



**ANA CLAUDIA FESTUCCI DE HERVAL**

**IMPACTO DOS CHOQUES NA COMPONENTE DISCRETA DA  
VOLATILIDADE :  
UMA ABORDAGEM VIA ONDALETAS**

**LAVRAS – MG  
2022**

**ANA CLAUDIA FESTUCCI DE HERVAL**

**IMPACTO DOS CHOQUES NA COMPONENTE DISCRETA DA VOLATILIDADE:  
UMA ABORDAGEM VIA ONDALETAS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Profa. Dra. Thelma Sáfydi  
Orientadora

**LAVRAS – MG  
2022**

**Ficha Catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca  
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Herval, Ana Claudia Festucci de.

Impacto dos choques na componente discreta da volatilidade : uma abordagem via ondaletas / Ana Claudia Festucci de Herval. – Lavras : UFLA, 2022.

68 p. : il.

Orientadora: Profa. Dra. Thelma Sáfadi.

Tese (doutorado) - Universidade Federal de Lavras, 2022.

Bibliografia.

1. Saltos de volatilidade.
2. Decomposição de ondaletas.
3. Mercado brasileiro de ações. I. Sáfadi, Thelma. II. Título.

**ANA CLAUDIA FESTUCCI DE HERVAL**

**IMPACTO DOS CHOQUES NA COMPONENTE DISCRETA DA VOLATILIDADE:  
UMA ABORDAGEM VIA ONDALETAS  
IMPACT OF SHOCKS ON THE DISCRETE COMPONENT OF VOLATILITY: A  
WAVELET APPROACH**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 03 de Junho de 2022.

Profa. Dra. Chang Chiann	USP
Prof. Dr. Ednilson Sebastião Ávila	UFLA
Profa. Dra. Izabela Regina Cardoso de Oliveira	UFLA
Prof. Dr. Paulo Henrique Sales Guimarães	UFLA

Profa. Dra. Thelma Sáfyadi  
Orientadora

**LAVRAS – MG  
2022**

*Dedico este trabalho ao meu amado pai, José Carlos,  
que não pode me ver obtendo este título, mas sempre foi um grande incentivador  
dos meus estudos e da minha carreira como professora.*

## AGRADECIMENTOS

Ao meu companheiro, amigo e marido, Leonilson, que sempre viu em mim um potencial que eu, muitas vezes, não enxerguei. Ele foi e é um verdadeiro propulsor do meu desenvolvimento e da minha carreira como professora e pesquisadora.

Ao meu irmão, Car, que esteve presente me ajudando desde os meus estudos para ingresso no doutorado, até a conclusão deste.

À minha mãe, Ana Rita, que foi um apoio inestimável, especialmente durante os últimos anos durante o isolamento social.

Ao meu filho, Isaac, que mesmo sem conseguir mensurar a importância do que estava acontecendo, foi muitas vezes paciente com todo o tempo em que tive que abdicar de estar com ele para trabalhar na tese.

Aos meus familiares: minhas irmãs Nina, Bia e Jaque, meu irmão Anderson, minha madrastra Rosângela, meus sogros Maria Lúcia e Marco Polo, grandes incentivadores e que se fizeram presentes, mesmo à distância.

Aos grandes amigos que fiz na UFLA, em especial, Renata e Denise, que foram muito mais do que companheiras de trabalho.

Aos colegas, amigos e professores do Grupo de Estudos em Séries Temporais, em especial, Marcela, Luiz Otávio, Paulo Henrique e Ednilson, pelas imensas colaborações e trocas ao longo desses anos.

À minha orientadora Thelma, que me orientou durante este período não apenas com relação à tese, mas me aconselhando e me fazendo olhar para a pesquisa e o trabalho com outros olhos.

Aos membros da banca examinadora, Profa. Chang, Profa. Izabela, Prof. Ednilson e Prof. Paulo Henrique, pela disposição em contribuir com este trabalho.

Aos demais professores do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, em especial ao Júlio, por muitas trocas e inspirações, e à Izabela, pela parceria e confiança com relação a um projeto muito especial para mim, o R-ladies.

À Universidade Federal de Lavras (UFLA) e ao Departamento de Estatística (DES) por me proporcionarem a realização deste doutorado.

À todos os parentes e amigos não citados anteriormente, mas não menos importantes, que me apoiaram nesta trajetória e se fizeram presentes com uma conversa amiga e, como não podia faltar, uma xícara de café.

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG).

*"Não basta a leitura sem a unção, não basta a especulação sem a devoção,  
não basta a pesquisa sem maravilhar-se, não basta a circunspeção sem o júbilo,  
o trabalho sem a piedade, a ciência sem a caridade, o estudo sem a graça".*  
São Boaventura

## RESUMO

A presença de saltos possui forte impacto na previsão da volatilidade de ativos financeiros. Estes saltos podem ser compreendidos como mudanças estruturais locais na série, devido a um movimento natural das negociações, ou ainda, como os choques frequentemente associados a uma questão comportamental dos investidores, em geral, causada por anúncios de notícias macroeconômicas. Desta maneira, é imperativo tentar compreendê-los bem, já que é um desafio tentar prevêê-los. A abordagem de ondaletas pode ser uma forte aliada em situações como esta, uma vez que detectam locais de salto com eficiência, além de possibilitarem uma interpretação econômica ao relacionar as escalas de tempo obtidas na decomposição com os horizontes de investimento do cotidiano dos operadores de bolsa de valores. O presente trabalho utilizou o limite universal de Donoho e Johnstone para detectar saltos nas séries obtidas com a transformada de ondaletas MODWT da série de preços. Os saltos detectados foram então mensurados e utilizados na estimação da Variação de Salto (componente discreta da Variação Quadrática, medida natural da variabilidade do processo de preço). Construída desta forma, é possível observar a contribuição dos saltos detectados em cada uma das escalas de tempo da decomposição de ondaletas para a composição da Variação de Salto, levando a uma melhor compreensão da relevância de cada nível/horizonte para esta componente descontínua da variabilidade do processo. Além disso, a série de retornos pode ser refinada com a eliminação dos saltos localizados nos diversos níveis, de forma que os modelos da família HAR ajustados a estas novas séries refletem o comportamento da Variância Integrada do processo, livre dos efeitos dos saltos. Para o estudo dos saltos e uma melhor compreensão de seu comportamento, uma aplicação em dados reais foi realizada a partir da série de log-preços das ações preferenciais da Petrobrás (PETR4), na frequência de 1 minuto, em um período com uma forte queda evidenciada por uma intervenção na presidência da estatal. A metodologia utilizada evidenciou que, particularmente para esta queda de preços citada, a variabilidade devida aos choques é impactada de forma que sua estimativa mais que triplica ao se considerar também os níveis de menor frequência, correspondentes a horizontes de investimentos que vão de minutos até 1 a 2 horas de negociação, o que destaca também o período de tempo que o efeito da notícia leva para diluir no mercado de ações. A aplicação dos modelos da família HAR evidenciou a presença de memória longa da volatilidade realizada: tanto no modelo HAR-RV tradicional, após eliminação dos saltos, quanto no HAR-J e HAR-CJ, através de suas componentes discretas. Considerar os modelos robustos a saltos permitiu evidenciar que eliminar os saltos a partir da decomposição de ondaletas, além da interpretabilidade que esta abordagem favorece, traz mais liberdade para aplicação de outros modelos de previsão de volatilidade que não se restrinjam à robustez a saltos.

**Palavras-chave:** Análise de saltos multiescala. Volatilidade Realizada. Retornos financeiros de alta frequência. Decomposição de ondaletas. Mercado brasileiro de ações.

## ABSTRACT

The presence of jumps has a strong impact on predicting the volatility of financial assets. These jumps can be understood as local structural changes in the series, due to a natural movement of the negotiations, or even as shocks often associated with a behavioral issue of investors, usually caused by macroeconomic news announcements. Therefore, it is imperative to try to understand them well, as it is a challenge to try to predict them. The wavelet approach can be a strong ally in situations like this, since it detects jump locations efficiently, in addition to allowing an economic interpretation by relating the time scales obtained in the decomposition with the investment horizons of the daily life of stock exchange operators. The present work used the universal limit of Donoho and Johnstone to detect jumps in the series obtained with the MODWT wavelet transform of the price series. The jumps detected were then measured and used in the estimation of the Jump Variation (the discrete component of the Quadratic Variation, a natural measure of the variability of the price process). Constructed in this way, it is possible to observe the contribution of the jumps detected in each of the wavelet decomposition time scales to the composition of the Jump Variation, leading to a better understanding of the relevance of each level/horizon for this discontinuous component of the variability of the process. Furthermore, the series of returns can be refined by eliminating the jumps located at the different levels, so that the HAR family models adjusted to these new series reflect the behavior of the Integrated Variance of the process, free from the effects of the jumps. For the study of jumps and a better understanding of their behavior, an application in real data was carried out from the log-price series of Petrobras preferred shares (PETR4), at a frequency of 1 minute, in a period with a strong fall evidenced by an intervention in the presidency of the state-owned company. The methodology used showed that, particularly for this price drop mentioned, the variability due to shocks is impacted in a way that its estimate more than triples when also considering the lower frequency levels, corresponding to investment horizons ranging from minutes to 1 to 2 hours of trading, which also highlights the length of time the news effect takes to dilute in the stock market. The application of the HAR family models showed the presence of a long memory of the realized volatility: both in the traditional HAR-RV model, after eliminating the jumps, and in the HAR-J and HAR-CJ, through their discrete components. Considering the jump robust models showed that eliminating jumps from the wavelet decomposition, in addition to the interpretability that this approach favors, brings more freedom to apply other volatility prediction models that are not restricted to jump robustness.

**Keywords:** Multiscale jump analysis. Realized Volatility. High-frequency financial returns. Wavelet decomposition. Brazilian Stock Market.

## LISTA DE FIGURAS

Figura 3.1 – Funções de ondaleta pai $\phi(t)$ (a) e ondaleta mãe $\psi(t)$ (b) da Haar. Algumas translações e dilatações da Haar: $\psi_{1,0}(t)$ (c), $\psi_{1,1}(t)$ (d), $\psi_{2,1}(t)$ (e) e $\psi_{3,1}(t)$ (f). . . . .	24
Figura 4.1 – Obtenção de séries de retornos atualizadas com a eliminação dos saltos obtidos a cada nível. . . . .	39
Figura 5.1 – Gráficos da série de preços (a) e dos retornos (b) das ações da PETR4 no período de 04/01/2021 a 25/06/2021 para a frequência de 1 minuto. . . . .	41
Figura 5.2 – Histograma (a) e QxQ plot (b) da série de retornos das ações da PETR4 no período de 04/01/2021 a 25/06/2021 para a frequência de 1 minuto. . . . .	42
Figura 5.3 – Gráficos da série de log-preços original e dos coeficientes de ondaletas dos 15 níveis de decomposição. . . . .	43
Figura 5.4 – Exemplificação do processo de localização de saltos para o dia 13/01/2021, com 11 saltos. (a) Gráfico dos coeficientes de ondaletas referente ao primeiro nível de decomposição. (b) Gráfico dos valores absolutos dos coeficientes de ondaletas com o limite universal em vermelho indicando as observações que representam salto. (c) Gráfico da série de log-preço com os respectivos locais de salto detectados. (d) Gráfico do retorno ajustado aos saltos localizados. . . . .	44
Figura 5.5 – Relação do número de saltos encontrados por dia e do número de dias que o apresentaram para cada nível de decomposição. . . . .	45
Figura 5.6 – Proporção de dias com e sem salto para cada nível da decomposição. . . . .	46
Figura 5.7 – Saltos detectados para o dia 15/03/2021, com 56 saltos localizados no nível $d_{10}$ . Em (a) a série de valores absolutos dos coeficientes de ondaletas com o respectivo limite universal em vermelho; em (b) a série de log-preços com a área hachurada destacando a região onde os saltos foram localizados. . . . .	46
Figura 5.8 – Gráficos dos histogramas das localizações de salto detectados nos níveis $d_1$ (em (a)) a $d_8$ (em (h)) com destaques amplificados para a região de maior frequência dos saltos localizados para a escala correspondente (ver Tabela 5.1). . . . .	48
Figura 5.9 – Gráficos referentes à variação de salto nível a nível da decomposição: (a) Valor estimado; (b) Valor acumulado. . . . .	50

Figura 5.10 – Gráfico da série de log-preços com as localizações dos saltos detectados em $d_1$ (a); e dos valores absolutos dos retornos (b) para os dias 22 a 26 de maio de 2021. . . . .	52
Figura 5.11 – Gráficos da série de log-preços para o dia 22/02/2021 com a identificação dos saltos localizados para os níveis $d_1$ (a), $d_2$ (b), $d_3$ (c), $d_4$ (d), $d_5$ (e), $d_6$ (f), $d_7$ (g). Para os níveis $d_4$ a $d_7$ , figuras (d) a (g), nota-se regiões (ou períodos) de salto, demarcados em vermelho pelo início e término do período referido, e o intervalo representado pela região hachurada em cinza. . . . .	53
Figura 5.12 – Gráficos das medidas realizadas: $RV$ , $BPV$ e $JV$ para o período detectado.	55
Figura 5.13 – Gráficos comparativos para as Variações de saltos estimadas por nível. . . . .	57
Figura 5.14 – Estimativas do ajuste do modelo HAR-RV tradicional para a série original e as séries ajustadas aos saltos a cada nível da decomposição. . . . .	58
Figura 5.15 – Estimativas do ajuste do modelo HAR-J para a série original e as séries ajustadas aos saltos a cada nível da decomposição. . . . .	59
Figura 5.16 – Estimativas do ajuste do modelo HAR-CJ para a série original e as séries ajustadas aos saltos a cada nível da decomposição. . . . .	60

## LISTA DE TABELAS

Tabela 5.1 – Relação entre nível de decomposição e Escala de tempo. . . . .	42
Tabela 5.2 – Valores da Variação de Salto estimada e acumulada a cada nível da decomposição da série de log-preço. . . . .	49
Tabela 5.3 – Número de saltos detectados nos dias da semana de 22 a 26 de fevereiro de 2021. . . . .	50
Tabela 5.4 – Descritiva das medidas realizadas $RV$ , $BPV$ e $JV$ . . . . .	54
Tabela 5.5 – Descritiva das variações de salto estimadas em cada nível da decomposição.	56
Tabela 5.6 – Relação da qualidade de ajuste $R^2$ em relação ao ajuste do modelo HAR-RV para os dados originais. . . . .	59

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>11</b>
<b>2</b>	<b>REVISÃO DE LITERATURA</b>	<b>15</b>
<b>3</b>	<b>REFERENCIAL TEÓRICO</b>	<b>20</b>
<b>3.1</b>	<b>Transformadas de Ondaletas</b>	<b>20</b>
<b>3.2</b>	<b>Decomposição da Variação Quadrática</b>	<b>27</b>
<b>3.3</b>	<b>Estimação da Variação de Salto via Ondaletas</b>	<b>29</b>
<b>3.4</b>	<b>Medidas Realizadas da Volatilidade</b>	<b>31</b>
<b>3.5</b>	<b>Modelos da família HAR</b>	<b>34</b>
<b>4</b>	<b>MATERIAIS E MÉTODOS</b>	<b>37</b>
<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>40</b>
<b>5.1</b>	<b>Variação de Saltos</b>	<b>41</b>
<b>5.2</b>	<b>Semana de 22 a 26 de fevereiro de 2021</b>	<b>50</b>
<b>5.3</b>	<b>Comparação de modelos da família HAR</b>	<b>54</b>
<b>5.3.1</b>	<b>Medidas realizadas</b>	<b>54</b>
<b>5.3.2</b>	<b>HAR-RV</b>	<b>56</b>
<b>5.3.3</b>	<b>HAR-J</b>	<b>58</b>
<b>5.3.4</b>	<b>HAR-CJ</b>	<b>60</b>
<b>6</b>	<b>CONCLUSÃO</b>	<b>62</b>
	<b>REFERÊNCIAS</b>	<b>64</b>

## 1 INTRODUÇÃO

A compreensão da dinâmica da volatilidade de ativos financeiros é de suma importância para gestão de risco de carteiras e alocação de portfólios, em especial em momentos de crises, como a situação político-econômica que o Brasil tem enfrentado nos últimos anos, ainda mais agravada com a crise global provocada pela Covid-19.

Como a volatilidade geralmente apresenta padrões diferentes durante períodos distintos, principalmente devido a choques exógenos, como crises financeiras e/ou mudanças políticas, o processo de difusão de salto em tempo contínuo é tradicionalmente utilizado; visto que a componente de difusão captura a variação suave do processo de preços, enquanto o componente de salto lida com as raras e grandes descontinuidades nos preços observados (LAURENT, 2018).

Andersen, Bollerslev e Diebold (2007) citam diversos trabalhos que evidenciam que este é um modelo realista para a série de preços de muitos ativos financeiros. Os autores constataram a importância da componente de salto e a melhora na previsão da volatilidade ao separar a componente contínua daquela devida a saltos, de forma que a Variação Quadrática (medida natural da variabilidade do processo de preço) é decomposta em Variância Integrada e Variação de Salto. Separar a Variação de Saltos na Variação Quadrática é importante, uma vez que ela é menos previsível devido ao fluxo de notícias do mercado (FAN; WANG, 2007).

Um salto pode ser considerado uma observação anormalmente grande em termos de uma magnitude que dependerá basicamente da volatilidade do processo da geração de dados, por isto pode ser entendido como uma grande mudança estrutural local na série de preços. Como eles representam incertezas inerentes ao mercado financeiro, sua previsão se torna naturalmente complexa, já que refletem a chegada de notícias impactantes ou negociações de ordens de tamanho anormal.

Apesar de sua importância e da ampla discussão acerca do assunto, não há um consenso na literatura a respeito de como identificar e de qual a melhor forma de lidar com estes saltos. É comum associá-los a anúncios de notícias macroeconômicas, porém outras vertentes apontam que podem ser devido à falta de liquidez do mercado ou, ainda, como assume a literatura de finanças comportamentais, podem ser causados pela conduta de determinados participantes do mercado. Hanousek, Kočenda e Novotný (2012) enfatizam que, independentemente da crença em torno da origem destes saltos, mais importante que tentar prevê-los, é preciso compreendê-los bem e estar pronto para reagir a eles adequadamente.

Diversas técnicas para detectar saltos podem ser empregadas, muitas delas baseadas em testes estatísticos que não detectam os saltos um a um, apenas os mensuram em um período determinado (HANOUSEK; KOČENDA; NOVOTNÝ, 2012). No sentido de obter a identificação exata do local de salto, uma técnica que pode ser empregada é dada a partir da limiarização (ou *thresholding*), em que um dado valor é considerado limítrofe para identificação de presença/ausência de saltos. Neste trabalho, a técnica abordada para a detecção de saltos é proveniente da decomposição da série de dados em vários níveis de detalhes, por meio da transformada de ondaletas, conforme demonstrado por Wang (1995).

A literatura em ondaletas tem se mostrado uma ferramenta útil aos economistas. Crowley (2007) aponta diversas contribuições para esta área, uma delas, em especial, dá-se pelo fato de que a maioria da literatura empírica considera apenas o comportamento agregado a um único horizonte de tempo, enquanto que, ao operar no mercado financeiro, os investidores podem estar interessados em diferentes horizontes de investimento dependendo da estratégia a ser utilizada, com horizontes intradiários (para um mesmo dia, em minutos ou horas) ou até mesmo horizontes diários, semanais, mensais e até anuais<sup>1</sup>. Uma exceção é o modelo autorregressivo heterogêneo (HAR, do inglês, *Heterogeneous AutoRegressive*, proposto por Corsi (2009)) que, apesar de considerar esta heterogeneidade dos investidores, o faz de maneira limitada ao domínio do tempo (BARUNIK; KREHLIK; VACHA, 2016).

Desta forma, o uso da transformada de ondaletas permite o estudo da influência destes diferentes horizontes de investimento, a partir de estimadores multiescala que decompõem a volatilidade em diferentes horizontes de investimentos. Apesar disto, não raro, apenas o nível de maior frequência é considerado (BARUNIK; VACHA, 2015) (BARUNIK; KREHLIK; VACHA, 2016), uma vez que é neste nível que se concentra a presença de ruídos da série. No entanto, Wang (1995) e Fan e Wang (2007) sugerem que outros níveis da decomposição podem identificar saltos importantes que seriam desconsiderados quando utilizado apenas o nível mais fino. Em seu estudo empírico seminal, Wang (1995) utilizou dados mensais de retorno do mercado de ações dos Estados Unidos no período de 1953 a 1991. Os dois picos evidenciados foram devidos à recessão de 1974 e ao *crash* da bolsa de Nova York em 1987. A série amostrada foi decomposta em 9 níveis de detalhes e estes picos foram detectados nos níveis 6 e 7, respectivamente. Este resultado impulsiona uma análise mais detalhada para detectar saltos em outros níveis da decomposição além daquele de maior detalhe.

---

<sup>1</sup> Para mais detalhes, consultar Corsi e Reno (2009), Gençay, Selçuk e Whitcher (2001) e Ramsey (1999).

Na literatura brasileira, os estudos com foco na detecção e compreensão de saltos ainda são incipientes. Os trabalhos mais comuns são aqueles que comparam modelos para a volatilidade considerando, entre eles, algum modelo robusto a saltos, como por exemplo Variação Bipower (em inglês, *BiPower Variation* – BPV), desenvolvido por Barndorff-Nielsen e Shephard (2004), como é o caso dos trabalhos de Santos e Ziegelmann (2014) e Araújo Júnior (2020). Um trabalho que se destaca é o de Pinto (2021) que considerou métodos de limiarização na decomposição em ondaletas e comparou modelos de memória longa para dados de alta frequência com e sem ajustes de saltos.

Sendo assim, a proposta desta tese se baseia na exploração da informação de salto contida em cada um dos horizontes de investimento obtidos a partir da análise de multirresolução através da decomposição de ondaletas ao propor que a Variação de Salto, ao invés de ser considerada como uma medida global, possa ser calculada para cada nível de decomposição e viabilizar uma discussão a respeito também da escolha de quantos e quais níveis de decomposição considerar. Esta abordagem permite avaliar o impacto destas diferentes variações na estimação da volatilidade do processo, dada pela Variância Integrada.

Outra grande vantagem desta abordagem se dá devido à interpretação de cada escala em associação aos horizontes de investimento, que fazem parte do cotidiano dos *traders* (especuladores do mercado). Além da estimação da Variação de Salto a cada nível, a metodologia também possibilita a criação de novas séries ajustadas aos saltos detectados, de maneira a explorar a contribuição de cada nível na melhora de ajuste de modelos não necessariamente robustos a saltos (como é o caso do modelo HAR tradicional) em comparação às suas alternativas que consideram saltos em suas modelagens. A metodologia proposta se baseia no que foi apresentado por Barunik e Vacha (2015) e Barunik, Krehlik e Vacha (2016), que diferentemente do trabalho de Fan e Wang (2007), utilizam a transformada não decimada. Esta escolha é importante, pois, para a interpretação dos locais de saltos em diversas escalas de tempo, é necessário que todas elas apresentem o mesmo comprimento que a série original, para que haja uma correspondência entre o salto detectado no nível e a série de interesse. Além disso, como a decomposição de ondaletas é utilizada apenas para fins de identificação dos locais de salto, não há prejuízos quanto à não ortogonalidade da transformada, uma vez que não será utilizada reconstrução a partir dos diversos níveis.

Desta maneira, como a abordagem tem um propósito maior no sentido da interpretação nível a nível de decomposição, um estudo empírico com dados reais foi conduzido a partir

de uma importante série de preços da bolsa de valores brasileira, as ações preferenciais da Petrobrás. A escolha deste ativo foi motivada por uma forte queda nos preços sofrida devido a uma interferência externa ao mercado, a troca do presidente da estatal por parte do presidente da República. A queda de preços teve alta repercussão e impactou até mesmo o Ibovespa.

Uma distinção pode ser feita uma vez que as definições de saltos apresentadas englobam tanto os saltos que são inerentes ao movimento do mercado, quanto aos choques que são devido à notícias impactantes. No contexto desta tese, o foco principal será na compreensão deste último, ou seja, nos choques de volatilidade que representam de fato um movimento anormal devido a uma causa externa ao mercado de ações<sup>2</sup>.

Após essa introdução que apresentou o problema de pesquisa e descreveu brevemente os objetivos deste trabalho, o capítulo 2 apresenta uma revisão da literatura que sumariza trabalhos que abordam os diferentes temas que vão compor a metodologia desta tese. O referencial teórico, apresentado no capítulo 3, traz as particularidades dos conceitos necessários para a aplicação da metodologia, como a base da teoria de ondaletas, a decomposição da variação quadrática em suas componentes contínua e discreta, as medidas realizadas para volatilidade e os modelos da família HAR. Em seguida, a série a ser explorada e a metodologia aplicada são descritas no capítulo 4. Os resultados da aplicação são apresentados no capítulo 5, que esmiúça a abordagem de ondaletas da estimação da variação de salto, a contribuição de cada nível, detalha o comportamento dos saltos na semana de maior impacto dos saltos e apresenta a comparação dos modelos para os dados originais e ajustados aos saltos. Por fim, as conclusões e as considerações finais do trabalho estão descritas no capítulo 6.

---

<sup>2</sup> Como grande parte da literatura utilizada neste trabalho não faz distinção destes termos, utilizando *jumps* para ambas as situações, optou-se por fazer o mesmo na presente tese, utilizando "salto" para ambos os casos, exceto aqueles referentes a um choque muito específico.

## 2 REVISÃO DE LITERATURA

Neste capítulo serão apresentados alguns trabalhos que abordam e dão suporte a esta tese. Eles estão ordenados inicialmente com os artigos seminais de pesquisa em alta frequência, seguido por trabalhos relacionados às modelagens de volatilidade, em especial, a volatilidade realizada. Depois, são apresentados os trabalhos com enfoque em abordagens de ondaletas no contexto da previsão da volatilidade e, por fim, o que há relacionado na literatura econométrica brasileira.

A disponibilidade de dados de alta frequência trouxe novos desafios para os pesquisadores, em especial na modelagem e previsão da volatilidade de ativos financeiros. Um dos primeiros trabalhos em dados de alta-frequência é de Hsieh (1991) que apresentou estimativas de variação de retornos diários considerando retornos intradiários do S&P500. Outros trabalhos relevantes se seguiram, como o de Zhou (1996) que utilizou dados de alta frequência no mercado de câmbio para explicar a autocorrelação negativa de retornos e estimar a volatilidade para estes dados; Andersen e Bollerslev (1997) e Andersen e Bollerslev (1998b) estudaram o comportamento da volatilidade intradiária, choques de volatilidade e persistência de longo prazo nas séries da volatilidade realizada do mercado de câmbio; e os efeitos da estrutura de mercado, efeitos sazonais intradiários e efeitos da volatilidade temporal foram analisados por Goodhart e O'Hara (1997). Outros trabalhos importantes estabeleceram propriedades teóricas e empíricas da estimação da variação quadrática, como os trabalhos de Andersen e Bollerslev (1998a), Andersen et al. (2001b), Andersen et al. (2001a), Barndorff-Nielsen e Shephard (2002a) e Meddahi (2002).

Esta nova classe de dados (dados de alta frequência ou intradiários) possibilitou uma nova abordagem a respeito do estudo e previsão da volatilidade, uma medida que não é diretamente observável da variância condicional latente. A conhecida Volatilidade Realizada (RV - do inglês *Realized Volatility*)<sup>3</sup>, dada pela soma de quadrados dos valores dos retornos intradiários é considerada, portanto, uma realização desta volatilidade oculta (SHIN, 2018).

A previsão da volatilidade realizada é, então, alternativa às previsões de volatilidade diárias em diversas aplicações financeiras. Os trabalhos de Giot e Laurent (2004), Louzis, Xanthopoulos-Sisinis e Refenes (2014), além de Wong, Chin e Tan (2016), por exemplo, compararam previsões de VaR (*Value-at-Risk*) baseados em modelos do tipo GARCH (sigla em

<sup>3</sup> Ao longo de todo o trabalho, serão adotadas as siglas em inglês em razão de facilitar o acesso e compreensão do que é empregado na literatura encontrada.

inglês para *Generalized AutoRegressive Conditional Heteroskedasticity*) e modelos de RV. Outros trabalhos, por exemplo, Čech e Baruník (2017) e Chiriac e Voev (2011) consideraram balanceamento de portfólio baseado nas previsões da matriz de covariâncias realizadas.

É possível encontrar na literatura alguns trabalhos de revisão a respeito de previsão de volatilidade, como a pesquisa comparativa realizada por Poon e Granger (2003) envolvendo diferentes abordagens de volatilidade. Andersen et al. (2006) retrataram modelos de previsão de volatilidade em casos uni e multivariados. McAleer e Medeiros (2008) se basearam nas características de memória longa das volatilidades realizadas. Andersen e Teräsvirta (2009) consideraram salto e ruído de microestrutura na revisão da previsão de volatilidade realizada, para os modelos MIDAS (*Mixed Data Sampling*, desenvolvido por Ghysels, Santa-Clara e Valkanov (2004)) e modelo de erro multiplicativo (MEM - *Multiplicative Error Model*).

As revisões mais recentes, de Bucci et al. (2017) e Shin (2018), se complementam, o último dando mais enfoque na classe de modelos HAR-RV (*Heterogeneous AutoRegressive Realized Volatility*), enquanto o primeiro trata de diferentes abordagens além do HAR, como GARCH, MIDAS, *Realized GARCH* e HEAVY (*High-frequency-based Volatility models*, de Shephard e Sheppard (2010)).

É importante destacar que, desde a proposta da volatilidade realizada por Andersen e Bollerslev (1998a), muitas outras versões desta foram sugeridas a fim de melhorar o estimador da variância integrada. Alguns deles são MEDRV e MINRV de Andersen, Dobrev e Schaumburg (2012), BPV de Barndorff-Nielsen e Shephard (2004), RTSCOV e RAVGCOV de Zhang, Mykland e Aït-Sahalia (2005), RKERNELCOV, de Barndorff-Nielsen et al. (2004) e RRTSCOV de Boudt e Zhang (2015), entre outros.

O modelo HAR proposto por Corsi (2009) é o mais amplamente utilizado no contexto de previsão de volatilidade realizada. Ele apresenta uma estrutura flexível, estimação tratável e um desempenho de previsão superior, o que resultou em diversos modelos estendidos a partir dele. Baseado na hipótese de mercado heterogêneo e no modelo HARCH (*Heterogeneous AutoRegressive Conditional Heteroskedasticity*) de Müller et al. (1997), é um modelo AR (autorregressivo) parcimonioso, que incorpora componentes de volatilidade diário, semanal e mensal.

Extensões do modelo HAR foram propostas ao logo dos últimos anos a fim de sanar algumas limitações. Por exemplo, o modelo proposto por Andersen, Bollerslev e Diebold (2007) que visa incorporar no modelo HAR convencional uma decomposição da volatilidade realizada

em um componente contínuo e outro discreto. Efeitos de alavancagem nas previsões da volatilidade foram considerados pelo modelo de Patton e Sheppard (2015) e no LHAR (*leverage HAR*, (CORSI; RENO, 2009)). O primeiro propõe uma variação de salto com sinal no modelo HAR, enquanto o segundo considera uma dependência de longo alcance no modelo base.

Os primeiros a sugerir um estimador de ondaletas para a variância integrada foram Høg e Lunde (2003) inspirados num estimador baseado na análise de Fourier. Já visando diminuir a discrepância entre a variância integrada e a variância realizada, Capobianco (2003) propôs o uso da transformada de ondaletas como um estimador comparável da variação quadrática e mostrou experimentalmente que a variância realizada convergia para a integrada através desta abordagem. Subbotin (2008) utilizou ondaletas para decomposição em tempo e escala da volatilidade e propôs um indicador de volatilidade para o mercado de ações que caracterize as flutuações tanto no domínio do tempo quanto da frequência. O trabalho de Gençay et al. (2010) explorou a linha de análise de multirresolução a partir das ondaletas para decompor a variância realizada e averiguar dependência da volatilidade em diferentes horizontes de tempo.

O uso de ondaletas para detectar localização de salto de forma eficiente foi demonstrado por Wang (1995). Alguns anos mais tarde, Fan e Wang (2007) iniciaram uma nova direção de pesquisa sobre estimativa de salto de preços e volatilidade em dados financeiros de alta frequência, a partir de métodos multiescala, propondo métodos não paramétricos para estimar saltos e variações de saltos para dados com ruídos de microestrutura de um modelo de difusão de salto. Eles aplicam o TSRV (ZHANG; MYKLAND; AÏT-SAHALIA, 2005) e o MSRV (ZHANG, 2006) em dados ajustados a saltos na estimativa da volatilidade integrada. Os autores apresentam o estimador de volatilidade realizada por ondaletas aos dados ajustados por salto e estabelecem uma conexão entre TSRV e a RV pela ondaleta Haar. Concluíram através das simulações que, de fato, os estimadores propostos superam os métodos existentes quando os dados contêm saltos e são equiparáveis caso contrário.

Ampliando a ideia de Fan e Wang (2007), Barunik e Vacha (2015) e Barunik, Krehlik e Vacha (2016) apresentam uma metodologia baseada em ondaletas para estimar a variância realizada que possibilita analisá-la no domínio do tempo-frequência. Em Barunik e Vacha (2015), é conduzido um estudo numérico do desempenho do estimador proposto, comparando com vários dos estimadores mais populares, como variância realizada, variação bipower, volatilidade realizada em duas escalas e kernels realizados. Segundo os autores, o estimador baseado em ondaletas provou ser o menos afetado pela presença de saltos e/ou ruído de microestrutura (em

diferentes níveis). Já em Barunik, Krehlik e Vacha (2016), os autores ampliam o estudo anterior propondo um estimador baseado no modelo *Realized* GARCH, de Hansen, Huang e Shek (2012) e no *Realized* GAS (*Generalized Autoregressive Score*) de Huang, Wang e Zhang (2014) através da volatilidade integrada decomposta, obtendo resultados superiores aos estimadores de referência.

Estudos relacionados a dados e modelagem econométrica em alta frequência têm sido mais frequentes no Brasil desde os anos 2010. Ao longo desta década, diversos pesquisadores focaram em estudar bases de dados de alta frequência do mercado brasileiro de ações, visando a previsão de volatilidade.

Santos e Ziegelmann (2014) comparam modelos MIDAS e HAR, além de uma combinação destes dois, para duas medidas realizadas de referência, a tradicional volatilidade realizada e a variação bipower, robusta a saltos. Apesar de seus resultados concluírem que os modelos são equivalentes para previsão da volatilidade futura, as previsões são melhores para a medida robusta a saltos.

Wink Junior e Pereira (2011) compararam dois modelos preditivos de volatilidade, HAR e MIDAS, e buscaram corrigir o viés de microestrutura através do filtro MA(q), utilizando dados intradiários com frequência de 5, 15 e 30 minutos para os cinco ativos mais líquidos da B3, concluindo a favor do primeiro para previsões fora da amostra e o segundo, dentro da amostra. Eles sugerem que trabalhos com dados brasileiros considerando efeitos de saltos e outros modelos contemplando efeito de microestrutura poderiam enriquecer a literatura econométrica brasileira.

Já em seu trabalho, Val (2016) compara a previsão dada por modelos GARCH com as obtidas em modelos que consideram dados intradiários, como o HAR e 2-Comp (modelo 2 Componentes, de Maheu e McCurdy (2007)), para os dois ativos mais líquidos do Ibovespa e confirma a superioridade dos modelos para dados de alta frequência para previsões tanto dentro quanto fora da amostra.

Outra comparação de projeções de volatilidade para, desta vez, três ações mais negociadas na nossa bolsa de valores foi realizada por Araújo (2016), utilizando os métodos HAR-RV, MIDAS-RV, ARFIMA - *AutoRegressive Fractionally Integrated Moving Average* e NN - *Nearest Neighbor*, além de propor modelos de combinação de volatilidades projetadas. Dentre os resultados obtidos, conclui que o método NN apresenta melhores resultados de projeção que os demais modelos utilizados.

Araújo Júnior (2020) buscou fazer uma análise econométrica do mercado de ações brasileiro, na qual investigou os principais fatos estilizados relatados pela literatura de finanças em alta frequência, além de comparar a qualidade do ajuste e a capacidade preditiva dos modelos de previsão HAR-RV e HEAVY para 20 maiores empresas do Ibovespa utilizando diversas medidas de volatilidade realizada, diferentes frequências e métodos de transformação de frequências. Uma das suas conclusões, em relação à previsão da volatilidade, foi que o tratamento de *outliers* não provocou impacto considerável na capacidade preditiva dos modelos HAR, e que as previsões são melhores para as maiores frequências utilizadas.

Utilizando abordagem de ondaletas na previsão de volatilidade, o trabalho de Pinto (2021) comparou os modelos ARFIMA e FI(E)GARCH para modelar a memória longa para dados de alta frequência com e sem ajustes de saltos, considerando para estes ajustes métodos de limiarização na decomposição em ondaletas, e obteve melhores previsões para as séries com ajustes de saltos nas diferentes aplicações utilizadas.

### 3 REFERENCIAL TEÓRICO

Neste capítulo, são apresentados os principais aspectos metodológicos para o desenvolvimento da presente tese. Inicialmente, serão abordados alguns conceitos da teoria de ondaletas, que será um importante instrumento para o estimador da Variação de Salto. Em seguida, são trazidos tópicos relacionados à decomposição da Variação Quadrática em suas componentes contínua e discreta e o processo de detecção de salto. Posteriormente, os fundamentos das medidas realizadas de volatilidade e os modelos da família HAR aplicados a elas concluem o capítulo.

#### 3.1 Transformadas de Ondaletas

De maneira simplificada, ondaletas são pequenas ondas que, independentemente da sua forma, começam e terminam em um ponto finito no tempo. Desta forma, apresentam um número definido de oscilações e duram por um determinado período de tempo ou espaço, o que as tornam capazes de aproximar localmente as variáveis no tempo (ou espaço), uma vez que tem a capacidade de serem manipuladas convenientemente, sendo "esticadas" ou "comprimidas" de modo a reproduzir a série em estudo. Seu interesse nas áreas de economia e finanças se dá basicamente pela sua capacidade de lidar com dados estacionários e não estacionários, sua localização no tempo e sua capacidade de decompor e analisar a oscilação de uma variável (CROWLEY, 2007).

Ramsey (1999) evidencia fortes contribuições da teoria de ondaletas na literatura econométrica que, no contexto deste trabalho, se destaca

1. Decomposição em escalas de tempo: as relações entre as variáveis podem ser encontradas no nível desagregado (escala) ao invés do nível agregado;
2. Previsão: desagregar previsão, estabelecendo aspectos da série global *versus* local.

Na análise de ondaletas, o objetivo é gerar o espaço integrável  $L^2(\mathbb{R})$  através de dilatações (ou compressões) e translações de uma única função,  $\psi$ , denominada ondaleta mãe, dada por

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right), \quad b \in \mathbb{R}, \quad a > 0. \quad (3.1)$$

O parâmetro  $a$ , correspondente à escala, refere-se ao processo de compressão/dilatação do sinal, que permite analisar aspectos locais e globais deste, de forma que ao se variar

as escalas, pode-se obter uma visão mais ou menos detalhada do sinal. Já o parâmetro  $b$ , correspondente à translação, está associado à informação de localização no tempo, de forma que, a medida em que o parâmetro varia, um sinal é analisado localmente em torno dele mesmo (OLIVEIRA, 2007).

Um caso particular e muito utilizado é obtido de  $\psi(t)$  por uma dilatação binária, tomando-se  $a = 2^j$ , e uma translação diádica, tomando-se  $b = k 2^{-j}$ ,  $j, k \in \mathbb{Z}$ , que fornecem as expressões

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad (3.2)$$

com  $j, k \in \mathbb{Z}$ . Neste caso, os parâmetros  $j$  e  $k$  representam escala e localização temporal, respectivamente. As funções

$$\{\psi_{j,k}(t); j, k \in \mathbb{Z}\} \quad (3.3)$$

formam uma base para  $L^2(\mathbb{R})$ , que não precisa ser necessariamente ortogonal. No entanto, a base ortogonal tem a vantagem de tornar possível a reconstrução perfeita do sinal original a partir dos coeficientes da transformada.

A partir da base em (3.3), para qualquer função  $f(t)$  de quadrado integrável sobre  $\mathbb{R}$ , ou seja, uma função cuja integral  $\int |f(t)|^2 dt$  existe, tem-se que

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{j,k} \psi_{j,k}(t), \quad (3.4)$$

onde  $c_{j,k}$  são chamados de coeficientes de ondaletas e são dados por

$$c_{j,k} = \int_{-\infty}^{\infty} f(t) \psi_{j,k}(t) dt. \quad (3.5)$$

Morettin (2014) lista algumas propriedades satisfeitas pelas funções  $\psi$  e  $\psi_{j,k}(t)$ , que são

P1.  $\int_{-\infty}^{\infty} \psi(t) dt = 0;$

P2.  $\int_{-\infty}^{\infty} |\psi(t)| dt < \infty;$

---

<sup>4</sup> A notação apresentada nesta seção caracteriza que frequências mais altas estão apresentadas nas escalas mais altas. No entanto, alguns (pacotes de) *softwares* utilizam a notação inversa, que pode ser encontrada em In e Kim (2013), na qual as primeiras escalas apresentam as frequências mais altas. Ao longo da aplicação desta tese, esta última foi a empregada e discutida por se caracterizar do pacote utilizado, no entanto, optou-se por manter esta notação na fundamentação teórica, uma vez que ela é mais atual e utilizada por diversos autores, como no livro Morettin (2014).

P3.  $c_\psi = \int_{-\infty}^{\infty} \frac{\Psi(\omega)}{\omega} d\omega < \infty$ , onde  $\Psi(\omega)$  é a transformada de Fourier;

P4.  $\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1$ , ou  $\int_{-\infty}^{\infty} |\Psi(\omega)|^2 d\omega = 2\pi$ ;

P5. Para algum  $r \geq 1$ , os primeiros  $r - 1$  momentos de  $\psi$  anulam-se, isto é  $\int_{-\infty}^{\infty} t^j \psi(t) dt = 0$ ,  $j = 0, 1, \dots, r - 1$  e  $\int_{-\infty}^{\infty} |t^j \psi(t)| dt < \infty$ .

A quinta propriedade, a respeito dos momentos nulos, permite caracterizar as ondaletas dentro de algumas famílias e o valor de  $r$  está relacionado ao grau de suavidade de  $\psi$  (quanto maior, mais suave).

Uma maneira de gerar ondaletas é através da ondaleta pai,  $\phi$ , também chamada de função escala, solução da equação

$$\phi(t) = \sqrt{2} \sum_k l_k \phi(2t - k) \quad (3.6)$$

que se relaciona com a ondaleta mãe,  $\psi$ , por

$$\psi(t) = \sqrt{2} \sum_k h_k \phi(2t - k). \quad (3.7)$$

Os coeficientes  $l_k$  e  $h_k$  são utilizados para calcular a Transformada de Ondaleta Discretas, DWT - *Discrete Wavelet Transform*, e são chamados, respectivamente de filtros passa-baixo (*low-pass*) e passa-alto (*high-pass*). Estes coeficientes são dados por

$$l_k = \sqrt{2} \int_{-\infty}^{\infty} \phi(t) \phi(2t - k) dt, \text{ e} \quad (3.8)$$

$$h_k = \sqrt{2} \int_{-\infty}^{\infty} \psi(t) \phi(2t - k) dt. \quad (3.9)$$

Análogo ao que acontece para a ondaleta mãe, a partir de dilatações e translações de  $\phi$ , a função 3.6 gera uma família de funções

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k), \quad (3.10)$$

com  $j, k \in \mathbb{Z}$ . O conjunto  $\{\phi_{j_o,k}(t), \psi_{j,k}(t); j, k \in \mathbb{Z}, j \geq j_o\}$ , permite escrever uma função  $f(t)$  do espaço  $L^2(\mathbb{R})$  como

$$f(t) = \sum_k c_{j_o,k} \phi_{j_o,k}(t) + \sum_{j \geq j_o} \sum_k d_{j,k} \psi_{j,k}(t), \quad (3.11)$$

cujos coeficientes de escala e ondaletas, respectivamente, são dados por

$$c_{j_o,k} = \int_{-\infty}^{\infty} f(t) \phi_{j_o,k}(t) dt, \quad (3.12)$$

$$d_{j,k} = \int_{-\infty}^{\infty} f(t) \psi_{j,k}(t) dt. \quad (3.13)$$

A ondaleta-pai (ou função de escala) representa essencialmente a parte de tendência suave (baixa frequência) da série, enquanto as ondaletas-mãe representam as partes detalhadas por escala (referente à alta frequência), observando a quantidade de alongamento da ondaleta ou "dilatação" (CROWLEY, 2007).

Há uma variedade de ondaletas, com diferentes formatos e características que justificam suas aplicações. A função de Haar é o exemplo mais antigo e simples de ondaletas, já as ondaletas de Daubechies, das quais a Haar é um caso particular, também são frequentemente utilizadas, uma vez que são ortogonais e de suporte compacto, como as *daublets*, *symmlets* e *coiflets*. Há outras tantas como a Morlet, Chapéu mexicano, Shannon, cada qual com suas peculiaridades. Mais detalhes podem ser obtidos em Sridhar, Kumar e Ramanaiah (2014).

Em algumas aplicações, especialmente na área de finanças, a utilização da ondaleta Haar é vista como equivocada, especialmente devido à suavização das funções que se costuma trabalhar. O que é visto como desvantagem para algumas aplicações, que é o fato de a transformada de Haar ser uma sequência de funções em forma de "quadrado", não é contínua e, portanto, não diferenciável e se assemelha a uma função degrau. Pode ser considerada uma vantagem para analisar sinais com transições bruscas, como a detecção dos saltos.

A função escala da função de Haar é  $\phi^{(H)}(t) = 1, 0 \leq t \leq 1$ , a ondaleta mãe é dada por

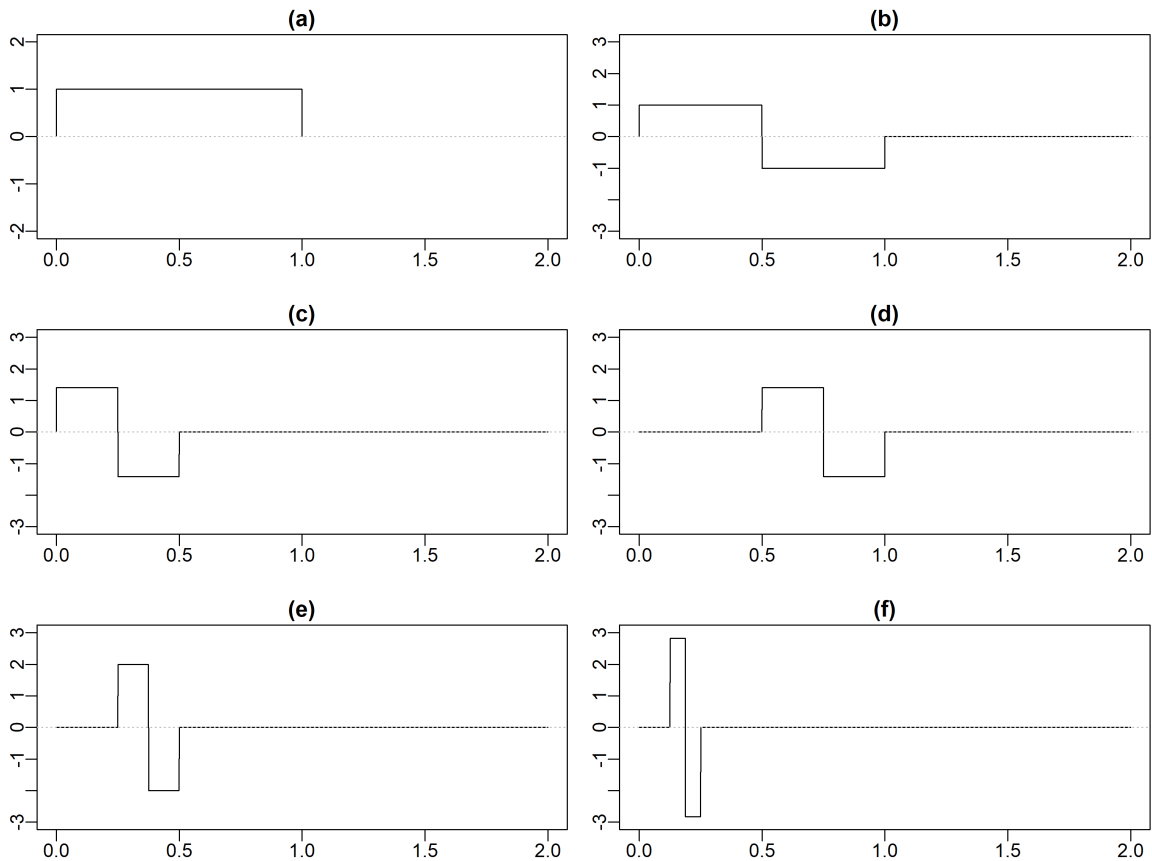
$$\psi^{(H)}(t) = \begin{cases} +1, & 0 \leq t < 1/2 \\ -1, & 1/2 \leq t < 1 \\ 0, & \text{caso contrário} \end{cases}, \quad (3.14)$$

e suas translações são dadas por

$$\psi_{j,k}^{(H)}(t) = \begin{cases} 2^{j/2}, & 2^{-j}k \leq t < 2^{-j}(k+1/2) \\ -2^{j/2}, & 2^{-j}(k+1/2) \leq t < 2^{-j}(k+1) \\ 0, & \text{caso contrário.} \end{cases} \quad (3.15)$$

Alguns exemplos das funções de Haar estão apresentadas na Figura 3.1.

Figura 3.1 – Funções de ondaleta pai  $\phi(t)$  (a) e ondaleta mãe  $\psi(t)$  (b) da Haar. Algumas translações e dilatações da Haar:  $\psi_{1,0}(t)$  (c),  $\psi_{1,1}(t)$  (d),  $\psi_{2,1}(t)$  (e) e  $\psi_{3,1}(t)$  (f).



Fonte: do autor (2022).

A análise de ondaletas possibilita a análise de multirresolução (MORETTIN, 2014), a partir da qual é possível analisar os dados em diversas escalas de resolução, como se fosse possível observar tais dados por um microscópio e estudar seu comportamento em diversas magnificações, ampliando o estudo da série no domínio do tempo, para contemplar também o domínio das frequências.

Para a decomposição da Variação Quadrática será utilizada a transformada de ondaleta contínua (CWT - *Continuous Wavelets Transform*), já que ela pode ser interpretada como a soma sobre todo o domínio de tempo da série, multiplicada por versões escalonadas e deslocadas da função de ondaletas escolhida, e produz uma série de coeficientes de ondaletas em função da escala (frequência) e localização (tempo) (ROCHA, 2009).

Desta forma, considerando  $f \in L^2(\mathbb{R})$ , pode-se definir a CWT de  $f$  com respeito à  $\psi$  como o produto interno de  $f$  pela função  $\psi_{a,b}$  dada pela expressão (3.1), ou seja,

$$(W_\psi f)(a, b) = |a|^{-1/2} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt, \quad b \in \mathbb{R}, a > 0. \quad (3.16)$$

Novamente, há algumas condições para  $\psi \in L^2(\mathbb{R})$ , como a admissibilidade, a ondaleta apresentar média zero e também a imposição de que a energia seja unitária, apresentadas, respectivamente, pelas expressões

C1.  $C_\psi := \int_{\mathbb{R}} |\hat{\psi}(s)|^2 \frac{1}{|s|} ds < +\infty$ , em que *chapéu* indica a transformada de Fourier;

C2.  $\hat{\psi}(0) = \int_{-\infty}^{\infty} \psi(s) ds = 0$ ;

C3.  $\int_{-\infty}^{\infty} |\psi(s)|^2 ds = 1$ .

A segunda e a terceira condição garantem que a ondaleta tenha alguns termos não nulos, mas que os que se distanciam de zero sejam cancelados<sup>5</sup>.

A CWT é muito importante na derivação do comportamento teórico no espaço tempo-escala. No entanto, ao se trabalhar com dados reais, é preciso alguma forma de amostragem para calcular os estimadores, por isso a necessidade de uma forma adequada de discretização. Ao utilizar transformada de ondaleta discreta de sobreposição máxima (MODWT - do inglês *Maximal Overlap Discrete Wavelets Transform*), que é uma forma especial de transformada de ondaleta discreta, restringem-se os parâmetros de escala  $j$  e translação  $k$  a apenas números inteiros (BARUNIK; VACHA, 2015).

A motivação para a utilização da MODWT se baseia essencialmente no uso de uma transformada que funcione o máximo possível como a DWT, mas que não seja afetada pela sensibilidade da DWT na escolha de um ponto de partida da série temporal, já que não utiliza o procedimento de decimação (conhecido como *downsampling*) garantindo que os vetores de ondaletas e os coeficientes de escala em todas as escalas tenham mesmo comprimento. Dessa forma, abdicando da ortogonalidade para não reduzir a amostragem, mas garantindo recursos como invariância de translação e capacidade de analisar qualquer tamanho de amostra. Além disso, a MODWT tem a propriedade de fase zero (que preserva as propriedades de fase da série temporal), que permite uma interpretação significativa do tempo em relação aos detalhes da ondaleta (KIM; IN, 2010).

<sup>5</sup> Mais detalhes sobre ondaletas e transformadas de ondaletas podem ser encontradas em Daubechies (1988), Daubechies (1992) e Gençay, Selçuk e Whitcher (2001).

Os filtros de ondaleta e escala MODWT denotados por  $g_l$  e  $h_l$ , tem comprimentos  $L$ ,  $l = 0, 1, \dots, L-1$ . Utilizando o filtro de ondaleta de Daubechies D(4) com o filtro de tamanho  $L = 4$ , há três propriedades básicas que ambos os filtros da MODWT devem cumprir. Para o filtro de ondaleta, são:

$$\sum_{l=0}^{L-1} h_l = 0, \quad \sum_{l=0}^{L-1} h_l^2 = 1/2, \quad \sum_{l=-\infty}^{\infty} h_l h_{l+2N} = 0, \quad N \in \mathbb{Z}_N, \quad (3.17)$$

enquanto para o filtro de escala MODWT, são:

$$\sum_{l=0}^{L-1} g_l = 1, \quad \sum_{l=0}^{L-1} g_l^2 = 1/2, \quad \sum_{l=-\infty}^{\infty} g_l g_{l+2N} = 0, \quad N \in \mathbb{Z}_N. \quad (3.18)$$

Com a ondaleta e os filtros apropriados, o cálculo dos coeficientes de escala e de ondaleta são obtidos via algoritmo piramidal (MALLAT, 1999) (PERCIVAL; WALDEN, 2000). Os primeiros coeficientes (tanto de escala, quanto de ondaleta) para o nível  $j = 1$  são calculados filtrando o processo  $x_t$  para  $t = 0, 1, \dots, N$  com os filtros de escala e de ondaleta MODWT dados por

$$W_{1,k} \equiv \sum_{l=0}^{L-1} h_l x_{k-l \bmod N} \quad \text{e} \quad V_{1,k} \equiv \sum_{l=0}^{L-1} g_l x_{k-l \bmod N}. \quad (3.19)$$

No segundo estágio do algoritmo da pirâmide,  $x_t$  é substituído pelos coeficientes de escala  $V_{1,k}$  e, após nova filtragem, são obtidos os coeficientes de ondaletas e de escala no segundo nível ( $j = 2$ )

$$W_{2,k} \equiv \sum_{l=0}^{L-1} h_l V_{1,k-l \bmod N}, \quad \text{e} \quad V_{2,k} \equiv \sum_{l=0}^{L-1} g_l V_{1,k-l \bmod N}. \quad (3.20)$$

E, assim, sucessivamente, obtém-se o  $j$ -ésimo nível dos coeficientes MODWT da forma:

$$W_{j,k} \equiv \sum_{l=0}^{L-1} h_l V_{j-1,k-l \bmod N} \quad \text{e} \quad V_{j,k} \equiv \sum_{l=0}^{L-1} g_l V_{j-1,k-l \bmod N}, \quad j = 1, 2, \dots, J^m, \quad (3.21)$$

em que  $J^m \leq \log_2(N)$  é o nível máximo de decomposição.

Pode-se então definir o vetor  $\mathcal{W}$  que consiste em  $J^m + 1$  subvetores de dimensão  $N$ , sendo os primeiros  $J^m$  correspondentes aos coeficientes de ondaletas da MODWT e o último,

aos coeficientes de escala da MODWT no nível  $J^m$ :

$$\mathcal{W} = [\mathbf{W}_1 \ \mathbf{W}_2 \ \dots \ \mathbf{W}_{J^m} \ \mathbf{V}_{J^m}]^T. \quad (3.22)$$

Na próxima seção, inicialmente introduzem-se noções básicas para compreensão dos modelos de difusão, conceitos como variação quadrática, variância integrada e variação de saltos e o uso da CWT para derivação da variação quadrática.

### 3.2 Decomposição da Variação Quadrática

Dada a velocidade e liquidez nos mercados financeiros, a partir do momento em que os preços são negociados, as operações tendem a ocorrer de maneira muito próxima do tempo contínuo; por isto, a literatura econométrica costuma descrever as séries de preços e retornos dos ativos como um processo de tempo contínuo (ARAÚJO, 2016).

Seja  $p_t$  o preço (ou log-preço) de um ativo financeiro no tempo  $t$ . O processo de difusão de salto em tempo contínuo tradicionalmente usado na precificação de ativos é dado pela equação diferencial estocástica

$$dp_t = \mu_t dt + \sigma_t dW_t + \xi_t dq_t, \quad (3.23)$$

em que  $W_t$  é um processo de movimento Browniano padrão. Os processos  $\mu_t$  e  $\sigma_t$  são previsíveis e representam a média e a volatilidade condicional instantânea do retorno, respectivamente, com  $\mu_t$  um processo de variação finita, enquanto  $\sigma_t$  é estritamente positivo e de quadrado integrável. O termo mais à direita, que será denotado por  $J_t$  a partir daqui, diferencia o processo com difusão de salto do processo de difusão em tempo contínuo convencional, uma vez que representa o componente de salto através de  $q_t$ , que é um processo de Poisson não correlacionado com  $W$  e governado pela intensidade de salto  $\lambda_t$ , ou seja,  $Prob(dq_t = 1) = \lambda_t dt$ , com  $\lambda_t$  positivo e finito. Essa suposição implica uma restrição comum na literatura de finanças de que só pode haver um número finito de saltos no caminho do preço por período de tempo. O fator de escala  $\xi_t$  controla a magnitude do salto no processo de preços/log-preços se ocorrer um salto no tempo  $t$ .

A expressão (3.23) também é chamada de "processo de preço/log-preço verdadeiro".

Seguindo a convenção usual de que um intervalo  $[t - h, t]$ , para  $0 \leq h \leq t \leq T$  representa a unidade de comprimento dentro de um dia e  $t = 1, 2, \dots, T$  representa cada dia da

amostra, então  $r_{t-h+(i/N)h}$  representa o  $i$ -ésimo retorno intradiário no intervalo  $[t-h, t]$ , para  $0 \leq h \leq t \leq T$ , dado por

$$r_{t-h+(i/N)h} = p_{t-h+(i/N)h} - p_{t-h+(i-1/N)h} \quad (3.24)$$

para  $i = 1, 2, \dots, N$ , onde  $N$  é o total de observações no intervalo  $[t-h, t]$ . É comum utilizar  $h = 1$ , de modo a conduzir a análise no horizonte de um dia, ou seja, neste caso  $N$  passa a ser o número de observações intradiárias.

O retorno continuamente composto ao longo dos intervalos  $[t-h, t]$ , para  $0 \leq h \leq t \leq T$ , dado por  $r_{t,h} = p_t - p_{t-h}$ . O retorno instantâneo pode ser unicamente decomposto em um componente médio previsível e integrado e uma inovação martingale local, com representação

$$r_{t,h} = r_t - r_{t-h} = \underbrace{\int_{t-h}^t \mu_s ds}_{\text{retorno médio esperado}} + \underbrace{\int_{t-h}^t \sigma_s dW_s}_{\text{Variância integrada}} + \sum_{t-h \leq l \leq t} J_l. \quad (3.25)$$

O interesse, em geral, está na estimação da Variância Integrada na expressão (3.25). No entanto, na prática, com dados observados, o preço/log-preço do ativo está latente, uma vez que o mesmo está contaminado por saltos e ruídos de microestrutura, de forma que é assumido que o preço/log-preço latente do ativo segue um processo padrão de difusão de salto contaminado com ruído de microestrutura. Seja  $y_t$  os log-preços observados evoluindo entre  $0 \leq t \leq T$ . Então, o processo de preço observado é dado por

$$y_t = p_t + \varepsilon_t, \quad (3.26)$$

considerando ruído de microestrutura  $\varepsilon_t$ , independente e identicamente distribuído, com média zero e variância  $\eta^2$ .

Neste caso, o processo de variação do retorno quadrático (chamada simplesmente de Variação Quadrática) ao longo do intervalo  $[t-h, t]$ , para  $0 \leq h \leq t \leq T$  associada ao processo de preço/log-preço  $p_t$  é dado pela soma da variância integrada difusiva e dos saltos quadrados acumulados:

$$QV_{t,h} = \underbrace{\int_{t-h}^t \sigma_s^2 ds}_{IV_{t,h}} + \underbrace{\sum_{t-h \leq l \leq t} J_l^2}_{JV_{t,h}}. \quad (3.27)$$

A Variação Quadrática  $QV_{t,h}$  associada ao processo de preço/log-preço  $p_t$  definida em (3.27) é uma medida natural da variabilidade do processo de preço/log-preço e é naturalmente decomposta em duas partes: a Variância Integrada do processo de preço latente e a Variação de Salto (usualmente, e também neste trabalho, denotados por  $IV_{t,h}$  e  $JV_{t,h}$ , do inglês *Integrated Variance* e *Jump Variation*, respectivamente) (ANDERSEN et al., 2001c) (BARNDORFF-NIELSEN; SHEPHARD, 2002b).

Observe que as inovações no componente  $\mu_t$  não afetam a variação do caminho da amostra do retorno, uma vez que a média é de ordem inferior em termos de propriedades de segunda ordem das inovações difusivas de  $\sigma_t dW_t$ . Desta forma, quando acumulados em muitos retornos de alta frequência em um curto intervalo de tempo de comprimento  $h$ , eles podem ser efetivamente desprezados (ANDERSEN; TERÄSVIRTA, 2009).

### 3.3 Estimação da Variação de Salto via Ondaletas

O interesse é decompor a Variação Quadrática no componente de Variação Integrada e Variação de Salto. Utilizando a transformada de ondaleta contínua (CWT), obtém-se

$$QV_{t,h} = \underbrace{\frac{2}{C_\psi} \int_{t-h}^t \int_0^\infty \int_{\mathbb{R}} \psi_{j,k}(s) \langle \psi_{j,k}, \sigma_s^2 \rangle dk \frac{1}{j^2} dj ds}_{IV_{t,h}} + \underbrace{\sum_{t-h \leq l \leq t} J_l^2}_{JV_{t,h}}, \quad (3.28)$$

onde  $\langle \psi_{j,k}, \sigma_s^2 \rangle$  é equivalente à expressão em (3.16).

A equação (3.28) decompõe a Variação Quadrática tanto no tempo quanto na frequência. Por decomposição no domínio da frequência obtemos  $j$  componentes que representam escalas que podem ser vistas como horizontes de investimento. Além disso, permite definir um modelo livre de medida da Variação Integrada em analogia ao estimador de Variância Realizada original.

A CWT permite a derivação do comportamento teórico no espaço tempo-escala, no entanto, ao se trabalhar com dados reais, se faz necessária alguma forma de discretização para o cálculo dos estimadores, que é feita, em geral, através da MODWT.

Para a estimativa da Variação de Saltos através da abordagem de ondaletas, Barunik, Krehlik e Vacha (2016) seguem conforme os resultados de Wang (1995) e a extensão de Fan e Wang (2007).

A ideia inicial é distinguir entre a parte contínua e a parte de saltos do processo de preço estocástico com ruído aditivo independente e identicamente distribuído. Para isto, assume-se

que o caminho da amostra do processo de preço da equação (3.26) terá um número finito de saltos.

Considere a transformada de ondaletas da série de log-preços (ou simplesmente, preços). Fan e Wang (2007) afirmam que para frequências mais altas desta transformada, o processo de  $p_t$  domina o ruído  $\varepsilon_t$  apenas próximo de um local de salto, do contrário é muito pequeno (ou seja, o ruído é relativamente pequeno e a dominância do preço é devido ao salto). Para detectar a parte dominante do processo  $y_t$ , o limite universal (*universal threshold*) de Donoho e Johnstone (1994) é o mais comumente utilizado na literatura, e portanto foi aplicado também neste trabalho.

A partir dos filtros de escala e de ondaleta MODWT, o cálculo dos coeficientes de escala e de ondaleta ( $d_{j,k}$ ) são obtidos via algoritmo piramidal (PERCIVAL; WALDEN, 2000) para a série de preços ou log-preços  $y_{t-h+(i/N)h}$ , para  $i = 1, 2, \dots, N$ , no intervalo  $[t - h, t]$ . Desta forma, os coeficientes  $d_{j,k}$  são obtidos para cada nível  $j$  da decomposição, em que  $J^m \leq \log_2 N$  é o nível máximo de decomposição para a série estudada.

Como foi utilizada a MODWT, existem  $N$  coeficientes de ondaletas em cada nível, que corresponde ao número de observações intradiárias. Assim, para cada nível  $j$  fixado, a identificação do salto é feita quando o valor absoluto dos coeficientes de ondaleta no respectivo nível é maior que o limite universal  $D_j(t)$ , ou seja, se  $|d_{j,k}| > D_j(t)$ , para cada valor de  $k$ ,  $k = 1, \dots, N$  para cada dia  $t = 1, 2, \dots, T$ , onde o limite  $D_j(t)$  é dado por

$$D_j(t) = s_j(t) \sqrt{2 \log N}, \text{ com } s_j(t) = \sqrt{2} \text{Mediana} \{|d_{j,k}|\} / 0,6745, \quad k = 1, \dots, N. \quad (3.29)$$

O fator  $\sqrt{2}$  em  $s_j(t)$  não aparece no original de Fan e Wang (2007). Barunik, Krehlik e Vacha (2016) o introduzem como um fator de correção devido ao uso da MODWT.

Quando há a presença de salto, a localização de salto é estimada por  $\hat{\tau}_l = k$ . Desta forma, para cada valor de  $k$ , o tamanho do salto  $\Delta_k J_t$  é estimado como

$$\Delta_k J_t = \left( y_{t-h+(\frac{k}{N})h} - y_{t-h+(\frac{k-1}{N})h} \right) \mathbb{I}_{\{|d_{j,k}| > D_j(t)\}}, \quad k = 1, \dots, N. \quad (3.30)$$

Neste caso, o tamanho do salto está sendo estimado especificamente pelo retorno. Ou seja, não se considera informações do preço à frente do salto localizado, apenas a imediatamente anterior. A Variação de Salto  $\widehat{JV}_{t,h}$  sobre o intervalo  $[t - h, t]$  no tempo discreto é estimada,

portanto, como a soma dos quadrados de todos os tamanhos de salto estimados

$$\widehat{JV}_{t,h} = \sum_{k=1}^N (\Delta_k J_t)^2. \quad (3.31)$$

O estimador dado em (3.31) é consistente da Variação de Salto do processo, ou seja,  $\widehat{JV}_{t,h} \xrightarrow{p} JV_{t,h}$  (BARUNIK; KREHLIK; VACHA, 2016).

Após detectar precisamente os saltos, é possível obter o processo de preço observado ajustado aos saltos ao se subtrair os saltos intradiários do processo a partir da série de retornos da seguinte forma

$$\Delta_k y_t^{(J)} = \Delta_k y_t - \Delta_k J_t, \quad (3.32)$$

em que,  $\Delta_k y_t = y_{t-h+(\frac{k}{N})h} - y_{t-h+(\frac{k-1}{N})h}$  é o  $k$ -ésimo retorno no intervalo  $[t-h, t]$ , para  $k = 1, \dots, N$ ,  $N$  total de observações intradiárias.

O processo ajustado aos saltos  $y_{t,h}^{(J)}$  converge em probabilidade para a parte contínua sem saltos, ou seja, para a Variância Integrada e pode ser utilizado então para estimação e previsão da volatilidade.

Conforme discutido anteriormente, é comum utilizar apenas o(s) primeiro(s) nível(is) de decomposição para detecção de saltos. Com o objetivo de expandir a análise para os demais níveis da decomposição e de verificar a contribuição dos saltos detectados para a composição da Variação de Salto estimada  $\widehat{JV}_{t,h}$ , a metodologia discutida será aplicada em diversos níveis  $j$  da decomposição de ondaletas MODWT. Fan e Wang (2007) sugerem um nível limite para ser explorado ao afirmar que, em níveis de alta resolução  $j_n$  com  $2^{j_n} \approx n/\log^2 n$ , os coeficientes de ondaletas são significativamente maiores do que os outros próximo aos locais de salto do processo de log-preço. Isto é concordante com o fato de que, em níveis de baixa resolução, as informações obtidas são mais relacionadas à tendência da série.

Uma outra abordagem para estimação da Variação Quadrática é dada a partir da teoria de medidas realizadas, apresentadas na próxima seção.

### 3.4 Medidas Realizadas da Volatilidade

Um simples estimador consistente da Variação Quadrática global sob a suposição de não contaminação de ruído no processo de preço/log-preço é fornecido pela conhecida Variância Realizada, ou Volatilidade Realizada, (RV) introduzida por Andersen e Bollerslev (1998a). A

variância realizada sobre  $[t - h, t]$ , para  $0 \leq h \leq t \leq T$ , pode ser estimada como

$$\widehat{RV}_{t,h} = \sum_{i=1}^N r_{t-h+(\frac{i}{N})h}^2 \quad (3.33)$$

em que  $N$  é o número de observações no intervalo  $[t - h, t]$  e  $r_{t-h+(\frac{i}{N})h}$  é o  $i$ -ésimo retorno intradiário no intervalo  $[t - h, t]$ .

O estimador em (3.33) converge em probabilidade para a Variação Quadrática, ou seja, para  $IV_{t,h} + JV_{t,h}$  quando  $N \rightarrow \infty$ <sup>6</sup>.

É usual considerar transformações de RV como  $\sqrt{RV}$  e  $\log(RV)$  e a notação RV é utilizada tanto para Variância Realizada, quanto para Volatilidade Realizada, assim como IV, para Variância Integrada e Volatilidade Integrada (SHIN, 2018). A mesma notação foi adotada no presente trabalho.

Dado que, na prática, em dados reais, o processo de preço/log-preço observado  $y_t$  é contaminado com ruído e saltos, é natural concluir que o principal objetivo se concentre na parte referente à Variância Integrada da Variação Quadrática. Barunik e Vacha (2015) cita uma ampla literatura subsequente que mostra a importância de se considerar os saltos e a presença de ruídos na estimação da Volatilidade Realizada.

Em frequências de amostragem muito altas, as observações de preços são contaminadas pelo ruído da microestrutura causado, entre outros casos, pelo *bid-ask spread*, que é a diferença entre a melhor oferta de venda e melhor oferta de compra. Ele induz um viés na medida de RV, que aumenta progressivamente, conforme a frequência de amostragem é aumentada (BANDI; RUSSELL, 2006) (BANDI; RUSSELL; YANG, 2008) (ZHOU, 1996). Para aliviar este problema, a abordagem geral é amostrar os preços em frequências mais baixas, normalmente de 5 a 30 minutos, o que, por outro lado, acaba por acarretar perda de informação, de maneira que, apesar da redução do ruído de microestrutura, há também uma perda significativa de eficiência (SHARMA, 2016).

Outro problema com a estimativa da RV é o viés induzido pelos saltos de preços, uma vez que representam as incertezas inerentes ao mercado financeiro. Como refletem a chegada de notícias impactantes ou negociações de ordens de tamanho anormal, a projeção da Variação de Saltos é complicada, o que leva a um aumento no interesse em decompor a Variação Quadrática

<sup>6</sup> Diversos trabalhos abordam esta convergência, por exemplo Andersen e Bollerslev (1998a); Andersen et al. (2001c); Andersen et al. (2003); Barndorff-Nielsen e Shephard (2002a); Barndorff-Nielsen e Shephard (2002b).

no componente de Variação Integrada e a Variação de Salto a fim de melhorar estas projeções. Estas duas fontes de variação no processo de preços diferem substancialmente e impactam a volatilidade futura de diferentes maneiras (BARUNIK; KREHLIK; VACHA, 2016).

Para contornar este problema, Barndorff-Nielsen e Shephard (2004) desenvolveram o estimador de Variação BiPower (BPV - *BiPower Variation*), que é um dos mais comumente utilizados e também considerado referência, visto que pode detectar a presença de saltos em dados de alta frequência. Este estimador permite suavizar o efeito de um salto em uma única observação de um retorno ao multiplicá-lo pelo retorno subsequente, caso este não seja afetado por um salto em si (diferentemente da RV tradicional, na qual o retorno é multiplicado por ele mesmo). Esta simples troca torna o estimador BPV consistente na presença de saltos. Desta forma, a ideia principal do estimador BPV é comparar duas medidas da variância, uma contendo a Variação de Salto e a outra robusta a saltos, contendo, portanto, apenas a parte referente à IV.

Complementarmente, Andersen et al. (2001a) propõem um ajuste do estimador original de Barndorff-Nielsen e Shephard (2004), que ajuda a torná-lo robusto para certos tipos de ruído de microestrutura; ao invés de multiplicar dois retornos subsequentes, os toma de maneira intercalada. A Variação BiPower sobre  $[t - h, t]$  pode então ser definida por:

$$\widehat{BPV}_{t,h} = \mu_1^{-2} \frac{N}{N-2} \sum_{i=3}^N \left| r_{t-h+(\frac{i-2}{N})h} \right| \cdot \left| r_{t-h+(\frac{i}{N})h} \right|, \quad (3.34)$$

em que  $\mu_a = \pi/2 = E(|Z|^a)$ , e  $Z \sim N(0, 1)$ ,  $a \geq 0$ .

O estimador (3.34) é um estimador consistente da Variância Integrada, ou seja,  $\widehat{BPV}_{t,h} \xrightarrow{p} \int_{t-h}^t \sigma_s^2 ds$ .

Assim, como  $\widehat{RV}_{t,h}$  fornece um estimador consistente da Variância Quadrática ( $IV + JV$ ), a Variação de Salto pode ser estimada consistentemente como a diferença entre a Variância Realizada e a Variação BiPower Realizada

$$\left( \widehat{RV}_{t,h} - \widehat{BPV}_{t,h} \right) \xrightarrow{p} JV_{t,h}. \quad (3.35)$$

O estimador  $\widehat{BPV}_{t,h}$  é uma referência na detecção de saltos. Várias medidas realizadas foram propostas para eliminar a interferência de ruídos de microestrutura e saltos na estimação da volatilidade. Algumas delas estão listadas no Quadro 3.1.

Quadro 3.1 – Resumo dos estimadores de volatilidade.

Autores	Estimador	Saltos	Ruídos
Andersen et al. (2003)	rCov		
Barndorff-Nielsen et al. (2004)	rKernelCov		X
Barndorff-Nielsen e Shephard (2004)	BPV	X	
Zhang, Mykland e Aït-Sahalia (2005)	rTSCCov		X
Zhang, Mykland e Aït-Sahalia (2005)	rAVGCov		X
Mancini e Gobbi (2012)	rThresholdCov		X
Boudt, Croux e Laurent (2011)	rOWCov	X	
Andersen, Dobrev e Schaumburg (2012)	medRV	X	
Andersen, Dobrev e Schaumburg (2012)	minRV	X	
Boudt e Zhang (2015)	rRTSCov	X	X

Fonte: Adaptado de Araújo (2016).

### 3.5 Modelos da família HAR

Andersen et al. (2003) mostraram que a modelagem direta de séries temporais da Volatilidade Realizada supera fortemente, em termos de previsão fora da amostra, os populares modelos GARCH e de volatilidade estocástica. De fato, desde que foi proposto por Corsi (2009), o modelo de volatilidade realizada autorregressivo heterogêneo (HAR-RV - *Heterogeneous AutoRegressive Realized Volatility*) para previsão de Volatilidade Realizada é um modelo chave, a partir do qual diversas modificações surgiram.

Uma das vantagens é que o modelo HAR é parcimonioso: apesar de poder ser considerado um modelo autorregressivo (AR) de ordem 22, já que emprega volatilidade realizada mensalmente (que corresponde a 22 dias úteis), ele o faz de maneira restrita a apenas três componentes, que aborda os comportamentos de volatilidade dos investidores de curto, médio e longo prazo, através de componentes de horizonte diário, semanal e mensal. Desta forma, o modelo HAR é dado por

$$y_{t+1} = \beta_0 + \beta_1 y_t^{(d)} + \beta_2 y_t^{(w)} + \beta_3 y_t^{(m)} + a_{t+1}, \quad (3.36)$$

em que  $y_t$  é alguma medida realizada definida (como a RV ou BPV definidas anteriormente),  $a_t$  é a série de erros da regressão e as medidas  $y_t^{(d)}$ ,  $y_t^{(w)}$  e  $y_t^{(m)}$  são, respectivamente, as médias móveis diária, semanal e mensal da medida realizada<sup>7</sup>.

<sup>7</sup> A notação  $y_t^{(d)}$ ,  $y_t^{(w)}$  e  $y_t^{(m)}$  é usual da literatura devido aos horizontes em inglês:  $d$  para *daily* (diário),  $w$  para *weekly* (semanal) e  $m$  para *monthly* (mensal).

Como para o mercado financeiro a semana é considerada com 5 dias de negociação e o mês com 22, tem-se que

$$y_t^{(d)} = y_t; \quad y_t^{(w)} = \frac{1}{5} (y_t + \dots + y_{t-4}); \quad y_t^{(m)} = \frac{1}{22} (y_t + \dots + y_{t-21}). \quad (3.37)$$

A construção do modelo desta maneira garante que os coeficientes  $\beta_1$ ,  $\beta_2$  e  $\beta_3$ , correspondentes aos horizontes diário, semanal e mensal, respectivamente, tenham uma interpretação econômica clara em termos de uma estrutura de previsão.

As características de memória longa persistente, comum nestas séries de alta frequência, são capturadas pela estrutura do *lag 22* em um modelo parcimonioso com apenas 3 parâmetros autorregressivos, além do que a soma dos parâmetros  $\beta_1 + \beta_2 + \beta_3$  ser próxima de 1 (CORSI, 2009).

Considerando a Volatilidade Realizada (que pode ser  $\sqrt{RV}$  e  $\log(RV)$ ) para o ajuste do modelo (3.36), obtém-se o modelo HAR-RV tradicional:

$$RV_{t+1} = \beta_0 + \beta_1 RV_t + \beta_2 RV_{t-5,t} + \beta_3 RV_{t-22,t} + a_{t+1}. \quad (3.38)$$

Algumas variações do modelo (3.36) que consideram o efeito dos saltos em suas modelagens são encontradas na literatura. O primeiro modelo a ser considerado, o HAR-J, estima a parte suave da variação de preço, a partir de uma medida realizada robusta a saltos, na qual a estimação de salto é dada por  $\hat{J}_t = \max [RV_t - BPV_t, 0]$ :

$$RV_{t,t+h}^{(J)} = \beta_0 + \beta_1 RV_t + \beta_2 RV_{t-5,t} + \beta_3 RV_{t-22,t} + \beta_4 J_t + \beta_5 J_{t-5,t} + \beta_6 J_{t-22,t} + \varepsilon_{t,t+h}, \quad (3.39)$$

no qual os componentes de salto podem ser considerados em horizontes diário, semanal e mensal, de maneira análoga a apresentada em (3.37) para as medidas realizadas.

Uma variação do modelo (3.39), o HAR-CJ proposto por Andersen, Bollerslev e Diebold (2007), considera que pequenos saltos podem estar associados a erros de medição ou até mesmo fazem parte do processo de variação contínua do caminho amostral, associando apenas valores grandes de  $J_t = \max [RV_t - BPV_t, 0]$  ao componente de salto a ser estimado. Por isto, separa a componente Contínua da componente de Salto. De maneira simplificada, o HAR-CJ substitui o estimador não robusto por um estimador robusto de salto quando um teste de salto mostra que

ocorreu um salto.

$$RV_{t,t+h} = \beta_0 + \beta_1 C_t + \beta_2 C_{t-5,t} + \beta_3 C_{t-22,t} + \beta_4 J_t + \beta_5 J_{t-5,t} + \beta_6 J_{t-22,t} + \varepsilon_{t,t+h}. \quad (3.40)$$

No próximo capítulo, encontram-se organizados de que maneira cada um dos conceitos abordados até aqui foram utilizados nesta tese e detalhes da amostra na qual foi aplicada a metodologia desenvolvida.

## 4 MATERIAIS E MÉTODOS

Para a aplicação da localização de saltos nos diversos níveis de decomposição foi utilizada a série de preços da PETR4 - Petrobrás (Petróleo Brasileiro S/A). A escolha deste ativo se deu devido à sua alta liquidez, grande participação e influência sobre o Ibovespa (principal indicador de desempenho das ações negociadas na bolsa de valores brasileira, que reúne as empresas mais importantes do mercado de capitais brasileiro). Outro fato que motivou a escolha deste ativo se deu a partir do objetivo de explorar a identificação de saltos de preços: a empresa passou por um período evidente de queda nos preços em fevereiro de 2021, devido ao anúncio da troca da presidência da mesma feita pelo então presidente do país. A notícia impactou não apenas o próprio ativo da Petrobrás, como também o Ibovespa, o que resultou em um grande salto no dia imediato e nos posteriores ao anúncio.

Os dados foram coletados via MetaTrader 5 no período correspondente à 04/01/2021 a 25/06/2021. A frequência utilizada foi de 1 minuto, já que ruídos, como a presença de saltos na série, tendem a ser maiores em alta frequência. A utilização de amostras maiores em uma frequência tão alta teria um alto impacto no número de níveis a serem analisados e, portanto, ficará a cargo de uma análise futura.

É importante salientar que há uma grande dificuldade na obtenção de dados brutos de séries intradiárias, os chamados dados “*tick-by-tick*”<sup>8</sup>. Desta forma, optou-se por tomar a frequência mais alta possível em intervalos regulares de 1 minuto.

A fim de evitar os efeitos do leilão de abertura e do preço *overnight* e considerando também a alta taxa de dados faltantes e presença de *outliers* nos minutos iniciais de abertura do mercado, a amostragem foi iniciada a partir das 10h20, desconsiderando os primeiros 19 minutos.

Outra questão importante referente aos dados se dá pelo fato de que a bolsa de valores brasileira tem seu horário de fechamento influenciado pelo da bolsa de valores americana, estendendo em 1 hora o horário de fechamento em determinados períodos do ano. Para fins de identificação dos saltos, não foi necessário desconsiderar tais observações.

Após o tratamento dos dados, a decomposição da série de log-preços é feita através da MODWT, utilizando o filtro Haar.

---

<sup>8</sup> Os dados “*tick-by-tick*” ocorrem em instantes aleatórios de tempo e podem ser amostrados em intervalos de tempo regulares ou até mesmo utilizando alguma modalidade de agregação presente na literatura.

A partir dos coeficientes obtidos na decomposição de ondaletas, prosseguiu-se para a detecção dos locais de saltos a partir do limite universal apresentado na expressão (3.29) para cada nível de decomposição. Como o limite de Donoho e Johnstone é fixado para o horizonte pré-definido, optou-se por utilizar o horizonte diário, fixando  $h = 1$ , de modo que corrobora com o horizonte da volatilidade realizada que é calculada para cada dia da amostra.

Com as localizações dos saltos determinadas, estima-se os tamanhos dos saltos identificados (conforme expressão (3.30)), cujas somas quadradas retornarão a estimação da variação de salto para cada nível, dada por (3.31). A soma das Variações de Salto por nível fornece a Variação de Salto total estimada.

Por fim, para se obter uma nova série de dados, dada pelos retornos ajustados aos saltos localizados em cada nível, prossegue-se para a eliminação dos saltos conforme a expressão (3.32).

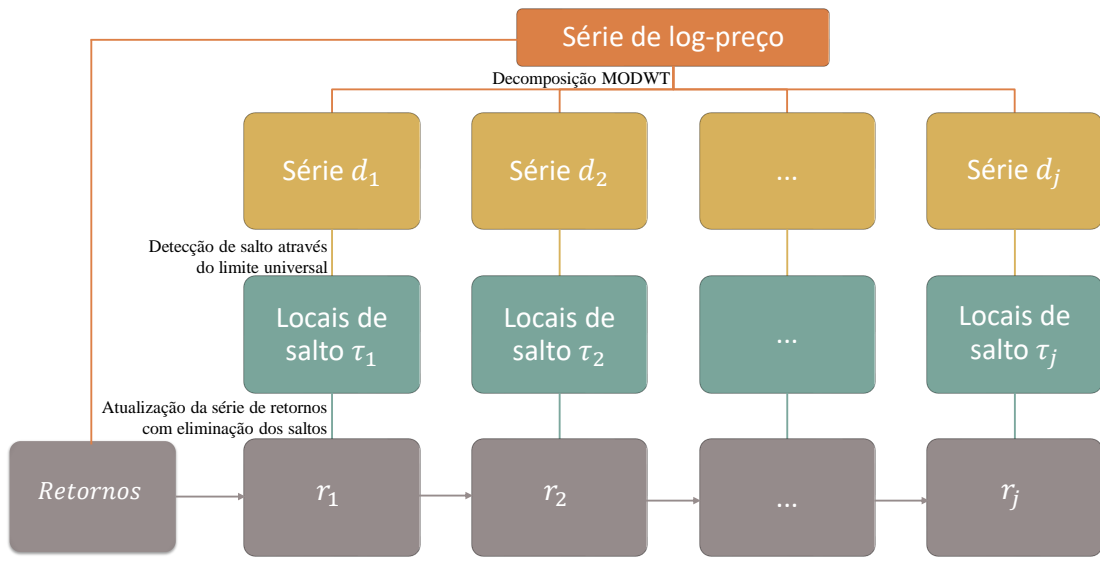
A Figura 4.1 apresenta o esquema que descreve as etapas descritas e as etapas da metodologia estão listadas a seguir:

- a) A série de log-preços é decomposta em  $j$  níveis de escala temporal, através da MODWT, com o filtro Haar;
- b) Para cada nível  $j$ :

Para cada dia  $t$ ,  $0 < t < T$ , com  $N$  observações intradiárias:

- os coeficientes  $d_{j,k}$ ,  $k = 1, 2, \dots, N$ , são comparados ao limite universal  $D_j(t)$ ,  $t = 1, 2, \dots, T$ ;
  - se  $|d_{j,k}| > D_j(t)$  então um salto é localizado e sua localização  $\tau_k$  é armazenada;
  - calcula-se o tamanho do salto  $\Delta_k J_t$  e a variação de salto do dia  $JV_t$  pela soma de todos os saltos detectados;
  - estima-se a Variação de Salto por nível a partir da soma de todos os saltos detectados no nível;
  - a série  $r_j$  é criada a partir dos saltos intradiários detectados no nível  $j$ ;
- c) A Variação de Salto  $JV^{(W)}$  da amostra é dada portanto pela soma das variações de saltos por nível ( $JV_{d_j}$ ).

Figura 4.1 – Obtenção de séries de retornos atualizadas com a eliminação dos saltos obtidos a cada nível.



onde  $r_j$  é dado pela série de retornos subtraindo-se os saltos localizados em  $d_1, d_2, \dots, d_j$ .

Fonte: do autor (2022).

Com a série de retornos original e as séries ajustadas aos saltos detectados nos diversos níveis da decomposição, é possível comparar o ajuste de diversos modelos de volatilidade realizada para estas diferentes séries. Os modelos da família HAR considerados são o original HAR-RV e dois outros que consideram a decomposição em componente discreta e contínua, HAR-J e HAR-CJ. A qualidade de ajuste dos diferentes modelos é comparada através da medida de  $R^2$ .

As análises foram realizadas no *software* R (R Core Team, 2020). Para a decomposição de ondaletas, o pacote utilizado foi *waveslim* (WHITCHER, 2020), enquanto para o cálculo das medidas realizadas e ajuste dos modelos da família HAR, o pacote utilizado foi *highfrequency* (BOUDT et al., 2020).

## 5 RESULTADOS E DISCUSSÃO

A aplicação empírica se baseou na análise da série de preços das ações da PETR4, na frequência de 1 minuto. A série amostrada no período de 04/01/2021 a 25/06/2021 compreendeu 119 dias de operação da bolsa de valores, resultando em 49611 observações após o tratamento dos dados. O número de observações por dia varia entre 285 (dia de Quarta-Feira de Cinzas, com meio expediente na bolsa de valores brasileira) a 455 (nos dias com horário de fechamento estendido).

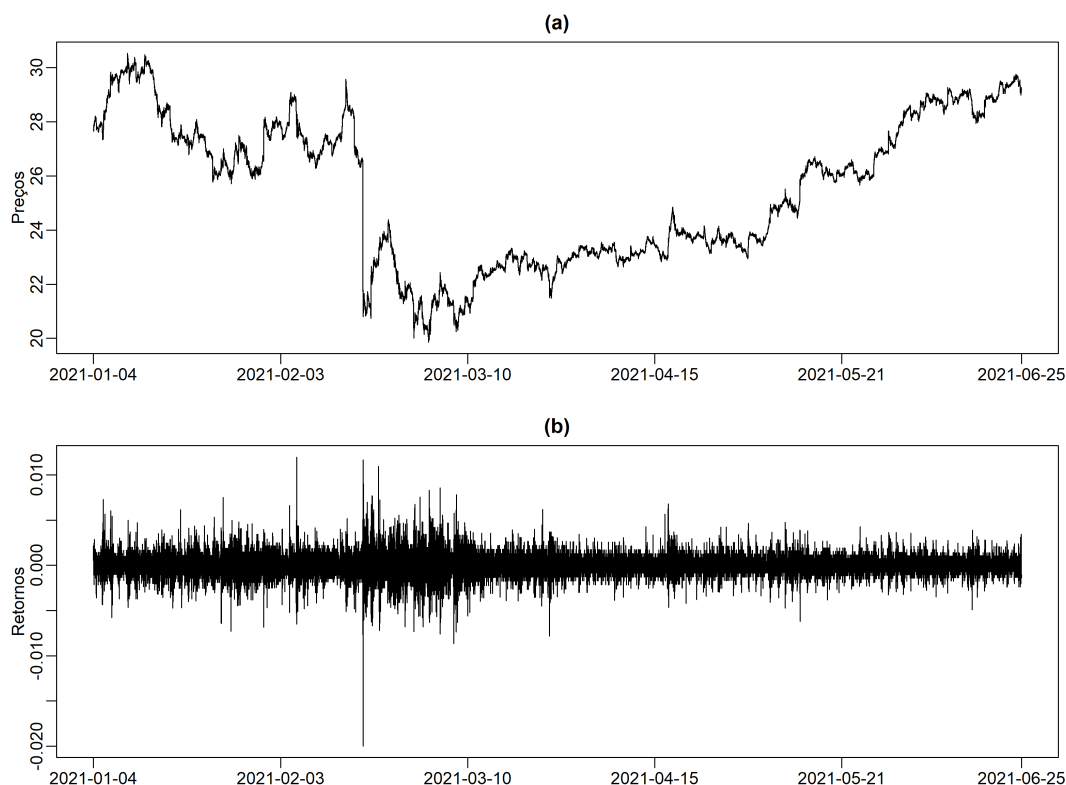
A Figura 5.1 apresenta o gráfico da série original com os preços praticados no período de coleta para a frequência de 1 minuto, além do gráfico da respectiva série de retornos. Estes gráficos evidenciam o período de queda que foi explorado neste trabalho ocorrido no dia 22/02/2021. O histograma e QxQ plot apresentados na Figura 5.2 demonstram uma característica particular dos retornos intradiários de possuírem caudas pesadas e não serem propriamente normalmente distribuídos. Este é um fato estilizado das séries intradiárias, cuja normalidade é dada sob a hipótese de ausência de saltos e ruídos de microestrutura<sup>9</sup>.

Como o número máximo de níveis de decomposição de ondaletas relaciona-se com o tamanho da amostra, foi possível obter 15 níveis de detalhes. A Tabela 5.1 apresenta a relação entre os níveis de decomposição da MODWT e as escalas de tempo correspondentes para uma amostra de frequência de 1 minuto. O nível  $d_1$  relaciona-se com o horizonte de tempo de 1 a 2 minutos de negociação; o nível  $d_2$ , com o de 2 a 4 minutos; e, assim, sucessivamente, de forma que, os níveis  $d_7$  a  $d_9$  relacionam-se com horizontes em frequências de horas de negociação; e, a partir do  $d_{10}$ , horizontes que correspondem a dias de negociação.

Os gráficos dos coeficientes de ondaletas obtidos a partir da transformada MODWT da série de log-preços para os 15 níveis de decomposição está apresentado na Figura 5.3. É possível notar que o choque do dia 22/02/2021 é refletida ao longo de quase todos os níveis de decomposição. Nos níveis de frequência mais baixa ( $d_{12} - d_{15}$ ), os coeficientes da decomposição fornecem informações relacionadas ao comportamento de tendência da série no período estudado.

<sup>9</sup> Para mais detalhes a respeito dos saltos estilizados das séries de alta frequência, pode-se consultar Araújo Júnior (2020).

Figura 5.1 – Gráficos da série de preços (a) e dos retornos (b) das ações da PETR4 no período de 04/01/2021 a 25/06/2021 para a frequência de 1 minuto.

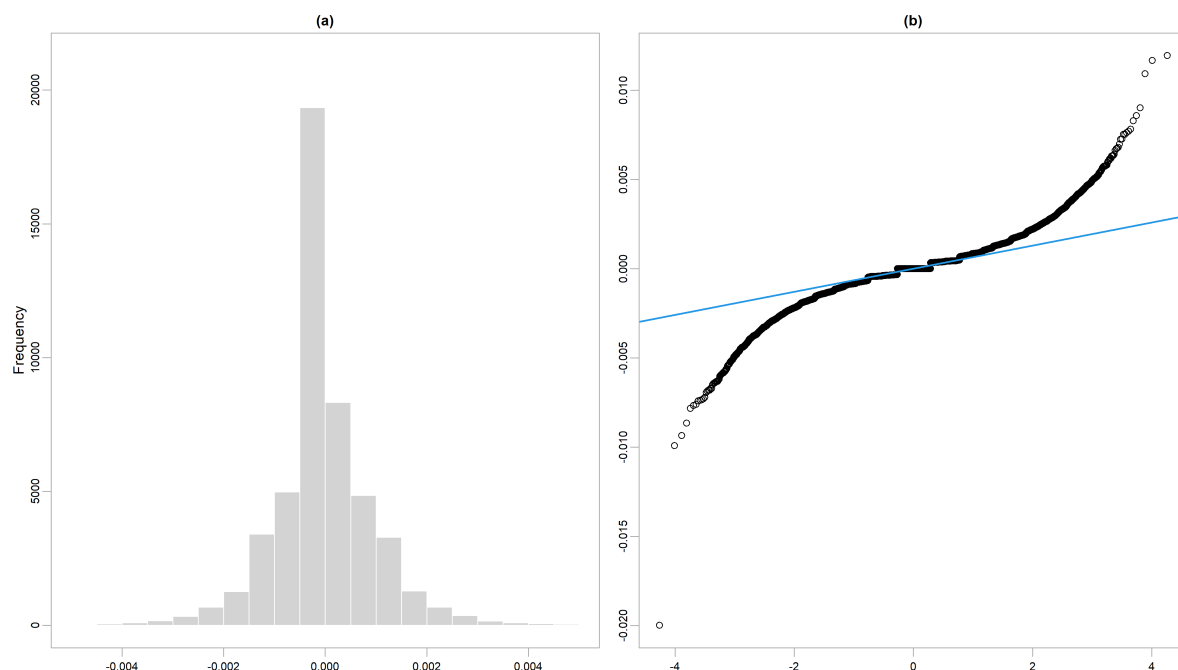


Fonte: do autor (2022).

## 5.1 Variação de Saltos

Os saltos são então localizados conforme a metodologia descrita no Capítulo 4. A Figura 5.4 ilustra o processo para o dia 13/01/2021, um dos dias com o maior número de saltos detectados, considerando o primeiro nível de decomposição. A imagem (a) apresenta a série  $d_1$  (série dos coeficientes do primeiro nível da decomposição da série de log-preços); a imagem (b), destaca em vermelho o limite  $D_j(t)$  e é possível observar as ocorrências que são consideradas saltos a partir dos valores da série de valor absoluto dos coeficientes  $d_1$  que excedem tal limite. As localizações destes valores são, portanto, os locais de salto detectados que estão destacados no gráfico da série de log-preço apresentada em (c). Para este dia específico, foram observados 11 saltos, precisamente às 10h20, 11h47, 13h17, 14h01, 14h10, 14h18, 14h37, 14h40, 14h56, 15h, 15h54. Uma vez encontradas as localizações dos saltos, o passo seguinte é calcular o tamanho dos saltos identificados, a fim de estimar a Variação de Salto e, assim, obter uma nova série dada pelos retornos ajustados aos saltos localizados neste primeiro nível da decomposição.

Figura 5.2 – Histograma (a) e QxQ plot (b) da série de retornos das ações da PETR4 no período de 04/01/2021 a 25/06/2021 para a frequência de 1 minuto.



Fonte: do autor (2022).

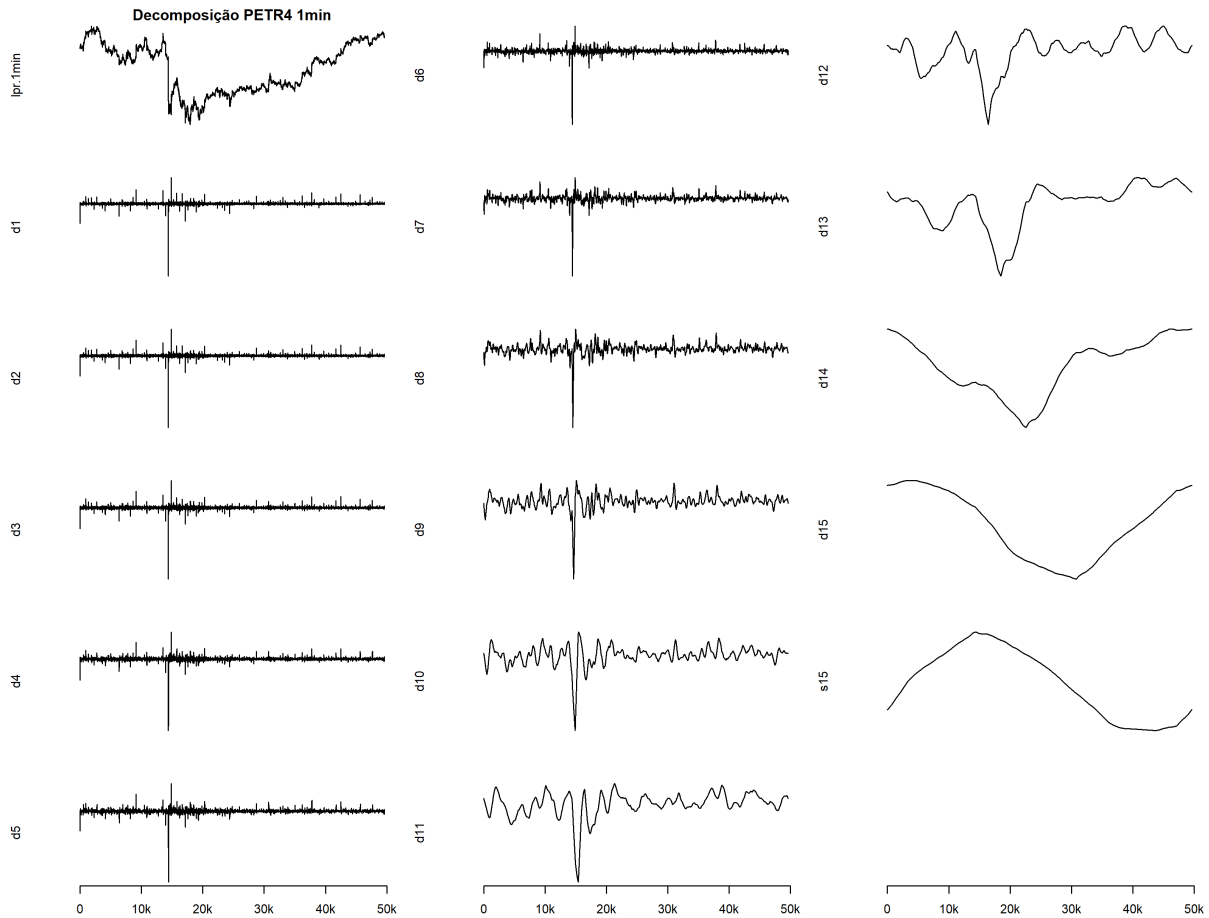
Tabela 5.1 – Relação entre nível de decomposição e Escala de tempo.

Nível	Escala de tempo
$d_1$	1 – 2 minutos de negociação
$d_2$	2 – 4 minutos de negociação
$d_3$	4 – 8 minutos de negociação
$d_4$	8 – 16 minutos de negociação
$d_5$	16 – 32 minutos de negociação
$d_6$	32 – 64 minutos de negociação
$d_7$	$\approx$ 1 – 2 horas de negociação
$d_8$	$\approx$ 2 – 4 horas de negociação
$d_9$	$\approx$ 4 – 8 horas de negociação
$d_{10}$	$\approx$ 1 – 2 dias de negociação
$d_{11}$	$\approx$ 2 – 4 dias de negociação
$d_{12}$	$\approx$ 4 – 8 dias de negociação
$d_{13}$	$\approx$ 8 – 17 dias de negociação
$d_{14}$	$\approx$ 17 – 34 dias de negociação
$d_{15}$	$\approx$ 34 – 68 dias de negociação

Fonte: adaptado de Crowley (2007).

A série de retornos ajustada aos saltos detectados em  $d_1$  para o dia 13/01/2021 está apresentada em (d), juntamente com os saltos detectados e eliminados.

Figura 5.3 – Gráficos da série de log-preços original e dos coeficientes de ondaletas dos 15 níveis de decomposição.

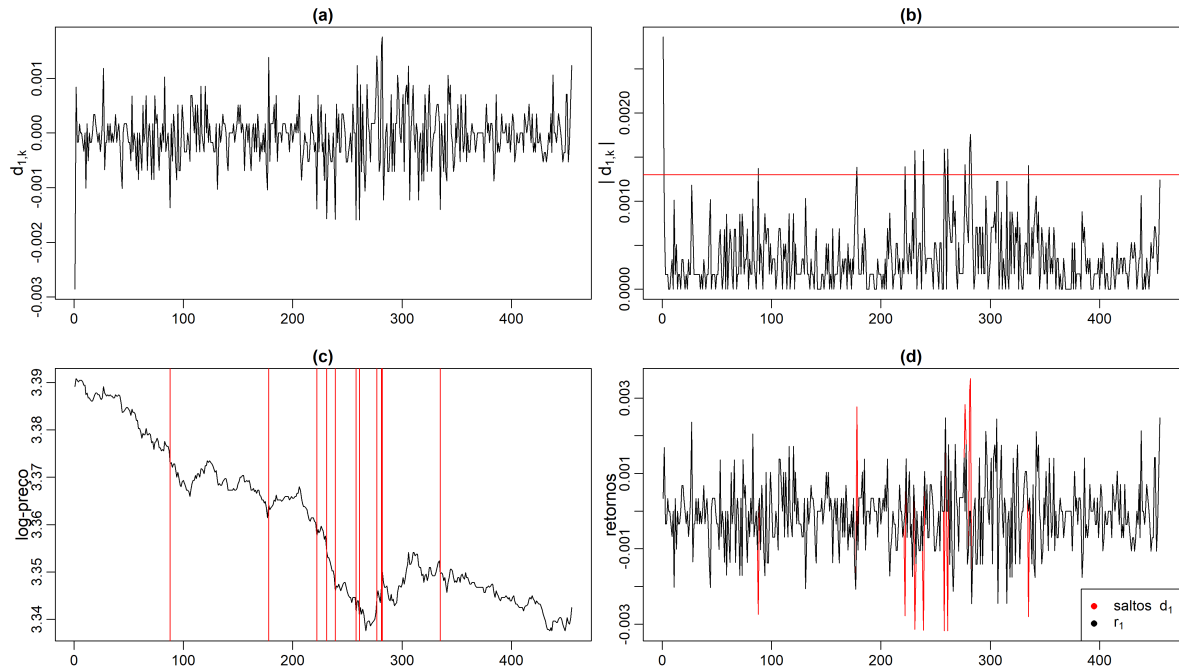


Fonte: do autor (2022).

De maneira análoga, a localização de saltos foi realizada também para os demais níveis de escala. A ideia é identificar os saltos que aparecem até o nível  $j_{max} = 8$ , no entanto, como o objetivo é explorar a presença de saltos nos diversos níveis, optou-se por avaliar também os níveis posteriores. Na Tabela 5.5 estão apresentados, para cada nível, a quantidade de saltos detectadas por dia e, entre parênteses, quantos (dos 119 dias amostrados) apresentaram aquela quantidade de saltos. É importante observar que, à medida que se caminha para frequências mais baixas, o número de dias sem salto aumenta e há menos dias com poucos saltos por dia, no entanto, alguns dias apresentam uma quantidade maior de saltos. Há, por exemplo, dias com 20, 40, 70 saltos nos níveis 4, 5, 6 e 7. Este maior volume de saltos pode estar relacionado mais com uma tendência do mercado do que com um comportamento de salto de fato.

Como o limite de Donoho e Johnstone é fixado para o horizonte escolhido de 1 dia, uma comparação natural que pode ser feita para os saltos localizados nos diferentes níveis é observar a proporção de dias que apresentaram saltos ou não, independentemente de quantos

Figura 5.4 – Exemplificação do processo de localização de saltos para o dia 13/01/2021, com 11 saltos. (a) Gráfico dos coeficientes de ondaletas referente ao primeiro nível de decomposição. (b) Gráfico dos valores absolutos dos coeficientes de ondaletas com o limite universal em vermelho indicando as observações que representam salto. (c) Gráfico da série de log-preço com os respectivos locais de salto detectados. (d) Gráfico do retorno ajustado aos saltos localizados.



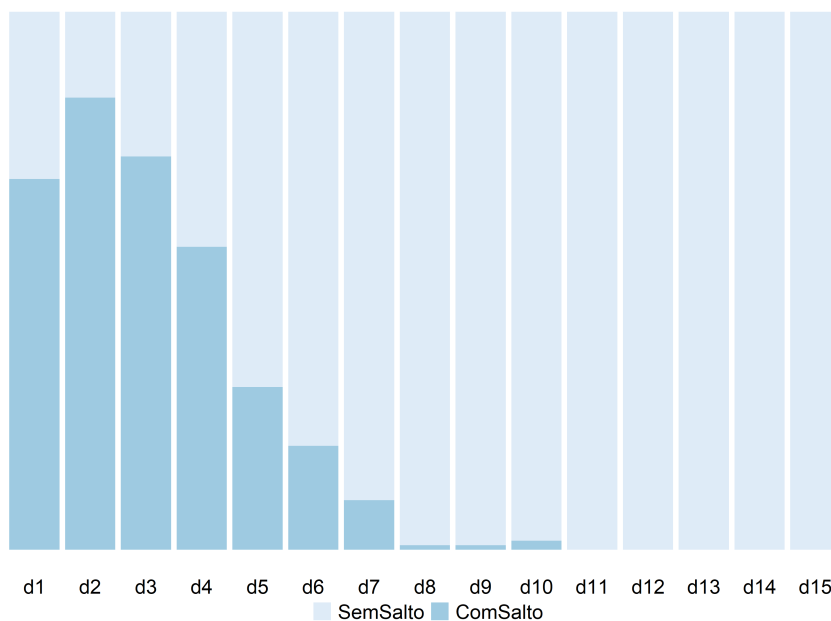
Fonte: do autor (2022).

saltos eles apresentaram. A Figura 5.6 apresenta esta proporção para cada uma das 15 escalas de tempo exploradas. Nota-se que, à medida que níveis mais suaves são explorados, o número de dias que apresentam saltos diminui. No entanto, conforme discutido anteriormente, para estes níveis de menor frequência, nos dias em que se verificou a presença de saltos, eles eram em maior número. Por exemplo, no nível  $d_7$  um único dia apresentou 79 saltos e no nível  $d_{10}$ , 56. Desta forma, é importante salientar que mais dias sem salto não indica, necessariamente, menos saltos, uma vez que se pode observar em um único dia uma quantidade maior de saltos (ver Tabela 5.5).

Estas situações em que muitos saltos são localizados no mesmo dia, em geral, consecutivamente, evidenciam mais a tendência do mercado na série do que de fato locais de salto, como é o caso do nível  $d_{10}$ . Neste nível, apenas dois dias apresentaram saltos: um deles, o dia 15/03/2021, com 56 saltos. A Figura 5.7 (a) apresenta a série de valores absolutos dos coeficientes de ondaletas para o nível  $d_{10}$  do dia 15/03/2021, destacado em vermelho o limite universal, de forma que todas as observações anteriores às 11h16 são consideradas saltos, definindo uma

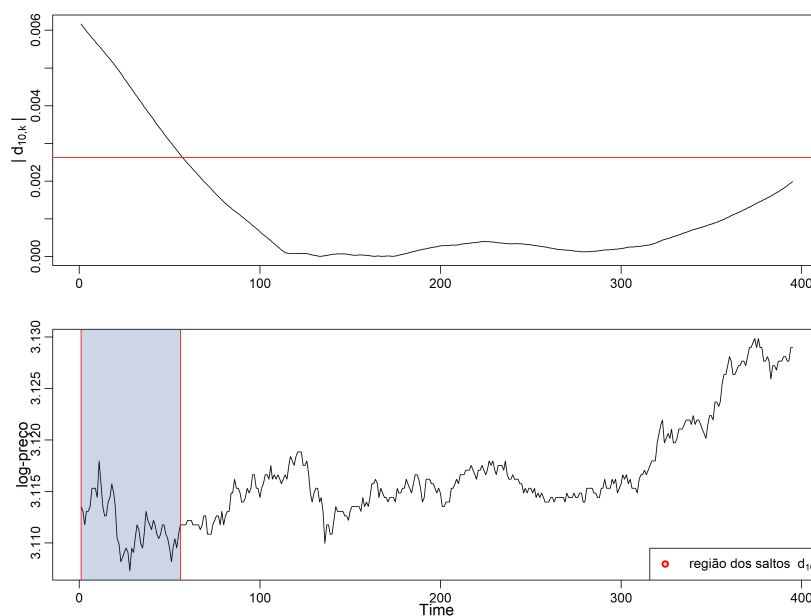


Figura 5.6 – Proporção de dias com e sem salto para cada nível da decomposição.



Fonte: do autor (2022).

Figura 5.7 – Saltos detectados para o dia 15/03/2021, com 56 saltos localizados no nível  $d_{10}$ . Em (a) a série de valores absolutos dos coeficientes de ondaletas com o respectivo limite universal em vermelho; em (b) a série de log-preços com a área hachurada destacando a região onde os saltos foram localizados.



Fonte: do autor (2022).

tempo equivalente (ver Tabela 5.1), conforme é possível confirmar nos histogramas apresentados na Figura 5.8.

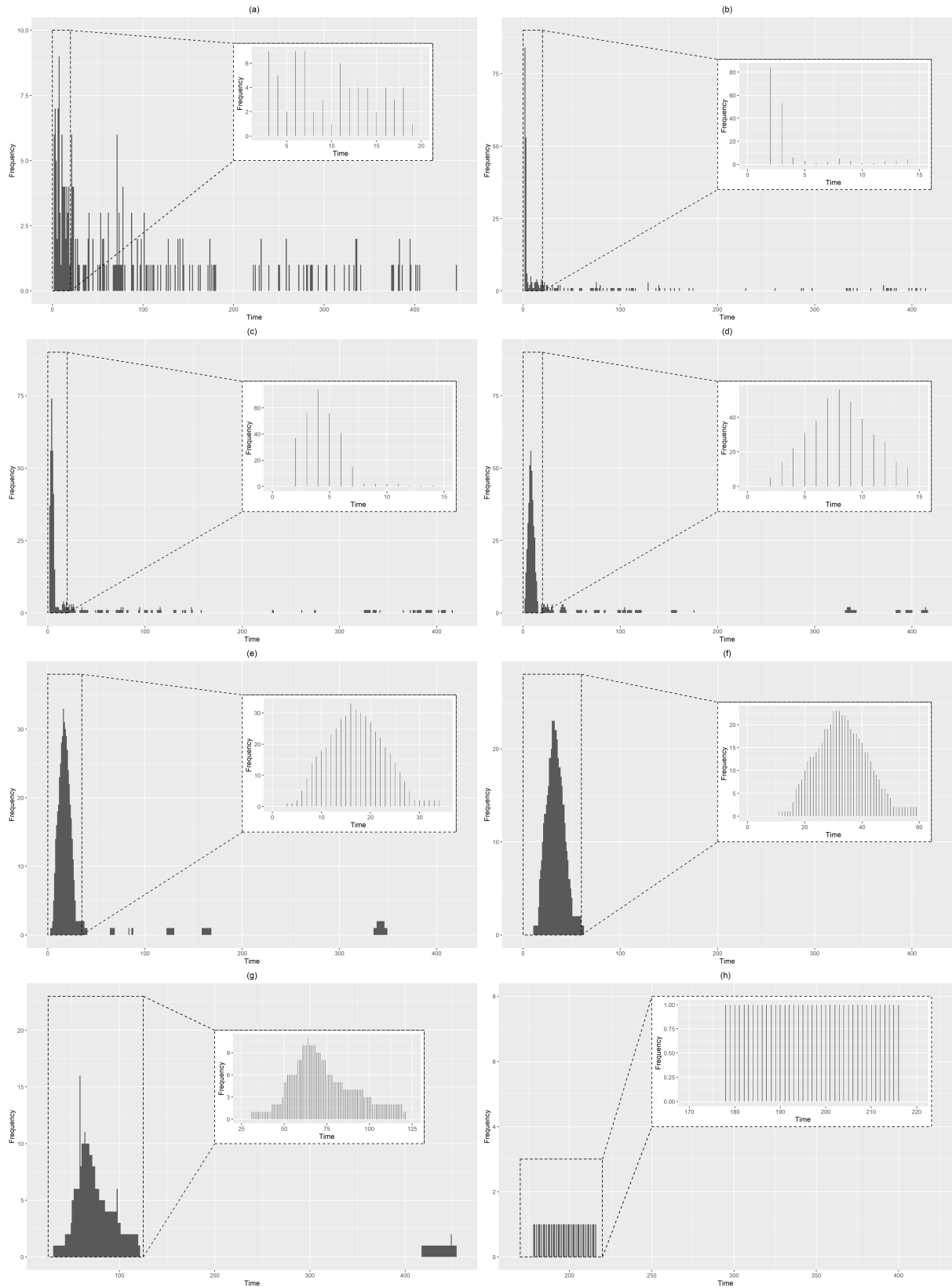
Para a construção destes histogramas, retirou-se da amostra as observações correspondentes ao dia de Quarta-Feira de Cinzas, uma vez que para ele o expediente começa às 13h, de forma que as observações iniciais e finais não correspondem ao mesmo instante de tempo dos demais dias. Além disso, em todos os níveis foram desconsiderados os saltos detectados na posição 1, pois eles refletem o efeito *overnight*, além de impossibilitar a estimação do tamanho do salto e do ajuste dos dados aos saltos.

Cada gráfico da Figura 5.8 apresenta a frequência dos saltos localizados em cada minuto dos dias da amostra, com um destaque amplificado da região onde a frequência foi a maior apresentada. A Figura 5.8 (a) apresenta o histograma para os saltos detectados no nível  $d_1$ ; a (b), os detectados no nível  $d_2$ ; e, assim, sucessivamente, até o nível  $d_8$ , na (h).

Tomando como exemplo a Figura 5.8 (e), correspondente aos saltos detectados no nível  $d_5$ . De acordo com a Tabela 5.1 esta escala corresponde à 16 - 32 minutos de negociação. O maior número de saltos detectados foi justamente no minuto 16, de acordo com a região em destaque inserida no gráfico. Enquanto para a Figura 5.8 (h), correspondente ao nível  $d_8$ , escala de 1 - 2 horas de negociação, os saltos detectados estão em torno de 120 a 240 minutos.

Esta observação enfatiza a necessidade de se analisar nível a nível os saltos detectados, já que em cada um deles é possível evidenciar a frequência de interesse.

Figura 5.8 – Gráficos dos histogramas das localizações de salto detectados nos níveis  $d_1$  (em (a)) a  $d_8$  (em (h)) com destaques amplificados para a região de maior frequência dos saltos localizados para a escala correspondente (ver Tabela 5.1).



Fonte: do autor (2022).

Uma vez localizados os saltos, seus tamanhos são estimados e, conseqüentemente, pode-se partir para a análise da Variação de Salto ( $JV$ , de *Jump Variation*). Na abordagem utilizada, foi possível estimar a Variação de Salto para cada nível, que fornece a variabilidade devida a saltos detectada nas diversas escalas de tempo para a série de log-preços da PETR4, e a soma delas compõe a Variação de Salto total da amostra. A Tabela 5.2 apresenta os valores estimados e acumulados da  $JV$  para cada nível de decomposição.

É possível observar que, até chegar ao nível  $d_8$  (suposto  $j_{max}$  segundo Fan e Wang (2007)), esta Variação de Salto mais que triplica, passando de  $5,2 \times 10^{-3}$  para  $17,3 \times 10^{-3}$ . A Figura 5.9 (a) mostra os valores estimados da Variação de Salto em cada nível, onde é possível observar que de fato o primeiro nível é o que apresenta maior  $JV$ . No entanto, ao analisar a Figura 5.9 (b), que fornece a Variação de Salto acumulada a cada nível, destaca-se a contribuição dos saltos detectados em todos os níveis para a composição de toda a  $JV$  Estimada.

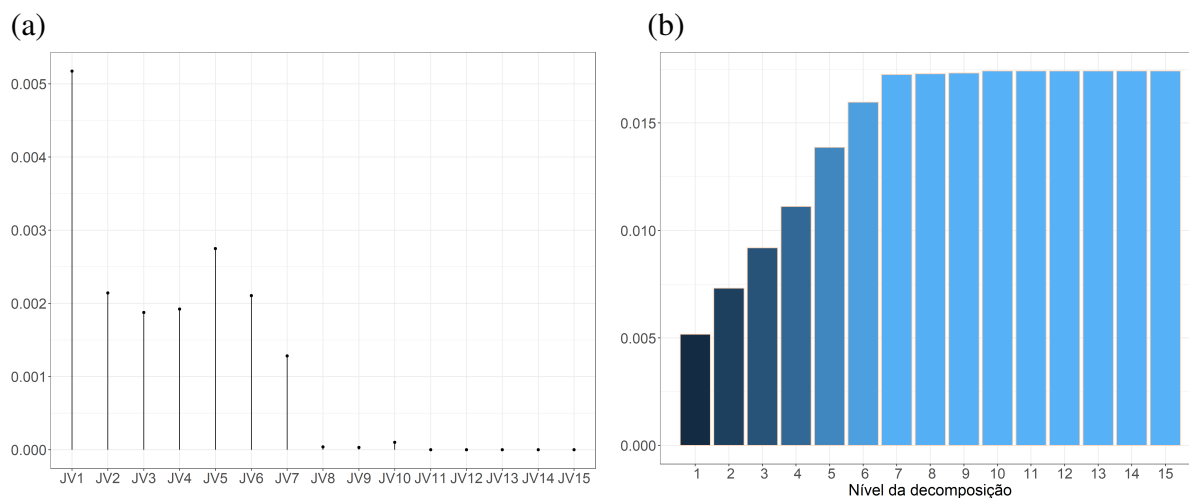
Tabela 5.2 – Valores da Variação de Salto estimada e acumulada a cada nível da decomposição da série de log-preço.

Nível	$JV$ Estimada	$JV$ Acumulada
$d_1$	0,005172	0,005172
$d_2$	0,002142	0,007314
$d_3$	0,001876	0,009190
$d_4$	0,001921	0,011111
$d_5$	0,002748	0,013859
$d_6$	0,002105	0,015964
$d_7$	0,001283	0,017247
$d_8$	$3,943 \times 10^{-05}$	0,017286
$d_9$	$3,152 \times 10^{-05}$	0,017318
$d_{10}$	$9,941 \times 10^{-05}$	0,017417
$d_{11}$	0	0,017417
$d_{12}$	0	0,017417
$d_{13}$	0	0,017417
$d_{14}$	0	0,017417
$d_{15}$	0	0,017417

Fonte: do autor (2022).

Estes resultados apontam para a contribuição que cada um dos diferentes níveis de decomposição da série de interesse tem na composição da Variabilidade devida à saltos. Além disso, a sugestão de se tomar o nível limite a ser explorado como o proposto por Fan e Wang (2007) parece ser um bom corte para se considerar. Na seção 5.3, mais detalhes a respeito destas Variações de Salto são discutidos, no contexto da medida realizada  $JV$  estimada.

Figura 5.9 – Gráficos referentes à variação de salto nível a nível da decomposição: (a) Valor estimado; (b) Valor acumulado.



Fonte: do autor (2022).

Um estudo mais minucioso a respeito da semana com o maior choque evidenciado na série foi conduzido e é apresentado na seção a seguir.

## 5.2 Semana de 22 a 26 de fevereiro de 2021

A fim de destacar a aplicação da detecção de saltos nos diferentes horizontes de investimento, representados aqui pelos níveis da decomposição, o comportamento dos saltos na semana de 22 a 26 de fevereiro de 2021 foi observado, semana seguinte à intervenção do então presidente do país na presidência da Petrobrás. O anúncio referido foi feito no final de semana, de forma que o maior impacto ocorreu no comportamento dos investidores na segunda e terça-feira subsequente e foi reduzindo ao longo da semana, conforme a Tabela 5.3.

Tabela 5.3 – Número de saltos detectados nos dias da semana de 22 a 26 de fevereiro de 2021.

Nível	22/02/2021	23/02/2021	24/02/2021	25/02/2021	26/02/2021
$d_1$	6	5	2	-	-
$d_2$	5	5	3	2	1
$d_3$	7	5	-	5	2
$d_4$	16	14	-	4	1
$d_5$	38	21	-	-	-
$d_6$	49	34	-	-	-
$d_7$	71	35	-	-	-

Obs. Os níveis a partir de  $d_8$  não apresentaram salto.

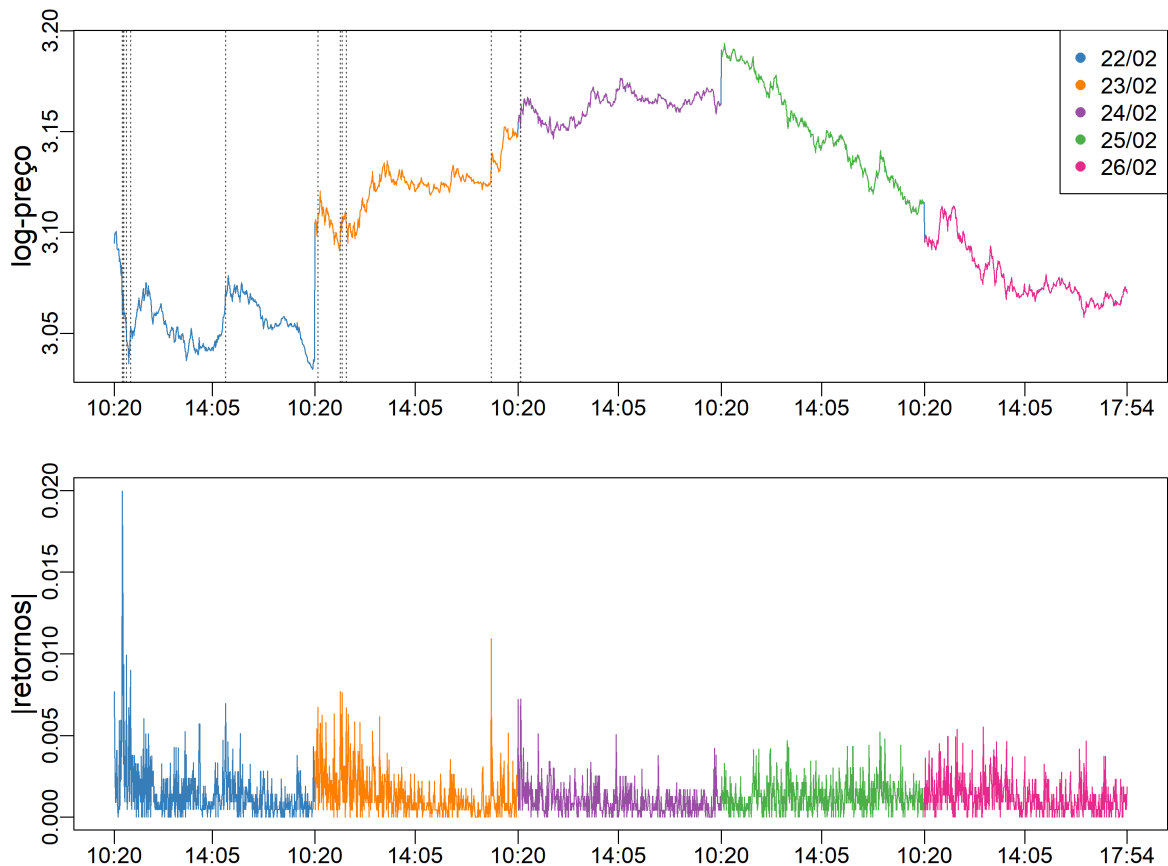
Fonte: do autor (2022).

Existe uma maior presença de saltos especialmente nos primeiros dias após o anúncio (da mudança da presidência da estatal) e nas primeiras horas destes dias, confirmando que os saltos aparecem devido a este comportamento mais reativo do mercado, caracterizando um choque na série de volatilidade. Também se nota que os saltos vão diminuindo ao longo dos dias, conforme o mercado vai se recuperando do efeito da notícia.

Há duas abordagens que se deseja destacar: analisar o comportamento dos saltos para um horizonte de tempo fixado, em diferentes dias da semana; e analisar o comportamento de um dia fixado (mais especificamente, a segunda-feira, 22/02/2021) para os diferentes horizontes de investimento. A primeira delas está apresentada na Figura 5.10 e a segunda, na Figura 5.11.

A Figura 5.10 apresenta a série de log-preço da semana do dia 22 a 26 de fevereiro de 2021, com os respectivos saltos detectados no primeiro nível de decomposição. Ao observar a Figura 5.10, é possível confirmar o agrupamento de saltos em horários próximos à abertura do mercado, nos três dias que apresentaram saltos. Além disso, o gráfico do valor absoluto dos retornos sinaliza os picos de volatilidade que caracterizam salto na série de log-preços.

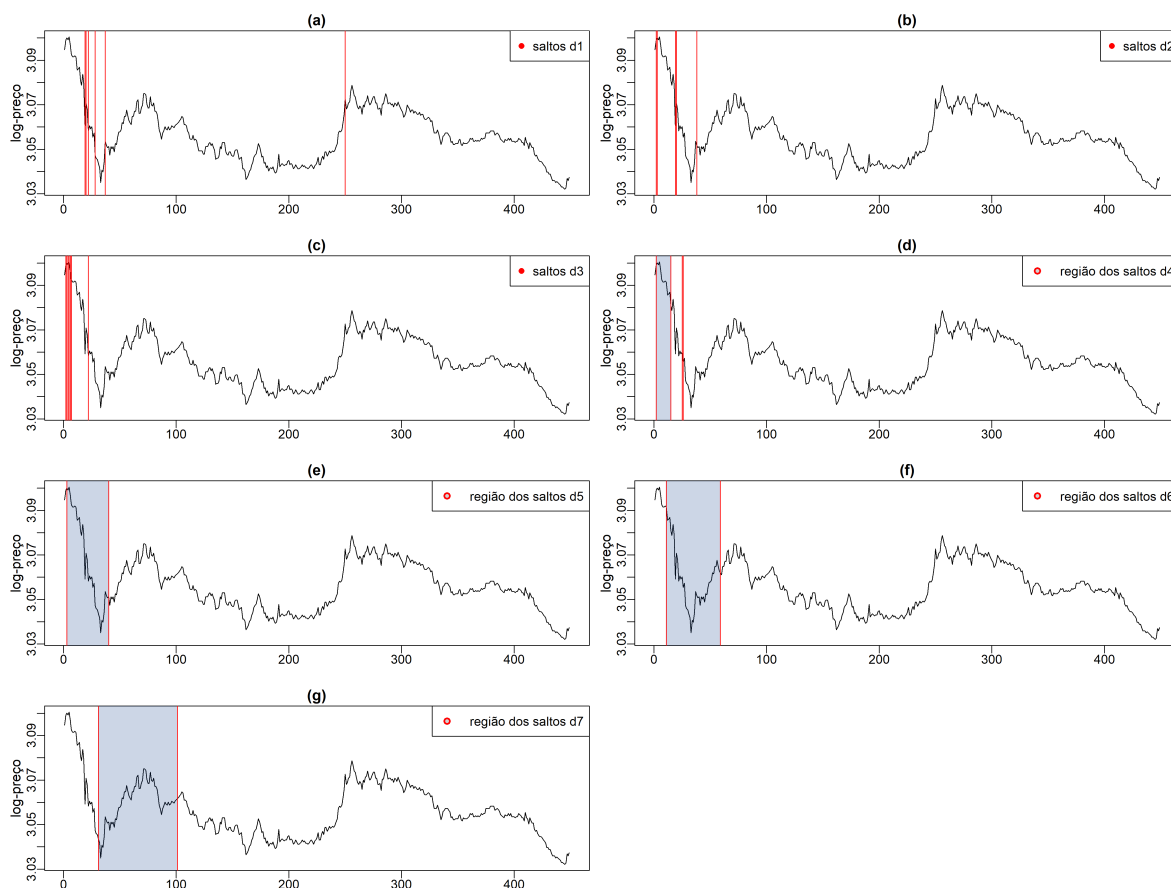
Figura 5.10 – Gráfico da série de log-preços com as localizações dos saltos detectados em  $d_1$  (a); e dos valores absolutos dos retornos (b) para os dias 22 a 26 de maio de 2021.



Fonte: do autor (2022).

Já a Figura 5.11 apresenta gráficos da série de log-preços para o dia de maior choque, 22/02/2021, destacando os saltos localizados nos níveis  $d_1$  (na Figura 5.11 (a)) até o  $d_7$  (Figura 5.11 (g)). Com exceção do primeiro nível, que detectou um salto isolado no período da tarde (Figura 5.11 (a)), todos os demais níveis detectaram saltos apenas nas primeiras duas horas de mercado. Além disso, nas frequências 4, 5, 6 e 7, são detectadas regiões (ou intervalos de tempo) de saltos, conforme sinalizado nas Figuras 5.11 (d), (e), (f) e (g) pelas regiões hachuradas em cinza com início e término do período de saltos detectados destacados em vermelho. Ou seja, nestas frequências não foram detectados saltos isolados, mas agrupamentos de saltos subsequentes minuto a minuto naquela escala/frequência. Somente após a 100ª observação, às 12h06, verificado na Figura 5.11 (g), referente ao nível 7 e à escala de 1 à 2 horas de negociação, a variabilidade devido ao choque tende a se estabilizar e o comportamento reativo do mercado se dilui.

Figura 5.11 – Gráficos da série de log-preços para o dia 22/02/2021 com a identificação dos saltos localizados para os níveis  $d_1$  (a),  $d_2$  (b),  $d_3$  (c),  $d_4$  (d),  $d_5$  (e),  $d_6$  (f),  $d_7$  (g). Para os níveis  $d_4$  a  $d_7$ , figuras (d) a (g), nota-se regiões (ou períodos) de salto, demarcados em vermelho pelo início e término do período referido, e o intervalo representado pela região hachurada em cinza.



Fonte: do autor (2022).

Para os demais dias da semana, o comportamento é semelhante ao apresentado na Figura 5.11, demonstrando que o efeito do choque da notícia se dilui ao longo do tempo, confirmado pela Tabela 5.3, com o número de saltos diminuindo com o passar dos dias (de 22 até 26) e com o passar das horas (da escala de 1-2 minutos para a de 1-2 horas).

Outra análise que foi possível realizar se dá a respeito de toda a variabilidade devida ao choque do dia 22/02/2021. Ela representa aproximadamente 16% de toda a Variação de Salto do nível  $d_1$ , e 26%, 8%, 8%, 41%, 57%, 43% dos níveis  $d_2$ ,  $d_3$ ,  $d_4$ ,  $d_5$ ,  $d_6$ ,  $d_7$ , respectivamente. Já com relação à variabilidade contida no dia 22/02/2021, as maiores contribuições são dos níveis  $d_5$  e  $d_6$ , com respectivamente 25% e 26% de toda a Variação de Salto presente neste dia, enquanto as menores contribuições foram dos níveis  $d_3$  e  $d_4$ , com cerca de 3% em cada.

### 5.3 Comparação de modelos da família HAR

A comparação que se segue é natural, uma vez que parte de uma série de retornos (original), para a qual pode ser ajustado o modelo HAR-RV tradicional e os modelos HAR-J e HAR-CJ, que fazem naturalmente uma decomposição da componente contínua e de salto para a série ajustada. A ideia é verificar se a eliminação de saltos por nível e as séries obtidas a partir delas, produzem melhores ajustes dos modelos da família HAR.

Para isto, inicialmente, é possível analisar as medidas realizadas estimadas.

#### 5.3.1 Medidas realizadas

A medida realizada  $RV$  é vista como uma medida natural da Variação Quadrática do processo e é uma medida diária obtida a partir das informações intradiárias. Como a  $BPV$  é uma medida realizada da Variância Integrada, tem-se por consequência que uma estimativa para a Variação de Salto do processo é dada por  $\widehat{JV} = \widehat{RV} - \widehat{BPV}$  (seção 3.4). Uma análise descritiva destas medidas está apresentada na Tabela 5.4 e o gráfico delas está apresentado na Figura 5.12.

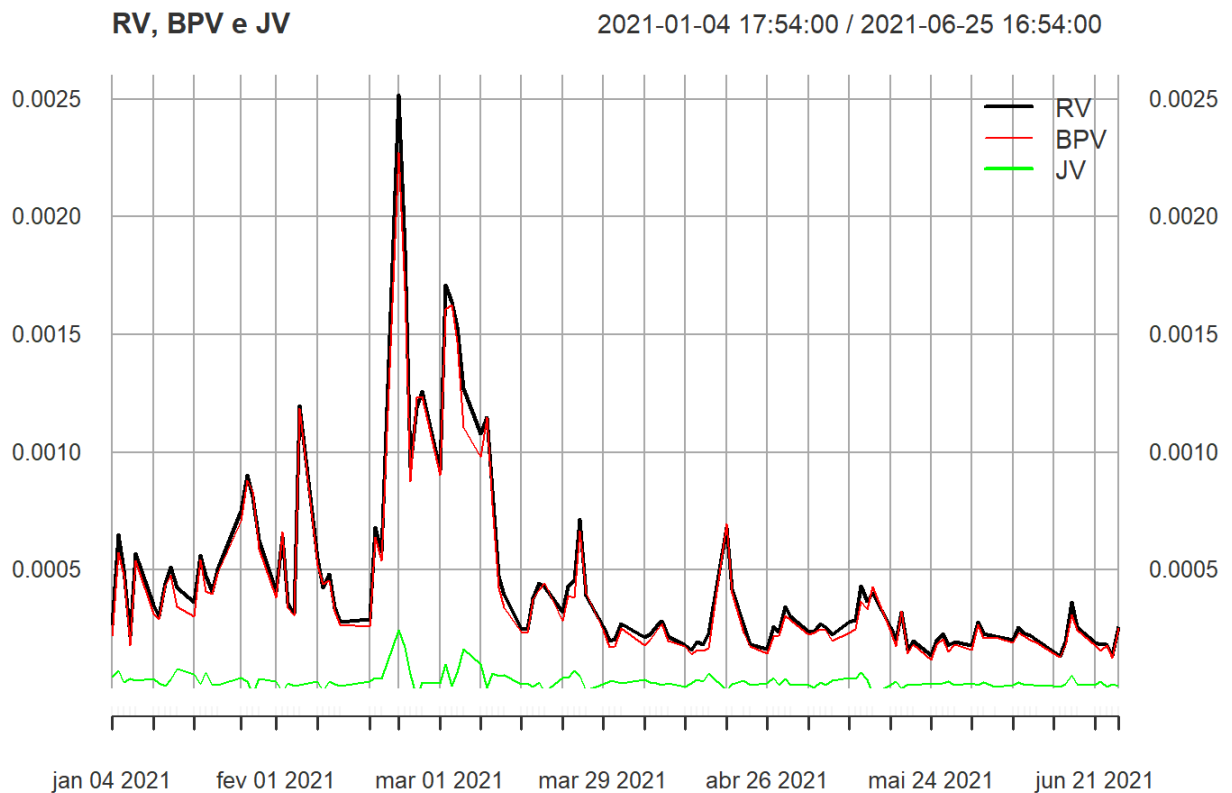
Tabela 5.4 – Descritiva das medidas realizadas  $RV$ ,  $BPV$  e  $JV$ .

	$RV$	$BPV$	$JV$
n.obs	119	119	119
Mínimo	0,13	0,12	-0,04
Máximo	2,52	2,27	0,24
1. Quartil	0,22	0,20	0,01
3. Quartil	0,49	0,47	0,04
Média	0,46	0,43	0,03
Mediana	0,30	0,28	0,02
D. Padrão	0,40	0,38	0,04
Assimetria	2,51	2,35	2,85
Curtose	7,27	6,01	12,78

Fonte: do autor (2022).

Uma maneira de comparar a metodologia desenvolvida, além da utilização da série de retornos ajustadas aos saltos, é a comparação da série  $JV$  obtida pela diferença  $RV - BPV$ , e a  $JV$  estimada pela metodologia de detecção de saltos por ondaletas. O primeiro ponto a destacar é que, segundo a metodologia desenvolvida neste trabalho, a série obtida tem o mesmo comprimento que a série de dados original, enquanto a série como medida realizada, limita-se

Figura 5.12 – Gráficos das medidas realizadas:  $RV$ ,  $BPV$  e  $JV$  para o período detectado.



Fonte: do autor (2022).

a um valor diário. Para poder efetuar tal comparação, então, realizou-se a abordagem contida em (3.31), de somar os quadrados dos saltos de cada dia.

A Tabela 5.5 apresenta uma análise descritiva destas  $JV$  estimadas. Comparando com a  $JV$  tradicional, é possível notar que, para qualquer nível de decomposição, a  $JV_{d_j}$ ,  $j = 1, 2, \dots, 15$ , apresenta maior média e desvio padrão, além de uma curtose muito mais acentuada.

A Figura 5.13 apresenta dois gráficos comparativos das Variações de Salto estimadas por nível da decomposição com a referência obtida com a  $BPV$ . Inicialmente, como a  $BPV$  pode ser vista como uma estimativa da Variância Integrada e  $\widehat{JV} = \widehat{RV} - \widehat{BPV}$ , pode-se pensar que a  $\widehat{IV} = \widehat{BPV} = \widehat{RV} - \widehat{JV}$ . Utilizando esta ideia, obtém-se duas estimativas extremas para a Variância Integrada do processo, uma subtraindo os saltos obtidos no primeiro nível  $IV_{d_1} = RV - JV_{d_1}$  e outra subtraindo os saltos obtidos até o décimo nível  $IV_{d_{10}} = RV - JV_{d_{10}}$ . Na Figura 5.13 (a) é possível observar que até mesmo considerando apenas os saltos no primeiro nível, ao comparar a  $IV_{d_1}$  com a  $BPV$  já apresenta uma diferença, especialmente em saltos

Tabela 5.5 – Descritiva das variações de salto estimadas em cada nível da decomposição.

	$JV_{d_1}$	$JV_{d_2}$	$JV_{d_3}$	$JV_{d_4}$	$JV_{d_5}$	$JV_{d_6}$	$JV_{d_7}$	$JV_{d_8}$	$JV_{d_9}$	$JV_{d_{10}}$
n.obs	119	119	119	119	119	119	119	119	119	119
Mínimo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Máximo	0,85	0,88	0,91	1,01	1,19	1,30	1,54	1,54	1,54	1,54
1. Quartil	0,00	0,01	0,01	0,01	0,02	0,02	0,02	0,02	0,02	0,02
3. Quartil	0,05	0,05	0,06	0,07	0,08	0,09	0,09	0,09	0,09	0,09
Média	0,04	0,05	0,06	0,07	0,08	0,08	0,09	0,09	0,09	0,09
Mediana	0,02	0,02	0,03	0,03	0,04	0,04	0,04	0,04	0,04	0,04
D,Padrão	0,09	0,10	0,11	0,12	0,14	0,15	0,18	0,18	0,18	0,18
Assimetria	6,08	5,73	5,27	5,05	5,27	5,32	5,61	5,62	5,62	5,62
Curtose	46,22	41,75	35,22	31,94	34,75	35,03	37,54	37,62	37,63	37,66

Fonte: do autor (2022).

mais aparentes, que caracteriza choques de volatilidade, como o do dia 22/02/2021. De maneira análoga, pode-se comparar apenas as Variações de Salto estimadas. Na Figura 5.13 (b), a série  $JV$  é a estimada a partir da  $BPV$ , enquanto as séries  $JV_{d_j}$ , são aquelas estimadas pelos saltos detectados até o nível  $d_j$ , para  $j = 1, 5$  e  $10$ . Novamente fica evidente que, quanto mais níveis de decomposição são considerados na estimativa da Variação de Salto, mais ela se distancia da estimada tradicionalmente a partir da  $BPV$ .

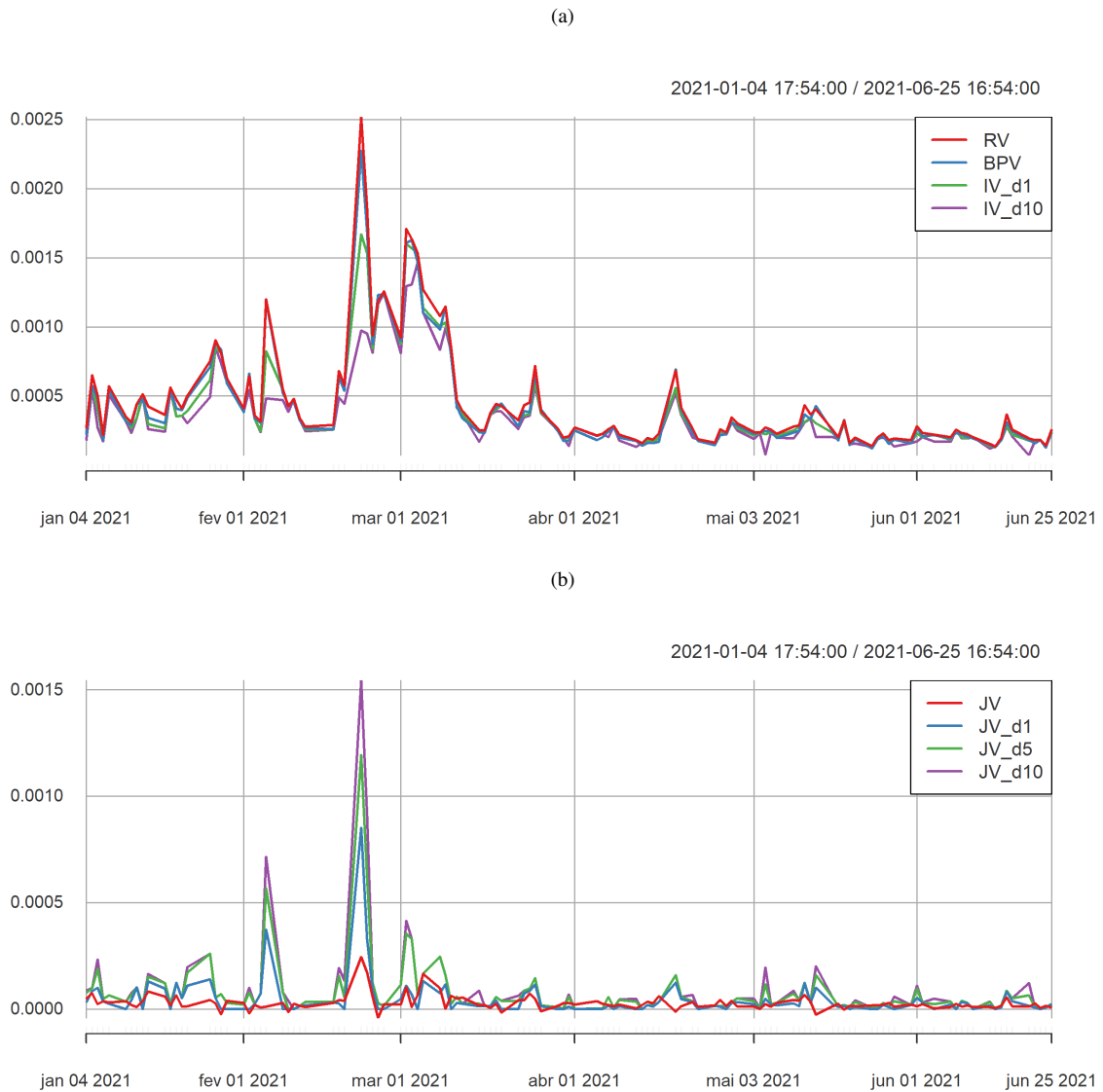
### 5.3.2 HAR-RV

O modelo HAR-RV compreende o ajuste da Volatilidade Realizada tradicional, dada pela soma dos quadrados dos retornos intradiários. Como este modelo não é robusto a saltos, espera-se que o ajuste seja melhor para os retornos ajustados aos saltos nos diferentes níveis de decomposição.

A Figura 5.14 resume as estimativas dos parâmetros e, entre parênteses, os respectivos valores-p para o modelo HAR-RV (apresentado em (3.38)), ajustado para as transformações logarítmicas das diferentes séries de retorno: retorno original; retorno  $r_1$ : ajustado aos saltos detectados no nível  $d_1$ ; retorno  $r_2$ : ajustado aos saltos detectados nos níveis  $d_1$  e  $d_2$ ; e, assim, sucessivamente, até retorno  $r_{10}$ : retorno ajustado aos saltos detectados nos níveis  $d_1$  até  $d_{10}$ .

Ao nível de significância de 1%, o coeficiente referente ao horizonte de tempo diário ( $\beta_1$ ) foi significativo para todos os modelos ajustados. É possível notar que, a medida que a série se torna mais livre de saltos (e, portanto, de ruídos) a significância do horizonte semanal ( $\beta_2$ ) vai aumentando, evidenciando um efeito de memória longa característico das séries de alta frequência, porém na componente semanal ao invés da componente mensal.

Figura 5.13 – Gráficos comparativos para as Variações de saltos estimadas por nível.



Fonte: do autor (2022).

Considerando a medida de qualidade de ajuste  $R^2$  ou  $R^2$  ajustado, pode-se considerar os ajustes de  $r_1$  a  $r_4$ , de forma que não há uma melhora evidente em relação a eliminar mais ou menos saltos. A principal contribuição fica a cargo da identificação do componente de memória longa a medida que mais saltos são localizados e eliminados, conforme comentado anteriormente.

É possível notar que a utilização dos retornos ajustados aos saltos traz uma melhora de até 6% na qualidade do ajuste ao considerar a eliminação de saltos até o sexto nível da decomposição. No entanto, do nível  $d_7$  em diante, o ajuste passa a ser "pior" que para os dados originais.

Figura 5.14 – Estimativas do ajuste do modelo HAR-RV tradicional para a série original e as séries ajustadas aos saltos a cada nível da decomposição.

		Ajustados aos saltos									
	Ret.original	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10
beta0	-0.840+	-0.747	-0.746	-0.752	-0.752	-0.760	-0.780	-0.801	-0.806	-0.807	-0.832
	(0.079)	(0.101)	(0.102)	(0.102)	(0.104)	(0.123)	(0.143)	(0.186)	(0.184)	(0.186)	(0.184)
beta1	0.707***	0.699***	0.691***	0.677***	0.675***	0.671***	0.634***	0.568***	0.564***	0.561***	0.563***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
beta2	0.090	0.154	0.165	0.187+	0.194*	0.196*	0.229*	0.293**	0.296**	0.299**	0.290**
	(0.263)	(0.141)	(0.127)	(0.069)	(0.040)	(0.034)	(0.017)	(0.006)	(0.006)	(0.005)	(0.006)
beta3	0.101	0.058	0.054	0.047	0.042	0.042	0.045	0.045	0.045	0.045	0.050
	(0.200)	(0.326)	(0.346)	(0.410)	(0.462)	(0.466)	(0.479)	(0.541)	(0.542)	(0.538)	(0.497)
R2	0.701	0.741	0.741	0.740	0.741	0.737	0.719	0.689	0.687	0.686	0.677
R2 Adj.	0.691	0.732	0.733	0.731	0.733	0.728	0.710	0.679	0.677	0.676	0.666

Fonte: do autor (2022).

### 5.3.3 HAR-J

Ao considerar modelos convencionais robustos aos saltos, é possível comparar se a eliminação dos saltos a cada nível está contribuindo de alguma maneira para o ajuste dos modelos.

O primeiro deles, HAR-J, apresentado em (3.39), estima a parte suave da variação de preço, considerando a Variação BiPower  $BPV$  para obter estimativas da componente de salto. A Figura 5.15 apresenta as estimativas dos parâmetros do modelo e seus respectivos valores-p para os 11 diferentes HAR-J ajustados aos retornos originais e aos retornos após eliminação dos saltos detectados nos diversos níveis de decomposição da série de log-preço.

Para o modelo HAR-J novamente o componente de volatilidade diária ( $\beta_1$ ) foi significativo ao nível de 1% de significância para todos os modelos ajustados.

Com relação à presença de memória longa evidenciada pelo componente mensal nos modelos do tipo HAR, esta foi observada apenas para os retornos originais (através do coeficiente  $\beta_3$ ) de forma que a metodologia empregada acabou por controlar este efeito.

Para o caso aplicado aos dados originais evidenciou-se um componente de memória longa no componente de salto, através do parâmetro correspondente ao horizonte mensal ( $\beta_6$ ).

Tabela 5.6 – Relação da qualidade de ajuste  $R^2$  em relação ao ajuste do modelo HAR-RV para os dados originais.

Série	$R^2$	$R^2/R^2.or$
ret.or	0,726	1,000
$r_1$	0,688	0,948
$r_2$	0,697	0,960
$r_3$	0,692	0,954
$r_4$	0,697	0,960
$r_5$	0,703	0,968
$r_6$	0,700	0,964
$r_7$	0,732	1,008
$r_8$	0,730	1,006
$r_9$	0,730	1,005
$r_{10}$	0,720	0,992

Fonte: do autor (2022).

Figura 5.15 – Estimativas do ajuste do modelo HAR-J para a série original e as séries ajustadas aos saltos a cada nível da decomposição.

	Ajustados aos saltos										
	Ret.original	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10
beta0	3.102 (0.220)	0.920 (0.764)	-0.056 (0.984)	-0.608 (0.851)	-0.499 (0.886)	-0.396 (0.915)	-0.744 (0.834)	-0.588 (0.889)	-0.605 (0.887)	-0.668 (0.878)	-0.645 (0.881)
beta1	0.577*** (0.000)	0.591*** (0.000)	0.503*** (0.000)	0.501*** (0.000)	0.508*** (0.000)	0.493*** (0.000)	0.434*** (0.001)	0.377* (0.017)	0.374* (0.018)	0.368* (0.020)	0.373* (0.018)
beta2	-0.030 (0.808)	0.169 (0.178)	0.102 (0.470)	0.151 (0.327)	0.126 (0.368)	0.107 (0.416)	0.081 (0.532)	0.074 (0.668)	0.076 (0.670)	0.077 (0.663)	0.057 (0.753)
beta3	0.779* (0.046)	0.333 (0.285)	0.384 (0.218)	0.277 (0.366)	0.306 (0.343)	0.352 (0.320)	0.399 (0.261)	0.481 (0.171)	0.480 (0.173)	0.478 (0.182)	0.496 (0.169)
beta4	1691.637 (0.293)	2123.464 (0.599)	2231.890 (0.494)	2351.379 (0.502)	2210.085 (0.485)	2874.403 (0.419)	3073.165 (0.368)	4028.087** (0.001)	3997.642** (0.001)	4072.739** (0.002)	4040.587** (0.002)
beta5	2038.199 (0.502)	1328.965 (0.524)	3510.374+ (0.061)	2986.534 (0.186)	3510.189 (0.183)	3630.444 (0.281)	5248.140 (0.119)	6643.980 (0.100)	6699.566+ (0.098)	6733.506+ (0.096)	6930.452+ (0.086)
beta6	-25003.297* (0.043)	-12048.642 (0.380)	-9081.041 (0.353)	-6379.711 (0.561)	-7268.381 (0.529)	-8601.566 (0.486)	-9262.097 (0.447)	-12439.150 (0.360)	-12408.021 (0.367)	-12267.795 (0.380)	-12707.951 (0.366)
R2	0.726	0.688	0.697	0.692	0.697	0.703	0.700	0.732	0.730	0.730	0.720
R2 Adj.	0.708	0.667	0.677	0.672	0.677	0.683	0.680	0.714	0.712	0.712	0.702

Fonte: do autor (2022).

No entanto, é possível notar que ele deixa de ser significativo ao aplicar a metodologia apresentada já no primeiro nível de decomposição. Apesar de o componente de salto diário ( $\beta_4$ ) passar a ser significativo ao considerar os retornos ajustados aos saltos nos níveis 7 em diante, isto não impactou a qualidade de ajuste que retornou maiores valores de  $R^2$  para a série  $r_7$ .

### 5.3.4 HAR-CJ

O HAR-CJ (apresentado em (3.40)), também separa componente contínua e de salto, com uma variação na estimativa da componente de salto considerando apenas saltos de maior magnitude. As estimativas para a aplicação deste modelo estão apresentadas na Figura 5.16.

Figura 5.16 – Estimativas do ajuste do modelo HAR-CJ para a série original e as séries ajustadas aos saltos a cada nível da decomposição.

	Ret.original	Ajustados aos saltos									
	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	
beta0	-0.902*	-0.826	-0.999	-0.946	-0.956	-0.976	-1.040	-1.442*	-1.444*	-1.739*	-1.778*
	(0.036)	(0.413)	(0.409)	(0.382)	(0.330)	(0.311)	(0.255)	(0.045)	(0.045)	(0.020)	(0.019)
beta1	0.722***	0.578***	0.542**	0.525**	0.526**	0.536**	0.496**	0.415**	0.412**	0.384**	0.391**
	(0.000)	(0.000)	(0.003)	(0.003)	(0.003)	(0.002)	(0.005)	(0.002)	(0.002)	(0.005)	(0.005)
beta2	0.016	0.255*	0.286**	0.315**	0.320***	0.308***	0.340***	0.322*	0.325*	0.312*	0.296*
	(0.869)	(0.010)	(0.009)	(0.003)	(0.001)	(0.001)	(0.000)	(0.018)	(0.019)	(0.018)	(0.026)
beta3	0.148+	0.067	0.052	0.044	0.037	0.038	0.041	0.119	0.118	0.134	0.138
	(0.053)	(0.367)	(0.470)	(0.560)	(0.638)	(0.612)	(0.626)	(0.224)	(0.227)	(0.192)	(0.173)
beta4	7223.359*	2884.002	2760.724	2655.420	2887.470	2992.486	3037.568	3911.069*	3920.785*	3964.477*	3980.102*
	(0.023)	(0.563)	(0.530)	(0.530)	(0.518)	(0.505)	(0.487)	(0.022)	(0.021)	(0.027)	(0.028)
beta5	-16959.476**	722.501	480.397	922.421	1134.470	1177.960	2139.742	1814.893	1791.765	2013.844	1975.910
	(0.002)	(0.719)	(0.784)	(0.644)	(0.610)	(0.581)	(0.389)	(0.569)	(0.575)	(0.493)	(0.497)
beta6	3138.582	-3849.350	-711.505	-3502.117	-3813.033	-2010.084	-1553.807	21307.363+	21209.155+	30795.714*	31108.970*
	(0.621)	(0.711)	(0.958)	(0.778)	(0.752)	(0.873)	(0.899)	(0.059)	(0.062)	(0.016)	(0.014)
R2	0.715	0.669	0.667	0.666	0.671	0.674	0.665	0.700	0.699	0.704	0.695
R2 Adj.	0.696	0.647	0.644	0.644	0.649	0.652	0.642	0.680	0.679	0.684	0.675

Fonte: do autor (2022).

Considerando o modelo HAR-CJ, a componente contínua de horizonte diário é significativa para todos os modelos ajustados (referente ao parâmetro  $\beta_1$ ), com a componente contínua semanal ( $\beta_2$ ) se mostrando também importante para as séries ajustadas aos saltos.

Considerando os dados originais, diferentemente do modelo HAR-J, não mais a componente de salto de memória longa refere-se ao horizonte mensal ( $\beta_6$ ), mas sim ao semanal ( $\beta_5$ ). Além disso, as componentes discretas de horizonte diário e semanal ( $\beta_4$  e  $\beta_5$ , respectivamente) se mostram significativas para os dados originais e também são controladas a partir da aplicação da metodologia proposta, já no primeiro nível da decomposição.

Uma vez mais nota-se que as componentes de salto de horizonte diário e mensal ( $\beta_4$  e  $\beta_6$ ) são significativas na aplicação do modelo às séries que eliminam os saltos em níveis maiores ou iguais a 7 (assim como aconteceu para o HAR-J com relação ao componente de salto diário).

Em termos de qualidade de ajuste, a eliminação de saltos em níveis de alta frequência não aprimora o ajuste do modelo aos retornos originais, reforçando este modelo como *benchmark* no ajuste de séries com presença de saltos.

## 6 CONCLUSÃO

A literatura econométrica apresenta diversas maneiras de se lidar com os saltos na estimação da volatilidade das séries financeiras, no entanto, é mais comum comparar modelos robustos a saltos do que de fato compreender o seu comportamento e suas implicações na estimação de sua variação.

O objetivo desta tese foi propor um estudo minucioso a respeito da Variação de Salto, através da utilização da abordagem de ondaletas para detectar saltos em diferentes escalas de tempo e num contexto em que a volatilidade apresenta diferentes padrões em períodos distintos, principalmente devido a choques exógenos, natural em períodos de crises e/ou mudanças políticas.

Consolidada como uma ferramenta de eliminação de ruído, foi apresentado que também é possível utilizar as transformadas de ondaletas para determinar saltos na série de ativos e, com isso, se aproximar do valor verdadeiro da volatilidade integrada, em especial em situações com um choque de volatilidade tão expressivo. O estudo foi conduzido para a série de preços da PETR4, no período de janeiro a junho de 2021, com destaque para um período evidente de queda de preços devido a uma interferência externa ao mercado.

Para a série analisada foi possível observar que a metodologia proposta proporciona de fato uma maior compreensão do comportamento dos saltos em diferentes escalas de tempo e como este pode afetar a dinâmica da volatilidade dos preços dos ativos. Ademais, do ponto de vista prático, a aplicação da metodologia pode ajudar os operadores de bolsa de valores (investidores ou *traders*) a encontrar informações importantes na volatilidade em diferentes horizontes de investimento, o que pode auxiliar nas estratégias com base no estudo das características da volatilidade dos ativos nos diferentes níveis de detalhe das séries temporais a serem exploradas.

Com relação às estimativas de Variação de Salto obtidas, apesar de a maior contribuição para esta componente da volatilidade vir dos saltos detectados no primeiro nível de decomposição, a incorporação das diferentes escalas temporais retorna uma Variação três vezes maior. Além disso, ao comparar estas Variações de Saltos dos diferentes níveis com a estimada pela referência (através da medida de volatilidade realizada de *BPV*), notou-se que a obtida pela abordagem de ondaletas parece deixar mais evidente os picos devidos aos choques. Estes dois resultados são relevantes uma vez que, quanto mais precisa for a estimação da Variação de Salto, melhor será a estimação da Variância Integrada.

Baseado nos objetivos propostos, foi realizado também o estudo comparativo considerando os dados ajustados aos saltos tanto para as abordagens padrão quanto para a de ondas. Foi possível notar que a aplicação dos modelos da família HAR evidenciou a presença de memória longa da volatilidade realizada: inicialmente ao considerar o modelo HAR-RV tradicional, na qual ela aparece após a eliminação dos saltos detectados nos diversos níveis de decomposição; enquanto ao se considerar os modelos que decompõem as componentes contínuas e discretas, a memória longa se mostrou presente na componente de salto semanal/mensal, que foi controlada com a metodologia aplicada. O resultado obtido com os modelos HAR-J e HAR-CJ que mostra que a componente de salto perde significância ao comparar o ajuste aos dados originais e o ajuste aos dados com saltos eliminados evidencia que a utilização da metodologia apresentada para eliminação de saltos permite que a etapa de previsão da volatilidade não se limite aos modelos robustos a saltos, visto que a componente discreta já foi controlada.

Uma limitação natural da metodologia proposta é devido ao fato de, até certo ponto, não ser possível uma generalização dos resultados, já que a presença de um choque significativo vai estar mais ou menos expressivo em uma série de dados em particular, para um específico período de tempo, para uma dada frequência escolhida. Desta forma, o que fica constatada é a necessidade de se utilizar a metodologia para justamente explorar a série de interesse conforme necessidade do pesquisador ou do especulador de mercado.

Além disso, maiores estudos podem ser conduzidos no sentido de se comparar as previsões de volatilidade fora da amostra considerando os dados ajustados aos saltos, no sentido de detectar alguma vantagem de utilizar a decomposição até um nível específico. Outra linha pode ser propôr estudos que examinem modelos que consideram ruído de microestrutura, no sentido de agregar ainda mais o aprimoramento da previsão da Variância Integrada do processo de preço.

## REFERÊNCIAS

- ANDERSEN, T. G.; BOLLERSLEV, T. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. **The Journal of Finance**, Wiley Online Library, v. 52, n. 3, p. 975–1005, 1997.
- ANDERSEN, T. G.; BOLLERSLEV, T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. **International economic review**, JSTOR, p. 885–905, 1998.
- ANDERSEN, T. G.; BOLLERSLEV, T. Deutsche mark–dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. **the Journal of Finance**, Wiley Online Library, v. 53, n. 1, p. 219–265, 1998.
- ANDERSEN, T. G. et al. Volatility and correlation forecasting. **Handbook of economic forecasting**, Elsevier, v. 1, p. 777–878, 2006.
- ANDERSEN, T. G.; BOLLERSLEV, T.; DIEBOLD, F. X. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. **The review of economics and statistics**, MIT Press, v. 89, n. 4, p. 701–720, 2007.
- ANDERSEN, T. G. et al. The distribution of realized exchange rate volatility. **Journal of the American statistical association**, Taylor & Francis, v. 96, n. 453, p. 42–55, 2001.
- ANDERSEN, T. G. et al. The distribution of realized stock return volatility. **Journal of financial economics**, Elsevier, v. 61, n. 1, p. 43–76, 2001.
- ANDERSEN, T. G. et al. The distribution of realized stock return volatility. **Journal of financial economics**, Elsevier, v. 61, n. 1, p. 43–76, 2001.
- ANDERSEN, T. G. et al. Modeling and forecasting realized volatility. **Econometrica**, Wiley Online Library, v. 71, n. 2, p. 579–625, 2003.
- ANDERSEN, T. G.; DOBREV, D.; SCHAUMBURG, E. Jump-robust volatility estimation using nearest neighbor truncation. **Journal of Econometrics**, Elsevier, v. 169, n. 1, p. 75–93, 2012.
- ANDERSEN, T. G.; TERÄSVIRTA, T. Realized volatility. In: **Handbook of financial time series**. [S.l.]: Springer, 2009. p. 555–575.
- ARAÚJO, A. C. de. **Combinação de projeções de volatilidade baseadas em medidas de risco para dados em alta frequência**. 287 p. Tese (Doutorado em Administração) — Universidade de São Paulo, São Paulo, 2016.
- ARAÚJO JÚNIOR, J. B. de. **Modelagem Econométrica em alta frequência em um mercado de ações emergentes**. 200 p. Tese (Doutorado em Ciências Contábeis) — Universidade de Brasília, Brasília, 2020.
- BANDI, F. M.; RUSSELL, J. R. Separating microstructure noise from volatility. **Journal of Financial Economics**, Elsevier, v. 79, n. 3, p. 655–692, 2006.
- BANDI, F. M.; RUSSELL, J. R.; YANG, C. Realized volatility forecasting and option pricing. **Journal of Econometrics**, Elsevier, v. 147, n. 1, p. 34–46, 2008.

BARNDORFF-NIELSEN, O. E. et al. Regular and modified kernel-based estimators of integrated variance: The case with independent noise. 2004.

BARNDORFF-NIELSEN, O. E.; SHEPHARD, N. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 64, n. 2, p. 253–280, 2002.

BARNDORFF-NIELSEN, O. E.; SHEPHARD, N. Estimating quadratic variation using realized variance. **Journal of Applied econometrics**, Wiley Online Library, v. 17, n. 5, p. 457–477, 2002.

BARNDORFF-NIELSEN, O. E.; SHEPHARD, N. **Measuring the impact of jumps in multivariate price processes using bipower covariation**. [S.l.], 2004.

BARUNIK, J.; KREHLIK, T.; VACHA, L. Modeling and forecasting exchange rate volatility in time-frequency domain. **European Journal of Operational Research**, Elsevier, v. 251, n. 1, p. 329–340, 2016.

BARUNIK, J.; VACHA, L. Realized wavelet-based estimation of integrated variance and jumps in the presence of noise. **Quantitative Finance**, Taylor & Francis, v. 15, n. 8, p. 1347–1364, 2015.

BOUDT, K. et al. **highfrequency: Tools for Highfrequency Data Analysis**. [S.l.], 2020. R package version 0.7.0.1. Disponível em: <<https://CRAN.R-project.org/package=highfrequency>>.

BOUDT, K.; CROUX, C.; LAURENT, S. Outlyingness weighted covariation. **Journal of Financial Econometrics**, Oxford University Press, v. 9, n. 4, p. 657–684, 2011.

BOUDT, K.; ZHANG, J. Jump robust two time scale covariance estimation and realized volatility budgets. **Quantitative Finance**, Taylor & Francis, v. 15, n. 6, p. 1041–1054, 2015.

BUCCI, A. et al. Forecasting realized volatility: a review. **Journal of Advanced Studies in Finance (JASF)**, ASERS Publishing, v. 8, n. 16, p. 94–138, 2017.

CAPOBIANCO, E. Multiscale stochastic dynamics in finance. In: **IEEE. 2003 IEEE International Workshop on Workload Characterization (IEEE Cat. No. 03EX775)**. [S.l.], 2003. v. 1, p. 222–227.

ČECH, F.; BARUNÍK, J. On the modelling and forecasting of multivariate realized volatility: Generalized heterogeneous autoregressive (ghar) model. **Journal of forecasting**, Wiley Online Library, v. 36, n. 2, p. 181–206, 2017.

CHIRIAC, R.; VOEV, V. Modelling and forecasting multivariate realized volatility. **Journal of Applied Econometrics**, Wiley Online Library, v. 26, n. 6, p. 922–947, 2011.

CORSI, F. A simple approximate long-memory model of realized volatility. **Journal of Financial Econometrics**, Oxford University Press, v. 7, n. 2, p. 174–196, 2009.

CORSI, F.; RENO, R. Har volatility modelling with heterogeneous leverage and jumps. **Available at SSRN 1316953**, Citeseer, 2009.

CROWLEY, P. M. A guide to wavelets for economists. **Journal of Economic Surveys**, Wiley Online Library, v. 21, n. 2, p. 207–267, 2007.

DAUBECHIES, I. Orthonormal bases of compactly supported wavelets. **Communications on pure and applied mathematics**, Wiley Online Library, v. 41, n. 7, p. 909–996, 1988.

DAUBECHIES, I. **Ten lectures on wavelets**. [S.l.]: SIAM, 1992.

DONOHO, D. L.; JOHNSTONE, J. M. Ideal spatial adaptation by wavelet shrinkage. **biometrika**, Oxford University Press, v. 81, n. 3, p. 425–455, 1994.

FAN, J.; WANG, Y. Multi-scale jump and volatility analysis for high-frequency financial data. **Journal of the American Statistical Association**, Taylor Francis, v. 102, n. 480, p. 1349–1362, 2007. Disponível em: <<https://doi.org/10.1198/016214507000001067>>.

GENÇAY, R. et al. Asymmetry of information flow between volatilities across time scales. **Quantitative Finance**, Taylor & Francis, v. 10, n. 8, p. 895–915, 2010.

GENÇAY, R.; SELÇUK, F.; WHITCHER, B. J. **An introduction to wavelets and other filtering methods in finance and economics**. [S.l.]: Elsevier, 2001.

GHYSELS, E.; SANTA-CLARA, P.; VALKANOV, R. The midas touch: Mixed data sampling regression models. 2004.

GIOT, P.; LAURENT, S. Modelling daily value-at-risk using realized volatility and arch type models. **Journal of Empirical Finance**, v. 11, n. 3, p. 379 – 398, 2004. ISSN 0927-5398. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S092753980400012X>>.

GOODHART, C. A.; O'HARA, M. High frequency data in financial markets: Issues and applications. **Journal of Empirical Finance**, Elsevier, v. 4, n. 2-3, p. 73–114, 1997.

HANOUSEK, J.; KOČENDA, E.; NOVOTNÝ, J. The identification of price jumps. Walter de Gruyter GmbH & Co. KG, 2012.

HANSEN, P. R.; HUANG, Z.; SHEK, H. H. Realized garch: a joint model for returns and realized measures of volatility. **Journal of Applied Econometrics**, Wiley Online Library, v. 27, n. 6, p. 877–906, 2012.

HØG, E.; LUNDE, A. Wavelet estimation of integrated volatility. **Unpublished paper: Aarhus School of Business**, 2003.

HSIEH, D. A. Chaos and nonlinear dynamics: application to financial markets. **The journal of finance**, Wiley Online Library, v. 46, n. 5, p. 1839–1877, 1991.

HUANG, Z.; WANG, T.; ZHANG, X. Generalized autoregressive score model with realized measures of volatility. **Available at SSRN 2461831**, 2014.

IN, F.; KIM, S. **An introduction to wavelet theory in finance: a wavelet multiscale approach**. [S.l.]: World scientific, 2013.

KIM, S.; IN, F. Portfolio allocation and the investment horizon: a multiscaling approach. **Quantitative Finance**, Taylor & Francis, v. 10, n. 4, p. 443–453, 2010.

LAURENT, S. **Estimating and forecasting ARCH models using G@RCH 8**. [S.l.], 2018.

LOUZIS, D. P.; XANTHOPOULOS-SISINIS, S.; REFENES, A. P. Realized volatility models and alternative value-at-risk prediction strategies. **Economic Modelling**, v. 40, p. 101 – 116, 2014. ISSN 0264-9993. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0264999314001229>>.

MAHEU, J. M.; MCCURDY, T. H. Components of market risk and return. **Journal of Financial Econometrics**, Oxford University Press, v. 5, n. 4, p. 560–590, 2007.

MALLAT, S. **A wavelet tour of signal processing**. [S.l.]: Elsevier, 1999.

MANCINI, C.; GOBBI, F. Identifying the brownian covariation from the co-jumps given discrete observations. **Econometric Theory**, JSTOR, p. 249–273, 2012.

MCALEER, M.; MEDEIROS, M. C. Realized volatility: A review. **Econometric Reviews**, Taylor & Francis, v. 27, n. 1-3, p. 10–45, 2008.

MEDDAHI, N. A theoretical comparison between integrated and realized volatility. **Journal of Applied Econometrics**, Wiley Online Library, v. 17, n. 5, p. 479–508, 2002.

MORETTIN, P. A. **Ondas e Ondaletas: da análise de Fourier à análise de ondaletas de séries temporais**. [S.l.]: Edusp–Editora da Universidade de São Paulo, São Paulo., 2014.

MÜLLER, U. A. et al. Volatilities of different time resolutions—analyzing the dynamics of market components. **Journal of Empirical Finance**, Citeseer, v. 4, n. 2-3, p. 213–239, 1997.

OLIVEIRA, H. de. **Análise de Fourier e Wavelets**. [S.l.]: Hélio Magalhaes de Oliveira, 2007.

PATTON, A. J.; SHEPPARD, K. Good volatility, bad volatility: Signed jumps and the persistence of volatility. **Review of Economics and Statistics**, MIT Press, v. 97, n. 3, p. 683–697, 2015.

PERCIVAL, D. B.; WALDEN, A. T. **Wavelet methods for time series analysis**. [S.l.]: Cambridge university press, 2000. v. 4.

PINTO, M. G. de F. **Memória longa em séries financeiras utilizando ondaletas e modelos de volatilidade condicional**. 152 p. Dissertação (Mestrado em Estatística) — Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2021.

POON, S.-H.; GRANGER, C. W. Forecasting volatility in financial markets: A review. **Journal of economic literature**, v. 41, n. 2, p. 478–539, 2003.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2020. Disponível em: <<https://www.R-project.org/>>.

RAMSEY, J. B. The contribution of wavelets to the analysis of economic and financial data. **Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences**, The Royal Society, v. 357, n. 1760, p. 2593–2606, 1999.

ROCHA, V. B. d. **Uma abordagem de wavelets aplicada à combinação de previsões: uma análise teórica e experimental**. 155 p. Dissertação (Mestrado em Ciências) — Universidade Federal do Paraná, Curitiba, 2009.

- SANTOS, D. G.; ZIEGELMANN, F. A. Volatility forecasting via midas, har and their combination: An empirical comparative study for ibovespa. **Journal of Forecasting**, Wiley Online Library, v. 33, n. 4, p. 284–299, 2014.
- SHARMA, P. Forecasting stock market volatility using realized garch model: International evidence. **The Quarterly Review of Economics and Finance**, Elsevier, v. 59, p. 222–230, 2016.
- SHEPHARD, N.; SHEPPARD, K. Realising the future: forecasting with high-frequency-based volatility (heavy) models. **Journal of Applied Econometrics**, v. 25, n. 2, p. 197–231, 2010. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.1158>>.
- SHIN, D. W. Forecasting realized volatility: A review. **Journal of the Korean Statistical Society**, Elsevier, v. 47, n. 4, p. 395–404, 2018.
- SRIDHAR, S.; KUMAR, P. R.; RAMANAI AH, K. Wavelet transform techniques for image compression-an evaluation. **International journal of image, graphics and signal processing**, Modern Education and Computer Science Press, v. 6, n. 2, p. 54, 2014.
- SUBBOTIN, A. A multi-horizon scale for volatility. **Available at SSRN 1101376**, 2008.
- VAL, F. d. F. **Ensaio sobre a precificação empírica de ativos, política monetária e suas inter-relações**. 134 p. Tese (Doutorado em Administração) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2016.
- WANG, Y. Jump and sharp cusp detection by wavelets. **Biometrika**, Oxford University Press, v. 82, n. 2, p. 385–397, 1995.
- WHITCHER, B. **waveslim: Basic Wavelet Routines for One-, Two-, and Three-Dimensional Signal Processing**. [S.l.], 2020. R package version 1.8.2. Disponível em: <<https://CRAN.R-project.org/package=waveslim>>.
- WINK JUNIOR, M. V.; PEREIRA, P. L. V. Modeling and forecasting of realized volatility: evidence from brazil. Sociedade Brasileira de Econometria, 2011.
- WONG, Z. Y.; CHIN, W. C.; TAN, S. H. Daily value-at-risk modeling and forecast evaluation: The realized volatility approach. **The Journal of Finance and Data Science**, v. 2, n. 3, p. 171 – 187, 2016. ISSN 2405-9188. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2405918816300423>>.
- ZHANG, L. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. **Bernoulli Society for Mathematical Statistics and Probability**, v. 12, n. 6, p. 1019–1043, 2006.
- ZHANG, L.; MYKLAND, P. A.; AÏT-SAHALIA, Y. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. **Journal of the American Statistical Association**, Taylor & Francis, v. 100, n. 472, p. 1394–1411, 2005.
- ZHOU, B. High-frequency data and volatility in foreign-exchange rates. **Journal of Business & Economic Statistics**, Taylor & Francis Group, v. 14, n. 1, p. 45–52, 1996.