

JHON HEIDER DOMINGOS REZENDE

**VOLATILIDADE IMPLÍCITA DO PETRÓLEO: ANÁLISES EM APRENDIZADO DE
MÁQUINA E MÉTODOS TRADICIONAIS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Economia, para obtenção do título de *Magister Scientiae*.

Orientador: Luciano Dias de Carvalho

Coorientadores: Eduardo José da Silva Luz
Júlio César Araújo da Silva Júnior

**VIÇOSA - MINAS GERAIS
2021**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

Rezende, Jhon Heider Domingos, 1996-
R467v Volatilidade implícita do petróleo [recurso eletrônico]:
2021 análises em aprendizado de máquina e métodos tradicionais /
Jhon Heider Domingos Rezende. – Viçosa, MG, 2021.
1 dissertação eletrônica (63 f.): il.

Orientador: Luciano Dias de Carvalho.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 57-63.

DOI: <https://doi.org/10.47328/ufvbbt.2021.042>

Modo de acesso: World Wide Web.

1. Petróleo - Comércio. 2. Macroeconomia. 3. Análise de séries temporais. 4. Mercado futuro. I. Universidade Federal de Viçosa. Departamento de Economia. Programa de Pós-Graduação em Economia. II. Título.

CDD 22. ed. 338.27282

JHON HEIDER DOMINGOS REZENDE

**VOLATILIDADE IMPLÍCITA DO PETRÓLEO: ANÁLISES EM APRENDIZADO DE
MÁQUINA E MÉTODOS TRADICIONAIS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Economia, para obtenção do título de *Magister Scientiae*.

APROVADA: 17 de junho de 2021.

Assentimento:

Jhon Heider Domingos Rezende
Autor

Luciano Dias de Carvalho
Orientador

Aos meus pais e irmãos.

AGRADECIMENTOS

A experiência em vivenciar um curso de pós graduação em Economia é marcada por uma gama de aprendizados e desafios. Essa jornada só é possível paralela à existência de uma rede sólida de apoio de pessoas que se importam conosco e estão prontas para partilhar a vitória ao fim desse caminho.

Assim, gostaria de registrar primeiramente o apoio incondicional de minha mãe, Maria Ediléia da Silva Rezende, e de meu pai, Pedro Domingos de Rezende, que nem por um segundo mediram esforços para que eu me tornasse um ser humano melhor a cada dia. Meu coração lhes pertence e assim o será até a eternidade! Tal como fizeram por mim, fizeram também por meus irmãos, Jámbertt e Jefferson, que cuidaram e me ensinaram tão bem até aqui. Eu amo todos vocês!

Ao Programa de Pós Graduação em Economia da Universidade Federal de Viçosa, à coordenação do programa e aos professores do Departamento de Economia, agradeço pelo esforço em melhorar o ensino e oferecer melhor infraestrutura aos alunos, assim como pela oportunidade de escuta.

Ao meu amigo, ex-orientador e agora colega de trabalho, Prof. Júlio César! Se não fosse por você eu não sei se teria chegado tão longe. Sou eternamente grato por todos os ensinamentos, escutas e cuidado que você teve comigo e com os meninos. Vamos voar longe, mas com certeza voaremos juntos! Agradeço também ao Prof. Eduardo por todo o apoio e tantos ensinamentos que nos proporcionou durante esse trabalho. Me inspirei em sua competência desde o primeiro instante em que nos conhecemos! Ao Prof. Luciano, que topou esse desafio e contribuiu de uma maneira inigualável para que pudéssemos chegar até aqui. Há poucos como vocês três!

Às pessoas que estão e continuam comigo muito antes da minha história em Viçosa começar, e a todas aquelas que entraram a partir dessa aventura: Ebinho, Marcos, Lucas, Letícia, Tati, Sofia, Kalango, Vidal, República Diboréstia, República Faixa de Gazza, ao grup R.H.T., Juliana, Salime, Ramiro, Mariana, Carmém, Luiz Bispo, aos queridos e queridas do forró viçosense e sanjoanense e tantos outros que não caberiam aqui. Eu amo todas e todos vocês!

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos.

“Entre tantos amores, o próprio”.

(Autor desconhecido)

RESUMO

REZENDE, Jhon Heider Domingos, M.Sc., Universidade Federal de Viçosa, junho de 2021. **Volatilidade implícita do petróleo: análises em aprendizado de máquina e métodos tradicionais.** Orientador: Luciano Dias de Carvalho. Coorientadores: Eduardo José da Silva Luz e Júlio César Araújo da Silva Júnior.

A volatilidade implícita é a medida utilizada em mercados financeiros para se calcular o risco futuro envolvido em um dado conjunto de ativos. Realizar previsões eficientes deste índice ao longo do tempo contribui para que se aprimore o processo de tomada de decisão de investidores. Este trabalho busca compreender se a rede neural recorrente profunda Deep Convolutional Long-short Term Memory (DCLSTM) produz previsões mais precisas da volatilidade implícita do mercado financeiro de petróleo norte-americano quando comparada aos modelos ARMA, ARMA-GARCH e VAR em econometria de séries de tempo. Os modelos selecionados como melhores candidatos de arquitetura da rede neural contaram todos com camada convolucional e foram treinados com adicional de um algoritmo genético. A rede neural apresentou performance de previsão da volatilidade implícita superior em todos os contextos quando comparada aos demais modelos.

Palavras-chave: Petróleo. DCLSTM. Macroeconomia. WTI. Séries de Tempo.

ABSTRACT

REZENDE, Jhon Heider Domingos, M.Sc., Universidade Federal de Viçosa, June, 2021. **Oil implied volatility: machine learning and traditional analyses.** Adviser: Luciano Dias de Carvalho. Co-advisers: Eduardo José da Silva Luz and Júlio César Araújo da Silva Júnior.

Implied volatility is the measure used in financial markets to calculate the future risk involved in a given set of assets. Making efficient forecasts of this index over time helps to improve the decision-making process of investors. This paper seeks to understand whether the Deep Convolutional Long-short Term Memory (DCLSTM) deep recurrent neural network produces more accurate predictions of the implied volatility of the US oil financial market when compared to the ARMA, ARMA-GARCH and VAR models in series econometrics of time. The models selected as the best neural network architecture candidates all had a convolutional layer and were trained with the addition of a genetic algorithm. The neural network showed superior implicit volatility prediction performance in all contexts when compared to other models.

Keywords: Oil. DCLSTM. Macroeconomics. WTI. Time series.

LISTA DE ILUSTRAÇÕES

Figura 1 – Cadeia LSTM	32
Figura 2 – Arquitetura de uma rede recorrente DLSTM.....	36
Figura 3 – Volatilidade Implícita e opções do United States Oil Fund.....	42

LISTA DE TABELAS

Tabela 1 – Estatística descritiva das variáveis em primeira diferença.....	43
Tabela 2 – Resultados das regressões utilizando MQO.....	44
Tabela 3 – Resultado das regressões quantílicas.....	47
Tabela 4 – Seleção dos melhores modelos pelo GA para a Eq. 3.1.....	49
Tabela 5 – Seleção dos melhores modelos pelo GA para a Eq. 3.1.....	49
Tabela 6 – Seleção dos melhores modelos pelo GA para a Eq. 3.1.....	50
Tabela 7 – Comparação entre os modelos para a Eq. 3.1.....	51
Tabela 8 – Comparação entre os modelos para a Eq. 3.2.....	52
Tabela 9 – Comparação entre os modelos para a E1. 3.3.....	53

SUMÁRIO

1	INTRODUÇÃO.....	11
1.1	O problema e sua importância.....	11
1.2	Objetivos.....	15
1.2.1	Objetivos Gerais.....	16
1.2.2	Objetivos Específicos.....	16
1.3	Hipóteses.....	16
2	REFERENCIAL TEÓRICO.....	17
2.1	Mercados Oligopolistas.....	17
2.2	Volatilidade Implícita e retorno.....	18
2.3	Volatilidade Implícita e retorno no mercado de petróleo.....	21
3	MÉTODOS.....	24
3.1	Dados.....	24
3.1.1	Volatilidade implícita e retornos – series históricas.....	24
3.1.2	Pré-processamento dos dados.....	25
3.2	Equações e hipóteses.....	30
3.3	Modelo LSTM.....	31
3.4	A rede neural recorrente LSTM profunda.....	35
3.5	Modelo ARMA.....	37
3.6	Modelo de Vetores Autorregressivos (VAR).....	37
3.7	Testes para diagnóstico de previsão.....	38
3.7.1	Erro Médio Absoluto (MAE).....	39
3.7.2	Erro Quadrático Médio (MSE) e Raiz do Erro Quadrático Médio (RMSE) ..	39
4	RESULTADOS EMPÍRICOS.....	41
4.1	Estatísticas descritivas.....	41
4.2	Resultados dos Mínimos Quadrados Ordinários.....	43
4.3	Resultados das regressões quantílicas.....	45
4.4	Resultados da rede neural DCLSTM.....	47
4.5	Comparação entre os modelos.....	51
	CONCLUSÕES	54
	REFERÊNCIAS.....	57

1. INTRODUÇÃO

1.1 O problema e sua importância

A volatilidade implícita é a medida utilizada em mercados financeiros para se calcular o risco futuro envolvido em um dado conjunto de ativos. Realizar previsões eficientes deste índice ao longo do tempo contribui para que se aprimore o processo de tomada de decisão de investidores. Este trabalho busca compreender se a rede neural recorrente profunda Deep Convolutional Long-short Term Memory (DCLSTM) produz previsões mais precisas da volatilidade implícita do mercado financeiro de petróleo norte-americano quando comparada a modelos tradicionais em econometria de séries de tempo.

Os choques de preços do petróleo observados ao longo do século XX evidenciaram o grau de dependência da dinâmica econômica global em torno do comportamento financeiro desta commodity. A série de conflitos geopolíticos constatados no oriente médio a partir da década de 1950 (e.g. da Crise de Suez em 1956 à Guerra do Golfo em 1991) resultou em cenários politicamente instáveis nos países desta região, promovendo restrições de oferta e consequente elevação no preço do petróleo¹. Durante a crise do subprime, o petróleo do tipo West Texas Intermediate (WTI) oscilou de 145 a menos de 30 dólares por barril entre julho e dezembro de 2008, demonstrando maior probabilidade na ocorrência de eventos extremos (risco de cauda) comparado ao mercado de ações. Eventos como estes expõem a importância de análises financeiras para o setor petrolífero (ABOURA; CHEVALLIER, 2013).

O petróleo está presente em mais de 50% da parcela total de commodities comercializadas em portfólios financeiros ao redor mundo (BAŠTA; MOLNÁR, 2018). O comportamento de variáveis financeiras do setor petrolífero exerce impacto direto sobre os investimentos das empresas (HENRIQUES; SADORSKY, 2011) e variáveis macroeconômicas (NARAYAN et al., 2014; RAFIQ et al., 2009), além de impactar balanças comerciais de países os quais possuem elevado grau de dependência monetária relacionada ao petróleo (BAŠTA; MOLNÁR, 2018). Assim, desdobramentos

¹ Para uma abordagem mais detalhada sobre a relação entre conflitos políticos e preço do petróleo, ver, por exemplo, Noguera-Santaella (2016).

econômicos advindos de mudanças nos níveis de oferta e demanda do petróleo revelam-se igualmente relevantes para compreender a importância desse mercado em contexto global.

A pandemia do novo coronavírus constatada oficialmente em março de 2020 contribuiu para oscilações no nível de petróleo disponível nos mercados globais. As restrições impostas pelos países com o intuito de frear o nível de contágio do vírus ao redor do mundo colapsou a demanda mundial por petróleo, reduzindo seu preço e aumentando o nível de incerteza do investidor quanto ao setor (RAJPUT et al., 2020). Aliado à COVID-19, as tensões políticas envolvendo Rússia e Arábia Saudita a despeito da partilha de mercado da commodity no mercado global resultaram em quedas recordes do preço do petróleo nos mercados internacionais (BIELISZCZUK et al., 2020). Nota-se, portanto, a sensibilidade que o mercado de petróleo possui em relação a choques externos e às expectativas dos agentes.

A volatilidade implícita é usada como proxy da percepção de risco (incerteza) dos agentes sobre os ativos financeiros. A teoria de precificação de ativos postula que a volatilidade implícita desempenha significativa capacidade preditiva sobre os retornos de ações, além de mensurar oscilações na volatilidade esperada dos ativos (MERTON,

1973) e na variância do prêmio de risco (DRECHSLER; YARON, 2011; DRECHSLER, 2013). Evidências empíricas na literatura sugerem uma relação negativa entre risco e retorno de ativos financeiros e são discutidas a partir de três hipóteses principais: alavancagem, feedback e comportamental.

A hipótese de alavancagem, inicialmente proposta por Black (1976), demonstra que variações negativas nos retornos (diferença entre o preço de hoje e o preço de ontem) de ativos financeiros incrementavam o nível de alavancagem financeira das empresas, ou seja, seu endividamento. Sob a ótica dos investidores, o risco de investimentos em empresas que apresentam crescimento de suas dívidas é maior, o que resulta em um nível de volatilidade maior de seus ativos. Por sua vez, a hipótese de feedback traz uma inversão dos fatos ao propor que variações positivas no nível de volatilidade dos ativos promoviam retornos negativos.

Ao formularem a hipótese de feedback, Poterba e Summers (1984) e Campbell e Hentschel (1991) indicam que a volatilidade pudesse ser um sinal de ajuste para o mercado. Assim, os autores postulam que se variações positivas contemporâneas na volatilidade produzem um acréscimo nos retornos futuros

esperados dos ativos, os preços desses ativos e, consequentemente, seus retornos, apresentariam queda no presente para se ajustar às mudanças futuras. Hibbert et al. (2008) argumenta que a literatura tem sido mais propensa a adotar as pressuposições teóricas da hipótese de alavancagem, dado que a hipótese de feedback envolve uma série de constatações econômicas e de dividendos dos ativos para que se estabeleça um cenário de validação da relação negativa entre risco e retorno. Além disso, a hipótese de feedback considera apenas o cenário de longo prazo desta relação.

Por sua vez, Low (2004) indica que a hipótese de alavancagem é uma explicação fraca para a relação risco-retorno. O autor identifica que a relação das variáveis é assimétrica e apresenta não-linearidade, melhor descrita como uma curva em formato de "S" com inclinação para baixo. Além de postular que os efeitos assimétricos de perdas nos ativos sejam advindos de uma causa comportamental no mercado, não dizendo respeito tão somente à percepção dos investidores quanto ao nível de endividamento das empresas. Apesar deste pressuposto, o autor não relaciona seus resultados com conceitos advindos de teorias comportamentais econômicas do investidor.

A conexão entre conceitos de economia comportamental como tentativa de se aprofundar na relação entre risco e retorno no mercado acionário é realizada por Hibbert et al. (2008). Os autores investigaram a relação de curto prazo entre o retorno do S&P 500 da Nasdaq 100 e variações no índice de volatilidade implícita do mercado em frequências diárias e intra diárias. Constatou-se que as hipóteses de alavancagem e feedback não explicam eficientemente a relação risco-retorno, propondo que a relação negativa entre volatilidade implícita e retorno seja consistente com os conceitos de representatividade, afeto e viés de extração encontrados na teoria de finanças comportamentais. A relação também se mostrou assimétrica e não-linear, tal como nos resultados de Low (2004).

Grande parte dos estudos que abordaram a relação entre risco e retorno, e discutiram as teorias subjacentes a esta relação, consideraram apenas análises com séries históricas do mercado acionário, tais como os apresentados até então. Contudo, é sabido que o mercado financeiro apresenta diversas categorias de investimento e instrumentos à disposição dos investidores. Dentre estas categorias, encontram-se os mercados futuros, onde as commodities são negociadas em formato de opções. Esta configuração do mercado abre margem para se verificar se as pressuposições teóricas de risco e retorno estabelecidas para o mercado acionário

americano são igualmente válidas para os ativos relacionados ao petróleo no mercado futuro.

O mercado de petróleo dispõe do CBOE Crude Oil Volatility Index (OVX)², responsável por mensurar a volatilidade implícita dos agentes sobre o petróleo do tipo West Texas Intermediate (WTI) com base nas opções do United States Oil Fund (USO). As séries históricas dessas variáveis possibilitaram com que a relação risco-retorno do petróleo fosse aprofundada na literatura por meio do uso de técnicas paramétricas e não paramétricas a fim de realizar previsões destas variáveis, assim como compreender em maior grau características estruturais das séries, tal como a presença de assimetria e não linearidade dessa relação (ABOURA; CHEVALLIER, 2013; PADUNGSAKSAWASDI; DAIGLER, 2014; AGBEYEGBE, 2015; SILVA JUNIOR, 2017).

Estudos que buscam realizar previsões e análises estruturais de séries de tempo fazem constante uso de modelos como os da família GARCH (SUN; YU, 2020) e da classe ARMA (SEDLÁKOVÁ, 2020; CHENG et al., 2020). Apesar de se constatar um grande número de análises que lançam mão destas abordagens metodológicas, desde a década de 1990 métodos em aprendizado de máquina têm se mostrado igualmente eficazes em descrever relações econômicas e financeiras (ZHANG et al., 1998). A principal justificativa para essa mudança de eixo metodológico ao longo do tempo concentra-se no fato de que modelos clássicos como ARMA e GARCH tem capacidade de atuação limitada sobre dados do gênero. Sagheer e Kotb (2019) postula que séries de tempo reais que apresentam alta complexidade e ruído não podem ser descritas por equações de modelos lineares como estes, já que essa formas funcionais dificultam a adequação dos modelos aos dados por meio de fortes restrições matemáticas.

Dadas as limitações das abordagens econométricas clássicas descritas anteriormente, modelos como support vector machine (YU et al., 2017b), least squares support vector regression (YU et al., 2017a) e (YU et al., 2016), programação genética (KABOUDAN, 2001) e de redes neurais artificiais (JAMMAZI; ALOUI, 2012) foram utilizados para realizar previsões do preço do petróleo no mercado financeiro. Este

² O OVX é construído a partir de base metodológica análoga ao *Volatility Index* da *Chicago Board Options Exchange* (VIX) e estima o nível de volatilidade implícita dos agentes com base nas opções do índice S&P 500 (ABOURA; CHEVALLIER, 2013).

trabalho, por sua vez, considera o objetivo de prever a variação da volatilidade implícita no mercado financeiro de petróleo, e não seu preço, por meio de uma rede neural recorrente profunda.

Modelar a variação e não a volatilidade em nível justifica-se a partir de três considerações. Primeiro, pesquisas quantitativas buscam compreender como mudanças na volatilidade esperada influenciam mudanças na avaliação de risco dos ativos (FLEMING et al., 1995; LOW, 2004; HIBBERT et al., 2008). Segundo, há alta probabilidade de se incorrer em regressões espúrias caso as estimativas considerem a volatilidade implícita e o preço em nível, já que ambas as variáveis em nível se comportam como um passeio aleatório. Terceiro, a volatilidade implícita diária em nível apresenta elevada autocorrelação capaz de afetar inferências em amostras finitas (FLEMING et al., 1995).

Realizar previsões da variação da volatilidade implícita do mercado de petróleo norte-americano por meio de novas abordagens metodológicas e discutir as pressuposições teóricas da relação risco-retorno a partir dessas previsões contribui para a literatura sob três aspectos. O primeiro deles sob a ótica acadêmico-científica, que usufrui de novos avanços quanto ao uso de modelos lineares e não lineares nas especificidades do campo de séries de tempo em finanças. Segundo, os resultados alcançados pelas modelagens do estudo fornecem novas evidências empíricas que podem ser utilizadas no processo de tomada de decisão dos investidores. Por fim, a pesquisa contribui para que novos estudos acerca do mercado global de petróleo sejam realizados, além de aprimorar o uso de seus instrumentos financeiros, tal como ocorre para o mercado de ações.

1.2 Objetivos

Dante do exposto nesta introdução, os objetivos do presente estudo podem ser divididos em gerais e específicos, a saber:

1.2.1 Objetivos gerais

Verificar se as previsões da variação da volatilidade implícita do petróleo WTI (OVX) por meio da rede neural Deep Convolutional Long-short Term Memory possuem maior acurácia e menor erro quando comparadas às previsões dos modelos ARMA, ARMA-GARCH e VAR.

1.2.2 Objetivos específicos

Discutir se os resultados obtidos pelos modelos de séries de tempo estão em consonância com as pressuposições teóricas acerca da relação risco-retorno abordadas na literatura.

1.3 Hipóteses

Duas hipóteses são levantadas para o presente trabalho:

- 1) A rede neural recorrente DCLSTM produz previsões com menor erro e maior acurácia quando comparadas às previsões dos modelos ARMA, ARMA-GARCH e VAR;
- 2) Os resultados da rede neural recorrente DCLSTM produzem coeficientes com sinais que confirmam uma relação negativa entre risco e retorno.

O restante do trabalho é organizado tal como se segue: o Capítulo 2 apresenta uma revisão dos estudos que abordaram a relação risco-retorno para o mercado de petróleo. O Capítulo 3 apresenta os modelos utilizados na pesquisa, tal como as séries históricas utilizadas no estudo e as transformações realizadas nos dados. A discussão dos resultados alcançados e a comparação entre cada modelo encontram-se Capítulo 4.

Por fim, as conclusões dos autores são expostas na Capítulo 5.

2. REFERENCIAL TEÓRICO

2.1 Mercados Oligopolistas

A dinâmica de comportamento do mercado mundial de petróleo é estruturalmente organizada sob as características de um oligopólio. Nesse tipo de competição, a concorrência constitui-se de poucas empresas produtoras de um bem comum que detêm a capacidade de controle de preço e restrição à entrada de novas empresas no mercado, sendo o grau de poder monopólio definido em função de uma estratégia cooperativa ou competitiva entre os integrantes do grupo (MAS-COELL et al., 1995).

A especificidade para o mercado mundial de petróleo diz respeito à formação de cartéis. A Organização de Países Exportadores de Petróleo (OPEP) é um cartel constituído por treze países membros, sendo responsável pela maior parte da produção e distribuição de petróleo ao redor do mundo. Diferentemente dos monopólios, os cartéis não detêm controle absoluto sobre seu mercado característico, sendo necessária a ponderação sobre como a decisão de preço tomada pelo cartel afeta os demais agentes que não integram o grupo. Além disso, os cartéis são caracterizados por instabilidade constante, já que há sempre a possibilidade de tentativa de rompimento da coalizão entre os membros como meio de obter lucros individualmente maiores (VARIAN, 2017)

Um mercado caracterizado enquanto oligopólio pode apresentar produtos que sejam ou não diferenciados entre as empresas que o constituem, importando apenas que a produção seja centralizada em um número pequeno de empresas. No longo prazo, os lucros para os membros do oligopólio podem ser substanciais, dado que barreiras à entrada dificultam e/ou impossibilitam a inserção de novas empresas no mercado (VARIAN, 2017). Essas e outras características do mercado oligopolista são mensuradas a partir de modelos matemáticos, tal como o modelo estático proposto por (BERTRAND, 1883) considera uma estrutura de competição oligopolística em que há escolhas simultâneas de preço pelas firmas com retornos constantes de escala.

Contudo, o presente trabalho não visa o aprofundamento de modelos específicos de oligopólio. Paralelamente, para o entendimento do arcabouço teórico que engloba a pesquisa, faz-se necessário a apresentação das bases teóricas necessárias ao entendimento do mercado de petróleo dentro dessa estrutura de

competição, tal como a demonstração de métodos distintos para problemas comuns ao setor petrolífero.

2.2 Volatilidade implícita e retorno

O risco é definido como a probabilidade de se incorrer em perdas ou a probabilidade de que eventos ruins ocorram. Os riscos podem estar atrelados à vida cotidiana das pessoas, ao mesmo tempo em que são associados à estruturas maiores, como os mercados financeiros. Assim, medidas de risco existem para que o processo de tomada de decisões dos indivíduos mitigue as chances de se incorrer em prejuízos. No âmbito de finanças, a proxy para o risco futuro de um ativo é conhecida como volatilidade implícita e a partir dela é possível compreender o comportamento que o retorno de um ativo financeiro apresenta ao longo do tempo.

A relação entre retorno e volatilidade em ativos financeiros é abordada na literatura a partir de diferentes teorias e abordagens. Black (1976) foi o primeiro a postular uma relação negativa entre essas duas variáveis, destacando que variações negativas nos preços das ações contribuíam para o aumento do nível de endividamento das empresas. Consequentemente, os investidores consideravam essas empresas mais arriscadas e seus preços se tornavam mais voláteis. Esta teoria foi documentada como teoria da alavancagem. Em contraposição, Christie (1982) e Schwert (1989a) destacam as dificuldades em se mensurar dados reais relativos à alavancagem das empresas, limitando a avaliação dos efeitos que retorno e volatilidade exercem sobre o seu nível de endividamento.

A continuidade de análises entre retorno e volatilidade de ativos possibilitou novas constatações acerca da dinâmica dessas variáveis. Como postulado por Bollerslev et al. (2009), além das observações quanto a relação negativa, outros estudos demonstraram a presença de assimetria entre retorno e volatilidade, evidenciando que informações negativas sobre os mercados exerciam impacto maior na volatilidade implícita quando comparado às informações positivas.

A partir da dificuldade em se mensurar o grau de endividamento das empresas, as formulações teóricas sobre a relação de retorno e volatilidade implícita partiram de outro pressuposto. Campbell e Hentschel (1991) sugeriram teoricamente que choques positivos na volatilidade implícita das ações resultavam em queda nos retornos. Segundo os autores, à medida em que as expectativas sobre os retornos

futuros de ações de uma empresa sobem em decorrência do aumento de sua volatilidade, os preços contemporâneos destas ações, tal como seus retornos, caem para que ocorra um ajustamento a esta mudança de expectativa no futuro. Assim, o aumento da volatilidade produz retornos negativos das ações no presente. Essa postulação ficou conhecida como hipótese feedback, também sugerida anteriormente por French et al. (1987).

Poterba e Summers (1984) destacam que a hipótese de feedback é dependente da existência de um prêmio de risco temporal para que exista uma conexão entre mudanças na volatilidade que afetem os retornos das ações. Tal como inferido na hipótese de alavancagem, a hipótese feedback também apresenta assimetria na relação de retorno e volatilidade implícita. A presença de assimetria é justificada a partir da observação de que o prêmio pelo risco tende a intensificar mudanças negativas no retorno esperado, ao passo em que ameniza as positivas (SILVA JUNIOR, 2017). Frequentemente a hipótese é testada a partir de regressões lineares.

A partir das pressuposições teóricas moldadas para explicar a relação negativa entre retorno e volatilidade implícita, a literatura apresentou uma diversidade de métodos a fim de fornecer evidências empíricas para a dinâmica desta relação. Como destacado por Hibbert et al. (2008), grande parte destes estudos consideram separadamente apenas uma das teorias produzindo resultados empíricos diversos.

A pesquisa de Schwert (1989b) aborda a relação de risco-retorno utilizando dados diários do S&P 500 destacando que a teoria de alavancagem oferece pouco suporte teórico para os resultados empíricos obtidos. Da mesma forma, Bollerslev e Zhou (2006) ressalta que a magnitude do impacto que o decréscimo de preços de ativos exerce sobre a volatilidade implícita é muito amplo para que este seja explicado exclusivamente por flutuações de alavancagem financeira. Estes resultados são semelhantes aos encontrados por Figlewski e Wang (2000), Andersen et al. (2001), Dennis et al. (2006), Kim e Kon (1994) e Tauchen et al. (1996) os quais destacaram que o efeito de alavancagem é visivelmente maior para índices de mercado (conjunto de ações) do que para ações quando analisadas isoladamente.

Enquanto os estudos apresentados focaram em verificar a veracidade de uma relação negativa entre retorno e volatilidade implícita, outros trabalhos reuniram esforços para verificar também a presença de assimetria. A partir de séries do mercado acionário japonês, Bekaert e Wu (2000) constataram que a hipótese de

feedback é mais propensa a produzir uma resposta assimétrica quando comparada a hipótese de alavancagem. As análises dos autores consideraram separadamente uma formulação teórica fraca e outra formulação teórica restrita para a relação das variáveis. A formulação restrita considera que retornos e volatilidade possuem uma relação negativa, enquanto a formulação fraca afirma que a relação negativa existe para os retornos e a volatilidade implícita (futura).

O uso da volatilidade implícita (VIX) como proxy para a volatilidade esperada tem sua primeira abordagem empírica realizada por Fleming et al. (1995). Os autores definem um modelo de regressão linear que considera simultaneamente os retornos do mercado e a volatilidade implícita em defasagens e passos futuros diversos. O modelo assume uma taxa de ajustamento de cerca de 40% descartando períodos de grandes quebras nas séries. Os resultados demonstraram que apenas o coeficiente de retornos contemporâneos foi estatisticamente significativo (e negativo). As demais variáveis explanatórias do modelo foram marginalmente significativas ou estatisticamente não significativas.

A contraposição às hipóteses de alavancagem e feedback surge a partir do trabalho de Low (2004). Fazendo uso de dados de retorno do S&P500 e do índice VIX, os resultados do autor sugerem a presença de um alto coeficiente quadrático na variável independente de retorno, indicando a presença de não linearidade em formato de 'S' para descrever a relação de risco-retorno no mercado acionário. O autor define o 'medo' dos investidores como a razão pelos acréscimos no VIX (convexidade) e a 'exuberância' destes atores como o motivo pelos decréscimos no índice (concavidade).

Low (2004) destaca ainda que a presença de assimetria na relação ocorre devido a fatores comportamentais, sem propriamente citá-los, além de afirmar que a hipótese de alavancagem é, na melhor das hipóteses, fraca para explicar a relação entre retorno e volatilidade. O aprofundamento quanto a menção de Low (2004) é posteriormente abordada por Hibbert et al. (2008) ao analisarem dados intra diários do S&P 500 (Nasdaq 100) referentes a retorno e volatilidade implícita dos ativos.

Hibbert et al. (2008) constataram que as hipóteses de alavancagem e de feedback não ofereciam suporte empírico suficiente ao examinar a dinâmica de curto prazo da relação entre retorno e volatilidade implícita. Assim, os autores propuseram que o comportamento dos investidores determinava em maior medida a dinâmica dessa relação. A abordagem comportamental oferece consistência com os resultados

dos autores quando considera as heurísticas de representatividade, afeto e extração de informações presentes na teoria comportamental em finanças (HIBBERT et al., 2008). Além disso, a análise sob a ótica de dados intra-diários reforça a presença de assimetria na relação das variáveis, tal como confirma seu comportamento negativo ao longo do tempo.

2.3 Volatilidade implícita e retorno no mercado de petróleo

A expansão de estudos interessados em aprofundar o entendimento quanto à relação de risco e retorno de ativos financeiros direciona-se para além do mercado acionário. A partir de abordagens metodológicas diversas como regressão quantílica e cópula condicional, por exemplo, a literatura apresenta trabalhos empíricos dedicados a analisar a relação de risco e retorno relacionada a mercados como o de euro, ouro, dólar e petróleo (ABOURA; CHEVALLIER, 2013; PADUNSAKSAWASDI; DAIGLER, 2014; DAIGLER et al., 2014; AGBEYEGBE, 2015; SILVA JUNIOR, 2017).

A literatura apresenta ainda tentativas de se realizar previsões de índices de volatilidade do petróleo por meio de modelos de séries temporais, como os da família GARCH (CHEONG, 2009; AGNOLUCCI, 2009; KANG et al., 2009; WEI et al., 2010; ALOUI; MABROUK, 2010; MOHAMMADI; SU, 2010). Apesar deste não ser o objetivo do presente trabalho, aplicações do gênero possibilitaram a inserção de outras abordagens metodológicas alternativas a modelos estatísticos ao lidar com questões econômicas relacionadas ao petróleo, como o uso de algoritmos de aprendizado de máquina.

Recentemente, o uso de redes neurais recorrentes (RNR) ganhou espaço na literatura dedicada a prever o comportamento e a estrutura de séries financeiras do mercado de petróleo. A rede neural recorrente baseada em células long short term memory (LSTM), proposta por Hochreiter e Schmidhuber (1997), é frequentemente usada para este objetivo dada sua capacidade de modelar dados em sequência, facilidade na incorporação de variáveis exógenas e extração de características dos dados que sejam relevantes no processo de modelagem (WU et al., 2019).

A capacidade de modelagem sequencial de células LSTM permite que se crie um estado de memória desempenhando o papel de retroalimentação entre a entrada e a saída dos dados (HOCHREITER; SCHMIDHUBER, 1997). Isso possibilita com que as informações inseridas em cada LSTM sejam associadas com as informações

subsequentes, ou seja, o passo anterior prevê o passo seguinte a partir da identificação de um padrão de comportamento. A partir dessa arquitetura a LSTM é capaz de atuar em tarefas de classificação de vídeo e imagem, reconhecimento de fala e tradução de textos. Isso demonstra a habilidade de aprendizado da rede sobre dependências entre sistemas não-lineares complexos, o que ocorre com séries de tempo em finanças.

A célula de rede LSTM destaca-se das demais redes neurais recorrentes dada a existência de portões que auxiliam o seu processo de retroalimentação. Estes portões permitem que informações passadas sejam ponderadas durante o processo de memória da rede, adicionando, descartando ou atualizando observações a fim de que se alcance maior acurácia nas previsões da sequência (HOCHREITER; SCHMIDHUBER, 1997). O processo de entrada da informação, sua ponderação e posterior saída ocorre sob operações (funções) não lineares e pode ser resumido como uma camada da célula de rede LSTM.

Contudo, séries de tempo construídas a partir de dados do mundo real frequentemente apresentam ruído, além de caracterizarem-se por uma estrutura de alta complexidade e serem multi-dimensionais. Essas características dificultam o processo de modelagem a partir de redes como as que utilizam células LSTM devido ao número limitado de operações não-lineares em sua arquitetura, o que impossibilita o alcance de um grau considerável de acurácia da rede ao modelar a dinâmica complexa e desconhecida de dados do gênero (LÄNGKVIST et al., 2014a).

O impasse relativo a incapacidade de se modelar dados com alta complexidade e ruído é contornado por meio de modelagens de aprendizagem profunda, como a proposta por Sagheer e Kotb (2019). Ao considerar que as operações de memória de unidades de rede LSTM originais ocorrem em um bloco, ou em uma camada, como descrito anteriormente, o termo "profunda" surge a partir do empilhamento de múltiplas camadas LSTM em uma mesma arquitetura de rede. Os autores desenvolvem uma abordagem de aprendizado profunda denominada como deep long short term memory (DSLTM) a fim de prever a produção de campos de petróleo.

A abordagem proposta por Sagheer e Kotb (2019) conta ainda com a aplicação de um algoritmo genético a fim de optimizar a busca por arquiteturas de rede adequadas, assim como os valores dos hiper-parâmetros da DLSTM (número de épocas, taxa de aprendizagem, tamanho do mini-lote, etc). A rede tem sua

performance comparada com modelos referência do campo de redes neurais e de estatística tradicional, a saber: ARIMA, Deep Gated Recurrent Unit (DGRU), Nonlinear Extension for linear ARPS (NEA) e High Order Neural Network (HONN). Todos os modelos foram aplicados sobre dois casos distintos de produção de petróleo em campos na China e na Índia. A rede neural DLSTM apresentou resultados superiores em todos as comparações após os testes de acurácia e de erro de previsão.

Como visto, a aplicação de algoritmos de aprendizado de máquina para previsão de variáveis como preço, produção e volatilidade no mercado financeiro de petróleo é vasta. Contudo, a literatura ainda carece do uso de redes neurais recorrentes como suporte empírico na tarefa de aprimorar e discutir as teorias de volatilidade implícita e retorno aplicadas ao contexto do mercado financeiro de petróleo. Essa aplicação possibilita o entendimento de características estruturais da relação entre as variáveis como assimetria, não-linearidade e defasagens ótimas, assim como colabora para que novos métodos sejam utilizados na tentativa de se realizar melhores previsões da volatilidade implícita nos diversos mercados.

3. MÉTODOS

Esta seção apresenta os procedimentos metodológicos utilizados na pesquisa. As previsões da volatilidade implícita do mercado de petróleo foram realizadas a partir da rede neural Deep Convolutional Long-short Term Memory (DCLSTM), junto com um mecanismo de atenção, e dos modelos ARMA, GARCH e VAR, tendo suas especificações apresentadas entre as subseções 3.3 e 3.6. As variáveis de risco e retorno utilizadas no estudo são construídas a partir das equações propostas por Low (2004) e Fleming et al. (1995) (seção 3.2) utilizando dados financeiros do mercado de petróleo norte-americano (seção 3.1). Após as transformações aplicadas aos dados, os modelos são utilizados para realizar previsões do índice OVX a fim de verificar qual destes métodos produz previsões com menor erro e maior acurácia. Os testes de raiz do erro quadrático médio (RMSE), erro médio absoluto (MAE) e erro quadrático médio (MSE) são apresentados na seção 3.7 e utilizados na pesquisa como métricas comparativas entre os modelos seguindo os procedimentos de Box et al. (1970).

3.1 Dados

A presente seção apresenta a série histórica da volatilidade implícita e das opções futuras de petróleo em suas escalas originais, assim como detalha o procedimento de pré-processamento dos dados enquanto etapa que antecede a modelagem em aplicações de modelos em séries temporais e em redes neurais artificiais.

3.1.1 Volatilidade implícita e retornos – séries históricas

Os dados utilizados no estudo são referentes aos valores de fechamento do índice OVX de volatilidade implícita do Chicago Board Exchange (Cboe) e das opções negociadas no United States Oil Fund (USO), ambos disponíveis na plataforma Yahoo Finance. Os dados refletem a dinâmica de precificação de futuros do petróleo no West Texas Intermediate (WTI). O recorte temporal das variáveis engloba o período de 10

de maio de 2007 a 31 de dezembro de 2019³ e possui um total de 3179 observações diárias. Deste total, 80% da amostra (2543) foram dedicados para o treinamento dos modelos e os outros 20% (636) dedicados a testar suas performances.

3.1.2 Pré-processamento dos dados

Os procedimentos a seguir foram realizados para os modelos de rede neural recorrente DCLSTM e de econometria clássica de séries de tempo (ARMA, ARMA-GARCH e VAR). A modelagem empírica a partir desses métodos estabeleceu-se em especificações univariadas - valores defasados da volatilidade implícita como preditores da volatilidade implícita contemporânea - e multivariadas, a fim de testar empiricamente pressuposições dadas pelas equações 3.1 e 3.2 na subseção 3.2.

3.1.2.1 Filtro de média móvel

A aplicação de filtros de média móvel nas séries de volatilidade implícita e retornos das opções no trabalho teve como objetivo suavizar as flutuações das séries e identificar a tendência das observações diárias ao longo do tempo. A estratégia contribuiu para a redução da presença de ruído nos dados e possibilitou que as observações fossem inseridas na etapa de treinamento da rede neural, dado que o algoritmo necessita de taxas de aprendizado baixas durante o processo de treino da rede (CHAKRA et al., 2013; SAGHEER; KOTB, 2019)⁴. Além disso, o filtro de média móvel possibilitou com que os dados de frequência diária fossem suavizados em frequência menores (característica de low pass), removendo flutuações de grandes proporções na série a partir do cálculo de uma média ponderada em uma faixa de cinco pontos passados na série⁵. Como destacado por Smith et al. (1997), o filtro de média móvel é, provavelmente, o melhor e mais simples filtro na tarefa de suavização

³ O início da amostra é referência à implementação do índice de volatilidade implícita OVX no mercado financeiro. Estas e outras informações sobre o índice podem ser consultadas na página oficial da bolsa americana na internet (CBOE, 2020).

⁴ O filtro de média móvel também foi aplicado aos dados utilizados pelos demais modelos para que os resultados pudessem ser comparados entre si.

⁵ Considerando que o mercado opera, em média, por cinco dias durante a semana, uma faixa com esta mesma extensão é considerada satisfatória para o cálculo da média ponderada.

de dados em séries de tempo e alcançou os propósitos para os dados do presente estudo.

3.1.2.2 Estacionarização da série

O comportamento histórico de dados em séries de tempo pode ser decomposto em componentes de tendência, sazonalidade e irregularidade. O entendimento acerca da trajetória dinâmica das observações a partir dessa decomposição aprimora a capacidade preditiva dos modelos e possibilita a modelagem das observações ao longo do tempo (ENDERS, 2012). Entretanto, modelagens em séries de tempo requerem que os dados sejam estacionários, o que não ocorre com dados financeiros. Dessa forma, é necessário que se aplique transformações nas observações para que estas cumpram a condição de estacionariedade. A medida mais comum para se alcançar a condição de estacionariedade consiste na diferenciação dos dados, subtraindo a observação do tempo passado ($t-1$) daquela do tempo presente (t). Ao assumir que as equações 3.1 a 3.2 estejam especificadas a partir da variação da volatilidade implícita ($\Delta OVX = OVX_t - OVX_{t-1}$) e do retorno logarítmico das opções USO ($[\ln(USOt) - \ln(USOt-1)]$ - onde $USOt$ representa a opção X negociada no período t), a condição de estacionariedade é atendida.

3.1.2.3 Normalização de escala dos valores

Redes neurais demandam que os dados sejam normalizados em uma escala específica de valores a fim de serem corretamente interpretados pela função de ativação da rede no formato de entrada (x) e saída (y). A função de ativação padrão da unidade de rede LSTM é uma tangente hiperbólica (\tanh), que aceita valores entre uma faixa de -1 a 1. Após as estimativas, os dados são convertidos novamente para sua escala original a fim de comparar os valores estimados e aqueles observados. Esse procedimento é feito a partir da função `scaler.fit_transform` do módulo `sklearn.preprocessing` no Python.

3.1.2.4 Treinamento e busca por modelos ótimos

Encontrar arquiteturas e hiper-parâmetros para modelos de redes neurais é um grande desafio hoje na literatura. Na maioria das vezes, o processo é feito de forma manual e empírica, o que nem sempre garante boas arquiteturas de redes (ELSKEN et al., 2019). Hoje existe um interesse grande em métodos para busca de arquiteturas de redes neurais. Neste trabalho, assim como em (SAGHEER; KOTB, 2019), utilizamos de um algoritmo genético para a tarefa de escolha das arquiteturas e alguns hiper-parâmetros. Contudo, propomos uma busca mais ampla e exploramos outras operações, como blocos de convolução, mecanismo de atenção e redes recorrentes bi-laterais.

Algoritmos genéticos são uma classe de algoritmos evolutivos desenvolvidos com base na teoria de seleção natural de Darwin (WHITLEY, 1994; DAVIS, 1991). Os algoritmos genéticos simulam a sobrevivência de indivíduos mais aptos entre uma população, ao longo das gerações. Para tal, aplica-se algumas operações nos indivíduos como seleção, cruzamento, mutação e elitismo.

Cada indivíduo da população é representado por um cromossomo, que codifica um genótipo com uma possível solução ao problema. Para este trabalho, cada indivíduo representa um modelo de rede neural treinado, com diversas arquiteturas, funções objetivos, tamanho do sinal de entrada, etc. A avaliação da aptidão dos indivíduos se dá apenas no conjunto reservado para o treino dos modelos, de forma que uma parte dos dados de treinamento (20%) é reservada para computar a aptidão por meio de uma métrica (RMSE) e o restante para criação dos modelos.

Cada iteração do algoritmo é chamada de geração. A cada geração, os indivíduos se cruzam, forçando reprodução. Ocasionalmente, os descendentes podem sofrer mutações para garantir maior diversidade na população. Esta mutação é um parâmetro do algoritmo genético. Indivíduos mais adaptados passam por uma seleção e os melhores tem maior chance de transmitir seus genes para gerações futuras, criando elitismo. A seleção dos melhores indivíduos ocorre por meio de um torneio. O toolbox DEAP (RAINVILLE et al., 2012) é utilizado aqui para implementação do algoritmo genético e suas operações.

Neste trabalho, o genoma dos indivíduos é codificado pelos parâmetros:

1) Número de Épocas (ep): quantidade de ciclos que a rede neural recorrente vai percorrer entre os dados durante seu processo de treinamento. O intervalo de busca pelo algoritmo genético está entre 100 e 500 épocas;

2) Camada convolucional (cc): define o aprendizado por meio de filtros não lineares que percorrem separadamente as informações de entrada da rede e auxiliam no processo de identificação de padrões da série de tempo ((LECUN et al., 2015a)). Elas são formadas por um banco de 30 filtros de kernel 5x1, seguido por um banco de 30 filtros com kernel 3x1, seguido por um banco de 2 filtros de kernel 1x1 com padding e ativação de Relu. Os parâmetros de busca pelo algoritmo genético foram definidos entre 0 e 10. Valores menores que 5 não incluem uma camada convolucional na rede e valores maiores ou iguais a 5 incluem uma camada convolucional na rede;

3) Número de camadas LSTM (lstm): número de modelos LSTM utilizados na rede neural profunda. O intervalo de busca pelo algoritmo genético está entre 1 e 3 camadas; 4) Bidirecional (bi): define se a LSTM será bidirecional ou não. Em uma LSTM bidirecional, a informação percorre a rede do início ao fim e, posteriormente, do fim ao início a fim de intensificar o processo de aprendizado ((SCHUSTER; PALIWAL, 1997; GRAVES; SCHMIDHUBER, 2005)). O intervalo de busca pelo algoritmo genético está entre 0 10. Valores menores que 5 indicam a opção 'Não' para a bidirecionalidade da LSTM e valores maiores ou iguais a 5 indicam a opção 'Sim' para a birecionalidade da LSTM;

4) Mecanismo de atenção (att): este mecanismo permite a seleção de informações importantes em dados ruidosos . A usabilidade deste mecanismo em modelos sequenceto-sequence como as estruturas de rede em LSTM tem grande vantagem ao considerar que modelos tradicionais de sequence-to-sequence comprimem toda a informação a ser codificada no decoder em uma única representação. Esta situação é conhecida como problema de gargalo. Assim, ao fazer uso do mecanismo de atenção, a ideia é construir uma conexão direta entre os principais insumos necessários para gerar um produto específico e gerar o score de atenção. Este score define o quanto importante uma informação é dentro de uma representação para o produto pretendido. Ao fim, é realizado um produto interno entre o vetor dos dados de entrada e o vetor do decoder sendo que a distribuição da atenção dada a informações importantes forma uma distribuição de atenção e a partir dela produz-se uma soma ponderada dos estados ocultos. Esta, por sua vez fornece as

informações (insumos) com maior nível de atenção para gerar as previsões de uma série (CHOROWSKI et al., 2015; VASWANI et al., 2017).

5) Número de neurônios (neuronio): responsáveis por atribuir um peso para as informações de entrada da rede. Eles são ativados por uma função (geralmente a função sigmoide, tal como apresentado na seção 3.3) e definem o processo de memorização da rede neural. A depender da complexidade da série temporal, a quantidade de neurônios inserida na rede pode ser maior ou menor (RIO et al.). Neste caso, a quantidade de neurônios é definida para cada camada LSTM escolhida pelo algoritmo genético, tal que cada uma delas é suscetível a um intervalo de busca entre 2 e 64 neurônios.

6) Dropout (drop): parâmetro responsável por diminuir a probabilidade da rede neural incorrer em ajuste excessivo (overfitting) nos dados. Uma situação de ajuste excessivo ocorre quando a rede neural perde sua capacidade de generalização. Isso resulta em um bom desempenho de predição da rede apenas sobre o conjunto de teste, mas não o conjunto de validação do modelo (VASCO, 2020). Neste trabalho, definimos o dropout entre 0 e 50% a ser definido pelo algoritmo genético;

7) Função de perda (lost): função que minimiza os erros incorridos pela rede neural durante seu processo de treinamento. As funções Mean Absolut Error (MAE) e Mean Square Error (MSE) foram as opções de função de perda para o algoritmo genético. Valores menores que 5 definem a MAE como função de perda e valores maiores ou iguais a 5 definem a MSE;

8) Defasagens (def): quantidade de períodos passados que a rede neural define para calcular o valor ótimo de suas previsões⁶. A janela de defasagem aqui proposta concentra valores entre 1 e 20 períodos defasados a serem definidos pelo algoritmo genético;

O Algoritmo 1 apresenta o pseudocódigo do algoritmo genético utilizado neste trabalho. O algoritmo gera um número pré-definido de indivíduos, por meio da função

DEAP.toolbox.population e então os avalia por meio da função DEAP.toolbox.evaluate. É feito o torneio entre indivíduos e os vencedores são selecionados para o cruzamento. Esta seleção é feita pela função DEAP.toolbox.select. A função DEAP.algorithms.varAnd gera novos indivíduos a partir da mistura genética dos pais, com probabilidade de mutações. O algoritmo roda

⁶ O formato de uma variável com n defasagens é igual a y_{t-n}

iterativamente, de acordo com o número de gerações, no caminho da função objetivo (ou fitness). Neste trabalho, visamos indivíduos que forneçam um modelo com menor valor de RMSE no conjunto de validação. Ou seja, a função objetivo é minimizar o RMSE.

O melhor modelos selecionado pelo algoritmo genético é empregado no restante da análise aqui proposta.

3.2 Equações e hipóteses

As equações teóricas propostas na literatura como meio de avaliar a relação entre volatilidade implícita e retornos de ativos financeiros são baseadas nos trabalhos de Black (1976), Tversky e Kahneman (1979), Christie (1982), French et al. (1987) e Campbell e Hentschel (1991). As equações M1 e M2 consideram as pressuposições teóricas apresentadas no trabalho de Low (2004). Elas descrevem em que medida o índice de volatilidade implícita do petróleo (OVX) se altera, em termos percentuais, quando há um choque nos retornos contemporâneos R_t e nos retornos contemporâneos ao quadrado R_t^2 das opções do United States Oil Fund (USO), tal como se segue:

$$M1 \equiv \% \Delta OVX_t = \beta_0 + \beta_1 R_t + \varepsilon_t \quad (3.1)$$

$$M2 \equiv \% \Delta OVX_t = \beta_0 + \beta_1 R_t + \beta_2 R_t^2 + \varepsilon_t \quad (3.2)$$

onde $\% \Delta OVX$ representa a variação percentual da volatilidade implícita, ou seja, $(OVX_t / OVX_{t-1}) - 1$; R_t representa o retorno logarítmico das opções USO ($\ln(USO_t) - \ln(USO_{t-1})$) e R_t^2 representa o retorno logarítmico ao quadrado das opções USO ($(\ln(USO_t) - \ln(USO_{t-1}))^2$). Os parâmetros β_0 e β_1 representam, respectivamente, o coeficiente angular e o coeficiente dos retornos nas duas equações. O coeficiente β_2 diz respeito ao coeficiente da não linearidade na equação, caso ela exista.

A Equação 3.3 surge a partir do trabalho de Fleming et al. (1995), adaptada segundo os critérios de Silva Junior (2017). A equação define a relação intertemporal entre o índice OVX e os retornos das opções de petróleo dois passos a trás e dois passos a frente no tempo. A equação inclui ainda um vetor de valores absolutos para os retornos contemporâneos que define se a volatilidade implícita do mercado petrolífero sofre alterações após choques no retorno das opções independente de sua direção (negativa ou positiva).

$$M3 \equiv \Delta OVX_t = \alpha_0 + \alpha_1 R_t + \alpha_2 R_{t-1} + \alpha_3 R_{t-2} + \delta_1 R_{t+1} + \delta_2 R_{t+2} + \delta_3 |R_t| + \varepsilon_t \quad (3.3)$$

onde ΔOVX representa a variação absoluta da volatilidade implícita (OVX_t/OVX_{t-1}); R_{t+1} e R_{t+2} representam, respectivamente, os retornos obtidos com as opções USO dois e três dias a frente; $|R_t|$ representa o valor absoluto do retorno contemporâneo sobre o USO; $\alpha_0, \alpha_1, \alpha_2$ e α_3 representam, respectivamente, os coeficientes angular, dos retornos contemporâneos, de uma e duas defasagens. Por sua vez, δ_1 e δ_2 representam os coeficientes de um e dois passos a frente, respectivamente, enquanto δ_3 representa o coeficiente dos retornos absolutos⁷.

3.3 Modelo LSTM

O comportamento em cadeia de uma rede neural recorrente (RNN)⁸ permite categorizar os dados de entrada de uma rede neural fornecendo uma estrutura interna confiável requerida pela tarefa de predição, além de apresentar a capacidade de descrever a dinâmica de comportamento dos dados (HÜSKEN; STAGGE, 2003). Sagheer e Kotb (2019) expõem que ao longo do processamento de uma sequência temporal em RNN, o produto da rede é obtido a partir das observações no tempo (t) e daquelas observadas no período anterior (t-1). Contudo, o desempenho das RNN tradicionais é limitado ao considerar que estas não apresentam eficiência com séries de dados contendo um grande número de observações.⁹ Isso ocorre devido à sua restrição de memorização durante o processo de aprendizado, tal como ilustrado por Pascanu et al. (2013) e Sutskever (2013).

⁷ O cálculo dos retornos com defasagem maior que um segue o mesmo raciocínio. Assim, de modo geral tem-se que $R_{t-n} = \ln(USO_t) - \ln(USO_{t-n})$ enquanto os retornos futuros são descritos como

$R_{t+n} = \ln(USO_{t+n}) - \ln(USO_t)$

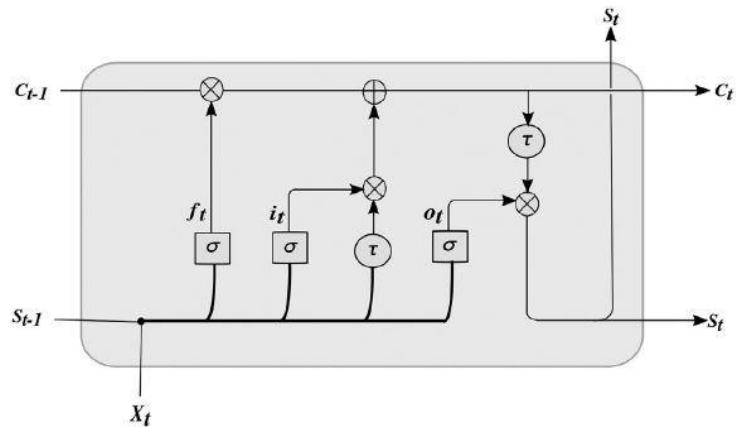
⁸ Uma rede neural recorrente realiza predições a partir da entrada de novas informações, mas também do produto passado gerado por ela mesma. Um exemplo clássico de um modelo de redes neurais é o usado na barra de pesquisa do Google. À medida que a frase está sendo escrita, a ferramenta de busca sugere a palavra seguinte com base na que foi digitada anteriormente, além de se basear nos registros feitos por outros usuários ao longo do tempo. Isso difere uma rede neural recorrente das redes neurais artificiais tradicionais, as quais dependem exclusivamente da entrada de informação para realizar previsões.

⁹ Uma rede neural recorrente tradicional é eficiente ao traduzir uma frase curta do inglês para o português, por exemplo. Porém, sua capacidade preditiva de uma série financeira de 3000 observações diárias é limitada.

Assim, as RNN tradicionais são adequadas apenas em tarefas que requerem informações de curto prazo para que o aprendizado seja alcançado. Entretanto, em certas ocasiões o aprendizado da rede somente será obtido caso ela seja capaz de absorver dependências de longo prazo sobre os dados. Uma das variantes das RNN capaz de cumprir com esse objetivo é unidade de rede denominada Long ShortTerm Memory (LSTM). Proposta por Hochreiter e Schmidhuber (1997), a unidade de rede LSTM é desenvolvida para solucionar o problema dos gradientes explosivos no processo de treinamento da rede. Assim, as LSTM truncam os gradientes sem incorrer em danos, permitindo o aprendizado de dependências de longo prazo ao produzir o que é intitulado como Carrossel de Erro Constante (CEC) dentro de unidades especiais. Tais unidades são não lineares e exercem a tarefa de abrir ou fechar os portões ao longo da rede, permitindo a regulagem do fluxo de erro durante o processo de aprendizado (SUTSKEVER, 2013). Além disso, destaca-se a capacidade das unidades de redes LSTM em aproximar as informações de longo prazo com atrasos significativos, acelerando o algoritmo de RNN (GREFF et al., 2016).

As unidades de redes LSTM constituem-se enquanto uma estrutura em cadeia a partir de quatro camadas de rede neural que interagem entre si. Ao observar a Figura 1 a seguir percebe-se uma linha horizontal contínua. Esta linha horizontal, presente no topo da cadeia, é denominada como estado da célula e exerce função semelhante a uma correia transportadora, percorrendo toda a cadeia a partir de interações lineares menores. A LSTM cumpre a tarefa de regular informações ao estado da célula por meio de três estruturas conhecidas como portão de esquecimento (f_t), portão de entrada (i_t) e portão de saída (o_t). A figura 1 ilustra o diagrama de funcionamento da rede.

Figura 1 – Cadeia LSTM



Fonte: Sagheer e Kotb (2019).

Os portões, por sua vez, regulam as informações por meio das camadas de rede identificadas a partir de três funções sigmôides e uma função tangente hiperbólica¹⁰. O processo de regulagem envolve as decisões de descartar ou não a informação candidata ao estado da célula; decidir qual informação será armazenada no estado da célula; e, por fim, qual será apresentada enquanto produto do estado da célula. Além das quatro camadas, o processo de manipulação da memória também conta com o auxílio de operações pontuais, como a adição de vetores, por exemplo. Assim, ressalta-se que as informações candidatas à célula são obtidas a partir de um vetor de entrada contemporâneo (X_t) e outro que representa o produto da célula anterior (S_{t-1}), gerando o produto em S_t . O processamento da LSTM é detalhado a seguir.

É importante ressaltar que as informações são inseridas na rede neural de forma conjunta, diferentemente de uma regressão clássica onde se estabelece uma condição de igualdade entre variável independente e variáveis dependentes. Ou seja, a volatilidade implícita e os retornos são lidos pela rede neural simultaneamente. Posteriormente, na etapa de programação, define-se a variável a ser predita pelo modelo (y = volatilidade implícita) que é denominada como variável alvo.

Na etapa de descarte, a camada sigmoide, juntamente a uma operação de multiplicação pontual, estabelece qual informação será excluída da célula produzindo números entre 0 (zero) e 1 (um) que representarão o quanto de cada componente será liberado. Quanto mais próximo de zero, menor a chance da informação ser candidata a acoplar a célula, enquanto o contrário ocorre quando o valor está próximo de um. Assim, o portão de descarte (f_t) cumpre essa tarefa da seguinte forma:

$$f_t = \sigma(X_t U^f + S_{t-1} W^f + b_f) \quad (3.4)$$

onde σ representa a função sigmoide que converte os dados para uma escala entre 0 e 1; X_t a informação de entrada da rede; U^f o peso da informação de entrada; S_{t-1} o estado oculto da etapa anterior; W^f o peso do estado oculto; e b_f um parâmetro de viés da equação.

¹⁰ A função sigmoide converte os dados de entrada da rede neural recorrente em uma escala que varia entre 0 e 1 e a função tangente hiperbólica posteriormente converte para uma escala entre -1 e 1. Após a etapa preditiva, os dados são convertidos para sua escala original para que sejam comparados com os valores reais.

Em seguida, a tarefa consiste em determinar qual informação será armazenada no estado da célula a partir de dois momentos: primeiro, o portão de entrada (i_t), por meio de uma camada sigmoide, decide quais valores serão atualizados; segundo, a camada tangente hiperbólica cria um vetor de novos valores candidatos \hat{C}_t , tal como se segue:

$$i_t = \sigma(X_t U^i + S_{t-1} W^i + b_i) \quad (3.5)$$

onde i_t é o portão de entrada; σ representa a função sigmoide; U^i o peso da informação de entrada; S_{t-1} o estado oculto da etapa anterior; W^i o peso do estado oculto dada pelo portão de entrada; e b_i um parâmetro de viés da equação de entrada.

$$\tilde{C}_t = \tanh(X_t U^c + S_{t-1} W^c + b_c) \quad (3.6)$$

onde \tilde{C}_t é o vetor de novos valores candidatos; \tanh é a função de ativação tangente hiperbólica não linear aplicada para normalizar os dados em uma escala entre -1 e 1; U^c o peso da informação de entrada; S_{t-1} o estado oculto da etapa anterior; W^c o peso do estado oculto; e b_c um parâmetro de viés da equação de atualização.

A partir da etapa anterior há uma atualização do estado da célula antiga C_{t-1} em um novo estado de célula C_t . Nesse estágio, o antigo estado C_{t-1} é multiplicado por (f_t) concretizando o processo de esquecimento. Em seguida, adiciona-se o produto entre o portão de entrada (i_t) e o vetor de novos valores que foram criados anteriormente \tilde{C}_t . A denotação matemática para estes cálculos é apresentada a seguir:

$$C_t = C_{t-1} \otimes f_t \oplus i_t \otimes \tilde{C}_t \quad (3.7)$$

Por fim, decide-se a parcela informacional a ser gerada como produto pelo portão de saída. Este produto será baseado no estado da célula em uma versão filtrada. O primeiro passo é decidir quais partes do estado da célula vamos produzir pela camada sigmoide. Em seguida, o estado da célula percorre a camada tangente hiperbólica (que produz valores entre -1 e 1) e é multiplicado pela saída do sigmoide produzindo apenas as parcelas que foram decididas. Matematicamente, a representação é tida como:

$$o_t = \sigma(X_t U_o + S_{t-1} W_o + b_o) \quad (3.8)$$

$$S_t = o_t \otimes \tanh(C_t) \quad (3.9)$$

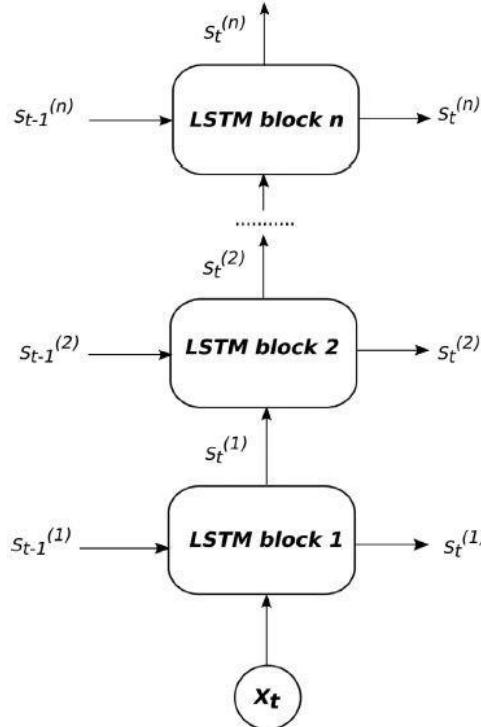
onde t é o portão de saída; σ a função sigmoide; U^o o peso da informação de saída; S_{t-1} o estado oculto da etapa anterior; W^o o peso do estado oculto; b_0 um parâmetro de viés da equação de saída; S_t é o estado ocultado atualizado e \otimes um ponto de multiplicação entre as equações.

3.4 A rede neural recorrente LSTM profunda

A literatura expõe que o aumento de profundidade de uma rede neural é uma maneira eficaz de aperfeiçoar o processo de aprendizagem geral da rede (LECUN et al., 2015b). Nesse sentido, o presente trabalho se inspira na metodologia proposta por Sagheer e Kotb (2019) baseada em uma LSTM profunda (Deep LSTM - DLSTM) desenvolvida com o intuito de otimizar previsões de séries temporais. O novo modelo dos autores empilha blocos de unidades de redes LSTM um sobre o outro em uma rede neural recorrente profunda a fim de combinar as utilidades e benefícios de uma única camada de LSTM.

A proposta em empilhar múltiplas unidades de redes LSTM em uma arquitetura hierárquica é prover mecanismos nas camadas inferiores que separem os fatores de variação presentes nos dados de entrada, posteriormente combinando o resultado dessa separação em camadas superiores (SAGHEER; KOTB, 2019). Em contextos em que a sequência de dados apresenta grandes extensões e alta complexidade, arquiteturas de redes profundas apresentaram performances melhores devido a uma representação mais compacta (LÄNGKVIST et al., 2014b; UTGOFF; STRACUZZI, 2002; HERMANS; SCHRAUWEN, 2013).

Figura 2 – Arquitetura de uma Rede Recorrente DLSTM



Fonte: Sagheer e Kotb (2019).

Ao observar a arquitetura de uma rede recorrente DLSTM, o vetor de entrada X_t é computado no primeiro bloco LSTM juntamente com o estado oculto anterior $S_{t-1}^{(1)}$, em que o sobrescrito (1) se refere ao primeiro bloco LSTM. O produto da primeira LSTM será considerado o estado oculto no tempo t , $S_t^{(1)}$, sendo calculado conforme explicitado na subseção 3.3. Dessa forma, o segundo bloco LSTM faz uso do estado oculto em $S_t^{(1)}$ junto com o estado oculto em $S_{t-1}^{(2)}$ para compor o produto $S_t^{(2)}$, que segue adiante no próximo período de tempo e em direção ao terceiro bloco de LSTM.

O procedimento é repetido até que o último bloco LSTM seja compilado na pilha. A Figura 2 ilustra a representação da arquitetura de uma DLSTM.

Os autores expõem que a arquitetura de uma rede profunda nesses moldes permite que cada bloco da cadeia processe parte do objetivo e repasse isso adiante para cada um dos blocos subsequentes até que a última camada acumulada forneça o produto final, além de permitir que cada estado oculto em cada um dos níveis da cadeia opere em diferentes escalas de tempo (SAGHEER; KOTB, 2019). Por fim, esses benefícios exercem considerável impacto em cenários com dependência de longo prazo ou na manipulação de modelos multivariados em séries de tempo (SPIEGEL et al., 2011).

3.5 Modelo ARMA

O modelo autorregressivo de média móvel (ARMA) será utilizado como benchmark de previsão. Ele é definido como a combinação de um processo de média móvel (MA) com uma equação em diferença linear (ENDERS, 2012). Considere a equação em diferença de n-ésima ordem:

p

$$OVX_t = a_0 + \sum_{i=1}^p a_i OVX_{t-i} + x_t \quad (3.10)$$

tal que, OVX_t seja a volatilidade implícita contemporânea do petróleo; OVX_t a volatilidade implícita do petróleo com i defasagens; a_0 uma constante, e a_i sejam os coeficientes autorregressivos da volatilidade implícita nos períodos de defasagem $i = 1, 2, \dots, p$. O termo x_t é conhecido como processo forçado com a capacidade de assumir diversas formas, tais como qualquer função do tempo, valores contemporâneos e defasados de outras variáveis e/ou distúrbios estocásticos (ENDERS, 2012).

Assumindo que $\{x_t\}$ seja um MA de ordem q , MA(q), tem-se que:

$$OVX_t = a_0 \sum_{i=1}^p a_i OVX_{t-i} + \sum_{i=0}^q \beta_i \epsilon_{t-i} \quad (3.11)$$

Caso as raízes unitárias na equação 3.11 não estejam sobre o círculo unitário, $\{OVX_t\}$ será denominada como um modelo autorregressivo de média móvel (ARMA) para OVX_t . Assim, a parte autorregressiva do modelo é representada pela equação em diferença 3.10, tal como o processo de média móvel pela sequência de $\{x_t\}$. Caso a parte homogênea da equação em diferença contenha p defasagens e o modelo para x_t q defasagens, tem-se um modelo ARMA (p, q). Ressalta-se que caso uma ou mais raízes características do processo $\{OVX_t\}$ seja maior ou igual à unidade, esta sequência é tida como um processo integrado e é denominada como um processo autorregressivo integrado de média móvel (ARIMA).

3.6 Modelo de vetores autorregressivos (VAR)

Quando não há total confiança sobre a exogeneidade de uma variável na série temporal, as variáveis do modelo são tratadas simetricamente (ENDERS, 2012). Supondo as sequências temporais entre volatilidade implícita $\{OVX_t\}$ e retornos das opções do petróleo $\{R_t\}$, considere que a volatilidade implícita $\{OVX_t\}$ seja afetada pelas realizações contemporâneas e defasadas dos retornos no mercado $\{R_t\}$, e que

os retornos $\{R_t\}$ sejam afetados pelas realizações contemporâneas e defasadas da volatilidade implícita $\{OVX_t\}$. Dessa forma, tem-se o sistema bivariado:

$$OVX_t = b_{10} - b_{12}R_t + \gamma_{11}OVX_{t-1} + \gamma_{12}R_{t-1} + \epsilon_{OVXt} \quad (3.12)$$

$$R_t = b_{20} - b_{21}OVX_t + \gamma_{21}OVX_{t-1} + \gamma_{22}R_{t-1} + \epsilon_{Rt} \quad (3.13)$$

dado que a volatilidade implícita OVX_t e os retornos das opções R_t sejam ambos estacionários; ϵ_{OVXt} e ϵ_{Rt} sejam ruídos brancos com erros padrões de σ_{OVX} e σ_R , respectivamente; e ϵ_{OVXt} e ϵ_{Rt} sejam ruídos brancos não correlacionados.

As equações 3.12 e 3.13 representam um vetor autorregressivo (VAR) de primeira ordem, dado que a defasagem máxima das equações é de um período. A estrutura desse sistema permite que a volatilidade implícita y_t afete os retornos das opções R_t e que os retornos das opções R_t afetem a volatilidade implícita OVX_t . Porém, as equações não estão em seus formatos reduzidos, dado que OVX_t tem um efeito contemporâneo sobre R_t e R_t tem um efeito contemporâneo sobre OVX_t . Por meio de manipulações algébricas é possível reescrever o sistema de maneira mais compacta:

$$x_t = A_0 + A_1x_{t-1} + e_t \quad (3.14)$$

tal que a_{i0} seja um elemento i do vetor A_0 , a_{ij} o elemento na linha i e coluna j da matriz A_1 , e e_{it} o elemento i do vetor e_t . Sua generalização multivariada pode ser representada como:

$$x_t = A_0 + A_1x_{t-1} + A_2x_{t-2} + \dots + A_px_{t-p} + e_t \quad (3.15)$$

sendo x_t um vetor ($n \times 1$) contendo cada uma das n variáveis incluídas no VAR; A_0 um vetor ($n \times 1$) de interceptos; A_i matrizes de coeficientes de ordem ($n \times n$); e e_t um vetor ($n \times 1$) de termos de erro.

3.7 Testes para diagnóstico de previsão

Esta seção é dedicada a apresentar as métricas responsáveis por mensurar a qualidade de previsão dos modelos. De forma generalizada, os testes apresentados a seguir fornecem a informação do quão distante as previsões feitas por um modelo estão dos reais valores observados em um dado conjunto de dados. Além disso, os

resultados dos testes auxiliam o processo de comparação de performance de diferentes classes de modelos utilizando a mesma amostra de dados, o que ocorre nesse trabalho que faz uso de uma rede neural e modelos tradicionais em séries de tempo. Neste trabalho utilizou-se os testes de Raiz do Erro Quadrático Médio (RMSE), Erro Médio Absoluto (MAE) e Erro Quadrático Médio (MSE).

3.7.1 Erro Médio Absoluto (MAE)

O Erro Médio Absoluto (MAE) considera a diferença entre os valores previstos e os valores reais, independente de sua direção (negativa ou positiva). Por ser absoluto, o erro médio calculado ao longo de toda a amostra considera que todas as diferenças individuais entre valores previstos e valores observados possuem o mesmo peso entre si (WILLMOTT, 1981). O cálculo final do MAE retorna valores na mesma unidade dos dados da amostra e pode ser expresso como:

$$MAE = \frac{1}{n} \sum_{j=1}^n |\hat{y}_j - y_j| \quad (3.16)$$

onde n é o total da amostra; j representa a j-ésima observação; \hat{y}_j é o j-ésimo valor b previsto; e y_j é o j-ésimo valor observado no conjunto.

3.7.2 Erro quadrático médio (MSE) e Raiz do erro quadrático médio (RMSE)

O Erro Quadrático Médio (MSE) mede o quanto próximos os valores estimados estão dos dados observados. Seu cálculo leva em consideração a média da diferença quadrática entre os valores observados e os valores previstos em uma amostra. Assim, é possível dizer que ela é responsável por medir a variância dos resíduos, tal como visto a seguir:

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2} \quad (3.17)$$

Por sua vez, a Raiz do erro quadrático médio (RMSE) testa a diferença da distribuição de erros de previsão dos modelos (erros para previsões fora da amostra).

Diferentemente do MSE, o RMSE agrega a magnitude de erros de diferentes instantes do tempo em uma única estatística de comparação e fornece o desvio padrão dos resíduos, e não sua variância (HAMILTON, 1994). A equação do teste pode ser descrita como:

$$RMSE = \sqrt{\sum_{t=1}^T (\hat{y}_t - y_t)^2 \div T} \quad (3.18)$$

tal que a previsão com menor estatística RMSE apresenta maior acurácia. A RMSE compara o valor previsto \hat{y}_t pelo modelo para a variável y_t durante o período T. Assim, considerando que 636 observações da volatilidade implícita e dos retornos das opções (20% da amostral total) foram separadas para testar os modelos, o cálculo do RMSE considera como \hat{y}_t os 636 valores e como y_t as 636 observações separadas como conjunto de teste da modelagem.

Quando comparados, as medidas de RMSE e MAE apresentam algumas características distintas entre si. Considerando que os erros são elevados ao quadrado (assim como no caso do MSE) os desvios que apresentam maior magnitude em relação aos valores originais recebem maior peso. Isso possibilita evitar possíveis vieses para modelos que apresentem maior sensibilidade a erros maiores.

A literatura não apresenta uma medida de erro ideal de comparação de performances entre modelos de séries de tempo, sendo essa escolha baseada na especificidade dos dados e do problema de pesquisa imposto. Apesar de se constatar uma grande utilização de medidas de erro percentuais, nossa amostra apresenta valores iguais a zero produzidos em função das técnicas de estacionarização das séries. Assim, o cálculo de medidas de erro percentuais neste contexto se torna inviável, optando assim pela forma convencional dos erros apresentados anteriormente.

4. RESULTADOS EMPÍRICOS

Esta seção dedica-se a apresentar os resultados quantitativos e visuais produzidos pela rede neural DCLSTM e pelos modelos estatísticos utilizados no estudo. As estatísticas descritivas das variáveis também são apresentadas, assim como a associação teórica de sua trajetória amostral com a relação risco-retorno em mercados financeiros.

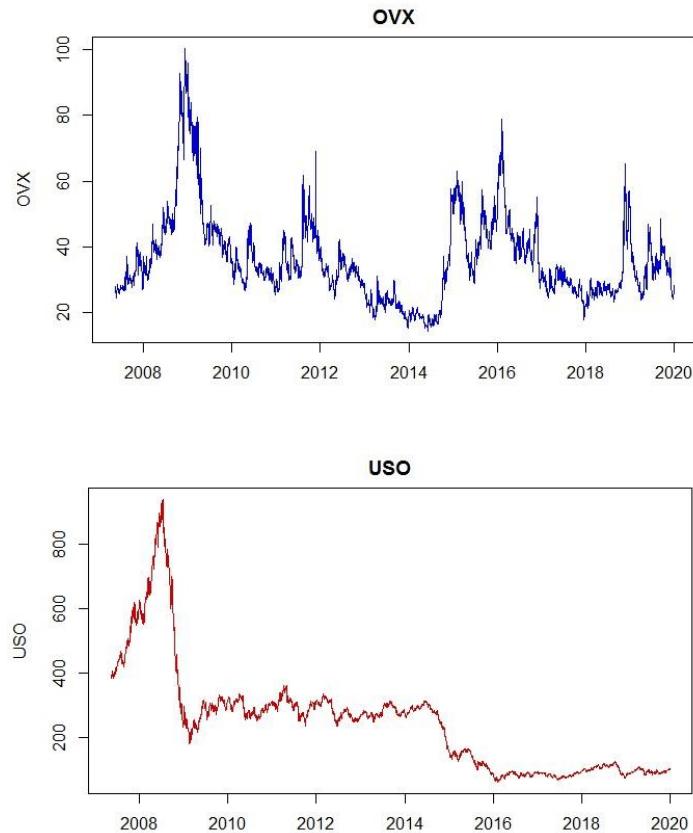
Os resultados da modelagem são apresentados individualmente para cada modelo e posteriormente têm sua performance comparada entre si. Ressalta-se que a performance alcançada por cada um dos modelos embasou-se em seu desempenho a partir dos dados de teste e não nos de treino. A literatura indica que uma avaliação robusta de performance de modelos preditivos deve embasar-se em dados não conhecidos pelo modelo (teste), ao contrário da utilização de dados já utilizados (treino) na modelagem (HYNDMAN, 2016).

Após aplicar os procedimentos descritos na subseção 3.1.2, os modelos referentes às equações 3.1 e 3.2 receberam um total de 3179 observações. Deste total, 80% da amostra (2543) foram dedicados para o treinamento dos modelos e os outros 20% (636) dedicados a testar suas performances. Para a Equação 3.3 os mesmos procedimentos foram aplicados. Contudo, o número de observações foi de 3180 devido à especificação de sua forma funcional, apesar da proporção dedicada para treino e teste dos modelos ter sido à mesma das equações anteriores, diferenciando apenas o número absoluto de observações para o subconjunto de treino (2544)

4.1. Estatísticas Descritivas

A trajetória temporal entre o índice de volatilidade implícita (OVX) e as opções (USO) para o mercado de petróleo norte-americano ilustra a relação negativa riscoretorno de ativos financeiros, inicialmente proposta nos trabalhos de Black (1976), Poterba e Summers (1984) e Campbell e Hentschel (1991). Parte desse comportamento inverso é nítido em períodos de choques no preço do barril de petróleo. A figura 3 ilustra este comportamento.

Figura 3 – Volatilidade implícita (OVX) e opções do United States Oil Fund (USO).



Fonte: elaboração dos autores a partir de dados do Yahoo Finance (2020).

O maior valor amostral para o índice OVX (100.42) ocorre em 11 de dezembro de 2008, no desdobrar da crise econômica mundial daquele ano. O aumento na medida de risco do mercado de petróleo americano ocorre simultaneamente a uma queda expressiva e contínua das opções do petróleo no WTI, atingindo em 18 de fevereiro de 2009 o menor valor desde o início da amostra (182.88).

Após o choque mundial no preço da commodity em 2008, a série aparenta sofrer uma variação em seu nível médio¹¹, nível este que se perdura até o segundo semestre de 2014, quando o preço do petróleo sofre novas quedas em escala global. O valor mínimo de 63.68 das opções ocorre no dia 10 de fevereiro de 2016, o qual antecede novo choque na volatilidade implícita do mercado no dia 12 de fevereiro de

¹¹A literatura em séries temporais também aborda essa variação como mudança e/ou quebra estrutural na série. O teste de Chow (1960) é comumente utilizado como ferramenta formal para identificações de possíveis quebras nas séries de tempo. Para mais detalhes sobre o tema veja Enders (2012).

2016. A partir desse ponto, a série indica percorrer uma nova trajetória em um nível médio menor que o anterior.

Após os procedimentos de estacionarização descritos na subseção 3.1.2.2, os dados tiveram sua magnitude alterada. A Tabela 1 apresenta a estatística descritiva das variáveis diferenciadas e os resultados do teste de raiz unitária Augmented Dickey-Fuller.

Tabela 1 – Estatística descritiva das variáveis em primeira diferença

	Média	Mediana	Mín.	Máx.	Desvio Padrão	Assimetria	Curtose	JB	ADF
ΔOVX	-0.00002	-0.0013	-0.1313	0.1186	0.0197	0.6801	3.5260	490.22#	229.484*
R_{USO}	-0.0004	0.0003	-0.0460	0.0503	0.0094	-0.3232	1.8088	490.22#	187.290*

Nota: # e (*) indicam significância estatística ao nível de 1% para os testes de JB e ADF, respectivamente.

O cálculo da curtose revela uma distribuição com caudas pesadas nas variações da volatilidade implícita (ΔOVX), o que não ocorre em grande proporção para os retornos das opções R_{USO}). Por sua vez, os valores de assimetria são baixos para ambas as variáveis, apesar de se identificar uma distribuição assimétrica à direita para a volatilidade implícita diferenciada, enquanto se tem uma distribuição assimétrica à esquerda para os retornos. O teste de Jarque-Bera (JB) confirma a não normalidade dos dados. Por meio do teste de Augmented Dickey-Fuller (ADF) constata-se ausência de raiz unitária nos dados após as transformações. Detalhes sobre o teste ADF podem ser obtidos em Dickey e Fuller (1979) e Dickey e Fuller (1981).

4.2 Resultados dos Mínimos Quadrados Ordinários

A Tabela 2 apresenta os resultados dos modelos de regressão por mínimos quadrados ordinários. As estimativas descrevem empiricamente a relação dos retornos em diferentes períodos no tempo associados à mudanças na variação absoluta e percentual da volatilidade implícita representada pelo índice OVX.

Tabela 2 – Resultados das regressões utilizando Mínimos Quadrados Ordinários

	$R^2(\%)$	Int.	R_t	R_{t-1}	R_{t-2}	R_{t+1}	R_{t+2}	$ R_t $	R_t^2
M1	14.24	-0.0003 (-0.707)	-0.7619*** (-20.570)						
M2	16.81	-0.0019*** (-4.645)	-0.6924*** (-18.559)					17.1716*** (8.908)	
M3	19.42	-0.1849*** (-8.573)	-35.7134*** (-11.552)	-1.6232 (-0.527)	4.1328* (1.706)	2.0960 (0.680)	5.3599** (2.216)	23.2688*** (10.337)	

Nota: (***) indica significância estatística ao nível de 1%, (**) significância estatística de 5% e (*) significância estatística de 10%.

Inicialmente, todos os modelos apresentam os coeficientes dos retornos contemporâneos negativos e estatisticamente significativos, corroborando a hipótese de alavancagem proposta inicialmente por Black (1976) e Christie (1982). A pressuposição de uma relação não linear entre os retornos e a volatilidade implícita pela equação 3.2 é corroborada ao observar um coeficiente de retornos ao quadrado estatisticamente significativo e que contribui para um ajuste maior do modelo aos dados. O resultado sugere que esta especificação funcional possa ser mais adequada para explicar a dinâmica entre retorno e volatilidade implícita do mercado de petróleo.

Além disso, a significância estatística do coeficiente de retornos ao quadrado indica a presença de bruscas oscilações dos retornos das opções de petróleo ao longo do tempo. Essa interpretação provém da pressuposição de que caso exista heterocedasticidade condicional em uma série (volatilidade), ela possa ser representada matematicamente por um coeficiente não linear. Dessa forma, ao incluir no modelo a informação de que a correlação entre os ativos se inverte ao longo do tempo (não linearidade), o efeito explicativo sobre a volatilidade futura (implícita) é maior.

Os resultados da Equação 3.3 mantêm a validação de uma relação negativa entre retornos e volatilidade implícita. Apenas os coeficientes de retornos contemporâneos foram consistentemente significativos, enquanto os retornos defasados em dois períodos e em dois passos a frente apresentaram significância menor. Os retornos defasados em um período e em um passo a frente não foram significativos para a amostra.

A inclusão dos retornos absolutos produz um coeficiente consistentemente significativo e aumenta o poder explicativo do modelo. Assim, se o retorno do mercado de petróleo é positivo, o efeito sobre a variação da volatilidade implícita será de -12.45 (-35.71 + 23.26) sugerindo que um acréscimo nos retornos dos ativos é acompanhado de um declínio de seu risco futuro. Por outro lado, se o retorno do mercado é negativo seu impacto sobre a volatilidade implícita será de -58.97 (-35.71 - 23.26), impactando positivamente a variação da volatilidade implícita.

Assim, é possível dizer que movimentos negativos no mercado financeiro de petróleo exercem maior impacto sobre as variações da volatilidade implícita desse mercado quando comparado a movimentos positivos (aversão ao risco). Esses resultados corroboram a presença de assimetria na relação das variáveis para o mercado de petróleo e estão em consonância com os resultados encontrados por Silva Junior (2017).

4.3 Resultados das regressões quantílicas

A Tabela 3 apresenta os resultados das estimações por regressão quantílica realizadas para as equações 3.1, 3.2 e 3.3. O procedimento foi realizado com o objetivo de verificar se a relação entre retornos e volatilidade implícita tem um comportamento distinto ao longo dos diferentes quantis da distribuição da volatilidade, além de verificar se as pressuposições de assimetria e não linearidade da relação permanecem válidas nestas condições. Assim, em um cenário de não-linearidade, o coeficiente de retornos ao quadrado seria estatisticamente significativo em todos os quantis ou mais significativo nos quantis maiores. O mesmo ocorre para a assimetria ao considerar o coeficiente de retornos absolutos. Caso não haja a presença de nenhuma dessas características nos dados, ambos os coeficientes não seriam estatisticamente significativos em nenhum dos quantis ou fracamente significativos apenas nos quantis menores.

A reestimação da Equação 3.1 evidencia que o coeficiente dos retornos contemporâneos permanece negativo e estatisticamente significativo para todos os quantis da amostra, em concordância com a teoria discutida na subseção anterior. À medida que se avança ao longo dos quantis o impacto negativo dos retornos sobre a variação percentual da volatilidade implícita tende a ser maior, com exceção do quantil 0.90 que é menor que o anterior. Contudo, esses resultados não oferecem suporte

empírico necessário para discorrer sobre aspectos de não linearidade e assimetria na relação entre volatilidade implícita e retornos.

Os resultados da reestimação da Equação 3.2 revelam comportamentos distintos para o coeficiente de retornos ao quadrado ao longo dos quantis. Observa-se um coeficiente positivo e estatisticamente significativo a partir da mediana (0.5) até o maior quantil da distribuição (0.95). Contudo, para o menor quantil (0.05) o coeficiente é negativo e estatisticamente significativo. Os coeficientes referentes aos quantis 0.10 e 0.25 não apresentaram significância estatística. Apesar de não ser possível descrever um padrão, os resultados indicam a presença de não linearidade em grande parte da relação entre a variação percentual da volatilidade implícita e os retornos do mercado de petróleo.

Para o coeficiente dos retornos contemporâneos obtidos pela reestimação da Equação 3.3 foram encontrados valores negativos e estatisticamente significativos ao longo de toda a distribuição. Dentre os retornos que sofreram defasagens ou que estão a passos a frente, apenas aqueles defasados em dois períodos e de dois passos a frente apresentaram significância estatística, ambos os casos para um quantil menor (0.25). Os resultados se assemelham à estimativa apresentada na Tabela 2, mas se aprofundam ao constatar que o impacto destes últimos retornos possui relevância estatística apenas em uma porção inferior da distribuição. Por sua vez, o coeficiente dos retornos absolutos é consistentemente significativo ao longo de toda a distribuição da volatilidade e indica a presença de assimetria na relação entre volatilidade implícita e retornos.

Tabela 3 – Resultados das regressões quantílicas

Quantil	0.05	0.10	0.25	0.50	0.75	0.90	0.95
M1							
<i>Intercepto</i>	-0.0261*** (-31.7559)	-0.0203*** (-37.4361)	-0.0111*** (-28.2413)	-0.0017*** (-4.3813)	0.0092*** (18.7096)	0.0209*** (26.7448)	0.0299*** (23.9580)
R_t	-0.6189*** (-8.2397)	-0.6620*** (12.1469)	-0.6830*** (-17.7627)	-0.7593*** (-19.1827)	-0.8068*** (-16.2117)	-0.7732*** (-9.6681)	-0.8395*** (-6.5209)
M2							
<i>Intercepto</i>	-0.0254*** (-29.2294)	-0.0199*** (-34.2544)	-0.0117*** (-24.3969)	-0.0031*** (-6.9451)	0.0061*** (11.6241)	0.0155*** (21.2032)	0.0222*** (21.8012)
R_t	-0.8824*** (-7.0190)	-0.7269*** (-9.9679)	-0.6724*** (-12.0574)	-0.6999*** (-13.3053)	-0.7343*** (-10.9365)	-0.6738*** (-6.5129)	-0.5600*** (-5.7142)
R_t^2	-14.0077** (-2.0557)	-7.9044 (-1.4124)	6.2239 (1.4299)	20.0478*** (4.8866)	37.5024*** (6.2669)	65.1011*** (7.6222)	70.6487*** (13.0514)
M3							
<i>Intercepto</i>	-0.8497*** (-16.5407)	-0.6236*** (-17.6982)	-0.3937*** (-17.5812)	-0.1747*** (-10.5128)	0.0195 (0.9149)	0.2738*** (7.9905)	0.4998*** (10.4096)
R_t	-46.4693*** (-6.3857)	-37.9011*** (-6.2894)	-33.1611*** (-8.2381)	-37.9054*** (-11.7657)	-32.2249*** (-8.6277)	-25.9017*** (-4.3479)	-27.5851*** (-3.8431)
R_{t-1}	3.8224 (0.5888)	-3.2507 (-0.5820)	-0.7869 (-0.2144)	4.1739 (1.4691)	0.4723 (0.1384)	-3.6423 (-0.6909)	-0.2437 (-0.0366)
R_{t-2}	3.4384 (0.6977)	5.8481 (1.3536)	7.1823** (2.5336)	3.7374* (1.6874)	4.2782 (1.6402)	5.5636 (1.3510)	-3.1474 (-0.6141)
R_{t+1}	6.8961 (1.0762)	2.4732 (0.4614)	-4.9752 (-1.4422)	3.3548 (1.1843)	0.5041 (0.1524)	0.9819 (0.1924)	0.6587 (0.0985)
R_{t+2}	7.0929 (1.4988)	5.7575 (1.4755)	5.7147** (2.2545)	0.8756 (0.3992)	-1.1149 (-0.4244)	2.4581 (0.6403)	5.0964 (0.9339)
$ R_t $	-34.5452*** (-5.4041)	-24.8131*** (-3.9413)	1.5383 (0.3909)	22.8867*** (7.8268)	46.8269*** (13.8099)	68.8879*** (12.7879)	78.6831*** (14.4542)
$R^2(\%)$	4.43	7.31	13.08	17.02	14.73	9.16	5.93
$R^2(\%)$	4.38	7.29	13.05	16.79	14.17	8.42	5.24
$R^2(\%)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Nota: (***) indica significância estatística ao nível de 1%, (**) significância estatística de 5% e (*) significância estatística de 10%.

4.4 Resultados da rede neural DCLSTM

Esta subseção é dedicada a apresentar e discutir os resultados das estimações para as equações 3.1, 3.2 e 3.3 utilizando a rede neural recorrente Deep Convolutional Long Short Term Memory (DCLSTM). Ao final do procedimento de modelagem dos dados utilizando a rede neural, são selecionados três modelos candidatos pelo

algoritmo genético para cada equação contendo hiper parâmetros, funções de custo, arquiteturas de rede e demais parâmetros mencionados na 3.1.2.4 com valores e características distintos. As tabelas 4, 5 e 6 apresentam estes modelos para cada uma das equações da pesquisa. O modelo escolhido para cada caso é destacado em negrito.

Para os experimentos com o algoritmo genético, foi considerado uma população de tamanho 30 e evoluiu-se por 10 gerações. os parâmetros de mutação e seleção foram, respectivamente, 95% e 5%.

Isso possibilita verificar qual configuração da rede incorre em uma menor medida de erro para que posteriormente seja comparada com os demais modelos de séries de tempo tradicionais. Após a seleção dos valores ótimos para cada modelo estimado, a decisão entre o melhor modelo dentre os demais leva em consideração o cálculo da menor raiz do erro quadrático médio (Root Mean Squared Error - RMSE) apresentado na 3.7. Portanto, a RMSE é a função objetivo a ser minimizada pelo algoritmo genético. A escolha do modelo ótimo para cada equação é destacada em negrito nas tabelas. Ressalta-se que o algoritmo genético é empregado apenas nos dados de treinamento.

Considerando o cenário em que a variação percentual da volatilidade implícita está em função apenas dos retornos, nota-se que o único parâmetro que se destoa consideravelmente dos demais é o número de épocas da rede, como visto na Tabela 4. A medida em que se acrescenta mais épocas ao modelo (ep), menor é seu RMSE. Nas três especificações o número de defasagens (def) se mantém consideravelmente próximo, tal como o número de camadas LSTM utilizadas na rede que apresenta acréscimo de apenas uma unidade do modelo escolhido para os demais. De acordo com Sagheer e Kotb (2019), a rede DCLSTM não necessita de um número de camadas elevado ao trabalhar com um número relativamente baixo de observações.

Quanto aos neurônios da rede, nota-se que os modelos com apenas uma cada LSTM utilizam menos neurônios à medida em que o número de épocas se eleva. Para o modelo com duas camadas, a quantidade de neurônios para cada camada mantém-se relativamente próxima uma da outra outra. Não há um padrão de definição ótima para os neurônios, sendo sua especificação construída em consonância com a observância do desempenho de outros parâmetros do modelo, o que corrobora mais uma vez a utilização do algoritmo genético na tentativa de evitar definições manuais errôneas deste e outros hiper parâmetros.

É importante ressaltar que todos os modelos candidatos apresentaram camada convolucional (cc) na construção da rede, assim como unidades LSTM bideracionais (bi) e o mecanismo de atenção (tt). Neste caso, o modelo escolhido com menor

Tabela 4 – Seleção dos melhores modelos pelo algoritmo genético para a Equação 3.1

ep	cc	Istm	bi	att	neuronio	drop	def	lost	RMSE
485	Sim	2	Sim	Sim	(29,48)	34	15	MAE	0.0123
121	Sim	1	Sim	Sim	54	7	15	MSE	0.0204
355	Sim	1	Sim	Sim	38	36	14	MSE	0.0139

Fonte: resultados obtidos a partir de estimação dos autores.

Ao analisar as redes neurais estimadas a partir da equação 3.2, nota-se que quando o termo de retornos ao quadrado é inserido na função de estimação o candidato a melhor modelo apresenta um RMSE relativamente próximo ao modelo escolhido na Tabela 4. Percebe-se ainda que acréscimos ao número de épocas e defasagens também resultam em um RMSE menor. Considerando os retornos ao quadrado na equação, os hiper parâmetros e as características de rede para as equações 3.1 e 3.2 são iguais. Contudo, ao considerar que RMSE do modelo ótimo da Tabela 4 é menor, suas previsões incorreriam em maior acurácia quando comparado ao segundo.

Diferente da primeira equação, um dos candidatos para a 3.2 não apresentou a característica de bidirecionalidade. Em comparação aos demais modelos que incluem esta característica nas unidades de rede LSTM, a rede apresentou desempenho inferior com um RMSE maior. Novamente, todas as redes apresentam camada convolucional em sua estrutura.

Tabela 5 – Seleção dos melhores modelos pelo algoritmo genético para a Equação 3.2

ep	cc	Istm	bi	att	neuronio	drop	def	lost	RMSE
485	Sim	2	Sim	Sim	(29,48)	34	15	MAE	0.0126
450	Sim	2	Não	Sim	(39,28)	21	13	MSE	0.0144
355	Sim	1	Sim	Sim	38	36	14	MSE	0.0135

Fonte: resultados obtidos a partir de estimação dos autores.

Por fim, ao contrário das especificações anteriores, a Tabela 6 considera a seleção de modelos para a Equação 3.3 considerando a variação absoluta (e não percentual) da volatilidade implícita. Assim, incorreria-se em erro comparar os valores

absolutos do RMSE dos candidatos a melhores modelos desta equação daqueles apresentados na Tabela 4 e na Tabela 5. Contudo, podemos destacar que os resultados das estimativas da rede da Equação 3.3 apresentam padrões de comportamentos distintos às demais equações.

Tabela 6 – Seleção dos melhores modelos pelo algoritmo genético para a Equação 3.3

ep	cc	Istm	bi	att	neuronio	drop	def	lost	RMSE
121	Sim	1	Sim	Sim	54	7	15	MSE	0.4906
450	Sim	2	Não	Sim	(39,28)	21	13	MSE	0.7275
495	Sim	3	Sim	Não	(38,13,26)	5	2	MSE	0.7238

Fonte: resultados obtidos a partir de estimação dos autores.

A começar pelo número de épocas, constata-se que o menor erro é obtido em uma quantidade de épocas consideravelmente menor do que os dois modelos selecionados para a equação de Fleming et al. (1995). Nota-se ainda que o aumento expressivo de defasagens do terceiro para o segundo modelo, sem alterações significativas no número de épocas, não é capaz de minimizar o valor de seu RMSE. Dessa forma, considera-se que o número de épocas também apresenta peso considerável sobre a performance ótima dos modelos ao considerar a volatilidade implícita em variação absoluta.

Neste caso, um número maior de camadas LSTM não foi capaz de melhorar o desempenho do modelo, considerando o RMSE elevado do modelo com 3 camadas quando comparado ao modelo com apenas uma camada. Além disso, a presença de bidirecionalidade nas unidades de rede LSTM novamente impactaram de forma positiva o desempenho do modelo, assim como é possível dizer que em conjunto, unidades de rede LSTM bidirecionais e o mecanismo de atenção, favoreceram o desempenho da rede.

Apesar do número de épocas para esta equação diferir das demais quanto à sua direção (neste caso quanto menor foi o número de épocas menor foi o erro), a definição de 15 defasagens para os retornos se manteve estável para os três cenários. A manutenção das camadas de convolução na rede nos três cenários revela um possível tipo de filtragem mais complexa que melhora a qualidade do sinal. É possível supor que os modelos tenham aprendido a filtrar ruídos ou algum tipo de artefato.

A partir desse resultado, somos capazes de pontuar algumas características que podem ser relevantes para a discussão teórica e prática no que se refere à relação entre volatilidade implícita e retornos aplicada ao mercado de petróleo. Este e outros pontos são abordados na seção 4.5.

4.5 Comparação entre os modelos

Nesta subseção são apresentados ao leitor os resultados comparativos entre a rede DCLSTM e os modelos clássicos de séries de tempo já discutidos entre as seções 3.3 e 3.6. Da mesma forma como realizado na subseção 4.4, os resultados são apresentados separadamente de acordo com a especificação funcional definida pelas equações 3.1, 3.2 e 3.3. Tratando-se dos modelos da classe ARMA, estimou-se os dois melhores candidatos para fins comparativos e o melhor dentre eles foi utilizado como benchmark para a especificação do modelo ARMA-GARCH.

A Tabela 7 apresenta as estimativas de todos os modelos considerando a especificação definida pela Equação 3.1. A performance da rede DCLSTM é consideravelmente maior quando consideramos as três métricas comparativas entre os modelos. Considerando que os dados tiveram o mesmo pré processamento, tal como descrito na subseção 3.1.2, é possível afirmar que o erro incorrido pela rede DCLSTM é consideravelmente menor que os demais. Destaca-se que apesar do modelo ARMA-GARCH(2,0,4)(1,2) considerar a volatilidade na execução dos modelo, resultando em estimativas mais consistentes, o modelo ARMA(2,0,4) se sobressai quanto à sua performance em termos de previsão, principal tarefa de interesse no uso dos nossos modelos.

Tabela 7 – Comparação entre os modelos para a Equação 3.1

Modelos de previsão	MAE	MSE	RMSE
DCLSTM	0.0085	0.0001	0.01236
ARMA(4,0,4)	0.0129	0.0032	0.01799
ARMA(2,0,4)	0.0129	0.0003	0.01797
ARMA-GARCH(2,0,4)(1,2)	0.0143	0.0004	0.02019
VAR	0.0337	0.0016	0.03937

Fonte: resultados obtidos a partir de estimação dos autores.

O mesmo ocorre para as estimativas apresentadas na Tabela 8. Ao adicionar o termo de retornos ao quadrado, é possível observar um comportamento semelhante

nos valores absolutos alcançados pelos modelos para as três métricas. Assim, em termos de previsão a adição dos retornos ao quadrado não produz impacto significativo sobre os resultados.

Tabela 8 – Comparação entre os modelos para a Equação 3.2

Modelos de previsão	MAE	MSE	RMSE
DCLSTM	0.0085	0.0002	0.01255
ARMA(4,0,4)	0.0129	0.0032	0.01799
ARMA(2,0,4)	0.0129	0.0003	0.01797
ARMA-GARCH(2,0,4)(1,2)	0.0143	0.0004	0.02019
VAR	0.0338	0.0016	0.03948

Fonte: resultados obtidos a partir de estimação dos autores.

Ao analisar a performance dos modelos apresentada na Tabela 9, nota-se que apesar da mudança de magnitude dos resultados dos erros devido à especificação considerar a variação absoluta da volatilidade implícita e não sua variação percentual, a rede DCLSTM ainda se sobressai sobre os demais modelos. O padrão de comportamento para a classe de modelos ARMA também se mantém. O modelo

ARMA-GARCH(2,0,4)(1,2) produz estimativas mais consistentes ao considerar a volatilidade da série na execução do modelo, porém incorre em um erro de previsão maior em todas as métricas.

Além dos resultados observados a partir das métricas de erro, é possível considerar alguns aspectos que surgem a partir de um panorama geral de comportamento dos dados em função dos modelos utilizados. Primeiro, o número de defasagens estimado pela rede DLSTM foi consideravelmente maior do que aqueles obtidos por meio dos demais modelos de estatística clássica e também se diferem dos resultados alcançados por mínimos quadrados ordinários. Isso mostra que, ao contrário das pressuposições feitas Low (2004), Fleming et al. (1995) e Hibbert et al. (2008), retornos quinzenais exercem impactos maiores sobre a volatilidade implícita do que retornos constatados na própria semana de negociação do mercado.¹²

¹²Considerando que a semana de operações de mercados financeiros tem duração de cinco dias úteis, este seria o número máximo de defasagens a impactar os retornos dentro de um intervalo de tempo semanal.

Tabela 9 – Comparação entre os modelos para a Equação 3.3

Modelos de previsão	MAE	MSE	RMSE
DCLSTM	0.3241	0.2407	0.4906
ARMA(1,0,4)	0.4189	0.3932	0.6271
ARMA(3,0,4)	0.4187	0.3936	0.6274
ARMA-GARCH(2,0,4)(1,2)	0.4709	0.5227	0.7229
VAR	0.6523	0.7116	0.7944

Fonte: resultados obtidos a partir de estimação dos autores.

Segundo, séries financeiras são marcadas pela presença de não linearidade e assimetria, como discutido anteriormente neste trabalho. Dessa forma, optar por modelos que não considerem tais características resultaria em performances inferiores àquelas alcançadas por modelos que englobem estes fatores em suas estimações. Sob a luz dessa explicação é possível compreender a performance inferior da classe de modelos ARMA quando comparado à rede, sendo o primeiro um modelo da classe de modelos lineares, ao contrário da rede neural que se adequa a uma estrutura de dados não lineares e assimétricos.

Terceiro, a utilização de mecanismos capazes de automatizar o processo de seleção e otimização de parâmetros é visto como um aliado na busca por modelos preditivos que incorram em erros menores. A utilização do algoritmo genético em paralelo ao uso do mecanismo de atenção contribui significativamente para um alcance adequado de estimativas dos hiper parâmetros da rede. Por outro lado, as estimações envolvendo mínimos quadrados ordinários e os demais modelos de estatística tradicional ainda dependem de análises manuais de outros instrumentos empíricos, tal como, por exemplo, o critério de informação Akaike. Além disso, a definição de parâmetros do modelo é realizada sob pressuposições teóricas que frequentemente não se aplicam ao mundo real.

5. CONCLUSÕES

A volatilidade implícita tem sido objeto de estudo em diversas abordagens empíricas na literatura que buscaram entender seu comportamento ao longo do tempo e sua relação com os retornos de ativos financeiros nos mercados. As três principais hipóteses teóricas que abordaram a relação entre estas duas variáveis foram as hipóteses de alavancagem, de feedback e a hipótese comportamental.

Estas três pressuposições teóricas foram amplamente discutidas e testadas empiricamente ao longo do tempo, mas principalmente sob a ótica dos mercados de ações. Posteriormente, as análises concentraram-se não só em verificar a validade destas hipóteses para outros mercados de relevância global (futuros - petróleo, ouro e outras commodities), assim como buscaram aplicar diversos modelos empíricos e testar suas performances.

Modelos dos campos de econometria de séries de tempo e de aprendizado de máquina são constantemente abordados na literatura como instrumentos capazes de modelar a relação entre retorno e volatilidade implícita e realizar previsões destas variáveis, especificamente da volatilidade implícita. Contudo, se verificou que em diversos contextos os modelos de aprendizado de máquina se sobressaíam sobre os de estatística clássica dada sua capacidade de modelar dados com alta complexidade, não-linearidade e ruído. Assim, este trabalho buscou verificar se a rede neural DCLSTM apresentava performance maior em modelar e prever a volatilidade implícita do mercado financeiro de petróleo norte-americano com relação aos retornos das opções negociadas neste mercado futuro. Os resultados da rede foram comparados aos modelos ARMA, ARMA-GARCH e VAR.

Os resultados empíricos do estudo revelaram não só a diferença de performance entre os modelos de séries de tempo utilizados na pesquisa, assim como demonstraram características estruturais do mercado financeiro de petróleo norte-americano por meio das análises de regressão. A presença de assimetria na relação entre volatilidade implícita e retornos neste mercado indica a aversão ao risco que os investidores possuem. Isso é constatado ao perceber que quando os investidores incorrem em prejuízos de seus ativos, o impacto desse efeito sobre a volatilidade implícita (risco) do petróleo é maior quando comparada a situação em que eles auferem ganhos. Em outras palavras, as pessoas buscam maior proteção de seus

recursos quando os perdem, do que arriscam mais quando identificam ganhos em suas escolhas.

A partir das análises de regressão quantílica, outras características estruturais da relação de risco-retorno do mercado de petróleo foram evidenciadas. A relação manteve-se negativa para todos os quantis. Isso significa dizer que ao longo de toda a amostra de dados, quando o risco (retorno) varia positivamente (negativamente), os retornos (risco) das opções de petróleo no mercado futuro variam negativamente (positivamente). Na prática, à medida que o risco de se investir em uma opção no mercado futuro de petróleo aumenta, os retornos atrelados àquela opção tendem a ser negativos ao longo do tempo. As regressões quantílicas também possibilitaram confirmar que esse comportamento não ocorre de forma linear à medida que o intervalo de dados disponíveis é maior.

Quanto às performances entre os modelos de séries de tempo, as hipóteses iniciais do trabalho quanto ao desempenho superior da rede neural em comparação aos demais foi confirmada. Dadas as medidas de qualidade de previsão e acurácia apresentadas no estudo, o índice de erro de previsibilidade da rede DCLSTM foi menor do que todos os demais modelos para todas as equações propostas. Esse resultado evidencia a capacidade da rede neural profunda em lidar com séries financeiras de tempo que apresentem não linearidade, ruído e alta complexidade. Dessa forma, a rede neural não apresentaria dificuldades maiores em modelar séries de informações diárias de outros ativos e índices financeiros, por exemplo. Isso indica a eficiência de uso da rede em contextos diversos do mercado financeiro que envolvam processos de tomada de decisão.

Pode-se dizer que dentre os hiper parâmetros selecionados para treinar a rede DCLSTM, o número de épocas apresentou destaque dentre os demais. Isso se justifica dado que no momento em que se escolheu a melhor especificação para a rede neural, a variação no número de épocas definido era a principal responsável por diminuir a magnitude das métricas de erro. Assim, conclui-se que à medida que a rede neural completou mais ciclos de aprendizado utilizando a amostra deste estudo, maior foi sua eficiência de aprendizagem para prever resultados melhores.

Ressalta-se também o grau elevado de precisão da rede que considerou em sua estrutura unidades LSTM bidirecionais e a inclusão do mecanismo de atenção para selecionar as informações que apresentavam maior importância para prever os dados. Além disso, a inserção do algoritmo genético contribuiu em grande medida

para que os parâmetros e hiper-parâmetros fossem selecionados corretamente a fim de alcançar maior performance.

REFERÊNCIAS

ABOURA, S.; CHEVALLIER, J. Leverage vs. feedback: Which effect drives the oil market? *Finance Research Letters*, Elsevier, v. 10, n. 3, p. 131–141, 2013.

AGBEYEGBE, T. D. An inverted u-shaped crude oil price return-implied volatility relationship. *Review of Financial Economics*, Elsevier, v. 27, p. 28–45, 2015.

AGNOLUCCI, P. Volatility in crude oil futures: A comparison of the predictive ability of garch and implied volatility models. *Energy Economics*, Elsevier, v. 31, n. 2, p. 316–321, 2009.

ALOUI, C.; MABROUK, S. Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed garch models. *Energy Policy*, Elsevier, v. 38, n. 5, p. 2326–2339, 2010.

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.

BAŠTA, M.; MOLNÁR, P. Oil market volatility and stock market volatility. *Finance Research Letters*, Elsevier, v. 26, p. 204–214, 2018.

BEKAERT, G.; WU, G. Asymmetric volatility and risk in equity markets. *The review of financial studies*, Oxford University Press, v. 13, n. 1, p. 1–42, 2000.

BERTRAND, J. Théorie mathématique de la richesse sociale. *Journal des savants*, Paris, v. 67, n. 1883, p. 499–508, 1883.

BIELISZCZUK, B. et al. The collapse of oil prices: Economic and political implications. *Polski Instytut Spraw Międzynarodowych*, 2020.

BLACK, F. Studies of stock market volatility changes. 1976 *Proceedings of the American Statistical Association Business and Economic Statistics Section*, American Statistical Association, 1976.

BOLLERSLEV, T. et al. A discrete-time model for daily s & p500 returns and realized variations: Jumps and leverage effects. *Journal of Econometrics*, Elsevier, v. 150, n. 2, p. 151–166, 2009.

BOLLERSLEV, T.; ZHOU, H. Volatility puzzles: a simple framework for gauging return-volatility regressions. *Journal of Econometrics*, Elsevier, v. 131, n. 1-2, p. 123–150, 2006.

BOX, G. E.; JENKINS, G. M.; REINSEL, G. Time series analysis: forecasting and control holden-day san francisco. *BoxTime Series Analysis: Forecasting and Control Holden Day*1970, 1970.

CAMPBELL, J. Y.; HENTSCHEL, L. No news is good news: An asymmetric model of changing volatility in stock returns. 1991.

CBOE. Cboe Crude Oil ETF Volatility Index (OVX). 2020. Disponível em: <<http://www.cboe.com/products/vix-index-volatility/volatility-on-etfs/>> cboe-crude-oil-etf-volatility-index-ovx>.

CHAKRA, N. C. et al. An innovative neural forecast of cumulative oil production from a petroleum reservoir employing higher-order neural networks (honns). *Journal of Petroleum Science and Engineering*, Elsevier, v. 106, p. 18–33, 2013.

CHENG, J. et al. Application of arima model in financial time series in stocks. In: SPRINGER. *International Conference on Artificial Intelligence and Security*. [S.I.], 2020. p. 232–243.

CHEONG, C. W. Modeling and forecasting crude oil markets using arch-type models. *Energy policy*, Elsevier, v. 37, n. 6, p. 2346–2355, 2009.

CHOROWSKI, J. et al. Attention-based models for speech recognition. *arXiv preprint arXiv:1506.07503*, 2015.

CHOW, G. C. Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 591–605, 1960.

CHRISTIE, A. A. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial Economics*, Elsevier, v. 10, n. 4, p. 407–432, 1982.

DAIGLER, R. T.; HIBBERT, A. M.; PAVLOVA, I. Examining the return–volatility relation for foreign exchange: evidence from the euro vix. *Journal of Futures Markets*, Wiley Online Library, v. 34, n. 1, p. 74–92, 2014.

DAVIS, L. *Handbook of genetic algorithms*. [S.I.]: CUMINCAD, 1991.

DENNIS, P.; MAYHEW, S.; STIVERS, C. Stock returns, implied volatility innovations, and the asymmetric volatility phenomenon. *Journal of Financial and Quantitative Analysis*, JSTOR, p. 381–406, 2006.

DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, Taylor & Francis, v. 74, n. 366a, p. 427–431, 1979.

DICKEY, D. A.; FULLER, W. A. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: journal of the Econometric Society*, JSTOR, p. 1057–1072, 1981.

DRECHSLER, I. Uncertainty, time-varying fear, and asset prices. *The Journal of Finance*, Wiley Online Library, v. 68, n. 5, p. 1843–1889, 2013.

DRECHSLER, I.; YARON, A. What's vol got to do with it. *The Review of Financial Studies*, Society for Financial Studies, v. 24, n. 1, p. 1–45, 2011.

ELSKEN, T. et al. Neural architecture search: A survey. *J. Mach. Learn. Res.*, v. 20, n. 55, p. 1–21, 2019.

ENDERS, W. Applied econometric time series. *Privredna kretanja i ekonomska politika*, v. 132, p. 93, 2012.

FIGLEWSKI, S.; WANG, X. Is the "leverage effect" a leverage effect? 2000.

FLEMING, J.; OSTDIEK, B.; WHALEY, R. E. Predicting stock market volatility: A new measure. *Journal of Futures Markets*, Wiley Online Library, v. 15, n. 3, p. 265–302, 1995.

FRENCH, K. R.; SCHWERT, G. W.; STAMBAUGH, R. F. Expected stock returns and volatility. *Journal of financial Economics*, v. 19, n. 1, p. 3, 1987.

GRAVES, A.; SCHMIDHUBER, J. Framewise phoneme classification with bidirectional Lstm and other neural network architectures. *Neural networks*, Elsevier, v. 18, n. 5-6, p. 602–610, 2005.

GREFF, K. et al. Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, IEEE, v. 28, n. 10, p. 2222–2232, 2016.

HAMILTON, J. D. (1994), *Time Series Analysis*. [S.I.]: Princeton: Princeton University Press, 1994.

HENRIQUES, I.; SADORSKY, P. The effect of oil price volatility on strategic investment. *Energy Economics*, Elsevier, v. 33, n. 1, p. 79–87, 2011.

HERMANS, M.; SCHRAUWEN, B. Training and analysing deep recurrent neural networks. In: *Advances in neural information processing systems*. [S.I.: s.n.], 2013. p. 190–198.

HIBBERT, A. M.; DAIGLER, R. T.; DUPOYET, B. A behavioral explanation for the negative asymmetric return–volatility relation. *Journal of Banking & Finance*, Elsevier, v. 32, n. 10, p. 2254–2266, 2008.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural computation*, MIT Press, v. 9, n. 8, p. 1735–1780, 1997.

HÜSKEN, M.; STAGGE, P. Recurrent neural networks for time series classification. *Neurocomputing*, Elsevier, v. 50, p. 223–235, 2003.

HYNDMAN, R. Measuring forecast accuracy. In: M. GILLILAND, L. TASHMAN, U. SGLAV (EDS.). *Business Forecasting: Practical Problems and Solutions*. [S.I.], 2016. p. 177—183.

JAMMAZI, R.; ALOUI, C. Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. *Energy Economics*, Elsevier, v. 34, n. 3, p. 828–841, 2012.

KABOUDAN, M. Compumetric forecasting of crude oil prices. v. 1, p. 283–287, 2001.

KANG, S. H.; KANG, S.-M.; YOON, S.-M. Forecasting volatility of crude oil markets. *Energy Economics*, Elsevier, v. 31, n. 1, p. 119–125, 2009.

KIM, D.; KON, S. J. Alternative models for the conditional heteroscedasticity of stock returns. *Journal of Business*, JSTOR, p. 563–598, 1994.

LÄNGKVIST, M.; KARLSSON, L.; LOUTFI, A. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, Elsevier, v. 42, p. 11–24, 2014.

LÄNGKVIST, M.; KARLSSON, L.; LOUTFI, A. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, Elsevier, v. 42, p. 11–24, 2014.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.

LOW, C. The fear and exuberance from implied volatility of s&p 100 index options. *The Journal of Business*, JSTOR, v. 77, n. 3, p. 527–546, 2004.

MAS-COLELL, A. et al. *Microeconomic theory*. [S.I.]: Oxford university press New York, 1995. v. 1.

MERTON, R. C. An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 867–887, 1973.

MOHAMMADI, H.; SU, L. International evidence on crude oil price dynamics: Applications of arima-garch models. *Energy Economics*, Elsevier, v. 32, n. 5, p. 1001–1008, 2010.

NARAYAN, P. K. et al. Do oil prices predict economic growth? new global evidence. *Energy Economics*, Elsevier, v. 41, p. 137–146, 2014.

NOGUERA-SANTAELLA, J. Geopolitics and the oil price. *Economic Modelling*, Elsevier, v. 52, p. 301–309, 2016.

PADUNGSAKSAWASDI, C.; DAIGLER, R. T. The return-implied volatility relation for commodity etfs. *Journal of Futures Markets*, Wiley Online Library, v. 34, n. 3, p. 261–281, 2014.

PASCANU, R.; MIKOLOV, T.; BENGIO, Y. On the difficulty of training recurrent neural networks. p. 1310–1318, 2013.

POTERBA, J. M.; SUMMERS, L. H. The persistence of volatility and stock market fluctuations. 1984.

RAFIQ, S.; SALIM, R.; BLOCH, H. Impact of crude oil price volatility on economic activities: An empirical investigation in the thai economy. *Resources Policy*, Elsevier, v. 34, n. 3, p. 121–132, 2009.

RAINVILLE, F. D. et al. Distributed evolutionary algorithms in python (deap). 2012.

RAJPUT, H. et al. A shock like no other: Coronavirus rattles commodity markets. *Environment, Development and Sustainability*, Springer, p. 1–12, 2020.

RIO, C. P. D. V. D.; SUL, P. D.; VASSALLI, L. C. Aplicação de redes neurais Istm para a previsão de.

SAGHEER, A.; KOTB, M. Time series forecasting of petroleum production using deep Istm recurrent networks. *Neurocomputing*, Elsevier, v. 323, p. 203–213, 2019.

SCHUSTER, M.; PALIWAL, K. K. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, Ieee, v. 45, n. 11, p. 2673–2681, 1997.

SCHWERT, G. W. Why does stock market volatility change over time? *The journal of finance*, Wiley Online Library, v. 44, n. 5, p. 1115–1153, 1989.

SCHWERT, G. W. Why does stock market volatility change over time? *The journal of finance*, Wiley Online Library, v. 44, n. 5, p. 1115–1153, 1989.

SEDLÁKOVÁ, M. Application of arma and garch models on time series of komerční banka stocks. *ACC Journal*, Technická univerzita v Liberci, Česká republika, 2020.

SILVA JUNIOR, J. C. A. D. An s-shaped crude oil price return-implied volatility relation: Parametric and nonparametric estimations. *International Journal of Economics and Finance*, v. 9, n. 12, 2017.

SMITH, S. W. et al. *The scientist and engineer's guide to digital signal processing*. California Technical Pub. San Diego, 1997.

SPIEGEL, S. et al. Pattern recognition and classification for multivariate time series. In: Proceedings of the fifth international workshop on knowledge discovery from sensor data. [S.I.: s.n.], 2011. p. 34–42.

SUN, H.; YU, B. Forecasting financial returns volatility: A garch-svr model. *Computational Economics*, Springer, v. 55, n. 2, p. 451–471, 2020.

SUTSKEVER, I. Training recurrent neural networks. [S.I.]: University of Toronto Toronto, Ontario, Canada, 2013.

TAUCHEN, G.; ZHANG, H.; LIU, M. Volume, volatility, and leverage: A dynamic analysis. *Journal of Econometrics*, Elsevier, v. 74, n. 1, p. 177–208, 1996.

TVERSKY, A.; KAHNEMAN, D. Prospect theory: An analysis of decision under risk. *Econometrica*, v. 47, n. 2, p. 263–291, 1979.

UTGOFF, P. E.; STRACUZZI, D. J. Many-layered learning. *Neural Computation*, MIT Press, v. 14, n. 10, p. 2497–2529, 2002.

VARIAN, H. *Microeconomia*. [S.I.]: Elsevier Brasil, 2017.

VASCO, L. P. Um estudo de redes neurais recorrentes no contexto de previsões no mercado financeiro. Universidade Federal de São Carlos, 2020.

VASWANI, A. et al. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.

WEI, Y.; WANG, Y.; HUANG, D. Forecasting crude oil market volatility: Further evidence using garch-class models. *Energy Economics*, Elsevier, v. 32, n. 6, p. 1477–1484, 2010.

WHITLEY, D. A genetic algorithm tutorial. *Statistics and computing*, Springer, v. 4, n. 2, p. 65–85, 1994.

WILLMOTT, C. J. On the validation of models. *Physical geography*, Taylor & Francis, v. 2, n. 2, p. 184–194, 1981.

WU, Y.-X.; WU, Q.-B.; ZHU, J.-Q. Improved eemd-based crude oil price forecasting using lstm networks. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 516, p. 114–124, 2019.

YAHOO FINANCE, S. Yahoo Finance. 2020. Disponível em: <<https://finance.yahoo.com/>>.

YU, L. et al. A hybrid grid-ga-based lssvr learning paradigm for crude oil price forecasting. *Neural computing and applications*, Springer, v. 27, n. 8, p. 2193–2215, 2016.

YU, L.; XU, H.; TANG, L. Lssvr ensemble learning with uncertain parameters for crude oil price forecasting. *Applied Soft Computing*, Elsevier, v. 56, p. 692–701, 2017.

YU, L.; ZHANG, X.; WANG, S. Assessing potentiality of support vector machine method in crude oil price forecasting. *Eurasia J Math Sci Technol Educ*, v. 13, n. 12, p. 7893–7904, 2017.

ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, Elsevier, v. 14, n. 1, p. 35–62, 1998.