Universidade Federal do ABC
Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas
Trabalho de Graduação em Engenharia de Informação

Predição de volatilidade utilizando a transformada wavelet e os modelos ARIMA e GARCH

Randre Ferreira Cardoso

Santo André - SP
2024

Randre Ferreira Cardoso

Predição de volatilidade utilizando a transformada wavelet e os modelos ARIMA e GARCH

Trabalho de Graduação apresentado ao Curso de Graduação em Engenharia da Informação da Universidade Federal do ABC como requisito para obtenção do título de Engenheiro de Informação.

Orientador: Prof. Dr. Marco Cazarotto

Universidade Federal do ABC

Santo André - SP

2024

Resumo

O presente trabalho propõe uma execução híbrida para a previsão da volatilidade financeira com o emprego conjunto dos modelos ARIMA, GARCH e da transformada wavelet. A série temporal das ações da Petrobras foi decomposta em componentes de baixa e alta frequência, sendo utilizada a wavelet de Daubechies, e cada componente foi modelada separadamente. Também foi analisado o emprego do ARIMA e do GARCH separadamente. A abordagem híbrida apresentou um melhor desempenho na previsão, superando os modelos empregados isoladamente (ARIMA e GARCH). A aplicação prática destes modelos ao mercado financeiro brasileiro mostra o seu potencial de previsão da volatilidade em séries temporais financeiras.

Palavras-chave: séries temporais, volatilidade financeira, previsão, ARIMA, GARCH, wavelet.

SUMÁRIO

1. Introdução

A análise de séries temporais exerce um papel de destaque em diversas áreas da engenharia, especialmente no domínio dos sistemas de controle, telecomunicações e processamento de sinais (Gençay; Selçuk; Whitcher, 2002). Para além disso, a previsão da volatilidade financeira é um tema de elevado interesse tanto na academia como no mercado financeiro (Engle, 1982). De fato, a volatilidade, que expressa a variabilidade dos retornos de um ativo financeiro, é um dos principais componentes na avaliação de risco e na construção de estratégias de investimento (Engle, 1982). Nesse sentido, a precisão na previsão da volatilidade se torna essencial para a tomada de decisões inteligentes e eficazes (Rubio; Palacio; Mejía, 2023).

Neste trabalho, propõe-se uma abordagem híbrida para a previsão da volatilidade financeira, combinando a decomposição via transformada wavelet com os modelos *AutoRegressive Integrated Moving Average* (ARIMA) e *Generalized AutoRegressive Conditional Heteroskedasticity* (GARCH).

A decomposição wavelet é uma ferramenta que pode ser utilizada para a análise de séries temporais, na medida em que permite a separação dos dados em vários níveis de frequência, o que pode facilitar a identificação de padrões subjacentes (Alizadeh; Aissa; Nouira, 2018). Nesse trabalho, foi utilizada a decomposição wavelet com a wavelet Daubechies para decompor a série temporal de volatilidade em componentes de alta e baixa frequência.

Após a decomposição, foram aplicados os modelos ARIMA para capturar a estrutura temporal de cada componente. Os modelos ARIMA são modelos amplamente usados para modelagem de processos de séries temporais, devido à sua capacidade de representar processos autorregressivos e de médias móveis (Box; Jenkins, 1976). Todavia, estes modelos não costumam capturar adequadamente a heterocedasticidade frequentemente verificada em dados financeiros (Rubio; Palacio; Mejía, 2023).

Ademais, fez-se uso do modelo GARCH, este último especialmente adequado para modelar a volatilidade condicional, o que permite capturar as mudanças dinâmicas da volatilidade ao longo do tempo (Engle, 1982). No próximo capítulo, será mostrada uma revisão da bibliografia mostrando como esses modelos funcionam do ponto de vista da modelagem matemática e sua formulação teórica para a aplicação em séries temporais.

A metodologia proposta, descrita no capítulo 3, é validada com uma série temporal da ação da Petrobras. Através da predição da volatilidade, foi avaliada a performance dos modelos combinados, comparando as previsões com os dados reais. As métricas de avaliação utilizadas são o erro percentual médio absoluto (MAPE) e o coeficiente de determinação R^2.

2. Revisão Bibliográfica

Neste capítulo, discorreremos sobre uma parte da bibliografia existente em relação aos principais tópicos abordados nesse trabalho, que consiste no entendimento da volatilidade nos mercados, e a fundamentação teórica dos modelos ARIMA, GARCH e a transformada wavelet.

Como uma alternativa para usar separadamente o modelo ARIMA ou GARCH, o estudo *"Forecasting volatility by using wavelet transform, ARIMA and GARCH models"* (Rubio; Palacio; Mejía, 2023), apresenta uma análise da utilização conjunta da transformada wavelet, ARIMA e GARCH nas previsões de volatilidade financeira. Este estudo mostra-se particularmente interessante nas operações em mercados financeiros nos quais a volatilidade é um elemento crucial na avaliação do risco e nas decisões de investimento (Alizadeh; Aissa; Nouira, 2018). Os autores da pesquisa utilizam a transformada wavelet para decompor as séries temporais financeiras em suas componentes de diferentes escalas de tempo e, dessa forma, possibilitar análise de tendências de longo prazo em relação a movimentações de curto prazo. Logo após a decomposição, os autores aplicam os modelos ARIMA e GARCH em cada componente de forma separada, permitindo capturar a dinâmica de volatilidade da série prevista *versus* a série real. A abordagem híbrida proposta pelos autores combina os benefícios da transformada wavelet em tratar dados não estacionários e as capacidades dos modelos ARIMA e GARCH de capturar estruturas temporais lineares e padrões da volatilidade agrupada. O método resulta em previsões mais precisas de volatilidade, especialmente em mercados com alta incerteza e com frequentes mudanças de regime. O estudo busca evidenciar a efetividade desta combinação de técnicas em oposição a modelos tradicionais que não consideram a análise multiescala, sugerindo que a transformada wavelet integrada ao ARIMA e ao

GARCH é uma poderosa ferramenta para prever volatilidade em ambientes financeiros complexos. Os resultados indicaram que o modelo híbrido é melhor na detecção de padrões da volatilidade em diferentes escalas temporais, possivelmente isso se torna especialmente atrativo para gestores de risco e para investidores que buscam melhores previsões em mercados voláteis (Zhang; Zhu; Wang, 2019).

2.1 Volatilidade e Séries Temporais

A volatilidade consiste em uma medida estatística da dispersão dos retornos de um ativo financeiro, comumente utilizada como uma sinalização do risco atribuído a esse próprio ativo. Dentro dos mercados financeiros, a previsão da volatilidade é relevante para a tomada de decisões de investimento, bem como para gerenciamento de risco e precificação de derivativos (Hull, 2015). A previsão de séries temporais, que consiste na modelagem de dados que exibem uma dinâmica ao longo do tempo, é bastante utilizada para estimar a volatilidade futura de ativos financeiros (Amiri; Samadi; Mousavi, 2016).

A volatilidade de uma série temporal financeira é, com frequência, modelada como a raiz quadrada da variância dos retornos durante um período *t*. A previsão precisa da volatilidade é de grande importância para a gestão de risco, bem como para a formulação de estratégias de investimento (Poon; Granger, 2003).

A volatilidade é considerada um fator de risco crítico, pois períodos de alta volatilidade podem indicar maior incerteza nos mercados, o que afeta diretamente as decisões de investidores, gestores de portfólio e formuladores de políticas econômicas. No contexto da modelagem e previsão de séries temporais, a captura precisa da volatilidade é essencial para entender a dinâmica dos mercados financeiros e prever seus movimentos futuros (Percival; Walden, 2000).

2.2 Modelos ARIMA para Previsão de Séries Temporais

O modelo *AutoRegressive Integrated Moving Average* (ARIMA) é considerado uma das metodologias mais utilizadas para a modelagem de séries temporais (Rubio; Palacio; Mejía, 2023). O ARIMA integra componentes autorregressivos (AR), de médias móveis (MA) e de diferenciação (I) na captura da estrutura temporal dos dados

(Box; Jenkins, 1976). Por causa de sua flexibilidade e capacidade de modelar dependências lineares, o ARIMA é amplamente utilizado para previsão de séries temporais financeiras (Zhang; Zhu; Wang, 2019).

O modelo ARIMA (p,d,q) é utilizado para modelar séries temporais que apresentam dependências lineares e padrões sazonais ou de tendência (Rubio; Palacio; Mejía, 2023). O modelo ARIMA é composto pelos seguintes componentes:

- **AR(p)**: parte autorregressiva de ordem p.
- **I(d)**: número de diferenciações aplicadas para tornar a série estacionária.
- **MA(q)**: parte de médias móveis de ordem q.

Matematicamente, o modelo ARIMA é descrito conforme a Equação 1.

$$\Phi(L)(1-L)^d y_t = \Theta(L)\epsilon_t \tag{1}$$

onde:

- $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ é o polinômio autorregressivo.
- $\Theta(L) = 1 + \theta_1 L + \theta_2 L_2 + \cdots + \theta_q L^q$ é o polinômio de médias móveis.
- L é o operador defasagem (lag operator).
- ϵ_t é um termo de erro com média esperada zero, variância constante e os erros não possuem correlação ao longo do tempo, o que significa que não há covariância entre ϵ_t e ϵ_{t-k}. O termo de erro é geralmente assumido como um processo de ruído branco (Box; Jenkins, 1976).

Para exemplificar a implementação do modelo ARIMA foi desenvolvido o cálculo com base na Equação 1, utilizando os coeficientes *p* igual a 3, *d* igual 1 e *q* igual a 2. Sendo assim a construção do modelo poderia seguir os seguintes passos:

1. Escrever a componente AR conforme descrito na Equação 1. De forma que temos três termos autorregressivos.

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 \tag{2}$$

2. Aplicar a diferenciação que nesse exemplo será uma vez.

$$(1-L)y_t = y_t - y_{t-1} \tag{3}$$

3. Escrever a componente MA conforme descrito na Equação 1. Sendo assim, temos dois termos de média móvel.

$$\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 \quad (4)$$

4. Com base nas Equações 2, 3 e 4 expandir a Equação 1.

$$y_t - y_{t-1} - \phi_1(y_{t-1} - y_{t-2}) - \phi_2(y_{t-2} - y_{t-3}) - \phi_3(y_{t-3} - y_{t-4}) = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} \quad (5)$$

5. Simplificando a Equação 5.

$$y_t = (1 + \phi_1)y_{t-1} + (\phi_2 - \phi_1)y_{t-2} + (\phi_3 - \phi_2)y_{t-3} - \phi_3 y_{t-4} + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} \quad (6)$$

Com base na Equação 6, é possível observar que y_t depende dos valores passados $y_{t-1}, y_{t-2}, y_{t-3}$ e y_{t-4}, ajustados pelos coeficientes ϕ_1, ϕ_2 e ϕ_3. Sendo que Equação 6 também inclui uma contribuição dos termos de erro passados ϵ_{t-1} e ϵ_{t-2} ajustados pelos coeficientes θ_1 e θ_2 (Brockwell; Davis, 2016).

Apesar de conseguir capturar padrões temporais complexos, como sazonalidades e tendências, o modelo ARIMA apresenta limitações ao lidar com dados financeiros que exibem volatilidade heterocedástica, ou seja, variações de volatilidade ao longo do tempo (Alizadeh; Aissa; Nouira, 2018). Essa limitação motivou o desenvolvimento de modelos que consideram explicitamente a volatilidade variável, como o GARCH (Engle, 1982).

Ainda no ramo da economia, o modelo ARIMA também é considerado um dos mais adequados para a realização de previsões de indicadores econômicos, tais como o PIB, as taxas de desemprego, a inflação, e as taxas de juros. Por exemplo, os analistas podem utilizar o ARIMA para prever o crescimento econômico baseando-se em dados passados, ao mesmo tempo em que ajustam o modelo para a sazonalidade e para a tendência (Engle, 1982). Adicionalmente, na epidemiologia, o ARIMA tem sido utilizado para prever a incidência de doenças infecciosas, como a gripe ou o COVID-19, com base em dados históricos de casos, o que ajuda a planejar e alocar recursos de saúde de forma mais eficaz (Benvenuto; Giovanetti; Vassalo; Angeletti, 2020).

2.3 Modelos GARCH para Modelagem de Volatilidade

O modelo *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH) (Bollerslev, 1986), estende o modelo ARCH (Engle, 1982) para capturar a heterocedasticidade condicional presente em séries temporais financeiras. O modelo GARCH é capaz de modelar a volatilidade como uma função das variações passadas, oferecendo previsões mais precisas em ambientes onde a volatilidade não é constante ao longo do tempo. O modelo GARCH (p,q) generaliza o modelo ARCH *(AutoRegressive Conditional Heteroskedasticity)*, permitindo que a variância condicional dependa tanto em termos de erro passados quanto de volatilidades anteriores (Engle, 1982). A especificação do modelo GARCH (p,q) é dada conforme a Equação 7 e a Equação 8.

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta \sigma_{t-j}^2 , \tag{7}$$

$$\epsilon_t = \sigma_t z_t , \tag{8}$$

onde:

- σ_t^2 é a variância condicional no tempo t.
- $\alpha_0 > 0$, $\alpha_i \geq 0$ e $\beta_j \geq 0$ são parâmetros do modelo, com a condição de estacionaridade $\sum_i^q \alpha_i + \sum_j^p \beta_j < 1$.
- p: Refere-se ao número de defasagens dos quadrados dos erros passados (componentes ARCH).
- q: Refere-se ao número de defasagens das variâncias condicionais passadas (componentes GARCH).
- σ_t: é a volatilidade condicional (a raiz quadrada da variância condicional σ_t^2).
- z_t**:** é um termo de ruído branco que segue uma distribuição normal padrão, ou seja, $z_t \sim N(0,1)$.

O modelo GARCH é amplamente usado em finanças devido à sua capacidade de capturar clusters de volatilidade, que são períodos em que a volatilidade permanece elevada por vários dias (Zhang; Zhu; Wang, 2019). O modelo também é muito útil na área de precificação de opções, onde é necessário prever com precisão a volatilidade das ações. Os *traders* e analistas utilizam o GARCH para ajustar preços de opções e estratégias de *hedge*, tendo em conta grandes mudanças na volatilidade do ativo subjacente (Hull, 2015). Além disso, as instituições financeiras utilizam o GARCH para estimar e prever o valor em risco (VaR), uma importante métrica que reflete o risco de perda em uma carteira de investimentos, ajudando a estimar o risco de grandes perdas devido às flutuações do mercado (Jorion, 2007).

2.4 Transformada Wavelet na Análise de Séries Temporais Financeiras

A transformada wavelet é uma ferramenta matemática para a análise multirresolução de séries temporais, permitindo a decomposição dos dados em componentes de diferentes frequências (Rubio; Palacio; Mejía, 2023). Essa propriedade é particularmente útil para séries temporais financeiras, onde os sinais podem conter informações em várias escalas de tempo. As wavelets têm sido aplicadas em diversas áreas, incluindo processamento de sinais, compressão de imagens e análise financeira (Percival; Walden, 2000).

No contexto da predição de volatilidade, a transformada wavelet pode ser utilizada para decompor a série temporal original em diferentes níveis de detalhe, capturando tanto tendências de longo prazo quanto oscilações de curto prazo (Alizadeh; Aissa; Nouira, 2018).

A transformada wavelet é amplamente utilizada para a análise de séries temporais financeiras, por conseguirem decompor um sinal em diferentes escalas de tempo; e sendo muito úteis em detectar padrões e rompimentos estruturais nos dados financeiros (Percival; Walden, 2000). No trabalho *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics* (Gençay; Selçuk; Whitcher, 2002), é apresentada uma visão geral dos métodos utilizando wavelets e de outras técnicas de filtragem em finanças e em economia. Em seu trabalho, eles procuraram mostrar como tais ferramentas podem ser utilizadas para capturar as tendências de longo prazo e as flutuações de curto prazo das séries temporais financeiras. Os autores demonstram que a aplicação da transformada wavelet é superior às técnicas convencionais (como a transformada de Fourier), especialmente no tratamento das séries temporais não estacionárias e nas altamente voláteis.

No trabalho *Wavelet Methods for Time Series Analysis* (Percival; Walden, 2000), os autores apresentaram uma visão detalhada sobre as técnicas utilizadas com a wavelet aplicados à análise de séries temporais, abordando tanto os fundamentos matemáticos quanto as aplicações práticas dessa metodologia. Os autores discutiram como a transformada wavelet oferece uma abordagem para lidar com a não estacionaridade, um problema comum em séries temporais financeiras e em outras áreas. Eles mostraram que, ao contrário da transformada de Fourier, que analisa sinais apenas no domínio da frequência, a transformada wavelet permite decompor um sinal em diferentes escalas de tempo, mantendo a informação temporal. Essa capacidade em diferentes escalas é particularmente útil para identificar padrões de longo prazo, como tendências, e flutuações de curto prazo, como picos de volatilidade (Alizadeh; Aissa; Nouira, 2018). Além disso, os autores destacam a aplicabilidade dos métodos wavelet na detecção de rupturas estruturais e na análise de cointegração entre diferentes séries temporais, tornando esta abordagem extremamente relevante para a análise de dados financeiros e econômicos, onde as mudanças de regime e a variabilidade são frequentes (Rubio; Palacio; Mejía, 2023).

A transformada wavelet discreta (DWT) de uma série temporal *x(t)* é dada conforme a Equação 9.

$$W(j,k) = \sum_t x(t)\psi_{j,k}(t) \ , \tag{9}$$

onde:

- $W(j,k)$ são os coeficientes de wavelet.
- $\psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi(2^{-j}t - k)$ são as funções wavelet geradas por translação e dilatação da wavelet-mãe $\psi(t)$.
- $x(t)$ é a série temporal.
- j e k representam os níveis de escala e translação, respectivamente.

A DWT permite decompor a série temporal em diferentes escalas de detalhe (níveis de resolução), capturando tanto as tendências de longo prazo quanto as flutuações de curto prazo. Essa decomposição é essencial para analisar os sinais financeiros que contêm informações em várias frequências (Amiri; Samadi; Mousavi, 2016).

2.5 Modelos Híbridos: ARIMA-GARCH-Wavelet

O estudo *Wavelet Decomposition and Hybrid Forecasting Model Based on ARIMA and GARCH for stock prices* (Zhang; Zhu; Wang, 2019), traz uma proposta para previsão de preços de ações baseada em um modelo híbrido que junta a transformada wavelet com os modelos ARIMA e GARCH. A motivação do estudo é propor soluções para as limitações dos modelos tradicionais na representação das complexas dinâmicas das séries temporais financeiras que, muitas vezes, apresentam comportamento não estacionário, mudança de variação de volatilidade e diferentes características de frequência no tempo. Os resultados obtidos no decorrer do estudo demonstram que o modelo híbrido apresentado tem um desempenho superior, do que cada um dos modelos individuais (ARIMA e GARCH) aplicados de maneira direta sobre a série temporal de preços, sem decomposição wavelet. O modelo híbrido propõe previsões mais acuradas e mais robustas, especialmente em cenários de extrema volatilidade do mercado, onde a habilidade de captar mudanças rápidas na volatilidade e na tendência de preços é essencial (Zhang; Zhu; Wang, 2019). O estudo conclui que a decomposição wavelet não melhora apenas a habilidade dos modelos preditivos, dividindo as componentes de diferentes frequências, como também tem o potencial de atuar como uma ferramenta para tratar a não estacionalidade inerente das séries temporais financeiras (Rubio; Palacio; Mejía, 2023). A combinação de Wavelet, ARIMA e GARCH dá origem a uma estrutura preditiva que abrange mais e que é capaz de lidar com diferentes regimes de mercado e fornecer previsões mais robustas para os preços das ações (Zhang; Zhu; Wang, 2019).

O *A hybrid wavelet-ARIMA-GARCH model for stock market volatility prediction: Case study of the Canadian stock market* (Alizadeh; Aissa; Nouira, 2018), investiga a eficácia de um modelo híbrido que combina a transformada Wavelet, ARIMA e GARCH para prever a volatilidade do mercado de ações canadense. Este trabalho é motivado pela necessidade de melhorar a precisão das previsões de volatilidade, um aspecto crucial na gestão de risco e na formulação de estratégias de investimento em mercados financeiros. O estudo apresenta evidências empíricas de que o modelo híbrido Wavelet-ARIMA-GARCH supera os modelos individuais (ARIMA ou GARCH aplicados isoladamente) na previsão da volatilidade do mercado de ações canadense.

Os autores demonstram que a decomposição wavelet, ao separar os diferentes padrões de frequência, permite uma modelagem mais precisa dos componentes de volatilidade, resultando em previsões mais robustas (Alizadeh; Aissa; Nouira, 2018).

O estudo intitulado *A hybrid model of ARIMA-GARCH and wavelet transform methods for predicting natural gas price volatility* (Amiri; Samadi; Mousavi, 2016), explora a aplicação de um modelo híbrido que combina a transformada wavelet, ARIMA e GARCH para prever a volatilidade dos preços do gás natural. Dado que o mercado de energia, especialmente o de gás natural, é altamente volátil e sujeito a influências de uma variedade de fatores econômicos, políticos e ambientais, a previsão precisa da volatilidade é crucial para a gestão de risco e a formulação de estratégias comerciais eficazes (Amiri; Samadi; Mousavi, 2016). O estudo também revela que a decomposição wavelet permite que o modelo ARIMA e GARCH se concentrem nos componentes mais relevantes da série temporal, melhorando assim a precisão e a robustez das previsões. Esta abordagem é particularmente útil em mercados como o de gás natural, onde a volatilidade é influenciada por uma ampla gama de fatores exógenos e endógenos, incluindo condições climáticas, políticas energéticas e variações na demanda e oferta (Zhang; Zhu; Wang, 2019).

Modelos híbridos combinam as forças de diferentes abordagens para melhorar a precisão das previsões (Rubio; Palacio; Mejía, 2023). No contexto deste trabalho, a série temporal de retornos é primeiro decomposta em componentes de baixa e alta frequência usando a transformada wavelet. Em seguida, os componentes são modelados separadamente utilizando ARIMA para capturar a média dos dados e GARCH para modelar a volatilidade.

A formulação matemática deste modelo híbrido pode ser expressa como:

1. Decomposição da série $x(t)$ em componentes de detalhes (D1 e D2) e aproximação (A) via DWT.
2. Aplicação de ARIMA nos componentes de aproximação e GARCH nos componentes de detalhe.

Essa abordagem permite capturar diferentes dinâmicas de volatilidade presentes nas escalas de tempo subjacentes, resultando em uma previsão mais robusta (Rubio; Palacio; Mejía, 2023).

A transformada wavelet discreta (DWT) de uma série temporal $x(t)$ envolve a aplicação iterativa de dois conjuntos de filtros: um filtro passa-baixa (h) e um filtro passa-alta (g).

- Coeficientes de Aproximação: Representam os componentes de baixa frequência do sinal.
- Coeficientes de Detalhe: Representam os componentes de alta frequência do sinal.

Para um dado nível j, a DWT pode ser expressa para aproximação e detalhe, descritas na Equação 10 e na Equação 11, respectivamente.

$$A_j(n) = \sum_k h(k) \cdot A_{j-1}(n-k), \tag{10}$$

$$D_j(n) = \sum_k g(k) \cdot A_{j-1}(n-k), \tag{11}$$

onde:

- $A_{j-1}(n)$ representa os coeficientes de aproximação do nível anterior *(j−1)*.
- $h(k)$ é o filtro passa-baixa aplicado para obter os coeficientes de aproximação.
- $g(k)$ é o filtro passa-alta aplicado para obter os coeficientes de detalhe.
- A soma é realizada sobre todos os índices k que correspondem aos coeficientes do filtro.

Os filtros $h(k)$ e $g(k)$ são projetados para satisfazer certas condições de ortogonalidade e compactação de energia. Os filtros $h(k)$ e $g(k)$ são derivados da wavelet mãe $\psi(t)$ e da função de escala $\phi(t)$. Essas funções satisfazem as relações de reconstrução, que garantem que o sinal original possa ser reconstruído a partir dos coeficientes de aproximação e detalhe (Zhang; Zhu; Wang, 2019). A relação para a função de escala e a relação para a wavelet mãe podem ser descritas conforme a Equação 12 e a Equação 13, respectivamente.

$$\phi(t) = \sqrt{2}\sum_k h[k]\phi(2t-k) \tag{12}$$

$$\psi(t) = \sqrt{2}\sum_k g[k]\phi(2t - k) \quad (13)$$

Onde:

- $\phi(t)$ é a função de escala, que captura as componentes de baixa frequência do sinal.
- $\psi(t)$ é a wavelet mãe, que captura as componentes de alta frequência do sinal.
- $h[k]$ é o filtro passa-baixa.
- $g[k]$ é o filtro passa-alta.
- k é o índice de somatório que percorre os coeficientes do filtro.
- $\phi(2t - k)$ representa a função de escala dilatada e transladada.

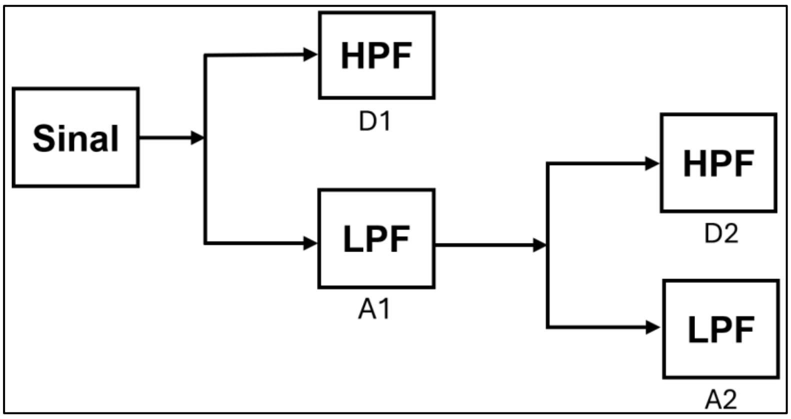

Figura 1: Transformada wavelet discreta de dois níveis. A1 e A2 são coeficientes de aproximação, D1 e D2 são coeficientes de detalhe, HPF é o filtro passa alta (em inglês: high pass filter) e LPF é o filtro passa baixa (em inglês: low pass filter).

A função de escala $\phi(t)$ é construída a partir de cópias escalonadas e deslocadas de si mesma, usando o filtro passa-baixa $h[k]$. A wavelet mãe $\psi(t)$ é construída a partir da função de escala $\phi(t)$, usando o filtro passa-alta $g[k]$ (Percival; Walden, 2000). No próximo capítulo será possível observar como foram aplicados esses conceitos nesse trabalho.

3. Metodologia

Neste capítulo será abordado como o trabalho foi montado, tendo como base o passo a passo da implementação, inicialmente, dos modelos individualmente e depois os utilizando de maneira híbrida.

3.1 Descrição dos dados utilizados

Neste estudo, utilizamos dados históricos de preços das ações da Petrobras (PETR4) como base para a análise e modelagem da volatilidade. Os dados foram obtidos através da plataforma Yahoo Finance, cobrindo o período de 01 de março de 1997 até 28 de junho de 2024. A frequência dos dados é diária, e os valores usados foram o do fechamento de cada dia. Para a análise, focamos nos retornos logarítmicos diários calculados conforme Equação 13 (Poon; Granger, 2003).

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad , \qquad (13)$$

onde P_t é o preço de fechamento da ação no dia t e P_{t-1} é o preço de fechamento do dia anterior.

3.2 Implementação dos modelos

A metodologia adotada envolve a aplicação de três abordagens principais para a modelagem e previsão da volatilidade: ARIMA, GARCH e a combinação de ambos com a transformada wavelet. A seguir, é descrito cada abordagem:

3.2.1 Modelo ARIMA

O modelo ARIMA foi aplicado para capturar a dinâmica linear dos retornos das ações. Utilizando o procedimento Box-Jenkins, os parâmetros p, d e q foram selecionados com base nos critérios de informação de Akaike (AIC) (Akaike, 1974) e Schwartz (BIC) (Schwartz, 1978).

3.2.2 Modelo GARCH

O processo de seleção dos parâmetros p e q do modelo GARCH seguiu critérios semelhantes aos do ARIMA, utilizando o AIC e a inspeção dos gráficos de autocorrelação dos quadrados dos resíduos.

3.2.3 Combinação dos modelos com a transformada wavelet

A abordagem híbrida envolve a combinação dos modelos ARIMA e GARCH com a transformada wavelet, permitindo a decomposição da série temporal em componentes de diferentes escalas de tempo. O processo foi realizado em três etapas:

1. Decomposição wavelet: Utilizando a transformada wavelet discreta (DWT) com uma wavelet-mãe adequada, como a Daubechies (db2), a série temporal foi decomposta em níveis de detalhe $D_1, D_2, \ldots, D_n$ e um nível de aproximação A.
2. Modelagem dos Componentes: Cada componente de detalhe foi modelado separadamente com o modelo GARCH para capturar a volatilidade em escalas de alta frequência, enquanto o componente de aproximação foi modelado com ARIMA.
3. Para fins de teste também, a componente de detalhe também foi modelada separadamente com o modelo ARIMA para capturar a volatilidade em escalas de alta frequência e o componente de aproximação foi modelado com GARCH.

4. Recomposição e Previsão: Após a modelagem dos componentes, as previsões individuais foram recombinadas para gerar a previsão final da volatilidade da série temporal original.

3.3 Validação dos Modelos

A validação dos modelos foi realizada em duas etapas:

1. **Validação Interna**: foi utilizada uma abordagem de janela móvel de volatilidade para avaliar a robustez dos modelos ao longo do tempo. Os dados foram divididos em conjuntos de treino e teste, e as previsões foram geradas iterativamente. O conjunto de treino foi o período do início da série que foi em de 01 de março 1997 até 30 de maio de 2024. O conjunto de teste foram os 21 dias úteis seguintes, se encerrando em 28 de junho de 2024.

2. **Comparação de Desempenho**: O desempenho dos modelos foi avaliado utilizando métricas como o erro percentual médio absoluto (MAPE) e o coeficiente de determinação R^2.

3.4 Implementação Computacional

Toda a implementação foi realizada em Python, utilizando as bibliotecas pandas, statsmodels, arch, e pywt para o processamento de dados, ajuste dos modelos e aplicação da transformada wavelet. Scripts separados foram desenvolvidos para cada etapa da metodologia: pré-processamento, modelagem ARIMA, modelagem GARCH, decomposição wavelet e recomposição das previsões. O esquema ilustrado na Figura 2 mostra o passo a passo que foi seguido para implementar a análise. Os resultados da implementação proposta estão descritos no capítulo seguinte, abordando separadamente cada modelo e depois discorrendo sobre a sua implementação híbrida.

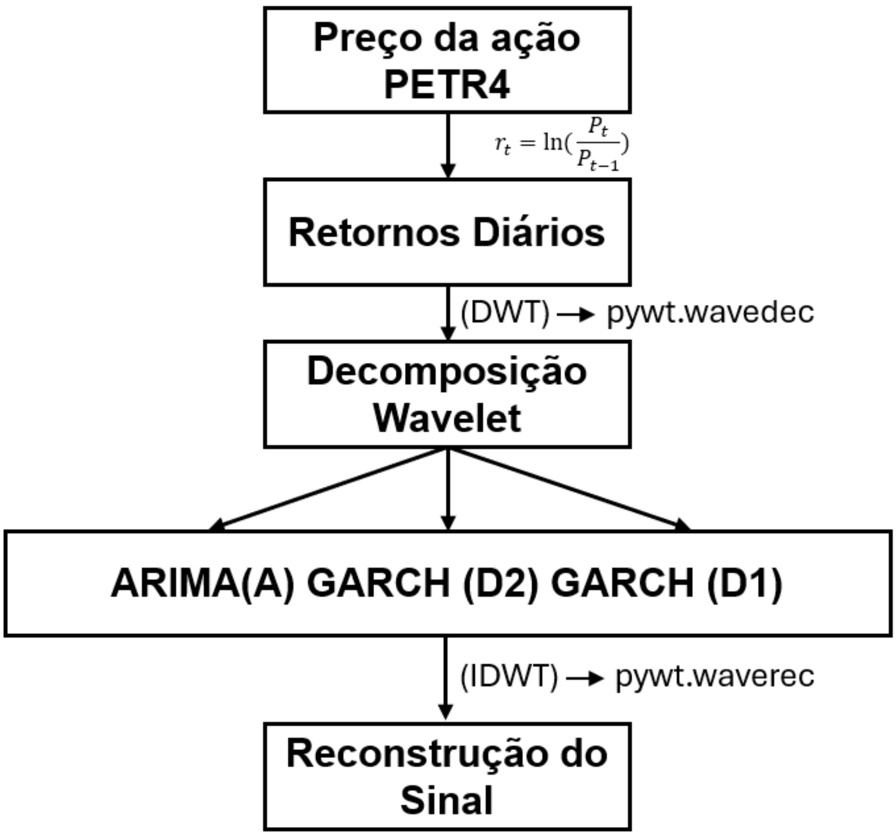

Figura 2: Fluxograma para representação da implementação dos modelos realizada em python.

4. Resultados e Discussões

Nesse capítulo iremos discorrer sobre os resultados encontrados e será possível observar o desempenho de cada modelo e qual que possuiu melhor aderência com os dados reais.

4.1 Desempenho dos Modelos ARIMA, GARCH e Híbrido

Neste capítulo, apresentamos e discutimos os resultados obtidos com a aplicação dos modelos ARIMA, GARCH e a combinação híbrida desses modelos com a transformada wavelet para a previsão de volatilidade das ações da Petrobras (PETR4).

4.1.1 Resultados do Modelo ARIMA

O modelo ARIMA foi ajustado para capturar a dinâmica linear dos retornos das ações da Petrobras.

O modelo ARIMA pressupõe que a série seja estacionária, o que significa que suas propriedades estatísticas, como a média e a variância, não mudam ao longo do tempo. Para essa análise é esperado que a série não seja estacionária, pois as séries de preços de ações costumam apresentar tendências de longo prazo, seja de baixa ou de alta, devido a fatores econômicos, políticos, e mudanças na percepção do valor das empresas (Alizadeh; Aissa; Nouira, 2018). Além disso, a variância dos retornos das ações não é constante ao longo do tempo. Durante períodos de crise financeira, por exemplo, a volatilidade tende a aumentar significativamente (Zhang; Zhu; Wang, 2019).

O teste de verificação de estacionaridade foi o teste ADF (Augmented Dickey-Fuller): se o valor p do teste for menor que um determinado nível de significância (por exemplo, 0,05), rejeita-se a hipótese nula de não estacionaridade.

Tabela 1: Resultado do Teste ADF para Preços de Fechamento

Resultados	
Estatística ADF:	3.111207
p-value:	1.000000
Nº de Observações	6114.000000
Valor Crítico (1%)	-3.431420
Valor Crítico (5%)	-2.862013
Valor Crítico (10%)	-2.567022

A Tabela 1 apresenta a estatística ADF, que é o valor do teste calculado. Este valor é comparado aos valores críticos para determinar se podemos rejeitar a hipótese nula de que a série tem uma raiz unitária, ou seja, de que a série não é estacionária. Os valores críticos indicam os limites inferiores da estatística ADF que permitiria a rejeição da hipótese nula a níveis de confiança de 1%, 5%, e 10%. Para que a série seja considerada estacionária, a estatística ADF deveria ser menor (mais negativa) do que esses valores críticos. Sendo assim, a estatística ADF é 3.111207, que é maior que todos os valores críticos. Portanto, não podemos rejeitar a hipótese nula. O valor p indica a probabilidade de obter a estatística do teste sob a hipótese nula. Um valor p alto (geralmente maior que 0,05) sugere que não podemos rejeitar a hipótese nula de que a série não é estacionária (Jorion, 2007). Neste caso, o valor p é 1.000000, o que significa que a evidência é extremamente forte contra a estacionaridade. Em outras palavras, a série é não estacionária.

Após a identificação dos melhores parâmetros utilizando os critérios de informação AIC e BIC, a especificação final selecionada foi ARIMA (2, 0, 3). O modelo demonstrou ser eficiente na captura das tendências de longo prazo e das oscilações regulares na série temporal, porém, apresentou dificuldades em prever mudanças abruptas de volatilidade, um comportamento esperado devido à natureza linear do ARIMA (Box; Jenkins, 1976).

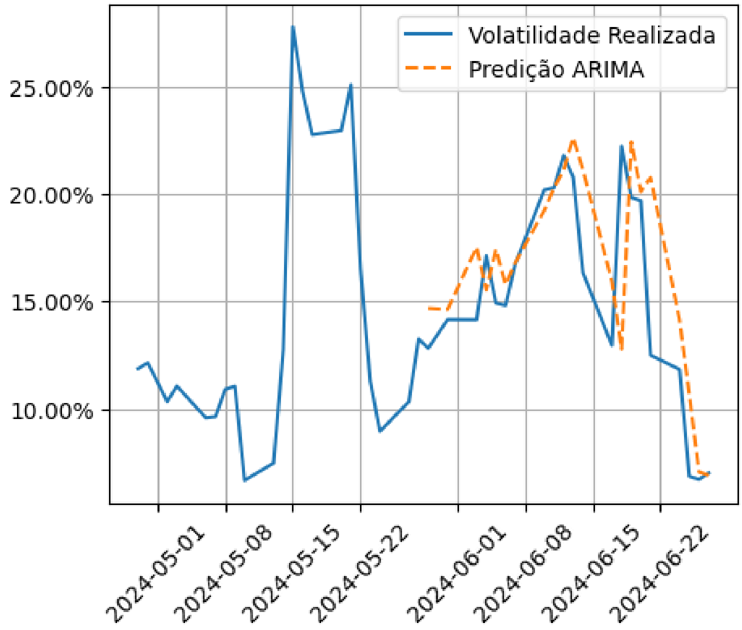

Figura 3: Real vs. ARIMA para volatilidade realizada PETR4.

4.1.2 Resultados do Modelo GARCH

O modelo GARCH é projetado para modelar a variância condicional de uma série temporal, ou seja, como a variância dos retornos em um período é condicionada pelas informações dos períodos anteriores (Engle, 1982). Para que a variância condicional seja significativa e interpretável, a série subjacente precisa ser estacionária, o que implica que suas propriedades estatísticas, como a média e a variância, não mudam ao longo do tempo (Rubio; Palacio; Mejía, 2023). Novamente, para essa análise não esperamos que a série seja estacionária, e como mostrado na tabela 1, a série de retornos é não estacionária. Para realizar a implementação também foram utilizados os critérios de informação AIC e BIC para determinar os coeficientes p e q do modelo GARCH, que foram 1 e 3 respectivamente. Os resultados podem ser observados na Figura 4.

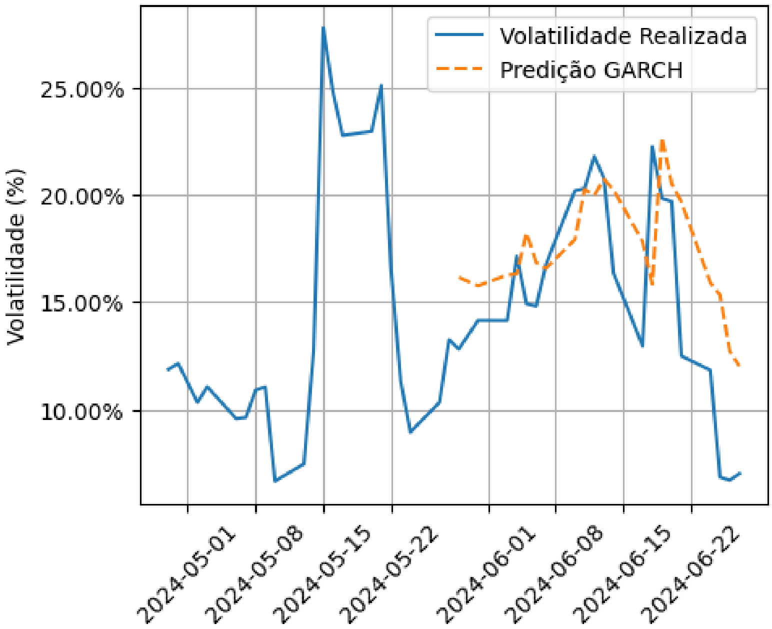

Figura 4: Real vs. GARCH para volatilidade realizada PETR4.

4.1.3 Resultados do Modelo Híbrido W-ARIMA-GARCH e W-GARCH-ARIMA

A wavelet utilizada para realizar a análise foi a Daubechies são eficazes na captura de características em diferentes escalas, o que é essencial em séries temporais que possuem estruturas em várias escalas de tempo (Rubio; Palacio; Mejía, 2023). Elas conseguem decompor o sinal em componentes que refletem tanto variações de curto prazo (capturando detalhes e possíveis ruídos) quanto tendências de longo prazo (aproximações) (Percival; Walden, 2000). Adicionalmente, a família Daubechies é ortogonal, o que significa que as wavelets e seus deslocamentos no tempo não se sobrepõem em termos de energia. Isso garante que, após a decomposição e reconstrução, não haja perda de informação. A ortogonalidade também permite a perfeita reconstrução do sinal após o processamento, o que é essencial para a remoção de ruído sem distorcer o sinal original (Gençay; Selçuk; Whitcher, 2002).

Para o modelo W-ARIMA-GARCH a decomposição wavelet permitiu separar a série temporal em componentes de baixa e alta frequência. Os valores encontrados na componente de aproximação foram modelados com ARIMA e o os componentes de detalhe com o GARCH. Ao contrário, para o modelo W-GARCH-ARIMA, os valores encontrados na componente de aproximação foram modelados com GARCH e o os componentes de detalhe com o ARIMA.

Os resultados da previsão são apresentados na Figura 5.

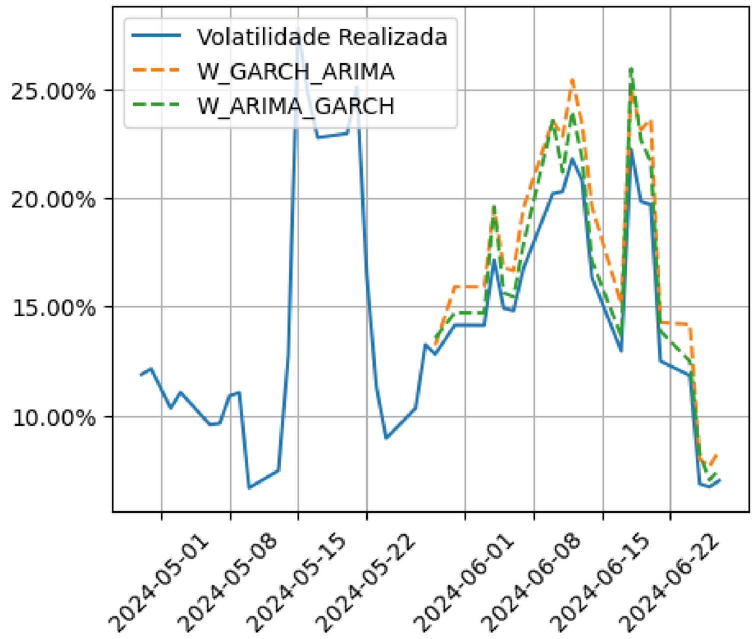

Figura 5: Real vs. W-ARIMA-GARCH e W-GARCH-ARIMA para volatilidade realizada PETR4.

4.2 Discussão dos Resultados

Os resultados indicaram que o modelo híbrido foi mais eficaz na previsão de períodos de alta volatilidade, pois a decomposição wavelet isolou as oscilações de alta frequência que o ARIMA e GARCH sozinhos não capturariam de forma tão eficiente. Conforme observado na tabela 2, os modelos híbridos tiveram o maior coeficiente de determinação (R^2) encontrado, indicando melhor ajuste do modelo a série original. Observando o erro percentual médio absoluto da tabela 2 (MAPE), o modelo W-ARIMA-GARCH teve o melhor desempenho por apresentar o menor erro dentre os modelos.

Tabela 2: Comparação das Métricas de Desempenho dos Modelos

Modelo	MAPE	R^2
ARIMA	16.94%	0.64
GARCH	28.64%	0.56
W-ARIMA-GARCH	8.57%	0.73
W-GARCH-ARIMA	14.94%	0.73

A análise comparativa dos modelos revelou que, embora o ARIMA e o GARCH sejam eficazes em suas respectivas áreas (tendência e volatilidade), a combinação híbrida oferece uma abordagem mais holística e precisa para prever a volatilidade em mercados financeiros (Zhang; Zhu; Wang, 2019). O uso da transformada wavelet foi fundamental para capturar informações em diferentes escalas temporais que os modelos tradicionais não conseguem abordar isoladamente, melhorando o desempenho preditivo.

Em termos práticos, a aplicação do modelo híbrido nas ações da Petrobras forneceu previsões de volatilidade mais ajustadas à realidade observada. Isso demonstra o potencial de uso de modelos híbridos em ambientes de mercado com volatilidade significativa e não linearidade, como é o caso do mercado de ações brasileiro (Medeiros; Veiga, 2009).

Os resultados obtidos corroboraram a literatura existente acerca da superioridade dos modelos híbridos na previsão da volatilidade de séries temporais financeiras. A combinação da modelagem estatística tradicional com técnicas de análise multirresolução (como a transformada wavelet) propicia uma análise mais completa e precisa das dinâmicas complexas em mercados financeiros (Alizadeh; Aissa; Nouira, 2018).

No entanto, é importante destacar que, apesar dos avanços obtidos com o modelo híbrido, existem desafios e limitações como a sensibilidade dos resultados à escolha da wavelet e a necessidade de ajustar parâmetros de forma iterativa. Além disso, a complexidade computacional do modelo híbrido é maior em comparação aos modelos tradicionais, o que pode limitar sua aplicabilidade em ambientes de tempo real ou em conjuntos de dados muito grandes (Rubio; Palacio; Mejía, 2023).

Por fim, os resultados sugerem que a aplicação dessa metodologia a outros ativos ou mercados pode trazer benefícios semelhantes, desde que os modelos sejam ajustados às características específicas de cada série temporal.

5. Conclusão

A integração das técnicas ARIMA, GARCH e transformada wavelet representa uma abordagem poderosa para a modelagem e previsão de volatilidade em séries temporais financeiras (Gençay; Selçuk; Whitcher, 2002). No caso das ações da Petrobras, o modelo híbrido demonstrou ser capaz de fornecer previsões mais precisas e detalhadas, capturando nuances que os modelos tradicionais isolados não conseguiram. Esses resultados ressaltam a importância de técnicas avançadas e multidisciplinares na análise financeira, abrindo caminho para futuras pesquisas e aplicações no campo.

Referências Bibliográficas

1. BOX, P.; JENKINS, M. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1976.
2. ENGLE, R. *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*. Econometrica, v. 50, n. 4, p. 987-1007, 1982.
3. BOLLERSLEV, T. *Generalized Autoregressive Conditional Heteroskedasticity*. Journal of Econometrics, v. 31, n. 3, p. 307-327, 1986.
4. GENÇAY, R.; SELÇUK, F.; WHITCHER, B. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. San Diego: Academic Press, 2002.

5. RUBIO, L.; PALACIO, A.; MEJÍA, A. *Forecasting volatility by using wavelet transform, ARIMA and GARCH models*. Eurasian Economic Review, v. 13, p. 803-830, 2023.
6. PERCIVAL, B.; WALDEN, T. *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press, 2000. 594 p.
7. ZHANG, X.; ZHU, X.; WANG, Y. *Wavelet decomposition and hybrid forecasting model based on ARIMA and GARCH for stock prices*. Procedia Computer Science, v. 162, p. 550-557, 2019.
8. ALIZADEH, S.; AÏSSA, S.; NOUIRA, A. *A hybrid wavelet-ARIMA-GARCH model for stock market volatility prediction: Case study of the Canadian stock market*. Journal of Computational and Applied Mathematics, v. 340, p. 653-671, 2018.
9. AMIRI, S.; SAMADI, S.; MOUSAVI, S. *A hybrid model of ARIMA-GARCH and wavelet transform methods for predicting natural gas price volatility*. Journal of Natural Gas Science and Engineering, v. 33, p. 291-302, 2016.
10. JORION, P. *Value at Risk: The New Benchmark for Managing Financial Risk.* McGraw-Hill. 2007
11. HULL, J. *Options, Futures, and Other Derivatives.* Pearson Education. 2015
12. POON, H.; GRANGER, J. *Forecasting Volatility in Financial Markets: A Review*. Journal of Economic Literature, 41(2), 478-539. 2003
13. BENVENUTO, D.; GIOVANETTI, M.; VASSALLO, L.; ANGELETTI, S.; CICCOZZI, M. *Application of the ARIMA model on the COVID-2019 epidemic dataset.* Data in Brief, 29, 105340. 2020
14. AKAIKE, H.**;** *A new look at the statistical model identification*. IEEE Transactions on Automatic Control, 1974.

15. SCHWARTZ, G.; *Estimating the dimension of a model.* The Annals of Statistics, (1978).

16. MEDEIROS, C.; VEIGA, A. *Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1) model.* Econometric Theory, 2009.