



**Universidade do Estado do Rio de Janeiro**

Centro de Ciências Sociais

Faculdade de Ciências Econômicas

Renan Scavone Fernandes Pereira


**Previsão da taxa de câmbio Real/Dólar norte-americano usando modelos  
de *Deep Learning***

Rio de Janeiro

2025

Renan Scavone Fernandes Pereira

**Previsão da taxa de câmbio Real/Dólar norte-americano usando modelos de *Deep Learning***



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Ciências Econômicas, da Universidade do Estado do Rio de Janeiro. Área de concentração: Economia Aplicada.

Orientador: Prof. Dr. Fernando Antonio Lucena Aiube

Rio de Janeiro

2025

CATALOGAÇÃO NA FONTE  
UERJ/REDE SIRIUS/BIBLIOTECA CCS/B

P436 Pereira, Renan Scavone Fernandes

Previsão da taxa de cambio Real/Dólar norte-americano usando  
modelos de Deep Learning / Renan Scavone Fernandes Pereira. –  
2025.

112 f.

Orientador: Prof. Dr Fernando Antônio Lucena Aiube  
Dissertação (Mestrado em Ciências Econômicas) – Universidade  
do Estado do Rio de Janeiro, Faculdade de Ciências Econômicas.

1. Câmbio - Teses. 2. Redes neurais (Computação) - Teses. I.  
Aiube, Fernando Antônio Lucena. II. Universidade do Estado do Rio  
de Janeiro. Faculdade de Ciências Econômicas. III. Título.

CDU 336.745:004.8

Bibliotecária: Lucia Andrade CRB7/5272

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta  
dissertação, desde que citada a fonte.

---

Assinatura

---

Data

Renan Scavone Fernandes Pereira

**Previsão da taxa de câmbio Real/Dólar norte-americano usando modelos de *Deep Learning***

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Ciências Econômicas, da Universidade do Estado do Rio de Janeiro. Área de concentração: Economia Aplicada.

Aprovada em 22 de agosto de 2025.

Banca Examinadora:

---

Prof. Dr. Fernando Antonio Lucena Aiube (Orientador)  
Faculdade de Ciências Econômicas – UERJ

---

Prof<sup>a</sup> Dra. Jessica Facioli  
Faculdade de Ciências Econômicas – UERJ

---

Prof. Dr. Edison Americo Huarsaya Tito  
Instituto Militar de Engenharia

---

Prof. Dr. Carlos Henrique Dias Cordeiro de Castro  
EPGE Escola Brasileira de Economia e Finanças

Rio de Janeiro

2025



## **AGRADECIMENTOS**

Manifesto minha gratidão ao Prof. Dr. Fernando Aiube, cujos ensinamentos transmitidos, dedicação e orientação foram indispensáveis para a realização deste estudo. Expresso, ainda, meu reconhecimento à Marinha do Brasil (MB) pelo valioso incentivo à minha capacitação profissional. Por fim, deixo meu agradecimento à minha família, cujo apoio, inspiração e encorajamento constante tornaram possível esta jornada.

## RESUMO

PEREIRA, Renan Scavone Fernandes. *Previsão da taxa de câmbio Real/Dólar norte-americano usando modelos de Deep Learning*. 2025. 112 f. Dissertação (Mestrado em Ciências Econômicas) – Faculdade de Ciências Econômicas, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2025.

A previsão da taxa de câmbio constitui um dos maiores desafios da literatura econômica e financeira, assumindo relevância estratégica para a formulação de políticas e a tomada de decisões voltadas à alocação eficiente de recursos. Nesta dissertação, investiga-se a previsão da taxa de câmbio Ptax de venda do Real brasileiro frente ao Dólar norte-americano (BRL/USD) por meio de modelos de *Deep Learning* (DL), avaliados em horizontes de 1, 10 e 20 dias úteis à frente e comparados ao modelo *Random Walk* (RW), amplamente reconhecido na literatura como um *benchmark* robusto em previsões cambiais. Utilizaram-se séries de dados com frequência diária, abrangendo o período de 2020 a 2024, divididas sequencialmente em conjuntos de treinamento, validação e teste, de forma a simular uma aplicação prática. O estudo adotou duas abordagens: a univariada, baseada exclusivamente na série da taxa de câmbio, e a multivariada, que incorporou variáveis econômicas e financeiras. Para cada horizonte de previsão e abordagem, testaram-se arquiteturas de DL com uma (1CR) ou duas camadas recorrentes (2CR). A seleção do modelo de DL foi orientada pelas métricas de erro no conjunto de validação, seguida da avaliação do desempenho no conjunto teste. Os resultados evidenciam que, na abordagem univariada, as previsões dos modelos de DL selecionados não superaram as do RW no conjunto de teste, período em que os dados são novos para os modelos. Apenas no horizonte de 10 dias úteis, o teste de Diebold-Mariano Modificado (DMM) indicou equivalência estatística. Entretanto, para 20 dias úteis, verificou-se deterioração acentuada do desempenho dos modelos de DL selecionados, com métricas de erro superiores em, no mínimo, 50% em relação ao RW. Por outro lado, na abordagem multivariada, o teste DMM, ao nível de 10% de significância, apontou que as previsões dos modelos de DL selecionados apresentaram um desempenho estatisticamente superior ao RW no horizonte de 20 dias úteis, nas duas arquiteturas avaliadas (1CR e 2CR). De forma consistente, os resultados do teste de flutuação de Giacomini e Rossi (2010) e do teste de racionalidade de Rossi e Sekhposyan (2016) indicaram, respectivamente, nas janelas temporais, uma acurácia igual ou superior das previsões dos modelos de DL multivariados selecionados e um menor viés e irracionalidade dos erros comparativamente ao RW. Conclui-se que, embora os modelos de DL não tenham superado o *benchmark* no curtíssimo prazo, a inclusão de variáveis econômicas e financeiras diárias proporcionou ganhos significativos no desempenho preditivo em horizontes mais longos, ainda no curto prazo. Os resultados reforçam o potencial dos modelos multivariados de DL para aprimorar a previsão da taxa de câmbio, contribuindo tanto para a literatura acadêmica quanto para aplicações práticas no mercado financeiro.

Palavras-chave: Taxa de câmbio Ptax. Previsão. Deep Learning.

## ABSTRACT

PEREIRA, Renan Scavone Fernandes. *Forecasting the exchange rate of the brazilian Real to the US Dollar using Deep Learning Models*. 2025. 112 f. Dissertação (Mestrado em Ciências Econômicas) – Faculdade de Ciências Econômicas, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2025.

Exchange rate forecasting is one of the greatest challenges in economic and financial literature, holding strategic importance for policy formulation and decision-making aimed at efficient resource allocation. This dissertation investigates the forecasting of the Ptax exchange rate for the Brazilian Real against the US Dollar (BRL/USD) using Deep Learning (DL) models, evaluated over time horizons of 1, 10, and 20 business days ahead, and compared to the Random Walk (RW) model, which is widely recognized in the literature as a robust benchmark for exchange rate forecasting. Daily data series covering the period 2020 to 2024 were used, sequentially divided into training, validation, and test sets to simulate a practical application. The study adopted two approaches: a univariate approach, based exclusively on exchange rate data, and a multivariate approach, which incorporated economic and financial variables. For each forecast horizon and approach, DL architectures with one (1CR) or two recurrent layers (2CR) were tested. DL model selection was guided by error metrics on the validation set, followed by performance evaluation on the test set. The results show that, in the univariate approach, the predictions of the selected DL models did not outperform those of RW on the test set, a period in which the data is new to the models. Only on the 10-business-days horizon did the Modified Diebold-Mariano (DMM) test indicate statistical equivalence. However, for the 20-business-days horizon, a sharp deterioration in the performance of the selected DL models was observed, with error metrics at least 50% higher than those of the RW. On the other hand, in the multivariate approach, the DMM test, at the 10% significance level, indicated that the forecasts of the selected DL models performed statistically better than those of RW over the 20-business-days horizon, in both architectures evaluated (1CR and 2CR). Consistently, the results of the Giacomini and Rossi (2010) fluctuation test and the Rossi and Sekhposyan (2016) rationality test indicated, respectively, in the time windows, equal or superior accuracy of the selected multivariate DL models' forecasts, and lower bias and irrationality in errors compared to the RW. It is concluded that, although the DL models did not outperform the benchmark in the very short term, the inclusion of daily economic and financial variables provided significant gains in predictive performance over longer horizons, even in the short term. The results reinforce the potential of multivariate DL models to improve exchange rate forecasting, contributing both to the academic literature and to practical applications in the financial market.

Keywords: Exchange Rate Ptax. Forecast. Deep Learning.

## LISTA DE ILUSTRAÇÕES

Figura	1 - Diagrama do modelo LSTM. . . . .	26
Figura	2 - Diagrama do modelo GRU. . . . .	28
Figura	3 - Gráfico das séries temporais coletadas no período de 2020 a 2024. . . . .	46
Figura	4 - Matriz de correlação das variáveis (preços) no período de 2020 a 2024. . .	48
Figura	5 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 1 dia útil à frente do modelo selecionado univariado com 1CR vs RW no período de teste . . . . .	54
Figura	6 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 1 dia útil à frente do modelo selecionado univariado com 2CR vs RW no período de teste . . . . .	55
Figura	7 - Gráficos das estatísticas dos testes de flutuação Giacomini e Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação . . . . .	57
Figura	8 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 1 dia útil à frente do modelo selecionado multivariado com 1CR vs RW no período de teste . . . . .	61
Figura	9 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 1 dia útil à frente do modelo selecionado multivariado com 2CR vs RW no período de teste . . . . .	62
Figura	10 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação . . . . .	64
Figura	11 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 10 dias úteis à frente do modelo selecionado univariado com 1CR vs RW no período de teste . . . . .	68
Figura	12 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 10 dias úteis à frente do modelo selecionado univariado com 2CR vs RW no período de teste . . . . .	69
Figura	13 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação . . . . .	71
Figura	14 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 10 dias úteis à frente do modelo selecionado multivariado com 1CR vs RW no período de teste . . . . .	75

Figura 15 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 10 dias úteis à frente do modelo selecionado multivariado com 2CR vs RW no período de teste . . . . .	76
Figura 16 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação . . . . .	78
Figura 17 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 20 dias úteis à frente do modelo selecionado univariado com 1CR vs RW no período de teste . . . . .	81
Figura 18 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 20 dias úteis à frente do modelo selecionado univariado com 2CR vs RW no período de teste . . . . .	82
Figura 19 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação . . . . .	84
Figura 20 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 20 dias úteis à frente do modelo selecionado multivariado com 1CR vs RW no período de teste . . . . .	87
Figura 21 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 20 dias úteis à frente do modelo selecionado multivariado com 2CR vs RW no período de teste . . . . .	88
Figura 22 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação . . . . .	90
Figura 23 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação nas previsões 1 dia à frente . . . . .	101
Figura 24 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação . . . . .	102
Figura 25 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação . . . . .	103
Figura 26 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação . . . . .	104
Figura 27 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação . . . . .	105

Figura 28 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação . . . . .	106
Figura 29 - Funções de perda dos períodos de treino e de validação do modelo de DL univariado com 1CR 1 dia à frente . . . . .	107
Figura 30 - Funções de perda dos períodos de treino e de validação do modelo de DL univariado com 2CR 1 dia à frente . . . . .	107
Figura 31 - Funções de perda dos períodos de treino e de validação do modelo de DL multivariado com 1CR 1 dia à frente . . . . .	108
Figura 32 - Funções de perda dos períodos de treino e de validação do modelo de DL multivariado com 2CR 1 dia à frente . . . . .	108
Figura 33 - Funções de perda dos períodos de treino e de validação do modelo de DL univariado de 1CR com previsão 10 dias úteis à frente . . . . .	109
Figura 34 - Funções de perda dos períodos de treino e de validação do modelo de DL univariado de 2CR com previsão 10 dias úteis à frente . . . . .	109
Figura 35 - Funções de perda dos períodos de treino e de validação do modelo de DL multivariado de 1CR com previsão 10 dias úteis à frente . . . . .	110
Figura 36 - Funções de perda dos períodos de treino e de validação do modelo de DL multivariado de 2CR com previsão 10 dias úteis à frente . . . . .	110
Figura 37 - Funções de perda dos períodos de treino e de validação do modelo de DL univariado de 1CR com previsão 20 dias úteis à frente . . . . .	111
Figura 38 - Funções de perda dos períodos de treino e de validação do modelo de DL univariado de 2CR com previsão 20 dias úteis à frente . . . . .	111
Figura 39 - Funções de perda dos períodos de treino e de validação do modelo de DL multivariado de 1CR com previsão 20 dias úteis à frente . . . . .	112
Figura 40 - Funções de perda dos períodos de treino e de validação do modelo de DL multivariado de 2CR com previsão 20 dias úteis à frente . . . . .	112

## LISTA DE TABELAS

Tabela	1 - Arquiteturas e parâmetros dos modelos de DL . . . . .	40
Tabela	2 - Métricas descritivas das séries temporais em preços . . . . .	47
Tabela	3 - Métricas de erro das previsões dos modelos univariados de DL com 1CR e do modelo RW 1 dia à frente . . . . .	51
Tabela	4 - Métricas de erro das previsões 1 dia à frente dos modelos univariados de DL com 2CR e do RW . . . . .	52
Tabela	5 - Métricas de erro das previsões 1 dia à frente do modelo RW e dos modelos de DL univariados selecionados em cada arquitetura . . . . .	53
Tabela	6 - Testes estatísticos para o modelo univariado de DL selecionado com 2CR e o modelo RW nas previsões 1 dia útil à frente . . . . .	56
Tabela	7 - Métricas de erro das previsões dos modelos multivariados de DL com 1CR e do modelo RW 1 dia útil à frente . . . . .	58
Tabela	8 - Métricas de erro das previsões dos modelos multivariados de DL com 2CR e do modelo RW 1 dia útil à frente . . . . .	59
Tabela	9 - Métricas de erro das previsões 1 dia útil à frente dos modelos de DL multivariados selecionados em cada arquitetura vs. modelo RW . . . . .	60
Tabela	10 - Testes estatísticos para comparação do modelo multivariado de DL selecionado com 2CR e do modelo RW nas previsões 1 dia útil à frente . . . . .	63
Tabela	11 - Métricas de erro das previsões dos modelos univariados de DL com 1CR e do modelo RW 10 dias úteis à frente . . . . .	65
Tabela	12 - Métricas de erro das previsões dos modelos univariados de DL com 2CR e do modelo RW 10 dias úteis à frente . . . . .	66
Tabela	13 - Métricas de erro das previsões 10 dias úteis à frente dos modelos de DL univariados selecionados em cada arquitetura vs. modelo RW . . . . .	67
Tabela	14 - Testes estatísticos para comparação do modelo univariado de DL selecionado com 1CR e do modelo RW nas previsões 10 dias úteis à frente . . . . .	70
Tabela	15 - Métricas de erro das previsões dos modelos multivariados de DL com 1CR e do modelo RW 10 dias úteis à frente . . . . .	72
Tabela	16 - Métricas de erro das previsões dos modelos multivariados de DL com 2CR e do modelo RW 10 dias úteis à frente . . . . .	73
Tabela	17 - Métricas de erro das previsões 10 dias úteis à frente dos modelos de DL multivariados selecionados em cada arquitetura vs. modelo RW . . . . .	74
Tabela	18 - Testes estatísticos para comparação do modelo multivariado de DL selecionado com 2CR e do modelo RW nas previsões 10 dias úteis à frente . . . . .	77
Tabela	19 - Métricas de erro das previsões dos modelos univariados de DL com 1CR e do modelo RW 20 dias úteis à frente . . . . .	79

Tabela 20 - Métricas de erro das previsões dos modelos univariados de DL com 2CR e do modelo RW 20 dias úteis à frente . . . . .	80
Tabela 21 - Métricas de erro das previsões 20 dias úteis à frente dos modelos de DL univariados selecionados em cada arquitetura vs. modelo RW . . . . .	81
Tabela 22 - Testes estatísticos para comparação do modelo univariado de DL selecionado com 1CR e do modelo RW nas previsões 20 dias úteis à frente . . . .	83
Tabela 23 - Métricas de erro das previsões dos modelos multivariados de DL com 1CR e do modelo RW 20 dias úteis à frente . . . . .	85
Tabela 24 - Métricas de erro das previsões dos modelos multivariados de DL com 2CR e do modelo RW 20 dias úteis à frente . . . . .	86
Tabela 25 - Métricas de erro das previsões 20 dias úteis à frente dos modelos de DL multivariados selecionados em cada arquitetura vs. modelo RW . . . . .	87
Tabela 26 - Testes estatísticos para comparação do modelo multivariado de DL selecionado com 1CR e do modelo RW nas previsões 20 dias úteis à frente . . .	89
Tabela 27 - Testes estatísticos para comparação do modelo univariado de DL selecionado com 1CR e do modelo RW nas previsões 1 dia à frente . . . . .	101
Tabela 28 - Testes estatísticos para comparação do modelo multivariado de DL selecionado com 1CR e do modelo RW nas previsões 1 dia à frente . . . . .	102
Tabela 29 - Testes estatísticos para comparação do modelo univariado de DL selecionado com 2CR e do modelo RW nas previsões 10 dias úteis à frente . . . .	103
Tabela 30 - Testes estatísticos para comparação do modelo multivariado de DL selecionado com 1CR e do modelo RW nas previsões 10 dias úteis à frente . . .	104
Tabela 31 - Testes estatísticos para comparação do modelo univariado de DL selecionado com 2CR e do modelo RW nas previsões 20 dias úteis à frente . . . .	105
Tabela 32 - Testes estatísticos para comparação do modelo multivariado de DL selecionado com 2CR e do modelo RW nas previsões 20 dias úteis à frente . . .	106



## LISTA DE ABREVIATURAS E SIGLAS

ADAM	<i>Adaptive Moment Estimation</i>
ADAMW	<i>Adaptive Moment Estimation with Decoupled Weight Decay</i>
BiGRU	<i>Bidirectional Gated Recurrent Unit</i>
BiLSTM	<i>Bidirectional Long Short-Term Memory</i>
DXY	<i>U.S Dollar Index</i>
CDS	<i>Credit Default Swap</i>
DL	<i>Deep Learning</i>
CRY	<i>Índice CRY</i>
GRU	<i>Gated Recurrent Unit</i>
LSTM	<i>Long Short-Term Memory</i>
ML	<i>Machine Learning</i>
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MSE	<i>Mean Squared Error</i>
BRL	<i>Real brasileiro</i>
ReLU	<i>Rectified Linear Unit</i>
RMSE	<i>Root Mean Squared Error</i>
RW	<i>Random Walk - Passeio Aleatório</i>
USD	<i>Dólar Norte-Americano</i>
VIX	<i>Cboe Volatility Index</i>

## SUMÁRIO

	<b>INTRODUÇÃO</b>	15
1	<b>REVISÃO DA LITERATURA</b>	18
2	<b>FUNDAMENTOS TEÓRICOS</b>	24
2.1	<b>Modelo <i>Random Walk</i> (Passeio Aleatório)</b>	24
2.2	<b>Long Short-Term Memory Unit (LSTM) e Bidirectional Long Short-Term Memory Unit (BiLSTM)</b>	24
2.3	<b>Gated Recurrent Unit (GRU) e Bidirectional Gated Recurrent Unit (BiGRU)</b>	27
2.4	<b>Conceitos relevantes aplicados ao <i>Deep Learning</i></b>	28
2.4.1	<u>Funções de ativação</u>	28
2.4.2	<u>Gradiente descendente</u>	29
2.4.3	<u><i>Overfitting e underfitting</i> e conceitos relacionados ao treinamento dos modelos de DL</u>	29
2.5	<b>Métricas de erro</b>	30
2.5.1	<u>Erro Médio Absoluto ou <i>Mean Absolute Error</i> (MAE)</u>	30
2.5.2	<u>Erro Quadrado Médio ou <i>Mean Squared Error</i> (MSE)</u>	31
2.5.3	<u>Raiz do Erro Quadrado Médio ou <i>Root Mean Squared Error</i> (RMSE)</u>	31
2.5.4	<u>Erro Médio Absoluto Percentual ou <i>Mean Absolute Percent Error</i> (MAPE)</u>	31
2.6	<b>Testes estatísticos</b>	31
2.6.1	<u>Teste de Diebold-Mariano (1995)</u>	32
2.6.2	<u>Teste de Diebold-Mariano Modificado (1997)</u>	33
2.6.3	<u>Teste de flutuação de Giacomini e Rossi (2010)</u>	33
2.6.4	<u>Teste de racionalidade de Rossi e Sekhposyan (2016)</u>	34
3	<b>METODOLOGIA</b>	37
3.1	<b>Coleta, preparação dos dados e treinamento</b>	37
3.2	<b>Modelos Preditivos</b>	39
3.3	<b>Avaliação do Desempenho</b>	41
3.4	<b>Seleção de variáveis explicativas</b>	42
4	<b>DESCRIÇÃO ESTATÍSTICA DOS DADOS</b>	45
5	<b>RESULTADOS</b>	50
5.1	<b>Resultados das previsões 1 dia à frente</b>	51
5.1.1	<u>Modelos univariados de DL vs. modelo RW nas previsões 1 dia à frente</u>	51
5.1.1.1	<u>Modelos univariados de DL com 1CR vs. modelo RW nas previsões 1 dia à frente</u>	51
5.1.1.2	<u>Modelos univariados de DL com 2CR vs. modelo RW nas previsões 1 dia à frente:</u>	52

5.1.1.3	Modelos de DL univariados selecionados vs. modelo RW nas previsões 1 dia útil à frente . . . . .	53
5.1.1.4	Testes estatísticos entre o modelo RW e o modelo de DL univariado selecionado com melhor desempenho no período de teste nas previsões 1 dia útil à frente . . . . .	56
5.1.2	<u>Modelos multivariados de DL vs. modelo RW nas previsões 1 dia útil à frente .</u>	58
5.1.2.1	Modelos multivariados de DL com 1CR vs. modelo RW nas previsões 1 dia útil à frente: . . . . .	58
5.1.2.2	Modelos multivariados de DL com 2CR vs. modelo RW nas previsões 1 dia útil à frente: . . . . .	59
5.1.2.3	Modelos de DL multivariados selecionados vs. modelo RW nas previsões 1 dia útil à frente . . . . .	60
5.1.2.4	Testes estatísticos entre o modelo de DL multivariado selecionado com melhor desempenho no período de teste vs. modelo RW nas previsões 1 dia útil à frente	63
5.2	<b>Resultados das previsões 10 dias úteis à frente . . . . .</b>	65
5.2.1	<u>Modelos univariados de DL vs. modelo RW nas previsões 10 dias úteis à frente:</u>	65
5.2.1.1	Modelos univariados de DL com 1CR vs. modelo RW nas previsões 10 dias úteis à frente: . . . . .	65
5.2.1.2	Modelos univariados de DL com 2CR vs. modelo RW nas previsões 10 dias úteis à frente: . . . . .	66
5.2.1.3	Modelos de DL univariados selecionados vs. modelo RW nas previsões 10 dias úteis à frente . . . . .	67
5.2.1.4	Testes estatísticos entre o modelo de DL univariado selecionado com melhor desempenho no período de teste vs. modelo RW nas previsões 10 dias úteis à frente . . . . .	70
5.2.2	<u>Modelos multivariados de DL vs. modelo RW nas previsões 10 dias úteis à frente</u>	72
5.2.2.1	Modelos multivariados de DL com 1CR vs. modelo RW nas previsões 10 dias úteis à frente: . . . . .	72
5.2.2.2	Modelos multivariados de DL com 2CR vs. modelo RW nas previsões 10 dias úteis à frente: . . . . .	73
5.2.2.3	Modelos de DL multivariados selecionados vs. modelo RW nas previsões 10 dias úteis à frente . . . . .	74
5.2.2.4	Testes estatísticos entre o modelo de DL multivariado selecionado com melhor desempenho no período de teste vs. modelo RW nas previsões 10 dias úteis à frente . . . . .	77
5.3	<b>Resultados das previsões 20 dias úteis à frente . . . . .</b>	79
5.3.1	<u>Modelos univariados de DL vs. modelo RW nas previsões 20 dias úteis à frente</u>	79
5.3.1.1	Modelos univariados de DL com 1CR vs. modelo RW nas previsões 20 dias úteis à frente: . . . . .	79

5.3.1.2	Modelos univariados de DL com 2CR vs. modelo RW nas previsões 20 dias úteis à frente: . . . . .	80
5.3.1.3	Modelos de DL univariados selecionados vs. modelo RW nas previsões 20 dias úteis à frente . . . . .	81
5.3.1.4	Testes estatísticos entre o modelo de DL univariado selecionado com melhor desempenho no período de teste vs. modelo RW nas previsões 20 dias úteis à frente . . . . .	83
5.3.2	<u>Modelos multivariados de DL vs. modelo RW nas previsões 20 dias úteis à frente</u>	85
5.3.2.1	Modelos multivariados de DL com 1CR vs. modelo RW nas previsões 20 dias úteis à frente: . . . . .	85
5.3.2.2	Modelos multivariados de DL com 2CR vs. modelo RW nas previsões 20 dias úteis à frente: . . . . .	86
5.3.2.3	Modelos de DL multivariados selecionados vs. modelo RW nas previsões 20 dias úteis à frente . . . . .	87
5.3.2.4	Testes estatísticos entre o modelo de DL multivariado selecionado com melhor desempenho no período de teste vs. modelo RW nas previsões 20 dias úteis à frente . . . . .	89
5.4	<b>Considerações finais sobre os resultados obtidos</b> . . . . .	91
	<b>CONCLUSÃO</b> . . . . .	94
	<b>REFERÊNCIAS</b> . . . . .	96
	<b>APÊNDICE A</b> – Testes estatísticos para comparação do desempenho das previsões do modelo RW e dos modelos de DL selecionados com segundo menor valor nas métricas de erro no período de teste . . . . .	101
	<b>APÊNDICE B</b> – Gráficos obtidos no treinamento dos modelos de DL selecionados com as menores métricas de erro no período de teste referente ao comportamento das funções de perda de treino e de validação . . . . .	107

## INTRODUÇÃO

A taxa de câmbio representa o preço de uma moeda estrangeira em relação à moeda nacional. Assim, ela exerce um papel fundamental na dinâmica macroeconômica, influenciando diretamente o comércio exterior, os fluxos de capitais e a formulação da política monetária. Em economias emergentes como a brasileira, sua oscilação pode afetar significativamente o nível de preços, a competitividade e as decisões de investimento. Nesse contexto, prever a trajetória da taxa de câmbio, ainda que em horizontes curtos, é uma tarefa de grande relevância para formuladores de políticas, analistas financeiros e agentes do mercado.

Durante a pandemia de COVID-19, o mercado cambial passou por uma intensa volatilidade. A incerteza econômica e o fluxo de capitais para ativos mais seguros levaram a uma expressiva desvalorização do Real brasileiro frente ao Dólar americano, impulsionada pela fuga de investidores estrangeiros e pela redução da liquidez global. Além disso, as medidas de estímulo fiscal e monetário adotadas pelo governo brasileiro para mitigar os efeitos da crise geraram preocupações quanto à sustentabilidade fiscal e à trajetória da dívida pública, aumentando a percepção de risco. Mais recentemente, o cenário político internacional tem sido marcado por conflitos armados e transformações geopolíticas relevantes, como as eleições nos Estados Unidos, que intensificaram disputas comerciais e elevaram tarifas aduaneiras — especialmente entre os EUA e a China. Esse ambiente contribuiu para tornar o comportamento da taxa de câmbio ainda mais complexo e volátil. Nesse cenário, reforça-se a necessidade de aperfeiçoamento contínuo dos modelos preditivos com objetivo de captar a complexidade do mercado cambial e oferecer suporte consistente à tomada de decisão de agentes econômicos, tanto no setor público quanto no privado.

Segundo Krugman, Obstfeld e Melitz (2015, p. 269) a taxa de câmbio é considerada um preço de ativo, sendo regulada pelos mesmos princípios que governam outros preços de ativos financeiros. Dessa forma, não existem oportunidades contínuas de lucro sem risco. Se existir uma diferença de preços, investidores exploram essa oportunidade por meio de arbitragem, levando a um equilíbrio nos preços. Por outro lado, pela hipótese de eficiência de mercado, não se esperam retornos superiores à média de forma consistente, já que os preços se ajustam rapidamente às novas informações. Neste contexto, Krugman, Obstfeld e Melitz (2015, prefácio) descrevem que as taxas de juros e as expectativas são fatores essenciais na determinação do câmbio.

As dinâmicas cambiais são influenciadas por uma multiplicidade de fatores, como choques externos, taxas de juros, fluxos de capitais e expectativas de mercado, o que torna a previsão da taxa de câmbio uma tarefa desafiadora. Embora modelos econométricos tradicionais sejam amplamente utilizados nesse contexto, eles frequentemente apresentam limitações na captura de padrões não lineares e de interações complexas entre variáveis. Desde os trabalhos pioneiros de Meese e Rogoff (1982, 1983), a literatura evidencia que modelos baseados em

fundamentos econômicos, em geral, não superam o desempenho preditivo do modelo *Random Walk* (RW), apontado como um *benchmark*, especialmente em horizontes de curto prazo.

Diante desse cenário e impulsionadas pelos avanços computacionais, as redes neurais artificiais têm se consolidado como uma alternativa promissora para modelagem e previsão em séries temporais financeiras. Nesse âmbito, as técnicas de aprendizado profundo, conhecidas como *Deep Learning* (DL), são capazes de capturar padrões complexos e não linearidades, oferecendo maior flexibilidade na modelagem de relações dinâmicas. Assim, o principal objetivo deste estudo é investigar se as previsões dos modelos *Long Short-Term Memory* (LSTM), *Bi-directional Long Short-Term Memory* (BiLSTM), *Gated Recurrent Unit* (GRU) e *Bidirectional Gated Recurrent Unit* (BiGRU), aplicados conforme as arquiteturas e configurações descritas na metodologia, são capazes de superar o desempenho das previsões do modelo RW, na previsão da taxa de câmbio Ptax de venda do Real brasileiro por Dólar americano.

A definição dos horizontes de previsão adotados neste estudo — 1, 10 e 20 dias úteis à frente — decorre de uma demanda prática observada no âmbito da administração pública, especificamente na Diretoria de Finanças da Marinha do Brasil, referente a otimização do envio de recursos financeiros ao exterior. Esses recursos destinam-se a cobrir despesas de custeio e aquisição de equipamentos de defesa, e sua liberação está sujeita a peculiaridades da execução orçamentária e financeira do governo federal, que muitos vezes impõem prazos operacionais de curto prazo. Nesse sentido, torna-se inviável a adoção de modelos baseados em fundamentos econômicos tradicionais como os modelos monetários, como a regra de Taylor combinada com a paridade da taxa de juros ou, ainda, como a paridade do poder de compra. Tais modelos utilizam variáveis macroeconômicas que, em sua maioria, são divulgadas com frequência mensal ou trimestral, o que compromete sua aplicabilidade em previsões de alta frequência.

Dessa forma, serão utilizadas variáveis econômicas e financeiras com disponibilidade diária nas previsões dos modelos multivariados com objetivo de capturar variações relevantes na taxa de câmbio de curto prazo. Entre elas, destacam-se as taxas de juros de curto e longo prazo, o índice de preços de *commodities* (CRY), os índices acionários Ibovespa e S&P 500, o *Credit Default Swap* (CDS) associado ao risco soberano e o índice de volatilidade Cboe (VIX) como proxy do risco de mercado e do risco sistêmico, além do *United States Dollar Index* (DXY), que reflete a força relativa do Dólar americano frente a uma cesta de moedas de países desenvolvidos. Essas variáveis permitem incorporar informações contemporâneas e relevantes ao comportamento diário da taxa de câmbio, tornando-as mais adequadas para os horizontes propostos.

Abrangendo o período de 2020 a 2024, adotam-se duas abordagens: a univariada, que utiliza apenas os valores passados da própria taxa de câmbio, e a multivariada, que incorpora variáveis econômicas e financeiras, além dos valores passados da taxa de câmbio. Para cada abordagem e horizonte de previsão, os modelos de DL serão avaliados em duas arquiteturas distintas: uma delas com uma camada recorrente e outra com duas. O desempenho dos modelos será comparado por meio das métricas de erro RMSE, MAE e MAPE, além da aplicação de

testes estatísticos.

Os resultados obtidos indicam que, embora as previsões dos modelos de DL não tenham superado as do *benchmark* tradicional no curtíssimo prazo, a inclusão de variáveis econômicas e financeiras de frequência diária gerou ganhos relevantes para horizontes mais longos, ainda no curto prazo. Esses achados evidenciam o potencial das abordagens multivariadas baseadas em redes neurais para aprimorar a previsão da taxa de câmbio, contribuindo para o avanço de um campo ainda pouco explorado na literatura brasileira e oferecendo subsídios tanto à formulação de políticas econômicas quanto a decisões voltadas à alocação eficiente de recursos.

Além desta introdução, o estudo está estruturado da seguinte forma: o Capítulo 1 apresenta a revisão da literatura sobre a previsão da taxa de câmbio; o Capítulo 2 descreve os modelos de previsão utilizados, os conceitos fundamentais de DL e os testes estatísticos aplicados para a comparação do desempenho preditivo; o Capítulo 3 detalha a metodologia adotada; o Capítulo 4 apresenta os dados utilizados na análise, acompanhados pelas respectivas estatísticas descritivas; o Capítulo 5 descreve a comparação dos resultados das previsões de 1, 10 e 20 dias úteis à frente do modelo RW com as dos modelos de DL univariados e multivariados, além das considerações finais sobre os resultados obtidos; e, na sequência, descrevem-se as conclusões dessa dissertação.

## 1 REVISÃO DA LITERATURA

A literatura apresenta diferentes abordagens para a previsão da taxa de câmbio, algumas das quais são destacadas a seguir, com ênfase especial naquelas aplicadas ao caso brasileiro. Porém, antes disso, cabe apresentar o conceito de taxa de câmbio Ptax e os motivos de sua utilização neste estudo.

O mercado de câmbio é dividido em dois tipos principais: à vista e a termo. No mercado à vista, a troca de moedas ocorre imediatamente, com um preço previamente estabelecido. Já no mercado a termo, o valor da moeda é determinado no presente, mas a liquidação da transação ocorre em uma data futura. Este estudo focará na previsão do mercado de câmbio à vista. Cabe observar que, conforme descrito por Dornbusch, Fischer e Startz (2013, p. 46), nos regimes de câmbio fixo os bancos centrais intervêm comprando e vendendo moedas estrangeiras para estabilizar a taxa de câmbio, enquanto nos regimes flutuantes, o mercado determina o valor da moeda. No entanto, muitos países, como o Brasil, adotam um regime intermediário, permitindo a flutuação cambial com intervenções ocasionais para evitar variações extremas.

Conforme o Estudo Especial n. 42 do Banco Central do Brasil (BRASIL. Banco Central do Brasil, 2019), a taxa de câmbio de referência do Real por Dólares americanos mais utilizada no mercado cambial brasileiro é a Ptax. Ela serve como base para contratos futuros, opções, derivativos e operações financeiras. Além disso, é referência para contratos em Real nas bolsas internacionais, e amplamente usada por analistas econômicos, pesquisadores e o público em geral. Sendo, por este motivo, a taxa de câmbio escolhida neste estudo preditivo. A Resolução n° 45 do Banco Central do Brasil (BRASIL. Banco Central do Brasil, 2020) dispõe que o referido órgão realize quatro consultas diárias, com duração de dois minutos cada, em horários aleatórios dentro dos intervalos: 10h-10h10, 11h-11h10, 12h-12h10 e 13h-13h10. Assim, a taxa de câmbio Ptax de compra e a taxa de câmbio Ptax de venda do Real/Dólar americano referentes a cada consulta corresponderão, respectivamente, às médias das cotações de compra e das cotações de venda efetivamente fornecidas pelos *dealers*, excluídas, em cada caso, as duas maiores e as duas menores. Por sua vez, as taxas Ptax de compra e taxas Ptax de venda de cada dia corresponderão às médias aritméticas das quatro taxas de compra ou das quatro taxas de venda do Real/Dólar americano, conforme o caso, obtidas na forma descrita acima, sendo divulgadas pelo Banco Central do Brasil (BCB) conjuntamente com o resultado da última consulta do dia.

Conforme descrito anteriormente, Krugman, Obstfeld e Melitz (2015, p. 269) apontam que a taxa de câmbio é considerada um preço de ativo, seguindo os mesmos princípios que governam os preços de ativos financeiros. Nesse sentido, aplicam-se também as conclusões de Fama (1965), que testa a validade empírica do modelo RW para o comportamento dos preços das ações e que aponta que os dados parecem oferecer um suporte consistente para esse modelo. O artigo utiliza dados diários das trinta ações que compõem o índice industrial *Dow Jones*, com



o período de análise variando de ação para ação, abrangendo, em geral, do final de 1957 até setembro de 1962. Argumenta-se que o uso do histórico de preços não seria capaz de contribuir para uma previsão confiável dos movimentos futuros e de aumentar o lucro esperado do investidor, dada a evidência de independência das mudanças dos preços, fato que é consistente com a hipótese de eficiência de mercado. Assim, considera-se que os argumentos a favor da análise gráfica são completamente insustentáveis, a menos que respaldados por dados empíricos.

Os reconhecidos estudos de Meese e Rogoff (1982, 1983) avaliaram a precisão preditiva de modelos estruturais e de séries temporais na previsão das taxas de câmbio, comparando-os ao modelo de RW. Os resultados indicaram a incapacidade dos modelos estruturais de explicar o comportamento das taxas de câmbio com base em fundamentos econômicos. Foram analisados modelos monetários de preços flexíveis (Frenkel-Bilson), preços rígidos (Dornbusch-Frankel) e um modelo de preços rígidos com conta corrente (Hooper-Morton), com a utilização de modelos VAR (vetores auto-regressivos) para identificar fatores que influenciam a taxa de câmbio. Utilizaram-se dados mensais de 1973 a 1981 para diferentes pares cambiais, considerando de acordo com cada modelo variáveis como diferenciais de juros de curto e de longo prazo, de oferta de moeda americana e estrangeira, de balanças comerciais, de inflação e de produção. Os resultados demonstraram que o RW é tão preciso quanto qualquer modelo estimado para previsões de 1 a 12 meses, com sua vantagem diminuindo conforme o horizonte se aproxima de 12 meses. Para prazos mais longos, embora algumas métricas de erro, como RMSE e MAE, possam apresentar melhorias ocasionais, os resultados continuam instáveis. Mesmo com o uso de valores futuros conhecidos das variáveis explicativas, os modelos estruturais não superam o RW, ressaltando a dificuldade inerente à previsão cambial.

Posteriormente, Rogoff (2008) destaca que alguns estudos trazem evidências de que modelos de taxa de câmbio baseados em fundamentos, especialmente monetários, parecem superar o RW na previsão de longo prazo. Esses modelos, que podem incluir funções de reação da regra de Taylor, explicam como as taxas de câmbio reagem a anúncios de juros do Banco Central, com valorização da moeda diante de expectativas de altas futuras. Considera-se que o debate passou a girar em torno da eficácia dos modelos estruturais em prever horizontes intermediários, de um mês a um ano, os mais relevantes para políticas econômicas. Salienta-se que o resultado do estudo de Meese e Rogoff (1983), o qual trouxe evidências de que os modelos estruturais não superam o modelo RW na previsão da taxa de câmbio, refere-se principalmente às moedas flutuantes de economias avançadas, como o Iene, a Libra e o Euro em relação ao dólar, mas não necessariamente às moedas de mercados emergentes, devido aos diferenciais de juros e inflação geralmente mais elevados. Ressalta-se também uma exceção no caso de países como Austrália, Nova Zelândia e Canadá, cujas taxas de câmbio parecem estar fortemente correlacionadas com os preços das *commodities* que compõem suas pautas de exportação.

Rossi (2013) revisa a literatura sobre previsão cambial desde Meese e Rogoff (1983) e compara o desempenho de diversos modelos em horizontes curtos (um mês ou trimestre) e longos (quatro anos) com o modelo RW sem *drift*, utilizando o RMSFE (*Root Mean Squared*

*Forecast Error*) como métrica. Reestimam-se modelos VECM recursivamente em uma janela móvel correspondente à metade da amostra e empregam-se testes estatísticos como Diebold-Mariano (1995) e West (1996), Clark-West (2006) e o teste de flutuação de Giacomini e Rossi (2010). O estudo considera taxas de câmbio em relação aos EUA para países como Austrália, Canadá, Japão e Reino Unido, reunindo indicadores mensais e trimestrais, incluindo taxas de juros *overnight*, Letras do Tesouro de 3 meses, Títulos do Tesouro de 5 anos, PIB/produção industrial, índice de preços ao consumidor (CPI) e estoque de moeda, além de dados anuais ou trimestrais sobre contas correntes, balança comercial, dívida pública e déficit/superávit do governo. As amostras variam de país para país, mas normalmente vão de 1973 à 2006. Os resultados indicam que preditores fundamentalistas tradicionais apresentam eficácia limitada no curto prazo, enquanto variáveis como a regra de Taylor e ativos estrangeiros líquidos mostram alguma previsibilidade em modelos lineares mais simples. Contudo, superar o modelo RW sem *drift* continua sendo um desafio, sem evidências consistentes de superioridade preditiva fora da amostra. Assim, as conclusões de Meese e Rogoff (1983) permanecem relevantes, sugerindo que a previsibilidade cambial é ocasional e de curta duração.

Em estudo aplicado no Brasil, Perdomo e Botelho (2007) revisitam a conclusão de Meese e Rogoff (1983) sobre a superioridade do modelo RW na previsão da taxa de câmbio. Tal estudo compara a eficácia preditiva do RW com as projeções de instituições financeiras e consultorias listadas no *ranking* Top-5 do BCB em três horizontes: curto (até um mês), médio (até quatro meses) e longo prazo (até doze meses). Os resultados indicam que o modelo RW superou o poder preditivo até mesmo dos modelos utilizados pelas instituições com melhor desempenho. Adicionalmente, constatou-se que a superioridade do modelo RW foi ainda mais pronunciada em horizontes estendidos. Assim, ressalta-se que, apesar dos resultados da literatura, agentes econômicos ainda utilizam métodos menos precisos, confiando em supostas vantagens informacionais.

No mesmo sentido, Hadad (2015) investiga no Brasil se modelos baseados em fundamentos macroeconômicos podem superar o RW na previsão da taxa de câmbio. Utilizando dados trimestrais de oito países analisados em pares, o estudo emprega modelos VAR, VECM e correção de viés para projeções de 1 a 6 trimestres, abrangendo períodos de crise e estabilidade. O teste de Diebold-Mariano foi aplicado para avaliar a superioridade preditiva entre os modelos. Concluiu-se que o modelo RW sem *drift* se mostrou o *benchmark* mais difícil de superar, especialmente até três trimestres, enquanto o modelo com *drift* foi facilmente superado. Além disso, os melhores modelos multivariados foram aqueles que não incluíram agregados monetários e taxas de juros internas e externas.

Kopp (2019) analisou o comportamento do mercado de câmbio brasileiro, testando a hipótese de Meese-Rogoff, descrita anteriormente. Utilizando dados mensais e trimestrais do período de janeiro de 2000 a julho de 2018, o estudo comparou previsões geradas por modelos de equilíbrio externo e pela regra de Taylor. Os resultados corroboram a hipótese de Meese-Rogoff, indicando que nenhum dos modelos econômicos superam a abordagem do pas-

seio aleatório em estimações fora da amostra para a taxa de câmbio.

Almeida (2022) analisou o comportamento da taxa de câmbio nominal do Real frente ao Dólar durante a pandemia de COVID-19, investigando possíveis desvios dos fundamentos globais. Utilizando 199 observações mensais desde janeiro de 2006, aplicou Mínimos Quadrados Ordinários e o Filtro de Kalman para estimar mudanças estruturais nos coeficientes das variáveis explicativas (CDS, DXY, CRY, VIX, SELIC e NTN-B). Observou-se que, a partir de 2020, houve alterações significativas nesses coeficientes, indicando uma modificação na relação econométrica entre as variáveis. A comparação entre modelos preditivos revelou perda de capacidade preditiva dos *benchmarks* tradicionais.

Genin (2023) investiga o poder preditivo de modelos de projeção da taxa de câmbio para cinco moedas em relação ao Dólar (Japão, Brasil, Colômbia, Chile e México) no período de 1995 a 2020, com dados trimestrais do Fundo Monetário Internacional e da Bloomberg. Foram utilizadas 104 observações trimestrais, reservando 75% dos dados para treinamento e 25% para teste. Inicia-se o estudo com os modelos estruturais clássicos testados por Meese e Rogoff (1982, 1983), através de modelos VAR e VECM, utilizando-se múltiplos critérios de seleção, incluindo medidas de função perda e de direção de mudança, esta última relacionada ao acerto na direção do ativo ao longo do tempo. Para considerar a inovação, incorpora-se a variável CDS de 5 anos como proxy da incerteza. Os resultados mostram que não há um modelo que seja universalmente superior nas projeções de um a seis trimestres, mas destacam a eficácia dos modelos monetários baseados em fundamentos em comparação com o RW a partir do segundo trimestre. O estudo explora também o uso de redes neurais como abordagem não linear. Neste caso, o horizonte de previsão foi definido em um trimestre. Aponta-se a superioridade preditiva da arquitetura LSTM em relação ao modelo RW, em termos de erros quadráticos médios, em todos os países analisados. Considera-se que uma explicação sugerida para o Puzzle de Meese e Rogoff (1983) seria a não-linearidade do processo gerador de dados. Contudo, argumenta-se que o modelo LSTM apresenta limitações na presença de *outliers* causados por eventos imprevisíveis, tais como a COVID-19.

Makika (2022) investiga a previsibilidade das taxas de câmbio utilizando redes neurais profundas em comparação com modelos lineares tradicionais. Foram aplicados modelos lineares ARIMA e ARFIMA e não-lineares Multilayer Perceptron (MLP), LSTM e GRU para prever as taxas diárias de câmbio Real/Dólar e Euro/Dólar com dados diários de dezembro de 2003 a maio de 2021, com horizontes de previsão dos preços diários de 1 e 7 passos à frente. A avaliação do desempenho utilizou MAE, MSE, RMSE e acurácia direcional. Os resultados demonstraram que os modelos não-lineares superaram os lineares em ambos os horizontes, sendo a estrutura GRU a mais eficaz em termos de erro e precisão preditiva.

Marins (2024) avalia a capacidade de previsão de curto prazo (um e três meses) do modelo de opções implícitas (*Option-Implied model*) para o Brasil e outros 13 países emergentes. As moedas investigadas são da Índia, México, Rússia, Chile, Singapura, África do Sul, Turquia, Colômbia, Malásia, Indonésia, Israel, Filipinas e Tailândia. O estudo compara modelos

*forward-looking* baseados em opções cambiais com modelos econômicos (regra de Taylor, modelos monetários e paridade de poder de compra) baseados em dados históricos e o modelo RW, no período de maio de 2008 a janeiro de 2022. Os resultados indicam que, em termos de RMSE, o modelo de opções implícitas supera estatisticamente os modelos econômicos e o modelo RW em quase todos os casos, apresentando as maiores reduções de erro para a maioria dos países, variando entre 21% e 32% para o horizonte de um mês à frente e entre 15% e 37% para o horizonte de três meses à frente. Utiliza-se o teste Diebold e Mariano (1995) com ajuste de West (1996) para testar a diferença entre as previsões dos modelos e as do RW.

Türkoğlu, Göççek e Yumrukuz (2024) avaliaram a eficácia de modelos de previsão para as flutuações da taxa de câmbio entre o Dólar e a Lira turca (USD/TRY), utilizando modelos como *Support Vector Machine* (SVM), XGBoost, LSTM e GRU com conjuntos de 96 e 21 variáveis explicativas (VIX, CDS, *Overnight Interest Rate*, US10Y *Treasury Yield*, entre outras). O estudo, que usou dados de 01/01/2010 a 30/04/2024 provenientes de fontes como Bloomberg e BDDK (Agência de Regulação e Supervisão Bancária da Turquia), concluiu que os modelos LSTM e GRU superaram os demais, com o GRU apresentando a maior precisão preditiva. A pesquisa evidenciou que modelos baseados em redes neurais são mais eficazes na captura das dinâmicas complexas da taxa de câmbio, ao contrário do SVM, que teve um desempenho inferior com dados de alta dimensionalidade, e do XGBoost, que apresentou poder preditivo moderado. Esses achados destacam a importância da escolha adequada de modelos e variáveis em previsões de séries temporais financeiras. Esses resultados contribuíram para o uso dos modelos LSTM e GRU neste estudo.

Siami-Namini, Tavakoli e Namin (2019) analisaram o desempenho dos modelos ARIMA, LSTM e BiLSTM na previsão de séries temporais de ações e índices de bolsa de valores, investigando se a estrutura bidirecional do BiLSTM melhora a precisão preditiva. Utilizando dados de séries temporais no período de 1985 a agosto de 2018, o estudo comparou a capacidade dos modelos de capturar padrões complexos e avaliou a influência do treinamento bidirecional. Os resultados indicaram que o BiLSTM superou o LSTM e o ARIMA, melhorando a acurácia preditiva em 37,78%, embora com maior tempo de convergência. Conclui-se que o BiLSTM é mais eficiente na modelagem preditiva, especialmente para séries temporais mais complexas. No mesmo sentido, García, Guijarro e Oliver (2024) realizaram uma análise comparativa entre modelos LSTM e BiLSTM para prever taxas de câmbio. Utilizando redes neurais recorrentes, previram os preços de fechamento diários de diversas moedas estrangeiras e do Bitcoin, com base em dados de 18/12/2017 a 16/01/2024. Os resultados em geral apontam evidências de superioridade do BiLSTM em relação ao LSTM na previsão de séries temporais financeiras, com redução nas métricas de erro MAE, MAPE e RMSE. Tais evidências motivaram a inclusão dos modelos BiLSTM e BiGRU neste estudo.

Ao integrar avanços recentes da literatura à previsão da taxa de câmbio Ptax (BRL/USD) no Brasil, este estudo amplia o escopo das pesquisas nacionais ao aplicar modelos de DL (LSTM, GRU e variantes bidirecionais) em múltiplos horizontes de curto prazo, comparando-os

ao modelo RW, conforme a tradição estabelecida por Meese e Rogoff (1983). A incorporação de técnicas capazes de capturar relações não lineares e dependências temporais complexas, em alternativa aos modelos econométricos tradicionais, aliada ao uso de métricas e de testes estatísticos robustos — como os de Diebold-Mariano Modificado (1997), de racionalidade de Rossi e Sekhposyan (2016), de flutuação de Giacomini e Rossi (2010) —, fortalece a avaliação comparativa ao permitir verificar a significância estatística das diferenças de desempenho entre os modelos, inclusive ao longo do tempo. Essa abordagem evidencia o potencial superior dos modelos de DL frente às abordagens tradicionais e oferece uma estrutura metodológica replicável e relevante para economias emergentes marcadas por elevada volatilidade cambial.

Os modelos de DL ou aprendizado profundo são compostos por redes neurais avançadas que possuem várias camadas de processamento, permitindo alcançar elevados níveis de abstração. Para Ciaburro e Venkateswaran (2017, pag. 7) essas redes neurais são amplamente aplicadas em tarefas complexas, como o reconhecimento e a classificação de imagens, além da identificação de escrita à mão, entre outras. Assim, esses modelos computacionais, inspirados no funcionamento do cérebro humano, são capazes de aprender e realizar tarefas complexas através do ajuste de pesos em suas conexões internas.

No contexto de redes neurais recorrentes ou *recurrent neural networks* (RNN), modelos como o LSTM e o GRU são projetados para lidar com dados sequenciais e temporais, incorporando mecanismos de memória de longo prazo que lhes permitem capturar dependências históricas nas séries temporais financeiras que são especialmente úteis na atividade de previsão. Diante do cenário econômico atual, marcado por elevada volatilidade na taxa de câmbio, este estudo busca avaliar a eficácia desses modelos e de suas variantes bidirecionais. Considerando a relevância do modelo RW na literatura, ele será utilizado como *benchmark* na análise comparativa de desempenho com os modelos baseados em DL. A seguir, serão apresentados detalhes adicionais sobre os modelos mencionados, bem como os principais fundamentos dos modelos de DL, métricas de erro de previsão e testes estatísticos relevantes para comparação dos desempenhos preditivos.

## 2 FUNDAMENTOS TEÓRICOS

### 2.1 Modelo *Random Walk* (Passeio Aleatório)

Rogoff e Stavrakeva (2008) revisitaram a literatura de Meese e Rogoff (1982, 1983) e apontaram que, na prática, o modelo de RW simples é uma alternativa muito mais difícil de vencer do que o modelo RW com *drift*. De acordo com o RW sem *drift*, o melhor preditor das taxas de câmbio de amanhã, ou de qualquer dia no futuro ( $s_{t+h}$ ), é a taxa de câmbio de hoje ( $s_t$ ). Portanto, as variações nas taxas de câmbio são completamente imprevisíveis, de forma que:

$$E_t[s_{t+h} - s_t] = 0, \quad (1)$$

onde  $E_t$  representa o valor esperado condicional às informações na data  $t$ .

Uma série  $Y_t$  descreve um RW sem *drift* se:

$$Y_t = Y_{t-1} + u_t, \quad (2)$$

onde  $u_t$  é ruído branco com variância  $\sigma^2$ .

### 2.2 Long Short-Term Memory Unit (LSTM) e Bidirectional Long Short-Term Memory Unit (BiLSTM)

Hochreiter e Schmidhuber (1997) propuseram o modelo LSTM como uma solução para as limitações das redes neurais recorrentes tradicionais, que enfrentavam dificuldades em aprender dependências de longo prazo devido ao problema do gradiente que decresce rapidamente ("*vanishing gradient*"). Para resolver isso, os autores introduziram unidades de memória com um mecanismo chamado "*constant error carousel*" (CEC), que permite manter o fluxo do gradiente ao longo do tempo. Além disso, o modelo incorporou portas multiplicativas que controlam a entrada, saída e retenção de informações na memória. A pesquisa utilizou dados artificiais para testar a eficácia do LSTM em comparação com métodos como "*backpropagation through time*" (BPTT) e redes de Elman, demonstrando um desempenho superior do mesmo ao aprender sequências longas de forma mais eficiente.

Joseph e Tackes (2024, p. 310) descrevem a arquitetura das redes neurais LSTM inspiradas em portas lógicas computacionais. Nessa arquitetura, as células de memória servem para armazenar informações de longo prazo e contam com três portas principais que regulam o fluxo de informação: a porta de entrada (*input gate*), que controla quanta informação será lida do estado oculto (*hidden state*) anterior e da entrada atual; a porta de esquecimento (*forget gate*), que decide quanta informação deve ser descartada da memória de longo prazo; e a porta de

saída (*output gate*), que define quanto do estado da célula atual (*cell state*) será utilizado para gerar o estado oculto atual, que representa a saída da célula. Essas portas são representadas matematicamente por meio das seguintes equações:

$$I_t = \sigma(W_{xi}x_t + W_{hi}H_{t-1} + b_i), \quad (3)$$

$$F_t = \sigma(W_{xf}x_t + W_{hf}H_{t-1} + b_f), \quad (4)$$

$$O_t = \sigma(W_{xo}x_t + W_{ho}H_{t-1} + b_o), \quad (5)$$

em que  $\sigma$  representa a função sigmoide,  $W_{xi}, W_{xf}, W_{xo}, W_{hi}, W_{hf}, W_{ho}$  são os pesos treinados, e  $b_i, b_f, b_o$  são os vieses treinados.

Introduz-se uma nova memória de longo prazo denominada estado da célula ou *cell state* ( $C_t$ ), cujas informações são atualizadas e esquecidas através das três portas citadas anteriormente. Considerando-se o estado da célula no instante anterior ( $C_{t-1}$ ), então a célula LSTM calcula um estado de célula candidato ( $\tilde{C}_t$ ), usando outra porta, mas desta vez com função de ativação tangente hiperbólica:

$$\tilde{C}_t = \tanh(W_{xc}x_t + W_{hc}H_{t-1} + b_c), \quad (6)$$

sendo  $W_{xc}$  e  $W_{hc}$  os parâmetros de pesos treinados, e  $b_c$  o parâmetro de viés treinado.

Dessa forma, a equação principal de atualização, que atualiza o estado da célula relacionado à memória de longo prazo da célula fica:

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t, \quad (7)$$

onde  $\odot$  representa produto de Hadamard, que significa a multiplicação elemento a elemento.

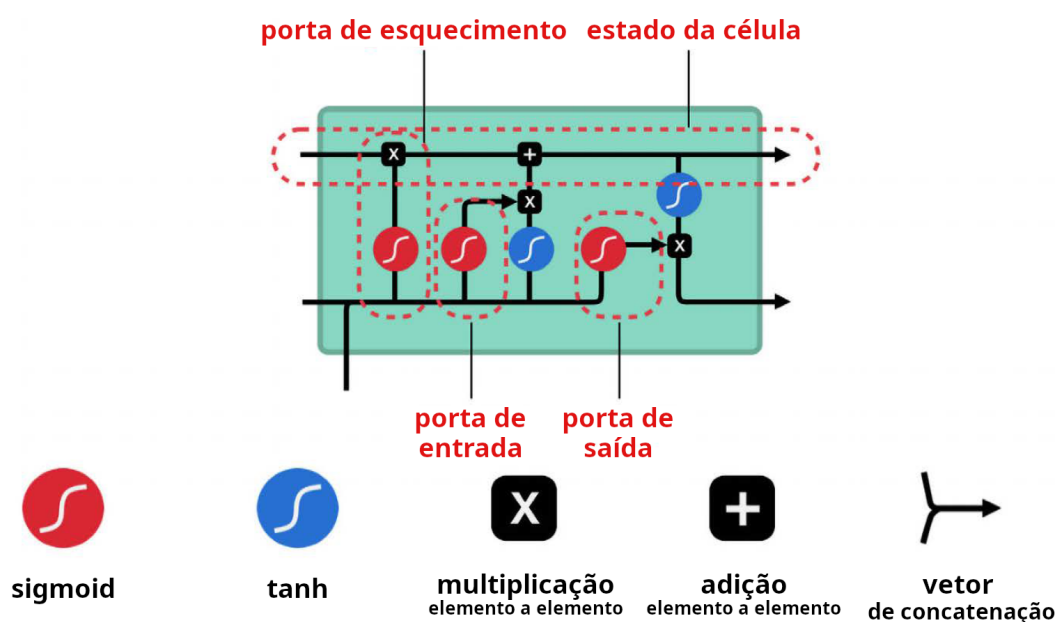
Utiliza-se a porta de esquecimento para decidir quanta informação do passo anterior será mantida, e a porta de entrada para decidir quanto do estado candidato atual será escrito na memória de longo prazo.

Por fim, usa-se o estado atual da célula ( $C_t$ ) recém-atualizado e a porta de saída para decidir quanta informação será repassada ao preditor por meio do estado oculto atual:

$$H_t = O_t \odot \tanh(C_t). \quad (8)$$

Dessa forma, a arquitetura LSTM possibilita a preservação de informações por longos períodos sem sofrer degradação, sendo amplamente utilizada em aplicações de séries temporais e processamento sequencial. Vale ressaltar que a função de ativação adotada neste estudo foi a *Rectified Linear Unit* (ReLU), que será descrita posteriormente. No entanto, a lógica apresentada para a arquitetura do modelo LSTM permanece inalterada. Para uma melhor ilustração, observa-se, a seguir, o diagrama do modelo LSTM.

Figura 1 - Diagrama do modelo LSTM.



**Nota:** Representação esquemática da arquitetura do modelo LSTM.

**Fonte:** JOSEPH; TACKES, 2024, p. 311. Tradução do autor.

De acordo com Schuster e Paliwal (1997), as Redes Neurais Recorrentes Bidirecionais (BRNNs) foram desenvolvidas para superar as limitações das RNNs tradicionais, permitindo que a rede utilize informações do passado e do futuro dentro de um intervalo de tempo. Para isso, a BRNN divide os neurônios de estado (*hidden units* ou *hidden states*) em dois conjuntos: um responsável pelo processamento na direção do tempo (*forward states*) e outro na direção oposta (*backward states*), sem conexões entre eles. Essa estrutura possibilita minimizar a função objetivo sem a necessidade de atrasos explícitos para incluir informações futuras.

O treinamento da BRNN segue um processo semelhante ao das RNNs convencionais, podendo ser realizado por meio de BPTT, sendo que a atualização dos estados *forward* e *backward* deve ocorrer simultaneamente. Em experimentos comparativos, a BRNN demonstrou desempenho superior a outras arquiteturas em tarefas de regressão e classificação, tendo sido testada tanto em dados artificiais quanto em problemas reais de classificação de fonemas.

Segundo Siarni-Namini, Tavakoli e Namin (2019), as redes neurais profundas BiLSTM são uma extensão dos modelos LSTM tradicionais, onde duas LSTMs processam os dados de entrada: uma no sentido direto e outra no sentido reverso. Essa abordagem tende a melhorar a captura de dependências de longo prazo, aumentando a precisão do modelo.



### 2.3 Gated Recurrent Unit (GRU) e Bidirectional Gated Recurrent Unit (BiGRU)

Joseph e Tackes (2024) descrevem a arquitetura das redes neurais GRU, propostas por Cho, Merriënboer e Gulcehre (2014) como uma variação das RNN que apresenta uma estrutura mais simples que a LSTM. A principal diferença é que a GRU elimina o componente de memória de longo prazo, utilizando apenas o estado oculto (*hidden state*) para propagar informações. Assim, o próprio estado oculto funciona como uma “rodovia de gradientes”. A arquitetura da GRU possui duas portas para regular o fluxo de informações: a porta de reinicialização e a porta de atualização.

A porta de reinicialização (*reset gate*) controla a quantidade de informação do estado oculto anterior a ser considerada na geração do estado oculto candidato no tempo atual, representando-se matematicamente por:

$$R_t = \sigma(W_{xr}x_t + W_{hr}H_{t-1} + b_r) \quad (9)$$

onde os parâmetros  $W_{xr}$  e  $W_{hr}$  são pesos treinados, enquanto  $b_r$  é o viés treinado.

Por sua vez, a porta de atualização (*update gate*) define o quanto do estado oculto anterior deve ser mantido e quanto do estado oculto candidato atual será incorporado no novo estado oculto, descrevendo-se por:

$$U_t = \sigma(W_{xu}x_t + W_{hu}H_{t-1} + b_u), \quad (10)$$

onde os parâmetros  $W_{xu}$  e  $W_{hu}$  são pesos treinados, enquanto  $b_u$  é o viés treinado.

Dessa forma, pode-se calcular o estado oculto candidato  $\tilde{H}_t$  da seguinte forma:

$$\tilde{H}_t = \tanh(W_{xh}x_t + W_{hh}R_t \odot H_{t-1} + b_h), \quad (11)$$

onde  $W_{xh}$  e  $W_{hh}$  são pesos treinados e  $b_h$  é o viés treinado. Utiliza-se a porta de reinicialização para regular o fluxo de informação do estado oculto anterior para o estado oculto candidato atual.

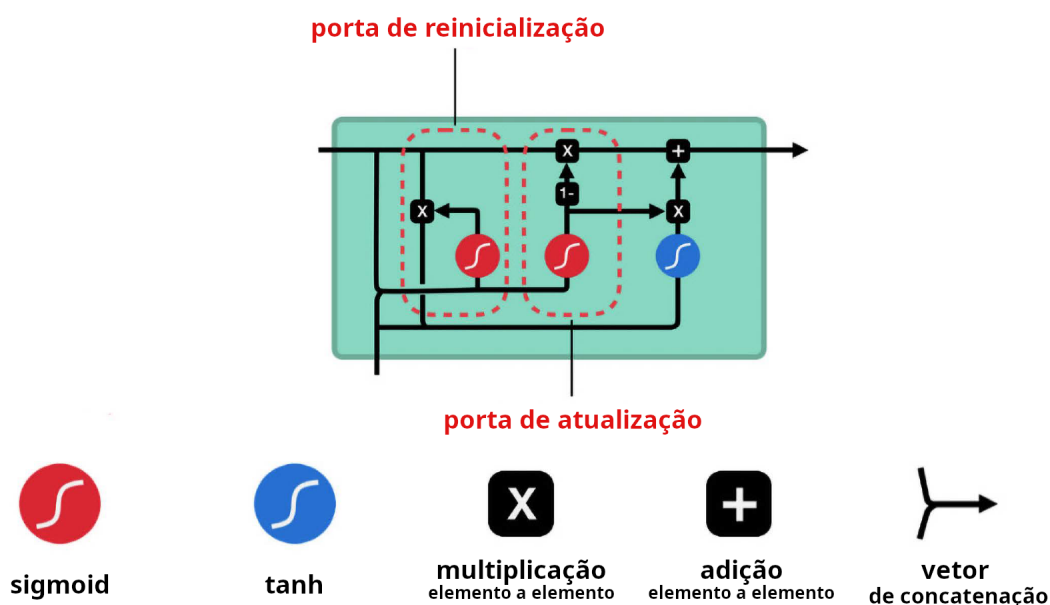
Assim, o estado oculto atual  $H_t$  (a saída que vai para o preditor) é calculado através da seguinte equação:

$$H_t = U_t \odot H_{t-1} + (1 - U_t) \odot \tilde{H}_t. \quad (12)$$

Basicamente, utiliza-se a porta de atualização (*update gate*) para determinar quanto do estado oculto anterior e quanto do estado oculto candidato atual será transmitido para o próximo instante de tempo ou para o preditor. Dessa forma, a rede GRU reduz a complexidade computacional em relação à LSTM, mantendo um desempenho eficiente para modelagem de séries temporais. Para uma melhor ilustração, pode-se observar, a seguir, o diagrama do modelo

GRU.

Figura 2 - Diagrama do modelo GRU.



**Nota:** Representação esquemática da arquitetura do modelo GRU.

**Fonte:** JOSEPH; TACKES, 2024, p. 313. Tradução do autor.

De forma equivalente ao apontando por Siarni-Namini, Tavakoli e Namin (2019) para o modelo BiLSTM, as redes neurais profundas BiGRU são uma extensão dos modelos GRU tradicionais, onde duas GRUs processam os dados de entrada: uma no sentido direto e outra no sentido reverso. Essa abordagem tende a melhorar a captura de dependências de longo prazo, aumentando a precisão do modelo. Assim, valem as mesmas considerações já descritas para as redes neurais recorrentes bidirecionais apontadas para o modelo BiLSTM.

## 2.4 Conceitos relevantes aplicados ao *Deep Learning*

### 2.4.1 Funções de ativação

Joseph e Tackes (2024, p. 281-283) estabelecem um paralelo entre as funções de ativação e os neurônios biológicos em que o axônio avalia as entradas para decidir se emitirá um sinal. De forma semelhante, as funções de ativação desempenham um papel crucial nas redes neurais, pois são responsáveis por permitir que essas redes modelem padrões não lineares. Elas são funções diferenciais não lineares que transformam um espaço vetorial não linearmente separável em um espaço linearmente separável — transformação fundamental para lidar com dados complexos.

A função sigmoide, também conhecida como função logística, é uma das funções de ativação mais antigas e amplamente utilizadas. Ela comprime os valores de entrada em um intervalo entre 0 e 1. No entanto, apresenta um problema conhecido como saturação: quando os neurônios operam em regiões planas da função, os gradientes tornam-se próximos de zero,

o que compromete o processo de aprendizado. Por essa razão, a função sigmoide é raramente utilizada em camadas ocultas de redes profundas, sendo mais comum em camadas de saída.

A função tangente hiperbólica (*tanh*) compartilha limitações semelhantes, pois também pode saturar para valores de entrada grandes positivos ou negativos, levando ao problema do *vanishing gradient* e dificultando o treinamento de redes neurais profundas.

Por outro lado, as funções de ativação ReLU tornaram-se um padrão para redes neurais profundas. Definida como  $g(x) = \max(x, 0)$ , essa função linear mantém valores positivos e anula valores negativos. Suas principais vantagens incluem baixo custo computacional, rápida convergência no treinamento e promoção da esparsidade em que uma parte dos neurônios pode ser desativada. A função de ativação ReLU também tem desvantagens, pois pode sofrer com o problema de "neurônios mortos", nos quais gradientes nulos impedem o aprendizado. Além disso, a média da saída tende a ser positiva, podendo introduzir viés nas camadas empilhadas. Porém, pelas suas vantagens de não saturação em relação a função de ativação tangente hiperbólica, a função ReLU será adotada neste estudo. Como a ReLU não tende gradualmente a um limite, ela evita o problema do *vanishing gradient*, tornando o treinamento de redes neurais profundas mais eficiente.

#### 2.4.2 Gradiente descendente

O gradiente descendente (*gradient descent*) é um algoritmo de otimização usado para minimizar uma função de custo, movendo-se iterativamente na direção do maior declive negativo. O tamanho do passo em cada iteração é regulado pela taxa de aprendizado (*learning rate*)  $\eta$ , que equilibra a estabilidade e a velocidade de convergência do modelo. O  $\theta$  é o vetor de parâmetros e a cada iteração, atualizam-se os parâmetros pela seguinte fórmula:

$$\theta_{i+1} = \theta_i - \eta \times \nabla f(a, b), \quad (13)$$

em que  $\nabla$  representa o gradiente da função.

#### 2.4.3 Overfitting e underfitting e conceitos relacionados ao treinamento dos modelos de DL

No contexto do aprendizado de máquinas, os fenômenos de *overfitting* e *underfitting* constituem um grande desafio a ser superado. Há a necessidade de que o modelo de DL, ao processar novos dados, apresente uma performance adequada. Enquanto a otimização matemática busca o máximo global nos dados disponíveis, o aprendizado de máquina visa minimizar o erro em um novo conjunto de dados usando o erro de treinamento como referência.

A regularização é apresentada como uma técnica para reduzir a complexidade do mo-

delo, favorecendo funções mais simples dentro do espaço de hipóteses, frequentemente por meio do decaimento de pesos. Podem-se citar os regularizadores L1 e L2, que adicionam termos à função de perda ( $L$ ) para penalizar pesos elevados, evitando o sobreajuste. Além disso, destaca-se o uso do *dropout*, que consiste em desligar aleatoriamente uma fração dos neurônios durante o treinamento, forçando o modelo a não depender excessivamente de unidades específicas e, assim, contribuindo para uma melhor capacidade de generalização.

Por outro lado, o *Early Stopping* é uma técnica que monitora a perda na validação e interrompe o treinamento quando há degradação do objetivo de minimização da função de perda. O parâmetro *patience* define quantas épocas sem melhora são toleradas antes de encerrar o treinamento. Tal procedimento foi utilizado neste estudo empírico de forma a evitar *overfitting*, em que o modelo perde a capacidade de generalização ao se ajustar demais aos dados de treino. No treinamento de redes neurais, uma etapa de treinamento representa uma única atualização dos parâmetros via gradiente, sendo que, no gradiente descendente estocástico em lote, essa atualização ocorre após cada lote de dados. O lote (*batch*) corresponde ao número de amostras processadas pelo modelo antes de calcular e atualizar os gradientes. Já uma época (*epoch*) ocorre quando o modelo percorre todos os dados do conjunto de treinamento, garantindo que todas as amostras tenham sido utilizadas para atualização dos parâmetros.

## 2.5 Métricas de erro

De forma geral, o erro corresponde à diferença entre o valor efetivamente observado da variável e o valor previsto pelo modelo no período  $t$ , tal que:  $e_t = (Y_t - \hat{Y}_t)$ .

### 2.5.1 Erro Médio Absoluto ou *Mean Absolute Error* (MAE)

Mede-se a magnitude média dos erros absolutos, sendo útil para entender o tamanho médio das discrepâncias. É ideal para comparar previsões em séries temporais na mesma escala.

$$MAE = \frac{1}{H} \sum_{t=1}^H |e_t|, \quad (14)$$

com  $|e_t|$  correspondente ao valor absoluto de  $e_t$ .

### 2.5.2 Erro Quadrado Médio ou *Mean Squared Error* (MSE)

O MSE penaliza os erros maiores devido ao uso do quadrado, destacando desvios significativos. É útil especialmente quando erros grandes precisam ser minimizados.

$$MSE = \frac{1}{H} \sum_{t=1}^H e_t^2. \quad (15)$$

### 2.5.3 Raiz do Erro Quadrado Médio ou *Root Mean Squared Error* (RMSE)

Fornecer uma interpretação mais intuitiva do erro, retornando valores na mesma unidade dos dados originais. Assim como o MSE, penaliza erros maiores, mas é mais útil na comparabilidade.

$$RMSE = \sqrt{\frac{1}{H} \sum_{t=1}^H e_t^2}. \quad (16)$$

### 2.5.4 Erro Médio Absoluto Percentual ou *Mean Absolute Percent Error* (MAPE)

Enquanto o erro absoluto e o erro quadrado são dependentes da escala, o erro percentual é uma medida de erro independente da escala.

$$MAPE = \frac{100}{H} \sum_{t=1}^H \frac{|e_t|}{y_t}. \quad (17)$$

Avalia-se o erro como uma porcentagem das observações reais, tornando-se independente da escala. Ideal para comparar séries temporais de diferentes escalas ou unidades.

## 2.6 Testes estatísticos

O teste de Diebold e Mariano (1995) é amplamente utilizado para comparar previsões de diferentes modelos, verificando se suas projeções são estatisticamente equivalentes ou apresentam diferenças significativas, sem necessariamente avaliar os modelos em si. Trata-se de uma ferramenta fundamental na análise comparativa da precisão preditiva.

Como a taxa de câmbio está sujeita a choques de diversas ordens, é essencial avaliar não apenas a precisão das previsões, mas também sua consistência ao longo do tempo. Assim, este estudo investiga também a racionalidade e estabilidade das previsões cambiais, aplicando os testes de racionalidade de Rossi e Sekhposyan (2016) e o teste de flutuação de Giacomini e

Rossi (2010).

Basicamente, o teste de flutuação de Giacomini e Rossi (2010) identifica possíveis mudanças na qualidade preditiva relativa ao longo do tempo entre dois modelos, detectando períodos em que as previsões podem se tornar relativamente menos confiáveis. Por sua vez, o teste de racionalidade Rossi e Sekhposyan (2016) examina a racionalidade das previsões, avaliando se os erros de previsão são imprevisíveis no sentido de que as previsões não tenham vieses e sejam eficientes, o que pode indicar ausência de choques estruturais e adequação do modelo para explicar o comportamento da taxa de câmbio em dada janela de tempo. Nesses dois últimos testes foram definidas janelas móveis de 20 dias úteis.

Por outro lado, para garantir uma avaliação robusta, utilizamos previsões diretas h-passos à frente, realizadas tanto no período de validação quanto no período de teste. Essa abordagem minimiza a propagação de erros acumulados, permitindo uma análise mais precisa da estabilidade e racionalidade das projeções ao longo do tempo.

### 2.6.1 Teste de Diebold-Mariano (1995)

O teste de Diebold e Mariano (1995) é utilizado para comparar previsões feitas por diferentes modelos, embora não analise os modelos propriamente ditos. Seu objetivo é verificar se as previsões fora da amostra de dois modelos são estatisticamente equivalentes ou apresentam diferenças significativas. Trata-se de um método simples de aplicar, mas considerado eficaz para realizar comparações entre projeções.

Conforme Diebold (2015), basicamente, o teste de DM avalia a igualdade de acurácia preditiva entre dois modelos, considerando diretamente os erros de previsão. O diferencial de perda em  $t$  entre os modelos 1 e 2 é definido como  $d_{12t} = L(e_{1t}) - L(e_{2t})$ , onde  $L(e_t)$  representa a função de perda como, por exemplo o erro quadrático. O teste tem como condição suficiente, mas não estritamente necessária, que  $d_{12,t}$  seja covariante estacionário. A estatística DM converge assintoticamente para a distribuição normal quando a suposição de estacionariedade é válida. Assume-se que  $E(d_{12t}) = 0$  sob a hipótese nula ( $H_0$ ) de acurácia equivalente entre as previsões contra a hipótese alternativa ( $H_A$ ) de que os modelos têm diferentes acurácias preditivas.

Pode-se testar a perda diferencial para todo o horizonte amostral  $T$  e avaliar a estatística  $DM_{12}$  resultante. Seja  $D_{12}$  a média amostral das perdas diferenciais, tal que  $D_{12} = \frac{1}{T} \sum_{t=1}^T d_{12t}$ , e  $\hat{\sigma}_{D_{12}}$  um estimador consistente do desvio padrão de  $D_{12}$ , então:

$$DM_{12} = \frac{D_{12}}{\hat{\sigma}_{D_{12}}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (18)$$

Adicionalmente, nos mesmos moldes que os trabalhos de Rossi (2013) e Saba (2015), será adotada no teste DM o estimador de *Newey-West* para a variância dos erros de forma a corrigir

possíveis problemas de heterocedasticidade e autocorrelação nos erros, o que torna o método mais robusto. Esta versão será chamada neste estudo simplesmente de teste DM.

### 2.6.2 Teste de Diebold-Mariano Modificado (1997)

Harvey, Leybourne e Newbold (1997) sugerem que melhores propriedades em amostras pequenas podem ser obtidas por meio de uma correção de viés na estatística do teste DM e de uma comparação da estatística corrigida com uma distribuição t de Student com  $(T - 1)$  graus de liberdade, em vez da distribuição normal padrão.

Iquiapaza, Bressan e Amaral (2010) descrevem que como as propriedades do teste DM são prejudicadas, especialmente, nos casos de não normalidade, Harvey, Leybourne e Newbold (1997) adaptaram-o para a distribuição t de Student:

$$DMm = \left[ \frac{T + 1 - 2h + T^{-1}h(h - 1)}{T} \right]^{1/2} DM \sim t_{T-1},$$

onde  $h$  é o número de passos à frente utilizados na previsão,  $T$  é o tamanho da amostra e  $DMm$  é a estatística de teste do teste Diebold-Mariano Modificado (DMM).

### 2.6.3 Teste de flutuação de Giacomini e Rossi (2010)

O teste de flutuação de Giacomini e Rossi (2010) investiga o desempenho preditivo local, fundamentando-se na premissa de que a qualidade preditiva dos modelos pode variar substancialmente ao longo do tempo. Os autores argumentam que medidas agregadas, como o erro absoluto médio (MAE) ou erro quadrático médio (MSE), podem mascarar informações relevantes sobre essa variação.

O teste de flutuação baseia-se na estatística de Diebold e Mariano (1995), aplicada em janelas móveis de tamanho  $m$ . Neste estudo adotou-se  $m = 20$ , aproximadamente um mês. A estatística de teste é calculada ao longo do tempo para detectar mudanças na capacidade preditiva relativa entre dois modelos concorrentes.

Formalmente, conforme resumido em Castro (2023) e Castro e Aiube (2023), define-se a perda relativa local como a sequência de diferenças de perda fora da amostra, calculadas em janelas móveis de tamanho  $m$ . O teste avalia o desempenho preditivo relativo entre modelos concorrentes ao longo do tempo, considerando uma função de perda  $L(\cdot)$  escolhida pelo pesquisador. Seja  $L^{(j)}(\cdot)$  a função de perda associada ao modelo  $j$ , com  $j = 1, 2$ , então a sequência das diferenças de perda fora da amostra é dada por:

$$\Delta L_{t,h} = L_{t,h}^{(1)} - L_{t,h}^{(2)} \quad (19)$$

No caso da perda quadrática associada ao erro médio quadrático de previsão, tem-se:

$$L_{t,h}^{(j)} = e_{t+h}^{(j)2}, \quad (20)$$

onde  $e_{t+h}^{(j)}$  representa os erros de previsão  $h$  passos a frente do modelo  $j$ .

A função de perda quadrada é a função de perda mais amplamente usada na prática, sendo implementada também neste estudo conforme o procedimento descrito abaixo.

Giacomini e Rossi (2010) definiram a perda relativa local como a sequência de diferenças de perda em janelas móveis de tamanho  $m$ :

$$m^{-1} \sum_{j=t-m+1}^t \Delta L_{j,h}, \quad t = m, m+1, \dots, P. \quad (21)$$

A hipótese nula ( $H_0$ ) testada é de mesma capacidade preditiva em cada período no tempo:  $E(\Delta L_{t,h}) = 0, \forall t$ , onde a hipótese alternativa pode ser  $E(\Delta L_{t,h}) \neq 0$  (alternativa bilateral).

Quando se considera a alternativa bilateral, a estatística do teste de flutuação é o maior valor sobre a sequência das perdas relativas redimensionadas:

$$\max_t |F_{t,m}^{OOS}|, \quad \text{onde } F_{t,m}^{OOS} = \hat{\sigma}^{-1} m^{-1/2} \sum_{j=t-m+1}^t \Delta L_{j,h} \quad \text{e } t = m, m+1, \dots, P, \quad (22)$$

com  $\hat{\sigma}^2$  sendo um estimador robusto de heterocedasticidade e autocorrelação (HAC) consistente da variância de longo prazo, conforme Newey e West (1987), e  $P$  correspondente ao tamanho da porção fora da amostra.

Dessa forma, a hipótese nula ( $H_0$ ) é rejeitada se:

$$\max_t F_{t,m}^{OOS} > k_{\alpha,\mu}, \quad (23)$$

onde  $k_{\alpha,\mu}$  é o valor crítico dependente do tamanho da janela móvel  $\mu$  e do número de diferenças de perda fora da amostra.

#### 2.6.4 Teste de racionalidade de Rossi e Sekhposyan (2016)

O teste de racionalidade proposto por Rossi e Sekhposyan (2016) avalia a capacidade preditiva de modelos em ambientes instáveis, expandindo a análise tradicional de viés e eficiência das previsões. Sua principal inovação reside na robustez diante de instabilidades amostrais, permitindo a detecção de quebras de racionalidade ao longo do tempo.

O teste examina a racionalidade das previsões em janelas móveis, identificando períodos específicos de irracionalidade dos erros de previsão que poderiam ser mascarados em análises



agregadas. Diferenciando-se da abordagem de Giacomini e Rossi (2010), que compara o desempenho relativo entre modelos sob instabilidade, o teste de Rossi e Sekhposyan avalia a habilidade preditiva absoluta, levando em conta o erro de estimação dos parâmetros. Aplicações empíricas demonstram sua eficácia, evidenciando, por exemplo, que previsões do Federal Reserve subestimaram consistentemente a inflação nos anos 1970 e a superestimaram nos anos 1980 devido a mudanças estruturais. Embora o uso de janelas móveis possa reduzir o poder estatístico em amostras pequenas, o teste demonstrou superioridade na detecção de irracionalidade localizada dos erros de previsão, tornando-se uma ferramenta avançada para análise preditiva sob instabilidade.

Rossi e Sekhposyan (2014), consideram o erro de previsão associado à previsão de  $h$  passos à frente feita no tempo  $t$ , denotando-o por  $v_{t+h}(\hat{\gamma}_{t,R})$ , em que  $\hat{\gamma}_{t,R}$  representa os parâmetros estimados do modelo. Descreve-se como exemplo para um modelo de regressão linear simples com um vetor de regressores defasados de  $h$  períodos,  $x_t$ , de dimensão  $(k \times 1)$ , onde  $E_t[y_{t+h}] = x_t'\gamma$ , a previsão no tempo  $t$  é dada por  $\hat{y}_{t+h|t} = x_t'\hat{\gamma}_{t,R}$  e o erro de previsão é dado por:

$$v_{t+h}(\hat{\gamma}_{t,R}) = y_{t+h} - x_t'\hat{\gamma}_{t,R}. \quad (24)$$

Esse erro representa a diferença entre o valor efetivo  $y_{t+h}$  e a previsão feita no tempo  $t$ , sendo fundamental para avaliar a qualidade das previsões e testar sua racionalidade. O teste de racionalidade das previsões baseia-se no trabalho de West e McCracken (1998), considerando-se a seguinte regressão geral:

$$v_{t+h}(\hat{\gamma}_{t,R}) = \hat{g}_t' \cdot \theta + \eta_{t+h}, \quad t = R, \dots, T, \quad (25)$$

em que  $\hat{g}_t \equiv g_t(\hat{\gamma}_{t,R})$  é um vetor  $(\ell \times 1)$  baseado nos dados do período  $t$ , que pode ser uma função do estimador  $\hat{\gamma}_{t,R}$ ,  $\theta$  é um vetor de parâmetros  $(\ell \times 1)$ , e  $v_{t+h}(\hat{\gamma}_{t,R})$  representa o erro de previsão estimado.

West e McCracken (1998) formulam a hipótese nula do teste como:  $H_0 : \theta = \theta_0$  vs.  $H_A : \theta \neq \theta_0$ , onde  $\theta_0 = 0$ .

Seja  $\hat{\theta}_P$  a estimativa de  $\theta$  na regressão, o teste de Wald correspondente é definido da seguinte forma:

$$\mathcal{W}_P = P(\hat{\theta}_P - \theta_0)' \hat{V}_{\theta,P}^{-1} (\hat{\theta}_P - \theta_0), \quad (26)$$

onde  $\hat{V}_{\theta,P}$  é um estimador consistente da variância de longo prazo de  $\sqrt{P}\hat{\theta}_P$ .

Para lidar com instabilidades, adota-se uma abordagem de regressão em janelas móveis. Seja  $\hat{\theta}_j$  a estimativa de  $\theta$  na regressão no instante  $j$  sobre janelas de tamanho  $m$ . Neste trabalho foi adotado  $m=20$ , aproximadamente igual ao período de um mês. Assim,  $\hat{\theta}_j$  é estimado sequencialmente para  $j = R + m, \dots, T$  usando as observações mais recentes. O teste de Wald

correspondente é:

$$\mathcal{W}_{j,m} = m\hat{\theta}'_j\hat{V}_\theta^{-1}\hat{\theta}_j, \quad \text{para } j = R+m, \dots, T. \quad (27)$$

Então, o teste de racionalidade é definido como:

$$\max_{j \in \{R+m, \dots, T\}} \mathcal{W}_{j,m}, \quad (28)$$

sendo utilizado para testar a hipótese nula:

$$H_0 : \theta_j = \theta_0 \quad \text{vs.} \quad H_A : \theta_j \neq \theta_0, \quad \forall j = R+m, \dots, T. \quad (29)$$

onde  $\theta_0 = 0$  e  $\theta_j$  representa o valor verdadeiro do parâmetro e  $\hat{V}_{\theta,P}$  é um estimador consistente da variância assintótica de  $\sqrt{P}\hat{\theta}_P$ . Conforme West e McCracken (1998), em casos especiais (como viés de previsão ou eficiência),  $\hat{V}_\theta$  é um estimador HAC da variância assintótica das estimativas dos parâmetros nas janelas móveis, implementado substituindo  $P$  por  $m$  na notação original.

A hipótese nula é rejeitada se  $\max \mathcal{W}_{t,m} > \kappa_{\alpha,\ell}$ , sendo  $\kappa_{\alpha,\ell}$  o valor crítico no nível de significância de  $100\alpha\%$ , com o número de restrições igual a  $\ell$ . O valor crítico para o teste pode ser obtido via simulação de Monte Carlo. Neste estudo, utilizou-se uma aproximação do teste tradicional de Mincer e Zarnowitz (1969), que conforme descrito por Rossi e Soupre (2017), segue uma distribuição assintótica qui-quadrado com  $\ell$  graus de liberdade. Descreve-se ainda que, ao contrário do teste de racionalidade que considera  $\max \mathcal{W}_{t,m}$ , o teste tradicional não é robusto a instabilidades pois foi projetado para avaliar a hipótese nula incondicional.

### 3 METODOLOGIA

Este capítulo descreve os procedimentos metodológicos adotados para a condução do estudo, que abrange desde a coleta e preparação dos dados até a modelagem preditiva e avaliação de desempenho. A pesquisa segue uma abordagem quantitativa, de natureza aplicada e caráter empírico, com foco na avaliação da capacidade preditiva de diferentes arquiteturas de modelos de DL para prever a taxa de câmbio Ptax de venda do Real brasileiro frente ao Dólar americano (BRL/USD), tendo como *benchmark* o modelo RW. O principal objetivo deste estudo é verificar se as previsões de algum dos modelos LSTM, BiLSTM, GRU e BiGRU, aplicados nas arquiteturas e parâmetros especificados, conseguem superar o desempenho do modelo RW, especialmente no período de teste.

Quanto aos procedimentos técnicos adotados, esta pesquisa é classificada como um estudo de caso, por buscar avaliar se a aplicação de modelos de DL sob certas especificações, no contexto do período analisado, é capaz de aprimorar a previsão da taxa de câmbio Ptax de venda, considerando cenários com e sem a inclusão de variáveis econômicas e financeiras como fatores explicativos.

Conforme Prodanov e Freitas (2013), esse procedimento objetiva orientar uma ou mais decisões e identificar seus motivos e resultados. A pesquisa também é classificada como descritiva, pois visa descrever as características de determinado fenômeno e estabelecer relações entre as variáveis. No contexto desta pesquisa, isso se reflete na análise da relação entre as variáveis explicativas e a taxa de câmbio Ptax de venda, além da comparação entre as previsões geradas pelos diferentes modelos.

Para um estudo mais abrangente, adota-se tanto a abordagem univariada, que considera apenas os valores passados da taxa de câmbio Ptax de venda, quanto a abordagem multivariada, que incorpora variáveis econômicas e financeiras (VIX, DXY, CRY, diferencial de juros, CDS de 5 e 10 anos, títulos do Tesouro dos EUA de 5 e 10 anos, índice Ibovespa e S&P 500), além dos valores passados da respectiva taxa de câmbio. Os horizontes de previsão considerados serão de 1, 10 e 20 dias à frente. Além disso, serão avaliadas duas variações de arquitetura nos modelos de DL em cada caso: uma com uma camada recorrente (1CR) e outra com duas camadas recorrentes (2CR), cujas especificações correspondem aos nomes atribuídos aos respectivos modelos de DL (LSTM, GRU, BiLSTM e BiGRU).

#### 3.1 Coleta, preparação dos dados e treinamento

A coleta e preparação dos dados deste estudo seguiram uma abordagem quantitativa, com a utilização do ambiente Google Colab para a implementação dos modelos preditivos e *download* dos dados em formato de planilha do Microsoft Excel. Os dados históricos da taxa

de câmbio Ptax de venda (BRL/USD) foram obtidos por meio da Interface de Programação de Aplicações (API) do Banco Central do Brasil. Além disso, foram coletadas séries temporais de variáveis econômicas e financeiras relacionadas à previsão cambial na plataforma Bloomberg. A base de dados considerada abrange o período de janeiro de 2020 a dezembro de 2024, com frequência diária.

Inicialmente, os dados foram carregados e pré-processados. O índice da série temporal foi convertido para o formato *datetime*, garantindo o correto manuseio temporal. Em seguida, foi realizada a seleção da variável-alvo, denominada "Ptax\_Venda", e das variáveis preditoras a serem utilizadas nos modelos multivariados.

A etapa de limpeza dos dados foi conduzida com rigor, mantendo valores ausentes como NaN, de forma a preservar a integridade das séries temporais e evitar imputações que pudessem introduzir viés. Na sequência, os registros com valores inconsistentes ou nulos foram removidos, assegurando que apenas dados válidos e representativos fossem utilizados no treinamento dos modelos.

As variáveis foram então normalizadas utilizando o método *MinMaxScaler*, que restringe os valores à faixa entre 0 e 1, favorecendo a estabilidade e a convergência dos algoritmos de aprendizado de máquina. Após essa etapa, os dados foram organizados em janelas móveis para formar sequências temporais, permitindo que os modelos capturassem padrões sazonais e de tendência relevantes para a previsão.

Para garantir uma avaliação robusta, o conjunto de dados foi dividido respeitando a ordem cronológica, destinando-se 80% ao treinamento, 10% à validação e 10% ao teste, procedimento semelhante ao adotado em diversos estudos preditivos, como os de Petracca (2024, p. 16) e Mbedzi (2022, p. 24). Tal procedimento está em conformidade com as recomendações de Joseph e Tackes (2024, p. 78), que destacam a importância de manter a estrutura temporal em problemas envolvendo séries temporais. A divisão entre validação e teste também evita o fenômeno conhecido como *data leakage*, no qual o modelo tem acesso a informações que não estariam disponíveis no momento da previsão. Um exemplo clássico de vazamento ocorre ao se aplicar técnicas de escalonamento antes da separação dos dados em treino e teste, comprometendo a avaliação realista do desempenho do modelo. O conjunto de validação permite o ajuste dos hiperparâmetros durante o treinamento e a seleção dos modelos de DL em cada arquitetura, enquanto o conjunto de teste, contendo dados completamente inéditos, é utilizado exclusivamente para avaliar o desempenho final das previsões dos respectivos modelos, simulando-se uma aplicação prática.

A dinâmica de treinamento dos modelos de redes neurais DL seguiu a lógica da otimização iterativa da função de perda com base nos dados de treino, acompanhada pela verificação do desempenho nos dados de validação para evitar o sobreajuste (*overfitting*). Os gráficos das funções de perda de treinamento e de validação dos modelos de DL com menores métricas de erro no período de teste podem ser observados no Apêndice B. O conjunto de teste foi mantido isolado até o momento final, servindo como referência para estimar a capacidade de generalização

dos modelos.

As previsões foram realizadas em três horizontes distintos: 1, 10 e 20 dias úteis à frente. Adotou-se a abordagem de previsão direta (*direct forecasting*), em que um modelo é treinado especificamente para cada horizonte, estimando diretamente o valor futuro desejado, sem recorrer a previsões intermediárias. Essa estratégia tem como objetivo minimizar o acúmulo de erros, sobretudo em horizontes mais longos, sendo recomendada para melhorar a precisão das previsões.

A definição do tamanho das janelas móveis para cada horizonte de previsão baseou-se na minimização das métricas de erro no período de validação, durante os estudos preliminares, em um procedimento voltado a simular uma aplicação prática. Para isso, foram conduzidos testes em que os modelos foram calibrados exclusivamente com informações do conjunto de treinamento e, em seguida, avaliados no conjunto de validação, isto é, nos dados já conhecidos no contexto da simulação prática. Esse critério de calibração permitiu selecionar, de forma sistemática, a configuração de janelas mais adequada para cada horizonte de previsão, antes de qualquer contato com os dados do período de teste. Como resultado, foram definidos tamanhos de janelas de 5, 30 e 40 dias úteis para os horizontes de 1, 10 e 20 dias úteis à frente, respectivamente. A etapa seguinte consistiu em verificar a capacidade de generalização dessa seleção, analisando o desempenho obtido no conjunto de teste.

### 3.2 Modelos Preditivos

O estudo avalia a eficácia de diferentes modelos preditivos, incluindo os modelos de DL LSTM, BiLSTM, GRU e BiGRU, bem como o modelo RW, usado como benchmark. Os modelos de DL foram aplicados tanto na abordagem univariada quanto multivariada. Na abordagem univariada, foi considerada apenas a série histórica da taxa de câmbio Ptax de venda, enquanto na abordagem multivariada, foram incorporadas variáveis econômicas e financeiras de frequência diária descritas na literatura, visando verificar sua contribuição para o aprimoramento das previsões.

Em cada combinação de modelo de DL, abordagem e horizonte de previsão foram adotadas duas arquiteturas distintas: a primeira com uma única camada recorrente (1CR) e a segunda com duas camadas recorrentes empilhadas (2CR). Foram conduzidos testes empíricos preliminares com objetivo de subsidiar a definição básica das arquiteturas, especificações e parâmetros dos modelos de DL. A estruturação dos modelos de DL foi orientada no sentido de redução das métricas de erro de previsão RMSE, MAE e MAPE, além de uma consistência entre os resultados obtidos nos períodos de validação e de teste, visando consistência das previsões e capacidade de generalização. A Tabela 1 descrita na sequência apresenta as configurações estruturais implementadas no estudo.

Tabela 1 - Arquiteturas e parâmetros dos modelos de DL

<b>Arquiteturas:</b>	Modelos de DL com uma camada recorrente ( <b>1CR</b> )	Modelos de DL com duas camadas recorrentes ( <b>2CR</b> )
<b>Camada de entrada</b>		
<b>Primeira camada recorrente</b>	Com 64 unidades, função de ativação ReLU e regularizador <i>kernel</i> L2 (0,01)	Com 64 unidades, função de ativação ReLU e regularizador <i>kernel</i> L2 (0,01)
<b>Segunda camada recorrente</b>	Não possui	Com 32 unidades e função de ativação ReLU
<b>Dropout</b>	Após a camada recorrente ( <i>dropout</i> : taxa = 0,1)	Após a segunda camada recorrente ( <i>dropout</i> : taxa = 0,1)
<b>Camada Densa 1</b>	Com 32 unidades, função de ativação ReLU e regularizador <i>kernel</i> L2 (0,01)	Com 16 unidades, função de ativação ReLU e regularizador <i>kernel</i> L2 (0,01)
<b>Camada Densa 2 (Camada de saída)</b>	1 unidade	1 unidade

**Nota:** A designação atribuída aos modelos de DL corresponde ao tipo de camada recorrente utilizada em cada arquitetura (1CR ou 2CR): LSTM, GRU, BiLSTM ou BiGRU.

**Fonte:** O autor, 2025.

A Tabela 1 resume as arquiteturas dos modelos de DL aplicados na previsão da taxa de câmbio Ptax de venda. Foram adotadas duas arquiteturas com o objetivo de avaliar o impacto da profundidade da rede neural na capacidade preditiva dos modelos.

Basicamente, a Arquitetura com 1CR, após a camada de entrada, é composta por uma única camada recorrente com 64 unidades, função de ativação ReLU e regularização *kernel* L2, seguida de uma camada densa com 32 unidades, função de ativação ReLU e regularização *kernel* L2, seguida por uma camada densa de saída. Já a Arquitetura com 2CR, após a camada de entrada, conta com duas camadas recorrentes: a primeira com 64 unidades e a segunda com 32 unidades, ambas utilizando função de ativação ReLU e regularização *kernel* L2. Além disso, inclui uma camada densa intermediária com 16 unidades, função de ativação ReLU e regularização *kernel* L2, antes da camada densa de saída.

Adicionalmente, cabe observar que ambas as arquiteturas utilizam a função de perda MSE. Os modelos de DL foram treinados com os otimizadores Adam ou AdamW, mantendo uma taxa de aprendizado (*learning rate*) inicial de 0,0001. Foram aplicadas técnicas de regularização como o *dropout* (0,1) e regularização *kernel* L2 (0,01), visando mitigar o risco de sobreajuste (*overfitting*). O *dropout* (0,1) significa que 10% das unidades daquela camada serão desativados de forma aleatória em cada iteração durante o treinamento, reduzindo a dependência do modelo de combinações específicas de unidades. Já o *kernel* L2 (0,01) adiciona um termo de penalização proporcional ao quadrado dos pesos da camada, incentivando o modelo a manter pesos menores e, assim, favorecendo soluções mais simples e com maior capacidade de

generalização.

Por outro lado, durante o treinamento, adotou-se o critério de *Early Stopping* (*patience* = 5 e "*min\_delta*" =  $1e-4$ ), para interromper o treinamento quando não ocorrer melhora significativa na perda de validação. O parâmetro "*min\_delta*" estabelece a variação mínima necessária na métrica monitorada para que uma melhoria seja considerada relevante. Adicionalmente, empregou-se o método "*ReduceLROnPlateau*" para ajuste dinâmico da taxa de aprendizado, configurado com *fator* = 0,5 (reduzindo a taxa pela metade), *patience* = 3 (número de épocas sem melhoria antes da redução) e "*min\_lr*" =  $1e-5$  (taxa de aprendizado mínima permitida).

O uso combinado dessas estratégias buscou promover uma estabilidade no processo de treinamento, boa capacidade de generalização e consistência das previsões nos diferentes horizontes temporais considerados.

### 3.3 Avaliação do Desempenho

A avaliação do desempenho será conduzida em duas etapas. Primeiramente, no período de validação, será selecionado o melhor modelo de DL com base nas menores métricas de erro desse período. Em seguida, esse modelo será comparado ao modelo RW no período de teste — fase em que os dados são inéditos para os modelos — simulando-se uma aplicação prática. Essa abordagem permite avaliar tanto a eficácia da seleção do modelo de DL quanto sua capacidade de generalização fora da amostra, além de possibilitar a comparação direta com o desempenho do modelo RW.

O desempenho preditivo dos modelos será mensurado pelas métricas de erro RMSE, MAE e MAPE. Adicionalmente, para cada abordagem e horizonte de previsão, realizam-se testes estatísticos entre o modelo RW e o modelo de DL selecionado, com melhor desempenho no período de teste, entre as arquiteturas aplicadas. O objetivo é verificar se o modelo de DL selecionado com melhor desempenho preditivo consegue superar, de forma estatisticamente significativa, a capacidade preditiva do modelo RW no período de teste. Dessa forma, a combinação das métricas de erro e dos testes estatísticos garantem uma análise abrangente e robusta da eficácia das previsões dos modelos avaliados.

Na comparação das previsões dos modelos selecionados de DL com o modelo RW serão utilizados quatro testes estatísticos. O teste de Diebold-Mariano (1995) será aplicado para verificar a existência de diferenças estatisticamente significativas entre as previsões dos modelos. O teste de Diebold-Mariano Modificado (1997) será utilizado como uma extensão do teste Diebold-Mariano, incorporando ajustes para o tamanho finito da amostra, tornando a comparação mais robusta. O teste de flutuação de Giacomini e Rossi (2010) será empregado para examinar se a diferença de acurácia preditiva entre os modelos varia ao longo do tempo, permitindo a detecção de possíveis instabilidades no desempenho. Por fim, o teste de racionalidade de Rossi e Sekhposyan (2016) será utilizado para avaliar se as previsões são eficientes e não

viesadas, identificando a presença de padrões sistemáticos nos erros, o que indicaria potencial de melhoria na capacidade preditiva.

### 3.4 Seleção de variáveis explicativas

A seleção das variáveis econômicas e financeiras com frequência diária enfrentou limitações devido à disponibilidade de dados, o que impediu a inclusão de variáveis comumente encontradas em modelos econômicos tradicionais, tais como: produção, inflação, agregados monetários, balança comercial, entre outras. No entanto, a seleção foi orientada pela relevância empírica das variáveis disponíveis e pela literatura associada à modelagem da taxa de câmbio, geralmente relacionada às condições econômicas e financeiras externas e domésticas.

Conforme descrito na revisão da literatura, Almeida (2022) considera entre as seguintes variáveis explicativas para a taxa de câmbio: CDS, DXY, CRY, VIX, SELIC e NTN-B; enquanto Türkoğlu, Göççek e Yumrukuz (2024) incluem, além do CDS e do VIX, as variáveis: *overnight interest rate*, títulos do Tesouro dos EUA de 10 anos, índice S&P 500, índice da bolsa de valores da Turquia (BIST100), entre outras. Segundo Krugman, Obstfeld e Melitz (2015), as taxas de juros e as expectativas nacionais são fatores essenciais na determinação do câmbio. Nesse sentido, foram realizados estudos preliminares com o objetivo de avaliar a contribuição das variáveis explicativas na redução das métricas de erro das previsões dos modelos multivariados durante o período de validação.

Dessa forma, as variáveis selecionadas foram: a taxa de câmbio Ptax de venda BRL/USD (Ptax\_Venda), o CDS Brasil de 5 anos (CDS\_5Y), o CDS Brasil de 10 anos (CDS\_10Y), o índice DXY (DXY\_INDEX), o índice de *commodities* (CRY\_INDEX), o índice Bovespa (IBOV), o índice S&P 500 (S&P500), o rendimento dos títulos do Tesouro dos EUA de 5 anos (TBOND\_5Y), o rendimento dos títulos do Tesouro dos EUA de 10 anos (TBOND\_10Y), o diferencial de juros entre Brasil e EUA (DIF\_JUROS) e o índice de volatilidade (VIX\_INDEX). A série de dados abrange o período de janeiro de 2020 a dezembro de 2024.

O *Credit Default Swap* (CDS) é um derivativo de crédito negociado em mercado de balcão em que investidores podem comprar ou vender proteção contra o risco de crédito soberano de um país ou de uma entidade corporativa. Fundamentalmente, o comprador de um CDS adquire uma proteção, pagando um prêmio regular ao vendedor, que garante o pagamento caso a entidade de referência entre em *default*. O CDS do Brasil funciona como uma *proxy* do risco-país, refletindo a probabilidade atribuída pelo mercado de que o país não honre suas obrigações financeiras. No presente estudo, são adotados os prazos de cinco e dez anos (CDS\_5Y e CDS\_10Y) por serem horizontes temporais considerados relevantes para investidores e para capturar diferentes perspectivas sobre a sustentabilidade da dívida soberana. Um aumento nos valores de CDS indica uma elevação da percepção de risco de crédito associada ao país em questão. Esse aumento pode ser desencadeado por diversos fatores, como deterioração das condições



econômicas, instabilidade política, choques externos negativos ou mudanças regulatórias desfavoráveis. A elevação do risco-país tende a pressionar o Real para uma desvalorização frente ao Dólar por diversos mecanismos como: fuga de capitais, redução do investimento direto estrangeiro, impacto nas expectativas e aumento da aversão ao risco, entre outros. Assim sendo, o sinal esperado da correlação entre o CDS Brasil e a taxa de câmbio R\$/US\$ é positivo.

O diferencial de juros de curto prazo entre Brasil e Estados Unidos (DIF\_JUROS) reflete a atratividade relativa dos investimentos em ativos denominados em Real em comparação aos ativos em Dólar. Essa disparidade nas taxas de juros exerce influência significativa sobre os fluxos de capitais internacionais e, conseqüentemente, sobre a taxa de câmbio (R\$/US\$). De acordo com Krugman, Obstfeld e Melitz (2015, p. 285), a condição de paridade descoberta de juros estabelece que o mercado cambial está em equilíbrio quando depósitos em moedas diferentes oferecem o mesmo retorno esperado. Desse modo, um aumento no diferencial de juros favorável ao Brasil, *ceteris paribus*, tende a tornar os ativos brasileiros mais atraentes, induzindo um influxo de capitais e, por conseguinte, uma valorização do Real frente ao Dólar. Portanto, o sinal esperado da correlação entre a taxa de câmbio (R\$/US\$) e o diferencial de juros de curto prazo é negativo.

O índice DXY ou *U.S. Dollar Index* mensura a variação do valor do Dólar americano em relação a uma cesta ponderada de moedas de economias desenvolvidas, incluindo o Euro, o Iene Japonês, a Libra Esterlina, o Dólar Canadense, o Franco Suíço e a Coroa Sueca. Um aumento nesse índice indica uma valorização generalizada do Dólar americano no cenário global, o que, por sua vez, tende a estar associada a uma desvalorização do Real. Isso ocorre porque um Dólar americano mais forte normalmente torna os ativos denominados em outras moedas, como o Real, relativamente menos atraentes para investidores internacionais. Conseqüentemente, o sinal esperado da correlação entre o índice DXY e a taxa de câmbio R\$/US\$ é positivo.

O índice CRY é uma sigla utilizada pela Bloomberg para se referir ao *CoreCommodity CRB Index*, que acompanha uma cesta diversificada de *commodities*, abrangendo setores como energia, metais, agricultura e pecuária. Dada a relevância do Brasil como exportador de diversas dessas matérias-primas, um aumento nos preços das *commodities* geralmente resulta em um fluxo maior de divisas para o país. Esse aumento na oferta de dólares tende a apreciar o Real em relação à moeda americana. Conseqüentemente, o sinal esperado da correlação entre o índice CRY e a taxa de câmbio R\$/US\$ é negativo.

As taxas de juros dos títulos do Tesouro dos EUA (*Treasury Bonds*) com vencimentos de 5 e 10 anos, doravante denominados TBOND\_5Y e TBOND\_10Y, respectivamente, são amplamente reconhecidas como o retorno de ativos de baixíssimo risco soberano. Elevações nesses rendimentos aumentam a atratividade dos títulos americanos, incentivando fluxos de capital para os Estados Unidos e, por conseguinte, exercendo pressão de depreciação sobre o Real brasileiro (R\$). Dessa forma, postula-se uma correlação positiva entre as variações dessas taxas de juros e a taxa de câmbio R\$/US\$.

O índice VIX (*Cboe Volatility Index*) quantifica a volatilidade implícita nos preços de

opções de curto prazo do índice S&P 500, sendo amplamente empregado como uma *proxy* da percepção de risco dos investidores no mercado acionário americano. Um aumento no VIX indica também uma elevação da aversão ao risco global, o que tende a impulsionar a alocação de capital para ativos considerados mais seguros, resultando em potenciais fluxos financeiros de saída de economias emergentes como a do Brasil. Consequentemente, espera-se uma correlação positiva entre as variações do índice VIX e a taxa de câmbio R\$/US\$.

O índice Bovespa (IBOV) representa o desempenho agregado das ações de maior liquidez e representatividade negociadas na bolsa brasileira B3. Um aumento no IBOV, *ceteris paribus*, indica um aumento da confiança dos investidores nas perspectivas da economia doméstica, o que pode fomentar a entrada de capital estrangeiro e, consequentemente, apreciar o Real brasileiro (R\$). Dessa forma, postula-se uma correlação negativa entre as variações do índice IBOV e a taxa de câmbio R\$/US\$.

Finalmente, o índice Standard & Poor's 500 (S&P 500) acompanha o desempenho das ações das 500 maiores empresas de capital aberto listadas nas bolsas de valores dos Estados Unidos, sendo amplamente reconhecido como um indicador do apetite global por risco. Uma elevação do S&P 500 sinaliza um aumento do otimismo em mercados de renda variável, o que pode estimular o interesse por investimentos em mercados emergentes, como o Brasil. Consequentemente, o fluxo de capitais tende a se direcionar para ativos com maior potencial de retorno, incluindo os brasileiros, apreciando o Real. Essa relação não é uma regra rígida e pode ser influenciada por diversos fatores. Desse modo, sob certas condições, o sinal para a correlação entre as variações do índice S&P 500 e a taxa de câmbio R\$/US\$ tende ser negativo.

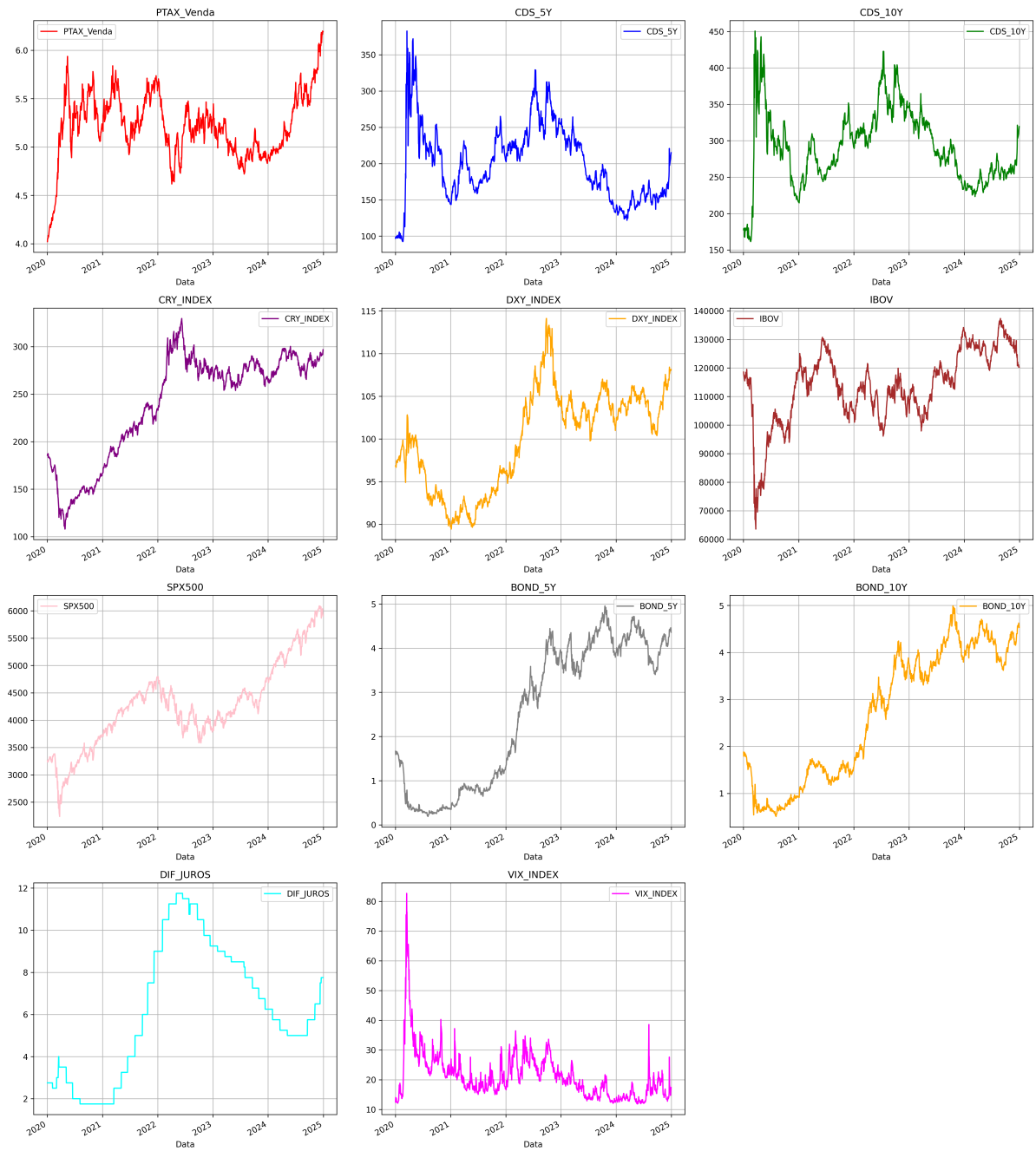
Cabe ressaltar que as relações esperadas entre a taxa de câmbio Ptax de venda e as diversas variáveis econômicas e financeiras, conforme descrito anteriormente, não são rígidas, estando sujeitas às influências do contexto socioeconômico e geopolítico, tanto em âmbito doméstico quanto internacional.

## **4 DESCRIÇÃO ESTATÍSTICA DOS DADOS**

As séries de dados diários das variáveis explicativas foram coletadas na plataforma Bloomberg, enquanto a série de dados da taxa de câmbio Ptax de venda foram obtidas por coleta direta na API do Banco Central. A coleta das séries de dados com frequência diária compreende o período de 01 de janeiro de 2020 a 31 de dezembro de 2024.

O conjunto de dados foi dividido sequencialmente e utilizado nos procedimentos de treinamento, validação e teste dos modelos de DL: LSTM, BiSLTM, GRU e BiGRU.

Figura 3 - Gráfico das séries temporais coletadas no período de 2020 a 2024.



**Nota:** Valores e datas das séries de dados consideradas no estudo.

**Fonte:** O autor, 2025.

Tabela 2 - Métricas descritivas das séries temporais em preços

Variável	Média	Variância	Desvio Padrão	Coefficiente de variação (%)	Curtose	Assimetria	Máximo	Mínimo
PTAX_Venda	5.22	0.12	0.34	6.50	1.12	-0.38	6.20	4.02
CDS_5Y*	201.20	2715.89	52.11	25.90	-0.02	0.45	382.72	92.35
CDS_10Y**	291.03	2530.51	50.30	17.28	0.20	0.27	450.61	161.41
CRY_INDEX	239.23	3119.19	55.85	23.35	-0.86	-0.71	329.59	108.00
DXY_INDEX	99.98	33.01	5.75	5.75	-0.99	-0.14	114.11	89.44
IBOV	113504.46	164197621.50	12813.96	11.29	1.06	-0.83	137343.96	63569.62
S&P500	4259.75	588694.26	767.26	18.01	-0.10	0.24	6090.27	2237.40
TBOND_5Y	2.51	2.62	1.62	64.42	-1.68	-0.16	4.96	0.19
TBOND_10Y	2.68	1.94	1.39	51.94	-1.57	-0.12	4.99	0.51
DIF_JUROS	6.20	9.99	3.16	50.94	-1.19	0.13	11.75	1.75
VIX_INDEX	21.40	68.45	8.27	38.67	11.48	2.57	82.69	11.86

**Nota 1:** CDS\_5Y\* e CDS\_10Y\*\* em pontos base.

**Nota 2:** Curtose expressa em excesso de curtose.

**Fonte:** O autor, 2025.

A Tabela 2 apresenta as estatísticas descritivas das séries temporais econômicas e financeiras utilizadas na previsão da taxa de câmbio Ptax de venda. A média, a variância e desvio padrão caracterizam a tendência central e a dispersão dos dados, enquanto a assimetria e a curtose fornecem informações sobre a forma das distribuições e o comportamento de suas caudas. O coeficiente de variação (%) permite avaliar a volatilidade relativa das séries em relação às suas médias, possibilitando comparações entre variáveis com magnitudes distintas.

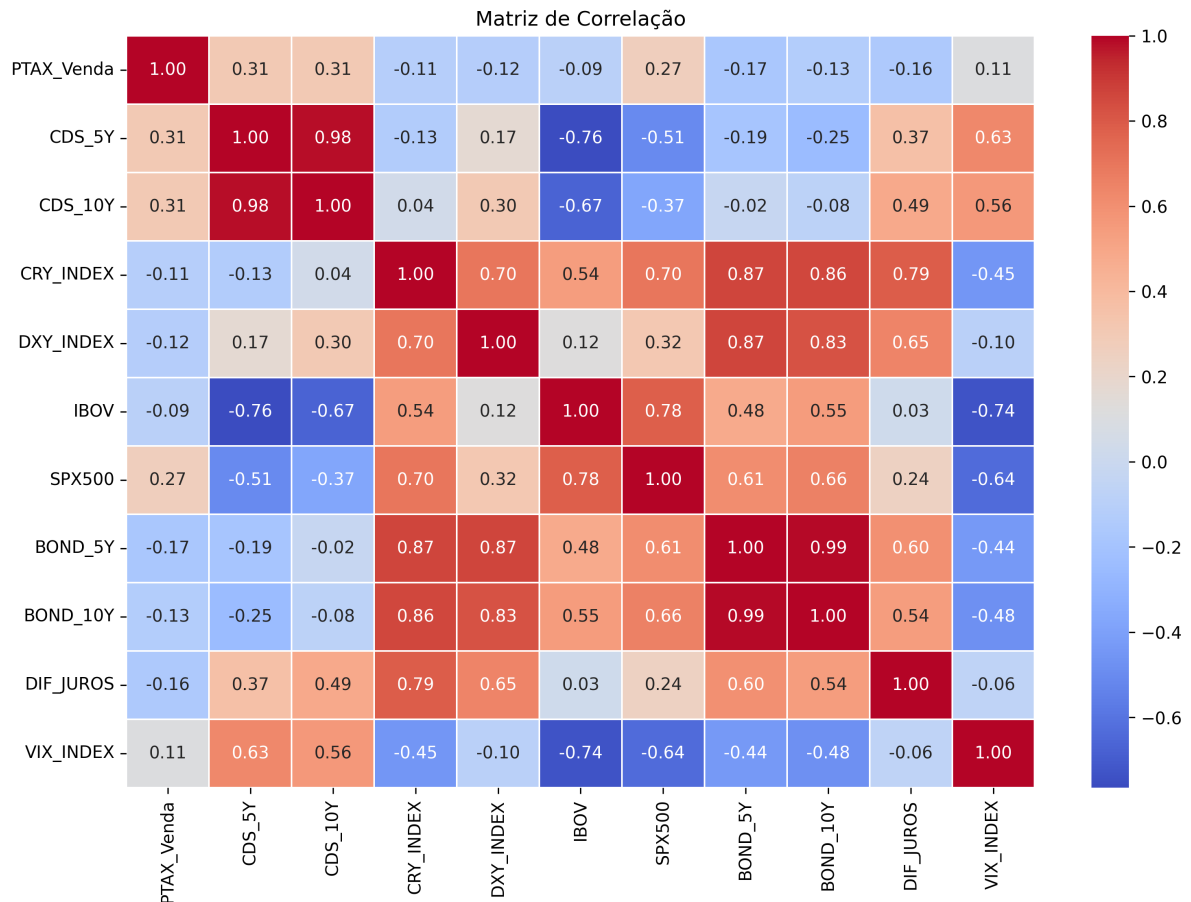
Observa-se que a série Ptax\_Venda apresenta média de 5,22, variância de 0,12 e leve assimetria negativa de -0,38, indicando maior concentração de valores acima da média e uma cauda mais longa à esquerda. A curtose positiva sugere presença de caudas mais pesadas que a distribuição normal. O coeficiente de variação relativamente baixo (6,50%) evidencia que a dispersão da série é pequena em relação à sua média, ao passo que outras variáveis, como TBOND\_5Y e DIF\_JUROS, apresentam coeficientes de variação elevados, indicando maior volatilidade relativa.

Destaca-se, ainda, o índice de volatilidade VIX, que apresenta curtose elevada (11,48) e assimetria positiva acentuada (2,57), refletindo episódios de elevada volatilidade nos mercados financeiros, em consonância com a representação gráfica da série S&P 500 apresentada na Figura 3. Esses resultados fundamentam a adoção de modelos de DL para captura de padrões complexos e não lineares, sem a necessidade de atender às premissas rígidas de modelos econométricos tradicionais.

Neste sentido, Joseph e Tackes (2024, p. 148, 591) apontam que nem sempre é essencial corrigir ou tornar uma série temporal estacionária no paradigma do *Machine Learning* (ML), que inclui o DL, pois algumas dessas situações podem ser tratadas usando o tipo certo de recursos no modelo. A decisão de transformar ou não a série deve ser tomada com base em experimentação, considerando que tais transformações podem trazer tanto benefícios quanto limitações. Além disso, o uso de estratégias como a janela móvel permite mitigar os efeitos da não estacionariedade, mantendo o modelo atualizado frente a mudanças no comportamento da série ao longo do tempo. Essa abordagem também evita o viés introduzido por janelas expansí-

veis, garantindo que o modelo trabalhe sempre com o mesmo comprimento da série.

Figura 4 - Matriz de correlação das variáveis (preços) no período de 2020 a 2024.



Fonte: O autor, 2025.

A matriz de correlação das variáveis consideradas no estudo, descrita na Figura 4, evidencia relações lineares estáticas entre a taxa de câmbio Ptax de venda e variáveis econômicas e financeiros no período de janeiro de 2020 a dezembro de 2024.

De modo geral, os sinais das correlações entre as variáveis tais variáveis seguem as expectativas teóricas e empíricas esperadas nas economias emergentes, em conformidade com o descrito por Almeida (2022). Entretanto, diante da multiplicidade de fatores socioeconômicos e geopolíticos que influenciam a taxa de câmbio, o sinal e a intensidade das correlações esperadas entre as variáveis não seguem padrões rígidos. Isso se observa nas séries descritas na Figura 3 e demonstra a elevada complexidade associada às inter-relações e dinâmicas da taxa de câmbio.

Neste caso específico, as correlações negativas entre a taxa de câmbio Ptax de venda e as variáveis explicativas índice DXY, TBOND\_5Y e TBOND\_10Y, bem como a correlação positiva entre a mesma taxa e o índice S&P 500, não corresponderam aos comportamentos esperados. Alguns fatores podem ter atuado nesse sentido, em primeiro lugar, em decorrência da pandemia de COVID-19, houve uma forte desvalorização do Real frente ao Dólar americano em 2020. Por outro lado, de 2021 até meados de 2022 ocorreu um *boom* no preço das *commodities* como se observa na série do índice CRY. Assim sendo, a correlação negativa e fraca entre

a Ptax de venda do dólar no Brasil e o índice DXY entre 2020 e 2024 sugere que, nesse período, a relação entre a força do dólar globalmente e a cotação do dólar no Brasil pode ter sido compensada por outros fatores. Quanto à correlação em relação aos TBOND\_5Y e TBOND\_10Y, observou-se, nesse período, um acelerado ciclo de alta das taxas de juros brasileiras, o que pode ter compensado o efeito esperado do aumento das taxas de juros norte-americanas na taxa de câmbio.

Por outro lado, conforme o esperado, verifica-se que a taxa de câmbio Ptax de venda apresentou correlação positiva em relação aos CDS Brasil de 5 e de 10 anos e ao índice VIX. Assim, há indícios de que desvalorizações da taxa de câmbio estiveram associadas, em certa medida, ao aumento da percepção ao risco fiscal do Brasil e ao risco dos investidores no mercado acionário norte-americano. Adicionalmente, observa-se uma correlação negativa da taxa de câmbio Ptax de venda com os índices CRY e IBOV e, ainda, com o diferencial de juros de curtíssimo prazo. Dessa forma, sugere-se que períodos de valorização dos preços das *commodities* e do mercado acionário brasileiro, além do aumento do diferencial de juros, coincidiram, em certa medida, com uma apreciação do Real.

No contexto das demais variáveis, destaca-se a correlação positiva entre os CDS e o diferencial de juros, o que sugere que o aumento do risco-país esteve associado a aumentos do diferencial de taxas de juros.

## 5 RESULTADOS

Neste Capítulo, são apresentados os resultados obtidos a partir da aplicação dos modelos de DL em comparação ao modelo RW na previsão da taxa de câmbio Ptax de venda diária do Real (BRL) em relação ao Dólar (USD). Basicamente, a análise comparativa será realizada em duas etapas. Primeiro, será selecionado um modelo de DL com base nas menores métricas de erro no período de validação. Em seguida, o desempenho das previsões do modelo de DL selecionado será comparado ao do modelo RW no período de teste, fase em que os dados são novos para os modelos. Assim, será possível avaliar capacidade de generalização dos modelos de DL, simulando-se, uma aplicação prática.

Para um estudo mais abrangente, serão adotadas tanto a abordagem univariada, que considera apenas os valores passados da taxa de câmbio Ptax de venda, quanto a abordagem multivariada, que incorpora variáveis econômicas e financeiras (VIX, DXY, CRY, diferencial de juros, CDS de 5 e 10 anos, títulos do Tesouro dos EUA de 5 e 10 anos, índice Ibovespa e S&P 500), além dos valores passados da respectiva taxa de câmbio. Os horizontes de previsão considerados serão de 1, 10 e 20 dias úteis à frente. Além disso, serão avaliadas duas variações de arquitetura nos modelos de DL em cada caso: uma com uma camada recorrente (1CR) e outra com duas camadas recorrentes (2CR). Ressalta-se que o nome atribuído aos modelos de DL corresponde ao tipo de camada recorrente (LSTM, GRU, BiLSTM e BiGRU) utilizada na respectiva arquitetura. Dessa forma, o desempenho preditivo poderá ser avaliado não só pela incorporação de variáveis econômicas e financeiras, em relação aos modelos univariados, mas também pela inclusão de uma camada recorrente adicional nas arquiteturas dos modelos de DL.

A análise comparativa terá como base inicial as métricas de erro RMSE, MAE e MAPE. Adicionalmente, para cada abordagem e horizonte de previsão, realizam-se testes estatísticos entre o modelo RW e os modelos de DL selecionados, que comparam as respectivas métricas de erro nas duas arquiteturas adotadas em cada abordagem e horizonte de previsão. Dessa forma, será possível verificar objetivamente se as previsões dos referidos modelos de DL, nas configurações especificadas, conseguiram superar a acurácia das previsões do modelo RW, em consonância com o objetivo principal definido para este estudo.

Quanto aos testes aplicados, os testes de Diebold e Mariano (1995) e Diebold-Mariano-Modificado (1997) permitem a comparação da acurácia das previsões dos modelos de DL em relação à do modelo RW, com a diferença de que o segundo inclui um ajuste para redução do viés relativo ao tamanho da amostra. Além disso, o teste de flutuação de Giacomini e Rossi (2010) permite descrever graficamente como varia a diferença de acurácia das previsões de dois modelos comparados ao longo do tempo. Sua estatística de teste (ET) é baseada no teste DM, aplicado em janelas móveis de 20 dias úteis neste estudo. Porém, as ET descritas nos testes estatísticos correspondem, nesse contexto, a janela de tempo com a maior diferença de acurácia preditiva entre os modelos comparados nos respectivos períodos analisados.



Por sua vez, o teste de racionalidade de Rossi e Sekhposyan (2016), aplicado para cada modelo individualmente, indica se os erros de previsão são imprevisíveis ou se apresentam padrões sistemáticos, o que pode indicar viés e limitações preditivas dos modelos ou a ocorrência de uma quebra estrutural. Nesse caso, a ET do teste em si considera a janela de tempo, também com 20 dias úteis, de maior irracionalidade nos erros de previsão do período analisado. Esse teste também permite descrever graficamente como varia a irracionalidade dos erros de previsão ao longo das janelas de tempo.

Conforme descrito no capítulo de Metodologia, este ensaio empírico seguiu a seguinte sistemática: realização do treinamento dos modelos de DL exclusivamente com o conjunto de dados segredados para este fim; utilização do período de validação para ajuste de hiperparâmetros e para a seleção dos modelos com as menores métricas de erro; e, em seguida, comparação das previsões dos modelos de DL selecionados, com melhor desempenho no período de teste, em relação às previsões do modelo RW, o que constitui o foco principal deste estudo.

Por fim, os resultados serão discutidos à luz das evidências empíricas e das contribuições da literatura, destacando as implicações práticas e os desafios associados ao uso de modelos de DL na previsão da taxa de câmbio.

## 5.1 Resultados das previsões 1 dia à frente

### 5.1.1 Modelos univariados de DL vs. modelo RW nas previsões 1 dia à frente

#### 5.1.1.1 Modelos univariados de DL com 1CR vs. modelo RW nas previsões 1 dia à frente

Tabela 3 - Métricas de erro das previsões dos modelos univariados de DL com 1CR e do modelo RW 1 dia à frente

Modelos	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
RW	<b>0.0308</b>	<b>0.0230</b>	<b>0.4476</b>	<b>0.0477</b>	<b>0.0332</b>	<b>0.5805</b>
LSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0568	0.0429	0.8291	0.0979	0.0794	1.3795
LSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0577	0.0440	0.8502	0.1162	0.0945	1.6311
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0590	0.0452	0.8727	0.1311	0.1068	1.8401
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0557	0.0424	0.8208	0.1051	0.0847	1.4662
GRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0532	0.0405	0.7826	0.1341	0.1055	1.8116
GRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0543	0.0414	0.7992	0.1197	0.0949	1.6335
GRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0567	0.0436	0.8417	0.1133	0.0922	1.5925
GRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0520	0.0399	0.7706	0.1097	0.0881	1.5196
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0368	0.0299	0.5819	<b>0.0538</b>	<b>0.0399</b>	<b>0.6978</b>
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0359	0.0286	0.5551	0.0552	0.0409	0.7157
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	<b>0.0351</b>	<b>0.0280</b>	<b>0.5454</b>	0.0587	0.0455	0.7913
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0360	0.0283	0.5481	0.0664	0.0522	0.9045
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0387	0.0307	0.5970	0.0602	0.0455	0.7966
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0382	0.0306	0.5948	0.0621	0.0469	0.8182
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0399	0.0319	0.6202	0.0681	0.0523	0.9105
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0394	0.0309	0.5986	0.0609	0.0465	0.8140

**Nota:** Os dois menores valores de cada coluna estão destacados em negrito.

**Fonte:** O autor, 2025.

A Tabela 3 apresenta as métricas de erro das previsões da taxa de câmbio Ptax de venda 1 dia útil à frente realizadas pelo modelo RW e pelos modelos de DL univariados com 1CR. No desempenho geral, observa-se que as previsões do modelo RW apresentaram as menores métricas de erro, tanto no período de validação quanto no período de teste. Entre os modelos de DL com 1CR, o BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16) foi selecionado, pois obteve as previsões com menores métricas de erro no período de validação. No período de teste, o seu desempenho preditivo foi relativamente próximo ao do modelo RW e ao do modelo do modelo de DL com as menores métricas de erro desse período, o BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16). Verifica-se, desde já, que o melhor desempenho entre os modelos DL no período de validação não implica necessariamente o mesmo resultado no período de teste.

As previsões dos modelos BiLSTM e BiGRU apresentaram menores métricas de erro do que os seus pares unidirecionais LSTM e GRU nos dois períodos analisados. Apesar disso, verifica-se que as métricas de erro das previsões dos modelos de DL unidirecionais foram superiores em relação às do modelo RW, especialmente no período de teste.

#### 5.1.1.2 Modelos univariados de DL com 2CR vs. modelo RW nas previsões 1 dia à frente:

Tabela 4 - Métricas de erro das previsões 1 dia à frente dos modelos univariados de DL com 2CR e do RW

Modelos (duas camadas recorrentes)	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
RW	<b>0.0308</b>	<b>0.0230</b>	<b>0.4476</b>	<b>0.0478</b>	<b>0.0332</b>	<b>0.5805</b>
LSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0753	0.0545	1.0041	0.1107	0.0928	1.6293
LSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0571	0.0433	0.8361	0.1041	0.0860	1.4921
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0549	0.0419	0.8103	0.0944	0.0766	1.3314
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0593	0.0444	0.8539	0.1225	0.1036	1.7961
GRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0532	0.0408	0.7871	0.1635	0.1265	2.1643
GRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0549	0.0415	0.7976	0.1498	0.1172	2.0096
GRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0544	0.0414	0.7984	0.1316	0.1054	1.8110
GRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0553	0.0416	0.7995	0.1464	0.1173	2.0147
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0408	0.0336	0.6545	0.0585	0.0440	0.7692
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0375	0.0298	0.5786	0.0612	0.0474	0.8242
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	<b>0.0368</b>	<b>0.0290</b>	<b>0.5631</b>	<b>0.0556</b>	<b>0.0416</b>	<b>0.7292</b>
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0417	0.0320	0.6170	0.0671	0.0541	0.9464
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0384	0.0308	0.5977	0.0635	0.0484	0.8428
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0389	0.0315	0.6143	0.0652	0.0487	0.8470
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0404	0.0323	0.6262	0.0693	0.0545	0.9473
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0454	0.0337	0.6488	0.0979	0.0802	1.3853

**Nota:** Os dois menores valores de cada coluna estão destacados em negrito.

**Fonte:** O autor, 2025.

A Tabela 4 apresenta as métricas de erro das previsões para 1 dia à frente realizadas pelo modelo RW e pelos modelos de DL univariados com 2CR. Verifica-se um padrão de desempenho similar ao observado para os modelos univariados com 1CR. O modelo RW também apresenta as menores métricas de erro tanto no período de validação quanto no período de teste. Os modelos BiLSTM e BiGRU continuam se destacando em relação aos modelos de DL unidirecionais. O BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16) foi novamente o modelo

com menores métricas de erro no período de validação. Porém, neste caso, o mesmo modelo também apresentou as menores métricas de erro entre os modelos de DL no período de teste.

De forma geral, o aumento da complexidade da rede neural, com a inclusão de uma segunda camada recorrente nos modelos de DL, não resultou em uma redução das métricas de erro. Na comparação com os modelos de DL com 1CR descritos na Tabela 3, observa-se que, em alguns casos, como no do modelo GRU, houve um aumento das métricas de erro no período de teste.

### 5.1.1.3 Modelos de DL univariados selecionados vs. modelo RW nas previsões 1 dia à frente

Tabela 5 - Métricas de erro das previsões 1 dia à frente do modelo RW e dos modelos de DL univariados selecionados em cada arquitetura

Arquitetura	Modelos e Parâmetros	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
-	RW	<b>0.0308</b>	<b>0.0230</b>	<b>0.4476</b>	<b>0.0478</b>	<b>0.0332</b>	<b>0.5805</b>
1CR	BiLSTM-adamw (LR0.0001 - Dropout0.1 - Batch16)	0.0351	0.0280	0.5454	0.0587	0.0455	0.7913
2CR	BiLSTM-adamw (LR0.0001 - Dropout0.1 - Batch16)	0.0368	0.0290	0.5631	0.0556	0.0416	0.7292

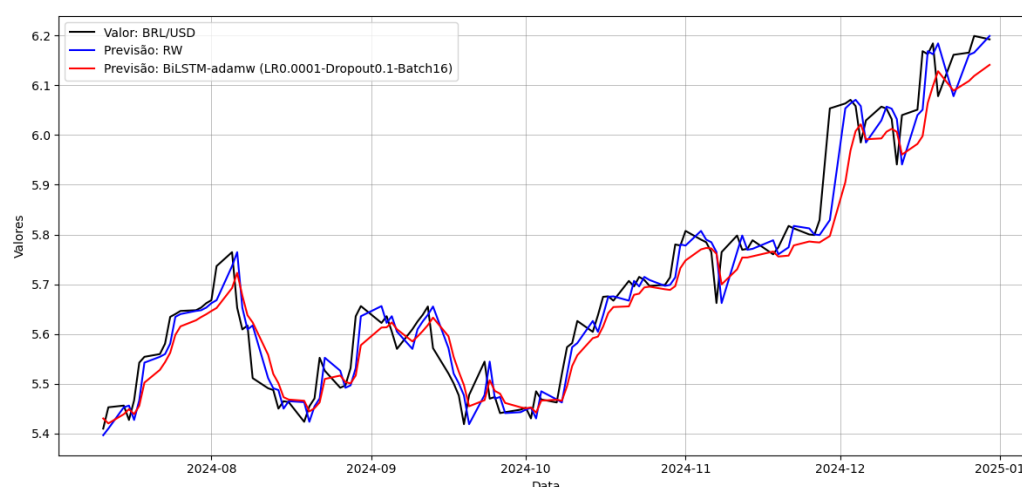
**Nota:** Os menores valores por coluna estão destacados em negrito.

**Fonte:** O autor, 2025.

A Tabela 5 apresenta as métricas de erro das previsões 1 dia útil à frente do modelo RW e dos modelos de DL univariados selecionados nas arquiteturas com 1CR e 2CR, pelo critério de menores métricas de erro no período de validação. De forma coerente com os resultados previamente descritos, entre todos os modelos, o RW obteve as menores métricas de erro em termos absolutos tanto no período de validação quanto no período de teste. O modelo BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16) foi o modelo selecionado nas duas arquiteturas e apresentou métricas de erros relativamente próximas às do RW nos dois períodos analisados.

Entre os dois modelos de DL univariados selecionados, o modelo com 2CR foi o que obteve as métricas de erro mais próximas às do RW no período de teste. Dessa forma, serão apresentados testes estatísticos com objetivo de verificar se a acurácia das respectivas previsões são estatisticamente equivalentes ou não. Porém, antes disso, serão apresentados os gráficos das previsões dos modelos de DL selecionados com 1CR e com 2CR em relação às previsões do modelo RW e às taxas de câmbio Ptax de venda efetivas do período de teste.

Figura 5 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 1 dia útil à frente do modelo selecionado univariado com 1CR vs RW no período de teste



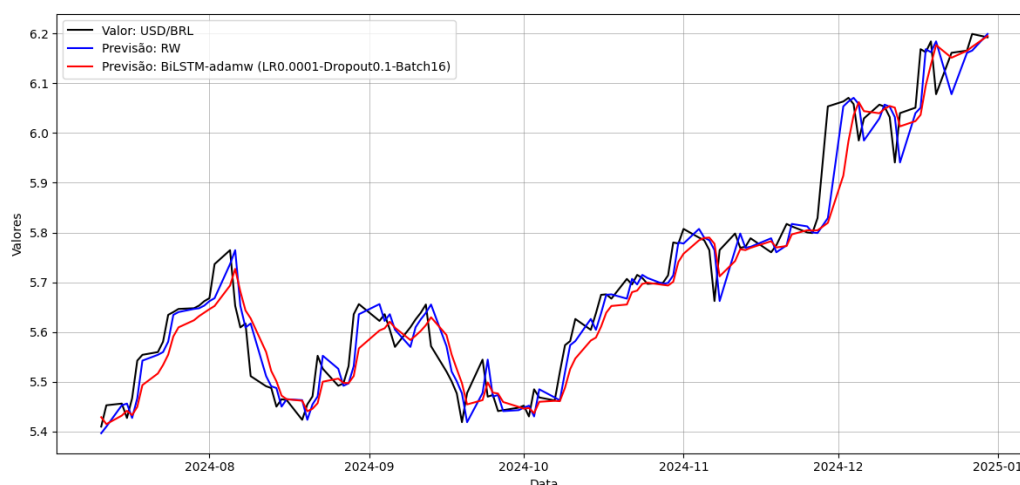
**Nota:** Referente ao modelo de DL univariado selecionado, descrito na legenda.

**Fonte:** O autor, 2025.

A Figura 5 apresenta a taxa de câmbio Ptax de venda (BRL/USD) no período de teste, comparando as previsões 1 dia útil à frente do modelo RW e as do modelo de DL selecionado BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16) com 1CR que adota uma abordagem univariada.

Observa-se uma elevada volatilidade ao longo do período de teste, com uma tendência de alta a partir do final da primeira metade do mesmo período. Visualmente, as previsões dos respectivos modelos indicam um desempenho relativamente próximo. No entanto, nota-se que o modelo selecionado BiLSTM exibe uma suavização maior em relação ao RW. No final do período de teste, a previsão do BiLSTM parece subestimar a magnitude do movimento ascendente da taxa de câmbio, possivelmente devido a um efeito de regularização excessiva ou à dificuldade do modelo univariado de captar padrões altamente dinâmicos. Em contraste, o RW, por sua natureza recursiva, ajusta-se mais rapidamente às mudanças bruscas do mercado, mantendo um alinhamento mais próximo com a série observada.

Figura 6 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 1 dia útil à frente do modelo selecionado univariado com 2CR vs RW no período de teste



**Nota:** Referente ao modelo de DL univariado selecionado, com 2CR, descrito na legenda.

**Fonte:** O autor, 2025.

A Figura 6 exibe a taxa de câmbio Ptax de venda (BRL/USD) no período de teste, comparando as previsões 1 dia útil à frente do modelo RW e as do modelo de DL selecionado BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16) com 2CR que adota uma abordagem univariada.

Assim como na análise anterior, observa-se uma alta volatilidade ao longo do período de teste, com uma tendência de valorização do Dólar mais evidente na segunda metade desse período. Dentro da sistemática adotada, a inclusão de uma segunda camada recorrente resultou na seleção de um modelo em que a suavização observada parece ter se reduzido ligeiramente em relação ao modelo de DL selecionado com 1CR. Nota-se uma maior proximidade das previsões às oscilações da série temporal, embora ainda apresente uma leve defasagem em momentos de maior volatilidade, conforme se observa no final do período de teste. Entretanto, o modelo RW continua exibindo um ajuste mais ágil às oscilações bruscas do mercado, reforçando sua eficácia para previsões de curtíssimo prazo em cenários de elevada instabilidade.

#### 5.1.1.4 Testes estatísticos entre o modelo RW e o modelo de DL univariado selecionado com melhor desempenho no período de teste nas previsões 1 dia útil à frente

Tabela 6 - Testes estatísticos para o modelo univariado de DL selecionado com 2CR e o modelo RW nas previsões 1 dia útil à frente

Testes	Período de Validação	Período de Teste
Diebold-Mariano (DM)	RW vs BiLSTM: Estatística = -2.8114, p-valor = 0.0049***	RW vs BiLSTM: Estatística = -2.0481, p-valor = 0.0405**
Diebold-Mariano Modificado (DMM)	RW vs BiLSTM: Estatística = -2.7993, p-valor = 0.0060***	RW vs BiLSTM: Estatística = -2.0394, p-valor = 0.0437**
Flutuação de Giacomini e Rossi (FGR)	RW vs BiLSTM: Estatística = -3.1345, p-valor = 0.0055***	RW vs BiLSTM: Estatística = -2.3538, p-valor = 0.0295**
Racionalidade de Rossi e Sekhposyan (RS)	RW: Estatística = 158.5189, p-valor = 0.0000***	RW: Estatística = 173.0916, p-valor = 0.0000***
Racionalidade de Rossi e Sekhposyan (RS)	BiLSTM: Estatística = 774.1853, p-valor = 0.0000***	BiLSTM: Estatística = 444.5458, p-valor = 0.0000***

**Nota 1:** Refere-se ao BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16).

**Nota 2:** \*\*\*, \*\* e \* indicam significância estatística ao nível de 1%, 5% e 10%, respectivamente.

**Fonte:** O autor, 2025.

A Tabela 6 apresenta os resultados dos testes estatísticos comparando as previsões do modelo RW com as do modelo selecionado BiLSTM (adamw - LR 0.0001 - Dropout 0.1 - Batch 16) com 2CR, que obteve as menores métricas de erro no período de teste entre as duas arquiteturas. Os resultados dos testes para o modelo selecionado com 1CR podem ser consultados na Tabela 27, disponível no Apêndice A.

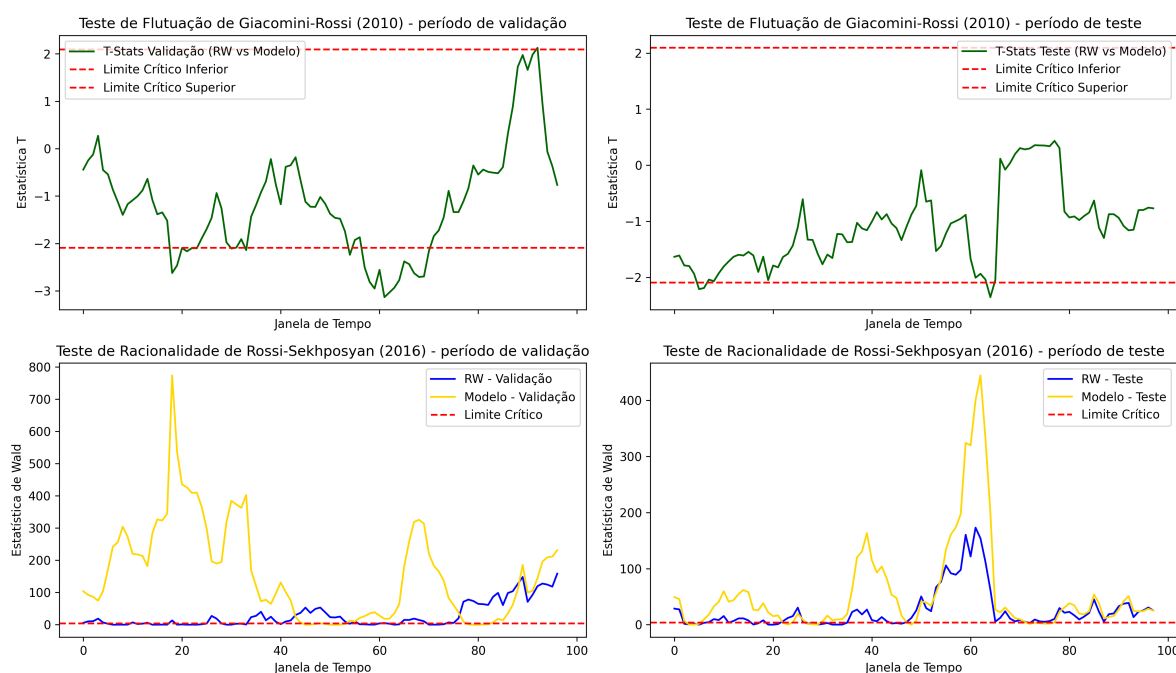
Os resultados dos testes Diebold-Mariano (DM) e Diebold-Mariano Modificado (DMM) na tabela acima indicam que as previsões dos modelos comparados apresentaram acurácias estatisticamente diferentes. As hipóteses nulas ( $H_0$ ) são rejeitadas nos dois testes ao nível de significância de 5%, tanto no período de validação quanto no de teste. Além disso, em termos de acurácia, os resultados sugerem que o modelo RW teve um desempenho estatisticamente superior ao BiLSTM selecionado, conforme evidenciado pelas estatísticas de teste (ET) negativas em ambos os períodos analisados.

Por outro lado, o teste de flutuação Giacomini e Rossi (GR), que avalia a diferença de acurácia das previsões ao longo do tempo entre o modelo de DL selecionado e o modelo RW, aponta para a rejeição da  $H_0$  de igualdade de desempenho das previsões. Neste caso, o referido teste aponta uma superioridade da acurácia das previsões do modelo RW em relação à do modelo de DL selecionado na janela de tempo em que a diferença de desempenho relativo foi mais extrema. A Figura 7, apresentada na sequência, ilustrará o comportamento do desempenho preditivo relativo ao longo das janelas de tempo dos períodos de validação e de teste.

Por sua vez, o teste de racionalidade de Rossi-Sekhposyan (RS), que é realizado para cada modelo individualmente, indica a rejeição da  $H_0$  de imprevisibilidade nos erros de previsão do modelo RW e do modelo de DL selecionado, tanto no período de validação quanto no de teste. Verifica-se então a presença de eventos de irracionalidade, nos quais os erros preditivos não são aleatórios e apresentam dependência temporal. No caso do BiLSTM, as ET são ainda mais elevadas do que as do modelo RW, especialmente no período de teste, evidenciando-se um maior viés em pelo menos uma janela de tempo em que os erros seguiram um padrão previsível. Tal resultado pode estar relacionado a limitações na capacidade preditiva do modelo de DL selecionado univariado, que utiliza apenas os preços passados da taxa de câmbio Ptax de venda

para a previsão de valores futuros. Além disso, eventos exógenos ocorridos nos períodos de previsão podem ter causado mudanças estruturais nos dados que resultaram em padrões nos erros de previsão.

Figura 7 - Gráficos das estatísticas dos testes de flutuação Giacomini e Rossi (2010) e testes de racionalidade de Rossi-Sekhsposyan (2016) ao longo dos períodos de teste e de validação



**Nota:** O "Modelo" descrito na legenda refere-se ao modelo de DL selecionado univariado, com 2CR, nas previsões 1 dia útil à frente. As janelas móveis possuem 20 dias úteis nos dois testes.

**Fonte:** O autor, 2025.

Conforme ilustrado na Figura 7, a evolução das ETs do teste de flutuação de GR revela que durante a maior parte do tempo a ET permaneceu dentro dos valores críticos, indicando uma relativa igualdade da acurácia preditiva entre os modelos nas janelas de tempo de 20 dias úteis, com curtos intervalos nos quais as previsões do RW apresentaram desempenho relativamente superior.

Por outro lado, a evolução da ET no teste de racionalidade de RS revela que, no período de validação, o modelo de DL selecionado apresentou diversas janelas de tempo em que o comportamento dos erros de previsões foi irracional, sugerindo falta de eficiência e viés nas previsões, com a ET superando intensamente o limite crítico. Por sua vez, a partir da segunda metade do período de teste, observa-se um aumento acentuado na irracionalidade dos erros de previsão do modelo de DL, seguido por uma redução também acentuada.

Uma possível explicação para esse comportamento é a influência de eventos exógenos relevantes. Destaca-se, por exemplo, o impacto das eleições presidenciais nos Estados Unidos, realizadas no final de 2024, que resultaram no fortalecimento global do dólar, como evidenciado pela trajetória do índice DXY descrita na Figura 3. Além disso, podem-se observar os menores patamares do diferencial de juros em 2024, desde meados de 2021, e o aumento dos CDS

em 2024 em virtude do aumento das preocupações fiscais e a crescente desconfiança quanto ao novo arcabouço fiscal brasileiro. Esses fatores podem ter provocado mudanças estruturais temporárias no comportamento da taxa de câmbio, comprometendo a acurácia e a eficiência das previsões dos modelos de DL.

De modo geral, os resultados obtidos com os modelos de DL univariados aplicados na previsão 1 dia útil à frente indicam que, mesmo com as diferentes especificações e parâmetros adotados neste estudo, sua eficácia para previsões de curtíssimo prazo pode ser limitada, principalmente em cenários de alta volatilidade.

### 5.1.2 Modelos multivariados de DL vs. modelo RW nas previsões 1 dia útil à frente

#### 5.1.2.1 Modelos multivariados de DL com 1CR vs. modelo RW nas previsões 1 dia útil à frente:

Tabela 7 - Métricas de erro das previsões dos modelos multivariados de DL com 1CR e do modelo RW 1 dia útil à frente

Modelos (uma camada recorrente)	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
RW	<b>0.0308</b>	<b>0.0230</b>	<b>0.4476</b>	<b>0.0478</b>	<b>0.0332</b>	<b>0.5805</b>
LSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0550	0.0419	0.8081	0.0938	0.0727	1.2682
LSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0580	0.0438	0.8440	0.0995	0.0773	1.3500
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0545	0.0415	0.8017	0.0850	0.0636	1.1186
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0615	0.0464	0.8926	0.1148	0.0947	1.6442
GRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0561	0.0420	0.8076	0.1777	0.1421	2.4365
GRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0514	0.0396	0.7694	0.1316	0.1023	1.7569
GRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0491	0.0381	0.7392	0.0933	0.0713	1.2394
GRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0474	0.0361	0.7004	0.1138	0.0879	1.5147
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	<b>0.0333</b>	<b>0.0257</b>	<b>0.5022</b>	0.1083	0.0856	1.4846
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0360	0.0277	0.5416	0.1289	0.1091	1.8990
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0363	0.0288	0.5615	0.0800	<b>0.0629</b>	<b>1.1012</b>
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0375	0.0289	0.5650	<b>0.0780</b>	0.0636	1.1205
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0631	0.0520	1.0063	0.0858	0.0696	1.2191
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0871	0.0629	1.1992	0.1756	0.1548	2.6844
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0600	0.0462	0.8879	0.1065	0.0900	1.5658
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0912	0.0768	1.4818	0.1213	0.1045	1.8322

**Nota:** Os dois menores valores de cada coluna estão destacados em negrito.

**Fonte:** O autor, 2025.

A Tabela 7 descreve as métricas de erro das previsões para 1 dia útil à frente realizadas pelo modelo RW e pelos modelos de DL multivariados com 1CR. Observa-se que, no desempenho geral, assim como descrito anteriormente na abordagem univariada, o modelo RW continua a exibir as menores métricas de erro, tanto no período de validação quanto no de teste.

Neste caso, o modelo BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16) é o modelo de DL com 1CR selecionado, visto que apresenta as menores métricas de erro no período de validação. Nesse mesmo período suas métricas de erro foram próximas às do RW, ao contrário do observado no período de teste em que as métricas de erro das previsões foram mais do que o dobro do que as observadas para o modelo RW.



Verifica-se que as métricas de erro das previsões dos modelos de DL multivariados não foram inferiores às apresentadas pelos modelos de DL univariados no período de teste, mesmo com a inclusão de variáveis econômicas e financeiras em tais modelos. Ocorreu, inclusive, um aumento nas métricas de erro das previsões do modelo BiGRU em relação à abordagem univariada, tanto no período de validação quanto no de teste.

No período de validação, os modelos BiLSTM continuaram a obter o melhor desempenho entre os modelos de DL. Apesar disso, no período de teste, os resultados das métricas de erro também foram superiores, em termos absolutos, aos obtidos nos respectivos modelos univariados.

### 5.1.2.2 Modelos multivariados de DL com 2CR vs. modelo RW nas previsões 1 dia útil à frente:

Tabela 8 - Métricas de erro das previsões dos modelos multivariados de DL com 2CR e do modelo RW 1 dia útil à frente

Modelos (duas camadas recorrentes)	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
RW	<b>0.0308</b>	<b>0.0230</b>	<b>0.4476</b>	<b>0.0478</b>	<b>0.0332</b>	<b>0.5805</b>
LSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0564	0.0437	0.8445	0.0942	0.0735	1.2891
LSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0633	0.0476	0.9146	0.1040	0.0846	1.4801
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0993	0.0846	1.6379	0.1008	0.0801	1.4065
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0703	0.0555	1.0706	0.0930	0.0727	1.2715
GRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0855	0.0697	1.3450	0.1374	0.1179	2.0440
GRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0814	0.0659	1.2718	0.1814	0.1588	2.7479
GRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0587	0.0435	0.8359	0.1163	0.0965	1.6702
GRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0480	0.0365	0.7068	0.1180	0.0940	1.6180
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0391	0.0299	0.5833	0.1063	0.0872	1.5200
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0366	0.0283	0.5525	0.0971	0.0760	1.3224
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	<b>0.0350</b>	<b>0.0278</b>	<b>0.5394</b>	0.0728	0.0529	0.9204
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0359	0.0279	0.5437	0.1258	0.0997	1.7295
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0517	0.0383	0.7350	0.1145	0.0991	1.7255
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.0413	0.0323	0.6299	0.0733	0.0550	0.9596
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0425	0.0330	0.6388	<b>0.0654</b>	<b>0.0478</b>	<b>0.8395</b>
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.0484	0.0371	0.7173	0.1154	0.0975	1.6863

**Nota:** Os dois menores valores de cada coluna estão destacados em negrito.

**Fonte:** O autor, 2025.

A Tabela 8 apresenta as métricas de erro das previsões para 1 dia útil à frente realizadas pelo modelo RW e pelos modelos de DL univariados com 2CR. Verifica-se um padrão de desempenho similar ao descrito para os modelos multivariados com 1CR. As previsões do modelo RW permanecem com as menores métricas de erro entre as previsões de todos os modelos. Algumas especificações de modelos BiLSTM e BiGRU resultaram em previsões com desempenho relativamente próximo às previsões do RW na fase de validação, o que não se repetiu, em geral, no período de teste.

O modelo BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16), tal como no caso univariado, volta a ser o modelo de DL selecionado com as menores métricas de erro no período de validação. Seus resultados são relativamente próximos ao modelo de DL com menores métricas

no período de teste, destacadas em negrito na Tabela 8. De forma geral, no período de teste, observa-se que a inclusão de uma camada adicional na arquitetura dos modelos multivariados gerou algumas reduções pontuais, em termos de métricas de erro, especialmente nos modelos BiGRU.

### 5.1.2.3 Modelos de DL multivariados selecionados vs. modelo RW nas previsões 1 dia útil à frente

Tabela 9 - Métricas de erro das previsões 1 dia útil à frente dos modelos de DL multivariados selecionados em cada arquitetura vs. modelo RW

Arquitetura	Modelos e Parâmetros	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
-	RW	<b>0.0308</b>	<b>0.0230</b>	<b>0.4476</b>	<b>0.0478</b>	<b>0.0332</b>	<b>0.5805</b>
1CR	BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.0333	0.0257	0.5022	0.1083	0.0856	1.4846
2CR	BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.0350	0.0278	0.5394	0.0728	0.0529	0.9204

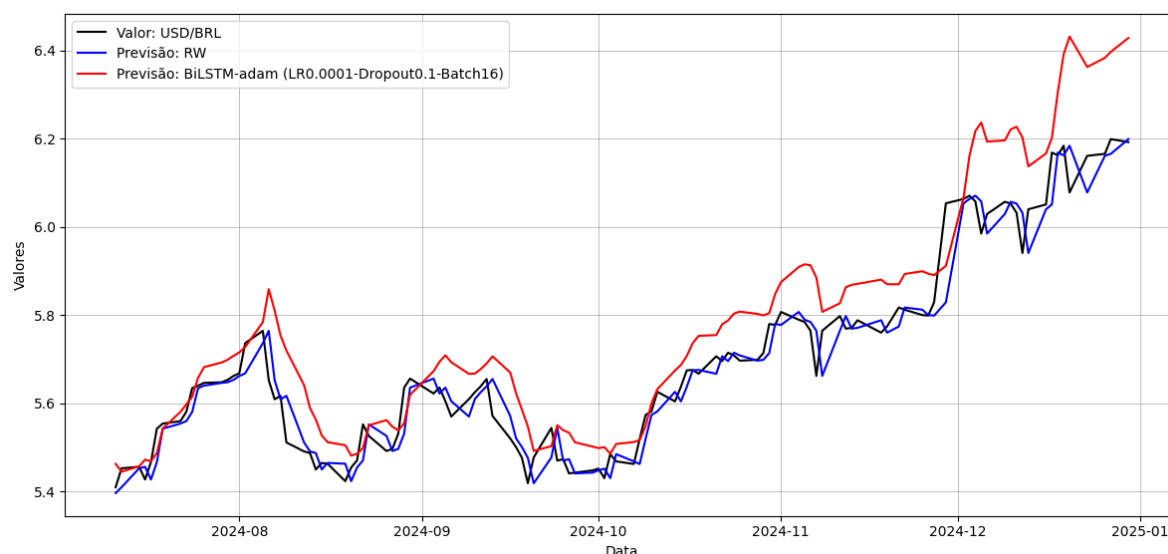
**Nota:** Menores valores destacados em negrito.

**Fonte:** O autor, 2025.

A Tabela 9 apresenta as métricas de erro das previsões 1 dia útil à frente do modelo RW e dos modelos de DL multivariados selecionados nas arquiteturas com 1CR e 2CR, pelo critério de menores métricas de erro no período de validação. Assim como no caso univariado, com o mesmo horizonte de previsão, o modelo RW apresenta as menores métricas de erro em termos absolutos, tanto no período de validação quanto no de teste. Entre os modelos de DL analisados, o modelo BiLSTM novamente se destaca, registrando as menores métricas de erro no período de validação em ambas as arquiteturas. No período de validação, as métricas de erro são ligeiramente menores do que às observadas no caso univariado. Contudo, no período de teste, os valores das métricas de erro são mais elevados e se distanciam ainda mais daqueles observados para o modelo RW, cujo desempenho aparenta ser o melhor entre todos os modelos analisados no horizonte de previsão de 1 dia útil à frente.

Entre os dois modelos de DL multivariados selecionados, o modelo com 2CR foi o que obteve as métricas de erro mais próximas às do RW no período de teste. Embora, neste caso, as métricas de erro não pareçam próximas às do modelo RW, serão apresentados testes estatísticos a fim de confirmar se os desempenhos das respectivas previsões não são estatisticamente equivalentes. Porém, antes disso, serão apresentados os gráficos das previsões dos modelos de DL selecionados multivariados com 1CR e com 2CR em relação às previsões do modelo RW e às taxas de câmbio Ptax de venda efetivas do período de teste.

Figura 8 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 1 dia útil à frente do modelo selecionado multivariado com 1CR vs RW no período de teste



**Nota:** Referente ao modelo selecionado multivariado, com 1CR, BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16).

**Fonte:** O autor, 2025.

A Figura 8 apresenta a taxa de câmbio Ptax de venda (BRL/USD) no período de teste, comparando as previsões 1 dia útil à frente do modelo RW e as do modelo de DL selecionado BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16) com 1CR que adota uma abordagem multivariada.

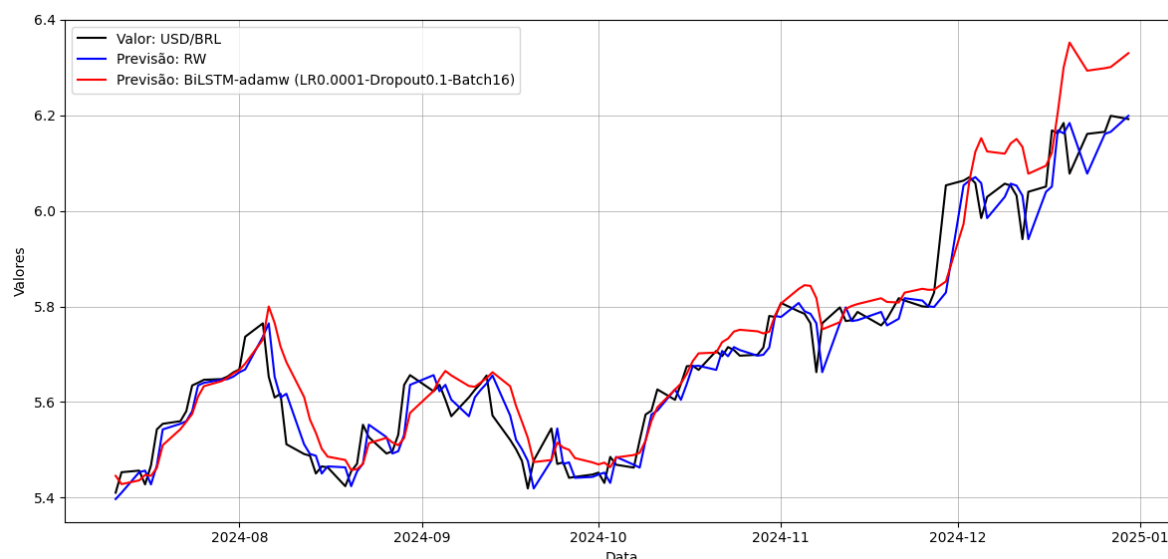
Observa-se que a introdução de variáveis explicativas no modelo de DL modifica seu comportamento em relação à versão univariada. Embora o modelo BiLSTM multivariado ainda apresente alguma suavização das oscilações de curto prazo, ele exibe um maior afastamento da trajetória real da taxa de câmbio Ptax de venda, especialmente em momentos de alta volatilidade. Esse efeito pode estar relacionado a uma maior dependência do modelo em relação às variáveis auxiliares, o que pode resultar em um descompasso temporal na resposta às mudanças abruptas do câmbio.

No início do período de teste, as previsões dos dois modelos comparados parecem próximas da série da taxa de câmbio Ptax de venda efetiva, com uma vantagem para as previsões do modelo RW. Entretanto, na segunda metade desse período, quando há um movimento mais acentuado e persistente de valorização do Dólar, nota-se um descolamento mais pronunciado das previsões do modelo BiLSTM em relação às do modelo RW e à taxa de câmbio Ptax efetiva. Esse comportamento pode indicar que o modelo de DL está atribuindo um peso excessivo às variáveis explicativas, o que pode comprometer sua capacidade de capturar variações bruscas que não estejam fortemente correlacionadas com os preditores utilizados.

Dessa forma, os resultados sugerem que, embora a abordagem multivariada possa enriquecer a modelagem da taxa de câmbio ao incorporar informações adicionais, sua implementação está sujeita a limitações quanto à disponibilidade de variáveis de alta frequência e

a capacidade das mesmas de incorporarem informações a respeito do comportamento futuro dessa variável.

Figura 9 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 1 dia útil à frente do modelo selecionado multivariado com 2CR vs RW no período de teste



**Nota:** Referente ao modelo selecionado multivariado, com 2CR, BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16).

**Fonte:** O autor, 2025.

A Figura 9 exibe a taxa de câmbio Ptax de venda (BRL/USD) no período de teste, comparando as previsões 1 dia útil à frente do modelo RW e as do modelo de DL selecionado BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16) com 2CR que adota uma abordagem multivariada.

Nota-se que, em relação ao modelo selecionado BiLSTM com apenas 1CR, o modelo selecionado com 2CR parece apresentar uma nova dinâmica das previsões. O modelo continua apresentando uma suavização das oscilações de curto prazo, característica comum das RNN, porém aparenta uma maior capacidade de acompanhar movimentos da taxa de câmbio Ptax de venda ao longo do tempo. Isso sugere que, dentro da sistemática adotada de seleção de modelo de DL, a profundidade adicional na rede pode ter contribuído para um aprendizado mais estruturado das sequências temporais.

Em relação à abordagem univariada com 2CR, no início do período de teste, ambas capturam de forma aproximada a respectiva taxa de câmbio, com diferenças sutis nas amplitudes das variações. Entretanto, à medida que a volatilidade aumenta, especialmente no final desse período, percebe-se que as previsões do modelo multivariado BiLSTM com 2CR exibem um comportamento distinto, com uma aparente superestimação da série real. Porém, esse viés parece menos acentuado do que no modelo multivariado selecionado com 1CR. Isso pode indicar que a maior profundidade da rede contribui para um ajuste mais refinado às dinâmicas não lineares do câmbio.

#### 5.1.2.4 Testes estatísticos entre o modelo de DL multivariado selecionado com melhor desempenho no período de teste vs. modelo RW nas previsões 1 dia útil à frente

Tabela 10 - Testes estatísticos para comparação do modelo multivariado de DL selecionado com 2CR e do modelo RW nas previsões 1 dia útil à frente

Testes	Período de Validação	Período de Teste
Teste de Diebold-Mariano (DM)	RW vs BiLSTM: Estatística = -2.3344, p-valor = 0.0196**	RW vs BiLSTM: Estatística = -2.8423, p-valor = 0.0045***
Teste de Diebold-Mariano-Modificado (DMM)	RW vs BiLSTM: Estatística = -2.3243, p-valor = 0.0219**	RW vs BiLSTM: Estatística = -2.8301, p-valor = 0.0055***
Teste de flutuação de Giacomini e Rossi (2010)	RW vs BiLSTM: Estatística = -2.9482, p-valor = 0.0083***	RW vs BiLSTM: Estatística = -2.6632, p-valor = 0.0154**
Teste de racionalidade de Rossi e Sekhposyan (2016)	RW: Estatística = 158.5189, p-valor = 0.0000***	RW: Estatística = 173.0916, p-valor = 0.0000***
Teste de racionalidade de Rossi e Sekhposyan (2016)	BiLSTM: Estatística = 374.6444, p-valor = 0.0000***	BiLSTM: Estatística = 373.6391, p-valor = 0.0000***

**Nota 1:** Refere-se ao BiLSTM (adamw-LR0.0001 - Dropout0.1 - Batch16).

**Nota 2:** \*\*\*, \*\* e \* indicam significância estatística ao nível de 1%, 5% e 10%, respectivamente.

**Fonte:** O autor, 2025.

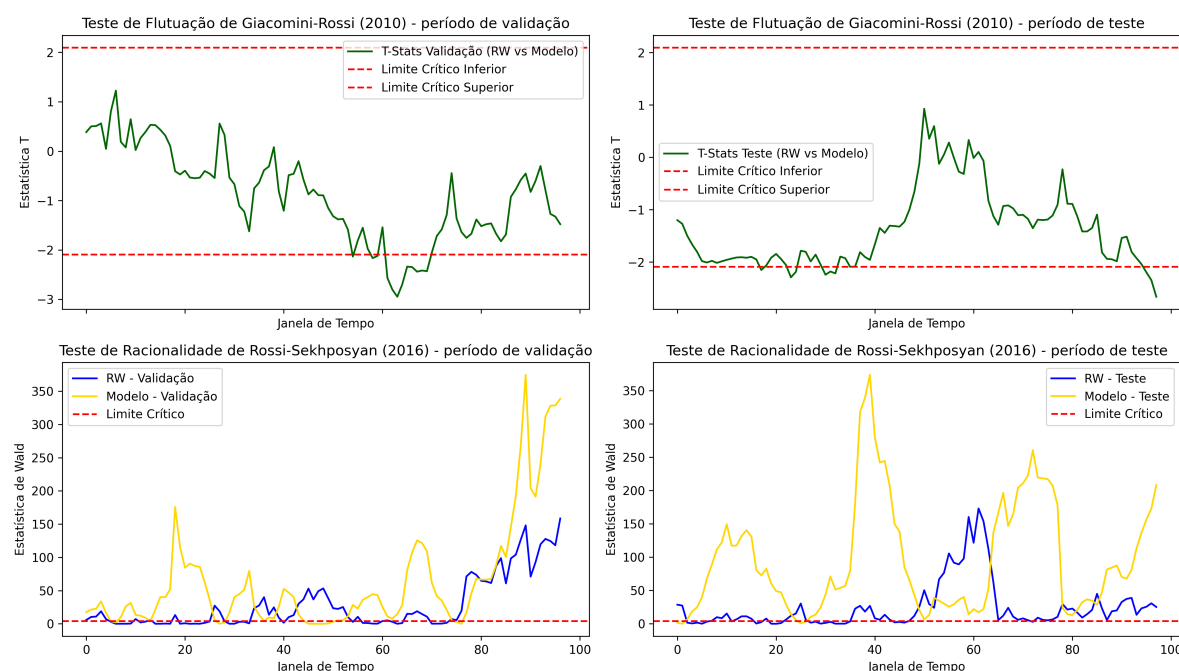
A Tabela 10 apresenta os resultados dos testes estatísticos comparando as previsões do modelo RW com as do modelo selecionado BiLSTM (adamw - LR 0.0001 - Dropout 0.1 - Batch 16) com 2CR, que obteve as menores métricas de erro no período de teste entre as duas arquiteturas. Os resultados dos testes para o modelo selecionado com 1CR podem ser consultados na Tabela 28, disponível no Apêndice A.

Assim como observado no caso univariado, os testes DM e DMM na tabela acima indicam que as previsões dos modelos apresentaram desempenhos, em termos de acurácia, estatisticamente diferentes. As respectivas  $H_0$  foram rejeitadas em ambos os testes ao nível de significância de 5%, tanto no período de validação quanto no de teste. Além disso, os resultados sugerem que o modelo RW apresentou desempenho estatisticamente superior ao modelo BiLSTM selecionado, conforme evidenciado pelas ET negativas e significativas.

Por outro lado, o teste de flutuação GR também indica a rejeição de  $H_0$ , sugerindo que a acurácia das previsões foi diferente em pelo menos uma janela de 20 dias úteis em ambos os períodos analisados. As respectivas ET negativas destacam a superioridade preditiva relativa do modelo RW nas janelas de tempo em que essa diferença preditiva foi mais pronunciada nos respectivos períodos, corroborando os resultados obtidos nos testes DM e DMM.

Em relação ao teste de racionalidade de RS, observa-se a rejeição da  $H_0$  de imprevisibilidade dos erros de previsão tanto para o modelo RW quanto para o modelo de DL multivariado selecionado, nos dois períodos analisados. Essa presença de irracionalidade indica que os erros preditivos podem não ser puramente aleatórios e que apresentam dependência temporal em pelo menos uma janela de tempo. No caso do modelo BiLSTM, as ET foram ainda mais elevadas do que as do modelo RW, especialmente no período de teste, sugerindo um viés mais acentuado em suas previsões. Esse comportamento pode indicar a necessidade de melhoria do modelo ou estar relacionado a mudanças estruturais na relação de curto prazo entre as variáveis explicativas, possivelmente influenciadas por fatores externos, como as já citadas eleições americanas.

Figura 10 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação



**Nota:** O "Modelo" descrito na legenda refere-se ao modelo selecionado multivariado, com 2CR, nas previsões 1 dia útil à frente. As janelas móveis possuem 20 dias úteis nos dois testes.

**Fonte:** O autor, 2025.

Conforme ilustrado na Figura 10, observa-se que, durante a maior parte do período de validação, a ET do teste de flutuação GR permaneceu dentro dos limites críticos, indicando uma equivalência preditiva entre os modelos nessas janelas de tempo de 20 dias úteis. Entretanto, na segunda metade do período de validação, as previsões do modelo RW demonstraram alguns episódios de acurácia relativamente superior, de forma semelhante ao observado no caso univariado. No período de teste, a ET começa próxima ao limite crítico inferior, cruzando-o algumas vezes e sugerindo novamente uma superioridade da acurácia das previsões do modelo RW. Na parte intermediária das janelas de tempo do período de teste, as previsões permanecem dentro dos limites críticos até que, no período final, as previsões do modelo RW apresentam um desempenho estatisticamente superior.

Por outro lado, a evolução da ET no teste de racionalidade de Rossi e Sekhposyan (2016) revela que o modelo de DL selecionado passou por diversos episódios em que os erros de previsões foram irracionais, especialmente no final do período de validação e em múltiplos momentos ao longo do período de teste. Esses episódios evidenciam ineficiência e viés, com a ET frequentemente ultrapassando de forma significativa o limite crítico. Tal comportamento pode ser atribuído a diversos fatores externos, já mencionados na análise univariada, que aumentaram a volatilidade da taxa de câmbio e resultaram em uma considerável depreciação do Real frente ao Dólar ao longo de 2024.

Os resultados obtidos com os modelos de DL multivariados aplicados à previsão de 1 dia útil à frente indicam que, embora os modelos de DL tenham um grande potencial, mesmo

com as diferentes especificações adotadas neste estudo, sua aplicabilidade para previsões de curtíssimo prazo da taxa de câmbio continua sendo um desafio, mesmo com a inclusão de variáveis explicativas.

## 5.2 Resultados das previsões 10 dias úteis à frente

### 5.2.1 Modelos univariados de DL vs. modelo RW nas previsões 10 dias úteis à frente:

#### 5.2.1.1 Modelos univariados de DL com 1CR vs. modelo RW nas previsões 10 dias úteis à frente:

Tabela 11 - Métricas de erro das previsões dos modelos univariados de DL com 1CR e do modelo RW 10 dias úteis à frente

Modelos (uma camada recorrente)	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
RW	<b>0.1140</b>	<b>0.0889</b>	<b>1.6886</b>	<b>0.1461</b>	<b>0.1251</b>	<b>2.1680</b>
LSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.1390	0.1005	1.8942	0.3462	0.2966	5.0549
LSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1415	0.1021	1.9238	0.3511	0.3061	5.2265
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1422	0.1012	1.9064	0.3247	0.2814	4.8018
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1353	0.1060	2.0150	0.3512	0.3086	5.2715
GRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.1335	0.0991	1.8724	0.2879	0.2545	4.3537
GRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1318	0.1014	1.9246	0.2830	0.2514	4.3060
GRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1352	0.1023	1.9381	0.2768	0.2453	4.1992
GRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1251	0.0987	1.8764	0.2717	0.2388	4.0841
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	<b>0.1018</b>	<b>0.0888</b>	<b>1.7065</b>	0.1554	0.1336	2.3090
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1295	0.0973	1.8391	0.3124	0.2679	4.5700
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1142	0.1024	1.9753	<b>0.1450</b>	<b>0.1227</b>	<b>2.1356</b>
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1251	0.0938	1.7741	0.2771	0.2447	4.1978
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.1243	0.0951	1.8036	0.2517	0.2222	3.8053
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1219	0.0932	1.7677	0.2576	0.2270	3.8842
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1207	0.0930	1.7640	0.2445	0.2158	3.6955
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1223	0.0935	1.7742	0.2571	0.2257	3.8603

**Nota:** Os dois menores valores de cada coluna estão destacados em negrito.

**Fonte:** O autor, 2025.

A Tabela 11 apresenta as métricas de erro das previsões 10 dias úteis à frente realizadas pelo modelo RW e pelos modelos de DL univariados com 1CR. Pela primeira vez, com exceção do MAPE no período de validação, o RW não obteve as menores métricas de erro de previsão entre todos os modelos analisados em ambos os períodos. No entanto, seus resultados permaneceram próximos aos do modelo de DL com menores métricas de erro no período de teste.

O modelo BiLSTM (adam - LR 0.0001 - Dropout 0.1 - Batch 16) registrou as menores métricas de erro no período de validação, sendo, portanto, selecionado para a análise comparativa com o RW no período de teste. No período de teste, os erros de previsão do modelo RW foram ligeiramente inferiores aos do modelo selecionado.

Neste caso, com algumas exceções, os modelos bidirecionais (BiLSTM e BiGRU) demonstraram métricas de erro relativamente próximas às dos modelos de DL unidire-

cionais (LSTM e GRU). No período de teste, a diferença de desempenho entre os modelos de DL e o RW tornou-se mais evidente, sugerindo que, de forma geral, os modelos de DL univariados, nas diferentes especificações adotadas, tiveram dificuldades em captar o comportamento dinâmico da taxa de câmbio.

Por fim, cabe observar que as métricas de erro apresentaram um aumento conforme o horizonte de previsão se ampliou para 10 dias úteis à frente, como é comum de se observar na previsão de séries temporais.

### 5.2.1.2 Modelos univariados de DL com 2CR vs. modelo RW nas previsões 10 dias úteis à frente:

Tabela 12 - Métricas de erro das previsões dos modelos univariados de DL com 2CR e do modelo RW 10 dias úteis à frente

Modelos (duas camadas recorrentes)	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
RW	<b>0,1140</b>	<b>0,0889</b>	<b>1,6886</b>	<b>0,1461</b>	<b>0,1251</b>	<b>2,1680</b>
LSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0,1714	0,1215	2,2834	0,4538	0,4005	6,8434
LSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0,1563	0,1107	2,0829	0,4350	0,3918	6,7069
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0,1828	0,1329	2,5017	0,4361	0,3959	6,7841
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0,1516	0,1075	2,0215	0,3930	0,3473	5,9360
GRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0,1467	0,1054	1,9850	0,3156	0,2831	4,8514
GRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0,1264	0,0971	1,8403	0,2920	0,2577	4,4077
GRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0,1333	0,0986	1,8640	0,2962	0,2641	4,5228
GRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0,1248	0,0973	1,8476	0,2835	0,2490	4,2584
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0,1335	0,0985	1,8578	0,3588	0,3112	5,3094
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0,1397	0,1026	1,9354	0,5182	0,3950	6,6785
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0,1333	0,0989	1,8657	0,4326	0,3884	6,6448
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	<b>0,1123</b>	<b>0,0873</b>	<b>1,6613</b>	0,4563	0,3580	6,0564
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0,1353	0,0979	1,8458	0,2961	0,2641	4,5229
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0,1253	0,0974	1,8484	0,3095	0,2699	4,6080
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0,1254	0,0928	1,7532	<b>0,2807</b>	<b>0,2450</b>	<b>4,1846</b>
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0,1231	0,0928	1,7575	0,2955	0,2570	4,3886

**Nota:** Os dois menores valores de cada coluna estão destacados em negrito.

**Fonte:** O autor, 2025.

A Tabela 12 apresenta as métricas de erro das previsões para 10 dias úteis à frente realizadas pelo modelo RW e pelos modelos de DL univariados com 2CR. Verifica-se que o RW obteve as menores métricas de erro no período de teste, enquanto o modelo BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32) apresentou as menores métricas de erro de previsão no período de validação, embora com valores próximos aos do RW. De forma geral, no período de teste, as previsões dos modelos de DL univariados exibiram métricas de erro de previsão visivelmente superiores às do RW. A inclusão de uma camada adicional nos respectivos modelos de DL, mantidos os parâmetros da arquitetura com 1CR, não resultou em melhorias nas métricas de erro no período de teste. Além disso, observa-se que as previsões dos modelos BiGRU apresentaram métricas de erro relativamente próximas às do modelo GRU nos dois períodos de análise, o que não parece se repetir para o caso dos modelos BiLSTM e LSTM.

Verifica-se que as previsões 10 dias à frente dos modelos de DL univariados com 2CR



apresentaram métricas de erro superiores às obtidas pelos modelos de DL com 1CR. Assim, apesar das diferentes especificações adotadas, tais modelos apresentaram limitações na captura dos padrões cambiais do período analisado.

### 5.2.1.3 Modelos de DL univariados selecionados vs. modelo RW nas previsões 10 dias úteis à frente

Tabela 13 - Métricas de erro das previsões 10 dias úteis à frente dos modelos de DL univariados selecionados em cada arquitetura vs. modelo RW

Arquitetura	Modelos e Parâmetros	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
-	RW	0.1140	0.0889	1.6886	<b>0.1461</b>	<b>0.1251</b>	<b>2.1680</b>
1CR	BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	<b>0.1018</b>	0.0888	1.7065	0.1554	0.1336	2.3090
2CR	BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1123	<b>0.0874</b>	<b>1.66132</b>	0.4563	0.3580	6.0564

**Nota:** Menores valores destacados em negrito.

**Fonte:** O autor, 2025.

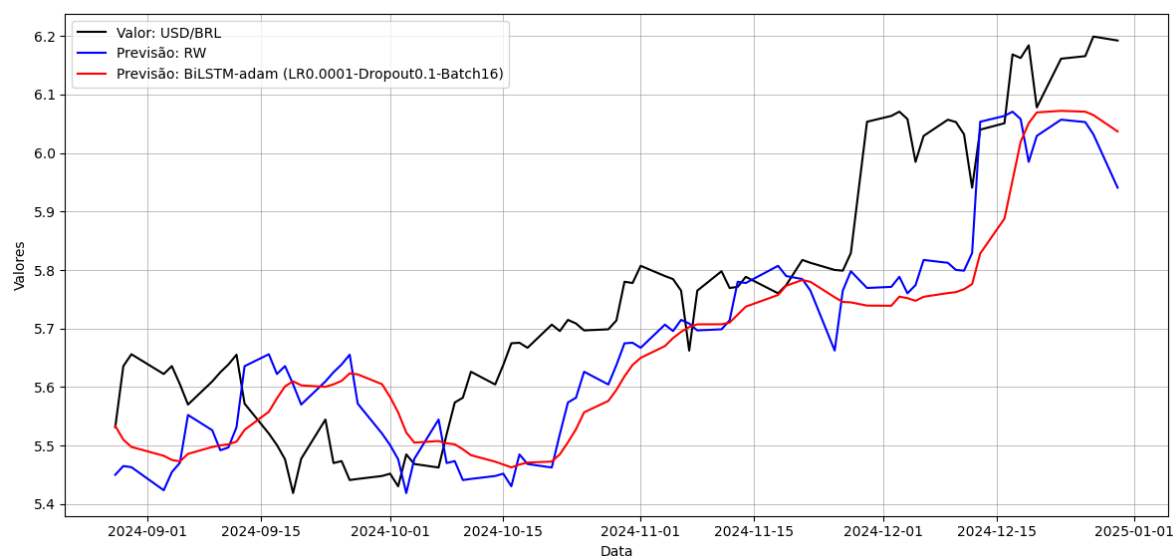
A Tabela 13 apresenta as métricas de erro das previsões 10 dias úteis à frente do modelo RW e dos modelos de DL univariados selecionados nas arquiteturas com 1CR e 2CR, pelo critério de menores métricas de erro no período de validação. No período de validação, os modelos de DL obtiveram menores métricas de erro em comparação ao modelo RW. O modelo com 1CR apresentou o menor RMSE, enquanto o modelo com 2CR obteve os menores valores de MAE e MAPE. Já no período de teste, o modelo RW obteve os menores erros de previsão nas três métricas.

O RMSE mede o erro médio em unidades absolutas e é sensível a grandes desvios (*outliers*). Assim, no caso do modelo de DL com 1CR, foram obtidas previsões mais precisas em termos absolutos, reduzindo discrepâncias extremas e de forma mais robusta na média global dos erros. O MAPE, por sua vez, mede o erro médio em termos percentuais e é mais adequado para comparar erros relativos, especialmente em séries temporais com valores menores, onde desvios proporcionais podem ter maior impacto. Dessa forma, o menor MAPE do modelo com 2CR sugere que ele teve um desempenho relativamente melhor na minimização de erros em relação aos valores previstos.

Por outro lado, o MAE mede o erro médio absoluto sem penalizar tanto os desvios extremos como o RMSE, refletindo a precisão geral das previsões. Logo, o menor MAE do modelo com 2CR no período de validação indica que, em média, suas previsões estiveram mais próximas dos valores reais em relação aos demais modelos.

Entre os modelos de DL selecionados, o modelo com 1CR apresentou as menores métricas de erro durante o período de teste. Dado que seus resultados foram próximos aos do modelo RW, serão realizados testes estatísticos para verificar se os desempenhos das respectivas previsões podem ser considerados equivalentes em termos de acurácia preditiva. Antes disso, serão apresentados os gráficos das previsões dos dois modelos de DL selecionados e do modelo RW.

Figura 11 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 10 dias úteis à frente do modelo selecionado univariado com 1CR vs RW no período de teste



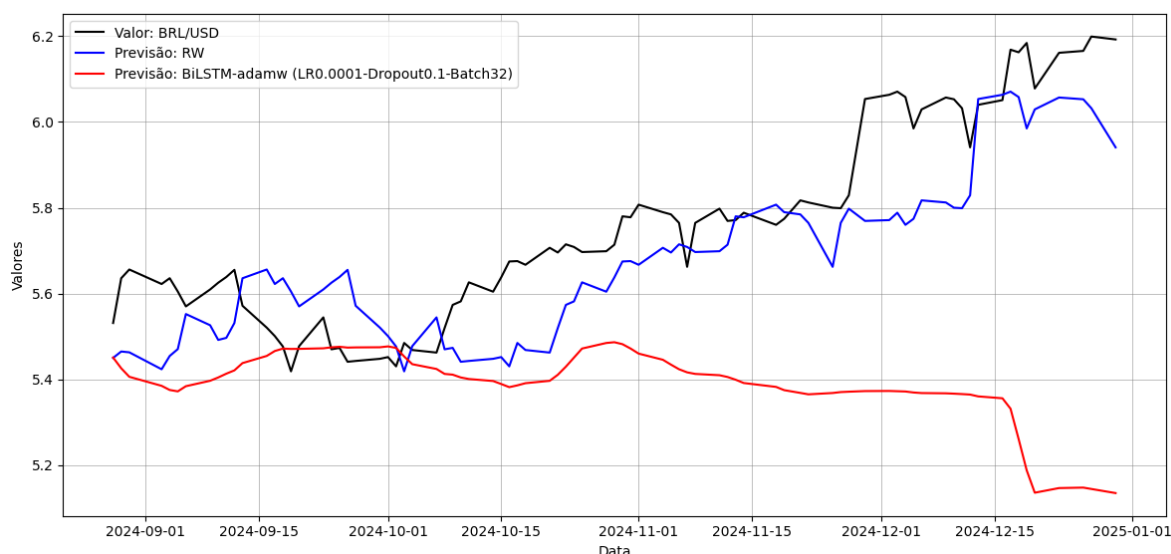
**Nota:** Referente ao modelo de DL univariado selecionado, com 1CR, descrito na legenda.

**Fonte:** O autor, 2025.

A Figura 11 apresenta a taxa de câmbio Ptax de venda (BRL/USD) no período de teste, comparando as previsões 10 dias úteis à frente do modelo RW e as do modelo de DL selecionado BiLSTM (adam - LR 0.0001 - Dropout 0.1 - Batch 16) com 1CR que adota uma abordagem univariada.

Observa-se que, no período de teste, o modelo de DL apresenta um comportamento mais suavizado em relação à trajetória efetiva da taxa de câmbio Ptax de venda. Essa suavização parece decorrer da dificuldade do modelo univariado em captar oscilações de curto prazo. As previsões do referido BiLSTM parecem relativamente próximas às do modelo RW, porém com suavização e certo atraso em relação a este. Isso evidencia os desafios em lidar com reversões abruptas, especialmente em cenários de elevada volatilidade.

Figura 12 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 10 dias úteis à frente do modelo selecionado univariado com 2CR vs RW no período de teste



**Nota:** Referente ao modelo de DL univariado selecionado, com 2CR, descrito na legenda.

**Fonte:** O autor, 2025.

A Figura 12 exibe a taxa de câmbio Ptax de venda (BRL/USD) no período de teste, comparando as previsões 10 dias úteis à frente do modelo RW e as do modelo de DL selecionado BiLSTM-adamw (LR 0.0001 - Dropout 0.1 - Batch 32) com 2CR que adota uma abordagem univariada.

Observa-se que, diferentemente dos modelos anteriores, o BiLSTM apresenta um viés acentuado e crescente no final do período de teste, sugerindo uma perda de estabilidade nas previsões.

Embora as funções de perda no treinamento e na validação, disponíveis no Apêndice B, apresentem um decaimento aparentemente normal, sem sinais evidentes de *overfitting* ou *underfitting*, verifica-se uma inconsistência no desempenho do modelo entre os períodos de validação e de teste. É possível que a rede neural tenha convergido para um mínimo local subótimo, resultando em pesos que não favorecem uma generalização adequada. Além disso, o período de teste coincidiu com choques externos, como as eleições americanas, que impactaram as expectativas e a valorização da taxa de câmbio. Mudanças abruptas na estrutura dos dados, incluindo tendência e volatilidade, podem não ter sido capturadas adequadamente pelo modelo de DL durante o respectivo treinamento.

Apesar das diversas especificações dos modelos de DL com 2CR, a ausência de variáveis exógenas pode ter limitado a capacidade da rede de capturar padrões de tendência e volatilidade, resultando na subestimação da taxa de câmbio. Estudos futuros podem explorar calibrações adicionais de parâmetros, arquiteturas híbridas ou estratégias de treinamento com objetivo de aprimorar a adaptação dos modelos de DL univariados a períodos de instabilidade no mercado. Esse exemplo evidencia que a seleção do modelo é, por si só, um grande desafio. Além das

inúmeras combinações possíveis de arquiteturas e hiperparâmetros, um desempenho favorável no período de validação não garante resultados satisfatórios no período de teste e nos períodos posteriores. A comparação dos resultados das previsões dos respectivos modelo de DL com o modelo RW (*benchmark*) pode ser um indicativo da inadequação dos mesmos nos casos de afastamentos substanciais.

#### 5.2.1.4 Testes estatísticos entre o modelo de DL univariado selecionado com melhor desempenho no período de teste vs. modelo RW nas previsões 10 dias úteis à frente

Tabela 14 - Testes estatísticos para comparação do modelo univariado de DL selecionado com 1CR e do modelo RW nas previsões 10 dias úteis à frente

Testes	Período de Validação	Período de Teste
Diebold-Mariano (DM)	RW vs BiLSTM: Estatística = 1.1986, p-valor = 0.2307	RW vs BiLSTM: Estatística = -1.2034, p-valor = 0.2288
Teste de Diebold-Mariano-Modificado (DMM)	RW vs BiLSTM: Estatística = 1.0597, p-valor = 0.2924	RW vs BiLSTM: Estatística = -1.0657, p-valor = 0.2897
Flutuação de Giacomini e Rossi (2010)	RW vs BiLSTM: Estatística = 5.0494, p-valor = 0.0001***	RW vs BiLSTM: Estatística = -3.2396, p-valor = 0.0043***
Racionalidade de Rossi e Sekhposyan (2016)	RW: Estatística = 5779.2941, p-valor = 0.0000***	RW: Estatística = 2980.7553, p-valor = 0.0000***
Racionalidade de Rossi e Sekhposyan (2016)	BiLSTM: Estatística = 3707.2103, p-valor = 0.0000***	BiLSTM: Estatística = 4836.8838, p-valor = 0.0000***

**Nota 1:** Refere-se ao BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32).

**Nota 2:** \*\*\*, \*\* e \* indicam significância estatística ao nível de 1%, 5% e 10%, respectivamente.

**Fonte:** O autor, 2025.

A Tabela 14 apresenta os resultados dos testes estatísticos comparando as previsões do modelo RW com as do modelo selecionado BiLSTM (adamw - LR 0.0001 - Dropout 0.1 - Batch 16) com 1CR, que obteve as menores métricas de erro no período de teste entre as duas arquiteturas. Os resultados dos testes para o modelo selecionado com 2CR podem ser consultados na Tabela 29, disponível no Apêndice A.

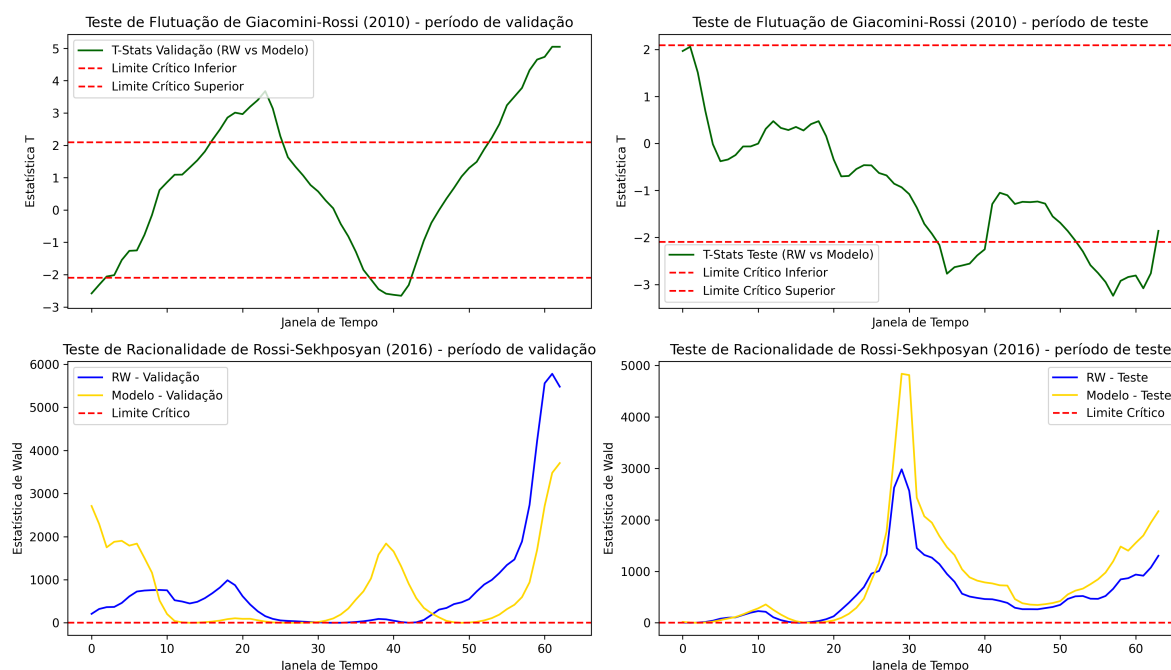
Os resultados dos testes DM e DMM na tabela acima indicam que as previsões dos modelos apresentaram acurácias estatisticamente equivalentes. As  $H_0$  não são rejeitadas nos dois testes ao nível de significância de 5%, tanto no período de validação quanto no de teste.

Por outro lado, o teste de flutuação GR aponta para a rejeição da  $H_0$  de igualdade de desempenho relativo das previsões. Porém, a ET considera a janela de tempo de maior diferença de acurácia entre as previsões dos modelo RW e as do respectivo modelo de DL selecionado com 1CR. Dessa forma, no período de validação, a ET positiva indica a superioridade da acurácia das previsões do modelo de DL selecionado na janela de tempo de maior diferença de acurácia, enquanto que a ET negativa, no período de teste, indica a superioridade da acurácia das previsões do modelo RW também na janela de tempo de maior diferença de acurácia. A Figura 13, apresentada na sequência, ilustrará o comportamento do desempenho preditivo relativo ao longo dos períodos de validação e de teste.

O teste de racionalidade de RS aponta a rejeição de  $H_0$  de imprevisibilidade dos erros de previsão no modelo RW e no modelo de DL selecionado, tanto no período de validação e quanto no de teste. Isso indica a presença de eventos de irracionalidade, nos quais os erros preditivos não foram aleatórios e apresentaram dependência temporal. Neste caso, as ET do modelo de

DL são menores que as do RW no período de validação e maiores no período de teste.

Figura 13 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação



**Nota:** O "Modelo" descrito na legenda refere-se ao modelo selecionado univariado, com 1CR, nas previsões 10 dias úteis à frente. As janelas móveis possuem 20 dias úteis nos dois testes.

**Fonte:** O autor, 2025.

Conforme ilustrado na Figura 13, a evolução das ETs no teste de flutuação de GR aponta que, durante a maior parte do tempo, a ET permaneceu dentro dos limites críticos, indicando uma equivalência estatística entre as previsões dos dois modelos comparados. No período de validação, observa-se uma alternância no desempenho relativo das previsões, com uma leve superioridade inicial do modelo RW e uma superioridade do modelo de DL no período final. Por sua vez, na segunda metade do período de teste, observam-se janelas de tempo em que as previsões do modelo RW demonstram uma acurácia superior.

Por outro lado, a evolução da ET no teste de racionalidade de RS revela que, no período de validação, o modelo de DL selecionado apresentou três momentos de maior irracionalidade nos seus erros de previsão, sugerindo falta de eficiência e viés, com a ET ultrapassando significativamente o limite crítico, especialmente no final desse período. Por sua vez, o modelo RW exibiu erros de previsão ainda mais irracionais no final do período de validação. No período de teste, as ETs de ambos os modelos apresentam um comportamento semelhante, com um pico de irracionalidade dos erros de previsão nas janelas centrais, com as ET do modelo de DL superando as ET do modelo RW.

De modo geral, os resultados obtidos nos modelos de DL univariados aplicados à previsão de 10 dias úteis à frente indicam que, embora os modelos de DL tenham um grande potencial, nas diversas variações adotadas neste estudo, sua aplicabilidade às previsões de curto

prazo pode ser restrita, especialmente em contextos de alta volatilidade.

## 5.2.2 Modelos multivariados de DL vs. modelo RW nas previsões 10 dias úteis à frente

### 5.2.2.1 Modelos multivariados de DL com 1CR vs. modelo RW nas previsões 10 dias úteis à frente:

Tabela 15 - Métricas de erro das previsões dos modelos multivariados de DL com 1CR e do modelo RW 10 dias úteis à frente

Modelo	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
RW	<b>0.1140</b>	<b>0.0889</b>	<b>1.6886</b>	<b>0.1461</b>	<b>0.1251</b>	<b>2.1680</b>
LSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.2259	0.1850	3.5072	0.4277	0.3934	6.7467
LSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.3347	0.2949	5.6180	0.6654	0.6421	11.0748
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.2817	0.2446	4.6544	0.5669	0.5179	8.8736
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1705	<b>0.1287</b>	<b>2.4325</b>	<b>0.1431</b>	<b>0.1109</b>	<b>1.8799</b>
GRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.2982	0.2604	4.9567	0.7167	0.6918	11.9333
GRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.3039	0.2681	5.1071	0.6647	0.6361	10.9596
GRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.2597	0.2168	4.1128	0.5439	0.5162	8.8858
GRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.2483	0.2101	3.9901	0.5989	0.5695	9.8035
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.2278	0.1948	3.7039	0.2415	0.2045	3.4928
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.2374	0.2065	3.9309	0.2628	0.2395	4.1221
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.2710	0.2279	4.3278	0.5938	0.5590	9.6105
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	<b>0.1529</b>	0.1388	2.6785	0.4074	0.3232	5.5089
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.2405	0.2029	3.8539	0.5268	0.4934	8.4773
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.2571	0.2230	4.2439	0.5287	0.5048	8.6988
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.2800	0.2441	4.6459	0.5665	0.5402	9.3036
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.2396	0.1983	3.7601	0.5195	0.4947	8.5200

**Nota:** Os dois menores valores de cada coluna estão destacados em negrito.

**Fonte:** O autor, 2025.

A Tabela 15 apresenta as métricas de erro das previsões para 10 dias úteis à frente realizadas pelo modelo RW e pelos modelos de DL multivariados com 1CR. No resultado geral, as previsões do modelo RW exibiram as menores métricas de erro no período de validação. Entre os modelos de DL, no período de validação, o BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32) obteve o menor RMSE, enquanto o modelo LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32) obteve os menores MAE e MAPE. Desse forma, ambos os modelos foram pré-selecionados entre os modelos de DL com 1CR.

No período de teste, com exceção do modelo LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32), as métricas de erro das previsões dos modelos de DL foram superiores às apresentadas pelos modelos univariados com o mesmo horizonte de previsão de 10 dias úteis à frente. Esse desempenho pode estar relacionado à maior complexidade das dinâmicas introduzidas pelas variáveis explicativas, bem como a eventuais mudanças no comportamento dessas variáveis em função do choque externo ocorrido no período.

Adicionalmente, observa-se que os modelos de DL bidirecionais não apresentaram vantagens claras em relação aos modelos unidirecionais. O modelo LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32) demonstrou consistência em suas previsões tanto no período de vali-

dação quando no de teste, destacando-se entre os modelos de DL especialmente nesse último período.

#### 5.2.2.2 Modelos multivariados de DL com 2CR vs. modelo RW nas previsões 10 dias úteis à frente:

Tabela 16 - Métricas de erro das previsões dos modelos multivariados de DL com 2CR e do modelo RW 10 dias úteis à frente

Modelo	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
RW	<b>0.1140</b>	<b>0.0889</b>	<b>1.6886</b>	<b>0.1461</b>	<b>0.1251</b>	<b>2.1680</b>
LSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.2308	0.1880	3.5637	0.3676	0.3316	5.6753
LSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1216	0.0901	1.7006	0.4024	0.3854	6.7879
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.2370	0.2005	3.8114	0.2562	0.2114	3.5923
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.2865	0.2486	4.7311	0.4009	0.3687	6.3306
GRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.2653	0.2290	4.3561	0.5252	0.4979	8.5719
GRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.2591	0.2281	4.3455	0.5433	0.5192	8.9455
GRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.2677	0.2316	4.4055	0.5432	0.5189	8.9413
GRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.2571	0.2233	4.2516	0.5722	0.5435	9.3563
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.2435	0.1940	3.6690	0.6180	0.5892	10.1452
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	<b>0.1125</b>	<b>0.0840</b>	<b>1.5919</b>	<b>0.1046</b>	<b>0.0808</b>	<b>1.4158</b>
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1391	0.1236	2.3861	0.2071	0.1494	2.5402
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1279	0.0949	1.7903	0.1611	0.1446	2.4944
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.1695	0.1356	2.5679	0.4287	0.4013	6.8994
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1754	0.1284	2.4170	0.4087	0.3761	6.4494
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.2207	0.1718	3.2460	0.5529	0.5311	9.1588
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.2332	0.1914	3.6272	0.6187	0.5947	10.2517

**Nota:** Os dois menores valores de cada coluna estão destacados em negrito.

**Fonte:** O autor, 2025.

A Tabela 16 apresenta as métricas de erro das previsões para 10 dias úteis à frente realizadas pelo modelo RW e pelos modelos de DL multivariados com 2CR. No resultado geral, o modelo BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32) apresentou as menores métricas de erro tanto no período de validação quanto no de teste, demonstrando consistência e capacidade de generalização nas suas previsões.

Com algumas exceções, as métricas de erro das previsões 10 dias à frente dos modelos de DL multivariados com 2CR foram ligeiramente inferiores às observadas na arquitetura com 1CR. Cabe observar que, enquanto na arquitetura com 1CR o modelo de DL selecionado foi um dos LSTM, nessa arquitetura com 2CR foi um dos BiLSTM. Desse modo, na sistemática adotada, não se identificou uma superioridade consistente dos modelos de DL bidirecionais em relação aos unidirecionais, sendo que os resultados variaram de acordo com a arquitetura empregada.

### 5.2.2.3 Modelos de DL multivariados selecionados vs. modelo RW nas previsões 10 dias úteis à frente

Tabela 17 - Métricas de erro das previsões 10 dias úteis à frente dos modelos de DL multivariados selecionados em cada arquitetura vs. modelo RW

Arquitetura	Modelos e Parâmetros	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
-	RW	0.1140	0.0889	1.6886	0.1461	0.1251	2.1680
1CR (pelo critério RMSE)	BiLSTM-adamw (LR0.0001-Dropout0.1-Batch32)	0.1529	0.1388	2.6785	0.4074	0.3232	5.5089
1CR (pelo critério MAPE)	LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1705	0.1287	2.4325	0.1431	0.1109	1.8799
2CR	BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	<b>0.1125</b>	<b>0.0840</b>	<b>1.5919</b>	<b>0.1046</b>	<b>0.0808</b>	<b>1.4158</b>

**Nota:** Menores valores destacados em negrito.

**Fonte:** O autor, 2025.

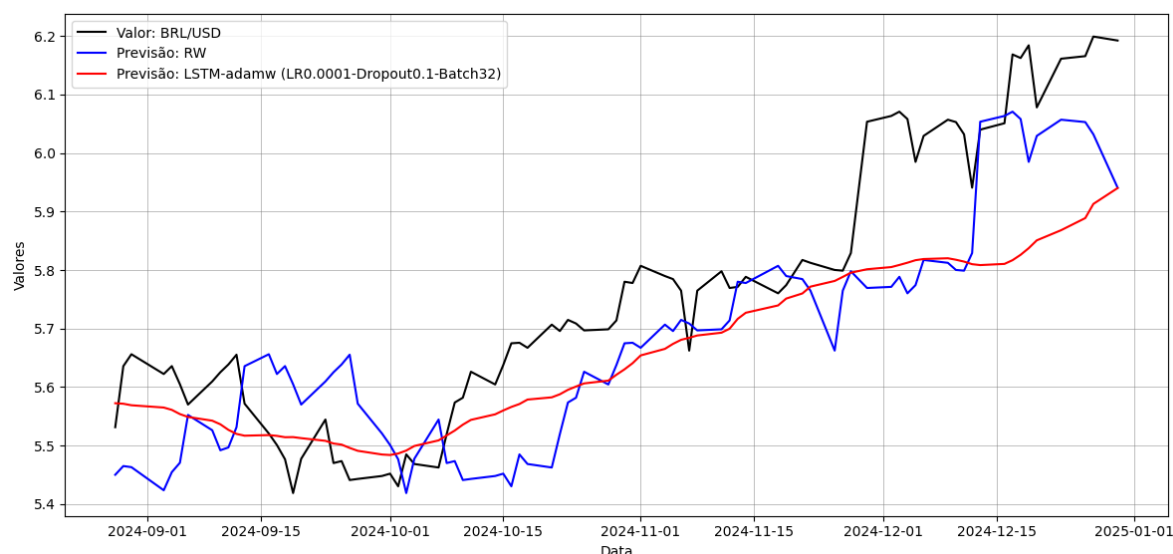
A Tabela 17 apresenta as métricas de erro das previsões 10 dias úteis à frente do modelo RW e dos modelos de DL multivariados selecionados nas arquiteturas com 1CR e 2CR, pelo critério de menores métricas de erro no período de validação.

Seguindo a sistemática adotada, entre todos os modelos, o BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32) com 2CR obteve as menores métricas de erro em suas previsões nos dois períodos analisados. O desempenho em relação ao modelo RW parece mais próximo no período de validação do que no período de teste. Por outro lado, entre os modelos com 1CR, o LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32) apresentou métricas de erro de previsão relativamente próximas às do modelo RW no período de teste.

Entre todos os modelos, o modelo de DL com 2CR apresentou as menores métricas de erro durante o período de teste. Dado que suas métricas de erro foram inferiores às das previsões do modelo RW, serão apresentados testes estatísticos com objetivo de verificar se as acurácias das previsões são estatisticamente diferentes. Antes disso, serão apresentados os gráficos das previsões dos dois modelos selecionados, com melhor desempenho no período de teste, em relação às previsões do modelo RW e às taxas de câmbio Ptax de venda efetivas.



Figura 14 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 10 dias úteis à frente do modelo selecionado multivariado com 1CR vs RW no período de teste



**Nota:** Menores valores destacados em negrito.

**Fonte:** O autor, 2025.

A Figura 14 apresenta a taxa de câmbio Ptax de venda (BRL/USD) no período de teste, comparando as previsões 10 dias úteis à frente do modelo RW e as do modelo de DL selecionado LSTM (adamw - LR0.0001 - Dropout 0.1 - Batch 32) com 1CR que adota uma abordagem multivariada.

Observa-se que as previsões do referido modelo LSTM multivariado apresentam uma trajetória mais suavizada em relação à série efetiva da taxa de câmbio Ptax de venda, sugerindo menor capacidade de captura de oscilações de curtíssimo prazo. Porém, tais previsões parecem acompanhar a tendência da taxa de câmbio Ptax de venda no período.

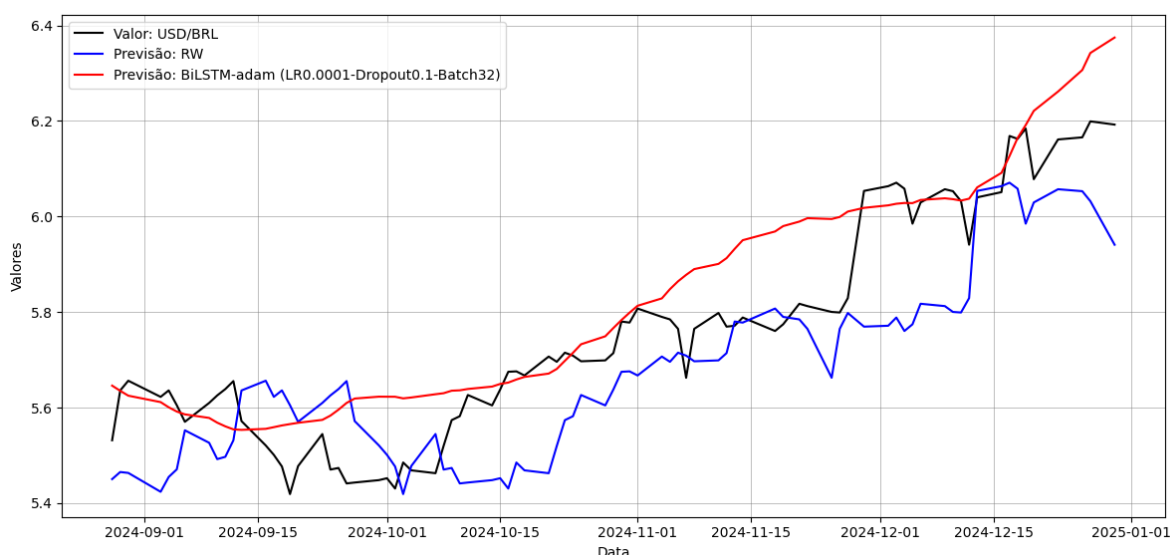
Esse resultado pode estar associado ao tamanho do lote (*batch size*) de 32, que influencia as atualizações dos pesos durante o treinamento. Em modelos recorrentes, tamanho de lotes maiores tendem a proporcionar ajustes mais estáveis dos gradientes, reduzindo a variabilidade nas atualizações, o que pode melhorar a generalização mas limitando a detecção de padrões dinâmicos e de curta duração. Por outro lado, o resultado também pode estar relacionado ao impacto das técnicas de regularização, como o *dropout* e a penalidade L2 aplicada aos pesos do *kernel*, que atuam limitando a magnitude dos ajustes durante o treinamento, resultando em previsões menos extremas e, conseqüentemente, mais suavizadas. O uso de uma taxa de aprendizado mais baixa (LR = 0.0001) também pode ter contribuído para uma convergência mais estável durante o treinamento, ao reduzir o impacto das oscilações no ajuste dos pesos da rede neural ao longo do processo de otimização.

Adicionalmente, o modelo de DL selecionado aparenta exibir uma tendência sistemática de subestimação da trajetória efetiva da taxa de câmbio Ptax de venda, especialmente na segunda metade do período de teste, em que ocorreu uma apreciação mais acelerada da taxa de

câmbio.

Experimentos empíricos preliminares indicaram que, mantendo-se os demais parâmetros constantes, o uso de taxas de aprendizado maiores resultou em um treinamento menos estável e em modelos de DL cujo desempenho foi comparável ao do modelo RW apenas no período de validação, evidenciando uma capacidade limitada de generalização no período de teste. Esses resultados ressaltam a importância de buscar um equilíbrio entre os parâmetros que favoreçam a capacidade de generalização dos modelos de DL ao se adotar uma sistemática de escolha baseada no desempenho na validação. Contudo, recomenda-se a realização de experimentos adicionais em estudos futuros para verificar se essa limitação decorre, sobretudo, das características específicas dos dados utilizados neste estudo, e não apenas da configuração dos parâmetros.

Figura 15 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 10 dias úteis à frente do modelo selecionado multivariado com 2CR vs RW no período de teste



**Nota:** Menores valores destacados em negrito.

**Fonte:** O autor, 2025.

A Figura 15 apresenta a taxa de câmbio Ptax de venda (BRL/USD) no período de teste, comparando as previsões 10 dias úteis à frente do modelo RW e as do modelo de DL selecionado BiLSTM (adam - LR 0.0001 - Dropout 0.1 - Batch 32) com 2CR que adota uma abordagem multivariada.

Observa-se que o respectivo modelo de DL selecionado apresenta uma trajetória que acompanha a tendência altista observada na série efetiva da taxa de câmbio Ptax do período de teste. Em contraste com o modelo de DL selecionado com 1CR, percebe-se que suas previsões não se aproximam tanto das previsões geradas pelo modelo RW. Esse comportamento pode indicar uma maior capacidade de retenção de tendências de curto prazo, o que reforça a importância de se testar empiricamente diferentes configurações de modelos de DL.

#### 5.2.2.4 Testes estatísticos entre o modelo de DL multivariado selecionado com melhor desempenho no período de teste vs. modelo RW nas previsões 10 dias úteis à frente

Tabela 18 - Testes estatísticos para comparação do modelo multivariado de DL selecionado com 2CR e do modelo RW nas previsões 10 dias úteis à frente

Testes	Período de Validação	Período de Teste
Diebold-Mariano (DM)	RW vs BiLSTM: Estatística = 0.1073, p-valor = 0.9146	RW vs BiLSTM: Estatística = 1.8696, p-valor = 0.0615*
Teste de Diebold-Mariano-Modificado (DMM)	RW vs BiLSTM: Estatística = 0.0948, p-valor = 0.9247	RW vs BiLSTM: Estatística = 1.6456, p-valor = 0.1037
Flutuação de Giacomini e Rossi (2010)	RW: Estatística = 13.3080, p-valor = 0.0000***	RW: Estatística = 18.8424, p-valor = 0.0000***
Racionalidade de Rossi e Sekhposyan (2016)	BiLSTM: Estatística = 3223.8987, p-valor = 0.0000***	BiLSTM: Estatística = 1722.5386, p-valor = 0.0000***
Racionalidade de Rossi e Sekhposyan (2016)	RW: Estatística = 5779.2941, p-valor = 0.0000***	RW: Estatística = 2980.7553, p-valor = 0.0000***

**Nota 1:** Refere-se ao modelo BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32).

**Nota 2:** \*\*\*, \*\* e \* indicam significância estatística ao nível de 1%, 5% e 10%, respectivamente.

**Fonte:** O autor, 2025.

A Tabela 18 apresenta os resultados dos testes estatísticos comparando as previsões do modelo RW com as do modelo selecionado BiLSTM (adamw - LR 0.0001 - Dropout 0.1 - Batch 16) com 2CR, que obteve as menores métricas de erro no período de teste entre as duas arquiteturas. Os resultados dos testes para o modelo selecionado com 1CR, com os menores erros de previsão no período de teste nessa arquitetura, podem ser consultados na Tabela 30, disponível no Apêndice A.

Os resultados dos testes DM e DMM apresentados na Tabela 18 indicam que as previsões dos modelos apresentaram acurácias estatisticamente equivalentes. Considerando o nível de significância de 5%, as  $H_0$  de equivalência preditiva não foram rejeitadas em ambos os testes e nos dois períodos analisados. No período de teste, os p-valores obtidos nos testes DM e DMM ficaram próximos aos níveis de significância de 5% e 10%, respectivamente; ainda assim, não foram suficientes para evidenciar diferenças estatisticamente significativas de acurácia entre os modelos. Quando comparado com os resultados obtidos no modelo de DL univariado selecionado, observa-se que, nesse caso, os p-valores ficaram mais próximos da rejeição da  $H_0$ , o que dá indícios de que a inclusão de variáveis explicativas pode ter contribuído para melhorar o desempenho preditivo.

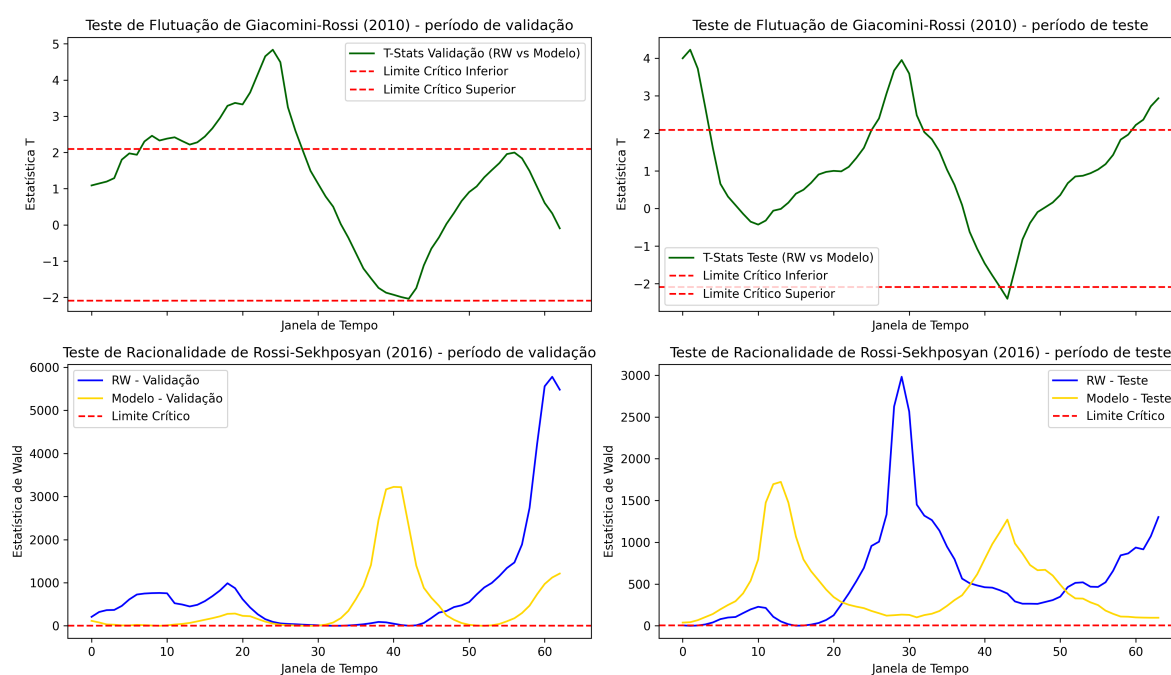
Por outro lado, o teste de flutuação GR aponta, ao nível de significância de 5%, para a rejeição da  $H_0$  de igualdade de acurácia das previsões dos dois modelos comparados. Tanto no período de validação quanto no de teste, a ET positiva indica a superioridade da acurácia das previsões do modelo de DL na janela de tempo de maior diferença de acurácia entre os modelos comparados. A Figura 16 ilustra o comportamento das ET ao longo dos períodos de validação e teste.

O teste de racionalidade de RS indica, ao nível de significância de 5%, a rejeição de  $H_0$  de imprevisibilidade dos erros de previsão no modelo RW e no modelo de DL selecionado, tanto no período de validação e quanto no de teste. O que aponta a presença de eventos de irracionalidade, nos quais os erros preditivos não foram aleatórios e apresentaram dependência temporal. Neste caso, as ET das previsões do RW foram maiores que as do modelo de DL selecionado nos dois períodos analisados, o que descreve que as suas previsões apresentaram vieses

ainda maiores em pelo menos uma das janelas de tempo dos respectivos períodos analisados.

Esses resultados sugerem que a inclusão de variáveis explicativas na rede neural tem potencial para melhoria da capacidade de preditiva, em pelo menos algumas janelas de tempo, tornando-a mais competitiva em relação ao desempenho das previsões do modelo RW. Apesar disso, a estabilidade e a racionalidade das previsões ainda são um desafio.

Figura 16 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação



**Nota:** O "Modelo" descrito na legenda refere-se ao modelo selecionado multivariado, com 2CR, nas previsões 10 dias úteis à frente. As janelas móveis possuem 20 dias úteis nos dois testes.

**Fonte:** O autor, 2025.

A Figura 16 descreve a evolução das ETs do teste de flutuação de GR nas diferentes janelas de tempo e mostra que, no início do período de validação, o limite crítico superior foi ultrapassado, indicando nesses casos uma superioridade da acurácia das previsões do modelo de DL em relação às do modelo RW. Posteriormente, a ET permaneceu dentro dos limites críticos, sugerindo igualdade de desempenho preditivo. No período de teste, durante a maior parte do tempo, a ET permaneceu dentro dos limites críticos. Ocorreram três intervalos em que se apontam a superioridade da acurácia das previsões do modelo de DL e um curto intervalo em que as previsões do modelo RW demonstraram um melhor desempenho.

A evolução da ET no teste de racionalidade de RS indica que, no período de validação, o modelo de DL selecionado apresentou dois momentos mais visíveis de irracionalidade nos erros de previsão, refletindo falta de eficiência e viés, com a ET ultrapassando significativamente o limite crítico, especialmente no início da segunda metade das janelas de tempo e no final do período. Por sua vez, o modelo RW demonstrou erros de previsão ainda mais irracionais que os do modelo de DL no final do período de validação. No período de teste, as ETs de ambos

os modelos exibiram dois picos de irracionalidade nos erros de previsão: as do modelo RW na região central e no final do gráfico, e as do modelo de DL no primeiro e terceiro quartil.

Os resultados obtidos com os modelos multivariados de DL para a previsão de 10 dias úteis à frente sugerem que, nas configurações aplicadas, sua eficácia em horizontes de curto prazo ainda é desafiadora, especialmente quando avaliada com base em critérios estatísticos convencionais. Embora o modelo de DL selecionado com melhor desempenho tenha apresentado menores métricas de erro, não foi possível rejeitar pelo teste DMM a hipótese de equivalência de acurácia preditiva em relação ao modelo RW, mesmo ao nível de significância de 10%, ainda que o p-valor associado à ET tenha sido muito próximo a esse nível. Porém, considerando o cenário de alta volatilidade e o fato de que o modelo não ter sido retreinado a cada nova previsão, há indícios de que suas previsões estiveram próximas de superar a acurácia das previsões do modelo RW. Além disso, aspectos relacionados à incerteza e a choques externos podem ter influenciado significativamente a dinâmica da taxa de câmbio no curto prazo.

### 5.3 Resultados das previsões 20 dias úteis à frente

#### 5.3.1 Modelos univariados de DL vs. modelo RW nas previsões 20 dias úteis à frente

##### 5.3.1.1 Modelos univariados de DL com 1CR vs. modelo RW nas previsões 20 dias úteis à frente:

Tabela 19 - Métricas de erro das previsões dos modelos univariados de DL com 1CR e do modelo RW 20 dias úteis à frente

Modelo (uma camada recorrente)	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
RW	0.1929	0.1612	3.0359	<b>0.2365</b>	<b>0.2137</b>	<b>3.6417</b>
LSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.1539	<b>0.1091</b>	<b>2.0344</b>	0.5062	0.4748	8.0803
LSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1947	0.1610	3.0279	0.4714	0.4352	7.3940
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1867	0.1481	2.7771	0.4527	0.4194	7.1276
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1923	0.1545	2.8999	0.4448	0.4080	6.9265
GRU (adam - LR0.0001 - Dropout0.1 - Batch16)	<b>0.1456</b>	0.1103	2.0693	0.3778	0.3344	5.6557
GRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1681	0.1197	2.2307	0.3760	0.3359	5.6860
GRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1497	0.1095	2.0480	0.3702	0.3313	5.6100
GRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1583	0.1145	2.1380	0.3743	0.3313	5.6030
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.1767	0.1293	2.4130	0.4045	0.3622	6.1340
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1981	0.1659	3.1206	0.4712	0.4374	7.4366
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1597	0.1139	2.1244	0.3558	0.3197	5.4163
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1531	0.1119	2.0902	0.4162	0.3562	6.0071
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.1516	0.1123	2.1006	0.3553	0.3138	5.3070
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1578	0.1158	2.1624	0.3768	0.3341	5.6521
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1524	0.1134	2.1207	<b>0.3509</b>	<b>0.3101</b>	<b>5.2459</b>
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1556	0.1144	2.1366	0.3980	0.3579	6.0632

**Nota:** Os dois menores valores de cada coluna estão destacados em negrito.

**Fonte:** O autor, 2025.

A Tabela 19 apresenta as métricas de erro das previsões para 20 dias úteis à frente realizadas pelo modelo RW e pelos modelos de DL univariados com 1CR. No período de validação,

o modelo LSTM (adam - LR0.0001 - Dropout0.1 - Batch16) apresentou os menores valores de MAE e MAPE, o que sugere que suas previsões estiveram, em média, mais próximas dos valores reais em comparação aos demais modelos, além um desempenho superior na minimização dos erros relativos. Por sua vez, o modelo GRU (adam - LR0.0001 - Dropout0.1 - Batch16) registrou o menor RMSE no mesmo período, destacando-se por gerar previsões mais precisas em termos absolutos, com menor incidência de discrepâncias extremas. Dessa forma, ambos os modelos de DL foram pré-selecionados entre os modelos de 1CR. No período de teste, entretanto, as previsões do modelo RW exibiram as menores métricas de erro entre todos os modelos. Adicionalmente, não foram observadas vantagens claras entre as previsões de nenhum dos modelos de DL analisados.

### 5.3.1.2 Modelos univariados de DL com 2CR vs. modelo RW nas previsões 20 dias úteis à frente:

Tabela 20 - Métricas de erro das previsões dos modelos univariados de DL com 2CR e do modelo RW 20 dias úteis à frente

Modelo (duas camadas recorrentes)	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
RW	0.1929	0.1612	3.0359	<b>0.2365</b>	<b>0.2137</b>	<b>3.6417</b>
LSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.2003	0.1660	3.1207	0.4519	0.4160	7.0642
LSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1975	0.1620	3.0427	0.4614	0.4240	7.1995
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1717	0.1323	2.4779	0.4905	0.4610	7.8451
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.2386	0.2073	3.9079	0.5830	0.5511	9.3864
GRU (adam - LR0.0001 - Dropout0.1 - Batch16)	<b>0.1662</b>	<b>0.1190</b>	<b>2.2188</b>	0.4214	0.3818	6.4716
GRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1987	0.1469	2.7410	0.4385	0.4026	6.8339
GRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	<b>0.1614</b>	<b>0.1168</b>	<b>2.1789</b>	<b>0.4035</b>	<b>0.3607</b>	<b>6.1051</b>
GRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1971	0.1458	2.7211	0.4249	0.3882	6.5865
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.2260	0.1944	3.6648	0.5186	0.4774	8.1037
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1996	0.1610	3.0221	0.4494	0.4156	7.0603
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1915	0.1530	2.8697	0.4613	0.4124	6.9818
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.2445	0.2123	4.0017	0.5268	0.4897	8.3240
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.1683	0.1224	2.2837	0.4541	0.4184	7.1041
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1708	0.1263	2.3593	0.4276	0.3889	6.5935
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1703	0.1269	2.3714	0.4251	0.3804	6.4384
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.2148	0.1936	3.6611	0.4730	0.4319	7.3244

**Nota:** Os dois menores valores de cada coluna estão destacados em negrito.

**Fonte:** O autor, 2025.

A Tabela 20 apresenta as métricas de erro das previsões para 20 dias úteis à frente realizadas pelo modelo RW e pelos modelos de DL univariados com 2CR.

No período de validação, as previsões do modelo GRU (adamw - LR0.0001 - Dropout0.1 - Batch16) apresentaram as métricas de erro entre todos os modelos. No entanto, no período de teste, foram as previsões do modelo RW que exibiram esse resultado. Nesse período, as métricas de erro do modelo GRU foram quase o dobro das observadas para as previsões do modelo RW. Acrescenta-se que, novamente, não foram identificadas vantagens claras entre as previsões dos modelos de DL analisados.

5.3.1.3 Modelos de DL univariados selecionados vs. modelo RW nas previsões 20 dias úteis à frente

Tabela 21 - Métricas de erro das previsões 20 dias úteis à frente dos modelos de DL univariados selecionados em cada arquitetura vs. modelo RW

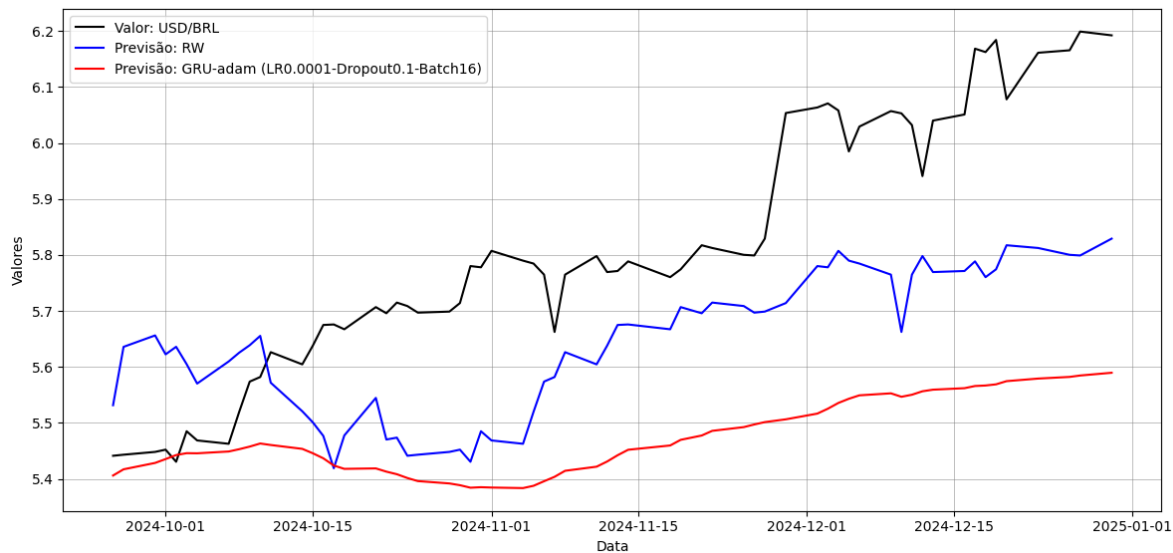
Arquitetura	Modelos e Parâmetros	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
-	RW	0.1929	0.1612	3.0359	<b>0.2365</b>	<b>0.2137</b>	<b>3.6417</b>
1CR (pelo critério RMSE)	GRU (adam - LR0.0001 - Dropout0.1 - Batch16)	<b>0.1456</b>	<b>0.1103</b>	2.0693	0.3778	0.3344	5.6557
1CR (pelo critério MAPE)	LSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.1539	0.1091	<b>2.0344</b>	0.5062	0.4748	8.0803
2CR	GRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1614	0.1168	2.1789	0.4035	0.3607	6.1051

Nota: Menores valores destacados em negrito.

Fonte: O autor, 2025.

A Tabela 21 apresenta as métricas de erro das previsões 20 dias úteis à frente do modelo RW e dos modelos de DL univariados selecionados nas duas arquiteturas (1CR e 2CR) pelo critério de menores métricas de erro no período de validação. Entre os modelos de DL selecionados, o modelo GRU (adam - LR0.0001 -Dropout0.1 - Batch16) com 1CR apresentou as menores métricas de erro durante o período de teste, porém com valores superiores aos do modelo RW. Assim, serão analisados os resultados dos testes estatísticos para verificar se a hipótese de equivalência de acurácia preditiva pode ser rejeitada. No entanto, antes disso, serão apresentados os gráficos das previsões dos dois modelos selecionados, com melhor desempenho no período de teste, e do modelo RW.

Figura 17 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 20 dias úteis à frente do modelo selecionado univariado com 1CR vs RW no período de teste



Nota: Menores valores destacados em negrito.

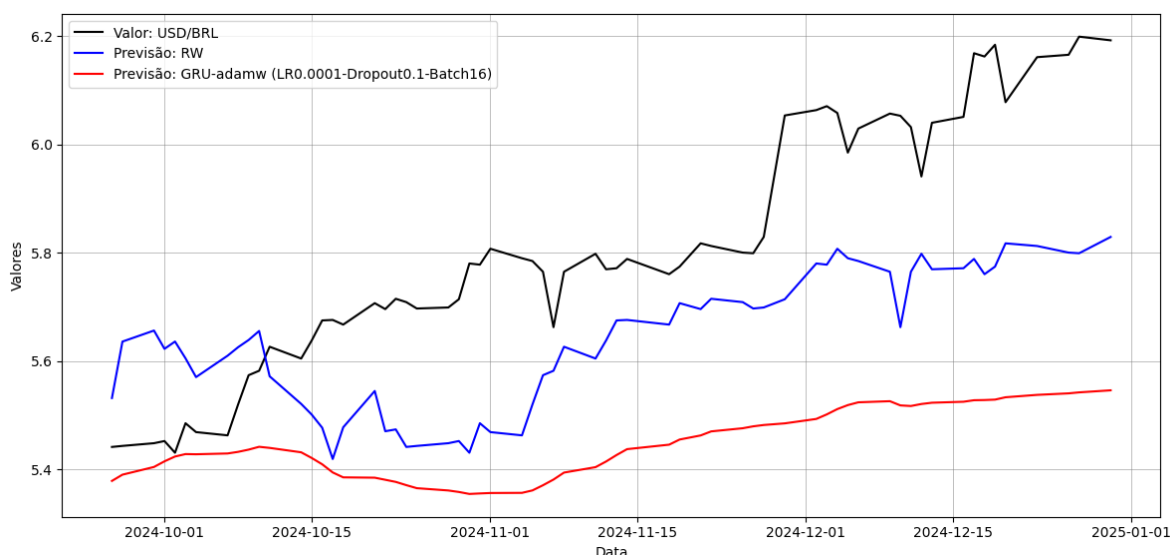
Fonte: O autor, 2025.

A Figura 17 apresenta a taxa de câmbio Ptax de venda (BRL/USD) no período de teste, comparando as previsões 20 dias úteis à frente do modelo RW e as do modelo de DL selecionado GRU (adam - LR0.0001 - Dropout0.1 - Batch16) com 1CR que adota uma abordagem univariada.

Observa-se um comportamento marcadamente distinto entre os modelos. O RW captura razoavelmente as oscilações da taxa de câmbio, mantendo uma resposta às mudanças abruptas no câmbio, apesar da defasagem inerente à sua própria característica. Em contrapartida, o referido modelo GRU apresenta uma suavização bastante pronunciada, resultando em previsões que subestimam significativamente a série.

A suavização excessiva do modelo GRU pode ser atribuída a uma combinação de fatores, incluindo problemas de convergência para um mínimo local no processo de otimização dos pesos da rede neural. Além disso, como o modelo de DL é univariado, ele só considera os preços passados para fazer as suas previsões, o que pode limitar sua capacidade preditiva nesse contexto de elevada volatilidade.

Figura 18 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 20 dias úteis à frente do modelo selecionado univariado com 2CR vs RW no período de teste



**Nota:** Menores valores destacados em negrito.

**Fonte:** O autor, 2025.

A Figura 18 apresenta a taxa de câmbio Ptax de venda (BRL/USD) no período de teste, comparando as previsões 20 dias úteis à frente do modelo RW e as do modelo de DL selecionado GRU (adamw - LR0.0001 - Dropout0.1 - Batch16) com 2CR que adota uma abordagem univariada.

Verifica-se um comportamento muito similar ao observado nas previsões do modelo univariado selecionado com 1CR descrito na Figura 17. O modelo GRU selecionado exibe uma trajetória suavizada e com viés de subestimação ao longo do período de teste. Esse comportamento sugere que, apesar da presença de 2CR e das diferentes configurações testadas, a rede neural ainda apresentou limitações na captação de oscilações de curto prazo, especialmente em períodos de maior volatilidade. Além disso, esse resultado pode indicar um ajuste inadequado no processo de otimização dos pesos da rede neural durante o treinamento, somado ao impacto de fatores externos ocorridos no período sobre a estrutura dos dados e à ausência de variáveis



explicativas no contexto univariado.

#### 5.3.1.4 Testes estatísticos entre o modelo de DL univariado selecionado com melhor desempenho no período de teste vs. modelo RW nas previsões 20 dias úteis à frente

Tabela 22 - Testes estatísticos para comparação do modelo univariado de DL selecionado com 1CR e do modelo RW nas previsões 20 dias úteis à frente

Testes	Período de Validação	Período de Teste
Diebold-Mariano (DM)	RW vs GRU: Estatística = 3.8531, p-valor = 0.0001***	RW vs GRU: Estatística = -4.5324, p-valor = 0.0000***
Teste de Diebold-Mariano-Modificado (DMM)	RW vs GRU: Estatística = 2.6411, p-valor = 0.0105**	RW vs GRU: Estatística = -3.1293, p-valor = 0.0027***
Flutuação de Giacomini e Rossi (2010)	RW vs GRU: Estatística = 20.9228, p-valor = 0.0000***	RW vs GRU: Estatística = -30.1642, p-valor = 0.0000***
Racionalidade de Rossi e Sekhposyan (2016)	GRU: Estatística = 1797.0981, p-valor = 0.0000***	GRU: Estatística = 29918.5294, p-valor = 0.0000***
Racionalidade de Rossi e Sekhposyan (2016)	RW: Estatística = 6708.8322, p-valor = 0.0000***	RW: Estatística = 7400.7754, p-valor = 0.0000***

**Nota 1:** Refere-se ao modelo GRU (adam - LR0.0001 - Dropout0.1 - Batch16).

**Nota 2:** \*\*\*, \*\* e \* indicam significância estatística ao nível de 1%, 5% e 10%, respectivamente.

**Fonte:** O autor, 2025.

A Tabela 22 apresenta os resultados dos testes estatísticos comparando as previsões do modelo RW com as do modelo de DL selecionado GRU (adam - LR0.0001 - Dropout0.1 - Batch16) com 1CR, que obteve as menores métricas de erro no período de teste entre as duas arquiteturas. Os resultados dos testes estatísticos referentes ao modelo de DL selecionado com 2CR podem ser consultados na Tabela 31, disponível no Apêndice A.

Os resultados dos testes DM e DMM apresentados na tabela acima indicam que as previsões dos modelos não tiveram acurácias estatisticamente equivalentes. Com base no nível de significância de 5%, as  $H_0$  foram rejeitadas em ambos os testes, tanto no período de validação quanto no de teste. No período de validação, os p-valores dos testes DM e DMM, associados às ETs positivas, apontam a superioridade da acurácia das previsões do modelo de DL. Por outro lado, no período de teste, as ETs negativas indicam que as previsões do modelo RW foram superiores em termos de acurácia.

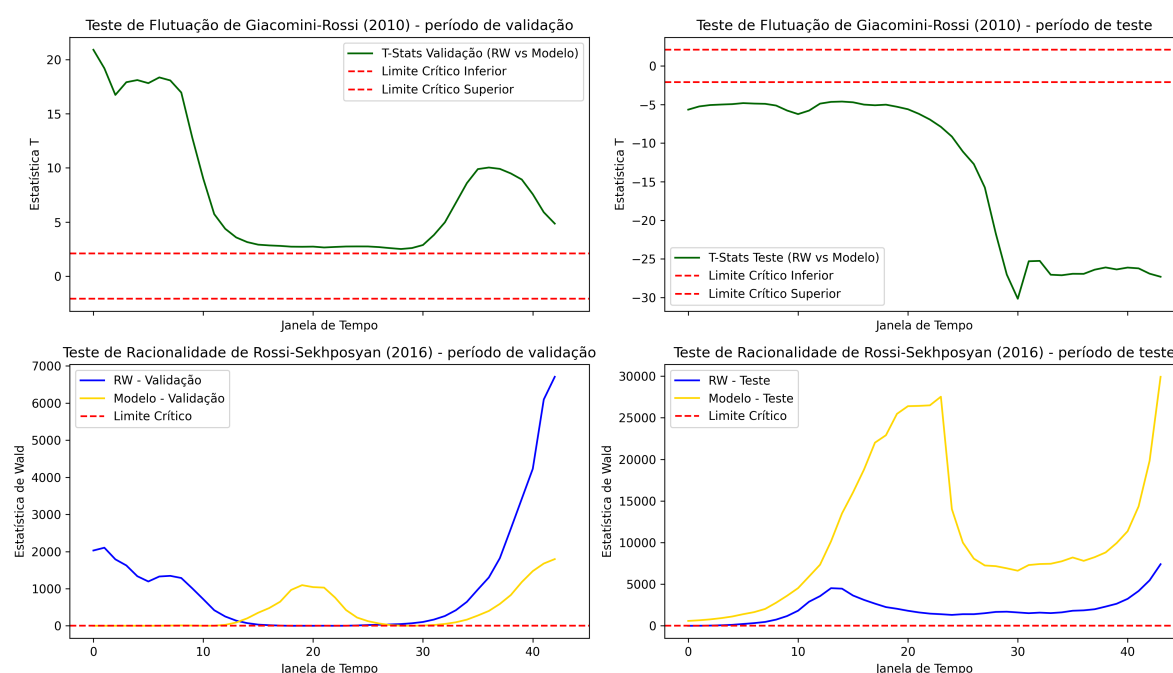
No mesmo sentido, o teste de flutuação GR também assinala a rejeição da  $H_0$  de igualdade de acurácia das previsões dos dois modelos comparados. Assim, no período de validação, a ET positiva evidencia a superioridade da acurácia das previsões do modelo de DL na janela de tempo de maior diferença de acurácia entre os modelos comparados. Já no período de teste, a ET negativa aponta a favor da acurácia das previsões do modelo RW também na respectiva janela de tempo de maior diferença preditiva. A Figura 19, apresentada a seguir, ilustra o comportamento da diferença de acurácia nas janelas de tempo ao longo dos períodos de validação e teste.

O teste de racionalidade de RS aponta a rejeição da  $H_0$  de imprevisibilidade dos erros de previsão no modelo RW e no modelo de DL selecionado, tanto no período de validação e quanto no de teste. Neste caso, as ET das previsões do RW foram maiores que as do modelo de DL selecionado no período de validação e menores no período de teste. Observa-se que a ET do modelo de DL no período de teste foi notadamente alta, o que reflete o viés de subestimação

já observado no gráfico das previsões.

Os resultados indicam que, embora tenham demonstrado menores métricas de erro no período de validação, as previsões dos modelos de DL univariados selecionados entre as diferentes especificações adotadas não conseguiram superar as previsões 20 dias à frente do modelo RW no período de teste.

Figura 19 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhpsoyan (2016) ao longo dos períodos de teste e de validação



**Nota:** O "Modelo" descrito na legenda refere-se ao modelo selecionado univariado com 1CR, nas previsões 20 dias úteis à frente. As janelas móveis possuem 20 dias úteis nos dois testes.

**Fonte:** O autor, 2025.

Conforme evidenciado na Figura 19, a evolução das ETs no teste de flutuação de GR revela que, no período de validação, o limite crítico superior foi ultrapassado de forma mais expressiva na parte inicial e na parte final das janelas de tempo, evidenciando-se nessas janelas uma superioridade da acurácia das previsões do modelo de DL. Em contraste, no período de teste, as ETs ultrapassaram o limite crítico inferior, indicando uma superioridade da acurácia das previsões do modelo RW, com um crescimento expressivo (em módulo) na segunda metade das janelas de tempo.

A análise das ETs nas janelas de tempo do teste de racionalidade de RS revela que, no período de validação, as ETs referentes às previsões do modelo de DL selecionado ultrapassaram significativamente o limite crítico, sobretudo nas janelas de tempo do centro e do final desse período. Em contraste, as previsões do modelo RW apresentaram erros de previsão ainda mais irracionais do que as do modelo de DL no final das janelas de validação. Já no período de teste, as ETs do modelo de DL foram consideravelmente superiores às do modelo RW, principalmente ao final da primeira metade e ao final das janelas de tempo.

Os resultados dos modelos de DL univariados nas diferentes configurações testadas sugerem que as respectivas acurácias das respectivas previsões decaem com o aumento do horizonte de previsão para 20 dias úteis à frente, especialmente quando comparadas às do modelo RW (*benchmark*).

Assim, em um cenário caracterizado por alta volatilidade e pela utilização apenas dos valores passados da própria taxa de câmbio Ptax de venda, os resultados das previsões 20 dias úteis à frente indicam que o desempenho do modelo de DL univariado esteve distante de superar o modelo RW. Nesse sentido, será verificado se a estratégia de inclusão de variáveis explicativas pode representar um caminho promissor, conforme os indícios já observados na previsão multivariada com horizonte de 10 dias à frente.

### 5.3.2 Modelos multivariados de DL vs. modelo RW nas previsões 20 dias úteis à frente

#### 5.3.2.1 Modelos multivariados de DL com 1CR vs. modelo RW nas previsões 20 dias úteis à frente:

Tabela 23 - Métricas de erro das previsões dos modelos multivariados de DL com 1CR e do modelo RW 20 dias úteis à frente

Modelo	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
RW	0.1929	0.1612	3.0359	<b>0.2365</b>	<b>0.2137</b>	<b>3.6417</b>
LSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.3881	0.3608	6.8317	0.9339	0.9041	15.4441
LSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.2835	0.2265	4.2463	0.5260	0.5021	8.5593
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.2850	0.2534	4.7826	0.7418	0.6992	11.8985
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	<b>0.1524</b>	<b>0.1119</b>	<b>2.0961</b>	<b>0.1480</b>	<b>0.1258</b>	<b>2.1861</b>
GRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.3439	0.3091	5.8382	0.8237	0.7954	13.5812
GRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.3677	0.3361	6.3575	0.8650	0.8390	14.3354
GRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.3158	0.2785	5.2523	0.7903	0.7577	12.9226
GRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.3183	0.2810	5.3002	0.7920	0.7643	13.0489
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.3096	0.2787	5.2654	0.5599	0.5278	8.9820
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	<b>0.1795</b>	<b>0.1314</b>	<b>2.4599</b>	0.4684	0.4150	7.0172
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.2479	0.1993	3.7357	0.5784	0.5371	9.1248
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.2751	0.2192	4.1075	0.7853	0.7552	12.8888
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.2377	0.1769	3.3015	0.6719	0.6457	11.0202
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.1958	0.1378	2.5704	0.5374	0.5129	8.7486
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.2228	0.1580	2.9418	0.6415	0.6103	10.3998
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.2969	0.2383	4.4670	0.8375	0.8145	13.9252

**Nota:** Os dois menores valores de cada coluna estão destacados em negrito.

**Fonte:** O autor, 2025.

Tabela 23 apresenta as métricas de erro das previsões para 20 dias úteis à frente realizadas pelo modelo RW e pelos modelos de DL multivariados com 1CR. Em ambos os períodos, de validação e de teste, o modelo LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32) destacou-se por apresentar as menores métricas de erro, demonstrando consistência e capacidade de generalização em suas previsões.

Os demais modelos de DL apresentaram métricas de erro superiores, inclusive quando comparados ao caso univariado com o mesmo horizonte de previsão, assim como já observado

nas previsões 10 dias à frente dos modelos de DL multivariado. Esse comportamento pode estar relacionado à maior complexidade das dinâmicas introduzidas pelas variáveis explicativas e à eventuais mudanças nas dinâmicas das respectivas variáveis em função do choque externo observado no respectivo período.

### 5.3.2.2 Modelos multivariados de DL com 2CR vs. modelo RW nas previsões 20 dias úteis à frente:

Tabela 24 - Métricas de erro das previsões dos modelos multivariados de DL com 2CR e do modelo RW 20 dias úteis à frente

Modelo	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
RW	0.1929	0.1612	3.0359	<b>0.2365</b>	<b>0.2137</b>	<b>3.6417</b>
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.2912	0.2474	4.6541	0.6904	0.6618	11.2876
BiGRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.2423	0.1944	3.6440	0.6682	0.6406	10.9273
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.2642	0.2051	3.8382	0.7870	0.7687	13.1548
BiGRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.3157	0.2642	4.9655	0.8741	0.8522	14.5746
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.1482	<b>0.1147</b>	<b>2.1611</b>	0.6612	0.6362	10.8603
BiLSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	0.2137	0.1637	3.0615	0.3455	0.3091	5.2321
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	<b>0.1427</b>	<b>0.1131</b>	<b>2.1374</b>	<b>0.1937</b>	<b>0.1515</b>	<b>2.5538</b>
BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.3993	0.3582	6.7638	0.9848	0.9626	16.4708
GRU (adam - LR0.0001 - Dropout0.1 - Batch16)	0.3282	0.2948	5.5678	0.6676	0.6345	10.8096
GRU (adam - LR0.0001 - Dropout0.1 - Batch32)	0.3082	0.2736	5.1629	0.6629	0.6298	10.7289
GRU (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.3129	0.2775	5.2358	0.6745	0.6433	10.9641
GRU (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.2830	0.2449	4.6137	0.6422	0.6065	10.3247
LSTM (adam - LR0.0001 - Dropout0.1 - Batch16)	0.3247	0.3006	5.6932	0.7031	0.6662	11.3487
LSTM (adam - LR0.0001 - Dropout0.1 - Batch32)	<b>0.1460</b>	0.1164	2.1992	0.3294	0.2972	5.1243
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	0.1719	0.1203	2.2414	0.2923	0.2578	4.3668
LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.2974	0.2431	4.5618	0.5516	0.5351	9.1476

**Nota:** Os dois menores valores de cada coluna estão destacados em negrito.

**Fonte:** O autor, 2025.

Tabela 24 apresenta as métricas de erro das previsões 20 dias úteis à frente realizadas pelo modelo RW e pelos modelos de DL multivariados com 2CR. Tanto no período de validação quanto no de teste, o modelo BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16) destacou-se por apresentar as menores métricas de erro em suas previsões, demonstrando consistência e capacidade de generalização em suas previsões. Novamente, os demais modelos de DL apresentaram métricas de erro superiores, inclusive quando comparados ao caso univariado com o mesmo horizonte de previsão. Da mesma forma que no caso anterior com 1CR, é possível alguns dos demais modelos de DL tenham convergido para um mínimo local subótimo no processo de treinamento. Por outro lado, fatores externos podem ter alterado as dinâmicas de curto prazo entre as variáveis explicativas e a taxa de câmbio, o que poderia justificar o aumento das métricas de erro de previsão especialmente no período de teste.

No entanto, cabe salientar que, dentro da sistemática adotada, tanto os modelos DL selecionados com 1CR quanto os com 2CR apresentaram consistência entre o desempenho apresentado no período de validação e no período de teste.

5.3.2.3 Modelos de DL multivariados selecionados vs. modelo RW nas previsões 20 dias úteis à frente

Tabela 25 - Métricas de erro das previsões 20 dias úteis à frente dos modelos de DL multivariados selecionados em cada arquitetura vs. modelo RW

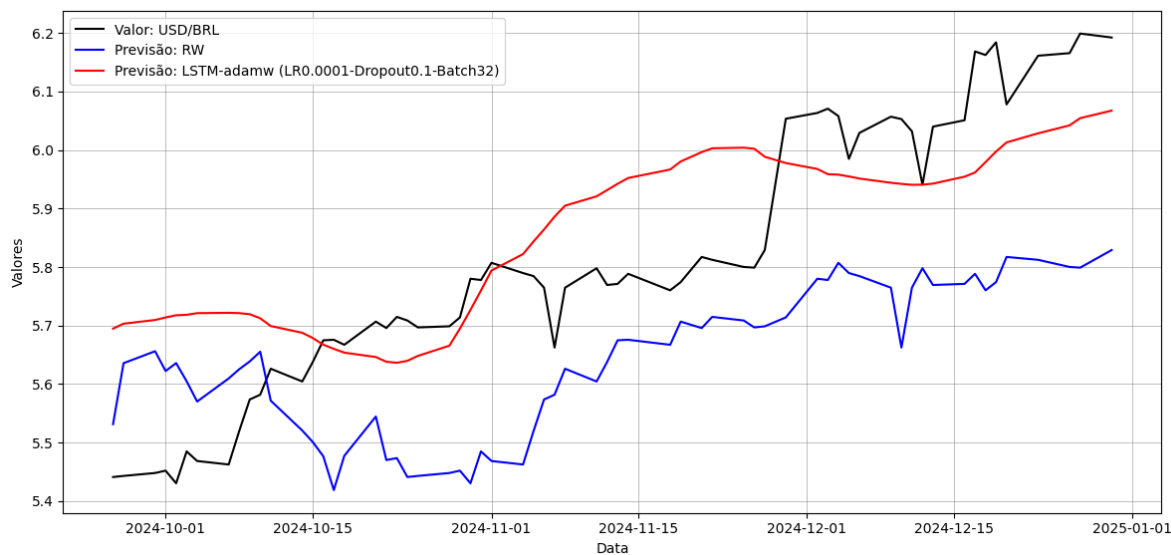
Arquitetura	Modelos e Parâmetros	RMSE (Val.)	MAE (Val.)	MAPE (Val.)	RMSE (Teste)	MAE (Teste)	MAPE (Teste)
-	RW	0.1929	0.1612	3.0359	0.2365	0.2137	3.6417
1CR	LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32)	0.1524	<b>0.1119</b>	<b>2.0961</b>	<b>0.1480</b>	<b>0.1258</b>	<b>2.1861</b>
2CR	BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16)	<b>0.1427</b>	0.1131	2.1374	0.1937	0.1515	2.5538

**Nota:** Menores valores destacados em negrito.

**Fonte:** O autor, 2025.

A Tabela 25 apresenta as métricas de erro das previsões 20 dias úteis à frente realizadas pelo modelo RW e pelos modelos de DL multivariados selecionados nas arquiteturas com 1CR e 2CR, com base no critério de menores métricas de erro no período de validação. Entre os modelos de DL analisados, o modelo LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32) com 1CR apresentou as menores métricas de erro no período de teste. Como os seus resultados também foram inferiores aos do modelo RW, serão realizados testes estatísticos para avaliar se a acurácia das respectivas previsões pode ser considerada superior a das previsões do modelo RW. Porém, antes disso, serão realizadas as análises gráficas das previsões dos modelos de DL selecionados nas duas arquiteturas e das previsões do modelo RW.

Figura 20 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 20 dias úteis à frente do modelo selecionado multivariado com 1CR vs RW no período de teste



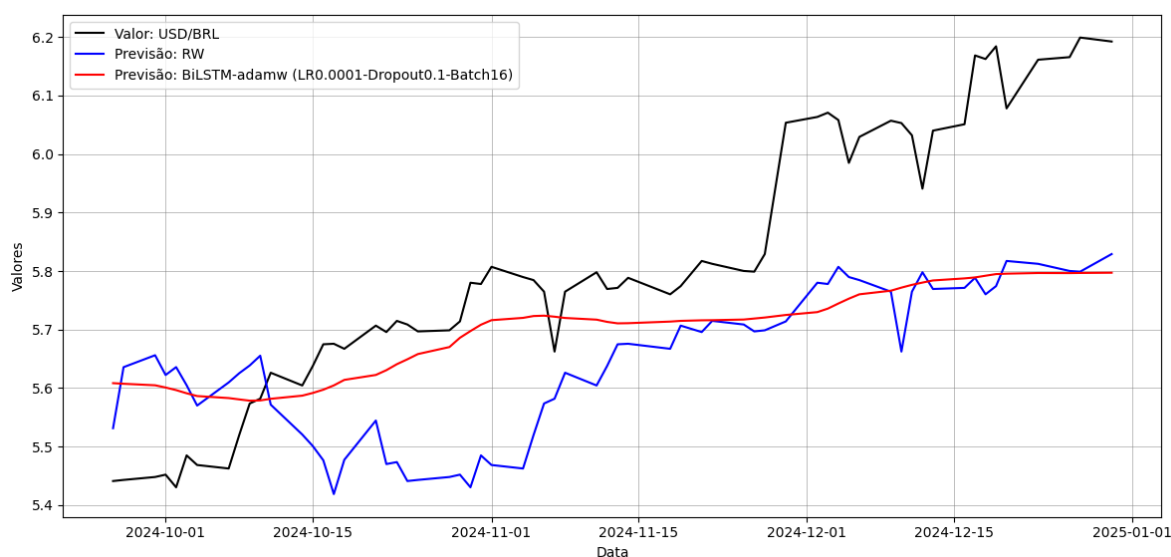
**Nota:** Menores valores destacados em negrito.

**Fonte:** O autor, 2025.

A Figura 20 apresenta a taxa de câmbio Ptax de venda (BRL/USD) no período de teste, comparando as previsões 20 dias úteis à frente do modelo RW e as do modelo de DL selecionado LSTM-adamw (LR0.0001-Dropout0.1-Batch32) com 1CR que adota uma abordagem multivariada.

Observa-se que, em comparação ao modelo com 2CR descrito na sequência, a previsão gerada pelo respectivo LSTM com 1CR aparenta um comportamento mais próximo da tendência de apreciação observada na série da taxa de câmbio Ptax de venda, embora exiba um viés alternado de superestimação em determinados períodos e subestimação em outros. Dessa forma, mesmo no caso multivariado, a capacidade dos modelos de DL adotados de captarem as oscilações de curtíssimo e de curto prazo permanece desafiadora.

Figura 21 - Gráfico das previsões da taxa de câmbio Ptax de venda (BRL/USD) 20 dias úteis à frente do modelo selecionado multivariado com 2CR vs RW no período de teste



**Nota:** Menores valores destacados em negrito.

**Fonte:** O autor, 2025.

A Figura 21 apresenta a taxa de câmbio Ptax de venda (BRL/USD) no período de teste, comparando as previsões 20 dias úteis à frente do modelo RW e as do modelo de DL selecionado BiLSTM (adamw - LR0.0001 - Dropout0.1 - Batch16) com 2CR que adota uma abordagem multivariada.

Observa-se que o modelo BiLSTM apresenta uma suavização das previsões com maior proximidade da taxa de câmbio Ptax de venda efetiva na primeira metade do gráfico. Por outro lado, o modelo RW apresenta maior variabilidade, embora com maior erro em relação à trajetória efetiva, especialmente na primeira metade do período de teste. A suavização do BiLSTM levou a previsões similares às do modelo RW no final do período de teste, momento em que o câmbio apresentou uma apreciação mais acelerada.

Neste caso, o modelo de DL multivariado parece priorizar padrões de prazo mais longo em detrimento de oscilações de curtíssimo prazo.

### 5.3.2.4 Testes estatísticos entre o modelo de DL multivariado selecionado com melhor desempenho no período de teste vs. modelo RW nas previsões 20 dias úteis à frente

Tabela 26 - Testes estatísticos para comparação do modelo multivariado de DL selecionado com 1CR e do modelo RW nas previsões 20 dias úteis à frente

Testes	Período de Validação	Período de Teste
Diebold-Mariano (DM)	RW vs LSTM: Estatística = 2.9638, p-valor = 0.0030***	RW vs LSTM: Estatística = 2.6422, p-valor = 0.0082***
Teste de Diebold-Mariano-Modificado (DMM)	RW vs LSTM: Estatística = 2.0315, p-valor = 0.0466**	RW vs LSTM: Estatística = 1.8242, p-valor = 0.0729*
Flutuação de Giacomini e Rossi (2010)	RW vs LSTM: Estatística = 7.5445, p-valor = 0.0000***	RW vs LSTM: Estatística = 8.9324, p-valor = 0.0000***
Racionalidade de Rossi e Sekhposyan (2016)	LSTM: Estatística = 2609.9446, p-valor = 0.0000***	LSTM: Estatística = 2025.6315, p-valor = 0.0000***
Racionalidade de Rossi e Sekhposyan (2016)	RW: Estatística = 6708.8322, p-valor = 0.0000***	RW: Estatística = 7400.7754, p-valor = 0.0000***

**Nota 1:** Refere-se ao modelo LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32).

**Nota 2:** \*\*\*, \*\* e \* indicam significância estatística ao nível de 1%, 5% e 10%, respectivamente.

**Fonte:** O autor, 2025.

Tabela 26 apresenta os resultados dos testes estatísticos comparando as previsões do modelo RW com as do modelo selecionado LSTM (adamw - LR0.0001 - Dropout0.1 - Batch32) com 1CR, que obteve as menores métricas de erro no período de teste entre as duas arquiteturas.

Os resultados dos testes DM e DMM apresentados na tabela acima indicam que as previsões dos modelos não tiveram desempenhos estatisticamente equivalentes no período de validação, uma vez que as  $H_0$  foram rejeitadas em ambos os testes, considerando o nível de significância de 5%. As ET positivas apontam que as previsões do modelo de DL apresentaram melhor desempenho nesse período. Por outro lado, no período de teste, apenas o teste DM indica a rejeição da  $H_0$  ao nível de significância de 5%. Quando realizado o ajuste pelo tamanho da amostra com o teste DMM, não se pode rejeitar a hipótese de equivalência preditiva em um nível mais exigente. Porém, ao nível de significância de 10%, o mesmo teste aponta a superioridade da acurácia das previsões 20 dias úteis à frente do modelo de DL selecionado com 1CR em relação às previsões do modelo RW. Ressalta-se que o mesmo resultado foi obtido nas previsões do modelo de DL selecionado com 2CR, conforme se observa na Tabela 32 incluída no Apêndice A.

Quando comparado com os resultados obtidos no modelo de DL univariado selecionado, observa-se que há indícios de que a inclusão de variáveis explicativas pode ter contribuído para melhorar o desempenho preditivo.

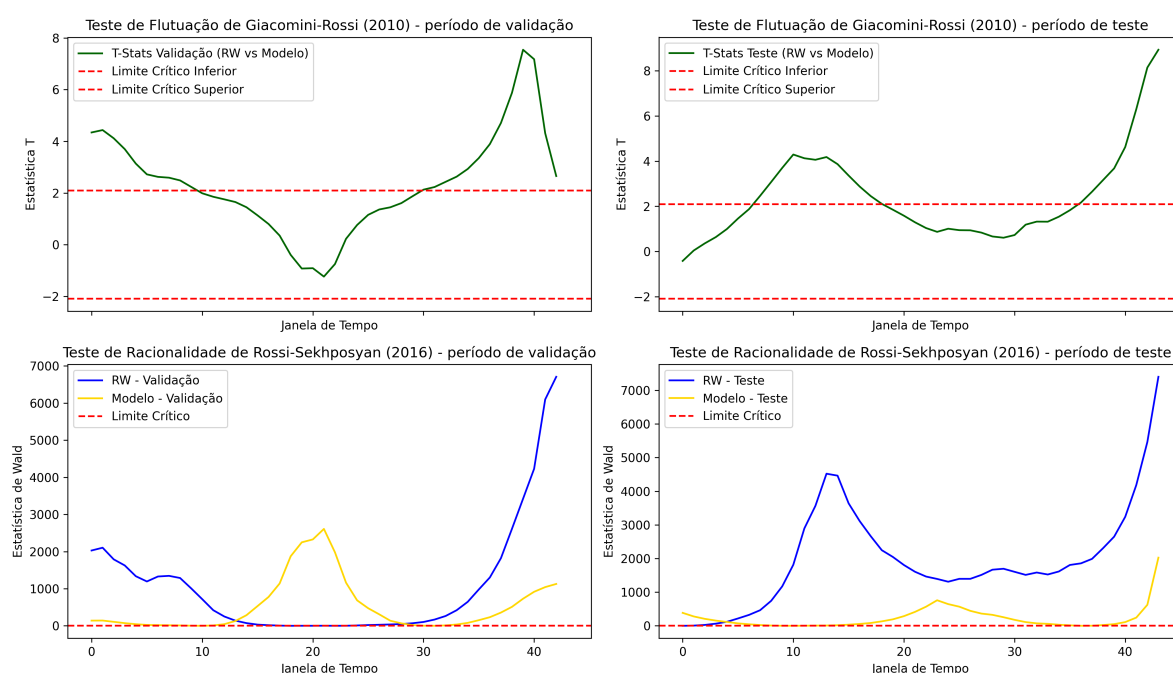
Adicionalmente, o teste de flutuação GR aponta a rejeição da  $H_0$  de igualdade de acurácia das previsões dos dois modelos comparados. Dessa forma, tanto no período de validação quanto no período de teste, a ET positiva evidencia a superioridade da acurácia das previsões do modelo de DL na janela de tempo de maior diferença de desempenho entre os modelos comparados. A Figura 22, apresentada a seguir, ilustra o comportamento da diferença de acurácia das respectivas previsões ao longo das janelas de tempo dos períodos de validação e teste.

O teste de racionalidade de RS indica a rejeição da  $H_0$  de imprevisibilidade dos erros de previsão, tanto no modelo RW quanto no modelo de DL selecionado, nos períodos de validação e de teste. Nesse caso, as ETs das previsões do modelo RW foram maiores que as do modelo de DL selecionado em ambos os períodos, o que sugere um maior viés nas previsões em pelo

menos uma janela de tempo.

Esses resultados reforçam a possibilidade de que a inclusão de variáveis econômicas e financeiras de frequência diária na rede neural tenha potencial para aprimorar a capacidade de modelagem da taxa de câmbio Ptax de venda, tornando-a mais competitiva em relação ao modelo RW, embora a estabilidade e a racionalidade das previsões ainda apresentem limitações. Destaca-se, ainda, a possibilidade de aprimorar as arquiteturas adotadas, incorporar novos modelos e mecanismos avançados, bem como de replicar o estudo em períodos posteriores, de forma a validar a robustez e a consistência dos resultados obtidos.

Figura 22 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação



**Nota:** O "Modelo" descrito na legenda refere-se ao modelo selecionado multivariado com 2CR, nas previsões 20 dias úteis à frente. As janelas móveis possuem 20 dias úteis nos dois testes.

**Fonte:** O autor, 2025.

Conforme ilustrado na Figura 22, a evolução das ET no teste de flutuação de GR mostra que, no período de validação, o limite crítico superior foi ultrapassado tanto na parte inicial como na parte final das janelas de tempo, indicando uma superioridade da acurácia das previsões do modelo de DL. No período de teste, a ET ultrapassou o limite crítico superior duas vezes, apontando nesses casos a superioridade da acurácia das previsões do modelo de DL, com crescimento expressivo no final das janelas de tempo.

A evolução das ET no teste de racionalidade de RS indica que, no período de validação, as ETs referentes ao modelo de DL selecionado ultrapassaram significativamente o limite crítico, especialmente nas janelas de tempo centrais e, com menor intensidade, nas finais. Por outro lado, no mesmo período, o modelo RW exibiu erros de previsão ainda mais irracionais que os do modelo de DL no final das janelas de tempo. No período de teste, as ETs do modelo



RW foram substancialmente superiores às do modelo de DL selecionado, especialmente no início do segundo quarto das janelas de tempo e, de forma ainda mais expressiva, ao final desse período.

É interessante notar que, de forma geral e coerente, tanto no período de validação quanto no de teste, as janelas de tempo que apresentaram erros de previsão mais irracionais segundo o teste de racionalidade de RS corresponderam à perda de acurácia das previsões dos respectivos modelos nas janelas de tempo identificadas pelo teste de GR.

Assim, os resultados obtidos revelam o potencial dos modelos de DL multivariados para superarem o desempenho das previsões do modelo RW, mesmo em horizontes de curto prazo. Esses resultados sugerem que a incorporação de variáveis econômicas e financeiras contribui para capturar melhor as dinâmicas que influenciam a taxa de câmbio, ampliando a capacidade preditiva dos modelos.

#### 5.4 Considerações finais sobre os resultados obtidos

Em relação ao comportamento da taxa de câmbio Ptax de venda, observa-se uma expressiva valorização do Dólar norte-americano frente ao Real brasileiro nos últimos meses de 2024, período correspondente ao conjunto de teste. Com base nas séries apresentadas na Figura 3, tal movimento pode ser atribuído a um conjunto articulado de choques macroeconômicos e financeiros, além de movimentos de preferência por liquidez. Nesse período, registrou-se um aumento dos *yields* dos títulos americanos de cinco e dez anos, o que elevou a atratividade relativa de ativos denominados em dólar e contribuiu para redirecionar fluxos internacionais para fora dos mercados emergentes. Simultaneamente, houve um pico na volatilidade global (medida pelo VIX) e um alargamento do prêmio de risco soberano brasileiro (CDS), ambos indicativos clássicos de aversão ao risco, favorecendo a busca por ativos mais seguros. Ademais, embora o diferencial de juros entre Brasil e Estados Unidos tenha aumentado, o mesmo esteve em patamares inferiores a períodos anteriores, o que pode ter limitado o incentivo ao *carry trade* em Real. Paralelamente, o desempenho relativo das ações brasileiras (Ibovespa em comparação ao S&P500) apresentou variações desfavoráveis ao Brasil, deteriorando os retornos esperados e intensificando a pressão cambial. A incerteza política associada ao calendário eleitoral norte-americano atuou como gatilho adicional, elevando a aversão ao risco e a volatilidade. Além disso, verificou-se uma valorização generalizada do Dólar frente a outras moedas fortes, refletida na alta do índice DXY, o que indica que o fenômeno não foi exclusivo em relação ao Real, mas parte de um reposicionamento global em direção ao Dólar como ativo de reserva e proteção.

Portanto, o padrão de valorização do Dólar resulta da interação entre o aumento dos *yields* externos, a intensificação da aversão ao risco (VIX e CDS), a compressão do diferencial de juros em relação a períodos anteriores, a fraqueza relativa dos ativos domésticos e a valorização

do Dólar frente a outras moedas, além da influência das eleições americanas. Esses canais explicativos também se refletem nos resultados empíricos das previsões dos modelos multivariados de DL no horizonte de 20 dias, os quais foram capazes de capturar essas dinâmicas conjuntas e oferecer previsões com acurácia superior à do RW. Por outro lado, no horizonte de 1 dia à frente, o ruído associado aos choques observados pode ter limitado a capacidade dos modelos de gerar ganhos significativos de acurácia. Já no horizonte de 10 dias à frente os resultados foram próximos de superar o *benchmark* apenas no caso multivariado.

De forma mais detalhada, seguindo a sistemática descrita no início deste Capítulo, os resultados da abordagem univariada apontaram que, no período de teste, as previsões dos modelos de DL selecionados, independentemente da arquitetura (1CR ou 2CR), não superaram a acurácia das previsões do modelo RW em nenhum dos horizontes de previsão analisados. Apenas no horizonte de previsão de 10 dias úteis à frente, o teste de DMM indicou uma equivalência estatística, ao nível de significância de 5%, entre as previsões do modelo RW e as do modelo de DL selecionado que obteve o melhor desempenho no período de teste. Com o aumento do horizonte de previsão para 20 dias úteis à frente, observou-se uma deterioração significativa na acurácia dos modelos de DL univariados, cujas métricas de erro de previsão superaram em ao menos 50% as do modelo RW no período de teste.

Na abordagem multivariada, verificou-se que, no horizonte de 1 dia útil à frente, as previsões do modelo RW apresentaram uma acurácia superior em relação às dos modelos de DL, segundo o teste DMM, ao nível de significância de 5%. No entanto, nas previsões 10 e 20 dias úteis à frente, os testes indicaram uma equivalência estatística entre as previsões. Considerando-se o nível de significância de 10%, o mesmo teste passa a indicar que a acurácia das previsões dos modelos de DL multivariados supera a das previsões do modelo RW no horizonte de 20 dias úteis à frente. Esse resultado é corroborado pelo teste de flutuação de GR e pelo teste de racionalidade de RS. Esses dois últimos testes indicaram, de forma coerente, um desempenho igual ou superior dos modelos de DL selecionado e um menor viés e irracionalidade nos erros de previsão em relação ao obtido pelo modelo RW, respectivamente, nas janelas de tempo de 20 dias úteis dos períodos de validação e de teste.

Além disso, tais evidências sugerem que a inclusão de variáveis econômicas e financeiras com frequência diária contribui para aprimorar a acurácia preditiva, mesmo em horizontes tradicionalmente considerados curtos, como o de 20 dias úteis à frente (aproximadamente um mês).

Esses resultados estão em consonância com a literatura seminal de Meese e Rogoff (1982; 1983), que demonstrou a robustez do modelo RW como *benchmark* na previsão cambial, sobretudo em horizontes curtos. Embora pesquisas posteriores, como Rogoff (2008), tenham identificado ganhos pontuais de modelos estruturais em horizontes mais longos, a consistência de resultados favoráveis ao modelo RW, inclusive em estudos aplicados ao Brasil, como os de Perdomo e Botelho (2007) e Kopp (2019), reforça o desafio de superá-lo com abordagens baseadas exclusivamente em fundamentos econômicos. Por outro lado, estudos recentes, como

os descritos por Genin (2023) e Makika (2022), evidenciam o potencial de abordagens não lineares e baseadas em DL na previsão da taxa de câmbio. O primeiro, por exemplo, mostra que as previsões de modelos LSTM têm potencial de superar as do RW em determinados contextos, o que é coerente com os resultados alcançados por este estudo.

No que tange à literatura que aborda as arquiteturas de modelos de DL na atividade de previsão, é apontado por Siami-Namini, Tavakoli e Namin (2019) e García, Guijarro e Oliver (2024) que modelos bidirecionais, como o BiLSTM, podem ampliar a acurácia preditiva em séries temporais complexas. No entanto, neste estudo, não se identificou um padrão consistente de superioridade dos modelos de DL bidirecionais, sendo que o desempenho variou em função da inclusão de uma camada recorrente adicional na arquitetura dos modelos de DL.

Assim, ainda que as previsões dos modelos de DL não tenham apresentado desempenho sistematicamente superior ao do modelo RW em todas as configurações testadas, os resultados desta dissertação evidenciam ganhos relevantes e ressaltam o potencial da abordagem, sobretudo na formulação multivariada e em horizontes de 20 dias úteis à frente, onde se verificaram ganhos significativos em termos de acurácia em relação ao *benchmark*. Pesquisas futuras podem explorar aprimoramentos como mecanismos de *Attention*, modelos híbridos, retreinamento dinâmico e seleção adicional de variáveis explicativas, entre outros. Apesar de mais custosas em termos computacionais e de implementação, tais estratégias podem aumentar a capacidade dos modelos de DL de capturar a dinâmica complexa e não linear da taxa de câmbio.

## CONCLUSÃO

Esta dissertação investigou, por meio de um estudo aplicado, o desempenho de diferentes arquiteturas e especificações de modelos de DL na previsão da taxa de câmbio Ptax de venda do Real brasileiro em relação ao Dólar americano (BRL/USD), considerando horizontes de 1, 10 e 20 dias úteis à frente. Foram avaliados os modelos LSTM, BiLSTM, GRU e Bi-GRU, tanto na abordagem univariada, baseada apenas na série da respectiva taxa de câmbio, quanto na multivariada, que inclui variáveis explicativas econômicas e financeiras. As séries utilizadas possuem frequência diária, abrangendo o período de janeiro de 2020 a dezembro de 2024, e foram divididas sequencialmente em conjuntos de treinamento, validação e teste. Para cada abordagem e horizonte de previsão, consideraram-se duas arquiteturas de modelos de DL: uma com uma camada recorrente (1CR) e outra com duas camadas recorrentes (2CR). Como referência comparativa, adotou-se o modelo RW, amplamente reconhecido na literatura desde Meese e Rogoff (1982, 1983) como um *benchmark* robusto na previsão cambial, especialmente em horizontes curtos.

Na abordagem univariada, os resultados indicaram que, no período de teste, nenhuma das previsões dos modelos de DL selecionados superou a acurácia das do modelo RW nos horizontes analisados. Apenas no horizonte de previsão de 10 dias úteis à frente, o teste de Diebold-Mariano Modificado (1997), ao nível de significância de 5%, apontou equivalência estatística entre as previsões do RW e as do modelo de DL com melhor desempenho. No horizonte de 20 dias úteis, observou-se uma deterioração significativa da acurácia dos modelos de DL univariados, cujas métricas de erro superaram em pelo menos 50% as do modelo RW, independentemente da arquitetura ou das especificações adotadas.

Na abordagem multivariada, o teste de Diebold-Mariano Modificado (1997), ao nível de significância de 5%, indicou que a acurácia das previsões do modelo de DL selecionado foi inferior à do modelo RW no horizonte de previsão de 1 dia útil à frente. Para os horizontes de 10 e 20 dias úteis à frente, os testes apontaram equivalência estatística entre as previsões dos modelos comparados. No entanto, ao nível de significância de 10%, houve evidência de superioridade da acurácia das previsões do modelo de DL no horizonte de 20 dias úteis à frente, em ambas as arquiteturas (1CR e 2CR). Esse resultado foi corroborado pelo teste de flutuação de Giacomini e Rossi (2010), que identificou, nas janelas de tempo, desempenho igual ou superior das previsões dos modelos de DL em relação ao RW, e pelo teste de racionalidade de Rossi e Sekhposyan (2016), que evidenciou menor viés e menor irracionalidade nos erros de previsão dos modelos de DL multivariados também em relação ao *benchmark*.

Os resultados obtidos ressaltam a importância de se testar empiricamente diferentes configurações de modelos de DL, bem como de incluir variáveis explicativas no processo de modelagem, mesmo em horizontes tradicionalmente considerados curtos. Observou-se que, conforme o horizonte de previsão se estende de 10 para 20 dias úteis à frente, as previsões

dos modelos de DL multivariados apresentaram acurácia significativamente superior as do modelo RW. Tais resultados estão em consonância com parte da literatura recente, como em Genin (2023), que destaca o potencial das previsões de modelos LSTM para superar as previsões do modelo RW na atividade de previsão cambial. Adicionalmente, embora estudos como Siami-Namini, Tavakoli e Namin (2019) e García, Guijarro e Oliver (2024) indiquem que arquiteturas bidirecionais, como o BiLSTM, podem aumentar a acurácia preditiva em séries temporais complexas, neste trabalho não se identificou um padrão consistente de superioridade dessas arquiteturas, sugerindo que tais ganhos podem depender da estrutura dos dados e da profundidade do modelo.

Cabe ressaltar que o modelo RW mantém-se como um *benchmark* robusto, em consonância com os achados clássicos de Meese e Rogoff (1982, 1983). Ainda assim, os avanços recentes no campo do DL e a crescente disponibilidade de dados econômicos em alta frequência abrem perspectivas promissoras para a construção de modelos mais acurados e adaptáveis, sobretudo em economias emergentes como a brasileira.

Resta reconhecer que o presente estudo enfrentou limitações associadas a restrições computacionais, ao tempo disponível para implementação e à limitação de variáveis econômicas relevantes com frequência diária, fatores que restringiram tanto a complexidade das arquiteturas de DL quanto a aplicação de técnicas ainda mais sofisticadas e que estão em constante evolução. Isso pode ter limitado a performance dos modelos, evidenciando a importância de futuros trabalhos explorarem arquiteturas ainda mais robustas e avançadas, maior disponibilidade de dados em alta frequência e maior poder computacional para calibrações mais profundas. Além disso, recomenda-se avaliar o desempenho dos modelos em diferentes períodos, dado que os resultados aqui obtidos podem ter sido influenciados pelo contexto da pandemia de COVID-19 e seus efeitos subsequentes.

Pesquisas futuras podem ampliar esses resultados ao incluir novas variáveis explicativas, incorporar análise de sentimento e métodos de extração de informação, adotar modelos híbridos, realizar retreinamento contínuo após cada previsão e integrar mecanismos de *Attention*. Apesar do maior custo computacional e da complexidade de implementação, tais estratégias tendem a fortalecer a capacidade dos modelos de DL em capturar a dinâmica complexa e não linear da taxa de câmbio.

Por fim, considera-se que esta pesquisa contribui para o avanço de um campo ainda pouco explorado na literatura brasileira ao evidenciar, de forma empírica e metodológica, condições sob as quais as previsões dos modelos de DL podem superar as do modelo RW (*benchmark*) na previsão cambial, oferecendo caminhos concretos para investigações futuras em um campo de elevada relevância analítica e econômica.

## REFERÊNCIAS

ALMEIDA, J. P. A. O. d. *Análise sobre o comportamento do câmbio nominal brasileiro durante a pandemia*. Dissertação (Mestrado Profissional em Economia) — Escola de Economia de São Paulo, Fundação Getúlio Vargas, São Paulo, 2022. Disponível em: <<https://repositorio.fgv.br/server/api/core/bitstreams/a421d18e-dfc6-489b-98d8-bee50bea819b/content>>. Acesso em: 22 out. 2024.

BRASIL. Banco Central do Brasil. *A taxa de câmbio de referência Ptax. Estudo Especial n° 42/2019*. Brasília, 2019. Disponível em: <[https://www.bcb.gov.br/conteudo/relatorioinflacao/EstudosEspeciais/EE042\\_A\\_taxa\\_de\\_cambio\\_de\\_referencia\\_Ptax.pdf](https://www.bcb.gov.br/conteudo/relatorioinflacao/EstudosEspeciais/EE042_A_taxa_de_cambio_de_referencia_Ptax.pdf)>. Acesso em: 20 jun. 2024.

BRASIL. Banco Central do Brasil. *Resolução BCB n° 45, de 24 de novembro de 2020*: Dispõe sobre a metodologia de apuração da taxa de câmbio real/dólar americano divulgada pelo banco central do brasil (ptax). Brasília: Banco Central do Brasil, 2020. Disponível em: <<https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o%20BCB&numero=45>>. Acesso em: 20 jun. 2024.

CASTRO, C. *Previsão da série temporal da inflação usando modelos orientados por score dinâmico e métodos de aprimoramento de acurácia preditiva: o caso do Brasil*. Tese (Doutorado em Ciências Econômicas) — Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2023. Disponível em: <<https://www.bdt.d.uerj.br:8443/bitstream/1/20904/2/Tese%20-%20Carlos%20Henrique%20Dias%20Cordeiro%20de%20Castro%20-%202023%20-%20Completa.pdf>>. Acesso em: 15 out. 2024.

CASTRO, C. H. D. C. D.; AIUBE, F. A. L. Forecasting inflation time series using score-driven dynamic models and combination methods: The case of brazil. *Journal of Forecasting*, v. 42, p. 369–401, 2023. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/for.2908>>. Acesso em: 25 mar. 2025.

CHO, K.; MERRIËNBOER, B. V.; GULCEHRE, C. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078v3 [cs.CL]*, 2014. Disponível em: <<https://arxiv.org/abs/1406.1078v3>>. Acesso em: 10 jun. 2024.

CIABURRO, G.; VENKATESWARAN, B. *Neural Networks with R: Smart Models Using CNN, RNN, Deep Learning, and Artificial Intelligence Principles*. 1. ed. Birmingham, UK: Packt Publishing Ltd, 2017. ISBN 978-1-78839-787-2.

DIEBOLD, F. X. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 33, n. 1, p. 1–1, 2015. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/07350015.2014.983236>>. Acesso em: 05 ago. 2024.

DIEBOLD, F. X.; MARIANO, R. S. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, v. 13, n. 3, p. 253–263, July 1995. Disponível em: <<https://users.ssc.wisc.edu/~bhansen/718/DieboldMariano1995.pdf>>. Acesso em: 01 out. 2024.

DORNBUSCH, R.; FISCHER, S.; STARTZ, R. *Macroeconomia*. 11. ed. Porto Alegre, RS, Brasil: AMGH, 2013. Recurso eletrônico. ISBN 978-85-8055-185-3.

FAMA, E. F. The behavior of stock-market prices. *The Journal of Business*, v. 38, n. 1, p. 34–105, 1965. Disponível em: <<http://www.jstor.org/stable/2350752>>. Acesso em: 23 out. 2024.

GARCÍA, F.; GUIJARRO, F.; OLIVER, J. Foreign exchange forecasting models: Lstm and bilstm comparison. *Engineering Proceedings*, v. 68, n. 1, p. 19, 2024. Disponível em: <<https://www.mdpi.com/2673-4591/68/1/19>>. Acesso em: 10 dez. 2024.

GENIN, C. d. S. V. *Modelos de projeção de câmbio: uma investigação múltipla com séries de tempo, modelos estruturais e model selection*. Tese (Doutorado em Administração de Empresas) — Universidade Presbiteriana Mackenzie, São Paulo, 2023. Disponível em: <<https://dspace.mackenzie.br/items/dd0a7822-1792-4402-b377-5b9ae53ccf9a>>. Acesso em: 05 jan. 2025.

GIACOMINI, R.; ROSSI, B. Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, Wiley Online Library, v. 25, n. 4, p. 595–620, 2010. Disponível em: <<https://centreforfinance.org/researchdetail.htm?research=289>>. Acesso em: 15 out. 2024.

HADAD, E. J. *Projeção de taxas de câmbio: É possível superar o modelo de passeio aleatório?* Tese (Doutorado em Administração de empresas) — Universidade Presbiteriana Mackenzie, São Paulo, 2015. Disponível em: <<https://adelfa-api.mackenzie.br/server/api/core/bitstreams/7e51c897-0d55-469c-93d7-6bd908a42f3d/content>>. Acesso em: 17 dez. 2024.

HARVEY, D.; LEYBOURNE, S.; NEWBOLD, P. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, Elsevier, v. 13, n. 2, p. 281–291, 1997. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0169207096007194>>. Acesso em: 05 dez. 2024.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, MIT Press, v. 9, n. 8, p. 1735–1780, 1997. Disponível em: <<https://didawiki.cli.di.unipi.it/lib/exe/fetch.php/magistraleinformatica/aa2/lstm.pdf>>. Acesso em: 23 mai. 2024.

IQUIAPAZA, R. A.; BRESSAN, A. A.; AMARAL, H. F. Previsão não-linear de retornos na bovespa: volume negociado em um modelo auto-regressivo de transição suave. *Revista de Administração Contemporânea*, Associação Nacional de Pós-Graduação e Pesquisa em Administração, Rio de Janeiro, v. 14, n. 1, p. 149–171, 2010. Disponível em: <<https://www.redalyc.org/pdf/840/84012377009.pdf>>. Acesso em: 07 out. 2024.

JOSEPH, M.; TACKES, J. *Modern Time Series Forecasting with Python: Industry-ready Machine Learning and Deep Learning Time Series Analysis with PyTorch and pandas*. 2. ed. Birmingham, UK: Packt Publishing Ltd, 2024. ISBN 978-1-83588-318-1.

KOPP, V. A. I. *Comportamento da taxa de câmbio no Brasil: evidências empíricas*. Dissertação (Mestrado Profissional em Finanças e Economia) — Escola de Economia de São Paulo, Fundação Getulio Vargas, São Paulo, 2019. Disponível em: <<https://repositorio.fgv.br/items/1dc7369b-4743-424d-a883-9a71e019a607>>. Acesso em: 19 nov. 2024.

KRUGMAN, P. R.; OBSTFELD, M.; MELITZ, M. J. *Economia Internacional*. 10. ed. São Paulo: Pearson Education do Brasil, 2015. Tradução de Ana Julia Perrotti-Garcia. ISBN 978-85-4301-506-4.

MAKIKA, H. *Previsão de séries temporais com aprendizagem profunda: uma aplicação para taxa de câmbio*. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, Campinas, 2022. Disponível em: <<https://repositorio.unicamp.br/acervo/detalhe/1259464>>. Acesso em: 11 dez. 2024.

MARINS, J. T. M. *Predictability of Exchange Rate Density Forecasts for Emerging Economies in the Short Run*. Brasília, 2024. 3–53 p. Disponível em: <<https://www.bcb.gov.br/content/publicacoes/WorkingPaperSeries/WP588v2.pdf>>. Acesso em: 28 nov. 2024.

MBEDZI, T. A. *Share Price Prediction for Increasing Market Efficiency using Random Forest*. Dissertação (Master of Science Degree in e-Science) — University of Venda, Department of Mathematical and Computational Sciences, Faculty of Science, Engineering and Agriculture, Thohoyandou, Limpopo Province, South Africa, 2022. Disponível em: <<https://univendspace.univen.ac.za/server/api/core/bitstreams/06b74889-c9fa-4a24-9bf6-35b3a22756b8/content>>. Acesso em: 15 jan. 2025.

MEESE, R.; ROGOFF, K. S. *The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?* Cambridge, MA, USA, 1982. (International Finance Discussion Papers, 204). Disponível em: <<https://www.federalreserve.gov/pubs/ifdp/1982/204/ifdp204.pdf>>. Acesso em: 26 mai. 2024.

MEESE, R. A.; ROGOFF, K. Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, Elsevier, v. 14, n. 1-2, p. 3–24, 1983. Disponível em: <[https://rogooff.scholars.harvard.edu/sites/g/files/omnuum5901/files/rogooff/files/51\\_jie1983.pdf](https://rogooff.scholars.harvard.edu/sites/g/files/omnuum5901/files/rogooff/files/51_jie1983.pdf)>. Acesso em: 26 mai. 2024.

MINCER, J. A.; ZARNOWITZ, V. *The evaluation of economic forecasts*. Cambridge, MA, USA, 1969. 3–46 p. Disponível em: <<https://www.nber.org/system/files/chapters/c1214/c1214.pdf>>. Acesso em: 29 out. 2024.

NEWAY, W. K.; WEST, K. D. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, v. 55, n. 3, p. 703–708, 1987. Disponível em: <<https://www.jstor.org/stable/1913610>>. Acesso em: 25 jan. 2025.

PERDOMO, J. P. J.; BOTELHO, F. B. Messe-rogooff revisitados: uma análise empírica das projeções para a taxa de câmbio no brasil. *Encontro Nacional de Economia da Associação Nacional dos Centros de Pós-Graduação em Economia–ANPEC*, v. 35, 2007. Disponível em: <<https://www.anpec.org.br/encontro2007/artigos/A07A038.pdf>>. Acesso em: 10 jun. 2024.

PETRACCA, L. *Time Series Forecasting using Transformers with Sentiment Analysis on Financial Data*. Dissertação (Master in Digital Innovation) — Universidad Politécnica de Madrid, E.T.S. de Ingenieros Informáticos, Dpto. Lenguajes y Sistemas Informáticos e Ingeniería de Software, Madrid, Spain, 2024. Disponível em: <[https://oa.upm.es/82874/1/TFM\\_LUCA\\_PETRACCA.pdf](https://oa.upm.es/82874/1/TFM_LUCA_PETRACCA.pdf)>. Acesso em: 20 jan. 2025.

PRODANOV, C. C.; FREITAS, E. C. d. *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico*. 2. ed. Novo Hamburgo, RS, Brasil: Editora Feevale, 2013. ISBN 978-85-7717-158-3.



ROGOFF, K. Comment on "exchange rate models are not as bad as you think". In: ACEMOGLU, D.; WOODFORD, M.; ROGOFF, K. (Ed.). *NBER Macroeconomics Annual 2007, Volume 22*. Chicago: University of Chicago Press, 2008. p. 443–452. Disponível em: <<http://www.nber.org/chapters/c4076>>. Acesso em: 16 nov. 2024.

ROGOFF, K.; STAVRAKEVA, V. *The Continuing Puzzle of Short-Horizon Exchange Rate Forecasting*. Cambridge: Harvard University, 2008. Disponível em: <<http://www.nber.org/papers/w14071>>. Acesso em: 16 nov. 2024.

ROSSI, B. Exchange rate predictability. *Journal of Economic Literature*, American Economic Association, Nashville, TN, USA, v. 51, n. 4, p. 1063–1119, 2013. Disponível em: <<https://www.aeaweb.org/articles?id=10.1257/jel.51.4.1063>>. Acesso em: 10 jan. 2025.

ROSSI, B.; SEKHPOSYAN, T. *Forecast Rationality Tests in the Presence of Instabilities, with Applications to Federal Reserve and Survey Forecasts*. Barcelona, Spain, 2014. Disponível em: <<https://bw.bse.eu/wp-content/uploads/2015/09/765-file.pdf>>. Acesso em: 27 dez. 2024.

ROSSI, B.; SEKHPOSYAN, T. Forecast rationality tests in the presence of instabilities, with applications to federal reserve and survey forecasts. *Journal of Applied Econometrics*, v. 31, n. 3, p. 507–532, 2016. Disponível em: <<https://onlinelibrary.wiley.com/doi/epdf/10.1002/jae.2440>>. Acesso em: 27 dez. 2024.

ROSSI, B.; SOUPRE, M. Implementing tests for forecast evaluation in the presence of instabilities. *The Stata Journal*, SAGE Publications, Los Angeles, CA, USA, v. 17, n. 4, p. 850–865, 2017. Disponível em: <<https://journals.sagepub.com/doi/pdf/10.1177/1536867X1801700405>>. Acesso em: 10 jan. 2025.

SABA, N. d. M. *Avaliando o desempenho preditivo de modelos de taxa de câmbio real efetiva: análise do caso brasileiro*. Dissertação (Mestrado em Economia) — Escola de Economia de São Paulo, Fundação Getulio Vargas, São Paulo, Brasil, 2015. Disponível em: <<https://repositorio.fgv.br/server/api/core/bitstreams/b4716efa-0691-4bd1-8494-c192eaa579f3/content>>. Acesso em: 30 nov. 2024.

SCHUSTER, M.; PALIWAL, K. K. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, v. 45, n. 11, p. 2673–2681, 1997. Disponível em: <<https://www.cin.ufpe.br/~fnj/RNA/bibliografia/BRNN.pdf>>. Acesso em: 03 out. 2024.

SIAMI-NAMINI, S.; TAVAKOLI, N.; NAMIN, A. S. The performance of lstm and bilstm in forecasting time series. *2019 IEEE International Conference on Big Data*, p. 3285–3292, 2019. Disponível em: <<https://par.nsf.gov/servlets/purl/10186554>>. Acesso em: 19 nov. 2024.

TÜRKOĞLU, F.; GÖÇEÇEK, E.; YUMRUKUZ, Y. Predictive abilities of machine learning and deep learning approaches for exchange rate prediction. *BDDK Bankacılık ve Finansal Piyasalar Dergisi*, Bankacılık Düzenleme ve Denetleme Kurumu, v. 18, n. 2, p. 186–210, 2024. Disponível em: <[https://www.bddk.org.tr/Content/docs/bddkDergiTr/dergi\\_0036\\_07.pdf](https://www.bddk.org.tr/Content/docs/bddkDergiTr/dergi_0036_07.pdf)>. Acesso em: 15 jan. 2025.

WEST, K. D. Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 1067–1084, 1996. Disponível em: <<https://users.ssc.wisc.edu/~kwest/publications/1990/Asymptotic%20Inference%20about%20Predictive%20Ability.pdf>>. Acesso em: 19 dez. 2024.

WEST, K. D.; MCCRACKEN, M. W. Regression-based tests of predictive ability. *International Economic Review*, v. 39, n. 4, p. 1067–1087, 1998. Disponível em: <<https://users.ssc.wisc.edu/~kwest/publications/1990/Regression%20Based%20Tests%20of%20Predictive%20Ability.pdf>>. Acesso em: 13 dez. 2024.

**APÊNDICE A – Testes estatísticos para comparação do desempenho das previsões do modelo RW e dos modelos de DL selecionados com segundo menor valor nas métricas de erro no período de teste**

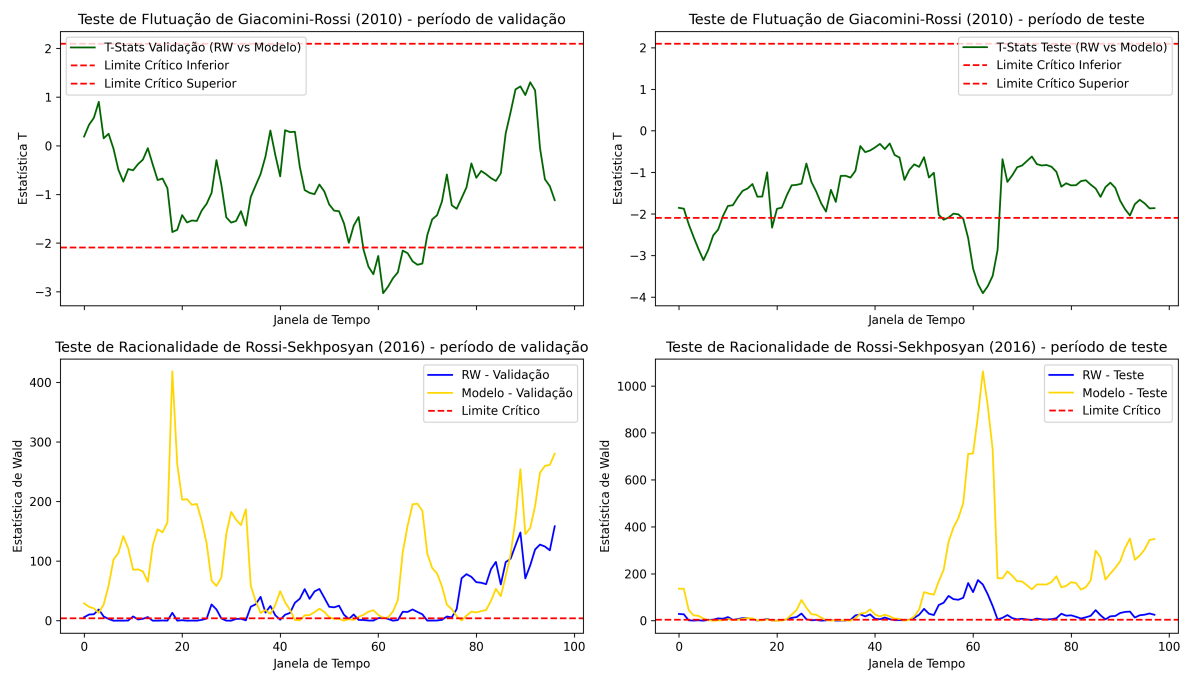
**Tabela 27 - Testes estatísticos para comparação do modelo univariado de DL selecionado com 1CR e do modelo RW nas previsões 1 dia à frente**

Testes	Período de Validação	Período de Teste
<b>Diebold-Mariano (1995)</b>	RW vs BiLSTM: Estatística = -2.2474, p-valor = 0.0246	RW vs BiLSTM: Estatística = -2.7020, p-valor = 0.0069
<b>Diebold-Mariano Modificado (1997)</b>	RW vs BiLSTM: Estatística = -2.2376, p-valor = 0.0272	RW vs BiLSTM: Estatística = -2.6904, p-valor = 0.0082
<b>Flutuação de Giacomini e Rossi (2010)</b>	RW vs BiLSTM: Estatística = -3.0298, p-valor = 0.0069	RW vs BiLSTM: Estatística = -3.9080, p-valor = 0.0009
<b>Racionalidade de Rossi e Sekhposyan (2016)</b>	RW: Estatística = 158.5189, p-valor = 0.0000	RW: Estatística = 173.0916, p-valor = 0.0000
<b>Racionalidade de Rossi e Sekhposyan (2016)</b>	BiLSTM: Estatística = 418.5315, p-valor = 0.0000	BiLSTM: Estatística = 1061.9670, p-valor = 0.0000

**Nota:** Refere-se ao modelo BiLSTM (adamw-LR0.0001 - Dropout0.1 - Batch16)

**Fonte:** O autor, 2025.

**Figura 23 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação nas previsões 1 dia à frente**



**Nota:** Referente ao modelo selecionado univariado com 1CR, nas previsões 1 dia útil à frente. As janelas móveis possuem 20 dias úteis nos dois testes.

**Fonte:** O autor, 2025.

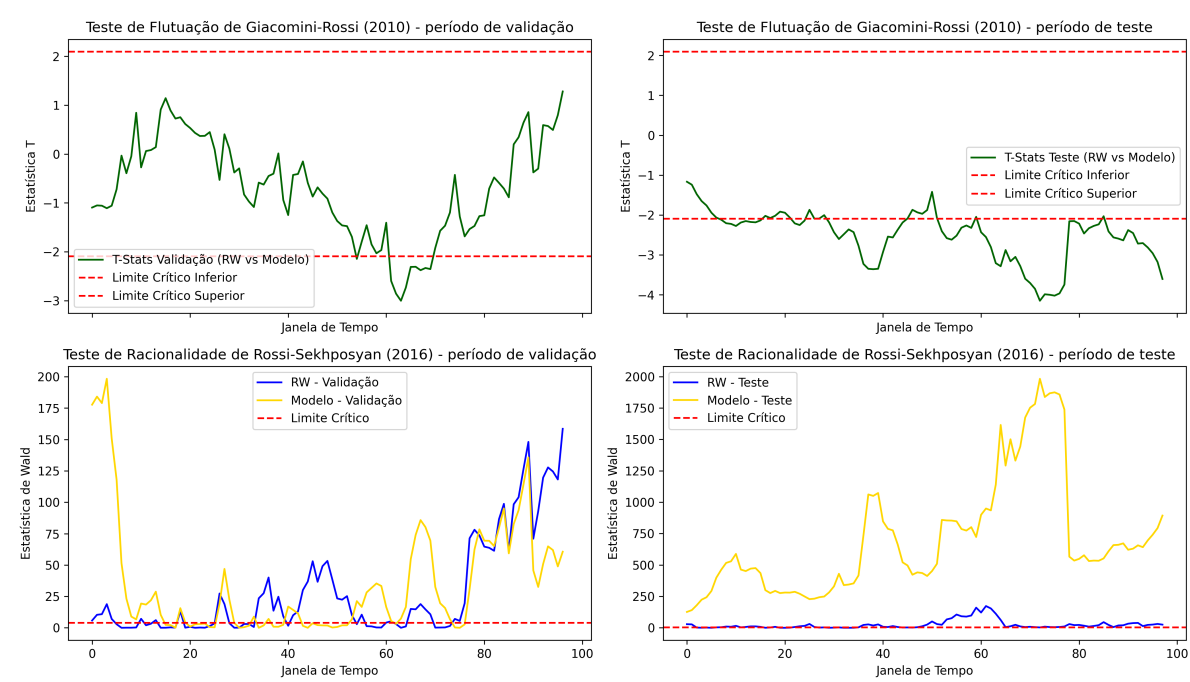
Tabela 28 - Testes estatísticos para comparação do modelo multivariado de DL selecionado com 1CR e do modelo RW nas previsões 1 dia à frente

Testes	Período de Validação	Período de Teste
Diebold-Mariano (DM)	RW vs BiLSTM: Estatística = -1.3222, p-valor = 0.1861	RW vs BiLSTM: Estatística = -3.6847, p-valor = 0.0002
Teste de Diebold-Mariano-Modificado (DMM)	RW vs BiLSTM: Estatística = -1.3165, p-valor = 0.1906	RW vs BiLSTM: Estatística = -3.6689, p-valor = 0.0004
Flutuação de Giacomini e Rossi (2010)	RW vs BiLSTM: Estatística = -3.7007, p-valor = 0.0073	RW vs BiLSTM: Estatística = -3.6689, p-valor = 0.0005
Racionalidade de Rossi e Sekhposyan (2016)	RW: Estatística = 158.5189, p-valor = 0.0000	RW: Estatística = 173.0916, p-valor = 0.0000
Racionalidade de Rossi e Sekhposyan (2016)	BiLSTM: Estatística = 198.3556, p-valor = 0.0000	BiLSTM: Estatística = 1982.2648, p-valor = 0.0000

**Nota:** Refere-se ao modelo BiLSTM (adam-LR0.0001 - Dropout0.1 - Batch16).

**Fonte:** O autor, 2025.

Figura 24 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação



**Nota:** Referente ao modelo selecionado multivariado com 1CR, nas previsões 1 dia útil à frente. As janelas móveis possuem 20 dias úteis nos dois testes.

**Fonte:** O autor, 2025.

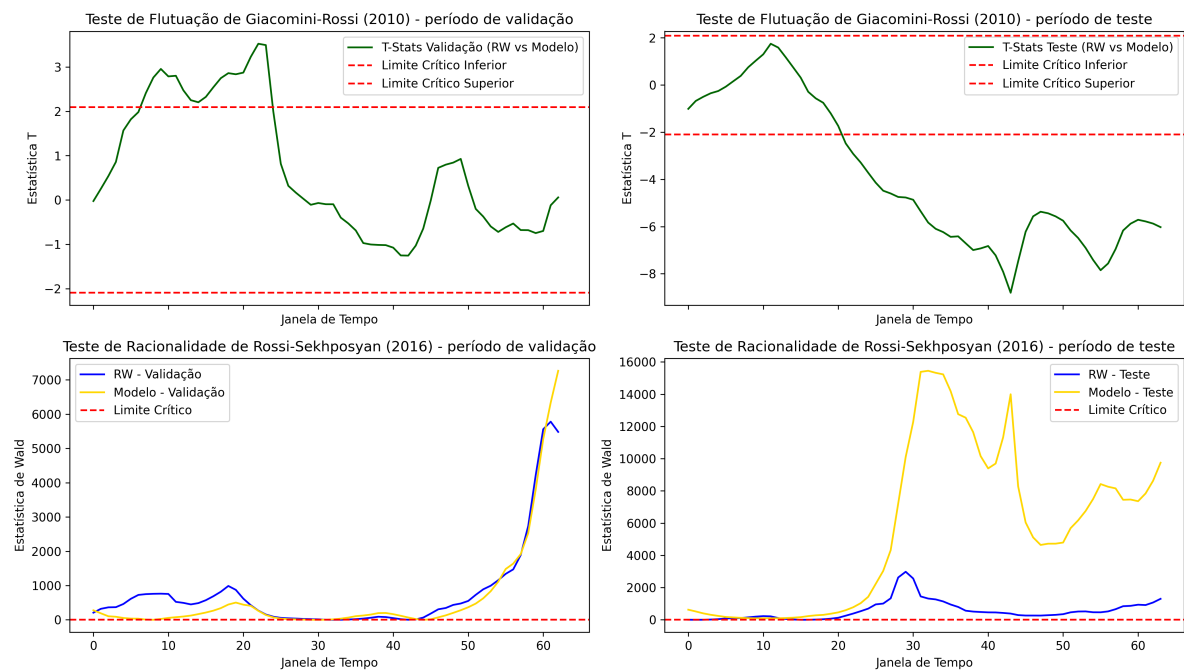
Tabela 29 - Testes estatísticos para comparação do modelo univariado de DL selecionado com 2CR e do modelo RW nas previsões 10 dias úteis à frente

Testes	Período de Validação	Período de Teste
Diebold-Mariano (DM)	RW vs BiLSTM: Estatística = 0.3266, p-valor = 0.7440	RW vs BiLSTM: Estatística = -3.1469, p-valor = 0.0017
Teste de Diebold-Mariano-Modificado (DMM)	RW vs BiLSTM: Estatística = 0.2887, p-valor = 0.7735	RW vs BiLSTM: Estatística = -2.7866, p-valor = 0.0066
Flutuação de Giacomini e Rossi (2010)	RW vs BiLSTM: Estatística = 3.5232, p-valor = 0.0000	RW vs BiLSTM: Estatística = -8.8055, p-valor = 0.0000
Racionalidade de Rossi e Sekhposyan (2016)	RW: Estatística = 5779.2941, p-valor = 0.0000	RW: Estatística = 2980.7553, p-valor = 0.0000
Racionalidade de Rossi e Sekhposyan (2016)	BiLSTM: Estatística = 7263.5454, p-valor = 0.0000	BiLSTM: Estatística = 15451.7924, p-valor = 0.0000

**Nota:** Refere-se ao modelo BiLSTM-adamw (LR0.0001-Dropout0.1-Batch32).

**Fonte:** O autor, 2025.

Figura 25 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação



**Nota:** Referente ao modelo selecionado univariado com 2CR, nas previsões 10 dias úteis à frente. As janelas móveis possuem 20 dias úteis nos dois testes.

**Fonte:** O autor, 2025.

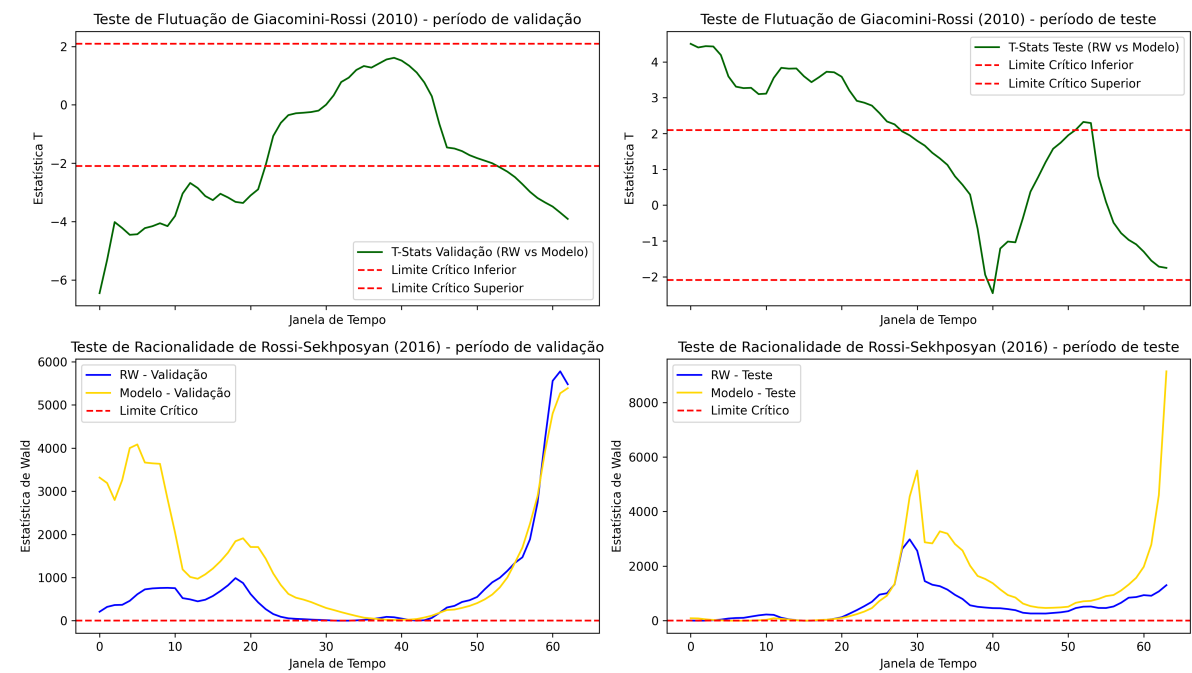
Tabela 30 - Testes estatísticos para comparação do modelo multivariado de DL selecionado com 1CR e do modelo RW nas previsões 10 dias úteis à frente

Testes	Período de Validação	Período de Teste
Diebold-Mariano (DM)	RW vs LSTM: Estatística = -2.3417, p-valor = 0.0192	RW vs LSTM: Estatística = 0.1469, p-valor = 0.8832
Teste de Diebold-Mariano-Modificado (DMM)	RW vs LSTM: Estatística = -2.0704, p-valor = 0.0416	RW vs LSTM: Estatística = 0.1300, p-valor = 0.8968
Flutuação de Giacomini e Rossi (2010)	RW vs LSTM: Estatística = -6.4545, p-valor = 0.0000	RW vs LSTM: Estatística = 4.5054, p-valor = 0.0007
Racionalidade de Rossi e Sekhposyan (2016)	RW: Estatística = 5779.2941, p-valor = 0.0000	RW: Estatística = 2980.7553, p-valor = 0.0000
Racionalidade de Rossi e Sekhposyan (2016)	LSTM: Estatística = 5388.9182, p-valor = 0.0000	LSTM: Estatística = 9139.8267, p-valor = 0.0000

**Nota:** Refere-se ao modelo LSTM-adamw (LR0.0001 - Dropout0.1 - Batch32) com duas camadas recorrentes.

**Fonte:** O autor, 2025.

Figura 26 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação



**Nota:** Referente ao modelo selecionado multivariado com 1CR, nas previsões 10 dias úteis à frente. As janelas móveis possuem 20 dias úteis nos dois testes.

**Fonte:** O autor, 2025.

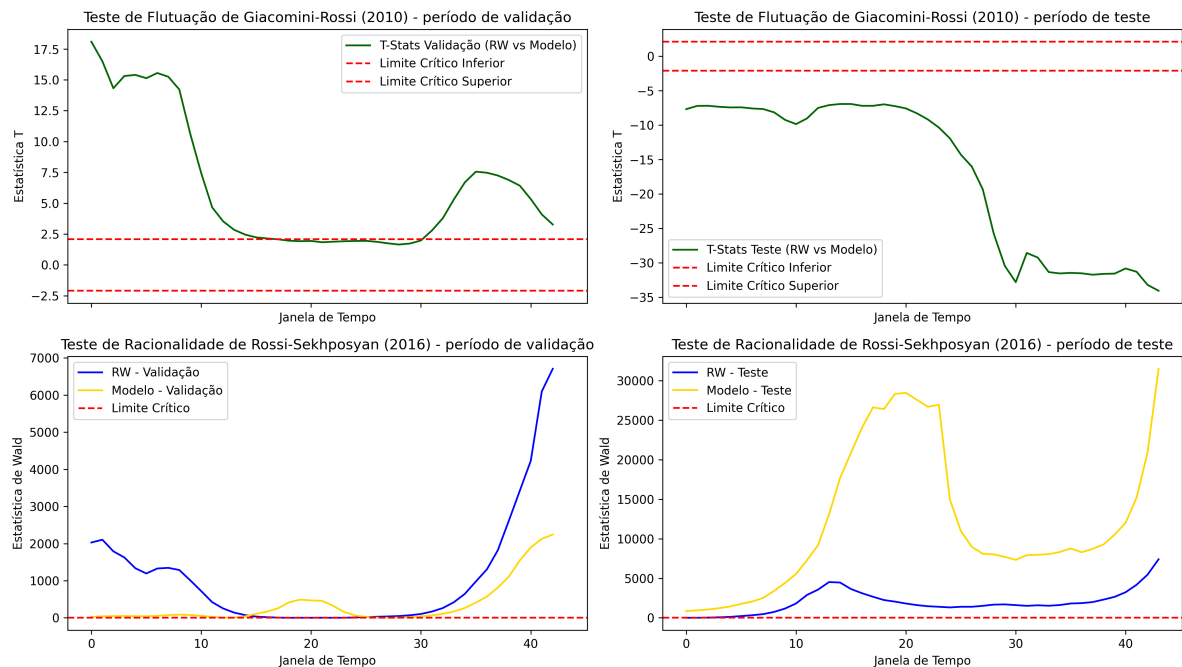
Tabela 31 - Testes estatísticos para comparação do modelo univariado de DL selecionado com 2CR e do modelo RW nas previsões 20 dias úteis à frente

Testes	Período de Validação	Período de Teste
<b>Diebold-Mariano (DM)</b>	RW vs GRU: Estatística = 3.7469, p-valor = 0.0002	RW vs GRU: Estatística = -5.0249, p-valor = 0.0000
<b>Teste de Diebold-Mariano-Modificado (DMM)</b>	RW vs GRU: Estatística = 2.5683, p-valor = 0.0127	RW vs GRU: Estatística = -3.4694, p-valor = 0.0010
<b>Flutuação de Giacomini e Rossi (2010)</b>	RW vs GRU: Estatística = 18.0825, p-valor = 0.0000	RW vs GRU: Estatística = -34.0440, p-valor = 0.0000
<b>Racionalidade de Rossi e Sekhposyan (2016)</b>	GRU: Estatística = 2245.1203, p-valor = 0.0000	GRU: Estatística = 31505.8133, p-valor = 0.0000
<b>Racionalidade de Rossi e Sekhposyan (2016)</b>	RW: Estatística = 6708.8322, p-valor = 0.0000	RW: Estatística = 7400.7754, p-valor = 0.0000

**Nota:** Refere-se ao modelo GRU-adamw (LR0.0001 - Dropout0.1 - Batch16) com 20 dias úteis de intervalo para validação e teste.

**Fonte:** O autor, 2025.

Figura 27 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação



**Nota:** Referente ao modelo selecionado univariado com 2CR, nas previsões 20 dias úteis à frente. As janelas móveis possuem 20 dias úteis nos dois testes.

**Fonte:** O autor, 2025.

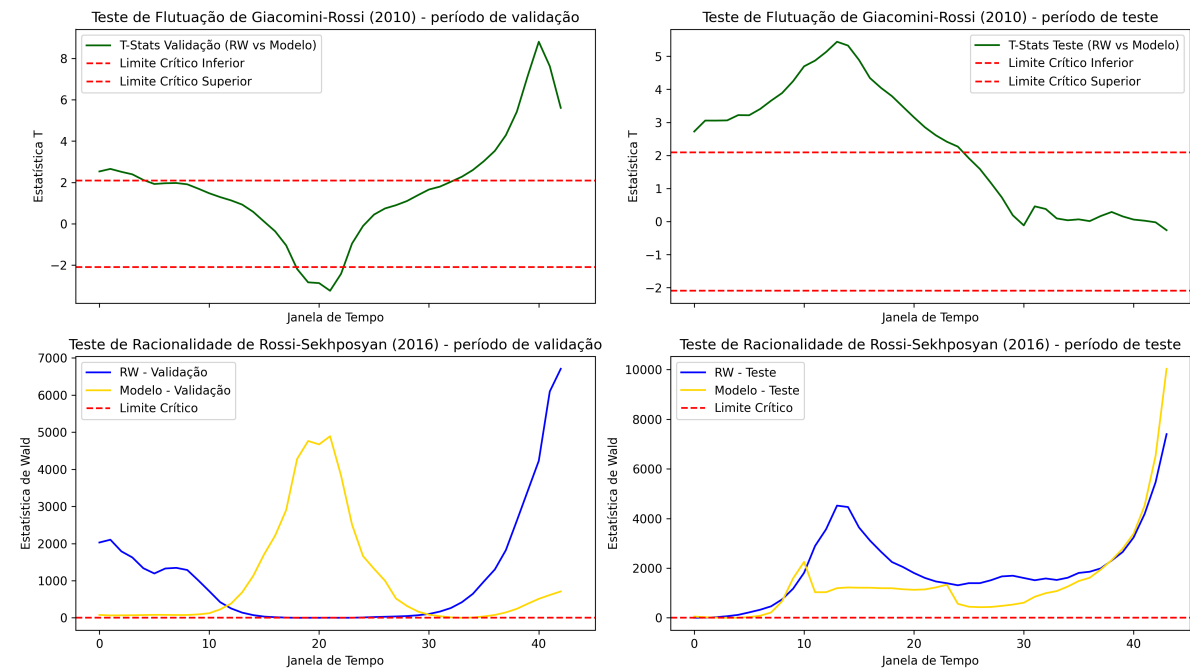
Tabela 32 - Testes estatísticos para comparação do modelo multivariado de DL selecionado com 2CR e do modelo RW nas previsões 20 dias úteis à frente

Testes	Período de Validação	Período de Teste
Diebold-Mariano (DM)	RW vs BiLSTM: Estatística = 2.7615, p-valor = 0.0058	RW vs BiLSTM: Estatística = 2.5029, p-valor = 0.0123
Teste de Diebold-Mariano-Modificado (DMM)	RW vs BiLSTM: Estatística = 1.8928, p-valor = 0.0631	RW vs BiLSTM: Estatística = 1.7281, p-valor = 0.0890
Flutuação de Giacomini e Rossi (2010)	RW vs BiLSTM: Estatística = 27.7071, p-valor = 0.0000	RW vs BiLSTM: Estatística = 9.3324, p-valor = 0.0000
Racionalidade de Rossi e Sekhposyan (2016)	BiLSTM: Estatística = 4891.8103, p-valor = 0.0000	BiLSTM: Estatística = 10032.5969, p-valor = 0.0000
Racionalidade de Rossi e Sekhposyan (2016)	RW: Estatística = 6708.8322, p-valor = 0.0000	RW: Estatística = 7400.7754, p-valor = 0.0000

**Nota:** Refere-se ao modelo BiLSTM-adamw (LR0.0001 - Dropout0.1 - Batch16) com 20 dias úteis de intervalo para validação e teste.

**Fonte:** O autor, 2025.

Figura 28 - Gráficos das estatísticas dos testes de flutuação Giacomini-Rossi (2010) e testes de racionalidade de Rossi-Sekhposyan (2016) ao longo dos períodos de teste e de validação



**Nota:** Referente ao modelo selecionado multivariado com 2CR, nas previsões 20 dias úteis à frente. As janelas móveis possuem 20 dias úteis nos dois testes.

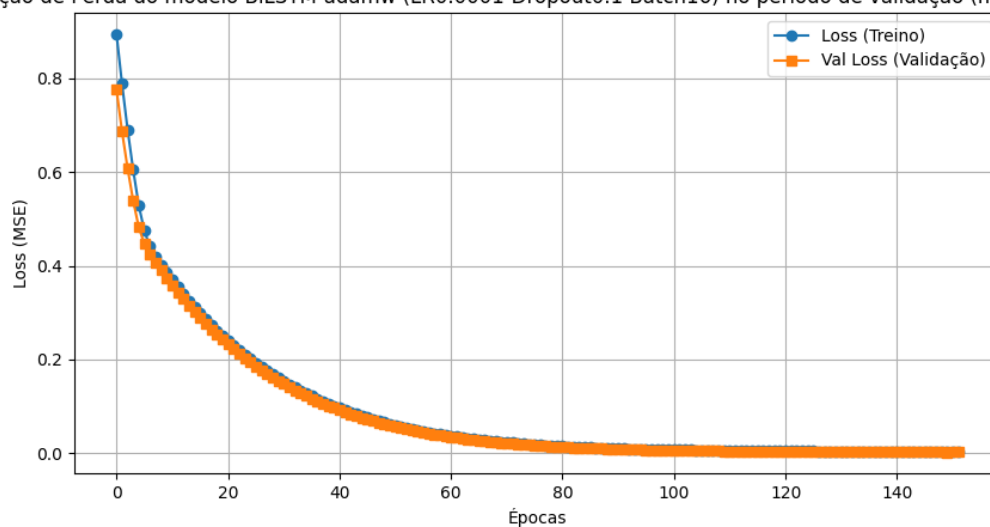
**Fonte:** O autor, 2025.



**APÊNDICE B** – Gráficos obtidos no treinamento dos modelos de DL selecionados com as menores métricas de erro no período de teste referente ao comportamento das funções de perda de treino e de validação

Figura 29 - Funções de perda dos períodos de treino e de validação do modelo de DL univariado com 1CR 1 dia à frente

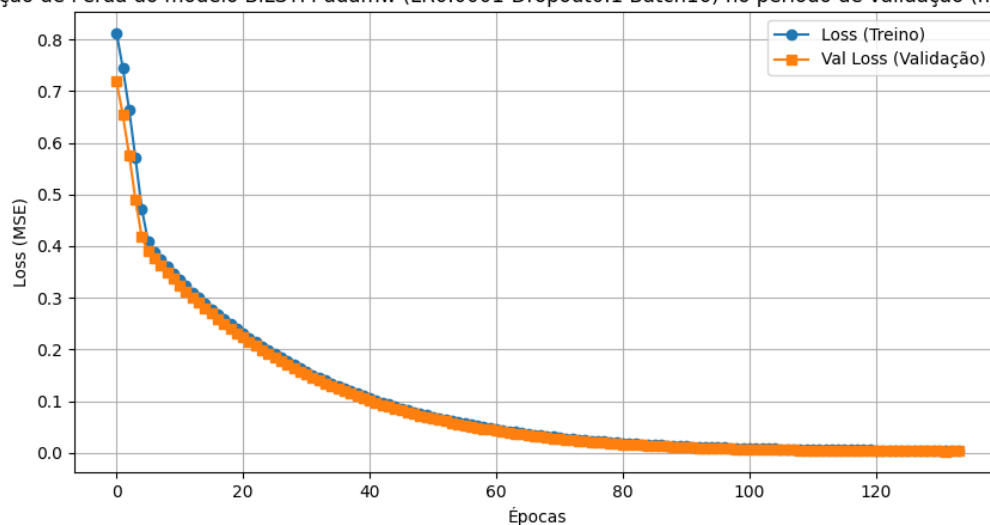
Função de Perda do modelo BiLSTM-adamw (LR0.0001-Dropout0.1-Batch16) no período de validação (menor RMSE)



**Fonte:** O autor, 2025.

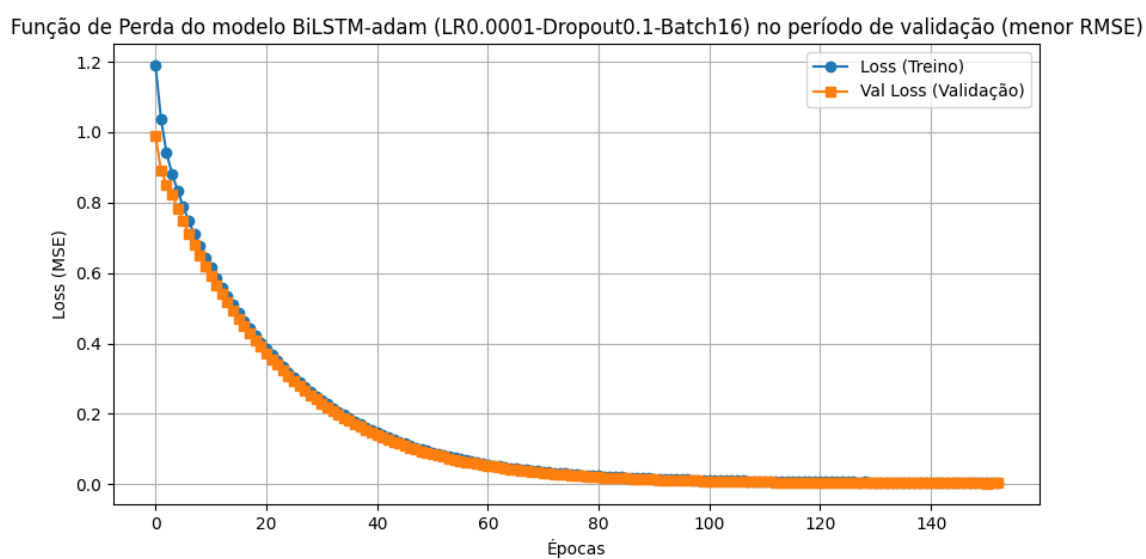
Figura 30 - Funções de perda dos períodos de treino e de validação do modelo de DL univariado com 2CR 1 dia à frente

Função de Perda do modelo BiLSTM-adamw (LR0.0001-Dropout0.1-Batch16) no período de validação (menor RMSE)



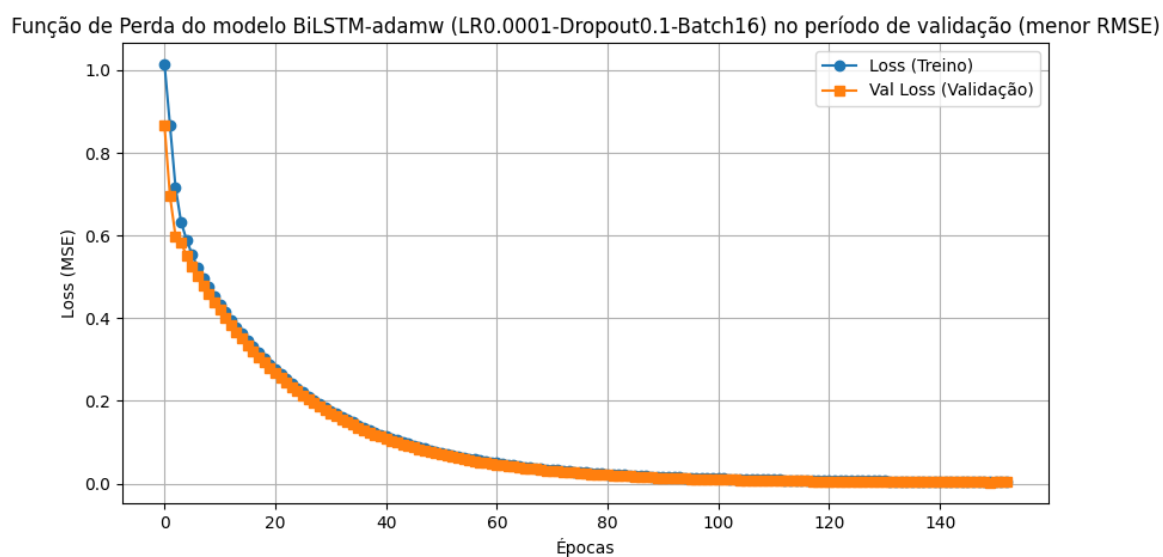
**Fonte:** O autor, 2025.

Figura 31 - Funções de perda dos períodos de treino e de validação do modelo de DL multivariado com 1CR 1 dia à frente



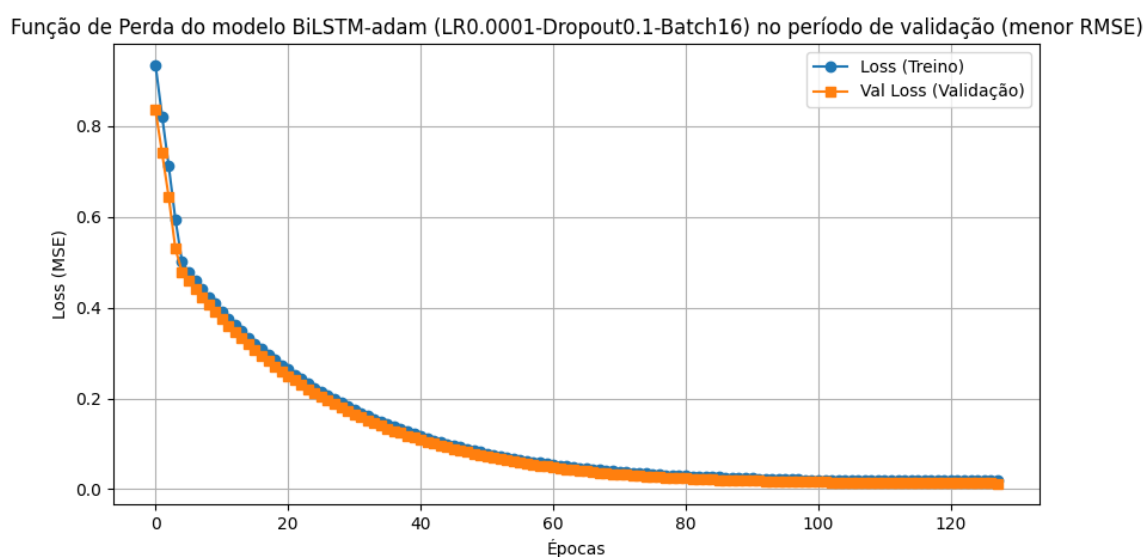
Fonte: O autor, 2025.

Figura 32 - Funções de perda dos períodos de treino e de validação do modelo de DL multivariado com 2CR 1 dia à frente



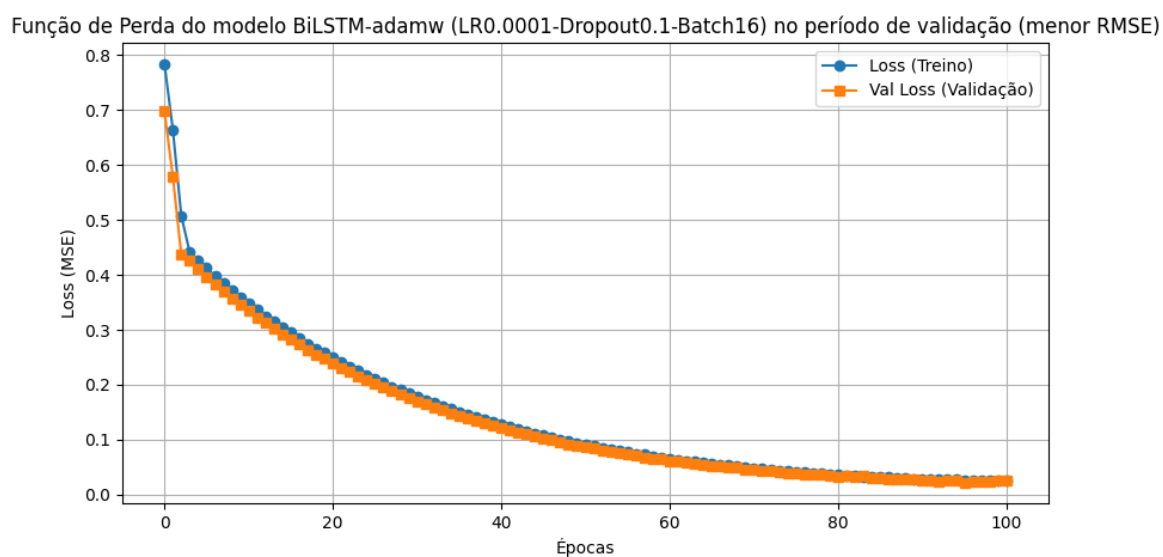
Fonte: O autor, 2025.

Figura 33 - Funções de perda dos períodos de treino e de validação do modelo de DL univariado de 1CR com previsão 10 dias úteis à frente



Fonte: O autor, 2025.

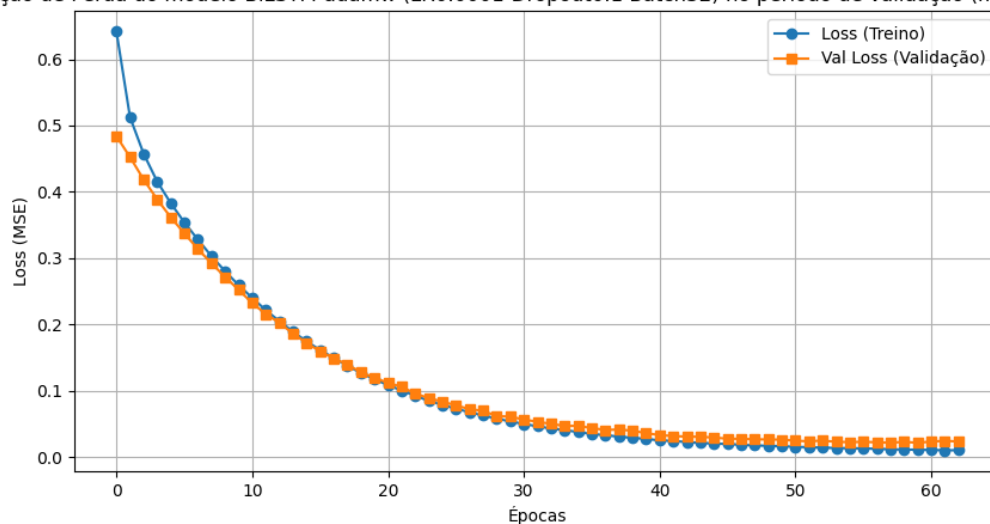
Figura 34 - Funções de perda dos períodos de treino e de validação do modelo de DL univariado de 2CR com previsão 10 dias úteis à frente



Fonte: O autor, 2025.

Figura 35 - Funções de perda dos períodos de treino e de validação do modelo de DL multivariado de 1CR com previsão 10 dias úteis à frente

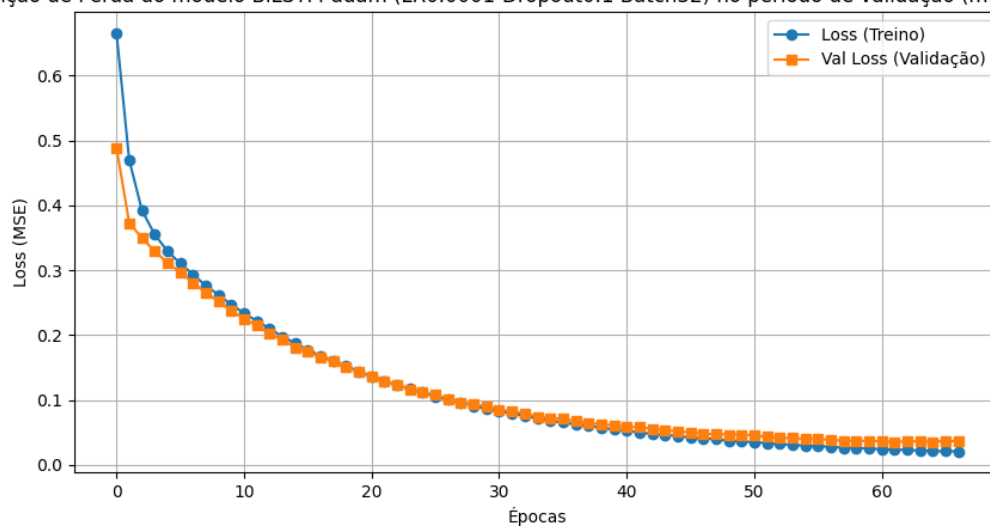
Função de Perda do modelo BiLSTM-adamw (LR0.0001-Dropout0.1-Batch32) no período de validação (menor RMSE)



Fonte: O autor, 2025.

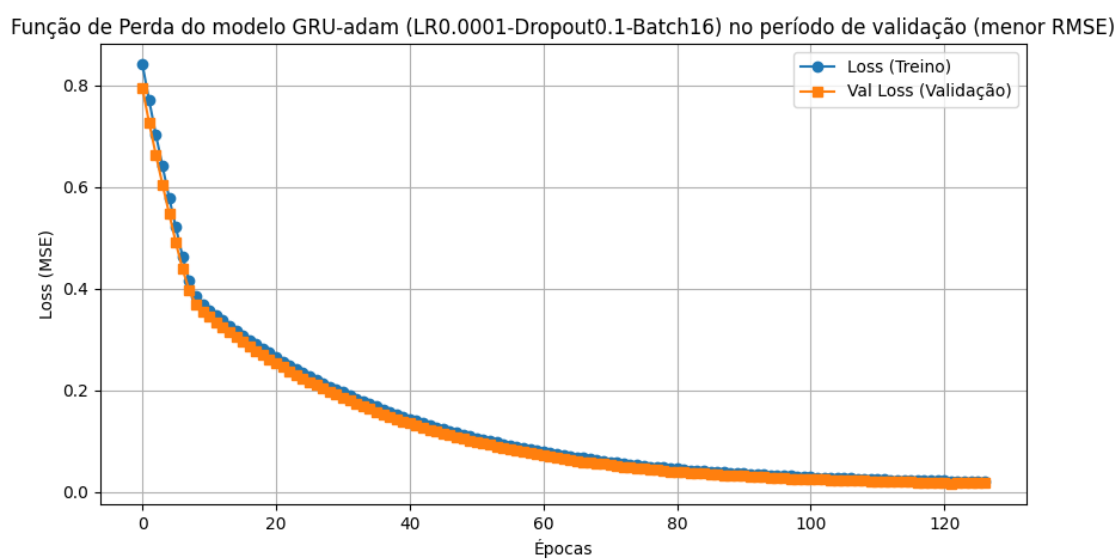
Figura 36 - Funções de perda dos períodos de treino e de validação do modelo de DL multivariado de 2CR com previsão 10 dias úteis à frente

Função de Perda do modelo BiLSTM-adam (LR0.0001-Dropout0.1-Batch32) no período de validação (menor RMSE)



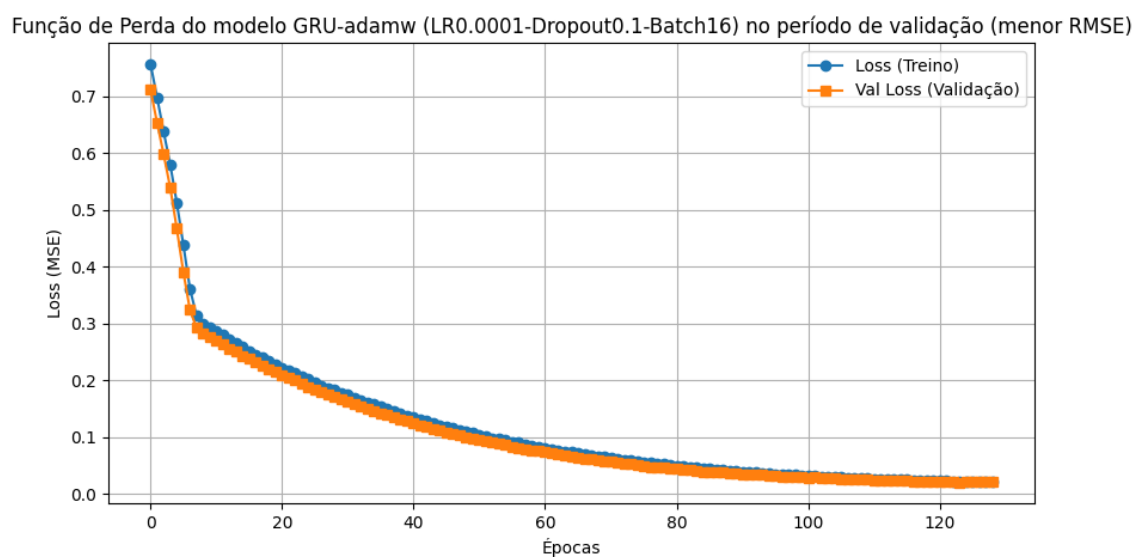
Fonte: O autor, 2025.

Figura 37 - Funções de perda dos períodos de treino e de validação do modelo de DL univariado de 1CR com previsão 20 dias úteis à frente



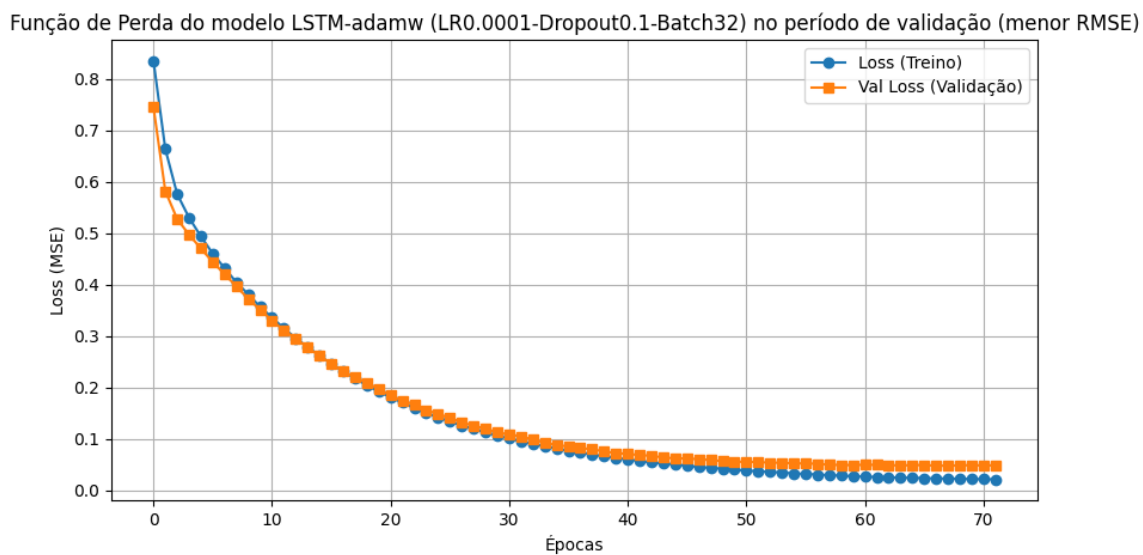
Fonte: O autor, 2025.

Figura 38 - Funções de perda dos períodos de treino e de validação do modelo de DL univariado de 2CR com previsão 20 dias úteis à frente



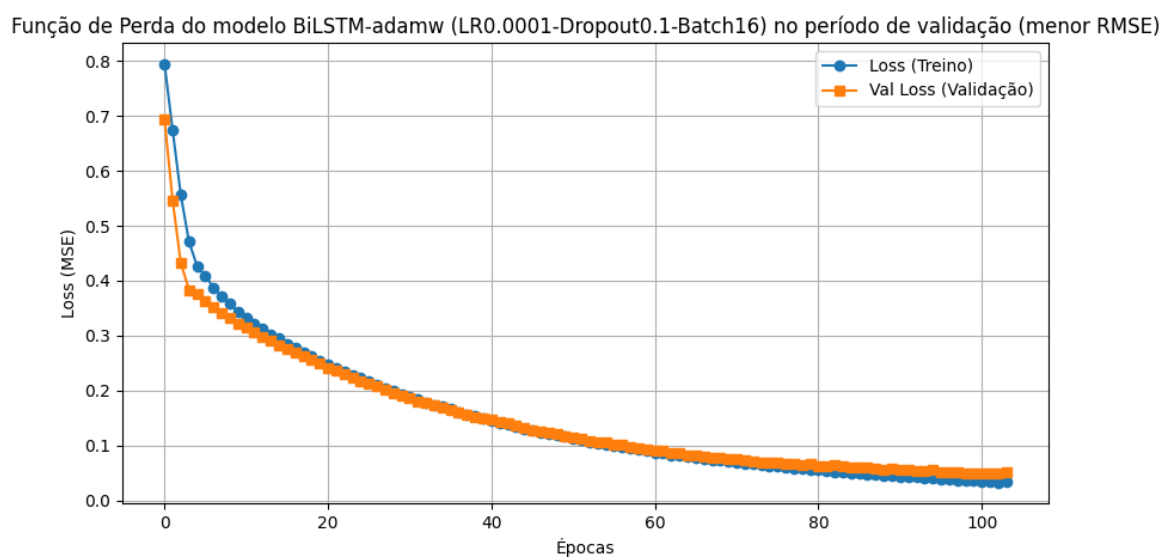
Fonte: O autor, 2025.

Figura 39 - Funções de perda dos períodos de treino e de validação do modelo de DL multivariado de 1CR com previsão 20 dias úteis à frente



Fonte: O autor, 2025.

Figura 40 - Funções de perda dos períodos de treino e de validação do modelo de DL multivariado de 2CR com previsão 20 dias úteis à frente



Fonte: O autor, 2025.