

INSTITUTO POLITÉCNICO DE LISBOA
INSTITUTO SUPERIOR DE CONTABILIDADE E ADMINISTRAÇÃO
DE LISBOA



ISCAL

A CERTEZA DA INCERTEZA NO MERCADO
DE ETF'S – ANÁLISE COMPARATIVA DO
MODELO GARCH E ALGORITMO DE
INTELIGÊNCIA ARTIFICIAL XGBOOST

Bruno Miguel Lourenço Fernandes

Lisboa, junho de 2024

INSTITUTO POLITÉCNICO DE LISBOA
INSTITUTO SUPERIOR DE CONTABILIDADE E ADMINISTRAÇÃO DE
LISBOA

A CERTEZA DA INCERTEZA NO MERCADO DE
ETF'S – ANÁLISE COMPARATIVA DO MODELO
GARCH E ALGORITMO DE INTELIGÊNCIA
ARTIFICIAL XGBOOST

Bruno Miguel Lourenço Fernandes

Dissertação submetida ao Instituto Superior de Contabilidade e Administração de Lisboa para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Análise Financeira, realizada sob a orientação científica de Prof. Doutor José Nuno Sacadura, Professor Adjunto, da área de Economia/Finanças.

Constituição do Júri:

Presidente: Doutor Joaquim Paulo de Carvalho

Arguente: Doutora Sónia Bentes

Vogal: Doutor José Nuno Sacadura

Lisboa, junho de 2024

Declaro ser o autor desta dissertação, que constitui um trabalho original e inédito, que nunca foi submetido (no seu todo ou qualquer das suas partes) a outra instituição de ensino superior para obtenção de um grau académico ou outra habilitação. Atesto ainda que todas as citações estão devidamente identificadas. Mais acrescento que tenho consciência de que o plágio – a utilização de elementos alheios sem referência ao seu autor – constitui uma grave falta de ética, que poderá resultar na anulação da presente dissertação

Agradecimentos

Ao professor, Doutor José Nuno Sacadura, orientador desta dissertação, pela sua disponibilidade e sentido crítico bem como pela partilha de conhecimento, inclusive nas suas aulas, que contribuíram quer para o desenvolvimento deste trabalho como também para que conseguisse desenvolver novas competências profissionais.

Aos professores, Doutor Paulo Viegas Ferreira de Carvalho, Doutor Fernando Manuel Monteiro Martim, Doutor António Gil Margalho Nogueira e Doutor Joaquim António Martins Ferrão, agradeço a forma como transmitiram os conhecimentos e partilharam experiências com toda a turma.

Aos meus colegas de mestrado, sobretudo ao Pedro, Afonso e Leonardo, com quem tive o prazer de trabalhar, conviver e partilhar experiências, memórias que certamente irão permanecer ao longo da vida.

Aos meus familiares que me apoiaram incondicionalmente e tornaram possível que me dedicasse quase em exclusivo ao mestrado, sem eles não seria possível.

Por fim um agradecimento especial à minha equipa do Bankinter e ao meu Diretor que estiveram sempre disponíveis.

Resumo

Levar a cabo decisões de investimento acertadas é extremamente importante para um investidor ou gestor de investimentos. Deste modo os investigadores têm-se debruçado com maior foco, na última década, na realização de estudos sobre a volatilidade dos instrumentos financeiros. Estes, desempenham um papel de extrema relevância na definição da estratégia de negociação ou até na determinação do momento apropriado para negociar, mas sobretudo na caracterização e análise de risco de um determinado valor mobiliário. À medida que a tecnologia vai avançando, surgem ferramentas e modelos que auxiliam os analistas. É nesse sentido que a motivação para esta investigação nasce, cujo intuito é a abordagem de diferentes estratégias, aplicando o modelo heterocedástico GARCH e o algoritmo de *Machine Learning* XGBOOST, a series temporais relacionadas com *Exchange Traded Funds* (ETF's) de mercados emergentes e desenvolvidos, de modo a prever a volatilidade destes ativos financeiros, durante o início da invasão russa ao território da Ucrânia, com recurso à linguagem de programação *Python* para uma amostra de dados cujo horizonte temporal é de 01 de janeiro de 2012 a 24 de abril de 2022 e fonte é o Yahoo Finance. Os resultados apresentados, com especial foco no algoritmo XGBOOST, sugerem que a utilização de algoritmos de *Machine Learning* permite alcançar métricas de erro substancialmente inferiores, comparativamente ao modelo GARCH. Paralelamente é possível constatar igualmente que a volatilidade nos mercados emergentes é superior à apresentada nos mercados desenvolvidos, resultado do reduzido desenvolvimento das empresas, insegurança dos investidores e o peso que os eventos adversos, com origem em países desenvolvidos, têm sobre estes mercados.

Palavras-chave: ETF's, mercado, previsão, volatilidade

Abstract

Making the right investment decisions is extremely important for an investor or investment manager. Therefore, researchers have focused more intensively over the past decade on conduction studies on the volatility of financial instruments. These play an extremely relevant role in defining trading strategy or even determining the appropriate timing for trading, but above all in characterizing and analyzing the risk of a particular security. As technology advances, tools and models emerge to assist analysts. It is in this sense that the motivation for this research arises, with the aim of approaching different strategies, applying the GARCH heteroskedastic model and the XGBOOST Machine Learning algorithm to time series related to ETF's from emerging and developed markets, in order to predict the volatility of these financial assets during the onset of the Russian invasion of Ukrainian territory, using the Python programming language for a data sample with a time horizon from January 1, 2012, to April 24, 2022 sourced from Yahoo Finance. The results presented with special focus on the XGBOOST algorithm, suggest that the use of machine learning algorithms allow us to achieve substantially lower error metrics compared to the GARCH model. At the same time, it is also possible to observe that volatility in emerging markets is higher than that in developed markets, because of reduced company development, investor insecurity, and the impact that adverse events, originating from developed countries, have on these markets.

Key Words: ETF's, forecasting, market, volatility

Índice

Agradecimentos.....	v
Resumo.....	vi
Abstract	vii
Índice	viii
Índice de Figuras.....	x
Índice de Tabelas	xii
Lista de Abreviaturas.....	xiii
1. Introdução.....	1
1.1 Questão de Investigação.....	3
1.2 Objetivo Geral.....	3
1.2.1 Objetivos Específicos.....	3
1.2.2 Objetivos Teóricos.....	3
1.2.3 Objetivos Metodológicos.....	3
1.2.4 Objetivos Práticos.....	4
1.3 Hipóteses.....	4
1.4 Justificação e Relevância do Estudo.....	4
2. Revisão da Literatura.....	6
2.1 Mercados Financeiros.....	6
2.1.1 Mercados Desenvolvidos.....	6
2.1.2 Mercados Emergentes.....	7
2.2 Sentimentos do Investidor.....	7
2.3 Crises Financeiras.....	8
2.4 Volatilidade dos Retornos Financeiros.....	9
2.5 Heterocedasticidade Condicional.....	12
2.6 Modelos de Regressão.....	13
2.6.1 Modelo ARCH.....	13
2.6.2 Modelo GARCH.....	15
2.6.3 Algoritmo XGBOOST.....	17
2.7 Machine Learning.....	17
2.8 Critérios de Seleção de Variáveis.....	18
2.8.1 AIC.....	18
2.8.2 BIC.....	18

2.9 Métricas de Avaliação da Performance dos Modelos	19
2.9.1 MAPE	19
2.9.2 RMSE.....	20
3. Método e Processo de Investigação.....	21
4. Descrição dos dados.....	22
4.1 Produtos selecionados	22
4.1.1 SPDR S&P 500 ETF Trust.....	22
4.1.2 iShares Core S&P 500 ETF	23
4.1.3 Vanguard Total Stock Market Index Fund ETF Shares.....	25
4.1.4 Vanguard Emerging Markets Stock Index Fund.....	26
4.1.5 iShares Core MSCI Emerging Markets ETF.....	27
4.1.6 Vanguard Total International Bond Index Fund	28
4.2 Periodicidade dos dados	29
4.3 Seleção do modelo de estimação.....	29
5. Implementação, Análise e Resultados: Estudo Empírico.....	29
5.1 Implementação e Análise - SPDR S&P 500 ETF Trust.....	30
5.1.1 Resultados: SPDR S&P 500 ETF Trust	45
5.1.2 Resultados: iShares Core S&P 500 ETF	45
5.1.3 Resultados: Vanguard Total Stock Market Index Fund ETF Shares	46
5.2 Implementação e Análise - Vanguard Emerging Markets Stock Index Fund.....	46
5.2.1 Resultados: Vanguard Emerging Markets Stock Index Fund.....	49
5.2.2 Resultados: iShares Core MSCI Emerging Markets ETF	50
5.3 Análise Vanguard Total International Bond Index Fund.....	50
5.3.1 Resultados: Vanguard Total International Bond Index Fund	53
6. Conclusões.....	54
6.1 Conclusões Gerais.....	54
6.2 Estudos futuros	55
Referências Bibliográficas.....	56

Índice de Figuras

Figura 1.1 - N° de citações e publicações por ano referentes inteligência artificial aplicada aos mercados de capitais.....	5
Figura 4.1 - Sectores - SPDR S&P 500 ETF Trust.....	23
Figura 4.2 - Top 9 Empresas - SPDR S&P 500 ETF Trust.....	23
Figura 4.3 - Sectores - iShares Core S&P 500 ETF.....	24
Figura 4.4 - Top 9 Empresas - iShares Core S&P 500 ETF.....	24
Figura 4.5 - Sectores - Vanguard Total Stock Market Index Fund ETF Shares.....	25
Figura 4.6 - Top 9 Empresas - Vanguard Total Stock Market Index Fund ETF Shares.....	25
Figura 4.7 - Sectores - Vanguard Emerging Markets Stock Index Fund.....	26
Figura 4.8 - Top 7 Empresas - Vanguard Emerging Markets Stock Index Fund.....	26
Figura 4.9 - Sectores - iShares Core MSCI Emerging Markets ETF.....	27
Figura 4.10 - Top 9 Empresas - iShares Core MSCI Emerging Markets ETF.....	27
Figura 4.11 - Sectores - Vanguard Total International Bond Index Fund.....	28
Figura 4.12 - Bond Ratings - Vanguard Total International Bond Index Fund.....	28
Figura 5.1 - 5 Primeiros registos da amostra.....	30
Figura 5.2 - 5 Últimos registos da amostra.....	30
Figura 5.3 - Estatística descritiva, Período Estável - SPDR S&P 500 ETF Trust.....	31
Figura 5.4 - Gráfico Rendimentos Acumulados - SPDR S&P 500 ETF Trust.....	32
Figura 5.5 - Gráfico Retornos Diários - SPDR S&P 500 ETF Trust.....	32
Figura 5.6 - Gráfico Distribuição Retornos Diários - SPDR S&P 500 ETF Trust.....	33
Figura 5.7 - Gráfico Autocorrelação Parcial do Quadrado dos Retornos - SPDR S&P 500 ETF Trust.....	33
Figura 5.8 - Gráfico GARCH vs Volatilidade Continua dos Retornos Diários (treino) - SPDR S&P 500 ETF Trust - Período Estável.....	35
Figura 5.9 - Gráfico Previsão Volatilidade - SPDR S&P 500 ETF Trust (PERÍODO ESTÁVEL) - Modelo GARCH.....	35
Figura 5.10 - Variáveis Adicionadas - SPDR S&P 500 ETF Trust (PERÍODO ESTÁVEL) - Algoritmo XGBOOST.....	36
Figura 5.11 - Top Importância das Variáveis - SPDR S&P 500 ETF Trust - Algoritmo XGBOOST (Período Estável).....	36
Figura 5.12 - Gráfico XGBOOST VS Volatilidade Continua dos Retornos Diários (treino) - SPDR S&P 500 ETF TRUST - PERÍODO ESTÁVEL.....	37

Figura 5.13 - Gráfico Previsão Volatilidade - SPDR S&P 500 ETF TRUST (Período Estável) - Algoritmo XGBOOST	37
Figura 5.14 - 5 Últimos Registos da Amostra (Período de Instabilidade)	38
Figura 5.15 - Estatística Descritiva, Período de Instabilidade - SPDR S&P 500 ETF TRUST	38
Figura 5.16 - Gráfico Rendimentos Acumulados - SPDR S&P 500 ETF TRUST - Período de Instabilidade.....	39
Figura 5.17 - Gráfico Retornos Diários - SPDR S&P 500 ETF TRUST (Período de Instabilidade)	40
Figura 5.18 - Gráfico Distribuição Retornos Diários - SPDR S&P 500 ETF TRUST (Período de Instabilidade)	40
Figura 5.19 - Gráfico Autocorrelação Parcial do Quadrado dos Retornos - SPDR S&P 500 ETF TRUST (Período de Instabilidade)	40
Figura 5.20 - Gráfico GARCH VS Volatilidade Continua dos Retornos Diários (treino) - SPDR S&P 500 ETF TRUST - Período de Instabilidade.....	41
Figura 5.21 - Gráfico Previsão Volatilidade - SPDR S&P 500 ETF TRUST (Período de Instabilidade) - MODELO GARCH.....	42
Figura 5.22 - Variáveis Adicionadas - SPDR S&P 500 ETF TRUST (Período de Instabilidade) - Algoritmo XGBOOST.....	43
Figura 5.23 - Top Importância das Variáveis - SPDR S&P 500 ETF TRUST - ALGORITMO XGBOOST (Período de Instabilidade)	43
Figura 5.24 - Gráfico XGBOOST VS Volatilidade Continua dos Retornos Diários (treino) - SPDR S&P 500 ETF TRUST - Período de Instabilidade.....	44
Figura 5.25 - Gráfico Previsão Volatilidade - SPDR S&P 500 ETF TRUST (Período de Instabilidade) - Algoritmo XGBOOST.....	44
Figura 5.26 - 5 Primeiros Registos da Amostra - VANGUARD EMERGING MARKETS STOCK INDEX FUND.....	47
Figura 5.27 - 5 Últimos Registos da Amostra - VANGUARD EMERGING MARKETS STOCK INDEX FUND.....	47
Figura 5.28 - Gráfico Retornos Diários Vanguard Total International Bond Index Fund.....	51
Figura 5.29 - Gráficos Rendimentos Acumulados - Vanguard Total International Bond Index Fund.....	51
Figura 5.30 - Gráfico Distribuição Retornos Diários Vanguard Total International Bond Index Fund.....	52
Figura 5.31 - Gráfico Autocorrelação Parcial do Quadrado dos Retornos - VANGUARD TOTAL INTERNATIONAL BOND INDEX FUND.....	52

Índice de Tabelas

Tabela 4.1 -Seleção de ETF'S.....	22
Tabela 5.1 - Volatilidade SPDR S&P 500 ETF Trust (PERÍODO ESTÁVEL)	32
Tabela 5.2 - Parâmetros Estimados - GARCH (1,1) VS GARCH (5,5) - SPDR S&P 500 ETF Trust (Período Estável).....	34
Tabela 5.3 - Volatilidade SPDR S&P 500 ETF TRUST (Período de Instabilidade)	39
Tabela 5.4 - Parâmetros Estimados - GARCH (1,1) VS GARCH (4,4) - SPDR S&P 500 ETF TRUST (Período de Instabilidade).....	41
Tabela 5.5 - Resultados Métricas MAPE E RMSE - SPDR S&P 500 ETF TRUST.....	45
Tabela 5.6 - Resultados Métricas MAPE E RMSE - ISHARES CORE S&P 500 ETF.....	45
Tabela 5.7 - Resultados Métricas MAPE E RMSE - VANGUARD TOTAL STOCK MARKET INDEX FUND ETF SHARES	46
Tabela 5.8 - Volatilidade Vanguard Emerging Markets Stock Index Fund (Período Estável). 48	
Tabela 5.9 - Parâmetros Estimados - GARCH (1,1) - VANGUARD EMERGING MARKET'S STOCK INDEX FUND (Período Estável).....	48
Tabela 5.10 - Volatilidade VANGUARD EMERGING MARKET'S STOCK INDEX FUND (Período de Instabilidade)	49
Tabela 5.11 - Parâmetros Estimados - GARCH (1,1) - VANGUARD EMERGING MARKET'S STOCK INDEX FUND (Período de Instabilidade)	49
Tabela 5.12 - Resultados Métricas MAPE E RMSE - Vanguard Emerging Markets Stock Index Fund.....	49
Tabela 5.13 - Resultados Métricas MAPE E RMSE - iShares Core MSCI Emerging Markets ETF.....	50
Tabela 5.14 - Volatilidade Vanguard Total International Bond Index Fund.....	51
Tabela 5.15 - Parâmetros Estimados - GARCH (1,1) - Vanguard Total International Bond Index Fund.....	522
Tabela 5.16 - Resultados Métricas mape e rmse - Vanguard Total International Bond Index Fund.....	53

Lista de Abreviaturas

AIC. - Akaike Information Criterion

ARCH. - Autoregressive Conditional Heteroscedasticity

ARMA. - Auto Regressive Moving Average

AUM's. - Assets Under Management

BIC. - Bayesian Information Criterion

COVID. - Coronavirus

ETF's. - Exchange Traded Funds

GARCH. - Generalized Autoregressive Conditional Heteroscedasticity

GBDT. - Gradient Boosted Decision Trees

GED. - Generalized Error Distribution

MAPE. - Mean Absolute Percentage Error

RMSE. - Root Mean Squared Error

XGBOOST. - Extreme Gradient Boosting

1. Introdução

Desde o início da sociedade que analisar informações históricas é uma das principais fontes de sabedoria para os líderes e sábios da humanidade. Este conceito aplica-se igualmente às ações porque existe uma tendência de que as ações apresentam características semelhantes ao longo do tempo.

A previsão do valor de mercado é de extrema importância para ajudar a maximizar o lucro da compra de opções de ações, mantendo o risco em níveis mais reduzidos. No entanto, a complexidade e o dinamismo das bolsas de valores tornam a previsão de comportamento dos preços de mercado de ações uma tarefa difícil. Inerente a esta temática surge o conceito de volatilidade, considerada por muitos como um dos fatores críticos no domínio das finanças, sendo que a possibilidade de a prever com rigor pode antecipar estratégias de rentabilidade e minimizar o risco de perdas financeiras. Existem investigadores de diversas áreas que procuram resolver esse problema, apresentando uma variedade de abordagens para encontrar uma solução.

Bressan (2004) refere que “diversas técnicas de previsão auxiliam a tomada de decisões por parte dos agentes envolvidos em atividades que necessitam de planeamento, avaliação de políticas e redução de incerteza”. Nesse contexto, o mesmo autor Bressan (2004) argumenta que uma das possíveis alternativas para diminuir a incerteza no processo de tomada de decisões económicas é a utilização de modelos de previsão de séries temporais.

Amparados na análise unicamente da variável em si, tais modelos são elaborados por meio de processos estocásticos especiais que têm o objetivo de estimar o valor futuro da variável em questão com base apenas em valores passados.

Tsay (2005) destaca a relevância da análise de séries temporais financeiras na compreensão e previsão de movimentos financeiros. Esta análise é crucial para uma variedade de aplicações, incluído a previsão de preços de ativos, gestão de riscos, tomada de decisões de investimento e modelação de volatilidade.

Ao examinar padrões passados de dados financeiros, os analistas podem identificar tendências, ciclos e sazonalidades que podem influenciar os mercados. Para além disso a análise de séries temporais permite a deteção de padrões complexos e não lineares nos dados financeiros, fornecendo *insights* valiosos para *traders* e investidores.

Nos dias de hoje, as séries temporais possuem diversas aplicações em problemas do mundo real como projeção de dados financeiros, corporativos e macroeconómicos, na previsão meteorológica, na estimativa de produção para a indústria, entre outros.

De forma que os investidores e analistas financeiros tomem decisões de investimento, tendo por base a avaliação e respetiva mitigação de risco, é fundamental a identificação dos fatores exógenos que tomam preponderância na volatilidade. Walther *et al.* (2019) verificaram que os fatores económicos e financeiros globais têm maior influência na volatilidade do que os fatores individuais de cada um dos países.

Dentro das abordagens possíveis para a resolução destas questões temos a análise preditiva que possibilita a antecipação de resultados por intermedio da tecnologia, incluindo técnicas como a mineração de dados, modelação, inteligência artificial, algoritmos de cálculos estatísticos e *machine learning*.

Por outro lado, existem igualmente técnicas e modelos com algumas décadas de existência que também permitem a modelação e análise de dados, nomeadamente de series temporais.

Como exemplo, é possível referir os modelos *Autoregressive Conditional Heteroscedasticity* (ARCH) de Engle (1982) e *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH) de Bollerslev (1986), onde os retornos são calculados por um processo estocástico, com uma volatilidade variável no tempo.

É no seguimento destas diferentes técnicas e modelos que surge a temática abordada na presente dissertação, na qual serão analisadas e previstas as volatilidades de algumas ações de mercados emergentes e em desenvolvimento, num contexto temporal/histórico bastante impactante como é o período da invasão russa que se iniciou a 24 de fevereiro de 2022.

Esta dissertação será dividida em 6 capítulos principais e respetivos subcapítulos (quando existentes): a presente introdução onde serão descritos todos os objetivos que se pretendem atingir com o projeto, a identificação da questão de investigação, respetivas hipóteses e justificação e relevância do estudo. Seguir-se-á a revisão da literatura dividida em 5 subcapítulos. De seguida será exposto o método e processo de investigação, posteriormente constará o índice provisório do projeto de dissertação, o cronograma de trabalhos a respetiva conclusão e por último encontrar-se-ão as referências bibliográficas.

1.1 Questão de Investigação

Será possível prever a volatilidade dos retornos de *Exchange Traded Funds* (ETF's) disponíveis em mercado, para o período que marcou o início da invasão russa (27 de janeiro de 2022 a 24 de abril de 2022), utilizando modelos autorregressivos de heterocedasticidade condicional bem como modelos de inteligência artificial?

1.2 Objetivo Geral

O presente projeto terá como objetivo geral analisar e prever a volatilidade dos retornos de ETF's em mercado, através da aplicação do modelo GARCH bem como o algoritmo *Extreme Gradient Boosting* (XGBOOST), com recurso a aprendizagem de máquina, vulgo *Machine Learning*.

1.2.1 Objetivos Específicos

Neste projeto serão apresentadas algumas abordagens para previsão da volatilidade de series temporais.

Deste modo, em matéria de objetivos específicos, pretende-se avaliar o comportamento das previsões de acordo com diferentes parâmetros e modelos aplicados ao problema, comparando os resultados previstos com os valores reais, tendo em consideração bases de dados de teste.

1.2.2 Objetivos Teóricos

Como objetivos teóricos, pretende-se explorar a implementação de modelos de autorregressivos e de inteligência artificial, a series temporais, efetuando a comparação entre eles, uma vez que pese embora existam algumas abordagens à temática definida, a literatura que aborde este tipo de comparação é reduzida.

1.2.3 Objetivos Metodológicos

Metodologicamente pretende-se aplicar os modelos de regressão tendo como base a linguagem de *Python*, vocacionada para toda a tipologia de programação, mas com recursos bastante importantes no tratamento de dados estatísticos e geração de gráficos.

1.2.4 Objetivos Práticos

Como guia ou recomendação para os possíveis utilizadores da informação gerada, desde investidores diretos a sociedades gestoras de patrimónios, brokers ou até instituições financeiras (bancos), pretende-se aconselhar a utilização prudente da informação que será gerada pela presente investigação dado que poderá influenciar decisões de investimento, o que no futuro poderá levar a perdas financeiras.

1.3 Hipóteses

Hipótese 1: Pretende-se confirmar com o presente projeto que períodos de instabilidade como é o caso do Covid-19 e a invasão da Rússia têm um impacto significativo na volatilidade dos ETF's, corroborando a influência dos fatores externos às transações de mercado com impacto positivo ou negativo face aos resultados esperados.

Hipótese 2: Pretende-se confirmar que os modelos e algoritmos de previsão são menos eficazes em períodos de maior volatilidade.

Hipótese 3: Pretende-se igualmente comparar os resultados obtidos com os diferentes modelos de regressão, confirmando o recurso à inteligência artificial como método mais sofisticado e célere para obter resultados aproximados.

1.4 Justificação e Relevância do Estudo

Para evidenciar a importância desta temática, é possível constatar o crescente número de citações sobre o recurso a inteligência artificial para previsões no âmbito do mercado de capitais a partir da Figura 1.1

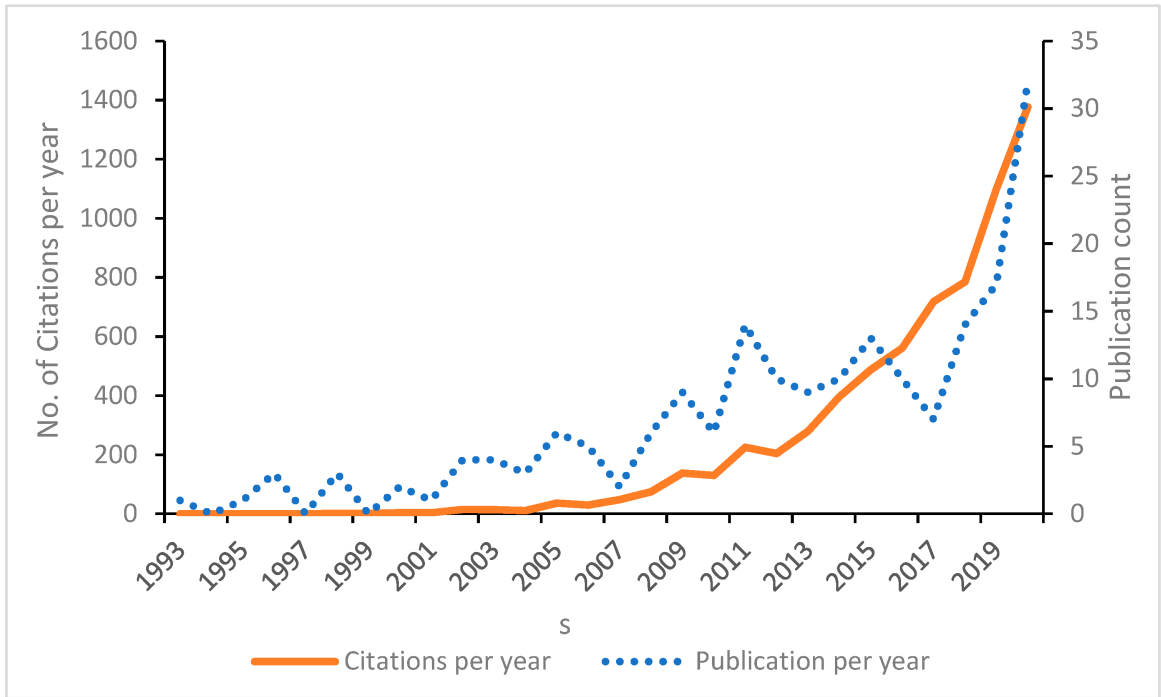


FIGURA 1.1 - N° DE CITAÇÕES E PUBLICAÇÕES POR ANO REFERENTES INTELIGÊNCIA ARTIFICIAL APLICADA AOS MERCADOS DE CAPITAIS

Fonte: <https://www.mdpi.com/1911-8074/14/11/526>

2. Revisão da Literatura

A revisão da literatura estará dividida em nove subcapítulos: conceitos aplicados aos mercados financeiros, investidores, crises financeiras, volatilidade, heterocedasticidade condicional, modelos de regressão aplicados no presente projeto de investigação, *machine learning*, critérios de seleção de variáveis e por fim, métricas de avaliação da *performance* dos modelos.

2.1 Mercados Financeiros

Os mercados financeiros oferecem oportunidades de investimento em países com uma economia desenvolvida, para investidores ou entidades que pretendam investir em substituição dos seus clientes. Os mercados proporcionam a oportunidade de obter um retorno elevado, no entanto, acompanhado de riscos também eles elevados.

É possível afirmar que os mercados financeiros representam o meio onde são realizadas as transferências ou fluxos de recursos monetários, entre os agentes financeiros/económicos de forma direta ou indireta.

Neste estudo e de forma mais abrangente os mercados financeiros são designados como o local, onde compradores e vendedores se encontram com o objetivo de transacionar ativos financeiros.

Os mercados financeiros representam a estrutura de uma economia capitalista, proporcionando a empresários e empresas um método de se financiarem e obterem liquidez. No que respeita ao financiamento, podem recorrer à emissão de dívida ou até à cessão de participações do negócio a eventuais compradores, com o intuito de sustentar os seus projetos e iniciativas, reduzindo o risco associado a estas transações (Ferreira, 2010).

2.1.1 Mercados Desenvolvidos

Os mercados desenvolvidos, como descritos por autores como *ver* Stiglitz (2003) são caracterizados por economias altamente industrializadas, instituições financeiras estáveis e padrões de vida elevados. Essas economias tendem a ser maduras e consolidadas com acesso, relativamente fácil, ao capital e uma base de consumidores estável. Krugman (2003) aborda algumas perspetivas sobre os desafios enfrentados por estes mercados, incluindo desigualdade

económica e políticas fiscais, destacando a importância de políticas económicas eficazes para a estabilidade.

2.1.2 Mercados Emergentes

A temática dos mercados emergentes é igualmente explorada na literatura, autores como Raghuram (2010) e Soto (2000) que exploram temáticas como a inclusão financeira e a reforma institucional, descrevem os desafios que estes mercados enfrentam no seu caminho para um desenvolvimento económico sustentável. Estas economias em transição, são caracterizadas por um rápido crescimento industrial e de urbanização, muitas vezes lidam com instabilidade política, infraestruturas subdesenvolvidas e volatilidade económica.

2.2 Sentimentos do Investidor

Nos mercados financeiros contemporâneos, os sentimentos dos investidores são reconhecidos como influências significativas na tomada de decisões e na dinâmica dos preços dos ativos. Shiller (2015), destaca a importância dos sentimentos, como o excesso de confiança e o medo, na formação de bolhas e crises financeiras. Este autor argumenta que essas emoções coletivas muitas vezes desviam os preços dos ativos dos seus fundamentos intrínsecos, criando oportunidades e riscos para os próprios investidores.

A Neuroeconomia, um campo interdisciplinar que combina a neurociência, economia e psicologia, oferece *insights* sobre os processos neurobiológicos subjacentes ao comportamento financeiro. Coates (2012) explora o fato das reações fisiológicas do corpo, como a libertação de hormonas do stress, influenciarem as decisões de investimento, podendo levar a comportamentos irracionais e a tomada de riscos excessivos em momentos de alta volatilidade do mercado.

Thaler e Sunstein (2008), propõem intervenções comportamentais conhecidas como “empurrões”, para auxiliar os investidores a ultrapassar estas interferências cognitivas e emocionais, argumentando que pequenas alterações no ambiente de escolha podem orientar os investidores para decisões mais informadas e alinhadas com os seus objetivos financeiros a longo prazo.

2.3 Crises Financeiras

As crises financeiras são eventos recorrentes na história econômica global, e compreender as suas causas, impactos e formas de mitigação é crucial para a estabilidade e o desenvolvimento dos mercados financeiros. Alguns autores financeiros têm contribuído significativamente para o entendimento destes fenômenos como é o caso de Roubini e Mihm (2008) que destacam a interconexão entre o colapso do mercado imobiliário dos Estados Unidos e a crise bancária internacional, explorando os fatos que levaram à bolha econômica imobiliária, como as políticas de crédito e os derivados financeiros complexos, com consequências devastadoras em termos de desemprego, falências de empresas e instabilidade econômica global.

Alguns economistas, como é o caso de Krugman (2020) propõem políticas econômicas para lidar com as consequências das crises financeiras, incluindo estímulos fiscais, regulação financeira mais estrita e medidas para combater a desigualdade econômica. *Krugman* argumenta que abordagens coordenadas e pragmáticas são essenciais para restaurar a confiança dos investidores e promover a recuperação econômica sustentável após uma crise.

➤ Impacto da Invasão da Rússia na Ucrânia nos mercados financeiros

A Invasão da Rússia na Ucrânia teve um impacto significativo nos mercados financeiros, gerando volatilidade e incerteza global. Autoridades e especialistas econômicos têm analisado os efeitos dessa crise geopolítica nos mercados financeiros mundiais.

Vlassopoulos e Tsileponis (2023) destacam que a incerteza geopolítica gerada pela invasão afetou os mercados financeiros de várias formas, incluindo volatilidade nos preços dos ativos, mudanças nos fluxos de capital e ajustes nas estratégias de investimento.

A persistência da incerteza geopolítica, combinada com a aplicação contínua de sanções econômicas e as respostas políticas em evolução, mantém a volatilidade nos mercados financeiros. Investidores e empresas enfrentam desafios contínuos para antecipar e mitigar os riscos associados a essa crise. As respostas políticas e econômicas dos países envolvidos e da comunidade internacional têm sido cruciais para moldar o cenário financeiro global.

Portanto, mesmo após a invasão inicial em 2022, o impacto da crise continua a ser uma área de interesse e pesquisa.

2.4 Volatilidade dos Retornos Financeiros

A análise da volatilidade representa uma ferramenta de gestão essencial para a avaliação da frequência e intensidade das oscilações nos preços dos ativos e a incerteza associada ao investimento financeiro num período específico do tempo. Quando a volatilidade de um ativo financeiro aumenta, também aumenta a expectativa de variação no valor desse ativo tal como o seu nível de risco. Por outro lado, uma volatilidade mais reduzida indica que as variações no valor do ativo são menores.

“Resultados empíricos mostram que a volatilidade de uma ação é muito maior quando a bolsa está aberta do que quando está fechada. Isso sugere que, até certo ponto, a própria negociação causa a volatilidade do preço das ações” Hull (2013, p.257).

Medir a volatilidade não é o mesmo que medir o preço de um ativo, enquanto o preço é algo que pode ser observado num determinado instante (observação direta), a volatilidade instantânea não é observável e requiere a variação de preços ao longo do tempo.

No que respeita aos tipos de volatilidade, usualmente são três:

- (i) **Volatilidade Histórica** – também designada de volatilidade estatística, medida de dispersão dos retornos de um ativo ou índice de mercado durante um período predefinido. Normalmente é calculada como o desvio padrão dos retornos passados do instrumento financeiro (quanto maior for o valor do desvio padrão, maior será a volatilidade)

A utilização do desvio padrão será a forma mais comum de efetuar o cálculo, no entanto, não é a única. Em particular para a volatilidade histórica, também ela conhecida como volatilidade real ou realizada do ativo, pode ser modelada utilizando modelos ARCH/GARCH

É importante realçar que, pese embora a volatilidade e o desvio padrão sejam estatísticas que quantificam as variações (neste caso, dos retornos), estes conceitos não são sinónimos. Enquanto o segundo pode ser aplicado de forma mais abrangente (peso, altura, quantidade, volume, etc.), a volatilidade poderá ser compreendida como o desvio padrão ao longo do tempo.

A fórmula para o cálculo da volatilidade histórica é:

$$\sigma = \sqrt{\frac{\sum_{t=1}^n (x_t - \bar{x})^2}{n}} \quad (2.1)$$

Onde,

x_t é o logaritmo natural do quociente entre duas cotações sucessivas, $\ln(P_t|P_{t-1})$

\bar{x} é a média aritmética de x_t

n é o número de observações

No cálculo da volatilidade, caso sejam consideradas cotações diárias (ano civil), semanais ou até mensais, poderá ser efetuada a respetiva anualização, multiplicando o resultado por $\sqrt{365}$, $\sqrt{52}$ ou $\sqrt{12}$.

Em determinadas análises deve-se dar maior a importância ou peso aos dados históricos mais recentes comparativamente aos dados mais antigos. Essa abordagem pode ser implementada com recurso a ponderadores. Desta forma, a volatilidade histórica ponderada σ_S pode se calculada pela seguinte expressão:

$$\sigma_S = \sqrt{\sum_{t=1}^n (nw_t X_t - \bar{x})^2} \quad (2.2)$$

Onde

$w_1 w_2, \dots, w_n$ são as ponderações atribuídas aos acontecimentos e $\sum w_t = 1$

Compreende-se, no entanto, que ao atribuir maior peso às variações mais recentes na volatilidade ponderada, esta tende a ser maior do que aquela obtida quando todos os eventos recebem a mesma ponderação. Além disso, observa-se que existe um valor médio ao longo do tempo para o qual os eventos atuais têm uma tendência a convergir, chamada de reversão à média. Se os níveis atuais estiverem abaixo do nível médio de longo prazo, tenderão a aumentar, enquanto, se estiverem acima, tenderão a diminuir, embora a velocidade dessa reversão possa variar, sendo mais rápida em períodos de maior volatilidade. A fórmula que descreve este fenómeno é:

$$\sigma^* = \sigma_{recente} + \alpha^*(\sigma_{recente} - \sigma_{longo}) \quad (2.3)$$

com,

α^* a velocidade de reversão.

- (ii) **Volatilidade Implícita** – também conhecida como volatilidade projetada, refere-se à percepção do mercado sobre quão volátil uma ação poderá ser no futuro. Esta é expressa como uma percentagem anualizada. No fundo a volatilidade implícita é aquela que equilibra o preço de mercado da opção, com o preço teórico derivado de um modelo de avaliação de opções. Neste processo, o preço da opção é considerado como input, com a volatilidade a ser determinada como resultado. Pese embora seja muito comum utilizar o preço das opções no fecho do mercado, uma abordagem mais precisa envolve o uso de um valor médio entre os preços de compra e de venda (*bid e ask*). Em termos de cálculo é usual a utilização do algoritmo de *Newton-Raphson* (Rouah & Vainberg, 2007).

Encontrar a volatilidade implícita poderá ser bastante útil quando se deseja compreender a percepção do mercado sobre um ativo subjacente num dado omento, ou quando se pretende comparar os preços das opções para diferentes preços de exercício e prazos de vencimento. É frequentemente utilizada para desenvolver estratégias de compra ou venda. Por exemplo, se as opções de compra estão subvalorizadas e as opções de venda estão sobrevalorizadas devido a diferenças na volatilidade, pode ser vantajoso para o investidor comprar opções de compra e vender opções de venda.

- (iii) **Volatilidade Futura** – a volatilidade futura, ou prevista, refere-se à expectativa de mudanças nas cotações no futuro e é considerada a mais desafiadora de se estimar. Uma abordagem comum para estimar essa volatilidade é através de modelos como os do tipo ARCH, os quais reconhecem que as volatilidades não permanecem constantes ao longo do tempo, assumindo que a volatilidade é influenciada pelo seu próprio histórico.

2.5 Heterocedasticidade Condicional

Em estatística, caso a variância de um conjunto de dados permaneça constante ao longo do tempo, a série é considerada homocedástica (Bollerslev, 1986), caso contrário será considerada heterocedástica.

Assim sendo, quando uma série temporal apresenta uma variância que não é constante, i.e., flutua ao longo do tempo relativamente a períodos anteriores, essa mesma série é considerada heterocedástica condicional.

O comportamento específico da variância ao longo do tempo pode ser observado, por exemplo, nas séries de retornos financeiros, que se manifesta pela existência de aglomeração de volatilidade ao longo do tempo, ou seja, “períodos em que exibem oscilações significativas por um período prolongado seguido por um período de relativa tranquilidade” (Gujarati & Porter, 2008, p. 773).

Uma vez que os retornos próximos no tempo tendem a mostrar níveis semelhantes de variabilidade, os choques que ocorrem no tempo t vão influenciar a volatilidade nos n períodos seguintes, explicando a persistência da volatilidade. Por outro lado, a longo prazo, assumindo que a variância incondicional seja finita, os períodos de alta volatilidade tendem a ser seguidos por períodos de menor volatilidade e vice-versa. Essa característica da volatilidade é conhecida como reversão à média (Mandelbrot, 1963)

Por exemplo, as eleições tendem a gerar instabilidades momentâneas, que se refletem em indicadores mais instáveis/voláteis em momentos específicos da série temporal analisada. Esses factos remetem para modelos cuja variância depende fortemente dos seus valores anteriores e dos valores de retornos recentes (Leblang & Mukherjee, 2004). Estes modelos são referidos na literatura como modelos de heterocedasticidade condicional onde a variância incondicional permanece constante, alterando apenas a variância condicional da série.

2.6 Modelos de Regressão

A grande maioria dos casos práticos relacionados com séries temporais de cariz financeiro (preços de ações, taxas de inflação, taxas de câmbio, etc.) contradizem a suposição de variância constante devido à presença de volatilidade.

Modelos de tendência linear ou até modelos *Autoregressive integrated moving average* (ARIMA) falhariam em ajustar-se corretamente, pois todos assumem homocedasticidade do erro (Verbeek, 2017).

A possível existência de autocorrelação da variância no tempo, com os seus valores desfasados em um ou mais períodos, foi a motivação por detrás do modelo de ARCH, no qual a variância do erro está relacionada ao erro ao quadrado no tempo anterior, bem como o modelo GARCH, no qual a variância do erro está relacionada aos erros ao quadrado de vários períodos no passado (Gujarati & Porter, 2008).

É importante referir que o autorregressivo descreve o mecanismo de feedback que permite a incorporação de observações passadas no modelo. Condicional implica que as observações atuais dependem das observações imediatamente anteriores e heterocedasticidade significa a flutuação da variância ao longo do tempo.

2.6.1 Modelo ARCH

Em 1982, Robert Engle, vencedor do Prémio Nobel de Ciências Económicas em 2003, considerou ser possível construir um modelo paramétrico no qual a variância seria condicionada por uma equação algébrica, modelando não apenas a média, mas também a variância condicionada (Engle 1982). Em resumo, “enquanto os modelos convencionais de séries temporais e econometria operam sob a suposição de variância constante, o processo ARCH permite que a variância condicional mude ao longo do tempo como uma função de erros passados, mantendo a variância incondicional constante” (Bollerslev, 1986, p. 1).

A estimação de modelos ARCH não pode ser feita através do método dos Mínimos Quadrados Ordinários, uma vez que este método minimiza a soma dos quadrados dos resíduos, que depende apenas dos parâmetros da equação da média e não da variância condicional.

Deste modo, a estimação de modelos ARCH é efetuada através da máxima verosimilhança, sob a hipótese de que os erros são distribuídos condicionalmente.

Tal como referido anteriormente o modelo ARCH considera que a variância condicional t (σ_t^2) é função dos erros anteriores (passados) ao quadrado (ε_{t-1}^2) podendo ser representado pela fórmula base (2).

Um processo X_t , diz-se um processo ARCH de ordem q se,

$$X_t = \sigma_t \varepsilon_t, \text{ com } \sigma_t^2 = \alpha_0 + \sum_{i=1, q} \alpha_i \varepsilon_{t-i}^2 \quad (2.4)$$

Sendo α_0 e α_i parâmetros do modelo que satisfazem um conjunto de condições ($\alpha_0 > 0$ e $\alpha_i \geq 0, i = 1, 2, \dots, q$) para garantir que a variância não condicionada é finita.

Habitualmente considera-se que a variável aleatória ε_t segue uma distribuição normal. Quando temos σ_t constante ao longo do tempo, afirma-se que X_t é um ruído branco.

Como ideia base é referido que para este modelo, X_t é não correlacionado, mas dependente e essa mesma dependência pode ser descrita através de uma função quadrática simples. Através da estrutura deste modelo é verificado considerando valores elevados de X_t irá implicar uma variância condicional σ_t^2 de X_t também elevada.

Este modelo apresenta algumas limitações descritas na literatura sendo que de acordo com Tsay (2005), pese embora o modelo ARCH seja de fácil implementação, apresenta algumas desvantagens dado que requer diversos parâmetros para descrever a evolução da volatilidade. Existem, contudo, outras limitações inicialmente referidas por Brooks (2002), nomeadamente:

- Modelo pressupõe que os choques positivos ou negativos produzem o mesmo efeito na volatilidade, dado que depende do quadrado dos choques anteriores;
- Não fornece formas de compreensão ou identificação da fonte das variações de uma série temporal financeira.
- Tende a devolver previsões exageradas da volatilidade porque responde lentamente a grandes alterações isoladas

Para que fossem ultrapassadas algumas das desvantagens do modelo, foi desenvolvido o modelo GARCH a abordar no próximo tópico.

2.6.2 Modelo GARCH

Para superar algumas das limitações do modelo anterior, Bollerslev (1986) propôs uma generalização do modelo ARCH que se assemelha à abordagem Auto Regressive Moving Average (ARMA), denominando-o de Modelo de Heterocedasticidade Condicional Autorregressiva Generalizada (GARCH).

Aqui, a variância condicional depende não apenas de choques aleatórios ao quadrado que ocorrem nos momentos imediatamente anteriores, mas também das variâncias condicionais dos modelos imediatamente anteriores.

Embora a estrutura dos modelos GARCH e dos modelos ARMA seja muito semelhante, existe uma diferença bastante relevante entre ambos: enquanto os retornos dependem apenas dos retornos de períodos anteriores no modelo ARMA, os modelos ARCH/GARCH consideram que os retornos também dependem da variância (volatilidade) observada no passado, bem como dos erros associados ao processo anterior (Bollerslev, 1986).

Em comum, os modelos ARMA e GARCH assumem que os mercados são eficientes e que os retornos das ações seguem processos estocásticos, sendo, portanto, imprevisíveis a médio e longo prazo. Além disso, os modelos estocásticos simples pressupõem que os retornos não estão correlacionados e que a sua média é zero (Wooldridge, 2015). O modelo GARCH para variância condicional pode ser considerado como um processo ARMA nas inovações ao quadrado, embora não nas variações, como as equações podem parecer sugerir.

O efeito de feedback proporcionado pela introdução de variâncias condicionais desfasadas torna o modelo GARCH mais simples (com menos parâmetros) em comparação com o modelo ARMA. A grande vantagem deste modelo é que, para estimar um determinado parâmetro, considera diferentes pesos para cada observação na série temporal, dando maior peso às mais recentes.

O modelo GARCH permite a presença de componentes autorregressivas e médias móveis na variância heterocedástica de ativos financeiros. Embora os modelos GARCH tenham uma

ampla gama de aplicações, combinando várias características observadas em séries temporais e produzindo resultados satisfatórios, existem limitações (Brooks & Lee, 1997 *cit in* Marques, 2017):

- É necessário que a série apresente uma variância condicional que flutue ao longo do tempo.
- Esses modelos muitas vezes falham em capturar eventos altamente inesperados que podem levar a mudanças estruturais significativas, bem como fenômenos irregulares, como quedas sucessivas e suas implicações.
- Efeitos assimétricos na volatilidade, por exemplo, 'boas notícias' (tranquilidade, com diminuição da volatilidade) ou 'más notícias' (turbulência, com aumento da volatilidade), não podem ser modelados pela abordagem GARCH.

No que concerne à expressão que sumariza o modelo GARCH, é possível constatar que a variância condicional t (σ_t^2) é função dos erros anteriores (passados) ao quadrado (ε_{t-1}^2), mas também dos seus próprios valores passados (σ_{t-j}^2), conforme representado na equação (3)

$$\sigma_t^2 = \alpha_0 + \sum_{i=1,q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1,p} \beta_j \sigma_{t-j}^2 \quad (2.5)$$

Sendo α_0 e α_i e β_j parâmetros do modelo.

O modelo representado acima, tal como referido anteriormente é um modelo autorregressivo de Heterocedasticidade condicional generalizada de ordem (p,q) e representa-se como GARCH (p,q), onde q é o grau do processo ARCH e o p é definido como o grau do processo GARCH. Deste modo é possível concluir que para p=0, o modelo GARCH (0, q) equivale ao modelo ARCH (q). Assim, para que o modelo apresente covariância estacionária e, nesse sentido, variância incondicional ou pelo menos uma tendência de convergência, é fundamental que sejam verificadas as seguintes condições:

$$\alpha_0 > 0$$

$$\alpha_i \geq 0 \quad (i = 1, 2, \dots, q)$$

$$\beta_j \geq 0 \quad (j = 1, 2, \dots, q)$$

2.6.3 Algoritmo XGBOOST

O XGBOOST é uma técnica de aprendizagem de máquina em código aberto (open source) proposta por Chen e Guestrin (2016) que permite ajustar relações não lineares complexas nos dados, implementar eficientemente algoritmos *Gradient Boosted Decision Trees* (GBDT) e efetuar diversas melhorias nos algoritmos e na engenharia. O XGBoost é um dos algoritmos de impulso, que utilizam dados brutos como base, efetuam o respetivo treino para obter uma função objetivo minimizada e utilizam isso para encontrar o melhor conjunto de parâmetros. Este, utiliza várias árvores de decisão, acrescentando termos regulares à função objetivo do GBDT para ajudar a reduzir o risco de sobre ajuste.

Este modelo ou técnica tornou-se muito popular devido à sua eficácia em várias tarefas, nomeadamente a previsão em séries temporais financeiras. No entanto, é importante observar que não se trata de uma abordagem direta para lidar com séries temporais, como os modelos autorregressivos tradicionais (por exemplo, ARMA), uma vez que utiliza aprendizagem de máquina para poder ser aplicado à resolução de problemas.

De modo a serem obtidos mais detalhes sobre os modelos e algoritmo, referidos nos pontos anteriores, deverá ser consultado o capítulo 4 – Modelo de Análise e Operacionalização de Variáveis no documento final de Dissertação.

2.7 Machine Learning

Machine Learning ou Aprendizagem de Máquina, tem vindo a ganhar cada vez mais relevo no mundo incluindo em Portugal, nos últimos anos. Esta área da inteligência artificial concentra-se no desenvolvimento de algoritmos e técnicas que permitem aos computadores aprender com os dados, sem necessidade de serem programados especificamente para tal.

2.8 Critérios de Seleção de Variáveis

Nos modelos de regressão são vários os critérios de seleção das variáveis, tais como o R-Quadrado, o *Akaike Information Criterion* (AIC), o *Bayesian Information Criterion* (BIC), C_p de *Mallows*, entre outros (Zhang, 2016)

2.8.1 AIC

O AIC é uma medida estatística utilizada para avaliar a qualidade relativa de um modelo estatístico. Foi proposto pelo estatístico japonês *Hirotsugu Akaike*. O AIC oferece uma forma de comparar diferentes modelos, considerando a sua capacidade de ajuste aos dados e a sua complexidade.

De forma simples, este critério penaliza modelos mais complexos, ou seja, aqueles que têm um elevado número de parâmetros, para evitar o sobre ajuste ou *overfitting* aos dados. É calculado utilizando a seguinte fórmula:

$$AIC = 2k - 2 \ln(L) \quad (2.6)$$

Onde:

k é o número de parâmetros no modelo

L é a função de verosimilhança do modelo que mede a probabilidade dos dados observados sob o modelo

Quanto menor o valor do AIC, melhor o modelo é considerado. Assim, o AIC pode ser utilizado para selecionar o modelo mais apropriado entre um conjunto de modelos candidatos (Yanagihara *et al.*, 2012).

2.8.2 BIC

O BIC, também conhecido como o Critério de Schwartz, surgiu em 1978 e representa outra das medidas estatísticas utilizada para avaliar a qualidade relativa dos modelos.

Esta, foi proposta pelos estatísticos Gideon Schwarz e Hirotsugu Akaike.

O BIC é semelhante ao AIC, mas aplica uma penalização ainda mais forte aos modelos complexos. Assim, o BIC tende a selecionar modelos mais simples, sendo calculado usando a seguinte fórmula:

$$BIC = 2 \ln(L) + k \ln(n) \quad (2.7)$$

Onde:

k é o número de parâmetros no modelo

L é a função de verossimilhança do modelo que mede a probabilidade dos dados observados sob o modelo

n é o tamanho da amostra.

Tal como acontece com o AIC, quanto menor o valor de BIC, melhor será o modelo (Konishi & Kitagawa, 2008).

2.9 Métricas de Avaliação da Performance dos Modelos

As Métricas de Avaliação da Performance dos modelos, também designadas na literatura por Métricas de Erro, são amplamente utilizadas na avaliação da eficácia e precisão dos modelos de regressão, nomeadamente no que concerne a modelos preditivos para séries temporais (Choudhary, Kumar, Sharma e Sharma, 2022), dado que permitem calcular o desvio do valor previsto em contraste com o valor real. Existem várias métricas, contudo, abordamos nos próximos tópicos aquelas que foram utilizadas na avaliação dos modelos aplicados nesta investigação.

2.9.1 MAPE

O *Mean Absolute Percentage Error* (MAPE) é uma métrica que dimensiona o desvio absoluto médio de modo a obter unidades percentuais em vez de unidades da variável, pese embora sejam utilizados valores absolutos para evitar os erros se anularem e erros relativos para que seja possível a comparação da previsão entre modelos de séries temporais. Em suma mede a precisão de um

modelo de previsão ao calcular a média dos erros percentuais absolutos entre as previsões e os valores reais.

O cálculo do MAPE é efetuado através da seguinte fórmula:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \hat{y}_i}{y_i} \times 100 \quad (2.8)$$

Onde:

y_j representa o valor real

\hat{y}_i representa o valor previsto

n é o número total de observações

Quanto menor o valor do MAPE, melhor será a precisão do modelo de previsão, i.e., para um MAPE de 5%, teremos em média um modelo cujas previsões estão a 5% de distância dos valores reais.

2.9.2 RMSE

O *Root Mean Squared Error* (RMSE) ou erro Quadrático Médio da Raiz, é outra métrica amplamente utilizada na avaliação de modelos de previsão e análise de dados. Esta, mede a média dos erros quadráticos entre as previsões e os valores reais (Jilmin, 2016), sendo calculada pela fórmula abaixo:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.9)$$

Onde:

y_i representa o valor real

\hat{y}_i representa o valor previstos

n é o número total de observações

O RMSE é bastante útil pois penaliza fortemente grandes erros de previsão, já que os erros são elevados ao quadrado antes de serem somados e a raiz quadrada é extraída.

Isso significa que os erros maiores têm um impacto proporcionalmente maior na métrica.

Assim como o MAPE, quanto menor for o valor do RMSE, melhor será a precisão do modelo.

3. Método e Processo de Investigação

O processo de investigação será conduzido em função das 3 fases que a literatura da metodologia recomenda (e.g. Quivy & Campenhoudt, 1998), em particular (i) romper com as ideias preconcebidas, (ii) construir a investigação e (iii) experimentar.

A primeira fase iniciou-se com a seleção da questão de investigação e continua neste momento na etapa de exploração, i.e., na recolha da literatura a estudar e sua tipologia.

De seguida será abordada a problemática e dar-se-á início à construção do modelo de análise de modo a definir propostas que expliquem o fenómeno em estudo, efetuar a previsão e preparação do plano de investigação, as operações a realizar e as respetivas consequências ou resultados esperados.

Por último será efetuada a observação, ou seja, a experimentação e testagem dos dados, que posteriormente serão analisados e tendo em consideração os resultados obtidos serão retirados as conclusões que se pretendem com o projeto.

4. Descrição dos dados

No que concerne aos dados, e uma vez que a presente dissertação tem como finalidade o tratamento e análise de series temporais, mais concretamente series relacionadas com o histórico de preços de títulos cotados, será utilizada como fonte de dados o *Yahoo Finance*, que permite a qualquer pesquisador, efetuar o *download* da informação para um ficheiro com extensão .csv, ou através de distintas linguagens de programação, cujo tratamento da informação se torna bastante mais simplificado.

4.1 Produtos selecionados

Foram selecionados seis ETF's, três de Mercados Emergentes e três de Mercados Desenvolvidos, sendo que o critério de seleção foi o volume de *Assets Under Management* (AUM's). Para os ETF's de Mercados Desenvolvidos a escolha foi realizada através do website <https://www.etf.com/etf-education-center/etf-basics/most-popular-etfs> e para os ETF's de Mercados Emergentes a fonte foi o website <https://www.investopedia.com/articles/etfs-mutual-funds/052116/5-largest-developed-market-etfs-efa-vea.asp>. Os valores mobiliários selecionados encontram-se referidos na Tabela 4.1 - SELEÇÃO DE ETF'S infra.

TABELA 4.1 - SELEÇÃO DE ETF'S

Tipo de Mercado	Mercado/Praça	ISIN	Designação	Ticker	Moeda	Composição do Portfolio
Desenvolvido	NYSEArca	US78462F1030	SPDR S&P 500 ETF Trust	SPY	USD	Ações
Desenvolvido	NYSEArca	US4642872000	iShares Core S&P 500 ETF	IVV	USD	Ações
Desenvolvido	NYSEArca	US9229087690	Vanguard Total Stock Market Index Fund ETF Shares	VTI	USD	Ações
Emergente	NYSEArca	US9220428588	Vanguard Emerging Markets Stock Index Fund	VWO	USD	Ações
Emergente	NYSEArca	US46434G1031	iShares Core MSCI Emerging Markets ETF	IEMG	USD	Ações
Emergente	NasdaqGM	US92203J4076	Vanguard Total International Bond Index Fund	BNDX	USD	Obrigações

4.1.1 SPDR S&P 500 ETF Trust

ETF composto por ações de diferentes setores, com a respetiva ponderação descrita na Figura 4.1 e o Top das nove empresas com maior relevo na Figura 4.2.

Sector Weightings (%)

Sector(s)	SPY
Basic Materials	2.09%
Consumer Cyclical	10.79%
Financial Services	12.53%
Real Estate	2.31%
Consumer Defensive	5.93%
Healthcare	12.56%
Utilities	2.13%
Communication Services	8.90%
Energy	3.71%
Industrials	8.23%
Technology	30.83%

FIGURA 4.1 - SETORES - SPDR S&P 500 ETF TRUST

Fonte: Yahoo Finance

Top 9 Holdings (30.30% of Total Assets)

[Get Quotes for Top Holdings](#)

Name	Symbol	% Assets
Microsoft Corp	MSFT	7.18%
Apple Inc	AAPL	6.17%
NVIDIA Corp	NVDA	4.56%
Amazon.com Inc	AMZN	3.75%
Meta Platforms Inc Class A	META	2.54%
Berkshire Hathaway Inc Class B	BRK-B	1.74%
Alphabet Inc Class C	GOOG	1.63%
Eli Lilly and Co	LLY	1.40%
Broadcom Inc	AVGO	1.33%

FIGURA 4.2 - TOP 9 EMPRESAS - SPDR S&P 500 ETF TRUST

Fonte: Yahoo Finance

4.1.2 iShares Core S&P 500 ETF

ETF composto por ações de diferentes setores com a respetiva ponderação descrita na Figura 4.3 e o Top das nove empresas com maior relevo na Figura 4.4.

Sector Weightings (%)

Sector(s)	IVV
Basic Materials	2.09%
Consumer Cyclical	10.79%
Financial Services	12.53%
Real Estate	2.32%
Consumer Defensive	5.93%
Healthcare	12.56%
Utilities	2.13%
Communication Services	8.89%
Energy	3.71%
Industrials	8.23%
Technology	30.82%

FIGURA 4.3 - SECTORES - ISHARES CORE S&P 500 ETF

Fonte: Yahoo Finance

Top 9 Holdings (30.26% of Total Assets)

[Get Quotes for Top Holdings](#)

Name	Symbol	% Assets
Microsoft Corp	MSFT	7.17%
Apple Inc	AAPL	6.16%
NVIDIA Corp	NVDA	4.56%
Amazon.com Inc	AMZN	3.75%
Meta Platforms Inc Class A	META	2.54%
Berkshire Hathaway Inc Class B	BRK-B	1.74%
Alphabet Inc Class C	GOOG	1.62%
Eli Lilly and Co	LLY	1.40%
Broadcom Inc	AVGO	1.33%

FIGURA 4.4 - TOP 9 EMPRESAS - ISHARES CORE S&P 500 ETF

Fonte: Yahoo Finance

4.1.3 Vanguard Total Stock Market Index Fund ETF Shares

ETF composto por ações de diferentes setores com a respectiva ponderação descrita na Figura 4.5 e o Top das nove empresas com maior relevo na Figura 4.6.

Sector(s)	VTI
Basic Materials	2.30%
Consumer Cyclical	10.49%
Financial Services	12.93%
Real Estate	2.93%
Consumer Defensive	5.75%
Healthcare	12.76%
Utilities	2.18%
Communication Services	8.20%
Energy	3.91%
Industrials	9.32%
Technology	29.22%

FIGURA 4.5 - SETORES - VANGUARD TOTAL STOCK MARKET INDEX FUND ETF SHARES

Fonte: Yahoo Finance

Top 9 Holdings (25.30% of Total Assets)

[Get Quotes for Top Holdings](#)

Name	Symbol	% Assets
Microsoft Corp	MSFT	6.29%
Apple Inc	AAPL	5.80%
NVIDIA Corp	NVDA	3.07%
Amazon.com Inc	AMZN	3.07%
Meta Platforms Inc Class A	META	1.84%
Alphabet Inc Class C	GOOG	1.47%
Berkshire Hathaway Inc Class B	BRK-B	1.45%
Eli Lilly and Co	LLY	1.17%
Broadcom Inc	AVGO	1.12%

FIGURA 4.6 - TOP 9 EMPRESAS - VANGUARD TOTAL STOCK MARKET INDEX FUND ETF SHARES

Fonte: Yahoo Finance

4.1.4 Vanguard Emerging Markets Stock Index Fund

ETF composto por ações de diferentes setores com a respectiva ponderação descrita na Figura 4.7 e o Top das sete empresas com maior relevo na Figura 4.8.

Sector Weightings (%)

Sector(s)	VWO
Basic Materials	8.43%
Consumer Cyclical	11.83%
Financial Services	21.90%
Real Estate	2.72%
Consumer Defensive	6.19%
Healthcare	4.33%
Utilities	3.69%
Communication Services	8.27%
Energy	6.14%
Industrials	8.32%
Technology	18.18%

FIGURA 4.7 - SETORES - VANGUARD EMERGING MARKETS STOCK INDEX FUND

Fonte: Yahoo Finance

Top 7 Holdings (15.11% of Total Assets)

[Get Quotes for Top Holdings](#)

Name	Symbol	% Assets
Taiwan Semiconductor Manufacturing Co Ltd	2330.TW	5.85%
Tencent Holdings Ltd	0700.HK	2.91%
Alibaba Group Holding Ltd Ordinary Shares	9988.HK	2.05%
Reliance Industries Ltd	RELIANCE.NS	1.55%
HDFC Bank Ltd	HDFCBANK.NS	1.11%
Infosys Ltd	INFY.NS	0.94%
Vale SA	VALE3.SA	0.70%

FIGURA 4.8 - TOP 7 EMPRESAS - VANGUARD EMERGING MARKETS STOCK INDEX FUND

Fonte: Yahoo Finance

4.1.5 iShares Core MSCI Emerging Markets ETF

ETF composto por ações de diferentes setores com a respectiva ponderação descrita na Figura 4.9 e o Top das nove empresas com maior relevo na Figura 4.10 - Top 9 Empresas - iShares Core MSCI Emerging Markets ETF

Sector Weightings (%)		
Sector(s)		IEMG
Basic Materials		8.53%
Consumer Cyclical		12.11%
Financial Services		20.54%
Real Estate		2.27%
Consumer Defensive		5.84%
Healthcare		4.81%
Utilities		2.75%
Communication Services		8.46%
Energy		4.70%
Industrials		8.34%
Technology		21.65%

FIGURA 4.9- SETORES - ISHARES CORE MSCI EMERGING MARKETS ETF

Fonte: Yahoo Finance

Top 9 Holdings (18.28% of Total Assets)

[Get Quotes for Top Holdings](#)

Name	Symbol	% Assets
Taiwan Semiconductor Manufacturing Co Ltd	2330.TW	5.75%
Samsung Electronics Co Ltd	005930.KS	3.52%
Tencent Holdings Ltd	0700.HK	3.03%
Alibaba Group Holding Ltd Ordinary Shares	9988.HK	1.91%
Reliance Industries Ltd	RELIANCE.NS	1.15%
ICICI Bank Ltd	ICICIBANK.NS	0.75%
Infosys Ltd	INFY.NS	0.74%
SK Hynix Inc	000660.KS	0.72%
HDFC Bank Ltd	HDFCBANK.NS	0.70%

FIGURA 4.10 - TOP 9 EMPRESAS - ISHARES CORE MSCI EMERGING MARKETS ETF

4.1.6 Vanguard Total International Bond Index Fund

ETF composto por obrigações de diferentes setores com a respectiva ponderação descrita na Figura 4.11 e os Ratings das obrigações na Figura 4.12

Sector Weightings (%)	
Sector(s)	BNDX
Basic Materials	0.00%
Consumer Cyclical	0.00%
Financial Services	100.00%
Real Estate	0.00%
Consumer Defensive	0.00%
Healthcare	0.00%
Utilities	0.00%
Communication Services	0.00%
Energy	0.00%
Industrials	0.00%
Technology	0.00%

FIGURA 4.11 - SETORES - VANGUARD TOTAL INTERNATIONAL BOND INDEX FUND

Fonte: Yahoo Finance

Bond Ratings	
Sector(s)	BNDX
US Government	78.36%
AAA	24.43%
AA	26.79%
A	27.57%
BBB	19.16%
BB	0.00%
B	0.00%
Below B	1.72%
Others	0.33%

FIGURA 4.12 - BOND RATINGS - VANGUARD TOTAL INTERNATIONAL BOND INDEX FUND

4.2 Periodicidade dos dados

De modo a obter resultados com representatividade e igualmente com horizontes temporais distintos, todos os dados extraídos da *Yahoo Finance* terão por base ETF's admitidos à negociação há pelo menos 10 anos, tendo sido extraídos dados de dois períodos distintos, (i) período pré pandemia do Coronavírus (COVID) de 01-01-2012 a 31-10-2019 e (ii) período com maior instabilidade onde se inclui a pandemia COVID e o início da invasão russa, de 01-01-2012 a 24-04-2022.

4.3 Seleção do modelo de estimação

Dado que se pretende obter a estimativa ou previsão da volatilidade dos rendimentos dos ETF's e uma vez que existem diversos modelos de regressão com recurso a aprendizagem de máquina, foi selecionado o algoritmo XGBOOST bem como o modelo econométrico GARCH, ambos implementados com recurso à linguagem de programação PYTHON.

5. Implementação, Análise e Resultados: Estudo Empírico

Para cada um dos seis produtos selecionados foram efetuadas duas análises comparativas, aplicando o modelo de GARCH e o algoritmo XGBOOST: uma análise para o período que designámos de “Período Estável”, com dados extraídos de 01 de janeiro de 2012 a 31 de outubro de 2019 e uma segunda análise para o período que designámos de “Período de Instabilidade” compreendido entre 01 de janeiro de 2012 a 24 de abril de 2022, contendo deste modo, informação para dois períodos de maior volatilidade, a pandemia do Coronavírus e o início da invasão Russa.

As imagens colocadas nos próximos tópicos foram geradas com recurso a algoritmos PYTHON, que permitiram uma extração bastante mais rápida e organizada de toda a informação.

Dado que o processo de extração e tratamento da informação foi exatamente o mesmo para todos os produtos, existe um maior detalhe no ponto 5.1, sendo que os restantes terão um resumo da informação que permitirá ao leitor da dissertação efetuar a respetiva comparação e inclusive e atingir os mesmos resultados que retirámos deste projeto.

5.1 Implementação e Análise - SPDR S&P 500 ETF Trust

- Período Estável – Modelo GARCH

A análise dos dados da amostra iniciou-se com a extração dos primeiros e últimos dados da amostra conforme detalhado na Figura 5.1 e Figura 5.2.

	Open	High	Low	Close	Adj Close	Volume
Date						
2012-01-03	127.760002	128.380005	127.430000	127.500000	102.045258	193697900
2012-01-04	127.199997	127.809998	126.709999	127.699997	102.205315	127186500
2012-01-05	127.010002	128.229996	126.430000	128.039993	102.477432	173895000
2012-01-06	128.199997	128.220001	127.290001	127.709999	102.213348	148050000
2012-01-09	128.000000	128.179993	127.410004	128.020004	102.461418	99530200

FIGURA 5.1 - 5 PRIMEIROS REGISTOS DA AMOSTRA

	Open	High	Low	Close	Adj Close	Volume
Date						
2019-10-24	300.910004	301.070007	299.459991	300.369995	280.672607	35453100
2019-10-25	299.739990	302.200012	299.679993	301.600006	281.821930	45205400
2019-10-28	302.940002	303.850006	302.910004	303.299988	283.410431	42147000
2019-10-29	303.000000	304.230011	302.859985	303.209991	283.326385	44284900
2019-10-30	303.429993	304.549988	301.989990	304.140015	284.195374	49643900

FIGURA 5.2 - 5 ÚLTIMOS REGISTOS DA AMOSTRA

De seguida, através da função “*skim*” foi possível verificar a composição da amostra: 1970 registos para um total de seis variáveis, sem valores nulos e sem valores em percentagem, com maior detalhe, no que concerne à estatística descritiva das mesmas, incluindo os percentis e um pequeno histograma, na Figura 5.3.

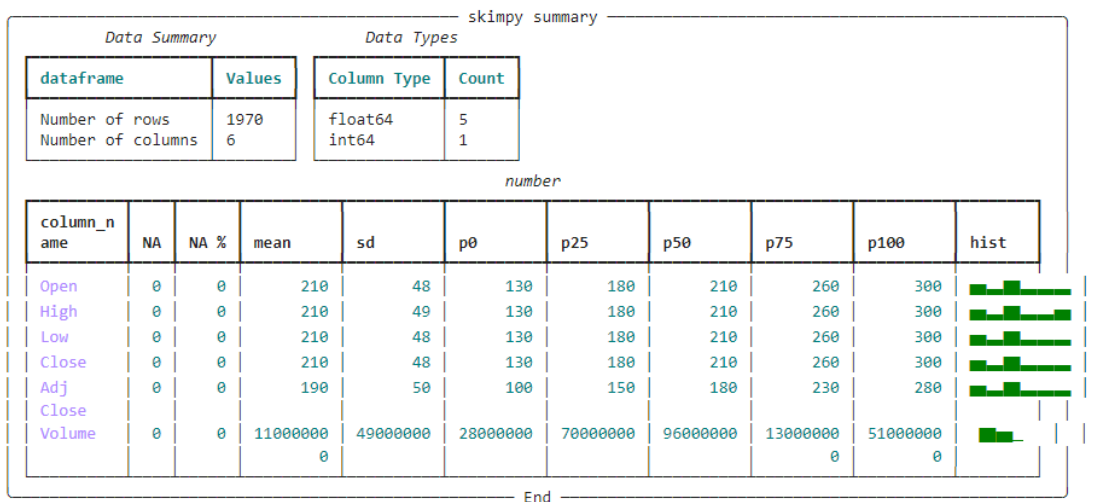


FIGURA 5.3 - ESTATÍSTICA DESCRITIVA, PERÍODO ESTÁVEL - SPDR S&P 500 ETF TRUST

O próximo passo foi a remoção das tendências da série de modo que seja possível modelar os dados. A forma mais simples é através do cálculo da variação percentual dos preços diários de fecho de mercado. Para tal, converte-se o preço diário num rendimento ou rentabilidade diária, através da seguinte formula:

$$\% \text{ retorno} = \left(\frac{P_2}{P_1} - 1 \right) * 100 \quad (5.1)$$

onde P2 é o preço do momento t, e P1 o preço no momento t-1 (dia anterior). Este processo foi facilmente implementado com recurso à função `DataFrame.pct_changes()`.

Após o cálculo das percentagens de retorno, foi gerado o gráfico dos rendimentos acumulados, conforme Figura 5.4, onde é possível verificar alguma linearidade dos mesmos, contudo, com alguns picos em determinados períodos.

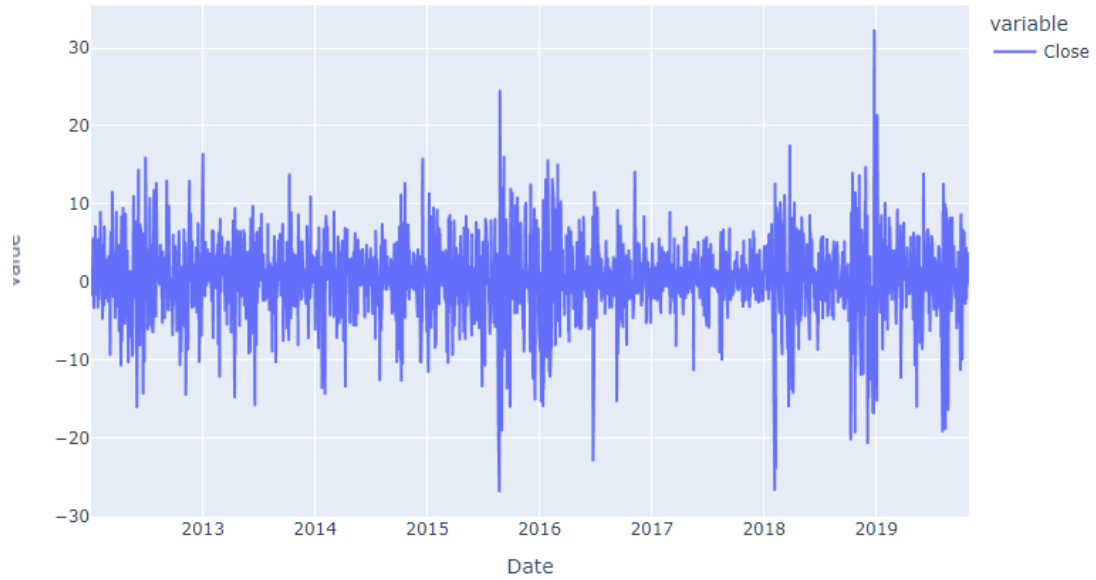


FIGURA 5.42 - GRÁFICO RENDIMENTOS ACUMULADOS - SPDR S&P 500 ETF TRUST

O passo seguinte foi o cálculo das volatilidades diária, mensal e anual, cuja informação se encontra disponível na Tabela 5.1 - Volatilidade SPDR S&P 500 ETF Trust (PERÍODO ESTÁVEL), bem como a geração do gráfico de retornos ou rendimentos diários, Figura 5.53

TABELA 5.1 - VOLATILIDADE SPDR S&P 500 ETF TRUST (PERÍODO ESTÁVEL)

	Volatilidade Diária %	Volatilidade Mensal %	Volatilidade Anual %
ETF	0.82	3.74	12.96

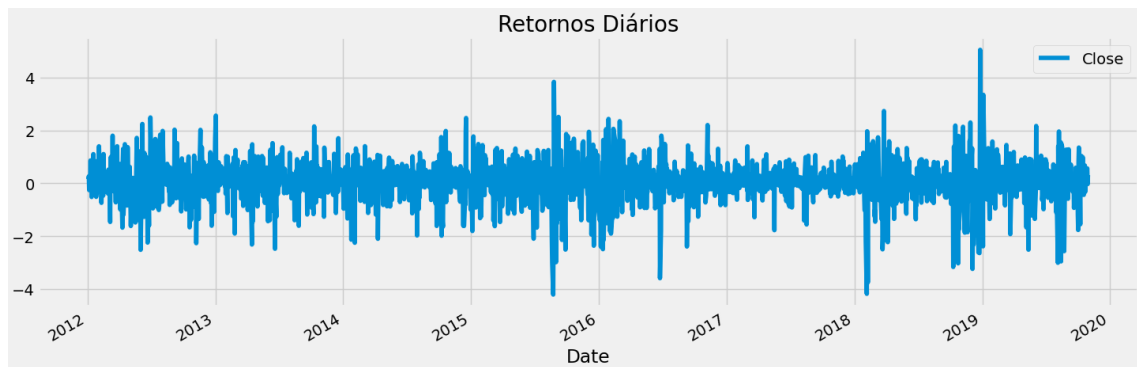


FIGURA 5.53 - GRÁFICO RETORNOS DIÁRIOS - SPDR S&P 500 ETF TRUST

A tendência da distribuição dos retornos diários é bastante evidente na Figura 5.6, com uma distribuição aproximadamente normal.

Distribuição dos Retornos Diários

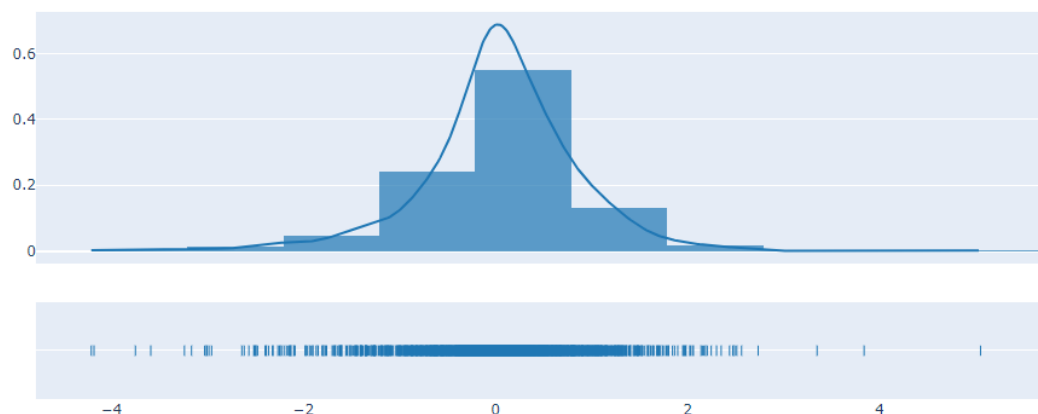


FIGURA 5.6 - GRÁFICO DISTRIBUIÇÃO RETORNOS DIÁRIOS - SPDR S&P 500 ETF TRUST

De modo a verificar a autocorreção parcial do quadrado dos retornos, foi utilizada a função *plot_pacf* que permite construir um gráfico (Figura 5.7), tornando bastante visual o resultado. Segundo a literatura, a existência de resultados positivos permite prever a volatilidade utilizando um modelo de GARCH.

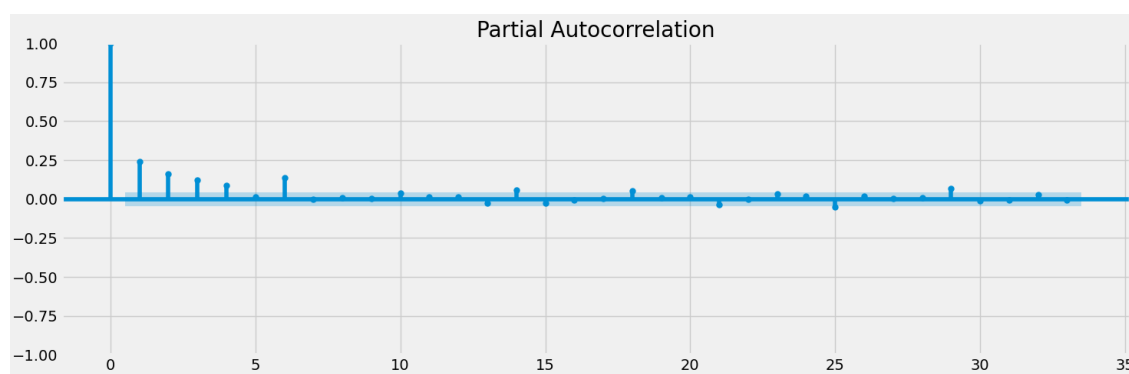


FIGURA 5.7 4- GRÁFICO AUTOCORRELAÇÃO PARCIAL DO QUADRADO DOS RETORNOS - SPDR S&P 500 ETF TRUST

Finalmente, passámos a construção do modelo GARCH, tendo sido selecionada a distribuição *Generalized Error Distribution* (GED) que pertence à família de distribuições exponenciais, oferecendo melhores resultados do que a *t-student*. Pese embora o gráfico de autocorreção parcial nos levasse a definir um modelo GARCH (4,4) ou até GARCH (5,5), devido ao número de resultados positivos mais significativos, de acordo com os testes realizados, o AIC e BIC são

inferiores num modelo GARCH (1,1) e deste modo decidimos avançar com este último, conforme é possível constatar as diferenças na Tabela 5.2.

TABELA 5.2 - PARÂMETROS ESTIMADOS - GARCH (1,1) VS GARCH (5,5) - SPDR S&P 500 ETF TRUST (PERÍODO ESTÁVEL)

Modelo	ω	α	β	Log L	AIC	BIC
Garch (1,1)	0,0392 [0,0213;0,05628]	0,1902 [0,136;0,244]	0,7607 [0,706;0,815]	-2122,62	4255,23	4283,16
		(1) 0,1632 [0,08701;0,239]	(1) 0,0000 [-0,928;0,928]			
		(2) 0,1407 [-0,002476; 0,282]	(2) 0,2909 [-0,141; 0,723]			
Garch (5,5)	0,0995 [-0,02012;0,219]	(3) 0,1087 [-0,07274; 0,290]	(3) 4,3896e-16 [-0,495; 0,495]	-2120,92	4267,84	4340,45
		(4) 0,0739 [-0,106; 0,254]	(4) 4,7460e-17 [-0,291; 0,291]			
		(5) 1,0632e-16 [-0,116;0,116]	(5) 0,1008 [-6,270e-02; 0,264]			

Na Tabela 5.2, Log L, representa o valor do logaritmo da função de verosimilhança, AIC é o critério de informação da *Akaike* e BIC o critério de informação *Baysiana*. Conforme é possível constatar, todos os parâmetros estimados do modelo, são estatisticamente significativos a um nível de significância de 5% (o intervalo de confiança a 95% encontra-se abaixo de cada um dos valores estimados), o Log L do modelo GARCH, sendo mais próximo do zero, seria um valor preferível, contudo os valores de AIC e BIC são inferiores no modelo GARCH (1,1).

Após a definição do modelo e dos respectivos dados de treino e de teste, gerámos um gráfico (Figura 5.8) que permite comparar o nosso modelo GARCH, com a volatilidade contínua dos retornos diários da média móvel, para os dados de treino.

GARCH vs Volatilidade Contínua dos Retornos diários - Treino

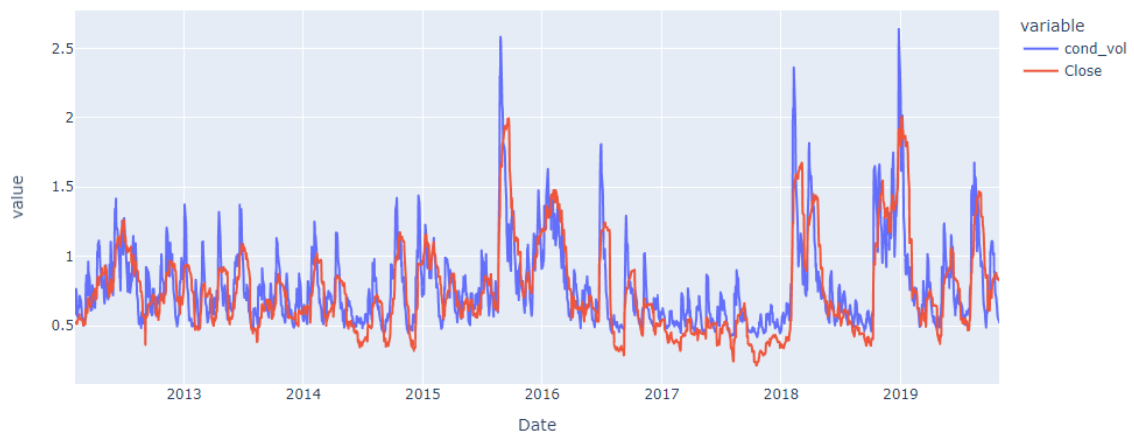


FIGURA 5.8 - GRÁFICO GARCH VS VOLATILIDADE CONTÍNUA DOS RETORNOS DIÁRIOS (TREINO) - SPDR S&P 500 ETF TRUST - PERÍODO ESTÁVEL

Foram considerados os últimos 60 dias da amostra, para a base de treino, sendo que após aplicação do modelo, geramos o gráfico que se encontra na Figura 5.9 e que demonstra a comparação entre os dados reais de teste e os previstos pelo modelo GARCH (1,1).

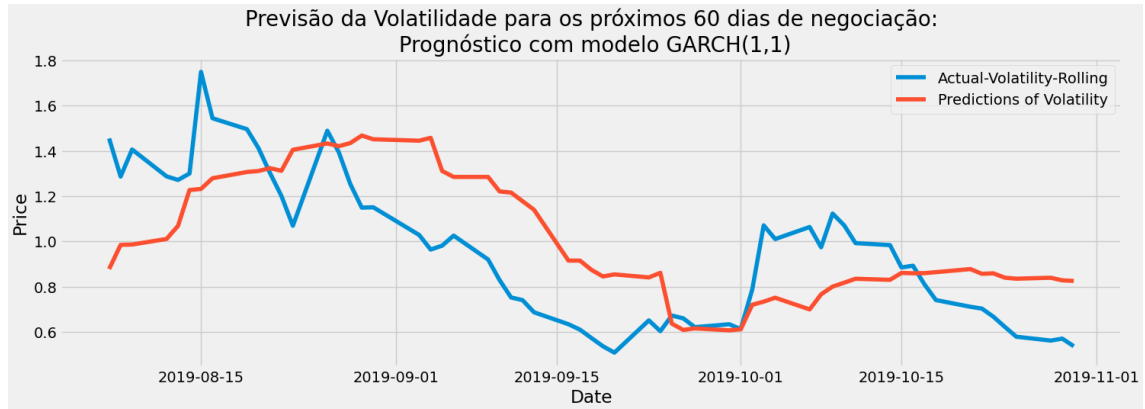


FIGURA 5.9 - GRÁFICO PREVISÃO VOLATILIDADE - SPDR S&P 500 ETF TRUST (PERÍODO ESTÁVEL) - MODELO GARCH

- Período Estável – Algoritmo XGBOOST

Os dados utilizados para a implementação do algoritmo XGBOOST foram exatamente os mesmos que foram definidos no modelo GARCH pelo que toda a componente de estatística descritiva não se torna relevante.

De modo a ser possível aplicar este algoritmo à análise de séries temporais, mais concretamente, na análise da volatilidade, é necessário incluir novas variáveis, nomeadamente a média movel a 22 dias (considerados 22 dias uteis em cada mês), bem como o dia da semana, o mês, o trimestre e o ano de cada uma das linhas da amostra, resultando na informação presente na Figura 5.10 onde prev1, prev2, prev3 e prev4 são as quatro médias moveis geradas para cada um dos dias.

	dayofweek	quarter	month	year	dayofyear	dayofmonth	prev1	prev2	prev3	prev4
21	4	1	2	2012	34	3	NaN	NaN	NaN	NaN
22	0	1	2	2012	37	6	0.520120	NaN	NaN	NaN
23	1	1	2	2012	38	7	0.524116	0.520120	NaN	NaN
24	2	1	2	2012	39	8	0.524086	0.524116	0.520120	NaN
25	3	1	2	2012	40	9	0.512461	0.524086	0.524116	0.520120
...
1904	2	3	7	2019	212	31	0.474883	0.468718	0.466092	0.512923
1905	3	3	8	2019	213	1	0.534944	0.474883	0.468718	0.466092
1906	4	3	8	2019	214	2	0.536627	0.534944	0.474883	0.468718
1907	0	3	8	2019	217	5	0.555362	0.536627	0.534944	0.474883
1908	1	3	8	2019	218	6	0.810250	0.555362	0.536627	0.534944

1888 rows x 10 columns

FIGURA 5.10 - VARIÁVEIS ADICIONADAS - SPDR S&P 500 ETF TRUST (PERÍODO ESTÁVEL) - ALGORITMO XGBOOST

De seguida passámos a implementação do algoritmo XGBOOST, tendo visualizado as variáveis com maior relevo na amostra, de acordo com a Figura 5.11.

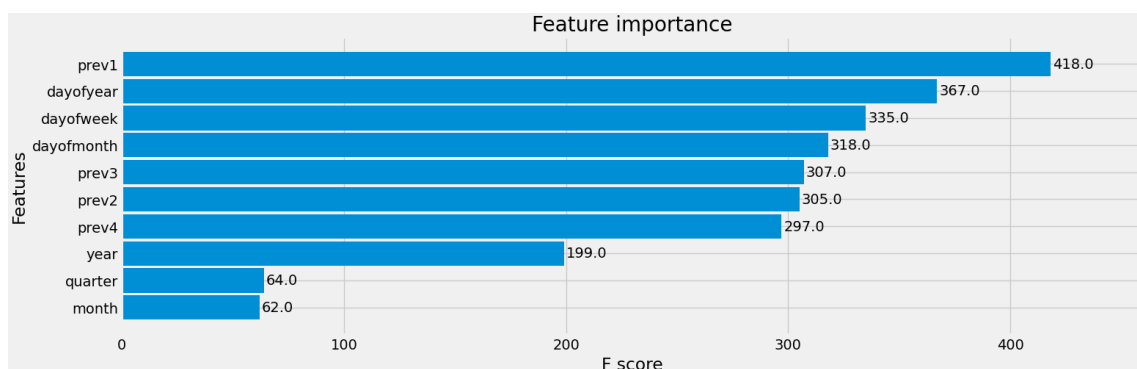


FIGURA 5.11- TOP IMPORTÂNCIA DAS VARIÁVEIS - SPDR S&P 500 ETF TRUST - ALGORITMO XGBOOST (PERÍODO ESTÁVEL)

Da mesma forma que comparámos o modelo GARCH com a volatilidade contínua dos retornos diários da média movel, também foi efetuada a comparação com os resultados do algoritmo XGBOOST para os dados de treino, conforme Figura 5.12

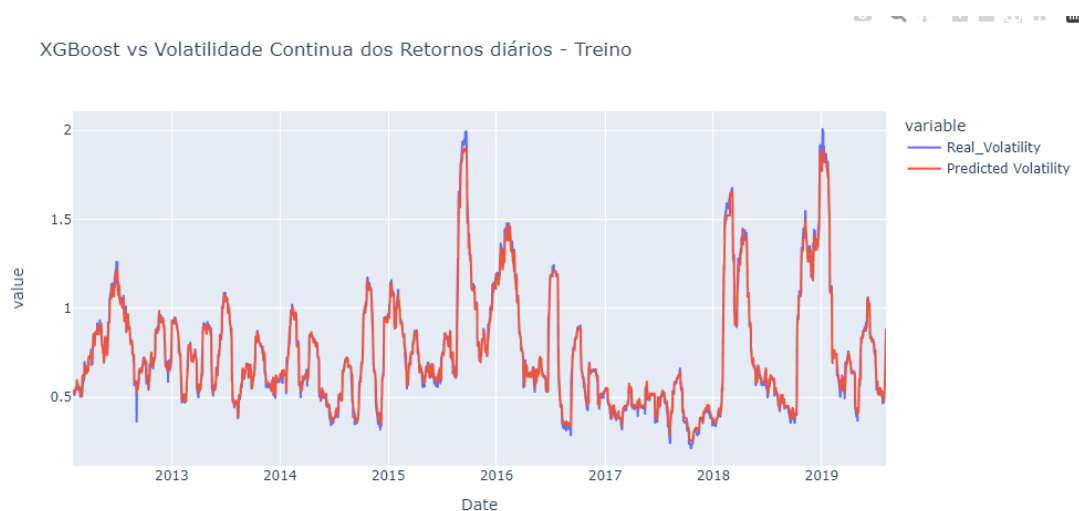


FIGURA 5.12- GRÁFICO XGBOOST VS VOLATILIDADE CONTINUA DOS RETORNOS DIÁRIOS (TREINO) - SPDR S&P 500 ETF TRUST - PERÍODO ESTÁVEL

Após aplicação do algoritmo para os dados de teste, foi gerado o gráfico de previsão de volatilidade (Figura 5.13).

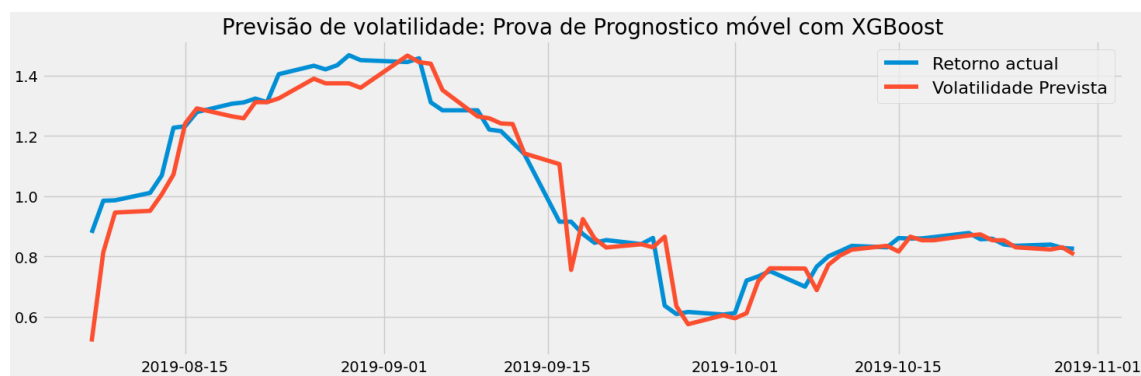


FIGURA 5.13- GRÁFICO PREVISÃO VOLATILIDADE - SPDR S&P 500 ETF TRUST (PERÍODO ESTÁVEL) - ALGORITMO XGBOOST

- Período de Instabilidade – Modelo GARCH

Tal como referido anteriormente, a implementação do modelo GARCH é exatamente igual para todos os produtos e independente dos períodos temporais, pelo que descrevemos apenas a informação que altera, tendo em consideração o período em tratamento.

De modo a constatar que a amostra selecionada é a correta, verificámos se os últimos registos da amostra correspondiam às últimas datas do período selecionado (ver Figura 5.145.14)

Date	Open	High	Low	Close	Adj Close	Volume
2022-04-18	436.809998	439.750000	435.609985	437.970001	425.844696	66002500
2022-04-19	437.859985	445.799988	437.679993	445.040009	432.718964	77821000
2022-04-20	446.920013	447.570007	443.480011	444.709991	432.398102	65224400
2022-04-21	448.540009	450.010010	437.100006	438.059998	425.932220	85417300
2022-04-22	436.910004	438.079987	425.440002	426.040009	414.244995	132471800

FIGURA 5.145 - 5 ÚLTIMOS REGISTOS DA AMOSTRA (PERÍODO DE INSTABILIDADE)

A estatística descritiva dos dados é demonstrada na Figura 5.15 abaixo, onde consta o número de linhas da amostra, neste caso 2594, que difere da amostra para o Período Estável que era de 1970, bem como a média, desvio padrão e percentis.

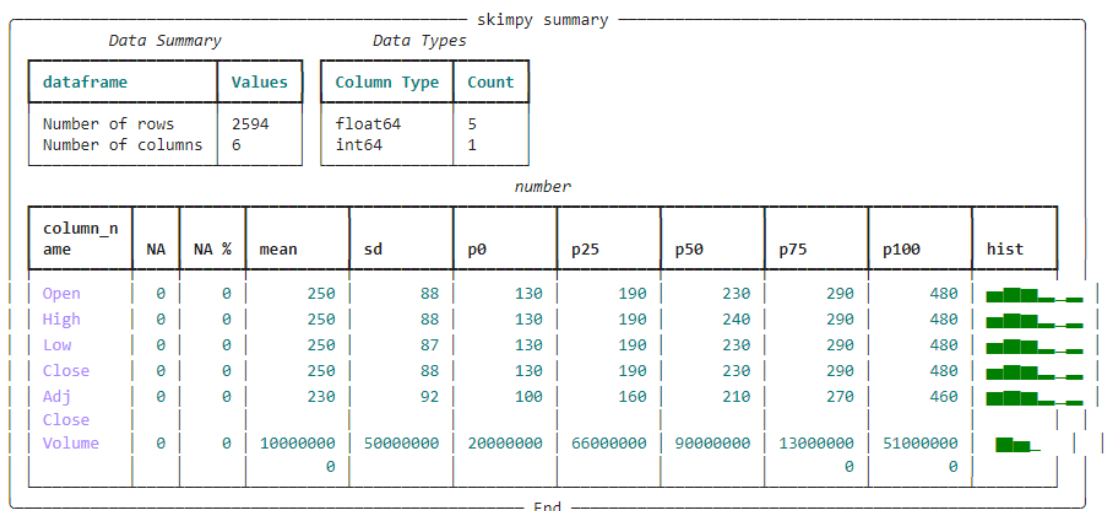


FIGURA 5.15 - ESTATÍSTICA DESCRITIVA, PERÍODO DE INSTABILIDADE - SPDR S&P 500 ETF TRUST

No gráfico representativo dos rendimentos acumulados (Figura 5.16) é possível verificar que em 2020, altura em que o surto da Covid-19 se espalhou pelo mundo, existiu uma variação bastante acentuada dos rendimentos diários acumulados.

Rendimentos acumulados

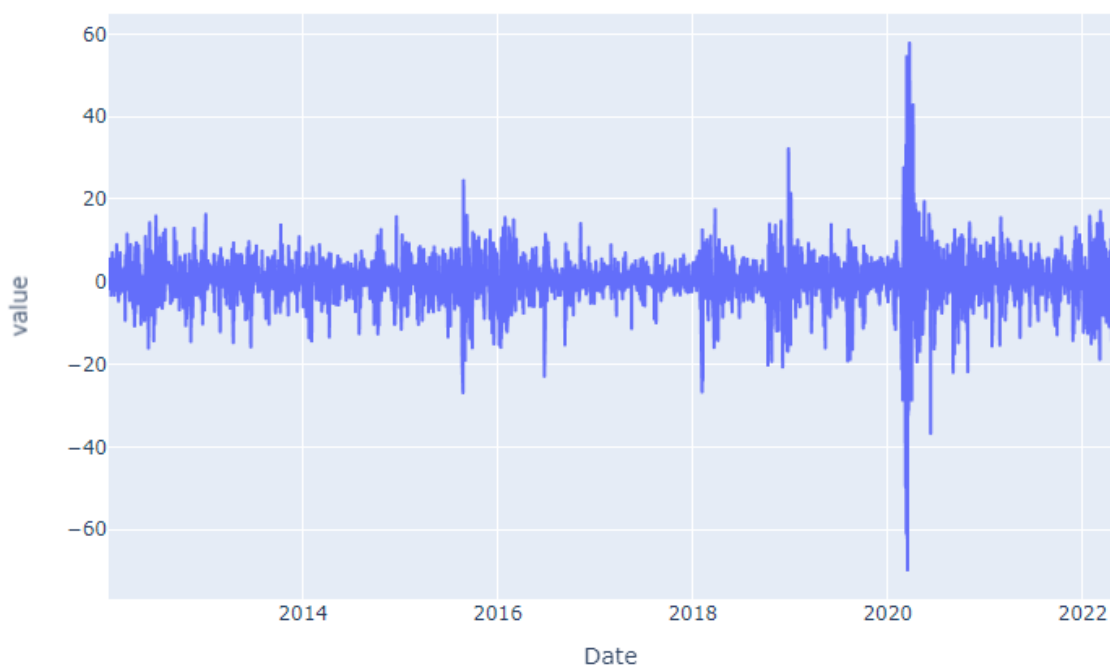


FIGURA 5.16 - GRÁFICO RENDIMENTOS ACUMULADOS - SPDR S&P 500 ETF TRUST - PERÍODO DE INSTABILIDADE

Também as volatilidades merecem ser referidas, uma vez que diferem das calculadas para o Período de Estabilidade (ver Tabela 5.3)

TABELA 5.3 - VOLATILIDADE SPDR S&P 500 ETF TRUST (PERÍODO DE INSTABILIDADE)

	Volatilidade Diária %	Volatilidade Mensal %	Volatilidade Anual %
ETF	1.03	4.72	16.35

Já em relação aos retornos diários, também o período marcado pela pandemia do Covid é bastante vincado no gráfico representativo destes retornos (Figura 5.176)

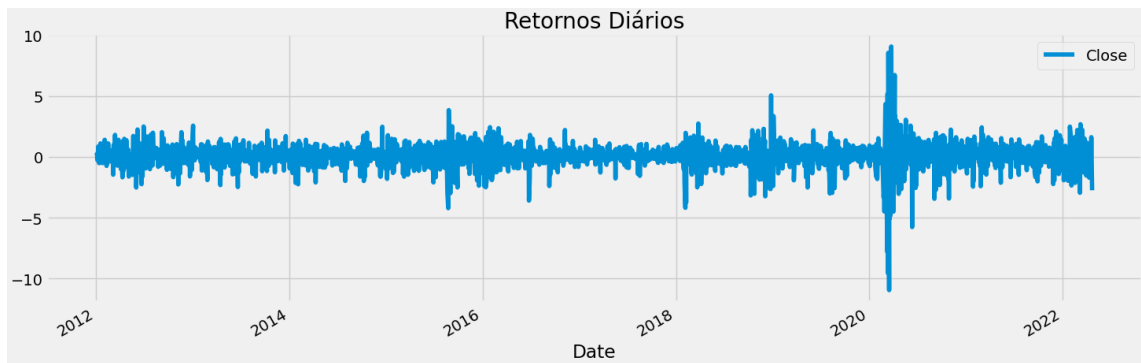


FIGURA 5.176 - GRÁFICO RETORNOS DIÁRIOS - SPDR S&P 500 ETF TRUST (PERÍODO DE INSTABILIDADE)

A distribuição dos retornos diários para o Período de Instabilidade é também próxima a uma distribuição normal conforme é possível verificar na Figura 5.18

Distribuição dos Retornos Diários

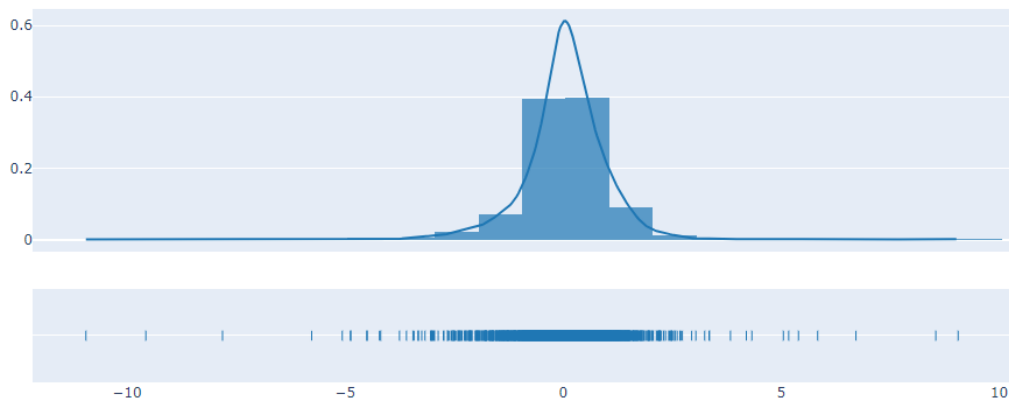


FIGURA 5.18 - GRÁFICO DISTRIBUIÇÃO RETORNOS DIÁRIOS - SPDR S&P 500 ETF TRUST (PERÍODO DE INSTABILIDADE)

Em termos de autocorrelação parcial a Figura 5.19, apresenta os resultados

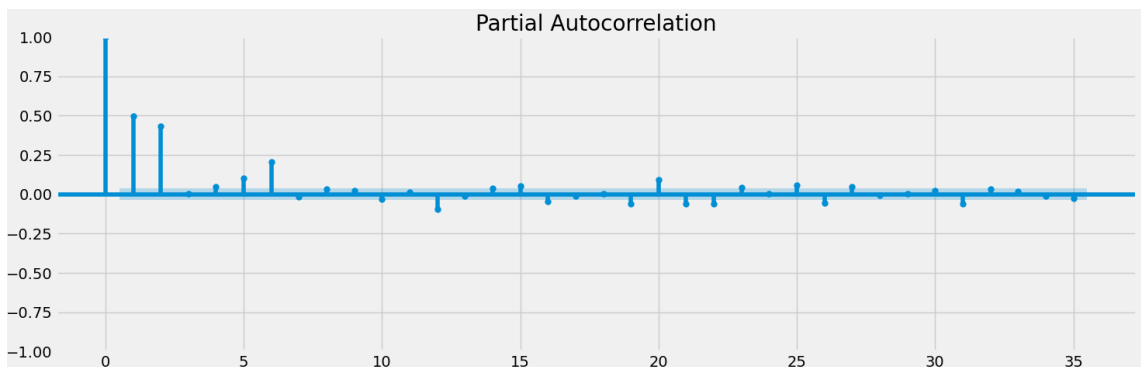


FIGURA 5.19 - GRÁFICO AUTOCORRELAÇÃO PARCIAL DO QUADRADO DOS RETORNOS - SPDR S&P 500 ETF TRUST (PERÍODO DE INSTABILIDADE)

Após implementação dos modelos GARCH, neste caso GARCH (1,1) e GARCH (4,4), os parâmetros estimados encontram-se referidos na Tabela 5.4.

TABELA 5.4 - PARÂMETROS ESTIMADOS - GARCH (1,1) VS GARCH (4,4) - SPDR S&P 500 ETF TRUST (PERÍODO DE INSTABILIDADE)

Modelo	ω	α	β	Log L	AIC	BIC
Garch (1,1)	0,0372 [0,0235;0,05214]	0,2064 [0,155;0,257]	0,7624 [0,714;0,811]	-3029,6	6069,19	6098,5
Garch (4,4)	0,0960 [0,05350;0,139]	(1) 0,1668 [0,09807;0,236]	(1) 0,0000 [-0,358;0,358]	-3028,17	6078,34	6142,8
		(2) 0,1913 [0,09989; 0,283]	(2) 0,2504 [-0,05723; 0,558]			
		(3) 0,0973 [0,007466; 0,187]	(3) 4,2165e-17 [-0,268; 0,268]			
		(4) 0,0624 [-0,006107; 0,131]	(4) 0,1521 [-0,05343; 0,358]			

Após a definição do modelo e dos respectivos dados de treino e de teste, gerámos um gráfico que permite comparar o nosso modelo GARCH, com a volatilidade contínua dos retornos diários da média móvel, para os dados de treino (Figura 5.20)

GARCH vs Volatilidade Contínua dos Retornos diários - Treino

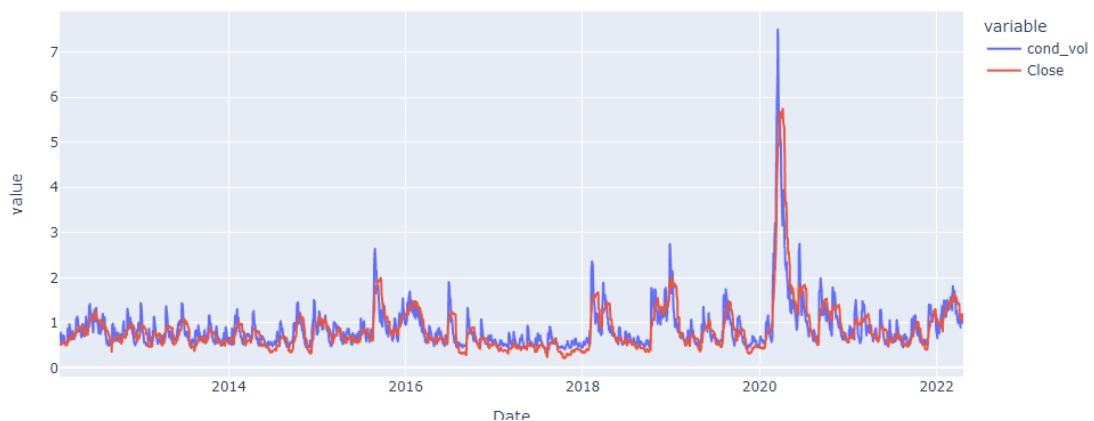


FIGURA 5.20 - GRÁFICO GARCH VS VOLATILIDADE CONTINUA DOS RETORNOS DIÁRIOS (TREINO) - SPDR S&P 500 ETF TRUST - PERÍODO DE INSTABILIDADE

Finalmente, construímos o gráfico de previsão da volatilidade para os próximos 60 dias de negociação, i.e., para o período de teste, Figura 5.21.

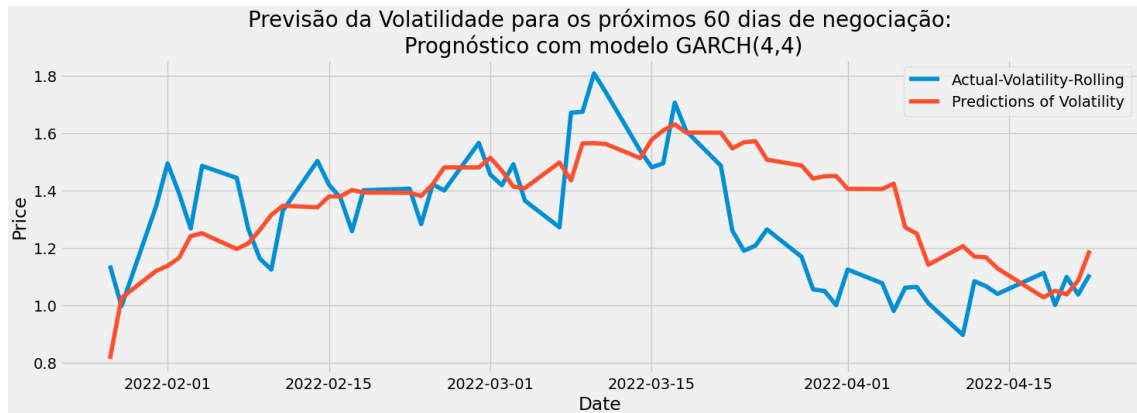


FIGURA 5.21 - GRÁFICO PREVISÃO VOLATILIDADE - SPDR S&P 500 ETF TRUST (PERÍODO DE INSTABILIDADE) - MODELO GARCH

- Período de Instabilidade – Algoritmo XGBOOST

Tal como foi efetuado para o Período Estável incluímos novas variáveis, nomeadamente a média movel a 22 dias (considerados 22 dias uteis em cada mês), bem como o dia da semana, o mês, o trimestre e o ano de cada uma das linhas da amostra, resultando na informação presente na Figura 5.22, onde prev1, prev2, prev3 e prev4 são as quatro médias moveis geradas para cada um dos dias.

	dayofweek	quarter	month	year	dayofyear	dayofmonth	prev1	prev2	prev3	prev4
21	4	1	2	2012	34	3	NaN	NaN	NaN	NaN
22	0	1	2	2012	37	6	0.520120	NaN	NaN	NaN
23	1	1	2	2012	38	7	0.524116	0.520120	NaN	NaN
24	2	1	2	2012	39	8	0.524086	0.524116	0.520120	NaN
25	3	1	2	2012	40	9	0.512461	0.524086	0.524116	0.520120
...
2528	3	1	1	2022	20	20	0.990739	0.984530	0.977028	0.988502
2529	4	1	1	2022	21	21	0.974173	0.990739	0.984530	0.977028
2530	0	1	1	2022	24	24	1.032495	0.974173	0.990739	0.984530
2531	1	1	1	2022	25	25	0.948104	1.032495	0.974173	0.990739
2532	2	1	1	2022	26	26	0.929038	0.948104	1.032495	0.974173

FIGURA 5.22 - VARIÁVEIS ADICIONADAS - SPDR S&P 500 ETF TRUST (PERÍODO DE INSTABILIDADE) - ALGORITMO XGBOOST

De seguida passámos a implementação do algoritmo XGBOOST, tendo visualizado as variáveis com maior relevo na amostra, de acordo com a Figura 5.23

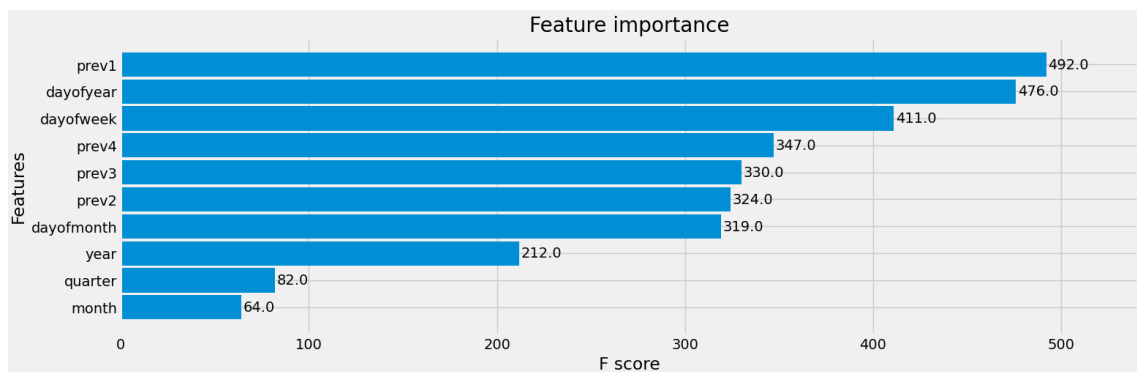


FIGURA 5.23 - TOP IMPORTÂNCIA DAS VARIÁVEIS - SPDR S&P 500 ETF TRUST - ALGORITMO XGBOOST (PERÍODO DE INSTABILIDADE)

Da mesma forma que comparámos o modelo GARCH com a volatilidade contínua dos retornos diários da média móvel, também foi efetuada a comparação com os resultados do algoritmo XGBOOST para os dados de treino, conforme Figura 5.24

XGBoost vs Volatilidade Contínua dos Retornos diários - Treino

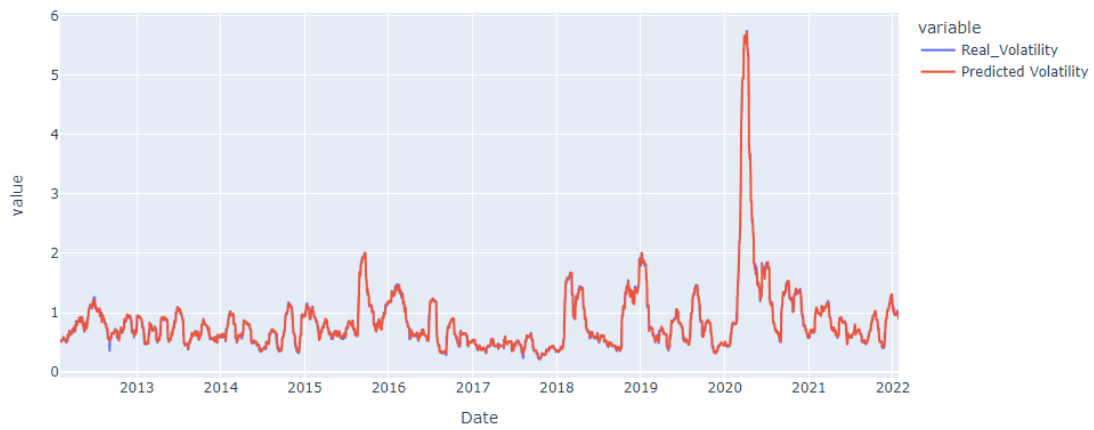


FIGURA 5.24 - GRÁFICO XGBOOST VS VOLATILIDADE CONTINUA DOS RETORNOS DIÁRIOS (TREINO) - SPDR S&P 500 ETF TRUST - PERÍODO DE INSTABILIDADE

Por último gerámos o gráfico de previsão para os dados de teste (Figura 5.25)

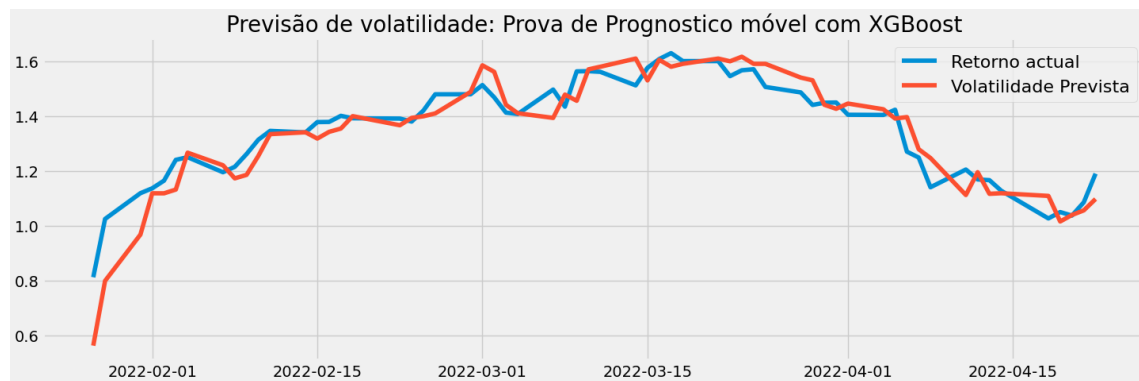


FIGURA 5.25 - GRÁFICO PREVISÃO VOLATILIDADE - SPDR S&P 500 ETF TRUST (PERÍODO DE INSTABILIDADE) - ALGORITMO XGBOOST

5.1.1 Resultados: SPDR S&P 500 ETF Trust

Tal como descrito na revisão bibliográfica, foram selecionadas as métricas MAPE e RMSE para efetuar a comparação dos resultados obtidos pelo modelo GARCH em contraste com o algoritmo XGBOOST, sendo que os resultados se encontram expressos em percentagem (%), na Tabela 5.5.

TABELA 5.5- RESULTADOS MÉTRICAS MAPE E RMSE - SPDR S&P 500 ETF TRUST

Período	Medidas	GARCH DADOS TREINO	XGBOOST DADOS TREINO	GARCH PREVISÕES	XGBOOST PREVISÕES
Estável	MAPE	21,83	3,8	23,66	5,83
	RMSE	20,62	3,82	27,56	8,2
Instável	MAPE	22,98	3,48	12,07	4,64
	RMSE	29,27	3,99	20,39	7,24

Da análise efetuada é importante referir que os valores calculados para as métricas MAPE e RMSE são bastante inferiores no algoritmo XGBOOST, o que já se antevia comparando os gráficos de previsão para ambas as realidades. O algoritmo XGBOOST apresenta resultados aproximados da realidade. É de ressaltar que sendo um algoritmo de aprendizagem de máquina, o mesmo é treinado para aprender com os erros anteriores, tornando-se cada vez mais eficiente. Paralelamente é possível verificar que, embora o período de instabilidade apresente uma volatilidade superior, as medidas MAPE e RMSE são significativamente melhores quer no modelo GARCH como no algoritmo XGBOOST.

5.1.2 Resultados: iShares Core S&P 500 ETF

O resultado das medidas MAPE e RMSE encontram-se disponível na Tabela 5.6

TABELA 5.6- RESULTADOS MÉTRICAS MAPE E RMSE - ISHARES CORE S&P 500 ETF

Período	Medidas	GARCH DADOS TREINO	XGBOOST DADOS TREINO	GARCH PREVISÕES	XGBOOST PREVISÕES
Estável	MAPE	21,95	3,61	24,03	7,83
	RMSE	20,66	3,64	27,46	9,93
Instável	MAPE	22,43	3,83	11,87	4,84
	RMSE	29,02	4,69	20,37	7,35

Tal como na análise para o *SPDR S&P 500 ETF Trust*, também aqui é possível verificar que as métricas são inferiores no algoritmo XGBOOST, destacando-se uma melhoria nos resultados no período de instabilidade.

5.1.3 Resultados: Vanguard Total Stock Market Index Fund ETF Shares

O resultado das medidas MAPE e RMSE encontram-se disponível na Tabela 5.7

TABELA 5.7- RESULTADOS MÉTRICAS MAPE E RMSE - VANGUARD TOTAL STOCK MARKET INDEX FUND ETF SHARES

Período	Medidas	GARCH DADOS TREINO	XGBOOST DADOS TREINO	GARCH PREVISÕES	XGBOOST PREVISÕES
Estável	MAPE	21,47	3,42	22,5	5,45
	RMSE	20,4	3,56	25,96	7,75
Instável	MAPE	21,9	3,51	11,78	5,13
	RMSE	28,41	4,13	20,78	8,42

A par das análises anteriores, é possível constar o mesmo nível de resultados nas medidas MAPE e RMSE e respetivo comportamento face aos Períodos Estável e de Instabilidade.

5.2 Implementação e Análise - Vanguard Emerging Markets Stock Index Fund

Tal como foi evidenciado no ponto 5.1 descrevemos a implementação e análise para o ETF *Vanguard Emerging Markets Stock Index Fund*, dado que existem alguns aspetos a destacar tratando-se de um produto de um mercado emergente, cujo comportamento é distinto do mercado desenvolvido. De todo o modo nos pontos que se seguem apenas iremos apresentar os resultados tal como efetuámos para os dois restantes ETF's de mercados desenvolvidos.

- Período Estável – Modelo GARCH

O primeiro aspeto a destacar é o crescimento bastante mais desacelerado das cotações destes ETFs.

A Figura 5.26 apresenta as cotações das primeiras datas da amostra e a Figura 5.27 as últimas datas. É possível deste modo verificar que num prazo de mais de sete anos as cotações apenas alteraram 7,25% aproximadamente, o que contrasta com o aumento 138,5 % de aumento nas cotações do *ETF SPDR S&P 500 ETF Trust*. Este contraste deve-se ao ritmo de crescimento das empresas nos mercados desenvolvidos ser exponencial em comparação com o mercado emergente onde o crescimento das empresas se encontra bastante estagnado ao longo do tempo.

	Open	High	Low	Close	Adj Close	Volume
Date						
2012-01-03	39.209999	39.480000	39.119999	39.330002	27.964165	29435300
2012-01-04	38.970001	39.250000	38.869999	39.220001	27.885946	13972300
2012-01-05	38.930000	39.099998	38.619999	39.029999	27.750860	18686900
2012-01-06	38.970001	38.990002	38.500000	38.570000	27.423784	21649100
2012-01-09	38.889999	39.020000	38.700001	38.880001	27.644211	22563300

FIGURA 5.26 - 5 PRIMEIROS REGISTOS DA AMOSTRA - VANGUARD EMERGING MARKETS STOCK INDEX FUND

	Open	High	Low	Close	Adj Close	Volume
Date						
2019-10-24	41.860001	41.880001	41.689999	41.740002	36.450066	11299600
2019-10-25	41.720001	41.990002	41.689999	41.930000	36.615990	9776500
2019-10-28	42.180000	42.349998	42.160000	42.259998	36.904160	9763100
2019-10-29	42.000000	42.119999	41.959999	42.020000	36.694580	8959700
2019-10-30	41.900002	42.180000	41.740002	42.180000	36.834309	13014400

FIGURA 5.27 - 5 ÚLTIMOS REGISTOS DA AMOSTRA - VANGUARD EMERGING MARKETS STOCK INDEX FUND

A volatilidade (Tabela 5.8) é outro aspeto de relevo dado que se apresenta superior face aos ETF's de mercados desenvolvidos, pese embora o período temporal seja o mesmo.

TABELA 5.8 - VOLATILIDADE VANGUARD EMERGING MARKETS STOCK INDEX FUND (PERÍODO ESTÁVEL)

	Volatilidade Diária %	Volatilidade Mensal %	Volatilidade Anual %
ETF	1.15	5.27	18.27

Após implementação do modelo GARCH, neste caso GARCH (1,1), os parâmetros estimados encontram-se referidos na Tabela 5.9

TABELA 5.9 - PARÂMETROS ESTIMADOS - GARCH (1,1) - VANGUARD EMERGING MARKETS STOCK INDEX FUND (PERÍODO ESTÁVEL)

Modelo	ω	α	β	Log L	AIC	BIC
Garch (1,1)	0,0569 [0,01147;0,102]	0,1081 [0,05670;0,160]	0,8510 [0,776;0,926]	-2970,91	5951,81	5979,74

Conforme é possível constatar, todos os parâmetros estimados do modelo, são estatisticamente significativos a um nível de significância de 5% (o intervalo de confiança a 95% encontra-se abaixo de cada um dos valores estimados).

- Período Estável – Algoritmo XGBOOST

Relativamente ao algoritmo XGBOOST não existe nenhuma particularidade a destacar, tendo o mesmo sido implementado de igual modo.

- Período de Instabilidade – Modelo GARCH

No que respeita ao período de instabilidade destacamos uma vez a volatilidade calculada, que é superior à dos ETF's disponíveis em mercado desenvolvido, mesmo num período temporal de elevada instabilidade (Tabela 5.10).

TABELA 5.10 - VOLATILIDADE VANGUARD EMERGING MARKETS STOCK INDEX FUND (PERÍODO DE INSTABILIDADE)

	Volatilidade Diária %	Volatilidade Mensal %	Volatilidade Anual %
ETF	1.27	5.82	20.17

Em relação aos parâmetros estimados para o modelo GARCH os resultados encontram-se na Tabela 5.11.

TABELA 5.11 - PARÂMETROS ESTIMADOS - GARCH (1,1) - VANGUARD EMERGING MARKETS STOCK INDEX FUND (PERÍODO DE INSTABILIDADE)

Modelo	ω	α	β	Log L	AIC	BIC
Garch (1,1)	0,0648 [0,02544;0,104]	0,1118 [0,07234;0,151]	0,8460 [0,790;0,902]	-4011,99	8033,97	8063,28

- Período de Instabilidade – Algoritmo XGBOOST

Relativamente ao algoritmo XGBOOST não existe nenhuma particularidade a destacar, tendo o mesmo sido implementado de igual modo.

5.2.1 Resultados: Vanguard Emerging Markets Stock Index Fund

O resultado das medidas MAPE e RMSE encontram-se disponível na Tabela 5.12.

TABELA 5.12 - RESULTADOS MÉTRICAS MAPE E RMSE - VANGUARD EMERGING MARKETS STOCK INDEX FUND

Periodo	Medidas	GARCH DADOS TREINO	XGBOOST DADOS TREINO	GARCH PREVISÕES	XGBOOST PREVISÕES
Estável	MAPE	14,03	3,29	15,07	4,35
	RMSE	18,1	4,96	17,98	6,37
Instável	MAPE	14,73	3,25	13,05	6,27
	RMSE	24	5,44	36,57	21,68

Semelhante às análises anteriores, é possível constar o mesmo nível de resultados nas medidas MAPE e RMSE e respetivo comportamento face aos Períodos Estável e de Instabilidade, contudo, a diferença entre os resultados das medidas nas previsões do modelo Garch vs algoritmo XGBOOST é inferior, e no período de instabilidade o RMSE para o XGBOOST é bastante mais elevado do que aquele que foi calculado para os ETFs de mercados desenvolvidos (21,68 em contraste com resultados inferiores a 10,00)

5.2.2 Resultados: iShares Core MSCI Emerging Markets ETF

O resultado das medidas MAPE e RMSE encontram-se disponível na Tabela 5.13.

TABELA 5.13 - RESULTADOS MÉTRICAS MAPE E RMSE - ISHARES CORE MSCI EMERGING MARKETS ETF

Período	Medidas	GARCH DADOS TREINO	XGBOOST DADOS TREINO	GARCH PREVISÕES	XGBOOST PREVISÕES
Estável	MAPE	15,11	2,66	17,18	4,71
	RMSE	19,49	3,97	19,50	6,29
Instável	MAPE	15,46	5,34	14,07	7,65
	RMSE	26,53	9,72	38,40	16,56

A par das análises anteriores, é possível constar o mesmo nível de resultados nas medidas MAPE e RMSE e respetivo comportamento face ao Períodos Estável e de Instabilidade.

5.3 Análise Vanguard Total International Bond Index Fund

No que respeita ao Vanguard Total International Bond Index Fund,

existem alguns aspetos de relevo que devem ser referidos uma vez que se trata de um ETF composto por Obrigações o que difere de todos os outros produtos que eram compostos por Ações.

A volatilidade dos retornos, Tabela 5.14, é bastante inferior a todos os anteriores produtos o que por si só faz com que os gráficos dos rendimentos diários ou acumulados, Figura 5.28 e Figura 5.29 sejam bastante distintos dos restantes ETFs

TABELA 5.14 - VOLATILIDADE VANGUARD TOTAL INTERNATIONAL BOND INDEX FUND

	Volatilidade Diária %	Volatilidade Mensal %	Volatilidade Anual %
ETF	0.19	0.86	2.98

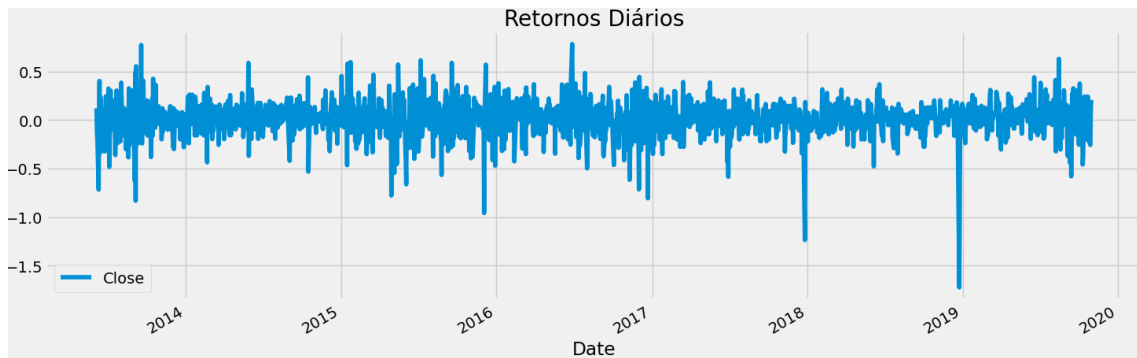


FIGURA 5.28 - GRÁFICO RETORNOS DIÁRIOS VANGUARD TOTAL INTERNATIONAL BOND INDEX FUND

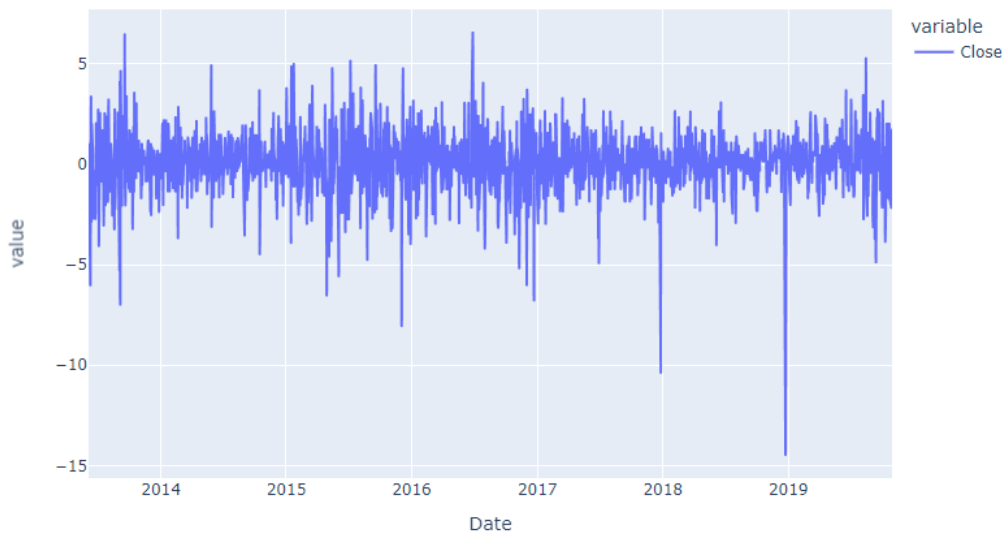


FIGURA 5.29 - GRÁFICOS RENDIMENTOS ACUMULADOS - VANGUARD TOTAL INTERNATIONAL BOND INDEX FUND

O ambiente de taxas de juro negativas ou perto de zero, faz com que estes produtos não fossem tão atrativos aos investidores e como tal os seus retornos e respetiva volatilidade é bastante diminuta.

A par das considerações acima, também a distribuição dos retornos diários (Figura 5.30) e o gráfico de autocorrelação parcial (Figura 5.31) têm alterações relevantes.

Distribuição dos Retornos Diários

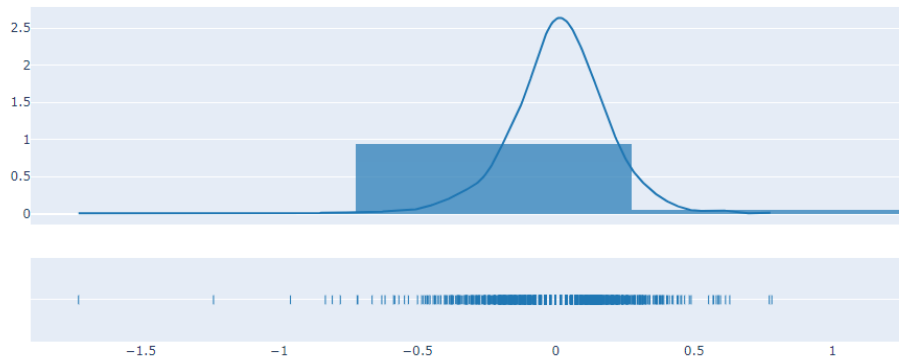


FIGURA 5.30 - GRÁFICO DISTRIBUIÇÃO RETORNOS DIÁRIOS VANGUARD TOTAL INTERNATIONAL BOND INDEX FUND

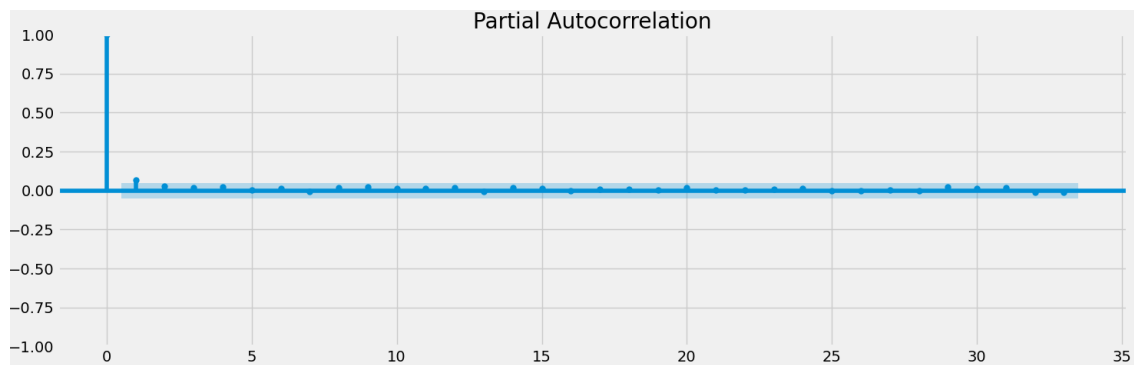


FIGURA 5.31- GRÁFICO AUTOCORRELAÇÃO PARCIAL DO QUADRADO DOS RETORNOS - VANGUARD TOTAL INTERNATIONAL BOND INDEX FUND

Os parâmetros estimados aquando da implementação do modelo e Garch (1,1), encontram-se descritos na Tabela 5.15.

TABELA 5.15- PARÂMETROS ESTIMADOS - GARCH (1,1) - VANGUARD TOTAL INTERNATIONAL BOND INDEX FUND

Modelo	ω	α	β	Log L	AIC	BIC
Garch (1,1)	0,0030150 [-0,001386;0,007416]	0,1401 [-0,01569;0,296]	0,7805 [0,532;1,029]	552,171	-1094,34	-1067,41

5.3.1 Resultados: Vanguard Total International Bond Index Fund

O resultado das medidas MAPE e RMSE encontram-se disponível na Tabela 5.16

TABELA 5.16 - RESULTADOS MÉTRICAS MAPE E RMSE - VANGUARD TOTAL INTERNATIONAL BOND INDEX FUND

Periodo	Medidas	GARCH DADOS TREINO	XGBOOST DADOS TREINO	GARCH PREVISÕES	XGBOOST PREVISÕES
Estável	MAPE	22,14	4,11	15,05	5,37
	RMSE	4,58	0,95	4,14	1,9
Instável	MAPE	24,27	3,67	18,84	15,96
	RMSE	7,29	1,08	9,71	7,87

6. Conclusões

6.1 Conclusões Gerais

Esta dissertação tinha como objetivos principais prever a volatilidade dos retornos de ETF's disponíveis em mercados desenvolvidos comparativamente aos disponíveis em mercados emergentes para o período inicial da invasão russa e comparar a eficácia do modelo GARCH com o algoritmo de inteligência artificial XGBOOST.

Relativamente às previsões, quer através do modelo GARCH como através do XGBOOST as previsões realizadas permitem traçar uma tendência/comportamento, i.e., quando a volatilidade real aumenta os modelos preveem igualmente um aumento e quando a volatilidade real desce os modelos comportam-se de igual modo quanto aos seus valores previstos, contudo, não foram alcançados valores iguais aos valores reais, existindo métricas de erro como o MAPE e o RMSE que estiveram praticamente sempre acima de 15 a 20% especialmente no modelo GARCH (com resultados menos satisfatórios nos dados de teste).

De todo o modo, efetuando uma comparação entre modelos, através da implementação do algoritmo XGBOOST, as métricas MAPE e RMSE apresentam resultados substancialmente mais satisfatórios, uma vez que quer nos dados de treino quer nos dados de teste os resultados estão na grande maioria das vezes abaixo de 10%, sendo que em alguns casos estão inclusive abaixo dos 5%, o que demonstra que o recurso à inteligência artificial em aprendizagem de máquina, é uma ferramenta muito útil, com resultados promissores e que poderão auxiliar analistas, intermediários financeiros, gestores de risco e até os próprios investidores a encontrarem o melhor momento para investir ou vender determinados produtos financeiros.

Paralelamente, o estudo realizado permitiu-nos retirar conclusões adicionais, nomeadamente no que respeita à volatilidade dos ETF's disponíveis em mercados emergentes vs mercados desenvolvidos e também comparar a evolução dos preços desses ativos.

Quer para os ETFs de mercados desenvolvidos como para os de mercados emergentes a volatilidade é superior quando temos em consideração eventos adversos. Como exemplo temos o SPDR S&P 500 ETF Trust (mercado desenvolvido) cujas volatilidades diárias, mensais e anuais são de 0,82%, 3,74% e 12,96% respetivamente, no período estável, e de 1,03%, 4,72% e

16,35% no período de instabilidade. Já em relação aos mercados emergentes o comportamento é semelhante, sendo que por exemplo para o Vanguard Emerging Markets Stock Index Fund as volatilidades no período de estabilidade são 1,15%, 5,27% e 18,27% e quando a amostra tem em consideração o período de instabilidade, aumentam para 1,27%, 5,82% e 20,17%. É igualmente de referir que se forem comparados os dois produtos/mercados, a volatilidade para ETF's de mercados emergentes é superior aos ETF's dos mercados desenvolvidos, devido à insegurança dos investidores e instabilidade das próprios emittentes dos subjacentes, que está bastante presente nestes mercados. Este tópico pode ser correlacionado igualmente à evolução dos preços dos ETF's. Quando estamos perante um produto de um mercado desenvolvido como é o caso do ETF SPDR S&P 500 ETF Trust verificamos um crescimento no preço de aproximadamente 138,5% o que comparando com um ETF de mercado emergente, por exemplo o Vanguard Emerging Markets Stock Index Fund, o crescimento é de apenas 7,25%, um contraste bastante acentuado que se deve não só mas também ao ritmo exponencial de crescimento das empresas nos mercados desenvolvidos em comparação com o mercado emergente onde o crescimento das empresas se encontra bastante estagnado ao longo do tempo.

6.2 Estudos futuros

Como reflexão sobre o trabalho realizado é possível enumerar algumas sugestões para trabalhos futuro de investigação cuja temática seja semelhante.

Uma vez que este estudo aborda apenas a comparação entre um modelo estatístico de regressão, GARCH e um algoritmo de inteligência artificial poderá ser alargada a comparação, incluindo outros modelos estatísticos mais evoluídos e recentes e algoritmos também distintos de modo a perceber se os resultados serão semelhantes.

Paralelamente o estudo poderá ser alargado a outras tipologias de produtos financeiros (Ações, Fundos de Capital de Risco, Fundos de Investimento Imobiliários, Seguros de Investimento, entre outros)

Referências Bibliográficas

- Adambekova, A. A., & Andekina, R. E. (2013). Financial Market and its Definitions: Transformation of Scientific Concepts. *World Applied Sciences Journal (Education, Law, Economics, Language and Communication)* 27, 12-16
- Bentes, Sónia (2011) – Sobre a Medição da Volatilidade nos Mercados Bolsistas Internacionais: Evidência dos Países do G7. Lisboa: *Edições Colibri/Instituto Politécnico de Lisboa*. 230 pp. ISBN 978-989-689-124-4
- Bressan, A. A. (2004). Tomada de decisão em futuros agropecuários com modelos de previsão de séries temporais. *RAE-Eletrônica*. 3(1). Disponível em <https://www.scielo.br/j/raeel/a/pyJSSHNTjgMQdFHtFdnYXJ/?format=pdf&lang=pt>
- Brooks, C (2002) *Econometrics for Finance*, Cambridge University Press
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, pp. 307-327. doi:10.1016/0304-4076(86)90063-1
- Chen, T. e Guestrin, C. (2016), ‘Xgboost: A scalable tree boosting system’, 785–794
- Choudhary, A., Kumar, S., Sharma, M., & Sharma, K. P. (2022). A Framework for Data Prediction and Forecasting in WSN with Auto ARIMA. *Wireless Personal Communications*, 123(3), 2245–2259.
- Coates, J (2012), *The Hour Between Dog and Wolf: Risk Taking, Gut Feelings and the Biology of Boom and Bust*, Fourth Estate
- Domingos, P. (2015), *O Algoritmo Mestre: Como a Busca pelo Algoritmo de Aprendizagem Definitivo Transformará o Nosso Mundo*, Lua de Papel.
- Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4). 987–1007, ISSN 00129682, 14680262. Disponível em: <http://www.econ.uiuc.edu/~econ536/Papers/engle82.pdf>

- Ferreira, D. (2010). *Futuros e Outros Derivados Ganhar (e não perder) nas bolsas e nos mercados OTC Evitar rogue traders*. (2ª Edição). Lisboa: Edições Sílabo
- Gujarati, D., & Porter, D. (2008). *Basic Econometrics* (5th ed.). McGraw-Hill.
- Hull, J. C. (2013). *Fundamentals of Futures and Options Markets: International Edition*. Pearson Education Limited.
- Jilmin, J. (2016, 23 de Março). MAE and RMSE — Which Metric is Better? - Human in a Machine World. Acedido a Disponível em: <https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-isbetter-e60ac3bde13d>
- Konishi, S., & Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer Science + Business Media
- Krugman, Paul. (2003) *The Great Unraveling: Losing Our Way in the New Century*. W. W. Norton & Company,
- Krugman, Paul. (2020) *End of Illusion: How to Recognize Economic Illusions and Build a Fairer World*. Penguin Random House
- Leblang, D., & Mukherjee, B. (2004). Presidential Elections and the Stock Market: Comparing Markov-Switching and Fractionally Integrated GARCH Models of Volatility. *Political Analysis*, 12(3), pp. 296-322. doi:10.1093/pan/mp020. Disponível em: <https://journals.sagepub.com/doi/full/10.1177/21582440231181352>
- Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business*, 36(4), 394-419. Disponível em: www.jstor.org/stable/2350970
- Marques, M. I. (2017). *Aplicação dos Modelos GARCH, EGARCH e TARARCH no DAX-30*. Instituto Politécnico de Lisboa. Fonte: hdl.handle.net/10400.21/7211
- Nouriel, R & Mihm, S. (2010) *Crisis Economics: A Crash Course in the Future of Finance*. Penguin Press
- Quivy, R. & Campenhoudt, L. V. (1998). *Manual de Investigação em Ciências Sociais*. Lisboa: Gradiva. Disponível em:

<https://tecnologiamidiainteracao.files.wordpress.com/2018/09/quivy-manual-investigacao-novo.pdf>

Raghuram, R. (2010) *Fault Lines: How Hidden Fractures Still Threaten the World Economy*. Princeton University Press,

Reinhart, C M. & Rogoff. K. S. (2012) *This Time Is Different: Eight Centuries of Financial Folly*. Princeton University Press

ROUAH, Fabrice Douglas, VAINBERG, Gregory – Option pricing models and volatility using Excel-VBA. New Jersey: *John Wiley & Sons, Inc.*, 2007. 441 pp. ISBN 978-0-471- 79464-6.

Samsudin, N. I., Mohamad, A., & Sifat, I. M. (2021). Implied Volatility of Structured Warrants: Emerging Market Evidence. *The Quarterly Review of Economics and Finance*, 80, pp. 464-479. doi: 10.1016/j.qref.2021.03.016

Shiller, Robert J. (2005), *Irrational Exuberance*, *Princeton University*, Princeton, New Jersey.

Soto, H. (2000), *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*. Basic Books

Stiglitz, Joseph E. (2003), *Globalization and its Discontents*. New York/London, WW Norton Company

Thaler, R. H. & Sustein R. C. (2008), *Nudge: Improving Decisions About Health, Wealth, and Happiness*, Penguin Books

Tsay, R. S. (2005). *Analysis of financial time series* (Vol. 543). John Wiley & Sons.

Verbeek, M. (2017). *A Guide to Modern Econometrics* (5th ed.). John Wiley & Sons

Vlassopoulos, T., & Tsileponis, N. (2023). The Economic Consequences of the Russia-Ukraine Conflict. *Journal of Economic Perspectives*, 37(2), 127-148.

Walther, T., Klein, T., & Bouri, E. (2019). Exogenous Drivers of Bitcoin and Cryptocurrency Volatility – A Mixed Data Sampling Approach to Forecasting. *Journal of International Financial Markets, Institutions and Money*, 101133. <https://doi.org/10.1016/j.intfin.2019.101133>

Wooldridge, J. (2015). *Introductory Econometrics: A Modern Approach* (6th ed.). Cengage.

Yanagihara, H., Kamo, K., Imori, S., & Satoh, K. (2012). Bias-corrected AIC for selecting variables in multinomial logistic regression models. *Linear Algebra and Its Applications*, 436(11), 4329–4341. <https://doi.org/10.1016/J.LAA.2012.01.018>

Zhang, Z. (2016). Variable selection with stepwise and best subset approaches. *Annals of Translational Medicine*, 4(7), 1–6. <https://doi.org/10.21037/atm.2016.03.35>