

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

GABRIEL SOARES MIRANDA

**ESTIMAÇÃO DA VARIÂNCIA REALIZADA DO ÍNDICE
EWZ UTILIZANDO REDES NEURAIS ARTIFICIAIS**

SÃO PAULO

2020

GABRIEL SOARES MIRANDA

ESTIMAÇÃO DA VARIÂNCIA REALIZADA DO ÍNDICE EWZ UTILIZANDO REDES NEURAIS ARTIFICIAIS

Dissertação apresentada à Escola de Economia de São Paulo da Fundação Getúlio Vargas, como requisito para a obtenção do título de Mestre em Economia.

Área de concentração:
Engenharia Financeira.

Orientador:
Prof. Dr. Afonso de Campos Pinto

SÃO PAULO

2020

Miranda, Gabriel Soares.

Estimação da variância realizada do índice EWZ utilizando redes neurais artificiais / Gabriel Soares Miranda. - 2020.

78 f.

Orientador: Prof. Dr. Afonso de Campos Pinto.

Dissertação (mestrado profissional MPFE) – Fundação Getulio Vargas, Escola de Economia de São Paulo.

1. Redes neurais (Computação). 2. Análise de variância. 3. Volatilidade (Finanças). 4. Índices de mercado de ações. 5. Mercado financeiro. I. Pinto, Afonso de Campos. II. Dissertação (mestrado profissional MPFE) – Escola de Economia de São Paulo. III. Fundação Getulio Vargas. IV. Título.

CDU 007.522.2

GABRIEL SOARES MIRANDA

ESTIMAÇÃO DA VARIÂNCIA REALIZADA DO ÍNDICE EWZ UTILIZANDO REDES NEURAIS ARTIFICIAIS

Dissertação apresentada à Escola de Economia de São Paulo da Fundação Getúlio Vargas, como requisito para a obtenção do título de Mestre em Economia.

Área de concentração:
Engenharia Financeira.

Data da Aprovação: 24/11/2020

Banca Examinadora:

Prof. Dr. Afonso de Campos Pinto
(Orientador)
EESP-FGV

Prof. Dr. Élia Yathie Matsumoto
EESP-FGV

**Prof. Dr. Flavio Almeida de
Magalhães Cipparrone**
USP-Escola Politécnica

Aos meus queridos pais, Magnólia e Luiz, e à minha amada noiva Heloisa. Esta conquista também é de vocês.

AGRADECIMENTOS

Gostaria de agradecer primeiramente à minha mãe, responsável por me dar uma boa educação e por me apoiar em todas as etapas até aqui. Um agradecimento também ao meu pai por todo o suporte dado ao longo da minha jornada. Os valores ensinados e o amor compartilhado foram meu alicerce desde o início. Todas as minhas conquistas foram graças a eles.

Gostaria de agradecer também à minha companheira Heloisa. Seu carinho, sua paciência, sua compreensão, seu incentivo e seu amor foram fundamentais em vários momentos ao longo do tempo que estamos juntos, em especial para a realização deste mestrado.

Obrigado aos meus tios, Nivaldo e Mauro, que acreditaram em mim, investiram em minha educação e me mostraram que era possível ir além do que era esperado.

Um muito obrigado ao meu professor e orientador Afonso por todos os ensinamentos e toda a atenção dada ao longo do curso e em especial durante a realização deste trabalho.

Agradeço também ao demais professores da FGV e aos meus colegas de classe com quem compartilhei aprendizados, alegrias e preocupações ao longo do curso.

RESUMO

Este trabalho apresentou a aplicação de uma nova técnica de estimação da variância realizada (RV) a partir de redes neurais artificiais (RNA) e comparou seu desempenho com os modelos de regressão linear múltipla (RLM) e de volatilidade implícita (IV) como preditor, a qual é obtida a partir do *Volatility Index*, VIX. O desempenho dos modelos é medido pelo R^2 ajustado. A nova técnica proposta possui duas abordagens, a primeira utiliza a RV como vetor resposta da RNA enquanto a segunda utiliza o prêmio de risco de variância (VRP), o qual é composto pela combinação do VR e do VIX. As informações de mercado utilizadas neste trabalho são compreendidas entre 06 de Janeiro de 2016 e 01 de Maio de 2020 e foram coletadas a partir de um terminal Bloomberg. O ativo escolhido para o estudo é o índice MSCI Brazil, EWZ, negociado na bolsa de valores americanas NYSE. Os modelos são testados em quatro cenários distintos e observa-se que os desempenhos superiores ocorrem quando não há períodos de crise. Evidenciamos que o modelo de RNA com o VRP como vetor resposta apresenta o melhor desempenho entre os modelos estudados, e que este mesmo modelo, além do cenário sem crise, apresenta um desempenho superior quando há um período de crise em ambos os dados de treinamento e teste quando comparados com os cenários onde se tem uma crise apenas em um ou apenas em outro. Como o modelo de RNA com a RV como vetor resposta apresenta um desempenho consideravelmente inferior ao seu par, é válido ressaltar a importância do VIX em aumentar a acurácia deste modelo de RNA que utiliza o VRP como vetor resposta ao exercer um papel de normalizador.

Palavras-chave: Rede neural artificial. VIX. Variância realizada.

ABSTRACT

The present dissertation is focused on applying a new technique for estimating realized variance (RV) using Artificial Neural Network (ANN) and comparing its performance against other benchmark models as Multiple Linear Regression (MLR) and Implied Volatility (IV) like a predictor, which is obtained from the Volatility Index (VIX). All models performance is measured by adjusted R^2 . The new proposed technique has two approaches, the first one uses the RV as the desirable information while the second one uses the Variance Risk Premium (VRP), which is the result of RV and VIX combination. The market information used in this work comprehends between January 6th 2016 and May 1st 2020, and it has been collected through a bloomberg terminal. The underlying asset chosen was the MSCI Brazil index, EWZ, which is traded on NYSE. All models have been tested in four different scenarios and higher performances are found in periods without crisis. It is shown in this work that the ANN model second approach has the best overall performance, and this same model, besides the scenario with no crisis, presents better results when it has crisis periods in both training and testing phases instead of when it has crisis periods only in training or only in testing phases. As the ANN model first approach considerably underperforms its pair, the second approach, we can highlight the importance of VIX in enhancing the model by playing a normalizer role in the process.

Keywords: Artificial neural network. VIX. Realized variance.

LISTA DE ILUSTRAÇÕES

Figura 1 – Modelo de um Neurônio Não Linear	25
Figura 2 – Gráfico da Função de Limiar	27
Figura 3 – Gráfico da Função de Sigmóide	27
Figura 4 – Exemplos de Arquiteturas de RNAs	28
Figura 5 – Fluxo de informações de um neurônio na camada de saída	31
Figura 6 – Fluxo de informações de um neurônio na camada intermediária	33
Figura 7 – Exemplo de janela móvel para amostra inicial de 500 dias.	44

LISTA DE TABELAS

Tabela 1	– Corte da base de entrada das opções.	38
Tabela 2	– Exemplo de tratamento para dois <i>bids</i> consecutivos para um vencimento em uma data base	41
Tabela 3	– R^2 ajustado dos modelos RNA_{RV} e RNA_{VRP} para amostra inicial de 400 dias nos cenários 1 e 2.	51
Tabela 4	– R^2 ajustado para os modelos VIX , RNA_{RV} , RNA_{VRP} , RLM_{VR} e RLM_{VRP} no cenário 1	53
Tabela 5	– R^2 ajustado para os modelos VIX , RNA_{RV} , RNA_{VRP} , RLM_{VR} e RLM_{VRP} no cenário 2	54
Tabela 6	– R^2 ajustado para os modelos VIX , RNA_{RV} , RNA_{VRP} , RLM_{VR} e RLM_{VRP} no cenário 3	55
Tabela 7	– R^2 ajustado para os modelos VIX , RNA_{RV} , RNA_{VRP} , RLM_{VR} e RLM_{VRP} no cenário 4	57
Tabela 8	– Modelos com maior R^2 ajustado para cada cenário e para cada configuração de número de opções por dia utilizadas.	58
Tabela 9	– Dias disponíveis para cada quantidade de opções diárias utilizadas e diminuição média dos dias disponíveis a cada aumento no número de opções.	58
Tabela 10	– Maiores R^2 ajustados do modelo VIX para cada cenário e para cada configuração de número de opções por dia utilizadas.	59
Tabela 11	– R^2 ajustado do modelo VIX para cada cenário e para cada combinação de amostra inicial e de número de opções por dia utilizadas.	60
Tabela 12	– Maiores R^2 ajustados do modelo RNA_{VRP} para cada cenário e para cada configuração de número de opções na camada de entrada.	61
Tabela 13	– R^2 ajustado do modelo RNA_{VRP} para cada cenário e para cada combinação de amostra inicial e de número de opções utilizadas na camada de entrada da RNA.	63
Tabela 14	– R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 400 dias nos cenários 3 e 4.	71
Tabela 15	– R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 500 dias nos cenários 1 e 2.	73
Tabela 16	– R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 500 dias nos cenários 3 e 4.	74
Tabela 17	– R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 600 dias nos cenários 1 e 2.	75

Tabela 18 – R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 600 dias nos cenários 3 e 4.	76
Tabela 19 – R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 700 dias nos cenários 1 e 2.	77
Tabela 20 – R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 700 dias nos cenários 3 e 4.	78

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO BIBLIOGRÁFICA	15
2.1	Modelos GARCH	15
2.2	Modelos Estocásticos	16
2.3	Predição da Volatilidade a Partir dos Preços de Opções e das IVs	17
2.4	Modelos que utilizam Redes Neurais Artificiais	17
3	FUNDAMENTAÇÃO TEÓRICA	20
3.1	Volatilidade Implícita	20
3.1.1	Volatilidade Implícita a Partir de Modelos de Apreçamento de Opções	20
3.1.2	Volatilidade Implícita a Partir do VIX	22
3.2	Variância Realizada (RV)	23
3.3	Prêmio de Risco de Variância (VRP)	23
3.4	Redes Neurais Artificiais	25
3.4.1	O Algoritmo <i>Backpropagation</i>	29
3.4.2	O Otimizador L-BFGS	35
4	METODOLOGIA	37
4.1	Dados de entrada	37
4.2	Tratamento dos Dados	39
4.2.1	Tratamento das Opções	39
4.3	Seleção das opções	42
4.4	Pré processamento dos dados	43
4.5	Cálculo da Variância Realizada dos Próximos T Dias	43
4.6	Janela Móvel	44
4.7	Modelos de previsão da variância realizada	44
4.7.1	Variância a partir do VIX	45
4.7.2	Regressão Linear Múltipla	46
4.7.3	Redes Neurais Artificiais	46
5	RESULTADOS	50
5.1	Otimização do coeficiente de penalização (λ)	50
5.2	Análise de desempenho da estimação da variância realizada	52
5.2.1	Cenário 1 - Sem filtro de datas	52
5.2.2	Cenário 2 - Filtro <i>Joesley Day</i>	54
5.2.3	Cenário 3 - Filtro <i>Covid</i>	55

5.2.4	Cenário 4 - Filtros <i>Joesley Day</i> e <i>Covid</i>	56
5.3	Comparativo entre os cenários de teste para os modelos <i>VIX</i> e <i>RNA_{VRP}</i> . .	57
5.3.1	Análise comparativa do modelo <i>VIX</i> para os 4 cenários disponíveis	58
5.3.2	Análise comparativa do modelo <i>RNA_{VRP}</i> para os 4 cenários disponíveis	61
6	CONCLUSÃO	64
	REFERÊNCIAS	66
	Apêndices	70
	APÊNDICE A VARIAÇÃO DO COEFICIENTE DE PENALIZAÇÃO(λ) . . .	71

1 INTRODUÇÃO

Quando se fala em risco no mercado financeiro muitas vezes se discute a volatilidade dos retornos dos ativos. Por ser uma variável não observável, a volatilidade pode ser estimada de diversas formas. Os seus modelos têm como objetivo principal prover uma métrica que pode ser utilizada na gestão de riscos financeiros, no auxílio da seleção de carteiras de ativos e na formação de preços de derivativos (GALDI; PEREIRA, 2007). Dessa forma, diferentes abordagens têm sido utilizadas para a sua estimação nas últimas décadas, como as econométricas que se baseiam em modelos autorregressivos, as estocásticas que introduzem a ideia de que a volatilidade é regida pelo seu próprio processo estocástico, as implícitas que utilizam os preços dos derivativos do ativo base e as de redes neurais, mais recentes, que se baseiam em modelos de *Machine Learning* (ML).

Essa estimativa da volatilidade é fundamental para o gerenciamento de risco de uma carteira. Neste trabalho, os termos volatilidade e variância serão discutidos no mesmo contexto, entendendo-se que a variância é obtida a partir do quadrado da volatilidade estimada. Seguindo essa linha, uma ferramenta comumente utilizada pelo mercado para este gerenciamento de risco supracitado é o prêmio de risco de variância (VRP), que pode ser negociado através de um *swap* de variância, o qual representa a recompensa que o investidor recebe por assegurar outro investidor contra movimentos extremos do mercado. As pessoas buscam a proteção contra movimentos adversos e indesejáveis do mercado financeiro, assim como o fazem em outros campos, através de seguros de carro ou de vida (BESADA, 2019). Este *swap* descrito acima é um produto negociado entre duas partes, em que o pagamento por uma parte, e por consequência o recebimento pela outra, é calculado através da diferença entre a variância esperada sob a medida neutra a risco e a variância estimada sob a medida física do mundo real observável. Em outras palavras, é a diferença entre a variância futura esperada pelo mercado determinada a priori e a variância realizada (RV) no período estipulado estimada a posteriori. Vale ressaltar que analogamente também existe o prêmio de risco de volatilidade, que pode ser negociado através do *swap* de volatilidade.

No que diz respeito à expectativa desta volatilidade a ser realizada no futuro, o VIX, *Volatility Index*, surge como uma alternativa para sua estimativa. Ele é um índice calculado a partir dos preços de contratos derivativos, neste caso opções, do índice norte-americano SP500¹, e reflete a expectativa do mercado em relação à volatilidade que será realizada

¹ Standard Poor's 500 é um índice composto pelos preços das 500 empresas americanas listadas em bolsa com os maiores valores de mercado. A participação de cada empresa na composição deste índice é balanceada pelo seu valor de mercado, isto é, quanto maior o valor de mercado da empresa maior o seu peso no índice.

nos próximos 30 dias corridos. A sua metodologia de cálculo já é aplicada em outros ativos como por exemplo o índice EWZ ², formando o VXEZW. Isso pode ser estendido e aplicável a qualquer outro ativo que possua opções negociados em bolsa. Neste trabalho o termo VIX será utilizado no sentido de sua metodologia e não diretamente relacionado ao VIX do SP500. Seguindo este contexto, a diferença entre o VIX calculado em uma determinada data e a volatilidade realizada do ativo base estimada depois de 30 dias corridos caracteriza o prêmio de risco de volatilidade.

Por conta da aversão ao risco e a tendência de superestimar a probabilidade de eventos extremos, estes *swaps* são bastante demandados e, algumas vezes, sobreprecificados. Assim, neste trabalho é proposto um modelo de estimação da variância realizada com o objetivo de superar o poder preditivo do VIX, cujo modelo de cálculo sera considerado como um *benchmark* na etapa de comparação, para que se tenha apreçamentos mais acurados. A abordagem aqui proposta vai em direção a uma tendência observada nos últimos anos, não apenas no mercado financeiro mas em diversos outros campos, que é a utilização do aprendizado de máquina (ML) na predicabilidade de um determinado objeto de estudo. A evolução do poder computacional das últimas décadas viabilizou a aplicação de algoritmos mais robustos em diferentes modelagens computacionais, dando ao ML cada vez mais espaço para mostrar os benefícios de seu uso.

Será apresentado um modelo de *Multilayer Perceptron* (MLP), considerado uma classe de rede neural artificial (RNA), em que se inserem os preços das opções do índice EWZ e se obtém a estimação da variância que será realizada nos próximos 30 dias corridos. A escolha da estimação da variância ao invés da volatilidade se deu pelo fato de que a variância possui uma fundamentação teórica mais significativa. Isso se dá pois a melhor forma de se precificar um produto é através de um portfólio replicante, e o *swap* que pode ser replicado usando opções com diferentes *strikes* é o swap de variância (DERMAN; KANI, 1994). Serão utilizadas duas abordagens, a primeira é a estimação feita diretamente em relação à variância realizada e a segunda é a estimação da variância realizada feita através da previsão do prêmio de risco de variância.

Além dessa introdução, o trabalho conta com cinco capítulos. No capítulo 2 desenvolve-se a revisão bibliográfica. No capítulo 3 é feita uma fundamentação teórica dos conceitos e das ferramentas utilizadas no trabalho. No capítulo 4 são detalhados a utilização e o tratamento dos dados, a seleção das opções e por fim como os modelos são treinados e testados. No capítulo 5 são apresentados os resultados dos testes, comparando-se o desempenho do modelo proposto com o do modelo *benchmark*. Por fim, o capítulo final 6 contém as conclusões e sugestões de possíveis trabalhos futuros.

² O EWZ é um Exchange-Traded Fund(ETF), que contém uma cesta de ações de médio e grande porte da bolsa brasileira. Mais informações em <https://www.ishares.com/us/literature/fact-sheet/ewz-ishares-msci-brazil-etf-fund-fact-sheet-en-us.pdf>

2 REVISÃO BIBLIOGRÁFICA

Este capítulo tem por objetivo revisar brevemente a literatura existente sobre a modelagem e a previsão das volatilidades implícitas (IVs) e realizadas (RVol), bem como descrever alguns estudos sobre modelos de Aprendizado de Máquinas (ML) aplicados a finanças.

2.1 Modelos GARCH

No passado, os primeiros modelos baseados em séries temporais para previsão de RVol mostraram que as medidas de volatilidade histórica (HV) apresentavam predicabilidade das futuras RVols. Engle (1982) introduz o modelo auto regressivo de heteroscedasticidade condicional (ARCH), baseado em informações passadas recentes, para as médias e variâncias da inflação no Reino Unido. Bollerslev (1986) generaliza o modelo de Engle ao adotar uma estrutura de defasagem para a variância.

Desde então, outros trabalhos apresentaram ramificações destes modelos auto regressivos, principalmente pela constatação de que eles são muito rígidos para serem utilizados em séries temporais, principalmente para períodos mais longos, devido ao fato da persistência estimada das variâncias condicionais serem próximas a um (BAUWENS; HAFNER; LAURENT, 2012). Dando continuidade aos modelos clássicos de ARCH e GARCH, Nelson (1991) introduziu um modelo exponencial de ARCH (EGARCH) que permite correlação entre os retornos e as inovações de volatilidade (diferença entre a melhor previsão de volatilidade futura e a volatilidade observada), elimina a necessidade de restrições de desigualdade nos parâmetros e permite interpretações de persistência dos choques na volatilidade. Engle e Lee (1993) sugeriram um modelo de duas equações, cada uma representando os componentes *short-run* e *long-run*¹ da volatilidade.

Paolella, Haas e Mittnik (2004a) apresentaram um modelo que permite escolher entre diversos regimes de mercado em cada instante de tempo. Porém, este modelo assume que as variáveis usadas para decidir entre estes regimes são todas independentes ao longo do tempo. Para corrigir este problema, Paolella, Haas e Mittnik (2004b) propuseram um modelo *Markov-Switching*, onde os parâmetros do modelo GARCH mudam de acordo com um processo de Markov. Uma extensão deste trabalho foi proposta por Paolella e Haas (2012).

¹ O termo *short-run* diz respeito a um intervalo de tempo em que a quantidade de pelo menos um dado de entrada é fixada e a quantidade dos outros dados pode ser variada. Enquanto o termo *long-run* diz respeito a um intervalo de tempo em que a quantidade de todos os dados de entrada pode ser variada.

2.2 Modelos Estocásticos

Este grupo de modelos assume que a volatilidade é regida pelo seu próprio processo estocástico. A primeira abordagem deste tema foi introduzida por Taylor (1982) como uma aproximação de Euler do modelo de difusão do ativo objeto. Assumindo que os preços das opções seguem um movimento Browniano, Heston (1993) apresentou um modelo onde a volatilidade segue um processo de Ornstein–Uhlenbeck, onde o termo de difusão é a raiz quadrada da variância do ativo e o *drift* apresenta reversão à média.

Danielsson (1994) apresentou um modelo de volatilidade estocástica (SV) que utiliza um método de Monte Carlo, chamado de Máxima Verossimilhança Simulada, para obter a densidade marginal dos preços, que é usada para estimar as dinâmicas dos preços diários dos ativos. Este método também foi utilizado por Durbin e Koopman (1997) para obter estimativas dos hiper parâmetros do modelo de espaço de estados, o qual é aplicado em observações que possuem distribuições não Gaussianas. Visando a modelagem da dinâmica das IVs das opções de ativos, Bedendo e Hodges (2009) utilizaram pela primeira vez o filtro de Kalman para uma melhor estimação simultânea dos parâmetros do modelo que atualiza a curva de volatilidade.

Outros modelos estocásticos surgiram mais recentemente a partir do trabalho de Gatheral, Jaisson e Rosenbaum (2014). Os autores mostraram que a log-volatilidade se comporta como um movimento Browniano fractal com expoente de Hurst H da ordem de 0,1 para qualquer escala de tempo razoável, enquanto que os modelos baseados no movimento Browniano não fractal possuem expoente H igual a 0,5. Dando continuidade a este trabalho, Bayer, Friz e Gatheral (2016) mostraram que o modelo proposto por Gatheral, Jaisson e Rosenbaum (2014) escrito em uma medida física P leva naturalmente a um modelo de precificação de opções sob uma medida Q que é consistente com a formato da superfície de IV observada. Os autores aplicam o modelo para um caso específico onde é assumida uma mudança determinística de medida entre P e Q , sendo assim considerado uma extensão não Markoviana do modelo de Bergomi, que passou a ser chamado de *rough* Bergomi, ou rBergomi. Mesmo com uma limitação de eficiência computacional, devido à complexidade do modelo, os autores mostram que o rBergomi consegue aproximar seus resultados à superfície de volatilidade do SPX com mais precisão. Adicionalmente os autores observaram independência entre o modelo rBergomi e a estrutura a termo da volatilidade do VIX.

Com o intuito de apresentar uma solução à limitação computacional que modelos de volatilidade estocástica enfrentam, Horvath, Muguruza e Tomas (2019) introduziram um método de calibração para aproximar as funções complexas do modelo eficientemente baseado em redes neurais profundas. Os autores exploram algumas configurações do método de redes neurais de modo a se obter uma aproximação das funções complexas do modelo estocástico em modo *off-line*, o que é difícil de se representar ou demanda um consumo

elevado de tempo quando comparado com outros métodos.

2.3 Predição da Volatilidade a Partir dos Preços de Opções e das IVs

Existem diversos trabalhos que abordam a predicabilidade da RVol a partir dos preços de opções e da IV. Jeff (1998) apresenta resultados que indicam que a IV é um preditor viesado de RVol, mas que também contém informações relevantes a respeito da volatilidade futura. Busch, Christensen e Nielsen (2007) afirmam que as IVs contém informações incrementais sobre a volatilidade futura e que fornecem previsões não viesadas para mercados de câmbio e ações. Andersen, Frederiksen e Staal (2007) sugerem que previsões de RVols baseadas em RVols passadas são não viesadas, são mais eficientes que o modelo de IV do VIX e podem ser tão eficientes quanto o modelo de IV do Black & Scholes (BS). Bandi, Russell e Yang (2008) mostram que os modelos de previsão de RVol desempenham com maior acurácia quando são abastecidos com as IV.

2.4 Modelos que utilizam Redes Neurais Artificiais

O ML demonstrou sucesso e estabeleceu uma forte presença em muitas aplicações importantes de finanças nos últimos tempos (PRADO, 2019).

Hutchinson, Lo e Poggio (1994) apresentam um dos primeiros trabalhos de apreçamento de derivativos baseados em redes neurais artificiais. Embora o uso de funções paramétricas são preferíveis para o apreçamento de derivativos, os autores demonstram que redes de aprendizado de máquina não paramétricas são uma boa alternativa para os casos onde os primeiros métodos citados falham. No que se refere ao futuro uso desse modelos, eles pontuam a necessidade da especificação de diversos dados de entrada, principalmente os que não são capturados por modelos paramétricos, como os retornos dos ativos, volatilidade do mercado e outras medidas de condições de mercado. Para que isso seja possível, os autores afirmam que a modelagem dos processos de volatilidade dos retornos do ativo base serão necessárias.

Hamid (2004) apresenta o uso de RNAs para prever variáveis gerais de mercado. Ele estuda o caso de previsão da volatilidade do Índice S&P500 usando RNAs e chega à conclusão de que esta abordagem possui um melhor desempenho preditivo das RVols quando comparada com as previsões das IV. Ele afirma, também, que previsões a partir de RNAs não são significantemente diferentes das RVols, enquanto as previsões a partir das IVs são significantemente diferentes das RVols em dois a cada três casos.

A combinação de modelos é uma abordagem muito utilizada na literatura. Donald-

son e Kamstra (1997) propõem a combinação de uma RNA com um modelo GARCH para capturar o efeito da volatilidade nos retornos dos preços de ações do mercado de Nova Iorque, Tóquio e Toronto. Eles demonstram que, tanto nos dados da amostra de treinamento quanto nos dados fora dela, a rede neural captura efeitos de volatilidade que não são capturados pelos modelos GARCH, EGARCH e GJR. Esta linha de abordagem é embasada por Hastie, Tibshirani e Friedman (2009) quando apresentam em seu livro uma metodologia de empilhamento de algoritmos, isto é, a combinação de dois modelos de famílias distintas. Neste contexto, Kristjanpoller e Minutolo (2015) apontam que o GARCH é uma das famílias de modelos mais utilizadas para prever a volatilidade do preço de um ativo, embora apresentem erros de previsão consideravelmente altos, e assim, propõem uma alternativa para a abordagem deste tipo de previsão, a qual é baseada em um modelo híbrido de RNA e GARCH. Neste trabalho, os autores alimentam a RNA com os valores previstos pelo modelo GARCH e com outras variáveis adicionais de forma que se tenha uma melhor previsão da volatilidade dos preços de ouro. Um dos pontos positivos desta abordagem é a habilidade que a RNA tem de aprender com os erros de previsão gerados pelo modelo GARCH. Os autores concluem que a RNA é capaz de melhorar as previsões de um modelo GARCH.

Uma continuação do trabalho de Kristjanpoller e Minutolo (2015) é apresentada por Kristjanpoller e Minutolo (2018), onde o modelo RNA-GARCH é complementado com um pré processamento para prever a volatilidade dos preços do Bitcoin. Os autores aplicam a Análise de Componente Principal (PCA) em alguns indicadores técnicos, como média móvel, momento, índice de força relativa, entre outros. Esses indicadores e suas transformações, obtidas a partir do PCA, se tornam entradas da RNA, complementando os valores previstos pelo modelo GARCH. Essa abordagem apresenta uma melhora na acurácia do modelo.

RNAs também podem ser combinadas com modelos estocásticos. O trabalho de Horvath, Muguruza e Tomas (2019) apresenta uma estrutura de calibração de modelos de volatilidade estocástica para o apreçamento de derivativos usando redes neurais profundas (*deep learning*). O grande diferencial deste trabalho está na aplicação desta calibração para modelos que se baseiam em volatilidades estocásticas que seguem um movimento browniano fractal, onde o expoente de Hurst é menor que 0,5. Estes modelos refletem funções complexas que são muito difíceis de se representar ou que consomem muito tempo de processamento, inviabilizando o uso deles no dia a dia da indústria. Os autores solucionam este problema ao usar uma rede neural profunda para gerar aproximações destas funções de apreçamento complexas, através da calibração dos seus parâmetros. A ideia é que este processo de calibração seja feito *offline* e o seu resultado seja aplicado *online*, de forma que se tenha superfícies de volatilidade em poucos milissegundos.

Seguindo a linha de utilização de RNAs para a previsão de informações do mercado

financeiro, Piccoli (2014) propõe a construção de uma rede para a previsão do movimento dos contratos de dólar futuro. O autor utiliza diversas variáveis quantitativas como informações de entrada do modelo e varia o conjunto de parâmetros das RNAs previamente determinados de forma que todas as redes sejam testadas e então seleciona aquela com o melhor desempenho. A partir da rede escolhida, o autor especifica estratégias de *trading* e compara os seus resultados financeiros com o *benchmark* do mercado, CDI². O autor conclui que a utilização das RNAs promove resultados superiores ao *benchmark* e permite o desenvolvimento de uma ferramenta que pode auxiliar empresas que necessitam fazer *hedge* de ativos e passivos, bem como *players* do mercado que necessitam rentabilizar carteiras. Já no que diz respeito à modelagem e previsão da volatilidade realizada de retornos de ativos Brasileiros, Carvalho et al. (2006) concluem que suposições de normalidade usualmente utilizadas em gerenciamento de risco para descrever distribuições de retornos diários padronizados pelas volatilidades não são compatíveis com as volatilidades estimadas nos modelos EWMA ou GARCH. Eles afirmam que ao se utilizar dados com frequência alta, ou seja, dados intra diários, encontra-se normalidade nos retornos padronizados e que as distribuições das volatilidades de ações Brasileiras se aproximam de uma log normal. Por fim, os autores estimam um modelo para o log das volatilidades realizadas que difere dos modelos de outros estudos e não encontram evidências de memória longa, o que torna um modelo simples de memória curta suficiente para modelar e prever séries diárias de volatilidade.

O trabalho, aqui apresentado, propõe uma nova técnica para a estimação da variância realizada de um ativo a partir dos preços diários de fechamento de suas opções. Esta técnica é baseada em redes neurais artificiais e possui duas abordagens distintas, a primeira é o treinamento das redes para obter a RV diretamente e a segunda é para obter a RV a partir do VRP. Nesta última abordagem, a RV é obtida através da combinação do VRP com o VIX. Os desempenhos das RNAs são comparados com os desempenhos do VIX e de dois modelos de regressão linear múltipla, sendo o primeiro uma regressão dos preços das opções em relação à RV e o segundo em relação ao VRP. Os desempenhos dos modelos são medidos pelo R^2 ajustado. As RNAs treinadas para estimar a RV a partir do VRP apresentam o melhor resultado geral entre modelos testados, sugerindo que modelos de RNAs podem ser mais eficientes que aqueles utilizados atualmente no mercado. Tais resultados também sugerem que a utilização do VIX, combinado com os preços de opções, exerce um papel de normalizador para a RNA, sendo capaz de melhorar seu desempenho.

² O CDI, sigla para Certificado de Depósito Interbancário, é um título emitido pelos bancos para a realização de operações de empréstimos entre si de curtíssimo prazo.

3 FUNDAMENTAÇÃO TEÓRICA

Este capítulo descreve as medidas de volatilidades implícita e variância realizada abordadas neste trabalho, bem como as definições do VIX e VRP. Além de detalhar de forma sucinta a dinâmica de uma RNA, o algoritmo *Backpropagation* e o otimizador L-BFGS, usados neste trabalho.

3.1 Volatilidade Implícita

A volatilidade permeia o mundo das finanças. Ela é considerada o grau de variação do preço de um ativo em uma série temporal. A variação dos fatores de risco econômicos são fundamentais para o apreamento de ativos financeiros. Neste contexto a volatilidade é um fator chave para problemas de apreamento de opções e alocações de portfólio (ANDERSEN; BOLLERSLEV, 1998).

A volatilidade implícita é uma métrica utilizada na estimação da volatilidade futura, representando a expectativa do mercado em relação ao futuro movimento do preço de um ativo. Nas subseções a seguir são apresentados exemplos de como ela pode ser obtida.

3.1.1 Volatilidade Implícita a Partir de Modelos de Apreçamento de Opções

Dado que IV não é um fator diretamente observável no mercado, ela pode ser determinada a partir de modelos de apreamento de opções. Certos fatores servem de parâmetros de entrada para os modelos de forma que se obtenham a IV quando se tem o prêmio da opção e vice-versa. O modelo Black & Scholes, apresentado inicialmente por Black e Scholes (1973) e estendido por Merton (1973), em virtude da sua simplicidade e tratabilidade matemática, é largamente aceito e utilizado no mercado para o apreamento de opções, sendo assim, é uma ferramenta bastante utilizada na determinação das IVs de um ativo.

De acordo com os autores as hipóteses do modelo são:

1. O preço de um ativo segue um movimento *Browniano* geométrico com tendência μ e volatilidade σ .
2. Vendas a descoberto é possível.
3. Não há custos de transações ou taxas, e os ativos são perfeitamente divisíveis.
4. Não há pagamentos de dividendo durante a vida do derivativo.

5. Não existe oportunidade de arbitragem.
6. As transações ocorrem de forma contínua.
7. A taxa de juros livre de risco r é constante e a mesma para todos os vencimentos.

Conforme descrito acima, o modelo parte da premissa de que o preço do ativo S segue um movimento *Browniano* geométrico W ,

$$dS = \mu S dt + \sigma S dW, \quad (3.1)$$

e $f(S, t)$ é o preço de um derivativo em função do preço do ativo subjacente e do tempo, em que ao se aplicar o lema de Itô¹, tem-se

$$df = \left(\frac{\partial f}{\partial S} \mu S + \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} \sigma^2 S^2 \right) dt + \frac{\partial f}{\partial S} \sigma S dW. \quad (3.2)$$

A partir das Equações 3.1 e 3.2, e sabendo que a variação do preço do ativo e do derivativo são perfeitamente correlacionadas em um período de tempo pequeno, Merton (1973) aplica a metodologia do portfólio replicante² para obter a equação diferencial de Black-Scholes-Merton

$$\frac{\partial f}{\partial t} + rS \frac{\partial f}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} = rf. \quad (3.3)$$

A Equação 3.3 corresponde aos diferentes derivativos baseados no ativo subjacente e ao se resolvê-la³ chega-se aos seguintes preços de opções de compra e venda:

$$c = S_0 N(d_1) - K e^{-rT} N(d_2) \quad (3.4)$$

$$p = K e^{-rT} N(-d_2) - S_0 N(-d_1), \quad (3.5)$$

onde c e p são os preços de opções de compra e venda de uma opção europeia, respectivamente, S_0 é o preço da ação no tempo zero, K é o preço de exercício da opção, r é a taxa de juros livre de risco anualizada, com capitalização contínua, σ é a volatilidade do preço da ação, T é o tempo para o vencimento da opção,

$$d_1 = \frac{\ln S_0/K + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2 = \frac{\ln S_0/K + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

e $N(x)$ é a função de distribuição acumulada da normal padrão.

Como a volatilidade σ da série de preços utilizada na fórmula não é diretamente observada, a sua estimação *a priori* é necessária. Os agentes do mercado costumam retirar

¹ Mais detalhes sobre o lema de Itô podem ser obtidos em Itô (1944)

² Para mais informações sobre a metodologia do portfólio replicante consultar Hull (2005)

³ A demonstração da solução da equação diferencial de Black-Scholes-Merton pode ser consultada em Hull (2005)

implicitamente esta volatilidade através dos preços observados das opções. Uma vez que não é possível isolar σ nas equações acima, utiliza-se processos iterativos para se buscar σ que satisfaça a equação com os dados lidos do mercado (S_0 , K , r , T e c ou p , dependendo da opção)(HULL, 2005).

3.1.2 Volatilidade Implícita a Partir do VIX

O índice de volatilidade VIX, *Volatility Index*, foi criado em 1993 e é calculado pela *Chicago Board Options Exchange* (CBOE) através das opções do índice americano S&P500. O VIX, conhecido como índice do medo, foi originalmente criado com base nas opções do S&P500 refletindo uma expectativa livre de risco da volatilidade para os próximos 30 dias, e hoje já existem outros índices de volatilidade tanto para outros ativos quanto para outras janelas de volatilidade. O VIX é calculado com base na seguinte fórmula:

$$\sigma^2 = \frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} e^{RT} Q(K_i) - \frac{1}{T} \left[\frac{F}{K_0} - 1 \right]^2 \quad (3.6)$$

onde:

$$\sigma : \frac{VIX}{100} \rightarrow VIX = 100 \sigma$$

T : Tempo de maturidade

F : Valor do índice futuro do S&P500

K_0 : Primeiro preço de exercício abaixo de F

K_i : Preço de exercício da i -ésima opção fora do dinheiro, sendo uma opção de compra se $K_i > K_0$, uma opção de venda se $K_0 > K_i$ ou ambas se $K_0 = K_i$

ΔK_i : Intervalo entre preços de exercício $\Delta K_i = \frac{K_{i+1} - K_i}{2}$

R : Taxa de juros livre de risco até a maturidade

$Q(K_i)$: Ponto médio entre o preço de compra e o preço de venda (*spread bid-ask*)

Diferentemente da metodologia do modelo Black & Scholes que considera apenas uma série de preços de opções, o VIX usa uma soma ponderada desses preços. Além desta diferença, estes dois modelos mencionados possuem outras. Um problema comum no cálculo da volatilidade implícita de um ativo é o *smile* de volatilidade, gerado pelo fato de que opções com o mesmo vencimento podem apresentar volatilidades diferentes conforme se muda o preço de exercício. Em tese, a volatilidade deveria ser igual para todos os preços de exercício, porém, observa-se que opções mais distantes do dinheiro possuem volatilidades maiores que as opções dentro ou próximas do dinheiro. Este problema não é observado no VIX uma vez que utiliza-se tanto opções no dinheiro quanto opções fora do dinheiro, e com ponderações que tornam o índice mais próximo das opções no dinheiro.

3.2 Variância Realizada (RV)

Se por um lado as volatilidades implícitas são expectativas futuras, por outro as volatilidades realizadas são observações históricas.

Na literatura há diversos trabalhos que apresentam diferentes abordagens para a estimativa da volatilidade realizada, uma vez que não exista uma definição concreta sobre o seu valor exato. Diferentes frequências de amostras intra diárias do preço do ativo podem ser usadas para se chegar neste valor estimado de volatilidade. Porém, esta abordagem de se utilizar amostras intra diárias não será utilizada neste trabalho por conta das características dos dados das opções a serem utilizados, os quais possuem frequência diária.

Assim, serão utilizados os preços de fechamento do índice EWZ para computar a sua variância realizada (RV). Portanto a RV dos próximos T dias a partir da data base⁴ t será calculada com a frequência de um dia:

$$RV_t = \frac{1}{T-1} \sum_{i=1}^T (r_{t+i} - \hat{r}_t)^2, \quad (3.7)$$

onde \hat{r}_t é o retorno diário médio dos próximos T dias a contar da data base t e,

$$r_j = \frac{p_j - p_{j-1}}{p_{j-1}} \quad (3.8)$$

é o retorno diário do índice EWZ e p_j é o preço de fechamento do índice EWZ j dias a frente da data base t .

3.3 Prêmio de Risco de Variância (VRP)

Eraker (2008) afirma que, na média, opções de índices possuem valores de volatilidade implícita que excedem os valores do desvio padrão não condicional anualizado dos seus respectivos ativos subjacentes. O autor faz uma comparação entre a expectativa de volatilidade anualizada do mercado, refletida pelo índice VIX, e o desvio padrão não condicional anualizado dos retornos do S&P500 entre os anos de 1990 e 2007. A primeira apresenta o valor médio de 19% e o segundo de 15,7%.

Neste contexto de divergência entre a volatilidade esperada pelo mercado e a volatilidade realizada do ativo temos o prêmio de risco de volatilidade. Analogamente, o VRP é definido como a diferença entre a variância esperada sob a medida neutra a risco e a variância esperada sob a medida física do mundo real observável, sendo esta segunda equivalente à variância dos retornos realizados do ativo. Como prática de mercado, a variância neutra a risco é medida por VIX^2 .

⁴ Define-se o termo data base como o dia de referência que está sendo analisado em dada etapa do processo.

De acordo com Bai, Zhang e Zhou (2019), em um determinado instante de tempo t denota-se o preço de um ativo $S(t)$, uma taxa de juros constante r , e $S(t) > 0$ com probabilidade 1 para todo t . Define-se, então, $g(r(t, \tau))$ como o *payoff* de um contrato *swap* de variância⁵ em T de uma função do retorno sobre o período $[t, T]$ e duplamente diferenciável, onde $r(t, \tau)$ é o τ -período de retorno dado por $r(t, \tau) \equiv \ln[S(t + \tau)] - \ln[S(t)]$. Sendo Q e P as medidas de probabilidade neutra a risco e observável, respectivamente, tem-se a seguinte equação para definir o prêmio de risco de variância:

$$VRP_t = E_t^Q[g_v(r)] - E_t^P[g_v(r)], \quad (3.9)$$

onde

$$g_v(r) := 2(e^r - 1 - r).$$

O valor esperado de $g_v(r)$ na medida neutra a risco Q , $E_t^Q[g_v(r)]$, pode ser replicado por uma sequência de opções, como feito por Bakshi e Madan (2000):

$$E_t^Q\{e^{-rt}g_v(r)\} = \int_{S_t}^{\infty} \frac{2(1 - \ln[\frac{K}{S(t)}])}{K^2} C(t, \tau, K) dK + \int_0^{S_t} \frac{2(1 + \ln[\frac{S(t)}{K}])}{K^2} C(t, \tau, K) dK. \quad (3.10)$$

Para estimar a integral na Equação 3.10, pode-se usar opções de compra e venda fora do dinheiro (OTM). Bams, Blanchard e Lehnert (2017) usaram uma aproximação trapezoidal para estimar estas integrais usando dados discretos. Este método pode gerar duas fontes de vieses, a primeira é o intervalo de discretização dos preços de exercício e a segunda é a assimetria do domínio de integração. O primeiro viés pode ser considerado relativamente o mesmo para a série de derivativos, uma vez que geralmente a maioria dos ativos possuem as mesmas diferenças entre os preços de exercício. Já o segundo viés pode ser corrigido através do uso do mesmo número $n(\geq 2)$ de opções de compra e venda OTM, para cada ativo, dia e vencimento. Dessa forma é obtém-se variâncias para diferentes vencimentos, sendo especificamente uma maior que 30 dias e outra menor que 30 dias, e então as interpola-se linearmente. O resultado deste processo é o mesmo obtido pela construção do VIX. Vale ressaltar a presença dos termos de ponderação pelo preço de exercício $\frac{2-2\ln[\frac{K}{S(t)}]}{K^2}$ e $\frac{2+2\ln[\frac{S(t)}{K}]}{K^2}$ na Equação 3.10, assim como o termo $\sum_i \frac{2\Delta K_i}{K_i^2}$ na Equação 3.6 da construção do VIX.

O valor esperado de $g_v(r)$ na medida de probabilidade física P , $E_t^P\{g_v(r)\}$, é estimado como a variância total realizada de 30 dias corridos, ou aproximadamente 22 dias de negociação, anualizada ($RV_{t+1}^{(22)}$). Confirma-se, assim, o prêmio de risco de variância como a diferença entre as expectativas de variâncias livre de risco e física. A equação do VRP pode ser escrita como:

$$VRP_t = VIX^2 - E_t[RV_{t+1}^{(22)}]. \quad (3.11)$$

⁵ Para mais detalhes sobre a função $g(r(t, \tau))$, consultar Demeterfi et al. (1999).

Especula-se que o prêmio de volatilidade é causado pelos investidores que possuem aversão a retornos negativos e alta volatilidade dos ativos, e que estão, portanto, dispostos a pagar um prêmio por um seguro de portfólio composto por opções de venda. Este seguro também é representado por um contrato *swap* (*swap* de variância), através do qual o VRP é negociado no mercado *over-the-counter* (OTC). No vencimento, o agente comprador deste contrato recebe $E_t[RV_{t+1}^{(22)}] - VIX^2$.

3.4 Redes Neurais Artificiais

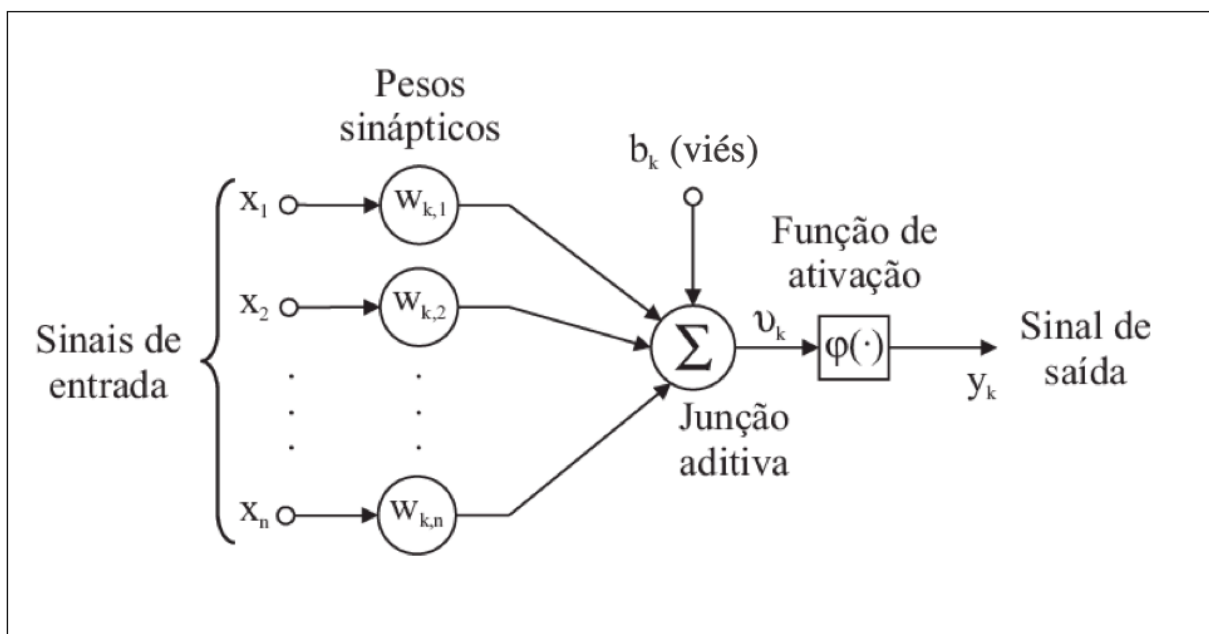
Uma rede neural artificial (RNA) é um modelo de inferência que busca representar o processo de raciocínio do cérebro humano durante a execução de alguma atividade (HAYKIN, 2007).

Os neurônios são as unidades básicas das RNAs. Eles se agrupam em uma ou mais camadas e são conectados entre si, desempenhando o papel de calcular determinadas funções matemáticas, que no geral são não lineares. Estas conexões entre os neurônios estão associadas a diferentes pesos que armazenam o conhecimento experimental e ponderam as entradas de cada unidade (BRAGA; LUDERMIR; CARVALHO, 2000).

Haykin (2007) afirma que uma rede neural é um processador paralelamente distribuído e composto por unidades de processamento simples, as quais possuem uma propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso.

A Figura 1 apresenta um modelo não linear de neurônio (HAYKIN, 2007).

Figura 1 – Modelo de um Neurônio Não Linear



Fonte: Haykin (2007)

Três componentes fundamentais de um neurônio estão representados na Figura 1:

1. As sinapses de entrada, ou ligações conectoras, onde cada uma é caracterizada pelo seu peso. Mais especificamente, um sinal x_j em uma conexão de entrada j conectada a um neurônio k é multiplicada pelo peso sináptico $w_{k,j}$.
2. Uma junção aditiva para os sinais de entrada, já considerando seus respectivos pesos, e um viés externo b_k (denominado limiar de ativação). As operações descritas aqui constituem uma combinação linear.
3. Uma função de ativação $\varphi(\cdot)$ para limitar a amplitude da saída y_k do neurônio onde

$$y_k = \varphi(v_k) = \varphi\left(\sum_{j=1}^m w_{k,j}x_j + b_k\right) \quad (3.12)$$

A função de ativação $\varphi(v)$ define a saída do neurônio em termos do campo local induzido v . Haykin (2007) identifica dois tipos básicos de funções de ativação:

1. Função de Limiar:

$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases} \quad (3.13)$$

$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v = 0 \\ -1 & \text{se } v < 0 \end{cases} \quad (3.14)$$

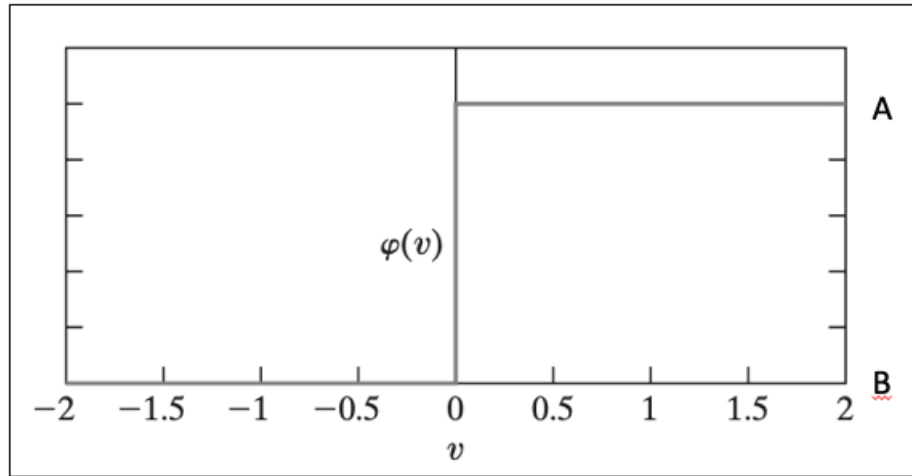
2. Função de Sigmóide:

$$\varphi(v) = \frac{1}{1 + \exp(-av)} \quad (3.15)$$

$$\varphi(v) = \tanh(v) \quad (3.16)$$

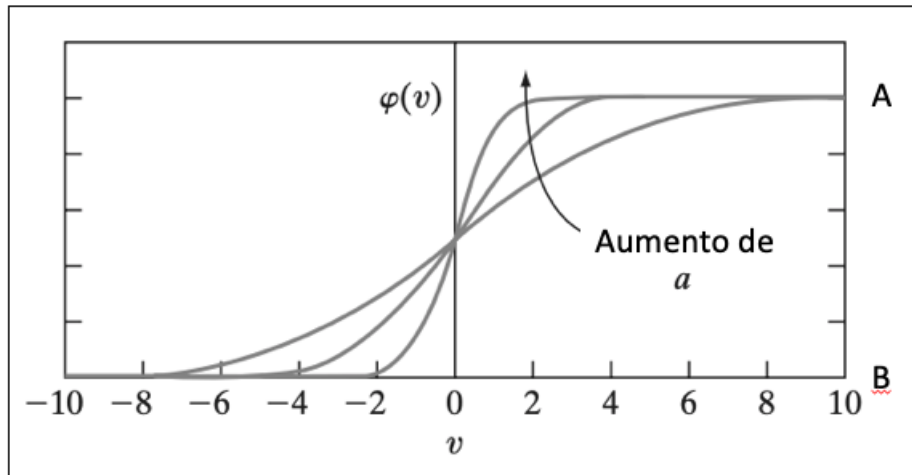
As Figuras 2 e 3 representam, respectivamente, as funções de limiar e de sigmóide. As letras A e B destas imagens assumem, respectivamente, os valores 1 e 0 para as Equações (3.13) e (3.15), e 1 e -1 para as Equações (3.14) e (3.16).

Figura 2 – Gráfico da Função de Limiar



Fonte: Elaborada pelo autor.

Figura 3 – Gráfico da Função de Sigmóide



Fonte: Elaborada pelo autor.

Neste trabalho serão utilizadas duas funções de ativação:

1. Função Unidade Linear Retificada (ReLU) na camada escondida: Função linear que retorna o mesmo valor de entrada, caso este seja maior ou igual a zero, ou zero, caso este seja negativo.

$$\varphi(v) = \begin{cases} v & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases} \quad (3.17)$$

2. Função Identidade na camada de saída: Função linear que retorna exatamente o mesmo valor de entrada.

$$\varphi(v) = v \quad (3.18)$$

Como mostrado na Figura 4 (HAYKIN, 2007), existem algumas maneiras de combinar os neurônios para se formar arquiteturas de RNAs. Cada uma delas contém os seguintes parâmetros:

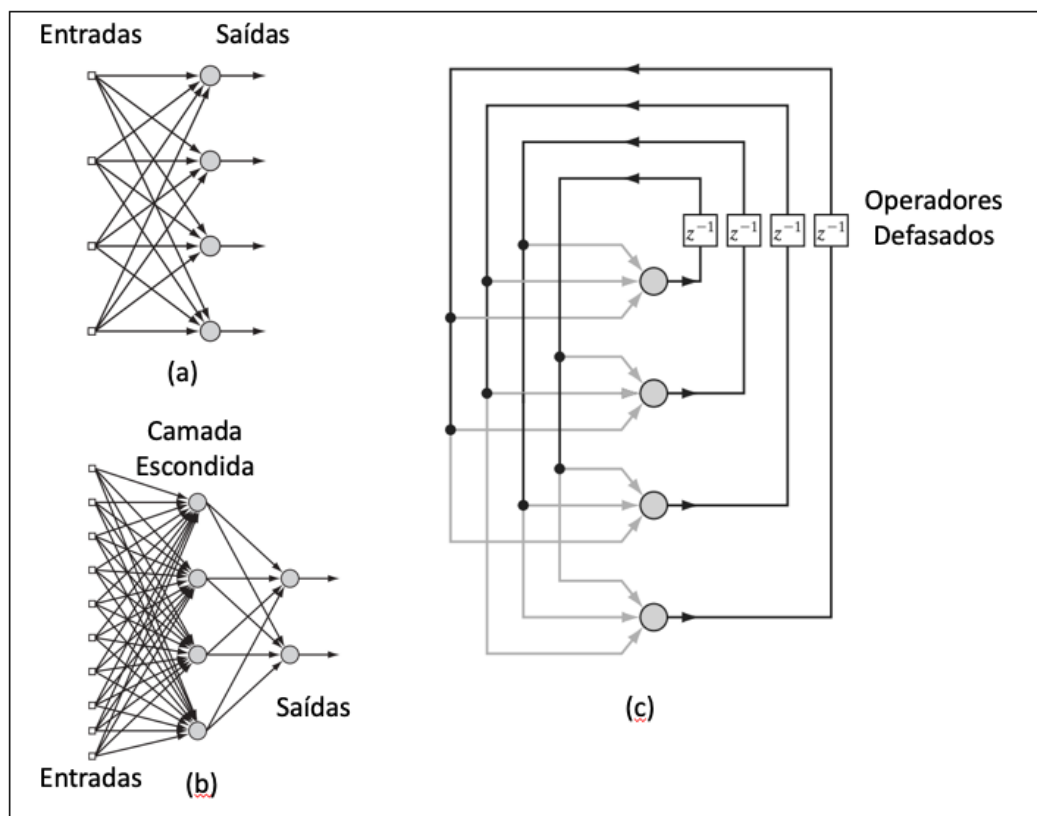
- Quantidade de camadas.
- Quantidade de neurônios em cada camada.
- Tipo e peso das conexões entre os neurônios.
- Topologia da RNA.

De acordo com Haykin (2007), em geral é possível identificar três classes fundamentais de arquiteturas de RNAs, representadas na Figura 4:

1. Redes acíclicas de uma única camada como exemplificado na estrutura (a).
2. Redes acíclicas com múltiplas camadas conforme a estrutura (b).
3. Redes recorrentes ilustradas na estrutura (c).

Neste trabalho serão utilizadas redes acíclicas com múltiplas camadas.

Figura 4 – Exemplos de Arquiteturas de RNAs



Fonte:

Adaptado de Haykin (2007).

O processo de aprendizagem de uma máquina corresponde a um algoritmo de treinamento que seleciona as informações relevantes que foram apresentados à RNA e as armazena nos pesos sinápticos de suas conexões. Este processo realiza ajustes nos pesos das conexões de forma que se tenha, ao final da iteração, todo o conhecimento armazenado neles. Dessa forma, podemos afirmar que os pesos das conexões representam a informação do ambiente em que a RNA está atuando.

Haykin (2007) faz uma analogia do processo de aprendizagem da máquina com o dos humanos, separando ele em dois tipos, o primeiro com professor e o segundo sem professor. O primeiro é também chamado de aprendizado supervisionado. Já o segundo pode ser separado em dois subgrupos, sendo eles o aprendizado por reforço e o aprendizado não supervisionado. Nesta analogia do professor, o que difere um grupo do outro é a presença, no aprendizado supervisionado, de um agente externo que é capaz de apontar quais são as respostas ótimas para um determinado conjunto de informações de entrada.

Como evidenciado no item anterior, as redes acíclicas podem ter uma ou mais camadas. Cybenko (1989) demonstrou pela primeira vez que uma rede acíclica com uma única camada escondida, isto é, duas camadas no total, é suficiente para aproximar uniformemente qualquer função matemática, sendo ela contínua ou não. O Multilayer Perceptron (MLP) é um caso particular de redes acíclicas com múltiplas camadas.

Os MLPs possuem três características básicas (HAYKIN, 2007):

- O modelo de cada neurônio da rede inclui uma função de ativação não linear que é diferenciável.
- A rede contém uma ou mais camadas que são escondidas de ambos os nós de entrada e saída.
- A rede possui um alto nível de conectividade, cuja extensão é determinada pelos pesos sinápticos da rede.

Haykin (2007) afirma que ao se combinar essas características com um processo de aprendizagem de máquinas apropriado, tem-se uma ferramenta com uma significativa eficiência computacional, a qual pode ser utilizada para a resolução de diversos problemas. Um método popular para o treinamento do MLP é o *Backpropagation* (BP).

3.4.1 O Algoritmo *Backpropagation*

Em aprendizado de máquinas, o *Backpropagation* é um algoritmo amplamente utilizado para o treinamento de RNAs acíclicas com aprendizado supervisionado, motivo pelo qual este algoritmo foi escolhido neste trabalho. O BP corresponde a um algoritmo

de aprendizado supervisionado por correção de erros, cujo treinamento pode ser separado em duas fases (HAYKIN, 2007):

1. Fase de propagação: os pesos sinápticos são fixados e os sinais de entrada são propagados através da rede, de camada em camada, até que a saída seja atingida.
2. Fase de retro-propagação: um sinal de erro é produzido através da comparação entre os valores de saída e os valores desejados. Este sinal de erro resultante é propagado através da rede, novamente de camada em camada, mas desta vez o sentido é da saída para a camada de entrada. Esta fase também é conhecida como fase de atualização de pesos, uma vez que o gradiente da função de perda na camada final é utilizado para aplicar recursivamente a regra da cadeia para atualizar os pesos da rede.

Para calcular o gradiente da função de perda na camada final, é necessário primeiramente conhecer essa função. Mas antes disso, denota-se

$$\tau = \{x(n), d(n)\}_{n=1}^N \quad (3.19)$$

como a amostra de treinamento usada para treinar a rede de modo supervisionado, onde $x(n)$ representa os vetores de estímulos na camada de entrada e $d(n)$ representa os vetores resposta desejados como saída da rede. Assim, na camada de saída, como ilustrado na Figura 5 (HAYKIN, 2007), observamos que o erro $e_j(n)$ do neurônio j , após a n -ésima iteração, pode ser definido por:

$$e_j(n) = d_j(n) - y_j(n), \quad (3.20)$$

onde $d_j(n)$ representa o j -ésimo elemento de $d(n)$ e $y_j(n)$ representa o sinal de saída gerado pela RNA.

Define-se a energia instantânea do erro do neurônio j como uma função quadrática da forma

$$\epsilon_j(n) = \frac{1}{2} e_j^2(n). \quad (3.21)$$

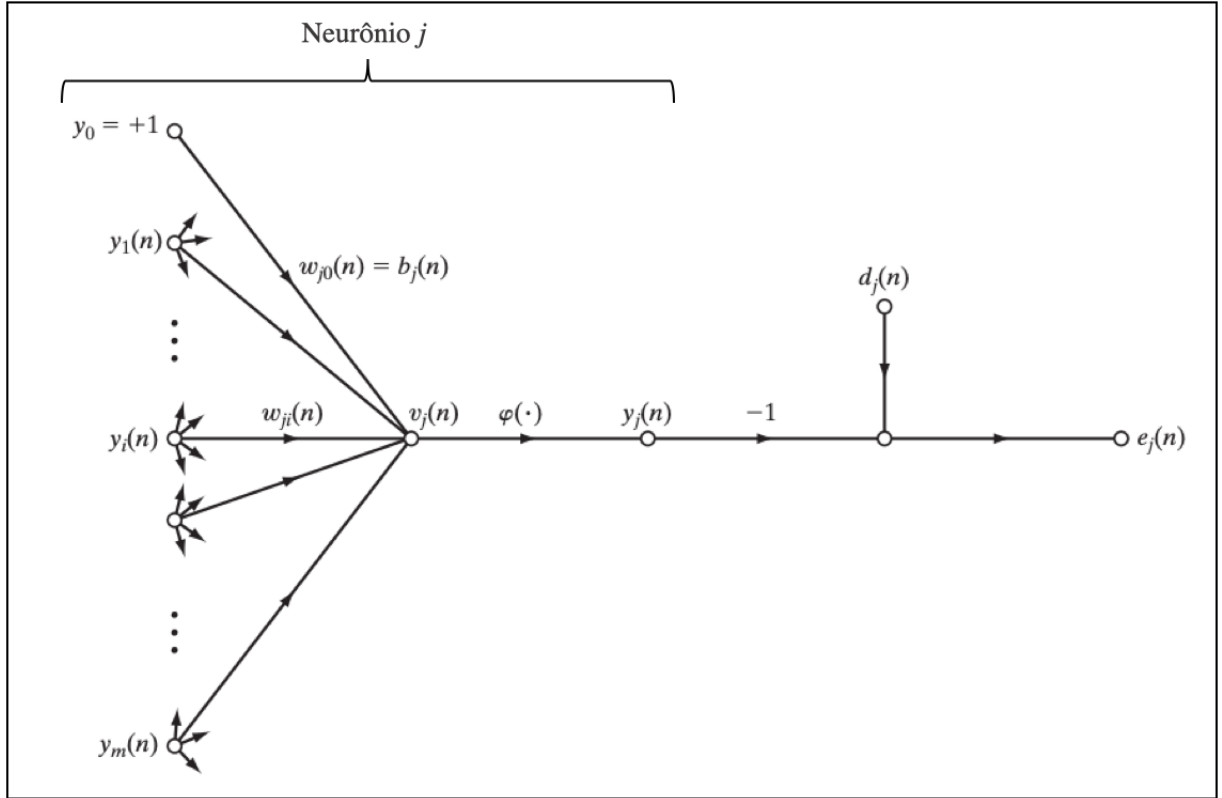
Somando-se a energia instantânea de todos os neurônios na camada de saída, pode-se expressar a energia instantânea de erro total de toda a rede como

$$\epsilon(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n), \quad (3.22)$$

onde C é o conjunto de todos os neurônios da camada de saída. Considerando a amostra de treinamento com N casos de exemplo, a energia instantânea de erro total média da amostra de treinamento, conhecida como o risco empírico, é definida por

$$\epsilon(\hat{N}) = \frac{1}{N} \sum_{n=1}^N \epsilon(n). \quad (3.23)$$

Figura 5 – Fluxo de informações de um neurônio na camada de saída



Fonte: Adaptado de Haykin (2007).

De acordo com a Figura 5, observamos que o neurônio j é alimentado por um conjunto de sinais de funções produzidos pela camada de neurônios anterior. Dessa forma, gera-se um campo local induzido $v_j(n)$:

$$v_j(n) = \sum_{i=0}^m w_{ji}(n)y_i(n), \quad (3.24)$$

onde m corresponde à quantidade de entradas provenientes da camada anterior. Este campo local induzido será a informação de entrada para função de ativação $\varphi(\cdot)$ do neurônio j , a qual irá gerar uma saída $y_j(n)$:

$$y_j(n) = \varphi_j(v_j(n)) \quad (3.25)$$

$$y_j(n) = \varphi_j \left(\sum_{i=0}^m w_{ji}(n)y_i(n) \right) \quad (3.26)$$

O algoritmo *Backpropagation* aplica uma correção $\Delta w_{ji}(n)$ no peso sináptico $w_{ji}(n)$. Esta correção é proporcional à variação de $\epsilon(n)$, uma vez que ao se definir $y_j(n)$, verifica-se que $e(n)$ e, conseqüentemente, $\epsilon(n)$, são funções apenas dos pesos sinápticos do neurônio j . Portanto, a variação de $\epsilon(n)$ é proporcional às suas derivadas parciais em função dos pesos sinápticos que correspondem ao gradiente de $\epsilon(n)$. De acordo com a regra da cadeia,

podemos expressar o gradiente de $\epsilon(n)$ como

$$\frac{\partial \epsilon(n)}{\partial w_{ji}(n)} = \frac{\partial \epsilon(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)}, \quad (3.27)$$

onde,

$$\frac{\partial \epsilon(n)}{\partial e_j(n)} = \frac{\partial \left(\frac{1}{2} \sum_{j \in C} e_j^2(n) \right)}{\partial e_j(n)} = e_j(n), \quad (3.28)$$

$$\frac{\partial e_j(n)}{\partial y_j(n)} = \frac{\partial (d_j(n) - y_j(n))}{\partial y_j(n)} = -1, \quad (3.29)$$

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \frac{\partial \varphi_j(v_j(n))}{\partial v_j(n)} = \varphi'_j(v_j(n)), \quad (3.30)$$

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = \frac{\partial \sum_{i=0}^m w_{ji}(n) y_i(n)}{\partial w_{ji}(n)} = y_i(n), \quad (3.31)$$

e, portanto,

$$\frac{\partial \epsilon(n)}{\partial w_{ji}(n)} = -e_j(n) \varphi'_j(v_j(n)) y_i(n). \quad (3.32)$$

De acordo com Haykin (2007), a derivada parcial da Equação 3.32 representa um fator de sensibilidade que determina a direção na qual $\epsilon(n)$ cresce. Dessa forma, o ajuste $\Delta w_{ji}(n)$ deve ser realizado na direção oposta do gradiente para que seja possível localizar o ponto de mínimo global da superfície de erro (gradiente descendente).

Segundo Haykin (2007), a correção $\Delta w_{ji}(n)$ aplicada em $w_{ji}(n)$ pode ser definida através da regra Widrow-Hoff tal que

$$\begin{aligned} \Delta w_{ji}(n) &= -\eta \frac{\partial \epsilon(n)}{\partial w_{ji}(n)} \\ \Delta w_{ji}(n) &= -\eta \left[-e_j(n) \varphi'_j(v_j(n)) y_i(n) \right] \\ \Delta w_{ji}(n) &= \eta \delta_j(n) y_i(n) \end{aligned} \quad (3.33)$$

onde η corresponde à taxa de aprendizagem do algoritmo *Backpropagation* e $\delta_j(n)$ corresponde ao gradiente local do neurônio j

$$\begin{aligned} \delta_j(n) &= \frac{\partial \epsilon(n)}{\partial v_j(n)} \\ \delta_j(n) &= \frac{\partial \epsilon(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \\ \delta_j(n) &= e_j(n) \varphi'_j(v_j(n)) \end{aligned} \quad (3.34)$$

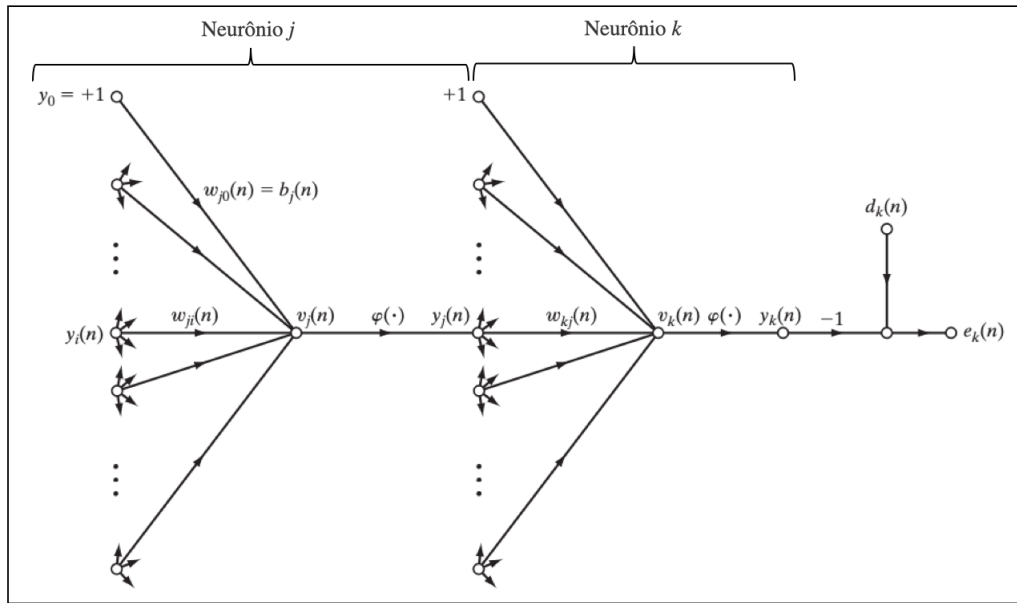
A Equação 3.33 é utilizada para ajustar os pesos sinápticos dos neurônios da camada de saída diretamente a cada iteração. Porém, esta mesma equação não pode ser utilizada de forma direta para calcular os ajustes dos pesos dos neurônios das camadas

intermediárias, uma vez que os neurônios dessas camadas não possuem um vetor resposta desejado. Assim, adaptações à Equação 3.33 são necessárias para que este sinal de erro seja calculado recursivamente, através da propagação dos outros sinais de erro dos neurônios diretamente conectados a esta camada intermediária.

A Figura 6 representa o fluxo de informação de um neurônio de camada intermediária conectado a outro neurônio de camada de saída. E com base nela e na Equação 3.34, um neurônio j de uma camada intermediária tem definido como seu gradiente local

$$\begin{aligned}\delta_j(n) &= -\frac{\partial \epsilon(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \\ \delta_j(n) &= -\frac{\partial \epsilon(n)}{\partial y_j(n)} \varphi'_j(v_j(n)).\end{aligned}\quad (3.35)$$

Figura 6 – Fluxo de informações de um neurônio na camada intermediária



Fonte: Adaptado de Haykin (2007).

Para calcular a derivada parcial $\frac{\partial \epsilon(n)}{\partial y_j(n)}$, o erro $\epsilon(n)$, obtido a partir da Figura 6, é definido como

$$\epsilon(n) = \frac{1}{2} \sum_{k \in C} e_k^2(n), \quad (3.36)$$

e diferenciando a Equação 3.36 em relação ao sinal y_j e aplicando a regra da cadeia, tem-se

$$\begin{aligned}\frac{\partial \epsilon(n)}{\partial y_j(n)} &= \sum_k e_k \frac{\partial e_k(n)}{\partial y_j(n)} \\ \frac{\partial \epsilon(n)}{\partial y_j(n)} &= \sum_k e_k \frac{\partial e_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_j(n)}.\end{aligned}\quad (3.37)$$

Da Figura 6, notamos que

$$\begin{aligned} e_k(n) &= d_k(n) - y_k(n) \\ e_k(n) &= d_k(n) - \varphi_k(v_k(n)) \end{aligned} \quad (3.38)$$

e

$$v_k(n) = \sum_{j=0}^m w_{kj}(n) y_j(n), \quad (3.39)$$

onde m é o número total de informações de entrada do neurônio k , e, portanto

$$\frac{\partial e_k(n)}{\partial v_k(n)} = -\varphi'_k(v_k(n)) \quad (3.40)$$

e

$$\frac{\partial v_k(n)}{\partial y_j(n)} = w_{kj}(n). \quad (3.41)$$

Assim, usando as equações 3.40 e 3.41, chega-se na derivada parcial desejada da Equação 3.35,

$$\frac{\partial \epsilon(n)}{\partial y_j(n)} = - \sum_k e_k(n) \varphi'_k(v_k(n)) w_{kj}(n) \quad (3.42)$$

$$\frac{\partial \epsilon(n)}{\partial y_j(n)} = - \sum_k \delta_k(n) w_{kj}(n). \quad (3.43)$$

Usando a Equação 3.43 na Equação 3.35, chega-se na fórmula do *Backpropagation* para o gradiente local $\delta_j(n)$, o qual é calculado recursivamente a partir dos pesos sinápticos $w_{kj}(n)$ e do gradiente local $\delta_k(n)$ das camadas posteriores (no caso da Figura 6, a camada de saída). O gradiente é definido como

$$\delta_j(n) = \varphi'_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n). \quad (3.44)$$

Com todas as equações definidas, o treinamento de uma RNA através do algoritmo *Backpropagation* pode ser resumido, segundo Haykin (2007), como:

1. Inicialização: Assumindo que não exista nenhuma informação prévia disponível, os pesos sinápticos são inicializados com valores aleatórios segundo uma distribuição normal com média zero e desvio padrão de modo que os valores iniciais dos campos locais induzidos ($v_j(0)$) não estejam na região de saturação da função de ativação $\varphi(\cdot)$.
2. Processamento de propagação: Apresenta-se à rede um conjunto de exemplos de treinamento, denominados $(x(n), d(n))$. Para cada vetor de entrada $x(n)$ aplicado na camada inicial, calcula-se um vetor resposta $y(n)$ sequencialmente, camada por camada, da entrada até a saída, enquanto os pesos sinápticos $w_{ji}(n)$ permanecem inalterados. Computa-se, então, o sinal de erro na camada final a partir da diferença entre o vetor resposta desejado $d(n)$ e o vetor resposta obtido na camada final $y(n)$.

3. Processamento de retropropagação: Com o sinal de erro obtido no passo anterior, calcula-se, para cada camada da rede, os gradientes locais $\delta(n)$ e o novo valor dos pesos sinápticos $w_{ji}(n+1)$, começando na camada de saída e terminando na camada de entrada.
4. Iteração: Repete-se as etapas 2 e 3 até que se atinja os critérios de parada ou até que todos os vetores de entrada, de todos os conjuntos de exemplos de treinamento, disponíveis sejam processados.

O processo do algoritmo *Backpropagation* descrito acima reflete a sua forma mais simples, a qual proporciona um desempenho baixo quando comparado com suas formas modificadas. Essas modificações podem ser feitas através da utilização de técnicas heurísticas ou através de métodos de otimização numérica, como mostrado por Demuth et al. (2014). Neste trabalho o método de otimização utilizado será o L-BFGS.

Outros detalhes devem ser discutidos paralelamente à escolha do método de treinamento de uma rede neural. São eles:

- Normalização das informações de entrada e saída.
- Como analisar e tratar o *overfitting*, que é o efeito da redução da capacidade de generalização da RNA por conta da memorização dos ruídos, como por exemplo a utilização de um polinômio de grau demasiadamente alto utilizado numa interpolação polinomial:
 1. Pode-se separar a amostra em duas partes, uma de treinamento e uma de validação, sendo esta segunda utilizada para acompanhar e encerrar o processo de treinamento quando se tem uma deterioração nas respostas após um certo tempo de convergência. Neste trabalho será introduzida uma abordagem de janela móvel a ser explicada mais a frente.
 2. Aplicação do método de regularização Baynesiana, que consiste em adicionar um termo de penalização à função custo, de forma que o algoritmo de estimação faça com que os parâmetros irrelevantes converjam para zero, reduzindo assim o número de parâmetros efetivos utilizados no processo. Neste trabalho, esse parâmetro será variado entre os valores pertencentes a $[10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3]$.

3.4.2 O Otimizador L-BFGS

Durante o processamento do algoritmo *Backpropagation*, apresentado no item anterior, temos um otimizador responsável pela eficiência da atualização dos pesos sinápticos. Neste trabalho optou-se pelo método *Limited-memory Broyden-Fletcher-Goldfarb-Shanno* (L-BFGS), que corresponde a uma aproximação do método BFGS através do uso de

uma quantidade limitada de memória computacional. Este método é da família quasi-Newtoniana, considerada uma generalização dos métodos secantes de encontrar raízes da primeira derivada de funções multidimensionais. Por este motivo são caracterizados como métodos de segunda ordem. Métodos desta família utilizam técnicas simples, chamadas de *Hill Climbing*, de busca local sem o armazenamento do caminho percorrido até a solução corrente, armazenando-se somente a solução propriamente dita como estado, sem a necessidade do conhecimento da direção e magnitude do gradiente.

O objetivo do método L-BFGS neste trabalho é melhorar a correção do peso sináptico realizada em 3.33, aplicando uma atualização da forma:

$$\Delta w_{ji}(n) = \eta G(n), \quad (3.45)$$

onde a direção descendente de G é definida por:

$$G(n) = -[\tilde{H}(n)]^{-1} \frac{\partial \epsilon(n)}{\partial w_{ji}(n)}, \quad (3.46)$$

e onde $\tilde{H}(n)$ é a aproximação da matriz Hessiana da função da energia instantânea do erro quadrático $\epsilon(n)$. A diferença entre o método BFGS e L-BFGS é que o primeiro armazena toda a matriz $\tilde{H}(n)$ em cada iteração para ser utilizada na próxima, e o segundo armazena apenas uma parte das informações e calcula uma aproximação da matriz Hessiana novamente. Por um lado, ao utilizar o método L-BFGS pode haver uma perda de acurácia na otimização ao não se armazenar a matriz completa em cada iteração, mas por outro pode haver um ganho de acurácia ao se calcular uma nova matriz em cada iteração.

Esta alternativa de usar o modelo L-BFGS ao BFGS está relacionada à redução do custo computacional quando se usa o primeiro, por conta da não necessidade de se armazenar todos esses dados da matriz em cada iteração. Esse ganho é relevante quando se tem uma quantidade elevada de pesos sinápticos e vieses ao longo das camadas da RNA, o que por consequência aumenta o tamanho da matriz em questão. Vale ressaltar que o uso de otimizadores quasi-Newtonianos, mesmo o L-BFGS, requer maior poder computacional quando comparado com otimizadores estocásticos como o Gradiente Descendente Estocástico (SGD) e o Adam. Porém quando comparado ao SGD, o L-BFGS pode apresentar uma maior estabilidade por conta do uso da busca linear e possui menos hiper parâmetros para se ajustar. Dadas as condições citadas e o fato de a base de dados deste trabalho ser pequena, optou-se pelo uso do otimizador L-BFGS. Para mais detalhes sobre a iteração que computa a aproximação da matriz Hessiana citada, consultar Nocedal (1980).

Feita a fundamentação da IV, VRP, RV e RNAs, bem como seu algoritmo de treinamento *Backpropagation* e seu otimizador L-BFGS, tem-se todo o arcabouço teórico necessário para a implementação de um modelo de RNA para a estimação da RV e para a comparação deste com outros modelos *benchmark*. Tal processo de implementação e comparação dos resultados originados deste modelo com os outros será apresentado no capítulo seguinte.

4 METODOLOGIA

Com a teoria fundamentada, o próximo passo é estruturar, tratar e selecionar os dados que serão utilizados como informações de entrada do modelo de RNA e de regressão linear múltipla (RLM) para a estimação da variância realizada. Cada um destes dois modelos apresentam duas abordagens, a primeira é o treinamento e regressão a partir somente da RV e a segunda através do VRP, o qual é uma combinação da RV com o VIX. Teremos assim dois modelos de RNA e dois modelos de RLM.

O quinto modelo a ser processado será o próprio VIX como preditor da RV dos 30 dias corridos seguintes. O próximo passo será então comparar os cinco modelos entre si, de forma que os dois modelos de RLM e o VIX sejam considerados modelos *benchmark*. Esta comparação será feita a partir dos valores de R^2 de cada modelo nos cenários estudados. Este processo de estimação da variância a ser realizada nos próximos 30 dias corridos a partir dos cinco modelos, bem como a comparação entre eles, será descrito neste capítulo.

4.1 Dados de entrada

O conjunto de dados de entrada consiste em:

- Livro de ordens diário (*BID* e *ASK*), capturados no fechamento do pregão, das opções do índice EWZ¹, negociadas na bolsa de valores americana NYSE, no período de 06 de Janeiro de 2016 a 01 de Abril de 2020. A tabela ilustrada na Figura 1 exemplifica estes dados de entrada.
- Preços diários de fechamento do índice EWZ, negociado na bolsa de valores americana NYSE, no período de 06 de Janeiro de 2016 a 01 de Maio de 2020.
- Dados diários da taxa básica de juros americana de um dia (*Fed Funds*).

A escolha do índice EWZ como ativo objeto deste trabalho se deu por conta dos seguintes motivos: maior volume de negociação das opções do índice EWZ quando comparado com as opções do índice IBOVESPA negociadas na B3, vencimentos semanais das opções do EWZ frente aos vencimentos mensais das opções do IBOVESPA e o fato de o índice EWZ ser um ativo negociado em bolsa de valores estrangeira que representa bem o mercado brasileiro.

¹ O EWZ é um *Exchange-Traded Fund* (ETF), que contém uma cesta de ações de médio e grande porte da bolsa brasileira. Mais informações em <https://www.ishares.com/us/literature/fact-sheet/ewz-ishares-msci-brazil-etf-fund-fact-sheet-en-us.pdf>

Segue abaixo uma ilustração da tabela contendo exemplos das informações de entrada das opções, onde as datas são representadas pelo ano, mês e dia, respectivamente:

Tabela 1 – Corte da base de entrada das opções.

Data	Bid	Ask	Vencimento	Tipo	Strike
2016-01-05	6.10	6.35	2016-02-05	C	14.0
2016-01-05	0.01	0.07	2016-02-05	P	14.0
2016-01-05	5.60	5.85	2016-02-05	C	14.5
2016-01-05	0.01	0.08	2016-02-05	P	14.5
2016-01-05	5.15	5.35	2016-02-05	C	15.0
2016-01-05	0.02	0.09	2016-02-05	P	15.0
2016-01-05	4.65	4.85	2016-02-05	C	15.5
2016-01-05	0.04	0.10	2016-02-05	P	15.5
2016-01-05	4.15	4.40	2016-02-05	C	16.0
2016-01-05	0.06	0.12	2016-02-05	P	16.0

As colunas da Tabela 1 e o que representam são:

- *Data*: o dia do pregão em que os dados foram capturados.
- *Bid*: o maior preço de comprador no final do pregão.
- *Ask*: o menor preço de vendedor no final do pregão.
- *Vencimento*: data em que o contrato de opção em questão expira.
- *Tipo*: C representa as opções de compra, *Calls*, e P representa as opções de venda, *Puts*.
- *Strike*: preço de exercício do contrato de opção em questão.

No modelo aqui apresentado, há 4 informações a serem escolhidas pelo usuário, que serão parâmetros de entrada para o modelo. São elas:

1. *T*: Janela de previsão em dias corridos. Neste trabalho ela foi fixada em 30 dias.
2. ΔK : Intervalo entre preços de exercício. Valor fixado em \$0,50 neste trabalho.
3. *N*: Número de opções a serem utilizadas em cada data base. Vale ressaltar que este número é constante para cada teste. Neste trabalho serão testados os casos de 9, 10, 11, 12 e 13 opções para cada data base, que serão explicados mais a frente.
4. Cenário *Z*: Filtro de datas considerando acontecimentos históricos, sendo eles:
 - a) 1: Sem filtro de datas.

- b) 2: Filtro *Joesley Day*. Remoção do intervalo entre as datas 18/04/2017 e 22/05/2017, que são próximas ao *Joesley Day*²
- c) 3: Filtro *Covid*. Remoção das datas posteriores a 10/02/2020, data que antecede em 30 dias o início do impacto nos mercados mundiais da pandemia referente ao coronavírus *SARS-CoV-2*³.
- d) 4: Filtro *Joesley Day e Covid*. Combinação dos cenários 2 e 3 com a remoção do intervalo entre as datas 18/04/2017 e 22/05/2017 e a remoção das datas posteriores a 10/02/2020.

4.2 Tratamento dos Dados

Antes de implementarmos o modelo da RNA, do VIX e da RLM, é necessário que se faça um tratamento prévio dos dados de entrada. O objetivo desta etapa é obter os valores resposta esperados, que no caso é a variância realizada do índice EWZ nos próximos T dias, e um conjunto de preços de N opções, $O_t(K_h, T)$, para cada data base disponível t , onde $N = 2n + 1$ é o número constante de opções escolhidas para o modelo, T é o prazo para o vencimento das opções e K_h os preços de exercícios das opções ao redor do preço de exercício médio K_0 ,

$$\{K_{-n}, K_{-n+1}, \dots, K_{-1}, K_0, K_1, \dots, K_{n-1}, K_n\}, \quad (4.1)$$

com a condição de que $\Delta K = K_i - K_{i-1}$ seja o mesmo para todas as datas disponíveis.

4.2.1 Tratamento das Opções

Com a posse dos dados iniciais das opções, representados parcialmente na Tabela 1, inicia-se o seu tratamento conforme as seguintes etapas:

1. Definição de preço das opções: Será considerado como preço (*mid*) a média aritmética dos valores do *bid* e do *ask* do fechamento diário de cada derivativo. Em caso de

² *Joesley Day* é uma menção ao dia 17/05/2017, em que houve a divulgação de um áudio gravado pelo dono da empresa JBS, Joesley Batista, no qual ele registrou o até então Presidente Michel Temer dando aval para comprar o silêncio do ex-deputado Eduardo Cunha, que estava preso por conta de desdobramentos da Lava Jato. O grau de incerteza em relação ao mercado brasileiro aumentou e o índice IBOVESPA chegou a cair mais de 12% em menos de 1 hora.

³ COVID-19 é a doença causada pelo coronavírus *SARS-CoV-2*, que em Dezembro de 2019 teve sua primeira transmissão identificada em Wuhan na China. A doença possui uma taxa de contágio elevada, passando de pessoa para pessoa, o que causou a sua disseminação ao redor do mundo. Em meados de Março os mercados globais foram impactados pela realização da gravidade da doença, que levaria a diversas medidas governamentais de restrições de circulação de pessoas e paralisações de diversos serviços ao redor do mundo. Até o momento da escrita deste trabalho ainda não se tem uma vacina contra o vírus, acumula-se um total de 30 milhões de pessoas infectadas e 930 mil pessoas mortas no mundo.

ausência de medida, ou seja, sem interesse de compra ou venda no mercado, o valor do *bid* ou do *ask* considerado para o cálculo do *mid* será zero.

2. Cálculo da diferença entre o *mid* das *calls* e das *puts* (*CPdif*): Para cada preço de exercício de uma mesma data base e de um mesmo vencimento, calcula-se o *CPdif* das opções subtraindo o preço *mid* das *calls* pelo preço *mid* das *puts*.
3. Definição do preço de exercício médio (K_0): Para cada data base e cada vencimento, e considerando todos os preços de exercício disponíveis para elas, seleciona-se o *CPdif* absoluto mínimo, e o seu *strike* correspondente será o *strike* médio K_0 de cada vencimento para cada um destes dias.
4. Cálculo do preço a termo, *forward*, do ativo base: Com $C - P$ sendo a diferença entre o preço da *Call* e da *Put*, também chamado de *CPdiff* neste trabalho, e um K_0 determinado para cada vencimento dentro de cada data base, calcula-se o *forward*, F , como

$$F = K_0 + (C - P)e^{RT}, \quad (4.2)$$

onde e é a constante de Euler, R é a taxa de juros básica americana, *Fed Funds*, para cada data base e T é o prazo, em dias corridos, da data base até o vencimento da opção analisada. A aplicação deste preço *forward* será explicada mais a frente.

5. Separação das opções fora do dinheiro para as *calls* e *puts*: Tanto o modelo do VIX quanto o modelo de RNA aqui apresentado, utilizam apenas as opções fora do dinheiro (OTM) para as *calls* e para as *puts*, sendo que para cada preço de exercício, para cada vencimento e para cada data base tem-se ou uma *call* ou uma *put* como opção correspondente. Vale ressaltar que a única exceção é para a opção que contém o *strike* médio, a qual é uma opção no dinheiro (ITM), e tem-se tanto uma *call* quanto uma *put* em sua representação. A regra para esta separação é:
 - $K_i > K_0$ -> Seleciona-se somente as *calls*.
 - $K_i < K_0$ -> Seleciona-se somente as *puts*.
 - $K_i = K_0$ -> Seleciona-se a *call* e a *put* (opções ITM).
6. Tratamento das opções no dinheiro (ITM): Como temos tanto uma *call* quanto uma *put* para o caso em que $K_i = K_0$, combina-se as duas através de uma média aritmética de seus *mid* para se obter uma opção final. Vale ressaltar que isto é feito para cada vencimento dentro de cada data base.
7. Remoção das opções mais fora do dinheiro (mais longe do K_0) a partir do aparecimento de dois *bids* zerados consecutivos: A partir do *strike* médio, K_0 , os *bids* são lidos consecutivamente nos dois sentidos, tanto para o lado crescente do *strike* quanto

para o lado decrescente. Quando se tem dois *bids* zerados consecutivos, ignora-se todas as opções com *strike* deste segundo *bid* em diante. Exemplo:

Tabela 2 – Exemplo de tratamento para dois *bids* consecutivos para um vencimento em uma data base

Data	Vencimento	i	Strike (Ki)	CPdif	Mid	Bid	Ask
2019-06-25	2019-07-12	-6	27.00	2.30	0.08	0.06	0.10
2019-06-25	2019-07-12	-5	27.50	1.80	0.12	0.13	0.11
2019-06-25	2019-07-12	-4	28.00	1.25	0.82	0.90	0.74
2019-06-25	2019-07-12	-3	28.50	1.12	0.92	1.01	0.83
2019-06-25	2019-07-12	-2	29.00	1.10	1.32	1.45	1.19
2019-06-25	2019-07-12	-1	29.50	0.40	1.52	1.67	1.37
2019-06-25	2019-07-12	0	30.00	0.06	2.32	2.55	2.09
2019-06-25	2019-07-12	1	30.50	0.90	1.82	2.00	1.64
2019-06-25	2019-07-12	2	31.00	1.52	1.12	1.23	1.01
2019-06-25	2019-07-12	3	31.50	2.30	0.60	0.66	0.54
2019-06-25	2019-07-12	4	32.00	2.83	0.12	0.00	0.24
2019-06-25	2019-07-12	5	32.50	3.80	0.04	0.00	0.08
2019-06-25	2019-07-12	6	33.00	3.95	0.01	0.00	0.02

Para o exemplo fictício representado na Tabela 2, o vencimento, para a data base em questão, sofreria um corte tal qual deixaria o seu conjunto de opções sendo

$$\{K_{-4}, K_{-3}, K_{-2}, K_{-1}, K_0, K_1, K_2, K_3, K_4\}. \quad (4.3)$$

8. Seleção das opções com vencimento próximo ou correspondente a T dias corridos a frente da data base: O vencimento das opções é semanal, portanto há diferentes vencimentos para uma mesma data base e é necessário selecionar o vencimento que corresponda com a janela pré determinada pelo usuário, T , que no caso deste trabalho foi fixada em 30 dias. Há casos em que uma data base possui um conjunto de opções com vencimento 30 dias corridos a sua frente, D_{30} , porém há outros em que não. Para o primeiro caso, seleciona-se diretamente o conjunto de opções com vencimento 30 dias corridos a frente da data base. Para o caso em que não há um vencimento correspondente a 30 dias corridos a frente da data base, escolhe-se os dois vencimentos mais próximos, um inferior a 30 dias e o outro superior.
9. Seleção do *strike* médio para cada data base: Para as datas base com apenas um vencimento, casos em que houve correspondência de 30 dias corridos, o K_0 será o *strike* médio já determinado para essas opções. Já para o caso em que há dois vencimentos distintos para uma mesma data base, o *strike* médio escolhido será o do conjunto de opções com vencimento mais próximo de D_{30} , e se os dois vencimentos estiverem equidistantes, escolhe-se o *strike* médio que possuir o menor *CPdiff*.

10. Interpolação do *mid* para cada data base que possuir dois vencimentos: Com base nas distâncias entre os dois vencimentos e D_{30} , interpola-se linearmente os preços das opções de mesmo *strike*. Com isso, tem-se apenas um conjunto de opções com vencimento D_{30} para cada data base.
11. Acréscimo de *strikes* para se ter o mesmo intervalo, ΔK , entre todas as opções: Como os dados de entrada não possuem o mesmo intervalo entre *strikes* subjacentes⁴, realiza-se um acréscimo de opções fictícias para preencher os intervalos que não atendam ao parâmetro de entrada, ΔK , do usuário, que neste trabalho foi escolhido como \$0,50. Tomamos um exemplo onde, para uma mesma data base, temos uma opção com *strike* \$42,00 e *mid* \$0,20, e uma opção subjacente com *strike* \$43,00 e *mid* \$0,10. Cria-se uma opção com *strike* \$42,50 e *mid* \$0,15. O cálculo do *mid* é feito através de uma interpolação linear com base nos *strikes* originais. Vale ressaltar que os limites dos *strikes* inferiores e superiores foram determinados com base nos preços de exercício originais, selecionados pela regra dos dois *bids* consecutivos explicado anteriormente.
12. Normalização do *mid* para remover o impacto temporal sobre o valor do dinheiro: O valor do dinheiro muda com o tempo, isto é, \$10,00 em 2008 é diferente de \$10,00 em 2015 (LAW; SMULLEN, 2008). Para remover este efeito dos dados de entrada e se obter um estimador de preço neutro a risco, é feita uma normalização nos preços das opções selecionadas, agora chamado de mid_{antigo} , com base no preço de exercício delas, K_i :
$$mid^{(i)} = \frac{mid_{antigo}^{(i)}}{K_i^2}, \quad (4.4)$$
com $i \in \{-n, -n + 1, \dots, n - 1, n\}$, sendo $N = 2n + 1$ o número total de opções disponíveis em cada data base. A definição do estimador neutro a risco pode ser encontrada em Bahra (1997).

4.3 Seleção das opções

Com os dados tratados, há uma etapa de filtro das opções antes que se iniciem os processamentos dos modelos. Essa etapa de filtragem é feita com base nos dois parâmetros de entrada do usuário:

- O cenário $Z \in \{1, 2, 3, 4\}$, representando os seguintes cenários de mercado: nenhum filtro, *Joesley Day*, *Covid* e *Joesley Day e Covid* respectivamente.

⁴ Opções com preço de exercício mais próximos do preço de exercício médio de seu vencimento tendem a ter um intervalo ΔK de \$0,50, enquanto opções com *strike* mais afastado do K_0 tendem a ter um ΔK de \$1,00.

- O número de opções N por data base.

Com o primeiro filtro de datas feito a partir do cenário, analisa-se o a quantidade de opções em cada data base para que se tenha N opções ao final do segundo filtro. Avalia-se se o número máximo i , em K_i , de cada data base t é maior ou igual a n , em $N = 2n + 1$. Caso i máximo seja menor que n , esta data base é removida do conjunto de dados de entrada, e caso contrário, considera-se i máximo como n . Dessa forma teremos n opções com *strike* abaixo do *strike* médio K_0 , uma única opção que possui o *strike* médio e n opções com *strike* acima do K_0 , totalizando N opções para cada data base que atende aos requisitos mínimos já explicados.

4.4 Pré processamento dos dados

Com os dados das opções tratados, faz-se uma padronização dos dados de entrada, neste caso os preços *mid* das opções, para cada data base t . Optou-se por este pré-processamento de dados pois sabe-se empiricamente que a distribuição dos preços (*mid*) OTM das opções apresentam uma forma próxima a de uma Gaussiana centrada no preço da opção com *strike* médio. Assim, ao se padronizar os dados para cada data base, obtemos um conjunto de dados distribuídos na forma de uma normal padrão, com média 0 e desvio padrão 1, que por sua vez ajuda os modelos a terem um desempenho melhor. Este processo é realizado da seguinte forma:

$$z_{i,t} = \frac{x_{i,t} - u_t}{s_t}, \quad (4.5)$$

com $i \in \{-n, -n + 1, \dots, n - 1, n\}$, sendo $N = 2n + 1$ o número total de opções disponíveis em cada data base, $z_{i,t}$ o novo valor do preço *mid*, $x_{i,t}$ o preço *mid* a ser padronizado, u_t a média dos preços *mid* para uma data base específica t e s_t o desvio padrão da amostra em t .

4.5 Cálculo da Variância Realizada dos Próximos T Dias

A Variância Realizada (RV) será a medida a ser prevista pelos modelos, portanto conhecê-la é fundamental tanto para efeitos de medida de desempenho quanto para o treinamento da RNA. O objetivo desta etapa é obtê-la a partir dos preços do índice EWZ. Por conta da limitação de dados de mercado das opções e, por consequência, da decisão de usar o preço de fechamento das mesmas ao invés dos preços *intraday*, será aplicado a mesma lógica na frequência de preços do índice EWZ para que as informações estejam alinhadas.

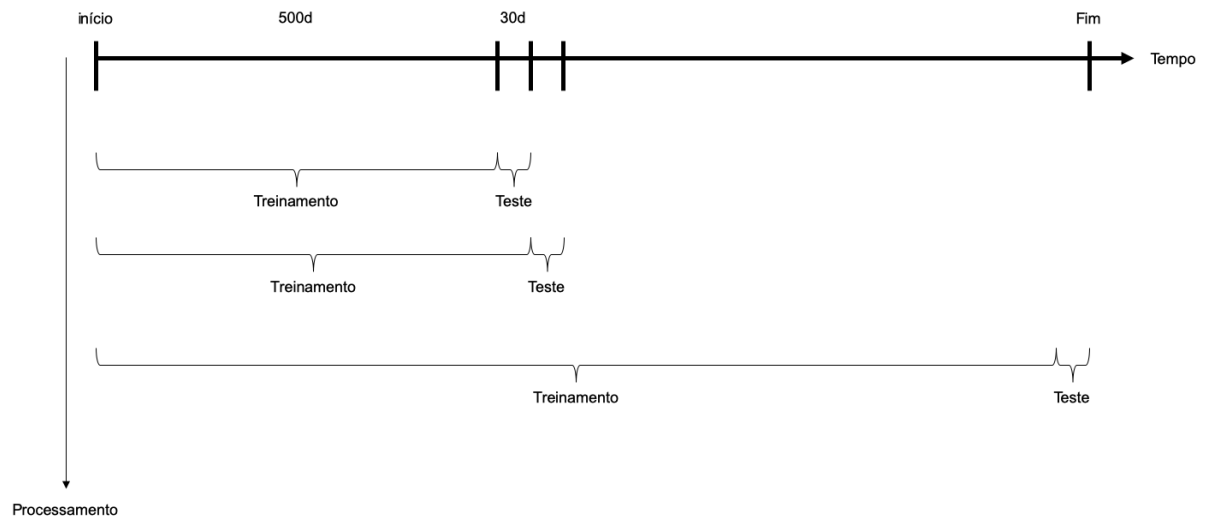
Com o cálculo realizado a partir da descrição feita em 3.2 e do parâmetro de entrada T , igual a 30 dias corridos para este trabalho, tem-se os valores da variância realizada para cada data base a ser prevista pelo modelo.

4.6 Janela Móvel

Como o problema descrito neste trabalho é baseado em séries temporais, adotou-se uma abordagem de janela móvel para o treinamento dos modelos de regressão linear múltipla (RLM) e de RNA, uma vez que à medida que o tempo passa, mais informações podem ser aproveitadas constantemente. Esta janela móvel leva em conta uma janela estabelecida em 30 dias, e um valor para a amostra inicial dos dados para cada processamento, que neste trabalho foi testada para 400, 500, 600 e 700 dias. Vale ressaltar que a escolha do tamanho da janela móvel foi arbitrária e decidiu-se não testar diferentes combinações de amostras para se evitar o *overfitting*, uma vez que a quantidade de dados disponíveis para este trabalho é limitada.

Começa-se com uma amostra de datas base, 500 dias por exemplo, e o primeiro passo é treinar os modelos de RLM e RNA com estas informações e testá-los com os próximos 30 dias de dados. O segundo passo é treinar novamente o modelo com todos os dados já utilizados na etapa anterior, no caso deste exemplo 530 dias, e testá-lo com os próximos 30 dias, que compreende do 531º dia até o 560º dia, e assim sucessivamente conforme ilustrado na Figura 7.

Figura 7 – Exemplo de janela móvel para amostra inicial de 500 dias.



Fonte: Elaborada pelo autor.

4.7 Modelos de previsão da variância realizada

Neste trabalho foram comparados cinco formas diferentes de se obter uma previsão da variância realizada: VIX como preditor da RV, RLM em relação à própria RV, RLM em relação ao VRP e duas RNAs, com a RV como vetor resposta de uma e o VRP de outra. A implementação de cada um destes cinco modelos serão descritos nos próximos itens.

4.7.1 Variância a partir do VIX

O VIX do EWZ (VXEWZ) é um índice lido no mercado através da negociação de seu ETF na bolsa americana *Chicago Board Options Exchange* (CBOE). Porém, como ele será utilizado tanto como um modelo de comparação com a RNA quanto como um parâmetro de cálculo do VRP, neste trabalho optou-se por realizar o seu cálculo manual para que se tenha um alinhamento de todos os dados de entrada dos diferentes modelos. Vale ressaltar que o modelo oficial do VIX utiliza o *BID* e o *ASK* de opções do EWZ lidos do mercado a cada minuto, cuja utilização neste trabalho se tornou inviável por conta da limitação de acesso a estes dados históricos.

Como explicado em 2.3 e 3.1.2, a volatilidade implícita a partir do VIX pode ser utilizada como um preditor da volatilidade realizada, isto é, o VIX representa a expectativa do mercado em relação à volatilidade anualizada que será realizada nos próximos T dias. Ele é lido como um número na base 100, ou seja, um VIX de valor 50 representa a esperança de uma volatilidade anualizada de 50%.

Com os dados das opções filtrados e com o cálculo do preço *forward* feito, conforme explicado no Item 4.2.1 de tratamento das opções, realiza-se o cálculo manual do VIX_t para cada data base t da seguinte forma:

$$VIX_t = 100\sqrt{\sigma_t^2}. \quad (4.6)$$

Com σ_t^2 sendo definido como

$$\sigma_t^2 = \frac{2}{\frac{T}{360}} \sum_i \Delta K_i e^{\frac{R_t T}{360}} mid_{i,t} - \frac{1}{\frac{T}{360}} \left[\frac{F_t}{K_{0,t}} - 1 \right]^2, \quad (4.7)$$

onde T é a janela em dias, i representa cada opção dentro das N possíveis da data base t , ΔK é o intervalo fixo entre um *strike* e outro, R_t é a taxa básica de juros americana na data base em questão, $mid_{i,t}$ é o preço *mid* determinado no Item 4.2.1 para cada opção i , F_t é o preço *forward* da data base t e $K_{0,t}$ é o preço de exercício médio da data t .

Como a medida utilizada neste trabalho é a variância e o VIX pode ser entendido como o desvio padrão anualizado dos retornos do ativo na base 100, tomaremos a expectativa da variância diária a ser realizada nos próximos T dias, Var_{VIX} , como:

$$Var_{VIX} = \left(\frac{VIX}{100} \frac{1}{\sqrt{365}} \right)^2. \quad (4.8)$$

Vale ressaltar que para transformar o VIX em diário devemos multiplicar o VIX original por $\sqrt{365}$ uma vez que o seu modelo é baseado no calendário de 365 dias corridos. Por simplicidade, o modelo de variância a partir do VIX será chamado de VIX neste trabalho a partir deste ponto. O seu valor será tanto um método de previsão da variância

realizada, modelo considerado como um *benchmark* para efeitos comparativos com a RNA, quanto uma informação a ser calculada em conjunto com a RV para se obter o VRP, a ser explicado mais a frente.

4.7.2 Regressão Linear Múltipla

Como modelo *benchmark* usaremos a RLM, que segundo Heij et al. (2004) é descrita como um modelo que explica uma variável y , dependente, a partir de um conjunto de variáveis x_1, x_2, \dots, x_k , $k > 0$, conforme,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon, \quad (4.9)$$

em que β_i são os coeficientes da regressão, onde $i = 0, 1, \dots, k$, β_0 é o termo constante e ϵ são fatores residuais que não estão incluídos no modelo. O seu objetivo é encontrar o conjunto de valores de β_i que minimize a soma dos quadrados dos resíduos e , ou seja, minimize $\sum_{j=1}^n e_j^2$, com n sendo o número de dias utilizados no modelo. Para isso, será utilizado o método dos mínimos quadrados (MMQ).

Com os dados pré processados e seguindo a dinâmica de janela móvel, faz-se uma RLM das informações das opções para a variância realizada nos T dias a frente de cada data base, medida a posteriori. Ou seja, o y da RLM é representado pelo RV_t , a variância diária realizada dos próximos T dias a partir de t , calculado em 4.5, e os x_{i+1} , com $i = 0, 1, \dots, N - 1$, são os preços *mid* padronizados de cada opção da data base t . Para cada etapa de processamento da janela móvel, treina-se o modelo com um determinado tamanho de amostra, que aumenta em 30 dias a cada passo, e testa-se o seu desempenho sobre as informações disponíveis nos 30 dias seguintes. O desempenho deste modelo, aqui chamado de RLM_{VR} será considerado na comparação com a RNA.

Para a segunda abordagem de regressão, em relação ao VRP ao invés da RV diretamente, utilizou-se o modelo VIX para o cálculo da variável dependente, da seguinte forma:

$$VRP = VIX - RV, \quad (4.10)$$

onde o VRP é a variável dependente, o VIX é o modelo explicado no Item 4.7.1, equivalente ao quadrado do VIX , e o RV é a variância realizada dos próximos 30 dias corridos. Este outro modelo de RLM será chamado de RLM_{VRP} .

4.7.3 Redes Neurais Artificiais

O terceiro e último modelo de estimação da variância realizada a ser apresentado é o de Redes Neurais Artificiais, foco deste trabalho. O modelo de RNA aqui implementado é o Multi Layer Perceptron (MLP), descrito no Item 3.4. A linguagem utilizada para a implementação dos algoritmos deste trabalho foi *Python* com o auxílio dos pacotes do *Sklearn* para as etapas de *Machine Learning*.

Como já explicado em 3.4 há diversos parâmetros a serem selecionados no processamento das redes, também chamados de hiperparâmetros. Uma prática comum em modelagem é otimizar um grande número de combinações de hiperparâmetros de modo que se obtenha a configuração com o melhor desempenho, porém isso é bastante sensível quando se trabalha com séries temporais financeiras por conta da facilidade de se gerar *overfitting*. Com o objetivo de se evitar o problema recém destacado, neste trabalho optou-se por variar somente alguns dos hiperparâmetros possíveis, conforme descrito a seguir:

- Número de camadas escondidas: Optou-se por utilizar apenas uma camada escondida.
- Número de neurônios nas camadas escondidas: A escolha do número de neurônios é variável e corresponde à quantidade de dados das opções na camada de entrada. Caso opte-se, por exemplo, por utilizar N igual a 17 opções, então haverá 17 neurônios na camada escondida.
- Algoritmo de treinamento: *Backpropagation*.
- Otimizador: L-BFGS.
- Função de ativação da camada escondida: ReLu.
- Função de ativação da camada de saída: Função identidade. Equivalente a apenas replicar o valor na camada de saída, como $f(x) = x$.
- Taxa de penalização (λ): Testa-se alguns valores para a taxa de penalização. $\lambda \in 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3$.

Com os parâmetros determinados acima, os dados já tratados e pré processados e o conceito de janela móvel estabelecido, inicia-se os processamentos dos algoritmos da RNA. Esta fase foi dividida em duas abordagens distintas. A primeira é baseada na regressão dos preços *mid* das opções em relação à RV, modelo chamado de RNA_{RV} , e a segunda é baseada na regressão dos preços *mid* das opções em relação ao VRP, modelo chamado de RNA_{VRP} .

Regressão via MLP em relação à RV

O processamento do MLP é realizado em cada etapa de treinamento da janela móvel. Isto é, suponhamos uma configuração em que tenha-se um total fixo de 17 opções por dia e uma amostra inicial de 400 dias com dados. A camada de entrada receberá 400 vetores, x_1, x_2, \dots, x_{400} , com 17 elementos cada, aos quais serão aplicados pesos sinápticos aleatoriamente inicializados, representado pela matriz W_1 de tamanho 17x17, para cada data base. Isso significa que todos os 17 valores de entradas, multiplicados por diferentes pesos, alimentam todos os 17 neurônios da camada escondida. A saída de cada neurônio

passará pela função de ativação ReLu, $\varphi_1(x) = \max(0, x)$, e a combinação dessas saídas será multiplicada pela matriz de peso sináptico W_2 , de tamanho 17x1 para cada data base, e também inicializada aleatoriamente.

O resultado desta operação será um valor na camada de saída, cujo próximo processamento será passar pela função de ativação final, representada pela função identidade $\varphi_2(x) = x$. Assim, obtém-se um valor estimado \hat{y} do valor real conhecido y , que representa a variância realizada do índice EWZ dos próximos 30 dias corridos a contar de cada data base em questão. Considerando o exemplo de apenas uma data base da situação descrita, a iteração da camada de entrada até a camada de saída, que produz o valor estimado \hat{y} , pode ser descrita pela equação

$$\hat{y} = \varphi_2 \left(\sum_{j=0}^{17} w_{2,j} \varphi_{1,j} \left(\sum_{i=0}^{17} w_{1,ij} x_i \right) \right). \quad (4.11)$$

Através da função perda do erro quadrático, descrita em 3.4.1, calcula-se o erro total médio das 400 amostras iniciais de treinamento, propaga-se este valor retroativamente, da camada de saída até a camada de entrada, atualizando os pesos sinápticos W_1 e W_2 . Com os novos pesos sinápticos, um novo \hat{y} é calculado e o processo é repetido até que se atinja um número máximo de iterações ou um critério de convergência pré estipulado.

Uma vez atingido o critério de parada, tem-se uma função $f(x)$ que, ao receber o vetor dos preços das opções x de um determinado dia, retorna a estimativa da variância realizada do índice EWZ dos próximos 30 dias corridos. A validação destes valores é realizada através da janela móvel de 30 dias, explicada anteriormente. E a o desempenho do modelo é medido através da comparação com o valor real, pelo coeficiente de determinação ajustado (R^2 ajustado), a ser discutido no capítulo de resultados.

Regressão via MLP em relação ao VRP

Esta segunda abordagem contém uma pequena diferença quando comparada à primeira, que é a variável dependente da regressão ser o VRP ao invés da RV. Tomemos o exemplo descrito acima, em que se escolhe 17 opções como informação de entrada. Nesta segunda abordagem, teremos os mesmos 17 dados das opções na camada de entrada e a única mudança será no vetor y , representado pelo VRP, como descrito no Item 4.7.1. Toda a etapa de treinamento do modelo será semelhante ao que foi descrito no item anterior, com esta única diferença na variável dependente. Porém na etapa de validação dos testes, os resultados do modelo serão subtraídos pelo valor do modelo VIX , e a comparação final será feita em relação à RV.

Este capítulo detalhou o processo de estimação da RV a partir de cinco modelos distintos começando pela coleta de dados de mercado, cujo passo a passo é resumido abaixo:

- Coleta dos preços de fechamento do índice EWZ e da suas opções e dos valores diários da taxa básica de juros norte americana.
- Definição dos parâmetros de entrada do usuário, sendo eles T a janela de previsão em dias corridos, ΔK o intervalo entre os preços de exercícios das opções, N o número de opções a serem utilizadas em cada data base e o cenário Z que corresponde ao filtro a ser aplicado nas datas disponíveis.
- Tratamento das informações de todas as opções.
- Seleção das opções a serem utilizadas com base no cenário Z escolhido, no número N de opções por data base e nos requisitos mínimos que devem ser atendidos em cada data base.
- Pré processamento dos dados que consiste na padronização dos preços *mid* das opções selecionadas.
- Cálculo diário da variância realizada nos próximos 30 dias corridos.
- Definição da abordagem de janela móvel para o treinamento e teste dos modelos RLM , RLM_{VRP} , MLP e MLP_{VRP} .
- Cálculo da RV em relação ao VIX.
- Processamentos dos modelos RLM e RLM_{VRP} .
- Processamentos dos modelos MLP e MLP_{VRP} .

O próximo passo será analisar os resultados obtidos em cada um dos cinco modelos e compará-los. Esta comparação será feita em relação ao desempenho dos modelos, o qual será medido pelo R^2 ajustado. Modelos com os maiores valores desta medida serão considerados superiores aos outros que obtiverem valores menores. As análises destes resultados, bem como as comparações de desempenho dos cinco modelos, serão feitas no capítulo a seguir.

5 RESULTADOS

Uma vez aplicada a metodologia de processamento dos cinco modelos apresentados neste trabalho, inicia-se a etapa de análise de resultados e de comparação entre os modelos. Antes de se iniciar as análises das saídas dos modelos, é importante definir a medida de desempenho a ser utilizada, a qual será o R^2 ajustado. Quanto maior for o seu valor, melhor será o desempenho do modelo que o detém. Dado que a interpretação do nível do R^2 ajustado pode variar entre diferentes áreas de estudo, neste trabalho considera-se que R^2 ajustados inferiores a 0 não são aceitáveis, de 0 a 0,4 são fracos, entre 0,4 e 0,7 são moderados e acima de 0,7 fortes ¹.

Com isso, primeiro será descrito como o coeficiente de penalização (λ) das RNA interferem no desempenho destes modelos e como isso foi tratado neste trabalho. Segundo, o desempenho dos cinco modelos aplicados serão comparados entre si para cada cenário pré determinado. Por fim serão feitas duas análises comparativas, sendo a primeira do modelo VIX para os quatro cenários disponíveis e a segunda para o modelo RNA_{VRP} também para os quatro cenários disponíveis. A escolha destes dois modelos para se fazer uma análise mais específica se deu pelo fato de eles serem os dois com melhores desempenho.

5.1 Otimização do coeficiente de penalização (λ)

A Tabela 3 mostra o R^2 ajustado das duas abordagens de RNA, o modelo RNA_{RV} e o modelo RNA_{VRP} , para a combinação entre a amostra inicial de 400 dias, o coeficiente de penalização $\lambda \in \{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 100, 1000\}$ e número de opções na camada de entrada $N \in 17, 19, 21, 23, 25, 27$ para os cenários 1 e 2. Os cenários 3 e 4 destas mesmas configurações são apresentados no A, assim como as demais configurações formadas pelas combinações entre o coeficiente de penalização $\lambda \in \{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 100, 1000\}$ e o número de opções $N \in 17, 19, 21, 23, 25$ com amostra inicial de 500 dias, o número de opções $N \in 17, 19, 21, 23$ com amostra inicial de 600 dias e o número de opções $N \in 17, 19, 21$ com amostra inicial de 700 dias para os cenários 1, 2, 3 e 4.

Nota-se que para λ inferiores a 1 o desempenho do modelo é ruim, uma vez que o R^2 ajustado é negativo para a todas das configurações. Quanto menor o λ , maior é a variância dos resultados pelo fato de o modelo criar fronteiras mais curvadas e que se adaptam melhor aos dados *in sample*, o que aumenta a probabilidade de se ter *overfitting*. Por outro lado, um λ maior diminui a variância dos resultados, aumenta a penalização dos pesos e combate o viés dos dados *in sample*, mas também aumenta a probabilidade de

¹ https://www.morningstar.com/InvGlossary/r_squared_definition_what_is.aspx

se ter um *underfitting*.

Tabela 3 – R^2 ajustado dos modelos RNA_{RV} e RNA_{VRP} para amostra inicial de 400 dias nos cenários 1 e 2.

Amostra inicial 400 dias					
# opções	λ	Cenário 1		Cenário 2	
		RNA_{VR}	RNA_{VRP}	RNA_{VR}	RNA_{VRP}
17	0.001	-38.895	-34.349	-479.637	-165.410
17	0.01	-44.573	-3.355	-49.664	-103.377
17	0.1	-5.382	-6.203	-4.480	-3.497
17	1	-0.158	0.135	-0.203	-0.251
17	10	-0.013	0.117	-0.014	0.089
17	100	-0.016	0.106	-0.032	0.108
17	1000	-0.014	0.112	-0.017	0.106
19	0.001	-132.521	-124.672	-8.762	-7.795
19	0.01	-46.042	-47.368	-8.552	-24.633
19	0.1	-8.359	-16.980	-3.408	-8.424
19	1	-1.713	-1.681	-0.499	0.202
19	10	-0.052	0.005	-0.022	0.133
19	100	-0.013	0.013	-0.039	0.120
19	1000	-0.013	0.025	-0.023	0.131
21	0.001	-45.203	-20.014	-61.286	-87.602
21	0.01	-17.922	-20.479	-41.957	-67.409
21	0.1	-61.551	-11.454	-17.928	-13.512
21	1	-0.202	-3.139	-0.301	-0.358
21	10	-0.031	0.171	-0.041	0.040
21	100	-0.021	0.165	-0.019	0.024
21	1000	-0.015	0.157	-0.016	0.015
23	0.001	-35.954	-80.550	-23.789	-19.727
23	0.01	-102.116	-80.743	-22.437	-9.596
23	0.1	-15.437	-7.781	-64.369	-21.705
23	1	-0.030	-1.704	-0.044	-0.039
23	10	-0.006	0.129	-0.039	0.132
23	100	-0.006	0.158	-0.022	0.121
23	1000	-0.023	0.138	-0.011	0.145
25	0.001	-185.114	-209.737	-158.465	-244.565
25	0.01	-204.913	-211.807	-228.222	-109.873
25	0.1	-18.541	-6.020	-5.049	-133.138

continua na próxima página

Amostra inicial 400 dias					
# opções	λ	Cenário 1		Cenário 2	
		RNA_{VR}	RNA_{VRP}	RNA_{VR}	RNA_{VRP}
25	1	-3.611	0.027	0.310	0.273
25	10	-0.024	0.160	-0.033	0.177
25	100	-0.038	0.163	-0.045	0.154
25	1000	0.000	0.189	-0.020	0.175
27	0.001	-31,315.623	-56,406.921	-1,923.022	-3,070.725
27	0.01	-94,745.300	-265,642.987	-1,206.610	-1,763.271
27	0.1	-2,243.125	-2,179.872	-209.932	-251.554
27	1	-284.897	-34.471	-5.773	-3.946
27	10	-3.009	-0.958	-1.337	-0.390
27	100	-2.804	-0.583	-1.210	-0.268
27	1000	-2.888	-0.810	-0.796	-0.128

Neste trabalho optou-se por otimizar esse λ deixando-o variável, de modo que para se comparar ambos modelos de RNA com os demais, seleciona-se sempre a configuração com o melhor desempenho, ou seja, nas próximas análises, o R^2 ajustado observado em cada modelo de RNA será o maior valor obtido entre todos os λ s possíveis para o modelo em questão sob as suas respectivas configurações de amostra inicial, número de opções N e cenário Z , como exemplificado através dos números destacados na tabela acima.

5.2 Análise de desempenho da estimação da variância realizada

Este item irá comparar o desempenho dos modelos RNA_{RV} e RNA_{VRP} com o desempenho do modelo *benchmark* VIX e os modelos de regressão linear RLM_{VR} e RLM_{VRP} nos diferentes cenários e sob as diferentes configurações de número de opções na camada de entrada e de tamanho da amostra de treinamento inicial.

5.2.1 Cenário 1 - Sem filtro de datas

O primeiro cenário de teste levou em conta todo o período disponível, que compreende de 6 de Janeiro de 2016 a 1 de Maio de 2020. A Tabela 4 compara a melhor configuração da RNA com os demais modelos já citados.

Tabela 4 – R^2 ajustado para os modelos VIX , RNA_{RV} , RNA_{VRP} , RLM_{VR} e RLM_{VRP} no cenário 1

Amostra Inicial (dias)	# opções	VIX	RNA_{VR}	RNA_{VRP}	RLM_{VR}	RLM_{VRP}
400	17	0.085	-0.013	0.135	-0.109	0.043
500	17	0.076	-0.017	0.122	-0.114	0.036
600	17	0.071	-0.024	0.118	-0.122	0.029
700	17	-0.018	0.002	0.019	-0.018	-0.015
400	19	0.002	-0.013	0.025	0.013	0.036
500	19	0.111	-0.012	0.133	0.001	0.152
600	19	0.105	-0.020	0.127	-0.006	0.146
700	19	0.102	-0.023	0.123	-0.008	0.144
400	21	0.138	-0.015	0.171	0.033	0.165
500	21	0.132	-0.020	0.165	0.035	0.163
600	21	0.125	-0.028	0.159	0.027	0.156
700	21	0.120	-0.033	0.155	0.023	0.152
400	23	0.128	-0.006	0.158	0.049	0.154
500	23	0.118	0.026	0.150	0.049	0.150
600	23	0.115	0.034	0.147	0.047	0.148
400	25	0.163	0.000	0.189	-0.190	0.116
500	25	0.157	-0.008	0.183	-0.198	0.111
400	27	-0.195	-2.804	-0.583	-44.927	-10.750

Os resultados apresentados na Tabela 4 mostram que os modelos possuem um desempenho fraco, dado que o R^2 é inferior a 0,4 em todos os casos. Os resultados sugerem que o modelo RNA_{RV} e RLM_{VR} não possuem um desempenho aceitável, ou seja, regredir o preço das opções diretamente em relação à variância realizada não demonstra ser uma boa abordagem com as configurações pré estabelecidas neste trabalho para este cenário. Porém nota-se que ao regredir o preço das opções em relação ao prêmio de risco de variância (VRP), abordagem representada pelos modelos RNA_{VRP} e RLM_{VRP} , tem-se um ganho de desempenho aceitável, uma vez que o R^2 ajustado passa a ser positivo e na maioria dos casos superior ao R^2 ajustado do modelo *benchmark*.

Era esperado que à medida que o número de opções na camada de entrada aumentasse, obtiver-se-ia um desempenho superior, porém os resultados não são conclusivos quando se olha apenas para este cenário, dado que a configuração com 23 opções apresenta um desempenho inferior ao da configuração com 21 opções. Isso pode representar um *outlier*, mas para que isso seja embasado, se faz necessário analisar outros cenários ou obter mais dados com mais alternativas de variação do número de opções na camada de entrada. É importante destacar que quando se utilizam 27 opções na camada de entrada o desempenho dos modelos passam a ser inaceitáveis por conta da diminuição expressiva do tamanho da amostra, de 1029 para 537, quando se muda de 17 para 27 opções.

5.2.2 Cenário 2 - Filtro *Joesley Day*

O segundo cenário de teste possui um filtro de dias próximos ao evento *Joesley Day*, como explicado anteriormente. Neste cenário, as datas inicial e final são as mesmas do cenário 1, porém desconsidera-se o intervalo entre os dias 18/04/2017 e 22/05/2017. Os resultados dos modelos são apresentados na Tabela 5.

Tabela 5 – R^2 ajustado para os modelos VIX , RNA_{RV} , RNA_{VRP} , RLM_{VR} e RLM_{VRP} no cenário 2

Amostra Inicial (dias)	# opções	VIX	RNA_{RV}	RNA_{VRP}	RLM_{VR}	RLM_{VRP}
400	17	0.082	-0.014	0.108	-0.144	0.031
500	17	0.076	-0.020	0.101	-0.149	0.026
600	17	0.068	-0.028	0.093	-0.160	0.017
700	17	0.064	-0.033	0.088	-0.164	0.012
400	19	0.115	-0.022	0.202	-0.105	0.095
500	19	-0.003	-0.009	0.035	-0.028	-0.005
600	19	0.103	-0.023	0.217	-0.062	0.122
700	19	0.098	-0.028	0.219	-0.066	0.117
400	21	0.006	-0.016	0.040	-0.033	-0.003
500	21	0.129	-0.009	0.155	-0.072	0.112
600	21	0.125	-0.015	0.150	-0.080	0.106
700	21	0.115	-0.025	0.189	-0.090	0.097
400	23	0.125	-0.011	0.145	-0.028	0.111
500	23	0.118	-0.019	0.137	-0.036	0.105
600	23	0.114	-0.023	0.133	-0.040	0.101
400	25	0.160	0.310	0.273	-0.147	0.134
500	25	0.157	0.384	0.175	-0.150	0.132
400	27	-0.141	-0.796	-0.128	-1.397	-0.488

Os resultados deste cenário se assemelham com os resultados obtidos no cenário anterior, onde o desempenho dos modelos, quando aceitáveis, são fracos. Nota-se o mesmo padrão de ganho de desempenho quando se regride o preço das opções em relação ao VRP ao invés da variância realizada diretamente. Para o cenário com filtro *Joesley Day*, a mudança ocorre apenas nos dados de treinamento, uma vez que a janela de datas removida fica dentro da amostra inicial mínima de 400 dias dentro da configuração com 17 opções na camada de entrada, e por consequência dentro das demais também. Para 25 opções, o desempenho do modelo RNA é superior aos demais, o que sugere ser um *outlier* por conta da ausência de desempenhos assim positivos em outras configurações. Vale ressaltar que a configuração com 27 opções continua apresentando um desempenho não aceitável pelos mesmos motivos já citados.

5.2.3 Cenário 3 - Filtro Covid

O terceiro cenário possui o mesmo início do cenário 1, 06/01/2016, porém a janela de teste vai até 10/02/2020. Os resultados dos modelos são apresentados na Tabela 6.

Tabela 6 – R^2 ajustado para os modelos VIX , RNA_{RV} , RNA_{VRP} , RLM_{VR} e RLM_{VRP} no cenário 3

Amostra Inicial (dias)	# opções	VIX	RNA_{VR}	RNA_{VRP}	RLM_{VR}	RLM_{VRP}
400	17	0.135	-0.845	0.030	-1.133	-0.133
500	17	0.075	-0.477	0.103	-0.616	0.067
600	17	0.294	-0.507	0.240	-0.655	0.110
700	17	-0.094	-1.996	-0.481	-1.546	-0.319
400	19	0.146	-1.068	0.031	-1.045	-0.134
500	19	0.163	-0.354	0.083	-0.508	0.101
600	19	0.334	-0.219	0.128	-0.440	0.176
700	19	-0.086	-1.564	-0.840	-1.439	-0.326
400	21	0.123	-0.909	-0.002	-1.122	-0.211
500	21	0.187	-0.697	0.064	-0.429	0.114
600	21	0.332	-0.478	0.143	-0.362	0.184
700	21	-0.113	-2.091	-0.667	-1.487	-0.376
400	23	0.181	-1.004	0.005	-1.525	-0.415
500	23	0.188	-0.510	0.075	-0.688	-0.034
600	23	0.094	-1.318	-0.191	-1.457	-0.317
400	25	0.190	-0.635	-0.041	-5.690	-1.162
500	25	0.347	-0.680	0.134	-6.985	-1.262
400	27	-0.305	-2.205	-0.630	-35.407	-5.293

Os resultados do cenário com filtro *Covid* divergem dos outros dois primeiros. Nota-se um desempenho superior do modelo *benchmark* frente às outras alternativas e também frente ao seu próprio desempenho nos outros cenários. O primeiro sinal mostrado por este cenário é de que a expectativa do mercado para a variância realizada se aproxima da realidade quando não se tem momentos de crise, como esperado. Tal fato poderá ser corroborado durante o teste do cenário 4, baseado na situação em que não se tem ambos os impactos do *Joesley Day* e do *Covid*. Neste cenário 3, as regressões feitas diretamente em relação à variância realizada, tanto pelas RNA quanto pelas RLM, continuaram a apresentar desempenhos não aceitáveis, reforçando a tese de que estes modelos não funcionam para as configurações aqui estabelecidas.

Já as regressões em relação ao VRP apresentaram piora em relação aos dois últimos cenários. A princípio tal fato não era esperado uma vez que se tenha removido uma janela de crise do conjunto dos dados de treinamento e teste. Porém tal fato pode ter sido impactado por dois motivos. O primeiro, como no cenário de filtro *Joesley Day*, é a diminuição do conjunto de dados para treinamento e teste do modelo. E o segundo pela

presença do intervalo de datas da crise *Joesley Day* na amostra inicial de treinamento, que pode ter enviesado os modelos de regressão a um treinamento com período de crise enquanto não se tinha nenhum nos dados de teste. Tal fato pode ser corroborado pelo aumento do desempenho dos modelos RNA_{VRP} e RLM_{VRP} à medida que se aumenta o tamanho da amostra inicial, uma vez que, neste caso, quanto maior o seu tamanho, menor é o impacto da crise *Joesley Day* no treinamento dos modelos. É válido destacar que a exceção ocorre em algumas configurações em que a amostra inicial é máxima, como nas combinações entre amostra inicial e número de opções 700 e 17, 700 e 19, 700 e 21 e 600 e 23, devido ao fato de se ter um número muito reduzido na amostra de teste dos modelos, o que pode distorcer o desempenho, tanto positiva quanto negativamente. Esta mesma exceção também impacta o modelo *VIX* negativamente, o que neste caso sugere que a amostra reduzida de treinamento contém informações de um período em que o mercado não apresentou um comportamento esperado pelos investidores. Nota-se também que o desempenho da configuração com 27 opções se mantém num patamar não aceitável pelos mesmos motivos já citados.

5.2.4 Cenário 4 - Filtros *Joesley Day* e *Covid*

O quarto cenário é a combinação dos cenários 2 e 3, onde tem-se o período do dia 06/01/2016 até 10/02/2020, com a remoção da janela de dias entre as datas 18/04/2017 e 22/05/2017. Os resultados dos modelos para este cenário são apresentados na Tabela 7.

Este último cenário de análise é desprovido de informações de crise, tanto nos dados de treinamento quanto nos dados de teste. Tal fato permite a análise dos modelos em uma situação neutra de acontecimentos atípicos em que o mercado se comporta de maneira totalmente inesperada com uma mudança abrupta. Neste cenário, é possível corroborar algumas das afirmações feitas nas análises anteriores, como o fato dos modelos RNA_{RV} e RLM_{VR} não apresentarem um desempenho aceitável para as configurações estabelecidas neste trabalho. Com a remoção das duas janelas de crise citadas acima, a amostra total atinge o menor tamanho dentre todos os cenários, o que mantém o desempenho inaceitável das combinações entre amostra inicial e número de opções de 700 e 17, 700 e 19, 700 e 21 e 600 e 23, reduz o desempenho da combinação 500 e 25 para o nível de não aceitável para o modelo *VIX*.

Tabela 7 – R^2 ajustado para os modelos VIX , RNA_{RV} , RNA_{VRP} , RLM_{VR} e RLM_{VRP} no cenário 4

Amostra Inicial (dias)	# opções	VIX	RNA_{VR}	RNA_{VRP}	RLM_{VR}	RLM_{VRP}
400	17	0.074	-0.386	0.259	-0.565	0.051
500	17	0.157	-0.312	0.274	-0.212	0.278
600	17	0.307	-0.293	0.411	-0.248	0.318
700	17	-0.111	-1.448	-0.054	-0.898	-0.146
400	19	0.107	-0.248	0.191	-0.555	0.050
500	19	0.176	-0.305	0.215	-0.190	0.271
600	19	0.360	-0.230	0.402	-0.213	0.327
700	19	-0.109	-1.229	-0.188	-0.845	-0.138
400	21	0.122	-0.331	0.216	-0.743	-0.046
500	21	0.170	-0.055	0.394	-0.206	0.177
600	21	0.388	-0.056	0.410	-0.267	0.243
700	21	-0.090	-1.020	-0.095	-0.877	-0.205
400	23	0.142	-0.119	0.268	-0.352	0.043
500	23	0.321	-0.179	0.325	-0.304	0.191
600	23	-0.114	-1.216	-0.186	-1.120	-0.338
400	25	0.180	-0.182	0.197	-3.620	-0.257
500	25	-0.001	-0.878	0.062	-8.003	-0.756
400	27	-0.243	-0.694	-0.214	-0.671	-0.300

Neste cenário aparecem pela primeira vez desempenhos de nível moderado, quando o R^2 ajustado é superior a 0,4, obtidos através do modelo RNA_{VRP} . Este modelo apresenta desempenho superior em relação aos outros em quase todas as combinações em que se tem níveis aceitáveis de desempenho. Este fato descarta a hipótese, levantada na análise dos cenários 2 e 3, de que a remoção de alguma das janelas de crise diminui o tamanho da amostra a ponto de se ter poucos dados e invalidar o uso dos modelos de regressão sob as circunstâncias apresentadas neste trabalho. Ao mesmo tempo reafirma a hipótese de relação entre os períodos de crise e como estes períodos impactam as amostras de treinamento e teste, a ser discutido a seguir. A configuração com 27 opções se mantém não aceitável, em linha com todos os cenários anteriores.

5.3 Comparativo entre os cenários de teste para os modelos VIX e RNA_{VRP}

Conforme observado no Item 5.2 e na Tabela 8, os modelos que mais apresentaram desempenhos aceitáveis foram o VIX e o RNA_{VRP} , os quais terão suas variações entre cenários discutidas neste item.

Tabela 8 – Modelos com maior R^2 ajustado para cada cenário e para cada configuração de número de opções por dia utilizadas.

# opções	Cenário 1		Cenário 2		Cenário 3		Cenário 4	
	Modelo	R^2	Modelo	R^2	Modelo	R^2	Modelo	R^2
17	RNA _{VRP}	0.135	RNA _{VRP}	0.108	VIX	0.294	RNA _{VRP}	0.411
19	RNA _{VRP}	0.152	RNA _{VRP}	0.219	VIX	0.334	RNA _{VRP}	0.402
21	RNA _{VRP}	0.171	RNA _{VRP}	0.189	VIX	0.332	RNA _{VRP}	0.410
23	RNA _{VRP}	0.158	RNA _{VRP}	0.145	VIX	0.188	RNA _{VRP}	0.325
25	RNA _{VRP}	0.189	RNA _{VR}	0.384	VIX	0.347	RNA _{VRP}	0.197
27	VIX	-0.195	RNA _{VRP}	-0.128	VIX	-0.305	RNA _{VRP}	-0.214

A Tabela 8 apresenta a seleção dos modelos que apresentaram melhor desempenho para cada configuração de número de opções por dia utilizadas, independente do número de dias da amostra inicial. Outra informação relevante para a análise dos modelos é a relação entre o aumento do número de opções diárias utilizadas na camada de entrada e a redução da quantidade de dias disponíveis para o processamento dos modelos, tanto para o cálculo do modelo *VIX* quanto para as etapas de treinamento e teste da RNA. Tal informação está representada na Tabela 9, bem como o percentual médio de diminuição de dias disponíveis na amostra total para cada aumento no número de opções diárias utilizadas.

Tabela 9 – Dias disponíveis para cada quantidade de opções diárias utilizadas e diminuição média dos dias disponíveis a cada aumento no número de opções.

# opções	Cenário 1	Cenário 2	Cenário 3	Cenário 4	Diminuição média
17	1,029	1,004	1,000	975	-
19	1,010	985	983	958	-1.8%
21	972	947	946	921	-3.8%
23	884	859	861	836	-9.1%
25	765	740	745	720	-13.7%
27	537	515	520	498	-30.3%

5.3.1 Análise comparativa do modelo *VIX* para os 4 cenários disponíveis

O modelo *VIX* é utilizado como baliza de desempenho neste trabalho pelos motivos já citados anteriormente. Portanto é necessário entender como é seu desempenho nos 4 cenários aqui analisados. A Tabela 10 resume o resultado deste modelo considerando a melhor configuração de amostra inicial.

Tabela 10 – Maiores R^2 ajustados do modelo VIX para cada cenário e para cada configuração de número de opções por dia utilizadas.

# opções	VIX			
	Cenário 1	Cenário 2	Cenário 3	Cenário 4
17	0.085	0.082	0.294	0.307
19	0.111	0.115	0.334	0.360
21	0.138	0.129	0.332	0.388
23	0.128	0.125	0.188	0.321
25	0.163	0.160	0.347	0.180
27	-0.195	-0.141	-0.305	-0.243

Percebe-se que os cenários 1 e 2 apresentam desempenhos muito próximos, o que de fato é esperado uma vez que a amostra *out of sample* é muito semelhante. O que difere um cenário do outro é a remoção da janela de dias da crise *Joesley Day*, que está contida na amostra inicial. Como o modelo VIX não é treinado como as regressões e o seu cálculo é baseado no preço de fechamento das opções de apenas um dia, o impacto da mudança do cenário 1 para o cenário 2 é mínimo. Já o impacto da mudança do cenário 2 para o cenário 3 parece ser evidente. Apesar de parecer claro o impacto de uma crise no modelo *benchmark* quando se retira o período do *Covid* da amostra de teste, e o seu R^2 ajustado é dobrado nesta mudança, deve-se atentar à possibilidade de criação de viés por conta da limitação de dados disponíveis. Do cenário 3 para o cenário 4 não eram esperadas mudanças significativas, porém há uma queda de desempenho no cenário 4 quando se calcula o modelo VIX com 25 opções, possivelmente por conta da diminuição significativa da amostra sendo testada, o que pode distorcer o R^2 ajustado tanto positiva quanto negativamente.

É importante ressaltar que, para o modelo VIX , quanto maior o tamanho da amostra inicial de treinamento das regressões, mais enviesado pode estar o seu R^2 ajustado, uma vez que estas informações da amostra inicial, dados *in sample*, não são computadas em seu cálculo. Para analisar este impacto é necessário observar a evolução do R^2 ajustado tanto para a variação do número de opções utilizadas quanto para a variação da amostra inicial, apresentado na Tabela 11.

Tabela 11 – R^2 ajustado do modelo VIX para cada cenário e para cada combinação de amostra inicial e de número de opções por dia utilizadas.

VIX					
Amostra Inicial (dias)	# opções	Cenário 1	Cenário 2	Cenário 3	Cenário 4
400	17	0.085	0.082	0.135	0.074
400	19	0.002	0.115	0.146	0.107
400	21	0.138	0.006	0.123	0.122
400	23	0.128	0.125	0.181	0.142
400	25	0.163	0.160	0.190	0.180
400	27	-0.195	-0.141	-0.305	-0.243
500	17	0.076	0.076	0.075	0.157
500	19	0.111	-0.003	0.163	0.176
500	21	0.132	0.129	0.187	0.170
500	23	0.118	0.118	0.188	0.321
500	25	0.157	0.157	0.347	-0.001
600	17	0.071	0.068	0.294	0.307
600	19	0.105	0.103	0.334	0.360
600	21	0.125	0.125	0.332	0.388
600	23	0.115	0.114	0.094	-0.114
700	17	-0.018	0.064	-0.094	-0.111
700	19	0.102	0.098	-0.086	-0.109
700	21	0.120	0.115	-0.113	-0.090

Como já mencionado, existe uma compensação entre aumentar o número de opções utilizadas no cálculo do modelo e a quantidade de dados para este mesmo cálculo, ou seja, quando se aumentam as informações disponíveis por dia, diminuem-se as informações no conjunto total dos dados. Com base nos valores apresentados na Tabela 11, evidencia-se que aumentar o número de opções por dia melhora o desempenho do modelo até o ponto em que a perda de informações totais supera os ganhos da primeira. Para a amostra inicial de 400 dias, a vantagem de se aumentar a quantidade de opções por dia vai até o número 25, depois disso a perda de eficiência fica evidente em todos os cenários. Um efeito de distorção semelhante ocorre quando se tem uma amostra inicial de 500 dias e se aumenta o número de opções por dia de 21 para 23. No cenário 3 o desempenho permanece próximo e no cenário 4 há um aumento significativo do mesmo, o que pode ser considerado um aparecimento de viés, o qual é corroborado no cenário 3 quando se tem o mesmo tamanho de amostra inicial e aumenta-se a quantidade de opções de 23 para 25. Quando se trabalha com uma amostra inicial de 600 dias, poderíamos concluir que há um aumento expressivo no desempenho do modelo na passagem do cenário 2 para o cenário 3, porém isto não seria correto devido ao desempenho observado nas configurações de 400 e 500 dias de amostra inicial ser muito divergente, como já explicado. Os resultados sugerem que a diminuição de dias utilizados no modelo VIX cria um viés positivo nos cenários 3 e 4 para uma amostra inicial de 600 dias, e um viés negativo nos mesmos cenários mas para uma amostra inicial

de 700 dias.

Fica evidente que há uma grande sensibilidade em se variar o número de opções por dia e o tamanho da amostra inicial por conta da alta probabilidade de se criar viés nos resultados. Tal problema ocorre pela quantidade limitada de dados utilizada neste trabalho.

5.3.2 Análise comparativa do modelo RNA_{VRP} para os 4 cenários disponíveis

Com as análises do modelo VIX feitas, partimos para o detalhamento do modelo RNA_{VRP} , objeto principal deste trabalho. A Tabela 12 evidencia alguns pontos já discutidos anteriormente, como o descarte dos resultados gerados quando se usam 27 opções na camada de entrada, ocasionado pelo fato de haver uma diminuição relevante da quantidade de dias disponíveis, cerca de 30% para todos os cenários de acordo com a Tabela 9, para o processamento do modelo, que acaba produzindo resultados enviesados.

Outro ponto evidenciado pela Tabela 12 é o aparente aumento do desempenho à medida que se aumentam as opções utilizadas por dia nos cenários 1 e 2, o que não é verificado nos cenários 3 e 4. Tal fato deve ser analisado com mais detalhes para que esta afirmação seja corroborada. Embora o modelo RNA_{VRP} seja diferente do modelo VIX no que diz respeito à utilização da amostra inicial, uma vez que o primeiro utiliza estas informações desde o início até o último treinamento da rede enquanto o segundo simplesmente a descarta, é importante levarmos em conta a quantidade de dias utilizada no início do treinamento do modelo para inferir determinadas afirmações. A Tabela 13 trás a evolução do R^2 ajustado do modelo RNA_{VRP} para as diferentes combinações entre os dias na amostra inicial e a quantidade de opções utilizadas na camada de entrada da RNA_{VRP} para os diferentes cenários.

Tabela 12 – Maiores R^2 ajustados do modelo RNA_{VRP} para cada cenário e para cada configuração de número de opções na camada de entrada.

RNA_{VRP}				
# opções	Cenário 1	Cenário 2	Cenário 3	Cenário 4
17	0.135	0.108	0.240	0.411
19	0.133	0.219	0.128	0.402
21	0.171	0.189	0.143	0.410
23	0.158	0.145	0.075	0.325
25	0.189	0.273	0.134	0.197
27	-0.583	-0.128	-0.630	-0.214

Observa-se na Tabela 13 que, embora existam algumas combinações de amostra inicial e de número de opções nos cenários 1 e 2 que fogem do padrão das demais, assim como no modelo VIX , é possível inferir que à medida que se aumentam as opções utilizadas, tem-se um desempenho mais elevado do modelo nos cenários 1 e 2. Na mudança do cenário

1 para o cenário 2, por um lado era esperado um ganho de desempenho pelo fato de se remover um período da amostra de treinamento em que o mercado financeiro sai de sua normalidade e apresenta distorções inesperadas e abruptas. Porém, por outro lado poderia haver uma correlação entre este período de mudança abrupta do *Joesley Day* e o período de início do *Covid*, a qual poderia ser capturada pela RNA, e ao invés de se produzir melhores resultados no cenário 2, o cenário 1, por conter ambos períodos de crise, poderia apresentar um melhor desempenho. Mas com base na Tabela 13, não se nota evidência suficiente para afirmar que o modelo possui um melhor desempenho no cenário 1 ao invés do 2, o que também vale para o contrário.

Assim como poderia ocorrer um aumento no desempenho do modelo entre os cenários 1 e 2, também havia a possibilidade de melhora entre os cenários 1 e 3, pelo fato de este último passar a desconsiderar um período de crise, onde o mercado se comporta de maneira inesperada. Mas por outro lado, o cenário 3 possui informações de uma crise em todas as suas etapas de treinamento, referente ao período *Joesley Day*, o que pode levar a remoção do período *Covid* a ter um impacto negativo, visto que a RNA utiliza informações de crise em todo treinamento mas não possui nenhuma outra em sua etapa de teste. Vale destacar que o fato de não existir evidências de que a remoção do período *Joesley Day* dos dados de treinamento piora o desempenho do modelo quando se tem o período *Covid*, também não serve como evidência do contrário. Ou seja, é possível que a remoção do período *Covid* sem que se remova o período *Joesley Day* piore o desempenho do modelo. Ao cruzar os resultados apresentados nas Tabelas 11 e 13, podemos notar que de fato há um decréscimo significativo no desempenho do modelo, evidenciado principalmente pelas configurações de amostra inicial 400 e 500. Embora se tenha um R^2 ajustado superior aos dos outros cenários na configuração de amostra inicial 600 e número de opções igual a 17, é possível inferir que este caso é um *outlier* a partir da combinação dos resultados do modelo *VIX* com o modelo *RNA_{VRP}* para as configurações de amostra inicial 600 e número de opções 19 e 21. A evidência neste caso de amostra inicial 600 é de que qualquer viés *out of sample* capturado utilizando 19 e 21 opções também deveria ser capturado ao se utilizar 17.

Tabela 13 – R^2 ajustado do modelo RNA_{VRP} para cada cenário e para cada combinação de amostra inicial e de número de opções utilizadas na camada de entrada da RNA.

RNA _{VRP}					
Amostra Inicial (dias)	# opções	Cenário 1	Cenário 2	Cenário 3	Cenário 4
400	17	0.135	0.108	0.030	0.259
400	19	0.025	0.202	0.031	0.191
400	21	0.171	0.040	-0.002	0.216
400	23	0.158	0.145	0.005	0.268
400	25	0.189	0.273	-0.041	0.197
400	27	-0.583	-0.128	-0.630	-0.214
500	17	0.122	0.101	0.103	0.274
500	19	0.133	0.035	0.083	0.215
500	21	0.165	0.155	0.064	0.394
500	23	0.150	0.137	0.075	0.325
500	25	0.183	0.175	0.134	0.062
600	17	0.118	0.093	0.240	0.411
600	19	0.127	0.217	0.128	0.402
600	21	0.159	0.150	0.143	0.410
600	23	0.147	0.133	-0.191	-0.186
700	17	0.019	0.088	-0.481	-0.054
700	19	0.123	0.219	-0.840	-0.188
700	21	0.155	0.189	-0.667	-0.095

Por fim, o cenário 4 é o primeiro a conter desempenhos moderados para o modelo RNA_{VRP} . Como destacado anteriormente, é possível que se tenha viés positivo nas amostras *out of sample* para as configurações com amostra inicial de 600 dias no cenário 4, porém o desempenho elevado deste modelo, quando comparado ao modelo *benchmark* e aos outros cenários, nas configurações com amostra inicial de 400 e 500 dias corrobora a expectativa inicial de que quando não há indícios de crises nas amostras de treinamento e teste, o modelo tende a apresentar um bom desempenho. É importante destacar a superioridade do modelo baseado em RNA e o ganho em se regredir os preços das opções em relação ao prêmio de risco de variância (VRP) ao invés da variância diretamente. Esta alternativa de se obter a variância realizada através da regressão da diferença entre o índice VIX elevado ao quadrado, VIX^2 , e a variância realizada pode ser entendida como um procedimento de normalização, cujo efeito geralmente aumenta o desempenho de algoritmos em modelos não lineares.

6 CONCLUSÃO

O objetivo deste trabalho foi implementar dois modelos de Redes Neurais Artificiais, sendo uma regressão dos preços de fechamentos de opções do índice EWZ em relação à variância realizada dos retornos diários do índice EWZ, modelo RNA_{VR} , e uma regressão em relação ao prêmio de risco de variância (VRP), modelo RNA_{VIX} , composto pela diferença entre o quadrado do índice VIX e a variância realizada (RV), do índice EWZ, e compará-los com o modelo *benchmark* representado pelo quadrado do índice VIX, chamado aqui de modelo VIX , e com dois modelos de regressão linear múltipla (RLM), sendo o primeiro uma regressão dos preços de fechamentos de opções do índice EWZ em relação à RV dos retornos diários do índice EWZ, modelo RLM_{VR} , e o segundo em relação ao VRP, modelo RLM_{VIX} . Optou-se pelo uso do índice EWZ como ativo base deste trabalho por 3 motivos. O primeiro é o maior volume de negociação das opções do índice EWZ quando comparadas com o as opções do índice IBOVESPA, o segundo é o fato dos vencimentos das opções do EWZ serem semanais frente aos vencimentos mensais das opções do IBOVESPA e o terceiro é fato do índice EWZ ser um ativo líquido negociado em bolsa de valores estrangeira que representa bem o mercado brasileiro.

Foram testados 4 cenários distintos, o primeiro de 06/01/2016 a 01/04/2020, o segundo com a mesma data de início e fim porém com a remoção do período compreendido entre 18/04/2017 e 22/05/2017, evento chamado de *Joesley Day*, o terceiro com a mesma data de início do primeiro porém com término em 10/02/2020, representando a remoção do período chamado neste trabalho de *Covid*, e o quarto período sendo a combinação dos segundo e terceiro cenários. Para o primeiro, segundo e quarto cenários, o modelo RNA_{VIX} possui o melhor desempenho entre os modelos testados, enquanto que para o terceiro cenário o modelo *benchmark VIX* é o que detém desempenho mais elevado. Embora o modelo RLM_{VIX} tenha desempenho aceitável, não foi destaque em nenhum dos cenários apresentados, corroborando a tese de que a regressão via RNA supera a RLM. É valido destacar que quando não há presença de períodos de crise na amostra de treinamento e teste os modelos apresentam um desempenho superior, como esperado.

Os modelos de regressão em relação à variância realizada, RNA_{VR} e RLM_{VR} , não apresentaram um desempenho mínimo aceitável para as configurações testadas neste trabalho, o que não deve ser entendido como modelos não aplicáveis para o problema em questão, mas sim como modelos que necessitam de uma amostra de dados mais robusta, com mais opções diárias disponíveis e uma janela de dados mais ampla. A utilização do quadrado do VIX foi determinante para se ter uma melhora no desempenho das regressões, uma vez que quando se regride em relação ao VRP tem-se o VIX exercendo um papel de

normalizador, cujo efeito geralmente é positivo, principalmente em modelos não lineares. De acordo com Bandi, Russell e Yang (2008), os modelos de previsão de RVol desempenham com maior acurácia quando são abastecidos com as IV, e podemos utilizar esta afirmação de forma análoga para corroborar esse aumento de acurácia no modelo aqui apresentado.

Embora se tenha atingido uma melhora geral em relação ao modelo *benchmark*, alguns pontos discutidos foram inconclusivos, principalmente pela limitação do dados utilizados neste trabalho. Para uma análise mais robusta e precisa, seria fundamental possuir tanto uma quantidade maior de opções, como 30, 40, 50 e até 60 opções por dia, quanto uma janela mais ampla de preços de fechamento de mercado. Devido ao acesso limitado a estas informações, algumas análises ficaram pendentes, portanto uma possível primeira extensão deste trabalho seria uma coleta maior de dados para que tais pendências fossem analisadas e concluídas.

Outra abordagem para a extensão deste trabalho seria a utilização de outras informações no modelo da RNA, como a inclusão de dados defasados de variância realizada, dados defasados de preços de opções, indicadores técnicos do ativo base ou até mesmo outros indicadores de mercado a princípio exógenos ao índice EWZ. Um trabalho a ser utilizado como exemplo é o de Kristjanpoller e Minutolo (2018), onde os autores aplicam o PCA em alguns indicadores técnicos do ativo base e transformam essas novas informações em dados de entrada da RNA. Como uma outra abordagem alternativa usando RNA, existe a possibilidade de geração de sinais ao invés de se estimar a variância a ser realizada, como por exemplo a previsibilidade de casos em a variância realizada é superior ou inferior a um patamar pré determinado da variância obtida através da volatilidade implícita do VIX.

REFERÊNCIAS

- ANDERSEN, T.; BOLLERSLEV, T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, v. 39, n. 4, p. 885–905, 1998.
- ANDERSEN, T. G.; FREDERIKSEN, P.; STAAL, A. D. The information content of realized volatility forecasts. 2007.
- BAHRA, B. Implied risk-neutral probability density functions from option prices: theory and application. Bank of England Working Paper, 1997.
- BAI, H.; ZHANG, X.; ZHOU, H. Option return predictability and variance risk premium. *Available at SSRN 3414841*, 2019.
- BAKSHI, G.; MADAN, D. Spanning and derivative-security valuation. *Journal of Financial Economics*, v. 55, n. 2, p. 205 – 238, 2000.
- BAMS, D.; BLANCHARD, G.; LEHNERT, T. Volatility measures and value-at-risk. *International Journal of Forecasting*, Elsevier, v. 33, n. 4, p. 848–863, 2017.
- BANDI, F. M.; RUSSELL, J. R.; YANG, C. Realized volatility forecasting and option pricing. *Journal of Econometrics*, v. 147, n. 1, p. 34–46, 2008.
- BAUWENS, L.; HAFNER, C.; LAURENT, S. *Handbook of Volatility Models and Their Applications*. [S.l.]: John Wiley Sons, Inc, 2012.
- BAYER, C.; FRIZ, P.; GATHERAL, J. Pricing under rough volatility. *Quantitative Finance*, v. 16, n. 6, p. 887–904, 2016. Disponível em: <<https://EconPapers.repec.org/RePEc:taf:quantf:v:16:y:2016:i:6:p:887-904>>.
- BEDENDO, M.; HODGES, S. D. The dynamics of the volatility skew: A kalman filter approach. *Journal of Banking Finance*, v. 33, n. 6, p. 1156–1165, 2009.
- BESADA, F. L. *Prêmio de risco de volatilidade (VRP) no mercado brasileiro*. Dissertação (Mestrado) — Escola Brasileira de Economia e Finanças da Fundação Getúlio Vargas, <https://hdl.handle.net/10438/27761>, 2019.
- BLACK, F.; SCHOLES, M. The pricing of options and corporate liabilities. *Journal of political economy*, The University of Chicago Press, v. 81, n. 3, p. 637–654, 1973.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, v. 31, n. 3, p. 307–327, 1986.
- BRAGA, A. d. P.; LUDERMIR, T. B.; CARVALHO, A. C. P. d. L. F. *Redes Neurais Artificiais: Teoria e Aplicações*. Brasil: LTC, 2000.
- BUSCH, T.; CHRISTENSEN, B. J.; NIELSEN, M. O. The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, v. 160, n. 1, p. 48–67, 2007.

- CARVALHO, M. C. et al. Modeling and forecasting the volatility of brazilian asset returns: a realized variance approach. *Revista Brasileira de Finanças*, Sociedade Brasileira de Finanças, v. 4, n. 1, p. 321–343, 2006.
- CYBENKO, G. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, Springer, v. 2, n. 4, p. 303–314, 1989.
- DANIELSSON, J. Stochastic volatility in asset prices. estimation with simulated maximum likelihood. *Journal of Econometrics*, v. 64, p. 375–400, 1994.
- DEMETERFI, K. et al. A guide to volatility and variance swaps. *The Journal of Derivatives*, Institutional Investor Journals Umbrella, v. 6, n. 4, p. 9–32, 1999.
- DEMUTH, H. B. et al. *Neural Network Design*. 2nd. ed. Stillwater, OK, USA: Martin Hagan, 2014. ISBN 0971732116.
- DERMAN, E.; KANI, I. The volatility smile and its implied tree. *Goldman Sachs Quantitative Strategies Research Notes*, 1994.
- DONALDSON, R. G.; KAMSTRA, M. An artificial neural network-garch model for international stock return volatility. *Journal of Empirical Finance*, Elsevier, v. 4, n. 1, p. 17–46, 1997.
- DURBIN, J.; KOOPMAN, S. J. Monte carlo maximum likelihood estimation for non-gaussian state space models. *Biometrika*, v. 84, n. 3, p. 669–684, 1997.
- ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, v. 50, n. 4, p. 987–1007, July 1982.
- ENGLE, R. F.; LEE, G. G. A permanent and transitory component model of stock return volatility. 1993.
- ERAKER, B. The volatility premium. *Manuscript*, Citeseer, 2008.
- GALDI, F. C.; PEREIRA, L. M. Valor em risco (var) utilizando modelos de previsão de volatilidade: Ewma, garch e volatilidade estocástica. *BBR-Brazilian Business Review*, FUCAPE Business School, v. 4, n. 1, p. 74–95, 2007.
- GATHERAL, J.; JAISSON, T.; ROSENBAUM, M. *Volatility is rough*. [S.l.], 2014.
- HAMID, S. A. Primer on using neural networks for forecasting market variables. 2004.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning: data mining, inference, and prediction*. [S.l.]: Springer Science & Business Media, 2009.
- HAYKIN, S. *Neural Networks: A Comprehensive Foundation (3rd Edition)*. EUA: Prentice-Hall, Inc., 2007. ISBN 0131471392.
- HEIJ, C. et al. *Econometric methods with applications in business and economics*. 1. ed. [S.l.]: Oxford University Press, 2004. ISBN 0199268010,9780199268016,0191533238,9780191533235.
- HESTON, S. L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, v. 6, n. 2, p. 327–343, 1993.

- HORVATH, B.; MUGURUZA, A.; TOMAS, M. Deep learning volatility. *Available at SSRN 3322085*, 2019.
- HULL, J. C. *Options, Futures and Other Derivatives*. [S.l.: s.n.], 2005.
- HUTCHINSON, J. M.; LO, A. W.; POGGIO, T. A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance*, Wiley Online Library, v. 49, n. 3, p. 851–889, 1994.
- ITô, K. Stochastic integral. *Proceedings of the Imperial Academy*, The Japan Academy, v. 20, n. 8, p. 519–524, 1944. Disponível em: <<https://doi.org/10.3792/pia/1195572786>>.
- JEFF, F. The quality of market volatility forecasts implied by sp 100 index option prices. *Journal of Empirical Finance*, v. 5, n. 4, p. 317–345, 1998.
- KRISTJANPOLLER, W.; MINUTOLO, M. C. Gold price volatility: A forecasting approach using the artificial neural network–garch model. *Expert Systems with Applications*, v. 42, n. 20, p. 7245 – 7251, 2015. ISSN 0957-4174.
- KRISTJANPOLLER, W.; MINUTOLO, M. C. A hybrid volatility forecasting framework integrating garch, artificial neural network, technical analysis and principal components analysis. *Expert Systems with Applications*, v. 109, p. 1 – 11, 2018. ISSN 0957-4174. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S095741741830294X>>.
- LAW, J.; SMULLEN, J. *A dictionary of finance and banking*. [S.l.]: Oxford University Press, 2008.
- MERTON, R. C. Theory of rational option pricing. *The Bell Journal of economics and management science*, JSTOR, p. 141–183, 1973.
- NELSON, D. B. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, v. 59, n. 2, p. 347–370, 1991.
- NOCEDAL, J. Updating quasi-newton matrices with limited storage. *Mathematics of computation*, v. 35, n. 151, p. 773–782, 1980.
- PAOLELLA, M.; HAAS, M. Mixture and regime-switching garch models. In: BAUWENS, L.; HAFNER, C. M.; LAURENT, S. (Ed.). *Handbook of volatility models and their applications*. Hoboken, NJ: Wiley, 2012, (Wiley handbooks in financial engineering and econometrics, 3). p. 71–102.
- PAOLELLA, M.; HAAS, M.; MITTNIK, S. Mixed normal conditional heteroskedasticity. *Journal of Financial Econometrics*, v. 2, n. 2, p. 211–250, 2004.
- PAOLELLA, M.; HAAS, M.; MITTNIK, S. A new approach to markov-switching garch models. *Journal of Financial Econometrics*, v. 2, n. 4, p. 493–530, 2004.
- PICCOLI, D. M. *Aplicação de redes neurais para previsão de contrato de dólar futuro no mercado brasileiro*. Dissertação (Mestrado) — Escola de Economia de São Paulo da Fundação Getúlio Vargas, <http://hdl.handle.net/10438/11502>, 2014.
- PRADO, M. L. d. Ten applications of financial machine learning. 2019. Disponível em: <<https://academic.microsoft.com/paper/2963557995>>.

TAYLOR, S. J. Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. *Time Series Analysis : Theory and Practice*, v. 1, p. 203-226, 1982.

Apêndices

APÊNDICE A – VARIAÇÃO DO COEFICIENTE DE PENALIZAÇÃO(λ)

Conforme destacado em 5.1, seguem abaixo as tabelas que evidenciam o desempenho satisfatório dos dois modelos MLP e MLP_{VRP} somente a partir dos valores de λ maiores ou iguais a 1 para os diferentes cenários e configurações analisados.

Tabela 14 – R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 400 dias nos cenários 3 e 4.

Amostra inicial 400 dias					
# opções	λ	Cenário 3		Cenário 4	
		MLP	MLP _{VRP}	MLP	MLP _{VRP}
17	0.001	-11,235.968	-17,763.514	-52,571.083	-17,987.328
17	0.01	-9,348.423	-6,077.113	-3,958.680	-11,159.563
17	0.1	-1,059.513	-2,620.714	-527.169	-407.667
17	1	-7.643	-22.459	-14.298	-59.725
17	10	-0.845	-0.032	-0.571	0.259
17	100	-1.034	0.007	-0.458	0.183
17	1000	-0.985	0.030	-0.386	0.222
19	0.001	-1,490.383	-520.775	-790.180	-669.222
19	0.01	-678.987	-1,016.216	-851.113	-1,161.835
19	0.1	-421.185	-426.928	-391.699	-572.443
19	1	-11.210	-13.371	-6.730	-7.960
19	10	-1.068	-0.129	-0.338	0.056
19	100	-1.201	0.031	-0.248	0.131
19	1000	-1.219	-0.100	-0.472	0.191
21	0.001	-1,772.437	-2,594.837	-2,099.517	-2,996.890
21	0.01	-1,710.759	-1,975.428	-1,443.151	-2,302.221
21	0.1	-309.480	-370.988	-579.321	-462.956
21	1	-8.730	-15.738	-11.557	-12.002
21	10	-1.101	-0.095	-0.813	0.216
21	100	-0.941	-0.002	-0.599	0.133
21	1000	-0.909	-0.014	-0.331	0.216
23	0.001	-1,084.615	-1,153.328	-1,018.394	-851.174
23	0.01	-890.470	-887.967	-721.296	-985.352

continua na próxima página

Amostra inicial 400 dias					
# opções	λ	Cenário 3		Cenário 4	
		MLP	MLP _{VRP}	MLP	MLP _{VRP}
23	0.1	-484.181	-408.242	-333.865	-272.445
23	1	-13.604	-8.976	-6.320	-15.566
23	10	-1.150	-0.375	-0.119	0.040
23	100	-1.004	0.005	-0.378	0.061
23	1000	-1.092	-0.112	-0.304	0.268
25	0.001	-47,451.844	-28,128.802	-10,630.376	-20,516.647
25	0.01	-43,157.514	-27,986.202	-17,929.776	-6,596.109
25	0.1	-480.159	-361.367	-420.298	-2,332.139
25	1	-15.938	-42.073	-13.802	-6.518
25	10	-0.683	-0.199	-0.252	0.163
25	100	-0.635	-0.041	-0.182	0.096
25	1000	-0.845	-0.067	-0.259	0.197
27	0.001	-107,033.005	-151,755.459	-1,848.836	-1,243.168
27	0.01	-26,570.294	-16,027.729	-1,324.711	-1,336.227
27	0.1	-6,694.651	-7,685.056	-112.095	-238.890
27	1	-10.582	-12.965	-6.007	-4.333
27	10	-3.473	-1.484	-1.240	-0.456
27	100	-2.511	-0.868	-1.215	-0.361
27	1000	-2.205	-0.630	-0.694	-0.214

Tabela 15 – R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 500 dias nos cenários 1 e 2.

Amostra inicial 500 dias					
# opções	λ	Cenário 1		Cenário 2	
		MLP	MLP_{VRP}	MLP	MLP_{VRP}
17	0.001	-40.628	-36.600	-480.888	-162.495
17	0.01	-46.293	-3.264	-46.807	-103.152
17	0.1	-4.163	-10.925	-4.166	-1.261
17	1	-0.142	0.122	-0.215	-0.250
17	10	-0.017	0.110	-0.020	0.082
17	100	-0.020	0.100	-0.039	0.101
17	1000	-0.019	0.104	-0.024	0.100
19	0.001	-68.299	-121.470	-13.634	-13.230
19	0.01	-40.911	-23.477	-12.954	-19.466
19	0.1	-22.674	-6.959	-6.508	-15.193
19	1	-0.604	-4.218	-0.350	0.035
19	10	-0.024	0.133	-0.022	0.011
19	100	-0.012	0.130	-0.009	0.023
19	1000	-0.031	0.120	-0.017	0.010
21	0.001	-44.725	-19.441	-50.152	-68.465
21	0.01	-16.757	-19.437	-43.458	-49.596
21	0.1	-61.381	-12.646	-122.104	-10.080
21	1	-0.233	-3.173	-0.205	0.126
21	10	-0.031	0.165	-0.012	0.155
21	100	-0.026	0.160	-0.009	0.139
21	1000	-0.020	0.152	-0.036	0.138
23	0.001	-33.888	-83.484	-20.374	-18.644
23	0.01	-109.746	-70.180	-22.758	-6.009
23	0.1	-15.684	-4.905	-64.054	-22.020
23	1	0.026	-1.563	-0.023	-0.012
23	10	-0.012	0.122	-0.049	0.126
23	100	-0.012	0.150	-0.033	0.112
23	1000	-0.026	0.131	-0.019	0.137
25	0.001	-186.361	-212.098	-349.636	-314.422
25	0.01	-198.928	-210.801	-332.263	-294.430
25	0.1	-16.319	-5.200	-10.609	-118.804
25	1	-3.685	0.010	0.384	0.108

continua na próxima página

Amostra inicial 500 dias					
# opções	λ	Cenário 1		Cenário 2	
		MLP	MLP _{VRP}	MLP	MLP _{VRP}
25	10	-0.032	0.155	-0.037	0.175
25	100	-0.047	0.157	-0.050	0.152
25	1000	-0.008	0.183	-0.024	0.171

Tabela 16 – R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 500 dias nos cenários 3 e 4.

Amostra inicial 500 dias					
# opções	λ	Cenário 3		Cenário 4	
		MLP	MLP _{VRP}	MLP	MLP _{VRP}
17	0.001	-12,414.898	-20,374.278	-59,999.838	-20,083.709
17	0.01	-10,662.995	-7,149.394	-4,110.471	-12,668.786
17	0.1	-1,163.540	-2,988.754	-557.828	-166.833
17	1	-9.303	-27.374	-16.847	-67.315
17	10	-0.477	0.103	-0.457	0.263
17	100	-0.764	0.046	-0.348	0.182
17	1000	-0.585	0.098	-0.312	0.274
19	0.001	-771.361	-808.190	-626.798	-626.869
19	0.01	-1,145.430	-976.867	-615.292	-914.847
19	0.1	-479.910	-303.366	-312.260	-701.971
19	1	-11.030	-12.518	-15.548	-7.943
19	10	-0.592	0.083	-0.305	0.087
19	100	-0.354	-0.036	-0.638	0.086
19	1000	-0.874	-0.082	-0.434	0.215
21	0.001	-1,925.214	-2,315.558	-1,522.681	-2,396.177
21	0.01	-1,684.068	-2,545.017	-2,468.792	-1,726.376
21	0.1	-337.275	-758.437	-671.747	-906.282
21	1	-11.025	-21.933	-17.112	-12.173
21	10	-0.795	-0.084	-0.055	0.208
21	100	-0.797	0.064	-0.303	0.016
21	1000	-0.697	0.051	-0.120	0.394
23	0.001	-1,172.006	-1,050.941	-745.738	-874.482
23	0.01	-792.535	-831.294	-907.956	-701.535
23	0.1	-378.881	-353.352	-272.875	-349.781
23	1	-13.281	-14.085	-3.492	-14.203
23	10	-0.681	0.022	-0.243	0.173

continua na próxima página

Amostra inicial 500 dias					
# opções	λ	Cenário 3		Cenário 4	
		MLP	MLP _{VRP}	MLP	MLP _{VRP}
23	100	-0.598	0.075	-0.633	-0.020
23	1000	-0.510	0.066	-0.179	0.325
25	0.001	-59,061.010	-35,941.013	-69,387.228	-61,027.652
25	0.01	-51,881.514	-36,644.064	-63,800.160	-59,125.794
25	0.1	-341.166	-350.533	-2,257.300	-1,324.596
25	1	-5.489	-43.947	-11.997	-54.296
25	10	-0.754	-0.053	-1.055	0.062
25	100	-0.680	0.134	-0.878	-0.109
25	1000	-0.845	-0.060	-1.024	-0.134

Tabela 17 – R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 600 dias nos cenários 1 e 2.

Amostra inicial 600 dias					
# opções	λ	Cenário 1		Cenário 2	
		MLP	MLP _{VRP}	MLP	MLP _{VRP}
17	0.001	-40.953	-36.859	-484.890	-163.856
17	0.01	-46.675	-3.255	-47.098	-103.566
17	0.1	-2.962	-10.781	-3.607	-0.987
17	1	-0.148	0.118	-0.195	-0.261
17	10	-0.024	0.103	-0.028	0.075
17	100	-0.026	0.094	-0.048	0.093
17	1000	-0.026	0.098	-0.033	0.092
19	0.001	-100.604	-116.386	-11.477	-8.516
19	0.01	-42.235	-28.945	-8.889	-11.339
19	0.1	-19.862	-6.296	-2.824	-10.792
19	1	-0.423	-4.594	-0.352	0.217
19	10	-0.030	0.127	-0.023	0.125
19	100	-0.020	0.123	-0.032	0.129
19	1000	-0.037	0.113	-0.036	0.119
21	0.001	-44.747	-19.278	-49.773	-65.679
21	0.01	-16.436	-18.928	-35.307	-48.428
21	0.1	-61.529	-12.470	-122.395	-8.123
21	1	-0.239	-3.190	-0.165	0.097
21	10	-0.039	0.159	-0.018	0.150
21	100	-0.033	0.152	-0.015	0.133

continua na próxima página

Amostra inicial 600 dias					
# opções	λ	Cenário 1		Cenário 2	
		MLP	MLP _{VRP}	MLP	MLP _{VRP}
21	1000	-0.028	0.144	-0.044	0.132
23	0.001	-20.475	-68.459	-19.012	-18.055
23	0.01	-99.669	-68.379	-21.166	-4.224
23	0.1	-14.483	-4.587	-64.157	-21.247
23	1	0.034	-1.577	-0.023	-0.008
23	10	-0.018	0.119	-0.055	0.122
23	100	-0.019	0.147	-0.038	0.108
23	1000	-0.033	0.127	-0.024	0.133

Tabela 18 – R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 600 dias nos cenários 3 e 4.

Amostra inicial 600 dias					
# opções	λ	Cenário 3		Cenário 4	
		MLP	MLP _{VRP}	MLP	MLP _{VRP}
17	0.001	-15,198.436	-24,965.170	-69,866.034	-23,386.308
17	0.01	-13,063.605	-8,681.087	-4,766.149	-14,682.531
17	0.1	-1,420.872	-3,675.184	-557.305	-149.258
17	1	-10.953	-33.816	-14.896	-78.370
17	10	-0.507	0.240	-0.397	0.411
17	100	-0.784	0.068	-0.293	0.251
17	1000	-0.622	0.112	-0.328	0.319
19	0.001	-755.535	-827.132	-1,502.366	-1,014.927
19	0.01	-769.623	-511.016	-1,060.298	-1,431.121
19	0.1	-492.867	-369.182	-390.373	-696.779
19	1	-14.538	-17.029	-10.325	-6.371
19	10	-0.415	0.128	-0.230	0.402
19	100	-0.219	-0.022	-0.288	0.155
19	1000	-0.790	-0.039	-0.470	0.239
21	0.001	-2,104.411	-2,596.381	-1,765.549	-2,464.786
21	0.01	-1,900.049	-2,889.718	-1,784.555	-1,900.688
21	0.1	-356.269	-770.405	-720.004	-816.571
21	1	-12.445	-23.461	-14.000	-18.497
21	10	-0.623	-0.035	-0.056	0.320
21	100	-0.789	0.069	-0.214	0.098
21	1000	-0.478	0.143	-0.254	0.410

continua na próxima página

Amostra inicial 600 dias					
# opções	λ	Cenário 3		Cenário 4	
		MLP	MLP _{VRP}	MLP	MLP _{VRP}
23	0.001	-1,803.137	-1,191.944	-1,260.069	-1,823.686
23	0.01	-1,147.888	-940.751	-1,560.650	-1,046.983
23	0.1	-311.276	-572.777	-536.498	-518.699
23	1	-14.375	-14.249	-7.424	-31.403
23	10	-1.318	-0.191	-1.615	-0.253
23	100	-1.479	-0.289	-2.321	-0.897
23	1000	-1.711	-0.409	-1.216	-0.186

Tabela 19 – R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 700 dias nos cenários 1 e 2.

Amostra inicial 700 dias					
# opções	λ	Cenário 1		Cenário 2	
		MLP	MLP _{VRP}	MLP	MLP _{VRP}
17	0.001	-81.230	-70.225	-487.415	-165.078
17	0.01	-66.261	-70.748	-47.323	-103.527
17	0.1	-5.781	-5.959	-3.042	-2.117
17	1	-0.142	-0.217	-0.174	-0.255
17	10	0.000	0.019	-0.033	0.069
17	100	-0.013	0.005	-0.053	0.088
17	1000	0.002	0.016	-0.038	0.086
19	0.001	-93.462	-115.560	-10.994	-8.115
19	0.01	-41.930	-29.560	-8.355	-10.472
19	0.1	-19.426	-5.471	-1.998	-9.544
19	1	-0.387	-4.287	-0.344	0.219
19	10	-0.036	0.123	-0.028	0.120
19	100	-0.023	0.119	-0.038	0.124
19	1000	-0.042	0.110	-0.042	0.113
21	0.001	-44.321	-19.138	-41.791	-56.021
21	0.01	-16.133	-18.518	-27.784	-42.028
21	0.1	-61.514	-10.952	-121.706	-4.344
21	1	-0.211	-3.217	-0.144	0.189
21	10	-0.045	0.155	-0.031	0.140
21	100	-0.038	0.147	-0.025	0.123
21	1000	-0.033	0.139	-0.054	0.122

Tabela 20 – R^2 ajustado dos modelos MLP e MLP_{VRP} para amostra inicial de 700 dias nos cenários 3 e 4.

Amostra inicial 700 dias					
# opções	λ	Cenário 3		Cenário 4	
		MLP	MLP_{VRP}	MLP	MLP_{VRP}
17	0.001	-22,336.559	-2,423.206	-166,713.290	-55,936.386
17	0.01	-3,326.151	-4,126.460	-11,355.880	-34,822.960
17	0.1	-210.131	-1,809.361	-1,116.550	-763.945
17	1	-65.059	-14.224	-26.198	-183.338
17	10	-1.996	-0.487	-1.583	-0.054
17	100	-2.755	-0.481	-1.448	-0.464
17	1000	-2.148	-0.670	-1.529	-0.360
19	0.001	-2,681.975	-3,124.493	-3,488.481	-2,328.484
19	0.01	-3,169.345	-1,178.168	-2,393.513	-3,182.256
19	0.1	-1,011.846	-634.485	-666.120	-1,252.413
19	1	-25.550	-38.383	-20.556	-14.383
19	10	-1.881	-0.840	-1.229	-0.188
19	100	-1.564	-1.135	-1.602	-0.733
19	1000	-2.816	-1.095	-2.006	-0.532
21	0.001	-3,873.849	-6,142.743	-1,107.029	-2,055.640
21	0.01	-5,234.399	-5,564.386	-1,398.293	-1,948.787
21	0.1	-938.266	-756.705	-899.727	-569.490
21	1	-16.265	-19.640	-21.079	-16.571
21	10	-2.358	-0.969	-1.365	-0.163
21	100	-2.762	-0.836	-1.020	-0.095
21	1000	-2.091	-0.667	-1.263	-0.333