodology. *Physica A*, Elsevier B.V., v. 473, p. 467–487, 2017. ISSN 0378-4371. Citado 2 vezes nas páginas 37 e 41.
- 28 KANTELHARDT, J. W. et al. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, v. 316, p. 87–114, 2002. Citado 2 vezes nas páginas 38 e 40.
- 29 OSSADNIK, S. M. et al. Correlation approach to identify coding regions in DNA sequences. *Biophysical Journal*, v. 67, n. February, p. 64–70, 1994. Citado na página 38.



- 30 FERREIRA, P.; DIONÍSIO, A.; MOVAHED, S. M. S. Assessment of 48 stock markets using adaptive multifractal approach. *Physica A*, v. 486, p. 730–750, 2017. Citado 2 vezes nas páginas 38 e 53.
- 31 KUMAR, A. et al. Multifractal detrended fluctuation analysis of alpha and theta EEG rhythms with musical stimuli. *Chaos, Solitons and Fractals*, Elsevier Ltd., v. 81, p. 52–67, 2015. ISSN 0960-0779. Citado 2 vezes nas páginas 38 e 42.
- 32 STANLEY, H. E.; MEAKIN, P. Multifractal phenomena in physics and chemistry. *Nature*, v. 335, n. 29, p. 405–409, 1988. Citado na página 38.
- 33 VICSEK, T.; BARABÁSI, A.-L. Multifractality of self-affine fractals. *Physical Review A*, v. 44, n. 4, p. 2730–2733, 1991. Citado na página 38.
- 34 SIMONS, M.; STANLEY, H. E.; GOLDBERG, A. L. Mosaic organization of DNA nucleotides. *Physical Review E*, v. 49, n. 2, p. 1685–1689, 1994. Citado na página 38.
- 35 HURST, H. E. C. M. G.; SC, D.; INST, F. P. The problem of long-term storage reservoirs. *Hydrological Sciences Journal*, v. 6024, 1956. Citado na página 39.
- 36 KANTELHARDT, J. W. *Fractal and multifractal time series*. [S.l.: s.n.], 2008. 1–59 p. Citado na página 40.
- 37 KARP, A.; Van Vuuren, G. Investment implications of the fractal market hypothesis. *Annals of Financial Economics*, v. 14, n. 01, p. 1950001, 2019. ISSN 2010-4952. Citado 2 vezes nas páginas 48 e 50.
- 38 FROMM, J. *The emergence of complexity*. [S.l.]: Kassel University, 2004. ISBN 3899580699. Citado na página 52.
- 39 PEARSON, K. Note on regression and inheritance in the case of two parents. *Royal Society*, v. 58, n. May, p. 240–242, 1895. Citado na página 52.
- 40 ZHAO, X.; SHANG, P.; HUANG, J. Several fundamental properties of DCCA cross-correlation coefficient. *Fractals*, v. 25, n. 2, p. 1–11, 2017. Citado 3 vezes nas páginas 52, 53 e 54.
- 41 STANLEY, B. P.; E., H. Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. *Physics Review Letters*, v. 100, n. 8, 2008. Citado na página 52.
- 42 ZEBENDE, G. F. DCCA cross-correlation coefficient: quantifying level of cross-correlation. *Physica A*, Elsevier B.V., v. 390, n. 4, p. 614–618, 2011. ISSN 0378-4371. Citado na página 53.
- 43 ZEBENDE, G. F.; SILVA, A. M. Detrended multiple cross-correlation coefficient. *Physica A*, Elsevier B.V., v. 510, p. 91–97, 2018. ISSN 0378-4371. Citado na página 53.
- 44 LORENZ, E. N. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, v. 20, p. 130–141, 1963. Citado na página 55.
- 45 RUELLE, D.; TAKENS, F. On the nature of turbulence. v. 20, n. 1948, p. 303–322, 1971. Citado na página 55.

- 46 TAKENS, F. Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence*, v. 898, p. 366, 1981. Citado na página 57.
- 47 DEYLE, E. R.; SUGIHARA, G. Generalized theorems for nonlinear state space reconstruction. v. 6, n. 3, 2011. Citado na página 57.
- 48 SUGIHARA, G. et al. Detecting causality in complex ecosystems. *Science*, v. 338, n. 6106, p. 496–500, 2012. ISSN 1549-6325; 1549-6333/e. Citado na página 59.
- 49 GRANGER, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, v. 37, n. 3, p. 424–438, 1969. Citado na página 59.
- 50 GRANGER, C. W. J. Spurious regressions in econometrics. *Journal of Econometrics*, v. 2, p. 111–120, 1974. Citado na página 59.
- 51 EBISUZAKI, W. A method to estimate the statistical significance of a correlation when the data are serially correlated. *Journal of Climate*, v. 10, n. 9, p. 2147–2153, 1997. ISSN 08948755. Citado na página 70.
- 52 VAPNIK, V.; CHERVONENKIS, A. Y. On the uniform convergence of relative frequencies of events to their probabilities. *Teor. Veroyatnost. i Primenen.*, v. 16, n. 2, p. 283–305, 1971. Citado na página 72.
- 53 HUANG, W.; NAKAMORI, Y.; WANG, S.-y. Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, v. 32, p. 2513–2522, 2005. Citado na página 72.
- 54 VAPNIK, V. *Statistical learning theory*. [S.l.]: John Wiley & Sons, Inc., 1998. 768 p. ISBN 0471030031. Citado na página 73.
- 55 VAPNIK, V. *The nature of statistical learning theory*. [S.l.]: Springer-Verlag New York, Inc., 1995. 314 p. ISBN 0387987800. Citado na página 74.
- 56 SMOLA, A. J.; SCHÖLKOPF., B. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. [S.l.]: The MIT Press, 2002. 648 p. ISBN 9780262194754. Citado 2 vezes nas páginas 75 e 83.
- 57 KHALID, S.; KHALIL, T.; NASREEN, S. A survey of feature selection and feature extraction techniques in machine learning. *Proceedings of 2014 Science and Information Conference, SAI 2014*, n. August 2014, p. 372–378, 2014. Citado na página 83.
- 58 KUMAR, M. et al. Genetic algorithm: review and application. *SSRN Electronic Journal*, v. 2, n. 2, p. 451–454, 2020. ISSN 1556-5068. Citado na página 85.
- 59 LEE, M. C. Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Systems with Applications*, Elsevier Ltd, v. 36, n. 8, p. 10896–10904, 2009. ISSN 09574174. Citado na página 92.
- 60 QIAN, B.; RASHEED, K. Hurst exponent and financial market predictability. *Proceedings of the Second IASTED International Conference on Financial Engineering and Applications*, p. 203–209, 2004. Citado na página 98.

## Apêndice A – Fluxograma - Modelo Quantitativo



## Apêndice B – Fatos Estilizados

Neste apêndice, serão apresentadas algumas das propriedades estatísticas conhecidas como fatos estilizados. A análise foi realizada nas séries temporais das quatro ações alvo deste trabalho no mesmo período de tempo utilizado no estudo da fractalidade (jan/2018 a nov/2018). Todos os retornos foram calculados pela equação 4 (*log*-retornos) a partir do preço de fechamento ajustado a proventos. Os resultados foram separados de acordo com as propriedades do retorno e da volatilidade.

### B.1 Propriedades do retorno

#### B.1.1 Ausência de autocorrelação (linear)

Autocorrelação é uma análise estatística que correlaciona dados de uma mesma série temporal com o intuito de encontrar padrões de repetição. A partir desta medição, é possível mensurar o quanto o valor da realização de uma variável aleatória é capaz de influenciar os seus vizinhos. Em outras palavras, dado um valor alto de uma variável o quanto o mesmo condiciona valores também altos em sua vizinhança.

Seja uma série temporal de retornos de uma ação  $\{r_t\}_{t=1}^T$ . O coeficiente de correlação entre  $r_t$  e  $r_{t-k}$  é dito de autocorrelação de  $k$ -ésima ordem e é denotado por:

$$\rho_k = \frac{Cov(r_t, r_{t-k})}{\sigma_{r_t} \cdot \sigma_{r_{t-k}}} = \frac{E[r_t r_{t-k}] - E[r_t]E[r_{t-k}]}{\sigma_{r_t} \cdot \sigma_{r_{t-k}}}, \quad (61)$$

onde  $E[\cdot]$  é o valor esperado e  $\sigma$  é o desvio padrão.

Por definição, a autocorrelação pode assumir valores no intervalo de  $[-1, 1]$ , onde:  $-1$  significa anti-correlação;  $1$ , correlação perfeita;  $0$ , ausência de correlação.

A função de autocorrelação de  $r_t$  é dada pelo conjunto de autocorrelações  $\{\rho_k\}$  e pode ser definida pela distância, ou atraso, com que se deseja medi-la. Quando a defasagem é zero, o valor será 1 uma vez que a variável está sendo correlacionada com ela mesma.

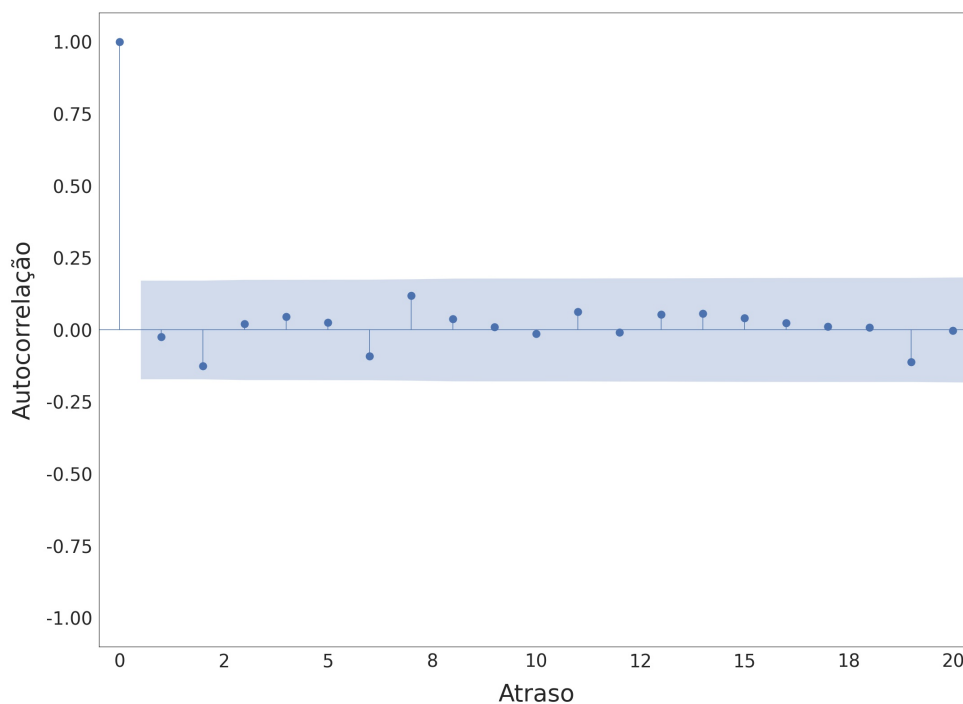
Em geral, a autocorrelação de  $k$ -ésima ordem de  $r_t$  pode ser definida como:

$$\rho_k = \frac{\sum_{t=k+1}^T (r_t - \bar{r}) \cdot (r_{t-k} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}, \quad (62)$$

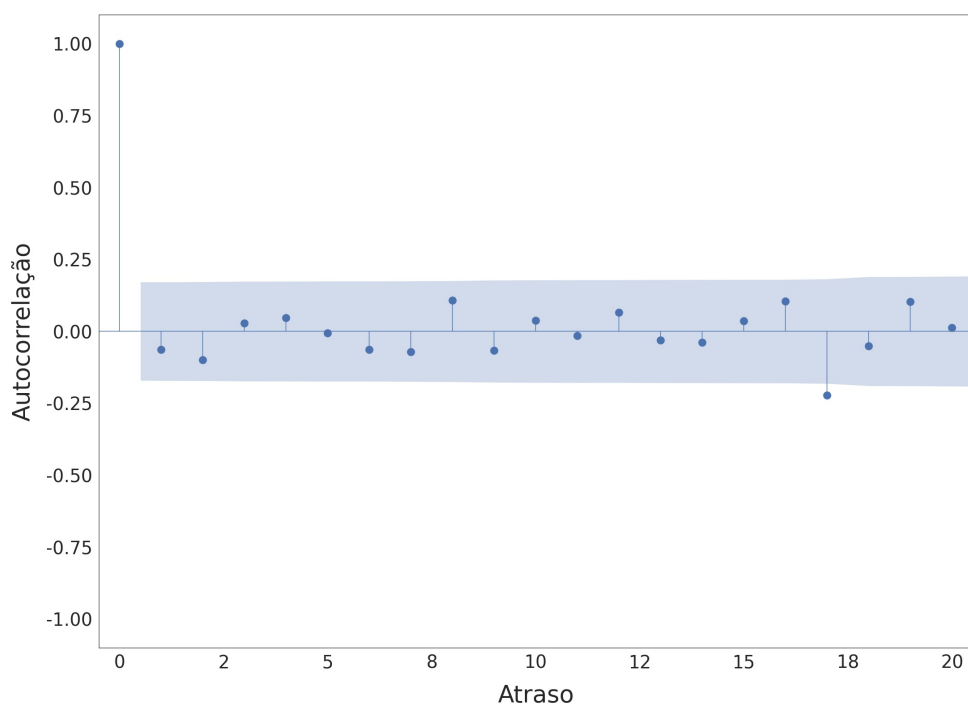
onde  $k$  é a defasagem e é dada por  $0 \leq k \leq T - 1$ .

Para retornos com escalas de tempo muito pequenas, abaixo de 20 minutos, foi constatado que os valores de autocorrelação são significativos [19]. Acima desse tempo, a autocorrelação tende rapidamente a valores próximos de zero.

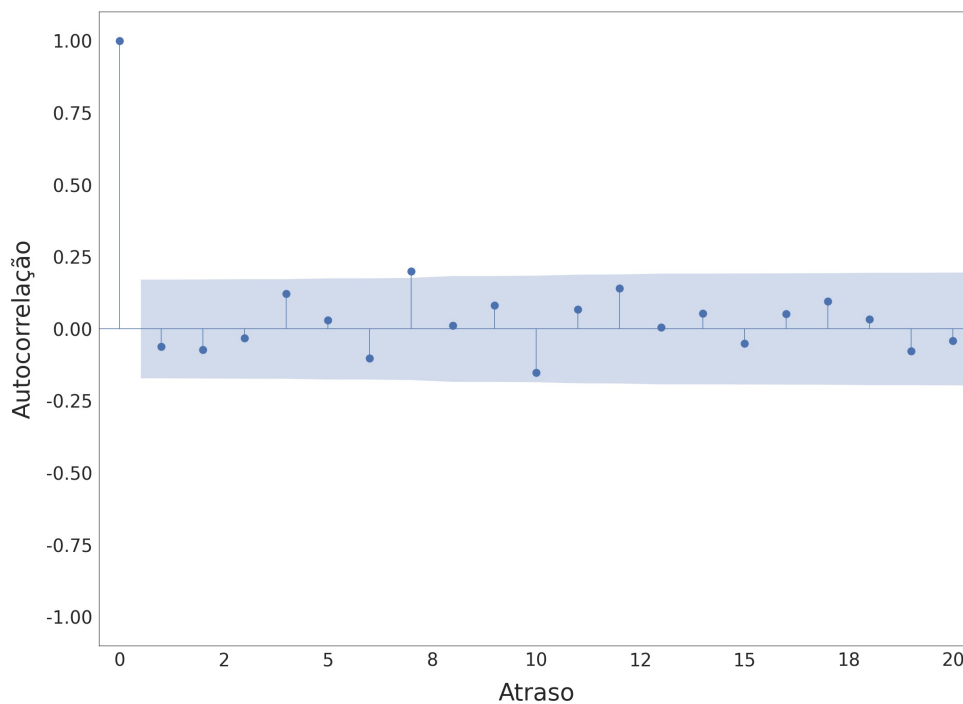
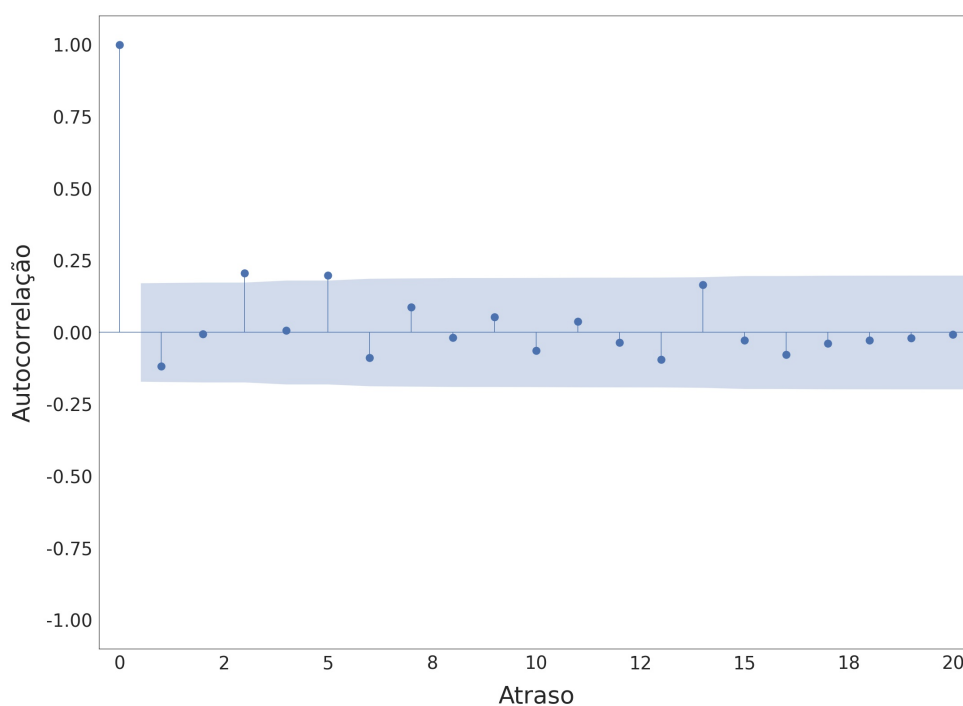
Figura 31 – Gráficos da autocorrelação dos *log*-retornos, com frequência diária, das ações alvo para diferentes atrasos ( $0 \leq k \leq 20$ ): (a) CYRE3, (b) EMBR3, (c) ITUB4 e (d) PETR4.



(a) CYRE3: autocorrelação dos *log*-retornos.



(b) EMBR3: autocorrelação dos *log*-retornos.

(c) ITUB4: autocorrelação dos *log*-retornos.(d) PETR4: autocorrelação dos *log*-retornos.

Fonte: Frederico Frias, 2021

A ausência de autocorrelação é verificada nos dados diários das ações alvo. É possível constatar na figura 31 que os resultados flutuam em torno de zero para defasagens de tempo iguais ou superiores a um dia.

### B.1.2 Caudas gordas

Quando uma distribuição de probabilidade exibe caudas maiores comparativamente à distribuição gaussiana, se diz que esta possui caudas gordas. Nestes casos, a curva da função de distribuição de probabilidade exibe um achatamento em relação à curva de distribuição normal.

Em estatística, uma medida que traduz este achatamento é a curtose, definida por:

$$k = \frac{\sum_{t=1}^T (r_t - \bar{r})^4}{\sigma_{r_t}^4} - 3. \quad (63)$$

Para uma distribuição mesocúrtica, distribuição gaussiana, a curtose é nula ( $k = 0$ ). Distribuições ditas leptocúrticas, isto é, distribuições que apresentam caudas gordas, o valor de  $k$  é maior do que zero.

Segundo [13], quanto maior a frequência dos dados maior é o grau de leptocurtose observado. É difícil de definir a forma exata da distribuição de probabilidade dos retornos, mas o aspecto de sua cauda é semelhante à de Pareto ou à de uma lei de potência, o que denota a probabilidade relativamente alta de se obter retornos anormais.

Todos os valores de curtose medidos, mostrados na tabela 18, são maiores que zero, indicando que a função de densidade de probabilidade de cada ativo apresenta um formato leptocúrtico (caudas gordas).

Tabela 18 – Valores de curtose dos *log*-retornos, com frequência diária, das ações alvo.

Ação	Curtose
CYRE3	2.10
EMBR3	7.59
ITUB4	0.31
PETR4	5.76

Fonte: Frederico Frias, 2021

### B.1.3 Assimetria ganho/perda

Em diversos tipos de ativos financeiros, como ações ou índices de ações, é observado que os retornos dos preços são mais acentuados nas perdas do que nos ganhos, com valores negativos mais altos do que os positivos. Esta propriedade estatística caracteriza um viés à

esquerda na curva de distribuição de probabilidade. Em outras palavras, a cauda esquerda da distribuição é mais longa do que a cauda direita, denotando uma maior concentração de retornos nesta região.

Pela equação 64, é possível mensurar a assimetria das caudas de uma distribuição de probabilidade:

$$s = \sum_{t=1}^T \frac{(r_t - \bar{r})^3}{\sigma_t^3}. \quad (64)$$

A distribuição normal é simétrica ( $s = 0$ ). Distribuições com assimetria negativa ( $s < 0$ ) possuem caudas mais pesadas à esquerda, enquanto distribuições positivas ( $s > 0$ ), caudas mais pesadas à direita.

Os resultados para cada um dos ativos, mostrados na tabela 19, estão de acordo com a propriedade de assimetria dos fatos estilizados. A maioria dos ativos, exceto CYRE3, apresentou valores negativos, o que evidencia perdas (retornos negativos) mais acentuadas do que ganhos (retornos positivos).

Tabela 19 – Valores de assimetria dos *log*-retornos, com frequência diária, das ações alvo.

Ação	Assimetria
CYRE3	0.45
EMBR3	-0.90
ITUB4	-0.07
PETR4	-0.80

Fonte: Frederico Frias, 2021

#### B.1.4 Gaussianidade agregativa

Diversos fenômenos físicos são modelados pela distribuição normal, também chamada de gaussiana. A vasta aplicabilidade desta distribuição de probabilidades é decorrente do teorema do limite central que afirma que a distribuição de variáveis aleatórias independentes tende a uma gaussiana a medida em que o tamanho da amostra aumenta [18].

A gaussianidade agregativa é a propriedade estatística, observada empiricamente, que mostra a convergência da distribuição de probabilidades para uma gaussiana a medida em que os intervalos de tempo considerados no cálculo dos retornos ficam cada vez maiores.



A função de densidade de probabilidade de uma distribuição gaussiana é dada por:

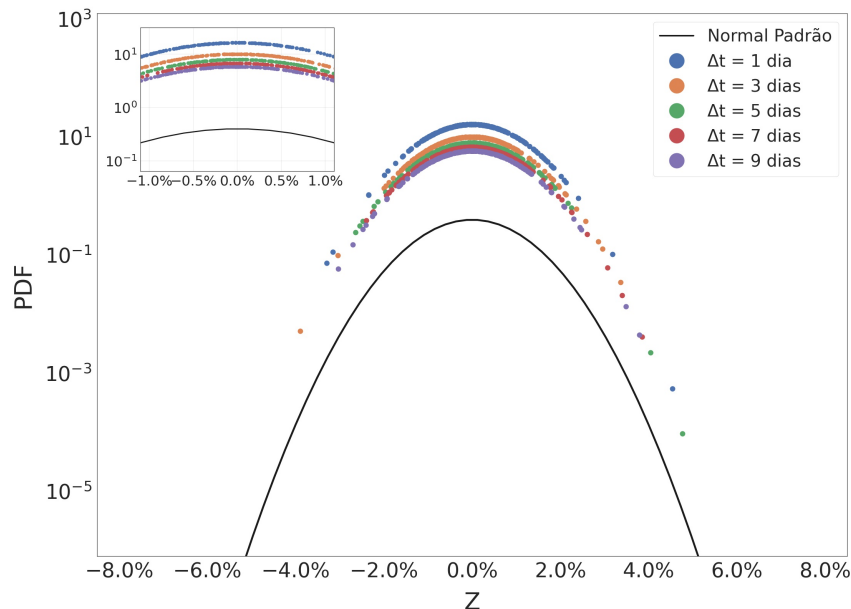
$$P(r_{t,\Delta t}) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(r_{t,\Delta t}-\mu)^2}{2\sigma^2}}, \quad (65)$$

onde a variável aleatória é o retorno do preço do ativo financeiro ( $r_{t,\Delta t}$ ) e  $\mu$  e  $\sigma$  são, respectivamente, a média e o desvio padrão dos retornos.

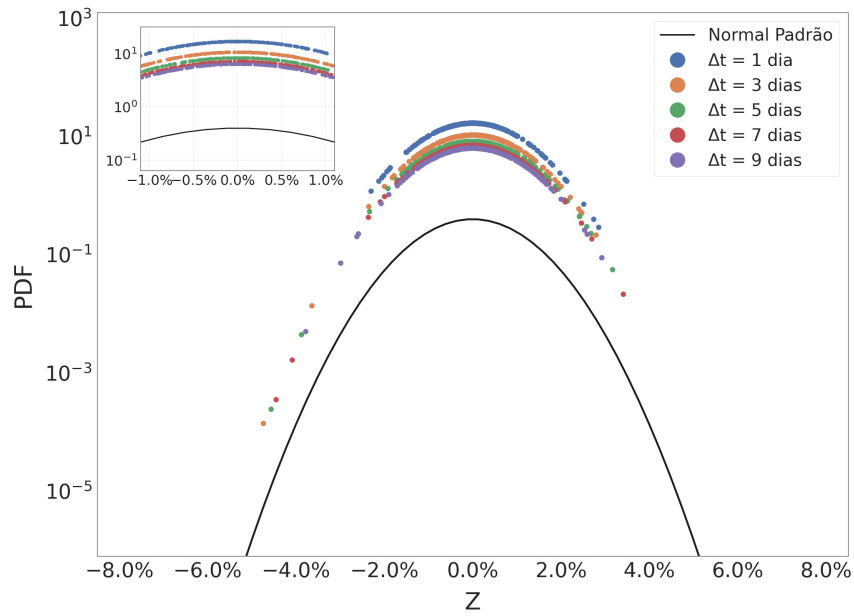
Para a constatação da propriedade da gaussianidade agregativa, deve-se gerar uma distribuição de probabilidade para diferentes intervalos de tempo  $\Delta t$  e compará-las com uma gaussiana normal padrão ( $\mu = 0$  e  $\sigma = 1$ ). Segundo esta propriedade, com o aumento do espaçamento do tempo utilizado no cálculo dos retornos, a distribuição de probabilidades dos dados empíricos tende a uma gaussiana [19].

Isto pode ser percebido ao analisar a figura 32: a medida que  $\Delta t$  fica maior, a distribuição dos retornos se aproxima da gaussiana normal padrão (linha em preto). Isso ocorre para as quatro ações alvo, evidenciando esta propriedade.

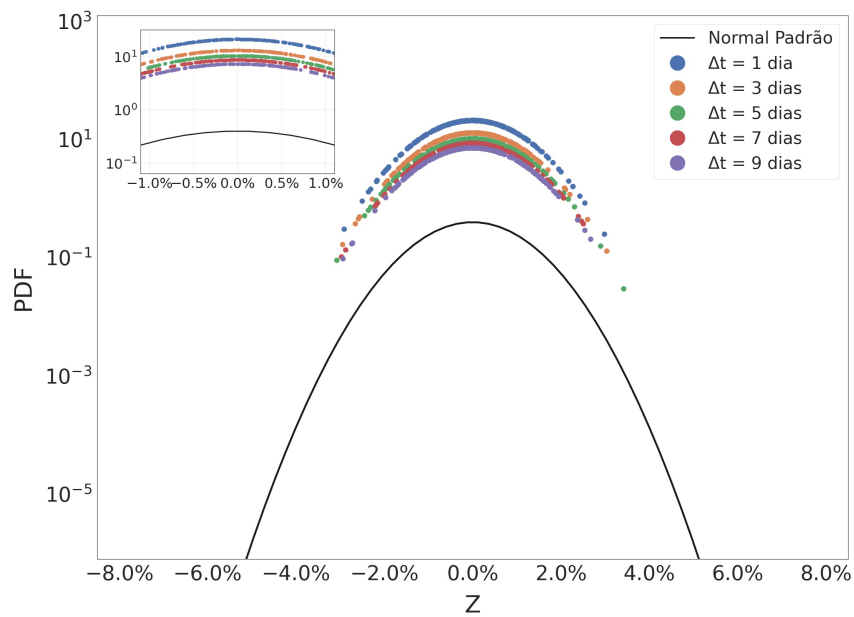
Figura 32 – Gráficos semi-*log* da função de densidade de probabilidade (PDF) dos *log*-retornos normalizados ( $Z$ ), com frequência diária, das ações alvo para diferentes intervalos de tempo ( $\Delta t$  1 a 9 dias): (a) CYRE3, (b) EMBR3, (c) ITUB4 e (d) PETR4.



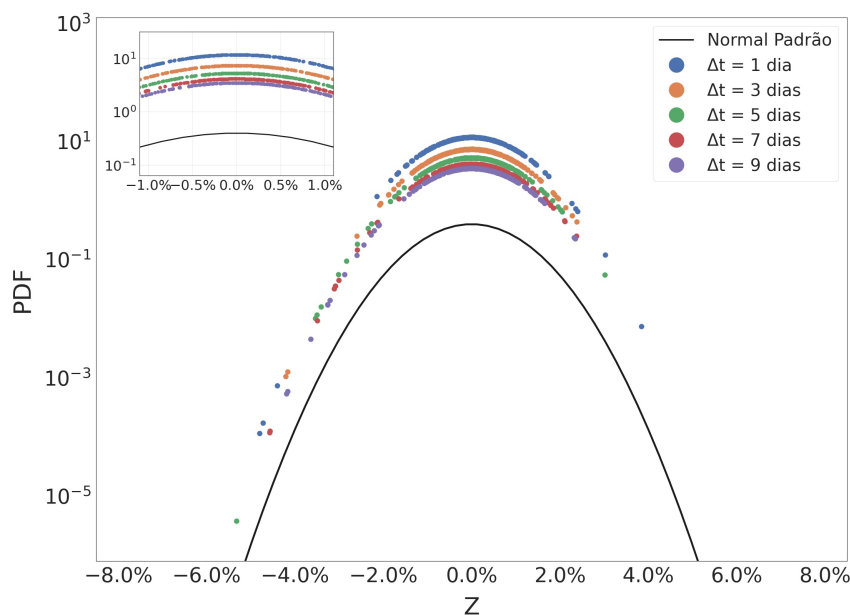
(a) CYRE3: função de densidade de probabilidade dos *log*-retornos ( $\Delta t$  de 1 a 9 dias).



(b) EMBR3: função de densidade de probabilidade dos *log*-retornos ( $\Delta t$  de 1 a 9 dias).



(c) ITUB4: função de densidade de probabilidade dos *log*-retornos ( $\Delta t$  de 1 a 9 dias).



(d) PETR4: função de densidade de probabilidade dos *log*-retornos ( $\Delta t$  de 1 a 9 dias).

Fonte: Frederico Frias, 2021

## B.2 Propriedades da volatilidade

### B.2.1 Agrupamento de volatilidade / Intermitência

A variação do preço de um dado ativo financeiro possui correlação de longo alcance no seu segundo momento, indicando a presença de um processo estocástico conhecido como volatilidade. Este processo é uma medida da dispersão dos retornos e não é diretamente observável. Existem alguns modelos utilizados para mensurar a volatilidade. Um dos estimadores mais simples é o desvio padrão do histórico de retornos considerando um período de tempo específico. Este modelo atribui pesos iguais a todas as observações, ao contrário do modelo de média móvel exponencialmente ponderada, do inglês, *exponentially weighted moving average* (EWMA), que é uma média móvel ponderada por um fator de decaimento exponencial, onde os retornos mais recentes possuem maior relevância. No caso de instrumentos financeiros não lineares, as ditas opções, a volatilidade está implícita no preço do ativo objeto, sendo calculada pela inversão da fórmula de *Black-Scholes*.

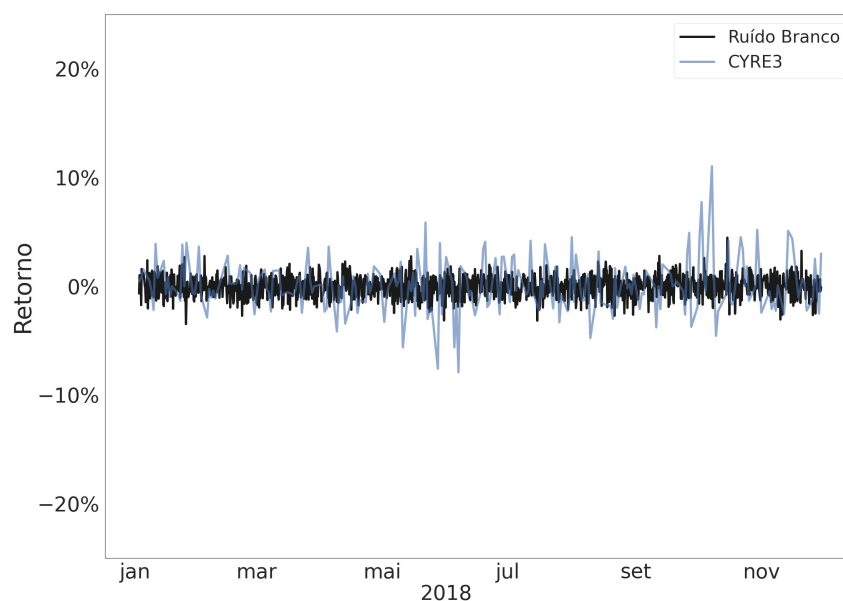
A volatilidade é uma medida de grande importância na mensuração do risco, além de ser crucial na modelagem de derivativos financeiros. Em momentos de incerteza, as flutuações do preço de um ativo apresentam níveis altos, indicando elevada volatilidade.

O inverso ocorre em períodos de calma, onde a volatilidade é baixa. Esta ruptura na intensidade dos retornos é denominada de intermitência [19]. Este comportamento ocorre quando a série temporal dos retornos sofre uma descontinuidade provocada por elevadas oscilações após um período de volatilidade baixa.

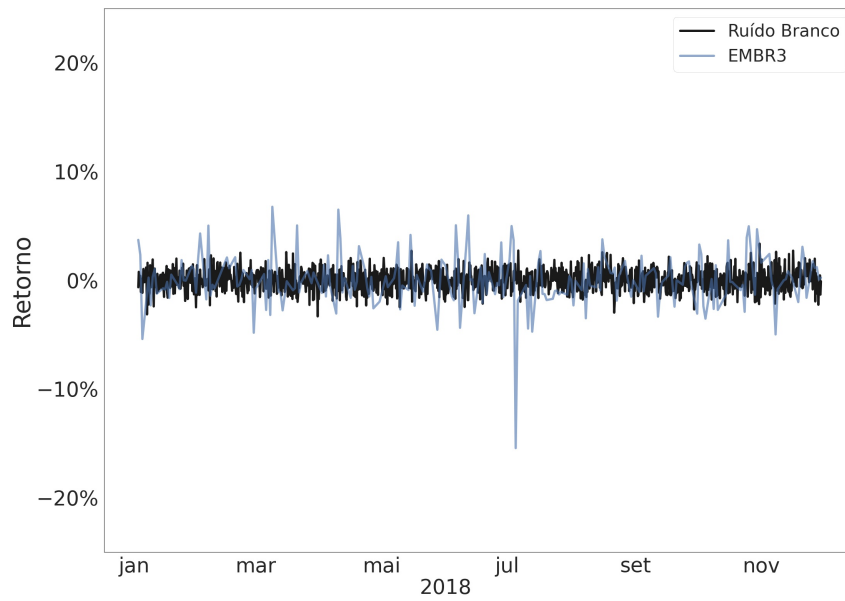
Estudos analíticos comprovam a existência de autocorrelação positiva em grandes intervalos de tempo consecutivos. A concentração de eventos de alta volatilidade é chamada de agrupamento de volatilidade [18]. Uma forma de se observar os grupos de alta volatilidade é comparando a série temporal de retornos do ativo estudado com um ruído branco gaussiano. Este último é um sinal aleatório, com distribuição normal com média zero, caracterizado por apresentar uma densidade espectral de potência constante, ou seja, possui a mesma intensidade em diferentes frequências.

Na figura 33, é possível observar a presença de regiões com concentrações de picos (alta volatilidade) em comparação ao sinal do ruído branco, comprovando a existência da propriedade de agrupamento de volatilidade nas séries temporais estudadas. Além disso, os retornos possuem valores de alta intensidade, indicando um período de incerteza, como também valores de baixa intensidade, denotando períodos de relativa calma, sendo possível constatar a propriedade conhecida com intermitência.

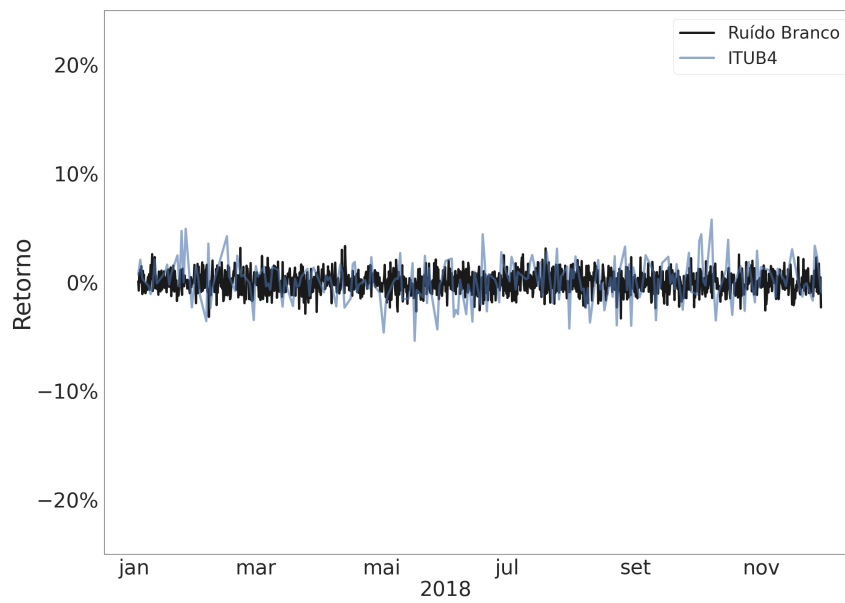
Figura 33 – Gráficos dos *log*-retornos, com frequência diária, das ações alvo comparados com um sinal de ruído branco gaussiano: (a) CYRE3, (b) EMBR3, (c) ITUB4 e (d) PETR4.



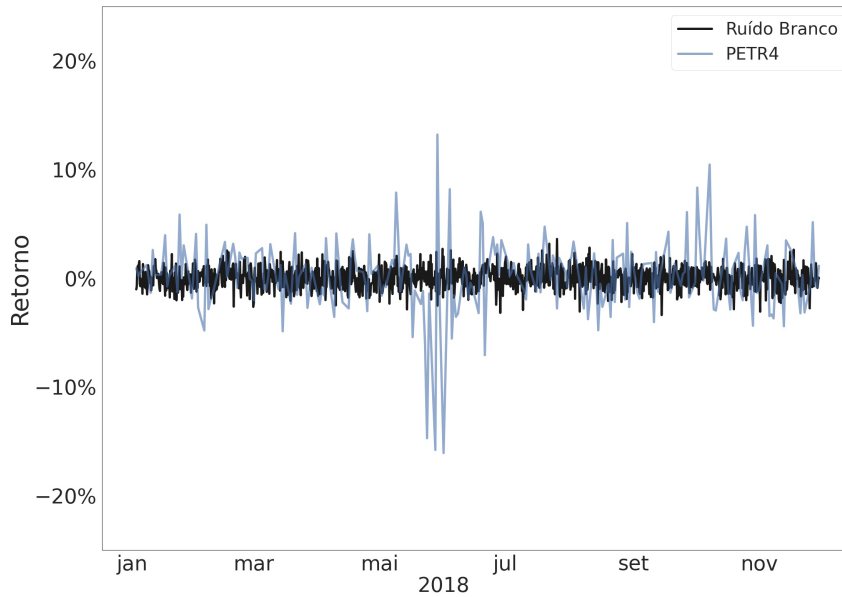
(a) CYRE3: *log*-retornos versus ruído branco.



(b) EMBR3: *log-retornos versus ruído branco.*



(c) ITUB4: *log-retornos versus ruído branco.*

(d) PETR4: *log*-retornos *versus* ruído branco.

Fonte: Frederico Frias, 2021

### B.2.2 Efeito de alavancagem

O efeito de alavancagem é a observação da correlação negativa entre a volatilidade e o retorno de um ativo. Este efeito ocorre na maioria das medições de volatilidade, o que caracteriza altas flutuações no preço do papel em momentos em que o mesmo está desvalorizado, possivelmente em consequência do aumento da alavancagem financeira da empresa. Segundo [19], este efeito pode ser medido por:

$$L(\tau) = \text{corr}(|r_{t+\tau, \Delta t}|^2, r_{t, \Delta t}). \quad (66)$$

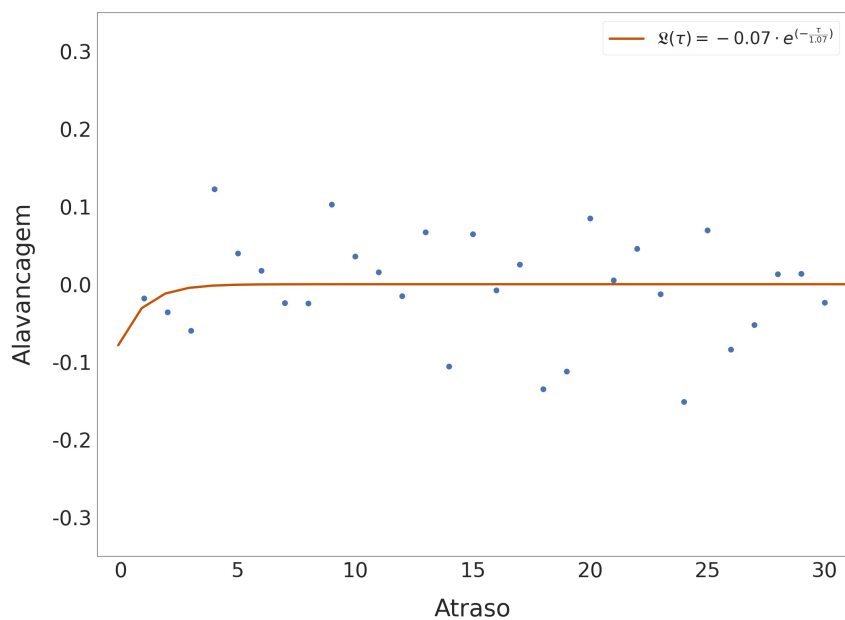
A dependência dos retornos é não linear, como pode ser observado pela equação 66. Inicialmente, a correlação assume valores negativos, tendendo a zero a medida que a defasagem ( $\tau$ ) de tempo aumenta. A interpretação é de que retornos negativos implicam no aumento da volatilidade e retornos positivos, a diminuem. Contudo, não é possível afirmar o mesmo no sentido inverso (a volatilidade alavancando o retorno), uma vez que  $L(\tau)$  é assimétrico, sendo insignificante para  $\tau < 0$ .

A função de correlação pode ser ajustada por uma exponencial da seguinte forma:

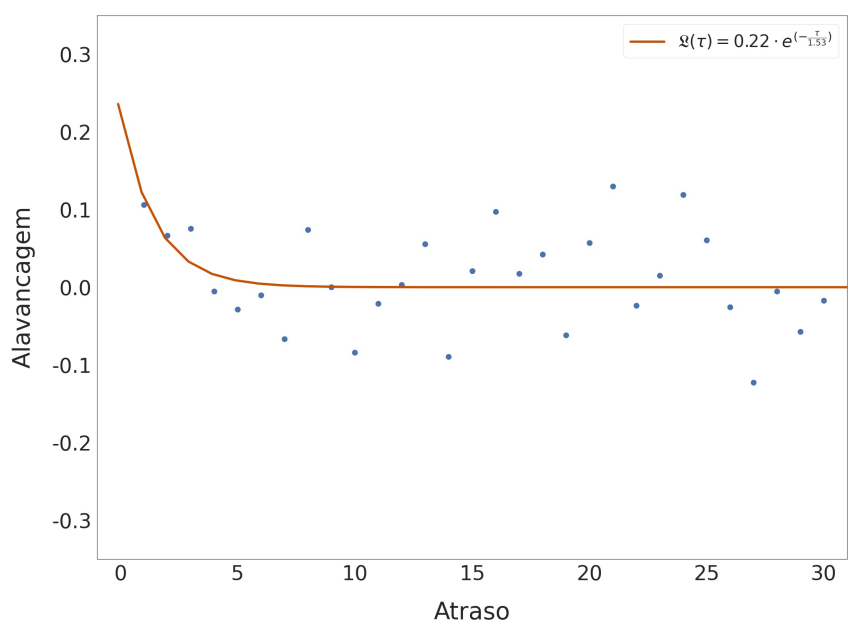
$$L(\tau) = -A \cdot e^{\frac{\tau}{T}}, \quad (67)$$

onde  $A$  é a amplitude,  $\tau$  é o atraso de tempo e  $T$  é o período.

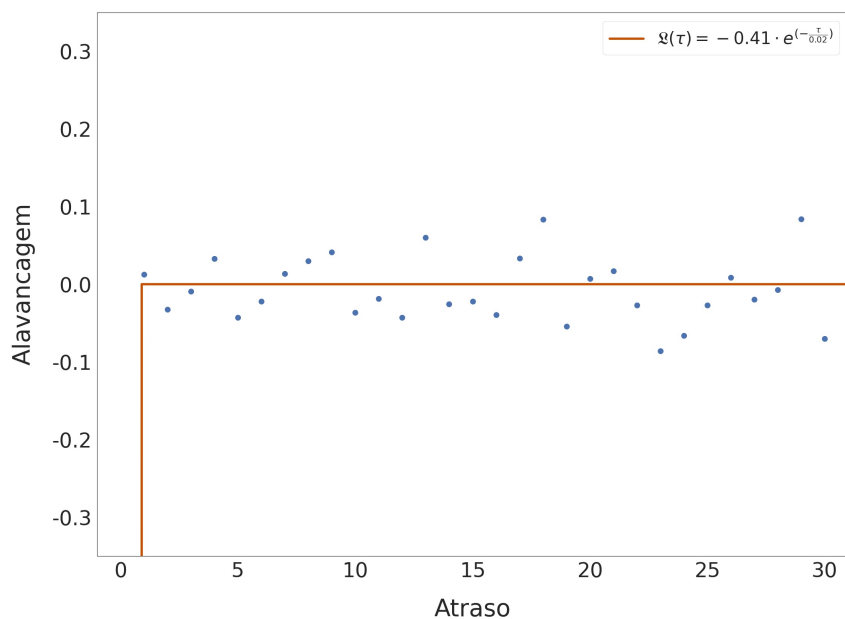
Figura 34 – Gráficos do efeito de alavancagem dos *log*-retornos, com frequência diária, das ações alvo ajustados por uma curva exponencial: (a) CYRE3, (b) EMBR3, (c) ITUB4 e (d) PETR4.



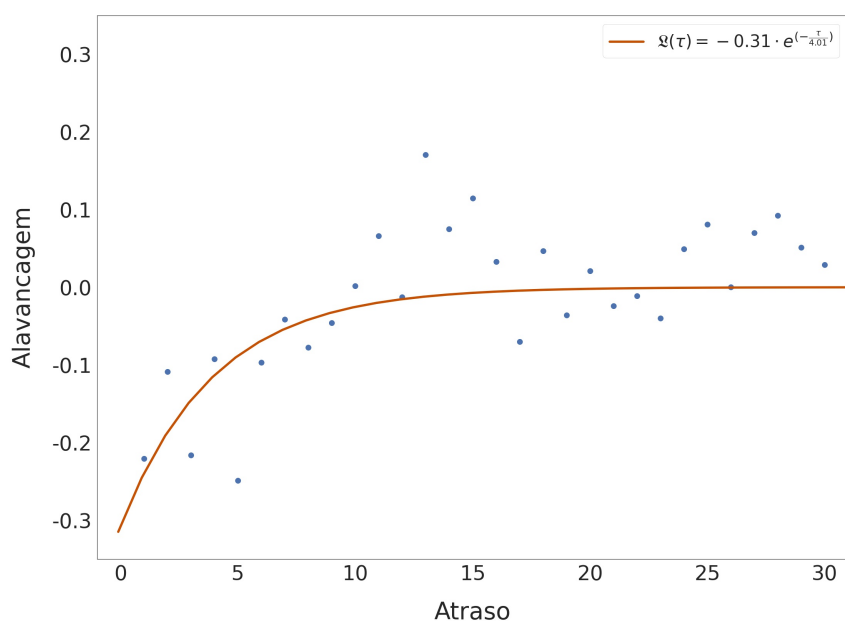
(a) CYRE3: efeito de alavancagem ajustado por uma curva exponencial.



(b) EMBR3: efeito de alavancagem ajustado por uma curva exponencial.



(c) ITUB4: efeito de alavancagem ajustado por uma curva exponencial.



(d) PETR4: efeito de alavancagem ajustado por uma curva exponencial.

Fonte: Frederico Frias, 2021

A figura 34 evidencia esta propriedade, ao verificar-se que as correlações têm valores iniciais negativos (exceto a EMBR3), decaindo a zero após um certo período de tempo  $T$ . Os tempos de decaimento são curtos, chegando no máximo a quatro dias (caso da PETR4). Além disso, as amplitudes  $A$  dos quatro ajustes estão bem próximas umas das outras.