



Universidade de Brasília

Faculdade de Administração, Contabilidade, Economia e Gestão de
Políticas Públicas
Departamento de Economia

Dissertação de Mestrado

**Identificação de sinais de alertas antecipados na
curva de juros prefixada do Brasil com uso de redes
neurais**

por

Guilherme Costa Chadud Moreira

Brasília, 30 de março de 2026

Identificação de sinais de alertas antecipados na curva de juros prefixada do Brasil com uso de redes neurais

por

Guilherme Costa Chadud Moreira

Dissertação apresentada ao Departamento de Economia da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Economia.

Orientador: Prof. Dr. Daniel Oliveira Cajueiro

Brasília, 30 de março de 2026

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

MM838i Moreira, Guilherme Costa Chadud
Identificação de sinais de alertas antecipados na curva
de juros prefixada do Brasil com uso de redes neurais /
Guilherme Costa Chadud Moreira; orientador Daniel Oliveira
Cajueiro. Brasília, 2026.
77 p.

Dissertação(Mestrado em Economia) Universidade de
Brasília, 2026.

1. Curva de Juros. 2. Redes Neurais. 3. Sistema de
Alerta. 4. LSTM. I. Cajueiro, Daniel Oliveira, orient. II.
Título.

Dissertação submetida ao Programa de Pós-Graduação em Economia do Departamento de Economia da Universidade de Brasília como parte dos requisitos para a obtenção do título de Mestre em Economia.

Texto aprovado por:

Prof. Dr. Daniel Oliveira Cajueiro
Orientador, UnB

Prof. Dr. Douglas Sad Silveira
Examinador externo à Instituição, UFJF

Prof. Dr. Herbert Kimura
Examinador externo ao Programa, UnB

Agradecimentos

Agradeço à minha família, aos amigos e aos professores pelo apoio e pela motivação ao longo desta etapa de intenso aprendizado. Registro, ainda, meu agradecimento ao professor Daniel Oliveira Cajueiro pela orientação e pelo suporte no desenvolvimento deste trabalho, bem como pelas valiosas contribuições dos membros da banca examinadora.

Resumo

Este estudo busca identificar sinais de alerta antecipado na estrutura a termo da curva de juros prefixada brasileira, no período entre 2015 e 2025, com a utilização da técnica de redes neurais, LSTM (*Long Short Term Memory*). O estudo analisa a presença de comportamentos anômalos na curva de juros, de modo a visualizar na estrutura temporal dos dados, pontos que se diferenciem do padrão recente das maturidades estudadas. Neste sentido, é desenvolvido um sistema de alerta antecipado (*Early Warning System* – EWS) com o uso de redes neurais do tipo LSTM, aplicado à dinâmica da curva de juros prefixada em múltiplas maturidades, com uma janela expandida para os dados e *threshold* dinâmico ao longo do tempo. Os resultados são comparados com o modelo estatístico tradicional ARIMA, utilizando a mesma metodologia de desenvolvimento do sistema de alerta. As evidências obtidas indicam uma maior capacidade do modelo de redes neurais em identificar alterações em padrões mais persistentes, enquanto o modelo tradicional captura choques pontuais e menos estruturais. Com isso, o projeto contribui para o debate sobre a capacidade de identificar sinais de mudança na estrutura da curva de juros, permitindo algum tipo de sinalização aos agentes de mercado, em um contexto marcado por grandes desafios preditivos de diferentes características.

Palavras-chave: Curva de Juros; Redes Neurais; Sistema de Alerta; LSTM.

Abstract

This study aims to identify early warning signals in the term structure of the Brazilian fixed-rate yield curve between 2015 and 2025 through the application of Long Short Term Memory (LSTM) neural networks. The analysis investigates the presence of anomalous behavior in the yield curve, seeking to detect deviations from the recent pattern of the maturities under study within the temporal structure of the data. In this context, an Early Warning System (EWS) is developed using LSTM neural networks and applied to the dynamics of the fixed-rate yield curve across multiple maturities, employing an expanding data window and a time-varying threshold. The results are compared with those obtained from the traditional statistical ARIMA model, using the same methodological framework for the development of the warning system. The empirical evidence indicates that the neural network model exhibits a greater capacity to identify more persistent structural changes, whereas the traditional model tends to capture more localized and less structural shocks. Overall, the study contributes to the debate on the ability to detect signals of change in the structure of the yield curve, potentially providing informative signals to market participants in a context characterized by significant predictive challenges of different natures.

Keywords: Yield Curve; Neural Networks; Early Warning System; LSTM.

Sumário

1	Introdução	1
1.1	Contexto	1
1.2	Motivação	4
1.3	Objetivo	5
1.4	Esboço do Trabalho	6
2	Referencial Teórico	7
2.1	Modelagem da Curva de Juros	7
2.1.1	Modelo de Nelson-Siegel-Svensson	7
2.1.2	Modelo de Diebold-Li	10
2.1.3	Metodologia ANBIMA	12
2.2	Modelos de Redes Neurais em Aprendizado Profundo	14
2.2.1	Redes Neurais Recorrentes	15
2.2.2	Modelo LSTM	16
3	Metodologia	20
3.1	Base de Dados	20
3.2	Tratamento das Variáveis	22
3.3	Definição do Modelo	25
3.3.1	Validação Temporal	26

3.3.2	Previsão com LSTM	27
3.3.3	<i>Benchmark</i> Operacional	28
3.3.4	Cálculo do Desvio	29
3.3.5	Construção do <i>Threshold</i> Dinâmico	29
3.3.6	Geração de Alerta	30
4	Resultados	31
4.1	Resultados das Projeções	31
4.2	Resultado do Sistema de Alerta	34
4.3	Análise Descritiva dos Períodos de Alerta	36
4.4	Mudanças da Dinâmica da Curva de Juros	39
4.5	Análise Comparativa LSTM vs ARIMA	42
4.6	Avaliação de Desempenho dos Modelos	47
4.7	Efeito <i>Lead-Lag</i> dos Modelos	51
4.8	Análise de Sensibilidade dos Parâmetros	53
5	Conclusão	57
6	Apêndice	60
	Referências Bibliográficas	60

Lista de Tabelas

3.1	Estatísticas descritivas por vértice da curva	22
3.2	Hiperparâmetros do modelo LSTM	27
4.1	Períodos analisados e durações (em semanas)	36
4.2	Proporção dos vértices acionados no EWS	37
4.3	Frequência relativa dos segmentos da curva no EWS	38
4.4	Número de vértices presentes no EWS, por segmento	38
4.5	Volatilidade da curva de juros, por grupo	40
4.6	Inclinação da curva de juros, por grupo	41
4.7	Frequência dos eventos, por grupo	42
4.8	Períodos analisados e respectivas durações (em semanas), modelo ARIMA . . .	43
4.9	Comparação das métricas de volatilidade entre os modelos LSTM e ARIMA . .	44
4.10	Comparação das métricas de inclinação entre os modelos LSTM e ARIMA . .	46
4.11	Frequência dos eventos entre os modelos LSTM e ARIMA	46
4.12	RMSE e MAE agregado por maturidade, (EWS do LSTM)	48
4.13	RMSE e MAE agregado por semana, (EWS do LSTM)	48
4.14	RMSE e MAE agregado por maturidade, (EWS individual)	49
4.15	RMSE e MAE agregado por semana, (EWS individual)	49
4.16	$\Delta RMSE$ e ΔMAE dos modelos	50
4.17	Média e mediana das variáveis de sensibilidade	54

6.1	Resultados dos cenários simulados para diferentes combinações de W , percentil e limiar de maioria	60
-----	---	----

Lista de Figuras

2.1	Estrutura interna de uma arquitetura RNN. Fonte: Olah, 2015.	15
2.2	Estrutura interna de um modelo LSTM. Fonte: Olah, 2015.	19
3.1	Série histórica das taxas de juros prefixadas, por vértice	21
3.2	Série histórica semanal das taxas de juros prefixadas, por vértice	23
3.3	Série histórica semanal das taxas de juros prefixadas diferenciadas, por vértice .	25
4.1	Série projetada das taxas de juros prefixadas, por vértice	32
4.2	Desvios absolutos médios das projeções para a curva agregada	33
4.3	Resultado do indicador EWS acima do limite	35
4.4	Resultado do indicador EWS associado ao desvio médio	35
4.5	Proporção dos vértices acionados no EWS	37
4.6	<i>Boxplot</i> da volatilidade da curva, por grupo	40
4.7	<i>Boxplot</i> da inclinação da curva, por grupo	41
4.8	<i>Boxplot</i> da volatilidade da curva, por grupo, do modelo ARIMA	45
4.9	<i>Boxplot</i> da inclinação da curva, por grupo, do modelo ARIMA	46
4.10	<i>Lead-lag</i> dos desvios para RMSE médio	52
4.11	<i>Lead-lag</i> dos desvios para MAE médio	52

Abreviações e Siglas

ADF	<i>Augmented Dickey-Fuller</i>
AICc	<i>Akaike Information Criterion corrected</i>
ANBID	Associação Nacional dos Bancos de Investimento
ANBIMA	Associação Brasileira de Entidades dos Mercados Financeiros e de Capitais
ANDIMA	Associação Nacional das Instituições do Mercado Aberto
AR	<i>AutoRegressive Model</i>
ARIMA	<i>AutoRegressive Integrated Moving Average</i>
DL	Diebold-Li
DNS	<i>Dynamic Nelson-Siegel</i>
ETTJ	Estrutura a Termo da Taxa de Juros
FSN-ECM	<i>Functional Signal Plus Noise - Equilibrium Correction Model</i>
IPCA	Índice Nacional de Preços ao Consumidor Amplo
LSTM	<i>Long Short Term Memory</i>
MAE	<i>Mean Absolute Error</i>
MLP	<i>Multilayer Perceptron</i>
MLSTM	<i>Multivariate Long Short Term Memory</i>
NTN-B	Notas do Tesouro Nacional - Série B
NS	Nelson-Siegel
NSS	Nelson-Siegel-Svensson

PCA-NN	<i>Principal Component Analysis for Neural Network</i>
RNN	<i>Recurrent Neural Network</i>
RMSE	<i>Root Mean Squared Error</i>
S&P 500	<i>Standard & Poor's 500 Index</i>
TANH	Tangente Hiperbólica

Capítulo 1

Introdução

1.1 Contexto

A curva de juros prefixada, também conhecida como Estrutura a Termo da Taxa de Juros (ET TJ) Pré, é composta por títulos públicos federais que pagam uma determinada taxa no momento da contratação, para diferentes horizontes de vencimento. Dessa forma, a curva de um determinado dia demonstra, sob a expectativa do mercado, a relação entre as taxas de juros e os prazos de maturidade. Os títulos públicos usados como referência são do Governo Federal e a responsabilidade de suas emissões é do Tesouro Nacional, garantindo assim, um risco de crédito similar.

A avaliação da curva de juros prefixada, e sua representação gráfica, é de fundamental importância para o entendimento das expectativas de investimentos no país, assim como, na percepção que o mercado tem quanto aos riscos inerentes à estes investimentos e da política monetária adotada no país. Neste sentido, a projeção da curva de juros é um importante balizador para os agentes do mercado financeiro e para os formuladores de política econômica no que diz respeito a diferentes abordagens, sejam elas macroeconômica, institucional, política, financeira ou ainda, uma combinação destas.

Dentre os seus formatos mais usuais destacam-se: i) curva crescente, sinalizando que os

juros sobem quanto mais longo for o prazo de vencimento, indicando uma economia, aparentemente, mais saudável; ii) curva invertida, com juros de curto prazo mais altos que os de longo, podendo ser uma evidência de recessão naquela economia; e iii) plana, com juros em patamares muito próximos para os diferentes prazos observados, sinalizando um período de pouca previsibilidade para aquela economia ou, ainda, um período de transição.

No Brasil, o cálculo e divulgação dos dados diários da ETTJ (Pré, IPCA e Implícita) são de responsabilidade da ANBIMA (Associação Brasileira de Entidades dos Mercados Financeiros e de Capitais), criada em 2009 a partir da junção das entidades ANBID (Associação Nacional dos Bancos de Investimento) e ANDIMA (Associação Nacional das Instituições do Mercado Aberto). Portanto, a ANBIMA atua como a principal referência para a consulta financeira desses dados em nível nacional.

A busca por entender o comportamento do mercado financeiro e as projeções financeiras tem sido um constante objetivo de modelagens em pesquisas de mercado e trabalhos acadêmicos. Com os avanços computacionais e de inteligência artificial o tema passou a ser tratado com uma gama maior de ferramentas incluindo modelos de aprendizado de máquina, porém, sem deixar de considerar os modelos estatísticos tradicionais como referência. Entretanto, em que pese uma maior quantidade de modelos e dados disponíveis, muitos trabalhos focam suas análises no segmento de renda variável, mercado futuro e, mais recentemente, criptoativos, podendo o segmento de renda fixa ser mais explorado visando novas implementações.

No aspecto metodológico da construção da curva de juros, o trabalho de Nelson e Siegel, 1987 foi um dos principais pilares para o avanço no tema, assim como, as contribuições posteriores de Svensson, 1994. Tempos depois, Diebold e Li, 2006 fazem uma reinterpretação da curva de juros de maneira dinâmica, contribuindo para a literatura de forma significativa ao interpretarem os fatores de curto, médio e longo prazo como inclinação, curvatura e nível, respectivamente.

Vale salientar que, apesar dos autores mencionados servirem de referência metodológica para este trabalho, estudos anteriores já investigavam a dinâmica dos fatores da estrutura a

termo da curva de juros com outras abordagens, entre eles destacam-se Vasicek, 1977, Cox, Ingersoll, Ross et al., 1985 e Campbell, 1986.

No âmbito das previsões das taxas de juros a literatura é ampla e variada. Cajueiro, Divino e Tabak, 2009 utilizam os modelos *Functional Signal Plus Noise - Equilibrium Correction Model* (FSN-ECM) e o modelo de Diebold e Li, 2006 (DL) para fazer a previsão de 12 taxas de juros no Brasil, com o horizonte temporal de 1, 3, 6 e 12 meses. Por sua vez, em Caldeira, Moura e Savino Portugal, 2010 o modelo DL é estimado pela máxima verossimilhança com filtro de Kalman no contexto de previsão da curva de juros brasileira. O trabalho de Fernandes, Nunes e Reis, 2021 demonstra a dinâmica de nível, inclinação e curvatura da taxa de juros nominal no Brasil, adotando apenas indicadores macroeconômicos com o uso do modelo de Nelson-Siegel.

No estudo de taxas pré-fixadas brasileiras, Faria, 2024 busca avaliar a estrutura a termo da taxa de juros, construída a partir dos títulos públicos federais pré-fixados do Brasil com uma metodologia não paramétrica de regressão *ridge* por *kernel* desenvolvida em Filipovic, Pelger e Ye, 2022. Com essa proposta, o trabalho analisa a viabilidade da utilização do modelo em comparação ao modelo de Svensson, 1994, usado como referência metodológica para a ANBIMA. Arantes, Portugal e Caldeira, 2014 também propõem estudo em relação a estrutura a termo da taxa de juros brasileira, para isso, investigam a capacidade preditiva dos modelos de redes neurais em comparação com os modelos tradicionais de previsão para curva de juros.

Em relação ao aspecto da utilização de redes neurais, Urolagin, Sharma e Datta, 2021 adotam a arquitetura LSTM Multivariada (MLSTM) em conjunto com as transformações de Mahalanobis e *Z-score* para avaliar as previsões do preço do petróleo no mercado futuro. Fischer e Krauss, 2018 usam a arquitetura LSTM para previsões fora da amostra de ações que constituíram o S&P 500 durante o período de 1992 até 2015 e comparam o desempenho preditivo com modelos de aprendizado de máquina.

No quesito de análises da curva de juros usando a abordagem de aprendizado de máquina, Suimon et al., 2020 constroem um modelo de três fatores para representar a curva de juros japonesa capaz de identificar os fatores como nível, inclinação e curvatura. Em Sima, 2022 o

trabalho é conduzido com a presença de um algoritmo de rede neural com componentes principais (*Principal Component Neural Network*, PCA-NN) na detecção de anomalia no mercado financeiro, em especial, na estrutura a termo de taxa de juros dos bancos europeus.

Portanto, apesar das pesquisas da curva de juros serem amplamente realizadas ao longo da história, tanto em âmbito nacional quanto internacional, a sua estimação, projeção e interpretação enfrentam importantes desafios que oferecem complexidade aos resultados encontrados. Em um contexto brasileiro, fatores de ordem política exercem grande influência no mercado de juros, com notícias relevantes sendo divulgadas na mídia impactando diretamente na perspectiva do mercado quanto ao futuro. Mais ainda, sinalizações do governo quanto ao cumprimento da política monetária e fiscal também podem ser de grande relevância para o investidor. Esses fatores, dentre diversos outros, conferem aos dados uma volatilidade significativa, impactando diretamente nas modelagens.

1.2 Motivação

A análise da curva de juros possibilita avaliar a presença de anomalias nas taxas praticadas no mercado de modo a identificar na estrutura temporal dos dados, pontos que se diferenciem do padrão capturado por modelos. As análises de anomalias ganharam mais evidência com o avanço dos modelos de redes neurais, em que, uma maior quantidade de dados não lineares puderam ser interpretados com um ganho de desempenho computacional e qualidade preditiva.

No âmbito das projeções é relevante analisar a capacidade de criação de mecanismos que evidenciem uma mudança de comportamento, de acordo com métricas pré estabelecidas e com certa previsibilidade. Neste contexto, os trabalhos que abordam o desenvolvimento de sistemas de alertas antecipados também são de grande valia.

Os modelos com estas características possuem uma vasta literatura e aplicabilidade em temas envolvendo questões climáticas e análise de falência de bancos e empresas, como em Rosa e Gartner, 2018, porém, ainda não são frequentes no estudo de anomalias para outros segmentos

do mercado financeiro, em especial, para renda fixa.

O presente estudo contribui com esse tópico que ainda pode ser mais discutido no caso brasileiro, que diz respeito à identificação prévia de sinais de alerta em relação à mudança de comportamento na curva de juros local, com a utilização de um modelo de redes neurais e outro estatístico tradicional. Ainda que alguns estudos discutam a identificação de anomalias, a maioria tem foco no segmento de renda variável com a observação de pontos históricos de distorção, o que leva a uma interpretação limitada dos casos.

1.3 Objetivo

O objetivo central deste trabalho é a identificação de sinais de alerta antecipado na estrutura a termo da curva de juros prefixada, avaliando sob a ótica do modelo LSTM, a capacidade de percepção prévia a uma mudança de comportamento na dinâmica dos juros. Comparativamente, é usado o modelo tradicional estatístico ARIMA com o mesmo intuito de identificação e, ainda, contrastar os desempenhos sinalizados.

Para o cumprimento deste objetivo, o estudo desenvolve uma metodologia própria de alerta de sinais baseado na capacidade preditiva do modelo, considerando um horizonte de até 4 semanas e janelas expandidas de tempo com um histórico fixo de datas, para simular o real desempenho frente aos dados que aparecem em cada período. De tal modo, o método conhecido como validação cruzada e amplamente discutido em Hyndman et al., 2025, permite avaliar como a metodologia proposta se comporta a medida que os dados vão sendo revelados.

A divisão temporal dos dados é similar a realizada em Jacomo Seraphim Nogueira, 2026 que utiliza janelas expandidas para realizar uma análise comparativa entre diversas técnicas estatísticas, aprendizado de máquina e aprendizado profundo para a previsão da despesa primária do governo central. Conforme destaca o trabalho, apesar de produzir erros maiores comparativamente à ajustes simples, esse método consegue aproximar mais de uma previsão real, dado que, as novas informações somente são incorporadas com o passar dos períodos.

As pesquisas em redes neurais possuem uma ampla literatura a respeito das diferentes aplicações e interpretações com diversas bases de dados e estruturas temporais. No quesito de escolha de modelos, a família LSTM é notada por possuir uma boa capacidade de armazenamento de memória tanto de curto prazo quanto de longo prazo e foi proposta originalmente em Hochreiter e Schmidhuber, 1997. Assim, é relevante avaliar o comportamento deste modelo na capacidade preditiva das maturidades escolhidas para a curva de juros no Brasil.

Os resultados obtidos contribuem para o debate sobre a capacidade de identificar sinais de uma mudança na curva de juros, permitindo algum tipo de sinalização aos agentes de mercado, em um contexto marcado por grandes desafios preditivos de diferentes características. Ainda, é possível comparar o desempenho da metodologia proposta com um modelo tradicionalmente usado em abordagens estatísticas.

1.4 Esboço do Trabalho

Este estudo possui além desse capítulo introdutório, o Capítulo 2 responsável por descrever o referencial teórico usado no decorrer do trabalho, o Capítulo 3 em que é apresentada a metodologia do estudo incluindo as discussões de base de dados, tratamento das variáveis e definição do modelo, o Capítulo 4 com uma ampla discussão dos resultados encontrados e, por fim, o Capítulo 5 responsável pela conclusão e consolidação dos tópicos apresentados.

Capítulo 2

Referencial Teórico

2.1 Modelagem da Curva de Juros

Essa seção apresenta os principais estudos que serviram como referencial teórico para a construção dos modelos de curva de juros com ajustes paramétricos. Neste sentido, a Seção 2.1.1 aborda o modelo de Nelson-Siegel-Svensson, fundamental na compreensão da estrutura da curva de juros utilizada, destacando tanto o desenvolvimento do modelo de Nelson-Siegel (NS), quanto a contribuição realizada por Svensson no estudo de curva de juros. Na Seção 2.1.2, descrevemos o modelo proposto em Diebold e Li, 2006, que amplia a discussão da estrutura a termo da curva de juros ao propor um modelo autorregressivo para a previsão dos fatores da curva. A Seção 2.1.3, por sua vez, descreve a metodologia da ANBIMA, utilizada como referência na estimação da estrutura a termo da curva de juros no Brasil.

2.1.1 Modelo de Nelson-Siegel-Svensson

O trabalho de Nelson e Siegel, 1987 desenvolve um modelo paramétrico de modo a representar de forma parcimoniosa e relativamente simples, as curvas de juros com diferentes comportamentos e formatos, sendo elas, em geral, monotônicas, com curvaturas e em forma de "S". O resultado apresentado pelos autores demonstra uma alta correlação entre as curvas de juros pre-

vistas pelo modelo e as taxas de juros de longo prazo de títulos públicos americanos, reforçando a ideia de que o modelo é eficiente na identificação de características que relacionem taxas de juros com prazos de maturidade. Os resultados foram encontrados a partir da formulação da função de taxa a termo exibida a seguir.

$$f(m) = \beta_0 + \beta_1 e^{-m/\tau_1} + \beta_2 \frac{m}{\tau_1} e^{-m/\tau_1} \quad (2.1)$$

Na equação, m indica o prazo e $\tau_1 > 0$ é o parâmetro de decaimento. A expressão mostra que $f(m)$ combina um termo constante (β_0), um termo exponencial decrescente ($\beta_1 e^{-m/\tau_1}$) e um termo que cresce com m mas é amortecido exponencialmente ($\beta_2 \frac{m}{\tau_1} e^{-m/\tau_1}$).

Conforme destacado em Faria, 2024, a função paramétrica desenvolvida por Nelson-Siegel permite interpolar a estrutura a termo da curva de juros, identificando diferentes formas que a curva pode assumir no decorrer dos anos. O modelo, que utiliza primordialmente três fatores, tem a forma funcional da taxa à vista dada por:

$$y(t, \tau) = \beta_0 + \beta_1 \left(\frac{1 - e^{-\frac{\tau}{\lambda}}}{\frac{\tau}{\lambda}} \right) + \beta_2 \left(\frac{1 - e^{-\frac{\tau}{\lambda}}}{\frac{\tau}{\lambda}} - e^{-\frac{\tau}{\lambda}} \right) \quad (2.2)$$

sujeito às seguintes restrições:

$$\beta_{0,t} > 0, \quad \beta_{0,t} + \beta_{1,t} > 0, \quad \lambda_t > 0 \quad (2.3)$$

Posteriormente, em Diebold e Li, 2006, a equação foi reinterpretada de forma dinâmica para avaliação do comportamento da curva de juros ao longo do tempo, onde os autores fornecem uma importante contribuição ao concluir que os fatores de longo, curto e médio prazo podem ser interpretados como fatores de nível, inclinação e curvatura, onde:

- $y(t, \tau)$ é o rendimento no tempo t para um título com maturidade τ ;
- β_0 representa o fator de nível, associado ao componente de longo prazo da curva;

- β_1 representa o fator de inclinação, que controla o declive da curva entre o curto e o longo prazo;
- β_2 representa o fator de curvatura, responsável por capturar a forma em “sino” da curva de juros;
- λ é o parâmetro de decaimento exponencial, que determina a velocidade de convergência da função.

A Equação 2.2 passou a ser utilizada como referência para estimar a curva de juros observada, ajustando os parâmetros β_0 , β_1 , β_2 e λ de forma a minimizar o erro quadrático entre o rendimento modelado e o rendimento observado.

Assim, diversos outros estudos passaram a utilizar o modelo de Nelson-Siegel como referência para novas implementações e abordagens. Dentre os trabalhos seguintes, Svensson, 1994 adiciona um novo termo ao modelo originalmente proposto, o parâmetro λ_2 . A inclusão flexibiliza o modelo ao permitir a captura de uma segunda curvatura na curva de juros com intuito de encontrar uma melhor adequação do modelo em diferentes instantes de tempo, incluindo períodos de curto prazo e longo prazo.

A extensão do modelo de Nelson-Siegel proposta por Svensson, 1994 e estimada com quatro fatores, é conhecida como modelo de Nelson-Siegel-Svensson (NSS) dado por:

$$y(t, \tau) = \beta_{0,t} + \beta_{1,t} \left(\frac{1 - e^{-\frac{\tau}{\lambda_{1,t}}}}{\frac{\tau}{\lambda_{1,t}}} \right) + \beta_{2,t} \left(\frac{1 - e^{-\frac{\tau}{\lambda_{1,t}}} - e^{-\frac{\tau}{\lambda_{2,t}}}}{\frac{\tau}{\lambda_{1,t}}} \right) + \beta_{3,t} \left(\frac{1 - e^{-\frac{\tau}{\lambda_{2,t}}} - e^{-\frac{\tau}{\lambda_{1,t}}}}{\frac{\tau}{\lambda_{2,t}}} \right) \quad (2.4)$$

obedecendo às restrições:

$$\beta_{0,t} > 0, \quad \beta_{0,t} + \beta_{1,t} > 0, \quad \lambda_{1,t} > 0, \quad \lambda_{2,t} > 0 \quad (2.5)$$

onde além dos coeficientes $\beta_{0,t}$, $\beta_{1,t}$ e $\beta_{2,t}$ já estimados no modelo NS, temos:

- $\beta_{3,t}$ representa o segundo fator de curvatura, que permite capturar formatos mais complexos da curva;
- $\lambda_{1,t}$ e $\lambda_{2,t}$ são parâmetros de decaimento exponencial, controlando a localização dos picos de curvatura ao longo dos prazos.

Logo, percebe-se que a Equação 2.4 se reduz ao modelo original de Nelson-Siegel quando $\beta_{3,t} = 0$ e $\lambda_{2,t} = \lambda_{1,t}$. Assim como no modelo original, os parâmetros $\beta_{0,t}$, $\beta_{1,t}$, $\beta_{2,t}$, $\beta_{3,t}$, $\lambda_{1,t}$ e $\lambda_{2,t}$ são estimados de modo a minimizar o erro quadrático entre os rendimentos observados e os modelados. A flexibilidade adicional introduzida por $\beta_{3,t}$ e $\lambda_{2,t}$ permite representar curvas de juros com múltiplos pontos de inflexão, sendo amplamente utilizada por instituições do sistema financeiro, dentre elas a ANBIMA conforme será detalhado adiante.

2.1.2 Modelo de Diebold-Li

A partir do modelo NSS (Nelson e Siegel, 1987 e Svensson, 1994), uma das contribuições mais significativas para o estudo da estrutura a termo da curva de juros foi desenvolvido por Diebold e Li, 2006. No trabalho apresentado, os autores fazem uma valiosa contribuição para a reinterpretação dos parâmetros do modelo Nelson-Siegel, introduzindo dinamicidade ao modelo original ao permitir que os parâmetros estimados variassem ao longo do tempo (*Dynamic Nelson-Siegel, DNS*). Neste sentido, é apresentado, com mais detalhes do que na Seção 2.1.1, como o conjunto de coeficientes estimados contribui para o entendimento da estrutura da curva de juros.

- $\beta_{0,t}$ é uma constante que independe do fator tempo e não apresenta decaimento mesmo no limite, ou seja, $y_t(\infty) = \beta_{0,t}$. Portanto, é associado ao componente de longo prazo da curva. Mais ainda, os autores destacam que qualquer variação em $\beta_{0,t}$ altera igualmente em todas as maturidades o rendimento na curva de juros, uma vez que, o parâmetro é igual em todo o horizonte de tempo, sendo então responsável pelo nível da curva de juros estimada;

- $\beta_{1,t}$ é acompanhado do fator $\left(\frac{1-e^{-\frac{\tau}{\lambda}}}{\frac{\tau}{\lambda}}\right)$ que começa em 1 mas tem um decaimento monotônico e acelerado até 0, sendo interpretado como um componente de curto prazo. O coeficiente $\beta_{1,t}$ é então caracterizado como o componente de inclinação da curva, dado que, as taxas de curto prazo carregam um maior peso de $\beta_{1,t}$ e, portanto, uma variação neste coeficiente altera de forma mais significativa a parte curta da curva comparada com a parte longa. Em Diebold e Li, 2006 uma aproximação é calculada como a diferença entre os rendimentos de 10 anos (em meses) e de 3 meses, isto é, $y_t(120) - y_t(3)$;
- $\beta_{2,t}$ é seguido do fator $\left(\frac{1-e^{-\frac{\tau}{\lambda}}}{\frac{\tau}{\lambda}} - e^{-\frac{\tau}{\lambda}}\right)$ que inicia em 0, aumenta, e retorna para 0, sendo compreendido como um fator de médio prazo. O coeficiente $\beta_{2,t}$ é interpretado como responsável pela curvatura da curva de juros. Nele, uma variação em $\beta_{2,t}$ tem pouco efeito em maturidades curtas e longas, porém afeta mais significativamente nas variações na parte intermediária da curva que carrega um peso maior do coeficiente. Os autores definem a aproximação desse componente como sendo duas vezes a diferença do rendimento de 2 anos (em meses) menos os rendimentos de 3 meses e de 10 anos (em meses), ou seja, $2y_t(24) - y_t(3) - y_t(120)$.

Os resultados encontrados demonstram que a curva média é tipicamente crescente e côncava, a dinâmica das taxas de juros é persistente mais no longo prazo do que no curto prazo, e que a volatilidade das taxas é maior no curto prazo em comparação com as de maior prazo. Com base nessas conclusões relevantes, Diebold e Li, 2006 estimam os coeficientes para diferentes datas e aplicam o modelo autorregressivo AR(1), conectando os estimadores ao longo do tempo. Esta abordagem, utilizando apenas os três fatores propostos em Nelson e Siegel, 1987 e obtidos a partir de um fator de decaimento λ fixo, permitiu inferir o comportamento da curva de juros de forma dinâmica e com bons resultados quando comparado com outros modelos estatísticos, principalmente para maturidades maiores. O modelo AR(1) é projetado como um processo univariado, conforme a seguir.

$$\hat{y}_{t+h|t}(\tau) = \hat{\beta}_{0,t+h|t} + \hat{\beta}_{1,t+h|t} \left(\frac{1 - e^{-\frac{\tau}{\lambda}}}{\frac{\tau}{\lambda}} \right) + \hat{\beta}_{2,t+h|t} \left(\frac{1 - e^{-\frac{\tau}{\lambda}}}{\frac{\tau}{\lambda}} - e^{-\frac{\tau}{\lambda}} \right) \quad (2.6)$$

onde, $\hat{\beta}_{0,t+h|t} = \hat{c}_i + \hat{\gamma}_i \hat{\beta}_{i,t}$, para $i = 0, 1, 2$, e $\hat{c}_i$ e $\hat{\gamma}_i$ são obtidos a partir da regressão de $\hat{\beta}_{i,t}$ no intercepto e $\hat{\beta}_{i,t-h}$.

2.1.3 Metodologia ANBIMA

A ANBIMA, Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais, foi criada em 2009 a partir da junção de duas outras entidades, a ANBID (Associação Nacional dos Bancos de Investimento) e a ANDIMA (Associação Nacional das Instituições do Mercado Aberto), e tem entre um dos seus principais objetivos a análise e distribuição de dados no mercado financeiro brasileiro, possuindo uma vasta base de dados do mercado financeiro e de capitais e, portanto, sendo uma das referências diárias para consulta de informações financeiras.

Dentre os dados disponibilizados diariamente pela ANBIMA, este trabalho foca especificamente na divulgação das informações da Estrutura a Termo da Curva de Juros (ETTJ), em que a entidade utiliza uma metodologia específica para divulgar os parâmetros estimados da curva $\beta_1, \beta_2, \beta_3, \beta_4, \lambda_1$ e λ_2 , assim como os valores da ETTJ IPCA, ETTJ Pré e Inflação Implícita para diferentes horizontes de tempo. Na sequência é descrita de forma sucinta a metodologia utilizada pela ANBIMA para a divulgação diária dos dados da ETTJ Pré, objetivo de análise deste estudo.¹

Inicialmente, para a formulação da ETTJ é adotada a premissa básica de que o preço de um título de renda fixa é equivalente ao fluxo de caixa do ativo, trazido a valor presente por uma função de desconto, em que o risco de crédito do Governo Federal (emissor do título) já está embutido nos juros. A fórmula para o cálculo é a seguinte:

¹A metodologia completa utilizada pela ANBIMA pode ser consultada no link: https://www.anbima.com.br/pt_br/informar/curvas-de-juros-fechamento.htm. Acesso em outubro de 2025.

$$P_{i,t} = \sum_{j=1}^{K_i} F_{i,j,t} b_t(T_{i,j}) + \varepsilon_{i,t} \quad (2.7)$$

- $P_{i,t}$ é o preço do título i na data t ;
- $F_{i,j,t}$ é o j -ésimo fluxo de pagamento, cupom ou amortização, do título i , na data t ;
- $T_{i,j}$ é o prazo, em anos (base 252 dias úteis), do pagamento j ;
- K_i é o número de pagamentos do título i ;
- $b_t(T_{i,j})$ é a função de desconto discreta para prazo $T_{i,j}$, na data t ; e
- $\varepsilon_{i,t}$ é o erro do modelo para o título i em t .

Por sua vez, a função de desconto é tal que:

$$b(T_{i,j})_t = \frac{1}{(1 + r_t(T_{i,j}))^{T_{i,j}}} \quad (2.8)$$

onde, $r_t(T_{i,j})$ é a taxa por prazo utilizada.

Em seguida, a metodologia da ANBIMA propõe a utilização do modelo NSS para a obtenção da taxa de juros $r_t(\tau)$, na data t e prazo τ . As vantagens deste modelo, o comportamento dos seus componentes e as interpretações derivadas do mesmo foram discutidas nas Seções 2.1.1 e 2.1.2.

Adicionalmente, é incluído um ponderador para W_i , calculado como o inverso da *duration*, para penalizar mais os erros de curto prazo, dado que, pequenas variações nos preços de ativos de curta duração implicam em maiores oscilações em suas taxas. Isso ocorre como resultado da diferença de sensibilidade de variações de taxas e de preço para ativos com diferentes maturidades.

$$\min \sum_{i=1}^N W_i \left(P_{i,t} - \sum_{j=1}^{K_i} F_{i,j,t} b_t(T_{i,j}) \right)^2 \quad (2.9)$$

Devida a complexidade do modelo e a sua não linearidade, a metodologia da ANBIMA destaca a utilização de um algoritmo genético para estimação dos parâmetros do modelo NSS. O algoritmo tem por objetivo encontrar candidatos a soluções ótimas para o problema e que, ao final das iterações, convirja para a solução ótima. Com isso os resultados tendem a ter uma menor volatilidade e melhor aderência aos dados, reduzindo o risco de encontrar um mínimo que seja local e não global, ou seja, uma falsa convergência. A partir daí, os parâmetros β podem ser refinados por uma estimação tradicional, uma vez que, o algoritmo genético já definiu uma região possível de otimização. A ETTJ, divulgada diariamente, é então obtida a partir da estimação dos 4 parâmetros β e dos 2 parâmetros de decaimento λ_1 e λ_2 , totalizando 6 parâmetros.

2.2 Modelos de Redes Neurais em Aprendizado Profundo

Nesta seção será abordado o modelo de redes neurais utilizado para a construção do sistema de alertas baseado em detecção de anomalias na ETTJ. Apesar da vasta gama de arquiteturas desenvolvidas e aprimoradas no âmbito do aprendizado de máquinas, em especial, de aprendizado profundo, esta seção terá enfoque no modelo LSTM (*Long Short Term Memory*) usado como modelo principal neste trabalho.

Para tanto, na Seção 2.2.2 são discutidos os principais componentes do modelo LSTM, desenvolvido por Hochreiter e Schmidhuber, 1997. Ressalta-se que essa seção não tem por objetivo exaurir os diversos e complexos tópicos de aprendizado profundo, sendo recomendados os trabalhos de Prince, 2023, Goodfellow, Bengio e Courville, 2016 e Paiva e Pereira, 2022 para uma abordagem mais ampla.

Ademais, com objetivo de formar uma base teórica para o modelo LSTM é exibida previamente, na Seção 2.2.1, a estrutura básica de uma rede neural recorrente (*Recurrent Neural Network, RNN*).

2.2.1 Redes Neurais Recorrentes

O modelo LSTM, e suas variações, são uma arquitetura do tipo rede neural recorrente (RNN) e, portanto, faz-se necessária uma breve explicação sobre o conceito e funcionamento dessa rede, usada em diversas aplicações como reconhecimento de padrões em texto, genomas, dados sequenciais e análise de imagens.

A partir do trabalho de Elman, 1990, as RNNs passaram a ter como principal diferenciação em relação às redes *feedforward* a capacidade de conectar a suas próprias saídas ao estado oculto da célula, funcionando como uma memória. Isto é, as redes recorrentes conseguem reutilizar informações da própria sequência ao contrário das redes *feedforward*. Ainda, as RNNs adotam o mesmo parâmetro de peso nas suas camadas, sendo ajustados pelos processos de retropropagação e gradiente descendente com intuito de otimizar o aprendizado por reforço e reduzindo os erros em previsões futuras.

Conforme exposto em Goodfellow, Bengio e Courville, 2016 a memória adicionada às redes neurais é preservada no estado oculto da rede, passando a informação reservada para outras etapas, caracterizando uma dependência temporal dos acontecimentos, pois, um determinado evento no tempo é uma função que depende de um ou mais eventos que vieram antes. A Figura 2.1, apresentada em Olah, 2015 auxilia no entendimento da arquitetura.

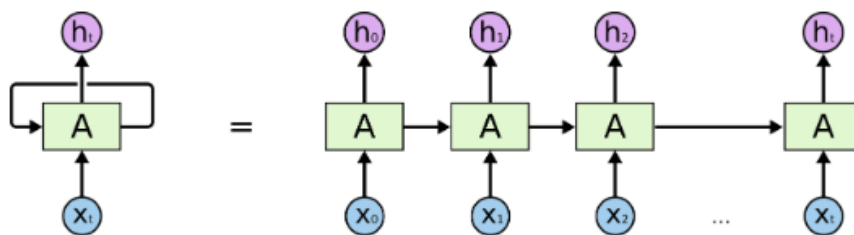


Figura 2.1: Estrutura interna de uma arquitetura RNN. Fonte: Olah, 2015.

As RNNs possuem também 3 tipos de estruturas, sendo elas: i) um para muitos (uma única entrada com muitas sequências de saída); ii) muitos para um (muitas entradas com uma sequência de saída); e iii) muitos para muitos (muitas entradas com muitas sequências de saída).

Entretanto, em função do desempenho preditivo destes modelos estar intrinsecamente ligado à distância entre os dados relevantes e o momento em que se deseja aplicar a predição, a medida que o intervalo de distância cresce, maior é a dificuldade em aprender as dependências de maior alcance temporal.

Em termos algébricos, este problema está ligado à dissipação do gradiente (*Vanishing Gradient Problem*) ou à explosão do gradiente (*Exploding Gradient Problem*) das RNNs. Conforme explicitado em Paiva e Pereira, 2022, essa limitação dificulta o treinamento da rede, uma vez que, durante a fase de retropropagação do gradiente ocorre um alto número de multiplicações entre o sinal do gradiente e a matriz de peso da camada recorrente. Assim, os valores da multiplicação vão tender a zero ou a infinito, sendo prejudicial, em ambos os casos, para o desempenho da rede.

Portanto, as limitações exibidas restringem a aplicação da arquitetura em situações de curtíssimo prazo, uma vez que, o gradiente é o valor utilizado para atualizar os pesos da rede neural e, se este valor tende a zero ou a um número muito alto, então não há uma contribuição significativa para o aprendizado da rede, reduzindo a capacidade de retenção da memória em períodos maiores.

2.2.2 Modelo LSTM

A estrutura LSTM, por sua vez, é caracterizada por conseguir capturar a memória de curto e longo prazo e, assim, há um maior fluxo de informações percorrendo pelas camadas do modelo permitindo, no geral, um melhor poder preditivo. Neste sentido, o principal objetivo da estrutura é entender como as memórias se relacionam entre si, isto é, qual memória do curto prazo será lembrada e qual memória de longo prazo irá continuar influenciando o momento atual.

A utilização dos portões (*gates*) são essenciais para esta atividade, controlando a informação que é transmitida entre os canais. Os 3 portões que compõem a estrutura LSTM são expostos a seguir com suas definições e implementações seguindo o trabalho de Olah, 2015. Mais a frente, a relação das camadas da célula é detalhada para um determinado período de tempo t .

O primeiro portão a ser discutido é o *forget gate*, responsável por decidir quais informações da célula permanecem relevantes, removendo as que não são mais consideradas úteis. Esse processo inicia-se por meio de dois valores de entrada, sendo eles x_t , entrada no momento em questão e h_{t-1} , saída da célula no momento anterior. Estes valores são então multiplicados por matrizes de peso, com a adição do viés na sequência. O resultado desta operação é passado por uma função de ativação com saída binária. Caso a saída seja 0 para um determinado estado de célula, a informação é esquecida, e caso seja 1 a informação é mantida para uso futuro.

A segunda etapa é realizada utilizando o *input gate*, portão responsável por definir quais dados da memória de curto prazo devem ser adicionados a célula. Nele a informação é controlada por meio da função sigmoide que filtra os valores a serem lembrados, utilizando as entradas x_t e h_{t-1} . Em seguida, um vetor é gerado utilizando a função *tanh* que fornece as saídas de -1 a +1, contendo todos os valores possíveis de h_{t-1} e x_t . A partir deste conjunto de dados, os valores do vetor e dos valores controlados são multiplicados para se obter as informações úteis do modelo.

O terceiro portão é denominado *output gate* e é responsável por decidir quais partes da célula são determinantes no estado atual para gerar o resultado. O processo retira as informações que são úteis do estado da célula atual e as apresenta como uma saída. Inicialmente, é gerado um vetor a partir da função *tanh* na célula. Em seguida, a informação é controlada usando a função sigmoide, responsável por filtrar os valores a serem lembrados com as entradas h_{t-1} e x_t . Os valores do vetor e os dados regulados são multiplicados para serem encaminhados como uma saída e entrada para a próxima célula.

Após o entendimento do funcionamento dos portões da estrutura LSTM, é importante detalhar como ocorre a interação entre as etapas presentes e apresentar as funções de ativação. Para isso, novamente recorre-se a explicação desenvolvida em Olah, 2015, assim como, a imagem explicativa para o caso geral de um modelo LSTM.

A célula da estrutura tem como primeiro passo a definição da informação que será mantida para a sequência do modelo e quais informações serão descartadas. Essa definição tem como

ponto central a função sigmóide presente no *forget gate*.

A função sigmóide é uma função matemática que converte qualquer valor real para uma escala entre 0 e 1, possuindo um formato em "S". A capacidade de comprimir valores para a escala mencionada permite sua ampla utilização em redes neurais como função de ativação, ao transformar a combinação linear dos dados em atributos percentuais e, por isso, muito utilizada em modelos de classificação. A fórmula mais usual é dada por:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.10)$$

Nesta etapa, a função recebe os valores da célula passada h_{t-1} e os valores da célula atual x_t , tendo como resultado um valor no intervalo entre 0 e 1, onde 0 representa esquecer a informação analisada e 1 representa lembrar toda a informação. Em conjunto às informações existe um vetor de pesos W_f e um viés b_f .

Na sequência é avaliado qual nova informação será retida pela célula. Este passo ocorre no *input gate* e é dividido em duas partes. A primeira é uma função sigmóide responsável por decidir qual informação do estado atual será atualizada, de forma similar ao caso descrito anteriormente. Já a segunda é definida por meio da função de ativação *tanh* (tangente hiperbólica), com interpretação muito similar à função sigmóide, com a diferença de mapear valores para a escala entre -1 e 1 e auxiliar na convergência durante a fase de treinamento dos dados, sendo capaz de introduzir a não-linearidade no modelo e com uma função dada por:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.11)$$

Na segunda parte é criado um vetor de novos valores candidatos (C_t) tal que pode ser adicionado ao novo estado da célula. A partir do resultado obtido no *forget gate* o estado da antiga célula, C_{t-1} , é atualizado com o que deve ser mantido e, na sequência, é adicionado o vetor da célula atual, C_t . Com isso temos os valores atualizados da célula antiga e da atual.

Por fim, resta definir o resultado da célula. A saída tem como referência o estado da célula,

porém, filtrado. Primeiramente, uma camada com a função sigmóide é rodada para decidir quais informações seguem para o resultado. Posteriormente, a função tangente hiperbólica é aplicada à célula para atribuir valores entre -1 e 1. Os valores entre as duas funções de ativação são multiplicados e apenas as informações relevantes são destinadas ao resultado.

Por se tratar de um estrutura complexa, a Figura 2.2, exibida em Olah, 2015, ajuda a compreender melhor o funcionamento do modelo utilizado como base neste trabalho, onde σ é a representação da função sigmóide.

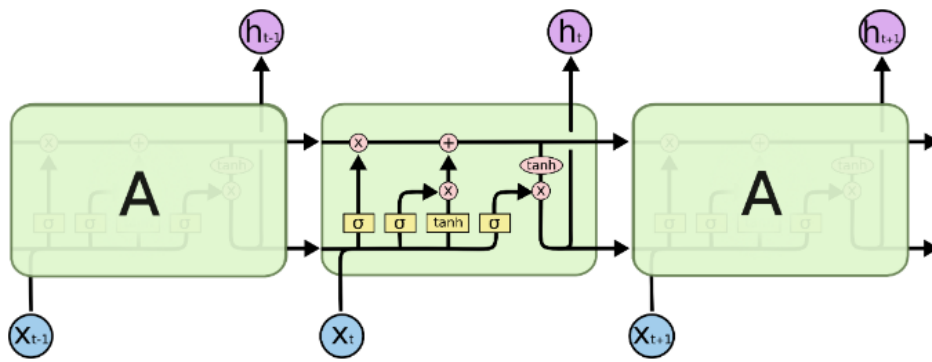


Figura 2.2: Estrutura interna de um modelo LSTM. Fonte: Olah, 2015.

Capítulo 3

Metodologia

3.1 Base de Dados

Para este trabalho, que tem por objetivo desenvolver um sistema de alerta antecipado a partir da identificação de pontos anômalos na curva de juros prefixada, foram analisados os dados da estrutura a termo das taxas de juros prefixadas, com base na divulgação diária das informações pela ANBIMA¹. A metodologia usada pela Associação é amplamente discutida na Seção 2.1.3.

A ANBIMA divulga diariamente, em seu portal eletrônico, as informações dos coeficientes estimados a partir dos dados coletados junto aos agentes do mercado financeiro. Neste sentido, podem ser observados os coeficientes $\beta_1, \beta_2, \beta_3, \beta_4, \lambda_1$ e λ_2 e com base nestes estimadores são obtidas as taxas de juros para diferentes maturidades.

Na publicação dos coeficientes existe uma particularidade em que a ANBIMA mantém disponível para visualização e *download* apenas as informações dos últimos 5 dias úteis. De tal modo, para consultar um período maior de tempo, entrou-se em contato com a Associação para o fornecimento dos coeficientes diários durante o período entre 01/01/2015 e 17/09/2025, contemplando mais de 10 anos de dados. As informações foram gentilmente cedidas para a análise conduzida neste trabalho.

¹Disponível no site da ANBIMA: https://www.anbima.com.br/pt_br/informar/curvas-de-juros-fechamento.htm.

Uma vez obtidos os coeficientes $\beta's$ e $\lambda's$ do período desejado utilizou-se a metodologia de Svensson, 1994, discutida na Seção 2.1.1, e que também é adotada pela ANBIMA, para se obter as taxas diárias das maturidades desejadas.

Em relação aos prazos, foram selecionados os 9 principais vértices para a taxa prefixada usados pela ANBIMA², sendo eles: 21 dias úteis, 42 dias úteis, 63 dias úteis, 126 dias úteis, 252 dias úteis, 504 dias úteis, 756 dias úteis, 1.008 dias úteis e 1.260 dias úteis. Desta forma, estudou-se o comportamento da curva de juros com base nas expectativas do mercado para diferentes horizontes temporais, desde o curto prazo (1 mês, 2 meses e 3 meses) até prazos mais longos de até 5 anos. As séries históricas das taxas de juros, para as 9 maturidades selecionadas, são apresentadas na Figura 3.1.

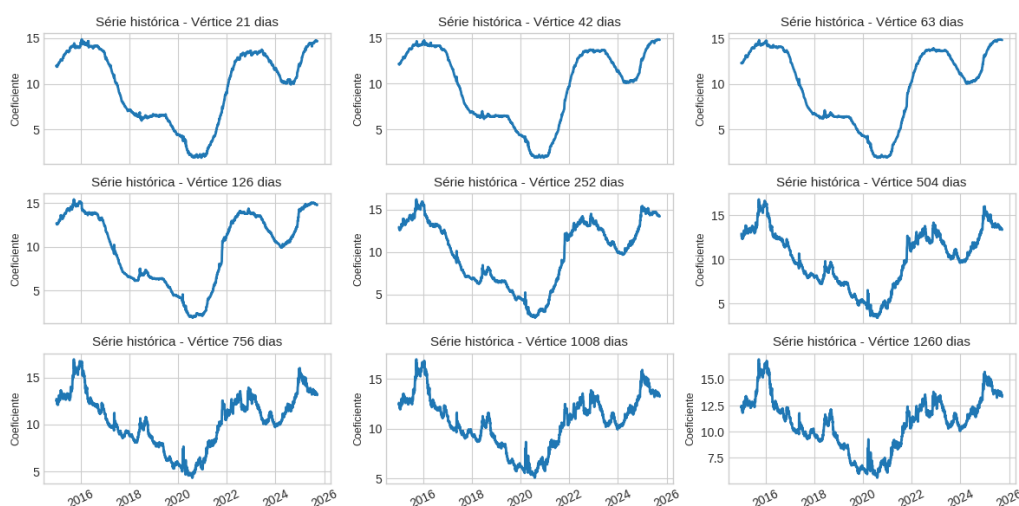


Figura 3.1: Série histórica das taxas de juros prefixadas, por vértice

A imagem elucida um comportamento relativamente parecido entre as maturidades, com ciclos de aumento e redução das taxas em períodos bem definidos, atingindo os maiores valores em 2016 e 2025, enquanto os períodos de baixa mais acentuada foram observados no ano de 2020. Adicionalmente, é apresentada na Tabela 3.1 as estatísticas descritivas dos dados, por maturidade, onde é possível corroborar um comportamento com maior volatilidade nas maturi-

²Circular nº 3.361, elaborado pelo Banco Central do Brasil (BCB)

dades mais curtas, enquanto nas mais longas o comportamento aparenta ser mais suavizado. O desvio-padrão e as taxas mínimas e máximas da tabela retratam esta interpretação.

Tabela 3.1: Estatísticas descritivas por vértice da curva

	21 dias	42 dias	63 dias	126 dias	252 dias	504 dias	756 dias	1008 dias	1260 dias
Obs.	2.686	2.686	2.686	2.686	2.686	2.686	2.686	2.686	2.686
Média	9,6674	9,7133	9,7471	9,8076	9,8871	10,1319	10,4126	10,6486	10,8295
D.P.	4,0139	4,0729	4,1073	4,1001	3,8459	3,2659	2,8953	2,6760	2,5383
Mín.	1,9208	1,9085	1,9010	1,8754	2,2307	3,3720	4,3377	5,0571	5,6032
25%	6,4606	6,4277	6,4010	6,3500	6,6068	7,5877	8,2883	8,8175	9,1770
50%	10,5870	10,6449	10,7103	10,7770	10,7987	10,4289	10,5620	10,8121	11,0747
75%	13,3892	13,5746	13,6693	13,7212	13,2499	12,7163	12,5703	12,5446	12,5421
Máx.	14,8842	14,8536	14,9227	15,4083	16,2190	16,8104	16,9396	16,9500	16,9314

3.2 Tratamento das Variáveis

A partir da série histórica dos dados descrita na seção anterior, foram realizados os tratamentos prévios das variáveis com dois objetivos bem claros: i) reduzir o ruído diário das séries, tornando-as estacionárias; e ii) melhorar o tempo de processamento computacional das informações.

Para cumprir os objetivos propostos, em um primeiro momento foi realizada a média semanal dos dados por maturidade. Cabe ressaltar que utilizar os dados do mercado financeiro em periodicidades que não sejam diárias é uma prática comum na literatura sobre o tema, principalmente quando se analisa mais de uma série temporal, como é o caso deste trabalho. Ao utilizar a média semanal é possível manter o comportamento dos dados em um intervalo relativamente pequeno de dias (no máximo 5 dias úteis, para uma semana sem feriados) e, ainda, reduzir de forma significativa o ruído diário e o número de observações.

Em Filho, 2017 o trabalho investiga os fatores que determinam a taxa de juros real de longo prazo no Brasil, em específico os títulos atrelados à inflação, NTN-B³. Para esse objetivo o

³Apesar de NTN-B (Nota do Tesouro Nacional Série B) ser o antigo nome do título público federal, hoje oficialmente chamado de Tesouro IPCA+, é ainda o nome mais usual no mercado financeiro.

autor usa os dados do título público e das demais variáveis nacionais e internacionais em uma escala de média mensal. Já em Ouyang, Yang e Lai, 2021, os autores utilizam o modelo LSTM para estudar um sistema de alerta de risco sistêmico para o mercado chinês baseado em dados mensais de 51 instituições financeiras.

Portanto, ao transformar os dados para semanais há um ganho claro de processamento computacional, uma vez que, como será visto nas próximas seções, utilizou-se a metodologia de validação cruzada com janelas expandidas, e um histórico fixo de datas, para as 9 maturidades. Em tentativas iniciais, os dados diários demandavam muito processamento para cada vértice, dificultando o andamento da análise desejada.

A Figura 3.2, com as séries históricas semanais, corrobora o argumento descrito, onde é possível perceber um comportamento histórico muito próximo do observado com as séries diárias durante todo o período de análise.

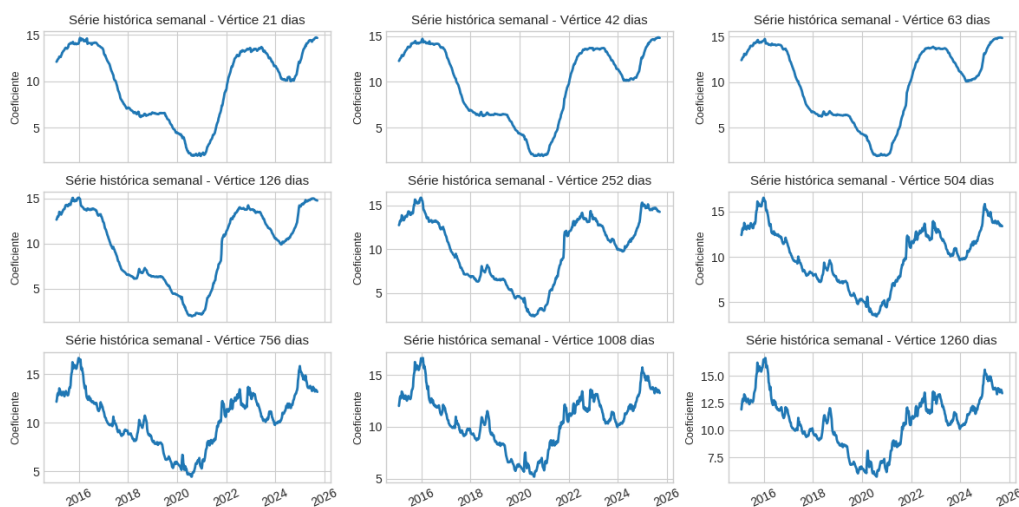


Figura 3.2: Série histórica semanal das taxas de juros prefixadas, por vértice

Na sequência, visando uma melhor performance do modelo, avaliou-se a estacionariedade das séries diárias e semanais a partir do teste Dickey-Fuller Aumentado (ADF) para um nível de significância de 5%. O teste ADF atua como uma versão estendida do modelo Dickey-Fuller simples ao considerar termos adicionais de defasagem da variável de interesse, com intuito de

capturar e eliminar a autocorrelação dos resíduos.

A hipótese nula (H_0) afirma que há raiz unitária na série temporal e, portanto, ela não é estacionária. A hipótese alternativa (H_1) garante o inverso, sendo a série estacionária. O critério de seleção do teste estipula que caso o p-valor seja menor que um determinado nível de significância pré-estabelecido, então há a rejeição da hipótese nula e a série é estacionária. Por outro lado, caso o p-valor seja superior ao nível de significância, a hipótese nula não é rejeitada e a série temporal é tratada como não estacionária.

No caso dos valores diários, todas as maturidades apresentaram um elevado p-valor, indicando a necessidade de diferenciação. Para as séries semanais os p-valores foram em um nível relativamente mais baixo, porém, com a maioria das séries históricas indicando a presença de raiz unitária. Logo, foram obtidas as primeiras diferenças dos vértices estudados.

Após tornar as séries em primeira diferença, as análises subsequentes foram feitas em termos de variação das taxas médias semanais. Novamente, ressalta-se que, com este procedimento não houve perda significativa de informações e, ainda, observou-se um melhor desempenho em termos de processamento e qualidade dos resultados expostos nas próximas seções.

Ademais, em que pese os modelos de redes neurais, como o LSTM, serem capazes de tratar os dados em nível e, serem relativamente eficientes para tratar dados não lineares, a diferenciação dos dados não é um impeditivo para sua utilização. Em trabalhos como o de Yadav, Jha e Sharan, 2020 os autores utilizam o modelo LSTM com diferença para avaliação preditiva do mercado de ações indiano.

A Figura 3.3 ilustra as séries semanais após a diferenciação. Como é possível observar, os dados apresentam um comportamento oscilando próximo de zero e mais estável ao longo do tempo, como é de se esperar de séries após a diferenciação. Entretanto, também é possível visualizar algumas exceções, como no início do ano de 2020 e no final do ano de 2021, fornecendo possíveis evidências de períodos onde há alguma anomalia sistêmica.

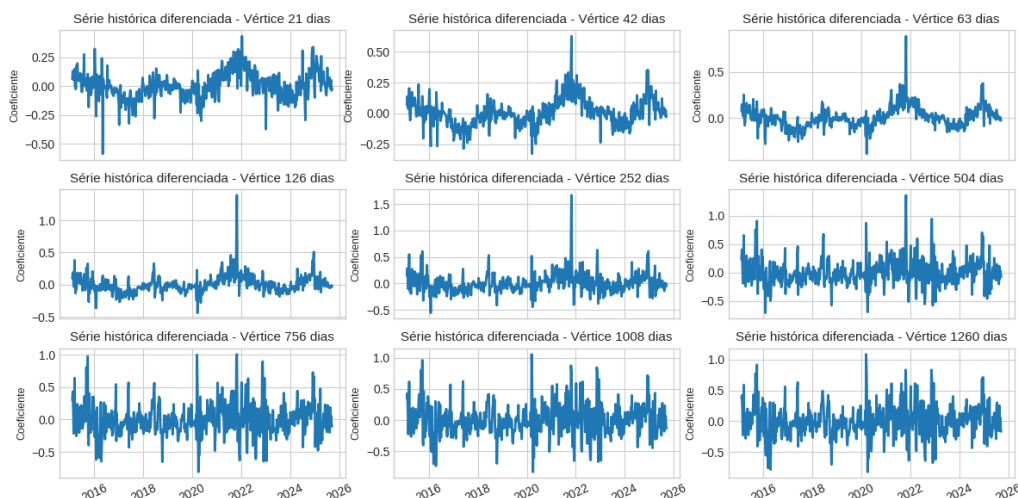


Figura 3.3: Série histórica semanal das taxas de juros prefixadas diferenciadas, por vértice

Ante o exposto, optou-se por trabalhar com médias semanais para reduzir ruído de alta frequência e tornar o sistema computacionalmente viável, preservando a dinâmica relevante das séries temporais para sinais de alerta. Ressalta-se que, semanas com feriados nacionais foram tratadas naturalmente pelo procedimento de reamostragem, agregando apenas os dias úteis disponíveis, sem introdução de viés temporal. Por fim, após a agregação semanal, o modelo passou a trabalhar com primeiras diferenças, possibilitando que os alertas captassem mudanças anormais na inclinação da curva em vez de desvios de nível, o que reduziu significativamente a instabilidade dos erros.

3.3 Definição do Modelo

Esta seção descreve a metodologia completa de desenvolvimento do modelo utilizado para a construção do sistema de alerta antecipado (*Early Warning System – EWS*) com o uso de redes neurais do tipo LSTM, aplicado à dinâmica da curva de juros prefixada em múltiplas maturidades. Toda a construção do modelo foi realizada em Python e, para a sua arquitetura foi usada a biblioteca *NeuralForecast* (mais detalhes em Olivares et al., 2022), utilizada como referência em Hyndman et al., 2025.

Conforme definido na seção anterior, as séries originais diárias foram agregadas para a periodicidade semanal tal que, $y_{i,t}$ = média semanal da série i na semana t . Na sequência, as séries são ordenadas temporalmente e transformadas em primeiras diferenças:

$$\Delta y_{i,t} = y_{i,t} - y_{i,t-1} \quad (3.1)$$

onde, i indexa a maturidade e t o tempo. Como é usual nesse tipo de operação, as observações iniciais sem diferença definida são descartadas. Essa transformação é fundamental para reduzir não estacionariedade e focar a modelagem nas variações da curva.

A partir dessa etapa, o método gerado combina previsão fora da amostra em diferenças, definição de um desvio operacional *ex-ante*, construção de *threshold* dinâmico e agregação temporal dos sinais de alerta. As subseções a seguir descrevem as etapas do processo de construção do modelo.

3.3.1 Validação Temporal

A avaliação do modelo é conduzida por meio de uma validação temporal do tipo janela expandida (*expanding window*), assim como em Jacomo Seraphim Nogueira, 2026. Para a construção das janelas definimos um conjunto de datas de corte (*cutoffs*) tal que cada *cutoff* possui no mínimo $T_{min} = 52$ observações semanais, aproximadamente um ano, e há informação suficiente para projeções até um horizonte $H = 4$ semanas à frente, cerca de um mês.

Formalmente, podemos definir que, para cada *cutoff* (c), o conjunto de treino é dado por:

$$T_c = (i, t, \Delta y_{i,t}) : t \leq c. \quad (3.2)$$

O procedimento descrito garante que, para cada iteração criada, o modelo utilize exclusivamente a informação disponível até o instante c , evitando que informação futura seja transmitida naquela janela de dados. Uma vez concluída todas as etapas explicitadas nas próximas subseções, a janela adiciona 1 semana e realiza o mesmo procedimento novamente. De tal modo,

com o passar do tempo maiores são os períodos disponíveis para testar o modelo.

3.3.2 Previsão com LSTM

O modelo base utilizado neste trabalho é o LSTM para múltiplas variáveis, com o uso da biblioteca desenvolvida em Olivares et al., 2022. Para a construção da arquitetura, foram utilizadas as seguintes informações: histórico de entrada de 52 semanas, horizonte de previsão $H = 4$ semanas, camadas ocultas com 32 unidades; normalização dos dados do tipo min-max; e número limitado de 20 iterações para evitar sobreajuste (*overfitting*). A Tabela 3.2 resume as informações da arquitetura utilizada, após uma vasta gama de testes e simulações com diferentes hiperparâmetros. Os demais parâmetros do modelo usados nesta metodologia são definidos pelo padrão da biblioteca *NeuralForecast*.

Tabela 3.2: Hiperparâmetros do modelo LSTM

Hiperparâmetro	Valor selecionado	Descrição
<i>input_size</i>	52	Comprimento máximo da sequência para retropropagação durante o treinamento
<i>scaler_type</i>	min-max	Tipo de escalonamento utilizado para normalização das entradas temporais
<i>encoder_hidden_size</i>	32	Número de unidades do estado oculto do LSTM
<i>decoder_hidden_size</i>	32	Número de unidades da camada oculta do decodificador MLP ^a
<i>encoder_n_layers</i>	1	Número de camadas do codificador LSTM
<i>decoder_layers</i>	1	Número de camadas do decodificador MLP
<i>max_steps</i>	20	Número máximo de etapas de treinamento
<i>random_seed</i>	10	Semente para garantir a reprodutibilidade

^a Arquitetura baseada em um *Simple Multi-Layer Perceptron (MLP)*.

Conforme mencionado anteriormente, este procedimento empírico adota uma estratégia de divisão temporal com janelas expandidas para a estimação dos modelos. Para cada *cutoff* T_c , o conjunto de treino incorpora todas as observações disponíveis até aquele momento, caracterizando uma abordagem do tipo janela expandida no nível dos dados.

No entanto, a arquitetura LSTM empregada opera sobre sequências temporais de dimensão fixa, de modo que cada instância de treino é formada por subsequências compostas pelas 52 semanas imediatamente anteriores ao ponto de previsão. Assim, embora a base utilizada para estimação cresça cumulativamente ao longo do tempo, o modelo possui memória temporal finita

em cada sequência, sendo treinado a partir de múltiplas janelas internas deslizantes extraídas do histórico completo disponível.

Essa combinação entre aprendizado cumulativo e memória local permite ao modelo incorporar informação histórica de longo prazo por meio do aumento progressivo do conjunto de treino, ao mesmo tempo em que preserva sensibilidade a padrões recentes. De tal modo, acredita-se que a arquitetura é particularmente adequada à construção de um sistema de alerta antecipado, uma vez que os desvios sinalizados refletem alterações estatisticamente relevantes em relação aos padrões históricos aprendidos até cada data de referência. Em seguida, o modelo gera previsões condicionais para cada maturidade e horizonte fora da amostra de treino no instante c .

$$\Delta \hat{y}_{i,c+h}, h = 1, \dots, H \quad (3.3)$$

3.3.3 Benchmark Operacional

Para definir uma métrica de comparação para os valores projetados, não foram utilizados os dados realizados para avaliar o erro de previsão, diferentemente de abordagens tradicionais *ex-post*. Esse ajuste metodológico é de fundamental importância para que o sistema simule uma situação real de alerta antecipado, ou seja, onde não se tem o valor realizado.

Alternativamente, foi definido um *benchmark* operacional *ex-ante*, elaborado a partir do comportamento recente (últimas 4 semanas) da própria série. Assim, para cada maturidade estudada, as projeções das variações são comparadas à mediana das variações observadas nas últimas 4 semanas, utilizada como referência comum para todos os horizontes de previsão em cada *cutoff*.

A escolha da mediana é baseada no fato de ser uma medida de centralidade e não influenciada por valores extremos, conferindo robustez a choques pontuais. Já a escolha do período de 4 semanas, teve como critério adotar o período recente de, aproximadamente, 1 mês para avaliar o distanciamento em relação ao histórico recente da curva prefixada.

Para cada maturidade i e *cutoff* c , foram calculados:

$$\Delta y_{i,c}^{ref} = \text{mediana}(\Delta y_{i,t} : t \in [c - W + 1, c]), \quad (3.4)$$

onde, W são as semanas utilizadas ($W = 4$, neste caso). Logo, o *benchmark* representa uma estimativa do comportamento recente da variação da série e, portanto, uma referência robusta para comparação dos valores estimados pelo modelo. Vale ressaltar que essa medida desenvolvida não deve ser interpretada como erro de previsão no sentido clássico, mas como um desvio em relação ao regime recente da série.

3.3.4 Cálculo do Desvio

Uma vez definido o ponto de comparação o sistema de alerta tem como base a discrepância entre as projeções do modelo e o *benchmark* recente para aquela janela de tempo.

$$e_{i,c,h} = \Delta \hat{y}_{i,c+h} - \Delta y_{i,c}^{ref} \quad (3.5)$$

O desvio apresentado na Equação 3.5 possui três características fundamentais, sendo elas: é ex-ante, pois depende apenas de informação disponível até o *cutoff*; é comparável ao longo do tempo, pois se ancora em um comportamento recente da série; e mede desvios anômalos da previsão em relação ao regime atual. Logo, cada desvio calculado é indexado à data c e ao horizonte de previsão correspondente $c + h$.

3.3.5 Construção do *Threshold* Dinâmico

Na sequência, foi definido o *threshold* (valor limite) que serve como base para a identificação de valores anômalos do estudo. Os desvios operacionais descritos na subseção anterior são acumulados ao longo das janelas de tempo criadas, formando um histórico por maturidade e horizonte de tempo, ou seja, para cada par (i, h) é definido uma janela móvel com tamanho $K_t = 52$. Posteriormente, o *threshold* é calculado como o percentil 95 da distribuição do valor absoluto dos desvios, tal que:

$$\theta_{i,h,c} = Q_{0.95}(|e_{i,t,h}| : t \in [c - K_t + 1, c]). \quad (3.6)$$

A opção por utilizar os desvios absolutos tem com fundamentação garantir a simetria na detecção de desvios positivos e negativos, evitando vieses direcionais. Com isso, buscou-se um *threshold* que se adapte aos diferentes regimes de volatilidade de cada maturidade e horizonte ao longo da análise realizada, porém, preservando o caráter *ex-ante* do sistema. De tal modo, diferentemente de abordagens usuais que adotam limites estáticos definidos em um único momento do tempo, o procedimento proposto permite que o *threshold* se ajuste dinamicamente à evolução da série.

3.3.6 Geração de Alerta

Por fim, uma vez desenvolvida a metodologia para as previsões do modelo, foi criada a regra base para o sistema de alerta identificar um padrão fora do comportamento esperado. Para cada maturidade i , *cutoff* c e horizonte de tempo h , o alerta é acionado sempre que o desvio absoluto projetado ultrapassar o *threshold* de sua correspondência:

$$\mathbf{1}_{i,c,h} = \begin{cases} 1, & \text{se } |e_{i,c,h}| > \Theta_{i,c,h}, \\ 0, & \text{caso contrário.} \end{cases} \quad (3.7)$$

Cabe salientar que a escolha de um *threshold* elevado (95º percentil) implica que somente os desvios suficientemente relevantes comparativamente ao histórico recente da série são sinalizados. Esses alertas constituem a base para etapas posteriores de agregação transversal (entre maturidades) e temporal (entre horizontes), formando o indicador final de EWS apresentado nos resultados.

Capítulo 4

Resultados

4.1 Resultados das Projeções

Nesta seção, são apresentados os principais resultados obtidos com a metodologia explicitada anteriormente. Apesar do sistema de alerta operar em diferença, para a exibição dos resultados, as previsões são reconstruídas em nível auxiliando a interpretação econômica e visualização.

Sendo $y_{i,c}$ o último nível observado no *cutoff* analisado, a reconstrução segue a formulação dada por:

$$\hat{y}_{i,c+h} = y_{i,c} + \sum_{j=1}^h \Delta y_{i,c+j}, \quad (4.1)$$

onde, i representa cada uma das maturidades, c o *cutoff* e h o horizonte de tempo projetado. Essa etapa permite visualizar, de forma mais clara e ampla, o impacto na curva de juros prefixada.

A Figura 4.1 apresenta as projeções para cada uma das 9 maturidades analisadas (21 dias, 42 dias, 63 dias, 126 dias, 252 dias, 504 dias, 756 dias, 1008 dias e 1260 dias). Como a metodologia adotada contempla o avanço da janela em um período à frente, uma mesma semana futura pode ser projetada a partir de diferentes *cutoffs*, fazendo com que a mesma data prevista apareça múltiplas vezes na exibição dos resultados. Para facilitar a visualização gráfica deste conjunto

de informações, as cores dos valores projetados foram suavizadas (em laranja) de modo que não sobrepujassem os valores do último valor observado no momento da previsão (em azul).

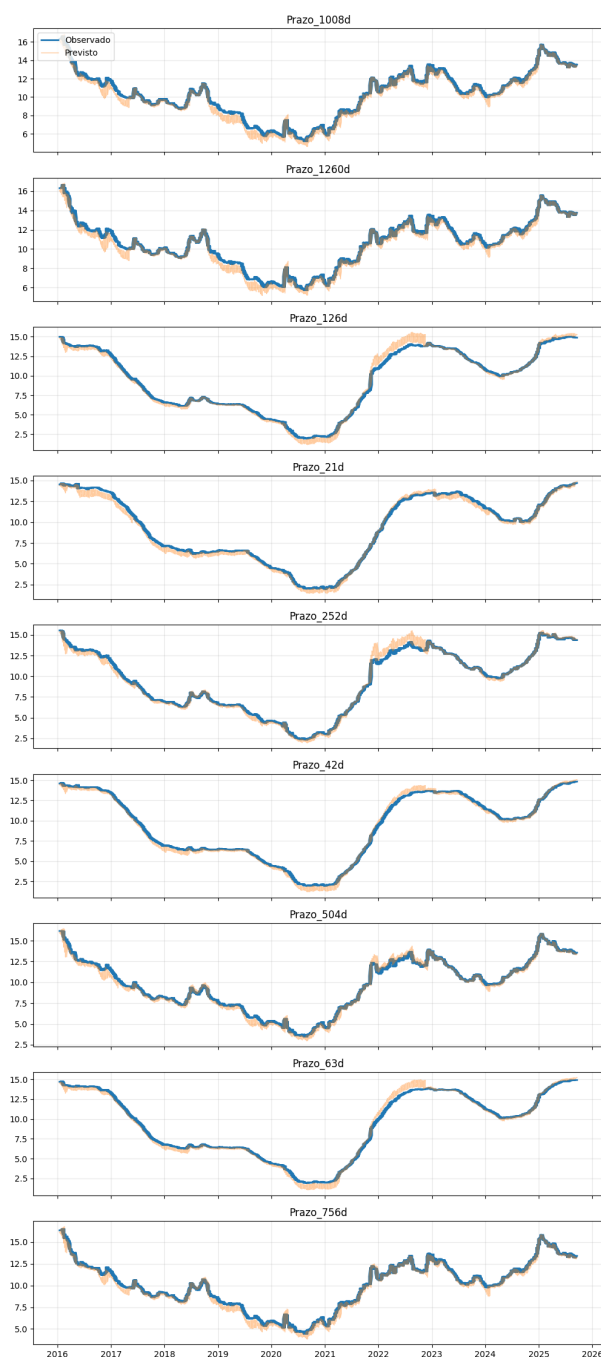


Figura 4.1: Série projetada das taxas de juros prefixadas, por vértice

Pela apresentação dos resultados, é possível observar que, durante boa parte do período

analisado, as projeções reconstruídas permanecem próximas do último valor observado no momento da previsão, corroborando a ideia abordada na parte metodológica de que a transformação em média semanal e a diferenciação dos valores não causam prejuízo na trajetória reconstruída da série. Além disso, o modelo LSTM demonstra consistência ao gerar projeções coerentes entre diferentes maturidades, refletindo padrões comuns aprendidos a partir da estrutura temporal da curva.

Ademais, os gráficos evidenciam períodos ao longo do tempo onde as projeções ficaram consistentemente abaixo ou acima do valor de referência, o que fornece evidência visual da capacidade do modelo em destacar desvios comparativamente ao comportamento recente dos vértices.

Na sequência, analisamos os desvios absolutos médios das projeções para a curva agregada, ou seja, são visualizados os desvios em conjunto das maturidades para o período de previsões estudado. A Figura 4.2 demonstra esse comportamento.

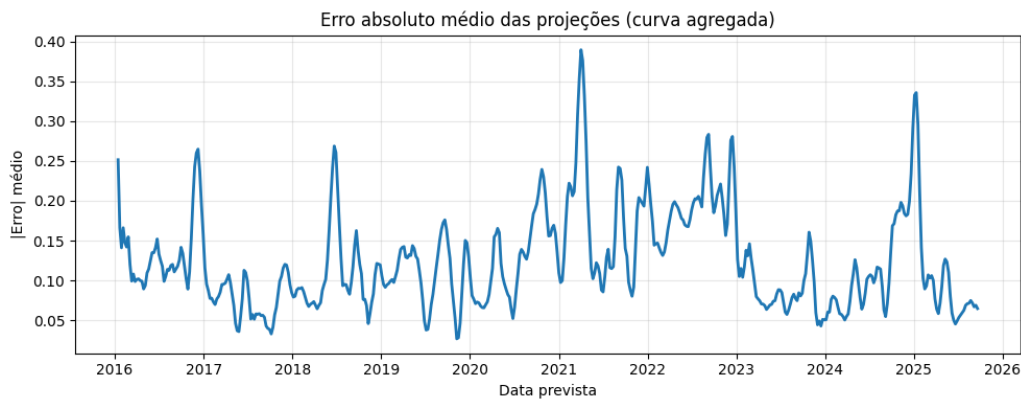


Figura 4.2: Desvios absolutos médios das projeções para a curva agregada

Como é possível observar, o desvio médio absoluto possui alguns saltos muito claros, indicando um distanciamento médio relevante em alguns períodos, como por exemplo: i) no final de 2016; ii) entre 2018 e 2019; iii) no início de 2021; e iv) no início de 2025. Esses e outros períodos serão analisados com maiores detalhes na sequência.

4.2 Resultado do Sistema de Alerta

Para avaliar os resultados do sistema de alerta de forma integrada, uma nova etapa foi acrescentada à análise, que consiste em calcular a proporção de maturidades que ultrapassam o seu *threshold* para uma mesma data prevista. O ponto crucial que se deseja capturar é de que períodos com estresse sistêmico tendem a afetar de forma simultânea diferentes segmentos da curva de juros, de modo que, um indicador agregado possui maior capacidade em perceber esses eventos conjuntos.

Tendo em vista que cada semana é projetada até 4 vezes à frente (em $H = 1, 2, 3$ e 4) e para as 9 maturidades, concluímos que a mesma semana pode ser projetada um total de até 36 vezes considerando todos os vértices. Em cada uma das projeções é verificado o desvio e o *threshold* do período. Caso o desvio absoluto, por maturidade, seja superior ao *threshold* considerado, é atribuído o valor 1, caso contrário o valor 0 é assumido.

Na sequência, é criado um indicador que avalia a proporção de maturidades que, em ao menos um dos horizontes projetados, superou o seu *threshold*, tal que:

$$alerta_t = \frac{1}{N} \sum_{i=1}^N \max_h alerta_{i,h,t}, \quad (4.2)$$

onde, N é o total de maturidades, h o horizonte projetado e t o período em questão.

Por fim, é estabelecida a proporção de 50%, como limite para que seja acionado o sinal de alerta, ou seja, se para uma semana específica, mais da metade das maturidades apresentaram, em alguma de suas projeções, um comportamento não esperado, então esse período é sinalizado. Logo, quanto mais próximo de 1 maior é a evidência de que o período seja identificado como anômalo, enquanto a aproximação de 0 indica um comportamento dentro dos padrões aprendidos.

Em resumo, os sinais individuais de alerta, estimados por maturidade e horizonte de previsão, são agregados no tempo por meio da proporção de maturidades da curva que apresentam comportamento anômalo. Posteriormente, é definido o indicador agregado de *Early Warning*

System (EWS) que destaca quando essa proporção ultrapassa um determinado limiar, buscando capturar episódios de estresse sistêmico na estrutura a termo da taxa de juros. A Figura 4.3 exibe os resultados pelo indicador desenvolvido, sinalizando os períodos que ultrapassam o limite estabelecido.

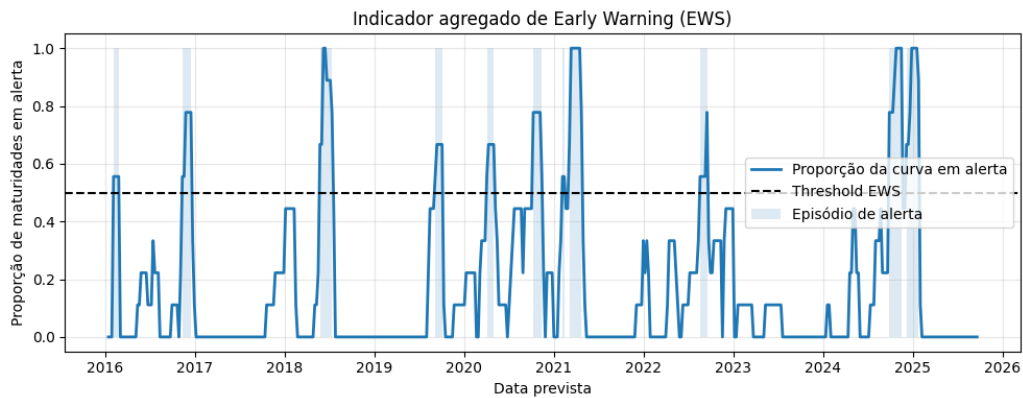


Figura 4.3: Resultado do indicador EWS acima do limite

Por sua vez, a Figura 4.4 demonstra o indicador de alerta associado ao desvio médio agregado visualizado anteriormente na Figura 4.2. Essa combinação permite corroborar com a ideia de que os picos observados no gráfico do desvio, representam períodos de alerta. Ademais, ilustra o conceito fundamental de um sistema de alerta que é ser informativo caso o alerta anteceda mudanças relevantes no comportamento da curva.

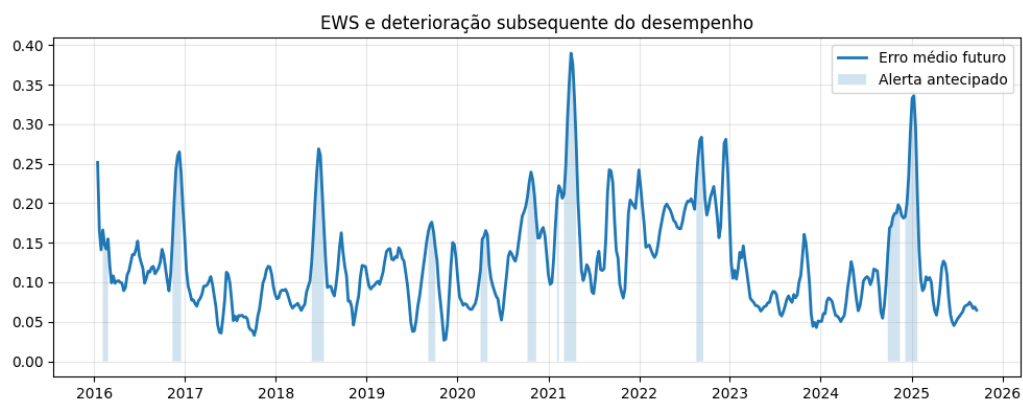


Figura 4.4: Resultado do indicador EWS associado ao desvio médio

4.3 Análise Descritiva dos Períodos de Alerta

Nesta seção são avaliados os períodos, em semanas, que apresentaram a sinalização de alerta. O objetivo principal dessa análise é de que os episódios de estresse são caracterizados não apenas pela ativação do alerta, mas por sua persistência ao longo do tempo.

Tabela 4.1: Períodos analisados e durações (em semanas)

Início	Fim	Duração
05/02/2016	26/02/2016	4
11/11/2016	16/12/2016	6
25/05/2018	13/07/2018	8
06/09/2019	04/10/2019	5
03/04/2020	01/05/2020	5
09/10/2020	13/11/2020	6
05/02/2021	12/02/2021	2
05/03/2021	23/04/2021	8
19/08/2022	16/09/2022	5
27/09/2024	15/11/2024	8
06/12/2024	24/01/2025	8

Conforme é observado na Tabela 4.1, 11 períodos foram selecionados durante toda a análise, sendo o menor de apenas 2 semanas, enquanto os maiores foram de 8 semanas. Os períodos sinalizados demonstram que há uma persistência temporal na identificação de alertas, de modo que, não são caracterizados como ruídos que tendem a acontecer de forma isolada.

Sob a ótica do impacto das maturidades individualmente, avalia-se nos períodos em que o EWS foi acionado aqueles vértices que tiveram uma contribuição mais significativa ao longo do tempo, ou seja, dado que houve alerta sistêmico em 65 períodos, com que frequência cada maturidade estava em alerta. A Tabela 4.2 apresenta a proporção de semanas em que cada vértice foi acionado, condicional à ocorrência do EWS.

O vértice de 504 dias úteis (2 anos) foi o que esteve mais presente nos períodos de alerta, com um total de 89,23% das vezes, seguido pelo vértice de 756 dias úteis (3 anos) presente em 87,69% dos períodos. Por outro lado, o prazo de 21 dias úteis (1 mês) esteve sinalizado em

metade dos eventos, com um total de 50,77% sendo o de menor presença dentre as 9 maturidades, o período de 1260 dias úteis (5 anos) foi o segundo com menor frequência aparecendo em 69,23%.

Tabela 4.2: Proporção dos vértices acionados no EWS

Maturidade	Proporção
Prazo 21d	50,77%
Prazo 42d	72,31%
Prazo 63d	80,00%
Prazo 126d	83,08%
Prazo 252d	84,62%
Prazo 504d	89,23%
Prazo 756d	87,69%
Prazo 1008d	73,85%
Prazo 1260d	69,23%

Da tabela apresentada, infere-se que as maturidades intermediárias exibiram uma presença maior nos episódios de alerta, sugerindo que os períodos de estresse sinalizados são impactados de forma mais consistente nesse trecho da curva. A Figura 4.5 ilustra a presença dos vértices nos episódios de alerta, ordenados de forma decrescente.

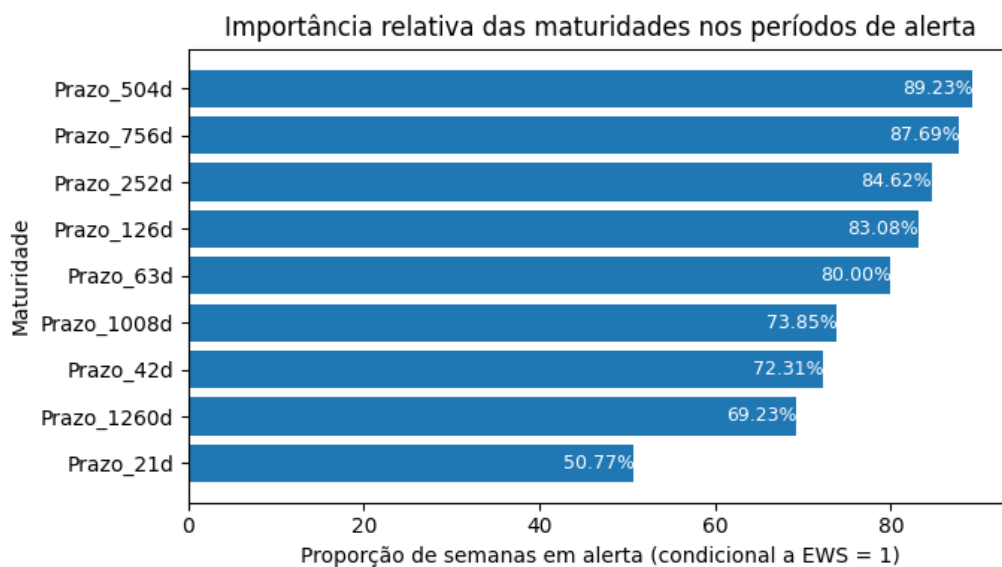


Figura 4.5: Proporção dos vértices acionados no EWS

Os resultados visualizados anteriormente motivam uma análise agregando os vértices por segmento, de modo que seja identificado quais trechos da curva, em conjunto, impactaram de forma mais significativa o sistema antecipado de alertas. As maturidades foram agrupadas em 3 segmentos da curva, cada um deles contendo 3 vértices, sendo: curto prazo (21 dias, 42 dias e 63 dias), médio prazo (126 dias, 252 dias e 504 dias) e longo prazo (756 dias, 1008 dias e 1260 dias). De forma análoga à realizada anteriormente, a Tabela 4.3 exibe a frequência relativa dos segmentos criados nos períodos em que os alertas foram identificados.

Tabela 4.3: Frequência relativa dos segmentos da curva no EWS

Segmento	Frequência Média
Curto Prazo	67,69%
Médio Prazo	85,64%
Longo Prazo	76,92%

De tal forma, nas semanas sinalizadas pelo EWS, em média, 67,69% das maturidades de curto prazo estavam em alerta, enquanto as de médio e longo prazo estavam em 85,64% e 76,92% das vezes, respectivamente. Os dados demonstram que os episódios de alerta se manifestam com maior intensidade no segmento com as maturidades de médio prazo, corroborando a ideia apresentada anteriormente quando avaliou-se os pontos da curva isolados. Paralelamente, avalia-se o comportamento das maturidades dentro de cada segmento com o objetivo de identificar se, nos períodos de alerta, os vértices dentro de um mesmo segmento reagem de forma similar ou não. A Tabela 4.4 demonstra o número de vértices, por segmento, presentes nos períodos em que o EWS foi acionado.

Tabela 4.4: Número de vértices presentes no EWS, por segmento

Nº de Vértices	Curto Prazo	Médio Prazo	Longo Prazo
Nenhum vértice	9	0	7
1 vértice	7	4	8
2 vértices	22	20	8
3 vértices	27	41	42

Conforme observado, no curto prazo os vértices possuem um comportamento mais heterogêneo, de modo que em 9 períodos o alerta foi acionado sem que nenhuma das maturidades tenha participado. Por outro lado, os segmentos de médio e longo prazo aparentam um comportamento mais similar entre os pontos da curva que compõem cada segmento, isso pode ser identificado a medida em que, quando houve a presença do alerta, sempre ocorreu a participação de pelo menos 1 vértice de médio prazo na sinalização. Mais ainda, em 41 períodos (aproximadamente 63% dos casos) os 3 vértices estavam em alerta evidenciando que além de serem significativos para o EWS os vértices intermediários tendem a ter um comportamento mais homogêneo. Situação similar é observada no segmento de longo prazo, porém, ressalta-se que nos horizontes mais longos, em 7 oportunidades o alerta foi acionado sem que houvesse a presença de algum dos seus componentes.

4.4 Mudanças da Dinâmica da Curva de Juros

Nesta seção, o objetivo é analisar se os períodos sinalizados pelo sistema de alerta estão associados a mudanças de regime na dinâmica da curva de juros prefixada, ou seja, se os períodos destacados na Tabela 4.1 retratam intervalos de tempo com características típicas de estresse. Para essa análise são abordados dois fatores de suma importância no estudo de curva de juros, sendo eles a volatilidade e a inclinação da curva em dois grupos distintos, um com sinalização de alerta e o outro sem.

A volatilidade entre as maturidades é realizada calculando-se o desvio-padrão da curva de juros, em nível, em cada semana. Este procedimento tem o intuito de mensurar a dispersão da curva, atuando como uma *proxy* para uma possível incerteza estrutural, e foram obtidas a média e mediana de cada grupo.

Por sua vez, a inclinação da curva é analisada utilizando os dois extremos dos vértices estudados, sendo a parte curta representada pelo vértice de 21 dias (1 mês) e a parte longa pelo vértice de 1260 dias (5 anos). De forma similar ao cálculo da volatilidade, também foram obti-

das a média e mediana de cada grupo para comparação. Os resultados descritivos da volatilidade são exibidos na Tabela 4.5.

Tabela 4.5: Volatilidade da curva de juros, por grupo

EWS	Vol. média	Vol. mediana
0	0,9638	0,8578
1	1,3967	1,2823

Os dados demonstram que a volatilidade média para o grupo com a presença do alerta é, aproximadamente, 46% maior em comparação ao grupo que não possui o alerta. Analogamente, a volatilidade mediana também foi maior, em aproximadamente 49%, para os períodos que sinalizaram alerta. Esta interpretação pode ser melhor compreendida a partir da visualização gráfica de *boxplot* na Figura 4.6, que expõe as diferenças em quartis.

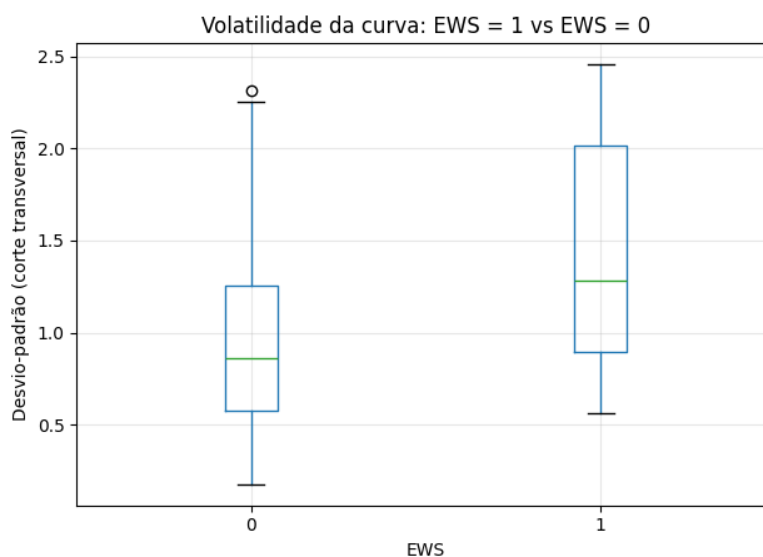


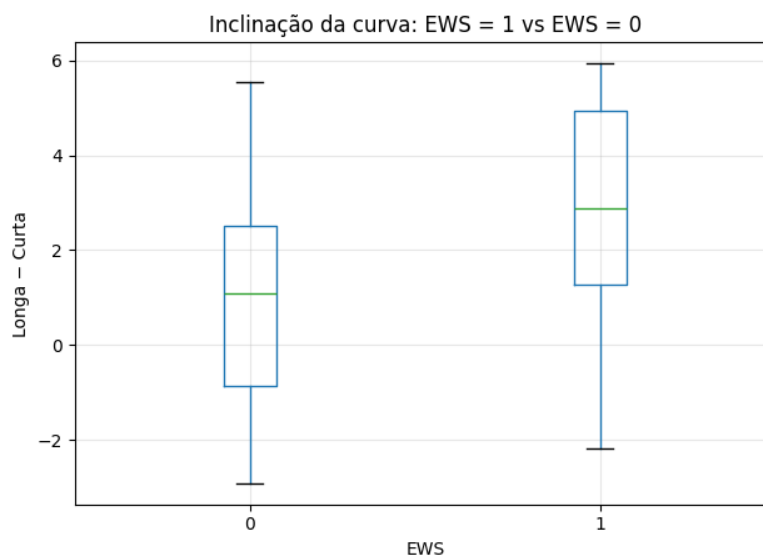
Figura 4.6: *Boxplot* da volatilidade da curva, por grupo

Os resultados indicam que durante os períodos sinalizados pelo EWS, a curva de juros perde consistência, com as maturidades reagindo de modo mais heterogêneo, comparativamente ao outro grupo. Na sequência, são expostas na Tabela 4.6 as estatísticas da inclinação da curva de juros, segmentadas pelos grupos.

Tabela 4.6: Inclinação da curva de juros, por grupo

EWS	Inclinação média	Inclinação mediana
0	1,0361	1,0788
1	2,6584	2,8734

Similarmente ao observado nas estatísticas da volatilidade, novamente o grupo com a sinalização do sistema de alerta obteve valores maiores que o grupo sem o alerta. Os valores demonstram um aumento de quase 156% na inclinação média e de 166% na inclinação mediana. A visualização gráfica na Figura 4.7 exibe essas variações em termos de quartis.

**Figura 4.7:** Boxplot da inclinação da curva, por grupo

Os resultados evidenciam um aumento na diferença entre as extremidades da curva de juros, sugerindo que ou ocorreu um aumento nos prêmios de risco de vértices mais longos ou uma queda dos prêmios mais curtos. Ambas as situações são características de momentos em que o mercado financeiro transita para um comportamento de estresse em relação ao observado em períodos de maior estabilidade.

Cabe ainda destacar que foram analisadas 506 observações semanais, onde o EWS desenvolveu sinalizou 65 semanas como períodos de alerta, enquanto 441 semanas não foram sinali-

zadas. A proporção de semanas identificadas é algo próximo a 13% e representa uma proporção satisfatória diante da complexidade dos dados.

Tabela 4.7: Frequência dos eventos, por grupo

EWS	Observações	Proporção
0	441	87,15%
1	65	12,85%

Ante o exposto, conclui-se que as semanas identificadas pelo sistema de alerta apresentam, em média, maior volatilidade na curva de juros e alterações sistemáticas na sua inclinação. Tais evidências sugerem que o indicador de EWS não captura flutuações aleatórias, mas está associado a períodos de estresse estrutural e reprecificação conjunta das maturidades da curva prefixada.

4.5 Análise Comparativa LSTM vs ARIMA

Esta seção traz uma análise entre os sinais de alerta antecipado obtidos via LSTM comparativamente com um modelo estatístico tradicional, no caso, o ARIMA (*AutoRegressive Integrated Moving Average*), com objetivo de avaliar a capacidade do sistema de alerta antecipado derivado de cada estrutura. Esse comparativo é de suma importância para o estudo, uma vez que permite visualizar as características dos resultados encontrados em ambos os modelos e, mais ainda, identificar se o sistema se mantém coerente com a utilização de outra arquitetura.

Para essa comparação, foi utilizado o modelo ARIMA com seleção automática dos parâmetros (AutoARIMA) da biblioteca *StatsForecast* desenvolvida por Azul Garza, 2022. Os parâmetros para cada estimação foram selecionados usando o critério de informação AICc (*Akaike Information Criterion*), adotado como padrão na biblioteca. Para a minimização do AICc é utilizado um algoritmo de busca eficiente das ordens do ARIMA, que ao invés de testar todas as combinações de valores, começa com modelos mais simples, ajusta gradualmente os parâmetros e, por fim, mantém apenas as que acrescentam melhoria relevante.

Cabe ressaltar que o restante da metodologia adotada foi a mesma para o desenvolvimento do alerta antecipado usando LSTM, e amplamente discutida no Capítulo 3. Logo, todos os parâmetros operacionais (horizonte de previsão, validação temporal, *benchmark* operacional, construção do desvio, definição do *threshold* dinâmico e geração de alerta) foram mantidos idênticos, sendo o único ponto de diferença o modelo que gera os valores previstos, $\hat{\Delta}y$. Com isso, as diferenças observadas decorrem, exclusivamente, da arquitetura proposta por cada modelo gerador das previsões.

Haja vista que o objetivo central desta seção é a comparação do sistema de alerta para cada um dos modelos, são exibidos, a seguir, os principais resultados do sistema de alerta antecipado usando o ARIMA, com as devidas comparações em relação ao modelo LSTM. A análise comparativa inicia-se com a Tabela 4.8 que apresenta os períodos, em semanas, que sinalizaram alerta a partir do modelo ARIMA.

Tabela 4.8: Períodos analisados e respectivas durações (em semanas), modelo ARIMA

Início	Fim	Duração
22/01/2016	11/03/2016	8
25/03/2016	25/03/2016	1
08/04/2016	15/04/2016	2
09/06/2017	09/06/2017	1
23/06/2017	07/07/2017	3
24/11/2017	24/11/2017	1
08/06/2018	20/07/2018	7
21/09/2018	21/09/2018	1
13/03/2020	20/03/2020	2
03/04/2020	22/05/2020	8
19/03/2021	19/03/2021	1
02/04/2021	23/04/2021	4
17/09/2021	17/09/2021	1
29/10/2021	29/10/2021	1
12/11/2021	10/12/2021	5
26/08/2022	26/08/2022	1
09/12/2022	09/12/2022	1
23/12/2022	30/12/2022	2
03/05/2024	03/05/2024	1
20/12/2024	31/01/2025	7

Observa-se, na Tabela 4.8, que um total de 20 períodos foi selecionado, sendo os menores de apenas 1 semana, enquanto os maiores chegam a 8 semanas. Dos eventos sinalizados, 13 apresentaram curta duração, representando 65% das ocorrências. Em comparação com os resultados obtidos pelo modelo LSTM (Tabela 4.1), identificou-se um total de 11 períodos, dos quais apenas 1 deles obteve sinalização de curta duração, com 2 semanas, não havendo nenhuma sinalização de 1 semana.

Logo, há evidências de que, em comparação ao modelo ARIMA, os períodos sinalizados pelo LSTM apresentam maior persistência temporal na identificação de alertas, capturando comportamentos de estresse que não se restringem a oscilações semanais. Por sua vez, o sistema baseado no ARIMA aparenta reagir de forma mais intensa a choques pontuais nos desvios de previsão, porém sem consolidá-los em regimes persistentes ao longo do tempo.

Essa primeira comparação evidencia diferenças entre o sistema de alerta antecipado baseado em redes neurais e aquele fundamentado em um modelo tradicional, uma vez que o LSTM tende a se organizar em blocos temporais mais longos, evitando sinais isolados e de difícil interpretação, o que resulta em menor ruído. Essa característica sugere que o LSTM identifica comportamentos mais persistentes, enquanto o ARIMA apresenta maior sensibilidade a choques transitórios de curto prazo.

Na sequência, discute-se a volatilidade e inclinação médias do ARIMA, com a avaliação comparativa entre as duas arquiteturas. A Tabela 4.9 exibe os resultados comparativos dos modelos, acrescentando ao resultado da Tabela 4.5 as informações do modelo ARIMA.

Tabela 4.9: Comparação das métricas de volatilidade entre os modelos LSTM e ARIMA

Modelo	EWS	Volatilidade Média	Volatilidade Mediana
LSTM	0	0.9638	0.8578
	1	1.3967	1.2823
ARIMA	0	0.9863	0.8732
	1	1.2750	1.2957

Nota-se que em ambos os casos o EWS se mostrou relevante, uma vez que há uma mu-

dança significativa no comportamento da curva quando há a sinalização do alerta ($EWS = 1$). Ao avaliar a volatilidade média, percebe-se que há um aumento de aproximadamente 46% no modelo LSTM e de 29% no ARIMA, indicando um melhor contraste entre momentos de normalidade e de estresse para o modelo LSTM.

A Figura 4.8 exibe a diferença de volatilidade, por grupo, no modelo ARIMA. A visualização gráfica, comparativamente com a Figura 4.6 ajuda a explicitar essa diferença de captura de momento por meio do *boxplot*, onde observa-se uma maior presença de eventos anômalos no ARIMA quando não há a sinalização de alerta.

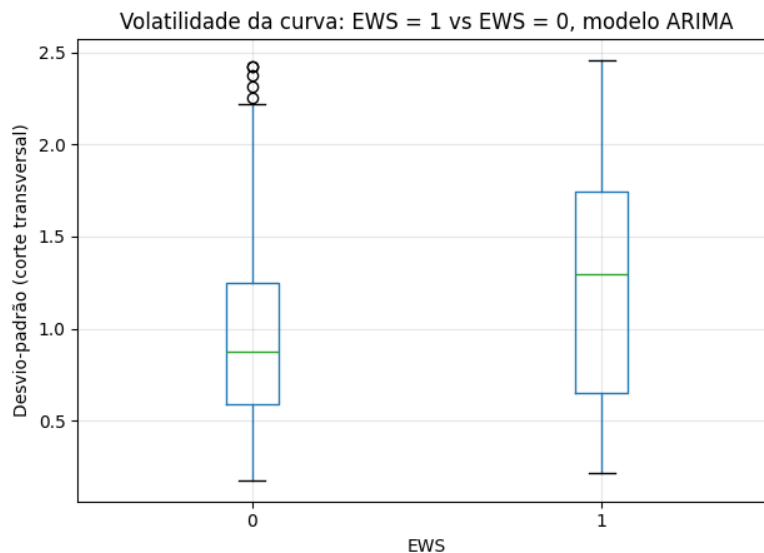


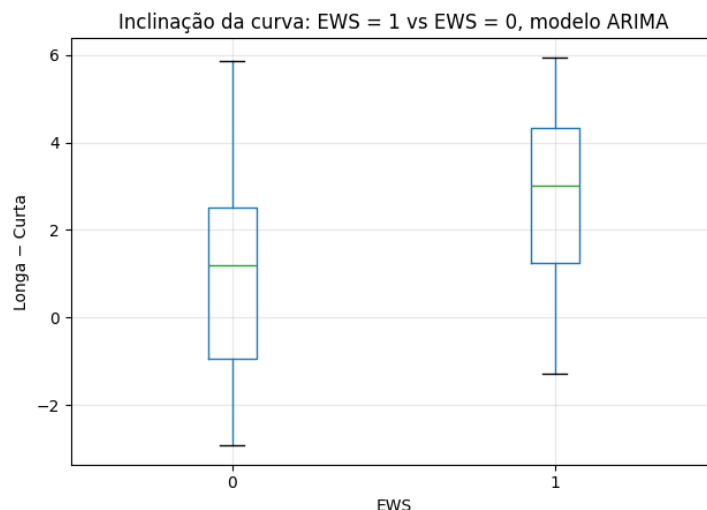
Figura 4.8: *Boxplot* da volatilidade da curva, por grupo, do modelo ARIMA

Em relação à inclinação da curva, a Tabela 4.10 apresenta os resultados dos dois modelos, complementando as informações da Tabela 4.6. Assim como no caso da volatilidade, os grupos com sinalização de alerta ($EWS = 1$) apresentam maior inclinação, refletindo uma diferença mais significativa de *spread* entre maturidades curtas e longas. Para o modelo LSTM, o aumento da inclinação é de aproximadamente 156%, enquanto, para o modelo ARIMA, é de 159%. O resultado levemente superior do ARIMA sugere que esse modelo tende a capturar, ligeiramente melhor, alguns pontos de maior oscilação.

Tabela 4.10: Comparação das métricas de inclinação entre os modelos LSTM e ARIMA

Modelo	EWS	Inclinação Média	Inclinação Mediana
LSTM	0	1.0360	1.0788
	1	2.6584	2.8734
ARIMA	0	1.0523	1.1960
	1	2.7284	3.0159

A Figura 4.9 apresenta, em complemento à Figura 4.7, a diferença gráfica de inclinação no modelo ARIMA.

**Figura 4.9:** Boxplot da inclinação da curva, por grupo, do modelo ARIMA

Em relação à frequência dos eventos, a Tabela 4.11 consolida as informações dos dois modelos. Como é possível visualizar, a proporção de períodos dos modelos com sinalização de alerta foi muito próxima, evidenciando que a comparação entre os modelos é apropriada.

Tabela 4.11: Frequência dos eventos entre os modelos LSTM e ARIMA

Modelo	EWS	Observações	Proporções
LSTM	0	441	87,15%
	1	65	12,85%
ARIMA	0	448	88,54%
	1	58	11,46%

Com os resultados obtidos, conclui-se que a principal diferença entre os modelos não está na quantidade de alertas encontrados, e sim, na forma como os alertas são consolidados ao longo do tempo. Os resultados indicam que, condicionados ao alerta, ambos os modelos identificam períodos associados a maior volatilidade e inclinação da curva de juros. Contudo, os alertas gerados pelo modelo LSTM apresentam maior contraste entre regimes, especialmente no que tange à volatilidade, sugerindo maior capacidade de discriminar períodos de estresse persistente, em comparação ao modelo ARIMA, cujos alertas parecem mais associados a choques transitórios.

4.6 Avaliação de Desempenho dos Modelos

Esta seção analisa o desempenho preditivo dos modelos em termos das métricas tradicionais de avaliação de desempenho, RMSE (*Root Mean Squared Error*, ou Raiz do Erro Quadrático Médio) e MAE (*Mean Absolute Error*, ou Erro Médio Absoluto). Embora a metodologia proposta tenha natureza *ex-ante*, a comparação com indicadores tradicionais *ex-post*, como RMSE e MAE, permite verificar se os alertas emergem justamente quando a capacidade dos modelos se deteriora, fornecendo evidência adicional que o EWS não é arbitrário.

Com essa finalidade, foram desenvolvidas duas análises complementares. A primeira consiste na avaliação condicional ao EWS do modelo LSTM, isto é, analisa-se o desempenho dos modelos nos períodos em que o sistema de alerta baseado no LSTM é acionado. Essa etapa tem por objetivo investigar se, quando o sistema baseado no LSTM identifica um regime de instabilidade, a previsibilidade da curva se deteriora para todos os modelos, inclusive para o ARIMA. Nesse contexto, o ARIMA atua como um modelo externo de validação.

A Tabela 4.12 apresenta o RMSE e o MAE calculados inicialmente para cada maturidade individualmente e, em seguida, agregados dentro de cada regime ($EWS = 0$ e $EWS = 1$). Dessa forma, a unidade estatística básica é a observação por maturidade em cada semana, permitindo capturar diferenças estruturais de volatilidade ao longo da estrutura a termo.

Tabela 4.12: RMSE e MAE agregado por maturidade, (EWS do LSTM)

Modelo	EWS	RMSE	MAE
ARIMA	0	0,3266	0,2103
	1	0,4804	0,3293
LSTM	0	0,4450	0,3144
	1	0,7306	0,5579

Por sua vez, a Tabela 4.13 apresenta as mesmas métricas, porém com agregação transversal prévia das maturidades em cada semana. Ou seja, para cada semana constrói-se inicialmente um erro médio da curva como um todo e, posteriormente, calcula-se o RMSE e o MAE condicionados ao regime de alerta. Assim, enquanto a primeira tabela enfatiza a dimensão transversal (maturidade por maturidade), a segunda enfatiza a dimensão temporal agregada da curva.

Tabela 4.13: RMSE e MAE agregado por semana, (EWS do LSTM)

Modelo	EWS	RMSE	MAE
ARIMA	0	0,2749	0,2101
	1	0,4275	0,3293
LSTM	0	0,3927	0,3139
	1	0,6846	0,5579

A Tabela 4.12 demonstra que o erro de previsão se eleva sistematicamente quando o EWS está ativo ($EWS = 1$). O aumento das métricas nos períodos sinalizados reforça que o sistema de alerta não está capturando ruído estatístico, mas sim intervalos nos quais a previsibilidade da curva se deteriora de forma generalizada, resultado também corroborado pelo modelo ARIMA. Observa-se ainda que o LSTM apresenta uma variação mais acentuada do que o ARIMA quando $EWS = 1$, indicando que o modelo de redes neurais é mais sensível a comportamentos instáveis e que os alertas tendem a surgir nos períodos em que as dinâmicas previamente aprendidas se tornam menos adequadas.

A Tabela 4.13 reforça essa interpretação sob a ótica da curva agregada. Mesmo após a agregação transversal das maturidades, os períodos classificados como alerta continuam associados

a maiores erros médios. Isso indica que a deterioração não é localizada em vértices específicos, mas reflete um regime mais amplo de instabilidade.

Na segunda análise, a avaliação é realizada de forma condicional ao EWS de cada modelo individualmente. Assim, é examinado se, quando o próprio modelo identifica um regime anômalo segundo seus próprios critérios, ocorre aumento das métricas de erro. As Tabelas 4.14 e 4.15 seguem a mesma lógica metodológica das anteriores, diferindo apenas quanto à definição do regime de alerta.

Tabela 4.14: RMSE e MAE agregado por maturidade, (EWS individual)

Modelo	EWS	RMSE	MAE
ARIMA	0	0,3122	0,2046
	1	0,5651	0,3903
LSTM	0	0,4450	0,3144
	1	0,7306	0,5579

Tabela 4.15: RMSE e MAE agregado por semana, (EWS individual)

Modelo	EWS	RMSE	MAE
ARIMA	0	0,2666	0,2042
	1	0,5099	0,3893
LSTM	0	0,3927	0,3139
	1	0,6846	0,5579

De forma análoga à primeira interpretação dos resultados, verifica-se aumento significativo das métricas quando há sinalização de alerta para ambos os modelos. A oscilação é particularmente expressiva na agregação por semana, sugerindo que os regimes identificados correspondem a mudanças sistêmicas no comportamento da curva. No caso do ARIMA, a variação mais acentuada é esperada, dado que o regime com EWS acionado é definido com base nos próprios critérios do modelo. Isso tende a produzir maior contraste entre períodos normais e anômalos.

Adicionalmente, para oferecer robustez a ideia apresentada de que o EWS separa regimes com dinâmicas estatisticamente distintas, calcula-se a diferença média das métricas entre os

períodos com e sem alerta, para cada modelo. Formalmente tem-se:

$$\begin{aligned}\Delta RMSE &= RMSE(EWS = 1) - RMSE(EWS = 0) \\ \Delta MAE &= MAE(EWS = 1) - MAE(EWS = 0)\end{aligned}\tag{4.3}$$

A Tabela 4.16 é construída com base nas informações obtidas nas Tabelas 4.12 e 4.14 demonstrando que, para ambos os modelos e sob as definições de regime, $\Delta RMSE$ e ΔMAE são positivos, tal que os alertas não são considerados aleatórios, mas sim associados a deterioração efetiva do desempenho preditivo.

Tabela 4.16: $\Delta RMSE$ e ΔMAE dos modelos

EWS	Modelo	$\Delta RMSE$	ΔMAE
EWS LSTM	ARIMA	0,1538	0,1190
	LSTM	0,2857	0,2436
EWS Individual	ARIMA	0,2529	0,1857
	LSTM	0,2857	0,2436

Comparando as magnitudes, observa-se que $\Delta RMSE$ e ΔMAE do LSTM são superiores aos do ARIMA tanto sob o EWS do LSTM quanto sob o EWS individual. Esse resultado sugere que os regimes identificados pelo LSTM apresentam maior contraste em termos de perda de capacidade preditiva. Embora o ARIMA apresente menores erros médios nas duas métricas avaliadas, o foco do estudo não é exclusivamente acurácia média, mas a identificação de rupturas na dinâmica da curva. Nesse sentido, o LSTM parece gerar um sistema de alerta mais informativo para fins de monitoramento, pois os regimes sinalizados estão associados a deteriorações mais intensas do desempenho.

Ademais, o fato de o ARIMA também apresentar aumento de erro quando condicionado ao EWS do LSTM reforça que o sistema de alerta não decorre de eventual má especificação do modelo de redes neurais, mas sim de mudanças estruturais na dinâmica da curva de juros.

4.7 Efeito *Lead-Lag* dos Modelos

A análise *lead-lag* é fundamental para avaliar a propriedade antecipatória do sistema de alerta, ao investigar se o desvio preditivo aumenta antes, no momento ou após o disparo do alerta. Esse tipo de avaliação é central para a interpretação de sistemas de alerta antecipado, pois permite distinguir entre um mecanismo genuinamente prospectivo e um sinal meramente reativo.

Conceitualmente, espera-se que um EWS bem calibrado apresente níveis relativamente baixos de erro no período pré-alerta, seguidos por crescimento progressivo das métricas à medida que se aproxima o momento de disparo, resultando em níveis elevados durante o episódio anômalo e eventual reversão gradual no período pós-alerta.

Conforme observado anteriormente, em particular na Tabela 4.1, o sistema desenvolvido identifica períodos persistentes de anomalia, caracterizados por semanas consecutivas com sinalização ($EWS = 1$). Assim, na análise *lead-lag*, as semanas não foram tratadas como eventos independentes, mas agrupadas em episódios contínuos de alerta. A interpretação passa, portanto, a considerar o comportamento médio dos desvios antes, durante e após cada episódio de alerta, independentemente de sua duração.

Essa estratégia evita contaminação estatística decorrente da sobreposição de informações, uma vez que semanas consecutivas sinalizadas poderiam inflar artificialmente o efeito pós-alerta. Ao trabalhar com episódios agregados, preserva-se a coerência temporal da transição entre regimes normal e anômalo.

Adicionalmente, como o sistema foi construído com horizonte de previsão de 4 semanas à frente, uma mesma semana pode ser projetada múltiplas vezes sob diferentes horizontes. Para ajustar este ponto na análise *lead-lag*, os horizontes foram agregados, de modo que a interpretação recaia sobre os desvios médios ao redor dos pontos de alerta. Com esses dois ajustes, agregação por episódio e agregação por horizonte, minimiza-se a distorção causada pela persistência do alerta e pela sobreposição de projeções. Dadas essas definições metodológicas, as Figuras 4.10 e 4.11 exibem o comportamento médio do RMSE e do MAE em janelas de 4

semanas antes e 4 semanas após o início de cada episódio de alerta.

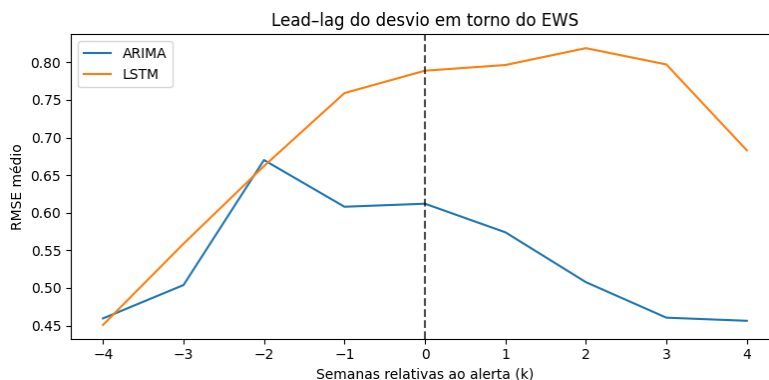


Figura 4.10: *Lead-lag dos desvios para RMSE médio*

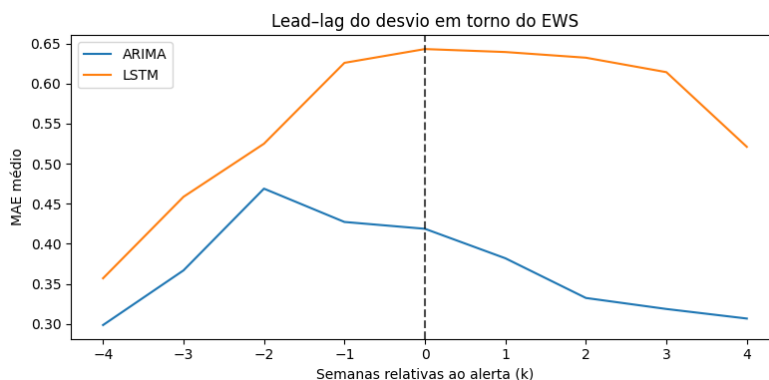


Figura 4.11: *Lead-lag dos desvios para MAE médio*

Observa-se, em ambos os modelos, crescimento sistemático das métricas no período pré-alerta, indicando que o disparo ocorre após acumulação gradual de deterioração preditiva. Esse padrão reforça a interpretação do EWS como indicador de mudança de regime, e não como resposta a choques isolados.

No caso do LSTM, a elevação no período pré-alerta é mais pronunciada, com aumento mais acentuado do RMSE e do MAE médios. Além disso, as métricas permanecem elevadas por um número considerável de semanas após o início do episódio, apresentando redução apenas a partir da terceira semana no período pós-alerta. Esse comportamento sugere que os episódios identificados pelo LSTM estão associados a regimes de deterioração mais persistentes.

O modelo ARIMA apresenta dinâmica distinta. O crescimento do desvio médio na fase pré-alerta é mais moderado, atingindo seu pico aproximadamente 2 semanas antes do início formal do episódio. Durante o período de alerta observa-se relativa estabilização, seguida por redução mais rápida no pós-alerta quando comparada ao LSTM. Esses resultados indicam que os episódios identificados pelo ARIMA tendem a estar associados a deteriorações mais transitórias.

Em conjunto, os resultados destacam diferenças relevantes na natureza dos regimes identificados pelos dois modelos. Enquanto o ARIMA sinaliza episódios associados a aumentos de erro de menor persistência, o LSTM identifica regimes nos quais a deterioração preditiva é mais intensa e prolongada. Assim, embora o LSTM apresente erros médios absolutos superiores aos do ARIMA, a dinâmica temporal observada sugere maior contraste entre regimes normal e anômalo, reforçando sua utilidade como instrumento de monitoramento de mudanças estruturais na dinâmica da curva de juros, corroborando as interpretações observadas anteriormente.

4.8 Análise de Sensibilidade dos Parâmetros

A avaliação de sensibilidade dos parâmetros selecionados na parte metodológica é crucial para analisar a robustez do sistema de alerta antecipado desenvolvido. Esta seção, que encerra a parte de resultados, tem por objetivo apresentar um comparativo entre diferentes combinações dos parâmetros e demonstrar que os resultados encontrados são estruturalmente estáveis, não dependendo de escolhas individuais.

Inicialmente, foram identificados 3 limiares considerados críticos, sendo eles: i) quantidade de semanas usadas como *benchmark*, e sua mediana (W); ii) percentil para o *threshold* (θ); e iii) percentual de maturidades utilizadas como gatilho para o acionamento do EWS. Diante desses 3 parâmetros, foram realizadas combinações de valores abaixo e acima das referências usadas no modelo deste trabalho, produzindo 27 cenários distintos. Para tanto, considerou-se 2, 4 e 8 semanas como referência para o *benchmark* e sua mediana, os percentis 90, 95 e 99 da distribuição dos desvios e, por fim, os gatilhos de 40%, 50% e 60% das maturidades em alerta

para uma determinada data.

Buscou-se, assim, comparar esse conjunto de combinações com os valores selecionados para a arquitetura base do estudo que, conforme descrito na seção metodológica, utilizou a mediana das últimas 4 semanas ($W = 4$), percentil 95 como *threshold* ($\theta = 0.95$) e gatilho de 50% dos vértices em cenário de estresse.

Em seguida, foram estruturadas as seguintes variáveis para controle da robustez dos parâmetros, sendo algumas delas já discutidas anteriormente: percentual de semanas sinalizadas pelo EWS, intensidade média dos alertas, duração média dos eventos, número de episódios em alerta e variação das métricas de erro RMSE e MAE ($\Delta RMSE$ e ΔMAE). Adicionalmente, foram calculados os valores de média e mediana dessas variáveis para fins de comparação com o modelo base. A tabela com os dados completos dos 27 cenários desenvolvidos pode ser consultada no Apêndice (Capítulo 6). Por sua vez, os resultados agregados (média e mediana), assim como os dados do modelo LSTM padrão, são exibidos na Tabela 4.17 a seguir.

Tabela 4.17: Média e mediana das variáveis de sensibilidade

	Taxa EWS	Intens. Média	Dur. Média	Nº Ep.	$\Delta RMSE$	ΔMAE
LSTM	12,85%	0,76	5,91	11	0,29	0,24
Média	15,70%	0,75	6,12	13,3	0,25	0,21
Mediana	13,44%	0,76	5,86	11	0,24	0,21

A variável *taxa EWS* representa o percentual de semanas sinalizadas como alerta dentro de uma amostra de 506 semanas. Nesse sentido, os 65 eventos sinalizados usando o modelo LSTM correspondem a 12,85% do total, enquanto a média entre os cenários foi de 15,70% e a mediana de 13,44%, indicando que o modelo não é excessivamente sensível, posicionando-se muito próximo da mediana dos cenários analisados.

A *intensidade média* é um indicador que varia entre 0 e 1, em que os valores mais próximos de 1 indicam que uma maior proporção de maturidades está simultaneamente em alerta quando o EWS é acionado. Os valores de média e mediana são novamente bastante próximos ao obtido com o LSTM base e indicam que, em média, aproximadamente 7 das 9 maturidades encontram-

se em alerta quando o EWS é ativado.

Conforme discutido nas seções anteriores, é esperado observar semanas consecutivas em estado de alerta, demonstrando que a sinalização não ocorre de forma pontual, mas de maneira persistente, uma característica desejável para EWS, pois reduz a possibilidade de ruídos espúrios. Nesse contexto, as variáveis de *duração média* e *número de episódios* permitem analisar a dinâmica temporal dos períodos sinalizados. Observa-se que a duração média do LSTM é de aproximadamente 6 semanas, valor levemente superior à mediana e ligeiramente inferior à média dos 27 cenários. O número de episódios, por sua vez, é exatamente a mediana (11 episódios) e é discretamente inferior à média. A proximidade dessas métricas em relação à mediana reforça a robustez estrutural do sistema, sugerindo que os episódios identificados não são artificialmente prolongados nem excessivamente curtos.

As variáveis $\Delta RMSE$ e ΔMAE , construídas com base na Equação 4.3, mensuram a diferença nas métricas de erro entre períodos com $EWS = 1$ e períodos com $EWS = 0$, permitindo avaliar o comportamento dos desvios em períodos anômalos. Os resultados indicam que, no modelo base, ambas as métricas apresentam valores superiores à média e à mediana dos demais cenários, ainda que em magnitudes moderadas. Esse comportamento sugere que os parâmetros selecionados tendem a concentrar os alertas em períodos caracterizados por desvios relativamente mais intensos, comparativamente às demais combinações testadas.

Uma vez analisado o comportamento do modelo base frente aos demais cenários, insta destacar os padrões estruturais decorrentes das variações paramétricas, expostas no Apêndice. No que se refere ao efeito percentil (0.90, 0.95 e 0.99) observa-se que percentis mais elevados reduzem a taxa de alertas e o número de episódios, ao mesmo tempo em que ampliam os diferenciais das métricas de erro. Tal resultado é consistente, pois um *threshold* mais elevado torna o critério de acionamento mais restritivo, provocando alertas menos frequentes e com valores mais extremos em relação ao diferencial de erro.

O percentual de vértices utilizado para acionar o EWS em uma determinada data foi outro parâmetro variado na análise de sensibilidade (40%, 50% e 60%). Os resultados evidenciam

que valores mais altos tornam o sistema mais rígido, reduzindo a taxa de alerta e o número de episódios. Em contrapartida, elevam a intensidade média e os diferenciais de erro, indicando que os episódios tendem a ser mais concentrados e intensos. Novamente, esse padrão é metodologicamente coerente para um EWS, pois critérios mais exigentes selecionam apenas situações de maior estresse sistêmico.

O último parâmetro avaliado em termos de sensibilidade foi o efeito da janela de *benchmark* (2, 4 e 8 semanas), onde janelas maiores de 8 semanas tendem a gerar menores variações da métrica RMSE, porém, com episódios mais longos. Em geral, os resultados sugerem que janelas mais curtas tendem a deixar o sistema mais reativo, enquanto janelas mais longas o tornam mais suavizados com diluição da intensidade dos desvios. Logo, entende-se que a escolha de 4 semanas é um ponto intermediário bem comportado entre reatividade e estabilidade.

Ante o exposto, conclui-se que o sistema apresenta boas propriedades estatísticas e que os parâmetros selecionados conferem robustez aos resultados obtidos. Ademais, a dinâmica dos episódios de alerta permanece estruturalmente estável sob múltiplas combinações de parâmetros, reforçando a consistência do sistema proposto.

Capítulo 5

Conclusão

O presente estudo investigou a identificação de sinais de alerta antecipado na estrutura a termo da curva de juros prefixada brasileira no período de 2015 a 2025, considerando 9 maturidades distintas e avaliando a capacidade de detectar, de forma prospectiva, mudanças de comportamento na dinâmica dos juros. Para esse fim, utilizou-se como modelo principal uma arquitetura de redes neurais do tipo LSTM, cuja estrutura permite capturar dependências temporais de curto e longo prazo. Como referência comparativa, empregou-se o modelo estatístico tradicional ARIMA, com o objetivo de contrastar desempenho preditivo e capacidade de detecção de períodos anômalos.

A contribuição central do trabalho reside no desenvolvimento de uma metodologia própria de sistema de alerta antecipado (EWS), construída a partir dos desvios das projeções dos modelos em horizonte de até 4 semanas à frente, combinada com janelas temporais expandidas para simular a incorporação progressiva de novas informações. Em vez de adotar exclusivamente a abordagem tradicional *ex-post* de comparação entre valores previstos e observados, foi implementado um *benchmark* operacional baseado na mediana das variações recentes, permitindo avaliar os desvios de forma dinâmica e adaptativa. A definição de *thresholds* por meio do percentil 95 da distribuição acumulada dos desvios absolutos conferiu ao sistema um critério estatístico consistente para o acionamento dos alertas.

Os resultados indicam que o modelo LSTM identificou episódios de alerta caracterizados por persistência temporal, sugerindo que os sinais capturados não se configuram como ruídos isolados, mas como regimes de deterioração mais prolongada da previsibilidade da curva. A análise por maturidade revelou maior participação dos vértices intermediários nos períodos de estresse, reforçando a relevância do segmento médio da curva na transição entre regimes.

A avaliação das propriedades econômicas dos episódios sinalizados demonstrou que as semanas com $EWS = 1$ estão associadas, em média, a maior volatilidade e a alterações mais acentuadas na inclinação da curva de juros, corroborando a interpretação de que o sistema identifica períodos de mudança estrutural e não apenas flutuações aleatórias.

Na comparação entre modelos, observou-se que o ARIMA apresentou menores métricas médias de erro (RMSE e MAE), porém seus episódios de alerta foram, em geral, mais curtos e associados a movimentos mais transitórios. O LSTM, por sua vez, apresentou maior contraste entre regimes normal e anômalo, tanto em termos de diferenciais de erro quanto na dinâmica temporal evidenciada pela análise *lead-lag*. Essa evidência sugere que, embora o modelo de redes neurais apresente maior erro médio, seus alertas estão vinculados a deteriorações preditivas mais persistentes, o que pode torná-lo particularmente informativo para fins de monitoramento de mudanças estruturais.

A análise de sensibilidade reforçou a robustez do sistema proposto, demonstrando que a dinâmica dos episódios de alerta permanece estruturalmente consistente sob diferentes combinações de parâmetros, incluindo variações no *benchmark*, no percentil do *threshold* e no percentual de maturidades exigidas para o acionamento do EWS.

Em síntese, o estudo evidencia que a aplicação de modelos com capacidade de memória temporal, como o LSTM, pode contribuir de forma relevante para a identificação antecipada de mudanças na dinâmica da curva de juros. Embora o modelo estatístico tradicional apresente desempenho preditivo médio superior, o sistema baseado em redes neurais demonstrou maior capacidade de distinguir regimes de estresse persistente.

Como extensões futuras, destaca-se a possibilidade de explorar outras arquiteturas de redes

neurais e suas particularidades, bem como de comparar o desempenho do sistema de alerta com modelos não lineares alternativos e de aplicar a metodologia a diferentes segmentos da estrutura a termo, como a curva indexada ao IPCA, ampliando a compreensão sobre a dinâmica dos regimes de juros no Brasil. Adicionalmente, pode-se utilizar dados de outros mercados financeiros que operam em horários distintos do brasileiro, como o asiático, a fim de investigar possíveis efeitos de contágio entre mercados e seus impactos sobre a capacidade de antecipação do sistema.

Capítulo 6

Apêndice

Tabela 6.1: Resultados dos cenários simulados para diferentes combinações de W , percentil e limiar de maioria

W	Percentil	Limiar	Taxa EWS	Intens. Média	Duração Média	Nº Episódios	Δ RMSE	Δ MAE
2	0,9	0,4	35,97%	0,71	6,50	28	0,21	0,16
2	0,9	0,5	28,85%	0,78	5,62	26	0,24	0,18
2	0,9	0,6	22,73%	0,84	5,00	23	0,22	0,17
2	0,95	0,4	24,31%	0,66	5,86	21	0,23	0,19
2	0,95	0,5	17,00%	0,76	4,53	19	0,27	0,22
2	0,95	0,6	12,85%	0,83	4,06	16	0,28	0,22
2	0,99	0,4	13,44%	0,61	4,86	14	0,31	0,24
2	0,99	0,5	7,51%	0,75	4,22	9	0,32	0,25
2	0,99	0,6	5,14%	0,84	4,33	6	0,37	0,29
4	0,9	0,4	31,62%	0,66	6,40	25	0,21	0,17
4	0,9	0,5	19,57%	0,79	7,07	14	0,23	0,19
4	0,9	0,6	15,22%	0,86	7,70	10	0,24	0,21
4	0,95	0,4	19,37%	0,66	7,00	14	0,22	0,19
4	0,95	0,5	12,85%	0,76	5,91	11	0,29	0,24
4	0,95	0,6	9,88%	0,83	5,56	9	0,28	0,24
4	0,99	0,4	10,47%	0,62	4,42	12	0,26	0,22
4	0,99	0,5	7,11%	0,70	4,50	8	0,33	0,29
4	0,99	0,6	4,55%	0,79	4,60	5	0,37	0,32
8	0,9	0,4	28,46%	0,66	6,00	24	0,13	0,12
8	0,9	0,5	17,98%	0,78	8,27	11	0,18	0,16
8	0,9	0,6	13,24%	0,86	9,57	7	0,19	0,17
8	0,95	0,4	15,61%	0,70	7,90	10	0,21	0,19
8	0,95	0,5	13,64%	0,74	8,63	8	0,22	0,20
8	0,95	0,6	9,68%	0,82	9,80	5	0,27	0,25
8	0,99	0,4	11,26%	0,70	6,33	9	0,20	0,20
8	0,99	0,5	8,89%	0,77	5,63	8	0,25	0,25
8	0,99	0,6	6,72%	0,84	4,86	7	0,29	0,28

Referências Bibliográficas

- Arantes, Breno De Oliveira, Portugal, Marcelo Savino e Caldeira, João Fróis (2014). *Previsão Da Estrutura A Termo Da Taxa De Juros Brasileira Usando Redes Neurais Artificiais*. Anais do XLI Encontro Nacional de Economia [Proceedings of the 41st Brazilian Economics Meeting] 128. ANPEC - Associação Nacional dos Centros de Pós-Graduação em Economia [Brazilian Association of Graduate Programs in Economics]. URL: <https://ideas.repec.org/p/anp/en2013/128.html>.
- Azul Garza Max Mergenthaler Canseco, Cristian Challú-Kin G. Olivares (2022). *StatsForecast: Lightning fast forecasting with statistical and econometric models*. PyCon Salt Lake City, Utah, US 2022. URL: <https://github.com/Nixtla/statsforecast>.
- Cajueiro, Daniel O., Divino, Jose A. e Tabak, Benjamin M. (2009). *Forecasting the Yield Curve for Brazil*. Working Papers Series 197. Central Bank of Brazil, Research Department. DOI: None. URL: <https://ideas.repec.org/p/bcb/wpaper/197.html>.
- Caldeira, Joao, Moura, Guilherme V e Savino Portugal, Marcelo (2010). “Efficient yield curve estimation and forecasting in Brazil”. *Revista Economia*, January/April.
- Campbell, John Y (1986). “Bond and stock returns in a simple exchange model”. *The Quarterly Journal of Economics* 101.4, pp. 785–803.
- Cox, John C, Ingersoll, Jonathan E, Ross, Stephen A et al. (1985). “A theory of the term structure of interest rates”. *Econometrica* 53.2, pp. 385–407.

- Diebold, Francis X. e Li, Canlin (fev. de 2006). “Forecasting the term structure of government bond yields”. *Journal of Econometrics* 130.2, pp. 337–364. ISSN: 0304-4076. DOI: 10.1016/j.jeconom.2005.03.005.
- Elman, Jeffrey L. (mar. de 1990). “Finding Structure in Time”. *Cognitive Science* 14.2, pp. 179–211. ISSN: 1551-6709. DOI: 10.1207/s15516709cog1402_1.
- Faria, Luis Giovanni (2024). “Aprendendo a curva de juros brasileira”. *Brazilian Review of Finance* 22.1, pp. 1–29.
- Fernandes, Marcelo, Nunes, Clemens e Reis, Yuri (abr. de 2021). “What Drives the Nominal Yield Curve in Brazil?” *Brazilian Review of Econometrics* 40.2, pp. 267–284. ISSN: 1980-2447. DOI: 10.12660/bre.v40n22020.79438.
- Filho, Adonias (nov. de 2017). “What Drives Long Term Real Interest Rates in Brazil?” *Brazilian Business Review* 14.6, pp. 624–635. ISSN: 1808-2386. DOI: 10.15728/bbr.2017.14.6.5.
- Filipovic, Damir, Pelger, Markus e Ye, Ye (2022). “Stripping the Discount Curve - a Robust Machine Learning Approach”. *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.4058150.
- Fischer, Thomas e Krauss, Christopher (out. de 2018). “Deep learning with long short-term memory networks for financial market predictions”. *European Journal of Operational Research* 270.2, pp. 654–669. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2017.11.054.
- Goodfellow, Ian, Bengio, Yoshua e Courville, Aaron (2016). *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press.
- Hochreiter, Sepp e Schmidhuber, Jürgen (nov. de 1997). “Long Short-Term Memory”. *Neural Computation* 9.8, pp. 1735–1780. ISSN: 1530-888X. DOI: 10.1162/neco.1997.9.8.1735.
- Hyndman, Rob J. et al. (2025). *Forecasting: Principles and Practice, the Pythonic Way*. Online edition. Accessed on 3 January 2026. Melbourne, Australia: OTexts. URL: <https://otexts.com/fpppy/>.

- Jacomo Seraphim Nogueira, Eduardo (2026). “Previsão da Despesa Primária do Governo Central: Uma Análise Comparativa entre Técnicas Estatísticas, Aprendizado de Máquina, Aprendizado Profundo e Combinação de Previsões”. *CADERNOS DE FINANÇAS PÚBLICAS* 26.01. URL: <https://publicacoes.tesouro.gov.br/index.php/cadernos/article/view/284>.
- Nelson, Charles e Siegel, Andrew (fev. de 1987). “Parsimonious Modeling of Yield Curves”. *The Journal of Business* 60, pp. 473–89. DOI: 10.1086/296409.
- Olah, Christopher (2015). *Understanding LSTM Networks*. URL: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- Olivares, Kin G. et al. (2022). *NeuralForecast: User friendly state-of-the-art neural forecasting models*. PyCon Salt Lake City, Utah, US 2022. URL: <https://github.com/Nixtla/neuralforecast>.
- Ouyang, Zi-sheng, Yang, Xi-te e Lai, Yongzeng (abr. de 2021). “Systemic financial risk early warning of financial market in China using Attention-LSTM model”. *The North American Journal of Economics and Finance* 56, p. 101383. ISSN: 1062-9408. DOI: 10.1016/j.najef.2021.101383.
- Paiva, Eduardo e Pereira, Fernando (mai. de 2022). “Deep Learning para Processamento de Linguagem Natural”. Em: *Tópicos Especiais em Sistemas de Informação: Minicursos SBSI 2022*. SBC, pp. 21–50. DOI: 10.5753/sbc.8690.0.2.
- Prince, Simon J.D. (2023). *Understanding Deep Learning*. The MIT Press. URL: <http://udlbook.com>.
- Rosa, Paulo Sérgio e Gartner, Ivan Ricardo (2018). “Financial distress em bancos brasileiros: um modelo de alerta antecipado”. *Revista Contabilidade & Finanças* 29, pp. 312–331.
- Sima, Daniel (2022). *Reconstruction-based Anomaly Detection on Financial Market Data*. <https://hdl.handle.net/10589/215104>. Accessed: 17 Sep. 2025.

- Suimon, Yoshiyuki et al. (abr. de 2020). “Autoencoder-Based Three-Factor Model for the Yield Curve of Japanese Government Bonds and a Trading Strategy”. *Journal of Risk and Financial Management* 13.4, p. 82. ISSN: 1911-8074. DOI: 10.3390/jrfm13040082.
- Svensson, Lars E.O. (set. de 1994). “Estimating and Interpreting Forward Interest Rates: Sweden 1992 - 1994”. DOI: 10.3386/w4871.
- Urolagin, Siddhaling, Sharma, Nikhil e Datta, Tapan Kumar (set. de 2021). “A combined architecture of multivariate LSTM with Mahalanobis and Z-Score transformations for oil price forecasting”. *Energy* 231, p. 120963. ISSN: 0360-5442. DOI: 10.1016/j.energy.2021.120963.
- Vasicek, Oldrich (1977). “An equilibrium characterization of the term structure”. *Journal of financial economics* 5.2, pp. 177–188.
- Yadav, Anita, Jha, C K e Sharan, Aditi (2020). “Optimizing LSTM for time series prediction in Indian stock market”. *Procedia Computer Science* 167, pp. 2091–2100. ISSN: 1877-0509. DOI: 10.1016/j.procs.2020.03.257.